Information Visualization Applied to Computer Network Security
A Case Study of a Wireless Network of a University

Luiz F. de Camargo\textsuperscript{1(✉)}, Alessandro Moraes\textsuperscript{1}, Diego R. C. Dias\textsuperscript{2},
and José R. F. Brega\textsuperscript{1}

\textsuperscript{1} Universidade Estadual Paulista, São Paulo, Brazil
camargo.luizfelipe@gmail.com, sanfatec@gmail.com, remo@fc.unesp.br

\textsuperscript{2} Universidade Federal de São João del-Rei, São João del-Rei, Brazil
diegodias@ufs.j.edu.br

Abstract. Computer networks are becoming increasingly vital to the activities of organizations, and their monitoring is necessary to ensure proper functioning. The use of the human cognitive process in decision-making through information visualization (IV) is a viable option for large amounts of data, such as those generated in network monitoring. Considering the need to monitor modern computer networks and the quality gain when using visualization techniques, the objective was to conduct a review study to understand the process of building a monitoring tool using information visualization resources and, from this review, follow with a case study through a tool for visualization application in a university wireless network management. To this end, a systematic literature review was performed, and then a survey of requirements was conducted with the university network managers. With the analysis of the data from the review and the survey, a tool was specified and developed, evaluated in several units of the university. In this way, the results from the review and the requirements survey are observed in time, which allowed the development of a solution using the observed trends, validating them in the use and evaluation of the tool. The main contribution of the work is the resulting tool and its impact on the management of the university wireless network, facilitating the activities of managers.

Keywords: Information visualization · Networks · Security · Management · Network monitoring

1 Introduction

Computer Networks have been increasingly becoming a vital part of the technological structures of any organization. Through them, data is received and transmitted every moment, to the most diverse places. All this flow must be transmitted securely, guaranteeing the availability of the networks for transmission whenever necessary, with the degree of security that the information demands [12].
To monitor networks and ensure their security and availability, the computer network administrator should use software that serves as tools for monitoring their network, obtaining real-time and historical information on the status of services and equipment that make up their network. However, effective monitoring to ensure the full operation of a network by recording all incidents that have occurred generates a considerable amount of data, causing problems for the network administrator when analyzing and obtaining new information from this data. To solve this problem, one can make use of the techniques of information visualization, using the human view to help the interpretation of data [7]. As for the project presented in this work, its main objective was to elaborate the architecture and implement a web-based application for information visualization related to the wireless network through dashboards, allowing the visualization of status information of access points and controllers that make up the network in question. As a secondary objective of the proposed project, it aims to carry out a systematic review of content about the use of information visualization applied to computer network security.

Several authors have conducted studies analyzing this area, Guimaraes et al. conducted a review on network management, classifying 285 articles published between 1985 and 2013, according to different taxonomies [6]. Dasgupta et al. sought to understand the human factor in data flow analysis [1]. Very similar to the proposal presented in this paper, using the D3.js library to treat network events through a dashboard one can cite the work done by McKenna et al. [9].

2 Background

Information security is the protection of information from various threats that seek to jeopardize the continuity of a business. Through information security, we seek to minimize the risk arising from these threats, always maximizing the return on business opportunities and investments. Security has three main pillars: confidentiality, integrity, and availability. Information security, even if focused on computer networks, is a vast subject, in a simplified manner, has as it concerns the access of information transmitted over the network by an unauthorized person for reading or modification [12].

Information visualization seeks to represent data sets as images, helping in understanding and making their interpretation more efficient. The use of visualization is appropriate when there is a need to increase human cognitive capabilities instead of replacing them with computational decision-making methods. The creation of a visualization tool should be accomplished by answering three questions: why performs the task, what data is displayed in visualizations, and how to construct the expression language as a design option [10]. Colin Ware [13] justifies the use of visualization through the potentiality of the human brain to process information: “Why should we be interested in visualization? Because the human visual system is a pattern seeker of enormous power and subtlety. The eye and the visual cortex of the brain form a massively parallel processor that provides the highest-bandwidth channel in human cognitive centers”.
Stephen Few [4] defines a dashboard as follows: a dashboard is a visual display of the most information needed to achieve one or more objectives that fit entirely on a single computer screen to be monitored at a glance. The term dashboard indicates an indicator panel, such as the indicator panel of a car (speedometer and engine temperature indicator), of an aircraft (flight altitude and wind speed indicator), among other vehicles. Dashboards can be classified into three main groups, according to the target audience to be reached: operational, tactical, and strategic.

3 Systematic Literature Review

This section outlines the systematic literature review undertaken to guide the project.

3.1 Review Objectives and Research Issues

The purpose of this review was to obtain an overview of the use of information visualization techniques in computer network security and, through this, to understand the field of research of information visualization deeply applied to computer network security. After several discussions and previous research, we reached the following main research question: how to build tools that use information visualization techniques applied to the computer network security context? We expanded this central question on the following questions: What programming language is usually used to build this type of tool? What visual techniques are usually used in this type of tool? What data are used in the construction and validation of this type of tool? What is the specific objective of the tool?

3.2 Search Strategy

For the present work, we defined the following search string: ("information" OR "data" OR "analysis") AND ("visualization" OR ("security" OR "defense")) AND ("network" OR "internet") AND ("security" OR "defense") OR "availability" OR "security visualization") AND ("security" OR "defense") AND ("visualization" OR "data" OR "data visualization") AND "visual analytics".

We considered as valid studies only articles published in the English language - journals and conferences. There were no restrictions on the date of publication, considering articles published until the search and identification phase of the studies on 12/12/2018. We applied the search string in the ACM Digital Library (24 studies), IEEE Xplore Digital Library (94 studies), Science Direct (3 studies), Scopus (164 studies), and Engineering Village (205 studies) databases.
3.3 Inclusion and Exclusion Criteria

We defined the following criteria for this review:

- Inclusion - “Affinity of the study with the desired themes”; and “Presents a practical application”;
- Exclusion - “Lack of affinity of the study with the desired themes”; “It is not an article”; “Study is not completely available in the English language”; and “It deals with a tool already analyzed”.

In the present work, due to the satisfactory quantity of studies returned, we chose not to perform the quality evaluation.

3.4 Data Extraction

For the present study, the following data composed the extraction form:

- Types of network addressed;
- Visualization techniques used;
- Objectives of visualizations: the motivations were summarized into three categories: “Category 1 - Analysis of abnormal activity in large volumes of historical or long-term data”, “Category 2 - Understanding network behavior and detect anomalies”, and “Category 3 - Detect, analyze and respond to specific attacks”;
- Data source used in visualizations;
- Network layer used; There was development of the tool;
- Name of the tool; Languages used in the construction of the tool; and
- It is a review study (survey and systematic review): It seeks to identify the studies considered as secondary, i.e., previous studies that have already sought to conduct a review work on an area.

3.5 Analysis and Discussion of Results

The number of studies obtained at the end of the review according to the classification can be seen in Table 1.

Types of Computer Networks. Thus, we sought to identify the relationship of the studies with the existing types of networks, with greater focus on the set of protocols “Transmission Control Protocol/Internet Protocol” (TCP/IP) (57 studies) and its specific parts (Border Gateway Protocol (BGP) - 3 studies, Mobile, wireless and cellular - 1 study, Wireless - 3 studies, 4G LTE - 1 study, Voice over Internet Protocol (VoIP) - 1 study).
Table 1. Numbers of studies during the review phases.

| Categories                                                                 | Quantities |
|---------------------------------------------------------------------------|------------|
| Identified studies: these are the studies returned by the selected search system, either manual or electronic | 490        |
| Duplicates studies: are those studies that are present in more than one selected database, are counted only once | 219        |
| Unique studies: Identified studies - Duplicates studies                    | 271        |
| Unsorted studies: studies that objectively do not meet the inclusion criteria. They are excluded during the selection phase | 165        |
| Selected studies: studies that apparently meet the inclusion criteria are included in the selection phase | 106        |
| Excluded studies: are those that, after evaluation of the full text, do not meet the inclusion criteria, are excluded during the extraction phase | 33         |
| Secondary studies: studies that already proposed, within the theme, a review of content, either through a systematic review, a survey, among other techniques | 7          |
| Included studies: studies that, after evaluation of the full text, meet the inclusion criteria, are maintained throughout the process and are used in the extraction phase | 66         |

Network Layer. Within the use of the TCP/IP protocol set, we sought to identify with which network layers the data was related. It is possible to notice the tendency of using data from Layer 2 (51 studies), the network layer, more specifically, data in network flow format (Netflow). The other layers showed little relevance: Layer 1 was present in 7 studies, Layer 3 was present in 3 studies, and Layer 4 was present in 9 studies. However, three studies did not report the network layer used.

Purpose of Visualization. We defined the categories as follows:

- “Category 1 - Analysis of abnormal activity in large volumes of historical or long-term data” - 17 studies;
- “Category 2 - Understanding network behavior and detecting anomalies” - 31 studies; and
- “Category 3 - Detecting, analyzing and responding to specific attacks” - 18 studies.

Data Sources. A total of 56 different data sources were used for data extraction and visualization generation. In Fig. 1, all cited sources are visualized through the word cloud technique. A significant presence of Visual Analytics Science and Technology (VAST) and NetFlow data sources can be verified.
Fig. 1. Data sources used in the studies displayed using the word cloud technique.

**Visual Techniques.** We applied the most diverse visual techniques in the tools addressed in the participating studies, grouping similar techniques we can cite the presence of 83 different visual techniques used in the constructions of visualizations. Figure 2 shows a visualization using the word cloud technique to demonstrate all the techniques present in the studies.

**Development and Language.** Among the 66 studies evaluated, the vast majority (62) developed a new tool, and only 4 of them dealt with the implementation of an existing tool. The programming languages used in the development of each tool were analyzed, but 30 studies had no mention of the programming language, and seven studies reported using two combined languages during the development—a tendency of using Java language for server-side applications and Javascript for applications executed in the client-side. We verified the use of several libraries and frameworks for visual development in the analyzed studies. The most used was the D3.js library written in Javascript and cited in 6 studies. D3.js allows the creation of visualizations through the native support of Web browsers to design patterns of Hypertext Markup Language, version 5 (HTML5).

### 3.6 Review Considerations

From the study carried out, we noticed the possibility of deepening the review, applying a quality assessment of the studies, expanding the number of studies evaluated, and selecting those with greater relevance. It is noticeable that the scenario leads to the development of a tool for monitoring computer networks based on the TCP/IP protocol set, as it is a more widespread network standard in use. A tool to understand the behavior of a network and detect the presence of anomalies (attacks, excessive use of the network, unavailability intervals), also offering the functions of detection of specific attacks and analysis on historical
Fig. 2. Trends being shown through the word cloud technique.

data. The general trend is the development of a tool that uses Layer 2 data, with network flow data; however, the use of another network layer can be considered an opportunity in the research area.

4 Specification

We proposed the development of a tool for network monitoring, presenting an interface in a dashboard format, applicable to the reality of the university and its wireless network structure. We also sought to identify possible limitations in the current monitoring solution through studies on the application of information visualization applied to computer network security. In this way, this paper presents a tool for network monitoring that offers an integrated view of the various aspects that make up the situation of a wireless network, being able to explore them interactively, using techniques of information visualization.

Nowadays, we have been using the Zabbix tool to monitor our university wireless network. However, some limitations are found, mainly regarding the access to the information within the Zabbix interface, since the menus have several levels and are not intuitive, requiring a great effort to locate the desired information. Due to the complexity of the university wireless network, since it is composed of about 32 units distributed in 24 locations, it was decided to initially develop a tool using the context of only one unit and then expand the tool to the other units.

4.1 Requirements Analysis

To obtain a complete dashboard-based monitoring tool, we sought to identify, within the structure of the university, the organizational levels and map the data
present on their respective dashboards. A brief presentation of the user profiles obtained with the requirements survey:

**Strategic Level.** Target audience: Rectory, Computer Networks Group. Objectives: monitor problems, balance loads, and know the service offered.

Data: Controllers: status, associated access points, usage load, associated users, and log. Traffic: total value, the value per unit, and the value per user segment. Users: total connected, total from other institutions, total per segment, and total visitors.

**Tactical Level.** Target audience: IT directors and network administrators of each unit. Objectives: to know the service offered.

Data: Traffic: total value in the unit, and the value per building or department. Users: total in the unit, total per building or department, total coming from other institutions, and total visitors.

**Operational Level.** Target audience: Analysts and assistants in the network areas. Objectives: monitor problems, balance loads, and know the service offered.

Data: Access points: status, usage load, associated users, log, and invalid logins. Traffic: total per access point, and total per user.

4.2 Data Collection and Organization

Zabbix Application Programming Interface (API) allows, through POST-type HTML requests with JavaScript Object (JSON) content. Zabbix stores the collected data (JSON format) on its database. However, the data format is not compatible with the format used by the D3.js library. Because of that, the treatment of this data was created through a module developed in PHP language.

4.3 Visual Representations

In this section we detailed the visual techniques selected to build the tool:

Bubble or proportional area graph - used for value comparison and proportion demonstration, allow a quick view of the relative size of objects compared to others, without the need to use scales [8].

Bar graph or column graph - one of the most basic graphs considered, allows easy comparison between different categories that are represented close, with alignment on the axis that represents the value zero and with the representation of the data through its height for the columns and its length for the bars. The use of different colors for highlights or different categories can help according to the desired narrative [8].

A bar graph or stacked columns - using for comparison between data categories; however, these categories can be broken down into segments and compared as part of a whole [8].
A line or time-series graph - displays information such as a series of data points connected by straight line segments. Similar to a scatter plot, except that the measuring points are ordered (usually by their x-axis value) and connected with straight-line segments [8].

4.4 Selected Technology

D3.js is a library written in JavaScript language for generating graphics or visualizations, manipulating and bringing to life data through HTML, Scalable Vector Graphic (SVG), and Cascading Style Sheets (CSS) technologies. It has an emphasis on current Web standards, to use all the capabilities of modern Internet browsers without using proprietary standards.

5 Comparison with Other Tools

In this section, we detailed some tools used to meet the demands of the scenario analyzed, evaluating the advantages and disadvantages compared to our solution. We addressed four tools: Grafana, Kibana, Splunk, and Zabdash. There were performed a simple installation and a brief testing period to obtain more information about each tool.

5.1 Grafana

Grafana is a free tool with a commercial version for creating visualizations in dashboard format obtaining data from several data sources, including Zabbix [5]. The installation of Grafana for testing was performed on the Linux operating system, Ubuntu distribution, version 18.10, through DEB package added to the system after downloading on the page of the manufacturer according to instructions available.

Grafana is a very flexible tool for data visualization, with a plugin for integration with Zabbix, which makes it one of the best options for network monitoring data visualization. However, the graph options and the possibilities of interaction with them represent limitations in comparison with our tool. Grafana supports the following types of graphs: time-series graph with lines, bars, and points, simple status, value table, time heat maps, and alert list.

5.2 Kibana

Kibana is a free data visualization tool that makes up the Elastic Stack or ELK (Elasticsearch, Logstash, and Kibana) software package. It generates visualizations from the data collected by the Logstash tool, treating and manipulating them in the Elasticsearch tool. The projects of the three tools were united in June 2012, giving rise to the current software package [3]. We carried out the tests on the Kibana tool using a virtual machine offered by the Bitnami group with the Elastic Stack (ELK) software package already implemented.
Kibana tool proves to be robust for performing all kinds of data analysis. However, it requires the creation of a scenario prepared using the ELK set. For the need of our university, it is unusual to apply such a complicated solution since there is already a monitoring tool implemented (Zabbix). The proposal of this work is more advantageous because it works as a layer developed on Zabbix. Kibana supports the following types of graphs: area chart, heat map, horizontal bars, line chart, pie chart, vertical bars, value table, pointer, target, and simple metrics.

5.3 Splunk

Splunk is a tool that captures, indexes and correlates various data. Among them from network equipment, in real-time in a searchable repository, from which it can generate graphs, reports, alerts, panels, and views. Splunk version 3.0 was released on August 6, 2007, and that was the first version available to the public [11]. The installation of Splunk Enterprise Trial for testing was performed on the Linux operating system, Ubuntu distribution, version 18.10, through DEB package added to the system after downloading on the page of the manufacturer according to instructions available.

Splunk software proved to be a good option for data analysis, including network data, performing all phases: data collection, treatment, and visualization. However, it has the disadvantage of being a commercial tool, having a free version with certain limitations that may be unfeasible in some specific scenarios. The university should prioritize the use of free software tools, a fact that, together with the limitations presented in the free version, makes it impossible to use Splunk to meet the needs. Splunk supports the following types of graphs: line graph, area graph, column graph, bar graph, pie chart, scatter plot, bubble graph, a single value, radial pointer, fill pointer, marker pointer, grouping chart, and choropleth chart.

5.4 Zabdash

Zabdash is an extension to Zabbix that proposes the addition of a dashboard view generated from the data available in the standard Zabbix interface. It is free and opensource hosted on GitHub and SourceForge. The installation of Zabdash for testing was performed on the Linux operating system, in Ubuntu distribution, version 18.04, through the installation of a folder within the Zabbix structure previously installed through the operating system repository.

Zabdash is a good option for extensive integration with Zabbix and easy installation. However, it has as a limitation the lack of flexibility, because it already has the graphics and panels predefined, without the possibility of customization, showing itself to be an option not very viable for the university scenario. Zabdash supports the following types of graphs: time-series graph, pie chart, and bar graph.

We can conclude that the analyzed tools are options for the application of information visualization in network management and monitoring. Meanwhile,
within the needs raised in the university scenario, the specification presented in this work is more appropriate. Figure 3 shows the screens of the evaluated tools.

Our tool is more advantageous than others since it depends only on Zabbix (a monitoring solution that is already in use, free, easy to deploy, and applies the best visualizations practices). The need for development can be considered a negative point in the comparison, but the use of the D3.js library to speed up the development process and the possible customization options helped to base the decision for development.

6 Development

We defined a modular tool development, adding new functionalities on demand for each module creation. The following modules compose the tool:

- Login: to manage access to the tool;
- Selection of units: to choose the desired unit;
- Data organization: to process data from Zabbix;
- Settings: to customize and storage of settings; and
- Visualization: to create and display graphics.

Figure 4 depicts the modules that make up the application. Figures 5 and 6 depict some of the design decisions made in the development process.

7 Evaluation

We performed tests and simulations using inspection methods such as heuristic evaluation, cognitive walk through, and consistency. The IT university team supported that task.
A second IT team unit performed usability tests, mainly evaluating the functioning of the modules that manage usage in multiple units, collecting the opinion of users involved in the tests, to improve the system for the next stage of testing [2]. We considered that test as a target audience sample of the system.

7.1 Questionnaire

We created a questionnaire based on the QUIS questionnaire, created in 1987 by a multidisciplinary group of studies of the University of Maryland Human-Computer Interaction Lab. That is suitable for interface evaluation. The original evaluation model takes into account nine factors. However, we decided to simplify some questions and omit sections that did not apply to the tool context. The questions that made up the QUIS questionnaire had numerical answers ranging
Fig. 6. Design decisions - Lightbox with a time series graphic; Detail E showing highlight time series; and Detail F showing a button to access the equipment log.

from “1” as the most negative aspect to “9” as the most positive aspect, also counting with the option “Not applicable”.

We created a tutorial teaching how to use the tool. This tutorial is available on the web pages in an integrated manner with the tool.

The initial audience to apply the questionnaire was defined in 12 university units, selecting units with different profiles. The evaluation request was sent to the network manager in each unit. The feedback was obtained from 10 users, belonging to 8 units distributed in 7 cities, performing the requested tests, and answering the questionnaire. There was difficulty in the adherence of professionals to use and evaluate the tool, even after several contacts, through different means of communication (calls, e-mail messages, and instant messages).

7.2 Questions and Answers

The questionnaire was composed of an identification step, and eight parts focused on different aspects of the interface. The online survey platform Google Forms was used to apply the questionnaire.

We requested the following information to identify the user: e-mail (through login), name, age, and gender. Part 1 of the questionnaire sought to identify how long the user used the tool. Part 2 sought to identify the areas of activity of the users.

Figure 7 depicts an overview of the other parts of the questionnaire.
7.3 Results

We concluded that the tool pleased in several aspects of the users. The users’ sample was of male professionals aged between 30 and 54 years, infrastructure professionals (wireless network, wired network, data centers, and VoIP telephony). They used the tool for one hour per week.

The positive points highlighted by the users were: ease of use, the learning process, the design of the screens, highlights presented, terminology, messages
generated by the tool, simplicity in the control options, and absence of failures and errors.

As negative points, we can mention the lack of flexibility of the tool, slowness in some situations, lack of instructions for error correction and feedback from the computer during operations.

Through the process of evaluation of the tool, we can say that the users approved the tool in most of the aspects evaluated. However, there are still improvements to be made. To improve the evaluation process, allowing a more complete and statistically reliable evaluation, it would be necessary for more users to test the tool, including the participation of professionals external to the university.

8 Conclusion

The main result of this work was the tool developed. The secondary objective was to perform a systematic review, which supported the relevance of the work.

Our tool deals with a specific network structure due to the demand presented by the Computer Network Group. It was built using light and flexible programming languages, focused on the web, using its library for generating graphics, facilitating and accelerating the professionals’ duties. The tool allows real-time monitoring of the network situation and also an exploration of historical data, to search for anomalies.

8.1 Relevance

This project was relevant mainly because it dealt with the concrete application of several technologies and areas of Computing. The tool obtained at the end of the project aims to be used in all university units, improving the supply of information to network administrators.

The comparison of the tool developed with the other options available in the market added relevance to the work, since that showed the positive and negative points of all the solutions, basing the decision to create a new tool customized according to professionals’ needs.

The process of interactive and incremental development combined with the evaluation of the tool by a sample of future users, with positive results, reinforces the relevance to the university, with mostly positive feedback.

8.2 Limitations

We can cite flexibility as a limitation of the tool. That was developed for a specific scenario, not allowing its use in another context (network infrastructure). The problem of flexibility was also noted during the evaluation process carried out by users since the views, and specific parameters that make them up are predefined and do not allow significant customizations.
8.3 Final Considerations

In this work, we presented the whole research process about our tool, its implementation proposal in a university, specification, development process, and the evaluation of users from different university units. The sequence employed in the development of this work allowed the desired results to be achieved, with the final effect of providing the university with a tool for monitoring the wireless network.

References

1. Dasgupta, A., Arendt, D.L., Franklin, L.R., Wong, P.C., Cook, K.A.: Human factors in streaming data analysis: challenges and opportunities for information visualization. In: Computer Graphics Forum, September 2017. https://doi.org/10.1111/cgf.13264
2. Dias, C.C.: Usabilidade na web: criando portais mais acessíveis. Alta Books (2007)
3. Elastic.co: Kibana User Guide [4.6] — Elastic (2018). https://www.elastic.co/guide/en/kibana/current/index.html
4. Few, S.: Information Dashboard Design. O’Reilly, Sebastopol (2006)
5. Grafana Labs: Grafana documentation — Grafana Documentation (2018). http://docs.grafana.org/
6. Guimaraes, V.T., Freitas, C.M.D.S., Sadre, R., Tarouco, L.M.R., Granville, L.Z.: A survey on information visualization for network and service management. IEEE Commun. Surv. Tutor. 18(1), 285–323 (2016). https://doi.org/10.1109/COMST.2015.2450538
7. Jacobs, K.: Data-Driven Security: Analysis, Visualization and Dashboards, 1st edn. Wiley, Hoboken (2014). https://doi.org/10.1017/CBO9781107415324.004
8. Kirk, A.: Data Visualization: A Successful Design Process: A Structured Design Approach to Equip You With the Knowledge of How to Successfully Accomplish any Data Visualization Challenge Efficiently and Effectively. Packt Publishing, Birmingham (2012)
9. McKenna, S., Staheli, D., Fulcher, C., Meyer, M.: BubbleNet: a cyber security dashboard for visualizing patterns. In: Proceedings of the Eurographics/IEEE VGTC Conference on Visualization, EuroVis 2016, pp. 281–290. Eurographics Association, Goslar (2016). https://doi.org/10.1111/cgf.12904
10. Munzner, T., Maguire, E.: Visualization Analysis & Design. AK Peters, Natick (2015). https://doi.org/10.1002/9781119978176
11. Splunk: Authentication - Splunk Documentation (2018). http://docs.splunk.com/Documentation/CIM/4.11.0/User/Authentication
12. Tanenbaum, A.S., Wetherall, D.J.: Computer Networks, vol. 52. Pearson Prentice Hall, Upper Saddle River (2011). https://doi.org/10.1016/j.comnet.2008.04.002
13. Ware, C.: Information Visualization: Perception for Design (Interactive Technologies). Morgan Kaufmann, Burlington (2004)