# Tribhuvan University
## Institute of Engineering, Pulchowk Campus

**A**

**THESIS PROPOSAL**

**ON**

# "Task Scheduling in Grid Computing Using Genetic Algorithm"


**Submitted By:**

Ujjwal Prajapati
(2069/MSCS/669)


**Submitted To:**

**Department of Electronics and Computer Engineering**

May 2014

# THESIS PROPOSAL

# ON

# "Task Scheduling in Grid Computing Using Genetic Algorithm"

Ujjwal Prajapati
(2069/MSCS/669)

**Department of Electronics and Computer Engineering**
**IOE, Pulchowk Campus**

May 2014

# Acknowledgement

I would like to express my gratitude to all the faculty members of Electronics and Computer Engineering Department who are directly/indirectly involved in inspiring and motivating me to carry out this thesis. A special word of gratitude to my program co-ordinator **Dr. Sanjeeb Prasad Panday** and **Professor Dr. Shashidhar Ram Joshi** whose continuous support and guidance has led me to come up with this title of thesis.

Also, I would like to express a word of thanks to all my classmates for their continuous idea sharing and discussion, which have definitely paved the path to undertake this thesis.

Abstract

Task scheduling is a key problem in Grid computing in order to benefit from the large computing capacity of such systems. The need of allocating a number of tasks to different resources for the efficient utilization of resources with minimal completion time and economic cost is the essential requirement in such systems. The problem is multi-objective in its general formation, with the objectives being the minimization of makespan and flowtime of the system along the economic cost. An optimal scheduling could be achieved minimizing the completion time and economic cost using the heuristic approach, which is chosen to be Genetic Algorithm. The ability of Genetic Algorithm to simultaneously search different regions of a solution space makes it possible to find a diverse set of solutions for difficult problems. Each individual is represented as possible solution. The solutions are the schedulers for efficiently allocating jobs to resources in a Grid system.

**Keywords**

Task Scheduling, Grid Computing, Distributed Computing, Makespan, Economic Cost, Genetic Algorithm

## Table of Contents

## List of Tables

## List of Figure

**FIGURE** **PAGE**

# CHAPTER ONE

## INTRODUCTION

# 1. Introduction

A computational grid is a large scale, heterogeneous collection of autonomous systems, geographically distributed and interconnected by heterogeneous networks. A computational grid contains resource management, task scheduling, security problems, and information management and so on. Task scheduling is one of the fundamental issues which play an important role in the operation of distributed computing systems. Task scheduling in distributed computing systems is defined as the process of assigning the tasks of a distributed application into the available processors, and specifying the start execution time of the tasks assigned to each processor [5]. The problem of task allocation in distributed computing system is the need to allocate a number of tasks to different processors for execution.

A task is an atomic unit to be scheduled by the scheduler and assigned to a resource. A task scheduling is the mapping of tasks to a selected group of resources which may be distributed in multiple administrative domains. In the case of static scheduling, information regarding all resources in the Grid as well as all the tasks in an application is assumed to be available by the time the application is scheduled. The application centric objective function in Grid computing could be either *makespan*, which is the time spent from the beginning of the first task in a job at the end of the last task of the job, or *economic cost* that an application needs to pay for resource utilization [3].

Job Scheduling is known to be NP-complete; therefore the use of heuristics is the de facto approach in order to cope in practice with its difficulty [4]. The meta-heuristics run on static instances of the problem and therefore in this approach static schedulers are obtained. A Genetic Algorithm is a meta-heuristic search technique which allows for large solution spaces to be partially searched in polynomial time, by applying evolutionary techniques from nature.

# CHAPTER TWO

## PROBLEM DEFINITION

## 2.1 Problem Statement

Job sharing is one of the major difficult tasks in a computational grid environment. Unlike scheduling problems in conventional distributed systems, this problem is much more complex as new features of Grid systems such as its dynamic nature and the high degree of heterogeneity of jobs and resources must be tackled. When there are more tasks than available resources, the problems arises for the minimization of the completion time and utilize the resources effectively with minimum cost.

## 2.2 Objectives

- To implement Genetic Algorithm for the multi-objective optimization of makespan, flowtime and economic cost of the system
- To see and compare the possible performance change considering different encoding schemes, operators and parameter tuning

## 2.3 Scope of Work

In any of the Grid computing where task scheduling is a necessarily complicated, this solution could be implemented. The solution could be run in Grid Scheduler or Grid Resource Manager which provides the functionality for discovery and publishing of resources along with scheduling, submission and monitoring of jobs.

# CHAPTER THREE

## RESEARCH METHODOLOGY

## 3.1. Literature Review

In Distributed Computing System, an allocation policy may be either static or dynamic, depending upon the time at which the allocation decisions are made. In a static task allocation, the information regarding the tasks and processor attributes is assumed to be known in advance, before the execution of the tasks. Distributed Computing Systems have become a key platform for the execution of hydrogenous applications. The major problem encountered when programming such a system is the problem of task allocation. Task allocation problem is known to be NP-hard problem in complexity, where required an optimal solution to the problem. The easiest way to finding an optimal solution to the problem is an exhaustive enumerative approach. But it is impractical, because there are $n^m$ ways of allocation m-tasks to n-processors.

*Ahmed Younes. Hamed* [1] presents a genetic algorithm, considering distributed computing system with heterogeneous processors in order to achieve optimal cost by allocating the tasks to the processors, in such a way that the allocated load on each processor is balanced. The algorithm is based on the execution cost of a task running on different processors and the task communication cost between two tasks to obtain the optimal solution. The proposed algorithm tries to minimize the processor execution cost and inter processor communication.

*Javier Carretero, Fatos Xhafa* [4] presents an extensive study on the usefulness of Genetic Algorithms for designing efficient Grid Schedulers when makespan and flowtime are minimized under hierarchic and simultaneous approaches. Two encoding schemes have been considered and most of GA operators for each of them are implemented and empirically studied.

*Mohammad I. Daoud and Nawwaf Kharma* [5] proposed customized genetic algorithm to produce high-quality task schedules for Heterogeneous Distributed Computing Systems. Also, the performance of the scheduling algorithm is compared to two leading scheduling algorithms which is based on both randomly generated task graphs and task graphs of certain real-world numerical applications, exhibits the supremacy of the new algorithm over the older ones, in terms of schedule length, speedup and efficiency.

*Prateek Kumar Singh, Neelu Sahu* [2] proposed compact genetic algorithm, which aims to generate an optimal schedule so as to get the minimum completion time while completing the tasks.

## 3.2. Methodology

A Genetic Algorithm is a meta-heuristic search technique which allows for large solution spaces to be partially searched in polynomial time, by applying evolutionary techniques from nature. Genetic Algorithm is high level algorithms that integrate other methods and genetic operators, therefore in order to implement it for a problem, we have to use the template for the method and design the inner methods, operators and appropriate data structures.

**begin**

  **Initialization:** Generate the initial population *P (t=0)* of n individuals

  **Fitness:** Evaluate the fitness of each of the population.

    Evaluate (P (t))

  **while** (not termination condition) **do**

    **Selection**: Select a subset of *m* pairs from *P (t)*.

      Let $P_1$ (t) = Select (P (t))

    **Crossover:** With probability $p_c$, cross each of the m chosen pairs.

      Let $P_2$ (t) = Cross ($P_1$ (t)) be the set of offspring.

    **Mutation:** With probability $P_m$, mutate each offspring in $P_2$ (t).

      Let $P_3$ (t) = mutate ($P_2$ (t))

    **Fitness:** Evaluate the fitness of each offspring. Evaluate ($P_3$ (t))

    **Replacement:** Create a new generation from individuals in P (t) and $P_3$ (t).

      Let P (t+1) = Replace (P (t), $P_3$ (t)); t = t+1

  **fwhile**

  **return** Best found solution;

**end**

# CHAPTER FOUR

## EXPECTED OUTPUT

## 4.1. Expected Output

| Tasks | Processors | System's Cost |
|---|---|---|
| $t_4$, $t_7$ | $P_1$ | 459 |
| $t_2$, $t_3$, $t_8$, $t_9$ | $P_2$ | |
| $t_1$, $t_5$, $t_6$ | | |

Table 1 Expected Output for Minimization of Cost

| Instance | Min-min | GA |
|---|---|---|
| u_c_hihi.0 | 8460675.000 | 7730973.882 |
| u_c_hilo.0 | 164022.440 | 157255.844 |
| u_c_lohi.0 | 275837.340 | 252907.580 |
| u_c_lolo.0 | 5546.260 | 5310.615 |

Table 2 Expected Output for Minimization of Makespan

| Instance | GA |
|---|---|
| u_c_hihi.0 | 1073774996 |
| u_c_hilo.0 | 28314677.9 |
| u_c_lohi.0 | 35677170.8 |
| u_c_lolo.0 | 942076.61 |

Table 3 Expected Output for Minimization of Flowtime

## 4.2. Time Schedule

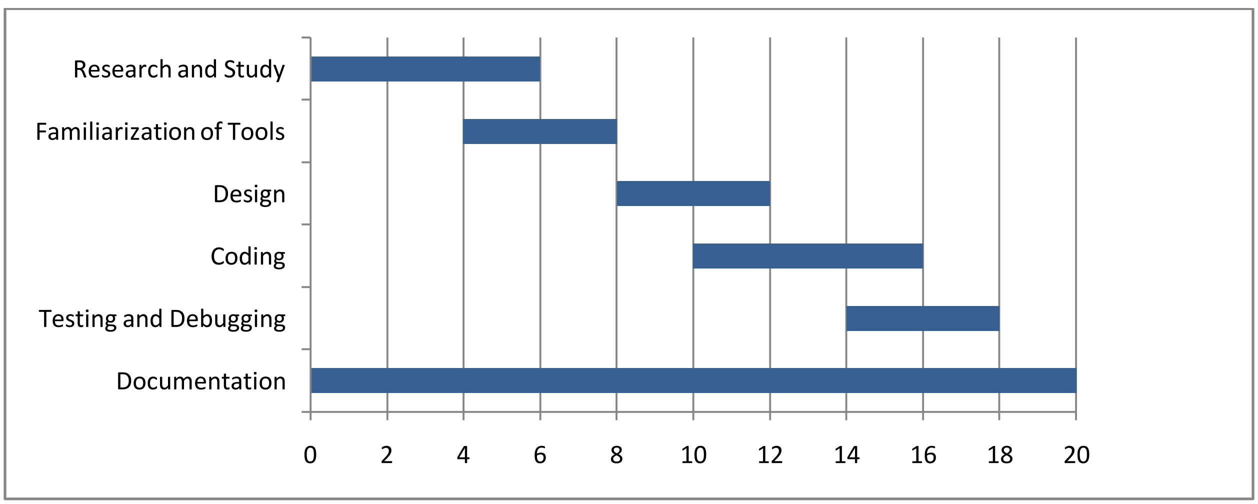


Figure 1 Gantt chart

# CHAPTER FIVE

## REFERENCES

# 5. References


[1] Ahmed Younes. Hamed "Task Allocation for Minimizing Cost of Distributed Computing Systems using Genetic Algorithms", Vol 2, Issue 9, September 2012

[2] Prateek Kumar Singh, Neelu Sahu "Task Scheduling in Grid Computing Environment Using Compact Genetic Algorithm", Vol3, Issue 1, January 2014

[3] Fangpeng Dong and Selim G. Akl "Scheduling Algorithms for Grid Computing: State of the Art and Open Problems", January 2006

[4] Javier Carretero, Fatos Xhafa "Genetic Algorithm Based Schedulers for Grid Computing Systems", Vol 3, No. 6, December 2007

[5] Mohammad I. Daoud and Nawwaf Kharma, "An Efficient Genetic Algorithm for Task Scheduling in Heterogeneous Distributed Computing Systems", July 2006