1-12-2006

A Survey, Taxonomy, and Analysis of Network Security Visualization Techniques

Rawiroj Robert Kasemsri

A Survey, Taxonomy, and Analysis of Network Security Visualization Techniques

by

Rawiroj Robert Kasemsri

Under the Direction of Ying Zhu

ABSTRACT

Network security visualization is a relatively new field and is quickly gaining momentum. Network security visualization allows the display and projection of the network or system data, in hope to efficiently monitor and protect the system from any intrusions or possible attacks. Intrusions and attacks are constantly continuing to increase in number, size, and complexity. Textually reading through log files or other textual sources is currently insufficient to secure a network or system. Using graphical visualization, security information is presented visually, and not only by text. Without network security visualization, reading through log files or other textual sources is an endless and aggravating task for network security analysts. Visualization provides a method of displaying large volume of information in a relatively small space. It also makes patterns easier to detect, recognize, and analyze. This can help security experts to detect problems that may otherwise be missed in reading text based log files.

Network security visualization has become an active research field in the past six years and a large number of visualization techniques have been proposed. A comprehensive analysis of the existing techniques is needed to help network security designers make informed decisions about the appropriate visualization techniques under various circumstances. Moreover, a taxonomy of the existing visualization techniques is needed to classify the existing network security

visualization techniques and present a high level overview of the field.

In this thesis, the author surveyed the field of network security visualization. Specifically, the author analyzed the network security visualization techniques from the perspective of data model, visual primitives, security analysis tasks, user interaction, and other design issues. Various statistics were generated from the literatures. Based on this analysis, the author has attempted to generate useful guidelines and principles for designing effective network security visualization techniques. The author also proposed a taxonomy for the security visualization techniques. To the author's knowledge, this is the first attempt to generate a taxonomy for network security visualization. Finally, the author evaluated the existing network security visualization techniques and discussed their characteristics and limitations. For future research, the author also discussed some open research problems in this field. This research is a step towards a thorough analysis of the problem space and the solution space in network security visualization.

INDEX WORDS: Network security, Security visualization, Taxonomy, Anomalies, Security information.

A SURVEY, TAXONOMY, AND ANALYSIS OF NETWORK SECURITY VISUALIZATION TECHNIQUES

by

Rawiroj Robert Kasemsri

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science

in the College of Arts and Sciences

Georgia State University

2005

A SURVEY, TAXONOMY, AND ANALYSIS OF NETWORK SECURITY VISUALIZATION TECHNIQUES

by

RAWIROJ ROBERT KASEMSRI

Major Professor:	Ying Zhu
Committee:	G. Scott Owen
	Saeid Belkasim
	Raheem Beyah

Electronic Version Approved:

Office of Graduate Studies
College of Arts and Sciences
Georgia State University
December 2005

ACKNOWLEDGEMENTS

I would like to thank my parents and friends for all their supports throughout the course of my study and the time it took to complete this thesis. I would like to highly thank Dr. Ying Zhu, my thesis advisor, for his support, guidance, motivation, and most of all, patience, all the qualities upon which this thesis can not have been on tract or completed without.

I would also like to thank my committee members, Dr. G. Scott Owen, Dr. Saeid Belkasim, and Dr. Raheem Beyah for their inputs and continuing suggestions. Lastly, I would like to thank Dr. Raj Sunderraman for his support and guidance throughout the course of the program.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF CHARTS

1. INTRODUCTION

1.1 THE NETWORK SECURITY PROBLEMS

The network security problems pertain to difficulties securing the systems. Problems arise when unauthorized users try to access the system illegally. Because data is vulnerable and contain confidential information most of the time, having a security system for the network is essential. Also, due to the increase in the number of attacks and intrusions, different methods and techniques of security must be applied. This, in return, also causes more loops and open areas for attacks. Allowing data and information to be accessed illegitimately is absolutely not acceptable in network security.

Network security, then, refers to the process of securing a network or system against these illegitimate accesses or intrusions. An intrusion occurs when a legitimate user or unauthorized user gains an illegal access. An attack occurs when damage has been done to the network or system. Anomalous activities are defined as accesses that are suspicious, but not necessarily an intrusion or an attack. Benign accesses are accesses that do not harm the network or the system while malicious accesses are simply harmful attacks [12].

Network security is important because data is vulnerable and confidential. Data is the main component in the network and system. Therefore, having data exposed to unauthorized users is strictly prohibited. Network security protects data from illegal accesses and unauthorized manipulations and modifications.

Some of the most common network attacks are: denial of service attacks, dictionary attacks, man in the middle attacks, IP spoofing, and password attacks. Furthermore, there are also attacks that target areas in the network or system. ARP flooding causes a denial of service

by flooding switches. ARP redirect redirects network traffic through a different computer. IP spoofing connects in spite of access restrictions, and broadcast IP spoofing causes a denial of service through forced replies. TCP hijacking targets by applying a man in the middle attack. TCP reset causes a denial of service attack. TCP client port spoofing attacks through data ports. There are many other attacks similar to the ones described.

Due to the increase in the number of network attacks, different solutions to network security problems must evolve periodically. Networks and systems are becoming increasingly more complex [27]. However, there is no absolute way to secure a network or system completely or indefinitely. There is bound to be an intrusion at some point in time. Currently, different solutions to network security problems are used. Patches is one of the solutions being used today in network security, which refers to the modification of codes due to bugs or misfeatures. Vulnerability scanning refers to the scanning of the system for weakness. Firewall, a dedicated gateway machine with special security precautions on it, is used to protect machines. Intrusion detection systems, or IDS, are tools that are used to secure a network or system by identifying intrusions or illegal accesses within the network or system. Whatever the technique or tool of securing a network or system may be, everything still rely heavily on human detection. Most of these techniques and tools must have the user analyze and detect the anomalies or intrusions. Moreover, a huge number of them require the user to sound the alarm. Most of the solutions proposed generally show intrusions or attacks, but humans are still required to respond.

1.2 THE TASKS OF A NETWORK SECURITY ANALYST

The tasks of a network security analyst are crucial and very detailed. A mistake can lead to security breaches or data leak from the network or system. Most of the time, securing something is difficult because there is no telling when an attack can occur.

Network security analyst must possess highly monitoring capabilities. They have to be able to monitor network traffic over periods of time, and must also be able to detect any anomalies and to take action, sounding the alarm, when fitting. They must be able to run diagnostic software and testing as well. Updates are another essential, since new attacks are constantly evolving. Because of this, network security analyst must be periodically informed about new attacks, and be aware of new tools and solutions to these attacks. Moreover, they must be able to distinguish between regular accesses, intrusions, and attacks. They must be able to identify which processes or accesses are benign, and which are malicious. Being able to identify false alarms from normal traffic or from intrusions is another valued asset. Having the ability to differentiate false positives and false negatives is another important task for network security analysts. It is very difficult to judge the quality of the output of the system, whether it is benign, a false alarm, or an intrusion [3]. False positives occur when normal traffic is identified as intrusions, or more commonly false alarms. False negatives occur when the system fails to identify an intrusion or attack.

In most of the security visualization techniques proposed in the research papers, a huge number of them require complete and thorough understanding of the tool and technique used. In other words, security visualization techniques are not designed for public use. Users must be trained to master these techniques. They must know all the different aspects and components of the security visualization technique they work with. To be a network security analyst for these security visualization techniques require deep knowledge and understanding of the techniques, as

well as the network they monitor and work with. It comes to no surprise that most of the current network security visualization techniques require thorough training for security analysts prior to using the techniques [5].

1.3 NETWORK SECURITY VISUALIZATION

Network security is defined as the measures taken to protect a network or system, or communications pathway, from unauthorized access to, and accidental or willful interference of, regular operations. Network security visualization, then, pertains to the display and representation of data within a network or system, as a means to protect or secure from unauthorized and illegal accesses. It is another approach of displaying data non-textually, and a more effectively way of securing a system or network through monitoring and detecting anomalous activities [18]. Network security visualization is broken down into the areas of monitoring the network or system, detecting anomalies, finding intrusions, assessing attacks, sounding alarms, and analyzing patterns or other information generated.

Network security visualization is part of the larger field of information visualization. Information visualization refers to the presentation of data to users and analysts to provide insights. Information visualization refers to the use of interactive, sensory representations, typically visual, of abstract data to reinforce cognition. Information visualization was built on theories of information design, computer graphics, human-computer interactions, and cognitive science. The advantage of information visualization is that it allows the transformation and representation of abstract data into a form humans can understand and analyze. Important aspects of information visualization are the interactivity and dynamics of the visual representation.

Visualization of network security is essential to all network systems. Being able to visualize and monitor the system helps eliminate the element of surprise in intrusions and the attacks. Visualization in network security allows patterns to be studied and analyzed at a less difficult level. With proper security visualization techniques, large volumes of information, such as raw data, can be displayed in a relatively small space. Analyzing large amounts of data in log files is a tremendous task for network security analysts. Thus, having a method of displaying those large volumes of data in a smaller space, and not using only textural information, helps make securing the network or system more efficient. Having network security analysts spending less time browsing through log files and more time detecting and analyzing data accounts for easier and better use of time, technology, and resources. It is a fact that false positives are a big issue in network security. Using visualization, this element can be minimized, though not completely eliminated [23]. With visualization, millions of security log entries can be correlated by day, into a cohesive, understandable display. Moreover, visualization of network security can help uncover misconfigurations, which are often undiscovered by the human brain. In addition, visualization can also expose security policy violations that have taken place, and signify normal and anomalous activities. In other words, many different areas of network security are important and complicated uniquely. Having a network security analyst handling every little detail effectively is impossible. Visualization provides a way to divide and handle different components, leaving some for the network security analyst to assess. By doing this, the network security analyst has fewer areas to assess, and more time to efficiently handle the more important areas, such as sounding the alarm or differentiating between benign and malicious accesses [7].

1.4 MOTIVATION AND PURPOSE

The research consisted of different areas of analyzing and the interpretations of many security visualization techniques. The main task was to survey the existing network security visualization techniques. This provides a high level overview of this field. Since network security visualization is a relatively new field, there has been very little or no survey at all in this area.

This research serves different purposes. One is to provide understanding and information regarding security visualization techniques. However, the main purpose is for security visualization designers. This research is a step towards creating a "cookbook" for network security visualization. Security visualization designers can conduct researches about what visualization techniques are available, what types of visualization methods used, the kinds of graphical visualization used, the types of data mainly used, the types of security scanning applied, the classifications of the security visualization techniques, and more. These information can be very useful before a security visualization designer wishes to design a security visualization tool.

Moreover, security visualization designers can be informed about the pros and cons regarding different issues. For example, a security visualization designer can ask questions like, what is used more in security visualization techniques, glyphs, color maps, or histograms? What types of data, or security information, are mainly used? What is the appropriate way to visualize certain security information? How to deal with various design issues, such as multi-variant data? Are there many security visualization techniques that implement user interfaces, such as detail on demand and zooming capabilities? To create such a "cookbook," it is necessary to survey the existing network security visualization techniques.

For researchers, this work provides the first taxonomy of network security visualization techniques. Taxonomy would provide researchers with a high-level overview of the field and a classification of the existing techniques. This would help them put their research in the context. There have been several taxonomies for information security. However, these taxonomies are general ones, trying to cover the entire field of information visualization with a single proposed taxonomy. The field is relatively large and complex. Thus, this taxonomy for the entire field is not sufficient. The taxonomy in network security visualization is an extension of the general taxonomies.

This research was conducted by analyzing and evaluating each of the network security visualization techniques. Each of the existing network security visualization techniques was evaluated. Strengths, weaknesses, advantages, and limitations were pointed out and studied in details. Identifying which stage of the network security problem solving is the visualization most helpful was another task. In other words, the question of how visualization aid in such network security problems was analyzed. Other questions were analyzed in addition, such as when will the security visualization be more helpful? How does having visualization for this technique benefit the network security analysts? How to evaluate the effectiveness of network security visualization techniques? What kinds of network security techniques are expected to be more effective than the others? What types of data will be used more than the others? These questions were analyzed from the survey, producing a guide for network security designers, analysts, and researchers.

1.5 LEADING RESEARCH GROUPS AND RESEARCHERS

The leading research groups in the field of network security visualization include: University of Illinois at Urbana-Champaign (http://www.ncassr.org/projects/sift/), Virginia Tech (http://infovis.cs.vt.edu/), Utah State University (http://www.cs.usu.edu/~erbacher/), Georgia Institute of Technology (http://www.csc.gatech.edu), and University of Maryland at Baltimore County (http://userpages.umbc.edu/~jgood/research/tnv/).

Security visualization is a relatively new field. Current network security visualization techniques are giving rise to new techniques, with much improvements and modifications to meet more demands and requirements. In the past few years, not much work has been done on the field of security visualization. However, as time progresses, more interests in this field are being developed. From the chart below, it can be seen that not many research papers included in this survey came from the year before 2002-2004. It is also very obvious that most papers tend to be published in 2004. The field of security visualization will continue to grow in the following years, attracting more researchers and designers to study and explore continually in this field.

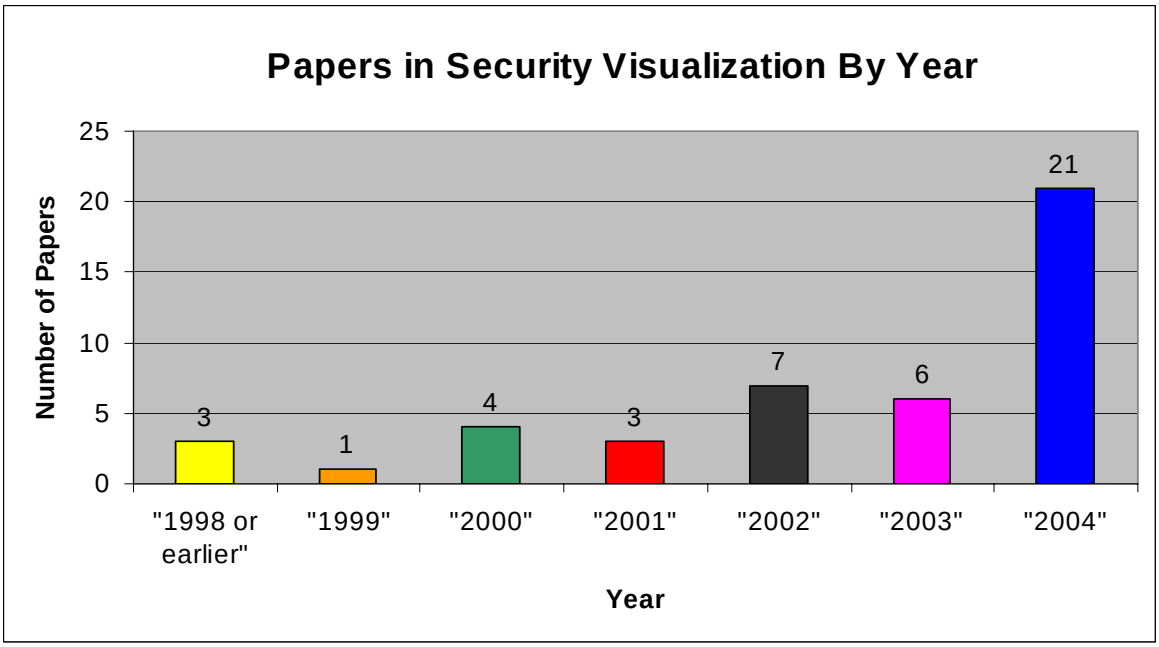

Chart 1. Papers in security visualization by year, used in the survey.

1.6 METHODOLOGY

The research was conducted specifically using information provided in other security visualization techniques. Over thirty research papers from various conferences and journals were chosen for this research. A survey was done for all the research papers on security visualization techniques. Upon analyzing each of the papers, different questions, topics, relevance information, predictions, and details were studied and investigated.

As mentioned, over thirty papers were collected from various conferences and journals. For each paper, a summary of different sections was written. These sections include: security

information being visualized (data), details on the visualization technique or techniques used, explanations of the visualization, the goal and purpose of the visualization techniques, how the network security analyst detect anomalies, the problems pertaining to this security visualization technique, comparisons of the security visualization techniques, the different user interfaces implemented and available, design issues, and results.

Furthermore, statistics on different various areas were collected and kept from each of the papers. The survey and taxonomy came from integrating the summaries of the papers. The visualization techniques were evaluated from the perspective of human cognition. In other words, the visualization techniques were evaluated as a form of problem detection and problem identification. In addition, the visualization techniques were evaluated in the form of visual search.

1.7 MAJOR CONTRIBUTIONS

This research contributes mainly to the area of network security visualization. The survey was conducted on different papers on network security visualization techniques. A survey is useful for providing information, statistics, and correlation among the techniques. The research also yields taxonomy for network security visualization techniques. As mentioned, taxonomy has been provided for the entire field of information visualization. This is too general and limited. It cannot effectively be applied to the field of network security visualization. Thus, this research provides new taxonomy, spawning detail classifications, focusing more on the area of network security visualization. The evaluation of network security visualization in this research is based upon the perspective of human cognition. All of the details above have never been done in the field of network security visualization.

1.8 OVERVIEW

Different chapters of this research concentrate on different areas of network security visualization techniques. Chapter 2 focuses on security visualization, including different network security visualization techniques, security data, tasks, and design issues. Chapter 3 provides information regarding the taxonomy, the classification of data. Chapter 4 pertains to evaluation. Finally, this research concludes in Chapter 5.

2. SECURITY VISUALIZATION

2.1 VISUALIZATION TECHNIQUES

There are many security visualization techniques, each possessing unique ways of displaying data. Some of the most common techniques include the usage of: glyphs, color maps, parallel coordinate plots, histograms, and scatterplots.

Each security visualization technique uses different types of methods in displaying data. Some of them even combines different techniques to display data, such as the techniques illustrated in [15, 22]. The measure of how efficient these security visualization techniques are depend on a variety of things, such as the type of data being displayed, how the data is displayed, and the nature of the tasks of the security visualization techniques themselves.

2.1.1 SCATTERPLOTS

Scatterplots can be thought of as a technique that combines nodes and lines to represent hosts with their connections. Usually, it is in the form of an abstract visual representation [13,

7]. Most lines often represent connections, and other icons, like circles or squares, represent hosts. Moreover, colors and shadows are used to represent different connection variations. The scatterplot can be visualized by its structure and clustering, allowing the user to decide which processes and requests are benign and malicious [5]. For the most part, scatterplots allows the user to be able to detect similarities and differences between different processes [16].

Using mostly IP addresses as their main data, scatterplots tend to display their data in two-dimensional space. The IP addresses are usually sorted by how similar they are to the monitored system. The more similar the IP address of a host to that of the monitored system is, the closer it is. Furthermore, scatterplots usually come with a detail window, since displaying information with lines and nodes are insufficient. Detail windows may include information like ports, protocols, time, date, alert IDs, and more [18].

Scatterplots are mainly used because of its advantages in representing data. Large amounts of data, such as in log files, can be visualized using graphs with lines and nodes. Each element in a scatterplot can mean different things, and tend to vary in shapes, sizes, and colors. Each of these attributes provide a clearer understanding of information rather than texturally going through a log file.

It is very useful to be able to represent different data, like the IP address or port, with a circle or square. Lines representing connection types is another issue that provides more consistency in visualization. With this approach, large amounts of information can be visualized in a single graphical screen.

The drawbacks in scatterplots also exist in many different ways. At times, studying lines and icons on the screen makes it as difficult as reading a textual log file. Also, information can be lost when lines or nodes overlap. Many of the crucial information can be misunderstood, hidden, or even obscured due to the nature of scatterplots.

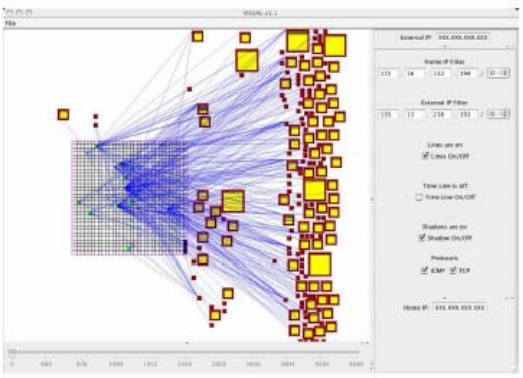

Figure 1. From [7], VISUAL displaying network data. Each small square represents home hosts, and the larger squares represent set of home hosts. The technique mainly shows network traffic, and communication between internal (on the grid) and external (in space) hosts.

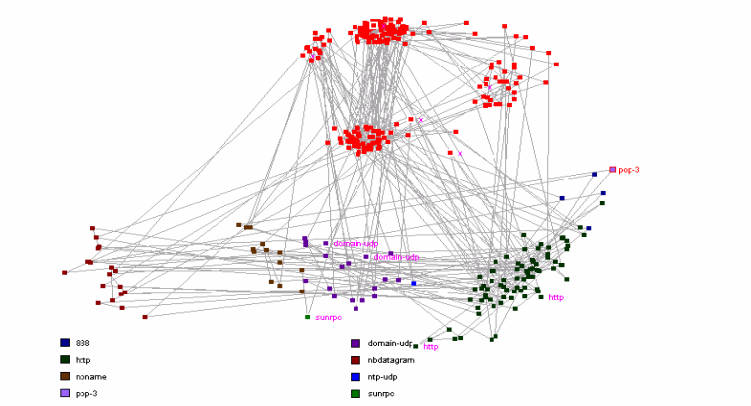

Figure 2. A scatterplot from [15]. Each data point is colored to a protocol attribute. Data points are connected by time, displaying sequence of events with similar properties. This can show similarities and the relationships.

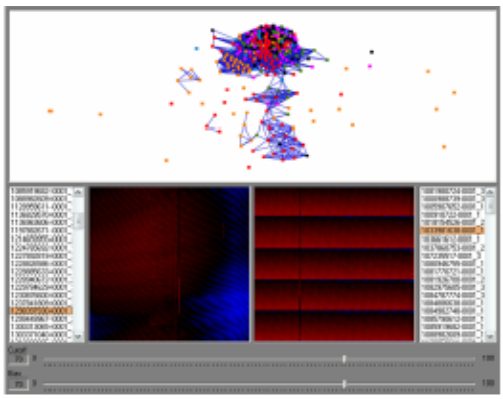

Figure 3. An example from [29], which shows both the high level and low level views. The scatterplot on top contain nodes, each representing a scan. Nodes with a higher match attract each other, and are clustered, considering they are from the same source. Each node or cluster of nodes will yield their patterns in the grids, which can be compared.

2.1.2 COLOR MAPS

When dealing with security visualization, the ability to differentiate between objects, usually nodes or hosts, is essential. Color maps make it easy for users to visualize information in security visualization.

Color maps are usually associated with some type of calculated value. In [9], typicality scores were used, and different range of colors is mapped to different typicality scores. The same approach is seen in [2, 3], where token scores derived from certain mathematical formulas are mapped to different colors. A threshold value is set, and anything above or below that value is mapped to certain colors. The main colors that are used are red, blue, green, orange, and yellow. Most of the time, red is used to signify attacks or intrusions. Blue and green are used to represent the type or connection that is normal in behavior. Orange and yellow are often used to display connections that are benign, but can possibly be an intrusion. In some cases, where there is a black background in color maps, white is also used to bring a stand out effect, making it easy to capture any anomalies.

In many areas, color maps assure to display a variety of data, but the majority, unsurprisingly, is the IP addresses and ports. In [19], the color map displays the ports in its grid. Each dot on the grid is denoted by a coordinate, given X = port / 256 and Y = port mod 256 where port is the port number (in 2 bytes). X, the horizontal, is the high bye of the port number and Y, the vertical, is the low byte. Each color of the dot is denoted by the value of the port. Thus, this is how color maps tend to represent its data. The changes and variation of each dot can be viewed by different colors, with respect to time. Here, different colors account for different variance. Black means no variance or changes, blue means small level of variance, red refers to large variance, and white denotes the most variance. In some cases, such as in [23], the IP addresses are displayed by color maps, and different colors can mean different values of the

data. Depending on the positions of the IP addresses determined by their prefixes, certain colors are assigned to each dot. Here, the brighter the background color maps as a whole, the more the changes that occurred. Different color maps use a variety of methods to represent its data, but they all share one common ground, the use of colors to represent data, or patterns.

Moreover, color maps tend to have users detect the patterns rather than alarming an intrusion. The drawbacks of color maps are that it is rather hard to display large amounts of data on a screen and have the user detect patterns without any complications or misconceptions. There is bound to be some kind of misinterpretations when mainly colors are used instead of text. The benefit of color maps, however, is that it is easier to detect any anomalies or intrusions when only colors are used. For example, when monitoring the data for any attacks, it is less difficult to see the color red rather than some text.

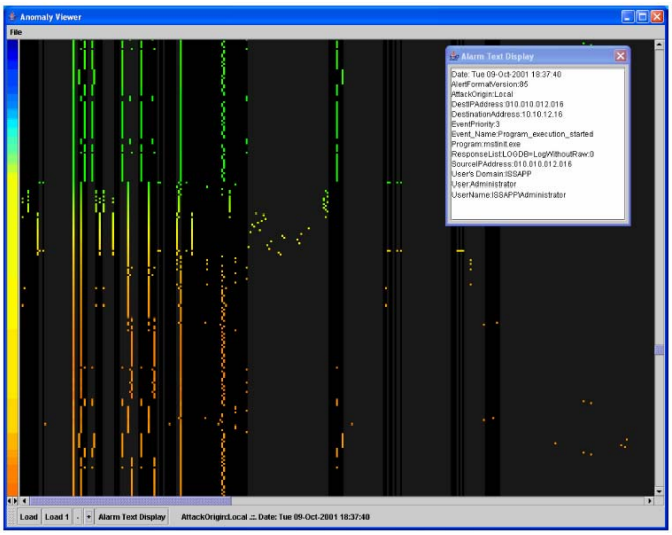

Figure 4. A color map example from [9]. The alarms are in rows, in chronological order from top to bottom. Each column represents a token (from typicality value computed). Pixels that are illuminated in the columns indicate the presence of that row's token in the alarm, and the color represents the typicality scores computed. In this case, black and gray are the absence of the tokens. The color codes on the left represent

the time of one hour, where dark blue is on the top of the hour and dark red is the bottom of the hour. A detail window on the top right provides description in details.

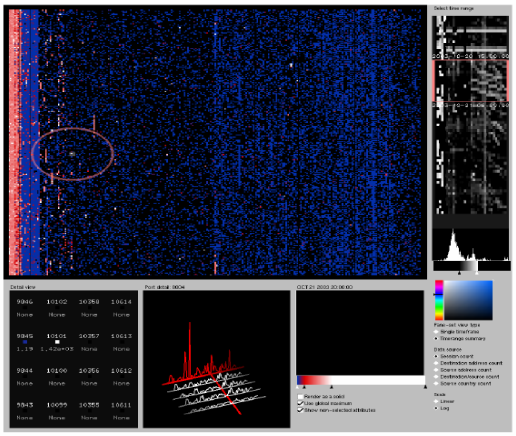

Figure 5. PortVis, from [19], contains three main frames: timeline, hour (main), and port. The main visualization here is color map. It is a 256 x 256 grid. Each dot on the grid is denoted by a coordinate, given X = port / 256 and Y = port mod 256 where port is the port number (in 2 bytes). X is the high byte of the port number while Y is the low byte. The color of each dot is denoted by the value of the port (black indicates no data). The red circle indicates the location of the selector, used in magnification.

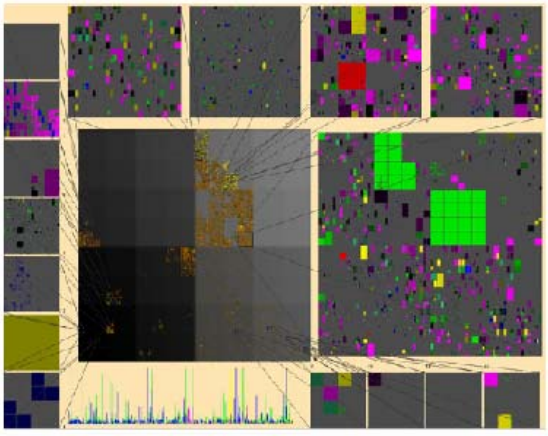

Figure 6. A 512 x 512 pixel square from [23]. Yellow represents Origin AS changes that occur on the current day while brown are the changes that occurred on the previous day. The size of the rectangles is the size of

the block of the IP address (prefix with smaller mask maps to larger rectangles). The position of each colored square is determined by the IP prefix (similar IP addresses with significant bits are close to each other), mask size, brightness, and type of change. The lines are drawn from the IP prefix to the AS number if there is the Origin AS change. The color of each line is the type of change (blue are H-type changes, pink are OS-type changes, and so forth).

2.1.3 GLYPHS

Glyphs are something that aids in visualization. An icon or a font can be considered a glyph. Glyphs are used mainly to represent an entity, just like a circle can represent a node in a network.

The use of glyphs can vary from one technique to another. Each glyph, in multidimensional data, can be mapped to multivariable entities. For example, the glyphs' location, 3D size, shape, color, transparency and opacity, and more, can yield different information for an entity [1]. Usually, these attributes are mapped to a glyph before it is displayed visually, so that some kind of comparison technique can be applied. In a network, each glyph can be connected to another glyph. Each glyph usually represents a node or a host. Most of the time, the IP address is the data that often gets mapped to the glyphs, such as in [12, 20], and glyphs of similar properties tend to be clustered or placed very closely to each other. This is done more in two-dimensional space rather than 3D. Because of this, scatterplots or some kind of color map techniques can also be rendered coherently together. In [11, 12], each glyph represents a remote connection, and the lines connecting the glyphs are the types of connection and the direction of the traffic. Colors, thickness of lines, type of lines, and length of lines are used to display different information about the connection. However, the glyphs are the main components that connect different entities of visualization.

Though glyphs are useful in security visualization in many areas, they alone display information insufficiently. Most of the time, the user cannot rely only on glyphs to display data, and capture patterns to form opinion on anomalies. In some cases, mapping multi-attributes values to a single glyph may seem like an efficient way to display data, when glyphs are clustered together. However, some other methods, such as clustering algorithm or how to map each glyph to coordinate attributes can complicate things as well. An example in [1] shows that each glyph is mapped to variables X, Y, and Z. But, how to map each entity (location, 3D size, shape, color, transparency and opacity), to a certain variable is not standardized. Therefore, various mapping to various variables will produce different graphs, and thus, different outcomes. Information, then, can be lost or hidden. The benefits of glyphs, however, is that it allows a combination of different techniques, such as color maps and scatterplots, to aid in visualization, which can bring more accurate results. Furthermore, using glyphs can also help in displaying large amounts of data on the screen, and also in the areas of comparing and contrasting similar and different elements.

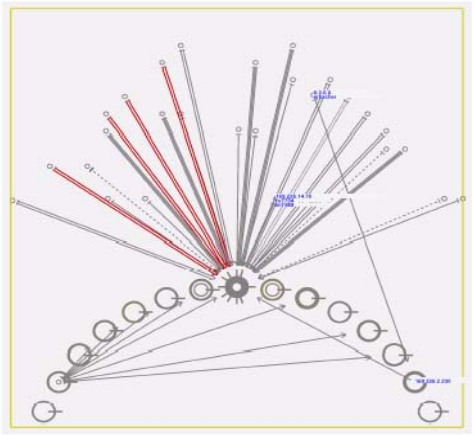

Figure 7. From [12], each big glyph represents a system. The center glyph is the monitored system (main server), and the others are connecting systems (local systems). The small circles in space are the nodes, or

hosts, positioned by their IP addresses and in a f irst-come-first-serve basis. The lines represent the type of connection as well as the direction of the traffic. Thicker lines indicate that there are more connections. Yellow lines indicate uncritical connections while red represents unusual or unexpected activities. Blue is used as an identifier. Color (by gray level) on the frames indicates the time of day (white at noon, black at midnight, and yellow represents PM). Nodes, or hosts, are sorted by their IP addresses, the most similar ones closer to the monitored system (glyph in the middle).

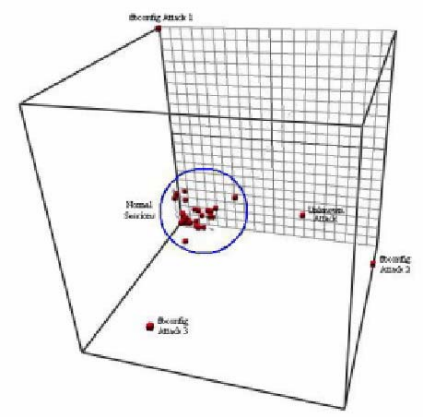

Figure 8. In this example, as in [1], the SFA, or Stereoscopic Field Analyzer, uses glyph-based volume rendering to visualize the multivariable entities, in multidimensional data. Up to nine attributes can be mapped for the entities, including a glyph's location, 3D size, shape, color, transparency and opacity, and more. These entities, or values, are assigned to variables (X, Y, and Z) coordinates, which then can be plotted. Note that the display is based on how the attributes are mapped to the variables. The plot can change when there is a change in the values assigned to the variables. In this case, similar glyphs tend to cluster together, and their relationships can be studied.

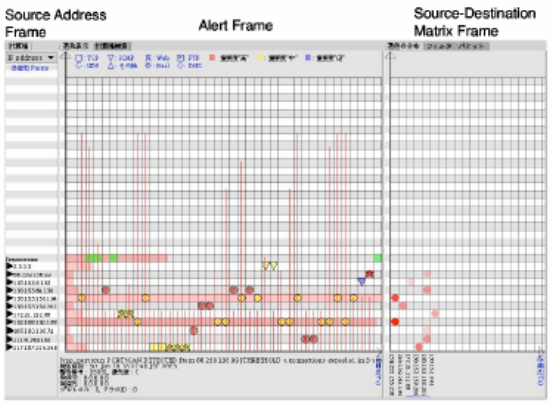

Figure 9. SnortView, from [18], contained three main frames: source address, alert, and source-destination matrix. All data are sorted by time. The IP addresses in the source address frame are listed and sorted vertically. The vertical axis in the alert frame represents the list of source IP addresses and the horizontal axis is the time. Each colored glyph, or icon, is a NIDS alert. Each color displays different priorities, red being the highest and blue is the lowest. Each shape (square, circle, star, and more) of the glyph represents the type of attack. There is a detail window on the bottom as well to provide information.

2.1.4 HISTOGRAMS

In security visualization, histograms are subject to be attached to many of the techniques. The difference between a histogram and a graph is that graphs are usually limited in information display. Histograms can provide more information, mainly activity details of a host, and different elements on a histogram can mean different things from one histogram to another. A graph can also be thought of as a combination of histograms, such as the technique mentioned in [6].

In histograms, the axes represent different things. From [6, 19], however, the main attributes usually come down to hosts and activities over time. The horizontal axis usually represents the hosts and the period of time of the activities of each host. The vertical axis

represents the activity patterns of each host. Each bar then is a histogram, and the patterns generated from this can be further studied.

The main functionality of histograms is that it displays activity patterns, which is very useful when dealing with anomaly based security scanning. Unlike graphs or scatterplots, histograms present its information in patterns that can be compared and contrasted by the user. Similarly to glyphs, histograms are also subject to be combined with other security visualization techniques, such as color maps or scatterplots, to produce more accurate and reasonable results as a whole.

Histograms lack in the area of conciseness. As mentioned, histograms generate and show a pattern, usually activity patterns. Because of this, histograms cannot show more information other than predictions to what something may be. Histograms need to be combined with other security visualization techniques to yield more reasonable outcomes. They also become very unhandy when dealing with signature based security scanning.

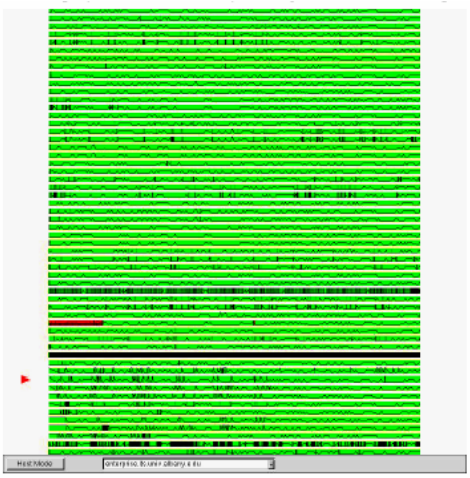

Figure 10. The graph consisted of histograms. The horizontal lines represent the hosts, and the period of time of the activities of each host. The vertical lines represent the activities of each host, thus giving a pattern. Each bar represents each host, and is a histogram. Hosts of similar patterns can be identified, and hosts of dissimilar patterns can also be spotted easily. The histograms, as a whole, form a graph. The visualization was retrieved from [6].

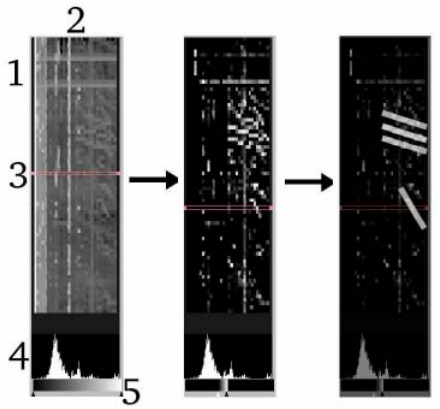

Figure 11. Another example from PortVis, [19]. The histogram, located on the bottom of the visualization, represents the frequency of the port activity levels. The gradient editor, located below the histogram, is used to emphasize a particular spike on the histogram, so that possible port scan attacks can be discovered. The histogram here is used with color maps as well as other security visualization techniques in PortVis.

Figure 12. MieLog, [22], contains different visualization areas. The histogram is represented as horizontal lines. The lines vary in length. The zoomed in screen, located left of the message screen, shows the patterns generated by the histograms. Spikes can be easily observed, leading to the detection of anomalies.

2.1.5 PAPRALLEL COORDINATE PLOT SYSTEMS

A parallel coordinate plot system is similar to a graph physically, but can display information similarly to a histogram. A parallel coordinate plot is made when each data point is projected as a line joining the components of the vector to a set of parallel coordinates [4]. The coordinates, however, is determined by different aspects. In [29], the X coordinate in the grid is denoted by the third byte of the destination IP address, and the Y coordinate is denoted by the fourth byte of the destination IP address. This can enable the user to visualize the correlation between the variables, the patterns that were generated, and the similarities or differences among the data sets. The parallel coordinate plots are usually used to present similarities and differences in the values of the data being visualized [15].

From [4], information that are included in the dimensions of a parallel coordinate plot can be things like date, URL, authorized username, binned status, bytes, and useragent. The plot can show relationships among the data. However, the IP addresses and ports are the two main data that are visualized most in a parallel coordinate plot system. Lines typically tend to represent network flows [27]. Color is often used in the plots to distinguish different traffic, directions of the traffic, or other network information. The vertical axis is usually represented by data like the IP address, or port number. The horizontal axis usually represents things like activity patterns or sequence of packets transferred in a network [10].

There are many drawbacks to the parallel coordinate plot systems. First, information cannot be sufficiently displayed using only the plots. Some sort of explanations, such as detail windows, must be implemented. The user cannot rely only on the plots to study the patterns and information generated. Also, due to the large amounts of data needed to be displayed by security visualization techniques, the coordinate plot systems immaturely cannot handle this problem. Information can be disregarded, hidden, or misunderstood when there is an increase in the number of data. This problem is also similar to the problems associated with other techniques, such as scatterplots and color maps.

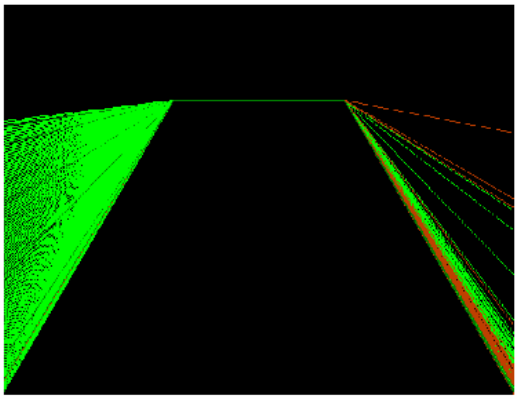

Figure 13. A parallel coordinate plot that shows relationships from the IP address to the ports. In this case, it is external port to external IP to internal IP to internal port. Variables in the plot include source and destination IP addresses, source and destination ports, and protocols (TCP or UDP). Orange is used for UDP traffic, and green for TCP. Each line represents the relationship of the variables (plot of TCP/UDP and inbound/outbound from home network). This illustration is from [10].

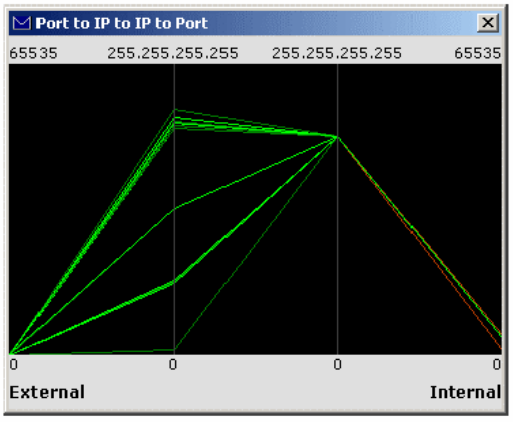

Figure 14. Another picture of a parallel coordinate plot from [10]. This is an external port to external IP to internal IP to internal IP parallel coordinate plot.

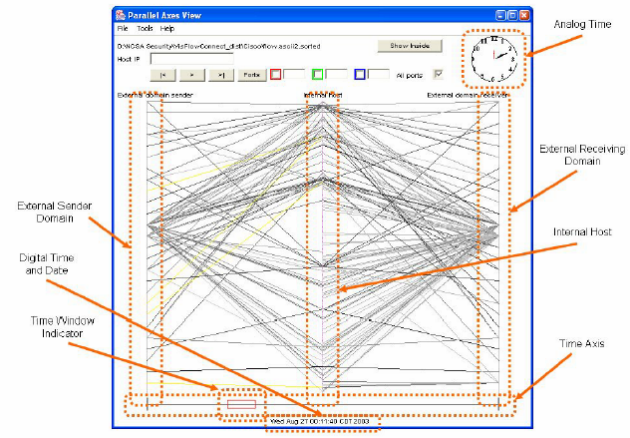

Figure 15. An example of a parallel coordinate plot from [27]. There are three vertical parallel lines that represent the external domain sender (traffic coming into network), the internal host (machines on internal network), and the external domain receiver (outgoing traffic). The horizontal axis is time. Each point is ordered by their IP address (lowest number IP on top). Traffic flows from left to right. The multitude of lines represents network flows. The darker and thicker the line, the larger the amount of traffic it contains.

2.1.6 OTHERS

Other than the mentioned security visualization techniques, there are more techniques that do not fit in any of the categories observed in this research. Mainly, these security visualization techniques are combined in conjunction with other fields, including cluster algorithms, self-organizing map algorithms, neural networks, data mining, and other types of methods pertaining to intrusion detection systems.

Even though many of these techniques provide visualization, many of them are ineffective in many areas. In [14], MineSet, a data-mining tool, was used to visualize records of connection data. However, only an abstract object and information model was constructed. The same idea was illustrated in [17, 21, 28], where intrusions were proposed to be recognized by

applying the self-organizing maps (SOM), the network-based detector using self-organizing maps (NSOM), and the resilient propagation neural network (RPROP). Data, supposedly, can be assumed to be able to be projected in plane graphs and maps for analytical reviews. However, some of these unique security visualization techniques using data mining are useful and not abstract completely. The Fuzzy Intrusion Recognition Engine, or FIRE, as proposed in [26], is a distributed intrusion detection system that uses fuzzy agents to detect anomalies that uses MineSet and GGobi. MineSet, as mentioned earlier, is a data-mining tool. GGobi is a visualization system to view and manipulate data. In FIRE, visualization is presented realistically on a two-dimensional screen, and data such as time, IP addresses, and port numbers are displayed using a scatterplot.

These techniques boast about its advantages and capabilities, but in the end, the abstract cannot complete the task physically like other security visualization techniques. Security visualization alone is a complicated subject. Therefore, combining it with other areas will make complications greater in detail. Accuracy in data, information lost or unassisted, and knowing whether or not the algorithms or methods produced the right outcome are some of the problems associated with these techniques. However, these techniques may also bring effectiveness. Abstracts may bring good predictions and ideas to new security visualization techniques. But, it is a fact that the majority of these techniques do not produce effective outcomes in their results. Thus, only ideas and predictions can be used, not concrete results.

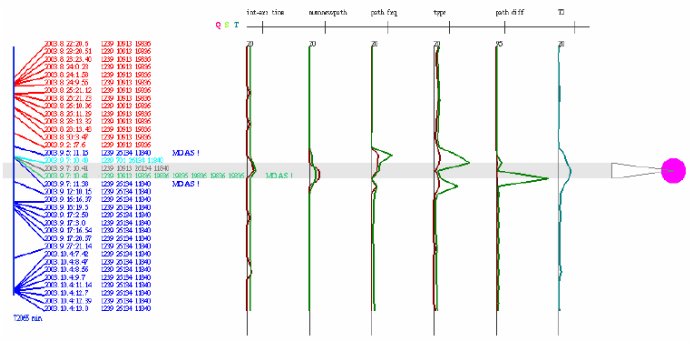

Figure 16. An example of a security visualization technique that uses data mining, from [24]. Anomalies can be detected by viewing the vertical lines, which contains messages from data mining. Deviations in the lines can alert the user that there may be an intrusion or attack. The anomalies are highlighted in gray.

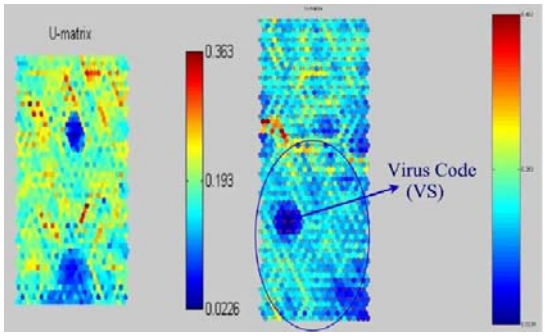

Figure 17. From [28], a security visualization technique that incorporates self-organizing maps (SOMs) and a unified distance matrix, or u-matrix. This technique displays neurons on the grids. Different colors are used to represent different distances between neurons. The display allows virus patterns to be visualized.

2.1.7 CHART

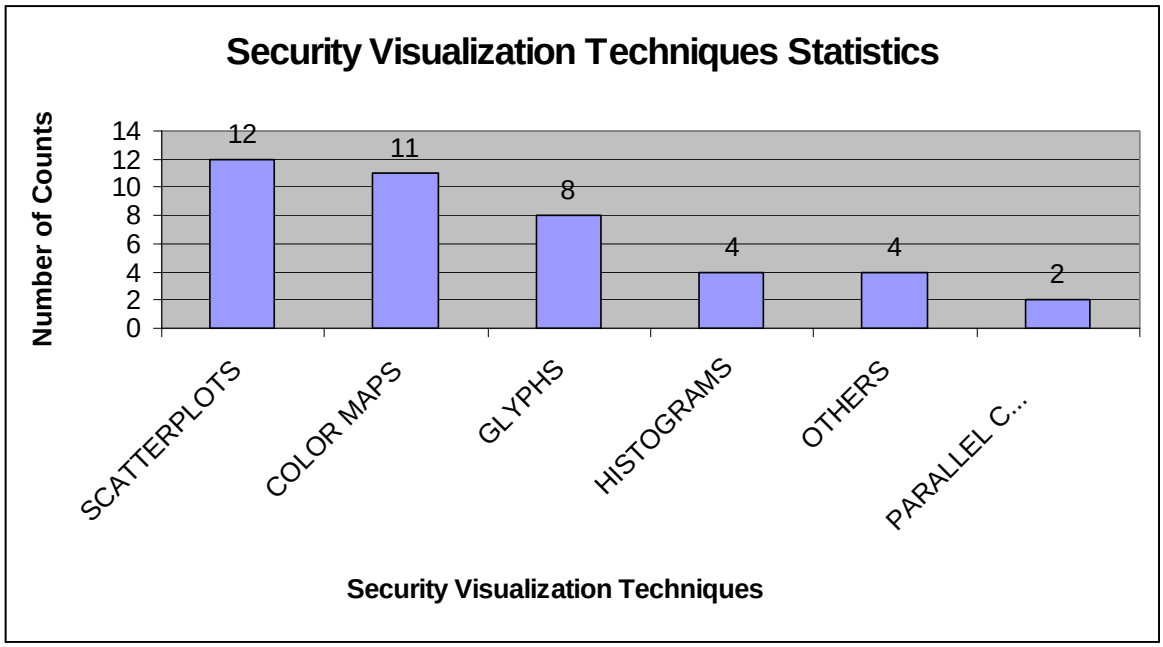

Chart 2. Security visualization techniques statistics comparing different types of visualization methods used.

2.1.8 DISCUSSION

It can be denoted that different security visualization techniques use different methods of displaying data with a common goal, to efficiently detect an anomaly. From the research, it can be concluded that the different security visualization techniques vary in number from one technique to another.

Scatterplots and color maps account for more than 50 percent of the security visualization techniques. This number comes with no surprise, since these two techniques are the main components in security visualization. Other techniques tend to combine it with these techniques. Such techniques, for example, are ones that have glyphs combined with color maps, or color

maps with histograms. Undoubtedly, scatterplots and color maps are the most common security visualization techniques used throughout.

Other techniques, such as glyphs and histograms, are also more commonly used than the other techniques, such as parallel coordinate plot systems. Glyphs are the next more commonly used security visualization technique. A group of icons, circles, squares, rectangles, or even node-like shapes can be considered a glyph system. Because of this, the number of counts for glyphs is high, but not as high as scatterplots and color maps. If scatterplots and color maps are considered an essence in security visualization techniques, or a mandatory entity in all visualization, then glyphs can also be concluded to be the most commonly used technique. Histograms come next on the list. The use of histograms, at this point, is not entirely great yet, but it can be predicted that in the near future, histograms can contribute more to security visualization. Currently, the amount of information a histogram displays is limited. Thus, not many security visualization techniques have included the use of histograms in their visualization yet.

The last two security visualization techniques are the ones that account for the least in this research. The others refer to techniques with association with other methods, such as data mining, neural networks, cluster algorithms, self-organizing maps algorithms, and other intrusion detection methods. They account for nearly 10 percent of the techniques. This is surprising, however, when compared to other techniques. This is relatively high for this research, since in most of these methods, visualizations were very limited in many areas. Lastly, the parallel coordinate plot system accounts for the least in all techniques. Shockingly, it accounts for less than 5 percent of the total techniques. The reason being so is predicted that parallel coordinate plot systems tend to overlap with scatterplots, and more security visualization techniques tend to

prefer scatterplots. Similar types of information and visualization can be drawn between the two techniques. Because of this, there is less use of parallel coordinate plot systems. Overall, the most surprising information gathered from the statistics chart is that the parallel coordinate plot system techniques account for less than the others (data mining, neural network, cluster algorithms, self-organizing maps algorithms, and methods of intrusion detection systems) techniques.

There are still many empty voids in these security visualization techniques, which can be assisted. Most of the security visualization techniques are in two-dimensional rather than 3-D. The first problem is that security visualization is still in its infant state. Not much development has been implemented in this area. Having a 3-D screen, at this point, can bring more complications. Information misconception is one of the main problems. Data can be hidden, unvisualized, and occluded. Displaying data on a 3-D screen still seems like a difficult task to accomplish. Two-dimensional screens are more common currently in security visualization techniques because all entities are presented on a plane. Even though some information or data can still be lost or visualized inaccurately, it is less complicated to use two-dimensional screens. It can be predicted surely that in the near future, more visualization improvements will take place, and the use of more 3-D screens will come into effect, even though current techniques find two-dimensional screens much simpler and more fitting.

2.2 SECURITY DATA

Different security visualization techniques tend to use different types of data in their visualization. Each visualization technique displays different types of data. Some of the data most commonly used in security visualization techniques include the source and destination IP

addresses, the source and destination ports, protocols, usernames, time and date, number of bytes transferred and received, types of connection, and more.

2.2.1 LOG FILES

Log files are used heavily throughout many of the security visualization techniques. Log files contain data information, usually network traffic data, such as IP addresses, ports, or connection information, which is used by the techniques. Usually, log files are great in size, and are rather difficult to be displayed graphically. Reading them textually is an overwhelmingly aggravating task.

Log files are being used in security visualization because they are the raw data that is fed in to most of the security visualization techniques to be visualized. They contain a huge amount of information, and being able to display many of its details is one of the essential goals of security visualization.

```
Jan 9 12:15:12 visualizer-s.cs.albany.edu xinetd[899]: START: pop3 pid=28097 from=169.226.2.54
Jan 9 12:15:12 visualizer-s.cs.albany.edu xinetd[28097]: USERID: pop3 WIN32 : Analyst
Jan 9 12:16:31 broomstick.cs.albany.edu in.telnetd[16593]: connect from root@cs.albany.edu
Jan 9 12:16:31 cs.albany.edu in.telnetd[16593]: connect from root@cs.albany.edu
Jan 9 12:22:29 visualizer-s.cs.albany.edu CROND[28100]: (root) CMD ( /sbin/rmmod -as)
Jan 9 12:25:31 broomstick.cs.albany.edu in.telnetd[16628]: connect from cdial20.infoblvd.net
Jan 9 12:25:31 cs.albany.edu in.telnetd[16628]: connect from cdial20.infoblvd.net
Jan 9 12:26:02 cs.albany.edu named[25266]: dangling CNAME pointer (google.lb.google.com)
Jan 9 12:29:45 cs.albany.edu in.telnetd[16654]: connect from Workstation72.ctg.albany.edu
Jan 9 12:29:51 von.cs.albany.edu in.rlogind[5625]: connect from pb@broomstick.cs.albany.edu
Jan 9 12:30:13 visualizer-s.cs.albany.edu xinetd[899]: START: pop3 pid=28101 from=169.226.2.54
Jan 9 12:30:13 visualizer-s.cs.albany.edu xinetd[28101]: USERID: pop3 WIN32 : Analyst
Jan 9 12:31:30 cs.albany.edu named[25266]: Cleaned cache of 799 RRs
Jan 9 12:31:30 cs.albany.edu named[25266]: NSTATS 979061490 977153081 Unknown=6 A=393521 NS=3
CNAME=98 SOA=9575 PTR=73966 MX=15120 TXT=10 AAAA=42 AXFR=32 ANY=12019
Jan 9 12:31:30 cs.albany.edu named[25266]: XSTATS 979061490 977153081 RR=198301 RNXD=66697
RFwdR=150932 RDupR=302 RFail=619 RFErr=0 RErr=17 RAXFR=32 RLame=16943 ROpts=0 SSysQ=23483 SAns=373313
SFwdQ=131146 SDupQ=30183 SErr=0 RQ=504450 RIQ=0 RFwdQ=131146 RDupQ=2489 RTCP=1069 SFwdR=150932
SFail=3460 SFErr=0 SNaAns=68541 SNXD=241409
Jan 9 12:32:28 visualizer-s.cs.albany.edu CROND[28103]: (root) CMD ( /sbin/rmmod -as)
Jan 9 12:34:07 karp.cs.albany.edu in.telnetd[27063]: connect from nas-70-57.albany.navipath.net
Jan 9 12:34:17 cs.albany.edu named[25266]: dangling CNAME pointer (gd25.doubleclick.net)
Jan 9 12:42:29 visualizer-s.cs.albany.edu CROND[28105]: (root) CMD ( /sbin/rmmod -as)
Jan 9 12:45:12 visualizer-s.cs.albany.edu xinetd[899]: START: pop3 pid=28106 from=169.226.2.54
Jan 9 12:45:12 visualizer-s.cs.albany.edu xinetd[28106]: USERID: pop3 WIN32 : Analyst
Jan 9 12:52:29 visualizer-s.cs.albany.edu CROND[28108]: (root) CMD ( /sbin/rmmod -as)
Jan 9 12:52:33 karp.cs.albany.edu in.telnetd[27137]: connect from 169.226.14.70
Jan 9 13:00:12 visualizer-s.cs.albany.edu xinetd[899]: START: pop3 pid=28109 from=169.226.2.54
Jan 9 13:00:12 visualizer-s.cs.albany.edu xinetd[28109]: USERID: pop3 WIN32 : Analyst
Jan 9 13:02:29 visualizer-s.cs.albany.edu CROND[28111]: (root) CMD ( /sbin/rmmod -as)
Jan 9 13:03:29 visualizer-s.cs.albany.edu CROND[28113]: (root) CMD (run-parts /etc/cron.hourly)
Jan 9 13:08:30 cs.albany.edu in.telnetd[16702]: connect from cm-24-29-78-15.nycap.rr.com
Jan 9 13:11:43 karp.cs.albany.edu in.telnetd[27175]: connect from grande.cs.albany.edu
Jan 9 13:12:29 visualizer-s.cs.albany.edu CROND[28115]: (root) CMD ( /sbin/rmmod -as)
Jan 9 13:14:55 cs.albany.edu named[25266]: dangling CNAME pointer (md1.doubleclick.net)
```

Figure 18. An example of a log file, over the course of one hour from a lightly loaded environment, from [13].

Network security visualization techniques heavily use log files. Moreover, each of the technique tends to use different types of log files. Some techniques use regular system log files or log files from a web server. Some techniques use NIDS log files, and some even use their own type of log files, such as NetFlow log files. There are exactly 21 counts of log files used in the survey, and seven different types of log files surveyed. According to the statistics chart, more than 30 percent of all the log files used were regular system log files. Almost 20 percent use log files from a web server. The rest of the log files, such as NIDS log files from Real Secure and Black Ice detectors, NetFlow, Firewall, BGP (Border Gateway Protocol), and TCP dump, contribute to nearly half of all the log files used.

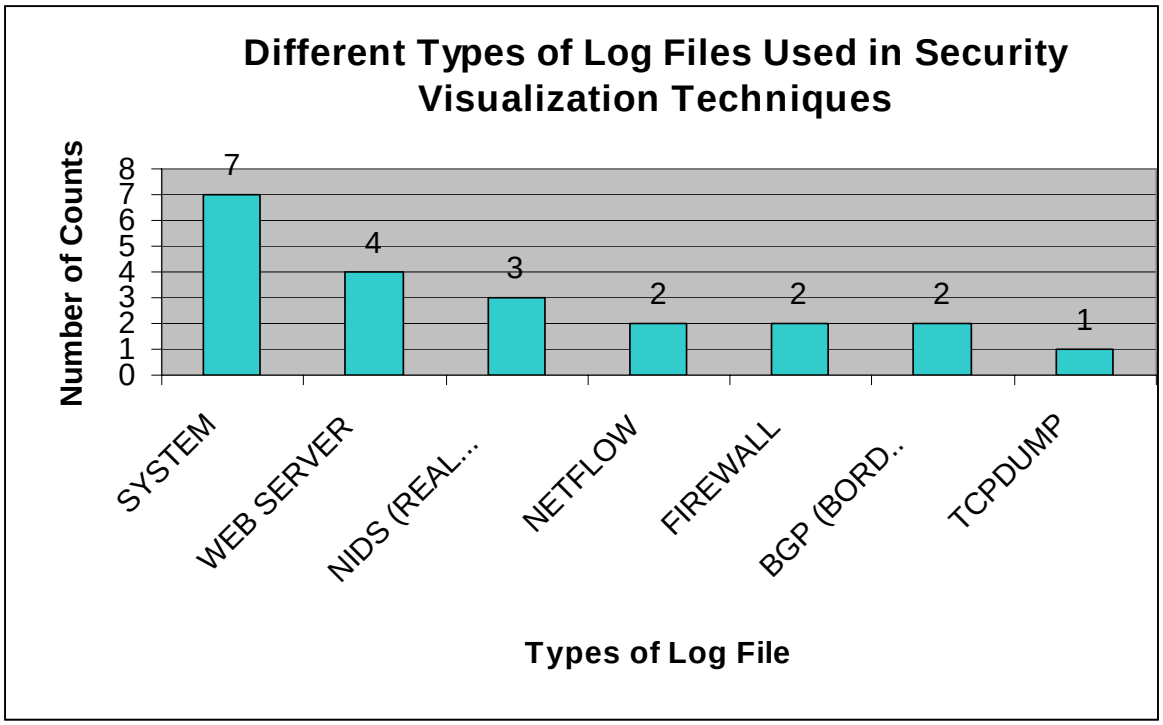

Chart 3. Different types of log files used in security visualization techniques.

2.2.2 IP ADDRESSES

The IP addresses are something that most security visualization techniques rely on. Ranging from 0 to 255, the IP address is a 32-bit host address defined by the Internet Protocol. It is represented using in dotted decimal notation, and uniquely identifies a node on the Internet. The IP addresses are usually divided by source and destination in security visualization, and each is used heavily. The IP address provides uniqueness in security visualization.

In security visualization, the IP address usually represents a host or a node of the connection. Most of the time, it signifies the presence or existence of the attacker or the host being attacked.

There are different ways how network security visualization techniques encode the IP addresses. The main ones are limited to: 1.) locating the IP addresses in space through the use of graphs, plots, or histograms, 2.) applying X-Y coordinate systems to map the IP addresses based on a set and consistent method (such as different byte values of the IP addresses), 3.) using different colors to denote different IP addresses (usually done in color maps), and 4.) basing the IP addresses depending on the value of the IP addresses (usually done in scatterplots), where similar IP addresses are clustered together.

IP addresses are very important to security visualization. As data, they represent the uniqueness of each individual in the system. If an anomaly or an attack is detected, the IP address represents the individual who is responsible for such intrusion. It can spell out the location of the attacker, the properties and characteristics, the frequencies of the attacks if a record was taken for identification, and in some cases, the associations of the attacker to the system. The IP addresses provide a sense of character in the system, and gives details about the different components within the system. Because of all this, the IP address is most commonly visualized as data in security visualization.

From the chart, more than 45 percent of the methods of encoding the IP addresses fall into the category of locating the IP addresses in dimensional space. This refers to simply putting the IP addresses in space, and representing them as dots or circles, and using lines, graphs, plots, or histograms, to visualize the IP addresses. The remaining methods fall into the rest of the categories, where the IP addresses are visualized through X-Y coordinates, colors, and similarities of the IP addresses. In X-Y coordinates, usually the IP addresses are visualized through their values. For example, the X coordinate is represented by the third byte of the IP addresses while the Y coordinate is represented by the fourth byte. In using colors to represent

the IP addresses, usually different colors denote different types of categories, such as red can mean attacks, blue means intrusions, or white means benign. Lastly, the IP addresses can be visualized through similarities. Here, the IP addresses are placed according to how similar their numbers are. Similar IP addresses are clustered together, forming patterns. Sometimes, the location of the IP addresses are also based on how similar their values are to a monitored system, which aids in visualizing the system as a whole.

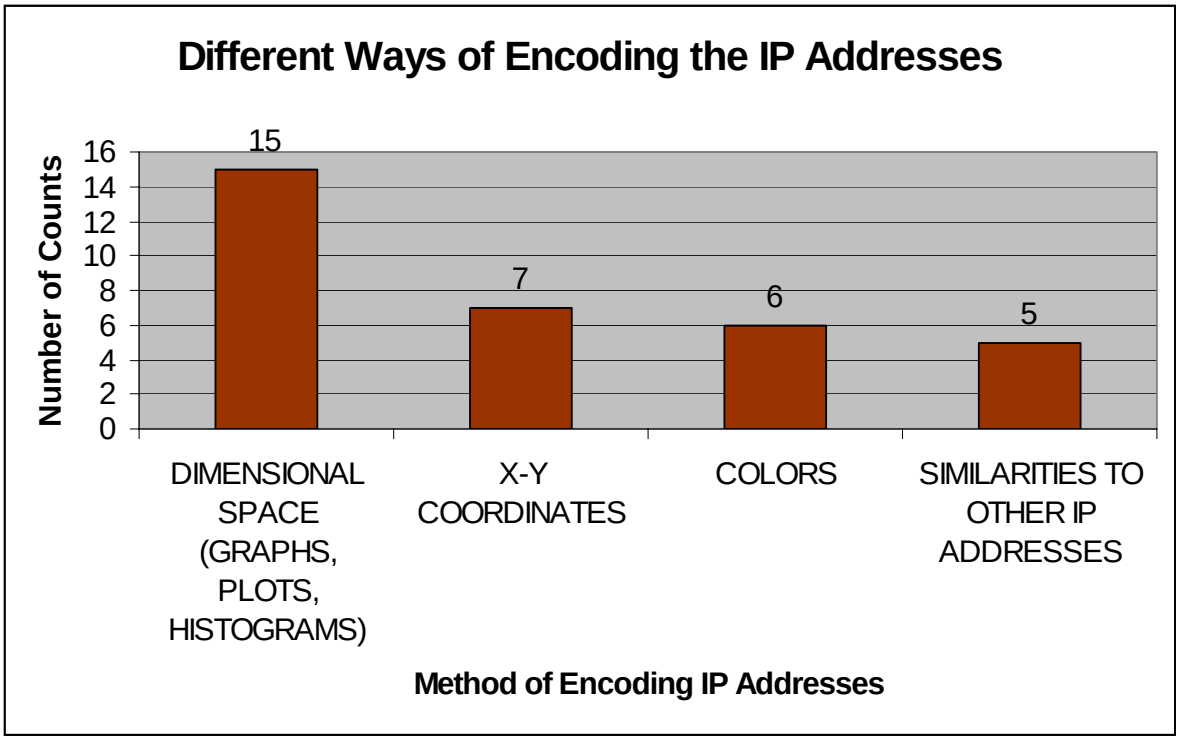

Chart 4. Different methods of encoding the IP addresses.

2.2.3 PORTS

A port is a computer circuit consisting of the hardware and associated circuitry that links one device to another. It is a logical channel or channel endpoint in a communications system. In TCP/IP and UDP networks, the port number identifies what type of port it is.

Port is something crucial in security visualization. When an attack occurs, for example a port scan attack, the attacker scans the ports that are available or vulnerable, and attacks it. Therefore, keeping the ports secure is vital. Attackers tend to target ports because it is the channel through which communication flows.

Ports information tends to be visualized differently. Though they are not as commonly used as IP addresses, ports also are important in security visualization. Some of the more common techniques of visualizing ports are the usage of colors, histograms, and X-Y coordinates. In the usage of colors, different colors are applied to different types of ports. This allows security analysts to visualize the different port types. In histograms, such as the example given in PortVis [19], the port activities are visualized. Different spikes represent different types of activities, depending on the histograms. Also, similarly as the IP addresses in X-Y coordinates, the ports are located based on their values.

From the chart, it appears that each different methods of visualizing ports in security visualization techniques are scattered almost equally. However, the use of X-Y coordinate accounts for the least in this area, but unsurprisingly not very much behind the other methods. Each of the methods accounts for approximately 37.5 percent except for the X-Y coordinate, accounting for 25 percent.

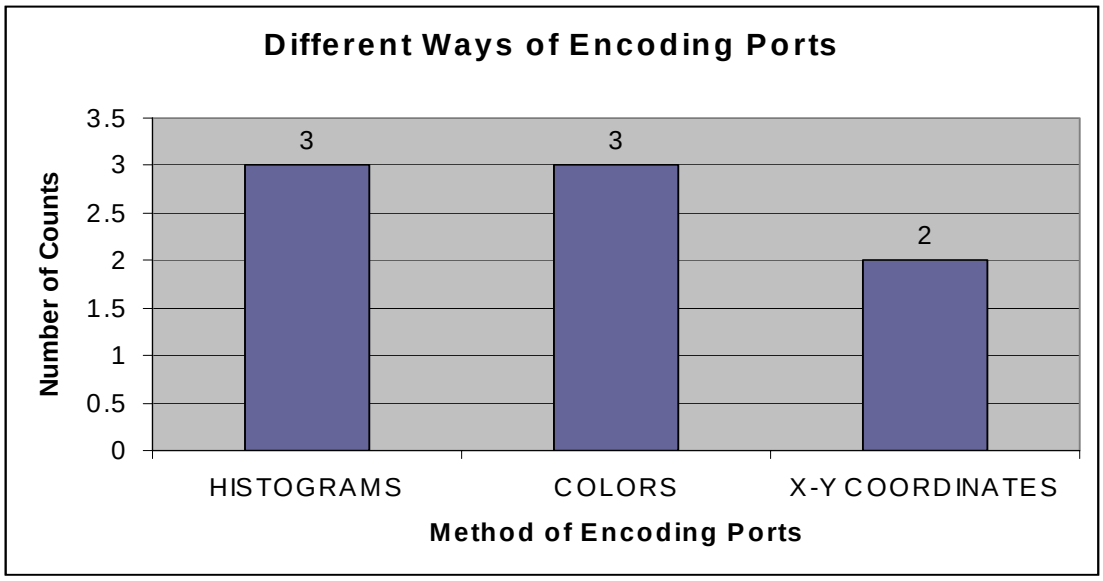

Chart 5. Different methods of encoding ports.

2.2.4 PROTOCOLS

A protocol is a set of rules describing how to transmit data, usually across a network. TCP and UDP are the two main types of protocols used in security visualization. TCP, or Transmission Control Protocol, is one of the main protocols in TCP/IP networks. It deals mainly in terms of packets, and is responsible for enabling connections and exchanges of data streams between hosts. TCP is also responsible for the delivery of data, and the transmission of packets, in the order they are sent.

UDP, or User Datagram Protocol, is one of the other main protocols aside from TCP. UDP is a connectionless protocol that also runs on top of IP networks. However, UDP acknowledges very few error recovery services. It also offers a direct way to send and receive

datagrams over an IP network. UDP is mainly used for broadcasting messages over a network or system.

Therefore, the understanding and methods of securing protocols are vital, due to the nature of the activities of protocols. In security visualization, protocols contribute to the passing and transfer of information, which is security information. Detecting security breaches in protocols, as a result, is another important task in security visualization.

2.2.5 USERNAMES

Usernames are like login names that uniquely identifies the names of the users. This is because users need to identify themselves in many areas, especially in network connections. Usually, usernames are associated with passwords for logging. Usernames are often short strings of alphanumeric characters. Common choices are first name, initials, or some combination of first name, last name, initials and an arbitrary number.

Usernames are used very rarely in security visualization. This may be because the IP addresses already uniquely identify the users, rather than having usernames. The IP addresses are more useful than the usernames.

2.2.6 TIME AND DATE

Time and date usually refer to the time and date of the connection. Time stamps are usually tagged to the connection information. It is important to be able to keep tract of the time and date of each connection, assuming that these information will become useful when analyzing attacks or intrusions.

Time and date are important in security visualization for a variety of reasons. First, the time and date allows security analysts to note the time and date the intrusions or attacks occurred. Moreover, having records of such data is crucial when dealing with future attacks. Similarities in time and date of attacks, as well as other related information, can be gathered. Furthermore, in a network, the ability to monitor every component at all time is ideal. Usually, traffic depends on time. Busier traffic runs during rush hours, while slow or normal traffic runs during off peak hours. Because some security visualization techniques use network traffic data, knowing how much traffic runs on certain time intervals is an asset. This also aids in the ability to detect intrusions or possible false positives. For example, in a monitored system, if there is normal traffic at specific times and busy traffic at certain times, it can be infer that having heavy traffic during certain times is usual, and not dangerous. This saves security analysts the work of monitoring such system and finding out they are just false alarms. In a way, time and date also serve as unique patterns for security analysts to compare and contrast.

2.2.7 CONNECTIONS INFORMATION

The types of connection refer to the different methods of connecting, or the different ways to connect. Information such as these can be the number of bytes transferred or received, the servers information, how the user connects to the network, or other connection information. Again, these information are useful in analyzing attacks or intrusions.

2.2.8 CHART

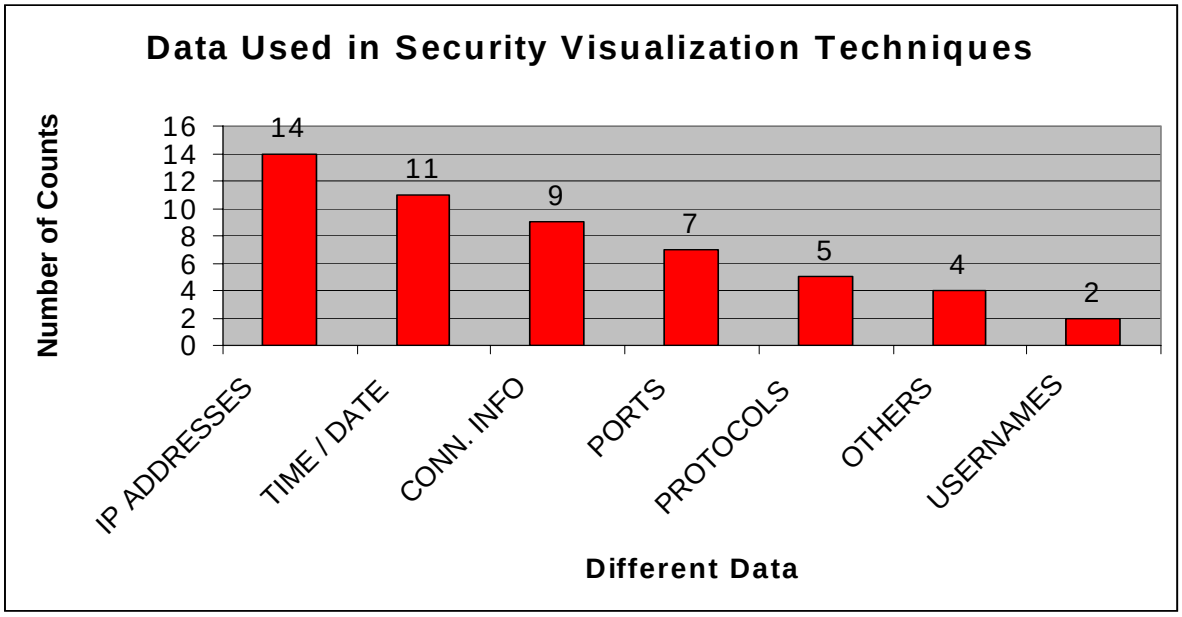

Chart 6. Statistics of data used in security visualization techniques.

2.2.9 DISCUSSION

From the chart generated above, it appears that the IP address and the time and date accounts for almost 50 percent of the data used in security visualization. The IP addresses, unsurprisingly, is the data used most frequently. The usernames, as predicted, are the ones that are least likely to be used. The connection information, ports, and protocols are used generally at the same level throughout.

The others field in the data list refers to data generated by neural networks, data mining, and other intrusion detection systems. Even though data from these methods are not used very often in security visualization, it can also be concluded that this field is still more than the usernames.

The IP address can be thought of as something that is almost required in every security visualization techniques. It is the main data in visualization. Displaying the IP address is required, in order to provide the ability to be able to identify an individual in the connection, which differentiates different entities overall.

2.3 TASKS

Different security visualization techniques focus on different tasks. It is not proper to conclude that all security visualization techniques have one goal in common, finding anomalies or finding intrusions. Some of the techniques focus on other areas in security visualization. The main tasks of security visualization techniques observed were narrowed down to three categories: finding intrusions, finding false alarms (false positives), and training classifiers. Each security visualization technique can have one or more tasks. However, at this point in the field of security visualization, it is not too general to limit each technique to one task. Having multiple tasks for a single technique is too broad and complex. The focus should be limit to one task for one technique, so that more effectiveness can be derived singly rather than being able to handle multiple tasks but not very effective in any way.

2.3.1 FINDING MALICIOUS ACTIVITIES

Finding malicious activities is the most common task in security visualization. Being able to visualize anomalies or intrusions is the ideal goal. This can be done by signature based security scanning or anomaly based security scanning. Finding intrusions in security visualization techniques vary from one technique to another. Some techniques, such as the technique in [9], uses color maps to find an alarm or an intrusion. Different colors are used to

display different entities, and the intrusion or the alarm is displayed as a certain color. Other techniques, such as MieLog in [22], use histograms to display possible attacks. The long spikes in the histograms in the outline area make the attacks seem obvious. Some of the common attacks, or intrusions, that current security visualization techniques are able to detect are: denial of service (DoS) attacks, ffbconfig attacks, dictionary attacks, portsweep attacks, protocol attacks such as TCP SYN flooding, port scan attacks, and other network intrusion detection alerts.

2.3.1.1 SIGNATURE BASED

Detecting malicious activities through signature based security scanning, though not as common as anomaly based security scanning, is still a useful way to find intrusions, or malicious activities. Here, basically the signature streams of the data are compared. Usually there is a list of signatures contained in a system. The technique sounds the alarm if there is a match between the signatures in the list to a signature obtained from the requests.

2.3.1.2 ANOMALY BASED

Detecting malicious activities through anomaly based is more common in the security visualization techniques. In an anomaly based security scanning, the patterns of the requests are compared to a normal pattern. Detection occurs when there is an outliner, or any shift or apparent differences in the graphs or patterns. Detecting malicious activities through any types of graphs, patterns, or plots can be considered an anomaly based security scanning.

2.3.1.3 CHART

From the chart provided, the statistics obtained was not surprising. More than 73 percent fell into the category of anomaly based. Security visualization techniques that detect malicious activities tend to use anomaly based security scanning rather than signature based security scanning. This is mainly because most of the techniques use some form of pattern recognition to detect intrusions or attacks. Signature based security scanning is also used to detect malicious activities, however, but not as commonly used as anomaly based.

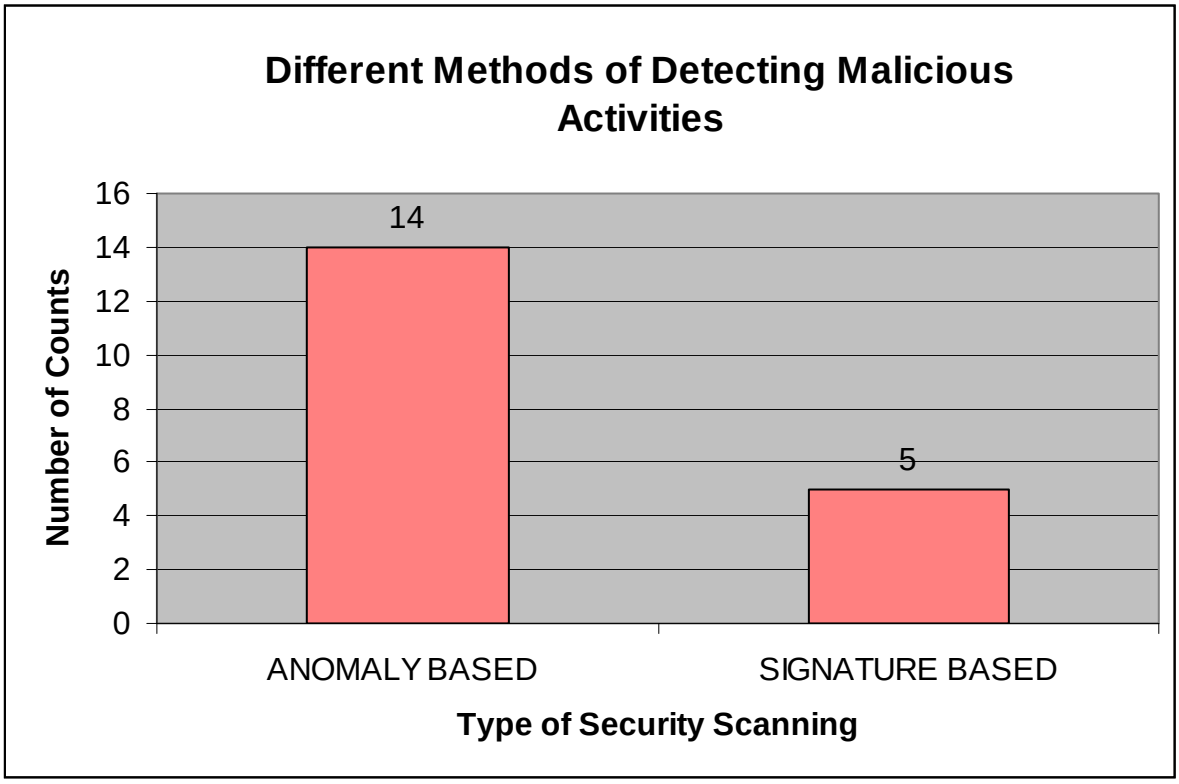

Chart 7. Different ways of detecting malicious activities, through anomaly based or signature based security scanning.

2.3.2 FINDING FALSE ALARMS

Finding false alarms, or false positives, is another task in security visualization techniques. By simplified definition, false positives occur when normal traffic is detected as an attack. False negatives occur when the system, or technique, fails to alarm real attacks. It is more common, in security visualization, to detect false positives than false negatives, since false negatives occur when a technique cannot function effectively. As mentioned in closely over 80 percent of all papers observed, false positives is a big problem in security visualization. Since data files, such as log files, are huge in size, detecting false alarms are not any less. Being able to decrease, if not eliminate, false alarms can also be an ideal goal in security visualization. False alarms account for much time and effort of users in security visualization; time that could have been spent finding real attacks or intrusions. As a result, some of the security visualization techniques turn their focus on finding false positive rather than finding intrusions or attacks. In [25], a clustering algorithm is used to implement a graph, which shows an abstraction of network traffic. The nodes on the graph represent the computers, and the edges represent the communication between the computers. The visualization of the traffic depends on the location of the nodes, which are grouped according to their communication strengths. The main visualization goal of this graph clustering and drawing is to detect false positives rather than finding intrusions or attacks. In [21], the technique implements two tasks, finding false positives and finding intrusions. Normal traffic is displayed as nodes that cluster together. Abnormal traffic does not cluster. However, the cluster algorithm, using SOMs, can detect which unclustered nodes are attacks, and which are false alarms. This is, more or less, as useful and beneficial as finding intrusions in security visualization.

2.3.3 TRAINING CLASSIFIERS

Another important task in security visualization is training classifiers. This refers to being able to classify the data as raw data or classified data. Raw data, of course, are data that are straight from log files, or are things like raw IP addresses, ports, and others. Classified data are data that have been classified to categories, such as attacks, possible attacks (alarms or suspect), or that of non-intrusive nature. Usually, these classified data are the output of other network intrusion detectors. NIVA, from [20], uses network intrusion detectors, like Black Ice or Real Secure, to classify its data. Another example from [29] illustrates the use of classified data. On the detail scatterplot window, the technique allows for the display of the IP addresses being clustered together. On the lower view, the IP addresses are viewed on two 256 x 256 grid, side by side in pairs. The technique assumes that the IP addresses have been scanned thoroughly before being displayed. This is an example of classified data. Thus, training classifiers, though the statistics for such task is relatively low, can be thought of as another functional task in security visualization. Overall, training classifiers is a task that is not at the same level as detecting intrusions or finding false positives. This is due to a variety of reasons. In the future, there can be predictions that the number of training classifiers in security visualization techniques will increase.

2.3.4 CHART

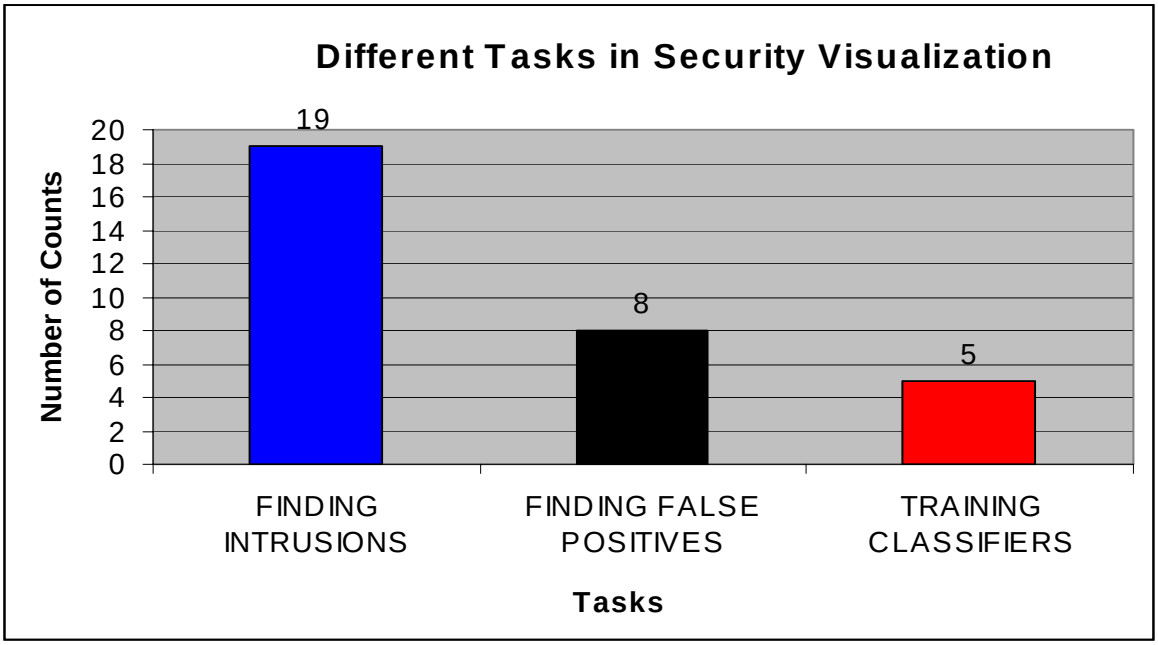

Chart 8. Comparing different tasks of each security visualization techniques.

As can be observed from the chart, the task of finding intrusions in security visualization accounts for nearly 60 percent of the tasks. Finding false positives account for approximately 25 percent, and training classifiers are about 15 percent. Moreover, detecting intrusions and false positives, together, account for almost 85 percent of the total tasks. It comes to no surprise that finding intrusions is the task most security visualization techniques target. Training classifiers are still at its development stage, where finding false positives are continuing to grow, in some cases may be at the same level as finding intrusions, in the future.

2.4 DESIGN ISSUES

2.4.1 HIGH DIMENSION DATA

These security visualization techniques require their display of data to be as efficient as possible, possessing simplicity but at the same time, containing sufficient amounts of information. High dimension data is not considered a problem in security visualization, but more of a difficulty.

Different security visualization techniques deal with their essence of high dimension data in different ways. The amount of information to be displayed on the screen cannot be insufficient, nor should it be too complicated for users to understand and analyze. When there is too little information displayed on the screen, there are too many holes and voids in the data display, and not enough data can be visualized. When there is too much information, data can be hidden or occluded. The key is, then, to have balance between the two.

Some visualization, for example SnortView in [18], uses five parameters in its data, which are: time of access, type of access, source of access, destination of access, and details of access. It contains three main frames: source address, alert, and source-destination matrix frames. Each frame is designed to visualize different parameters. The IP addresses are sorted and displayed in the source address frame. Colored icons, which represent NIDS alert, as well as time, are displayed in the alert frame. It was mentioned that the information displayed is sufficient. However, according to SnortView, the source-destination matrix was not needed initially. It was implemented after the two frames because the technique needed a way to display both source and destination IP addresses in conjunction, joined by vertical and horizontal lines. So, SnortView has three main frames to display its different parameters of the data being visualized.

PortVis, [19], contains five to six visualization areas. PortVis uses raw data from an external text file. The data contained network traffic information. However, the idea is that the

data is reduced to a set of counts of entities rather than the content itself. For example, instead of a list of each TCP session, the data contains information on the counts of how many TCP sessions exist. The main parameters used in PortVis are: protocol, port, and hour as the main fields. The other remaining five fields are: session count, source address, destination address, source/destination address pairs, and source countries. The three main visualization areas are timeline, hour, and port. The timeline contains a histogram and the gradient editor to display the activities of the ports. The hour, or main, visualization displays dots on the grid where the location of each dot is denoted by X = port / 256 and Y = port mod 256 where port is the port number. The port number uses the 2-byte classification, where X is the high byte of the port number and Y is the low byte of the port number. Moreover, the port activity visualization is used to display the port activities in histograms. Different visualization areas in PortVis allow different parameters of the data to be visualized.

According to the statistics chart, different groups of the number of parameters included in security visualization techniques are shown. It seems safe to say that the majority of the techniques display zero to three parameters, or well over 56 percent. The group of four to six and seven to nine parameters account for approximately 40 percent. There is only one technique that includes ten or more parameters in its method. It is proper to include that at this point in security visualization, the number of parameters is still relatively not very large in most techniques. As more techniques advance, more parameters will certainly be used, yielding more complicated visualization.

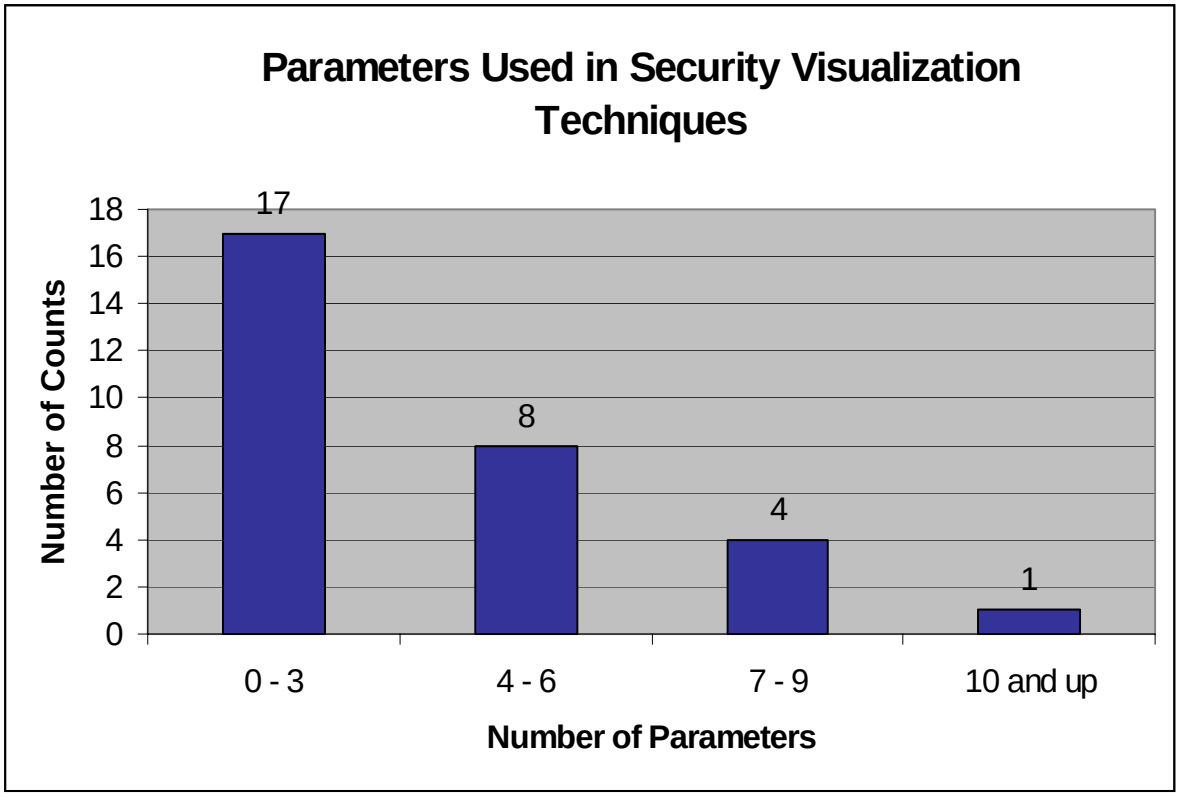

Chart 9. Statistics of the number of parameters used as data in each of the security visualization techniques.

2.4.2 MULTIPLE LEVELS OF DETAIL

Security visualization, in almost all cases, is complex and rather complicated to understand and analyze. Because of this, having a single detail window to display information is sometimes ineffective. Visualization of data, connections, nodes, and others are usually done separately in the same technique. Different levels of details require different visualization methods. For example, a visualization of the IP addresses and their connections cannot be used to visualize port activities. There is a need to provide more than a single window to display all information.

Most of the security visualization techniques implemented more than one window to display its data. For example, PortVis, from [19], have about five to six visualization areas. In its timeline area, the port information is displayed in columns and histograms. In the hour, or main, area, the 256 x 256 grid shows the port location in terms of their X and Y coordinates, using a color map. The magnification visualization area shows the highlighted area from the main area, so that more details can be zoomed in. The port activity visualization area displays the port activities in scatterplots and graphs. A user-interface visualization area also allows the user to change several parameters and control the appearance of the main display, as well as other selections. Imagine if all of these data is displayed in one single screen. It is not possible to do so without complications and aggravation. PortVis is a security visualization technique that exemplifies different multiple levels of details effectively.

VisFlowConnect, from [27], is another good example that uses multiple levels of details. The technique contains four views: the global view, domain view, internal view, and host statistics view. Each view provides different information in their display. In the global view, traffic information, or network flows, are represented in multitude of lines. Line colors and thickness represents the type of traffic and the amount of traffic. The visualization displays the orders of the IP addresses, where the lowest number IP addresses are on top. In the domain view, most of the properties are obtained from the global view, except for the zoom in and out capabilities. This provides information more closely. The internal network view provides information like the global view with some additions. The view allows the IP addresses to be selected, and information is displayed. Lastly, the host statistics view provides a table-like detail data about the traffic for specific machines, such as the number of bytes that flow in and out of a machine in a time window. This provides a more precise and accurate view of the data. So, each

view in VisFlowConnect provides different levels of details. Each visualization area provides information that would be hard to display and view in other areas, or in a combined visualization area.

According to the chart, most security visualization techniques tend to contain at least two windows in its visualization to display multiple levels of details. Having one, single window for visualization accounts for less than 14 percent of the total visualization techniques. The majority of the visualization contains two windows. This is rather surprising because most of the time, it is still insufficient to have only two windows. The prediction was that at least three to four windows are considered sufficient. However, combining the fields of three and four and more windows account closely to 50 percent of the visualization techniques, which is acceptable. It can still be noted that security visualization techniques today are still limited in the multiple levels of detail display. Having two windows to display different levels of information, though is the majority in this case, is still not sufficient enough. As more security visualization techniques continue to grow and develop, multiple windows would surely appear in their views or visualization areas.

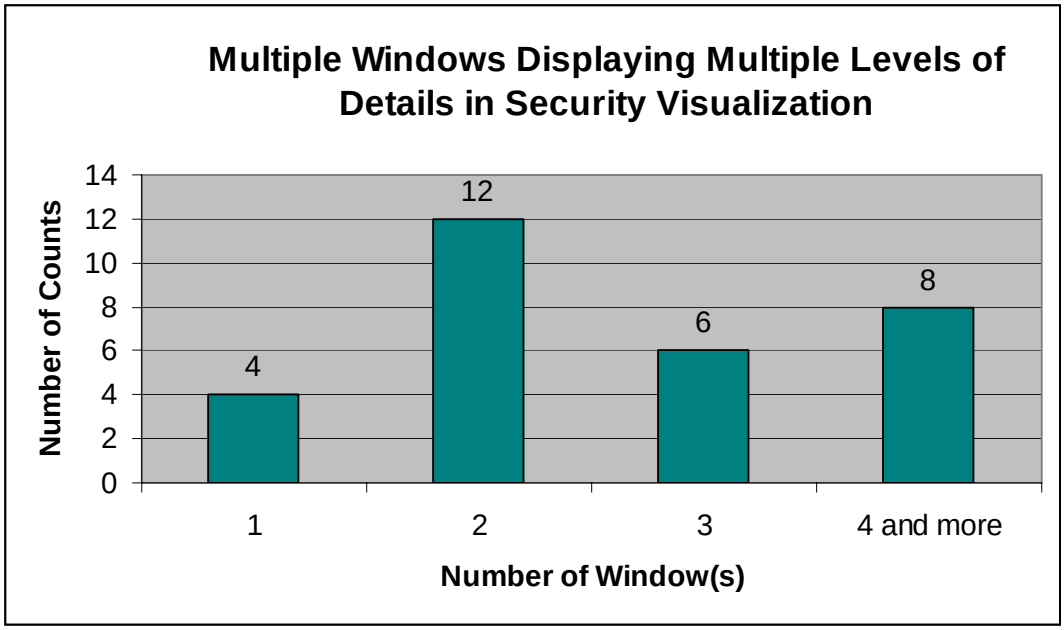

Chart 10. Statistics of the number of windows used in each of the security visualization techniques.

2.4.3 USER INTERACTION TECHNIQUES

Most security visualization techniques today are still in its developing phase when it comes to user interactions. A little more than half of all the security visualizations observed in this research fails to provide user interaction. However, when there is a user interaction in any security visualization techniques, the most dominant one tend to be the zoom in and out capabilities.

Security visualization techniques currently do not allow users to customize. They are more of a visualization technique, to display only, and not to interact. They are more towards the display program and not applications.

Except for the zooming capabilities, another more common user interaction technique tends to be the detail windows on demand. This is a window that is generally located somewhere on the visualization technique, and displays textual information, such as IP addresses, port

numbers, time and date, and number of bytes transferred. This is due to the fact that sometimes, having just graphical visualization, color maps, or plots, are not enough. Users still depend on textual information at some point. However, displaying log files is too difficult, especially when everything is in text. So, there must be some kind of balance, both texts and graphical visualizations. Another reason for having the detail windows is that security visualization techniques today might not be that developed yet, up to a point where having text information is unnecessary.

Other user interaction techniques, such as the ability to change colors in the lines to distinguish details, switching back and forth between different data to visualize, or having 3-D screens instead of two-dimensional screens so the users can manipulate the visualization, are seen in these security visualization techniques, but just in its beginning phases. User interaction techniques in security visualization continue to grow steadily, as more security visualization techniques increase in its complexities.

3. TAXONOMY

Data that was gathered and observed from the research can be divided into two main groups, network traffic data, and processed or intrusion detection systems data. Furthermore, there are two ways to further divide network traffic data. They can be classified by signature based visualization methods, anomaly based visualization methods, or both signature and anomaly based visualization methods. Another classification can be done by abstract visualization or concrete visualization.

3.1 NETWORK TRAFFIC DATA

Network traffic data pertains to raw data, which are generally the different types of data used in the security visualization techniques. Most of the security visualization techniques, as will be discussed later, use network traffic data. Data such as the IP addresses, ports, protocols, time and date, connection information, and usernames can be considered network traffic data. Such data in the form of different log files are also considered to be network traffic data. Specifically, these data are real data, not processed, filtered, or categorized initially before being used in the security visualization techniques.

For example, in [11], the security visualization technique proposed uses network traffic data, not classified or processed data. The security information that was visualized was data from a university system. The data are information contained in log files. The log files are combined into a single database. They contain information, relating to time and date, host information, and IP addresses. This security visualization technique, an anomaly based visualization method, uses glyphs to represent the hosts in the network, using the IP address as a mean to position everything in space. There is a monitored system with an IP address. Each glyph's position is determined by their IP address. The more similar the IP address of a glyph is to that of the monitored system, the closer the position of that glyph is to the monitored system. In this case, the glyphs that have their rightmost number of the IP addresses are placed on the same ring of the monitored system.

VisFlowConnect, [27], is another security visualization technique that uses network traffic data. The security information that was visualized was NetFlow records, stored in a form of log file. A NetFlow record represents a distinct network flow and certain characteristics about that flow. It records information on unidirectional end-to-end transactions between two machines. The fields from these records that were used in VisFlowConnect were: source IP

address and port number, destination IP address and port number, number of bytes transferred in the flow, number of packets transferred in the flow, protocol used, and start and end times. In this technique, a signature based visualization method, the multitude of lines represent the network connections. The three vertical parallel axes represent the external domain sender (incoming traffic), the internal host (machines on internal network), and the external domain receiver (outgoing traffic). In this technique, each point in the axis is ordered by the IP address of the machine or domain. The lowest number IP addresses are on top. The IP addresses, then, is used to categorize the positions of each machine on the network.

As mentioned earlier, network traffic data can be categorized into two classifications of visualization methods. The first classification is by dividing them into signature based, anomaly based, or both signature and anomaly based visualization methods. This other classification is by dividing them into either the category of abstract visualization or concrete visualization.

3.1.1 CLASSIFICATION BY SIGNATURE BASED, ANOMALY BASED, OR BOTH VISUALIZATION METHODS

3.1.1.1 SIGNATURE BASED VISUALIZATION METHODS

In signature based security scanning, the data streams of the requests are compared with the signatures in a list contained in the system. If there is a match between the data stream and one of the signatures in the system, it is more likely that there is an illegal access, or an intrusion.

By definition, signature based security scanning use detection models concentrating on the intrusive signal only with only a rudimentary, undeveloped model of benign processes. The problem with signature based security scanning is that only after the request is granted or

accepted can the detection test take place. At some point in time, intrusions may have already occurred. Detection can only be done afterwards.

SnortView, [18], is a good example of a security visualization technique that is classified as a signature based visualization method. The security information that was visualized in SnortView was Network-based Intrusion Detection Systems (NIDS) log files. More specifically, the information in the log files that were focused on were: time of access, type of access, source of access, destination of access, and details of the accesses. Because SnortView uses comparisons between the data streams and its signatures in its system to detect intrusions, it is a signature based visualization method.

Another example of a security visualization technique that is a signature based visualization method is exemplified in [1]. The security information that was visualized was the source and destination IP addresses and the source and destination ports. IDEVAL, a type of log file that contains intrusion and attacks such as ffbconfig and portsweep, was also used. The technique uses glyphs as its graphical representation. The signatures of the data streams obtained are tested and compared in its library, IDEVAL, making it a signature based visualization method.

3.1.1.2 ANOMALY BASED VISUALIZATION METHODS

An anomaly based security scanning, on the other hand, compares the patterns of the request to the normal pattern. Usually, this is done by an anomaly chart or graph. So, when there is an apparent shift in the graphs or chart, it can be concluded that there is an intrusion.

Anomaly based security scanning concentrate on the benign model of signal only, and ignores the malicious processes. Anomaly based security scanning rely on monitoring unusual

behaviors, and it marked by the models of automated learning. Moreover, it is a fact that anomaly based security scanning focuses on monitoring unusual behaviors of the systems, so it is capable of detecting new attacks of unusual patterns, though they may not be illegal. Because of this, there is a high rate of false alarms, or benign processes.

The security visualization technique proposed in [9] is an example of an anomaly based visualization method. The security information that was visualized was the representation of text descriptor of a RealSecure alarm log file. The log files are formatted into binary values. These binary values are formatted into ones and zeroes, where a one represents the presence of a token in an alarm (the multivariate Bernoulli event representation). The technique uses two sets of data, a notional data set, which is created and controlled by the laboratory, and a data set from an operational environment. In this technique, statistics (frequencies of the symbols) of the host alarm streams generated are observed. Typicality scores are calculated for each alarm. It is the sum of the number of times the symbols appeared in a period of time in comparison to other symbols in the network. If the alarm symbol appears often, its typicality value is high. In contrast, if the alarm symbol appears less often or unusual, its typicality value is low, and is concluded as anomalous. Typicality scores, or Tr, are displayed in color patterns, from red, yellow, and green, where each color marks different typicality scores. Red marks low typicality scores, yellow marks intermediate typicality scores, green marks high typicality scores, and gray and black marks the absence of the token. Raw data is used to calculate the typicality scores, which in turn displays the data, and the patterns generated. This technique is an anomaly based visualization method because the suspicious patterns of the data are compared to the regular patterns.

Another illustration of an anomaly based visualization method that uses network traffic data is exemplified in [29]. The security information visualized was data network scans, including source and destination ports, pairs of destination IP addresses, and packets of arrival times. The technique uses two grids to display graphical patterns of the pairs of the IP addresses. There is a detail window on top to show clusters of the nodes, where each node represents a scan. The IP addresses are used to mainly position the nodes. In the detail window, the X coordinate is the third byte of the destination IP address while Y is the fourth byte of the destination IP address. The patterns can be compared side by side in the windows. Nodes will cluster if they are from the same source, yielding patterns. This, then, is an anomaly based visualization method because the patterns are compared in order to detect any anomalies.

3.1.1.3 BOTH SIGNATURE BASED AND ANOMALY BASED VISUALIZATION METHODS

There are some cases in this classification where some security visualization techniques can be considered both signature based and anomaly based visualization methods. This can be because of a variety of reasons. Some security visualization techniques may possess these two qualities. Usually, a security visualization technique can be signature based because of its library of signatures that are used to compare with the data streams. It is also anomaly based most of the times because its output may be in the form of graphical patterns which are generally used to compare with regular patterns. In this case, the security visualization technique is considered both a signature based and an anomaly based visualization method.

NVisionIP, in [8] exemplifies both signature based and anomaly based visualization qualities. The security information visualized was NetFlow files, which is a type of log file. A

NetFlow file or record consists of records of port connections between machines over a period of time. NetFlow records provide information about network traffic flows, in sequences of packets over a period of time. NVisionIP allows the visualization of multidimensional data space. The statistics of each IP address in the network is calculated. The results are processed by a module before it is projected. The IP addresses, then, determines the position of the nodes on the screen. The subnets are located on the X-axis, and the hosts are represented on the Y-axis. This is how the network traffic data is used in NVisionIP. NVisionIP is a signature based visualization method because the signatures of the outputs are compared to the signatures in the library. It is also an anomaly based visualization method because the output patterns are compared to regular patterns in the visualization graphs.

Another example is the security visualization technique in [4]. The security information that was visualized was log files from a web server. The data used were: IP address, remote username, authenticated username, time, http request, http status code, number of bytes, referring URL, and user agent. A parallel coordinate plot was used to visualize the data. The dimensions of the plots were generated by some of the data, including date, remsys/request/url, authuser, binned status, bytes, and useragent. The parallel coordinate plot is made when each data point is projected as a line joining the components of the vector to a set of parallel coordinates. This enables the user to visualize the correlation between the variables, the patterns that were generated, and the similarities or differences among the data sets. This technique is considered an anomaly based visualization method the patterns plotted in the graph is compared to the normal pattern. It is also a signature based visualization method because there is a comparison made between the signatures of the data streams. Thus, it is both a signature based and an anomaly based visualization method.

3.1.1.4 CHART

Types of Security Scanning	Number of Counts
SIGNATURE BASED ONLY	5
BOTH	4
ANOMALY BASED ONLY	21

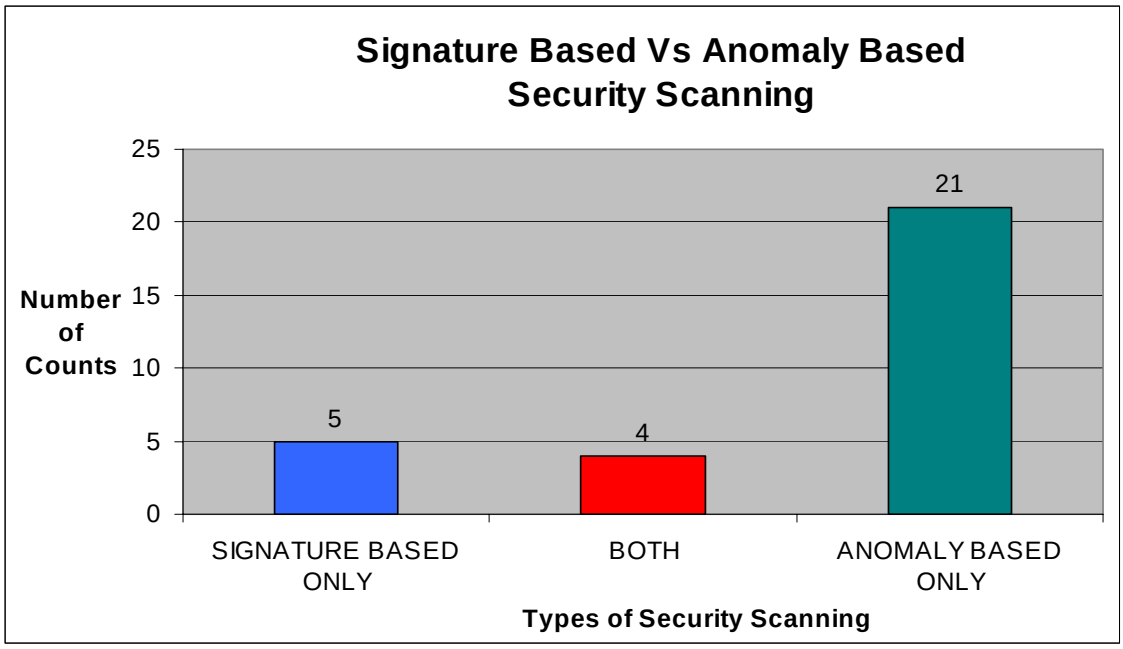

Chart 11. Classifying security visualization techniques as signature based, anomaly based, or both signature and anomaly based security scanning.

3.1.1.5 DISCUSSION

From the chart, it appears that about 70 percent of all security visualization methods are anomaly based only. This is relatively high, but not surprisingly so. Most security visualization techniques tend to use some form of graphs or pattern like visualization, such as multitude of lines, to display data. Moreover, the techniques tend to allow patterns to be compared and analyzed, allowing the user to judge and identify the intrusions themselves. Because most security visualization techniques require users to identify attacks or intrusions themselves, most of the techniques must possess some kind of pattern comparison scheme to allow users to recognize and contrast.

Signature based visualization methods only account for less than 17 percent of all security visualization techniques. Most security visualization techniques must have a way to display its data to the users to detect anomalies. This is done generally by anomaly based. Because of this, the number of security visualization techniques that are signature based only is relatively and unsurprisingly low. Furthermore, less than 14 percent of all security visualization techniques are both signature based and anomaly based, which is very normal, considering security visualization techniques today are still classified, and not to the complex point where having both signature based and anomaly based qualities is required.

3.1.2 CLASSIFICATION BY ABSTRACT OR CONCRETE VISUALIZATION

3.1.2.1 ABSTRACT VISUALIZATION

Abstract visualization pertains to visualization that bears little or no resemblance to the physical thing. As a result, more effort is needed to interpret the visualization. Abstract

visualization methods generally include techniques that require abstract views, proposed plans of visualization, and most of the time, assumptions about the data. Some security visualization techniques do not even implement any tools for its data, but just proposed solutions that can be done generally or easily. This is an example of an abstract visualization method. Moreover, some security visualization techniques yield generalized plans with an implemented tool. Without feeding real data to provide results, such techniques are also considered abstract visualization methods as well.

In [14], the technique heavily uses abstract views and proposed plans. The security information visualized was log files. Mainly, it was data pertaining to firewall logs. The log consisted of information of connections through the firewall. The records indicate connections that have been accepted, rejected, and denied. The technique focuses on studying the behavior of the system. It identifies what is normal behavior, and any deviation from that pattern is considered anomalous. Based on relay events, or an event where a host makes an outbound connection after accepting an inbound connection, the patterns were generalized. The technique simply calls for a model proposed to go about establishing a pattern. There is the formation of the object model and the information model. The object model displays major entities, such as users, processes, resources such as files and sockets, and interactions. The information model displays the properties of objects, including file permissions, transactions, and time intervals of file access by a process. Using these models, the system activities are compared with the patterns, and with a predefined confidence level set, anomalous activities can be alarmed. The technique, as mentioned, uses many assumptions, plans, and abstract visual representation. There is no tool created to visualize the data, just entity models to support the theories. This technique, then, can be considered an abstract visualization method.

Another good example of an abstract visualization is found in [13]. The security information visualized was from a one week 500, 000 records from a university principal server and other workstations. The log files used contained information relating to time and date, host information, and IP addresses. In the form of glyphs, the technique uses nodes and lines to represent hosts with their connections. The small circles in space represent remote connections while the larger circle is the principal system that is being monitored. The colors represent the connection variations. Red means failed connections, yellow represents lost NFS mount, and so forth. Each line being displayed represents a certain type of connection, such as Telnet connections in one solid lines, ftp connections in long dash lines, anonymous ftp connections in short dash lines, and so forth. The visual aspect of this security visualization technique, though meaningful, is still abstract. Hence, this technique displays an overall abstract visualization of its data and detection of attacks.

3.1.2.2 CONCRETE VISUALIZATION

Concrete visualization means that the visualization is physically easy to understand and interpret. Generally, this classification of visualization methods is easy to identify, since the data used is raw data, or network traffic data. Security visualization techniques that display data realistically in its techniques, or tools, can be considered a concrete visualization method. Anything that is, for example, in abstract forms or generalized plans of visualization, would not belong in this classification.

The security visualization technique in [10] can be considered a concrete visualization technique. The information visualized was information pertaining to TCP, UDP, IP addresses, and Ethernet header. Using a parallel coordinate plot system, the technique shows relationships

from the IP addresses to the ports. The variables in the plots included the source and destination IP addresses, the source and destination ports, and the protocol (TCP or UDP, and inbound or outbound from the home network). The combinations of the data sets yielded the plots to the following combinations, which are external IP to internal IP, external IP to internal port, external port to internal port, external IP to external port to internal port to internal IP, and external port to external IP to internal IP to internal port. The technique possesses concrete visualization, meaning that the data it displays can be physically related to and understood. Real data are used to be visualized, not abstract views.

Another example of a concrete visualization method is found in [15]. The information visualized was log files generated by a firewall. This technique contains many methods, including the spring layout algorithm, the self-organizing map algorithm, and the parallel coordinate plot system. In the spring layout method, data points are placed in random order. Each similar data points are pulled close together and dissimilar ones are pushed farther apart. This forms large clusters of data points where there are high similar activity patterns. In the self-organizing method, the vector is the coordinate of each event, as a point. The nodes are in circles, and the weight in sub-circles. The vectors are presented in layers before an output map is created. Colors, shapes, and textures are used to represent different attributes in the output map (color map). Lastly, the parallel coordinate graph displays divergence, trends, and correlation. The vertical axis is the dimension of the n-dimensional space, spaced apart at a certain constant distance. The attributes are on top and bottom of the axes. The horizontal lines connect the attribute values to present similarities and differences. Because the data in each sub-methods in this technique can be visualized physically without having any abstract relevance or assumptions, it is a concrete visualization method.

3.1.2.3 CHART

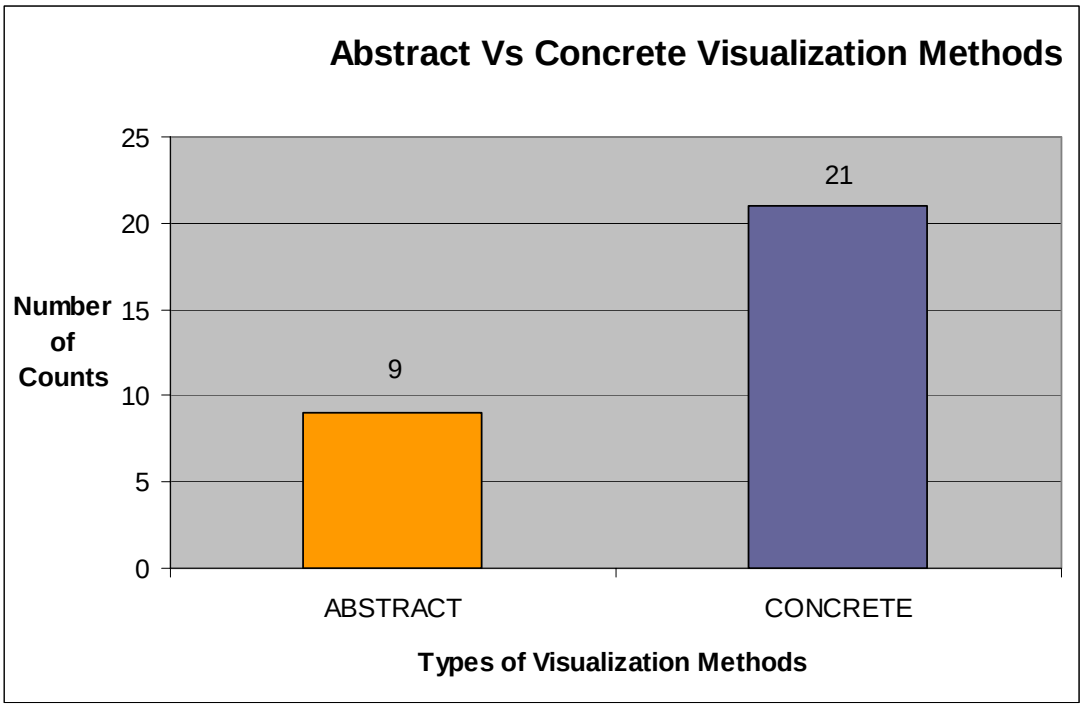

Chart 12. Classifying security visualization techniques as an abstract or a concrete visualization method.

3.1.2.4 DISCUSSION

It comes to no surprise, from the chart above, that the majority of the security visualization techniques are concrete visualization methods. Concrete visualization methods account for approximately 70 percent of the total number of security visualization techniques. The other 30 percent belong to abstract visualization.

In this area, unlike the signature based or anomaly based visualization methods classification, there is no overlapping of both abstract and concrete visualization. All of the security visualization techniques, if classified by concrete or abstract visualization, are either an abstract visualization method or a concrete visualization method.

As mentioned, concrete visualization, unsurprisingly, account for more percentage than abstract visualization. This is because though security visualization is at its developing phase, there are calls for physical security visualization techniques, things that can really be used as tools. Abstract visualization still exists for the same reason, that security visualization has not reach its peak stage. As a result, there are still plans, assumptions, and abstracts to further predict and propose new techniques. It can be predicted that the number of concrete visualization methods will increase proportionally to the decrease in the number of abstract visualization methods.

3.2 PROCESSED OR INTRUSION DETECTION SYSTEMS DATA

As mentioned earlier, security visualization techniques can be divided into two main groups. The first group refers to network traffic data, which can then be divided and sub-classified into either: 1.) signature based, anomaly based, or both signature based and anomaly based visualization methods, or 2.) abstract visualization or concrete visualization methods. The latter group, which is focused on in this section, refers to processed or intrusion detection systems data.

Processed or intrusion detection systems data are generally data that are classified. In some security visualization techniques, data are sometimes classified as benign, suspect, or intrusive. Visualization of such filtered data is based upon those classifications. In other

security visualization techniques, the data used is the output of the other intrusion detection systems. This can also be considered a security visualization technique that uses processed or intrusion detection systems data. Generally, data that are not raw data or from network traffic are considered as processed or intrusion detection systems data. Most of the time, security visualization technique that embed other methods, such as data mining, neural networks, self-organizing map algorithms, cluster algorithms, or other algorithms, tend to include processed data in its visualization. There are also a number of security visualization techniques that use data from other intrusion detection systems. These data are called IDS data.

NIVA, Network Intrusion Visualization Application, in [20], is an example of a security visualization technique that uses processed or intrusion detection systems data. The security information visualized was log files. However, these log files are from different types of intrusion detectors, such as Real Secure and Black Ice. NIVA uses data from these intrusion detectors and uses links and colors to signify attacks. The application consists of a rendering window and a graphical user interface window. The rendering window, 3-D, consists of glyphs connected by links. Each glyph's location is based upon it's IP address where the closer the glyphs, the more similar their IP addresses are. The link color represents the severity of the attacks. Yellow is moderate, while red is the most crucial. The main point here is that the data NIVA uses is not network traffic data as seen earlier. It uses output data that was fed into other intrusion detectors.

Another good example of a security visualization technique that uses processed or intrusion detection systems data is exemplified in [24]. The security information that was visualized was the Border Gateway Protocol, or BGP. The BGP is used to communicate between ASes, or the Internet. Routers use BGP to communicate in the network reachability

information. The technique consists of four main components: the real-time data collection, anomaly detection engine, data server and communication, and interactive visualization client. The real-time data collection records raw BGP data for analysis by three different ways, by BGP table and dump, BGP updates, and routing table from the router. The anomaly detection engine examines BGP updates for pre-selected prefixes. The data server and communication transfers data processed by the anomaly detection engine to the interactive visualization client. The interactive visualization client is the visualization itself. The update messages are color codes and drawn according to the timestamp. Anomalous updated messages are highlighted in gray. Using data mining, BGP data was used in this technique. The technique did not use data pertaining to network traffic or raw data. The data used was filtered and processed by data mining.

3.3 CHART

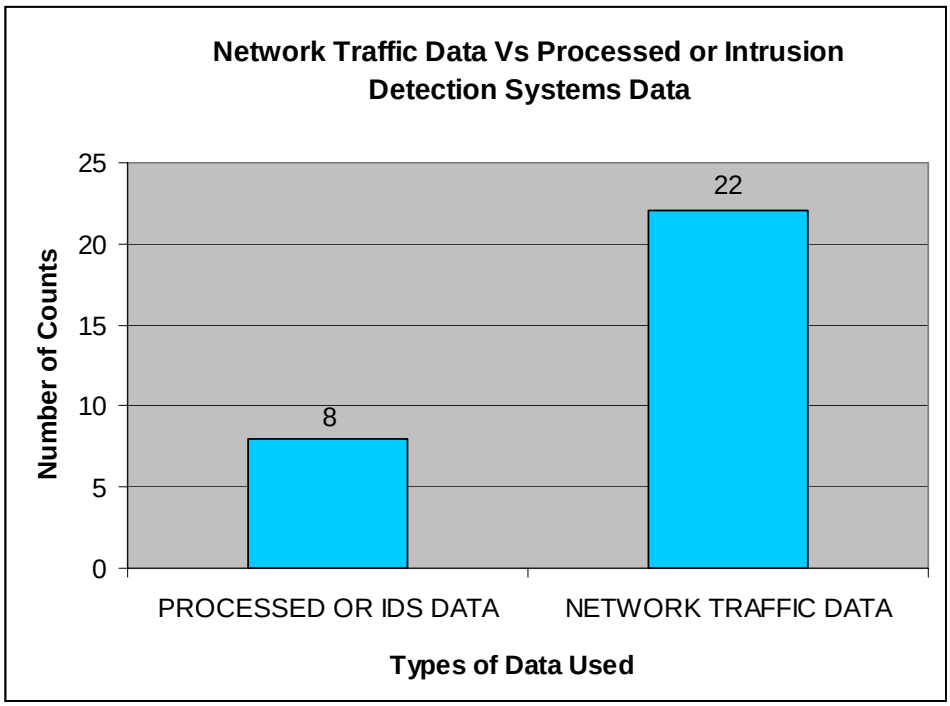

Chart 13. Statistics of the classification of data used, as network traffic data or as processed or intrusion detection systems data.

3.4 DISCUSSION

From the papers reviewed and analyzed in this research, the statistics in this area is not surprising. The majority of all security visualization techniques use network traffic data, or raw data. Network traffic data account for well over 73 percent of all security visualization technique data. Processed or intrusion detection systems data account for less than 27 percent.

Most security visualization use network traffic data because these data are realistic, physical, and more easily to be visualized. Things such as IP addresses, ports, protocol, time and date, and usernames are well displayed in these security visualization tools. Having these data provides an easier time for visualization, since real and unfiltered data are used. This allows the

user to be able to visualize all components realistically, providing the information data obtained from the sources.

Processed data or intrusion detection systems data are also used, nevertheless. However, they are not used as much as network traffic data for various reasons. The first is that security visualization techniques today are not greatly advanced. Because of this, data that are used must be limited to raw network traffic data, or other information that can be obtained from the network. Having data from neural networks, data mining, intrusion detection systems data, or even data that is classified into different groups requires more components and additions to the techniques. As more and more security visualization techniques develop, however, it can be inferred that different types of data, not necessarily from the proposed group, can be evolved. Some data that will be used in future techniques might be the output data from other security visualization techniques as well. As security visualization techniques continue to evolve and advance, so will the data that will be used.

4. ANALYSIS

4.1 CHARACTERISTICS OF CURRENT NETWORK SECURITY VISUALIZATION TECHNIQUES

4.1.1 DATA CENTERED VISUALIZATION

The network security visualization techniques observed based their visualization solely on the security information it uses. In other words, data was the key to determine how the visualization was for the techniques [30]. For example, in visualizing the IP address, it might be

more suitable to use techniques such as glyphs or color maps. When visualizing log files, for instance, most of the security visualization techniques tend to use scatterplots instead. The focus is on the data used, and the visualization of the entire technique depends on the type of data being used.

4.1.2 FOCUS ON PROBLEM DETECTION

In this area, network security visualization techniques rely on problem detection for its visualization. Visualization, then, is the solution to those problems. For example, if the problem pertains to information regarding ports, then visualization of the technique would have to focus on the area of ports. Visualization of ports would then be the solution to this problem in security visualization.

4.1.3 ANOMALY BASED PROBLEM DETECTION

The majority of the network security visualization techniques use anomaly based problem detection. In this area, mainly some sorts of patterns are used in comparison and contrast. Generally, there is a regular standardized pattern that is used to compare to the anomalies. When there is a difference between the detected pattern and the regular pattern, it can be inferred that there might be an intrusion or a possible attack.

4.1.4 2D VISUALIZATION TECHNIUQES ARE DOMINANT

Two-dimensional visualization is current the sufficient method of display for almost all of the network security visualization techniques. The reason for this is that two-dimensional display is less complicated, at this point, and requires less implementation. However, this

display can be very limited, result in data occlusion, or even data loss. It is marked that in the near future, most of the network security visualization techniques will embed the use of 3D visualization in its display.

4.1.5 MULTIPLE WINDOWS

Most of the network security visualization techniques tend to contain more than one window of display in visualizing its data. Because data visualization in these techniques is rather large and complicated, usually a single window is not sufficient to display all data. Therefore, dividing data display into different multiple windows are something that is used in many of the network security visualization techniques. This might be associated with the fact that most network security visualization techniques use two-dimensional display rather than 3D. In 3D, relatively more information can be visualized in a single screen. Two-dimensional display, as mentioned, is limited in many areas.

4.1.6 FEW PARAMETERS

Network security visualization techniques currently do not have the ability to use many parameters in its data field or in its visualization. Relatively very small numbers of parameters are used in most of the techniques. Security visualization is in fact a very large field of data. Therefore, using multiple parameters is something that would make these security visualization techniques more efficient. However, due to certain limitations and constraints in today's architecture, this issue has put a hold on visualization. But, it can be argued that as more parameters continue to be used in the network security visualization techniques, the more efficient the technique will be in securing and displaying data.

4.2 LIMITATIONS OF CURRENT NETWORK SECURITY VISUALIZATION TECHNIQUES

4.2.1 NETWORK SECURITY VISUALIZATION TECHNIQUES NOT INTUITIVE

These network security visualization techniques are generally designed for security network analysts. It is rather hard to understand and perceive such contents alone, mainly due to the nature of the field. Moreover, these network security visualization techniques, even for network security analysts, are not something one can perceive generally and immediately. It takes training and thorough analysis prior before being able to fully understand the mechanics of each technique. Since the field is currently new, most of the time, there have not been many simple explanations and researches done. Most of the researches in this area are complex and detailed.

4.2.2 LIMIT ON USER INTERACTIONS

In most of the network security visualization techniques observed, very few of them have implemented any sort of user interaction interface in its technique. The ability for users to change and modify parameters is very limited. Zooming capabilities are one of the most common user interaction techniques found in most of the network security visualization techniques. Demand on detail is the next more common user interaction technique, though not much of it has been implemented as well. In this stage, the network security visualization techniques focus mainly on displaying data and the anomalies, in hope to detect intrusions or

possible attacks. Because of that, the area of user interaction interfaces remains limited and constrained.

4.2.3 VERY LITTLE USER STUDY

In recent security visualization workshop (VizSEC 2005), less than 20 percent of the papers submitted include user studies. The percentage is even lower for the submissions in the previous years. This prediction is also seen in the papers being analyzed in this research. Less than three percent of the papers analyzed included a user study. This area, in fact, is also being neglected as well.

4.2.4 EFFECTIVENESS OF VISUALIZATION IS UNCLEAR

In nearly all papers surveyed, each paper starts by stating that visualization can help improve the performance of security experts. However, there is very little study to prove this. There is no comprehensive user study that quantifies and compares the performance difference of security problem solving using log files versus using visualization. Therefore, it is still not clear how much benefits a security expert or analyst can get from visualization. However, in theory, the visualization argues that its method of displaying data is more effective than other techniques in visualizing data and detecting intrusions.

4.2.5 TECHNIQUES NOT OPTIMIZED FOR HUMAN PERCEPTION

In most, if not all, of the network security visualization techniques, design decisions are generally not made based on human perceptions. It is rather also very difficult to conclude if this is a problem or not, since human perceptions can also have faults and mistakes. Constructively,

human perceptions are mainly more relied on. These network security visualization techniques remain at fault for not being able to be optimized for human perception.

4.3 OPEN RESEARCH PROBLEMS

4.3.1 WHAT ARE THE BEST WAYS TO MAP VARIOUS SECURITY DATA ITEMS TO VISUAL PRIMITIVES?

There is no possible formula to map security data to the visualization techniques. Each visualization techniques depend on various areas. Some data, such as the IP address, can be argued to be more efficiently visualized through the use of glyphs and scatterplots. Ports, however, can be better visualized through color maps. Each entire visualization depends on many areas, though data is the main component is many network security visualization techniques.

4.3.2 CAN WE USE THE EXISTING RESEARCH IN HUMAN VISUAL COGNITION TO BUILD A SOLID FOUNDATION FOR INFORMATION SECURITY VISUALIZATION?

Human perception is something that can be applied to the field of security visualization. The research in the area of human visual cognition is very complex, and can be used in security visualization as well. However, the problem of stability is an issue. Since patterns do not remain fixed, new patterns can generate new parameters for human perception. Also, differences in human perception, in relations to attributes such as gender, age, and quality and quantity of experience would also vary in details.

4.3.3 WHAT ARE THE EFFECTIVE VISUALIZATION TECHNIQUES FOR DIFFERENT TYPES OF NETWORK SECURITY PROBLEM DETECTION?

Effectiveness of the network security visualization techniques are mapped to the solution to intrusion detection, reduction in the number of false positives and negatives, or sounding the alarm when there is an attack. For example, in the problem of port scan attacks, the visualization of the port numbers, using color maps, can help visualize the possible attacks. Depending on the problem, the visualization is solely based in retrieving solutions.

4.3.4 HOW CAN VISUALIZATION BE USED FOR NETWORK SECURITY PROBLEM IDENTIFICATION AND PROJECTION?

Visualization can present data, as well as other parameters, on the screen. The task of monitoring becomes easier when data is not represented as texts only. Moreover, the network security visualization techniques continue to implement more understandable essence in its display. For example, having the IP addresses labeled next to each glyphs provides an easier time for network security analysts to conclude which unauthorized users are possible candidates for attacks. This help solves the problem of identification, where projection is the display of the data itself.

4.3.5 WHAT USER INTERACTIONS SHOULD BE DEVELOPED TO SUPPORT HIGH LEVEL SECURITY PROBLEM SOLVING?

As mentioned earlier, the limitations on user interactions remain a task to be developed. However, in regards to the field of network security visualization, user interactions in the

techniques should be heavily increased. Mainly, the areas of information display should be increased. Other areas can include the ability for users to manipulate and modify data, animation capabilities with respect to time, and a better way to allow users to correlate large amounts of data in different visualization methods.

4.3.6 ARE CONCRETE VISUALIZATION TECHNIQUES MORE EFFECTIVE THAN ABSTRACT VISUALIZATION TECHNIQUES? WHEN SHOULD WE USE ONE?

Concrete visualization itself provides a more detailed and physical display of the data. It can be analyzed realistically. Abstract visualization does not allow users to be able to conclude much without having to response to many assumptions. In the area of security visualization, concrete visualization is more preferred. Abstract visualization should be used to predict and infer data. Concrete visualization should be used when dealing with realistic and physical data and visualization.

4.3.7 CAN WE COME UP WITH A STANDARD USER STUDY METHODOLOGY TO EVALUATE DIFFERENT SECURITY VISUALIZATION TECHNIQUES?

Basically, this can be done depending on each user. The methodology can be based on different important areas. Mainly, the different security visualization techniques can be analyzed through its effectiveness in meeting its tasks and goals, difficulties in visualization data, the efficiency itself as a tool for security visualization, user interaction capabilities, and the ability to function in relation with other security visualization techniques.

4.3.8 HOW TO MEASURE THE LONG TERM USABILITY OF THE SECURITY VISUALIZATION SYSTEMS?

Since the security information, or data, is constantly changing, the usability of the security visualization techniques should be measured by how stable it can remain. Advances and updates in attacks can also add to the complexities of this field. So, the long term usability of each security visualization technique depends on how much it can remain sufficient and effective around the constantly changing field of security visualization.

4.3.9 WHAT IS THE FALSE POSITIVE AND FALSE NEGATIVE RATE FOR NETWORK SECURITY VISUALIZATION?

In many of the network security visualization techniques, false positives account for more than other areas. In other words, false alarms are the main problem. They are costly, take much effort and time, and inhumanely are aggravating for network security analysts. False negatives also are a problem, but do not measure up to the problem of false positives. It was mentioned in almost all papers that false alarms are one of the biggest concerns in all security visualization techniques. The rate for each one, however, continues to increase regardless.

4.3.10 HOW TO TRAIN SECURITY EXPERTS TO USE THE VISUALIZATION SYSTEM?

This depends mainly on the security visualization techniques and also on the type of the system. There is no standard way to be able to use all systems with a single training. Each security visualization technique is designed specifically to meet different tasks. It takes thorough understanding of those tasks and the technique itself to be able to use the system. Other

components, such as data and how to detect the anomalies, also account for how effective a security expert can use the technique.

4.3.11 HOW CAN WE DESIGN EFFECTIVE SECURITY VISUALIZATIONS FOR DIFFERENT GROUPS OF USERS (AVERAGE USERS, NOVICE USERS, EXPERTS)?

This area depends on many different various components. The issues of data, design, and tasks are complex in details, and there is no single way to effective design a security visualization technique. Each security visualization technique is designed specifically to meet a purpose, a problem, a task, or a difficulty. The target is in those areas rather than the users. Different users, average, novice, or experts, are mapped to security visualization techniques, and not the techniques to the users. The designs are focused on different issues, where users are relatively small in significance when compared to other components.

5. CONCLUSION

In this thesis, the author had conducted a survey of the network security visualization techniques. From this study, the author has the following findings, which are discussed in details.

Network security data has its unique characteristics. Some data is more important in security analysis than others. The author had conducted a preliminary analysis of the security data in Chapter 2 Section 2 and Chapter 3. For raw network traffic data, the author had found that the most important data include the IP addresses, ports, time and data, and connections information. For processed or intrusion detection systems date, it is still not very clear what data

is more important for security problem solving. More thorough analysis of intrusion detection systems data set is needed.

Due to the uniqueness of the network security data and network security analysis tasks, there is a need for general guidelines for mapping security data to visual primitives. The author has analyzed the existing visualization techniques and developed several general principles in Chapter 2 Section 1.

Current security visualization techniques are largely data centered. This is because we have a much better understanding of security data than tasks of security analysis, especially the high level tasks. Based on the analysis of the existing network security visualization techniques, the author had hypothesized that task-oriented security visualizations will be more effective than data centered visualizations. However, much work needs to be done to better understand the nature of the tasks of network security analysts and experts.

As a result of the analyzing the current visualization techniques, the author has developed a taxonomy for network security techniques in Chapter 3. As far as the author knows, this is the first taxonomy for security visualization techniques. This taxonomy will help researchers classify and understand various visualization techniques. This taxonomy can also help researchers compare and study similar techniques in groups.

Finally, the author has pointed out the characteristics of the existing techniques, the major limitations, and some important open research problems in Chapter 4. This information can help researchers better understand the state of the art in network security visualization and future research directions.

The field of network security visualization is evolving rapidly and new techniques are constantly being proposed. The analysis and the taxonomy shall also evolve with the advance of

the field. This research is the first attempt to conduct a comprehensive analysis of existing network security visualization techniques. The author has analyzed existing visualization techniques from different perspectives. This is the first step towards developing a set of useful guidelines and principles to help designers design effective visualization. The taxonomy will also help researchers gain insight into this research field.

REFERENCES

[1] T. Atkison, K. Pensy, C. Nocholas, D. Ebert, R. Atkison, C. Morris (2001). "Case Study: Visualization and Information Retrieval Techniques for Network Intrusion Detection." *Joint Eurographics - IEEE TCCG Symposium on Visualization (VisSym 2001),* IEEE.

[2] Axelsson, S. (2004). "Combining a Bayesian Classifier with Visualisation: Understanding the IDS." *Proceedings of the ACM CCS Workshop on Visualization and Data Mining for Computer Security,* ACM.

[[3] Axelsson, S. (2004). "Visualising the Inner Workings of a Self Learning Classifier: Improving the Usability of Intrusion Detection Systems." *Technical Report*. Department of Computing Science, Chalmers University of Technology.

[4] Axelsson, S. (2003). "Visualisation for Intrusion Detection: Hooking the Worm." *Proceedings of the 8th European Symposium on Research in Computer Security (ESORICS 2003)*. Gjovik, Norway, Springer-Verlag.

[5] Axelsson, S. (2004). "Visualising Intrusions: Watching the Webserver." *Proceedings of the 19th IFIP International Information Security Conference (SEC2004)*. Tolouse, France.

[6] R. F. Erbacher, M. Garber (2004). "Fusion and Summarization of Behavior for Intrusion Detection Visualization." *Proceedings of the IASTED International Converence On Visualization, Imaging, and Image Processing*. Marbella, Spain.

[7] R. Ball, G. Fink, C. North (2004). "Home-Centric Visualization of Network Traffic for Security Administration." *Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security* Washington, DC, USA, ACM Press.

[8] R. Bearavolu, K. Lakkaraju, W. Yurcik, H. Raje (2003). "A Visualization Tool for Situational Awareness of Tactical and Strategic Security Events on Large and Complex Computer Networks." *IEEE Military Communications Conference (MILCOM 2003)*, IEEE. 2: 850.

[9] J. Colombe, G. Stephens (2004). "Statistical Profiling and Visualization for Detection of Malicious Insider Attacks on Computer Networks." *Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security*. Washington, DC, USA, ACM Press.

[10] G. Conti, K. Abdullah (2004). "Passive Visual Fingerprinting of Network Attack Tools." *Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security*. Washington, DC, USA, ACM Press.

[11] R. F. Erbacher, K. L. Walker, D. Frincke (2002). "Intrusion and Misuse Detection in Large-Scale Systems." *IEEE Computer Graphics & Applications*. 22 (1): 38-48.

[12] R. F. Erbacher, Z. Teng, S. Pandit (2002). "Multi-Node Monitoring and Intrusion Detection." *Proceedings of the IASTED International Conference on Visualization, Imaging, and Image Processing*. Malaga, Spain, IASTED.

[13] Erbacher, R. F. (2001). "Visual Behavior Characterization for Intrusion Detection in Large Scale Systems." *Proceedings of the IASTED International Conference on Visualizaation, Imaging, and Image Processing*. Marbella, Spain. September 3-5, 2001, pp. 54-59.

[14] S. C. Fortier, L. A. Shombert (2000). "Network Profiling and Data Visualization." *Proceedings of the 2000 IEEE Workshop on Information Assurance and Security*. United States Military Academy. West Point, NY, IEEE.

[15] L. Girardin, D. Brodbeck (1998). "A Visual Approach for Monitoring Logs." *Proceedings of the 12th Systems Administration Conferences (LISA '98)*. Boston, MA, USA, USENIX Association.

[16] Goldring, T. (2004). "Scatter (and other) Plots for Visualizing User Profiling Data and Network Traffic." *Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security*. Washington, DC, USA, ACM Press.

[17] C. Jirapummin, N. Wattanapongsakorn, P. Kanthamanon (2002). "Hybrid Neural Networks for Intrusion Detection System." *Proceedings of the 2002 International Technical Conference on Circuits/Systems, Computers and Communications (ITC-SCCC 2002)*. Phuket, Thailand.

[18] H. Koike, K. Ohno (2004). "SnortView: Visualization System of Snort Logs." *Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security*. Washington, DC, USA, ACM Press.

[19] J. McPherson, K. Ma, et al. (2004). "PortVis: A Tool for Port-Based Detection of Security Events." *Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security*. Washington, DC, USA, ACM Press.

[20] K. Nyarko, T. Capers, et al. (2002). "Network Intrusion Visualization with NIVA, an Intrusion Detection Visual Analyzer with Haptic Integration." *Proceedings of the 10th*

Symposium on Haptic Interfaces for Virtual Environment & Teleoperator Systems (HAPTICS '02), IEEE Computer Society.

[21] K. Labib, R. Vemuri (2002). "NSOM: A Real-Time Network-Based Intrusion Detection System Using Self-Organizing Maps." *Networks and Security*, 2002.

[22] T. Takada, H. Koike (2002). "MieLog: A Highly Interactive Visual Log Browser Using Information Visualization and Statistical Analysis." *Proceedings of the 16th Systems Administration Conference (LISA '02)*. Philadelphia, PA, USENIX.

[23] S. T. Teoh, K. Ma, et al. (2002). "A Visual Technique for Internet Anomaly Detection." *IASTED International Conference on Computer Graphics and Imaging (CGIM '02)*, IASTED.

[24] S. T. Teoh, K. Zhang, et al. (2004). "Combining Visual and Automated Data Mining for Near-Real-Time Anomaly Detection and Analysis in BGP." *Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security*. Washington, DC, USA, ACM Press.

[25] J. Toelle, O. Niggemann, et al. (2000). "Supporting Intrusion Detection by Graph Clustering and Graph Drawing." *Third International Workshop on the Recent Advances in Intrusion Detection*. Toulouse, France, Lecture Notes in Computer Science, Springer-Verlag.

[26] J. Xin, J. E. Dickerson, J. A. Dickerson (2003). "Fuzzy Feature Extraction and Visualization for Intrusion Detection." *The 12th IEEE International Conference on Fuzzy Systems (FUZZ '03)*, IEEE.

[27] X. Yin, W. Yurcik, et al. (2004). "VisFlowConnect: NetFlow Visualizations of Link Relationships for Security Situational Awareness." *Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security*. Washington, DC, USA, ACM Press.

[28] Yoo, I. (2004). "Visualizing Windows Executable Viruses Using Self-Organizing Maps." *Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security*. Washington, DC, USA, ACM Press.

[29] C. Muelder, K. Ma, T. Bartoletti (2004). "A Visualization Methodology for Characterization of Network Scans." *Workshop on Visualization for Computer Security (VizSEC 2005).*

[30] A. Komlodi, J. R. Goodall, W. G. Lutters (2003). "An Information Visualization Framework for Intrusion Detection." *CHI '04 Extended Abstracts on Human Factors in Computing Systems*.