FILE SHARING

60-367-01

November 27th 2016

Rahul Sharma

Table of Contents

1. Abstract

This paper explores the different aspects of file sharing. This paper will go into detail about the different types of file sharing and its brief history. It will also discuss how file sharing is used throughout the years. This paper will take a look at examples of popular file sharing methods such as BitTorrent and Gnutella as well as their clients. New forms of file sharing which were introduced recently will also be discussed. Examples of these would-be technologies like Dropbox and Google Drive. The purpose of this paper is to inform the reader about the various technologies and protocols used in file sharing systems.

2. Introduction

File sharing is the practice of transmitting files from one computer to another computer over a network. File sharing is often used to distribute or provide access to digital media such as computer programs, audio, video and images. It allows for the sharing of different documents onto a cloud-based server. Different types of file sharing methods are peer-to-peer, file sync, and sharing services. File sharing is legal in both Canada and the US. However, file sharing copy righted material is considered illegal.

3. Types of File Sharing

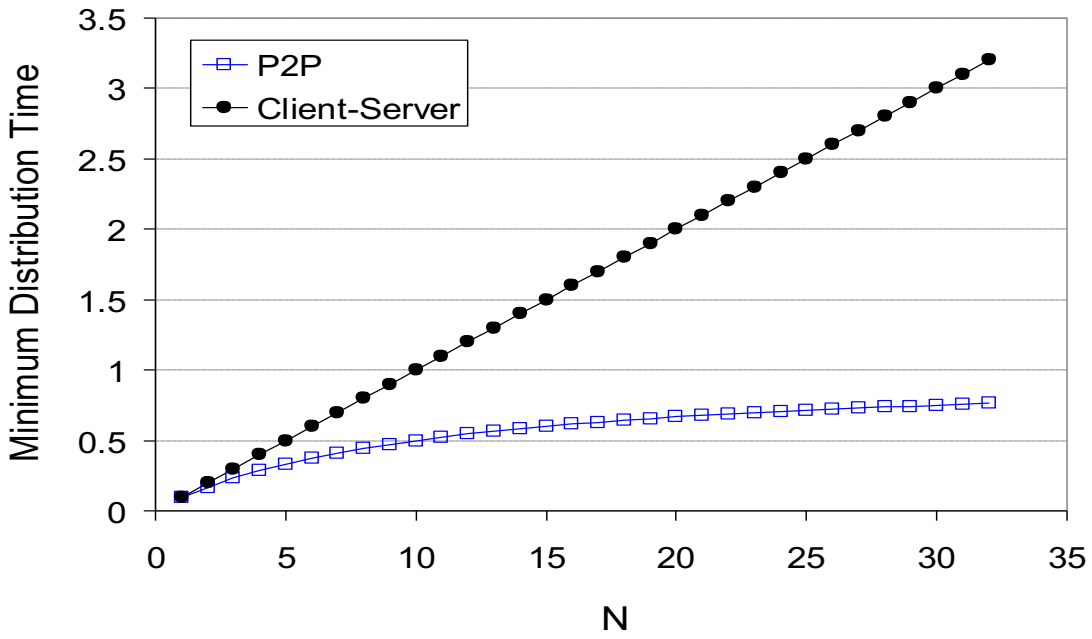
Peer-to-peer (P2P) file sharing is based on P2P application architecture. It uses a software program to locate computers that contain the file a user is searching for. These 'computers' are known as peers. In essence, every peer is treated like a server, by which users can exchange files. A seed is a computer that has a complete copy of a file that is being shared. A leach is a computer that is taking a file that is being shared by a seed. P2P file sharing works as follows:[1]

1. *Computer A* runs a P2P program and sends out a request for the file the user wants to obtain.
2. To locate this file, the software queries other computers that are connected to the internet and running the file sharing software.
3. When the software finds a computer that has the file on its hard drive, the download begins.
4. Others using the file-sharing software can obtain files they want from your computer's hard drive

The P2P model works more efficiently than the client-server model. With more users in the network, a particular file will share at a faster pace. Instead of taking a file from just one user, P2P allows you to take different parts of the file from hundreds of other users. The diagram below displays the minimum distribution time between P2P and Client-Server technologies. The

[1] Http://www.howstuffworks.com/hsw-contact.htm, "How BitTorrent Works," HowStuffWorks, 2005, 4, accessed November 21, 2016, http://computer.howstuffworks.com/bittorrent1.htm.

minimum distribution time is represented in the vertical axis. The N in the diagram is represented in the horizontal axis. The N represents the number of client's data is being distributed to.



Another type of file sharing is the file sync and sharing services method. In this technique, files are uploaded on a shared directory between different users. Files are shared via the cloud. Google Drive and Dropbox are popular file sync sharing services. These files are easily accessible on different platforms.

4. Modern History

Modern file sharing began in the early 2000s. This is when file sharing protocols and file formats were first introduced. Projects like Gnutella and BitTorrent were created for the primary reason to share files between different users. Initially, P2P servers had a central server where all the data of different peers was stored. This model was later changed because if a central server is down, communication between different peers would be disabled completely. In 1999 Napster released, a new way to share MP3 files based on a centralized P2P network. Napster was later shut down in 2001 after a legal battle with the RIAA (Record Industry Association of America). In 2000, Gnutella and LimeWire launched. LimeWire was later shutdown in 2010 due to a

permanent injunction by a New York based federal court. In 2001, BitTorrent was released in order to fill in the void left by Napster. In 2003, 'The Pirate Bay' launched from Sweden. It later became one of the most popular sites on the internet and was raided multiple times by the police.[2]

5. BitTorrent

5.1 – Type

- Developer: BitTorrent
- Software License: Proprietary, Adware

5.2 – Protocol

BitTorrent is a communications protocol for P2P file sharing. It is most commonly used to transfer large files such as video and audio. To send or receive files, a user must use a BitTorrent Client. A client is a program that implements the BitTorrent protocol. The BitTorrent protocol coordinates file transfer among all peers sharing a torrent. Torrent files are files that contain text and point out the trackers for a download to begin. Torrent files are commonly used in file sharing. Common clients used include uTorrent, Deluge, Flud, Xunlei and various others. The BitTorrent Protocol is used to reduce the impact of large files being shared over a server, thus lowering the total bandwidth used. Rather than downloading a file from a single peer, the protocol allows the user to download multiple peers who are uploading and downloading a file. This also increases the speed of the download. The file being distributed is divided into segments called pieces. Each piece is protected by a cryptographic hash contained in the torrent descriptor, ensuring that any modification can be easily detected.[3] Pieces are often downloaded non-sequentially and are later re-arranged by the BitTorrent Client. Throughout the download, pieces are the same size allowing for a file to be paused and resumed at a later time.

[2] @geekdotcom, "LimeWire, Napster, The Pirate Bay: A Brief History of File Sharing | Gadgets | Geek.com," @geekdotcom,6, accessed November 21, 2016, http://www.geek.com/gadgets/limewire-napster-the-pirate-bay-a-brief-history-of-file-sharing-1359473/.
[3] Bram Cohen, "BitTorrent.org," BitTorrent.org,6, accessed November 21, 2016, http://www.bittorrent.org/beps/bep_0003.html.

5.3 – Technologies

The BitTorrent protocol has developed various technologies over the years.

5.3.1 - Trackers

A tracker is a special type of server that assists communication between peers. The tracker server keeps track of where file copies reside on peer machines, then they help coordinate efficient transmission and reassembly of the copied file. Public trackers are used by adding the tracker address to an existing torrent by anyone. Private trackers restrict usage by having a user register to the website. Multi-trackers for a single torrent were introduced to increase efficiency. If one tracker fails, other trackers can continue to maintain the swarm for one torrent. Multi-trackers may increase download speed - but can often be poorly implemented leading to overhead-traffic. On May 2nd, 2005, trackerless torrents were introduced. These are implemented through a distributed hash table. In trackerless torrents, peer IP addresses are stored in the DHT, then are accessed using the BitTorrent info hash as the key.

5.3.2 - Web Seeding

This technology was implemented in 2006 and allows for clients to download pieces of torrent files. A website may distribute a torrent for a particular file and make that file available for download from that same web server. This simplifies long-term seeding and load balancing through using existing web hosting setups. Hash web seeding is serves content through info-hash and piece number rather than the filename.

5.3.3 - RSS Feeds

Broadcatching which is the downloading of digital content that has been made available over the Internet using RSS. It combines RSS feeds with the BitTorrent protocol to create a content delivery system. The RSS feed will track the content, while BitTorrent is continuing to ensure content integrity with cryptographic hashing the data. As a result, feed subscribers will receive uncorrupted content.

5.3.4 - Encryption

Some ISPs have chosen to slow down BitTorrent Transfers. The reason for this is because BitTorrent makes up a large portion of total traffic. As a result, BitTorrent has decided to disguise its traffic. Protocol encryption (PE), message stream encryption (MSE), and protocol header encryption (PHE) are features included with BitTorrent to enhance privacy. These methods make it harder for torrents to be identified by ISPs.

5.4 – uTorrent

uTorrent is a BitTorrent client owned fully by BitTorrent. It is one of the most popular clients with over 100 million users. [4] It was created in 2004 by Ludvig Strigeus. It comes with BitTorrent technology including trackers, seeding and encryption. The application is an executable file which is available for Windows, Mac OSX and Linux.

6. Gnutella

6.1 – Type

- Developer: Justin Frankel
- Software License: GNU General Public License.

6.2 – Protocol

Pronounced “new-tella”, Gnutella is a protocol for distributed searches. It is a large peer-to-peer network which was the first decentralized network of its kind. In this model, every client is a server that provides a client-side interface for users to issue queries, view search results, accept queries from other servers, check for matches against their local data set, and respond with applicable results. The Gnutella protocol consists of a set of descriptors used for communication between different servers.

[4] @torrentfreak, "Spotify Reminded of UTorrent Past After Branding Grooveshark 'Pirates' - TorrentFreak," TorrentFreak, 2014, 8, accessed November 21, 2016, https://torrentfreak.com/spotify-reminded-of-utorrent-past-after-branding-grooveshark-pirates-141112/.

Descriptor	Description
Ping	Used to actively discover hosts on the network. A server receiving a Ping descriptor is expected to respond with one or more Pong descriptors.
Pong	The response to a Ping. Includes the address of a connected Gnutella server and information regarding the amount of data it is making available to the network
Query	The primary mechanism for searching the distributed network. A server receiving a Query descriptor will respond with a QueryHit if a match is found against its local data set.
QueryHit	The response to a Query. This descriptor provides the recipient with enough information to acquire the data matching the corresponding Query
Push	A mechanism that allows a firewalled server to contribute file-based data to the network

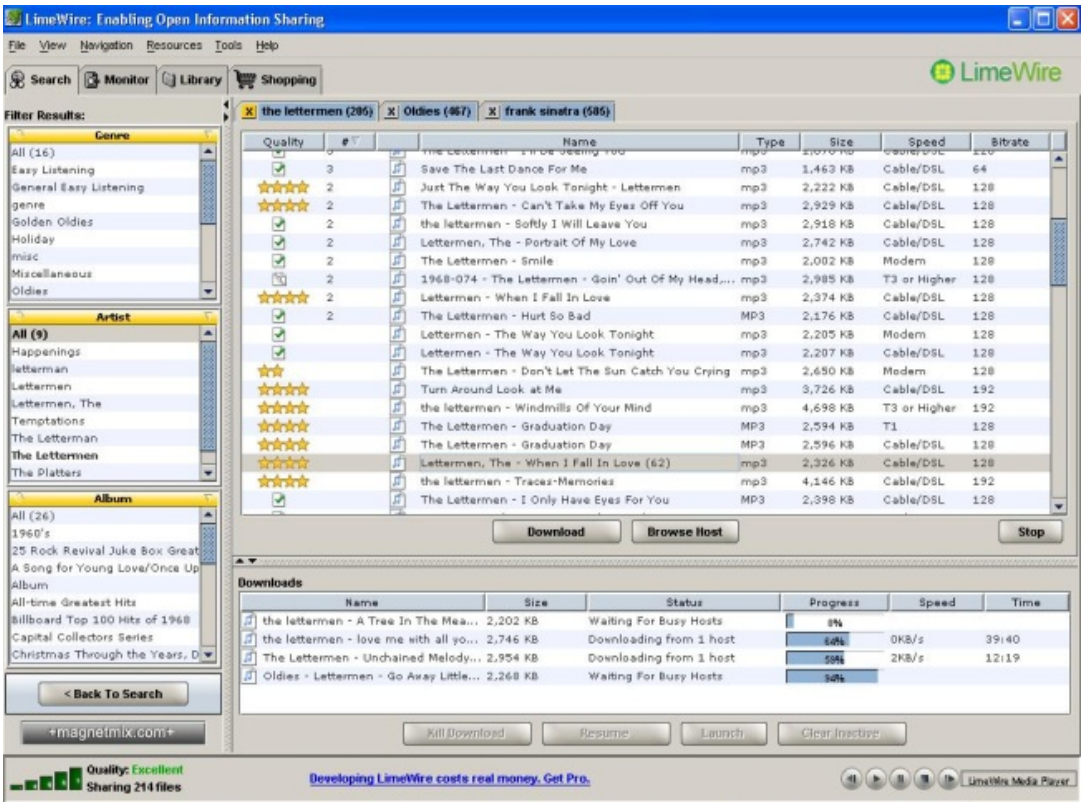
[5]

[5]"Efficient Massive Contents Distribution Strategy for P2P Using Sensor Smart Network," International Journal of Distributed Sensor Networks, 2015, 9, accessed November 21, 2016, http://dsn.sagepub.com/content/11/10/569063.full

The server establishes a connection by creating one with a server that already is on the network. The TCP/IP connection to the server is created once the address of another server is received.

6.3 – LimeWire

Like BitTorrent, Gnutella uses external clients to facilitate the download of peer to peer data. A Gnutella client is similar to a search engine, in which users can search for a file which is currently being shared. LimeWire is a popular client used on the Gnutella server. It released in May of 2000 and shut down in 2010 after lawsuits arose by the Federal Court regarding copyright claims[6]. LimeWire is a program a user can install onto their PCs. The user then has the ability to search through the Gnutella server, and find files. The files are downloaded through uploads (people who are sharing the file).



An image of the LimeWire interface[7]

[6] Josh Halliday, "LimeWire Shut down by Federal Court," The Guardian, 2010, 10, accessed November 21, 2016, https://www.theguardian.com/technology/2010/oct/27/limewire-shut-down.
[7] Ezra Lowman, "Ezra," Prezi.com, 2013, 10, accessed November 21, 2016, https://prezi.com/xbx7ebcu6whd/ezra/.

6.3.1 – LimeWire Technologies

LimeWire uses the Digital Audio Access Protocol (DAAP) for sharing of its library. DAAP was introduced by iTunes as it is a specialized HTTP server.[8] It sends a list of queries and streams the requested query identified by the user. DAAP also notifies the client if there are any changes in the server. As a result, after downloading a file on LimeWire, the user is given the ability to stream it straight from the LimeWire program. In addition, LimeWire allows connections to be encrypted with transport layer security (TLS). TLS is an cryptographic protocol that provides communication security over a computer network. This provided additional security for each person using LimeWire.

7. Relationships Among Services

There are key similarities and differences between Gnutella, LimeWire and BitTorrent. First off, Gnutella, LimeWire and BitTorrent are similar because both are P2P file sharing servers. Gnutella and BitTorrent are both protocols, unlike LimeWire and uTorrent which are clients. Gnutella and BitTorrent are both have a decentralized network. They differ in the fact that BitTorrent has many more features such as trackers and RSS feeds. In BitTorrent, trackers are used to communicate among peers. In Gnutella, the main server is used for communication once the user receives a valid TCP/IP connection. These programs are similar due to the fact that they are both clients and are used to P2P file sharing. Since uTorrent is a product of BitTorrent, it comes with all of its features. This is different due to the fact that LimeWire does not have all of these features as it is a product of Gnutella. LimeWire also allows you to stream the file downloaded (usually audio, video) from the program, whereas, uTorrent downloads it into a folder which is then accessible.

8. Legal Issues and Other Risks

P2P file sharing has brought up numerous amounts of legal issues throughout the course of its existence. Software like BitTorrent and Gnutella are legal because they are just a tool to share

[8] "Digital Audio Access Protocol (DAAP)," DAAP - Digital Audio Access Protocol, 10, accessed November 21, 2016, http://daap.sourceforge.net/docs/.

files. The user is responsible for what he or she uses the software for. Under copyright law, it is illegal to download or share copyrighted materials such as music or movies without the permission of the copyright owner. [9] For example, LimeWire was shut down because a U.S Federal court judge issued an injunction forcing the stoppage of "the searching, downloading, uploading, file trading...and/or all functionality".[10] This occurred after various record companies sued LimeWire for providing their music illegally. However, this injunction just opened the doors for other software which have similar features. Many of these P2P software services have been shut down, but eventually, they are resilient as they get back up on their feet. An example of this would be ThePirateBay. Its slogan is "the most resilient BitTorrent site" because it has been shut down many times but finds itself still standing. Other risks with P2P file sharing include the risk of viruses. Since a user is downloading files from other users, they are always prone to downloading viruses which can affect their PCs.

9. Cloud Based File Sharing

Cloud computing is a technique in which data can be stored and accessed over the internet instead of using a computer's hard drive. [11] Cloud-based file sharing is gaining popularity quickly. It allows a user to store files onto the internet, that can be later be shared with different users. This file sharing technique makes it easier to access files without allocating space on hard drive storage. This is often used by companies to help make files easily attainable and to save space on their main hard drives. The cloud also allows for portability, meaning that a user is able to access a file from any location (if he is given access to the file to begin with)[12]. There are many companies who give access to cloud-based services. Dropbox, Oracle Cloud, Google Drive, and Microsoft OneDrive are some of the larger ones.

[9] Libtutorials101rhc Follow, "Week 15 Copyright Plagiarism(," Share and Discover Knowledge on LinkedIn SlideShare, 2016, 11, accessed November 21, 2016, http://www.slideshare.net/libtutorials101rhc/week-15-copyright-plagiarism.
[10] By Greg Sandoval October 26, 2010 1:02 PM PDT @sandoCNET, "Judge Slaps Lime Wire with Permanent Injunction," CNET, 2010, 12, accessed November 22, 2016, https://www.cnet.com/news/judge-slaps-lime-wire-with-permanent-injunction/.
[11] By Eric Griffith May 3, 2016 10 Comments, "What Is Cloud Computing?," PCMAG, 12, accessed November 22, 2016, http://www.pcmag.com/article2/0,2817,2372163,00.asp.
[12] "Everything You Need to Know about Cloud-based File Sharing," SearchCloudStorage, 12, accessed November 22, 2016, http://searchcloudstorage.techtarget.com/feature/Everything-you-need-to-know-about-cloud-based-file-sharing.

9.1 – Dropbox

Dropbox is a file hosting service that consists of cloud based servers. It offers cloud storage, file synchronization, personal cloud and client software. Dropbox was founded in 2007, by MIT students Drew Houston and Arash Ferdowsi. One of their main technologies is the 'sync' feature. This allows for synchronization of data across various platforms, allowing users to collaborate amongst each other. There are three types of syncing: delta, LAN and streaming. Delta syncing speeds up the transfer of updating files by uploading and downloading modified portions of the file instead of transferring entire files on every change. For example, if the user edits a file that's already in his or her Dropbox, the system will detect this and make changes for the bits that have changed. LAN sync is used for new files in a local area network. This bypasses the need to download the file from the cloud if they are being synced between devices on the same router. As a result, the process is sped up. Lastly, streaming sync was introduced to improve the latency and accelerate the transfer of larger files by downloading on multiple drives.[13]

10. File Sharing Moving Forward

File sharing has evolved throughout the years. Early on, P2P file sharing was extremely popular, however it caused lots of legal problems. Different programs that implemented P2P file sharing were constantly shut down then re-created with a different program name or URL. Nevertheless, P2P file sharing will continue to be used. The future of file sharing is in file sync and cloud based technology. In the last five years, it has gained massive popularity. Companies like Dropbox, Google, Microsoft and many others are profiting from their cloud-based services. Having data stored on the internet, has increased production for many users due to its convenience. Both P2P and file sync/cloud-based technologies will continue to impact the internet.

[13] Lee Bell, "Https://www.theinquirer.net/inquirer/feature/2431930/the-technology-behind-dropbox-sync," The Inquirer, October 30, 2015, 13, accessed November 22, 2016, https://www.theinquirer.net/inquirer/feature/2431930/the-technology-behind-dropbox-sync.

11. References

- Bell, Lee. "Https://www.theinquirer.net/inquirer/feature/2431930/the-technology-behind-dropbox-sync." The Inquirer. October 30, 2015. Accessed November 22, 2016. https://www.theinquirer.net/inquirer/feature/2431930/the-technology-behind-dropbox-sync.
- By Greg Sandoval October 26, 2010 1:02 PM PDT @sandoCNET. "Judge Slaps Lime Wire with Permanent Injunction." CNET. 2010. Accessed November 22, 2016. https://www.cnet.com/news/judge-slaps-lime-wire-with-permanent-injunction/.
- By Eric Griffith May 3, 2016 10 Comments. "What Is Cloud Computing?" PCMAG. Accessed November 22, 2016. http://www.pcmag.com/article2/0,2817,2372163,00.asp.
- Cohen, Bram. "BitTorrent.org." BitTorrent.org. Accessed November 21, 2016. http://www.bittorrent.org/beps/bep_0003.html.
- "Digital Audio Access Protocol (DAAP)." DAAP - Digital Audio Access Protocol. Accessed November 21, 2016. http://daap.sourceforge.net/docs/.
- "Efficient Massive Contents Distribution Strategy for P2P Using Sensor Smart Network." International Journal of Distributed Sensor Networks. 2015. Accessed November 21, 2016. http://dsn.sagepub.com/content/11/10/569063.full.
- "Everything You Need to Know about Cloud-based File Sharing." SearchCloudStorage. Accessed November 22, 2016. http://searchcloudstorage.techtarget.com/feature/Everything-you-need-to-know-about-cloud-based-file-sharing.
- Halliday, Josh. "LimeWire Shut down by Federal Court." The Guardian. 2010. Accessed November 21, 2016. https:/www.theguardian.com/technology/2010/oct/27/limewire-shut-down.
- Http://www.howstuffworks.com/hsw-contact.htm. "How BitTorrent Works." HowStuffWorks. 2005. Accessed November 21, 2016. http://computer.howstuffworks.com/bittorrent1.htm.
- Libtutorials101rhc Follow. "Week 15 Copyright Plagiarism(." Share and Discover Knowledge on LinkedIn SlideShare. 2016. Accessed November 21, 2016. http://www.slideshare.net/libtutorials101rhc/week-15-copyright-plagiarism.
- @geekdotcom. "LimeWire, Napster, The Pirate Bay: A Brief History of File Sharing | Gadgets | Geek.com." @geekdotcom. Accessed November 21, 2016. http://www.geek.com/gadgets/limewire-napster-the-pirate-bay-a-brief-history-of-file-sharing-1359473/.
- Lowman, Ezra. "Ezra." Prezi.com. 2013. Accessed November 21, 2016. https://prezi.com/xbx7ebcu6whd/ezra/.
- @torrentfreak. "Spotify Reminded of UTorrent Past After Branding Grooveshark 'Pirates' - TorrentFreak." TorrentFreak. 2014. Accessed November 21, 2016. https://torrentfreak.com/spotify-reminded-of-utorrent-past-after-branding-grooveshark-pirates-141112/.