

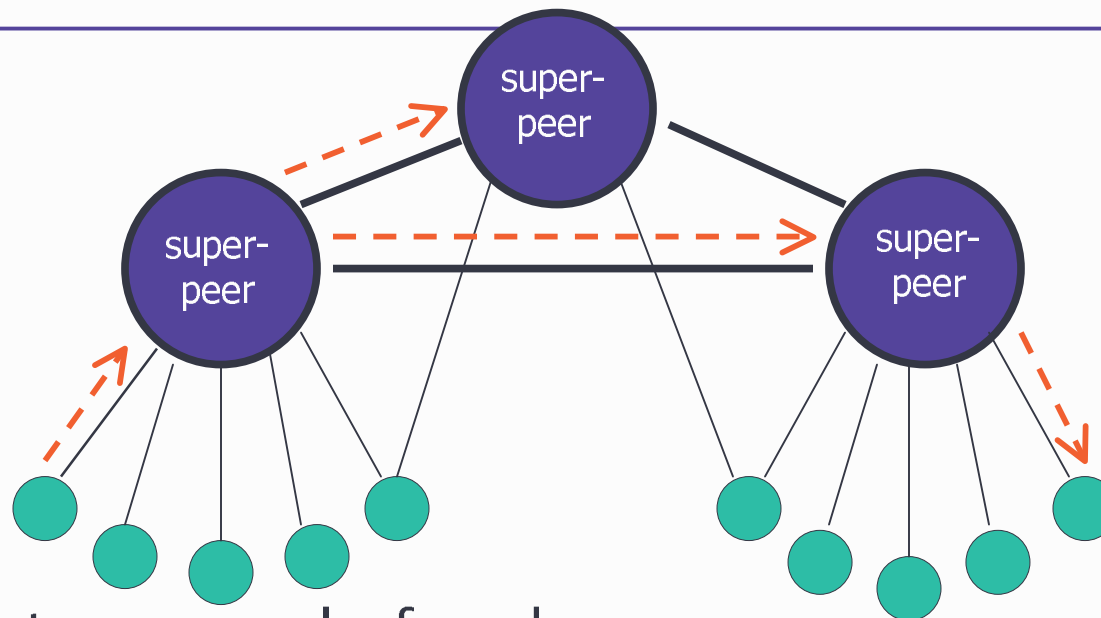
The Case for Service Provider Deployment of Super-Peers in Peer-to-Peer Networks

Sumeet Singh, Sriram Ramabhadran, Florin Baboescu, Alex C. Snoeren

University of California, San Diego

presented by
sumeet singh

A Typical P2P Topology



- a) Most peers are leaf-node
- b) Super-peers are explicit entry points for leaf-nodes.
- c) All protocol control traffic is routed via the super-peers.
- d) **Being a Super-Peer has incentives & disincentives:**
 - Resource Overhead (BAD)
 - High Out degree = Better Query Response Rate (GOOD)
 - A Tension Exists: *To be or, not to be*

The Opportunity!

- Service Providers can leverage the unique capabilities of super-peers to transparently influence individual node behavior without the explicit knowledge or consent of the user.

Talk Organization

- We address two separate issues
 1. Incentives for
 - a) service providers to deploy super-peers.
 - ISP's
 - Content Distributors / Publishers
 - b) end-users to utilize these super-peers
 2. Technical issues in implementing a super-peer
 - Topic Based Search Optimization
 - Preliminary Evaluation Results

P2P Traffic Characterization

- A Large volume of ISP backbone traffic is P2P.
- There are various popular P2P protocols, and not all are open.
- P2P applications **do not exploit topological locality**.
- New generation of P2P applications **use any available port**.
- Hard to directly apply traditional Traffic Engineering
 - For example, Proxy based caching for web cannot be applied, as in P2P traffic there is no standard protocol, no use of standard port numbers, and the object sizes are much larger

P2P Traffic Engineering (P2PTE)

- What does this mean to an ISP?
 - Better utilization of the ISP network.
 - Better performance for the end-user; possibly driving customers' selection of an ISP.
- By providing a policy-driven framework in the super-peers, ISPs can adequately control the control traffic at the application level.

P2PTE – Methodology

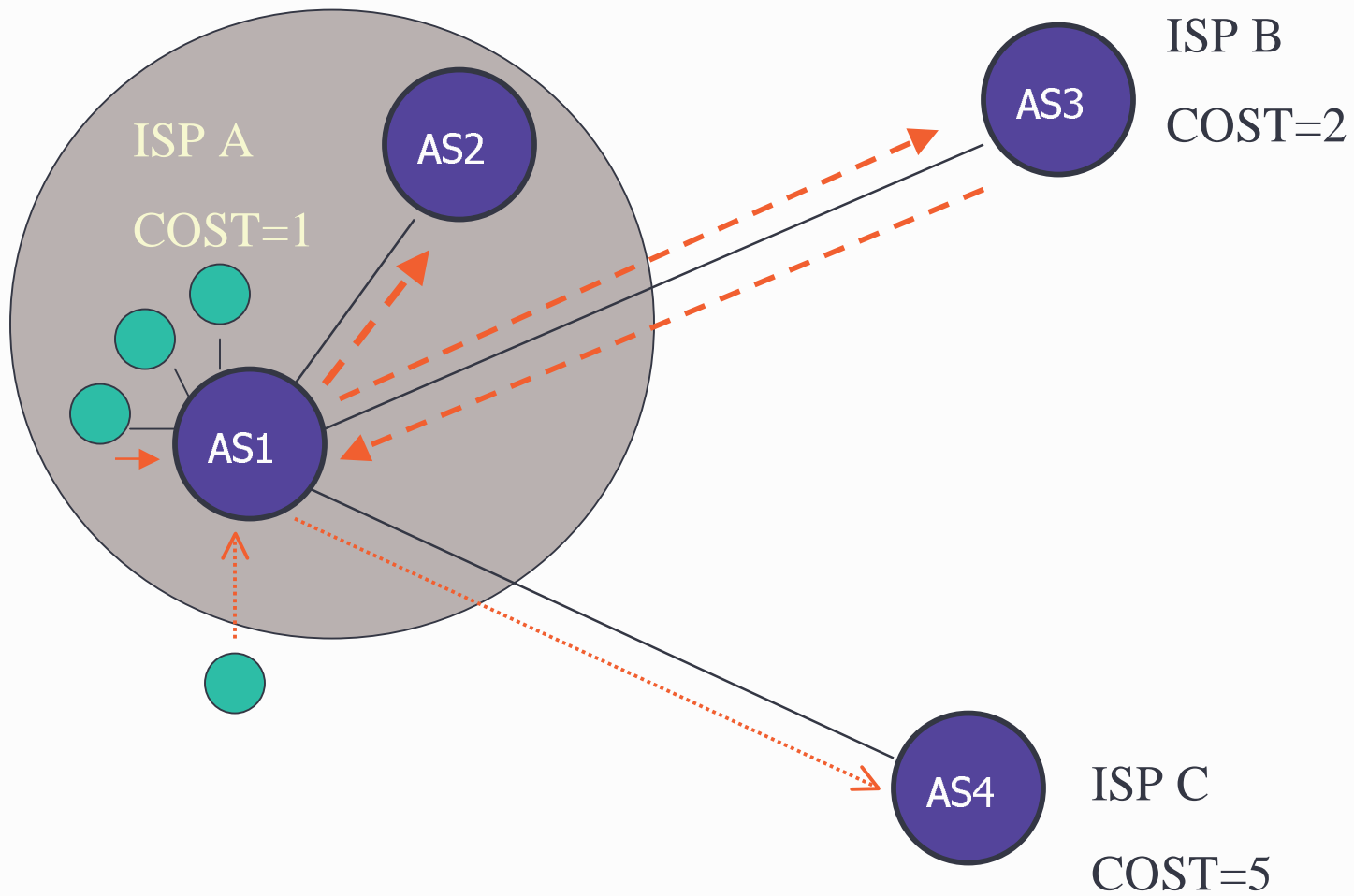
Method:

- Preferentially forward the queries to networks with LOW cost.

Goal:

- To achieve a significant number of query responses (if possible within the networks with low cost).

P2PTE / Sample Policy

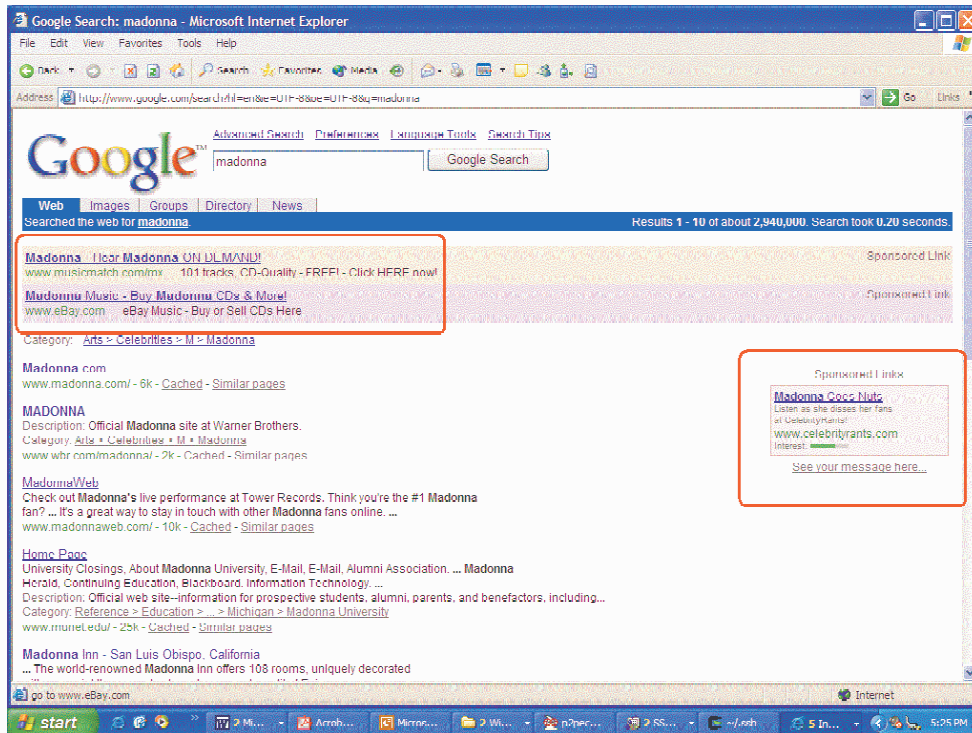


Selective Content Distribution

What is it?

- Super-Peers can provide links to the push content as part of a response to queries made for similar items.
- This information is introduced transparently in the network, the clients do not need to upgrade their software or subscribe to an additional service.
- Content Distributors can charge publishers for providing Push Content via the Super-Peers.

Selective Content Distribution



It is part of
our daily
WWW
experience

Content Distributors can provide a similar service in P2P networks by utilizing a push-based model.

End-user perspective

- The Key Question:
will an end-user subscribe to a model in which a service provider provides and potentially monitors, controls, modifies, or even sells connectivity to a peer-to-peer network?

Incentives for the end-user

- A super-peer architecture can provide improved search capabilities
- Service-Providers can implement policies that provide an enhanced level of download performance.
- Super-peers can provide users with novel aggregated views of the content in the network
- Super-peers can act as application-level bridges between different peer-to-peer protocols

Providing Efficient Search Services

- Traditional Methodology: cache/replicate actual data to improve search performance.
- We propose an alternate strategy; by exploiting the following observations:
 - There is likely to be significant locality in the type of content requested by the individual peers, as well as the type of content that the peer shares.
 - Super-peers are in a unique position to generate aggregated views of the content in the network.
 - **Most Queries are for Hay not Needles.** Thus it is more of a problem of trying to find the right haystack, not the needle in the haystack.

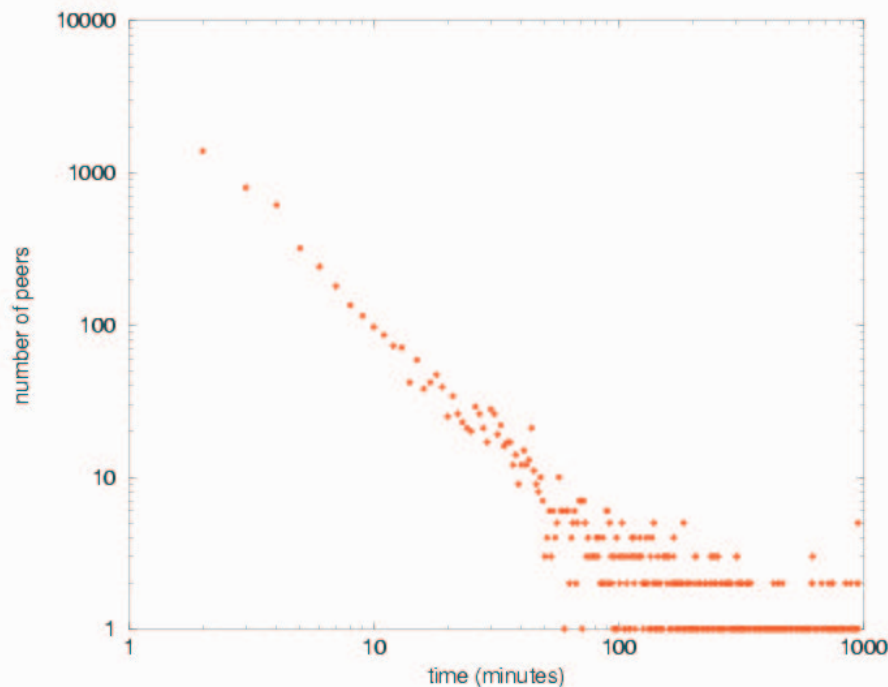
Topic-Based Search Optimization

- We Propose:
Cache pointers to source of content – as opposed to content itself.
- Difficult to track each individual object, track topics instead.
- Topics are created dynamically by analyzing the data in Query and the Query Response messages.
- Example Meta Data, contained in QueryResponse:
For Music File: Title, Artist, Album, Release Year, Category

Topic-Based Search Optimization : An Example

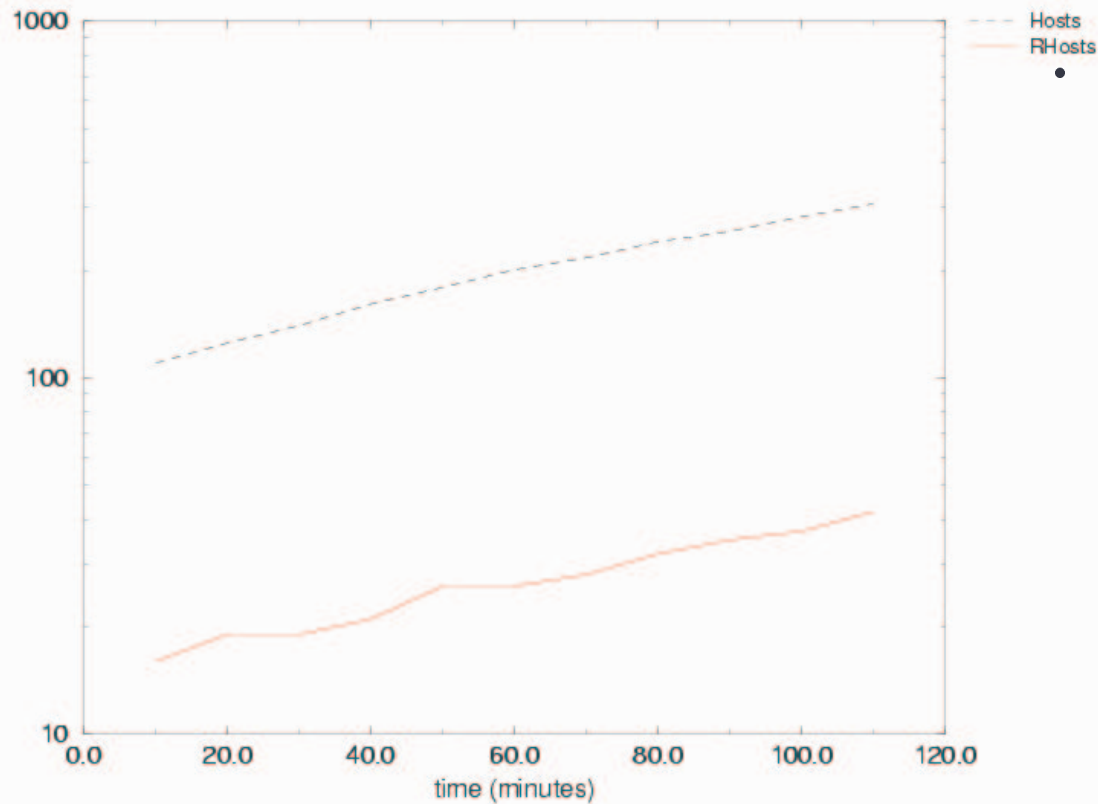
- Let us consider two queries
 - A query for “Pink Floyd”
 - A query for “The Wall”
- If both these queries were being routed via the Super-Peer,
 - The super peer could route the queries to peers with interest in the topic “Pink Floyd”
 - peers can benefit from queries by my themselves earlier, as well as other peers.

Evaluation– Duration of Peer Relationships



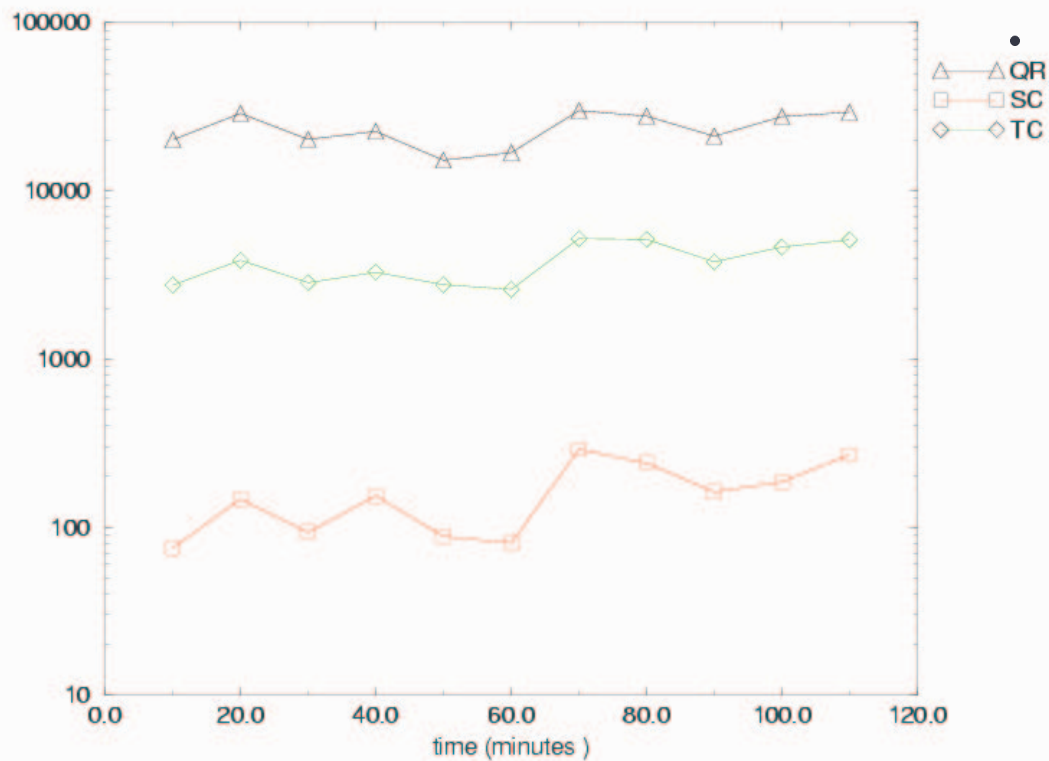
- Super-Peer activity for 960 minutes.
 - 5563 unique connections
 - 500 > 1 hour
 - 2000 between 2–5 minutes
 - 3000 between 5–60 minutes

Evaluation – Topic Query Response Rates



- Approximately 15% of the peers respond positively when sent a Topical Query based on their search history.

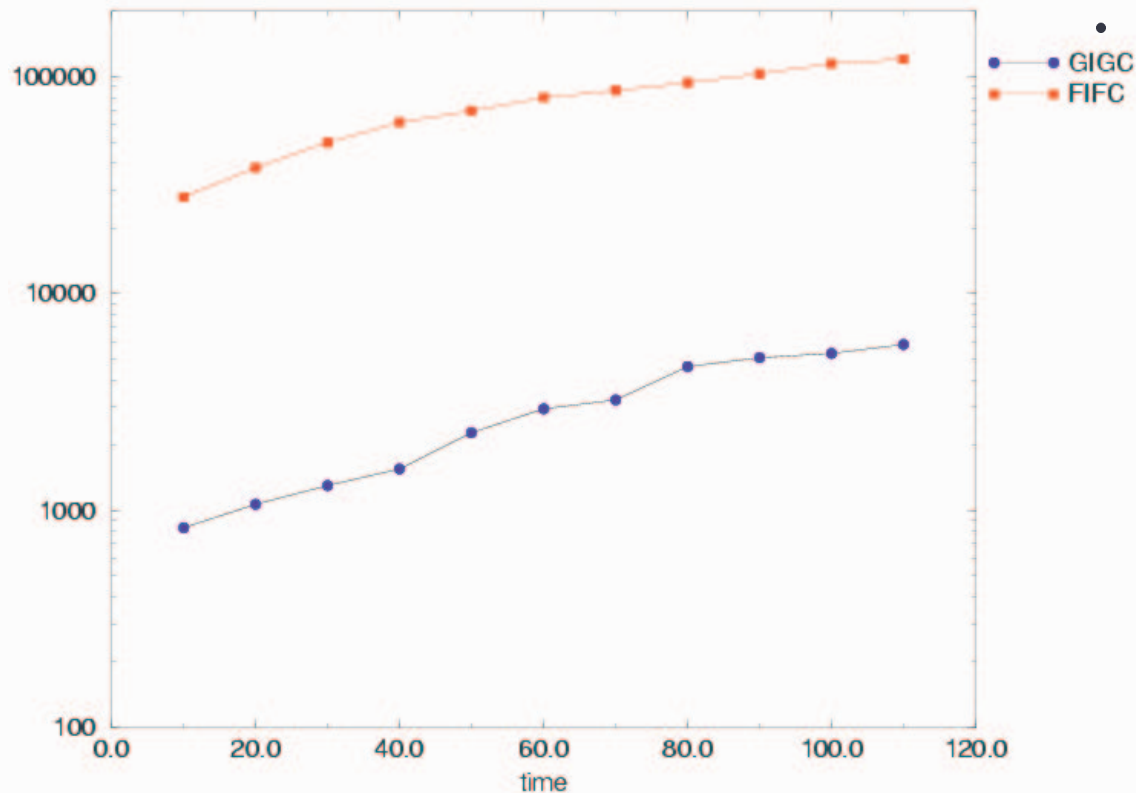
Evaluation – Expected Hit Rates



- 25% of the Queries match at least 1 Topic of Interest.

We only need for 15% of these Queries to return +ve results to perform as well as caching actual objects.

Evaluation- Expected Cache Size



- Expected Cache Size, is substantially larger (in order of gigabytes) if we cache the actual objects as opposed to maintaining pointers to the objects.

GIGC : Number of Active Groups

FIFC : Number of Objects in File Cache

Value Added Service Content Aggregation

Pre-cached Query Results for Topic: 70s Music

1970	abba	waterloo	waterloo	8
1970	bad company	guitar rock	feel like making love	12
1970	billy joel	the stranger	just the way you are	6
1970	bing crosby	christmas classics	frosty the snowman	8
1970	bing crosby	christmas classics	let it snow!	6
1970	black sabbath	masters of reality	sweat leaf	8
1970	bob marley	catch a fire	kinky reggae	16
1970	buggles	the age of plastic	video killed the radio star	22
1970	carly simon w/james taylor	no. cakes	mockingbird	8
1970	david allan coe & johnny cash	john r cash	cocaine carolina	6
1970	david bowie		changea	8
1970	elvis presley	amazing grace	put your hand in the hand	6
1970	gordon lightfoot	summertime dream	wreck of the edmund fitzgerald	6
1970	harry chapin	heads & tails	taxi	6
1970	hcl chocolate		you sexy thing	6
1970	janes taylor	gorilla	how sweet it is to be loved by you	40
1970	janes taylor	greatest hits	mexico	10
1970	jin croce	original oldies 60's & 70's	line in a bottle	6
1970	joe feat. mystikal & cnk	isi's rmx	stutter	10
1970	joni mitchell	ladies of the canyon	big yellow taxi	6
1970	marvin gaye	what's going on	mercy mercy me	16
1970	meatloaf	bat out of hell	paradisebythedashboardlights	8
1970	neil diaronc	greatest hits	cracklin rocc	6
1970	nino ceta	the godfather	godfather theme.mp3	6
1970	pink floyd	the wall	hey you	44
1970	pink floyd	the wall	is there anybody out there?	6
1970	pink floyd	the wall	run like hell	6
1970	rod steuart	storyteller (disc 3)	you're on my heart	6
1970	rolling stones		it's only rock 'n roll	8
1970	rolling stones	sticky fingers	can't you hear me knocking	6
1970	rolling stones	sticky fingers	dead flowers	6
1970	rolling stones	sticky fingers	wild horses	8
1970	stevie wonder	songs in the key of life	isn't she lovely	22
1970	the foundations	original oldies 60's & 70's	build me up buttercup	16
1970	the rolling stones	some girls	beast of burden	21
1970	various artists	from philly with love	me and mrs. jones	6

- The figure shows a list of 70s objects being shared in the P2P Network (Artist, Album, Song Name)
 - Users can query the Super-Peer for a topic to perform general searches similar to the one shown here ("70s music"), and chose to download an object from a catalog, as opposed to repeatedly searching for that might not be available.
 - Existing P2P Systems do not support general queries.