Expansion of Data Center's Energetic Degrees of Freedom to Employ Green Energy Sources

Stefan Janacek[1], Wolfgang Nebel[2]

Abstract

Rising power consumption of data centers is a topic of great concern and therefore several power saving technologies exist. This paper describes the idea of a data center overall power saving and controlling strategy, allowing the data center to enter optimized minimum power states but also to control its own power consumption to apply demand response management. Therefore, the degrees of freedom a virtualized data center has are modelled and the methodology used to control its energy state is described, taking into account the IT hardware like servers and network gear as well as the influence of cooling devices and power distribution devices. In the paper, we describe our models used for our simulations, the methodology and the power saving potential of our system. We formulate the problem to control the data center's power consumption by applying different consolidation strategies as an extended bin packing optimization problem, where virtual machines must be packed on a specific number of servers. External constraints like the time-flexibility of the solution and the influence on supporting devices are applied by using cost functions. We present a greedy solver for this problem and show first results and the potential of the entire approach.

1. Introduction

Information and Communication Technologies (ICT) and especially data centers play a significant role in our today's world. Growing markets as cloud computing and on-demand services fortify this trend. As a result, the power demand of ICT components, including data centers, kept on rising during the last few years. Since the energy costs have also become a major economical factor, power saving and efficiency technologies for data centers have emerged. Among them are technologies like virtualization [1], server consolidation [2], and application load scheduling to times of lower energy prices [3, 4]. A fairly recent trend is to enable the data center to benefit from renewable energy sources [5 - 8], allowing it to operate at full load in times of high availability and cutting its load otherwise. Unfortunately, this methodology needs to alter the running applications, stopping their execution in the worst case or it needs a network of connected data centers in different geographical locations. This approach may be viable for many scenarios; however, often this is not possible. Instead, this paper proposes the idea of expanding the degrees of freedom a data center already has without altering any of its running applications. The goal of the methodology proposed in this paper is to let the data center mostly operate in a minimum energy state; however, to allow demand response management, it should be able to enter a specific energy state, hence be able to control its power consumption. This could, for example, be used to follow an external power profile induced by a Smart Grid. There are also data center internal motivations to control the power consumption of a data center subspace (for example a room or a cage). Especially when load balancing techniques are applied, a data center may have significant diverse power states in different rooms, leading to inefficient global device states or even harming the grid stability. In these cases, the possibility to control the power consumption of a subset of the data center's devices can become necessary. For this, the data center's existing degrees of freedom are identified and expanded to be able to reach a high power consumption variability, while still

[1] OFFIS e.V., 26121 Oldenburg, Germany, stefan.janacek@offis.de

[2] University of Oldenburg, 26111 Oldenburg, Germany, wolfgang.nebel@uni-oldenburg.de

keeping the applications unchanged. A key aspect is the modeling and description of interdependencies between different device categories in data centers. A base technology for the methodology proposed is server virtualization that enables the data center to live-migrate virtual machines (VM) across different physical machines (PM, the terms *physical machine* and *server* in this paper mean the same thing). By using this technology, the migration of running applications encapsulated in VMs to different PMs can be used to intensely influence the server's power consumption, also affecting the amount of cooling and UPS load needed, thus changing the entire power consumption of the data center. Here, very dense states with minimal power consumption are possible, as well as loosely packed states with a higher consumption but also with increased flexibility. The methodology assumes a virtualized data center, where the following operations are allowed as adjustments: 1) VM migrations, 2) server switch-ons, 3) server switch-offs. Each action takes a specific amount of time. In this paper we address the problem of finding a suitable VM allocation on the existing servers to either enter a minimal power state or to enter a state approaching a specific power demand while taking into account possible side effects that may occur in combination with the data center's hardware devices and the time it needs to enter this state. We formulate the problem as an NP-hard extended bin packing problem [9] with a global cost function, where VMs (items) must be allocated to the PMs (bins). This approach is not new [10]; however in the approach presented here it is not always optimal to just minimize the number of active PMs to reach the desired state.

To the best of our knowledge, this is the first approach that researches a methodology that is able to control the data center's power demand using these adjustments while also taking into account interdependencies of the data center's hardware devices. To be able to evaluate the specific load states of a data center in terms of power demand, a data center simulation is presented that models the server's power demand, the efficiency of uninterruptible power supply (UPS) devices, cooling power demands via approximating meta-models and network flows in a sample simulated data center. The rest of this paper is organized as follows: Section 2 lists the related work, in Section 3 the models and architecture of the simulation is described, while Section 4 shows the problem formulation and the methodology used. In Section 5, we present first results and analyze the potential of the approach. We conclude in Section 6 and describe our next steps and further research.

2. Related work

The area of research this paper addresses is also focus of other researches. General server power models can be found in [11, 12] while [13] already proposes additional models for racks and cooling units. Energy models for data centers are found in [14, 15]. Our research partly bases on these results. In [5], the authors propose the idea to combine a data center with a local power network that includes renewable energy sources. Such a power network is, however, less complex than a smart grid, since it only consists of power producers. The authors also cover the aspect of the intermittency of these power producers. They propose to shift the work load to other data center locations, each profiting from individual energy advantages. A similar approach is covered in [6], including weather conditions at different locations. [16] proposes a *service request routing* for data centers to distribute the load according to the electric grid in a smart grid. In [7], the authors present the idea of a carbon-aware data center operation. They propose three key ideas to implement this concept: on-site and off-site renewable energies and *Renewable Energy Certificates (REC)*. In our research, the usage of RECs is, however, not a legitimate concept. Modeling of thermal behavior of data center components, especially of servers, has been researched before. In [17], the thermal load of processors and micro controllers is considered. [18] handles thermal predictions of processors and combines it with a Dynamic Voltage and Frequency Scaling (DVFS) technique.

Thermal modeling of a server rack is arranged in [19]. [20] presents a dynamic model for the temperature and cooling demand of server racks that are enclosed in hot aisle containment. The correlation of power consumption and temperature of server internal coolers is investigated in [21]. As a result, the authors state that it is possible to save power under certain conditions, when the Computer Room Air Conditioning (CRAC) adapts itself to a higher temperature level and the server coolers compensate this by applying a higher rotation frequency. They also model the time that cool air needs to travel from CRAC units to a specific server rack. However, a detailed correlation to server load is not handled. [22] handles the planning of VM migrations under consideration of VM interdependencies like communication, security aspects and other SLAs. These are not considered in this paper, since it aims at showing the concept to maximize the degrees of freedom. However, the methodology proposed here can easily be adopted to also support VM interdependencies, if needed.

3. Models and simulation architecture

The methodology described in this paper uses a data center simulation that is able to model the power consumption of IT hardware, in this case the PMs, and the supporting devices such as cooling and UPS devices. Figure 1 shows the architecture of the information flow of the models used for the simulation of the hardware devices in the data center. The simulated data center that is used for the evaluations in this paper consists of 960 PMs in 96 racks with 8 UPS devices. The devices are located in two different rooms. For each simulation, the number of VMs is static, meaning there are no VMs coming into the simulation or leaving it.

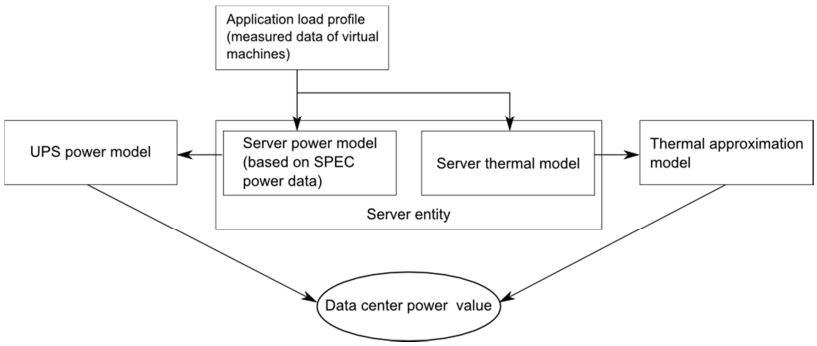

Figure 1: Model architecture for the simulation of the data center's hardware devices

The application load profiles of the VMs consist of load measurements of real applications hosted in a mid-sized data center.

Server models and application load

The simulated data center consists of heterogeneous server models, the model data is based on the publicly available results of the SPEC power benchmark and on own measurements [11]. The total power consumption of a PM is split into two parts: the minimum static power consumption P_{st} that describes the consumption in idle mode and the dynamic power consumption P_{dy} that is influenced by the utilization of the PM. As an indicator of the utilization, the CPU load is used as the only value; it has already been shown that it has strong correlations with the power consumption [8, 11]. The total power consumption of a PM is given by:

$$P_t = P_{st} + P_{dy}$$

The CPU utilization of a PM is calculated by adding all of the VM's utilizations at each instant of time. Let n be the total number of VMs on a server at an instant of time and $C_{VMi}(t)$ the CPU utilization of the VM i, the total CPU utilization $C_{PM}(t)$ of the PM at the time t is calculated as:

$$C_{PM}(t) = \sum_{i=1}^{n} C_{VMi}(t)$$

Our measurements showed that the variability in RAM allocations is very small; hence it is assumed that each VM has a static memory allocation. This value is retrieved by finding the maximum RAM allocation the VM had during the measurement duration. Each PM can operate a maximum number of VMs at each instant of time; this number is limited by the resource usage of each VM. Relevant values are the CPU load $C_{VMi}(t)$ at each time t and the RAM allocation M_{VMi} (as this value is static, is has no reference to time), where these in sum must not exceed the PM's physical resources C_{PM} and M_{PM}:

$$\forall\, t\colon\ C_{PM} \geq \sum_{i=1}^{n} C_{VMi}(t)\ \ and\ M_{PM} \geq \sum_{i=1}^{n} M_{VMi}$$

The RAM allocation of VMs forms a hard and static boundary regarding the maximum number of VMs of each PM. Overprovisioning of RAM is not assumed. Finally, the total power consumption of a PM $P_{PM}(t)$ is calculated using the power models published in [8, 11] using the CPU utilization of the PM at the time t.

VM allocation state

A VM allocation state A defines the power state of each PM (on or off) and for each PM that is powered on the list of VMs hosted on this PM. A state is legal, if all VMs can access the resources they need for their operation at the current time. To cross from one state to another, VMs will be migrated and PMs can be switched on or off respectively.

UPS models

The data center simulation uses a basic UPS model scheme that evaluates the efficiency for a specific UPS device. For most UPS, the efficiency increases with rising load. Hence, the UPS should always be operated with the best efficiency factor, for example, at least with 80% load. The methodology proposed in this paper uses the UPS model to find an allocation that leads to an improved UPS efficiency factor, compared to other methodologies that do not consider UPS power consumption. It is assumed that each UPS device has at least a minimum power consumption P_{Umin}, even if the devices (servers) attached to it are powered off. It is also assumed that UPS devices are not turned off if unused. Regarding this information, we formulate the following UPS power model that is used for the data center simulation: Let P_U be the total power consumption of all devices the UPS powers (servers) including the UPS device's own consumption and P_D the power consumption of all devices attached to the UPS. The efficiency factor function $i(P_D)$ defines the UPS efficiency at the power load P_D. Then $P_U(P_D)$ can be calculated as:

$$P_U(P_D) = \begin{cases} P_{Umin}, & if\ P_D < P_{Umin} \\ P_D + \left(1 - i(P_D)\right) * P_D, & else \end{cases}$$

Thermal models

The thermal models needed in this simulation need to evaluate 1) the power consumption of the cooling devices depending on the workload of servers in different data center locations (room, racks, cages) and 2) the time period the air takes to flow from the server outlets to the air-cooling device and the CRAC units need to adapt itself to the new heat situation. The main challenge is to develop fast models, since traditional (and accurate) approaches like computational fluid dynamics (CFD) simulations are too slow for the needs in this simulation. Therefore, a similar methodology as described in [23] is used. It is assumed that the heat produced by the servers Q is equal to the power consumption of these devices so that $Q = P_{servers}$. Based on these models, we define the

function $P_{th}(P_{servers})$ that calculates the needed cooling power for a given server power consumption at the time *t*.

Network topology model

The network of the simulated data center is modeled as a graph while the used topology is VL2 (see Figure 2). It is assumed that the network connections between the different switch layers have different bandwidth sizes, allowing different amounts of parallel network traffic. In this paper, the network graph is used to determine the amount of live-migrations of VMs that can be performed in parallel. To be able to reach a different VM allocation state, often several migrations will occur; if most of them can be run in parallel, the target allocation state can be reached in less time.

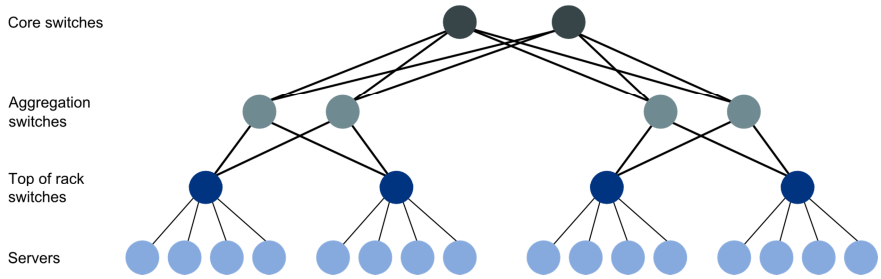

Figure 2: Network topology (VL2)

The following rules apply to parallel migrations: 1) Each PM may only be the source or the target of one migration at the same time. 2) Each switch node in the network graph can only handle as much migrations so that the maximum bandwidth is not exceeded for more than 50%. In our model, this ensures that the running applications can still access the network safely; in other network scenarios this value might be changed according to the real conditions. When a new allocation state should be entered or evaluated, our algorithm calculates the needed migrations to cross from the current state to the new state and finds its involved PMs and their network paths respectively. It also evaluates which PMs can be switched off or have to be switched on. It then creates a migration plan where as much parallel migrations as possible are scheduled. Based on this information, the algorithm calculates the amount of steps $s(A_{current}, A_{target})$ that is needed to migrate from the current allocation state $A_{current}$ to the new state A_{target} where each step takes a constant amount of time (defined by the duration of migrations and server switches).

4. Problem formulation and methodology

The goal of the presented methodology is to let the data center migrate from a current allocation and power state to another state with a specific power consumption, either a minimal or a given consumption under the consideration of the time it needs to enter the desired state. As stated in the introduction, we formulate the problem as a combinatorial NP-hard multidimensional bin packing problem with a cost function. The classic one dimensional bin packing problem aims to distribute a number of items into a finite number of bins where the optimization goal is to minimize the number of bins used. However, applying this approach to the problem described here may lead to inefficient solutions. If the methodology just minimizes the number of servers, power savings will occur for the IT hardware but not for the supporting devices like UPS and cooling. These may run into significantly inefficient states, destroying the savings achieved by switching off servers. Similarly, if a specific power consumption should be approached, the modifications caused by the reactions of the supporting devices may lead to severe deviations. To eliminate these problems, a new approach is presented that still uses the bin packing representation of the problem; however, instead of trying to minimize the amount of bins used, a cost function is used to rate the effectiveness of the entire solution regarding power consumption and the time needed to reach the

new state. The formal definition of the problem is as follows: Given is a set $V = \{v_1, \ldots. v_m\}$ of VMs in the data center with resource demand vectors $r_{v1}, r_{v2}, r_{v3}, \ldots, r_{vm}$ and a set $S = \{s_1, \ldots., s_k\}$ of PMs available with resource capacity vectors of $x_{s1}, x_{s2}, x_{s3}, \ldots., x_{sk}$. Find an allocation A of all elements in V to an arbitrary number δ of elements in S so that for each $s \in S$:

$$\sum_{i=1}^{j} r_{vi} \leq x_s + b$$

where b is a buffer value used to prevent overloading a PM and j is the number of VMs on the PM s. The optimization goal is, in contrast to the classic bin packing problem not to minimize the number of used PMs, but instead to maximize the fitness of the allocation $f(A)$. This function evaluates the allocation A in terms of the proximity towards the desired power consumption (minimal or target value); the time it needs to enter this allocation is then considered when a new solution is chosen. The function is presented in detail in the following.

Fitness function

To measure the fitness of each allocation, first the total data center power consumption $P_{DC}(t, A)$ under the allocation A is calculated.

$$P_{DC}(A) = P_U\left(\sum_{i=1}^{\delta} P_{PMi}\right) + P_{th}(\sum_{i=1}^{\delta} P_{PMi})$$

Next, the duration in steps to migrate from the data center's current allocation state $A_{current}$ to the solutions state $A_{solution}$ is retrieved using the network graph.

$$d = s(A_{current}, A_{solution})$$

In normal operation state, the methodology tries to let the data center operate in an energy efficient state, hence the optimization goal is to minimize $P_{DC}(A)$. The fitness function is then defined as follows:

$$f(A) = \frac{1}{P_{DC}(A)}$$

If the methodology is used to apply demand response management, target power consumption for the data center is given as $P_{DCtarget}$. In that case, the optimization goal is to minimize the deviance to the given consumption a:

$$a = |P_{DCtarget} - P_{DC}|$$

In this case, the fitness function uses a instead of $P_{DC}(A)$.

The second optimization goal is always to minimize the amount of steps d needed to reach the new allocation state, since the new state should always be reached with as few operations as possible. When two allocations are compared, first the fitness value is used and as a second condition the number of steps d is compared, for example if a solution needs a significantly lower amount of steps and the fitness is only marginally worse, this solution is preferred.

The methodology described in this paper works as follows: starting from an initial state in the data center, a first fit decreasing (FFD) algorithm is used to create a first solution. This is densely packed, but as already described not the optimal solution. This solution is used as a starting value for a heuristic search algorithm. The main challenges for this algorithm are 1) the creation of fast and convenient heuristics to evaluate each sub-step on the way to a better solution; 2) apply these heuristics to find neighbor states in the global neighborhood.

5. Analysis and potential

We evaluate a sample scenario with the data center described in Section 3. At first, the operation state of the data center is in the initial non-optimized state where each PM is powered on. In this case, the simulated data center had a power consumption of about 145kW. After applying the FFD optimization, which is analog to traditional power saving methodologies only taking the server hardware into account, the power consumption was 78kW (see Figure 3 at point **A**). However, using the algorithm that also accounts for the efficiency factors of the infrastructure devices, the power consumption could be decreased to 66kW, additionally saving about 15% energy. The algorithm's run time for this case was about 2 minutes on an Intel Core i5 (2.5 GHz) computer. At the time point **C**, a demand response request (DRR) is received, the data center enters the given power state and at time point **E**, it goes back to the optimized state (**F**).

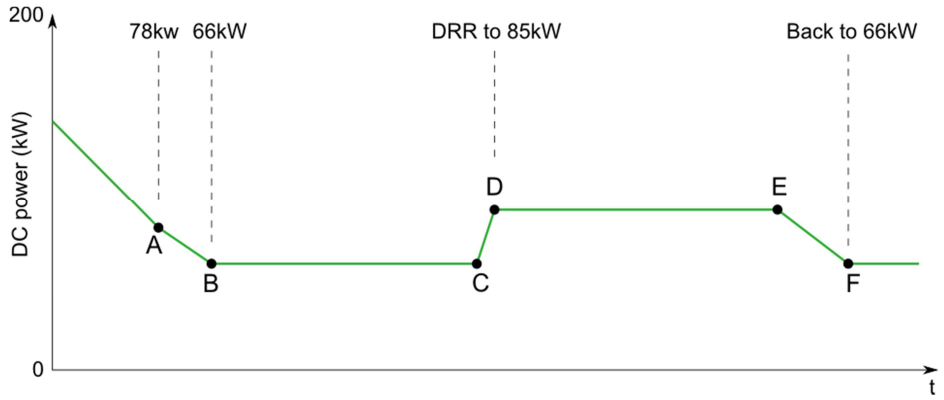

Figure 3: Schematic view of the results and the potential of the power saving and control strategy using the methodology described in this paper

The methodology described in this paper is thus not only able to reach additional power savings but also to allow the data center to apply demand response management.

6. Conclusion

In this paper, a methodology was described that allows the data center to improve its energy efficiency by taking into account the IT hardware (servers) and the infrastructure devices (UPS, cooling) when finding VM allocations. This approach leads to an additional saving potential of about 15%. The methodology is also able to find fast transitions from the current allocation state to specific power states, enabling the data center to apply demand response management. Our future research will create more detailed thermal models for different cooling strategies (free cooling, chillers, etc.) and an improved method to retrieve optimal parallel migration plans using the network model. We are also working on evolutionary algorithms to find an allocation state near the optimum while still completing in realistic time frames. Since the problem to solve is very complex and it is generally hard to determine the "real" optimum, a competitive analysis is planned for the evaluation of the algorithm. It is also planned to integrate a load forecasting method from [2] into the methodology that is used to predict VM application load, thus allowing the methodology to act proactive.

References

[1] Barham, P. (et al.): Xen and the Art of Virtualization. In: Proceedings of the Nineteenth ACM Symposium on Operating Systems Principles, pp. 164-177. ACM, New York, 2003

[2] Hoyer, M. (et al.): Proactive Dynamic Resource Management in Virtualized Data Centers. In: Proceedings of the 2nd International Conference on Energy-Efficient Computing and Networking, pp.11-20. ACM, New York, 2011

[3] Dalvandi, A., Gurusamy, M., Kee Chaing Chua (2013): Time-Aware VM-Placement and Routing with Bandwidth Guarantees in Green Cloud Data Centers, Cloud Computing Technology and Science (CloudCom), 2013 IEEE 5th International Conference on , vol.1, no., pp.212,217, 2-5 Dec.

[4] Mukherjee, T. (et al.): Spatio-temporal Thermal-aware Job Scheduling to Minimize Energy Consumption in Virtualized Heterogeneous Data Centers. In: Computer Networks, Special Issue on Resource Management in Heterogeneous Data Centers, Volume 53, Issue 17, pp. 2888-2904. Elsevier North-Holland, Inc., New York, 2009

[5] Ghamkhari, M., Mohsenian-Rad, H.: Optimal Integration of Renewable Energy Resources in Data Centers with Behind-the-Meter Renewable Generator. In: IEEE International Conference on Communications (ICC), pp. 3340-3344. IEEE, New York, 2012

[6] Zhang, Y., Wang, Y., Wang, X.: GreenWare: Greening Cloud-Scale Data Centers to Maximize the Use of Renewable Energy. In: Lecture Notes in Computer Science, pp. 143-164. Springer, Heidelberg, 2011

[7] Ren, C., Wang, D., Urgaonkar, B., Sivasubramaniam, A.: Carbon-Aware Energy Capacity Planning for Datacenters. In: IEEE 20th International Symposium on Modeling, Analysis Simulation of Computer and Telecommunication Systems, pp. 391-400. IEEE, New York, 2012

[8] Janacek, S., Schomaker, G., Nebel, W.: Data Center Smart Grid Integration considering Renewable Energies and Waste Heat Usage, Energy-Efficient Data Centers, Lecture Notes in Computer Science Volume 8343, 2014, pp 99-109, Springer, 2013

[9] Epstein, L., Levin, A.: Bin packing with general cost structures, Mathematical Programming, April 2012, Volume 132, Issue 1-2, pp 355-391

[10] Carli, T., Henriot, S., Cohen, J., Tomasik, J.: A packing problem approach to energy-aware load distribution in Clouds, arXiv:1403.0493, 2014

[11] Janacek, S. (et al.): Modeling and Approaching a Cost Transparent, Specific Data Center Power Consumption. In: 2012 International Conference on Energy Aware Computing, pp. 68-73. IEEE, New York, 2012

[12] Pedram, M., Hwang, I.: Power and Performance Modeling in a Virtualized Server System. In: Proceedings of the 2010 39th International Conference on Parallel Processing Workshops, pp. 520-526, 2010

[13] Pakbaznia, E., Pedram, M.: Minimizing Data Center Cooling and Server Power Costs. In: Proceedings of the 14th ACM/IEEE international symposium on Low power electronics and design, pp. 145-150. ACM, New York, 2009

[14] Abbasi, Z., Varsamopoulos, G., Gupta, S.K.S.: Thermal Aware Server Provisioning and Workload Distribution for Internet Data Centers. In: Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing, pp. 130-141. ACM, New York, 2010

[15] Mukherjee, T., Banerjee, A., Varsamopoulos, G., Gupta, E., Rungta, S.: Spatio-temporal thermal-aware job scheduling to minimize energy consumption in virtualized heterogeneous data centers, Computer Networks, Volume 53, Issue 17, 3 December 2009, Pages 2888–2904

[16] Mohsenian-Rad, H., Leon-Garcia, A.: Coordination of Cloud Computing and Smart Power Grids. In: Proceedings of IEEE Smart Grid Communications Conference, pp. 368-372. IEEE, New York, 2010

[17] Wei, W., Lingling, J., Jun, Y., Pu, L., Sheldon, T.: Efficient power modeling and software thermal sensing for runtime temperature monitoring, ACM Transactions on Design Automation of Electronic Systems (TODAES), Volume 12 Issue 3, August 2007

[18] Cochran, R., Reda, S.: Consistent Runtime Thermal Prediction and Control Through Workload Phase Detection. In: Proceedings of the 47th Design Automation Conference, pp. 62-67. ACM, New York, 2010

[19] Jeonghwan, C. (et al.): Modeling and Managing Thermal Profiles of Rack-mounted Servers with ThermoStat. In: Proceedings of HPCA, pp. 205-215. IEEE, New York, 2007

[20] Zhou, R., Wang, Z., Bash, C.E., McReynolds, A.: Modeling and Control for Cooling Management of Data Centers With Hot Aisle Containment. In: ASME Conference Proceedings, pp. 739-746. ASME, New York, 2011

[21] Sungkap, Y., Lee, H.-H.S.: SimWare: A Holistic Warehouse-Scale Computer Simulator, Computer, Volume:45, Issue: 9, 2012

[22] Al-Haj, S., Al-Shaer, E.: A formal approach for virtual machine migration planning, Network and Service Management (CNSM), 2013 9th International Conference on, 2013

[23] Jonas, M., Gilbert, R.R., Ferguson, J., Varsamopoulos, G., Gupta, S.K.S.: A transient model for data center thermal prediction, Green Computing Conference (IGCC), 2012 International