Auction based Resource Allocation Strategy for Infrastructure as a Service

Thesis submitted in partial fulfillment of the requirements for the award of degree of

Master of Engineering

in

Information Security

Submitted by

Anita Kumari

(801433003)

Under the supervision of

Dr. Sushma Jain

Assistant Professor

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT

THAPAR UNIVERSITY

PATIALA – 147004

June 2016

Certificate

I hereby certify that the work which is being presented in the thesis entitled, "*Auction Based Resource Allocation Strategy for Infrastructure as a Service*", in partial fulfillment of the requirements for the award of degree of Master of Engineering in *Information Security* submitted in Computer Science and Engineering Department of Thapar University, Patiala, is an authentic record of my own work carried out under the supervision of *Dr. Sushma Jain* refers other researcher's work which are duly listed in the reference section.

The matter presented in the thesis has not been submitted for award of any other degree of this or any other University.

Anita

(Anita Kumari)

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.

(Dr. Sushma Jain)
Assistant Professor,
CSED

Countersigned by

Deepak Garg
13/6/16

(Dr. Deepak Garg)
Head
Computer Science and Engineering Department
Thapar University
Patiala

(Dr. S. S. Bhatia)
Dean (Academic Affairs)
Thapar University
Patiala

Acknowledgement

During the course of this thesis I have been lucky to be blessed by a lot of kind people. Let me take a moment to thank each one of them.

To begin with, I am thankful to God for his blessings and for consistently showing me the right direction.

I wish to express my sincerest gratitude towards the guidance and help that I have received from my supervisor Dr. Sushma Jain. I shall ever remain indebted for her consistent support and encouragement. She provided me help whenever needed, and also arranged for me the resources required to complete this thesis report on time.

I am also thankful to Dr. Deepak Garg, Head, Computer Science and Engineering Department, Thapar University for his help and cooperation. Along with that I express my gratitude to all the staff and faculty members of Computer Science and Engineering Department, Thapar University for providing me with the facilities required for this thesis.

I would like to thank all my friends especially Ritu for their support. Also I want to express my appreciation to every person who contributed with either inspirational or actual work to this thesis.

Last but not the least I am very grateful to all my family members for their inspiration and moral support which kept me motivated to purse my studies.

Anita

Anita Kumari

Abstract

The cloud providers provide various computational resources like CPU power, memory, and bandwidth to the users through Virtual Machines (VM). Various VMs differ in their resource capacity. Each cloud provider has a set of VMs which are available to users at different prices. Providers have a limited set of resources. Therefore, it is necessary to efficiently allocate available resources to the rational users and minimize the cost of resources as to gain maximum profit from the available resources. Allocation of Resource in cloud computing is one of the most challenging issues in resource management for both cloud and grid computing. In this thesis, a Vickrey-Clarke-Groves (VCG) auction based approach is used to deal with the problem of resource allocation at Infrastructure as a Service (IaaS) layer of cloud among rational users and to incentivize selfish providers. The proposed mechanism uses a reverse auction, which efficiently allocates resources, calculates payment, and also achieves social benefit among the users.

The mechanism tried to solve resource allocation problem with payment calculation. The proposed mechanism fulfills properties of incentive compatibility, individual rationality and maximizes social welfare. The main target of the service providers is to maximize the utilization of their resources, maximizing the profit through minimizing the cost of the resources, achieve a social benefit, and maximizing the efficiency of the whole process. The proposed approach has been simulated using CloudSim and NetBeans IDE. Experimental analysis has proved the effectiveness of the system. This model is compared with fixed price model and has given a better performance in terms of pricing and utility.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

In past years, advancement in computational power, memory space, and networking infrastructure have permitted to share, generate and process increasing amounts of data and information in different ways. Computing technology had developed and introduced several new applications that are used by designers for achieving their business objectives. These applications have leaded us to new demands for powerful computing infrastructure.

1.1 Cloud Computing

All of us in this universe need a significant change of their working area. As the time passes rapidly, within seven decades from the creation of the very first digital computer, [1] people are living in the age of the Internet, and one cannot predict one's life without Internet. As the network bandwidth is increasing, the internet is becoming a necessary utility for day to day work. The internet offers its consumers, many conveniences (services) such as Emails, News provider, Banking Solutions, Infrastructure provider, Social Networking, Stock exchange, Entertainment Provider, and E-Business etc. By the presence of this emerging technology different researcher perceive cloud computing as one of the new areas of service computing environment. Cloud computing found its origins from cluster as well as grid computing. All these are a promising standard to provide information computing as a service to end user. But the job of these service providers is not just to provide some features and facilities to the consumers, but also provide these services in an efficient manner. The main motive behind these computing standards is to provide computing services to its consumers as an assured Quality of Service (QoS). The way of computing also modifies with the varying technology, which gives indications to the development of service-oriented computing.

To realize cloud computing there is a need to associate it with extensively used computing terminology such as Cluster Computing and Grid Computing. A huge number of researchers explain Clusters and Grid Computing in dissimilar ways. According to [2], Grid Computing is a kind of distributed computing architecture in which clients obtain resources from geologically located service providers and permit to communicate with them. Cloud computing is a big-sized framework comprised of virtualized computing machines, storage space, computing platform, software application, computing power, and service platforms are provided on request to outside clients over an internet linked medium in a business oriented environment. According to [3], a system in which a number of computing machine work as a stand-alone single computing structure are distributed and parallel system known as clusters. From the descriptions, the cloud computing intersections several features of these computing technologies so it is clear that Cloud Computing traces its roots from Grid and Cluster Computing which are well-known utility computing standards. Grid Computing traces its root from Electric Grids where it offers electricity as a metered service by the execution of the computing Grids; consumers are priced for the services they use according to Pay-per-use model.

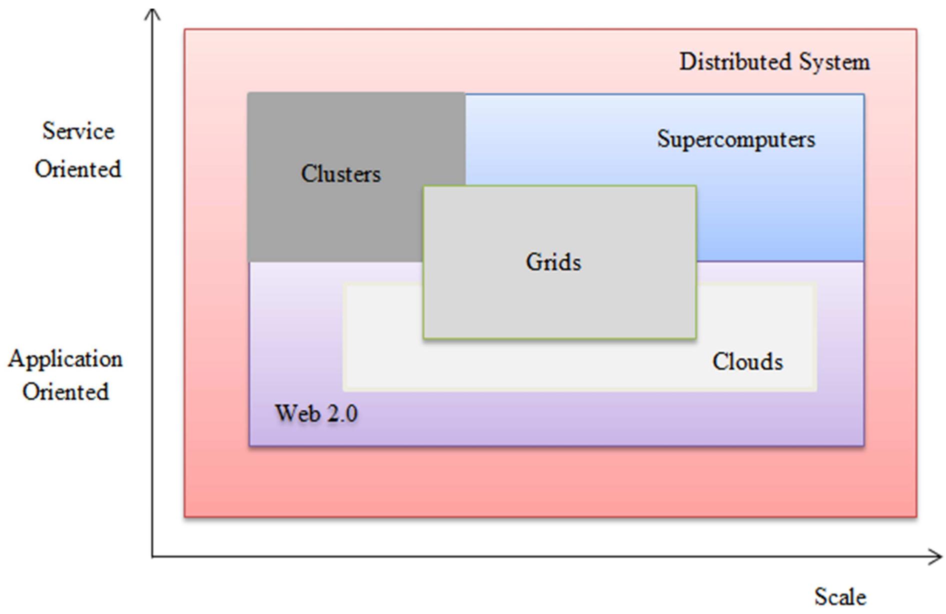


Figure 1.1 Overview of Cloud Computing and Grid Computing

Figure 1.1 shows the connection between cloud and other fields, how they scale up. These utility facilities come below in the distributed system, in this distributed system cloud computing machine interacts, communicate, and synchronize with other machines through a network. Every machine distinguishes its work and accomplishes the jobs given to it, even if a machine requirement some resources to compute that task, then that machine request for resources on network. Supercomputers, Cluster, Grid, and Cloud Computing are connected with each other because cloud is obtained from the grid and clusters.

1.1.1 Evolution of Cloud Computing

The vision of global network came before the world in the form of internet in 1969 which was actually a research project at Advanced Research Projects Organization under the Ministry of Defence, United State. The internet was used for scientific and military applications. In 1988, internet was used for commercial purposes by providing services like email and telnet. Services provided by Cloud Service Providers (CSP) are mainly accessible through internet. Many authorities believe that the vision of cloud computing was the perspective of American computer scientist John McCarthy that he had presented in sixties. In the last sixty years, there have been several evolutions in the past of computers from very big size computers in 1970s to small personal computers in 1999. After that the distributed computing came.

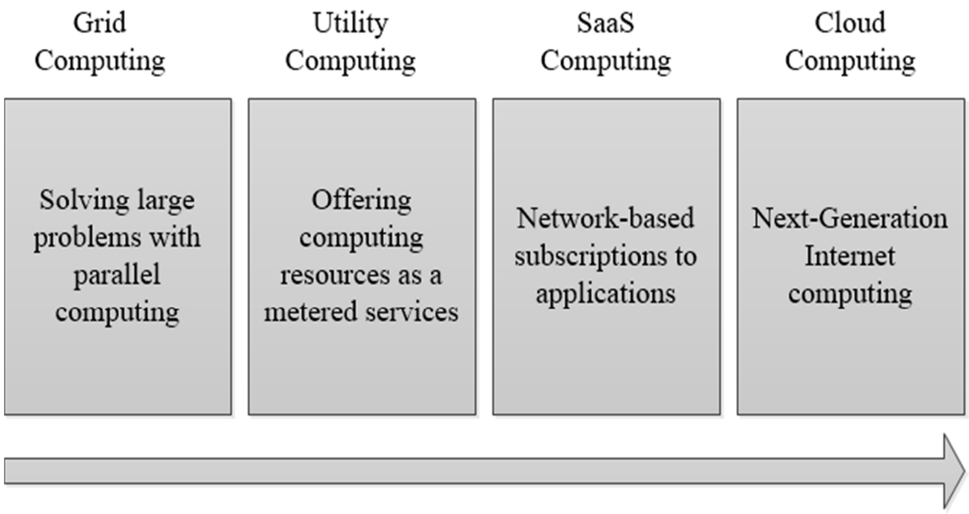


Figure 1.2 Evolution of Cloud Computing

For computational purposes companies want to start compact machines so they took an initiative to migrate from centralized computing to distribute computing in 2010. This cost effective method of implementing computational necessities was known as Cloud Computing. The procedure of such computing, Computing-as-a-Service, was provided by utilizing internet [2].

In Figure 1.2, Cloud Computing has progressed through various parts which contain SaaS, Grid Computing, and Utility Computing. With the purpose of fulfilling a common task a grid computing is the accumulation of different computing resources and services which are collected from multiple sites these resources are distributed and heterogeneous.

But the scope is limited in the research work and scientific applications in Grid Computing, which is the limitation of Grid Computing. Metered services are provided by Utility Computing based on the consumption of the users liking traditional electricity consumption.

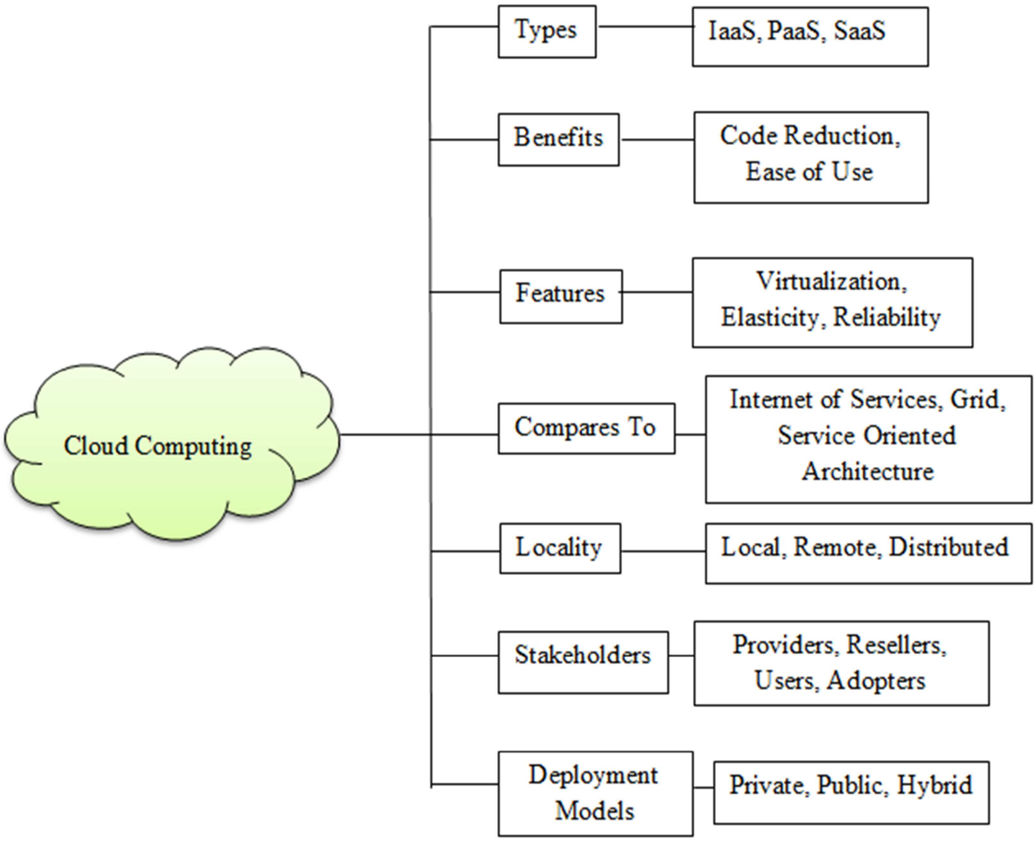


Figure 1.3 Features of Cloud Computing

The third part of the evolution "Software-as-a-Service" (SaaS) progressed in 2010 which both framework and application are kept on the server and users made able to connect to the remote server using the internet. One part of cloud computing services is SaaS. Other parts of these services are IaaS and PaaS. There was no customary description of this computing standard, so Institute of Standards and Technology has suggested a description in 2011 which includes essential characteristics and deployment models of cloud. Figure 1.3 shows the features of cloud computing.

1.2 Deployment Models of Cloud Computing

Figure 1.4 shows the different types of deployment model of Cloud services [4]. The details of models are described below:

- **Public Cloud** - It is also identified as external Cloud which provides services like platform, information, business process and resources which is publically available and distributed by all Cloud users via internet. Using the internet, these Services are offered by a third party vendor. This deployment model mechanism works on pay per use billing model. Cost effectiveness and scalability is the main advantages of public cloud. Google App Engine (GAE) and Microsoft Azure (MA) are a few examples of public clouds.
- **Private Cloud –** This type of cloud model is controlled within an organizational basis for hosting of private computational resources for the use of only particular certified users. This is for private use of the business replicating the Cloud. Only authorized users are able to utilize the cloud resources and it is more secure Cloud. The infrastructure and maintenance cost is same as public cloud, but in scalability and sharing of costs are different. Private Cloud service providers are Amazon Virtual Private Clouds and Rackspace Private.

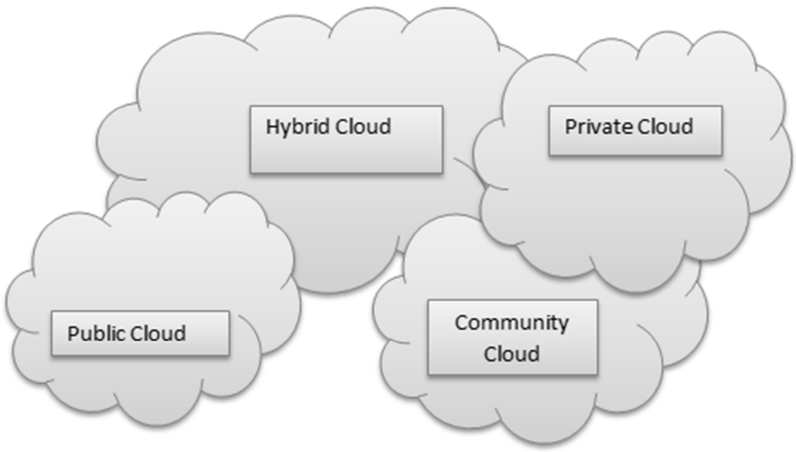


Figure 1.4 Deployment Model of Cloud Computing

- **Community Cloud** - Various organizations work organized on several project works. The organizations use the facilities of unrestricted cloud which are able to share among various organizations for particular interests. It is a shared effort which benefits in distribution of infrastructure among particular community. It is also accomplished by third party.
- **Hybrid Cloud** - It is the fusion of private Cloud platform with the public cloud provider to solve several computing requirements. It is generally designed for single business. The autonomous infrastructures of private and public cloud connect with each other over encoded link. The complex data need not to be

visible to third party as it is kept on public and public Cloud which is used to deliver computational resources. The most common attribute for hybrid cloud is elasticity.

1.3 Characteristics of Cloud Computing

Many researchers explain various essential characteristics of cloud computing [5] are described below:

- **Reliability** - Service providers duplicate the content stored on servers to provide reliability by verifying destructed services. There is no damage of content and it tries to avoid delay in cloud services to the users as it duplicate content on the servers specifically used for duplication purposes. So the data still persist even if certain problem happens in any server.
- **Location Independence** - Information uploaded by cloud users is preserved on unidentified servers and their locality is hidden from user. But their data is able to be read from anywhere, anytime via internet. Cloud customers have no idea of procedural complexity of preserving and accessing data from the servers. Through these qualities of cloud computing, user's applications and information are preserved at optimal sites that assist in decreasing access time, fast regaining of data and better level of correctness for the information.
- **Flexibility** - Flexibility offered by cloud in retrieving data as it provides an attribute of retrieving documents anywhere, anytime just by linking with internet through several devices like smartphones, laptops, personal computers tablets, mobile phones etc. User's data is preserved centrally which assists them in distribution of information among friends, family members and their coworkers in organization. Memory (storage space) is purchased from cloud resource suppliers according to the requirements of the users and they are able to purchase more space if necessary, any time and have the facility to even release space if not necessary anymore.
- **Decreased Cost -** The main essential attribute is pay-as-you go prototype that is used by various SaaS providers to offer services to their clients of cloud computing. The principle of this model is 'the more you use, the more it will cost'. Users spend according to their utilization. Small business industrialist has the option to simply balance the growth by hiring infrastructure from cloud

service providers without concerning about maintenance of resource price. They acquire current technology and they do not have to concern about software up-gradation. The only need of them is the internet connectivity to associate themselves to the servers located remotely and access their services.

- **Large Storage** - These days, huge amount of data is produced and it is increasing exponentially which has become a complex task. Maintaining such a huge amount of data has several phases such as exploring, storage, security and managing of bulk of data. Cloud computing has provided its solution by providing huge memory space to store data so that dynamic scaling become efficient without effecting performance of the cloud system.

1.4 Cloud Computing Services

Figure 1.5 discusses various layers linked with the different models of Cloud Service [6]. The details of models are described below:

- **Software-as-a-Service (SaaS)** - These types of services is retrieved over internet; in this user runs application on client side. The user does not recognize where the application is running but the user is able to still store, install its data. These services are to be retrieved anytime when necessary. Users only interact with the interface which is distributed by the cloud service provider and get access to information through a web browser. Today in the market, many facilities are provided as a SaaS provider and Microsoft, Google, Amazon is big giants of IT.
- **Platform-as-a-Service (PaaS)** - Some research scholars consider PaaS as a part of SaaS with several less services than SaaS. PaaS offers facilities which effort on development of applications and providing computational resources online which support developer in handling software or hardware. In this service, users deploy and develop or modify applications. Through PaaS enterprise, testing of applications, deployment becomes easy and cost effective. Cloud Service providers offer memory (storage space), network, operating system, office suits etc.
- **Infrastructure-as-a-Service (IaaS)** - These services provide computer infrastructure to the consumer. With this processing power, softwares, connections, servers, network and other computational resources, etc., are

offered as a service by cloud service providers. IaaS is the control on resources by the Cloud user, which is the main difference between PaaS and IaaS. They acquire full control to accomplish the resources in IaaS. They are able to define the factors according to them such as different types of servers etc. This kind of facility is proper for those corporations which have software packages existing with them. So they place this software on cloud. Examples of IaaS are Rackspace, Amazon Elastic Compute Cloud.

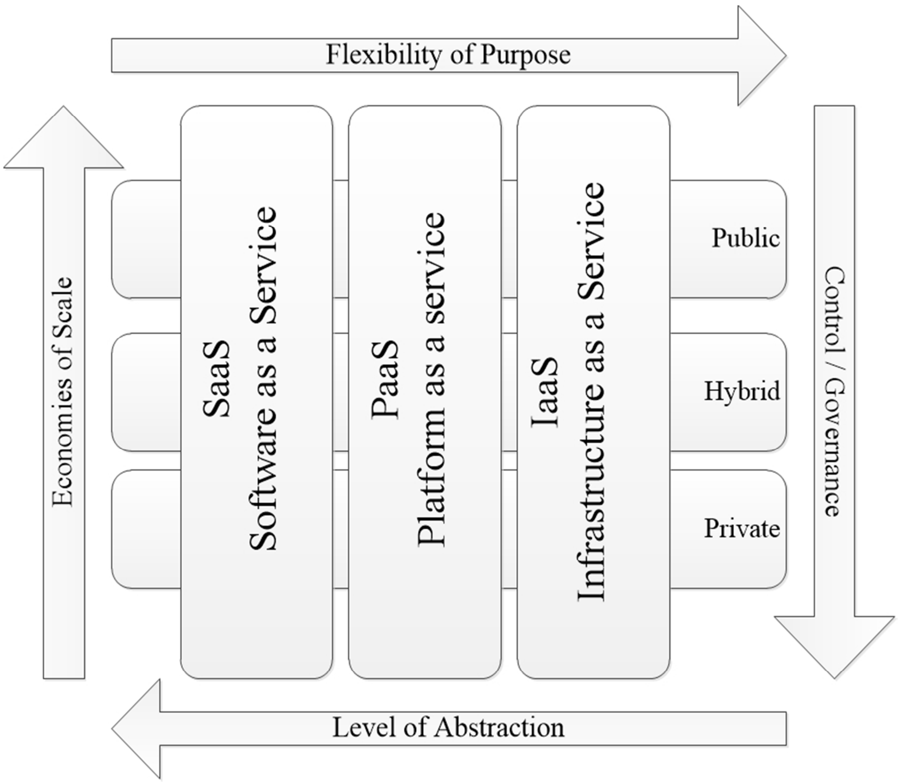


Figure 1.5 Services of Cloud Computing

1.5 Cloud Computing Applications

Cloud technology computation and users are rising day to day over every field as it is flexible to adopt. It is being adopted in every field, both big and small. People are shifting towards cloud due to its low initial investment. Organizations around the technology fields have realized the cloud computing importance, as it is allowing them to solve major problems associated when using internet for providing services. The important reasons that favors cloud computing are:

- Adopt new technology to provide computing services.

- Reduce initial infrastructure cost, increase resources like storage, computation power as needed.
- Increasing size of data associated with user.
- This section provides few applications of cloud computing in science and productivity field.

1.5.1 Scientific Applications

Cloud computing application uses high computing power and throughput. Experiments and research associated with scientific field demands huge input resources and for very short period. Cloud technology proves best for such types of services. Few successful application and services of cloud are discussed below.

- **ECG Analysis through Cloud –** Regular health monitoring is a new craze in field of health. In cloud health service, cloud analyses the Electro Cardio Graph (ECG), and provides prediction about heart related diseases. An instrument is used to check the heart of patient and regularly get data about its condition. This data is transferred to mobile phone. It then forwards it to cloud which provides different information after analysis.
- **Satellite Image Cloud Processing –** Non-spatial and spatial data is gathered by satellite. This data collection involves huge data processing that turns it into useful Geographic Information System (GIS). A cloud application made up by Indian Space Research Organization (ISRO) to analyze these images [7].

1.5.2 Productivity Applications

Productivity based application implement cloud computation for few commonly used platform services. Website development, document storage, work automation is some common task done efficiently using cloud platform. Some productivity applications of cloud computing are:

- **Drop Box** - Drop box provides file hosting as cloud service by Drop box, Inc. offering cloud storage synchronizing files, and client side software. It also allows its users to make dedicated folder for each computer, which is later synchronized so it appears as same folder irrespective of which computer is used to see it.

- **RoboEarth** - This project is managed by the Eindhoven University of Technology (EUT), Netherlands, to create a world wide web of robots with a big database so robots could share data about objects, surrounding and tasks. Researchers of A-Star Social Robotics laboratory (ASORO), Singapore, have developed cloud computing base that enables robots to create 3-D maps of their surrounding environment at faster rate than they would have with onboard computers.

1.6 Research Issues in Cloud Computing

There are various open challenges [8][9] to be addressed so as to become completely accepted by the business and to build this model more protected, well-organized, cost effective for user's requests.

- **Fault Tolerance and Scalability** - Fault tolerance is the best captivating characteristic of Cloud, which is works on demand scalability. So the cloud service should be considered scalability along with today's workload. In this situation it becomes essential for application services to bear failure. Cloud bears so much due to hardware downtime. Therefore, it becomes a big Issue for researchers to purpose fault tolerant applications which are achieved easily.
- **Energy Management -** Due to increasing requirements for calculation power by network requests and scientific purposes, big datacenter is created which are utilizing power source. A big amount of cost is paid on the electrical power consumption. So it is a big challenge for researchers. There is an essence for energy efficient methods which could help in decrease of operating cost and redeemable energy.
- **Security of Data** - Users load their information and content in cloud which is kept on the existing servers at several unidentified sites, other countries, due to which information is pretentious both by the laws of individual countries and cloud service provider's policies. It is necessary for protected data from unauthorized access and maintains the integrity of user's data. So the issue is increasing data security, integrity and provides a trustable cloud environment.

- **Interoperability** - With the aid of independent manager from one cloud to another for obtaining essential performance and still preserving transparency is a main research challenge for Cloud researchers.

1.7 Cloud Resource Management

In [10] the authors have suggested that the resource allocation is done to efficiently handle fluctuation of workload while guaranteeing QoS to the users. Available network and computing resources are usually limited and need to be effectively distributed among the end users. To provide efficient management of resources, issues like resource mapping, resource allocation, resource provisioning and resource adaption must be considered. Lack of maturity in virtualization tools and processors prevents cloud computing growth. Recent work done by [11] has examined future and current challenges to provide better service to providers and users of cloud computing. Not much has been done for understanding present operational challenges user face when they try to execute software in the cloud. In [12] the study shown that the problem associated with efficient management of energy of different types of resources at Internet hosting servers. The main problem was to measure the resource demand of every application at its present request load and to efficiently allocate resources.

Research organization includes monitoring process, resource discovery and allocating processes as shown in Figure 1.6 below. Process management involves managing physical resources like network bandwidth, disk space, and CPU cores. Resources are divided and distributed among virtual machines executing potentially heterogeneous workloads. We study the taxonomy of resource management in Figure 1.6 below.

The basic element of managing resources is the discovery process. It has included searching the appropriate types of resources available that meet the application requirements. Cloud service provider manages these resources. The process is taken by user broker or resource broker to search available resources. Discovery involves brief description of available resources. According to [13], resource discovery gives a method for resource management system (RMS) to determine the state of the resources that are being managed and other interoperable RMS. Resource discovery executes by disseminating resources and provides state information of resources to the server.

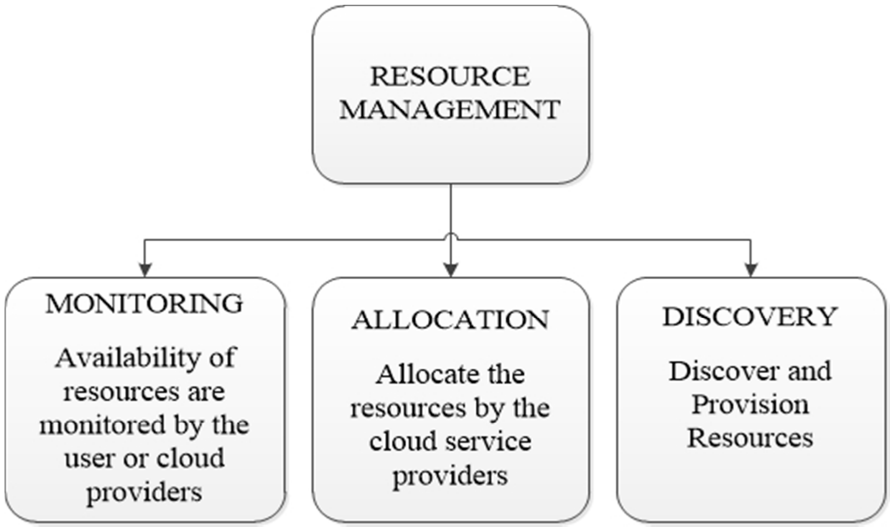


Figure 1.6 Elements of Resource Management

The allocation process involves assignment of available source over the internet to cloud application. The allocation of resources is done to meet user request and payment system. In this process, dispatching and scheduling is applied for allocating resources. The scheduler schedules required services and other resources to the end user. Then contributor allocates the required resources to the client.

Resource monitoring as elaborated in [14] has proved to be an vital tool to control and manage software and hardware infrastructures. It has also given a Key Performance Indicators (KPIs) and data for both application and platform in cloud that is used for data collection in assisting decision method of resource allocation. It has been the main component for monitoring the resource state in the occurrence of an error whether at the service layer or at the physical layer.

Group Discovery and Allocation as the provisioning process, while one single process is being monitored. However, given three processes are interrelated in order to provide resources to the end users. Discovery process is the important element in resource management. It contains discovering the suitable services that are demanded by users and essential by applications. It has organized by service provider. The procedure has used by broker to discover resources. Discovery process has included full explanation of existing resources.

1.7.1 Modules of Resource Management System

The basic modules of cloud resource management system are [15]:

- **Service Level Agreement (SLA) Management** - The SLA Management module prepares a contract among the consumer and the cloud resource supplier for services demanded by the consumer. The contract comprises of the asset requests of a consumer such as control processing unit (CPU), memory (storage space) and other architectural design of a virtual machine (VM). It also comprises QoS requests like task turn-around time, execution time, cost of operating etc. In this agreement the consumer is able to settle with cloud resource supplier for amount of the facilities delivered by them. This component interconnects with price manager before a contract is prepared. After completing an agreement the consumer sends demands to the scheduler.
- **Admission Controller -** The demand prepared by the consumer through SLA is authenticated dependent on the existence of demanded assets and other limitations identified in the SLA. If the prerequisite cannot be fulfilled then the demand is dismissed by the cloud resource provider. This circumvents provisioning of illegal tasks onto the cloud assets.
- **Pricing** - Pricing is estimated for dynamic tasks dependent on its utilization of assets. The consumed assets cost and time is evaluated and priced to the consumers in real time.
- **Task Scheduler** - The task scheduler provides the demanded job on the existing VM or provides a demanded VM on the Physical Machine. In both case there is a line to identify the importance of the job. At a time, only one task is performed on a VM. The job of scheduler is to select on which Physical machine or Virtual machine the task has to be provided. For this motive, the scheduler gives information of the datacenter such as existing CPU, storage space (memory), bandwidth etc. and this content taken from a load balancer. After execution of tasks lodge by the consumer, the pricing component is used to make the bill for the assets utilized by the consumer and this is interconnected to the consumer through the SLA management component.
- **Load Balancer -** The load balancer is used to discover the underutilized and over utilized physical resource or virtual resource. It equilibriums the workload between Virtual machines to overpower the assets utilization. In this method, the energy is preserved by utilizing assets.

1.7.2 Types of Resources

Within a computer system a resource is limited availability and these resources are any physical and virtual mode. Each device linked to a computer system is an asset. Each internal device module is an asset. Various logical and physical assets are discussed below [10]:

A. **Physical Resources**

Memory, peripheral devices and processor etc. are physical assets. Physical resources differ equally from processor to processor. For instance, a usual Personal Computer system might have 640 K of memory storage space, one Winchester disk drive of 20 MByte, one floppy diskette drive, only one keyboard, and only one video display controller or device. A usual mainframe computer has numerous parallel processors, hundreds of disk drives, memory storage space is in the tens of millions of bytes from, hundreds of computer terminals, external storage medium such as tapes, and other important motive peripheral devices, and is linked to a worldwide network with thousands of new homogeneous computers. Cloud service providers provide asset provisioning strategies to customers, viz. long-term on-demand and short-term registration strategies. Outsourcing methods are utilized to abduct benefit of cloud computing hardware platform for providing availability and scalability abilities to the web requests installed on it; this would surely grow the numeral of cloud customers and hence raise the asset utilization of clouds. Now several important physical assets are discussed below:

- **Central Processing Unit (CPU)** - It executes most of the execution inside a computer system. The CPU utilization is the big issue in cloud computing, which refers to a computer system utilization of processing asset, or the amount of workload controlled by a CPU. Real CPU utilization in cloud computing differs depending on the quantity and type of controlled computing works. Assured jobs need full CPU time, while others need less CPU time because of non-CPU asset requests. Better CPU utilization makes it simply to devour great amounts of execution power for batch process execution, data exploration, and high operational performance computing requirements.
- **Memory** - The cloud computing architecture requests for a clustered computing structure of the memory assets in the way of virtual objects. Departed are the

times when memory management was completed consuming the static approaches. As the cloud computing atmosphere is dynamic and unstable, there is a vital requirement to imprint the dynamic allocation of memory technology in the cloud computing based network. The larger number of hubs in cloud computing servers joined with the fast acceptance of virtualization techniques also generates vast request for memory.

- **Storage** - It is preservation of data stored on other sites maintained by a third party. Instead of preserving data to computer's hard diskette drive or other internal data repository device, we preserve it to a distant database. The Internet builds the link between the system and the data repository. Cloud computing repository systems usually dependent on hundreds of data repository servers. Because computers sometimes need management or repair, it is essential to preserve the similar data on a number of machines. Such process is known as redundancy. Without duplicity or redundancy, a cloud computing repository system could not guarantee customers that they could get their data at certain time. Most systems preserve the similar information on servers that utilize distinct power sources. Customers get their information even if single power source fails. The two largest matters about cloud computing repository are consistency and security of data.
- **Workstations** - IT supervisors are observing a technology where fast PCs are be categorized as workstations. Larger CPUs, quick graphics, and RAM up to 20 GB–these are devices that are layout to do a lot of internal processing. The obstacle is to take benefit of cloud techniques and finances and usage of workstations to take high-performance computing (HPC) capabilities to the company consumer or other consumer linked to the Internet with better connection. It has been observed that the consumers acquire the better services every time, accurate operational working of workstation is very much demanded. The higher configuration systems should be used as workstations so that they tolerate the overloaded computer network. The problem here is to observe that workstations work by itself, without human interaction. Over this, monitoring network as a service offers by clouds.
- **Network elements** - Handling lots of network components such as hubs, switches, bridges etc. result in to unsustainable management costs, demanding computerized techniques for usual system management jobs. The computerized techniques

requirement to assign with large number of monitoring network areas of various instructions of sizes greater than recent systems. Through this, clouds offer communication between nodes as a service.

- **Sensors/actuators** - Applications including Internet-connected entities such as sensors, devices, has just given to increase to the concept of clouds, which are supported as big sized networks of distributed objects with accessible devices and storage capacity.

B. Logical Resources

Logical resources are abstracted to gain provisional control of physical resources. It supports application development and effective communication protocol. Logical resources are important in cloud computing as follows:

- **Operating system** - It offers users the "logical" view and well supported environment for managing physical resources and mechanism along with policies to control different objects and resources. Operating system allows managing file, device, performance, security and fault tolerance, thus facilitates effective utilization of resources.
- **Energy** - The basic technique implemented to minimize consumption of energy is focusing the workload so to have minimum of switching and physical nodes. This method requires maintaining performance and power trade-off, as overall application performance is degraded with workload consolidation.
- **Network throughput/bandwidth** - When building services over cloud there is always concern of measuring the data throughput being measured bits per second of a network access or communication link. A trivial method to perform measurement is sending a large file to another system and calculates the time being taken for transferring complete file or copying the file. Network efficiency is increased by improving throughput. Congestion is prevented by managing bandwidth and protocol, usually by refusing and accepting cell. Most critical problem is bandwidth allocation and is concerned of successful integration between services of different type.
- **Information security** - End-users need to feel easiness when using a cloud service that contains their processes, software and data; there is need of highly reliable services that provide assurance of availability, safety, privacy and

protection. Hence there is need for considering different security issues, like non-repudiation, authentication, integrity and data confidentiality.

- **Delays** - Even a small difference of data transfer creates significant change in service, when the discussion is about data specific traffic, the user experience along with cloud services, or the method to of trade. A cloud computing service provider must make accurate choices to scale up or down resources in data-centers by taking into consideration other utility criteria like the interval in virtual resource setup, migrating available processes, resource consumption, etc.
- **APIs (Applications Programming Interfaces)** - It acts as a communication link between different software components through different protocols. It includes different object classes, specification for routines, data structures and variables. Different forms of specification for API exist like POSIX as international standard, vendor documentation MS API, programming language libraries like Java API.
- **Protocols** - Different protocols include functionality for device identification and build communication with other along with formatting rules specifying data packaging mechanism as they are transferred. Several different network protocols are being developed in cloud environments for various purposes. Various examples of protocols discuss user identity management protocol (U-IDM protocol) for cloud service providers and cloud computing customers. It authorizes and authenticates different customers and providers.
- **Network loads** - Cloud applications present variable workloads. Therefore it is necessary to perform an analysis of cloud services, capabilities etc. to recognize common problems, explore load balancing, patterns, and approaches which could potentially facilitate efficient resource management and improve process efficiency.

1.8 Resources Allocation

It is a procedure of allocating the existing resources in a commercial and efficient manner. It is the task of scheduling of the existing resources and existing activities needed by those events while considering both the resource obtainability and the project time. It resolves such problem by permitting the resource providers to handle the resources for each distinct demand of resource.

Resource allocation [16] is a commercial and effective way of allocating the resources. Resource allocation is used to schedule the resources and existing activities while considering both the resource accessibility and the project time. It solves the problem of distinct demands of users for distinct resources.

Resource Allocation Strategy (RAS) is a task for distributing and utilizing limited resources in the boundary of cloud computing environment to meet the requirements of the cloud application.

1.8.1 Significance of Resource Allocation

By managing resource accurately, the resource services are utilized in an efficient manner. Provisioning of resources solves this issue by managing the resources for every individual.

Resource Allocation Strategy (RAS) is the integration of cloud resource provider activities for making the use of cloud resources and allocation of scarce resources in cloud platform so as to reach the needs of the cloud service application.

Points to avoid for optimal resource allocation:

- When two service applications on cloud attempt to access the same type of resource at the same time which causes resource contention situation to occur.
- Problem of shortage of resources occurs in case of restricted resources.
- Resource fragmentation occurs when there are private resources.
- Over-provisioning of computing resources occurs when the service application gets additional resources than the demanded one.
- Under-provisioning of computing resources occurs when the service application is allocated with less numbers of resources than the demanded.

1.8.2 Auction Model and VCG Technique

- **Auction Model –** An auction model is defined as the technique in which service providers provide services in an online auction and consumers bid on services according to their requirements. The consumer with the maximum bid wins the service. And eBay is the largest site. This model generally provides best work for consumers who have a list that varies, and when the service provider does not essence to retail at a fixed price with the purpose to make revenue. But sometimes this model is not beneficial for retailing services.

There are several sites available which auction translation and web designing services etc. via freelance. A large number of people visit of this an auction sites. These sites offered cloud services of consumers.

- **VCG Technique –** Vickrey–Clarke–Groves (VCG) auction is an auction which comprises of sealed-bid auction of various numbers of entities. In this technique, the bidding is done by users which submit bidding value mentioning the value of the item. During this process the bidder has no idea about the bidding value quoted by other participants of the auction. This auction mechanism allocates the assets in an optimal routine. The optimal method must be socially optimal as it demands the price which is equal to the loss they have employed to other participants caused for them. This auction technique provides bidders to quote the valid or true price value. It is possible because of the optimal price quoting strategy. This mechanism is the overview of Vickrey auction for considering several items at a time.

1.8.3 Advantages of Resource Allocation

Resource allocation is beneficial while utilizing cloud computing irrespective of business market or organization size. Although, it is limited as this technology is still in evolving stage. The comparative perspective of advantages and limiting factors of allocating resources in cloud is given below:

A. Advantages

- Resource allocation allows user to access the ubiquitous services without installing software. This facilitates development and hosting of services through internet.
- Accessing resources from anywhere irrespective of geographical region is another major benefit. The applications and data is able to be accessed from anywhere in the world, on any system.
- No technical effort from user side is needed.
- Resources of cloud are shared through internet to avoid resource scarcity.

1.9 Structure of Thesis

This thesis is structured in the following manner:

Chapter 2 – This chapter contains the literature survey which includes evolution of resource allocation strategies in different cloud computing techniques. It has also contains the work of many researchers in the area of cloud computing and resource management.

Chapter 3 – This chapter explained the problem statement and objectives of proposed work in the field of resource allocation in cloud computing.

Chapter 4 – This chapter offers solution to the problem which is explained in previous Chapter by proposing an algorithm for resource allocation.

Chapter 5 – In this chapter implementation of the proposed algorithm is discussed via simulation tool CloudSim and Netbeans IDE. It's also includes experiment results and comparison with the other present technique**.**

Chapter 6 – This chapter includes the conclusion, summary and future scope of the work proposed in this thesis.

Chapter 2

Literature Survey

In this chapter different issues related to associated research of resource allocation methods of Grid and Cloud Computing are discussed and brief comparison between the available methods has been done.

2.1 Resource Allocation Strategies

The tree diagram in Figure 2.1 has depicted the taxonomy of Resource Allocation Strategies (RAS). Vinothina *et al.* [17] have discussed the RAS in cloud in the following section.

- **Execution Time** - Several resource allocation mechanisms has introduced in cloud. Jiyani *et al.* [18] have considered actual job execution time and pre-emptible arrangement for allocating the resources. It has resolved the issue of resource contention and increased utilization of resource by using diverse modes of renting computing resources. Majumdar [20] has suggested that calculating the implementation period for a job is a tough task for a cloud user and errors are made very frequently. Melendez and Majumdar [19] have proposed a matchmaking strategy which was based on the criteria of any-Schedule-ability for assigning jobs to unclear resources in diverse environment.
- **Policy –** Shin and Akkan [21] have proposed a method in which as centralized customer and cloud resource management have shortfall in management of users in scalable way, computing resources and organization level safety policy has projected a distributed customer and virtualized cloud resource management for IaaS by accumulating a new level layer called domain in between the cloud user and the virtualized cloud resources. Based on the method of role based access control (RBAC), virtual cloud resources were allocated to users through domain layer.
- **Virtual Machine (VM)** - The system possessed a set-up of virtual machines capable of live relocation across multi- domain infrastructure. A number of research scholars have introduced efficient resource allocations for real time

jobs on systems having multiple CPUs. In studies done by many researchers, they have scheduled tasks on preset number of CPUs. Guo *et al.* [22] have described method to allocate virtualized cloud resources amongst selfish VMs in a not so cooperative environment. The non-cooperative environment consists of the VMs concern basically about its own profit. Stochastic approximation approach is used to design and examine QoS performance.

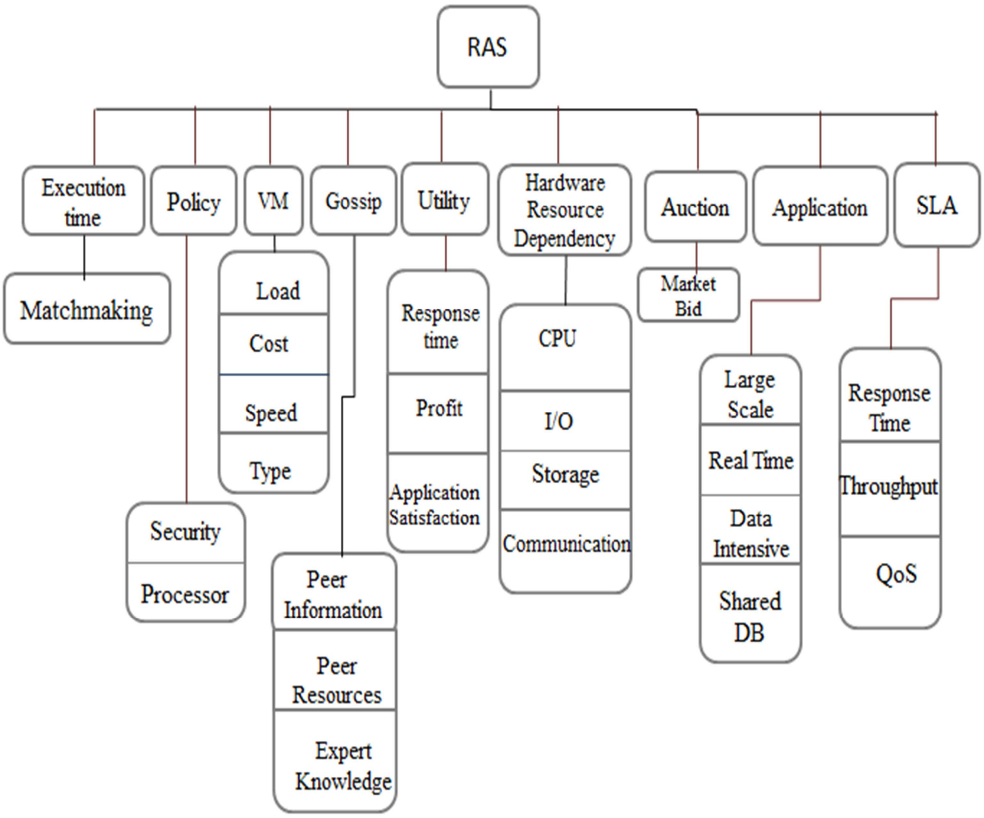


Figure 2.1 Resource Allocation Strategies in Cloud Computing

- **Gossip –** Wuhib and Stadler [23] have reported a gossip-based protocol for allocating the resources in big scale cloud environment. It has performed a main role within distributed middleware design for big clouds. The system was designed as a dynamic set of nodes which represents the cloud environment. Every node in the cloud environment has a definite processor and memory capacity. Niyato *et al.* [24] have suggested Gossip based VM management along with virtual machine allocation and cost management. Using this method of allocation, the organizations cooperated with each other to share the available resources, thereby reduces the cost.

- **Utility Function –** Gmach *et al*. [25] have proposed a method that manages VMs in IaaS by improving objective function like cost minimization function, cost performance function and achieving QoS objectives. Zhu *et al.* [26] have suggested an objective function which was defined as a usefulness property chosen dependent on units of implementation time, number of QoS parameters, goals achieved and profit etc. Some works are there that has dynamically assigned CPU resources to convene QoS objectives by primarily allocating requirements to high priority service applications. The authors have not tried to get the most out of the objectives. Therefore, Dorian and Freisleben [27] have suggested a Utility based resource allotment for VMs which uses live VM migration as a resource distribution mechanism. This has controlled the cost performance trade-off by altering VM utilities.
- **Hardware Resource Dependency –** Hu *et al.* [28] have suggested Multiple Job Optimization (MJO) to enhance the hardware utilization. Jobs were categorized by physical-resource reliance. These are CPU bound, Network Input/output bound, Disk I/O bound and storage memory bound. MJO scheduler has identified the kind of jobs and parallel jobs of unusual categories. Resources were allocated based on these categories. This system has mainly focused only on CPU and I/O resource.
- **Auction** - Resource allocation in cloud by auction mechanism has introduced by Lin in [29]. The suggested method was dependent on sealed-bid auction. The resource supplier gathers all the consumers' bids and finds the resource price. The resource was shared among the first k^{th} highest bidders below the cost of the (k+1)th highest bid value. This has made the cloud provider decision rule easy and clear the allocation rule by converting the problem of resource allocation into ordering problem.
- **Service Application** - Resource Allocation strategies have also suggested based on the nature of the cloud applications by Ali *et al*. [31]. Huu and Montagant [30] have proposed that allocation strategies for virtualized infrastructure are modeled for workflow based cloud applications where resources are assigned using the workflow of the application. For applications based on workflow, the application logic has interpreted to generate an execution schedule estimate. This has helped the user to approximate the

accurate amount of resources that was used for each run of the application. Services group optimization, FIFO, Optimized and Naive are the four strategies which have modeled to allocate resources and schedule computing jobs.

- **SLA** - The works related SLA consideration is still used in cloud. In order to attain the SaaS provider's goal, a range of RAS particular to SaaS layer in cloud has introduced. With the appearance of SaaS, cloud applications are moving away from personal computer based to network delivered-hosted services. Moreover, Lee [32] has described the issue of profit driven cloud service request task scheduling in cloud by taking into account the goals of both communicating parties such as service providers and cloud users. Chiu *et al.* [33] have suggested that organizing the cloud resources for SaaS tasks was demanding for SaaS providers. Hence a framework for resource administration for SaaS providers to proficiently manage the cloud service levels of their clients has contributed by Lui *et al.* [33]. All the above described has primarily focused on SaaS provider's profit and considerably decreases wastage of resource and SLA violations.

2.2 Existing Economic Models in Cloud Computing

2.2.1 Price Based model

- Buyya *et al.* [34] have proposed a design which has simulated and developed a cloud asset pricing model. This design has satisfied main basic constraints: the dynamic characteristic of model to give better satisfaction guarantee of Quality of Services (QoS) – from user perspective and profitability constraint – from cloud service provider perspective. Financial base theory and usage of cloud assets as principal resource to capture the real value of the cloud computing commodities have implemented. The authors have studied the results of four different metric which have introduced and guaranteed QoS and price rate as follows: (a) Moore's law dependent asset value depreciation; (b) new ubiquitous technology dependent volatility technique for catering price change; (c) a different finance option price based model that combines above two methods; (d) the age effect resources and its depreciation. The authors have presented that the cloud attributes drawn to finance economic model and studied the result of cloud commodity pricing for different parameters, like resource age, QoS and contract period.

- Mihailesu and Teo [35] have presented a pricing mechanism which was appropriate for coherent client request comprises of several resource. The proposed technique strategy-proof dynamic strategy has compared with fixed cost mechanism. The number of effective requests was improved with the help of dynamic pricing. The techniques of resource allocation have been discussed for normal cloud and federated clouds. In dynamic pricing techniques the coherent clients supply and use a resource, which was why this technique was more appropriate for federated clouds. The utilization of user was also constant. This scheme was capable of making a balance between the completed requested and number of assigned resources based on market scenario. The authors have proved that this scheme works better in case of average customer welfare and it has raised the benefit of seller up to 200%. Rate of effective has also risen up to 200%. Using hash table of distributed nature the authors have produced a peer-to-peer network and also it has provisioned delivery of cloud resources and assignment using this mechanism.
- Wee [36] has learned the Spot Instance (SI) casted by Amazon Web Service (AWS) in excess of one year time period. A few of the cloud services has constant price in AWS. All the services rented by AWS to its consumers based on fixed rate before SI. This was the reason why the author has analyzed content of AWS instances. More enhanced price rate was attained for an occurrence of service if the load on cloud is transferred from peak time to idle time. The author description has proposed that the cost saving %age is 3.7% in average case. But in 52.3% was saved when dynamic pricing was used instead of fixed mechanism. This dynamic strategy has launched by AWS in which pricing was done dynamically. There exited no additional cost in this approach in which the cost of transfer was taken. Optimal reliability strategy has utilized to manage the cloud resources when the cost was increased to its highest level.
- Li *et al.* [37] have proposed an architecture which was called cloud banking model contained resource agency. From the global point of view, it has offered the guidance and analysis for each applicants, a price fetched up-to-date repetitious algorithm has been specified, it has examined the previous usage

proportion of the resource, and repetition has presented prices continually, have acquired the availability of assets next time, the ending price to consumers were computed. Experimentation displayed, based on past resources utilization, the pricing algorithm was dynamic in nature and has provided benefit for provider and more helpful to preserve the benefits of applicants of resources in the cloud.

- Yang *et al.* [38] have presented a pricing strategy; bidding process has used to calculate the autonomous pricing and quoting in which the value of resources was fixed by the cloud service providers. And the decision making approach was used to set prices.
- Chen and Yeh [39] have defined the market mechanism application for resource allocation in cloud computing services, which needed bridging finance and relative technical challenges associated with software agent. Dynamic change in the resource availability with time has required more complicated treatment. Here author implemented a mechanism for allocation and rating mechanism with market based model for resource allocation in cloud. Service providers and buyers have determined their charges and get prices using k-pricing that sets transaction charge individually for matched service-buyer pair. These methods adaptively meet customer/service provider constraints and requirements set by bundled services. The authors have defined the market-based method for cloud environment. It makes possible for buyers to assign arbitrary services to variety of service providers. The given allocation method and pricing method have developed a simulation environment with software agent to analyze the proposed model performance. The basic experiment has shown that model executes properly. The authors have focused on the trade exchange behavior, specially the service provider agent and buyer agent interaction. Market based method has shown better promise to enhance pricing and resource allocation. Present mechanisms, however, are inadequate to handle wide scale configuration with strategic service providers and buyers who benefit by manipulating the mechanism.
- Zhang *et al.* [40] have formularized the problem of allocating virtual machine in clouds as a combinatorial auction base problem and offered a mechanism with group price for solving it, so the cloud provider expresses the discounted

price of each variety of imported virtual machine. The authors have identified the theoretical attributes of the proposed mechanism along with individual rationality, truthfulness and individual rationality. Regressive simulation results have presented that the method prosed yields efficient allocation and traceable computation while generating more value for cloud providers s compared to fixed price mechanism.

- Li *et al*. [41] have proposed the group-auction dependent method for cloud based market to effectively allocate resources. In recent market system, asset providers offer assets as virtual machine. End users provide their bids. The system has given assistance to decide how a provider should efficiently allocate resources to users. Group action has also reduced the cost and improved the sales volume for participation. To decide resource allocation, the optimization problem was defined as Distributed Constraint Optimization Issue, and a distributed message passing method, max-sum algorithm was implemented by using group action formation game to search for early optimal solution within quadratic time complexity. In comparative study, authors have deployed the system in two scenarios. On grounds of acceptance rate and resource efficiency, the authors have compared this algorithm with the random algorithm and greedy algorithm. The important participants of this research paper are as follows: (a) suggested a budget-balanced methodology dependent on group-buying the cloud instance market. (b) Used the distributed algorithm max-sum to search for optimally allocating resources.

2.2.2 Auction Based Model

- Wei *et al.* [42] have introduced an auction mechanism which is based on dynamic strategy to solve the problem of resources allocation by estimating the ability in the cloud computing atmosphere. Truth-offering feature bears when we employ second auction mechanism into the allocation problem. The main focus of this feature is to help cloud service providers to increase revenue. The authors also have proposed a Gaussian distribution algorithm to allocate the resources to the consumers in efficient way.
- Prasad *et al.* [43] have talked about multiple resources obtaining from various cloud sellers contributing in bidding. This was completed by allocating dynamic pricing for these multiple resources. Since the authors have taken into

account the number of resources to be obtained from various cloud sellers in an auction, the issue became one of the combinatorial auctions. In the Combinatorial Auction Branch on Bids model the authors have first processed the user demands, then have examined the auction and announced a set of sellers bidding for the auction as victors.

- Zaman and Grouse [44] have designed an auction-based approach for dynamic VM provisioning and resource allocation that has considered the user request, when taking provisioning results. This mechanism has proved truthful (i.e., a consumer take advantage of its utility simply by bidding its correct valuation for the demanded package of VMs). And has estimated the proposed mechanism by implementing excessive simulation examinations utilizing real amount of work traces. The experiments have shown the better income for the cloud service provider and enhance the utilization of cloud resources.
- Xing-wei *et al.* [45] have introduced the microeconomic and genetic algorithms, and an enhanced periodical auction framework was built, and then ribbon capability was used to define the storage capacity for specific users. The optimized resource allocation solution has been pursued based on inherited algorithm. Experimental results have shown that the suggested model is both effective and feasible, which has maximized the vendor total trading amounts as well as reduced the executing time of winner determination.
- Wang *et al.* [46] have introduced definite microeconomic methods. These methods were to solve the resource management and resource allocation problem in the cloud atmosphere. By merging the concept of batch matching into the reverse auction technique, the authors have proposed a reverse batch matching auction technique. The approach of double penalty and the tracking down of service quality efficiently prohibit the existence of trading fraud which was based on auction. There are three estimation conditions as well as the market effectiveness, consumer satisfaction and quality of service which was used as the optimization purposes and the insusceptible evolutionary procedure was applied to find the solution of resource allocation in optimum manner. The simulated results have shown that this proposed mechanism is both effective and feasible.

- Bonacquisto *et al.* [47] have proposed a present marketplace of assets where the request and the proposal of assets were compared in auction-based deals. Particularly, the authors have observed at this market from the cloud provider point of view, which has required an approach to optimally assign their unemployed computing capacity. And the authors have also suggested an adaptive approach that, appropriately personalized to the providers' business purpose, provides them to increase the revenue in the obtaining auctions point of view. In addition, the effect of resource overfilling onto the utilization level of data centers in cloud has been examined by means of broad simulations.
- Ozer and ozturan [48] have proposed a mathematical model based on auction for offering economically effective resources allocation in these atmospheres. This model has enhanced multi-unit unbiased auction model which was combinatorial by presenting a fine price strategy which outcomes improved consumption of assets and hence increased revenue liable on the application field. This model has expressed via integer linear programming. The authors have proposed new five heuristic procedures based on meta-heuristic methods. And have presented a method called intelligent neighbor selection. The routines of these algorithms are matched with a marketable mixed integer programming (MIP) resolved on produced test parameters. The results offered by the progressive algorithm are as effective as the results offered by the MIP resolve for these test parameters.
- Wang *et al.* [49] have proposed an auction based technique in which the resources are time by time auctioned to house requests of users. The authors have focused on two important tasks of maximizing income and auction reliability. This model has comprised allocation methods based on capacity, which estimate the number of instances to be bid, in a specific duration. And also auction strategy, based on active payment method which corresponds to the suggested allocation scheme. This model is valid according to two dimensions and optimal if the demand is high. Additionally, by determining definite optimization design, the complexity of this model is reduced. Broad simulation has shown that this model keeps track of market variations, it generates better income than pricing on demand.

- YOU *et al.* [50] have proposed a mechanism to enhance utilization of resources in case of big sized data centers, which is known as RAS-M (Resource allocation strategy) on the basis of market scenario. The authors have first constructed the model of this approach, in which a function is designed based on quality of service parameters. The balanced state of this model is described and optimality proved. The Second approach is GA-based adjusting of cost price has been reported to handle the issue of attaining equilibrium level for this model. Simulation and experimental analysis have shown that the state of equilibrium is attained. This state validates this model and proves that this model is effective in nature to achieve its objectives.
- Zhao *et al.* [51] have studied the online procedure of taking auction mechanism according to allocation objectives. This online behavior of auction is in synchronization with asynchronous requests of clients. After wisely describing truth situations under the online gaining auction standard, the authors have proved that reliability is assured by a price-based assignment rule and compensation rule. The reliability characteristics has truly transformed the mechanism scheme problem into an online procedure model problem, with a borderline pricing function for cloud resources as variables fixed by cloud providers for this auction. The authors have derived the minimal pricing utility for the online procedure. The authors have also shown the inexpensive proportion of the communal cost of our process in contradiction of the offline VCG mechanism and of the cloud resource sharing cost of this procedure in contrast to that of the offline optimum auction.
- Aida and Fujiwara [52] have proposed a market strategy to allocate services to applicants efficiently. The mechanism has enabled users (1) to order a combination of cloud services for co-allocations and workflows and (2) to reserve current/future cloud services in a forward/spot market. The estimation has shown that the methodology works well in probable setting. And has offered a marketing procedure for allocating services to subscribers efficiently. This mechanism has allowed the users (1) to batch a combination of provided services for co-allocation and workflows and (2) to preserve current/future services in a present spot market. The analysis has presented good results proving that mechanism works in given probable setting.

- Lan *et al*. [53] have discussed a method for allocating resources using multi-unit uninterrupted double action mechanism and relative pricing method. Firstly, resource allocation strategy is constructed using continuous double action model which includes three agents. After that the existing double action refined scoring algorithm, and has provided new parameters to find the pricing algorithm which would be suitable in cloud. In each interval, cloud node agents and user agents find their bidding values and demand values respectively. This technique has taken into focus both fair chances for users to allocate resources along with supply of resources through cloud nodes and requirement of cloud nodes and users.
- Bello and Wakil [54] have proposed concept of achieving flexibility in the pricing models. It is done for SaaS users. Several use cases are implemented to get the possible timings in the pricing scheme. The discounts were given to the users on the basis of utilization of resources done by user. The concept has given good performance in terms of pricing in cloud computing.
- Shang *et al.* [55] have studied the current cons in cloud market and proposed a pricing scheme which is based on double auction. Using this strategy a uniform cloud market has been developed. This idea has created a good competitive scenario to other cloud service providers. Focus of this scheme was on enhancing flexibility of pricing model.
- Son *et al*. [56] has introduced a proposed model to resolve the issue of allocation of resources on remote location and data centres. It has solved the issue of limited computing resources in cloud. It is also based on SLA. It also focuses on assuring SLA in cloud computing.
- Meng and Mihailescu [57] have proposed a price determining technique which uses marginal bid. A technique called double auction is used to allocate resources in cloud computing. This model has better flexibility, but there is no analysis given to analyse the system’s performance.

Chapter 3

Research Gap and Problem Statement

The previous chapter described various techniques which have been discussed in the literature survey for resolving the resource allocation problem in grid computing and cloud computing. Relied on literature survey, several gaps discovered in cloud computing has been known in this chapter and the problem statement is discussed below. Thesis Objectives are also enlisted at the end of this chapter.

3.1 Gap Analysis

The literature survey in the previous chapter consists of various resource allocation market based models. There are a number of gaps discovered based on that literature survey which are explained below:

- Maximum techniques are biased towards service provider. But these techniques did not give benefits to service provider and consumer simultaneously in the uncontrolled cloud market.
- Existing resource allocation method is static in nature. Advance reservation of resources was used in resource monitoring, but client requests are dynamic in nature and these requests cannot be controlled by static models.
- Maximum cloud service provider companies have used a price model which is based on fixed pricing to allocate the price of their resources. But in real market space these fix price based models do not fit in every situation. Consumer requests are uncertain and dynamic, but the nature of these models is static. So, the resources are not used in an optimal way which is provided by the provider.
- Obtaining of resources was done manually. As there are many cloud service providers, consumers have to go to all the sellers in order to find out the efficient model which is a very heavy task. Maximum work in literature survey has assumed single cloud service provider who fulfills the necessities of clients. So there is demand for scalable and automated method for obtaining resources.
- The work in using an auction based technique in dynamic pricing and allocation of market based takes into account the single type of resources.

- Auction based methods suggested in literature survey does not assume minimum limitation on cloud service provider so as to fulfill clients' demands.
- In this manner, cloud service providers may fraud consumers and get increased revenue by offering low scale services.
- Nowadays, in the literature, fines are compulsory on the cloud service provider for not fulfilling devoted services to the client.

Now, all the previous models have its research gap in between them. Table 3.1 has shown requirements of different cloud providers and their clients. A literature survey of the research in this area based on factors of customers and the cloud service providers is presented in table 3.2. The outlook of cloud providers and the users in several price based models is given in table 3.2. Taking these properties into account the table 3.2 is constructed.

Table 3.1 Different Perspectives of Providers and Consumers of Cloud Computing

Provider's point of view	**Consumer's point of view**
Optimum Resource Utilization	Best cost and running time
Increased benefit and revenues	Price cut and Rewards
Minimization of cost problem	Consumer's demand fulfillment based on QoS

Table 3.2 Comparison between Different Economic Models

Author	**Economic Model**	**Provider's point of view**	**Consumer's point of view**
Buyya [34]	Commodity Market Model	✓	✓
Mihailescu[35]	Market Model	X	✓
Wee [36]	Real Time Pricing	X	✓
Yang [38]	Bidding	X	✓
Wei [42]	Second Priced Auction Model Using Sealed Bid	X	✓
Zaman [44]	Combinatorial Auction Models	✓	X

Wang [46]	Reverse Auction with Batch Making Technique	X	✓
Wang [45]	English Combinatorial Auction	✓	✓
Aida [52]	Double Side Auction Model using K pricing	X	✓
Walki [54]	Commodity Market Model	X	✓
Shang [55]	Double Side Auction Model	X	✓
Son [56]	SLA Based Model	✓	X
Mihailescu[57]	VCG Based Model	✓	✓

3.2 Problem Statement

Optimal allocation of resources is one of the most vital issue in cloud computing. Cloud service providers should be providing resources to allocate in efficient manner otherwise, there could exist the issue of over utilization or under-utilization. In the cloud market multiple providers are there who want to retail their resources. Every seller wants to retail resources to consumer and gets increase revenue. In this situation, it becomes tough for consumers to select the perfect between the multiple cloud service providers. And cloud providers to select the consumer who needs more resources which is also a major task. So the issue is to conclude a suitable cloud service provider for every consumer who could retail several resources in minimum feasible price with respect to his demands and find appropriate consumer for cloud service provider which gives importance to his resources mostly.

The framework should fulfill all the demands of consumers at a specific time. It should also carry out some minimum limitations on the cloud service providers to satisfy demands otherwise fines the cloud provider who disrupts some conditions. Thus, to outline, this thesis provides the benefit to cloud providers as well as cloud consumers with the help of resource allocation algorithm which is market based.

The purpose of the thesis is to offer a way out for allocation of resources issue in a cloud environment via auction technique which is based on an economic model in the market based allocation of resource. The thesis fulfills the demands of cloud consumers who need resources in an ordered way to finish their job within the reasonable price they are presenting. It also allocates resources in an optimal manner of cloud suppliers who are rivalry among themselves, because in cloud market space multiple providers are there. Also cloud suppliers want increase revenue and a pricing technique is designed and fines the cloud providers disrupt indicated conditions.

3.3 Objectives

The objectives of this thesis are explained below:

- To study the current resource allocation methods for optimum resource allocation.
- To design an auction based algorithm to allocate resources in an efficient way.
- To estimate the performance of the proposed algorithm.

Chapter 4

Proposed Work

In the previous chapter the various objectives, gaps and problem statement have been described in this thesis. This chapter describes the proposed method, solution of problem statement and the architecture of the proposed work. It also includes the resource allocation algorithm based on the proposed model.

4.1 Proposed Model

The proposed work is merging the Reverse Auction which is an Auction Technique and Vickrey-Clarke-Groves (VCG) mechanism which provides the resource allocation scheme and payment method on a cloud. After this process the provider provides the resources to particular consumer and the consumer can use the cloud resources.

4.1.1 Architecture of Proposed Model

There are the four components shown in figure 4.1. In this architecture consumer sends a request for the resource to resource allocation module mentioning his service requirements to the providers' side. The cloud service providers send quotations for that requested resources by the consumer to the Resource Allocation Module (RAM). RAM works as intermediate between the consumers and providers which consists of reverse auction technique. This is a form of auction in which job of consumer and provider is reversed and VCG technique is used to calculate the payment of particular resources availed by the cloud provider. RAM initiates a bidding process and call for the bidding values from the providers. All the providers give their quotations to the broker mentioning about their quoted bid value. The broker then checks which provider has sufficient number of resources and lower bid value for its resource. The broker then sorts the list of bidding values given by the provider in ascending or descending order and checks for the lowest bid among all. The provider which has lower bid value and has sufficient resources will then assigned to the consumer. After completion of bidding process the provider provides the resources to particular consumer. If the winner provider is unable to provide requested resource instance, then some amount of fine or penalty will be imposed on that provider.

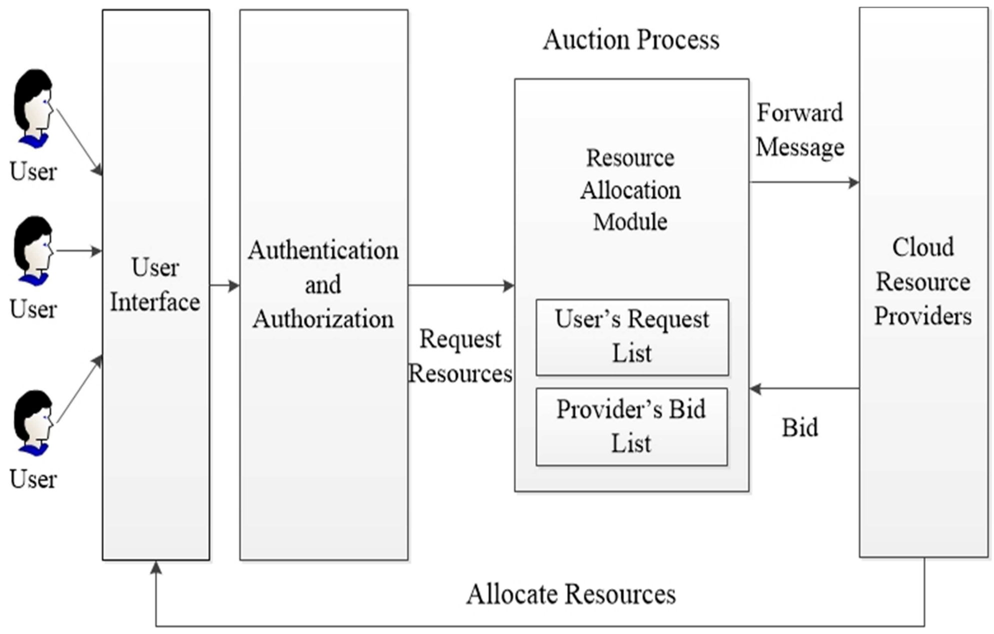


Figure 4.1 Architecture of Proposed Model

The brief explanation of all the components of the proposed model is given below:

- **Consumer** - Consumers are the cloud users who need assets to fulfill their prerequisite needed to finish the job. The cloud users are the users which have their registered account with the available cloud providers. The customer or the end user request for a resource available on the cloud provider, then the authentication of this user is checked by the cloud provider. If the cloud provider verifies that the user is a valid user, then the access is given to him else access denied.
- **Resource Allocation Module** – This component comprises of the three participants cloud user, broker and the cloud provider. This module is responsible for the selection of provider according to the bidding process. It assigns a provider to particular user.
- **Cloud Service Provider** - This component has resources such as storage space (memory), applications, utility software, etc. all are available on the cloud. Cloud provider requires selling these resources to increase revenue. When a broker sends message, then provider will take part in the bidding process. If sufficient quantities of resources are available on a cloud provider, then it can quote prices for its resources. The provider can take part in more than one bidding processes given that it has sufficient resources available.

There is no limit on the number of bidding process the provider can take part in.

4.1.2 Resource Allocation Module (RAM)

This is the main component of proposed model which helps in allocating the resources available on the cloud provider to the customers or cloud users. It consists of main component called a broker as shown in figure 4.2 which communicates with the users and the cloud providers. The main task of the broker is to select an optimal cloud resource for the user. To select one cloud provider from the list, the broker uses bidding process. The bidding process is accomplished using reverse auction in which pricing is calculated using VCG strategy. Once the price has been decided then the broker informs the cloud user which has requested for the resource. In this reverse auction is a category of auction in which supplier and consumer role is reversed. Consumer requests resources and supplier, usually bid for offered resources to consumer. The auction winner offers resources to consumers. The suggested model comprises of multiple cloud providers and multiple consumers who need assets to finish their job. The consumers lodge their necessities to the broker. There are various types of assets existing on cloud providers. After getting requests from consumers' side, broker forward messages to cloud resource providers who start bidding. The whole auction process is managed and controlled by the broker. If sufficient resources are available on a cloud provider, then it can take part in the auction process and quote prices for its resources.

After completion of bidding which is given by providers, broker will start mapping consumers and sellers after computing the lowest bid seller and maximum resource requesting consumer. If the lowest bid seller has sufficient resources that win the auction and then he will be allocated first consumer to provide resources. Otherwise, examine for the next resource seller. But if the auction victor is not capable to fulfill consumer necessity, then he will be fined by broker.

The following scenario explains the process:

In proposed process, there are several cloud providers who holds different type of computational resources and they want to give their resources to consumers according to consumer requirements.

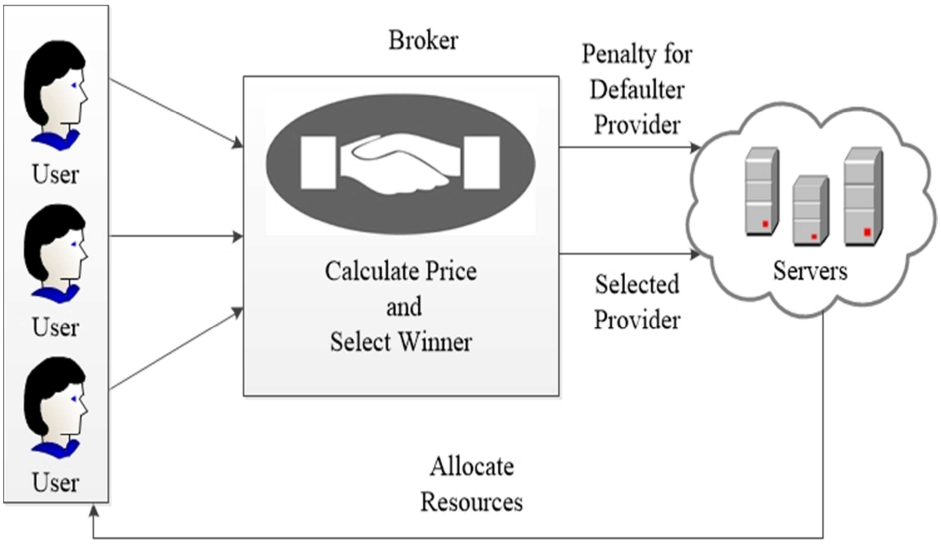


Figure 4.2 Architecture of Resource Allocation Module

The main objective of cloud providers is to take full advantage of their resource utilization and therefore increasing the profit by providers through decreasing the resource cost and a social benefit achieve and in this manner increasing the efficiency of the complete system.

Supposing there is a set of resource providers called N and n ε N who has a large number of computing resources. Cloud resource providers provide computing resources to end users from the pool of different types of virtual machines (VM). Considered, there are three types of VMs available of each cloud resource provider. Small, medium and large are three different types of VMs.

Each VM differs in its memory (ms), CPU power (cp) and storage capacity (st). The capacity of these VMs is pointed out in the table I which is being taken according to the real market scenario. And each type of VM assigned an id.

Table 4.1 Different Types of VMs

ID	VM Type	CPU Power	Memory	Storage
1	Large	40 GHz	32	2 TB
2	Medium	20 GHz	16	1 TB
3	Small	10 GHz	8	500 GB

Following are the variables used in this process:

k	-	Particular id of the VM
P_K	-	Property function of VM
Cp, mr, st	-	Different values for different type of VMs
B_i	-	Resource request issued by i^{th} consumer
$Inst_i$	-	Number of instances
K^{th}	-	Type of virtual machine requested by the i^{th} user
vl_i^k	-	Corresponding value of the resources for the customer

Here,
$$B_i = \{inst_i,\ k,\ vl_i^k\} \quad (1)$$

For example, let there be a user who issues a request $B_i = \{3, 2, 10\}$. It means the user has solicited for 3 instances of type 2 i.e. *Medium* type of VM and he has set the value of the resources as 10 which is the maximum price he is ready to pay. Based on the demand of the user, the broker then initiates a bidding process. After that he calls for the bidding values from the providers who have minimum 3 instances of type 2 VMs. The providers then give its bid values to the broker who acts as an intermediate entity for bidding process. The provider who has quoted a lower price and has 3 instances of type 2 virtual machines gets selected. If no such provider is present, then user's request gets discarded. If the provider does not fulfill the job, a penalty is imposed on the provider.

4.2 Execution Flow of the Proposed Model

The working of the proposed model is as follows:

Step 1 - All the users of cloud and the cloud providers gets them registered with the broker.

Step 2 - The cloud users specify their requirements and send it to the broker. The requirement involves his budget, time duration, and resource specifications.

Step 3 - The broker collects all the data received from cloud clients, and informs all the cloud providers registered on the broker about the user requirements.

Step 4 - The broker starts the bidding process and set a time limit for the auction procedure. All the providers have to quote their prices during that span of time. Only those providers who bid their values within the time will be taken.

Step 5 - After getting the request from the broker, all the providers' checks for the available resources. If the provider has sufficient amount of resources, then it sends the bidding value to the broker for that particular resource. Any provider with no resources available at that time will be removed from the list.

Step 6 - When the time is over, the broker stops the auction.

Step 7 - Broker will sort the bid values received from providers.

Step 8 - Now, the provider who has quoted lowest bid value becomes the winner of the auction. If any provider has lesser number of resources than requested, then it will search for another user with smaller request.

Step 9 - The broker then computes the payment for the requested resource using payment equation specified in VCG algorithm and informs the user.

Step 10 - If the provider has more resources available, then it will compete for another user.

Step 11 - Now the resource is allocated to the user and the utility of the selected provider is calculated.

Step 12 - If the provider is not able to satisfy the users need, then some amount will be imposed on that provider as penalty amount.

Step 13 - The users whose requirements are not satisfied by the providers is assigned another provider with enough number of resources.

4.3 Resource Allocation

Resources are allocated based on auctioning mechanism and payment for all the resources are calculated using VCG mechanism. The whole procedure is explained further in this chapter.

Parameters used are mentioned as follows:

i	-	Represents user where $i=1,2,\ldots..n$
t_i	-	Time taken by user i
c	-	Unit storage cost of a storage node
C	-	Cost variable where $C=ct_i$
V	-	Value for each participant i
B_i	-	Benefits of each participant i
U_i	-	Utility of resources for a specific user i
μ	-	Resource allocation function
n	-	Provider number
N	-	Number of providers
b	-	Bidding variable by provider
x & y	-	Are some variable used for specifying user
b_k	-	Bid value specified for virtual machine k
$V_i(b_i, \mu)$	-	Value for each participant i with respect to bidding value b_i and allocation function μ
$C_i(b_i, \mu)$	-	Cost of the resource with respect to bidding value b_i and allocation function μ
v	-	Value of a Virtual machine for user
l	-	Constant factor
$payment_i(b)$	-	Represent payment function which calculates payment with respect to bid value

μ_{-i} - Value of resource allocation function when i is not participating

4.3.1 Resource Allocation Strategy

Resources are allocated by the cloud provider with the help of VCG strategy. VCG strategy executes in following manner:

1) A resource evaluation method is used which is represented by μ_i where i denotes the user.
2) The objective of this strategy is to enhance the system efficiency and social benefit of the proposed system. It is done by calculating $\mu(b)$ as:

$$\mu^*(b) = \arg\max \sum_i V_i\ (\ b_i, \mu\ ) \tag{2}$$

The equation (2) defines the allocation strategy of the mechanism.

3) Each user, after receiving the resources pays the price equal to the loss he imposes on others due to his presence.

$$payment_i\ (b)\ = \max \sum_{j \neq i} V_i(b_j, \mu_{-i})\ -\ \sum_{j \neq i} V_j(b_j,\ \mu) \tag{3}$$

This rule is the essence of VCG mechanism and it can be implemented according to the resource allocation rule. Only through effective truth revealing strategy, the mechanism of VCG auction is meaningful. The three main properties to be verified in VCG strategy are incentive compatible, individually rational, maximized social welfare.

The main guiding ideology of the VCG based auction mechanism does not vary with application. Hence, the Resource Allocation Strategy for the defined model can be expressed as:

$$\begin{aligned} \mu^*\ (b) &= \arg\max \sum_i V_i(b_i, \mu) \\ &= \arg\max \sum_i (V\ - C_i(b_i, \mu)) \qquad (4) \\ &= \arg\max\ (\mathrm{v}l_i\ - \sum_i C_i(b_i, \mu)) \end{aligned}$$

The value *v* of a VM for the user is fixed. So the above equation can be modified as:

$$\begin{aligned} \mu^*\ (\mathrm{b}) &= \arg\max\ (\mathrm{v}l_i - \sum_i C_i(b_i, \mu)) \\ &= \arg\min \sum_i C_i(b_i, \mu) \qquad (5) \end{aligned}$$

$$= \arg\min \sum_i (b_i \times \mu_i)$$

The allocation of VMs can only be done through the bidding value as the actual cost value of the VM is a private function. The equation (5) shows that the allocation function tried to minimize the total cost of the system. This mechanism tried to allocate resources when the bidding value by resource provider is less than or equal to the user's assumed value, i.e. $b_i \leq v$. VMs which do not meet this criterion will not be selected due to their higher value.

4.3.2 Winner Selection

When the broker requests all the cloud providers for the user's requirements, the providers respond with the bidding value of their resources. After receiving all these values, the broker sorts them in increasing order. The provider who has quoted the lowest value for the requested resource will be selected by the broker. The selected provider is considered as the winner of the auction, which then assigned to the cloud user.

4.3.3 Payment Calculation

After selection of the winner, the broker calculates the payment for the requested resource using a payment function defined by truthful bidding in VCG strategy.

The truthful bidding is described below:

The resource allocation function defined in equation (5) tried to reduce the total cost of the system. But it is not necessary that the bidding value b_i is truthful, which means the bidders can quote false prices. Also, the system's cost will never achieve minimum value. So there is a method of incentive so that a bidder always reveals truth value. Payment function of the VCG mechanism implements this method. The payment function for different VMs can be expressed as follows:

$$payment_i^* (b) = \min \sum_{j \neq i} C_j\left(b_j,\ \mu_{-i}\right) - \sum_{j \neq i} C_j(b_j,\ \mu^*) \qquad (6)$$

In the above equation, $\mu = (\mu_1, \mu_2, \ldots \mu_n)$ represents bidding of VMs by resource provider and μ_{-i} is the allocation when VM i is not participating in the allocation mechanism. The utility of the profit that the resource provider gets when renting VM is the difference between the total payments he receives from the user and the cost of running the VM. So his profit can be expressed as:

If the provider reveals truth value of the VM i.e. $b_i = c_i$

$$U_i(b) = payment_i^*\ (b) - C_i(ct_i,\ \mu^*)$$

$$= \min \sum_{j \neq i} C_j(b_j, \mu_{-i}) - \sum_{j \neq i} C_j\left(b_j,\ \mu^*\right) - C_i(b_i,\ \mu^*)$$

$$= \min \sum_{j \neq i} C_j\left(b_j,\ \mu_{-i}\right) - \sum_j C_j\ (b_j,\ \mu^*) \quad (7)$$

The above equation shows that when the resource provider reveals truth value, the actual profit is the difference between the price of the system when resource provider *i* participates and it does not participate.

If the resource provider chooses to quote false value of the VM i i.e. $b_i \neq c_i$ the profit will be :

$$U_i\ (b) = payment_i^*\ (b) - C_i(c_i, \mu^*)$$

$$= \min \sum_{j \neq i} C_j\left(b_j,\ \mu_{-i}\right) - \sum_{j \neq i} C_j\left(b_j,\ \mu^*\right) - C_i(\ c_i,\ \mu^*)$$

$$= \min \sum_{j \neq i} C_j\left(b_j, \mu_{-i}\right) - \sum_j C_j\left(b_j, \mu^*\right) - \left(C_i(ct_i, \mu^*) - \sum_i C_j(b_j, \mu^*)\right) \quad (8)$$

The equation shows that profit that the provider gets through false bidding is not the actual profit and it is less than actual profit.

4.3.4 Penalty Estimation

Some amount as a penalty is imposed on the cloud provider if it does not fulfill user's requirement as decided while allocation of resources. These are called defaulter providers. The requests assigned to these providers will then transferred to another provider. The new provider will be taken as winner whose price is denoted by P_{new}. The price of the old one is denoted by P_{old}. The penalty P_f equation is as follows:

$$P_f = (P_{new} - P_{old}) \,/ \sum_{i=1}^{n-1} P_{new} - P_{old} \quad (9)$$

4.3.5 Utility Function

After allocating resources to the user, the provider utility is calculated using VCG based allocation algorithm presented in section 4.5. Utility function is based on the bidding value of the user i and value of the resource allocation function. The utility function is expressed as follows:

$$U_i(b) = \min \sum_{j \neq i} C_j\left(b_j,\ \mu_{-i}\right) - \sum_j C_j(b_j, \mu^*) \quad (10)$$

4.3.6 Sequence Diagram of Proposed Model

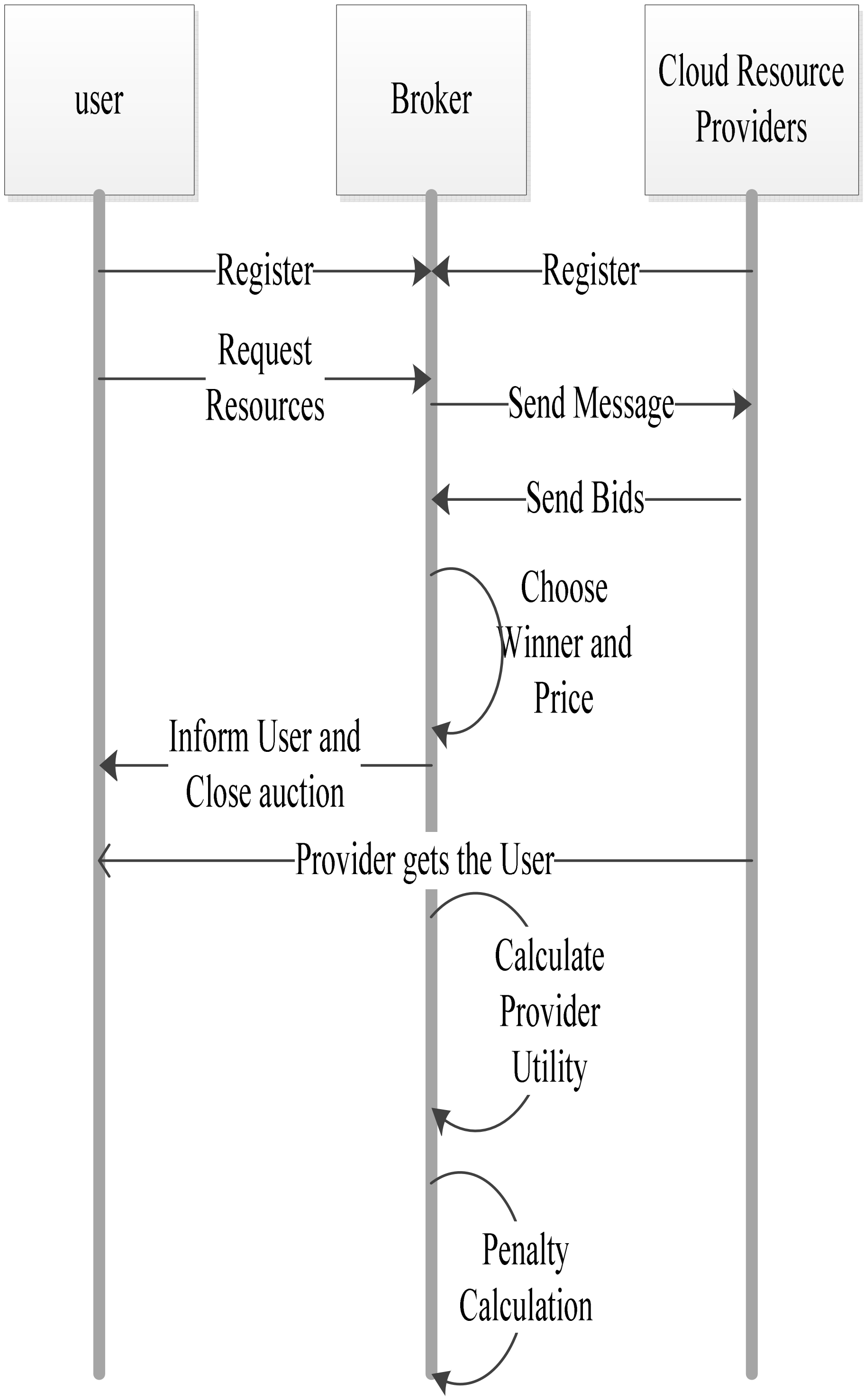


Figure 4.3 Sequence Diagram of Proposed Model

4.4 Flow Chat for Proposed Model

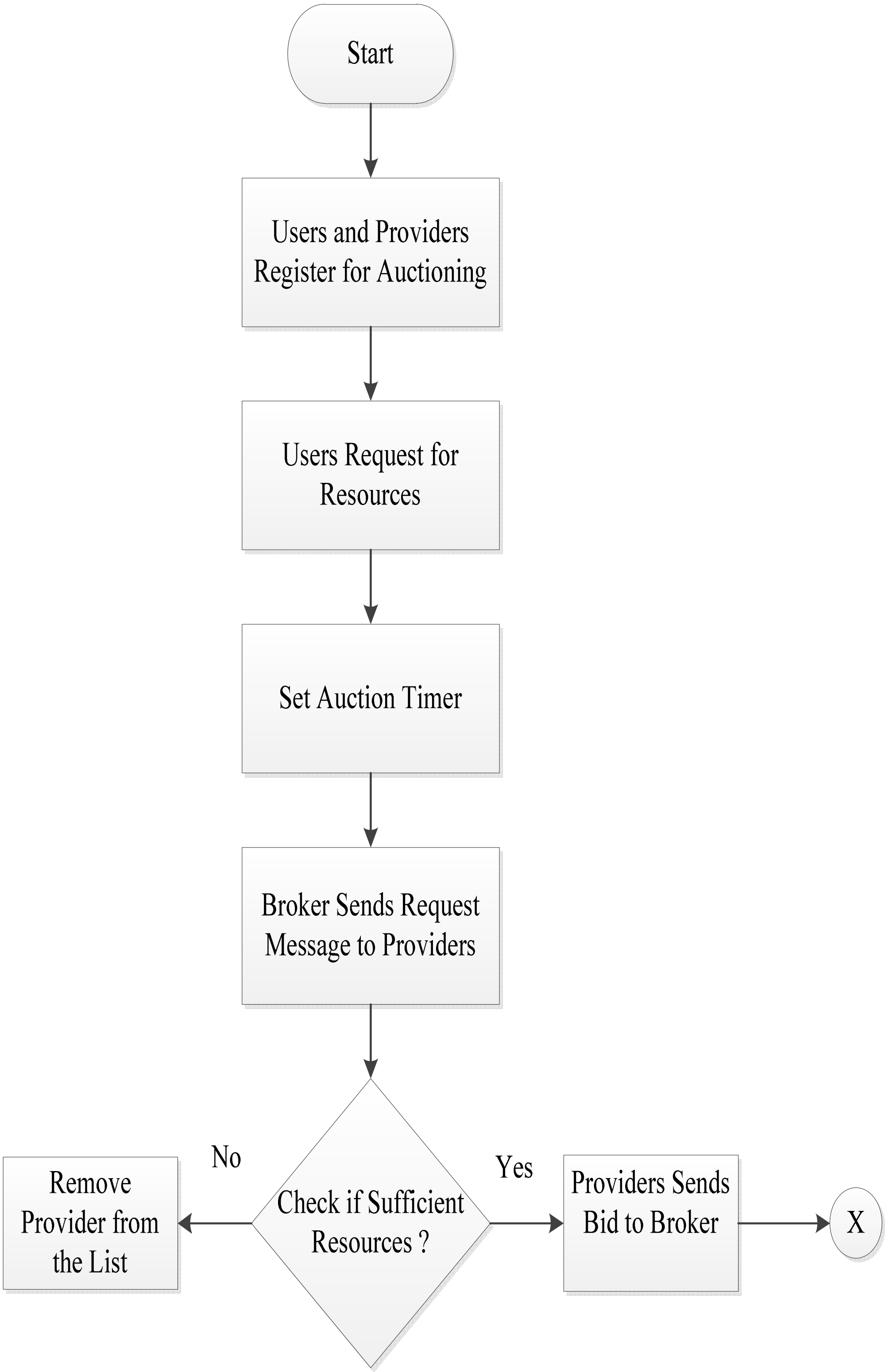


Figure 4.4 Flow Chart for Proposed Model (a)

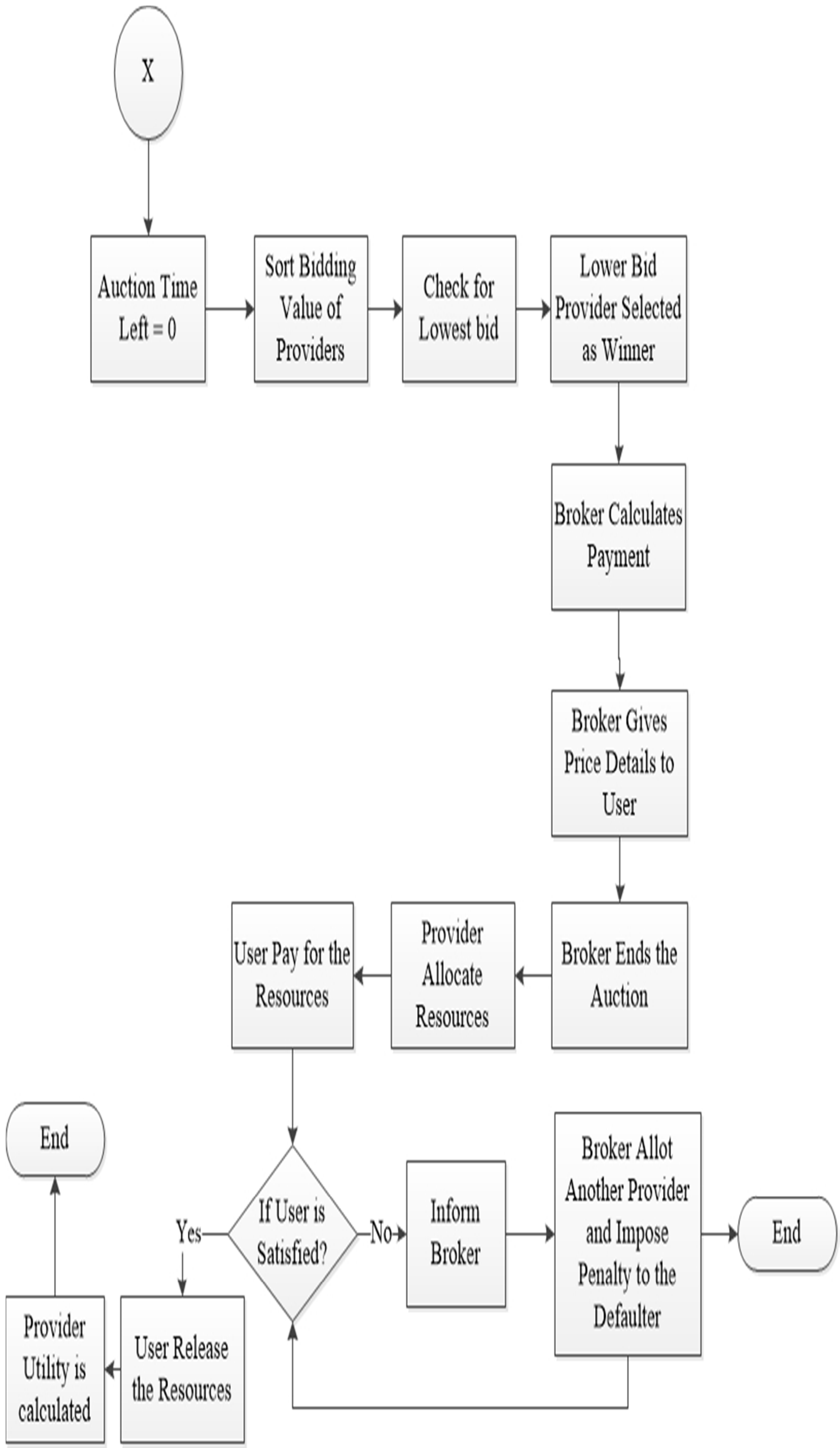


Figure 4.5 Flow Chart for Proposed Model (b)

4.5 Proposed VCG Based Resource Allocation of Virtual Machines

Input: User request *($inst_i$, k, vl_i^k)* to broker

Output: Payment value and winner provider.

1. The user will submit his request to the broker.
2. The broker will announce the start of the auction and VM type *k* to the resource providers. Let there are *n* providers and *N* is the set of providers. *N={1,2,...n}*
3. Cloud providers will send their bids b_i and a maximum number of VMs *max_vm* of *k* type to the broker.
4. Sort the list of cloud provider in ascending order according to their bid values.
5. for *i =1* to *i = n*
6. *If($inst_i$ > max_vm || b_i > vl_i)*
7. Reject the provider
8. End If
9. End for
10. A cloud provider with lowest bid value wins the auction.
11. Calculate payment:

$$payment_i^*(b) = \min \sum_{j \neq i} C_j(b_j, \mu_{-i}) - \sum_{j \neq i} C_j(b_j, \mu^*)$$

12. Calculate utility for winning seller:

$$U_i(b) = \min \sum_{j \neq i} C_j(b_j, \mu_{-i}) - \sum_j C_j(b_j, \mu^*)$$

13. *If($U_i(b)>0$)* then
14. Inform the winning resource provider and allocate resource according to equation (4).
15. *Else*
16. Failure of the auction.
17. End if

Chapter 5

Implementation and Experimental Results

In this chapter, the employment of the proposed model, an auction based algorithm to allocate resources in an efficient way and evaluate the performance of the proposed algorithm.

5.1 Experimental Setup

For simulating this proposed work, a cloud simulator is used to create a virtual cloud environment to get results and analysis. Simulating environment lets us to create cloud resources with different resource requirements and system configuration. The following are the requirements for our proposed model:

- CloudSim
- NetBeans IDE

5.1.1 CloudSim

CloudSim is a platform to generate a simulation environment for cloud computing framework. Following are mostly used the features of CloudSim:

- Simulation of large and small scale data centers on one machine.
- Able to create virtual machines, cloudlets, hosts and other cloud entities with different configuration.
- Able to use Different scheduling policies, topologies, network facilities.
- Able to implement different algorithm in a cloud environment.

5.1.2 NetBeans IDE

It is a platform which is utilized to generate software applications. This platform supports multiple languages C/C++, java, php and HTML, etc. It also has various plugins which is used for performing different operations. There are several libraries already implemented in NetBeans which can be used by developers to do complex tasks really fast. Java is a platform independent language which makes it widely used. CloudSim is a java based tool, so integration of NetBeans with CloudSim became possible.

5.2 Implementation of Proposed Model

The model proposed in this thesis is implemented in CloudSim and NetBeans framework by extending the CloudSim examples available in the toolkit. Firstly, some number of virtual machines is created to be allocated to the cloud users. Similarly, different resources are created in CloudSim and given different configurations. Variables are used to define the number of providers and users. Different ranges of resources are defined such that the bandwidth available from 200-1000 b/s etc. Table 5.1 shows the range of different type of resources.

Table 5.1 Range of Different Types of Resources

Sr. No.	Types of Resource	Range of Resource
1	OS	Unix, Linux, Window7,8,10
2	RAM	[256 - 1024] MB
3	APIs	Java, Android
4	CPU	Core 2duo, i3, i5, i7
5	Storage	512 MB – 1TB
6	Network Bandwidth	[120, 1000] b/s
7	MIPS	[100 – 2000]

5.2.1 Interface

A small demo tool is created to show the results and change in the behaviors of the results. Figure 5.1 depicts the interface of the tool of the proposed model. Firstly, Value of different parameters of the proposed work is entered. Figure 5.1 shows the entered values. The number of user taken is 14, Number of provider are 7 and total requested resource instances are 42 (total of all the requested). It shows 14 users and 7 providers have requested to participate in the auction process.

After taking values from the users and providers, the model has generated a list of user requests and available providers.

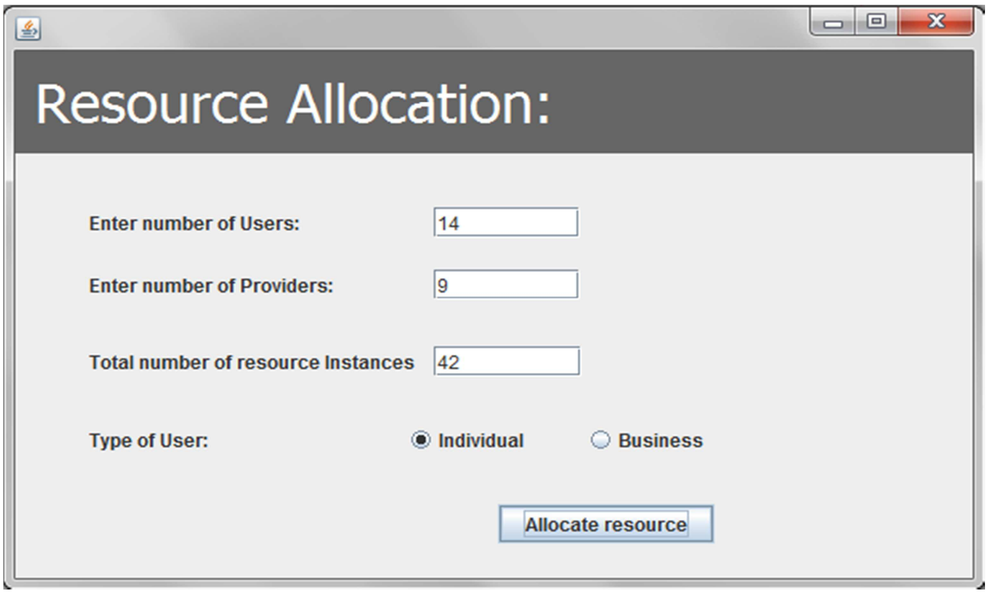
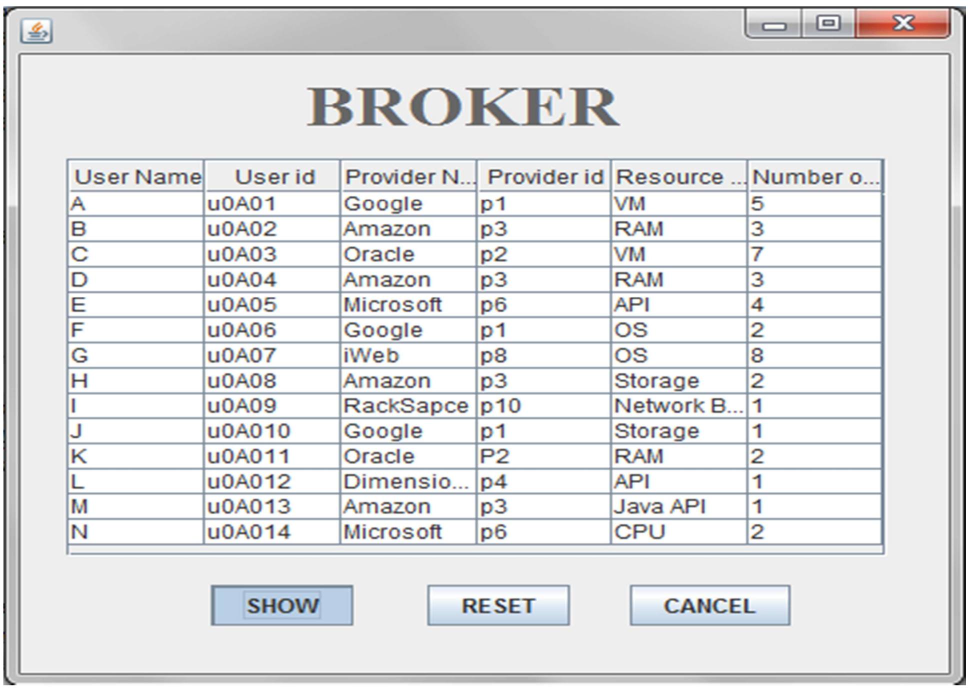


Figure 5.1 Interface of the Tool of Proposed Model

The model sorted them in ascending order of prices. The user selects the lowest bid provider to do their task. Then the providers are allotted different users. Figure 5.2 shows winners of all the requests of different users.

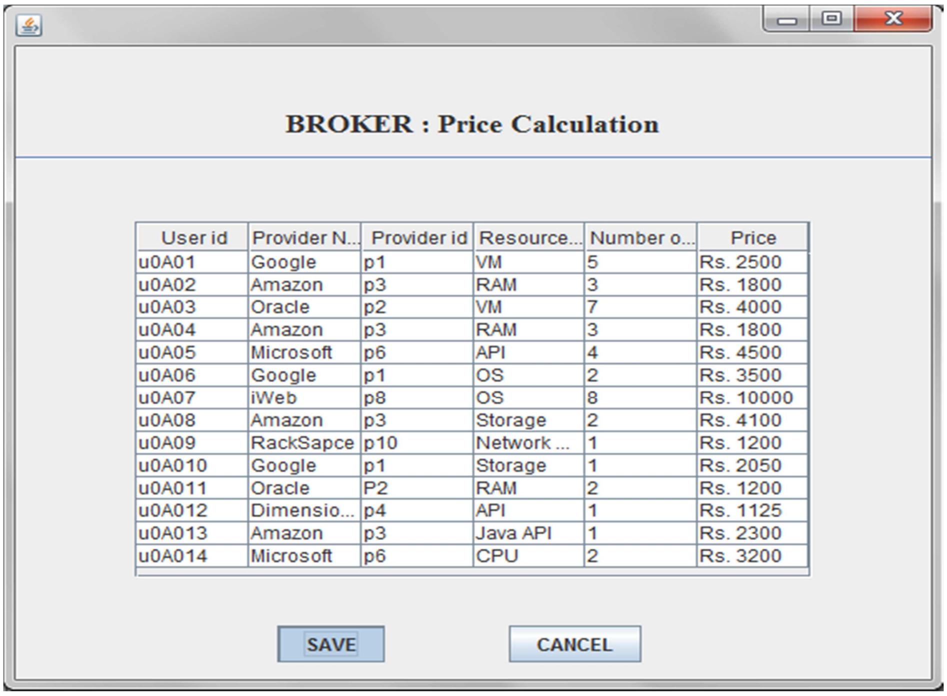


User Name	User id	Provider N...	Provider id	Resource ...	Number o...
A	u0A01	Google	p1	VM	5
B	u0A02	Amazon	p3	RAM	3
C	u0A03	Oracle	p2	VM	7
D	u0A04	Amazon	p3	RAM	3
E	u0A05	Microsoft	p6	API	4
F	u0A06	Google	p1	OS	2
G	u0A07	iWeb	p8	OS	8
H	u0A08	Amazon	p3	Storage	2
I	u0A09	RackSapce	p10	Network B...	1
J	u0A010	Google	p1	Storage	1
K	u0A011	Oracle	P2	RAM	2
L	u0A012	Dimensio...	p4	API	1
M	u0A013	Amazon	p3	Java API	1
N	u0A014	Microsoft	p6	CPU	2

Figure 5.2 Auction Winners

It also shows cloud user id, user name, winner provider id and its name, number of instances. This figure 5.2 shows the allotted providers to different users. Some providers are not available in the list because they did not have enough number of resources available with them at the time of auction. So their participation has been canceled by the broker.

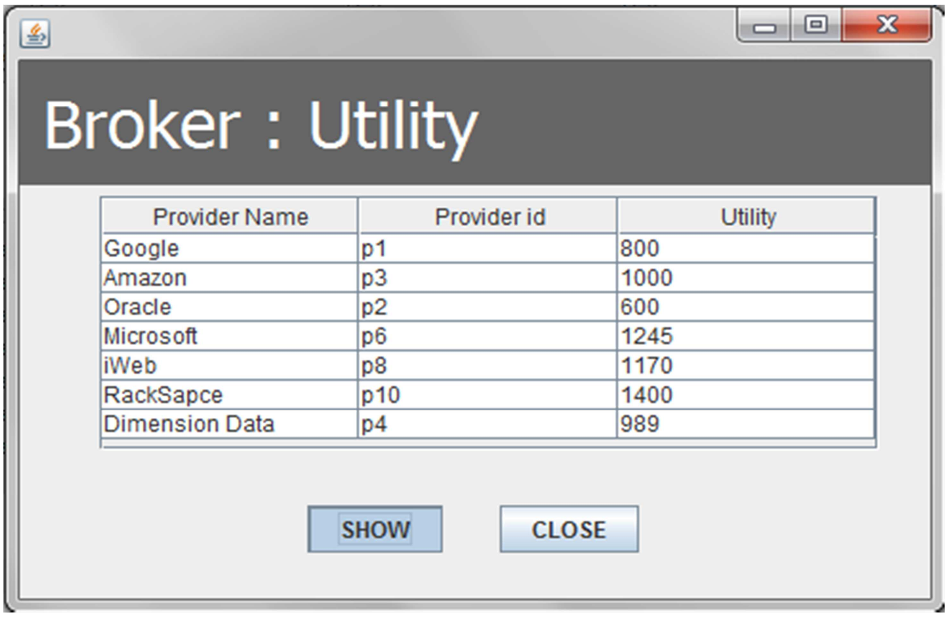
In figure 5.3, the payment of different users is calculated with respect to different allotted providers and the number of instances of a particular resource. The payment is calculated using VCG strategy for different resources.



BROKER : Price Calculation

User id	Provider N...	Provider id	Resource...	Number o...	Price
u0A01	Google	p1	VM	5	Rs. 2500
u0A02	Amazon	p3	RAM	3	Rs. 1800
u0A03	Oracle	p2	VM	7	Rs. 4000
u0A04	Amazon	p3	RAM	3	Rs. 1800
u0A05	Microsoft	p6	API	4	Rs. 4500
u0A06	Google	p1	OS	2	Rs. 3500
u0A07	iWeb	p8	OS	8	Rs. 10000
u0A08	Amazon	p3	Storage	2	Rs. 4100
u0A09	RackSapce	p10	Network ...	1	Rs. 1200
u0A010	Google	p1	Storage	1	Rs. 2050
u0A011	Oracle	P2	RAM	2	Rs. 1200
u0A012	Dimensio...	p4	API	1	Rs. 1125
u0A013	Amazon	p3	Java API	1	Rs. 2300
u0A014	Microsoft	p6	CPU	2	Rs. 3200

Figure 5.3 Price Calculation



Broker : Utility

Provider Name	Provider id	Utility
Google	p1	800
Amazon	p3	1000
Oracle	p2	600
Microsoft	p6	1245
iWeb	p8	1170
RackSapce	p10	1400
Dimension Data	p4	989

Figure 5.4 Utility of Providers

In figure 5.4 the utility of different providers is shown. After every allocation of resources of a particular provider, the utility increases or decreases with respect to reliability of the provider. The utility is a measure of resource usage by various cloud users.

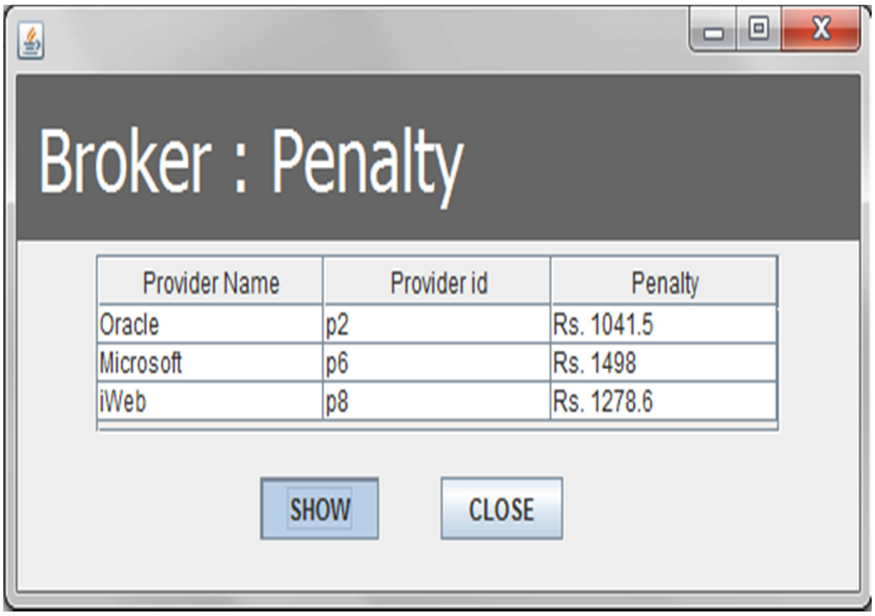


Figure 5.5 Penalties of Providers

After that, the penalty of different providers is calculated if applicable. The penalty of three providers for a particular request is shown in figure 5.5.

5.2.2 Experimental Results

In this segment the results with respect to different user request, system configuration, number of instances of a resource, capabilities of providers is shown.

Test Case 1: Execution time with respect to number of resource instances for fixed price model and VCG model.

In figure 5.6, the graph shows the variation of execution time for different cloud resources for VCG and fixed price model. When a user requests for more number of instances the execution time for model also increases. In this case VCG model takes less time to provide results of auction than the fixed price model.

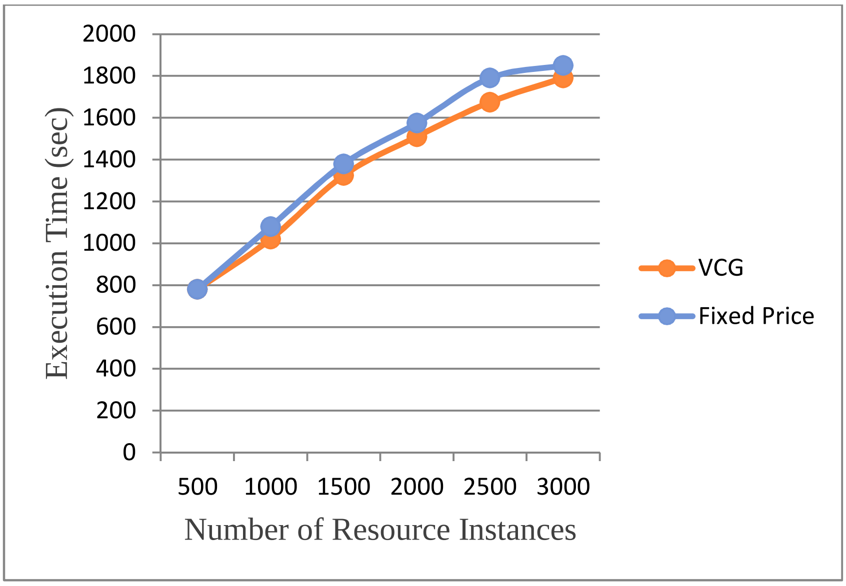


Figure 5.6 Execution Time Comparison w.r.t. number of Resource Instances for Fixed Price Model and VCG Model

Test Case 2: Cost (per unit) comparison with respect to number of resources for fixed price model and VCG model.

Above figure 5.7 shows that the cost (per unit) calculated by the proposed model varies when a large number of resource instances are entered by the broker. VCG model gives better per unit cost for increasing instances than the fixed price model.

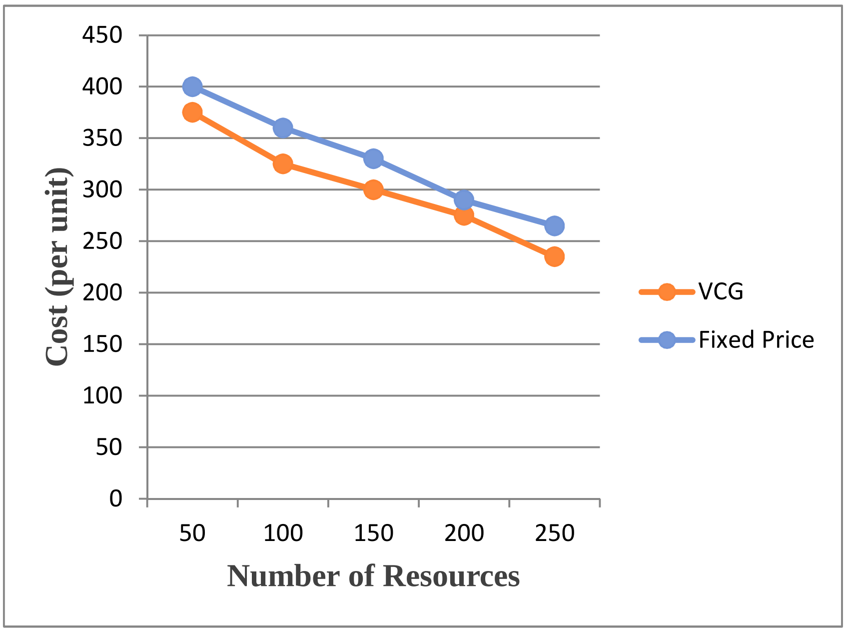


Figure 5.7 Cost (per unit) Comparison w.r.t. Number of Resources for Fixed Price Model and VCG Model

Test Case 3: Cost (per unit) comparison with respect to service time for fixed price model and VCG model.

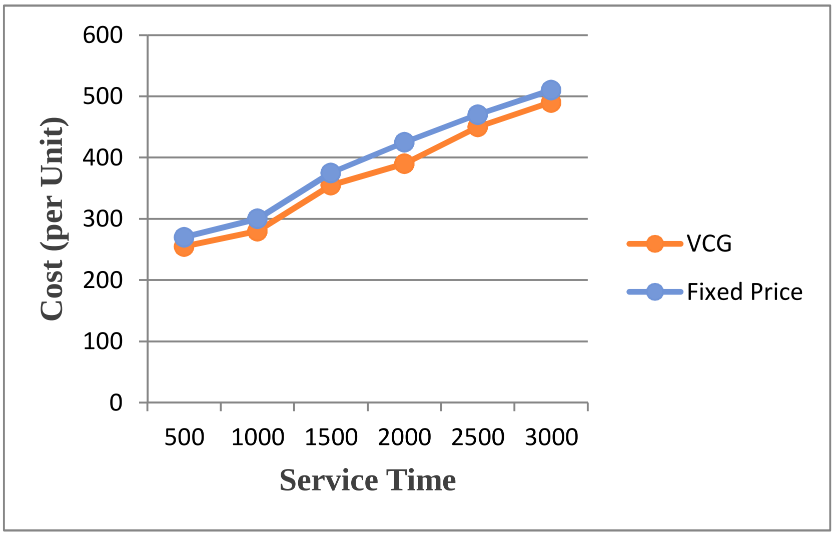


Figure 5.8 Cost (per unit) Comparison w.r.t Service Time for Fixed Price Model and VCG Model

The above figure 5.8 shows the cost variation with respect to service time of the requested resource. When a user requests for a resource for a longer period of time, the cost per unit for that resource also increases. The graph shows that the cost per unit increases slightly when the user needs to use the resource for a month or year.

Test Case 4: Utility comparison of provider for fixed price model and VCG model.

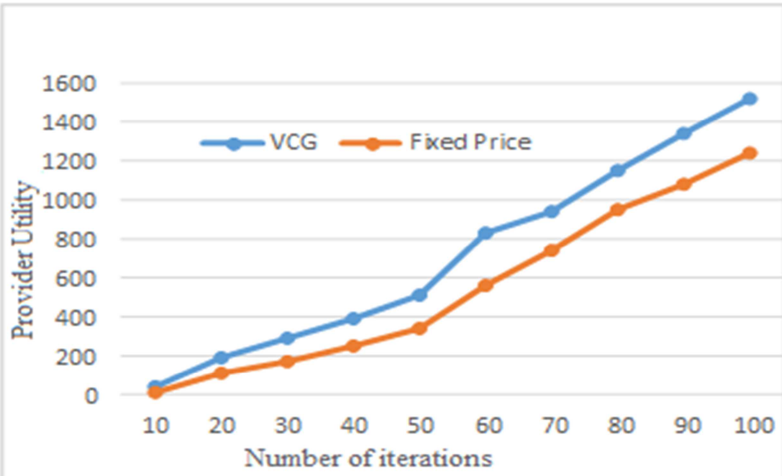


Figure 5.9 Utility Comparison of Provider for Fixed Price Model and VCG Model

In this test case, the utility of different providers calculated by the VCG model is compared with that of the fixed price model. When the VCG model is used to calculate pricing in an auction, the user pays the optimal price. The users also get benefited in this model. Thus, the utility of the provider also increases. The graph shows that the utility of the provider in VCG model is better than the fixed price model.

Test Case 5: Price comparison between fixed price model and VCG based model with respect to each user.

In the figure, the graph shows that the price paid by the cloud users in VCG model is less than that in case of the fixed price model. In the fixed price model, the cost of all the resources is fixed. But, in VCG model, price for different user request is calculated by considering some factors such as time, number of instances, bidding value, and resource allocation function.

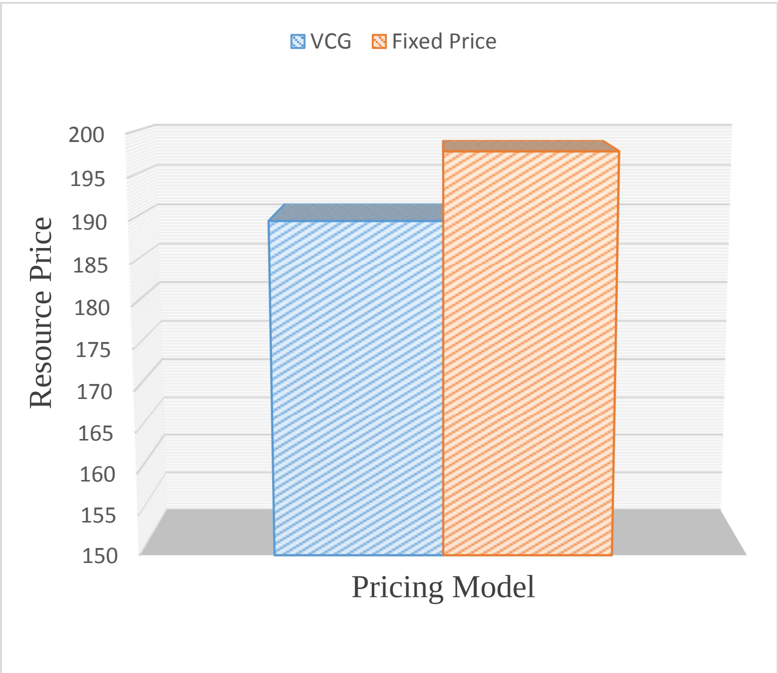


Figure 5.10 Price Comparison between Fixed Price Model and VCG Based Model w.r.t Each User

5.2.3 Results of Different Factors for both Fixed Price and VCG Model

Table 5.2 Execution Time for Fixed Price Model and VCG Model

No. of Resource Instances	VCG	Fixed Price
500	780	780
1000	1020	1080
1500	1325	1380
2000	1510	1575
2500	1675	1790
3000	1790	1850

Table 5.3 Cost w.r.t. Resources for Fixed Price Model and VCG Model

No. of Resources	VCG	Fixed Price
50	375	400
100	325	360
150	300	330
200	275	290
250	235	265

Table 5.4 Cost w.r.t. Service time for Resource for Fixed Price and VCG Model

Service Time	VCG	Fixed Price
500	255	270
1000	280	300
1500	355	375
2000	390	425
2500	450	470
3000	490	510

Table 5.5 Provider Utility w.r.t. Number of Iterations for Fixed Price and VCG Model

Utility	VCG	Fixed Price
10	50	20
20	200	110
30	300	200
40	400	290
50	500	390
60	850	590
70	950	790
80	1160	990
90	1370	1100
100	1550	1250

Chapter 6

Conclusion and Future Scope

6.1 Conclusion

There are various problem occurs in cloud computing, but the most challenging issue in cloud computing is how to allocate the resources. Many scholars have suggested different type of methodologies to solve resource allocation problem. So the problem with the fixed price model and resource allocation technique in the cloud is considered. A VCG auction-based mechanism for better resource allocation on an IaaS layer of cloud is introduced. The mechanism tried to solve resource allocation problem with payment calculation. A broker in the algorithm, based on VCG model, chooses the best winner of the auction, according to the demands of the user and budget mentioned by him. If the utility of the resource provider is greater than zero resources are allocated to the user. The proposed mechanism fulfills properties of incentive compatibility, individual rationality and maximizes social welfare. The main target of the cloud service providers is to maximize the utilization of their resources and thus maximizing the profit through minimizing the cost of the resources and achieve a social benefit and thereby maximizing the efficiency of the whole process. The reverse auction technique has given benefits to both provider and the user. This model is compared with fixed price model and has given a better performance in terms of pricing and utility.

6.2 Thesis Contribution

- In this thesis, literature review of a number of existing economic models and various gaps of existing model in resource allocation algorithms in Cloud Computing has been studied.
- A new VGC algorithm has been introduced to solve resource allocation problem which is based on an auction model. This algorithm has provided the optimal resource allocation to users and benefits to both cloud resource providers and users.
- For pricing of resources, dynamic pricing model is presented in this model.
- This algorithm is compared with the fixed price model. And results have shown that VCG provides best results.

6.3 Future Scope

- In the future, we would like to add different parameters like service initiation time, deadline and QoS parameters demanded by the user.
- A subscription based pricing can be considered in which some concession is given to the user if he is willing to use the service for long duration of time.
- The utility of provider and user can be compared using different models.
- A different strategy can be formulated in which the pricing is calculated based on the type of user i.e. individual user and business entity.

Video Presentation

Video Presentation Link → https://youtu.be/_6Oyt9vRuqU

References

[1] M. H. Weik, (1961). The ENIAC Story [Online], Available: http://ftp.arl.mil/mike/comphist/eniac-story.html

[2] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, “Cloud computing and emerging IT platforms: Vision, hype and reality for delivering computing as the 5th utility,” *Future Generation Computer Systems*, vol. 25(6), pp.599-616, 2009.

[3] G.F. Pfister, In Search of Clusters, 2nd ed., Prentice Hall, Upper Saddle River, USA, 1998.

[4] “Cloud Computing Technology- the study of its model,” [online] Available: www.websiteglobe.blogspot.in/2012/05/Cloud-computing-technologystudyofits.html. [Mar. 5, 2014].

[5] M. Ahmed, A. Chowdhury, and M. H. Rafee, “An advanced survey on cloud computing and state-of-the-art research issues”, *International Journal of Computer Science Issues*, vol. 9(1), pp. 201-2017, 2012.

[6] V. R. Srinivasa, N. K. R. Nageswara, and E. K. Kumari, “Cloud computing: an overview”, *Journal of Theoretical and Applied Information Technology*, vol. 9(1), pp. 71-76, 2009.

[7] K. Raghavendra, A. Akilan, N. Ravi, K. P. Kumar, and G. Varadan, “Satellite Data product Generation using Aneka Cloud,*” in Research Demo at 10th IEEE International Symposium on Cluster, Cloud and Grid Computing, CCGrid*, Melbourne, Australia, 2010.

[8] R. P. Padhy, M. R.Patra, and S. C. Satapathy, “Cloud computing: security issues and research challenges”, *International Journal of Computer Science and Information Technology & Security (IJCSITS)*, vol. 1(2), pp. 136-146, 2011.

[9] M. Nazir, “Cloud computing: overview & current research challenges,” *Journal of Computer Engineering (IOSR-JCE)*, vol. 8(1), pp. 14-22, 2012.

[10] S. S. Manvi and G. K. Shyam, “Resource management for Infrastructure as a Service (IaaS) in cloud computing: A survey,” *Journal of Network and Computer Application*, vol. 41, pp. 424-440, 2014.

[11] B. Urgaonkar, P. Shenoy, A. Chandra, and P. Goyal, “Dynamic provisioning of multi-tier internet applications,” *In Proceedings of Second International Conference on Autonomic Computing, ICAC*, , IEEE, pp. 217-228, 2005.

[12] J.S. Chase, D.C. Anderson, P.N. Thakar, and A.M. Vahdat, “Managing energy and server resources in hosting center,” *In proceedings of 11th IEEE/ACM international conference on grid computing*, vol .12(4), pp. 50-2, 2010.

[13] K. Krauter, R. Buyya, and M. Maheswaran, “A taxonomy and survey of grid resource management systems for distributed computing,” *Software: Practice and Experience*, vol. 32(2), pp. 135–164. 2002.

[14] G. Aceto, A. Botta, W. De. Donato, and A. Pescape, “Cloud monitoring: A survey,” *Computer Networks*, vol. 57(9), pp. 2093–2115, 2013.

[15] S.T. Selvi, C. Valliyammai, and V.N. Dhatchayan, “Resource Allocation Issues and Challenges in Cloud Computing,” *In proceedings of 2014 International Conference on Recent Trends in Information Technology*, 2014**.**

[16] V. Awasare and S. Deshmukh, “Survey and Comparative Study on Resource Allocation Strategies in Cloud Computing Environment,” *IOSR Journal of Computer Engineering*, vol. 16(2), pp. 94-101, 2014.

[17] V.Vinothina, R. Sridaran, and P. Ganapathi, “A Survey on Resource Allocation Strategies in Cloud Computing,” *International Journal of Advanced Computer Science and Applications*, vol. 3(6), pp. 97-104, 2012.

[18] J. Li, M. Qiu, J.W. Niu, Y. Chen, and Z. Ming, “Adaptive resource allocation for preemptable jobs in cloud systems,” *In proceedings of 10th International Conference Intelligent Systems Design and Applications* , IEEE, pp. 31-36, 2010

[19] J.O. Melendez and S. Majumdar, “Matchmaking with Limited knowledge of Resources on Clouds and Grids,” *In International Symposium Performance*

Evaluation of Computer and Telecommunication Systems (SPECTS), IEEE, pp. 102-110, 2010.

[20] S. Majumdar, "Resource Management on cloud : Handling uncertainties in Parameters and Policies," *CSI communications*, pp.16-19, 2011.

[21] D. Shin and H. Akkan, "Domain- based virtualized resource management in cloud computing," *In proceedings of 6th International Conference of Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom*), IEEE, pp. 1-6, 2010.

[22] Z. Kong, C.Z. Xu, and M. Guo , "Mechanism Design for Stochastic Virtual Resource Allocation in Non-Cooperative Cloud System," *In proceedings of 4th International Conference on Cloud Computing*, pp. 614-621, 2011.

[23] F. Wuhib and R. Stadler, " Distributed monitoring and resource management for Large cloud environments," *In International Symposium on Integrated Network Management (IM)*, IFIP/IEEE, pp. 970-975, 2011.

[24] D. Niyato, Z. Kun, and P. Wang, "Cooperative Virtual Machine Management for Multi-Organization Cloud Computing Environment," *In Proceedings of the 5th International ICST Conference on Performance Evaluation Methodologies and Tools ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering)*, pp. 528-537, 2011.

[25] D. Gmach, J. Roliaand, and L. cherkasova, "Satisfying service level objectives in a self-managing resource pool," *In Proceedings of Third international conference on self-adaptive and self-organizing system*, IEEE, pp. 243-253, 2009.

[26] X. Zhu, D. Young, B.J. Watson, Z. Wang, J. Rolia, S. Singhal, B. McKee, C. Hyser, D. Gmach, R. Gardner, and T. Christian, "Integrated capacity and workload management for the next generation data center," *In proceedings of 5th international conference on Automatic computing(ICAC'08)*, pp. 172-181, 2008.

[27] D. Minarolli and B. Freisleben, "Uitlity –based Resource Allocations for virtual machines in cloud computing," *In Symposium of Computers and Communications (ISCC)*, IEEE, pp. 410-417, 2011

[28] W. Hu, C. Tian, X. Liu, H. Qi, L. Zha, H. Liao, Y. Zhang, and J. Zhang, "Multiple Job Optimization in MapReduce for Heterogeneous Workloads," *In proceedings of 6th International Conference on Semantics, Knowledge and Grids* , IEEE, pp. 135-140. 2010.

[29] Wei-Yu Lin, "Dynamic Auction Mechanism for Cloud Resource Allocation," *In proceedings of 10th International Conference on Cluster, Cloud and Grid Computing*, IEEE/ACM , pp.591-592, 2010

[30] T. T. Huu and J. Montagnat, "Virtual Resource Allocations for workflow-based applications distribution on a cloud infrastructure," *In proceedings of International Conference on Cluster, Cloud and Grid Computing*, IEEE/ACM, pp.612-617, 2010

[31] W. Iqbal, M.N. Dailey, I. Ali, P. Janecek and D. Carrera, "Adaptive Resource Allocation for Back-end Mashup Applications on a heterogeneous private cloud," *In proceedings of International Conference Electrical Engineering/Electronics Computer Telecommunications and Information Technology (ECTI-CON)*, IEEE, pp. 317-321, 2010.

[32] Y.C. Lee, "Project driven service request scheduling in clouds," *In proceedings of International symposium on cluster & Grid Computing, (CC Grid)*,Melbourne, Australia, 2010.

[33] R.T. Ma, D.M. Chiu, J.C. Lui, V. Misra, and D. Rubenstein, "On Resource Management for Cloud users :a Generalized Kelly Mechanism Approach," *Electrical Engineering, Tech. Rep.*, 2010.

[34] B. Sharma, R. K. Thulasiram, P. Thulasiraman, S. K. Garg, and R. Buyya, "Pricing Cloud Compute Commodities: A Novel Financial Economic Model," *In Proceedings of 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*, IEEE Computer Society, pp. 451-457, 2012.

[35] M. Mihailescu and Y.M. Teo, “Dynamic resource pricing on federated clouds,” *In proceedings of International Symposium on Cluster Computing and the Grid*, IEEE Computer Society, pp. 513–517, 2010.

[36] S. Wee, “Debunking real-time pricing in cloud computing,” *In proceedings of International Symposium on Cluster, Cloud and Grid Computing (CCGrid),* IEEE/ACM, pp. 585-590, 2011.

[37] H. Li, J. Liu, and G. Tang, “A Pricing Algorithm for Cloud Computing Resources,” *In proceedings of International Conference on Network Computing and Information Security,* IEEE, vol. 1, pp. 69-73, 2011.

[38] J. Yang, S. Yang, M. Li, and Q. Fu, "An Autonomous Pricing Strategy toward Market Economy in Computational Grids," *In Proceedings of International Conference on Information Technology: Coding and Computing (ITCC),* IEEE , vol. 2, pp. 793-794, 2005.

[39] Y.M. chen and H.M. Yeh , “Autonomous adaptive agents for market-based resource allocation of cloud computing,” *In Proceedings of 9th International Conference on Machine Learning and Cybernetics,* vol. 6, pp. 2760-2764, 2010.

[40] Y. Zhang, B. Li, Z. Huang, J. Wang, J. Zhu, and H. Peng, “Strategy-Proof Auction Mechanism with Group Price for Virtual Machine Allocation in Clouds,” *In proceedings of Second International Conference in Advanced Cloud and Big Data*, IEEE, pp. 60-68, 2014.

[41] Y. Ma, B. Li, Y. Zhang, and J. Zhu, “Efficient Auction Mechanism with Group Price for Resource Allocation in Clouds,” *In proceedings of Second International Conference on Advanced Cloud and Big Data,* IEEE, pp. 85-92, 2014.

[42] W.Y. Lin, G.Y. Lin, and H.Y. Wei, “Dynamic Auction Mechanism for Cloud Resource Allocation,” *In proceedings of 10th International Conference on Cluster, Cloud and Grid Computing, IEEE/*ACM, pp. 591-592, 2010.

[43] G. Vinu Prasad, S. Rao, and A.S. Prasad, “ A combinatorial auction mechanism for multiple resource procurement in cloud computing,” *In*

proceedings of 12th International Conference in Intelligent Systems Design and Applications (ISDA), IEEE, pp. 337-344, 2012.

[44] S. Zaman and D. Grosu,"A Combinatorial Auction-Based Mechanism for Dynamic VM Provisioning and Allocation in Clouds," *Transactions on cloud computing*, IEEE, vol. 1(2), pp. 129-141, 2013.

[45] W. Xing-wei, W. Xue-yi, and H. Min, "A resource allocation method based on the limited English combinatorial auction under cloud computing environment" *In proceedings of 9th International Conference on Fuzzy Systems and Knowledge Discovery*, IEEE, pp. 905-909, 2012.

[46] X. Wang, J. Sun, M. Huang, C. Wu, and X. Wang, "A resource auction based allocation mechanism in the cloud computing environment," *In proceedings of 26th International Parallel and Distributed Processing Symposium Workshops and PhD Forum*, IEEE, pp. 2111-2115, 2012.

[47] P. Bonacquisto, G. Di Modica, G. Petralia, and O. Tomarchio, "A strategy to optimize resource allocation in auction-based cloud markets," *In proceedings of International Conference on Services Computing*, IEEE, pp. 339-346, 2014.

[48] A.H. Ozer and C. Ozturan, "An auction based mathematical model and heuristics for resource co-allocation problem in grids and clouds," *In proceedings of fifth international conference in Soft Computing, computing with words and Perceptions in System Analysis, Design and Control*, pp. 1-4, 2009.

[49] W. Wang, B. Liang, and B. Li, "Revenue Maximization with Dynamic Auctions in IaaS Cloud Markets," *In proceedings of 21st International Symposium in Quality of Service*, IEEE/ACM, pp. 1-6, 2013.

[50] X. YOU, X. XU, J. Wan, and D. YU, " RAS-M:Resource Allocation Strategy based on Market Mechanism in Cloud Computing," *In proceedings of Fourth ChinaGrid Annual Conference*, IEEE, pp. 256-263, 2009.

[51] J. Zhao, X. Chu, H. Liu, Y.W. Leung, and Z. li, "Online Procurement Auctions for Resource Pooling in Client-Assisted Cloud Storage Systems," *In*

proceedings of Conference in Computer Communications, IEEE, pp. 576-584, 2015.

[52] K. aida and I. Fujiwara, "Applying Double-sided Combinational Auctions to Resource Allocation in Cloud Computing;" *In proceedings of 10th Annual International Symposium on Applications and the Internet*, IEEE/IPSJ, pp. 7-14, 2010.

[53] Y. Lan, W. Tong, Z. Liu, and Y. Hou, "Multi-Unit Continuous Double Auction Based Resource Allocation Method," *In proceedings of third International Conference on Intelligent Control and Information Processing*, pp. 773-777 , 2012.

[54] S. A. Bello and G. A. Wakil, "Flexible Pricing Models for Cloud Services," *In Transactions on Networks and Communications*, vol. 2(5), pp. 15-28, 2014.

[55] S. Shang, J. Jiang, Y. Wu, Z. Huang, G. Yang, and W. Zheng, "DABGPM: A Double Auction Bayesian Game-Based Pricing Model in Cloud Market," *In Network and Parallel Computing*, pp. 155-164, 2010.

[56] S. Son, G. Jung, and S. Jun, "An SLA-based cloud computing that facilitates resource allocation in the distributed data centers of a cloud provider," *J. Supercomput.*, vol. 64, pp. 606–637, 2013

[57] T. Yong Meng and M. Mihailescu, "A strategy-proof pricing scheme for multiple resource type allocations," *In proceedings of International Conference on Parallel Processing (ICPP)*, pp. 172–179, 2009.

List of Publication

Anita Kumari and Dr. Sushma Jain, “Auction based Resource Allocation Strategy for Infrastructure as a Service”, 2016, International Conference on Computational Techniques in Information and Communication Technologies (ICCTICT), IEEE Explore,2016, **[To be published]**