Towards Operational Cost Minimization in Hybrid Clouds for Dynamic Resource Provisioning with Delay-Aware Optimization using minimizing communication cost algorithm.

Kavitha.S[1],Dhanya.D[2],

[1]P.G.Scholar Dept of computer science and Engineering,Mar Ephraem college of Engineering and Technology, Marthandam, India.

[2]Dept of computer science and Engineering, AsstProf,Mar Ephraem college of Engineering and Technology,Marthandam,India.

Email:Kavi6165@gmail.com

Abstract- Recently, hybrid cloud computing paradigm has be widely advocated as a promising solution for Software-as-a-Service (SaaS) providers to effectively handle the dynamic user requests. With such a paradigm, the SaaS providers can extend their local services into the public clouds seamlessly so that the dynamic user request workload to a SaaS can be elegantly processed with both the local servers and the rented computing capacity in the public cloud. However, although it is suggested that a hybrid cloud may save cost compared with building a powerful private cloud, considerable renting cost and communication cost are still introduced in such a paradigm. How to optimize such operational cost becomes one major concern for the SaaS providers to adopt the hybrid cloud computing paradigm. However, this critical problem remains unanswered in the current state of the art. In this paper, we focus on optimizing the operational cost for the hybrid cloud paradigm by theoretically analyzing the problem with a Lyapunov optimization framework. This allows us to design an online dynamic provision algorithm. In this way, our approach can address the real-world challenges where no a priori information of public cloud renting prices is available and the future probability distribution of user requests is unknown. We then conduct extensive experimental study based on a set of real-world data, and the results confirm that our algorithm can work effectively in reducing the operational cost.

Keywords: cloud computing, optimization, cost minimization

1 INTRODUCTION

Cloud computing has surged into popularity in the IT industry, which can provision computing resource with a cost-effective, elastic solution. In recent years, a hybrid cloud paradigm is widely advocated by the industry practitioners, where a software-as-a-service (SaaS) provider although owning a small local data center can extend its services into a public infrastructure-as-a-service (IaaS) cloud. With such a paradigm, a SaaS provider can scale up and down its computing capacity by renting different numbers of virtual machines (VMs) in the public cloud according to the dynamic user demand instead of relying only on the fixed capacity of local data center. This can handle the dynamics of user requests elegantly and cost-effectively. There are diverse successful user cases. For example, Open-Text, a leading enterprise content management software provider, employs the hybrid cloud model to demonstrate their software, which makes their salespeople more productive and results in increased company revenue. Oxford University uses hybrid cloud to support their database services, in which researchers are able to use the database service in their virtual machines provided on the public cloud which is connected to their private cloud. One of the world's leading game companies, SEGA, uses hybrid cloud to improve its development process. In fact, more and more leading IaaS cloud solutions (e.g., Amazon EC2 and VMWare vCloud) are now aiming at such a hybrid cloud paradigm. A SaaS provider can now quickly and seamlessly adopt such a computing paradigm with a set of handy tools from the IaaS providers. But charming as it looks, the cost-effectiveness of such a paradigm highly depends on how well the SaaS provider can optimize the cost caused by renting VMs from the public IaaS cloud. Acquiring public IaaS computing capacity may actually cause a considerable cost.

Minimizing the cost of hybrid cloud operation is actually a very challenging task. First of all, the end users will be driven away if a SaaS cannot meet the service level agreement (SLA). In other words, a SaaS provider has to maintain its computing capacity while limiting the number of the VMs to reduce the renting cost at the same time. However, the user requests are highly dynamic in nature. Their patterns cannot be known and even accurately predicted in advance. Moreover, the communication cost between the local servers and the public IaaS cloud cannot be ignored, which unfortunately inherits the dynamics if the number of the renting VMs are dynamically tuned. Finally, the prices of VMs in the public IaaS cloud are typically varying and unpredictable. All these dynamic factors can have a great impact on the cost, and hence bring great difficulty to the cost minimization task.

However, existing approaches on deciding the cost-efficient computing capacity of the cloud generally requires a priori knowledge of the user demand and the VM prices, or an accurate prediction. They also do not consider the dynamics of user requests. As a result, they are not specifically tailored for optimizing the cost of hybrid cloud operation.

2 PROPOSED METHODOLOGY

In this paper we adopt Lyapunov optimization into the optimal cloud resource provisioning problem by tailoring the framework according to specifics of the hybrid cloud settings.We tackle the cost minimization problem with a fast online algorithm for dynamic cloud resource provisioning in hybrid clouds.In this paper, we focus on optimizing the operational cost for the hybrid cloud paradigm by theoretically analyzing the problem with a Lyapunov optimization framework. This allows us to design an online dynamic provision algorithm. In this way, our approach can address the real-world challenges where no a priori information of public cloud renting prices is available and the future probability distribution of user requests is unknown. We then conduct extensive experimental study based on a set of real-world data, and the results confirm that our algorithm can work effectively in reducing the operational cost.

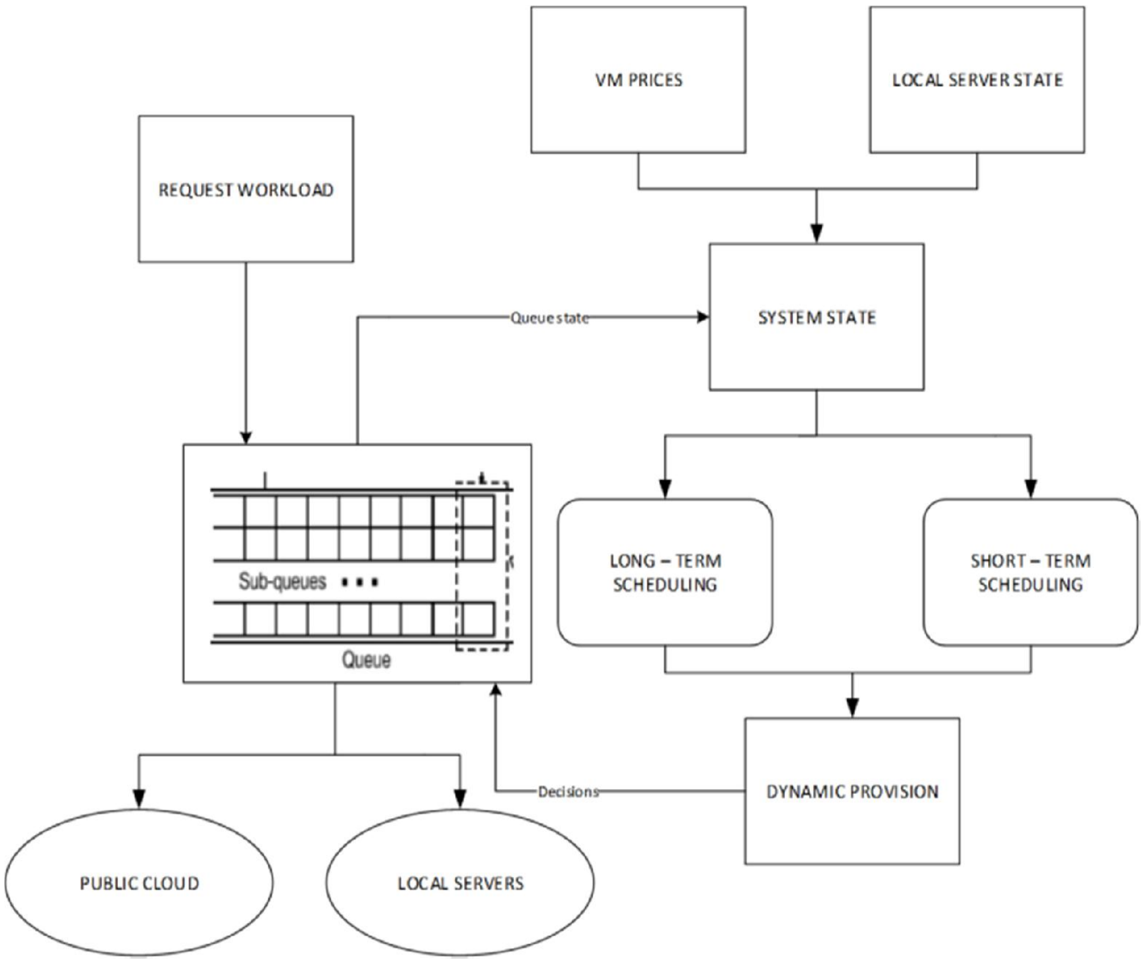


Figure 2.1: Architecture of ODPA

3 PROBLEM DESCRYPTION

In this section, we focus on the system model of the problem we are facing. We consider a SaaS provider operating with a small local data center (or local servers in our following discussions) and the provider can also extend its service capacity via renting the VMs from a public IaaS cloud. This is a typical hybrid cloud model. The public IaaS cloud will in general provision three types of VM services for the SaaS provider. The first is the reserved VM service, which is long-term service with a fixed VM numbers. The second is the on-demand VM service, where the number of VMs can be set instantaneously by the SaaS provider. The last is the on-spot VM services, which the SaaS provider can bid for. The price of the reserved service per unit is typically the lowest. The on-demand one is often the most expensive but it is charged in a pay-as-you-go fashion. The price of the on-spot one is dynamic according to the user bid (reflecting the user demand). Since renting the public cloud resources incurs monetary cost, the SaaS provider should reduce such cost. The key problem to the SaaS provider is how to decide the number of VMs it need to rent so that the performance requirement is satisfied and the cost is minimized. In the real-world scenario, the SaaS provide has to make decisions on the number of the VMs in some degrees of granularity and thus the reserved VM services can decided in long-term manner. The numbers of the other two types of VMs (the on demand ones and the on-spot ones) can be decided in a relatively short-term manner. In this way, how to determine the numbers of the above four types of VMs for the SaaS provider can be divided into a two-scale decision process.

- User Requests and SaaS Service Model

  The user requests to the SaaS provider usually have deadlines, allowing the request to wait for a maximum periods of time before it is scheduled. To serve the user requests, the SaaS provider allocates each request a certain number of VMs according to the request requirement, consisting of those in the local servers and those in the public IaaS cloud. Without loss of generality, we assume that VMs are homogenous in terms of service capacity (otherwise, the capacity of VMs can be normalized theoretically).

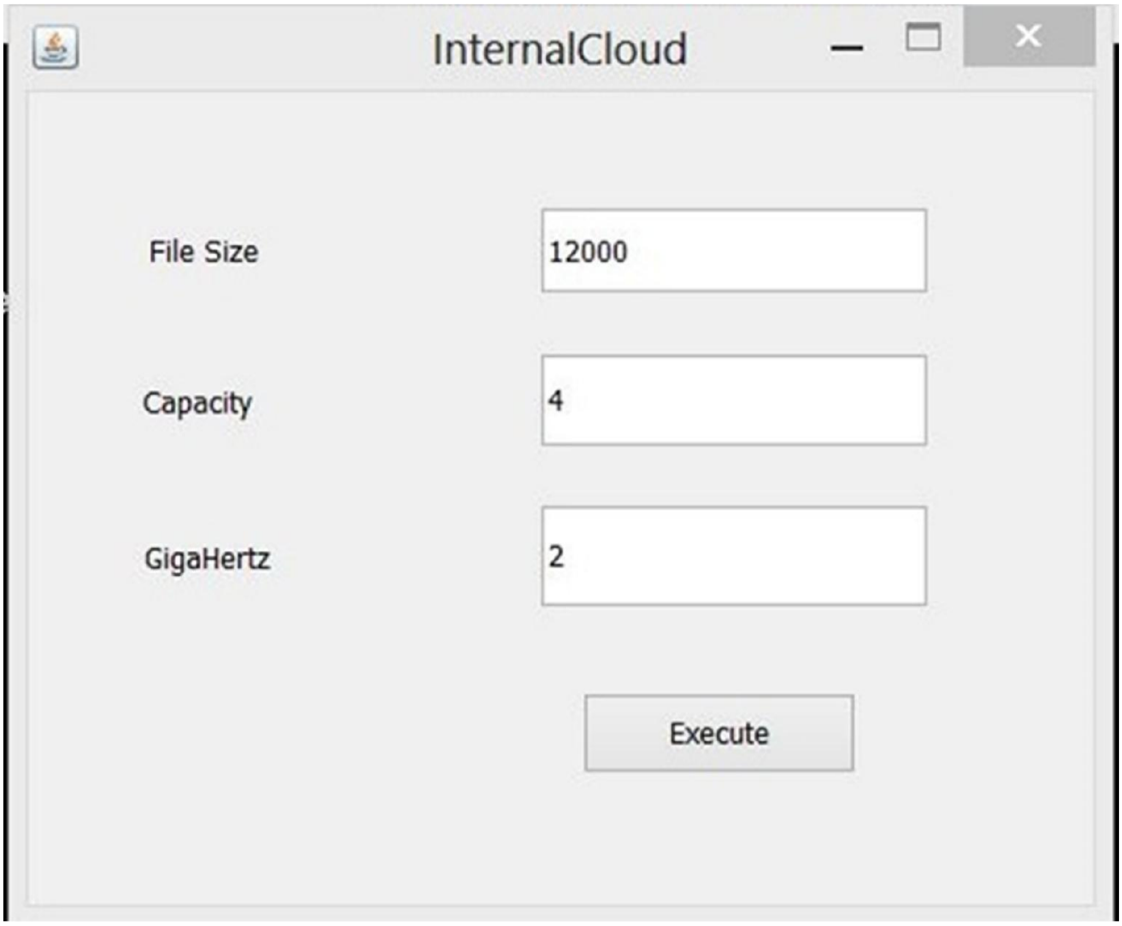
- Cost Model

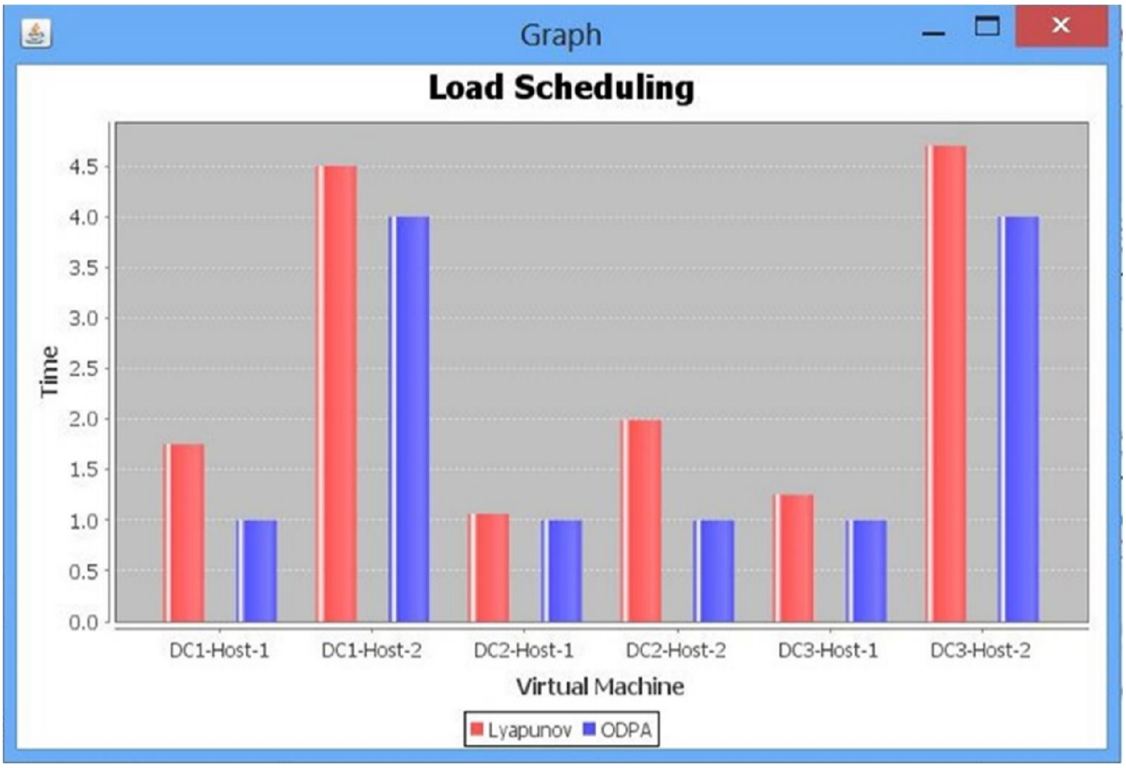
  We consider that the cost of the SaaS provider includes running the local servers, purchasing three types of VMs, and the communication cost between the VMs across the cloud and local servers.

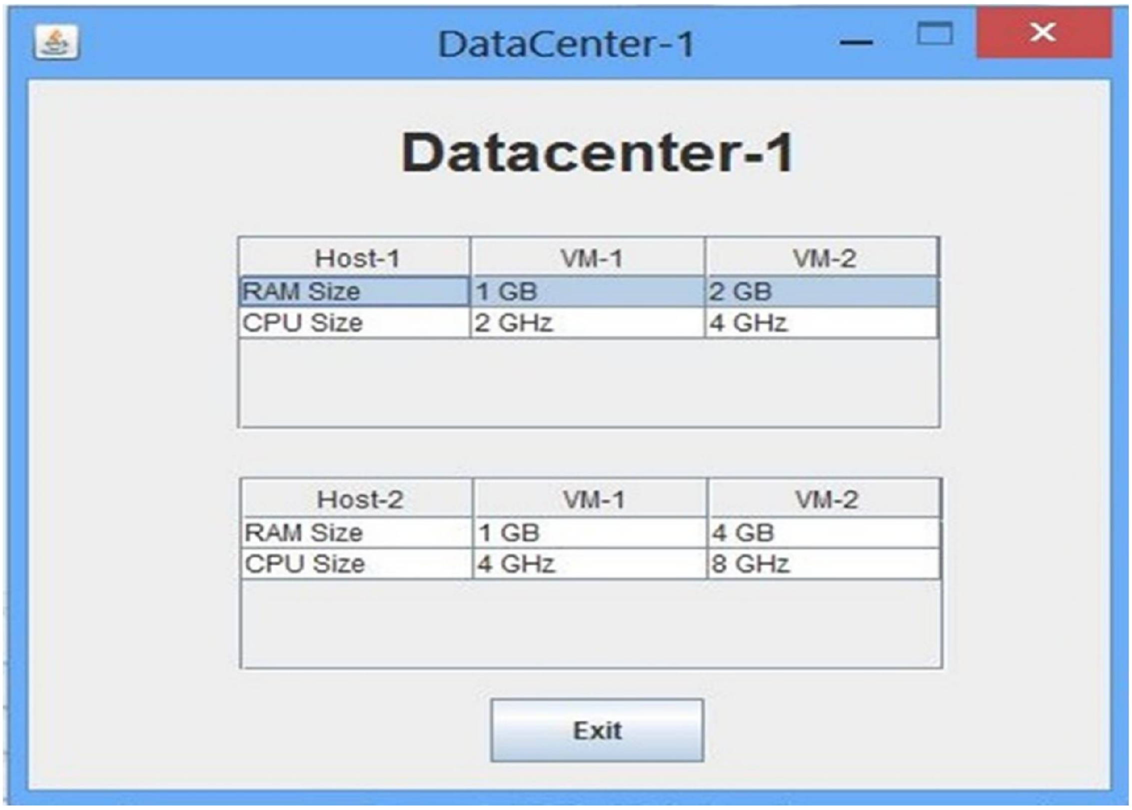
- Problem Formulation

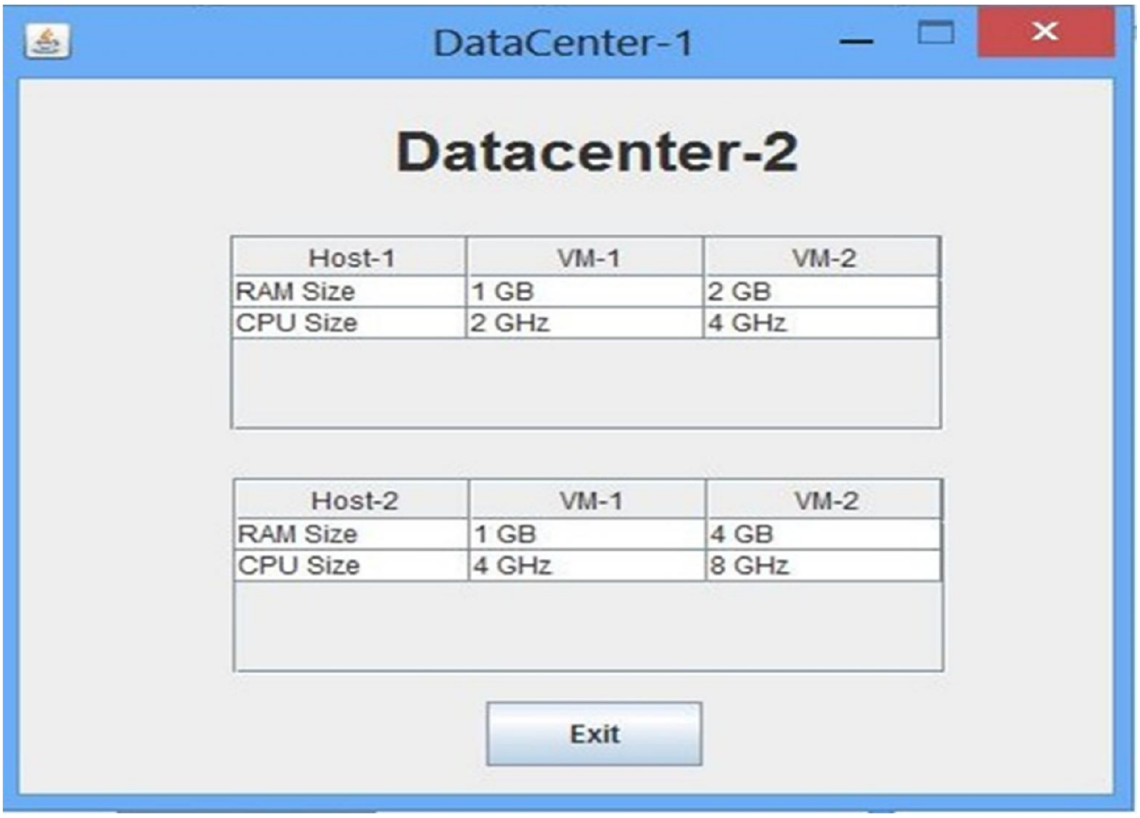
  The SaaS provider can optimize its cost via a two-scale decision process: decide the reserved VM number in each coarse-grained decision interval with length and decide the numbers of the on-demand and on-spot VMs in each fine-grained decision time slot in the interval, such that the time-average expecting cost is minimized.

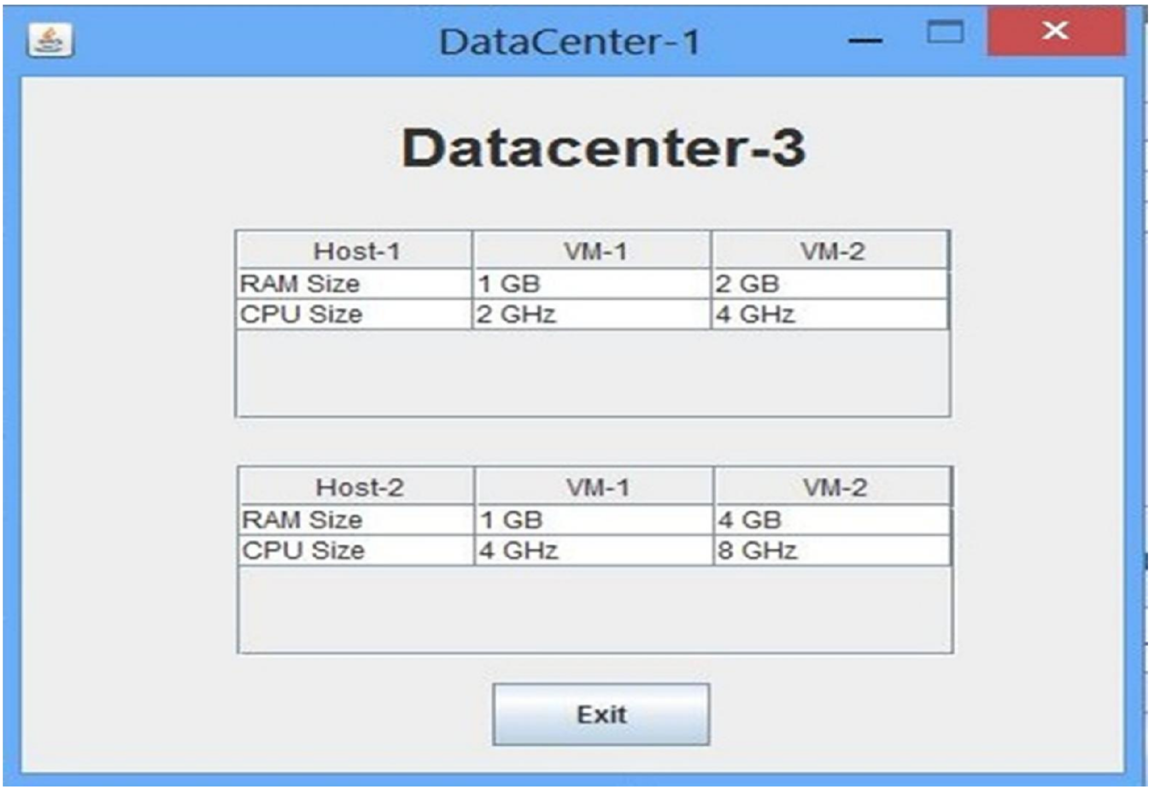
3 RESULT ANALYSIS

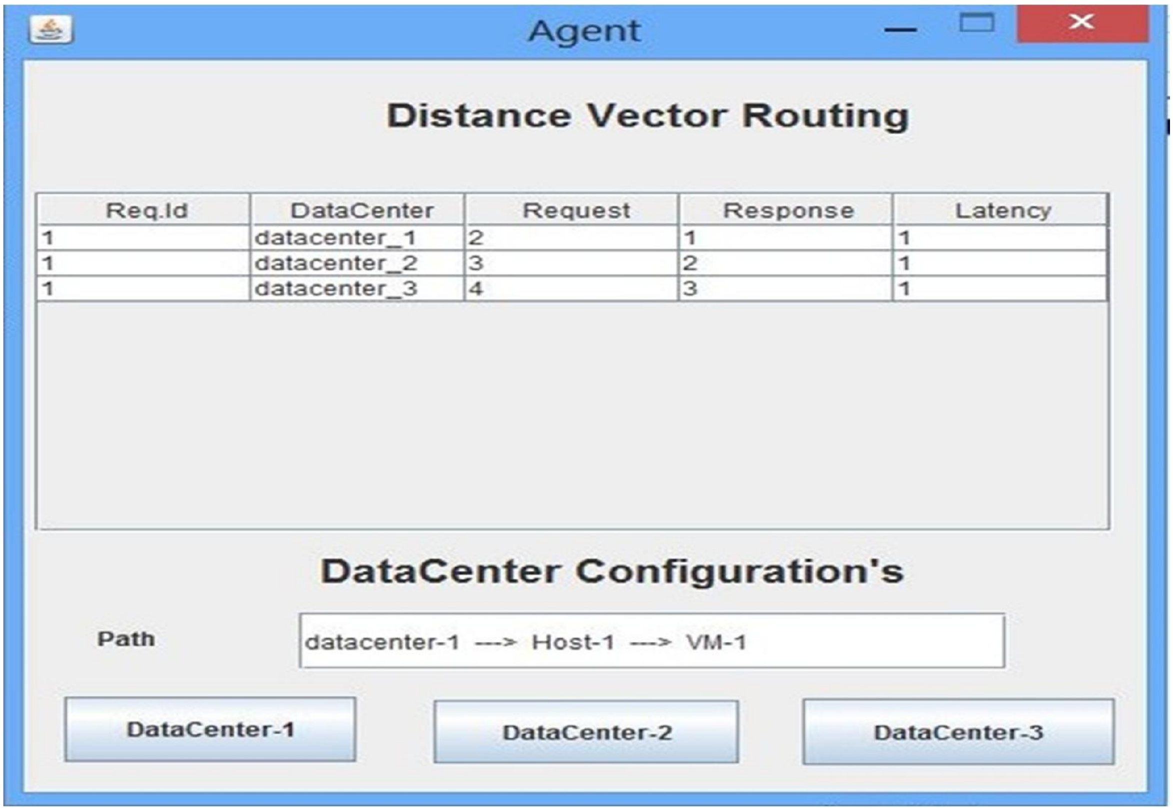

InternalCloud
File Size
12000
Capacity
4
GigaHertz
2
Execute


Graph
Load Scheduling
Time
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
DC1-Host-1
DC1-Host-2
DC2-Host-1
DC2-Host-2
DC3-Host-1
DC3-Host-2
Virtual Machine
Lyapunov
ODPA


DataCenter-1
Datacenter-1
Host-1 | VM-1 | VM-2
RAM Size | 1 GB | 2 GB
CPU Size | 2 GHz | 4 GHz
Host-2 | VM-1 | VM-2
RAM Size | 1 GB | 4 GB
CPU Size | 4 GHz | 8 GHz
Exit


DataCenter-1
Datacenter-2
Host-1 | VM-1 | VM-2
RAM Size | 1 GB | 2 GB
CPU Size | 2 GHz | 4 GHz
Host-2 | VM-1 | VM-2
RAM Size | 1 GB | 4 GB
CPU Size | 4 GHz | 8 GHz
Exit


DataCenter-1
Datacenter-3
Host-1 | VM-1 | VM-2
RAM Size | 1 GB | 2 GB
CPU Size | 2 GHz | 4 GHz
Host-2 | VM-1 | VM-2
RAM Size | 1 GB | 4 GB
CPU Size | 4 GHz | 8 GHz
Exit



5 CONCLUSION

This paper investigates how to optimize the monetary cost of purchasing cloud VMs for the hybrid cloud computing paradigm. Our work assumes an arbitrary request arriving probability and no accurate a priori knowledge of VM prices in the public cloud. We specifically tailor a theoretical model based on Lyapunov Optimization framework according to the real-world challenges of this problem. We then develop a method to minimize the time average cost with an online dynamic allocation algorithm. Both the theoretical analysis and the experimental study based on real-world data trace demonstrate the advantages of the algorithm. The evaluation shows that the online dynamic provision algorithm can achieve much lower cost than the conventional method and approach the ideal offline optimal method closely.

REFERENCE

[1] Li, Sinan, Yangzhong Zhou, Liangbao Jiao, Xiaodong Yan, Xiongfei Wang, and Michael R. Lyu. "Towards Operational Cost Minimization in Hybrid Clouds forDynamic Resource Provisioning with Delay-aware Optimization",IEEE TRANSACTIONS ON SERVICES COMPUTING, VOL. 8, NO. 3, MAY/JUNE 2015.

[2] Erol-Kantarci, Melike, and Hussein T. Mouftah. "Energy-Efficient Information and Communication Infrastructures in the Smart Grid: A Survey on Interactions and Open Issues." Communications Surveys & Tutorials, IEEE 17, no. 1,179-197, 2015.

[3] Champati, Jaya Prakash, and Ben Liang. "One-restart algorithm for scheduling and offloading in a hybrid cloud." In Proc. IEEE/ACM International Symposium on Quality of Service (IWQoS), 2015.

[4] Reaz, Abu Sayeem, Vishwanath Ramamurthi, Massimo Tornatore, and Biswanath Mukherjee. "Cloud-Integrated WOBAN: An offloading-enabled architecture for service-oriented access networks." Computer Networks 68, 5-19, 2014.

[5] Bhatt, Jignesh, Vipul Shah, and Omkar Jani. "An instrumentation engineer's review on smart grid: Critical applications and parameters." Renewable and Sustainable Energy Reviews 40, 1217-1239, 2014.

[6] Minhas, Umar Farooq. "Scalable and Highly Available Database Systems in the Cloud." PhD diss., University of Waterloo, 2013.

[7] Meng, Haiyan, Jing Li, Weiqing Liu, and Changchun Zhang. "MMSD: a Metadata-Aware Multi-Tiered Source Deduplication Cloud Backup System in the Personal Computing Environment." International Review on Computers and Software (IRECOS) 8, no. 2, 542-550, 2013.

[8] Kantarci, Burak, and Hussein T. Mouftah. "Designing an energy-efficient cloud network [Invited]." Journal of Optical Communications and Networking 4, no. 11, B101-B113, 2012.

[9] Wang, Chengwei, Karsten Schwan, Vanish Talwar, Greg Eisenhauer, Liting Hu, and Matthew Wolf. "A flexible architecture integrating monitoring and analytics for managing large-scale data centers." In Proceedings of the 8th ACM international conference on Autonomic computing, pp. 141-150. ACM, 2011.

[10] Chen, Jaime, Manuel Díaz, Luis Llopis, Bartolomé Rubio, and José M. Troya. "A survey on quality of service support in wireless sensor and actor networks: Requirements and challenges in the context of critical infrastructure protection", Journal of Network and Computer Applications 34, no. 4,1225-1239, 2011.

[11] Chowdhury, Pulak. "Energy-efficient next-generation networks (E2NGN)." PhD diss., University of California Davis, 2011.