ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 19, April 2016

PROFIT DRIVEN HYBRID CLOUD MANAGEMENT FOR MULTIMEDIA CLOUD COMPUTING

[1] K.Eesha,[2]C.Divya,[3]R.Deepika,[4]L.Sudha

[1,2,3] Student, S.A Engineering College, Chennai, Tamil Nadu, India.
[4]Assistant professor, S.A Engineering College, Chennai, Tamil Nadu, India.

[1] ishaswap72@gmail.com,[2] divyachandrasekar@outlook.com, [3]deepikarajkumart@gmail.com,
[4] sudhal@saec.ac.in

Abstract: It is known that with a hybrid cloud, a multimedia cloud service provider (MCSP) can quickly extend its services to multiple geographical locations with quality-of-service (QoS)guarantees. Meanwhile, to maximize its profit, the MCSP needs an online management mechanism to operate the hybrid cloud efficiently. In this paper, we study how to maximize an MCSP's profit from provisioning multimedia services to geographically distributed users with a hybrid cloud. We first design a service of provisioning model to manage the resources in the hybrid cloud. Here, in order to make the model practical and address the different situations in private and public clouds, we consider different time granularities for resource reservations. Then, we leverage the Lyapunov optimization technique to maximize the profit of MCSP and propose an online algorithm that can manage the hybrid cloud in the distributed manner. Specifically, the algorithm determines the access control and routing of each multimedia service request, and allocates the resources in the hybrid cloud accordingly. We also apply the –persistent technique to ensure that the worst-case latency of the provisioned requests is bounded. Finally, the proposed algorithm is evaluated with extensive simulations using both synthetical and real traces. Simulation results indicate that the algorithm can manage the hybrid cloud efficiently and maximize the profit of MCSP.

***Index Terms*—Datacenter management, hybrid cloud, Lyapunov optimization, QoS-aware requests scheduling.**

INTRODUCTION

With the rapid development of the Internet, especially the mobile networks, the demands on multimedia services, such as online gaming, social network, videoconference, etc., are growing exponentially. These services usually associate with a sequence of media processing tasks, e.g., caching, transcoding and rendering, which may consume large amounts of computing and storage resources. Then meanwhile, the tasks can be highly dynamic and exhibit huge peak resource requirements. Moreover, since the multimedia services can occupy a lot of computing resources and hence can consume a large amount of energy, how to support them efficiently on energy-constrained devices (e.g., smartphones)will be challenging. Therefore, in order to handle the multimedia services cost-effectively, more and more service providers adopt the cloud infrastructure, as it can respond to demands timely and allocate computing, communication and storage resources adaptively according to the requirements .Due to the fact that the computing tasks are off-loaded to the a cloud, the energy-consumption of the end-user devices can also be reduced significantly. The quality-of-service (QoS) of multimedia cloud computing depends heavily on the underlying network architecture. It is known that deploying datacenters (DCs) close to end users can reduce network latency, improve user experience, and help to promote the services to multiple geographical locations Hence, there is an increasing need to build geographically distributed (geo-distributed) cloud systems with multiple DCs. However, due to the high capital expenditure (CAPEX) and operational expenditure (OPEX) of geo-distributed multi-DC cloud systems, it is not realistic for most of the multimedia cloud service providers (MCSPs) to build the systems on their own. Therefore, the hybrid cloud in Fig. 1 becomes a cost-effective alternative , which uses both private and public DCs. Depending on the scale of its applications, an MCSP can own and operate a few geo-distributed DCs as the private cloud. Meanwhile, if the capacity of the private cloud is not enough, it can rent resources from a public cloud and create virtual machines (VMs) dynamically. With hybrid cloud, an MCSP can quickly extend its services to multiple geographical locations with QoS guarantees. However, in order to manage the hybrid cloud well and

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 19, April 2016

provide multimedia services to geo-distributed end users, one needs to properly address several issues. First of all, the latency that end users in different locations experience should be maintained carefully to ensure consistent QoS. That is to say, the worst-case service latency should be bounded for all the users. Secondly, we need to manage the servers in private cloud and the VMs in public cloud coordinately so that the service-level agreement (SLA)violations where the end users encounter can be minimized and the profit of the MCSP can be maximized. Thirdly, an efficient online algorithm that can achieve profit-driven request scheduling and resource management for the multimedia cloud computing is highly desired. Christo Ananth et al. [7] discussed about a method, In vehicular ad hoc networks (VANETs), because of the nonexistence of end-to-end connections, it is essential that nodes take advantage of connection opportunities to forward messages to make end-to-end messaging possible. Thus, it is crucial to make sure that nodes have incentives to forward messages for others, despite the fact that the routing protocols in VANETs are different from traditional end-to-end routing protocols. In this paper, stimulation of message forwarding in VANETs is concerned. This approach is based on coalitional game theory, particularly, an incentive scheme for VANETs is proposed and with this scheme, following the routing protocol is in the best interest of each node. In addition, a lightweight approach is proposed for taking the limited storage space of each node into consideration.

The theoretical analysis on offline optimization is then provided based on the assumption on the distribution of the requests' arrivals. Next, we apply the Lyapunov optimization technique on the model to maximize the profit of MCSP and propose an online algorithm that can manage the hybrid cloud in the distributed manner. Specifically, the algorithm determines the access control and routing of each multimedia service request, and allocates the resources in the hybrid cloud accordingly. Meanwhile, we leverage the –persistent technique to ensure that the worst-case latency of the provisioned requests is bounded. Finally, we evaluate our proposal with extensive simulations using both synthetical and realtraces.

I. PROBLEM STATEMENT

Description of the service provisioning model for resource management in hybrid cloud and formulate the problem of profit-driven request scheduling and resource management for multimedia cloud resource. These services usually associate with a sequence of media processing tasks ,e.g., caching ,transcoding and rendering, which may consume large amounts of computing and storage resources. Meanwhile, the tasks can be highly dynamic and exhibit huge peak resource requirements. Therefore, the hybrid cloud in Fig. 1 becomes a cost-effective alternative, which uses both private and public DCs. Depending on the scale of its applications, an MCSP can own and operate a few geo-distributed DCs as the private cloud. Meanwhile, if the capacity of the private cloud is not enough, it can rent resources from a public cloud and create virtual machines (VMs) dynamically.

II. SYSTEM ANALYSIS

A. EXISTING SYSTEM

The existing system consumes large amounts of computing and storage resources. Meanwhile, the tasks can be highly dynamic and exhibit huge peak resource requirements. Moreover, since the multimedia services can occupy a lot of computing resources and hence can consume a large amount of energy, how to support them efficiently on energy-constrained devices will be challenging. The quality-of-service (QoS) of multimedia cloud computing depends heavily on the underlying network architecture. It is not realistic for most of the multimedia cloud service providers (MCSPs) to build the systems on their own. Due to the fact that the computing tasks are off-loaded to the cloud, the energy-consumption of the end-user devices can also be reduced significantly. The quality-of-service (QoS) of multimedia cloud computing depends heavily on the underlying network architecture. It is known that deploying datacenters (DCs) close to end users can reduce network latency, improve user experience, and help to promote the services to multiple geographical locations. Therefore, the hybrid cloud becomes a cost-effective a iternative, which uses both private and public DCs(data center).There is a problem of request's queuing delay by adjusting the queue length's. The existing system might not be flexible and adaptive.

B. PROPOSED SYSTEM

The investigation of how to maximize an MCSP's profit from provisioning multimedia services to geo-distributed users with hybrid cloud. First design a service provisioning model to manage the resources in hybrid cloud. Here, in order to make the model practical and

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 19, April 2016

address the different situations in private and public clouds, consider different time granularities for resource reservations. The theoretical analysis on offline optimization is then provided based on the assumption on the distribution of the requests' arrivals. Next, apply the Lyapunov optimization technique on the model to maximize the profit of MCSP and propose an online algorithm that can manage the hybrid cloud in the distributed manner. Specifically, the algorithm determines the access control and routing of each multimedia service request, and allocates the resources in the hybrid cloud accordingly. Meanwhile, leverage the persistent technique to ensure that the worst-case latency of the provisioned requests is bounded. Finally, evaluate our proposal with extensive simulations using both synthetical and realtraces. An online algorithm for hybrid cloud management, which can maximize the MCSP's time-average profit and ensure that the worst case latency of all the provisioned requests is bounded. It operates in the distributed manner. There is no problem of requests queuing delta or request arrival pattern in this system.

.

C. SYSTEM ARCHITECTURE

we describe the service provisioning model for resource management in hybrid cloud and formulate the problem of profit-driven request scheduling and resource management for multimedia cloud computing.

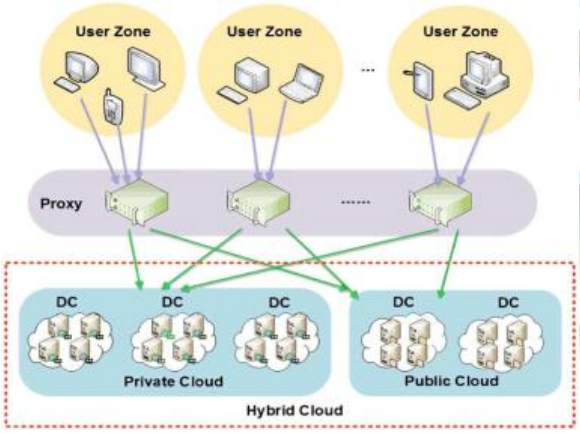

Fig. 1. Multimedia services in hybrid cloud environment.

III. SYSTEM IMPLEMENTATION

A. Multimedia Services in Hybrid Cloud

We consider a hybrid cloud as illustrated in Fig. 1, in which multimedia service requests are processed in the DCs and the proxies handle the access control and routing of the requests. In this work, we assume that the hybrid cloud operates as a discrete-time system and the service provisioning scheme in it can be changed every time interval Δt. Hence, we can normalize system time with the time slot (TS) Δt and obtain the normalized system time as t={1,2,3..}. Note that the actual choice of Δt depends on the control and management mechanism in the hybrid cloud. Basically, Δt should be at least longer than the maximum round-trip time (RTT) among the proxies and the DCs to ensure effective information exchanges, and the study in suggested that Δt should be on the magnitude of seconds or tens of seconds. We denote the DCs in the hybrid cloud with set $\mathcal{D}$ and assign an index j to each of them. Without loss of generality, we assume that the first $|\mathcal{M}|$ DCs in (i.e., the indices satisfy $j \in [1.|\mathcal{M}|]$) are private DCs (denoted with set $\mathcal{M}$), while the rest $|\mathcal{N}|$DCs in $\mathcal{D}$ are public ones (denoted with set $\mathcal{N}$). Hence, we have $\mathcal{D}=\mathcal{M} \cup \mathcal{N}$. In a private DC j, the MCSP installs S_j servers for multimedia cloud computing. On the other hand, the rented resources in a public DC j can support at most S_j VMs. Apparently, the servers and

VMs should carry homogenous resources for serving the multimedia service requests. Then, the resources in each server or VM in a DC j can be quantified as, e_j which means that each server or VM in the DC can process e_j requests within one TS. The MCSP provides multimedia services to end users located in a few geo-distributed zones (denoted with set $\mathcal{Z}$). We assign an index i to each zone. There is a service proxy in each zone, and the requests from end users are first sent to the proxies in their zones and then routed to the hybrid cloud for further processing. We use $A_i(t)$ to represent the number of requests that the proxy in zone i receives within TS t . We define A_{max} as the maximum arrival rate of the requests, i.e.

$$A_i(t) \leq A_{max}, \quad \text{for all i,t.}$$

For the system, we define Amax as the time-average expectation number of the incoming requests in zone i, i.e.

$$A_i = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} A_i(t).$$

There is an SLA on the service provisioning of the requests, which means that each request should be processed within duration, otherwise, it will be dropped due to SLA violation.

B. Requests Scheduling in Service Proxy

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 19, April 2016

Since the incoming requests in zone i will be aggregated at the proxy before being routed to one of the DCs for further processing, we introduce a decision variable $a_i(t)$ to denote the number of requests that the proxy in zone I admits in TS t.. Apparently, we have

$$0 \le a_i(t) \le A_i(t), \text{ for all } i \in \mathcal{Z} \quad (1)$$

and define the time-average expectation number of admitted requests in zone i by a_i. To avoid unnecessary latency, each proxy routes all the admitted requests to their corresponding DCs in the current TS. We also introduce another set of decision variables $\{a_{i,j}(t)\}$, each of which denotes the number of requests that DC j receives from zone i in TS t.. Then, we have

$$a_i(t) = \sum_{j \in \mathcal{D}} a_{i,j}(t), \quad \text{for all } i \in \mathcal{Z} \quad (2)$$

And we define $a_{i,j}$ as the time-average expectation number of requests that DC j receives from zone i , then we have $a_i = \Sigma_{j \in \mathcal{D}} a_{i,j}$, for all i.

C. Resource Management in Hybrid Cloud

It is known that when the workload is relatively low, we can switch the servers in the private DCs from working to idling and achieve reduction on OPEX (e.g., the energy cost).

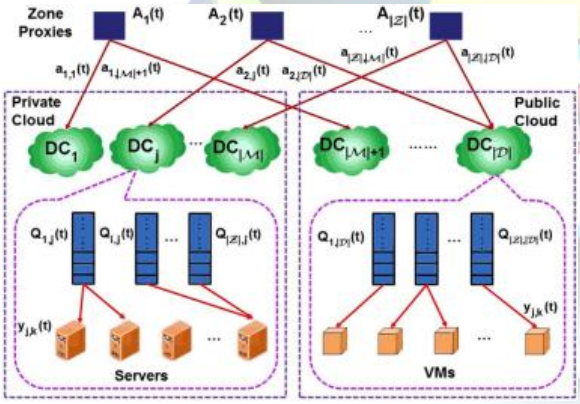

Fig. 2.Service provisioning model for multimedia services in hybrid cloud.

Since the TS in the system is usually much longer than the duration of the switching transition, we can ignore the transition over heading following analysis. Similarly, the VMs in the public DCs can also be released dynamically for cost saving. Then, the state of the *k*-th server/VM in DC j in TS t can be represented with a decision variable $y_{j,k}(t)$, such that

1, server/VM is in working state
0,server/VM is in idle/released state.

Meanwhile, we notice that the VMs provided by public DCs are usually charged based on a relatively long time granularity, e.g., on the magnitude of hours or even longer. Therefore, we define the time granularity for VM rentals as ,T_g which is also normalized with the duration of TS (i.e., Δt). Then, once a VM is rented, the MCSP needs to pay for a whole T_g at least. Hence, the MCSP should try to make full use of the VM within T_g, otherwise, its down payment would be wasted. This means that in the hybrid cloud, once a VM is rented, it will stay in the working state for at least T_g,

$$T_g \cdot y_{j,k}(t), \quad \text{for all } j \in N. \quad (4)$$

As shown in Fig. 2, we make the requests from different zones be buffered in independent queues in each DC for further processing .$Q_{i,j}(t)$ denotes the number of buffered requests from zone in TS in DC j.
The total number of working servers/VMs in DC j in TS t for request-processing is

$$X_j(t) = \sum_{i \in \mathcal{Z}} x_{i,j}(t). \quad (5)$$

.
To minimize SLA violations, we may purposely drop certain buffered requests in the DCs to make sure that the rest ones can be processed timely, i.e., within the worst-case latency *d*.

D. Profit-Driven Service Model

The MCSP's profit is the margin between revenue and cost. The revenue comes from serving the multimedia service requests, and hence it can be formulated as a function $f(\cdot)$ of the number of served requests. Here, we need to point out that the optimization techniques discussed in the following sections are applicable as long as $f(\cdot)$ is a concave and continuous function. Hence, for simplification, we use the formulation in and assume that $f(\cdot)$is a linear one as

$$R = \sum_{i \in \mathcal{Z}} f(a_i - b_i) = \sum \Pi_i \cdot (a_i - b_i)$$

where Π_i is the revenue per served request for zone i.
The cost includes three parts, i.e., the energy cost of private DCs, the payment for renting the resources from public DCs, and the communication cost for data

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 19, April 2016

transfers. The communication cost is from transmitting the data of multimedia services from the end users to the hybrid cloud, which is calculated as

$$C_t(t) = \sum_{i \in Z} \sum_{j \in Z} \gamma_{i,j} \cdot e_j \cdot x_{i,j}.$$

where $\gamma_{i,j}$ is the average communication cost per request for the situation in which the requests are from zone and get processed in DC j.
To this end, we can see that the total time-average cost is the summation of the three parts,

$$C = C_m + C_n + C_t$$

and in order to maximize the time-average profit of the MCSP, we need to solve the following optimization problem:

Maximize P= R – C.

Meanwhile, we should make sure that the lengths of all thequeues would not increase towards infinite.

IV. CONCLUSION

In the proposed scheme it was investigated how to maximize an MCSP's profit from provisioning multimedia services to geo-distributed users with hybrid cloud. First design a service provisioning model to manage the resources in hybrid cloud. Then, leverage the Lyapunov optimization technique to maximize the profit of MCSP and proposed an online algorithm that can manage the hybrid cloud in the distributed manner. Then application of the ε-persistent technique to ensure that the worst-case latency of the provisioned multimedia requests would be bounded. Finally, evaluation of the proposal with extensive simulations using both synthetical and real traces. Simulation results indicated that the proposed algorithm can manage the hybrid cloud efficiently and maximize the profit of MCSP.

V. FUTURE ENHANCEMENT

The algorithm can be operated in a truly online manner and can obtain request scheduling and resource management with knowledge on the request arrivals. The T-slot lookahead algorithm which is a quasi-online scheme can be made to know future request arrivals in prior or in advance.

VI. REFERENCES

[1]H. Li, L. Zhong, J. Liu, B. Li, and K. Xu, "Cost-effective partial migration of VoD services to content clouds," in Proc. CLOUD, Jul. 2011,pp. 203–210.

[2] R. Duan, R. Prodan, and X. Li, "Multi-objective game theoretic scheduling of bag-of-tasks workflows on hybrid clouds," IEEE Trans. Cloud Comput., vol. 2, no. 1, pp. 29–42, Jan. 2014.

[3] Y. Wen, X. Zhu, J. Rodrigues, and C. Chen, "Cloud mobile media:Reflections and outlook," IEEE Trans. Multimedia, vol. 16, no. 4, pp.885–902, Jun. 2014.

[4]W. Zhu, C. Luo, J. Wang, and S. Li, "Multimedia cloud computing,"IEEE Signal.Process. Mag., vol. 28, pp. 59–69, May 2011.

[5] I. Bisio, R. Pan, F. Lavagetto, M. Marchese, A. Sciarrone, C. Fra, and M. Valla, "Smartphone-based automatic place recognition with Wi-Fi signals for location-aware services," in Proc. ICC, Jun. 2012,
pp. 4943–4948.

[6] I. Bisio, F. Lavagetto, M. Marchese, and A. Sciarrone, "Comparison of situation awareness algorithms for remote health monitoring with
smartphones," in Proc. GLOBECOM, Dec. 2014, pp. 2454–2459.

[7] Christo Ananth, Kavya.S., Karthika.K., Lakshmi Priya.G., Mary Varsha Peter, Priya.M., "CGT Method of Message forwarding", International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE), Volume 1, Issue 1, August 2015,pp:10-15

[8] M. Neely, Stochastic Network Optimization With Application to Communication and Queueing Systems. San Rafael, CA, USA: Morgan
and Claypool, 2010.

[9] X. Nan, Y. He, and L. Guan, "Optimal resource allocation for multimedia cloud in priority service scheme," in Proc. ISCAS, May 2012,pp. 1111–1114.

[10] X. Nan, Y. He, and L. Guan, "Optimization of workload scheduling for multimedia cloud computing," in Proc. ISCAS, May 2013, pp.
2872–2875.

[11] X. Nan, Y. He, and L. Guan, "Towards optimal resource allocation for differentiated multimedia services in cloud computing environment,"
inProc. ICASSP, May 2014, pp. 684–688.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 19, April 2016

[12] Y. Wu, C. Wu, B. Li, X. Qiu, and F. Lau, "CloudMedia: When cloud on demand meets video on demand," in Proc. ICDCS, Jun. 2011, pp.
268–277.

[13] W. Gong, Z. Chen, J. Yan, and Q. Shuai, "An optimal VM resource allocation for near-client-datacenter for multimedia cloud," in Proc.
ICUFN, Jul. 2014, pp. 249–254.

[14] R. Van den Bossche, K. Vanmechelen, and J. Broeckhove, "Cost-optimal scheduling in hybrid IaaS clouds for deadline constrained workloads,"inProc. CLOUD, Jul. 2010, pp. 228–235.

[15] M. Neely, "Opportunistic scheduling with worst case delay guarantee sin single and multi-hop networks," in Proc. INFOCOM, Apr. 2011, pp.
1728–1736.

[16] X. Qiu, H. Li, C. Wu, Z. Li, and F. Lau, "Cost-minimizing dynamic migration of content distribution services into hybrid clouds," IEEETrans. Parallel Distrib. Syst., to be published.

[17] H. Xu and B. Li, "Joint request mapping and response routing for geo-distributed cloud services," in Proc. INFOCOM, Apr. 2013, pp.854–862.

[18] S. Ren, Y. He, and F. Xu, "Provably-efficient job scheduling for energy and fairness in geographically distributed data centers," in Proc.
ICDCS, Jun. 2012, pp. 22–31