TASK SCHEDULING IN CLOUD COMPUTING ENVIRONMENT USING HYBRID OF GENETIC ALGORITHM AND NAKED MOLE RAT ALGORITHM (GA-NMRA)

MOHAMMAD OZANIEZIE BIN ONASIS

FACULTY OF COMPUTING AND INFORMATICS
UNIVERSITI MALAYSIA SABAH

2022

UMS
UNIVERSITI MALAYSIA SABAH

TASK SCHEDULING IN CLOUD COMPUTING ENVIRONMENT USING HYBRID OF GENETIC ALGORITHM AND NAKED MOLE RAT ALGORITHM (GA-NMRA)

MOHAMMAD OZANIEZIE BIN ONASIS

THESIS SUBMITTED IN PARTIAL FULFILLMENT FOR THE BACHELOR OF COMPUTER SCIENCE WITH HONOURS (NETWORK ENGINEERING)

FACULTY OF COMPUTING AND INFORMATICS
UNIVERSITI MALAYSIA SABAH
2022

UMS
UNIVERSITI MALAYSIA SABAH

NAME : MOHAMMAD OZANIEZIE BIN ONASIS

MATRIC NUMBER : BI18110193

TITLE : TASK SCHEDULING IN CLOUD COMPUTING ENVIRONMENT USING HYBRID OF GENETIC ALGORITHM AND NAKED MOLE RAT ALGORITHM (GA-NMRA)

DEGREE : BACHELOR OF COMPUTER WITH HONOURS (NETWORK ENGINEERING)

VIVA'S DATE : 17 JANUARY 2022

CERTIFIED BY;

1. **SUPERVISOR** SIGNATURE

MR. NORDIN SAAD

DECLARATION

I hereby declare that the material in this thesis is my own except for quotations, equations, summaries, and references, which have been duly acknowledged.

11 FEBRUARY 2022

MOHAMMAD OZANIEZIE BIN ONASIS
BI18110193

ACKNOWLEDGEMENTS

First and forever, I would like to express my deepest gratitude to Allah S.W.T for his grace and love, I was able to complete my research work successfully.

I would also like to thank my supervisor, Mr. Nordin Saad, Assoc.Prof. Dr. Azali Saudi , all senior lecturers, panel, and examiner, at Universiti Malaysia Sabah, Faculty of Computing and Informatics for giving me the opportunity and teaching me to make this research a success. With his guidance, words of passion, and guidance I was able to understand deeply and successfully inspired me to complete this research. He taught me a lot about how to prepare literature reviews and gave a lot of insightful explanations about the methodology to further facilitate my research success.

My sincere thanks also to my parents Onasis bin Hj. Abd Husin and my mother, Kasmah binti Jenteri who has not given up on praying for my success. Thank you also to my friends Muhammad Syahril, Kamal Khairi, Mohd Erwan,and Mohd Izuddin for always supporting me. If I have not had the guidance and discussion with them, I would not be able to make this research a success.

Lastly, I would like to thank all involved directly or indirectly for the success of this research.

MOHAMMAD OZANIEZIE BIN ONASIS

11 FEBRUARY 2022

ABSTRACT

Cloud computing is on-demand service and resources available for the computing systems nowadays, especially in the data storage without interference from humans. Task scheduling and resource allocation are essential aspects of cloud computing. This research proposes task scheduling in cloud computing using a hybrid genetic algorithm and naked mole rat algorithm to solve the task scheduling problem. Genetic Algorithm (GA) was widely used because of its accuracy and simplicity. However, it will become slower in some instances that include a more significant problem size. Hence, Naked Mole Rat Algorithm (NMRA) can optimize the efficiency and performance because it provides an efficient scheduling mechanism. NMRA also will minimize the execution time and deadline. This research will compare hybrid Genetic Algorithm and Naked Mole Rat Algorithm (GA-NMRA) with other meta-heuristic algorithms. Other than that, this research will use Waterfall Model Methodology as its research methodology. Furthermore, this research will apply hybrid GA-NMRA for task scheduling in cloud computing environments. This research will conduct several experiments in Cloud Computing Environment Simulation comparing GA, NMRA and the hybrid GA-NMRA to get research results. The result will show that GA-NMRA will improve the quality of service, minimize the time execution and deadline for a given task.

ABSTRAK

Penjadualan Tugas di dalam Persekitaran Perkomputeran Awan menggunakan Hibrid Algoritma Genetik dan Algoritma Tikus Mondok Telanjang (GA-NMRA)

Pengkomputeran awan ialah perkhidmatan atas permintaan dan sumber yang tersedia untuk sistem pengkomputeran pada masa kini, terutamanya dalam storan data tanpa gangguan daripada manusia. Penjadualan tugas dan peruntukan sumber adalah aspek penting dalam pengkomputeran awan. Kajian ini mencadangkan penjadualan tugas dalam pengkomputeran awan menggunakan hibrid algoritma genetik dan algoritma tikus mondok telanjang (GA-NMRA) untuk menyelesaikan masalah penjadualan tugas. Algoritma Genetik (GA) digunakan secara meluas kerana ketepatan dan kesederhanaannya. Walau bagaimanapun, ia akan menjadi lebih perlahan dalam beberapa keadaan yang merangkumi saiz masalah yang lebih ketara. Oleh itu, Algoritma Tikus Mondok Telanjang (NMRA) boleh mengoptimumkan kecekapan dan prestasi kerana ia menyediakan mekanisme penjadualan yang cekap. NMRA juga akan meminimumkan masa pelaksanaan dan tarikh akhir. Penyelidikan ini akan membandingkan Hibrid Algoritma Genetik dan Algoritma Tikus Mondok Telanjang (GA-NMRA) dengan algoritma meta-heuristik yang lain. Selain itu, kajian ini juga akan menggunakan metodologi model airterjun sebagai metodologi kajian. Seterusnya, projek ini akan digunakan GA-NMRA hibrid untuk penjadualan tugas dalam persekitaran pengkomputeran awan. Kajian ini juga akan melakukan beberapa eksperimen didalam Simulasi Persekitaran Perkomputeran Awan, membandingkan GA, NMRA dan hybrid GA-NMRA untuk mendapatkan hasil kajian. Hasilnya akan menunjukkan bahawa GA-NMRA akan meningkatkan kualiti perkhidmatan, meminimumkan pelaksanaan masa dan tarikh akhir untuk tugasan yang diberikan.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Introduction

In a world that is in track of becoming a digital society, the Internet and its services has been a core necessity for most people. As the technology of computing has been more advanced than the last decade, it helps us to conduct our daily chores easily. One of the breakthroughs of computing technology is Cloud Computing.

Cloud Computing can be defined as the delivery of different services through the Internet, including data storage, servers, databases, networking, and software. This technology makes it possible to save files to a remote database and retrieve it on demand from any part of the world. It also enables us to control and access any device connected to the cloud from any part of the world. Examples of Cloud Computing service providers are Google (Google Drive), Amazon (Amazon Web Services (AWS)) and Microsoft (Microsoft Azure).

Task scheduling is an important aspect of cloud computing. While task scheduling is one of the ways that will increase cloud computing performance, workload division will be supplied equitably and effectively for all resources to reduce waiting time, execution time, maximum throughput, and effectively undesired resource utilization. The fundamental issue with task scheduling is the time spent waiting and executing tasks. For example, the difficulty with First Come First Serve (FCFS) is that tasks have a long waiting time, whereas Shortest Job First (SJF) has a shorter waiting time but a longer Total Finish Time (TFT) than FCFS.

In this project, I will do research base on the combination of two algorithms which is Genetic Algorithm (GA) and Naked Mole-Rat Algorithm (NMRA) and this algorithm will become a Hybrid Genetic Algorithm and Naked Mole-Rat Algorithm (GA-NMRA).

Hopefully, this algorithm will counter the highest waiting time and longer total finish time for the task scheduling

1.2 Problem Statement

Following to the introduction part, we already know that cloud computing users are increasing each day. Jamali S et al.,2016 stated that millions of users send requests to cloud resources every single day. Scheduling these jobs is difficult for the cloud system. The latter also explains that algorithm that is widely used for task scheduling is rule-based scheduling algorithm. For example, deterministic scheduling algorithms and exhaustive algorithms. The problem with this algorithm is when it comes to the large scale of task scheduling. The performance of this algorithm was very far from the optimal performance. Thus, meta-heuristic algorithms need to be used to solve this problem. Natesan G et al., 2019 mentioned that the quality of the result or the speed of convergence can be improved when a meta-heuristic algorithm is paired with other meta-heuristic algorithms based on population or local-search-based meta-heuristic algorithms. Therefore, this paper proposed the hybrid Genetic Algorithm and Naked Mole-Rat Algorithm (GA-NMRA) for task scheduling in a cloud computing environment.

Problems that were identified:

i. Rule-based or static scheduling algorithm cannot afford large scale of task scheduling problem in cloud computing environment.

ii. Wrong job allocation scheduling has an impact on computer performance since it is impossible to build a parallel and synchronized process that runs at the same time.

iii. Scheduling algorithms that inflexible will take longer to process data, degrading system performance even more.

1.3 Project Objectives

The ais for Task Scheduling in Cloud Computing using Genetic Algorithm and Naked Mole Rat Algorithm (GA-NMRA) are including:

i. To utilise and implement Genetic Algorithm and Naked Mole Rat Algorithm (GA-NMRA) in Cloud Computing environment.
ii. To develop a web-based system for the algorithm simulation process and result arrangement.
iii. To implement and evaluate the handling performance of the algorithms in term of the best duration time on the CloudSim simulator.

1.4 Project Scope

This project will lead to detailed explanation on GA, NMRA and hybrid GA-NMRA. Some research will be conducted to come out with advantages of hybrid GA-NMRA. Furthermore, research also proceeds by exploring previous research using NMRA in task scheduling. It will lead to an improvise list to achieve the objective of this project.

The algorithm that has been proved to enhance task scheduling in cloud computing will benefits Jabatan Teknologi Maklumat dan Komunikasi (JTMK), Universiti Malaysia Sabah especially to their Network Administrator and System Administrator. It can be applied to ensure all university's system that will be transferred into cloud can have better performance. This project will lead to detailed explanations on GA, NMRA and hybrid GA-NMRA.

To obtain experimental findings that may be utilized to evaluate the project hypothesis, developed web-based application systems and some computations will be used. This project's potential is to have improved performance in terms of job scheduling in cloud computing environments, which are in great demand from consumers. This project's outcome will be displayed in a web-based application that will assist any developer in interpreting the results and determining any other parameter or situation to test the algorithm's performance.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

With the ubiquitous growth of Internet access and big data in their volume, velocity, and variety through the Internet, cloud computing becomes more and more proliferating in

the industry, academia, and society. Cloud computing is composed of distributed computing, grid computing, utility computing, and autonomic computing. Cloud computing provides on-demand computing and storage services with high performance and high scalability. Several computing paradigms have promised to deliver this utility computing.

Cloud computing is one such reliable computing paradigm. However, the rising energy consumption of cloud data centres has become a prominent problem. Task scheduling is an important step to improve the overall performance of the cloud computing. Traditional monitoring and management mechanisms are designed for enterprise environments, especially a unified environment. However, the large scale, heterogeneous resource provisioning places serious challenges for the management and monitoring mechanism in multiple data centres.

In recent years, the problem of task scheduling on a distributed environment has caught the attention of researchers. Task scheduling is considered a critical issue in the Cloud computing environment by considering different factors like completion time, the total cost for executing all users' tasks, utilization of the resource, power consumption, and fault tolerance.

2.2 Task Scheduling in Cloud computing

Task Scheduling are defined as a mechanism used to select resources and execute the task to get less waiting and execution time. There a two level of task scheduling the first level is set of policies to distribute virtual machine in host and the second level is set of policies to distributed task to virtual machine (Tahani Aladwani, 2020). The biggest challenge in cloud is to control the performance such as execution time, response time, waiting time, network, bandwidth, and service cost. The task scheduling has been developed in the 2013 to 2020 which is a 3% in year 2013 and in the year of 2014 is 8%, in 2015 is 5% and 8% contribution that has been provide by the researcher (Prashant B. Jawade, Sai Kumar D, Ramachandram S, 2021).

Tasks scheduling algorithms are defined as the mechanism used to select the resources to execute tasks to get less waiting and execution time.

In the cloud computing environment, there are two levels of scheduling algorithms:

i. First level: in host level where a set of policies to distribute VS in the host.
ii. Second level: in VM level where a set of policies to distribute tasks to VM.

Tasks scheduling algorithms are defined as a set of rules and policies used to assign tasks to the suitable resources (CPU, memory, and bandwidth) to get the highest level possible of performance and resources utilization.

The advantages of the task scheduling algorithms are:

i. Manage a performance and QoS for cloud computing.
ii. Manage memory and CPU.
iii. Maximizing resources.
iv. Minimizing total execution time.
v. Improving fairness.
vi. Increasing successfully completed task number.
vii. Real-time system.
viii. High throughput.
ix. Load balancing.

2.3 Genetic Algorithm

Genetic Algorithm is inspired by evolution that are a family of computational models and bring a potential of solution to a specific problem for a simple chromosome. This Algorithm also viewed as function optimizer, even though the range of problems to which this Algorithm applied are large. The implementation of this algorithm begins with a random population of chromosomes, and it will evaluate the structures and allocated and to find out the better solutions to the target problem so that more chances to reproduce than to a poorer solution.

The process of natural selection will start with the selection of fittest individual from a population and will produce offspring which inherit the characteristics of parents and will be added to next generation. There a three component that involve in Genetic Algorithm which is gene, chromosome, and population. Chromosome is a complicated long thread of DNA (deoxyribonucleic acid), and each trait is got from combination of the DNA itself. There a five phase that are considered in a genetic algorithm which is initial population, fitness function, selection, crossover, and mutation.

Selection phase will select the fittest individual and pass their genes to the next generation, the parents are selected base on their fitness score for a reproduction. The crossover is the most important phase, crossover point is chosen at random from the genes which their parents mated from. Crossover is needed to combine two string to get a better string. This recombination creates a different individual in the next generation by combine the material from two individual that are from previous generation. For example, before crossover the string 1 is 1110000 and the string 2 is 0000111 but after crossover happened string 1 will become 00001111 and string 2 will become 11110000 (Tom V. Mathew).

2.4 The Naked Mole Rate Algorithm

The developers of this algorithm were inspired by NMR's social actions to suggest a stochastic optimization algorithm. The algorithm is based on NMR mating patterns and has the following main characteristics:

i. NMRs are eusocial animals who live in a population of 295 people, with an average of 70–80 people. The current study employs a group of 50 NMRs for its experimental research.
ii. A female queen leads the party, which divides the population into breeders and workers. Breeders are the most active NMRs in the working population, and they are only in charge of mating, while workers oversee several tasks.
iii. Workers perform required duties, and breeders replace the best of them. To put it another way, high-performing workers become breeders, while low-performing breeders are returned to the worker pool.
iv. The best breeder among the breeder pool mates with the queen.

The four rules mentioned above have been idealised to form a naked mole-rat algorithm (NMR). There are three stages to the algorithm. The population of NMRs is initialised in the first phase, followed by the worker phase, and finally the breeder phase. The breeder process is chosen based on the likelihood of breeding.

a. Initialization: Initially, it generates a uniformly distributed random population of *n* NMR where each NMR in the range [1, 2 … n] is a *D*-dimensional vector. Here *D* represents the number of variables or parameters to be tested in the problem. Each NMR is initialized as:

$$NMRi,j = NMRmin,j + U(0,1) \times (NMRmin,j - NMRmax,j) \quad (1)$$

where $i \in [1, 2, \ldots n]$, $j \in [1, 2, \ldots D]$, $NMRi,j$ is the ith solution in the jth dimension, $NMRmin,j$, $NMRmax,j$ are the lower and upper bounds of the problem function respectively and $U(0,1)$ is uniformly distributed random number. After initialization, the objective function is evaluated, and its fitness is calculated. Based upon the fitness, *B* breeders and *W* workers are identified, and overall initial best solution *d* is

calculated. After initialization, the population of NMR is subjected to repeated cycles or iterations of the search process of worker and breeder phase.

b. Worker phase: In this phase, the workers tend to improve their fitness so that they get a chance to become a breeder and eventually mate with the queen. So here, the new solution of worker NMR is generated based upon its own experience and local information. Here the fitness of new NMR is evaluated and if the new mating fitness is better, the old solution is discarded and the new solution is memorized. Otherwise, the older solution is retained. After all the worker rats complete the search process, the final fitness of all of them is remembered. To produce a new solution from the old one, the NMR uses the following equation:

$$wit{+}1 = wit + \lambda(wjt - wkt) \quad (2)$$

where wit corresponds to the ith worker in the tth iteration, $wit{+}1$ is the new solution or worker, λ is the mating factor and wjt and wkt are two random solutions chosen from the worker's pool. The value of λ is obtained from a uniform distribution in the range of [0, 1].

c. Breeder phase: The breeder NMR also update themselves in order to be selected for mating and also to stay as a breeder. The breeder NMRs are updated based upon a breeding probability(bp) with respect to the overall best *d*. This bp is a random number in the range of [0, 1]. Some of the breeders may not be able to update their fitness and hence may be pushed back to the workers' category. The breeders modify their positions according to the equation given below:

$$bit{+}1 = (1 - \lambda)bit + \lambda(d - bit)$$

Here bit corresponds to the breeder i in the iteration t, λ factor controls the mating frequency of breeders and helps in identifying a new breeder $bit{+}1$ in the next iteration. To start with, the value of bp has been set to 0.5 as the initial value. For simplicity, we have assumed that there is only a single queen and best among the breeder mates with the queen. So here we find only the best breeding male who will breed with the female. The algorithm works by differentiating or identifying the

breeders and workers among the pool of NMRs. After an initial evaluation, the best breeder, and the best worker is selected. The fitness of workers is updated so that their fitness improves, and they may get a chance to become breeders. On the other hand, breeders also update their fitness based upon breeding probability so that they remain breeders. The breeder which becomes sterile will be pushed into the workers' category. The best breeder among the population serves as the potential solution to the problem under test.

2.5 Review on Previous Research - Task Scheduling in Cloud Computing Based on Hybrid Moth Search Algorithm and Differential Evolution

This paper presents an alternative method for cloud task scheduling problem which ais to minimize make span that required to schedule several tasks on different Virtual Machines (VS). The proposed method is based on the improvement of the Moth Search Algorithm (SA) using the Differential Evolution (DE). The SA simulates the behaviour of moths to fly towards the source of light in nature through using two concepts, the phototaxis and Levy flights that represent the exploration and exploitation ability respectively. However, the exploitation ability is still needed to be improved, therefore, the DE can be used as local search method. To evaluate the performance of the proposed SDE algorithm, a set of three experimental series are performed. The first experiment ais to compare the traditional SA and the proposed algorithm to solve a set of twenty global optimization problems. Meanwhile, in second and third experimental series the performance of the proposed algorithm to solve the cloud task scheduling problem is compared against other heuristic and meta-heuristic algorithms for synthetical and real trace data, respectively. The results of the two experimental series show that the proposed algorithm outperformed other algorithms according to the performance measures.

2.4 Review on Previous Research - An Efficient Approach to The Map-Reduce Framework and Genetic Algorithm Based Whale Optimization Algorithm for Task Scheduling in Cloud Computing Environment.

In the digital era, cloud computing has emerged a significant service in the IT sector. It ensures pooling of resources and provided services on-demand over the web. The scheduling of appropriate tasks is an important aspect in cloud computing in which much research is carried out. The users demand for resources are volatile in nature and hence when a large count of resources is requested, the computational overhead in cloud is supposed to effectively allocate resources and to complete these tasks. The research issue includes as how a VM can schedule these tasks in an effective manner. This paper proposes an efficient approach using the MAP reducing framework and GA-WOA for efficient scheduling of tasks in the given cloud. Initially, the task features are extracted from the client's task. Then, the features are reduced by using the MRQFLDA algorithm. After that, the large tasks are separated into sub-tasks using a map-reduce framework. Finally, the tasks are efficiently scheduled by using the GA-WOA algorithm. The experimental simulations are carried out using the CLOUDSIM environment. The results show that the proposed method GA-WOA outperforms the other methods in terms of various metrics used for the evaluation.

2.5 Summary

Based on my literature review on several papers regarding of task scheduling using algorithms in cloud computing environment, I would like to use the Naked Mole Rat Algorithm (GA-NMRA), based on the findings of previous research.