ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 9, September 2017

Approach of intelligent agent in Cloud resource management

[1] Manoj Sharma, [2]Dr. Manoj Kumar Sharma, [3] Dr. S. Srinivasan
[1] Research Scholar, Department of C.S.E., Suresh Gyan Vihar University Jaipur, Rajasthan
[2] Professor, Department of C.S.E., Suresh Gyan Vihar University Jaipur, Rajasthan
[3] Professor, PDM University, Bahadurgarh, Haryana.

***Abstract* - Rapid technological developments in Information Technology (IT) and ubiquitous Internet access are causing serious challenges in service provisioning and resource management landscapes. Cloud computing is proving to be a reliable technology to address these challenges. Service provisioning in the Cloud relies on Service Level Agreements (SLAs) representing a contract signed between the customer and the service provider including non-functional requirements of the service specified as Quality of Service (QoS) and penalties in case of violations. Flexible and reliable management of resources and SLA agreements are of paramount importance to both Cloud providers and consumers. On the one hand, providers have to prevent SLA violations to avoid penalties and on the other hand, they have to ensure high resource utilization to prevent costly maintenance of unused resources In this paper, we propose a novel Cloud management infrastructure, which is based on holistic monitoring techniques and mechanisms for low-level resource metrics to high-level SLA mapping, application scheduling and deployment, and the ability to monitor multiple application executing on the same host.**

Index Terms— Cloud computing; intelligent agent; load balancing; fitness value; load percentage;

1. INTRODUCTION

Cloud Computing is a reference model to provide IT infrastructure and applications as a service in a scalable manner. It is aimed at optimizing the resource utilization according to customized SLAs by means of virtualization and sharing resources. One formal definition was given in a report from the University of Melbourne as: "Cloud Computing is a type of parallel and distributed system consisting of a collection of interconnected and virtualized computers that are dynamically provisioned and presented as one or more unified resources based on a service level agreement".
Cloud computing comprises three IT layers: infrastructure, platforms and software. All of them are delivered as services over the Internet, hence providers are classified as:

- Infrastructure as a service (IaaS): offering web-access to processing, storage or connectivity. The key point is that hardware resources are abstracted and encapsulated by virtualization. However, end users have control over operating systems, storage and applications.

- Platform as a Service (PaaS): offering easy development environments, reusable components, libraries, collaboration services and workflow facilities to design, develop, test, deploy and host applications.

- Software as a Service (SaaS): offering services directly consumable by end- users. The main difference from conventional software suppliers is the deployment, licensing and billing model.

The Cloud involves three major actors: the Cloud user, the Cloud vendor and the original Cloud provider. The cloud Vendor is an organization that has a local tax registration and provides the Cloud services to the Cloud user according to expected levels of quality of experience (QoE) and quality of service (QoS) based on a service level agreement (SLA). However, the Cloud vendor is not necessarily the owner of the infrastructure, platform or software offered to the Cloud user. Here appears the third actor, the original Cloud provider, who is the organization who owns the SaaS, PaaS or IaaS. Clouds which are available to consumers in a pay-as-you-go manner are called Public Clouds or external clouds, while clouds which are owned, controlled and used by a single organization are denominated as Private Clouds or internal clouds. Single clouds from these types can be combined to get Hybrid Clouds and Federation of Clouds. Hybrid clouds are an alternative to get scalable IT resources provided by external clouds and to get security and privacy provided by internal clouds, at the same time. Federation Clouds are a group of single clouds collaborating to exchange data and processing resources through interfaces.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 9, September 2017

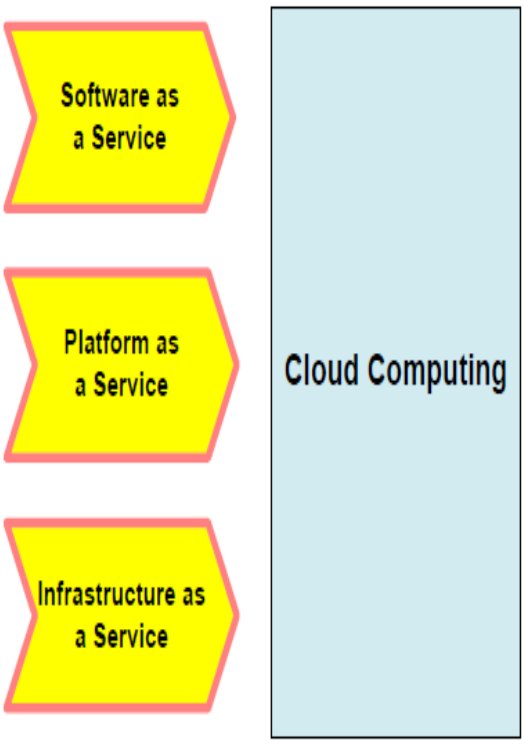

Figure 1 Cloud models

Market-Based Cloud

A Market-Based Cloud is a Cloud environment implemented according to the definition for a Market-Based System. The main components in a Market-Based Cloud are:

1.***Cloud users*** request job executions at an expected QoS parameters specified by a SLA,

2.Cloud providers or sellers own processing resources to trade,

3.Cloud brokers or buyer perform the Cloud trading on behalf of the Cloud users to select the most suitable Cloud provider for the job.

4.Market mechanism: establishes a set of rules to the trading interactions between buyers and sellers in order to define how buyers quote bid prices and seller quote offer prices.

This approach is aimed at analyzing to what extent the resource allocation based on market-based controls and reputation metrics can minimize SLA violation in Cloud architectures.

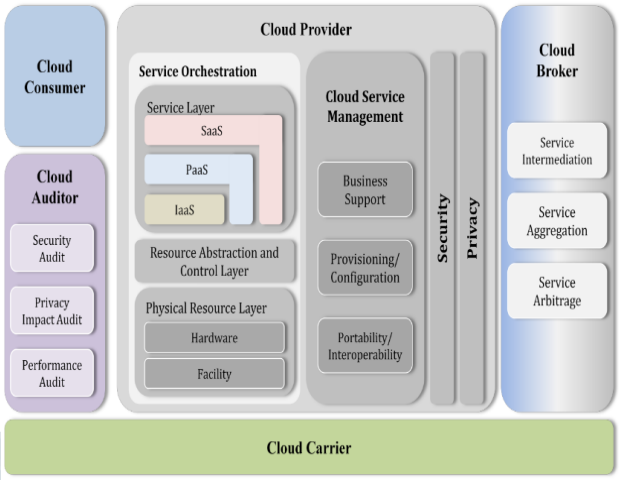

Figure 2: NIST Cloud Computing Reference Architecture

There are two approaches for managing resource allocation: centralized and decentralized [4]. In the first one, resource allocation is performed by a single node or a coordinator which is responsible for analyzing the user requirements and evaluating the current Cloud architecture configuration in order to decide how to optimize the utilization of resources from a global perspective. The key disadvantages for this classical approach include:

•The Cloud architecture must be static while the central coordinator is carrying out the calculus of the resource allocation algorithm

•The central coordinator knows the global state of the resources in the Cloud and

•The central coordinator must be connected to every single resource which leads to scalability problems.

On the other hand, a decentralized resource allocation mechanism is carried out by multiple self-interested agents. Each of them acts on behalf of a user at the decision making level while trying to maximize its own good without concern for the global benefit.

The main advantages of this mechanism are: no single point of failure, no performance bottleneck and the amount of exchange of messages to converge on an optimal solution. Some distributed trading mechanisms are auctions and bargaining.

II.RELATED WORK

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 9, September 2017

Information sharing involves platform-enabling data exchanging between senders and receivers [57]. Related works have proved the positive effects on lead-time, inventory and bullwhip effects which lead to redundant stock in SCM [19]. A multiagent system (MAS) is a system that consists of a number of agents whose key features are reactivity, sociality and proactivity.

These agents, able to cooperate, coordinate and negotiate, interact with one another by exchanging messages through some computer network infrastructure [10]. For the past years, certain works have kept assertive attitudes towards the cooperative/collaborative multi-agent models for they are helpful to information sharing in an autonomous and decentralized way [18,50] and seamlessly integrate heterogeneous components [20]. Jennings and Wooldridge [21] indicated that negotiatory or cooperative MAS technique is required in an open system which consists of different organizations, agenda and volatile guidance from users. The open system stated is just like an agile enterprise. Moreover, by comparing equation- based and agent-based modeling, Van Dyke Parunak et al. [5] concluded that MAS is most appropriate for domains with a high degree of localization, distribution and dominated by discrete decision. These domains certainly include SCM.

Communication between agents in the open systems (including the Internet) has been an issue [10]. Among the MAS based methods or applications, the peer-to-peer, hierarchical/clustering and integration approaches are mentioned in this research for further comparison. First, Vouros [53] proposed an efficient method combining both routing indices and token- based information to help search and share categorized information in a large-scale dynamic network in an efficient way. In this methodology, the large-scale dynamic environment, similar to our work, is also a key factor to information sharing. However, the undistinguished agents in [53] may limit its usage in specialized organizations within an enterprise. In this case, highly social clusters are present in the conjunction with functional site-roles and role-related information sharing processes that are predetermined in accordance with the assembling services whose order cannot be altered arbitrarily. Considering the limitation imposed on heterogeneous agents, Vouros also extended the work to [12], in which an overlay network was established by logical content providers and recommendation links between heterogeneous agents. Although this attempt further enhances the information sharing efficiency at an early stage, the overlaid topology/hierarchy-based peerto-peer approach incurs the costs upon index initialization or whenever changes occurred in the dynamic environment.

Next, the server-based hierarchy structure in the cooperative MAS is efficient because of the distinguished function agents planned. For applications using MAS to search and deliver information, Ardissono et al. [2] built a system with a recommendation engine which is adaptive to personalized layout and facilities. De Meo et al. [11] presented MAS for e-health services and compared three algorithms to search information. The stated web-based recommendation solutions are both user-centered deliveries in a proactive Business-to-Customer model, and their social attributes are emphasized as well. In the interaction designs, the interface agents were both built to collect the users' enquiries and present personalized content by cooperating with other agents. In this research, not only software to homogeneous customers is considered, but also group collaboration is enhanced with the system used in a Business-to-Business model. The register agent cooperates with personalized interface agent by dynamically presenting the company information the users belong to, and returns a feedback prompt to users about the results of their application. On the other hand, heterogeneous social/human agents, along with software agents, seamlessly collaborate to achieve the common goal in this study. Therefore, the job broker among human agents is built to keep information from overflowing to uninterested human roles. In addition, a token is also designed to mitigate the spreading of incorrect information generated by human error. Only with the token, invoked by the action of submitting, is execution of the function by a distinct human agent allowed. Common traits of these web-based hierarchical works are multi-user support and predefined information routing. Furthermore, a middle agent is usually found to coordinate the requesters and information providers, while the benefits and disadvantages of each typed middle agent coexist [10,12]. As for the integration approach, heterogeneous software connection or data exchange is usually realized with tools.

For instance, Ardissono et al. [2] used a JAVA-based technique to integrate heterogeneous software

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 9, September 2017

(CORBA) for better interoperability.

In this research, a XML-based Web Services technique is deployed for the specific requirements of the environment.

Other works based on the integration approach have discussed the cooperation among layered federal applications connected with the EAI bus such as the SCM system. These works, however, are beyond the scope of this research. In other words, web/server-based MAS technique is especially efficient in mediating information sharing between senders and receivers even with coexistent problems such as tradeoffs between the advantages and disadvantages of middle agents and lack of global view to the MAS.

III. QUALITY OF SERVICE IN CLOUD COMPUTING

In this section of the thesis the topic of resource management in Cloud Computing is discussed, including a clarification of what elasticity means in the context of a Cloud.

Scheduling of virtual resource and the algorithms used in IaaS implementations, one open source, another that accommodates advanced reservation and a commercial offering are discussed, in addition to monitoring tools used to detect environmental changes that provides input to Cloud resource schedulers.

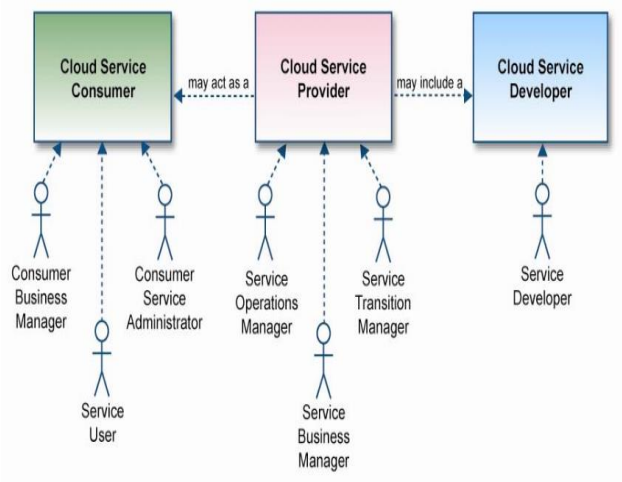

Figure 3: Cloud Actors

Resources management in distributed systems, in the most generalised of contexts, refers to the efficient allocation of workload to a shared computing resource. This is achieved by setting a goal, such as maximising resource utility or workload throughput, given a set of constraints, often technological and economical. Resource management involves the following:

•***Characterisation:*** Knowledge acquisition and understanding of the system workload and its resource requirements.
•***Allocation:*** The distribution of workload to resources across competing tasks or services.
•***Adaptation:*** The accommodation of system and environmental changes such as failures and changes in workload.

These three aspects of resource management are only possible through the monitoring of system resources. In Cloud Computing resource management differs to traditional distributed systems such as Grids and clusters due to Virtualization. VMs are multiplexed between the resources of a physical machine enabling multitenancy and resource sharing.

In addition, Virtualization provides a layer of abstraction above the physical resource, enabling a two tier approach to resource management in a IaaS provider. An application running on Cloud infrastructure is thus unaware of the underlying physical environment with which it is executing. This make the characterisation of workload difficult as the infrastructure layer in a Cloud is unaware of the attributes associated with an application running inside a VM and is an active area of research [19]. The goal of an IaaS provider is to maximise revenue which involves the minimising of power consumption and maximization of resource utility via the consolidation of VMs onto physical machines.

In Cloud Computing monitoring tools are used to characterise the current workload that a VM is placing on a physical machine. When additional VMs are brought online the current state of the Cloud is evaluated and the VMs are allocated to a suitably underutilised resource that meets the QoS requirements. In the event of hardware failure, fault tolerance is achieved via the migration of a VM to a functional resources. The live migration of running VMs also play a role in adapting to changes in demand. Workloads in Clouds do not remain static and change during periods of peak user activity. Thus a VM that is

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 9, September 2017

using many resources could have a potential QoS impact on other VMs running in the same physical machine and should be migrated to a resource with appropriate spare capacity.

IV.PROPOSED INTELLIGENT AGENT BASED FRAMEWORK

As shown in Fig. 4, proposed system consists of two main layers: the customers' applications layer and cloud provider's resources layer.

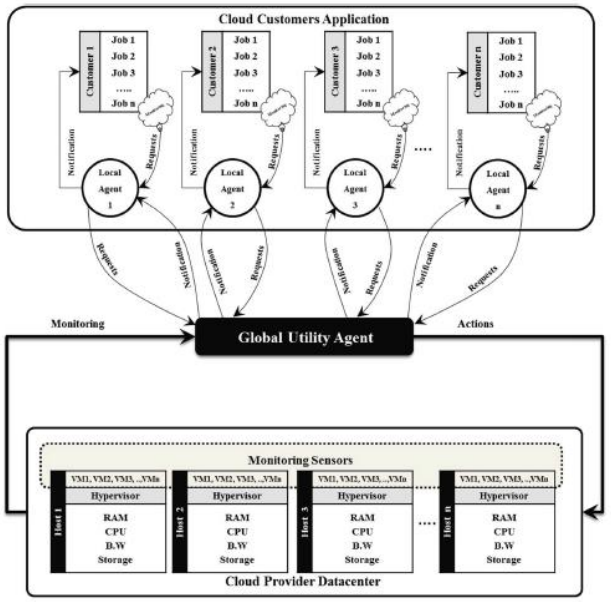

Figure 4 Working of proposed system

The provider's resources include a set of datacenters each with a large number of physical machines (hosts). Each host has specific characteristics (in terms of CPU, RAM, storage and BW) to host multiple VMs. Moreover, it has monitoring sensors to measure the utilization of its resources and the execution of the customers' applications on it. The applications layer consists of a set of customers having several jobs possibly spanning multiple VMs. Each job is associated with specific performance goals specified in the SLA (e.g., time to complete their tasks, number of requests for specific time periods). The multi-agent components include a global utility agent and a set of l ocal utility agents. While the global utility agent has the classical role of the "central broker" allowing it to manage all of the system resources, the local agents are assigned to each customer with the objective of improving the resource utilization without causing SLA violations. One of the advantages of using multiple agents is to allow for certain optimizations that can conducted "locally" (i.e., at the customer level) without burdening the global agent with the low-level details of these optimizations.
Here we discuss one such example where each

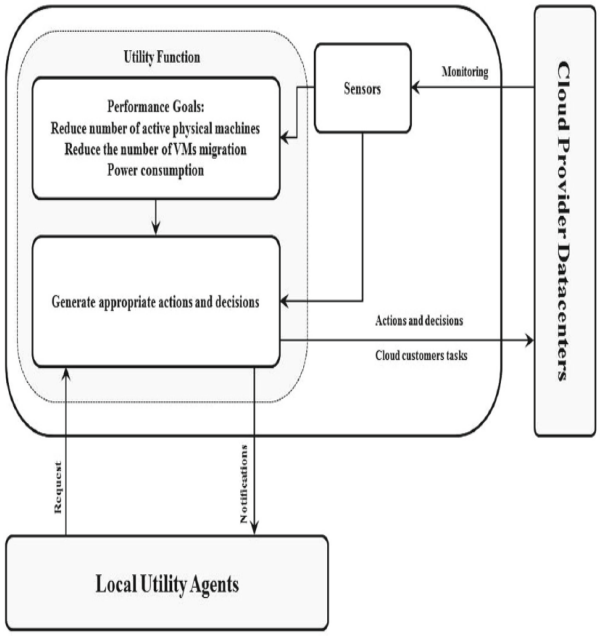

Figure 5 Working of proposed agent system.

V. CONCLUSION

This study presented the Dynamic Resources Provisioning and Monitoring (DRPM) system, a multi-agent system to manage the cloud provider's resources while taking into account the customers' quality of service (QoS) requirements as determined by the service-level agreement (SLA). Moreover, DRPM includes a new Virtual Machine (VM) selection algorithm called the Host Fault Detection (HFD) algorithm. The proposedDRPMsystem is evaluated using the CloudSim tool. The results show that the DRPM system allows the cloud provider to increases resource utilization and decreases power consumption while avoiding SLA violations.

REFERENCE

[1] Armbrust, M., Fox, A., Griffith, R., Joseph, A.D.,

Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., et al.: A view of cloud computing. Commun. ACM 53(4), 50–58 (2010)

[2]Beloglazov, A., Buyya, R.: Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in cloud data centers. Concurr. Comput. 24(13), 1397–1420 (2012)

[3]Bonvin,N., Papaioannou,T.G.,Aberer, K.:Autonomic SLA-driven provisioning for cloud applications. In: Proceedings of the 2011 11th IEEE/ACM international symposium on cluster, cloud and grid computing, IEEE Computer Society, pp. 434–443 (2011)

[4]Calheiros, R.N.,Ranjan, R.,Beloglazov, A.,DeRose, C.A.,Buyya, R.: Cloudsim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms. Software 41(1), 23–50 (2011)

[5]Chatterjee, S., Hadi, A.S.: Regression analysis by example.Wiley, Hoboken (2013)

[6]Duong, T.N.B., Li, X., Goh, R.S.M.: A framework for dynamic resource provisioning and adaptation in iaas clouds. In: Proceedings of the IEEE Third International Conference on Cloud Computing Technology and Science (CloudCom), 2011, pp. 312–319 (2011). IEEE

[7]Hategan, M., Wozniak, J., Maheshwari, K.: Coasters: uniform resource provisioning and access for clouds and grids. In: Proceedings of the Fourth IEEE International Conference on Utility and Cloud Computing (UCC), pp. 114–121 (2011). IEEE

[8]Herbst, N.R., Kounev, S., Reussner, R.: Elasticity in cloud computing: what it is, and what it is not. In: Proceedings of the 10th International Conference on autonomic computing (ICAC 2013), San Jose, CA (2013)

[9]Huang, H., Wang, L.: P&p: a combined push-pull model for resource monitoring in cloud computing environment. In: Proceedings of the Cloud Computing (CLOUD), 2010 IEEE 3rd International Conference on, pp. 260–267 (2010). IEEE

[10]Jararweh,Y., Jarrah, M.,Kharbutli, M.,Alsaleh,M.N., Al-Ayyoub, M.: CloudExp: a comprehensive cloud computing experimental framework. Simul. Model. Pract. Theory 49, 180–192 (2014)

[11]Marshall, P., Keahey, K., Freeman, T.: Elastic site: using clouds to elastically extend site resources. In: Proceedings of the 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, EEE Computer Society, pp. 43–52. I (2010)

[12]Siddiqui, U., Tahir, G.A., Rehman, A.U., Ali, Z., Rasool, R.U., Bloodsworth, P.: Elastic jade: dynamically scalable multi agents using cloud resources. In: Proceedings of the Cloud and Green Computing (CGC), 2012 Second International Conference on, pp. 167–172 (2012). IEEE

[13] 13. Vaquero, L.M., Rodero-Merino, L., Buyya, R.: Dynamically scaling applications in the cloud.ACMSIGCOMMComput.Commun. Rev. 41(1), 45–52 (2011).