APPLYING BUSINESS INTELLIGENCE IN INTERNET TRAFFIC ANALYSIS: A CASE STUDY OF INFORMATION AND COMMUNICATION TECHNOLOGY SILPAKORN UNIVERSITY

KAMPHON KORNANAN

A THEMATIC PAPER SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE (INFORMATION TECHNOLOGY MANAGEMENT) FACULTY OF GRADUATE STUDIES MAHIDOL UNIVERSITY 2016

Thematic Paper
entitled

APPLYING BUSINESS INTELLIGENCE IN INTERNET TRAFFIC ANALYSIS: A CASE STUDY OF INFORMATION AND COMMUNICATION TECHNOLOGY SILPAKORN UNIVERSITY

.....
Mr. Kamphon Kornanan
Candidate

.....
Lect. Sotarat Thammaboosadee,
Ph.D. (Information Technology)
Major advisor

.....
Lect. Chanattha Chansutthirangkool,
Ph.D. (Information Technology)
Co-advisor

.....
Prof. Patcharee Lertrit,
M.D., Ph.D. (Biochemistry)
Dean
Faculty of Graduate Studies
Mahidol University

.....
Asst. Prof. Supaporn Kiattisin,
Ph.D. (Electrical and Computer Engineering)
Program Director
Master of Science Program in Information Technology Management
Faculty of Engineering,
Mahidol University

Thematic Paper
entitled

APPLYING BUSINESS INTELLIGENCE IN INTERNET TRAFFIC ANALYSIS: A CASE STUDY OF INFORMATION AND COMMUNICATION TECHNOLOGY SILPAKORN UNIVERSITY

was submitted to the Faculty of Graduate Studies, Mahidol University
for the degree of Master of Science (Information Technology Management)
on
January 4, 2017

.....
Mr. Kamphon Kornanan
Candidate

.....
Lect. Taweesak Samanchuen,
Ph.D. (Electrical Engineering)
Chair

.....
Lect. Sotarat Thammaboosadee,
Ph.D. (Information Technology)
Member

.....
Asst. Prof. Sustarum Thammaboosadee,
Ph.D. (Political Science-International Relations)
Member

.....
Lect. Chanattha Chansutthirangkool,
Ph.D. (Information Technology)
Major advisor
Member

.....
Prof. Patcharee Lertrit,
M.D., Ph.D. (Biochemistry)
Dean
Faculty of Graduate Studies
Mahidol University

.....
Asst. Prof. Jackrit Suthakorn,
Ph.D. (Robotics)
Dean
Faculty of Engineering
Mahidol University

ACKNOWLEDGEMENTS

The success of this thematic can be succeeded by attentive support from advisor. I must thank for helped and lead me to the goal that was smooth. Thanks to the parents to encourage their support for the study. Thanks my cousin for encouraging continued. Thanks the support from class mates. Thanks to the supervisors a chance to study. I appreciate the everybody in work place for the chance in this study.

Thanks reference work of the study to the example in doing research, the idea of the study and an example in the way of a good education.

Kamphon Kornanan

APPLYING BUSINESS INTELLIGENCE IN INTERNET TRAFFIC ANALYSIS: A CASE STUDY OF INFORMATION AND COMMUNICATION TECHNOLOGY SILPAKORN UNIVERSITY

KAMPHON KORNANAN 5736273 EGIT/M

M.Sc. (INFORMATION TECHNOLOGY MANAGEMENT)

THEMATIC PAPER ADVISORY COMMITTEE: SOTARAT THAMMABOOSADEE, Ph.D., CHANATTHA CHANSUTTHIRANGKOOL, Ph.D.

ABSTRACT

Information and communication technology (ICT) Silpakorn University, who has provided the internet service has got the user complaints concerning internet service problem. Their staff in ICT could not solve the right problem. To find the right problem they should buy network monitoring application for displaying the internet usage. However, their IT budget was not approved for buying software. We are using a solution like network monitoring application by using statistical application for the analysis of network traffic. In analysis process, we used sample data in the period from March to May 2016.

The perspective presentation of traffic analysis would be appropriate in statistic and it was based on the network traffic variables. It showed the amount of bandwidth usage, the deployment of applications and inter traffic usage.

The result of network analysis was presented in the perspective of traffic analysis. Lab1 and Lab3 had the most bandwidth usage. Internet application traffic mostly used HTTP/HTTPS. The internet usage was usually used at 16.00. The result of traffic analysis dashboard evaluation was evaluated by Deputy Dean of Information Technology ICT. They were interested in implementation of methodology and complained about report layout arrangement. The internet traffic analysis was appropriate with the perspective presentation.

KEY WORDS: NETWORK TRAFFIC / BUSINESS INTELLIGENCE / DASHBOARD / TRAFFIC ANALYSIS

66 pages

การประยุกต์ใช้ธุรกิจอัจฉริยะในการวิเคราะห์ข้อมูลการจราจรเครือข่าย กรณีศึกษาของคณะเทคโนโลยีสารสนเทศและการสื่อสาร มหาวิทยาลัยศิลปากร

APPLYING BUSINESS INTELLIGENCE IN INTERNET TRAFFIC ANALYSIS: A CASE STUDY OF INFORMATION AND COMMUNICATION TECHNOLOGY SILPAKORN UNIVERSITY

กำพล กรนานันท์ 5736273 EGIT / M

วท.ม. (การจัดการเทคโนโลยีสารสนเทศ)

คณะกรรมการที่ปรึกษาสารนิพนธ์: โยฑศ์รัตต ธรรมบุษดี , Ph.D., ชนัฏฐา จันสุทธิรางกูร, Ph.D.,

บทคัดย่อ

คณะเทคโนโลยีสารสนเทศและการสื่อสารให้บริการงานเครือข่ายและรับคำร้องเรียนจากผู้ใช้งานเรื่องปัญหาการใช้อินเตอร์เน็ตแต่ไม่สามารถแก้ไขปัญหาได้ตรงจุด จึงรวบรวมข้อมูลการใช้เครือข่ายเพื่อหาต้นเหตุของปัญหาโดยนำโปรแกรมสำหรับวิเคราะห์ข้อมูลทางสถิติมาประยุกต์ใช้ในการวิเคราะห์ข้อมูลการใช้เครือข่ายที่ถูกเก็บไว้จากกลุ่มตัวอย่าง การใช้เครือข่ายของคณะเทคโนโลยีสารสนเทศและการสื่อสารในช่วง เดือนมีนาคมถึงเดือนพฤษภาคม ปี 2559 โดยใช้หลักสถิติในการวิเคราะห์ข้อมูล

ในการนำเสนอมุมมองของการวิเคราะห์เครือข่ายให้มีความเหมาะสมตามตัวแปรของข้อมูลเครือข่ายและสถิติ โดยแสดงตามปริมาณแบนวิดท์ที่ใช้, แสดงการใช้งานแอพลิเคชั่น , แสดงการใช้งานในกลุ่มเครือข่าย และ แสดงการใช้งานเครือข่ายภายนอกประเทศ

ผลการวิเคราะห์ข้อมูลพบว่า ส่วนใหญ่ผู้ใช้งานเครือข่ายอยู่ในกลุ่มห้อง Lab 1 และ Lab 3 ในส่วนของการใช้งานเครือข่ายนิยมใช้แอปพลิเคชั่น HTTP / HTTPS มากที่สุด ซึ่งอยู่ในช่วงเวลา 16.00 น. ส่วนใหญ่มักใช้งานภายในประเทศมากที่สุดผลที่ได้จากการประเมินการใช้งานแดชบอร์ดของการวิเคราะห์ข้อมูลเครือข่ายจากรองคณบดีฝ่ายเทคโนโลยีสารสนเทศ โดยได้รับคำแนะนำจากการปรับปรุงแดชบอร์ด เพื่อให้ง่ายต่อการเข้าใจข้อมูล จากผลลัพธ์การใช้โปรแกรมในการวิเคราะห์ข้อมูลมีความเหมาะสมและสามารถนำไปใช้ในการวิเคราะห์เครือข่ายแทนโปรแกรมเฉพาะได้

66 หน้า

CONTENTS

CONTENTS(cont.)

CONTENTS(cont.)

LIST OF TABLES

.

LIST OF FIGURES

LIST OF FIGURES (cont.)

CHAPTER I
INTRODUCTION

1.1 Statement of Problems

The Faculty of Information and Communications Technology (ICT), Silpakorn University, Phetchaburi Campus, consists of around 1,300 users which are staff, teachers, and students. Also, there are approximately 300 internet users per day. The Faculty of ICT has leased a 50 Mb/s internet speed for general use connecting to external sites, document delivery, mailing applications, and also teaching. However, the quota is provided without limitation of other network ports. To widely provide the internet and diversified uses, ICT faculty uses Fortinet FortiGate 600C to control network and store traffic data for creating a report. Unfortunately, the FortiGate 600C cannot show more than 24 hours historical usage data. From this reason, the researcher has chosen the Syslog server to keep traffic data by forwarding data to Kiwi Syslog server, and keep in text file format. In each day, there are more than 1 million lines recorded in the traffic log file.

With this amount of users, the daily requirements on the network are huge and in high volume. With the program frequently used for connecting the system to interface with the server, high volume of data transmission and users increasing over time causes problems of congestion in the network. Hence, the analysis of the network traffic requires network monitoring to display network traffic. However, the IT department has not been approved for the procurement program for the network from the board because it is expensive. In order to analyze the use of the network, leading business intelligence tool used to analyze and display network traffic analysis of the network in the form of a report in business intelligence dashboard to easily reading.

From motivation mentioned above, the presentation of information requires analyzing a large amount of information and presenting information in a format that is easily understandable. Simplifying for presenting and understanding the information of network traffic, the Business Intelligence (BI) dashboard is used for presenting

information to the network of the ICT faculty and replacing the network monitoring tool.

1.2 Objective

This topic aims to analyze the network log and create reports by use BI Dashboard for be applied and replace the software that must be purchased.

1.3 Scope of work

1. Collecting all forward traffic log in the network from March to May 2016.
2. Analyzing network log in the Faculty of Information and Communication Technology, Silpakorn University, Phetchaburi ICT Campus from 9.00 to 18.00.
3. Presenting network Bandwidth, Date, Time, Source, Destination and Application by the use of Business Intelligence Dashboard.

1.4 Benefit expected

In this topic, we will use the business intelligence for analysis and presentation of network log. We will obtain a network monitoring and analysis software.

CHAPTER II
LITERATURE REVIEW

This chapter provides the theory and the related research.

2.1 Network System

Network system is a group of computers and computer equipment is connected to the network so that users can connect to each other. There are many types of networks like LAN, MAN, and WAN. The network consists of a computer, network card, and intermediate for data transmission such as cable link, protocol (For communication, it is called as TCP / IP), network operating system for managing and controlling the use of network resources such as Windows Server, Novell Netware, Red Hat, fortigate. This feature saves the connection data with other computers in the network or IP Address of source and destination computer data known as traffic log which will be discussed in the next section.

2.1.1 LAN (Local Area Network)

Local Area Network is a group of the computers communicating between a line or wireless link and a server to share resources such as a printer or network storage [1]. Local Area Network is interconnecting computers within a limited area, about 1 or 2 kilometers, such as a school, laboratory or university campus.

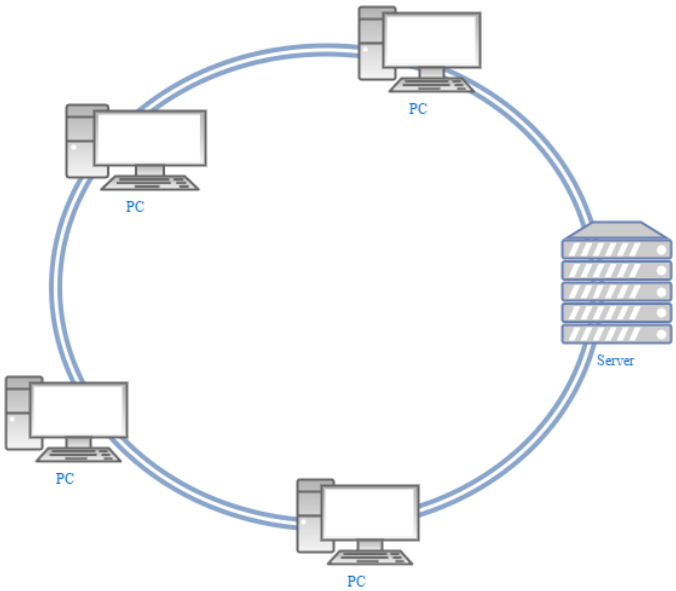


Figure 2.1 Local Area Network.

2.1.2 MAN (Metropolitan Area Network)

Metropolitan Area Network is a group of computers and devices that connects users with computer resource that cover by a large Local Area Network but smaller than the area covers by a Wide Area Network [1]. Metropolitan Area Network interconnects computers within a limited area, about 10 kilometers.

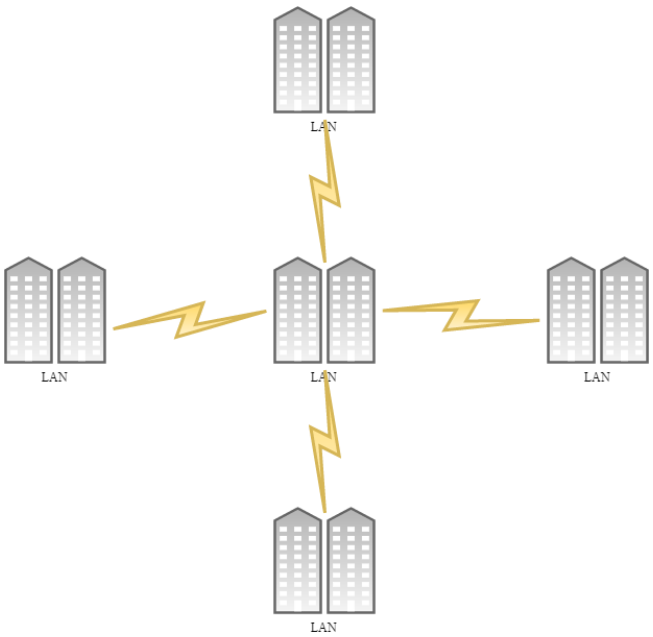


Figure 2.2 Metropolitan Area Network.

2.1.3 WAN (Wide Area Network)

Wide Area Network is a network that exists over a large-scale geographical area covering a country or around the world [1]. The channels for communication are cable and microwave satellite. The Wide Area Network connects different smaller networks including Local Area Network and Metropolitan Area Network.

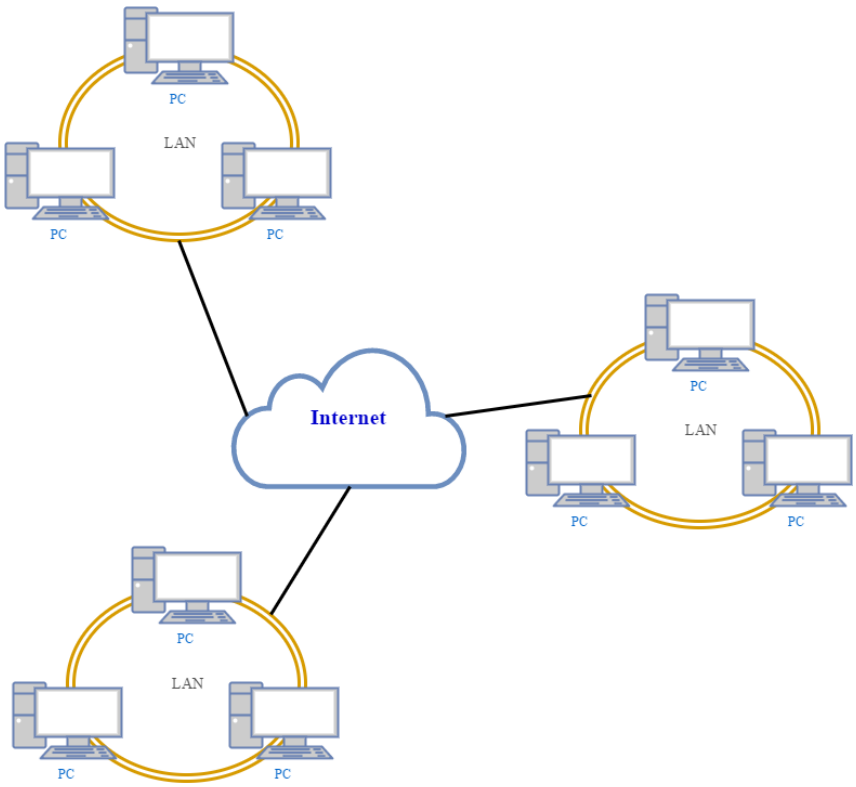


Figure 2.3 Wide Area Network.

2.1.4 TCP/IP (Transmission Control Protocol / Internet Protocol)

Transmission Control Protocol/Internet Protocol is a network communication language for communication between network device on the internet to communicate from source device to destination device and find network route for sending data by automatically [1]. The Transmission Control Protocol are two roles in network communication, Source of Transmission Control Protocol disport data into many packages and send to the destination by Internet Protocol and destination Transmission Control Protocol receive the package, check package and compose the received package into the data. Internet Protocol role is managing the sending package from source to destination device by using IP Address for identifying source and destination location. Internet Protocol Address (IP Address) is a unique

number of address that uses for data communications and the packet of data can send to the destination device.

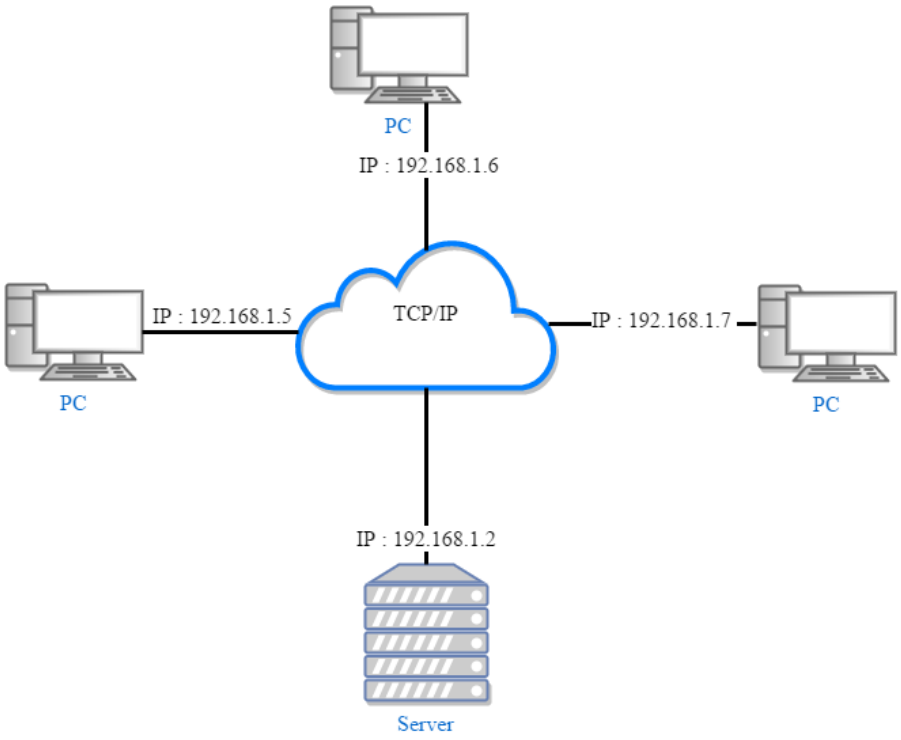


Figure 2.4 Transmission Control Protocol / Internet Protocol.

2.2 Firewall

Firewall is a system protection the network from internal and external threats that going in corporate security systems. Within the security policy based on threat prevention and firewall capabilities [2], it used to manage security on the network (e.g. managing the web filter, antivirus, intrusion protection, VPN, real-time monitor, log, and report). A general firewall is network firewall on the market the device offers a variety of brands (e.g. Cisco, Fortinet, Jupiter Network and WatchGuard Firewall). Another type of firewall is software firewall that is available in operating systems (e.g. FreeBSD, Linux, KerioOS and CentOS). It has the ability like the hardware firewall. The hardware firewall is the network security which in broadband router. We will discuss logs theory in traffic logs.

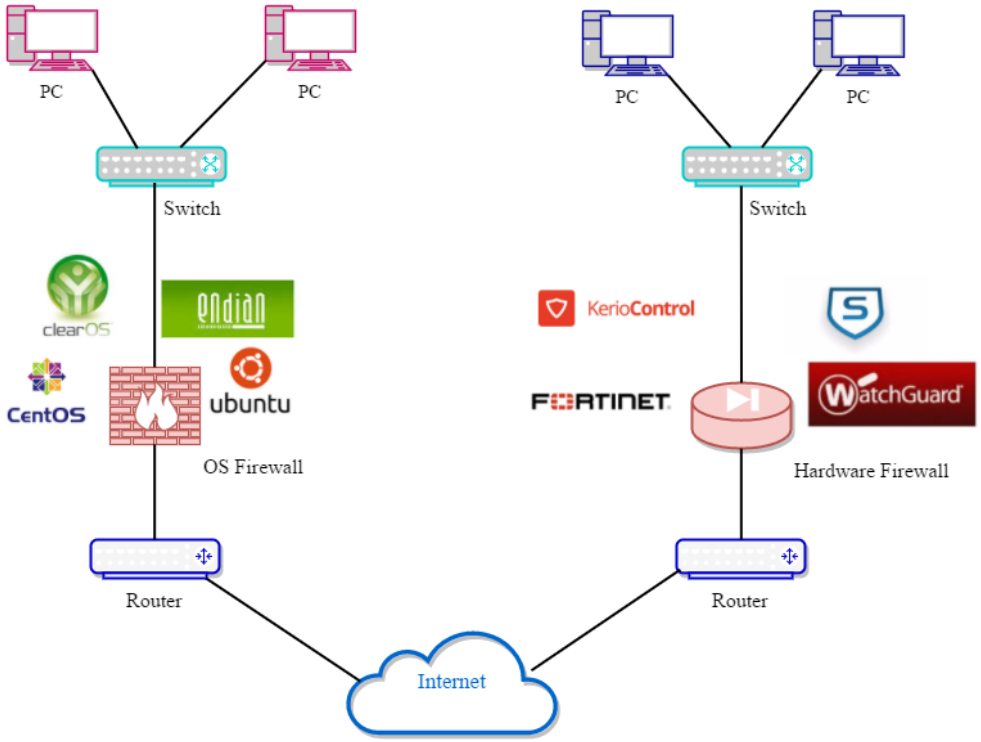


Figure 2.5 Hardware and software firewall.

2.3 Network Logs

Network usage data is recorded its flow through the network firewall policies, because of the demand of data traffic that goes through authentication login. The policy of the firewall controls all traffic trying to pass between the interface of the firewall and Fortigate interface of VLAN. Logging traffic policy data traffic follows firewall policy has logging enabled on Log Allowed traffic and Log Violation traffic [3] as exemplified in Figure 2.5.

- The packet comes with an inbound interface.
- Log packet sent to compare firewall policy such as URL filter.
- Traffic log packet was allowed just 1 firewall policy.
- Packet passes and sent out an interface.

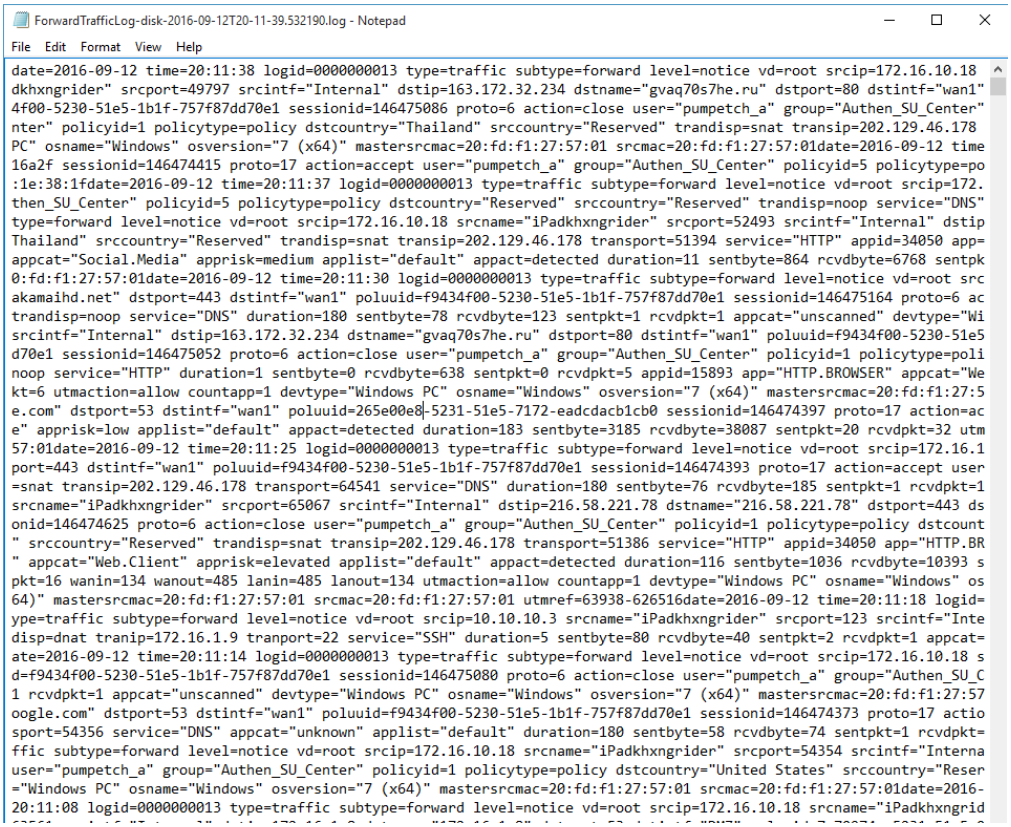


Figure 2.6 Fortigate traffic log.

Network log message is stored in the log file or any log device or event system memory. Connection information is stored in the form of a Syslog. Network log contains a type field that identifies log type and which log file it is stored following:

- Traffic

Records traffic flow information, such as an HTTP/HTTPS request and its response, if any.

- Security (UTM)

Records virus attack and intrusion attempts.

- Event

Records system and administrative events, such as downloading a backup copy of the configuration, or daemon activities.

Each log might contain a subtype field in log type for example:

- In traffic logs, the sub types are local, forward, multicast, and sniffer.

- In security (UTM) logs, some log entries have a subtype of DLP, Web Filter, Email or other sub types.
- In event logs, some log entries have a subtype of user, system, or other sub types.

For storing network log, the device and software have a standard log to store the network information which we will talk about Syslog protocol in the next theory.

2.4 Syslog Protocol

Syslog protocol is a standard log introduced by Berkeley Software Distribution (BSD) [4]. It is used for stored information web access, IP Address and Date and Time in order to trace and specify the user. The Syslog protocol is defined by RFC (Request for Comments) document published by the internet Engineering task force [5] [6] as follows:

- RFC 3164 The BSD Syslog Protocol (obsoleted by RFC 5424)
- RFC 3195 Reliable Delivery for Syslog
- RFC 5424 The Syslog Protocol
- RFC 5425 TLS Transport Mapping for Syslog
- RFC 5426 Transmission of Syslog Messages over UDP
- RFC 5427 Textual Conventions for Syslog Management
- RFC 5848 Signed Syslog Messages
- RFC 6012 Datagram Transport Layer Security (DTLS) Transport Mapping for Syslog
- RFC 6587 Transmission of Syslog Messages over TCP

The Syslog has many formats depending on Log Server such as 3com, Cisco, Pix firewall and kiwi log presented as follows.

2.4.1 Kiwi format

Kiwi is a network software management tool developed by SolarWinds and used for storage management and data in order to monitor the performance of the device

events with the information collected at the facility, hostname, and information about traffic and recorded in a database with a sample format shown in Figure 2.6 [7].

Format: Date (DD-MM-YYYY) [TAB] Time (HH:MM: SS) [TAB] Priority (Facility.Level) [TAB] Host name [TAB] Message text
Example: 22-07-2002 [TAB] 12:34:56 [TAB] Local5.Debug [TAB] firewall-inside [TAB] prot=UDP port=53 dst=203.25.36.47 src=192.168.1.2 bytes=64

Figure 2.7 Kiwi log format.

2.4.2 Sawmill format ISO

Sawmill is a log analysis tool of Flowerfire, Inc. available since 1997, which provides log analysis and reporting solutions for log files from web servers, media servers, security applications, mail servers, firewalls, proxy, gateways, and FTP servers [8]. Log files stored by Sawmill log in Sawmill format are shown in Figure 2.7.

Format: DateTime (YYYY-MM-DD HH:MM:SS) [TAB] Priority (Facility.Level) [TAB] Hostname [TAB] Message text
Example: 2002-07-22 12:34:56 [TAB] Local5.Debug [TAB] firewall-inside [TAB] prot=UDP port=53 dst=203.25.36.47 src=192.168.1.2 bytes=64

Figure 2.8 Sawmill log format.

2.4.3 Comma Separated Values (CSV)

CSV is a format for storing data characters in plain text comprising field is separated by a tab, space, comma (,). Comma Separated Values was first used in 1972, in IBM FORTRAN programming file type is standard RFC 4180 [9]. There are several certification programs support applications such as Notepad and Microsoft Excel which have the following format.

Format: DateTime (YYYY-MM-DD HH:MM:SS),Priority (Facility.Level),Hostname,Message text
Example: 2002-07-22 12:34:56,Local5.Debug,firewall-inside,"prot=UDP port=53 dst=203.25.36.47 src=192.168.1.2 bytes=64"

Figure 2.9 Comma Separated Values format.

2.5 Computer Crimes Act (CCA)

The cause of technology and the development of Thailand has changed to the better way of internet use in a country with the wide use but difficult to control internet using. The Ministry of Information and Communication and the draft adopted Act. PC on June 18th, 2550 called Computer-Related Crime Act B.E. 2007 [10] in order to force providers to keep the traffic log data according to the law, stated in part 2 of Competent Officials Section 26: Computer-Related Crime.

The Ministry of Information and Communication drafts the Act. Personal Computer on June 18th, 2550 called Computer-Related Crime Act B.E. 2007 [10] in order to force providers to keep the traffic log data according to the law, stated in part 2 of Competent Officials Section 26: Computer-Related Crime presented as follows:

"A service provider shall keep traffic data for not less than ninety days from the day when such data has been entered into a computer system. If necessary, the competent official shall, as a particular case and time, instruct any service provider to keep traffic data for over ninety days but not exceeding one year.

A service provider shall keep user's data as necessary for the purpose of identifying the user from the first day of such a service and store such user data for not less than ninety days from its expiry date.

The Minister shall prescribe the type of service providers, how and when the provisions of paragraph one shall apply by promulgation in the Government Gazette.

Any service provider, who fails to comply with this Section, shall be liable to a fine not exceeding five hundred thousand Baht."

Storage applications Network Traffic of the organization will need to be compliant data recording applications act with that action, the Computer Crime Act 2550, section 26, paragraph 3, announced by the Ministry of Information and Technology, communicate Rules retention of traffic data providers [11]. The organization is classified into the category, access service provider for institutions. Criteria for the storage of a computer system access for Institutions have responsibility for keeping the information below.

Table 2.1 The information for keeping in another log type.

Log type	Information
Network Log	1 . Access Logs Specific to Authentication and Authorization Servers such as TACACS (Terminal Access Controller Access-Control System) or RADIUS (Remote Authentication Dial-In User Service) DIAMETER (Used to Control Access to IP Routers or Network Access Servers)
	2. Date and Time of Connection of Client to Server
	3. User ID
	4. Assigned IP Address
	5.Calling Line Identification
E-mail servers Log	1.Simple Mail Transfer Protocol: SMTP Log includes; - Message ID - Sender's E-mail address - Receiver's E-mail Address - Status Indicator including Sent e-mail, Reply e-mail, and Delayed e-mail
	2. IP Address of Client Connected to Server
	3. Date and time of connection of Client Connected to server
	4. IP Address of Sending Computer
	5. User ID (if stored)

Table 2.1 The information for keeping in another log type (cont.).

	6 . The recorded data access via e-mail program from members of or access information e-mail to the members. The continued storage of data, data access, electronic mail, which is pulled to the service. (POP3(Post Office Protocol version3) Log or IMAP4 (Internet Message Access Protocol; Version4) Log)
Transfer files	1.Access server log
	2.Date and Time of Connection of Client to Server
	3.IP source Address
	4.User ID
	5.Path and Filename of data Object Uploaded or Downloaded
Web Server	1.Access server log
	2.Date and Time of Connection of Client to Server
	3.IP source Address
	4.Command information system applications.
	5.The data indicate the path to retrieve the data(URI: Uniform Resource Identifier) such as URL (Universal Resource Locator)
Usenet	1.Access log (NNTP(Network News Transfer Protocol) Log)
	2.Date and Time of Connection of Client to Server
	3.Port number (Protocol Process ID)
	4. Hostname
	5.Posted Message ID
Data from interactions on the Internet for example Internet Relay Chat (IRC) or Instance Messaging(IM) etc.	Date and time of connection client to server) and hostname and client IP address)

In the traffic information network used in the analysis, in order to be legal and safe for use in research. Thus, the Law on the reference to Computer Crimes Act because the data is expired or no longer needed to use and not to the extent permitted by law will not be used for illegal acts on the computer. The traffic information network used to last for more than 90 days and can be used for data analysis.

2.6 Data Visualization

Data presented in the report to show the results of the data analysis to be able to understand it simple to save time in describing the data and make the presentation interesting remember to read or view current presentation with graphs or charts in a variety of report formats used to consider the appropriateness of implementation, stating the purpose of the information is presented [12]. The benefits of the chart are follows

2.6.1 Column Charts

As illustrated in Figure 2.9 [13], the purpose of column charts is to show trends during the time, to group related data and compare data. It is useful for the discrete data or showing trends over time.

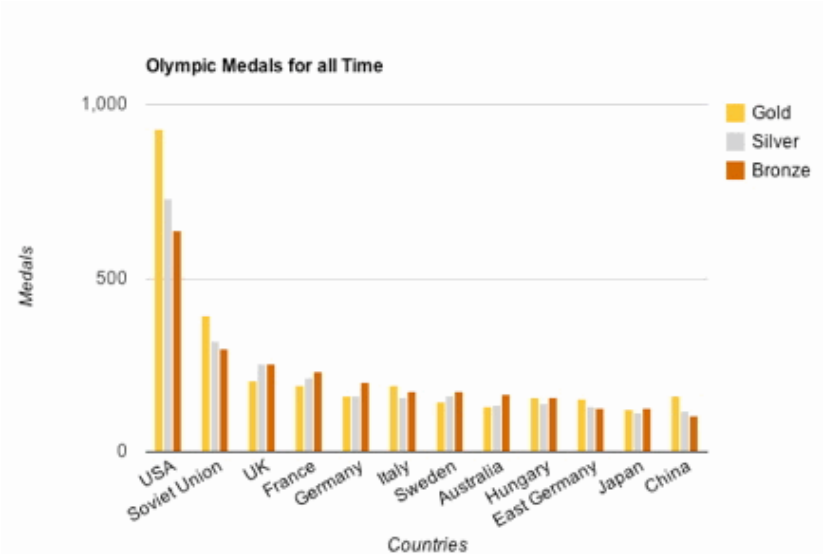


Figure 2.10 Example of Column chart.

2.6.2 Line Charts

The purpose of line charts is to show trends during the time. It is useful for showing trends over time and to compare multiple sets of data. As illustrated in Figure 2.10 [14].

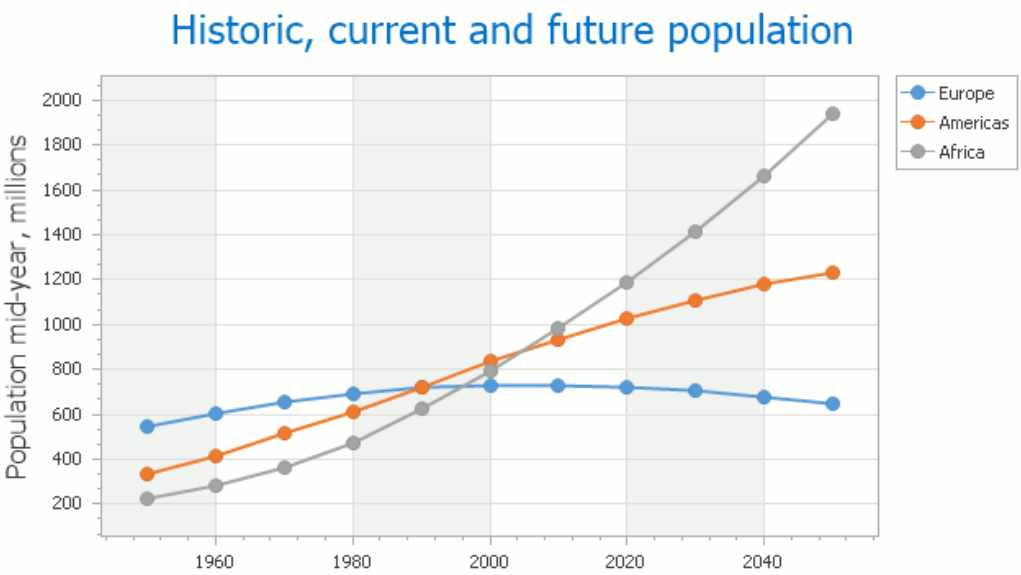


Figure 2.11 Example of Line chart.

2.6.3 Pie Charts

The purpose of pie charts is to show the relationship of all the parts or a portion highlighted. It is useful for highlighting the disproportionate use of the circle, to show the relationship all highlight the true value of chart types such stacked chart. An example of the pie chart is shown in Figure 2.11[15].

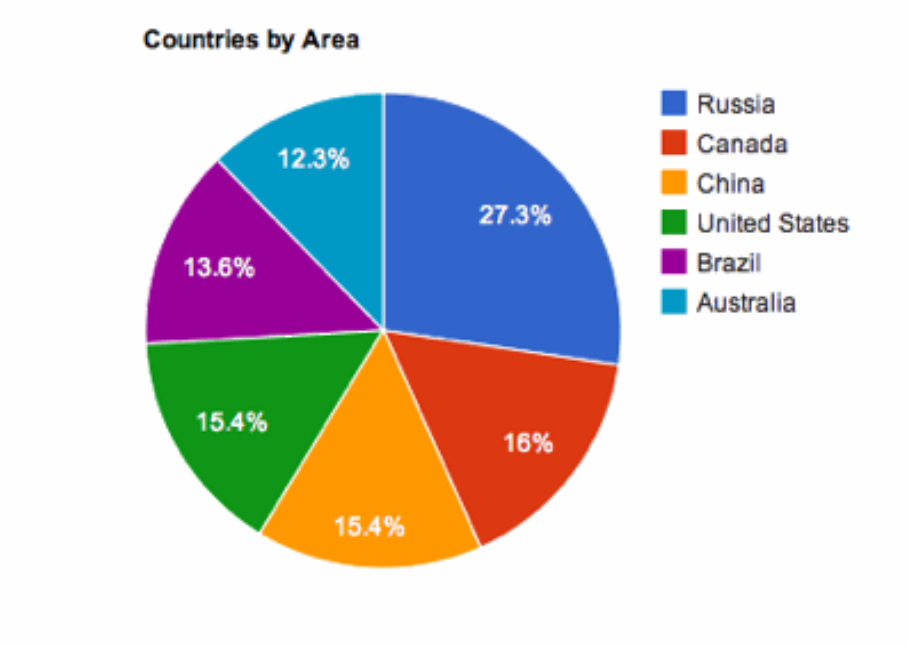


Figure 2.12 Example of the pie chart.

2.6.4 Bar Charts

The purpose of bar charts is to compare data and to group related data. It is useful for plotting a series of multiple data sets, as shown in Figure 2.12 [16].

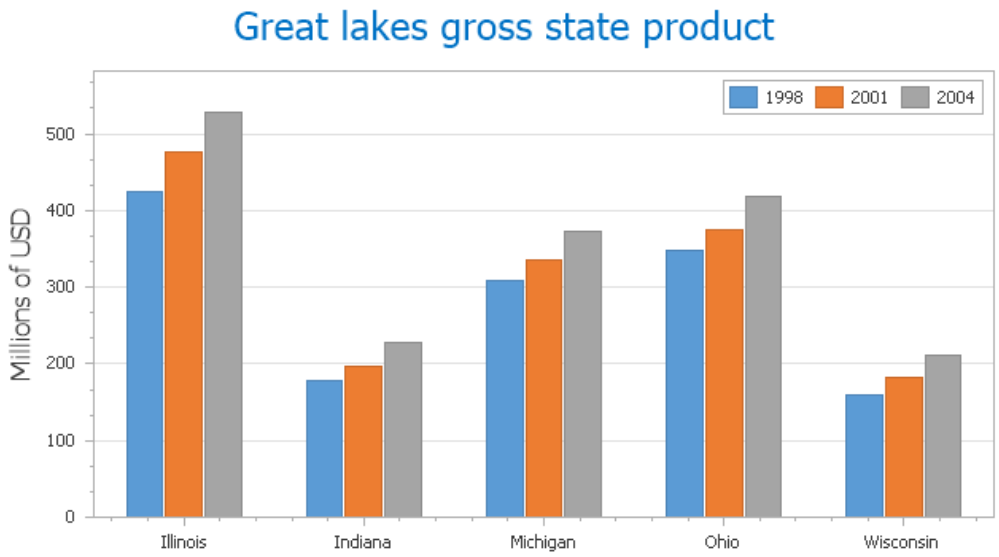


Figure 2.13 Example of a bar chart.

2.6.5 Area Charts

The purpose of area charts is to stress the importance of changes during time. It is useful for highlighting a number of changes over time. Area chart also uses the stack to show the relationship of parts to the whole. An example of the area chart is shown in Figure 2.13[17].

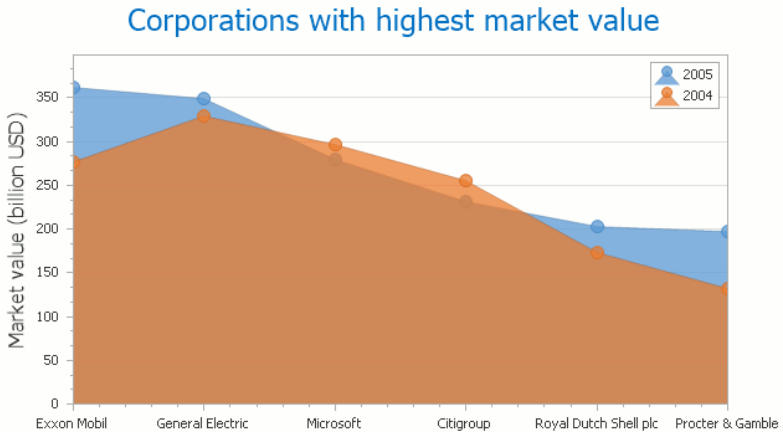


Figure 2.14 Example of the area chart.

2.6.6 Scatter Charts

As illustrated in Figure 2.14[18], the purpose of scatter charts is displaying the type of relationship between two variable of data. It is useful for the relationship between the two measured.

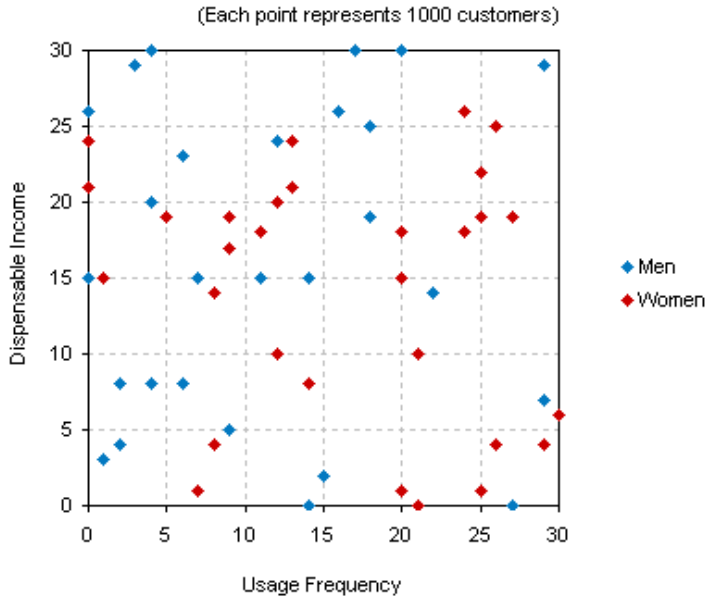


Figure 2.15 Example of scatter chart.

2.6.7 Bubble Charts

The purpose of bubble charts is to show the relationship and compare between multiple measures. It is useful for analysis pattern or correlations of multiple measures data as illustrated in Figure 2.15 [19].

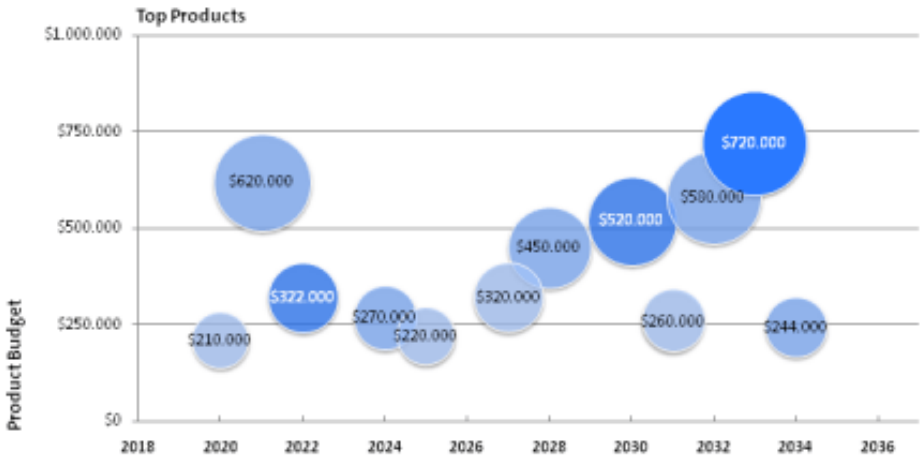


Figure 2.16 Example of a bubble chart.

2.6.8 Mixed Chart

The mixed chart is useful for comparing two measures during the composite chart contains a column chart and a line chart. The chart on the x-axis represents the column chart to measure one. Line chart shows a second measurement by default in the content pane are represented by a column and the second is represented by the y-axis on the left side of the chart shows the values for the column and the y-axis on the right side of the chart shows the values for the line. An example of the pie chart is shown in Figure 2.16[20].

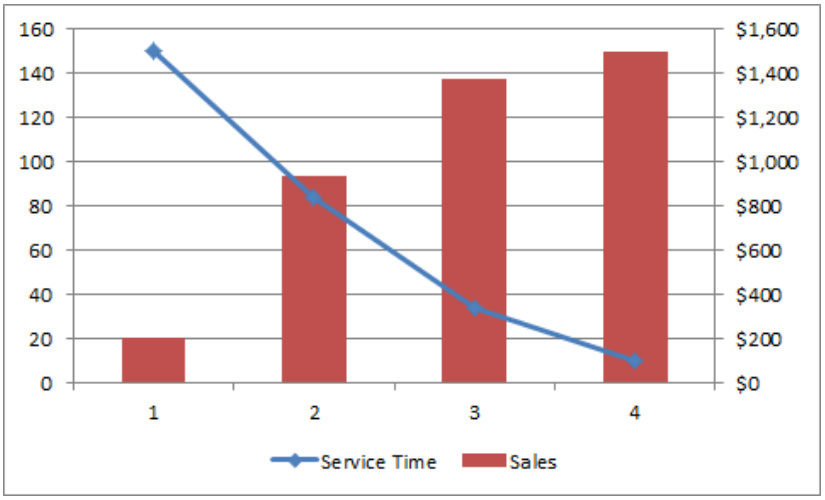


Figure 2.17 Example of the mixed chart.

2.6.9 Tree map

As illustrated in Figure 2.17 [21], the purpose of a treemap is specifying the type of high and low value. It is useful for the shows patterns of the high and the low.

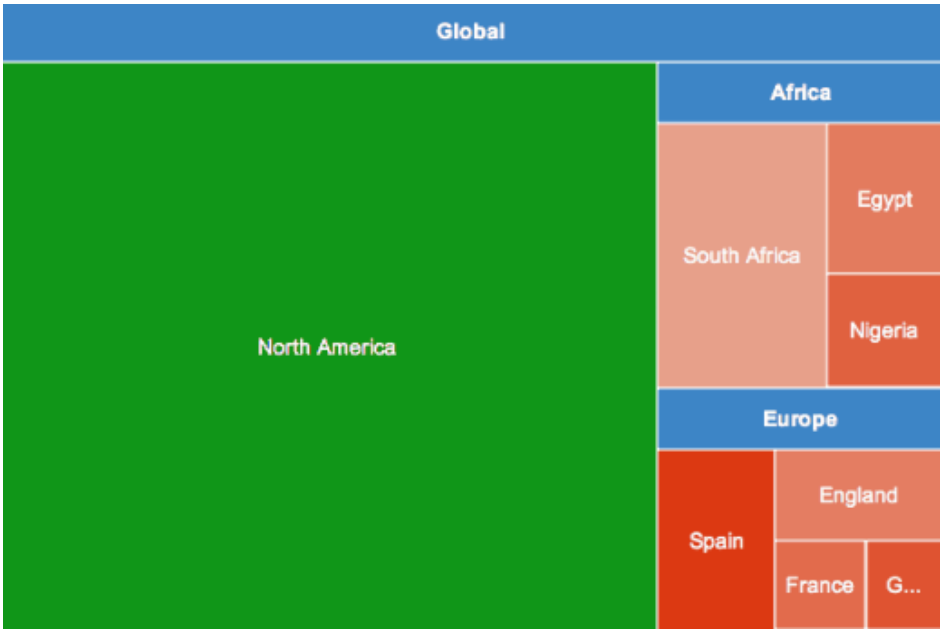


Figure 2.18 Example of the treemap.

In each graph model to take into account the benefits or advantages of each type of graph. In order to display data in graphs effectively. The advantages and features of graph types listed in Table 2.2 below.

Table 2.2 The advantage of each chart.

Type of Chart	The advantage
Bar chart	• Each type of data is displayed in the frequency distribution. • Display the relative number or proportion of different types. • Summarize large data in visual form. • Clarify the trends over a table. • Estimate key values are quickly. • Allow for visual verification of the authenticity and validity of the calculations. • Understandable easily because it is widely used in business and the media.
Column chart	• It shows the absolute numbers, so you do not need to convert percentages into numbers, as well as a pie chart. So the difference is more obvious.
Line chart	• Ability to compare multiple data series continues with ease. • Interim data can be inferred from the graph.
Pie chart	• Display the relative proportions of the various layers of information. • The size of the circle can be made proportional to the total volume it represents. • Summarize large data in visual form. • It will be seen more easily than other types of graphs. • Allow a visual inspection of the validity or accuracy of the calculations. • Require minimal further explanation. • Understandable because it is widely used in business and the media.
Area chart	• Visible more powerful than a line chart.

Table 2.2 The advantage of each chart(cont).

Scatter chart	• Trend in the relationship of the data. • Still, the value of information and the exact size of the sample. • Show minimum / maximum and outliers.
Bubble chart	• Compared with many charts and graphs. • Handy when your worksheet contains three data points on the negative, and multiple data sets.
Mixed chart	• It is very effective when you want to show the two different measures or different for the same variable.
Tree map	• Good catches the eye • The overall trend shows as well.

2.7 Business Intelligence (BI)

A software, technology, and examples supported the data in the database for creating a report in many formats and bring the report to support for decision planning or answer questions to executives by export data and present in report format ex. Bar Graph, Graph chart, Pie Chart, Table, [22] etc.

BI System will forecast business by use the instant data, data exported from Data warehouse and procedure data filter by use export-import, analysis, report support program. BI have a process for prepare data before analysis data below this.

2.7.1 BI Data Prepare Process

- Define data sources divided by internal resources and external resources. Determining resources must take into account the results you want. So that the information brought into line with the desired result.
- The design of the data warehouse (Data warehouse Design).
- Selectively modify data in the right condition and in accordance with the format of the data warehouse design to import the data warehouse by ETL (Extract, transform, load).

- The preparation of the data stored in the data warehouse informs Multidimensional Model or Cube data in various formats prior to creating report formats the tools help to Query Data.

For simply with the audience to data should custom in the report, if data have many data dimension BI dashboard is the answer for easy to understand the data.

2.7.2 BI Dashboard

The reports are obtained from the analysis of the relationship between the creator dashboards to let the audience get the purpose of presenting information that is included on a single page [23]. There are two types of dashboard together with a business role model identifier. The first is Strategic Dashboard, provide for the management to understand the condition of the organization helps to identify opportunities for expansion and improvement of the business. Information in the dashboard updates must be no more than monthly. Each one is Operational Dashboard, a dashboard that monitors real-time operation and alerts the user. The data dashboard is updated constantly the strategic analysis is not used. Dashboard present data in charts and graphs. For describing data and direct with the purpose of information we should know what is a good dashboard, we will discuss in data visualization.

2.8 Related Works

The first paper is researched by Gu Zhaojun, Li Yong, Niu Wenjing [24], in 2010. They research on Firewall Syslog log analysis to use with monitor the security of the network. The goal of the research is establishing a log's analysis system which is implemented by gathering and analyzing firewall logs, thus achieve to monitor the running status of the network. The network uses Cisco PIX firewall for security protect network and collecting Syslog logs by employing the thread pool technique (Programming) in operations research, systems analysis, logs designed. The MySQL database developed by java language with workflow four parts: the hardware firewall, the log gathering preprocessor, the database system and the log analysis. Inquiry engine displayed the data analysis in the TopN Sort (e.g. TopN sort of application, TopN sort of destination IP address and TopN sort of destination port). After import data and the

information analysis by TopN statistics and analyze security event section of traffic analysis, user browsing the website analysis, the most popular host analysis, security event detection and log's severity. In this study, the administrator can determine the status of the current network. In order to improve the performance of firewall with better immediately and improve the methods of statistical analysis and process improvement analysis examining security event to provide more efficiently.

Jian-hua Huang, Man-qi Zhang, Yuan-long Jiang [25], in 2012. They research about building a log (log generated) by the firewall and analysis, security of networks and SHERNET network applications (Shanghai education and research network). The Universities in Shanghai are connected to this network with multi-point connections and a lot of traffic used make congestion problem in network. Data analysis applications and the security system will help to maintain and manage the network to better recording systems and recording formats for networks log. In the research recording format use RFC3164 format. This patterns are recording 3 part as <priority>, <HEADER>, <MSG>. The usage network log collected by Syslog receiver, which is a part key system information was storing in the MySQL database and sent to the real time alarm functions status of the current network. The database was analyzed using TopN statistic analysis and analyze the network security events which detected in the management model of the network variety of device manufacturers. The form of recorded data whicht has led many to helping in the management information by parsing log. Dividing the data into two layers, layer 1 was storing in a database for statistical analysis. The analysis of the data stored in the database. The information will be useful to analyze such a facility, severity, source IP, Destination IP, protocol, destination port and traffic. When the results of the analysis and displays them on tables, line graphs, bar graphs and pie charts in perspective of Source IP, the volume of applications traffic. Data analysis should be improved to make the analysis more efficiently and accuracy. Tthis research can lead to data mining for used to study the behavior of network users as well.

According to the present theory and research related to Chapter 2 in the next chapter, we will talk about all the research process.

CHAPTER III
RESEARCH METHODOLOGY

This chapter discusses the methodology used in our research about BI dashboard creation and data analysis of network data usage of faculty ICT in March to May 2016, with the process works as described in following sections.

3.1 Preliminary Study

The first step is problem study of traffic log, network framework of ICT faculty and description of traffic log in the firewall.

3.1.1 Network Framework

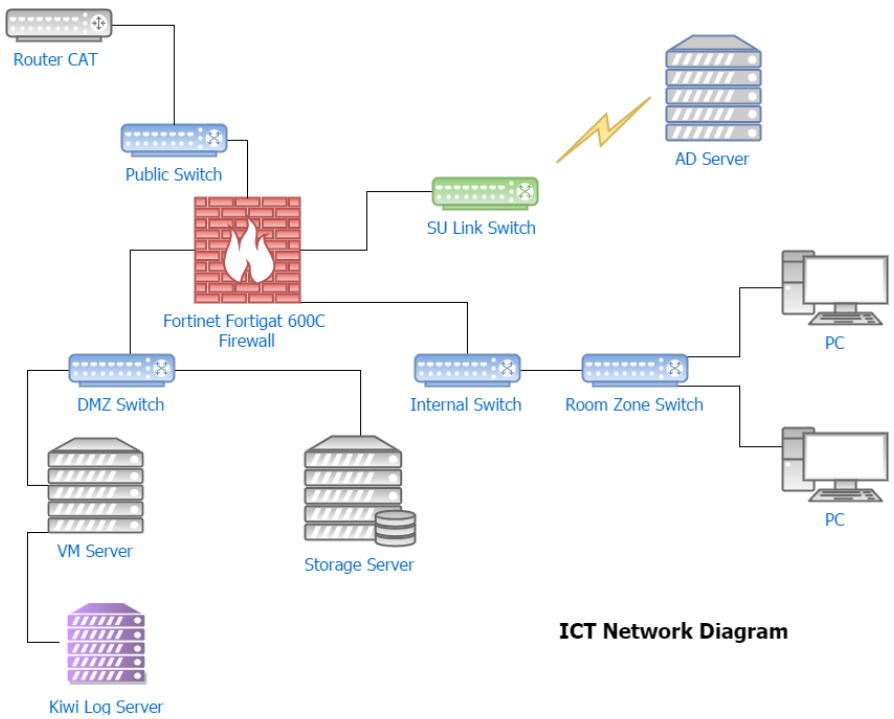


Figure 3.1 Network diagram of ICT faculty.

In our study, the router is Cisco2 8 0 0 , a gateway device which has configured by CAT TELECOM PUBLIC COMPANY LIMITED (CAT). Our network is used internet leased line of CAT in transfer speed 50 Megabits per second and connects to the public switch. The public switch is HP1910G which connected with Cisco 2800 and Fortinet Fortigate 600C firewall, for using any public IP Address. The public switch is between router and firewall. We use Fortinet fortigate 600C firewall for control network routing to any department switch such as office rooms, Lab Computer 1-6, professor rooms. In our network structure have 6 Virtual LAN created by Hp4210G internal switch and provide the packet to another switch. CCSU (Computer Center, Silpakorn University) link switch connected to SU Radius Server for use authentication internet by synchronizing database from CCSU. Hp1810G switch is a De-Militarized Zone (DMZ) use for safe connection server such as DHCP, DNS, etc. Dell PowerEdge R510 is a VM (Virtualization Machine) server contain Management server, DHCP server, DNS server, Kiwi Log server

Hp4210G internal switch creates VLAN and has PC in VLAN following

- VLan 1 IP group 172.16.10.xx Lab 1, PC 108 devices
- VLan 2 IP group 172.16.20.xx Lab2, PC 56 devices
- VLan 3 IP group 172.16.30.xx Lab3&6, PC 110 devices
- VLan 4 IP group 172.16.40.xx Lab 4-5, Ict building, PC 90 devices
- VLan 5 IP group 172.16.50.xx, PC 30 devices
- VLan 6 IP group 172.16.60.xx Office room, PC 20 devices

3.1.2 Logging Network traffic

3.1.2.1 FortiGate 600C Firewall

In recording log messages recording in RAW format have two part like HAED and BODY [26]

- **Header** is containing the date and time the log originated, log identifier, message identifier, administrative domain (ADOM), the log category, severity level, and where the log originated. These fields are common to all log types.
- **Body** is describing the reason why the log was created and actions taken by the FortiGate device to address it. These fields are varying by log type.

The records unit is done in order to record a sequence of actions below.

- The log has scanned when a packet comes to inbound or outbound interface of the firewall.
- During log scanning process, the fortigate unit operate compare policy route at the same time records the action and results of log scanned.
- Log messages (the action and results of log scanned) are sent to Kiwi Syslog server (or log device).

The traffic log from fortigate 600C has many types. Our work chooses traffic forward the message for analysis data in message log [27]. However, we should make an understanding with log meaning in the message.

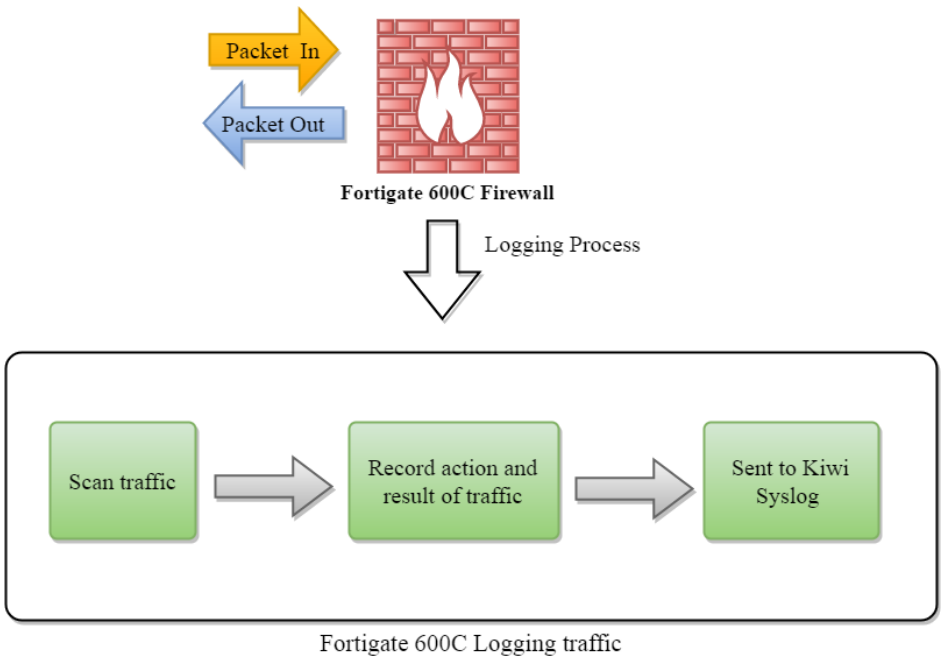


Figure 3.2 FortiGate firewall logging process.

3.1.2.2 Log Definition

Log Definition is the description or information of log type. It is describing log activities used for statistic and monitor purpose. We will explain traffic forward message of FortiGate Firewall [28] in Table 3.1.

Table 3.1 Name and definition of log fields.

Log field	Meaning
Type	Traffic
Subtype	Forward
Level	Notice
Date	The date at which the log was recorded.
Time	The time at which the log was recorded.
Status	The state of the traffic that is identified by the policy.
Vd	The virtual domain in which the logging occurred. If VDOMs are not configured, this will display "root".
Trandisp	Whether the packet is source NAT translated (*snat*) or destination NAT translated (*dnat*), both (*snat+dnat*) or neither (*noop*).
Srcip	The source of machine IP Address.
Srcname	The name of the device.
Srcport	The source port of the Transmitsion Control Protocol or User Datagram Protocol traffic.
Dstip	The destination Internet Protocol.
Dstname	The destination name, this can be a name or an Internet Protocol Address.
Dstcountry	Destination of the country that packet was sent to.
Srccountry	The source of the country that packet was sent to.
Dstport	The destination port number of the Transmitsion Control Protocol or User Datagram Protocol traffic. The destination port appears as zero for other types of traffic.
Tranip	The translated Internet Protocol in Network Address Translation mode. For Transparent mode, it is zero.
Transport	The translated port number in Network Address Translation mode. For Transparent mode, it is zero.
Transip	The translated source Internet Protocol in Network Address Translation mode. For Transparent mode, it is zero.

Table 3.1 Name and definition of log fields (Cont.)

Transport	The translated source port number in Network Address Translation mode. For Transparent mode, it is zero.
Service	The service of traffic where the event occurred.
Proto	The protocol number which applies to the session or packet.
Duration	Time of session works until end in seconds.
Policyid	The ID number of the firewall policy that applies to the session or packet.
Custom	The field was built to show data traffic by setting up their own.
Indentidx	ID policy identity, this data is zero if the firewall policy is not a policy-based identity; Otherwise, it will show the number of identity-based policies that match the traffic. This number is unique worldwide, it is only unique in the firewall policy defined.
Sentbyte	The number of sent bytes.
Rcvdbyte	The number of received bytes related to the log message.
Shaperdropsentbyte	Shaper dropped sent bytes.
Shaperdroprcvdbyte	Shaper dropped received bytes.
shaperperipdropbyte	Per IP dropped bytes.
Shapersentname	The name of the traffic shaper sending the bytes.
Shaperrcvdname	The name of the traffic shaper receiving the bytes.
Shaperperipname	The Per IP shaper name.
Sentpkt	The number of sent packets related to the log message.
Rcvdpkt	The number of received packets related to the log message.
Vpn	The name of the Virtual Private Network tunnel used by the traffic.
Vpntype	The type of Virtual Private Network tunnels that the traffic is flowing through.
Vpntunnel	The name of the Virtual Private Network tunnel that was used.

Table 3.1 Name and definition of log fields (Cont.)

Srcintf	The source of network interface. For outgoing traffic originating from the firewall.
Dstintf	The destination of network interfaces on fortigate firewall.
Sessionid	ID number of Session.
Appid	ID number of Application.
App	The name of the application that triggered the action within the control list.
Appcat	The application category that the application is associated with.
Applist	The name of the applications.
Appact	Policies action of an application.
User	Username.
Group	The group name.
Osname	Name of the device Operating System.
OSversion	The version number of the device Operating System.
Unauthuser	Unauthenticated username.
Unauthusersource	Method used to detect username.
Uttmaction	The Unified Threat Management action was taken by the system.
Filename	The name of the file that was transferred.
Virus	The name of the virus detected.
Attack	The action of the virus act.
Hostname	The hostname information.
Catdesc	The category description.
Sender	Sender or dispatcher.
Recipient	Receiver.
Mailcount	Mail number.
Spamcount	Spam number.
Dlprule	Data Loss Prevention rule.
Utmevent	The type of Unified Threat Management event taking place.

Table 3.1 Name and definition of log fields (Cont.)

Utmseverity	Unified Threat Management severity.
sha256	SHA256 hash.
Analyticssubmit	Whether analytics were submitted or not. Can be false or true.
Crscore	Client Reputation score.
Craction	Client Reputation action.

3.1.3 Kiwi Syslog Server

VM server (Dell PowerEdge R510) was created a Windows Server 2012 standard 64 bit OS and install Kiwi Syslog server in Windows Server. After that, we are configuring Kiwi log software to receiving log messages from Fortigate 600C and save to log files. Log recording process has two sessions for stored files, one is the original log file which is in text file format. Another one is a copy log files and an encryption file for protected editing and modify data in the log file. Log file storing are kept in 90 days for accuracy according to CCA 2007. The traffic data is the period from March to May 2016 for analysis, as captured in Figure 3.3.

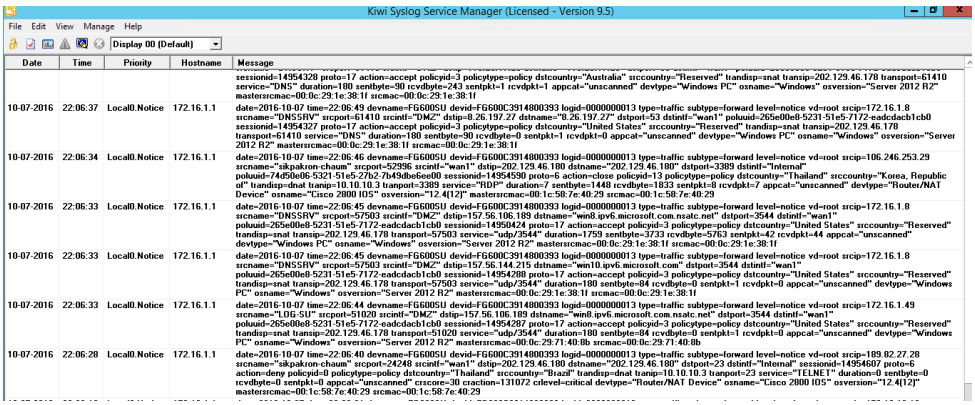


Figure 3.3 Kiwi Syslog service in windows server.

3.2 Requirement Analysis

In this process is collect information from Deputy Dean of IT ICT faculty SU, regarding the results of the analysis show that traffic log data was stored, to analyze

internet traffic and network resources to provide appropriate and adequate needs of users in the network. The network congestion problems increased over time, the information provides for the traffic usage analyzed and presented in the form of various reports as the network bandwidth over time, a group of application on the network and a group of the network using the traffic. The management recognizes displaying report the preliminary results show the following:

- The information required making the reader display the plenty of data.
- Be able to demonstrate the network usage.
- The reader can view the data in the overview.

3.3 Data Collection

This process we discussed how to collect the internet traffic for analyzing the traffic in the faculty of Information and Communication Technology Silpakorn University Phetchaburi ICT Campus between March to May 2016. Which collected traffic and stored by a firewall. It is stored in Raw log format. The firewall should not store data in them memory. If the firewall was setup to keeps log in them memory, it makes memory full quickly. So, the traffic was stored in the storage server by using software recorded data named SolarWinds Kiwi Syslog Server. Kiwi log server stored traffic in two files, each one is a normal log file and each one is encryption log file. All the day, the network traffic is saved in a file in Kiwi Syslog format. Next process is a process for the data separates. The separate of the network traffic contained in the message into data subsets. The information of network traffic, the log has kept many types. When analyzing the data into the statistical program, it making query data difficulty. With the problem of the log are forced by SolarWind Log Management and analysis is limited by a trail software, so we created a program for parsing log on each field. Each filter will be use as Tab for separated variables and the value is after equals mark (=). The data from a parsing program are stored in MySQL database server.

3.4 Network Log Data Structure

From section 3.3 we get log file. text from Kiwi log server and log files are in RFC 3164 format. Log files has three part in a message like **<PRI> HEADER MESSAGE.** We will explain each part following [29].

3.4.1 PRI

This field is containing two fields called the Syslog Facility and the Syslog Severity Levels The priority value ranges from 0-6. They have the list of facility available, the list of Severity Level and detailed explanation of the severity Levels are below.

Table 3.2 Priority Levels.

Level (0 is highest)	Name	Description
0	Emergency	The system is unusable or not responding.
1	Alert	Immediate action required. Used in security logs.
2	Critical	Functionality is affected.
3	Error	An error exists and functionality could be affected.
4	Warning	Functionality could be affected.
5	Notification	Information about normal events.
6	Information	General information about system operations. Used in event logs to record configuration changes.

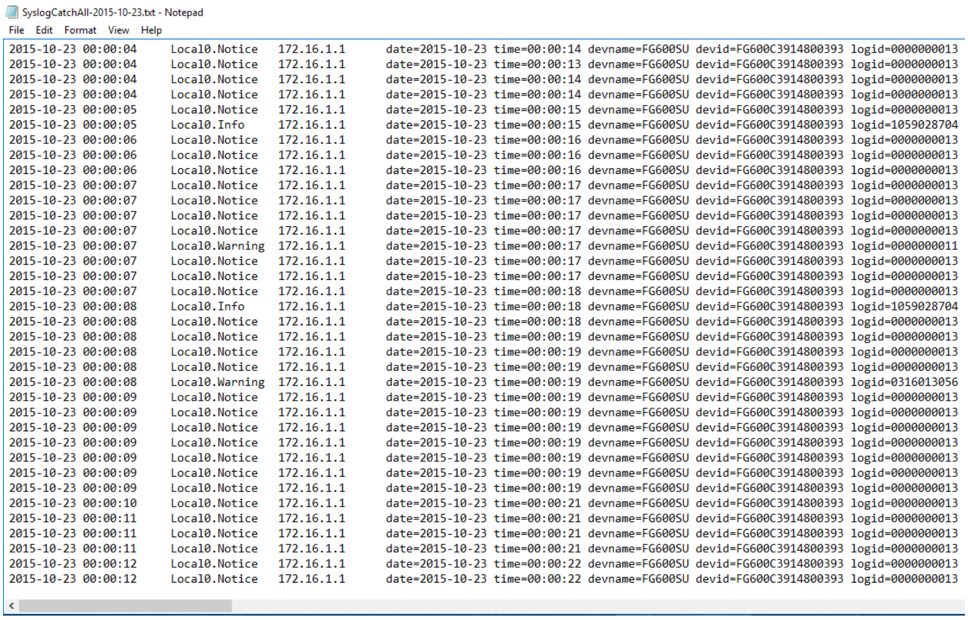
3.4.2 Header

This field is containing two fields called the TIMESTAMP and the HOSTNAME. The TIMESTAMP is the local time in the format of [mm dd hh:mm:ss] .The HOSTNAME is the indication of hostname or IP address of host device. If it does not have a hostname, then it will contain own IP Address of hostname.

3.4.3 Message

This field is containing the text of Syslog messages, along with some information of the packet in the network. The Syslog messages generated by Fortigate 600C in the following format.

Format: Date (MM-DD-YYYY) [TAB] Time (HH:MM: SS) [TAB] Priority (Facility. Level) [TAB] Hostname [TAB] Message text



```
SyslogCatchAll-2015-10-23.txt - Notepad
File  Edit  Format  View  Help
2015-10-23 00:00:04    Local0.Notice   172.16.1.1     date=2015-10-23 time=00:00:14 devname=FG600SU devid=FG600C3914800393 logid=0000000013
2015-10-23 00:00:04    Local0.Notice   172.16.1.1     date=2015-10-23 time=00:00:13 devname=FG600SU devid=FG600C3914800393 logid=0000000013
2015-10-23 00:00:04    Local0.Notice   172.16.1.1     date=2015-10-23 time=00:00:14 devname=FG600SU devid=FG600C3914800393 logid=0000000013
2015-10-23 00:00:04    Local0.Notice   172.16.1.1     date=2015-10-23 time=00:00:14 devname=FG600SU devid=FG600C3914800393 logid=0000000013
2015-10-23 00:00:05    Local0.Notice   172.16.1.1     date=2015-10-23 time=00:00:15 devname=FG600SU devid=FG600C3914800393 logid=0000000013
2015-10-23 00:00:05    Local0.Info     172.16.1.1     date=2015-10-23 time=00:00:15 devname=FG600SU devid=FG600C3914800393 logid=1059028704
2015-10-23 00:00:06    Local0.Notice   172.16.1.1     date=2015-10-23 time=00:00:16 devname=FG600SU devid=FG600C3914800393 logid=0000000013
2015-10-23 00:00:06    Local0.Notice   172.16.1.1     date=2015-10-23 time=00:00:16 devname=FG600SU devid=FG600C3914800393 logid=0000000013
2015-10-23 00:00:06    Local0.Notice   172.16.1.1     date=2015-10-23 time=00:00:16 devname=FG600SU devid=FG600C3914800393 logid=0000000013
2015-10-23 00:00:07    Local0.Notice   172.16.1.1     date=2015-10-23 time=00:00:17 devname=FG600SU devid=FG600C3914800393 logid=0000000013
2015-10-23 00:00:07    Local0.Notice   172.16.1.1     date=2015-10-23 time=00:00:17 devname=FG600SU devid=FG600C3914800393 logid=0000000013
2015-10-23 00:00:07    Local0.Notice   172.16.1.1     date=2015-10-23 time=00:00:17 devname=FG600SU devid=FG600C3914800393 logid=0000000013
2015-10-23 00:00:07    Local0.Notice   172.16.1.1     date=2015-10-23 time=00:00:17 devname=FG600SU devid=FG600C3914800393 logid=0000000013
2015-10-23 00:00:07    Local0.Warning  172.16.1.1     date=2015-10-23 time=00:00:17 devname=FG600SU devid=FG600C3914800393 logid=0000000011
2015-10-23 00:00:07    Local0.Notice   172.16.1.1     date=2015-10-23 time=00:00:17 devname=FG600SU devid=FG600C3914800393 logid=0000000013
2015-10-23 00:00:07    Local0.Notice   172.16.1.1     date=2015-10-23 time=00:00:17 devname=FG600SU devid=FG600C3914800393 logid=0000000013
2015-10-23 00:00:07    Local0.Notice   172.16.1.1     date=2015-10-23 time=00:00:18 devname=FG600SU devid=FG600C3914800393 logid=0000000013
2015-10-23 00:00:08    Local0.Info     172.16.1.1     date=2015-10-23 time=00:00:18 devname=FG600SU devid=FG600C3914800393 logid=1059028704
2015-10-23 00:00:08    Local0.Notice   172.16.1.1     date=2015-10-23 time=00:00:18 devname=FG600SU devid=FG600C3914800393 logid=0000000013
2015-10-23 00:00:08    Local0.Notice   172.16.1.1     date=2015-10-23 time=00:00:19 devname=FG600SU devid=FG600C3914800393 logid=0000000013
2015-10-23 00:00:08    Local0.Notice   172.16.1.1     date=2015-10-23 time=00:00:19 devname=FG600SU devid=FG600C3914800393 logid=0000000013
2015-10-23 00:00:08    Local0.Notice   172.16.1.1     date=2015-10-23 time=00:00:19 devname=FG600SU devid=FG600C3914800393 logid=0000000013
2015-10-23 00:00:08    Local0.Warning  172.16.1.1     date=2015-10-23 time=00:00:19 devname=FG600SU devid=FG600C3914800393 logid=0316013056
2015-10-23 00:00:09    Local0.Notice   172.16.1.1     date=2015-10-23 time=00:00:19 devname=FG600SU devid=FG600C3914800393 logid=0000000013
2015-10-23 00:00:09    Local0.Notice   172.16.1.1     date=2015-10-23 time=00:00:19 devname=FG600SU devid=FG600C3914800393 logid=0000000013
2015-10-23 00:00:09    Local0.Notice   172.16.1.1     date=2015-10-23 time=00:00:19 devname=FG600SU devid=FG600C3914800393 logid=0000000013
2015-10-23 00:00:09    Local0.Notice   172.16.1.1     date=2015-10-23 time=00:00:19 devname=FG600SU devid=FG600C3914800393 logid=0000000013
2015-10-23 00:00:09    Local0.Notice   172.16.1.1     date=2015-10-23 time=00:00:19 devname=FG600SU devid=FG600C3914800393 logid=0000000013
2015-10-23 00:00:09    Local0.Notice   172.16.1.1     date=2015-10-23 time=00:00:19 devname=FG600SU devid=FG600C3914800393 logid=0000000013
2015-10-23 00:00:09    Local0.Notice   172.16.1.1     date=2015-10-23 time=00:00:19 devname=FG600SU devid=FG600C3914800393 logid=0000000013
2015-10-23 00:00:10    Local0.Notice   172.16.1.1     date=2015-10-23 time=00:00:21 devname=FG600SU devid=FG600C3914800393 logid=0000000013
2015-10-23 00:00:11    Local0.Notice   172.16.1.1     date=2015-10-23 time=00:00:21 devname=FG600SU devid=FG600C3914800393 logid=0000000013
2015-10-23 00:00:11    Local0.Notice   172.16.1.1     date=2015-10-23 time=00:00:21 devname=FG600SU devid=FG600C3914800393 logid=0000000013
2015-10-23 00:00:11    Local0.Notice   172.16.1.1     date=2015-10-23 time=00:00:21 devname=FG600SU devid=FG600C3914800393 logid=0000000013
2015-10-23 00:00:12    Local0.Notice   172.16.1.1     date=2015-10-23 time=00:00:22 devname=FG600SU devid=FG600C3914800393 logid=0000000013
2015-10-23 00:00:12    Local0.Notice   172.16.1.1     date=2015-10-23 time=00:00:22 devname=FG600SU devid=FG600C3914800393 logid=0000000013
```

Figure 3.4 Example of Kiwi Syslog format.

3.4.4 Log Definition

From 3.4 network log data structure, we obtain data from Fortigate firewall and stored by Kiwi log server. So that, the network traffic definition is almost like as Fortigate firewall log definition. The log file contains information of multiple network applications. In order to understand and explanation the network traffic correctly, the various types of network traffic variable are displayed meaning of the traffic log as the Table 3.1.

3.5 Design of BI Dashboard

The dashboard in the design process created a dashboard for display correctly according the objectives and indicators, which are the three main steps below [30].

3.5.1 Foundation

To identify the target audience, understand what type of dashboard and valuable to the organization by use three key question. First question “Who is your audience?” the second question is “What value will the dashboard add?” and the third question is “What type of dashboard am I creating?” .If you can answer three question, you will get details of dashboard design. It brings the details for use in part 2, we will discuss the form and structure of dashboard.

3.5.2 Structure

This part offers ideas about the big-picture elements of dashboard-building blocks that will use to construct the dashboard. They can be broken into four categories.

- **Form**

The format of the dashboard that we want to deliver to the audience. The form has a factor that influences the dashboard form such as timeliness, aesthetic value, mobility, connectivity, data detail, data density, interactivity, and collaboration.

- **Structure**

The structure has grid overlay or layout that helps frame the content of dashboard to understand the picture and to fit a piece picture. A dashboard structure requires a deep understanding of how the system is measuring works. There is three model to tell the story of the dashboard: flow, relationships, and grouping.

- **Flow**

A sequence of events or actions across time.

- **Relationships**

The relationships between entities or measures. The relationships may be mathematical, geographical, organizational or functional.

- **Grouping**

The structure is group related information into categories or a hierarchy.

- **Design principles**

To design dashboard, it has a key design principle that uses when design dashboard: compactness/modularity, gradual reveal, guide attention, supports casual use, leads to action, customizable and explanation before information.

- **Functionality**

The common features that can make dashboard useful. It is having a few basics features such as drill down, filter, comparison, alert and export/print. The advanced features are a text-based summary, starring/tagging, annotation, save/track changes, advanced visualizations.

3.5.3 Information design

The information design part is guided tips for putting information on the page that effectively to the audience. For designing a clear, we concentrate on the chart, table, and visualizations that communication information by look to interface design and information display.

- **Interface design**

Simplicity is a primary goal of well-designed websites as the same dashboard. To create dashboard interface we can learn about organizing the page, color, and typography.

- **Information display**

In the dashboard, we want to create charts and tables that highlight the right information and easy to read, we should be understanding the data types and chart types to display the right information.

Overlay of traffic log in the network is typically used to display the information mostly by use TopN statistic in the display, such as TopN source, TopN IP address, TopN destination which results. Output can be no novelty or variety when we took more TopN at the time or in the information display to see a deeper than ever as TopN source by time, TopN session by time and. destination country.

In this study, use data networks to analyze and present data in a dashboard. The dashboard contains a report by presenting data which shows the graph below.

- Slicer report
- Stack bar chart
- Line chart
- Pies chart
- Sankey diagram
- Word cloud chart
- Map chart

3.6 Business Intelligence Development

After the design of BI dashboard for the traffic analysis, we used the statistical application for analyze network traffic and presenting on the dashboard. We use Power BI software, a statistical software of Microsoft Inc. which we used in this work. In this process we are working two step, the first is a query, analyze data and create a report from traffic analysis by using the Power BI Desktop. The second, we published files from Power BI Desktop to cloud storage of Microsoft which use to create dashboard as figure 3.5.

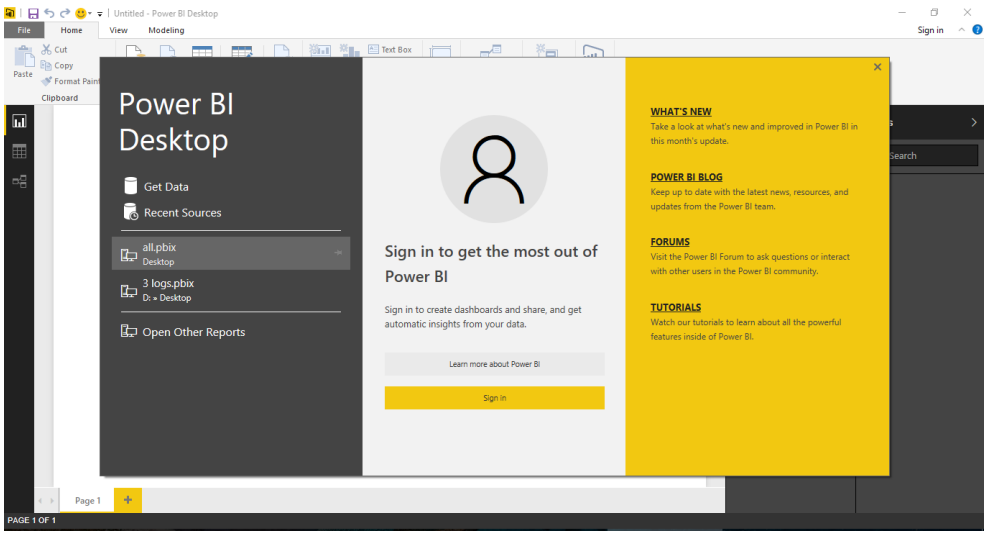


Figure 3.5 Power BI Desktop application.

3.7 Evaluation Form Designing

The finally, we would evaluate the dashboard which we created by using websites design metric for evaluating to the audiences, Executives IT departments or Dean of the Faculty of Information Communications and Technology. Which below:

Table 3.3 Evaluation Matrices of BI Dashboard

Gender	□ Male	□ Female			
Educational degree	□ Under BS	□ BS	□ MS or higher education		
Seniority	□ 25 to less than 30 years		□ 31 to less than 35 years		
	□ 36 to less than 40 years		□ 41 to less than 45 years		
	□ 46 to less than 50 years				
Below are the attitude and performance of BI Dashboards for network dashboard.					
	Very low	Low	Moderate	Good	Excellent
1. The accuracy of processing					
2. Appropriate to group reports					
3. Presentation format the report is accurate and complete					
4. Deployment type of data					
5. There is easy access to the report					
6. Responding to the report when choose materials					
7. It is convenient to download the data					
8. Report design is beautiful					
9. Can actually be implemented					

3.7 Research Timeline

In this process, we are showing a research process timeline from preliminary study process to policy planning process.

Table 3.4 Research timeline table

	2016									
	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Preliminary Study										
Requirement Analysis										
Data Collection										
Network Log Data Structure										
Design of BI Dashboard										
Policy Planning										
Evaluation Designing										
Documentation										

CHAPTER IV
RESULTS AND DISCUSSION

This chapter discusses the results from the bi dashboard which analyze network traffic data and presents information in the form of various reports. With the data analysis, the researcher studies theoretical related to network traffic analysis and presenting information in a report. The results will be presented in two parts.

- The first part is a presentation of the network traffic during March to May 2016.
- The second part is dashboard evaluated by Deputy Dean of IT ICT faculty SU.

4.1 Result of Traffic Analysis Dashboard

This section, we will present the results from the dashboard creation and present an analysis of the network traffic, which will be presented in the volume of applications in various categories during from March to May 2016. The result of network traffic analysis uses for decisions on IT development planning. The results are displayed in the different perspectives**.** From network traffic analysis, we will show the reports as the figure 4.1.

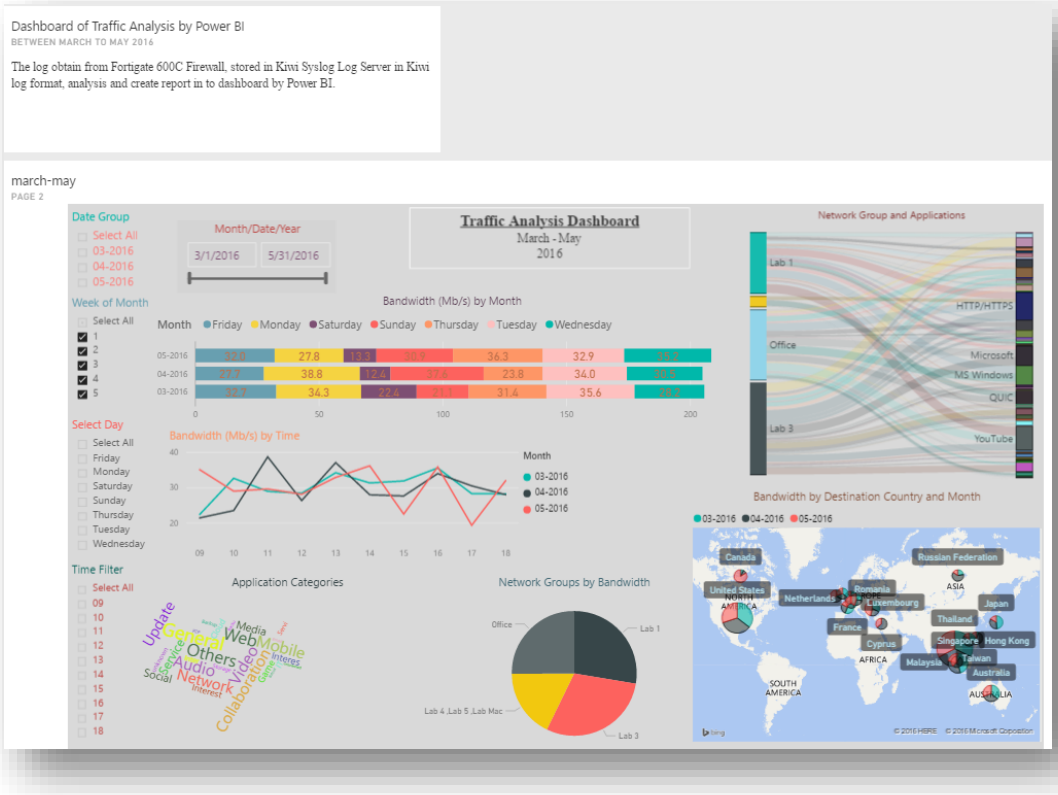


Figure 4.1 A Dashboard of Traffic Analysis.

4.1.1 Introduction on Reports

In the dashboard are presenting reports. Before summarizing the result of network traffic analysis we will explain the report below:

- **Slicer Report**

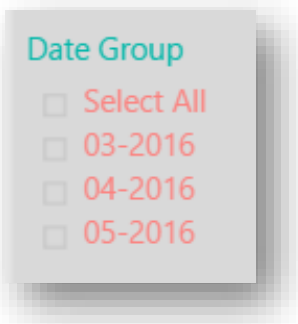


Figure 4.2 A slicer for select Month.

As the figure 4.2 shows the report, which uses selecting groups data to shown in dashboard such as 03-2016 is a group data of date in March, 04-2016 is a group data of date in April and 05-2016 is a group data of date in May.

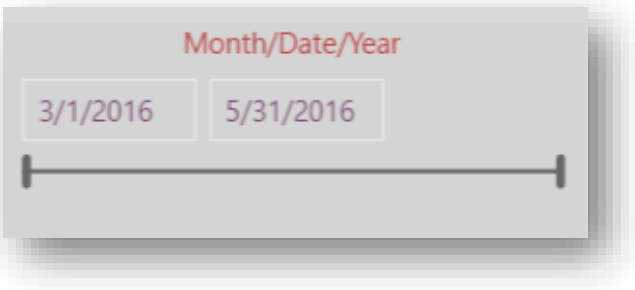


Figure 4.3 A slicer to filter date range.

Figure 4.3 is a slicer which uses for data filtering in date scope. This report use by fill date in the text box, pick a date from the calendar, or slider line to filter the date.

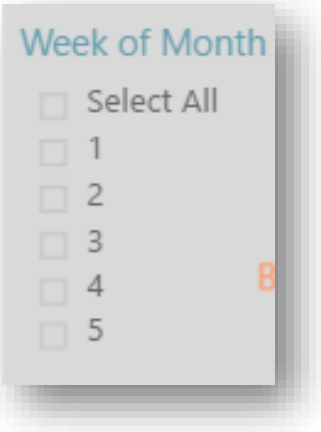


Figure 4.4 A slicer for select a week of the month.

As a captured in figure 4.4 is a slicer use for filter the week number of the month by select the check box (e.g. the first, second, third, fourth and fifth) the week of the month. This can select a single option or select all.

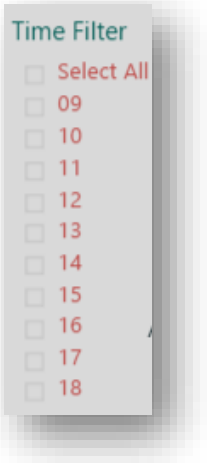


Figure 4.5 A slicer for select time.

Figure 4.5 is a slicer for select time group. In this topic, we grouped time from 9.00 to 18.00, during the time of network usage. It can select one or more to filter time.

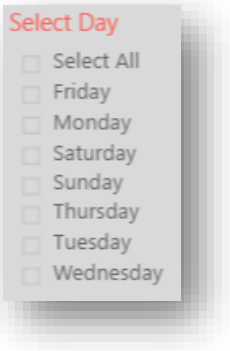


Figure 4.6 A slicer to filter day name.

A slicer as figure 4.6 used to data filtering by show as a day name such as Sunday, Monday, Tuesday, Wednesday, Thursday, Friday, Saturday or select all day. It can select by use check box and can select one or more option as needed.

- **Stacked Bar Chart**

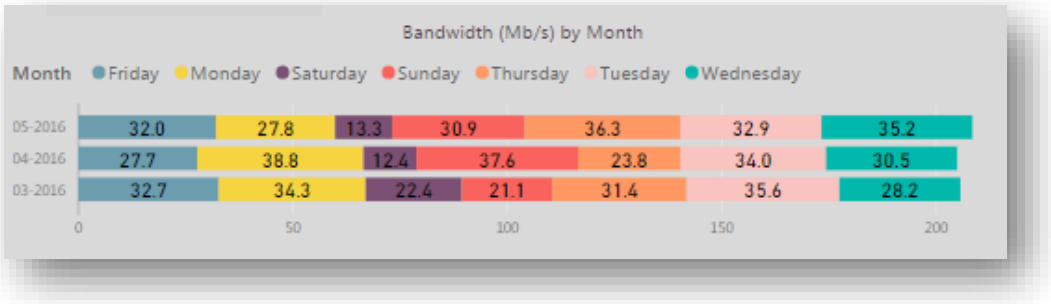


Figure 4.7 Bandwidth by day name in bar chart.

As figure 4.7 is a stacked bar chart, this shows the bandwidth by day name and information show in colors (e.g. blue is Friday, yellow is Monday, purple is Saturday, red is Sunday, orange is Thursday, pink is Tuesday and green is Wednesday).

- **Sankey Diagram**

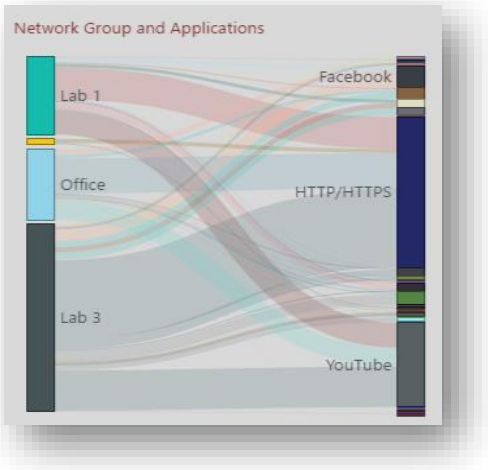


Figure 4.8 Sankey diagram showing network groups and applications by bandwidth.

In figure 4.8 is a diagram showing network groups and applications by bandwidth volume. There is show link with network group and applications, a link width Indicate the bandwidth volume of applications.

- **Word Cloud Chart**

Figure 4.9 Word cloud chart show application categories.

As a captured in figure 4.9, Word Cloud chart shows the text of application categories and size of texts Indicate the bandwidth volume. The bandwidth volume is remodeled by the variables filter selection such as day, time, day name, network group, etc.

- **Pie Chart**

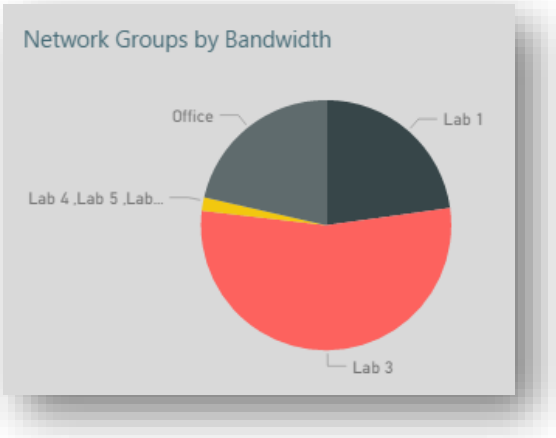


Figure 4.10 Pie chart show network groups by bandwidth.

The figure 4.10 is a pie chart, which displays network groups by bandwidth volume. The size of network groups is changed by variable from values filter.

The chart is consists of a network group such a Lab 1, Lab 2, Lab 3, Lab 4 / Lab 5 / Lab Mac and Office.

- Lab 1 is a network group of VLAN 10 have source IP Address 172.16.10.2 - 172.16.10.254 .
- Lab 2 is a network group of VLAN 20 have source IP Address 172.16.20.2 - 172.16.20.254 .
- Lab 3 is a network group of VLAN 30 have source IP Address 172.16.30.2 - 172.16.30.254 .
- Lab 4 / Lab 5 / Lab Mac is a network group of VLAN 40 have source IP Address 172.16.40.2 - 172.16.40.254 .
- The office is a network group of VLAN 60 have to source IP Address 172.16.60.2 - 172.16.60.254 .

- **Map Chart**

Figure 4.11 Map chart showing bandwidth by destination country and month.

As a figure 4.11 is a map chart, which shows the bandwidth by destination country and month. There is show in a group of March, April and May. The chart show values of the bandwidth in circle and circle is consist of the month as in color (e.g. green is March, gray is April, and red is May). Circle in the chart is the volume of bandwidth in each month and country, it can resize by variable filtering.

4.1.2 Summary Report of Data Analyze

After the introduction on the work of each report. Following is a summary of the results from the analysis presented in the report, and each type can be summed up in a variety of different perspectives. The analysis of data from the network, from March to May 2016, as follows.

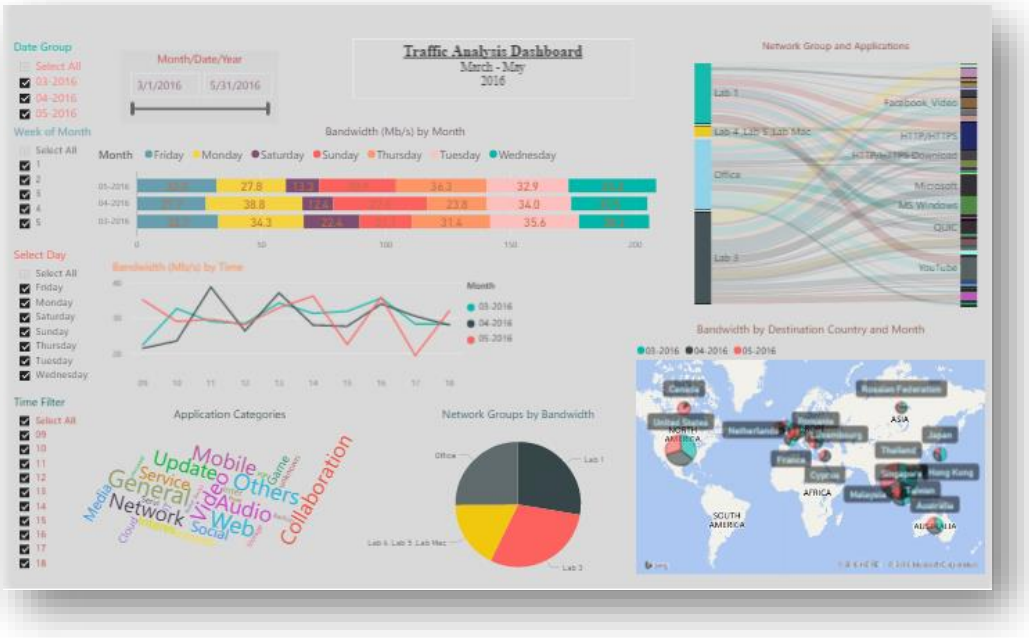


Figure 4.12 Traffic analysis dashboard from March – May 2016.

- **Summary Report of Bandwidth by Day Name**

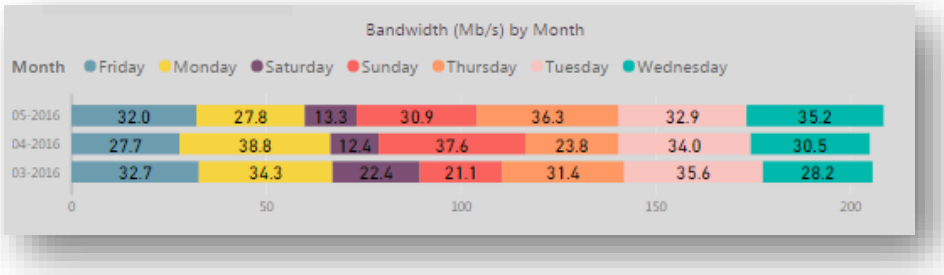


Figure 4.13 Bandwidth by Month and Day Name.

As a figure 4.12 is a traffic analysis dashboard from March – May 2016, which presented a report indicating the data has been analyzed. When viewed on a monthly basis to find the amount of network usage for each month. The report was comparison the data different in the during Monday - Sunday. This report will tell us the following in March, Tuesday is the network usage highest speed at 35.6 Mb/s. In April, the highest active on Monday at 38.8 Mb/s. In May, the most active on Thursday at a speed of 36.3 Mb/s as the figure 4.13. The average of the three months of the internet traffic highest is on Tuesday and the second is Wednesday.

- **Summary Report of Bandwidth by Time**

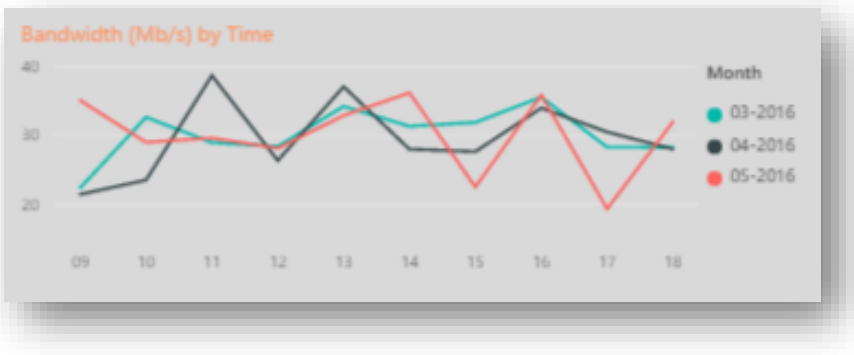


Figure 4.14 Bandwidth by time.

As the figure 4.14, the report of the bandwidth by time tell us following, at 16.00 in March with the network usage most at 35.63 Mb/s. In April, the highest active time at 11.00 the network usage at 38.79 Mb/s. In May, the time at 14.00 is highest active network usage at 35.92 Mb/s.

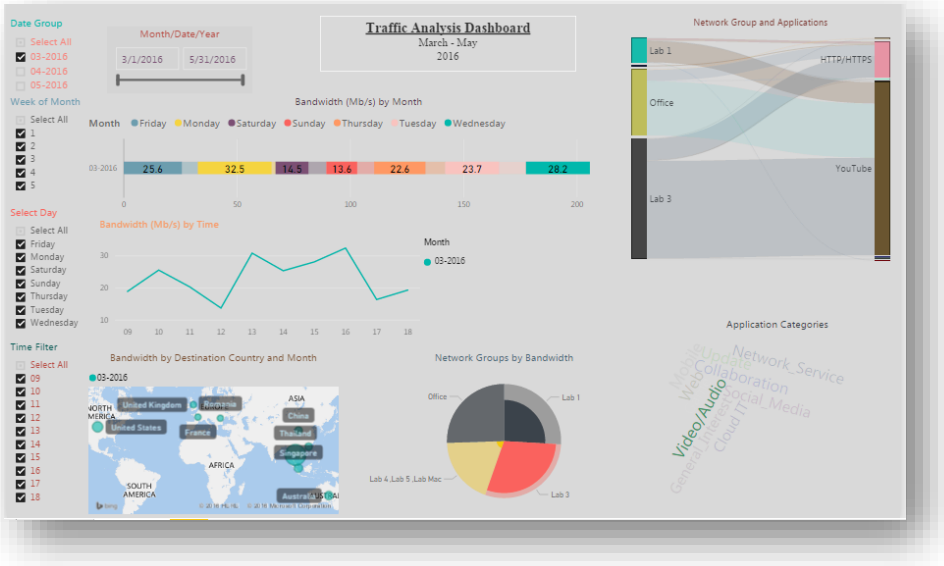


Figure 4.15 Filter report by Application Categories in March.

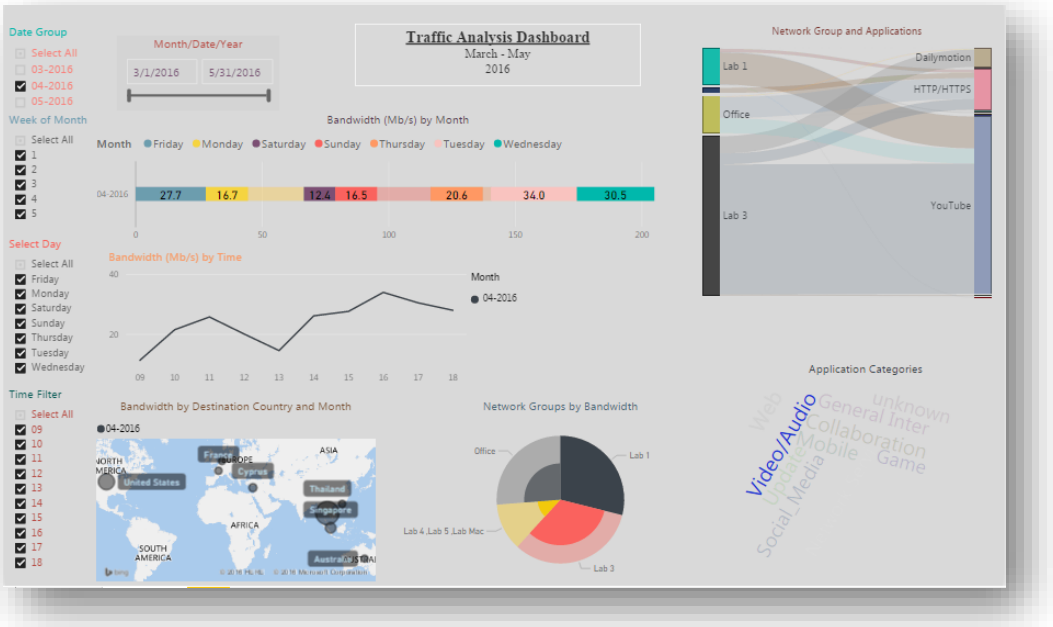


Figure 4.16 Filter report by Application Categories in April.

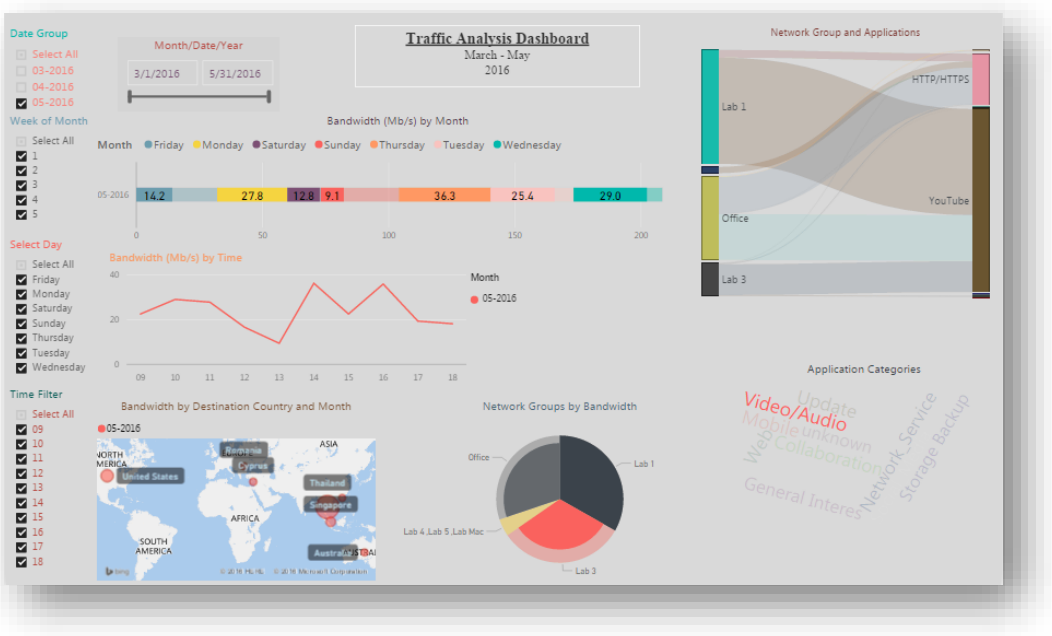


Figure 4.17 Filter report by Application Categories in May.

According to the network usage as the figure 4.15, figure 4.16, and 4.17. Graph of network usage has risen every month around 16.00. A similar manner not usual can explain to the cause of break-out sessions or the lessons allow student using the internet.

- **Summary Report of Bandwidth by Network Group**

The amount of network usage by specific groups of network usage segments are described separately in each of the months as figure 4.18, figure 4.19 and figure 4.20 following.

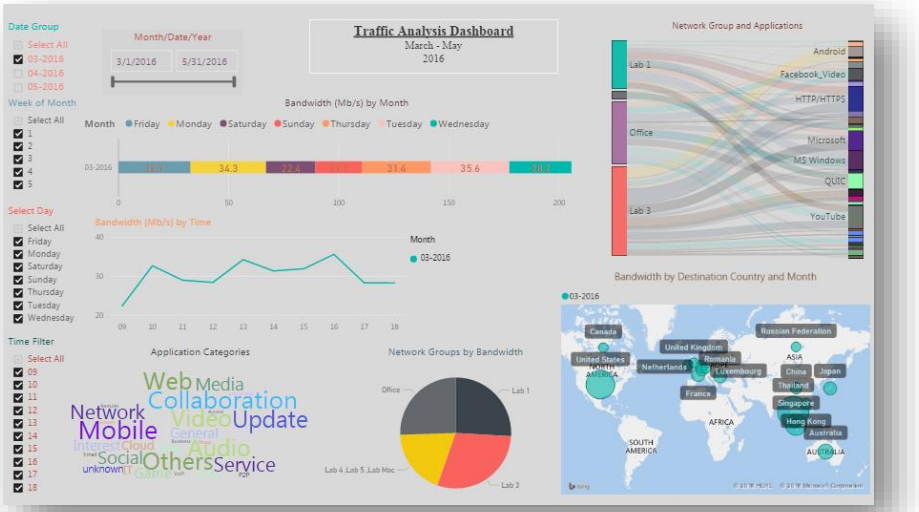


Figure 4.18 Dashboard of March.

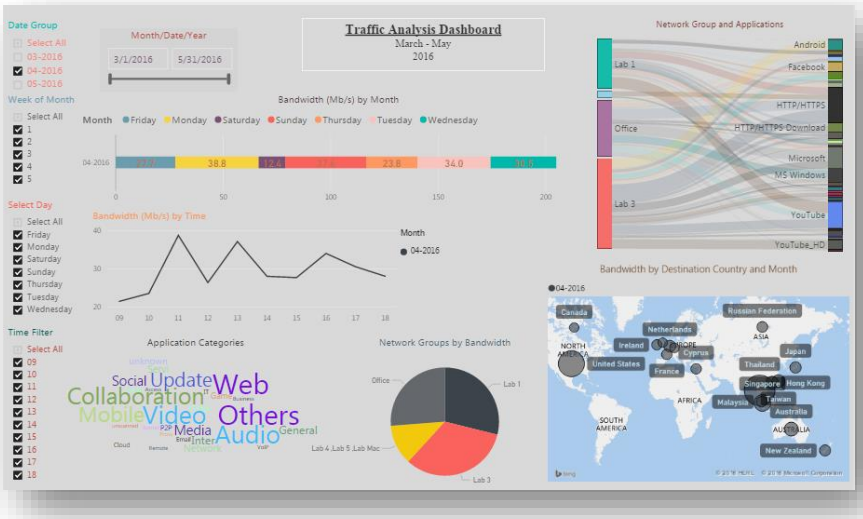


Figure 4.19 Dashboard of April.

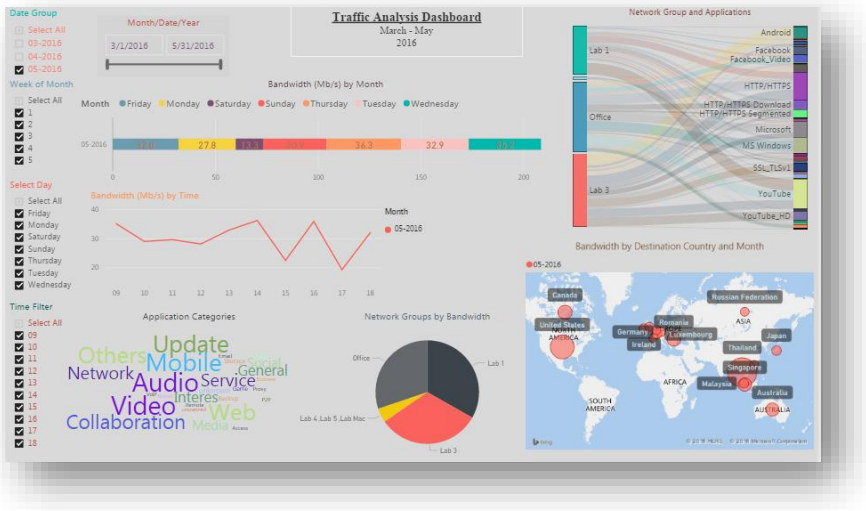


Figure 4.20 Dashboard of May.

As figure 4.18, figure 4.19 and figure 4.20, the network group which used traffic most in March is Lab 3, Lab 1, Office and Lab 4/ Lab 5/ Lab Mac.

In April the network group with the highest usage is Lab 3, Lab 1, Office and Lab 4/ Lab 5/ Lab Mac.

In May the highest usage network group is Lab 1, Lab 3, Office and Lab 4/ Lab 5/ Lab Mac.

Lab 3 is highest active the network usage, which have the demand of networks. It makes the network service most important as well and minor is Lab 1.

- **Summary Report of Bandwidth by Destination Country**

As figure 4.18, figure 4.19 and figure 4.20. Almost 5 countries, where user connected moistest. The first is Thailand, United State, Singapore, Australia and Japan respectively. Know that almost user use traffic in the local country and for a foreign country is United State, it makes even know a demand for more International Bandwidth. We should make sure the International Bandwidth is meet the needs of the user or not.

- **Summary Report of Bandwidth by Application**

The perspective of bandwidth by the application, we presented in Sankey Diagram and show in each month. The application used the network traffic most, which have data link bolder.

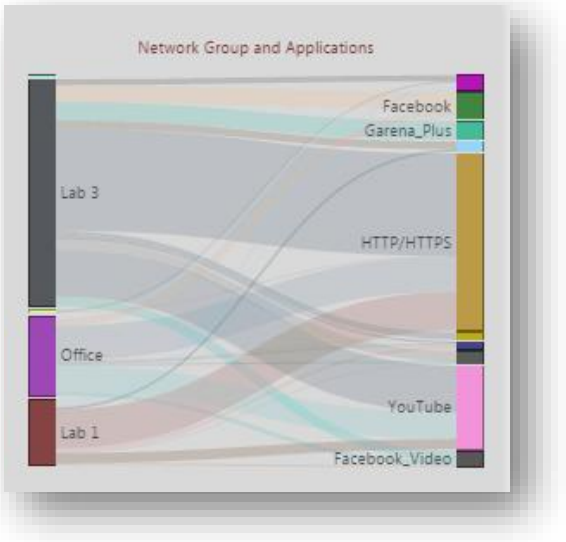


Figure 4.21 Network Group and Application in March.

As a figure 4.21 application usage in network traffic most in March, use applications is HTTP/HTTPS, YouTube, Facebook, Garena_plus, and Facebook_Video respectively.

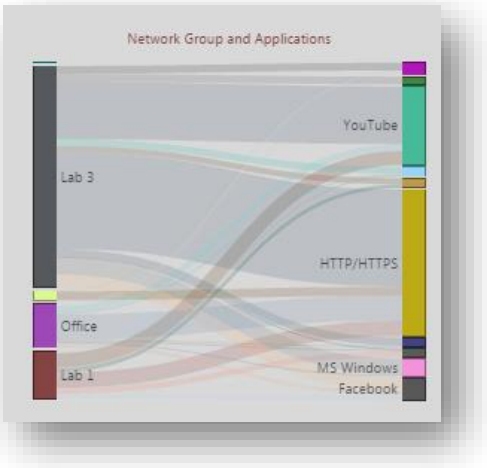


Figure 4.22 Network Group and Application in April.

As a figure 4.22 application usage in network traffic most in April are HTTP/HTTPS, YouTube, MS Windows, Facebook, and Facebook_Video respectively.

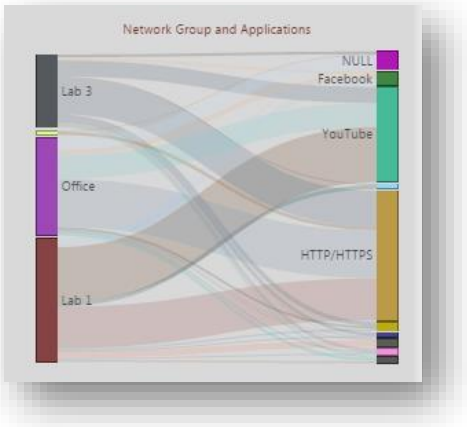


Figure 4.23 Network Group and Application in May.

As a figure 4.23 application usage in network traffic most in May are HTTP/HTTPS, YouTube, Facebook, HTTP/HTTPS Download and Microsoft respectively.

Figure 4.24 Applications used most in Word Cloud chart.

In this view, which displays the application was using the network traffic most. The demand for use the network applications are HTTP/HTTPS, YouTube is secondary, and the third is Facebook. However, HTTP/HTTPS application could drill down the category of application for finding the real type usage.

The summary in the dashboard, we get the results from each report which make conclusions from the dashboard. By using a variety of user requirements of network applications, the report summarized the results as follows.

- Most of the networks which use the Internet are students and staff. The nature of the Internet usage, it would be different with a group of users. Maybe suggest to the key of the Internet use for providing the internet with the usage priority.
- The internet mainly using are HTTP/HTTPS, YouTube, and Facebook, can indicate the use of an active and lead to serve the legitimate purpose of the provided the Internet services.
- Although the Internet is mainly used for domestic bandwidth. But the bandwidth outside the country may not be enough to use. It is an approach increase the network bandwidth with suitable the application networking.
- The daily bandwidth usage is not the same, but the report has shown the average bandwidth on Tuesday more than another day. That should manage distributed the classes, which spread computer using evenly to reduce the network bandwidth usage in a single day more than another day.

4.2 Evaluation of Dashboard Using

This section discusses the results of dashboard was obtained from Deputy Dean of IT, who evaluated dashboard by the assessment of the effects below.

1) The accuracy of processing in the dashboard was estimated to moderate, because of cannot verify the data.

2) Appropriate to group reports were assessed low. The group has not reported in the media, making it difficult to understand.

3) The presentation format is an accurate and complete assessment of the level-Moderate, the report is not in the same group making it difficult to understand the content.

4) Deployment type of data estimated to be in good.

5) Access to the report easy estimated to be in excellent.

6) The report responding when choosing materials estimated in good.

7) The data download conveniently was evaluated at a level low, because the data used in the dashboard are unfit to downloading.

8) Report design evaluated in the Moderate.

9) Can actually implement estimated in excellent.

The results of the evaluation are summarized average score was 3.33 of 5.00 point, the results in a normal level. We got the suggestion for improving the dashboard layout presentation and based on the same journey report layouts in the same group. So they make an understand with the information easily.

4.3 BI Dashboard Refinement

The suggestion from Deputy Dean of IT to refine BI Dashboard layout and report group, which make understand to the result easier. We have to split a report in the same group and separate report out to make the dashboard not too tight. After update dashboard from suggested, the result has been as below.

- After improving dashboard, we split out the report from 1 dashboard 1 page to 1 dashboard 3 page as figure 2.5, figure 2.6, figure 2.7.
- We are adjusting a report in the same group of data.
- As figure 2.6, we improve the value of data into a percentage of total bandwidth in each month.
- Figure 2.6, we are splitting pie chart for March, April, May.

To adjust the dashboard from a suggestion for improving the layout of the report in the dashboard. By modifying from one dashboard and one-page change to one dashboard 3 pages. In page 1, shows the network usage over time, shows daily network usage between Sunday - Monday. The network usage period from 9.00 – 18.00, which can be displayed by filtering monthly intervals by selecting the checkbox on the left of the report as Figure 2.5. In page 2 of dashboard displays network usage reports based applications. In the upper part of the page showing the network usage of network group (e.g. Lab1, Lab 2, Lab 3-6, Lab 4/ Lab 5/ Lab Mac and Office) by displayed based on usage each month in pies chart. The Sankey report shows the top 10 application of

network traffic usage. In the Word Cloud report show top 20 applications, which use the most network bandwidth as Figure 2.6. In page 3 of the dashboard shows the data of the destination network in the Word Cloud report, the tree map report shows the countries, where have been deployed with the top 10 most used as figure 2.7.

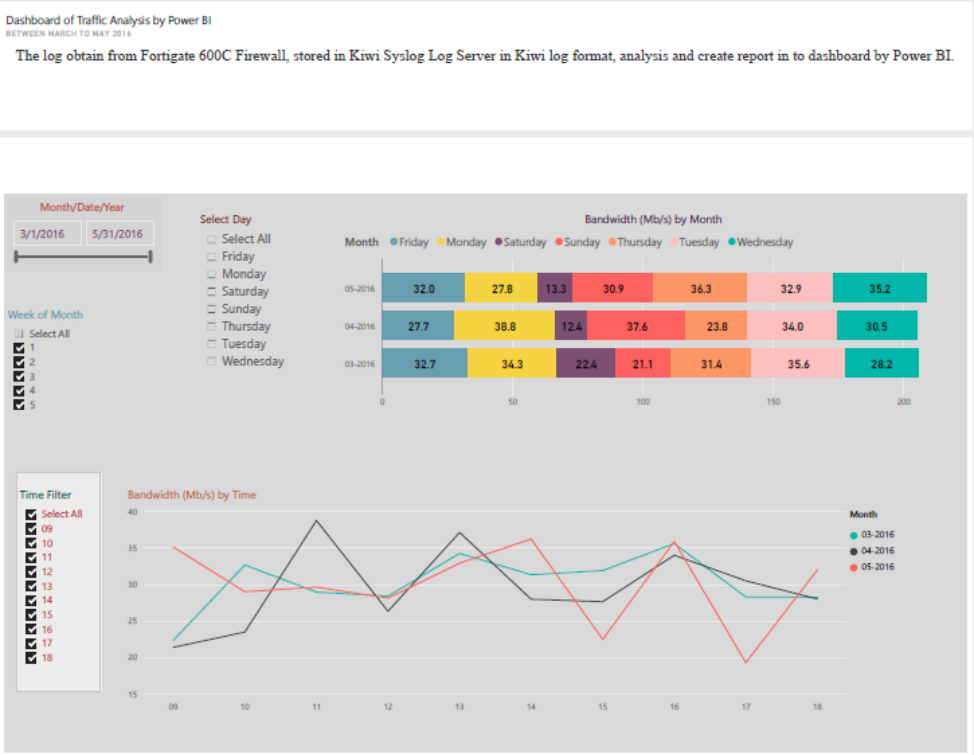


Figure 4.25 Dashboard after refine (1).

Figure 4.26 Dashboard after refine (2).

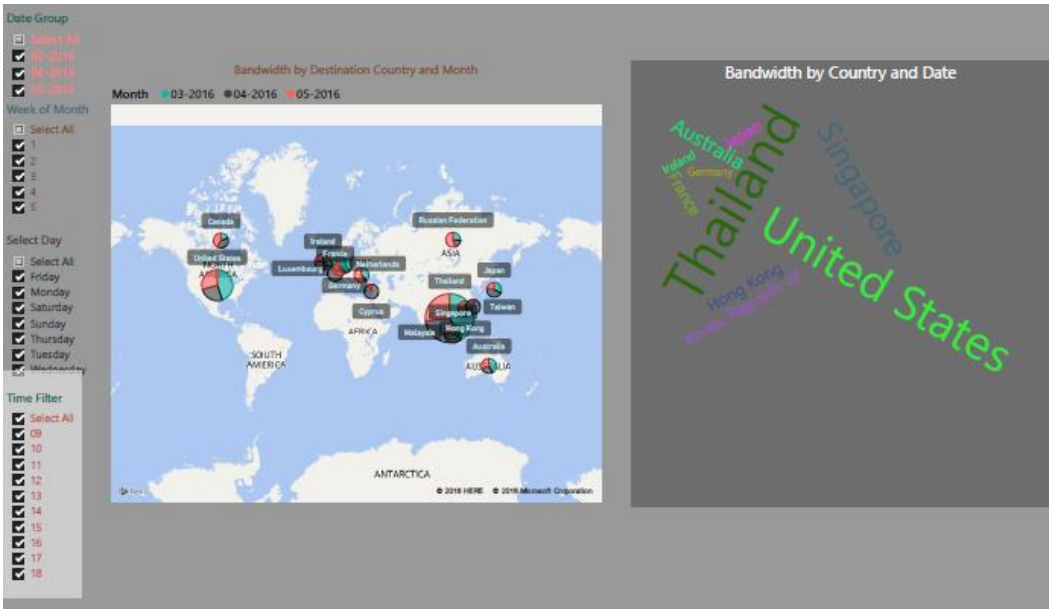


Figure 4.27 Dashboard after refine (3).

This chapter shows the network traffic analysis report from Chapter 3 and the report came out on the dashboard. By evaluating the user feedback, the data series used to offer attractive and got the lack of design. The detail of results of this research is summarized in the next chapter.

CHAPTER V
CONCLUSION

This thematic paper presents the visualization-driven by Business Intelligence to analyze the network traffic and leads to tentative ICT plan for ICT faculty, Silpakorn University. The conclusion is shown below.

5.1 Conclusion

The problem with using the Internet as soon as a complaint from users. IT departments need to resolve the problem, but cannot identify real of the source problem. To identify the problems by analyzing network data and presented to Deputy Dean of IT, the information should make understand easier. The network traffic is during March to May 2016, which was analyzed and presented in a dashboard.

The researchers studied works of logs server, logs management, analyzing methods, presenting the information in a dashboard, and use Top-N statistical to analysis the data as the related work guide in the operation.

In the methodology process, the researchers studied traffic format and get requirement from Deputy Dean of IT to identify the variables that want to on the dashboard. Then we keep data from Kiwi log server and parsing the traffic log in a format which easily for analyzed and put into a database. We use tools for statistical analysis connect to the database and analysis data, by using the network parameters (e.g. date, time, destination, application, service and received byte). The data were analyzed by statistical functions and display in the appropriate manner of presentation to Deputy Dean of IT and evaluate the dashboard.

The analysis of the traffic log, we obtained 3 pages and 1 dashboard. The dashboard can make an understand the results from the internet usage easier.

5.2 Suggestion and Future Work

The current study was limited to hardware and software for analysis a plenty data. This work needs a large storage and memory for stored data. It needs database server and statistics tool for analysis large data. To avoid them we should filter data and stored only important data, it will take up less space. And should use the high-performance computer for analyzing data.

The results of network traffic analysis, which have been monitoring the internet use through the network to finding a regular use by the aforementioned data were statistically analyzed. In our future work will including improved the network traffic management, to prepare a log for analysis in the accuracy. The traffic log has the plenty of data, we should have storage database which storage more data and query data quickly. In order to efficiently the information and present the information in real time. And additional the network security analysis by observing the transmission of information to a foreign country in behavior unique, it will be useful information for IT security for the organization in the future.

REFERENCES

[1] Gary A. Donahue (June 2007). Network Warrior. O'Reilly Media, Inc. p. 3-5.

[2] Gary A. Donahue (June 2007). Network Warrior. O'Reilly Media, Inc. p. 361.

[3] Fortinet Inc. Logging and Reporting for FortiOS 5.0: Fortinet, Inc.; 2014 December 16.

[4] GmbH RGAA. The Syslog Protocol: IETF Contributions published; 2009. [Accessed 28 May 2016] Available from: https://tools.ietf.org/html/rfc5424.

[5] William Stewart 2015, RFC's. Internet Request for Comments. [Accessed 18 June 2016] Available from: http://www.livinginternet.com/i/ia_rfc.htm.

[6] Society TI. The BSD Syslog Protocol: Cisco Systems; 2001. [Accessed 28 May 2016] Available from https://tools.ietf.org/html/rfc3164.

[7] SolarWinds I. Kiwi Syslog Server: SolarWinds, Inc.; 1998. [Accessed 28 May 2016] Available from: http://www.kiwisyslog.com/help/syslog/index.html?action_log_file_formats.htm.

[8] SAWMILL. LOG FORMATS SUPPORTED BY SAWMILL. [Accessed 28 May 2016] Available from: https://www.sawmill.net/log_formats.html.

[9] SolidMatrix Technologies I. Common Format and MIME Type for Comma-Separated Values (CSV) Files: The Internet Society 2005. [Accessed 11 July 2016] Available from: https://tools.ietf.org/html/rfc4180.

[10] Computer-Related Crime Act B.E. 2007. Thailand. 26. Sect. 3 (2007).

[11] Criteria for the storage of a computer system access. Thailand. 26. Sect. 3 (2007).

[12] IBM. Corp. Chart types: International Business Machines Corp. [Accessed 11 July 2016] Available from http://www.ibm.com/support/knowledgecenter/th/SSMR4U_10.2.1/com.ibm.swg.ba.cognos.dsk_ug.10.2.1.doc/c_dsk_charts_intro.html.

[13] Column charts. Google.2016. [Accessed 15 Aug 2016] Available from https://support.google.com/docs/answer/190718?hl=en

[14] Line Chart. Developer Express Inc. 1998-2016. [Accessed 15 Aug 2016] Available from https://documentation.devexpress.com/#WindowsForms/CustomDocument2976)

[15] Pie Chart. Developer Express Inc. 1998-2016. [Accessed 15 Aug 2016] Available from https://documentation.devexpress.com/#WindowsForms/CustomDocument2978

[16] Side-by-Side Bar Chart. Developer Express Inc. 1998-2016. [Accessed 15 Aug 2016] Available from https://documentation.devexpress.com/#WindowsForms/CustomDocument2972

[17] Area Chart. Developer Express Inc. 1998-2016. [Accessed 15 Aug 2016] Available from https://documentation.devexpress.com/#WindowsForms/CustomDocument2979

[18] Scatter Chart with Highlight Box to Group Data Points in Chart. Excel & VBA - da Tab Is On. [Accessed 15 Aug 2016] Available from http://www.databison.com/scatter-chart-with-highlight-box-to-group-data-points-in-chart/

[19] Bubble Chart. PresentationLoad GmbH. 2016. [Accessed 15 Aug 2016] Available from http://www.charteo.com/en/PowerPoint/Data-driven-Diagrams/Bubble-Chart-35.html

[20] Combination Chart. WWW.EXCEL-EASY.COM. 2010-2016. [Accessed 15 Aug 2016] Available from http://www.excel-easy.com/examples/combination-chart.html

[21] Tree map. Google.2016. [Accessed 15 Aug 2016] Available from https://support.google.com/docs/answer/190718?hl=en

[22] Wu J-Y. Computational Intelligence-Based Intelligent Business Intelligence System: Concept and Framework. Bangkok. Computer and Network Technology (ICCNT), 2010 Second International Conference on: IEEE. p. 334 - 8.

[23] N. Jacome-Grajales, G. Escobedo-Briones, J. Roblero and G. Arroyo-Figueroa . Application of Business Intelligence to the Power System Process Security. Innovative Computing Technology (INTECH), 2013. P.258 - 262.

[24] Zhaojun G, Yong L, Wenjing N, Tianjin C, editors. Analysis and implement of PIX firewall Syslog log. Information Management and Engineering (ICIME), 2010 The 2nd IEEE International Conference on; 2010 16-18 April 2010.

[25] Huang Jh, Zhang Mq, Jiang Yl, editors. The design and implement of the centralized log gathering and analysis system. Computer Science and Automation Engineering (CSAE), 2012 IEEE International Conference on; 2012 25-27 May 2012.

[26] Fortinet Inc. The FortiGate Cookbook 5.2: Fortinet, Inc; 2015.

[27] Fortinet Inc. FortiOS Log Reference Guide: Fortinet Technologies Inc.; 2014. p.13 - 28

[28] FortiGate Log Message Reference - FortiOS 5.0.10 p. 29.

[29] LogReferenceGuide Fortinet TechnologiesInc. p.11

[30] Juice I. A Guide to Creating Dashboards People Love to Use 2009-2010 [cited 2009 November].

APPENDIX

Evaluation of the dashboard analysis of network usage

This questionnaire is intended to be used to evaluate the use of the dashboard analysis of the network applications of information and communications technology during the months of March to May 2016 and bring the results of the analysis presented in a dashboard used to create policies to improve efficiency even more.

The explanation for those who complete the assessment dashboard.

This questionnaire is divided into two parts.

Part 1: the information of assessor.

Part 2: the evaluated of the trial dashboard according to the attitude of users.

Evaluation of Network Dashboard

Gender	□ Male	□ Female			
Educational degree	□ Under BS	□ BS	□ MS or higher education		
Seniority	□ 25 to less than 30 years		□ 31 to less than 35 years		
	□ 36 to less than 40 years		□ 41 to less than 45 years		
	□ 46 to less than 50 years				
Below are the attitude and performance of BI Dashboards for network dashboard.					
	Very low	Low	Moderate	Good	Excellent
1. The accuracy of processing					
2. Appropriate to group reports					

3. Presentation format the report is accurate and complete					
4. Deployment type of data					
5. There is easy access to the report					
6. Responding to the report when choose materials					
7. It is convenient to download the data					
8. Report design is beautiful					
9. Can actually be implemented					

Comment

……

……

……

……

BIOGRAPHY

NAME	Mr. Kamphon Kornanan
DATE OF BIRTH	1 June 1988
PLACE OF BIRTH	Prachuap Khiri Khan, Thailand
INSTITUTIONS ATTENDED	University of the Thai Chamber of Commerce, 2006-2010 Bachelor of Science (Computer Science) Mahidol University, 2014-2017 Master of Science (Information Technology Management)
HOME ADDRESS	258/3 Village No. 3 Aow-Noi Sub-district, Mueang Prachuap Khiri Khan, Prachuap Khiri Khan, 77000 Tel. 084-971-9912 E-mail: k_bbird@live.com
EMPLOYMENT ADDRESS	Faculty of Information and Communication Technology, Silpakorn University Phetchaburi IT Campus. No.1 Village No.3, Sampraya, Cha-Am, Phetchaburi 76120, Thailand. E-mail: kornanan_k@su.ac.th, k_bbird@live.com