Assisted Peer-to-Peer Search with Partial Indexing

Rongmei Zhang and Y. Charlie Hu

INFOCOM 2005

Present by C.H. Lin
Outline

• Introduction
• Assisted search with partial indexing
• Experiment results
• Conclusion
Introduction

• Centralized P2P
  – Napster
  – Maintain a full index of shared files
  – Search is performed on centralized server

• Disadv.
  – Not scalable
Introduction

• Unstructured P2P
  – Gnutella
  – Flooding based search mechanism

• Disadv.
  – Not scalable
Introduction

• Structured P2P
  – Chord, Pastry, CAN
  – DHT based search mechanism

• Disadv.
  – Can not support keyword search
Introduction

• Assisted search with partial indexing
  – Exploiting partial index to improve search efficiency in unstructured P2P overlay network
  – Index is built based on interest
Assisted search with partial indexing

• Interest
  – The interests of a peer are represented by the dominant properties of its data possession

• Popularity
  – Property popularity can be determined from observing passing traffic
  – Properties of local data that are seen the least frequently in passing queries are identified as “unpopular”
Assisted search with partial indexing

• Index overlay
  – Structured P2P network
  – Pastry

• Search overlay
  – Unstructured P2P network
  – Gnutella
Index overlay

• The index maintains two types of information:
  – the top interests of peers
  – globally unpopular data

• The partial index has three complementary purposes
  – assists peers to find other peers with similar interests
  – provides search hints for those data difficult to locate
  – improve the chances of finding unpopular data

• The index overlay is constructed according to the bootstrap mechanism of the corresponding structured P2P overlay
Search overlay

• Search overlay is created based on peer interests
• After joining the index overlay, a new peer can obtain the addresses of other peers with similar interests
• As the reply containing such nodes is received, the peer initiates connections to them
Search

• 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
Experiment results

- History-based [20]
  - Interest-based shortcuts

(a) Gnutella.

(b) Shortcuts.
Experiment results

![Graph showing experiment results with various lines representing different strategies: random, history-based, interest-based, interest+popularity-based x 2, and optimal. The y-axis represents success rate, and the x-axis represents simulation time (X10 minutes).]
Experiment results

![Graph showing experiment results with three lines representing random, history-based, and interest-based methods over simulation time. The search scope is on the Y-axis, and simulation time (X10 minutes) is on the X-axis.]
Experiment results

![Graph showing search delay over simulation time for different methods: random, history-based, interest-based, and interest+popularity-based x2. The x-axis represents simulation time (X10 minutes), and the y-axis represents search delay (seconds).]
Conclusion

• The assisted search protocol achieves higher search efficiency and scalability than a pure flooding-based or history-based search scheme.

• At the same time it also retains desirable features of search in unstructured overlays such as robustness.