The Journal of Systems and Software 84 (2011) 465–478

Contents lists available at ScienceDirect

The Journal of Systems and Software

journal homepage: www.elsevier.com/locate/jss

Resource discovery in a Grid system: Directing requests to trustworthy virtual organizations based on global trust values

K. Karaoglanoglou*, H. Karatza

Aristotle University of Thessaloniki, Department of Informatics, Thessaloniki 541 24, Greece

ARTICLE INFO

Article history:
Received 19 January 2010
Received in revised form 2 September 2010
Accepted 29 October 2010
Available online 11 November 2010

Keywords:
Trust management
Grid system
Global trust values
Directing requests

ABSTRACT

This paper studies the resource discovery problem in a Grid system, in which global trust values play a crucial role. The proposed mechanism suggests that routers and resources comprise virtual organizations (VOs) within a Grid system, where a router controls locally a number of resources in each virtual organization. Global trust values are assigned to the system's VOs. These trust values show whether a VO and subsequently its local resources are trustworthy or not. Our primary goal is to discover the appropriate resource for a specific request and then effectively direct this request to a trustworthy VO that controls locally the appropriate resource. Furthermore, the trust-aware resource discovery mechanism also manages the cases of dynamic changes in the trustworthiness of VOs. For instance, VOs that in the past were untrustworthy could now be trustworthy. The proposed mechanism is capable of detecting these dynamic changes, so that the directing of requests occurs in an up-to-date way. Finally, this paper presents the performance evaluation of the proposed trust-aware resource discovery mechanism by providing a number of simulation tests in Grid systems of different sizes.

1. Introduction

A Grid system is comprised by combined computer resources from multiple administrative domains that are applied to the solution of a demanding problem that requires a great amount of processing power and/or processing of large data-sets. In a more formal definition, a Grid system can be described as "a large-scale, geographically distributed, hardware and software infrastructure composed of heterogeneous networked resources owned and shared by multiple administrative organizations which are coordinated to provide transparent, dependable, pervasive and consistent computing support to a wide range of applications. These applications can perform distributed computing, high throughput computing, on-demand computing, data-intensive computing, collaborative computing or multimedia computing" (Bote-Lorenzo et al., 2004).

The definitions provided above, clearly state that the base of Grid technology is the concept of resource sharing. The shared resources in a Grid infrastructure could vary from plain desktop systems to clusters and from storage devices to large data-sets. Therefore, a main issue in such a system is how to manage all the types of resources and how to provide access to these remote resources either to execute a job or to have access to the resources' data. A mechanism provided by the Grid infrastructure should be available to discover an appropriate resource for such requests. Therefore, one of the main capabilities a Grid infrastructure needs to support is a resource discovery mechanism (Buyya and Venugopal, 2005).

When discovering the resource capable of satisfying a request, there are certain issues that need to be addressed. Suppose a resource provider P is part of a Grid system and is capable of satisfying the request from a resource requestor R. The issues that should be considered in this communication are to protect the local data in P from a malicious attack caused by the components of the request, and to ensure the integrity and secrecy of the request. Since successful transactions in a Grid system are crucial for the integrity of the whole system, it would be beneficiary to direct requests only to trustworthy resource providers. This will significantly decrease the possibility of endangering the integrity of the system.

There are mechanisms capable of dealing with the above implications in resource-sharing environments like sand-boxing (Chang et al., 2000), encryption (Schneier, 1996) and other access control and authentication techniques. These mechanisms, however, incur additional overhead. Incorporation of the trust factor into the resource discovery decisions could significantly enhance the integrity of the system and also minimize the additional overhead. A resource discovery mechanism, aware of the trust relationship between resource providers and resource requestors, can perform the directing procedure of the requests in a way that the possibility of unwanted implications can be minimized. Using and managing trust, Grid technology becomes more appealing. An efficient trust

* Corresponding author.
E-mail addresses: kkaraogl@csd.auth.gr (K. Karaoglanoglou), karatza@csd.auth.gr (H. Karatza).

0164-1212/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jss.2010.10.043

mechanism results in a safer communication between the Grid nodes, and increases the quality of service.

In this paper, we will introduce global trust values for all nodes or in our case VOs. Based on the availability of these values, we propose a trust-aware resource discovery mechanism that can effectively direct requests created in the Grid system to trustworthy resource providers. Moreover, the trust-aware resource discovery mechanism manages effectively the cases of dynamic changes in the trustworthiness of Grid nodes in order to maintain the directing of requests up-to-date.

The paper is organized as follows: Section 2 presents related word on resource discovery and trust management mechanisms. Section 3 describes the VO-based Grid model in which the proposed trust-aware resource discovery mechanism is based, and discusses important trust concepts used throughout the paper. Section 4 analyzes the procedures that the trust-aware resource discovery mechanism uses. Section 5 presents experiments and testing of the proposed mechanism. Finally, Section 6 presents our conclusions and discusses future research directions.

2. Related work

In this section, research work concerning the Grid resource discovery problem in general, as well as implementations of trust mechanisms is presented. Though significant work has been done towards the direction of trusted Grid computing (Song et al., 2005, 2006) focusing to the security implications that could arise in such systems, research concerning management of trust in the context of resource discovery in Grid systems is limited. Due to lack of research work concerning trust mechanisms in Grid systems, in this section we present related work regarding implementations of such mechanisms in P2P systems.

A noteworthy approach to the resource discovery problem is the matchmaking framework (Raman et al., 1998). The matchmaking framework was designed to solve real problems encountered in the deployment of Condor, a high throughput computing system. Several other research papers make use of the matchmaking framework trying to add new aspects in the existing mechanism (Raman, 2001; Tangmunarunkit et al., 2003; Zhu et al., 2004; Tangpongprasit et al., 2005; Maheswaran and Krauter, 2000; Vidal et al., 2006; Castano et al., 2004; Karaoglanoglou and Karatza, 2008). According to this framework, requestors and providers (resources) in a Grid system advertise their characteristics. A matchmaking service is responsible of finding a match between the advertisements and informing the relevant entities of the match.

In recent years, significant interest has focused on two systems: Grids and P2P systems. Both systems share the same main idea. They are both resource-sharing environments. Their difference is that they have followed different evolutionary paths. Grid systems are mainly used in complex scientific applications, while P2P systems are developed around mainstream services such as file-sharing. Research papers, concerning that field, suggest the use of existing protocols developed for P2P systems into Grid systems (Iamnitchi et al., 2002; Al-Dmour and Teahan, 2005; Ali et al., 2005; Talia et al., 2006; Basu et al., 2005; Zerfiridis and Karatza, 2003; Crespo and Garcia-Molina, 2002).

Another notable approach to the resource discovery problem is the semantic communities one (Li and Vuong, 2005; Zhu et al., 2005; Somasundaram et al., 2006). Motivation behind the semantic communities approach is that Grid communities and human communities consist of members that are engaged in sharing and communication. Main target in this approach is to create Grid communities based on similar-interests policies allowing community nodes to learn of each other without relying on a central meeting point.

Research work concerning the field of trust mechanisms is mainly focused to the ways trust is calculated in P2P systems. Several research papers provide the so-called trust functions or trust metrics in order to compute trust values for the nodes comprising P2P systems. Given the similarities between Grid and P2P systems, these trust functions can be adjusted for use in Grid systems.

In the 'NICE' trust model (Lee et al., 2003), after each interaction, a client-entity signs a cookie stating the quality of the transaction to the server-entity. The server-entity uses this signed cookie to prove its trustworthiness to other entities in the system. 'NICE' trust model urges entities to form cooperative groups. Entities from different groups may have different evaluations of trust for the same server-entity.

'PeerTrust' (Xiong and Liu, 2002, 2003) model is designed for decentralized P2P electronic communities. This trust model takes into consideration three factors: amount of satisfaction, number of interactions, and balance factor of trust. The amount of satisfaction a peer receives regarding its service is resulting from the interactions other peers have had with this peer. The better a peer fulfils its part of the service, the more satisfaction it will receive from the other peers. 'PeerTrust' utilizes the P-Grid storage structure (Aberer, 2001) so that information about transactions and complaints can be retrieved by any entity.

In the 'EigenTrust' model (Kamvar et al., 2003) the global reputation of each peer is given by the local reputation values assigned to this peer by other peers, weighted by the global reputations of the assigning peers. Normalizing local reputation values in a sensible manner so that malicious peers cannot easily subvert the system and using an efficient algorithm to aggregate the local reputation values, the model concludes to the global reputation value of a peer.

It is not in our intentions to develop a new trust metric for computing trust values for the nodes or resources of Grid systems, though some preliminary work towards incentive-based strategies for encouraging cooperation is presented. This paper is mainly focused in proposing and presenting an efficient way of dealing with global trust values of Grid nodes and how these values can be managed effectively for solving the Grid resource discovery problem.

3. Depiction of the Grid environment

3.1. The Grid-Router model

A Grid system can be seen as an environment comprised by routers and resources (Karaoglanoglou and Karatza, 2008; Li et al., 2002). Each router is in charge of its local resources and also connects with other routers within the Grid system. Fig. 1 presents a Grid system based on the Grid-Router model. The system is comprised by three routers, where each one controls its local resources. Fig. 1 also shows the way that the routers are connected (Router 1 connects directly to Routers 2 and 3).

In order to direct a request efficiently, each router in the Grid system maintains a Routing Table with size equal to the number of different resources in the network. Each data element in that table is the minimum distance measured in hops from that router to all the resources available in the system. The drawback in this approach is that the size of the Routing Tables could become unbearably large in cases of systems with a large number of different resources. It is obvious that maintaining information regarding the distances from every router to every different type of resource available, poses serious limitations for the system's scalability.

3.2. Achieving scalability

Instead of maintaining the distances to all resources available in the system, thus hindering the system's scalability, the proposed