Optimal Query Mapping in Mobile OLAP

Ilias Michalarias and Hans -J. Lenz

Freie Universität Berlin

{ilmich,hjlenz}@wiwiss.fu-berlin.de

Overview

I. mobile OLAP Research Context

II. Query Mapping in mOLAP

III. An Analytical Framework for Derivability Estimation

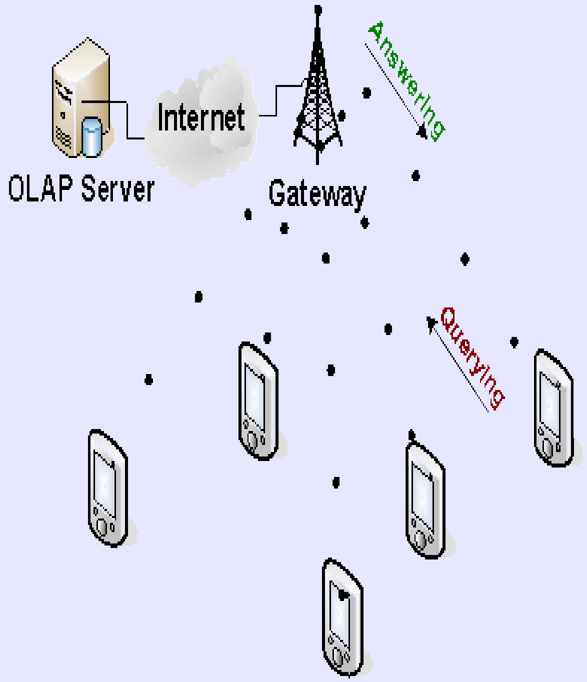
IV. Evaluation

V. Outlook

mOLAP Application Scenario

Freie Universität Berlin

- Brokers accessing the stock market gallery data mart:
  - At opening and closing times different stocks in different financial dimensions are analyzed by many traders using some portable device
  - Some of these stocks are more popular than others, similarly, some analytical dimensions are more important than others
  - In this situation, a data mart equipped with a broadcast gateway will be responsible for serving the incoming requests



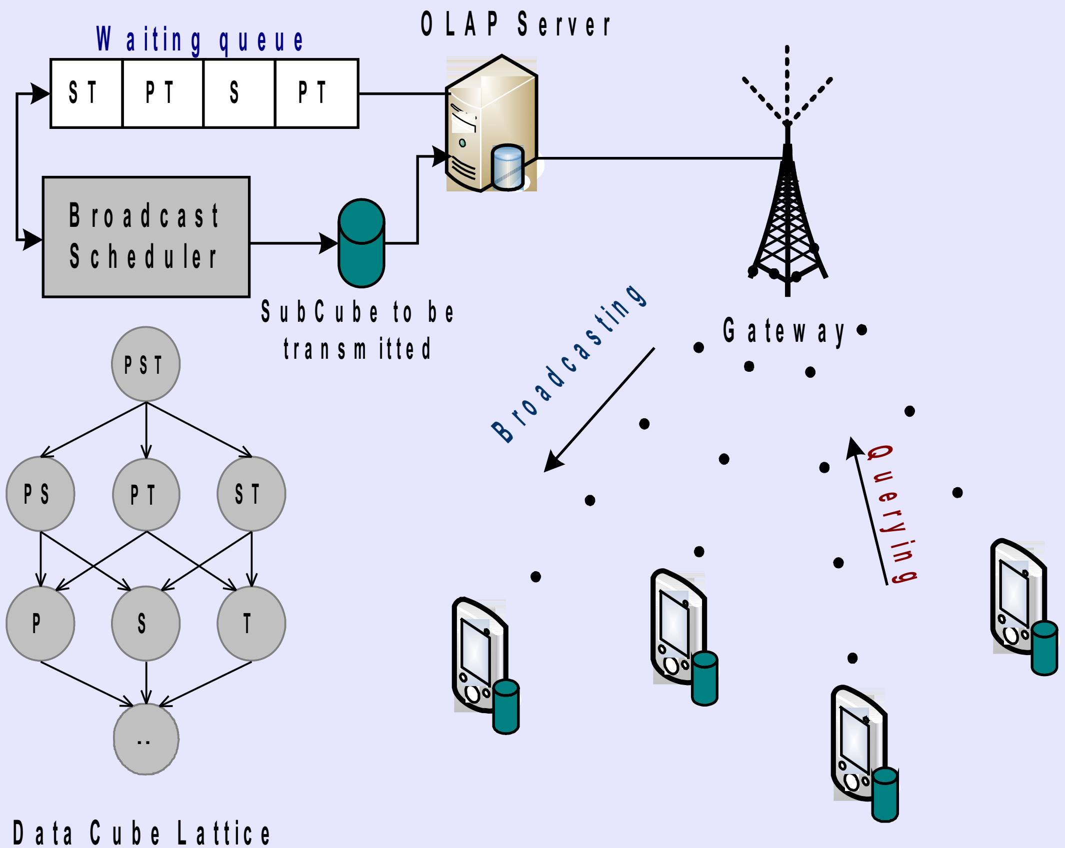
mOLAP Research Context

Freie Universität Berlin

- Major Issues:
  - Management of Multidimensional data in wireless networks
  - Providing equal/comparable functionality with desktop counterparts
  - Cope with limited resources such as bandwidth, energy and small screen size
- Fundamental Requirement:
  - Offline functionality
- Wireless bandwidth the usual bottleneck of the system
- Transmitted sub-cubes are items order of magnitude bigger than usually assumed by conventional broadcast systems e.g., web pages

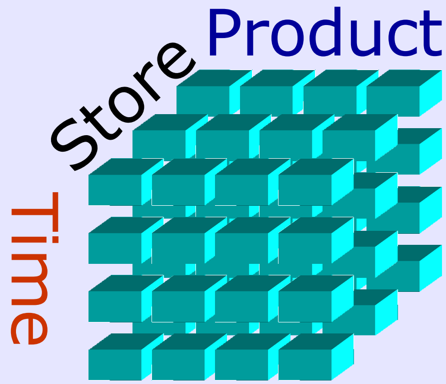
Architecture

- Client-Server Network Architecture
- Broadcast-based Dissemination
- Clients may have to locally process data

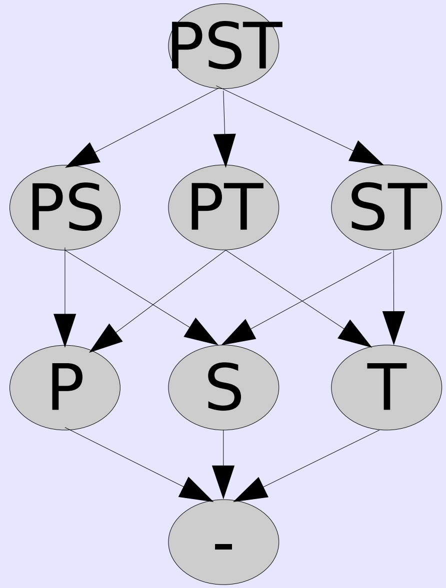


Optimization Options

- **Optimization by means of:**
  - **Subsumption**
    - In conjunction with wireless broadcast reduces the number of necessary transmissions [Sharaf et al., 2004] [Michalarias et al., 2005]
  - Compression
    - Receiving clients are served faster
    - Indirect reduction of waiting time for pending requests [Sharaf et al., 2003] [Michalarias et al., 2006, 2007]



Data Cube



Data Cube Lattice

Why Query Mapping?

Freie Universität Berlin

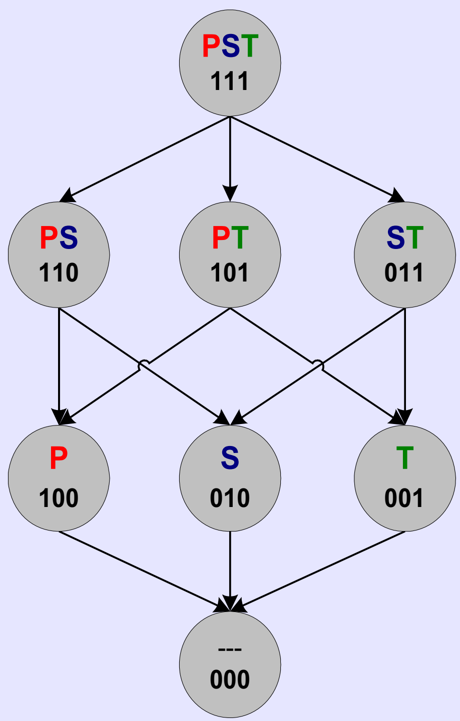
- Upon receipt, queries are mapped to the corresponding nodes of the aggregation lattice because:
  - Point-to-Point system proven inefficient
  - Broadcast systems perform better with a limited number of data items
  - Subsumption exploitation becomes higher and thus each transmission (broadcast) serves multiple requests

Aggregation Lattices

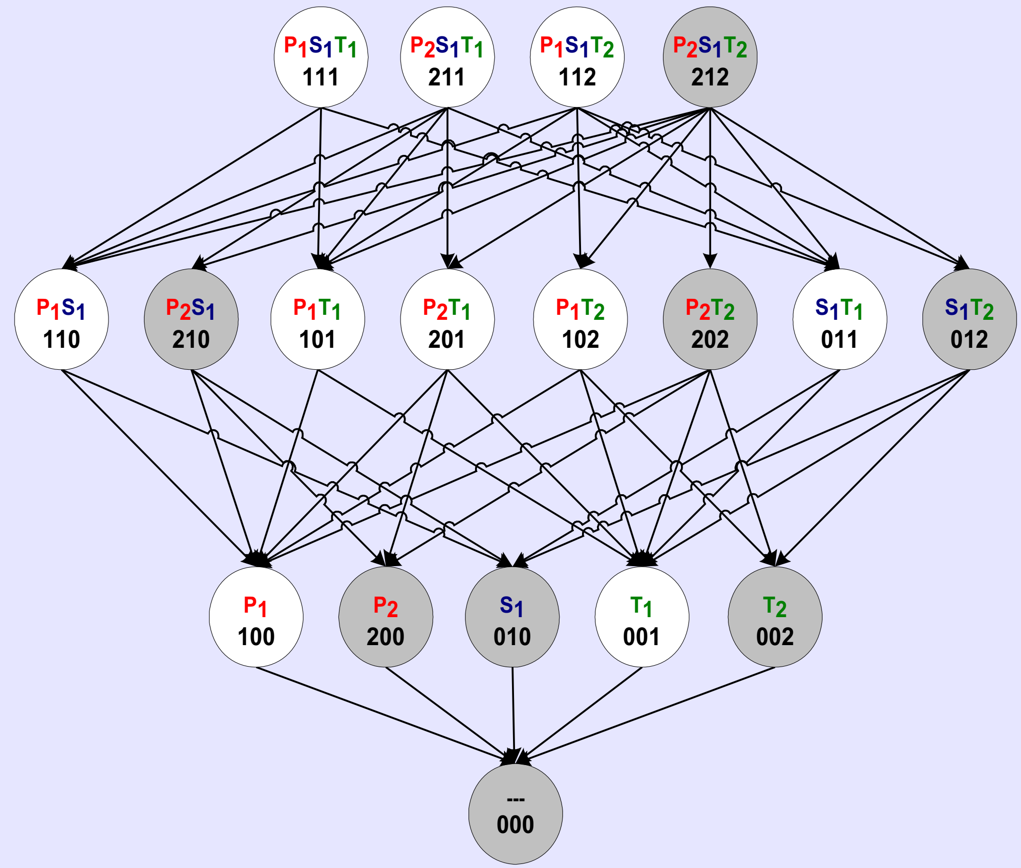
Hierarchies		
Product	**Store**	**Time**
ALL P_0		ALL T_0
î	ALL S_0	î
Category P_1	î	Year T_1
î	StoreID S_1	î
Code P_2		Day T_2

\# of DCL nodes= 2^n

\# of hDCL nodes= $\prod_{i=1}^{N} Gr(i)$



Data Cube Lattice *(DCL)*

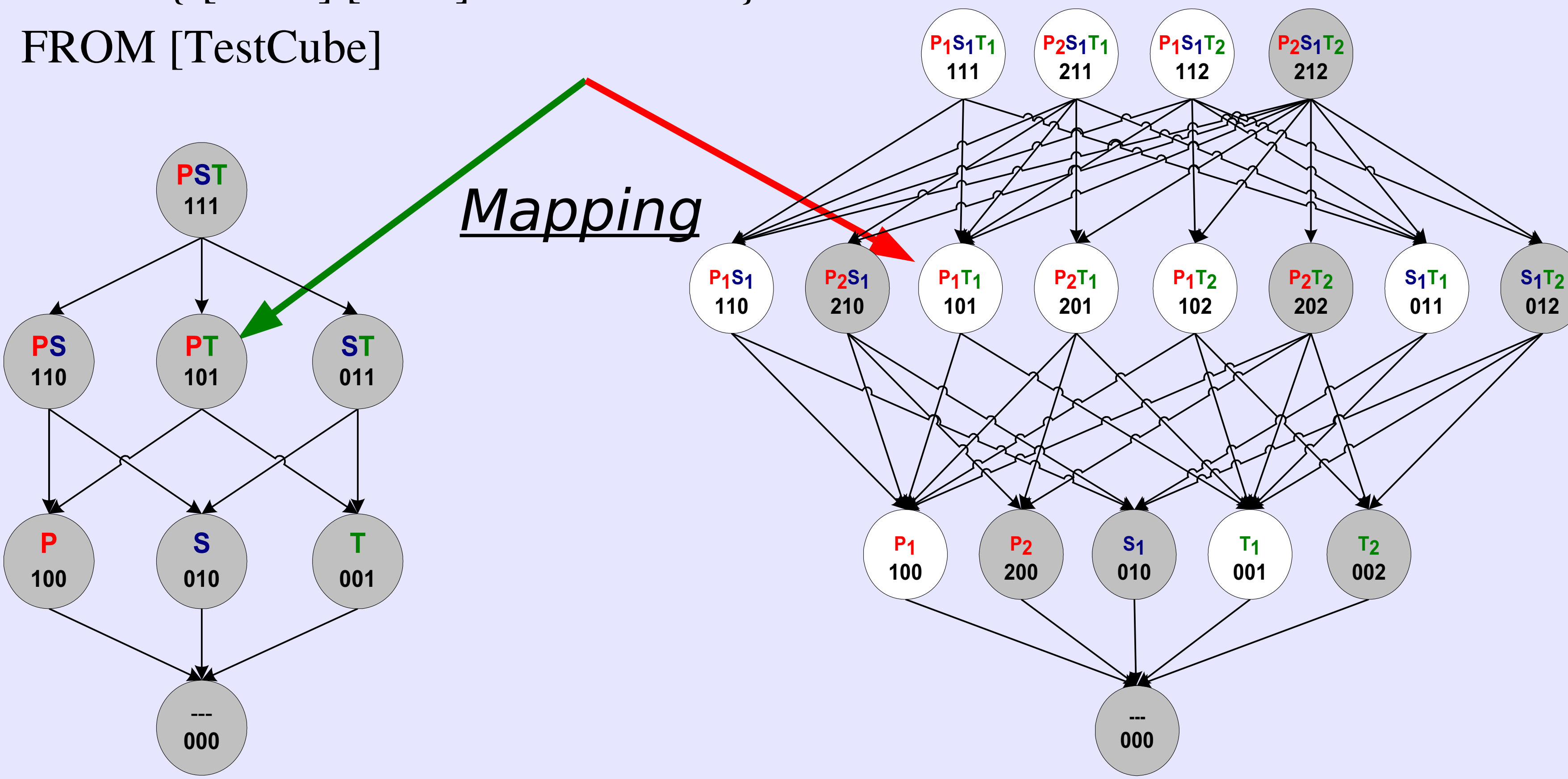


Hierarchical Lattice *(hDCL)*

Query Mapping to Aggregation Lattices

- An Example:

SELECT
{ [Product].[Category].[Drinks] } ON COLUMNS,
{ [Time].[Year].AllMembers } ON ROWS
FROM [TestCube]



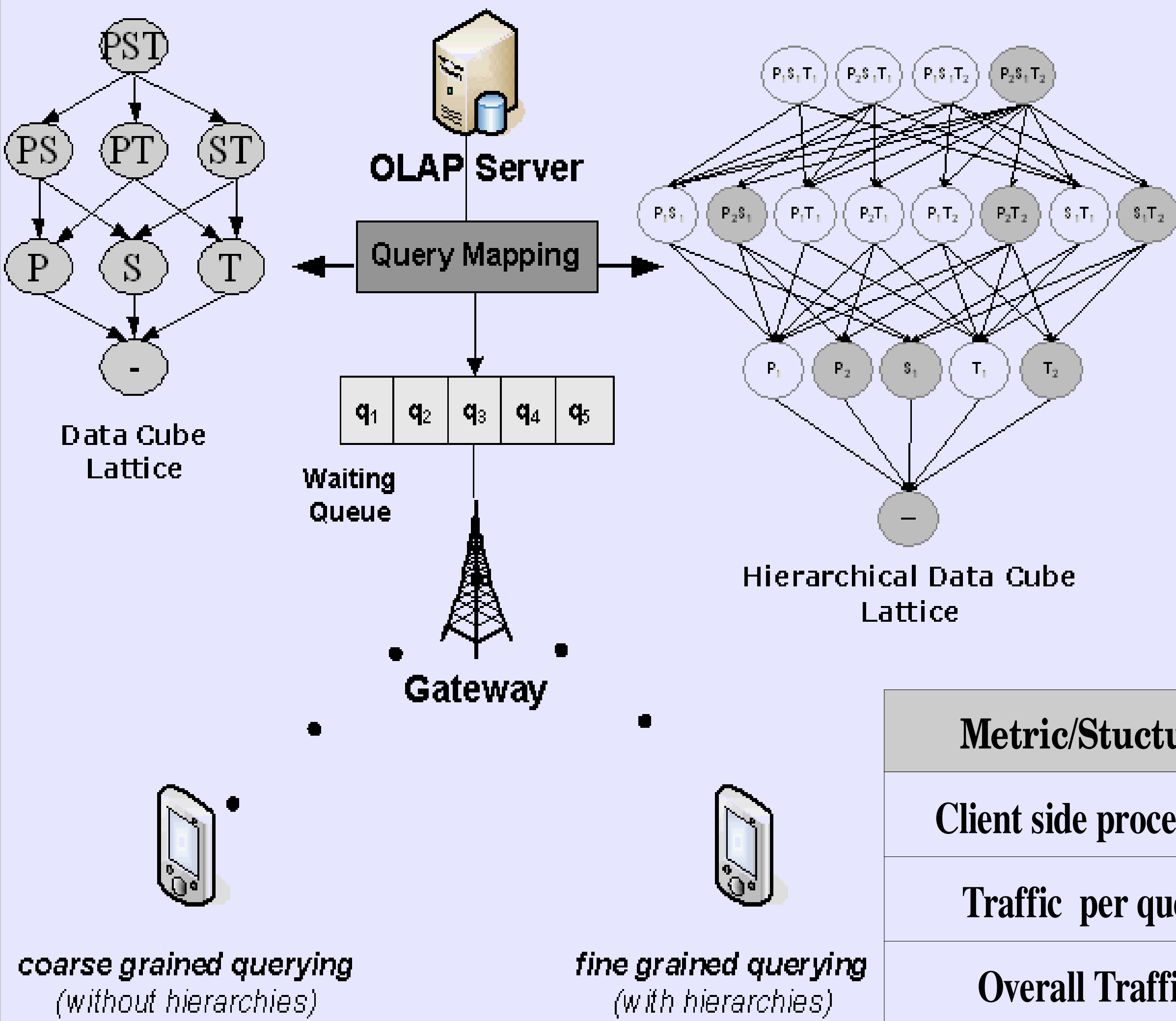
Data Cube Lattice *(DCL)*

Hierarchical Lattice *(hDCL)*

An Analytical Framework for mOLAP

- Problem: Query mapping (in mOLAP) using the DCL or the hDCL?
- Motivation
  - Previous approaches rather intuitive – no formal background
  - Related work does not provide any clear answer
- Objective
  - Analyze the impact of hierarchies on mOLAP
  - Optimize subsumption exploitation among sub-cubes
- Information provided
  - Subsumption probabilities in general
  - mOLAP specific derivation Probabilities

Trade-offs with query mapping



Metric/Stucture	DCL	hDCL
Client side processing	more	less
Traffic per query	more	less
Overall Traffic	?	?
# of data items	less	more
Offline functionality	enhanced	limited

An Analytical Framework for mOLAP

Freie Universität Berlin

- Its objective is to quantify the degree of subsumption exploitation in mOLAP
- Features
  - The mOLAP queue is modeled as a *multiset*
  - The query distribution is based on server collected statistics making the framework independent of the workload
  - Each probability is computed both for *DCL* and *hDCL* mapping
  - Provides the basis for additional computations

Probabilities

1. $P(e_a \succeq e_b)$: In a multiset Q of lattice nodes, the probability that a selected element $e_a \in Q$ is an ancestor of another selected element $e_b \in Q$.
2. $P(e_a \succeq Q)$: In a multiset Q, the probability that a selected element $e_a \in Q$ is an ancestor of every element in Q.
3. $P(\exists\ e : e \succeq Q)$: In a multiset Q, the probability that there exists one element, which is an ancestor of every element in Q.
4. $P(e_a \succeq q \subseteq Q)$: In a multiset Q, the probability that a selected element $e_a \in Q$ is an ancestor of exactly $|q|$ $(|q| - 1 + itself)$ elements of Q.
5. $P(e_a \succeq q^+ \subseteq Q)$: In a multiset Q, the probability that a selected element $e_a \in Q$ is an ancestor of at least $|q|$ $(|q| - 1 + itself)$ elements of Q.
6. $P(\exists\ e : e \succeq q \subseteq Q)$: In a multiset Q, the probability that there exists at least one element, which is ancestor of exactly $|q|$ $(|q| - 1 + itself)$ elements of Q.

Relevance to mOLAP

Subsumption AFTER scheduling decision approach (STOBS, SBS)

1st Step: Item to be transmitted decided by scheduling metrics (irrelevant to subsumption)

q_1	q_2	q_3	q_4	q_5	q_6	q_7	q_8	q_9

2nd Step: Subsumptions examined considering the already defined ancestor

q_1	q_2	q_3	q_4	q_5	q_6	q_7	q_8	q_9

Finally: Broadcast q_8 and serve $\{\mathbf{q_6}, q_2, q_4\}$

$P(e_a \succeq q \subseteq Q)$: In a multiset Q, the probability that a selected element $e_a \in Q$ is an ancestor of exactly $|q|$ $(|q| - 1 + itself)$ elements of Q.

Subsumption BEFORE scheduling decision approach (FCLOS)

1st Step: All elements ancestor candidates

q_1	q_2	q_3	q_4	q_5	q_6	q_7	q_8	q_9

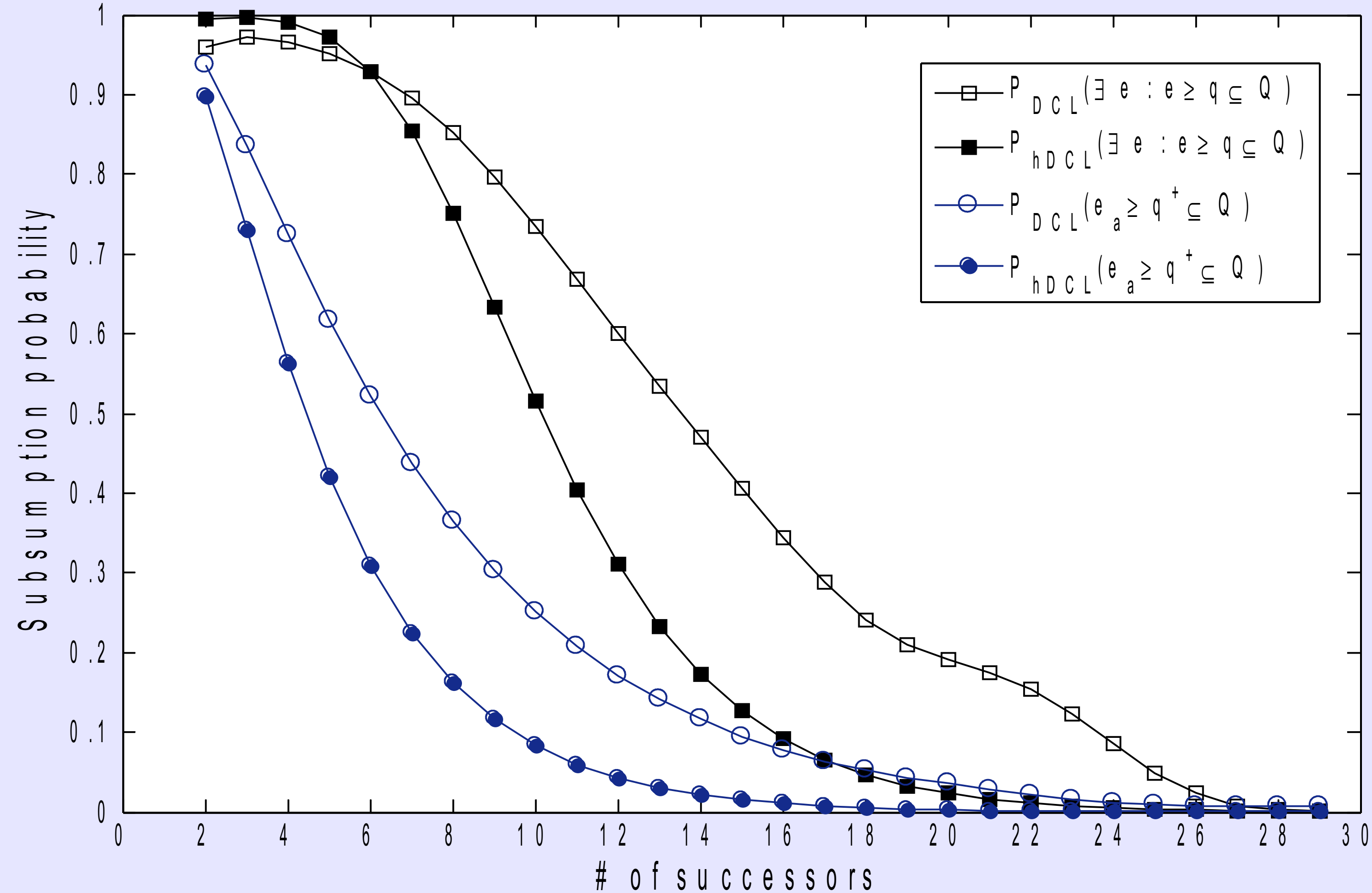
2nd Step: Find optimal grouping (after having checked all combinations)

q_1	q_2	q_3	q_4	q_5	q_6	q_7	q_8	q_9

Finally: Broadcast q_8 and serve $\{\mathbf{q_8}, q_1, q_3, q_9\}$

$P(\exists\ e : e \succeq q \subseteq Q)$: In a multiset Q, the probability that there exists at least one element, which is ancestor of exactly $|q|$ $(|q| - 1 + itself)$ elements of Q.

Analytical Results (Queue of 30 queries)

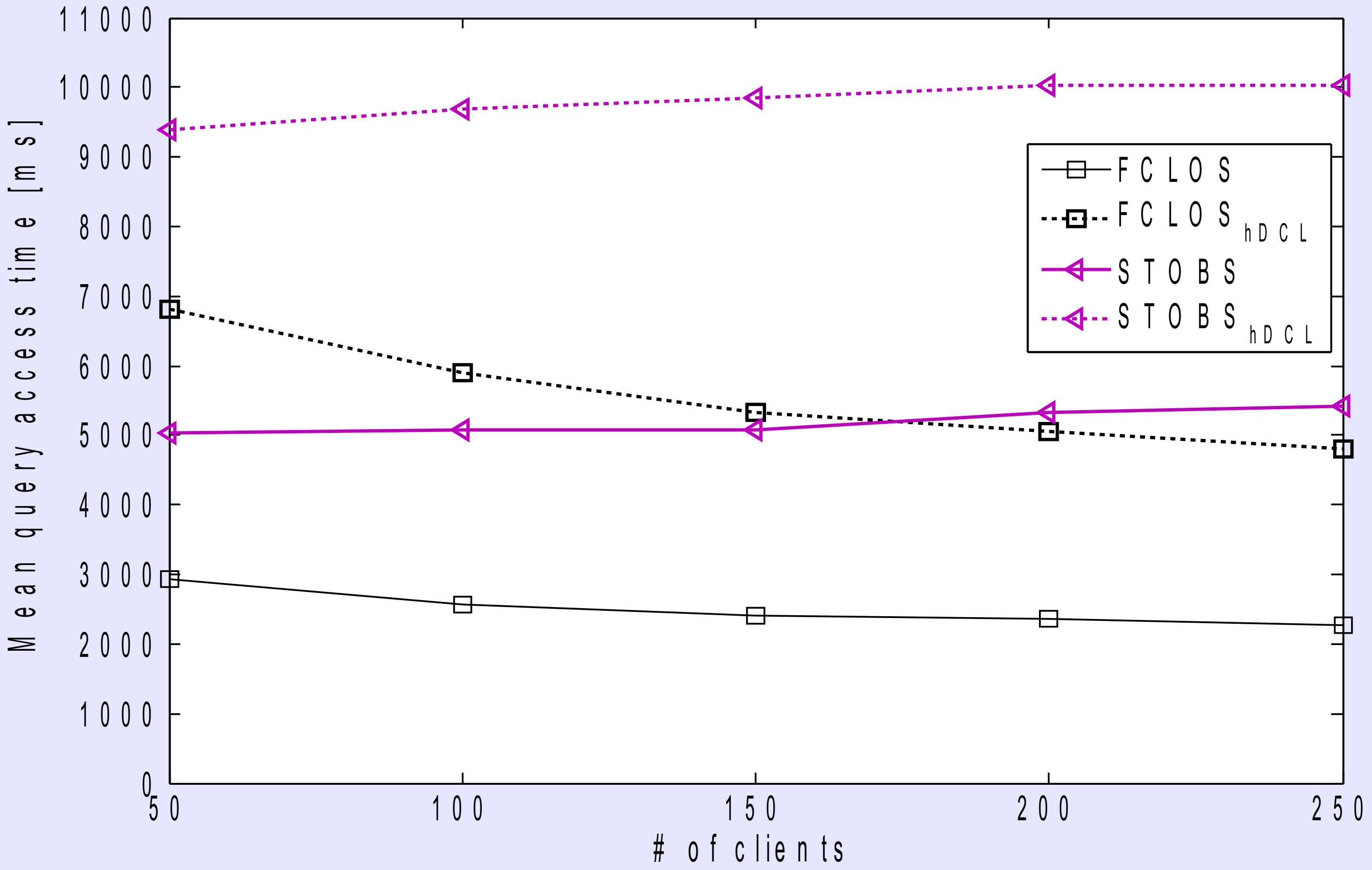


- P: In a multiset/queue Q, the probability that ***a selected*** element of Q is an ancestor of at least |q| elements of Q *(STOBS related)*
- P: In a multiset/queue Q, the probability that ***there is an*** element of Q is an ancestor of |q| elements of Q *(FCLOS related)*

Experimental Environment

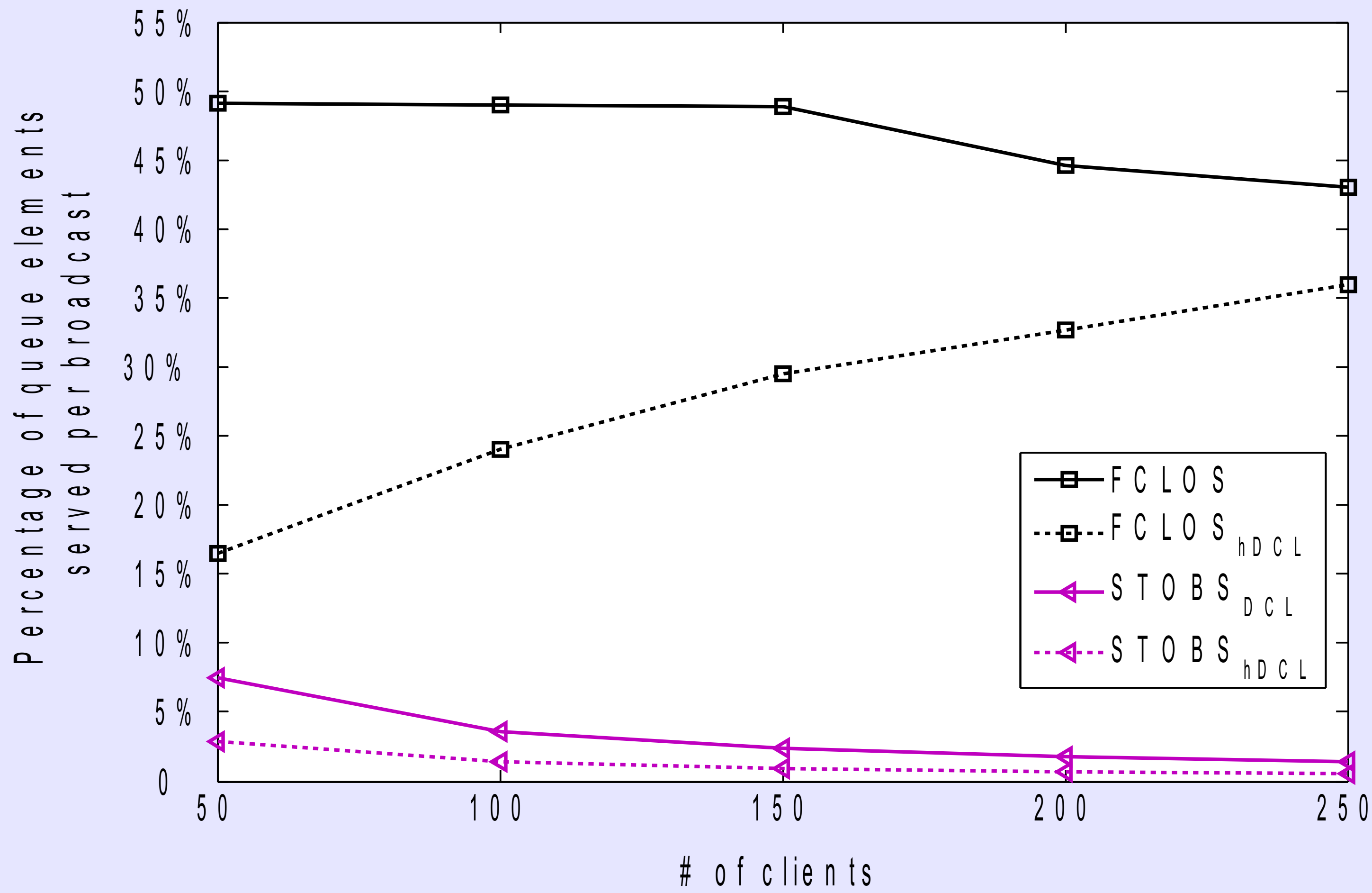
- mOLAP application experiment
- Mobile clients issue OLAP queries
- FCLOS and STOBS evaluated against their respective hDCL extensions
- Datasets used
  - A real data mart provided by an OLAP company
  - Semi-synthetic

Query Access Time



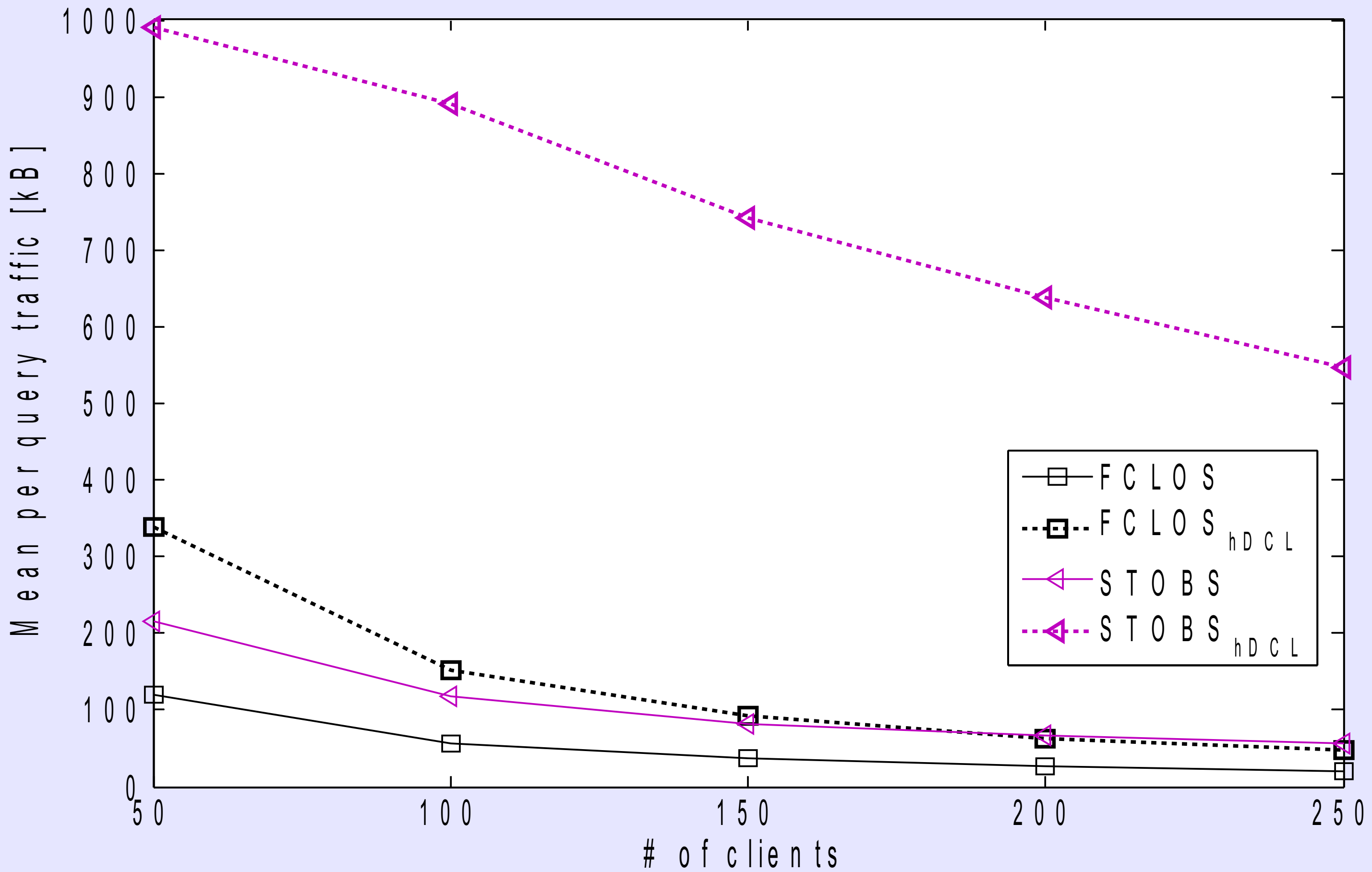
Access time reduced (by around 50%) with the DCL variant for both scheduling approaches

Subsumption Exploitation



- This metric translates more directly to the information provided by the analytical framework.
- **The analytical results are confirmed by the experiments (and vice-versa)**

Per Query Traffic



Per query traffic reduced with the DCL variant for both scheduling approaches

Conclusions

Freie Universität Berlin

- *DCL vs. hDCL in mOLAP:*
  - *DCL* provides optimal derivation exploitation
  - Don't use hierarchies
- mOLAP schedulers:
  - Pre-defining the node to be transmitted and then checking derivations **reduces** the number of simultaneously served requests
  - If every element of the queue is considered as candidate for transmission, the number of simultaneously served requests **increases**

mOLAP Literature

(1) M. A. Sharaf, Y. Sismanis, A. Labrinidis, P. Chrysanthis and N. Roussopoulos. *Efficient Dissemination of Aggregate Data over the Wireless Web*. International Workshop on the Web and Databases (WebDB), pages 93–98, June 2003.

(2) Y. Sismanis, N. Roussopoulos, A. Deligianannakis, and Y. Kotidis. *Dwarf: Shrinking the Petacube*. ACM SIGMOD, pages 464–475, 2002.

(3) I. Michalarias and H.-J. Lenz. *Dissemination of Multidimensional Data Using Broadcast Clusters*. In Distributed Computing and Internet Technology, volume 3816 of LNCS, pages 573–584. Springer, 2005.

(4) I. Michalarias, V. Boucharas and H.-J. Lenz. *Hybrid Scheduling for Aggregated Data Delivery in Wireless Networks*. In Proceedings of the 1st International Conference on Communications and Networking in China, 2006. IEEE.

(5) I.Michalarias and Arkadiy Omelchenko. *Compressed Aggregations for mobile OLAP Dissemination*. In Proceedings of the18th International Workshop on Database and Expert Systems Applications, pages 609-614, 2007.IEEE

Discussion

Freie Universität Berlin

Questions?