International Journal of Computer Science and Engineering Communications,
Vol.4, Issue.1, Page.1290-1293, (2016)
www.scientistlink.com

Hybrid Based Resource Reservation in the Cloud for Media Streaming Application

J.Mohana
Department of Computer Science and Engineering,
Ranipettai Engineering College, Vellore, India.
mohanajaivel@gmail.com

Abstract
Media streaming applications have recently attracted a large number of users in the Internet. With the advent of these bandwidth-intensive applications, it is economically inefficient to provide streaming distribution with guaranteed QoS relying only on central resources at a media content provider . Cloud computing offers an elastic infrastructure that media content providers (e.g., Video on Demand (VoD) providers) can use to obtain streaming resources that match the demand. Media content providers are charged for the amount of resources allocated (reserved) in the cloud. Most of the existing cloud providers employ a pricing model for the reserved resources that is based on non-linear time-discount tariffs (e.g., Amazon Cloud Front and Amazon EC2). Such a pricing scheme offers discount rates depending non-linearly on the period of time during which the resources are reserved in the cloud. In this case, an open problem is to decide on both the right amount of resources reserved in the cloud, and their reservation time such that the financial cost on the media content provider is minimized. Simple - easy to implement - algorithm for resource reservation and overload avoidance that maximally exploits discounted rates offered in the tariffs, while ensuring that sufficient resources are reserved in the cloud. Based on the prediction of demand for streaming capacity, our algorithm is carefully designed to reduce the risk of making wrong resource allocation decisions. The results of my numerical evaluations and simulations show that the proposed algorithm significantly reduces the monetary cost of resource allocations in the cloud and it is used to avoid overload that was accrued in the resource provisioning by using load balancer technique.
Index Terms:—Media streaming, cloud computing, non-linear pricing models, network economics

1 Introduction

MEDIA streaming applications have recently attracted large number of users in the Internet. In 2010, the number of video streams served increased 38.8 percent to 24.92 billion as compared to 2009 . This huge demand creates a burden on centralized data centers at media content providers such as Video-on-Demand (VoD) providers to sustain the required QoS guarantees[3] . The problem becomes more critical with the increasing demand for higher bit rates required for the growing number of higher definition video quality desired by consumers. In this paper, we explore new approaches that mitigate the new approaches that mitigate the cost of streaming distribution on media content providers using cloud computing. The delay sensitive nature of media streaming traffic poses unique challenges due to the need for guaranteed throughput (i.e., download rate no smaller than the video playback rate) in order to enable users to smoothly watch video content on-line[2]. Our main contribution in this paper is a practical—easy to implement—Prediction-Based Resource Allocation algorithm (PBRA) that minimizes the monetary cost of resource reservation in the cloud by maximally exploiting discounted rates offered in the tariffs, while ensuring that sufficient resources are reserved in the cloud with some level of confidence in probabilistic sense. We first describe the system model. We formulate the problem based on the prediction of future demand for streaming capacity. We then describe the design of our proposed algorithm for solving the problem . The results of our numerical evaluations and simulations show that the proposed algorithms significantly reduce the monetary cost of resource allocations in the cloud as compared to other conventional schemes.

2 System Organization

Recently, streaming resources (e.g., bandwidth) have become a feature offered by many cloud providers to content providers with intensive bandwidth demand. The streaming of media content to content viewers located at different geographical regions at guaranteed data-rate is a part of the service offered by the cloud provider. The common way of implementing this service in the cloud is by having multiple data-centers inside the networks of the access connection providers (e.g., Internet Service Providers, ISPs) located at appropriate geographical locations.

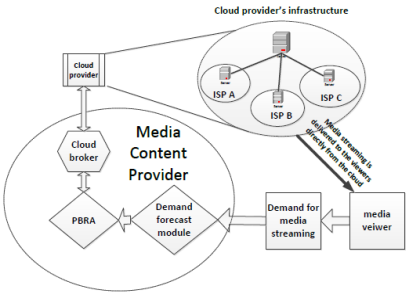


Fig 2 Structure of demand forecasting for media streaming

Cloud service providers may need to negotiate contracts with a number of ISPs to co-locate their servers into the networks of those ISPs. In this regard, another group of papers have focused on studying different types of contracts between cloud service providers and ISPs with the purpose of minimizing the expenses of cloud providers. However, an interesting design approach is to look at the resource reservation problem from the viewpoint of content providers. Obviously, content providers are more interested in minimizing their costs, i.e., the amount of money that they are charged directly by cloud providers.

To the best of our knowledge, very few studies have investigated the problem of optimizing resource reservation with the objective of minimizing the monetary costs for content providers. A good example is presented in, wherein a resource reservation optimization problem was formulated to minimize the costs of content providers, so called cloud consumers, using a stochastic programming model. In the process of problem formulation, uncertain demand and uncertain cloud providers' resource prices are considered. In contrast, the optimization problem formulated in my work takes into account a given probability distribution function obtained from aforementioned studies for the prediction of media streaming demands. Furthermore, the problem of cost minimization is addressed by utilizing the discounted rates offered in the non-linear tariffs. To the best of our knowledge, none of the previous papers has investigated the problem of cost minimization for media content providers in terms of monetary expenses by taking into account both the penalties caused by the over-provisioned or under-provisioned reserved resources, and the advance purchase of resources at cloud providers for just the right period of time.

2.1 VM Migration

Virtual machine monitors (VMMs) like Xen provide a mechanism for mapping virtual machines (VMs) to physical resources. This mapping is largely hidden from the cloud users. Users with the Amazon EC2 service, for example, do not know where their VM instances run. It is up to the cloud provider to make sure the underlying physical machines (PMs) have sufficient resources to meet their needs. VM live migration technology makes it possible to change the mapping between VMs and PMs While applications are running. The capacity of PMs can also be heterogeneous because multiple generations of hardware coexist in a data center.

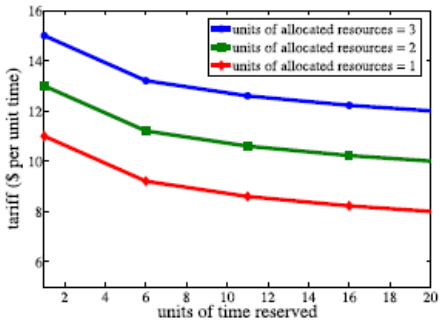
2.2 Prediction Based Resource Allocation

This approach optimally determine both the amount of reserved resources in the cloud and their reservation time - based on prediction of future demand for streaming capacity – such that the financial cost on the media content provider is minimized. The proposed algorithms exploit the time discounted rates in the tariffs, while ensuring that sufficient resources are reserved in the cloud without incurring wastage. I have evaluated the performance of our algorithms numerically and using simulations. The results show that our algorithms adjust the trade-off between resources reserved on the cloud and resources allocated on-demand. In future work, I shall perform experimental measurements to characterize the streaming demand in the Internet and develop our own demand forecasting module[1]. I shall also investigate the case of multiple cloud providers and consider the market competition when allocating resources in the clouds.

3 Results and Discussion

The system model that we advocate in this paper for media streaming using cloud computing consists of the following components (Fig. 1) _ Demand forecasting module, which predicts the demand of streaming capacity for every video channel

during future period of time. Cloud broker, which is responsible on behalf of the media content provider for both allocating the appropriate amount of resources in the cloud, and reserving the time over which the required resources are allocated. Given the demand prediction, the broker implements our proposed algorithm to make decision on resource allocations in the cloud. Both the demand forecasting module and the cloud broker are located in the media content provider site.

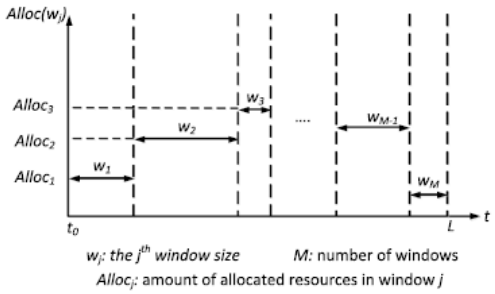


Cloud provider, which provides the streaming resources and delivers streaming traffic directly to media viewers. PBRA algorithm design.I summarize the assumptions that we use in our analysis as follows:

1) I assume that upon receiving the resource allocation request by the cloud provider from the media content provider, the resources required are immediately allocated in the cloud, i.e., updating the cloud configuration and launching instances in cloud datacenters incurs no delay.

2) Since the only resource that we consider in this work is bandwidth, it would be important to delve into the relation between the cloud provider and content delivery networks (CDN). However, I assume that the provisioning of media content to media viewers (clients of the media content provider) located at different geographical regions at guaranteed data-rate is a part of the service offered by the cloud provider. The common way of implementing this service in the cloud is by having multiple data-centres inside the networks of the access connection providers (e.g., ISPs) located at appropriate geographical locations .We assume that the media content provider is charged for the reserved resources in the cloud upon making the request for resource reservation (i.e., prepaid resources); and therefore, the media content provider cannot revoke, cancel, or change a request for resource reservation previously submitted to the cloud.

3) In clouds, tariffs (prices of different amount of reserved resources in $ per unit of reservation time) are often given in a tabular form. Therefore, the cloud service provider requires a minimum reservation time for any allocated resources, and only allow allocated resources in the cloud.



3.1 Hybrid Approach for Resource Provisioning

In this section, I consider the case, wherein the cloud provider offers two different types of streaming resource provisioning plans: the reservation plan and the on-demand plan. With the reservation plan, the media content provider reserves resources in advance and pricing is charged before the resources are utilized (upon receiving the request at the cloud provider, i.e., prepaid resources).

With the on demand plan, the media content provider allocates streaming resources upon needed. Pricing in the on-demand plan is charged by pay-per-use basis. In general, the prices (tariffs) of the reservation plan are cheaper than those of the on demand plan (i.e., time discount rates are only offered to the reserved (prepaid) resources). Amazon Cloud Front ,Amazon EC2 , Go Grid , MS Azure, Op-Source, and Terre-mark are examples of cloud providers which offer Infrastructure-as-a-Service (IaaS) with both plans. When the media content provider only uses the resource reservation plan, the under-provisioning problem can occur if the reserved (prepaid) resources are unable to fully meet the actual demand due to high fluctuating demand or prediction mismatch. Also, over-provisioning problem can occur if the reserved (prepaid) resources are more than the actual demand, in which parts of the reserved resources are wasted. However, when the cloud provider offers both the reservation plan and the on demand plan, the media content provider can allocate resources in the cloud more efficiently. In particular, the media content provider can use reservation plan to benefit from the time-discounted rate, while use the on-demand plan to dynamically allocate streaming resources to its clients at the moment when the reserved resources allocated using the reservation plan are unable to meet the actual demand and extra resources are needed to fit the fluctuated and unpredictable demands (e.g., flash crowd). We call this approach hybrid resource provisioning. This hybrid approach eliminates both the over-provisioning (over-subscribed) cost and the under-provisioning problem that may occur when using the reservation plan only. In this hybrid resource provisioning approach, tradeoff between the amount of resources allocated using the on demand plan and the amount of resources allocated using the reservation plan needs to be adjusted in which the hybrid approach can optimally perform. In this section, I propose an algorithm for this hybrid resource provisioning approach that maximally benefits from the time discounted rate offered in the resource reservation plan, while eliminating any over-provisioning cost of reserved resources such that the overall monetary cost of resource allocations in the cloud (including both the reserved resources and the on demand resources) is minimized.

4 Conclusion

This paper studies the problem of resource allocations in the cloud for media streaming applications. I have considered non-linear time-discount tariffs that a cloud provider charges for resources reserved in the cloud. I have proposed algorithms that optimally determine both the amount of reserved resources in the cloud and their reservation time—based on prediction of future demand for streaming capacity—such that the financial cost on the media content provider is minimized. The proposed algorithms exploit the time discounted rates in the tariffs, while ensuring that sufficient resources are reserved in the cloud without incurring wastage. I have evaluated the performance of our algorithms numerically and using simulations. The results show that algorithms adjust the tradeoff between resources reserved on the cloud and resources allocated on-demand.

References

[1] Cisco Systems Inc., San Jose, CA, USA, “Cisco visual networking index: Forecast and methodology, 2010–2015,” White Paper, 2010.
[2] Y. Liu, Y. Guo, and C. Liang, “A survey on peer-to-peer video streaming systems,” Peer-to-Peer Netw. Appl., vol. 18, no. 1, pp. 18– 28, 2008.
[3] Cisco Systems Inc., San Jose, CA, USA, “Data center virtualization and orchestration: Business and financial justification,” White Paper, 2007.