Literature Review in Visual Analytics for Malware Pattern Analysis

M. Wagner, W. Aigner, A. Haberson, A. Rind

St. Poelten University of Applied Sciences, Matthias Corvinus-Strasse 15, 3100 St. Poelten, AUSTRIA
IC\M/T-Institute of Creative\Media/Technologies

ABSTRACT

Due to the increasing number of malware, monitoring of vulnerable systems is becoming increasingly more important. This applies to networks, individual computers, as well as mobile devices. For this purpose, there are various approaches and techniques to detect or to capture malicious software. To support the analysts, visualizing the data and using visual analytics (VA) methods during data exploration are beneficial approaches. There are a number of different visualization methods available which provide interaction for data exploration. We conducted a literature survey to provide an overview of the currently existing visualization and interaction techniques for malware analysis from the view of VA. All found papers were divided into 3 main categories to present common characteristics. This report shows that the scope of malware analysis in combination with VA is still not very well explored. Many of the described approaches use only few interaction techniques and leave a lot of room for future research activities.

1 INTRODUCTION

An increasing number of malicious software (malware) samples is used for espionage and attacks against infrastructure. Thus, monitoring of vulnerable systems such as networks, individual computers, and mobile devices becomes more and more important (e.g., [1], [2]). For this, effective prevention environments are needed [3]. Lee et al. [4] show that the use of visualization speeds up the malware analysis process. This is where visual analytics (VA) comes in. VA, "*the science of analytical reasoning facilitated by interactive visual interfaces*" [5, p. 4], is a comparably young research field. A major tenet of VA is that analytical reasoning is not a routine activity that can be automated completely [6]. Instead it depends heavily on analysts' initiative and domain experience which they can exercise through interactive visual interfaces. Such visual interfaces, especially Information Visualizations, are high bandwidth gateways for perception of structures, patterns, or connections hidden in the data. Interaction is "*at the heart*" of Information Visualizations [7, p. 136] and allow the analytical reasoning process to be flexible and react to unexpected insights. The benefits of VA were specified by Keim et al.: "*Visual analytics combined analysis techniques with interactive visualizations for an effective understanding, reasoning and decision making on the basis of very large and complex datasets*" [8, p. 7]. Various approaches exist to detect malicious software based on their signature, their behavior, or heuristics [9]. By using VA methods, it is possible to recognize patterns in the malware samples' code or behavior and to help the analysts in gaining a better outcome from their work. Finally, it is possible to simplify the malware classification for analysts [10].

Currently, there is no literature available that reviews the field of malicious software detection tools from the view of VA. To close this gap, we provide a systematic overview and classification of the commonly used visualization and interaction techniques for malware detection Therefore, we conducted an extensive literature research and characterized all papers found along the aspects of supported interaction technique, dimensionality of the representation space, and visualization technique. This categorization schema provides a good overview of the currently used VA techniques for malware analysis and their deficits in regard of support for interaction.

2 METHOD

To get a good overview of visualization techniques for malware detection in the field of IT-security, we used different search engines (Google Scholar, Microsoft Academic Search, IEEE Xplore, ACM digital library). The searching activities were performed in the period from March 15, 2014 to April 6, 2014. First, the results of the literature research (papers) were divided in several categories to get an overview of the covered areas:

- **Survey:** Includes papers which give a good overview of malware detection techniques and/or visualization techniques. These papers will be used for the related work.
- **Malware detection on local hosts:** Papers which operates on dedicated hosts as their main topic and deal with visualizations and VA to detect malware on local hosts.
- **Binary and gray-scale images for malware detection:** Papers which describe the detection and the classification of malicious software based on image processing methods.
- **Network and intrusion detection:** Depends on papers in relation to network scanning, traffic analysis and pattern recognition in networks for malware detection and visualization.

To gain a better outcome of the literature research, the second stage was to search for authors of the current best matching papers in combination with the best matching keywords of the previous research. Additionally, we visited the homepages of the authors in this step. Thus, it was possible to find more than 130 different scientific papers. In order to sort out inappropriate papers, we read all the abstracts and conclusion of found papers, to find out if they really fit into our topic. This way, we reduced the findings to 25 papers. In the third stage, the selection criteria was that all papers have to be from well-known conferences (e.g., VizSec, VAST, TVCG) and publishers (e.g., ACM, IEEE, Springer). In addition to the results of the search engines, we visited the homepage of the VizSec (Visualization for Cyber Security) conference (http://www.vizsec.org/) to search for papers which were not found earlier. In this stage, we found two other papers which matched with our criteria, so we integrated them into our predefined categories.

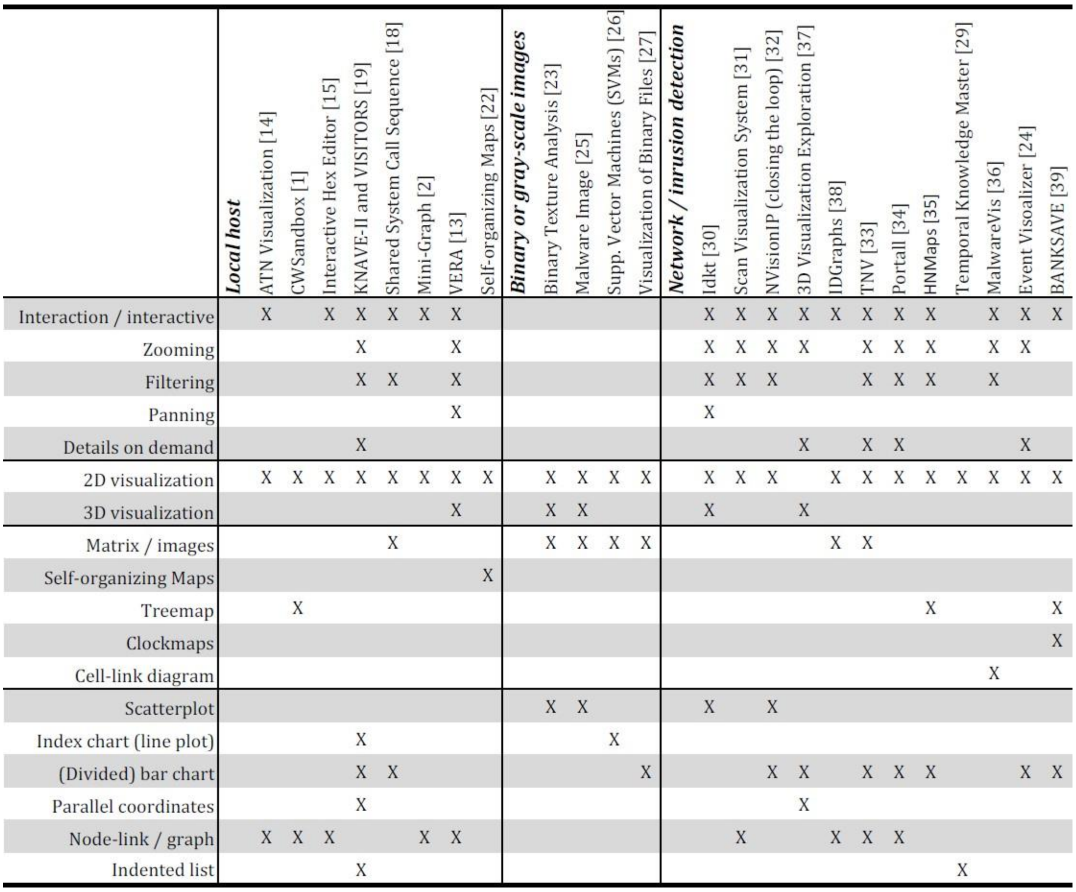
3 RELATED WORK

Previous works explores this area from different points of view. Lee et al. [4] concluded that it is necessary to use visualization for malicious software detection. They propose the use of visualization to recognize and extract unseen malware patterns. For the classification of malware the authors use a technique to extract the properties of malicious software provided by anti-virus companies. For the visualization of the collected data, Lee et al. use a dispersal pattern chart, which is a context-specific variant of scatterplots. The result of the performed test shows that it is possible to detect malicious software rapidly. Espionage and attack activities for example, can be detected by the use of signatures whose patterns are stored in a database. But this approach only works for well-known malware, while unknown malware will be overlooked [3]. Dornhackl et al. [3] introduce an approach for malware pattern extraction in their paper. They logged the system calls of malicious software which is executed on a host and analyze them. So it is possible to extract and define malicious system call patterns. Shiravi et al. [11] present a survey of 38 network security visualization systems, which are divided into five different groups of use-cases, but only some of the presented tools support methods for interaction and data exploration. Additionally, Conti dedicated a part of his book [12] on visualization for malware detection but focused on the network level.

4 RESULTS

In this section we present the approaches found by comparing the tools of the different categories and describe their advantages and disadvantages. Figure 1 shows example views for each category. Additionally, we respond to missing functionalities in relation to VA. All the used visualization techniques, dimensionality of the representation space, and the supported interaction techniques are listed in Table 1.

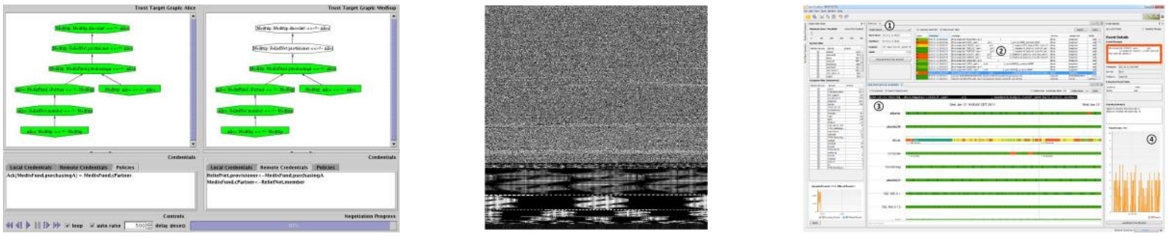
Table 1. Characterization of the different reviewed approaches.



	Local host	ATN Visualization [14]	CWSandbox [1]	Interactive Hex Editor [15]	KNAVE-II and VISITORS [19]	Shared System Call Sequence [18]	Mini-Graph [2]	VERA [13]	Self-organizing Maps [22]	***Binary or gray-scale images***	Binary Texture Analysis [23]	Malware Image [25]	Supp. Vector Machines (SVMs) [26]	Visualization of Binary Files [27]	***Network / inrusion detection***	Idkt [30]	Scan Visualization System [31]	NVisionIP (closing the loop) [32]	3D Visualization Exploration [37]	IDGraphs [38]	TNV [33]	Portall [34]	HNMaps [35]	Temporal Knowledge Master [29]	MalwareVis [36]	Event Visoalizer [24]	BANKSAVE [39]
Interaction / interactive		X		X	X	X	X	X								X	X	X	X	X	X	X	X		X	X	X
Zooming					X			X								X	X	X	X		X	X	X		X	X	
Filtering					X	X		X								X	X	X			X	X	X		X		
Panning								X								X											
Details on demand					X														X		X	X				X	
2D visualization		X	X	X	X	X	X	X	X		X	X	X	X		X	X	X		X	X	X	X	X	X	X	X
3D visualization								X			X	X				X			X								
Matrix / images						X					X	X	X	X						X	X						
Self-organizing Maps									X																		
Treemap			X																				X				X
Clockmaps																											X
Cell-link diagram																									X		
Scatterplot											X	X				X		X									
Index chart (line plot)					X								X														
(Divided) bar chart					X	X								X				X	X		X	X	X			X	X
Parallel coordinates					X														X								
Node-link / graph		X	X	X			X	X									X			X	X	X					
Indented list					X																			X			

4.1 Malware detection on local hosts

We found 8 different approaches for malware detection which operate on local hosts. All the presented tools use 2D visualization techniques. Only the *VERA* approach [13] uses 2D and 3D visualization techniques for the data representation. Most of the tools (*ATN* [14], *CWSandbox* [1], *Interactive Hex Editor* [15], *Mini-Graph* [2] and *VERA* [13]) represent the malware data by the use of node-link diagram or graph [16] visualization techniques. Additionally, *CWSandbox* also uses treemaps [17] for the data visualization. Only two tools (*Shared System Call Sequence* [18] and *VISITORS* [19]) of this Section use a kind of bar charts [20] for the visualization. *VISITORS* combined bar chart visualization techniques with index charts and parallel coordinates [21]. Additionally, *Shared System Call Sequence* use matrix [16] visualization techniques. So we can see that the most used visualization technique of that area are node-link diagrams followed by bar charts. The major disadvantage of all the presented tools is that no tool supports all the interaction techniques which are listed in Table 1. *CWSandbox* and *Self-organized Maps* [22] do not support any kind of interaction for the data exploration. All the other approaches of this section support interaction, only *VERA*, *VISITORS,* and *Shared System Call Sequences* supported some interaction techniques (see Table 1). Thus, we can see that interaction techniques are less used for the malware detection systems which operated on single hosts. This knowledge opens the possibility for further research in this area.



(1) ATN [14] (2) Binary Texture Analysis [23] (3) Event Visualizer [24]

Figure 1. Example images for the 3 different categories as listed in Table 1.

4.2 Binary and gray-scale images for malware detection

In this category, we present several approaches for malware detection which use binary or gray-scale images (*Binary Texture Analysis* [23], *Malware Images* [25], *SVMs* [26] and *Visualization of Binary Files* [27]). None of these approaches supports interaction with the data or other interaction techniques (e.g. zooming, filtering) as listed in Table 1. For the detection of the malicious software data, *Binary Texture Analysis* and *Malware Images* use scatterplots [28] (2D and 3D visualizations). Additionally, *Malware Images* also uses matrix visualization techniques [16]. The *SVMs* approach utilize index charts [16] and the *Visualization of Binary Files* use bar charts [20] for the data visualization. In this area it could be very helpful to explore the data by the use of interaction techniques to detect pattern in the generated images of malicious software. Based on the knowledge that no interaction techniques are currently used, this field provides the possibility for future research.

4.3 Network and intrusion detection

In the field of Network and intrusion detection most of the approaches use interaction techniques (see Table 1). Only the *Temporal Knowledge Master* [29] approach does not provide interaction techniques for the data analysis. Thus, 7 approaches (*IDkt* [30], *Scan Visualization System* [31], *NVisionIP* [32], *TVN* [33], *Portall* [34], *HNMaps* [35] and *MalwareVis* [36]) supported interaction, zooming and filtering for the data exploration. The *3D Visualization Exploration* [37] and the *Event Visualizer* [24] approaches support interaction and zooming. Additionally, *IDGraphs* [38] and *BANKSAVE* [39] are only described as interactive in the papers. Only *IDkt* [30] supports interaction, filtering, zooming and panning for the exploration or the malware data. In addition to the former described data exploration features, *3D Visualization Exploration*, *TVN*, *Portall,* and the *Event Visualizer* supports also details on demand. 11 out of 12 presented approaches use 2D space visualizations (see Table 1) and only one approach uses plain 3D space visualization. The most used visualization technique for malware data was the bar [20] technique, followed by node-link diagrams and graph [16] visualization techniques. Most of the approaches combined two or more visualization techniques for the data visualization (see Table 1). A very interesting data visualization, called cell-link diagram [36], was used by *MalwareVis*. Based on the evaluation of the approaches, we can see that most of the approach use many interaction techniques. Thus, we conclude that future research is possible in this area, but many other researchers work in this field currently.

5 CONCLUSION

Based on the findings of the performed literature research, we built up a categorized and structured overview of the common used visualization and interaction techniques which are used for malware detection in the field of IT-security. The most interesting area for future research is the malware detection on local hosts. In this field, there are many different approaches available but most of them do not provide many interaction techniques. Additionally, in the area of binary and gray-scale images there are no systems found which provided interaction. Thus, it is very clear that a lot of space for future research would be available in this field. In the area of Network and intrusion detection we located the most research activities. So we would focus on malware detection on local host area for future research in combination with VA.

Based on the provided results of this literature review we showed that there is a large potential available for future work in the field of malware detection in combination with VA methods for IT-security. This included creation of new prototypes, visualization techniques and systems. Based on future research it would be possible to integrate existing expert knowledge into a new malware detection system. This way, the analysts get the opportunity to analyze more data, use knowledge supported detection methods and it will be possible to get deeper into the data. By the integration of different zooming and filtering techniques, it will be possible to analyze and explore the data faster and more efficient. Finally it could be possible to simplify and speed up the malware detection and classification for the analysts.

6 ACKNOWLEDGEMENT

This work was supported by the Austrian Science Fund (FWF) via the KAVA-Time project no. P25489. Many thanks to G. Rottermanner for his feedback to our manuscript.

LITERATURVERWEISE

[1] P. Trinius, T. Holz, J. Gobel, und F. C. Freiling, „Visual analysis of malware behavior using treemaps and thread graphs“, in *6th Int. Workshop on Vis. for Cyber Security (VizSec'09)*, 2009, p. 33–38.

[2] C. L. Yee, L. L. Chuan, M. Ismail, und N. Zainal, „A static and dynamic visual debugger for malware analysis“, in *18th Asia-Pacific Conf. on Commun. (APCC)*, 2012, p. 765–769.

[3] H. Dornhackl, K. Kadletz, R. Luh, und P. Tavolato, „Malicious Behavior Patterns“, in *IEEE 8th Int. Symp. on Service Oriented System Engineering*, 2014, p. 384–389.

[4] D. Lee, I. S. Song, K. J. Kim, und J. Jeong, „A Study on Malicious Codes Pattern Analysis Using Visualization“, in *Int. Conf. on Information Science and Applications (ICISA)*, 2011, p. 1–5.

[5] J. J. Thomas und K. A. Cook, *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. National Visualization and Analytics Ctr, 2005.

[6] P. Wegner, „Why Interaction is More Powerful Than Algorithms“, *Commun. ACM*, Bd. 40, Nr. 5, p. 80–91, 1997.

[7] R. Spence, *Information Visualization: Design for Interaction*, Auflage: 2nd rev. ed. New York: Prentice Hall, 2006.

[8] D. Keim, J. Kohlhammer, G. Ellis, und F. Mansmann, Hrsg., *Mastering the information age: solving problems with visual analytics*. Goslar: Eurographics Association, 2010.

[9] Z. Bazrafshan, H. Hashemi, S. M. H. Fard, und A. Hamzeh, „A survey on heuristic malware detection techniques“, in *2013 5th Conf. on Information and Knowledge Technology (IKT)*, 2013, p. 113–120.

[10] M. Wagner, W. Aigner, A. Rind, H. Dornhackl, K. Kadletz, R. Luh, und P. Tavolato, „Problem Characterization and Abstraction for Visual Analytics in Behavior-based Malware Pattern Analysis“, in *Proceedings of the 11th Workshop on Vis. for Cyber Security (VizSec'14)*, 2014, p. 9–16.

[11] H. Shiravi, A. Shiravi, und A. A. Ghorbani, „A Survey of Visualization Systems for Network Security“, *IEEE Trans. Vis. Comput. Graph.*, Bd. 18, Nr. 8, p. 1313–1329, 2012.

[12] G. Conti, Security data visualization: graphical techniques for network analysis. San Francisco: No Starch Press, 2007.

[13] D. A. Quist und L. M. Liebrock, „Visualizing compiled executables for malware analysis“, in *6th Int. Workshop on Vis. for Cyber Security (VizSec'09)*, 2009, p. 27–32.

[14] D. Yao, M. Shin, R. Tamassia, und W. H. Winsborough, „Visualization of automated trust negotiation“, in *IEEE Workshop on Vis. for Computer Security (VizSec'05)*, 2005, p. 65–74.

[15] J. Donahue, A. Paturi, und S. Mukkamala, „Visualization techniques for efficient malware detection“, in *IEEE Int. Conf. on Intelligence and Security Informatics (ISI)*, 2013, p. 289–291.

[16] J. Heer, M. Bostock, und V. Ogievetsky, „A tour through the visualization zoo“, *Commun. ACM*, Bd. 53, Nr. 6, p. 59, 2010.

[17] B. Shneiderman, „Tree Visualization with Tree-maps: 2-d Space-filling Approach“, *ACM Trans Graph*, Bd. 11, Nr. 1, S. 92–99, 1992.

[18] J. Saxe, D. Mentis, und C. Greamo, „Visualization of Shared System Call Sequence Relationships in Large Malware Corpora“, in *Proceedings of the 9th Int. Symp. on Vis. for Cyber Security (VizSec'12)*, 2012, p. 33–40.

[19] A. Shabtai, D. Klimov, Y. Shahar, und Y. Elovici, „An Intelligent, Interactive Tool for Exploration and Visualization of Time-oriented Security Data“, in *Proceedings of the 3rd Int. Workshop on Vis. for Computer Security (VizSec'06)*, 2006, p. 15–22.

[20] W. S. Cleveland und R. McGill, „Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods“, *J. Am. Stat. Assoc.*, Bd. 79, Nr. 387, p. 531–554, 1984.

[21] A. Inselberg und B. Dimsdale, „Parallel Coordinates“, in *Human-Machine Interactive Systems*, A. Klinger, Hrsg. Springer US, 1991, p. 199–233.

[22] I. Yoo, „Visualizing Windows Executable Viruses Using Self-organizing Maps“, in Proceedings of the 2004 ACM Workshop on Vis. and Data Mining for Computer Security (VizSec'04), 2004, p. 82–89.

[23] L. Nataraj, V. Yegneswaran, P. Porras, und J. Zhang, „A Comparative Assessment of Malware Classification Using Binary Texture Analysis and Dynamic Analysis“, in *Proceedings of the 4th ACM Workshop on Security and Artificial Intelligence*, 2011, S. 21–30.

[24] F. Fischer, F. Mansmann, und D. A. Keim, „Real-time Visual Analytics for Event Data Streams“, in *Proceedings of the 27th Annual ACM Symp. on Applied Computing*, 2012, p. 801–806.

[25] L. Nataraj, S. Karthikeyan, G. Jacob, und B. S. Manjunath, „Malware Images: Visualization and Automatic Classification“, in *Proceedings of the 8th Int. Symp. on Vis. for Cyber Security (VizSec'11)*, 2011, S. 4:1–4:7.

[26] K. Kancherla und S. Mukkamala, „Image visualization based malware detection“, in *2013 IEEE Symp. on Computational Intelligence in Cyber Security (CICS)*, 2013, p. 40–44.

[27] K. Han, J. H. Lim, und E. G. Im, „Malware Analysis Method Using Visualization of Binary Files“, in *Proceedings of the 2013 Research in Adaptive and Convergent Systems*, 2013, S. 317–321.

[28] N. Elmqvist, P. Dragicevic, und J. Fekete, „Rolling the Dice: Multidimensional Visual Exploration using Scatterplot Matrix Navigation“, *IEEE Trans. Vis. Comput. Graph.*, Bd. 14, Nr. 6, p. 1539–1148, 2008.

[29] A. Shabtai, M. Atlas, Y. Shahar, und Y. Elovici, „Evaluation of a Temporal-abstraction Knowledge Acquisition Tool in the Network Security Domain“, in *Proceedings of the 4th Int. Conf. on Knowledge Capture*, 2007, p. 7–14.

[30] A. Komlodi, P. Rheingans, U. Ayachit, J. R. Goodall, und A. Joshi, „A user-centered look at glyph-based security visualization“, in *IEEE Workshop on Vis. for Computer Security (VizSec'05)*, 2005, p. 21–28.

[31] C. Muelder, K.-L. Ma, und T. Bartoletti, „A visualization methodology for characterization of network scans“, in *IEEE Workshop on Vis. for Computer Security (VizSec'05)*, 2005, p. 29–38.

[32] K. Lakkaraju, R. Bearavolu, A. Slagell, W. Yurcik, und S. North, „Closing-the-loop in NVisionIP: integrating discovery and search in security visualizations“, in *IEEE Workshop on Vis. for Computer Security (VizSec'05)*, 2005, p. 75–82.

[33] J. R. Goodall, W. G. Lutters, P. Rheingans, und A. Komlodi, „Preserving the big picture: visual network traffic analysis with TNV“, in *IEEE Workshop on Vis. for Computer Security (VizSec'05)*, 2005, p. 47–54.

[34] G. A. Fink, P. Muessig, und C. North, „Visual correlation of host processes and network traffic“, in *IEEE Workshop on Vis. for Computer Security (VizSec'05)*, 2005, p. 11–19.

[35] F. Mansmann, D. A. Keim, S. C. North, B. Rexroad, und D. Sheleheda, „Visual Analysis of Network Traffic for Resource Planning, Interactive Monitoring, and Interpretation of Security Threats“, *IEEE Trans. Vis. Comput. Graph.*, Bd. 13, Nr. 6, p. 1105–1112, 2007.

[36] W. Zhuo und Y. Nadjin, „MalwareVis: Entity-based Visualization of Malware Network Traces“, in *Proceedings of the 9th Int. Symp. on Vis. for Cyber Security (VizSec'12)*, 2012, p. 41–47.

[37] A. Oline und D. Reiners, „Exploring three-dimensional Vis. for intrusion detection“, in *IEEE Workshop on Vis. for Computer Security (VizSec'05)*, 2005, p. 113–120.

[38] P. Ren, Y. Gao, Z. Li, Y. Chen, und B. Watson, „IDGraphs: intrusion detection and analysis using histographs“, in *IEEE Workshop on Vis. for Computer Security (VizSec'05)*, 2005, p. 39–46.

[39] F. Fischer, J. Fuchs, F. Mansmann, und D. A. Keim, „BANKSAFE: A visual situational awareness tool for large-scale computer networks: VAST 2012 challenge award: Outstanding comprehensive submission, including multiple vizes“, in *IEEE Conf. on Visual Analytics Science and Technology (VAST)*, 2012, p. 257–258.