

# Consumer Context based Resource Allocation in Cloud Computing

Kamini Bharti

Student

Department of Computer Science, Guru Nanak Dev University, India

Kamaljit Kaur

Assistant Professor

Department of Computer Science, Guru Nanak Dev University, India

## ABSTRACT

Cloud Computing has recently emerged as a compelling paradigm for managing and delivering services over the internet. It is now considered as one of the rapidly developing platforms in the various fields related to computer science in one way or another. It is provided in a pay-per-use manner as per the demand of the consumer. Resource allocation in cloud computing is performed with the objective of minimizing the costs related with the management of resources at the provider side. It also involves fulfillment of customer demands and application requirements. Numerous techniques for resource allocation have been proposed and implemented so far in the world of cloud computing but none of which has considered introducing a parameter from the consumer context. Resource allocation being one of the most critical factors which defines the overall efficiency of the cloud, usually involves an attribute on which the whole process of the allocation of resources is based upon. In this report, a policy has been proposed & implemented, keeping in mind the importance of consumer context in cloud. The importance of context in the growth and success of a business has also been discussed.

## General Terms

Cloud Computing, Policy Evaluation, Simulation, Quality of Service

## Keywords

Cloud Computing, Resource Allocation, Consumer Context, Consumer Relations, CloudSim

## 1. INTRODUCTION

Cloud computing can be defined as a new way of computing dynamically scalable and virtualized resources which are provided as a service over the internet. It is a model for enabling on demand network access to a shared pool of resources like servers, applications, storage. Various services are provided that can be provisioned and released with minimal management effort [1]. It is a recent trend in IT that moves computing and data away from desktop into large data centers. The applications are delivered as services over the Internet [2]. Cloud computing can also be defined as a type of parallel and distributed system containing of a collection of inter-connected and virtualized computers that are dynamically provisioned, and offered as one or more unified computing resources based on service-level agreements established through negotiation between the service provider and consumers [7]. Cloud computing based on the offer of services, we get three different types of services: Infrastructure as a Service (IaaS) is one of the “Everything as a Service” trends. Platform as a Service (PaaS) cloud systems provide a software execution environment that application services can run on. Software-as-a-Service (SaaS) is based on

licensing software use on demand, which is already installed and running on a cloud platform.

## 2. RESOURCE ALLOCATION IN CLOUD COMPUTING

Resource Allocation (RA) is the process of assigning available resources to the needed cloud applications over the internet. Resource allocation is one of the challenging issues in cloud computing, where the scarce resources are distributed. Resource provisioning solves that problem by allowing the service providers to manage the resources for each individual module. Resource allocation is used to assign the available resources in an economic way. It is part of resource management. The benefit of resource allocation is that user neither has to install software nor hardware to access the applications, to develop the application and to host the application over the internet. There is no limitation of place and medium. User can reach our applications and data anywhere in the world, on any system. Cloud computing offers a cleverer computing archetype in which resources is leased on-demand. The key goals of the cloud resource suppliers and consumers are to allocate the cloud resources strongly and attain the maximum financial profit. Resource allocation is one of the important issues in cloud computing, where rare resources are distributed. From a consumer's viewpoint, resource allocation relates to how products and services are distributed in the midst of users. Proficient resource allocation results in a more industrious economy. Resource allocation to cloud users is a multilayered process due to the complexity of finest allocation of resources i.e., proficient allocation with restricted resources and highest profit [8].

In cloud computing, there are two different modes of renting the computing capacities from a cloud provider.

- Advance Reservation (AR): Resources are reserved in advance. They should be available at a specific time
- Immediate: When a client submits a request, either the resources are provisioned immediately, or the request is rejected, based on the resource availabilities [2].

## 3. CONSUMER CONTEXT FOR RESOURCE ALLOCATION IN CLOUD COMPUTING

Context is all about the whole situation relevant to an application and its set of users. All the important aspects of different situations cannot be enumerated, as this will change from situation to situation. Context is any information that can be used to characterize the condition of an entity such as a person, place, or object that is considered relevant to the interaction between a user and an application, including the

user and applications themselves. If a piece of information can be used to characterize the state of a participant in an interaction, then that information is context [12].

In our case, the context in terms of the user data was considered i.e. the consumer context. Currently various parameters such as the most requested lease in the user history & frequency in leases are taken into consideration to prepare offers or counter offers. The users are ranked on the basis of their dealings with the service providers. As the proposed system evolves, various aspects can be included in relation to the context whatever seems feasible in making the system more efficient.

### **3.1 The Importance of Using Consumer Context**

In context-aware system, context is utilized to provide relevant information and/or services to the user, where relevancy depends on the user's task.

There are three categories of features that a context-aware application can support:

- Presentation of information and services to a user
- Automatic execution of a service for a user
- Tagging of context to information to support later retrieval [13]

The job of being an IaaS provider can be seen as business in a major way as opposed to the techniques proposed so far, none of which take in consideration the consumer context which could also be a factor for its expansion or betterment. In the rules of business, analyzing the user data which can be historical or real-time actually helps in making the right decisions for making different policies of resource allocation & management. This is where context comes in play. It bridges the gap of human-computer interaction.

### **3.2 The Types of Leases Considered**

Two types of leases considered are:

1) Immediate Lease: It is the case in which a consumer might request for immediate allocation of resources. Through the consumer context the user can be offered the existing early slot as an offer or a counter offer can be made based on the parameters like a minute delay in the start time or reduced resources which might be available so as to absorb the user.

2) Advanced Reservation: There are periods which might be weekly, monthly etc. in which an organization may have a system overhaul or peak load times due to maintenance so they can reserve additional resources in advance. Again there can be offers & counter offers as per the context.

## **4. RELATED WORK**

Resource allocation is one of the major challenge in cloud computing. Proposed an adaptive resource allocation algorithm for the cloud system with preemptable tasks which addresses these challenges in the cloud computing. These algorithms adjust the resource allocation adaptively based on the updated of the actual task executions [2]. An adaptive resource allocation model that allocates the consumer's job to an appropriate data center is proposed. The method to adaptively find a proper data center is based on two

evaluations: 1) the geographical distance between a consumer and data centers 2) the workload of each data center. The proposed model allocated the consumer requests to the data center that is closed to consumer location [3]. CloudSim: an extensible simulation toolkit that permits modeling and simulation of Cloud computing systems and application provisioning environments [5]. Advance reservation algorithms in Haizea are extended to provide negotiation based allocation. Proposed algorithm to generate counter offers provides counter offers considering constraints flexibilities to maximize the chances of their acceptance. Using ranking algorithm, consumers will get suitable sorted offers according to their needs. These algorithms will not handle the situations when system has multiple requests for a single slot [4]. Backfilling and swapping concepts are applied for rescheduling the accommodated leases to make space for newly arrived leases. The information about leases which are to be rescheduled, the order of deciding of rescheduling is used by swapping. The idle resources, which cannot be filled up by consecutive lease is filled by Backfilling [6]. An economy based leasing algorithm is developed and integrated with Haizea. This economy based algorithm looks after both of the parties i.e. customer and service provider. It uses the concept of optimization techniques to optimize the costs. If an appropriate resource is not found, then the algorithm uses negotiation on budget and resource demand that increases the number of accepted lease. The proposed algorithm schedules lease based on its budget and cost of resource along with the capabilities [10]. The composition of different layers in the cloud such as IaaS and SaaS and its joint optimization for efficient resource allocation is proposed. The efficient resource allocation optimization problem is conducted by sub-problems. Through an iterative algorithm the proposed cloud resource allocation optimization algorithm is achieved [9]. This provides overview on the Cloud's anatomy, definition, characteristic, affects, architecture, and core technology. This clearly classifies the Cloud's deployment and service models, providing a full description of the Cloud services vendors and also addresses the customer-related aspects such as the Service Level Agreement, and security issues. Finally, it covers detailed comparisons between the Cloud Computing paradigm and other existing ones in addition to its significant challenges [11].

So far the algorithms or techniques have been based upon the cost involved, job duration, order of submission etc. i.e. only the technical aspects but none of the techniques involves the consideration of the consumer relations or the consumer context which is quite essential. Since, cloud computing is ultimately a business in itself and the level of commitment and trust a customer reflects for the service provider, should be valued as well, apart from the technical aspects. In our case, the context in terms of the user data was considered i.e. the consumer context. Currently various parameters such as the most requested lease in the user history & frequency in leases are taken into consideration to prepare offers or counter offers. The users are ranked on the basis of their dealings with the service providers. As the proposed system evolves, various aspects can be included in relation to the context whatever seems feasible in making the system more efficient & provide the service provider some metrics to evaluate his service based on the consumer relations & business rules.

## 5. PROPOSED WORK

The existing algorithms or techniques for resource allocation have only involved the technical aspects related to cloud computing but none has taken the view of cloud computing as a rapidly growing business in the technical world & taken into account the consumer context. In a business, the consumer relations play an important role and hence using the consumer context as an important attribute for the resource allocation could actually play a great role in making the cloud computing business a great success, but it has not yet been used as a valued attribute for resource allocation. In order to inculcate this aspect, a new allocation technique is to be formulated which uses an attribute of the consumer context as an important factor for determining which allocation policy is to be followed. In this way, users can be ranked on the basis of their loyalty to the company & can be provided more privileges, the existing user data can be utilized so as to identify the user interests in the different company policies & packages and hence it can be involved in the allocation policy of the service provider so as to maintain a level of customer satisfaction along with other important aspects of resource allocation.

## 6. SIMULATION & RESULTS

The performance measurement has been gathered by using the CloudSim framework [5].

The various components to be considered are:

### 6.1 The Datacenters

which consists of the actual hardware which will host the Virtual Machines.

*Assumption:* There are 4 Datacenters. Each Datacenter has 50 hosts, which means there are total 200 hosts.

- *Type 1 host* (25 units)
  - 4 PEs, each with MIPS = 5000
  - RAM = 16384 (host memory in MB)
  - Storage = 1000000 (host storage)
  - Bandwidth = 10000
- *Type 2 host* (25 units)
  - 2 PEs, each with MIPS = 5000
  - RAM = 16384 (host memory in MB)
  - Storage = 1000000 (host storage)
  - Bandwidth = 10000

### 6.2 Virtual Machines

Execute the lease which is allotted. Total 850 VMs have been considered.

*Assumption:*

- Size = 10000 (image size in MB)
- RAM = 512 (VM memory in MB)
- MIPS = 500, 1000, 2000
- Bandwidth = 1000
- pesNumber = 2 (number of CPUs)
- VMM = Xen (VMM name)

Priorities are assigned according to the criteria of

- Priority 1 if MIPS > 2000
- Priority 2 if MIPS > 1000
- Priority 3 if MIPS < 1000

### 6.3 Leases

Which are the request for resources by the consumers in the form of tasks to be executed through the VMs, these requests are referred to as cloudlets in CloudSim.

*Assumption:*

There are two types of leases to work with i.e. Advanced Reservation & Immediate. Three types of packages have been defined for the simulation purposes. Here Length resembles the RAM required for execution, FileSize/Output size denotes the hard disk space required & PesNumber denotes the number of CPU(s) required.

- Package v1:
  - Length = 40000
  - FileSize = 300
  - Outputsize = 300
- Package v2:
  - Length = 35000
  - FileSize = 200
  - Outputsize = 200
  - PesNumber = 2
- Package v3:
  - Length = 30000
  - FileSize = 100
  - Outputsize = 100
  - PesNumber = 2

### 6.4 Brokers

In the form of Web Services are responsible for

- a) Accepting the leases from the user
- b) Fetching the leases from the database
- c) Creating a list of the cloudlets yet to be executed
- d) Submitting the list of leases to the respective VMs associated with that Web Service.

*Assumption:*

- Three Web Services have been considered in the simulation.
- ARWebService1: For handling the previously present AR leases in the DB.
- ARWebService2: For handling the new AR requests from the consumer.
- iWebService: For handling the requests of immediate leases.

### 6.5 Consumer

Which enter the leases as per their requirement i.e the lease for advanced reservation or an immediate lease.

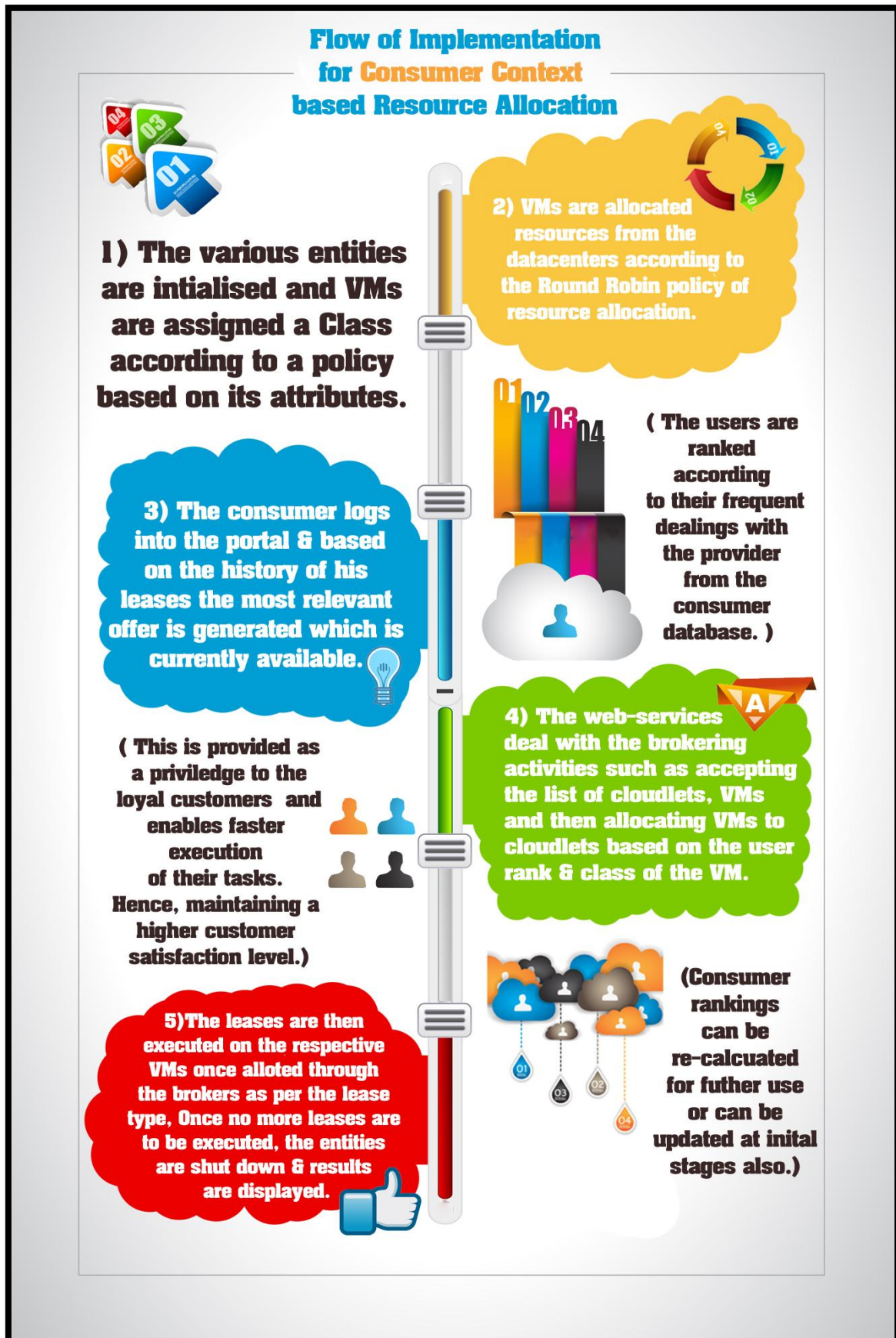


Fig 1.Flow of Implementation for Consumer Context based Resource Allocation

## 7. SIMULATION SCREENSHOTS

A user interface was added to the CloudSim simulator for enabling user interactivity.

### 7.1 Customer Registration Page:

Fig 2.Registration Page for a New user

This screenshot represents the login screen for a new user. Various fields related to the customers are accepted & added to the customer database of the service provider.

### 7.2 User ID Generation:

Fig 3.Generation of User ID after successful registration

After the successful registration of the customer a User ID is generated. It is to be used by the customer to login into the system along with the password given at the time of registration to enter the leases.

### 7.3 Login Page:

Fig 4.Customer Login Page

This screenshot represents the login page for the customer. The option to register for the new users has also been provided. If the User ID is not found/ password is incorrect, the appropriate message is displayed to the user.

### 7.4 Generated Offer Page for the Customer:

Fig 5.Generated offer page for the Customer

As the customer logs in to the system, the appropriate offer/package available for the customer is displayed as per the consumer context.

### 7.5 Other Available Offers/Immediate Lease page:

Fig 6.Other available offer/Immediate Lease page for the Customer

The alternate offers which are available are displayed on this page for the consumer not interested in the generated offer. The option to enter a customized immediate lease has also been provided.

## 8. RESULTS

### 8.1 The Comparison of The Execution Times of Leases through the Proposed Policy & the Default Policy of Resource Allocation in the Simulator i.e. FCFS VM to Host Allocation.

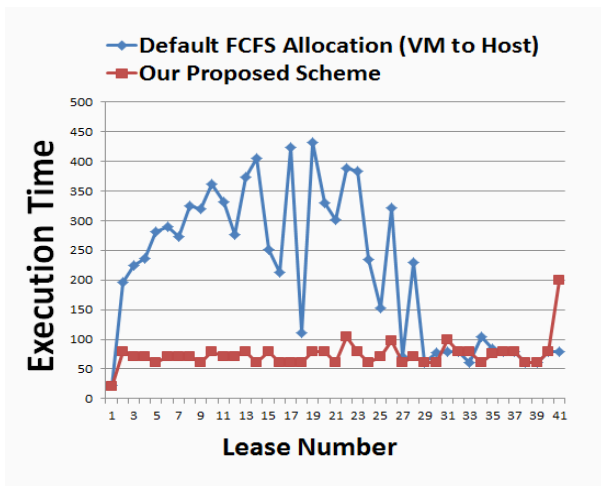


Fig 7. Comparison graph of Execution time of Leases

It can be seen that there is a good amount of improvement in the execution time of the leases as per the policy implemented. There is no overall overhead/spike seen by the inclusion of the consumer context parameter. Hence, the consumer context parameters can be included along with the existing policies with slight variation in results.

### 8.2 A Representation of the Consumer Satisfaction Levels according to the Consumer IDs

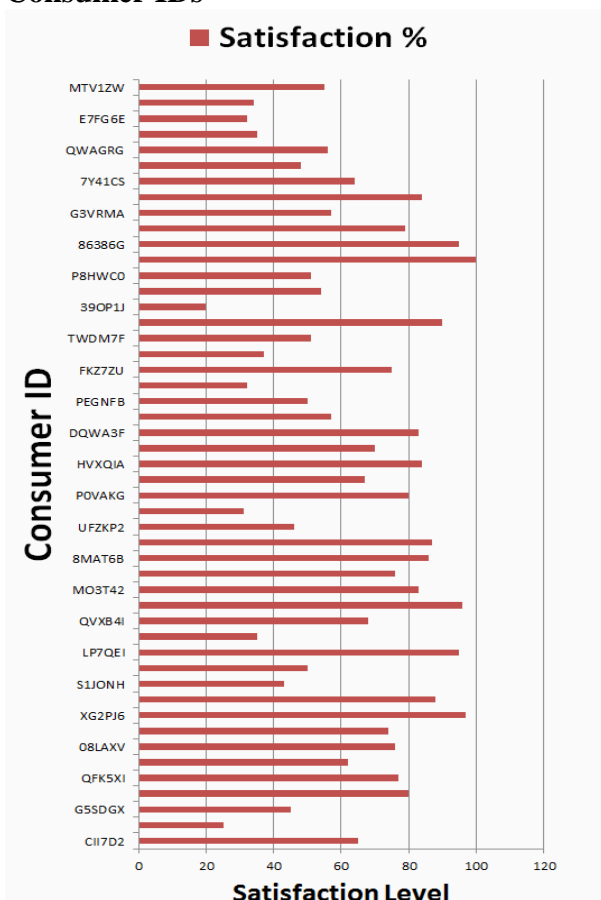


Fig 8. Consumer ID V/s Consumer Satisfaction Level

This graph represents the consumer satisfaction levels according to the various Consumer IDs. This is based on the satisfaction metric of the consumer which enables the service provider to evaluate his services and the consumer response. It can be used by the service provider to enquire what the consumers want or which new package to create so as to increase the consumer satisfaction.

### 8.3 Overall Execution Time according to the Policy of VMs Allocated to Web Services:

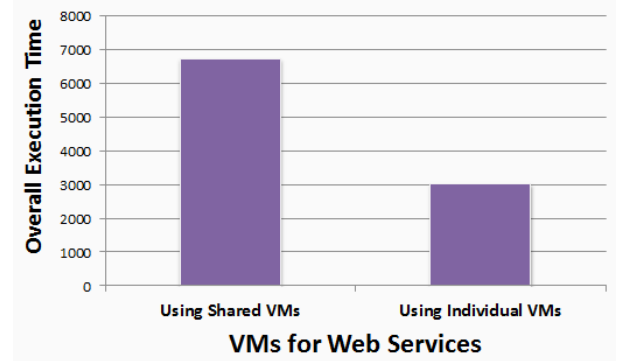


Fig 9. Overall Execution Time V/s Types of VMs for Web Services

This graph represents the increase in overall execution time if instead of individual/independent VMs, Web Services use shared VMs. On the other hand this also increases the possibility of failure of leases or low performance because of the ambiguity between the web services due to shared VMs. Hence, individual VMs have been used.

### 8.4 Representation of the Successful Lease Execution Probability according to the Consumer Classes.

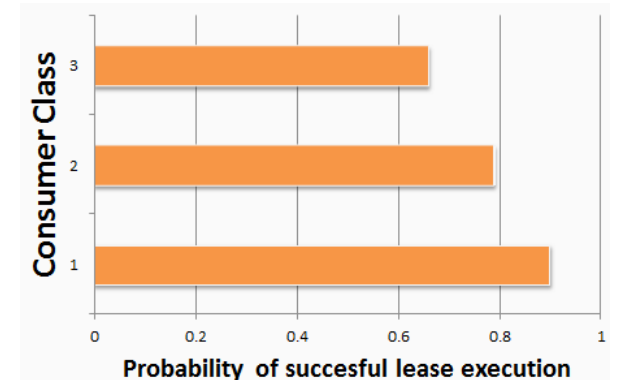


Fig 10. Consumer Class V/s Successful Lease Execution Probability

The Consumer Classes may be defined according to the policy of the service provider. For ex. Top 25% of the consumer ranks might be considered as Consumer Class 1. This graph represents the successful lease execution probability as per the Consumer Class to which the consumer may belong. Since higher class consumers are given first priority over VMs with faster processing power, their successful lease execution probability is more and decreases as the classes are descended.

## 9. CONCLUSION & FUTURE WORK

Through the proposed policy of resource allocation a new aspect of involving the consumer context parameter(s) for resource allocation has been implemented along with FCFS &

Round Robin allocation policies. The service provider may implement their own policies along with considering the consumer context parameters. The results are much improved from the default allocation policy implemented in the simulator i.e. FCFS policy for VM to host allocation. This policy also enables the maintenance of a consumer satisfaction index through which the company can improve their policies as per the response given by the consumers. This policy of including the parameter(s) related to the consumer context, enables new avenues for the provider to evaluate his/her service quality and develop them accordingly.

Through this implementation an interface has been also added to the CloudSim simulator which could be further developed in future to bring in more interactivity and may allow the implementation of the user created scenario in a simplified manner.

## **10. ACKNOWLEDGEMENT**

At the very outset I wish to extend my heartfelt thanks to the almighty GOD for his blessing along with gratitude in the completion of this paper. I wish to express my sincere thanks to Dr.Gurvinder Singh, Head of the Department (CSE), Guru Nanak Dev University, for providing me the opportunity to conduct the research paper. I wish to express my profound and deep sense of gratitude to Ms.Kamaljit Kaur, Assistant Professor, Dept. of CSE for sparing her valuable time to extend help in every step of my dissertation. I also express my sincere thanks to Mrs.Veena, Dept. of CSE, for her continuous encouragement & support through the research lab. I am obliged to all those who have contributed directly or indirectly in the completion of the work.

I take this opportunity to express my regards and obligation to my parents whose support and encouragement I can never forget in my life.

## **11. REFERENCES**

- [1] Munich, Gerald kaefler, Cloud computing architecture, IEEE Spectrum, February 2009.
- [2] Jiayin Li; Meikang Qiu; Jian-Wei Niu; Yu Chen; Zhong Ming, "Adaptive resource allocation for preemptable jobs in cloud systems," Intelligent Systems Design and Applications (ISDA), 10th International Conference on , vol., no., pp.31,36, Nov. 29 2010-Dec. 1 2010
- [3] Gihun Jung; Kwang-Mong Sim, "Agent-Based Adaptive Resource Allocation on the Cloud Computing Environment," Parallel Processing Workshops (ICPPW), 2011 40th International Conference on , vol., no., pp.345,351, 13-16 Sept. 2011
- [4] Janki Akhiani, Sanjay Chuadhary and Gaurav Somani, "Negotiation for Resource Allocation in IaaS Cloud", Bangalore Compute Conf., pp 15 ACM, 2011
- [5] Rodrigo N. Calheiros, Rajiv Ranjan, Anton Beloglazov, Cesar A. F. De Rose and Rajkumar Buyyal, "CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms", *Softw. Pract. Exper.*, Vol. 41, pg 23–50, 2011
- [6] Amit Nathani, Sanjay Chaudhary, Gaurav Somani, "Policy based resource allocation in IaaS cloud", *Future Generation Computer Systems*, Volume 28, Issue 1, Pages 94-103, January 2012
- [7] Dhaval Limbani, Bhavesh Oza, "A Proposed Service Broker Policy for Data Center Selection in Cloud Environment with Implementation", *IJCTA*, Vol 3 (3), pg 1082-1087, MAY-JUNE 2012.
- [8] N.R.Ram Mohan, E.Baburaj, "Resource Allocation Techniques in Cloud computing Research Challenges for Applications", *Computational Intelligence and Communication Networks*, 2012
- [9] Chunlin Li • Layuan Li, "Efficient resource allocation for optimizing objectives of cloud users, IaaS provider and SaaS provider in cloud environment", *Journal of Supercomputing* Vol 65, pp. 866–885, Springer, 2013
- [10] Hemant Kumar Mehta, Eshan Gupta, "Economy Based Resource Allocation in IaaS Cloud", *International Journal of Cloud Applications and Computing*, 3(2), 1-11 April-June 2013
- [11] Ahmed Shawish and Maria Salama, "Cloud Computing: Paradigms and Technologies", *Inter-cooperative Collective Intelligence: Techniques and Applications*, Studies in Computational Intelligence, pp. 495, Springer, 2014
- [12] Forbes, "How Cloud Computing Is Accelerating Context-Aware Coupons, Offers and Promotions"
- [13] TCS big data study sales, "Big Data Benefits and Challenges"