Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 5 Issue 02, February-2016

A Survey on Energy Aware Resource Allocation for Cloud Computing

Pragya Matre
Student M.E. CSE
UIT RGPV Bhopal, India

Dr. Sanjay Silakari
HOD CSE
UIT RGPV Bhopal, India

Uday Chourasia
Assistant professor CSE
UIT RGPV Bhopal, India

***Abstract* - Energy is been a critical issue in cloud computing. In this paper a survey is done on energy aware resource allocation in cloud. As resource allocation in cloud is a NP-Hard problem; so several exact solutions and approximate solutions are proposed. As in resource allocation approximate solutions are also useful so in future such solutions can be proposed.**

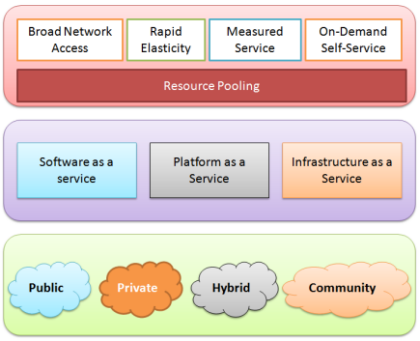
Keywords: NP-Hard, heuristic algorithms, energy in cloud.

I INTRODUCTION

Cloud Computing is a term used to describe both a platform and type of application. As a platform it supplies, configures and reconfigures servers, while the servers can be physical machines or virtual machines. On the other hand, Cloud Computing describes applications that are extended to be accessible through the internet and for this purpose large data centers and powerful servers are used to host the web applications and web services.

The cloud is a metaphor for the Internet and is an abstraction for the complex infrastructure it conceals. There are some important points in the definition to be discussed regarding Cloud Computing. Cloud Computing differs from traditional computing paradigms as it is scalable, can be encapsulated as an abstract entity which provides different level of services to the clients, driven by economies of scale and the services are dynamically configurable.

Cloud Computing refers to both the applications delivered as services over the Internet and the hardware and systems software in the datacenters that provide those services. The services themselves have long been referred to as Software as a Service (SaaS). The datacenter hardware and software is what we will call a Cloud. When a Cloud is made available in a pay-as-you-go manner to the general public, we call it a Public Cloud; the service being sold is Utility Computing. We use the term Private Cloud to refer to internal datacenters of a business or other organization, not made available to the general public. Thus, Cloud Computing is the sum of SaaS and Utility Computing, but does not include Private Clouds. People can be users or providers of SaaS, or users or providers of Utility Computing.



II ENABLING TECHNOLOGIES

Before going into the idea of cloud computing, two technologies will be introduced that made the way of distributed computing and therefore cloud computing realizable.

Virtualization

With virtualization, applications and infrastructure are independent, allowing servers to be easily shared by many applications where applications are running virtually anywhere in the world. This is possible as long as the application is virtualized. Virtualizing the application for the cloud means to package the bits of the application with everything it needs to run, including pieces such as a database, a middleware and an operating system. This self-contained unit of virtualized application can then run anywhere in the world. Virtualization also allows building so-called sandboxes. Sandboxes assure a higher degree of security and reliability by providing a mechanism to run programs safely. It is commonly used to "execute untested code or programs from unverified third-parties, suppliers and untrusted users.

Load Balancing

Load balancing is the key to success for cloud architectures. It is capable of distributing the working processes evenly between 2 or more computers, so that resources can be used efficiently and therefore increases performance and availability. A so-called load balancer is automatically able to deal with different amount of work capacity by adapting its distribution decisions according to the moments a request is made. A load balancing solution is often used in internet services, where the idea of load balancing is run by an application.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 5 Issue 02, February-2016

III LITERATURE REVIEW

A stochastic integer programming approach for the minimization of the total energy consumption of cloud servers subject to different running modes is presented in [26] where authors incorporate backlogging and penalty cost for unfinished requests and demonstrate that allowing servers to consolidate requests to later periods allows to reduce the total consumed energy. A practical technique for predicting performance interference due to shared processor caches among VMs consolidated on the same PM is presented in [27]. The results of these works could be used to complement our model integrating new aspects and improving our placement decisions for give performance and cost objectives.

Other interesting works, mainly focusing on performance and energy consumption evaluation of cloud systems, are the following. In [22], Sarji et al. propose two different models for the energy consumption of a cloud server when switching it to operational mode from a sleep mode or from an off state. The two models differ in terms of energy consumption and required time. They also propose an energy saving engine in the cloud provider able to decide when to trigger VMs migration and PMs state switching. In [23], Wu and Zhao provided performance models for live migration which can be used to predict the VM migration time given the application behavior and the resources available for migration. In [24], [25], the authors provide analytical models to investigate the more convenient strategies to manage a federation of two or more private or public IaaS clouds taking into account energy consumption and VM migration. Such works can be considered as a preliminary study on energy aware models of cloud systems where energy aspects have been treated considering the data center as a whole. In fact, while in [10], [19] authors model a cloud broker able to balance the workload among three server pools in different running status, the possibility to dynamically turn on and off the physical resources according to the load and the system status is not taken into consideration, as done in our work. Moreover, no resource allocation policies are taken into consideration within the same pool or among different pools. Nevertheless, the interacting model approach is very interesting and it could be useful to extend our model to more complex scenarios, for example trying to separate the arrival process, the virtual layer, and the physical layer portions of our SRN. In [9] the author presents an SRN model able to capture some performance indices related to an IaaS cloud. Cloud-specific concepts such as infrastructure elasticity are analyzed also providing an exhaustive set of performance metrics regarding both the system provider (e.g., utilization) and the final users (e.g., responsiveness).

Broad categories of approaches to reduce energy consumption

APPROACH NAME	DESCRIPTION	DRAWBACKS	REFERENCE
Efficient VM Migration for reducing energy consumption	In this approach different algorithms are proposed by various authors for reducing number of VM Migrations to save energy.	One of the drawbacks of this technique is VM migrations are necessary in case of load balancing and thus can't be reduced after a certain level.	[6] [23]
VM consolidation Approach	In this approach those physical machines on which load is below certain level, such PMs are switched to sleep mode and VMs running on those PMs are migrated to other PMs.	The drawback of this approach is this will result in increased VM Migrations which will ultimately consume energy.	[12] [17]
Efficient Resource Allocation	This approach is proactive in nature and thus it proves better in most of the cases.	But this approach posses considerable overhead over hypervisor.	[11] [3]

IV CONCLUSION

In this paper a survey is done on energy aware resource allocation algorithms for cloud. As in cloud there are number of physical machines and large number of VMs are allocated to these physical machines. So resource allocation is a challenging task. In this paper several energy aware algorithms are studied. And it is found that approximate algorithms have proven to be better as compared to other energy models.

V REFERENCES:

[1] D. Bruneo, "A stochastic model to investigate data center performance and QoS in IaaS cloud computing systems," IEEE Trans. Parallel Distrib. Syst., vol. 25, no. 3, pp. 560–569, Mar. 2014.
[2] R. Ghosh, F. Longo, V. K. Naik, and K. S. Trivedi, "Modeling and performance analysis of large scale IaaS clouds," Future Generation Comput. Syst., vol. 29, no. 5, pp. 1216–1234, 2013.
[3] D. Bruneo, A. Lhoas, F. Longo, and A. Puliafito, "Analytical evaluation of resource allocation policies in green IaaS clouds," in Proc. 3rd Int. Conf. Cloud Green Comput., Sep. 2013, pp. 84–91.
[4] D. Gupta, L. Cherkasova, R. Gardner, and A. Vahdat, "Enforcing performance isolation across virtual machines in Xen," in Proc ACM/IFIP/USENIX Int. Conf. Middleware, 2006, pp. 342–362.
[5] M. Armbrust, A. Fox, R. Griffith, A. D Joseph, R. Katz, A Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M Zaharia, "A view of cloud computing," Commun. ACM vol. 53, no. 4, pp. 50–58, Apr. 2010.
[6] M. Mishra and A. Sahoo, "On theory of VM placement: Anomalie in existing methodologies and their mitigation using a novel vecto based approach," in Proc. IEEE Int. Conf. Cloud Comput., 2011 pp. 275–282.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 5 Issue 02, February-2016

[7] H. Liu, H. Jin, X. Liao, C. Yu, and C.-Z. Xu, "Live virtual machin migration via asynchronous replication and state synchronization, IEEE Trans. Parallel Distrib. Syst., vol. 22, no. 12, pp. 1986–1999, dec. 2011.

[8] A. Verma, G. Dasgupta, T. K. Nayak, P. De, and R. Kothari "Server workload analysis for power minimization using consolidation, in Proc. Conf. USENIX, 2009, pp. 28–28.

[9] S. Pacheco-Sanchez, G. Casale, B. Scotney, S. McClean, G. Parr, and S. Dawson, "Markovian workload characterization for Qo prediction in the cloud," in Proc. IEEE Int. Conf. Cloud Comput., 2011, pp. 147–154.

[10] K. S Trivedi, Probability and Statistics with Reliability, Queuing and Computer Science Applications, second edition. Hoboken, NJ, USA: Wiley, 2011.

[11] A. Beloglazov, J. Abawajy, and R. Buyya, "Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing," Future Gener. Comput. Syst., vol. 28, no. 5, pp. 755–768, May 2012.

[12] R. Buyya, R. Ranjan, and R. Calheiros, "Modeling and simulation of scalable cloud computing environments and the cloudsim toolkit: Challenges and opportunities," in Proc. High Perform. Comput. Simul., Int. Conf., Jun. 2009, pp. 1–11.

[13] J. H Kim, S. M Lee, D. S Kim, and J. S Park, "Performability analysis of IaaS cloud," in Proc. Innovative Mobile Internet Serv. Ubiquitous Comput., Int. Conf., 2011, pp. 36–43.

[14] A. Iosup, S. Ostermann, M. Yigitbasi, R. Prodan, T. Fahringer, and D. Epema, "Performance analysis of cloud computing services for many-tasks scientific computing," IEEE Trans. Parallel Distrib. Syst., vol. 22, no. 6, pp. 931–945, Jun. 2011.

[15] V. Stantchev, "Performance evaluation of cloud computing offerings," in Proc. 3rd Int. Conf. Adv. Eng. Comput. Appl. Sci., oct. 2009, pp. 187–192.

[16] H. Khazaei, J. Misic, and V. Misic, "A fine-grained performance model of cloud computing centers," IEEE Trans. Parallel Distrib. Syst., vol. 24, no. 11, pp. 2138–2147, Sep. 2013.

[17] A. V Do, J. Chen, C. Wang, Y. C Lee, A. Zomaya, and B. B Zhou, "Profiling applications for virtual machine placement in clouds," in Proc. IEEE Int. Conf. Cloud Comput., 2011, pp. 660–667.

[18] I. Mitrani, "Service center trade-offs between customer impatience and power consumption," Perform. Eval., vol. 68, pp. 1222–1231, Nov. 2011.

[19] 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 Proc. IEEE Int. Symp. Dependable Comput., 2010, pp. 125–132.

[20] R. Ghosh, V. K. Naik, and K. S. Trivedi, "Power-performance trade-offs in IaaS cloud: A scalable analytic approach," in Proc. 41st Int. Conf. Dependable Syst. Netw. Workshops, 2011, pp. 152–157.

[21] R. Ghosh, F. Longo, V. Naik, and K. Trivedi, "Quantifying resiliency of IaaS cloud," in Proc. 29th IEEE Symp. Reliable
. 492–499. Distrib. Syst., 2010, pp. 343–347.

[22] I. Sarji, C. Ghali, A. Chehab, and A. Kayssi, "Cloudese: Energy efficiency model for cloud computing environments," in Proc. Int. Conf. Energy Aware Comput., 2011, pp. 1–6.

[23] Y. Wu and M. Zhao, "Performance modeling of virtual machine live migration," in Proc. IEEE Int. Conf. Cloud Comput., Jul. 2011, pp

[24] D. Bruneo, F. Longo, and A. Puliafito, "Evaluating energy consumption in a cloud infrastructure," in Proc. IEEE Int. Symp. World Wireless, Mobile Multimedia Netw., Jun. 2011, pp. 1–6.

[25] D. Bruneo, F. Longo, and A. Puliafito, "Modeling energy-aware cloud federations with SRNs," Trans. Petri Nets Other Models Concurrency, vol. 7400, pp. 277–307, 2012.

[26] J. Wang and S. Shen, "Risk and energy consumption tradeoffs in cloud computing service via stochastic optimization models," in Proc. IEEE 5th Int. Conf. Utility Cloud Comput., 2012, pp. 239–246.

[27] S. Govindan, J. Liu, A. Kansal, and A. Sivasubramaniam, "Cuanta: Quantifying effects of shared on-chip resource interference for consolidated virtual machines," in Proc. 2nd ACM Symp. Cloud Comput., 2011, pp. 22:1–22:14.

[28] Dario Bruneo, Audric Lhoas, Francesco Longo, and Antonio Puliafito "Modeling and Evaluation of Energy Policies in Green Clouds" in IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 26, NO. 11, NOVEMBER 2015.