TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 0?/?? pp???-???
Volume 18, Number 1, Feburary 2013

QoS-aware Virtual Machine Scheduling for Video Streaming Services in Multi-Cloud

Wei Chen and Junwei Cao*

Abstract: Video streaming services are trending to be deployed on cloud. Cloud computing offers better stability and lower price than traditional IT facilities. Huge storage capacity is essential for video streaming service. More and more cloud providers appear so there are increasing cloud platforms to choose. A better choice is to use more than one data center, which is called multi-cloud. In this paper a closed-loop approach is proposed for optimizing QoS and cost. Modules of monitoring and controlling data centers are required as well as the application feedback such as video streaming services. An algorithm is proposed to help choose cloud providers and data centers in a multi-cloud environment as a video service manager. Performance evaluation of the algorithm is included with different video service workload. Compared with using only one cloud provider, dynamically deploying services in multi-cloud is better in aspects of both cost and QoS. If cloud service costs are different among data centers, the algorithm will help to make choices to lower the cost and keep a high QoS.

Key words: cloud computing; dynamic scheduling; data centers; video streaming; service computing; performance evaluation; QoS

1 Introduction

Cloud computing is changing more and more services on Internet[1,2]. In the area of IaaS, Amazon is the most popular cloud provider, but more and more providers are coming into this area. The numbers of cloud providers will increase explosively in future. Netflix is a video streaming service provider and based on Amazon EC2. It has been proved that a video service based on cloud computing is feasible. But with more cloud providers, how to choose from the providers is becoming increasingly important. Different cloud providers may charge a different price and support different service item. One cloud provider may have several data centers to choose. The position of data center is also important for IO type service like streaming video. The quality of service (QoS) will decrease if the data center is far from the end users. In such a multi-cloud environment, applications based on cloud should make choices of how to use these resources. Security in cloud computing is also very important. Lots of works[3,4] have been done to resolve this problem. In multi-cloud, security problem is more important and difficult. With such standard security management, cooperation in multi cloud providers are realizable.

For a video service system based on cloud, the cost of renting storage and virtual machines (VM) are the main part of the total cost. The cost is dynamically changing with the need of applications. Less VMs than needed will result in a high resource occupancy rate. More VMs than needed will cause a waste of cost. The standard of the needed number is based on QoS. An appropriate resource occupancy rate of VM can reduce the packet loss or decoding delay in the video

• W. Chen and J. Cao are with the Research Institute of Information Technology, Tsinghua National Laboratory for Information Science and Technology, Tsinghua University, Beijing 10084, P. R. China. E-mail: chenw06@mails.tsinghua.edu.cn, jcao@tsinghua.edu.cn.

$*$ To whom correspondence should be addressed.

Manuscript received: year-month-day; revised: year-month-day; accepted: year-month-day

streaming service, which will help to improve the QoS. In the simulation, the QoS of service is calculated by the distance between the user and the server as well as the resource occupancy rate of the server. There are several important work on resource scheduling in multi-cloud computing environment. And the QoS guaranteeing is researched for lots of times. In this paper, authors referred their work and made distributions on resource scheduling of VM on multi-cloud.

2 Related Work

Cloud computing is rapidly developing and becoming more and more attractive. Low cost, high efficiency and scalability are very significant in the environment of big data, which is becoming a trend these years. Amazon, Google and similar technique companies are heavily pushing the develop of cloud computing. For SMEs(small and medium-sized enterprises), cloud computing is the first choice of decreasing the cost of IT.[5- 7]

In cloud computing, economics are becoming critical important for both the cloud providers and users. In [8], authors researched the problems of optimal multiserver configuration to maximizing the profit. A lot of factors are taken into account such as the amount of a service, the workload, the configuration of the multiserver system, the service level, the QoS, the cost of renting, the cost of energy consumption, and the profit of a service provider. By modeling this problem as an optimization problem, authors solved the problem and made a simulation on it.

Multi-cloud[9,10], which means building a hybrid platform for one vertical applications by more than one cloud services. These cloud services may be provided by different providers and the data centers are usually built in different locations. By using the location based feature of some application, the system can support the application nearby. The cost of the whole network will decrease and the quality of service will be improved. As related work, there are several works focused on multi-cloud and QoS in cloud computing, and they are introduced below.

In [10], authors research how to configure the virtual machine of users dynamically when there are several cloud platform and there are different prices to lower the cost of users. It brings a prediction model of the price of cloud services. Using the predicted prices, the system schedule the virtual machine to archive a lower total cost. In result, users can save up to 5% per day. This paper is a useful attempt in multi-cloud environment, which will be more and more popular in the next years. In our work, we will also take the prices of each cloud service into account. But the goal is to reach a better QoS (quality of service) and price at the same time.

An important work in [11] is trying to summarized a new optimization approach in clouds. In clouds, QoS guaranteeing is a significant work. In this paper, authors built a performance model to invest the cloud. A closed loop is set to control the QoS of cloud. While the cloud is serving, a sensor is used to observe the status of cloud. The observation result is compared with QoS goal. An optimization method is used to analyze and plan the next behaviour. Then the plan is executed by the effector to control a allocation of resources.

This work is enlightening and important. In an video streaming system, the QoS guaranteeing is very important. A closed-loop can ensure the QoS in an acceptable scope. The optimization model needs to make a correct instruction of increasing or decreasing resources. Compared to this work, the model described in our work extend the background to multi-cloud environment. QoS guaranteeing is also one of the indicator in our system. Cost control is the other one.

Video streaming technique has been developed for several years and can resolve lots of problems for the online video demand. But on a large scale situation, more targeted development and optimization are required. In [12], authors introduced key issues on video streaming. Application-layer QoS are specially discussed because it is very important in video application. CDN (content delivery network) is also a very important way to lifting the quality of video service. It is a buffer-like service which can support content delivery need. By CDN only, lots of problems are not solved very well, so some related technique based on CDN are develop.

In[13,14], authors studied the QoS for voice and video streaming on Internet. The QoS is affected by the transition delay and packet lost rate. Authors estimate the ”goodness” of a video transition from the perspective of the video stream, instead of the traditional way of relying on raw network performance detections. The estimates are used to make decisions of which path should be chosen.

In [15], authors researched the method of support video stream and decrease the cost for the video-on-demand application. In this research, authors used a novel queue network model to describe the users viewing behaviors. So they derive the equilibrium demand of upload bandwidth to satisfy the demand of smooth playback. Then, they take practical cloud parameters into account. Two optimization problems related to VM provisioning and storage rental are formulated and some efficient solutions are proposed. In cloud computing, users need to optimize the time and numbers of VM and storage to lower the cost. Thirdly, the designed a practical dynamic cloud provisioning algorithm and then implemented them. A video-on-demand provider can easily configure the cloud services to meet its demands based on their solutions. To test the performance of their algorithm, an evaluation based on real system implementations are token. Practical user dynamics observed in real-world video-on-demand system. The results confirmed the adaptability and effectiveness of their system in varying demands and guaranteeing smooth playback at any time.

Their work offers a good train of thought and practical help. Our work referred the analyzing model and method. But their work are based on normal one cloud environment, in which only one cloud provider is providing cloud service and no other choices are offered. So the price and location are not in consideration. In recent years, more and more cloud providers are starting their cloud services. Each developer will have several choice of using which provider and which data centers. Also they can choose two or more of them at the same time to support their need. In the algorithm designing chapter of this paper, the background is set to the environment in which several cloud providers can be chosen and the price and location of data centers are the most aspect for the choice.

For the type of video service, most providers are now using their own devices to support their business instead of using a cloud service. But along with the development of cloud computing, more and more service will be transferred to cloud platform. Just like only very small number of company will produce electricity when they need it. Netflix is a good example of providing video services via cloud platform. They used the Amazon Web Service to start their business. When the count of users are increasing rapidly, the resources can be ready for them very soon and when the users are decreasing, the cost can decrease at the same time.

3 Algorithm Design

3.1 Background

In this paper, we mainly concern the situation of multi cloud. There are several data center in several places and in each data center, we can use an elastic computing resources. In authors opinion, cloud computing is the trend of the network. More and more small and medium enterprises will choose cloud computing to build their network services instead of buying lots of facilities and employing lots of IT staff to managing them. But with single cloud provider, the network reliability and the price will be a potential risk. A mature large-scale service cannot build their service on one cloud provider. The main point of this paper, is how to improve the quality of service and lower the cost in the multi cloud environment.

In cloud computing, VM (virtual machine) is the unit of service provided for the users. When the service need more computing ability, users can ask for more VM. In one data center, the network bandwidth is wide enough so the data transaction between VMs are very fast and cheap. For the video service, the system will store a copy of video data in each data center and all the VMs in this data center will share this copy to provide service.

The internet out of data center is more complex. When the user is far from the server in data center, the quality of video service will decrease, because the delay time and packet loss rate will increase. At the same time, because of the retransmission and artificial refresh operation, the press on the service will also increase. So if there are lots of users around somewhere, a new data center nearby will help resolve the problem. But transferring the data to the new data center and renting storage space will cost a lot. How to make the decision is one of the target in this paper.

To be simple, we put the locations of users and servers on a 2-dimension map. Normally, in a big city, the population density is high and in other locations, the population density is low. We simulate a users distribution map and designed several cities on the map. The population distribution is generated randomly and we make the test based on this map. How to generate the map is not the content of this paper and the algorithm do not rely on the layout of the map, we can say it is enough to use the simulated map.

In the map, lots of available data centers are located somewhere. Parts of them are in big cities, which means they are near to a large numbers of users. Parts of them are far from big cities, which means the cost of the data center will be low so the price should be low. The service system of our video stream has the condition of use resource of all these data centers. If we choose to start a service from one data center, the system need to use the storage service on this data center first. Then the system can decide the number of VM in this data center.

The cost of the system is one of our target in this algorithm. The cost are composed by the VM cost and storage cost. When one cluster in a data center is begin to be used, the storage cost will occur. The VM cost is in direct proportion to the number of used VM.

Assume that the distant between one user and the server for him is x, the consumption of the server resources are f(x). Normally, f(x) is an increasing function. But the increasing amplitude is not big. In one data center, the system can ask for any number of VM. When the resource of the existing VM are nearly used out, for example bigger than 80%, the QoS (Quality of Service) will decrease. The system will ask for one more VM to serve the new customers when the existing resources are not enough. The QoS is relating to the distance (x) too. If the x is bigger than one threshold, the QoS of this user will decrease. The QoS is between 0 and 1.

Based on all these assumption, authors designed a load balancing algorithm for multi cloud model. This algorithm will lower the total cost under the condition of keeping the average QoS. When the number of users changed, the algorithm will make correct decision to archive this goal.

3.2 Model Description

Figure 1 shows the architecture of this system. In every available data center, there is a monitor and an executor. The monitor can collect the resource occupancy rate of all the VM in this data center. There is a local load balancing mechanism. When some VMs have high occupancy rates and others have low rates, the monitor will send message to the executor to redirect the connections of some users. So the occupancy rate can keep relatively balanced. The global optimizer only need to collect the average resource occupancy of VMs. The user's location is collected by the optimizer and the location is the most important reference to decide which data center should be used. When the optimizer find that VM need to be more or less, it will send an instruction to the executor.

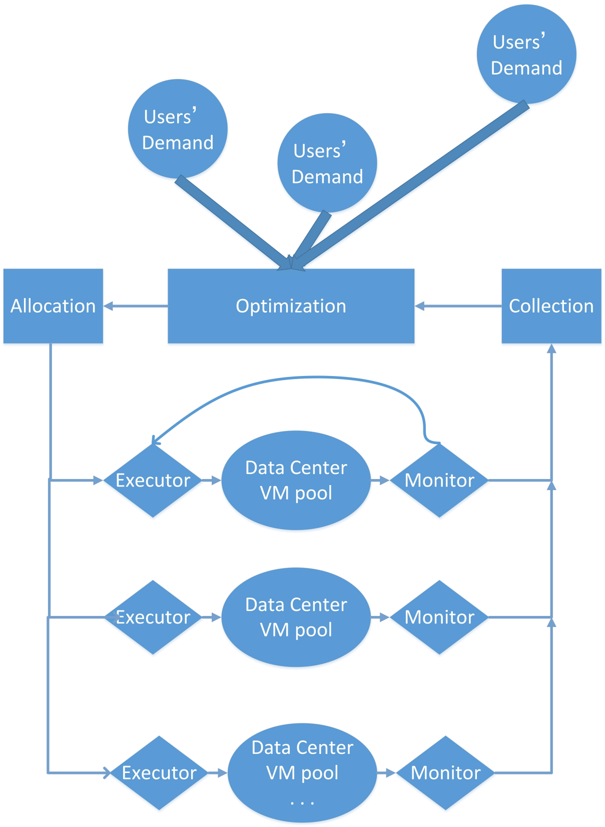


Fig. 1 Dynamic Scheduling Model for Video Service deployment in Multi-cloud

By this loop, the optimizer will has full control of this system. Cooperated with the algorithm in next section, the QoS can keep an acceptable value and the cost will be as low as possible.

3.3 Algorithm Implementation

The variable of this question is :

1. do or do not start a service in which data center
2. how many VMs should be used in each data center.
3. which VM will be distributed for each user.

By make decision of these problems in the algorithm, the system need to reach two goal:

1. The total cost be low.
2. The average QoS be high.

Normally, the QoS will be floating from 0 to 1. And the cost is bigger than 0. Decide two goals at the same is complex and hard to control. So we make an evaluation indicator according to the real demand. When then QoS is very low, the system will be not usable so the indicator will be punished. When the QoS is bigger than one threshold, the promotion of QoS makes little sense. So the indicator will consider the cost more and decrease the weight of QoS. For the others, these two will be considered at the same time.

So the indicator is :

$$f(cost, QoS) = \begin{cases} \frac{0.7*cost}{QoS^2}, & QoS \in (0, 0.7) \\ \frac{cost}{QoS}, & QoS \in [0.7, 0.9] \\ \frac{\sqrt{0.9*cost}}{\sqrt{QoS}}, & QoS \in (0.9, 1] \end{cases} \tag{1}$$

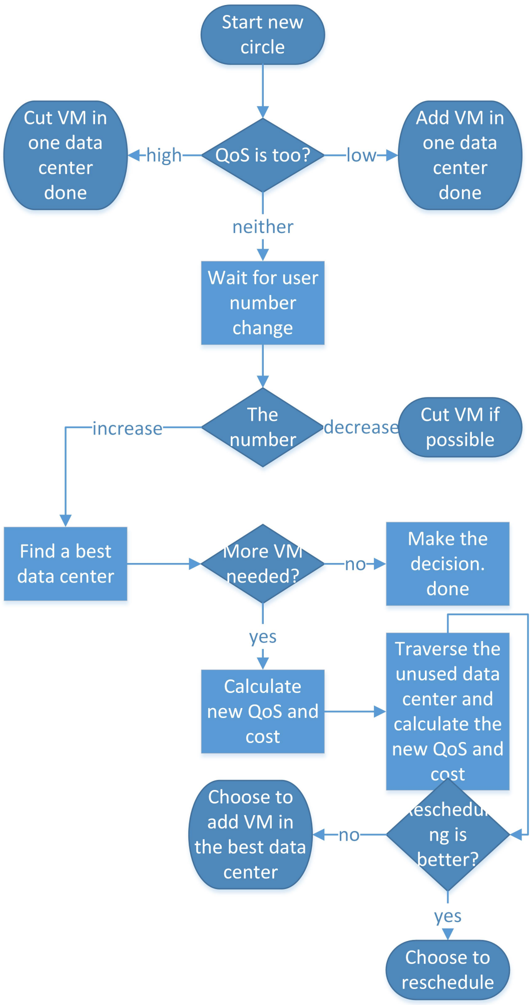

Start new circle
Cut VM in one data center done
high
QoS is too?
low
Add VM in one data center done
neither
Wait for user number change
increase
The number
decrease
Cut VM if possible
Find a best data center
More VM needed?
no
Make the decision. done
yes
Calculate new QoS and cost
Traverse the unused data center and calculate the new QoS and cost
Choose to add VM in the best data center
no
Reschedung is better?
yes
Choose to reschedule

Fig. 2 process of the algorithm

This algorithm needs to keep optimizing and making decision when the number of users changes. Fig. 2 is the process flow diagram. The algorithm can be described as below:

1. In the current status, check if the QoS is too low or high enough.
2. If the QoS is too low, add VM in the data center which has the highest resource occupancy rate. If the QoS is too high, try to decrease VM number in the data center which has the lowest resource occupancy rate.
3. Wait for the number of user changes. When the number of user decrease, check the data center which own the server of the leaving user. Decrease the VM number if possible.
4. When a new user comes, find a best data center for this user. If the existing VMs have enough resources to server this user, lead the user here, make this choice and end this circle. The best data center is chosen by considering both the price and the distance.
5. If the existing VM do not have enough resources, try to calculating the new average QoS and total cost after adding a VM in this data center.
6. Searching for the available data centers which has not been used, try to start using one. Once one data center is started, the existing users will be reevaluated which data center is the best. After the rescheduling, calculating the new average QoS and total cost. Choose a best candidate and compare with the result of last step.
7. Choose a better one and make this choice. End this circle.

4 Performance Evaluation

4.1 Experiment Data Set

From the design of the algorithm, we can find that the system will serve the nearby users better.

First, we can check the simulated map. We designed 9 cities and most users appears around the cities. The map size is 10000*10000 and each red point stands for one user appears in that location. The locations are randomly generated but it do not impact the result of the algorithm. In the real system, the manager can input the real data of users location to the algorithm.

When the number of users is 100, 1000 and 10000, the population distribution are show in Fig. 3. In the simulation environment, we set 18 cloud platform in the map. 9 of them are located in the 9 cities, the other 9 locate randomly on the map. The location shows in the picture.

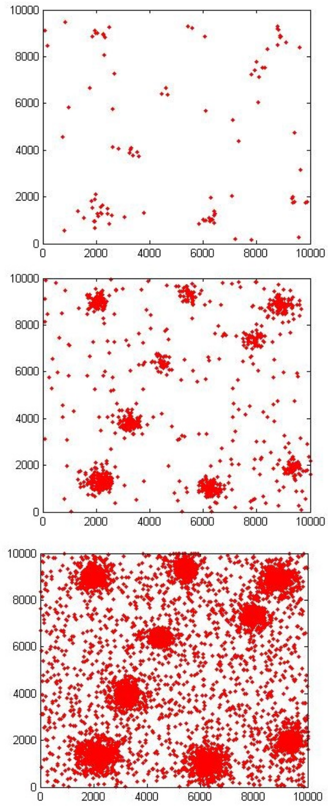


Fig. 3 Users' distribution on generated map of users' number 100,1000,10000

At the beginning, there are no users in our system, the system randomly choose a data center in the map and begin to service. The number of users increase from 0 to 10000, we monitor the average QoS and the total cost of the system.

4.2 Experimental Results without Price Difference

Firstly, we set the prices of all the data centers are the same. So the system will surely lead the user to the nearest opened data center. Besides, we set the storage cost of each data center is 10 times of the cost of one VM in the same data center. In this situation, the QoS changing curve along with the users' increasing is shown in Fig. 4 The inflection points appears several times in the figure. It is there because the rescheduling of all the links after a new data center begin to be used.

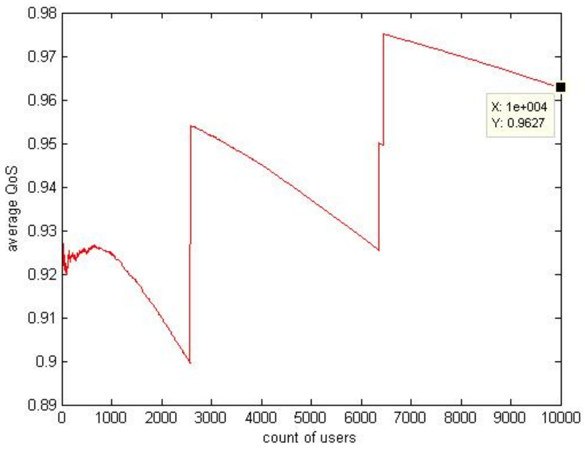


Fig. 4 QoS changes along with the increasing of users

Each time the rescheduling occurred, the QoS will be improved. It is reasonable because when one more data center is used, some nearby users will be able to connect to this data center and the QoS of these users will be improved. After the rescheduling, the QoS is decreasing slowly along with the increasing of the users number. It is because the more user connect, the heavier the loads of the data centers are.

The lowest point of QoS appears on the first inflection point and the value is 0.9. Since then, the QoS is always be bigger than 0.9. When the number of users reach 10000, the QoS is about 0.963.

The total cost will surely increase along with the number of users. So we inspect the average cost of each user. It means the totalcost/usernumber. When the user count is very small, the storage cost will appear too large. So we only show the changing curve of averagecost since the number of users is bigger than 20. The curve is shown as Fig. 5 The basic trend of the cost-per-user is to decrease, because the storage cost per user is lower. When the number reaches 10000, the cost-per-user is 0.69 which we will use to compare with later. The curve has lots of small wave, which

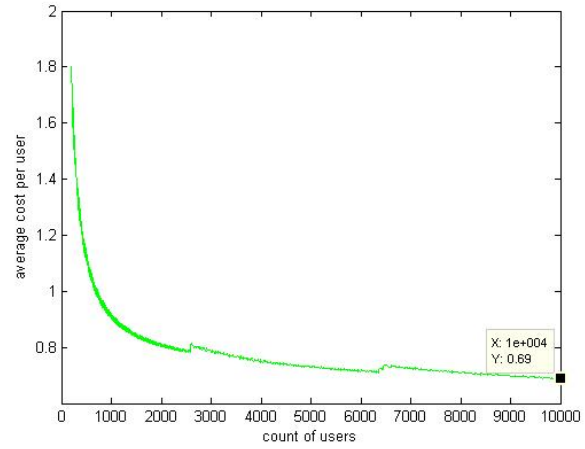


Fig. 5 Cost-per-user decrease along with the increasing of users

is caused by the cost increase of new VM and new data center. According to the simulation results, 4 data centers are used. Then number of VM in each data center is 73,66,68,98. The 4 data centers on map is shown as Fig. 6 As a comparison, we simulate the algorithm of using only one of the data centers and only add VM when the resource occupancy rate is larger than the threshold. The curve of the QoS is shown as Fig. 7 It is reasonable that the QoS is continuously decreasing because many users are too far from this data center and have no other choice but bearing the bad QoS.

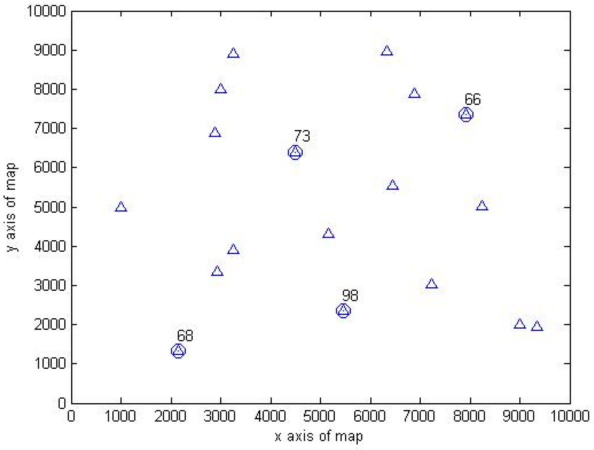


Fig. 6 The location of data centers which are used

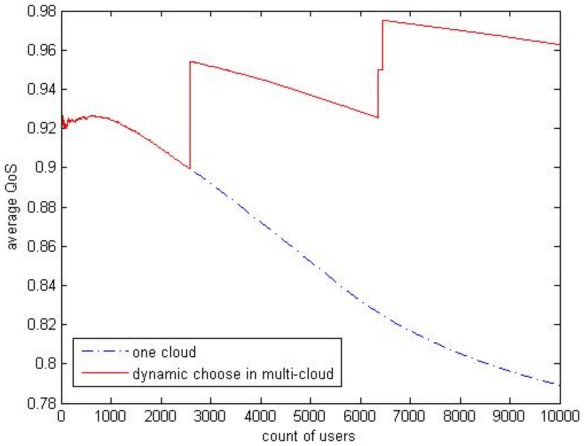


Fig. 7 QoS decrease along with the increasing of users

In what we are interested is, how about the performance of cost-per-user? The curse is shown as Fig. 8 From Fig. 8, we can find that, the cost-per-user is also continuously decreasing along with the user number. But when the number reaches 10000, the cost-per-user is 0.7241, which is bigger than that of our algorithm. Although the cost of storage in a new data center is large, the cost-per-user is lower by our algorithm. And the QoS is much better than only using one data center.

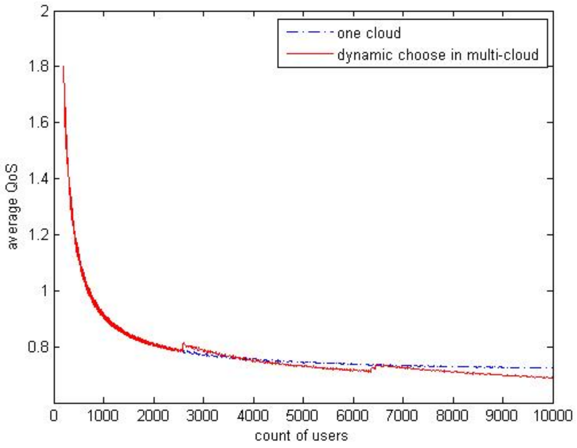


Fig. 8 Cost per user is decreasing but higher than last contrast

4.3 Experimental Results with Price Difference

In real system, the price of each data center will not be the same. Some cloud providers charge a higher price than others, and some data center is built in places where the electricity is cheaper. So the price of each data center is different. So now we bring a price coefficient for each data center. the coefficient changes from 0.85 to 2. The costs of storage and VM are all need to times by the coefficient. In our settings, the price coefficient in big cities is normally larger than that of others, but there are exceptions. The coefficients of data centers in cities change from 1 to 2, and those of the others change from 0.85 to 1.8. Based on it, the price coefficients are randomly generated. When price difference is brought into account, the decision of which data center should one user connect to is not an obvious thing. In this algorithm, the system will choose a cheaper one if the distances do not differ too big to heavily influence the QoS. In this situation, the changing curve of QoS is shown as Fig. 9

The QoS changing curve do not differ too much with that of Fig. 4. The lowest QoS is bigger than that of no price difference. But it seems not a certain thing and changes along with the price coefficients matrix. The cost-per-user changing curve is shown in Fig. 10. When the number of user reaches 10000, the cost-per-user is 0.6181. It is lower than that of no price difference. Although most data centers has a price coefficient larger

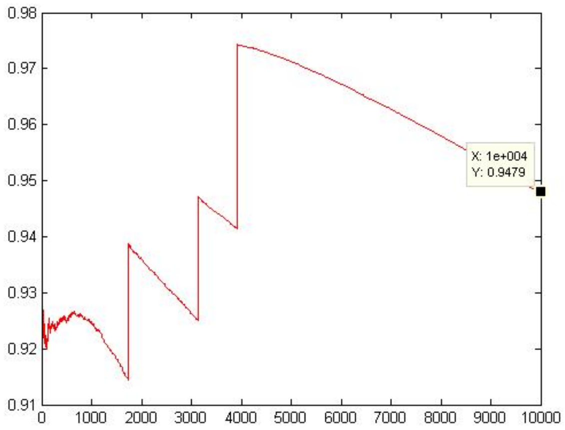


Fig. 9 QoS changes along with the increasing of users. The price coefficient do not make great changes.

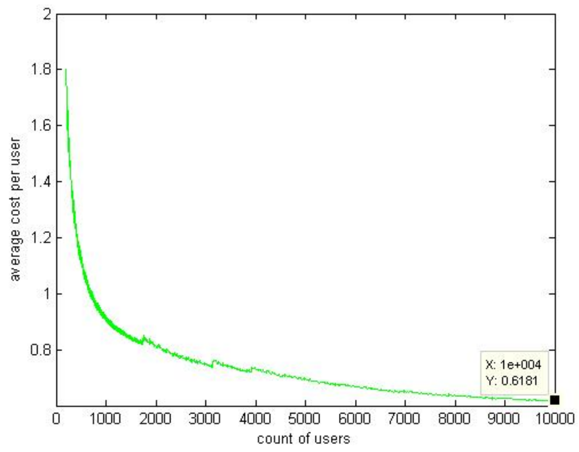


Fig. 10 Cost-per-user decreases along with the increasing of the users and lower than that without price coefficient

than 1, which means the prices of most data centers is higher than the first simulation environment, the cost decreases at the end. It confirms that, in our algorithm, the system will choose the cheaper data centers under the premise of keeping the QoS. According to the simulation results, 4 data centers are used. The 4 data centers on map is shown as Fig. 11

4.4 Experimental Results with Different Storage Price

In the simulation in last sections, we set the price of storage as 10 times of the price of VM. The ratio is decided by the storage size. If the whole size of videos is bigger, the ratio will be higher.

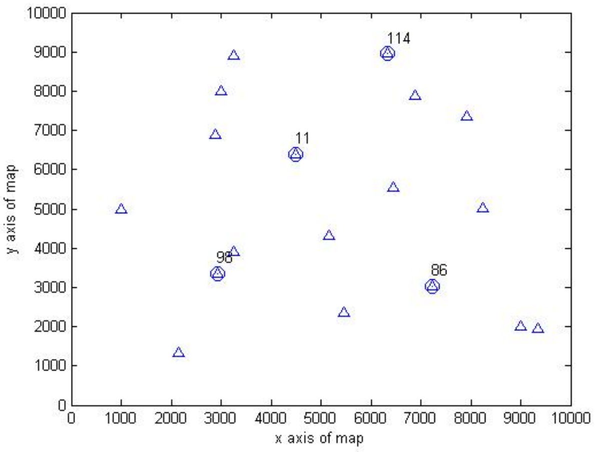


Fig. 11 The location of data centers which are used with price coefficient

In the next simulation, we will try different storage prices and analyze the performance of QoS along with the number of users.

1. we set the price of storage in each data center as 25 times of VM price. And all the prices of data center are the same. The performance of QoS is shown in Fig. 12

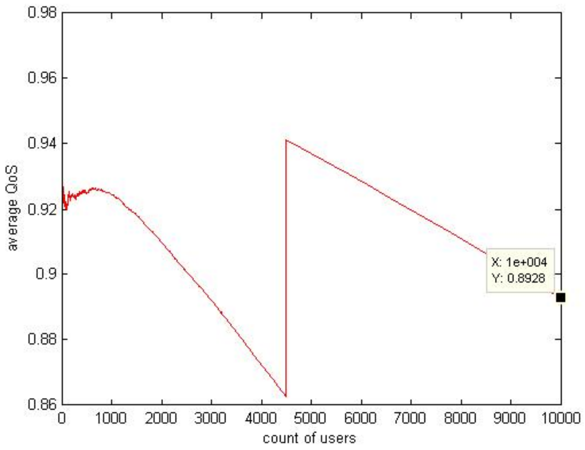


Fig. 12 QoS changes along with user count when the price of storage is very high,only 2 data centers are used in the end

As we can see, when the price of storage increase, the number of opened data center decreases. The number is 3 and less that of last section. This is easy to understand that the more expensive the fixed cost per data center is, the less data center will be started to use.

The used data centers on map is shown as Fig. 13

2. The price coefficient token into account. The QoS is shown in Fig. 14

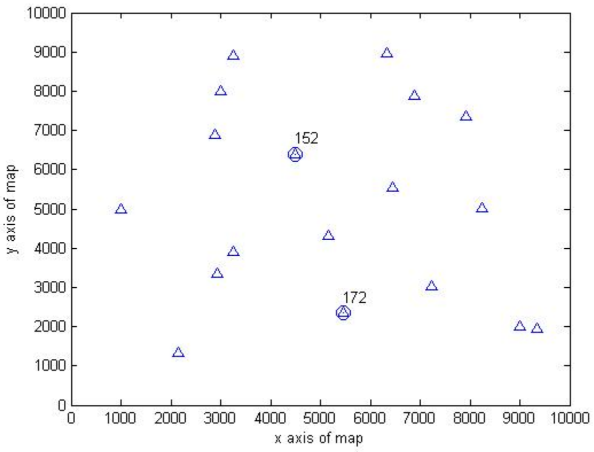


Fig. 13 The location of data centers which are used when the price of storage is very high

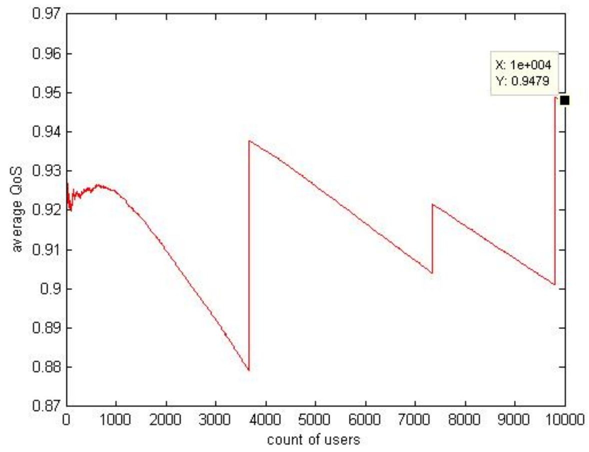


Fig. 14 QoS changes along with the user count with price coefficient when the price of storage is very high wit

3. No price coefficient used. The price of storage is cheaper to 2 times of that of VM. The QoS is shown in Fig. 15 The used data centers on map is shown as Fig. 16

4. The price coefficient is the same as last section. The price of storage is 2 times of that of VM. The QoS is shown in Fig. 17 The explain is the same as above. Almost all data centers are used.

5. No price coefficient used. The storage is free. The QoS is shown in Fig. 18

6. The price coefficient is the same as last section.The storage is free. The QoS is shown in Fig. 19 This is an extreme situation. When the storage is free, there is no fixed cost for each data center. All the data center will be used soon, and the QoS is much better than before.

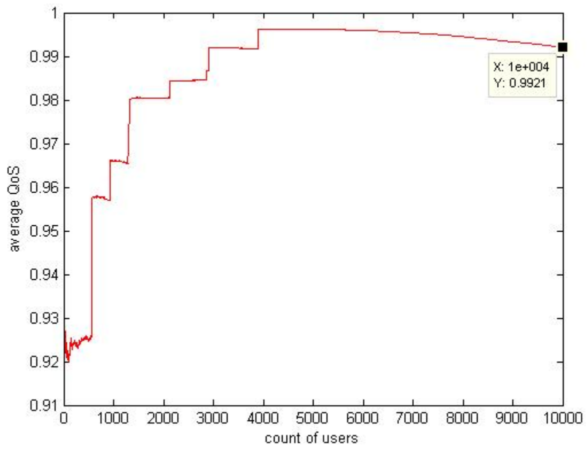


Fig. 15 QoS changes along with the user count when the price of storage is cheap

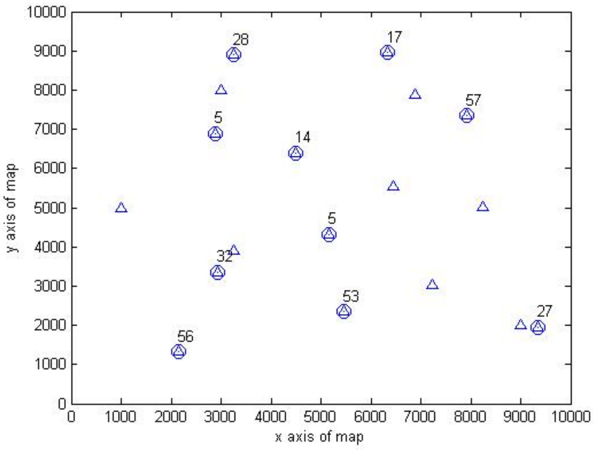


Fig. 16 locations of used data centers when the price of storage is cheap

5 Conclusions

In this paper, authors describe an algorithm of configure resources for a video stream service in the multi-cloud environment. Cloud providers are becoming more and more along with the technique developing. For a mature large-scale service, choosing more than one data center is a good choice. This algorithm is used to configure storage and VM resources in this situation. The main contribution of authors includes 2 points. First, authors described the algorithm and realized it. Second, authors made a simulation to validate the effectiveness of this algorithm.

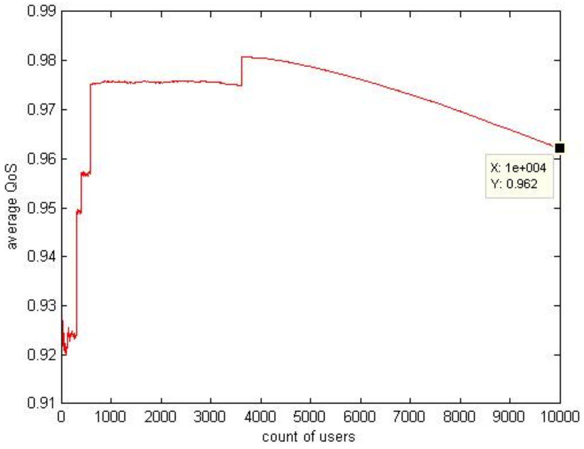


Fig. 17 QoS changes along with the user count when the price of storage is cheap with price coefficient

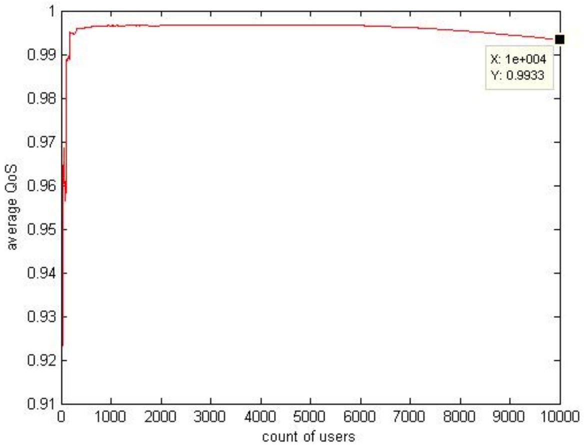


Fig. 18 QoS changs along with the user count when the price of storage is free

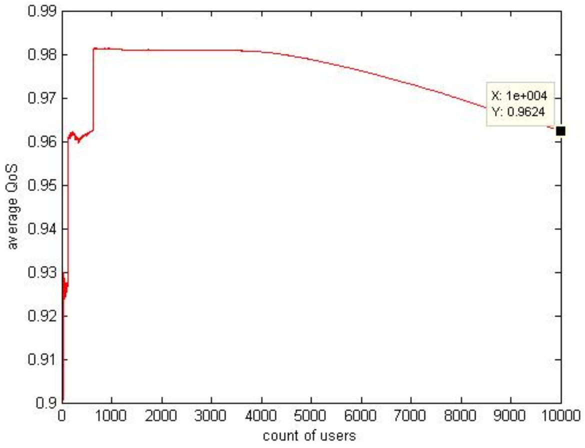


Fig. 19 QoS changes along with the user count when the price of storage is free with price coefficient

Acknowledgements

This work was supported in part by Ministry of Science and Technology of China under National 973 Basic Research Program (grants No. 2011CB302805 and No. 2013CB228206), National 863 Science and Technology Support Program (grant No. 2013BAH19F01) and National Natural Science Foundation of China (grant No. 61233016).

References

[1] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph,R. Katz, A. Konwinski, and M. Zaharia, ”A view of cloud computing,” *Communications of the ACM.*, vol. 53, no. 4, pp.50-58, 2010.

[2] I. Foster,Y. Zhao,I. Raicu, and S. Lu, Cloud computing and grid computing 360-degree compared,” presented at Grid Computing Environments Workshop, 2008, IEEE.

[3] M. A. AlZain, E. Pardede ,B. Soh, and J. A. Thom, ”Cloud computing security: from single to multi-clouds,” present at System Science (HICSS), 2012 45th Hawaii International Conference on. IEEE, 2012.

[4] Z. Chen, F. Han, J. Cao, and S. Chen, ”Cloud Computing-Based Forensic Analysis for Collaborative Network Security Management System,” *Tsinghua Science and Technology, Special Section on Cloud Computing*, vol. 18, no. 1, pp.40-50, 2013

[5] R. Buyya,C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility, *Future Generation Computer Systems* vol. 25, no. 6, pp.599-616, 2009.

[6] D. Kondo,B. Javadi,P. Malecot, F. Cappello, and D. P. Anderson, Cost-benefit analysis of cloud computing versus desktop grids, present at Parallel & Distributed Processing IPDPS 2009. IEEE International Symposium on. IEEE, 2009.

[7] D. Yuan, Y. Yang, X. Liu, and J. Chen, ”A cost-effective strategy for intermediate data storage in scientific cloud workflow systems,” presented at Parallel & Distributed Processing (IPDPS), 2010 IEEE International Symposium on. IEEE, 2010.

[8] J. Cao, K. Hwang, K. Li and A. Y. Zomaya, ”Optimal Multiserver Configuration for Profit Maximization in Cloud Computing,” *IEEE Trans. Parallel and Distributed Systems, Special Issue on Cloud Computing*, vol. 24, no. 6, pp. 1087-1096, 2013.

[9] A. Sampaio,N. Mendonca, ”Uni4Cloud: an approach based on open standards for deployment and management of multi-cloud applications,” presented at Processdings of the 2nd International Workshop on Software Engineering for Cloud Computing. ACM, 2011.

[10] J. L. Lucas Simarro,R. Moreno-Vozmediano,R. S. Montero, and I. M. Llorente, ”Dynamic placement of

virtual machines for cost optimization in multi-cloud environments," presented at High Performance Computing and Simulation (HPCS), 2011 International Conference on. IEEE, 2011.

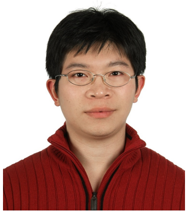
[11] J. Li, J. Chinneck, M. Woodside, M. Litoiu, G. Iszlai, "Performance model driven QoS guarantees and optimization in clouds" presented at Software Engineering Challenges of Cloud Computing, 2009. CLOUD'09. ICSE Workshop on. IEEE, 2009.

[12] D. Wu,Y. T. Hou, W. Zhu, Y. Q. Zhang, and J. M. Peha, "Streaming video over the Internet: approaches and directions," *Circuits and Systems for Video Technology, IEEE Transactions* vol. 11, no. 3 pp. 282-300 , 2001.

[13] S. Tao, R. Gurin, "Application-specific path switching: A case study for streaming video," presented at Proceedings of the 12th annual ACM international conference on Multimedia. ACM, 2004.

[14] S. Tao,K. Xu,A. Estepa, T. F. L. Gao, R. O. C. H. GUerin, J. Kurose, and Z. L. Zhang, "Improving VoIP quality through path switching," presented at INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE. IEEE, 2005.

[15] Y. Wu, C. Wu, B. Li, X. Qiu, and F. C. Lau, "Cloudmedia: When cloud on demand meets video on demand," presented at Distributed Computing Systems (ICDCS), 2011 31st International Conference on. IEEE, 2011.



Wei Chen born in 1988, he is now a master student of Research Institute of Information Technology, Tsinghua University. He received his bachelor degree in control theories and engineering in 2010 from Tsinghua University. Email: chenw06@mails.tsinghua.edu.cn. His research interests include cloud computing, mobile internet, online education.

Junwei Cao received his Ph.D. in computer science from the University of Warwick, Coventry, UK, in 2001. He received his bachelor and master degrees in control theories and engineering in 1996 and 1998, respectively, both from Tsinghua University, Beijing, China. He is currently a Professor and Vice Director, Research Institute of Information Technology, Tsinghua University, Beijing, China. He is also Director of Open Platform and Technology Division, Tsinghua National Laboratory for Information Science and Technology. Before joining Tsinghua University in 2006, he was a research scientist at MIT LIGO Laboratory and NEC Laboratories Europe for about 5 years. He has published over 150 papers and cited by international scholars for over 4,000 times. He is the book editor of Cyberinfrastructure Technologies and Applications, published by Nova Science in 2009. His research is focused on distributed computing and applications. Dr. Cao is a senior member of the IEEE Computer Society and a member of the ACM and CCF.

FIGURES. Below each figure, there should be a figure legend. The font is Times New Roman \zihao{5-}. A figure legend normally begins with a brief title describing its whole contents, and continues with a short description of each panel. The format of figure legends is "Fig. 1 Figure legend". The preferred figure formats is TIFF or JPEG with a preferred resolution of **600dpi** relative to the final figure size. When figures are divided into parts, each part should be labeled with a lower-case (a), (b), and so on, in the same font size as used elsewhere in the figure. All lettering in figures should be in lower-case, except for the first letter of each label which should be capitalized. Use **a single space** to separate a number and its units. Ensure that the labels are sufficiently large and clear to be readable when the figure is reproduced in the print version of the journal. Figures are referred to in the manuscript as "Fig. 1" and "Fig. 1a", or "Figure 1" at the beginning of a sen-tence. After a manuscript is accepted, we may ask the authors to provide high resolution figures. Failure to supply the file can significantly delay the publication of your work. If possible, please provide a set of high resolution figures along with your initial manuscript. The width of figures should better be 8, 12, or 17 cm, and the height should be less than 19 cm.

TABLES. The font is Times New Roman \zihao{5-}. Above each table, there should be a brief title describing its contents. Title format should follow "**Table 1 Table title**". All details, such as nonstandard abbreviations, and a description of standards of error analysis, should be included in footnotes. Tables are referred to in the manuscript as "Table 1". The concise type tables are recommended.

EQUATIONS. The font is Times New Roman \zihao{5}. Equations and mathematical expressions are identified by parenthetical numbers, such as (1), and are referred to in the manuscript as "Eq. (1)", or "Equation (1)" at the beginning of a sentence.