Allocation Strategy for Cloud Datacenter Based on Multi agent & CP approach

Merzoug.Soltane1, Kazar. Okba 2 , Ezziyyani.Mostafa3,Dardour.Makhlouf 4
1University of El_oued LINFI , El-oued ,Algeria
2University of Biskra , Biskra, Algeria
3University of Abdelmalek Essaâdi IBDD-TC, Tangier, Maroc
4University of Tebessa LRS ,Tebessa ,Algeria
{ merzoug-soltane@univ-eloued.dz ,ezziyyani@gmail.com,kazarokba@yahoo.fr, m.derdour@yahoo.fr }

Abstract. The massive diffusion of Cloud services in the internet led to increasing demands of services and cloud infrastructure are conduct to an increasing in energy consumption in data centers, which has presents interesting great challenge. In this paper, we propose new strategy of allocation resources, with intelligent management of resources our approach based on constraint programming (CP), according to the autonomous resources allocation using multi-agent systems (SMA). At first, we make a review of various solutions that have been proposed in order to optimize Cloud energy consumption; second, we offer a logical solution to manage physical and virtual resources in smarter data center. At the end, we conclude this paper by future work.

Keywords: Cloud Computing; SMA; multi agent system; allocation resources; resources management; energy consumption; smart data center.

1 Introduction

A cloud computing it is one of the latest technology used via the Internet, the major advantage of cloud service is the elasticity that allows a dynamic change in the number of resources based on the varying demand from a customer as well as a pay-as-you-go opportunity, both of which can lead to substantial savings for the customers. Appropriate management of resources in clouds is essential for effectively harnessing the power of the underlying distributed resources and infrastructure. The problems range from handling resource heterogeneity, allocating resources to user requests efficiently as well as effectively scheduling the requests that are mapped to given resource, as well as handling uncertainties associated with the workload and the system. As a researcher, one can understand the opportunities to dig further to carry-on with more innovations to contribute better solutions to the existing problems. On other hand, the remarkable increasing demands on cloud services increasing expansion of datacenter to huge datacenter that led to a rise in energy consumption by datacenter the create negative factors affect the environment. According to (Barroso

2013)[1], a typical data center needs 10 megawatts of power to operate. In 2011, the energy consumption of Datacenters is expected to exceed 100000000.00 kilowatt per hour (kWh).By 2013, the Datacenter in the US consumed 91 billion of electricity (kilowatt per hour); equivalent to the annual output of 500 megawatts, costing US businesses 13$ billion per year for electricity bills and generating nearly 100 million tons of carbon pollution per year.

The purpose of our research is to serve a smart configuration of the cloud infrastructure based on energy consumption; it mainly has a scientific interesting in the field of cloud to show the concept of smart management of resources in cloud datacenter and its importance on minimizing energy consumption, and maximizing the benefits of cloud service provider. In this paper, we propose new allocation strategy for cloud data center based on multi agent system and constraint programming. At first, a management system design of cloud infrastructure is put forward based on allocation of resource and energy consumption constraint. After that, we involve a multi-agent system to shift traditional resources management systems to autonomous resources management systems.

The rest parts of this paper are organized as follows. In section 2, we present the resource management in Cloud Computing and related works. In section 3, describes our cloud architecture with integration multi agent system. Section 4 describes our Allocation strategy based constraint programming. While the section 5, last section concludes our paper and presents some perspectives.

2 Resource Management in Cloud Computing

When discussing resources, a difference should be made between data-center resources such as available servers, storage space and network bandwidth and computing resources directly available to mobile devices. At data-center level there are multiple tiers for resource allocation and optimization: at cluster or supercomputer, virtual machine and operation system disk image levels. Moreover, for any of those levels different objectives can be pursued: increasing power usage efficiency, increasing or insuring a predefined level of availability for provided services, increasing performance, lowering the data-center air conditioning costs or a combination of them[2].

Resource allocation is the process of distributing available resources between the various applications running in a cloud environment. There are several problems addressed by an optimal resource allocation [3]:

- Resource contention: multiple consumers are trying to gain access to the same resources at the same time
- Scarcity of resources: there are limited resources. Different cloud implementations must cope with different levels of resource scarcity. The problem is of bigger importance for private clouds.
- Resource fragmentation: the resources are isolated. Even when there are enough resources, their potential consumers cannot gain access to them.

- Over-provisioning: resources are reserved for a client’s exclusive use in quantities exceeding its needs.
- Under-provisioning: opposed to over provisioning, not enough resources are reserved for exclusive use of a client.

3 Literature Review

This section presents our interested research about cloud data center architecture and allocation algorithms, according to the literature, the power consumption has attracts a lot of research s in the past few of years. Whereas, Liu et al [4], present Green Cloud architecture, which aims to reduce data center power consumption while guaranteeing performance from a user’s perspective. Using the recommendations developed in its open-source Cloud standards’ incubator. Tian et al [5]., proposed a dynamic and integrated load balancing algorithm for resource scheduling in Cloud data centers. Fumiko Satoh et al [6], also focus on reducing the usage of energy in data centers. But for the future energy management they develop an energy management System for cloud by the use of sensor management function with an optimized VM allocation tool. This system will help to reduce the energy consumption in multiple data centers and results shows that it will save 30% of energy. This system also used to reduce the energy in carbon emissions. Rasoul.B et al [7], proposed software architecture that calculates the energy consumption in datacenter and provide services to the users which uses energy efficiently. Beloglazov et al[8],focuses on virtual machine for the reduction of the energy consumption. An author proposed the dynamic reallocation technique for VMs and toggles off the unused servers which results, considerable energy saving in the real Cloud Computing data centers. Hulkury and al[9], used The Green Broker uses these directories and chooses the green offer and energy efficiency information and allocates the services to the private cloud. And finally give the result to the users. Garg and al [10], proposed a new architecture called as integrated green Cloud architecture (IGCA). It smartly includes client oriented in the Cloud Middleware that verifies the cloud computing is better than the local computing with QoS and budget.

4 Cloud datacenter Architecture

This section, we will explain the combination of our multi-agent system under the Cloud environment. Our SMA is dedicated, where 3 agents: Analyzer agent, agent scheduling, Controller agent. The figure above, showed the role of each agent in different layers of Cloud Datacenter. Indeed, the proposed architecture consists of three layers of the Cloud: the application layer, the network layer and the Datacenter layer. In which the figure expresses the relationship between the Cloud layers and the function of each agent in each layer.

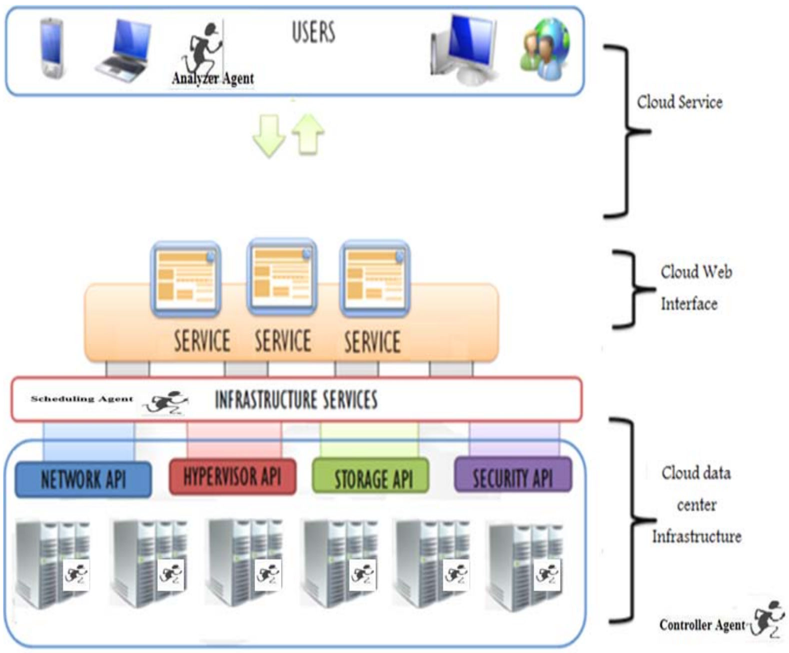


Fig 1: Intelligent architecture for cloud system

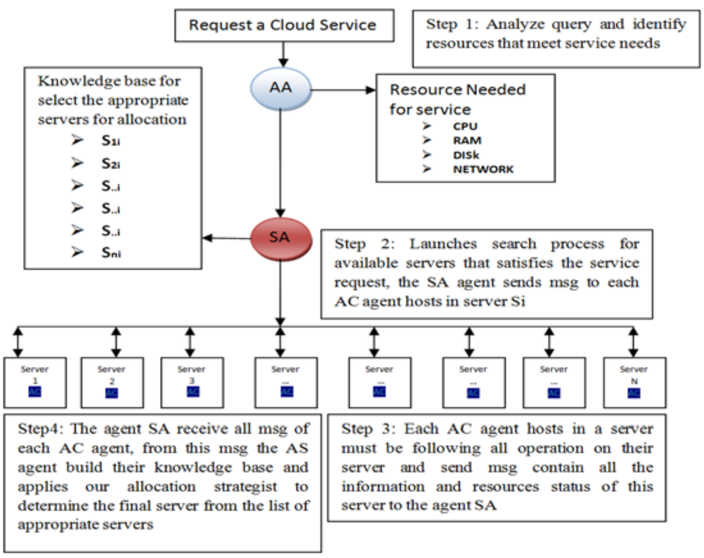


Fig 2 : communication between agents

If the user of service cloud select service and send request to our system, our system in time ti launches analyzer agent (AA) that last analyze user request and identify all resource needed like (CPU speed, RAM capacity, disk storage …..), when this agent finished her job send anther request to scheduling agent (SA), this agent depends on knowledge base to select and allocate resource, so at first, to select server for final allocation, this agent needed to create or update their knowledge base to select the appropriate servers for allocation ,so to create knowledge base SA agent launched research process for available servers that satisfies the service request, the SA agent sends msg to each AC agent hosts in server Si, Each AC agent hosts in a server must be following all operation on their server and send msg contain all the information about resources status in this server to the agent SA, The agent SA receive all msg of each AC agent, from this msg the SA agent build their knowledge base and applies our allocation strategist to determine the final server from the list of appropriate servers .

4.1 Agents description

This subsection includes a brief description of agent's components that are used for an efficient cloud management and the resources allocation. Some properties must be considered to adequately perform the cooperative activity Autonomy: An agent is said to be autonomous if it is able to make decisions for performing additional actions, or for changing its current task. Perception: the capacity of an agent to perceive its environment and consequently to update its mental state. Auto-Organization: an agent is able to auto-organize himself, when it has the capability to evaluate his interactions with others and add organizational links or remove some of them.

Analyzer Agent (AA)

An analyzer agent examines and extracts specifications from client requests to identify the major parameters: CPU speed, RAM, storage capacity, etc. Subsequently, these specifications will be communicated to the network agent. Figure 2 shows the architecture of the analyzer agent, on which the input is for client requests an acquiring, while output interface is designed to release the aforementioned specifications to the SA agent.

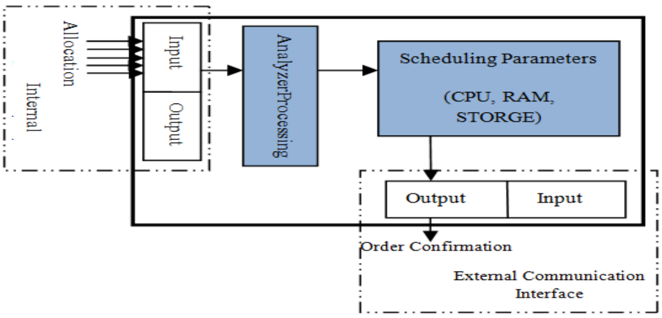


Fig 3 : Analyzer agent configurations

Scheduling Agent SA

To achieve an efficient allocation, the scheduling agent figures prominently position when synchronization processing and task performance presents its main purpose. Along with planning, these specifications of the allocation must comply with constraints defined by the allowance algorithm in order to maintain both energy consumption and the execution cost. Likewise, the SA agent collaborates with other agents in which planning decisions are endowed. The figure 3 shows a schematic representation of the SA agent. Where it has two communication interfaces, including external interface to communicate with the AC agent, while the external interface to communicate with the AA agent.

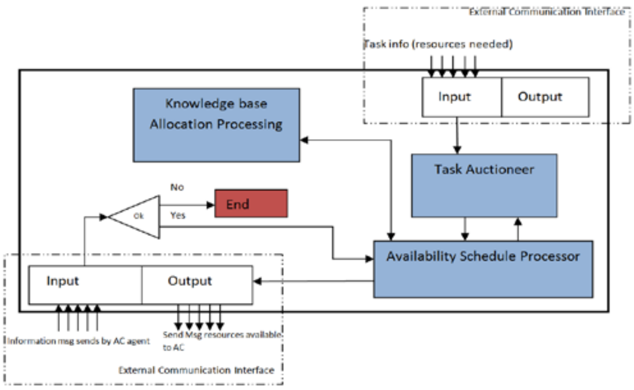


Fig 4. Scheduling agent configurations.

Controller Agent CA

This agent hosts in all work server in datacenter, his is intended to trigger monitoring method when resource availability statements are requested. For providing monitoring data based on the arrival a request, the AC agent provides a communication interface that carries monitoring method outcomes to the SA agent. Figure … shows a schematic representation of the CA agent. Whose internal external enables the communication with the SA agent, while the internal interface ensures communication and controlling and monitoring all resources in server Si.

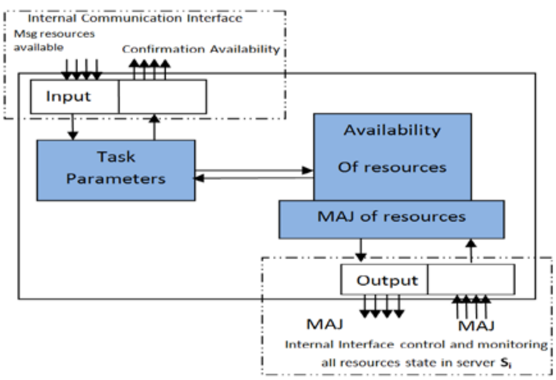
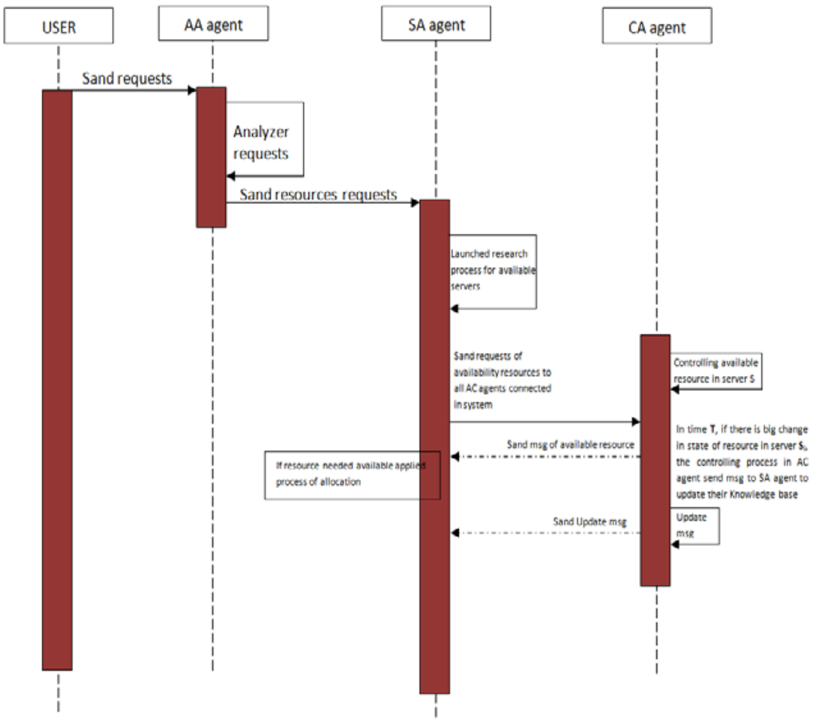


Fig5. Controller agent configurations.

Fig 6. Show interaction between agents



5 Strategy of allocation

1. Problem description

The hardware architecture consists of a set P ={p1, . . . , pk, . . . , pm} of m identical processors with fixed memory capacity mk and identical processing speed. They are all connected to a network with bandwidth δ .

A task T_i is defined through its temporal characteristics and resource needs: its period Ti (as a task is periodically activated), Tasks are periodically activated in an independent way, and they read and write data at the beginning and the end of their execution.

Finally, each processor is scheduled with a fixed priority strategy. A priority P_i is given to each task.

An allocation is a mapping A: $T \longrightarrow P$ such that the image of a task Ti is a processor p_k:

$$Ti \longrightarrow A(Ti) = p_k$$

The allocation problem consists in finding the mapping A which respects the whole set of constraints described in the immediate below. There are three classes of constraints the allocation problem must respect: timing, resource, and allocation constraints.

- **Allocation constraints:**

This set of constraints deal with the position (or relative position) of the tasks on the processors. Some tasks require specific processor characteristics to be executed (signal processor, compression processors, databases, etc.) and can only reside on a subset of the available processors. Others must not be put together on the same processor. Two sets of tasks may also have to be disjoint in any assignment [11].

- **Resource constraints:**

The memory usage of a processor cannot exceed a fixed capacity [10].

- **Timing constraints:**

A hard real-time system must respect all timing constraints to assure the security of the process. Temporal constraints define a schedulable allocation according to deadlines or due dates requirements [11].

An allocation is said to be valid if it satisfies allocation and resource constraints. It is schedulable if it satisfies timing constraints. Finally, a solution to our problem is a valid and schedulable allocation of the tasks.

2 Our solution for Problem of allocation

Constraint programming (CP) techniques have been widely used to solve a large range of combinatorial problems. A *constraint satisfaction problem* (CSP) consists of a set V of variables defined by a corresponding set D of possible values (the so-called *domain*) and a set C of constraints. A solution to the problem is an assignment of a value in D to each variable
in V such that all constraints are satisfied. This mechanism coupled with a backtracking scheme allows the search space to be explored in a *complete way*. For a deeper introduction on CP, we refer to [12].

We propose an approach based on multi agent system and integrate constraint programming, for this, we suppose a data center D consists of a group of physical servers S; wherein each server $Si \in S$ characterized by a number of processors (nbCpu); RAM capacity (RamC);storage unit (Disk); limited bandwidth(Net BW); a set of virtual machines (Vij). At virtual level, for each virtual machine $vj \in Vij$ includes virtual resources such as nb VCpu; RAMs and Disks.

Nevertheless, to achieve to an effective management, we associate these parameters by precondition equations as the following:

The Cpu allocation for the VMs must not be exceeded the entire number of physical Cpu on the server, as shown on the Eq.1.

$$\sum_{j=1}^{n}(vj.nbVCpu) \leq (S_i.nbCpu) \ldots\ldots\ldots\ldots\ldots (1)$$

On only one server, memory utilization rate by the VMs must not exceed the total rate of available RAM capacity on this server

$$\sum_{j=1}^{n}(vj.RamC) \leq (Si.RamC) \ldots\ldots\ldots\ldots\ldots\ (2)$$

The utilization rate of storage space by VMs on the server must not exceed the available memory space on the hard disk.

$$\sum_{j=1}^{Vj}(vj.VDisk) \leq (Si.Disk) \ldots\ldots\ldots\ldots\ldots..(3)$$

The number of Vji allocated by the server Si must not be exceeded approved VMs allowed by this server.

$$V_{ij} \leq \left(S_i.\mathrm{maxV_j}\right) \ldots\ldots\ldots\ldots\ldots\ (4)$$

Step 1: inputs initialization

```
  Nbr_ PM:  number of physique server in the data center;
  Nbr_ VM:  number of virtual server in the data center;
  LV : list of running VMs in each server on data center;
  LV.V : list of available VMs in each server on data center;
  LS.V : list of available servers in data centers;
  L: list of available server Ready for VM (Ri) allocation and Table of
all tuple M;
  M: The constraint to selecting server for allocation VM ;
  Free (): searching free resource in Server Si
  AC_Request: denotes a function that extracts the allocation request
                    specifications such as the identifying needed
                    resources;
```

Step 2: Analyzed Request

```
  If user send their request (user.request)
  AA agent (AC_Request)
  {
  AC_Request= Ri;
  AC_Request(Ri);
  Send(Ri)to SA agent ;
  }
```

Step 3: VM Allocation

```
  If (SA Receive Request (Ri)) SA launched process to create their
knowledge base
       {
  Send msg (Ri) (Discovery available resources in datacenter for Ri
request)

       }
  If (CA Receive msg (Ri)) launched search process
  {
  Nbr_ PM();
  Nbr_ VM ();
```

```
LV();
LV.V ();
LS.V ();
Send L(Ri)to SA agent
}

If (AC.free ()not null)
{
Sand msg (AC.Free) to SA agent SA agent update knowledge base
}
If (SA Receive L(Ri)) receive list of available Sj ready to allocate Ri
request

If there is a list of servers suitable for the allocation then
For each Si in Selected SerList
{
PiSj = Calculate power consumption (Ri);

}
Min Energy Consumption(PiSj)
Allocate a VM with lowest energy consumption.
Else

There is no ready server for allocation;
```

6 Conclusions

In this paper, we introduce to the smarter cloud resource management, where it mainly based on multi-agent system and constraint programming (CP). Before allocation, we were proposing a precondition constraint aimed select a ready server for the request allocation. Whereas, energy consumption criteria has been selected to identify a preferment server in order to satisfy the user request. As key feature, we involve 3 agents for autonomous cloud resources configuration and improve the allocation policies. In the future work, we attempt to simulate a real scenario by utilizing OMNET++ simulator and iCanCloud framework.

7 Reference

1 Baldonado, M., Chang, C.-C.K., Gravano, L., Paepcke, A.: The Stanford Digital Library Metadata Architecture. Int. J. Digit. Libr. 1 (1997) 108–121

2 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, no. 6, pp. 97-104, 2012

3 Andrei IONESCU "Resource Management in Mobile Cloud Computing" Informatica Economică vol. 19, no. 1/2015

4 Liu L, Wang H, Liu X, Jin X, He WB, Wang QB, et al. GreenCloud: a new architecture for green data center. Proceedings of the sixth international conference industry session on autonomic computing and communications industry session, ICAC-INDST'09. New York, NY: ACM; 2009. p. 29_38.

5 Tian WH, Zhao Y, Zhong YL, Xu MX, Jing C. A dynamic and integrated load balancing scheduling algorithm for Cloud data centers. In: the proceedings of CCIS2011, Beijing.

6 F. Satoh, H. Yanagisawa, H. Takahashi and T. Kushida, (Eds.), "Total Energy Management system for Cloud Computing", Proceedings of the IEEE International Conference of the Cloud Engineering (IC2E), (2013), March 25-27; Redwood City, CA.

7 R. Beik, (Ed.), "Green Cloud Computing: An Energy-Aware Layer in Software Architecture", Proceedings of the Spring Congress of the Engineering and Technology (S-CET), (2012), May 27-30; Xian.

8 A. Beloglazov and R. Buyya, (Eds.), "Energy Efficient Allocation of Virtual Machines in Cloud Data Centres", Proceedings of the 10th IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGrid), (2010) May 17-20; Melbourne, Australia.

9 M. N. Hulkury and M. R. Doomun, (Eds.), "Integrated Green Cloud Computing Architecture", Proceedings of the International Conference on Advanced Computer Science Applications and Technologies (ACSAT), (2012), Washington DC, USA.

10S. K. Garg, C. S. Yeo and R. Buyya, (Eds.), "Green Cloud Framework for Improving Carbon Efficiency of Clouds", Proceedings of the 17th International European Conference on Parallel and Distributed Computing (EuroPar), (2011) August-September. Bordeaux, France.

11 R. Bart´ak. Constraint programming: In pursuit of the holy grail. In proc of WDS99 1999

12 Pierre-E." How to Solve Allocation Problems with ConstraintProgramming" 2005