Peer to Peer Systems

Leila Moradi
Department of Information Systems and Management
Universiti Kebangsaan Malaysia, UKM
Bangi, Malaysia
sheidaa447@yahoo.com

Abstract

Peer-to-Peer technology is an emerging concept applied at different levels of the systems architecture. Its main attractions are direct interaction and data exchange between Peer systems. Although no formal definition exists, the concept is widely deployed in different ways. A peer-to-peer distributed network architecture is composed of participants that make a portion of their resources (such as processing power, disk storage, and network bandwidth) available directly to their peers without intermediary network hosts or servers. Peers are both suppliers and consumers of resources, in contrast to the traditional client-server model where only servers supply, and clients consume.This paper gives an overview of the different areas peer-to-peer technology is used and disusses about the various issues surrounding peer-to-peer systems.

1. Introduction

Peer-to-peer was popularized by file sharing systems like Napster. Peer-to-peer file sharing networks have inspired new structures and philosophies in other areas of human interaction. In such social contexts, peer-to-peer as a meme refers to the egalitarian social networking that is currently emerging throughout society, enabled by Internet technologies in general. In a peer-to-peer network, the users handle administration. This means that all the users need to be trained in how to share files, folders, and printers. In a peer-to-peer network, suddenly shutting down your computer can cause one of your colleagues to be unable to print. Some of active peer-to-peer technologies include:

- Many file sharing networks, including Gnutella, FastTrack, BitTorrent
- Completely decentralized networks of peers: Usenet (1979) and WWIVnet (1987).
- Software publication and distribution (Linux, several games)
- The JXTA extensions of Java.
- Research like the Chord project, the PAST storage utility, the P-Grid, and the CoopNet content distribution system.
- Distributed hash tables
- In bioinformatics, drug candidate identification. The first such program was begun in 2001 the Centre for Computational Drug Discovery at the University of Oxford in cooperation with the National Foundation for Cancer Research.

2. Architecture of P2P Systems

Peer-to-peer networks are typically formed dynamically by ad-hoc additions of nodes. In an 'ad-hoc' network, the removal of nodes has no significant impact on the network. The distributed architecture of an application in a peer-to-peer system provides enhanced scalability and service robustness.

Peer-to-peer systems often implement an Application Layer overlay network on top of the native or physical network topology. Such overlays are used for indexing and peer discovery. Content is typically exchanged directly over the underlying Internet Protocol (IP) network. Anonymous peer-to-peer systems are an exception, and implement extra routing layers to obscure the identity of the source or destination of queries.

P2P networks are typically used for connecting nodes via largely ad hoc connections. Sharing content files (see file sharing) containing audio, video, data or anything in digital format is very common, and real time data, such as telephony traffic, is also passed using P2P technology.

A pure P2P network does not have the notion of clients or servers but only equal peer nodes that simultaneously function as both "clients" and "servers" to the other nodes on the network. This model of network arrangement differs from the client-server model where communication is usually to and from a central server. A typical example of a file transfer that is not P2P is an FTP server where the client and server programs are quite distinct: the clients initiate the download/uploads, and the servers react to and satisfy these requests.

The P2P overlay network consists of all the participating peers as network nodes. There are links between any two

nodes that know each other: i.e. if a participating peer knows the location of another peer in the P2P network, then there is a directed edge from the former node to the latter in the overlay network. Based on how the nodes in the overlay network are linked to each other, we can classify the P2P networks as unstructured or structured.

2.0.1. Structured peer-to-Peer systems. Structured P2P network employ a globally consistent protocol to ensure that any node can efficiently route a search to some peer that has the desired file, even if the file is extremely rare. Such a guarantee necessitates a more structured pattern of overlay links. In structured peer-to-peer networks, connections in the overlay are fixed. They typically use distributed hash table-based (DHT) indexing, such as in the Chord system (MIT).

Distributed hash tables(DHT). Distributed hash tables (DHTs) are a class of decentralized distributed systems that provide a lookup service similar to a hash table: (key, value) pairs are stored in the DHT, and any participating node can efficiently retrieve the value associated with a given key. Responsibility for maintaining the mapping from keys to values is distributed among the nodes, in such a way that a change in the set of participants causes a minimal amount of disruption. This allows DHTs to scale to extremely large numbers of nodes and to handle continual node arrivals, departures, and failures. DHTs form an infrastructure that can be used to build peer-to-peer networks. Notable distributed networks that use DHTs include BitTorrent's distributed tracker, the Kad network, the Storm botnet, YaCy, and the Coral Content Distribution Network. Some prominent research projects include the Chord project, the PAST storage utility, the P-Grid, a self-organized and emerging overlay network and the CoopNet content distribution system (see below for external links related to these projects). DHT-based networks have been widely utilized for accomplishing efficient resource discovery[3][4] for grid computing systems, as it aids in resource management and scheduling of applications. Resource discovery activity involves searching for the appropriate resource types that match the users application requirements. Recent advances in the domain of decentralized resource discovery have been based on extending the existing DHTs with the capability of multi-dimensional data organization and query routing. Majority of the efforts have looked at embedding spatial database indices such as the Space Filling Curves (SFCs) including the Hilbert curves, Z-curves, k-d tree, MX-CIF Quad tree and R*-tree for managing, routing, and indexing of complex Grid resource query objects over DHT networks. Spatial indices are well suited for handling the complexity of Grid resource queries. Although some spatial indices can have issues as regards to routing load-balance in case of a skewed data set, all the spatial indices are more scalable in terms of the number of hops traversed and messages generated while searching and routing Grid resource queries.

Chord system. Chord is a peer-to-peer lookup algorithm for finding a single node in a structured network of peers as a rendezvous point for a given key, which is an index for a desired entity of information. It is one of the four original distributed hash table protocols, along with CAN, Tapestry, and Pastry. It was introduced in 2001 by Ion Stoica, Robert Morris, David Karger, Frans Kaashoek, and Hari Balakrishnan, and is developed at MIT. Using the Chord lookup protocol, node keys are arranged in a circle. The circle cannot have more than 2m nodes. The circle can have ids/keys ranging from 0 to 2m - 1. IDs and keys are assigned an m-bit identifier using what is known as consistent hashing. The SHA-1 algorithm is the base hashing function for consistent hashing. The consistent hashing is integral to the probability of the robustness and performance because both keys and IDs (IP addresses) are uniformly distributed and in the same identifier space. Consistent hashing is also necessary to let nodes join and leave the network without disrupting the network. Each node has a successor and a predecessor. The successor to a node or key is the next node in the identifier circle when you move clockwise. The predecessor of a node or key is the next node in the id circle when you move counter-clockwise. If there is a node for each possible ID, the successor of node 2 is node 3, and the predecessor of node 1 is node 0; however, normally there are holes in the sequence, so, for example, the successor of node 153 may be node 167 (and nodes from 154 to 166 will not exist); in this case, the predecessor of node 167 will be node 153. Since the successor (or predecessor) node may disappear from the network (because of failure or departure), each node records a whole segment of the circle adjacent to it, i.e. the K nodes preceding it and the K nodes following it. One successor and predecessor are kept in a list to maintain a high probability that the successor and predecessor pointers actually point to the correct nodes after possible failure or departure of the initial successor or predecessor.

2.0.2. Unstructured Peer-to-Peer Systems. An unstructured P2P network is formed when the overlay links are established arbitrarily. Such networks can be easily constructed as a new peer that wants to join the network can copy existing links of another node and then form its own links over time. In an unstructured P2P network, if a peer wants to find a desired piece of data in the network, the query has to be flooded through the network to find as many peers as possible that share the data. The main disadvantage with such networks is that the queries may not always be resolved. Popular content is likely to be available at several peers and any peer searching for it is likely to find the same thing. But if a peer is looking for rare data shared by only a few other peers, then it is highly unlikely that search will be successful.

Since there is no correlation between a peer and the content managed by it, there is no guarantee that flooding will find a peer that has the desired data. Flooding also causes a high amount of signaling traffic in the network and hence such networks typically have very poor search efficiency. Most of the popular P2P networks are unstructured. In pure P2P networks: Peers act as equals, merging the roles of clients and server. In such networks, there is no central server managing the network, neither is there a central router. Some examples of pure P2P Application Layer networks designed for file sharing are Gnutella (pre v0.4) and Freenet. There also exist hybrid P2P systems, which distribute their clients into two groups: client nodes and overlay nodes. Typically, each client is able to act according to the momentary need of the network and can become part of the respective overlay network used to coordinate the P2P structure. This division between normal and 'better' nodes is done in order to address the scaling problems on early pure P2P networks.

3. Peer-to-Peer-like Systems

In modern definitions of peer-to-peer technology, the term implies the general architectural concepts outlined in this article. However, the basic concept of peer-to-peer computing was envisioned in earlier software systems and networking discussions, reaching back to principles stated in the first Request for Comments, RFC 1. A distributed messaging system that is often likened as an early peer-to-peer architecture is the USENET network news system that is in principle a client-server model from the user or client perspective, when they read or post news articles. However, news servers communicate with one another as peers to propagate Usenet news articles over the entire group of network servers. The same consideration applies to SMTP email in the sense that the core email relaying network of Mail transfer agents has a peer-to-peer character, while the periphery of e-mail clients and their direct connections is strictly a client-server relationship. Tim Berners-Lee's vision for the World Wide Web, as evidenced by his Worldwide Web editor/browser, was close to a peer-to-peer design in that it assumed each user of the web would be an active editor and contributor creating and linking content to form an interlinked web of links. This contrasts to the broadcasting-like structure of the web as it has developed over the years. Advantages and weaknesses of P2P networks

4. Advantages and Weaknesses of P2P Networks

In P2P networks, all clients provide resources, which may include bandwidth, storage space, and computing power. As nodes arrive and demand on the system increases, the total capacity of the system also increases. This is not true of a client-server architecture with a fixed set of servers, in which adding more clients could mean slower data transfer for all users. The distributed nature of P2P networks also increases robustness, andin pure P2P systemsby enabling peers to find the data without relying on a centralized index server. In the latter case, there is no single point of failure in the system. It also help to reduce system costs and allows cost sharing by bundling resources from different sites and making use of existing infrastructure. Peer-to-peer systems are also more resilient because of their distributed and non-hierarchical nature. As with most network systems, unsecure and unsigned codes may allow remote access to files on a victim's computer or even compromise the entire network. In the past this has happened for example to the FastTrack network when anti P2P companies managed to introduce faked chunks into downloads and downloaded files (mostly MP3 files) were unusable afterwards or even contained malicious code.Consequently, the P2P networks of today have seen an enormous increase of their security and file verification mechanisms. Modern hashing, chunk verification and different encryption methods have made most networks resistant to almost any type of attack, even when major parts of the respective network have been replaced by faked or nonfunctional hosts. Internet service providers (ISPs) have been known to throttle P2P file-sharing traffic due to the high-bandwidth usage . Compared to Web browsing, e-mail or many other uses of the internet, where data is only transferred in short intervals and relative small quantities, P2P file-sharing often consists of relatively heavy bandwidth usage due to ongoing file transfers and swarm/network coordination packets. A possible solution to this is called P2P caching, where a ISP stores the part of files most accessed by P2P clients in order to save access to the Internet.

5. Conclusion

Although, Peer-to-peer technology and its potentials are still being discussed, there is a clear indication that the technology can have considerable impact on applications and distributed systems. The technology is a very heterogeneous work and research area. The concepts of peer-to-peer are used because they increase scalability, improve performance, flexible, dynamic and are inherently fault tolerant. Peer-to-peer concepts are found at the application layer, in distributed systems and even within the communication sub-system. The major application areas for peer-to-peer methods and technologies are Internet and Web based applications, distributed systems, data sharing, and collaboration. The term peer-to-peer is defined by its usage in different contexts and no formal definition exists. The application areas of peer-to-peer concepts are also too heterogeneous to clearly define a fixed set of attributes peer-to-peer systems have to adhere to. Many peer-to-peer systems are charac-

terised by how many of them they implement rather than by a specific well defined sub-set. Although Peer to peer system has many advantages, there are some weaknesses that endanger users data, these weaknesses mostly relate to security issues but modern and new peer to peer systems are more secure and safer. In all there is still room for development and improvements.

References

[1] D. Liben-Nowell, H. Balakrishnan: Observations on the dynamic Evolution of Peer-to-Peer Networks, Proceedings of the 1st International Workshop on Peer-to-Peer Systems, Cambridge, USA, 2002.

[2] K. Kant, R. Iyer, V. Tewari, On the Potential of Peer-to-Peer Computing: Classification and Evaluation, http://kkant.ccwebhost.com/download.htm, 2002

[3] O. Babaoglu, H. Meling, and A. Montresor, Towards Adaptive, Resilient and Self-Organizing Peer-to-Peer Systems. In International Workshop on Peer-to-Peer Computing, 2002

[4] Gnutella website. http://gnutella.wego.com.

[5] Napster. http://www.napster.com.

[6] Ohaha. http://www.ohaha.com/design.html.

[7] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Schenker, A scalable content addressable network. In Proceedings of ACM SIGCOMM, August 2001.

[8] Ion Stoica, Robert Morris, David Karger, M. Frans Kaashoek, and Hari Balakrishnan. Chord: A scalable peer-to-peer lookup service for internet applications. In Proc. ACM SIGCOMM 2001, August 2001. An early version appeared as LCS TR-819 available at http://www.pdos.lcs.mit.edu/chord/papers.

[9] Antony Rowstron Peter Druschel. Past: Persistent and anonymous storage in a peer-to-peer networking environment. In Proceedings of the 8th Conference on Hot Topics in Operating Systems (HotOS 2001), May 2001.

[10] MojoTainment, MojoNation Web Site, http://www.mojotainment.com/mojonation/index.html, 2002

[11] Marling Engle J. I. Khan. Vulnerabilities of P2P systems and a critical look at their solutions, May 2006

[12] Ramesh Subramanian and Brian Goodman (eds), Peer-to-Peer Computing: Evolution of a Disruptive Technology, ISBN 1-59140-429-0, Idea Group Inc., Hershey, PA, USA, 2005.

[13] Stephanos Androutsellis-Theotokis and Diomidis Spinellis. A survey of peer-to-peer content distribution technologies. ACM Computing Surveys, 36(4):335371, December 2004.

[14] P. Leigh, D. Benyola: Future Developments in Per Networking, Equity Research, White Paper, Raymond James Associates, 2001.

[15] Antony Rowstron Peter Druschel. Past: Persistent and anonymous storage in a peer-to-peer networking environment. In Proceedings of the 8th Conference on Hot Topics in Operating Systems (HotOS 2001), May 2001.

[16] Stoica, I., Morris, R., Karger, D., Kaashoek, F., and Balakrishnan, H. (2001). Chord: A Scalable Peer-To-Peer Lookup Service for Internet Applications. In Proc. of SIGCOMM, pages 149160, San Diego, CA.

[17] John F. Buford, Heather Yu, Eng Keong Lua P2P Networking and Applications. ISBN 30-12374-214-5, Morgan Kaufmann, December 2008.

[18] Ralf Steinmetz, Klaus Wehrle (Eds). Peer-to-Peer Systems and Applications. ISBN 3-540-29192-X, Lecture Notes in Computer Science, Volume 3485, September 2005.

[19] Biddle, Peter, Paul England, Marcus Peinado, and Bryan Willman, The Darknet and the Future of Content Distribution. In 2002 ACM Workshop on Digital Rights Management, November 2002.

[20] A. Mauthe: Multimedia Multipeer Communications, PhD Thesis, Lancaster University, 1998.

[21] Shuman Ghosemajumder. Advanced Peer-Based Technology Business Models. MIT Sloan School of Management, 2002.

[22] Rowstron, A. and Druschel, P. (2001). Pastry: Scalable DistributedObject Location and Routing for Large-scale Peer-to- Peer Systems. In Proc. of Middleware, pages Heidelberg, Germany.

[23] Detlef Schoder and Kai Fischbach, Core Concepts in Peer-to-Peer (P2P) Networking. In: Subramanian, R.; Goodman, B. (eds.): P2P Computing: The Evolution of a Disruptive Technology, Idea Group Inc, Hershey. 2005