



**ANALYZING THE IMPACT OF SWARM SIZE AND PEER  
UP LINK CAPACITY FOR BIT TORRENT NETWORKING  
PERFORMANCE**

**TEO BOON YAN**

**MASTER OF COMPUTER SCIENCE  
(INTERNETWORKING TECHNOLOGY)**

**2017**



**Faculty of Information and Communication Technology**

**ANALYZING THE IMPACT OF SWARM SIZE AND PEER  
UPLINK CAPACITY FOR BIT TORRENT NETWORKING  
PERFORMANCE**

**TEO BOON YAN**

**Master of Computer Science (Internetworking Technology)**

**2017**

**ANALYZING THE IMPACT OF SWARM SIZE AND PEER UPLINK  
CAPACITY FOR BIT TORRENT NETWORKING PERFORMANCE**

**TEO BOON YAN**

**A dissertation submitted  
in fulfillment of the requirements for the degree of Master in Computer  
Science (Internetworking Technology)**

**Faculty of Information and Communication Technology**

**UNIVERSITI TEKNIKAL MALAYSIA MELAKA**

**2017**

## DECLARATION

I declare that this dissertation entitled “Analyzing The Impact Of Swarm Size And Peer Uplink Capacity for Bit Torrent Networking Performance” is the result of my own research except as cited in the references. The dissertation has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature : .....

Name : Teo Boon Yan

Date : .....

## APPROVAL

I hereby declare that I have read this dissertation and in my opinion, this dissertation is sufficient in terms of scope and quality as a partial fulfillment of Master of Computer Science (Internetworking Technology).

Signature : .....

Supervisor Name : Dr. Wahidah Binti Md. Shah

Date : .....

## **DEDICATION**

I lovingly dedicate my dissertation to my beloved family for their love and care, for my supervisors and professors, who has provided me guidance and advice as well as my fellow classmates, my sincere gratitude towards all my colleagues and friends who each and everyone has provided encouragement and supported me each me step of the way.

## ABSTRACT

This dissertation is to document with the intention to prepare and provide an insight to Bit Torrent Networking. The topic for the dissertation is *Analyzing The Impact of Swarm Size and Peer Uplink Capacity for Bit Torrent Networking Performance*. The purpose of this dissertation is to give a better insight about Bit Torrent, which would be doing for my final year project. As Bit Torrent is a large portion of networking especially peer to peer networking, there is only a surface of understanding about Bit Torrent performance. This dissertation is would use lab experimental and software to simulate the Bit Torrent performance with a small, medium and large swarm of users. The methodology would use the simulation to explore the performance of Bit Torrent networking as the size of the swarm expands. Besides that, this dissertation project would also deep dive into what would happen to the Bit Torrent performance if the network bandwidth of the swarm is low and the impact it may have on the average Bit Torrent performance. With the initial studies and previous reference as well as the lab experiments, it would provide a better insight on the performance for Bit Torrent networking with the impact of different swarm size, data chunk size, the uplink capacity as well as the comparison with client-server architecture.

## ABSTRAK

*Dengan tujuan disertasi ini adalah untuk menyempurnakan Projek Penyelidikan dengan tujuan untuk menyediakan serta memberi gambaran tentang Bit Torrent. Tajuk untuk disertasi ini adalah Analysis Impak Size Kumpulan Dan Pemuatnaik Serakan Untuk Bit Torrent Prestasi Jaringan. Disertasi ini akan memberikan gambaran atau sudut pandangan and lebih mendalam tentang Bit Torrent yang akan saya sertakan untuk projek akhir tahun. Yang disekia maklum, Bit Torrent adalah sebahagian besar daripada jaringan network terutamanya dalam jaringan serakan, dalam hal ini hanya didapati pemahaman dasar tentang prestasi Bit Torrent. Disertakan disertasi ini akan mengemukakan eksperimen makmal serta simulasi maya untuk menyelidik prestasi Bit Torrent dengan siz kecil, sederhana serta besar untuk jaringan sesawang. Methodologi yang akan digunakan untuk mengkaji prestasi Bit Torrent sesawang dengan kekumpulan siz yang semakin meningkat. Selain itu, disertasi ini akan mengkaji dengan mendalam tentang prestasi Bit Torrent untuk size data, kebolehan muat naik serta banding dengan jaringan jenis server-klien. Dengan analisis pendahuluan serta rujukan penyelidikan lab makmal terdahulu, akan dapat memberi pandangan yang lebih baik tentang prestasi jaringan Bit Torrent serta impaknya dalam siz kumpulan, size data, kuasa muat naik dan banding dengan pelayan-pelanggan..*



## ACKNOWLEDGEMENTS

I would like to convey my upmost gratitude to everyone who enable me the possibility to complete this dissertation. A special appreciation I give to my dissertation, with much gratefulness for the crucial role of my supervisor Dr. Wahidah binti Md Shah, who gave the permission to use all required tools and application necessary to complete the task “*Analyzing The Impact Of Swarm Size And Peer Uplink Capacity For Bit Torrent Networking Performance*”.

Furthermore, I would also wish to acknowledge Professor Dr Burairah bin Hussin, Dr Othman bin Mohd and Dr Faizal Abdollah, whose contribution in stimulating suggestions and encouragement, helped me to coordinate my dissertation especially in writing this dissertation. A special thanks goes to my fellow team mates, who help me with the moral support and drive to succeed. Last but not least, many thanks go to my family members who inspire me to write this dissertation. I would also like to express my gratitude towards other professors and people around me, who helped me, improve my skills and potential towards writing this dissertation.

# TABLE OF CONTENTS

	PAGE
<b>DECLARATION</b>	
<b>APPROVAL</b>	
<b>DEDICATION</b>	
<b>ABSTRACT</b>	<b>i</b>
<b>ABSTRAK</b>	<b>ii</b>
<b>ACKNOWLEDGEMENTS</b>	<b>iii</b>
<b>TABLE OF CONTENTS</b>	<b>iv</b>
<b>LIST OF TABLES</b>	<b>vi</b>
<b>LIST OF FIGURES</b>	<b>vii</b>
<b>LIST OF EQUATIONS</b>	<b>viii</b>
<b>LIST OF ABBREVIATIONS</b>	<b>ix</b>
<b>CHAPTER</b>	
<b>1. INTRODUCTION</b>	<b>1</b>
1.1 Introduction	1
1.2 Research Background	2
1.3 Research Problem	5
1.4 Research Question	6
1.5 Research Objective	7
1.6 Research Method	7
1.7 Research Contribution	8
1.8 Research Scope	9
1.9 Research Result	9
1.10 Summary	10
<b>2. LITERATURE REVIEW</b>	<b>12</b>
2.1 Introduction	12
2.2 Machine Learning	13
2.3 Bit Torrent Taxonomy	18
2.4 Bit Torrent Operation	20
2.5 File Sharing	21
2.6 Bit Torrent Parameters	22
2.7 Service Control	22
2.8 Network Simulator, Emulators and Crawlers	23
2.8.1 Swarm	25
2.8.2 Seeders and Leechers	26
2.8.3 Piece and Data Chunk Size	27
2.8.4 Client Server Architecture	27
2.8.5 Download Speed	27
2.8.6 Upload Speed	28
2.8.7 Download Time	28
2.8.8 NS Simulator	29
2.9 Bit Torrent Dataset	30
2.10 Summary	32

<b>3. RESEARCH METHOD</b>	<b>33</b>
3.1 Introduction	33
3.2 General Research Methodology	33
3.3 Clarify and Understand Research Question	34
3.4 Gather Information	35
3.5 Literature Review	35
3.6 Exploratory Study	35
3.7 Quantitative Technique	35
3.8 Design Experiment	36
3.9 Data Collection	37
3.9.1 Bit Torrent Data Chunk Size With Reference To Data Completion Time	37
3.1.2 Bit Torrent Seed Uplink Speed With Reference To Data Completion Time	38
3.10 Analysis Of Experiment	38
3.11 Interpretation Of Findings	38
3.12 Document Findings	39
3.13 Summary	39
<b>4. ANALYSIS</b>	<b>40</b>
4.1 Introduction	40
4.2 Data Collection	40
4.3 Data Preparation	41
4.4 Data Setting	41
4.5 Data Preparation and Representation	44
4.6 Knowledge Representation	47
4.7 Summary	48
<b>5. RESULT AND DISCUSSION</b>	<b>49</b>
5.1 Introduction	49
5.2 Experiments Result for Seed Chunk Size Analysis	49
5.3 Experiments Result for Seed Uplink Capacity	50
5.4 Experiments Result for Bit Torrent Comparison with Client Server Paradigm	51
5.5 Summary	52
<b>6. CONCLUSION AND FURTHER DIRECTIONS</b>	<b>53</b>
6.1 Introduction	53
6.2 Summary of Research	53
6.3 Research Contribution	55
6.4 Research Limitation	56
6.5 Future Research	56
6.6 Summary	57
<b>REFERENCES</b>	<b>58</b>

## LIST OF TABLES

<b>TABLE</b>	<b>TITLE</b>	<b>PAGE</b>
1.1	The Research Questions About Bit Torrent Networking	6
1.2	Research Objectives Relating To Bit Torrent Networking.	7
1.3	Research Methodology About Bit Torrent Networking.	7
1.4	Research Contribution About Bit Torrent Networking.	8
1.5	The Research Scope Of The Dissertation	9
2.1	P2P File Sharing Networks And Running Applications	22
4.1	Table Displaying The Data Chunk Size In Kbytes For 250Kbytes And 500Kbytes With Respect To The Completion Time In Seconds.	41
4.2	Table Displaying The Data Chunk Size in Kbytes for 1000Kbytes, 2500Kbytes and 5000Kbytes With Respect To The Completion Time In Seconds	42
4.3	Table Showing Speed uplink capacity (Kbps) for 500Kbps and 1000Kbps with The Completion Time (Seconds)	43
4.4	Table Showing Speed uplink capacity (Kbps) for 2500Kbps and 5000Kbps with The Completion Time (Seconds).	44

## LIST OF FIGURES

<b>FIGURE</b>	<b>TITLE</b>	<b>PAGE</b>
2.1	Figure of Taxonomy for Bit Torrent	19
3.1	The General Research Methodology	34
4.3	Bit Torrent and Client Server Paradigm Comparison Of Number of Chunks Received	47

## **LIST OF ABBREVIATIONS**

NS	-	Network Simulator
TCP	-	Transmission Control Protocol
P2P	-	Peer to peer
UDP	-	User Datagram Protocol
QoE	-	Quality Of Experience
IP	-	Internet Protocol

## LIST OF EQUATION

<b>EQUATION</b>	<b>TITLE</b>	<b>PAGE</b>
1	Euqation 1	37
2	Equation 2	38
3	Equation 3	45

## CHAPTER 1

### INTRODUCTION

#### 1.1 Introduction

Peer to peer (P2P) internetworking is a large part of the overall internet. Peer to peer networking mechanisms encompasses the network as users can transfer data to another user on the network on the same level and also linking with other peers on the network with the similar data sources which encompasses globally linking all the users. This enables the users to harness the resources from various sources while sharing the data at the same time with other users globally. Peer-to-peer network is a type of used communication the practice of peer to peer type of file distribution is commonly used to disseminate data or electronic data through the Internet. Peer to peer (P2P) network is one of the most frequently used protocol for transferring large data files. For example, video clips, songs, movie files or digital audio files. The basic concept of the peer to peer networking is where a user could send or receive files through the network using peer-to-peer sharing. The peer-to-peer computing or networking is a type of disseminated application construction that partitions tasks or jobs among all the peers or connected network clients. Peers or clients are all equally privileged and all have equal potential to participate in the application. This form of network is known as a peer-to-peer network of nodes in the network. Peers encompasses a section of their properties for example the linkage bandwidth or the processing power which is directly available to other network participants or peers without a server.



All peers are both the providers or contributor to the network as well as the consumer of resources in the network which is different from the traditional method of client-server model which the intake and the supply of the resources are separated and controlled by the server. The emerging collaboration of peer to peer systems are successful in advance of the age of peers doing the same things while distributing data among the resources which are looking for diversified peers which can bring in new or exclusive resources and abilities to a virtual network where it engaged a larger task which is beyond what is able to be achieved by selected individual peers which intern benefits all the connected peers. As peer to peer becomes more popular the usage of peer to peer networking has been incorporated into several different technologies and protocols which uses the peer to peer networking structure.

## **1.1 Research Background**

The background about this dissertation is to provide an in-depth understanding about the peer to peer networking, in particularly the Bit Torrent protocol of peer to peer networking's behavior through lab experimental simulation. In networking, there are many different types and structures of networking methods. One of a fast developing and trending peer to peer networking structure is Bit Torrent.

Peer to peer is a major form of sharing data and being distributed over the Internet and a major contributor is Bit Torrent. As Bit Torrent is the most popular form of protocol for sharing huge data and peer-to-peer networks have been able to share files over a large portion of the Internet. The boom started around 2004, where Bit Torrent was responsible for almost a third of all Internet traffic. With rapid growth as of 2013, Bit Torrent was

responsible for three percent all worldwide bandwidth, and the amount is over half of six percent of sum of the bandwidth which is specialized for file sharing (Bryan & Willis 2012).

The protocol which Bit Torrent use to transmit files is a client application which implements the Bit Torrent paradigm. These applications are inclusive of Bit Torrent,  $\mu$ Torrent, Transmission, qBittorrent, Vuze, Deluge, and BitComet (Chandra et al. 2011). The Bit Torrent enables the host to look for peers which are called seeds that can transmit data and is able to produce a list of all the data possible for transfer.

Bit Torrent's beginning is actually discovered by a programmer named Bram Cohen, who was an ex-University at Buffalo's student, who used to initiate the standard in 2001. Bit Torrent clients are compatible for a variety of computer systems. (Computing 2015).

Since 2013, Bit Torrent has risen to an amount over 15 million simultaneous users at any time. By the year 2012, Bit Torrent is used by over hundred and fifty million users. The number of bit torrent users is estimated to be over two hundred and fifty million users (Griffin & Kesidis 2011). This huge figure compose of a large number of users and this is a huge growing figure, but we yet to have a good understanding on how the network will perform in small, medium or large number of users. We are also do not know the behavior exactly if the users have limited networking connection speed if it would affect the overall Bit Torrent networking performance.

The BitTorrent standard can be utilized bringing down the overall cost of server as well as the consequences for networking of distributing huge data. Besides installing a file from a server the BitTorrent standard also permits hosts to connect to a group of hosts to transfer data from each other at the same time. The protocol is another way to the a single source, several mirroring server sources mechanism for sharing files and can function effectively in networks with significantly lower bandwidth. By utilizing the BitTorrent

protocol, a several clients, such as personal computers can be used to replace large servers while efficiently sharing files to many recipients. Sudden jumps in the internet traffic can be prevented by using this lower bandwidth in a given location. It also helps to maintain internet traffic higher for most of the users, regardless if they use the BitTorrent protocol or not.(Chandra et al. 2011).

A host that wants to upload a document would need to first create a simple torrent descriptor file that they distribute by conventional means such as email (Hanle n.d.). The host would then make the data itself available through a BitTorrent node acting as a seed. Those with the torrent descriptor file can give it to their own BitTorrent nodes, which acting as peers or leechers download it by connecting to the seed and/or other peers.

The details on the distributed file transmission is the data being transmitted is sectioned into several segments called pieces. Every piece would receive a new sample of the file it becomes a source for other nodes or users, thus relieving the initial seed from having to send that piece to every computer or user wishing a copy. By using BitTorrent the duty of disseminating the file is shared by those who need it, this is only possible for a particular seed to send only one copy of the file itself and at the end share it with an uncountable number of peers.

Every slice or segment is controlled by an encrypted hash which is found in the torrent descriptor (Schollmeier et al. 2002). This ensures that changes of the pieces can be tracked and detected. Thus, this avoids accidental and malicious changes of any of the segments received at other nodes. If a node begins with an original copy of the torrent descriptor, it can check the authenticity of the entire file it receives.

Segmented pieces are normally installed randomly and are reset into the proper sequence from BitTorrent Client that checks which segment it needs. The segments are made

up of the same dimension and size in a installation. Because of this method, the installation of any data may be stopped at any time and continue at a later time and not damaging previously downloaded data. This makes Bit Torrent very useful in transferring large size files (Wolisz 2011). This makes it possible for the client to be available for segments and install them at once and its better than stopping the installation and wait for the next segment in line, that normally decreases the total time of the download.

When a node has finished downloading a document, it becomes an additional seed. This then changes its state from peers to seeders quantified by the overall strength of the file.

The segmented pieces of BitTorrent can cause a wide spreading of a file over many computer nodes. When there are more peers that join the swarm, the possibility of a complete download by any particular node is higher. The same as the norm Internet distribution methods, it allows a huge decrease in cost for the source's hardware and bandwidth cost.

## **1.2 Research Problem**

The research problem in this dissertation is to define and analyze the lack of understanding about the behavior of Bit Torrent networking in various different scenarios. In particularly the behavior of the Bit Torrent network with growing size of the packet or chunk size, if there are variation in the uplink speed and the comparison with client-server based networking (Rossi 2012).

As Bit Torrent is a growing technology being used these days, there are several uncertain factors. As technology advances, the data size of data or chunks of data being shared and transmitted through the internet is also increasing. With this increasing trend of the chunk or data size there is much to understand about the relationship of the Bit Torrent

network if the data size or chunk size of the data is increasing. The phenomenon and the relationship ratio of the Bit Torrent performance on the growing data trend.

Bit Torrent usage usually use  $\mu$ TP for its message. For a gain of band-widths, Bit Torrent used is UDP, that enables checking of the internet traffic source address. Thus may lead to Denial-of-service attacks, where hosts using Bit Torrent clients that are used as amplifiers for an attack to another service (Melakessou & Engel 2009).

Several research about Bit Torrent has shown that a large percentage of files obtainable for download through Bit Torrent have malicious files.

There is not much indication on how the Bit Torrent will perform in small, medium or large number of users or large swarms. It is also an ambiguity if the users have low uploading speed if it would slow down the Bit Torrent performance.

Hence in this dissertation would further analyst the performance for different number of swarms and also if uploading speeds are slow for the swarm.

### **1.3 Research Question**

Referring back to the Research Problems in Section 1.3, there are three research questions which will be addressed. The research problems are comprised which form the research problem statement are as follows:

Table 1.1: The Research Questions About Bit Torrent Networking.

Q1. How is the chunk size in the Bit Torrent network affect the Bit Torrent network performance?
Q2. How will the uplink speed affect the Bit Torrent network performance?
Q3. How will seed uplink capacity in Bit Torrent compare to client server paradigm performance?

#### 1.4 Research Objective

Referring to the research questions formulated in Section 1.4, the research objective which have been developed as follows:

Table 1.2: Research Objectives Relating To Bit Torrent Networking.

O1. To study and understand about Bit Torrent network characteristic.
O2. To design experiment and gather data for Bit Torrent networking.
O3. To analyse the impact of Bit Torrent network with respect to chunk size, uplink speed and Bit Torrent comparison with client server paradigm.

#### 1.5 Research Method

Based on the research objective in Section 1.5, the research method have been developed are as follows:

Table 1.3: Research Methodology About Bit Torrent Networking.

M1. Prior to the analysis setup, a preparation of literature survey is used to gather and obtain information relating to the Bit Torrent networking about the
---

behaviour of Bit Torrent network, the structure of Bit Torrent networking and approaches of analysing Bit Torrent networking.
M2. Using simulation methodology to analyse Bit Torrent network and gather data sets on the classification and behaviour of the Bit Torrent network.
M3. Discuss and provide understanding about the Bit Torrent networking behaviour to evaluate the chunk size, uplink behaviour and comparing the Bit Torrent network to client server paradigm.

## 1.6 Research Contribution

From the research method shown in Section 1.6, the research contribution relating to this dissertation are as follows:

Table 1.4 Research Contribution About Bit Torrent Networking.

C1. An enhanced prediction about the behaviour of Bit Torrent networking with growing size chunks of data in Bit Torrent network with respect to the completion data transfer time.
C2. Improve the understanding and data set for uplink speeds for Bit Torrent networking which can provide a better knowledge about the behaviour of Bit Torrent with the increasing trend of uplink speed.
C3. Provide an understanding about the behaviour of Bit Torrent network with respect to client-server paradigm.

## 1.7 Research Scope

To achieve the Research Objectives, this study will focus on several aspects as stated in the following table scope as shown below:

Table 1.5: The Research Scope Of The Dissertation

S1. The scope of the dissertation will emphasis on datasets containing increasing trend of chunk size in data with respect to the completion time, the uplink speed for Bit Torrent networking, the comparison of uplink speed between Bit Torrent networking and client server archicture paradigm.
S2. To predict and understand the pattern or the relationship of the Bit Torrent network with the increasing pattern of the Bit torrent data size and variation of uplink speed in Bit Torrent network and in comparison with client server paradigm.
S3. Using simulation tools to measure and provide data analysis of the performance of Bit Torrent structure and the results to create statistical analysis experiment.

### 1.8 Research Results

From the research, there is a pattern which can be observed from the generated data sets which would be able to give a better analysis and understanding about the Bit Torrent network behavior in various circumstances. The analysis can be used to predict the behavior of the Bit Torrent network structure to optimize and produce the best possible Bit Torrent structure with the data sizes and uplink speeds to have an optimum completion time for data transfer within the Bit Torrent network structure to share data between the peers. This dissertation is useful to provide a better understanding in the advancing networking of peer to peer architecture for future evaluation and setup of peer to peer networking architecture. A better optimum set for peer to peer networking can be set for better performance in the peer-to-peer network. For additional research can be done to improve the performance and compatibility of the network to suit to the future setup of peer-to-peer networking to enhance and optimize the speed, data transfer rates, completion time, allocation of resources and completion time. Peer to peer network can be harness to promote a better setup of data transfer access to enhance its advantage of data sharing and mass data transfer for a better enhanced overall network. A better understanding of the peer-to-peer networking can improve and clarify the behavior and relationship of the parameters which affects the Bit