Examensarbete MEE06:11

P2P Networking and Technology Enablers in Business Applications

Mahesh Hariharan

This thesis is presented as part of the Degree of Master of Science in Electrical Engineering with emphasis in Telecommunications

Blekinge Institute of Technology
2005/2006

Master of Science in Electrical Engineering with emphasis in Telecommunications
Blekinge Institute of Technology
School of Engineering
Department of Telecommunication Systems
Examiner: Markus Fiedler, Blekinge Institute of Technology

ABSTRACT

The usage of Peer to Peer Networks over the Internet has been growing by exponentially. Apart from the hype surrounding P2P, it has remarkable ramifications on the way the Internet could be used. This is an area which is not explored as well as we would want to. This thesis examines the architectural differences in P2P networks and generic application domains where the principles of P2P are exploited.

The usage of P2P in different business verticals and technology enablers that go along with it are presented. The focus is on several case studies each addressing a different use case. The common thread running through all of these use cases is the ability to resolve a business issue.

Finally the focus is on how P2P networks might probably change the way the Internet behaves in the near future.

Keywords: Peer to Peer, Business Applications, Technology Enablers, Collaboration, Distributed Processing, Content Distribution, Knowledge Management

INDEX

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am extremely grateful to Dr. Markus Fiedler, School of Engineering, BTH for his valuable comments on this report.

I also wish to thank the management of Mandriva SA, Paris, France for having provided the opportunity to work with them during this period.

1. INTRODUCTION

1.1 Background

Peer to Peer (P2P) networks have revolutionized computing. As a phenomenon, P2P has caught up very quickly on the Internet. The powerful usage of P2P has tremendous repercussions on the way business is conducted over the Internet.

The ubiquitous nature of the Internet and the ease at which P2P networks can be deployed makes it a force to reckon with.

However, apart from file sharing which P2P fortunately or unfortunately has become synonymous with, there are several other business application areas for the same. Large and medium sized enterprises are looking at P2P as a means to optimise their network resources. Small sized businesses are using P2P to keep application costs low.

P2P networks can be used in applications ranging from conserving Internet bandwidth, accelerating the speed of information dissemination, routing voice calls as well as for sharing computer infrastructure resources.

1.2 Motivation

There have been many studies and investigations into P2P networks both from an academic perspective as well as industry reportage. However most of these publications focus on a niche area of P2P applications. Most usage of P2P networks is in silos which tend to be isolated in their behaviour. They tend to resolve a particular issue or bring in certain feature sets to applications that would not exist otherwise.

This thesis introduces a horizontal perspective on vertical applications within industry. There is a need to examine P2P applications used in different verticals with a benchmark of commonality within them. The focus is not just pure technology or operations specific. The report describes novel usage of the concept of P2P networks within an application space keeping in mind the larger picture of a common base that these applications originate from.

1.3 Structure

The scope of this thesis encompasses a review of the technology behind Peer to Peer Networking and a focus on several case studies on different verticals and P2P applications therein. The major attention areas are business focused with a technology perspective.

We start from the networking basics and move on to middleware concepts and then finally on to the actual deployment of P2P networks in business applications.

We talk about a brief history of P2P and the different network topologies and communication models in use. We focus on a generic review on JXTA, an open standard P2P protocol. We then talk about traffic management in P2P networks and the practicalities and economics of the same.

Next is a brief discussion about generic application domains for P2P networks ranging from collaboration, distributed processing, content dissemination and knowledge management. Several case studies are included for review. Each case study focuses on one particular business application. We first talk about how P2P is used in Supply Chain Management information systems. We then analyse issues regarding Network Gaming using a peer enabled network. Further we look into a case study involving Video on Demand capabilities using P2P. Finally, on a business productivity perspective, we look at Groove, a collaboration and knowledge management tool using P2P.

A large part of the discussion focuses on an analytical case study that proposes a solution for large scale content dissemination for Mandriva, a computer software product company. Taking this ahead, we discuss about Skype using Voice over Internet Protocol using P2P.

Further discussion revolves on the possible future work that could be carried out using this report as a basis. We also discuss the likely trends that might evolve in P2P networking and applications.

2. PEER TO PEER NETWORKING BASICS

2.1 What is P2P?

Networking [1] is "A system of computers interconnected by telephone wires or other means in order to share information". Traditional approaches towards networking have always involved connecting one or more computers physically or otherwise to access anything ranging from processing power, memory resources and storage space to bandwidth and real time information. There has always been some sort of a central resource linking the elements of a network.

This is true for centralised databases and data warehousing to simple user oriented applications which run on the Internet. This approach which involves computers that utilize a central repository to run applications is known as the Client-Server Model.

A different approach towards connecting computers and applications online is the Peer-to-Peer Networking model. Popularly known as P2P, this system has revolutionized the way information sharing is done over the Internet. In this thesis we examine the use of P2P and present an analysis of the different business applications using P2P as well as the various end-user specific parameters.

In very simple terms, a peer-to-peer (or P2P) computer network is a network that relies on the computing power and bandwidth of the participants in the network rather than concentrating it in a relatively few servers [2].

In a P2P network each of the participants or peers is autonomous and depends on the other peers for functionality. Every peer receives and theoretically contributes an equal amount of content to the other.

A P2P network operates on the fringes of the Internet. This means that the P2P network utilizes resources of the individual peers rather than the connection itself. On the Internet, however, this would refer to a network of users who are part of the P2P network by virtue of using a common application to connect to each other say for example KaZaA [19]. KaZaA is a P2P application that runs on the FastTrack protocol.

From an operational perspective, a P2P network might use network resources ranging from unused processor cycles, storage space to available bandwidth depending upon the application running in a P2P fashion. Applications like KaZaA use bandwidth to look up media. Some other applications like SETI@Home [6] utilize the unused processor cycles to analyse and search for information. This is also known as Grid Computing wherein different computers can be linked together to form one large processing unit that can be used for resource intensive tasks like number crunching.

A fundamental aspect of P2P computing is that a fully decentralised P2P network cannot be controlled by any one single peer or for that matter any kid of network agent. Such a kind of network is continuously self evolving and is self stabilizing.

Two factors of important consideration here are the scalability and reliability issues due to the inherent dynamically changing structure of a P2P network.

2.2 P2P versus Client-Server Model

The advantages of a client-server model range from easy-updating capability, single source of availability, ease of administration to considerable control over access and restriction of content dissemination. The disadvantages of the traditional Client Server model are that this system is not scalable and has a major bottleneck when it comes to using it for a large number of clients. Bottlenecks might be due to bandwidth over-usage and they are seen from large system response times, high transaction times and errors in committing and rollback as well as system downtime due to overloading of resources.

A P2P network brings together users of different demographics and this in nature brings in resources and content from disparate sources which might not usually be available in a normal network. A P2P network has a tremendous advantage as compared to client server networks in terms of resource utilization and efficient distribution of bandwidth.

Collaboration is the key to the performance of a P2P network. This is in direct contrast to a traditional client server set-up wherein to access any kind of data, a client would need an intermediary, in this case a server that acts as a facilitator to other clients and/or services running on other servers or clients.

From a business service continuity perspective [22], the advantages of a P2P network over a Client Server network are illustrated below in Table 1.1

	Peer to Peer Network	Client Server Network
Infrastructure	Built entirely using the resources of the connected peers.	Relies entirely on the resources of the main central server.
Scalability	Network resources can be ramped up and down as required	With a single centralised resource, this cannot be done without downtime. (Especially useful in installing new equipment)
Availability	No single point of failure (except in P2P networks with a centralised indexing mechanism)	If the server goes down, the network ceases to operate.
Problem Resolution	It would be difficult to troubleshoot malfunction of the network since connectivity is between several clients.	Since the connectivity is between the client and the server, it is easier to focus on either to resolve issues.

Table 2.1: Major differences between Peer to Peer and Client Server Networks

With a pervasive presence of computers, P2P networking is bound to be used in most applications that require resources beyond the capacity of a traditional client server model.

2.3 P2P COMMUNICATION MODELS

From an application perspective, P2P communication approaches can be subdivided into centralised, de-centralised/pure and hybrid P2P topologies. The choice of an underlying topology would depend on the network parameters in focus including availability.

2.3.1 Centralised Directory Model

Centralised P2P applications use a centralised server to keep a track of the different users as well as the content being offered. These are the so called "First generation" P2P protocol based applications like that of the erstwhile Napster. As illustrated in Figure 2.1, the centralised repository only maintains a dynamically updated content redirection service that connects two different peers. The content is never stored in the intermediary.

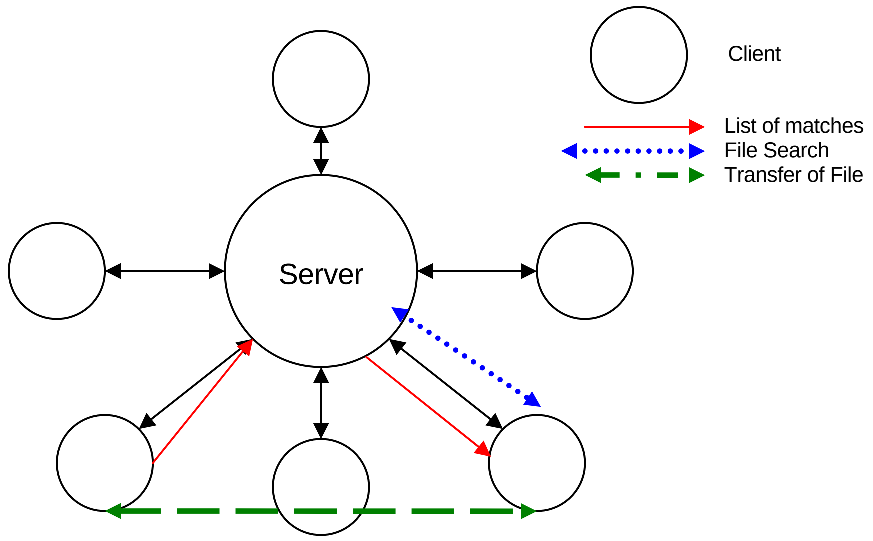

Figure 2.1: A Centralised P2P system with a server to facilitate data transfers

The above figure illustrates the centralised P2P distribution model. As shown in the figure, the Server is the single point of contact for ALL the clients and any querying done from any client has to pass via the server.

The server can support simultaneous file searching queries, responding with a list of matches as well as initiating and terminating the actual file transfer. This methodology is convenient to setup. However scalability is an issue with the fact that with growing number of peers, the load on the server might increase, inordinately resulting in unavailability of service to many clients.

2.3.2 Pure P2P Applications

Decentralised applications as illustrated in Figure 2.2 do not use any kind of a central repository to maintain information about the number of users, the shared content as well as the status. Each peer is in theory connected to several other peers via the nearest available peer. A common example of such a protocol is Gnutella.

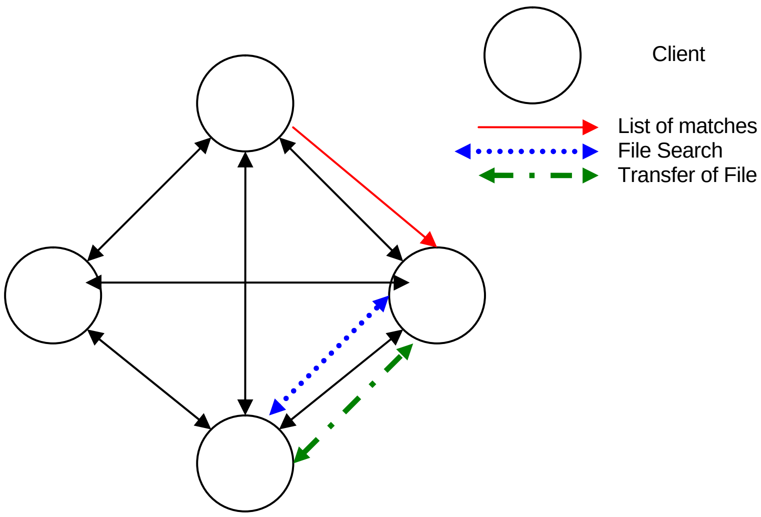

Figure 2.2: A De-centralised P2P system with peers directly connected to facilitate data transfers

2.3.3 Hybrid applications

These applications use a combination of both the technologies. This would be required in cases where a centralised index server would be required to co-ordinate searches and to maintain a list of active users. However, the server would just act as a pointer to the source of data and would not be actively involved in the data transfer which is done independently from the server.

2.4 P2P Network Topologies

Of interest at this juncture, from a network design perspective are the different network topologies that could be used to support a P2P network apart from the simple centralised and decentralised ones.

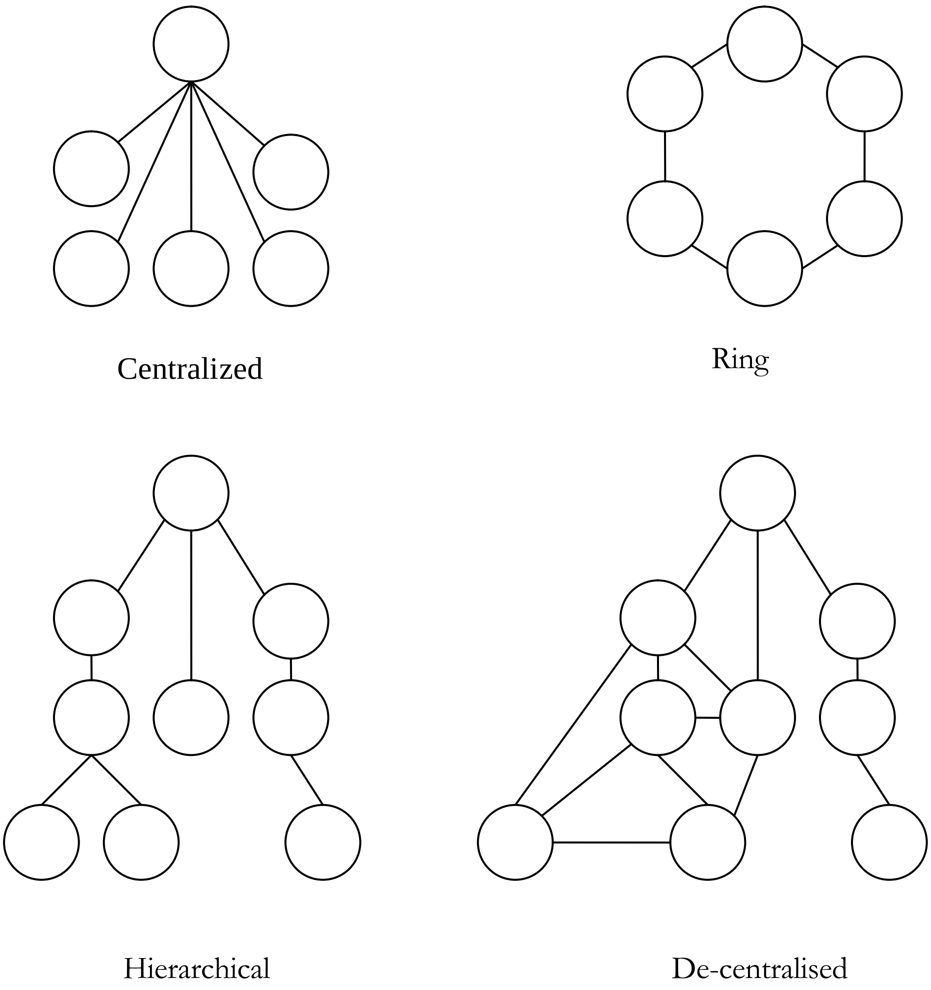

Figure 2.3: A schematic representation of the basic network topologies that could be used in a P2P network

The use of a P2P topology as illustrated in Figure 2.3 would depend upon the ultimate requirement of the content sharing environment. Also of importance is the consideration of hardware and software resources apart from bandwidth consumption and, finally, cost. The cases illustrated in Figure 2.3 are discussed in the following subsections.

2.4.1 Centralised Model

The central server maintains an index of all files stored on each node. This directory is constantly updated as and when peers connect or disconnect from the network. The index serves as a central point of control and routing. The flip side to it is that it also is a single point of failure.

2.4.2 Ring Model

There is no central server in such a model. Instead each peer is connected to each other via the next available peer. Information is routed by queries that hop from one peer to the other and the Chord [22] model of a P2P network is a strong example.

2.4.3 Hierarchical Model

In this model, we have the concept of "super-peers" [24] wherein each peer has a super-peer which is one level above that peer in hierarchy. This is a mix of the pure and hybrid P2P models. A well-known example of such a model is the DNS (Domain Name System). The DNS lookups are propagated across a hierarchy in a network. They return an IP address for a name lookup which is cascaded down across the hosts connected to the network.

The recent trend is towards a super peer hierarchical model [24]. Every peer in a super peer network would have its own super peer. Each super peer acts as a local centralised server that manages searching and linking peers under its hierarchy. This is then propagated across similarly with further peers downstream. Each super peer is a node that processes information from its localized cluster [24]. A super peer network combines the strengths of both centralised searching capabilities and the autonomy and load balancing techniques of decentralised networks. [24]

2.4.4 Multiple Network Topologies

Another interesting point to note here is the fact that the network topologies in a P2P network are not particularly restricted to the above basic four, but could also be a combination of one or more.

For example, a network might be a composite one which would include a Centralised + Ring combination or a Centralised + Decentralised format.

The different topologies might be used depending upon the environment variables as well as the functionality desired into the network. This is also of interest when factors like load-balancing and clustering are of importance. As suggested before, the topology to be used depends entirely on the application parameters.

Combining two different architectures makes a powerful choice in terms of manageability of the network, scalability, coherence, fault tolerance as well as security. For example a combination network might look as shown in Figure 2.4 below.

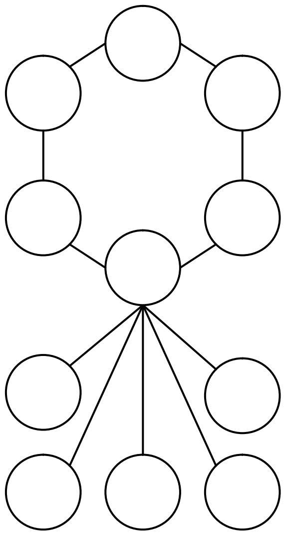

Figure 2.4: A representation of a Ring +Centralised combo network topology

The combination network as illustrated in Figure 2.4 would be used in a scenario where the peers in the ring network would be allowed to access the external centralised network only when the resource, in this case say content is not available with the other peers in the same ring. This mechanism would allow a system administrator to restrict P2P access beyond certain parameters and conserve network bandwidth OR to use this as a security firewall mechanism.

The following Table 2.1 illustrates currently available popular applications for the above discussed P2P network topology models in Figure 2.3 and Figure 2.4.

P2P Network Topology	**Example**
Centralised	Direct Connect, KaZaA [Public Domain]
Ring	Groove Networks [Secure Networks]
Hierarchical	Domain Name System [Legacy]
De-centralised	Emule, Edonkey [Anonymous usage]

Table 2.2: A list of popular P2P applications and the corresponding network topologies in use.

2.5 P2P Platforms

Having looked at different network topologies and communication models in P2P, we now focus on P2P platforms.

A P2P platform provides the middleware that talks to disparate applications and connects them in order to help data interchange.

JXTA is a popular P2P platform based on open standards. One of the major reasons for choosing JXTA as the candidate for our case study is the fact that JXTA is a completely open standard P2P platform that could potentially reside at the base of horizontal business applications over different vertical domains. This is illustrated by the framework shown in Figure 3.1 [3].

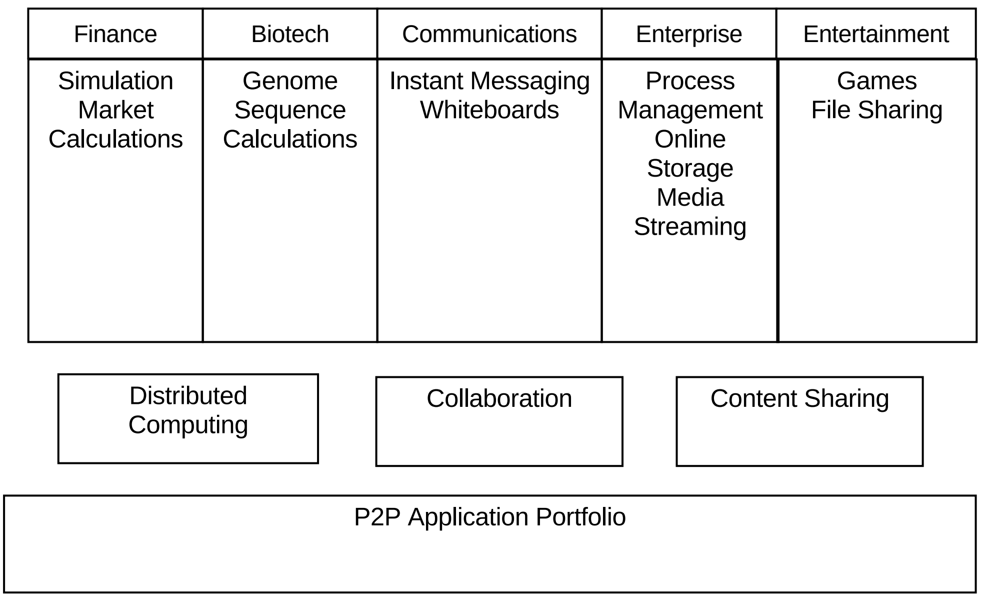

Figure 2.5: P2P Application Portfolio Framework

2.6 JXTA: A Generic Overview

JXTA in itself is not a P2P application. JXTA is a framework that enables running P2P applications over it. According to the Jxta.org website, "JXTA™ technology is a set of open protocols that allow any connected device on the network ranging from cell phones and wireless PDAs to PCs and servers to communicate and collaborate in a P2P manner." [4]

The JXTA framework provides a conceptual pipeline framework to connect different peers within the network. This data pipe allows one to one as well as one to many links between peers on the edge of the networks in an abstract fashion.[4] This brings in a great deal of flexibility in terms of building business applications since JXTA works around practical considerations of firewalls and network routes.

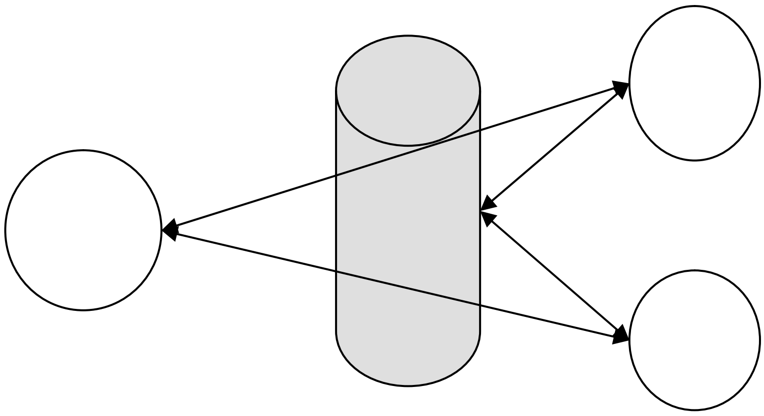

Figure 2.6: Graphical representation of JXTA data pipe

JXTA uses a terminology that includes a definition of Peer and Peer Group services. These services run as independent listeners that respond to the queries from the objects describing the different service parameters including system specifications and executable code.

These services are specified using XML, an open standard like that brings in a considerable amount of flexibility in cross-platform data transfer and compatibility. These services can then be called upon by different objects using a combination of a descriptor language and high level code.

JXTA is the underlying platform that provides an easy-to-use interface to develop P2P applications. This makes it quite attractive to develop business applications without having to worry about constraints like application portfolios or scenarios.

The following protocols are currently defined in JXTA [4]

• Peer Discovery Protocol
• Peer Resolver Protocol
• Peer Information Protocol
• Rendezvous Protocol
• Pipe Binding Protocol
• Endpoint Routing Protocol

Each of the above protocols specifies how the framework is defined in order to facilitate P2P applications over different platform independent scenarios. JXTA as a specification can be used as the underlying layer to develop new applications. The Peer Discovery Protocol is used to discover existing peers in the network. The Resolver Protocol assists in identifying the network parameters and locating the other peers. The Peer Information Protocol shares information about the data served on a particular peer. The Rendezvous protocol helps in information propagation. The Peer Binding Protocol binds one or more peers to a data pipe thereby facilitating data transfer. Finally the Endpoint Protocol determines the best routing mechanism using knowledge gathered from neighbouring peers.

Using a combination of unique identifiers and XML message advertisements, the JXTA framework provides an open standard, asynchronous communication channel using data pipes to enable building P2P applications on top of it.

3. EVOLUTION OF DISRUPTIVE TECHNOLOGIES

The term P2P has been probably coined just about a few years back but the underlying concept has been around for long.

P2P Networking has been around for quite a while and has been in use from the days of the infancy of the Internet. It was used for a more simple application though — just to exchange messages on a bulletin board, known as Usenet. [25]

DNS is a legacy application which is based on P2P networking concepts and is still in use today. Simple Mail Transfer Protocol (SMTP) works in conjunction with a relay server that examines the email address for the domain host and forwards it to the destination host using DNS lookups. SMTP might be described as loosely based on P2P since it is not self-organizing in nature.

Instant Messaging that became popular in the early 90s was also a form of P2P. ICQ [26] was the earliest IM application that was extremely popular. It was used to exchange short messages between users and later was developed further to exchange media also.

This technology was then adapted to create software programs that allowed users to share media. It just brings in the storage resources of all the users connected to the network and creates one massive storage archive of media. This concept has caught on so well that different models with slightly tweaked technical modifications have cropped up.

P2P reached cult status with the growing popularity of Napster, one of the earliest available programs for file sharing. Today, popular file sharing networks include KaZaA [19], Gnutella, iMesh [27], DirectConnect [28] and many more. The architecture varies and so do the file-sharing mechanisms. There are tweaks that work around firewalls and also try to ensure anonymity, but the underlying concept remains the same — sharing resources. The ease of use and more so, the ease of availability of media are behind the popularity of these networks. This method of acquiring data has revolutionized the Internet and is amazingly simple to use.

3.1 Managing P2P Network Costs

P2P in itself is a very useful technology but it needs to be channelled in the correct direction. In most P2P applications, the network is highly dynamic; in essence it is constantly changing, as the nodes join and leave frequently. Add to the fact that there are no central servers and the clients are unreliable because there is no central control to prevent peers from disconnecting from the P2P network. The last but not the least issue is that of security. The P2P system is built in such a way that the peers are either clients or servers depending upon whether they are downloading or uploading data.

In most cases up to 75% of the bandwidth usage in commercial broadband service providers is for P2P traffic [5]. In such a case, the usage of P2P actually denies a good Quality of Service to traditional FTP, HTTP and streaming services. This is a matter of tremendous concern to non-P2P users as well as ISPs.

One way of working around this is to use P2P Content Caching mechanisms which set in place a content cacher that runs on a particular port and listens to P2P traffic. All P2P requests are automatically routed via the cacher and if a local copy of the request is already available, the request is not sent beyond the local server. From an ISP point of view, this results in tremendous cost savings and also clears networks from congestion. The productivity of the system administrators is also increased as they can spend their time in efficient bandwidth management. This does bring into question the legality of caching copyrighted content on an ISP's servers. However, laws in the US absolve an ISP of any liability from copyright infringement on the part of its users.

In keeping with this trend, in the UK, Cachelogic, a company that provides caching software , NTL, one of the largest British ISPs and BitTorrent Inc. have signed up for a joint technology trial in caching P2P traffic. [39]

Most service providers have some sort of P2P traffic control policies in place. This might range from tracking and accounting Quality of Service levels for VoIP services running on SIP [34], H.323 [33] and other voice protocols to blocking file uploads or provisioning a percentage of the bandwidth for P2P applications. A commercial concept of tiered services brings in accountability to both users as well as service providers with dynamic allocation and billing of provisioned bandwidth.

3.2 Network Economics

At this stage of the report, we can now focus briefly on the economics behind a P2P network, incentives to share and how a network effect [17] affects traditional business models.

The concept of P2P as a disruptive technology has just about taken shape in the past decade. Technologically speaking, P2P is advanced not in terms of what it does, but in terms of what it can achieve in a manner that challenges existing norms and systems in place.

In any P2P network, there has to be an incentive for all the users to share resources. This is required to prevent free-riding and what is popularly known as bandwidth "leeching" wherein certain users simply download and never upload. Most P2P applications have some sort of restrictions in place which make it obligatory for users either to share bandwidth and/or storage space. Some applications like BitTorrent go one level further and set the downlink speed more or less equal to the uplink speed.

P2P networks challenge the very foundations of traditional usage of networks. From a network economics perspective, a P2P network is a new contender to the somewhat deeply entrenched traditional client server model. For any new concept to gain acceptance in a market, a critical mass is required.

In terms of network usage, P2P networking has reached and exceeded the critical mass required for acceptance. As opposed to traditional networks, P2P networks will not reach saturation or have any constraints in terms of resources. This is due to the unique feature of P2P that uses the network bandwidth and storage resources from all the peers involved rather than extract undue assets from any central resource. Certain advanced P2P networks like Pastry [20] are self organizing in nature and this is something that sets it apart from any other activity on the Internet.

4. P2P APPLICATION DOMAINS

P2P applications can and are being used for a variety of domains ranging from simple file sharing to resource-sharing for high-end computing. Of particular focus and of potential economic viability is the use of P2P in the following areas:

- Collaboration
- Distributed Processing
- Content Distribution
- Knowledge Management

An interesting fact is that each of these application domains has an overlap with the other when it comes to practical scenarios as illustrated by Figure 5.1. The dependencies on each domain would again be derived from that specific application domain. For example, a document management system that runs on a central Intranet would constitute part of a *Knowledge Management System*. The same system would be classified as a *Distributed Process* if the users of the system utilize their own computing resources and stored the output in the centralised database. This content when used by other concurrent users constitutes a *Content Distribution* scenario which again is a sub function of the *Knowledge Base* and also uses a *Distributed Process* for the content dissemination.

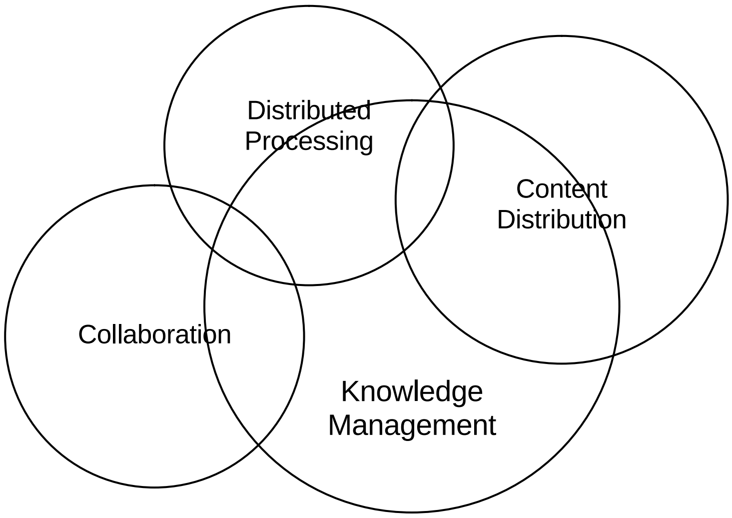

Figure 4.1: A representation of overlaps between P2P application domains

We look at each of the above areas in detail in the following subsections.

4.1 Collaboration

Collaborative applications are being used increasingly in workspaces and businesses. This has been made possible partly due to the proliferation of the Internet into our lives and the ubiquitous nature of the networked systems.

4.1.1 Functional Perspective

From a purely functional perspective, collaboration is effective in project management when the team is spread over geographically disparate locations. Collaborative applications help in change management and version control in terms of editing data or keeping track of new inputs. This has been done traditionally using systems like the *Concurrent Versioning System* (CVS). A step further would be a P2P version of such a system that would help change management in a real time scenario instead of a timed one that works on a scheduled basis using nightly builds.

4.1.2 Business Application

Collaboration is used by many corporates and organizations. Popular proprietary systems like Lotus Notes have been in use for many years now that include collaboration as a key feature. Also in place are applications like Documentum [40] which are used in key workflow based document management systems. Software applications enable employees to form ad hoc virtual workspaces, where individuals can share schedules and documents, conduct voice, video and text conversations, and perform other productivity tasks.

4.1.3 Case Study: Collaborative Workflow

As the use of the Internet grew, businesses operating in diverse geographies started exploring possibilities of operating umbrella applications that would link all their systems in the different locations. Thus the concept of collaborative workflow management emerged.

Traditional methods of distributed information retrieval and analysis used to be with the help of database systems that used nightly backups from different locations. However these were not "real time" applications and were simple carousel type collect and retrieve applications.

With new age acronyms like "Just in Time" [29] and "Six Sigma" [30] quality initiatives, real time collaboration is increasingly being used by large corporations with global identities.

Real time collaboration uses distributed information processing to gather, analyse and process data from various sources. Of particular interest is the Supply Chain Management (SCM) initiative since it involves various entities both upstream as well as downstream. A P2P system linking the different stakeholders of a SCM module would improve speed and streamline processes.

4.1.3.1 Traditional SCM systems versus P2P-based SCM

Enterprise Resource Planning (ERP) applications have specific modules in place to take care of SCM. Such a system has various stakeholders ranging from suppliers, independent equipment vendors, the organization itself, downstream value add partners, distribution channel partners and finally the retailers and end users. This is a generic view for any kind of manufacturing industry that sells to end users in a retail market. This might obviously change with different industry segments.

A traditional SCM system often uses a central data warehouse to collate data from all the different entities of the network as illustrated in Figure 4.2. This is done normally to separate the processing of the data from the actual collecting of information. Large SCM systems do an enormous amount of number crunching and this is generally done with the aid of mainframe systems and the like. From a functionality perspective, this system is excellent in terms of throughput and speed of analysis of data.

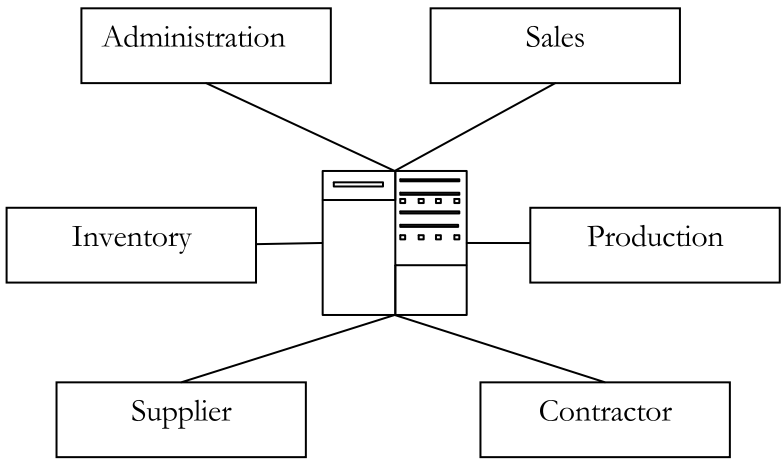

Figure 4.2: A representation of a traditional Supply Chain Management System with a centralised data warehouse

Business Logic however comes into play here and the connections are architected in such a way so as not to compromise data security and each sub system functions independently of the other and data interchange is generally on a "need to know" basis. For example, a supplier need not access any data about administrative issues but the opposite is true since administration would want to know payment schedule data from a supplier.

A P2P-based supply chain is an innovative approach and is far ahead of its time. Instead of using a centralised data repository, each of the entities in the SCM system can be connected to each other as individual peers as illustrated in Figure 4.3 which is discussed in the following subsections.

4.1.3.2 P2P Models in a Supply Chain Management Environment

Common models used in a SCM scenario with a P2P backbone are as illustrated in Table 4.1

One to One
A One to One model is used in a situation when there are just two entities that need to collaborate with each other. This is, at a macro level where both the entities can exchange information and data can be processed simultaneously to synchronize results at both ends. This simplified model can be replicated at each stage in order to collaborate effectively across upstream as well as downstream.
Hub & Spoke
A Hub & Spoke model is used where instead of each peer being connected to other, a collection of peers are connected to one single hub which acts as a “super peer” which is then connected to similar other hubs in a P2P manner.
Multiple Linked
A Multiple Linked model uses a true P2P network topology discussed earlier. Each of the entities is connected to each of the others to form a true functional P2P network model.
Agent Based
Agent Based models are used when an intermediary might be required to connect two peers. This might be a necessity when one or the other peers work on different protocols either in the network or application layer.

Table 4.1: A list of P2P models used in Supply Chain Management Scenarios

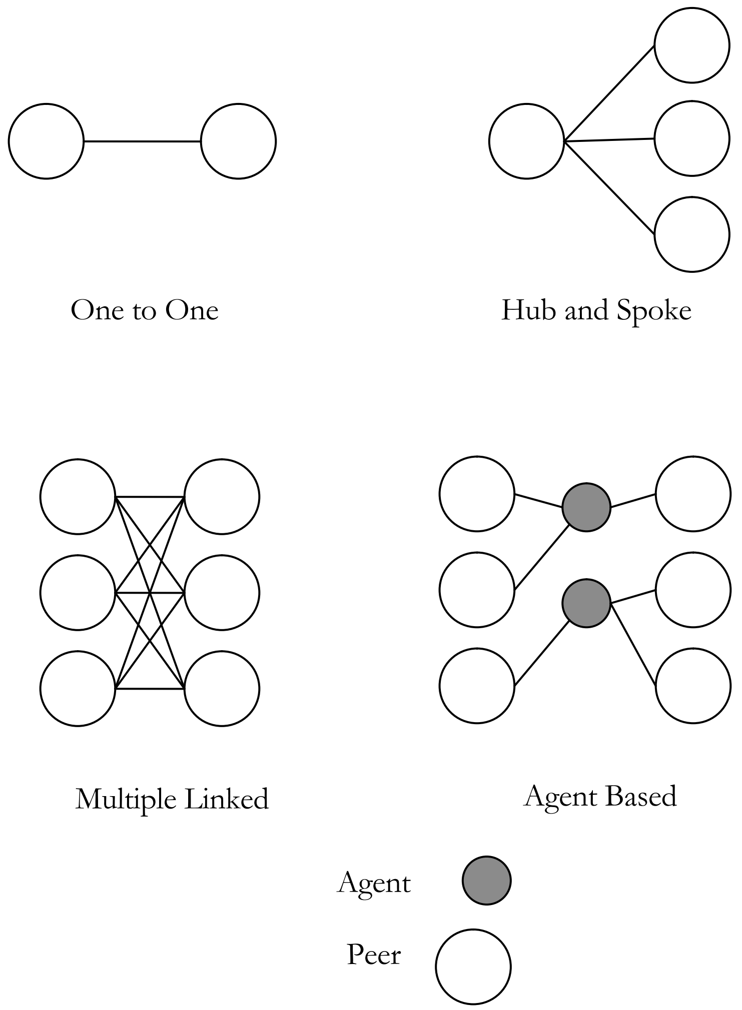

Figure 4.3: A schematic representation of a P2P-based Supply Chain Management System

Figure 4.3 above illustrates the different basic models of a P2P-based SCM system. Each of the different scenarios is used on a stand-alone or a combination basis depending upon the business logic needs. Also of particular consideration would be the practicalities of linking each of the entities in the network in that particular combination.

4.1.3.3 Practical Scenarios

A computer manufacturing company might have several plants based in different parts of the world. Each plant would be responsible for manufacturing different parts of a computer. One plant might specialize in hard disks, the other in memory chips and still another in motherboards or processors.

But in reality, most of the work is done via contract manufacturing and the different parts of the system are assembled together at one single place. This is similar to what Dell Computers does.

Collaborative systems here are used to determine the level of inventory, the safety level for ordering, co-ordination of various parts for assembly, final roll out and packaging as well as shipping mechanisms.

Wal-Mart, for example, achieved a 2 % improvement in retail store in-stock, a 14 % decrease in store-level inventory as well as a 32 % increase in sales and a 17 % increase in retail turns in their project. [8]

P2P systems are used in information sharing. Traditional centralised methods of procurement have now given way to one-one communication between suppliers, vendors and companies who can now share information with one another in the way the business needs deem fit.

Electronic marketplaces are being increasingly used by large companies like General Electric and the trend is also towards reverse auctions and global tendering. A P2P system that gives access to information in real time to all the stakeholders involved would be far more dynamic than a centralised warehouse repository which would need more infrastructure and policy issues in place.

This kind of a system is far more scalable and gives more control to the parties involved in the transaction. Modern ERP systems like SAP now support real time information processing from the supplier, vendor, shop floor, packaging and shipping up to administration.

From a practical perspective, many organizations are using P2P-based links in their SCM system without converting their entire SCM system on a P2P backbone but using such a network topology in parts. This is done to maintain scalability and still keep the system going on without any interruptions.

4.2 Distributed Processing

Distributed processing in terms of P2P would be the effective sharing of computing resources for large-scale resource-hungry applications ranging from complex arithmetical calculations to parallel processing.

4.2.1 Functional Perspective

Distributed processing applications share computing resources like idle CPU time, memory cycles and disk storage as well as in some cases Internet bandwidth. Resources are collated from the different peers and used to perform resource intensive applications. A common use of such a system would be analysis of weather and climate data. A popular project that has caught the fancy of many people is the SETI project [6] that assists in detecting possible extra terrestrial life.

4.2.2 Business Application

Distributed processing can result in tremendous cost savings for businesses which would otherwise need to invest in expensive hardware and/or software for their mission-critical applications. A common use of distributed processing is clustering where in theory several computers can be linked together to form one large system that would have the combined processing power and memory resources of all of them. This is used by many companies to run database servers and in some cases also file storage.

4.2.3 Case Study: P2P Network Gaming

Figure 4.4: A schematic representation of a gaming console and a 3D rendered gaming screen [51]

The gaming industry by itself is a huge market raking in more than US$31 billion globally. [14] The advent of the Internet expanded the reach of gaming and brought in a whole new dimension of collaborative gaming more popularly known as network gaming.

With increasing complexity in games including the use of 3D sound and high resolution graphics and imagery, games today have a great deal of sophistication both in terms of their content as well as the gamers who play them in the first place. Apart from traditional computer based games and old world arcade games, we now have console based units more popularly the Playstations and the Xboxes of the world. Nearly all the new generation games feature modular locales and graphics and multimedia which are either loaded from a hard disk / DVD or from the Internet.

Network gaming goes one step ahead and pull in real time content off the servers. It also allows players to connect with one another and use either text messaging or VoIP systems to communicate with each other.

A P2P-based network game is simply a logical step ahead as we will see in the forthcoming subsections.

An advanced game today comes with its own set of proprietary encoded information ranging from the different avatars, graphics, images, voice covers to the background music and the different online props. All this information is stored in traditional media like CDs or DVDs. On loading a game, this information is pulled through and rendered using 3D graphic card acceleration and surround sound output.

Certain games also come with built in update features that allow them to connect to a central server and download new features to the game. Taking this step further, games can also logon to a P2P network and discover other users on the same system who would be running the game.

The gaming application would then connect to different peers on the network and once the connections are established, the game could then be played in a multiplayer mode. Resource sharing here is of extreme importance since for effective networking gaming, bandwidth, disc storage and latency are all required in varying measures.

This kind of a P2P application would constantly use peer discovery agents to locate new game specific data and the game would constantly evolve as it is being played. This would introduce a lot of dynamic content into gaming and the game would be different each time it is played with different peers being connected to it.

Ultima Online, a role playing game by Origin Systems is a very good example of a peer based network game. Players connect to one of at least 25 geographically dispersed server clusters and communicate with the "persistent world" through the server [15]. Apart from that there are projects like JXTA chess [41] that work on the principle of peer discovery. The P2P system uses the bandwidth of the individual users to share game content and imagery. The client software at each user's end processes the graphics locally. Another such example of a real time virtual reality environment game using P2P would be Solipsis. [42]

This type of a new model in gaming can also mean further revenue in terms of subscription fees and network usage fees. Also in place are incentive based systems where the users are rewarded with either monetary incentives or new game features in exchange for letting other peers download from them. [43]

There are constantly evolving models that encourage game-based competition and scoring patterns that are tied down to exclusive feature downloads on the P2P network.

Pervasive gaming, a new trend, is the new age of network gaming. Pervasive games are defined as "new game experiences that are tightly interwoven with our everyday lives through the items, devices and people that surround us and the places that we inhabit." [16]

Pervasive gaming along with P2P networks bring in a great deal of business sense in terms of scalability and ease to market scenarios.

4.3 Content Distribution

Content Distribution in relation to P2P would be the use of the P2P mechanism to effectively disseminate content over the Internet. The focus would be on optimising bandwidth usage.

4.3.1 Functional Perspective

Traditional content distribution over the Internet would be a huge resource overhead in terms of bandwidth as well as infrastructure resources. An alternative is to use P2P to spread the content over the Internet on a geometrical proportion. This would optimise bandwidth from the provider's perspective. Popular P2P applications like KaZaA and BitTorrent do exactly that. Instead of having one single source for the content, we have several peers who mirror the information.

4.3.2 Business Application

Start-ups like the Swedish company Joltid [7] have in place proprietary technologies that allow content distributors to use P2P for effective and faster dissemination of data. Also some software product companies are using P2P as an effective method to distribute installations over a large user base. Content aggregation which is fast becoming the norm for service providers also uses P2P technologies. BitTorrent is being used by many open source vendors to distribute their software. Details of which we see in Chapter 5.

4.3.3 Case Study: Video on demand using P2P

Traditional Video on Demand (VoD) services work using a remote video server and a remote backbone which is connected to a subscriber network via a switching office.

The subscribers access content via a set top box which is programmed to decrypt the encrypted signals from the remote backbone. A set top box might also feature a VoD browser that is displayed onscreen for a user to select content as illustrated in Figure 4.5 below.

Figure 4.5: Sample Video on Demand Browser [44]

This system of television is in place in several countries in the world and it works on normal cable television networks. A new system on VoD using P2P networks is being built and feasibility studies and consumer trials are in place in Europe. A notable example would be that of the British Broadcasting Corporation (BBC) [45] that plans to use Peer to Peer technologies for remote video and audio streaming.

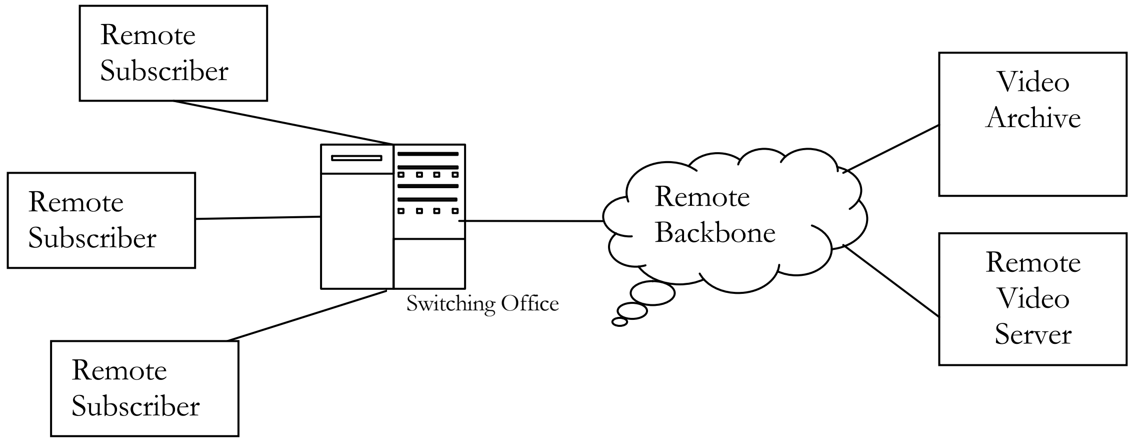

Figure 4.6: A representation of a Video on Demand Service Architecture

The BBC in conjunction with Kontiki (a commercial P2P services provider) has developed what it calls an interactive media player, or iMP. This software service allows users to choose content from a repository of TV and Radio programmes. The iMP extends the capabilities of a normal VoD player. It also functions as a P2P client in the background sharing content with other concurrent users as represented in Figure 4.6.

An innovative concept is the use of Digital Management Rights (DRM) software that validates content usage on a time expiry model. The BBC player automatically de-authorizes content that is more than 7 days old. Also in place is a geographical IP detection mechanism that restricts the usage of this system only to the residents of the United Kingdom.

The iMP uses what BBC calls an EPG or an Electronic Program Guide that dynamically lists content available from the past one week and future content for the next. Downloads can be scheduled and they become automatically available to other users on a P2P mechanism.

The iMP is still not publicly available. However a pilot trial was scheduled to run from September 2005 onwards until the end of the year. Users will have the facility to search and filter about 190 hours of TV programmes and 310 radio programmes, plus local content and selected feature films.

The BBC iMP went live in September 2005 and has presumably been successful enough. No official results however are available. Recent news reports [46] confirm that the BBC has stopped the trial period and has now signed a new deal with a UK trade organization to commercially broadcast BBC owned and copyrighted content on the P2P based player by the end of this year.

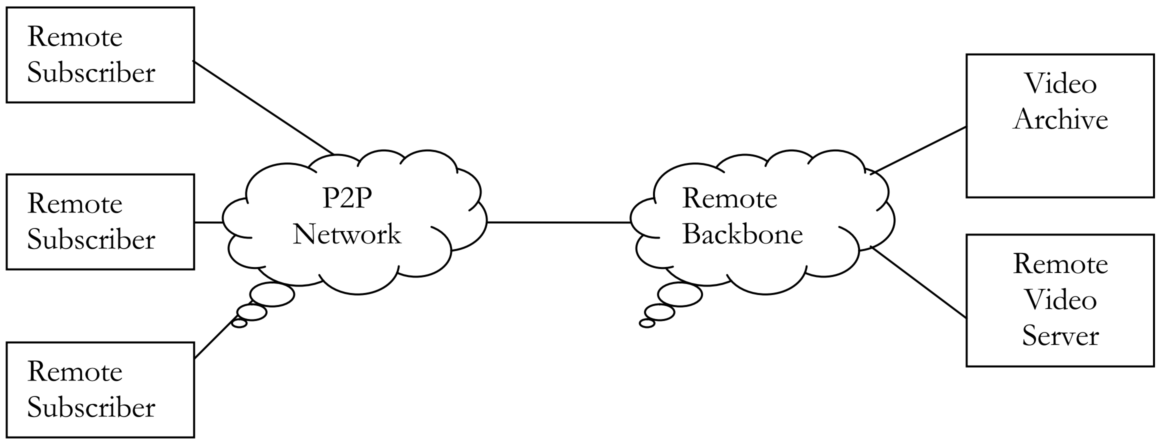

Figure 4.7: A representation of a P2P-based Video on Demand Service

The major difference between a standard VoD service and that of a P2P-based one is that of the subscriber user agent. A P2P-based VoD service downloads content in digital format on to the multimedia computer of the user while a traditional VoD supports screening on a television supported by a set top box as illustrated by Figure 4.7.

4.4 Knowledge Management

Knowledge Management is the effective use of content and information productively to attain common business and educational goals.

4.4.1 Functional Perspective

Information is redundant and is just data if it's not used properly. Knowledge Management is the use of software tools that automate collection of information as well as its dissemination. Of important focus here is the way the information is shared and not just the information itself. Knowledge management makes it easier to store and retrieve data, while at the same time giving access to the people with the proper credentials. This also ensures confidentiality as well as laying out processes for proper use of the information.

4.4.2 Business Applications

Businesses have traditionally used some sort of Knowledge Management application ranging from proprietary systems to popular products like Vignette [47]. Company wide portals on Intranets are used for simplifying tasks like human resource management and allocation as well as routine administrative tasks like holidays and time sheets. A company portal can serve as a single point of interaction for the employees, customers, partners as well as suppliers. Information is shared with each entity in accordance with his/her credentials. Knowledge Management can also be used in data mining and information retrieval. A P2P-based system for Knowledge Management would automate many processes in the system ranging from data inputs to "view" based information retrieval from a data warehouse.

4.4.3 Case Study: Collaborative Workspaces

P2P-based collaborative tools increase productivity by shortening time spans between changes made by a project team in a heterogeneous environment. Previous work [11] refers to collaboration scenarios covering different time spans: ad-hoc, short term and long term.

Ad-hoc collaboration refers to temporary and limited collaboration, short term refers to using common resources over a period of time to achieve a certain goal, and long term collaboration involves periods where system resources and users do not fluctuate heavily.

We look at a case study that includes a very popular commercially available P2P collaboration tool that is being used by project teams where the users are at disparate locations.

Groove Networks is a collaboration tool that runs on a pure P2P network. This software application recently acquired by Microsoft maps out content on the users' desktops and automatically synchronizes it to match changes made by others.

It also brings in a good deal of utilities ranging from instant messaging to file sharing. Groove Networks is the company behind Groove Virtual office, a P2P-based knowledge management and collaborative workspace tool. [12]

This software allows users to work with changes from others as if they were on the same physical location. The Groove Networks architecture works using a relay server that acts as an intermediary between several users. A schematic representation of the Groove architecture is as follows:

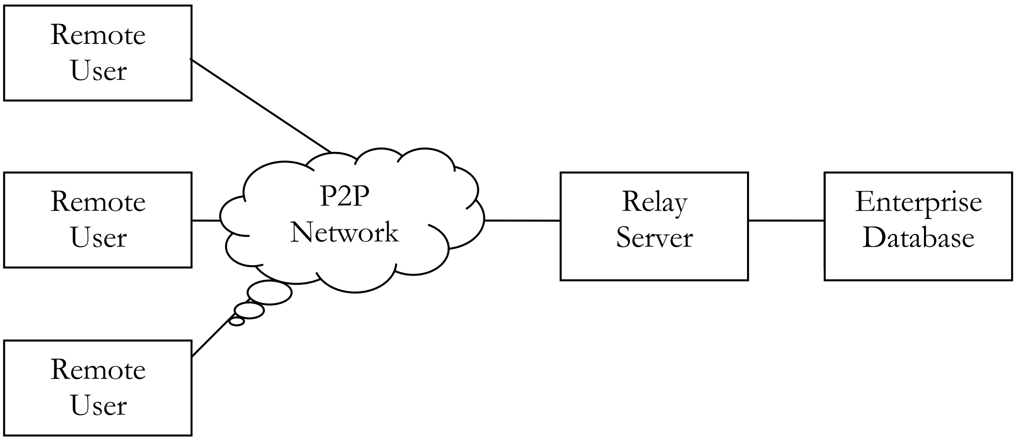

Figure 4.8: A schematic representation of the Groove Networks architecture

The unique feature about Groove is that it supports disconnected users and their content is synchronized as soon as they are online. The entire workspace profile is stored on the client and the relay server only stores changes in content and never the entire workspace. Any changes made by a particular user are immediately reconciled on the clients connected.

Groove uses an architecture that runs on a decentralised network using Internet Business Servers (IBSs) [13] that communicate using Conversational Process Managers (CPM) that provide decentralised, P2P process management.

A service discovery agent runs on each peer that automatically determines services running on other peers on the network using a single common port.

Figure 4.9 illustrates how a P2P network brings in connectivity to users in three different domains, all at the fringes of the network. The Document repository and database objects are available within the P2P network and are available seamlessly to users in all domains.

The services discovery agent is responsible for the discovery, initiation, control and termination of all services available within the P2P network.

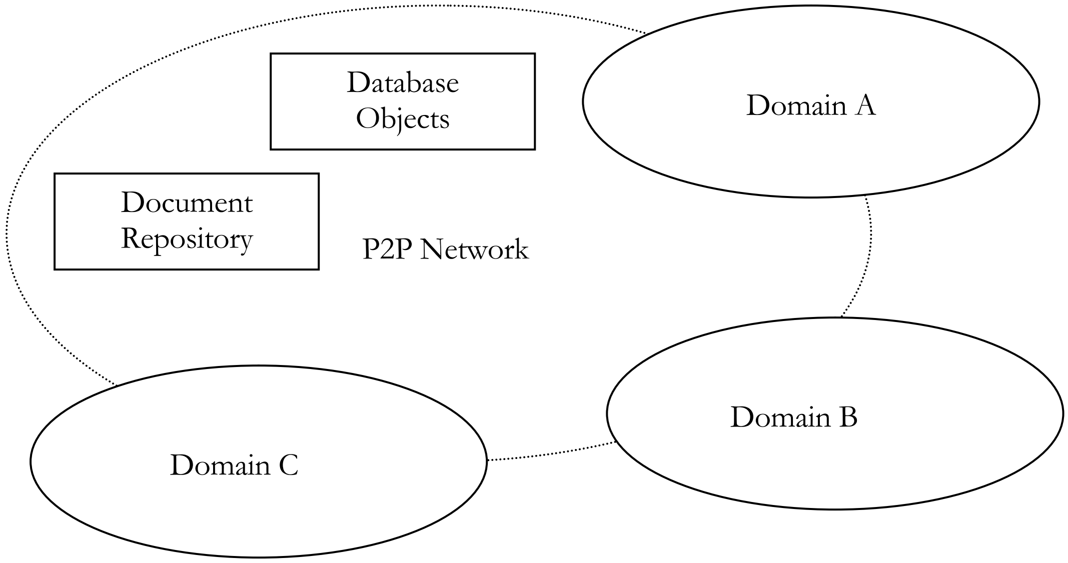

Figure 4.9: A schematic representation of a Knowledge Database running on a P2P Network

This brings in decentralization of resources within the network with different services like file sharing, print services, transaction servers, databases and many more being able to run independently.

This architecture is particularly relevant in offshore development scenarios and many large global companies are actively using or investigation options using commercially available software that integrates traditional mailing solutions, workgroup software, and a secure VPN-enabled P2P network.

4.5 Intermediate Summary

At this stage in the report, we have reviewed Peer to Peer networking as a technology which is a direct precursor to business applications.

We have dealt with the differences between P2P and traditional client server models, the background on P2P, its history and evolution and the various P2P network topologies. We have also focused on the major application areas under which P2P applications operate.

We have also focused on a P2P framework and how applications are built around it and particular business functions that are addressed in the market today using P2P applications.

We have looked at case studies which clearly demonstrate how the principle of P2P networking can be applied to different domains.

In the next chapter, we shall look at an investigative study in choosing the right P2P application for large scale content dissemination over the Internet. In subsequent chapters, we describe innovative usage of P2P and the future of applications running on P2P.

5. ANALYTICAL CASE STUDY

5.1 Statement of problem

Mandriva is the company behind Mandriva Linux their flagship open source product. Mandriva Linux being an open source software product is freely available on the GPL license over the Internet. The free-to-use download version of Mandriva Linux is available on global mirrors using FTP. The challenge is to identify an appropriate P2P application for use by the subscribers of the commercial download service for the optimised commercial version of Mandriva Linux.

5.2 Purpose of Study

Peer to Peer applications have a major benefit in decentralizing resources. This has a direct correlation with economic viability since there is no single entity which owns the resources. The total cost of ownership and responsibility is low. When there is one central repository of information, there is a huge pressure on it in terms of resources. This is especially true when it comes to high volume of usage.

For example Mandriva Linux has an active community of 4 million users. In even a hypothetical situation, if each of the 4 million users were to download the three binary images of the free download version of the Open Source operating system we would have to deal with a volume usage of 4000000 x 3 x 700 = 8400 Terabytes of data transfer which is staggering.

If those users would like to drain the images within 1 hour = 3600 s, we obtain 8400 TB/3600 s = 2.39 GBps = 19.1 Gbps -- which is hardly possible.

Hence plain HTTP/FTP download on a single server is simply ruled out.

Easier work arounds include having multiple FTP mirrors. At the same time the commercial versions of Mandriva Linux cannot be set-up on multiple mirrors to avoid risk of loss of revenue.

P2P applications bypass this bottleneck by using a very neat system of dividing the resource allocation between one and more number of users. P2P transfers scale well even on low bandwidth connections. For example if for example there is one single ISO image of a particular application available for download from a single server and there are twenty users who simultaneously need to download it.

A quick back of the envelope calculation shows us that for twenty users each downloading 700 MB of data the total bandwidth usage is 14 GB. The same if done via P2P will actually relieve the source of the bandwidth strain and with a random function that is used will divide the 700 MB amongst the 20 users to set the actual bandwidth used to be only 700 MB in total and the rest of the file is actually downloaded amongst the 20 downloaders since each of them ideally has 1/20th part of the file already on their systems and they don't need to reach the source but rather exchange the information amongst themselves. So in a nutshell we actually have saved bandwidth to the tune of 95 %.

The Mandriva Linux edition consists of CD binary images ranging from 3 to 8 depending on the flavour, each of circa 700 MB each. The challenge here is to use a suitable P2P application that can allow the 4 million users of Mandriva Linux [9] a fast, reliable and convenient method of access as well as provide the paying user community a fast, fair and reliable service. Mandriva uses differential service models for its users. The online community that pays subscription fees is given priority in terms of access and early previews as well as commercial tool sets bundled with the operating-system editions. The general open source community has access to the free versions of the software.

5.3 Methodology for Selection of Applications

To determine which application could be used to deliver the Linux distribution across the Internet, we consider the use of four popular applications and decide on basis of performance metrics as well as cost overheads. The four applications are chosen on the basis of end user feedback in terms of popularity as indicated by the survey found in Appendix 1.

- Limewire [31] on Gnutella
- iMesh [27]
- BitTorrent [32]
- KaZaA [19]

We now look at an overview of each P2P application on the basis of the following basic requirements:

- The ability to discover other peers
- The ability to query other peers
- The ability to share content with other peers

5.3.1 Limewire

Limewire is a lightweight P2P client. A feature is "Swarm" download [31] that allows a user to download from multiple sources simultaneously. Anonymity features are available that allow users to mask their identity using third party web proxies. It uses GWebCache, a distributed connection system for searching over a Gnutella based network.

Figure 5.1: Limewire running on Gnutella [48]

Advantages:

- No bundled ad-ware or spy-ware
- Very fast on local networks
- High level of usability features
- Anonymity built-in

Disadvantages:

- Bandwidth usage is high when the number of search hits are less
- Available only on Windows

5.3.2 iMesh

iMesh [27] is another standard P2P application utility that helps users to download content off other users simultaneously sharing information. It connects to its own network as well as to Gnutella to search and look up content.

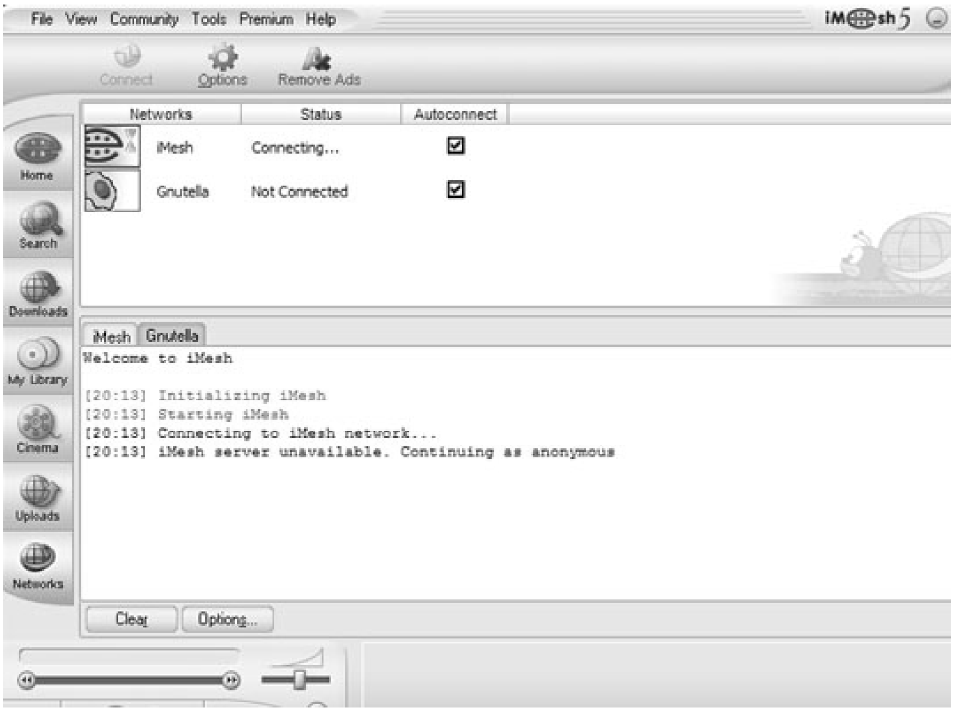

Figure 5.2: iMesh initialising and connecting to the iMesh Network

Advantages:

- Sleek interface
- Built-in media browser
- Multiple network capabilities (can connect to its own network as well as Gnutella)
- Simultaneous downloads from multiple locations

Disadvantages:

- Searches are slow as compared to other tools

5.3.3 KaZaA

One of the earliest P2P applications around, KaZaA was and is to a certain extent popular even now.

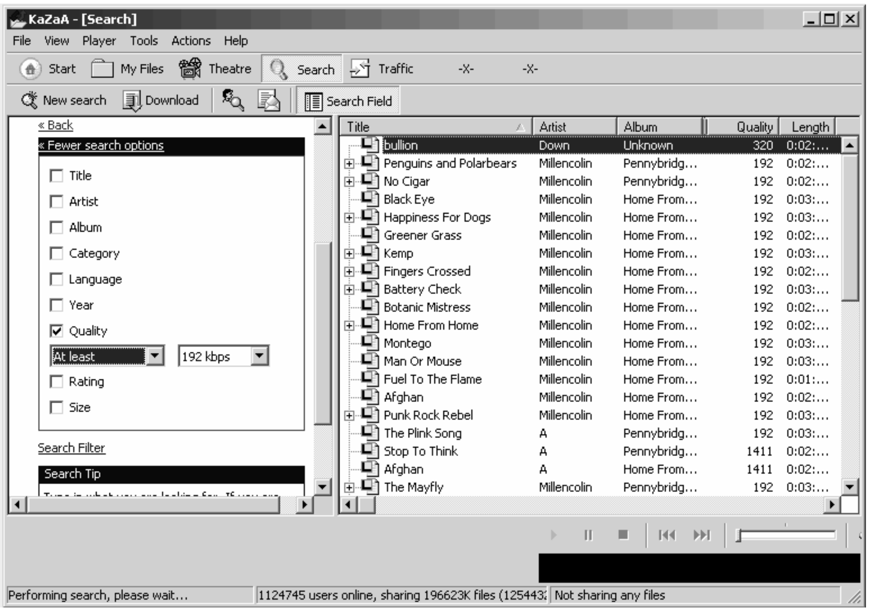

Figure 5.3: KaZaA connected through its proprietary Fast Track protocol [48]

Advantages:

- Fast client
- Built in media browser
- Multi-lingual user interfaces
- Quick search capabilities

Disadvantages:

- Ad-ware comes installed with the client
- Sponsored results dilute the quality of searches

5.3.4 BitTorrent

BitTorrent [32] is a neat utility designed to download any media off a web based location .BitTorrent works on a unique principle of simultaneous upload and download. This is done to maintain a collaborative scheme of access to media.

BitTorrent works on the principle of give and take, i.e. the download speed of the media is directly proportional to the upload speed. BitTorrent was designed by Bram Cohen and is one of the world's most widely used software application for distributed downloads with minimal overheads on the bandwidth of the source. BitTorrent distributes file parts across a network and prevents bottlenecks associated with having everything centralised on a single server.

Advantages:

- Thin client, does not require too much system resources
- Uses unused upload bandwidth of the users to facilitate download for other users

Disadvantages:

- Since the speed of download depends on the upload speed contributed by the user, lower end users on dialup connections or on low bandwidth connections would find it difficult to download larger files in a short period of time especially in asymmetrical situations (e.g. ADSL).

5.4 Selection of Parameters

We compare the network parameters as well as usability issues in each application. To compare the network parameters, we use free-to-use tools to measure bandwidth usage and monitor network traffic.

Apart from an analysis of network parameters in focus, we also need to look at usability features of the P2P application. This would be done by analysing user feedback from a sample population with a fair mix of novice as well as experienced users. The survey sheet is listed in the Appendix.

We compare the following parameters in each case:

- Speed
- Reliability
- Accuracy
- Ease of use
- End user infrastructure requirements
- Security
- Platforms supported
- Bandwidth Constraints
- Provider Infrastructure support required
- Volumes supported
- Openness of Protocol
- Future support
- Legality

5.5 Survey Results

The respondents were sent an email with a referrer link on it. The email was sent to 56 people of which 31 chose to respond. Once the fields were filled, the total score was calculated and sent via email.

Each parameter was rated on a scale of 1 to 10 with higher figures for better performance.

A graph plotted to compare the selective total scores reveals BitTorrent as the most favoured application by all the users in the selected sample. The results are shown in Figure 5.4.

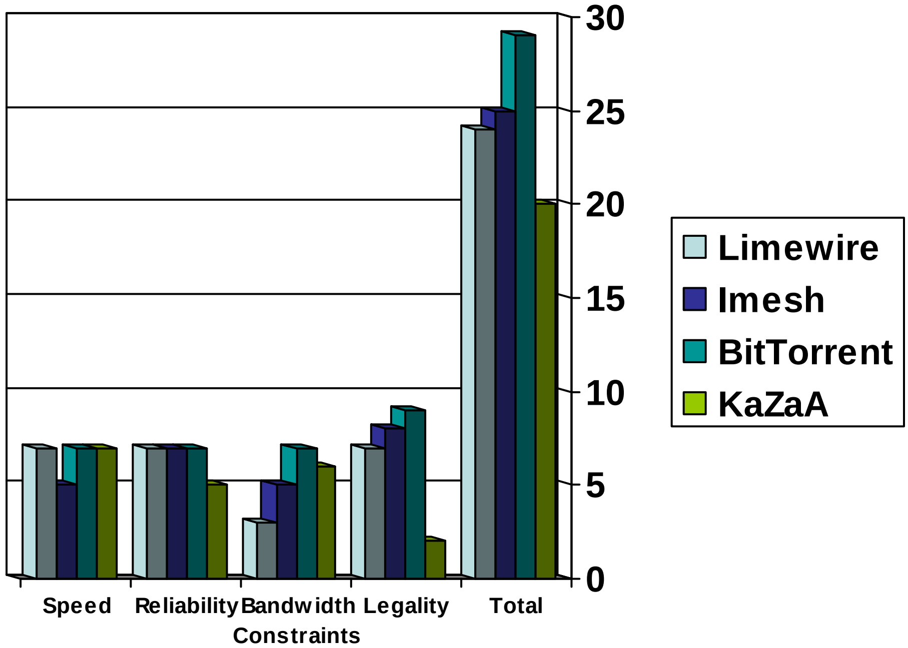

Figure 5.4: A graph showing a comparison in scoring patterns as results of the survey

The major drawback in Limewire was speed. iMesh lost due to bundled ad-ware/spy-ware and so did KaZaA. BitTorrent was the most reliable application though it did force users to upload simultaneously. However, that did not prevent 29 users to rate it at 9 for legality which was the major decisive factor.

5.6 Findings & Discussions

A comparison of the considered P2P applications has proven that BitTorrent is the best choice for content distribution where the media size and the demand for the media are huge.

A major advantage of BitTorrent is its legality. Since there is no "search" function built into it, the user would have to know the source of the .torrent file beforehand. Taking also into account the fact that there is no attempt made to conceal the origin and source of the media, this system is win-win for both copyright holders as well as end users who want to share their data.

This along with the ease of set-up of a BitTorrent enabled media source has led many Linux distribution companies and other free OS projects to use BitTorrent exclusively for distributing their content.

In terms of load-balancing, BitTorrent by its inherent nature doesn't stress the original source. When each user has one part of a file and they upload to one another, in a matter of time, each user ends up getting the entire file. A major advantage is that when the number of users increases, the speed of the download also increases due to the "upload-download" concept wherein the speed of downloading is set to be proportional to the speed of uploading. With hashing functions available, users of BitTorrent can be assured of the reliability of the data. The BitTorrent interface is extremely easy to use and in theory an infinite amount of volume of data can be supported.

BitTorrent is of particular interest for Mandriva because it decentralises the entire procedure of downloading the operating system. This also fosters in the community and those new to it a sense of collaboration which in itself is an essence of the Open Source spirit.

5.7 How BitTorrent is used at Mandriva

A Mirror Server is generally used which has the exact replica of the content in the original server. A commonly used mechanism is the "Rsync" script. This is set-up in the Linux cron job and as a nightly backup, the entire distribution and the RPM repository is mirrored across the world.

The procedure for initiating the Mandriva Linux OS for Internet based distribution is as follows:

- The system administrator at the company "seeds" the content on to the web server.
- The "seed" is a ".torrent" file which points to a "tracker" file that contains a list of all available peers containing the complete or part of the file.
- The "tracker" contains the IP address, files available and file sizes of the peer as well as a "Hash" code that is nothing but a sequence of numbers used to uniquely identify that particular user.

This particular "tracker" is then responsible for determining the nearest peer to the downloading peer and the BitTorrent client simultaneously starts uploading and downloading the distribution binary image.

6. P2P USAGE IN DISPARATE VERTICALS

6.1 Voice over IP over P2P

When one talks about P2P applications, one normally refers to data or file exchange. However, information need not necessarily be confined to just files i.e. text or binary data, it can even be voice. Voice information encoded over the IP network is exactly similar to any other plain text or binary information sent over the Internet.

The basic principle behind VoIP as illustrated in Figure 6.1 is to digitise the voice packets and transmit them via the IP backbone. At both ends we have an Analogue-Digital Converter (ADC) and a Digital to Analogue Converter (DAC) that does the translation.

"TCP/IP networks are made of IP packets containing a header (to control communication) and a payload to transport data: VoIP uses it to go across the network and come to destination." [10]

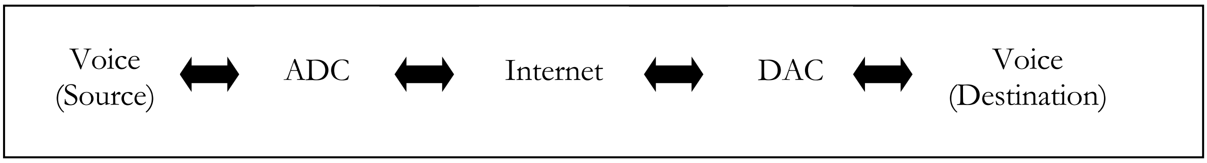

Figure 6.1:Voice over IP basic principles

VoIP calls are routed via the Internet using the H.323 [33] protocol stack. This protocol stack works on OSI protocol layers 2 – 5. It is obvious that such a system requires a residential gateways and expensive Customer Premises Equipment (CPE). This is also quite an interesting business case for exclusive VoIP service providers.

6.1.1 Evolution of Voice over IP Systems

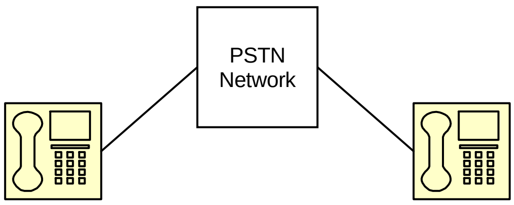

Figure 6.2: Plain Old Telephone System with PSTN Network

Figure 6.2 illustrates how a normal Packet Switched Telephone Network works. An analog telephone when used to call another party connects via the local telephone exchange and the call is routed via the PSTN exchange. This mechanism at the last mile end is completely analog in nature.

A Voice over IP (VoIP) system uses an intermediate voice server or digital PABX that runs on a circuit switching mechanism instead of the packet switched network. The analogue voice information is converted into digital information. A typical voice channel requires about 64kbps bandwidth. Each channel can support about 8 calls simultaneously from a speech coding perspective.

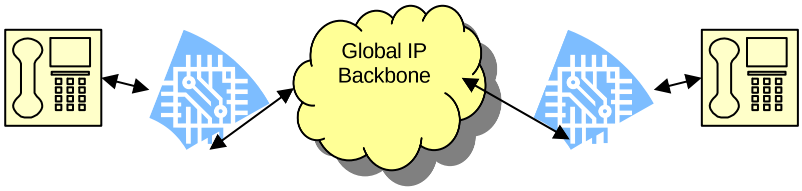

Figure 6.3: Voice over IP system with computer enabled analogue to digital conversion

Compression codecs from vendors [49] can be used which can enhance voice quality and at the same time use up less bandwidth. This VoIP system is used by large corporations for direct inward dialling or to route internal calls over Integrated Private Lease Circuits.

6.1.2 Next Step: VoIP over Peer to Peer

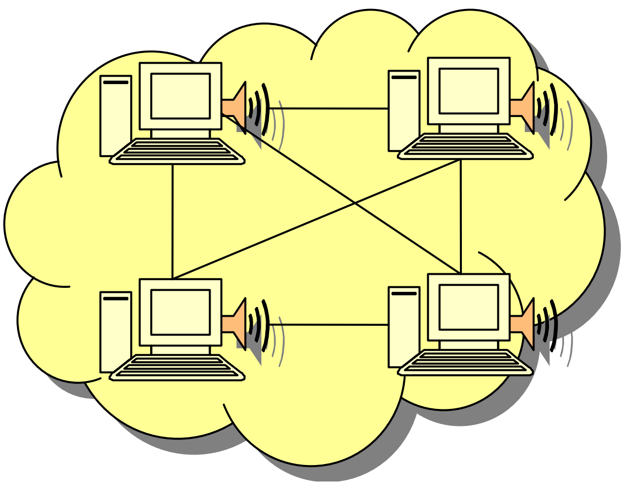

Figure 6.4: Peer to Peer based Voice network running over the global IP network

Figure 6.1 and Figure 6.2 illustrate the use of PSTN and VoIP networks. Most Telecom Service Providers use a combination of both of the above to serve an optimum mix of cost effective and reliable calling services.

Cellular Service Providers also use VoIP to connect long distance calls as it is extremely cost effective. However all of the above scenarios entail using the PSTN network, one way or the other. This does not in any way bring down costs to a level that is negligible.

The next logical step would be to use the P2P functionality over a VoIP network and build applications that use bandwidth optimally and route calls faster without using a centralised mechanism but running on the global Internet backbone as illustrated in Figure 6.4. This also helps the business application without compromising on the end results.

Such an application would use the multimedia features of a standard PC connected to a P2P network and send and receive voice data over the P2P network similar to that of any standard P2P application uploading or downloading content.

6.1.3 P2P-SIP Architecture

Previous studies [34] have shown that Session Initiation Protocol (SIP) could be used to transport voice across the Internet.

A P2P-SIP architecture has advantages that range from basic user registration and call set-up as well as advanced services such as offline message delivery, voice/video mails and multi-party conferencing [35].

Such an architecture would also be interoperable with any existing SIP gateway and also brings in the scalability and reliability aspects of a standard P2P network.

Traditional H.323 [33] based IP telephony systems use a combination of user registration and the DNS system to locate the other users. On the other hand, a P2P-SIP Architecture uses the concept of a ring-networked Distributed Hash Table [22] like Chord [22].

6.2 Case Study: Skype, VoIP application over P2P

We now look at examples of non conventional usage of P2P. Our previous case studies focused on P2P networks in specialized computer applications. The following case study is unique in that it talks about synergy between the concepts of P2P, file-sharing and VOIP.

Skype [37] is a P2P application that combines the use of disruptive technologies with traditional VoIP features. It runs just like any other instant messaging application on a computer desktop. It also features text messaging.

This is not so for the other Instant Messaging clients where the voice quality is not pretty good. However, dial up users might notice a small degree of latency in the call and Skype works better on broadband connections. Still, it delivers a reasonable quality even on UMTS networks, [50] Thus, a UMTS-enabled device with Skype installed can in principle replace a mobile phone.

Skype scales pretty well even when the user has a low bandwidth connection. It uses an intelligent routing protocol to utilize the best available route to the other user and also uses compression algorithms.

Figure 6.5: Skype running on its own proprietary P2P network, showing about 1.75 million users simultaneously logged on [37].

Skype is superior to other Voice over IP clients in the sense that, it does not need a central server and, resources allocated to a particular call are not server centric.

When a user is signed on to Skype, the application listens on a particular port (which is user-configurable) for any calls. When one user tries to call the other, a direct connection between both is established. This ensures quicker call completion rates and better voice quality. Something unique about Skype is that it uses a non-fire-walled user as an intermediary when a particular user is behind a firewall and cannot connect directly.

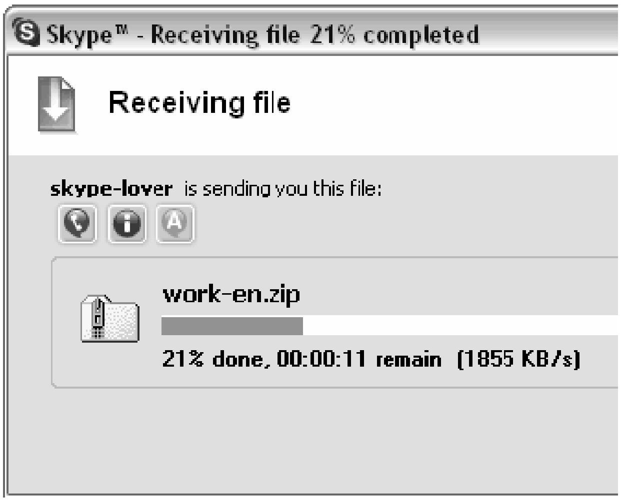

Figure 6.6: Skype running with a P2P enabled file transfer [37]

Skype marries the file sharing technology with voice in a very elegant manner. As expected Skype also facilitates one-to-one file sharing.

Skype uses a proprietary technology to look up other users on its network [37].

The latest versions of Skype now are available in different operating system platforms and have also gone many steps forward in introducing interfaces with video calling as well as with traditional PSTN and GSM phone lines. Previous studies [38] have shown that the voice quality in Skype is better than that in other existing VoIP instant messaging clients like MSN or Yahoo!

7. CONCLUSION

P2P technologies as we see today are evolving continuously even as we speak. With new protocols and clients being developed, this technology would go much beyond the parameters of the Internet as we know it.

But again, due to different mechanisms and different proprietary as well as open source protocols, there would soon be a move to an open standard for P2P applications on the Internet much along the lines of open web standards.

The use of P2P is not limited to file sharing alone as many people might believe; rather the near future would see an efficient use of P2P enabling technologies on a lot of cost saving ventures as well as in a considerable simplification of content dissemination.

P2P has changed the way we look at content. The move now would be towards looking at media content as a service rather than a product. Just as text and images are available on the Internet, so would be audio and video content. Due to obvious conflict of interests amongst major media companies and the actual artists and the end users, of course, a legal framework for usage of P2P applications shall take some time to come.

Future work in this area could be to look at Return on Investment parameters based on a comparison between running the same application on a traditional Client Server Model and running it as a P2P enabled one. Data modelling can be done to visualize how network parameters would interact and we could analyse service level parameters based on empirical data.

P2P in itself is only a novel use of technology. Business applications working around P2P need to be innovative enough to offer value for money for consumers. Next generation content networks must address the issue of unique and dynamic content on occasionally connected and mobile content providers [18].

Apart from technological advances in the P2P arena, there is always a lingering fear of regulatory hurdles. This is expected to be an eternal cat-and-mouse game with each side trying to better the other. In the end, what needs to be seen is whether technology or policy prevails.

Appendix

Survey Screen

The screenshot below is a replica of the P2P application preferences survey conducted in Mandriva.

Mandrakesoft Quick P2P survey

This is a mini survey to collect information on the suitability of a P2P application for enabling Mandrakelinux downloads on the Internet

Please rate the following programs on a scale of 1 to 10. (1 = Poor, 10 = Excellent)

	Limewire	Imesh	BitTorrent	KaZaA	Total
Speed					
Reliability					
Accuracy					
Ease of use					
End user infrastructure requirements					
Security					
Platforms supported					
Bandwidth Constraints					
Provider Infrastructure support required					
Volumes supported					
Openness of Protocol					
Future support					
Legality					

Calculate Total Scores | Submit

If you have any questions please contact me at mahesh@mandrakesoft.com

References

[1] Definition of networking from http://www.dictionary.com as on November 2, 2005
[2] http://www.wikipedia.org as on November 2, 2005
[3] Milojicic, D., Kalogeraki, V., Lukose, R., Nagaraja, K., Pruyne, J., Richard, B., Rollins, S., Xu, Z.: Peer-to-peer computing. Technical report, HP (2002)
http: //www.hpl.hp.com/techreports/2002/HPL-2002-57.pdf
[4] http://www.JXTA.org as on November 2, 2005
[5] A. Parker. The true picture of peer-to-peer filesharing, 2004. http://www.cachelogic.com/
[6] SETI @ Home project at http://setiweb.ssl.berkeley.edu/ as on November 2, 2005
[7] http://www.joltid.com as on November 2, 2005
[8] J. Småros & K. Främling, Peer-to-peer information systems: An enabler of collaborative planning, forecasting and replenishment, Logistics Research Network 6th Annual Conf., Edinburgh, UK, 2001.
[9] Numbers as in Autumn 2004
[10]VoIP Howto, Linux Documentation Project Arcomano, 2002
[11] Hauswirth, Podnar & Decker, Proceedings of the 3rd International Workshop on Distributed and Mobile Collaboration (DMC 2005) Linkoping, Sweden, 2005.
[12]http://www.groove.net/home/index.cfm as on July 03, 2006
[13]Chen, Hsu & Dayal (Software Technology Laboratory, HP Labs, HPL-2001-14)
[14]http://en.wikipedia.org/wiki/Video_game_industry as on November 2, 2005
[15] Peer-to-Peer Suitability for Collaborative Multiplayer Games , Saddik & Dufour, MCR Lab, University of Ottawa in Proceedings of the Seventh IEEE International Symposium on Distributed Simulation and Real Time Applications, Delft, Netherland 25th - 28th Oct. 2003.
[16]IPerG, a consortium of which the Swedish Institute of Computer Sciences is a part of
[17] Network Effects, Nonlinear Pricing and Entry Deterrence, by Arun Sundararajan, Working Paper EC-03-17, Stern School of Business, N.Y.U.
[18]Ram Sadasiv,Quantum Stream at The O'Reilly Peer-to-Peer and Web Services Conference Washington, D.C. -- November 5-8, 2001
[19]http://www.KaZaA.com as on November 2, 2005
[20] A. Rowstron and P. Druschel, "Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems". IFIP/ACM International Conference on Distributed Systems Platforms (Middleware), Heidelberg, Germany, pages 329-350, November, 2001
[21]Office of Government Commerce, UK, http://www.itil.co.uk/ as on July 03, 2006
[22] I Stoica, et al "Chord: A Scalable Peer-to-peer Lookup Protocol for Internet Applications". IEEE/ACM Transactions on Networking, Vol 11, No 1 (Feb 2003)
[24]Yang et al, Designing a Super-Peer Network, 19th International Conference on Data Engineering (ICDE'03), Stanford University
[25] Stephen Daniel, James Ellis, and Tom Truscott (1980). USENET - A General Access UNIX® Network
[26]http://www.icq.com as on November 2, 2005
[27] http://www.imesh.com as on November 2, 2005
[28] Jon Hess, Neomodus, 1999 http://www.neo-modus.com as on July 3, 2006
[29] Just in Time manufacturing has roots in Japanese mass production initiatves. Also has been used extensively by automotive manufacturers including Toyota and Ford over the years
[30]Six Sigma, lean manufacturing process, Bill Smith, Motorola 1986 http://www.motorola.com/content/0,,3074-5804,00.html#ss
[31] Swam Download Features : http://www.limewire.com/english/content/home.shtml as on November 2, 2005

[32]http://www.bittorrent.com, http://bitconjurer.org/ Bram Cohen as on November 2, 2005
[33] J. Toga and J. Ott. ITU-T standardization activities for interactive multimedia communications on packetbased networks: H.323 and related recommendations. Computer Networks and ISDN Systems, 31(3):205– 223, Feb. 1999.
[34] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. R.Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler. SIP: session initiation protocol. RFC 3261, IETF, June 2002.
[35] Kundan Singh and Henning Schulzrinne, "Peer-to-peer Internet Telephony using SIP", Columbia University Technical Report CUCS-044-04, New York, NY, Oct 2004.
http://www.cs.columbia.edu/~library/TR-repository/reports/reports-2004/cucs-044-04.pdf
[36] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Shenker. A Scalable Content Addressable Network. In Proceedings of the ACM SIGCOMM, San Diego, CA, USA, August 2001.
[37] http://www.skype.com/ as on November 2, 2005
[38] Baset & Schulzrinne, Columbia University, September 15, 2004
http://www.cs.columbia.edu/techreports/cucs-039-04.pdf as on July 3, 2006
[39] Cachelogic press release on their website at http://www.cachelogic.com/news/pr100206.php as on April 18, 2006
[40] EMC Software Products web site at
http://software.emc.com/products/content_management/documentum_family/architecture/architecture.htm as on April 18, 2006
[41] http://chess.jxta.org/ as on July 3, 2006
[42] Keller J. and Simon G., Toward a Peer-to-peer Shared Virtual Reality. In 22nd International Conference on Distributed Computing Systems (ICDCS02), pages 595 -601. IEEE Computer Society, Jul 2002.
[43] Chiranjeeb Buragohain, Divyakant Agrawal, Subhash Suri; A Game Theoretic Framework for Incentives in P2P Systems, at the proceedings of the Third International Conference on P2P Computing (P2P2003), Linkoping Sweden, 2003
[44] Pic Source : http://www.informitv.com as on November 2, 2005
[45] BBC Press Release :
http://www.bbc.co.uk/pressoffice/pressreleases/stories/2005/05_may/16/imp.shtml
as on May 16, 2005
[46] http://www.theregister.co.uk/2006/06/08/bbc_pact_deal/ as on July 3, 2006
[47] http://www.vignette.com/ as on July 3, 2006
[48] http://www.mpex.net as on November 2, 2005
[49] Vendors like Skype use proprietary voice codecs from commercial providers like Global IP Sound; http://www.globalipsound.com/
[50] T. Hoßfeld, A. Binzenhöfer, M. Fiedler and K. Tutschku. Measurement and Analysis of Skype VoIP Traffic in 3G UMTS Systems. Proceedings of IPS MoMe 2006, Salzburg, February 2006.
[51]Pic Source : http://www.playstation.com.au as on November 2, 2005