

User Issues in Visual Monitoring Environments

Robert F. Erbacher
Department of Computer Science, LI 67A
University at Albany-SUNY
1400 Washington Avenue
Albany, NY 12222, USA
erbacher@cs.albany.edu

Abstract

Visual monitoring environments, such as intrusion detection systems, debugging environments, and feature extraction systems, require that a user familiar with the target domain examine, most often continuously, the visual representation of the underlying data. This improves the efficiency of the analysis but requires that the visualization expert work with the user to provide the information in an efficient form. How the users employ the environment and the type and quantity of data will also affect aspects of the environment. The goal is to develop a user centric view when designing the software and meet the unique needs of the user at hand. Our work with intrusion and misuse detection has led to the need to develop techniques geared for these users. This requires us to give up some typical metaphors familiar to visualization experts that would not be acceptable to the expected user base. We will discuss the issues involved in developing visualization techniques when the user is not a visualization expert, has preconceived notions or expectation of the visualization environment, and has needs that fall outside the normal expectations of the visualization expert.

Keywords: Visualization, Perception, Large Scale Database Visualization, On-Line Monitoring

1. Introduction

Most environments developed by or with the aid of visualization experts are ultimately designed to be used by non-experts in their daily tasks. Every application will have its own particular requirements, however, there are basic philosophies and requirements that must be examined to meet the basic needs of the target user, including:

- How will the environment be used?
- What are the user's expectations?
- What are the user's preconceived notions?
- For what duration will the environment be used?
- How often will the environment be used?
- What will be the users primary task or focus?
- How much data will the user need to monitor?
- What types of interaction will the user want or need to perform?

Often the issues raised by these questions will contradict each other. In such circumstances, the environment must meet the minimal needs of the user, provide for the most important characteristics, and allow the user to change the focus of the application to meet the needs of the moment. Linked multiple views are effective at handling the constraints of such systems [18].

This will require that the environment be configurable and highly adaptable at runtime.

Our recent work has focused on the development of capabilities for use by non-experts in intrusion and misuse detection, link and feature detection, and program debugging. Each of these areas has interrelationships with the other areas, however, the goals of the problem domain are completely different, as are the target user base and the expected use of the environment. Consequently, we have had to deal with many of the above questions and developed very different environments and capabilities to deal with the differing needs of the user for each problem domain.

2. Previous Work

The prior work of Erbacher et al. [6, 5] describes the basic visualization environment used in this research. As we continue to analyze the data collected by the Hummer intrusion detection system [17] and other available data we have realized the need to focus the visualization techniques on the needs of the expected users. In particular, our users are systems administrators or security experts. In the case of system administrators, the users will be monitoring the network environment continuously while performing other duties. Thus, the visualization environment will not be their primary focus. Rather they will be concentrating on other tasks and will need to examine the visualization environment at intervals to determine the activity on the network. This means the environment should consist of a small window in the corner of the system administrator's window, similar to how many users display the system load and a clock. The visualization techniques employed must clearly display the network environment in such a small display.

There are four principal issues that need to be considered when developing an environment designed for users who are not visualization experts. This includes task-based visualization techniques, user-based visualization techniques, human perception, and preconceived expectations, such as cultural metaphors.

While there has been quite a bit of work on perception and the human visual system [19, 12, 14, 16, 9, 10] that identify characteristics to be used as a basis for the development of visualization techniques, little has been done to actually compare and contrast visualization techniques. Little work has been done to compare visualization techniques in a formal setting. The information exploration shoot-out, organized by Georges G. Grinstein and supported by the National Institute of Standards and Technology (NIST) [11], provides the only attempt to develop a body of knowledge that identifies what visualization techniques definitively work better on a given data set. The idea was to identify which researcher's techniques were effective at identifying intrusions in a network data set. Beyond this limited work and the work on perception there has been no examination of the needs of the user or the task under analysis. There have been extensive studies on user interfaces, how appropriate different user interfaces are and how to apply them. Similar such studies are critical for the promotion of appropriate visualization techniques and environments.

Most visualization techniques are developed with little thought to the end user, especially when the end user is not a visualization expert. This is particularly a problem with information visualization in which the data representations tend to be far more abstract than in scientific visualization. In scientific visualization the data itself provides a basis and mechanism for analysis. This foundation does not exist in information visualization. As the quantity of data needing more abstract representation increases (e.g., census data and network data) the need for a better understanding of the needs of such environments is critical. Our situation is complicated by not only the need to develop techniques effective for the given task but we must also develop techniques that will be applicable to the user base.

If we take network visualization as an example, we must consider the needs of the user and the task at hand. While there are many techniques that are applicable for representing the information, some are more suitable for the user than others. Generally, most users conceptualize networks as a set of nodes and links. While any representation is better than the basic log files to

identify an attack [1]. These log files can incorporate hundreds if not thousands of messages per day. Maintaining the typical network representation metaphor aids the user in understanding the data and the resultant analysis. Most previous work involving visualization related to networks has emphasized graphics that depict network performance and bandwidth usage [2, 13, 3, 15, 19], even down to the router [3], individual packets [8], and individual e-mail messages [4]. The techniques developed for these purposes do not provide sufficient detail or handle sufficient numbers of nodes and attributes in combination for our needs. The work by Eick et al. [4] strictly deals with e-mail and subsequently resolves many fewer nodes and attributes than is needed for intrusion detection. Other work has been geared towards visualizing systems for program analysis and program development [7]. These environments typically deal with small numbers of processors that are working on a single task and thus have a common grounding. This research into network usage has not been applied to network intrusions. It does, however, provide a starting point.

Currently, system administrators are required to analyze log files to detect and locate system intrusions and misuse. The amount of data available results in system administrators not fully collecting or monitoring all available information for all systems under the administrators control, rather focussing on primary systems and servers and examining minimal information from the remaining systems. In fact, system administrators in general do not collect or analyze any data related to Microsoft or Apple based operating systems at all; even though it is possible for these systems to be the targets of break-ins, subversion, and misuse. Network traffic itself is only analyzed intermittently. Log file analysis is becoming the greatest time consumer for system administrators. If they were to analyze all the available data they would truly be fighting a losing battle as more information is generated than can possibly be analyzed in a textual format during a reasonable period of time. Identifying actual intrusions and misuses requires that the intentions of the user be known during examination of the user's activity. This is currently unfeasible and results in missed attacks and many false alarms. This situation is

only likely to get worse and with the globalization of e-commerce and interest in Internet voting the potential for serious damage increases. Ultimately, the goal must be to identify attempted break-ins or attacks before the attack is successful so the situation can be monitored and a response initiated before damage occurs. Current log file analysis only reveals that an attack has occurred in the past. At this point, it may not be possible to determine if the attack was successful or not since hackers generally subvert the log reporting facilities as one of their first actions. This leads to extensive amount of analysis being required to determine the integrity of each system. It is imperative that we reduce the number of false alarms and increase the number of true attacks detected. This provides us with a direction in which we can aid the user base. Clearly, real-time analysis is needed which can aid the system administrator in identifying intrusions as they are occurring and provide sufficient information to determine if an intrusion has already occurred or may be likely to occur. We must also provide information representing complex interactions and behaviors that can be indicative of a break-in after an attack has been successful, something not directly identifiable through log files.

3. Task Focus and Large Scale Database Visualization

In intrusion and misuse detection [1] our target users are system administrators and security experts. In this scenario, we principally have non-experts using the visualization environment to monitor activity on the network. This is just one of many tasks the user is performing so the visualization environment will not be the primary focus of the user. It must also be persistent for continuous analysis. Generally, we need an environment and visualization techniques that allows the needed information to be presented in a fairly small window, incorporating it into the administrator's already extensive tool chest. However, depending on the number of systems under the system administrator's auspices there can be an enormous amount of information visually necessary for complete analysis. The main computer system infrastructure for our

university consists of one primary UNIX system, several dozen non-primary UNIX workstations, and hundreds of other remote connects, all of which must be monitored. This can lead to the need to visually represent hundreds of nodes and links simultaneously, generally a task that most visualization environment would allocate to the entire screen. As mentioned, however, we can only allocate a small portion of the screen real-estate to this single task. Clearly, the problem is in and of itself a large-scale database problem, an open and critical issue in the visualization field [21]. However, the problem is complicated by domain needs. This problem can be solved in part by providing a multilevel approach.

At the basic or scaled down view the user will be unable to “perceive” all the details inherent to the analysis process. Sufficient detail must be provided in the scaled down window in order for the user to determine when “interesting” activity is occurring. This must be done very quickly, using perceptual characteristics, as the user will likely just glance occasionally at the visualization environment or use VCR-like controls to view the recent history of the environment to determine if anything has occurred of interest since they last reviewed the information. With intrusion detection, timeliness is critical as any delay in detection provides time for damage to be incurred on the system.

When activity of interest has been determined, the user will need to examine the data in greater detail. Scaling up the window should allow more features and attributes to be enabled, allowing the system administrator to analyze in greater detail activity on the network. The visualization techniques can be modified themselves as well. At the reduced level, information related to non-critical systems can be reduced and condensed, any remote connections at all should be a wake up signal so the details of those connections are unimportant. The critical systems, however, will need constant monitoring to identify the meaning of the numerous connections. The principal UNIX system for the university has ~150 connections ongoing during the afternoon. This is reduced to ~30 connections late at night, principally connections from local individuals logged in continuously and consequently do not represent truly “active” connections.

4. User Based Metaphors

A second issue is the preconceived notions of the user and what they are expecting and willing to use in a visualization environment. There are many ways of creating a visual representation of the network topology and the interaction between systems. A two dimensional connectivity grid is applicable as well as glyphs that include node connectivity incorporated as part of the glyph. However, these techniques are not as acceptable to system administrators. They have preconceived notions of the network and the interaction among the systems. They require a familiar model in the visualization environment. Changing this model would not be readily accepted. Even should we develop techniques better suited for the representation of the intrusions and misuse data; if they are unfamiliar to the system administrator or deviate too much from the expected model then the visualization will ultimately not be used. Our current representation uses glyphs representing the individual systems with directed lines showing activity from one system to another [6, 5]. This connectivity relates to virtual connections over the network infrastructure and is not related directly to the hardware infrastructure. This is acceptable to system administrators as it is an abstract metaphor they are familiar with. System administrators will often draw diagrams of the network using a similar representation. However, the difference between the hardware connectivity and virtual connectivity must be made distinct and clear or the visualization could be misinterpreted. Therefore, it is critical to visually represent the locality of the nodes. Systems that are local to the network and are being monitored by the system administrator must be maintained distinct from external nodes for which limited information may be available.

Application of techniques not based on the user’s needs and expectations would clearly fail in this scenario. These results drive the need for a better understanding of what techniques work better on given types of data and what types of techniques are better for a given user base. The intersection between these two sets can then be examined for solutions to a given problem task.

5. Culturally Based Metaphors

An area closely related to user based metaphors is that of culturally based metaphors. As users with different training will perceive visualizations differently so to will users from differing cultural backgrounds. It is well known that different cultures perceive colors differently and have other ingrained concepts that must be taken into account if a tool is going to be deployed globally. This may not be an issue if the user base is expected to have an expertise in the field and will know the intended meaning behind the provided representation. It is when the user base has had no particular training in such differing metaphors that the intended meaning of a visual representation can be lost and misinterpreted.

This clearly shows the need for customization of the environment such that different default configurations can be used based on the intended audience. This will also aid in resolving any other local issues that may be present, such as individuals with color blindness.

6. Perception and Information Analysis

The last topic of interest to our discussion is the need for the inclusions of perceptually oriented techniques within visualization environments. Perception has become recognized as being of critical importance within the visualization community. However, the visualization techniques as of yet have not been applied in a substantive way [18]. Perception is discussed in most visualization papers but mostly as an after thought. Perception theories are not incorporated into the foundation of the visualization techniques. User studies are talked about but rarely performed.

Perception needs to be taken more seriously and integrated into the early development of visualization techniques. Combined with user studies, the use of perception should greatly improve the applicability of developed visualization techniques, tools, and methodologies.

In our work, the reduced view of the network visualization must present important aspects of the

network infrastructure to the user in a glance. Many of the features will be very reduced in size. The details will be limited when presented in this form. Keeping in mind that with the other tools absorbing the system administrator's attention to our tool will likely only get a glance every now and then. The tool must display important information in such a way as to draw the user's attention when events of interest occur. The activity on the network for the most part will be very dynamic. This flickering of continuous and energetic activity must "not" draw the attention of the user unless something untoward has occurred.

To this end, the environment for the most part represents details in pastels [20], which provide the information in an effective way but do not draw the user's attention. When user set limits occur on activity or when unusual or unexpected events occurs then brighter colors are used to display the information, making the events stand out in the corner of the user's eye and drawing their attention. Basically, the idea is to reduce the contrast provided in the visual display, except when an important event occurs, so as not to weary the user or unnecessarily distract said user.

7. Example

Figure 1 provides an example of the visualization display currently under development for the network intrusion and misuse detection environment. The display shows some of the difficulties inherent to the environment. The number of connections occurring simultaneously is rather high, upwards of 150 at times. This view only shows a single system in the network being monitored. In general, all the systems under the auspices of a system administrator would need to be monitored.

The connections include anonymous ftp's, privileged ftp's, and telnet connections. Since a single event in and of itself will not be sufficient to identify the ultimate goal of a user we must provide massive amounts of information to the system administrator such that the ultimate goals of the user can be determined.

The overall activity is comprehensible without much difficulty. As no serious events are occurring the environment has no glyphs that

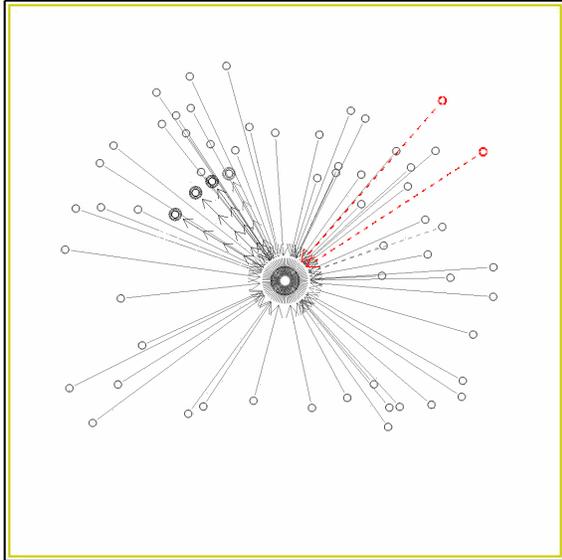


Figure 1: High activity on the University's main computer system. The nodes are dimmed so questionable activity can be given greater emphasis.

stand out significantly from the rest of the display. The red nodes are systems for which a reverse hostname look up failed. This is important but not critical so it is differentiated but not too strongly. We can use brighter reds and thicker lines for systems that it is critical to draw the user's eyes to.

8. Conclusions

Many of the issues discussed in this paper are applicable to a wide range of problem areas and data sets, particularly those requiring constant monitoring and those used by non-visualization experts. Clearly, there is much work to be done in the promotion of the discussed issues of perception, task based visualization, and user based visualization. While some of these topics have been discussed frequently in the visualization field, they have not been taken to heart and worked on extensively. This needs to improve if visualization is to progress beyond the basic research being done today and become a mainstream tool for the exploration and analysis of data.

Greater effort needs to be placed on developing user centric views to the development of visualization environments. Merely having a

capability is insufficient. That capability must be usable. This requires that the visualization expert incorporate techniques specifically geared towards aiding the user in their particular task and designed for their comprehension, even if the visualization expert does not like it.

References

1. Rebecca Gurley Bace, *Intrusion Detection*, Macmillan Technical Publishing, 2000.
2. Richard Becker, Stephen Eick, and Allan Wilks. "Graphical methods to analyze network data," In *IEEE International Conference on Communications ICC '93 Proceedings*, Geneva, Switzerland, pp. 946-95, May 1993.
3. Kenneth Cox, Stephen Eick, and Taosong He, "3D geographic network displays," *ACM Sigmod Record*, Vol. 25, No. 4, pp. 50, December 1996.
4. Stephen G. Eick and Graham J. Wills, "Navigating Large Networks with Hierarchies," In *Visualization '93 Conference Proceedings*, San Jose, California, pp. 204-210, October 1993.
5. Robert F. Erbacher and Deborah Frincke, "Visual Behavior Characterization for Intrusion and Misuse Detection," *Proceedings of the SPIE '2001 Conference on Visual Data Exploration and Analysis VIII*, San Jose, CA, January, 2001, (To Appear).
6. Robert F. Erbacher and Deborah Frincke, "Visualization in Detection of Intrusions and Misuse in Large Scale Networks," *Proceedings of the International Conference on Information Visualization '2000*, London, UK, July, 2000, pp. 294-299.
7. Robert F. Erbacher, "Visual Assistance for Concurrent Processing," University of Massachusetts at Lowell Doctoral Dissertation (1998 cs-3), Lowell, MA 01854.
8. Deborah Estrin, Mark Handley, John Heidemann, Steven McCanne, Ya Xu, and Haobo Yu, "Network Visualization with Nam, the VINT Network Animator," *IEEE Computer*, Vol. 33, No. 11, pp. 63-68, November 2000.

9. James J. Gibson, *The Senses Considered as Perceptual Systems*, Greenwood Press, 1966.
10. James J. Gibson, *The Ecological Approach to Visual Perception*, Lawrence, Erlbaum Associates, 1986.
11. Georges Grinstein, "Workshop on Information Exploration Shootout Project and Benchmark Data Sets: Evaluating How Visualization does in Analyzing Real-World Data Analysis Problems," *Proceedings of the IEEE Visualization '97 Conference*, IEEE Computer Society Press, Phoenix, AZ, pp. 511-513, 1997.
12. Markus Gross, *Visual Computing, The Integration of Computer Graphics, Visual Perception and Imaging*, Springer-Verlag, 1994.
13. Taosong He and Stephen G. Eick, "Constructing Interactive Visual Network Interfaces," *Bell Labs Technical Journal*, Vol. 3, No. 2, pp. 47-57, April-June 1998.
14. William R. Hendee and Peter N.T. Wells, *The Perception of Visual information*, Springer-Verlag, 1994.
15. Eleftherios E. Koutsofios, Stephen C. North, Russel Truscott, and Daniel A. Keim, "Visualizing Large-Scale Telecommunication Networks and Services," *Proceedings of the IEEE Visualization '97 Conference*, IEEE Computer Society Press, San Francisco, CA, pp. 457-461, 1999.
16. David Marr, *Vision*, W.H. Freeman and Co., 1982.
17. Polla, D., J. McConnell, T. Johnson, J. Marconi, D. Tobin, and D. Frincke, "A FrameWork for Cooperative Intrusion Detection," *21st National Information Systems Security Conference*, pp. 361-373, October 1998.
18. Jonathan C. Roberts, "Multiple-View and Multiform Visualization," In *Proceedings of Visual Data Exploration and Analysis VII*, volume 3960, pages 176-185. IS&T and SPIE, January 2000.
18. Bernice Rogowitz, Personal Communication at IEEEVis 2000.
19. Robert Spence, *Information Visualization*, Addison-Wesley, 2001.
20. Edward Tufte, *Envisioning Information*, Graphics Press, 1990.
21. Research topics list, Information Visualization '2000, <http://www.infovis.org/infovis2000/restopics.html>.