Chapter 19
An Ontology-Based P2P Network for Semantic Search

Tao Gu
University of Southern Denmark, Denmark

Daqing Zhang
Institut Telecom SudParis, France

Hung Keng Pung
National University of Singapore, Singapore

ABSTRACT
This article presents an ontology-based peer-to-peer network that facilitates efficient search for data in wide-area networks. Data with the same semantics are grouped together into one-dimensional semantic ring space in the upper-tier network. This is achieved by applying an ontology-based semantic clustering technique and dedicating part of node identifiers to correspond to their data semantics. In the lower-tier network, peers in each semantic cluster are organized as Chord identifier space. Thus, all the nodes in the same semantic cluster know which node is responsible for storing context data triples they are looking for, and context queries can be efficiently routed to those nodes. Through the simulation studies, the authors demonstrate the effectiveness of our proposed scheme.

INTRODUCTION
In recent years, the use of context information has attracted a lot of attention from researchers and industry participants in ubiquitous and pervasive computing. Users and applications are often interested in searching and utilizing widespread context information. Context information is characterized as an application’s environments or situations (Dey et al., 2000). With the vast amount of context information spread over multiple context spaces and the increasing needs of cross-domain context-aware applications, how to provide an efficient context search mechanism is challenging in the context-aware research community.

One approach is to use a centralized search engine to store context data and resolve search requests. Although this approach can provide fast responses to a context query, it has limitations such as scalability, a single processing bottle-
neck and a single point of failure. Peer-to-peer (P2P) approaches, on the other hand, have been proposed to overcome these obstacles and are gaining popularity in recent years. P2P systems such as Gnutella (Gnutella) and Freenet (Freenet) allow nodes to interconnect freely and have low maintenance overhead, making it easy to handle the dynamic changes of peers and their data. The past years have seen an increased focus on decentralized P2P systems (Han, et al., 2006, Li, et al., 2006, Liu, et al., 2004, Morselli, et al., 2005). However, a query has to be flooded to all the nodes in a network including the nodes that do not have relevant data. The fundamental problem that makes search in these systems difficult is that data are randomly distributed in the network with respect to their semantics. Given a search request, the system either has to search a large number of nodes or run a risk of missing relevant data. Other P2P systems such as Chord (Stoica, et al., 2001), CAN (Ratnasamy, et al., 2001), Pastry (Rowstron, et al., 2001) and Tapestry (Zhao, et al., 2004) typically implement distributed hash tables (DHTs) and use hashed keys to direct a search request to the specific nodes by leveraging a structured network. In these systems, a data object is associated with a key which can be produced by hashing the object name. A node is assigned with an identifier which shares the same space as the keys. Each node is responsible for storing a range of keys and corresponding objects. When a search request is issued from a node, the search message is routed through the network to the node responsible for the key. They can guarantee to complete search in a logarithmic number of steps. Over years, many applications have been developed, such as file sharing (LimeWire) and content distribution (Castro, et al., 2003).

In this system, context data are represented by a collection of RDF (RDF) triples. Peers with the same semantics are grouped together into a semantic cluster in the upper-tier network. All the semantic clusters are constructed as a one-dimensional semantic ring space. This is achieved by dedicating part of hashed node identifiers to correspond to their data semantics. Data semantic is extracted according to a set of schemas. Peers in each semantic cluster can be organized as a structured P2P network such as Chord identifier space in the lower-tier network. Thus, all the nodes in the same semantic cluster know which node is responsible for storing context data triples they are looking for, and context queries can be efficiently routed to those nodes.

The rest of the article is organized as follows. Section 2 presents the detail of the two-tier semantic P2P network. Section 3 evaluates the performance of our system using simulation and presents the results. Section 4 reviews related works, and finally Section 5 concludes our work.

THE TWO-TIER SEMANTIC P2P NETWORK

In this section, we first present an overview of the two-tier semantic P2P network, followed by a description of technical details. For ease of discussion, we use the terms node and peer interchangeably for the rest of the article.

OVERVIEW

In this network, a large number of nodes storing context data are grouped and self-organized into a two-tier semantic P2P network, in accordance with their semantics. A node can act as producer, consumer or both. Producers provide various context data for sharing whereas consumers obtain context data by submitting their context queries and receiving results. Each node maintains a lo-