REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE MOHAMED BOUDIAF - M'SILA

FACULTE DES MATHEMATIQUES ET INFORMATIQUE

DEPARTEMENT D'INFORMATIQUE

N° :………………………………………..

DOMAINE : MATHEMATIQUES ET INFORMATIQUE.

FILIERE : INFORMATIQUE

OPTION : SYSTEMES D'INFORMATIONS AVANCES.

Mémoire présenté pour l'obtention
Du diplôme de Master Académique

Par: BOUDIAF HAMZA

Intitulé

PLACEMENT DES MACHINES VIRTUELLES DANS LES ENVIRONNEMENTS CLOUD ET LES DATA CENTERS VIRTUALISES

Soutenu devant le jury composé de : 10/06/2017.

BOURAHLA MUSTAPHA	Université de M'sila	Président
SAYAD LAMRI	Université de M'sila	Rapporteur
GARNA ABDERRAHIM	Université de M'sila	Examinateur

Année universitaire : 2016 /2017

I would like to lovingly dedicate this dissertation to my parents and all my family including my wife and my best friends.

A C K N O W L E D G M E N T S

This dissertation would not have been successfully completed without the contributions of many people. First of all, I would like to thank my thesis advisor,Dr. SayadLamri, who has given me the opportunity to undertake a Master andprovided me with invaluable guidance and advice throughout my Master training. I would like to express mygratitude to all my Master committee members, for their constructive comments and suggestionson improving my work.

TABLE OF CONTENTS

DEDICATION I

ACKNOWLEDGMENTS II

LIST OF FIGURES III

LIST OF TABLES .... *IV*

LIST OF ABBREVIATIONS IIV

Table of Contents V

Introduction ....

CHAPTER IData Center Environment 12

1 Introduction 13

2 Data Center Architecture 13

2.1 Data Center Architecture 13

2.2 Data Center Energy 15

3 Clouds Computing 16

3.1 The Main Benefits Of Cloud Computing 17

3.2 Cloud Essential Characteristics 17

3.3 Cloud Service Models 19

3.4 Cloud Deployment Models 20

4 Virtualization 23

4.1 Need's of Virtualization 23

4.2 Types of virtualization 23

5 Conclusion 24

CHAPTER 2 Energy-Efficient VM Placement first fit decrease Algorithm for Cloud Data Center 25

1 Introduction 26

2 Relatedwork 27

3 Virtual machine placement Algorithms. 28

4 Classifications of VM Placement Algorithms 28

4.1 Static VM Placement 28

4.2 Dynamic VM Placement 28

5 Virtual machine placement problem in cloud 29

6 Virtual machine placement optimization 30

7 Metaheuristic approach 30

8 System Model 30

9 Energy Model 32

10 Conclusion 34

CHAPTER 3 implementation and Test 36

1 Introduction 37

2 Development environment 37
2.1 Operating system 37
2.2 Microsoft Visual Studio 2012 37
3 Experiments 38
3.1 Test 39
3.2 Results 40
4 Interfaces of the application 41
4.1 Home page 41
4.2 Creation of servers 41
4.3 Apply the FFD algorithm 42
5 Interfaces of code source 43
5.1 Affect method 43
5.2 First Fit Decrease Algorithm applied 43
6 Conclusion 44
Global conclusion 45
Bibliography 46

LIST OF TABLES

2.1 Table2.1 Energy-efficient virtual machine placement (EVMP) 34

3.1 Table 3.1 Types of heterogeneous physical machines 39

3.2 Table 3.2 Types of virtual machines 39

LIST OF FIGURES

1.1 Fat-tree topology 14
1.2 VL2 topology 15
1.3 Energy consumption at different levels in computing systems 16
1.4 NIST Cloud Computing Definitions 16
1.5 On-demand self-service..17
1.6 Broad network access...18
1.7 Cloud resource pooling...19
1.8 Cloud service models..20
1.9 Private cloud 21
1.10 Community cloud 21
1.11 Public cloud...22
1.12 Hybrid cloud..22
2.1 Classifications of VM Placement...29
2.2 Hosted Virtualization 29
2.3 The IAAS system model 31
2.4 The system model...32
3.1 First Fit Decrease algorithm results..40
3.2 First Fit algorithm results..40
3.3 Home page..41
3.4 Creation of servers 41
3.5 VM first fit decrease representation 42
3.6 VM first fit decrease results...42
3.7 Affect method...43
3.8 First Fit Decrease Algorithm applied 43

LIST OF ABBREVIATIONS

DC	DATA CENTER
VM	VIRTUAL MACHINE
VMs	VIRTUAL MACHINE SET
PM	PHYSICAL MACHINE
PMs	PHYSICAL MACHINE SET
(SaaS)	SOFTWARE AS A SERVICE
(PaaS)	PLATFORM AS A SERVICE
(IaaS)	INFRASTRACTUREAS A SERVICE
CPU	CENTRAL PROCESSING UNIT
MEM	MEMORY
NET	NETWORK BANDWIDTH
CPU_i^u	AVERAGE CPU UTILISATION OF A PHYSICAL SERVER i
MEM_i^u	AVERAGE MEM UTILISATION OF A PHYSICAL SERVER i
NET_i^u	AVERAGE NET UTILISATION OF A PHYSICAL SERVER i
IR_i^u	THE OVERALL AVERAGE UTILISATION
SM_i	THE BALANCED UTILISATION AND ENERGY CONSUMPTION
TCO	TOTAL COST OF OWNERSHIP
UPS	UNINTERRUPTIBLE POWER SUPPLIES
PDU	POWER DISTRIBUTION UNITS

INTRODUCTION

Data center is a place gathers a hundreds of thousands of servers, interconnected via switches, routers and high-speed links. Cloud computing offers a variety of resources such as CPU and storage, systems software, and applications as services anytime and anywhere over the Internet. To reduce the number of physical machines (PM) in datacenters virtualization technologies are widely used, it allows users to run different applications through virtual machines (VM) and make it serves as a basic unit to perform various tasks in the same machine. Although virtualization technologies can improve energy consumption by increasing PM's utilization, energy cost of datacenters is still rapidly growing up. Today, large companies such as Amazon, Google, Facebook, and Yahoo! routinely use data centers for storage, web search, and large-scale computations. According to [1], in 2006, the energy consumption of datacenters in U.S. was 61 billion kilowatt-hour (kWh), and in 2013, U.S. data centers consumed an estimated 91 billion kilowatt-hours of electricity, equivalent to the annual output of 34 large (500-megawatt) coal-fired power plants. The annual electricity consumption of data centers is projected to increase to roughly 140 billion kilowatt hours by 2020, the equivalent annual output of 50 power plants, costing American businesses $13 billion annually in electricity bills and emitting nearly 100 million metric tons of carbon pollution per year [2]. As a result, improving energy allows of no delay to cloud data centers.

Industry and academia are attracted by the power costs of cloud data centers; they focus on the biggest power consumers (i.e., computer servers and cooling systems), yet without taking the networking part into consideration cause the network elements consume 10-20 % of the total power in data center revealed by recent studies.

The virtual machine placement defined as the process of choosing the foremost appropriate host for each virtual machine request. In our work, we use a virtual machines placement algorithm for cloud data center, considering the multi-dimensional resource constraints, such as CPU, memory, network bandwidth, and so on. We focus on virtual machine placement problem, the initial static allocation algorithm proposed.

The dissertation organized as follows, the first chapter presents general knowledge of the system environment like data center, cloud and virtualization technology ; the second chapter represents virtual machine placement, the assignment of the virtual machine in the system energy environment, the approach chosen and the related study. The last chapter focuses on the tools used to develop the project and shows results.

CHAPTER 1

DATA CENTER ENVIRONMENT

CHAPITRE 1

DATA CENTER ENVIRONMENT

1 Introduction

In recent years, the power costs of cloud data centers attracted significant attention from both industry and academia. They focus their early works on data center energy exactly on the biggest power consumers. The world of data center divides to two parts the first one is physical designing it focus on where we place servers, routers, switches, cables and how we assure cooling. The second one is technology of data centers and how built its topology (it is our work domain attention). In this chapter, we try to cover all the needs to built data center topology and assure the access to it.

2 Data Center Architecture

2.1 Data Center Architecture

The switch-centric and the server-centric are mainly two class's categorized data centers. Switches are the dominant components for interconnection in switch-centric data center, while in server-center data center, servers with multiple Network Interface Cards (NIC) take part in routing and packet forwarding decisions. Switch-centric data centers include VL2 [13], Portland [21], Fat-Tree [16] .Switch-centric data centers include Dcell [16], Bcube [14].

The fat-tree topology, depicted in Figure (1.1), consists of k pods, each of which consisting of k/2 edge switches and k/2 aggregation switches. Edge and aggregation switches connected as a close topology and form a complete bipartite in each pod. Also each pod is connected to all core switches forming another bipartite graph. Fat-Tree built with k-port identical switches in all layers of the topology and each of which supports k3/4 hosts. Fat-tree IP addresses are in the form 10: pod: subnet: hosted. In fat-tree, the address lookup is implemented by a two table lookup approach instead of the longest prefix matching. Address lookups are done in two steps; first, the lookup engine does a lookup on the first level table to find the longest matching prefix. Then the matched address is used to index the second level table which holds the information of the IP address and output port to reach the intended destination. [14]

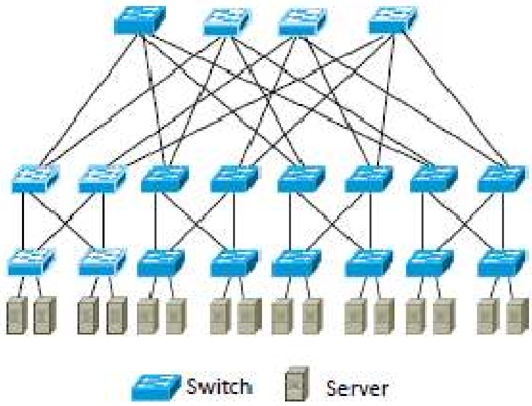

Figure 1.1 Fat-tree topology [14]

VL2 was proposed in [12] and considered as a solution to overcome some of the critical issues in conventional data centers such as over-subscription, agility and fault tolerance.VL2 supports VM migration from server to server without breaking the TCP connection and keeping the same address. VL2 implements a clos topology between core and aggregation layers to provide multipath and rich connectivity between the two top tiers. VL2 employs Valiant Load Balancing (VLB) to evenly load balance traffic flows over the paths using Equal Cost Multi Path (ECMP).

Portland was proposed in [21] whose DCN topology is based on a fat-tree network topology. Portland consists of three layers: edge, aggregation and core. It is built out of low cost commodity switches. Portland and fat-tree both differ in the addressing scheme for packet routing but both at the end aim at providing agility among services running on multiple machines. Both reduce broadcast by intercepting Address Resolution Protocol (ARP) requests and employ a unicast query through a centralized lookup service. Portland implements hierarchical Pseudo MAC (PMAC) addresses for layer 2 routing and forwarding protocol. Port Land assigns a unique PMAC address to each end host.

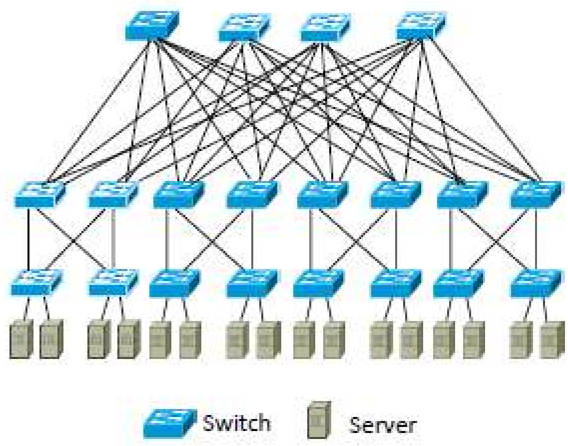

Figure 1.2: VL2 topology [14]

2.2 Data Center Energy

Hardware efficiency and the resource management system deployed on the infrastructure and the efficiency of applications running in the system determine energy consumption. The interdependence of different levels of computing systems in regard to energy consumption is shown in Figure 1.3. Energy efficiency impacts end-users in terms of resource usage costs, which are typically determined by the Total Cost of Ownership (TCO) incurred by the resource provider. Higher power consumption results not only in boosted electricity bills but also in additional requirements to the cooling system and power delivery infrastructure, i.e., Uninterruptible Power Supplies (UPS), Power Distribution Units (PDU), and so on. [14]

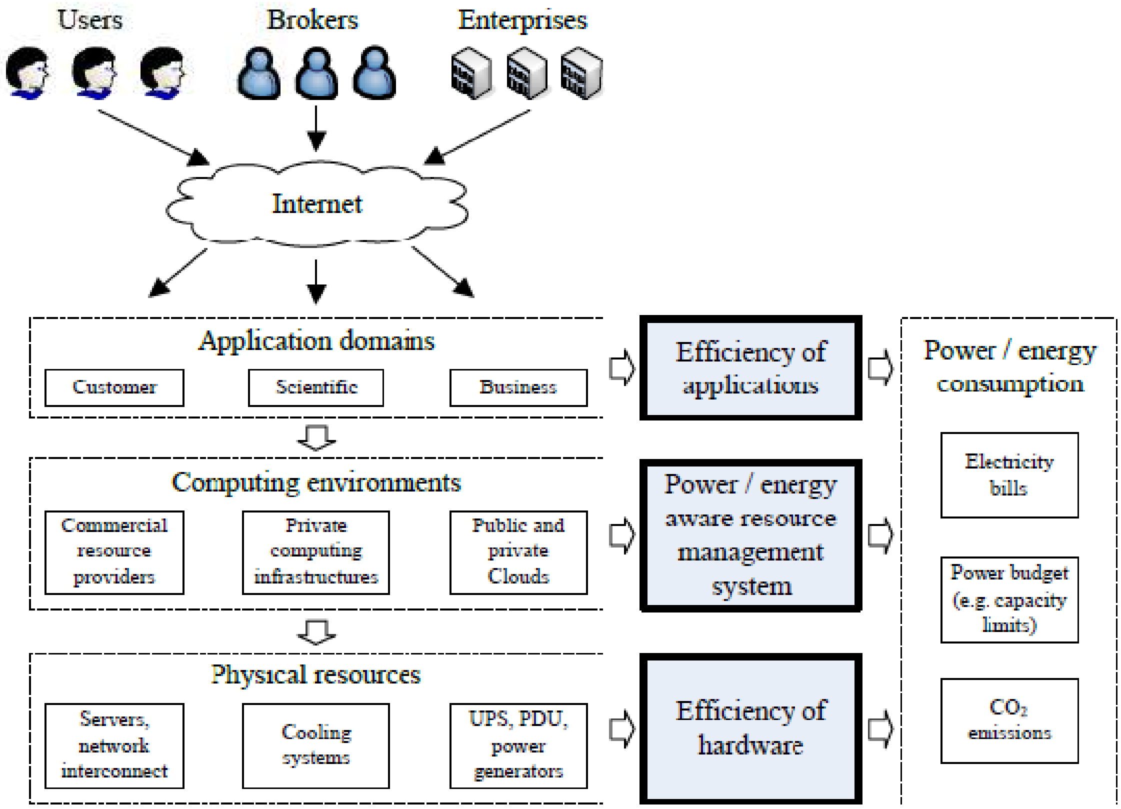

Figure 1.3 Energy consumption at different levels in computing systems [14]

3 Cloud Computing:

Cloud computing gets its name as a metaphor for the Internet in the same time the Internet is represented in network diagrams as a cloud.

Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

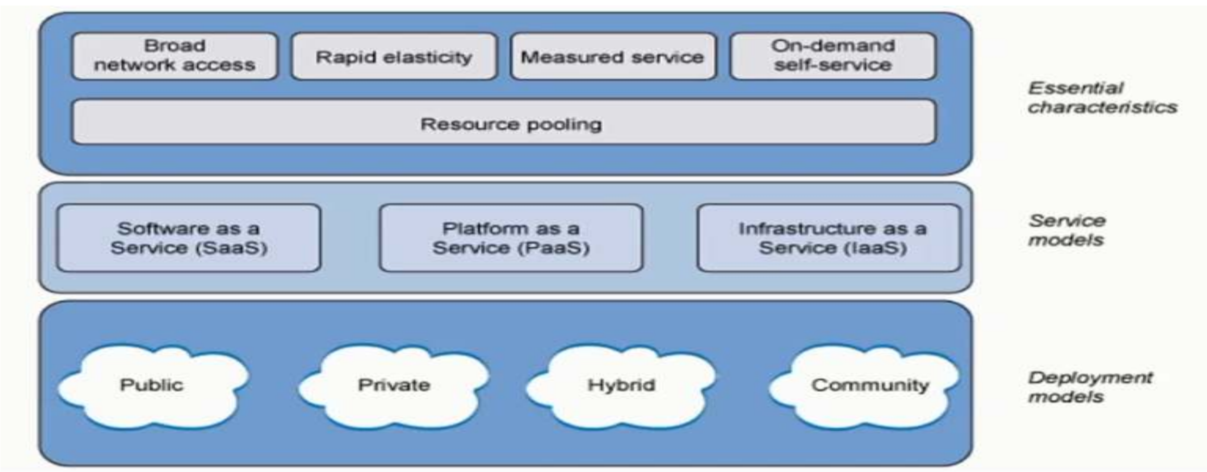

Figure 1.4: NIST Cloud Computing Definitions [30]

3.1 The Main Benefits Of Cloud Computing

Cloud computing boasts several attractive benefits for businesses and end users. Three of the main benefits of cloud computing are:

3.1.1 Self-service provisioning:

End users can spin up compute resources for almost any type of workload on demand. This eliminates the traditional need for IT administrators to provision and manage compute resources.

3.1.2 Elasticity:

Companies can scale up as computing needs increase and scale down again as demands decrease. This eliminates the need for massive investments in local infrastructure, which may or may not remain active.

3.1.3 Pay per use:

Compute resources are measured at a granular level, allowing users to pay only for the resources and workloads they use.

3.2 Cloud Essential Characteristics:

3.2.1 On-demand self-service:

A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service provider.

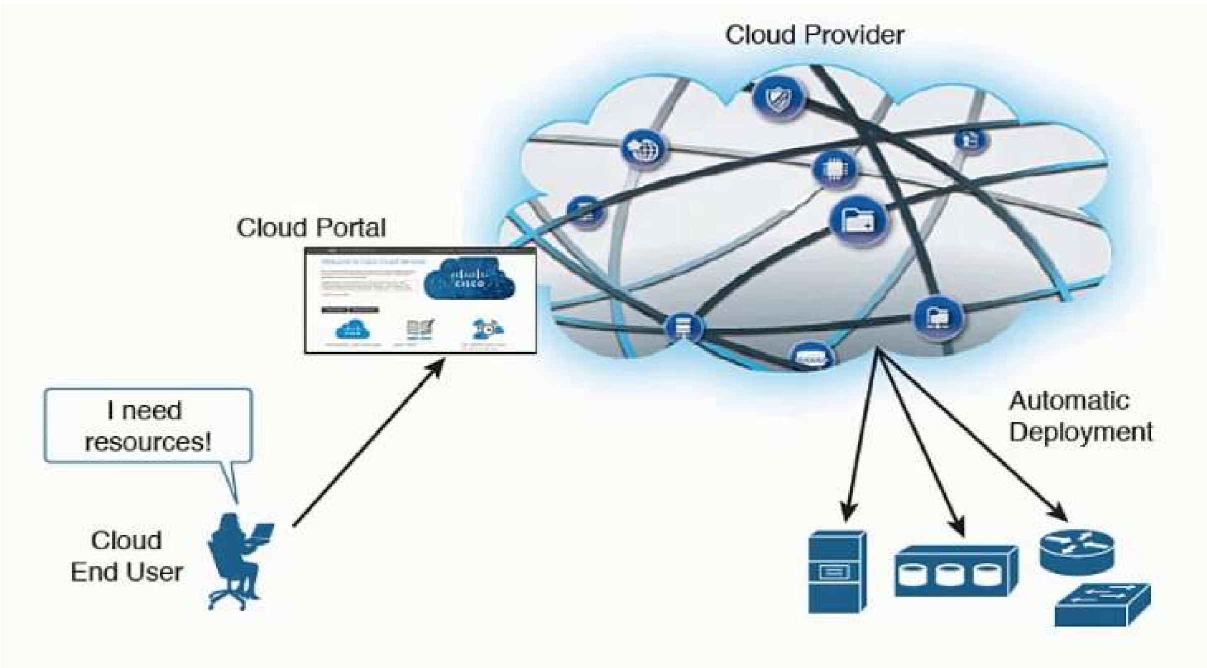

Figure 1.5: On-demand self-service [30]

3.2.2 Broad network access:

Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms (e.g., mobile phones, tablets, laptops, and workstations).

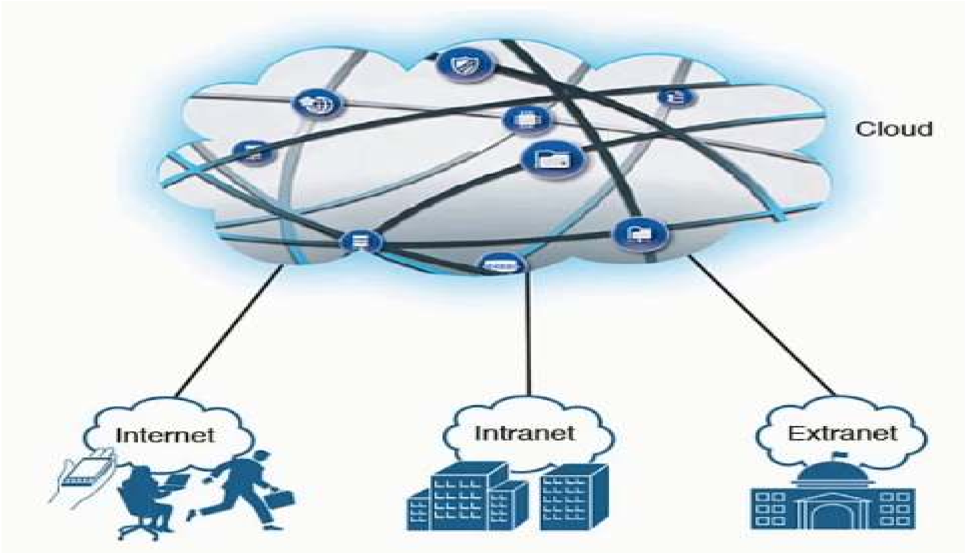

Figure 1.6: Broad network access [30]

3.2.3 Resource pooling:

The provider's computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. There is a sense of location independence in that the customer generally has no control or knowledge over the exact location of the provided resources but may be able to specify location at a higher level of abstraction (e.g., country, state, or datacenter). Examples of resources include storage, processing, memory, and network bandwidth.

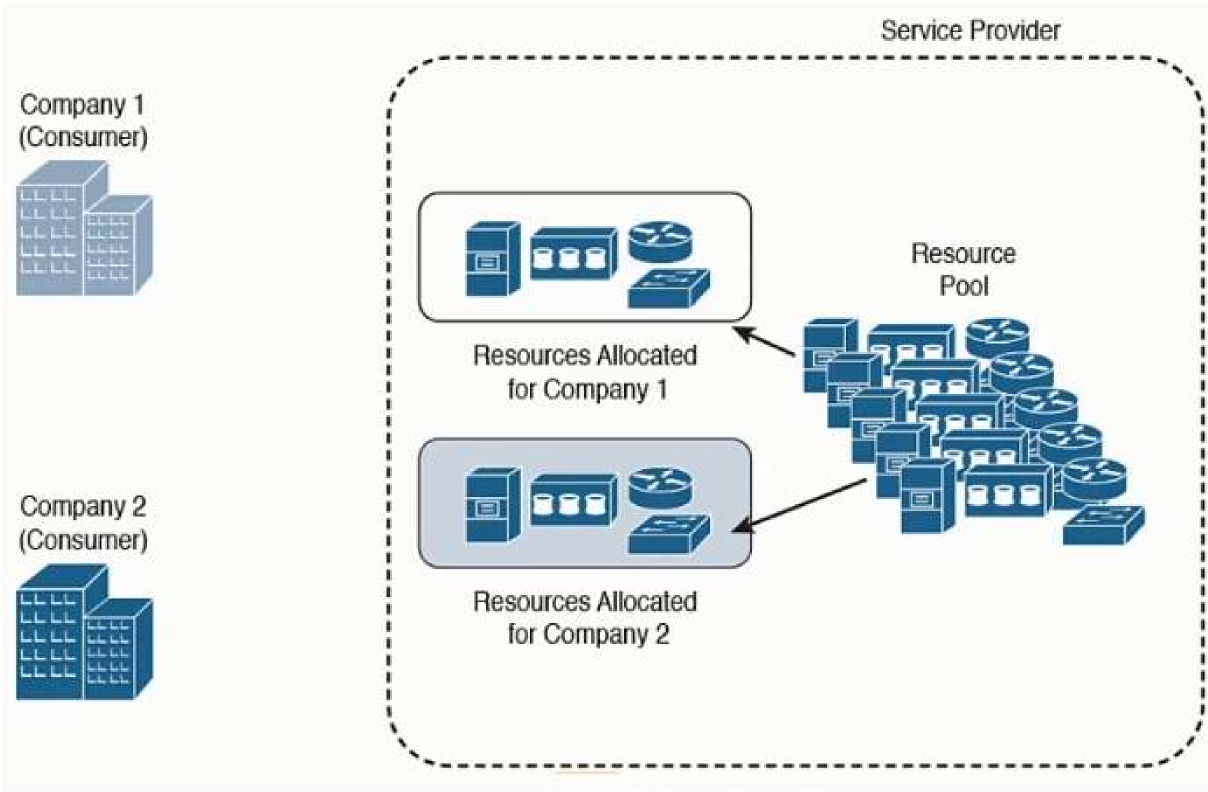

Figure 1.7: Cloud Resource pooling [30]

3.2.4 Rapid elasticity:

Capabilities can be elastically provisioned and released, in some cases automatically, to scale rapidly outward and inward commensurate with demand. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be appropriated in any quantity at any time.

3.2.5 Measured service:

Cloud systems automatically control and optimize resource use by leveraging a metering capability1 at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts). Resource usage can be monitored, controlled, and reported, providing transparency for both the provider and consumer of the utilized service. [30]

3.3 Cloud Service Models:

3.3.1 Software as a Service (SaaS):

Client rent the hardware like servers, memory, and CPU and network connection. Client should install the OS, remotely turns on ad turns off the server and he should install the applications (Client rents just the hardware, and this kind of service is the cheapest).

3.3.2 Platform as a Service (PaaS):

The provider provides the hardware and the infrastructure and he is the only one who turns servers on or off. In this case the client does not care about licenses he just installs the applications and backup the system and data (this kind of service is more expensive than the previous).

3.3.3 Infrastructure as a Service (IaaS):

The client needs just the Internet connection to access and manage the applications. All Storage, networks, and other fundamental computing resources are assured and controlled by the provider.

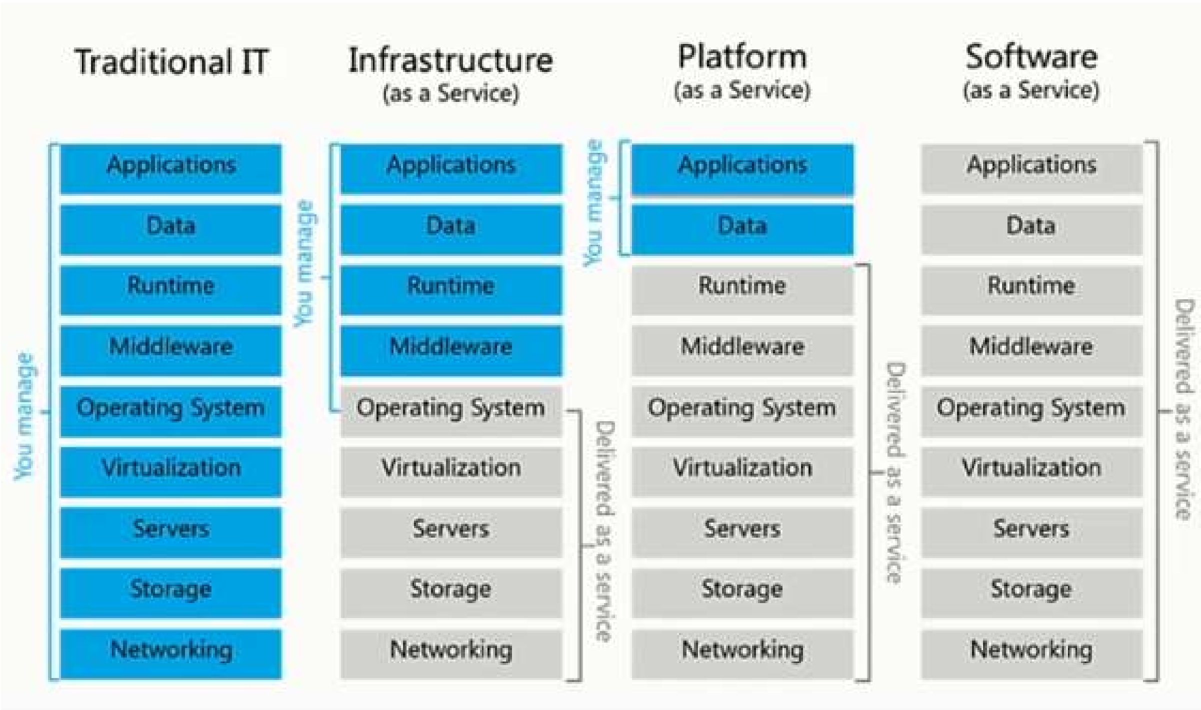

Figure 1.8 Cloud Service Models [30]

3.2 Cloud Deployment Models:

3.4.1 Private cloud:

The cloud infrastructure is provisioned for exclusive use by a single organization comprising multiple consumers (e.g., business units). It may be owned, managed, and operated by the organization, a third party, or some combination of them, and it may exist on or off premises.

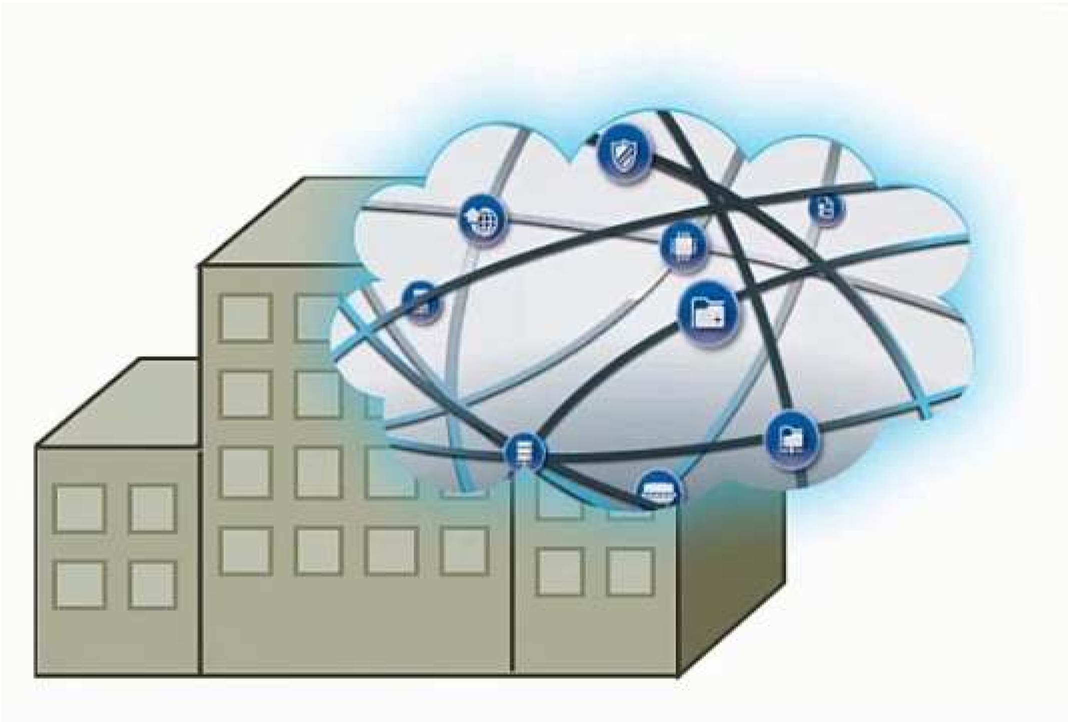

Figure 1.9 Private cloud [30]

3.4.2 Community cloud:

The cloud infrastructure is provisioned for exclusive use by a specific community of consumers from organizations that have shared concerns (e.g., mission, security requirements, policy, and compliance considerations). It may be owned, managed, and operated by one or more of the organizations in the community, a third party, or some combination of them, and it may exist on or off premises.

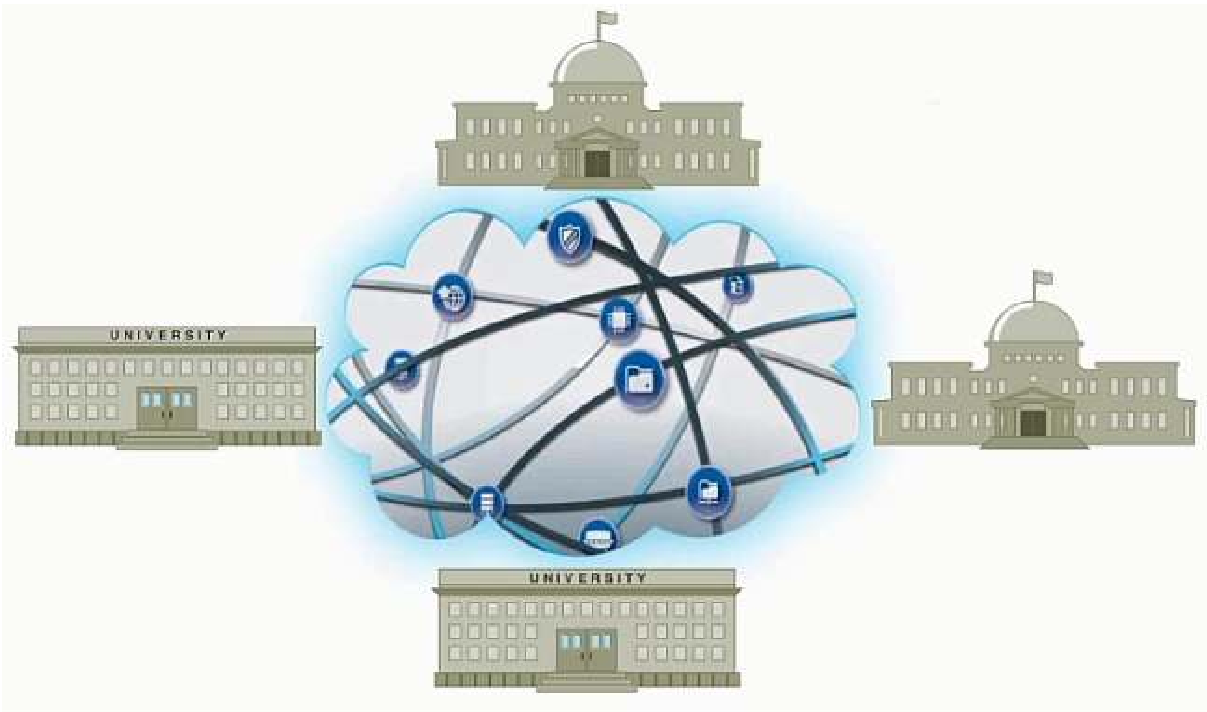

Figure 1.10: Community cloud [30]

3.4.3 Public cloud:

The cloud infrastructure is provisioned for open use by the general public. It may be owned, managed, and operated by a business, academic, or government organization, or some combination of them. It exists on the premises of the cloud provider.

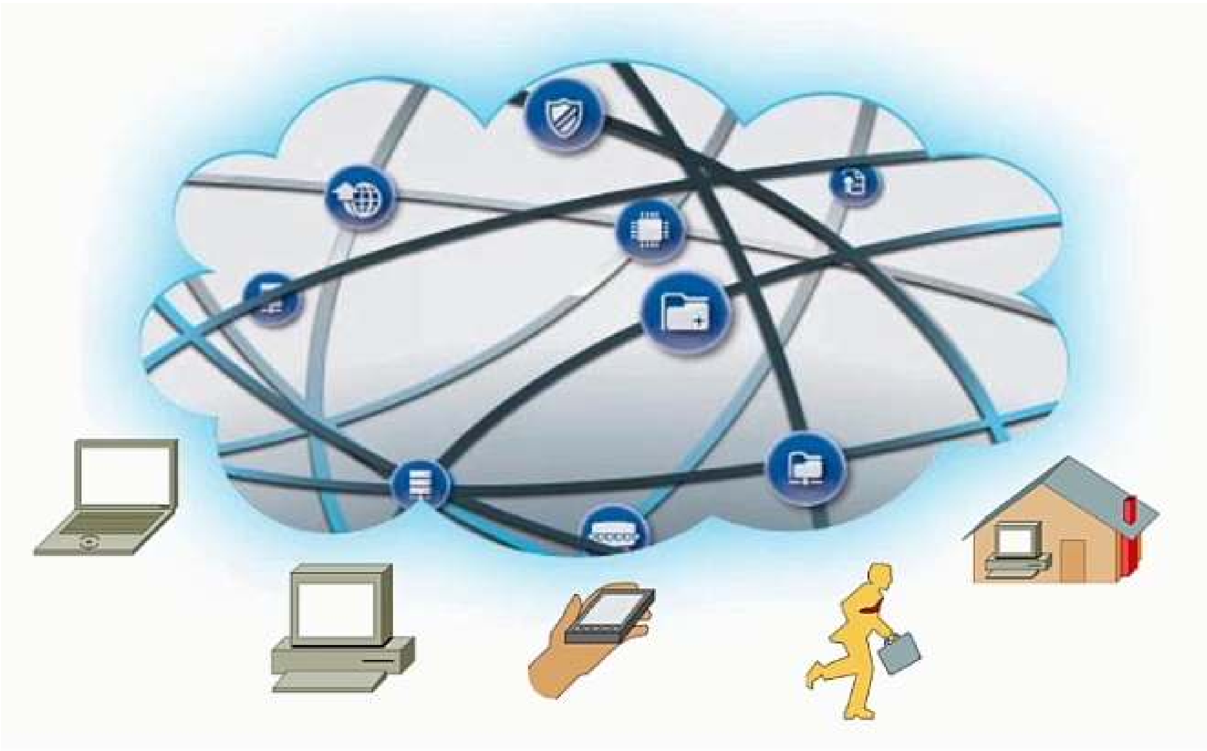

Figure 1.11: Public cloud [30]

3.4.4 Hybrid cloud:

The cloud infrastructure is a composition of two or more distinct cloud infrastructures (private, community, or public) that remain unique entities, but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load balancing between clouds). [30]

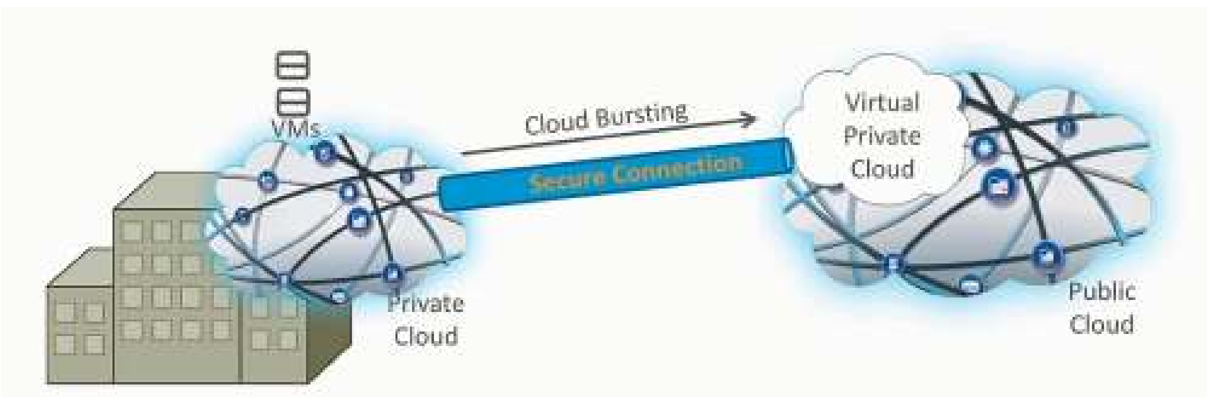

Figure 1.12: Hybrid cloud [30]

4 Virtualization:

4.1 Need's of Virtualization:

In data center we used the virtualization method to allow the sharing of one physical server among multiple virtual machines (VM). With different objectives, we can implement virtualization in both the server and switch domain. Server domain virtualization usually sharing limited resources among different applications.

Virtualization in the Data Center Network domain, on the other hand, aims to implement logically different addressing and forwarding mechanisms on the same physical infrastructure. Data Center Network virtualization separates logical networks from the underlying physical network; letting each virtual network (VN) can implement customized network protocols and management policies [28]. Also, since VNs are logically separated from one another, implementing performance isolation and application QoS is facilitated. Management procedures of VNs will be more flexible because every VN has its own control and management system. Furthermore, isolation offered in network virtualization environments can also minimize the impact of security threats.

Virtualization enables services to be moved between servers, and virtualization has multiple VMs which can serve different applications multiplexed to share one server. Data center resources are underutilized since the average traffic load accounts for about 30% of its resources [28], datacenters can migrate virtual machines to consolidate workloads on a set of servers and then shutdown underutilized servers to save a great power. The migration of VMs is optimized by selecting the VMs to be relocated on the basis of heuristics related to utilization thresholds.

4.1 Types of virtualization

There are six areas of IT where virtualization is making headway:

4.2.1. Network virtualization is a method of combining the available resources in a network by splitting up the available bandwidth into channels, each of which is independent from the others and can be assigned -- or reassigned -- to a particular server or device in real time. The idea is that virtualization disguises the true complexity of the network by separating it into manageable parts; much like your partitioned hard drive makes it easier to manage your files.

4.2.2. Storage virtualization is the pooling of physical storage from multiple network storage devices into what appears to be a single storage device that is managed from a central console. Storage virtualization is commonly used in storage area networks.

4.2.3. Server virtualization is the masking of server resources -- including the number and identity of individual physical servers, processors and operating systems -- from server users. The intention is to spare the user from having to understand and manage complicated details of server resources while increasing resource sharing and utilization and maintaining the capacity to expand later.

The layer of software that enables this abstraction is often referred to as the hypervisor. The most common hypervisor -- Type 1 -- is designed to sit directly on bare metal and provide the ability to virtualize the hardware platform for use by the virtual machines (VMs). KVM virtualization is a Linux kernel-based virtualization hypervisor that provides Type 1 virtualization benefits similar to other hypervisors. KVM is licensed under open source. A Type 2 hypervisor requires a host operating system and is more often used for testing/labs.

4.2.4. Data virtualization is abstracting the traditional technical details of data and data management, such as location, performance or format, in favor of broader access and more resiliency tied to business needs.

4.2.5. Desktop virtualization is virtualizing a workstation load rather than a server. This allows the user to access the desktop remotely, typically using a thin client at the desk. Since the workstation is essentially running in a data center server, access to it can be both more secure and portable. The operating system license does still need to be accounted for as well as the infrastructure.

4.2.6. Application virtualization is abstracting the application layer away from the operating system. This way the application can run in an encapsulated form without being depended upon on the operating system underneath. This can allow a Windows application to run on Linux and vice versa, in addition to adding a level of isolation.

5 Conclusion

In this chapter, we have looked at general knowledge of data center world and cloud computing to get the readers familiars with this kind of subject and prepare them to the next chapter where we explain how the system model and the energy model work.

CHAPTER 2

Energy-Efficient VM Placement first fit decrease Algorithm for Cloud Data Center

CHAPITRE 2

Energy-Efficient VM Placement first fit decrease Algorithm for Cloud Data Center

1 Introduction

The data centers share resources in the granularity of VM not only have high resource utilization, but also have better isolation of the application. The stability, efficiency of resource, users' satisfaction and operating costs of cloud services are directly related to the virtual machine placement problem. Which prove the important of the study of cloud computing data center virtual machine placement problem.

One of an important part of the operating costs is the energy consumption especially with the expending scale of the cloud data centers.

Power consumption is mainly due to servers, networking devices, and cooling systems. There are two main approaches for reducing the energy consumption of data centers: (a) Shutting down devices or (b) scaling down performance.

The energy consumption of a tenant placed into data center can save by an effective approach which is virtualization. For each virtual machine (VM) tenant expresses computation requirement. for each pair of VMs bandwidth requirement. To increase utilization of data centers virtualization can group VMs into fewer servers (physical machines (PMs).

However, the assignment of VM is restricted by server computation resource such as CPU, memory, and storage or hard disks. Most tenants can't be placed into a single server; they must be placed in multiple servers. VMs of a tenant placed in different servers always need to communicate to each other which mean the need to allocate network resource and link bandwidth.

This chapter talks about the resource allocation problem for virtual networks in cloud data centers and how we deal with all the constraints to assign VM to PM?

2 RELATED WORKS

Research on the virtual machine placement has been the key of cloud resource scheduling; many scholars have carried out a lot of work, and put forward their own solutions for this problem. Here, we selected a small part of the solution as representatives, to make a brief summary and analysis.

Song et al. [21/4] proposed a virtual machines allocation mechanism based on two-layer on-demand resource for data center. The proposed local and global resource allocation algorithms with feedback optimize the allocation of resources. Li et al. [5] presented multiple objective genetic algorithms with the application service level objectives, to develop a framework for the virtual machine placement strategy.

In [6], they propose heuristic virtual machine placement algorithm. The loads on the physical machine balance the objectives of maximizing the number of served applications and minimizing the cost of migration. This work focuses on scalability, but in the case that all of the virtual machines can be placed, and the minimization of power and migration cost are the goals, there is no effect on the placement of virtual machines on a physical server..

Ardagna et al. [9] propose the VM placement solutions based on SLAs to maximize the profits of the cloud computing system. The issues they raised consider only the soft contract service level agreements.
The several solutions above are from the perspective of improving service, with emphasis on meeting the quality of service (QoS). As for the virtual machine placement algorithms, they use different strategies, basically focusing on the improving resource utilization without violating SLAs. However, these programs do not consider the energy consumption. Energy consumption has become an important part of operating costs, thus energy-efficient virtual machine placement algorithm is a critical research.

Cardosa et al. [11] studied the energy-aware virtual machine allocation method in virtualized enterprise computing environments, using a priori objective function. This method can effectively reduce energy consumption, but not for cloud computing data centers

3 Virtual machine placement Algorithms

Virtual Machine placement is the process of selecting the most suitable Physical Machine (PM) for a given Virtual Machine (VM). So a VM placement algorithm aims at determining the most optimal VM to PM mapping whether it is an initial VM placement or a VM migration for placement re-optimization.

Depending on the application performance requirements we allocate computing resources reasonably. Which give as possibility to integrate the virtual machine onto fewer physical machines. So we can avoid the idle physical machine, lose resources and energy, which can greatly reduce energy consumption.

4 Classifications of VM Placement Algorithms

4.1 Static VM Placement

Depending on the goal of placement, a VM placement algorithm can be broadly categorized into two types:

4.1.1 Power-based approach: Aims to obtain a VM-PM mapping which results into a system that is energy-efficient with utmost resource utilization, [1, 13].

4.1.2 QoS-based approach: Aims to obtain a VM-PM mapping to ensure maximal fulfillment of quality of service requirements, [7, 27, 28].

In our work we focus on power-based approach.

4.2 Dynamic VM Placement

Depending on the necessity of migration, (transferring a VM from one PM to another PM) a VM Placement algorithm can be classified into 2 types:

4.2.1 Static VMP- VM placement does not consider either the states of the virtual machines and physical machines, or the arrival rate of the user requests.

4.2.2 Dynamic VMP- achieves optimum solutions from the already present mapping of VMs at minimal cost.

Furthermore, the dynamic VM placement algorithms can be categorized as reactive and proactive VM placement.

a) Proactive VMP: changes the VM's Physical Machine of an initial placement before the system reaches a certain condition or satisfies a specific criteria.

b) Reactive VMP: changes to an initial placement after the system reaches a certain undesired state. The change to this original placement could be because of the performance, maintenance, power or load problems or some SLA violations.

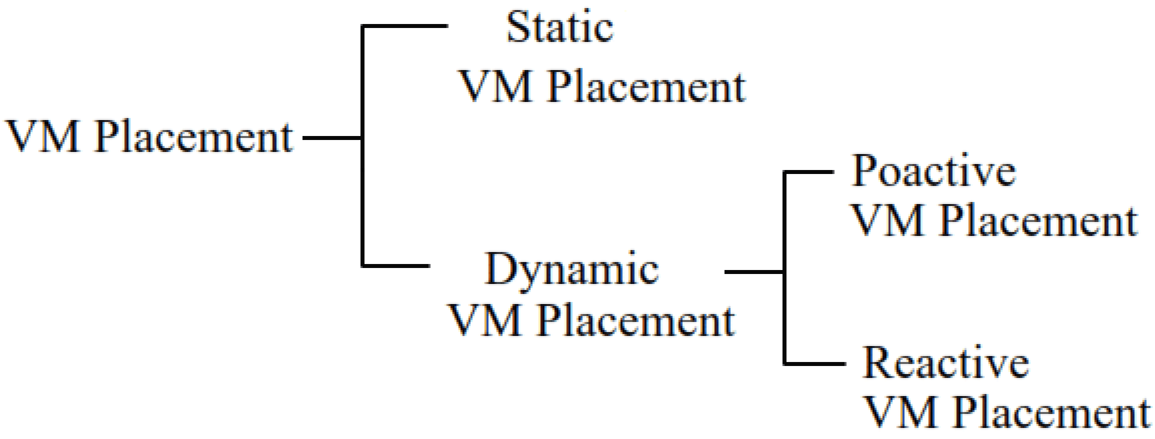

Figure2.1 Classifications of VM Placement [24].

5 Virtual machine placement problem in cloud

In general, each VM shares physical hardware resources with the other VMs and to be more secure, it is kept isolated from the other VMs [24]. As shown in Figure. 2.1

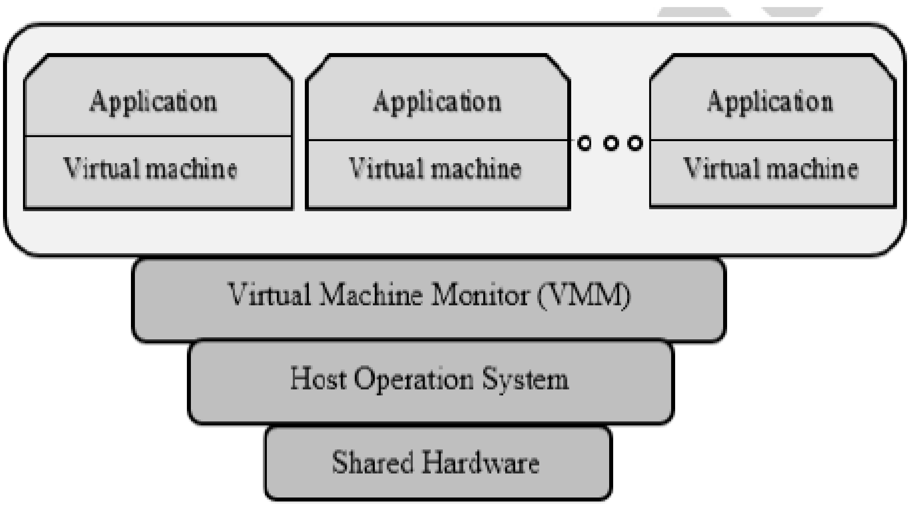

Figure2.2 Hosted Virtualization [24]

The problem mapping tasks on distributed services belongs to a class of problems known as NP-hard problem. For such problems, no known algorithms are able to generate the optimal solution within polynomial time. In cloud environments, Virtual machine placement decisions must be made in the shortest time possible because there are many users competing for resource and time.

6 Virtual machine placement optimization

After acquiring information about virtual machines in the cloud a set of appropriate candidates is highlighted.

The resource selection mechanism elects the candidate solution that fulfils all requirements and optimizes the usage of the infrastructure.

The resource selection may be done using in optimization algorithm.

Many optimization strategies may be used, from simple and well known techniques such as simple metaheuristic algorithms, genetic algorithm, Ant colony and particle swarm optimization (PSO) for cloud.

7 Metaheuristic approach

Metaheuristic is an approach deals with optimization problems using metaheuristic algorithms and no classic method can solved. Numerous simple and primary VM placement algorithms exist one of them first fit decrease.

First fit decreasing: each PM or VM is represented with different resource capacities and decreasing ordered.

8 System Model

Our work base on infrastructure as a service (IaaS) model of the system is shown in Figure 2.3: Users submit applications through their terminals; cloud service suppliers provide a virtual machine as unit of the required computing resources to users. Users rent these resources and pay the provider according to the amount of occupied resources and the length of occupied time. IAAS package the hardware equipment and other basic resources as a service to users. In IAAS environment, users can do almost anything they want to do, but they must consider how to make multiple machines work together. The biggest advantage of IAAS is that it allows users to dynamically apply or release nodes, bill according to usage. The number of running server in IAAS reaches as much as hundreds of thousands, so that the resources users can apply for is almost unlimited. Meanwhile IAAS is shared by the public, which has a higher efficiency of resource usage. [23]

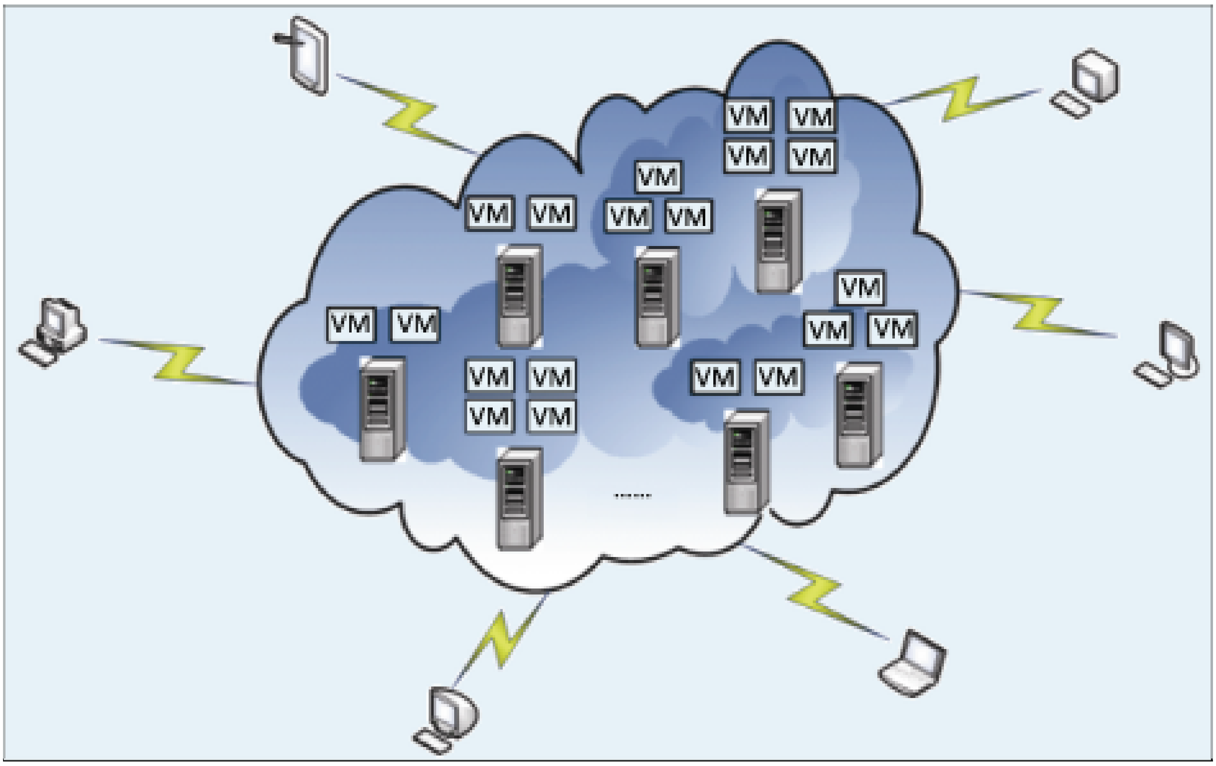

Figure 2.3: The IAAS system model [23]

The system model we use is shown in Figure. 2.4 The virtual machine placement in the cloud data center; before assign the virtual machine to a physical machine to complete the mapping of VMs and PMs we allocate the computing resources to the applications.

We implement the first fit decrease algorithm. Energy consumption is an important goal, we save the energy through letting more physical machines off (or letting idle physical machine into hibernation). User submits application requests, we complete the mapping between virtual machines and physical machines according to the virtual machine that the application request for.

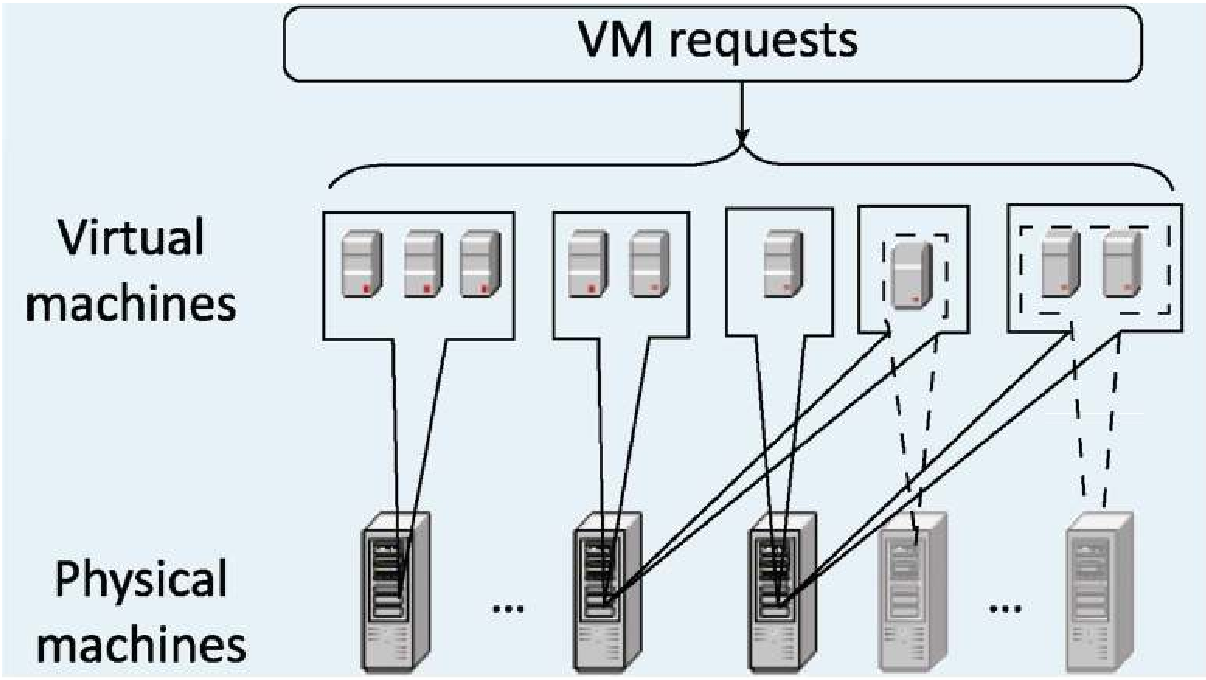

Figure2.4 The system model

9 Energy Model

Generally, CPU, memory, hard disk capacity and network bandwidth are factors affect energy consumption decide the main cost of each node. CPU consumes a major portion of the energy it has a top priority. We assume that the physical machine of the cloud data center has three modes: idle mode, active mode and sleep mode. Numerous documents show that the average energy consumption of idle physical machine is 70 % of that running at full speed. Thus, in order to reduce the overall energy consumption, we need to convert the idle physical machine into sleep mode. The CPU utilization has a classical linear relationship with the workload of the entire system. That is to say, the CPU utilization is on behalf of energy efficiency to some extent. Based on this fact, we use the same energy model in the Ref. [8], as defined in Eq. (1) in the following. [23]

$$P(u) = K \cdot P_{max} + (1 - K) . P_{max} . u \qquad (1)$$

P max represents the maximum power of the server with the peak workload, k denotes the proportion of the power the idle server consumed, then 1−k represents the consumption proportion server with workload, and u is the CPU utilization.

$$E = \int_{t0}^{t1} P\big(u(t)\big)dt \qquad (2)$$

The CPU utilization may change over time and load. Thus, the CPU utilization is a function of time, so we use u (t) to represent. Therefore, the total energy E of a physical node can be defined as a integral of the energy consumption function over a period of time, as shown in Eq. (2).

The overall average utilization of the physical machine i is defined as follows, as shown in formula (3), taking the average utilization of CPU, memory, network bandwidth.

$$IR_i^u = \frac{CPU_i^u + MEM_i^u + NET_i^u}{3} \qquad (3)$$

The definition of unbalanced multi-dimension utilization of the physical machine i, we refer to Ref. [7], as the following Eq. (4) shown

$$IUV_i = \frac{(CPU_i^u - IR_i^u)^2 + (MEM_i^u - IR_i^u)^2 + (NET_i^u - IR_i^u)^2}{3} \qquad (4)$$

The roughly steps we mentioned in the virtual machine placement selection algorithm are: (1) Select physical machines which can be set, i.e. CPU, memory and network bandwidth can meet the physical requirements of the virtual machine; (2) select the most appropriate physical machines can be set in accordance with the energy and standard balanced utilization. Therefore, the selection criteria is very important, we choose two factors here, the balanced utilization and the energy consumption to define the selection criteria, as shown in the Eq. (5).

$$SM_i = w_1.IUV_i + w_2.P_I(u) \qquad (5)$$

In Eq. (5), w1; w2 represent the power imbalance rate and energy consumption values, the values must meet w1+w2 =1. For example, we set that w1; w2 equal to 0.5, representing the same weight of the unbalanced utilization and energy consumption. The specific algorithm is shown in the following Table 2.1:

Algorithm 1 energy-efficient virtual machine placement(EVMP)
Input: a new VM, a list of servers S-list in the cloud data center
Output: a VM-PM match
SM=max
for each server in the S-list
if the server i has enough resource for the VM then
calculate the selection criteria value of the server i in Eq(5)
if(SM_i <SM) then
SM= SM_i
Assign the VM to the server i

Table2.1 Energy-efficient virtual machine placement (EVMP) [23].

In this case the major problem of constraints is done but we face the problem of losing resources. We modify the EVMP algorithm by EVMP first fit decrease algorithm; we order the VMs from Max to Min and pass to assign VMs to server.

First Fit Decreasing product: Each physical machine or the VM is represented with different resource capacities and each resource capacity is normalized in the percentage of utilization.

10 conclusion

In this chapter, it has been presented the system model, the energy model and shows the energy-efficient virtual machine placement algorithm how facilitate to gather all the constraints in one variable. It lights the way to optimize the energy that which prove it in the next chapter.

CHAPTER 3

IMPLEMENTATION AND TEST

CHAPITRE 3

IMPLEMENTATION AND TEST

1 Introduction

In this final chapter, we perform experiments to evaluate EVMP first fit decrease algorithm described in last section. In our experiments, we use some metrics to evaluate the performance.

2 Development environment

2.1 Operating system

We used the Microsoft Windows 7, it has the advantage of being the most used OS, it is easy to use, high efficiency and HCI.

2.2 Microsoft Visual Studio 2012

Microsoft Visual Studio is an integrated development environment (IDE) from Microsoft. It is used to develop computer programs for Microsoft Windows superfamily of operating systems, as well as web sites, web applications and web services. Visual Studio uses Microsoft software development platforms such as Windows API, Windows Forms, Windows Presentation Foundation, Windows Store and Microsoft Silverlight. It can produce both native code and managed code.

Microsoft VisualC# (pronounced "C sharp") is a programming language that is designed for building a variety of applications that run on the .NET Framework. C# is simple, powerful, type-safe, and object-oriented. The many innovations in C# enable rapid application development while retaining the expressiveness and elegance of C-style languages.

As an object-oriented language, C# supports the concepts of encapsulation, inheritance, and polymorphism. All variables and methods, including the Main method, the application's entry point, are encapsulated within class definitions. A class may inherit directly from one parent class, but it may implement any number of interfaces. Methods that override virtual methods in a parent class require the override keyword as a way to avoid accidental redefinition. In C#, a struct is like a lightweight class; it is a stack-allocated type that can implement interfaces but does not support inheritance.

In addition to these basic object-oriented principles, C# makes it easy to develop software components through several innovative language constructs, including the following:

- Encapsulated method signatures called *delegates*, which enable type-safe event notifications.
- Properties, which serve as accessors for private member variables.
- Attributes, which provide declarative metadata about types at run time.
- Inline XML documentation comments.
- Language-Integrated Query (LINQ) which provides built-in query capabilities across a variety of data sources.

If you have to interact with other Windows software such as COM objects or native Win32 DLLs, you can do this in C# through a process called "Interop." Interop enables C# programs to do almost anything that a native C++ application can do. C# even supports pointers and the concept of "unsafe" code for those cases in which direct memory access is absolutely critical.

The C# build process is simple compared to C and C++ and more flexible than in Java. There are no separate header files, and no requirement that methods and types be declared in a particular order. A C# source file may define any number of classes, structs, interfaces, and events.

3 Experiments

We use some metrics to evaluate the performance. The first metric is the total power consumption of the data center. Thesecond metric is the imbalance utilization value of the resources in the cloud datacenter. For that the different types of applications require different resources, whichinevitable results grate differences in the utilization of resources as the time went.

3.1 Test

We propose a data center consisting of 10 heterogeneous physicalmachines. These PMs has three types, as shown in Table 3.1

Table of PM			
NAME	CPU/MIPS	MEM	Network/ (Mbit s-1)
SER1	1000	64	200
SER2	1000	64	200
SER3	1000	64	200
SER4	1000	64	200
SER5	1000	64	200
SER6	1000	64	200
SER7	1000	64	200
SER8	1000	64	200
SER9	1000	64	200
SER10	1000	64	200

Table 3.1 Types of heterogeneous physical machines

In our simulated experiments, there are four different types of VMs. The parametersof the VMs are described in Table 3.2

Table of VM			
NAME	CPU/MIPS	MEM	Network/ (Mbit s-1)
VM1	100	1	10
VM2	700	1	70
VM3	300	2	30
VM4	400	4	40
VM5	700	8	50
VM6	700	7	60
VM7	500	5	10
VM8	600	5	40
VM9	200	4	20
VM10	700	3	70
VM11	300	3	20
VM12	500	3	50

Table 3.2 Types of Virtual machines

3.2 Result

The experiment is proven, the SM wasted in FFD less than the other. It is as shown in the comparison between Figure 3.1 FFD algorithm results and Figure 3.2 FF algorithm results.

Server Name	SM()	After	SM()
SERVER-1	908808,9	After	3122,781
VM 1	7345	-	1
VM 9	28127,22	-	1
VM 3	61735,55	-	1
VM 4	110020	-	1
VM 7	147847,2	-	1
VM 8	230908,9	-	1
VM 5	319702,2	-	1
SERVER-2	908808,9	After	250736,6
VM 6	327327,2	-	1
VM 2	330745	-	1
SERVER-3	908808,9	After	576348,3
VM 10	332460,6	-	1
SERVER-4	908808,9	After	908808,9
SERVER-5	908808,9	After	908808,9
SERVER-6	908808,9	After	908808,9
SERVER-7	908808,9	After	908808,9
SERVER-8	908808,9	After	908808,9
SERVER-9	908808,9	After	908808,9
SERVER-10	908808,9	After	908808,9

Figure 3.1 First Fit Decrease algorithm results

Server Name	SM()	After	SM()
SERVER-1	932808,9	After	3222,197
VM 10	332460,6	-	1
VM 2	330745	-	1
VM 8	230908,9	-	1
VM 9	28127,22	-	1
VM 1	7345	-	1
SERVER-2	932808,9	After	27912,19
VM 6	327327,2	-	1
VM 5	319702,2	-	1
VM 7	147847,2	-	1
VM 4	110020	-	1
SERVER-3	932808,9	After	871073,3
VM 3	61735,55	-	1
SERVER-4	932808,9	After	932808,9
SERVER-5	932808,9	After	932808,9
SERVER-6	932808,9	After	932808,9
SERVER-7	932808,9	After	932808,9
SERVER-8	932808,9	After	932808,9
SERVER-9	932808,9	After	932808,9
SERVER-10	932808,9	After	932808,9

Figure 3.2 First Fit algorithm results

4 Interfaces of the application

4.1 Home page

It represents the first page loaded, it contain a menu to add PMs or VMs and choose thealgorithm applied. It contains six-list box represent PMs created and its SM, VMs created and its SM before and after ordered. It is shown in Figure. 3.3Home page.

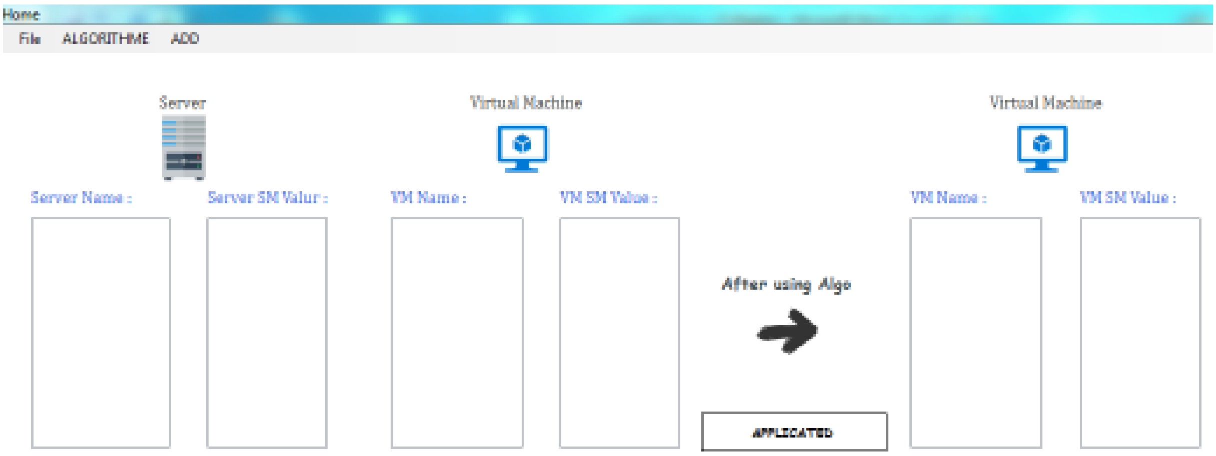

Figure 3.3 Home page

4.2 Creation of servers

In the home pagewe click on ADD and choose servers its load a new form shows the different characteristics of server like it shown in Figure. 3.4Creation of servers.With the same instructions we can create VMs.

Figure 3.4 Creation of servers

4.3 Apply the FFD algorithm

After the creation of PMs,VMs and choose the algorithm from the menu as mention before we click on the button APPLICATED the VMs ordered in list boxas shown in Figure. 3.5VM first fit decrease representation. In addition, new form loaded it presents the results (each VM where it assigned) as shown in Figure. 3.6 VM first fit decrease results.

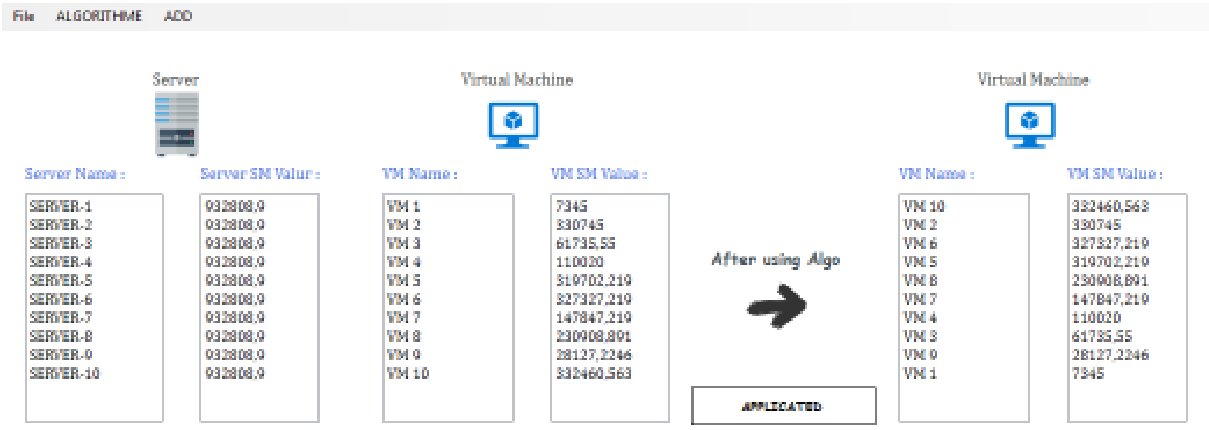

Figure 3.5VM first fit decrease representation

Server Name	SMO	After	SMO
SERVER-1	908808,9	After	3122,781
VM 1	7345	-	1
VM 9	28127,22	-	1
VM 3	61735,55	-	1
VM 4	110020	-	1
VM 7	147847,2	-	1
VM 8	230908,9	-	1
VM 5	319702,2	-	1
SERVER-2	908808,9	After	250736,6
VM 6	327327,2	-	1
VM 2	330745	-	1
SERVER-3	908808,9	After	576348,3
VM 10	332460,6	-	1
SERVER-4	908808,9	After	908808,9
SERVER-5	908808,9	After	908808,9
SERVER-6	908808,9	After	908808,9
SERVER-7	908808,9	After	908808,9
SERVER-8	908808,9	After	908808,9
SERVER-9	908808,9	After	908808,9
SERVER-10	908808,9	After	908808,9

Figure 3.6VM first fit decrease results

5 Interfaces of code source

5.1 Affect method

It represents how VMs are affected it shown in Figure. 3.7 Affect method.

```
public bool affect(VirtualMachines v)
{

    if (this.SMv > v.SM())
    {
        this.List_of_vm.Add(v);
        v.Set(this.NAME);
        this.SMv -= v.SM();
        return true;
    }
    else
        return false;
}
```

Figure 3.7 Affect method

5.2 First Fit Decrease Algorithm applied

It represents how we applied our algorithm after we choose it on the VMs as shown in Figure. 3.8First Fit Decrease Algorithm applied.

```
private void button3_Click_1(object sender, EventArgs e)
{
    /***************************************************************/
    foreach (VirtualMachines v in List_of_vm)
    {
        for (int j = 0; j < List_of_server.Count; j++)
        {
            bool f = List_of_server.ElementAt(j).effect(v);

            if (f)
            {
                //listBox2.Items.Add(tx_nvm.Text);
                MessageBox.Show("Vm has been Added to  " + List_of_server.ElementAt(j).SName());
                // listBox3.Items.Add(vm.Get());
                break;
            }
            else
                MessageBox.Show("Vm dosn't added to " + List_of_server.ElementAt(j).SName() + " \n  plz check other serv

        }
    }
    Form4 f4 = new Form4(List_of_server);
    f4.List_of_server = List_of_server;
    f4.ShowDialog();

    /***************************************************************/
```

Figure 3.8 First Fit Decrease Algorithm applied

6 Conclusion

In this final chapter, we showed how we realize the optimization with the first fit decrease on Energy-Efficient VM Placement algorithm and proved by the comparison of the two algorithm after solve le problem of constraints such as CPU, memory, network bandwidth, and so on. The experimental results show that the EVMP-FFD contribute a lot to energy saving. It gives the priority to the max virtual machine than return and complete with small virtual machine with the aims of minimize the energy don't starts an necessaries physical machine and minimize the waste of resources.

GLOBAL CONCLUSION

We use the experimental of Energy-Efficient VM Placement Algorithmsfor Cloud Data Center and apply a metaheuristic optimization approach to get a suitable solution. In general, finding an optimal solution or even sub-optimal solutions is not easy task.

However,the experimental did not consider other part of the energy consumption, such as the cooling power consumption. The energy model does not capture the energy consumption accurately.

1 Future work

Virtual Machine Placement as we mention before is divided to two parts we deal with the first one, which is the initial placement of VMs, and we let our work open to research with a proposition for the second part which is the migration of virtual machine. Our works needs two thresholds, the upper and lower utilization thresholds for the servers and test the servers each time to detect if the resource utilization of the server is up or down the thresholds to decide the migration of VMs (migration of some VMs to get the utilization of resource under the upper thresholds or migration of all VMs and switch off the server of cores to minimize the power consumption.

BIBLIOGRAPHY

[1] Amazon, E., “Amazon elastic compute cloud (Amazon EC2),” Amazon Elastic Compute Cloud (Amazon EC2), 2010

[2] Anton Beloglazov “Energy-Efficient Management of Virtual Machines in Data Centers for Cloud Computing” February 2013,……..

[3] Ardagna, D., Panicucci, B., Trubian, M., et al.: Energy-aware autonomic resource allocation in multitier virtualized environments. IEEE Trans. Serv. Comput. 5(1), 2–19 (2012)

[4] Ballani, H., Costa, P., Karagiannis, T., and Rowstron, A., “Towards predictable datacenter networks,” ACM SIGCOMM Computer Communication Review, Vol. 41, ACM, 2011, pp. 242–253.

[5] Beloglazov, A. and Buyya, R., 2010, November. Adaptive threshold-based approach for energy-efficient consolidation of virtual machines in cloud data centers. In Proceedings of the 8th International Workshop on Middleware for Grids, Clouds and e-Science (Vol. 4). ACM

[6] Beloglazov, A., Abawajy, J., Buyya, R.: Energy-aware resource allocation heuristics for Efficient management of data centers for cloud computing. Future Gener. Comput. Syst. 28 (5), 755–768 (2012).

[7] Beloglazov, A. and Buyya, R., 2013. Managing overloaded hosts for dynamic consolidation of virtual machines in cloud data centers under quality of service constraints. Parallel and Distributed Systems, IEEE Transactions on, 24(7), pp.1366-1379.

[8] Beloglazov, A. and Buyya, R., 2010, May. Energy efficient resource management in virtualized cloud data centers. In Proceedings of the 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing (pp. 826-831). IEEE Computer Society.

[9] Benson, T., Akella, A., Shaikh, A., and Sahu, S., “CloudNaaS: a cloud networking platform for enterprise applications,” Proceedings of the 2nd ACM Symposium on Cloud Computing, ACM, 2011, p. 8

[10] Bobroff, N., Kochut, A. and Beaty, K., 2007, May. Dynamic placement of virtual machines for managing sla violations. In Integrated Network Management, 2007. IM'07. 10th IFIP/IEEE International Symposium on (pp. 119-128). IEEE

[11] Cardosa, M., Korupolu, M.R., Singh, A.: Shares and utilities based power consolidation in virtualized server environments. In: IFIP/IEEE International Symposium on Integrated Network Management (IM 2009), pp. 327–334. IEEE (2009)

[12] Greenberg, A., Hamilton, J. R., Jain, N., Kandula, S., Kim, C., Lahiri, P., Maltz, D. A., Patel, P., and Sengupta, S., “VL2: a scalable and flexible data center network,” ACM SIGCOMM Computer Communication Review, Vol. 39, ACM, 2009, pp. 51–62

[13] Greenberg, A., Hamilton, J. R., Jain, N., Kandula, S., Kim, C., Lahiri, P., Maltz, D. A., Patel, P., and Sengupta, S., "VL2: a scalable and flexible data center network," ACM SIGCOMM Computer Communication Review, Vol. 39, ACM, 2009, pp. 51–62.

[14] Guo, C., Lu, G., Li, D., Wu, H., Zhang, X., Shi, Y., Tian, C., Zhang, Y., and Lu, S., "BCube: a high performance, server-centric network architecture for modular data centers," ACM SIGCOMM Computer Communication Review, Vol. 39, No. 4, 2009, pp. 63–74

[15] Guo, C., Lu, G.,Wang, H. J., Yang, S., Kong, C., Sun, P.,Wu,W., and Zhang, Y., "Secondnet: a data center network virtualization architecture with bandwidth guarantees," Proceedings of the 6th International COnference, ACM, 2010, p. 15.

[16] Guo, C., Wu, H., Tan, K., Shi, L., Zhang, Y., and Lu, S., "Dcell: a scalable and faulttolerant network structure for data centers," ACM SIGCOMM Computer Communication Review, Vol. 38, ACM, 2008, pp. 75–86.

[17] Jun Liu Optimizing the Energy Consumption of Servers and Networks in Cloud Data Centers, pp. 60, 2015.

[18] K.T. Rao, P.S. Kiran, and L.S.S. Reddy, "Energy Efficiency in Datacenters through Virtualization: A Case Study," Global Journal of Computer Science and Technology, vol 10, No 3, 2010.

[19] L. A. Barroso and U. Holzle, "The case for energy-proportional computing," Computer, vol. 40, no. 12, pp. 33–37, 2007.

[20] Li, Q., Hao, Q., Xiao, L., Li, Z.: Adaptive management and multi-objective optimization of virtual machine placement in cloud computing. Chin. J. Comput. 34(12), 2253–2264 (2011)

[21] M. Yue, "A simple proof of the inequality FFD (L)< 11/9 OPT (L)+ 1,for all l for the FFD bin-packing algorithm," Acta Mathematicae Applicatae Sinica (English Series), vol. 7, no. 4, pp. 321–331, 1991

[22] Niranjan Mysore, R., Pamboris, A., Farrington, N., Huang, N., Miri, P., Radhakrishnan, S., Subramanya, V., and Vahdat, A., "Portland: a scalable fault-tolerant layer 2 data center network fabric," ACM SIGCOMM Computer Communication Review, Vol. 39, ACM, 2009, pp. 39–50.

[23] Song, W., Xiao, Z., Chen, Q. and Luo, H., 2014. Adaptive resource provisioning for the cloud using online bin packing. Computers, IEEE Transactions on, 63(11), pp.2647-2660.

[24] Song, Y., Sun, Y., Shi, W.: A two-tiered on-demand resource allocation mechanism for VM-based data centers. IEEE Trans. Serv. Comput. 6(1), 116–129 (2013)

[25] Tang, C., Steinder, M., Spreitzer, M., et al.: A scalable application placement controller for enterprise data centers. In: Proceedings of the 16th International Conference on World Wide Web, pp. 331–340. ACM (2007)

[26] X. Fan,W. D.Weber, and L. A. Barroso, “Power provisioning for a warehouse-sized computer,” in Proceedings of the 34th Annual International Symposium on Computer Architecture (ISCA), 2007, pp. 13–23

[27] Xiuyan Lin, Zhanghui Liu, and Wenzhong Guo Energy-Efficient VM Placement Algorithms for Cloud Data Center

[28] Zhang, Y., Y. Li, and W. Zheng, Automatic software deployment using user-level virtualization for cloud-computing. Future Generation Computer Systems, 2013. 29(1): p. 323-329.

Websites

[29] Thibodeau, P., “Data centers are the new polluters,” http://www.computerworld.com/article/2598562/data-center/data-centers-are-the-new-polluters.html, Accessed: 07-15-2015.

Videos

[30] CCNA Data Center DCICT 200-155 LiveLessons

ملخص

مراكز البيانات هي أماكن تحتوي على مئات الآلاف من الخوادم، مترابطة عبر مفاتيح والموجهات والروابط عالية السرعة لتخزين كميات كبيرة من البيانات واستضافة تطبيقات الخدمة على نطاق واسع. سحابة هو نموذج الحوسبة، الذي يوفر مورد الحوسبة كخدمة من خلال الشبكة. تماما مثل المياه والكهرباء التي نستخدمها يوميا، سحابة تسمح للمستأجر لاستخدام الموارد الحوسبة بطريقة مريحة وبناء على الطلب. ويقوم مقدمو خدمات الحوسبة السحابية بنشر مراكز البيانات بسرعة في جميع أنحاء العالم. وتستهلك مراكز البيانات هذه كميات كبيرة من الطاقة، مما يساهم في ارتفاع التكاليف التشغيلية. وبالتالي، فإن تحسين استهلاك الطاقة للخوادم والشبكات في مراكز البيانات يمكن أن يقلل من التكاليف التشغيلية. يسمح المحاكاة الافتراضية لمراكز البيانات هذه بتوطيد خدماتها على عدد أقل من الخوادم الفعلية مما هو مطلوب أصلا. قد يكون هناك مئات من هذه الخدمات تعمل في مركز بيانات واحد. خوارزمية التنسيب فم المقترحة تنظر في قيود الموارد متعددة الأبعاد، مثل وحدة المعالجة المركزية والذاكرة ونطاق ترددي الشبكة، وهلم جرا. الخوارزميات المقترحة تسهم ليس فقط الكثير لتوفير الطاقة، ولكن أيضا محاولة أفضل لتلبية نوعية الخدمة (كوس). ولذلك، فإننا نقدم وفورات كبيرة في تكاليف التشغيل والاستفادة الكاملة من مختلف الموارد في مركز البيانات السحابية. الخوارزمية لديها آفاق واعدة في التطبيق.

كلمات البحث: الحوسبة السحابية، وضع الجهاز الظاهري.

ABSTRACT

Data centers are places contain hundreds of thousands of servers, interconnected via switches, routers and high-speed links for storing large volumes of data and hosting large-scale service applications. Cloud is the computing paradigm, which provides computing resource as a service through network. Just like the water and the electricity we use daily, Cloud allows the tenant to use computing resource in a convenient and on-demand way. Cloud computing service providers are rapidly deploying data centers across the world. These data centers consume significant amounts of energy, contributing to high operational costs. Thus, optimizing the energy consumption of servers and networks in data centers can reduce operational costs. Virtualization allows these data centers to consolidate their services onto a lesser number of physical servers than originally required. There may be hundreds of such services running in a single data center. The proposed VM placement algorithm considers the multi-dimensional resource constrains, such as CPU, memory, network bandwidth, and so on. The proposed algorithms contribute a lot to energy saving. Therefore, we make significant savings in operating cost and make full use of various resources in the cloud data center. The algorithm has promising prospect in application.

Keywords: Cloud Computing, Virtual Machine Placement.

RESUME

Les centres de données sont des endroits contenant des centaines de milliers de serveurs, interconnectés via des commutateurs, des routeurs et des liaisons à grande vitesse pour stocker de gros volumes de données et héberger des applications de service à grande échelle. Cloud est le paradigme informatique, qui fournit des ressources informatiques en tant que service par réseau. Tout comme l'eau et l'électricité que nous utilisons quotidiennement, Cloud permet au locataire d'utiliser des ressources informatiques de manière pratique et à la demande. Les fournisseurs de services de cloudcomputing déploient rapidement des centres de données à travers le monde. Ces centres de données consomment des quantités importantes d'énergie, ce qui contribue à des coûts opérationnels élevés. Ainsi, l'optimisation de la consommation d'énergie des serveurs et des réseaux dans les centres de données peut réduire les coûts opérationnels. La virtualisation permet à ces centres de données de consolider leurs services sur un nombre moindre de serveurs physiques que prévu initialement. Il peut y avoir des centaines de ces services fonctionnant dans un seul centre de données. L'algorithme de placement de VM proposé considère les contraintes de ressources multidimensionnelles, telles que la CPU, la mémoire, la bande passante du réseau, etc. Les algorithmes proposés contribuent l'économie d'énergie. Par conséquent, nous faisons d'importantes économies en coûts d'exploitation et utilisons pleinement diverses ressources dans le centre de données de cloud computing. L'algorithme a une perspective prometteuse en application.

Mots-clés: Cloud Computing, Virtual Machine Placement.