

Characterization and Comparison of Cloud Energy Efficient Scheduling Algorithms

Rupali Gurav¹, Dipti Patil²

P.G. Student, Department of Information Technology, PIIT College, Panvel, Maharashtra, India

Assistant Professor, Department of Information Technology, PIIT College, Panvel, Maharashtra, India

ABSTRACT: Cloud computing is a new trend emerging in IT environment with huge requirements of infrastructure and resources. One of the main challenges in cloud computing is to increase the availability of computational resources, while minimizing system power consumption and operational expenses. It is challenging to effectively utilize the green energy since its generation depends on dynamic natural conditions. Efficient task scheduling and resource provisioning scheme ensure energy efficiency to cloud user's on-demand basis in pay-as-you-say-manner. This paper analyzes various energy efficient task scheduling algorithms as found in research history.

KEYWORDS: Cloud Computing, Energy Efficiency, Data Centers.

I. INTRODUCTION

Cloud Computing

The cloud computing is based on the concept of *dynamic provisioning*, which is applied to services, computing capability, storage, networking, and information technology infrastructure to meet user requirements. The resources are made available for the users through the Internet and offered on a pay-as-use basis from different Cloud computing vendors [1].

Cloud computing is a new paradigm for enterprises that can effectively facilitate the execution of tasks. Task scheduling is an important issue which is greatly influencing the performance of cloud computing environment. The cloud service provider and clients have different objectives and requirements [2].

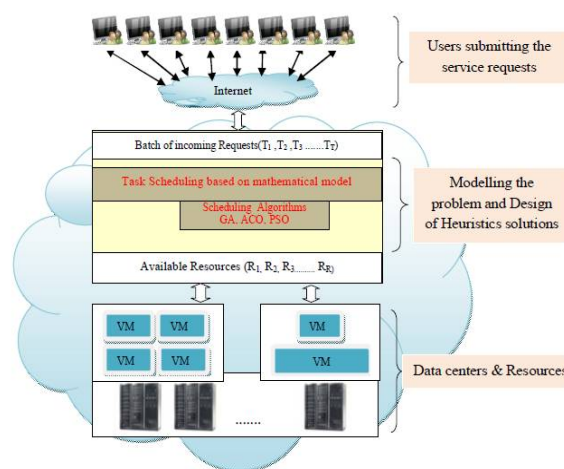


Fig. Cloud Computing Architecture

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

Energy Efficient Cloud Computing

Cloud computing infrastructures are designed to support the accessibility and deployment of various services oriented applications by the users. Cloud computing services are made available through the server farms or data centers. To meet the growing demand for computations and large volume of data, the cloud computing environments provides high performance servers and high speed mass storage devices. These resources are the major source of the power consumption in data centers along with air conditioning and cooling equipment. Due to high energy consumption by data centers, it requires efficient technology to design green data center. On the other hand, Cloud data center can reduce total energy consumed through task consolidation and server consolidation using the virtualization by workloads can share the same server and unused servers can be switched off.

II. MOTIVATION

Energy efficiency is increasingly important for cloud computing, because the increased usage of cloud computing infrastructure, together with increasing energy costs. There is a need to reduce the greenhouse gas emissions call for the energy efficient technologies that decrease the overall energy consumption of computation, storage and communication equipment. The current research trends on energy efficient resource allocation have identified the following key area for energy-saving techniques in cloud computing infrastructure:

- **Powering down:** Switching off the entire system when not in use or in idle state can be considered a key area of Energy Aware Computing.

- **Dynamic voltage and frequency scaling (DVFS):**

The DVFS technique is used to reduce the heat generated by the chip in two different ways. The power saving can be possible by adjusting automatically the operating frequency of the processor with the help of system clock available on board which also reduces the heat generated by the chip on operation.

- **Task Consolidation:** Srikantaiah et al.[2] have discussed an approach to switch off the idle machine by finding the minimum number of appropriate machine to which the task to be allocated.

- **Resource Scaling:** In this approach the minimum number of resources are assigned to the set of tasks to meet the deadline in such a way that the task will completed before the deadline to minimize the energy.

III. EXISTING ENERGY EFFICIENT SCHEDULING ALGORITHMS

Dynamic Energy-Aware Capacity Provisioning for Cloud Computing Environments

In this paper, Qi Zhang provided a control-theoretic solution to the dynamic capacity provisioning problem that minimizes the total energy cost while meeting the performance objective in terms of task scheduling delay. Specifically, they modeled this problem as a constrained discrete-time optimal control problem, and use Model Predictive Control (MPC) to find the optimal control policy. They mitigate this concern by designing a dynamic capacity provisioning system that controls the number of active servers in the data center according to (1) demand fluctuation, (2) variability in energy prices and (3) the cost of dynamic capacity reconfiguration. Their solution is based on the well-established Model Predictive Control framework, and aims to find a good trade-off between energy savings and capacity reconfiguration cost.

They used the Auto-Regressive Integrated Moving Average (ARIMA) model to predict the time series G_r^k which represents the usage of resource type r in all the machines at time k . They divided time into intervals of equal duration.

Task Scheduling and Server Provisioning for Energy-Efficient Cloud-Computing Data Centers

The proposed approach was formulated as an integer programming problem to minimize the cloud computing data center energy consumption by scheduling tasks to a minimum numbers of servers while keeping the task response time constraints. The average task response time and the number of active servers needed to meet such time constraints are bounded through the use of a greedy task scheduling scheme. They proposed the most-efficient server- first task-scheduling scheme to minimize energy expenditure as a practical scheduling scheme.

The proposed scheme minimizes the average task response time and, at the same time, minimizes server-related energy expenditure. The scheme presented using Matlab simulation with independent exponential distributed task arrivals and compares the performance of the proposed scheme with that of a random-based task-scheduling scheme.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

Energy-Efficient Scheduling of Urgent Bag-of-Tasks Applications in Clouds through DVFS

They proposed an approach for reducing the power required to execute urgent, CPU-intensive Bag-of- Tasks applications on cloud infrastructures. It exploits intelligent scheduling combined with the Dynamic Voltage and Frequency Scaling (DVFS) capability of modern CPU processors to keep the CPU operating at the minimum voltage level (and consequently minimum frequency and power consumption) that enables the application to complete before a user-defined deadline.

Their approach applies DVFS at the middleware/Operating System level rather than at CPU level

It is composed of a virtualized IaaS cloud data center that supports a PaaS layer that is available to multiple users for execution of urgent, CPU-intensive BoT applications. The data center is composed of a number of (potentially heterogeneous) physical servers (hosts). Virtual machines are executed in the servers, and they are managed by a virtual machine manager installed on each host.

The rationale behind proposed scheduling algorithm is to dynamically scale the frequency and voltage of CPUs assigned to a virtual machine in such a way that tasks sequentially executed in the VM complete before their deadlines. This type of application is found in domains such as disaster management and healthcare. They proposed a cloud aware scheduling algorithm that applies DVFS to enable deadlines for execution of urgent CPU-intensive Bag-of-Tasks jobs to be met with reduced energy expenditure.

Heterogeneity-aware Workload Placement and Migration in Distributed Sustainable Datacenters

They proposed a holistic heterogeneity-aware cloud workload placement and migration approach, sCloud, which aims to maximize the system goodput in distributed self-sustainable datacenters. They proposed a heterogeneity-aware cloud workload placement and migration approach in distributed self sustainable datacenters. It maximizes the system performance in terms of system goodput, based on the time varying green power supply, heterogeneous workload characteristics and QoS requirements. They designed an optimization-based algorithm to dynamically place transactional requests to distributed datacenters, with respect to their green power supplies. In order to further improve the system goodput, integrate another online algorithm to dynamically migrate batch jobs across distributed datacenters when the green power supplies vary widely at different locations. They have implemented sCloud in university cloud testbed and performed extensive evaluations with real world weather conditions and workload traces.

Energy-Efficient Resource Allocation and Provisioning Framework for Cloud Data Centers

The proposed framework: i) predicts the number of virtual machine (VM) requests, to be arriving at cloud data centers in the near future, along with the amount of CPU and memory resources associated with each of these requests, ii) provides accurate estimations of the number of physical machines (PMs) that cloud data centers need in order to serve their clients, and iii) reduces energy consumption of cloud data centers by putting to sleep unneeded PMs.

They developed a prediction approach that combines machine learning clustering and stochastic theory to predict both the number of VM requests and the amount of cloud resources associated with each request and proposed adaptive enhancements to their predictor that make the prediction parameters tunable in real time based on the actual request load. This increases the prediction accuracy over time and avoids the need for frequent model training that other machine learning approaches, such as Neural Network, suffer from. They proposed an integrated resource provisioning framework relies on the proposed prediction approaches to make suitable energy-aware resource management decisions. They used real Google data traces to evaluate the effectiveness of our framework.

A Genetic Algorithmic approach for Energy Efficient Task Consolidation in Cloud Computing

They proposed the GA-based task scheduling model and introduced a suitable codification scheme for chromosome. They have generated a POP_SIZE number of random initial population and calculating the fitness value of individuals. Then, using roulette wheel selection method, parents are selected to produce off springs using single point crossover with probability 0.8. Some of the individuals are subjected to the mutation with a probability 0.2. The population for the next generation is selected again through roulette wheel selection method. The constant population size has been maintained for a fixed number of iterations. The individual from the last generation with minimum energy value is selected to allocate the tasks to VMs.

Energy-efficient Scheduling Policy for Collaborative Execution in Mobile Cloud Computing

They formulated the task scheduling problem as a constrained stochastic shortest path problem over an acyclic graph. They applied the "LARAC" algorithm to obtain the task scheduling policy for Markovian chain model. Their investigation suggested a one-climb policy, in which there exists at most onetime migration from the mobile device to the cloud. Collaborative execution can significantly reduce energy consumption on a mobile device.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

Dynamic Heterogeneity-Aware Resource Provisioning in the Cloud

They present Heterogeneity-Aware dynamic capacity provisioning scheme for cloud data centers. Specifically, they first use the K-means clustering algorithm to divide workload into distinct task classes with similar characteristics in terms of resource and performance requirements. Then present a technique that dynamically adjusting the number of machines to minimize total energy consumption and scheduling delay.

pMapper: Power and Migration Cost Aware Application Placement in Virtualized Systems

Akshat Verma and his team have presented an application placement controller *pMapper* that minimizes power and migration costs, while meeting the performance guarantees. *pMapper* differs from all existing literature because it addresses the problem of power and migration cost aware application placement in heterogeneous server clusters that support virtualization with live VM migration.

Sr. No.	Paper Title	Algorithm or Framework Used	Assumptions	Future Work
1.	Dynamic Energy-Aware Capacity Provisioning for Cloud Computing Environments	Model Predictive Control (MPC)	Machines are homogeneous	To handle machine heterogeneity, Understand the interplay between the scheduler and the capacity controller
2.	Task Scheduling and Server Provisioning for Energy-Efficient Cloud-Computing Data Centers	server- first task-scheduling scheme, greedy task-scheduling scheme	Independent exponential distribution task arrivals, Used Central Scheduler	Only Focused on task deadline constraint can be possible to focus on other objective.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

Sr. No.	Paper Title	Algorithm or framework used	Assumptions	Future Work
3.	Energy-Efficient Scheduling of Urgent Bag-of-Tasks Applications in Clouds through DVFS	Dynamic Voltage and Frequency Scaling (DVFS) capability of modern CPU	Jobs with soft deadlines, a dedicated amount of servers are reserved, Targeted CPU-intensive applications	Improve algorithm to support other types of applications, such as workflows and MapReduce
4.	Heterogeneity-aware Workload Placement and Migration in Distributed Sustainable Datacenters	sCloud	Not Specified	Integrate the energy storage technique with sCloud for sustainable computing in green datacenters
5.	Energy-Efficient Resource Allocation and Provisioning Framework for Cloud Data Centers	Framework to predicts the number of virtual machine (VM) requests	Not Specified	To investigate whether VM requests follow certain regular daily trends and rely on that to further improve the workload prediction module
6.	A Genetic Algorithmic approach for Energy Efficient Task Consolidation in Cloud Computing	Genetic Algorithm	Not Specified	The change in no. of initial population, no. of generation, crossover and mutation probability

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

Sr. No.	Paper Title	Algorithm or Framework Used	Assumptions	Future Work
7.	Energy-efficient Scheduling Policy for Collaborative Execution in Mobile Cloud Computing	LARAC" algorithm	The tasks of the application have been replicated on the cloud side before the application execution. output data of the application must be displayed on the mobile device.	The task topology can be extended into more generic graphs, Structural properties for optimal task scheduling policy
8.	Dynamic Heterogeneity-Aware Resource Provisioning in the Cloud	Capacity Provisiong(CBP) Based and Capacity Scheduling(CBS)	Machines heterogeneity are naturally characterized	Consider Workload heterogeneity at fine-grained (i.e., per-task) level
9.	pMapper: Power and Migration Cost Aware Application Placement in Virtualized Systems	minPowerParity(mPP), pMapper	For each application being placed, they characterize the incremental power drawn by it at various workloads on all the server platforms	Consider task arrival rate and scheduling delay objectives.

Table: Comparison of different energy efficient Algorithms

IV. CONCLUSION

The cloud environment provides heterogeneous hardware and resource demands; therefore, it is important to exploit these features to make energy efficient cloud-based resource schedulers. In this paper we have discussed different energy efficient task scheduling algorithm and frameworks to better understand the existing research in this area along with possible future work to improve the existing work.

REFERENCES

- [1] Dilip Kumar, "Energy Efficient Resource Allocation for Cloud Computing", *International Journal of Artificial Intelligent Systems and Machine Learning*, CIIT, vol.6 no 1, January 2014, pp.32-38.
- [2] Srikantaiah, S., Kansal, A., and Zhao, F. (2008), "Energy aware consolidation for cloud computing", *Proceedings of the 2008 conference on Power aware Computing and systems*, volume 10, USENIX Association.
- [3] Qi Zhang, M. F. Zhani, Shuo Zhang, Quanyan Zhu, Raouf Boutaba and Joseph L. H., "Dynamic Energy-Aware Capacity Provisioning for Cloud Computing Environments".
- [4] Ziqian Dong, Roberto Rojas-Cessa and Ning Liu., "Task Scheduling and Server Provisioning for Energy-Efficient Cloud-Computing Data Centers", *Conference Paper- July 2013*, DOI: 10.1109/ICDCSW.2013.68.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

- [5] Rodrigo N. Calheiros and Rajkumar Buyya, "Energy-Efficient Scheduling of Urgent Bag-of-Tasks Applications in Clouds through DVFS", 2014 *IEEE 6th International Conference on Cloud Computing Technology and Science*.
- [6] Dazhao Cheng, Changjun Jiangy and Xiaobo Zhou., "Heterogeneity-aware Workload Placement and Migration in Distributed Sustainable Datacenters".
- [7] Mehdiar Dabbagh, Bechir Hamdaoui, Mohsen Guizaniy and Ammar Rayes, "Energy-Efficient Resource Allocation and Provisioning Framework for Cloud Data Centers".
- [8] Dilip Kumar, Bibhudatta Sahoo, Bhaskar Mondal and Tarni Mandal, "A Genetic Algorithmic approach for Energy Efficient Task Consolidation in Cloud Computing", *International Journal of Computer Applications (0975 8887)*, Volume 118 - No. 2, May 2015.
- [9] Weiwen Zhang¹, Yonggang Wen¹, and Dapeng Oliver Wu²., "Energy-efficient Scheduling Policy for Collaborative Execution in Mobile Cloud Computing", *2013 Proceedings IEEE INFOCOM*.
- [10] Qi Zhang, M. F. Zhani, Shuo Zhang, Quanyan Zhu, Raouf Boutaba and Joseph L. H., "Dynamic Heterogeneity-Aware Resource Provisioning in the Cloud" , *IEEE TRANSACTIONS ON CLOUD COMPUTING*, VOL. 2, NO. 1, JANUARY-MARCH 2014.
- [11] Akshat Verma¹, Puneet Ahuja², and Anindya Neogi, "pMapper: Power and Migration Cost Aware Application Placement in Virtualized Systems".