International Journal of Hybrid Information Technology
Vol. 9, No.12 (2016), pp. 321-332
http://dx.doi.org/10.14257/ijhit.2016.9.12.29

New Approach for Virtual Machines Consolidation In Heterogeneous Computing Systems

Jan Fesl[1,2,3*], Jiří Cehák[1] , Marie Doležalová[1,3] and Jan Janeček[2]

[1]University of South Bohemia, Faculty of Science, Institute of Applied Informatics, Branišovská 31a, České Budějovice, 370 05
[2]Czech Technical University in Prague, Faculty of Electrical Engineering, Department of Computer Science, Karlovo Náměstí 13, Prague, 121 35
[3]Biology Centre CAS, Laboratory of Analytical Biochemistry & Metabolomics, Branišovská 31, České Budějovice, 370 05
jfesl@prf.jcu.cz, jiri.cehak@outlook.cz, d.marienka@seznam.cz, janecek@fit.cvut.cz

Abstract

The energy consumption is one of the most important factors in the virtual machines deployment in the current data centres. Various studies proved that the energy aware management of the virtual machines can reduce the total energy consumption about tens of percents. We developed the new approach, based on the distributed algorithm, which is able to consolidate the virtual machines between various virtualization nodes without the central coordinator. The input data for this algorithm is collected online from the electronic wattmeters, which are placed before the energy input of each virtualization node.

***Keywords**: consolidation, virtual machine, distributed, energy aware, heterogeneous computing system*

1. Introduction

The distributed virtualization systems, in principle depicted in the Figure 1, are today used in all commercial data centres in the world. The main blocks of such systems are the virtualization nodes (VN), network area storages (NASes) and management nodes, which are interconnected via the high throughput communication network. The virtualization nodes are the high performance computers on which are the virtual machines executed. Network area storages serve for storing of the virtual machines hard disk drives images. The separated storing of the virtual images from the virtualization nodes is a necessary aspect for their efficient live migration [1]. The management node(s) serves as the central coordinator of all actions in such distributed system. The poor simultaneous coordinated execution or stopping of various virtual machines can cause that some virtualization nodes can contain many running virtual machines and can be overloaded. On the other hand, there can exist some virtualization nodes, which contain only a small amount of the running virtual machines and can be underloaded. The running of such virtual machines is much more expensive than it has to be, because they can run on the other not overloaded virtualization nodes and the underloaded nodes can be further hibernated or switched off after the latest virtual machine migration. The high energy consumption of the virtualization nodes has the direct impact on the amount of heat, which is produced. The heat production further also affects the activity of the data centre air condition system, which fundamentally participates on the total energy consumption [2].

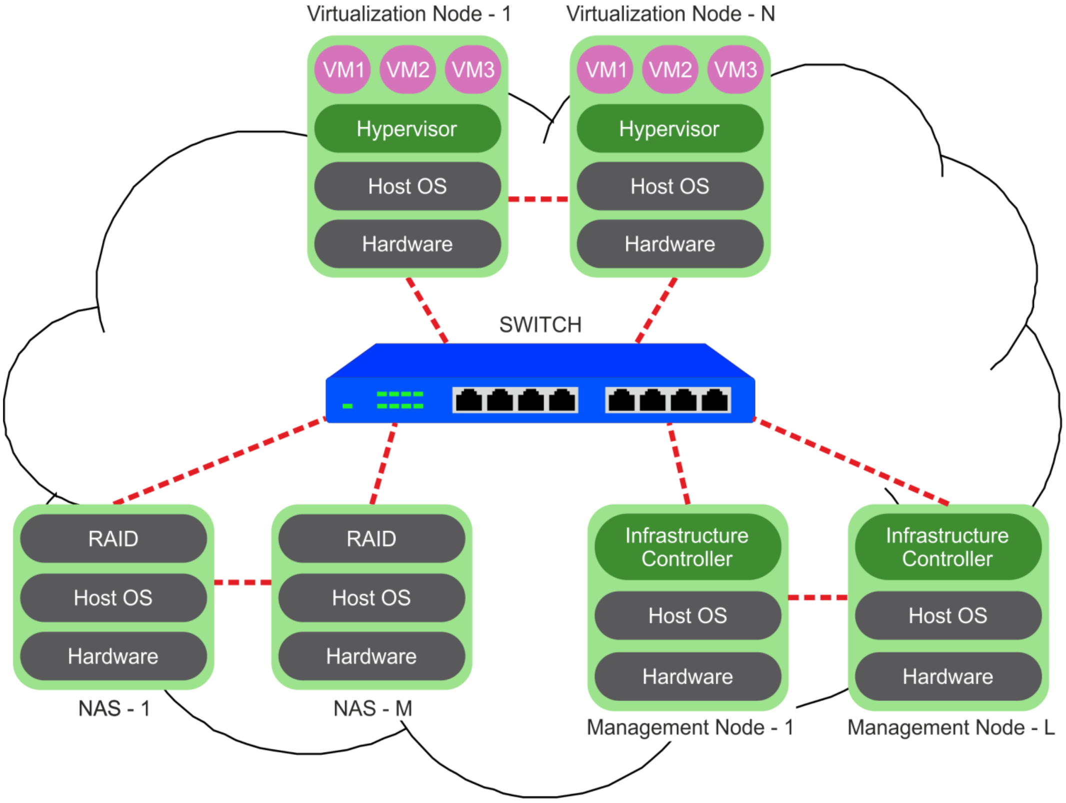

Figure 1. Common Architecture of the Distributed System for Virtualization, N = Max Count of the Virtualization Nodes, M = Max Count of the Nases and L = Max Count of the Management Nodes

Contribution of the proposed solution

We studied various factors [3] which can indicate the computer system state. The CPU utilization is often elected as the prime utilization indication factor, but its power consumption function can oscillate and has many step changes. The total energy consumption is a more proper indicator, does not influence the measurement, its function shape is much smooth and takes in care more aspects. We created a measurement system, which is able to read the energy consumption of all virtualization nodes and cooperates directly with the virtualization nodes hypervisors. This system uses the online measured data for the virtual machines consolidation. The next aspect is the heterogeneity of the virtualization nodes. It is possible, that some virtualization nodes can be older or not so energy efficient than others. We created the classification function for a real computing system, which can evaluate its energy efficiency according to its utilization level. This value is further used as the metrics in the new proposed algorithm. To avoid the central node failure, we tried to develop the fully distributed solution.

2. Related Work Overview

There have been published many papers about this topic in the last years. The first work about this topic was introduced by Nathui and Schwan [4], who proposed the virtual power management approach, which allows settings the various power management policies. Stoess [5] introduced the framework for the energy management for the modular operating systems, which contains another model for the virtual machines consolidation. Verma [6] proposed architecture of the power-aware application consolidation framework – pMapper. This framework allows setting the various scenarios for the virtual machines in use. The main goal of this framework stands in the power minimization by a fixed performance requirement. The relations between the energy consumption and resources utilization was studied by Shrikantaiah [7]. The direct relation between the power consumption and CPU utilization was in detail studied by Fan [9] and more specified by Beloglazov [10]. Beloglazov further created the complex classification and overview of the various virtual machines consolidation algorithms [11] and provided their comparison. Farahnakian [12] applied the reinforcement learning technique to improve the virtual machines consolidation efficiency. Cao [13] introduced a framework with the specific heuristics for the minimum power and maximum utilization policy, which cares about the service level agreement (SLA) for the specific virtual machine deployment. Another similar solution was proposed by Dupont [14]. The optimization for the virtual machine consolidation based on the improved genetic algorithm was published by Xu [15], another approach based on the ant colony optimization algorithm proposed Feller in [16], similar works were presented by Mills [17], Bobroff [18] and Borgetto [19]. Kansal [20] used the firefly optimization algorithm and introduced the novel migration technique. The real implementations of some algorithms exist today for the Xen hypervisor – EnaCloud [21], the implementation for the hypervisor KVM is described in [22].

3. Distributed Virtual Machine Consolidation Algorithm (DVMCA)

3.1. Virtualization Nodes Consumption Suitability Metrics

The novel proposed mechanism takes into account the heterogeneity of the virtualization nodes. The requirement for this feature is logical, because often happens, that some group of the virtualization nodes has another hardware than the other groups. That means, some virtualization nodes are more energy aware than the others. The main factor in the energy consumption is the global CPU utilization which is associated with the air condition system activity [10]. We introduced new metrics, which characterizes the node energy consumption efficiency. The total energy (TC) consumption value can be calculated as follows. The variable u is from the interval of 0-1 (0 means idle, 1 means fully utilized) and P(u) is the global power consumption value for the concrete utilization.

$$TC = \int_0^1 P(u)\,du \tag{1}$$

The efficiency (E) of the virtualization node N, which has total K physical CPU cores, can be calculated as follows.

$$E(N) = \frac{K}{TC} = \frac{K}{\int_0^1 P(u)\,du} \tag{2}$$

The function P(u) for the specific virtualization node is typically unknown. This can be experimentally measured and completed by the linear approximation. The CPU utilization can be increased between 0 and 100% by the 10% step. The equation (1) can be retransformed into this shape.

$$E(N) = \frac{K}{0.1*\sum_{i=1}^{10} P(\frac{i-1}{10})+1/2*(P(\frac{i}{10})-P(\frac{i-1}{10}))} \tag{3}$$

The consumption model is created for each virtualization node in the distributed system. The virtualization nodes (V) are then descendent sorted via their computed efficiency value. The system utilization stress can be reached by using [23] or [24].

Problems in virtual machines consolidation

1) Underloaded virtualization node detection – means the indication of the low utilized node

2) Overloaded virtualization node detection – means the indication of the high utilized node

3) Virtual machine placement - means the finding of the best node for the virtual machine execution

Tools for virtual machines consolidation

1) Virtualization node switching on/off

2) Virtualization node hibernation/wake up

3) Virtual machine live migration – means the transfer of some virtual machine to another virtualization node without stopping the virtual machine

3.2. The Distributed Algorithm Principle

The main goal of this idea is depicted in the Figure 2 below. All nodes are sorted via their efficiency. Every node contains the virtual machines consolidator (VMC). VMC is a module, which directly cooperates with the local hypervisor and remote hypervisors. The next device, which is interconnected to the VMC, is an electronic wattmeter. VMC reads periodically the current value of the power and knows its energy power consumption function course for its node. The value will be further used for the node state detection. The next algorithm steps are done simultaneously on all virtualization nodes.

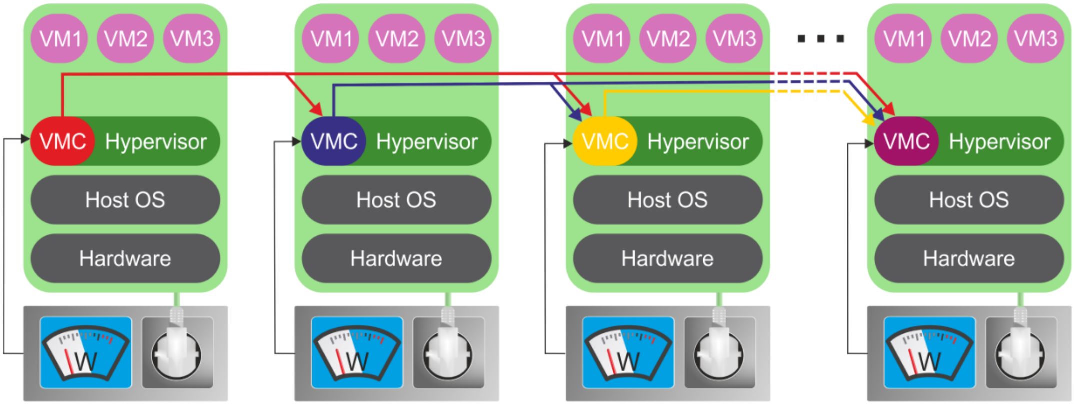

Figure 2. Distributed System Virtual Machines Consolidation Principle

Every virtualization node N knows its power consumption efficiency E value and knows all virtualization nodes, which have worse E value than it. The connection to these nodes is depicted with the lines with arrows between the virtualization nodes. If some node N with better E value is not overloaded, it can offer its resources to another node M with worse E value. The offer from N to M contains the specification of N free resources - the number of free CPU cores and size of free operating memory. If M has some virtual machines, which can be run under the offered resources of N, all these machines are migrated from M to N. N selects the M node, which has the worst possible E value.

During the migration phase of the virtual machines N and M mustn't accept offers from other virtualization nodes. If M further contains no running virtual machines, then it can be switched off or hibernated.

The solution can be further optimized as follows. After the consolidation finishing, it can happen, that some nodes with high E value, but with low number of CPU cores are fully utilized and other nodes with high number of CPU cores and lower E value are not fully utilized. N therefore asks all other computing nodes M (with worse E value and sufficient computing resources) to migrate all its virtual machines to M. M is selected from all nodes with the worse E value, but first are asked the nodes with better E value. If M has enough free resources, all virtual machines from N are migrated to M and N can be switched off. During the migration phase of the virtual machines N and M mustn't accept offers from other virtualization nodes. This case can happen only if there are the great resource differences between the virtualization nodes.

3.3. Nodes States Detection Policies

The main aspect for the virtualization node state detection is the energy power consumption, which is measured by the locally connected wattmeter. The proper absolute power consumption value (for the state detection) is for each virtualization node different and depends on its hardware configuration. The data is collected continuously and is evaluated in the periodic intervals. The absolute under/over load power consumption value (for the state detection) must be measured experimentally and depends on the specific system utilization. There [25] were published 4 various evaluation metrics – MAD, IQR, LR and LRR for the node state detection. The next possible policy is MEAN – the mean of all values in the last monitored time interval.

3.4. Virtual Machine Selection Policy for Migration

The best-fit (BF) selection policy was selected for the virtual machines selection. That means, the group of virtual machines, which can maximally utilize the offered free resources, is selected.

4. Measurements and Results Evaluation

4.1. Measurements Specification

The system measurements were realized by the using of the university virtualization system. The system consists of four various groups of the virtualization nodes. All servers contain the Microsoft Windows Server 2012 R2 operating system with the Hyper-V hypervisor.

Server categories:

1) SuperServers – 2 physical CPUs Intel Xeon v2, with 20 CPU physical CPU cores, 128 GBs RAM, 10 Gbit/s ethernet LAN connection, 2x Xeon Phi 5110P computing card, based on SuperMicro technology
2) Servers – 2 physical CPU Intel Xeon v3, 12 CPU cores, 32 GB RAM, , 10 Gbit/s ethernet LAN connection, 2x Xeon Phi 5110P computing card, based on SuperMicro technology
3) Old Servers – 2 physical CPU Intel Xeon v1, 8 CPU cores, 16 GB RAM, Asus technology, 1 Gbit/s ethernet LAN connection
4) Swap Servers – 1 physical Intel i5 CPU, 16 GB RAM, assembled from the parts of the various manufactures

The images of all virtual machines were stored on the network area storage, with the 10 Gb/s LAN connection. The power consumption data were measured by the EnerGenie PWM-LAN electronic wattmeters and the data collection was realized by the using of the in-house developed application EnergyMeter, which can be downloaded from its GitHub directory [26] .

4.2. Energy Consumption Efficiency Measurements

The energy consumption model was experimentally measured for all virtualization nodes in our system. The measurement results were consistent in all server categories and can be very good reproduced. The Figures 3-6, show the energy consumption of every server group. The system idle interval determines the minimum consumption power of the system without the stress. The virtualization nodes with the Xeon Phi computing cards have the higher idle power consumption value and this can handicap these nodes by their efficiency value computation. This can be solved in the equation (3) by the K variable substitution to K+L, where is the L the count of additional CPU cores.

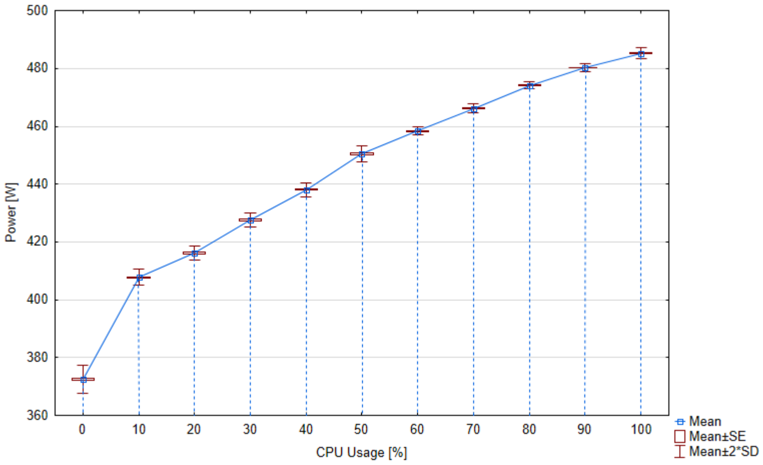

Figure 3. Super Server Consumption Evaluation Model

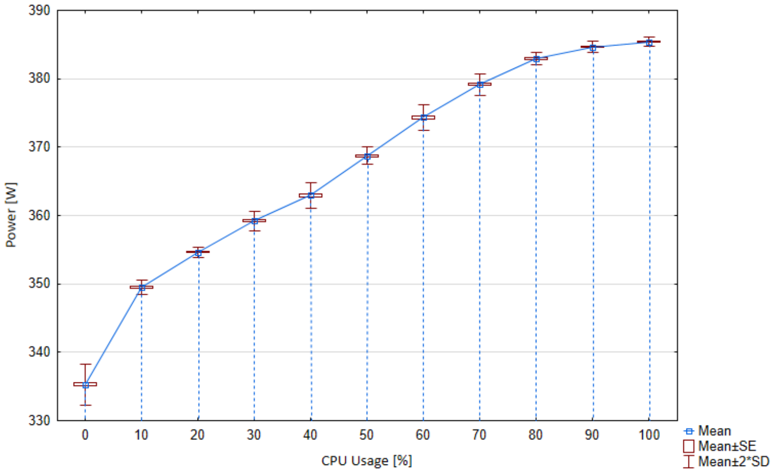

Figure 4. Server Consumption Evaluation Model

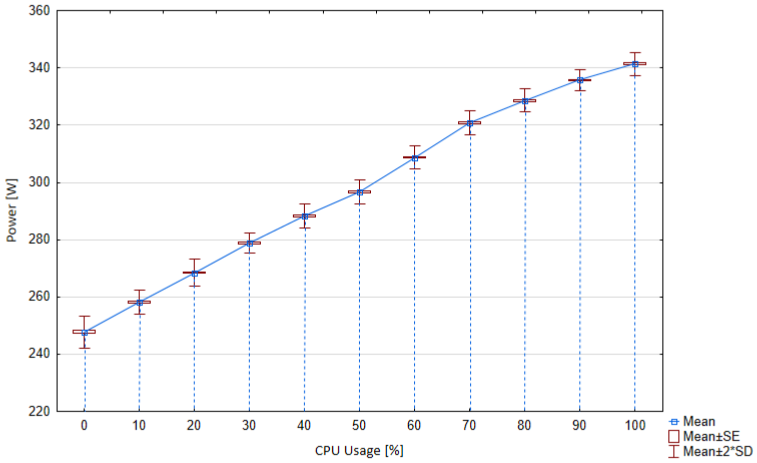

Figure 5. Old Server Consumption Evaluation Model

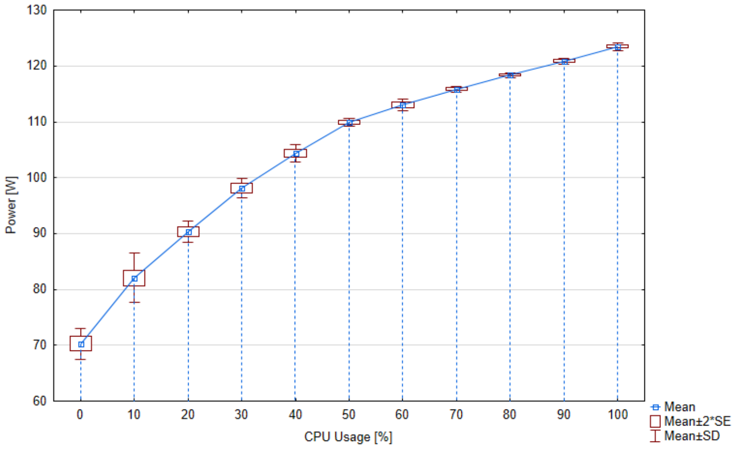

Figure 6. Swap Server Consumption Evaluation Model

The performance models showed the differences between all virtualization nodes. The Old Server nodes ha the same CPU cores count as the Server Nodes. The Server Nodes contains extra two additional Xeon Phi computing cards and the power consumption is similar. The air condition system by the Old Server has the similar power consumption as its physical CPUs. The comparison of all energy consumptions between all virtualization nodes groups is depicted in the Figure 7. The minimum idle states consumptions were between 70 – 260 Watts.

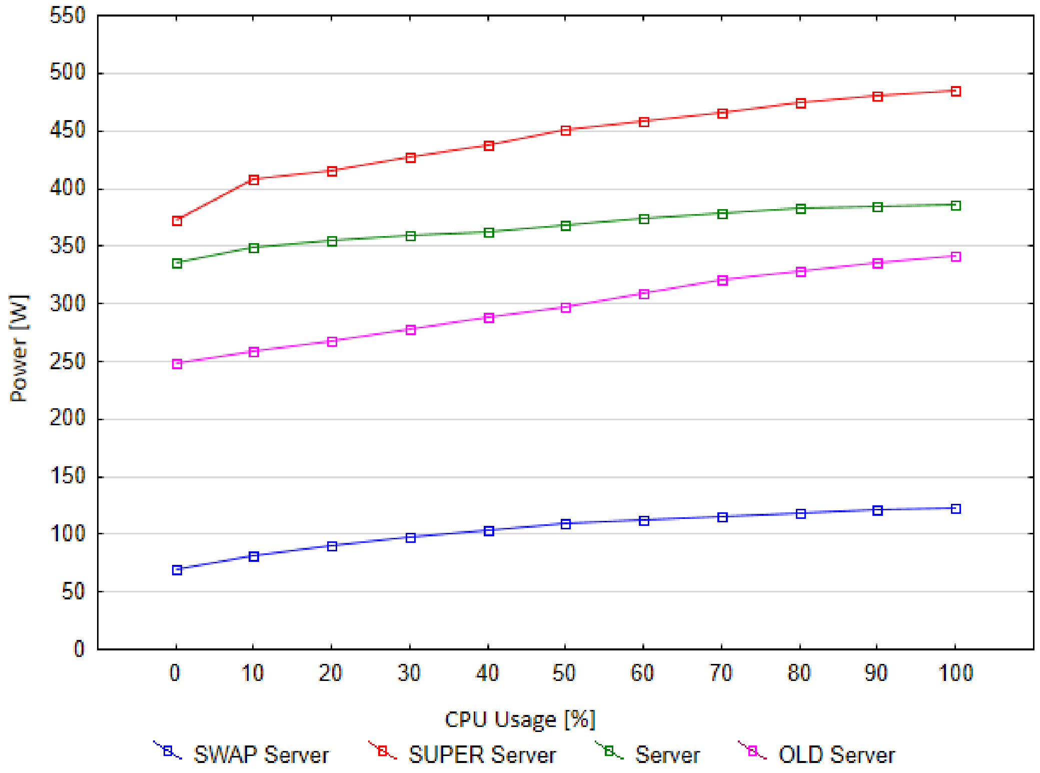

Figure 7. Power Consumptions Comparison (All Server Groups)

4.3. Experimental Algorithm Evaluation

The testing evaluation scenario was performed as follows. We prepared the group of hundreds of virtual machines (with exactly the same virtual hardware configuration, 2 CPU cores and 4 GBs RAM with the same 64bit operating system Windows 10 Education). The total energy consumption was computed as the sum of values, which were measured by the standalone electronic wattmeter. Every virtualization node energy input was monitored by its own device. One additional device was connected to the central network switch. All virtual machines were 6x automatically executed or stopped in the specific order by the virtualization system. One measurement batch took about one-hour time. The virtual machines were placed equally between all running virtualization nodes. The reference energy consumption value was the total power consumption of the not coordinated system – that means, the virtual machines ran still on the same virtualization nodes and no virtualization node was switched off by its idle time. The next measurement was similar, but the idle nodes were switched off. The last four measurements were performed by the using of the proposed algorithm. That means, the virtual machines were consolidated and idle nodes were switched off. The each virtualization node state evaluation was performed periodically every 5 minutes. The data was processed by the using of the MEAN, MAD, IQR and LR evaluation metrics. The results are depicted in the Figures 8 and 9. It is necessary to say, that our results are relevant to our specific system configuration and depend on the virtual machines state changes count.

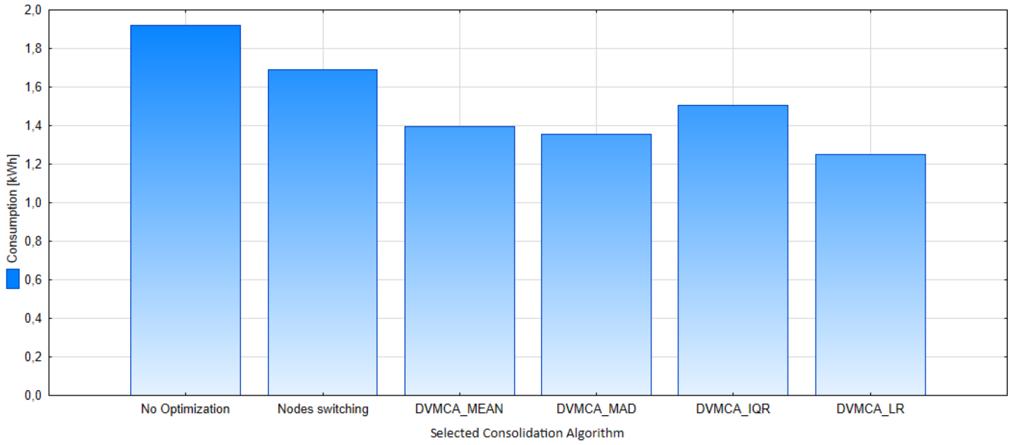

Figure 8. System Energy Consumption by the Specific Consolidation Algorithm Selection

The total consumption of the not coordinated system (total 15 virtualization nodes) was about 1.92 KWh, which means the 40% global average utilization of all nodes. The idle state nodes switching optimization reduced the energy consumption about 12%. This value is dependent on the virtual machines on/off switching changes count. The MEAN or MAD metrics optimization saved 27.2 or 29.4% of energy in comparison to the not coordinated system. The IQR metrics saved only 21.7% of energy. The best metrics was the LR metrics, which saved 34.8% of energy.

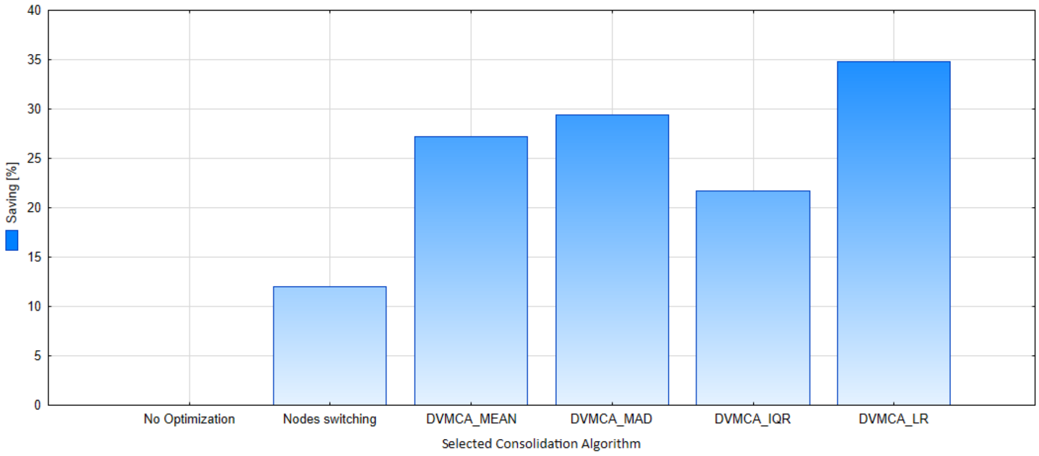

Figure 9. Energy Consumption Saving By the Specific Consolidation Algorithm Selection

5. Conclusion

We introduced the new fully distributed virtual machine consolidation algorithm, which is able to consolidate the virtual machines by the using of the real measurements obtained from the electronic wattmeters. The power consumption value seems to be the more suitable factor than the CPU utilization, which is used in other consolidation algorithms, because it has the smoother shape and does not step change. The real energy consumption measurements respect the various hidden additional aspects like the additional expanding cards, air condition system *etc.*, which further participate on the global system consumption. The testing was realized on the real distributed virtualization system and verified the abilities of the proposed solution.

Acknowledgement

The authors would like to acknowledge to the University of South Bohemia, Faculty of Science for providing the technical background, which was necessary for this paper creation. All authors were financed by the internal grants of the University of South Bohemia.

The authors would like further acknowledge to Michal Konopa for his valuable comments and advices.

References

[1] C. Clark, K. Fraser, S. Hand, J. G. Hansen, E. Jul, C. Limpach, I. Pratt and A. Warfield, “Live migration of virtual machines”, in Proceedings of the 2nd USENIX Symposium on Networked Systems Design and Implementation (NSDI), **(2005)**, pp. 273–286

[2] J.M. Pearson, “Large-scale Distributed Systems and Energy Efficiency: A Holistic View”, ISBN-10: 1118864638, Wiley, **(2015)**.

[3] J. G. Koomey, “Estimating total power consumption by servers in the US and the world”, Lawrence Berkeley National Laboratory, Tech. Rep., **(2007)**.

[4] R. Nathuji and K. Schwan, “Virtualpower: coordinated power management in virtualized enterprise systems”, ACM SIGOPS Oper Syst Rev., vol. 41, no. 6, **(2007)**, pp. 265C278.

[5] J. Stoess, C. Lang and F. Bellosa, “Energy management for hypervisor-based virtual machines”, Proceeding ATC’07 USENIX annual technical conference on proceedings of the USENIX annual technical conference, **(2007)**.

[6] A. Verma, P. Ahuja and A. Neog, “pMapper: power and migration cost aware application placement in virtualized systems”, Proceeding middleware ’08 proceedings of the 9th ACM/IFIP/USENIX international conference on middleware. Springer, New York, **(2008)**, pp. 243–264.

[7] S. Srikantaiah, A. Kansal and F. Zhao, “Energy aware consolidation for Cloud computing“, Proceedings of the 2008 conference on power aware computing and systems", San Diego, **(2008)**.

[8] E. Elnozahy, M. Kistler and R. Rajamony, “Energy-efficient server clusters,” in Power-Aware Computer Systems, ser. Lecture Notes in Computer Science, B. Falsafi and T. Vijaykumar, Eds. Springer Berlin Heidelberg, vol. 2325, **(2003)**, pp. 179–197.

[9] X. Fan, W. D. Weber and L. A. Barroso, “Power provisioning for a warehouse-sized computer”, in Proceedings of the 34th Annual International Symposium on Computer Architecture (ISCA), vol. 2007, **(2007)**, pp. 13–23.

[10] A. Beloglazov, R. Buyya, Y. C. Lee and A. Zomaya, “A taxonomy and survey of energy efficient data centers and cloud computing systems”, Adv. Comput., vol. 82, no. 2, **(2011)**, pp. 47–111.

[11] A. Beloglazov, “Energy-efficient management of virtual machines in data centers for cloud computing”, PhD thesis, Department of Computing and Information Systems, The University of Melbourne, **(2013)**.

[12] F. Farahnakian, P. Liljeberg and J. Plosila, “Energy-Efficient Virtual Machines Consolidation in Cloud Data Centers using Reinforcement Learning”, 22nd Euromicro International Conference on Parallel, Distributed, and Network-Based Processing, **(2014)**.

[13] Z. Cao and S. Dong, “An energy-aware heuristic framework for virtual machine consolidation in Cloud computing”, The Journal of Supercomputing, Springer, vol. 69, Issue 1, **(2014)**, pp. 429–451.

[14] C. Dupont, T. Schulze and G. Giuliani, “An energy aware framework for virtual machine placement in cloud federated data centres”, e-Energy’12 proceedings of the 3rd international conference on future energy systems: where energy, computing and communication meet. ACM, New York, **(2012)**.

[15] J. Xu and J.A.B. Fortes, “Multi-objective virtual machine placement in virtualized data center environments”, 2010 IEEE/ACM international conference on green computing and communications and international conference on cyber, physical and social computing, Hangzhou, **(2010)**, pp. 179–188.

[16] E. Feller, L. Rilling and C. Morin, “Energy-aware ant colony based workload placement in Clouds”, Proceeding GRID’11 proceedings of the 2011 IEEE/ACM 12th international conference on grid computing. IEEE Computer Society, Washington, DC, **(2011)**, pp. 26–33.

[17] K. Mills, J. Filliben and C. Dabrowski, “Comparing VM-placement algorithms for on-demand Clouds”, 2011 IEEE third international conference on Cloud computing technology and science (CloudCom), Athens, **(2011)**, pp. 91–98.

[18] N. Bobroff, A. Kochut and K. Beaty, “Dynamic placement of virtual machines for managing SLA violations”, 10th IFIP/IEEE international symposium on integrated network management, Munich, **(2007)**, pp. 119–128.

[19] D. Borgetto, H. Casanova and A. Costa, “Energy-aware service allocation”, Future Generation Computer Systems (FGCS). Elsevier press, Amsterdam, vol. 28, no. 5, **(2012)**, pp. 769–C779.

[20] N. Kansal and I. Chana, “Energy-aware Virtual Machine Migration for Cloud Computing - A Firefly Optimization Approach”, Journal of Grid Computing, Springer, vol. 14, Issue 2, **(2016)**, pp. 327–345

[21] B. Li, J. Li, J. Huai, T. Wo, Q. Li and L. Zhong, “EnaCloud: An Energy-saving Application Live Placement Approach for Cloud Computing Environments”, International Conference on Cloud Computing, IEEE, **(2009)**.

[22] S. Akiyama, T. Hirofuchi, R. Takano and S. Honiden, “MiyakoDori: A Memory Reusing Mechanism for Dynamic VM Consolidation”, Fifth International Conference on Cloud Computing, IEEE, **(2012)**.

[23] The Spec Power Benchmark, available online, http://www.spec.org/power_ssj2008/.

[24] J. Fesl., Exact Vitualization BenchMark (evbench), available online, http://https://github.com/UniThinkTeam/EVBench.

[25] R. N. Calheiros, R. Ranjan, A. Beloglazov, C. A. F. D. Rose and R. Buyya, “CloudSim: A toolkit for modeling and simulation of Cloud computing environments and evaluation of resource provisioning algorithms”, Software: Practice and Experience, vol. 41, no. 1, **(2011)**, pp. 23–50.

[26] J. Fesl, LAN Energy meter data collection tool, available online, http://https://github.com/UniThinkTeam/EnergyMeter.

Authors

Jan Fesl, he was born in Ceske Budejovice, Czech Republic in 1982. His M.Sc. Diploma received in Computer Science in 2007 at the Czech Technical University of Prague, Czech Republic. Currently, he is an assistant professor at the University of South Bohemia and he is the Ph.D. student at the Czech Technical University of Prague. His current research is focused on the computer networks and distributed computing systems.

Jiří Cehák, he was born in Ceske Budejovice, Czech Republic in 1990. He is the M.Sc. student at the University of South Bohemia. Main areas of his research are the bioinformatics and computer graphics.

Marie Doležalová, she was born in Hradec Králové, Czech Republic in 1981. She is the M.Sc. student at the University of South Bohemia. Main areas of her research are the bioinformatics and computer graphics.

Jan Janeček, he is the associate professor at the Czech Technical University in Prague, head of the network research group. His current research is focused on the computer networks, distributed computing and internet of things.