Indian Journal of Science and Technology, *Vol 8(15), 45371, July 2015*

*ISSN (Print) : 0974-6846*
*ISSN (Online) : 0974-5645*

# Usable, Flexible and Adaptive Network Data Visualization Design for Multiple Levels of Computer Users

**Doris Hooi-Ten Wong*** **and Chong-Meng Chee**

Universiti Tunku Abdul Rahman, Jalan Universiti, Bandar Barat, 31900 Kampar, Perak Darul Ridzuan, Malaysia; doriswong@utar.edu.my

## ABSTRACT

Numerous Network Data Visualization Tools have been developed to visually analyze and visualize multivariate data. However, these conventional network data visualization tools are typically designed with network administrators (advanced users) in mind. In this paper, we construct an adaptive visualization tool in order to solve the demands of different computer users. We adopted three supervised algorithms for our framework, namely Naive Bayes (NB), C4.5, and Support Vector Machine (SVM). Our experiment showed that the proposed framework not only managed to produce a usable interface but also has better visualization compared to existing network data visualization applications. Moreover, it is able to present comprehensive network data and is capable of adapting to user feedback during the network data visualization process. This intelligence enables the framework to adjust to the needs of different computer users when they perform network data visualization.

**Keywords:** Network Data Visualization, Statistical Analysis and Learning, Visualization

## 1. Introduction

A large amount of Network Data Visualization Tools such as Cichlid[4], Net[3], Watch Point[15], NTOP[18], Nodemap[16], NAV[1], VISUAL[2], SCPD[13], PortVis[14], NVisionIP[12], and NIVA[19], have been widely deployed over the past decade due to the growing demand for network data visualization. These tools typically present network data without taking into consideration the level of the computer user who is operating the program.

The existing Network Data Visualization Tools typically involve the presentation of network data regardless of the level of network data expertise among different levels of computer users. However, with the increasing demand for network data visualization by computer users of various levels, there is a need for new network data visualization tools. This is because the existing tools are useless and meaningless to beginner users who find it difficult to understand complicated and complex visualization. Our previous work has addressed issues including user-centered visualization that shows relevant network information to relevant computer users by classifying network data into three different levels, and providing statistical analysis and learning in visualization[20].

This paper has been organized as follows: We will discuss the existing problems. Furthermore, our proposed framework and methodology will be discussed in the following section. Details on evaluation of the results of our proposed framework will be discussed as well. We concluded the paper with a summary and some research contributions which can lead to future improvements.

**Author for correspondence*

## 2. Existing Problems

A number of visualization tools designed to visualize network data in many ways. For instance, the intrusion detection report from the system when the malicious attacks are detected[13]. However, none of the existing visualization tools are able to present adaptive visualization to different levels of users – a capability that we believe will serve as a unique complement for existing visualization tools. In this section, we will discuss three main problems.

### 2.1 Interface Usability

Due to the growth of communication network; currently, network data visualization is required by different levels of computer users. However, each user may require a different interface according to his or her perspective, and level of network knowledge. This is because different levels of computer users may require different interfaces according to their different network knowledge level. Different types of users employ different information to tackle different types of problems[10]. Visualization with a comprehensive interface can adapt to various computer users based on their expertise and feedback.

### 2.2 Visualization Flexibility

Understanding the meaning of large amounts of network data is a difficult task if the visualization is unclear and ambiguous to the computer user[11]. Appropriate network visualization for computer users has become an issue of concern[8]. The visualization without illustrate all network data to particular computer user can improve level of understanding among the diverse computer users. Flexible visualization can be achieved based on user feedback.

### 2.3 System Adaptability

Current network data visualization presents network data to different levels of computer users, who may give feedback to the system based on their gained knowledge and experience[9]. An adaptive system manages to improve the system by automatically generating appropriate visualization for different levels of computer users based on different data requirements and different feedbacks.

## 3. Proposed Framework Overview

Figure 1 in[21] shows the structure of the proposed framework which have been implemented into the proposed network data visualization tool, namely dVisStair. The framework is designed to heighten the performance of network data visualization application from the following three main perspectives: interface usability, visualization flexibility, and system adaptability.

## 4. Methodology

Visualization often tends to be a cyclic process, where each of the iteration will work to provide feedbacks into the system. A typical and conventional problem is the presentation of all types of visualization to beginner users regardless of their ability to understand. Complicated visualization seem only understandable to advanced users. Thus, the most effective way of providing visualizations is to present one that best fits the users' expertise level. Figure 1 shows the entire process from network data collection to user profile collection and the feedback learning loop. The first step involves classifying the network data into three different levels, which applies to three different categories of users. Meanwhile, user classification will cooperate with data classification in order to avoid unfeasible of visualization. So, once both classifications have been completed, a rule-based system will match them automatically, generating the proposed visualization which will be fed to the users. The feedback learning loop allows the user to provide feedback to the system. The feedback will be processed; thus, producing defined feedback. As a result, the most adaptive

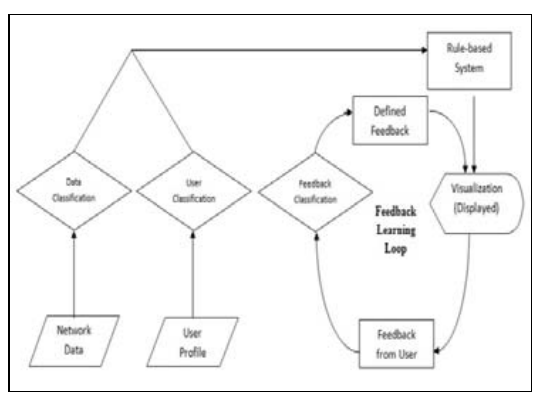


**Figure 1.** Entire cyclic process.

visualization will be generated and displayed to the users. This enhances the feedback learning loop by gaining the feedbacks from the users to be reflected back into the visualization system.

## 4.1 Data Classification

The process of capturing, decoding and analyzing the network data will be conducted simultaneously. In order to ensure different levels of users have comprehensive visualization on network data, classification of data must be applied. According to the Gulf of Execution problem, it is stated that "*It is difficult for users to ask for information or data for what he or she wants to visualize, because they do not, in generally, know what is available*"[5].

IECA is based on one of the classification algorithms, called C4.5. It is a decision tree based classification algorithm that chooses the maximum information gain ratio as the criterion of best split to generate the most meaningful classification automatically.

Two criteria are used in C4.5: information gain and gain ratio. Let *RF(C,S)* denote the relative frequency of cases in that belong to class. The information content of a message that identifies the class of a case in is then as follows (:

$$I(S) = -\sum_{j=1}^{x} RF(C_j,S)\,\log_2\left(RF(C_j,S)\right) \tag{1}$$

After $S$ is partitioned into subsets $S_1$, $S_2$, …, $S_k$ by a test $A$, the information gained is as follows (2):

$$G(S,A) = I(S) - \sum_{i=1}^{k} \frac{|S_i|}{|S|} I(S_i) \tag{2}$$

where $I(S)$ is the entropy of the $S$.

The following is the information due to the split of $S$, $S(S,A)$ on the basis of the value of the categorical attribute:

$$S(S,A) = \sum_{i=1}^{k} \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|} \tag{3}$$

Based on the algorithm, a process has to be run to classify future objects (network data) into different classes (levels). Firstly, one has to decide and assign the type of network data to be grouped as primary, intermediate or tertiary data level. Thus, the algorithm will ensure that different network data is classified completely into different levels. The IECA ensures that different network data is being classified into different levels. This will result in the best network data understanding among different levels of computer users.

## 4.2 User Classification

When dealing with a large group of users, it is necessary to classify the users. Thus, user classification is performed quickly by applying supervised algorithm. Different levels of computer users require different visualization interface. EACA is introduced and is based on one of the important classification algorithms, called NB.

- Assume $A_i$ attributes where $i = 1,2,3,\ldots, n$ and which take values $a_i$ where $i = 1,2,3,\ldots, n$.
- Assume $C$ as class label and $E = (a_1, a_2, \ldots, a_n)$ as testing instance. $E$ will be classified into class $C$ with the maximum posterior probability. Bayes' rule for this classification is (4);

$$P(C|E) = \arg\max_{c} P(C)P(E|C) \tag{4}$$

Being modified (5) as follows to define the NB;

$$P(C|E) = \arg\max_{c} P(c)\prod_{i=1}^{n} P(A_i|C) \tag{5}$$

For our scenario; $t_i$ are the features collected from the computer users.

$d = t_1$
$t_I = (t_1, t_2, t_3, t_4)$
$t_1$= experience
$t_2$= frequency usage per hour
$t_3$= interaction time
$t_4$= age

Feature $t_i$ is the attribute feature collected from the different users respectively. We compute probabilities for each class $C_j$ in(6). For each of class $C_j$, the probability is calculated that the features d belong to that class in (7) and (8).

$$C \in (beginner, intermediate, advanced) \tag{6}$$

$$P(C_j|d) = P(C_j|t_i) \tag{7}$$

$$P(C_j|t_i) = P(C_j|t_1 \ldots t_4) \tag{8}$$

The NB classifier classifies new instances (users) by applying equation (9) to compute the probability that feature $t_i$ is a member of class $C_j$. The new instance is assigned to the class with the higher probability from equation (10).

$$P\left(C_j \middle| t_i\right) = \frac{P\left(t_i \middle| C_j\right) P\left(C_j\right)}{P\left(t_i\right)} \tag{9}$$

$$P\left(C_j \middle| t_1 \ldots t_4\right) = \frac{P\left(t_1 \ldots t_4 \middle| C_j\right) P\left(C_j\right)}{P\left(t_1 \ldots t_4\right)} \tag{10}$$

By combining equation (9) and (11), the probability of feature $t_i$ ($t_1$, $t_2$, $t_3$, $t_4$) given class $C_j$, defined in (12), as a class membership for all of the attributes measurements found in the feature. When many attribute measurements have high rates of membership in class $C_j$, the probability that feature $t_i$ is a member of class $C_j$ increases. Conversely, when few attribute measurements have high rates of membership in class $C_j$, the probability that feature $t_i$ is a member of class $C_j$ decreases.

$$P\left(C_j \middle| t_i\right) = \frac{\prod_{i=1}^{4} P\left(t_i \middle| C_j\right) P\left(C_j\right)}{P\left(t_i\right)} \tag{11}$$

$$P\left(t_i \middle| C_j\right) = \prod_{i=1}^{4} P\left(t_i \middle| C_j\right) \tag{12}$$

In this method, we construct a rule which will allow us to assign future objects (computer users) to different classes (beginner, intermediate or advanced), given only the vectors of variables a.k.a classification features describing the future objects; as well as optimizing the computer users' classification. NB may not be the best possible classifier in any particular application, but it can usually be relied on to be robust and to do quite well[7]. NB assists in developing the EACA, thus providing a comprehensive algorithm that ensures complete classification among different levels of computer users.

### 4.3 Feedback Classification

Next, we will discuss the feedback learning loop which we believe is unique compared to other conventional network data visualization applications. In order to improve the visualization, a capability that enables the system to learn users' feedbacks was added. This allows the users to submit their feedback back into the system, which is useful for improving and providing adaptive visualization to the users. Regular iterations of this feedback learning loop will refine the visualization, making it more adaptive to the user. And by saving this refined relational data, future feedbacks which contain the same requirements can recall the user defined feedback, providing the user with the most adaptive and comprehensive visualization. In order to learn from the users' feedbacks, SALA will be applied to generate adaptive visualization for the computer users and improve the usability, flexibility and adaptability of the program[9]. This algorithm which is based on SVM[6] is considered a must try as it offers one of the most robust methods among all well-known algorithms. This research applies supervised methods to determine and learn what the user needs to view and what the user can understand. There are explicit and implicit information from the features to the classification task at hand. The classifier must then be able to take advantage of this information to successfully classify and learn feedbacks with as few labeled training examples as possible.

## 5. Evaluation

A one-way ANOVA was chosen and conducted for the performance evaluation. A research supported that 40 respondents were sufficient for the evaluation[17]. 40 respondents (computer users) were gathered to conduct the evaluation. They were from different backgrounds: HCI experts with interfaces experience, graphic design experts, network security, network officers, and network administrators. The respondents can be grouped into three levels: advanced, intermediate and beginner level. They helped us to evaluate the dVisStair, iNetmon and Rumint applications. Respondents were given a simple task of rating the usability, flexibility, and adaptability of these applications. Table 1 shows the Likert Scale range. Mean results of interface usability, visualization flexibility, and system adaptability will be discussed in the Results and Discussions section.

**Table 1.** Likert Scale

| Scale | Rating |
|---|---|
| 1 | Strongly Agree |
| 2 | Agree |
| 3 | Neutral |
| 4 | Disagree |
| 5 | Strongly Disagree |

### 5.1 Assessment on Interface Usability

The purpose of this assessment is to determine how useful the interface is in ensuring the full expression of the network data to different levels of computer users. A usable interface will maximize the interface usability among different levels of computer users.

Most visualization works aim to produce a good interactive system with usable interface. However, current technologies have only managed to generate demos that are attractive but have complicated interfaces[22].The most well developed visual display is ineffective if we have not carefully considered the users' levels, needs, and system interfaces. Therefore, interface usability plays a vital role in increasing the speed of tasks carried out among computer users. When usability was evaluated, the focus was on improving the user interface. We compared the application interfaces generated among computer users. dVisStair showed three different interfaces based on different levels of computer users. On the other hand, iNetmon showed only the interfaces that can be understood by network administrators. This is similar to the Rumint application.

In our scenario, experts were used to evaluate early prototypes. Then, end users were used to evaluate a refined version. We did not use expert review exclusively as we believe that it is a complement to formal user studies which cannot be replaced.

### 5.2 Assessment on Visualization Flexibility

The purpose of this assessment is to determine the efficiency of the application in presenting effective visualization of network data to different levels of computer users (compared with iNetmon and Rumint). With effective visualization, both little or large amounts (flexible) of network data can be presented to computer users without burdening them; especially for beginner users.

Most visualization works aim to produce flexible and easy visualization. If the displayed visualization is too complicated or too limited, it might become meaningless to the computer users. Therefore, visualization flexibility is important not only to improve visual representation, but also to ensure clear understanding among computer users. The objective of evaluating visualization flexibility is to improve computer users' understanding of network data. dVisStair managed to present a visualization that is simple and flexible to computer users of different levels. iNetmon and Rumint are also able to display different visualizations but still leave computer users feeling confused and unable to understand the information from the visualization.

### 5.3 Assessment on System Adaptability

The purpose of this assessment is to determine whether the system is able to adapt automatically to feedback from individual users. Only an adaptive system can fulfill users' needs. Most current systems only provide customization for users. This results in complications for users who are not familiar with the network data. Therefore, system adaptability is needed to adopt into the tool. The aim of evaluating the system adaptability is to produce an adaptive system that is able to generate an adaptive interface together with visualization for the user automatically.

## 6. Results And Discussions

The evaluation from the previous section has proven that, the more feedback we get, the more adaptive visualization can be achieved in order to present greater adaptive visualization to the users. dVisStair has improved interface usability and is more user-friendly. On top of that, the visualization flexibility has been improved as the system has learnt from the users' feedback and is therefore, able to provide better visualization to the users. System adaptability has also been achieved as the system managed to learn the user's feedback intelligently and customize adaptive visualization for the computer users 12. In our scenario, post hoc comparisons using the Tukey HSD test indicated that the mean score for dVisStair on interface usability (Mean = 3.86, SD = 0.93) was significantly different from iNetmon (Mean = 2.35, SD = 0.93). However, the Rumint (Mean = 1.96, SD = 0.97) did not significantly differ from iNetmon. The mean of interface usability is shown in Figure 2. Based on the Tukey HSD test, dVisStair rated as the most flexible visualization (Mean = 3.52, SD = 0.78). iNetmon has the second highest rate (Mean = 2.56, SD = 0.74). The application flexible visualization was Rumint (Mean = 2.13, SD = 0.83). The mean of visualization flexibility is shown in Figure 3. Figure 4 shows the mean of system adaptability of the network data visualization applications. As shown in the Tukey HSD test, the achievements of the dVisStair application (Mean = 3.33, SD = 0.93) were significantly different from iNetmon (Mean = 2.00, SD = 0.80) and Rumint (Mean = 2.09, SD = 0.72).

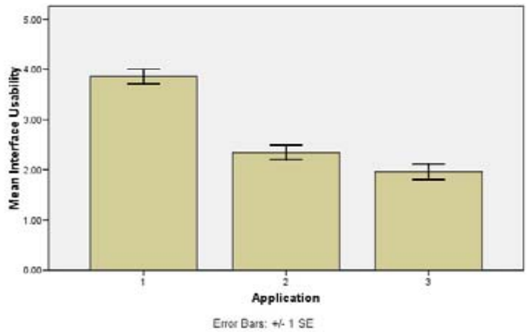


Figure 2. Mean of Interface Usability.

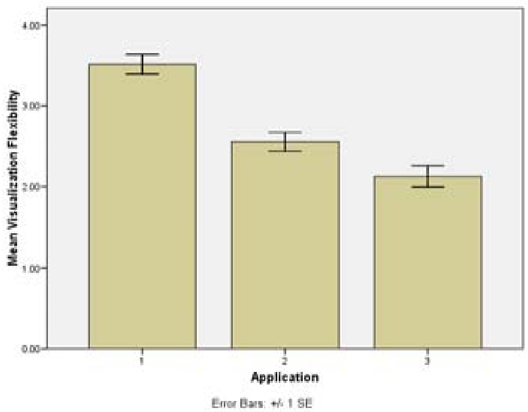


**Figure 3.** Mean of Visualization Flexibility.

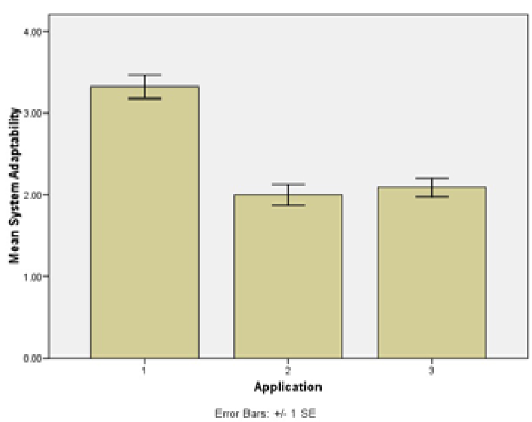


**Figure 4.** Mean of System Adaptability.

## 7. Conclusion

The capability to illustrate different network data visualization has become extremely important due to the demands from different levels of computer users. We have combined EACA, IECA, and SALA which work together with rule-based algorithm to produce adaptive comprehensive visualization for all users. Evaluations have been carried out among 40 respondents. The results proved that by using the supervised algorithms in our proposed framework, the dVisStair has improved in many aspects; the most important aspects being interface usability, visualization flexibility, and system adaptability. The main contribution of this research work is to ensure that different interfaces and visualizations are available for different levels of computer users. Furthermore, we believe that the results of this research will be a powerful driving force for the network data visualization world to serve the needs of different levels of computer users.

## 8. References

1. Allen M, Mclachlan P. NAV Network Analysis Visualization. University of British Columbia. 2009
2. Ball R, Fink G A, North C, Home-Centric Visualization of Network Traffic For Security Administration. Proceedings of The 2004 ACM Workshop on Visualisation and Data Mining For Computer Security (Vizsec/Dmsec 04); 2004. p. 55–64.
3. Becker RA, Eick SG, Wilks AR. Visualizing Network Data. IEEE Transactions on Visualization and Computer Graphics. 1995; 1(1):16–28.
4. Brown JA, Mcgregor AJ, Braun HW. Network Performance Visualization: Insight through animation. Proceedings of PAM2000 Passive and Active Measurement Workshop; 2000.
5. Card S, Robertson G, Mackinlay J. The Information Visualizer. An Information Workspace. Proceedings of ACM Human Factors In Computing Systems Conference (Chi'91); 1991. p. 181–8.
6. Chen Y, Pedrycz W, Watada J. A Fuzzy Regression Based Support Vector Machine (SVM) Approach to Fuzzy Classification. ICIC Express Letters-An International Journal of Research and Surveys. 2010; 4(6):2355–62.
7. Domingos , Pazzani M. On The Optimality of The Simple Bayesian Classifier Under Zero-One Loss. Machine Learning. 1997; 29:103–30.
8. El-Shehaly M, Gracanin D, Abdel-Hamid A, Matkovic K. A visualization framework for traffic data exploration and scan detection. 3rd International Conference on New Technologies, Mobility and Security; 2009. p. 1–6.
9. Ghazarian A, Noorhosseini SM. Automatic detection of users' skill levels using high-frequency user interface events. Journal of User Modeling and User-Adapted Interaction. 2010; 20(2): 109–-46.
10. Heer J, Van Ham F, Carpendale S, Weaver C, Isenberg P. Creation and collaboration: Engaging new audiences for

information visualization. Lecture Notes in Computer Science. 2008; 92–133.

11. Keim DA. Information visualization and visual data mining. IEEE Transactions on Visualization and Computer Graphics. 2002; 8(1):1–8.
12. Lakkaraju K, Yurcik W, Lee AJ. NVisionIP: Net-Flow Visualizations of System State For Security Situational Awarenes. Proceedings of The 2004 ACM Workshop on Visualisation And Data Mining For Computer Security (Vizsec/Dmsec '04); 2004. p. 65–72.
13. Lau. S. The Spinning of Potential Doom. Commun ACM. 2004; 47(6): 25–26.
14. Mcpherson J, Ma KL, Krystosk P, Bartoletti T, Christensen M. Portvis: A Tool For Port-Based Detection of Security Events. Proceedings of The 2004 ACM Workshop on Visualisation And Data Mining For Computer Security (Vizsec/Dmsec '04); 2004. p. 73–81.
15. Wildpackets. Watchpoint. 2010 Jan. Available from: http://www.wildpackets.com
16. Newton M. 2009 May. Available from: http://nodemap.internode.on.net/
17. Nielsen J. Qualitative testing. 2006. Available from: http://www.useit.com
18. Ntop. 2009 May. Available from: http://www.ntop.org/documentation.html.
19. Nyarko K, Capers T, Scott C, Ladeji-Osias K. Network Intrusion Visualization With NIVA, An Intrusion Detection Visual Analyzer With Haptic Integration. 10th Symposium Proceeding Haptic Interfaces For Virtual Environment and Teleoperator Systems; 2002. p. 277–84.
20. Wong DHT, Ramadass S. Chai KS. Towards An Innovative Framework For Network Data Visualization. ICIC Express Letters-An International Journal of Research and Surveys. 2012; 6(2): 425–30.
21. Wong DHT, Ramadass S. Adaptive Interfaces For Multiple Levels of Users In Network Data Visualization. 15th International Workshop on Advanced Image Technology; 2012. p. 844–8.
22. Tory M, Moller T. Evaluating visualizations: Do expert reviews work? IEEE Computer Graphics and Applications. 2005; 25:8–11.