A Survey on Forecasting Analysis for Resource Allocation Strategies in Cloud for Media Streaming Applications

P Dileep Kumar Reddy
Lecturer, Dept. of CSE, JNTUA, AP

Y.Sajid Hussain
M. Tech, Dept. of CSE, JNTUA, AP

***Abstract*— Cloud computing is a new technology that has become a huge demand in video streaming applications. Users can access video streaming applications from anywhere with the help of cloud. Now a days, more companies and marketing firms are able to hire of cloud resources for storage and other computational purposes, so that their infrastructure costs can be significantly reduced. Furthermore, they may make use of the entire company have access to applications on the basis of pay-as-you-go model. The resource allocation mechanism which is based on reservation method can be improved further with effective forecasting model that can provide good approximation to improve resource allocation strategies. This can help to improve customer needs and the growth of video streaming businesses that are based on cloud computing. The amount of resources reserved in the cloud are charged by the media content providers. Resource allocation is performed with the objective of reducing the costs associated with it. Time Series Analysis implemented to predict the allocated resources in the cloud and to reduce the cost. The other challenges of reserved resources meets the requirements of the customer and VoD application requirements. In this paper, various strategies are discussed in detail**.

Index Terms— **Cloud Computing, Media Streaming Applications, Time Series Analysis, and Resource Allocation**
.

I. INTRODUCTION

Huge number of users in the Internet is attracted by the media streaming application. As comparison to recent years, the sum of video streams supplied has increased 38.8% to $ 24.92 billion. Due to this massive demand creates supporting a burden to centralized data centers on media content providers, such as Video on Demand (VoD) to support the required QoS guarantees.

Video delivery systems, such as Youtube, Netix, Hulu, etc., have gained unprecedented popularity on the Internet nowadays. According to Cisco Visual Networking Index: Forecast and Methodology," globally, Internet video traffic will be 55% of all consumer Internet traffic in 2016, up from 51% in 2011. Most internet users to view live and on-demand videos through a web browser and HTTP-based video streaming, can other people videos through client software to download the video service provider. Other video delivery systems include on-line gaming services such as OnLive.

Despite the popularity of internet video and the increasingly growing demand for better video quality, most internet video services remains best effort systems. Since video ows are delay sensitive, the Quality of Service (QoS) guarantee to the end user, the video media servers to the end-user must be at a rate which is greater than the video bit rate (at least in the long term). However, QoS usually not guaranteed in the current Internet VoD systems, particularly because of the limited output bandwidth video servers. Most video service providers to meet the bandwidth capacity of their streaming servers quality providence. However, over-provisioning is costly and equally ineffective sometimes because a large amount of server capacity is not used during off-peak hours, Whereas, to join in the event of a shaft crowd, as a large number of users on the system, the expected capacity cannot even sufficient.

Certain VoD systems choose a peer-to-peer (P2P) architecture, where the end users can help to deliver the video content servers to each other. While leveraging user upload band- width 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 QoS, what is missing from today's video delivery systems is a refined scheme with a flexible allocation mechanism the online user demand economic is accurately predicted, this can vary the resource provisioning overtime.

II. .RELATED WORK

The demand of the user and the utilization of CPU is widely analyze [1]. Y. Lee et al. proposed a method called prediction called Radial Basis Function(RBF) networks to forecast the demand request by the user in web applications[2]. The demand activities of a user in P2P streaming using non-stationary time series model was predicted in [3]. The proposed literature concentrates on resource allocation and its impacts on cloud users, cloud

All Rights Reserved © 2017 IJSETR

providers and with forecasting technique to demonstrate the proof of concept.

III. Cloud Resource Allocation Strategies

Based on such demand forecast, the developed resource allocation mechanisms for data centers in the cloud to satisfy video application requirements with the minimum cost. The study gives information about a new type of service where companies like Netflix can make reservations for (outgoing) bandwidth guarantees from the cloud. The objective is to reserve as little bandwidth as possible while still guaranteeing the video delivery quality. An auto-scaling system has been proposed that can adjust resource allocation for each cloud tenant (a video company or channel) to closely match its short term demand prediction, accommodating both the estimated demand expectation and unexpected demand variation. When multiple data centers coexist in the cloud, the optimal direction policy of tenant demands to different data centers has been derived and proposed a practical heuristic solutions and evaluated the methods through extensive trace-driven simulations.

Resource Allocation Strategy (RAS) is to be satisfied with the integration of cloud provider activities for the use and allocation of scarce resources within the boundaries of cloud environment with a view to the needs of the cloud application. It requires the type and amount of the resources that are necessary for each application to complete a user command. An optimal RAS the following criteria occur as follows:

a**) Resource contention** arises when two applications trying to get to the same source at the same time.

b**) scarcity of resources** arises if there are limited resources.

c) **Resource fragmentation** arises if resources are isolated. [There will be plenty of resources but unable to allocate the necessary application].

d) **Over-provisioning** of resources arises if the application gets surplus funds than the requested one.

e) **Under-provisioning** of resources arises if the application is assigned fewer numbers of funds than demand.

IV. Cloud Workload Management and Resource Allocation

Cloud bandwidth reservation is becoming technically feasible. There have been proposals on data center traffic engineering to offer elastic bandwidth guarantees for egress traffic from virtual machines (VMs) [11]. The idea of virtual networks is implemented to connect the VMs of the same tenant in a virtual network with bandwidth guarantees [7, 11]. Furthermore, it is proposed to explicitly rate control to allocate bandwidth based on current deadlines. Such research progress, the cloud more attractive to bandwidth-intensive applications such as video-on-demand and MapReduce calculations that depend on the network to created large amounts of data at high prices. Netflix as a major provider VoD, moved to store data, video encoding and streaming servers, Amazon AWS in 2010 [4].

Virtualization techniques for supporting cloud-based IPTV services are also being developed by major U.S. VoD providers such as AT&T [5]. Novel flow control algorithms and improvements over TCP have been presented to ensure QoS-aware video delivery from cloud data centers. Furthermore, video demand forecasting techniques have been proposed, such as the non-stationary time series models introduced in [18, 19, 20], and video access pattern extraction via principal component analysis in [12]. All these recent advances will help the realization of video delivery from data centers, and will enable efficient and quality-assured management of video workload in the cloud. Predictive and dynamic resource provisioning has been proposed mostly for VMs and web applications with respect to CPU utilization [9, 10, 20]. Work exploits the unique characteristics of VoD bandwidth demands and is distinct from the foregoing works in three aspects. First, our bandwidth workload consolidation is as simple as solving convex optimization for a load direction matrix. Leverage the fact that unlike VM, demand of a VoD channel can be fractionally split into video requests. Second, the system forecasts not only the expected demand but also the demand volatility, and thus can control the risk factors more accurately. In contrast, most previous works [9, 10] assume a constant demand variance. Third, exploit the statistical correlation between bandwidth demands of different video channels to save resource reservation that considers VM consolidation with independent random bandwidth demands.

V. Measurement, Prediction in VoD Systems

Video-on-Demand systems have gained enormous popularity on today's Internet. Some examples of production systems include Metacafe, Netflix, Hulu, Youtube, etc. Since the inception of using a mesh-like P2P architecture as a solution to live media streaming, exemplified in CoolStreaming, peer assistance has also been introduced into video-on-demand services to increase system scalability. Significant research efforts [13], [14], [15], [16] have been devoted to the measurements of video streaming systems, with a focus on the user behavior, scalability and system performance. A number of coding and ad hoc optimization techniques [8], [15] have also been proposed to enhance their performance.

The importance of bandwidth demand estimation to capacity planning in Internet VoD systems has been recognized recently. It is shown that estimating time-varying demands in a large-scale IPTV network can help the system optimally place content on its geographically distributed servers [6]. Toward this goal, the recent demand history is used as an estimate of future demand in each video channel [6] and demands for previously released movies are used as the demand prediction for newly released movies. Apparently, this simple method does not yield accurate forecasts. Because measurements show that video workload

All Rights Reserved © 2017 IJSETR

illustration regular daily schedule [6, 19, 22], various techniques have been recently proposed to predict large-scale VoD traffic. A Autore-progressive Integrated Moving Average (ARIMA) model is introduced in [22] in order to predict the evolution of populations video streaming systems. To account for diurnal effects, seasonal ARIMA models have been introduced in the previous work [18, 19] to predict non-stationary demand evolution at a fine granularity.

However, most existing forecast methods for video demand assume a constant forecast error variance, and thus provision a constant amount of resource cushion for quality assurance. In fact, the measurements of the UUSee system show that bandwidth demand is subject to rapid changes in some periods, while remaining tranquil and highly predictable in other periods.

Volatility reduction in the mixed traffic of multiple channels is similar to the idea of statistical multiplexing and resource overbooking [21] in shared hosting platforms, where the resources are booked to satisfy a certain percentile of demand in each application instead of its worst-case demand, so as to enhance resource utilization. However, the volatility reduction discussed here is novel in three aspects. First, we are concerned with forward-looking resource allocation and volatility forecasts for future demand, while in [21] the resource usage of each application is profiled in an online and fixed manner, ignoring the change of demand patterns over time. Second, study focuses on large-scale VoD systems, where the concurrent number of users can ramp up by several hundreds or thousands in tens of minutes. In this scenario, any fixed resource usage profiling for small video channels (e.g., those with a user population of 20 in [21]) will be insufficient. Last but not least, it do not assume independence between the demands of channels. Instead, it accurately quantify the conditional demand variance in each channel, which enables the use of financial instruments such as hedging and diversification to achieve cost-effective server management with service level guarantees.

Principal Component Analysis (PCA) has been proposed in [12] to extract video demand evolution patterns over longer periods (of weeks or months) and forecast coarse-grained daily populations. An approach is utilize to find the common factors that drive the demand evolution of all coexisting video channels using PCA at a fine granularity and then make forecasts for individual video channels as regressions from factor forecasts obtained from the seasonal ARIMA model. Such an approach combines the strengths of both PCA and the seasonal ARIMA model. Unlike [12], these approach makes short-term predictions with a lead time of 10 minutes, enabling fine-grained auto-scaling of resource allocation.

Recently, there has been an increasing interest in applying statistical learning tools to diagnose, predict and improve the reliability of large-scale distributed systems. Mahimkar et al. [17] focus on characterizing performance issues in a large-scale IPTV network, and propose Giza, which applies multi-resolution data analysis to localize regions and physical components in the IPTV distribution hierarchy that are experiencing performance problems. It represents one of the first attempts to characterize the dependencies among the demand statistics and performance metrics in a large-scale video delivery system. A systematic time-series analysis approach is introduced that learns the user behavior and system dynamics from online measurements, and utilizes the learned rules to predict demand and performance for the future.

VI. Conclusion

In this paper we studied the video streaming applications that cloud based. The VoD related applications over cloud can have subscribers in large scale. However, it is important to have quality of services. In cloud paradigm, an effective resource allocation strategy is required for achieving user satisfaction and maximizing the profit for cloud service providers. This paper summarizes the classification of RAS and its impacts in cloud system and time series to have accurate prediction of streaming forecast for different videos associated with a VoD system. This research can be extended further to investigate the prediction methods in the context of cloud based Internet Protocol Television (IPTV).

VII. References

[1] 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.

[2] Y. C. Lee and Y. Albert, “Zomaya: Rescheduling for reliable job completion with the support of clouds,” in Future Generation Computer Systems, vol. 26, no. 8, pp. 1192–1199, 2010.

[3] D. Niu, Z. Liu, B. Li, and S. Zhao, “Demand forecast and performance prediction in peer-assisted on-demand streaming systems,” in Proc. of IEEE Infocom conference, pp 421–425, 2011.

[4] Four Reasons We Choose Amazon's Cloud as Our Computing Platform. The Netix\Tech" Blog, Dec. 14 2010.

[5] V. Aggarwal, X. Chen, V. Gopalakrishnan, R. Jana, K. K. Ramakrishnan, and V.A. Vaishampayan. Exploiting Virtualization for Delivering Cloud-based IPTV Services. In Proc. of IEEE INFOCOM Workshop on Cloud Computing, 2011.

[6] D. Applegate, A. Archer, V. Gopalakrishnan, S. Lee, and K. K. Ramakrishnan. Optimal Content Placement for a Large-Scale VoD System. In Proc. ACM International Conference on Emerging Networking Experiments and Technologies (CoNEXT), 2010.

All Rights Reserved © 2017 IJSETR

[7] H. Ballani, P. Costa, T. Karagiannis, and A. Rowstron. Towards Predictable Data center Networks. In Proc. of SIGCOMM'11, Toronto, ON, Canada, 2011.

[8] B. Cheng, X. Liu, Z. Zhang, H. Jin, L. Stein, and X. Liao. Evaluation and Optimization of a Peer-to-Peer Video-on-Demand System. J. Syst. Archit., 54(7):651{663, Jul. 2008.

[9] D. Gmach, J. Rolia, L. Cherkasova, and A. Kemper. Workload Analysis and Demand Prediction of Enterprise Data Center Applications. In Proc. of IEEE Symp. Workload Characterization, 2007.

[10] Z. Gong, X. Gu, and J. Wilkes. PRESS: PRedictive Elastic ReSource Scaling for Cloud Systems. In Proc. of IEEE International Conference on Network and Services Management (CNSM), 2010.

[11] C. Guo, G. Lu, H. J. Wang, S. Yang, C. Kong, P. Sun, W. Wu, and Y. Zhang. Second-Net: a Data Center Network Virtualization Architecture with Bandwidth Guarantees.In Proc. of ACM International Conference on Emerging Networking Experiments andTechnologies (CoNEXT), 2010.

[12] G. Gursun, M. Crovella, and I. Matta. Describing and Forecasting Video Access Patterns. In Proc. of IEEE INFOCOM Mini-Conference, 2011.

[13] C. Huang, J. Li, and K. W. Ross. Can Internet Video-on-Demand be Pro_table? In Proc. of SIGCOMM'07, Kyoto, Japan, August 2007.

[14] Y. Huang, T. Z. J. Fu, D.-M. Chiu, J. C. S. Lui, and C. Huang. Challenges, Design and Analysis of a Large-scale P2P-VoD System. In Proc. of SIGCOMM'08, Seattle, Washington, August 2008.

[15] Z. Liu, C. Wu, B. Li, and S. Zhao. UUSee: Large-Scale Operational On-Demand Streaming with Random Network Coding. In Proc. of IEEE INFOCOM, 2010.

[16] J.-G. Luo, Q. Zhang, Y. Tang, and S.Q. Yang. A Trace-Driven Approach to Evaluate the Scalability of P2P-Based Video-on-Demand Service. IEEE Trans. Parallel Distrib. Syst., 20(1):59{70, Jan. 2009.

[17] A. Mahimkar, Z. Ge, A. Shaikh, J. Wang, J. Yates, Y. Zhang, and Q. Zhao. Towards Automated Performance Diagnosis in a Large IPTV Network. In Proc. Of SIGCOMM'09, Barcelona, Spain, 2009.

[18] D. Niu, B. Li, and S. Zhao. Understanding Demand Volatility in Large VoD Systems. In Proc. of the 21st International workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV), 2011.

[19] D. Niu, Z. Liu, B. Li, and S. Zhao. Demand Forecast and Performance Prediction in Peer-Assisted On-Demand Streaming Systems. In Proc. of IEEE INFOCOM Mini-Conference, 2011.

[20] D. Niu, H. Xu, B. Li, and S. Zhao. Quality-Assured Cloud Bandwidth Auto-Scaling for Video-on-Demand Applications. In Proc. of IEEE INFOCOM, Orlando, Florida,
2012.

[21] B. Urgaonkar, P. Shenoy, and T. Roscoe. Resource Overbooking and Application Pro_ling in Shared Hosting Platforms. In Proc. of USENIX Symposium on Operating Systems Design and Implementation (OSDI), 2002.

[22] C.Wu, B. Li, and S. Zhao. Multi-Channel Live P2P Streaming: Refocusing on Servers. In Proc. of IEEE INFOCOM, 2008.

All Rights Reserved © 2017 IJSETR