International Journal of Engineering Research and Development
e-ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com
Volume 10, Issue 3 (March 2014), PP.59-62

CRAM: Efficient Dynamic Resource Scheduling in Cloud Computing

M.Gomathy[1]
[1]Shrimati Indra Gandhi College, Lecturer, Department of Computer Science, Trichy, Tamilnadu, India.

Abstract: - Cloud computing has emerged as the default paradigm for a variety of fields especially considering the resources and infrastructures consumption in case of distributed access. The solution has however placed a lot of emphasis of a cloud server with variety of demands of which quality of service reminds a paramount strategy. There are a lot of strategies is in place for this quality cost which is regulated by service level agreement [SLA].An SLA is an agreement between client and server which when violated will impose penalties for the infringement or violation performance evaluation place a key role allowing system managers to evaluate the effects of different resource management strategies on the data center functioning and to predict the corresponding cost/benefits. In order to deal with very large systems composed of hundreds or thousands of resources. The system should allow to easily implementing different strategies and should have policies to represent different working conditions. Keeping this is mind a rewards scheme called as SRNS-Stochastic Reward Nets is utilized which is dynamic in nature to the status of the requests made and job allocated. The proposed model is scalable enough to represent system 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. The resources offered by other public cloud system through a sharing and paying model is also controlled in term of performance metrics like utilization, responsiveness and load burst. This model is analytical in nature unlike the existing approaches which are assumptive and simulation based in nature.

Keywords:- *Cloud Computing, Stochastic Reward Nets, Cloud-Oriented Performance Metrics, resiliency, responsiveness.*

I. INTRODUCTION

Cloud computing is a computing paradigm in which different computing resources such as infrastructure, platforms and software applications are made accessible over the internet to remote user as services.

Infrastructure-as-a-Service (IaaS) clouds are becoming a rich and active branch of commercial services. Uses of IaaS clouds can provision “processing, storage, networks and other fundamental resources” on demand, that is, when needed, for as long as needed, and paying only for what is actually consumed. However, the increased adaption of clouds and perhaps even the pricing models depend on the ability of (perspective) cloud users to benchmark and compare commercial cloud services.

.

The typical performance evaluation approaches in quality of service such as simulation or on-the-field measurements cannot be easily adapted. Simulation does not allow conducting comprehensive analyses of the system performance due to the great number of parameters that have to be investigated. In order to implement particular resource management techniques such as VM multiplexing or VM live migration that, even if transparent to final user, has to be considered in the design of performance models in order to accurately understand the system behavior.

Finally, different clouds, belonging to the same or to different organizations, can dynamically join each other to achieve a common goal, usually represented by the optimization of resources utilization. This mechanism, referred to as cloud federation, allows providing and releasing resources on demand thus providing elastic capabilities to the whole infrastructure. All the parameters do not or may not conform to real time situations or exigencies and are thus only reflective in nature.

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

II. RELATED WORK

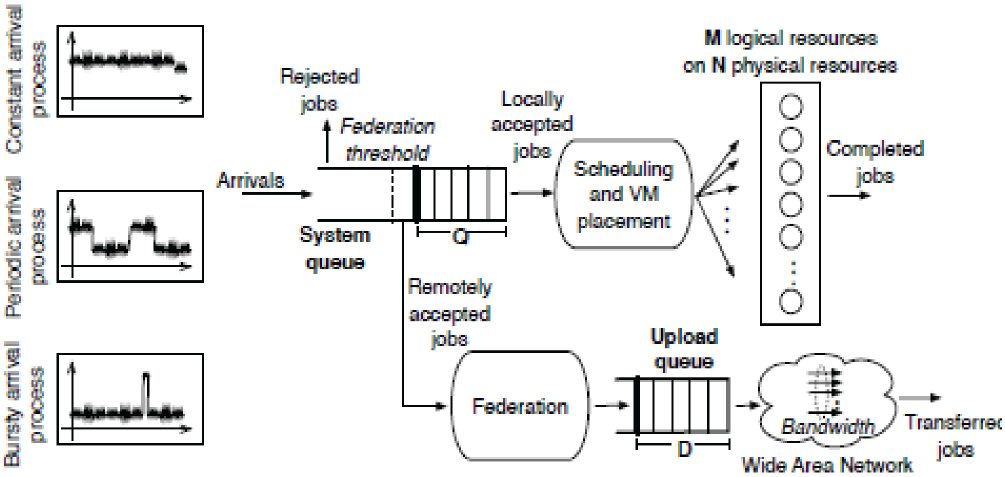
Cloud systems differ from traditional distributed systems. First of all, they are characterized by a very large number of resources that can span different administrative domains. Moreover, the high level of resource abstraction such as VM multiplexing [2] or VM live migration [3] that, even if transparent to accurately understand the system behavior. Finally, different clouds, belonging to the same or to different organizations, can dynamically join each other to achieve a common goal, usually represented by the optimization of Resources utilization. This mechanism, referred to as cloud federation, allows providing and releasing resources on demand thus providing elastic capabilities to the whole infrastructure.

For these reasons, typical performance evaluation approaches such as simulation or on-the-field measurements cannot be easily adopted. Simulation [5],[6] does not allow to conduct comprehensive analyses of the system performance due to the great number of parameters that have to be investigated. On-the-field experiments [7],[8] are mainly focused on the offered QoS, they are based on a black box approach that makes difficult or correlate obtained data to the internal resource management strategies implemented by the system provider. On the contrary, analytical techniques represent a good candidate thanks to the limited solution cost of their associated models.

III. PROPOSED MODEL

The proposed Model presents a stochastic model, based on Stochastic Reward Nets (SRNs), that exhibits the above mentioned features allowing capturing 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.

Cloud based systems are inherently large scale, distributed, almost always virtualized, and operate in automated shared environments. Performance and availability of such systems are affected by a large number o parameters including characteristics of the physical infrastructure (e.g., number of servers, number of cores per server, amount of RAM and local storage per server, configuration of physical servers, network configuration, persistent storage configuration), characteristics of the virtualization infrastructure (e.g., VM placement and VM resource allocation, deployment and runtime overheads),failure characteristics (e.g., failure rates, repair rates, modes of recovery),characteristics of automation tools used tom manage the cloud system, and so on. Because of this, any naïve modeling approach will quickly run into state explosion and/or intractable solution.



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 investigating 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).

3.1 Cloud Model and Resources

Clouds are modeled in the server client mode by a datacenter component for handling service requests. These requests are application elements sandboxed within VMs, which need to be allocated a share of processing power on Datacenter's host components. By VM processing, it means that a set of operations related to VM life cycle: provisioning of a host to a VM, VM creation, VM destruction, and VM migration. A Datacenter is composed by a set host, which are responsible for managing VMs during their life cycles. Host is a component that represents a physical computing node in a cloud: it is assigned a pre-configured processing capability (expressed in million of instruction per second-MIPS), memory, storage, and a scheduling policy for allocating processing cores to virtual machines. The Host component implements interfaces that support modeling and simulation of both single-core and multi-core nodes.

3.2 Client Requests

Clients are registered with the server for utilizing the resources and application. Needs to access the resources and applications are implemented and customary SLA's are put in place to handle the request handling mechanisms.

Each Request is considered as a job to be completed. A job is usage of a app or a resource or both and sending back the processed requested data. The job arrival or rather the request process constituter three different scenarios. In the first one (Constant arrival process) the arrival process be a homogeneous Poisson process with uniform rate.

However, in 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 and for these reasons, in order to take into account the dependencies of the job arrival rate on both the days of a week, and the hours of a day, in the second scenario is the periodic arrival process. This is also chosen to model the job arrival process as Markov Modulated Poisson Process (MMPP).

3.3 Cloud Federation and Monitoring

This is the analytical algorithm part, where parameters like weight; intermediate requests like waiting time, bandwidth calculation completion time are all executed here. This is based on the inputs the optimal solution is arrived federation with other clouds is modeled allowing tokens in place p_{queue} to be moved, through transition t_{upload}, in the upload queue represented by place p_{send}. In accordance with the assumptions made before, transition t_{upload} is enabled only if the number of tokens in place p_{queue} is greater than Q and the number of tokens in place P_{send} is less than D. Moreover, in order to take into account the federated cloud availability, concurrent enabled transition t_{upload} and t_{drop} are managed by setting their weights.

3.4 Virtual Multiplexing

VM, whose management during its life cycle is the responsibility of the host component. As discussed earlier, a host can simultaneously instantiate multiple VMs and allocate cores based on predefined processor sharing policies(space-shared, time-shared).Every VM component has access to a component that stores the characteristics related to a VM's internal scheduling policy, which is extended from the abstract component called VM Scheduling.

The model supports VM scheduling at two levels: First, at the host level and second, at the VM level. At the host level, it is possible to specify how much of the overall processing power of each core in a host will be assigned to each VM. At the VM level, the VM assign specific amount of the available processing power to the individual task units that are hosted within its execution engine.

IV. CONCLUSION

SRNs allow us to define reward functions that can be associated to a particular state of the model in order to evaluate the performance level reached by the system during the sojourn in order in that state performance metrics able to characterize the system behavior from both the provider and the user point-of-views. the 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 and all the analysis are done by autonomic techniques able to change on-the fly the system configuration in order to react to change on the working conditions. This can be extended to models which represent PaaS and SaaS Cloud Systems and to integrate the mechanisms needed to capture VM migration and the data center consolidation aspects that cover a crucial role in energy saving policies.

In future the model can be implemented to include intra and reservoir clouds where other performance parameters like application takes sufficient memory and bandwidth issues are considered. This model may be replicated in web services domain where similar features are used and service agreements are made between the website client and web service provider so the concept can be extended to cover such models and issues as well in the future.

REFERENCES

[1]. R. Buyya et al., "Cloud computing an 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]. Amazon Web Services (AWS), Online at http://aws. amazon.com.

[4]. Microsoft Azure, http://www.microsoft.com/azure/.

[5]. Google App Engine, Online at http://code.google.com/appengine/.

[6]. Brock, M.; Goscinski, A.; Grids vs. Clouds Future Information Technology (FutureTech), 2010 5th IEEE International Conference2010 pp 1-6.

[7]. 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.

[8]. Daryl C. Plummer, Thomas J. Bittman, Tom Austin, David W. Cearley, David Mitchell Smith "Cloud Computing: Defining and Describing an Emerging Phenomenon".

[9]. High-Performance Cloud Computing: A View of ScientificApplications by Christian Vecchiola, Suraj Pandey, and Rajkumar Buyya.

[10]. B. Rochwerger et al., "Reservoir – when one cloud is not enough," Computer, vol.44 , no.3, pp.44-55,march 2011.

[11]. Market-Oriented Cloud Computing: Vision, Hype, and Reality for Delivering IT Services as Computing Utilities Buyya, R. Chee Shin Yeo Venugopal, S.

[12]. Performance Evaluation of Cloud Computing Offerings Vladimir Stantchev, SOA and Public Services Research Group, TU Berlin.

[13]. Schaper, J Cloud Services Digital Ecosystems and Technologies (DEST), 2010 4th IEEE International Conference 2010 pp 91-92.

[14]. R. Buyya, R. Ranjan, and R. Calheiros," Modeling and simulation of scalable cloud computing environments and cloudsim toolkit: Challenges and opportunities," in High performance Computing Simulation, 2009.HPCS'09. International Conference on, June 2009.pp.1-11.

[15]. Srinivasa, K.G.; Siddesh, G.M.; Cherian, S.;" Fault-Tolerant Middleware for Grid Computing "Srinivasa, K.G. Siddesh, G.M Cherian, S.High Performance Computing and Communications (HPCC), 2010. 12th IEEE International Conference pp 635 – 640.

[16]. Xingchen Chu Nadiminti, K. Chao Jin Venugopal, S.Buyya, R, Univ. of Melbourne, Melbourne "Aneka: Next- Generation Enterprise Grid Platform for e-Science and e- Business Applications".

[17]. 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.

[18]. Zhao, Peng; Huang, Ting-lei; Liu, Cai-xia; Wang, Xin; Research of P2P architecture based on cloud computing IEEE International Conference.

[19]. http://www.vmware.com/files/pdf/techpaper/VMW-TWP-vSPHR-SECRTY-HRDNG-USLET-101-WEB-1.pdf.

[20]. http://www.vmware.com/files/pdf/vc_dbviews_40.pdf.