모바일 클라우드 환경에서 오프로딩 프레임워크 리소스 스케줄링에 관한 연구

Misbah Liaqat[3], 손윤식[1], 오세만[1], 김순곤[2], 김성진[3], 고광만[3]
[1]동국대학교 컴퓨터공학과,
[2]중부대학교 컴퓨터게임학과,
[3]상지대학교 컴퓨터정보공학부
golpiler@gmail.com[3], sgkim@joongbu.ac.kr[2], kkman@sangji.ac.kr[3]

A Study on the Offloading Framework Resource Scheduling in Mobile Cloud Environments

Misbah Liaqat[3], Younsik Son[1], Seman Oh[1], Soongohn Kim[2], Seongjin[3], Kwangman Ko
[1]Dept. of Computer Engineering, Dongguk University
[2]Dept. of Computer Game Engineering, Joongbu University
[3]Dept. of Computer Engineering, Sangji University

Abstract

Virtualization was devised as a resource management and optimization technique for mainframes having scaleless computing capabilities. The resource scaling can be done with a variety of virtualization methods such as VM creation, deletion, and migration. In this paper, we designed to achieve the load balancing, several load balancing schemes such as Minimum Execution Time (MET), Min-Min scheduling, Cloud Analyst have been reported in literature in addition to a comprehensive study on First Come First Serve (FCFS) and Round-robin schedulers.

1. Introduction

Compared to the advantages of virtualization, the load of virtualization and the complexity of utilization of resource provided by physical server are increasing. Virtualization technology is demanding monitoring and performance evaluation of resources due to application performance characteristics and performance influences.

2. Related Works

OpenStack is an open source cloud platform for building private and public cloud computing environments developed by NASA and RackSpace. It is composed of Nova controlling virtual machine, Glance managing virtual disk image, Swift storing data, Keystone, and Horizon, which makes open stacks easier to manage application integration[1].

Open stacks are modular in many components, and message-based architecture allows each component to be easily scalable because of its low coupling. These components provide various functions of computation, storage and network virtualization, And convenient development and management tools, and it features low maintenance costs. Load balancing is a technique to control all physical resources in a cloud data center to have a constant load. Round Robin[2], Min-Min scheduling[3,4,5], Max-Min algorithm[6], open stack scheduler[7], Min-min Algorithm[8] The study was cited as a static load balancing method for a single resource by placing it with a virtual machine.

3. Resource Scheduling Technique on Offloading Framework

Virtualization was devised as a resource management and optimization technique for mainframes having scaleless computing capabilities. Virtualization in mainframes results in efficient management of coarse-grained resources with limited overhead. In a virtualized data center architecture, multiple clients often share same hardware resources with the help of virtualization techniques. Moreover, hardware resources provisioned for a data center client can be scaled dynamically according to varying workload. The resource scaling can be done with a variety of virtualization methods such as VM creation, deletion, and migration. A workload can be consolidated or migrated onto a lesser number of resources using VM migration. The resultant resource set provides energy efficiency and higher resource utilization. The CPU

power along with other computing resources such as memory and I/O can be scaled gracefully with the help of virtualization technologies. When a hardware resource is underutilized due to lesser client requests, it represents an opportunity for resource consolidation. The workload of underutilized hardware is transferred to another suitable hardware with the help of hypervisor. The workload consolidation and migration technique is depicted in Figure 1.

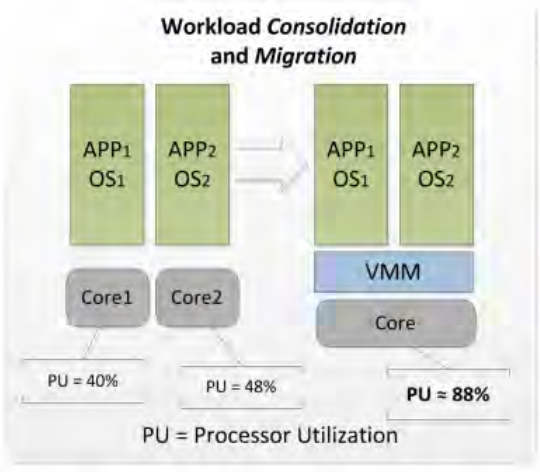

Figure 1. CPU Utilization Monitoring and Scheduling

Load balancing, a deployed function, plays its vital role in cloud and cloud data center domains for efficient resource management. Load balancing ensures even distribution of resources among a set of users in a uniform way such that underlying servers do not become overloaded and idle at any time within cloud operation time line. Overlooking load balancing establishment abruptly decreases system throughput due to overloaded servers and ultimately leads to SLA violation. It has become an integral part of all distributed internet based systems as distributed computing comes with the challenges of high resource demands that overload servers. Load balancer increases the capacity and reliability of applications by decreasing the burden on a server. Load balancer starts with identification of hot spot, an overloaded server, and start migrating its load on a server which has sufficient resources such that the resources are evenly distributed. However, the criterion of where, which, and how to migrate workloads from the physical servers pose challenges that cloud operator has to consider during all these decision makings. The VectorDot scheme has considered the current load on the communication paths connecting physical servers and network attached storage[9]. Furthermore, VectorDot has addressed the overloaded servers, switches, and storage entities while meeting the desired objective function. Moreover, using constraint programming paradigm, tasks are migrated within nodes located in a cluster and has proved that consolidation overhead is indomitable while choosing a new configuration and also it is affected from the total migration time with that configuration[10]. Furthermore, employed Entropy has significantly reduced total VM migration duration in addition to the total number of nodes acquiring low performance overhead. Consequently, the authors of [11] has accurately projected the total migration cost in order to have an accurate estimation guess of migration time, so that sufficient resource can be prepared and reserved on the basis of VMs count and the performance degradation period instigated by VM migration, that is higher than actual total migration duration. Moreover, the proposed scheme has also presented the migration cost based on the migrating VM configuration and size. However, in order to balance the load based on migration techniques number of migration should be controlled because it effects the performance of other VMs running on source and destination hosts. Therefore, there is a need to balance the load based on the aforementioned research gap.

4. Conclusion

In this paper, to achieve the load balancing, several load balancing schemes such as Minimum Execution Time (MET), Min-Min scheduling, Cloud Analyst have been reported in literature in addition to a comprehensive study on First Come First Serve (FCFS) and Round-robin schedulers.

References

[1] Ben Cheng; Bill Buzbee, "A JIT Compiler for Android's Dalvik VM" Google. Retrieved, 2015.

[2] Ken Pepple, Deploying Open Stack, 2011.

[3] A. K. Sidhu, S. Kinger, "Analysis of load balancing techniques in cloud computing," International Journal of Computers & Technology, Vol. 4, Issue 2, pp. 737 - 741. 2013.

[4] H. Chen, F. Wang, N. Helian, G. Akanmu, "User-priority guided min-min scheduling algorithm for load balancing in cloud computing," in Parallel Computing Technologies (PARCOMPTECH), 2013

National Conference on IEEE, pp. 1 - 8, 2013
[5] S.-C. Wang, K.-Q. Yan, W.-P. Liao, S.-S. Wang, "Towards a load balancing in a three-level cloud computing network," in Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on, Vol. 1, IEEE pp. 108 - 113, 2010
[6] K. Etminani, M. Naghibzadeh, "A min-min max-min selective algorithm for grid task scheduling," in Internet, 2007. ICI 2007. 3rd IEEE/IFIP International Conference in Central Asia on, IEEE, pp. 1 - 7, 2007.
[7] O. Elzeki, M. Reshad, M. Elsoud, "Improved max-min algorithm in cloud computing," International Journal of Computer Applications, Vol. 50, No. 12, pp.22-27, 2012.
[8] O. Litvinski, A. Gherbi, "Openstack scheduler evaluation using design of experiment approach," in: Object/Component/Service-Oriented Real-Time Distributed Computing (ISORC), 2013 IEEE 16th International Symposium on, IEEE, pp. 1 - 7, 2013.
[9] G. Patel, R. Mehta, U. Bhoi, "Enhanced load balanced min-min algorithm for static meta task scheduling in cloud computing," Procedia Computer Science 57, pp. 545 - 553, 2015.
[10] 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. IEEE, 2011, pp. 275 - 282.
[11] F. Hermenier, X. Lorca, J.-M. Menaud, G. Muller, and J. Lawall, "Entropy: a consolidation manager for clusters," in Proceedings of the 2009 ACM SIGPLAN/SIGOPS international conference on Virtual execution environments. ACM, 2009, pp. 41 - 50.
[12] M. Zhao and R. J. Figueiredo, "Experimental study of virtual machine migration in support of reservation of cluster resources," in Proceedings of the 2nd international workshop on Virtualization technology in distributed computing. ACM, 2007, p. 5. [40] D. Chernicoff, The Shortcut Guide to Data Center Energy Efficiency . Realtime Publisher, 2009.