Peer-to-Peer Networking

Is this the Future of Network Computing?

U Ananth Kini and Samarth M Shetty

U Ananth Kini and Samarth M Shetty are computer science BE graduates from Sri Jayachamarajendra College of Engineering, University of Mysore. They are currently doing a project on peer-to-peer networking at SERC, Indian Institute of Science, Bangalore. Their interests include networking, operating systems and embedded systems.

Peer-to-Peer (P2P) networking in recent times has been touted as the 'killer application' that is poised to shape the Internet's future. The purpose of this article is to define P2P and explain its working. We also describe various models of P2P and diverse applications of this innovative networking concept.

Introduction

Whenever confronted with a new word, probably the first object we refer to is the dictionary .The dictionary says, "**Peer** (pîr) *n.* Is a person who has equal standing with another or others, as in rank, class, or age". The keyword to be watched out for is 'equal'. If 'equality' is what is being stressed in P2P (Peer-to-Peer abbreviated) then naturally what exists at present must have some degree of 'inequality'. The present architecture in network computing is the well-known client-server model. A brief discussion of client-server model is necessary to understand the P2P model.

Let's assume the common situation where the sociable Mr. X wants to post his personal web page on the net. The first step of course is to type out the web page with accompanying decorations. The very next step is to upload the web pages onto a server. Mr. A, who is interested in visiting Mr. X's web page simply types out X's URL. The browser software contacts the server to obtain the required HTML file.

In the above example, client can be identified as the machine running the browser, whereas the web server housing Mr. X's page is the server of the client-server model. It is to be observed that it is the client, which is the initiator of the requests. Never

In a P2P network any computer is allowed to act as a source (initiator) or a sink (consumer) as the need be.

does the server spontaneously make a contact with the client. So, the 'inequality' that exists in a client-server model is that each machine in such a network can either be identified as a client or a server.

The P2P model erases the 'inequality' by designating each machine with the capability of a client as well as a server. In other words, any member is allowed to act as a source (initiator) or a sink (consumer) as the need be. Such a capability makes perfect sense since the ultimate aim of P2P network is to let the resource reside at the point of origin.

Need for Peer-to-Peer Networks

The three principal elements of the Internet can be identified as – data, computing resources and bandwidth. All of these are vastly under utilized, probably due to the traditional client-server computing model.

- First of all, no single search engine or portal can locate and maintain the ever-increasing amount of data on the web up-to-date. Moreover, a huge amount of data is transient and needs to be constantly updated by techniques such as web crawling. According to some researchers [2] for every *megabyte* of data produced only one *byte* is accessible to the search engines. This is indicative of the extent to which useful data remains hidden from those who need it.

- Secondly, new generation processors and storage devices continue to operate at breakneck speeds. These devices however are subject to immense workloads since the data tends to be centralized in present computation environment.

- Finally, laying of fiber optic cables is a common sight these days. These cables, no doubt are aimed at increasing the bandwidth and consequently reducing congestion. On the contrary, there may not be an appreciable increase in performance, if every other person goes to sites like Google for content, Amazon for books and so on.

These problems are, to some extent, alleviated in P2P model and hence P2P is expected to make the Internet more responsive.

TCP/IP protocol naturally suits the P2P networking model.

Services Expected of a P2P Protocol

The current network scenario is dominated by TCP/IP protocol. Fortunately, this protocol naturally suits the P2P model also, for the simple reason that once the caller party establishes the connection, free exchange of data can take place symmetrically i.e., there is no distinction in the communication capabilities of the caller and called party. Such a feature is necessary to sustain a P2P type communication, which typically involves query-reply, request-response and so on. Apart from drawing support from already existing TCP/IP protocol, there is also a need for providing the services described below around which P2P will pivot. They are as follows:

1. *Membership Service or Subscription Service*

The membership service or subscription service is used by the current members to reject or accept new subscriptions to a group. Peers wishing to join a peer group must first locate a current member, and then request to join. The application to join is either rejected or accepted by the collective set of current members by means of a voting procedure. The other alternative is to elect a designated group representative to accept or reject new membership applications. Also, a peer may subscribe to more than one peer-group simultaneously. For example, a person may be a member of Napster (Mp3 file sharing) as well as Gnutella (General file sharing) P2P service at the same time.

2. *Discovery Service*

The discovery service is used by peer members to search for peer-group resources. Only the peers that are currently logged on will be the ones that are searched. For example, in Napster [1] a dedicated server handles the discovery service. A peer, as soon as it logs on, registers its presence with the server, along with an index of Mp3 files it has to offer. A peer on the lookout

Peer monitoring is useful when features such as reliability and guaranteed service times are to be provided to the subscriber of a P2P network.

for a Mp3 file shall query the server, to obtain a list of peers possessing the required song. It will directly contact the peers mentioned in this list to download the file containing the song. There are a few more methods through which peers can be discovered. We will deal with them in the next section.

3. *Peer Monitoring Service*

Peer monitoring implies keeping a close track of a peer's status. Such a service is useful when features such as reliability and guaranteed service times are to be provided to the subscriber of a P2P network. For example, a failure in the peer system must be detected as early as possible so that corrective actions can be taken. It is sometimes better to shut down an erratic peer and transfer its responsibilities to another peer.

4. *Access Service*

The access service is used to validate requests made by one peer to another. The peer requiring data from another peer provides its credentials and particulars about the request being made. The access service has to determine if the access is permitted and if the request is warranted.

Classification of Architectures in P2P Networks

P2P networks can be classified depending on their basic organization. The classification that follows will be taking into account two factors. Firstly, the methodology used in detecting the presence of other peers and second the manner in which peer contents are discovered. We describe below four architectures.

1. *Pure P2P*

A pure P2P application has no central server whatsoever, as seen in *Figure* 1. It dynamically discovers other peers on the network and interacts with each of them for sending or receiving content.

This can be done in two ways:

***Method* 1.** It can employ network broadcasting and discovery

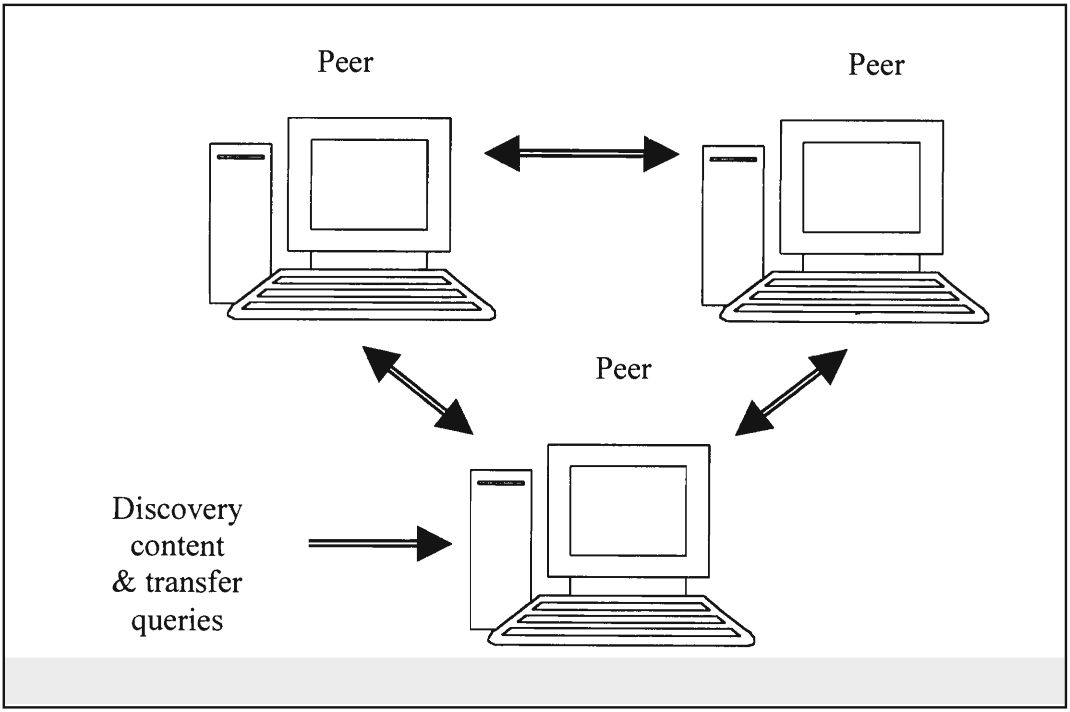

Figure 1. Pure P2P organization.

techniques such as IP multicast to discover the other peers. Using IP multicast can be difficult since it is not widely deployed on the Internet, but it can be useful in intranet where the network is more controlled and infrastructure required for multicast is known to exist.

Method 2. A peer can use information stored in a local configuration table to discover other peers (for example, a configuration entry that tells it who to talk with). In this case, the applications use a scheme such as a well-known node approach, where each peer knows about at least one other peer. This means that each peer maintains a table of peers that it can contact. So a peer discovery initiated by a peer will begin by probing the peers that it is aware of. These in turn contact the peers specified in their local tables and so on. Hence the search gradually diffuses across the network until the required peer is located.

A peer can use information stored in a local configuration table to discover other peers.

Analysis: The strength of this type of application is that it does not rely on any one server to be available for registration of its location in order for other peers to find it. At the same time, the

A peer application may use a log server to download a list of other peers on its network that it can use to query for content.

lack of a central log server poses a problem because a relatively small number of clients can be discovered, thereby limiting the application's reach. The slowly spreading discovery process may imply intolerable delays in a large network. Hence this model may not be highly scalable, especially with Method 2.

2. *P2P with a Simple Log Server*

This architecture, as depicted in *Figure* 2, works just like the pure P2P architecture except that it relies on a central server for discovery of other peers. In this model, the application usually notifies the central log server of its existence at login time. The peer application then uses this log server to download a list of other peers on its network that it can use to query for content. When content is needed, it goes through the list and contacts each peer individually with its request.

Analysis: We observe a considerable speed up in discovery since a peer can now obtain a list of online peers from the log server. But it must still check the listed peers whether they possess the required contents. This may lead to going through irrelevant peers. Requesting content from each individual peer can be quite expensive from a network resource perspective. However it is highly scalable, as its reach is fast and wide.

3. *P2P with a Log and Lookup Server*

This model, similar to the one shown in *Figure* 2, extends the log server so that it also includes content lookup services. In this case, the peer application not only registers with a log server, but it also uploads a list of its contents at regular intervals. When an application is looking for some particular content, it queries the central server rather than sending a query to each client. The central server then responds with a list of the clients that contain the requested content, and the peer application can contact those clients directly to retrieve the content.

Analysis: Quite often this approach will scale better than the previous options because it reduces the number of queries going

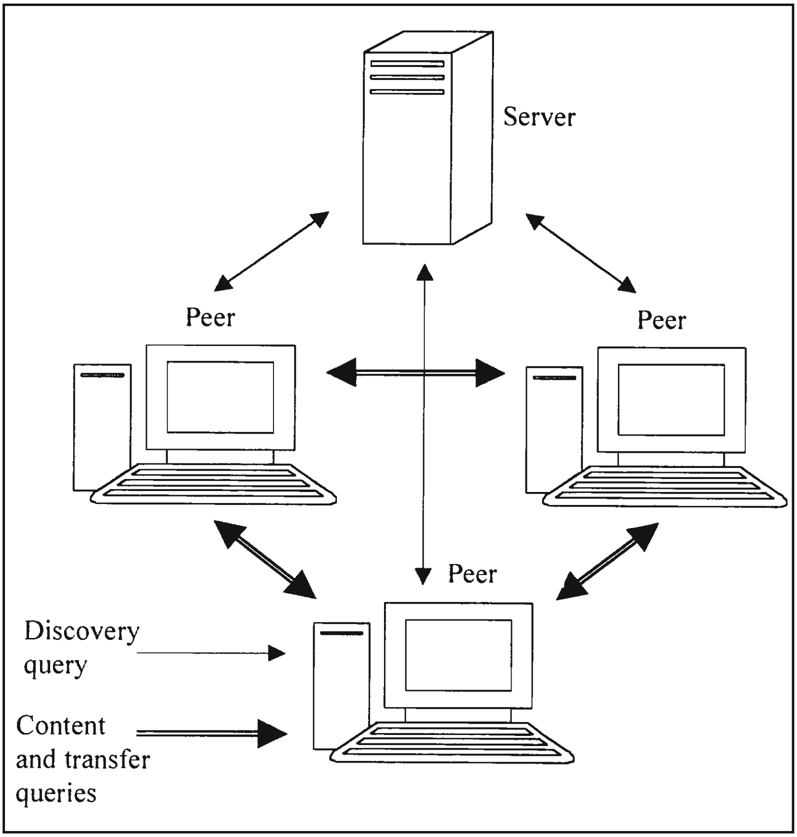

Figure 2. P2P with a log server.

over the network (arguably one of the scarcest resources) i.e. irrelevant peers need not be probed for content as in previous case. However, this saving will incur a cost on the server. Servers are now more involved in the process of content sharing and the peer's demands will use significant resources of the server.

4. *P2P with a Log, Lookup and Content Server*

Just to show that this can actually come full circle, a system can be designed so that the peers can upload the content to the server as well, if you so choose (see *Figure* 3). This approach effectively becomes the client/server model because the peers are no longer contacting other peers for content. Each peer registers with a server (if needed), queries it for data, and transfers any desired content down from the server. The server quickly becomes the bottleneck and is easily overwhelmed by the peers (clients).

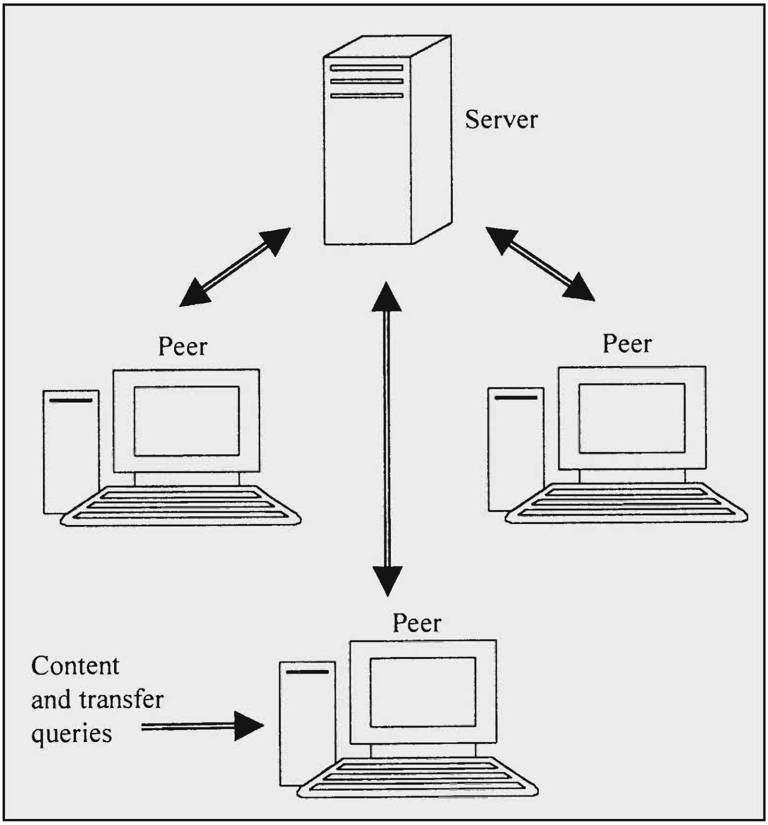

Figure 3. P2P with log, lookup and content server.

Paradigms of P2P

A more skeptical look at the concept of P2P will hint the presence of P2P-like concepts from a time when applications like the USENET [6] were introduced. If so, what could have pushed P2P into limelight all of a sudden? Napster [2] in fact gave wide publicity to P2P. In the following section we will have a quick look at the ingenuity of SETI@home [3]. A similar discussion of Napster can be found in *Resonance* [1]. These two applications deserve a discussion, simply because they symbolize two fundamental facets of P2P, content sharing and computation sharing, respectively.

Napster gave wide publicity to P2P.

SETI@home [3]

'Star Trek', 'Star Wars', 'Alien', have one thing in common.

They all deal with alien civilizations and their relationships to humans. The popularity of these movies suggests that many of us dream of one day meeting up with an (friendly) alien race. Enthusiasts all over the world are currently engaged in several programs that are looking for the evidence of life elsewhere in the cosmos. Collectively, these programs are called SETI (Search for Extra-Terrestrial Intelligence). SETI@home is one of a number of projects searching for evidence of extraterrestrial intelligence, but it is unique in its approach that makes it one of the most exciting Internet projects ever.

SETI@home is the most powerful computer in the world; SETI@home currently has a processing speed of about 15 Teraflops, whereas the IBM's ASCI White, one of the leading supercomputers is rated at 12 Teraflops. It is probably the cheapest super computer existing at present costing about $0.5million so far, whereas IBM's ASCI white costs $110 million. [3]

SETI@home works as follows. Volunteers join it by downloading a program on their computers that is akin to a screen saver. It wakes up during idle processor cycles of the machine in which it resides. It performs pre-specified computations on a data provided by the SETI@home server. The results of this analysis are ultimately sent back to the SETI@home team. This result is combined with the processed data from many other SETI@home participants to help in the search for extraterrestrial signals. When it exhausts the current set of data, it downloads a chunk of data for analysis. The data processing does not occur in 'real time' so that interesting signals must be followed up at a later date. (See *Box* 1).

Box 1. Data Processing logic of SETI@home

We expect that extra terrestrials would want to send us a signal in the most efficient manner that would allow us to easily detect the message. Now, it turns out that sending a message on many frequencies is not efficient. It takes lots of power. If one concentrates the power of the message into a very narrow frequency range (narrow bandwidth) the signal is easier to weed out from the background noise. This is especially important since we assume that they are far enough away that their signal will be very weak by the time it gets to us. So SETI@home does not look for broadband signal, instead it focuses on a very specific frequency message. The SETI@home screen saver acts like tuning your radio set to various channels, and looking at the signal strength meter. If the signal strength goes up, it attracts our attention.

Advantages of P2P

We list below some of the prominent advantages of P2P networking.

1. ***Abundant availability of resources:*** P2P places huge quantity

of resources at our disposal, may it be computational resources or content.

2. ***Enhanced Load Balancing:*** If we consider a situation where a piece of data is present only at a particular peer, it is possible that the peer is overburdened with requests. P2P can circumvent this problem by providing multiple copies of data. Also, explicit caching algorithms, where intermediate peers cache frequently used data, can be devised to help distribute the content more evenly. Thus query load is more evenly balanced.

3. ***Redundancy and fault tolerance:*** Due to fast duplication of data in P2P model, it is possible that even if one peer does not possess the required data, some other peer would have similar copy. If a peer goes down after receiving a chunk of data for computation, there is a very good possibility that a similar machine is present elsewhere in the network that can take up the terminated job.

4. ***Content based addressing:*** In the present Internet scene, a person is required to type out the address of a particular site. There may be very little correspondence between the site name and its contents. In P2P the exact address of a node storing a particular content remains transparent to the user. The user queries the network for the content and P2P software translates the requests into specific nodes that hold the content. This procedure can lead to a grouping of addresses based on the content the respective nodes store. The segregation of content into specialized groupings distributed over P2P networks can lead to more refined data repository.

5. ***Improved search:*** The search engines rely on web crawlers that scour the Internet for content and store them in massive, searchable databases. Such indexing only includes contents from publicly operating servers, and databases do not undergo immediate updates when any of those servers or links go down. By contrast, in P2P, a node is indexed only when it is online.

Suggested Reading

[1] V Rajaraman, Electronic commerce, Emerging Applications and Some Legal Issues, *Resonance*, Vol. 6, No.8, pp.18-27, Aug 2001.

[2] http://www.howstuff works.lycos.com/napster.htm

[3] SETI@home, http://www.setiathome.ssl.berkeley.edu

[4] Lance Olson, NET P2P: Writing Peer-to-Peer Networked Apps with the Microsoft. NET Framework, http://www.msdn.microsoft.com

[5] Li Gong, Project JXTA: A Technology Overview, Sun Microsystems Inc, http://www.jxta.org

[6] Andrew S Tanenbaum, *Computer networks*, Prentice Hall Of India, New Delhi, 1999.

Hence, the P2P index always synchronizes with the current status.

Disadvantages of P2P

P2P network has the following disadvantages.

1. ***Spurious content, poor connection:*** Due to lack of central authority, the quality of the content posted on the peer group is questionable. For example the Mp3 version of the same song may be available as a copy with a very good sound quality and another copy may be filled with noisy glitches. But for the P2P search both versions are part of the same search and indistinguishable, until actually heard. Also, slow and error prone dial up connections used by some of the peers may disrupt the normal functioning of the network. Such poor quality of service may plague P2P networks if the peer contents or their capabilities are not subject to some kind of quality check from time to time.

2. ***May serve as haven for crackers ('hackers'):*** Handing over of a peer's resource partly or fully to the peer group may result in unscrupulous elements getting hold of local configuration information .The information may be used to spread viruses and other malicious code to a peer or to the whole network.

3. ***Infringement of copyright laws:*** Due to free exchange of data over the peer group, such situations are bound to occur.

Conclusion

P2P networking is providing many new opportunities of sharing resources – both data and computing over the Internet. Even though the idea is not new, new improvements in software has eased the realization of P2P networks. Many innovative ideas within the realm of P2P are on the horizon.

Address for Correspondence
U Ananth Kini
#46, Ramnath Krupa
Dena Bank Colony
2nd Main, 3rd Cross
Ganganagar
Bangalore 560 032, India.
Email:
kiniananth@rediffmail.com
kiniananth@hotmail.com

Samarth M Shetty
C-325, 9th Main
JP Nagar
Mysore 570 008, India.
Email:s_shetty@rediffmail.com