International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2017 IJSRCSEIT | Volume 2 | Issue 2 | ISSN : 2456-3307

Request and Redirection in Content Delivery Network using Load Balancing System

Gayathri V, Vasumathi N, L. Thangapalani

Department of Computer Science and Engineering, Prince Dr. K. Vasudevan College of Engineering and Technology, Chennai, Affiliated Anna University, Tamilnadu, India

ABSTRACT

A Cloud technology provides a new opportunity for Video Service Providers to running a virtual machine and hosting video applications in a cost effective manner. Under this project, a VSP may rent virtual machines VMfrom multiple geodistributed data centers that are close to video request to run their services. Cloud Data Center are located in different location.Based on the user request we predict geographical location request and Redirect the nearby data centers. If the server traffic is high then by using load balancing technology their request is redirect to next nearby Cloud data center in virtual machine. A cloud provider deploys its applications in geographically distributed CDNs to improve Stability and Reliability. For cost and performance, each CDN provides services through multiple CDN that deliver traffic between millions of user and the CDNs provider. The geographical diversity of the bandwidth and energy cost brings the CDCs provider a big challenge of how to minimize the bandwidth and energy cost of the CDCs provider. So that the performance will increase and traffic delay is low. We propose a systematic method called Cost Aware Workload Scheduling and Admission Control for Distributed Cloud Data Center(CAWSAC). This scheduling strategy can intelligently dispatch requests, and achieve lower cost and higher throughput for the CDNs provider.

Keywords : Cloud Data Centers(CDC),Video Service Provider (VSP), Cloud Service Provider (CSP), Workload scheduling

I. INTRODUCTION

Cloud data centers (CDCs) are shared to concurrently operate multiple applications that provide services to global users. It consumes tens of megawatts of power for running and cooling tens of thousands of servers. To achieve low latency and high availability, applications are replicated and deployed in multiple CDCsdistributed in different locations. However, requests must first go through the wide-area network (WAN) consisting of multiple available ISPs and then arrive in distributedCDCs. For example, Google's WAN provides great amount of applications including mail, search and video to global users. Existing cloud providers deliver at least petabyte traffic per day. Thus, the CDCs provider suffers huge ISP bandwidth cost to deliver traffic. Besides, the bandwidth cost of each ISP is specified based on the service-level agreement (SLA) signed with the CDCs provider, with some ISPs being much cheaper than others are. However, recent work ignores the diversity in the bandwidth cost and capacities of ISPs, and cause high cost and request loss. In addition, CDCs are located in different areas where the energy cost is regional, i.e., the energy cost in distributed CDCs also exhibits geographical diversity. Therefore, it is challenging to minimize the total cost of the CDCs provider in a market where the bandwidth and energy this process takes place by two methods. The first process executes the revenue based workload admission control method according to outside arrival requests, and provides the input for the second stage.

The second process runs the cost-aware workload scheduling to specify the optimal workload assignment that can minimize the total cost of the CDCs provider.

Extensive trace-driven experiments based on the real-life workload in Google production cluster are

conducted to evaluate the proposed cost request and redirection methods, which just redirect the quality of the videos, which is not take place then it, redirect to next link process.

II. METHODS AND MATERIAL

1. Related Works

Here, we discuss the related work and presents the contribution of the workload admission control and Request and redirection in comparison to existing works.

A) Content Admission Control

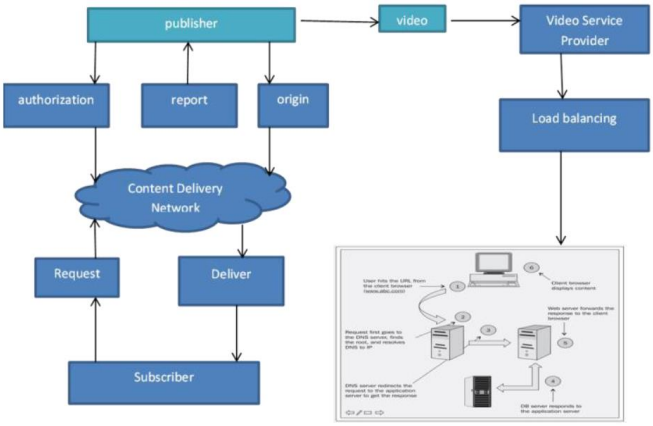
Admission control is to protect servers from overload and to guarantee the performance of applications. It used to provide joint response routing and request mapping in geographically distributed CDCs. The reinforcement learning method and cascade neural networks are integrated to improve the scalability and agility of the system. The present a simple algorithm that drops excessive workload provided that the performance is met. However, these works only consider a single application and ignore the resource conflict among multiple applications. Our revenue-based workload admission control method can judiciously admit requests by considering priority, revenue and the expected response time of different request flows.

B) Performance Modelling

Theypropose a stochastic reward netsbased analytical model to evaluate the performance of cloud infrastructure. The behavior of a cloud system is quantified and evaluated using the defined performance metrics. However, these works cannot accurately model the energy cost of distributed CDCs. In this concept, we model the energy cost based on servers and consider its geographical diversity for minimizing the total cost of the CDCs provider.

2. Architecture of Distributed CSP & VSP

Content Delivery Network (CDN) to host video services on their various datacenters distributed in various regions. And then CDN to predict user location and then provide the user priority request. With our approach the video service provider provide to different quality videos like High, Medium, Low is able to provide an efficient. Cost effective and quality service to any number of clients. The Cloud Data Center handle a resource renting from multiple CSPs and load balance the user requests to these resources in a nearly virtual machine. And then virtual machine has been created in a different location by admin controller. Renting each virtual machine hosted in cloud for particular time period . The framework is capable of handling heterogeneous types of user requests, workloads and Manage user requirements. For that propose an algorithm to solve the jointed stochastic problem to balance the cost saving and performance monitoring. Then CDN to predict user location and then provide the user priority request.With our approach the video service provider provide to different quality videos like High,Medium,Low is able to provide an efficient. The Cloud Data Center handle a resource renting from multiple CSPs and load balance the user requests to these resources in a nearly virtual machine.



A) Algorithm Dyreceive

(Dynamic Request Redirection and Resource Provisioning for Cloud-Based Video Services under Heterogeneous Environment) Allocate resources for cloud based video services on user request from multiple regions to distributed data centers and dynamically computes the near and optimal virtual machine. Video Service Providers (VSP) for running compute-intensive video applications in a cost effective manner. VSP may rent virtual machines (VMs) Multiple geo-distributed datacenters that are close to video requestors to run their services. Optimizing the number of VMs of each type rented from datacenters located in different regions in a given time frame

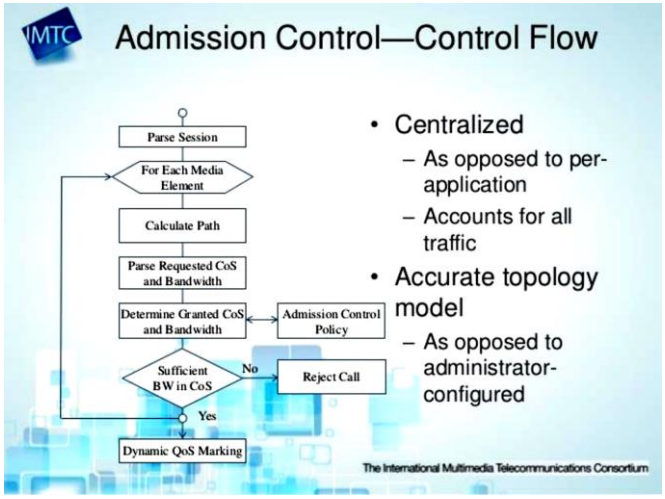
becomes essential to achieve cost effectiveness for VSPs.

B) Procedure

1. Input:
2. sk, ωc, Wmax, ℓc a c r, Amax rc, pkd(τ), ρk d, drent, V, a, b, u, v (∀c ∈ C, ∀d ∈ D, ∀r ∈ R, ∀k ∈ K);
3. Output:
4. n c,k d (τ), λcrd(τ) (∀c ∈ C, ∀d ∈ D, ∀r ∈ R, ∀k ∈ K);
5. Initialization step: Let τ = 0, st = cputime, and set Q c d(0) = 0, Hc d(0) = 0, (∀c ∈ C, ∀d ∈ D), ddec(0) = 0;
6. while the service of VSP is running do
7. calculate time slot τ ,τ = (curtime − st)/60s;
8. estimate the decision overhead ddec(τ) based on ddec(t), t ∈ [τ − 5, τ − 1];
9. Resource provisioning:
10. foreach datacenter d ∈ D d
11. if (τ mod md) == 0 then
12. Observing the queue backlogs Q c d (τ), H c d (τ) and the VM price p k d (τ) at current time;
13. Getting the VM provisioning strategy (n c,k d (τ)) by solving the problem (24) using CVX tool;
14. Request redirection:
15. if request arrives at system then
16. foreach r ∈ R, c ∈ C do
17. Observing the queue backlogs Q c d (τ), Hcd(τ) , the network delay drd and estimating the computation delay dcomp(τ) at current time;
18. Getting the request redirection strategy λ c rd (τ) by solving the problem (21) using (22);
19. Update the queues Q c d (τ), Hcd(τ) according to queue dynamic equation (12)(13) respectively.
20. Record the decision-making time consumed at current time slot ddec (τ).

C) Admission Control Problem

The workload admission control method inclines to higher admit higher requests. If there are not enough server to execute higher priority request, the request may experience extremely long response time and brings less are more revenue to CDCs provider. In this case to maximize the total admission control method can refuse some of higher priority request intelligently admit lower priority once there is more request.



The CDN provider and the value of penalty function in every time slot using the DYRECEIVE algorithm, to solve the Admission control problem.

c	n	$M_{c,n}$	$b^a_{c,n}$($/hour)	$b^s_{c,n}$($/hour)	RT^{user}_n(ms)	$\mu_{c,n}(10^3$ requests/second)
c=1	n=1	1000	0.0007	0.00007	1	9
	n=2	800	0.0013	0.00013	1.25	7.5
	n=3	800	0.0020	0.0002	2	3
	n=4	1200	0.0033	0.00033	2.5	1.5
		M_1=3800				
c=2	n=1	1800	0.0001	0.00001	1	4.5
	n=2	1800	0.0002	0.00002	1.25	3.75
	n=3	1200	0.0003	0.00003	2	1.6
	n=4	2500	0.0004	0.00004	2.5	1
		M_2=7300				
c=3	n=1	1000	0.0003	0.00003	1	10.5
	n=2	1000	0.0007	0.00007	1.25	6
	n=3	1200	0.0010	0.0001	2	3
	n=4	2000	0.0017	0.00017	2.5	1.6
		M_3=5200				

Using DYRECEIVE algorithm, the cost of energy will be reduced and also it avoids the user requests standing in queue.

D) Modules

- **CDN Admission control**
- **CDN Request and Response**
- **Resource Allocation**
- **Workload scheduling**

CDN Admission Control

A content delivery network (CDN) is a system of distributed servers (network) that deliver web pages and other Web content. CDN providers use additional techniques to optimize the delivery of files in optimal data center. Policy file will be generated for user request for dynamic request redirection and enabling good quality of service.

CDN Request and Response

The Video service provider request for the CDN to host their application in the cloud. And then admission controller accept the request for each VSP. The video service provides choose the Virtual instances on various data centers and request the CDN to host theirServices. The video service provider application has the various type of videos such as the high quality, medium quality and the low quality videos. The rent for data center usage will be calculated by CDN and offered to video service provider. This bill generation is done for usage configured by the VSP. As our approach enables dynamic request redirection based on geographic location and type of user request VM usage will be very optimal which results in less cost for the CDN.

Resource Allocation

The video service application deployment is done on various data centers. If the VSP is satisfied with the bill generation process he can proceed with the banking process. The banking gateway is connected when transaction is initialized and OTP will be generated and send to VSP mail ID which he can validate it in upcoming process to complete the transaction. If the transition is successfully made he will get access to various data center and virtual instances. She/he can now deploy his own video service application in the CDN by packaging the contents and sending to various data centers. Finally allocate the resource for Data center. Then the services as started and made available to all user through CDN.

Workload Scheduling

Request scheduling and resource allocation in the cloud can be classified based on different perspectives of cloud providers and cloud users. There are many efforts on designing Scheduling strategies for cloud providers. For single datacenters,improving resourceutilization and fairness are often the focus. For multiple datacenters, some work propose scheduling strategies to minimize the cost of electricity use through balancing load among geographically located datacenters. It systematically handles resource renting from multiple CSPs and schedules user requests to these resources in a nearly optimal manner. In particular, the framework is capable of handling heterogeneous types of user requests, workloads and QoE requirements. Users from different regions obtain various services like video streaming from CDN by the policy the video service provider already generated. Once the VSP receives a request, the request will be dynamically redirected to an optimal datacenter like that High quality, Medium quality, Low quality, based on previous user survey and workloads in geographical location.

E) Evaluation

Workload Admission Control

The revenue based workload admission control method inclines to higher admit higher requests. If there are not enough server to execute higher priority request, the request may experience extremely long response time and brings less are more revenue to CDCs provider. In this case to maximize the total revenue based admission control method can refuse some of higher priority request intelligently admit lower priority once there is more request

Workload Scheduling

The average workload scheduling method equally allocates requests among multiple ISPs and does not consider the bandwidth capacities of ISPs. If the total occupied bandwidth of requests scheduled to a specific ISP exceeds the bandwidth capacity of the ISP, some requests must be refused and cannot traverse this ISP. Besides, requests admitted by each ISP are also equally allocated to multipledistributed CDCs. Therefore, requests scheduled to each CDCmay exceed the capacity of available servers in that CDC. The penalty cost for lower priority requests is lessthan that of higher priority requests. If more admitted requestsare allocated to execute in distributed CDCs, these requests willadd additional bandwidth and energy cost to the total cost ofthe CDCs provider

III. CONCLUSION AND FUTUER WORK

To allocate resources for cloud based video service on user request from multiple regions. Distributed data centers and dynamically computes the near and optimal virtual machine. Using load balancing system, user request standing in queues are avoided. A revenue-

based workload admission control method. VM scheduling indistributed CDCs where multiple heterogeneous VMs concurrently run in a server. To extend our work to consider finer-grained metrics, e.g., memory and storage constraints. Also consider VM scheduling in distributed CDCs where multiple heterogeneous VMs concurrently run in a server.

IV.REFERENCES

[1] Bo hui Li, Haitao Yuan, Jing Bi,Wenjie Jiang,Wei Tan (2016) 'CAWSAC: Cost-Aware Workload Scheduling and Admission Control for Distributed Cloud Data Centers', IEEE Transaction on Automation Science and Engineering.

[2] Jianying Luo, Lei Rao, Xue Liu (2013) 'Data center Energy Cost Minimization: a Spatio-Temporal Scheduling Approach', Proceedings IEEE INFOCOM, pp 340-344.

[3] Joe Wenjie Jiang, TianLan, Sangtae Ha, Minghua Chen, Mungchiang (2012) 'Joint VM Placement and Routing for Data center Traffic Engineering', pp 2876-2880.

[4] Srinivas Narayana, Wenjie Jiang, Jennifer Rexford, Mung Chiang (2013) 'Joint Server Selection and Routing for Geo-replicated Services' IEEE/ACM 6th International Conference on Utility and Cloud Computing, pp 423-428.

[5] Jianying Luo, Lei Rao, Xue Liu (2014) 'Temporal Load Balancing with Service Delay Guarantees for Data center Energy Cost Optimization', IEEE Transactions on Parallel and Distributed Systems, Vol. 25, Issue 3, pp 775-784.

[6] Gemma Reig, Jordi Guitart (2012) 'On the Anticipation of Resource Demands to fulfill the QoS of Web Applications', IEEE 13th International Conference on Grid Computing, pp 147- 155.

[7] Qi Zhang, Mohamed FatenZhani, Raouf Boutaba, Joseph L. Hellerstein (2014) 'Dynamic Heterogeneity-Aware Resource Provisioning in the Cloud', IEEE Transactions on Cloud Computing, Vol. 2, Issue 1, pp 14-28.

[8] Xingguan Zuo, Guoxiang Zhang, Wei Tan (2014) 'Self-Adaptive Learning PSO- Based Deadline Constrained Task Scheduling for Hybrid IaaS Cloud', IEEE Transactions on Automation Science and Engineering, Vol. 11, Issue 2 , pp 564-573.

[9] Manish Bansal, Kiavash Kianfar, Yu Ding, Erick Moreno-Centen (2013) 'Hybridization of Bound-and Decompose and Mixed Integer Feasibility Checking to Measure Redundancy in Structured Linear Systems', IEEE Transactions on Automation Science and Engineering, Vol. 10, Issue 4, pp 1151-1157.

[10] Chi-Yao Hong (2013) 'Achieving High Utilization with Software-Driven WAN'.

[11] Z. Zhu, J. Bi, H. Yuan, Y. chen (2011) 'SLA based Dynamic Virtualized resources Provisioning forShared Cloud data centers', IEEE 4th Int. Conf. Cloud Comput. , pp 630-637.

[12] R. Ghosh, F. Longo, V. K. Naik, K. S. Trivesdi(2013) 'Modeling and Performance analysis of large scale IaaS clouds', Future Generation Comput. Syst., vol.29, no.5. pp. 1216-1234.

[13] Y. Xia, M. Shou, X. Luo, Q. Zhu, J. Li, Y. Huang (2013) 'Stochastic modeling and Quality Evaluation of IaaS clouds', IEEE Trans. Autom.Sci. Eng., Vol. 12, no. 1, pp.1131-1146.

[14] Y. Guo, Y, Fang (2013) 'Electricity Cost saving strategy in data centers by using energy storage', IEEE Trans. Parallel Distrib. Syst., pp. 1149-1160.

[15] S. Agarwal, M. Kodialam, T. Lakshman(2013) 'Traffic Engineering in Software defined Networks', Proc. 32nd IEEE Inst. Conf.Comput.Commun. , pp. 2211-2219.