ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 2, No 5, May 2013

Efficient and Reliable Resource Management Framework for Public Cloud Computing

M.Sasitharagai[1], T.Rajendran[2] and M.Malarmathi[3]
[1]PG Scholar
[2]Professor & Dean
Department of Computer Science and Engineering,
Angel College of Engineering and Technology, Tirupur,
[3]Assistant Professor, Department of Information Technology, SNS College of Engineering, Coimbatore

Abstract **- The problem of dynamic resource management for a large-scale cloud environment is mitigated with optimized high throughput performance. The resource management framework consists of, Gossip protocol that ensures fair resource allocation among sites by calculating Memory Load Factor and CPU Load Factor and routing table for dynamically managing the tasks. A request partitioning approach based on gossip protocol is proposed that facilitates the cost-efficient and online splitting of user requests among eligible Cloud Service Providers within a networked cloud environment. Following the outcome of the request partitioning phase, the embedding phase - where the actual mapping of requested virtual to physical resources is performed that allows for efficient and balanced allocation of cloud resources. Finally, a thorough evaluation of the overall framework on a simulated cloud environment is made, which offers reliable and dynamic resource management.**

Keywords **- Cloud Computing, Resource Allocation, Gossip Protocol, Routing table, Resource mapping**

I. INTRODUCTION

Cloud computing offers easy accessible computing resources of variable size and capabilities. This standard allows applications to rent computing resources and services on-demand, benefiting from dynamic allocation and the economy of scale of large data centers [1]. We focus the problem of resource management for a large-scale cloud environment. Such an environment includes the physical infrastructure and associated control functionality that enables the provisioning and management of cloud services. In recent years, Cloud Computing has become a hotspot for business institutions and research institutions. It concerns mainly about how the computing resources are virtualized [2], and how to efficiently schedule user's tasks and make a reasonable distribution of system resources for realizing the resource load balance is also the key factors of raising the cloud computing platform's performance and service quality.

A. RESOURCE PROVISIONING

Generally, the services are provided to the clients by the Cloud Service Provider (CSP). Such services are namely, Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) and all these resources can be scheduled to the clients in cloud environment in a balanced and a cost-efficient manner with the application of scheduling algorithms, which balances the load condition on inputs.

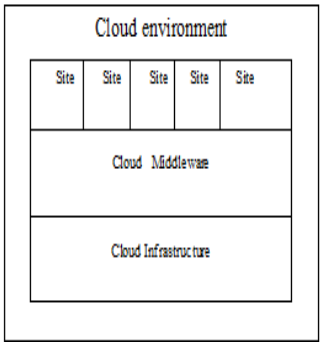

Fig.1. Overall architecture of the cloud environment

The CSP owns and administers the physical infrastructure [3], on which cloud services are provided. It offers hosting services to site owners

through a middleware that executes on its infrastructure. This work contributes towards engineering a middleware layer that performs resource allocation in a cloud environment, with the following design goals:

1)Performance objective: We consider computational and memory resources and the objective is to achieve reasonable fairness among sites for computational resources under memory constraints.

2) Adaptability: The resource allocation process must dynamically and efficiently adapt to changes in the demand from sites.

3) Scalability: The resource allocation process must be scalable both in the number of machines in the cloud and the number of sites that the cloud hosts.

Along with the above mentioned design goals, the middleware layer should be capable enough to deal with the dynamic inputs and also if the user is not able to access the needed resource, the resource management framework should assign the task to a virtualized server, where the resource allocation can be fairly achieved, in turn makes the cloud environment a balanced one. The paper is organized as follows. Section 2 discusses related work; Section 3 introduces the proposed system, where it explains about dynamically managing the services in large cloud environment and Section 4, demonstrates the performance analysis and finally Section 5 concludes the paper.

II. RELATED WORK

A. CLOUD ENVIRONMENT

Cloud computing uses parallel computing to solve big problems and the calculated resources can be measured in services to users of the utility computing platform. This can be achieved through virtualization technology [4]. The issue arising from Web Service applications is concerned and an optimal approach scheme is implemented for discovering the most suitable web service according to the consumer's functional and quality requirements [5]. The new emergence of cloud computing technologies provides method to deal with complex applications which needs high performance. This can be made possible through inter-connected and virtualized computers which dynamically provision as per user's requirements [6].

B.SCHEDULING OF RESOURCES

Task scheduling is very important to scientific workflows and task scheduling is a challenging problem too. Considering the balanced resource allocation on dynamic inputs, the existing techniques available are by Fetahi Wuhib, Rolf Stadler and Mike Spreitzer [3] that uses configuration matrix that controls the module scheduler and request forwarder components. For this work, computational resources (i.e., CPU) and memory resources of cloud are considered, which are available on the machines in the cloud infrastructure. Each machine runs a machine manager component given in [7] that computes the resource allocation policy. The resource allocation policy is computed by a protocol P* that runs in the resource manager component. The computed allocation policy is sent to the module scheduler for making decisions. In this paper, a novel approach is adopted for designing and developing a broker-based architecture and its QoS-based selection algorithm for evaluating services in cloud environment [8].

C.MAPPING OF RESOURCES

Recently network virtualization [9] has been proposed as a promising way to overcome the current ossification of the Internet by allowing multiple heterogeneous virtual networks (VNs) to coexist on a shared infrastructure. A major challenge in this respect is the VN embedding problem that deals with efficient mapping of virtual nodes and virtual links onto the substrate network resources. This literature enables us with an idea of establishing the link between the task and resources in terms of routing table.

Virtualizing and sharing networked resources is a computing and networking architectures. Embedding multiple virtual networks on a shared substrate is a challenging problem on cloud computing platforms and large-scale sliceable network testbeds [10]. This literature aids us to analysis the virtual component in the server in case of resource failure. A major challenge in building the diversified Internet is to

perform efficient and on-demand VN assignment. Resource matching, splitting, embedding and binding steps required for virtual network provisioning are proposed and evaluated. Splitting of the virtual network provisioning request across multiple infrastructure providers is solved using both max-flow min-cut algorithms and linear programming techniques. Virtual network embedding is formulated and solved as a mixed integer program with the aim of decreasing embedding cost for infrastructure providers while increasing the acceptance ratio of requests. Performance of the splitting and embedding algorithms is reported.

III. PROPOSED SYSTEM

We propose a dynamic resource management for a large-scale cloud environment with scalable and optimized high throughput performance. This can be done through a distributed middleware framework called Resource Management Framework with prominent elements. 1. Gossip protocol that ensures fair resource allocation among applications. 2. Dynamically managing the service resources for tasks with different service domains by adapts to the allocation to load changes using routing table

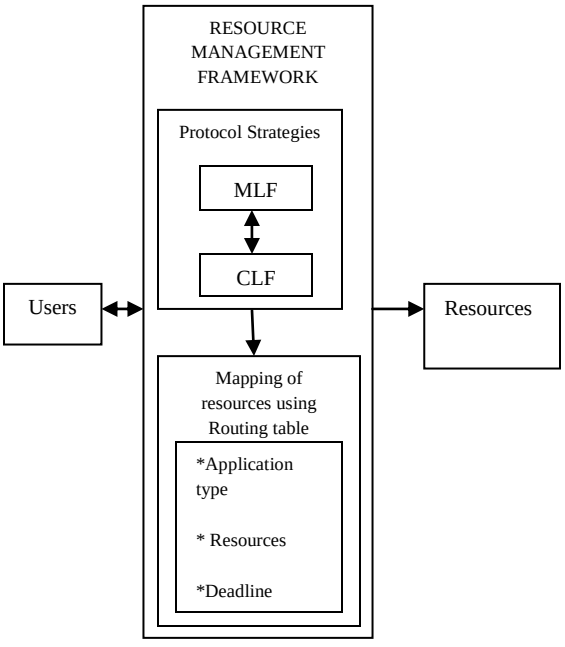

Fig.2. Architectural diagram

A.DISTRIBUTED GOSSIP PROTOCOL FOR SCHEDULING OF RESOURCES

Formalizing the resource allocation problem in dynamic large cloud environment is a tedious process but making this process to be efficient, we need a robust gossip protocol to perform dynamic resource management by satisfying performance objectives.

In this paper, we introduce a robust gossip protocol (P*) [11], for calculating aggregates in a proactive manner. We assume that each node maintains a local approximate of the aggregate value. The core of the protocol is a simple gossip-based communication scheme in which each node periodically selects some other random node to communicate with. During this communication the nodes update their local approximate values by performing some aggregation specific and strictly local computation based on their previous approximate values. This local pair wise interaction is designed in such a way that all approximate values in the system will quickly converge to the desired aggregate value.

For this work, we consider a cloud as having computational resources (i.e., CPU) and memory resources [3], which are available on the machines in the cloud infrastructure. We formulate the resource allocation problem as that of maximizing the cloud utility under CPU and memory constraints. The solution to this problem is a routing table that controls the scheduler and request forwarder components. This can be done by applying the gossip protocol for dynamically managing the service resources for tasks with different service domains by adapting to the allocation to load changes using routing table and virtualized server. This states that the cost of change from the current configuration to the new configuration must be minimized, when the appropriate resource is not mapped within the specified deadline.

Based on our work thus far, we believe that, for a gossip protocol running in large-scale dynamic environments, the advantages of continuous execution with dynamic input outweigh its potential drawback of analyzing the behavior of the protocol [12]. Compared to sequential execution, continuous execution does not need global synchronization,

whereas, sequential execution assume static input and produce a single output value. Whenever the input changes, protocol is restarted and produce a new output value, which requires global synchronization.

Application placement in datacenters is often modeled through mapping a set of applications onto a set of machines such that some utility function is maximized under resource constraints. This approach has been taken in [13]. An alternative way of computing a feasible configuration is given in [14], which mainly focus on the application placement in datacenters. The information resources in web systems are generally distributed, dynamic and varied. As these computing atmospheres are open, information resources can be connected or disconnected at any time [15].

The Process Flow can be organized as:

1. Initial scheduling of the task through gossip protocol specifications like Memory and CPU constraints
2. Scheduled output from gossip system is passed into request partitioning method
3. Request partitioning Method
4. Mapping the Task and resources

B.ROUTING TABLE MANAGEMENT

We implement a novel request partitioning approach based on gossip protocol that facilitates the cost-efficient and online splitting of user requests among eligible Cloud Service Providers (CSPs) within a networked cloud environment through establishing the clusters for resources based on the configuration. Routing table is formed to map the dynamic task to the efficient resources in terms of task properties and previous determined factors from the gossip protocol. After incorporating all the aspects of scheduling properties in routing table, then we do the scheduling as per reference in it. Following the outcome of the request partitioning phase, the embedding phase - where the actual mapping of requested virtual to physical resources is performed can be realized through the use of a distributed intra-cloud resource mapping approach that allows for efficient and balanced allocation of cloud resources. Resource Management Algorithm is applied here for mapping the virtual machine to physical resources.

In our work, we find that each of the users demanding resources is fairly satisfied by the constraints with the help of a routing table, where the resources are grouped as clusters, which provides efficient and a dynamic mapping approach to the appropriate resource.

IV. PERFORMANCE ANALYSIS

To evaluate the performance of our resource management framework, cloud computing simulation software CloudSim is used. CloudSim is a framework for simulation and modeling of cloud computing environments mainly used for resource allocation and scheduling. It includes data centers, Virtual machines, brokers, cloudlets and hosts.

A.SIMULATION OF DISTRIBUTED GOSSIP PROTOCOL

Upon simulating the gossip protocol, we have simulated various users demanding resources and their respective MLF is calculated using CloudSim2.1. The respective Memory Load Factor (MLF) of each of the user is calculated in accordance with 2 aspects:

1. Memory utilization
2. Execution time

Correspondingly, CPU Load Factor (CLF) can also be shown in accordance with MLF, whereas MLF and CLF are the inputs to routing table, where resource type information, deadline constraints are the parameters stored in additional in routing table for performing efficient and dynamic mapping.

Also, we have simulated various users demanding resources with the help of a routing table management. Generally, the routing table is used to store the information of same type as clusters, so that, mapping strategy can be performed faster. Typical routing table entries used for simulation is given as follows,

TABLE I
ROUTING TABLE

USER	MLF	CLF	DEADLINE	TIME
1	20	43	7	10
2	16	34	8	9
3	20	37	9	12

In accordance with routing table values, the resource management framework performs the effective mapping strategy as shown below,

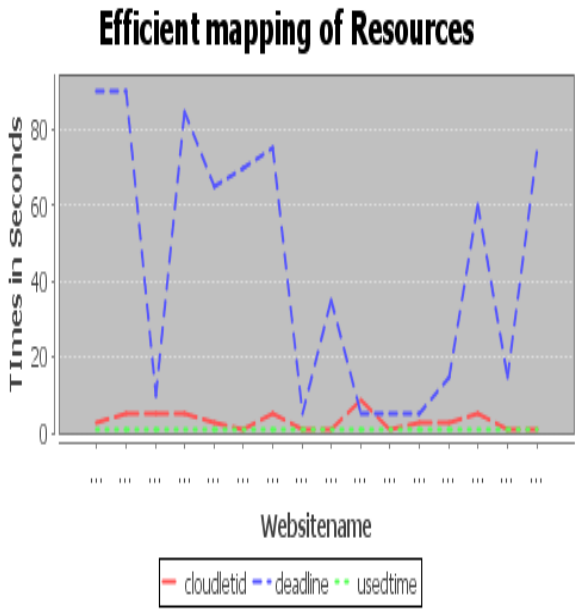

Fig.3. Execution time of each user

The above graph clearly illustrates that, (Fig.3) yields better result by reducing the execution time of each user in getting serviced. This is achieved with the help of routing table, which prioritize the user needs and maps with the stored cluster information.

V. CONCLUSION

With this paper, we make a significant contribution towards engineering a resource management middleware for a site-hosting cloud environment. We identify a key component of such a middleware and presented P* protocol that can be used to meet our design goals for resource management: fairness of resource allocation with respect to sites, efficient adaptation to load changes and scalability of the middleware layer in terms of both the number of machines in the cloud as well as the number of hosted sites. Also, the resource management protocol dynamically manages the service resources for tasks with different service domains by adapting to the allocation to load changes using routing table, where the efficient mapping of resources is achieved with improved load balancing and scalable performances in terms of execution time.

REFERENCES

[1] Navendu Jain, Ishai Menache, Joseph (Seffi) Naor and Jonathan Yani, "A Truthful Mechanism for Value-Based Scheduling in Cloud Computing"

[2] Linan Zhu, Qingshui Li and Lingna He, "Study on Cloud Computing Resource Scheduling Strategy Based on the Ant Colony Optimization Algorithm," in IJCSI International Journal of Computer Science Issues, vol. 9, issue 5, no. 2, September 2012.

[3] Fetahi Wuhib, Rolf Stadler and Mike Spreitzer, "A Gossip Protocol for Dynamic Resource Management in Large Cloud Environments," IEEE Trans. Network and Service Management, vol. 9, no. 2, pp. 213-225, June 2012.

[4] Application Architecture for Cloud Computing, IBM WHITE PAPER

[5] T.Rajendran and P.Balasubramanie, "An Efficient Architecture for Agent-Based Dynamic Web Service Discovery with QOS, Journal of Theoretical and Applied Information Technology," Vol 15. No. 2, pp 86-95, May2010.

[6] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I.Brandic, "Cloud computing and emerging it platforms: Vision,hype, and reality for delivering computing as the 5th utility," Future Generation Computer Systems, 25(6):599–616, 2009.

[7] F. Wuhib, R. Stadler, and M. Spreitzer, "Gossip-based resource management for cloud environments," in 2010 International Conference on Network and Service Management.

[8] T.Rajendran and P.Balasubramanie, "An Optimal Broker-Based Architecture for Web

Service Discovery with QoS Characteristics, International Journal of Web Services Practices, "Vol. 5, No.1, pp. 32-40, July 2010.

[9] M., Chowdhury, M. R. Rahman, R. Boutaba, "ViNEYard: Virtual Network Embedding Algorithms with Coordinated Node and Link Mapping, Networking" IEEE/ACM Transactions on vol.20, no.1, pp.206-219, Feb. 2012,doi:10.1109/TNET.2011.2159308.

[10] Xiang Cheng, Sen Su, Zhongbao Zhang, Hanchi Wang, Fangchun Yang, Yan Luo, and Jie Wang, "Virtual network embedding through topology-aware node ranking," *SIGCOMM* Comput. Commun. *Rev.*, vol. 41, no. 2, pp. 38-47, April 2011, doi:http://doi.acm.org/10.1145/1971162.1971168.

[11] M. Jelasity, A. Montresor, and O. Babaoglu, "Gossip-based aggregation in large dynamic networks," ACM Trans. Computer Syst., vol. 23, no. 3, pp. 219–252, 2005.

[12] F. Wuhib, M. Dam, and R. Stadler, "A gossiping protocol for detecting global threshold crossings," IEEE Trans. Network and Service Management, vol. 7, no. 1, pp. 42–57, Mar. 2010.

[13] D. Carrera, M. Steinder, I. Whalley, J. Torres, and E. Ayguade, "Utility-based placement of dynamic web applications with fairness goals," in Network Operations and Management Symposium, 2008. NOMS 2008. IEEE, April 2008, pp. 9 –16.

[14] C. Tang, M. Steinder, M. Spreitzer, and G. Pacifici, "A scalable application placement controller for enterprise data centers," in 2007 International Conference on World Wide Web.

[15] T.Rajendran and P.Balasubramanie, "An Efficient Multi-Agent-Based Architecture for Web Service Registration and Discovery with QoS", European Journal of Scientific Research, Vol.60 No.3, pp.439-450,2011.

AUTHOR'S PROFILE

M.Sasitharagai is a final year student doing Master of Engineering in Computer Science at Angel College of Engineering and Technology, Tirupur and received her Bachelor of Engineering degree in Computer Science from Maharaja Engineering College, Avinashi in the year 2011. Her area of interest is Cloud Computing. She has presented over 7 papers in National and International Conferences. She has also published 5 papers in International Journals.

Dr.T.Rajendran completed his PhD degree in 2012 at Anna University, Chennai in the Department of Information and Communication Engineering. Now he is working as a Dean for Department of CSE & IT at Angel College of Engineering and Technology, Tirupur, Tamilnadu, India. His research interest includes Distributed Systems, Web Services, Network Security and Web Technology. He is a life member of ISTE & CSI. He has published more than 51 articles in International/ National Journals/Conferences. He has visited Dhurakij Pundit University in Thailand for presenting his research paper in an International conference. He was honored with Best Professor Award 2012 by ASDF Global Awards 2012, Pondicherry.

M.Malarmathi received her Bachelor of Engineering degree in Information Technology stream from Anna University Chennai and Master of Technology in Information Technology from Anna University Coimbatore. Now she is working as an Assistant professor in the Department of Information Technology at SNS College of Engineering, Coimbatore. Her area of interest is Peer to Peer networks. She has presented many papers in National conferences and she is a member of Computer Society of India.