

# Analysis of BitTorrent Flow Behavior on Large-Scale Networks

Chen Liang

College of Computer Science & Engineering  
Southeast University, Nanjing, P.R.China

2007.04

# Outline

- Background
- Trace Collection
- Overview of BitTorrent Traffic
- Characteristics of BitTorrent Flow
- Characteristics of BitTorrent Peer
- Conclusion

# Background

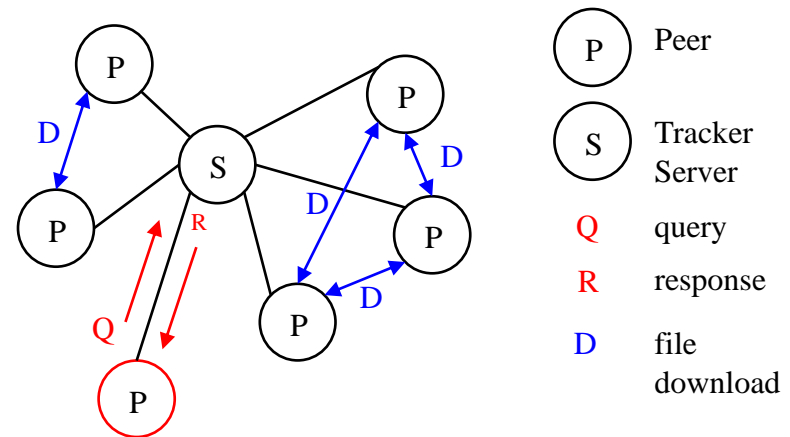
- Peer-to-Peer(P2P) file sharing applications have evolved to the major traffic sources in the Internet. In particular, the BitTorrent file sharing system is causing high amounts of traffic volume in today's networks.
- The characteristics of BitTorrent traffic have strong influence on the characteristics of the whole Internet traffic.
- Previous researches all use default port number to identify BitTorrent traffic, which makes the result sometimes inaccurate.
- Understand the characteristics of BitTorrent flow behavior accurately is important.

# What is BitTorrent?

- BitTorrent is a file distribution system based on the P2P paradigm.
- It capitalizes the resources of peer nodes to efficiently distribute large contents.
- Unlike other well-known P2P applications, such as eDonkey or Gnutella, which strive to quickly locate hosts that hold a given file, the sole objective of BitTorrent is to quickly replicate a single large file to a set of clients.

# How BitTorrent Works?

- A BitTorrent network contains all the currently active peers and a central component, called *tracker* server.
- Two kinds of peers
  - *seeds*
  - *leechers*



Structure of BitTorrent Network

- The tracker is not involved in the actual distribution of the file; instead, it keeps meta-information about the peers that are currently active and acts as a rendezvous point for all the clients of the torrent.

# How BitTorrent Works? (2)

- A user joins an existing torrent by downloading a torrent file, which contains the IP address of the tracker.
- Upon joining the torrent, a new client uses HTTP to receive from the tracker a list of active peers to connect to.
- Then the new client connects to each active peer using BitTorrent protocol to download the file he wants.

# Trace Collection

- Trace was collected by CERNET Eastern (North) Regional Center at Nov 10, 2005, using Watcher1.1.
- Watcher is a General-purpose Measurement Platform on Large-scale High-speed Network, which is designed and implemented by CERNET Eastern (North) Regional Center. It can capture and analyzes network traffic on any number of 1Gb channels.
- We only capture TCP traffic and identify BitTorrent traffic by examining the packet payload.

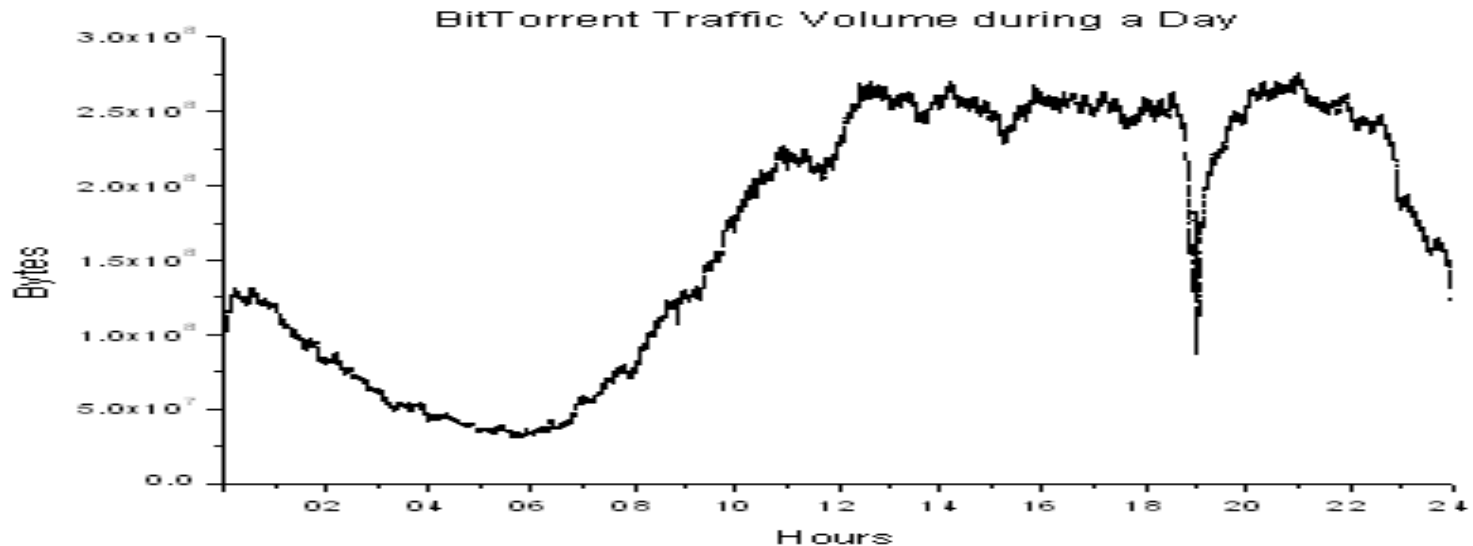
# Trace Description

- Start Time: 2005/11/10 00:00
- Duration: 24 hours
- Available Bandwidth: 1G\*2\*3(6Gb)
- Capture Length: 60Bytes
- Number of TCP Flow: 814.9M
- bps: 2054.5M
- pps: 419.1K

# Overview of BitTorrent Traffic

	Num. of Flows	Num. of Pkts	Num. of Bytes	Avg. Flow Length (pkt)	Avg. Flow Length (Byte)	Avg. Pkt. Size (Byte)
TCP	814.9 M	34.5G	21.7T	43.4	26.6K	627.5
Bit-Torrent	19.5M	19.0G	13.4T	999.8	686.2K	702.7
Bit-Torrent/ TCP	0.024	0.55	0.62	23.0	25.8	1.1

# Overview of BitTorrent Traffic



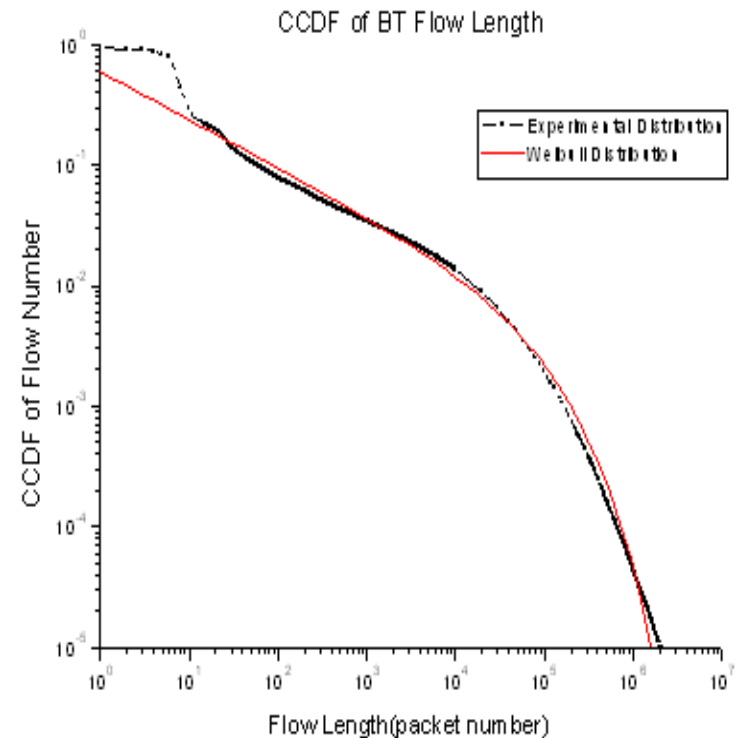
- The overall volume varies much throughout the day.
- The largest volume is six times larger than the smallest volume (at 6a.m.).
- Mainly because of human behavior.

# Characteristics of BitTorrent Flow

- BitTorrent Flow Length
- BitTorrent Flow Duration
- BitTorrent Flow Rate

# BitTorrent Flow Length

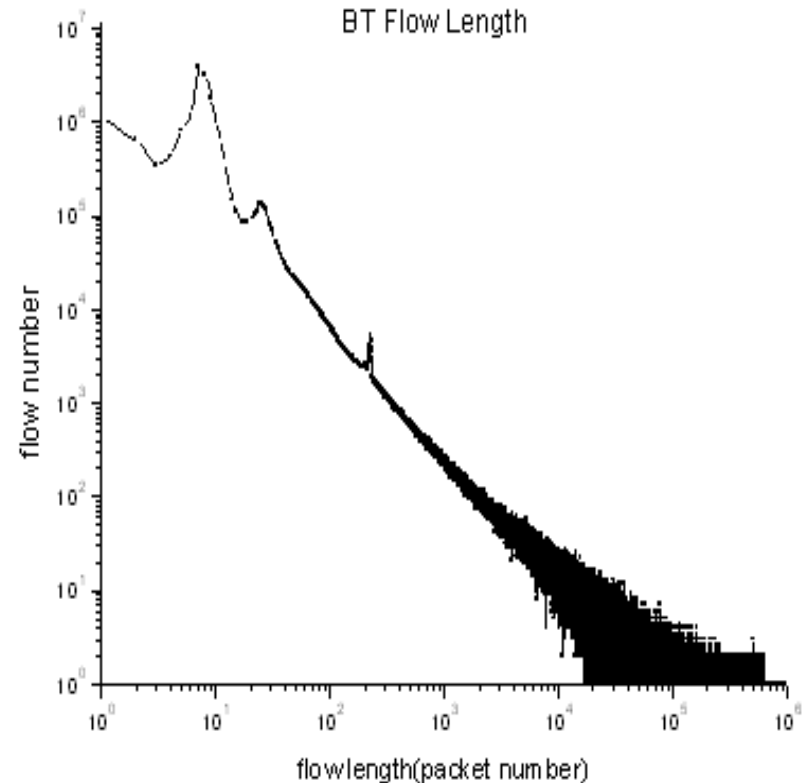
- Heavy-tailed
- The flows containing less than 100 packets represent more than 90% number of flows, and they contribute to less than 2% of the BitTorrent traffic volume.
- We call the flows containing more than 100 packets as “*elephant flow*”.
- Weibull distribution fits the distribution of BitTorrent flow length very well.



CCDF of BitTorrent Flow Length

# BitTorrent Flow Length (2)

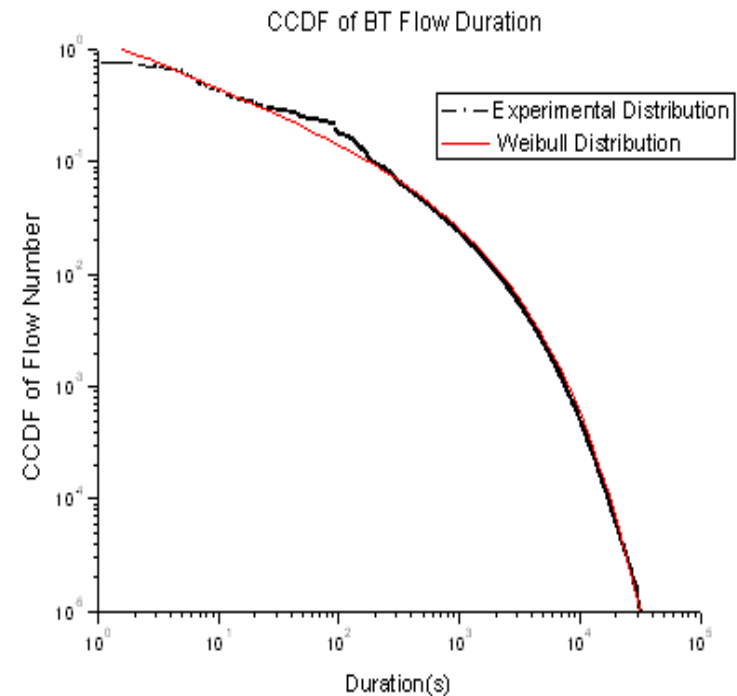
- Not exactly heavy-tailed.
- The pike when flow length is about 7 is caused by BitTorrent Protocol connection refused.
- The pike when flow length is about 30 and 120 is caused by route circulation.



Distribution of BitTorrent Flow Length

# BitTorrent Flow Duration

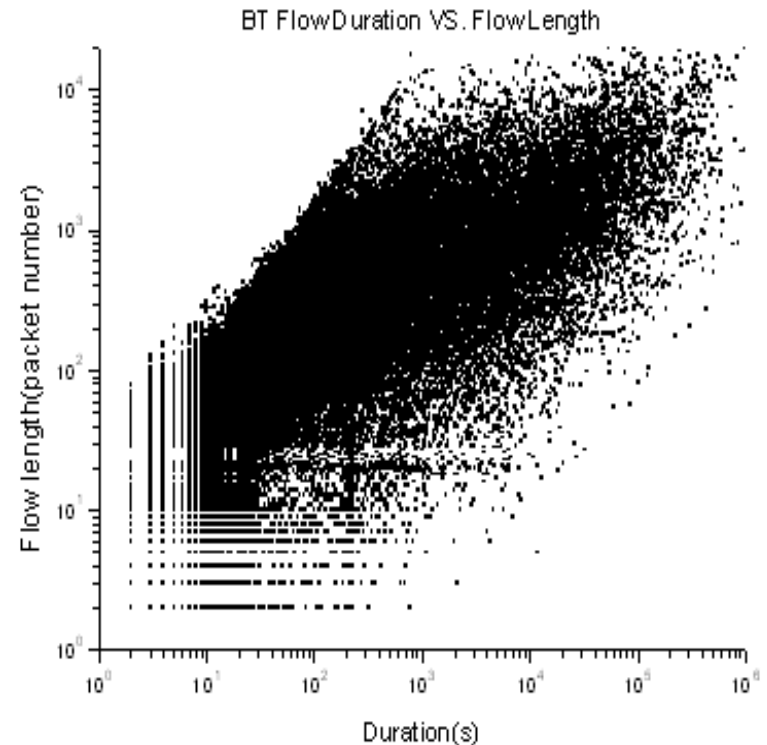
- The flows whose duration are less than 100 seconds represent more than 90% number of flows, and they contribute to less than 2% of BitTorrent flow traffic.
- We call the flows lasting more than 100 seconds as “*tortoise flow*”.
- Weibull distribution also fits the distribution of BitTorrent flow duration well.



CCDF of BitTorrent Flow Duration

# Flow Duration versus Flow Length

- Generally, the longer a flow is, the bigger it is.
- “*elephant flow*” corresponding to “*tortoise flow*”.
- If we could discover these “*elephant flow*” in time, we could limit the total BitTorrent traffic volume or optimize the BitTorrent network.



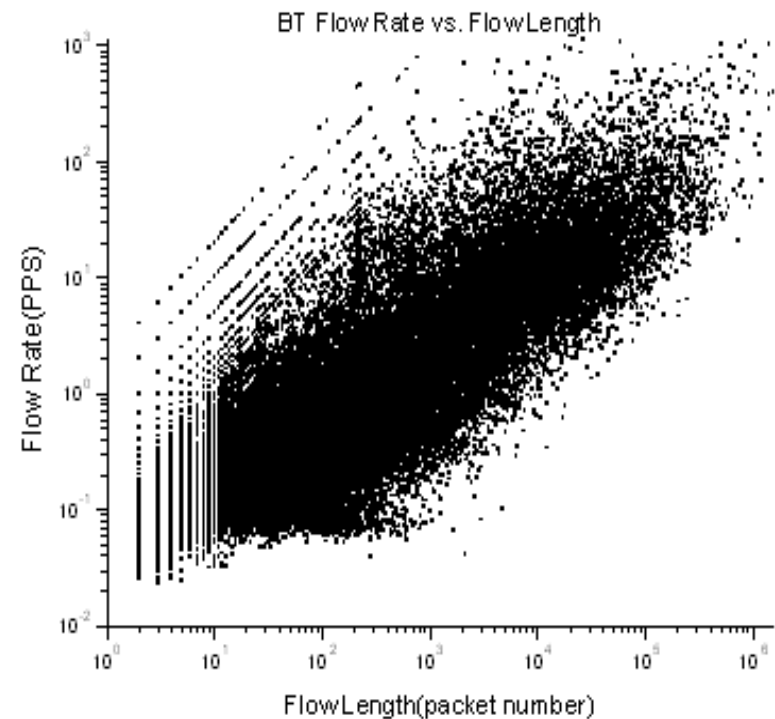
BitTorrent Flow Duration versus Flow Length

# BitTorrent Flow Rate

- 95% of BitTorrent flows' rate are less than 10pps.
- Average rate is 3.3pps, much slower than average TCP flow rate, which is about 6-7pps.
- Long idle periods during the connections.
- Mainly because of BitTorrent's "choke/unchoke" policy.

# Flow Length versus Flow Rate

- In general, there is a strong correlation between flow length and rate.
- Two reasons:
  - User behavior
  - “choke/unchoke” policy



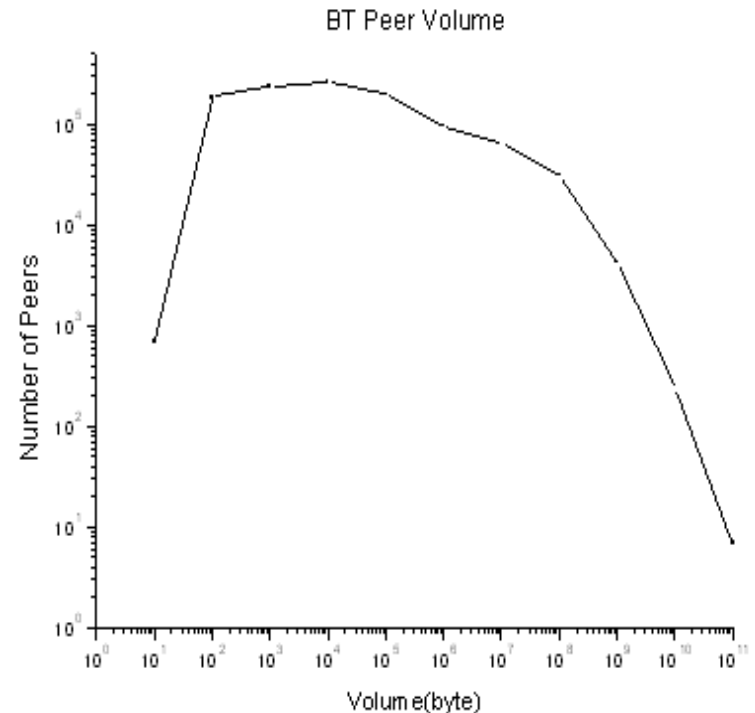
BitTorrent Flow Length versus Flow Rate

# Characteristics of BitTorrent Peer

- Use unique IP address as the identification of each peer.
- More than 1M peers in our trace
- Volume of BitTorrent Peer
- Connection of BitTorrent Peer

# Volume of BitTorrent Peer

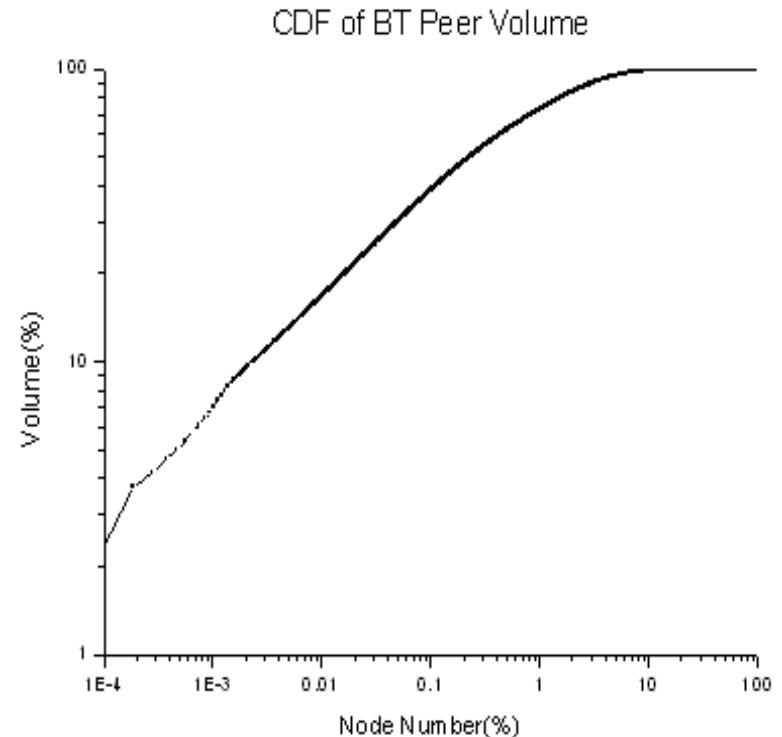
- Volumes are mainly between 1KB and several hundreds of MB.
- Peers whose volume are less than 1KB mainly due to connection attempt.
- Peers whose volume are more than 1GB generally have very large number of connections, and most of them are *seeds*.



Distribution of BitTorrent Peer Volume

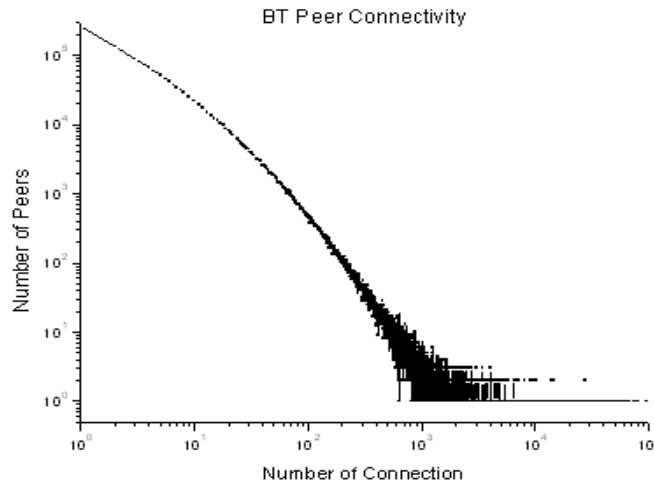
# Volume of BitTorrent Peer (2)

- 1% of total peers contribute to almost 80% BitTorrent traffic.
- 10% of total peers contribute to more than 90% BitTorrent traffic.
- Very skewed distribution

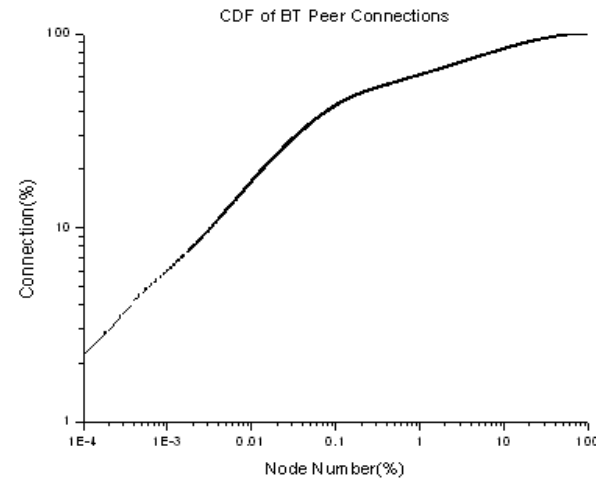


CDF of BitTorrent Peer Volume

# Connections of BitTorrent Peer



Distribution of BitTorrent Peer Connection



CDF of BitTorrent Peer Connection

- Heavy-tailed
- CDF is very similar with that of BitTorrent peer's volume.
- 1% of total peers contribute to almost 80% connections.
- 10% of total peers contribute to more than 90% connections.

# Volume versus Connection

- Within the first 1% rank of BitTorrent peer connection, 75% peers transfer more than 500MB volume.
- Peers transferring more than 1GB data all within the first 1% rank of BitTorrent peer connection.
- 87% peers within the first 1% rank of BitTorrent peer connection also within the first 1% rank of BitTorrent peer volume.
- There are few peers in the BitTorrent network, which have very high connectivity and volume. They are *seeds* or are transferring large files.
- If we can control these peers effectively, we can control the volume of the whole BitTorrent network.

# Conclusion

- Our study indicates firstly that BitTorrent traffic accounts for more than 60% of the total network traffic.
- Secondly, most of the BitTorrent connections are very short and represent a small volume, and both flow length and duration can be approximated by a Weibull distribution.
- Thirdly, due to the “choke/unchoke” policy, the average flow rate is slower than the average TCP flow rate.
- Lastly, very few users contribute to the most significant part of the number of connections and the traffic volume.
- Thus, if we can discover and control the few “*elephant flows*” or “*heavy-hitter users*”, we could control the total BitTorrent traffic.