REDUCING ENERGY CONSUMPTION AND INDEPENDENT TASK ORIENTED SCHEDULING APPROACH FOR VIRTUALIZED CLOUDS

G.Tharanya*, G.Vijayasekaran, R.Rajakumar*****

*PG Student, Sir Issac Newton College of Engineering & Technology,Nagapattinam,tharanyacse.90@gmail.com
**Assistant Professor, Sir Isaac Newton College of Engineering and Technology,Nagapattinam,gunavijay87@yahoo.co.in
***Vice-Principal, Annai College of polytechnic,Kumbakonam,raja2012mtech@gmail.com

Abstract— **Reducing energy consumption is a major design constraint for modern heterogeneous computing systems to minimize electricity cost, improve system reliability and protect environment. Conventional energy-efficient scheduling strategies developed on these systems do not sufficiently exploit the system elasticity and adaptability for maximum energy savings, and do not simultaneously take account of user expected finish time. In this paper, we develop a novel scheduling strategy rolling-horizon scheduling architecture for real-time task scheduling in virtualized clouds. Then a task-oriented energy consumption model is given and analyzed. Based on our scheduling architecture, we develop a novel energy-aware scheduling algorithm named EARH for real-time, aperiodic, independent tasks. The EARH employs a rolling-horizon optimization policy and can also be extended to integrate other energy-aware scheduling algorithms. Furthermore, we propose two strategies in terms of resource scaling up and scaling down to make a good trade-off between task's schedulability and energy conservation. Extensive simulation experiments injecting random synthetic tasks as well as tasks following the last version of the Google cloud trace logs are conducted to validate the superiority of our EARH by comparing it with some baselines. The experimental results improve the scheduling quality and effectively enhance the system elasticity.**

Keywords — **Virtualized cloud, real-time, energy-aware, scheduling, rolling-horizon, elasticity.**

1 INTRODUCTION

Some public cloud applications and service provider environments use cluster-like architectures, with massively parallel workload and data distribution characteristics. Common cluster applications include data aggregation and "big data" analytics applications like Hadoop, Needlebase, Platform Computing Symphony software, or Vertica software. As a request comes in, a task scheduler spawns multiple jobs to multiple servers—causing a flurry of network traffic that does not go up to the network core but out to peer servers. Social networking sites use services like me cache for distributing memory objects to alleviate database load and speed up performance. Applications like Swift and services like Amazon's Simple Storage Service (S3) distribute storage across multiple nodes. Distributing storage, distributing databases across multiple servers, sending requests to multiple servers, and accumulating the responses are all E/W traffic intensive.

For example, consider how you plan a vacation. You visit the dynamic travel website of your choice and enter your variable data (when you want to travel, where, whether you need a hotel, flight, or a car).

The site pulls together the appropriate responses from multiple databases, along with related ads, and shows you the options within a matter of seconds. Not only is this process very heavy in E/W traffic flow because it pulls data from multiple servers, it is also latency sensitive. If a travel website cannot serve the data to you within a matter of seconds, you're likely to go to a competitor.

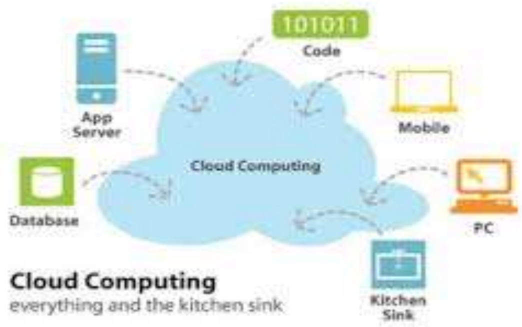

Figure 1.1.2.Cloud Sink

1.2 Research scope

The focus of this work is on energy-efficient resource management strategies that can be applied on a virtualized data center by a Cloud provider (e.g. Amazon EC2). The ability to migrate VMs between physical hosts with low overhead gives flexibility to a resource provider as VMs can be dynamically reallocated according to current resource requirements and the allocation policy. Idle physical nodes can be switched off to minimize energy consumption. In this paper we present a decentralized architecture of the resource management system for Cloud data centers and propose the development of the following policies for continuous optimization of VM placement:

- ❖ Optimization over multiple system resources at each time frame VMs are reallocated according to current CPU, RAM and network bandwidth utilization.
- ❖ Network optimization – optimization of virtual network topologies created by intercommunicating VMs. Network communication between VMs should be observed and considered in reallocation decisions in order to reduce data transfer overhead and network devices load.
- ❖ Thermal optimization – current temperature of physical nodes is considered in reallocation decisions. The aim is to avoid "hot spots" by reducing workload of the overheated nodes and thus decrease error-proneness and cooling system load.

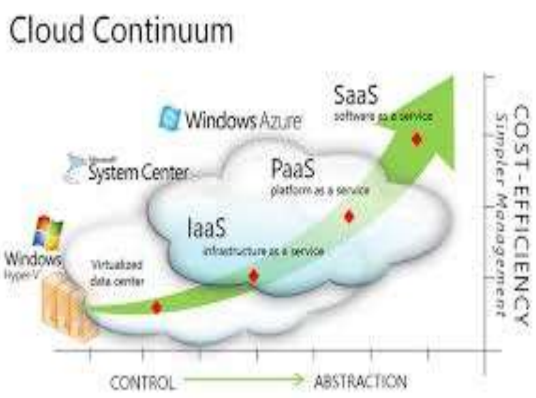

Figure 1.2.1 Cloud Services

1.3 Research challenges

The key challenges that have to be addressed are:

1) How to reduce functional costs while maintaining vital Quality of Service?
2) How to optimally solve the trade-off between energy savings and delivered performance?
3) How to determine when, which VMs, and minimize energy consumption by the system, while minimizing overhead?
4) How to develop efficient scheduling and scalable algorithms for resource allocation?
5) How to develop comprehensive solution by combing several allocation policies with different objectives?
6) How to save energy by dynamic VMs consolidation?

1.4 Cloud Service Provider

A cloud provider is a company that offers some component of cloud computing – typically Infrastructure as a Service (IaaS), Software as a Service (SaaS) or Platform as a Service (PaaS) – to other businesses or individuals. A service provider that offers customers storage or software services available via a private (private cloud) or public network (cloud)

Usually, it means the storage and software is available for access via the Internet. Cloud computing providers are notable entities who have verifiably significant production cloud computing service offerings. Cloud Service Providers offer Powered hybrid, private, and public cloud solutions with stronger service-level agreements (SLAs), better customer service, and proven expertise. Configurations range from on-premises, managed cloud to off-premises public cloud solutions. Cloud Service Providers deliver Powered services to help you deploy IT-as-a-service solutions faster and gain the financial advantage of greater efficiency, control, and choice.

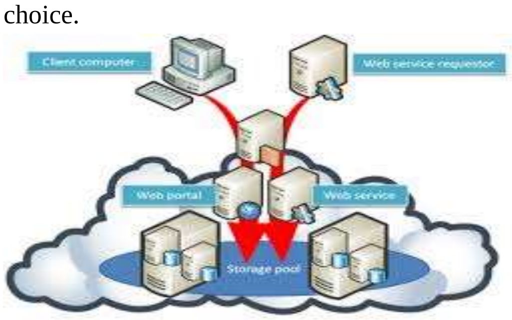

Figure 1.3.1 Cloud Storage Pool

1.5 Cloud Consumer

A cloud consumer is an organization that has a formal contract or arrangement with a cloud provider to use IT resources made available by the cloud provider. Specifically, the cloud consumer uses a cloud service consumer to access a cloud service. A cloud consumer is the principal stakeholder that uses the cloud computing services. A cloud consumer represents a person or organization that maintains a business relationship with, and uses the service from, a cloud provider. The end user that actually uses the service, whether it is Software, Platform or Infrastructure as a Service

Figure 1.5.1.Business with Clouds

1.6 Resource provisioning System

A Service Level Agreement is an important document that is used to define the level of a service that exists between a service provider and their customer. The agreement is generally in expressed in a simple language so that it can be clearly understood by the customer. The document may also include more technical terms for defining the service. The Service Level Agreement is often part of a wider service contract. A Service Level Agreement can either be an informal contract between parties or a legally binding contract. The SLA may address several areas including the availability of the service, the performance of the service, how it will operate, and priorities, responsibilities of involved parties, guarantees and warranties. As well as defining key areas, the Service Level Agreement may also specify a level of service, including targets and a minimum level that can be reached. Some of the common uses for a Service Level Agreement would be for telecom companies, IT service providers, Internet Service Providers (ISP) and outsourcing. In resource provisioning for cloud computing, an important issue is how resources may be allocated to an application mix such that the service level agreements (SLAs) of all applications are met. In this module, we implement scheduler, Real time controller and VM controller to allocate resources to VM and physical machines.

1.7 Scheduler

Scheduling refers to the set of policies to control the order of work to be performed by a computer system. There have been various types of scheduling algorithm existing in distributed computing system, and job scheduling is one of them. The main advantage of job scheduling algorithm is to achieve a high performance computing and the best system throughput. Scheduling manages availability of CPU memory and good scheduling policy gives maximum utilization of resource. We compared three algorithms Time Shared, Space shred and generalizes priority algorithm. The scheduler consists of a rolling-horizon, a real-time controller, and a VM controller. The scheduler takes tasks from users and allocates them to different VMs. The rolling-horizon holds both new tasks and waiting tasks to be executed.

1.8 Real time controller

The real-time controller determines whether a task in the rolling-horizon can be finished before its deadline. If not, the real-time controller informs the VM controller, and then the VM controller adds VMs to finish the task within timing constraint. If no schedule can be found to satisfy the task's timing requirement although enough VMs have been added, the task will be rejected. Otherwise, the task will be retained in the rolling-horizon.

1.9 VM controller

VMController is a general purpose open source and cross-platform virtual machine controller. At its current state, VirtualBox is the only supported hypervisor but we plan to support any other hypervisor with exposed APIs like VMWare, in future. With VMController you can write powerful scripts to control, manage, monitor, operate and administrate multiple virtual machines using powerful APIs via XMLRPC. The scheduling decision for the tasks in the rolling-horizon is updated their execution orders, start times, allocated VMs and new active hosts. If no schedule can be found to satisfy the task's timing requirement although enough VMs have been added, the task will be rejected.

1.10 EARH implementation

The idea to use a rolling horizon approach on supply chain optimization and planning problems has been applied in different contexts. However, as far as we know it has not been applied for ship scheduling and our type of application. Even though a rolling time horizon is usually used when uncertainties in data exist, a rolling time horizon is also applicable to reduce problem size. Where a rolling horizon is used to solve capacitated lot sizing problems, and they apply a rolling horizon method on a staff scheduling application in a starting phase, followed by an improvement phase where some time periods are selected and re-optimized. Both a forward and backward rolling horizon method for scheduling of multipurpose plants. A rolling horizon approach is used to optimize a hierarchical planning system. In this module, we implement rolling-horizon optimization and then integrate an energy-efficient scheduling algorithm. Based on these concepts we scaling up and scaling down VM to migrate tasks from one VM to another VM.

1.11 Evaluation criteria

Guarantee Ratio (GR) defined as: GR=Total count of tasks guaranteed to meet their deadlines/Total count of tasks; Total Energy Consumption (TEC) is: Total energy consumed by hosts. Energy Consumption per Task (ECT) calculated

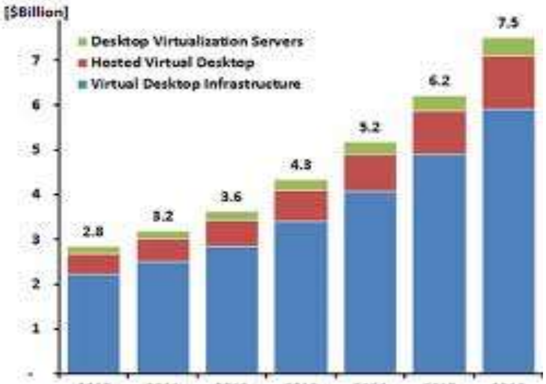

Figure 1.11.1 Evaluation

as: ECT= Total energy consumption/Accepted task count; In this module, we evaluate the performance of the system using Guarantee Ratio (GR), Total Energy Consumption (TEC), Energy Consumption per Task (ECT) and Resource Utilization (RU).

1.11 Contributions

Dynamic scheduling is that the task arrival is uncertain at run time and allocating resources are tedious as several tasks arrive at the same time. In case of dynamic scheduling information of the task components/task is not known before hand. Thus execution time of the task may not be known and the allocation of tasks is done on fly as the application executes Project focus on scheduling periodic and independent real-time tasks. Dynamic approach to create virtual clusters to deal with the conflict between parallel and serial Tasks. Tasks are dynamically available for scheduling over time by the scheduler. It is more flexible than static scheduling, to be able of determining run time in advance. It is more critical to include load balance as a main factor to obtain stable, accurate and efficient scheduler algorithm. A new energy-aware scheduling method that take into consideration the make span conservation and energy reduction. The scheduling is devised with measure to identify the degree of task computation efficiency relative to the application completion time. The degree is used as a utilization value to identify a level of virtue for executing task on processor, and implied an effectively energy consumption of that processor.In this approach, the task load is adjusted automatically without running time prediction. Implement EARH scheduling strategy to provide best tradeoff in task scheduling. Extend this work to analyze the deadlines and also get ACK for task completion. Dynamic scheduling algorithms for this scheduling mechanism have been introduced to generate scheduling with the shortest average execution time of tasks.

Conclusion

Firstly, we proposed a rolling-horizon scheduling architecture, and then a task oriented energy consumption model was analyzed and built. On this basis, we presented a novel energy-aware scheduling algorithm named EARH for real-time tasks, in which a rolling-horizon policy was used to enhance the system's schedulability. Additionally, the resource scaling up and resource scaling down strategies were developed and integrated into EARH, which can flexibly adjust the active hosts' scale so as to meet the tasks' real-time requirements and save energy. The EARH algorithm is the first of its kind reported in the literature; it comprehensively addresses the issue of real-time, schedulability, elasticity, and energy saving. To evaluate the effectiveness of our EARH, we conduct extensive simulation experiments to compare it with other algorithms. The experimental results indicate that EARH can efficiently improve the scheduling quality of others in different workloads and is suitable for energy aware scheduling in virtualized clouds.

References

[1] Adel Nadjaran Toosi, Rodrigo N. Calheiros, Rajkumar Buyya , "Interconnected Cloud Computing Environments: Challenges, Taxonomy, and Survey" Journal, ACM Computing Surveys (CSUR) , Volume 47 Issue 1, July 2014.

[2] Moreno, Ismael Solis , Yang, Renyu ; Xu, Jie ; Wo, Tianyu," Improved energy-efficiency in cloud datacenters with interference-aware virtual machine placement" Autonomous Decentralized Systems (ISADS), 2013 IEEE Eleventh International Symposium on march 2013.page NO:1-8

[3] Amir Vahid Dastjerdi , Sayed Gholam Hassan Tabatabaei, A dependency-aware ontology-based approach for deploying service level agreement monitoring services in Cloud, Journal Volume42,April2012 Pages 501-518.

[4] Josep Ll. Berral, Ricard Gavalda, Jordi Torres, Adaptive Scheduling on Power-Aware Managed Data-Centers Using Machine Learning, Proceeding GRID '11 Proceedings of the 2011 IEEE/ACM 12th International Conference on Grid Computing, Pages 66-73.

[5] Kumar, K. , Jing Feng ; Nimmagadda, Y. ; Yung-Hsiang Lu, Resource Allocation for Real-Time Tasks Using Cloud Computing, Computer Communications and Networks (ICCCN), 2011 Proceedings of 20th International Conference on April 2014,page 1-7.

[6] Nik Bessis, , Florin Pop, Valentin Cristea , Meta-scheduling issues in interoperable HPCs, grids and clouds , Journal, International Journal of Web and Grid Services, Volume 8 Issue 2, August 2012, Pages 153-172.

[7] H. Cademartori, Green Computing Beyond the Data Center, 2007. Available: http://www.powersavesoftware.com/Download /PS WP GreenComputing EN.pdf

[8] Sheheryar Malik, Fabrice Huet, "Adaptive Fault Tolerance in Real Time Cloud Computing," services, pp.280-287, 2011 IEEE World Congress on Services, 2011.

[9] Íñigo Goiri , Josep Ll. Berral , J. Oriol Fitó, Ferran Juliá ,Energy-efficient and multifaceted resource management for profit-driven virtualized data centers, Journal, Future Generation Computer Systems, Volume 28 Issue 5, May, 2012, Pages 718-731 .

[10] Qi Zhang, Lu Cheng, Raouf Boutaba," Cloud computing: state-of-the-art and research challenges", Journal of Internet Services and Applications, May 2010, Vol 1, Issue 1, pp 7-18.

AUTHORS

G.Tharanya, she is received B.E., computer science and engineering degree from Anjalai Ammal-Mahalingam (AAMEC) Engineering College, Kovilvenni in 2011 and also currently pursuing doing M.E computer science and engineering degree from Sir Isaac Newton Engineering and Technology, Nagapattinam, Tamilnadu

E-mail: tharanyacse.90@gmail.com

G.Vijayasekaran obtained his Bachelor's degree in Information and Technology from Maamallan Institute of Technology, Chennai. Then he obtained his Master's degree in Computer Science and Engineering from MNM Jain Engg. College, Chennai Affiliated to Anna University Currently, he is an Assistant professor in Department of Computer Science and Engineering at Sir Isaac Newton College of Engineering and Technology, Nagapattinam

E-mail: gunavijay87@yahoo.co.in

R.Rajakumar received his M.Sc [IT], M.Phil [CS], M.Tech [CSE] Degree from various recognized Universities. Currently he is an working as a Assistant professor in Department of Computer Science and Engineering in Annai Group of Institutions, Kumbakonam

Email:
raja2012mtech@gmail.com

