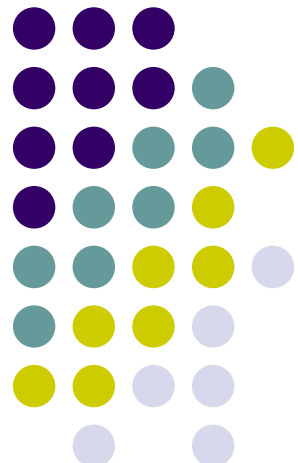
Availability in BitTorrent Systems

Yin-Yeh Tseng

2007/04/26

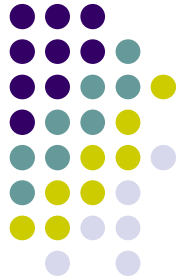


Outline

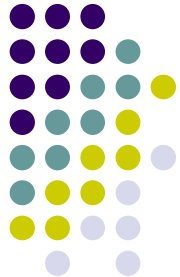
- Introduction
- BitTorrent overview
- Tracker reliability
- Multiple trackers
- Distributed hash table (DHT)
- References

Introduction

- Peer-to-peer networking model has been widely applied in file-sharing applications because of its benefits on scalability, service survivability, and low-cost deployment.
- BitTorrent is a popular P2P-based system designed for the distribution of large files. It constituted about 30% of internet backbone traffic in June 2004.

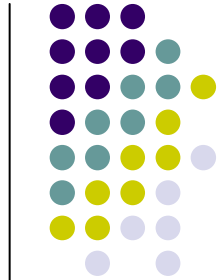


- Although BitTorrent eliminates a single point of congestion as regards data traffic, it continues to have a single point of failure.
- The BitTorrent systems employ a centralized tracker to enable coordination between peers.
- If the tracker fails or is unreachable, the system becomes unavailable to new peers, so they can not obtain the file or contribute resources to the system.

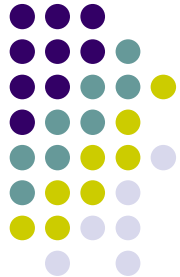


- Two recent trends have emerged to tackle the problem of tracker availability.
  - Multiple trackers
  - Distributed Hash Tables(DHT)

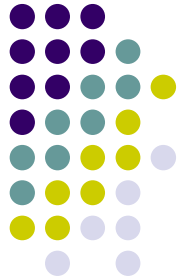
BitTorrent overview



- The tracker maintains the set of active peers, also called the swarm, interested in a specific file.
- A peer joins the swarm by announcing itself to the tracker, which returns a small random subset of peers from the swarm.
- Peers use this subset to connect to other peers to obtain missing pieces of the file.

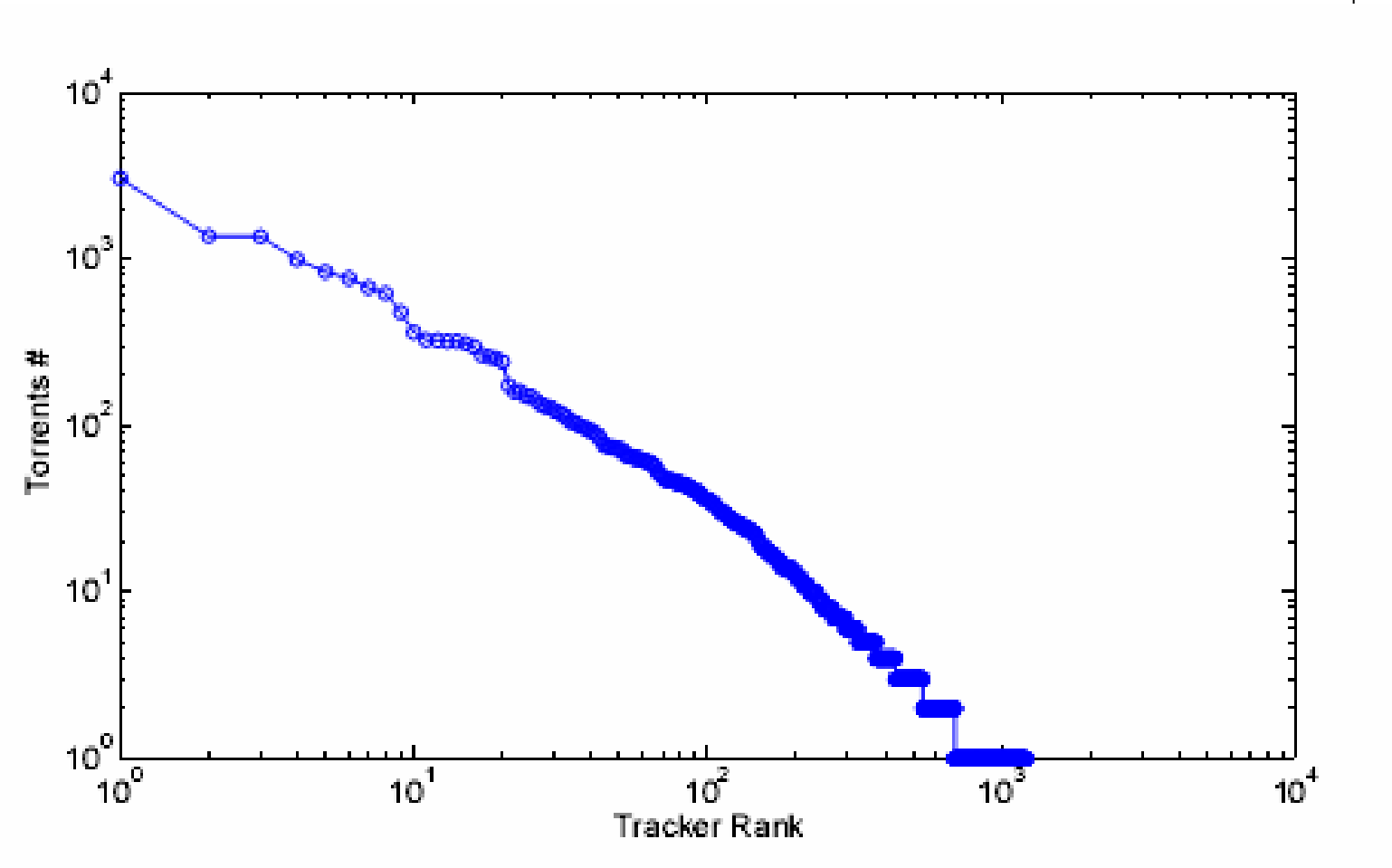


- Each peer periodically reports its information, such as connecting port, upload/download volume to the tracker.
- To share a file through BitTorrent, peers first create a torrent file.
- A torrent contains meta information about the files to be shared in the *info* section and about the tracker in the *announce* section.

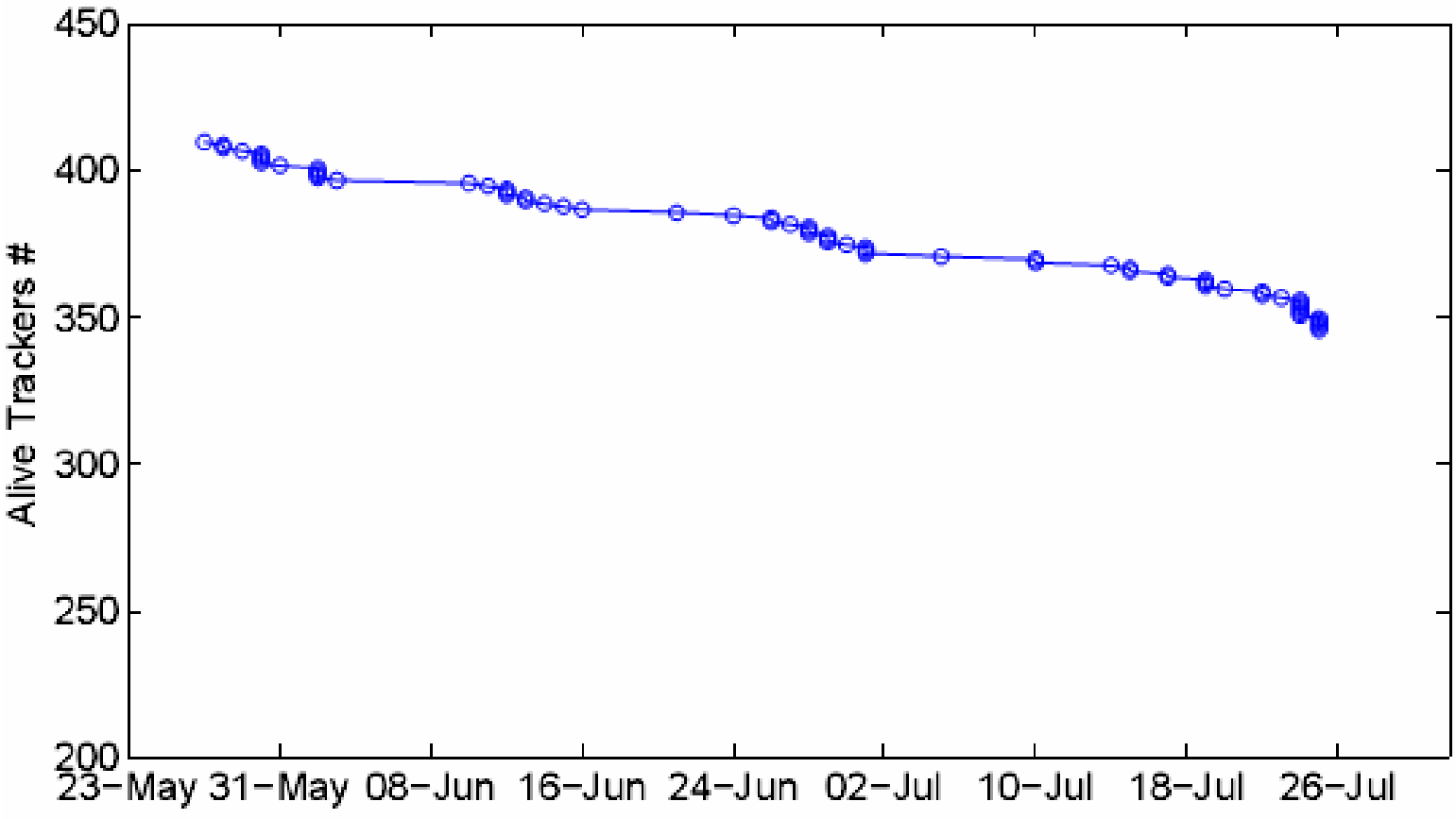


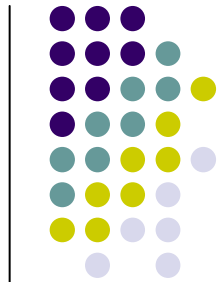
TRACKER RELIABILITY

Popularity of the trackers



Number of live Trackers

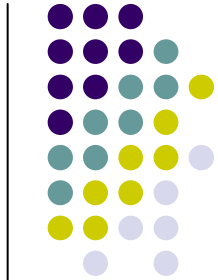




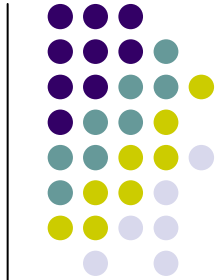
Multiple trackers

Multi-Tracker Feature

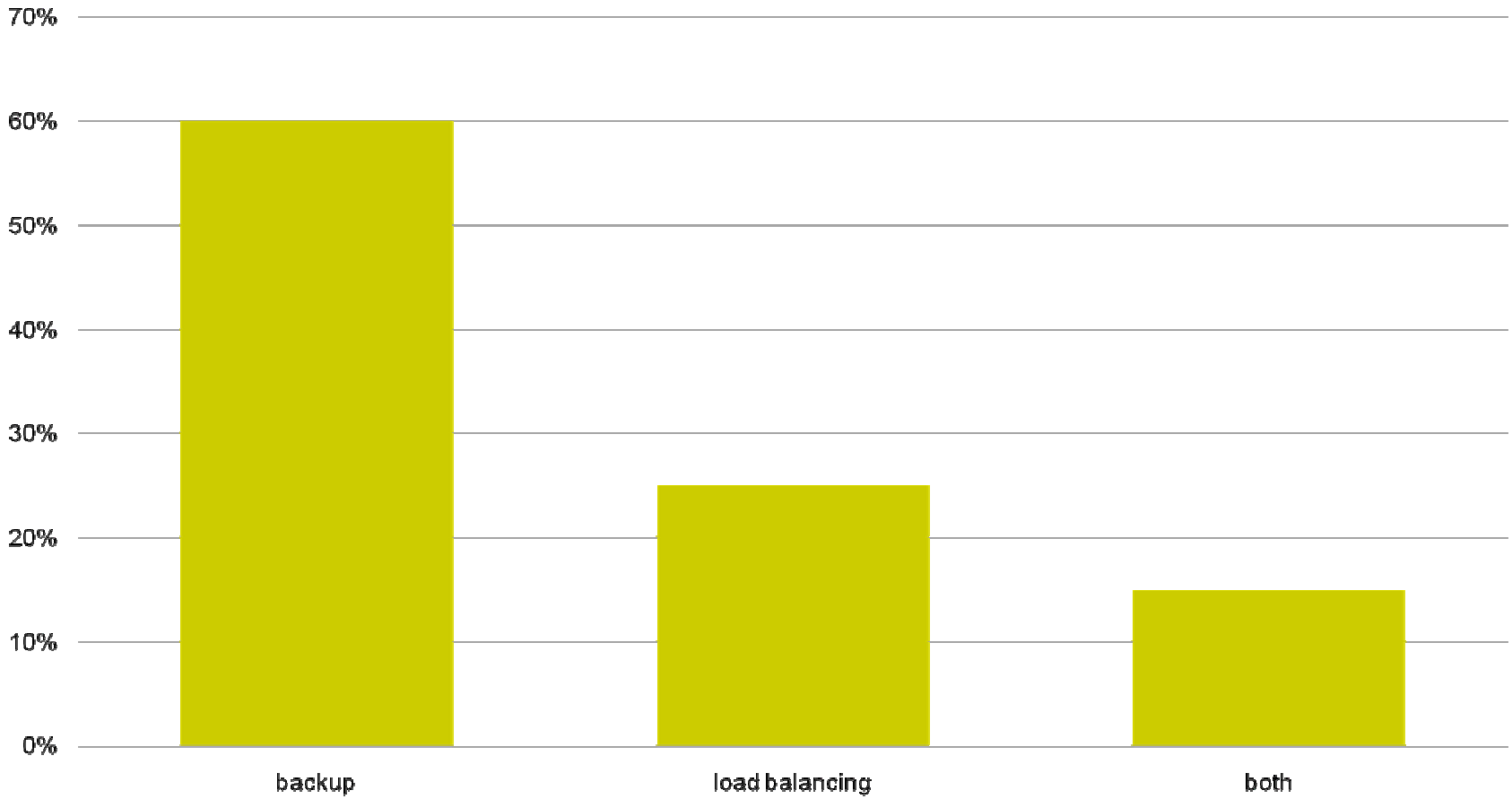
- Multi-Tracker feature allows two or more trackers to take care of the same content.
- In addition to the announce section, a new section, *announce-list*, has been introduced.
- It contains a list of lists of tracker URLs.



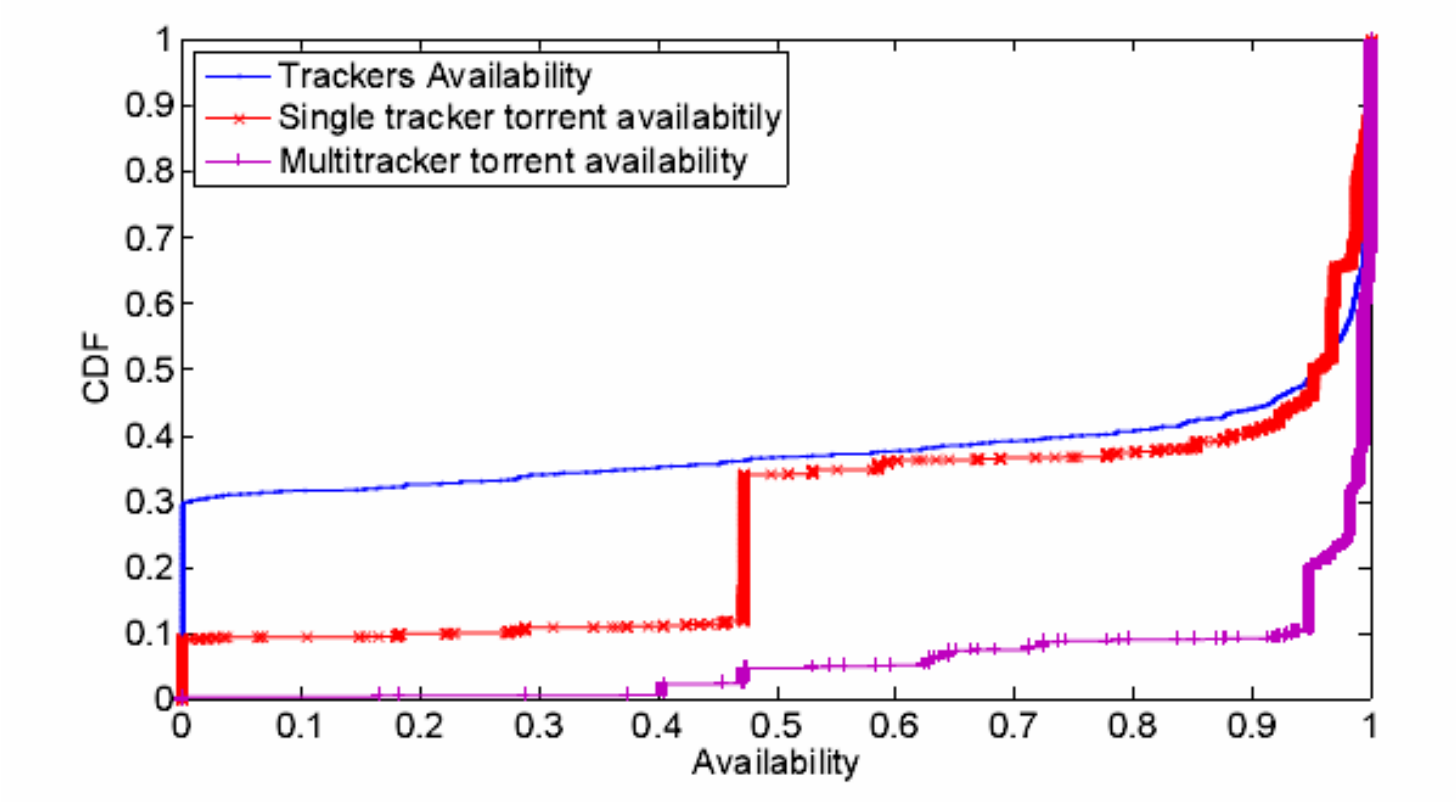
- Trackers in the same list have load-balance purpose: a peer randomly chooses one of them and sends it an announce request.
- All the trackers in the same list exchange information about the peers they know.

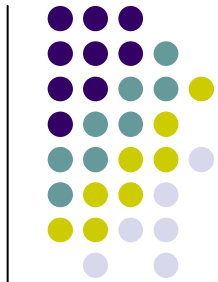


- The different lists of trackers are intended for backup purpose
- Trackers in different lists do not share information


70%
60%
50%
40%
30%
20%
10%
0%
backup
load balancing
both

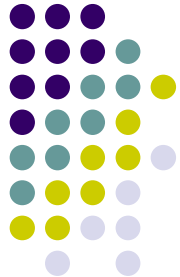
Trackers and Torrents Availabity



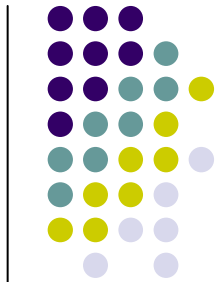


DISTRIBUTED HASH TABLE (DHT)

Overview

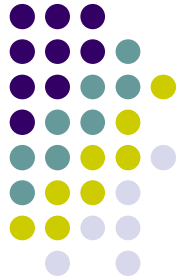


- Each node has a globally unique identifier known as the “node ID.” Node IDs are chosen at random from the same 160-bit space as BitTorrent infohashes.
- Every node maintains a routing table of known good nodes.



- A good node is a node has responded to one of our queries or sent us a query within the last 15 minutes.
- The routing table is updated from time to time . Peers in a swarm can exchange DHT related information.

Torrent File Extensions

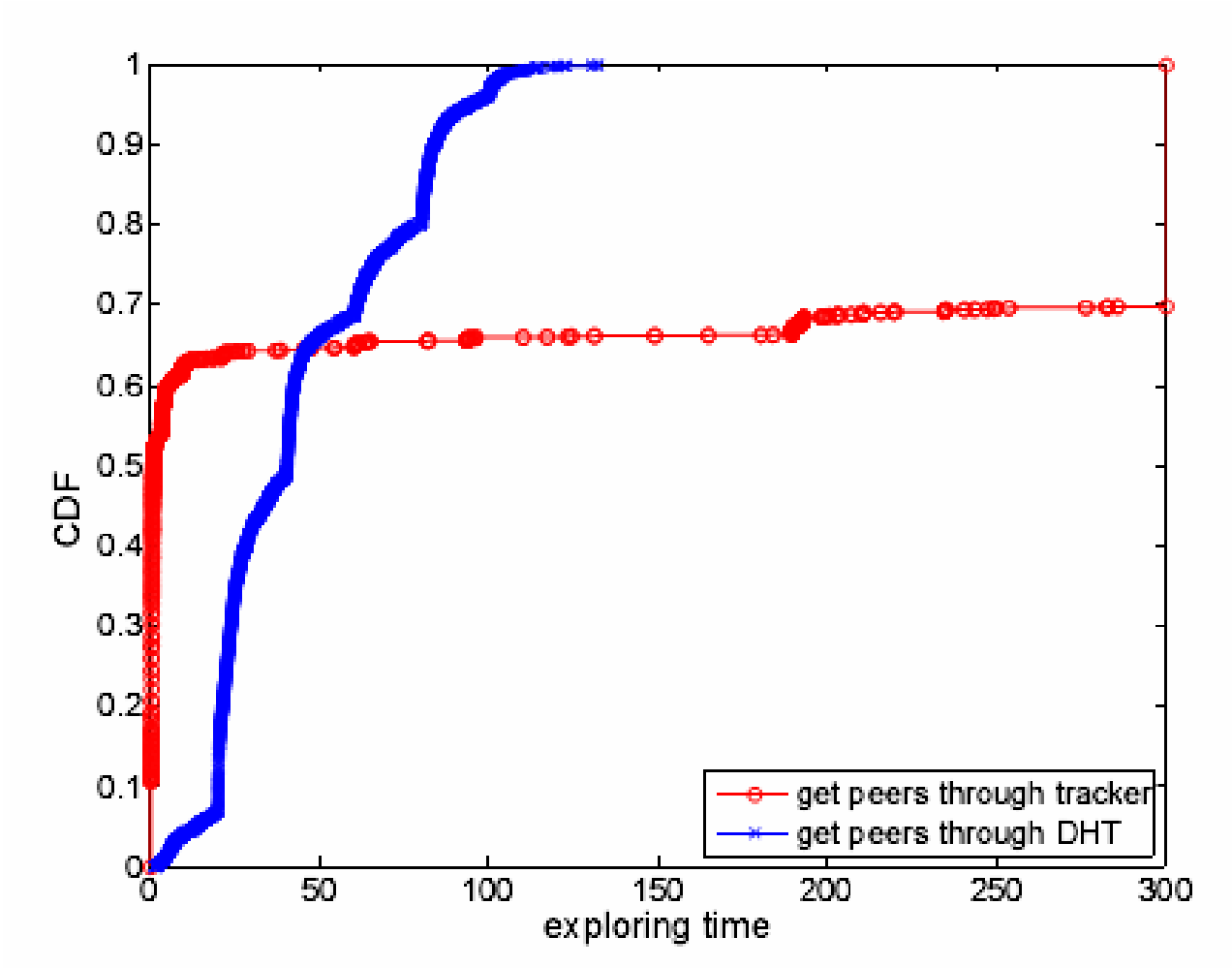


- A trackerless torrent has a *nodes* key. This key should be set to the K closest nodes in the torrent generating client's routing table.
- These DHT nodes act as bootstrap nodes.

DHT Operation

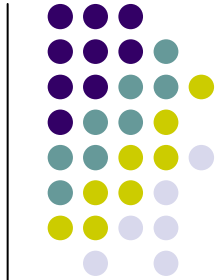
- To create a new torrent, the node insert some good DHT nodes from his routing table .
- These nodes are used in order to initialize the client routing table.
- In order to download the content, the BitTorrent client can send requests to the set of DHT nodes in its routing table closest to the infohash.

The cumulative distribution of the time needed to fine the first valid peer in a swarm



Conclusion

- the problem of tracker availability
- two recent trends to tackle the problem of tracker availability:
  - Multiple trackers
  - Distributed Hash Tables(DHT)



- DHT improves information availability, but induces a higher response latency.
- Current trend of combining multiple trackers and DHT can provide high information availability with low information response latency.

References

- G. Neglia, G.Reina, H. Zhang, D.Towsley, A. Venkataramani, and J.Danaher, “Availability in BitTorrent Systems,” in *Proc. of IEEE INFOCOM 2007*
- *“DHT protocol specification,”* http://www.bittorrent.org/Draft_DHT_protocol.html
- “Multitracker description,” http://wiki.depthstrike.com/index.php/P2P:Protocol:Specifications:Multitracker