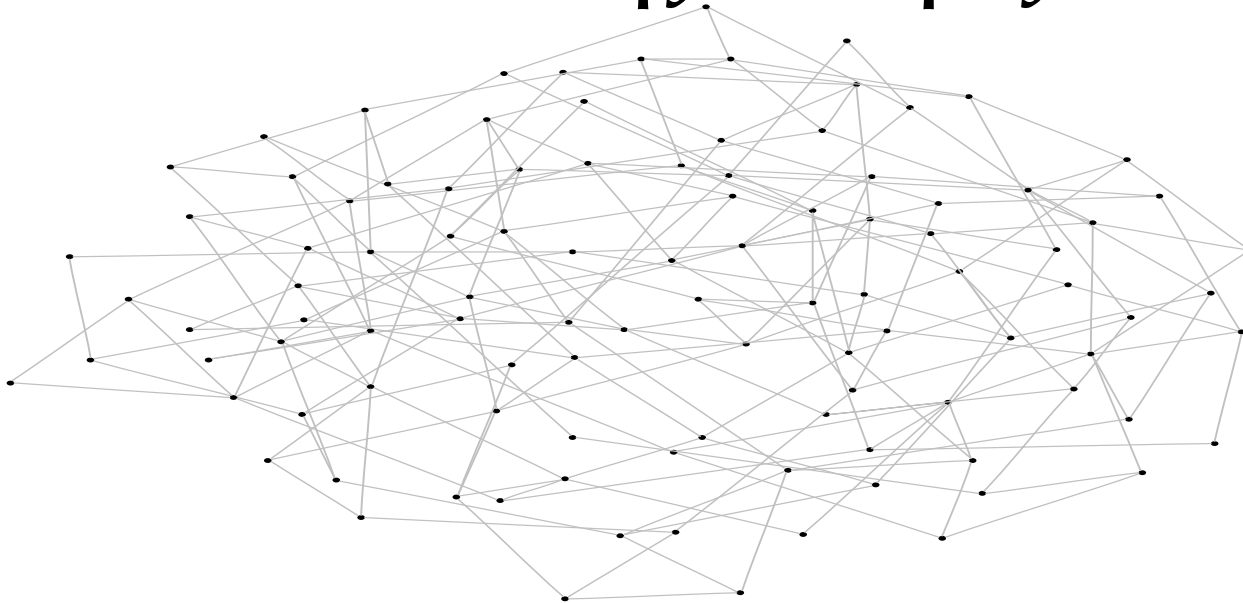


“Peer-to-Peer Systems” / “Συστήματα Ομοτίμων”

**Δημήτρης Ζείναλιπούρ
Επισκέπτης Λέκτορας**



<http://www.cs.ucy.ac.cy/~dzeina/>

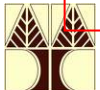


Στόχοι Διάλεξης

- Μια γενική εισαγωγή στα Συστήματα Ομοτίμων (**Ελληνικές Διαφάνειες**).
- Αναφορά σε ένα από τα πολλά ερευνητικά προβλήματα: “Information Retrieval in Peer-to-Peer Systems” (**Αγγλικές Διαφάνειες**)

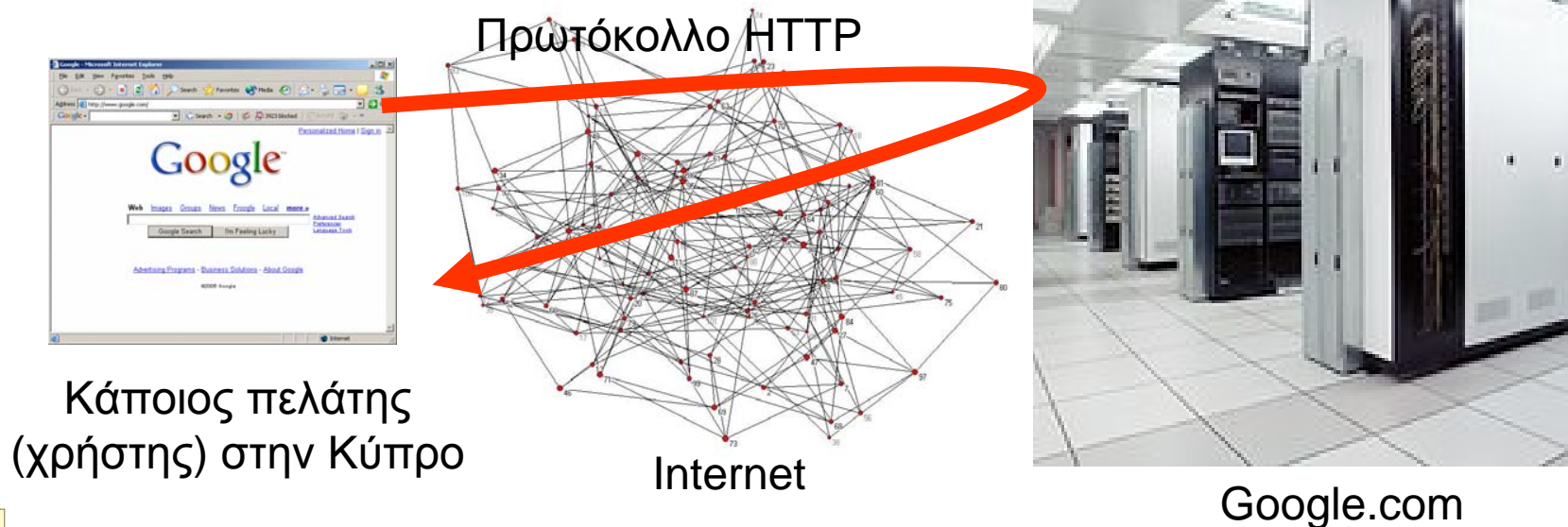


Μέρος 1: Εισαγωγή στα Συστήματα Ομοτίμων



Από συστήματα Πελάτη-Εξυπηρετητή σε Συστήματα Ομοτίμων

- Οι περισσότερες υπηρεσίες στο Διαδίκτυο ακολουθούν το μοντέλο Πελάτη-Εξυπηρετητή (π.χ. HTTP, FTP, DNS, POP3,...)



Χαρακτηριστικά συστημάτων Πελάτη-Εξυπηρετητή (Π/Ε)

Πλεονεκτήματα

- Κεντρικοποιημένη διαχείριση δεδομένων,
- Ασφάλεια δεδομένων (**Security**),
Αυτά τα χαρακτηριστικά είναι ιδανικά για πολλές διαδικτυακές βάσεις δεδομένων (τράπεζες, email, web κτλ)

Μειονεκτήματα

- Χρειάζονται ακριβές υπολογιστικές υποδομές
- Σε περίπτωση βλάβης του εξυπηρετητή χάνουμε την υπηρεσία (**single point of failure**).
- Χρειάζονται συνεχή διαχείριση (**administration**).
- Μπορούν να λογοκριθούν (**copyright**) και να ελεγχθούν (από καθεστώτα, κτλ)

Τα P2P Systems προσπαθούν να ξεπεράσουν αυτά τα προβλήματα.



Τι είναι το Peer-to-Peer (P2P)?

- **Φοιτητής:** “Online Gaming” και “Multimedia downloads”
- **Υπάλληλος Εταιρείας:** “Instant Messaging”
- **Warner Bros:** Η φθηνή (για την εταιρεία) διανομή ταινιών σε πελάτες. **Sony:** Φθηνό Online Gaming.
- **Άλλοι:** Αποδοτικό & Φθηνό VOIP (Skype)
- **Διάφοροι Ορισμοί φαίνεται να συμφωνούν στα πιο κάτω:**
 - Ο Διαμοιρασμός Πόρων (**Resource Sharing** - CPU cycles, DISK, Bandwidth, etc.)
 - Απευθείας Επικοινωνία (**Direct Communication**) μεταξύ Ίσων (peers)
 - Αποκέντρωση (**decentralization**) & αυτόματη οργάνωση (**self organization**).



Τι είναι το Peer-to-Peer (P2P)?

- Οι χρήστες του συστήματος έχουν διπλό ρόλο: δηλ. είναι **Πελάτες** και **Εξυπηρετητές** την ίδια στιγμή.
- Καθώς ο χρήστης A εξυπηρετείτε από τον χρήστη B, κάποιος χρήστης Γ μπορεί να εξυπηρετείτε από τον A.
- Δεν υπάρχει κεντροποιημένη διαχείριση...το οποίο δημιουργεί ένα αίσθημα ελευθερίας.
- ...Παράλληλα όμως δημιουργεί και πολλά προβλήματα (όπως αυτό της παράνομης ανταλλαγής τραγουδιών)



Στόχοι και Πλεονεκτήματα του P2P

- **Edge-Computing**
 - Αξιοποίηση αχρησιμοποίητου **bandwidth, storage, processing power** στα **άκρα (edge)** του Internet
- **Scalability**
 - Δεν υπάρχει συμφόρηση (bottleneck) σε κάποιο κεντροποιημένο κόμβο. Επομένως τα συστήματα αυτά μπορούν να μεγαλώνουν “απεριόριστα”.
 - Oct 12th Gnutella (Limewire.com): 2,219,539 κόμβοι
- **Reliability (Αξιοπιστία)**
 - No single point of failure, Γεωγραφική Κατανομή Περιεχομένου (CDNs)
- **Ease of administration**
 - Οι κόμβοι οργανώνονται μεταξύ τους αυτόματα (**self-organization**).
 - Αυτόματα επίσης γίνεται το **replication** και **load balancing** καθώς επίσης τέτοια συστήματα παρέχουν **fault tolerance**.
- **Anonymity – Privacy**
 - ...κάτι το οποίο δεν είναι εύκολο σε ένα centralized σύστημα



Λόγοι Ανάπτυξης του Peer-to-Peer?

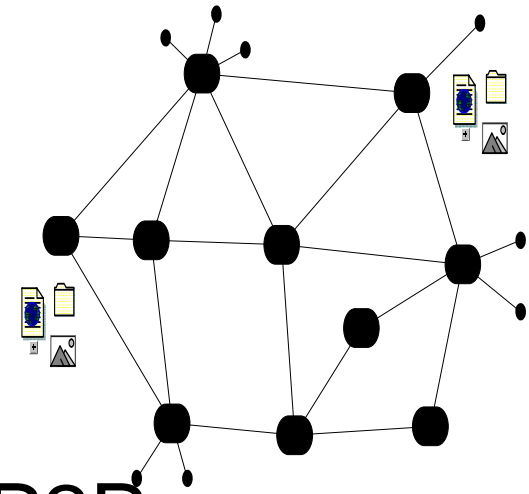
(*στην μορφή που το ορίζουμε σήμερα)

- Ο **μεγάλος αριθμός PCs** με πολύ ψηλή υπολογιστική ισχύ διαθέσιμα στα άκρα (edges) του Internet.
- Επίσης, υπάρχει πολύ **ψηλότερο bandwidth** διαθέσιμο στα άκρα του Internet (ADSL, Satellite, Cable, Ethernet LANs, etc.)
- Επομένως, γίνεται εφικτό να αξιοποιήσουμε το ***Storage, Cycles, Content*** και να έχουμε αλληλεπίδραση χρηστών στα άκρα (edges) του Internet, χωρίς την χρήση κεντρικών servers .
- * Σημειώστε ότι το άλλα Internet Services από το 1980- π.χ. DNS, BGP, Usenet etc, ορίζουν και αυτά ένα P2P μοντέλο ανταλλαγής πληροφοριών



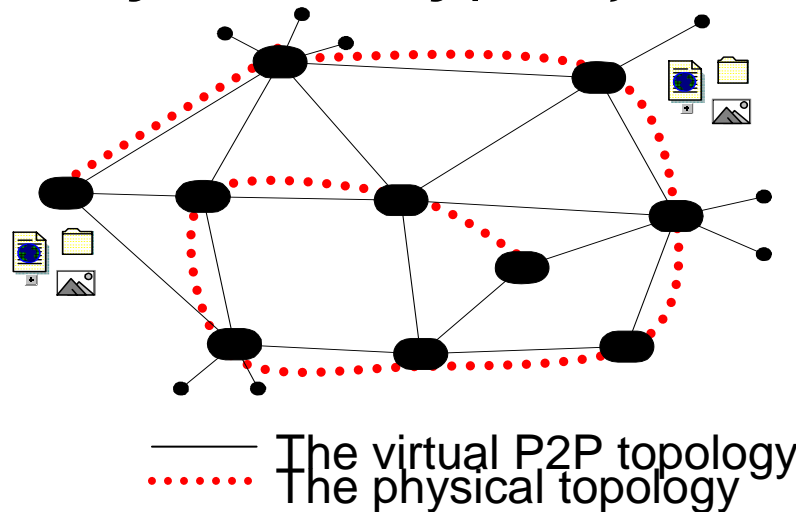
Βασικά Θέματα που απασχολούν την Ερευνητική Κοινότητα Peer-to-Peer ?

1. **Data Placement:** Που πρέπει να τοποθετηθούν τα δεδομένα?
2. **Search Mechanisms:** Με τι μηχανισμούς θα βρούμε την πληροφορία?
3. **Overlay Structure:** Πως θα οργανώσουμε την δομή του P2P overlay δικτύου, έτσι ώστε να μπορούμε να πετύχουμε το 1-2 πιο αποδοτικά (χρόνος, χώρος, ανωνυμία, κτλ)



P2P Συστήματα και Overlay Networks

- **P2P Συστήματα** είναι δομημένα πάνω από **Overlay Networks (χρησιμοποιούνται και οι δυο όροι)**.
- Δηλαδή οι peers, εγκαθιδρύουν TCP ή UDP socket connections με άλλους κόμβους. Αυτό δημιουργεί ένα νοητό “virtual” γράφο διασύνδεσης.
- Ο γράφος αυτός **δεν** αντιπροσωπεύει τις πραγματικές συνδέσεις μεταξύ των κόμβων αυτών.



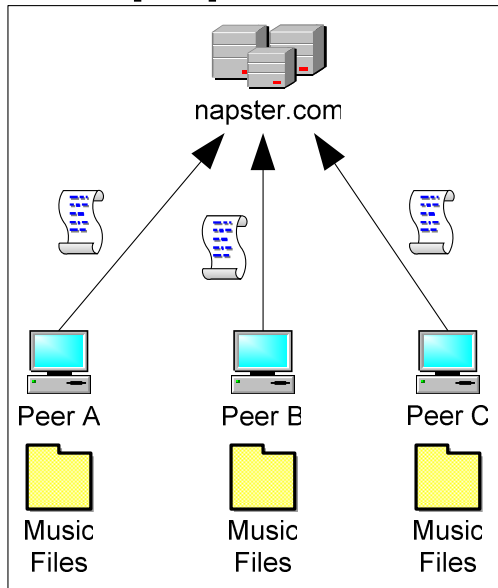
Εφαρμογές Συστημάτων Peer-to-Peer

A) Ανταλλαγή Αρχείων (File-Sharing)

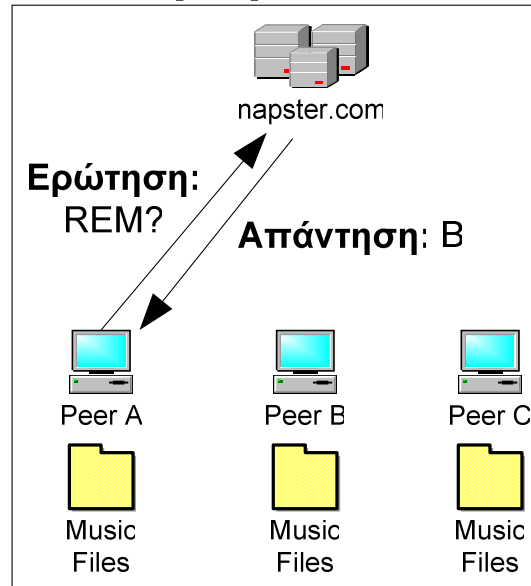
1) Napster:

- Δημιουργείται το 1999 από ένα 18-χρονο φοιτητή.
- Ο κάθε χρήστης εκτελεί τις ακόλουθες τρεις διαδικασίες

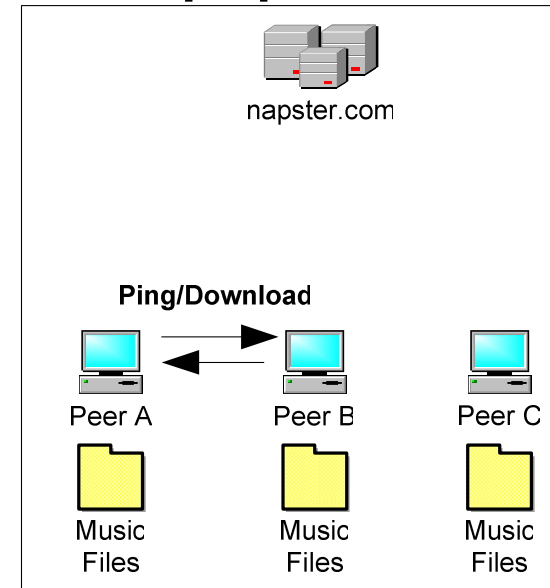
Αποστολή Λίστας Τραγουδιών



Αναζήτηση Τραγουδιών



Ανάκτηση Τραγουδιών



Παράθυρο Αναζήτησης Napster

Napster v2.0 BETA 7

File Actions Help

Home Chat Library Search Hot List Transfer Discover Help

Artist: Find it!

Title: Clear Fields

Max Results: Advanced >>

Filename	Filesize	Bitrate	Freq	Length	User	Connection	Ping
incomplete_other_artist\Tito Puentes Golden Latin Jazz Allstars - Oye Como ...	3,696,640	128	44100	3:51	bdenzler	DSL	343
incomplete_other_artist\[Marty Robbins] The Fastest Gun Around.mp3	542,304	128	44100	0:39	bdenzler	DSL	343
incomplete_other_artist\Ravi Shankar - Chants Of India 04 - Asato Maa.mp3	2,449,408	128	44100	2:35	bdenzler	DSL	343
other_artist\Engelbert Humperdinck - White Christmas.mp3	9,277,648	320	44100	3:52	bdenzler	DSL	343
other_artist\Grateful Dead - Franklin's Tower - Reggae Style.mp3	4,635,458	128	44100	4:48	bdenzler	DSL	343
Unknown Artist - You seriously have to listen to this.mp3	462,848	318	16000	0:17	sam113...	Cable	383
MP3z\artist - 'The Way Life Is' By Drag-On featuring Case.mp3	4,726,784	128	44100	4:54	burg651	Cable	386
MP3z\artist - 'Opposite Of H2O' By Drag-On featuring Jadakiss.mp3	3,540,992	128	44100	3:41	burg651	Cable	386
Various Artist - Perfect Day 97.mp3	3,722,344	128	44100	3:53	falkstad	ISDN-128K	398
Liszt\Liszt - Etude 'Un sospiro' - Cziffra-artist.mp3	2,752,512	128	44100	2:53	Iskjdfkjl...	Unknown	504
Music\Waiting To Exhale - Original Soundtrack Album - Various Artist - Count...	3,199,083	96	44100	4:26	Jzfork9	56K	511
Track 03_artist.mp3	4,054,332	128	44100	4:13	immusic...	Cable	514
Track 02_artist.mp3	6,228,974	128	44100	6:26	immusic...	Cable	514
Track 01_artist.mp3	4,731,426	128	44100	4:54	immusic...	Cable	514
Track 04_artist.mp3	4,514,505	128	44100	4:41	immusic...	Cable	514
Track 05_artist.mp3	4,105,323	128	44100	4:16	immusic...	Cable	514
mixer in track 01_Artist_0721011750.mp3	180,686	128	44100	0:17	immusic...	Cable	514
Album\Reflex - Keep In Touch-Artist.mp3	7,041,024	160	44100	5:49	rotimca	56K	527

Returned 100 results.

Get Selected Songs Add Selected User to Hot List

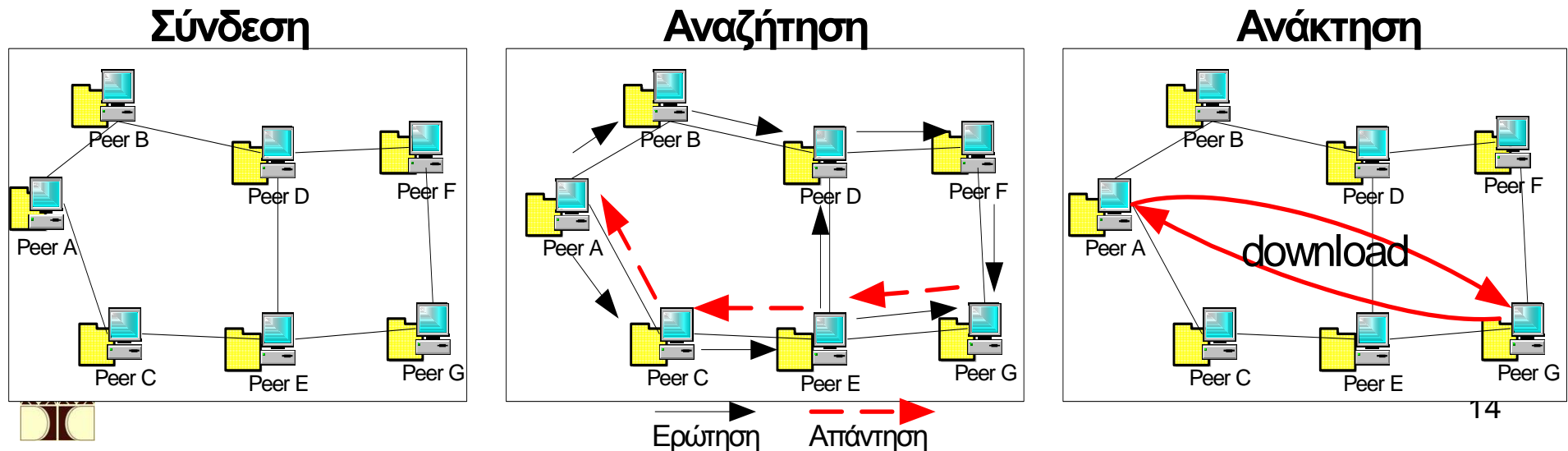
Online (keyscreen): Sharing 491 files. Currently 740,043 files (2,991 gigabytes) available in 5,873 libraries.



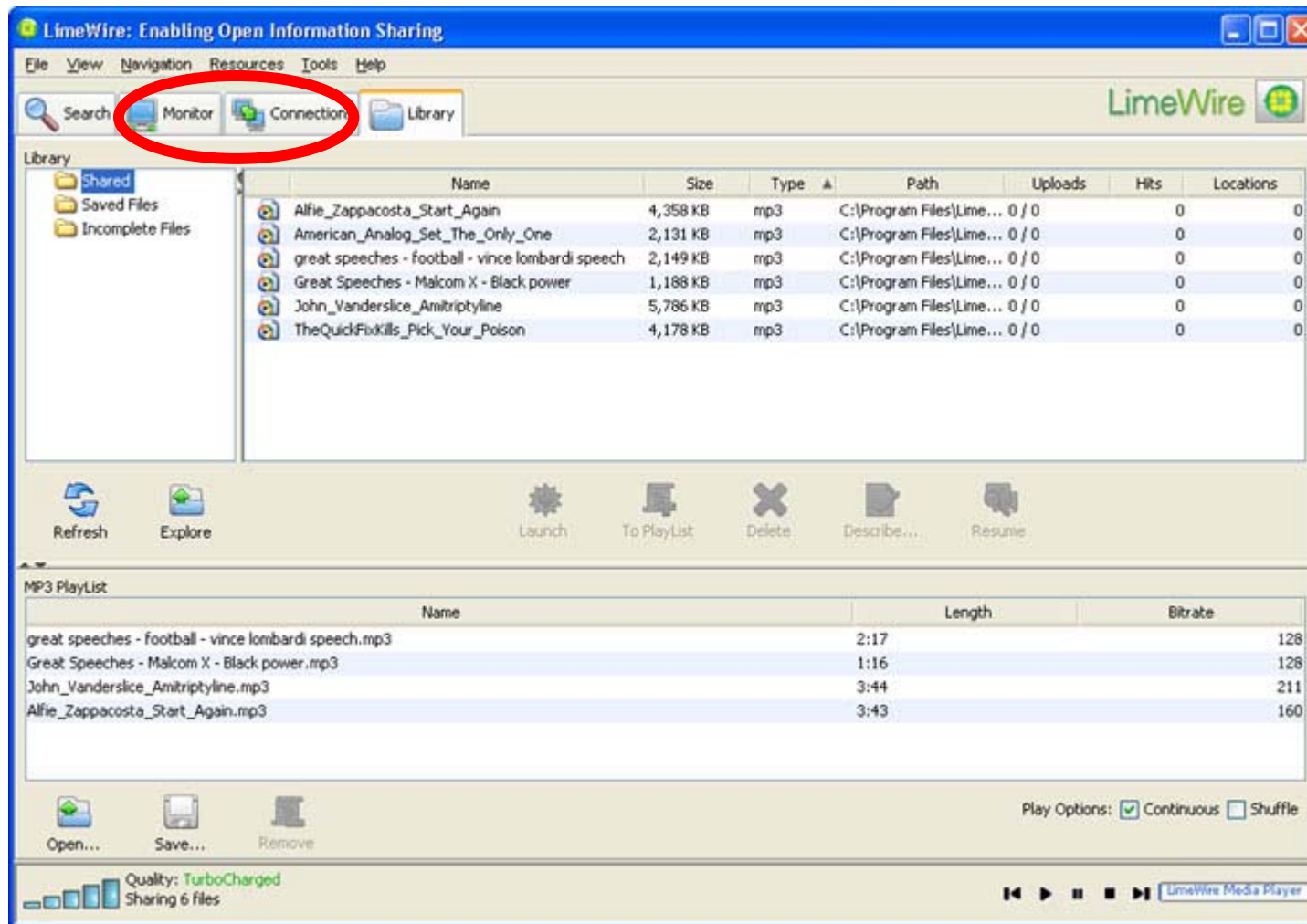
Εφαρμογές Συστημάτων P2P

2) Gnutella:

- Το (αρχικό) Napster κλείνει το 2001 μετά από απόφαση του δικαστηρίου για παράνομη ανταλλαγή αρχείων.
- Στις αρχές του 2000 δημιουργείτε ένα νέο σύστημα στο οποίο δεν υπάρχει κεντροποιημένη διαχείριση της λίστας των αρχείων.
- Με αυτό τον τρόπο δημιουργείτε ένα εντελώς κατακεκομμένο σύστημα (που θεωρητικά δεν μπορεί να κλείσει κανείς...)



Παράθυρο Αναζήτησης Gnutella (LimeWire)



The screenshot displays the LimeWire application window titled "LimeWire: Enabling Open Information Sharing". The interface includes a menu bar (File, View, Navigation, Resources, Tools, Help) and a toolbar with buttons for Search, Monitor, Connection, and Library. The "Monitor" button is circled in red. Below the toolbar, the "Library" section shows a list of files with columns for Name, Size, Type, Path, Uploads, Hits, and Locations. The "MP3 PlayList" section at the bottom shows a list of files with columns for Name, Length, and Bitrate. The status bar at the bottom indicates "Quality: TurboCharged" and "Sharing 6 files".

Name	Size	Type	Path	Uploads	Hits	Locations
Alfie_Zappacosta_Start_Again	4,358 KB	mp3	C:\Program Files\Lime... 0 / 0	0	0	0
American_Analog_Set_The_Only_One	2,131 KB	mp3	C:\Program Files\Lime... 0 / 0	0	0	0
great speeches - football - vince lombardi speech	2,149 KB	mp3	C:\Program Files\Lime... 0 / 0	0	0	0
Great Speeches - Malcom X - Black power	1,188 KB	mp3	C:\Program Files\Lime... 0 / 0	0	0	0
John_Vanderslice_Amitriptyline	5,786 KB	mp3	C:\Program Files\Lime... 0 / 0	0	0	0
TheQuickFixKills_Pick_Your_Poison	4,178 KB	mp3	C:\Program Files\Lime... 0 / 0	0	0	0

Name	Length	Bitrate
great speeches - football - vince lombardi speech.mp3	2:17	128
Great Speeches - Malcom X - Black power.mp3	1:16	128
John_Vanderslice_Amitriptyline.mp3	3:44	211
Alfie_Zappacosta_Start_Again.mp3	3:43	160



Εφαρμογές Συστημάτων P2P

B) Διαδικτυακή Τηλεφωνία (Internet Telephony)

– Skype



- Δωρεάν συνομιλία με άλλους χρήστες οπουδήποτε στον κόσμο.
- δημιουργήθηκε από τους ιδρυτές του εργαλείου ανταλλαγής αρχείων Kazaa)
- Η Αρχιτεκτονική Π/Π χρησιμοποιείται για να δρομολογεί έξυπνα τα πακέτα φωνής μεταξύ των διάφορων Πελατών

Γ) Διαδικτυακά Παιχνίδια (Online Gaming)

Δ) Πάταξη του Spam (π.χ. SpamNet)

E) Instant Messaging (e.g. IRC, MSN & Yahoo Msgers),

Z) Content Distribution Networks , P2P Web Caching,

H) Application-Level Multicast (e.g. Narada)



Classification of P2P Systems

(Based on the Degree of Centralization)

A) Centralized Approaches

- Centralized Indexes e.g. Napster,

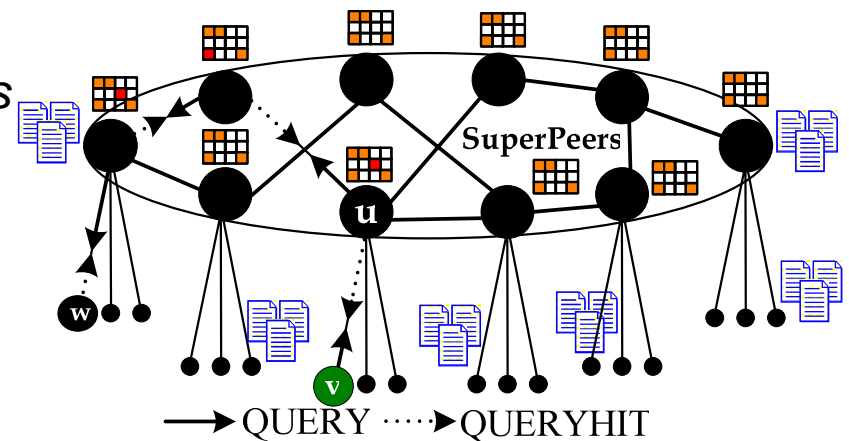
B) Purely Distributed Approaches

- Each node has only local knowledge.
- e.g. Gnutella, Fasttrack (Kazaa)

C) Hybrid Approaches

- One or more peers have partial indexes of the contents of others.

e.g. Limewire's Ultrapeers (Superpeers):
Automatically chosen by the system based on their capacities (storage, bandwidth, etc.) and availability (connection time)



Classification of P2P Systems

(Based on the **Network Structure**)

Based on how nodes are linked to each other we have:

A) Unstructured P2P Systems

Neighbors are selected arbitrarily e.g. Gnutella, Fasttrack

Pros: + *Easy to setup!*

Cons: - *Topology is not optimized – expensive search*
- *Appropriate when searching for popular items.*

B) Structured P2P Systems – Distributed HashTables (DHTs)

- *Nodes and Content are hashed to some ID space*
- *Each nodeid maintains links to a few well chosen other nodeids*
- *Each contentID is assigned to the closest nodeID*

e.g. Chord, Pastry, Tapestry, CAN, etc. (put(contentID, content))

Pros: + *Optimized Topology, search: $O(\log n)$*

+ *Appropriate when searching for a specific item*

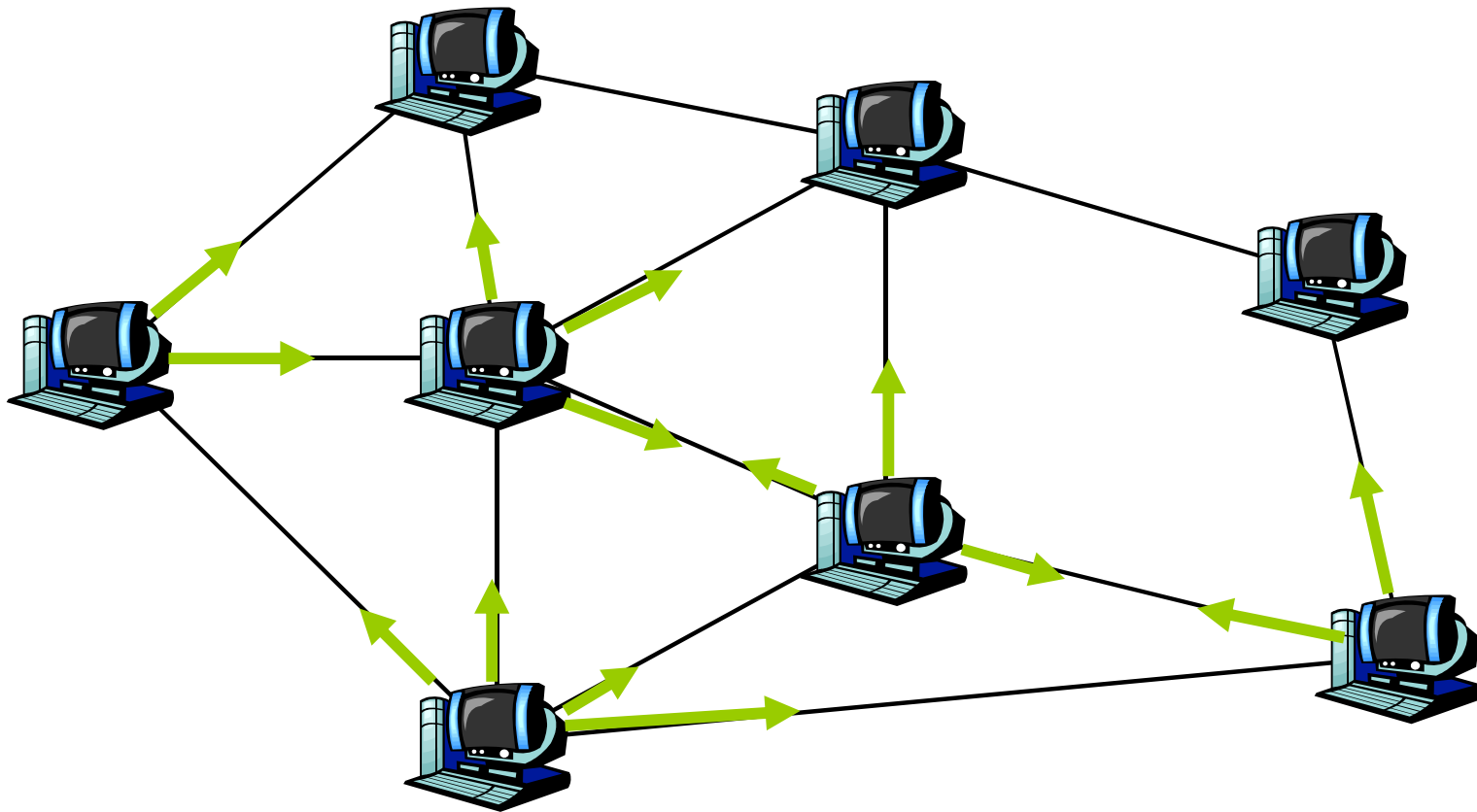
Cons: - *Churn, system does not converge to a steady state*



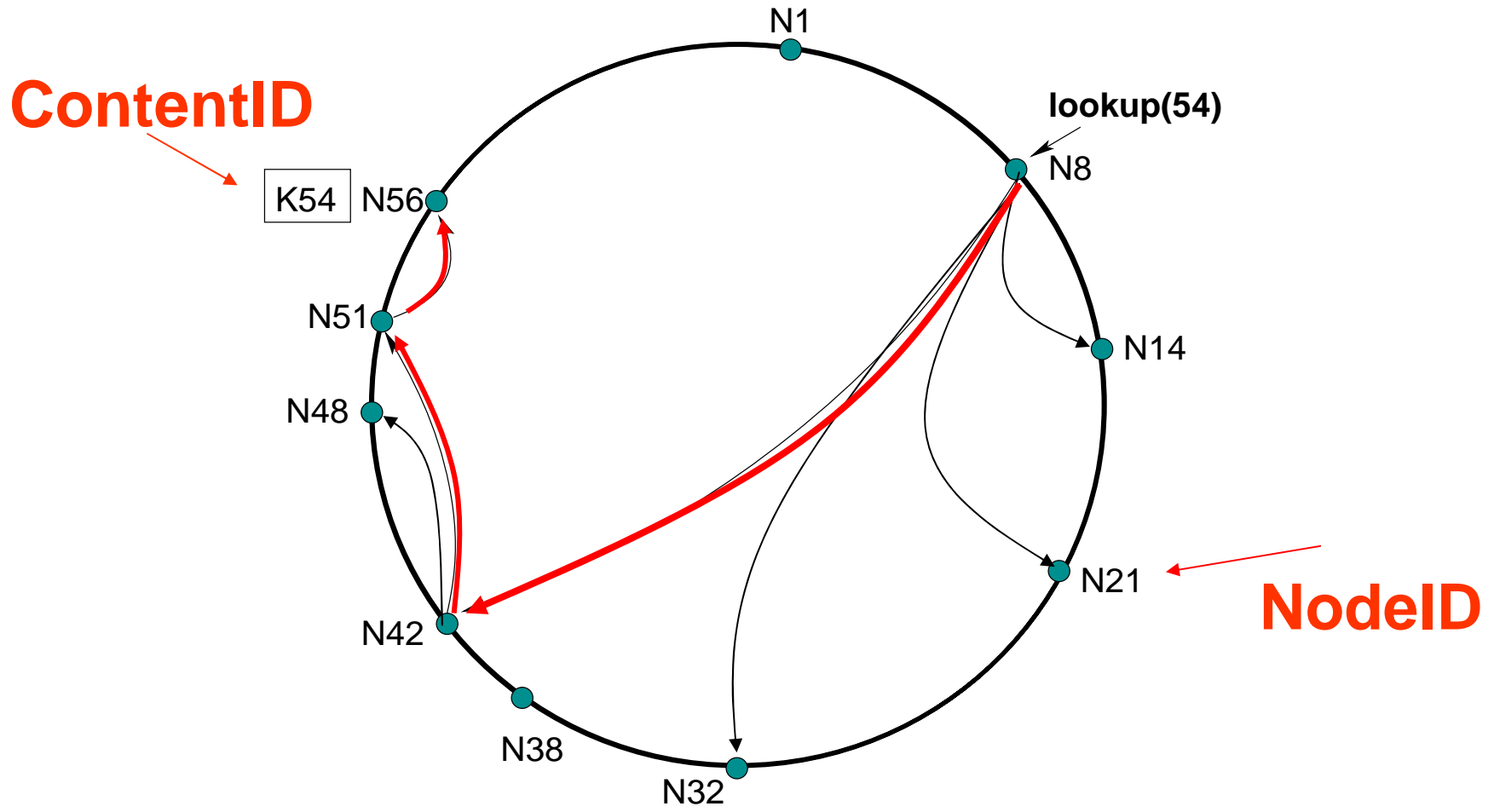
Classification of P2P Systems

(Based on the Network Structure)

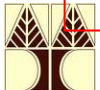
Search in an **Unstructured P2P System**: Flooding



Search in a *Structured P2P System*: Route to closest Hash



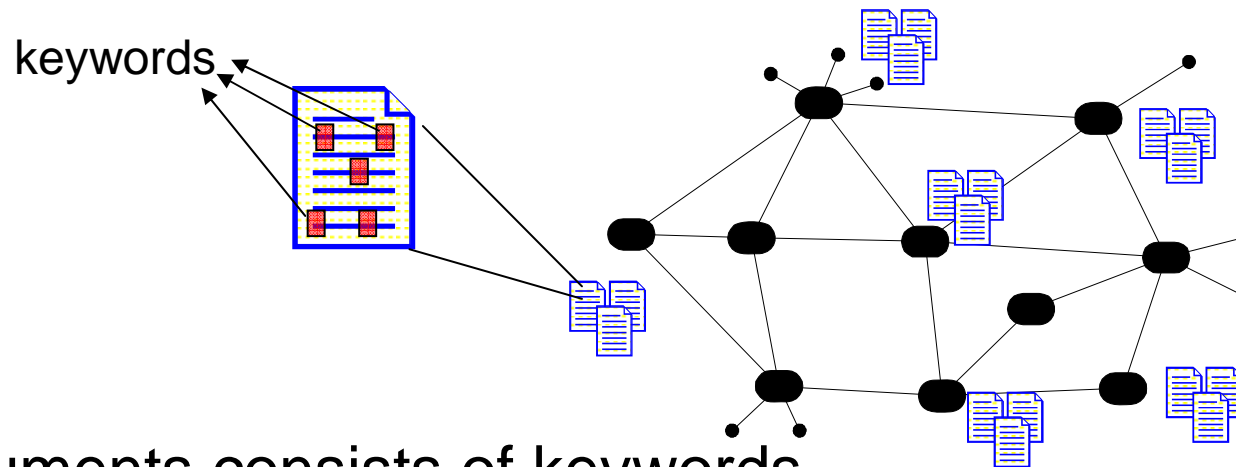
Μέρος 2: Information Retrieval in P2P Systems



Information Retrieval in P2P

Problem:

“How to efficiently retrieve Information in P2P systems where each node shares a collection of documents?”



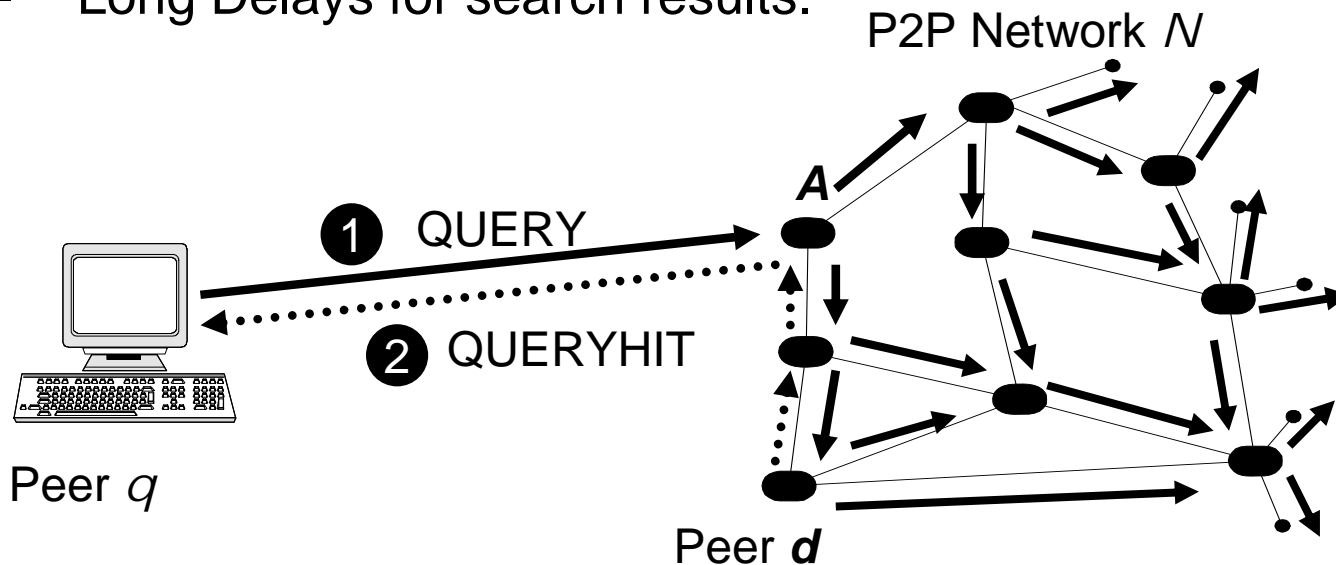
- Documents consists of keywords.
- We are not only interested in “Rem.mp3” but in Google-like keyword searches: “annual growth report cyprus”
- **“Information retrieval (IR) is the science of searching for information in documents (text, sound, images or data).**



Search Technique 1 - BFS

1. Breadth-First Search (Gnutella)

- **Idea:** Each Query Message is propagated along all outgoing links of a peer using TTL (time-to-live).
- TTL is decremented on each forward until it becomes 0
- Technique for I.R in P2P systems such as Gnutella.
- **Highlights**
 - The physical network is overloaded (even with TTL=7)
 - Long Delays for search results.

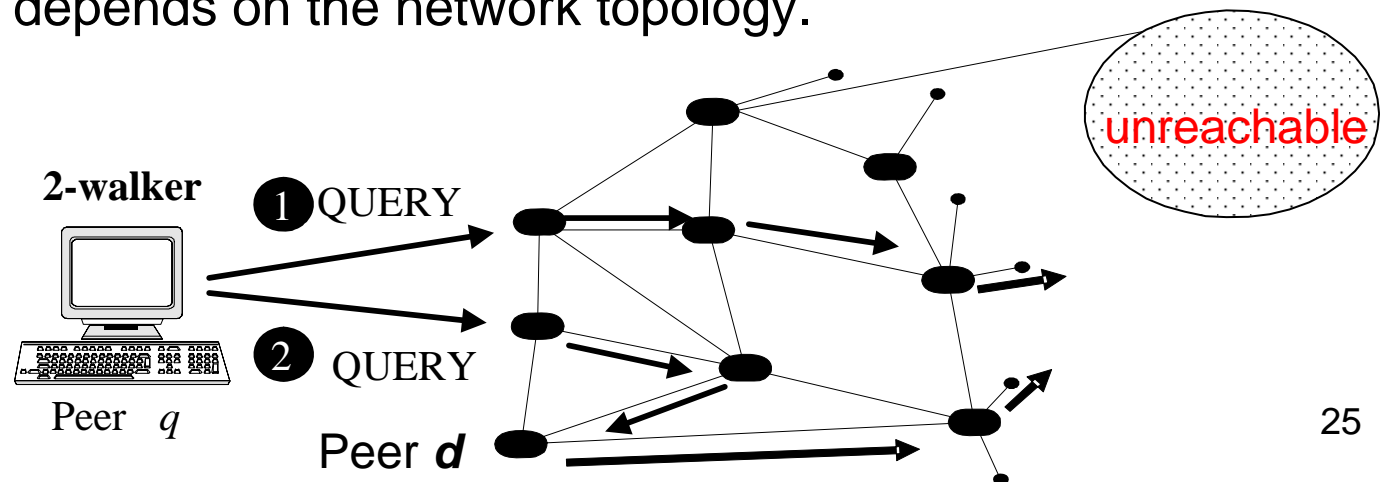


Search Technique 3 - RANDOM

3. Searching Using Random Walkers

[Q. Lv et al P. Cao, E. Cohen, K. Li, and S. Shenker. ICS2002]

- **Idea:** Each Query Message is forwarded to 1 neighbor
- With k walkers after T steps we reach \approx the same nodes as 1 walker after kT steps. (They use 16-64 walkers)
- **Highlights**
 - Network Traffic reduced (from BFS) by 2 orders of magnitudes
 - Increases the user-perceived delay (from 2-6 hops to 4-15 hops)
 - This algorithm is probabilistic and the likelihood to locate the objects depends on the network topology.

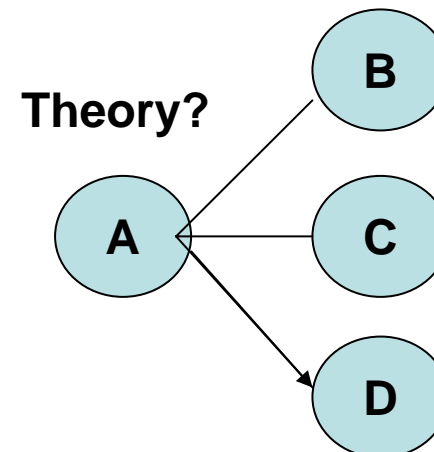


Search Technique 4 - LI

5. Searching using Local Indices [Arturo Crespo and Hector Garcia-Molina, ICDCS 2002.]

- **Idea: Create indices which contain “statistics” that reveal the “direction” towards the documents.**
- **Highlights**
 - Not Scalable, Expensive Routing Updates but better than replicating data indexes.
 - Assumes a static environment.

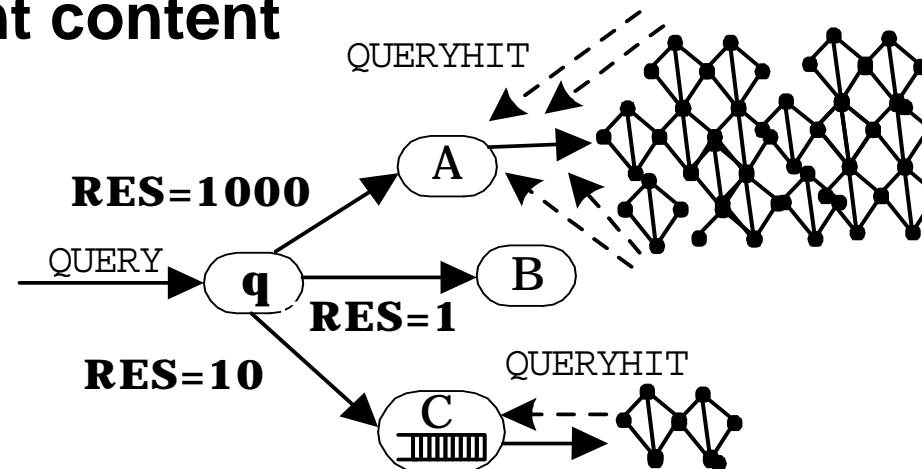
@A	Number of Documents	Database Related	Network Related	Theory Related
A	300	20	80	0
B	100	20	0	10
C	1000	0	300	0
D	200	100	0	150



Search Technique 5 - >RES

6. Directed BFS and the >RES Heuristic

- >RES: The Most Results Heuristic.
- **Idea:** Forward the query to your neighbors based on aggregate statistics (e.g. num of results a peer returned, shortest queue, shortest response time)
- **>RES works well because:**
 - It captures stable/large network segments.
 - Potentially less overloaded peers
- **Drawback: >RES doesn't route queries to the most relevant content**



Search Techniques: Remark

- On 1st June we crawled the Gnutella P2P Network for 5 hours with 17 workstations.
- We analyzed 15,153,524 query messages.
- **Observation:** High locality of specific queries...similar to the WWW.
- *We try to exploit this property for more efficient searches*

#	Query	Occurrence	%	#	Query	Occurrence	%
1	divx avi	588,146	3,88%	11	divx	24,363	0,16%
2	spiderman avi	50,175	0,33%	12	spiderman	23,274	0,15%
3	p__ mpg	39,168	0,25%	13	xxx avi	22,408	0,14%
4	star wars avi	38,473	0,25%	14	capture the light	21,651	0,14%
5	avi	29,911	0,19%	15	buffy mpg	20,365	0,13%
6	s__ mpg	27,895	0,18%	16	g__ mpg	20,251	0,13%
7	Eminem	27,440	0,18%	17	buffy avi	19,874	0,13%
8	eminem mp3	25,693	0,16%	18	t__ mpg	19,492	0,12%
9	dvd avi	25,105	0,16%	19	seinfeld vivid	18,809	0,12%
10	b____	24,753	0,16%	20	xxx mpg	18,686	0,12%



Search Technique 6 - ISM

Intelligent Search Mechanism (ISM)

[CIKM'02, IS'05]

Query	GUID	Connection & Hits	Timestamp
Athens 2004 Olympics	G439ID	(peer1,20), (peer4,50),...	100002222
VLDB Canada 2004	F549QL	(peer2,10)	100065652
***	***	***	***
Florida storm	PN329D	NULL	100022453

|L|-dim space: {athens,2004,olympics,vldb,canada,florida, storm}
e.g. If q ="athens olympics" \Rightarrow q (vector of q) = [1,0,1,0,0,0,0]

b) Cosine Similarity – The Similarity Function

$$sim(q, q_i) = cos(q, q_i) = \frac{\sum(\vec{q} * \vec{q}_i)}{\sqrt{\sum(\vec{q})^2} * \sqrt{\sum(\vec{q}_i)^2}}$$

c) RelevanceRank – Ranking Neighbors by similarity

$$RR(peer_i, q) = \sum_{q_j = \text{"Queries answered by } peer_i\text{"}} sim(q_j, q)^\alpha * results(q_j)$$



Search Technique 7 - PlanetP

4. Using Randomized Gossiping to Replicate Global State [F.M Cuenca-Acuna, Thu D. Nguyen HPDC-12]

- **Idea: Advertise a summary of a Peer's content to the neighborhood (using Bloom Filters).**

- **Bloom Filters are used for Membership Queries**

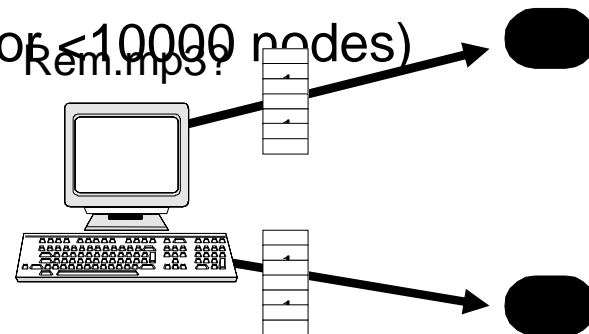
PeerA: Does PeerB maintain “rem.mp3”?

PeerA: `lookup_locally(BloomfilterB, “rem.mp3”)`

PeerA: Answer NO (definite) or MAYBE (False Positives are possible)

- **Highlights**

- No Data Replication Required
- Not Scalable (Shown to work well for <10000 nodes)



Συμπαγείς Κατάλογοι (Compact Membership Directories - Bloom Filters)

Πρόβλημα

- Έστω ότι ο **A** θέλει να στείλει στον **B** μία συλλογή $D=\{d_1, d_2, \dots, d_n\}$ από n στοιχεία.
- Αυτή η συλλογή μπορεί να είναι για παράδειγμα τα ονόματα όλων των αρχείων που έχει ο **A**, έτσι ώστε να ξέρει και ο **B** τι αρχεία έχει ο **A**.
- Αντί να στείλει όλη την λίστα ο **A** στον **B** (το οποίο είναι ακριβό), ο **A** μπορεί να στείλει ένα συμπαγή κατάλογο (Bloom Filter)

Bloom Filter: είναι ένα διάνυσμα **V** (**vector**) από **m bits**, στο οποίο μπορεί να συμπεριληφθεί το περιεχόμενο της συλλογής **D**, με τον ακόλουθο τρόπο:



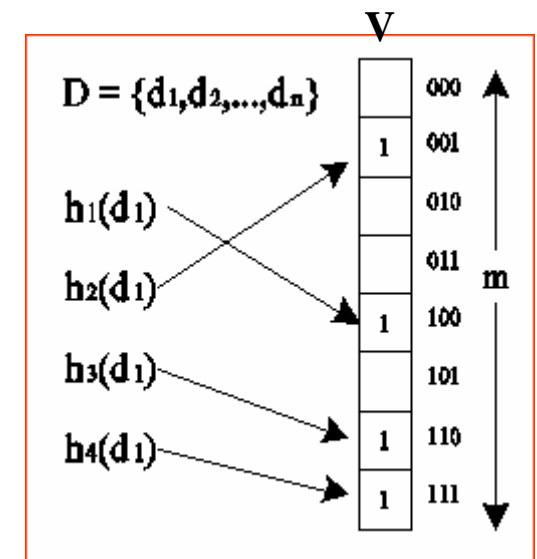
Συμπαγείς Κατάλογοι (Compact Membership Directories - Bloom Filters)

Bloom Filter

1. Θέτουμε σε ένα διάνυσμα V μεγέθους m όλα τα στοιχεία ίσο με «0».
2. Χρησιμοποιώντας K διαφορετικά hash functions, κάνουμε hash όλα τα n στοιχεία της συλλογής $D=\{d_1,d_2,\dots,d_n\}$.
3. Όπου γίνεται hash θέτουμε την τιμή του πίνακα ίσο με «1»
4. Ο A αποστέλλει το V (όχι το D) στον B .
5. Αν ο B θέλει να δει αν ο A έχει ένα αρχείο «test.c», τότε χρησιμοποιεί τα ίδια K hash functions πάνω στο όνομα του αρχείου και ελέγχει αν όλα τα στοιχεία του διανύσματος V είναι ίσο με «1».

Αν όχι τότε ο A σίγουρα δεν έχει το αρχείο,
Ειδ' αλλιώς πολύ πιθανόν ο A το έχει!

- Για μεγάλες τιμές του m και K τα bloom filters δουλεύουν πολύ αποδοτικά
- Έχουν εφαρμογές σε καταναμημένα συστήματα (web caches, p2p systems, etc)



Search Techniques: Remark

- All techniques discussed so far do not take into account **anonymity**.
- While A downloads (or queries) through B, B knows exactly what we are looking for
- Freenet
 - Designed to provide anonymity.
 - ...although it is not widely used today.



Search Technique 8 - Freenet

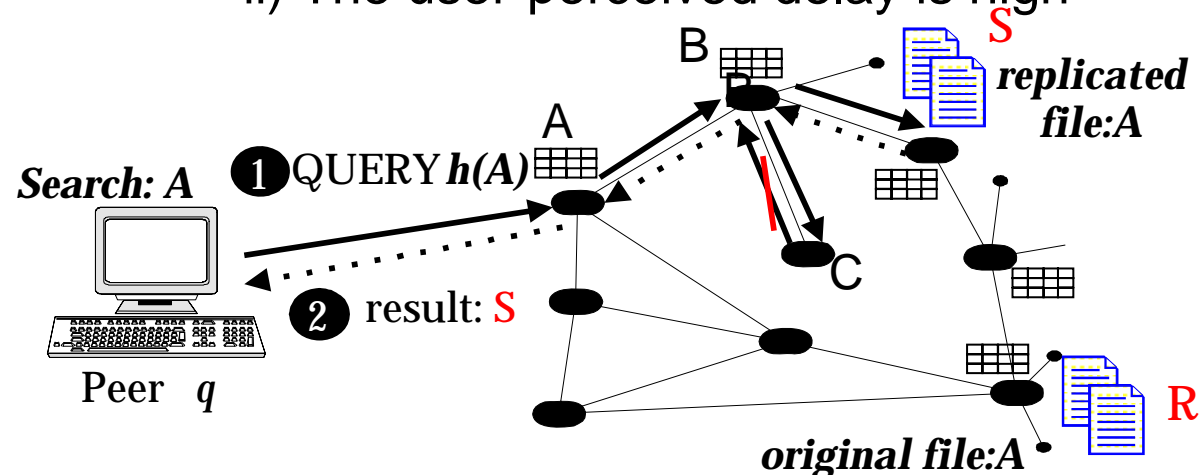
7. Depth-First-Search and Freenet

[I. Clarke O. Sandberg, B. Wiley, and T.W. Hong, LNCS 2009]

Idea: Looking for A? Search for hash(A). Queries are routed based on the “key closeness” in a DFS manner.

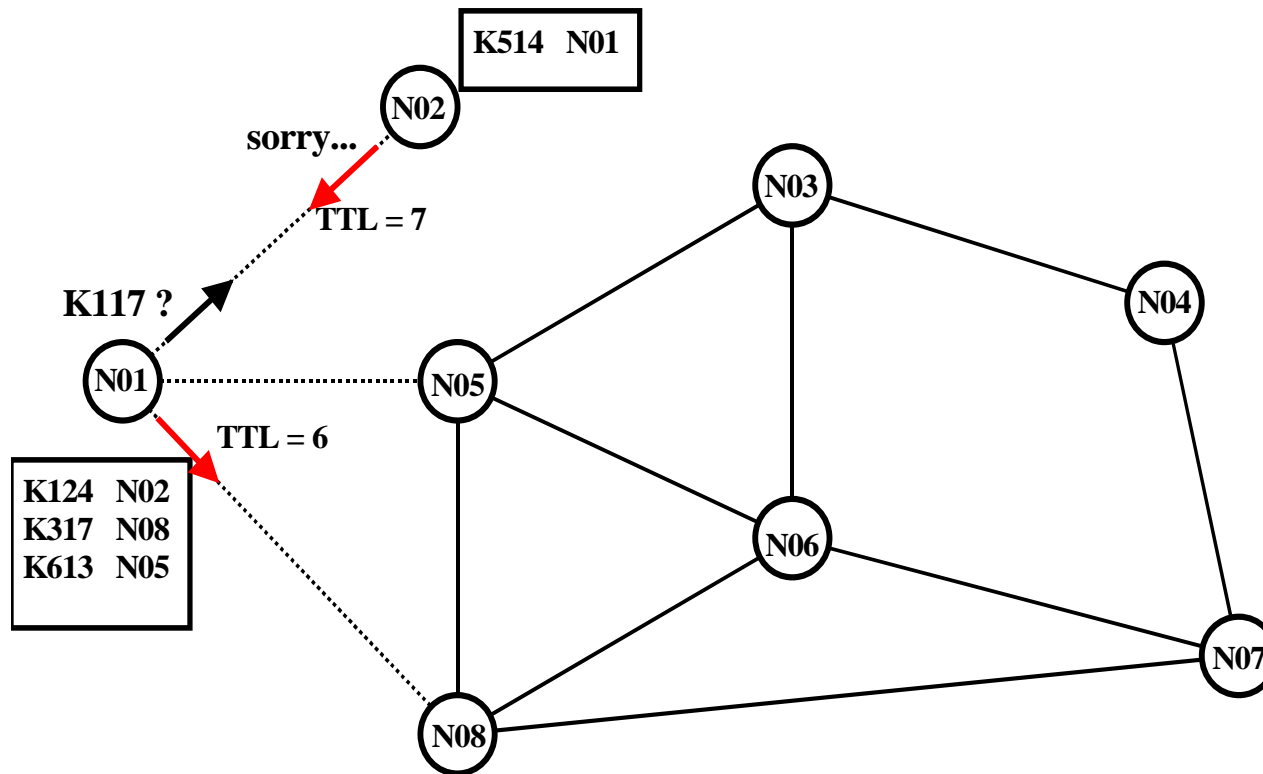
Highlights:

- Uses caching of key/object for future requests.
- Data Replication along the QueryHit path provides **availability**
- Anonymity of Searcher and Publisher.
- Drawbacks: i) **Searches ONLY based on Object Identifier.**
ii) The user-perceived delay is high

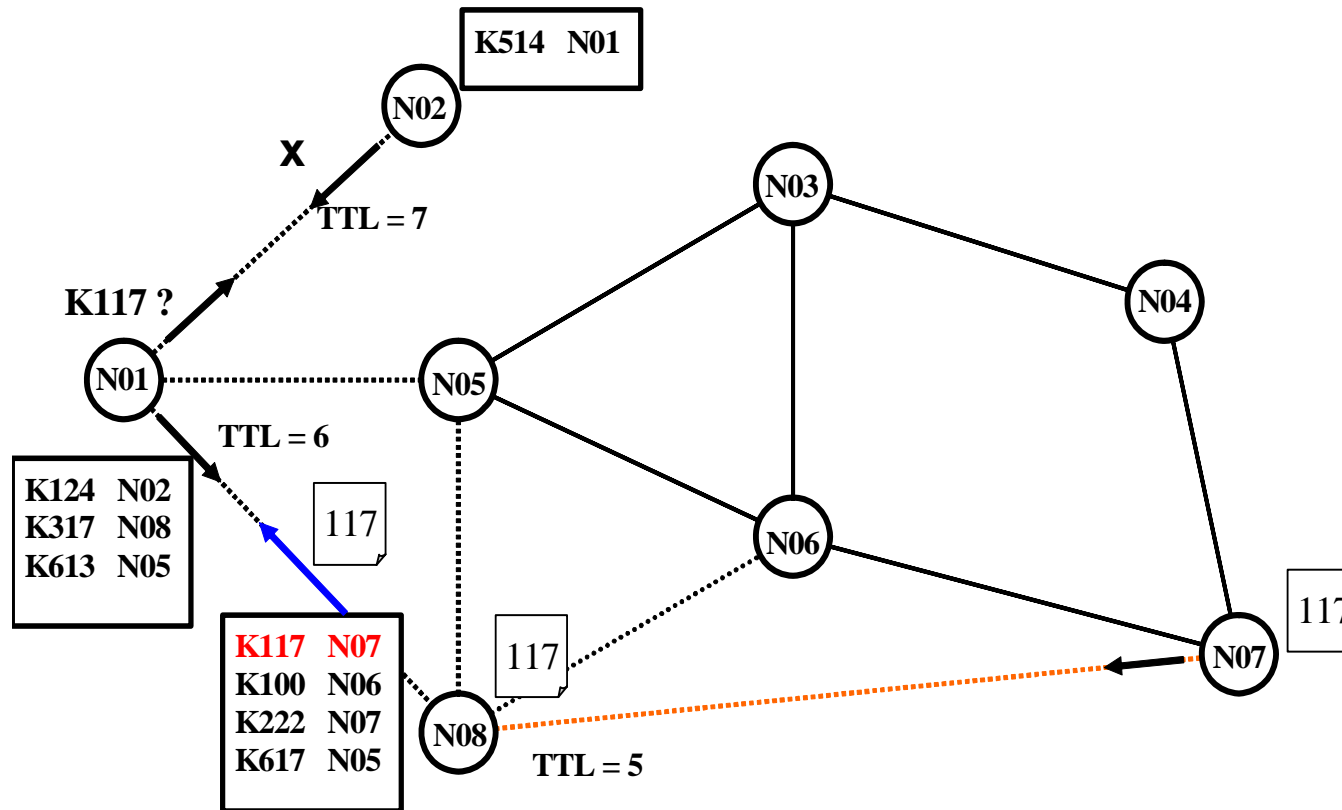


Search Technique 8 - Freenet

Search: Queries are routed based on the “key closeness” in a DFS manner.



Search Technique 8 - Freenet



Item Found:

- The key K117 is added to the key table of N08, while the answer is routed back.
- Document 117 is replicated along the Queryhit path



Search Technique 8 - Freenet

- pros:
 - complete decentralization
 - fault tolerance/self-organization
 - anonymity
 - scalability (to some degree)
- cons:
 - questionable efficiency & performance
 - rare keys disappear from the system



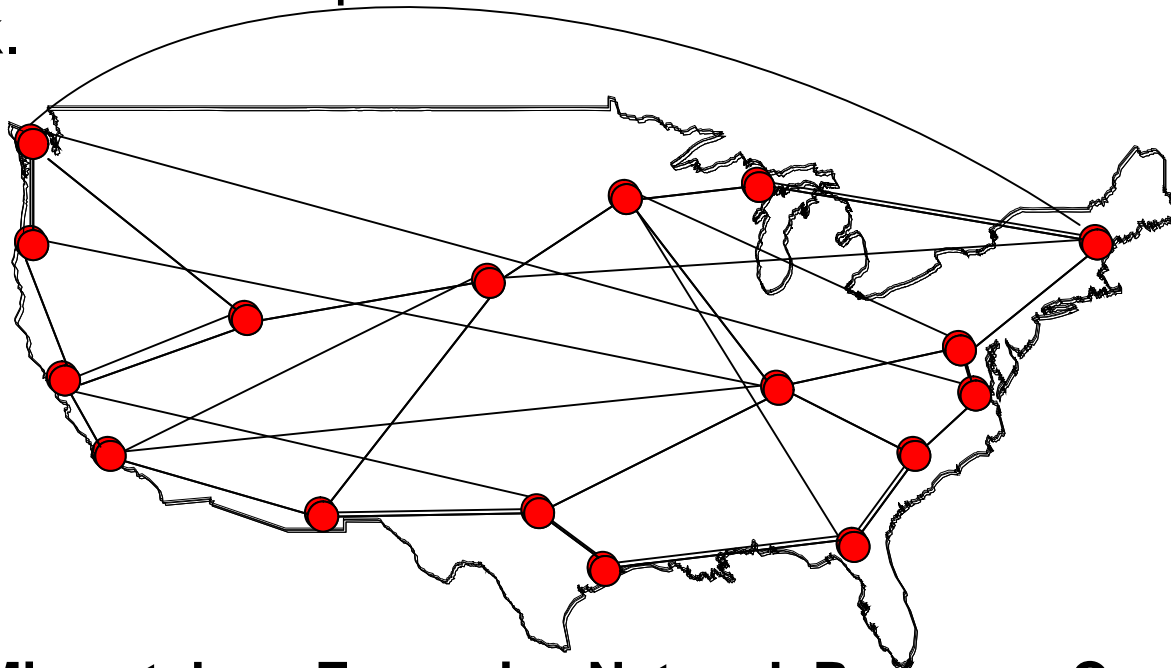
Network-Efficient Topologies

- P2P Networks are usually network-agnostic. Therefore they don't take into regards the **efficiency of the overlay topology**.

Example

- Assume that we have a US Newspaper Agency that makes its news scripts available in a P2P distribution network.

The Agency
Topology
Network
Network

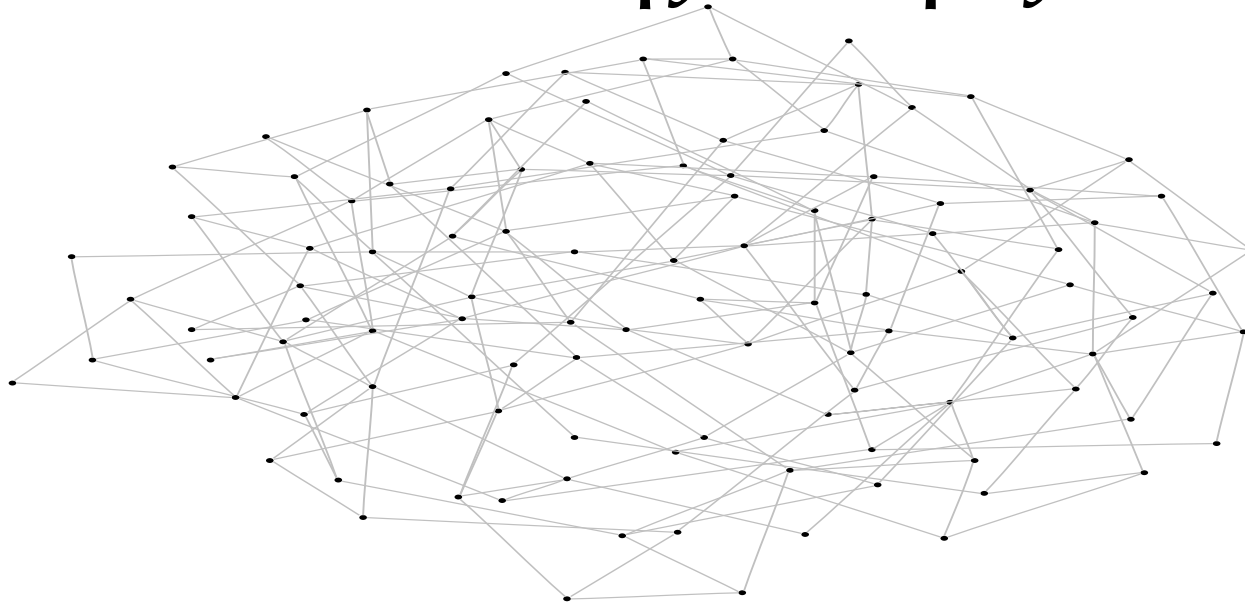


Network Mismatch => Excessive Network Resource Consumption



“Peer-to-Peer Systems” / “Συστήματα Ομοτίμων”

**Δημήτρης Ζείναλιπούρ
Επισκέπτης Λέκτορας**



<http://www.cs.ucy.ac.cy/~dzeina/>

