

Optimized Data Center in IaaS Using Cloud Computing Systems

K. Hemapriya¹, J. Deepa², S. Kaviarasan³

Assistant Professor, Department of CSE, Panimalar Institute of Technology, Chennai, India¹

Assistant Professor, Department of CSE, Panimalar Institute of Technology, Chennai, India²

Assistant Professor, Department of CSE, Panimalar Institute of Technology, Chennai, India³

ABSTRACT: Cloud Computing provides scalable computing and storage resources. More and more data-intensive applications are developed in this computing environment. Evaluation of cloud data center performance is a key attribute in propagation of IaaS cloud services. This evaluation is necessary to quantify the cost benefit and Quality Of Service (QoS) of Cloud services. Cloud system managers find it difficult to maintain performance data due to the huge number of parameters that need to be monitored in a cloud system. Different applications have different *quality-of-service (QoS)* requirements. In this paper, we present an analytical model, based on *Stochastic Reward Nets (SRNs)*, that is both scalable to model systems composed of thousands of resources and flexible to represent different policies and cloud-specific strategies. Several performance metrics are defined and evaluated to analyze the behavior of a Cloud data center: *utilization, availability, waiting time, and responsiveness*. We are also notifying the resiliency analysis to take into account the load bursts. Finally, a general approach is presented that, starting from the concept of system capacity, can help system managers to opportunely set the data center parameters under different working conditions.

KEYWORDS: Cloud computing, stochastic reward nets, cloud-oriented performance metrics, resiliency, responsiveness.

1. INTRODUCTION

Cloud computing is the use of computing resources (hardware and software) that are delivered as a service over a network (typically the Internet). The name comes from the common use of a cloud-shaped symbol as an abstraction for the complex infrastructure it contains in system diagrams. Cloud computing entrusts remote services with a user's data, software and computation. Cloud computing consists of hardware and software resources made available on the Internet as managed third-party services. These services typically provide access to advanced software applications and high-end networks of server computers. The goal of cloud computing is to apply traditional supercomputing, or high-performance computing power, normally used by military and research facilities, to perform tens of trillions of computations per second, in consumer-oriented applications such as financial portfolios, to deliver personalized information, to provide data storage or to power large, immersive computer games. The cloud computing uses networks of large groups of servers typically running low-cost consumer PC technology with specialized connections to spread data-processing chores across them. This shared IT infrastructure contains large pools of systems that are linked together. Often, virtualization techniques are used to maximize the power of cloud computing. *On-demand self-service* is a consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service's provider. Resource pooling is where the provider's computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. Measured service is Cloud systems automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts). Capabilities can be rapidly and elastically provisioned, in some cases automatically, to quickly scale out and rapidly released to quickly scale in. The characteristics of cloud computing can be given by the Service Models. Cloud Computing comprises three different service models, namely *Infrastructure-as-a-Service (IaaS)*, *Platform-as-a-Service (PaaS)*, and *Software-as-a-Service (SaaS)*. The three service models or layer are completed by an end user layer that

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2015

encapsulates the end user perspective on cloud services. If a cloud user accesses services on the infrastructure layer, for instance, she can run her own applications on the resources of a cloud infrastructure and remain responsible for the support, maintenance, and security of these applications herself. If she accesses a service on the application layer, these tasks are normally taken care of by the cloud service providers.

In section 2, we summarize previous work and point out main ideas that emerge from it. In section 3, proposed system is discussed. In section 4, system architecture. In section 5, experimental results are discussed. In section 6, we summarize our proposed work.

II. PREVIOUS WORK

Here we review previous work, highlighting the concepts that will be exploited in our work.

Live virtual machine migration via asynchronous replication and state synchronization gives the design and implementation of a novel approach, namely, CR/TR-Motion, which adopts checkpointing/recovery and trace/replay technologies to provide fast, transparent VM migration for both LAN and WAN environments. With execution trace logged on the source host, a synchronization algorithm is performed to orchestrate the running source and target VMs until they reach a consistent state. CR/TR-Motion can greatly reduce the migration downtime and network bandwidth consumption.

The CloudSim toolkit supports modelling and creation of one or more virtual machines (VMs) on a simulated node of a Data Center, jobs, and their mapping to suitable VMs. It also allows simulation of multiple Data Centers to enable a study on federation and associated policies for migration of VMs for reliability and automatic scaling of applications. The dependability of cloud services can be analyzed with long-term performance traces from Amazon Web Services and Google App Engine, currently two of the largest commercial clouds in production. We find that the performance of about half of the cloud services we investigate exhibits yearly and daily patterns, but also that most services have periods of especially stable performance. Last, through trace-based simulation we assess the impact of the variability observed for the studied cloud services on three large-scale applications, job execution in scientific computing, virtual goods trading in social networks, and state management in social gaming. We show that the impact of performance variability depends on the application, and give evidence that performance variability can be an important factor in cloud provider selection. The Performance evaluation of cloud computing is done for the platform and not for a particular service instance. the approach focuses on NFPs (nonfunctional properties) of individual services and thereby provides a more relevant and granular information. An experimental evaluation in Amazon Elastic Compute Cloud (EC2) verified this approach.

III. PROPOSED SYSTEM

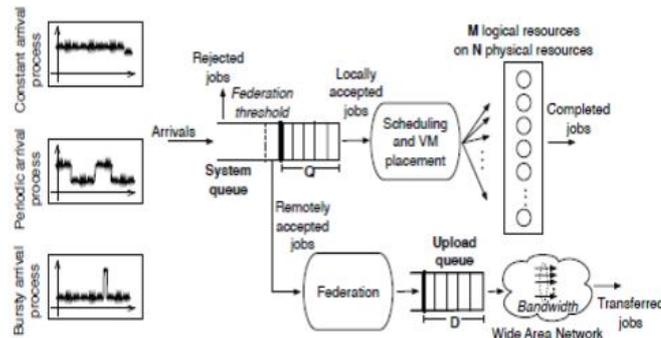
In our proposed system, we present a stochastic model, based on Stochastic Reward Nets (SRNs), that exhibits the above mentioned features allowing to capture the key concepts of an IaaS cloud system. The proposed model is scalable enough to represent systems composed of thousands of resources and it makes possible to represent both physical and virtual resources exploiting cloud specific concepts such as the infrastructure elasticity. With respect to the existing literature, the innovative aspect of the present work is that a generic and comprehensive view of a cloud system is presented. Low level details, such as VM multiplexing, are easily integrated with cloud based actions such as federation, allowing to investigate different mixed strategies. An exhaustive set of performance metrics are defined regarding both the system provider (e.g., utilization) and the final users (e.g., responsiveness). To provide a fair comparison among different resource management strategies, also taking into account the system elasticity, a performance evaluation approach is described. Such an approach, based on the concept of system capacity, presents a holistic view of a cloud system and it allows system managers to study the better solution with respect to an established goal and to opportunely set the system parameters.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2015

IV. SYSTEM ARCHITECTURE



Here, we consider an IaaS cloud system composed of physical resources. Job requests (in terms of VM instantiation requests) are enqueued in the system queue. Once the queue limit is reached, further requests are rejected. The system queue is managed according to a ROUND ROBIN scheduling policy. When a resource is available, a job is accepted and the corresponding VM is instantiated. We assume that the instantiation time is negligible and that the service time (i.e., the time needed to execute a job) is exponentially distributed. According to the VM multiplexing technique, the cloud system can provide a number of logical resources. In this case, multiple VMs can be allocated in the same physical machine (PM). The performance degradation of multiplexed VMs depends on the multiplexing technique and on the VM placement strategy. The system is able to optimally balance the load among the PMs with respect to the resources required by VMs. Cloud federation allows the system to use, in particular situations, the resources offered by other public cloud systems through a sharing and paying model. In this way, elastic capabilities can be exploited to respond to particular load conditions. Job requests can be redirected to other clouds by transferring the corresponding VM disk images through the network. A job is redirected only if it arrives when the system queue is full. A redirected job is inserted in the upload queue waiting for the VM transfer completion. There is a maximum number of concurrent redirected jobs (elasticity level). Finally, the server status becomes normal.

V. EXPERIMENTAL RESULTS

HOME PAGE:

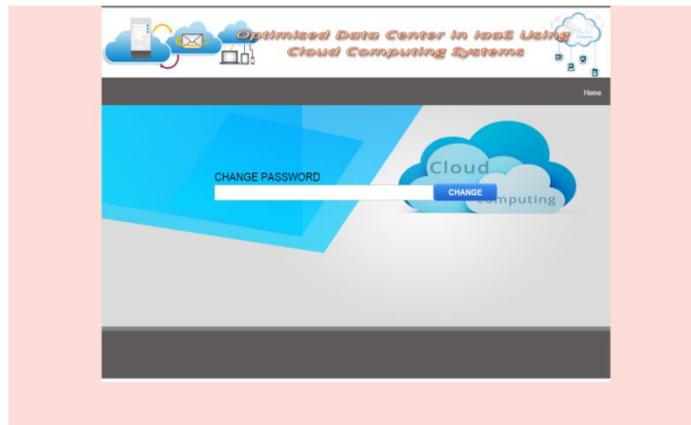


International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2015

CHANGE PASSWORD:



MONITOR SEVER:



GRAPHS:



International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2015

LOCATIONS:

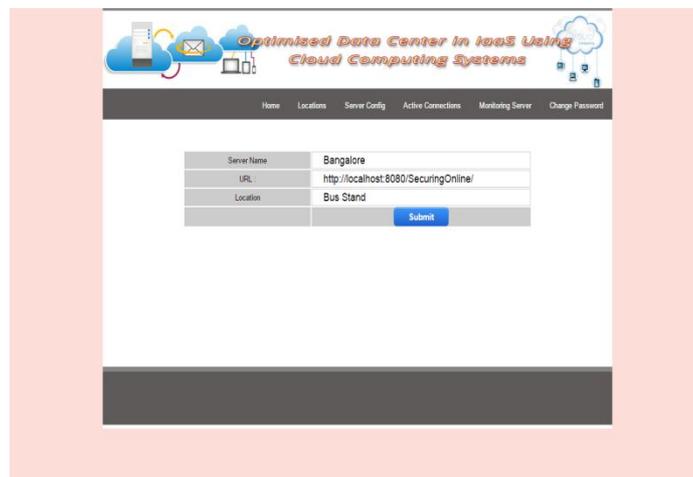


Optimised Data Center in IaaS Using Cloud Computing Systems

Home Add Locations

S NO	Server	URL	Location	Total	Available	Status	Edit	Delete
1	Bangalore	http://localhost:8080/SecuringOnline/	Bus Stand	3	3	OverLoad	Edit	Delete
2	Chennai	http://localhost:8080/SecuringOnline/	Anna Nagar	3	1	Normal	Edit	Delete
3	Tamil Nadu	http://localhost:8080/SecuringOnline/	Pune	3	0	Idle	Edit	Delete

EDIT LOCATIONS:



Optimised Data Center in IaaS Using Cloud Computing Systems

Home Locations Server Config Active Connections Monitoring Server Change Password

Server Name	Bangalore
URL	http://localhost:8080/SecuringOnline/
Location	Bus Stand
Submit	

ADD LOCATIONS:



Optimised Data Center in IaaS Using Cloud Computing Systems

Home

Server Name :

URL :

Location :

Submit

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2015

ACCESS SERVER:



VI. CONCLUSION

In this paper, we have presented a stochastic model to evaluate the performance of an IaaS cloud system. Several performance metrics have been defined, such as availability, utilization, and responsiveness, allowing to investigate the impact of different strategies on both provider and user point-of-views. In a market-oriented area, such as the Cloud Computing, an accurate evaluation of these parameters is required in order to quantify the offered QoS and opportunely manage SLAs. Future works will include the analysis of autonomic techniques able to change on-the-fly the system configuration in order to react to a change on the working conditions. We will also extend the model in order to represent PaaS and SaaS Cloud systems and to integrate the mechanisms needed to capture VM migration and data center consolidation aspects that cover a crucial role in energy saving policies.

REFERENCES

- [1] R. Buyya *et al.*, "Cloud computing and emerging it platforms: Vision, hype, and reality for delivering computing as the 5th utility," *Future Gener. Comput. Syst.*, vol. 25, pp. 599–616, June 2009.
- [2] X. Meng *et al.*, "Efficient resource provisioning in compute clouds via vm multiplexing," in *Proceedings of the 7th international conference on Autonomic computing*, ser. ICAC '10. New York, NY, USA: ACM, 2010, pp. 11–20.
- [3] H. Liu *et al.*, "Live virtual machine migration via asynchronous replication and state synchronization," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 22, no. 12, pp. 1986–1999, dec. 2011.
- [4] B. Rochwerger *et al.*, "Reservoir - when one cloud is not enough," *Computer*, vol. 44, no. 3, pp. 44–51, march 2011.
- [5] R. Buyya, R. Ranjan, and R. Calheiros, "Modeling and simulation of scalable cloud computing environments and the cloudsim toolkit: Challenges and opportunities," in *High Performance Computing Simulation, 2009. HPCS '09. International Conference on*, june 2009, pp. 1–11.
- [6] A. Iosup, N. Yigitbasi, and D. Epema, "On the performance variability of production cloud services," in *Cluster, Cloud and Grid Computing (CCGrid), 2011 11th IEEE/ACM International Symposium on*, may 2011, pp. 104–113.
- [7] V. Stantchev, "Performance evaluation of cloud computing offerings," in *Advanced Engineering Computing and Applications in Sciences, 2009. ADVCOMP '09. Third International Conference on*, oct. 2009, pp. 187–192.
- [8] S. Ostermann *et al.*, "A Performance Analysis of EC2 Cloud Computing Services for Scientific Computing," in *Cloud Computing*, ser. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering. Springer Berlin Heidelberg, 2010, vol. 34, ch. 9, pp. 115–131.
- [9] H. Khazaei, J. Mistic, and V. Mistic, "Performance analysis of cloud computing centers using m/g/m/m+ τ queuing systems," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 23, no. 5, pp. 936–943, may 2012.
- [10] R. Ghosh, K. Trivedi, V. Naik, and D. S. Kim, "End-to-end performance analysis."
- [11] G. Ciardo *et al.*, "Automated generation and analysis of Markov reward models using stochastic reward nets." *IMA Volumes in Mathematics and its Applications: Linear Algebra, Markov Chains, and Queueing Models*, vol. 48, pp. 145–191, 1993.
- [12] D. Gupta, L. Cherkasova, R. Gardner, and A. Vahdat, "Enforcing performance isolation across virtual machines in xen," in *Proceedings of the ACM/IFIP/USENIX International Conference on Middleware*, New York, NY, USA: Springer-Verlag New York, Inc., 2006, pp. 342–362.
- [13] M. Armbruster *et al.*, "A view of cloud computing," *Commun. ACM*, vol. 53, pp. 50–58, Apr. 2010.
- [14] J. N. Matthews *et al.*, "Quantifying the performance isolation properties of virtualization systems," in *Proceedings of the 2007 workshop on Experimental computer science*, ser. ExpCS '07. New York, NY, USA: ACM, 2007.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2015

- [15] M. Mishra and A. Sahoo, "On theory of vm placement: Anomalies in existing methodologies and their mitigation using a novel vector based approach," in *Cloud Computing (CLOUD), 2011 IEEE International Conference on*, July 2011, pp. 275–282.
- [16] A. V. Do *et al.*, "Profiling applications for virtual machine placement in clouds," in *Cloud Computing (CLOUD), 2011 IEEE International Conference on*, July 2011, pp. 660–667.
- [17] A. Verma *et al.*, "Server workload analysis for power minimization using consolidation," in *Proceedings of the USENIX Annual technical conference*, Berkeley, CA, USA, 2009, pp. 28–28.
- [18] G. Balbo *et al.*, *Modelling with Generalized Stochastic Petri Nets*. John Wiley and Sons, 1995.
- [19] R. Sahner, K. S. Trivedi, and A. Puliafito, *Performance and reliability analysis of computer systems: an example based approach using the SHARPE software package*. Kluwer Academic Publishers, 1995.
- [20] J. K. Muppala, K. S. Trivedi, and S. P. Woolet, "On modeling performance of real-time systems in the presence of failures." *Readings in Real-Time Systems*, pp. 219–239, 1993.