On maximizing reliability of grid transaction processing system considering balanced task allocation using social spider optimization

Dharmendra Prasad Mahato¹, Ravi Shankar Singh
Department of Computer Science and Engineering, Indian Institute of Technology (BHU), Varanasi 221005, India

1. Introduction

Grid computing is a geographically distributed architecture providing a large number of heterogeneous resources. Maximizing reliability in these environments is a complex task. The grid technology is limited not only to scientific computing but also it is expanding in large scale to business applications by reducing the users' interventions as much as possible [1]. The problem is further complicated by the fact that these resources may fail at any point of time. Load balanced task allocation becomes a challenging job in such a complex and dynamic environment as both the application and computational resources are heterogeneous. Thus a load balanced task allocation strategy which allocates the transactions based on failure possibility of the grid constituents becomes very important for the reliable execution of the grid transaction. This paper deals with the problem of load balanced task allocation in grid transaction processing system with the goal of optimizing the system availability, reliability, and miss ratio. The excessive load on the system can degrade the performance as well as the reliability of transaction processing in such environment. A lot of work has been done on reliability maximization while task allocating with load balancing and without load balancing in different distributed computing environments including the grid system. But, in the literature, we rarely find any example of load balanced task allocation in grid computing system. Maximization of reliability by optimal task allocation in a grid computing system deserves a special attention. The method should be implemented by determining the node at which a specific transaction is to be executed in order to optimize the reliability. As per our knowledge and literature review, no methods or algorithms seem to be available in the literature for this problem.

A transaction which is a short sequence of interactions with the database uses operations such as finding a record or modifying an item by representing one meaningful activity in the user’s environment [2]. In grid computing system, a transaction has a peculiar nature regarding their computation operation and their completion on various nodes in the system. Transaction processing by using grid computing system consists of various service calls executed by different peers of the grid [3]. When the number of users increases beyond limit, the grid computing may be used to solve the problem of scalability and reduced completion time. Transactions in e-commerce, VisaNet, PayPal and financial trading systems are the kind of transaction processed with the grid. Transaction processing in a grid computing system deserves a special attention. The method should be implemented by determining the node at which a specific transaction is to be executed in order to optimize the reliability. As per our knowledge and literature review, no methods or algorithms seem to be available in the literature for this problem.

This paper deals with the problem of task allocation in the grid transaction processing system. There has been quite some research on the development of tools and techniques for grid computing systems, yet some important issues, e.g., service reliability with load balanced transaction allocation in grid computing system, have not been sufficiently studied. Load balanced transaction allocation becomes a challenging job in such a complex and dynamic environment as both the application and computational resources are heterogeneous. The problem is further complicated by the fact that these resources may fail at any point of time. The problem of finding an optimal task allocation solution is known to be an NP-hard. We propose grid transaction allocation based on social spider optimization (LBGTA_SSO) method for this problem. The LBGTA_SSO is based on the cooperative behavior of social spiders to find a collection of task allocation solutions. We also derive reliability formulae for grid transactions considering resource availability. For comparison we modify some existing algorithms to obtain the task allocation algorithms. The results show that our algorithm works better than the modified existing algorithms.

ARTICLE INFO

Keywords:
Reliability
Social spider optimization
Load balanced task allocation

ABSTRACT

This paper deals with the problem of task allocation in the grid transaction processing system. There has been quite some research on the development of tools and techniques for grid computing systems, yet some important issues, e.g., service reliability with load balanced transaction allocation in grid computing system, have not been sufficiently studied. Load balanced transaction allocation becomes a challenging job in such a complex and dynamic environment as both the application and computational resources are heterogeneous. The problem is further complicated by the fact that these resources may fail at any point of time. The problem of finding an optimal task allocation solution is known to be an NP-hard. We propose grid transaction allocation based on social spider optimization (LBGTA_SSO) method for this problem. The LBGTA_SSO is based on the cooperative behavior of social spiders to find a collection of task allocation solutions. We also derive reliability formulae for grid transactions considering resource availability. For comparison we modify some existing algorithms to obtain the task allocation algorithms. The results show that our algorithm works better than the modified existing algorithms.
ing short-running computations on the data volumes, and a long-lived transaction which consists of operations representing long-running computations on the data volumes [3].

Transaction processing, while using business applications in grid technology, is widely needed in order to make them reliable and real-time, with several transparencies on location, replica, concurrency, and failure, as an effective means by sharing a large number of resources among different organizations [4]. Grid transaction processing system is the platform that needs to share, access, transport, process, and manage large data collections distributed worldwide. They give large e-commerce and financial transaction support to its customers to handle operations where speed and scalability are essential. They combine high-end computing technologies with high-performance networking and wide-area storage management techniques. In such a platform, transaction processing plays a vital role. In grid transaction processing system, the transactions arrive at any node and may execute at several nodes before committing [4–6]. Due to dynamic nature of grid computing system, some nodes add while some nodes leave the system. Therefore some nodes are heavily loaded while others are lightly loaded. As a result, the waiting time at heavily loaded nodes increases [7] and therefore most of the transactions have to face rollback or abort consequences. When the queues at the heavily loaded nodes are full, transactions’ requests cannot be added further to the queues of these nodes. The situation in this scenario causes unavailability which usually occurs during the peak time when several transactions’ requests arrive in a short period. As a result, most of the transactions cannot be completed in time because they have their respective deadlines to meet. Therefore, load balanced task allocation becomes important for transaction processing in grid computing system [8].

The load balanced task allocation can be performed by two distinct ways; The first way balances the load of the system after task allocation which creates interprocessor overhead. Due to this overhead, the reliability of the system starts decreasing. In this situation, the system most likely requires an optimal task allocation method [9]. But, for transactions in grid computing systems which have deadline constraints, this approach is not suitable as the delay due to the interprocessor overhead in the system is caused by the deadline-miss fault [10]. Whereas the second way balances the load before task allocation where the delay remains minimum. The failure caused due to the deadline-miss fault in this case is minimum as compared to the former way of the task allocation. Thus, load balanced task allocation is such a method which allocates the tasks after balancing the loads of the assigned nodes. By using this method, the availability of resources is maximized reducing the amount of downtime. Maximization of resource availability maximizes the reliability of the system [11].

To achieve performance, reliability and consistency, data must be readily accessible in a data warehouse, backup procedures must be in place and the recovery process must be in place to deal with system failure, human failure, computer viruses, software applications or natural disasters. The reliability analysis and evaluation in grid transaction processing system becomes a challenging issue and attracts more and more attentions of researchers. Therefore, load balanced task allocation mechanism for maximizing reliability in the grid processing system runs in the following order: (i) it receives new incoming transactions with their assigned deadlines, (ii) it checks for available resources, (iii) it selects the appropriate nodes according to availability and performance criteria and (iv) it produces a planning of jobs to selected nodes within their deadlines [12], (v) it finds out the availability, and lastly (vi) it finds out the reliability.

In this paper, we propose a load balanced grid transaction allocation algorithm based on social spider optimization technique. Social spider optimization [13] technique is a novel swarm optimization technique which is based on the simulation of the cooperative behavior of social spiders. In this optimization technique, the individuals simulate a group of spiders which interact with each other based on the biological laws of the cooperative colony. This technique is able to combine the power of general meta-heuristics with problem-specific heuristics to search a large possible solution within a reasonable computing time. To illustrate the proficiency and robustness of the proposed algorithm, it is compared with other well-known evolutionary methods.

The remainder of this paper is organized as follows. Section 2 reviews the related work in the field of task allocation, load balancing and transaction processing system. Section 3 presents the reliability-aware load balanced task allocation model. Section 4 gives notations, definitions, and assumptions used in the paper. Section 5 formulates the load balanced task allocation problem for maximizing system availability and reliability. Section 6 gives a brief introduction to the SSO algorithm along with our proposed algorithm. Section 7 discusses in detail how to apply the proposed algorithm to solve the problem. Section 8 reports experiment results. Section 9 concludes the paper and point our future work, and Section 10 gives the compliance with ethical standards regarding the research in this paper.

2. Related work

Load balancing [14–18] is an important issue in distributed computing system, because the resources are geographically distributed computers or clusters which are aggregated logically to serve as a unified computing resource. Load balancing technique, when applied in a distributed system like grid, produces services with high availability, scalability, and predictability characteristics [19–21]. The conventional load balancing algorithms can not be harnessed successfully due to the obstacles caused by the real-time, heterogeneity, autonomy and dynamic nature of the transaction oriented grid service [22].

Maximization of reliability in such an environment using task allocation is another issue. Most of the solutions of this problem in the literature are based on simple system models, which do not accurately reflect load balancing and transaction in grid computing systems. In these systems, the load burden and subsequently failures of nodes, networks, and the deadline-miss are unavoidable and may have an adverse effect. Therefore, experiments in [23–25] show that the consideration of reliability in these systems is essential for the next generation of accurate and efficient schedules.

Shatz et al. [23] dealt with the task allocation problem, an essential phase in distributed software design, with the goal of maximizing the system reliability. A quantitative problem model, algorithms for optimal and suboptimal solutions, and simulation results were provided and discussed. Shatz et al. [26] proposed reliability-oriented task-allocation for heterogeneous distributed-computer systems with hardware redundancy. Redundancy of both processors and communication links was considered. A quantitative allocation model was derived and used to present and discuss models and algorithms for systems with level-two or level-three redundancy.

Yen and Chen [27] developed methods for assessing the reliability of the real-time system in terms of the timeliness and accuracy of the output, to aid decision making. They used two commonly coordination structures, parallel and pipeline. Some adaptive schemes were also used.

With the goal of maximizing the reliability, Attiya and Hamam [28] addressed the problem of task allocation in heterogeneous distributed systems. They first developed an allocation model for reliability based on a cost function representing the unreliability caused by the execution of tasks on the system processors and the unreliability caused by the interprocessor communication time subject to constraints imposed by both the application and the system resources. It then presented a heuristic algorithm derived from the well-known simulated annealing (SA) technique to quickly solve the mentioned problem.

Vidyathi and Tripathi [29] proposed a task allocation algorithm for maximizing reliability in distributed system using genetic algorithm. Kang et al. [30] proposed task allocation algorithm for maximizing
دریافت فوری
متن کامل مقاله

امکان دانلود نسخه تمام متن مقالات انگلیسی
امکان دانلود نسخه ترجمه شده مقالات
پذیرش سفارش ترجمه تخصصی
امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
امکان دانلود رایگان ۲ صفحه اول هر مقاله
امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
دانلود فوری مقاله پس از پرداخت آنلاین
پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات