

Semantic Overlay Networks

Karl Aberer and Philippe Cudré-Mauroux

School of Computer and Communication Sciences
EPFL – Switzerland
{karl.aberer, philippe.cudre-mauroux}@epfl.ch

1 Introduction

In a handful of years only, Peer-to-Peer (P2P) systems have become an integral part of the Internet. After a few key successes related to music-sharing (e.g., Napster or Gnutella), they rapidly developed and are nowadays firmly established in various contexts, ranging from large-scale content distribution (BitTorrent) to Internet telephony (Skype) or networking platforms (JXTA). The main idea behind P2P is to leverage on the power of end-computers: Instead of relying on central components (e.g., servers), services are powered by *decentralized overlay* architectures where end-computers connect to each other dynamically.

So far, much effort has been devoted to the development of distributed hash tables (DHTs) to index data in a totally decentralized way. Though extremely robust and scalable, these systems suffer from simplistic data models, which mainly consist of unstructured collections of key-value pairs. More recently however, a significant number of uncorrelated research efforts focused on enriching P2P systems with more expressive data models and query languages. As a result, various Semantic Overlay Networks (SON) supporting relational, semi-structured or triple-based collections of data over decentralized P2P networks started to appear. SON systems can be considered as truly unique as they aim at managing structured data in very large-scale, decentralized, heterogeneous and highly dynamic environments.

2 Contents

We start with a brief review of current P2P architectures. We then focus on data integration and query resolution in various SON:

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the VLDB copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Very Large Data Base Endowment. To copy otherwise, or to republish, requires a fee and/or special permission from the Endowment.

**Proceedings of the 31st VLDB Conference,
Trondheim, Norway, 2005**

Data Integration in SON :

- Integration Based on Source Descriptions
- Translation Rules
- Opinion-based Semantic Integration
- Mapping Tables.

Query Resolution in SON :

- Query Rewriting Using Local Views
- Routing Indices
- Distributed Table Lookup
- Semantic Gossiping.

We conclude the tutorial with an overview of current research trends in the area of Semantic Overlay Networks as well as a description of potential future applications:

Current Research Directions :

- Information Retrieval in SON
- Dissemination Services
- Emergent Semantics
- Corpus-Based Semantic Integration.

We illustrate each approach with a high-level description of a particular system. Since decentralization principles were key drivers in ensuring scalability of P2P systems, we also discuss in detail for each system its degree of decentralization (e.g., lack of central component) on both systemic and semantic levels. In the end, we wish this tutorial to represent a solid introduction to recent information management techniques developed for structured data in very large-scale, heterogeneous environments. As IT needs slowly drift from information presentation to information manipulation, one might recognize in Semantic Overlay Networks one of the few viable paradigms capable of underpinning very large scale data processing applications while still relying on standard information infrastructures.