2010 12th International Conference on Computer Modelling and Simulation

Case Study: Visualization Methodology For Analysing Network Data

Doris Wong Hooi Ten and Sureswaran Ramadass

National Advanced IPv6 Centre (NAv6)
Universiti Sains Malaysia
Penang, Malaysia
{doris, sures}@nav6.org

Abstract— **Existing monitoring methods have been used for network traffic and network malicious such as using a network monitoring system or network intrusion system. These systems mainly focus to capture traffic and anomaly data. After the capturing process, the data is visualized without doing any analysis on these captured network data. Besides, there are a number of tools with visualization techniques (e.g. Scatterplot, bar graph) that can be applied in the network monitoring system. The one-dimensional (1D), two-dimensional (2D), three-dimensional (3D) and high multidimensional data can be presented by using relevant visualization techniques. Most of the visualization techniques exist in generating the captured network data into informative graphical 2D or 3D view. However, there are no tools or systems that are able to identify as well as analyses the types of given network data and presents it into the robust and meaningful illustration based on user requirements. This paper explored a new algorithm that would be used to solve the constraints which have been mentioned above. We also presented how the algorithm can facilitate the current visualization methodology. By the end of this, our proposed methodology should be able to analyse different types of network data automatically and effectively.**

Keywords- network traffic, network malicious, information visualization, automated visualization methodology, visualization technique.

I. INTRODUCTION

Monitoring a network is a very common activity by administrator in either network traffic detection or network intrusion detection. This process is always performed to determine what is being existed on the network [1]. For instance, network administrator would frequently monitor the organisation network for anomaly detection if they look for malicious detection from their organisation network. However, if there are any anomaly detections, the monitoring system only will show all the results from the system to the network administrator without doing any analysis. Most of the visualization techniques exist in generating the captured network data into informative graphical 2D or 3D view, but the systems are incapable of identify and analyse any types of given network data and present it in the robust and meaningful representation [2].

This research focuses on the data that captured by using the network monitoring system as input data. Different attributes of network data can be utilise to analyse the types of data that enter to the analysis and intelligence module part. However, the system must be able to obtain required attributes and apply them into the written algorithm.

In order to accomplish this, automated concepts need to be developed for ensuring the automated analysis of data will be taken place in the analysis and intelligence process. This paper was organised as follows. In section II of this paper, we presented existing visualization techniques and problems. In section III, we discussed the development of visualization methodology architecture. Besides, we discussed the advantages of a proposed algorithm in section IV. Finally, we discussed expected result of proposed visualization methodology in section V and following by a conclusion of the paper in section VI.

II. EXISTING VISUALIZATION TECHNIQUES AND PROBLEMS

There are number of techniques in the visualization area that can be applied on the network monitoring system. There are one dimensional (1D), 2D, 3D and high multidimensional data, which can be presented by using relevant visualization [3].

Visualization techniques such as the 3D Scatterplot, line graphs, survey plots and bar charts are some examples of technique for 2D data representation. By using the 2D data representation on large volume of data, variables would be increased and crowded, underlying information of these data may not be able to be presented efficiency and clearly. Thus, network administrators may not be able to conduct any measures accordingly on the network faulty.

In order to obtain 3D visualization technique, it can be achieved by expanding 2D technique. By adding z-axis (depth) in 2D data, the third dimension can be easily achieved in scatter-plots, bar charts and line graph. Animation is another visualization technique in displaying 3D data in a more interesting manner. Researchers have utilized the MATLAB to generate 3D data by using high-performance language. Computation, visualization and programming have been incorporated together and created a user friendly environment to improve a researcher working with the data presenting and displaying.

There are few techniques namely, Andrews Plots, Parallel Coordinate Plots and m-and-n plots to display the

978-0-7695-4016-0/10 $26.00 © 2010 IEEE
DOI 10.1109/UKSIM.2010.94

multidimensional data [4]. The data result will be displayed through concurrent multiple views that are linked between drawing lines and the points from another different view, which is corresponding to the same specification [4]. However, a researcher still needs to address the multidimensional data visualization problems, which are associated to the technique of extracting low-dimensional displaying. Likewise, it cannot be employed automatically for high-dimensional data if the data set size is too large.

III. VISUALIZATION METHODOLOGY ARCHITECTURE

In this paper, we proposed visualization methodology architecture for network monitoring, which focuses on second module in the architecture namely, analysis and intelligence module. Generally, visualization tends to be an iteration process. There are two agents in the data procurement module to solve the system overhead problems in data searching and data capturing process. Data mining technique such as association rules (defined policy) is applied in the analysis and intelligence module to verify the usefulness of the tool [5]. The proposed visualization methodology architecture has been shown in Figure 1.

A. Data Procurement Module

Besides the agent being responsible to collect, structure and transfer, there are two new basic agents that we proposed in this section. The agent Seeker is used to search and capture for particular data from the network monitoring system which includes data from the internet and intranet. These captured data will be stored in an intermediate database. Another proposed agent, Transferor takes responsible to transfer the captured data into the analysis and intelligence module. Data is transferred by Transferor from the intermediate database to the analysis and intelligence module for further processing and analysing. With this methodology which divides capturing and transferring activities into two different agents, we believe the system congested can be avoided and also will be additionally efficiency [6].

B. Analysis and Intelligence Module

Data mining is an essential practice and approach in the analysis and intelligence module [7]. This practice is applied in analysis and intelligence module where extracting of hidden predictive information from the large database. The proposed algorithm will be discussed more in the next section. Based on the different selected criteria, data will be restructured, rearranged and reclassified. Data from the intermediate database will be classified based on expertise level and data type. These activities aim to address a breakdown problem in the system. An update of captured data from the database can be classified from time to time continuously. In order to produce comprehensive output (input for visualization methodologies repository), process classification will base on the created algorithm to produce the output. The output will be used as an input in the visualization methodologies repository. Input will be evaluated according to the data type and later will be fixed with the appropriate selected visualization techniques. The system will make sure the visualization techniques are matching accordingly. The agent named receptionist R1 works to record and keep track of the updates from the process analysis. At the same time, receptionist R2 takes responsibility to ensure the consistency and efficiency on every single updated output and input between output from the analysis process and input for the visualization methodologies repository.

C. Data Visualization Module

Hypercube technologies, 3D sound, virtual reality and internal network space are different visualization and environment modules, which will be provided in this section. This may be helpful for the network administrator to choose base on their own understanding degree. An agent named Carrier involves in this module. Carrier plays the role to carry the selected data to the mapping process before displaying the intuitive dynamic data result. The network administrator is permitted to choose for details viewing and navigation. Single view of the data display will be created. Besides that, network administrator is allowed to interact with the data in a real-time environment. Interactivity and modification can be done between network administrator and display data. Multimedia elements such as text, video and animation also are included in this section. The integration between the comprehensive data and multimedia elements can produce an intuitive result to the network administrator. [6].

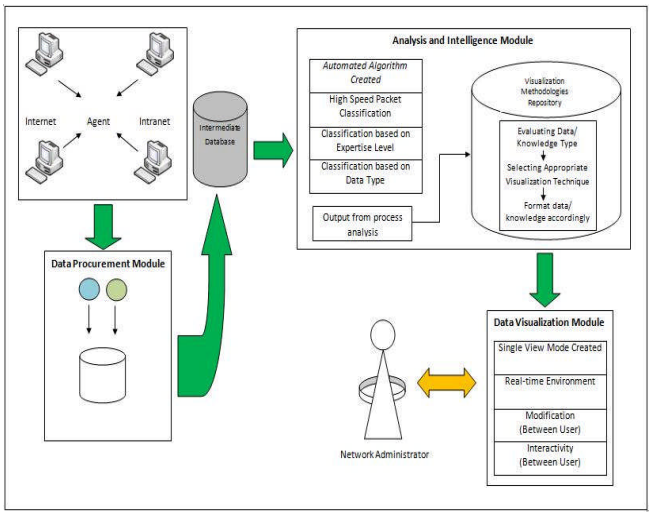

Figure 1. Proposed Visualization Methodology Architecture.

IV. ADVANTAGES OF PROPOSED ALGORITHM

We are currently focusing in the analysis and intelligence module and an automated analysis algorithm was developed to enhance the existing data visualization tools. Previously, the entire captured network data is only being displayed without any proper analysis from the system. We would like to propose a new automated algorithm which can be fully utilised in the network monitoring system. Basically, the conventional network monitoring system will monitor the system and capture all the data from the system. Then, the

data will be displayed to the network administrator as one overview result which includes all the captured data. Based on the displayed result, there are some of the results, which might not be relevant to some of the network administrators. Different levels of users require for the different kind of data.

Therefore, the proposed method will be able to analyse the captured data and run the sufficient analysis on the data. We need the process because many of the visualization tools that never analyse the data before representing the result to the different types of the network administrator. Network administrator will confuse with the complicated displayed data. In the development phase, Programming Language C or C++ will be used to develop the algorithm. Cichlid program code will be used as one of the reference codes for the algorithm development [8]. We will present our well-developed algorithm in the following working papers to show the progress of the development. This proposed algorithm will be able to analyse the data base on the predefined attributes (such as an expertise level, data type, level of details). This process will automatically present the data in the most sufficient understandable view. Overall, higher understandable of the data will be achieved. Figure 2 has shown the analysis and intelligence module part.

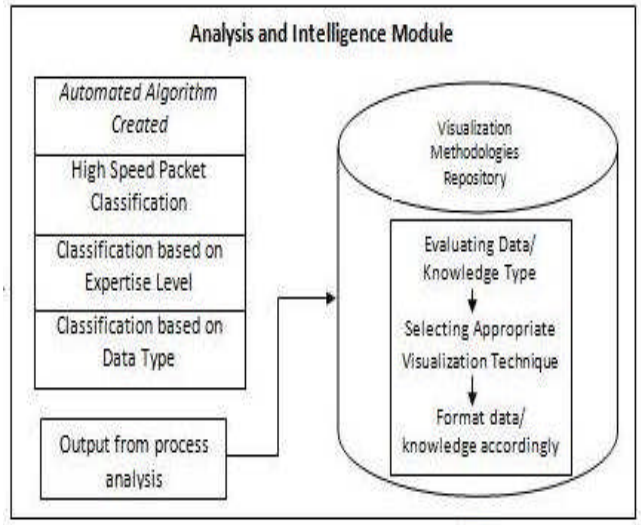

Figure 2. Analysis and Intelligence Module.

V. Expected Result Of Proposed Visualization Methodology

There are several anticipation results that we are expecting from our proposed visualization methodology. With our proposed methodology, network administrators are able to view a more understandable network traffic results which presented based on requirements from different level of network administrators. For example, the existing data visualization tools would only present all the network traffic data without classifying based on network administrators' requirements. Our proposed visualization methodology allows different types of network administrators to send their requirement to the system. Then, the system will classify based on given attributes from network administrators. Additionally, the system allows appropriate matching between visualization technique and the classified network traffic data. Finally, network administrator will be able to visualize the displayed data with the features provided such as interactive, real-time environment and modification also being allowed.

VI. Conclusion

While we are trying to develop the automated data analysis system, we also hope that the proposed algorithm will be able to be applied into different fields such as education, engineering and environmental management. It should be able to capture and analyse the data based on the predefined criteria or attributes. Based on the result from the data analysis, we will map the data to the best interpretation visualization display. We believe that our proposed algorithm will be able to present network traffic data to different level of users in different perspective view. It will improve and enhance the network traffic monitoring tasks.

Acknowledgment

Sincere thanks to National Advanced IPv6 (NAv6), Universiti Sains Malaysia (USM) colleagues for their willingness to spare and contribute their time, guidance, sharing their knowledge and support. This paper will be a part of the new research in NAv6 lab. Without their guidance this paper would not have been possible.

References

[1] F. Fan, and E. S. Biagioni, "An approach to data visualization and interpretation for sensor networks," In Sprague Jr, R. H. (Ed.) Proceedings of the Hawaii International Conference on System Sciences. Big Island, HI, 2004.

[2] C. Muelder, K. L. Ma, and T. Bartoletti, "A visualization methodology for characterization of network scans," IEEE Workshop on Visualization for Computer Security 2005, VizSEC 05, Proceedings. Minneapolis, MN, 2005.

[3] D. A. Keim, "Information visualization and visual data mining," IEEE Transactions on Visualization and Computer Graphics, 8(1), 1-8, 2002.

[4] A. Buja, J. A. Mcdonald, J. Michalak, and W. Stuetzle, "Interactive data visualization using focusing and linking," Proceedings IEEE Visualization'91, 156-163, 1991.

[5] D. S. C. Russo, P. Gros, P. Abel, D. Loisel, J-P Paris, "Using Virtual Reality for Network Management: Automated Construction of Dynamic 3D Metaphoric Worlds," 1999.

[6] D. W. H. Ten, S. Manickam, S. Ramadass, H. A. A. Bazar, "Study on Advanced Visualization Tools In Network Monitoring Platform," Third UKSim European Symposium on Computer Modeling and Simulation (EMS '09), IEEE Xplore Digital Library, 25-27 Nov 200, pp. 445 - 449, doi: 10.1109/EMS.2009.24.

[7] S. Sumathi, and S. N. Sivanandam, "Data Mining: An introduction - Case study. Studies in Computational Intelligence," 2007C. Cranor, Y. Gao, T. Johnson, V. Shkapenyuk, O. Spatcheck, "Gigascope: High Performance Network Monitoring with an SQL Interface," SIGMOD' 02, ACM, 2002.

[8] J. A. Brown, A. J. McGregor, and H-W Braun, "Network Performance Visualization: Insight through animation," Proceedings of the Passive and Active Measurement Workshop, 2002.