Sustainable Computing: Informatics and Systems 41 (2024) 100942

Contents lists available at ScienceDirect

Sustainable Computing: Informatics and Systems

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

Priority based job scheduling technique that utilizes gaps to increase the efficiency of job distribution in cloud computing☆

Saydul Akbar Murad [a,d], Zafril Rizal M. Azmi [a,*], Abu Jafar Md. Muzahid [a], Md. Murad Hossain Sarker [b], M. Saef Ullah Miah [a], MD. Khairul Bashar Bhuiyan [c], Nick Rahimi [d], Anupam Kumar Bairagi [e]

[a] Faculty of Computing, College of Computing & Applied Sciences, Universiti Malaysia Pahang, Pekan, Pahang, 26600, Malaysia
[b] Comilla University (CoU), Comilla, 3506, Bangladesh
[c] Brac University, Dhaka, 1212, Bangladesh
[d] School of Computing Sciences & Computer Engineering, University of Southern Mississippi, Hattiesburg, 39401, USA
[e] Khulna University, Khulna, 9208, Bangladesh

ARTICLE INFO

Keywords:
Cloud computing
Job scheduling
Backfilling
Resource management
Gap searching
SJF
LJF
FCFS

ABSTRACT

A growing number of services, accessible and usable by individuals and businesses on a pay-as-you-go basis, are being made available via cloud computing platforms. The business services paradigm in cloud computing encounters several quality of service (QoS) challenges, such as flow time, makespan time, reliability, and delay. To overcome these obstacles, we first designed a resource management framework for cloud computing systems. This framework elucidates the methodology of resource management in the context of cloud job scheduling. Then, we study the impact of a Virtual Machine's (VM's) physical resources on the consistency with which cloud services are executed. After that, we developed a priority-based fair scheduling (PBFS) algorithm to schedule jobs so that they have access to the required resources at optimal times. The algorithm has been devised utilizing three key characteristics, namely CPU time, arrival time, and job length. For optimal scheduling of cloud jobs, we also devised a backfilling technique called Earliest Gap Shortest Job First (EG-SJF), which prioritizes filling in schedule gaps in a specific order. The simulation was carried out with the help of the CloudSim framework. Finally, we compare our proposed PBFS algorithm to LJF, FCFS, and MAX–MIN and find that it achieves better results in terms of overall delay, makespan time, and flow time.

1. Introduction

With the rapid adoption of virtualized Information technology (IT) components such as databases, servers, and on-demand storage, a growing variety of cloud services are being delivered to consumers and businesses under a billing model over the wired and wireless Internet. This is because cloud computing, which is the most popular model for resource delivery, allows any resource to be delivered as a service, including networks, platforms, software, and applications. Similar to social economy services, quality of service (QoS) requirements in the cloud, such as cost, energy consumption, security, and reliability, are the main challenges for cloud service providers, which are receiving more and more attention from enterprises and academia [1–3]. Among these obstacles, the flow time, makespan time, and resource utilization are three of the most important concerns for cloud customers and providers.

Currently, popular enterprise schedulers, like First Come First Served (FCFS) [4], are widely used to manage jobs because of their simplicity and flexibility of implementation. FCFS schedules jobs based on the order in which they are submitted and the availability of resources required for each job. Although FCFS offers several benefits and is widely used, its fairness attribute often results in low utilization of resources because jobs cannot be executed if there are older jobs in the queue, even if the resources are available that are required. In addition, the FCFS approach is ineffective for scheduling interactive

☆ All persons who have made substantial contributions to the work reported in the manuscript (e.g., technical help, writing and editing assistance, general support), but who do not meet the criteria for authorship, are named in the Acknowledgments and have given us their written permission to be named. If we have not included an Acknowledgments, then that indicates that we have not received substantial contributions from non-authors.

* Corresponding author.
E-mail addresses: zafril@ump.edu.my (Z.R.M. Azmi), mrumi98@gmail.com (A.J.M. Muzahid), mh6367828@gail.com (M.M.H. Sarker), md.saefullah@gmail.com (M.S.U. Miah), kbashar707@gmail.com (M.K.B. Bhuiyan), nick.rahimi@usm.edu (N. Rahimi), anupam@cse.ku.ac.bd (A.K. Bairagi).

https://doi.org/10.1016/j.suscom.2023.100942
Received 10 April 2023; Received in revised form 17 November 2023; Accepted 22 November 2023
Available online 30 November 2023