AFRL-IF-RS-TM-2007-4
Final Technical Memorandum
February 2007

ATTACK ANALYZER: A NETWORK ANALYSIS AND VISUALIZATION TOOL

Advanced Technical Concepts

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

STINFO COPY

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE
ROME RESEARCH SITE
ROME, NEW YORK

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release by the Air Force Research Laboratory Rome Research Site Public Affairs Office and is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-IF-RS-TM-2007-4 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE DIRECTOR:

/s/	/s/
ANNA L. LEMAIRE Work Unit Manager	IGOR G. PLONISCH, Chief Strategic Planning & Business Operations Division Information Directorate

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
FEB 2007	Final	May 06 – Sep 06

4. TITLE AND SUBTITLE

ATTACK ANALYZER: A NETWORK ANALYSIS AND VISUALIZATION TOOL

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA8750-06-1-0035

5c. PROGRAM ELEMENT NUMBER
62702F

6. AUTHOR(S)

Russell L. Kahn

5d. PROJECT NUMBER
558B

5e. TASK NUMBER
II

5f. WORK UNIT NUMBER
RS

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Advanced Technical Concepts, Inc.
352 Ford Hill Road
Berkshire NY 13736-2135

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AFRL/IFB
26 Electronic Parkway
Rome NY 13441-4514

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFRL-IF-RS-TM-2007-4

12. DISTRIBUTION AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED. PA# 07-075

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The massive amounts of data that confront systems analysts as they monitor computer networks for security violations can be overwhelming. As a result, analysts may overlook critical details that may signal network break-ins or other system intrusions. This flood of data can consume systems analysts' time and lead to missed security violations and in extreme cases could lead to the complete collapse of a computer network or networks creating dangers to those who depend on them. The author addresses this problem with the development of a prototype visualization tool that attempts to clarify when a computer networks' security may be compromised. The tool itself, "Attack Analyzer" is described and the methodology and fieldwork testing used to create it are detailed. The tool uses a somewhat novel top-down, or deductive approach, moving from the general to the specific, rather than a bottom-up, or inductive method.

15. SUBJECT TERMS
Security violations, computer network security, visualization tool

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Anna L. Lemaire
U	U	U	UL	12	19b. TELEPHONE NUMBER (*Include area code*)

TABLE OF CONTENTS

LIST OF FIGURES

I. Introduction

Systems Analysts are faced with an overwhelming amount of data that needs to be monitored, analyzed, and acted upon in a brief time period. Martin L. Sheppard, a principal security engineer at AFRL noted in an interview that in just two days he was faced with reviewing and analyzing almost 65,000 records, in the form of 7500 rows of data grouped by such factors as IP address, signatures, alerts, violations, time, and severity level.

This paper describes a concept for a first generation "kinetic text" system using a multi-media tool. This visualization tool, referred in the rest of this paper as "Attack Analyzer," is based loosely around the Visalert intrusion detection system developed by Utah State and the University of Utah, although it switches the views of some of the variables and provides more of a "top down" approach, focusing on attack awareness as opposed to Visalert's "bottom up" approach, which concentrates on system alerts and lower level warnings.

II. Discussion of the Problem

Attack Analyzer is a graphical system designed to present data on potential attacks in insightful ways that provide multiple views for system analysts. The Attack Analyzer system advances earlier tools by utilizing attack analysis awareness tools being developed by the network fusion group at the Air Force Research Laboratory (AFRL) in Rome, NY, where the author spent nine weeks in the summer of 2006. Attack Analyzer allows for the manipulation of attack data in real time so as to provide detailed insights into packet-level network traffic and to provide insight into how to detect and respond to network security threats and to understand the pattern of such events.

Attack Analyzer expands on earlier network alert systems (e.g., Visalert, Rumint, IDS Rainstorm, Time-Based Network Traffic Visualizer) (University of Utah) adding real-time interactivity, addition of independent and dependent variables, and audio media with expert analysis and input. The prototype being developed is a Flash based, but final versions of the tool may be built with other animation tools, notably Java.

The objective of the visualization tool is to improve situation awareness, showing as many attack factors as possible and showing relationships among those factors. It provides a snapshot of activities at a point in time, with the user choosing the time frame and setting criteria for severity. In addition, audio inputs would be available for the user to use for journaling observations at points in time and at the end of each viewed period. The same system analyst could review these journals at a later time or another system analyst could review it on a later shift. In addition, expert analysis on typical sets of issues could be provided via audio outputs, with icons and related audio analysis could be provided at key intersections of events. Thus, for instance, pre-recorded expert advice might indicate to the system analyst what to look for when a set of criteria are present.

In *Understanding the Cyber Defender: A Cognitive Task Analysis of Information Assurance Analysts* (D'Amico, et. al), system's analysts' methods were studied and were determined to fall into three levels:

1. Data is associated with suspicious activity,
2. The data is further reviewed to find patterns of suspicious activities and to determine the severity of the attack,
3. Finally inferences are drawn regarding the vulnerability of the target and the extensiveness of the threat.

This prototype is designed to take system's analysts directly to Level 3 allowing them to immediately draw inferences regarding the severity of an attack and to consider methods for responding to it. The use of filters is meant to allow analysts to further refine and focus concerns and thus return to Level 2 to find patterns of suspicious activities and to determine the severity of the attack.

III. Methodology

Attack Analyzer when through three phases of development, with each phase ending with an in-depth review by the network fusion group at AFRL. A discussion of each phase, with a review of the discussions that led to the succeeding prototype, follows:

Phase 1. The initial view, as shown in Figure 1 involved a W^3 vision, indicating where, when, and what information regarding the attack. The tool involves multi-media views with sight, sound, and motion,

provided multiple views using mouseovers and mouse clicks; allowed for asynchronous and synchronous observations, interactivity, rule-based clustering to indicate major threats, and drill downs to data. Additionally, it provides for journaling and expert advice via audio inputs.

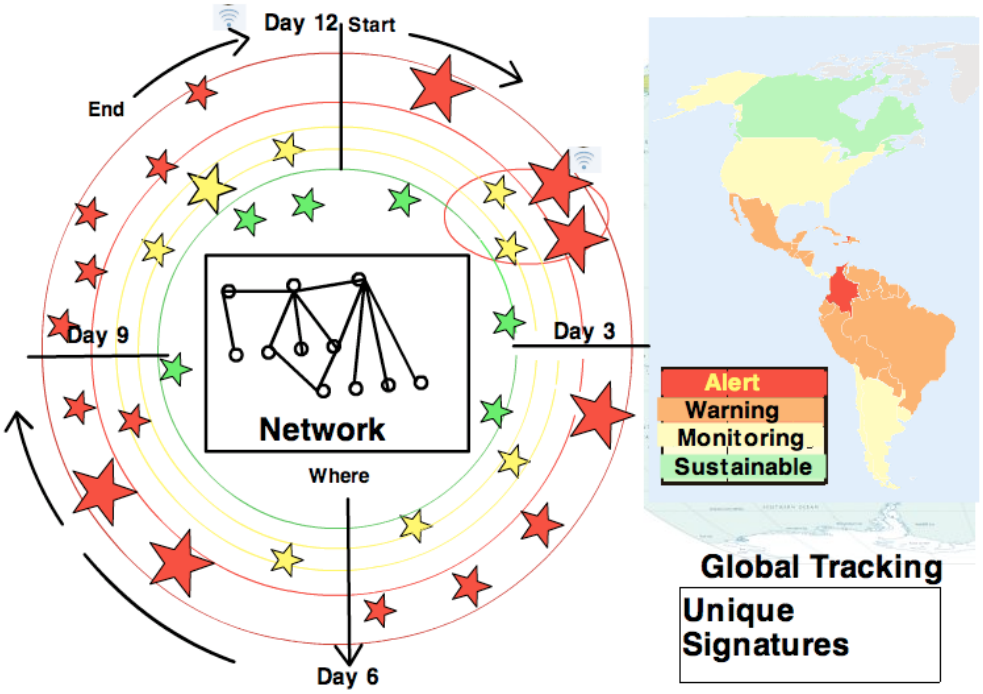

Figure 1: First Prototype: A Bottom Up Approach

The first module provided a three-dimensional variable awareness –when, where, and what. These three dimensions are considered most critical for developing situational awareness (Foresti, Agutter, Livnat, and Moon, p. 52). The graphic interface design is based around the concept of assigning graphical attributes that provide the most relevant information for developing situational awareness in an interactive, dynamic format.

These three variables may be defined as

When – The point in time when the alert was noted
Where – The local network node that the alert refers to (e.g., an IP address)
What – The type of alert (e.g., snort, ethereal, TcpDump) and the severity level of the alert

By providing a three-dimensional analytical tool we can now visualize alert instances across a continuum. Adding an interactive component allows system analysts to uncover relationships, meanings, and underlying issues. This systematic visualization tool reflects the idea that effective data representation should be consistent with users' cognitive representations. Conversely, failure to use perceptual principles can lead to wasted effort, lost time, and lack of success in determining the cause of an alert and in developing an appropriate response.

"Where" was indicated via a network topology in the center of the screen, with icons indicating network interconnections. To the right of the screen and outside the main diagram was a world map indicating the location of the intruder in a geopolitical space. The map also indicated the degree of stability for the country that was the source of the intrusion, broken into "Alert," "Warning," "Monitoring," and "Sustainable" components. A country's degree of stability was color coded, red, orange, yellow, and green.

"When" was indicated on the main "stage" through the concept of an analog "clock" with the user moving through time by moving around the circle, based on the location on the circle. In the sample view, 12 days were displayed with time broken up into four quadrants, each representing activities over a three-day period. Thus, the location around the "clock" quickly suggested key points of activity.

"What" was indicated by both an icon for the attack and the star's size and color and the location of the star along concentric rings. The size indicates volume of the attack, color indicates the severity of the attack, and the location on the concentric circles indicates the type of attack (e.g., ftp, torrent bit stream, etc.). The severity is determined by the systems analyst, who might consider such issues as signature, source, methods, or location on the network. In addition, the system would include oval clusters that would group a series of attack, thus indicating relationships among the activities.

In the field test and discussion that followed the completion of the first prototype the group suggested a reduction in the reliance on time and physical space and a move away from alerts and toward an emphasis on attack methods, confidence and certainty metrics, and the sequencing of each attack. It was noted that within the group, attacks are measured by the stage that they are in, rather than by time or location of the intrusion. It was also noted that a geopolitical view of the assailant (via a color-coded world map) could be problematic as intruders' locations cannot be easily determined and are often disguised because attackers often commandeer a rogue site that they break into and use for subsequent attacks.

The key issues noted in the discussion – a focus on attacks and the stages in an attack and use of a "ground truth file" generated by the fusion group. They recommended a top-down approach, as seen from the fusion development group and focusing on known assailant's techniques (such as reconnaissance) and data that has been processed and developed into a taxonomy rather than the bottom-up approach, based on alerts and raw data, which is often the view seen by system's analysts. Thus, subsequent prototypes involved more of an deductive approach, moving from the general to the specific, as opposed to the initial prototype, which was an inductive view, moving from specifics and determining more general themes based on that information.

Phase 2. Prototype 2 (as shown in Figure 2) focused on stages of an attack, from the taxonomy described and illustrated in "Hacking Exposed, Network Security Secrets & Solutions" (McClure, Stuart; et. al)."

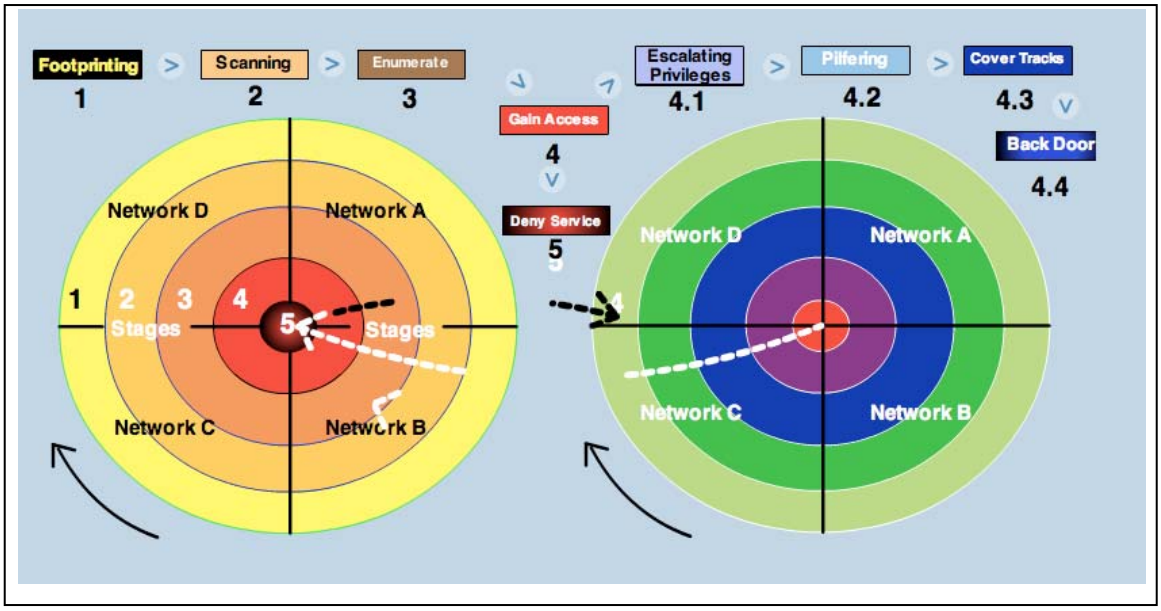

Figure 2: Second Prototype: An Attack Scenario and a Top-Down Approach

The initial view included an integrated set of two concentric circles representing stages in an attack. The stages included were recommended by my point of contact and involved a series of actions along with a loop-back based on what actions were taking place. The concentric circles indicated the stage of the attack with the center circle indicating access to computing system and denial of service to other users.

Moving around the circle indicated confidence levels of each attack, or a sense of whether the attack was a true attack or a false positive – a misidentified attack.

The two circles were distinguished by showing an initial attack in a circle with stages in increasingly "warm" colors (yellow to dark red) and the second set of potential stages shown in increasingly "cool" colors (light green to dark purple). The one exception was a red circle in the center of the second set of attacks, indicating that the attacker has gained access to the network. This redundancy in colors was meant to indicate that an identical stage could be reached via these techniques as could occur in the first set of attacks, shown in the circle on the left.

The circles are further broken down to four quadrants with each quadrant representing a different level of network, most to least important. The person using the tool could define these networks. Thus, the user could determine how to distinguish the four networks so that it provided the most value in visualizing potential intrusions.

Eliminated in this prototype was any specific reference to time, alerts, or geopolitical locations. The focus was on the stage of the attack at a point in time, as determined by the Network Fusion group. The circles provide a situational awareness at a point in time with the focus on the stage of the attack, the confidence level that an attack is truly occurring and the series of stages that an attack has proceeded through. Prototype 2 showed three main variables: What, Where, and How. What was show in several ways: Stages of an attack are shown with circles located at the stage and by the size of circles, with size representing the volume of a particular attack. In addition, the position in each quadrant indicates the confidence level. Where is shown via the indications on what Network (or Networks) are being effected. How is designated by showing all techniques involved in the current attack.

The discussion following the second full field test1 focused on a more micro view of the tool. For instance, it was agreed that instead of showing all phases of an attack it would be best to show the latest stage of each attack that is being studied with further "drill downs" to more detailed information, such as earlier stages shown in the evidence file on mouse "click downs." In addition, a need for a robust set of filters was requested, so that the system's analyst could see multiple views, depending on his/her interest or concern. To further simplify the view of potential attacks it was agreed to eliminate colors from the concentric rings and add colors to the circles that represent potential attacks. Colors in the circles would represent the "score" or confidence level, gauged on a range linked to the scores recorded for each attack, with values ranging from 0.1 to 1.0. Colors would be the same as those used by the Homeland Security Advisory System:

Color	Homeland Security Value	Score or Confidence
Red	Severe	.90 -1.0
Orange	High	.80 - .89
Yellow	Elevated	.70 - .79
Blue	Guarded	.60 - .69
Green	Low	.01 - .59

Phase 3. The third prototype synthesized the discussions, literature reviews, and design considerations. This tool simplifies the visualization further, in part by reducing stages to the four types identified in the fusion group's list of evidence – reconnaissance, intrusion, privilege escalation, and achievement of the goal (usually denial of service). The final prototype eliminated all loop backs, replaced colors for each stage for gradations of grey, and added colored circles reflecting the score of each attack. The resulting tool is shown below in Figure 3, which shows the view when a user has moused over a circle (attack).

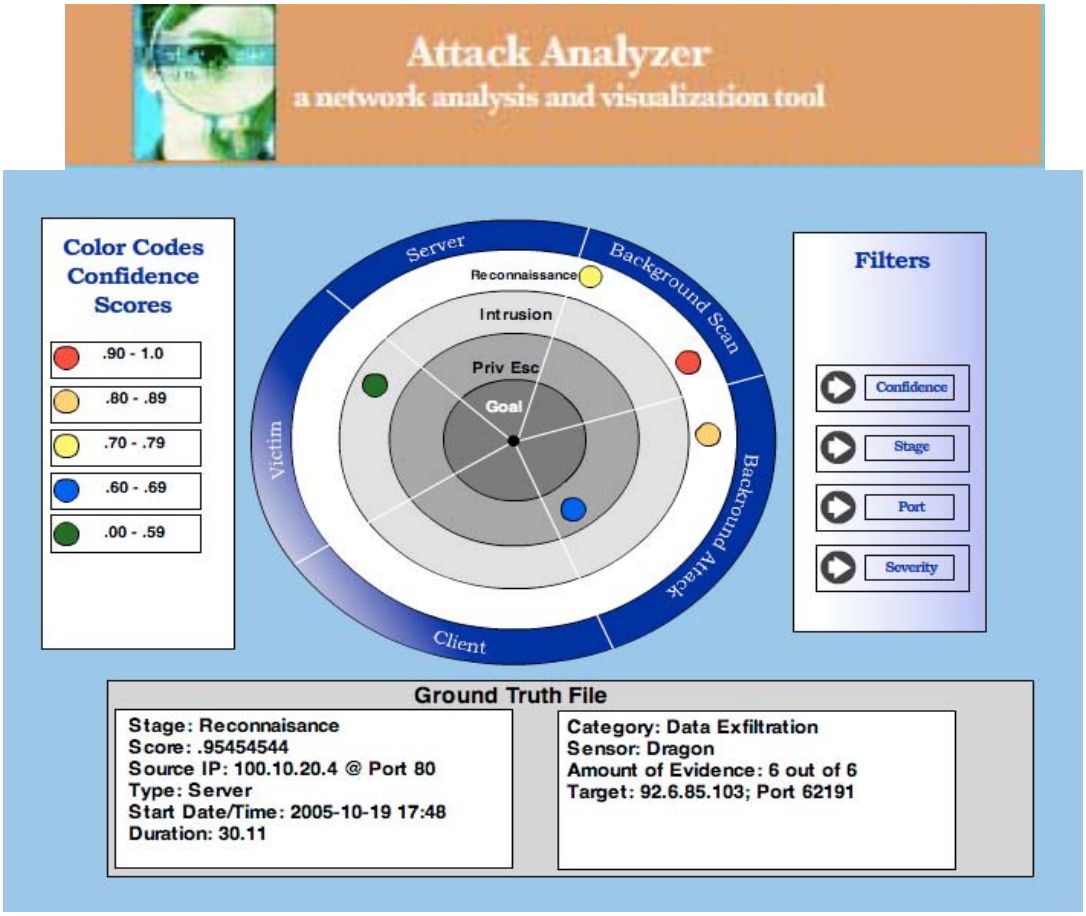

Figure 3: Prototype 3; A Simplified Attack Scenario with Filters

IV. Results

Visualization is based on the concept that less is more. That is, that by honing in on just those areas that are of interest to the user (in this case the system's analyst) he or she will be able to make more sense of the data and be able to respond to it more intelligently and more intuitively. which determined that the IA analytical process moves through three stages of situational awareness: perception, comprehension and projection.

Perhaps the most serious concern of system's analysts is that they are overloaded with information at a time when they can least afford it – when an attack is taking place. The idea of a visualization tool is to do the analysis and disagregation ahead of time, so that when an attack takes place it can be immediately perceived and assessed. It's important to cut to information relevant and important and remove data that is not critical to an analysis. This allows for analytical activities ahead of time and allowing for intuitive actions at the time of an attack. If you get too caught up in generating information you can drown in the data (Blink).

In *Understanding the Cyber Defender: A Cognitive Task Analysis of Information Assurance Analysts*, systems analysts methods were studied and were determined to fall into three levels:

1. Data is associated with suspicious activity,
2. The data is further reviewed to find patterns of suspicious activities and to determine the severity of the attack,
3. Finally inferences are drawn regarding the vulnerability of the target and the extensiveness of the threat.

Audio Cues. Although it is not shown in the final prototype, in addition to the interactive mouse tools there could be a series of audio cues available throughout the system. The audio could be played by any mp3 player (such as an ipod) or over the speaker system on a computer. Audio would be used in three formats:

1. At the outset an audio cast would be made available (perhaps through a podcast device – ipod or other mp3 player) or built into the Flash file, to explain strategies for using the visualization tool. This would include common analysis techniques for different network topologies as well as techniques for troubleshooting specific issues that a network administrator may already be facing, such as virus intrusions, torrent streams, and system slowdowns. Thus, a system analyst would be provided with strategies for analysis before the analyst starts the program. These audio cues could be provided by experts in each field thus providing input at an expert level not otherwise available.
2. During the exploration of the visualization tool the user could click on mp3 icons to get further information about a specific alert and some issues to look for when these alerts come up on the screen. In particular, patterns could be described that indicate particular problems. The cues could also provide ideas for handling various situations once a pattern is discovered.
3. At the end of the exploration, the user can create a new audio log file, summarizing the current situation, what's been noted, what analysis has found thus far and notes regarding patterns and concerns of note. The audio file would end with a description of needs for pattern matching in future reviews and thoughts about needs for future analysis. In particular, the analyst may note thoughts about extending the time line forward or needs for looking backward ("backfilling").

Drill Down Capabilities and Filters. A second generation tool would allow for data drilling by clicking on a W^3 variable and generating "front end" statistics such as IP addresses and actual locations (starts and finish). Graphs might also be generated allowing for manipulating data at all levels to provide rapid answers to ad hoc queries. In particular, this would allow for determining a fourth key variable – "who" and for generating detailed data views. At a basic level, a user could click at any point on the viztool to find more specific data behind each moment in time and alert.

Synchronous with the movement of the mouse over time, type of alert, or location would be correlated movement in the two gauges and graphical bar chart at the bottom of the interface. Thus, dynamically the SA can see changes in a number of variables across several

Visual Devices – Iconography and Design Theory. The visual display of quantitative data is a challenge with the ultimate test involving what is the most illustrative of the issue without being chaotic. Key issues involve representation of the data (Berryman):

- Use of icons
- Use of color
- Use of space
- Placement of data
- Movement across the space
- Navigation schema
- Use of contrast (simultaneous, hue, saturation, value)
- Use of perspective, layout, grid system

In addition, key considerations would include collection of data, methods for providing audio cues, content of audio cues, methodology for creating and displaying information. The general look and feel of the tool should incorporate basic design elements of emphasis, balance, rhythm, and unity.

Design creates meaning not only through the work itself and where elements are placed, but through a complex social relationship that involves at least two elements besides the element itself and the designer: 1) how viewers interpret or experience the design and 2) the context in which an design is seen.

The approach to the Attack Analyzer applies a design interface that considers a number of the key aspects of Human Centered Design (Cooley).

Coherence – Embedded meanings clarified
Inclusiveness – Inviting

Malleability – Ability to sculpt the environment to suit one's needs
Engagement – A sense that one is being invited to participate
Ownership – Created and own parts of the system
Responsiveness – System responds to your needs
Purpose – System capable of responding to user's purpose
Panoramic – Windows through which one can take a wider view
Transcendence – User should be engaged, enticed, provoked to go beyond

V. Conclusion

In addition to the prototypes shown in this paper, the final tool would include a reasoned-based front-end model to the tool explaining the analytical basis and underlying concepts behind the diagnostic tool. The front end was designed to run as an HTML based website utilizing extensive ActionScript (the Flash programming language) and basic visualization and animation tools inherent to Web design. The front end includes three components:

1. An introduction to the tool
2. A breakdown of the toolbox components
3. A disagregation of the analytical methods with in-depth discussions of each

Because the objective of this project was to create mockups and an initial design, it isn't possible to provide conclusive answers to these questions. Most critically, the visualization tool does not, in its present configuration, contain working linkages to the fusion group's tracking system. Thus, full-scale in situ field tests are not yet possible. Thus, for example, it is not yet viable to determine, with certainty, whether the system will hold up when used in the field under stressful situations, in difficult climatic situations and in multi-cultural ad multi-ethnic environments. However the theoretical underpinnings of the model, based around human-centered design principles and concerns noted in the literature – confidence, purity, cost utility, and timeliness – would indicate the potential for a positive correlation in these areas.

However, discussions, interviews, development of models and field testing amongst a small control group has provided important information about the composition of a kinetic text system and reasons why it might improve understanding of and clarification about key relationships. Only a large scale field test with a working visualization tool can determine its true potential value.

The visualization tool allows organization and illustration within a key dimension and provides access to relational and multidimensional data. As you move down a hierarchy the tool can reveal patterns and illuminate relationships. The tool allows flexibility for exploration at any level that interests the analyst. With this data drilling software the user selects database information from any category. In particular, the tool provides for a number of elements not seen thus far in other tools: real time dynamic interaction, front end details and descriptions including analytical techniques, use of audio inputs from the analyst for journaling and for handoffs of the tool to the analyst following him/her at the helm of the network. Output audio feeds from experts in the field providing in-time tips and techniques at the point that the tool is being used. Thus, analysts can be viewing the data while hearing (via headphones or speakers) information on what to look for on the screen.

The purpose of the profiler tool is two-fold, on the one hand it would look at just the intruder's activities on the network – all activities emanating from their IP address. Thus an analyst could develop a better sense of the intruder's patterns, proclivities, methods of attack, and computing resources. Since a key concern for any tool is avoiding false-positives, this should improve the chances of determining the level of the threat. In some situations it might be possible, using this ethnographic model, to determine who the individual might be and who controls him/her and thus create a portfolio of the individual, complete with data about her/his education, methods, organization, motivations, attributes, and resources. In some cases, a timeline of expected future activities might be generated based on the portfolio.

On the other hand, using pattern-matching techniques, "Attack Analyzer" might also be used to look for other predators with similar profiles, thus potentially connecting networks of intruders who may be working together or under the same umbrella organization. A profile might eventually include enough detail about the organization – social, military, economic, and political, to provide for a military response to the cyber threat. The level of threat could better be determined by such factors as whether the threat is from a criminal group, foreign intelligence group, hackers, insiders, a terrorist cell, or some other attacker.

References

Berryman, Greg. 1990. Notes on Graphic Design and Visual Communication. Crisp Publications.

Cooley, Mike. 2000. Information Design: Human-Centered Design. Cambridge, Massachusetts: The MIT Press.

Creswell, John W. 1998. Qualitative Inquiry and Research Design: Choosing Among Five Traditions. Thousand Oaks, CA: Sage Publications, Inc.

D'Amico, Anita; Daniel Tesone, Kirsten Whitley, Brianne O'Brien, Emilie Roth; "Understanding the Cyber Defender: A Cognitive Task Analysis of Information Assurance Analysts" Report No. CSA-CTA-1-1 June 2005.

Stefano Foresti, James Agutter, Yarden Livnat, and Shaun Moon University of Utah, "Visualization for Cybersecurity" IEEE Computer Graphics and Applications; March/April 2006

Huxley, Lesly and Jacobs, Neil. 2002. "From static content to dynamic communities: The evolution of networked educational resources." Online Information Review, 26 (1): 19-30.

Light, A. and Wakeman, I. 2001. "Beyond the interfaces: users' perceptions of interaction and audience on websites." Interacting with Computers, 13 (3): 325-351.

McClure, Stuart, Joel Scambray; George Kurtz, 2003 4th ed. "Hacking exposed: network security secrets and solutions; Berkeley, Ca, McGraw-Hill/Osborne.

Salerno, Dr. John J., George Tadda, Douglas Boulware, Michael Hinman, 2006. "Achieving Situation Awareness In A Cyber Environment."

University of Utah, "Intuitive Visual Representation of Network Events to Improve Decision Making"; ARDA P2INGS PROJECT Center for the Representation of Multi-Dimensional Information. www.cromdi.utah.edu