ASSESSING REPUTATION AND TRUST IN SCIENTIFIC GRIDS

Andrea Bosin, Nicoletta Dessi
Dipartimento di Matematica e Informatica,Universita' degli Studi di Cagliari
72,Via Ospedale, Cagliari, Italy

Balachandar Ramachandriya Amarnath
Dipartimento di Matematica e Informatica Universita' degli Studi di Cagliari, 72,Via Ospedale, Cagliari, Italy
and INFN sezione di Cagliari, Monserrato,Italy

Keywords: Grid Computing, Trust, Reputation, Data mining.

Abstract: Up to now, reputation and trust are widely acknowledged to be important for business environments, but little attention has been placed in security aspects of Grids devoted to scientific applications. This paper addresses these aspects and proposes both a Reputation Management System and a mechanism for assessing the reputation of computing resources belonging to a scientific Grid. Because the overall performance of the Grid depends on the quality of service given by each single resource, the resource reputation is a measure of trustworthiness, in the sense of reliability and is asserted on a set of properties, namely the resource operative context, that express the resource capability in providing available and reliable services. Unlike the business applications, we are not interested in assessing the reputation of a single resource but in finding a set of resource that have similar capabilities of successfully executing a specific job. Toward this end, the paper proposes a mechanism that assesses reputation by clustering Grid resources in groups that exhibit similar behavioural patterns and share similar operative contexts. Simulation results show the effectiveness of the proposed approach and the possible integration of such a model in a real Grid.

1 INTRODUCTION

Grid computing involves the formation of virtual organizations where different institutions pool together their computational resources and diverse and large groups of geographically distributed users seek to share and use networked resources in coordinated fashion (Foster et al,2002). From user's perspective, a Grid is a collaborative problem solving environment that must guarantee the quality of job execution. In scientific Grids, a Grid scheduler hides the complexities of the Grid computing environment from a scientist. It discovers resources that the user can access, maps application jobs to resources, starts execution and monitors the execution progress of tasks. As such, the Grid scheduler provides a reliable local distributed execution environment and maximizes resource utilization of a local site, but it does not enforce absolute control over the reputation of distributed resources. Usually, reputation is what is generally said or believed about a person or thing character or standing . In business transactions, the concept of reputation has been proposed as a mechanism for addressing the problem of evaluating whether a network node acts consistently. Usually operative on P2P networks, a reputation system calculates the global reputation score of a peer by considering feedbacks from all other peers who have interacted in the past, with this peer. In this paper, we apply a similar reference scheme for evaluating reputation in scientific Grids: reputation is not a static property of a single resource but rather it is an assessment based on experience that is shared through the Grid and it augments and decays with time and frequency of interactions. Due to the diversity of computational resources and the complex interactions that can occur between different nodes, analytical modelling of resource reputation appears impracticable and often not effective since the Grid scheduler is not interested in assessing the reputation of each single resource but in finding a set of resources that have similar capabilities of successfully executing a

Bosin A., Dessi N. and Ramachandriya Amarnath B. (2008).
ASSESSING REPUTATION AND TRUST IN SCIENTIFIC GRIDS.
In *Proceedings of the Tenth International Conference on Enterprise Information Systems - ISAS*, pages 235-238
DOI: 10.5220/0001691202350238

specific job. Towards this end, this paper proposes an heuristic mechanism that computes the single resource reputation by considering a set of attributes that we refer as Resource Operative Context **(**ROC**)** . Usually derived from log-files, ROC attributes allows for evaluating the resource behaviour with time. To avoid information queries on single resource behaviour and to obtain information faster and more reliable, we propose to cluster resources that exhibit similar behavioural patterns and share similar operative contexts. Based on this information, the Grid scheduler may limit the number of resources that are potentially capable of providing efficient job execution. The paper presents preliminary results concerning the implementation of the proposed approach in a simulated Grid environment.

2 RELATED WORK

A survey of existing methodologies for trust and reputation in grid environment is presented in (G.C. Silaghi et al, 2007) that discusses many reputation based trust management systems and their suitability for grids. The complexities of enterprise grid environment and the importance of QoS in grid are discussed in (D.A.Menasce, & E.Casalicchio, 2004) that remarks the need of considering SLAs and cost constraints in the grid scheduling. Concepts related to trust and reputation are validated by mathematical definitions in (W.Xinhua et al, 2005). (B.Ma et al, 2006) present an approach to compute and compare the trustworthiness of entities in the same autonomous and different domains. In (Thamarai Selvi Somasundaram et al,2007), a general purpose trust management system for computational grids is presented , based on several information that can be obtained directly by the grid scheduler. A model of trust aspects in executing collaborative distributed services is presented in (C. Argiolas et al. ,2008)

3 ASSESSING REPUTATION

The reputation of a resource reveals its reliability in executing jobs. Previous section emphasizes the importance of resource's past behaviour for assessing its reputation. However, in an open and decentralized scientific Grid, there is not a centralized authority which collects and maintains reputation information. Additionally, users only submit their jobs and are not asked for a feedback. This makes it impossible for the grid scheduler to have updated reputation information about the whole network, since some Grids may have thousand of resources. Possible approaches could be using distributed data structures or evaluating reputation by local knowledge on the interested resource. Our proposal considers the evaluation of the reputation as a centralized activity supported by a Reputation Management System (RMS) , depicted in figure 1, that periodically interacts with every grid resource. RMS is intended to support the grid scheduler in selecting resources by assessing their reputation and to calculate the delay in executing each user job. The RMS can be configured to be either proactive, active, or passive. In a proactive configuration , the RMS will actively stop all communication with resources that are not conformant to the level of reputation required by the Grid scheduler. With active and passive status, the RMS will not intervene as strongly as with proactive configuration but it will report non compliant resource behaviours to the Grid Scheduler. In a passive configuration, the RMS will only log task behaviour while the resource conformance to the agreed reputation level will be verified at a later time. The RMS includes a Performance Module (PM) that estimates the resource's reputation as follows. It takes periodic data on resources from the scheduler log file and structures this information in relational tables whose attributes express basic characteristic of the resource behaviour that we globally refer as the Resource Operative Context (ROC). Attributes differ in a number of ways. For example , they can be of different types (i.e. quantitative or qualitative) and may contain explicit relationships to one another. A Database Management System is responsible for the storage and the management of ROC information. Hence, each ROC is an instance of a relation table that represents the resource operative status with respect to time. As mentioned earlier, we are not interested in evaluating the most accurate reputation of each single resource, but in clustering resources that exhibit similar behaviours in executing jobs. To reduce the effort required to achieve this, the RMS uses an application that consists of a computation that queries periodically the databases and carries out further analysis on the retrieved data. Specifically, this analysis is based on a more general technique known as instance-based learning which uses retrieved data as a training set for the k-means algorithm (J.A.Hartigen & M.A.Wong, 1979) which is a popular clustering technique. Taking into consideration the scale of the considered attribute, each cluster can be labelled by a qualitative reputation level (i.e. very high, high, medium, sufficient, bad) that classifies each resource according to its past behaviour.

To assure better availability and reliability in executing job, the RMS interacts with the Grid scheduler that periodically asks for resource reputation levels as a decision support parameter in scheduling resources for task submissions.

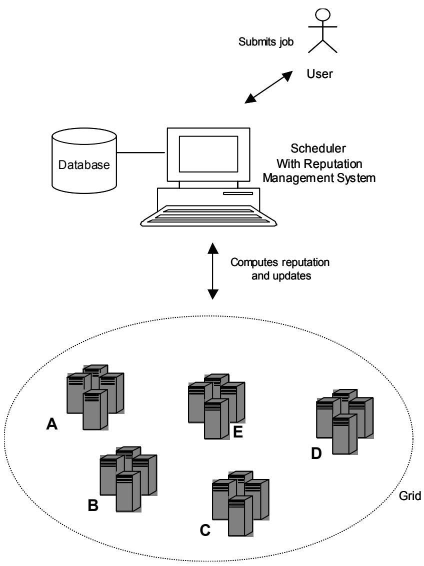

Figure 1: The Reputation Management System.

4 A CASE STUDY

Towards a validation of the proposed approach, we implemented and tested a prototypical version of the RMS in a scientific Grid that we simulated by using GridSim (S. Venugopal et al., 2004).We simulated 25 grid resources and submitted 1500 jobs to each resource. For the purpose of simulation, we created grid resources that are similar in nature with respect to operating system (in our case, Solaris operating system), and consist of same amount of computing power. Moreover, we simulated jobs with similar complexity so that every job has the same length, file size and also produces an equal volume of output. Even if not fully compliant with a real scientific Grid environment, these requirements allows us to evaluate the behaviour of each individual resource in executing jobs. The RMS database was simulated by creating a *resource_model* file for every resource. For the sake of simplification, we considered two attributes namely the number of successful jobs and the execution time taken by a resource to complete a job. The number of successful jobs determines the reliability and quality of the resource in executing a given job successfully. In real grid environment, a job will fail due to many reasons that often depend on the improper maintenance of the site by the resource provider. This behaviour is highly unwanted in the grid environment and it will affect the reputation of the resource significantly.

The execution time is the amount of time taken by the grid resource to complete the execution of the submitted job. This attribute reveals the reliability of the resource in executing the job in time because, during the registration in a grid, a participating resource provider would agree to contribute a fixed amount of computing power. However, if it fails in providing the committed computing power while executing a job, the execution will take more time than estimated. This delay in job execution will eventually lower the reputation of the resource.

Step 1: Introducing Job Failures

Our first experiment aimed at assessing the reliability of a resource with respect to successful execution of submitted jobs. We submitted 1500 jobs to every resources using traditional First Come First Served scheduling mechanism and we randomly simulated jobs whose execution will fail by a resource. Figure 2 shows the experiment results.

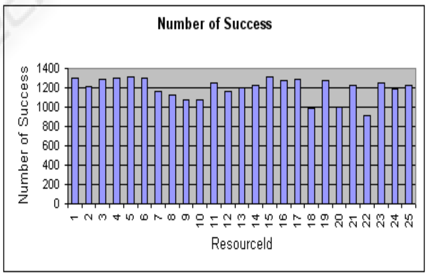

Figure 2: Number of Successes by each resource.

Step 2 : Introducing Delay in Job Execution

In our second experiment we tried to evaluate the reliability of the resource with respect to executing the job in time. With our experience in grid, some computing resources shut down during job execution and this issue originates a delay that affects the reputation of the resource. In this experiment, we simulated random delays in job executions as follows. The resource with id 25 was given no delay and it executed all the jobs in 331ms. The delay caused by the remaining resources was evaluated with respect to this execution time, being jobs of similar complexity. Figure 3 shows the results of this experiment.

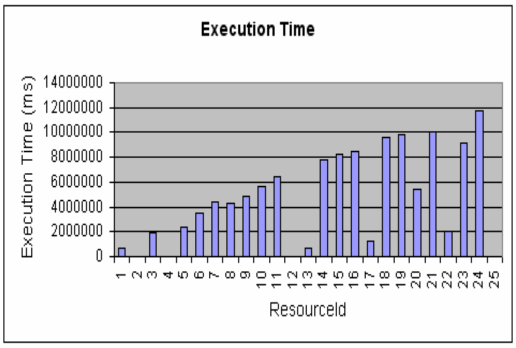

Figure 3: Execution time taken by each resource.

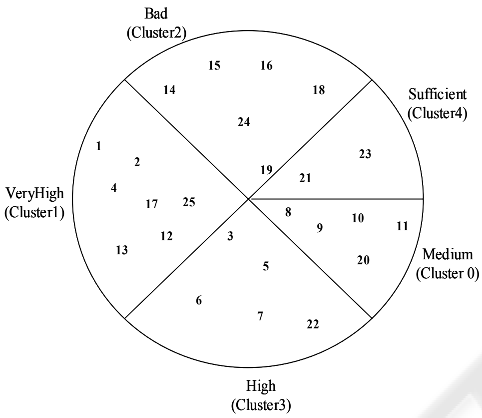

Figure 4: Reputation Levels.

Step 3: Assessing Reputation Levels

Results of previous steps provided a basis for grouping resources into five clusters, each labelled with a reputation level (i.e. very high, high, medium, sufficient, bad). In detail, we stored both the values of success and execution time attributes in a single dataset that served as a training set for the application of the k-means algorithm. Every cluster represents the category of resource which will eventually help the grid scheduler to make resource selection based on their reputation score. Figure 4 shows the experiment results.

5 CONCLUSIONS

The key objective of this paper was to incorporate the reputation in decision making for scheduling resources . The proposed RMS considers the past resource behaviour and takes advantage from a data mining approach to extract knowledge from the operative context that measures the resource capability in providing available and reliable services. The merit of the proposed approach lies in the fact that the reputation is automatically asserted from data that directly incorporates information about the resource behaviour. As future work, we plan to further specify possible properties of the ROC and to implement the proposed reputation model in a real scientific Grid.

ACKNOWLEDGEMENTS

Part of this work was supported by MIUR under the project PON-Cybersar.

REFERENCES

Foster, I., Kesselman, C, Jeffrey M. Nick, & Steven Tuecke (2002). The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration, *A Draft Document*. Version: 6/22/2002

G.C. Silaghi, A.E. Arenas & L.M.Silva (2007). Reputation-based trust management systems and their applicability to grids, *CoreGRID Technical Report* Number TR-0064.

W.Xinhua, G.Xiaolin, & C.Feifei (2005). A Trust and Reputation Model of grid Resources for Cooperating Application, *Proceedings of the First International Conference on Semantics, Knowledge, and Grid SKG 2005.*

S. Venugopal, R. Buyya & L. Winton (2004). A Grid Service Broker for Scheduling Distributed Data-Oriented Applications on Global Grids, *Proceedings of the 2nd International Workshop on Middleware for Grid Computing* , ACM Press, USA.

Thamarai Selvi Somasundaram, R. et al. (2006). A Framework for Trust Management System in Computational Grids, *Proceedings of the 2nd National Conference on Advanced Computing (NCAC 2007)*, Anna University, India.

D.A.Menasce, & E.Casalicchio (2004), QoS in Grid Computing, *IEEE Internet Computing*, 85-87.

B.Ma, J.Sun, & C.Yu (2006). Reputation-based Trust Model in Grid Security System, *Journal of Communication and Computer*, USA, Volume 3, No.8 (Serial No.21).

J.A.Hartigen & M.A.Wong (1979). A K-Means Clustering Algorithm, *Journal of Applied Statistics*, Volume 28, pp 100-108.

C.Argiolas,N.Dessì,M.G.Fugini(2008) .Modeling Trust Relationships in Collaborative Engineering Projects, *Proceedings of the 7nd International Workshop on Conceptual Modelling Approaches for E-Business (eComo 2008), April 22-25 2008 ,Klagenfurt (Austria).*