Volume 4, Issue 9, September 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Comparative Study on Peer-to-Peer Architectural Styles

Aditya Mandhare
Computer Engineering Department,
University of Mumbai, India

Abstract— With rapid increase in the use of the internet, large amount of data is transferred over the network. This data may include, documents, images, audio and video files. However it is required that the internet should be robust and scalable and that dependency between different components be reduced. This need has led to the development of a new architectural style called the peer-to-peer style. It is a decentralized and distributed architectural style, and has enabled desiring parties to retain autonomous control over the aspects of their participation. It improves reliability and eliminates the need of a full time system administrator. The paper discusses various peer-to-peer architectural styles and also makes a comparison between them.

Keywords— Peer-to-Peer; Architectural style; Decentralized Architecture; Resource trading; content directory; node; supernode.

I. INTRODUCTION

The term "peer-to-peer" (P2P) refers to a class of systems and applications that employ distributed resources to perform a critical function in a decentralized manner. P2P architectural style consists of a network of loosely coupled autonomous peers, each peer acting both as a client and a server. Unlike the client-server style where state and logic are centralized on the server, P2P decentralizes both information and control. Peer-to-peer (P2P) systems have become a popular medium through which to share huge amounts of data. P2P systems distribute the main costs of sharing, data disk space for storing files, and bandwidth for transferring them across the peers in the network, thus enabling applications to scale without the need for powerful, expensive servers.

Their ability to build an extremely resource-rich system by aggregating the resources of a large number of independent nodes makes peer-to-peer systems more capable than many centralized systems at relatively lower cost. In fact, peer-to-peer technology is one of the most important and suitable networking technologies for ubiquitous communications since it supports easily one-to-one communication between devices, free and extensible distribution of resources and distributed search for enormous amount of resources.

P2P is a class of applications that take advantage of resources storage, cycles, content, human presence available at the edges of the Internet. Because accessing these decentralized resources means operating in an environment of unstable connectivity and unpredictable IP addresses, peer-to-peer nodes must operate outside the DNS and have significant or total autonomy of central servers". In a short word, P2P is a special distributed system on the application layer, where each pair of peers can communicate with each other through the routing protocol in P2P layer.

II. ADVANTAGES OF PEER-TO-PEER STYLES

Peer-to-Peer networks have several advantages some which make them popular in various web applications.

A. Easy installation and configuration of computers on this network

The ease of installation and configuration makes these networks a popular choice. It also makes it cost efficient. Therefore there has been a surge of applications using this technology in recent times.

B. The resources and contents are shared by all peers

This is in contrast to a client server system in which a single server shares all the files. However in peer-to-peer style, all the peers participate in resource sharing.Improved reliability as it does not depend on a central server.

C. Every user is an administrator of his machine and there is no need of a centralized administrator

Here, since there is no central server, and data is distributed between peers, the reliability of the system is improved and failure of one peer does not bring down the entire system.

D. Networks are cheaper to build and maintain

Here every user is client as well as the server of his machine. This eliminates need of an administrator. It makes system easy to manage.

E. Networks are cheaper to build and maintain

The overall cost of building the system is comparatively less than a client-server system. This is because a single party is not responsible for maintaining the central server and access control. Each user is client as well as server and maintains his machine.

III. DRAWBACKS OF PEER-TO-PEER STYLES

Peer-to-Peer networks also have certain drawbacks which raise concerns over their usage.

A. Access Control

The whole system is decentralized and is difficult to administer. Thus it becomes difficult to control the overall accessibility of the entire system.

B. Security Issues

In these systems security is very less as malwares like, viruses, worms and other spywares can be easily transmitted over the network. This may harm the system on which these malwares are being downloaded.

C. Copyright Issues

Due to the use of peer-to-peer networks, there is no central server where files are placed. As a result illegal sharing of music files and video files take place violating the copyright rules. Also since there is no central server, this activity cannot be tracked.

D. Data Recovery

Since there is no central server, data recovery and backup is very difficult. In these systems, each user needs to have his own backup system.

IV. PEER-TO-PEER STYLES

Depending upon the way in which the peers interact with each other the different types of peer-to-peer styles are as follows:

A. Hybrid Client-Server / Peer-to-Peer

This type of architectural style consists of a central server, known as peer and content directory. Each peer must first register itself with the central server. While registering itself with the central server or logging in later, the peer informs the server of its internet address and the files present with the peer which it is willing to share. A complete record of which files are present and the server on which they are present is maintained by the peer and content directory.

Now when a peer needs to download a file from the network, it queries the server as to where on the internet can a given file be obtained. The peer is then informed by the server about the locations where the file is present. The peer then chooses a location and makes a call directly to that peer and downloads the file. Thus, architecturally the system can be seen as a hybrid of client-server and pure peer-to-peer (P2P). The 'peers' act as clients and the 'peer and content directory' act as server during the registration and querying process. Once the peer knows about the file location, the file transfer takes place between the peers without further inclusion of the 'peer and content directory'.

B. Pure Decentralized P2P

In this architectural style, there is no central server. All the peers are equal in capability. All of them share equal responsibility. It ensures that a single central server does not have to bear the brunt of the entire network. This improves performance and robustness. The peers use different ways to discover other peers which hold the requested file. This may increase network traffic in some cases but, also improve robustness of the system. Failure of a single peer does not have an adverse impact and the system can still continue to function without much deterioration in performance. Different algorithms are used for searching the peers. In order to make searching efficient, time out mechanisms and hop count limits are used. This ensures efficiency of the system.

C. Overlayed P2P

An overlay network is a layer of virtual network topology on top of the physical network, which directly interfaces to users. With the rapid advancement of Internet and computing technology, much more aggregate information and computing resources are available from clients or peers than from a limited number of centralized servers. Overlay networks provide us with the following advantages and opportunities to better utilize the increasingly growing Internet information and resources. Overlay networks allow both networking developers and application users to easily design and implement their own communication environment and protocols on top of the Internet, such as data routing and file sharing management.

Data routing in overlay networks can be very flexible, quickly detecting and avoiding network congestions by adaptively selecting paths based on different metrics, such as probed latency. The end-nodes in overlay networks are highly connected to each other due to flexible routing. As long as the physical network connections exist, one end-node can always communicate to another end-node via overlay networks. Thus, scalability and robustness in overlay networks

are two attractive features. The high connectivity of increasingly more end-nodes to join overlay networks enables effective sharing of a huge amount of information and resources available in the Internet.

D. Resource Trading P2P

The main goal of resource trading p2p is to support speedy replication of large files on individual peers, upon demand. It is an attempt to maximize use of all available resources in the network of interested peers and to minimize the burden on any one participant. Thus it also promotes scalability. In this style, a file is distributed in parts to many peers. This distributes both the processing and network load. A peer does not obtain a requested large file from a single resource, rather it obtains pieces of the file from many peers. These pieces are then reassembled. In this style the requesting peer does not only download the pieces, but is also responsible for uploading the portions of the file that it has, to other interested peers. Thus, along with downloading pieces of a file, the peer also acts as a 'seed' which uploads content for other peers to download. The operating context is however one in which many peers are simultaneously interested in obtaining a copy of the file.

V. APPLICATIONS USING EACH STYLE

A. Napster

In Napster, a large cluster of dedicated central servers maintain an index of the files that are currently being shared by active peers. Each peer maintains a connection to one of the central servers, through which the file location queries are sent. The servers then cooperate to process the query and return a list of matching files and locations to the user. On receiving the results, the peer may then choose to initiate a file exchange directly from another peer. In addition to maintaining an index of shared files, the centralized servers also monitor the state of each peer in the system, keeping track of metadata such as the peers' reported connection bandwidth and the duration that the peer has remained connected to the system. This metadata is returned with the results of a query, so that the initiating peer has some information to distinguish possible download sites.

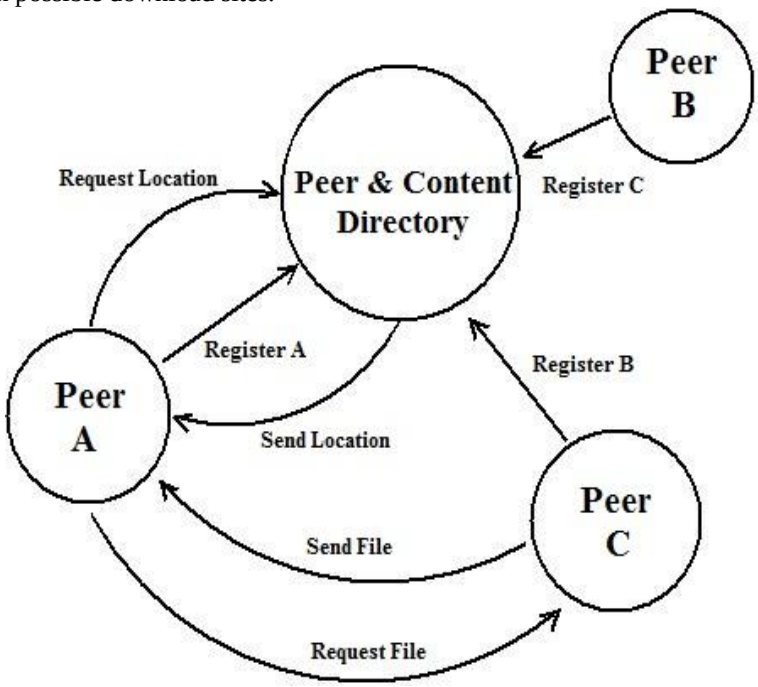

Figure 1. Notational View of the operation of Napster

As we can see in the figure 1, the peers A, B and C have registered themselves with the central server. Now as Peer A wishes to download a file, it contacts the central server. The central server informs A about location of Peer C, who has the requested file. Peer A then directly contacts Peer C who responds by providing the requested file.

B. Gnutella

In Gnutella, the peers in the system form an overlay network by forging a number point-to-point connections with a set of neighbours. Now if a peer wishes to locate a file, it sends a query packet to all of its neighbours. This flooding is done in a controlled manner. Upon receiving a query packet, a peer checks if any locally stored files match the query. If a match is found, the peer sends a query response packet back towards to the query originator. Whether or not a file match is found, the peer continues to flood the query through the overlay.

To help maintain the overlay as the users enter and leave the system, the Gnutella protocol includes ping and pong messages that help peers to discover other nodes in the overlay. Pings and pongs behave similarly to query/query response packets: any peer that sees a ping message sends a pong back towards the originator, and also forwards the ping onwards to its own set of neighbours. Ping and query packets thus flood through the network; the scope of flooding is controlled with a time-to-live field that is decremented on each hop.

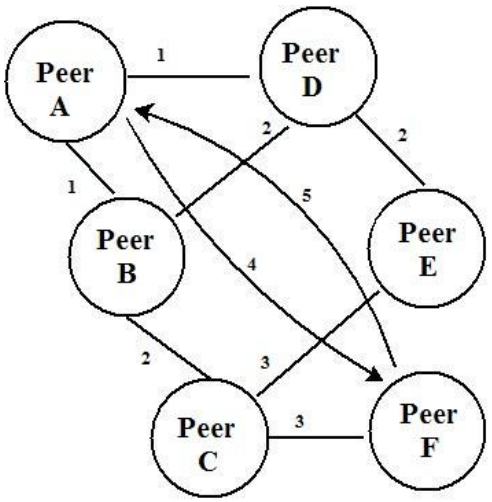

Figure 2. Notational interactions between peers using Gnutella

Consider the figure 2. In this, 'peer A' wishes to obtain a resource. It sends a request to peers it knows. Thus, it sends a request to peers 'B' and 'D'. Now these peers will check if they possess the requested resource. If they do not, they forward this request to other peers they know. Thus 'peer B' forwards the request to peers 'C' and 'D', whereas peer D forwards it to peer 'E' and 'B'. Again, peers 'C' and 'E' do not possess the requested resource. Peer 'C' propagates this query to peer 'F'. Peer 'F' responds positively. This information about peer 'F' is relayed back to peer 'A' via peers 'C' and 'B'. Now peer 'A' will directly contact peer 'F' and then download the required file from it.

C. Skype

Skype is a peer-to-peer VoIP client. The users of Skype can send text messages and make voice calls. Skype is an overlay peer-to-peer network. There are two types of nodes in this overlay network, ordinary hosts and super nodes. An ordinary host is a Skype application that can be used to place voice calls and send text messages. A super node is an ordinary host's end-point on the Skype network. Any node with a public IP address having sufficient CPU, memory, and network bandwidth is a candidate to become a super node.

Thus an ordinary host gets promoted to a super node if it has the required specifications. The supernode provides directory services and call routing services. An ordinary host must connect to a super node and must register itself with the Skype login server for a successful login. The Skype login server is an important entity in the Skype network. User names and passwords are stored at the login server. The login server performs user authentication during login. This server also ensures that Skype login names are unique across the Skype name space. Apart from the login server, there is no central server in the Skype network. Online and offline user information is stored and propagated in a decentralized fashion and so are the user search queries.

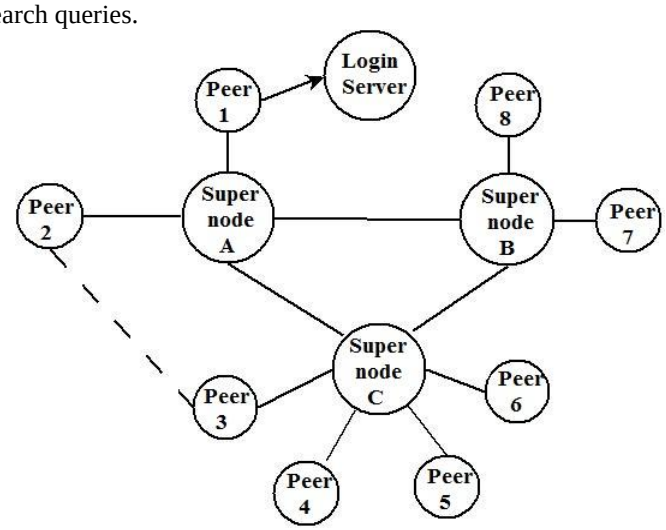

Figure 3. Notational instance of Skype architecture.

Figure 3 shows how the application functions in more detail. The figure shows that Peer 1 is logging into the Skype server. That server then tells the Skype peer, the address of Supernode A. The Peer 1 then contacts Supernode A. To see if any of its buddies are online, a query is issued to the supernode. When a Skype call is made, the call may proceed from the peer to the supernode, and then directly to the destination peer, or to another supernode, and then to the receiving peer. However, if both peers are on public network and not behind firewall, then a direct interaction can be established as in the case of Peers 2 and 3. The Skype calls are encrypted and this maintains privacy as theses calls are relayed through supernode intermediaries.

D. Bit Torrent

Bit Torrent in itself is only a file-downloading protocol. In Bit Torrent, files are split up into chunks. These chunks may amount to thousand per file. The downloaders of a file barter for chunks of it by uploading and downloading them in a tit-for-tat-like manner to prevent parasitic behaviour. A parasitic node can exploit all members of the swarm, and achieve download speeds far in excess of regular downloaders. Overall, tit-for-tat fails because each peer selfishly and locally seeks to maximize its profits, without concern for the global good and often without conferring with other nodes. Each peer is responsible for maximizing its own download rate by contacting suitable peers, and peers with high upload rates will with high probability also be able to download with high speeds. After downloading of a file is complete, it may act as a seed by staying online for a while and sharing the file for free, i.e., without bartering. Responsibility for the discovery of content is outside the scope of Bit Torrent. To find a file in Bit Torrent, users access web sites which act as global directories of available files. Some of these web sites acting as global directories include Suprnova.org, Youceff.com , Piratebay.org.

Bit Torrent makes use of a machine called 'Tracker', in order to oversee the process by which a file is distributed to a set of interested peers. However this tracker is not responsible for actually performing the file transfer. The peers interact with the Tracker, to identify other peers with which they communicate to effect the download. The meta-data associated with each file provides information about, how the file is pieced and also about the attributes of those pieces. It also tells about the location of the tracker.

Each peer runs a Bit Torrent application. This application decides, what piece of the file should be downloaded next, and from which peer to download. Several algorithms are used to facilitate this type of file sharing. The aim of these algorithms is to maximize the use of resources available at all peers. However, Bit Torrent keeps a strict check, to ensure that a peer is not manipulated. It may be possible that a peer is manipulated to only download content, and does not participate in the uploading process. Such a peer is then penalized by other peers, through deprioritization of access to the pieces it needs.

VI. COMPARISON OF STYLES

Each of the P2P styles discussed above viz. Hybrid, Pure Decentralized, Overlayed P2P and Resource Trading P2P have some advantages and drawbacks. The architectural cleanness of Napster (hybrid style) is also its downfall. If for example a highly desirable file becomes available, the server will be swamped with requests seeking its location. Thus there is a possibility of server going down. If the server goes down due technical or any legal matters, the peers lose the ability to find other peers.

Gnutella (pure decentralized) is highly robust. Removal of one peer does not diminish the ability of remaining peers to perform. The removal of a peer may make a specific resource unavailable if the peer was the unique source of that resource. However if the resource is available with several sources, it is still conceivable from one of the sources. Thus removal of many peers wont hamper the performance. However Gnutella has a few drawbacks like, when a new peer comes to a network, how does it find peers to which its queries can be sent ? Also the peers lack global knowledge. They have knowledge only about, the requests received, requests which they have issued or passed along. As a result the system may continue to query for a long time even after the reply is received by the original peer who started the query. If this is not properly handled, the system may as a whole become inefficient. However efficient time out mechanism may prove to be a good solution, in improving the efficiency of this system. Another problem with Gnutella is that there is no guarantee that the file received is not malicious. The downloaded file may turn out to be a virus or a worm.

Skype (overlayed P2P) addresses the problem of discovery of peer. Te network is not flooded with requests in order to locate a buddy (peer) as it would in case of Gnutella. Also since the directories are replicated and distributed in the form of supernodes, the problems of scalability and robustness as encountered in case of Napster are also solved. Also encryption facility provides security to the calls made. Due to the Login Server which restricts the participants, malicious clients can be prevented from entering the network. Bit Torrent (resource trading) style possesses all the qualities of a good P2P style. It also accounts for the possibility of any given peer dropping out of the process at any time, thus ensuring robustness.

VII. CONCLUSIONS

This paper presents an overview of the research relating to P2P architectural styles. It has compared the various advantages and drawbacks of these architectural styles. It has also stated the applications of each style. P2P is an important technology that has already found its way into existing products and research projects. P2P network is factually a network for distributed object storage, searching and sharing. It would emerge as a solution to certain inherent problems in distributed systems. It may not be the only solution for such problems. Also, it may not be appropriate for all problems, but it will continue to be a strong alternative for scalability, anonymity, and fault resilience requirements. P2P architectural styles and the applications developed using them have an opportunity for deployment in the future

ACKNOWLEDGMENT

I express sincere regards to all the researchers who have done pioneering work in the field of Peer-to-Peer architectures. It was the hard work and relentless efforts put by them, that today P2P architectures have become popular. I would also like to thank all the authors whose papers have been taken as a reference and as a guidance material for this paper. Truly, their research has proved to be of great value to this paper.

REFERENCES

[1] An Analysis of the Skype Peer-to-Peer Internet Telephony Protocol Salman A. Baset and Henning Schulzrinne Department of Computer Science Columbia University, New York NY 10027 {salman,hgs} @cs.columbia.edu September 15, 2004.

[2] A Measurement Study of Peer-to-Peer File Sharing Systems Stefan Saroiu, P. Krishna Gummadi, Steven D. Gribble {tzoompy, gummadi, gribbleg} @cs.washington.edu. Technical Report # UW-CSE-01-06-02 Department of Computer Science & Engineering University of Washington Seattle, WA, USA, 98195-2350

[3] The Bit Torrent P2P File-Sharing System: Measurement And Analysis. J.A. Pouwelse, P. Garbacki, D.H.J. Epema, H.J. Sips. Department of Computer Science, Delft University of Technology, the Netherlands, j.a.pouwelse@ewi.tudelft.nl.

