Semantic based Resource Provisioning and Scheduling in Inter-cloud Environment

V. Nelson
Department of Computer Science
Pondicherry University
Pondicherry, India
mrnelsone@gmail.com

V. Uma
Department of Computer Science
Pondicherry University
Pondicherry, India
umabskr@gmail.com

Abstract—**The resources held by a single cloud are usually limited and at peak period, the organization may not be able to give the guaranteed services due to insufficient provisioning of resources. So it is essential to organize cloud systems that complement each other such as to procure resources from other participating cloud systems. However, it is difficult to provide the right resources from various cloud providers because management policies and descriptions about various resources are different in each organization. Having these differences, it is hard to provide interoperability among them. Representing cloud environment through ontology can conceptualize common attribute among the various resources semantically. Considering this fact, we propose an Inter-cloud Resource Provisioning System (IRPS) in which the resources and tasks are described semantically and stored using resource ontology and the resources are assignedusing a set of inference rules and a semantic scheduler.**

Keywords— Intercloud, semantics, inference, ontology, resource sharing, sceduling, meta – data repository.

I. INTRODUCTION

Inter-cloud systems are unavoidable as it is very difficult for a single cloud provider to satisfy their customer requirements to the maximum. Cloud has limits and so it is indispensable to organize that cloud systems complement each other such as to procure resource from other cloud systems (Inter-cloud) connected via fast internet. However, it is difficult to provide the right resources from various providers because the management policies and descriptions about various resources are different in each organization. These differences hinder the uniform view of various resources. Ontology based resource description is used to solve these problems. All the cloud providers participating in the Inter-cloud systems describe their resources semantically and store it in their respective meta-data repository. Our proposal is to deal with the cloud providers to provide automated resource allocation and scheduling. Additionally, the cloud service providers should try to maximize their profit by optimizingthe utilization of resources, maximize the customer satisfaction by meeting all the requirements and in case of conflict penalties should be granted to the customers. Under these conditions, the Inter-cloud resource provisioning system has been proposed andit is expected to be availed from technologies such as inferencing, semantics and virtualization. This model provides a framework which schedules the customer requests considering Service Level Agreement (SLA), state of Service Provider (SP) resources and SP level preferences for cloud customers. Also this model can re-schedule the customers' requests based on their priority and reservations. The overallsystem is described semantically to assign resources based on a set of inference rules. The usage of virtualization technology enables dynamic resource distribution whichhelps in adapting to changing environment.

The paper is organized as follows. Section II reviews the related works. Section III deals with the proposed system and the semantic scheduler which is used to allocate timeslots for the given tasks. Section IV tells about the overall process flow. Finally section V is the conclusion of this paper.

II. RELATED WORK

In Inter-cloud each organization describes their resources and application requirement using their own language which actually makes the interoperation very costly. Since Inter-cloud is the most beneficial for the market oriented computing, interoperability problem should be addressed. A framework is proposed to facilitate themanagement of intercloud components and to provide the guaranteed end to end services to meet the SLA requirements [1]. The representation of Cloud computing environment using ontology can capture thesame attributes among the resources and to describe relations among them semantically [2].The emerging problem of semantic interoperability between heterogeneous cooperating Cloudstowards a platform has been focused in RASIC. In this work, they tried to pave the way towards a Reference Architecture for Semantically Interoperable Clouds (RASIC) [3].

It is mentionable that the Intercloud Resource sharing has been tried out in a cloud environment without using semantic concepts [4]. Another framework called SERAhas been proposed [5] andthis workbrings up a service provider management framework, which reduces costs and at the same time fulfills the guaranteed service. The basics of Intercloud interoperability problem have been discussed in the paper [6]. This workalso introducesarchitecture of Intercloud Standards. The next work supports an ontology based service discovery in cloud environment to provide deployment of appliances based on QOS [7]. Ontology based resource matchmaking algorithm [8-9] has been presented by a set of rules which identify the perfect resource to fulfill the user requirements. Previously Distributed resource allocation process is carried out using semantics multi-agent technology which makes the integration of different service providers in the cloud environment easier [10]. In another work scheduling ontology in general has been discussed [11]. RDF query language has been used to retrieve information from the ontology [12].

III. PROPOSED WORK : INTER-CLOUD RESOURCE PROVISIONING SYSTEM (IRPS)

This section gives an overview of the architecture of the IRPS. The architecture comprises two clouds of different service providers in a federated cloud environment. The cloud service consumer selects the appropriate cloud for their task with the help of cloud brokers. Cloud brokers acts like a mediator. The Cloud Co-coordinator (CC) publishes the available supply of resources currently with the respective cloud. Also it updates the resource availability after every change in the resource pool. The Inter-cloud directory brings the service providers and consumers together. It helps the cloud customers to locate the published services by the cloud vendors.

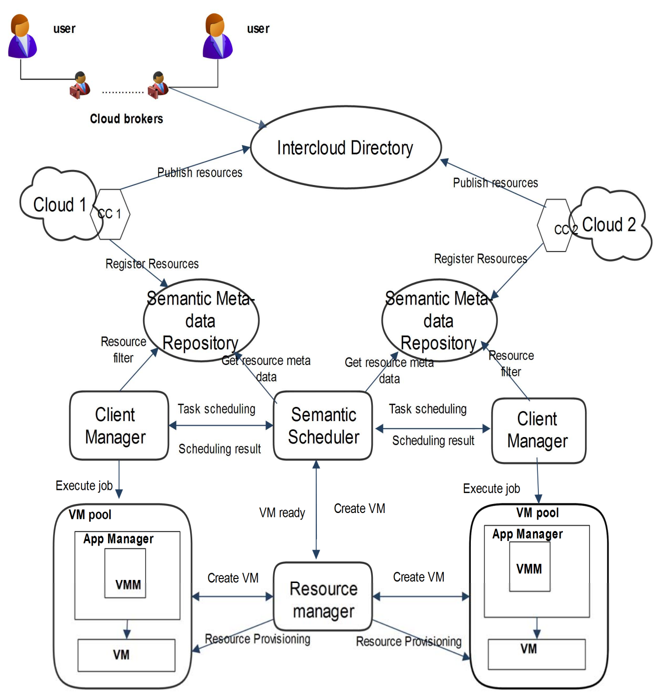

Fig.1. Inter-cloud Resource Provisioning System

We have the common Semantic Scheduler (SS) and Resource Manager for the federated cloud environment. The SS allocates the required resources to each task according to its priority, requirements and the system status in orderto favor customers with more priority. The Client Manager (CM) manages the execution of client's task and takes decisionswhen some unanticipated events, such as SLA violations occur.

The Resource Manager (RM) produces virtual machines (VM) to perform clients 'tasks based on the minimum resource allocationsuggested by the Semantic Scheduler. Application Manager (AM) supervises the resource usage in order to encounter the SLA violations. All the components are semantically described and stored in the Semantic Metadata Repository (SMR). Metadata is data that describe the information about resources. Resource discovery is made easy with semantic description of resources.

Fig.1 depicts the task life cycle of the Inter-cloud Resource Provisioning System (IRPS) and how the different components interact among each other. There are two clouds represented in this federated environment (cloud1 and cloud2). Each cloud has its own Semantic Meta data Repository (SMR), where the semantic descriptions of each component are stored. When a task arrives to the system, a CM is established to manage its execution. The CM queries the SMR and selects the potential nodes with minimum requirements for running the task and registers the task description in SMR. Then, it requests a time slot to the Semantic Scheduler. When the task gets the time slot, the Scheduler contacts the Resource manager and requests the VM creation to execute the task. TheRM creates VM after receiving the Semantic Scheduler request and an AM that monitors the SLA fulfillment for this task and informs the Semantic Scheduler when they are ready. The Semantic Scheduler sends the message to the CM with the access code to the VM. After getting the access code, the task is submitted to the assigned VM by the CM. When the task starts running, it is monitored by the corresponding AM to encounter SLA breaches. If the cloud is not able to meet the resource requirement, the request for additional resource will be sent to the other cloud through RM and get it sanctioned.

A. SMR and Semantic Descriptions:

SMR makes sure that cloud users and vendors are relying on the common service meaning, regardless of the differences, the organizations make. So this reduces most of the communication gap that exist among various cloud providers.

All the components in IRPS are semantically described and registered into the Semantic Metadata Repository (SMR). This provides the common view of all the resources in the Inter-cloud environment. In this context, metadata stores information about computational resources. For E.g. the date when a resource was in use, how long a specific resource is allowed to access, location of the resource, storage capacity, memory availability, OS type etc. An RDF model is created using the data obtained from the Inter-cloud resource ontology. The resource properties and RM descriptions are added to the RDF model and are registered to the SMR. The task requirements and the client descriptions are added to another RDF model and are registered into the SMR. SPARQL [12] is created using the task requirements given by the user and is used to choose all the potential hosts from the SMR that fulfill the resource requirements. The query results are again registered in the SMR in order to update the repository.

B. *Semantic Scheduler*

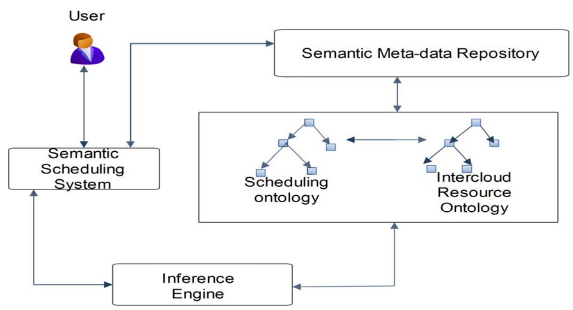

Fig. 2 Semantic Scheduler

Fig. 2 shows the process of scheduling for the requested task using inference engine. When CM requests the time slot for the task to the semantic scheduling system, the semantic scheduling system retrieves the task description from the SMR. The inference engine performs the reasoning act and extracts the task schedule, mapping the Scheduling ontology [11] and the Inter-cloud Resource Ontology. The task schedule will be sent to the Client Manager and gets stored in the Semantic Meta-Data Repository.

C. *Inference*

In this paper we extendedthe inference engine algorithm [5] which reasons using ontology form of knowledge representation. The inference engine checks the given conditions and examines the given rules. If it goes true for the given tasks, appropriate actions will be triggered and executed. These rules implement the scheduling policies.

Inference Engine Algorithm for Semantic Scheduler:

1: Define t as Task; c as Condition; r asRules and v as Values.

2: Assign (t_i , v)

3:**Pre-condition:**SMR is correct.

4: **Post-condition:**All scheduled tasks have been executed and all the meta-data in therepository have been updated.

5: t_i=v; v={0,1,2,3..}

6: c_j={FCFS algorithm, FCFS failure, Task Re-Allocation failure}

7: r_k={FCFS, Task Re-allocation, Task cancellation}

8: if resource availability is true,

9: if time slot in a machine for a task is using FCFS algorithm, fire FCFS (R1).

else

10: if FCFS cannot find a timeslot in a machine, fire Task re-allocation (R2).

11: if Task re-allocation fails to find a contact CM of cloud 2 and do the previous steps again.

12: else solution, fire Task cancellation (R3).

13: Register the results and set SMR updated.

We have proposed an inference engine algorithm for scheduling the given tasks considering the resource availability. This algorithm provides the optimal solution by maximizing the efficiency of resource usage. Finally the task scheduling is done using inference engine algorithm. The execution time of the tasks will be monitored by the Semantic Scheduler. When the time slot is reached, it will contact the RM and CM to execute the task.

IV. PROCESS FLOW

A. *Outer Flow*

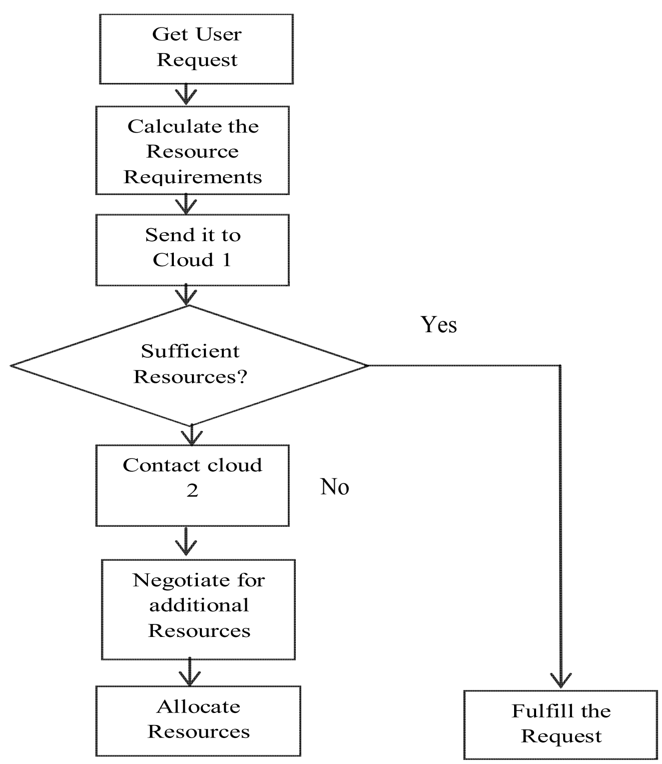

Fig. 3. Outer flow of IRPS

Process flow helps in understanding the flow throughout the system. Fig 3 shows the general view of Inter-cloud Resource Provisioning System according to the user requirements. The overall aim of the system is to allocate resources to the user tasks and complement additional resources when the task runs short of resources. When the first cloud happens to face the peak period and unable to accommodate the user request, the request will be redirected to check the availability of resources in the second cloud. Additional resources will be negotiated and allocated to satisfy the SLA requirements.

B. Inner Flow

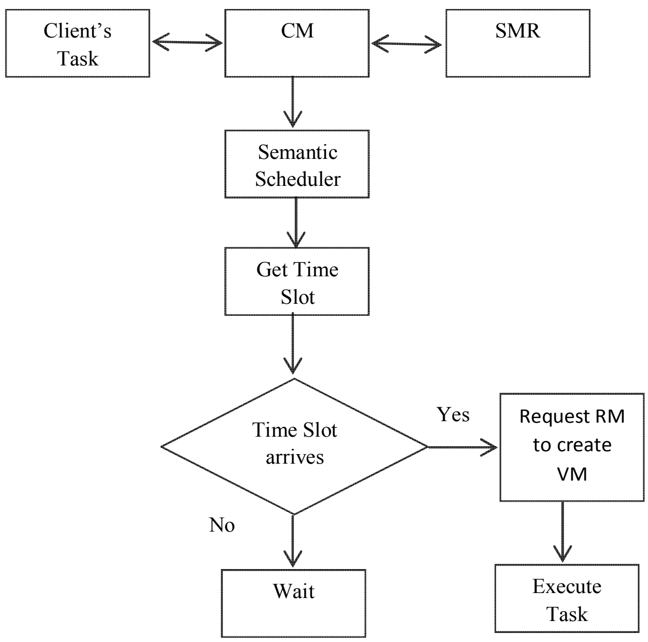

Fig.4. Inner flow of IRPS

Fig. 4 shows explain the task execution process inside the system. When the client's tasks are submitted to the cloud system, a CM is created and it queries the SMR to select the potential nodes. The CM requests time slots for the given tasks and stores. When the time slot arrives, RM is requested to create VM to execute the task. Otherwise this process is waited until the corresponding time slot arrives.

V. EVALUATION STUDY

The proposed work can be implemented using different existing semantic framework. The semantic part of IRPS can be implemented and tested in Intercloud environment. SESAME [13] is a semantic framework consists of a repository, a query engine and an administration module for adding and deleting RDF data. It can support excessive querying of RDF data using RQL query language and understands the semantics of most of the RDF schema classes and properties. Thus it supports basic inferencing.

The existing system (SERA) [5] uses Jena framework. Usage of SESAME in our proposed system instead of Jena helps in achieving faster response time. Having less number of triples (RDF), SESAME is expected to perform much faster than Jena [14].

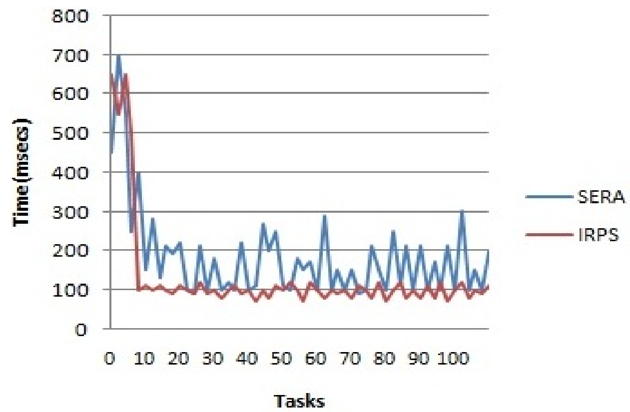

Fig.5. SPARQL Query Time

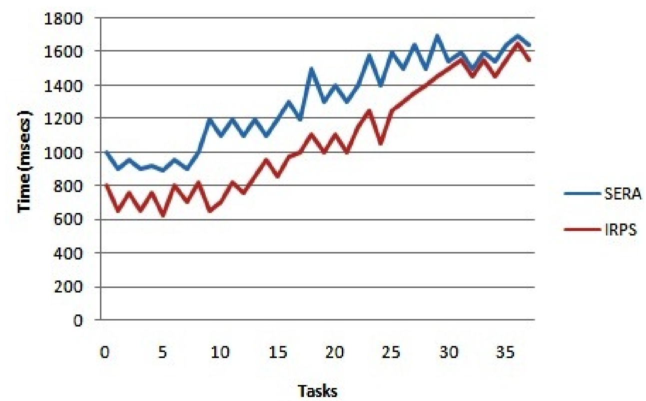

Fig.6. Inference Time

Fig.5 and Fig. 6 shows the comparison graph between the SERA and IRPS. Using SESMAE in IRPS, it is expected to have faster performance as the graphs show.

VI. CONCLUSION

This paper introduces Inter-cloud Resource Provisioning System (IRPS) which enables the fulfillment of customer requirements to the maximum by providing additional resources to the cloud system participating in a federated environment. The tasks are scheduled to allocate resources by the semantic scheduler and an inference engine. Inference processing is done using an inference engine algorithm. When a cloud system runs short of resources, it is complemented by the other cloud system in the federated environment so as to fulfill the customer requirements. Semantics in the federated environment is a powerful idea. So this idea will solve the interoperability problem which the federated cloud environment suffers.

REFERENCES

[1] R. Buyya, R. Ranjan, and R. N. Calheiros, "InterCloud: Utility-Oriented Federation of Cloud Computing Environments for Scaling of Application Services," in Proceedings of the 10th International Conference on Algorithms and Architectures for Parallel Processing (ICA3PP 2010), Busan, South Korea, 2010.

[2] HyunjeongYoo, CinyoungHur, Seoyoung Kim, and YoonheeKim , “An ontology-based resource selection service on science cloud.”International Journal of Grid and Distributed Computing Vol. 2, No. 4, December, 2009.

[3] NikolaosLoutas,VassiliosPeristeras,ThanassisBouras,EleniKamateriDimitriosZeginis,“Towards a reference architectureor semantically interoperable clouds.”IEEE Second International Conference on Cloud Computing Technology and Science(CloudCom), 2010.

[4] Michael D. Brookbanks, Brendan F. Coffey, Christopher J. Dawson, Thirumal Nellutla, Robert C. Patterson, JR., James W. Seaman, International Business Machine Corporation, “Intercloud resource sharing within a cloud computing environment”, 2011.

[5] JorgeEjarque, Marc de Palol, InigoGoiri, Ferran Julia, JordiGuitart, Jordi Torres and Rosa M. Badia, “Using semantics for resource allocation in computing service providers,” Barcelona Supercomputing Center and UniversitatPolitecnica de Catalunya 2009.

[6] David Bernstein Erik Ludvigson Krishna Sankar Steve “Blueprint for the inter-cloud – protocols and formats for internet and web applications and services.” Diamond Monique Morrow Cisco Systems, Inc. , ICIW '09.

[7] A. V. Dastjerdi, S. G. H. Tabatabaei, and R. Buyya, "An Effective Architecture for Automated Appliance Management System Applying Ontology-Based Cloud Discovery," in Proceedings of the 10th IEEE/ACM International Symposium on Cluster, Cloud and GridComputing (CCGrid 2010),Melbourne, Australia, 2010.

[8] H. Tangmunarunkit, S. Decker and C. Kesselman. “Ontology-based resource matching in the grid - the gridmeets the semantic web.” Springer International Semantic Web Conference, 2003.

[9] Neela Narayanan V and Kailash S. “Resource matchmakingin grid - semantically.” In The 9th International Conference on Advanced Communication Technology, 2007.

[10] JorgeEjarque, RaulSirvent and Rosa M. Badia “A Multi-agent approach for semantic resource allocation.” IEEE CloudCom 2010.

[11] Dnyanesh Rajpathak1, Enrico Motta1 and Rajkumar Roy, “A generic task ontology for scheduling applications.” Cranfield University, UK 2007.

[12] Marcelo Arenas, Claudio Gutierrez, Jorge Perez, “ On the Semantics of SPARQL.”

[13] Aimilia Magkanaraki, Grigoris Karvounarakis, Ta Tuan Anh, Vassilis Christophides,Dimitris Plexousakis, “Ontology Storage and Querying.” Technical Report no:308, Apr 2002.

[14] Christian Bizer, Andreas Schultz “The Berlin SPARQL Benchmark.” Web-based Systems Group,Freie Universität Berlin.2009.