Home-Centric Visualization of Network Traffic for Security Administration

Robert Ball, Glenn A. Fink, Anand Rathi, Sumit Shah, and Chris North
Department of Computer Science
Virginia Polytechnic Institute and State University
Blacksburg, Virginia 24061
{rgb6,arathi,sshah,finkga,north}@vt.edu
http://infovis.cs.vt.edu/

Abstract

Today's system administrators, under the burden of rapidly increasing network activity, need the ability to rapidly understand what is happening on their networks. While text-based systems are able to assist with awareness by alerting the user to potential problems, they are not good at helping the user form an accurate mental model of the situation. Because text-based systems cannot provide a dynamic overview of the whole network, network administrators need visualization tools. We present VISUAL (Visual Information Security Utility for Administration Live), a network visualization tool that allows users to perceive patterns in communications between their home (or *internal*) networks and external hosts.

We use a new visualization technique that gives a quick overview of current and recent communication patterns in the monitored network to the users. While many tools can detect and show fan-out and fan-in, VISUAL shows network events graphically, in context. Visualization helps users comprehend the intensity of network events on a much more intuitive level than text-based tools are able to provide. VISUAL provides insight for networks with up to 2,500 home hosts and 10,000 external hosts, shows the relative activity of hosts, displays them in a constant relative position, and reveals the ports and protocols used.

Keywords: Visualization Networks, Information Visualization, Large Data Set Visualization, Security, Network Communication

1 Introduction

In today's connected world, hackers cause damage worth millions of dollars. From denial of service attacks to corporate espionage, hackers and their tools cause companies and individuals loss of productivity and theft of critical data. According to a recent survey conducted by the FBI [FBI 2003], 80 percent of 503 companies surveyed responded that they had suffered financial losses due to cyber crimes. The tab on these losses was a staggering 450 million dollars, with only 40 percent of the victims quantifying their losses.

A significant portion of the effort to secure a network involves using traditional solutions such as IDS (Intrusion Detection System) and firewalls. These solutions produce megabytes of textual log data and packet traces (see figure 1) per day that network administrators must sift through to detect potentially malicious traffic.

Network administrators try to recognize intruders attempting to break-in or users trying to gain unauthorized access. Network administrators also use log files and packet traces to analyze how, when, and where machines are being used and for what intent. These tasks must not dominate network administrators' schedules, since they have many other pressing duties. Any new tool to help administrators reduce the amount of effort spent monitoring the network must be easy to use since training dollars are often scarce.

Many different techniques are currently used for analyzing network traffic. Various visualizations show individual computers, or show traffic between a single host and external systems. However, there are not currently any techniques that allow the home network or the external network to be very large. We believe our approach is the most scalable of the available concrete network visualizations.

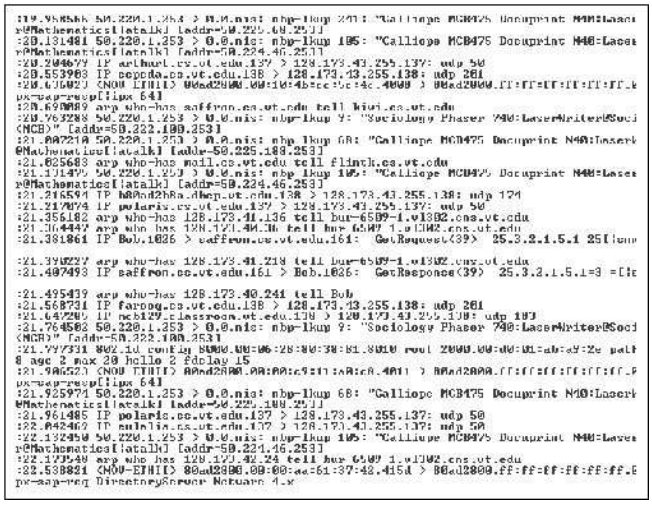

Figure 1: An example of a packet trace file to be visualized.

Network administrators are interested foremost in their own network: what is happening in it and how it is affected by the "unsafe" external Internet. By visualizing communications to reveal the subjects, objects, and duration of conversations, network administrators will be able to identify patterns that may be difficult to detect by conventional methods. One of the hardest parts of securing a network is constructing an accurate mental model of what is happening so that appropriate action can be taken.

As textual data is hard for people to sort through, this problem can be minimized by visualizing network traffic data. We hypothesize that by visualizing packet traces, network administrators can quickly and efficiently identify communication patterns in their networks. Currently, network administrators have to sift through large packet traces and log files to gain insight into their networks. If they find something suspicious, they may set up network monitoring programs to get more information to act. This procedure can be time-consuming and inefficient. In a medium-sized (Class B) network, log files and packet traces may easily approach gigabytes of information each day.

By visually observing the communication patterns in network traffic, network administrators can clearly identify the abnormal events

in their networks. Visual Information Security Utility for Administration Live (VISUAL) aids network administrators by showing a home-centric overview of their network. Aside from seeing abnormal traffic, VISUAL allows network administrators to develop an accurate mental model of what is normal on their own network so that they can diagnose problems better.

In this paper we first describe related work in section 2. We then explain our system design in section 3. A usability study that we conducted is in section 4. Future work is explained in section 5 and conclusions in section 6.

2 Related Work

We classify network awareness tools according to purpose (security or other), type (visual, textual, etc.), form (abstract or concrete), data (direct from the monitored network or post-processed), and perspective (the level that the observations apply to). In this section we compare VISUAL with other related work. A summary of the comparison is shown in Table 1.

Teoh, *et al.*, [Teoh et al. 2003] have developed a visual IDS that allows users to interactively explore connection log data. They provide a variety of plots that enable users to view high-dimensional data and discover anomalous behavior. Their approach relies on abstract presentations of the data that may require significant expertise to interpret. According to our classification, Teoh's work is a general-purpose, visual approach, abstract presentation of direct data from an internetwork perspective.

Network Intrusion Visualization Application (NIVA) [Nyarko et al. 2002] is an intrusion detection data visualizer with haptic integration that allows the user to interactively detect and analyze structured attacks over time using three dimensional space. NIVA's novel haptic interface allows users to "feel" virtual objects to analyze intrusion detection data. According to our classification, NIVA is a security-purpose, visual and other approach, concrete type presentation of post processed data from an individual network perspective.

Erbacher and Frincke [Erbacher 2003] discuss a visualization technique they have developed for the Hummer IDS [Frincke et al. 1998] [Frincke 2000]. The system generates an event list for visualization by processing an alert database from Hummer. Erbacher and Frinke arrange host dots in five concentric circles, where the center is the home host, and each enclosing circle matches one less octet than the previous one. They visualize the network data by drawing connection lines between host dots for traffic where line color represents the time of day. The authors claim the system is capable of visualizing traffic in real-time. By our classification, Erbacher and Frincke's work is a security-purpose, visual, concrete presentation of post processed data from an individual host perspective. VISUAL differs from Erbacher, et al. in that we have a *home-centric* perspective of networks instead of a single or few computers. We are able to visually show thousands of home hosts instead of just one or a few.

Luc Girardin [Girardin 1999] uses self-organizing maps to show an overview of network activity. He has implemented a neural network to reduce the dimensionality of the space of network and logging information down to two dimensional topological maps that illustrate the state of a network. Girardin's work uses foreground and background colors, sizes, and relative positions on the map display to display both network state and the deviance of current state (i.e., quantized error) from the normal. The method does not require any foreknowledge of the network whose state is to be displayed, but the resulting self-organizing maps are somewhat difficult to read. We believe that though the maps were an attempt to make the abstract network data more concrete by making it visible the result is nearly as abstract as traditional charts and graphs. The approach shows great promise though. We would classify this method as a general-purpose, visual, abstract presentation of direct data from an individual network perspective.

EtherApe [Cota 2001] is an open-source tool that features link layer, IP and TCP modes. EtherApe displays network activity graphically, animating host and link sizes according to traffic levels. It uses color and size (thickness) to encode information about protocol and traffic intensity. EtherApe uses Berkeley Packet Filtering (BPF) to narrow the scope of what is displayed. Our main criticism of EtherApe is its inability to scale beyond about 30 hosts/connections simultaneously. At this point and beyond, the otherwise neat display becomes very garbled with text labels, etc. Our classification of EtherApe is a general-purpose, visual, concrete presentation of direct data from an individual host perspective.

In addition to the above papers, the following two papers have contributed to the network visualization community in their unique techniques. We recommend the following two papers as helpful guides in visualizing networks and network traffic. Cheswick [Cheswick et al. 2000] uses a simulated spring-force algorithm to present graphs from their database. With the amount of data that they present, they show a colorful tree that represents the Internet. They explain how for intranets they were able to use their tool to find potential problems in routing. Herman [Herman et al. 2000] portrays a summary of various techniques for visualizing graphs. Hermann's paper surveys the many different graph presentation techniques that exist in a concise manner. Their topics range from visualizing trees, general graphs, spanning trees, to layout issues and solutions.

VISUAL's purpose is to provide more concrete visualizations that will require much less training to interpret. In addition, VISUAL is also more scalable in that it can show up to approximately 10,000 external hosts and 2,500 internal hosts. VISUAL is a security-purpose, visual, concrete presentation of direct data from a home-centric perspective.

Name	Purpose	Type	Form	Data	Perspective
Teoh, *et al.*	general	visual	abstract	direct	internetwork
NIVA	security	visual and other	concrete	post-processed	single network
Erbacher *et al.*	security	visual	concrete	post-processed	single or few hosts
Girardin	management	visual	abstract	direct	single network
EtherApe	general	visual	concrete	direct	single host
VISUAL	security	visual	concrete	direct	home network

Table 1: Summary of VISUAL compared to related work.

3 System Design

As figure 1 shows, the type of data that we wish to visualize comes from network monitoring tools such as TCPDump or Ethereal that represent the individual packets sent between computers. The contents of the data include: Source IP, destination IP, source port, destination port, protocol, and time recorded among other things.

VISUAL's design follows Shneiderman's information visualization mantra [Shneiderman 1996]: "Overview first, zoom and filter, then details on demand." We wish to quickly show an overview of network traffic for a small to mid-size network. To do this, we show a home-centric ("us versus them") perspective of the network. In other words, VISUAL shows a representation of all the home (internal) IPs as a large box with all the home IPs represented as small squares (see figure 2). By showing the network traffic as home centric, fan-in and fan-out become apparent in a way that helps users develop an accurate mental model of network activity. A user can see which home IPs received a great deal of connections (fan-in) and which external IPs were hit the most from the internal IPs (fan-out). Although there are many ways to detect these phenomena, we believe to present them in a concrete visualization is the most effective for seeing them in context and to inform the user of total network state.

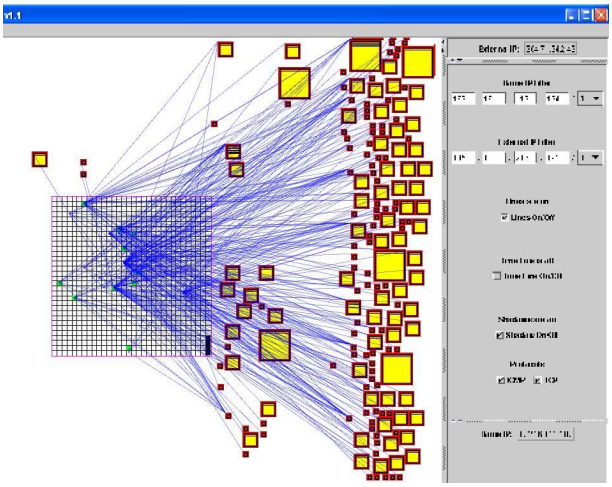

Figure 2: An example of VISUAL displaying 80 hours of network data from a home network of 1020.

This type of visualization also shows relative amounts of activity among external IPs in that each external IP is represented as a square whose size is proportional to how many packets were sent/received relative to how many packets were sent/received by other external IPs.

Our home-centric perspective is based on the assumption that network administrators are most concerned with the security of the hosts for which they are responsible. These hosts are predefined in a text file. After reading the list of home IP addresses, VISUAL loads and displays the network traffic. This view constitutes the "overview" of the network data (see figure 2).

As VISUAL is intended to be an overview of "us versus them," we determined the following list of display priorities that are intended to offer key insights to the administrators about the traffic going through their networks:

1. The hosts that communicate with hosts on the home network.
2. Amount of communication that takes place between an external host and the home network (that is, the number of packets exchanged).
3. Methodology of communication (for example, what protocols and ports were used during this communication.)
4. Representation of time during the capture of packets.

3.1 External IP Layout

Our goal is to visualize traffic from external hosts (4 billion possible Internet Protocol (IP) addresses in IP version 4) that communicate with the home network. We attempt to maintain a constant position for each external host. We map every external host to a square object in the display area. In other words, if a given host appears in two different data sets, its screen position would be approximately the same in both plots. Therefore, we implement a mapping function that gives every external IP address a unique position in the display area. We use techniques from Keim [Keim and Herrmann 1998] that help to display large amounts of spatially referenced data on a limited-size screen display.

We developed a mapping function that maps an IP address to the available screen space as monitors do not have 4 billion pixels to display every possible IP address in IPv4. In IPv4 each IP address consists of 32 bits. In dotted quad notation these bits are represented as A.A.A.A, where A ranges from 0 to 255. In our mapping function, the first two numbers of the IP address (high 16 bits) determine its X coordinate on the screen and the last two numbers (low 16 bits) determine the Y coordinate. For example, in figure 3 there are 183 external IPs shown as small, medium, and large yellow squares with a red border. Each of these squares is positioned in the display area based on its IP address.

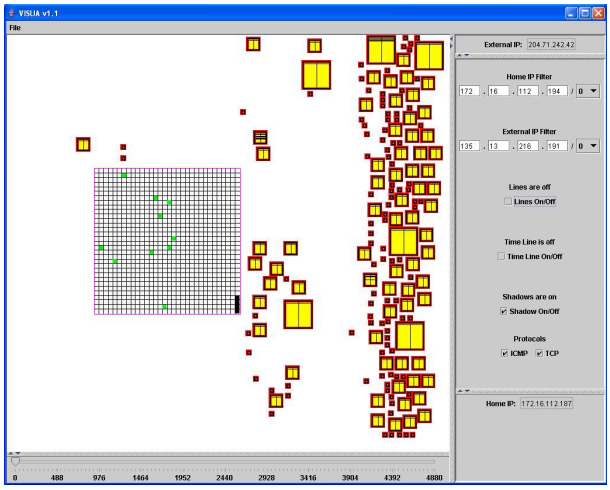

Figure 3: Mapping function example. Lines not displayed to show the mapping function better.

This simple mapping scheme however would suffer from a drawback that IP addresses that are similar (e.g., 192.168.0.1 and 192.168.0.2) map into points very close to each other. This generally happens when hosts on the same external network appear in the same plot. To overcome these problems of mapping to the same point and overlapping, we use adaptive techniques from [Robertson et al. 1998]. These techniques guarantee that no two markers will overlap each other if there is enough screen space available.

Simply put, if two external IPs map to the same space in the display area then the one plotted last is moved to the nearest available position.

The algorithm that guarantees no overlap follows the pattern of pushing down and then to the right (increasing the x coordinate and resetting the y coordinate to 0) when there is not any more space on the bottom of the display area. If an IP is pushed to the bottom right corner of the display area, then the algorithm tries again at the top left corner of the display area. If there is no extra space in the display area, the algorithm scales all IPs down (e.g., 2 pixels by 2 pixels instead of 3 pixels by 3 pixels) and repositions all the IPs again.

A feature of this mapping is that external IPs from the same Class B or C network are clustered in vertical lines. This fact may help network administrators to see additional patterns. The mapping scheme takes advantage of the fact that there is likely to be empty areas due to non-uniform communication to spread out the dense clusters. As a demonstration of scale, figure 2 shows approximately 1020 home hosts, 183 external hosts and 915 communications. In general, we find our mapping approach scales to be able to display approximately 10,000 external hosts.

At this time we are not giving any special treatment to nonroutable (private) IP addresses, broadcast, multicast, or any other special address classes. They will appear on the place their address maps to on the screen. Relative positioning greatly helps users because as the data changes from day to day, the user can find a given external IPs in the same relative position. Since computers do not frequently change IP addresses (at least within a tightly constrained range) our simple mapping approach will help administrators detect patterns from session to session.

3.2 Home Network Layout

The large square grid in the display area represents the home network. The home network is divided up into many smaller squares that each represents a computer on the home network. The home grid is automatically positioned in an empty area of the display space.

There is the possibility, given enough data, that overlap of an external IP and the home network will happen. If an external hosts, would be plotted over the home network, it is moved to a nearby location. The home network is also movable to avoid overlap. VISUAL cannot guarantee non-overlap if there is sufficient data to fill up the display area with external hosts. Thus we have made the home network resizable.

3.3 Communication

The next insight that we desire is to support who is communicating with whom. In other words, who in our home network of computers is communicating with what external computers? In order to facilitate this, a representation of the home network is shown with lines from individual computers in the home network communicating to other external hosts.

Each line represents a communication of one or more packets sent between communication end points. A line between computers represents only the fact that there was some communication between the computers, not how much. To see how much traffic was communicated between the two computers, see section 3.4. The reason for having only one line for each connection is to minimize occlusion.

Line color shows the direction of traffic. For example, figure 4 shows an external host sending packets to many different internal computers without any reply. Communication sent from an external host to an internal host without response, is presented in red. Bidirectional communication is shown in blue. Communication that left the home network without receiving a reply, is appears in green.

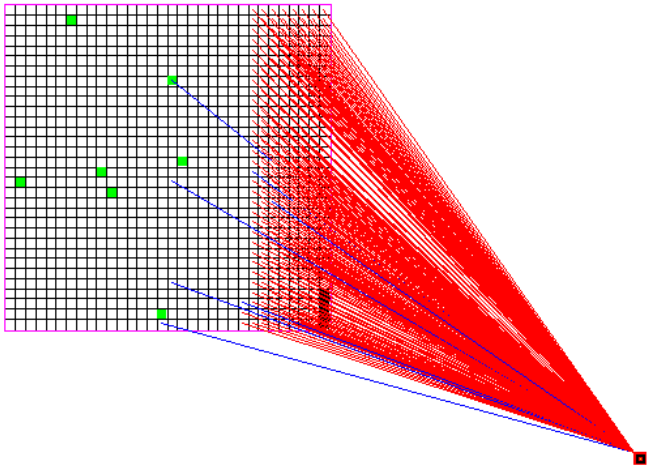

Figure 4: Example of fan-in. An external computer performing a ping sweep on a subnet in the home network.

In the external-external view, we summarize the internal traffic (home computers communicating with other home computers) by shading home computers green if they communicated with other home computers (see figure 4 or 5). For instance, if most computers in the home network showed up as green, then most computers communicated internally. On the other hand, if few squares are green, as in figures 4 and 5, then there were few computers communicating internally. As is the case with the lines, a green square does not represent the amount of activity, but simply that there was activity. Different shades of green indicate the proportional degree of traffic similar to how the size of external hosts indicate proportional degree of external traffic.

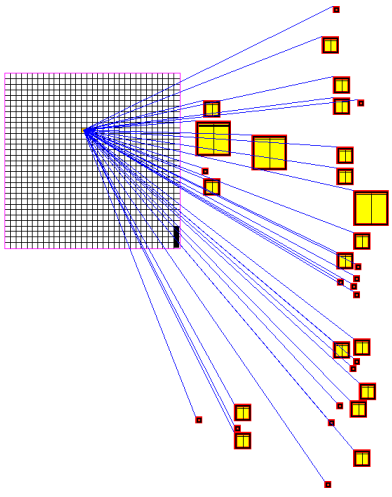

Figure 5: Example of fan-out. Many different external computers are downloading information from a server on the home network.

VISUAL emphasizes internal to external and vice-versa traffic because this was found to be the most important in requirements anal-

ysis with network administrators. Administrators we interviewed indicated that they generally monitor the internal-external traffic until they notice a potential security hazard. Then they turn their attention inward to locate and eradicate problems within their own bailiwick. To view internal traffic, VISUAL provides an internal-internal mode where all internal hosts are plotted twice: once within the home grid, and a second time outside it as if they were external hosts. In this view, external hosts are not plotted. In this way, we can quickly identify which internal host is responsible for which traffic.

As mentioned earlier, fan-in and fan-out are evident with our visualization. Figure 4 shows an external host performing a ping sweep (a method of checking what computers are there). As can be clearly seen from the figure, the external host systematically goes through every IP address in a particular subnet in our home network. This is easily recognizable by most users (see the usability study, section 4 for details), without regard to experience in networking.

As an example of fanout, consider figure 5. In figure 5, several external IPs are contacting a particular computer in our home network. Although this could be a small denial-of-service attack, it happens to be a public FTP server that external hosts are accessing.

3.4 Amount of Activity

After seeing who is communicating with whom, the insight we desire is to know how many packets were communicated between computers during the time period that the data set is recorded. Thus we size the marker of every external host that communicated with the home network in proportion to how many packets it received/sent during the time frame compared to the other computers that communicated with the home network.

Figure 5 shows different sized squares. The largest squares contributed about 5 percent of the of the packets for this data set while the smallest squares contributed to less than 0.001 percent. The results of these combined communications are aggregated. For example, figure 5 shows different sizes of external hosts. There are 3 large external hosts that represent the external computers that communicated with the home network the most, several medium-sized squares that contributed less traffic proportionally, and small-sized squares that contributed very little in comparison.

3.5 Port Visualization

The next insight is visualizing what ports an external host that communicated with an internal host used. To show each port number on the screen space as an icon or as text for each square would clutter the screen space and cause confusion. Instead the destination and source ports are shown visually as horizontal lines within the external host's marker. We draw a horizontal line through the external host's marker to represent each port used. There are 65,535 possible ports that a computer can currently use. The position of the horizontal line is normalized to the size of the marker. Low numbered ports are appear towards the top of the marker. By showing visually the relative position of the ports, one can know approximately how many ports an external IP used and where in the port space it was during communication with the home network.

We show horizontal lines for both the destination and source ports of the used by the external hosts. The home network computers do not display the port visualization to prevent a confusing display.

Figure 6 shows an external host that is communicating on a range of ports with many different home computers. The external host has many different lines that represent ports at the beginning of the port space, the first third and the first half.

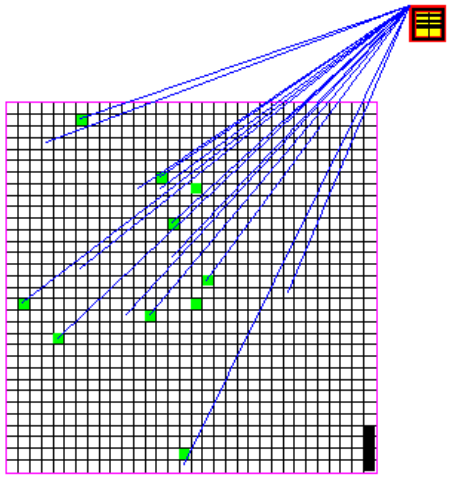

Figure 6: Display of one external IP using multiple ports.

3.6 Interactive Filtering

After seeing the overview, the user may want further insight into fewer computers. A user may select a single computer (see figure 7), or multiple computers (see figure 8), to see only the traffic that occurred with those particular computers and filter out all other data. For example, if the home network consists of 1020 computers and thousands of lines and outside traffic is visualized, then it may be desirable to only see a subset of the entire data set. By selecting particular computers - either an outside computer, or a computer in the home network - only those computers are shown.

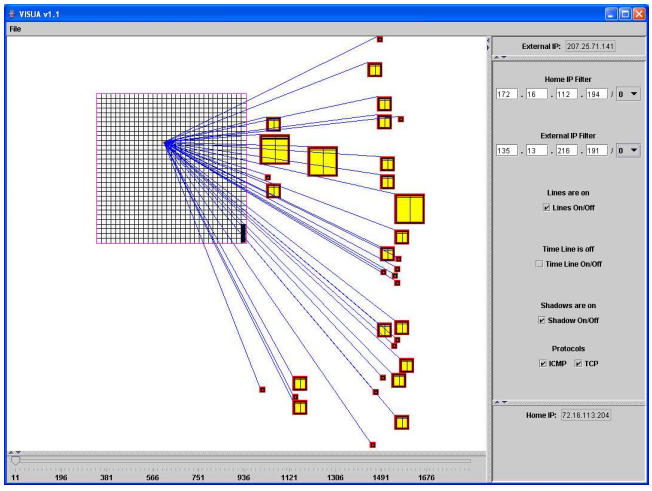

Figure 7: An example of a single selection. Only one home computer is selected, which is shown as an orange square.

A user may select interactively by either clicking on a single computer or selecting a range of computers. The user may select more computers CTRL clicking. A selected computer that is clicked again while holding down CTRL is deselected. Also, a range of computers may be selected by holding down shift and moving a dynamic box around the desired area, like in Windows®Desktop.

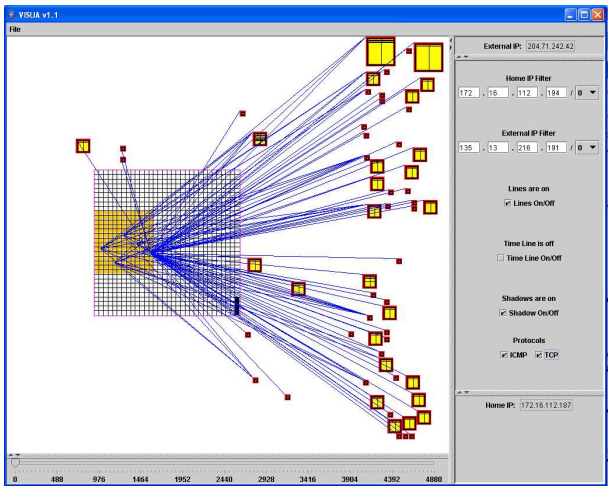

Figure 8: An example of multiple selection. All the traffic for the selected home computers is displayed. The selected home computers are represented as orange squares.

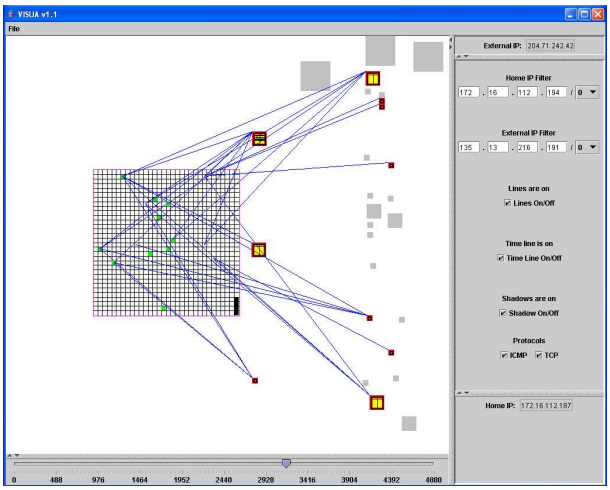

Figure 9: Time line example using shadows. The light-gray squares represent inactive external hosts that have been active in the last 200 seconds. The yellow squares represent active external hosts during the time segment.

The user may also select particular computers to scrutinize by using a Classless Internet Domain Routing (CIDR) bit mask filter. For example, if the user wanted to see only external IP addresses whose first octet was 125, then the user would set the external IP filter to 125.x.x.x /8.

3.7 Time Line

Although the size of the squares of the outside computers shows how many packets were received and sent relative to other outside computers, the size does not show *when* the packets were received relative to other packets. For example, a large external marker (an external host that sent/received many packets from the home network) may have communicated only at the beginning of the data set or during the entire time.

The user may activate a time-slider to show the sequence of packets from beginning to end. Only the outside computers and packets that were active at an instance of time are shown with the time-slider. All other lines and outside computers that were not active during that particular instance are hidden.

3.8 Shadows

The timeline feature helps the user quickly track the flow of events as they happened chronologically. We introduce a "shadow" that follows [Nowell et al. 2001]'s idea of "change blindness." That is, as the time-slider is moved across time the external IPs that communicate with the home network appear and disappear. According to [Tversky 1992] we expect it would be difficult for the human mind to keep track of what external host markers were actually on and which were off. VISUAL helps the user remember which external hosts just communicated with the home network by representing a light-gray square, or a shadow, where the external host was for a short period of time.

If an external host is communicating intermittently with the home network it would disappear and reappear. We have added "shadows" to compensate for pauses in data and also to help the user see what has been recently actively but not currently active. With shadows enabled an intermittently communicating external host would flash normal and then light-gray rather than disappearing completely (see figure 9). The default time window for when shadows are shown is 200 seconds, but can be changed according to preference.

3.9 Filters

Other available filters are check boxes located in the control panel on the right side of the application that toggle features on and off. There are check boxes for toggling the following features on or off:

- Lines
- Time line
- Shadows in the time line
- Different protocols (e.g., TCP, UDP, ICMP, etc.)

The user can also control the physical size of the home network grid. Independently of this feature, there is also a feature that grows or shrinks the markers of the external hosts. Host markers of either kind can be reduced to a single pixel or allowed to overlap one another to avoid moving markers from their original spaces by to the overlap algorithm. We intend to implement a fish-eye view of the display area to help view more detail on individual hosts.

3.10 Details on Demand

At any time, the user can also get details of a computer by selecting it. The details include:

- The host's IP address
- The IP addresses of all the computers that the selected computer communicates with
- The TCP/UDP ports (both source and destination), the host uses to communicate

- The percentage of the overall traffic this particular computer contributes to the overall data set within the analysis time period.

4 Usability Study

We conducted a usability study to determine the effectiveness of VISUAL to help users get insight into traffic patterns on a network. Our main objective was to verify if users could determine if abnormality exists in the data set.

Since network administrators are familiar with their own networks and have a mental model of what the communication patterns *should* look like, they can use VISUAL to see whether communication patterns look normal or not. However, test users, having only a single look at a test network cannot draw such conclusions. For our usability study, the *normal* versus *abnormal* behavior was achieved by comparing the two or three data sets the users looked at. As we only had a limited amount of time with each user, we could not show the users more than three data sets. Part of their task was to identify behavior that is markedly different between the datasets.

We conducted our usability study with nine graduate and undergraduate university students. During each session we had each user answer several biographical questions aimed at determining their experience with computers and network security. The users described themselves on a range from "power users" to "occasional users" of computers. We did not have any network administrator as part of our usability study because we wanted to test untrained users. When asked about the interface, none of our users mentioned any discomfort with the interface of VISUAL nor complained about any usability issues.

We explained to each user how VISUAL works then allowed them to become comfortable with VISUAL for about five minutes with a training data set that did not have any abnormalities in it. Once they were familiar with VISUAL, we presented them two testing data sets and asked them to:

1. Describe anything striking about this data set.
2. List the IP addresses of four external hosts that appear to be involved in normal (repeated, frequent in time) communication with the home network.
3. List the IP addresses of four external hosts that only communicate from time to time with the home network.
4. List the IP addresses of four home network hosts that make the largest number of connections to external hosts.
5. List the IP addresses of four external hosts that connect to the largest number of different home network hosts.

All data sets came from the same network [Lippmann et al. 2000]. The second data set that the users were shown contained large amounts of data compared to the first data set, but was still normal for the network. The third data set had normal network traffic except it contained a ping sweep and was slightly smaller than the second data set.

The users wrote out their responses for what they deemed abnormal in their own words. Every user was able to quickly find the same set of abnormalities in the data, Even though each user used different words in their answers their answers agreed. In the second data set the only abnormality was that there were three servers that received most of the traffic. Although this is actually a normal trait for the network that we used, the three servers stood out as different from the first data set. In the third data set every user was able to quickly focus on a ping sweep as abnormal. Although not all of the users knew what a ping sweep was, they were still able to see that it was a different traffic pattern than the rest of the data presented.

Users were also able to quickly identify which external hosts communicated most with the home network (based on amount of traffic sent/received), which external hosts communicated most often (based on time spread), which internal computers communicated the most with one another, and which external hosts communicated to the most internal hosts (based on number of connections). However, users had difficulty using the timeline to identify external computers that we described as "only communicating from time to time" with the home network in the third question. Part of the problem was with how we framed the question. We did not define "from time to time" quantitatively, and some users were confused. Also, unless there was something striking about the intermittent communication, users had difficulty characterizing the communication and identifying hosts involved in it.

Another problem users experienced was that communication lines occluded some host markers in the overview window. We fixed the occlusion problem by making the lines translucent so that the host markers showed through.

During our usability study, we had not enabled the feature that allowed viewing of traffic strictly between internal hosts. Users and subsequent interviewees have expressed their desire to see this internal traffic as well. Subject matter experts evaluated VISUAL and determined that their investigation of potential intrusions would require an internal-only view of traffic superior to what VISUAL offered in the study.

One user said VISUAL made it, "easy to make sense of data and see general trends." This user said VISUAL would be less usable for "fine-grained" view of network traffic data. This observation fits our intentions well for VISUAL at this stage and underscores the importance of providing drill down to the packet level for analysis purposes in the future.

We feel that our usability test successfully demonstrated that a traffic pattern visualizer such as VISUAL can provide insight into network traffic data without requiring any training. The participants also provided many useful comments that we will use to improve VISUAL and work based on it in the future.

5 Future Work

VISUAL is an proof-of-concept program that has allowed us to test the premise that visualizations of network traffic data help users rapidly form accurate mental models of network events with little or no training. We plan to continue improving VISUAL based on the feedback received from more usability studies.

Currently, VISUAL is useful for small networks of fewer than about 12,500 nodes (of which 2,500 are internal). We believe the concept could be scaled to larger networks of perhaps a few hundred thousand hosts. One technique that may be useful is shrinking the host markers to single pixels and implementing a distorted lens (fisheye) to view and select individual computer markers. Our forthcoming project, Network Eye, is designed to explore larger scale visualization techniques.

VISUAL currently relies on a preprocessor to digest network packet traces. We accepted this limitation early in development to save

time, but it has always been our intention to allow VISUAL to accept tcpdump [tcp 2004] format packet traces and the like. VISUAL would be much more useful to system administrators as a real-time network traffic visualizer.

We believe that the view of purely internal traffic VISUAL provides is a good approach since it uses the same presentation that our usability study confirmed as good for internal–external traffic. However, this mode and the interactions that allow switching between modes must be tested. We envision doing an expert review and another usability test with this feature.

A user-suggested idea for the timeline is to change the alpha channel for external host markers so that they are fully opaque when active and slowly fade away (increasing transparency) until they disappear over time [North et al. 2001]. It would also be helpful to place a histogram over the timeline's scale to show the distribution of the relative amount of activity at each moment. The implementation would be similar to [Li and North 2003] and would enable users to more rapidly locate areas of interest when viewing in timeline mode.

6 Conclusions

In this paper we discussed our contributions to network security visualization which include the following:

- A new visualization of network traffic that allows rapid perception.
- Usability information that shows how the tool helps users make accurate mental models of network state.
- Network visualization scalability to over 10,000 nodes.
- Temporal visualization of traffic.

Network data analysis is a very important task from an administrator's point of view. A significant amount of time and manpower is devoted to sift through text-only log files and packet traces generated by tools used to secure networks. VISUAL has demonstrated that visualization considerably reduces the time required for data analysis of network traffic and at the same time provides insights which might otherwise be missed during textual analysis.

References

CHESWICK, B., BURCH, H., AND BRANIGAN, S. 2000. Mapping and visualizing the internet. In *Proceedings of the 2000 USENIX Annual Technical Conference*, USENIX Assoc.

COTA, J. T., 2001. Implementacion de un monitor y analizador grafico de red en el entorno gnome, July.

ERBACHER, R. F., AND FRINCKE, D. 2000. Visualization in detection of intrusions and misuse in large scale networks. In *Proceedings of the International Conference on Information Visualisation*, IEEE Computer Society, International Conference on Information Visualization (IV2000), IEEE, 294–302.

ERBACHER, R. F. 2003. Intrusion behavior detection through visualization. In *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*, IEEE Computer Society, IEEE, 2507–2513.

FBI, 2003. Cybercrimes on the rise.

FRINCKE, D. A., TOBIN, D., MCCONNELL, J. C., MARCONI, J., AND POLLA, D. 1998. A framework for cooperative intrusion detection. In *Proc. 21st NIST-NCSC National Information Systems Security Conference*, NIST, 361–373.

FRINCKE, D. 2000. Balacing cooperation and risk in intrusion detection. *ACM Transactions on Information and System Security (TISSEC) 3*, 1 (February), 1–29.

GIRARDIN, L. 1999. An eye on network intruder-administrator shootouts. In *Proceedings of the 1st USENIX Workshop on Intrusion Detection and Network Monitoring*, USENIX Assoc.

HERMAN, I., MELANCON, G., AND MARSHALL, M. S. 2000. Graph visualization and navigation in information visualization: a survey. *IEEE Transactions on Visualization and Computer Graphics 6*, 1, 24–43.

KEIM, D. A., AND HERRMANN, A. 1998. The gridfit algorithm: An efficient and effective approach to visualizing large amounts of spatial data. In *Proceedings of the Conference on Visualization '98*, IEEE Visualization, 181–188.

LI, Q., AND NORTH, C. 2003. Empirical comparison of dynamic query sliders and brushing histograms. In *Proceedings of IEEE Symposium on Information Visualization*, IEEE Computer Society, 147–153.

LIPPMANN, R., HAINES, J., FRIED, D., KORBA, J., AND DAS, K. 2000. The 1999 darpa off-line intrusion detection evaluation. *Computer Networks*, 34, 579–595.

MIT. Mit lincoln laboratory network intrusion datasets. http://www.ll.mit.edu/ist/ideval/data/data_index.html.

NORTH, C., FAROOQ, U., AND AKHTER, D. 2001. Datawear: Revealing trends of dynamic data in visualizations. In *LBHT Proc. IEEE Symposium on InfoVis 2001*, IEEE computer Society, IEEE, 8–11.

NOWELL, L., HETZLER, E., AND TANASSE, T. 2001. Change blindness in information visualization: A case study. In *Proceedings of 2001 Information Visualization*, IEEE Computer Society.

NYARKO, K., CAPERS, T., SCOTT, C., AND LADEJI-OSIAS, K. 2002. Network intrusion visualization with niva, an instrusion detection visual analyzer with haptic integration. In *Proceedings of the 15th annual ACM symposium on User interface software and technology*, Palgrave Macmillan, IEEE, 277–284.

ROBERTSON, G., CZERWINSKI, M., LARSON, K., ROBBINS, D. C., THIEL, D., AND VAN DANTZICH, M. 1998. Data mountain: Using spatial memory for document mangament. In *Proceedings of the 11th annual ACM symposium on User interface software and technology*, ACM Press, ACM, 153–162.

SHNEIDERMAN, B. 1996. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the IEEE Symposium on Visual Languages '96*, IEEE Computer Society, IEEE, 336–343.

2004. Tcpdump public repository, June.

TEOH, S. T., MA, K.-L., AND WU, S. F. 2003. A visual exploration process for the analysis of internet routing data. In *Proceedings of the IEEE Conference on Visualization 2003*, IEEE Computer Society.

TVERSKY, B. 1992. Distortions in cognitive maps. *Geoforum 23*, 2, 131–138.