



A Stochastic Model to Investigate Data Center Performance and QoS in IaaS Cloud Computing Systems

B. Sundarraj¹, K.G.S. Venkatesan², Vimal Chand³

Assistant Professor, Dept. of C.S.E., Bharath University, Chennai, Tamilnadu, India¹

Associate Professor, Dept. of C.S.E., Bharath University, Chennai, Tamilnadu, India²

Dept. of C.S.E., Bharath University, Chennai, Tamilnadu, India³

ABSTRACT:-Cloud data center management is a key problem due to the numerous and heterogeneous strategies that can be applied, ranging from the VM placement to the federation with other clouds. Performance evaluation of Cloud Computing infrastructures is required to predict and quantify the cost-benefit of a strategy portfolio and the corresponding Quality of Service (QoS) experienced by users. Such analyses are not feasible by simulation or on-the-field experimentation, due to the great number of parameters that have to be investigated. 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. A resiliency analysis is also provided to take into account 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: Performance evaluation, Cost-benefit, Quality of service, Cloud-specific, Resiliency analysis

I. 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 [2].

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 [3]. 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 [5].

II. LITERATURE SURVEY

Live migration of virtual machines (VM) across physical hosts provides a significant new benefit for administrators of data centers and clusters. Previous memory-to-memory approaches demonstrate the effectiveness of live VM migration in local area networks (LAN), but they would cause a long period of downtime in a wide area



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

network (WAN) environment [9]. This paper describes the design and implementation of a novel approach, namely, CR/TR-Motion, which adopts check pointing/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 [8].

Cloud computing aims to power the next generation data centers and enables application service providers to lease data center capabilities for deploying applications depending on user QoS (Quality of Service) requirements. Cloud applications have different composition, configuration, and deployment requirements. Quantifying the performance of resource allocation policies and application scheduling algorithms at finer details in Cloud computing environments for different application and service models under varying load, energy performance, and system size is a challenging problem to tackle [11].

Cloud computing is an emerging infrastructure paradigm that promises to eliminate the need for companies to maintain expensive computing hardware. Through the use of virtualization and resource time-sharing, clouds address with a single set of physical resources a large user base with diverse needs. Thus, clouds have the potential to provide their owners the benefits of an economy of scale and, at the same time, become an alternative for both the industry and the scientific community to self-owned clusters, grids, and parallel production environments. For this potential to become reality, the first generation of commercial clouds needs to be proven to be dependable. In this work we analyze the dependability of cloud services. Towards this end, we analyze long-term performance traces from Amazon Web Services and Google App Engine, currently two of the largest commercial clouds in production [12].

Advanced computing on cloud computing infrastructures can only become viable alternative for the enterprise if these infrastructures can provide proper levels of non functional properties (NPFs). A company that focuses on service-oriented architectures (SOA) needs to know what configuration would provide the proper levels for individual services if they are deployed in the cloud. In this paper we present an approach for performance evaluation of cloud computing configurations. While cloud computing providers assure certain service levels, this is typically done for the platform and not for a particular service instance [15].

Cloud Computing is emerging today as a commercial infrastructure that eliminates the need for maintaining expensive computing hardware. Through the use of virtualization, clouds promise to address with the same-shared set of physical resources a large user base with different needs. Thus, clouds promise to be for scientists an alternative to clusters, grids, and supercomputers. However, virtualization may induce significant performance penalties for the demanding scientific computing workloads. In this work we present an evaluation of the usefulness of the current cloud computing services for scientific computing. We analyze the performance of the Amazon EC2 platform using micro-benchmarks and kernels. While clouds are still changing, our results indicate that the current cloud services need an order of magnitude in performance improvement to be useful to the scientific community [18].

III. PROBLEM IDENTIFICATION

In cloud environment the server utilization is not properly maintained. For example, a web site has three web servers; each server is capable of handling 1 lacks requests at a time. It is not mandatory that all servers need to be online even when there is no enough request. This leads to high cost, because deploying the IaaS need more cost. This project explains the way of solving this problem in the efficient way. Another problem occurs if the request limit rises immediately, this may lead to crash all the active servers at a time. Recently “Flipcart.com” site has been crashed due to high bandwidth at a time [14].

IV. EXISTING SYSTEM

In order to integrate business requirements and application level needs, in terms of Quality of Service (QoS), cloud service provisioning is regulated by Service Level Agreements (SLAs) : contracts between clients and providers that express the price for a service, the QoS levels required during the service provisioning, and the penalties associated with the SLA violations. In such a context, performance evaluation plays a key role allowing system managers to

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

evaluate the effects of different resource management strategies on the data center functioning and to predict the corresponding costs/benefits [19].

Cloud systems differ from traditional distributed systems. First of all, they are characterized by a very large number of resources that can span different administrative domains. Moreover, the high level of resource abstraction allows to implement particular resource management techniques such as VM multiplexing or VM live migration that, even if transparent to final users, have to be considered in the design of performance models in order to accurately understand the system behaviour [23].

n-the-field experiments are mainly focused on the offered QoS, they are based on a black box approach that makes difficult to correlate obtained data to the internal resource management strategies implemented by the system provider.

Simulation does not allow conducting comprehensive analyses of the system performance due to the great number of parameters that have to be investigated [25].

V. PROPOSED SYSTEM

Here is the proposed system based on Stochastic Reward Nets (SRNs), that exhibits the above mentioned features allowing capturing the key concepts of an IaaS cloud system [29]. 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 is defined regarding both the system provider (e.g., utilization) and the final users (e.g., responsiveness) [30].

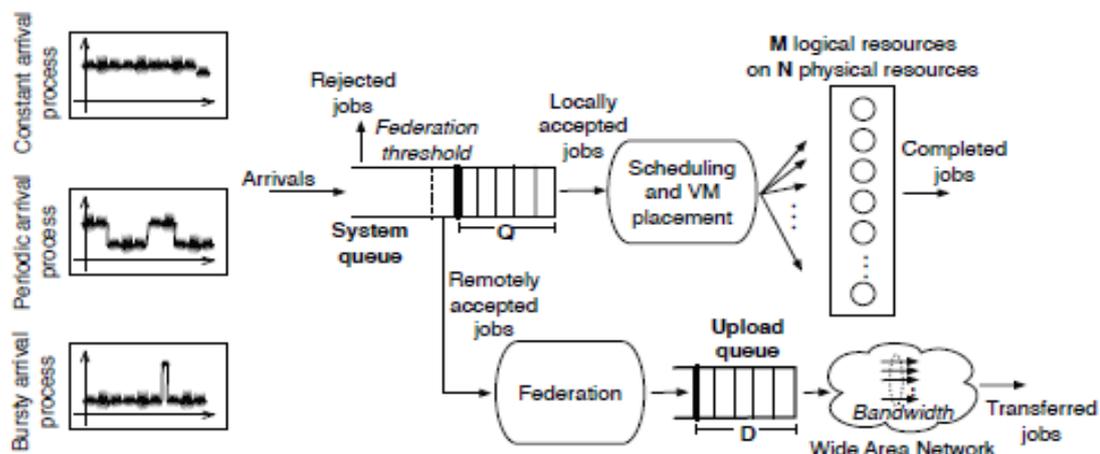


Fig. 1. An IaaS cloud system with federation.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

VI. MODULES

There are five modules in this system:

1. System Queuing
2. Scheduling Module
3. VM Placement Module
4. Federation Module
5. Arrival Process

A. System Queuing:

Job requests (in terms of VM instantiation requests) are en-queued in the system queue. Such a queue has a finite size Q , once its limit is reached further requests are rejected. The system queue is managed according to a FIFO scheduling policy [31].

B. Scheduling Module:

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 with mean $1/\mu$ [36].

C. VM Placement:

According to the VM multiplexing technique the cloud system can provide a number M of logical resources greater than N . In this case, multiple VMs can be allocated in the same physical machine (PM), e.g., a core in a multicore architecture. Multiple VMs sharing the same PM can incur in a reduction of the performance mainly due to I/O interference between VMs [35].

D. Federation Module:

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 in order to respond to particular load conditions. Job requests can be redirected to other clouds by transferring the corresponding VM disk images through the network [37].

E. Arrival Process:

Finally, we respect to the arrival process we will investigate three different scenarios. In the first one (Constant arrival process) we assume the arrival process be a homogeneous Poisson process with rate λ . However, large scale distributed systems with thousands of users, such as cloud systems, could exhibit self-similarity/long-range dependence with respect to the arrival process. The last scenario (Bursty arrival process) takes into account the presence of a burst with fixed and short duration and it will be used in order to investigate the system resiliency [38].

VII. CONCLUSION

A stochastic model was presented here to evaluate the performance of an IaaS cloud system. Several performance metrics have been defined, such as availability, utilization, and responsiveness, allowing investigating 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. This can also extend the model in order to



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

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.

VIII. ACKNOWLEDGEMENT

The author would like to thank the Vice Chancellor, Dean-Engineering, Director, Secretary, Correspondent, HOD of Computer Science & Engineering, **Dr. K.P. Kaliyamurthie**, Bharath University, Chennai for their motivation and constant encouragement. The author would like to specially thank **Dr. A. Kumaravel, Dean School of Computing**, for his guidance and for critical review of this manuscript and for his valuable input and fruitful discussions in completing the work and the Faculty Members of Department of Computer Science & Engineering. Also, he takes privilege in extending gratitude to his parents and family members who rendered their support throughout this Research work.

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, December - 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 cloud sim toolkit: Challenges and opportunities," in *High Performance Computing Simulation, 2009. HPCS '09. International Conference on*, PP. 1–11, June - 2009.
6. [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. K.G.S. Venkatesan and M. Elamurugaselvam, "Design based object oriented Metrics to measure coupling & cohesion", *International journal of Advanced & Innovative Research*, Vol. 2, Issue 5, PP. 778–785, 2013.
8. Teerawat Issariyakul • Ekram Hoss, "Introduction to Network Simulator NS2".
9. S. Sathish Raja and K.G.S. Venkatesan, "Email spam zombies scrutinizer in email sending network Infrastructures", *International journal of Scientific & Engineering Research*, Vol. 4, Issue 4, PP. 366–373, April 2013.
10. G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE J. Sel. Areas Communication.*, Vol. 18, No. 3, PP. 535–547, Mar. 2000.
11. K.G.S. Venkatesan, "Comparison of CDMA & GSM Mobile Technology", *Middle-East Journal of Scientific Research*, 13 (12), PP. 1590–1594, 2013.
12. P. Indira Priya, K.G.S. Venkatesan, "Finding the K-Edge connectivity in MANET using DLTRT, *International Journal of Applied Engineering Research*, Vol. 9, Issue 22, PP. 5898–5904, 2014.
13. K.G.S. Venkatesan and M. Elamurugaselvam, "Using the conceptual cohesion of classes for fault prediction in object-oriented system", *International journal of Advanced & Innovative Research*, Vol. 2, Issue 4, PP. 75–80, April 2013.
14. Ms. J. Praveena, K.G.S. Venkatesan, "Advanced Auto Adaptive edge-detection algorithm for flame monitoring & fire image processing", *International Journal of Applied Engineering Research*, Vol. 9, Issue 22, PP. 5797–5802, 2014.
15. K.G.S. Venkatesan. Dr. V. Khanna, "Inclusion of flow management for Automatic & dynamic route discovery system by ARS", *International Journal of Advanced Research in computer science & software Engg.*, Vol. 2, Issue 12, PP. 1–9, December–2012.
16. Needhu. C, K.G.S. Venkatesan, "A System for Retrieving Information directly from online social network user Link", *International Journal of Applied Engineering Research*, Vol. 9, Issue 22, PP. 6023–6028, 2014.
17. K.G.S. Venkatesan, R. Resmi, R. Remya, "Anonymizing Geographic routing for preserving location privacy using unlinkability and unobservability", *International Journal of Advanced Research in computer science & software Engg.*, Vol. 4, Issue 3, PP. 523–528, March–2014.
18. Selvakumari. P, K.G.S. Venkatesan, "Vehicular communication using Fvmr Technique", *International Journal of Applied Engineering Research*, Vol. 9, Issue 22, PP. 6133–6139, 2014.
19. K.G.S. Venkatesan, G. Julin Leeya, G. Dayalin Leena, "Efficient colour image watermarking using factor Entrenching method", *International Journal of Advanced Research in computer science & software Engg.*, Vol. 4, Issue 3, PP. 529–538, March–2014.
20. K.G.S. Venkatesan. Kausik Mondal, Abhishek Kumar, "Enhancement of social network security by Third party application", *International Journal of Advanced Research in computer science & software Engg.*, Vol. 3, Issue 3, PP. 230–237, March–2013.
21. Annapurna Vemparala, Venkatesan.K.G., "Routing Misbehavior detection in MANET'S using an ACK based scheme", *International Journal of Advanced & Innovative Research*, Vol. 2, Issue 5, PP. 261–268, 2013.
22. K.G.S. Venkatesan. Kishore, Mukthar Hussain, "SAT : A Security Architecture in wireless mesh networks", *International Journal of Advanced Research in computer science & software Engineering*, Vol. 3, Issue 3, PP. 325–331, April–2013.
23. Annapurna Vemparala, Venkatesan.K.G., "A Reputation based scheme for routing misbehavior detection in MANET'S", *International Journal of computer science & Management Research*, Vol. 2, Issue 6, June - 2013.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

24. K.G.S. Venkatesan, "Planning in FARS by dynamic multipath reconfiguration system failure recovery in wireless mesh network", International Journal of Innovative Research in computer & comm. Engineering, Vol. 2, Issue 8, August -2014.
25. R. Ghosh, K. Trivedi, V. Naik, and D. S. Kim, "End-to-end performability analysis for infrastructure-as-a-service cloud: An interacting stochastic models approach," in *Dependable Computing(PRDC), 2010 IEEE 16th Pacific Rim International Symposium on*, PP. 125 –132, December – 2010.
26. K.G.S. Venkatesan, AR. Arunachalam, S. Vijayalakshmi, V. Vinotha, "Implementation of optimized cost, Load & service monitoring for grid computing", International Journal of Innovative Research in computer & comm. Engineering, Vol. 3, Issue 2, PP. 864 – 870, February -2015.
27. R. Karthikeyan, K.G.S. Venkatesan, M.L. Ambikha, S. Asha, "Assist Autism spectrum, Data Acquisition method using Spatio-temporal Model", International Journal of Innovative Research in computer & communication Engineering, Vol. 3, Issue 2, PP. 871 – 877, February -2015.
28. K.G.S. Venkatesan, B. Sundar Raj, V. Keerthiga, M. Aishwarya, "Transmission of data between sensors by devolved Recognition", International Journal of Innovative Research in computer & comm. Engineering, Vol. 3, Issue 2, PP. 878 – 886, February -2015.
29. K.G.S. Venkatesan, N.G. Vijitha, R. Karthikeyan, "Secure data transaction in Multi cloud using Two-phase validation", International Journal of Innovative Research in computer & comm. Engineering, Vol. 3, Issue 2, PP. 845 – 853, February -2015.
30. K.G.S. Venkatesan, "Automatic Detection and control of Malware spread in decentralized peer to peer network", International Journal of Innovative Research in computer & comm. Engineering, Vol. 1, Issue 7, PP. 15157 – 15159, September -2013.
31. Sathish Raja, S K.G.S. Venkatesan, "Electronic Mail spam zombies purify in email connection", International Journal of Advanced Research in Computer Science Engineering & Information Technology, Vol. 1, Issue 1, PP. 26 – 36, June – 2013.
32. K.G.S. Venkatesan. Dr. V. Khanna, S.B. Amarnath Reddy, "Providing Security for social Networks from Inference Attack", International Journal of Computer Science Engineering & Scientific Technology, March –2015.
33. K.G.S. Venkatesan, Dr. Kathir. Viswalingam, N.G. Vijitha, " Associate Adaptable Transactions Information store in the cloud using Distributed storage and meta data manager", International Journal of Innovative Research in computer & communication Engineering, Vol. 3, Issue 3, PP. 1548 – 1555, March -2015.
34. K.G.S. Venkatesan, Dr. V. Khanna, Jay Prakash Thakur, Banbari Kumar, "Mining User profile Exploitation cluster from computer program Logs", International Journal of Innovative Research in computer & communication Engineering, Vol. 3, Issue 3, PP. 1557 – 1561, March -2015.
35. Ms.J.Praveena, K.G.S.Venkatesan, "Advanced Auto Adaptive edge-detection algorithm for flame monitoring & fire image processing", International Journal of Applied Engineering Research, Vol. 9, Issue 22, PP. 5797 – 5802, 2014.
36. K.G.S.Venkatesan, "Planning in FARS by dynamic multipath reconfiguration system failure recovery in wireless mesh network", International Journal of Innovative Research in computer & comm. Engineering, Vol. 2, Issue 8, August -2014.
37. 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.
38. 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.
39. H. Khazaei, J. Mistic, and V. Mistic, "Performance analysis of cloud computing centers using m/g/m/m+r queuing systems," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 23, no. 5, pp. 936–943, may 2012.