

Assisted Peer-to-Peer Search with Partial Indexing

IEEE INFOCOM 2005

Presented by C.L. Wang
2006/01/25

Outline

- Introduction
- Design Concept
 - Definitions
 - Overlay Initialization
 - Resolving Queries
- Performance
- Conclusions

Introduction

- P2P systems can be classified into two categories:
 - Structured P2P Overlays
 - Unstructured P2P Overlays

- This approach is opposite in that search is performed in the unstructured P2P overlay with the assistance of a structured index overlay.

Introduction

- Unstructured P2P Overlays
 - The search algorithm
 - Data is replicated by the querier upon a successful search.
- Structured P2P Overlays
 - Exploring locality in peer interests through interest-based clustering
 - Selective indexing of unpopular data

Features of the hybrid Overlay

- Peers communicate their interests via the index overlay and the search overlay is constructed based on peer interests.
- The interest-based registries can be consulted for hints about where to forward the query for a second try.
- Peers also identify the properties from their local data repository that are globally unpopular.

Design Concept

- Definitions
 - Properties (keywords)
 - Local interests
 - Popularity of data
 - Partial index
- Overlay Initialization
- Resolving Queries

Properties (keywords)

- Peers can uniquely identify the data items that they possess.
- A data item is also associated with a number of *properties* (keywords).

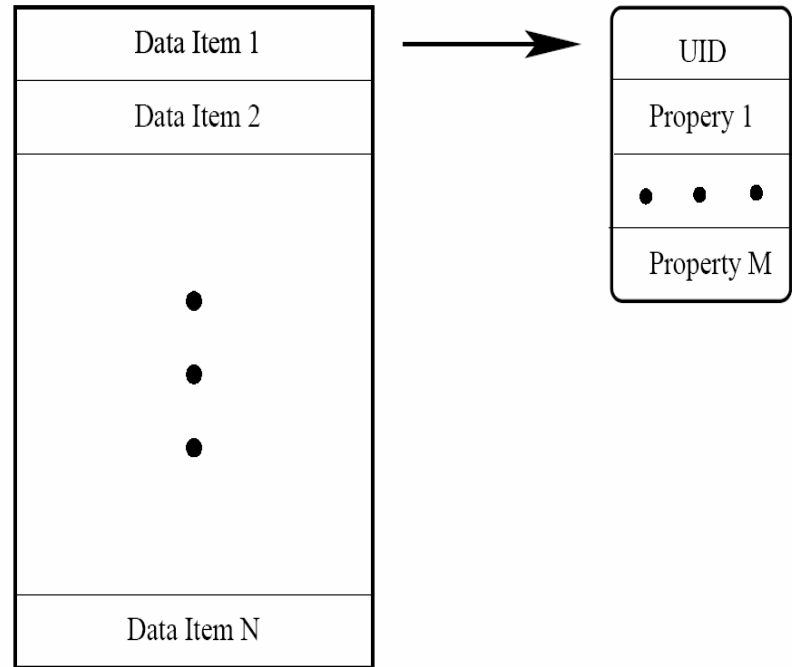


Fig. 1. Data possessed by a peer. Each data item contains one or more properties.

Local interests

- Merged properties of local data items, sorted according to the frequency of appearance.
- Top ranked properties become the local interests of the peer.

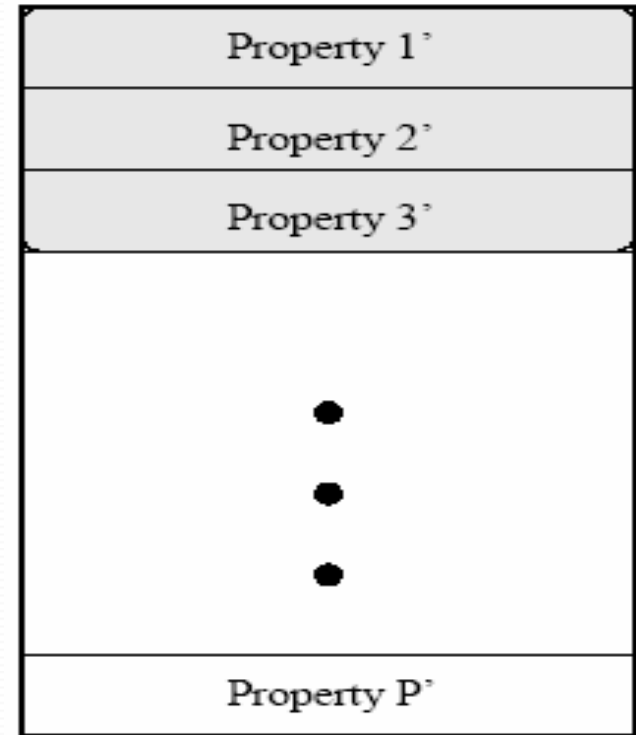


Fig. 2. shows the local interests of the peer in Fig. 1.

Popularity of data

- Property popularity can be determined from observing passing traffic.
- Data popularity is observed by individual peers.
- Those properties of local data that are seen the least frequently in passing queries are identified as “unpopular”.

- Merged properties of local data items, sorted according to the frequency of appearance in received queries.
- Bottom ranked properties become the locally-observed unpopular properties.

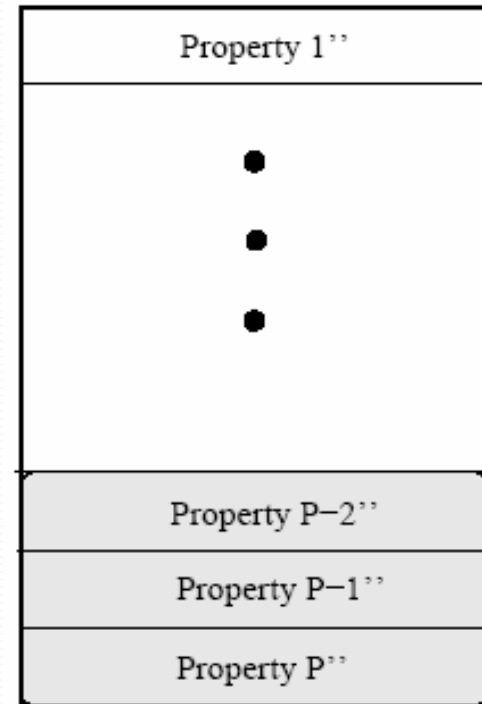


Fig. 3. shows locally-observed unpopular properties

Partial index

- We propose a **DHT-based** partial indexing scheme.
- The index only maintains information about the **top interests** of peers and **unpopular data**.
- Instead of indexing all data items owned by each peer, only a portion of them are registered with the index.

Purposes of partial index

- Help peers to join the index overlay of the similar interests
- Provide hints for second search
- Help improve the chances of finding unpopular data

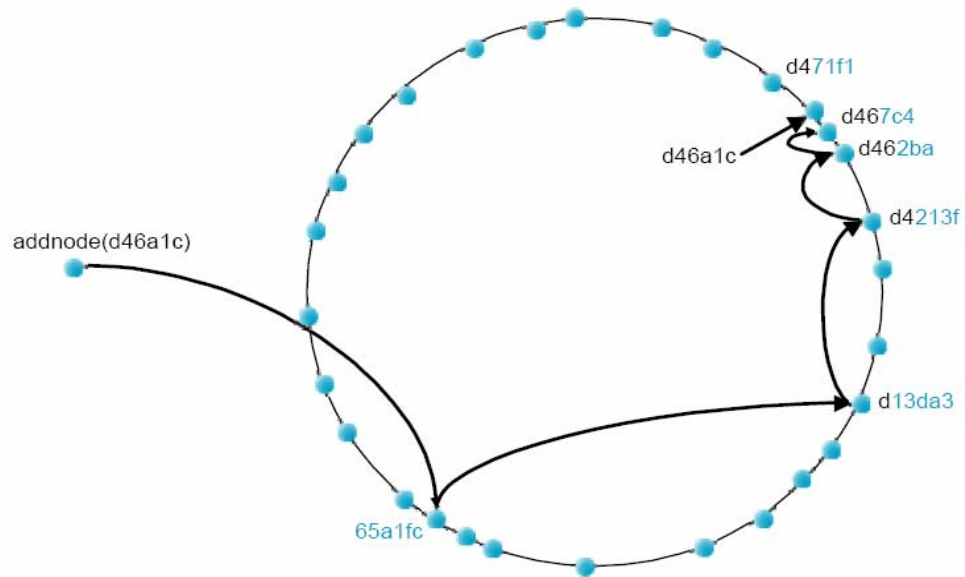
Overlay Initialization

- Join the index overlay
- Obtain the addresses of other peers with similar interests
- Construct search overlay

Join the index overlay

- The index overlay is constructed according to the DHT-based structured P2P overlay.

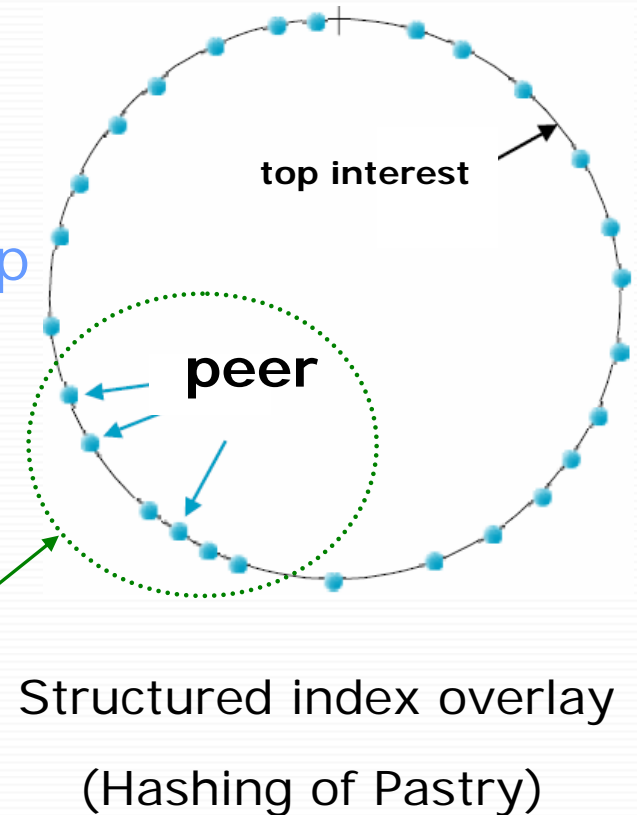
Peer Joining (Pastry)



Other peers with similar interests

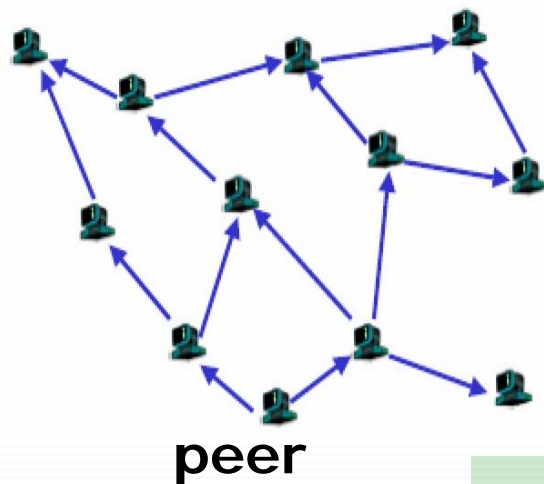
- A new peer can obtain the addresses of other peers with similar interests through a **lookup (key)** operation.
- Peers register their top interests through an **insert (key, value)** operation.

Peers with similar interests

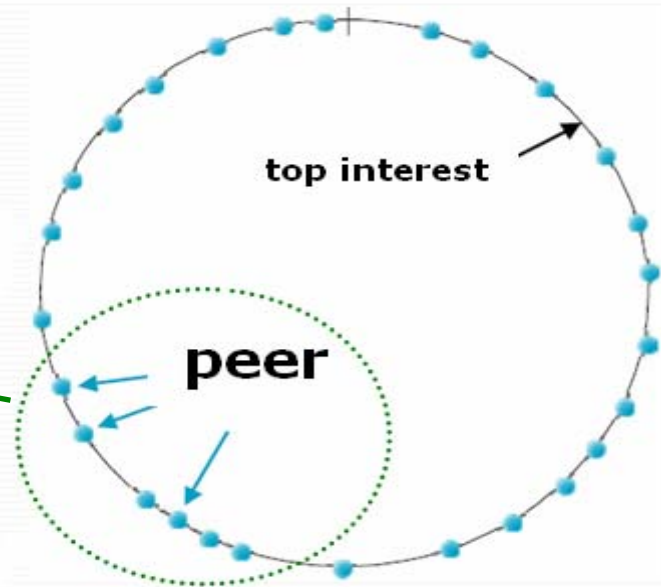


Construct search overlay

- It uses **flooding** for finding other peers and searching resources in the network.



Peers with similar interests



Resolving Queries

- Issued to the search overlay
- Seeking search guidance from the index overlay.

Issued to the search overlay

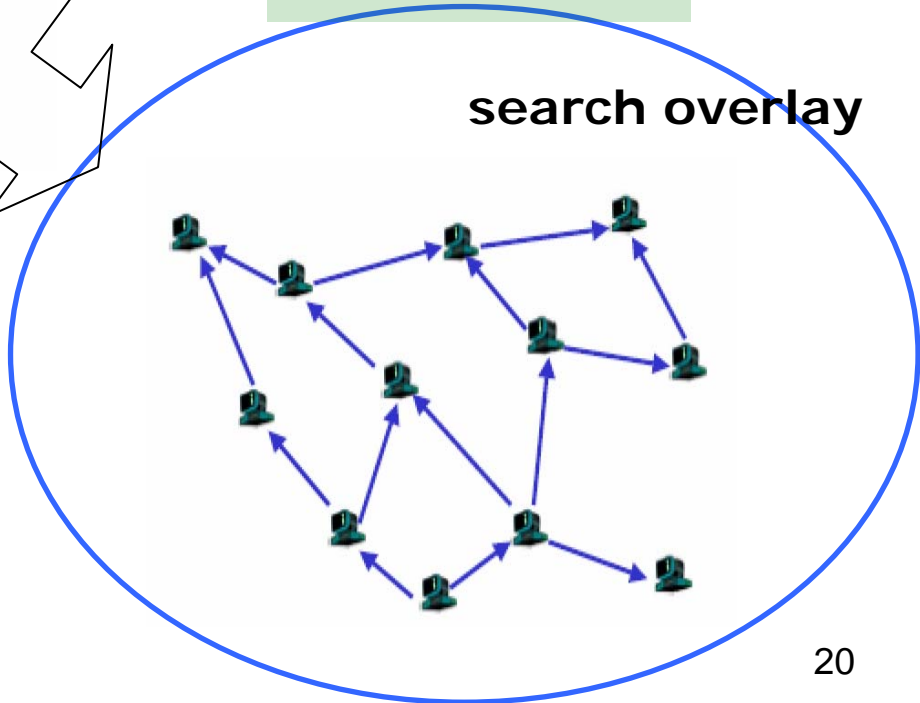
- A query is first issued to the search overlay.
- If the **first try** in the **search overlay** yields no hits at all or the peer is not satisfied with the results.
- The peer has a **second chance** by seeking search guidance from the **index overlay**.

Seeking search guidance

- An unpopular property as observed locally by a peer is registered to the index overlay.
- Together with interest-based registries, (un)popularity-induced registries can also be returned as potential destinations when the index overlay is queried for search hints.



Peers with similar interests



1. Issued to the search overlay
2. Seeking search guidance from the index overlay

Performance

- Search Delay
 - the time elapsed until the first reply is received at the querier

- Overhead
 - in the index overlay
 - in the search overlay

Search Delay

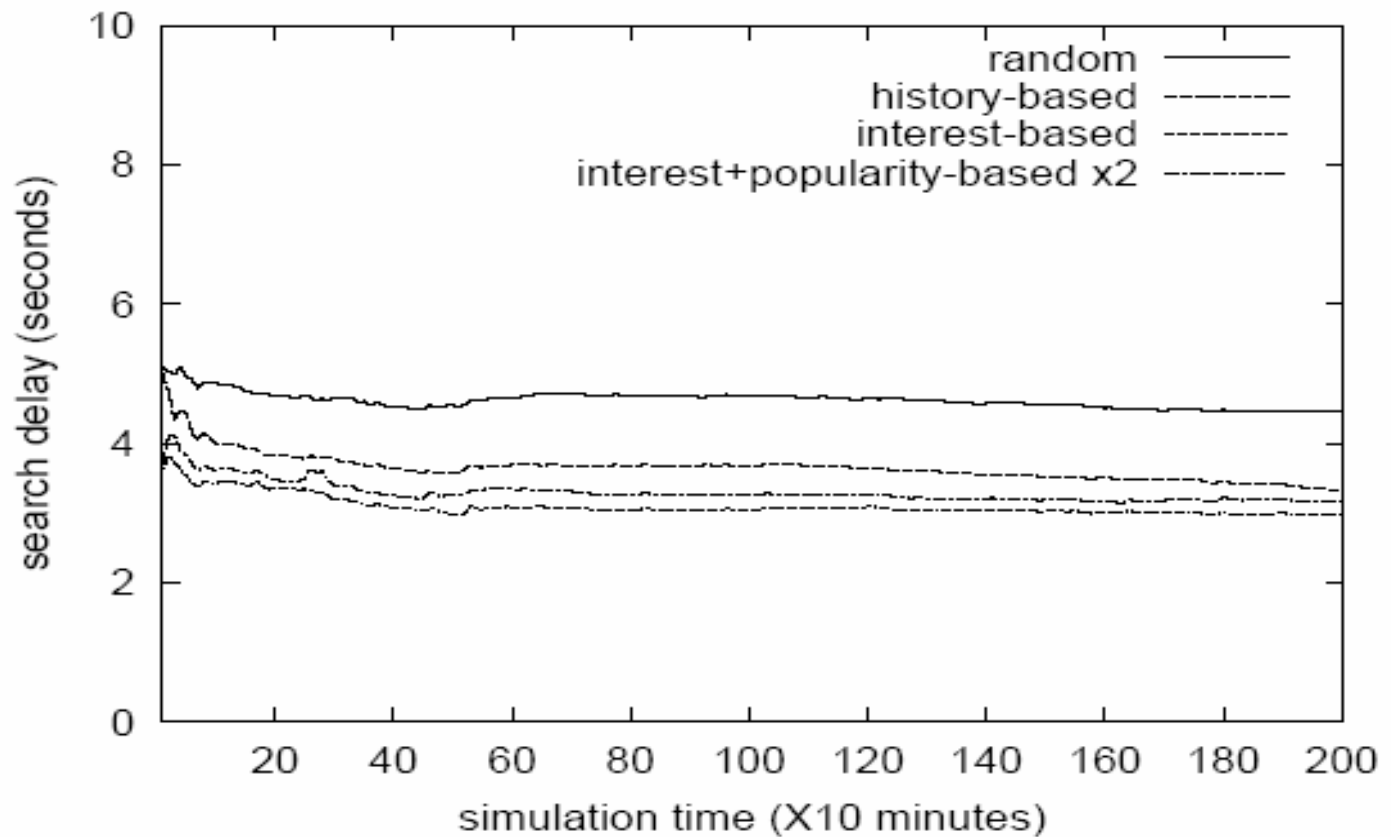
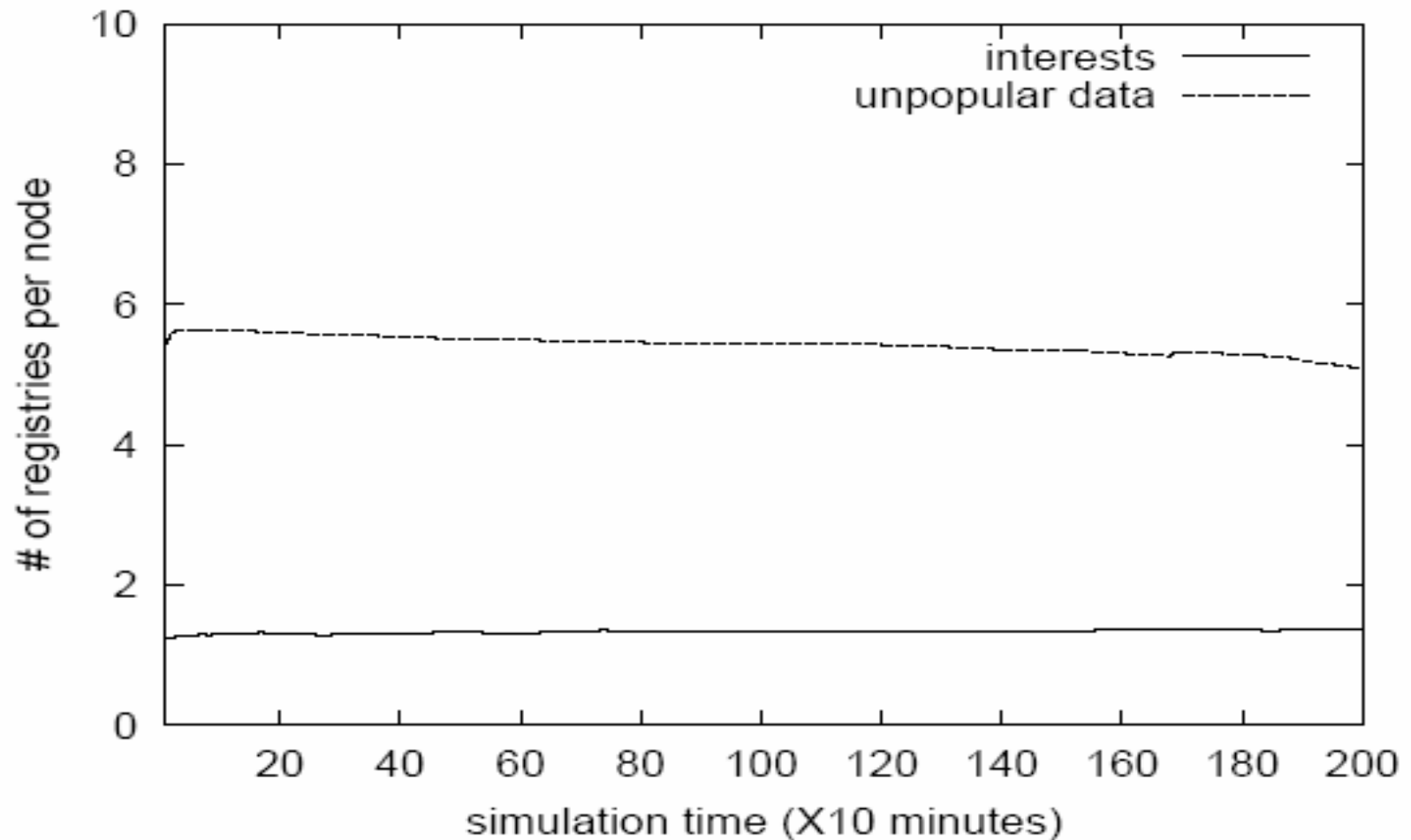
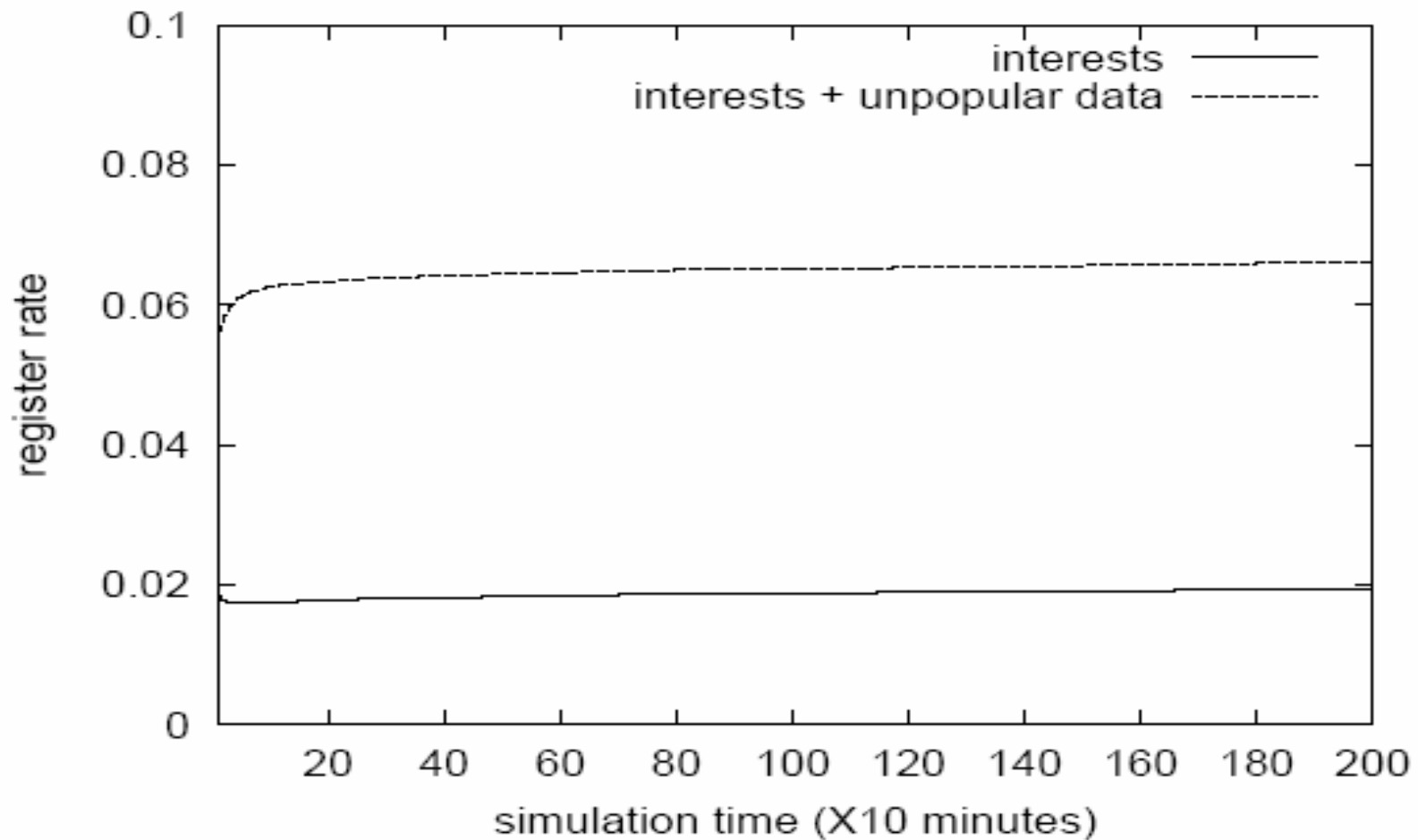


Fig. 20. Search delay

Overhead in the index overlay



Overhead in the search overlay



Conclusions

- We have presented a new protocol for P2P search based on peer interests and data popularity.
- The assisted search protocol leverages the advantages of both unstructured and structured P2P systems.
- The assisted search protocol achieves higher search efficiency and scalability than a pure flooding-based or history-based search scheme.