FDMC: Framework for Decision Making in Cloud for Efficient Resource Management

Alexander Ngenzi1, Selvarani R2, Suchithra R3
1Department Computer Science Engineering, Jain University, Bangalore, India
2Computer Science Engineering, Alliance University, Bangalore, India
3Department Master of Science in Information Technology, Jain University, Bangalore, India

ABSTRACT
An effective resource management is one of the critical success factors for precise virtualization process in cloud computing in presence of dynamic demands of the user. After reviewing the existing research work towards resource management in cloud, it was found that there is still a large scope of enhancement. The existing techniques are found not to completely utilize the potential features of virtual machine in order to perform resource allocation. This paper presents a framework called FDMC or Framework for Decision Making in Cloud that gives better capability for the VMs to perform resource allocation. The contribution of FDMC is a joint operation of VM to ensure faster processing of task and thereby withstand more number of increasing traffic. The study outcome was compared with some of the existing systems to find FDMC excels better performance in the scale of task allocation time, amount of core wasted, amount of storage wasted, and communication cost.

Article Info
Article history:
Received Aug 30, 2016
Revised Oct 19, 2016
Accepted Nov 2, 2016

Keyword:
Cloud computing
Decision making
Resource allocation
Resource management
Virtual machine

1. INTRODUCTION
With the increasing demands of data and service availability, the present era of cloud computing is also undergoing a challenging situation [1]. It is found that majority of the existing studies are carried out considering deployment model and service model of the cloud. The deployment model will consist of public, private and hybrid cloud [2] while the service model will consist of PaaS (Platform-as-a-Service), IaaS (Infrastructure-as-a-Service), and SaaS (Software-as-a-Service) [3]. All the services of cloud computing reach its user using resource virtualization mechanism [4]. However, the challenge in the virtualization process in cloud is its higher dependencies on storage and channel / processing capacity as compared to conventional physical server. The host machine can easily deploy virtual machine by replicating it among them in order to exploit the complete utilization of resources. However, this theoretical claim is very different from practices user experiences [5]. In order to cater up the dynamic needs of the user, an effective resource management is highly required in cloud environment [6]. The term resource will mean both software as well as hardware that are required to perform a specific task e.g. storage, CPU, network, memory, channel capacity etc [7], [8]. The phenomenon of resource management essentially deals with various deployment strategies, selection criteria of various software and hardware in order to conform a better and fault tolerance services to the users [9]. As the work load is characterised by heavy fluctuation over cloud, an effective resource management is highly required to be implemented. Resource allocation is one of the prime process in resource management that leads to mapping of physical servers and workloads to the Virtual Machine (VM) [10]. Various components of resource management over cloud environment are load balancer, SLA

Corresponding Author:
Alexander Ngenzi,
Computer Science Engineering,
Jain University,
Bangalore, India.
Email: alexanderngenzi@gmail.com
A framework for decision making in cloud for efficient resource management. The study mainly captures the memory and computational resource utilization. Xu and Li [20] have presented a scheme that performing monitoring of an essential resources over the cloud using agent-based approach. The study mainly captures the memory and computational resource utilization. Xu and Li [20] have presented a framework that considering steady matching modelling in order to decouple various procedures of VM mapping principle over physical servers. Using experimental based approach, the proposed system was tested on 20 node clusters. Yu et al [21] have developed a technique of resource allocation of heterogeneous types for catering up multiple demands on the cloud-based networks. The prime aim of the techniques was to resist utilization of skewed resources over physical server using resource-based prediction approach. The study outcome was assessed using arrival rate, number of active physical server, utilization, skewness, and delay. Study on dynamic allocation of resources was also carried out by Yang et al [15]. The presented technique performs autonomous migration of the jobs among the VMs depending on the amount of load. The result was assessed using time with increasing size of problem and CPU utilization using Open Nebula. Selvarani and Sadhasivam [16] have presented a task scheduling scheme over cloud in order to perform mapping of the required resources. The cost of resources as well as performance of computation is estimated by the presented technique and its outcome was analyzed with respect to time and cost. Usage of optimization scheme was also seen in literatures towards resource management policy formulation. The work done by Liu et al [17] has implemented ant colony optimization as well as genetic algorithm that significantly assist in performing faster search toward optimal results. The study outcome was compared with individual outcomes from Genetic algorithm and ant colony optimization too. Zhu et al [18] have introduced a framework that performs provisioning of resources and effective scheduling of task. The authors have incorporated a software-engineering based technique in order to perform scheduling. Panda et al [19] have discussed about an algorithm that targets multiple environments of cloud based on smoothening concept. The evaluation of the study was carried out using bigger dataset of heterogeneous types. Caton et al. [20] have presented a very unique technique that uses the potential networking attributes of social network in order to carry out resource allocation in cloud. The implementation of the technique was carried out over open source framework using stochastic modelling of node participation process. Yu et al [21] have developed a technique of resource management over vehicular network with an aid of VM migration schemes. A hierarchical architecture is presented in this scheme that performs sharing of the resources over vehicle cloud network. The study outcome was evaluated with respect to rate of arrival of local VMs. Meera and Swamynathan [22] have introduced a scheme that performing monitoring of an essential resources over the cloud using agent-based approach. The study mainly captures the memory and computational resource utilization. Xu and Li [23] have presented a framework that considering steady matching modelling in order to decouple various procedures of VM mapping principle over physical servers. Using experimental based approach, the proposed system was tested on 20 node clusters. Yu et al [24] have presented a technique of sharing and effective management of cloud resources (exclusively for channel capacity). The architecture adopted by the author is nearly similar to social internet of vehicles.
The following are the problem identifications of the proposed study:

a. The existing research work towards resource allocation doesn’t focus on synchronous factor of VMs for which reason task scheduling among the VM is still unsolved issue.

b. The decision of resource allocation are mainly implemented using deterministic approach which is known for its overhead creation over highly distributed network like cloud.

c. At present, there are no studies which exploit the joint operations of all the available VMs in order to exploit the full potentials of VM.

Therefore, in spite of multiple research contribution in existing system there is no a single benchmarked model to be spoken of in this regard. The potential characteristics of VM were never fully exploited and were always seen as a single entity whereas a novel framework with chain of VM with some significant behavioural characteristics modelling may add new functionalities to minimize task completion time. Hence, there is a need of a novel model that performs autonomous decision making practices in the dynamic environment of cloud in order to facilitate resource allocation. The next section presents a novel framework that addresses the research problems brief in this section followed by algorithm discussion.

1.3. The Proposed Solution

The main purpose of the proposed system is to design a framework called FDMC of Framework for Decision Making in cloud environment for enhancing the efficiency in resource management. This work is a continuation of our prior techniques [25]. The entire implementation is focussed on exploiting the potential utilization of VM. Figure 1 highlights the proposed scheme of FDMC.

![Figure 1. Schematic Diagram of Proposed FDMC](image)

The proposed framework is essentially divided into two core components i.e. i) joint operation scheme and ii) decision making scheme. The joint operation scheme is mainly designed to exploit the full potentials of job processing by Virtual machine (VM) in order to ensure better synchronocity among the VMs which indirectly communicate with the data centers. This component primarily performs the task of resource allocation to the VM with the help of task allocator, host machine, and task request. It then interacts with modelling of virtual machine on the basis of its operational state of busy or idle. The significant step performed by Joint operation scheme is also to filter the VM based on its operational status in order to perform dynamic job swapping among the eligible VMs. The VMs are considered eligible if they have lower memory demands and more communication vectors. The next component is decision making modelling which is responsible for awarding the VM that participate in successful task scheduling process. However, in order to undertake decision, the proposed system do have dependencies on knowledge-based attribute e.g. list of eligible VM as well as capabilities of eligible VM basedon this prior status. The system than computes incentives using empirical means precisely for the VM under observation and then performs allocation of necessary amount of incentives to the VMs. One interesting fact of the proposed technique is that although incentives are allocated to all types of eligible VM, but it has its own selection process. If in case a VM takes up a job but could not process it completely, the incentive is completely withdrawn. However, if such VM
choose to jointly work with other VM to process the same task than task completion time will be minimized. Moreover, in this process large participation of the VM can be ensured in order to gain the incentive and hence the communication performance can be greatly increased without any inclusion of any external agents. Hence, the proposed decision making model automatically computes and allocates incentive based on the degree of task being completed which is consistently monitored and simultaneously the respective status of the VM is updated. The next section discusses about the algorithm implementation and result outcomes of proposed system.

2. ALGORITHM IMPLEMENTATION

The proposed algorithm is mainly responsible for allocation of incentive only for the system which encourages grouping or joint usage of the VM over cloud environment. The algorithm takes the input of J (Job Number), R (Request types), VM, (VM of i types), H (Number of Host Machine), S (Switches), Cv (Communication vector of host), memreq (memory required), which after processing gives the output of inc (incentive).

Algorithm for incentive allocation
Input: J, R, VM, H, S, C, memreq
Output: inc
Start
1. init J, R, VM, H, S, C, /i=4
2. def availability -> Cv > memreq
3. If availability=1
4. set status=1
5. else, set status=0
6. for i=1:H
7. for j=1:H
8. groupmat=Cv(i)+Cv(j)
9. end
10. end
11. for i=1:H-1
12. if η(i)<memreq
13. inc=(η(i)-memreq)*Cv(i)/η(i)
14. else if η(i)=memreq
15. inc=δ.Cv(i)/η(i) //δ → constant
16. else if η(i)>memreq
17. inc=δ-(η(i)-Cv(i)/η(i))*Cv(i)/η(i)//δ → constant
End

The algorithm defines its initial types of the jobs with respect to the number and uses a particular identification to recognize it. It also maintains a matrix of 4 different types of request (R) along with 4 different types of VM configuration in the form of matrix i.e. small, medium, large, and extra large. The system considers 400-1000 host machines with 4 types of VMs where the creation of VM is essentially carried out using job number and method deployed. The algorithm also formulates communication vector which consists of all the potential configurations of the existing VMs with a defined memory (Line-2). A host machine is only declared to be available if the communication vector (Cv) is found to be more than memory demands (memreq) (Line-2). It then iterates to check for the entire host H for its availability status (Line-3-Line-5). The next step of the algorithm is to filter only the available VMs and motivate for joint operation by allocation of incentives. All the joint operations of the VMs are carried out by adding the communication vectors (Line-6 to Line-10). After a list of host is filtered, the proposed algorithm proceeds towards incentive computation using the attribute of memory demands. The system check for the entire host if there is any node within the list of filtered VM which is found to have less memory demands (Line-12) and it computes the incentive on the empirical basis of η (feasible joint operation with ith host), memory demands (memreq), and communication vector (Cv) (Line-12 and Line-13). Similarly, it also checks for the condition if any filtered VMs has equivalent memory demands (Line-14) as well as for the condition if filtered VM is found to possess more memory availability (line-16). The algorithm has respective calculation of incentives for all the above three conditional cases, where the first condition (η (i) <memreq) represents busy state of VM, second condition (η (i) =memreq) represents ideal operation of condition of VM, while the third condition (η (i)>memreq) represents idle state of VM that also states its readiness to take up new incoming
After computing the incentive, the proposed system also computes the maximum permissible value of incentive for all the available host machines. The parameter $\delta$ is constant, which can be altered to any networking coefficient in future. Once the updating of the host machine is carried out, the system then check the condition specified in (Line-2) for performing job allocation in cloud environment. The entire phenomenon of job allocation is carried out considering the busy and idle state of VMs, where the focus is mainly to split the VMs with busy status to share its processing task with other VMs with ideal or nearly idle state. The algorithm therefore is able to process more number of incoming jobs and is quite independent of any load of the traffic over cloud environment. Hence, irrespective of any cardinality of the communication vector, the algorithm always explore the best capable VM and optimize both over and underutilize VM by its dynamic incentive allocation policy.

3. RESULT ANALYSIS

The proposed study was implemented over Matlab considering more than 500 numbers of virtual machines in the design principle. We also formulate three different types of Jobs i.e. Job-1, Job-2, and Job-3 which is responsible for executing 4, 8, and 12 job request to the proposed algorithm. The study also considers 4 different types of VM configuration based on its cores of storage (small storage of 100 GB, medium storage of 200 GB, large storage of 400 GB, and extra large storage of 1000 GB). The study of the proposed system is also assessed using standard performance parameters e.g. task allocation time, amount of core wasted, amount of storage wasted, and communication cost. For an effective benchmarking, the study outcome was also compared with the recent and similar kind of work being carried out by Wang et al [26], Nurmi et al [27], Shu et al [28], and Ren et al [29]. The discussion of the comparative performance analysis is as follows:

![Figure 2. Comparative Analysis of Task Allocation Time](image)

Figure 2 shows that proposed FDMC has accomplished the minimal task allocation time for all the three types of jobs. The prime reason behind this is the approach of Wang et al [26] uses round robin over the rack servers and not on the VM which consumes time. Nurmi et al [27], too, perform similar task but over the host machine leading to more time consumption during task allocation with increase of job types. The work carried out by Shu et al [28] uses multi-objective optimization model and therefore, it corresponds with similar performance outcome for Job-1 with proposed FDMC. However, with increasing job types (2 and 3), it increases its task allocation time owing to inclusion of iterative steps. However, the performance of Ren et al [29] is better than Wang et al [26], Nurmi et al [27], and Shu et al [28] with respect to job-1 and job-2 types. However, due to extra computational steps for cost estimation, it couldn’t further optimize for Job-3.

Figure 3 discusses about the comparative performance analysis for core wasted during the process of resource allocation over cloud environment. The outcome shows proposed FDMC is found with higher degree of utilization of cores resulting in superior outcomes in comparison to others. The problems with other existing schemes are usage of cores to resposit temporary buffers with indefinite time of garbage disposal of memory (as it depends upon incoming task only). The another problem with existing mechanism is that time consumption to perform decision of resource allocation is found directly proportional to incoming jobs that
results in over utilizing of the core by allocating all the current job threads. It thereby results in core wastage. We perform further confirmation of this by assessing extent of storage wastage too.

Figure 4 shows the amount of storage being wasted. The outcome shows that proposed FDMC has lower wastage of storage as compared to other existing systems. This is because majority of the existing systems are dependent on larger queues as they will normally take more task allocation time and such maintainace of longer queues are capable of handling more jobs but at the cost of storage.

![Figure 3. Comparative Analysis of Core Wasted](image1)

![Figure 4. Comparative Analysis of Storage Wasted](image2)

Figure 5 shows comparative analysis of communication cost

![Figure 5. Comparative Analysis of Communication Cost](image3)

Basically, we define communication cost as the total amount of resources being spent in order to accomplish a task of job processing over a cloud environment. The outcome shows that proposed FDMC is better than Wang et al [26], Nurmi et al [27], and Shu et al [28], but not better than Ren et al [29]. This is because the approach of Ren et al [29] is based on optimization principle of Lyapunov technique that minimizes the queue length on the basis of cost parameter. Although, this is one significant advantage of Ren et al [29] but this advantage comes at the cost of over memory utilization. Hence, in this regard, the proposed FDMC can be stated as the better alternative of work carried out by Ren et al [29]. Jennings and Stadler [30] Surveyed study into resource-management (RM) for cloud atmospheres. As working on this study, they were astonished by the quantity of current outcomes that they establish, and the document produces therefore

_FDMC: Framework for Decision Making in Cloud for Efficient Resource Management (Alexander Ngenzi)_
superior than estimated. The paper has proposed by Younge et al [31] a novel structure is illustrated that offers a competent green improvements inside a scalable Cloud computing design. Utilizing power-aware forecast methods, variable resource organization, live relocation, and a minimum virtual machine propose, generally scheme efficiency will be greatly enhanced in a data center stands Cloud with minimum presentation transparency.

4. CONCLUSION

The concept of cloud computing has gained a faster pace in the commercial enterprise market encapsulating more number of users registering for services. With inclusion of mobile network, the generation of data and job request has exponentially increased which are required to be processed at faster speed. Although, there is a word pervasiveness connected with cloud architecture due to its process of virtualization, but there are many challenging factor which drags back the success factor of cloud computing. This paper raises the problem of resource allocation which is mandatory to be implemented on the process of virtualization over cloud environment. After reviewing the existing mechanism of addressing the problems of resource management in cloud, it was found that existing techniques are never benchmarked, they don't exploit the full potential of VM, and normally adopts deterministic approaches of resource allocation which is directly associated with traffic intensity. However, this paper has laid down a rationale of decision-making process which rewards the VM with incentive if the VMs perform jointly with each other. A simple modelling is discussed in this regard along with the algorithm implementation considering the dynamic behaviour of the users online. Different from existing techniques, the proposed algorithm formulation is more non-deterministic type that gives better edge to the resource attribute calculations. The study outcome was compared with some of the standard and significant research contribution in recent times to find proposed system offer better task scheduling process with respect to the performance parameters considered for the study. The entire implementation discussed in this paper significant assists in server consolidation process and therefore, our future work will be on the direction of enhancing the present work to offer better resource management scheme.

REFERENCES
[1] Z. Mahmood, Cloud Computing: Challenges, Limitations and R&D Solutions, Springer, 2014.
[2] J. Bond, The Enterprise Cloud: Best Practices for Transforming Legacy IT, O'Reilly Media, Inc., 2015.
[3] B. Furht, A. Escalante, Handbook of Cloud Computing, Springer Science & Business Media, 2010.
[4] L. Wang, R. Ranjan, J. Chen, B. Benatallah, Cloud Computing: Methodology, Systems, and Applications, CRC Press, 2011.
[5] K. Chandrasekaran, Essentials of Cloud Computing, CRC Press, 2014.
[6] T. Wenhong, Z. Yong, Optimized Cloud Resource Management and Scheduling: Theories and Practices, Morgan Kaufmann, 2014.
[7] N. Sabharwal, P. Wali, Cloud Capacity Management: Capacity Management, Apress, 2013.
[8] S.U. Khan, A.Y. Zomaya, Handbook on Data Centers, Springer, 2015.
[9] Aljawarneh, Shadi, Cloud Computing Advancements in Desi...
[19] S. K. Panda, S. Nag and P. K. Jana, “A Smoothing Based Task Scheduling Algorithm for Heterogeneous Multi-Cloud Environment”, *IEEE-International Conference on Parallel, Distributed and Grid Computing*, 2014.

[20] S. Caton, C. Haas, K. Chard, K. Bubendorfer, O. Rana, “A Social Compute Cloud: Allocating and Sharing Infrastructure Resources via Social Networks”, *IEEE Transactions on Services Computing*, 2014.

[21] R. Yu, Y. Zhang, S. Gjessing, “Toward Cloud-based Vehicular Networks with Efficient Resource Management”, *arXiv*, 2013.

[22] A. Meera, S. Swamynathan, “Agent based Resource Monitoring system in IaaS Cloud Environment”, *Elsevier-Science Direct, International Conference on Computational Intelligence: Modeling Techniques and Applications, Procedia Technology*, vol.10, pp.200 – 207, 2013.

[23] H. Xu, B. Li, “Anchor: A Versatile and Efficient Framework for Resource Management in the Cloud”, *IEEE Transactions on Parallel and Distributed Systems*, 2013.

[24] R. Yu, X. Huang, J. Kang, J. Ding, S. Maharanj, “Cooperative Resource Management in Cloud-Enabled Vehicular Networks”, *IEEE Transactions On Industrial Electronics*, 2015.

[25] A. Ngenzi, R. Selvarani, S.R. Nair, “Dynamic Resource Management in Cloud Data Centers for Server Consolidation”, *arXiv*, 2015.

[26] W. Wang, G. Casale, “Evaluating Weighted Round Robin Load Balancing for Cloud Web Services”, *IEEE-International Symposium on Symbolic and Numeric Algorithms for Scientific Computing*, 2014

[27] D. Numni, R. Wolski, C. Grzegoreczyk, “The Eucalyptus Open-source Cloud-computing System”, *IEEE/ACM International Symposium on Cluster Computing and the Grid*, 2009.

[28] W. Shu, W. Wang, and Y. Wang, “A novel energy-efficient resource allocation algorithm based on immune clonal optimization for green cloud computing”, *Springer-EURASIP Journal on Wireless Communications and Networking*, vol.64, 2014.

[29] S. Ren and M. Schaar, “Dynamic Scheduling and Pricing in Wireless Cloud Computing”, *IEEE Transactions on Mobile Computing*, 2013.

[30] B. Jennings, and R. Stadler, “Resource management in clouds: Survey and research challenges”, *Journal of Network and Systems Management*, Vol. 23, No. 3, pp. 567-619, 2015.

[31] A.J. Younge, V. Laszewski, G. Wang, L.L Alarcon & W. Carithers, “Efficient resource management for cloud computing environments”, *In Green Computing Conference, International*, pp. 357-364, 2010.

**BIOGRAPHIES OF AUTHORS**

**Alexander Ngenzi** received his Bachelor of Science in Electronics and Telecommunication Engineering and Master of Science in Operational Communications from former Kigali Institute of Science and Technology (KIST) which is now the College of Science and Technology (CST), University of Rwanda (UR) in 2007 and 2011 respectively. Currently, He is doing his PhD in the department of Computer Science and Engineering at Jain University since 2013. He is also a Lecturer in the department of Computer and Software Engineering (CSE), University of Rwanda since 2011. Till date, he has published three papers in both Scopus and Thomson Reuters indexing journals. He presented and published one conference paper in ANAED indexing Journals. His area of specialization is Computer Science Engineering. His research interests include; Cloud Computing, Server Consolidation, Virtualization, Computer networks and Telecommunication Networks. His area of specialization is Computer Science and Engineering.

**Professor R. Selvarani**’s areas of specialization are Computer Science and Engineering. She was awarded a doctoral degree by the Jawaharlal Nehru Technological University, Hyderabad. She was given the “Best Teacher” award a couple of times. She served as Professor and Head of Dept. of CSE for more than a decade in various well reputed engineering institutions in Bangalore and also as Dean Research for about four years in a leading engineering institution in Bangalore. She has published her research work in several peer reviewed journals such as Elsevier, Info comp, Inder science, ACM SIGSOFT, IEEE Conf. etc. and supervised doctoral scholars. Her several book chapters and journal publications are indexed in major indexing systems like Scopus & Thomas Reuters. She has a patent in software architecture and design domain. Her publication in Elsevier Journal was selected in terms of having “the best content” in the area of Information Technology for the year 2012 by VERTICAL NEWS, USA. She is carrying out collaborative research with Leeds Metropolitan University, UK. Prof. Selvarani is listed in Who’s Who for science and Technology, USA Prof. Selvarani’s current research interest includes Machine Learning, Internet of Things Software design quality estimation, Service Oriented Cloud Applications; Software Safety critical Systems, QoS in Distributed Networks.
Professor Suchithra R received her B.COM, MCA and Ph.D degrees in Manonmaniam Sundaranar University, Tirunelveli respectively. She is a Professor and Head of MS (IT) Department, Jain University Bangalore, August 2010 till date. She was a Lecturer, Dayananda Sagar Institutions, Aug 2009 – 2010, Lecturer Dept. of MCA, Oxford Institutions -Sep 2006 to Aug 2009, Lecturer Dept. of MCA, AMC Engineering College -Sep 2005 to Aug 2006, Lecturer Dept. of Computer Science and Engineering, PSN Engineering College -Sep 2003 to Aug 2005 respectively. She published more than 15 papers in both Scopus and Thomson Reuters indexing Journals. Her research interests include; Health Care and Cloud Computing, Green Computing, User Interface Design, Precision Agriculture and Cloud, Cloud Security, Social Media Data and Sentimental Analysis, Big Data and Internet of Things. Her area of specialization is Computer Science and Information Technology.