video search & epidemic protocols
in Zero-Server P2P Systems

tribler

Dr. Ir. Johan Pouwelse
Assistant Professor
j.a.pouwelse@TUDelft.nl
Delft University of Technology

TU Delft

P2P NEXT

FREEBAND

NWO
Netherlands Organisation
for Scientific Research

stw

fusion

2007 - 2013

Sixth Framework Programme

Contents

- Two different search approaches

- Query-response based systems

- Internet-deployments

  Napster, Gnutella, Kazaa

- Academic research

- Epidemic protocols

  - Fundamentals

  - Examples

  - Buddycast case study
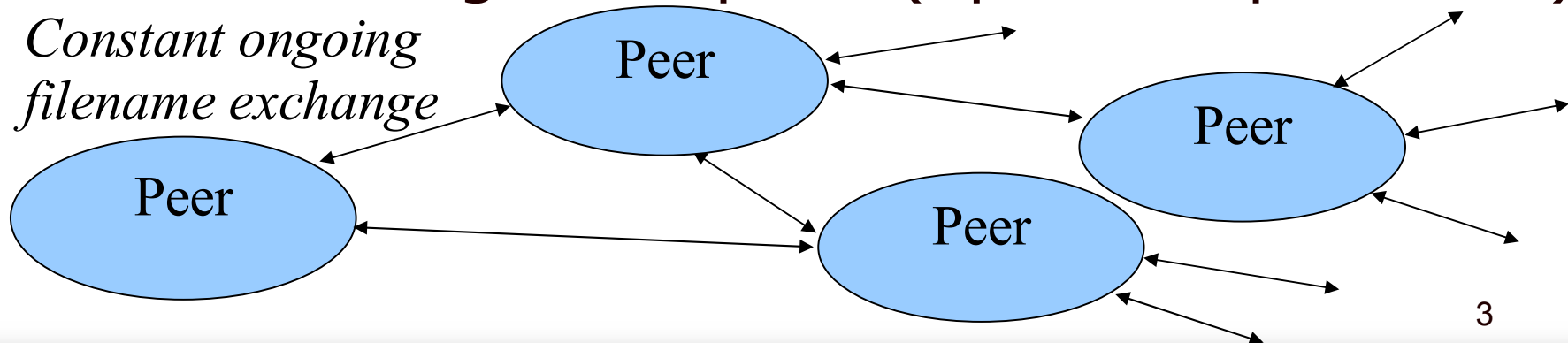
- P2P Video search: open research challenges

Two different search approaches

Proactive Pull

Real-time query & response

Peer

Query
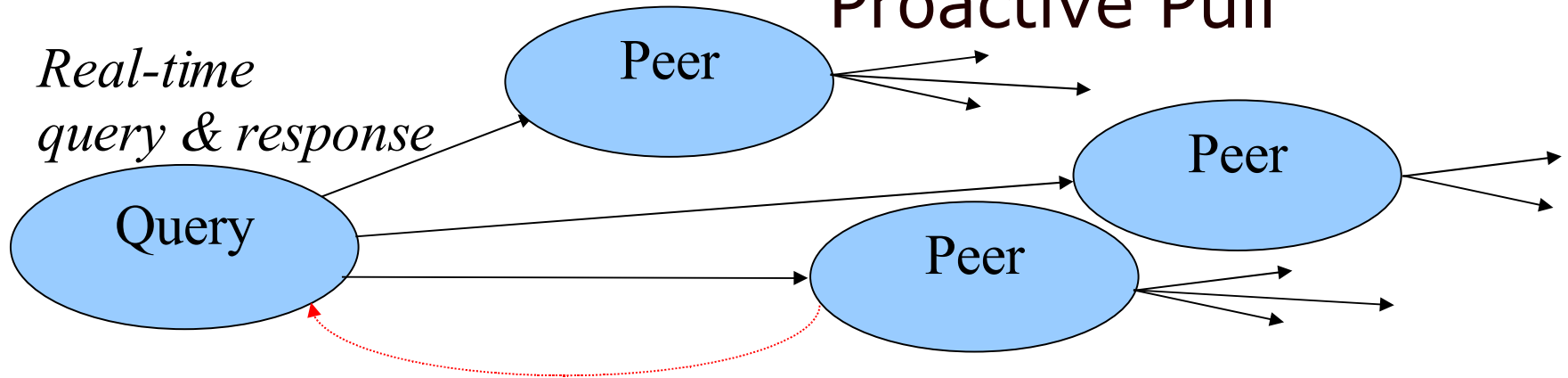
Peer

Peer

Background push (Epidemic protocols)

Constant ongoing filename exchange

Peer

Peer

Peer

Peer

Peer

P2P keyword search approaches

- Central server, flooding, superpeers
- Semantic routing
- DHT
"keyword lookup can be layered on top of the distributed hash model, it is an open question if indexing can be done efficiently"*

- Cubit, 2008, edit distance.

*Source"Looking up data in P2P systems",
  H Balakrishnan,F Kaashoek, D. Karger , R. Morris, I. Stoica,
  Communications of the ACM, February 2003

File   View   Player   Tools   Actions   Help

Web   My Kazaa Lite   Theater   Search   Traffic   Kazaalite.com Forums

New search   Download   Search Field

Search

Search for video files

Search for:

two towers

Found 231 files. Press Search More to get more results.

(•) All   ( ) Title   ( ) Artist

Search More   Stop Search

« Back

More search options »

Search filter

Search Tip

Type in what you are looking for. If you are looking for a specific media type you may check the corresponding button to get more search options.

| Title	User	Size
lords of the ring(3of3)	kazaalite...	1
[tmd]The.Lord.Of.The.Rings.The.Two.Towers.(NEWFXP).ts.(1of3)-real deal	(zenit)@...	1
Lord of the Rings The Two Towers (1of3)	kaak@K...	1
Lord Of The Rings: The Two Towers Preview (No Fake!) DIVX	2 Users	1
(tmW)Lord.Of.The.Rings.--.The.Two.Towers.(CD3of3)	3 Users	1
(tmW)Lord.Of.The.Rings.--.The.Two.Towers.(CD2of3)	2 Users	1
The Lord Of The Rings The Two Towers (2002) cd 1 of 3	6 Users	7
The Lord Of The Rings The Two Towers cd 3 of 3	9 Users	2
Lord Of The Rings(full)	4 Users	7
lord of the rings-the two towers cd2of3.mpg	2 Users	6
Lord of The Rings- The Two Towers (Part 2/2	3 Users	1
[TMD]The.Lord.Of.The.Rings.The.Two.Towers.(TCF).INTERNAL.CAM.(3of3)	15 Users	1
The.Lord.Of.The.Rings.The.Two.Towers.INTERNAL.CAM-TCF.CD1	2 Users	6
the lord of the rings ll	4 Users	7
The Lord Of The Rings - The Two Towers [INTERNAL.CAM-TCF.CD3]	2 Users	5
The Lord of the Rings - The Two Towers (2002 FULL MOVIE) (DivX5) (640x352)	9 Users	1
Lord Of The Rings The Two Towers cd 2	6 Users	7
The Lord Of The Rings The Two Towers (2002) cd 2 of 3	2 Users	7
Lord of the Rings- Two Towers FULL (1/2) [DivX]	6 Users	6
the_two_towers-trailor_fs.l	3 Users	4
Lord of the Rings: The Two Towers	13 Users	7
Two towers trailer	2 Users	3
Lord Of The Rings - The Two Towers Trailer Large	kvist@K...	3
(Smr)lord of the rings - The two towers (Real. Ripped From Studio;)) 1 of 3	5 Users	1
Sagan om de två tornen-Trailer	matzz@...	3
Lord of the Rings 2 - The Two Towers	4 Users	6

Semantic routing
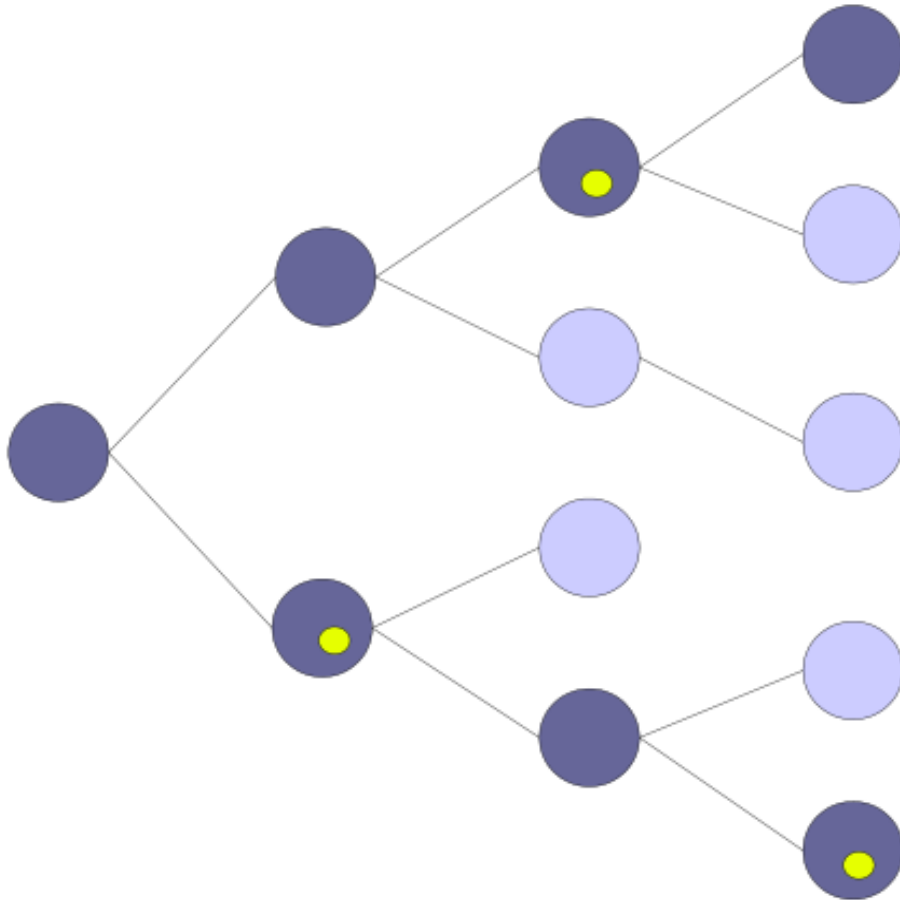
A query is forwarded to nodes that are likely to have knowledge about a reply based on semantics.
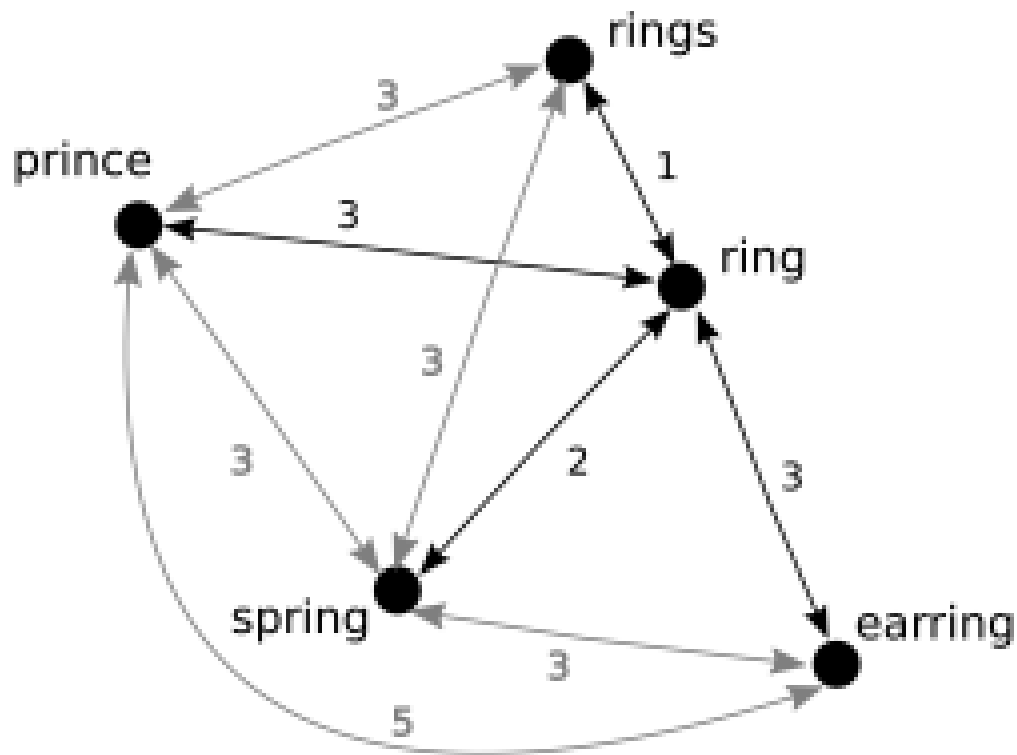
Explored in: Socialised.net, Neurogrid, Alpine, Limewire

Multi-keyword search using DHT and Bloom filters
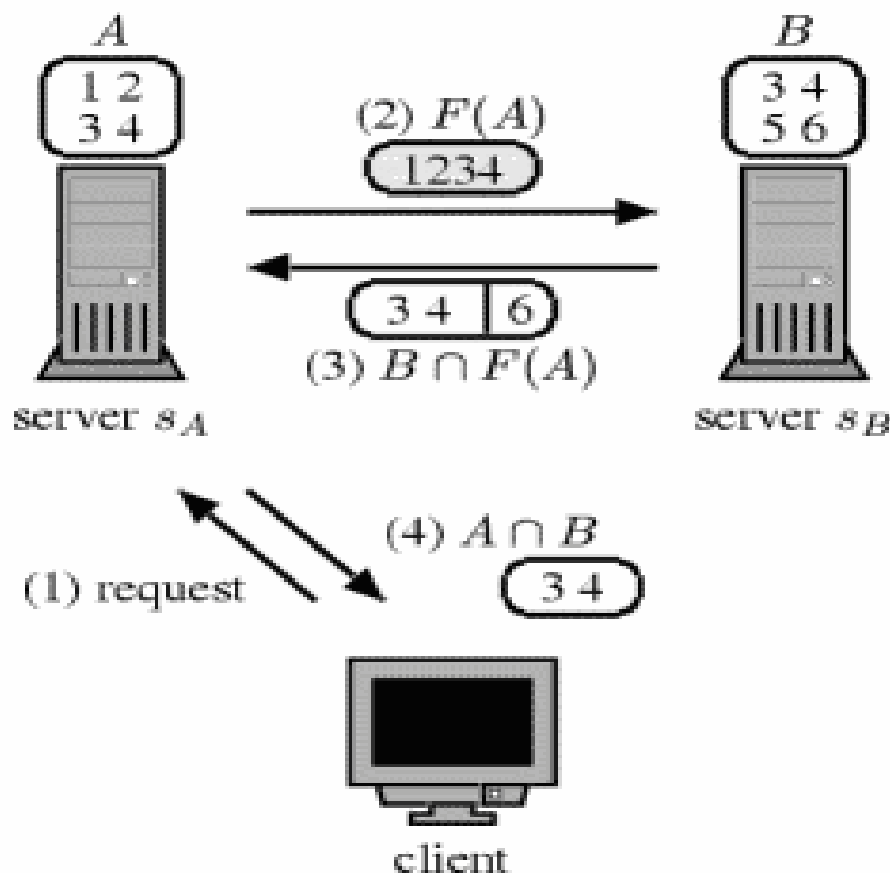
- s_A and s_B: servers storing the document ID lists for keywords k_A and k_B.

- A and B are the sets of document IDs matching the keywords k_A and k_B.

3. F(A) is a Bloom filter representation of A.

Source: Patrick Reynolds and Amin Vahdat,
"Efficient Peer-to-Peer Keyword Searching," Middleware 2003

Cubit: approximate keyword search

1. Edit nearness

2. No semantics

3. Released in 2007

Source: http://www.cs.cornell.edu/~bwong/Cubit

Cubit search

Epidemic protocol fundamentals

- Each peer has information regarding the state of the entire distributed system

- Each peer periodically chooses a peer with which to share this its information

- New information is merged with existing information to obtain a new state

- Design space: target peer, info exchange, retention

- Challenges
  - Availability
  - Connectivity
  - Security

Epidemic protocol examples

- Xerox Clearinghouse service (1987); replica consistency

- Newscast (2002)*, vicinity, T-MAN, BAR gossip

- BuddyCast (2006); *first Internet-deployed protocol*

  - Limit randomness

  - Avoid one-time encounters

  - Both peer and content discovery

  - Apply collaborative filtering

  - Exploit the TeraByte hard-disk age

  - Preserve state across sessions

*Source: M. Jelasity, M. van Steen. "Large-Scale Newscast Computing on the Internet." Technical Report IR-503, October 2002.

Preserve state across sessions

- Storage of state in relational database
- Long-term peer identifier (e.g. PermID)
- Robust against IPv4 changes
- PermID: Non-spoofable and compact
- Store:
  content, peers, profiles, social network...

Bob → Alice: PID_b

Bob ← Alice: PID_a

Bob → Alice: Rb

Bob ← Alice: Sa (Ra, Rb, PID_b), PUB_a, Ra, PID_b

Bob → Alice: Sb (Ra, Rb, PID_a), PUB_b, PID_a
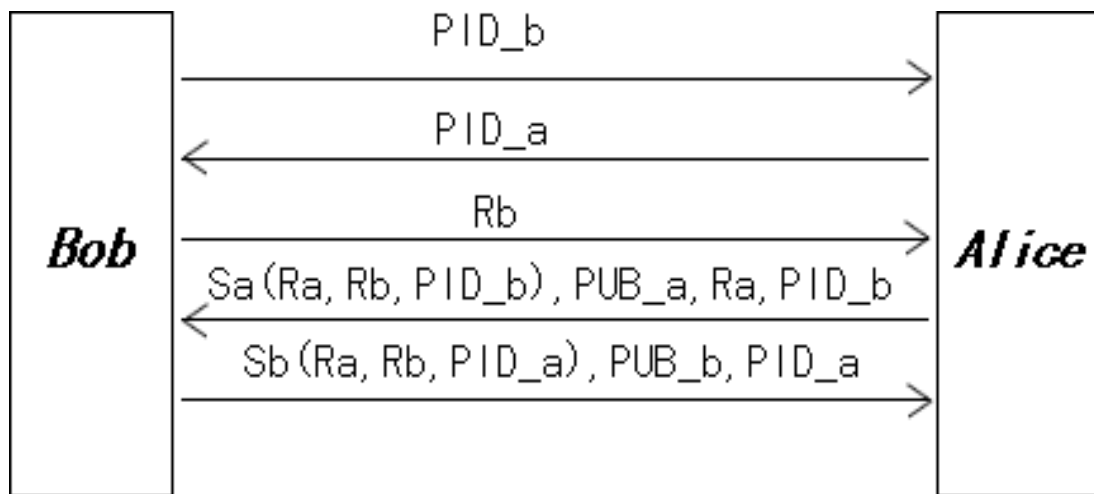
*R:Random number, S:Signature, PID:BT PeerID, PUB:PermID

Peer

peer_id

permid
name
ip
port
thumbnail
oversion
similarity
friend
superpeer
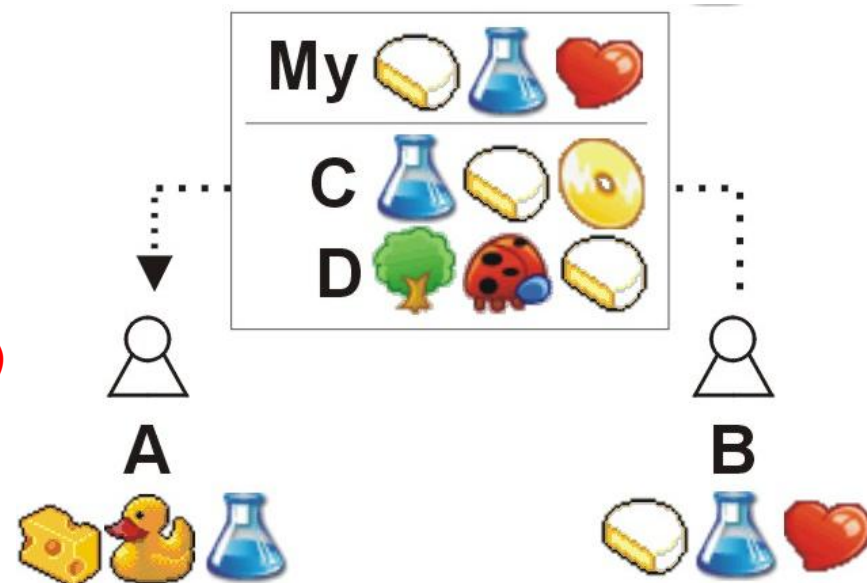last_seen
last_connected
last_buddycast
connected_times
buddycast_times
num_peers
num_torrents
num_prefs
num_queries

MyInfo

entry

value

Category

category_id

name
description

Torrent

torrent_id

infohash
name
torrent_file_name
length
creation_date
num_files
thumbnail
insert_time
secret
relevance
source_id (FK)
category_id (FK)
status_id (FK)
num_seeders
num_leechers
comment

TorrentStatus

status_id

name
description

TorrentSource

source_id

name
description

TorrentTracker

tracker
torrent_id (FK)

announce_tier
ignored_times
retried_times
last_check

BarterCast

peer_id_from
peer_id_to

downloaded
uploaded
last_seen
value

Preference

peer_id (FK)
torrent_id (FK)

MyPreference

torrent_id (FK)

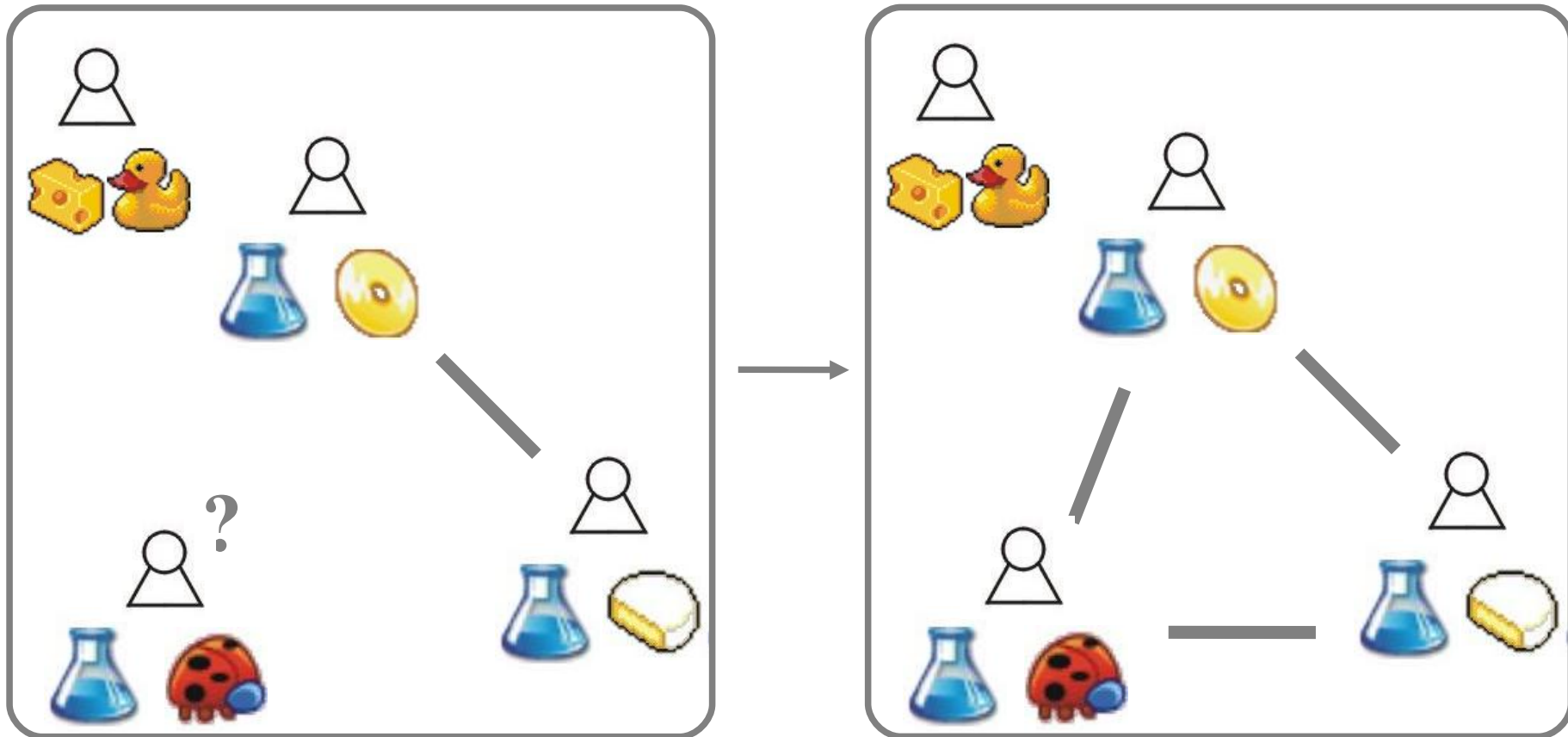destination_path
progress
creation_time

Semantic clustering

- Organize P2P network
- References to related items
- Used in Buddycast
- Message layout
  - Your preferences
  - Your similar peers *(taste buddies)*
  - Their preferences
- Emergence
  - Peer discovery
  - Content discovery
  - Database of tastes

Semantic discovery

BuddyCast message:
peers and Bittorrent items

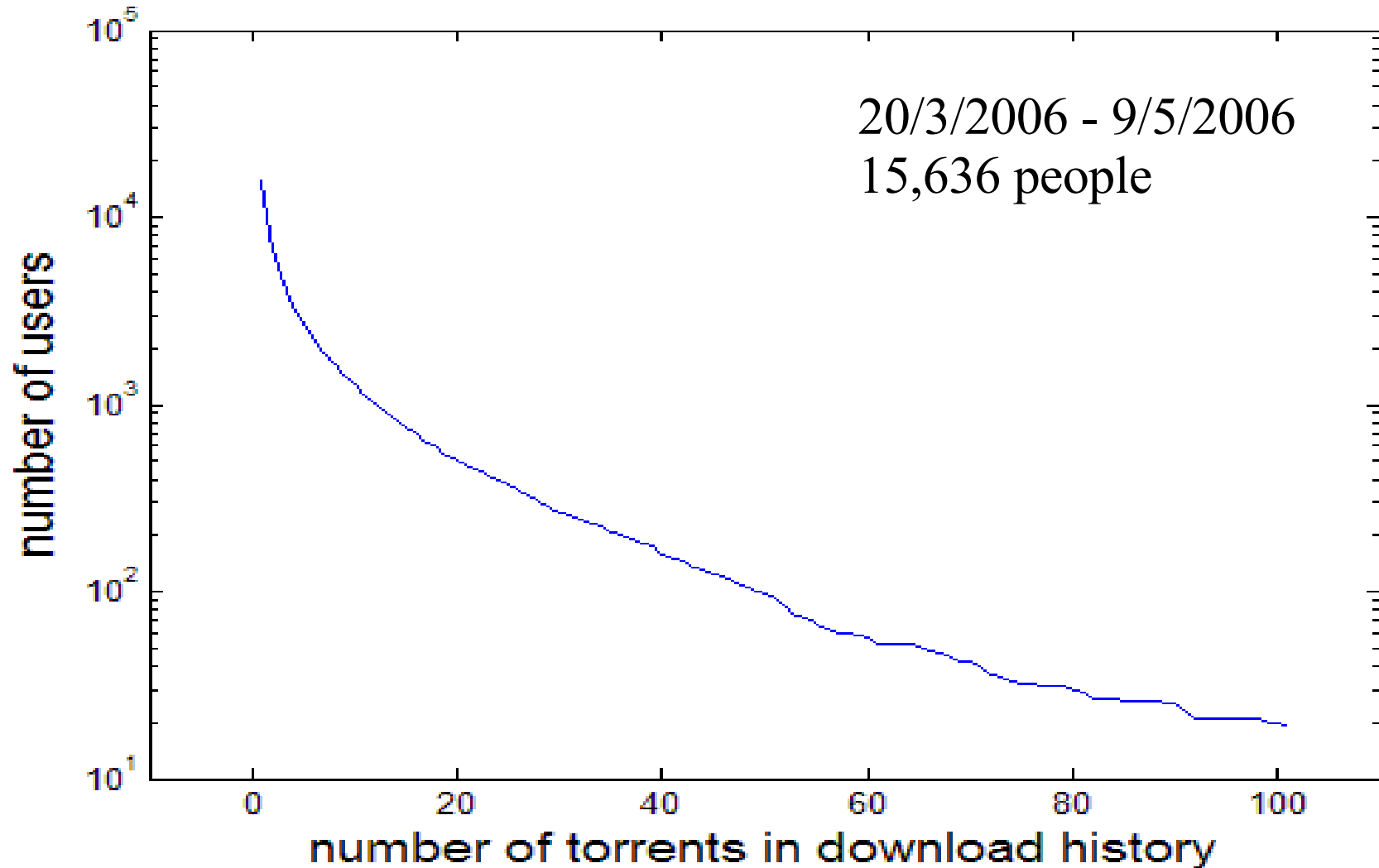

Recent downloads (max 25)
{SHA1,seed/leech,last_seen}

Recent discovered (max 25)
{SHA1,seed/leech,last_seen}

Taste Buddy (max 10)
{IP,Port,Similarity,last_seen)

Random Peer (max 10)
{IP,Port,Similarity,last_seen)
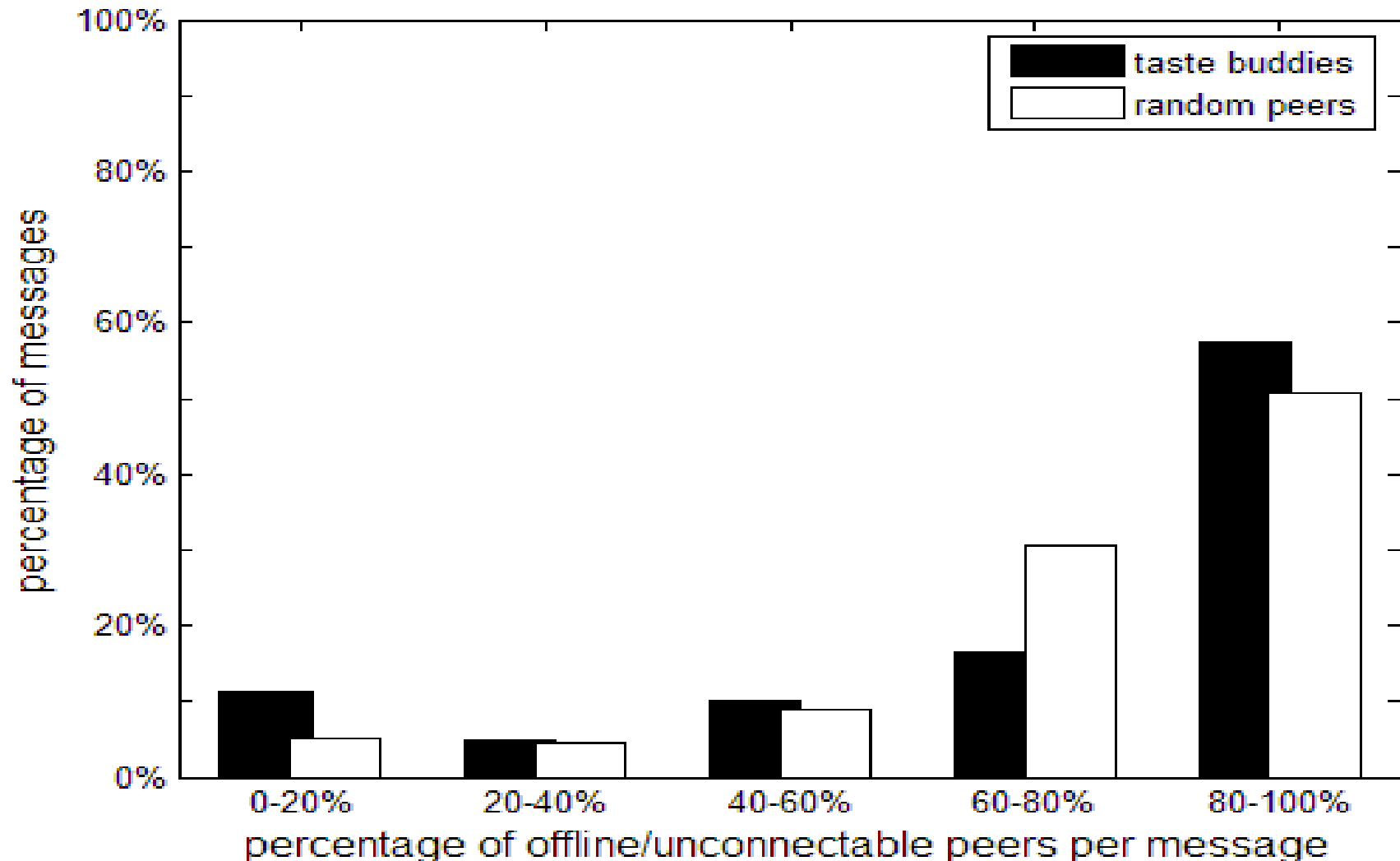
BuddyCast trace of P2P downloads

20/3/2006 - 9/5/2006
15,636 people

number of users (y-axis, from 10^1 to 10^5)

number of torrents in download history (x-axis, 0 to 100)

Buddycast 1: no online check (2006)

Improvement: Keep 30 Open TCP connections
CC: Connection Candidate Cache   (size=100)
      Potential targets for next connection
CT: Connected Taste buddies
      Most similar discovered peers (size=10)
CR: Connected Random peers
      Last connected peers   (size=10)
CU: Connected NATed/Firewalled peers
      Connected peers with   (size=10)

```
loop:
 wait(ΔT time units) {wait=15 seconds}
 Q ← Top similar peer or likely online peer from CC
 connectPeer(Q)
 remove Q from CC
 if Q is connected successfully then:
  buddycast msg send ← createBuddycastMsg()
  send buddycast msg send to Q
  receive buddycast msg recv from Q
  CC ← fillPeers(buddycast msg recv)
  addConnectedPeer(Q) {add to CT,CR or CU using sim()}
```
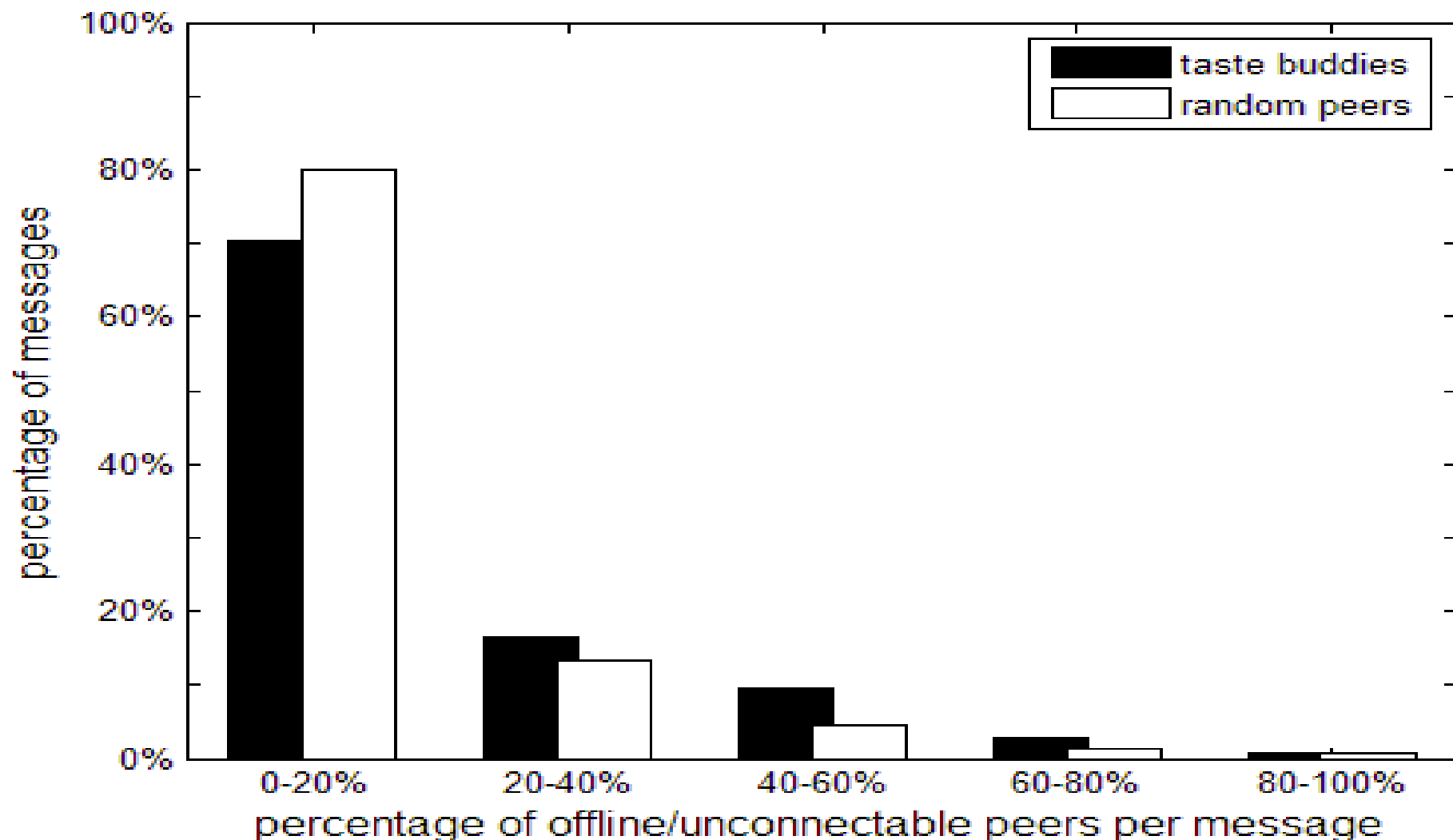
19

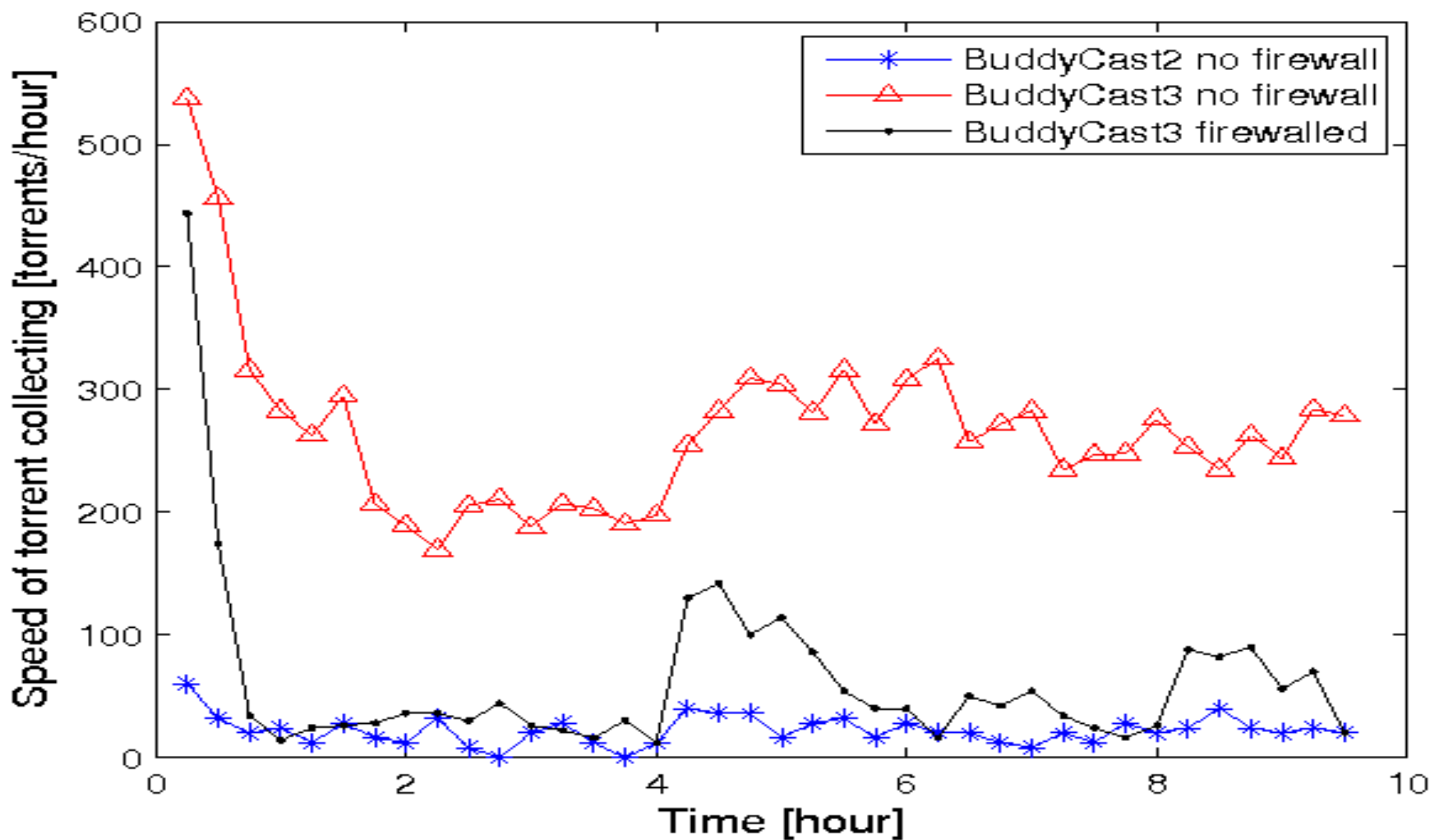# Buddycast 2: **with** online check

Epidemic protocol and keyword search

- Peers gather peers, items, profiles (size=100k)

- Keyword search

    ▪ Local database

    ▪ N peers around you (x 100K items)

      "remote search"

    ▪ Validate validity of P2P items

- Hybrid solution:

  proactive pull + constant background push

Buddycast3: dynamic speed in 1ˢᵗ hour

Tribler 4.1.9

File   Tools   Help

files   persons   profile   library   subscr   friends

family filter: ON OFF

discovered files: 7258
discovered persons: 3083
Person connected:
65.191.76.140:6881 36.765
(peer deemed unreachable)

tribler

Discovered files   thumbnails ▾

legal torrents   🔍   All ▾

stop   clear results   Searching for 'legal torrents'... (112 results)

compressed

[Exclusive] Tiesto -
Club Life 026
09-28-07 -

video

DivX for Windows
(incl. DivX Player)
v6.6.2 -LegalTorrents

PeerIt.com Peer to
Peer Signup and Sell
Example

Mac Tips -
Transmission

What is a Torrent? Is
BitTorrent Legal?

TUTORIAL: How to
use BitTorrent and
download Torrents

Fairy From
Wonderland -Tian
Wai Fei Xian episode

Fairy From
Wonderland -Tian
Wai Fei Xian episode

Fairy From
Wonderland Tian Wai
Fei Xian episode 1

Fairy From
Wonderland Tian Wai
Fei Xian episode 1

Fairy From
Wonderland Tian Wai
Fei Xian episode 1

Free Legal Movies
Videos To
Download?

Fairy From
Wonderland Tian Wai
Fei Xian episode 1

other

Freeware Is Better
Than Torrents -
Duncan's Freeware

other

The Art of Writing
Speaking the English
Language eBooK

compressed

Top 100 Fonts
(Graffiti, Horror,
Handwritten More) -

other

Cool Youtube
Downloader Full -
LegalTorrents

compressed

DVD Shrink
v3.2.0.15 -
LegalTorrents

other

The Complete Works
of William
Shakespeare eBooK

[www.torsky.org]
Jassie from Barely
Legal

◀ 1 2 3 4 ▶

Fairy From Wonderland Tian Wai Fei Xian

options ▾

info   files

size: 11:43
creation   23-10-2006
popularity: 15385
fit to taste: ♡

play

Description

from youtube: the first episode of Fairy From
Wonderland or Tian Wai Fei Xian with english
subs---i
dont think this is illegal t...

People who like this also like

No similar files found yet.

Similar Files

No similar files found yet.

Video search: open research challenges

- Spam & fraud prevention

- Rich metadata: tagging and ratings, subtitles, annotations..

- Moderation

- Relevance ranking

  - Semantic matching; word boundary, swarm or file name..

  - Popularity and age

  - Perceptual quality and download speed

  - Bonus or penalty based on collaborative filtering

  - Implicit feedback usage; Clicklogs

Final remarks

- What is P2P Video without search capability?

- Maturing field and transition:

  - From simulations to prototypes

  - From models to datasets

- Deployed self-organizing systems

- Importance of on-line video

  - Youtube influences wars and elections

  - Media is where society thinks

  - Mediaspace controlled by the people

- New business opportunities: reach millions at near-zero cost