Faculty of Information and Communication Technology

PASSIVE MEASUREMENT METHOD FOR UNKNOWN NETWORK PROTOCOL IDENTIFICATION AND CLASSIFICATION

NORAYU ABD GHANI

MASTER OF SCIENCE IN INFORMATION AND COMMUNICATION TECHNOLOGY

2010

PASSIVE MEASUREMENT METHOD FOR UNKNOWN NETWORK PROTOCOL IDENTIFICATION AND CLASSIFICATION

NORAYU ABD GHANI

A thesis submitted

in fulfillment of the requirements for the degree of Master of Science

in Information and Communication Technology

Faculty of Information and Communication Technology

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2010

DECLARATION

I declare that this thesis entitle "Passive Measurement Method for Unknown Network Protocol Identification and Classification" is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature : .....

Name : **NORAYU ABD GHANI**

Date : 22 July 2010

TABLE OF CONTENT

LIST OF TABLE

LIST OF FIGURE

ACKNOWLEDGEMENT

A journey is easier when you travel together. Interdependence is certainly more valuable than independence. This thesis is the end of my long journey in obtaining my Master Degree in IT by research work whereby I have been accompanied and supported by many people with words of encouragement.

It is a pleasant aspect that I have now the opportunity to express my gratitude for all of them. I am deeply indebted to my supervisor Prof. Dr. Shahrin Sahib from the Faculty of Information Technology and Communication, Technical University of Malaysia whose help, stimulating suggestions and encouragement helped me in all the time of research and not to get lost during the development of this thesis not only in technical but emotional support. His wide knowledge and his logical way of thinking have been of great value for me. His understanding, encouraging and guidance have provided a good basis for the present thesis. Also thanks to my co-supervisor, Dr. Mohd Faizal Abdollah, his supervision helped me in finalize the research and writing of this thesis. My sincere thanks also goes to Prof. Dr Nanna Suryana Herman whose sincerely encourage, support and been an inspiration on how to make things perfect.

I feel a deep sense of gratefulness for my parents who formed part of my vision and taught me the good things that really matter in life. My acknowledgement would not be completed without expressing special respect, love and thank to my husband Fazli Musa who is always give me his endless support enabled me to complete this work.

ABSTRAK

Pemantauan trafik rangkaian lazimmnya dimanfaatkan oleh pentadbir rangkaian untuk mencapai sasaran prestasi dan keselamatan rangkaian. Bagaimanapun, melaksanakan pemantauan trafik rangkaian menghadapi beberapa cabaran, di antarnya mengenalpasti dengan tepat trafik, alat yang sesuai dan strategi pemantauan. Perisian penganalisa protokol rangkaian adalah salah satu alat yang popular dalam membantu pentadbir rangkaian untuk melaksanakan pemantauan trafik. Oleh kerana itu, ketepatan dalam mengenalpasti dan mengklasifikasikan trafik rangkaian dapat membantu mempermudahkan pemantauan trafik rangkaian, disamping itu operasi rangkaian dapat difahami dengan lebih baik. Oleh kerana itu setiap paket yang berada dalam rangkaian perlu dikenalpasti dengan tepat untuk mencapai pulangan atas pelaburan yang optimum. Bagaimanapun, keupayaan perisian penganalisa protokol rangkaian boleh menjadi satu cabaran kepada pentadbir rangkaian. Mencerap protocol rangkaian yang tidak dapat dikenali adalah satu cabaran untuk mencapai kecekapan dalam menyediakan perkhidmatan rangkaian. Oleh itu kajian ini memfokuskan untuk mengenalpasti dan mengklasifikasikan protokol rangkaian yang tidak dikenali dalam rangkaian UTeM. Kajian ini mencadangkan satu metodology umum untuk mengenal pasti dan mengklasifikasikan trafik rangkaian tersebut. Walaupun vendor menyatakan keupayaan perisan yang disediakan dapat mengenalpasti trafik rangkain dengan tepat, kajian ini menunjukkan sebaliknya, dan perisian yang berbeza mengklasifikan protokol trafik secara berbeza.

ABSTRACT

Network traffic monitoring is a way for enterprises to meet performance, security and compliance goals. Yet implementing network traffic monitoring tools can also pose a series of challenges that range from difficulty in identifying exact network traffic to trouble finding the right tools and strategies for monitoring. Software protocol analyzer is a popular tool in helping network administrator to perform network traffic monitoring. In view of the fact that, accuracy in identification and classification of network packet could advanced network monitoring, and better understanding of the operational networks applications. Therefore, every packets running on the network should be able to be recognized and accurately defined to optimize network resources' usage and return of investment. Anyhow, the capability of network protocol analyzer in decoding network traffic could be a challenge to the network administrator. Capturing network traffic with unknown network protocol is a challenge to provide efficient and accurate network service. This work is focusing on to identify and reclassify the unknown network protocol in UTeM network. UNTICED methodology proposed in this research able to accurately identify and reclassify unknown network protocol in the university network. While many software protocol analyzer vendor claims to provide accurate protocol classification, research finding confirms that different software protocol analyzer classified protocol differently. For this reason, the accuracy of network protocol analyzer claimed is to confirm tool dependent.

CHAPTER 1

INTRODUCTION

1.1 Overview

This chapter describes an overview of the thesis. The context of this chapter is structured into sections that explain the detail of this research. First, it begins with a background evolution of internet and network. Secondly, the motivation that leads to the problem statement is discussed. Next, research objectives are identified then followed by the research strategies which are derived through these objectives. Beside, the scope of research and experiment implementations is included to provide an understanding of the research area. On top of that, the overall layout of the research with short description of each chapter is explained, lastly; a conclusion section is provided to conclude this chapter

1.2 Background

Evolution of Internet has been accompanied by the development of various network applications such as text-based utilities, web, electronic commerce, video, voice and multimedia streaming (Nakamura et al., 2003). The main trust behind these activities is the exploitation of high-speed and sophisticated communication technologies, which commence serious challenges to network bandwidth management as well as network design (Oppenheimer, 2004) (Moore & Zuev, 2005) on local area networks. This scenario present challenges for network monitoring (Distinct, 2008). Obviously, all of these applications causing the growing fraction of all traffic carried by the network, and because of that every packets in network traffic becomes harder to classify. Furthermore, protocol misclassification was mentioned in (Cisco, 2006) that inherited wrong service classifications. Identifying and classifying the types of network traffic that compete for limited bandwidth is the first step toward understanding and solving performance problems and security issues as well.

Using special network measurement hardware or software, information about network packets could be collected. Information collected is useful in a variety of work such as helping in troubleshooting, protocol debugging, workload characterization, and performance evaluation and improvement. Specific tools used to capture and analyze network traffic so that data collected from the network could be presenting in a specific pattern for a specific purpose as wished by an organization.

Network traffic for Internet Protocol (IP) (Kozierok, 2005) packet classification (Wang, 2009) could be defined as assortment of packets according to defined rules. Rules for packet classification (Zhu et al., 2007) can be based on several fields in the IP header, classification determine the per-hop behaviors and traffic conditioning functions such as shaping and dropping that are to be applied to the packet. Specific method (Willianson, 2001) in measuring network traffic needs to be clarified to make sure the analysis done meet the research objectives

In order to analyze the network traffic, specific tool need to be use to capture the packets. This activity is call sniffing. The packet sniffer is a computer software or hardware that can intercept and log traffic passing over a digital network. A sniffer is also refers as a protocol analyzer or network analyzer. These terms are the names used interchangeably within the industry for tools that perform protocol and network analysis via packet captures.

Protocol analyzer is the handy tool help in these solutions; however, thousand of products in the market that cause different interpretation based on the vendor definition. An organization must be sure that the chosen network analyzer can actually see what is happening on the network; this required the accuracy of protocol interpretation by the network analyzer. Network applications depend on networking protocols to transmit data, and many of today's security issues are directly related to the improper use of these

protocols, for that reason, the accuracy of protocol analyzer in defining a protocol is necessity.

1.3 Unknown Protocol

Unknown protocol here refers to protocol unable interpreted by the protocol analyzer. Study show that the unknown protocol might be caused by P2P application, fragmented packet, user behavior or a network attack (Tolga, 2010), (Microsoft Support, 2005), (Karagiannis et al., 2004). Every protocol analyzer has their own packet decoder, the packet decoders are the essential engines to decode packets, the decoder attempts to find a match in its defined set of IP protocols and sub protocols for each packets (Capsa, 2009). If a match cannot found, then the packet is marked as unknown. The sub protocol is the next layer down within each IP Protocol. The unknown packets may be unnecessary or unexpected traffic (Ma et al., 2006). Cited in (Microsoft Help and Support, 2005), unknown traffic was appeared in one server in which the server has no capabilities to identify all those packet's protocols. They consider the following network traffic to be unknown or unidentified:

i. Network traffic that does not match any protocol definition, this traffic typically includes primary connections.

ii. Network traffic that no application filters takes responsibility for. This traffic typically includes secondary connection.

1.4 Motivation

In today's computer world, evolution of Internet in line with the development of network application has brought difficulty for accurately recognized every packet in the network. Every network applications have their own implication to the design of the network such as bandwidth, hardware, and security management. Network administrators have the responsibility to allocate every network resources in efficient and effective way. Therefore, every packets running on the network should be recognizable, and accurately defined to optimize the usage of network resources, and investment; furthermore, user will experience better performance.

Network traffic characteristics depend on many factors including network technology, user and application behavior (Viipuri, 2004). Unfortunately, running well-known applications on not-so-well-known ports or used others protocol as wrappers to pass through the firewall is a common scenario, this especially true for most of peer-to-peer application. Moreover, the increasing of network application and traffic caused some of the traffic unidentifiable, thus classified as unknown protocol by network monitoring tools. Those tools are developing to display packets exchanges textually in order to show the operations of various protocols. However, a common problem with many protocol analyzers is inability to accurately identify (Caceres, 1989) network protocol so that those unidentifiable protocol been classified as unknown network protocol. Therefore, looking exclusively at port number to identify packet is not the best way to reveal the correct

identification of unknown packets (Moore & Papagiannaki, 2005). This situation increases the possibility for the existence of unknown packets in the network traffic flow, which need to classify. Effective network management and resources management could be achieved by accurately identify network traffics.

Reported by (Madhukar & Williamson, 2006), the internet unknown traffic is keep on increasing every year. There has been a variety of works on analyzing and parsing network protocols with varying degrees of automation (Gopalrathnam et al., 2006). Two trades off for the unknown network protocol management, improving bandwidth management by accurately either identify those unknown packet or by removing the unknown packet, so that the bandwidth could be free out for bandwidth intensive application, at the same time, could improve the network performance.

Network packets may have different classification criteria and the classification of packets could be made based on header fields such as IP Source Address, Destination Address, and protocol or fields in the packet payload such as port number (Moore & Zuev, 2005). The analysis of network traffic mostly depends on the tool used to capture packets for further analysis (Shahrin et al., 2006).

In general, organization invest huge amount of resources in network technology to compete in business world, definitely they wish to gain higher return of investment. A healthy network is crucial to ongoing business operations. The evolvement in peer-to-peer

(P2P) technology could be seen as one of the factors that affect the network performance of organization (F5 Network, 2007) (CAIDA, 2010). Studies show that, employee behavior is believed to as one of the factor that contributes (KJ et al., 2007) to inefficient and ineffective of the running network (Wayne, 2005). By accessing to P2P technology and the nature of P2P application cause the increasing number of unknown network traffic (Ramco & Paris, 2004). Current survey shows that employees are using company resources to access peer-to-peer (P2P) applications on company time and exposing organizations to serious and potentially risks (Madhukar & Williamson, 2006).

In order to help the network administrator to react quickly and accurately to network problem and security issue, the accuracy of network protocol identification is essential. Hence, accurate identification and classification (Dainotti et al., 2008) of network protocol can benefit for those organizations in terms of return of investment. Thus, the shortcoming of protocol analyzer to identify all network traffic accurately motivates this research, so that the unknown network protocol could be identified, that will benefit in network monitoring activities.

1.5 Problem Statement

IT personnel are under extreme pressure to respond to incidents that affect the network performance, organization profitability, and security as soon as they occur. Therefore, accurately identify all network traffics and its protocol is essential to improve

network performance and services. Cited in Beale et al. (2004) vendor tends to develop the application for network monitoring tools based on their interpretation and definition. If the protocol analyzer could not identify the packets according to their way of decoding, these packets are group as unknown. Viipuri (2004), Ramco & Paris (2004), and Gopalrathnam (2006) have done the improvement of algorithm however there are still some packets unidentifiable. Most of the software protocol analyzers merely classify these packets under a group of protocol classification call unknown or *Others* protocol (IBM, 2009).

This research attempts to discover factors that related to the existence of *unknown* packets and to identify and classify the *unknown* packets. The *unknown* packets may be unnecessary or unexpected traffic. Effective network management and resources management could achieve by accurately identify those traffics. Effective network management is predicates on the ability to detect, diagnose, and resolve problems quickly while resources management is a concept that used to improve performance of variety of systems. For this reasons, analyzing network protocol and traffic is an essential solution. Analyzing network protocol is depending on the network protocol analyzer. Protocol analyzer is a tool, which is necessary for network administrator to know about the network operation and condition. Packet captures tools were developing to display packets exchanges textually in order to show the operations of various protocols.

Each application on the network is associated with network port to permit the communication. The theory is if all packets are from known port and from known

applications, there should be all identified packets in the network. Than all traffics in the network should generated from known applications and the respective port number, since each applications are associated with particular port, all packets must be on known port and known application. It cannot be on *unknown* port or *unknown* application. On the other hand, this means that every applications are identifiable and have their specific port number as identification criteria.

Intensive and strong analyses need to be carried out to investigate the existence of these packets to identify the source of *unknown* and the implication of the *unknown* to the network. If the *unknown* essentially undesirable of course their occupancy truly waste the network resources such as bandwidth.

In general, organization invest huge amount of resources in network technology to compete in business world, definitely they wish to gain higher return of investment. A healthy network is crucial to ongoing business operations. A minute of downtime can cause significant financial losses. Furthermore, time is money, so the minute of achieving efficient and effective employment of network technology service as soon as its impact on revenue or cost savings will be felt. The evolvement in peer-to-peer (P2P) technology could be seen as one of the factors that gives an impact to the network performance of an organization. Since some studies show that, employee behaviors are believed to as one of the factor that contributes to inefficient and ineffective of the running network. By accessing to P2P technology and the nature of P2P application cause the increasing number

of *unknown*. Current survey show that employees are using company resources to access peer-to-peer (P2P) applications on company time and exposing organizations to serious and potentially risks and, seeing as P2P is primarily used to exchange pirated audio, video, and software files or inappropriate content. The effect of running P2P applications, downloading large files, and allowing fellow P2P users to upload files from employee's shared folders can also slow down network performance, negatively influencing the functioning of business-critical applications. P2P contribute most of organization Internet traffic (Ramco & Paris, 2004), (Madhukar & Williamson, 2006). Indirectly, the unknown network protocol that is believed contributed by P2P application also as a contribution factors for unproductive used of network bandwidth, which could decrease network performance and disturb network services. For these reasons properly utilizing network bandwidth help in preventing below factors:

i. Loss of productivity

ii. Loss of revenue

iii. Loss of credibility

In other words, unknown network protocols give a negative impact to productivity and return of investment. Abuagla & Sulaiman (2009a), Antoniades et al. (2009), Gopalrathnam et al. (2006), Moore & Papagiannaki (2005) and Ramco & Paris (2004) had done some improvement to classify unknown network protocol, anyhow there still some portion of protocol unidentifiable. Consequently, this research focuses to identify and