AN EFFICIENT RESOURCE ALLOCATION FRAMEWORK FOR CLOUD USING SERIES QUEUE WITH BLOCKING MODEL

[1]Dr. S. Saravanan, [2]Dr. P. Anbalagan

[1,2]Assistant Professor
Department of Computer Science and Engineering
Annamalai University, Annamalainagar, Tamil Nadu, India

Abstract: In recent two decades cloud and its associated technologies has made a tremendous change in IT infrastructure. The usage of cloud has invoked many new technologies and raise in mobile and connected technologies. These cloud platforms are remote computers with many virtual servers. These cloud platforms requires the allocation of resources effectively. In order to address this issue, this paper proposes a novel and effective resource allocation and scheduling scheme which emphasis an efficient mathematical model. The key idea of the proposed scheme is to effectively utilize the series queue with blocking model. Further, to validate the efficacy of the proposed model a real time validation is carried out with the real infrastructure of 64 GB RAM, Intel Xeon Processor running in Cent Operating System. From the experimentation, it is confirmed that the proposed model is efficient and effective in resource scheduling and allocation. Further, an analysis is performed to prove the efficacy and the analysis results also reveals that the proposed scheme is effective and outperforms the state of the art method in terms of accuracy and fast processing.

IndexTerms **- Resource Allocation, Cloud Scheduling, Virtual Machine, Physical Machine**

I. INTRODUCTION

In recent years, Cloud computing services offer high performance computing on big data applications. Cloud computing provides an infrastructure platform and software resources to users in the form of leases [1]. Cloud providers place user jobs on suitable physical servers and dynamically allocate computing resources to user jobs in form of virtual machines. The cloud platform segments the physical resources such as CPU, memory and storage space into a limited number of virtual machines via virtualization techniques [2]. Energy consumption plays a major role in data centers, The Microsoft updates that the physical resources in data centers will account for 45% of the total cost and energy consumption will account for 15% [3]. In past few years energy consumption in data centers has been doubled. Therefore, energy consumption is one of the major issues in the data centers. The users can easily access the required cloud resources, while some of the underlying infrastructures are hidden from the users. Cloud providers have full control of the infrastructures. Cloud provider must schedule the user job in a way which minimizes the completion time [4].

The paper is organized as follows: section 2 deals with the literature review whereas the section 3 explains the proposed method. Section 4 exhibits the experimental setup. Section 5 analyzes the performance of the proposed method with the state of the art schemes. Finally, the paper is concluded in the section 6.

II. LITERATURE REVIEW

Cloud computing sets a replacement paradigm for infrastructure management by giving unprecedented prospects to deploy software system in distributed environments. Its goal is to share resources among the cloud users, cloud partners, and cloud vendors within the cloud price chain. With the exponential growth of cloud computing as an answer for providing flexible computing resource, a lot of and a lot of cloud applications emerge in recent years. For instance, Google, Microsoft, Yahoo, and IBM square measure apace deploying knowledge centers in numerous locations round the world to deliver cloud computing service. A way to build reliable and complicated cloud application has become associate degree imperative and crucial analysis drawback. In cloud computing, resource programming is that the method of distribution of accessible resources to the required user request over the net. The resource programming method in cloud computing in the main considers jobs, virtual machines and physical machines [5].

The resource programming method is extremely advanced because of the subsequent reasons: initial, the resources hosted beneath cloud computing have advanced provisioning, composition, configuration, and preparation needs [6]. These options need that the modeling language has the powerful expressibility and quantifiability, that not solely characterizes the structure of cloud computing, however additionally provides the suggestions to compose the model of basic parts. Ancient approaches cannot be simply applied in modeling dynamic resource programming method because of the variation of system and user needs [7].

Second, cloud computing is dynamic as a result of its resources and services may be mechanically elected at run-time support specific requests, and resource in cloud application is also vulnerable to unsure factors, like server failure, malicious threats, the failure of cloud service. It is difficult to set up resource programming strategy at the planning time and account for all dynamic behaviors throughout the execution, that results in a loss of responsibility management over cloud computing. Third, several applications, e.g., financial transactions and scientific computing, are real-time in nature, where the correctness depends not solely on the computation results, however additionally on the time instants that they become accessible[8].

Furthermore, it is essential to preserve the power to provision virtual machine inside the point. Finally, the characteristics of cloud computing, like fast snap and multi tenant bring on several difficult verification problems that have profound influence on the performance of system [9]. Most attempts to validate mechanisms cannot be much possible to verify the correctness

of accommodative resource programming strategy, in face of advanced and numerous behaviors in cloud computing. Resource programming involves the dynamic re-optimization and distribution of system resources at runtime.

Hardwiring these functionalities into the code of a cloud application ends up and that's difficult to take care of. Aspect-Oriented Programming (AOP) will improve the software system adaptivity and scale back the coupling between modules that is appropriate for describing the extremely dynamic cloud application. We are able to abstract the resource programming method as a separate side by exploitation of AOP, which is able to enhance the standard of product code and scale back the quality of resource programming method [10]. However, AOP is especially employed in the coding stage, rather than the entire method of package style and the necessities of aspect-oriented analysis often appears ambiguous and vague. Formal strategies can be used to analyze the resource scheduling method in cloud computing, so increasing the linguistics constraints.

Petri net could be a mathematically primarily based technique for modeling and verificatory package artifacts. The mixture of Petri nets with AOP is an efficient methodology, which might be accustomed to analyze resource-planning method. And also the formalization permits the analysis and verification of operational properties of resource scheduling model. Cloud computing present's variety of challenges concerning the management of tasks and resources, like price constraints and completion time needs [11]. From the literature, it is clearly stated that the existing schemes used for resource scheduling and allocation are not efficient and hence there is a need for an effective schemes, which can allocate the resources effectively. Hence, this paper proposes an effective resource allocation scheme.

III. PROPOSED METHODOLOGY

In this research, a novel and efficient resource allocation framework for cloud has been proposed. The proposed methodology utilizes the series queue with blocking model for effective resource allocation for the cloud virtual machines. Figure. 1 shows the general approach for jobs running in the job queue.

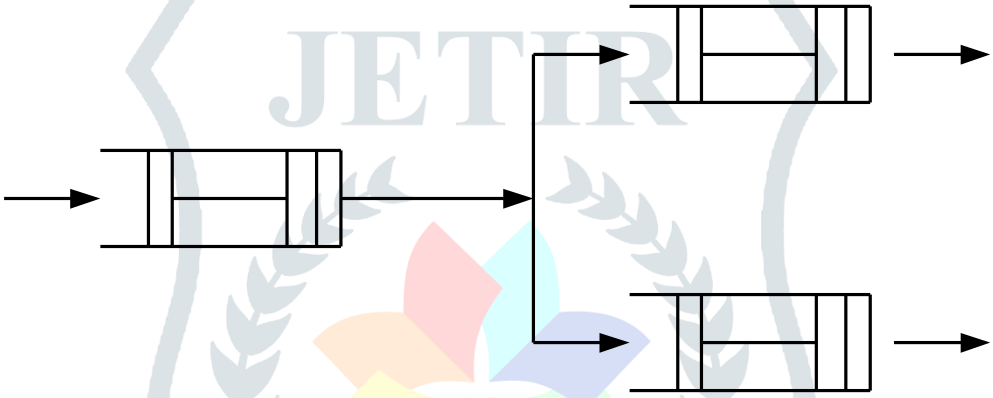

Fig.1 Job queue: general view

The key idea of series queue with blocking model is the job entering the server or cloud will be allocated to the two or more sibling running parallel to the each queues. Once the job gets finished then all the services are re-organized back and sent as the combined. Whenever the size of the queue exceeds, then the queue will apply restrictions in the blocking. These blockings are then handled using buffer where the job waits until the allocation engine is set free to proceed with the further jobs. The novelty of the proposed scheme is the model can work either with or without synchronizing the queue

Fig.2 Scheduler with blocking queue model

Furthermore, the model supports the schemes such as transfer blocking, rejection blocking where the blocking (Figure. 2) is effectively handled (Figure. 3).

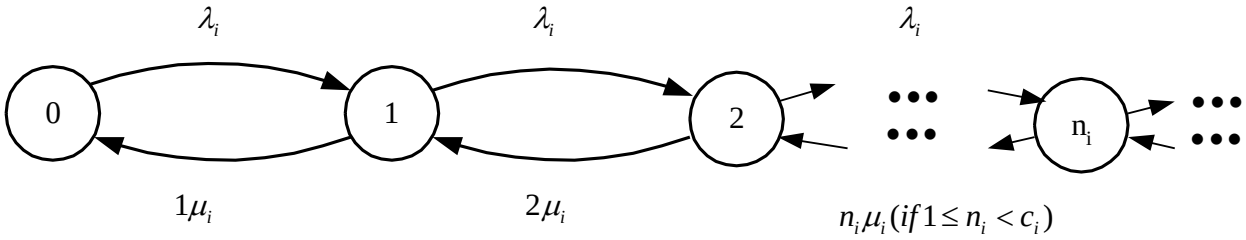

Fig.3 Stochastic balance of the scheduling queue

IV. EXPERIMENTAL SETUP

Figure. 4 shows the exact setup carried out for experimentation.

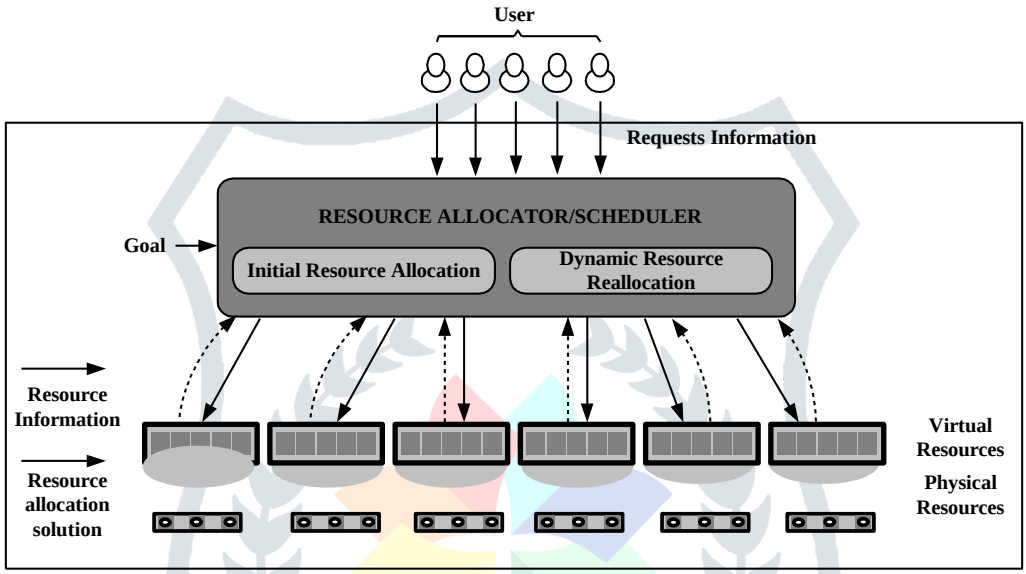

Fig.4. Experimental setup for the proposed model

V. PERFORMANCE ANALYSIS

To test the efficacy of the proposed scheme, a performance analysis is carried out with the state of the art schemes. Figures.5 – 9, shows the performance analysis results of the proposed scheme with the state of the art schemes.

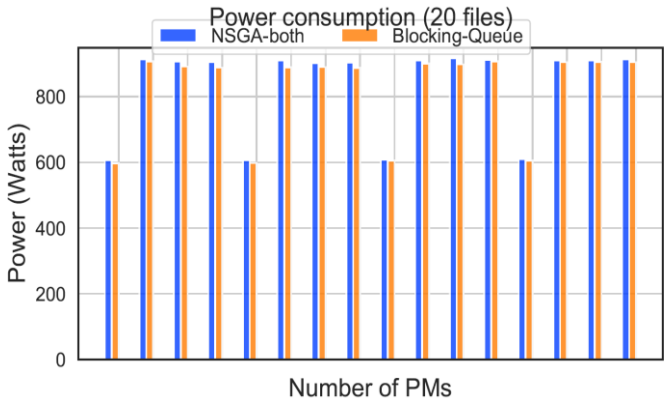

Fig.5. Performance analysis - File number

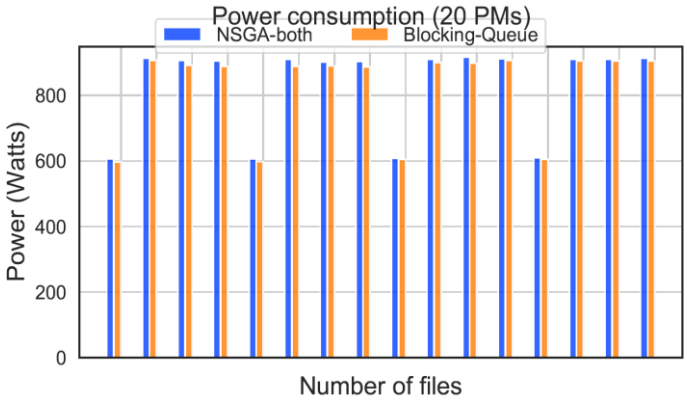

Fig.6. Performance analysis – Physical machine size

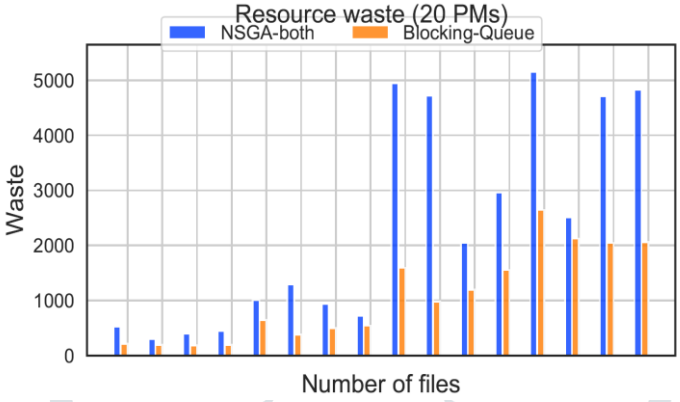

Fig.7. Performance analysis – resource wastage (20 PMs)

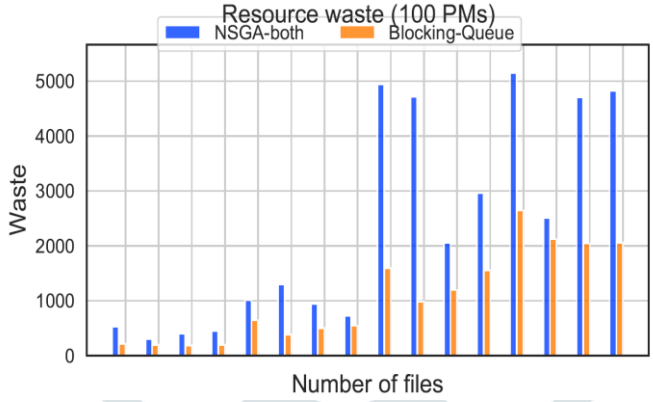

Fig.8. Performance analysis – resource wastage (100 PMs)

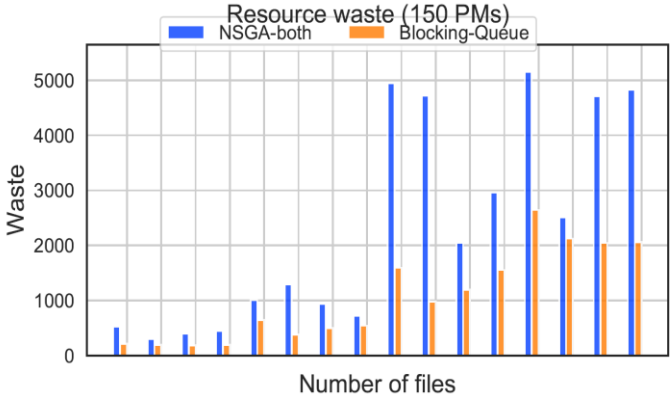

Fig.9. Performance analysis – resource wastage (150 PMs)

VI. CONCLUSION

This paper proposed a novel scheduling scheme for resource allocation. From the experimental results, it is confirmed that the proposed scheme is better in terms of fast allocation and processing. Further, it is also confirmed that the proposed scheme outperforms the state of the art schemes available in the literature. Experimental results produced 93% fast processing in allocation for both the virtual machine and physical machine allocation.

REFERENCES

[1] W. Xia and L. Shen, "Joint resource allocation using evolutionary algorithms in heterogeneous mobile cloud computing networks," in *China Communications*, vol. 15, no. 8, pp. 189-204, Aug. 2018.

[2] L. Zhao, M. Du and L. Chen, "A new multi-resource allocation mechanism: A tradeoff between fairness and efficiency in cloud computing," in *China Communications*, vol. 15, no. 3, pp. 57-77, March 2018.

[3] D. Lu, J. Ma and N. Xi, "A universal fairness evaluation framework for resource allocation in cloud computing," in *China Communications*, vol. 12, no. 5, pp. 113-122, May 2015.

[4] J. Son and R. Buyya, "Priority-Aware VM Allocation and Network Bandwidth Provisioning in Software-Defined Networking (SDN)-Enabled Clouds," in *IEEE Transactions on Sustainable Computing*, vol. 4, no. 1, pp. 17-28, 1 Jan.-March 2019.

[5] X. Wang, X. Wang, H. Che, K. Li, M. Huang and C. Gao, "An Intelligent Economic Approach for Dynamic Resource Allocation in Cloud Services," in *IEEE Transactions on Cloud Computing*, vol. 3, no. 3, pp. 275-289, 1 July-Sept. 2015.

[6] X. Xu, W. Dou, X. Zhang and J. Chen, "EnReal: An Energy-Aware Resource Allocation Method for Scientific Workflow Executions in Cloud Environment," in IEEE Transactions on Cloud Computing, vol. 4, no. 2, pp. 166-179, 1 April-June 2016.

[7] J. N. Khasnabish, M. F. Mithani and S. Rao, "Tier-Centric Resource Allocation in Multi-Tier Cloud Systems," in *IEEE Transactions on Cloud Computing*, vol. 5, no. 3, pp. 576-589, 1 July-Sept. 2017.

[8] H. Wang, Z. Kang and L. Wang, "Performance-Aware Cloud Resource Allocation via Fitness-Enabled Auction," in *IEEE Transactions on Parallel and Distributed Systems*, vol. 27, no. 4, pp. 1160-1173, 1 April 2016.

[9] Q. Li, J. Zhao, Y. Gong and Q. Zhang, "Energy-efficient computation offloading and resource allocation in fog computing for Internet of Everything," in *China Communications*, vol. 16, no. 3, pp. 32-41, March 2019.

[10] L. Wei, C. H. Foh, B. He and J. Cai, "Towards Efficient Resource Allocation for Heterogeneous Workloads in IaaS Clouds," in *IEEE Transactions on Cloud Computing*, vol. 6, no. 1, pp. 264-275, 1 Jan.-March 2018.

[11] K. Chard and K. Bubendorfer, "Co-Operative Resource Allocation: Building an Open Cloud Market Using Shared Infrastructure," in *IEEE Transactions on Cloud Computing*, vol. 7, no. 1, pp. 183-195, 1 Jan.-March 2019.