RESEARCH ARTICLE OPEN ACCESS

Cloud for Media Streaming

Pradeep Landge[1], Ashish Naware[2], Pooja Mete[3], Saif Maniyar[4], Mrs. Sweta Kale[5]

1,2,3,4,5(Department of Information Technology, RMD Sinhgad School of Engineering, Savitrabai Phule Pune University,PUNE)

Abstract:

Cloud computing is changing the way that video services are offered, enabling elastic and efficient resource allocation through auto-scaling. In this system, a new framework of cloud workload management for multimedia delivery services, demand forecast, predictive resource allocation and quality assurance as inter-dependent components. Based on the trace analysis of a production VoD system, propose time-series techniques to predict media resource demand from online monitoring, and determine resource reservations from multiple data centres. Prediction-Based Resource Allocation algorithm (PBRA) that maximize utilization offered in the tariffs, while ensuring that required resources are reserved in the cloud. Demand forecasting module, which predicts the user demand of streaming capacity for each media channel during future time period. Cloud broker is responsible from the side of the media content provider for both allocating the proper amount of resources in the cloud, and reserving the time period over which the required resources are allocated. This system has focused on improving the QoS and maximizing the utilization of resources for a particular time, such as Internet VoD systems, especially in the context of emerging cloud based services.

Keywords — **Media streaming, Cloud Computing, Non-linear pricing models.**

I. INTRODUCTION

Most Internet users watch live and on-demand videos through a web-browser and http-based video streaming, others may watch videos through client software downloadable from the video service provider. Other video delivery systems include on-line video gaming services such as OnLive. Despite the popularity of Internet video and the ever-increasing demand for improved video quality, most Internet video services remain best-effort systems. Since video flows are delay-sensitive, to guarantee the Quality of Experience (QOE) for an end-user, the video must be delivered from media servers to the end-user at a rate no less than the video bit rate (at least in the long run). However, QOE is usually not guaranteed in current Internet VOD systems, mainly due to the bounded egress bandwidth from video servers. Most video service providers over-provision the bandwidth capacity of their streaming servers to provide quality assurance. However, over-provisioning is costly and even ineffective sometimes, since a large amount of server capacity is unused during nonpeak hours, whereas in the event of a flash crowd, when a large number users join the system, the provisioned capacity may not even be sufficient.

Certain VOD systems adopt a Peer-to-Peer (P2P) architecture (e.g., Cool Streaming , PPLive , UUSee), where end-users can help the servers deliver video content to each other. While leveraging user upload bandwidth can alleviate the burden on media servers to some extent, the user resources are not dedicated and their contribution is not reliable. As a result, most P2P video systems are in fact peer-assisted systems, where the servers still play a major role in streaming, and thus face the same issue of how much server capacity should be provisioned. Because of the above reasons, in order to guarantee the QoE, what is missing from today's video delivery systems is a refined scheme to accurately predict the online user demand together with a flexible allocation mechanism that

can economically vary the resource provisioning over time.

II. GOALS & OBJECTIVES

To develop scheme to decide right amount of resources reserved in the cloud and their reservation time.

To ensure proposed system significantly reduces the cost of resource allocation as compared to conventional schemes.

III. EXISTING SYSTEM

For the resources reserved on the basis of discontinuous time-discount tariffs for Ex.Amazon cloudfront & amazonEC2.This cost scheme offers various discount rates based on the discontinuous time period during which the resources are reserved on cloud.here the problem is to decide the good or right amount of reserved resources and also their reservation time like the financial cost of the media content provider.

IV. SYSTEM MODEL

Main components of system model-

- **Media viewer:** It will request for media and media streaming is delivered to viewer directly from the cloud. Viewer can request media from web browser or Android application.
- **Media content provider:** It is used for prediction of future demand for video streaming capacity is required to help for resource reservation planning.
- **Demand forecasting module:** It 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 reservation time over which the required resources is allocated. Given the demand prediction, the broker implements our algorithm to make decision on resource allocations[1] in the cloud. Both the demand forecasting module and the cloud broker is present in the media content provider site. Cloud provider, which provides the streaming resources and delivers the streaming data directly to the media viewers.

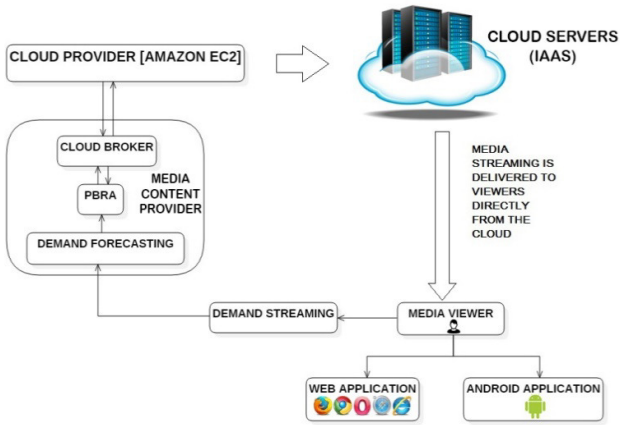

Fig. 1. System Model

- **Cloud broker:** It is responsible on behalf of the media content provider for both allocation of the appropriate amount of resources in the cloud, and reservation time over which the required resources is allocated. Given the demand prediction, the broker implements our algorithm to make decision on resource allocation in the cloud. The demand forecasting module and the cloud broker are present in the media content provider site.
- **Cloud provider:** It provides the streaming resources and delivers streaming traffic directly to media viewers.The cloud provider charges media content providers for the reserved resources according to the period of time during which the resources are retained in the cloud.
- **Prediction-Based Resource Allocation algorithm (PBRA):** It is the algorithm which minimizes the monetary cost of resource reservation[1] in the cloud by increase discount rates given by the tariffs, while ensuring that sufficient resources are reserved in the cloud with some extent level of confidence in probabilistic sense. Describe the design of our algorithm for solving the problem.

V. ALGORITHM DESIGN

The media content provider can predict the demand for streaming capacity of a video channel (i.e., the statistical expected value of the demand E[D(t)] is known) over a future period of time L

using one of the methods. The content provider reserves resources in the cloud on the basis of the predicted demand. The algorithm is based on time-slots with variable durations (sizes). In every timeslot, the media content provider makes a decision to reserve amount of resources in the cloud. The amount of resources to be reserved and the period of time over which the reservation is made (duration of time-slots) changes from one time-slot to another, and are determined in our algorithm to yield the minimum overall monetary cost .

Alternatively call a time-slot a window, and denote the window size (duration of the time-slot) by w. Since the actual demand changes during a window size, while allocated resources in the cloud remain the same for the whole window size (according to the third assumption above), the algorithm needs to reserve resources in every window j that are sufficient to handle the maximum predicted demand for streaming capacity during that window with some probabilistic level of confidence η.

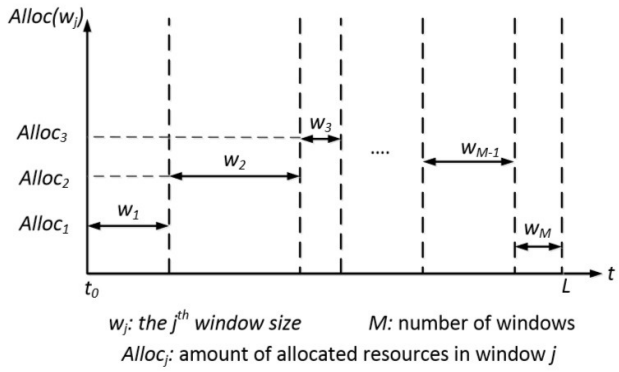

Fig. 2. PBRA algorithm design

The amount of reserved resources in window j by Alloc j. Since the decision on the amount of reserved resources is affected by the wrong prediction of future streaming demand, our on-line algorithm is designed to obtain accurate demand prediction (by enabling a mechanism that continuously updates the demand forecast module on the basis of the actual demand received at the media content provider overtime)in order to decrease the risk of making wrong resource reservation decisions.

The monetary cost of the reserved resources during window j by Cost(wj,Allocj), and can be computed as-

$$Cost(wj, Allocj) = tariff(wj, Allocj) \times wj$$

where tariff(wj , Allocj) represents the price charged by the cloud provider for amount of resources Allocj reserved for period of time (window size) wj. Note that the values of tariff and Cost in any window j depend on both the amounts of allocated resources (Allocj) and the time period over which resources are reserved (wj). Also note that the algorithm runs on-the-fly. More specifically, the demand forecast module predicts streaming capacity demand in the upcoming period of time L and adds this information to the algorithm. The algorithm upon receiving the demand prediction, computes the proper size of window j (i.e., w∗ j), and the right amount of reserved resources in window j (i.e., Alloc∗ j), such that the cost of the reserved resources during window j (i.e Cost (wj , Allocj) in (2)) is minimized; or equivalently.

Hence, the objective of our algorithm is to minimize Cost(wj,Allocj) ∀j, subject to Probability(D(t) ≤ Alloc (t)) ≥ η, ∀ t ∈ L. In other words, our objective is to minimize the monetary price of reserved resources such a way that the amount of reserved resources at any instant of time is guaranteed to meet the actual demand with probabilistic confidence equals to η. As earlier, D(t) is a random process that follows a log normal distribution with mean E [D(t)] and variance (σ) characterized, respectively. Thus, using the constraint above, and for any window size wj, now compute the minimum amount of required reserved resources during window j(Allocj) by solving the following formula for Allocj.

$$\int_{0}^{Alloc_j} \frac{1}{x.\sigma\sqrt{2\pi}} \left(e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu_{max}}{\sigma}\right)^2} \right) dx = \eta$$

where μmax is the maximum value of the predicted streaming traffic demand during the window j (i.e μmax = argmax(E[D(t)]) ∀ t ∈ wj).

Note that the equation follows from the log normal probabilistic allocation of the demand for streaming capacity.

As earlier, the cloud service provider often requires a as small as reservation time for any allocated resources is wmin, and only allows discrete levels (categories) of reservation times for any amount of allocated resources in the cloud. Therefore, assume that any reservation time period required at the cloud has to be in multiplicative order of wmin (i.e wj = k · wmin, where k is +ve integer). Thus, the algorithm employs a demo window (wh) to assist in making optimum decision on the size of each window j. In particular, for every window j, the algorithm starts an recursion process with a trial window of size wh=wmin, and computes the cost rate (Xh=tariff(wh,Alloch), where h is repetition index), and Alloch is computed by solving Eq. for Alloc.

Recall that due to the time discount rates offered in the tariffs, increasing the time during which the allocated resources are reserved may lead to less pricing cost (higher discounted rate) on the media content provider. However, increasing the window size (time-slot) significantly may also result in high over-provisioning (over-subscribed) price as the media content provider has to allocate resources in the cloud that meet the highest demand during the window time period. Thus, in order to recognize whether the price is decreasing or increasing with increasing the window size, the demo window size (wh) is increased one wmin unit in every iteration (i.e., wh = wh+wmin) and the cost rate of this new trial window size is computed (Xh+1). The algorithm keeps increasing the trial window size until wh = L in order to scan the whole period of time over which the demand was predicted (L) (Fig. 3), and finds the value of wh that yields the minimum cost that is the optimum requied size of window j (w∗ j). Since L is the period of time over which the future demand is predicted, then wmin ≤ w∗ j ≤ L.

During every window, the media content provider receives the actual streaming demand for the video traffic, which may be different from the predicted demand. According to the real demand, Demand forecast module updates its prediction and adds the algorithm with a new predicted demand for other future period of time L. The algorithm upon receiving the updated demand prediction, computes the optimum size of the next window, and reserves optimum required resources in the next window, and so on.

given by the tariffs, while ensuring that sufficient resources are reserved in the cloud with some extent level of confidence in probabilistic sense. Describe the design of our algorithm for solving the problem.

VI. PROPOSED SYSTEM

The algorithm is for reservation of resources that maximally exploits the discounted rates in tariffs,ensuring that sufficient resources are reserved. Predicting the demand for streaming capacity is designed in a way to deduct the risk of wrong decision making for resource allocation.

VII. CONCLUSION

The system is focused on improving the Quality-of-Service (QoS) and reducing the operational cost in large-scale multimedia delivery systems, such as Internet Video on Demand (VoD) systems, especially in the context of emerging cloud-based services. finally, investigate related business models for VoD providers to use the cloud services.

VIII. FUTURE WORK

In the future work, formulate the bandwidth resource as a distributed optimization problem, for which to propose new distributed algorithms other than the traditional gradient algorithm, to ensure fast and robust convergence in very large systems.

ACKNOWLEDGEMENT

We take this opportunity to thank our project guide and Head of Department Mrs. Sweta Kale, for her valuable guidance and immense support in providing all the necessary facilities, moral support to encourage us, which were indispencible in the completion of this project. We are also thankful to all the staff members of the Department of Information Technology of RMD Sinhgad School Of Engineering, Warje for their valuable time,support, comments, suggestion and persuation. We would also like to thank Institute for providing

required facilities like internet access and impostant books.

REFERENCES

1. *S. Chaisiri, B-S Lee, and D. Niyato, "Optimization of Resource Provisioning Cost in Cloud Computing," in IEEE Transactions on Services Computing, vol. 5, no. 2, pp. 164–177, 2012.*

2. *S. Islam, J. Keung, K. Lee, and A. Liu, "Empirical Prediction Models for Adaptive Resource Provisioning in the Cloud," in Future Generation Computer Systems, vol. 28, no. 1, pp. 155–162, 2012.*

3. *W. Zhu, and C. Luo and J. Wang and S. Li, "Multimedia Cloud Computing," in IEEE Signal Processing Magazine, vol. 28, no. 3, pp. 59–69, 2011*

4. *A Survey on Peer-to-Peer Video Streaming Systems Yong Liu · Yang Guo · Chao Liang*