https://doi.org/10.48009/2_iis_2019_89-96

Issues in Information Systems
Volume 20, Issue 2, pp. 89-96, 2019

RESEARCH ON SECURITY VISUALIZATION: A SURVEY

Shuyi Liu, Vanderbilt University, Shuyi.liu@vanderbilt.edu
Wu He, Old Dominion University, WHe@odu.edu
Xin Tian, Kennesaw State University, xtian2@kennesaw.edu

ABSTRACT

With the advancement of technology, cyber-attacks and viruses are becoming more and more of prominent issues. Fortunately, nowadays there are a multitude of approaches to gain insight and deeper understanding of malware and network anomalies to fend them off. One of the most potent methods is through analyzing the network dataset for peculiar patterns and trends. As such, visualization plays a vital role in understanding and interpreting the status and security of a system. This paper aims to review and analyze the progress of security visualization over the past ten years, not only covering the tools and techniques introduced but also the applications and evaluation systems proposed.

Keywords: Security visualization, network security, information visualization, visualization techniques, visual analysis

INTRODUCTION

We live in a world where data and information are all around us. Throughout recent years, the volume of data has only been growing faster and faster. In fact, more data has been created in the past two years than in the entire previous history of the human race. It is predicted that by the year 2020, about 1.7 megabytes of new information will be generated every second for every human being on the planet (Marr, 2015). With such an overwhelming amount of information in need to be analyzed, it would be nearly impossible to discern with bare eye any meaning or insight among the vast volume of data. Fortunately, there is a solution to this problem: Visualization. We have all heard of the famous proverb "a picture is worth a thousand words." Visualization makes use of pictorial and graphical representations to interpret and analyze data, efficiently exposing the inherent underlying patterns. This powerful technique allows millions or even billions of data points to be simply summarized by just one graphical picture. There are many visualization tools and methods that are available to assist in visualizing big data. Most advanced visualization technologies are derived from the fundamental visualization techniques that many people are familiar with. These include scatter plots, radar charts, treemaps, parallel coordinates, line graphs, and composite bar charts.

Visualization is employed in a variety of fields, ranging from medicine and biology to business and economics. One field that is drastically growing in need for visualization is Cyber-security. According to Cybint News, there is a hacker attack every 39 seconds, affecting one in three Americans each year. One of the most efficient and effective ways to prevent cyber-attacks is through analyzing the network from data collected. Since cyber analysts work with a substantial amount of large network data on a daily basis, it was only natural that visualization was adopted. The purpose of security visualization is to assist cyber analysts in the task of not only perceiving patterns and trends to gain insight into security data but also providing details and specifications to enable the analysts to handle the problem. Through security visualization, networks can be thoroughly reviewed and scanned for cyber-attacks such as DDoS attacks and worm outbreak and treated to be prevented in the near future. Besides anomaly detection, security visualization also plays a vital role in Security metrics, Security monitoring, forensics, and malware analysis. This paper seeks to summarize and analyze the current progress of the field of security visualization by conducting a survey review of related papers on this matter.

Table 1. Visualization Tools/Techniques Details

Author & Year	Tool/Technique Name	Type	Data Source	Method	Application
Urbanski, 2011	Cover-VT	NETWORK ANALYSIS	GPS, IDS sensors	Geospatial map	Education
Ferebee, 2011	N/A	NETWORK ANALYSIS	Firewall log data, Google Maps API	Geospatial map	Business
Kan, 2010	NetVis	NETWORK ANALYSIS	Snort	Treemap	Administration
Jiawan, 2009	NetViewer	NETWORK ANALYSIS	WildPackets, OmniPeek	3D Coordinate System	Administration
Sarigiannidis, 2015	VisIoT	MALWARE & THREAT ANALYSIS	Firewall log data	"Core Circle"	Administration
Hao, 2015	N/A	SITUATIONAL AWARENESS	Firewall log data	Cluster tree	Administration
Kotenko, 2014	N/A	SITUATIONAL AWARENESS	Olympic Core Games System	Treemap	Administration
Novikova, 2013	N/A	MALWARE & THREAT ANALYSIS	Firewall, routers, IDS	Node link	Administration
Savola, 2011	N/A	SITUATIONAL AWARENESS	SuI (implemented with the REST interface)	Cluster tree	Administration
Harrison, 2011	N/A	NETWORK ANALYSIS	VAST 2010 Mini Challenge 2	Node link	Administration
Maple, 2010	N/A	MALWARE & THREAT ANALYSIS	Any IDS logs	Treemap, Node link	Administration
Nance, 2011	N/A	MALWARE & THREAT ANALYSIS	Individual and Business log files	Bipartite	Administration
Siadati, 2016	APT-Hunter	MALWARE & THREAT ANALYSIS	Login summaries logs	Node link	Business
Yelizarov, 2009	N/A	MALWARE & THREAT ANALYSIS	Firewall log	"3D Coordinate Histogram"	Administration
Alam, 2016	J-Viz	MALWARE & THREAT ANALYSIS	Any IDS logs	Canonical Node Link	Business
Glanfield, 2009	OverFlow	NETWORK ANALYSIS	SiLK (System for Internet-Level Knowledge)	Chord Diagram, Treemap	Business
Dang, 2015	N/A	MALWARE & THREAT ANALYSIS	Any IDS log	Radial Bipartite	Administration
Koniaris, 2013	N/A	MALWARE & THREAT ANALYSIS	Honeypot	Histogram	Business
Muallem, 2013	VGSE	SITUATIONAL AWARENESS	Maxmind, WhoIS, Google Maps API	Geospatial map	Business
Thomson, 2013	Pianola	NETWORK ANALYSIS	Any IDS log	Timeline Event Map	Administration
Landstorfer, 2014	Pixel Carpet	NETWORK ANALYSIS	SSH log	Pixel Map	Administration
Yoon, 2018	N/A	NETWORK ANALYSIS	NetInsider	Tomogram	Administration
Fu, 2017	N/A	MALWARE & THREAT ANALYSIS	Any IDS logs	RGB matrix	Administration
Papadopoulos, 2016	BGPGraph	MALWARE & THREAT ANALYSIS	BGP	Node link	Administration
Dumas, 2012	AlertWheel	MALWARE & THREAT ANALYSIS	Snort	Radial Bipartite	Administration

Section 2 reviews the key ideas that each article presents and analyzes the trends and insights discovered among papers Lastly, section 3 concludes the paper and provides suggestions for future research.

LITERATURE REVIEW AND DISCUSSION

The basis for this survey was fifty-four articles found on the IEEE Xplore and ACM databases. These fifty-four articles were selected based on adequacy after carefully reviewing several articles published in the past ten years displayed under the keywords "security visualization."

Table 2. Statistics and Classification of Articles

	# of Articles	Percentage
Type of Article		
Evaluation	9	17.31%
Survey	9	17.31%
Purpose/Application	4	7.69%
Tool/Model	30	57.69%
Type of Tool/Technique		
Network Analysis	9	36%
Malware & Threat Analysis	12	48%
Situational Awareness	4	16%
Type of Method		
Treemap	5	20%
Geospatial	3	12%
Node Link	5	20%
Bipartite	3	12%
Others	9	36%
Application		
Administrative	18	72%
Business	6	24%
Other	1	4%
Year Published		
2008	1	1.92%
2009	6	11.54%
2010	3	5.77%
2011	5	9.62%
2012	6	11.54%
2013	8	15.38%
2014	2	3.84%
2015	4	7.69%
2016	7	13.46%
2017	7	13.46%
2018	3	5.77%
Geographic Location		
United States of America	16	30.77%
China	10	19.23%
England	8	15.38%
Greece	3	5.77%
Russia	3	5.77%
Germany	2	3.84%
Korea	2	3.84%
Other	8	15.38%

For organizational purposes, the papers reviewed were classified into four categories: (1) introducing or explaining a security visualization tool/model, (2) describing methods of evaluating existing security visualization models, (3) informing potential practical applications of a specific security visualization tools or the field as a whole, and (4) summarizing and critiquing progress in the field of security visualization up until the time of the article's publication. Many articles fell into multiple categories but only the main classification was listed in Table 3.

Table 3. Record and Classification of Articles

Author & Year	Evaluation	Survey	Application	Tool
Haina, 2017			X	
Yang, 2016			X	
Sethi, 2016	X			
Safdar, 2018				X
Sethi, 2017	X			
Alshaikh, 2013	X			
Gates, 2013	X			
Karapistoli, 2012		X		
Harrison, 2012		X		
Shiravi, 2012		X		
Urbanski, 2011				X
Ferebee, 2011				X
Kan, 2010				X
Jiawan, 2009				X
Sarigiannidis, 2015				X
Hao, 2015				X
Kotenko, 2014				X
Kasture, 2014	IRRELEVANT			
Langton, 2013	X			
Novikova, 2013				X
Savola, 2011				X
Harrison, 2011				X
Maple, 2010				X
Nance, 2011				X
Siadati, 2016				X
Goodall, 2009	X			
Trinius, 2009		X		
Yelizarov, 2009				X
Jeong, 2008		X		
Alam, 2016				X
Webga, 2015	IRRELEVANT			
Glanfield, 2009				X
Garae, 2017			X	
Dang, 2015				X
Koniaris, 2013				X
Muallem, 2013				X
Yao, 2016		X		
Thomson, 2013				X
Landstorfer, 2014				X
Li, 2012		X		
Wong, 2010		X		
McKenna, 2015	X			
Read, 2009				X
Yoon, 2018				X
Fu, 2017				X
Arima, 2017	X			
Gonzalez-Granadillo, 2017			X	
Lin, 2018				X
Bi, 2017				X
Papadopoulos, 2016				X
Jackle, 2016				X
Li, 2013	X			
Dumas, 2012				X
Goodall, 2012		X		

It is very clear from Table 2 that a large majority (about 58%) of the relevant articles collected were introducing a certain visualization tool or technique. Only about 17% of the relevant articles resembled, however slightly, that of a survey review. Currently, there is an abundance of network visualization tools that each has their unique strengths and weaknesses. The problem nowadays is not being able to determine which tool is best suited for a specific task given the large volume of tools available. Introducing a new visualization tool, although is greatly commendable for

contributing to the growth of the field, will not fix this specific issue. Critical and in-depth survey reviews, on the other hand, will provide insight to allow one to determine which tool is most efficiently and effectively fit for the task. There are generally three types of visualization tools. The first type is Network Analysis tools which specifically focus on detecting possible attacks by mapping and monitoring the physical network. Another type is Malware and Threat Analysis tools which thrive in detecting and eliminating malware and threats. Lastly, Situational Awareness tools provide high-level abstract view of a system along with suggestions based on the trends and patterns detected, enabling them to be beneficial to both technical and non-technical people (Marr, 2015). These categories are not mutually exclusive; tools may fall under multiple types. Table 1 shows a beneficial to both technical and non-technical people (Marr, 2015). These categories are not mutually exclusive; tools may fall under multiple types. Table 1 shows a summary of the key visualization tools presented in the articles reviewed. Among the articles reviewed concerning visualization tools and techniques, a large majority of the tools were classified as either Network Analysis or Malware & Threat Analysis or even both whereas very few tools (only 16%) were under the category of Situational Awareness. Even though most security visualizations are designed for trained analysts or professionals, visualization generated by tools should strive to be able to be understood even by untrained people. In addition, sometimes it is necessary to have a broad abstract overview of the system to be able to instantly convey the current status. In this case, we are less concerned about an individual anomaly detection evaluation or a detailed summary of a specific network; we are more interested in having an extensive overarching view of the system for easy and broad understanding with the capability of providing helpful suggestions for convenience. With these reasons in mind, there is a pressing need for more systems to incorporate Situational Awareness visualization.

During the process of reviewing the multitude of newly designed tools and techniques, we noticed that there was a collective structure that they all seem to follow. From this understanding, we have proposed a generic pipeline presenting the process of visualizing security data, as shown in Figure 1.

The first step is to identify the problem and address questions that need to be answered. With this in mind, it will be unlikely to fall into the common pitfalls such as visualizing for the wrong reasons and visualizing mindlessly. Next step is to use a tool to monitor and obtain input security data from logs, whether if it is from network security data, organizational manage data, or personal manage data. There are various online tools available to assist with this task such as Snort, GFI LanGuard, and Microsoft Network Monitor.

At this point, you are entering the preprocessing stage, in which the objective is to filter out irrelevant information from log data to obtain the necessary information. The final data will then be stored in the database in preparation for the next step. Preprocessing can be done manually or with the help of a tool. One available tool online is Alteryx. This is where all the magic happens. With the filtered data in the database, we can apply the visualization tools to map the data to visual items. There are lots of visualization tools to get the job done, as mentioned in the literature review section. Regardless of which one you pick, it will most likely be derived from one of these basic visualization techniques: Treemaps, Scatter Plots, Radar Charts, Parallel Coordinates, Line Graphs, and/or Composite Bar Charts.

Next step is to render the graph as a unified visualization interface and apply finishing touches. This means applying scale transformations, translations, zooms, and clips to the graph to focus on the important parts. In addition, this is when we can add/adjust the color, size, and shape of the visual graph. The user should be able to interact with the graph after this step is done. There are lots of tools available online to help such as Prefuse, VRAY, and Maxwell.

Lastly, the visualization is to be interpreted and analyzed. Draw appropriate conclusions based on the underlying patterns detected. Try to see if the questions that were posed at the beginning of the process can be answered. This job is usually done by the analyst.

Another thing we noticed while reviewing the articles was that there were quite a lot of visualization techniques based on treemaps. Treemaps in this case are a structure of presenting data in the form of nested rectangles. They are very helpful for providing a quick overview of the primary behavior and hierarchy relationship within the sample. The size of the rectangle tiles typically represents the magnitude of its significance in the network. Most network security visualizations reviewed that implemented treemaps use the width of the rectangles to specify the amount of network activity for that region; the wider the rectangle the more API calls. From this, one can determine which section performs a specific operation the most frequently, which can be interpreted as the origin of malware (such as

DDoS) if the frequency of activity is abnormally high. Unfortunately, there are downsides to this method as it does not provide any sequential information, which is crucial in many cases for detecting and designing preventative treatments for anomalies. However, this does not make treemaps unviable as every visualization method has flaws of its own. For example, histograms are great for visualizing data as they scale well in terms of data quantity but unfortunately they are very limited in terms of dimensionality. On the other hand, parallel coordinates scale very well in dimensionality but are easily overwhelmed by large data streams. A visualization expert is one that is able to meticulously select the most suitable method among the vast pool available given the circumstance.

Much attention has also been directed towards RGB-coloring techniques for visualization. This is derived from the realization that color can be used to convey a variety of features and dimensionality to reduce the overall complexity of the model. For example, Fu et al. uses RGB byte value channels to represent key information such as string constants, API calls, and DLLs which directly reflect on the nature of malware. The saturation of the color is used to represent the concentration and distinct patterns of activities performed in that area. With this method, not only is it easier to distinguish between different sections within a network but also it is much easier to categorize the family of detected malware.

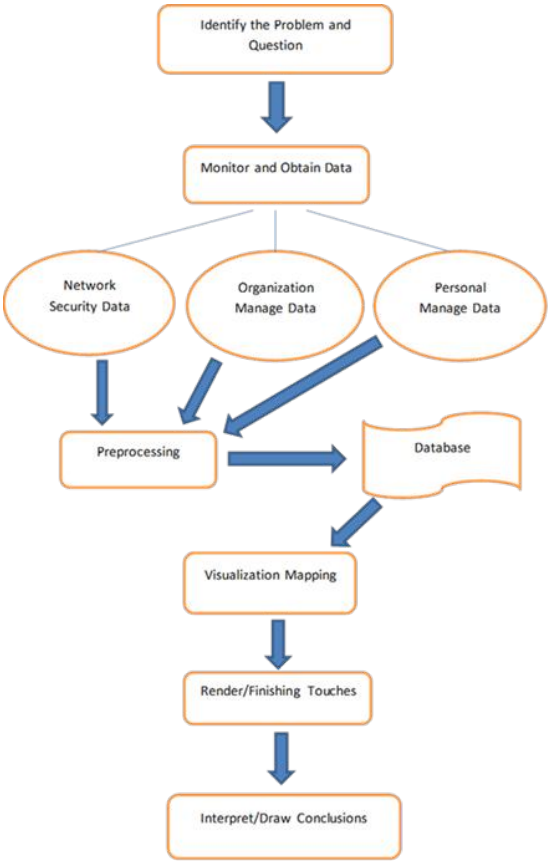

Figure 1. Proposed Generic Pipeline for Creating Security Visualization

Even though a vast majority of the tools and techniques introduced were for network administration purposes (no surprise), there seems to be a shifting focus on the system security of businesses and organizations. An example is an article by Ferebee et al. that introduces a weather-map based tool whose sole purpose is to abstract vulnerabilities up to the business service level to allow the organizations to assess how it will influence them. This trend is very reasonable since the safety of a company's networks and systems means everything; the growth and success of a company does not matter at all if it will fall at any second from a cyber-attack. Every year, companies invest a substantial amount of money on security software to protect their systems. It does not take a marketing genius to

realize that these tools are highly demanded and relevant. For these reasons, naturally the field has shifted its focus towards this area.

It is not uncommon at all to be given a network dataset that has dozens of variables. However, it is very likely that many of these variables either do not have any substantial contributions of their own or their contribution to the overall dataset is so similar to that of another variable that it is not unique. In this case, those variables should be removed to reduce the overall complexity. The entire study of Multidimensional scaling (MDS) is to determine these aforementioned variables. One of the most well-known MDS methods is Principal Component Analysis (PCA) which is implemented through orthogonal transformation. Jackle et al. introduce Temporal Multidimensional Scaling (TMDS), a novel statistics technique that excels in identifying patterns in multivariate data and reducing time-dependent dimensionality. Compared to PCA, this method is significantly more useful for security analysts since it thrives with time-dependent data, which most network data are. With two credible case study performed on the model that yielded promising results, we highly advocate for the use of this technique in the field of Cyber-security.

As for standards for evaluation of visualization, we believe that user-involvement should play a role in the evaluation process. We agree with Gates et al that only experts in the field should be considered for feedback; however this makes it difficult to obtain a large number of candidates to sample. Nevertheless, this does not mean that we should abandon the idea of including human cognitive assessment into the equation. We must remember that the fundamental purpose of visualization is to enable other humans to interpret the represented data's underlying trends and messages. Any algorithmic standard, no matter how much support it receives, will never be a better representation of human cognition than humans themselves. The ideal case that every visualization evaluation system should strive for is a perfect balance between user-involvement/feedback and methodological guidelines.

CONCLUSION AND FUTURE WORKS

As the number of security related events, including malware and viruses, generated in modern networks is on the rise, the pressing need for security visualization systems is felt more than ever. Over the past ten years, much progress has been made in the field of security visualization, including the introduction of a multitude of innovative visualization tools and the development of evaluation tools and standards for visualization. In this paper, we have delved into fifty-four recent articles related to security visualization and have critically analyzed the implications of each. Our hope is that this will shed light and motivate future researchers in this area. Potential future works that this paper calls upon are more critical survey reviews, more visualization tools that incorporate Situational Awareness, and more thorough standardized evaluation systems in the field.

ACKNOWLEDGMENTS

This research is supported in part by NSF under grant CNS-1659795. Special thanks to all the assistance and guidance from mentors and faculties from Old Dominion University. This work would not have been possible without the variety of valuable resources provided by Old Dominion University.

REFERENCES

Countries most affected by mobile malware 2018 | Statistic. (n.d.). Retrieved from https://www.statista.com/statistics/325201/countries-share-of-malicious-attacks/

Ferebee, D., Dasgupta, D., & Schmidt, M. (2011). Security Visualization: Cyber Security Storm Map and Event Correlation. 2011 IEEE Symposium on Computational Intelligence in Cyber Security (CICS). doi:10.1109/CICYBS.2011.5949412

Fu, J., Xue, J., Wang, Y., Liu, Z., & Shan, C. (2018). Malware Visualization for Fine-Grained Classification. IEEE Access, 6, 14510-14523.

Gates, C., & Engle, S. (2013). Reflecting on Visualization for Cyber Security. 2013 IEEE International Conference on Intelligence and Security Informatics. doi:10.1109/ISI.2013.6578842

Gordon, K. (n.d.). Topic: Internet usage in the UK. Retrieved from https://www.statista.com/topics/3246/internet-usage-in-the-uk/

Gordon, K. (n.d.). Topic: Internet usage in the United States. Retrieved from https://www.statista.com/topics/2237/internet-usage-in-the-united-states/

Jäckle, D., Fischer, F., Schreck, T., & Keim, D.A. (2016). Temporal MDS Plots for Analysis of Multivariate Data. IEEE Transactions on Visualization and Computer Graphics, 22, 141-150.

Marr, B. (2015, November 19). Big Data: 20 Mind-Boggling Facts Everyone Must Read. Retrieved from https://www.forbes.com/sites/bernardmarr/2015/09/30/big-data-20-mind-boggling-facts-everyone-must-read/#76f74d0517b1