Information Sciences 319 (2015) 113–131

Contents lists available at ScienceDirect

Information Sciences

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

Maximizing reliability with energy conservation for parallel task scheduling in a heterogeneous cluster

Longxin Zhang [a], Kenli Li [a,*], Yuming Xu [a], Jing Mei [a], Fan Zhang [b,c], Keqin Li [a,d]

[a] College of Information Science and Engineering, National Supercomputing Center in Changsha, Hunan University, Changsha 410082, China
[b] Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[c] Research Institute of Information Technology, Tsinghua University, Beijing 100084, China
[d] Department of Computer Science, State University of New York, New Paltz, NY 12561, USA

ARTICLE INFO

Article history:
Received 29 June 2014
Received in revised form 7 February 2015
Accepted 14 February 2015
Available online 21 February 2015

Keywords:
Cluster computing
Energy consumption
List scheduling
Parallel application
Reliability

ABSTRACT

A heterogeneous computing system in a cluster is a promising computing platform, which attracts a large number of researchers due to its high performance potential. High system reliability and low power consumption are two primary objectives for a data center. Dynamic voltage scaling (DVS) has been proved to be the most efficient technique and is exploited widely to realize a low power system. Unfortunately, transient fault is inevitable during the execution of an application while applying the DVS technique. Most existing scheduling algorithms for precedence constrained tasks in a multiprocessor computer system do not adequately consider task reliability. In this paper, we devise a novel Reliability Maximization with Energy Constraint (RMEC) algorithm, which incorporates three important phases, including task priority establishment, frequency selection, and processor assignment. The RMEC algorithm can effectively balance the tradeoff between high reliability and energy consumption. Our rigorous performance evaluation study, based on both randomly generated task graphs and the graphs of some real-world applications, shows that our scheduling algorithm surpasses the existing algorithms in terms of system reliability enhancement and energy consumption saving.

1. Introduction

Nowadays, big data applications are progressively becoming the major focus of attention due to the enormous increment of data generation and storage that has taken place in the recent years. The data size is constantly increasing, as of 2012 ranging from a few dozen terabytes to many petabytes of data in a single data set [37]. The major features of big data are high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery, and process optimization [5]. Many real-world areas such as telecommunications, health care, pharmaceutical, Internet search, financial and business informatics generate massive amounts of data. Tmall [38], which is the largest online shopping site in Asia, has become an indispensable part of daily life of Chinese. During the Chinese "Single Day" Double 11 shopping carnival, Alibaba made another record on November 11 2013: total transaction in a day hit 35 billion yuan (USD 5.71 billion). Merely 55 s after 0:00 on November 11, the transaction amount reached 100 million yuan (USD 16.3 million). At 0:06:07, it reached 1 billion yuan (USD 163 million). At 5:49 am, it reached 10 billion

* Corresponding author. Tel.: +86 731 88664161.
E-mail address: lkl@hnu.edu.cn (K. Li).

http://dx.doi.org/10.1016/j.ins.2015.02.023

yuan (USD 1.63 billion). About 188 million of transactions were conducted on this day. It contributes to the powerful support provided by Aliyun. In the data center of Aliyun, there are about 40 thousand jobs to be run on a cluster which is composed of 3000 nodes to process the 1.5 petabyte transaction records everyday, and then to output 20 terabyte results.

A data center garnered significant support and encouragement by its participants, who spanned industry, government labs, and academia [11]. To meet the needs of economic development, national centers for supercomputing have sprung up. The ranking of the top supercomputers in the world show a major trend in heterogeneous architectures. They are superior by power efficiency rather than speed. Such high end computing facilities can consume a very large amount of power, although they provide high performance computing solutions for scientific and engineering applications. For instance, operating a middle-sized data center (i.e., a university data center) demands 80,000 kW power [34]. In addition, high power consumption usually leads to expensive cooling costs. Furthermore, keeping computing facilities running on high power for a long time will result in high temperature of computing systems, which further degrades systems' availability and reliability.

While performance/hardware-cost has increased dramatically with the advancement of electronic technology, power consumption in computer systems has also increased according to Moore's law [36]. Such increased energy consumption causes severe ecological, economic, and technical issues. Hence, it is not very hard to image the size of adverse environmental footprint left by the heterogeneous computing systems (HCS) in a cluster. This issue has attracted extensive research activities in recent years, with the growing advocacy of green computing systems. Some hardware technologies [33], such as energy-efficient monitors, low-power microprocessors and processors consisting of multiple processing element cores and a selective-associative cache memory, are employed to address the energy consumption problems. Comprehensive surveys can be found in Refs. [33,2,4].

Task scheduling problems are classic and important. A large number of excellent algorithms are proposed. Huang et al. [15] proposed three types of fuzzy models to solve the Fuzzy Time-dependent Project Scheduling Problem (FTPSP) while guaranteeing resource constraints. Under the dynamic grid environment, Kołodziej and Khan [22] introduced a Hierarchic Genetic Strategy based Scheduler (HGS-Sched) to achieve fast reductions in makespan and flowtime in the concurrent search of the optimization domain. An adaptive scoring method was used to schedule jobs in grid environment by Chang et al. [7]. With the rapid development of society, reducing processor energy consumption has been a critically important and pressing research issue in recent years. The dynamic voltage and frequency scaling (DVFS) technique [35] is widely recognized as the basis of numerous energy management solutions. It exploits the fact that the dynamic power consumption is a strictly convex function of the CPU frequency, and attempts to conserve energy by reducing clock speed and supply voltage at active state. Benefiting from DVFS, various energy-aware task scheduling and resource management methods have emerged as promising studies for sustainable computing. Many excellent strategies and approaches have been developed, but their scope is restricted to unique processor systems [42], homogeneous computing systems [21,13,43], and battery based embedded systems [31].

Numerous algorithms have been devised to accomplish speedup for parallel applications in the form of directed acyclic graphs (DAG). It is generally appreciated that a task scheduling problem is NP-hard [12]. Usually, scheduling algorithms aim to map tasks onto proper processors and sort tasks in an appropriate sequence, so that task precedence constraints are met and the minimum scheduling length can be achieved. A significant number of existing studies are devised for homogeneous systems, such as the well known Dynamic-Level Scheduling (DLS) algorithm [27]. Recently, a few diverse list scheduling algorithms have been developed to handle heterogeneous systems, for instance, Constrained Earliest Finish Time (CEFT) algorithm [20], Critical-Path-On-a-Processor (CPOP) algorithm [32], and Heterogeneous Earliest Finish Time (HEFT) algorithm [32]. Among these algorithms, HEFT and CPOP have been proven to be very promising algorithms with their demonstrated low complexity and performance effective capability. It is widely accepted that a major challenge in scheduling is to diminish interprocessor communication cost. Node duplication is an effective solution that has been exploited to deal with the above described problem. Based on this, recently reported algorithms, such as HCPFD [14] and HLD [3], were proposed for HCS. They improve the performance by taking account of limited effective duplication into them. Idle time slots, scattered among the processors, are exploited for duplicating the immediate parent of a child task so as to make its start time earlier. The comparison analysis of HCPFD [14] and HLD [3] shows that they significantly outperform other algorithms, for instances, DLS [27] and HEFT [32]. However, the duplication technique aims to reduce the schedule length at the expense of sacrificing more energy and higher complexity. With respect to the promotion of system reliability, a number of hardware and software based techniques for hypercube (HC) fault tolerance was developed by Abd-El-Barr and Gebali [1], which can be exploited to enhance the system reliability and fault tolerance aspects of existing hypercube multi-computer networks (HCNs). Dogan and Özgüner developed three reliability cost functions that were incorporated into making dynamic level (DL) and introduced a Reliable Dynamic Level Scheduling (RDLS) algorithm [9,10]. Tang et al. proposed a Hierarchical Reliability-Driven Scheduling (HRDS) algorithm in a grid computing system [30]. Wang et al. developed scheduling heuristics which reduce energy consumption of parallel task execution in a cluster by using the DVFS mechanism [34].

Unfortunately, most of these approaches are on the basis of simple system models, which do not precisely reflect the real parallel computation systems. One of the assumptions, i.e., a node never fails during execution, may lead to some problems. In real systems, the transition faults in task execution are inevitable and may have an adverse impact on the running applications. Studies in [41] show that it is critical to design an accurate schedule with consideration of task reliability. Low power consumption and high system reliability, availability, and utility are the main concerns of modern high-performance computing system development. A number of studies [18,40,25] revealed the interplay between energy consumption and system reliability. However, these approaches are exclusively confined to the embedded systems. Li and Xu