Agent Based Framework for Scalability in Cloud Computing

Aarti Singh[1], Manisha Malhotra[2]

[1]Associate Prof. , MMICT & BM, MMU, Mullana
[2]Lecturer, MMICT & BM, MMU, Mullana

Abstract: Cloud computing focuses on delivery of reliable, secure, fault-tolerant, sustainable, and scalable infrastructures for hosting internet-based application services. These applications have different composition, configuration, and deployment requirements. Cloud service providers are willing to provide large scaled computing infrastructure at a cheap prices. Quantifying the performance of scheduling and allocation policy on a Cloud infrastructure (hardware, software, services) for different application and service models under varying load, energy performance (power consumption, heat dissipation), and system size is an extremely challenging problem to tackle. This problem can be tackle with the help of mobile agents. Mobile agent being a process that can transport its state from one environment to another, with its data intact, and is capable of performing appropriately in the new environment. This work proposes an agent based framework for providing scalability in cloud computing environments supported with algorithms for searching another cloud when the approachable cloud becomes overloaded and for searching closest datacenters with least response time of virtual machine (VM).

Keywords: Cloud Computing, Cloud Service Provider, Data Centers, Mobile Agent, Scalability.

1 Introduction:

Cloud computing is a paradigm that focuses on sharing data and computations over a scalable network of nodes, spanning across end user computers, data centers, and web services. A scalable network of such nodes forms a cloud. An application based on these clouds is taken as a cloud application. In recent years, most of the developed softwares are based on distributed architecture, such as service-oriented, P2P & cloud computing. With the development of computer hardware and networking, distributed architectures have also grown, especially service-based cloud computing has changed the traditional computer and centralized storage approach. It facilitates processing and storage capabilities as per the requirement. Infrastructure-as-a-Service (IaaS), provides Virtual Machines (VMs) fully satisfying the user requests in terms of resources. The resources of the providers are usually hosted in the form of a data center. The data center is a set of physical machines which are interconnected, virtualized, and geographically distributed. Since the customer may have different geographic location, a service provider should have distributed data centers throughout the world so as to provide services to the customers. In the cloud computing, the distance between datacenters leads to undesirable network latency, which in turn leads to delay in services. For example, a VM allocated in a data center, far away from customer location, the customer will suffer from delayed response due to the limitation of network resources. In addition, if the VM is heavily loaded it increases the response time to the customer , compared to if VM is having less workload. Thus, a provider is required to find suitable data centers for serving a customer based on the user location and workload of the data centers. Figure 1 given below provides the architecture of a data center. This will involve knowledge about existing data centers as well as dynamically checking status of their load whenever a request has to be served. This job can be performed well by employing mobile agents in clouds as they can move from one location to the other on the network without constraining network bandwidth. They can transport their state from one environment to another, with their data intact, and are capable of performing appropriately in the new environment. They also maintain the information about their cloud i,e resource entropy, capacity, size etc and can collect the same information from other clouds as per requirement.Scalability of the cloud services is an important concern. It should be transparent to users, so that users may work on a cloud requesting for services without bothering about how and from where they are

provided. For example, every cloud has only a finite amount of physical storage entities. Therefore, a cloud c1 may seek help from another cloud c2 for shared storage to fulfill some request on storage. Such sharing requirement may result in the data to migrate among multiple clouds. Similarly, a user might request for some services when resources of the cloud has already been exhausted, in that case rather than refusing the customer for the services, request should be redirected to some other cloud having the desired resources.

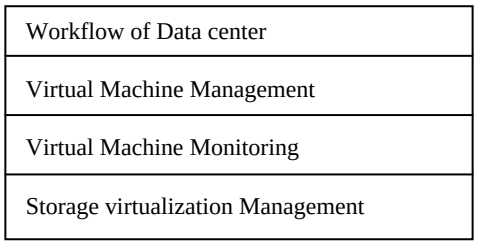

Fig 1: Architecture of data center

This work aims to focus on problems associated with scalabilty of cloud services and to propose an agent based mechanism for this problem. This paper is structured as follows: section 2 provides the review of relevant literature, section 3 provides the proposed work, and section 4 concludes the paper.

2 Related Work:

Zhang et al. [1] proposed an intelligent workload factoring service for enterprise customers to make the best use of public cloud services along with their privately-owned (legacy) data centers. The core technology of the intelligent workload factoring system is a fast frequent data item detection algorithm, which enables factoring incoming request only. It is only applied to data access. Amoretti et al. [2] presented a framework and a middleware that enables highly dynamic and adaptive cloud, characterized by peer providers and by services that can be replicated by means of code mobility mechanism. M. Stillwell [3] describes the heuristic resource allocation approach under homogeneous virtualized cluster environment. They assume the VM represents one computational job. So, they have described allocating VMs by the required resource such as number of CPUs for the jobs. N. Fallenbeck [4] proposed the approach for both serial and parallel job in virtualized environment dynamically. However, even though previous works consider different constraints for allocation, there is less consideration in terms of the geographical location of consumer and data center. J. Rao et al [5] proposed a reinforcement learning based approach to determine the memory and storage requirements of a VM. C. Zeng [6,7] delineate cloud service composition solution based on the agent paradigm, which has proved to be effective for distributed applications where no exact and complete system snapshot can be determined given the inherent dynamicity of the environment. Foster [8] highlighted the services distribution for the consumers by service providers in different geographical locations. [9,10] proposed a dedicated connection between the VM for virtual networking independence in real and virtual workspace. [11,12] delineate various applications and extensions designed to create virtually unique work environments in which resources are deployed in multiple environments. From literature review, it has been observed that scalability problem has not been paid much attention in cloud computing and there is scope of research in this direction. Next section provides the proposed agent based framework for ensuring scalability in cloud computing.

3 Proposed Work

This work focuses on ensuring scalability in cloud computing in situations where either the resources of the cloud have been exhausted and it can not provide services to more customer or the requested resources are not available with it. This work is being accomplished in two parts: first is to search another community cloud to satisfy the request in hand and secondly to search for closest datacenters with least response time of virtual machines (VM). The proposed framework makes use of mobile agents to achieve the goal. Mobile agent being a process that can transport its state from one environment to another, with its data intact, and is capable of performing appropriately in the new environment. The proposed framework associates a mobile agent with each public/ private cloud, which contains the information about that cloud such as various resources available with the cloud and it also keeps track of free and allocated resources. Thus, whenever a service request arrives to a cloud mobile agent checks the available free resources to decide whether the request can be served or not. The proposed framework comprises of cloud mobile agents and directory agent as shown in Fig. 2 given below.
Cloud mobile agent (MA_C): It is associated with every cloud and is responsible for maintaining resource information as well as their status at any point in time (whether free or allocated).

Directory Agent (DA_C): DA_C keeps record of all MA_C registered with it, along with capabilities of the clouds. Whenever a cloud is created its MA_c will have to get registered with a DA_c, which maintains their database necessary for providing scalability in service.

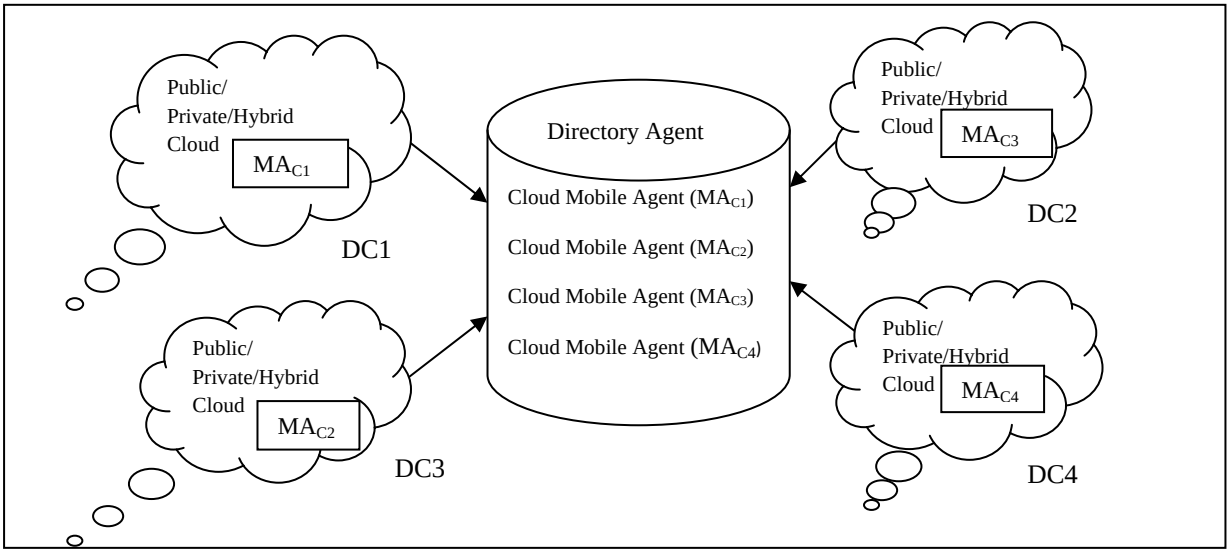

Fig 2: High level view of agents in proposed framework

Intially the MA_c sends a registration request to nearest DA_C along with information of the cloud with which it is associated, in response the DA_C sends back an acknowledgement signal indicating that the MA_c has got register with it. Fig. 3 given below explains this process.

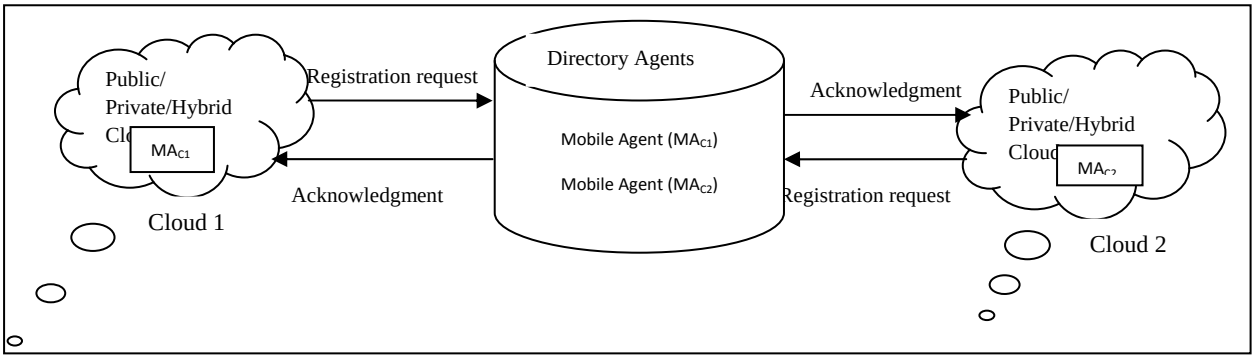

Fig 3: Registration Process

Now whenever a public/private cloud becomes too overloaded to handle another user request then the scalability feature is exercised to provide services to the user through some other cloud. In such situation MA_c gets activated and sends request to directory agent demanding for the list of other MA_c capable of providing the desired service. The directory agent on receiving this request searches its database and provides the list of competent MA_cs. On receiving the list from the DA_C the initiator MA_c sends service request to them and waits for their response. If some cloud possessing the required resources becomes ready to provide the services, it responds back to the initiator MA_c. Initiator MA_c scans all the received responses, performs negotiation with concerned MA_cs and then finally assigns the in hand request to the MA_c most suitable both, in terms of less cost and faster services.

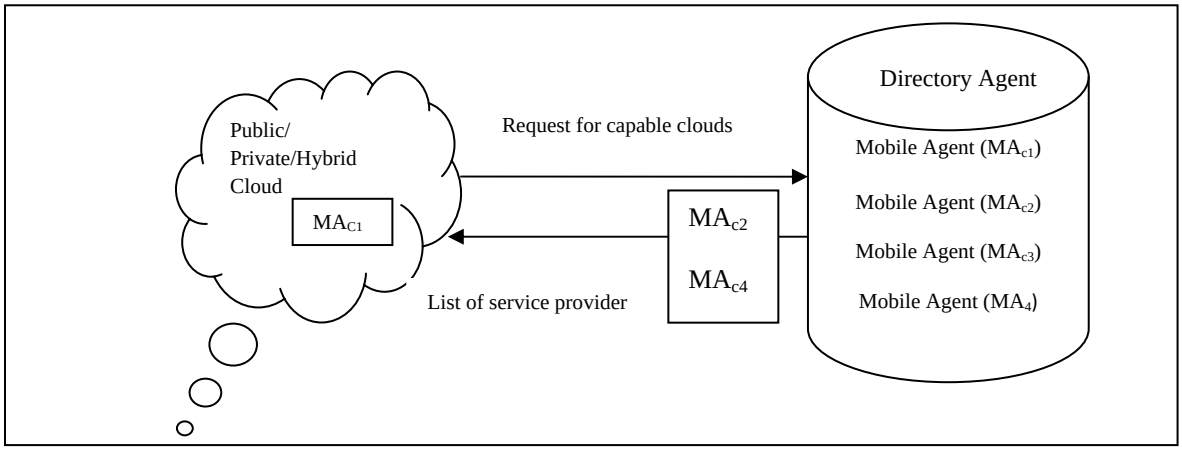

Figure 4: Process of searching for clouds capable of providing desired servies using directory agents

Next section provides the algorithms for the cloud mobile agent MA_c and the directory agent DA_C.

3.1 Algorithms

The algorithm for cloud mobile agent MA_c and the directory agent DA_C are as given below:

Cloud Mobile Agent (MA_c)

```
MAc( )
Input: service_request
Output:     service_delivery_acceptance          or
redirection
{ Call MAc_Intialization()
  On (service_request)
 { check available resources
    If (available resources > requested resources)
       Return (service_delivery_acceptance)
    Else
       Call redirection()
}
}
MAc_Intialization()
{
    Intialize registration with DAc
    Send cloud information to directory agent.
    Receive acknowledgement
   return ( )
 }

Redirection()
{
  Send request to DAc for providing list of MAcs
with desired resources
  Receive list of MAc s from DAc
 Send request for providing services to potential
MAcs
 Receive responses from the potential MAcs
 Perform negotiation
 Assign service task to most suitable MAc
 Redirect all further communication from  the user
to serving MAc
}
```

Di

```
DAC ()
Input: registration request, search_request_
for_ service_providers
Output: acknowledgement, list of MACs
{
  If (registration request)
 { receive specification of cloud capabilities
   Create entry for the new MAC
   Send an acknowledgement to requesting MAC
indicating registration
 }
If (search_request_ for_ service_providers)
{ receive the desired resources
   Search the database for the desired resources
   Return the list of all suitable MACs
}
```

4 Conclusion:

In this paper we have proposed an agent based framework to ensure scalability in cloud computing environments. Scalability of services is a desired feature in cloud environments and it can be achieved by employing mobile agents. Algorithm for implementing the agents involved are also provided, however implementation of this framework is still under progress. Future work aims to achieve this and also to extend this framework for ensuring reliability of the mobile agents involved.

Refrences:

[1] Zhang, H., Jiang, G., Yoshihira, K., Chen, H., Saxena, A.: Intelligent Workload Factoring for a Hybrid Cloud Computing Model. In California : International Workshop on Cloud Services, Los Angeles, July, 2009.

[2] Amoretti, M., Laghi, M. C., Tassoni,F., Zanichelli, F.: Service Migration within the Cloud: Code Mobility in SP2A. Proceedings of International Conference on High Performance Computing and Simulation (HPCS),2010, pp.196-202.

[3] Stillwell, M., Schanzenbach, D., Vivien, F., Casanova, H.: Resource Allocation using Virtual Clusters. Proceedings of the 9th IEEE Symposium on Cluster Computing and the Grid (CCGrid'09), May 2009..

[4] Fallenbeck, N., Picht, H.J., Smith, M. and Freisleben, B.: Xen and the Art of Cluster Scheduling. In Washington: 2nd IEEE International Workshop on Virtualization Technology in Distributed Computing (VTDC '06).

[5] Rao, J., Bu, X., Xu, C.Z., Wang, L. and Yin Vconf., G.: A Reinforcement Learning Approach to Virtual Machines Auto-Configuration. Proceedings of the 6th intenational conference on Autonomic computing, 2009.

[6] Zeng, C., Guo, X., Ou, W., and Han, D.: Cloud Computing Service Composition and Search Based on Semantic. In Berlin, Heidelberg: International Conference on Cloud Computing, , Eds. LNCS 5931, Springer-Verlag, 2009, pp. 290-300.

[7] Zou, G., Chen, Y., Yang, Y., Huang, R. and Xu, Y.: AI Planning and Combinatorial Optimization for Web Service Composition in Cloud Computing. Proceeding of the International Conference on Cloud Computing and Virtualization, 2010.

[8] Foster, I., Yong, Z., Raicu, I., Lu, S.:Cloud Computing and Grid Computing 360-Degree Compared. Workshop on Grid Computing Environments, 2008. GCE '08 , vol., no., pp.1-10, 12-16 Nov. 2008.

[9] Sundaraj, A., Dinda, P.: Towards Virtual Networks for Virtual Machine Grid Computing. In USA: 3rd conference on Virtual Machine Research And Technology Symposium - Volume 3 pp. 177-190, 2004.

[10] Keahey, K., Doering, K., Foster, I.: From Sandbox to Playground: Dynamic Virtual Environments in the Grid. In: 5th IEEE/ACM International Workshop on Grid Computing, 8 Nov. 2004 pp. 34- 42.

[11] Zhang, X., Keahey, K., Foster, I., Freeman, T.: Virtual Cluster Workspaces for Grid. Online available on: http://www.nimbusproject.org/files/VWCluster_TR_ANL_MCS-P1246-0405.pdf

[12] Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., Zaharia, M.: Above the clouds: A Berkeley View of Cloud Computing. Technical Report, February 2009.

[13] Buyya, R., Yeo, C.S., Venugopal, S., Broberg, J., Brandic, I.: Cloud computing and Emerging IT Platforms: Vision, Hype, and Reality for Delivering Computing. 5th utility Future Generation Computer Systems, Volume 25, Issue 6, June 2009, pp. 599-61.

[14] Sim, K.M., Shi, B.: Adaptive Commitment Management Strategy Profiles for Concurrent Negotiations. In Estoril Portugal :1st International Workshop on Agent-based Complex Automated Negotiations (ACAN) 2008, , held in conjunction with 7th Int. Conf. on autonomous agents and multi-agent systems (AAMAS), pp.16-23.

[15] Kang, J., Sim, K.M.: A Multi-criteria Cloud Service Search Engine. In Hangzhou, China : Proceeding of IEEE Asia-Pacific Services Computing Conf., Dec.6 - 10, 2010,.

[16] Gutierrez-Garcia, J.O., Sim, K.M.: Self-Organizing Agents for Service Composition in Cloud Computing. In USA: 2nd IEEE International Conference on Cloud Computing Technology and Science, 2010.

[17] Gutierrez-Garcia, J.O., Sim, K.M.: Agent-based Service Composition in Cloud Computing. In Jeju Island Korea: Proceeding of on Grid and Distributed Computing conference, December 13-15, 2010.

[18] Zhou, Y., Yang, Y., Liang, L., He, D., Sun, Z. : An Agent Based Scheme for Supporting Service and Resource Management in Wireless Cloud. Proceeding of 19th IEEE International Conference on Grid and Cloud Computing by pp.34-39.

[19] Ramaswamy, A., Balasubramanian, A., VijayKumar, P.: A Mobile Agent Based Approach of Ensuring Trustworthiness in the Cloud. In Chennai: IEEE-International Conference on Recent Trends in Information Technology, 2011 June 3-5 pp. 678-682.