CAESARA - Combined Architecture for Energy Saving by Auto-Adaptive Resource Allocation

Daniel Versick, Djamshid Tavangarian

Research Group Computer Architecture
University of Rostock
Joachim-Jungius-Str. 9
18059 Rostock
daniel.versick@uni-rostock.de
djamshid.tavangarian@uni-rostock.de

Abstract: Energy consumption of data centers increased continously during the last decades. As current techniques for improving their energy efficiency are usually limited to one type of data center components, it is difficult to employ them for a holistic optimization which may utilize synergies between all components. This paper introduces architecture and working principles of the novel energy management system CAESARA that applies a holistic view for evaluating an energy-efficient virtual machine placement in current virtualization environments for minimizing the number of running energy-consuming physical machines. In contrast to existing solutions, our placement algorithm does not only aggregate load to a minimum number of servers - it also discovers which servers act as most energy-efficient migration targets in conjunction with their environment (e.g. their cooling equipment). This paper presents the architecture of CAESARA and defines properties for an energy-efficient virtual machine placement algorithm. As the placement algorithm uses an energy-aware metric, it is the basis for an innovative system, which may optimize the energy efficiency of data centers in a holistic way by including any infrastructural energy consumption.

1 Introduction

The global energy consumption of IT equipment is steadily rising and produces an increasing portion of global energy production. Currently, data centers consume about 1.5 % of global electricity production, whereby their total energy usage has almost tripled between 2000 and 2010 (which corresponds to a rise of about 12 % per year) [Koo08][Koo11]. The increasing demand of computational power, especially in current Grid- and Cloud-Computing environments, is an important reason for the rising number of running computers as well as increasing energy consumption.

Due to the gain of carbon dioxide pollution in our environment, rising energy costs and the future high-growth IT market, approaches are needed to improve energy efficiency of the entire IT. Some methods, e.g. effective reuse of waste heat, correct dimensioning of IT components and employing energy-efficient hardware are state-of-the-art. Demand-driven adaptation of some IT resources is already deployed in a few areas of IT (e.g. computational resources in Cloud Computing). However, a holistic approach for demand-driven adaptation of all data center components, which would benefit from synergies between the components, is still missing. Such an approach opens a challenging field of research as the complexity of the underlying algorithms is tremendous.

In the present paper a novel approach is introduced which may be the basis for such a holistic adaptation of data center components depending on load situation within the data center. We propose the architecture of an energy management system (CAESARA - Combined Architecture for Energy Saving by Auto-adaptive Resource Allocation), that adapts performance and power consumption of all server systems including their air conditioning. A uniform metric quantifying an optimum power condition while considering all components is an important prerequisite for such an integral control. This contribution will firstly introduce the general architecture of CAESARA. Later it focuses on properties for a placement algorithm which dynamically aggregates virtual machines on few servers by employing real energy consumption of physical and virtual machines as metric and thus, is the basis for even incorporating the energy consumption of data center infrastructure devices such as air conditioning systems.

2 Related Work

To improve the energy efficiency of today's data centers, there are couple of methods that can be grouped into eight dimensions according to recommendations of the German Federal Ministry of Environment, Nature Conservation, and Nuclear Safety (BMU) [Fed10]: (1) applications and data, (2) virtualization, (3) IT hardware, (4) uninterruptable power supply, (5) air conditioning, (6) building planning and heat reuse, (7) purchasing power, and last but not least (8) management and accounting of energy efficiency. As methods for improving the energy efficiency in these eight fields are usually proposed independently of each other, currently a holistic optimization is complex and only feasible with appropriate knowledge by the data center operator. CAESARA has the goal to acquire such a holistic optimization automatically and to employ load migration techniques in conjunction with cooling power estimation. Existing approaches that only use load migration for improving energy efficiency are presented in the following.

Energy-efficient placement of virtual machines in large-scale Cloud Computing environments has become an essential research problem. One solution - EnaCloud [LLH$^+$09] - allows dynamic application live placement by considering energy efficiency in a Cloud platform. It supports application scheduling and live migration to minimize the number of running machines for saving energy.

In [VAN08] the authors present a framework, named pMapper, that tries to aggregate virtual machines onto a few servers by considering migration costs and estimated electrical energy consumption after migration. The authors conclude that guessing the future energy consumption is very difficult and vague. Thus, they use a simplified energy model not taking real energy consumptions into account. This seems to be a good approach but as we will show later in this contribution, it prevents a holistic optimization solution.

Current hypervisors usually have some add-ons for better utilization of resources in portfolio, e.g. vMotion by VMware is designed for right-sizing the IT infrastructure in case of dynamically changing demand [VMw08]. The tool's algorithm for target decision considers load of source and destination physical machine and free resources (memory and CPU) at the destination host for a best possible live migration of virtual machines.

In analogue to VMware, XenServer by Citrix [bCS12] reduces power consumption by dynamic server consolidation. XenServer's Host Power Management takes advantage of embedded hardware features to reduce data center electricity consumption by dynamically consolidating virtual machines on few systems and then powering off underutilized servers.

3 Energy Management System CAESARA

The work presented in this contribution is part of a joint project between industry and the University of Rostock. Goal of this project is the improvement of data center energy efficiency by enhancing utilization of physical data center components using dynamic load migration between devices within the data center (e.g. live migration of virtual machines between physical server systems). Depending on load and environmental parameters within the data center, the best placement of virtual machines (VMs) and storage data and a suitable configuration of the air conditioning systems as well as systems of the building automation is calculated in a dynamic and holistic way [VT10].

For collecting all necessary data as well as for calculating such an energy-efficient configuration, a central CAESARA appliance is used, which is a dedicated system within the data center. This contribution introduces the CAESARA architecture and especially focuses properties of a placement algorithm for computational load using a metric that estimates real energy consumption of server systems as well as data center infrastructure based on load situations.

In the following, data center resources are shortly presented as they may be observed and controlled by the CAESARA architecture.

Computing Resources In this context computing resources are mainly server systems. Server usually consume up to 40 % of total energy used in data centers [Fed10]. CAESARA employs state-of-the-art virtualization solutions for controlling and migrating load between computing resources. It acquires load information by using a small agent software (as a distributed part of CAESARA) that is installed within the hypervisor of every physical server. Thus, CPU and memory load of any virtual machine (VM) is collected by the operating system interface in order to forward this data to the CAESARA appliance. Additionally, power consumption measurements are used to acquire the power consumption of the physical system in dependency of load. After finding an energy-efficient placement of VMs by the CAESARA appliance, the agent receives data packets in return that instruct a load redestribution by live migration to other computers. Live migration enables VMs to be transferred at runtime to the calculated target servers while transparantly continuing the execution.

The agent software is based on *libvirt*, which is an abstraction layer for a large number of virtualization solutions. Thus, it runs on top of the most popular hypervisor systems, such as Xen [Xen12], KVM [bVMfLoxh12], and VMware [VVSfDfPS12].

Any unused servers, which don't execute VMs anymore, are then be switched off or suspended to energy-saving modes. This is done by the agent software after getting a suspend data packet from the CAESARA appliance. In case of increasing load, CAESARA reenables these computing resources. Currently, this is achieved by employing Wake-on-LAN packets, sent by the CAESARA appliance, which may wake up a computer when receiving it over the network. Other wake-up-solutions deploy network-controllable power distribution units to switch on server systems.

Air Conditioning Systems Current cooling systems consume up to 50 percent of the total energy consumption of data centers [Fed10]. CAESARA manages current load information of all computing resources and thus, realizes an estimation of energy consumption of these resources based on knowledge about their energy consumption as a function of load. As the majority of the electrical power of IT systems is converted into heat, the IT energy consumption correlates with heat production. Having further knowledge about the physical environment of IT systems (server area size, distribution of IT resources on this area, and room height), which is stored in a local administrator-managed database, CAESARA can estimate the amount of currently necessary cooling capacity of the air conditioning system for integrating into an energy-efficient VM placement algorithm as it is proposed in the next Section. Furthermore, the cooling estimation can be used to properly adapt the cooling power of air conditioning systems to the current demand.

4 CAESARA Software Architecture

The software architecture of CAESARA generally consists of two components: one E-Server for the data center which is a central software system running on the CAESARA appliance and a couple of E-Clients for any data center component which are e.g. the upper-mentioned agents running on the hypervisor hosts. Both components communicate using a web service interface which is adapted to the requirements of embedded systems (WS4D [Web12]). An E-Client collects load information and controls the power characteristics of its corresponding component. For computing ressources the E-Client normally is executed as software within the virtualization hypervisor. The E-Server stores all collected information in a database and computes an energy-optimized load placement. The evaluation of that placement is acquired by the software component Decision Agent as shown in Figure 1. As it clearly can be seen, the data flow between E-Clients and E-Server is circular resulting in an iterative optimization.

For optimizing the energy consumption of a data center in a holistic way, CAESARA is

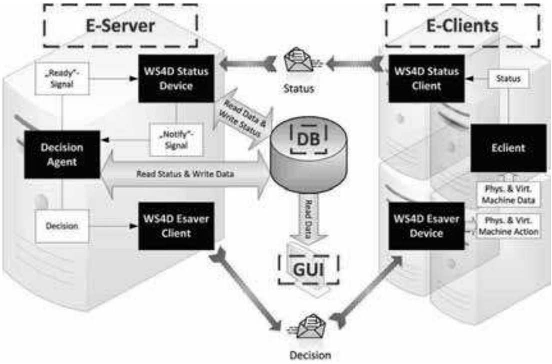

Figure 1: CAESARA Software Architecture

based on a configuration algorithm, referred to as Live Load Placement Algorithm (LLPA) which uses current load information of virtual machines to estimate an energy-efficient VM placement whereas estimated cooling power may also be included in the VM placement strategy. This holistic approach might find the best VM placement even in data centers with varying air-conditioning concepts in different rooms.

4.1 Properties of Live Load Placement Algorithm

In the following three important properties of the LLPA are presented and described:

Energy-aware Metric Current solutions for dynamically aggregating load within a data center and reducing the number of running hosts usually employ the free CPU and memory capacities of hosts and the demand of virtual machines for deciding which host is the best suiting one for a live migration. This is a way of reducing the energy consumption but it definitely does not find the most energy-efficient configuration as it even does not look at the energy consumptions at all. There may be another host that consumes less energy when executing that VM. CAESARA solves this problem by acquiring real power consumption measurements for decision. This property is called energy awareness. Furthermore, when employing real energy consumption for evaluating a data center configuration, it is easier to integrate other data center components in a holistic way. A measurable energy amount is consumed by every component whereas e.g. load characteristics of different device types are rarely comparable.

Consumption-estimative As any change of configuration in the data center produces costs (e.g. additional energy consumption or lower service quality), the number of changes has to be minimized. Therefore, the configuration algorithm should evaluate the best possible configuration by estimating its goodness before change. Our placement algorithm chooses a target system from a number of possible candidates for a virtual machine migration that has the lowest estimated power consumption when executing this virtual machine. The developed energy consumption estimation mechanism for VMs uses hardware and software components to measure the actual VM loads and corresponding whole system power consumption. As the resources processor (CPU), hard disk drive (HDD), and network interface controller (NIC) have an influence on power consumption, they are observed regarding to the type and duration of its utilization by a certain VM. Furthermore, these resources can be analyzed relatively easy by the E-Client on the hypervisor without needing to install an agent to the guest virtual machines. Other components like memory accesses also influence the energy consumption and will be respected by the presented approach to a certain degree. Based on individual energy models for each VM, their energy consumption as part of the total measured power consumption of the system can be calculated.

Total power consumption of a computer system comprises a CPU (P_C), HDD (P_H), NIC (P_N), and idle part (P_{Idle}) as well as some other components' consumptions which are firstly ignored. The latter corresponds to energy consumption while the system has no load. Thus, the power consumption can be seen as a sum of the partial consumptions of its components. As the consumption of components is not always a simple function of the load, our energy model approximates the energy consumption by polynomial regression methods with varying degrees of the polynomial function. Independent variables of our regression function are observable resource utilizations (CPU: x_C, HDD: x_H, NIC: x_N). This principle of modeling energy consumption is illustrated in the following equation.

$$\begin{aligned} P_{Sys} &= P_C + P_H + P_N + P_{Idle} \\ &= a_{1,C} * x_C + a_{2,C} * x_C^2 + ... + a_{m,C} * x_C^m + b_C \\ &\quad +a_{1,H} * x_H + a_{2,H} * x_H^2 + ... + a_{m,H} * x_H^m + b_H \\ &\quad +a_{1,N} * x_N + a_{2,N} * x_N^2 + ... + a_{m,N} * x_N^m + b_N \\ &\quad +P_{Idle} \end{aligned}$$

For estimating the power consumption of virtual machines in dependency of their load, our estimation system acquires system related parameters during two phases. The first phase (*init phase*) has to be started on every host before running any VM on the hosts (thus, at startup time of the physical machine) and is deployed to evaluate the $a_{i,j}$-parameters for the physical machine (PM). These *PM specific parameters* specify the power consumption of a physical machine in dependency of its load regardless of any virtual machine running on that physical host. Thus, after the first phase, a rough estimation of power consumption can be calculated as the overall load of a physical host can be estimated to a certain degree by knowledge about the load characteristics of all virtual machines. A more detailed view on power consumption is assembled by constantly measuring the energy consumption of the physical systems in dependency of the load situations within the virtual machines during the second phase (*runtime phase*). Thus, the same parameters, which are acquired for a physical machine during the init phase, are now calculated for any virtual machine on every host (*VM specific parameters*). Further details about the algorithm can be found in [WVT12].

For validating the accuracy of our power estimation we performed measurements on various platforms. Figure 2 shows the results measured on an AMD Phenom II X6 1090T (6 Cores @ 3,2 GHz) with 12 GByte DDR3 main memory. The bottom picture of Figure 2 depicts the estimated and measured energy consumption of a physical host while changing the load of three hosted virtual machines. The load variations of the virtual machines are shown in the upper picture of Figure 2. In our test measurements estimated (P_{EST}) and measured (P_{MEAS}) consumption differ by only 1.1 % in average. Thus, consumption estimation is a valuable way to reduce the number of necessary live migrations by computing an energy-efficient VM placement before actually configuring it. It is therefore a necessary property of an energy-efficient VM placement algorithm.

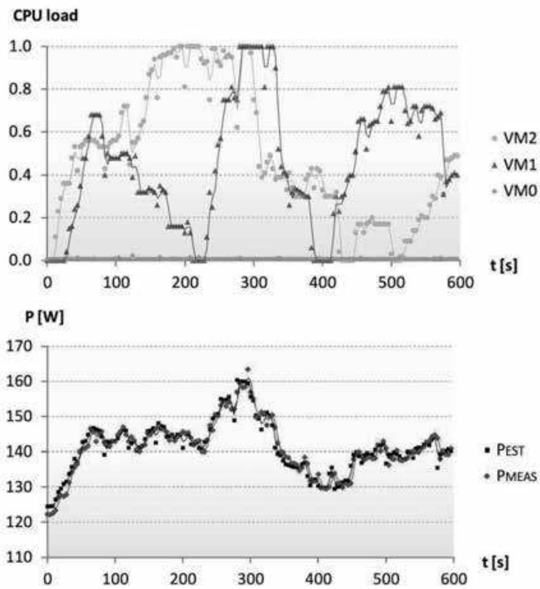

Figure 2: Predicted and measured energy consumption of a server hosting virtual machines with changing load

Proactive When only substantiating configuration information on current load situations, sudden load increase may lead to an overload problem. This problem can be reduced by predicting future load situations. This can be achieved e.g. by using autocorrelation techniques which detect cycles within the load. Such cycles are the basis for future load estimation.

CAESARA currently implements an energy-aware and consumption-estimative placement algorithm for finding an energy-efficient placement of virtual machines. As the placement problem of virtual machines can be reduced to the bin-packing problem known from complexity theory, it is NP-hard. Thus, evaluating the most energy-efficient placement leads to high computation effort which may not be feasible in every data center. The problem size (number of physical and virtual machines) may be quite high especially in current large data centers. For reducing this complexity, CAESARA implements a greedy heuristic (LLPA - Live Load Placement Algorithm) which is shortly described in the following.

The LLPA is a greedy algorithm which calculates the next system's state depending only on the current state. Thus, if a new virtual machine is started the system evaluates which of the currently running hosts has the lowest additional energy consumption when executing this virtual machine. Basic conditions like the free CPU and memory capacity have to be taken into account as well before starting a live migration. Eventually, a new physical host has to be started up in case that all running servers are highly loaded. The calculation of energy consumption may also include secondary energy consumptions as e.g. energy which is necessary for cooling the hardware. Furthermore, the system checks for any running VM after a user-defined time span if it is still running on the most energy-efficient host. This has to be done as load variations in the VMs may result in a necessary configuration changes.

5 Energy-Saving Potential

A first version of the CAESARA VM placement considering real energy consumptions of servers in conjunction with CPU and memory capacities was currently tested in a non-productive test environment with up to 5 server systems. Because of the dynamic on-demand allocation of novel server systems and the used live migration techniques it was possible to increase the load on the machines up to 95 % without getting into trouble. As conventional virtualized data centers operate the server systems at about 70 to 80 % load, this corresponds to a load increase of about 20 % using our approach which leads to about 20 % less server systems and the appropriate cooling power. However, until now only synthetical workloads have been employed which leads to the question if the results are transferable to real world scenarios. In future, it has to be analyzed how the algorithm behaves with real world workloads as they may have other load variations.

6 Conclusion

This contribution introduced a novel approach for holistically optimizing data centers by dynamic on-demand adaptation of resources. As current energy-optimizing systems only consider dedicated resource types, CAESARA tries to integrate all components into one metric.

It has been demonstrated that the complexity of such a system is tremendous which usually leads to NP-complete optimization problems. It has also been shown a greedy heuristic for VM placement that uses an energy-aware and consumption-estimative metric with a complexity far lower than NP. This metric finds a local optimum for physical hosts that are executing the currently running virtual machines in an energy-efficient manner. As the heuristic uses real energy consumption values for evaluation, an easy integration of secondary energy consumption of the data center infrastructure is possible.

Currently CAESARA has only been tested with synthetic workloads. For a significant valuation of energy-savings using CAESARA the test has to be repeated with real workload scenarios. This will be one of our next steps within this project.

Acknowledgement

This work was realized in context of the project "Improving Energy Efficiency of Data Centers" which was funded by the Ministry of Economics, Labor and Tourism of the German Federal State Mecklenburg-Western-Pomerania under grant number V-630-00008-2010-/044/047 by financial means of ESF and ERDF. We thank the funding authorities for their support and our industry partner INR Informationssysteme Rostock for the kind and fruitful cooperation in our joint project.

References

[bCS12] XenServer by Citrix Systems. Website. http://www.citrix.com/English/ps2/products/subfeature.asp?contentID=2300458, 2012.

[bVMfLoxh12] KVM: Kernel based Virtual Machine for Linux on x86 hardware. Website. http://www.linux-kvm.org, 2012.

[Fed10] Federal Ministry for the Environment, Nature Conservation and Nuclear Safety (BMU). BMU-Brochure: Energy-Efficient Data Centres. Best-Practice Examples from Europe, the USA and Asia. http://www.bmu.de/english/energy_efficiency/doc/45749.php, 2010.

[Koo08] Jonathan G. Koomey. Worldwide Electricity Used in Data Centers. *Environmental Research Letters, vol. 3, no. 034008*, 2008.

[Koo11] Jonathan G. Koomey. Growth in Data Center Electricity Use 2005 to 2010. *Analytics Press*, 2011.

[LLH+09] B. Li, J. Li, J. Huai, T. Wo, Q. Li, and L. Zhong. EnaCloud: An Energy-Ssaving Application Live Placement Approach for Cloud Computing Environments. *Proceedings of the 2009 IEEE International Conference on Cloud Computing, CLOUD 09, pages 17-24*, 2009.

[LPGM08] A.W. Leung, S. Pasupathy, G. Goodson, and E.L. Miller. Measurement and Analysis of Large-scale Network File System Workloads. In *USENIX 2008 Annual Technical Conference on Annual Technical Conference*, number June, pages 213–226. USENIX Association, 2008.

[MVT12] Robert Meschke, Daniel Versick, and Djamshid Tavangarian. Energy-Efficient Storage System. In *Proceedings of the 17th International SDPS Conference*. SDPS, 2012.

[VAN08] Akshat Verma, Puneet Ahuja, and Anindya Neogi. pMapper: power and migration cost aware application placement in virtualized systems. In *Proceedings of the 9th ACM/IFIP/USENIX International Conference on Middleware*, Middleware '08, pages 243–264, New York, NY, USA, 2008. Springer-Verlag New York, Inc.

[VMw08] VMware. How VMware Virtualization Right-sizes IT Infrastructure to Reduce Power Consumption. 2008.

[VT10] Daniel Versick and Djamshid Tavangarian. Reducing Energy Consumption by Load Aggregation with an Optimized Dynamic Live Migration of Virtual Machines. In *2010 International Conference on P2P, Parallel, Grid, Cloud and Internet Computing*, pages 164–170. Published by IEEE, November 2010.

[VVSfDfPS12] Servers VMware Virtualization Software for Desktops, Virtual Machines for Public, and Private Cloud Solutions. Website. http://www.vmware.com, 2012.

[Web12] Website. Webservices for Devices. http://ws4d.e-technik.uni-rostock.de/, March 2012.

[WVT12] Ingolf Wassmann, Daniel Versick, and Djamshid Tavangarian. Energy Consumption Estimation of Applications. To be published in Proceedings of the 28th Annual ACM Symposium on Applied Computing. 2012.

[Xen12] Xen.org, home of the Xen hypervisor, the powerful open source industry standard for virtualization. Website. http://www.xen.org, 2012.

[ZPL01] Y. Zhou, J.F. Philbin, and Kai Li. The Multi-queue Replacement Algorithm for Second Level Buffer Caches. In *Proceedings of the General Track: 2002 USENIX Annual Technical Conference*, pages 91–104, 2001.