

AN APPROACH TOWARDS MOST EFFICIENT PEER-TO-PEER SYSTEMS



Session 2004-2007

Project Advisor
Sir Muddathir Qamar

Submitted By
Riffat Ehsan
0402200-06

Department of Computer Science
University of Management and Technology
Lahore

A report submitted to the
Department of Computer Science
in partial fulfillment of the requirements for the
degree

Bachelor of Science

In

Computer Science

By

Riffat Ehsan
0402200-06

University of Management and Technology
January 23, 2008

Acknowledgements

I truly acknowledge the cooperation and help make by Sir Atif Mehdi, Head Department of Computer Science, University of Management and Technology. He has been a constant source of guidance throughout the course of this project. I would also like to thank my Project Advisor Sir. Muddathir Qamar for his help and guidance in understanding the project. I am also thankful to my friends and families whose silent support led me to complete my project.

Riffat Ehsan
(ID # 040220006)

Date:
Januray 23, 2008

Abstract

Peer-to-Peer systems are gaining very much popularity in the current era in the social, academic and commercial communities. Its attraction is because of the recognition of its key demanding features being provisioned against the client-server environment. To keep running system all the time efficiently and to keep persistent the data, some techniques are being found. In my thesis, I will discuss the possibilities or mechanism to store data fault tolerant in a peer-to-peer system. My project covers the issues regarding reliability in peer to peer systems. It is possible to achieve high reliability and performance even in adverse conditions. I will research on the existing mechanisms regarding the reliability and will use any simulator to check the reliability in peer to peer systems.

List of Figures

<u>Figure 1.1:</u> A peer to peer network.....	3
<u>Figure 1.2:</u> Client server system	3
<u>Figure 1.3:</u> A simple centralized network.....	4
<u>Figure 1.4:</u> A simple decentralized network	5
<u>Figure 1.5:</u> Unstructured P2P network.....	6
<u>Figure 1.6:</u> Structured P2P network.....	7
<u>Figure 2.1:</u> Napster network.....	14
<u>Figure 2.2:</u> KaZaA network.....	17
<u>Figure 2.2:</u> Gnutella network.....	18
<u>Figure 2.3:</u> CAN network.....	24
<u>Figure 2.4:</u> Wide and deep search in the JXTA network.....	26
<u>Figure 3.1:</u> KaZaA	31

Table of Contents

Statement of Submission	i
Acknowledgements	ii
Abstract.....	iii
List of Figures	iv
Table Of Contents.....	v
Chapter 1: Introduction to Peer to Peer Systems.....	1
1.1. Introduction.....	2
1.2. Peer to Peer Systems.....	2
1.2.1 Centralized network.....	4
1.2.2 Decentralized network	4
1.2.3 Unstructured P2P network	5
1.2.4 Structured P2P network.....	6
1.2.5 Hybrid P2P network.....	7
1.3. Working of Peer to Peer systems.....	7
1.4. Categories of P2P Networks.....	8
1. Categories based on Services.....	8
2. Categories based on Application.....	8
1.5. Advantages of peer-to-peer networks.....	9
1.6. Advantages over client/server network.....	10
Chapter 2: Latest Peer to Peer Systems Used Today.....	11
2.1 Introduction.....	12
2.2 Napster.....	13
2.2.1 Background.....	13
2.2.2 Original Program Information.....	14
2.2.3 Current status.....	15
2.3 KaZaA.....	15
2.3.1 Four types of additional software.....	16
2.3.2 Participation level.....	17
2.4 Gnutella.....	17
2.4.1 History.....	18
2.4.2 How it works.....	18
2.4.3 Protocol features and extensions.....	20
2.5 CAN.....	21
2.5.1 Layers.....	21
2.5.2 CAN application fields.....	22
2.6 JXTA.....	23
2.6.1 The origins of JXTA search.....	24
2.6.2 JXTA search.....	24
2.6.3 Wide and deep search methods.....	25
2.6.4 The future of JXTA search	26

Chapter 3: Comparison.....	27
3.1 KaZaA.....	28
3.1.1 Introduction.....	29
3.1.2 What Can Kazaa Do?	30
3.1.3 How Does Kazaa Work?	30
3.1.4 Peer-To-Peer (P2P) and How Kazaa Works.....	31
3.1.5 Kaza files.....	32
3.1.6 Is Kazaa Legal?.....	32
3.1.7 Privacy and Security Issues.....	33
3.1.8 Is Kazaa Spying us?	34
3.2 Gnutella.....	35
3.2.1 Gnutella.....	36
3.2.2 History of Gnutella.....	36
3.2.3 Gnutella's Architecture.....	38
3.2.4 Protocol features and extensions.....	38
3.2.5 Gnutella Clients.....	39
3.2.6 Limitations of the Protocol.....	39
3.2.7 The Future of Gnutella.....	42
3.2.8 Is Gnutella Legal?	43
Conclusion.....	44
References.....	47

TABLE OF CONTENTS

Chapter 1	3
Introduction to Peer to Peer Systems	3
1.1 Introduction	4
1.2 Peer to Peer Systems	4
1.2.1 Centralized network	6
1.2.2 Decentralized network	6
1.2.3 Unstructured P2P network:	7
1.2.4 Structured P2P network:	8
1.2.5 Hybrid P2P network:	9
1.3 Working of Peer to Peer systems	9
1.4 Categories of P2P Networks	10
1.4.2 Categories based on Application:	10
1.5 Advantages of peer-to-peer networks	12
1.6 Advantages over client/server network	12
Chapter 2	14
Latest Peer to Peer Systems Used Today	14
2.1 Introduction	15
2.2 Napster	16
2.2.1 Background	16
2.2.2 Original Program Information	17
2.2.3 Current status	18
2.3 KaZaA	18
2.3.1 Four types of additional software	19
2.3.2 Participation level	20
2.4 Gnutella	20
2.4.1 History	21
2.4.2 How it works	21
2.4.3 Protocol features and extensions	23
2.5 CAN	24
2.5.1 Layers	24
2.5.2 CAN application fields	25
2.6 JXTA	26
2.6.1 The origins of JXTA search	27
2.6.2 JXTA search	27
2.6.3 <i>Wide and deep search methods</i>	28
• Wide search	28
• Deep search	29
2.6.4 <i>The future of JXTA search</i>	29

Chapter 3.....	30
Comparison	30
3.1 Gnutella.....	32
3.1.1 <i>History of Gnutella</i>	32
3.1.2 Gnutella's Architecture	34
3.1.3 Protocol features and extensions	34
3.1.4 Gnutella Clients	35
3.1.5 <i>Limitations of the Protocol</i>	35
3.1.6 <i>The Future of Gnutella</i>	38
3.1.7 Is Gnutella Legal?	39
KaZaA	40
Decentralized P2P network	40
3.2 Introduction.....	41
3.2.1 KaZaA	42
3.2.2 What Can KaZaA Do?	42
3.2.3 How Does KaZaA Work?	42
3.2.4 Peer-To-Peer (P2P) and How KaZaA Works	43
3.2.5 KaZaA files.....	44
3.2.6 Is KaZaA Legal?	44
3.2.7 Privacy and Security Issues	45
3.2.8 Contrvrsial Issues.....	46
INTRODUCTION.....	47
OVERVIEW OF KAZAA DESIGN.....	47
SCREENSHOTS.....	59
Toolbars & Buttons	60
Corners & Borders	61
Layers	61
Refreshing The Skin	62
Image Files	62
Conclusion	64
References.....	67

Chapter 1

Introduction to Peer to Peer Systems

1.1 Introduction

“A peer-to-peer (or P2P) computer network is a network that relies on computing power at the ends of a connection rather than in the network itself.”[1]

It is a communications environment that allows all desktop and laptop computers in the network to act as servers and share their files with other users on the network. It's a way of organizing, and a way of thinking about organizing. It's also a political and social program for those who believe that in many cases, peer to peer modes are a preferable option. These systems are not restricted to technology or P2P file sharing as such, but covers every social process with a peer to peer dynamic, whether these peers are humans or computers. These networks are used for sharing content like audio, video, data or anything in digital format. P2P network can also mean grid computing. Peer-to-peer networks are generally simpler, but they usually do not offer the same performance under heavy loads.

P2P refers to technology that enables two or more peers to collaborate spontaneously in a network of equals (peers) by using appropriate information and communication systems without the necessity for central coordination.

1.2 Peer to Peer Systems

Every network requires some form of special software in order to control the flow of information between the users being networked. A Network Operating System, or "NOS", is installed on each computer requiring network access. It monitors and controls the exchange and flow of files, email, and other network information. Network Operating Systems are classified according to whether they are peer-to-peer or client-server NOSs.

Peer-to-peer NOSs like Windows 95, Windows98 and Windows for Workgroups are best for home & small office use. They do an excellent job of sharing applications, data, printers, and other local resources across a handful of computers.

Client-server NOSs such as Windows NT, Linux and Novel NetWare are better for larger scale organizations that require fast network access for video, publishing, multimedia, spreadsheet, database, and accounting operations.

Figures 1.1 shows five peer-to-peer workstations. All five computers can communicate directly with each other and share each other's resources. Since peer-to-peer

computers have their own hard disk drives that are accessible and sometimes shared by all of the computers on the peer-to-peer network, each computer acts as both a *client* (or node) and a *server* (information storage).

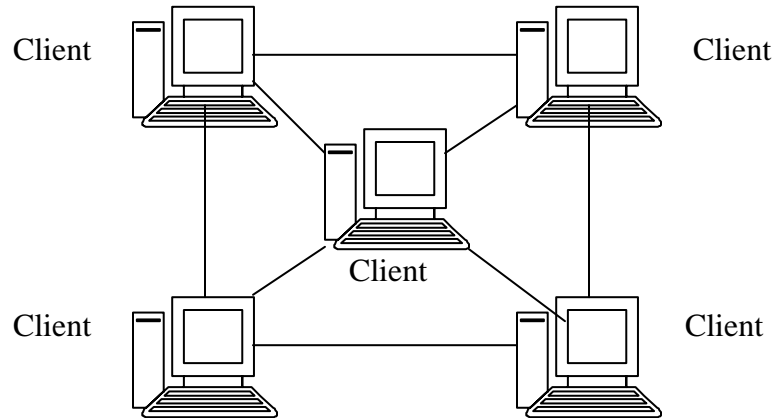


Figure 1.1: A peer to peer network

Figure 1.2 shows that the client computers are separate and subordinate to the file server. The primary applications and files used by each of the clients are stored on the file server. If the clients below want to communicate with each other, they must do so through the file server. A message from one client to another client is first sent to the file server, where it is then routed to its destination by the server. It is obvious that if there are tens or hundreds of client computers, a file server would be the only way to manage the complex and most times simultaneous operations that large networks with many clients would generate.

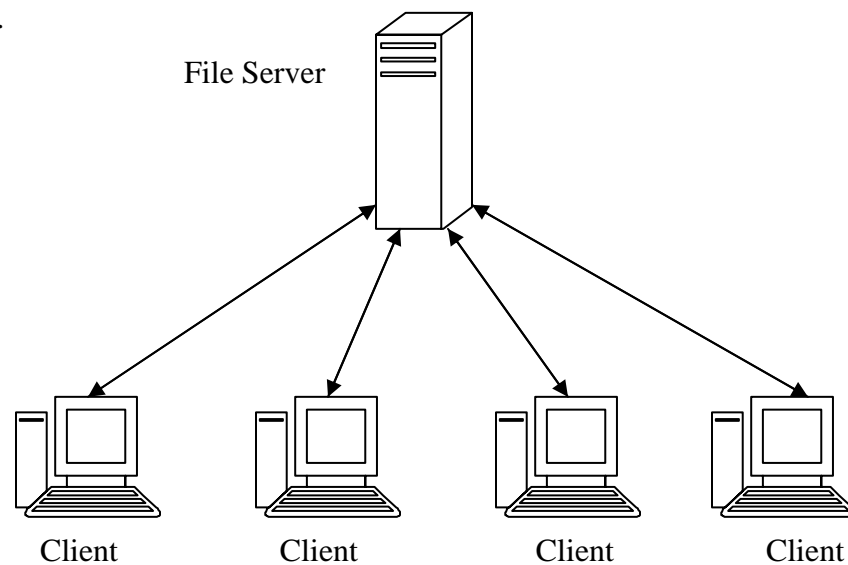


Figure 1.2: Client server system

1.2.1 Centralized network

A centralized network will have much of its links dispersed around one or a few nodes. Napster [3] is one of the Centralized P2P network.

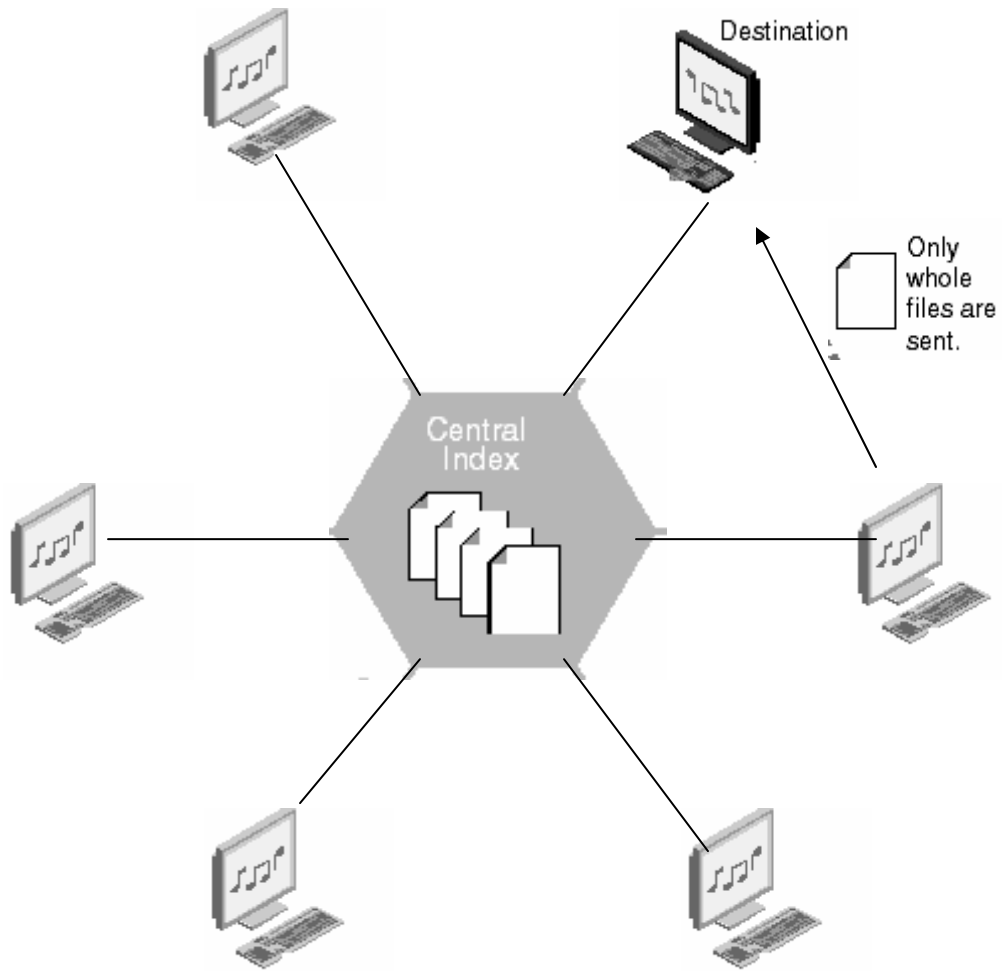


Figure 1.3: A simple centralized network

1.2.2 Decentralized network

A decentralized network is one in which there is little variation between the nodes of links each node possesses. KaZaA [3] is one of the Decentralized P2P network.

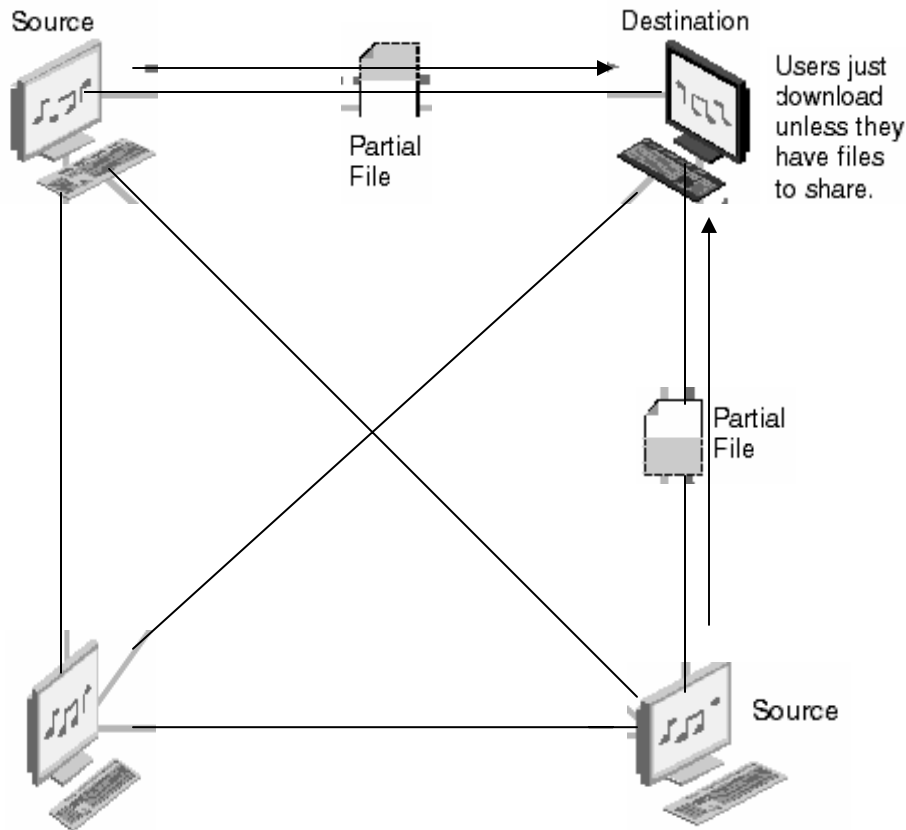


Figure 1.4: A simple decentralized network

1.2.3 Unstructured P2P network:

An unstructured P2P network is formed when the overlay links are established arbitrarily. Such networks can be easily constructed as a new peer that wants to join the network can copy existing links of another node and then form its own links over time. In an unstructured P2P network, if a peer wants to find a desired piece of data in the network, the query has to be flooded through the network in order to find as many peers as possible that share the data.

The main disadvantage with such networks is that the queries may not always be resolved. Popular content is likely to be available at several peers and any peer searching for it is likely to find the same thing, but if a peer is looking for rare data shared by only a few other peers, then it is highly unlikely that search will be successful. Since there is no correlation between a peer and the content managed by it, there is no guarantee that flooding will find a peer that has the desired data. Flooding also causes a high amount of

signaling traffic in the network and hence such networks typically have very poor search efficiency.

Most of the popular P2P networks such as Gnutella [3] is unstructured.

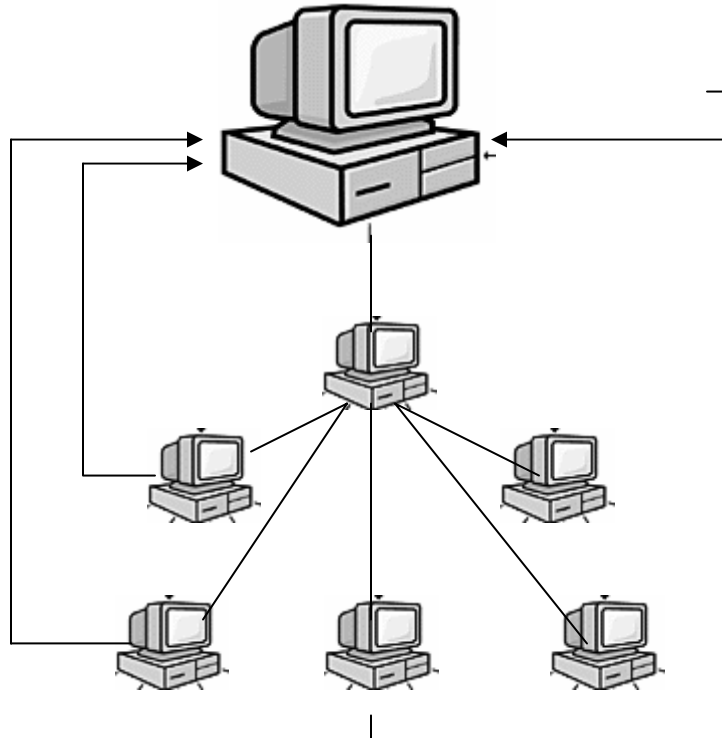


Figure 1.5: Unstructured P2P network

1.2.4 Structured P2P network:

Structured P2P networks employ a globally consistent protocol to ensure that any node can efficiently route a search to some peer that has the desired file, even if the file is extremely rare. Such a guarantee necessitates a more structured pattern of overlay links.

By far the most common type of structured P2P network is the distributed hash table (DHT), in which a variant of consistent hashing is used to assign ownership of each file to a particular peer, in a way analogous to a traditional hash table's assignment of each key to a particular array slot. Some of the well known Structured P2P networks are CAN [3], and Tulip [3].

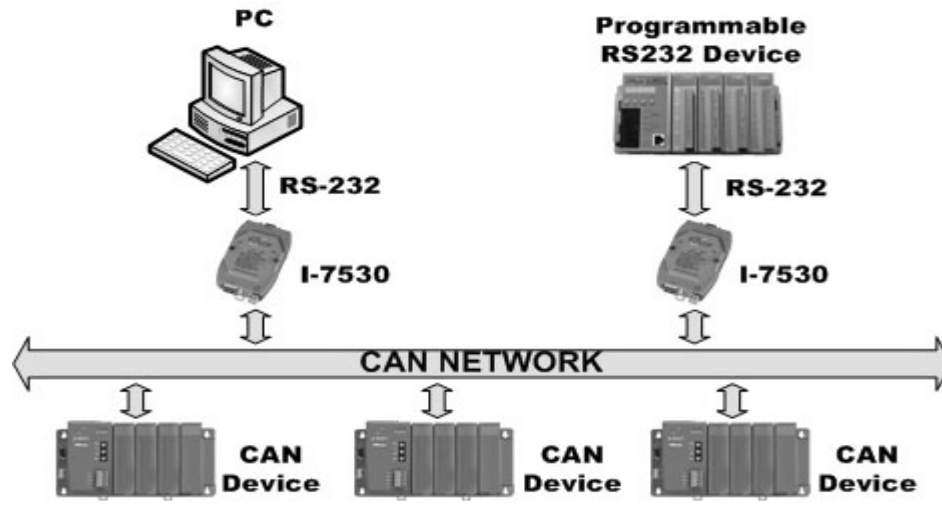


Figure 1.6: Structured P2P network

1.2.5 Hybrid P2P network:

- Hybrid P2P networks are the mixture or the combination of two networks , they can either be the centralized and structured or the decentralized and the unstructured networks combined.
- They serve as the open source platform.
- Hybrid P2P networks are appealing to businesses, because they still afford a level of control that is lost with pure P2P.
- Hybrid P2P networks such as JXTA [3]

1.3 Working of Peer to Peer systems

P2P is not only popular with home users but many small businesses have come to rely on this cost-effective solution for sharing files with co-workers and clients. P2P promotes the ease of working together when you're not physically located in the same office. In just seconds updated files and data can be shared with peers and confidential files can be blocked for security. Additionally, companies can also block access to Internet music and video files to assist in maintaining a work-oriented P2P network. Not only does this keep the company free and clear from legal issues regarding music downloading and sharing but it also keeps the corporate bandwidth usage down.

1.4 Categories of P2P Networks

There are many applications and services that claim to be peer-to-peer.

1.4.1 Categories based on Services:

The two major categories of peer-to-peer systems based on services are:

- 1) File sharing and
- 2) CPU sharing

a) **File sharing:**

File sharing is the practice of making files available for other users to download over the Internet and smaller networks. Usually file sharing follows the peer-to-peer (P2P) model, where the files are stored on and served by personal computers of the users.

b) **CPU sharing:**

It is also called "peer-to-peer computing" and "distributed computing," the latter term first coined in the 1970s. Grid computing is also known as "utility computing."

1.4.2 Categories based on Application:

There are 3 categories of peer to peer systems based on application.

- ❖ Distributed Computing
- ❖ Instant Messaging
- ❖ Affinity Communities

a) *Distributed Computing*

Distributed computing environments pool the processing power of many computers. One popular application involves utilizing spare CPU time on client machines across the network to complete tasks more quickly and make more efficient use of computing resources.

Grid computing promises to enable super computing capabilities at a fraction of current costs by breaking down large computations into much smaller ones that can be run on numerous laptops and PCs with spare CPU time.

Grid computing is becoming popular with financial services, biotech, and science firms that need intense processing power. It combines the idle or unused CPU processing power and/or free disk space of many computers in the network.

b) Instant Messaging

One very common form of P2P networking is Instant Messaging (IM) where software applications, such as MSN Messenger or AOL Instant Messenger, for example, allows interconnected users to exchange text messages and files synchronously. While most vendors offer a free version of their IM software others have begun to focus on enterprise versions of IM software as business and corporations have moved towards implementing IM as a standard communications tool for business.

c) Affinity Communities

Affinity communities are direct file sharing groups, such as Napster and Kazaa. They comprise a specific server like yahoo group, msn group or any other. P2P holds considerable potential for improved collaboration both within and outside the enterprise, and for software interaction, allowing programs to send data inputs and outputs from one application to another. The affinity communities emphasize knowledge integration over acquisition and learning.

1.5 Advantages of peer-to-peer networks

- An important goal in peer-to-peer networks is that all clients provide resources, including bandwidth, storage space, and computing power. Thus, as nodes arrive and demand on the system increases, the total capacity of the system also increases. This is not true of client-server architecture with a fixed set of servers, in which adding more clients could mean slower data transfer for all users.
- The distributed nature of peer-to-peer networks also increases robustness in case of failures by replicating data over multiple peers, and -- in pure P2P systems -- by enabling peers to find the data without relying on a centralized index server. In the latter case, there is no single point of failure in the system.
- When the term peer-to-peer was used to describe the Napster network, it implied that the peer protocol was important, but, in reality, the great achievement of Napster was the empowerment of the peers (i.e., the fringes of the network) in association with a central index, which made it fast and efficient to locate available content. The peer protocol was just a common way to achieve this.
- No need for a network administrator.
- Network is fast and inexpensive to setup and maintain.
- Each computer can make backup copies of its data to other computers for security.
- Peer-to-peer is, by far the easiest type of network to build for either home or office use.

1.6 Advantages over client/server network

Peer-to-peer networking has the following advantages over client/server networking:

- Content and resources can be shared from both the center and the edge of the network. In client/server networking, content and resources are typically shared from only the center of the network.
- A network of peers is easily scaled and more reliable than a single server. A single server is subject to a single point of failure or can be a bottleneck in times of high network utilization.