ABSTRACT

Grids can form the basis for pervasive computing due to their ability of being open, scalable, and flexible to various changes (from topology changes to unpredicted failures of nodes). However, such environments are prone to failures due to their nature and need a certain level of reliability in order to provide viable and commercially exploitable solutions. This is causing nowadays a significant research activity which is focused on the topic of achieving certain levels of Quality of Service (QoS) in highly unreliable environments (such as mobile and ad hoc Grids). This study will focus on the state-of-the-art analysis of the QoS aspects in Grids and how this is achieved in terms of technological means. A small survey and related work will be also presented. A more detailed analysis on the features of unreliable environments such as mobile Grids will be described. An innovative and efficient mechanism will be described, which is especially designed for such environments, in order to enhance them with the QoS attributes of reliability (fault tolerance through replication of tasks) and service differentiation to the Grid users through a simple task prioritization scheme. The results that this recent research work is presenting are promising for the future advancement of Grid commercialization in such environments.

INTRODUCTION

Grid computing has recently migrated from traditional high performance and distributed computing to pervasive and utility computing based on the advanced capabilities of the wireless networks and the lightweight, thin devices. This has as result the emergence of a new computing paradigm which is
Mobile Grid. Mobile Grid is a full inheritor of Grid with the additional feature of supporting mobile users and resources in a seamless, transparent, secure and efficient way. It has the ability to deploy underlying ad-hoc networks and provide a self-configuring Grid system of mobile resources (hosts and users) connected by wireless links and forming arbitrary and unpredictable topologies. However, it is also the basis and the enabling technology for pervasive and utility computing due to the ability of being open, highly heterogeneous and scalable.

This modern approach that combines thousands of parts into a large system provides generally less reliable platforms, which in combination with the long running codes results into application execution times that exceed the mean time to failure of the machines. For this reason fault tolerance is of vital importance in this new mobile grid paradigm since both mission-critical systems and computational intensive applications belong in the context of diverse, dependable and cross-organizational environments which is the case of the emerging mobile Grid. Mobile Grid computing is a typical example of highly unreliable computational environment and thus it is not expected to be fault free, despite the fact that individual techniques such as fault avoidance and fault removal (Lyu 1995) may additionally be applied to its resources. Therefore fault tolerance mechanisms need to be deployed to allow the Grid system to perform correctly in the presence of faults enhancing it with the appropriate reliability.

We present a fault tolerant model for task scheduling in mobile Grid systems based on the task replication concept. The basic idea is to produce and schedule in the Grid infrastructure multiple replicas of a given task. The number of replicas is calculated by the Grid middleware and is based on the failure probability of the Grid resources and the policy that is adopted for providing a specific level of fault tolerance. The adopted replication model is based on static replication (Nguyen-Tuong, 2000), meaning that when a replica fails it is not substituted by a new one. The introduction of task replicas causes an overhead in the workload that is allocated to the Grid for execution. Scheduling and resource management are important in optimizing Grid resource allocation (Ramamritham, Stankovic, & Shiah, 1990), and determining its ability to deliver the negotiated Quality of Service (QoS) to all users (“Scheduling Working Group,” 2001; Wang, Ramamritham, & Stankovic, 1995). The idea in this paper is to handle the additional tasks with a resource management scheme based on the knapsack problem formulation (Pisinger, 1995), where to each task a weight and a profit for its correct execution have been assigned. By this we allow for an efficient time scheduling of the tasks and their replicas so as to maximize the utilization of the Grid resources and the profit we can gain from the successfully executed ones. Moreover a prioritization scheme for the tasks is applied in order to allow an efficient scheduling complementing the QoS attributes of the Grid.

The remaining of the paper is structured as follows. Section 2 gives a short discussion on how QoS is perceived in Grid environments and what are the issues that affect it. Section 3 presents related work in the field of the fault tolerance in Grid computing as well as in Grid scheduling and formulation as Knapsack problem. Section 4 gives an overview of the problem formulation and the notation that is followed in the paper. Section 5 provides the concept and analysis of the task replication model adopted for achieving the desired fault tolerance in the Grid infrastructure. Section 6 enhances the qualitative parameters of Grid operation by introducing a simple and efficient prioritization scheme for the task scheduling. In section 7 we present the simulation results in terms of their efficiency as well as the overall evaluation of the proposed scheme in terms of performance for profit optimization. Finally, section 8 concludes with a discussion on future research and potentials for the current study.
Related Content

Software-Defined Networking for Scalable Cloud-based Services to Improve System Performance of Hadoop-based Big Data Applications
www.igi-global.com/article/software-defined-networking-for-scalable-cloud-based-services-to-improve-system-performance-of-hadoop-based-big-data-applications/153967?camid=4v1a

A Next Generation Technology Victim Location and Low Level Assessment Framework for Occupational Disasters Caused by Natural Hazards
www.igi-global.com/article/next-generation-technology-victim-location/52050?camid=4v1a

Multimedia Services Offer Mixing Telco and Internet Assets
www.igi-global.com/chapter/multimedia-services-offer-mixing-telco/64571?camid=4v1a

A Push-Based Prefetching for Remote Caching RAM Grid
www.igi-global.com/article/push-based-prefetching-remote-caching/37509?camid=4v1a