Peer-to-Peer and GRID Computing, 2G1526 Lecture 08

Jim Dowling and Seif Haridi

SICS, KTH

Goal

- To Understand:
  - Motivation for Super-Peer P2P Architectures
  - Super-Peer Topologies and Super-Peer Promotion
  - Case Study: Skype

Basic Idea

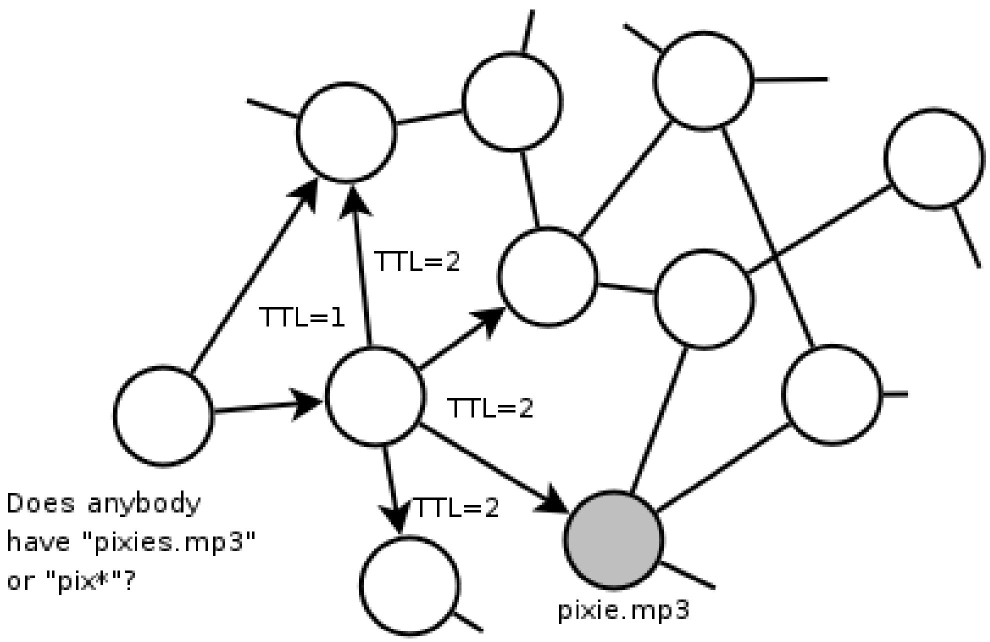
- Real-World P2P systems for the open Internet are heterogeneous
  - Peer resources (Bandwidth, CPU, Memory)
  - Peer session-time
- Use Peers with better “characteristics” to provide services to other peers in the system

Motivation: Search in P2P

- Centralised (Napster)
- Flooding (Gnutella)
  - Essentially a breadth-first search using TTLs
- Distributed Hash Tables

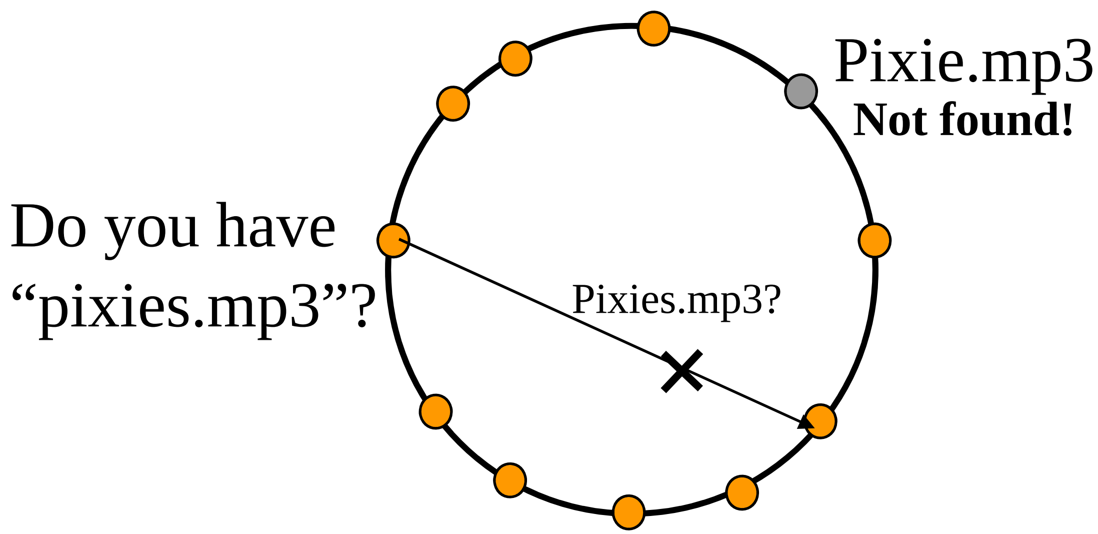
Searching in Gnutella

- Queries are flooded to neighbours, have a TTL, and are forwarded only once
  - Can we search using fewer packets?



Searching in DHTs

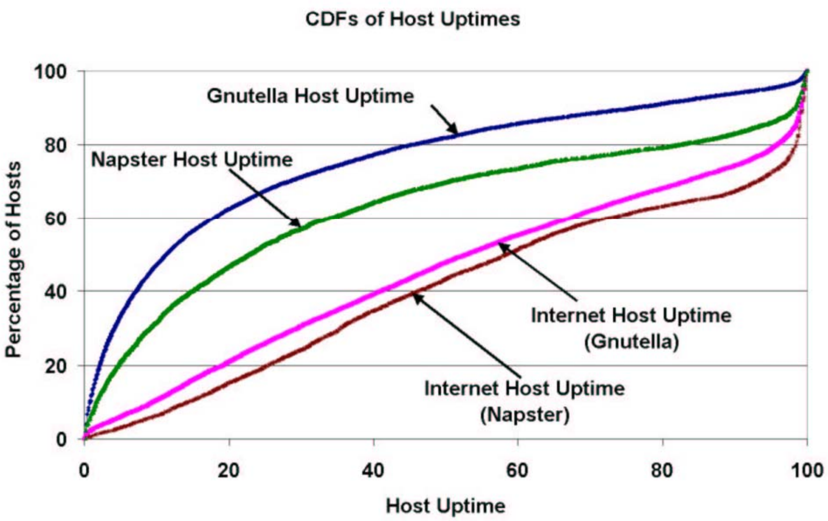
- Need to know the exact filename
  - Keys (filenames) map to node-ids.
    - Change in file name => search at different node
    - No wildcard matching: cannot ask for file “pix*”

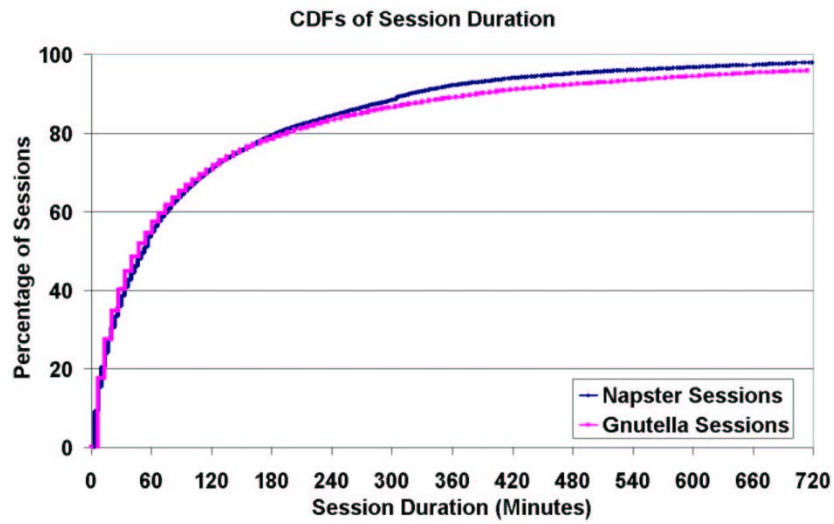


But….all peers are not equal

- Peers have heterogeneity with respect to:
  - Available Bandwidth
  - Average Session Time
  - Open IP address (vs. NAT-bound)
  - Latency
  - CPU/Memory

Peer Session Time Distribution

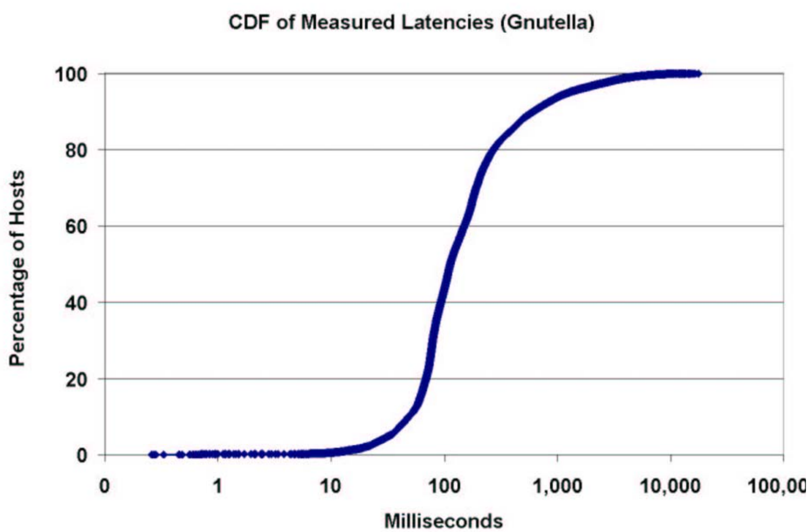


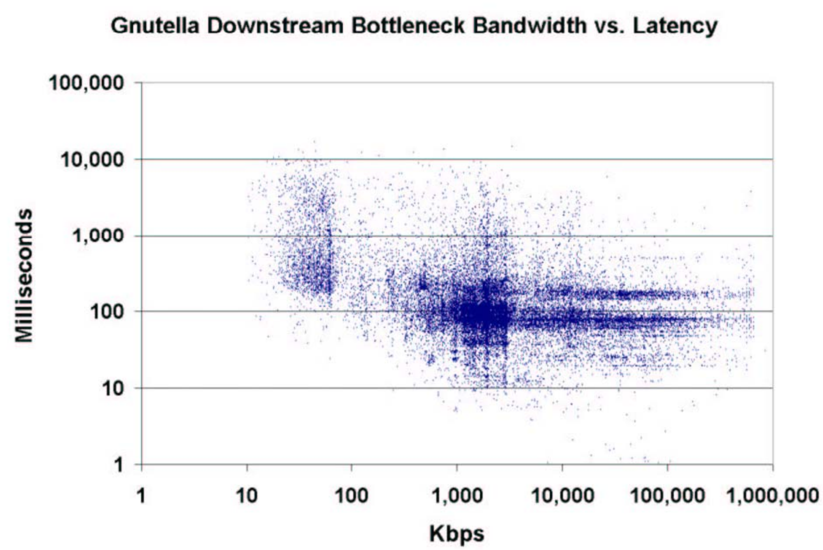


- Median session duration is 60 minutes for Napster and Gnutella

Saroiu et.al., “A Measurement Study of Peer-to-Peer File Sharing Systems”, MMCN, 2002

Peer Latency Distribution

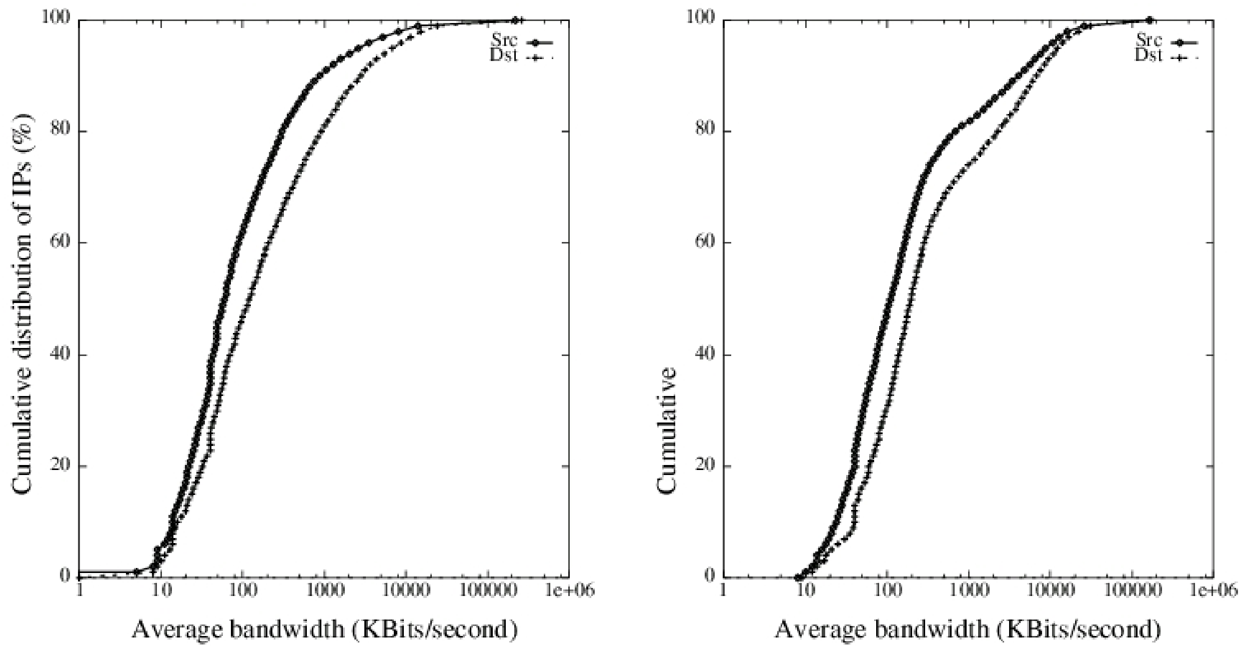




- 20% peers have a latency of at most 70ms and 20% have a latency of at least 280ms

Saroiu et.al., “A Measurement Study of Peer-to-Peer File Sharing Systems”, MMCN, 2002

Peer Bandwidth Distribution



- FastTrack: 33% IP addresses have mean downstream b/w 56Kbps or less; 50% have mean upstream b/w 56Kbps or less
- Direct Connect: 20% IP addresses have mean downstream b/w 56Kbps or less; 33% have mean upstream b/w 56Kbps or less

Sen and Wang, Analyzing peer-to-peer traffic across large networks, IEEE/ACM TON, 2004

Super-Peer Definition

- A super-peer is a node in a peer-to-peer network that operates both as a server to a set of clients, and as an equal in a network of super-peers. [Yang/Molina '03]

- Super-peers have high utility relative to non super-peers, where higher utility peers are "better" at providing super-peer service(s).

  Spare Bandwidth/CPU; Open IP Address; etc

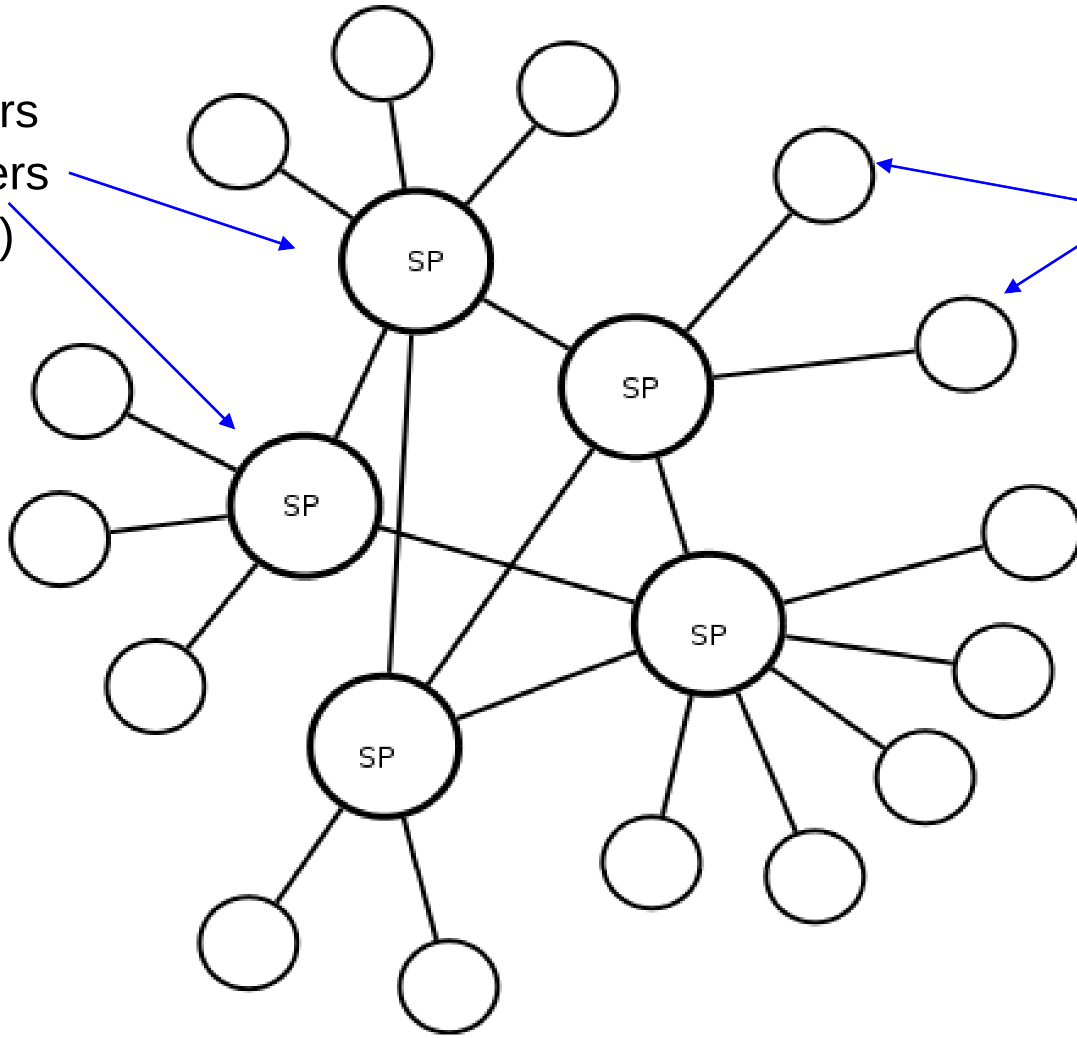
  - Measured peer utility can be used to rank peers to enable the best peers to be promoted to super-peers.

Goals of Super-Peer Networks

- Exploit heterogeneity in P2P Networks by using higher utility peers to provide services
- Super-Peers provide redundant instances of System Services giving a P2P system:
  - Scalability
  - Load balancing
  - Fail-over
  - Robust to node failures, message loss

Super-Peer Architecture

Super-Peers
(both servers
and clients)



Ordinary Peers
(are clients)

Super-Peer Design Issues

- What services do super-peers provide?
- Ordinary peer to super-peer connections
  - Redundancy, Fail-Over, Performance, Fairness
- (Intra-)Super-peer overlay network topology
  - Random, DHT, Gradient
- Super-peer promotion
  - Which/how many peers should be super-peers?

What Services do Super-Peer Systems provide?

- File Naming/Retrieval
  - Fast-Track
  - Kazaa
  - E-Donkey
- Voice Over IP (VoIP)
  - Skype
- Anything you want!
  - Sun’s JXTA framework for building super-peers

Ordinary Peer to Super-Peer (SP) Connections

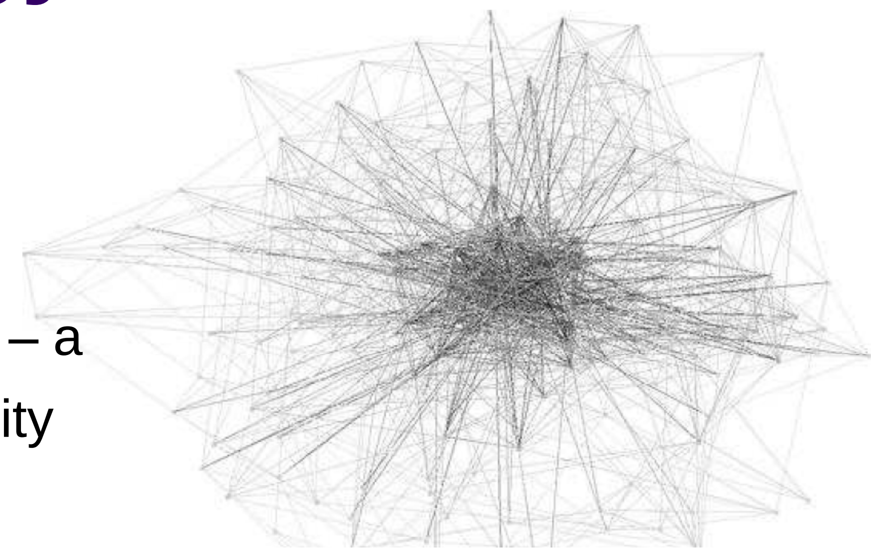
- 1 active SP connection per ordinary peer
  - Suitable for TCP traffic
  - Failover SP connections for fail-over
- >1 active SP connection per ordinary peer
  - More suitable for UDP traffic; possible with TCP
    - Requires session management for P2P routing
- Fairness allocating Ordinary Peers to SPs
  - Don’t overuse the SP’s resources

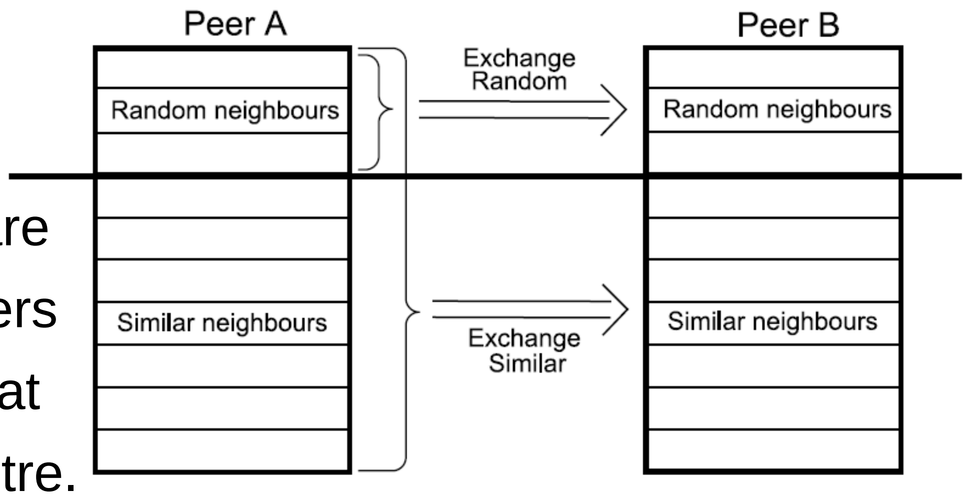
Super-Peer Connection Topology

- The type of intra-SP topology affects the type of distributed services they provide
- Random Network
  - Flooding, Gossiping and Random Walk
- DHT (don't know of any systems that do this)
- Gradient Topology
  - Gradient search, Gossiping and Random Walk

Gradient Topology

- No distinction between peer and super-peers
- Peers have 2 sets of neighbours – a set of neighbours with similar utility and a set of random neighbours
- Peers gossip to maintain their random and similarity sets
- Peers with highest utility levels are clustered in the centre, while peers with decreasing utility are found at increasing distance from the centre.

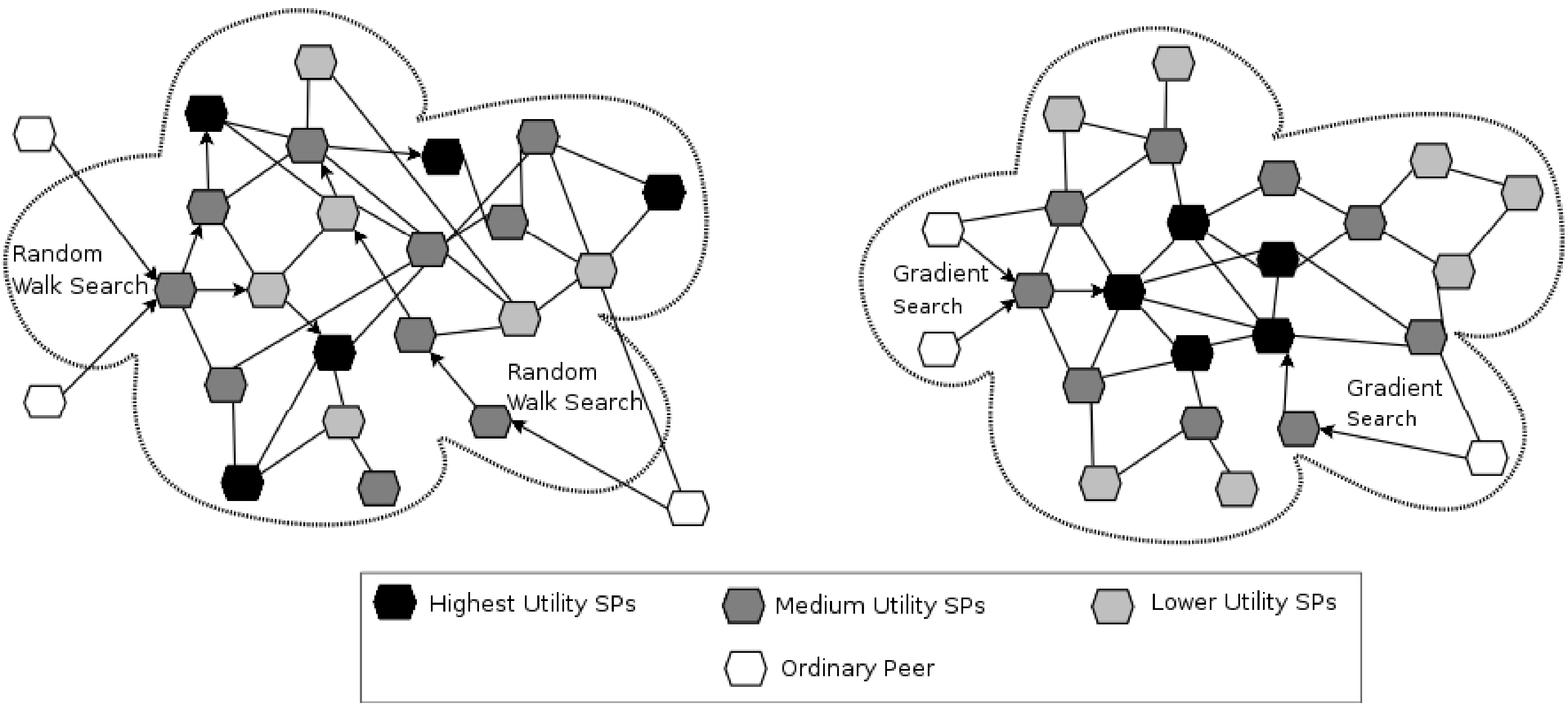




Search in Random and Gradient SP Topologies

Random Walk in a Random Topology

Gradient Ascent Search in a Gradient Topology



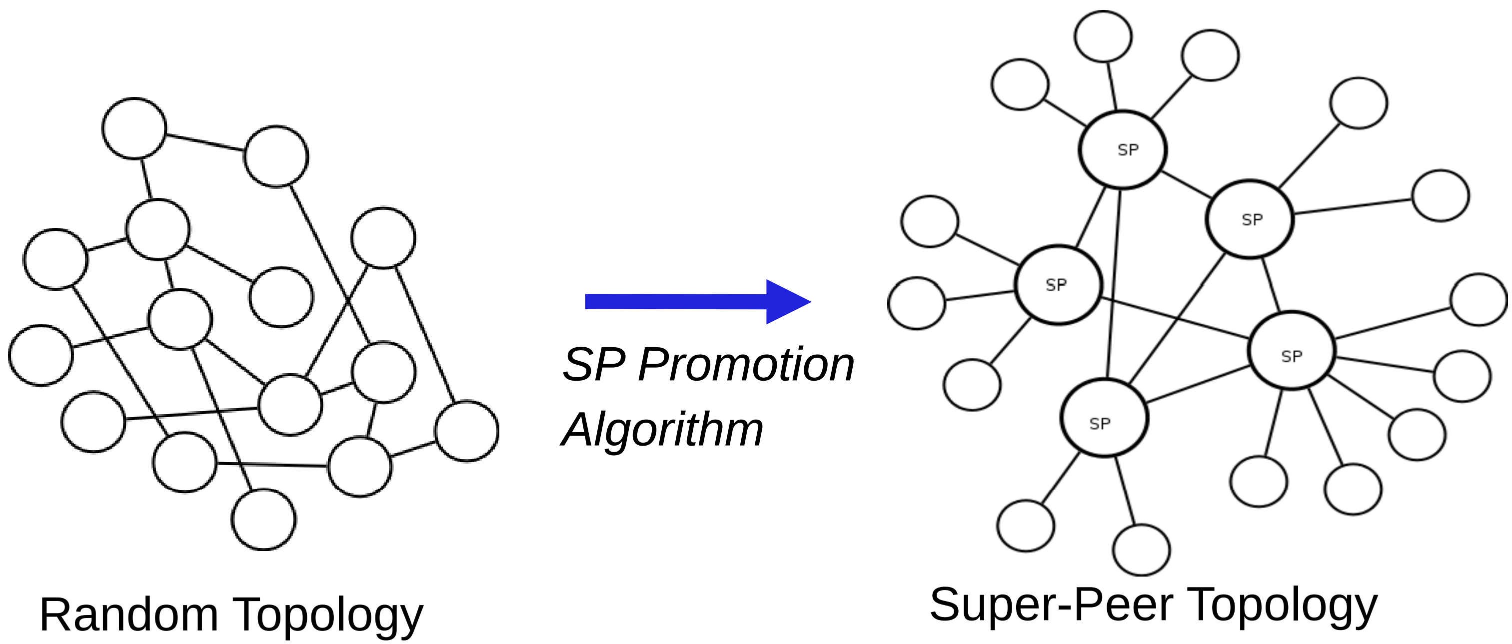
Super-Peer Promotion

- Issues:
  - What level of “utility” is required for a peer to become a super-peer (it is service-dependent)?
  - How many clients should a super-peer support?

- Options:
  1. Promote peers that meet local utility requirements
  2. Promote the top 'X' percent of peers with highest utility (requires global knowledge of peer utility distrib.)

Super-Peer Promotion Algorithm Properties

- Distributed Promotion > Centralised Promotion
- Session-start or Runtime > Bootstrap Time
- Fairness to Super-Peers vs. System Availability



Peer Heterogeneity and Promotion

- Select peers to be super-peers with the highest utility:
  - Session Time (Average/Median/Expected), Bandwidth, Latency, Open IP Address or NAT-bound

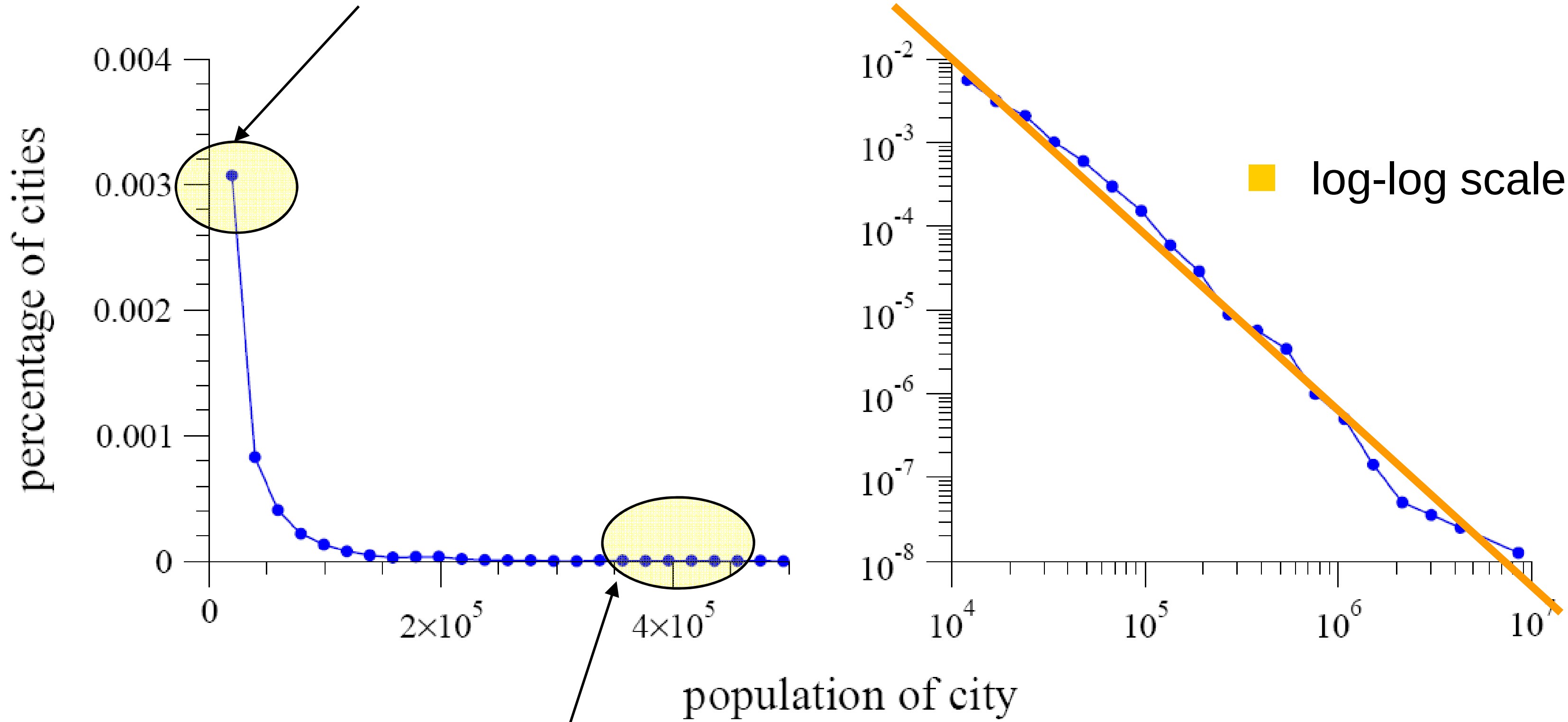
- If we have a model of peer heterogeneity, it helps super-peer promotion algorithms

Peer Heterogeneity and Power Laws

- What type of heterogeneity is found in peers over different characteristics, such as bandwidth, session-time, etc?
  - Depends on the system, but....
- Measurements of Gnutella showed all sorts of power law like relationships

Power Law Example

Lots of cities with a small population



Small number of cities with high population

Power Laws

- A power law distribution satisfies

$$\Pr\left(X \geq x\right) \approx Cx^{-\alpha}$$

normalization constant
(probabilities over all *x* must sum to 1)

power law exponent α

- Log-Log cumulative distribution function (CDF) is exactly linear:

$$\ln \Pr\left(X \geq x\right) \approx c - \alpha \ln x$$

Power Laws and Scale Invariance – Scale Free

- A power law is any polynomial relationship that exhibits the property of scale invariance.

- A power law looks the same no mater what range we look at it on (1 to 10 or 300 to 7000)
  - Only true of a power-law distribution!

Zipf's Law

- Zipf's law usually refers to the 'size' of an occurrence of an event relative to its rank r.
- The expected value of the rth ranked variable is E(Xr) given by $$E(Xr) \approx C.r^{-(b)}$$

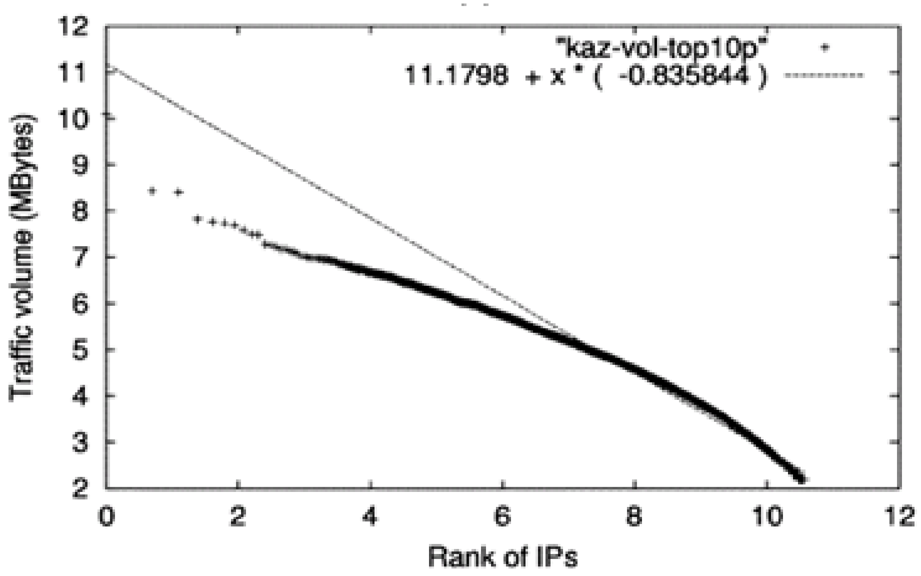
  with b close to unity, where C is a normalization constant and b is the slope of the graph

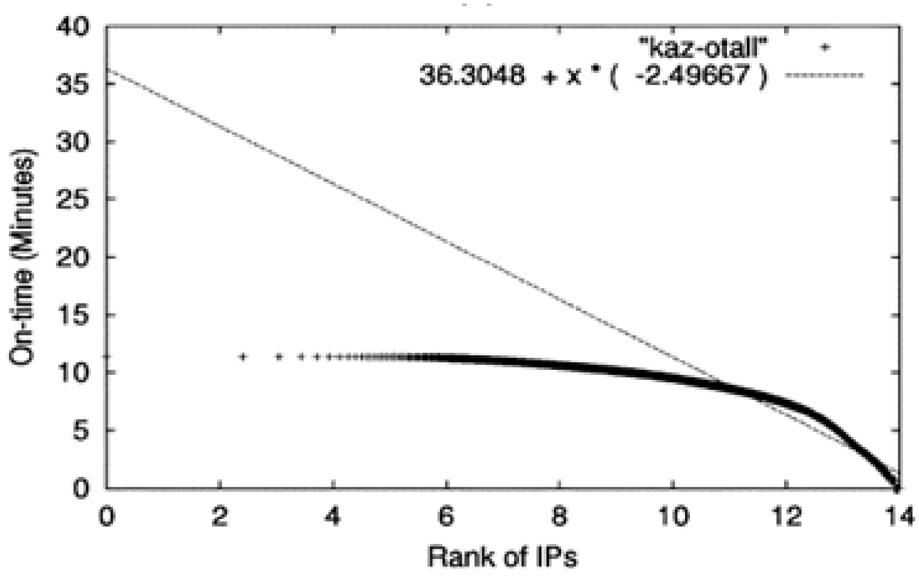
Pareto Distribution

- In 1897, Pareto asked how many people in northern Italy have an income greater than x?
- This is expressed as a cumulative distribution (the probability that a person earns X or more):

$$P(X \geq x) \approx C.x^{-(k)}$$

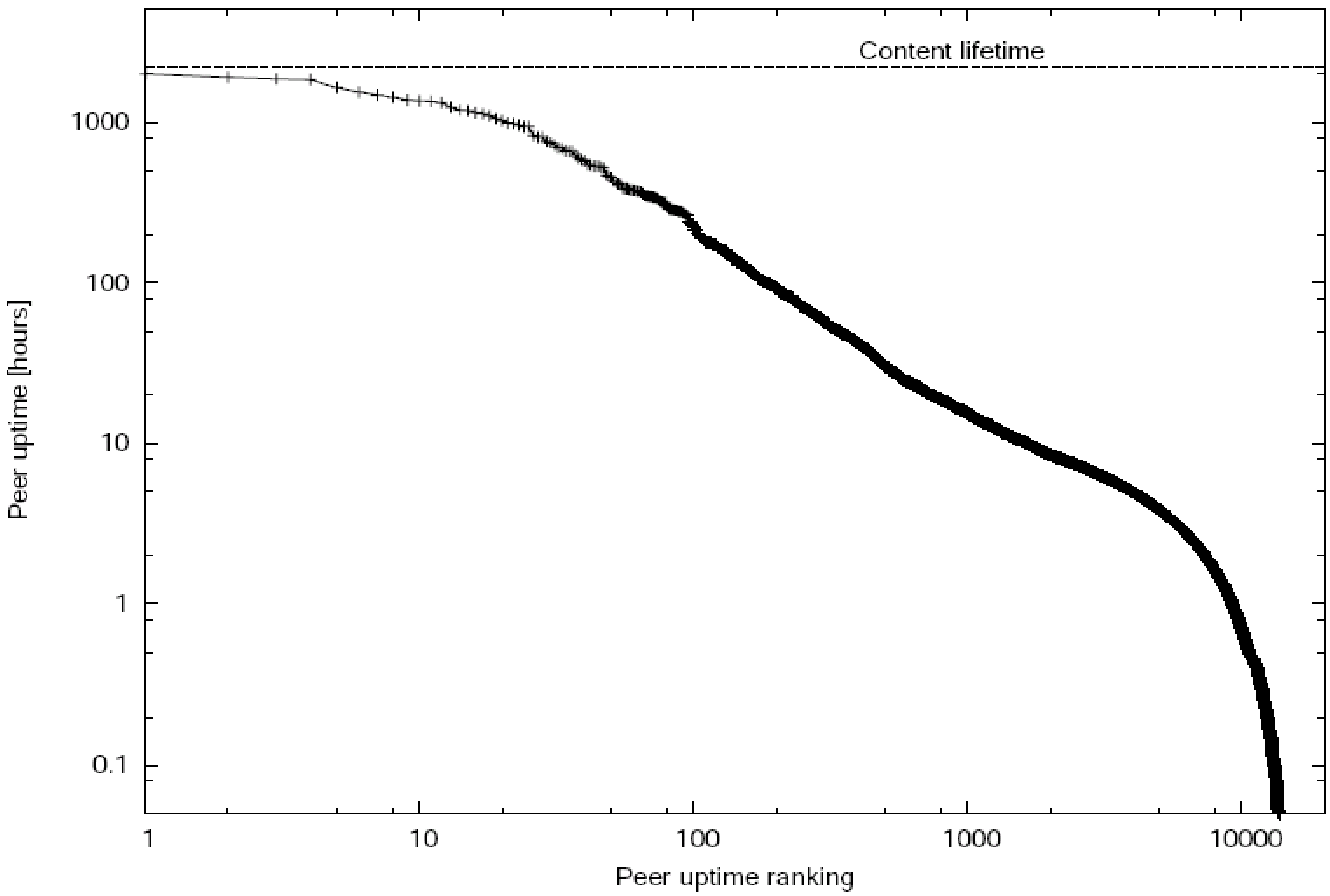
Fast-track: heavy-tailed, not Zipf, for n/w traffic and Session Time





Sen and Wang, Analyzing peer-to-peer traffic across large networks, IEEE/ACM TON, 2004

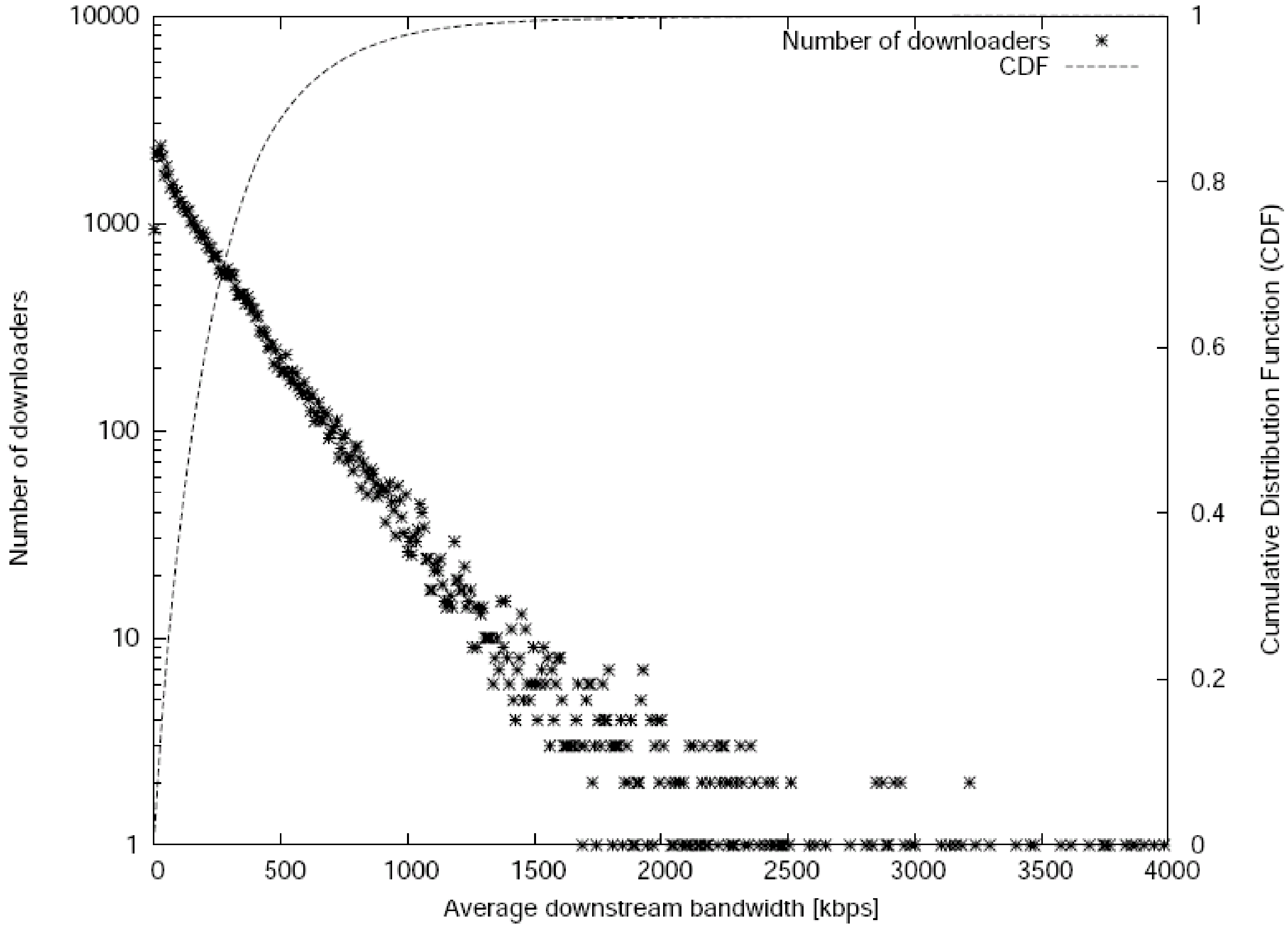
Bittorrent, also heavy-tailed, for Session Time Distribution



Log-Log plot of the uptime distribution of the 53,833 peers downloading ”Beyond Good and Evil”

Poulse et al., “The Bittorrent P2P File-sharing System: Measurements and Analysis”, IPTPS '06

Bittorrent Download Speed Distribution



Plot of the download speeds of 54,845 peers over 2 week period

Poulse et al., “The Bittorrent P2P File-sharing System: Measurements and Analysis”, IPTPS '06

Models of P2P Heterogeneity

- Distributions of peer utility levels help P2P systems set the minimal level required for peer promotion to a super-peer
- Peer Promotion algorithms can update these models at runtime
  - Centralised or Gossiping

Super-Peer Promotion using Local Knowledge

- Promotion algorithm makes a local decision about promotion using:
  - Measurements of the peer's local utility levels for Bandwidth, CPU, Memory
  - A Model of Peer Utility Distributions
  - Tests of the peer's local network interfaces for NAT-bound or Open IP Address

Super-Peer Election using Global Knowledge

- Too expensive solutions
  - Centralized Servers
  - Classic Election Algorithms

- Gossiping / Aggregation
  - Used by Gradient Topology

Utility Estimation in the Gradient Topology

- Aggregation gossiping algorithm runs over the random neighbours used to estimate:
  - Max Utility, Number of Peers in the System, Cumulative Histogram of Peer Utility Values
  - Peers use the local Cumulative Histogram of Peer Utility Values to estimate the utility level for the top ‘X’% of peers in the system.

Case Study: Skype

- Skype is a P2P VoIP telephony System
- Skype is based on a Super-Peer Architecture
  - Call-Establishment, NAT traversal
- Skype encrypts TCP/UDP Payloads
  - Analysis of Skype is limited

Logging In with Skype

- Contact Bootstrap Servers
  - Try different protocols in order:

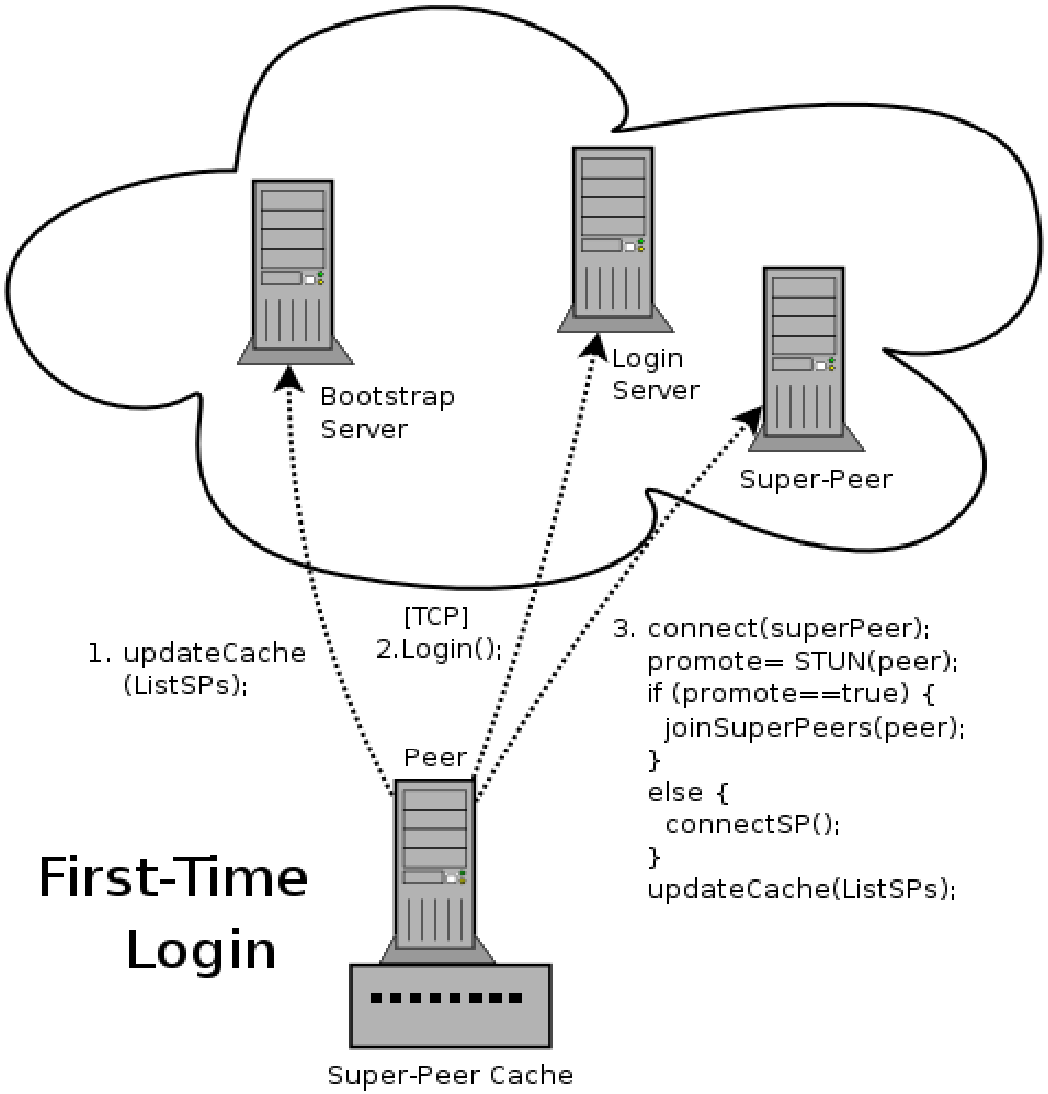
    UDP, TCP, HTTP Proxy, HTTPS Proxy
  - Fill Super-Peer Cache
    - Super-Peer Cache is stored in the <HostCache> entry

      in 'shared.xml'

      Linux: ~/.Skype/shared.xml
      Windows: C:\Documents and Settings\<user>\Application Data\Skype\shared.xml
- Login to Centralised Login Server using TCP

Baset et al., “An Analysis of the Skype Peer-to-Peer Internet Telephony Protocol”

Skype First-Time Login



Super-Peer Promotion in Skype

- Simple Traversal of UDP through NATs (STUN) Protocol run between a Peer and either a Super-Peer or Bootstrap Server:
  - STUN determines if peer has an open IP Address. Available Bandwidth at the peer is measured.

```
if ( (STUN_SP(NetworkIF) = OPEN_IP)
   && (bandwidth_available(NetworkIF) > minBW)) {
    promote_to_superpeer
}
```

  - Run every time a peer starts a Skype session

Guha et al., “An Experimental Study of the Skype Peer-to-Peer VoIP System”, Workshop on P2P '06

Question

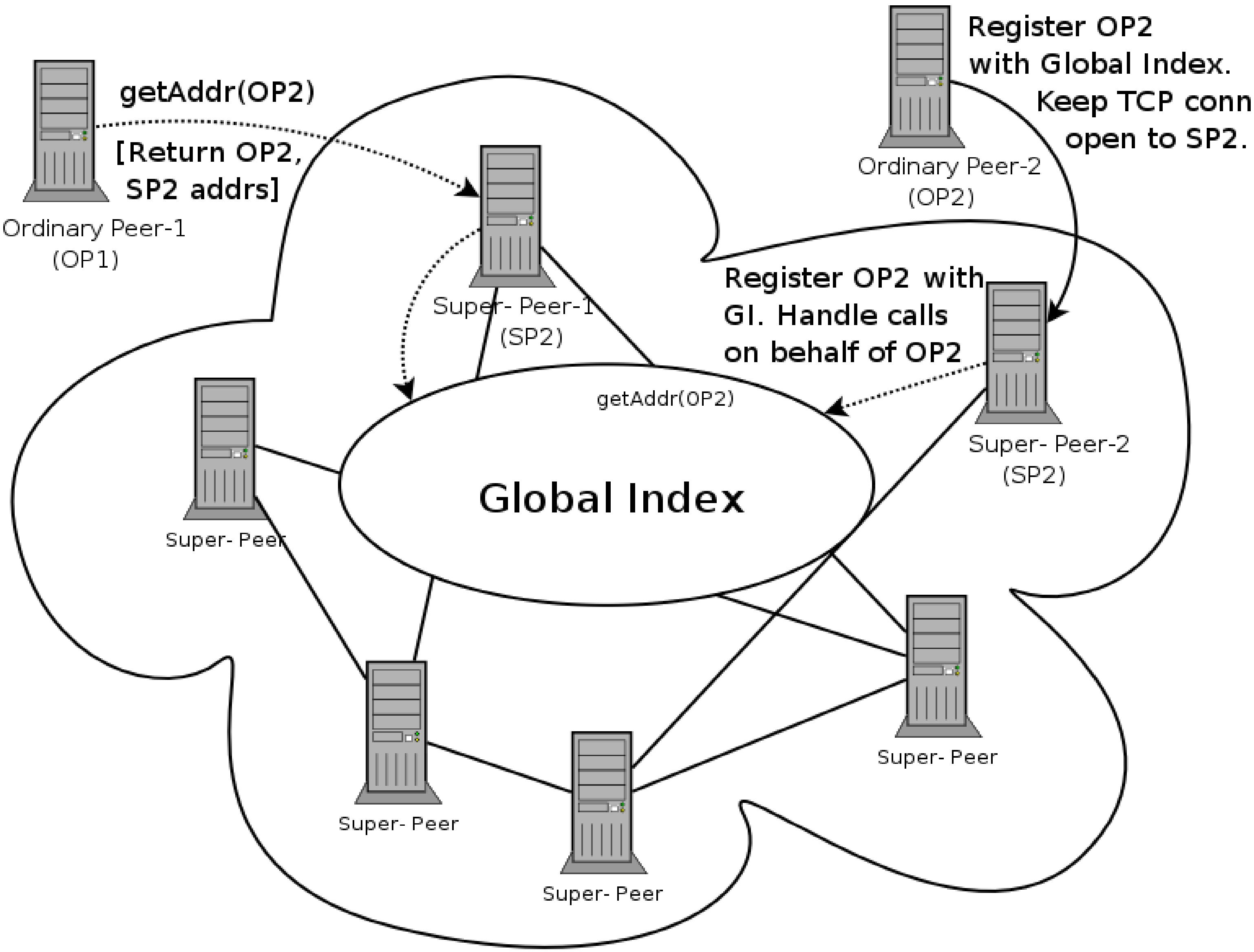
- How do Skype peers discover users on the Super-Peer Overlay?

Decentralized User Directory

- Super-peers manage a Global Index of users and hosts
- "The Global Index (GI) technology is a multi-tiered network where supernodes communicate in such a way that every node in the network has full knowledge of all available users and resources with minimal latency."[*]
  - Finds users who have logged in during the last 72 hours

[*] http://www.skype.com/products/explained.html

Skype Global Index



Super-Peer Overlay Topology

- Only speculative information on how super-peers are connected and how the Global Index is implemented
- Question: could the GI be
  - a decentralised protocol (DHT/Gradient)?
  - a hierarchical protocol (DNS)?
  - centralised?

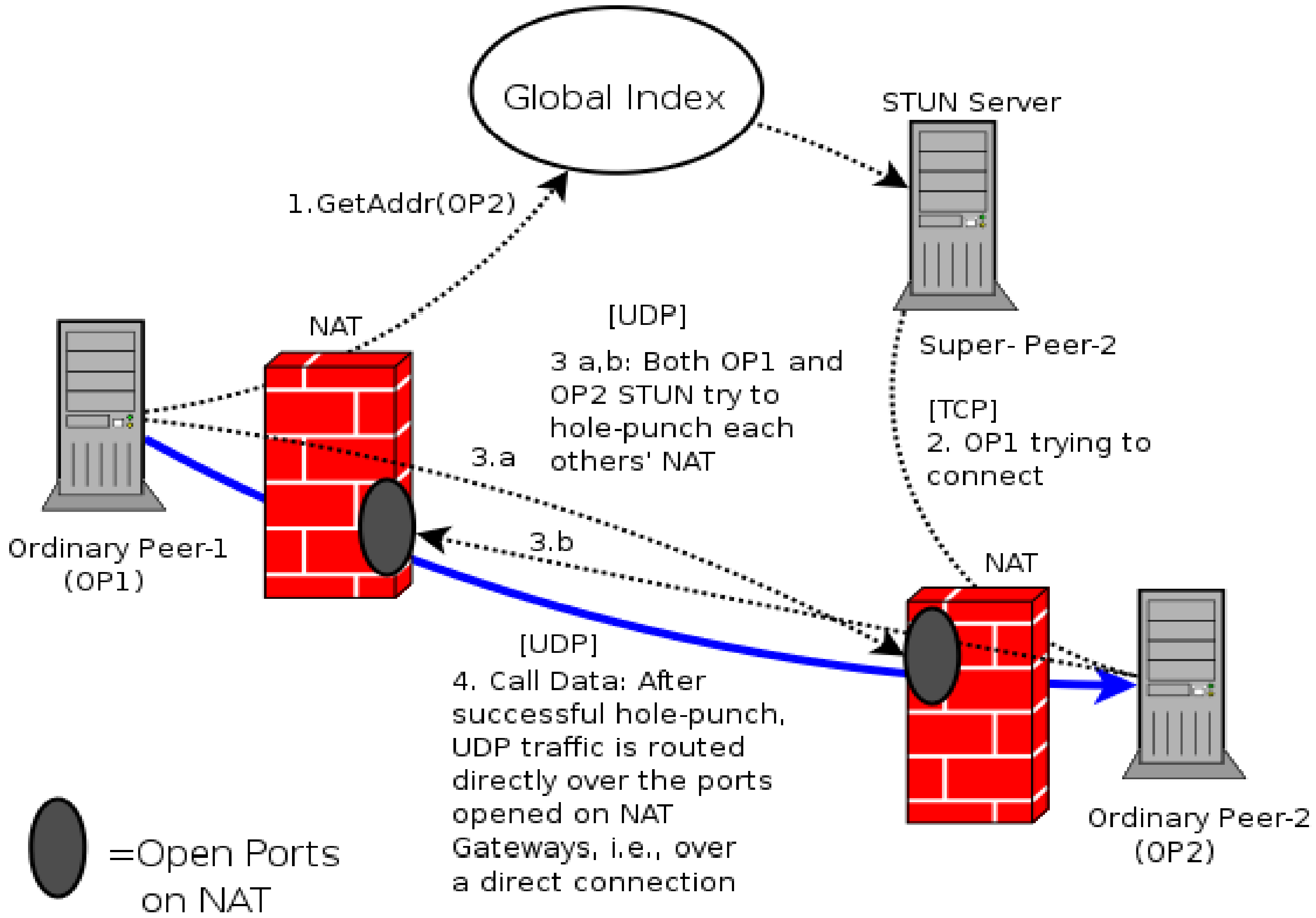
Call Setup in Skype

- If the destination peer has an open IP addr, call signalling is generally carried over a direct TCP connection
- How do peers accept incoming connections if they are behind a NAT Gateway?
  - Using an always-open TCP connection to a super-peer. The super-peer accepts the connection on behalf of the NAT'd peer, and the NAT'd peer negotiates how to setup the connection.

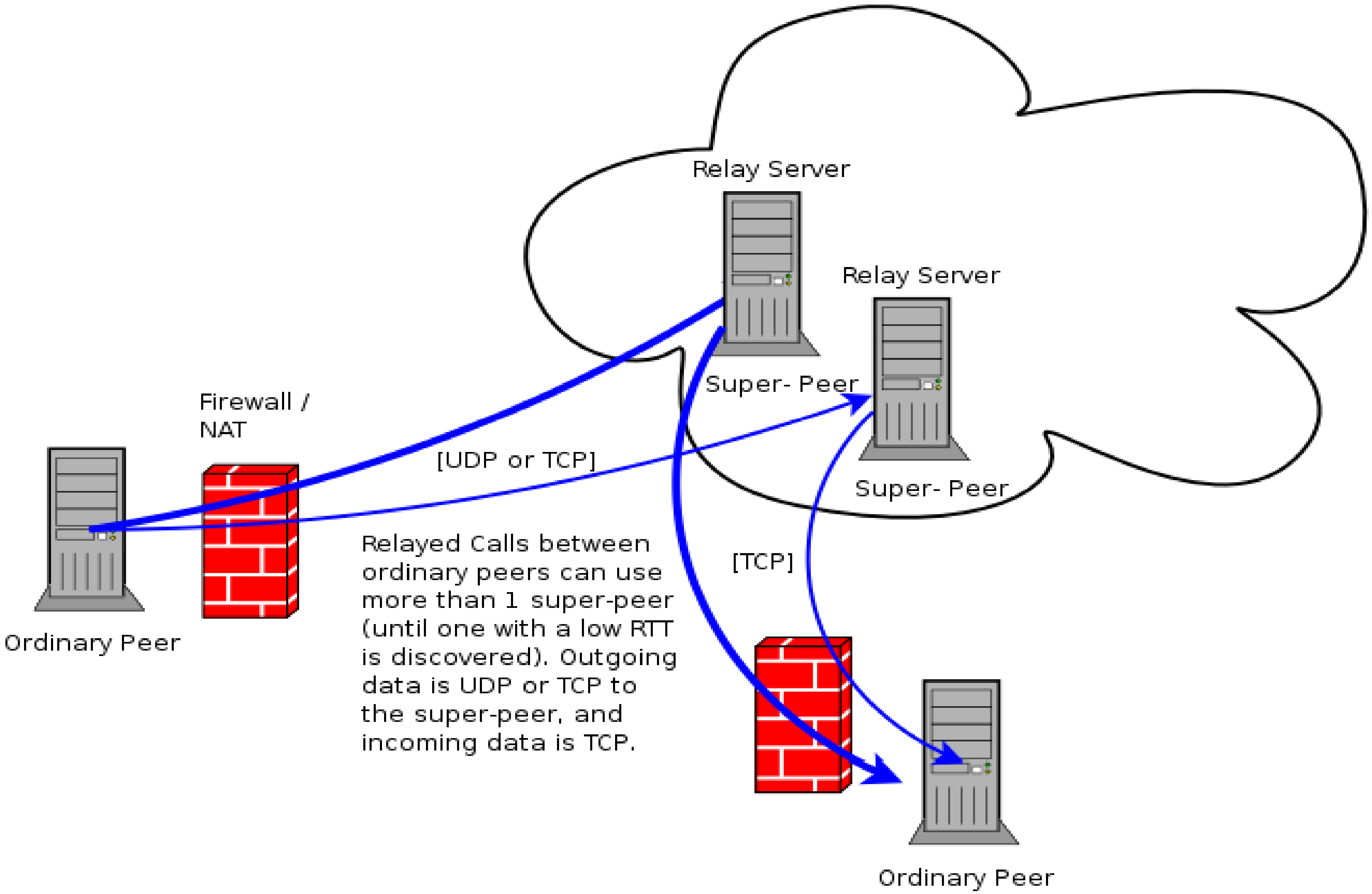
NAT Traversal in Skype

- There are 2 general techniques for routing UDP traffic to peers behind NAT Gateways
  - Simple Traversal of UDP through NATs (STUN)
  - Relaying of UDP/TCP traffic, e.g., Traversal Using Relay NAT (TURN)
- In Skype, super-peers provide both STUN and Relaying Services to help call establishment
  - Try STUN. If STUN fails, try Relaying.
    - Approx. 9 % of Skype traffic is relayed [Guha et. al]

STUN in Skype



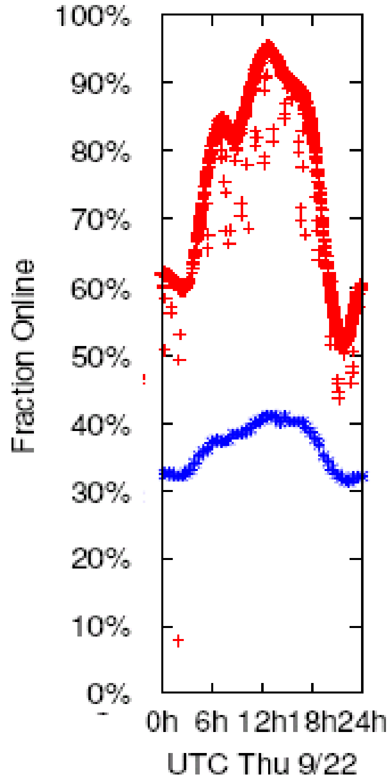
Relaying in Skype



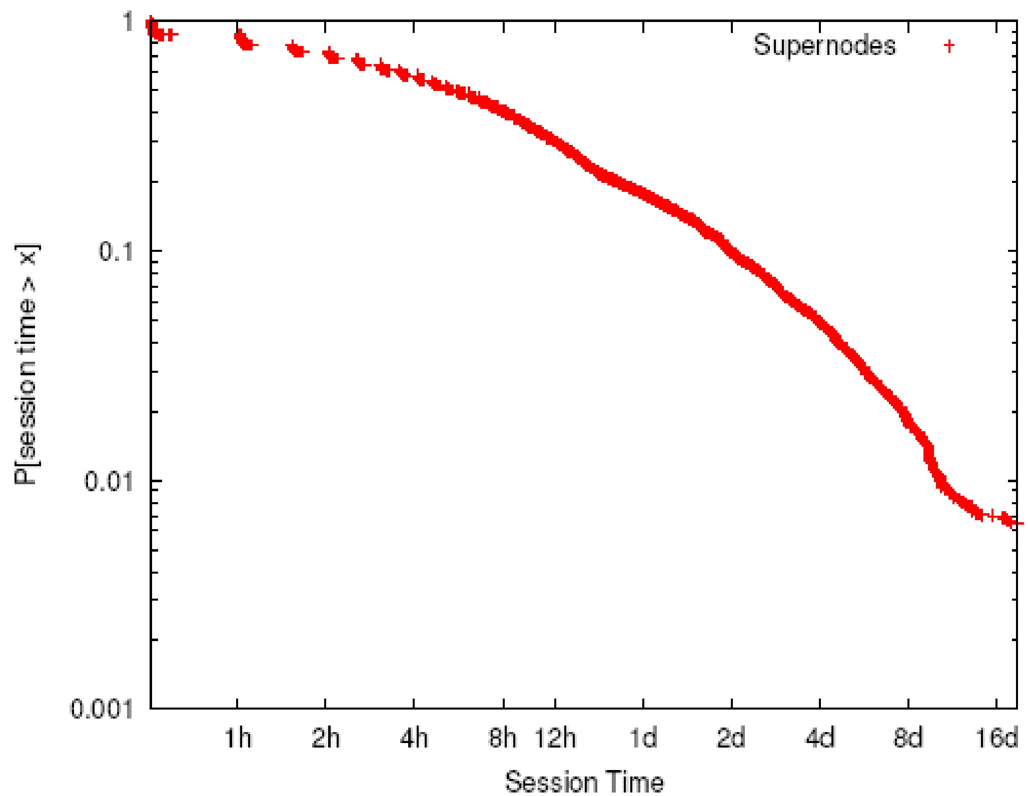
Why does Skype NAT Traversal work?

- There is (currently) an adequate ratio of peers with open IP addresses to NAT-bound peers
- There is low churn among Super-Peers
  - Super-peers show have median session times of several hours.
  - Super-peer session lengths are heavy-tailed and are not exponentially distributed.

Super-Peer Session Times in Skype



Online Time: Super-Peer (Red) versus Ordinary Peers (Blue)



Super-peer session times in Skype (Log-log plot of the CDF)

Guha et al., “An Experimental Study of the Skype Peer-to-Peer VoIP System”, Workshop on P2P '06

Other Skype Features

- User profiles and chat histories are cached/stored using the Global Index?
- Call conferencing
- Video Conferencing
  - Is video conferencing for NAT-bound peers feasible using current Skype technology?

Skype's Impact

- Skype demonstrates that:
  - P2P networks can overcome NAT traversal problems for UDP
  - Call Signaling and establishment is feasible using decentralised protocols and overlay P2P networks
  - P2P overlay networks can scale out to handle millions of users, with real-time services such as VoIP

Super-Peer References

- Yang and Molina. Improving Search in P2P Networks, Dist. Comp. Sci Conf, 2002.
- J. Sacha, J. Dowling, R. Cunningham, and R. Meier. Using aggregation for adaptive super-peer discovery on the gradient topology. SelfMan, 2006
- Lada A. Adamic. Zipf, Power-laws, and Pareto - a ranking tutorial, HP Technical Paper, 2002.
- G. Paolo Jesi, A. Montresor, and O. Babaoglu. Proximity-aware Superpeer Overlay Topologies, SelfMan, 2006.

P2P Measurement References

- Sen and Wang, Analyzing peer-to-peer traffic across large networks, IEEE Trans on Networking, 2004.
- Pouwelse et al., The Bittorrent P2P File-Sharing System: Measurements and Analysis, IPTPS 2005.
- Project JXTA 2.0 Super-Peer Virtual Network, Sun Microsystems.
- Lada A. Adamic, Zipf, Power-laws, and Pareto - a ranking tutorial, HP Labs
- Saroiu et.al., “A Measurement Study of Peer-to-Peer File Sharing Systems”, MMCN, 2002
- Ripeanu, et.al., “Mapping the Gnutella Network: Properties of Large-Scale Peer-to-Peer Systems”, Internet Computing Journal, 2002

Skype Case Study

- S. Guha, N. Daswani, and R. Jain. An Experimental Study of the Skype Peer-to-Peer VoIP System. Workshop on Peer-to-Peer Systems, 2006.
- S. Baset and H. Schulzrinne. An analysis of the skype P2P internel telephony protocol. IEEE Infocomm 2006.
- S. Ren, L. Guo, and X. Zhang. Asap: an as-aware peer-relay protocol for high quality voip. IEEE DC,2006.