IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 11, November 2015

Service Populating and QoS Aware Mechanism for Cloud Based Environment

Smita Patil[1], Arpit Solanki[2]

PG Scholar, CSE, RKDF School of Engineering, Indore, India [1]

Professor, CSE, RKDF School of Engineering, Indore, India[2]

Abstract: In today's world Cloud Computing and Mobile Computing are most popular trends. Cloud Computing provides a huge processing and computing capability to its users and Mobile Computing provides all time connectivity, mobility and software functionality to mobile devices like smartphones. In future these mobile devices are expected to move around the world and switch from one network to another. For such situation a mechanism would be required in order to provide all time connectivity with the network. So, it can be said that movement of users will add congestion to the network, which will result in degradation of Quality of service and hence Quality of Experience too. So a mechanism is required to manage the resources effectively while improving and maintaining the Quality of Service. This paper introduces a framework which will enable service running on local public Cloud to populate on another public Cloud as per the requirement. It also addresses the situation where frequent migration of services would add network congestion. This will prevent network from experiencing huge traffic load and will offer automated service and resource management.

Keyword: Quality of Service, Cloud Computing, Service migration, Mobile computing, network congestion.

I. INTRODUCTION

Cloud computing is found everywhere. Enterprises are regularly searching for a new and an advance method to elevate the profits and reduce their costs. Such enterprises require various techniques that let them elevate and do not load them financially. From the existing technologies, Cloud computing has arose as a promising and interesting solution which provides access to virtual computing resources, platforms, and applications in a pay-as-you-go manner and on the demand of users. Cloud service users can access what they need and pay only for what they actually use. Because of Cloud computing the delivery of IT services is being raised to a new height that brings the all the comfort of basic requirements such as water and electricity to its users. The various benefits of Cloud computing, such as effectiveness in terms of cost and access, scalability and ease of management, encourage more and more IT companies and service providers to follow it and offer their services via Cloud computing models. Clouds provide three types of services, namely Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). IaaS Clouds provide various computing resources like processing power, storage, networks and other fundamental computing resources. The underlying Cloud infrastructure is managed by a provider. However, flexibility provided to so that they can select their virtual machine images to deploy these applications. In the Platform as a Service model, providers supply clients with tools and services to develop software applications. In addition to the IaaS restrictions, PaaS users do not have the ability to manage or control their virtual machine images and servers. SaaS providers allow customers to use the applications such as web based email, calendar or other applications running on a Cloud infrastructure. Neither the infrastructure nor the applications are controlled by users in this model. In recent years cloud computing becomes popular due to its simple nature. Mobile computing makes user able to move from one location to other location at the time of accessing and using services from cloud. So laptops and other mobile devices become popular. Clients can access cloud services while moving from one geo location to another. So, Cloud Computing and Mobile Computing opened up the various areas of research. So many research are being done on this field.

II. RELATED WORK

Researchers at the University of Minnesota are developed a migration technique for virtual machines within a Cloud that provides heterogeneity and dynamism in network topology and job communication patterns to allocate virtual machines on the available physical resources [6]. The aim of that is to bring physically closer any virtual machines which exchange a lot of traffic with each other. So by using that technique they can use faster connections within the same network hierarchical level instead of allowing traffic go through slower connections between levels. Since Cloud is actually a network of computers with a hierarchical structure, it is obvious that sometimes, there can be a lot of traffic between different hierarchical levels, depending on where data is stored and processed within the infrastructure. By placing virtual machines that performs individual parts of a bigger task, closer to each other, will reduce this cross-boundary communication which goes through slower network links compared to the much faster links which exist within the same hierarchical boundaries. Faster communication for the two VMs is the advantage of this, which improves the overall performance and reduces the congestion within the network. It will make the use of Cloud resources more efficient, which results in lower costs for the provider and more savings for the clients. Another technique invents a method called decentralized affinity aware migration technique that incorporates heterogeneity and dynamism in network topology and job communication patterns to allocate virtual machines on available physical resources[7]. It monitors network affinity between pairs of virtual machines and it uses bartering algorithm coupled with migration to dynamically adjust virtual machine placement such that communication overhead is minimized. That is if the communication overhead between two virtual machine of different network is larger then there is need to place these two virtual machines closer to each other (i.e, on same rack, cluster, node or network links). It increases the performance and reduces the traffic, for that affinity aware bartering and migration algorithm is used.

Another research project by the University of Minnesota involves the reshaping of the physical footprint of virtual machines within a Cloud [8-10]. The aim is to lower operational costs for Cloud providers and improve hosted application performance, by accounting for affinities and conflicts between co-placed virtual machines. It is achieved by mapping virtual machine footprints and then comparing them. When similarities are found in the footprint then the virtual machines are migrated to the same physical location. Its aim is to build control systems for Cloud environments which generates such footprint reshaping to achieve various objectives such as lower power consumption, higher reliability and better performance. It also reduces costs for providers and make Cloud services cheaper for clients.

Another proposed architecture aimed at improving the performance of Cloud technologies is called Media-Edge Cloud (MEC). It is an architecture that aims to improve the QoS and Quality of Experience (QoE) for multimedia applications [11]. It is achieved by a Cloudlet of servers which is running at the edge of a bigger Cloud. The main objective of that is to handle requests closer to the edge of the Cloud which reduces the latency. Any other processing is needed, then requests are sent to the inner Cloud, so the Cloudlets are reserved for QoS sensitive multimedia applications. So the aim is to divide the network hierarchy within the Cloud, such a way that physical machines that are closer to the Clouds outer boundaries will handle QoS sensitive services. Since these machines reside on the border of the Cloud, the data has to travel less distance within the Cloud ti reah to the clients. It improves QoE for clients and also reduces network congestion within the Cloud.

But none of research had considered user mobility into account. Also, all the research assume that only one provider is in control of a Cloud. Because of this the providers could not share the resources in efficient manner. This could create problem in future when the mobility increase and the data will have to travel a very long distance in order to reach to its mobile user. This will add congestion to the network and increase load on a particular Cloud. One of the research has considered user mobility and provided a framework which improves the QoS and QoE of service while providing the efficient resource management[1].

But one of the potential problem which is not addressed in above research is that moving a service could add overhead to the network. The amount of traffic that would be generated depends on the size of service being moved. This means if a service migration adds negligible improvement to the QoS or it is moving frequently, then its migration will only be the cause of network load. So, a potential solution to this problem is to prevent the recently migrated service and also instead of migrating whole service only migrate an instance of service.

III.CLOUD BASED SERVICE FRAMEWORK IN FUTURE

At present, the Internet works in resource centric manner. This means that the user gets the service by directly connecting to the physical location of the service. This connection is established by typing the URL which in turn is converted into IP address of the server which contains the required service. Same technique is used by the Cloud service also. The users connect to the Cloud and are offered with the services hosted by that Cloud. The problem of this scenario is that the user has to know the name physical resource it wants to access. If that physical resource is having some issue or problem or failure then it would add little room for the redundancy. A solution to this use of DNS, but it is not affordable for the smaller providers who offer services at low cost.

In future, a new service-centric access could be provided to the users. This approach will not only focus on the QoS but also be capable of moving service across the boundary of Cloud. To support the service populating framework a new concept of Cloud would be needed i.e. Open Cloud would be required. Currently all Clouds and their policies are under the control of their providers, so no third party can populate their service on those Clouds. So for the population of service by the third party, open Cloud is needed and this open Cloud can be viewed as a resource pool. To support the concept of Open Cloud new architecture or framework is needed. The framework is having same abstraction as that of the OSI model as shown in figure 1.

A. Service Management Layer

It deals with service registration into the Cloud. It includes Service and Security Level Agreement (SSLA) between the service providers and Cloud providers. It also keeps record of unique service ID assigned to each of the service. This layer handles the billing information between service and Cloud. For example, if a service provider wants to publish its service then, it has to define some security and QoS parameters which should be satisfied for adequate execution of that service. The same parameters are considered at the time of migration.

B. Service Subscription Layer

This layer handles the subscription of user or client for a particular service. It handles the unique user ID and SLAs between user and service. The billing information between the user and service is also managed by this layer. For

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 11, November 2015

example, when a user wants or requests a service, it is considered as subscriber of that service.

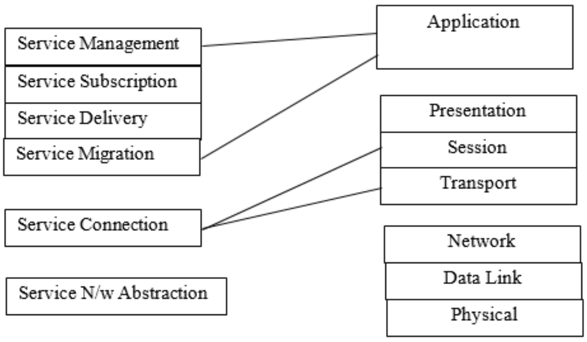

Figure 1: Abstraction of proposed framework

C. Service Delivery Layer

This layer is responsible for delivering the service to the user or client. The main logic required to process the QoS characteristics and mobility reside on this layer. This layer continuously monitors the QoS parameters to check whether it is as per the need or not. If it is not, then this layer marks that service as ready to migrate. After that it finds the Cloud which is more close to the client. Once such Clouds are marked, it sends the list of QoS parameters to those Clouds and whichever meets the requirement is chosen for service migration. It then gives instructions for migration to layer below it.

D. Service Migration Layer

This layer performs the task of service migration as per the instruction received from the above layer. Basically it allocates the resources required to run that service on new Cloud and then migrates the instance of service to the new Cloud and initiates the connection between the user and new Cloud.

E. Service Connection Layer

This layer performs the monitoring task. It monitors the connection between the user and service and sends the data to the above layer, using which migration decision is taken.

F. Service Network Abstraction Layer

This layer provides the transparency to the above layer, so that all network connection details could be known to the above layers.

IV. MECHANISM

All needed is a mechanism that is capable of gathering information about the network status and traffic as well as the information related to the Quality of Service, so that proper decision regarding migration of service could be taken. The mechanism which performs this task is known as QoS manager which is part of Service Connection Layer.

Next requirement is of a server, which will keep track of service IDs, and on which Cloud the particular service is registered. Also, it should keep track of user IDs and the details of how many users are accessing a specific service. This work is done by Service Tracking and Resolution or STAR. Using the information provided by the STAR the decision is taken that which Cloud is best for populating the particular service and meeting all the Quality of Service requirements. This STAR will be part of Service Delivery Layer.

One more mechanism would be required which will perform the actual task service migration. This would be done by Global Service Population Authority or GSPA. It is also part of Service Delivery Layer. GSPA is also responsible for updating the record of STAR.

The actual working of the proposed framework is as follows:

1. The user or subscriber will send the service request to the STAR.
2. The STAR will then find the cloud which is more close to the client requesting service. This will be done with the help of routing table, which is maintained by the STAR itself.
3. Once the start finds the appropriate Cloud, it sends the IP of that Cloud to the subscriber.
4. Using that IP address the user will access the service from that Cloud.

Above was the basic scenario. Now consider the scenario where subscriber or user changes its location.

1. User accesses service and also provides the information regarding the QoS experienced by it. Basically this information will be automatically read by the GSPA from the user's device.
2. If the QoS is as per the SSLA, then the user keeps accessing the service from the same Cloud. But if it is not, then, the GSPA searches for the Cloud which is near to the user's physical location. It is done by using the record stored in STAR.
3. Once the nearer Clouds are found GSPA checks which of them best suits the QoS requirement and migrates the instance of service to that Cloud.
4. Then it sends the updates to the STAR.

Next scenario is of frequent movement of user.

1. User accesses service and also provides the information regarding the QoS experienced by it. Basically this information will be automatically read by the GSPA from the user's device.
2. If the QoS is as per the SSLA, then the user keeps accessing the service from the same Cloud. But if it is not, then,
3. GSPA checks the previous migration time of the service.
4. If that time is less than the threshold value 't(threshold)' set by the GSPA i.e.

$$\textbf{t(migration)} < \textbf{t(threshold)}$$

then, the service is not migrated till t(migration) becomes greater than the t(threshold)

5. Otherwise, the GSPA searches for the Cloud which is near to the user's physical location. It is done by using the record stored in STAR.
6. Once the nearer Clouds are found GSPA checks which of them best suits the QoS requirement and migrates the instance of service to that Cloud.

7. Then it sends the updates to the STAR.

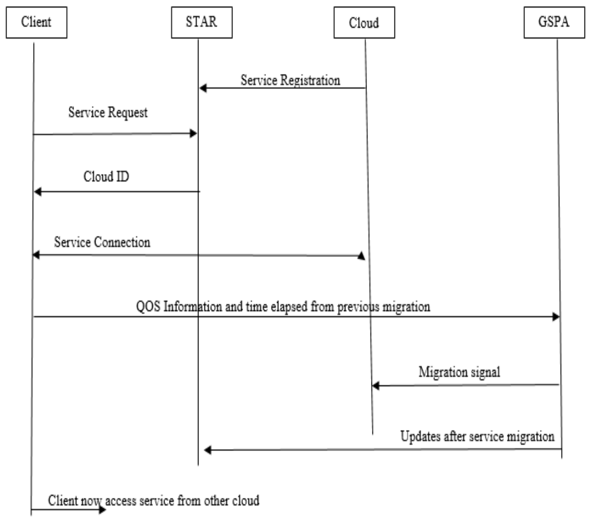

Fig 2: Working of proposed framework

V. CONCLUSION

Thus, this paper discussed the scenario of Cloud Computing in the future. Current service framework may create problem of network congestion in future when the demand for Cloud Computing will increase to a very large extent. To handle that situation a new framework and the concept of open Cloud has been introduced. It will be able to populate the service to the Cloud at other physical location as per the demand of QoS and network status. Also, it handles the situation where user frequently moves from one location to another by restricting the service migration for some period of time which is less than a threshold time set by GSPA. Thus, it has provided a potential solution which could solve the problem that may arise in future because of increasing demand of services by the users.

VI. FUTURE SCOPE

The current framework introduced maintains GSPA at each Cloud, which lacks in centralization. As the future work, the global or centralized GSPA and STAR could be maintained so that consistency could be maintained.

REFERENCES

[1] Fragkiskos Sardis, Glenford Mapp, Jonathan Loo" On the Investigation of Cloud BasedMobile Media Environment With Service Populating QoS Aware Mechanisms" IEEE trasaction on multimedia, vol.15, no.4, june 2013
[2] M. Mishra, A. Das, P. Kulkarni and A. Sahoo, "Dynamic Resource management using virtual machine migration," IEEE communication magazine, Sep 2012.
[3] V. S. Kushwah and A. Saxena, " A Security approach for Data migration in cloud computing," International Journal of Scientific and Research Publications, Volume 3, Issue 5, May 2013 1 ISSN 2250-3153
[4] P. Kulkarni, A. Tilak, N. Barhate and A. Basu, "Cloud Computing discussion paper for education, automotive and IT/ITeS sector," CII Pune: IT and ITeS Panel
[5] M. Shiraz, A Gani, R.H. Khokhar and R. Buyya, "A review on distributed application processing frameworks in smart mobile devices for mobile cloud computing," IEEE COMMUNICATIONS SURVEYS & TUTORIALS, VOL. 15, NO. 3, THIRD QUARTER 2013
[6] J. Sonnek, J. Greensky, R. Reutiman, and A. Chandra, "Starling: Minimizing communication overhead in virtualized computing platforms using decentralized affinity-aware migration," in Proc. 39th Int. Conf. on Parallel Processing (ICPP'10), San Diego, CA, USA, Sep. 2010.
[7] J. Sonnek and A. Chandra, "Virtual putty: Reshaping the physical footprint of virtual machines," in Proc. Workshop on Hot Topics in Cloud Computing (HotCloud'09), San Diego, CA, USA, Jun. 2009.
[8] C. A. Waldspurger, "Memory resource management in VMWare ESX server," in Proc. OSDI, 2002.
[9] T.Wood, G. Tarasuk-Levin, P. Shenoy, P. Desnoyers, E. Cecchet, and M. Corner, "Memory buddies: Exploiting page sharing for smart colocation in virtualized data centers," in Proc. 5th ACM Int. Conf. Virtual Execution Environments, 2009.
[10] D. Gupta, S. Lee, M. Vrable, S. Savage, A. C. Snoeren, G. Varghese, G. M.Voelker, and A.Vahdat, "Difference engine: Harnessingmemoryredundancy in virtual machines," in Proc. OSDI, 2008.
[11] W. Zhu, C. Luo, J. Wang, and S. Li, "Multimedia cloud computing," IEEE Signal Process. Mag., vol. 28, no. 3, pp. 59–69, May 2011.
[12] T. Brisko, RFC 1794, DNS Support for Load Balancing, IETF, 1995.