UNIVERSITI PUTRA MALAYSIA

AN EFFICIENT ALGORITHM FOR INDEPENDENT TASK SCHEDULING IN HETEROGENOUS COMPUTING SYSTEMS

ABDULRAHMAN KHALDOON MAHMOOD AL-QADHI

FSKTM 2017 16

AN EFFICIENT ALGORITHM FOR INDEPENDENT TASK SCHEDULING IN HETEROGENOUS COMPUTING SYSTEMS

BY:

ABDULRAHMAN KHALDOON MAHMOOD AL-QADHI

A PROJECT SUBMITTED TO FULFILL THE REQUIERMENTS FOR MASTER DEGREE OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY

UNIVERSITI PUTRA MALAYSIA

JUNE 2017

Abstract

Abstract submitted to the Faculty of Computer Science and Information Technology in Fulfillment of the Requirements for the Degree of Master of Science.

An Efficient Algorithm for Independent Tasks Scheduling in Heterogeneous Computing Systems

By

Abdulrahman Khaldoon Mahmood Al-Qadhi

June 2016

Chairman: Mr. Ahmad Alauddin Ariffin

Faculty:Computer Science and Information Technology

Task scheduling is critical in heterogenous systems, especially with the huge number of tasks transmitted over grid. Since heuristics are proposing methods for solving heterogenous computing systems, several techniques were proposed for the scheduling on grid computing systems to get better execution time. In this thesis, a proposed new heuristic algorithm named Two Stages TasksTransfer (TSTT) algorithm introduced as an enhancement of state of TPB scheduling algorithm. Scheduling problem addressed Heterogeneous Computing Scheduling Problem (HCSP) mathematical model, where the independent tasks assigned to heterogenous processors with different characteristics. Twelve datasets with different heterogeneity level examined using different heuristic algorithms to compare the performance with the new algorithm. The proposed algorithm showed its efficiency in term of makespan, resource utilization and load balancing metrics for set of tasks.

Acknowledgment

I would like to express my sincere gratitude to almighty god for helping me, giving me the strength and taking care of me through my study journey.

I would also like to thank my parents for their unlimited support. Without you I wouldn't have reached thus far. Thank you from the bottom my heart.

Special thanks for my sister for her support and for believing in me and for her continuous support.

I would love to express my gratitude to my supervisor Mr. Ahmad Alauddin Ariffin for his guidance through my research. As well as all computer science faculty lecturers who enriched me with their knowledge.

Finally, I would like to thank my friends and colleagues; You were my family here in in Malaysia, and you gave me hope, support and love. Thank you.

SUPERVISOR CONFIRMATION

Project Title:

AN EFFICIENT ALGORITHM FOR INDEPENDENT TASK SCHEDULING IN HETEROGENOUS COMPUTING SYSTEMS

Has been prepared by:

ABDULRAHMAN KHALDOON MAHMOOD AL-QADHI (GS42993)

Thesis has been supervised by:

MR. AHMAD ALAUDDIN ARIFFIN

Checked and approved by: Date: …….

…….
(Mr. Ahmad Alauddin Ariffin)
Master Project Supervisor
Department of Communication Technology and Network
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
43400 UPM Serdang, Selangor

Declaration

I hereby declare that the project based on my original work except for quotations and citations, which have been duly acknowledged. I also declare that it has not been previously or currently submitted for any other degree at UPM or any other institution.

_____

Abdulrahman Al-Qadhi

Date: ……..……

Table of Contents

List of tables

List of figures

List of Abbreviation

CPU	Central Processing Unit
EFT	Earliest Finishing Time
ETC	Estimated Time to Complete
GA	Genetic Algorithm
GAP	The gap between optimal and current solution
GHz	Gigahertz
GriPhyN	Grid Physics Network
GTSD	Grouped Tasks by Standard Deviation
HC	Heterogenous Computing
HCSP	Heterogenous computing scheduling problem
HPC	High Performance Computing
LB	Lower bound execution time
LHCGrid	Large Hadron Collider Grid
LJFR-SJFR	Long Job First – Short Job First
LSufferage	List Sufferage
MCT	Minimum Completion Time
meanCT	Mean Completion Time
MET	Minimum Execution Time
NP-hard	non-deterministic polynomial-time hard
OLB	Opportunistic Load balancing
PB	Penalty Based
Qos	Quality of Service
TPB	Tenacious penalty based
TSTT	Two Stages Task Transfer

CHAPTER 1

INTRODUCTION

This chapter provides an introduction to the research, problem statement, objectives, scope of the study and research contribution.

1.1. Overview

In the last decade, the high speed low-cost computers production led to boost the processing power through networks; this fact helps to grow in distribution of distributed systems to contain and tap these resources (Gogos et al., 2016a, Nesmachnow et al., 2010).

A heterogeneous computing (HC) referring to set of heterogeneous elements called computers, resources, machines or processors), these resources connected to each other by a network (public or private), for instance, Ethernet (Nesmachnow et al., 2010). Fig. 1.1 shows the heterogeneous computing environment.

Grid computing is a famous example of distributed computing resources working over loosely coupled virtual machines, created by connecting many heterogeneous computing resources with different characteristics. Grid infrastructure provides an easy low cost solutions to access for distributed computing resources distributed over the globe to solve large problems (Foster and Kesselman, 2003).

Fig. 1.1. HC environment (Izakian et al., 2009b)

The main issue when using a distributed computing systems environment is finding an efficient way for task scheduling to satisfy the user need of fast computing power, where tasks should be assigned to the resources in a clever way (Nesmachnow et al., 2010).

The quality of heterogeneous resources refers to the variety of capabilities for both resources and interconnected technology (e.g. data transmission, CPU's processing power and data transfer speed).

In general, HC system involves a parallel and/or distributed high-performance resources, cooperate in order to provide support to solve the intensive-computational applications. HC systems used also in scientific and high-performance applications to solve complex problems with high computational capabilities (Khokhar et al., 1993, Freund et al., 1998, Eshaghian, 1996, Braun et al., 2001).

Scheduling in HC system is critical because it is related to quality of service metrics, such as resource usage, efficiency, economic profit, etc. In other word, task scheduling is task planning or task assignment, which has been an important focus in this line of research in the last two decades (Qureshi et al., 2014).

1.2. Problem statement

An issue of independent task scheduling in heterogeneous systems is critically important. Because of the need for an efficient scheduling technique to solve complex computational problems in grid and other heterogeneous environments like cloud computing, many researches have been done in order to reach the optimality. A literature review conducted to have a clear view about the techniques used to solve the scheduling problem and provide a super solution with a lower execution time. Another main issue is finding an accredited comparison between these solutions without ignoring other important performance evaluation metrics like: resource utilization, and resource imbalance ratio, which can be play an essential role in evaluating the algorithm.

1.3. Objectives

The main objectives in this thesis are:

1- To find improved solution for heterogeneous computing scheduling problems.
2- To perform a performance evaluation for the proposed algorithm, and compare it with other scheduling algorithms (e.g. Suffefage, LSufferage, Penalty based, and Tenacious Penalty Based algorithms), using makespan, resource utilization and load imbalance ratio metrics for this purpose.

1.4. Scope of study

a- Independent tasks.

b- Each task submitted once to one processor.
c- Non- preemptive system.
d- Each processor executes one task at time.
e- Machines always available.

1.5. Research contribution

a- An efficient algorithm proposed and developed to schedule the tasks in distributed system considering all heterogeneity probabilities (Section 3.5).
b- Performance evaluation for a proposed algorithm conducted to show the efficiency of the proposed algorithm by comparing it with other well-known heuristic algorithms (Section 4.4.1, 4.4.2 and 4.4.3).

1.6. Thesis Organization

The next chapters will be organized as follows: The chapter 2 will discuss the related work to the study area heterogeneous computing scheduling. An extensive review conducted to show what the grid computing and what the researchers done to solve the scheduling problem in the grid. Chapter 3 discusses the methodology for conducting the research (i.e. datasets, model, algorithms and performance evaluation). Chapter 4 presents the results and discusses it to show the performance evaluation for the proposed algorithm. Finally, chapter 5 is the conclusion, which summarizes the problem and the proposed solution, and presents the future work.

REFERENCES

ALI, S., SIEGEL, H. J., MAHESWARAN, M. & HENSGEN, D. Task execution time modeling for heterogeneous computing systems. Heterogeneous Computing Workshop, 2000.(HCW 2000) Proceedings. 9th, 2000. IEEE, 185-199.

APTER, M. J. 1970. The computer simulation of behaviour.

ATIQUE, M. Efficient scheduling of independent tasks using modified heuristics. Computing and Communications Technologies (ICCCT), 2015 International Conference on, 2015. IEEE, 373-378.

BATISTA, D. M. & DA FONSECA, N. L. A brief survey on resource allocation in service oriented grids. Globecom Workshops, 2007 IEEE, 2007. IEEE, 1-5.

BIANCHI, L., DORIGO, M., GAMBARDELLA, L. M. & GUTJAHR, W. J. 2009. A survey on metaheuristics for stochastic combinatorial optimization. *Natural Computing,* 8, 239-287.

BOTE-LORENZO, M. L., DIMITRIADIS, Y. A. & GÓMEZ-SÁNCHEZ, E. Grid characteristics and uses: a grid definition. Grid Computing, 2004. Springer, 291-298.

BRAUN, T. D., SIEGEL, H. J., BECK, N., BÖLÖNI, L. L., MAHESWARAN, M., REUTHER, A. I., ROBERTSON, J. P., THEYS, M. D., YAO, B. & HENSGEN, D. 2001. A comparison of eleven static heuristics for mapping a class of independent tasks onto heterogeneous distributed computing systems. *Journal of Parallel and Distributed computing,* 61, 810-837.

BUYYA, R. 2002. Economic-based distributed resource management and scheduling for grid computing. *arXiv preprint cs/0204048*.

CAO, L., LIU, X., WANG, H. & ZHANG, Z. 2014. OPT-Min-Min scheduling algorithm of grid resources. *Journal of Software,* 9, 1868-1875.

CASANOVA, H., LEGRAND, A., ZAGORODNOV, D. & BERMAN, F. Heuristics for scheduling parameter sweep applications in grid environments. Heterogeneous Computing Workshop, 2000.(HCW 2000) Proceedings. 9th, 2000. IEEE, 349-363.

CHATURVEDI, A. K. & SAHU, R. 2011. New heuristic for scheduling of independent tasks in computational grid. *International Journal of Grid and Distributed Computing,* 4, 25-36.

CHENG, R. & GEN, M. 1997. Parallel machine scheduling problems using memetic algorithms. *Computers & Industrial Engineering,* 33, 761-764.

DAVIDOVIĆ, T., TEODOROVIĆ, D. & ŠELMIĆ, M. 2015. Bee colony optimization-Part I: The algorithm overview. *Yugoslav Journal of Operations Research,* 25, 33-56.

ESHAGHIAN, M. M. 1996. *Heterogeneous computing*, Artech House Publishers.

ETMINANI, K. & NAGHIBZADEH, M. A min-min max-min selective algorihtm for grid task scheduling. Internet, 2007. ICI 2007. 3rd IEEE/IFIP International Conference in Central Asia on, 2007. IEEE, 1-7.

EZZATTI, P., PEDEMONTE, M. and MARTÍN, Á., 2013. An efficient implementation of the Min-Min heuristic. Computers & Operations Research, 40(11), pp.2670-2676.

FOSTER, I. What is the grid? A three point checklist. GRIDtoday, 1 (6), 2002. Workshop on Dependable Distributed Data Management (WDDDM'09), 2002a. 14-17.

FOSTER, I. 2002b. The grid: A new infrastructure for 21st century science. *Physics today,* 55, 42-47.

FOSTER, I. & KESSELMAN, C. 2003. *The Grid 2: Blueprint for a new computing infrastructure*, Elsevier.

FREUND, R. & SUNDERAM, V. 1994. Special issue on heterogeneous processing. *Journal of Parallel and Distributed Computing,* 21, 255-256.

FREUND, R. F., GHERRITY, M., AMBROSIUS, S., CAMPBELL, M., HALDERMAN, M., HENSGEN, D., KEITH, E., KIDD, T., KUSSOW, M. & LIMA, J. D. Scheduling resources in multi-user, heterogeneous, computing environments with SmartNet. Heterogeneous Computing Workshop, 1998.(HCW 98) Proceedings. 1998 Seventh, 1998. IEEE, 184-199.

GALSTYAN, A., CZAJKOWSKI, K. & LERMAN, K. 2005. Resource allocation in the grid with learning agents. *Journal of Grid Computing,* 3, 91-100.

GHOSH, T. K., DAS, S., BARMAN, S. & GOSWAMI, R. A Comparison Between Genetic Algorithm and Cuckoo Search Algorithm to Minimize the Makespan for

Grid Job Scheduling. Advances in Computational Intelligence: Proceedings of International Conference on Computational Intelligence 2015, 2017. Springer, 141-147.

GHOSH, T. K., GOSWAMI, R., BERA, S. & BARMAN, S. Load balanced static grid scheduling using Max-Min heuristic. Parallel Distributed and Grid Computing (PDGC), 2012 2nd IEEE International Conference on, 2012. IEEE, 419-423.

GOGOS, C., VALOUXIS, C., ALEFRAGIS, P., GOULAS, G., VOROS, N. & HOUSOS, E. 2016a. Scheduling independent tasks on heterogeneous processors using heuristics and Column Pricing. *Future Generation Computer Systems,* 60, 48-66.

GOGOS, C., VALOUXIS, C., ALEFRAGIS, P., XANTHOPOULOS, I. & HOUSOS, E. 2016b. Scheduling independent tasks on heterogeneous computing systems by optimizing various objectives.

HAUPT, R. L. & HAUPT, S. E. 2004. *Practical genetic algorithms*, John Wiley & Sons.

HOOS, H. H. & STÜTZLE, T. 2004. *Stochastic local search: Foundations and applications*, Elsevier.

IBARRA, O. H. & KIM, C. E. 1977. Heuristic algorithms for scheduling independent tasks on nonidentical processors. *Journal of the ACM (JACM),* 24, 280-289.

IORDACHE, G., BOBOILA, M., POP, F., STRATAN, C. & CRISTEA, V. 2008. Decentralized grid scheduling using genetic algorithms. *Metaheuristics for Scheduling in Distributed Computing Environments.* Springer.

ITURRIAGA, S., NESMACHNOW, S., LUNA, F. & ALBA, E. 2015. A parallel local search in CPU/GPU for scheduling independent tasks on large heterogeneous computing systems. *The Journal of Supercomputing,* 71, 648-672.

IZAKIAN, H., ABRAHAM, A. & SNASEL, V. Comparison of heuristics for scheduling independent tasks on heterogeneous distributed environments. Computational Sciences and Optimization, 2009. CSO 2009. International Joint Conference on, 2009a. IEEE, 8-12.

IZAKIAN, H., ABRAHAM, A. & SNÁŠEL, V. 2009b. Metaheuristic based scheduling meta-tasks in distributed heterogeneous computing systems. *Sensors,* 9, 5339-5350.

IZAKIAN, H., LADANI, B. T., ZAMANIFAR, K. & ABRAHAM, A. A novel particle swarm optimization approach for grid job scheduling. International Conference on Information Systems, Technology and Management, 2009c. Springer, 100-109.

KAMALAM, G. & BHASKARAN, V. M. 2012. New enhanced heuristic min-mean scheduling algorithm for scheduling meta-tasks on heterogeneous grid environment. *Eur. J. Sci. Res,* 70, 423-430.

KAMALAM, G. & MURALIBHASKARAN, V. 2010. A new heuristic approach: Min-Mean algorithm for scheduling meta-tasks on heterogeneous computing systems. *International Journal of Computer Science and Network Security,* 10, 24-31.

KAZEM, A. A. P., RAHMANI, A. M. & AGHDAM, H. H. A modified simulated annealing algorithm for static task scheduling in grid computing. Computer Science and Information Technology, 2008. ICCSIT'08. International Conference on, 2008. IEEE, 623-627.

KHAN, S. & ARDIL, C. A Game Theoretical Energy Efficient Resource Allocation Technique for Large Distributed Computing Systems. PDPTA, 2009. 48-54.

KHAN, S. U. A goal programming approach for the joint optimization of energy consumption and response time in computational grids. Performance Computing and Communications Conference (IPCCC), 2009 IEEE 28th International, 2009. IEEE, 410-417.

KHAN, S. U. & AHMAD, I. 2009. A cooperative game theoretical technique for joint optimization of energy consumption and response time in computational grids. *IEEE Transactions on Parallel and Distributed Systems,* 20, 346-360.

KHOKHAR, A. A., PRASANNA, V. K., SHAABAN, M. E. & WANG, C.-L. 1993. Heterogeneous computing: Challenges and opportunities. *Ieee Computer,* 26, 18-27.

Kim, T. Y., & Oh, H. (n.d). Grid computing. Retrieved from http://selab.inha.ac.kr/selab2/index.html?p=research/grid

KRAUTER, K., BUYYA, R. & MAHESWARAN, M. 2002. A taxonomy and survey of grid resource management systems for distributed computing. *Software: Practice and Experience,* 32, 135-164.

LIN, S. C. & YEN, E. 2011. *Data Driven E-Science: Use Cases and Successful Applications of Distributed Computing Infrastructures (ISGC 2010)*, Springer Science & Business Media.

LIU, H., ABRAHAM, A. & HASSANIEN, A. E. 2010. Scheduling jobs on computational grids using a fuzzy particle swarm optimization algorithm. *Future Generation Computer Systems,* 26, 1336-1343.

Load Balance and Parallel Performance. (2010, January 19). Retrieved from https://software.intel.com/en-us/articles/load-balance-and-parallel-performance

MAHESWARAN, M., ALI, S., SIEGEL, H. J., HENSGEN, D. & FREUND, R. F. 1999. Dynamic mapping of a class of independent tasks onto heterogeneous computing systems. *Journal of parallel and distributed computing,* 59, 107-131.

MAKHORIN, A. 2006. GNU linear programming kit, version 4.9. GNU Software Foundation.

MITCHELL, M. 1998. *An introduction to genetic algorithms*. MIT press.

MONTANA, D. & ZINKY, J. 2008. Optimizing Routing and Backlogs for Job Flows in a Distributed Computing Environment. *Metaheuristics for Scheduling in Distributed Computing Environments.* Springer.

NESMACHNOW, S. 2010. Parallel evolutionary algorithms for scheduling on heterogeneous computing and grid environments.

NESMACHNOW, S. & CANABÉ, M. GPU implementations of scheduling heuristics for heterogeneous computing environments. XVII Congreso Argentino de Ciencias de la Computación, 2011.

NESMACHNOW, S., CANCELA, H. & ALBA, E. 2010. Heterogeneous computing scheduling with evolutionary algorithms. *Soft Computing,* 15, 685-701.

OMARA, F. A. & ARAFA, M. M. 2010. Genetic algorithms for task scheduling problem. *Journal of Parallel and Distributed Computing,* 70, 13-22.

PEDEMONTE, M., NESMACHNOW, S. & CANCELA, H. 2011. A survey on parallel ant colony optimization. *Applied Soft Computing,* 11, 5181-5197.

PINEL, F., & DORRONSORO, B., 2014. Savant: Automatic generation of a parallel scheduling heuristic for map-reduce. International Journal of Hybrid Intelligent Systems, 11(4), 287-302.

QURESHI, M. B., DEHNAVI, M. M., MIN-ALLAH, N., QURESHI, M. S., HUSSAIN, H., RENTIFIS, I., TZIRITAS, N., LOUKOPOULOS, T., KHAN, S. U. & XU, C.-Z. 2014. Survey on grid resource allocation mechanisms. *Journal of Grid Computing,* 12, 399-441.

RAJABIOUN, R. 2011. Cuckoo optimization algorithm. *Applied soft computing,* 11, 5508-5518.

RANJAN, R., HARWOOD, A. & BUYYA, R. 2006. A taxonomy of peer-to-peer based complex queries: a grid perspective. *arXiv preprint cs/0610163*.

RANJAN, R., HARWOOD, A. & BUYYA, R. 2008. Peer-to-peer-based resource discovery in global grids: a tutorial. *IEEE Communications Surveys & Tutorials,* 10.

ROCHE, T., SUMMERVILLE, M. 2016. The Future of Grid Computing in a Cloud-Enabled World, Retrieved from https://research.tabbgroup.com/report/v14-059-future-grid-computing-cloud-enabled-world.

SAID, M. P. & KOJIMA, I. 2009. S-MDS: Semantic monitoring and discovery system for the grid. *Journal of Grid Computing,* 7, 205-224.

SKILLICORN, D. B. Motivating computational grids. Cluster Computing and the Grid, 2002. 2nd IEEE/ACM International Symposium on, 2002. IEEE, 401-401.

TABAK, E. K., CAMBAZOGLU, B. B. & AYKANAT, C. 2014. Improving the performance of independenttask assignment heuristics minmin, maxmin and sufferage. *IEEE Transactions on Parallel and Distributed Systems,* 25, 1244-1256.

TANG, P., YEW, P.-C. & ZHU, C.-Q. Impact of self-scheduling order on performance on multiprocessor systems. Proceedings of the 2nd international conference on Supercomputing, 1988. ACM, 593-603.

TEODOROVIĆ, D., ŠELMIĆ, M. & DAVIDOVIĆ, T. 2015. Bee Colony Optimization-part II: The application survey. *Yugoslav Journal of Operations Research,* 25, 185-219.

WANG, L., SIEGEL, H. J., ROYCHOWDHURY, V. P. & MACIEJEWSKI, A. A. 1997. Task matching and scheduling in heterogeneous computing environments using a genetic-algorithm-based approach. *Journal of parallel and distributed computing,* 47, 8-22.

XHAFA, F. & ABRAHAM, A. 2008. *Meta-heuristics for grid scheduling problems*, Springer.

XHAFA, F., ALBA, E. & DORRONSORO, B. Efficient batch job scheduling in grids using cellular memetic algorithms. Parallel and Distributed Processing Symposium, 2007. IPDPS 2007. IEEE International, 2007a. IEEE, 1-8.

XHAFA, F., CARRETERO, J., BAROLLI, L. & DURRESI, A. 2007b. Immediate mode scheduling in grid systems. *International Journal of Web and Grid Services,* 3, 219-236.

YAGIURA, M. & IBARAKI, T. 2001. On metaheuristic algorithms for combinatorial optimization problems. *Systems and Computers in Japan,* 32, 33-55.

YANG, X.-S. & DEB, S. Cuckoo search via Lévy flights. Nature & Biologically Inspired Computing, 2009. NaBIC 2009. World Congress on, 2009. IEEE, 210-214.

YEO, C. S., BUYYA, R., DE ASSUNCAO, M. D., YU, J., SULISTIO, A., VENUGOPAL, S. & PLACEK, M. 2007. Utility computing on global grids. *The Handbook of Computer Networks,* 3.

YOUSIF, A., NOR, S. M., ABDULLAH, A. H. & BASHIR, M. B. 2014. A Discrete Firefly Algorithm for Scheduling Jobs on Computational Grid. *Cuckoo Search and Firefly Algorithm.* Springer.

ZHANG, Z., FENG, Y., ZHANG, S. & SUN, Y. 2015. Heuristic Grid Resource Scheduling Algorithm based on Group of Task and Secondary Distribution. *International Journal of Security and Its Applications,* 9, 427-438.