Ain Shams University
Faculty of Science

Task Scheduling Algorithms on Grid Computing Systems

A thesis

In partial fulfillment of the requirement for the degree of Master of Science in computer science

Submitted By:

Mohamed El sayed Abdel-Aziz

Supervised by:

Prof. Dr. Soheir M. Khamis

Math. Dept. - Faculty of Science

Ain Shams University

Dr. Naglaa M. Reda Taher

Math. Dept. - Faculty of Science

Ain Shams University

Prof. Dr. Abdel Nasser Tawfik

ECTP Founding Director.

MTI University

Submitted to:

Department of Mathematics,

Faculty of Science, Ain Shams University,

Cairo – Egypt

2015

Abstract

Scheduling tasks on heterogeneous resources distributed over a grid computing system is an NP-complete problem. The main aim for several researchers is to develop variant scheduling algorithms for achieving optimality. However, using of the full power of resources is still a challenge.

In this thesis, two heuristic algorithms called Sort-Mid and Range-Suffrage are proposed. Sort-Mid and Range-Suffrage aim to maximizing the resources utilization and minimizing the makespan.

The base step of Sort-Mid is to get the mean value of two consecutive middle values in the sorting list of completion time of each task. Then, the maximum value of these mean values is determined. Finally, the task has the maximum value is assigned to the machine having the minimum completion time. Experimental tests indicate that Sort-Mid utilizes the grid by more than 99% at 12 instances. In addition, Sort-Mid has good makespan in the 12 instances.

Range-Suffrage decision depends on detecting the maximum average value of completion times among certain tasks. These tasks are selected depending on their suffrage values. The task having the maximum average is assigned to the resource with the minimum completion time. Experimental results show that Range-Suffrage utilizes the grid by 97.9% in an instance, 98.9% in an instance, and more than 99% in other instances. On the other hand, Range-Suffrage has lower makespan than other algorithms in 7 instances.

Contents

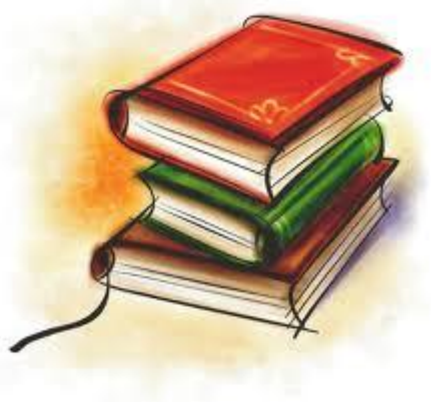
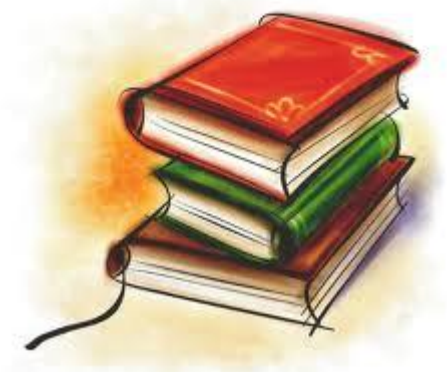


TABLE OF CONTENTS

Chapter 2 :

Acknowledgment



ACKNOWLEDGMENT

All my Praises and gratitude to almighty Allah who aided me to bring forth this thesis to light. This work would have been impossible without the mercy and guidance of Allah.

First and foremost, I would like to express my deepest gratitude to my supervisor Prof. Dr. *Soheir M. Khamis*, whose expertise, understanding and advices were a great help by leading me to this stage in my graduate studies. I thank her for her patience and excellent technical guidance in writing and presenting research.

To my supervisor Dr. *Naglaa M. Reda*, I express my deep and sincere gratitude for her outstanding level of knowledge, scientific understanding, encouraging, personal guidance and for having led me into a fascinating area of research.

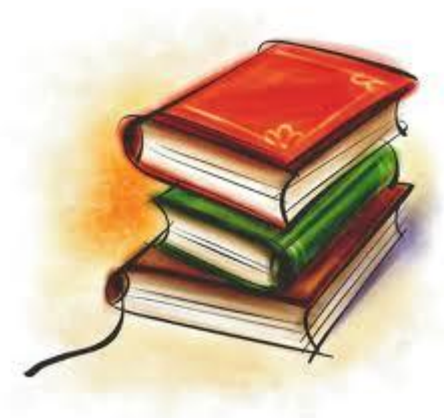
I would like to thank my supervisor Prof. Dr. *Abdel Nasser Tawfik* who has been supportive in many respects over the research years.

I am very thankful to all the staff members of the Egyptian Center for Theoretical Physic ECTP and the members of mathematics department in faculty of science - Ain Shams University for their help and support.

A great grateful to my family (all of them); they have always encouraged me, I indebted to all; my mother spirit, father, and brothers.

I would like to express my deep love to my sweet little daughter *Jodi*, who cheered me during the course of my graduate studies. Finally, I dedicate this research to my wife, for the encouragement during this study.

Summary



Summary

In the mid 1990s, the term grid computing was coined to describe technologies that would allow consumers to obtain computing power on demand. Grid computing is based on collecting resources to work all together as a single system. This grid system is able to execute applications that are too complex for a single resource to cope with.

The grid heterogeneous resources need to be orchestrated to solve complex problems in many scientific fields such as engineering, education, medical treatment, earthquakes, business, etc. Hence, grid can be considered as an effective solution to process and manage large data sets used in scientific application.

Scheduling plays an important role to measure the effectiveness of resources management to achieve high performance on grid computing. The problem of scheduling tasks on distributed resources belongs to a class of problems called NP-complete problems. The process of efficiently scheduling and distributing these massive calculations on resources, which have been assembled in a grid, is one of the most critical challenges facing grid computing. This is because grid computing is too complex due to the differences in the quality of the participating resources in their processing capability. Up to now, many researches have been developed to find strategies to solve task scheduling problem heuristically.

The aim of this thesis is to deal with task scheduling problem. We introduce two new techniques for improving the results of previous heuristic algorithms in maximizing resources utilization and minimizing the makespan. The suggested techniques are named Sort-Mid and Range-Suffrage.

The thesis consists of five chapters, one appendix, and a list of references.

Chapter one presents a detailed study of high-performance computing systems. A description of layers that composite a grid computing system is given. In addition, the most important advantages of grid computing features are given. Finally, the classifications of grid computing systems are categorized according to their objectives and the area at which the grid systems calculations are done.

Chapter two focuses on the scheduler which is the most important part of the grid computing system. The scheduler components and how these components interact together are explained. Then, the phases of grid scheduling are reviewed in detail. Furthermore, the scheduler organization, policies, challenges, and some grid scheduler systems are introduced.

Chapter three includes a formal description of the problem of task scheduling on a grid computing system. So, the classification of algorithms that deal with this problem is demonstrated. Most of related algorithms for solving the scheduling problem that are needed to evaluate our work are summarized.

Chapter four concentrates on our research which is restricted to finding an efficient method to solve the task scheduling problem. We suggest two new strategies. The first one is based on a sorting technique, so called Sort-Mid algorithm. In this algorithm, the main idea is to get the average of middle two consecutive completion times of resources in the sorted completion time matrix among all tasks. Then, it chooses the task having large average and assign it to the resource having minimum completion time. The second one depends on what is called the suffrage value. The main idea depends on specified range suffrage constraint, so the algorithm is called Range-Suffrage. The main step of the algorithm is to choose those tasks having this range and complete the algorithm as usual. The description of these two algorithms, pseudo codes, and the proof of the time complexities are given. An illustrative example for comparison between our two algorithms and other well-known previous algorithms is presented.

Chapter five contains an experimental testing for the new algorithms and other well-known algorithms in this field. The algorithms are executed on a benchmark simulation model which consists of 12 different 512×16 execution time completion matrices. All recorded tables and figures are presented and show the position of our new method with respect to the others.

In the appendix, we give the script codes of our implementation using the programming language Visual Basic for our proposed algorithms.

LIST OF FIGURES