BitTorrent

BitTorrent

Ankit Mishra
Information Technology, MNNIT
ankit14498@gmail.com

Harsh Mishra
Information Technology, MNNIT
imharshmishra07@gmail.com

Nitesh Sharma
Information Technology, MNNIT
nitesh786sharma@gmail.com

Ananya Pandey
Information Technology, MNNIT
ananya.01998@gmail.com

ABSTRACT

In recent years, peer-to-peer networking received a lot of attention due to ongoing battle with music and
movie industry. It is not a new concept, has existence from past 4 decades and can be traced back to
original implementation of the Internet.

It is a distributed P2P system, has the potential to change the model of broadcast media and file
distribution. It uses symmetric(tit-for-tat) transferring model in an attempt to reach Pareto efficiency.
Its protocol employs various mechanisms and algorithms in a continuous effort to try to ensure that
there are no other variations in the network arrangement which will make every downloader at least as
well off and at least one downloader strictly better off.

BitTorrrent base operation around the concept of a torrent file, a centralized tracker and an associated
swarm of peers. The centralized tracker provides the different entities with an address list over available
peers.

1 INTRODUCTION

Comparing to more common server-client solution, a P2P approach has several advantages including

- Increased robustness
- Resource providing such as bandwidth, storage space by peers

Authors addresses: Ankit Mishra, Information Technology, MNNIT; Harsh Mishra, Information Technology, MNNIT; Nitesh Sharma, Information Technology, MNNIT; Ananya Pandey, Information Technology, MNNIT.

BitTorrent has proven to be and efficient and reliable alternative to classical server-client solution.
"The emergence of peer-to-peer computing signifies a revolution in connectivity that will be as profound
to the Internet of the future as Mosaic was to the Web of the past" - Patrick Gelsinger, Vice President
and CTO, Intel Corp.

In simplest definition, P2P:
"A communications model in which each party has the same capabilities and either party can
initiate a communication session".

This means P2P computing is an alternative to traditional client/server architecture where there is a
single (or some cluster) server and many clients.

A more precise definition is stated :
"A distributed network architecture may be called a Peer-to-Peer (P-to-P, P2P,...) network, if the
participants share a part of their own (hardware) resources (processing power, storage capacity,
network link capacity, printers, ...). These shared resources are necessary to provide the Service and
content offered by the network (e.g. file sharing or shared workspaces for collaboration). They are
accessible by other peers directly, without passing intermediary entities. The participants of such a
network are thus resource (Service and content) providers as well as resource (Service and content)
requestors (Servant-concept)."

2 P2P NETWORK TOPOLOGIES

2.1 Pure P2P

"A distributed network architecture has to be classified as a "Pure" Peer-to-Peer network, if it is firstly a
Peer-toPeer network according to Definition 1 and secondly if any single, arbitrary chosen Terminal
Entity can be removed from the network without having the network suffering any loss of network
service."

2.2 Hybrid P2P

"A distributed network architecture has to be classified as a "Hybrid" Peer-to-Peer network, if it is firstly
a Peer-to-Peer network according to Definition 1 and secondly a

central entity is necessary to provide
parts of the offered network services."

3 INTRODUCTION TO BITTORRENT

It is a protocol which makes the distribution of files, mainly larger files, easier and less bandwidth
consuming for the publisher. It can be done by utilizing the upload capacity of the peers that are
downloading a file. Some increase in number of downloaders will only result in the load on the publisher
hosting the file.

Basic flow of BitTorrent:

- If we assume the upload capacity of the server is the same as the download capacity of a peer,
  the time for the download to finish will be two times the time if only one peer where
  downloading from the server.
- By splitting the file and send one part to each peer, and let the peers download the part they are
  missing from each other, both download time and load on the server is reduced.

BitTorrent protocol is much more complex than above key points but the idea is much similar.

4 AREA OF USAGE

It is possible to download Linux distributions using BitTorrent which enables much faster download than
the regular FTP or HTTP can provide. Internet browser like Opera can be downloaded.
In addition to increased download speed, it has benefit that it eases the pressure on the servers when a
new version is released.
BitTorrent also has potential business usage. Distribution of ISO-images, OS, large software and patches
can be done at higher speeds using BitTorrent. Within an organization one can also use the protocol to
distribute applications and updates more rapidly.
It is easy for a user to find different parts of the file and download them quickly. This can be called the
multiplier effect, and a slightly popular show or movie can become insanely popular within days, or
maybe within hours.

5 BITTORRENT ARCHITECTURE

Its architecture normally consists of following entities :

- A static metainfo file("a torrent file")
- A tracker
- An original downloader("seed")
- The end user downloader("leecher")

Steps in publishing a file using BitTorrent

(1) Create a metainfo file from the file that you want to publish. The torrent file contains the
filename, size, hashing information and the URL of the "tracker". The torrent file can be distributed by e-
mail, IRC, HTTP etc

(2) To download or "seed" a file, you need a BitTorrent client. The BitTorrent client is a free application
that administrates the download procedure.
(3) A BitTorrent download is started by opening the torrent file in the BitTorrent client.
(4) The tracker keeps a log of peers that are currently downloading a file, and helps them find each other.
The tracker is not directly involved in the transfer of data and does not have a copy of the file. The
tracker and the downloading users exchange information using a simple protocol on top of HTTP.
(5) The user gives information to the tracker about which file it's downloading, ports its listening on etc.
The response from the tracker is a list of other users which are downloading the same file and
information on how to contact them. This group of peers that all share the same torrent represents a
'swarm'.
(6) An original downloader known as a "seed" has to be started. A "seed" is a user that has the entire file.
A downloading user that has nothing or only parts of a file is called a "leecher". The "seed" must upload
at least one complete copy of the file. Once an entire copy is distributed amongst the other
downloaders, the 'seed' can stop uploading and the download will still continue for all downloaders as
long as there are enough people downloading the file, and all pieces of the file are available.
(7) For a popular file one complete copy from the seed 7 may be enough while for a less popular file,
continues uploading by the seed may be necessary. The result is that the bandwidth requirements for
the publisher are less if many people are downloading.

6 ALGORITHMS

Choosing peers to connect to is a two-sided problem. We need a way of finding the best sequence of
downloading the pieces. This is determined by the piece selection algorithm. A peer, who has the piece
you want, might not let us to download it. Strategies for peers not allowing other peers to download
from them is known as choking, and concerns resource allocation.

6.1 The Piece Selection Algorithm

The goal is to replicate different pieces on different peers as soon as possible. This will increase
the download speed, and also make sure that all pieces of a file is somewhere in the network if
the seeder leaves.

Several policies of algorithm:

6.1.1 Strict Policy. Once a sub-piece has been requested, the remaining sub-pieces for that particular piece are
requested before sub-pieces from any other piece. This helps us to get a complete piece as
quickly as possible.

6.1.2   Rarest First. The main policy in BitTorrent is that of "rarest first". This means that when a peer selects
the next piece to download, it selects the piece which the fewest of their peers have.

6.1.3   Random First Piece. Once you start downloading, you don't have anything to upload. It is important to get the
first piece as fast as possible, and this means that the "rarest first"-policy is not the most
efficient. Rare pieces tend to be downloaded slower, because you can download its sub-
pieces from only one (or maybe a few) other peers. As mentioned earlier, multiple peers with the same piece increase the download speed. The policy is then to select the first piece
randomly. When the first piece is complete, we change to "rarest first".

6.1.4   Endgame mode. Sometimes a piece might be downloaded from a peer with a slow transfer rate. This can
potentially delay the finishing of a download. To prevent this we have the "endgame mode".

7   RESOURCE ALLOCATION

No centralized resource allocation exists in BitTorrent. Every peer is responsible for maximizing its
download rate. A peer will try to download from whoever they can. To decide which peers to upload to,
a peer uses a variant of the "tit-for-tat" algorithm. The "tit-for-tat"-strategy comes from repeated game
theory, and is a strategy of cooperation based on reciprocity.

7.1   The Choking Algorithm

Choking is a temporary refusal to upload to another peer, but you can still download from
him/her. To cooperate peers allow uploading, and to not cooperate they "choke" the
connection to their peers.
The goal is to have several bidirectional connections at any time, and achieve "Pareto
efficiency.
So, a peer always unchokes a fixed number of its peers (the default is four). . Deciding which
peers to unchoke is determined only by the current download rates. It has been chosen to
use a 20-second average to decide this. Due to the usage of TCP it's not desirable to choke
and unchoke too rapidly. Thus, this is calculated every ten seconds. The result is that any peer will upload to peers which provide the best download rate.
It prohibits a large number of "free riders" which are peers who only download and don't
allow uploading. In order for a peer-to-peer network to be efficient all peers have to
contribute to the network. This restriction is not present in most other peer-to-peer
protocols and applications, and is one of the reasons why BitTorrent has become so popular.

7.2   Optimistic unchoking

BitTorrent also allows an additional unchoked peer, where the download rate criterion isn't
used. This is called optimistic unchoking.

8   IMPROVEMENTS

8.1   Bulk traffic marking

Version 4 of the protocol had one technical improvement which made it more business-friendly. The
BitTorrent traffic is now marked as bulk traffic which makes traffic shaping a lot easier.

8.2   Decentralized tracker

With one centralized tracker the BitTorrent network is not very fault tolerant. If the tracker goes
down the file will no longer be available, since there are no way the peers could know about each other. One centralized tracker also makes the network vulnerable to denial of service
attacks.
Version 4.1 was released by Cohen and the big newsflash was support for a decentralized
tracker.
The tracker is distributed in the sense that every client or node in the network now acts a
lightweight tracker. The solution is based on distributed hash tables (DHTs). This makes it
possible to share files with minimal resources, but no guarantees can be made as respect to
reliability.

9   DISTRIBUTED HASH TABLES (DHTS)

A DHT is a decentralized distributed system. The system consists of a set of participating nodes and
a set of keys. The DHT performs the function of a hash table. Key and value pair can be stored and a
value can be looked up if the correct key is provided. What separates a DHT from an ordinary hash
table is that the storage and lookups are distributed among the nodes (machines) in the network. All
nodes are peers that can join and leave the network freely. DHTs makes provable guarantees about
performance despite the seemingly chaos created by random joining and leaving peers.
   Any DHT protocol emphasizes the following characteristics:

- Decentralization
- Scalability
- Fault tolerance

10   FUTURE OF BITTORRENT

BitTorrent is a cheap and reliable alternative for normal business file distribution. It saves the content
publishers considerable money in bandwidth by utilizing their own customers in distributing the
companies' content. Large corporations like Blizzard Entertainment Inc, has already begun utilizing the

power of BitTorrent in the distribution of its patches for the online game, The World of Warcraft.

11 CONCLUSION

In just a few years BitTorrent has become the most popular protocol for peer-to-peer networking. It's
simple in use, and has powerful capabilities which enable faster download of files, and more fairness
among the peers, than most traditional peer-to-peer protocols. It has proven to be a well-designed and
powerful file sharing protocol, based on the extensive adaptation and usage by the Internet users. But, it
is by no means optimized, and thus has the potential to only get better as time goes by. Originally the
protocol involved a central tracker in order to let peers find other peers, i.e. a hybrid peer-to-peer
protocol. New versions of the protocol has removed this central tracker and distributed this function
onto the peers themselves. This improvement has made BitTorrent a pure peer-to-peer protocol. The protocol is highly self-configuring. Peers find each other using the centralized or distributed tracker,
without any user involvement. Peers choose which peers to download pieces of the file from using
policies and algorithms embedded in the protocol. BitTorrent seems to have the potential to
revolutionize the landscape of broadcast media and file distribution. The internet can become the
world's largest source for Video-on-Demand. We can envision a future where people won't watch entire
shows; they'll just download and watch the parts they care about. BitTorrent can also leverage the costs
of distributing shows and movies, making broadcasting possible for almost every Internet user. This can
have tremendous effects for the large networks, and especially the content providers. The industry
needs to change and adopt the new technology before they are over run.