



UNIVERSITI PUTRA MALAYSIA

**MODIFIED WORKFLOW SCHEDULING USING HYBRID PSO-GA
ALGORITHM IN CLOUD COMPUTING**

OMOTAYO PATRICK OKE

FSKTM 2019 34



UPM
UNIVERSITI PUTRA MALAYSIA
BERILMU BERBAKTI

**MODIFIED WORKFLOW SCHEDULING USING HYBRID
PSO-GA ALGORITHM IN CLOUD COMPUTING**

**OMOTAYO PATRICK OKE
GS48833**

**MASTER OF COMPUTER SCIENCE
UNIVERSITI PUTRA MALAYSIA**

2019



**Modified Workflow Scheduling Using Hybrid PSO-GA Algorithm
in Cloud Computing**

By

**Omotayo Patrick Oke
GS48833**

**Thesis Submitted To The School Of Graduate Studies, Universiti
Putra Malaysia, In Fulfilment Of The Requirements For The
Master Of Computer Science**

July 2019

DEDICATION

“With joy and in glorious heart, I dedicate this work to God of heaven and earth for His unfailing love shine upon my life that makes my studies goes smooth as possible. Also to my Father and Mother that gave me the moral backings. Finally, with warmest heart, I dedicate this work to my dearest, lovely and God sent wife and my wonderfully made children as they were there with me whenever I felt weary to attend classes, they will be with me to the school sitting in the car. Sweet-heart, a very big THANK to you and the children.”

Abstract of thesis presented to Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Master of Computer Science

**Modified Workflow Scheduling Using Hybrid PSO-GA Algorithm in
Cloud Computing**

By

**Omotayo Patrick Oke
GS48833**

July 2019

**Chairman: Mr. Mohd Noor Bin Derahman, Master
Faculty: Computer Science and Information Technology**

ABSTRACT

Cloud computing environment offers customers with multiple on-demand services and resource sharing. Business processes are managed using cloud-based workflow technology, which is one of the difficulties of using resources efficiently owing to the inter-task dependencies. A Hybrid GA-PSO algorithm is suggested in this paper to effectively allocate duties. The goal of the Hybrid GA-PSO algorithm is to reduce the makespan and cost and balance the load of dependent tasks in cloud computing environments over the heterogenous resources. Results of the experiment show that the PSO-GA algorithm reduces the complete workflow execution time, Compared to GA, PSO and GA-PSO. It also decreases the cost of implementation. It also increases the workflow application's load balancing over accessible resources. Finally, the findings acquired also showed that the suggested algorithm converges more quickly and with greater performance than other algorithms to ideal alternatives.

Abstrak tesis yang disampaikan kepada Senat Universiti Putra Malaysia sebagai pemenuhan keperluan untuk ijazah Sarjana Sains Komputer

Penjadualan Aliran Kerja Diperbodakan Menggunakan Algoritma PSO-GA Hybrid di Pengkomputeran Awan

Oleh

**Omotayo Patrick Oke
GS48833**

Julai 2019

**Pengerusi: Encik Mohd Noor Bin Derahman, Sarjana
Fakulti: Sains Komputer dan Teknologi Maklumat**

ABSTRAK

Persekitaran pengkomputeran awan menawarkan pelanggan dengan pelbagai perkhidmatan atas permintaan dan perkongsian sumber. Proses perniagaan diuruskan menggunakan teknologi aliran kerja berasaskan awan, yang merupakan salah satu daripada kesukaran menggunakan sumber yang berkesan disebabkan oleh kebergantungan antara tugas. Algoritma GA-PSO Hybrid dicadangkan dalam karya ini untuk mengagihkan tugas dengan berkesan. Matlamat algoritma Hybrid GA-PSO adalah untuk mengurangkan makepan dan kos dan mengimbangi beban tugas yang bergantung kepada persekitaran pengkomputeran awan ke atas sumber heterogonous. Keputusan eksperimen menunjukkan bahawa algoritma PSO-GA mengurangkan masa pelaksanaan aliran kerja yang lengkap, Berbanding GA, PSO dan GA-PSO. Ia juga mengurangkan kos pelaksanaan. Ia juga meningkatkan keseimbangan beban aplikasi aliran kerja ke atas sumber yang boleh diakses. Akhirnya, penemuan yang diperoleh juga menunjukkan bahawa algoritma yang dicadangkan itu lebih cepat dan dengan prestasi yang lebih baik daripada algoritma lain untuk alternatif yang ideal.

ACKNOWLEDGEMENTS

I had a lot of support and aid throughout the writing of this dissertation. First of all, I would like to thank my supervisor, Mr Mohd Noor Bin Derahman, whose knowledge has been invaluable especially in formulating the study subject and methodology.

I want to recognize my peers at the University Putra Malaysia. For their amazing cooperation. You have been very supportive of me and always ready to assist me. I would like to highlight Mr. Said Mahmoud in particular, I would like to thank you for your outstanding advice and support throughout my thesis.

I would also like to thank all the other lecturers who have been instrumental in their precious advice through this dissertation. You gave me the instruments I required to choose the correct path and finish my dissertation effectively.

I would also like to thank my parents for their wise advice and moral support. For me, you're always there. Finally, I want to acknowledge my dearest, lovely and God sent wife and my wonderfully made children as they were there with me all the way. Sweet-heart, a very big THANK to you and the children.

I certify that a Thesis Examination Committee has met on July 2019 to conduct the final examination of Omotayo Patrick Oke on his thesis entitled “Modified Workflow Scheduling Using Hybrid PSO-GA Algorithm in Cloud Computing” in accordance with the Universities and University Colleges Act 1971 and the Constitution of the University Putra Malaysia [(P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Master of Science

Members of the Thesis Examination Committee were as follows:

Mohd Noor Derahman (Mr.)

Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Supervisor)

Masnida Hussin (AP. Dr.)

Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Assessor)

Hazlina Hamdan (Dr.)

Master Program Coordinator
University Putra Malaysia
Faculty of Computer Science and
Information Technology
Date:

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfillment of the requirement for the degree of Master of Computer Science. The members of the Supervisory Committee were as follows:

Mohd Noor Derahman (Mr.)
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Supervisor)

Masnida Hussin (AP. Dr.)
Faculty of Computer Science and
Information Technology
Universiti Putra Malaysia
(Assessor)

I certify that a Thesis Examination Committee has met on July 2019 to conduct the final examination of Omotayo Patrick Oke on his thesis entitled “Modified Workflow Scheduling Using Hybrid PSO-GA Algorithm in Cloud Computing” in accordance with the Universities and University Colleges Act 1971 and the Constitution of the University Putra Malaysia [(P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Master of Science

Members of the Thesis Examination Committee were as follows:

Mohd Noor Derahman (Mr.)
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Supervisor)

Masnida Hussin (AP. Dr.)
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Assessor)

Hazlina Hamdan (Dr.)
Master Program Coordinator
University Putra Malaysia
Faculty of Computer Science and
Information Technology
Date:

DECLARATION

I hereby declare that the thesis is 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 concurrently submitted for any other degree at UPM or other institutions.

Signature:

Date: 01/07/2019

Name: **Omotayo Patrick Oke**

Matric No: GS48833

Table of Contents

CHAPTER ONE	1
1.0 Overview	1
1.2 Motivation	2
1.3 Research Problem	2
1.4 Research Objective	4
1.5 Scope of Research	4
1.6 Contributions of Research	4
1.7 Thesis Organization	5
CHAPTER TWO	6
LITERATURE REVIEW	6
2.0 Introduction	6
2.1 Workflow	6
2.2 Scientific Workflow	7
2.3 Workflow Classification	8
2.4 Workflow Life cycle	9
2.5 Workflow Scheduling	11
2.6 Workflow Scheduling Techniques	12
2.7 Related Works on Workflow Scheduling	14
CHAPTER THREE	19
WORKFLOW SCHEDULING PROBLEM IN HETEROGENEOUS ENVIRONMENTS	19
3.0 Introduction	19
3.1 Proposed Hybrid PSO-GA Algorithm	19
3.1.1 Initializing Population	20
3.1.2 Applying the GA Algorithm	21
i. Selection Operator	21
3.2 Simulation	22
3.3 Dataset	23
3.4 Performance Evaluation	23
CHAPTER FOUR	25
AN EFFICIENT PSO-GA ALGORITHM FOR WORKFLOW TASK SCHEDULING IN CLOUD ENVIRONMENTS	25
4.0 Introduction	25
4.1 Performance Analysis	25
4.2 The PSO-GA Algorithm	26

CHAPTER FIVE	28
RESULT, RECOMMENDATION AND FEATURE WORK, AND CONCLUSION.....	28
5.0 Introduction.....	28
5.1 Result Discussion.....	28
5.3 Recommendation and Feature Work.....	30
5.2 Conclusion	30
LIST OF FIGURES	32
Fig. 2.1: Known scientific workflows	32
Fig. 2.2: Workflow Life Cycle.....	32
Fig. 4.1: Makespan.....	32
Fig. 4.2: Load Balancing.....	32
Fig. 4.3: Execution Cost.....	32
REFERENCES.....	33
APPENDIX.....	38

CHAPTER ONE

INTRODUCTION

1.0 Overview

One of the emerging trends in IT environment is the cloud computing with various requirements and huge resources. As modern computing, applications are capturing large volume of data which would require powerful computing resource to process. Therefore, efficient load balancing guarantees an efficient resource utilization to cope with the increasing users' demands. More so the need to process task with very huge storage capacity are fast growing. For this reason, cloud computing is receiving attention due its high-performance in providing computing services and facilities to users. Cloud computing is an emerging technology in which user charged on pay-as-use basis. A cloud provider in cloud computing provides services on the basis of clients' requests. The services provided by cloud computing includes Software as a Service (SaaS), Infrastructure as a Service (IaaS), and Platform as a Service (PaaS) (Aljammal et al., (2007); Liu et al., (2015), Manasrah et al., (2017)). Workflow scheduling is one of the key issues in the management of workflow execution. Scheduling is a process that maps and manages execution of inter-dependent tasks on distributed resources. It introduces allocating suitable resources to workflow tasks so that the execution can be completed to satisfy objective functions specified by users. Proper scheduling can have significant impact on the performance of the system.

Fortunately, with the advent of clouds, workflows applications are executed in the cloud. The cloud computing employs a good number of scheduling technique to efficiently allocate task to available cloud resources. This include Min-Min, Max-Min, X-Suffrage, Genetic Algorithm, Particle Swarm Optimization algorithms, to a name a few.

1.2 Motivation

Due to its acceptance, capability and applicability by researcher to provide optimal decision for the multi objective workflow optimization, various applications are modeled as workflow applications with a set of tasks and the dependencies between them such that one task can only be execute after which its dependent tasks have completed their execution. Due to its significance, workflow applications are being used in a range of domains, such as astrophysics, bioinformatics, and disaster modeling and prediction. Moreover, complicated problems like complex scientific applications are emerging recently and can only be address through the combination of various methods and techniques in a single solution, as these applications are often multi-object scheduling problem.

1.3 Research Problem

Though workflow applications are the mechanism of a large-scale business process execution, which consisting of a set of events or tasks which are dependent on each other. However, the order of their execution must be considered when assigning the tasks to VM processors. Assigning the dependent tasks to the most appropriate VM processors is not that easy. The scheduling processes of the workflow applications are considered as a multi objective optimization problem. There is always a trade-off between minimize the money cost and the execution time for the work flow application with efficient load balancing over the VMs in the cloud environment to have optimal decision for the multi objective workflow optimization. The workflow scheduling problem is an inherited problem from the heterogeneous computing environments, for which different research efforts were made to address the scheduling problem (chen et al., (2009); Talukder et al., (2009);

Wieczorek et al., (2009)). However, heterogeneous computing environments are not easy to set up, and their capability of giving more uniform performance with less failure is quite limited in comparison to the cloud environments (Li et al., (2017), Stergiou et al., (2018)). Due to this, several metaheuristic algorithms were proposed to solve the scheduling problem of the workflow tasks and to get an efficient solution for tasks distribution over the different VMs in the cloud environment. For instance, Genetic Algorithm (GA) (Dasgupta et al., (2013)), Ant Colony Optimization (Z. Zhang & X. Zhang (2010)), Swarm Intelligence (Braun et al., (2001)), and Artificial Bee Colony (ABC) (Rana et al., (2014)) are few examples of the various proposed solutions of workflow scheduling problem but did not address the total monetary cost and the execution makespan. To this effect, we intend address the workflow scheduling problem in heterogeneous computing environments. Basically, our idea is not to only minimizing the workflow execution time, but also to consider the total monetary cost and the execution makespan in order to get an optimal decision for the multi objective workflow optimization problem.

- The benchmark paper carried out scheduling using GA (Generic Algorithm) and PSO (Particle Swarm Optimization) then proposed Hybrid in the order of GA-PSO for a better optimization but the identified problem is that after GA runs then PSO get value from GA to run in the hybrid.
- The problem is that it is known that standard PSO Algorithm get trap in local optimum solutions which cause premature convergence. The main reason for premature convergence is that the particles in the population are highly affected by particle's personal best and global best, especially in the later stage of iterations.

- Then if you try to make the velocity steps very small, such that particles cannot overfly the region but then it will take a lot of time to explore the whole region and solutions will take longer execution time.

1.4 Research Objective

The main objective of this research work is primarily to propose an efficient algorithm that addresses the workflow scheduling problem over an heterogeneous cloud environment, by;

1. Reducing the makespan of the workflow tasks
2. Reducing the cost of workflow execution
3. Balancing the load of the workflow tasks on heterogeneous VMs in the selected cloud.

1.5 Scope of Research

The scope of this research work is to address the workflow scheduling problem in heterogeneous environments by optimizing the workflow execution time, total monetary cost and the execution makespan of the workflow task.

1.6 Contributions of Research

With the proposition of Hybrid PSO-GA algorithm, the workflow scheduling problem was addressed by combining the individual strength of GA and PSO algorithms to address the workflow scheduling problem. The proposed algorithm was able to reduce the total make span execution time and also balance the load over the VMs with minimum total monetary cost.

1.7 Thesis Organization

Chapter 1 is an introductory chapter that discusses the problem statement of the research, the objective, and the scope of the research.

Chapter 2 is the background chapter which presents the fundamental concepts of workflow scheduling problem. workflow scheduling scientific problem, workflow patterns, classifications, workflow life cycle, workflow scheduling and related works. Literature review chapter that explains the previous existing techniques

Chapter 3 is the research methodology chapter. It describes how this research work was conducted. Also, the chapter discusses the different phases in this research work and the methodology followed during each phase. The measurement metrics and the datasets that have been used in the simulation of this research are presented.

Chapter 4 explains the detail steps of the proposed approach for addressing workflow scheduling problem. This chapter illustrates the phases of the proposed approach.

Chapter 5 presents the results of the simulation conducted to evaluate the performance of the proposed approaches compared with the most relevant existing approaches. The chapter also discusses the results with respect to different parameters. It also reflects the conclusions and the contributions of this research.

REFERENCES

- Abdulal., W., Jabas, A., Ramachandram, S., & Jadaan, O. A. (2012). Grid computing -technology and applications, widespread coverage and new horizons. In D. S. Maad (Ed.), (p. 89-110). InTech.
- Aljammal., A. H., Manasrah, A. M., Abdallah, A. E., & Tahat, N. M. (2017). A new architecture of cloud computing to enhance the load balancing. *International Journal of Business Information Systems*, 25(3), 393-405.
- Arabnejad, H., & Barbosa, J. G. (2014). A budget constrained scheduling algorithm for workflow applications. *Journal of grid computing*, 12(4), 665-679.
- Braun, T. D., Siegel, H. J., Beck, N., Bölöni, L. L., Maheswaran, M., Reuther, A. I., ... & Freund, R. F. (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(6), 810-837.
- Chitra, S., Madhusudhanan, B., Sakthidharan, G. R., & Saravanan, P. (2014). Local minima jump PSO for workflow scheduling in cloud computing environments. In *Advances in computer science and its applications* (pp. 1225-1234). Springer, Berlin, Heidelberg.
- Couvares, P., Kosar, T., Roy, A., Weber, J., & Wenger, K. (2007). Workflow management in condor. In I. J. Taylor, E. Deelman, D. B. Gannon, & M. Shields (Eds.), *Workflows for e-science*. (p. 357-375). Springer London.
- Dasgupta, K., Mandal., B., Dutta, P., Mandal., J. K., & Dam, S. (2013). A genetic algorithm (ga) based load balancing strategy for cloud computing. *Procedia Technology*, 10, 340-347.
- Daoud, M. I., & Kharma, N. N. (2011). A hybrid heuristic-genetic algorithm for task scheduling in heterogeneous processor networks. *J. Parallel Distrib. Comput.*, 71(11), 1518-1531.

de Carvalho, E. C. A., Jayanti, M. K., Batilana, A. P., Kozan, A. M. O., Rodrigues, M. J., Shah, J., . . . Pietrobon, R. (2010). Standardizing clinical trials workflow representation in uml for international site comparison. *Public Library of Science, One (PLOS One)*, 11(5), 1-8.

Fard, H. M., Prodan, R., Barrionuevo, J. J. D., & Fahringer, T. (2012, May). A multi-objective approach for workflow scheduling in heterogeneous environments. In *2012 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (ccgrid 2012)* (pp. 300-309). IEEE.

Ge, Y., & Wei, G. (2010, October). GA-based task scheduler for the cloud computing systems. In *2010 International Conference on Web Information Systems and Mining* (Vol. 2, pp. 181-186). IEEE.

Guirado, F., Roig, C., & Ripoll, A. (2013). Enhancing throughput for streaming applications running on cluster systems. *Journal of Parallel and Distributed Computing*, 73(8), 1092-1105.

Guo-ning, G., Ting-lei, H., & Shuai, G. (2010a). Genetic simulated annealing algorithm for task scheduling based on cloud computing environment. In *International conference on intelligent computing and integrated systems (iciss)*.

Guo, L., Zhao, S., Shen, S., & Jiang, C. (2012). Task scheduling optimization in cloud computing based on heuristic algorithm. *Journal of networks*, 7(3), 547.

Hoheisel, A., & Alt, M. (2007). Workflows for e-science: Scientific workflows for grids. In I. J. Taylor, E. Deelman, D. B. Gannon, & M. Shields (Eds.), (p. 190-207). Springer London.

Kumar, P., & Verma, A. (2012, August). Scheduling using improved genetic algorithm in cloud computing for independent tasks. In *Proceedings of the International Conference on Advances in Computing, Communications and Informatics* (pp. 137-142). ACM.

- Li, J., Liu, Z., Chen, X., Xhafa, F., Tan, X., & Wong, D. S. (2015). L-EncDB: A lightweight framework for privacy-preserving data queries in cloud computing. *Knowledge-Based Systems*, 79, 18-26.
- Li, P., Li, J., Huang, Z., Gao, C. Z., Chen, W. B., & Chen, K. (2018). Privacy-preserving outsourced classification in cloud computing. *Cluster Computing*, 21(1), 277-286.
- Mao, Y., Chen, X., & Li, X. (2014). Max–min task scheduling algorithm for load balance in cloud computing. In *Proceedings of International Conference on Computer Science and Information Technology* (pp. 457-465). Springer, New Delhi.
- Manasrah, A. M., Smadi, T., & Almomani, A. (2017). A Variable Service Broker Routing Policy for data center selection in cloud analyst. *Journal of King Saud University-Computer and Information Sciences*, 29(3), 365-377.
- Munir, E. U., Mohsin, S., Hussain, A., Nisar, M.W., & Ali, S. (2013). SDBATS: A novel algorithm for task scheduling in heterogeneous computing systems. In *Ieee 27th international parallel and distributed processing symposium workshops phd forum (ipdpsw)* (p. 43-53).
- Omara, F. A., & Arafa, M. M. (2010). Genetic algorithms for task scheduling problem. *Journal of Parallel and Distributed Computing*, 70, 13-22.
- Pandey, S., Wu, L., Guru, S. M., & Buyya, R. (2010, April). A particle swarm optimization-based heuristic for scheduling workflow applications in cloud computing environments. In *2010 24th IEEE international conference on advanced information networking and applications* (pp. 400-407). IEEE.
- Rana, M., Bilgaiyan, S., & Kar, U. (2014, July). A study on load balancing in cloud computing environment using evolutionary and swarm-based algorithms. In *2014 International Conference*

on Control, Instrumentation, Communication and Computational Technologies (ICCICCT) (pp. 245-250). IEEE.

Stergiou, C., Psannis, K. E., Kim, B. G., & Gupta, B. (2018). Secure integration of IoT and cloud computing. *Future Generation Computer Systems*, 78, 964-975.

Talukder, A. K. A., Kirley, M., & Buyya, R. (2009). Multi-objective differential evolution for scheduling workflow applications on global Grids. *Concurrency and Computation: Practice and Experience*, 21(13), 1742-1756.

Topcuoglu, H., & Wu, M.-Y. (2002). Performance-effective and low-complexity task scheduling for heterogeneous computing. *IEEE Transactions on Parallel and Distributed Systems*, 13, 260-274.

Verma, A., & Kaushal, S. (2015). Cost-time efficient scheduling plan for executing workflows in the cloud. *Journal of grid computing*, 13(4), 495-506.

Vydyanathan, N., Catalyurek, U., Kurc, T., Sadayappan, P., & Saltz, J. (2011). Optimizing latency and throughput of application workflows on clusters. *Parallel Computing*, 37, 694-712.

Wieczorek, M., Hoheisel, A., & Prodan, R. (2009). Towards a general model of the multi-criteria workflow scheduling on the grid. *Future Generation Computer Systems*, 25(3), 237-256.

Wu, Z., Liu, X., Ni, Z., Yuan, D., & Yang, Y. (2013). A market-oriented hierarchical scheduling strategy in cloud workflow systems. *The Journal of Supercomputing*, 63(1), 256-293.

Xu, H., Yang, B., Qi, W., & Ahene, E. (2016). A multi-objective optimization approach to workflow scheduling in clouds considering fault recovery. *KSII Transactions on Internet and Information Systems (TIIS)*, 10(3), 976-995.

Yu, J., Buyya, R., & Ramamohanarao, K. (2008). Workflow scheduling algorithms for grid computing. In *Metaheuristics for scheduling in distributed computing environments* (pp. 173-214). Springer, Berlin, Heidelberg.

Zhang, L., Chen, Y., Sun, R., Jing, S., & Yang, B. (2008). A task scheduling algorithm based on PSO for grid computing. *International Journal of Computational Intelligence Research*, 4(1), 37-43.

Zhang, Z., & Zhang, X. (2010, May). A load balancing mechanism based on ant colony and complex network theory in open cloud computing federation. In *2010 The 2nd International Conference on Industrial Mechatronics and Automation* (Vol. 2, pp. 240-243). IEEE.