CHAPTER I

INTRODUCTION

1.1 Background and Motivation

In recent years, various combined system architectures or operators such as cloud computing or computer grid systems [1], [2], [3], [4], and [5] have been introduced to solve increasingly complicated problems in science and commercial. With the fast-paced development of computational technology [6], [7], [8], and [9], the system architectures become more complex and require high computing power. Thus one mandate to the development for supporting algorithms is to minimize the complexity of computation processes, especially to reduce computing time and energy consumption.

Ordinary server assignment algorithms send tasks, data, and functions to be generated by users on client sites when parallel codes are queueing for execution by a cluster of servers. In practice, only allotted jobs are executed at the assigned servers. Moreover, if the job demands addition security measures, high transferring rate to send data across networks will incur as more processes are involved. Therefore, in systems which contain time complexity, efficient server assignment algorithms are needed to achieve optimal energy consumption.

Energy efficient server assignment problems have been widely studied [10], [11], [12], and [13] such as those for real-time 2-level heterogeneous grid system proposed by Terzopoulos and Karatza [14], i.e., Dynamic Voltage Scaling (DVS) and Dynamic Power Management (DPM). These studies focus on reduction of energy consumption with minimum performance degradation. In a typical distributed system, some data or functions are kept at a local scheduler and some are kept in specific places due to required security. A task scheduler is employed to determine the starting time of each task and the server to which the task should be sent.

In this study, two energy saving scheduling algorithms are proposed. The first algorithm is the Energy-Efficient Process Clustering Assignment (EPC) algorithm which aims to handle complex scenarios and find an optimal energy efficient assignment in a reasonable computing time. The important aspects to the development of this algorithm are matching the server to which the tasks should be sent and clustering tasks for execution on the same server. The second algorithm is the energy sufficiency level algorithm for distribution assignment (ESL) algorithm which aims to lower energy consumption, time complexity, and system finish time. Special provision to handle limited power supply scenario is also incorporated in ESL algorithm. The development of both algorithms is based on level based scheduling and idle slot reduction which can perform over limited or unlimited energy supplies.

1.2 Objective

The main objective of this study is to develop a scheduling assignment algorithm which minimizes energy consumption and system finish time for problems in distributed system.

1.3 Scope of work

The proposed algorithm works under two scenario cases.

1) The focus is given to the total energy consumption yielded by each algorithm. The total energy consumption is taken as a combination of execution energy, transmission energy, and idle energy.
2) Both the total energy consumption and system finish time are used as operating parameters. For this case, one processing unit is selected as a main processing unit which is dedicated to handle task scheduling, task execution, and wait time before the final result from the last task execution is sent over. Additional measure is provided for limited power supply.

The initial constraints imposed on the processing units in both scenario cases are i) all processing units are capable of executing one task at a time, ii) each processing unit is capable of executing more than one task, iii) there can be more

than one processing unit which can execute the same task and iv) the energy consumption for each processing unit to execute each task is different.

1.4 Contribution

This thesis consists of two energy saving assignment algorithms for distributed scheduling assignment. The EPC algorithm can be used to complete the assignment with optimized total energy consumption as the algorithm minimizes energy consumption of each processing unit during process execution, system idling, and data transmission. In the situation where battery supplied to the processing units are limited, the ESL algorithm is supplemented. Energy reserve function introduced in this algorithm is used to further reduce the total energy consumption when available battery supply is lower than the originally required value. Therefore, the work offers optimal total energy consumption and system finish time for distributed task scheduling.

1.5 Dissertation Organization

The thesis is organized as follows. A review of related works and their background theories are presented in the Chapter II. In Chapter III, descriptions of the problem formulation are presented. The first study case with scheduling to minimize total energy consumption is described and the detail with its results of the proposed Energy-Efficient Process Clustering Assignment (EPC) algorithm is discussed. The second study case with scheduling to minimize total energy consumption and scheduling length with limited power supply is described and the detail with its results of the proposed Energy Sufficiency Level Algorithm for Distribution Assignment (ESL) is described and its experimental results are discussed. Notable points of the proposed algorithm are discussed in the Chapter IV. Finally, the overall conclusion is summarizing in Chapter V.