An Information Visualization Framework for Intrusion Detection

Anita Komlodi, John R. Goodall, Wayne G. Lutters
Department of Information Systems, UMBC
1000 Hilltop Circle, Baltimore, MD 21250, USA
{ komlodi, jgood, lutters }@umbc.edu

ABSTRACT
This paper reports a framework for designing information visualization (IV) tools for monitoring and analysis activities. In this user study, the domain for these activities is network intrusion detection (ID). User-centered design methods have been widely used for many years, however, innovative IV displays are often developed with limited consideration of user needs in the context of real-life problems. While it can be argued that this is required to generate creative new solutions, the resulting tools often do not support actual users in their daily work. Several IV tools have been developed to support ID, but there is little evidence that these solutions address the needs of the users. We studied ID analysts' daily activities in order to understand their routine work practices and the need for designing IV tools. We developed a three-phase process model that frames corresponding requirements for IV tools. This model significantly extends the scope of contemporary IV for ID tools in novel ways.

Author Keywords
User studies, intrusion detection, information visualization, design implications, interaction design

ACM Classification Keywords
H.5.2. User Interfaces. User-Centered Design.

INTRODUCTION

As organizational dependence on information technology and network infrastructure increases, there is a correlated increase in the requirements for information assurance [6]. Even the best information security policies and prevention technologies will eventually fall to a determined attacker, which is why organizations rely on intrusion detection (ID) analysts [10]. These ID analysts monitor output from intrusion detection systems (IDS). They use IDS output in conjunction with other system, network, and firewall logs to keep abreast of system activity and potential attacks. These textual files can be enormous and quite complex, making manual review unfeasible, which often results in both undetected attacks and false alarms [2].

ID analysts monitor network activity using an IDS for evidence of actions that attempt to compromise the integrity, confidentiality, or availability of a computing or network resource [5]. The challenge of detecting network intrusions in a timely manner is one of both great difficulty and utmost importance. Finding specific evidence of attack activity in the enormous number of potentially relevant ID events presents an almost overwhelming task for a security analyst.

The most important source of information for analysts is the output of IDS's, which automatically identify potential attacks and produce descriptive alerts. Due to the complicated nature of detecting actual intrusions, most current IDS's place the burden of distinguishing an actual attack from a large set of false alerts on the ID analyst, resulting in a significant cognitive load. We believe that this load may be mitigated using information visualization (IV), which takes advantage of human perceptual abilities to amplify cognition. We conducted an exploration of the design space of IV for ID via a field study of practicing analysts that identified several design implications.

RELATED RESEARCH

Although IV seems like a natural choice for ID, until recently there has been little research into coupling the two technologies, and only a single, informal user study [11] into the work of ID analysts.

The limited number of efforts applying IV to the specific problem of ID usually lacked corresponding user studies to evaluate the need or effectiveness of the approach. Girardin and Brodbeck [4] and Fortier and Shombert [3] describe visualizations that use firewall log data to facilitate network profiling and log analysis. Erbacher, Walker, and Frincke [2] and Erbacher [1] describe a 2D glyph-based visual overview of a single host and a small network, respectively.

Nyarko, et al. [7] use input from existing IDS sensors and display that data using glyphs in 3D space. Solka, Marchette, and Wallet [9] also utilize an existing IDS for input and apply several known graphical techniques, such as parallel coordinate plots and circle plots, to visualize network traffic. All of these systems have made

assumptions about the nature of ID work and the needs of security analysts without empirical support. The sole known exception to this is Yurcik, et al. [11], which describes gathering requirements from a small sample security operators and attempting to incorporate those findings into their visualization prototype.

These papers present various visualization approaches to ID and demonstrate the need for more innovative tools. What is missing from these studies is an understanding of the unique needs of ID analysts. The lack of understanding of user needs creates limitations in these tools, such as an excessive focus on monitoring, but little support for analysis, overly complex displays for monitoring, inadequate interactivity and manipulability in the tools, and a lack of correlated multiple views and data sources. Clearly understanding the complexity of the ID task and how analysts accomplish their work is crucial to designing successful support tools. Thus, before designing visualization tools to support ID tasks, we need to understand how the human ID analysts currently interact with their IDS to successfully detect malicious or illicit activity.

METHODOLOGY

The objective of this research was twofold: (1) to gain an understanding of how ID analysts perform intrusion detection, and (2) to determine characteristics of ID tools that will address the current limitations in ID monitoring and analysis tools. The methods used are summarized in Table 1.

Method	No. of analysts
Prototype evaluation	4
Contextual interviews	9
Focus group	7

Table 1. Study methods and ID domain expert participants.

The sample for the study was comprised of sixteen security analysts (four interview participants also volunteered for the prototype evaluation) with ID expertise, either theoretical or practical. By deliberately choosing a sample with diverse experience in ID, the range of viewpoints was increased.

The data collection methods were selected to answer the two research questions. The need to understand general ID behaviors necessitated contextual interviewing. Specific visualization characteristics were explored via a focus group and a usability test of a functional prototype (a standard 3D glyph-based system). Analyst interaction with the prototype helped elicit more system specific needs than either the interviews or focus group alone, as it is hard for users to imagine IV tools without seeing them

RESULTS

The results of the study identified several important issues for the design of IV tools for ID. The participants were enthusiastic about visualization tools for ID, as demonstrated by the following participant:

> "I would opt for any type of graphical representation over text… because I can look at a graphic much easier than I can read text and I can think about or do other things if I am being distracted"

The analysis led to the development of a process model for intrusion detection work and related visualization needs. Table 2 presents the relationship between the typical tasks of analysts and the related requirements for IV tools.

Phase	Analyst Tasks	Visualization Needs
Monitoring	• Monitoring all attack alerts • Identifying potentially suspicious alerts	• An overview of the alert data • Simple displays • Support for pattern and anomaly recognition • Flexibility • Speed of processing
Analysis	• Analyzing alert data • Analyzing other related data • Diagnosing attack	• Multiple views, zoom, drill down, focus + context solutions • Correlation between displays, linked views • Filtering and data selection
Response	• Responding to attack • Documenting and reporting attack • Updating IDS	• Suggestion for response action • Incident reporting • Annotation/feedback to facilitate future analysis • Saving views • Historical display • Reporting data transfer

Table 2. ID tasks and visualization needs.

INTRUSION DETECTION TASKS

In order to design tools to support the tasks related to ID, it is imperative to first understand how the work is accomplished. All of the participants followed a similar, high-level process model consisting of the three phases shown in Table 2: monitoring, analysis and diagnosis, and response.

The first phase of ID is the surveillance of the network infrastructure and resources. For the analysts we interviewed, this consists of either real-time monitoring of IDS output or post-hoc examination of batch processed (usually daily) IDS output. To do so, the participants all relied on textual displays, either in the form of email notifications or IDS consoles that would display the most recent alerts in tabular format. Although the IDS is the primary focus of the monitoring phase, other monitoring systems, from simple "pings" to determine if a server is listening to collecting bandwidth and system usage, also play a role. From a security standpoint, these secondary systems are typically not used for detecting intrusions per se, but provide context for the analysis that takes place next. It should be noted that many analysts do have duties and responsibilities in addition to ID, and so often have limited time and attention to give to the continuous monitoring of the IDS.

The transition from monitoring to analysis and diagnosis is triggered by an event, usually an alert generated by the IDS

that is suspicious. While monitoring involves only the display of the output from the IDS or other monitoring devices, diagnosing that output involves not only the alert artifact itself, but a host of other sources of data that provide the contextual information necessary to determine whether or not the alert is an actual intrusion and if so, how severe it is. In some cases this is an easy decision; however, much more likely are situations when an alert needs to be investigated in more detail. To accomplish this, analysts rely on the alert itself, their own knowledge and experience, and the available contextual information relating to the alert, all of which must be fused together in a cognitively intensive process of diagnosing the accuracy and severity of an alert.

If the results of the analysis lead to a diagnosis that the alert does indeed represent an intrusive or malicious activity, the analyst must then determine the correct response. This includes reacting to, documenting, and reporting the attack. If an active response is required, the analyst must choose the most appropriate response based on prior experience and knowledge of the attack and the environment.

The following section describes the IV requirements to support these tasks. As mentioned earlier, most current IV tools focus solely on the monitoring phase, and do not consider the entire ID process as a whole.

INFORMATION VISUALIZATION REQUIREMENTS

Phase 1: Monitoring

Simple displays

Our participants preferred simple, 2D displays for this phase, as these allow for continuous monitoring without the need for focused attention, building on pre-attentive visual processing (the fast, parallel recognition of color, shape, and movement by humans). Visualization support for this phase must provide a starting point for recognizing and flagging events that require further analysis in a way that can be done quickly and effectively without requiring the analysts' full attention.

Overview displays: Data and visualization attributes

Similar to the findings in [11], displaying an overview of the current activity is essential. As one participant told us "people want the big picture." Graphical overviews can serve this purpose well. Attributes of the popular IDS Snort [8] (most IDS's have similar attributes) alerts are shown in Table 3. All participants were asked to select the most important data attributes to include in the visualization displays. There was general agreement about these attributes (shown in bold), with the date/time being unanimously considered the most important. These attributes are well suited to provide an overview for the monitoring phase. The rest of the attributes (not bolded) must be provided in a drill-down detail view to support later analysis.

Message Signature	Time to live	Push flag
Classification	Type of Service	Connection Reset flag
Priority	Snort Rule ID	Syn packet flag
Date	IP header length	Fin flag
Time	IP datagram length	Sequence number
Source IP	IP Flag	Ack number
Source port	Reset flag 1	Window size
Destination IP	Reset flag 2	Length of data within segment
Destination Port	Urgent flag	
Protocol	Ack flag	Data urgent Pointer

Table 3. IDS alert data attributes for overview display

Flexibility

The need for end-user customizability with the IV displays was an important finding for both the monitoring and analysis phases. As described above, ID requires a deep understanding of idiosyncratic local networks. Analysts have had to configure IDS's in order to identify attacks on their unique network. This flexibility must be reflected in the visualization displays. Participants were very much in favor of the ability to set up their own visualization display settings and they did not object to the added effort, but voiced the need for saving these settings and being able to reuse them later.

Phase 2: Analysis and Diagnosis

Filtering and interaction

While monitoring tools should require as little user interaction as possible, supporting analysis is a much more interactive activity. Due to the large size of the data sets, filtering is a very important function for IV tools for ID as a transitional mechanism from monitoring to analysis. Multiple discrete ranges need to be selected, and predefined and user-defined clusters should be able to be saved and reused in more complex displays. In addition, filtering data should provide a means for highlighting data without necessarily removing it from the display, as the data that is not the focus of the task is still important in providing vital contextual information for correctly diagnosing the alert.

Exploration

The analysis and diagnosis task requires support for user exploration that warrants markedly different IV displays than those used in monitoring. The need for simplistic displays for quickly identifying an alert in monitoring is replaced by a need for more powerful visualizations that can represent multidimensional data from multiple sources.

Multiple data sources and correlation

The analysis and diagnosis of an alert cannot be accomplished without also taking into account secondary data sources that supplement the information contained in the alert itself. A visualization tool must effectively fuse these disparate data sources together seamlessly in a single display, that can correlate all of the data together. An example of this is host information that determines if the target of an attack is vulnerable to the attack described in

the alert. The breadth of the data sources will depend on the organization, but will include both dynamically collected and static network-level and host-level data.

Multiple views and levels of data

In this phase, the ability to have multiple views of the same or related data becomes important. Analysts would like to utilize multiple displays at the same time, such as multiple displays each running the visualization tool on the same data, but with different data attributes or different time spans displayed. Another important need is to display several levels of data (i.e., network sessions, raw packets, host information), and allow users to drill down or zoom in on certain data items.

Phase 3: Response

The support necessary for responding to attacks extends IV displays beyond data manipulation and viewing. The ability to save views, keep histories of exploration and activity, and annotating alerts will all help analysts document and report incidents. These functions are often missing from IV tools, although they allow users to make the transition between exploring and finding information and using and reusing this information in their work. Suggesting possible responses for different types of attacks could greatly aid the speed and efficiency of responding to attacks; these suggestions could come from annotations of previously diagnosed similar attacks or from IDS developers.

CONCLUSION

The application of user-centered and ethnographically informed methods to the design of information visualization tools sheds light on the mismatch between innovative displays and the needs of real-life users. Novel IV solutions can only be successfully applied to real-world problems if designers understand the work they are designing support tools for. Our study addressed this disconnect between visualization tools and their context of use by exploring the design space of visualization tools for ID via a field study. We have identified several design implications based on a three-phase model of ID analysts' work. Our future work will incorporate our findings into the design of visualization tools to support ID analysts.

IV for ID tools should include simple, network-based displays for monitoring and more complex, linked, multiple displays are needed to support the diagnosis and analysis of attacks. These second set of displays should allow analysts to drill down and examine the attack activity in more detail and from several different views, and synthesize multiple data sources needed to put IDS data in the larger context of the analysts' environments. Current IV for ID tools focus on the monitoring phase with limited support for analysis and diagnosis of IDS alerts. Dynamic interaction and exploration capabilities in these tools are usually missing or limited, although it is crucial for their successful application. New IV tools incorporating these guidelines should be developed to support the entire process of detecting intrusions and other monitoring and follow-up analysis activities

The wider application of user-centered approaches in IV design is needed to ensure the utility of these tools to users. Both user-centered design and evaluation are often missing from the design of these tools, however, as our results show, this often leads to solutions that do not serve users' needs or are unusable by users.

ACKNOWLEDGMENTS

Thanks to Andrew Sears, Penny Rheingans, Jeff Campbell, Enrique Stanziola, Utkarsha Ayachit, and the Department of Defense for their contributions and support.

REFERENCES

1. Erbacher, R.F. Glyph-based generic network visualization. *Proc. SPIE Conference on Visualization and Data Analysis*, (2002), 228-237.
2. Erbacher, R., Walker, K., & Frincke, D. Intrusion and misuse detection in large-scale systems. *IEEE Computer Graphics & Applications* (2002), 1, 38-48.
3. Fortier, S.C. & Shombert, L.A. Network profiling and data visualization. *Proc. IEEE Systems Man and Cybernetics Society Information Assurance and Security Workshop*, (2000), 136-142.
4. Girardin, L. & Brodbeck, D. A visual approach for monitoring logs. *Proc. Systems Administration Conference (LISA)*, (1998), 299-308.
5. Heady, R., Luger, G. Maccabe, A., & Servilla, M. The architecture of a network level intrusion detection system. *Technical Report CS90-20*, Dept of Computer Science, University of New Mexico (1990).
6. Northcutt, S., Novak, J., & McLachlan, D. *Network intrusion detection: An analyst's handbook*. (2nd ed.) New Riders Publishers, Indianapolis, IN, USA, 2000.
7. Nyarko, K., Capers, T., Scott, C., Ladeji-Osias, K. Network intrusion visualization with NIVA, an intrusion detection visual analyzer with haptic integration. *Proc. 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, (2002), 277 –284.
8. Snort. http://www.snort.org/
9. Solka, J.L., Marchette, D.J., and Wallet, B.C. Statistical visualization methods in intrusion detection. *Computing Science and Statistics*, (2000).
10. Stallings, R. *Cryptography and network security: Principles and practice*. Prentice Hall, Upper Saddle River, NJ, USA, 1999.
11. Yurcik, W., Barlow, J., Lakkaraju, K., & Haberman, M. Two visual computer network security monitoring tools incorporating operator interface requirements. *CHI Workshop on HCI and Security Systems*. (2003)