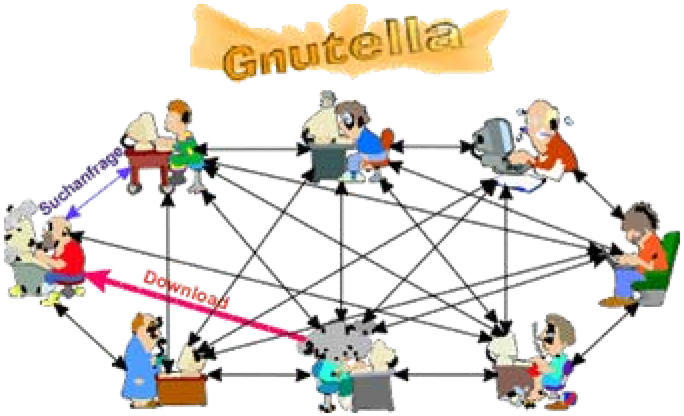
Fast and Low-Cost Search Schemes by Exploiting Localities in P2P Networks

Journal of parallel and distributed computing 2005

Presented by Ching Lan Wang
May 10, 2007

Outline

- Introduction
- Characterizing the localities in the peer community and individual peers
- The CAC-SPIRP algorithm : combining CAC and SPIRP techniques
- Experiments and performance evaluation
- Summary

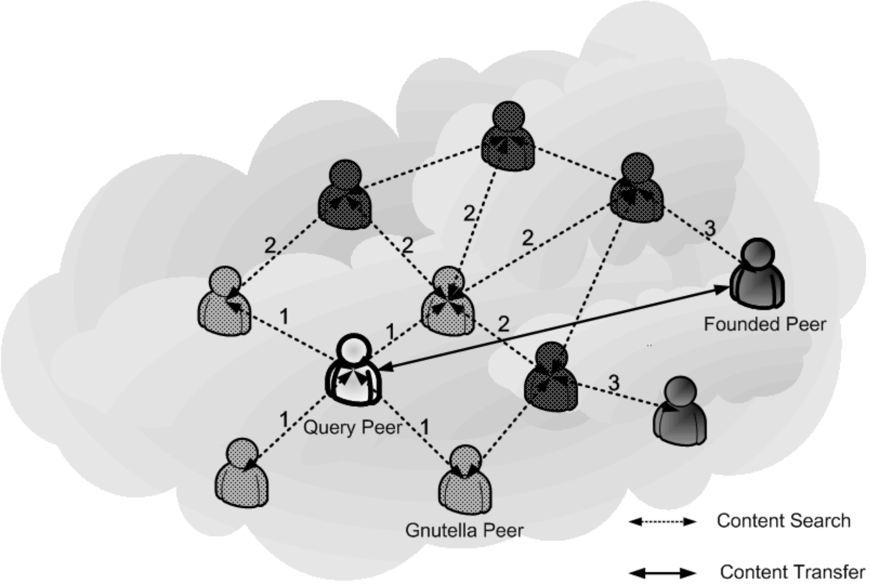


Introduction

- A lot of search schemes for decentralized, unstructured P2P networks have been proposed.
  - i.e., flooding, random walking
- The two performance objectives for designing and optimizing search algorithms in unstructured P2P networks
  - to improve search quality (individual peer)
    - to increase the number of effective results
    - to minimize the response time of each query
  - to reduce the total search cost of the peer community (all peers in the system)
    - to minimize the network bandwidth consumptions

Characterizing the localities in the peer community and individual peers

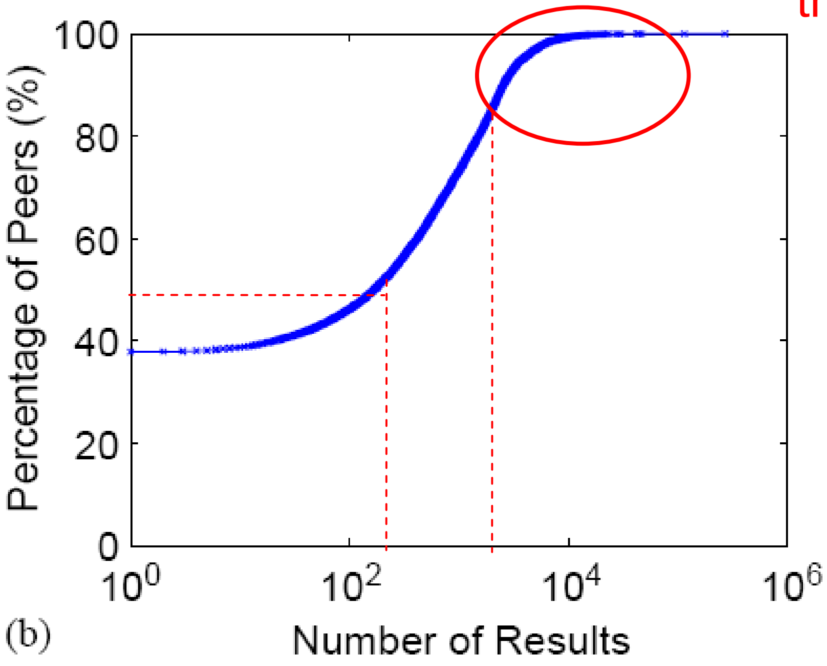
- The locality of content serving in the peer community
- The localities of search interests of individual peers



The locality of content serving in the peer community

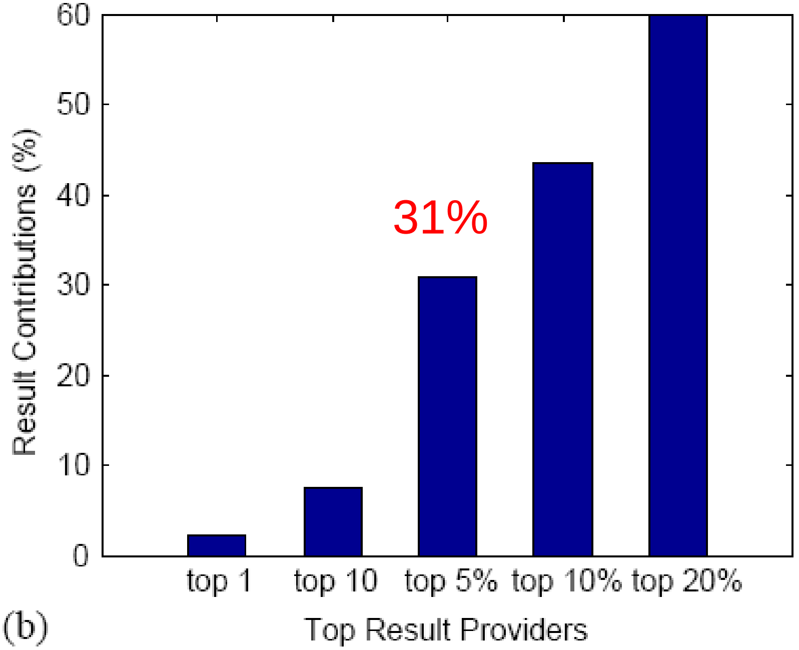
- Most search results are served by a small number of content-abundant peers.

10% peers can reply more than 2500 query results.



The localities of search interests of individual peers

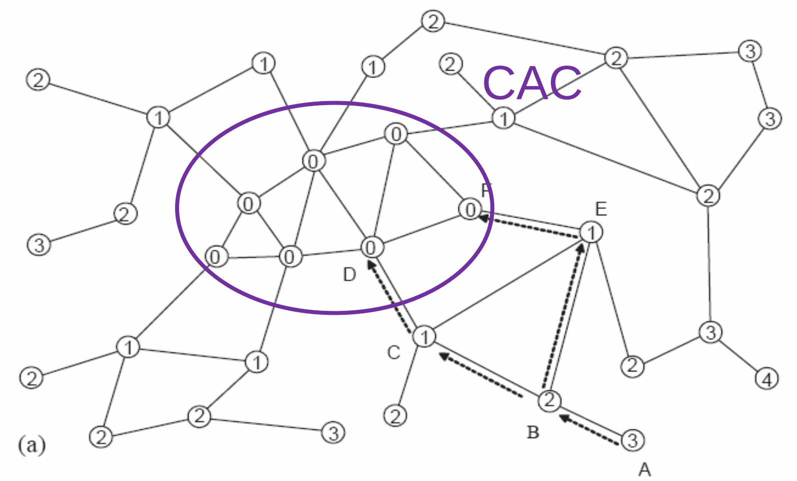
- A peer’s requests generally focus on a few interest topics, and it can be satisfied by a small number of peers with the same interests.



CAC technique

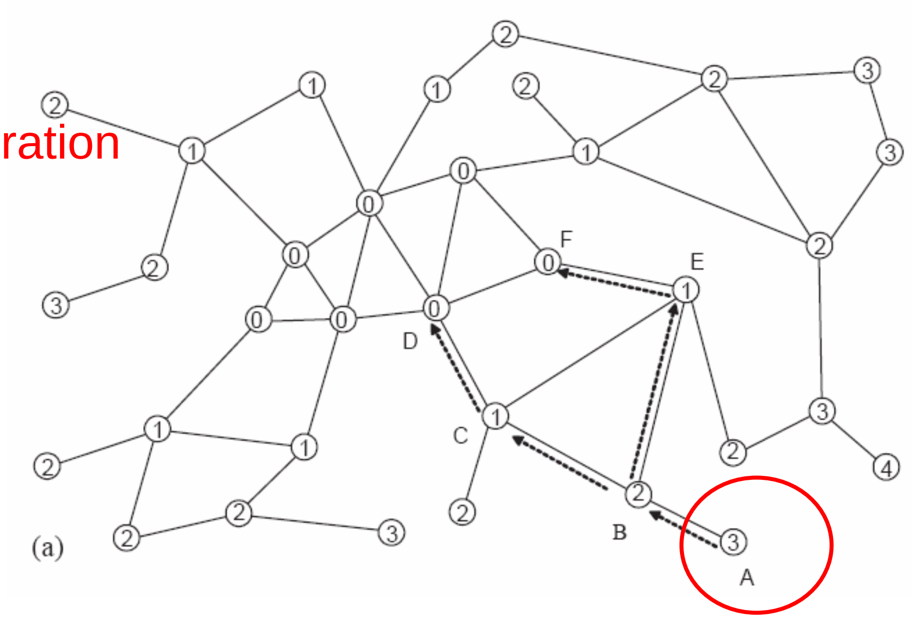
- CAC = Content-Abundant Cluster
- The basic idea of CAC technique is to have a collection of content-abundant peers in the peer community be self-organized into a CAC to actively serve contents for the entire P2P network.
- In CAC technique, we allow peers self-evaluate the quality of content service they can provide based on the history of their query-answering.
- Peers whose content service qualities reach a threshold are CAC member candidates and have the same possibility to join the CAC.

- The CAC is a connected overlay independent of the original P2P overlay.
- There are two types of links in P2P systems implementing CAC technique:
  - original P2P overlay link
  - CAC overlay link
- Each peer in the system is assigned a level:
  - the level of each CAC peer is defined as 0
  - the level of a non-CAC peer is defined as the number of hops from this peer to the nearest CAC peer.



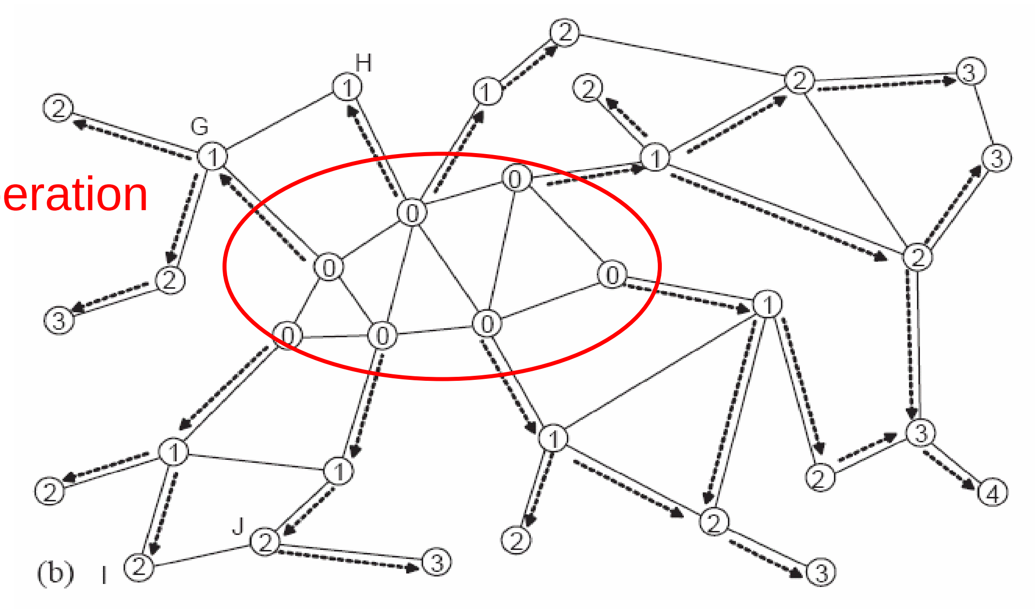
Query routing

- First, the query is up-flowed to and then flooded in the CAC again.
- Upon receiving the query, each CAC peer propagates it to level 1 peers immediately.



Up-flowing operation

- The query is propagated from lower level peers to higher level peers in the P2P overlay.
- Down-flooding is much more efficient than simply flooding the query in the P2P overlay because only links between two successive levels of peers are used for propagating queries, reducing a great amount of unnecessary traffic.



Down-flooding operation

Comparison with the super peer structure

- CAC has several advantages over the super peer structure due to several major differences between the two structures.
  - CAC is content-based, which has much richer resources than index-based super peers
  - The routing paths for up-flowing and down-flooding operations are on top of the P2P overlay and self-adaptive.

SPIRP technique

- SPIRP = Selectively Prefetching Indices from Responding Peers
- SPIRP technique is client oriented and motivated by the search interest localities of individual peers.
- Although the contents in a typical P2P network are huge and highly diverse, each peer's interests are limited and generally focused on a few topics.

- Each peer maintains a set of indices of files to be shared in the P2P network, called the outgoing index set.
- It also maintains a set of indices selectively prefetched from its responders, called the incoming index set.
- The responder set is also ranked as a priority queue, where the priority is defined as the number of queries a responder has responded so far.

SPIRP Operations

- Sending queries
  - As a requesting peer sends a query, it searches the incoming index set first.
  - If any indices match the query, the requesting peer checks if the corresponding responders are still *alive* and then returns the available matched results to the user.
  - If the query cannot be satisfied locally, the peer sends the query to the P2P network in a normal way.

- Prefetching and replacing indices
  - The requesting peer asks those high-priority responders, which are not in the incoming index set currently, to send their related indices until the incoming index set is full.

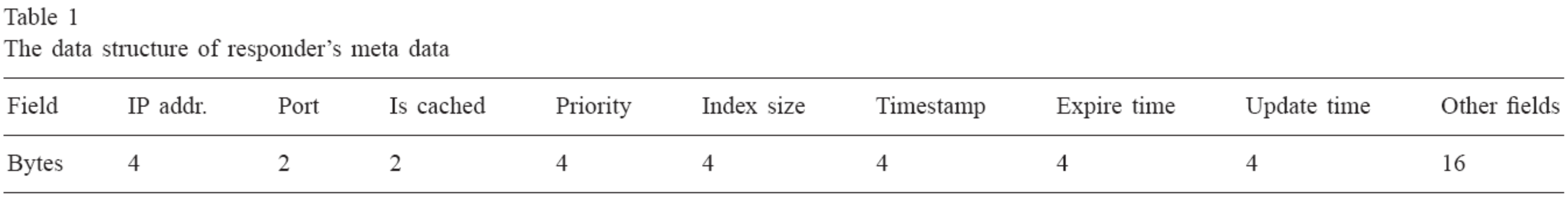


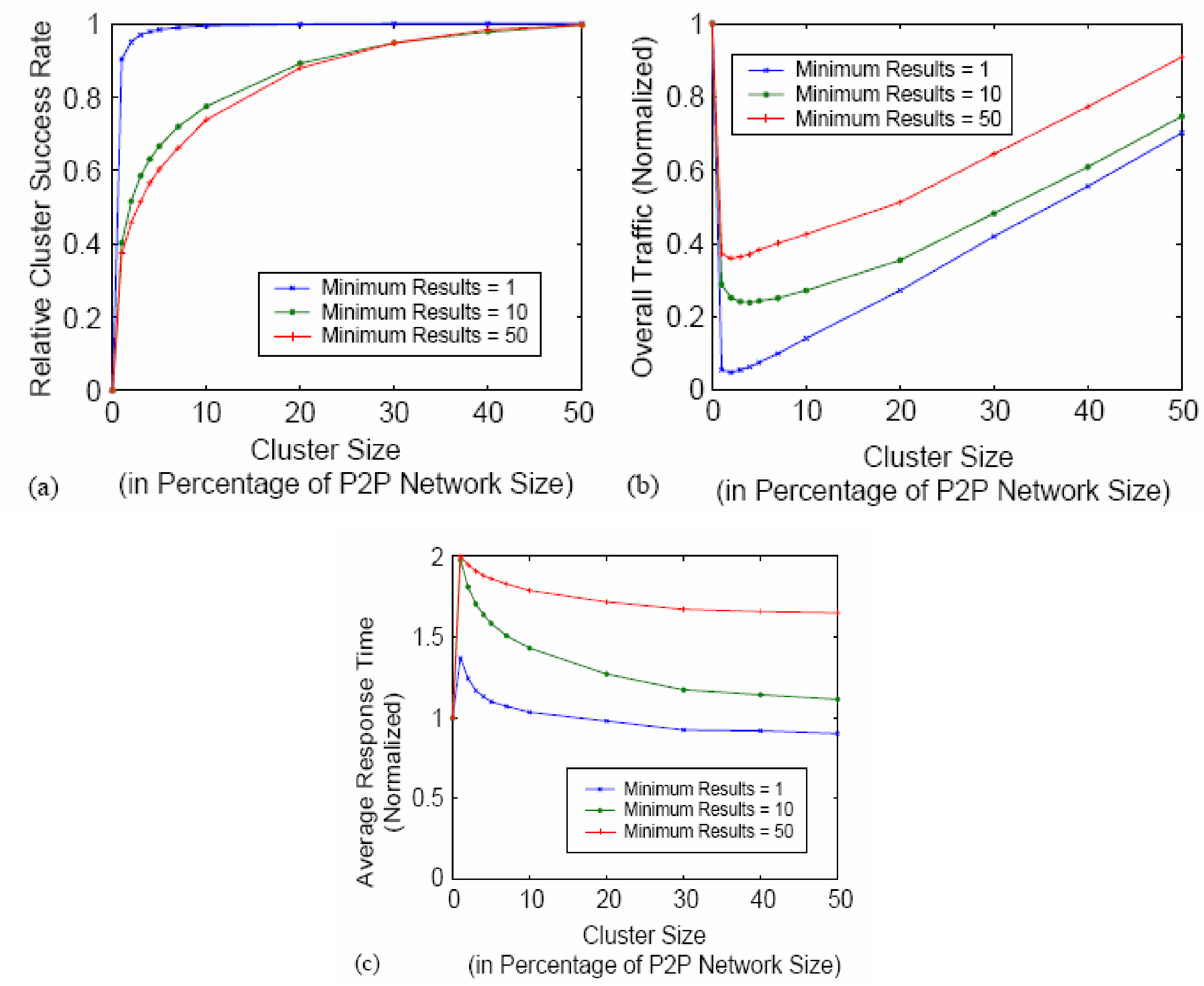
Table 1
The data structure of responder’s meta data

Field	IP addr.	Port	Is cached	Priority	Index size	Timestamp	Expire time	Update time	Other fields
Bytes	4	2	2	4	4	4	4	4	16

The CAC-SPIRP algorithm

- The CAC technique has its strong merits on reducing both bandwidth consumption and client response time when the requests success in the CAC, while the SPIRP technique shares the same advantage when the search interests is well exploited by the selective prefetching.
- The algorithm is simply to combine both CAC and SPIRP: the peers use SPIRP to prefetch file indices, and use CAC to route outgoing queries.

Performance evaluation of CAC technique

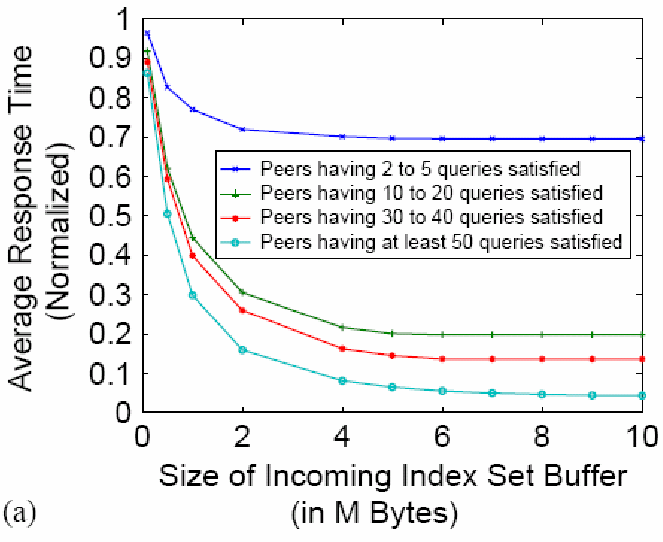


(a)

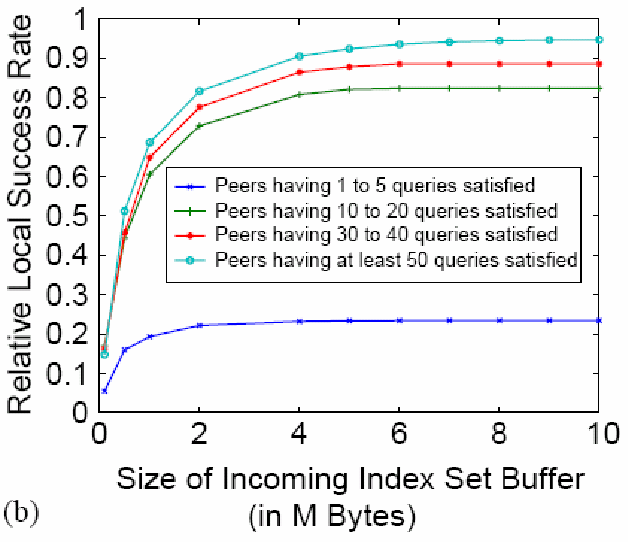
(b)

(c)

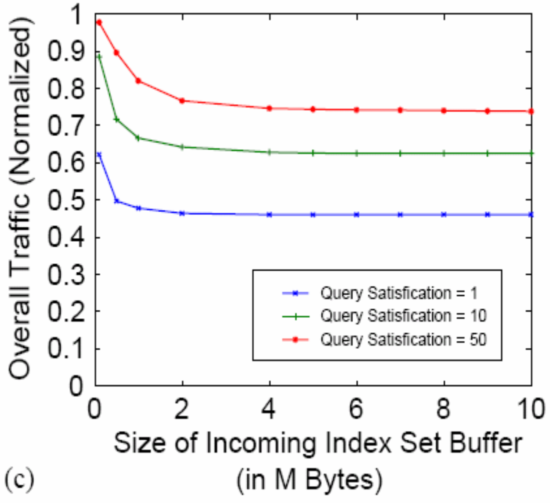
Performance evaluation of SPIRP technique



(a)

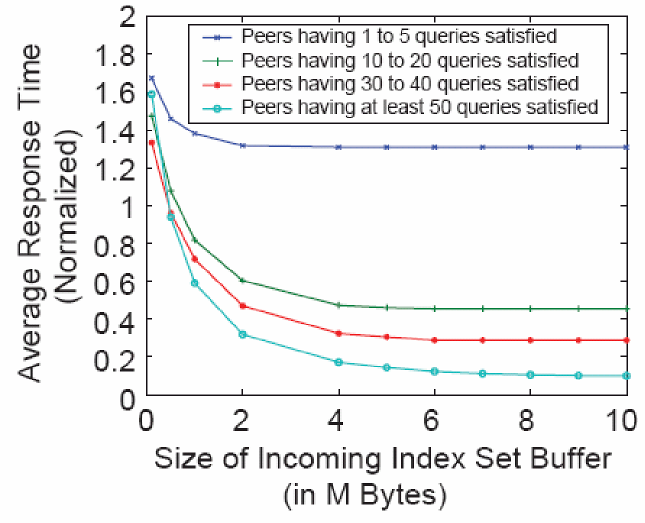


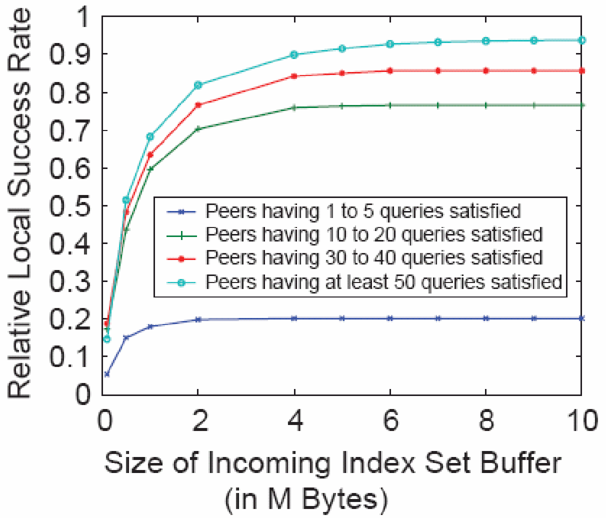
(b)

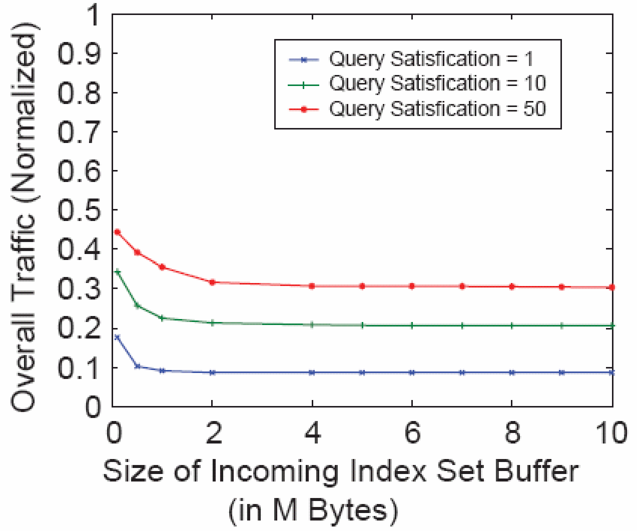


(c)

Performance evaluation of CAC-SPIRP algorithm







Summary

- Efficient content locating in unstructured P2P networks is a challenging issue because searching algorithm designers need to consider the objectives of both improving search quality and reducing search cost, which may have conflicting interests.
- It proposes CAC-SPIRP, a P2P searching algorithm aiming at both low traffic and low latency.
- By exploiting both the search interest localities of individual peers and the content locality in the peer community, we aim at achieving both objectives for significant performance improvements.