Dynamic, Extensible Query Processing in Super-Peer Based P2P Systems[*]

Christian Wiesner Alfons Kemper Stefan Brandl

University of Passau
Germany

1. Introduction

P2P applications have been quite successful, e.g., for exchanging music files, where simple attributes describing these resources are used. A lot of effort has been put into refining topologies and query routing functionalities of these networks. Less effort has been put into extending the query functionalities. Query processing techniques for P2P systems are still inefficient: sending (atomic) queries to the appropriate peers clearly fails for queries which need data from more than one peer to be executed. While quite a few database techniques can be re-used, a P2P data management infrastructure poses additional challenges. Because of the dynamic nature of P2P networks, we can neither assume global knowledge about data distribution, nor are static topologies and static query plans suitable. Unlike in traditional distributed database systems, we cannot assume a complete schema instance but rather work with a distributed allocation schema which directs query processing tasks from one node the neighbors. Therefore, our distributed routing indexes contain both metadata and content information and are hosted by a backbone of super-peers.

2. Extensible Distributed Query Processing

To enable dynamic, extensible, and distributed query processing in super-peer based P2P networks, where both standard query operators and user-defined code can be executed nearby the data, we distribute query processing to the (super-)peers. Therefore, super-peers provide functionality for the management of the index structures, query optimization, and query processing capabilities. Additionally, we expect that peers provide query processing capabilities to be a full member of the P2P network. These query processors can be dynamically extended by operators such that query evaluation plans (QEPs) with user-defined code, e.g., selection predicates, join predicates, etc., can be pushed from the client to the (super-)peers where they are executed. Furthermore, super-peers have to provide an optimizer for generating good query plans from the queries they receive. We utilize these distributed query processing capabilities at the super-peers and distribute the query stated by the user to the corresponding super-peers. This distribution process is guided by the routing index which is dynamic and corresponds to the data allocation schema in traditional distributed DBMSs. However, as the index is dynamic and dispersed, static query optimization is not possible. Thus, query optimization must also be dynamic and based on the allocation schema of the data known at the super-peer.

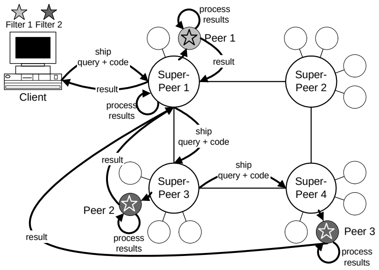

Figure 1. Pushing Code-Carrying QEPs

Figure 1 illustrates a super-peer based P2P network with extensible distributed query processing capabilities. The figure shows the architecture and the flow of messages in our approach where queries and code are pushed through the network. The client sends the query including user-defined operators to the first super-peer where the local routing indices are consulted and the query is split into two parts. The single parts including the user-defined filters are shipped to the neighbors where the plan optimization continues. This way, operators are pushed to the data sources.

3. The Demonstration

We present a P2P architecture using a super-peer topology where research labs and universities register their projects including description, people, papers, etc. Users of this network may ask for possible research partners, i.e., people who are involved in the same project and have written many papers, e.g., about *P2P*. As the data is specified using a *semantic* description, these queries could be executed in such a P2P network much more efficiently than by existing search engines. We demonstrate the processing of queries including user-defined operators and the distributed plan generation. Pushing user-defined code towards the data sources will be illustrated using, e.g., operators for the analysis and filtering of PDF and XML files.

References

[1] I. Brunkhorst, H. Dhraief, A. Kemper, W. Nejdl, and C. Wiesner. Distributed queries and query optimization in schema-based p2p-systems. In *International Workshop on Databases, Information Systems and Peer-to-Peer Computing*, Berlin, Germany, September 2003.

[*] A detailed description of the proposed techniques are presented in [1].