PeerDB: A P2P-based System for Distributed Data Sharing

Wee Siong Ng
Beng Chin Ooi
Kian-Lee Tan
Aoying Zhou

Presented by: Aseem Cheema
For CS-856 Web Data Management
Prof. Tamer Ozsu

Overview

- Introduction
- BestPeer – a framework for PeerDB
- PeerDB Node Architecture
- PeerDB Relation matching
- PeerDB Query processing
- PeerDB Auto-reconfiguration
- PeerDB Cache management
- PeerDB Performance Study
- Comments

Introduction

1. P2P vs DDBS

- ad-hoc / stable nodes
- incomplete / complete answers
- global shared schema
- content location
- coarse / fine granularity

Introduction

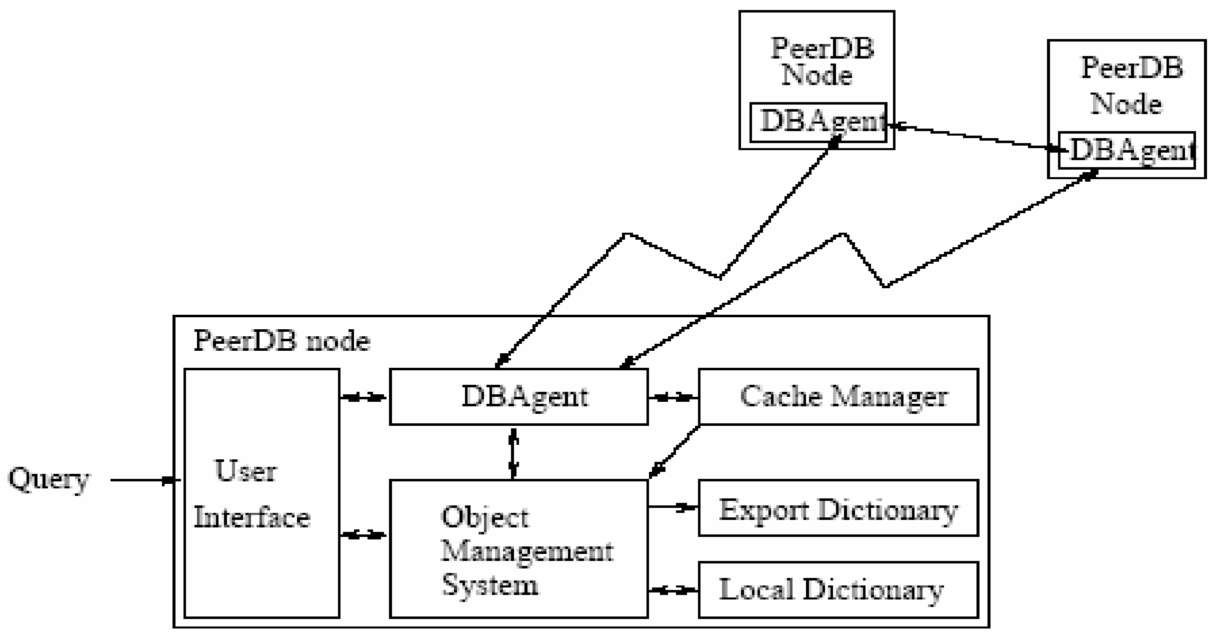
2. PeerDB is P2P DDBS

- Each node is a DBMS
- No global schema
- Incomplete answers possible
- Fine granularity, content searching

BestPeer – a framework for P2P

- LIGLO Servers
- Each node has private & shared data
- Mobile agents & P2P
- Fine granularity
- Share computational power
- Dynamic reconfiguration

PeerDB – Node Architecture



Peer DB – Node Architecture

- Data Management System – MySql
  - Meta-data (keywords) stored in Local Dictionary and Export Dictionary.
- DBAgent – for Mobile Agents
  - Master agent that manages the queries.
  - Dispatches worker agents to neighboring nodes.

PeerDB – Relation Matching

For each relation, meta-data (keywords provided by user) are maintained for relation name and attribute names.

Considering Query Q of form (R,A,C), R is searched against keywords for relation names, and V (A U C) is searched against keywords for attribute names.

$$\text{Match}(Q,D) = \frac{(wtr \cdot r) + (wta \cdot N_{match}(AUC,T))}{Wtr + (wta \cdot N(AUC))}$$

Set of relations above a threshold value are considered.

Peer DB – Query Processing

- Phase 1 – Relation matching technique is applied and relations returned to the user. User chooses the relations he is interested in.
- Phase 2 – Queries directed to the chosen peers and answers returned.

Peer DB – Query Processing

Local Query Processing

Phase 1
- Master Agent created to overlook the operation.
- Relation matching agents are dispatched (IP & TTL).
- Returns relations to user for selection.

Phase 2
- Data Retrieval Agent for user selected relations.
- Answers returned to the user.

Peer DB – Query Processing

Remote Query Processing

Phase 1
- Agent arrives, if new, TTL reduced by 1.
- Export dictionary searched & relations returned.
- If TTL > 0, more relation matching agents cloned.

Phase 2
- Agent formulates SQL query.
- Results returned & agent dropped.

Peer DB – Auto-Reconfiguration

Monitoring Statistics

- Reconfiguration based on the policy selected by the user.
- Relation information, number of answer objects returned.
- Temporal Locality using stack.

Peer DB – Cache Management

- Caching for a fixed period of time.
- LRU replacement policy.
- BPID to avoid multiple copies.

Peer DB – Performance Study Relation Matching Strategy

- Set of semantically related C categories.
- c keywords in each category.
- Created set of relations with 2-5 keywords.
- Attributes assigned 2-5 keywords.
- SELECT attribute_X FROM relation_i
  WHERE attribute_Y = value_1 and
  attribute_Z > value_2.

Peer DB – Performance Study Relation Matching Strategy

Threshold	Precision	Recall
0.1	0.33	0.85
0.3	0.36	0.78
0.5	0.50	0.57
0.7	1.0	0.28
0.9	1.0	0.21

Peer DB – Performance Study On PeerDB Performance

Evaluation Methodology

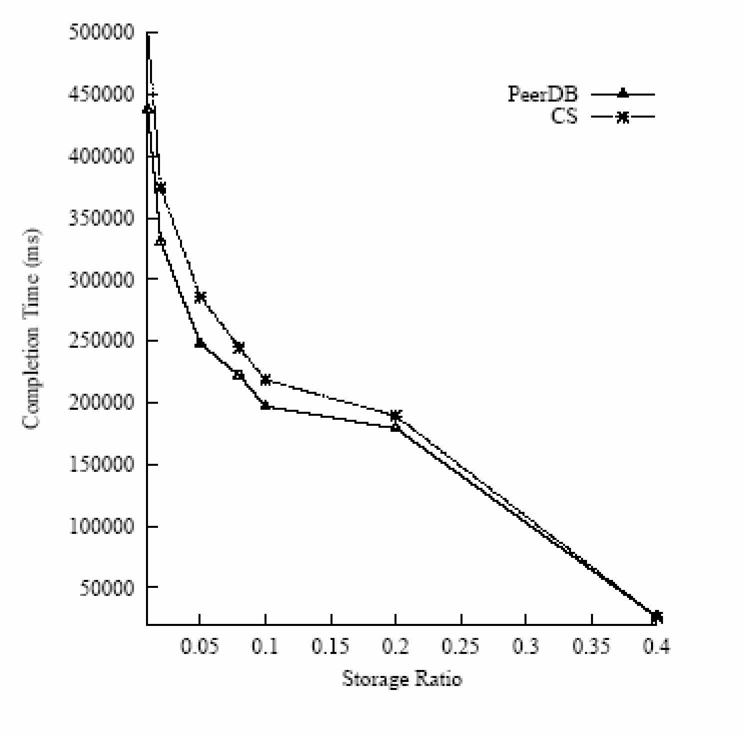
- P2P protocols & reconfiguration.
- Response Time / Rate.
- Quantity & Quality of answers.

Peer DB – Performance Study On PeerDB Performance

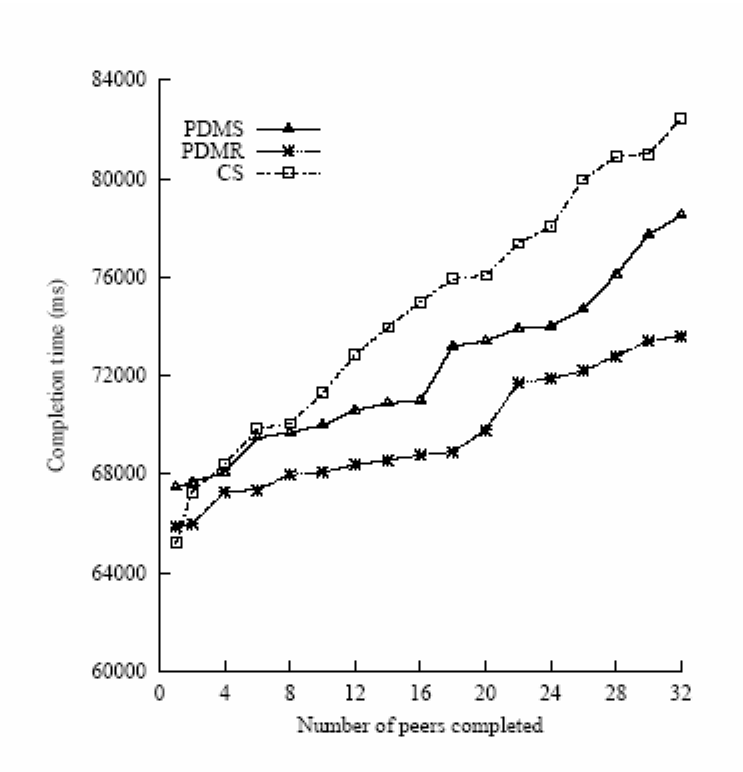
Experimental Setup

- 32 PCs, 200MHz, 64M, WIN NT 4.0
- 10,000 objects, 10 KB each, each node holes 1000 object.
- 80% queries directed at 20% of data.
- 15% queries directed at 20% of cold data.
- Average of at least 3 different executions.

Peer DB – Performance Study

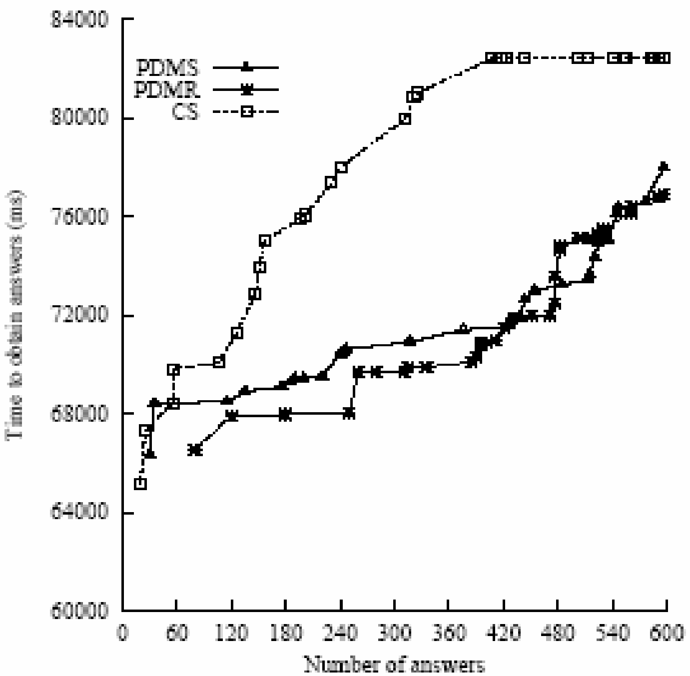


Effect of Storage Capacity

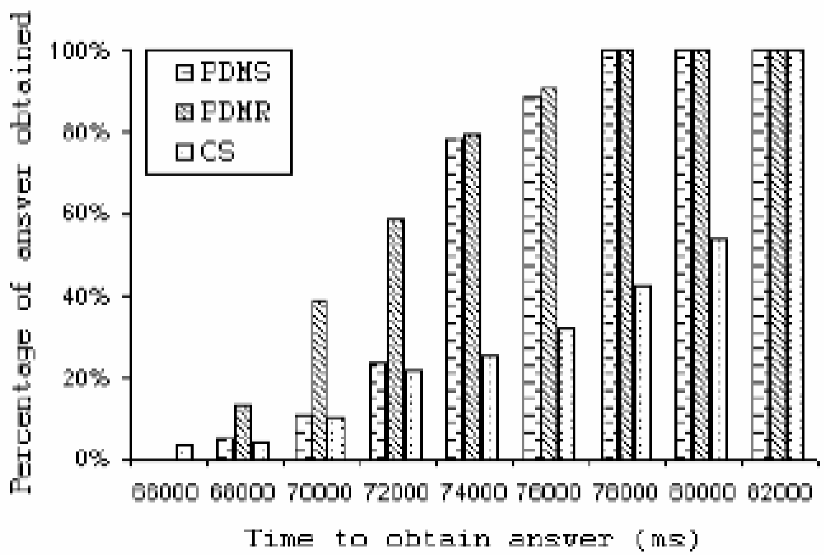


Rate of Returning Answers

Peer DB – Performance Study



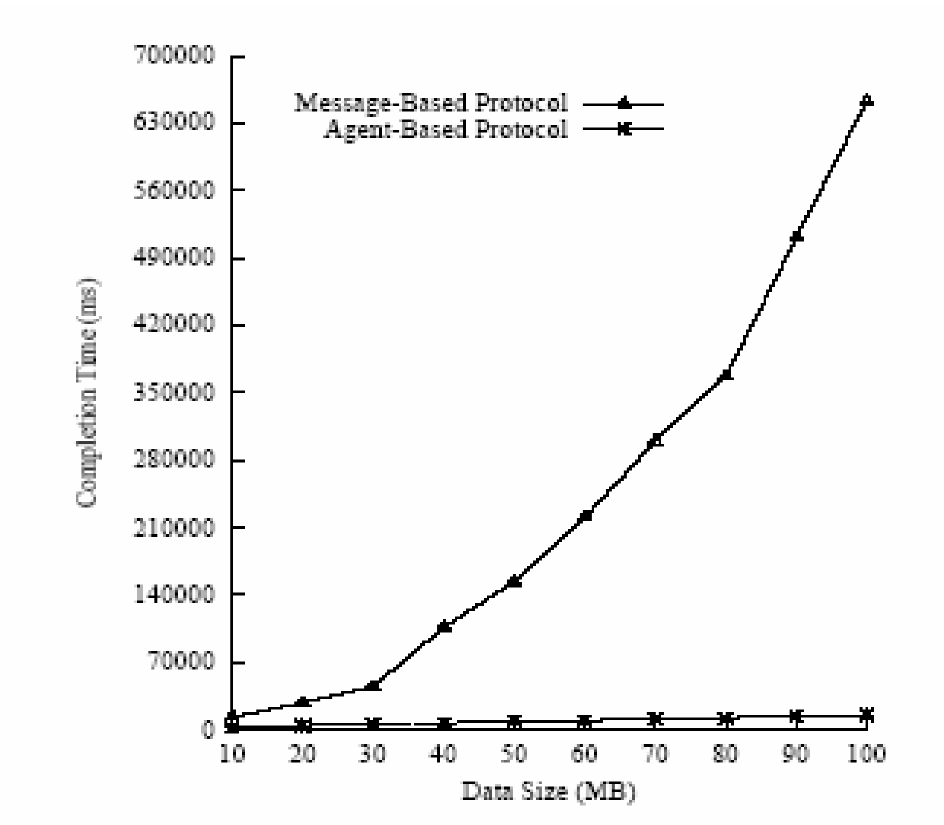
(a) First search query.



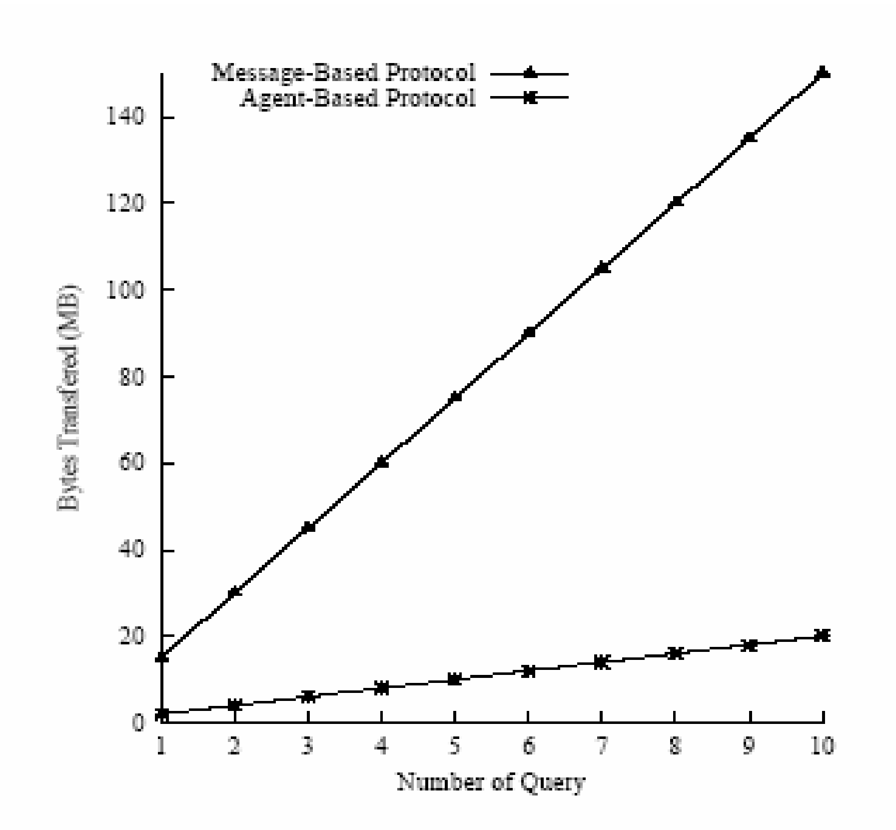
(b) % answers returned.

Number of Answers Returned.

Peer DB – Performance Study



Completion time vs Data Size



Communication Overhead

Peer DB – Comments

- Search Engine without a ranking algorithm ???
- User selection – scalability ???
- What CS ???

Peer DB – References

[1] W. S. Ng, B. C. Ooi, and K. L. Tan. BestPeer: A self configurable peer-to-peer system. In *Proceedings of the 18th International Conference on Data Engineering*, page 272, San Jose, CA, April 2002 (Poster Paper).

[2] www.BestPeer.com

[3] Marthie Schoeman, Elsabé Cloete. Architectural Components for Efficient Design of Mobile Agent Systems. In *Proceedings of the 2003 annual research conference of the South African institute of computer scientists and information technologists on Enablement through technology,* September, 2003.