IOSR Journal of Engineering (IOSR JEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
PP 67-70

An Overview of Energy-Efficient Cloud Data Centres

Prof. Anupam Chaube
G. H. Raisoni Institute of Information Technology, Nagpur

Abstract: *The designers of computing systems have interested on the improvements of computing performance that are driven by the demand of consumer, business, and scientific applications. However, further growth of computing performance has started to be limited due to the increasing energy consumption of cloud data centres as a result of carbon dioxide footprints and overwhelming electricity bills. Therefore, the designers of computing systems have started to improve the energy efficiency. In this study, we present an overview of energy efficient design of cloud virtualized data centres to guide future.*
Keywords: *cloud computing; energy consumption; data centre; virtualization*

I. Introduction

The designers of computing systems have interested on the improvement of the system performance. The performance per watt ratio has been constantly rising, but the total power drawn by computing systems is hardly decreasing. If this trend continues, the energy consumption cost of physical machines (PMs) in their lifetime will be more than the cost of hardware [1]. For large-scale computing infrastructures, the problem is even worse. Power consumption of USA data centres has increased by 62.5% from 2005 to 2013 and expected to increase by 150% in 2020 [2]. Most of the energy consumption of data centres is consumed by computing resources. Accordingly, resource management is important to ensure that the applications efficiently utilize the available computing resources.

Impact of energy efficiency on end-users is usually determined by the total resource usage cost of a resource provider. Results of high energy consumption increase not only electricity bills but also additional requirements to power delivery infrastructure and a cooling system. The cooling problem becomes crucial with the increasing density of computer components. Thus, heat dissipation should be improved [3]. The main goal of this work is to give an overview of the recent research works in energy-efficient cloud data centres. Furthermore, the aim is to show the recent level of development in the area and discuss directions for future work.

The rest of the paper is organized as follows. In Section II, we present an overview of the research in energy-efficient design of cloud data centres. In Section III, a conclusion and future works are presented.

II. Energy-Efficient Resource Management For Clouddata Centres

Most energy-efficient resource management approaches for cloud computing aim to improve consolidation of the workload into the minimum of PMs. Idle resources can be switched off, which reduces energy consumption, and increases resource utilization. However, the consolidation has to meet the servicelevel agreement (SLA) requirements, as well as minimize the performance degradation and energy consumption. In this section, we give an overview of different approaches of efficient management of trade-off between performance and energy in virtualized data centres. Table I shows the most important reviewed research works.

An approach for power-efficient resource management in data centres has been proposed for the first time in the context of virtualized systems by Nathuji and Schwan [4]. A new power management technique called “soft resource scaling” was applied in this research. Soft resource scaling uses the virtual machine manager (VMM)’s scheduling ability to provide a virtual machine (VM) less time for utilizing the resource. “Soft” scaling is suitable when hardware scaling provides a little decrease in energy or is not supported. It has been found that the combination of “soft” and “hard” scaling improves energy consumption because of typically limited number of hardware scaling states.

An approach for power management for a data centre by combination of different power management strategies has been proposed by Raghavendra et al. [5]. The authors find that it is more likely to apply different solutions from multiple vendors even though all aspects of power management can be handled by implementing a centralized solution. The existing solutions are classified by a number of attributes, such as hardware/software, local/global types of policies, the objective function, and performance constraints. The authors focus on individual solutions instead of solving the whole problem. They have applied a feedback control loop to coordinate the controllers’ actions. However, the proposed approach is unable to guarantee that system meet quality of service (QoS) requirements. Thus, the approach is not suitable for cloud computing providers, where a comprehensive support for SLA and more reliable QoS are essential, but for enterprise environments.

Table 1: Cloud Data Center Research Works

Authors	Resources	Goal	Techniques
Nathuji and Schwan [4]	CPU	Minimizing energy under performance constraints	Soft scaling, switching off PM, VM consolidation, DFVS
Raghavendra et al. [5]	CPU	Minimizing power and meet power budget	Switching off PM, VM consolidation, DVFS
Verma et al. [6]	CPU	Minimizing power with meeting performance requirements	Switching off PM, VM consolidation, DVFS
Kusic et al. [7]	CPU	Minimizing power with meeting performance requirements	Switching off PM, VM consolidation
Song et al. [8]	RAM, CPU	Minimizing energy with meeting performance requirements	Resource throttling
Stillwell et al. [9]	CPU	Minimizing energy with meeting performance requirements	VM consolidation, resource throttling
Cardosa et al. [10]	CPU	Minimizing power with meeting performance requirements	Soft scaling, DFVS
Gmach et al. [11]	RAM ,CPU	Minimizing energy with meeting performance requirements	VM consolidation, switching off PM
Kumar et al. [12]	RAM, CPU, Network	Minimizing energy with meeting performance and power budget constraints	DVFS, VM consolidation
Buyya et al. [15]	CPU	Minimizing energy with meeting performance requirements	Leveraging heterogeneity, DVFS
Beloglazove and Buyya [17, 18]	CPU, RAM	Minimizing energy with meeting performance requirements	Switching off PM, dynamic VM consolidation, DVFS
Horri et al. [19]	CPU, RAM	Minimizing energy with meeting performance requirements	Switching off PM, Dynamic VM consolidation
Fu and Zhou [20]	CPU, RAM	Minimizing energy with meeting performance requirements	Switching off PM, Dynamic VM consolidation
Arianyan et al. [21]	CPU, RAM	Minimizing energy with meeting performance requirements	Switching off PM, Dynamic VM consolidation
Han Et al [22]	CPU, RAM	Minimizing energy with meeting performance requirements	Switching off PM, Dynamic VM consolidation

The problem of dynamic VM placement has been solved by a heuristic bin packing algorithm by Verma et al. [6]. The pMapper application placement framework has been proposed to minimize the power consumption and maintain the SLA. It consists of an arbitrator, performance manager, power manager and migration manager. The authors claim that the proposed framework is general enough to combine different performance and power management strategies under SLA's restrictions. The problem has been formulated as a continuous improvement. The VM placement should be improved at each time frame to maximize the performance and minimize the power consumption. Moreover, the solution can be used to any type of the workload. However, SLA cannot be met because of unforeseeable workloads and instability. The authors have defined a bin packing algorithm using variable costs and bin sizes to solve the placement problem. A proposed algorithm and the pMapper architecture have been implemented with performing extensive experiments to evaluate the system efficiency. The authors have suggested numerous future works such as more advanced applications of idle states, consideration of memory bandwidth, and extending the theoretical formulation of the problem. An approach for performance and power efficient resource management in virtualized computing systems has been proposed by Kusic et al. [7]. SLA for each application in the computing system was defined as the request processing rate. The goal was to raise the profit of resource providers by reducing SLA violation and power consumption. A sequential optimization was used to solve the problem using the limited lookahead control (LLC). To improve the performance, neuralnetworks have been applied. However, 10 hosts were only used in the experiments, which was not sufficient to demonstrate the applicability of an approach.

An approach for the efficient resource allocation in multiapplication virtualized data centres has been proposed by Song et al. [8]. The goal was to reduce energy consumption by improving the resource utilization. According to the application priorities, the resources were allocated to applications using several VMs instantiated on different PMs. Random access memory (RAM) and CPU utilizations were only taken into consideration in the decisions. A drop in the performance occured for the low-priority applications in cases of limited resources, since critical applications were only used the

resources. VMs were preinstantiated on a set of PMs, and they were assigned by fractions of the total resources. The limitation of the proposed approach is that the migration of VMs is not used for adjusting the allocation at run-time. Moreover, machine learning is required to obtain the resource utilization functions. Such approach is appropriate for enterprise environment, where the priorities of applications can be clearly defined.

The resource allocation problem for high performance computing (HPC) applications in virtualized homogeneous clusters have been studied by Stillwell et al. [9]. The goal was to maximize the resource utilization by improving user-centric metric that was denoted as the yield, which was part of the maximum achievable computing rate. A mixed integer programming model has been proposed to define problem and to solve small instances of the problem, but not to solve largescale problem instances. It was assumed that the resource requirements of applications to be known in advance, which is not typical in practice. Moreover, CPU is only considered in the optimization.

An approach for the power-efficient VM allocation in virtualized enterprise computing environments has been proposed by Cardosa et al. [10]. They used min, max, and shares parameters. Min and max parameters specify minimum and maximum allowed amount of CPU resources allocated to a

VM. Shares parameter sets percentages, where CPU resources will be allocated to each VM that share the same resource. This approach is suitable for enterprise environments, where a comprehensive support for SLA and the priorities of applications are not essential. The proposed algorithm enhances the ability to reduce the resources required by a VM to the minimum and expand it when there are available spare resources to bring additional profit. One of the limitations of the proposed approach is that the VM migration is not used for adjusting the allocation at run-time. Another problem is that CPU is only considered by the model in the optimization. Moreover, static definition of the application priorities is required for the proposed approach, which makes this approach not applicable in real-world environments.

An approach for the energy-efficient dynamic VM consolidation for enterprise systems has been proposed by Gmach et al. [11]. In this research, combination of a migration controller and a workload placement controller has been proposed. The workload placement controller improves the allocation under QoS requirements using historical resource usage information collected from VMs in the data centre. Multi-objective improvement was performed to find a new VM placement which minimized the required VM migrations and the average number of PMs in the system. The migration controller was run in parallel to solve the underloading and overloading of PMs.

The GreenCloud project that aimed at developing of energy-efficient provisioning of cloud resources under QoS requirements has been proposed by Buyya et al. [15]. A realtime virtual machine model has been introduced. Three policies for proposed model have been proposed using dynamic voltage

and frequency scaling (DVFS) to improve the energy efficiency and maximize the request acceptance rate, while meeting QoS constraints. The policies are the lowest-DVS policy, the d-advanced-DVS policy, and the adaptive-DVS policy. The approach has been evaluated using the CloudSim toolkit.

A novel QoS-aware VMs consolidation approach has been proposed by Horri et al. [19] for cloud environments, which adopts a method based on resource utilization history of virtual machines. The authors have evaluated proposed algorithms using CloudSim simulator. Experimental evaluation shows that there is improvement in energy consumption and QoS metrics. In addition, it proves a trade-off between QoS and energy consumption in the cloud environment.

VM placement and selection for dynamic consolidation in cloud computing environment has been proposed by Fu and Zhou [20]. VM selection policy not only considers CPU utilization, but also defines a variable that represents the degree of resource satisfaction to select VMs. Moreover, authors have proposed a novel VM placement policy based on selecting PM with minimum correlation coefficient to be placed by a migratable VM. Results show that the proposed policies perform better than existing policies according to SLA violation percentage, VM migration time, and energy consumption.

Arianyan et al. [21] have proposed as an effective VM consolidation to save energy in cloud data centres. The authors have proposed a new holistic cloud resource management procedure as well as novel heuristics based on multi-criteria decision making method to solve underloaded PMs and VM placement problems. The results of simulations using Cloudsim simulator validates the applicability of the proposed policies which shows up reductions in number of VM migrations, SLA violation, and energy consumption in comparison with benchmark works.

III. Conclusions And Future Directions

Energy efficiency is the most important design requirements for modern computing systems, as they continue to consume large amounts of electrical power. Apart from high operating costs incurred by computing resources, this leads to significant emissions of carbon dioxide (CO2) into the environment. Most of the energy consumption of data centres is consumed by computing resources. Accordingly, resource management is important to ensure that the applications efficiently utilize the available computing resources. It is necessary to study the existing resource management approaches to facilitate further improvements in this area. Therefore, we have given an overviewed of a number of recent research works in energy-efficient resource management in virtualized data centres.

For future work, we suggest to improve network topologies established between VMs for virtualized data centres to reduce the load of network devices and the network communication overhead. We also propose to investigate cloud federations including geographically distributed data centres. Another research direction is to investigate the control of user over the energy consumption in data centres, in addition to support the flexible SLA negotiation between users and resource providers. However, many open research challenges have become more prominent in the age of cloud computing.

References

[1]. L. Barroso, "The Price of Performance, ACM Press, " Queue, vol. 3, no. 7, pp. 53, 2005.
[2]. W.D. Weber, X. Fan, L.A. Barroso, "Powering the data center," U.S. Patent No. 8,595,515. Washington, DC: U.S. Patent and Trademark Office, 2013.
[3]. A. Beloglazov, R. Buyya, Y.C. Lee, A. Zomaya, "A taxonomy and survey of energy-efficient data centers and cloud computing systems, " Advances in computers, vol. 82, no. 2, pp. 47-111, 2011.
[4]. R. Nathuji and K. Schwan, "VirtualPower: Coordinated power management in virtualized enterprise systems," ACM SIGOPS Operating Systems Review, vol. 41(6), pp. 265–278, 2007.
[5]. R. Raghavendra, P. Ranganathan, V. Talwar, Z. Wang, and X. Zhu, "No "power" struggles: Coordinated multi-level power management for the data center," SIGARCH Computer Architecture News, vol. 36, no. 1, pp. 48–59, 2008.
[6]. A. Verma, P. Ahuja, and A. Neogi, "pMapper: power and migration cost aware application placement in virtualized systems," in Proceedings of the 9th ACM/IFIP/USENIX International Conference on Middleware, pp. 243–264, 2008.