BORANG PENGESAHAN STATUS TESIS*

JUDUL: GENERALIZATION OF P2P TRAFFIC PATTERN

SESI PENGAJIAN: I / 2007

Saya SITI AZURA BINTI MD SAID

mengaku membenarkan tesis (PSM/~~Sarjana/Doktor Falsafah~~) ini disimpan di Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dengan syarat-syarat kegunaan seperti berikut:

1. Tesis dan projek adalah hakmilik Universiti Teknikal Malaysia Melaka.
2. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. ** Sila tandakan (/)

_____	SULIT	(Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)
_____	TERHAD	(Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan)
___/___	TIDAK TERHAD	

(TANDATANGAN PENULIS)

Alamat tetap : No 2, Lorong 2/92B,
Taman Kobena, Cheras 56100,
Kuala Lumpur

Tarikh : 13/11/2007

(TANDATANGAN PENYELIA)

Prof. Dr. Nanna Suryana Herman

Tarikh : 13/11/2007

CATATAN:

* Tesis dimaksudkan sebagai Laporan Projek Sarjana Muda (PSM)
** Jika tesis ini SULIT atau TERHAD, sila lampirkan surat daripada pihak berkuasa.

GENERALIZATION OF P2P TRAFFIC PATTERN

SITI AZURA BINTI MD SAID

This report is submitted in partial fulfillment of the requirements for the
Bachelor of Computer Science (Computer Networking)

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY
UNIVERSITI TEKNIKAL MALAYSIA MELAKA
2007

DECLARATION

I hereby declare that this project report entitled

GENERALIZATION OF P2P TRAFFIC PATTERN

is written by me and is my own effort and that no part has been plagiarized without citations.

STUDENT : _____ Date: 13/11/2007

(SITI AZURA BINTI MD SAID)

SUPERVISOR : _____ Date: 13/11/2007

(PROF DR. NANNA SURYANA HERMAN)

DEDICATION

To my lovely bonda Sarifa Fauzia Syed Ahmad,
and supportive ayahanda Md Said Ibrahim

ACKNOWLEDGMENTS

I want to express my sincere gratitude towards my project supervisor, Prof. Dr. Nanna Suryana Herman for his suggestions, feedbacks, and patience. Thanks a million to him for all of his support and encouragement.

To both my parents, Sarifa Fauzia Syed Ahmad and Md Said Ibrahim, thanks a lot for your support and energy that you gave to me. Thank you for the trust and not let me down.

Also a special word of thanks to lecturers and fellow friends of Faculty of Information and Communication Technology (FTMK) department at UTeM

Encik Mohd Faizal Abdollah
Ahmad Irwan Ab Ghani
Azril Naim Ahmad Zainudin
Mohamad Hafiz Mat Zani
Mohd Fairuz Hamzah
Sapri Sahabudin
Siti Aminah Masor

for their helps, guidance and supports when I need them. Thank you.

ABSTRACT

The project that I have chosen to evaluate is called Generalization of P2P Traffic Pattern. This project includes analyzing data, transferring files and monitoring the network traffic. Basically, this project is about analyzing the network traffic that happens in P2P applications. The analyzing process covers five different types of P2P application. Each P2P applications produced their own traffic pattern that are differ from one another. The P2P applications that I have analyzed are Ares, iMesh, eMule, BitComet and LimeWire. These five applications have been analyzed in a few context of packet such as capacity of the files as well as the protocols that being used. Analysis that has been done is executed using Ethereal or also known as Wireshark. This software allows me to execute all my activities and analysis. That activity has been done with the help of that software. This software has generated a graph for every situation that makes the analysis a lot more easily accordingly.

ABSTRAK

Projek yang saya pilih untuk diperhalusi diberi tajuk *Generalization of P2P Traffic Pattern*. Pelaksanaan dalam projek tersebut termasuklah menganalisa data, memindah fail, dan memantau keadaan rangkaian. Secara amnya, projek ini melibatkan penilaian dan analisa terhadap rangkaian trafik yang berlaku dalam applikasi antara *peer*. Proses analisa dalam projek ini meliputi lima jenis applikasi P2P yang berbeza. Setiap satu applikasi P2P akan menghasilkan pola trafik yang tersendiri yang berbeza antara satu applikasi berbanding applikasi yang lain. Applikasi P2P yang saya pilih untuk tujuan projek ini adalah Ares, iMesh, eMule, BitComet dan LimeWire. Kelima-lima applikasi ini telah di analisa dalam beberapa konteks paket seperti kapasiti fail serta protokol yang digunakan oleh setiap satu applikasi P2P tersebut. Analisa yang dijalankan menggunakan perisian Ethereal ataupun kini lebih dikenali sebagai Wireshark. Perisian tersebut membolehkan aktiviti-aktiviti menganalisa data dapat dijalankan dengan lebih lancar dan sempurna. Melalui perisian ini juga, gambarah graf dapat diaplikasikan dan proses analisa menjadi lebih mudah dan tersusun.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

P2P	-	Peer-to-peer
ICT	-	Information and Communication Technology
TTL	-	Time-to-live
HTTP	-	Hypertext Transfer Protocol
IP	-	Internet Protocol
R&D	-	Research and Development
PSM	-	*Projek Sarjana Muda*
TCP	-	Transmission Control Protocol
CASE	-	Computer-Aided Software Engineering
ISP	-	Internet Service Provider
FTP	-	File Transfer Protocol
IO	-	Input / Output
ARP	-	Address Resolution Protocol
MAC	-	Media Access Control
UDP	-	User Datagram Protocol
IGMP	-	Internet Group Management Protocol
ICMP	-	Internet Control Message Protocol

CHAPTER I

INTRODUCTION

1.1 Project Background

Peer-to-peer (P2P) systems are becoming increasingly popular as they enable users to exchange the digital information by participating in complex networks. P2P systems have become a popular medium through which to share huge amounts of data.

The P2P applications that I have decided to study and describe its behavior are based on five different network or protocol. The application involves are Ares, BitComet, eMule, iMesh, and LimeWire.

These P2P uses different networks and it works slightly different from one another. These P2P depends on every peer for whatever files user searches for. How the peer works? The explanation of how those P2P works has been elaborated for this project.

P2P computer network relies primarily on the computing power and bandwidth of the peer or client in the network rather than concentrating it in a relatively low number of servers. P2P are networking that typically connecting nodes or peers using ad hoc connection.

Each P2P have produced different pattern of behavior within the time that I have set to analyze it. The difference that occurs is what I have decided to analyze and it is where the difference of those P2Ps lies.

1.2 Problem Statements

In every project that needs to be develop always have at least one problem that will influence the motive of it. The type of network involves for this project are Ares, BitTorrent, eDonkey, FastTrack as well as Gnutella. This network has been analyzed using an analyzer, Ethereal.

The problem that I have been facing in order to develop the project is security. I have requested for a R&D laboratory for the analyzing purpose.

However, in a matter of security, I decided not to execute the application in the laboratory instead; I used my very own laptop for this purpose. This is because the P2P application not only downloading files, but also some unknown viruses that will and already corrupt my hard disk drive.

1.3 Objectives

The analysis covered:

- Generalize the P2P traffic pattern.
  After capturing the traffic pattern, there is a pattern of every situation captured.
- Find the P2P traffic pattern of :
  - before
  - current
  - after

  which shows the different pattern of each traffic.
- Find the fastest P2P application for downloaded files.

1.4 Scope

The scope that involves in this project is:

- Packet traffic analysis.
  Each application has produced slightly different packet traffic.
- Performance analysis.
  Analyzed the difference in performance of each application.
- Traffic pattern analysis.
  After capturing the traffic packet, the application produced a pattern.
- Generalization traffic pattern analysis.
  The pattern of the traffic has then been generalized in the end of the process.

1.5 Project Significance

The project significance would be the outcome or output that has been analyzed by me after the completion of the project. The results not only show the fastest P2P application, but also helps user to choose the best P2P for them to download files. The completion of this project will also help other researches out there to compare the behavior of P2P in broader scope.

1.6 Expected Output

After the completion, I have expected to produce the packet traffic of each and every application that I have chosen. After the packet traffic has been produced, I get to see the traffic pattern and generalized at the end of the process.

1.7 Conclusion

This chapter is simply a brief introduction to the project that I have been doing for the whole semester. By understanding the needs and details of the project, I keep my project activities on track. Also, I have reviewed some literatures as well as journals and it has been describe and elaborate in Chapter II.

CHAPTER II

LITERATURE REVIEW AND PROJECT METHODOLOGY

2.1 Introduction

Literature review is important to the one who wants to develop a system, make a research or even the one who analyze network performance. The more the one reviews the literature, the more knowledge they will gain.

I am required to review some journals as well as literatures to make me understand more on what I should do during the PSM duration. For my project, I have reviews more than ten journals in order to make me understand more the how the packet traffic moves.

Since P2P are quite popular and has been used widely nowadays, the journals and literatures of this topic can easily be found in the Internet. However, in P2P is not only about the packet traffic, but it also about the performance. It has been divided into categories which makes my findings a little difficult because the topics are somehow irrelevant.

In order to get the correct findings and journals, one will have to know the key-words of their findings. The related journals will then be the best reference for those who wish to get the different point of view or to get the idea on this project.

2.2 Facts and Findings

Fact-finding technique is the step where one collects and gathers information related to the project.

2.2.1 Domain

P2P traffic pattern is in ICT in Networking and Distributed Computing domain. The title itself, P2P does relate to distributed computing. P2P distribute the main costs of sharing data – disk space for storing files and bandwidth for transferring it across the peers in the network.

2.2.2 Existing Research

Research is important for those who wish to develop a system, analyze network traffic or to get the better ideas on how to start their project or even compares their project with others.

2.2.2.1 Readings: Peer-to-Peer Applications

P2P packet traffic can be captured using a number of simulations. "According to M. Bawa *et al.* (2003), the search techniques should be simple and practical enough to be easily incorporated into existing systems". Current successfully deployed P2P data-sharing systems follow very simple protocols.

Some review however states that, "A question naturally arises: how can we compare these technologies and evaluate which one is better given an application? For instance, some P2P concepts and technologies – are well suited to capture and support certain aspects (A. Perini, 2001)".

This is because each application is differ from one another. The packet that has been captured from application A will produce the different packet from application B.

P2P is a complex system. I will have to understand and characterized the behavior of the existing applications. Effective search techniques are needed to make provisions for the unreliable nature of the peers.

P2P traffic is a significant fraction of total workload. In April 2003, 20 – 40% of total bytes correspond to traffic. The case studies are on unstructured systems and structured ones.

Although these protocols are clearly suboptimal, they highlight how simplicity is the key to wide and rapid adoption. One will have to understand and characterizes the behavior of the existing P2P applications.

Any techniques should be adaptive, and tune itself according to the current state of the system. The research had also studied the ways to improve the existing systems by exploring the novel ways to organize and use unstructured P2P systems.

In particular, they had explored the possibility of a completely decentralized search network built in and ad hoc way. Unlike structured topologies, hosts are not restricted to certain neighbors. Instead, the protocol is devoted to incrementally improving the established network through self-supervision.

In structured P2P systems, the location of an object or resource is determined by globally agreed – upon scheme. They had also search and explored a new search protocol that can be viewed as a hybrid of structured and unstructured systems.

Based on Gnutella network, "Generated traffic split by message type over a 376 minute period. Note that overhead traffic forms more than 50% of traffic (R. Matei, 2002)."