Using Visualization for Situational Awareness Support to IA Analysts

Robert J. Bagnall

Secure Decisions

DRAFT

Publication Date: 31 July, 2006

TABLE OF CONTENTS

TABLE OF CONTENTS 2
TABLE OF FIGURES 4
1 Overview 5
2 Cognitive Task Analysis (CTA) 6
2.1 Resulting Goals of Visualization Within the IA Analysis Effort 6
2.1.1 Monitoring 6
2.1.2 Inspecting 6
2.1.3 Exploring 6
2.1.4 Forecasting 7
2.1.5 Educating 7
3 Findings 8
Help Mitigate Analysts' Challenges 8
3.1 Data Challenges 8
3.1.2 Massive Data 8
3.1.3 Complex Data 8
3.1.4 Unreliable Data 8
3.1.5 Low-level Data 9
3.1.6 Segmented Data 9
3.1.7 Missing or Insufficient Data 9
3.1.8 Time Delays in Receiving Data 9
3.2 Cognitive Challenges 10
3.2.2 Knowledge of Network Topology and Sensor Placement 10
3.2.3 Functions and Missions of Monitored Hosts 10
3.2.4 Areas of Vulnerability an Attacker Could Exploit 11
3.3 Ease of Use 11
3.3.1 User-customizable 11
3.3.2 Design With Expansion and Evolution In Mind 11
3.3.3 Unambiguous Visual Idioms 11
4 Improving Situational Awareness of IA Analysts 12
4.1 Multiple Views of a Single Data Model 12
4.1.1 Placement of Panels 12
4.1.2 Easy Manipulation of Display Data 12
4.2 Enhance Analyst's Skills 12
4.2.1 Detection of Relationships 12
4.2.2 Maintenance of Separate Mental Models of Multiple Attackers 13
4.2.3 Evaluation of Alternative Hypotheses 14
5 Design Concepts & Lessons Learned 15
5.1 Broad Situational Overview (The Dashboard) 15
5.1.1 Port Activity (for IP) 15
5.1.2 Top Ten Active Destination IPs 15
5.1.3 Top Ten Active Source IPs 15
5.1.4 Organization Hot IP List 16
5.1.5 Personal Hot IP List 16
5.2 E-Diary 16

5.3 Filter Status/Selection Status 17
5.4 Visual Representation Through Interactive 3rd Party Products: SecureScope 18
5.5 Web Browser 19
5.6 Scene Interaction Between 3rd Party Apps Within A Framework 20
5.7 Display Filtering 21
5.8 Reporting 22
5.9 Packet Data (payload) Analysis 22
5.10 Drag and Drop Capability 22
5.11 Range Filtering 23
6 Framework Concept Reviewer Feedback 24
6.1 General Discussion of Features 24
6.1.1 Summary of Visual Aides Usefulness 25
6.1.2 Summary of VIA Features Perceived Utility 25
6.2 Recommended Improvements to the Framework Concept/Feature Set 26
6.2.1 eDiary Recommendations 26
6.2.2 Report Generator Recommendations 26
6.2.3 Visualization Recommendations 26
6.2.4 Additional Recommendations 27

DRAFT

TABLE OF FIGURES

Figure 2 Dashboard 15
Figure 3 E-Diary 16
Figure 4 Filter Status Figure 10 Selection Status 17
Figure 9 SecureScope 18
Figure 14 Browser 19
Figure 16 Scene Interaction - Row Selection of Table Lens Highlights Node in Star Tree 20
Figure 17 Scene Interaction - Selection Filter Used to Highlight All Foreign Nodes that had >= 10000 Bytes Sent Back to them 21
Figure 18 Display Filter Panel Figure 19 Selection Filter Panel 22

DRAFT

1 Overview

The objective of this document is to help the reader understand the needs, concepts, difficulties, and goals behind developing concepts and fielding solutions related to improving situational awareness for Information Assurance (IA) analysts through visualization. The document explains in detail the core set of design principles that the VIA design team constantly referred back to when developing and fleshing out VIA's visualization concepts. It is this core set of design principles that kept the VIA design team focused on designing a system that will fulfill the current and future needs of the IA analyst.

The core of the document and conclusions presented herein is centered around a Cognitive Task Analysis study involving actual Computer Network Defense (CND) and IA analysts in real-world environments. The study engaged over 40 individual analysts and eight organizations – both commercial and government. Details of the CTA are described in Section 2 of this document.

This document, and the related concepts herein, focuses on the first and second stages of IA analysis – initial detection, followed by situation assessment and the eventual declaration of an incident. The concepts herein are designed to assist IA analysts in different roles during these two stages. First, analysis of suspicious activity from massive data sets. This is followed by technique analysis, impact assessment, and correlation against historical data to find new (or verify existing) patterns that illuminate suspicious activity.

DRAFT

2 Cognitive Task Analysis (CTA)

The Cognitive Task Analysis (CTA) is the study of an individual or team's mental processing, activities and communications within a specific work context. It is part of the larger field of cognitive engineering that includes CTA, cognitive work analysis, and knowledge elicitation. The output of a CTA is a detailed description of the tasks that an individual or team performs, the data on which they operate, the decisions they make, and the processes and activities (cognitive, communicative, and physical) that they engage in to reach those decisions.
CTA is focused on eliciting information from individuals about the thought processes they use in the course of completing specific tasks. It involves both observing individuals as they go about their work, and asking directed questions about the way in which they approach the problems, how they decide what step to take next, and what are the challenging tasks of their work. The focus is on the cognitive processes, rather than the mechanics of how tasks are completed in the current environment, and therefore tends to be less affected by tool bias. Researchers ask questions such as "Why did you use that tool? What caused you to use it? What information did you get from it? How did that change the way you think about the problem at hand? Would there be another way to get this information? What steps will you take next?" Such questions enable the researcher to separate the subject's thought process about tasks from the way existing tools require him to operate and perform that task.

DRAFT

2.1 Resulting Goals of Visualization Within the IA Analysis Effort

Secure Decisions performed an extensive study of existing visualizations along with what types of information various visualizations best convey. Five distinct components of data presentation came out of our findings. The components are:

2.1.1 Monitoring

A visualization might aid in monitoring an a continuous data flow for anomalies or the continued function or success of an automated process. This style of application is characterized by continually changing data, the need to comprehend at a glance the overall state, the need to easily recognize state changes, and the need to take actions based on perceived states.

2.1.2 Inspecting

Inspecting a given data set is another possible application for a visualization. This style of application is characterized by directed searches for specific details, requests to provide clarification, and finding data to validate hypotheses.

Inspecting is often associated with analytic discoveries, motivated by observation, where analysis is understood as comprehending the whole by breaking it down into its component parts and their inter-relations, often in support of an induction. Yet visualization can also be used to trigger inspection as well as display it in-process or its results. Anomalies can be effectively displayed visually as an indicator that further inspection is required.

2.1.3 Exploring

Exploring is often associated with conceptual discoveries through a synthesis of observed characteristics into a coherent whole, often in service of a deduction. Data visualization often

enables ad hoc “visual discovery” and recognition of patterns, trends, and anomalies during exploration.

2.1.4 Forecasting

A data presentation might be applied to forecasting a potential future state. This style of application is characterized by pattern matching, extrapolation, correlation, and attributing causation or potential causation to an event or pre-event indicators.

The goal of forecasting may be “descriptive” if it seeks to find the likely future state presuming the current progression continues without intervention. It may be “normative” if it seeks to determine a particular future state based on potential courses of action (e.g., correction or “what if?”).

2.1.5 Educating

Educating may be associated with training others to perform particular tasks, teaching others about particular concepts, communicating findings to colleagues or laypersons, and/or documenting decisions for review or justification. Here visualization can play a key role, particularly when applied in support of reporting or briefing non-technical personnel and senior staff. Visualization provides a highly-effective means of displaying situational awareness and conveying both meaning and impact to non-technical decision makers who require succinct, accurate displays of a given situation upon which they must choose some action.

DRAFT

3 Findings

Help Mitigate Analysts' Challenges

The following is a list of all of the challenges that were discovered as a result of the CTA. Although no visualization can hope to solve all obstacles to effective analysis, this list is tracked throughout the concept, design, testing, and implementation phases in order to keep a good perspective on what obstacles analysts need to overcome.

3.1 Data Challenges

3.1.1 Inefficient Access to Data

Because of the constant flow of suspicious network activity within the regular operational flow, analysts find expeditious data collection and data manipulation are imperative preliminary steps in the analytic process. Having a full possession of available data allows them to investigate and assess indicators and events quickly. The greatest internal obstacles experienced by analysts interviewed were related to limitations related to their access to data, lengthy periods of time to acquire data, and time latency related to reorganizing and filtering data prior to the interpretation process. We consistently observed analysts spending substantial amounts of time sorting, truncating, merging and filtering data rather than performing actual analysis on that data. Multiple data sources, including those from outside their own organization, further compounded the problem.

DRAFT

3.1.2 Massive Data

Voluminous data presents problems when the data volume exceeds the analyst's ability to mentally process it. Data overload on analysts is made more problematic when the volume exceeds the computing resources' processing and storage capabilities as well. In either case, the result is incomplete analysis and/or inaccurate results, which may lead to malicious activity or incidents being missed.

3.1.3 Complex Data

Indicators of an intrusion may occur in a variety of data sources, e.g. packet headers, network traffic patterns, policy violations, network topology, CERT advisories. It is left up to the analyst, armed with tools and scripts to help sort and filter data, to formulate associations between numerous data sources. Human analysts are, and will continue to be, the directly in the line of correlation of complex environment data.

3.1.4 Unreliable Data

Sensors such as Snort, ISS RealSecure, and other tools often produce a high number of false positive alert indicators, which lead an analyst to falsely investigate legitimate network activity. One of the most time-consuming activities of Level 1 analysts is the elimination of those false positive alerts from legitimate indicators within the sensor data. False positive alarms add significantly to the general workload and divert analysts' attention from investigating malicious network activity.

3.1.5 Low-level Data

Individual data points, without context or reference, have little meaning to the analyst. Meaning is only derived when a human being, in the absence of available tools, pieces together individual indicators or data elements into larger constructs, e.g. individual packets grouped together into session transcripts, or individual probes grouped into scans.

3.1.6 Segmented Data

Events and/or incidents may occur over more protracted periods of time than what current toolsets or analytical procedures actually support. Tools currently used provide reports in hourly and/or daily groupings, and do not analyze data over longer periods. This can lead to analysts missing hostile activity – known as a "low-n-slow" attack. Information may also come in narrow, segmented slices over a long period of time, making it difficult for analysts to fuse and observe an entire attack pattern. If the activity spans more than one shift, the information and resulting analysis can easily be lost. In many cases, analysts also only review data pertaining to their assigned list of sensors. This makes it impossible to see patterns across multiple sensors, which could indicate a larger problem, unless analysts are regularly communicating and sharing information and analyses across multiple sensors and multiple shifts..

3.1.7 Missing or Insufficient Data

Throughout the entire CTA, analysts never had a complete set of data to accompany an attack. There were several factors that contributed to the missing data problem. They work with data sets that have several, sometimes very important, pieces missing. Incomplete sessions, one direction of a two-way communication, lack of sensor coverage of part of the network, and techniques used to reduce data all omit valuable information from the analysis. In addition, organizations are limited in what data they can and do collect, since it is resource-intensive to collect and store massive information for long periods of time. As a result, organizations do not collect every type of data that could be used in an investigation, and they delete or overwrite historical logs that might have been useful in an investigation.

Reports escalated from other subordinate organizations about suspicious activity or confirmed incidents do not always provide the completeness of data required by analysts, e.g. the report may summarize an incident but not include all relevant historical packet data or indicators. An analyst may not possess or find enough supporting data to package up the incident to generate an incident report. Other analysts have indicated that their lack of access to packet payload data presents an obstacle in determining whether unusual activity is malicious.

3.1.8 Time Delays in Receiving Data

Analysts experience substantial time delays in gathering all the data required to comprehensively investigate suspicious activity and/or incidents. Bottlenecks in collecting the data resulted from the need for analysts to access several different data sources, using several different tools. Certain data was obtainable by request only and not directly accessible. The requesting and obtaining of specific information required reliance upon not only other tools and data sources, but additional human resources as well. Bringing such additional assets into the mix created time delays in attaining information. Depending on the type of data being requested, it is not unusual

DRAFT

for information to be obtained from two hours to a month after the analyst has requested it, and often too delayed in its arrival to be effective in thwarting any attacks.

3.2 Cognitive Challenges

3.2.1 Knowledge of "Normal" Environment Behavior

Analysts look for deviations from normal patterns as cues to attacker activity, yet what is normal for a specific enclave varies over time and is seldom well-understood. Analysts develop a sense of normal through first-hand experience and rely heavily on that knowledge; but they find it difficult to articulate what normal looks like. Moreover, they have significant trouble conveying "normal" between shifts or experience levels among analysts. New analysts are limited in their ability to detect meaningful anomalies until they can learn over time or gain access to knowledge from others about what constitutes normal. Consistently, however, gaining experience from others is difficult as it requires training time that most organizations cannot or do not provide due to manpower and data volume issues.

3.2.2 Knowledge of Network Topology and Sensor Placement

Analysts use their own mental model of the topology of the network they are monitoring to answer many questions they raise during the course of their analysis. Knowledge of network topology and sensor placement within it assists them in determining whether an IP address associated with unusual activity is inside or outside their network, where sensors would have to be placed to provide them with visibility into specific network traffic, and what path an attacker might take through the network once he has compromised a host.

Despite their reliance on network topology, analysts rarely have an accurate topology supplied to them by the enclave. Instead, they create a mental model of the topology of the network they are monitoring, based on a combination of factors such as patterns of "which device is talking to which", old topology diagrams, contact with the SysAdmins of the networks they are monitoring, and other sources. They piece together a picture of what the topology looks like, but they rarely get feedback regarding the accuracy and completeness of that picture. Major analysis decisions are made based on a set of assumptions about network topology that may or may not be correct. This is particularly true with commercial Managed Security Service Providers (MSSPs) who monitor the topologies of clients. In this case, the knowledge must be extended to numerous topology collection efforts across multiple clients, further hindering the accuracy and raising the potential for error during analysis.

3.2.3 Functions and Missions of Monitored Hosts

The function of a host being monitored influences the type of network traffic pattern that is considered normal, and what constitutes deviations from normal. We repeatedly observed that analysts need to know the specific function of a host, before determining whether activity to and from it is suspicious or legitimate. They will interpret the same traffic patterns in radically different ways depending on whether the monitored node is a personal workstation, a web server, or a mail server. In addition, their assessment of the impact of an attack on the node will vary depending on the functionality and criticality of the host.

DRAFT

Unfortunately, analysts often have limited access to information about the functionality of a host they are monitoring. They rarely have a priori knowledge of each host's function or mission within the enclave. When presented with unusual activity on a host, they have to contact the local system administrator for information about the host's function. Even then, they may not obtain the information they need.

3.2.4 Areas of Vulnerability an Attacker Could Exploit

Attempts to exploit the network occur all the time. The analyst is mainly concerned with those exploits that can achieve success by taking advantage of some vulnerability within the network. Therefore, the analyst wants to know where the vulnerabilities are, and how an attacker might utilize those vulnerabilities to penetrate the network. At present, vulnerability data can be made available to the analyst if requested; however, it is not always readily available. Emerging vulnerabilities also cause the list of vulnerable nodes to change frequently. Furthermore, the analyst is limited in his ability to mentally project all possible paths through vulnerable nodes on the network that an attacker could take in order to reach a targeted node.

3.3 Ease of Use

Any visualization solution must be easy to use and intuitive to the end-user. A piece of software can be incredibly powerful and useful, but if the end-user does not find it easy to work with, the software will become "shelfware". For this reason, careful attention must be paid to making sure that any functionality added to the system has "ease of use" in mind.

3.3.1 User-customizable

According to the CTA findings, different analysts perform data investigation in different ways. For this reason, it is essential that any visualization solution that is developed to aid analysis must allow analysts the freedom to investigate the data in various ways.

3.3.2 Design With Expansion and Evolution In Mind

Designing a new visualization system for analyst issues today may not necessarily help the analyst with tomorrow's problems. For this reason, the visualization solution design should be built on a framework whereby a combination of commercial, 3rd party, and other existing visualization tools can interoperate. This allows for growth and flexibility within the framework itself and does not tie down an organization or its analysts to any one solution. With the framework in place, if an analyst finds a different kind of visualization that he/she wants to utilize, only a minimal amount of integration is required to add that new visualization.

3.3.3 Unambiguous Visual Idioms

In selecting what visualization views to incorporate into a new design, only those views that convey a clear, precise meaning to the analyst must be selected. Ensuring only those views that convey an appropriate "message" about the data they are visualizing are integrated into the system helps preserve and enhance the analyst's situational awareness.

DRAFT

4 Improving Situational Awareness of IA Analysts

The main objective in designing an visualization solution is to increase an analyst's situational awareness. The overall goal is to increase analysts' situational awareness, resulting better, more informed decisions.

4.1 Multiple Views of a Single Data Model

Providing different mechanisms to view the data in different ways may reveal patterns in the data that would otherwise have not been detected. The conceptual design must show the same data simultaneously in multiple ways to enhance the end-user's situational awareness. This further allows them to alter their perspective and questioning of what they're seeing for indicators by using visual triggers.

4.1.1 Placement of Panels

To enhance the analyst's situational awareness, VIA was designed to be an interactive framework of panels. A notion of a "main" panel exists that is to be the focal point of the application. Whichever one of the available views currently occupies the main panel, the appropriate toolbar for that view is displayed in the main panel's toolbar area (lower part of the window). An analyst can place any view he likes into the main panel by double clicking a specific "hot" area in the desired view to display. Upon doing so, the current contents of the main panel will switch with the selected view's panel leaving the newly selected view in the main panel and the old main panel view in the newly selected view's panel. Keeping views visible that are relevant to the analyst helps the analyst keep the best possible situational awareness.

DRAFT

4.1.2 Easy Manipulation of Display Data

In order to learn the most about the data, an analyst needs to be free to manipulate the data (such as filtering) without paying too large of a price (time). Conducting the CTA helped identify the need for data manipulation controls to be easily accessible and quick to operate. As a result, slide bar controls were identified for use extensively in the conceptual design of display filtering and selection filtering control panels for analysts.

4.2 Enhance Analyst's Skills

4.2.1 Detection of Relationships

One of the fundamental cognitive skills that all effective analysts share is the ability to see patterns and connections in discrete events. At *Stage 1 – Detection* and *Stage 2 – Situational Assessment* of the analysis process, analysts string together packet headers or IDS alerts that seem to belong together. They may use time, source IP, destination IP, geographic location of source IP or attacker techniques as the common element that draws the disparate activities together. This piecing together of a realistic incident picture is not limited to correlation analysts. It is done in varying degrees by all analysts, even level 1 analysts who are working under strict time constraints. During *Stage 3 – Threat Assessment*, the incident handler or threat analyst looks

for connections and patterns among confirmed incidents. They use the same connecting elements as in Stages 1 and 2 to draw disparate incidents together, as well as common nation-state sponsors.

The exploration of data to find relationships is a lengthy, self-directed process. As they sort through the events, they look for connections. They may not initially know how events are connected, but they test various hypotheses as they inspect the data. They may sort data first by IPs, then by countries of origin. The first sorting is often done mentally, to see if any common element pops out at them; they then follow with whatever sorting tools they have available. These tools could be a series of Perl scripts, SQL queries to a database, or sorting columns on a display; occasionally they use visualization tools to visually group data based on multiple dimensions.

The tools used vary from operation to operation, with some providing more triggers or support than others to analysts. Yet overall, the cognitive skill involved in detecting relationships is so critical, that any procedures or aids that can expedite or enhance it would improve the IA analysis process. This is precisely the goal behind the conceptual design: to provide better indicators and potential early warning for analysts before incidents occur.

4.2.2 Maintenance of Separate Mental Models of Multiple Attackers

Analysts simultaneously track many potential attackers and incidents; each "track" is a mental model of an attacker. In *Stage 1 – Detection* and *Stage 2 – Situation Assessment*, they notice suspicious activity associated with a particular source IP at a specific point in time. They look for other prior activities from the same IP and set up the sensors to notify them whenever activity from the same IP occurs in the future. In doing this, they keep track of a potential attacker. The interesting activity that cued the level 1 analyst's attention may not initially warrant escalation; however, the data he accumulates through tracking may supply sufficient evidence to escalate in the future. The level 2 analyst keeps track of numerous escalated events in a similar manner. In *Stage 3 – Threat Assessment*, incident handlers and threat analysts keep track of multiple incidents, intrusion sets and problem sets using the same types of cognitive skills. As incidents are reported, they look for features that match other incidents. They hypothesize that several disparate incidents may, in fact, be the work of the same attacker. If there is sufficient evidence to support the hypothesis that the incidents are part of the same "track", they escalate to intrusion or problem sets.

At any one time analysts may be tracking dozens of interesting activities, events, incidents, or problem sets. Several analysts noted that keeping them separated in their minds can be challenging. Each track requires the analyst to create a mental model of an attacker, and for each model the analyst hypothesizes the attacker's identity, motive, and modus operandi. In some cases, particularly with incident data in Stage 3, the analyst combines separate tracks when the mental models are so similar that the analyst deduces that the two tracks have emanated from the same source. Likewise, the analyst may be tracking what initially appears to be one attacker or incident; however, data no longer seems to adequately fit the mental model of a single entity. The analyst then modifies the model, eliminates the model as no longer representative of an attacker, or divides it into two or more potential attackers.

Tracks can persist in an analyst's mind for several months, even when the track activity is infrequent. One level 2 analyst we interviewed spoke about "acquiring a target", mapping out what he thought the potential attacker's motive and techniques were, losing the target for a

DRAFT

month, and then “reacquiring the target” by noticing activity that fit the mental model of the attacker. For this to occur, it was necessary for the analyst’s mental model to persist over time, even with little activity.

Much of the simultaneous tracking of mental models is done mentally with few external aids. Analysts rely heavily on their memory of what they have seen in the past, so that they can recognize the target source IP when they see it, or recognize the profile of the attacker even if the source IP has changed. They keep personal notes in an electronic notebook or paper notebook, to which they refer when they see activity that matches one of their existing tracks. Threat analysts and some correlation analysts use graphical techniques to help them group attackers and depict relationships. A product called *Analyst Notebook*, which is in widespread use by law enforcement agencies to track patterns of criminal activity, is used by threat analysts to help them keep track of multiple intrusion or problem sets. However, it has scaling and user interaction problems that make it difficult to use when tracking many problem sets.

4.2.3 Evaluation of Alternative Hypotheses

Analysts generate and test alternative hypotheses as a method for determining the best explanation for unusual activity within their enclave or community. When faced with interesting activity such as a large volume of network traffic to and from a specific node on the network, an analyst will test the hypothesis that the traffic is legitimate by comparing the activity to known parameters of normal for a node with that organizational function. If that hypothesis is not confirmed, the analyst will consider other hypotheses such as that the activity represents a policy violation (e.g. use of Kazaa) by an authenticated user without malicious intent or that the activity represents the unauthorized use of a compromised box by an attacker who passed undetected through the perimeter defense. Each of these hypotheses requires the analyst to gather data to evaluate the hypothesis. The analyst will frequently consider several hypotheses simultaneously, mentally keeping track of the status of each evaluation.

When analysts consider multiple, often novel, alternative explanations of data indicators, they view the data from multiple perspectives. This prepares them cognitively to discover unusual patterns and trends, and fosters correlation.

DRAFT

5 Design Concepts & Lessons Learned

5.1 Broad Situational Overview (The Dashboard)

The dashboard view is actually comprised of five views. This view of information is designed to be high-level. It can be used as a quick synopsis of the general state of security (during briefings, meetings, and demonstrations to management) or projected on an overhead or large operational display board within the operation for managers and visitors.

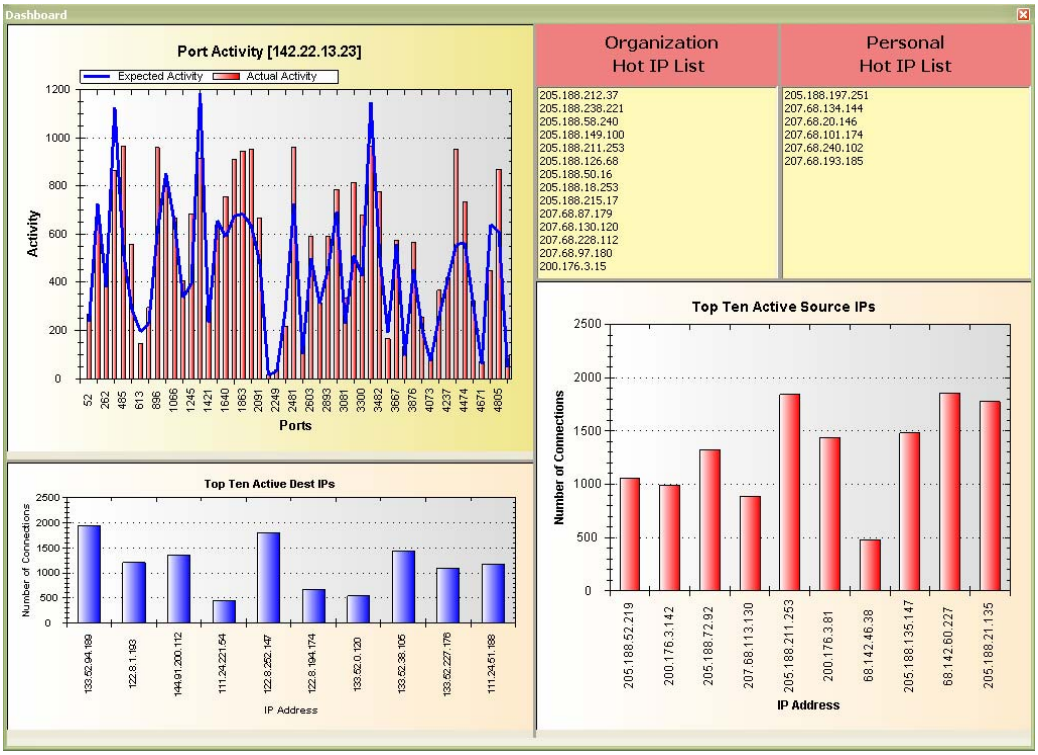

Figure 1 Dashboard

5.1.1 Port Activity (for IP)

Bar chart indicating specific port activity for a specific IP address as it compares to the expected port activity (blue line).

5.1.2 Top Ten Active Destination IPs

Bar chart indicating the organization's top ten most active destination IPs. In a production system, the number of IPs to display shall be configurable.

5.1.3 Top Ten Active Source IPs

Bar chart indicating the top ten most active source IPs. In a production system, the number of IPs to display shall be configurable.

5.1.4 Organization Hot IP List

Displays the IP addresses that the organization has deemed serious enough to be closely monitored.

5.1.5 Personal Hot IP List

IP addresses that are not currently on the organization's hot IP list, but have caught the attention of an analyst who feels he would like to closely monitor. Putting IP addresses in this list would help analysts who want to track suspicious IP activity across multiple data dumps. Rather than having to perform specific queries to see if the specified IP addresses exist in a new data dump, the conceptual design should be configured to automatically color code these entries when they are loaded into the system.

5.2 E-Diary

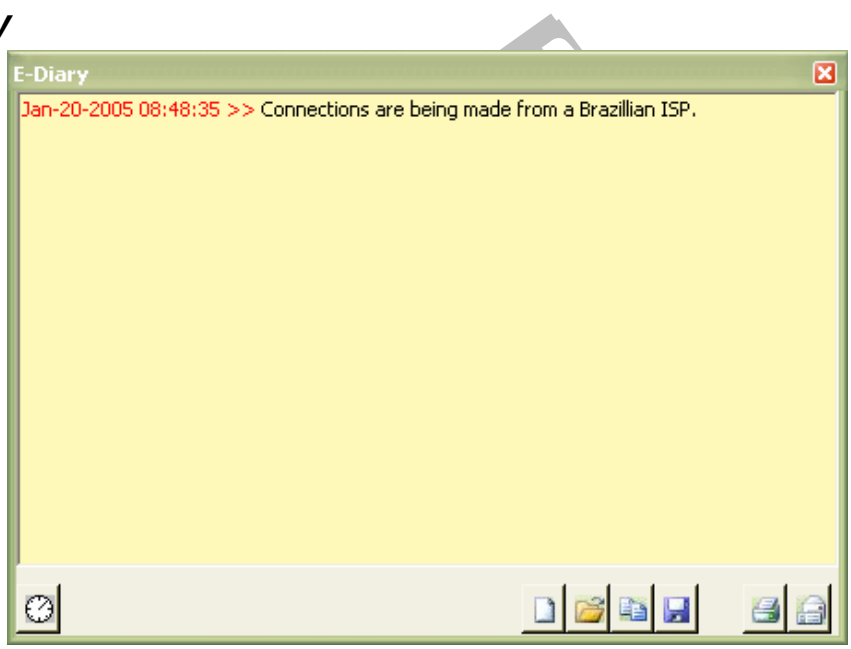

Figure 2 E-Diary

Provides a mechanism by which analysts can document their thoughts and findings without leaving the application. The E-Diary contains a timestamp button that allows an analyst to automatically timestamp an entry. Additional features, although not functional in the demonstration application, allow the analyst to save, email, open, copy, paste existing E-Diary transcripts. The intent of the E-Diary could be shared and used as a shift changeover communication tool.

5.3 Filter Status/Selection Status

Figure 3 Filter Status

Figure 4 Selection Status

The filter status panel and the selection status panel typically co-reside with one another in a single tab control panel. The filter status panel was designed to give the analyst a higher level of situational awareness by summarizing any data filters (query or display) that are currently in effect. In a fully functional system, this filter status panel would also show what criteria were placed on the query used fetch the starting dataset.

The selection status panel was designed to give the analyst a higher level of situational awareness by summarizing any data selection (highlight) filters that are currently in effect. If an analyst ever has a question why something is (or is not) highlighted in the scene, he can refer to this panel to see a quick summary of all of the selection parameters and current assignments.

5.4 Visual Representation Through Interactive 3rd Party Products: SecureScope

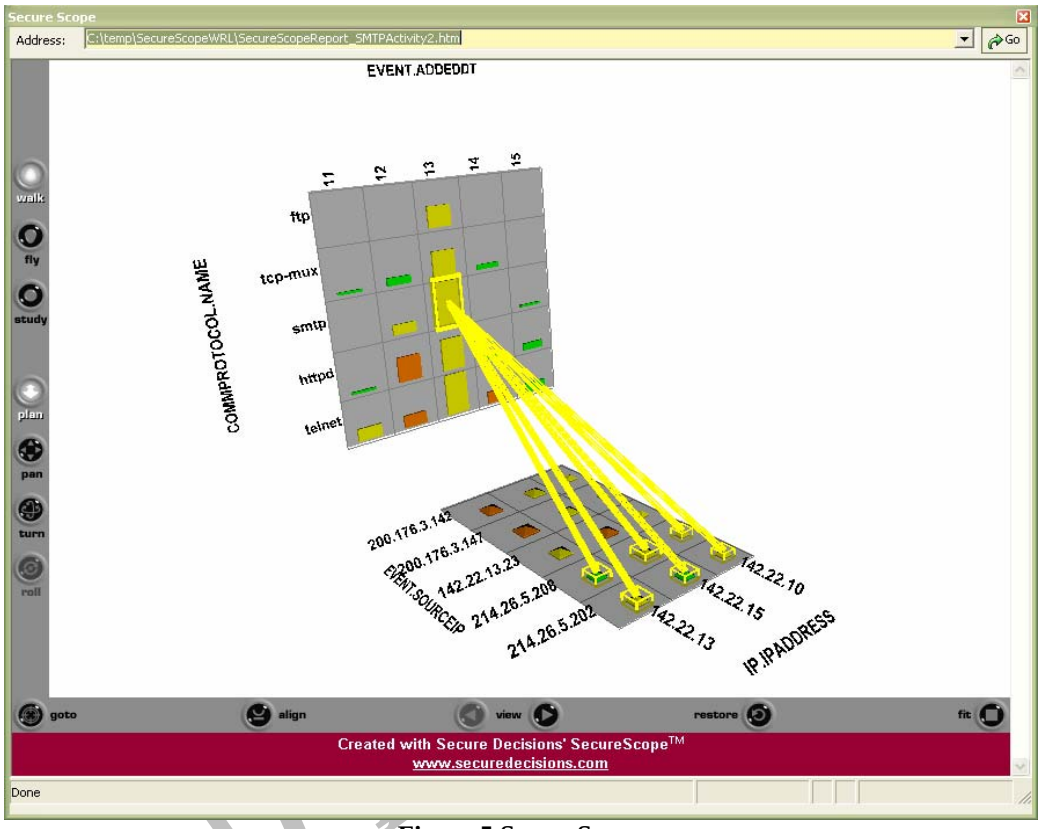

Figure 5 SecureScope

SecureScope developed by Secure Decisions does a good job of showing relationships among different data populations in relation to time. In the frequency-wall scene above, an analyst could very clearly see that two source IP addresses were using the SMTP protocol on hour 13 on multiple destination IP addresses.

5.5 Web Browser

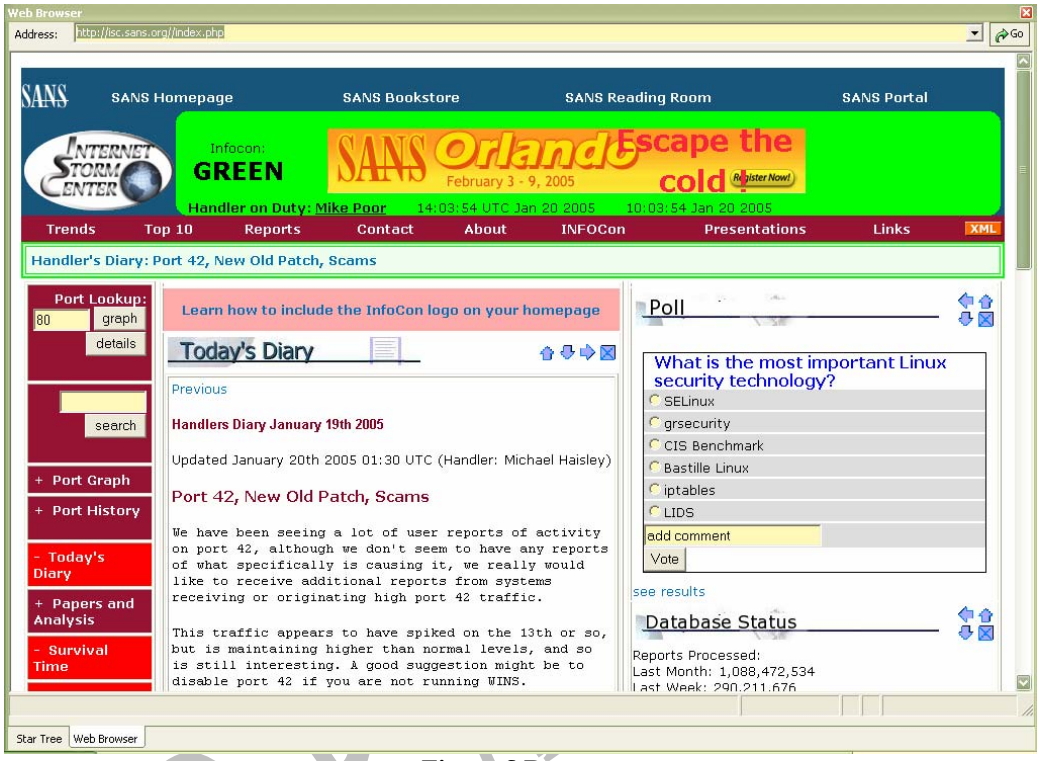

Figure 6 Browser

Through Secure Decision's CTA observations, it was evident that analysts frequently use external resources and controls like web browsers during their investigative process. For this reason, the concept system should be designed to have a web browser built into the application. The main panel display should be a tabbed window that always has the web browser as the second tab making it very accessible to the analyst without forcing the analyst to leave his investigative process.

5.6 Scene Interaction Between 3rd Party Apps Within A Framework

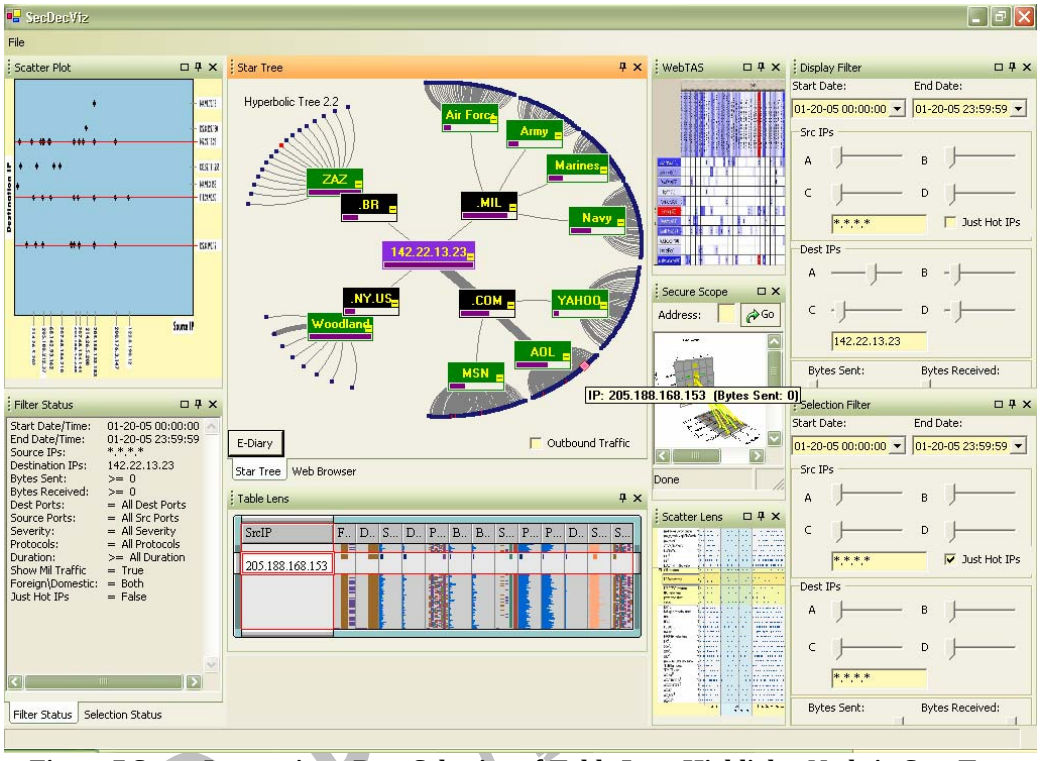

Figure 7 Scene Interaction - Row Selection of Table Lens Highlights Node in Star Tree

The capability of showing different views of the same data is powerful, but if one can interact with a selected view and have the actions be reflected in all views that would make for an extremely powerful tool. In the example above, a selection was made in the Table Lens which forced the Star Tree to highlight the node that corresponds to the Table Lens selection. To show the proper node was found, the mouse was hovered over the Star Tree node forcing the tool tip to be displayed which indicates the same source IP as the selected item in the Table Lens.

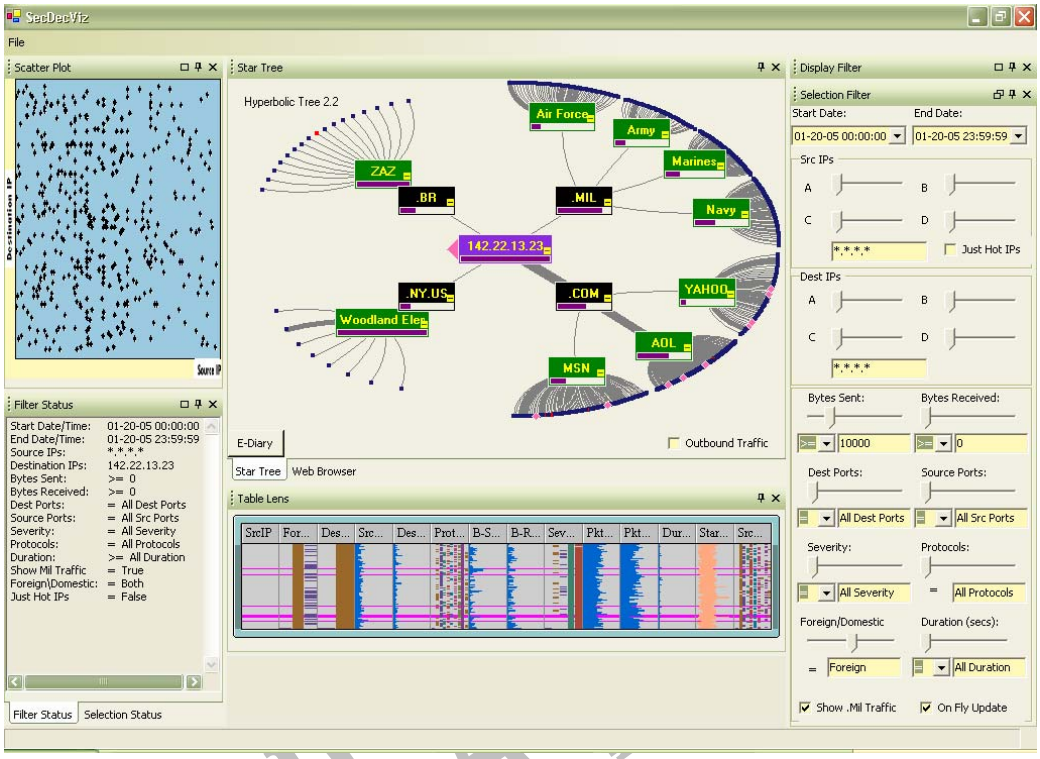

Figure 8 Scene Interaction - Selection Filter Used to Highlight All Foreign Nodes that had >= 10000 Bytes Sent Back to them

Another form of scene interaction comes in the form of the selection filter. In the above example, data records from both the Star Tree and the Table Lens were selected based upon the selection criteria of the source IP being foreign along with the number of bytes sent back to the source IP being greater than or equal to 10000. In a fully functional system, the Scatter Plot located in the upper left hand corner would also show the highlighted records.

5.7 Display Filtering

Display filtering refers to the ability of the analyst to manipulate the data that is currently loaded in the system. When display filtering is invoked, no additional data source queries need to be made to determine what data to display. The data is not really removed when it is filtered from the display, which leaves an analyst is free to explore the data by experimenting with defining various filter properties without taking the additional performance hit of having to perform a re-query each time.

The same easy to use filtering controls are also made available as selection (or highlight) controls. When performing data selection, data entries that meet the selection criteria are highlighted in each of the views. Again, these highlights interact across multiple visual tools within the framework and are completely controlled by the analyst.

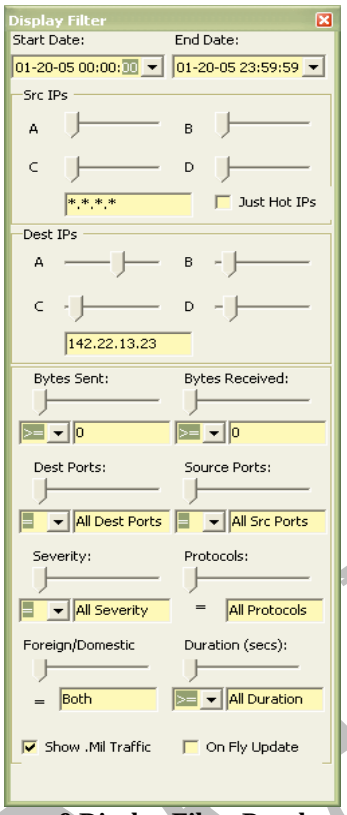

Figure 9 Display Filter Panel

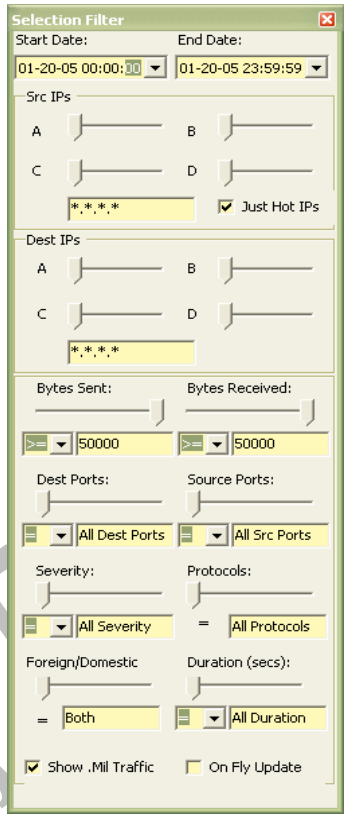

Figure 10 Selection Filter Panel

5.8 Reporting

Reporting is an extremely important feature that would definitely be part of a full prototype or final version. Much of an analyst's work involves gathering up supporting evidence to substantiate their claims of specific suspicious behaviors. It is envisioned that the conceptual system will allow an analyst to drag various pieces of information onto a report template reducing the analysts "tool" time and increasing the analysts' productivity. Since the type of reports required from analysts will change from site to site, the reporting mechanism should be very flexible. A mechanism should exist that allows for the analyst creation of new report templates that can be added to the system.

5.9 Packet Data (payload) Analysis

In the current demonstration version, inspection of the connection/session details were made available to the analyst through the connection/session dialog. The CTA also revealed that analysts like to investigate and study the individual packet information for the connections. A fully functional version shall have a packet viewer that shall allow an analyst the opportunity to inspect individual packets.

5.10 Drag and Drop Capability

In order to save the analysts as much time as possible, wherever appropriate, drag and drop functionality shall be made available. This way, if an analyst needs to put a specific source IP in

a report, he can just click and drag the source IP from one of the displays and drop it into the specific part of the report.

5.11 Range Filtering

The current display filtering (and selection) mechanisms of the system provide for quick trivial filtering of quite a few properties. One kind of filtering analysts will want to perform deals with specifying ranges of values. Ranges such as specific destination ports or specific set of protocols are often desired when filtering data. To accommodate for this, an advanced display filtering option shall be added to allow analysts to spend a little more time refining range type of display filtering. Any ranges that are set up shall be clearly displayed in the display filter status pane in order to keep the analyst aware of what filters are currently active.

DRAFT

6 Framework Concept Reviewer Feedback

6.1 General Discussion of Features

Discussions with ARL, JTF-NetDefense and JTF-J2 analysts were conducted during framework concept presentation regarding the features and benefits of the visualization framework as well as anticipated developmental and deployment challenges. Overall, the analysts' initial impression of the framework was that a fully implemented version would increase the analyst's situational awareness and be useful to the analyst. There was a great deal of appeal to having all analyst tools integrated together within a single flexible framework, and providing the ability to review data in a variety of different visualization techniques. The inter-connectivity between visualization tools providing for simultaneous propagation of data change throughout all the visualizations presented in the framework was highly desirable, as it was important to the analyst to be able to review the data from several different perspectives.

A significant focus of the discussion centered on the data volume problem. Presented with terabytes of data on a daily basis, these analysts are constantly searching for new methods for filtering down the data to something manageable. While they understand that handling voluminous data is a challenge for any analysis tool, nonetheless, the successful tool will require substantial and sophisticated data filtering capability to filter data down significantly before the tool can be considered effective. None of the sites perceived conceptual framework as a resource that had the ability to perform such complex filtering at this juncture. They all recommended that flexible querying and filtering mechanisms be incorporated into the framework to address the data volume problem and achieve significant data reduction.

Specific data sources analysts interviewed thought would integrate well into the framework were discussed. The analysts believed that NetFlow data would be a good data source, with the caution that NetFlow data never tells the whole story. Multiple data sources, such as the JCD, and data coming from multiple sensors, would be required in order to gain a complete picture of an incident or set of incidents. The ability of the solution to integrate with multiple data sources, both classified and unclassified, was a highly desirable feature.

NetDefense analysts expressed some specific concerns regarding the effectiveness of a visualization tool suite in their analysis environment. They indicated that they could realize some real value in using the conceptual framework if the it could be effectively used in two problem areas: The first area was reducing data volume through the use of automated data reduction techniques. This could produce a potential time savings of (up to) approximately 30 minutes per incident for the analyst. The second area was in supporting the type of analysis that NetDefense requires the most help with, which is when they don't know what they are looking for in the data. Typically, they either know what they are looking for, or they have an idea as to what they are looking for in the data. Using their current tool suite, by the time they have filtered down the data to something manageable, they are very close to an answer, and therefore the added value of for this type of analysis would be limited. If, however, the framework could

DRAFT

provide an effective means for visualizing the "I don't know what I'm looking for" scenario, the tool may have a good deal of value to the NetDefense analysts.

6.1.1 Summary of Visual Aides Usefulness

The Secure Decisions concept of the user interface visualization framework was presented to a body of analysts representing ARL, JTF-GNO/NetDefense and JTF-GNO/J2. The presentation focused on how conceptually the framework could help to address hard IA problems and provide insight and improved situational awareness when investigating various cyber incident (real and hypothetical) scenarios. Reviewing analysts were given a brief questionnaire, and asked their opinion on the usefulness of visual aids to increase situational awareness. Analysts were asked to rank the effectiveness of visual aids in specific focus areas, on a scale from 1 to 5, where 1 indicated they felt that a visual aide would not be effective at all, and 5 indicated they felt that a visual aide would be highly effective. While the response to the questionnaire was limited, the results of the evaluation are provided below.

	Score	Score	Score	Score	Score	Score	Ave.
Monitoring of Overall System Status	5	5	NF	5	5	3	4.6
Rapid Inspection of Large Volumes of Data	5	4	NF	5	5	5	4.8
Exploring Data for patterns or Trends	3	4	NF	4	5	5	4.2
Comparing Data from Two or More Sources	3	3	NF	5	5	5	4.2
Forecasting or Predicting	3	4	NF	5	3	5	4.0
Communicating Findings to Others	3	4	NF	4	5	4	4.0

NF = No feedback provided.

According to the survey results, the key value of visual aides to increase situational awareness was in the areas of rapid inspection of large volumes of data and monitoring of overall system status.

6.1.2 Summary of VIA Features Perceived Utility

During Secure Decisions presentation of the visualization and user interface framework concept, reviewing analysts were given a brief questionnaire, and asked their opinion on the perceived utility of proposed features that were discussed and/or demonstrated during the presentation. Analysts were asked to rank the utility of each feature on a scale from 1 to 5, where 1 indicated they felt that the feature was not useful, and 5 indicated they felt that the feature would be extremely useful. The results of the evaluation are provided below.

Feature	Score	Score	Score	Score	Score	Score	Ave.
Hyperbolic Tree	4	4	5	5	5	NF	4.6
Analyst e-Diary	4	4	NF	4	5	5	4.4
Dashboard	4	5	3	5	5	NF	4.4
Report Builder	3	NF	NF	4	5	5	4.25
Display / Selection Filter	5	NF	3	4	5	NF	4.25
Display / Selection Filter Status	NF	NF	3	4	5	NF	4.0
Scatter Plot	3	3	5	4	5	NF	4.0
Parabox	2	4	5	2	5	NF	4.0
Table Lens	NF	3	5	2	5	NF	3.75
Histogrid	4	2	4	4	5	NF	3.8
3-D SecureScope Frequency Wall	NF	NF	2	3	5	NF	3.33

DRAFT

Feature	Score	Score	Score	Score	Score	Score	Ave.
Line Chart	NF	NF	3	2	5	NF	3.33
Bar Chart	NF	NF	1	4	5	NF	3.33

NF = No feedback provided.

Several of the proposed visualization technologies were favorably reviewed by analysts. Specific comments regarding selected visualization technologies within the framework are discussed in the following sections.

6.2 Recommended Improvements to the Framework Concept/Feature Set

6.2.1 eDiary Recommendations

The concept of the eDiary, as discussed in paragraph 5.3, was favorably reviewed by both ARL and JTF-GNO NetDefense and J2 analysts. Additional recommendations to enhance this feature included:

a. Automating of chronologies within the diary, which would assist analysts in understanding the series of events that led up to a specific incident.
b. Consider connecting the framework to "Visual Links" (a competitor to "Analyst Notebook"). Doing so would allow the analyst to be able to treat data like a "blog".

DRAFT

6.2.2 Report Generator Recommendations

The concept of an automated report generator, as discussed in paragraph 7.1, was favorably reviewed by the analysts. Since reporting is an integral and time-consuming part of many analyst workflows, a mechanism for streamlining reporting and communication of results is highly desirable. Additional recommendations to enhance this feature included:

a. Provide the capability of tagging IP addresses to country flags, a feature similar in functionality to a tool known as "Analog"
b. Provide the capability to easily tie an alert back to the raw data, and provide this information in the generated report.
c. Provide the capability to easily collect and coordinate technical information from multiple sources (such as Excel, Advizor, raw NetFlow data, and other sources) into a Situational Awareness Report. This would be a good time saver and reduce errors by avoiding retyping of data.
d. Provide the capability of presenting summary statistics with a report.
e. Provide *Analyst Notebook* to support report generation

6.2.3 Visualization Recommendations

Analysts expressed interest in the framework providing the capability to visualize the following types of data that are utilized as an integral part of their investigations:

a. IP connections, FLAGs (including drill down), packet content
b. Interface to or share data with CMMA (a CJMTK product (commercial joint mapping toolkit))
c. NetFlows, firewall logs, and IDS logs

d. IP data
e. Provide a good visualization mechanism for figuring out Indications & Warnings, especially “low & slow” types of attacks.
f. De-clutter the initial views of the data.

6.2.4 Additional Recommendations

Specific recommendations as part of tool discussion consisted of the following:

a. The analyst will want to have the capability of declaring an IP or subnet as “hostile”. When declared, the analyst would want to be able to perform an historical search of the IP throughout the available data.
b. The analyst will want to have the capability to download lists of hot IPs (from an IDS for example). From there, the framework should have the ability to probe data over the past month, or other specified time period, to see if the IP is active again.
c. Outfit VIA with the ability to provide instant notification of a match of IPs. This would likely require searches through several different data sources.
d. Provide pattern modeling capability with rules notification
e. Provide support for real-time data visualization
f. Provide a system that pre-calculates link analysis data using the IPs in the alerts. This type of work could be accomplished offline, perhaps on a dedicated box. An analyst will likely want to see the communication patterns of the IPs in the alerts (predictable analysis behavior), so there would be a time savings if the system starts to work on this in advance of the analyst.
g. Provide ability for the analyst to define advanced SQL queries, advanced filtering mechanisms and drill-down capabilities such as :
    1. Elimination of payloads that are less than N bytes
    2. Elimination of scanning activity
    3. Filtering on specific FLAG settings, such as all “Pushes” for example
    4. Filtering on specific FLAG combinations in conjunction with specific time intervals
    5. Filtering on mil-to-mil traffic
    6. Filtering on mil-to-non mil traffic
    7. Filtering and querying based on subnets
    8. User-defined categories of data

h. Provide a capability where an analyst could select an IP address of interest, right-click on the address, and select from a group of (classified) text reports associated with the selected IP address
i. Provide a connection / cross cue from the framework to additional search engines
j. Provide a mechanism/tool for playing out hypothetical (attack) scenarios, similar in nature to the Secure Decisions-developed Excel spreadsheet/timeline presentation
k. Provide a connection from the framework to “Hermes”
l. Provide a connection from the framework to the JTAP database
m. Provide a connection from the framework to the JTID database. This would allow the framework the ability to perform a “JTID pull” on a specific IP address.

DRAFT

All organizations (ARL, JTF-J2 and JTF-NetDefense) discussed similar interests in the framework providing the ability to focus on a specific IP address, and finding all the activity (from multiple data sources) associated with that IP address.

Conclusion

The conceptual visualization framework is the result of a ground-up assessment of the needs and operating conditions of professional IA analysts working in live environments across multiple types of operations. The CTA study reiterated widely held beliefs surrounding IA analysis: that it is time-consuming work hampered by a lack of good visibility into massive data sets. Both tool function and flexibility are lacking for analysts, regardless of data type or operational environment. The framework was designed to address and solve those issues. It is in testing at JTF-NetDefense currently and is evolving toward the goal of Federal and commercial deployment.

DRAFT