# Management of Metadata in P2P Networks

*Valerie Bönström*

*boen@inf.fu-berlin.de*

*Institut für Informatik*

*Freie Universität Berlin*

## 1. Motivation

Peer-to-peer (P2P) systems have recently become very popular and a lot of research how to route queries in a P2P network has been done. But, especially regarding more complex data sets (such as RDF, XML, etc.) and more complex queries (such as range queries) many questions are still open. Existing approaches are not efficient (see [gnutella]), are based on a very restrictive data model (see [stoica], [rowstron], [ratnasamy]), demand knowledge about the content of the data (such as a global schema etc.) or do not support complex queries on the whole network (see [nejdl]). Therefore an approach is needed, how complex data sets can be efficiently stored to enable efficient processing of complex queries. These queries should be processable on the whole network without previous knowledge about the content of the data.

## 2. The Project

The described problem is considered within the framework of the „DFG-Schwerpunktprogramm: Basissoftware für selbstorganisierende Infrastrukturen für vernetzte mobile Systeme".

Within the project we develop lightweight, adaptive methods to disseminate, exchange and use information in ad hoc networks with mobile devices. Thereby we look at networks, which can frequently change. Our main focus will be the relation of the content and technical aspects in the network.

One main assumption is, that the communication of network members can be improved by using content description of data and services.

## 3. Current Focus

In context with the project my work focuses on how to use metadata and semantics of data to process complex queries in P2P overlay networks. First of all, I adress this problem in the specific context of hierarchically organized data, as popular complex data sets (such as XML documents) are often hierarchically organized. The idea is to store hierarchically organized data in that way, that its structure can be used to efficiently route and process complex queries. Therefore I investigate the hierarchical DHT-based approach to efficiently store hierarchically organized data and process complex queries on the whole network without previous knowledge about the content of the data. I analyze its performance regarding storage and network load compared to pure DHT-based systems and I will discuss possibilities of improvement. In that context tests will be done to analyze the influence on storage and network load of adding a hierarchical structure to a DHT – based network.

But to use the hierarchical organisation of data is only one approach to use metadata to process data in P2P nertworks. Furthermore I will consider and analyze different possibilities how to use, store and process metadata in P2P nertworks.

## 4. Literatur

[gnutella] Gnutella, http://www.gnutella.com

[freenet] Freenet website, http://freenet.sourceforge.net

[stoica] I. Stoica, R. Morris, D. Karger, M. F. Kaashoek, H. Balakrishnan: Chord: A scalable peer-to-peer lookup service for internet applications, Proceedings of the ACM SIGCOMM, 2001

[rowstron] A. Rowstron and P. Druschel: Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer

systems, Proceedings of the of the 18th IFIP/ACMInternational Conference on Distributed Systems Platforms (Middleware 2001), 2001

[ratnasamy] S. Ratnasamy, P. Francis, M. Handley, Karp, S. Shenker: A scalable content-addressable network, Proceedings of the ACM SIGCOMM, 2001

[zhao] B. Zhao, J. Kubiatowicz, A. Joseph: Tapestry: An infrastructure for fault-tolerant wide-area location and routing, http://www.cs.berkeley.edu/ravenben/publications/CSD-01-1141.pdf, 2001

[nejdl] W. Nejdl, M. Wolpers, W. Siberski, C. Schmitz, M. Schlosser, I. Brunkhorst, A. Lser: Super-peer-based routing and clustering strategies for rdf-based peer-to-peer networks, Proceedings of the 12th International World Wide Web Conference, Budapest, Hungary, 2003