Computers and Electrical Engineering 000 (2017) 1–17

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

Genetic-based algorithms applied to a workflow scheduling algorithm with security and deadline constraints in clouds[☆]

Henrique Yoshikazu Shishido[a,b,*], Júlio Cezar Estrella[a], Claudio Fabiano Motta Toledo[a], Marcio Silva Arantes[a]

[a] *Institute of Mathematics and Computer Science (ICMC), University of São Paulo (USP), SP, Brazil*
[b] *Department of Computing, Federal Technological University of Paraná, Cornélio Procópio, PR, Brazil*

ARTICLE INFO

Article history:
Received 1 May 2017
Revised 1 December 2017
Accepted 2 December 2017
Available online xxx

Keywords:
Workflow scheduling
Cloud computing
Security
Genetic algorithm
Particle swarm optimization
Deadline

ABSTRACT

There have been a number of metaheuristic scheduling techniques for cloud described in the literature, as well as their applications. The efficiency of metaheuristic techniques has been established in a wide range of workflow scheduling algorithms for cloud environments. However, it is still unknown whether the metaheuristic that is chosen, is suitable for solving the problem of optimization. This paper examines the effect of both Particle Swarm Optimization (PSO) and Genetic-based algorithms (GA) on attempts to optimize workflow scheduling. A security and cost-aware workflow scheduling algorithm was selected to evaluate the performance of the metaheuristics. Three algorithms were evaluated in three real-world workflows with a risk rate constraint that ranged between 0 and 1 with a 0.1 step. The findings indicate that GA-based algorithms significantly outperformed the PSO both in term of cost-effectiveness and response time.

1. Introduction

scientific workflow is a recent paradigm for distributed programming that is deployed in computational experiments in scientific A areas such as physics, astronomy, and biology. A workflow can be defined as a procedure involving a series of steps designed to simplify the complexity of executing and managing applications [1]. It is commonly represented as a Directed Acyclic Graph (DAG), where each node carries out a task, and each edge denotes a precedence or flow constraints between the tasks. Previously, workflows were deployed in computational grids. However, owing to the increase in the complexity of the scientific applications for handling big data, more powerful and scalable infrastructures are needed to run complex workflows within a reasonable amount of time [2].

Cloud computing is an infrastructure which can be rapidly and elastically provisioned on demand [3]. It can offer different virtual machine configurations that are capable of executing workflows. Clouds can be classified as private, community-based, public or hybrid clouds. Private cloud is owned and used by a single organization, and it does not charge the services it offers. Community-based clouds are designed for a specific community that has shared concerns. Public cloud shares services with multiple-tenants and these are charged on the basis of their use. Finally, hybrid cloud consists of a mix of public,

☆ Reviews processed and recommended for publication to the Editor-in-Chief by Guest Editor Dr. L. Bittencourt.
* Corresponding author at: Institute of Mathematics and Computer Science (ICMC), University of São Paulo (USP), SP, Brazil.
E-mail addresses: shishido@usp.br (H.Y. Shishido), jcezar@icmc.usp.br (J.C. Estrella), claudio@icmc.usp.br (C.F.M. Toledo), marcio@icmc.usp.br (M.S. Arantes).

https://doi.org/10.1016/j.compeleceng.2017.12.004

communal and private clouds, where public/communal cloud resources are integrated with private cloud to create a single environment. There is a wide range of cloud services (e.g., virtual machine (VM) types, storage policy, and cloud data center locations) [1,4]. For this reason, scheduling is a key factor in a workflow executed in a cloud environment and has been the subject of investigation in recent years [5]. Scheduling algorithms are designed to map out the tasks of the workflow for VMs based on scheduling criteria and subject to the constraints of the users. However, the heterogeneity of these tasks and wide range of cloud services, lead to an NP-hard optimization problem. Carrying out optimal scheduling within a reasonable time is a challenge because some variables have to be taken into account such as the fact that there are (a) several types of VMs with different capacities and prices, (b) tasks with heterogeneous loads, and (c) other attributes that impede the process of optimization.

In view of this, metaheuristic techniques have been employed in workflow scheduling in clouds to find a near-optimal scheduling scheme. However, research on the use of metaheuristics for workflow scheduling in cloud has been largely restricted to a single technique. In a recent study, it was found that Particle Swarm Optimization (PSO) is the most widely employed technique, and more than 50% of the scheduling algorithms were based on metaheuristics [6]. These findings suggest that there is a need to understand whether scheduling algorithms that are optimized by metaheuristics can achieve a better performance than any alternative and whether PSO is, in fact, a good choice.

This paper examines the effects of different metaheuristics on a workflow scheduling algorithm for cloud. Our major objective was to investigate the differences between the optimization obtained by Particle Swarm Optimization, the Genetic Algorithm (GA), and the Multi-Population Genetic Algorithm (MPGA). The approach adopted in this study involves a mixed methodology based on a security-aware and cost-aware workflow scheduling algorithm [7] that applied PSO to optimize the combinatorial scheduling scheme, and Genetic-based algorithms. Real-world workflows were deployed to evaluate the suitability of each metaheuristic. In addition, an analysis was conducted of the time spent until the stagnation of each optimization technique. Statistical Tests were conducted that assumed $p \leq 0.05$ to check if there is a significant difference between PSO and GA-based algorithms.

The findings should be a useful addition to solutions provided by workflow scheduling algorithms based on multidimensional optimization, and show an efficient method for achieving an optimization beyond that of PSO. However, the reader should bear in mind that this study does not claim that workflow scheduling problems optimized by GA-based algorithms always outperform PSO. As mentioned earlier, the optimization efficiency of GA-based and PSO algorithms has only been examined in a single workflow scheduling algorithm.

The remainder of the paper is structured as follows: Section 2 summarizes the related work. Section 3 formulates the problem and system models in detail. The methodology employed for the experiments is outlined in Section 4. Section 5 explains the experimental planning and examines the results of the evaluation. Finally, in Section 6 there are some concluding remarks and suggestions for future research in the area.

2. Related work

Scheduling scientific workflows is an important and challenging area of research in cloud computing. There have been numerous studies on the algorithms needed to find the optimal workflow scheduling in different scenarios and comply with the constraints of the users. For instance, workflow scheduling algorithms have been created to enable users to meet their deadline [4,5] although these incur high execution costs. Thus, other studies have addressed the issue of budgetary constraints in workflow scheduling in an attempt to limit the costs to a maximum value defined by the user [8]. However, none of the scheduling methods cited above has been able to find a near-optimal solution. Instead, they optimize the scheduling at the task level and thus ignore the benefits that can be obtained from taking account of the features of the entire workflow structure to achieve a near-global optimal solution. Other studies use metaheuristic algorithms such as Genetic Algorithms, Simulated Annealing (SA), Particle Swarm Optimization, Cat Swarm Optimization (CSO) and Ant Colony Optimization (ACO) to find approximate solutions that can be obtained faster than those of traditional exhaustive algorithms with regard to the computing time.

Masdari et al. [9] conducted an in-depth analysis of the PSO-based task and workflow scheduling schemes devised for a cloud environment. A dynamic scheduling of intensive workflows for reducing costs is put forward by Pandey et al. [10]. This scheme optimizes the cost of the task-resource mapping using PSO and takes into account the computation and data transmission costs. Wu et al. [11] employed a Revised Discrete PSO (RDPSO) algorithm to mitigate the high overhead of data transfers in cloud environments. The main goal of this scheme is to reduce the computation cost under a deadline constraint. Jianfang et al. [12] employed a discrete PSO to mitigate the security threats of the cloud environment. The purpose of this algorithm is to find feasible solutions that meet the demands of security, makespan and cost constraints. Energy-consumption is another issue in the cloud data center, and this is addressed in [13]. The authors combined PSO with dynamic voltage and a frequency scaling technique to reduce the costs and makespan, as well as the energy consumption of the cloud data center. Verma and Kaushal [14] introduced a bi-objective algorithm based on PSO (BPSO) to reduce the execution cost and makespan while meeting the deadline and overcoming the budgetary constraints. In Xue and Wu [15], there is a hybrid approach to PSO combined with a hill climbing algorithm to improve the ability to carry out a local search and maintain the diversity of the population.

Some workflow scheduling algorithms apply a basic GA, while others adapt it to improve the results. Furthermore, most of them use a heuristic to ensure there is a more suitable initial population so that better results can be achieved. Aryan and

Please cite this article as: H.Y. Shishido et al., Genetic-based algorithms applied to a workflow scheduling algorithm with security and deadline constraints in clouds, Computers and Electrical Engineering (2017), https://doi.org/10.1016/j.compeleceng.2017.12.004