

Energy Efficient Resource Management in Virtualized Cloud Data Centers

Anton Beloglazov* and Rajkumar Buyya

Cloud Computing and Distributed Systems (CLOUDS) Laboratory
Department of Computer Science and Software Engineering
The University of Melbourne, Australia
{abe, raj}@csse.unimelb.edu.au

Abstract

Rapid growth of the demand for computational power by scientific, business and web-applications has led to the creation of large-scale data centers consuming enormous amounts of electrical power. We propose an energy efficient resource management system for virtualized Cloud data centers that reduces operational costs and provides required Quality of Service (QoS). Energy savings are achieved by continuous consolidation of VMs according to current utilization of resources, virtual network topologies established between VMs and thermal state of computing nodes. We present first results of simulation-driven evaluation of heuristics for dynamic reallocation of VMs using live migration according to current requirements for CPU performance. The results show that the proposed technique brings substantial energy savings, while ensuring reliable QoS. This justifies further investigation and development of the proposed resource management system.

1. Introduction

Modern resource-intensive enterprise and scientific applications create growing demand for high performance computing infrastructures. This has led to the construction of large-scale computing data centers consuming enormous amounts of electrical power. Despite of the improvements in energy efficiency of the hardware, overall energy consumption continues to grow due to increasing requirements for computing resources. For example, in 2006 the cost of energy consumption by IT infrastructures in US was estimated as 4.5 billion dollars and it is likely to double by 2011 [2]. Apart from the overwhelming operational costs,

building a data center leads to excessive establishment expenses as data centers are usually built to serve infrequent peak loads resulting in low average utilization of the resources. Moreover, there are other crucial problems that arise from high power consumption. Insufficient or malfunctioning cooling system can lead to overheating of the resources reducing system reliability and devices lifetime. In addition, high power consumption by the infrastructure leads to substantial carbon dioxide (CO_2) emissions contributing to the greenhouse effect.

A number of practices can be applied to achieve energy-efficiency, such as improvement of applications' algorithms, energy efficient hardware, Dynamic Voltage and Frequency Scaling (DVFS) [13], terminal servers and thin clients, and virtualization of computer resources [1]. Virtualization technology allows one to create several Virtual Machines (VMs) on a physical server and, therefore, reduces amount of hardware in use and improves the utilization of resources. Among the benefits of virtualization are improved fault and performance isolation between applications sharing the same resource (a VM is viewed as a dedicated resource to the customer); the ability to relatively easy move VMs from one physical host to another using live or off-line migration; and support for hardware and software heterogeneity.

Traditionally, an organization purchases its own computing resources and deals with maintenance and upgrade of the outdated hardware, resulting in additional expenses. The recently emerged Cloud computing paradigm [4] leverages virtualization technology and provides the ability to provision resources on-demand on the pay-as-you-go basis. Organizations can outsource their computation needs to the Cloud, thereby eliminating the necessity to maintain own computing infrastructure. Cloud computing naturally leads to energy-efficiency by providing the following characteristics:

- Economy of scale due to elimination of redundancies.

*The author is in the first year of his candidature as a PhD student and Prof. Rajkumar Buyya is the supervisor. The work presented in this paper is a proposal for future research.

- Improved utilization of the resources.
- Location independence – VMs can be moved to a place where energy is cheaper.
- Scaling up and down – resource usage can be adjusted to current requirements.
- Efficient resource management by the Cloud provider.

One of the important requirements for a Cloud computing environment is providing reliable QoS. It can be defined in terms of Service Level Agreements (SLA) that describe such characteristics as minimal throughput, maximal response time or latency delivered by the deployed system. Although modern virtualization technologies can ensure performance isolation between VMs sharing the same physical computing node, due to aggressive consolidation and variability of the workload some VMs may not get the required amount of resource when requested. This leads to performance loss in terms of increased response time, time outs or failures in the worst case. Therefore, Cloud providers have to deal with energy-performance trade-off – minimization of energy consumption, while meeting QoS requirements.

1.1. 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 main instrument that we leverage is live migration of VMs. 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.

1.2. Research challenges

The key challenges that have to be addressed are:

1. How to optimally solve the trade-off between energy savings and delivered performance?
2. How to determine when, which VMs, and where to migrate in order to minimize energy consumption by the system, while minimizing migration overhead and ensuring SLA?
3. How to develop efficient decentralized and scalable algorithms for resource allocation?
4. How to develop comprehensive solution by combining several allocation policies with different objectives?

The remainder of the paper is organized as follows. In the next section we discuss related work followed by the proposed system architecture in Section 3. In Sections 4 and 5 we present the allocation policies and evaluation of the heuristics followed by conclusion and future work.

2. Related Work

Early work in energy aware resource management is devoted to mobile devices with the objective of improving battery lifetime [11, 18]. Later on, the context has been shifted to data centers [12, 7] and virtual computing environments such as Clouds. Nathuji and Schwan [10] have proposed an architecture of energy management system for virtualized data centers where resource management is divided into local and global policies. On the local level, the system leverages guest operating system’s power management strategies. Consolidation of VMs is handled by global policies applying live migration to reallocate VMs. However, the global policies are not discussed in detail considering QoS requirements. In contrast, our work focuses on global VM allocation policies considering strict SLA.

Kusic et al. [8] have stated the problem of continuous consolidation as a sequential optimization and addressed it using Limited Lookahead Control (LLC). The proposed model requires simulation-based learning for the application-specific adjustments. Due to complexity of the model the optimization controller’s execution time reaches 30 minutes even for a small number of nodes (e.g. 15), that is not suitable for large-scale real-world systems. On the contrary, our approach is heuristic-based allowing the achievement of reasonable performance even for large-scale as shown in our experimental studies.

Srikantaiah et al. [16] have studied the problem of requests scheduling for multi-tiered web-applications in virtualized heterogeneous systems in order to minimize energy consumption, while meeting performance requirements. To handle the optimization over multiple resources, the authors have proposed a heuristic for multidimensional bin packing problem as an algorithm for workload consolidation. However, the proposed approach is workload type and application dependent, whereas our algorithms are independent of the workload type and thus are suitable for a generic Cloud environment.

Song et al. [15] have proposed resource allocation to applications according to their priorities in multi-application virtualized cluster. The approach requires machine-learning to obtain utility functions for the applications and defined application priorities. Unlike our work, it does not apply migration of VMs to optimize allocation continuously (the allocation is static).

Cardosa et al. [6] have explored the problem of power-efficient allocation of VMs in virtualized heterogeneous computing environments. They have leveraged “min”, “max” and “shares” parameters of VMM that represent minimum, maximum and proportion of CPU allocated to VMs sharing the same resource. The approach suits only enterprise environments or private Clouds as it does not support strict SLA and requires knowledge of applications priorities to define shares parameter.

Verma et al. [17] have formulated the problem of dynamic placement of applications in virtualized heterogeneous systems as continuous optimization: at each time frame the placement of VMs is optimized to minimize power consumption and maximize performance. The authors have applied a heuristic for bin packing problem with variable bin sizes and costs. The authors have introduced the notion of cost of VM live migration, but the information about the cost calculation is not provided. The proposed algorithms, on the contrary to our approach, do not handle strict SLA requirements: SLA can be violated due to variability of the workload.

In contrast to the discussed studies, we propose efficient heuristics for dynamic adaption of allocation of VMs in run-time applying live migration according to current utilization of resources and thus minimizing energy consumption. The proposed approach can effectively handle strict QoS requirements, heterogeneous infrastructure and heterogeneous VMs. The algorithms do not depend on a particular type of workload and do not require any knowledge about applications executing on VMs.

Calheiros et al. [5] have investigated the problem of mapping VMs on physical nodes optimizing network communication between VMs, however, the problem has not been explored in the context of energy consumption minimization. Recently, a number of research works have been

done on thermal efficient resource management in data centers [14, 9]. The studies show that software-driven thermal management and temperature aware workload placement bring additional energy savings. However, the problem of thermal management in the context of virtualized data centers has not been investigated. Moreover, to the best of our knowledge there are no studies on a comprehensive approach that combines optimization of VM placement according to current utilization of the resources with network and thermal optimization for virtualized data centers. Therefore, the exploration of such approach is timely and crucial, especially considering rapid development of Cloud computing environments.

3. System Architecture

In this work the underlying infrastructure is represented by a large-scale Cloud data center comprising n heterogeneous physical nodes. Each node has a CPU, which can be multi-core, with performance defined in Millions Instructions Per Second (MIPS). Besides that, a node is characterized by the amount of RAM and network bandwidth. Users submit requests for provisioning of m heterogeneous VMs with resource requirements defined in MIPS, amount of RAM and network bandwidth. SLA violation occurs when a VM cannot get the requested amount of resource, which may happen due to VM consolidation.

The software system architecture is tiered comprising a dispatcher, global and local managers. The local managers reside on each physical node as a part of a Virtual Machine Monitor (VMM). They are responsible for observing current utilization of the node’s resources and its thermal state. The local managers choose VMs that have to be migrated to another node in the following cases:

- The utilization of some resource is close to 100% that creates a risk of SLA violation.
- The utilization of resources is low, therefore, all the VMs should be reallocated to another node and the idle node should be turned off.
- A VM has intensive network communication with another VM allocated to a different physical host.
- The temperature exceeds some limit and VMs have to be migrated in order to reduce load on the cooling system and allow the node to cool down naturally.

The local managers send to the global managers the information about the utilization of resources and VMs chosen to migrate. Besides that, they issue commands for VM resizing, application of DVFS and turning on / off idle nodes. Each global manager is attached to a set of nodes

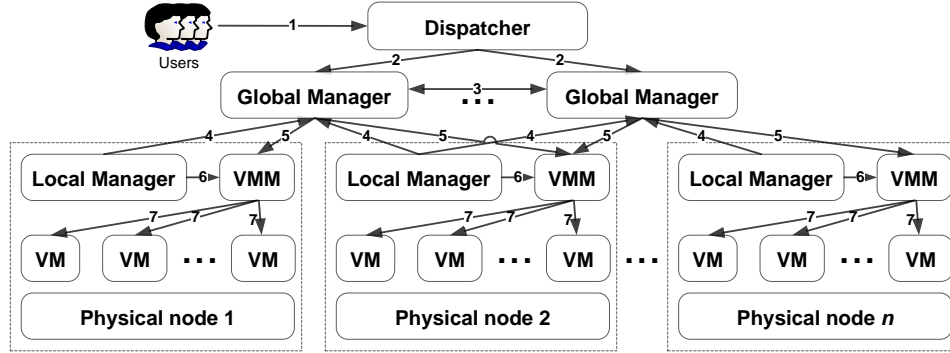


Figure 1. The system architecture

and processes data obtained from their local managers. The global managers continuously apply distributed version of a heuristic for semi-online multidimensional bin-packing, where bins represent physical nodes and items are VMs that have to be allocated. The decentralization removes a Single Point of Failure (SPF) and improves scalability. Each dimension of an item represents the utilization of a particular resource. After obtaining allocation decision, the global managers issue commands for live migration of VMs.

As shown in Figure 1, the system operation consists of the following steps:

1. **New requests for VM provisioning.** Users submit requests for provisioning of VMs.
2. **Dispatching requests for VM provisioning.** The dispatcher distributes the requests among the global managers.
3. **Intercommunication between global managers.** The global managers exchange information about utilization of resources and VMs that have to be allocated.
4. **Data about utilization of resources and VMs chosen to migrate.** The local managers propagate information about resource utilization and VMs chosen to migrate to the global managers.
5. **Migration commands.** The global managers issue VM migration commands in order to optimize current allocation.
6. **Commands for VM resizing and adjusting of power states.** The local managers monitor their host nodes and issue commands for VM resizing and changes in power states of nodes.
7. **VM resizing, scheduling and migration actions.** According to the received commands, VMM performs actual resizing and migration of VMs as well as resource scheduling.

4. Allocation policies

We propose three stages of VM placement optimization: reallocation according to current utilization of multiple system resources, optimization of virtual network topologies established between VMs and VM reallocation considering thermal state of the resources. Each of these stages is planned to be investigated separately and then combined in an overall solution. The developed algorithms have to meet the following requirements:

- Decentralization and parallelism – to eliminate SPF and provide scalability.
- High performance – the system has to be able to quickly respond to changes in the workload.
- Guaranteed QoS – the algorithms have to provide reliable QoS by meeting SLA.
- Independence of the workload type – the algorithms have to be able to perform efficiently in mixed-application environments.

The VM reallocation problem can be divided in two: selection of VMs to migrate and determining new placement of these VMs on physical hosts. The first part has to be considered separately for each optimization stage. The second part is solved by application of a heuristic for semi-online multidimensional bin-packing problem.

At the first optimization stage, the utilization of resources is monitored and VMs are reallocated to minimize the number of physical nodes in use and thus minimize energy consumption by the system. However, aggressive consolidation of VMs may lead to violation of performance requirements. We have proposed several heuristics for selection of VMs to migrate and investigated the trade-off between performance and energy savings. To simplify the problem for the first step we considered only utilization of CPU. The main idea of the policies is to set upper and lower

utilization thresholds and keep total utilization of CPU created by VMs sharing the same node between these thresholds. If the utilization exceeds the upper thresholds, some VMs have to be migrated from the node to reduce the risk of SLA violation. If the utilization goes below the lower thresholds, all VMs have to be migrated and the node has to be switched off to save the energy consumed by the idle node. Another problem is to determine particular values of the utilization thresholds. The results of the proposed algorithms evaluation are presented in Section 5.

Due to continuous reallocation, some intensively communicating VMs can be placed inefficiently leading to excessive load on the network facilities. Therefore, it is crucial to consider network communication behavior of VMs in reallocation decisions. The aim of the second proposed optimization stage is to place communicating VMs in a way minimizing the overhead of data transfer over network.

A cooling system of a data center consumes a significant amount of energy, therefore, the third proposed optimization stage is aimed at optimization of cooling system operation. Due to consolidation, some computing nodes experience high load leading to overheating and thus require extensive cooling. Monitoring of the nodes' thermal state using sensors gives an opportunity to recognize overheating and reallocate workload from the overheated node to allow the natural cooling. The network and temperature optimizations are subjects for the ongoing research work.

5. Evaluation

As the proposed system is targeted on a large-scale Cloud data center, it is necessary to conduct large-scale experiments to evaluate the algorithms. However, it is difficult to run large-scale experiments on a real-world infrastructure, especially when the experiments have to be repeated for different policies with the same conditions [3]. Therefore, simulation has been chosen as a way to evaluate the proposed heuristics. We have chosen CloudSim toolkit [3] as a simulation framework, as it is built for simulation of Cloud computing environments. In comparison to alternative simulation toolkits (e.g. SimGrid, GangSim), CloudSim supports modeling of on-demand virtualization enabled resource and application management. We have extended the framework in order to enable energy aware simulations as the core framework does not provide this capability. In addition, we have incorporated the abilities to account SLA violations and to simulate dynamic workloads that correspond to web-applications and online services.

The simulated data center consists of 100 heterogeneous physical nodes. Each node is modeled to have one CPU core with performance equivalent to 1000, 2000 or 3000 MIPS, 8 Gb of RAM and 1 TB of storage. Users submit requests for provisioning of 290 heterogeneous VMs that fill the full ca-

capacity of the data center. For the borderline policies we simulated a Non Power Aware policy (NPA) and DVFS that adjusts the voltage and frequency of CPU according to current utilization. We simulated a Single Threshold policy (ST) and two-threshold policy aimed at Minimization of Migrations (MM). Besides that, the policies have been evaluated with different values of the thresholds.

Table 1. Simulation results

Policy	Energy	SLA	Migr.	Avg. SLA
NPA	9.15 KWh	-	-	-
DVFS	4.40 KWh	-	-	-
ST 50%	2.03 KWh	5.41%	35 226	81%
ST 60%	1.50 KWh	9.04%	34 231	89%
MM 30-70%	1.48 KWh	1.11%	3 359	56%
MM 40-80%	1.27 KWh	2.75%	3 241	65%
MM 50-90%	1.14 KWh	6.69%	3 120	76%

The simulation results are presented in Table 1. The results show that dynamic reallocation of VMs according to current utilization of CPU can bring higher energy savings comparing to static allocation policies. MM policy allows to achieve the best energy savings: by 83%, 66% and 23% less energy consumption relatively to NPA, DVFS and ST policies respectively with thresholds 30-70% and ensuring percentage of SLA violations of 1.1%; and by 87%, 74% and 43% with thresholds 50-90% and 6.7% of SLA violations. MM policy leads to more than 10 times fewer VM migrations than ST. The results show the flexibility of the algorithm, as the thresholds can be adjusted according to SLA requirements. Strict SLA (1.11%) allow achievement of the energy consumption of 1.48 KWh. However, if SLA are relaxed (6.69%), the energy consumption is further reduced to 1.14 KWh.

6. Conclusion and Future Work

In this paper we have presented a decentralized architecture of the energy aware resource management system for Cloud data centers. We have defined the problem of minimizing the energy consumption while meeting QoS requirements and stated the requirements for VM allocation policies. Moreover, we have proposed three stages of continuous optimization of VM placement and presented heuristics for a simplified version of the first stage. The heuristics have been evaluated by simulation using the extended CloudSim toolkit. One of the heuristics leads to significant reduction of the energy consumption by a Cloud data center – by 83% in comparison to a non power aware system and by 66% in comparison to a system that applies only DVFS technique but does not adapt allocation of VMs in run-time. Moreover, MM policy enables flexible adjustment of SLA by setting appropriate values of the utilization thresholds: SLA can be relaxed leading to further improvement of en-

ergy consumption. The policy supports heterogeneity of both the hardware and VMs and does not require any knowledge about particular applications running on the VMs. The policy is independent of the workload type.

Table 2. Future research work and timeline

Timeline	Work description
01/2010 - 03/2010	Completion of the algorithms for the optimization over multiple resources.
04/2010 - 09/2010	Algorithms and a prototype for the network optimization.
10/2010 - 03/2011	Algorithms and a prototype for the temperature optimization.
04/2011 - 02/2012	Overall solution and a real implementation of the system as a module of a VMM, and experimental evaluation.

We propose a plan for the future research work that consists of several steps presented in Table 2. Once the algorithms for all of the proposed optimization stages are developed, they will be combined in an overall solution and implemented as a part of a real-world Cloud platform, such as Aneka¹.

The obtained results show that dynamic consolidation of VMs brings substantial energy savings while providing required QoS. Besides the significant reduction of operational costs, the project is socially valuable as it decreases carbon dioxide footprints and overall energy consumption by modern IT infrastructures.

References

[1] P. Barham, B. Dragovic, K. Fraser, S. Hand, T. Harris, A. Ho, R. Neugebauer, I. Pratt, and A. Warfield. Xen and the art of virtualization. In *Proceedings of the 19th ACM symposium on Operating systems principles*, page 177, 2003.

[2] R. Brown et al. Report to congress on server and data center energy efficiency: Public law 109-431. *Lawrence Berkeley National Laboratory*, 2008.

[3] R. Buyya, R. Ranjan, and R. N. Calheiros. Modeling and simulation of scalable cloud computing environments and the CloudSim toolkit: Challenges and opportunities. In *Proceedings of the 7th High Performance Computing and Simulation Conference (HPCS'09)*. IEEE Press, NY, USA, 2009.

[4] R. Buyya, C. S. Yeo, and S. Venugopal. Market-oriented cloud computing: Vision, hype, and reality for delivering it services as computing utilities. In *Proceedings of the 10th IEEE International Conference on High Performance Computing and Communications (HPCC'08)*. IEEE CS Press, Los Alamitos, CA, USA, 2008.

¹Aneka is a market-oriented Cloud development and management platform with rapid application development and workload distribution capabilities (<http://www.manjrasoft.com/>).

[5] R. N. Calheiros, R. Buyya, and C. A. F. D. Rose. A heuristic for mapping virtual machines and links in emulation testbeds. In *Proceedings of the 38th International Conference on Parallel Processing (ICPP 2009)*. IEEE CS Press, USA, 2009.

[6] M. Cardosa, M. Korupolu, and A. Singh. Shares and utilities based power consolidation in virtualized server environments. In *Proceedings of IFIP/IEEE Integrated Network Management (IM)*, 2009.

[7] J. S. Chase, D. C. Anderson, P. N. Thakar, A. M. Vahdat, and R. P. Doyle. Managing energy and server resources in hosting centers. In *Proceedings of the 18th ACM symposium on Operating systems principles*, pages 103–116. ACM New York, NY, USA, 2001.

[8] D. Kusic, J. O. Kephart, J. E. Hanson, N. Kandasamy, and G. Jiang. Power and performance management of virtualized computing environments via lookahead control. *Cluster Computing*, 12(1):1–15, 2009.

[9] J. Moore, J. Chase, P. Ranganathan, and R. Sharma. Making scheduling “cool”: temperature-aware workload placement in data centers. 2005.

[10] R. Nathuji and K. Schwan. Virtualpower: Coordinated power management in virtualized enterprise systems. *ACM SIGOPS Operating Systems Review*, 41(6):265–278, 2007.

[11] R. Neugebauer and D. McAuley. Energy is just another resource: Energy accounting and energy pricing in the nemesis OS. In *Proceedings of the 8th IEEE Workshop on Hot Topics in Operating Systems*, pages 59–64, 2001.

[12] E. Pinheiro, R. Bianchini, E. V. Carrera, and T. Heath. Load balancing and unbalancing for power and performance in cluster-based systems. In *Workshop on Compilers and Operating Systems for Low Power*, pages 182–195, 2001.

[13] G. Semeraro, G. Magklis, R. Balasubramonian, D. H. Albonesi, S. Dwarkadas, and M. L. Scott. Energy-efficient processor design using multiple clock domains with dynamic voltage and frequency scaling. In *Proceedings of the 8th International Symposium on High-Performance Computer Architecture*, pages 29–42, 2002.

[14] R. K. Sharma, C. E. Bash, C. D. Patel, R. J. Friedrich, and J. S. Chase. Balance of power: Dynamic thermal management for internet data centers. *IEEE Internet Computing*, pages 42–49, 2005.

[15] Y. Song, H. Wang, Y. Li, B. Feng, and Y. Sun. Multi-Tiered On-Demand resource scheduling for VM-Based data center. In *Proceedings of the 2009 9th IEEE/ACM International Symposium on Cluster Computing and the Grid-Volume 00*, pages 148–155, 2009.

[16] S. Srikantaiah, A. Kansal, and F. Zhao. Energy aware consolidation for cloud computing. *Cluster Computing*, 12:1–15, 2009.

[17] 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*, pages 243–264. Springer-Verlag New York, Inc., 2008.

[18] H. Zeng, C. S. Ellis, A. R. Lebeck, and A. Vahdat. ECOSystem: managing energy as a first class operating system resource. *ACM SIGPLAN Notices*, 37(10):132, 2002.