Resource Discovery Service while Minimizing Maintenance Overhead in Hierarchical DHT Systems

Riad Mokadem, Paul Sabatier University, France
Abdelkader Hameurlain, Paul Sabatier University, France
A Min Tjoa, Vienna University of Technology, Austria

ABSTRACT

Despite hierarchical Distributed Hash Table (DHT) systems have addressed flat overlay system problems, most of existing solutions add a significant overhead to large scale systems. This not only increases the bandwidth consumption but also affect the routing efficiency. This paper deals with a resource discovery while minimizing maintenance overhead in hierarchical DHT systems. The considered resources are metadata describing data sources. In the solution, only one gateway in one overlay is attached to the superior level overlay. It aims to reduce both lookup and maintenance costs while minimizing the overhead added to the system. The authors present a cost analysis for a resource discovery process and discuss capabilities of the proposed protocol to reduce the overhead of maintaining the overlay network. The analysis result proved that our design decrease significantly the maintenance costs in such systems especially when nodes frequently join/leave the system.

Keywords: Distributed Hash Tables, Grids, Hierarchical Model, Peer to Peer Systems, Resource Discovery

INTRODUCTION

A resource discovery is a real challenge in unstable and large scale environments. It consists to discover resources (e.g., computers, data) that are needed to perform distributed applications in such systems (Mastroianni et al., 2007). Throughout this paper, we focus on the discovery of metadata describing data sources in Grid systems.

Classical resource discovery approaches in Grids are either centralized or hierarchical and were proved inefficient as the scale of Grid systems rapidly increases. Excessive access to a centralized node generates bottleneck and its failure paralyzes all the system. The using of web services, inspired from hierarchical models, has been explored in several research works (University of Chicago, 2004). Although
the advantage of being Open Grid Service Architecture (OGSA) (Foster et al., 2004) compliant, i.e., each resource is represented as a web service, this strategy is not adapted for grid environments since the dynamicity node properties in large scale Grids (Pacitti & Valduriez, 2007). Large amounts of research works have been adopted Peer-to-Peer solutions to deal with resource discovery in Grid systems. P2P routing algorithms have been classified as structured or unstructured (Pacitti & Valduriez, 2007). The most popular service provided by P2P unstructured systems is the popular file sharing over the Internet as Gnutella (Klingberg & Manfredi, 2002). Although the good fault tolerance proprieties in these systems, a flooding –used in each search- is not scalable since it generates large volume of unnecessary traffic in the network. Structured Peer-to-Peer systems as DHT are self-organizing distributed systems designed to support efficient and scalable lookups in spite of the dynamic proprieties in such systems. Classical flat DHT systems organize nodes, having the same responsibility, into one overlay network with a lookup performance of O(log(N)), for a system with N peers. However, the using of a flat DHT does not consider the autonomy of virtual organizations and their conflicting interests nor the locality principle, a crucial consideration in Grids (Harvey et al., 2003). Moreover, typical structured P2P systems as Chord (Stoica et al., 2001) and Pastry (Rowston & Druschel, 2001) are very sensitive to the dynamics of the network. They suffer not only from temporary unavailability of some of its components but also from churn. It occurs in the case of the continuous leaving and entering of nodes into the system. Recent research works (Mastroianni et al., 2007) proved that hierarchical overlays have the advantages of faster lookup times, less messages exchanged between nodes, and scalability. They are valuable for small and medium sized Grids, while the super peer model (Yang & Garcia-Molina, 2003) is more effective in very large Grids. Several research works (Garces-Erice et al., 2003; Ganesan et al., 2004; Rhea et al., 2004; Sanchez-Artigas et al., 2005; Zols et al., 2006; Samad et al., 2008; Martinez-Yelmo et al., 2008) proved that hierarchical DHT systems based on the super peer concept can be advantageous for complex systems. A hierarchical DHT employ a multi level overlay network where peers are grouped according to a common property such as resource type or locality for a lookup service used in discovery (Garces-Erice et al., 2003). In this context, a Grid can be viewed as a network composed of several, proprietary Grids, virtual organizations (VO) (Foster et al., 2004) where every VO is dedicated to an application domain (e.g., biology, pathology). Within a group, one or more nodes are selected as super peers to act as gateways to nodes in the other groups. However, most of existing solutions neglect the churn effect and deal only with the improving performance of the overlay network routing. More, they generate significant additional overhead to large scale systems. Xu et al. (2003) demonstrated the high maintenance state needed (memory, CPU and bandwidth) when all peers in the overlay are attached to different levels of the hierarchy. Several proposals for reducing maintenance costs have appeared in the literature. Montresor (2004) proposed an algorithm SG-1 to find the optimal number of super peers in order to reduce its maintenance costs. It is based on the information exchange between super peers on their capacities through a gossip protocol (Artigas, 2006). Also, despite a good strategy to manage a churn in Samad et al. (2008) through a lazy update of the network access points, inter-organizations lookups were expensive because of the complex routing system. We also cited the self organizing distributed algorithm (Zols et al., 2009) in which all decisions taken by the peers are based on their partial view in the sense that the algorithm became fully decentralized and probabilistic. Hence, Most of these solutions add significant load at some peers which generates an additional overhead to large scale systems.

In this paper, we propose a single-gateway based hierarchical DHT solution (SG-HDHT) for an efficient resource discovery in Grids. We focus on the discovery of metadata describing the data sources in Grids. Our solution reduces
SEMCL: A Cross-Language Semantic Model for Knowledge Sharing
www.igi-global.com/chapter/semcl-cross-language-semantic-model/68217?camid=4v1a