Power-Aware Cooling Control Architecture for Container Data Center

Hiroyoshi Kodama, Masatoshi Ogawa, Hiroshi Endo, Toshio Sugimoto, Hiroyuki Fukuda, Masao Kondo, and Jun Tanaka
Fujitsu Laboratories Ltd.
Kanagawa, Japan
Email:{hkodama, ogawa.masatoshi, endo-hiroshi, tsugi, fukuda.hiro, condor, tanaka.jun.777}@jp.fujitsu.com

***Abstract*—In this paper, we propose a power-aware cooling control architecture for container data centers. In leading edge data centers, the Computer Room Air Conditioning (CRAC) unit controls the temperature and humidity. The CRAC uses large fans to take in hot air and blow out cold air. On the other hand, the server cooling is autonomously controlled with built-in server fans via many temperature sensors in the server. That is, both a large fan outside the server and a small fan in the server contribute to cooling the server. We consider that, if a large CRAC fan works mainly with the server cooling fan, the total electric power needed to cool the server can be reduced. For the same air flow, the fan speed of a big fan is less than that of a small fan. Our combined cooling system is based on keeping the temperature of the Central Processing Unit (CPU) fixed. The CRAC fan is controlled following the Model Predictive Control (MPC) using the CPU temperature. The server fan is programmed to a lower speed than the one that would be used by a standard controller by issuing commands to the Intelligent Platform Management Interface (IPMI). We verified the proposed system by using an actual container data center. The results show that the proposed system realizes power savings of more than 30% compared to the standard control system. In particular, the power-saving effect of the proposed system is large when the cold aisle is 25°C or more.**

Keywords- power-aware cooling system; container data center; CRAC; MPC; IPMI; CPU temperature.

I. INTRODUCTION

A variety of cloud computing services have been proposed in the last few years [1][2]. The Information and Communication Technology (ICT) equipment to provide cloud computing services is stored in leased facilities called Data Centers (DCs), which are strictly managed. The number of DCs increases as cloud computing spreads and develops. Therefore, power consumption increases, and lowering the power consumption of data centers is one of the pressing issues of the global environment.

A new cooling system for Container Data Centers (CDCs) that incorporates fan-less servers and a fresh air cooling method to minimize power consumption has been proposed [3]. As a result, a 22.8% energy saving was achieved with this system compared with the conventional container servers with built-in fans. However, the new system was not practical because it required special conditions, for example, fresh air cooling and the use of fan-less servers.

The typical server (x86 server) has built-in fans and cools itself automatically. For instance, the Baseboard Management Controller (BMC) can control the fan so that the fan speed in the server will increase as the temperature of the CPU increases. There are various places where the temperature has to be observed, for example, the CPU, the memory, the Power Supply Unit (PSU), and the exhaust of the server.

The CRAC controls the temperature, humidity, and flow of air using fans in the CDC. Two types of fans (built-in fans in the server and larger fans in the CDC) exist for the purpose of cooling the server. For the same air flow volume, the power consumption of a large fan is less than that of a small fan. However, these two types of fans are controlled independently. Therefore, we thought that we could obtain a further power saving by cooperatively controlling the two types of fans.

In this paper, we propose a cooling control architecture for the server and CRAC. The CRAC fan and the server fan are controlled by a manager server to bring the temperature below a constant temperature (for example, the CPU temperature) with the proposed architecture.

The paper is structured as follows. In Section II A, we explain power consumption and the control of a standard server fan. In Section II C, we explain our cooling control architecture. Section III discusses the experimental results. Finally, Section V concludes this study.

II. COOLING CONTROL ARCHITECTURE

A. Power consumption and control of server fan

The server has two or more internal cooling fans. These fans are controlled by an algorithm saved to the firmware based on the value of the temperature sensor in the server. In the case of the RX200S7 [4], which is made by Fujitsu, there are 6 tandem fans in the front. Figure 1 shows the relation between the server fan speed and the power consumption. In this case, the server fan speed was changed by us regardless of the sensor temperature, and the power consumption of the server was measured. When the server fan speed was raised to 80%, the power consumption increased by about 60 W. The power consumption of a fan is proportional to the cube of the rotation speed. When fan duty reaches 100%, as much as about 100 W is consumed by the fans only. The speed of the built-in fan of a standard server is controlled by the value of the temperature sensor in the server. The control algorithm is programmed into the BMC and executed. The rotation speed of the fan can be controlled with IPMI. However, the server vendor has not opened to the public the IPMI command for fan control. In this paper, we controlled the built-in server fan by using a IPMI vendor-specific command.

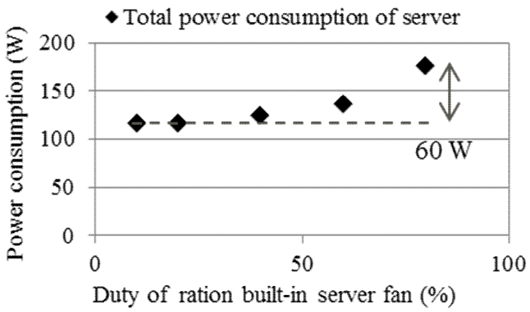

Figure 1. Power consumption of built-in server fan.

B. *Cooling Control method of CRAC fan*

The CRAC has a certain cooling capacity when the computer room is fully loaded with servers. The CRAC controls the air temperature and the air flow rate. However, when the cooling space is as narrow as that of the container data center, the air flow rate greatly affects the cooling capability. We have already proposed a cooling control method based on the Model Predictive Control (MPC) [5] for a container data center directly utilizing fresh air [6]. In our MPC method, the CRAC fan was controlled so that the highest CPU temperature in the rack might be made the predetermined CPU temperature.

C. *Concept Architecture*

For the same air volume, using the large fan rather than the small one saves energy. The air flow of CDC is ideal. In this case, Newton's law of cooling model is applicable for a closed space like the CDC. The equation for Newton's law of cooling is shown below.

$$\frac{dQ}{dt} = h \times A \times \left(T(t)_w - T_f\right) \qquad (1)$$

where Q is the thermal energy, h is the heat transfer coefficient, A is the heat transfer surface area, T_w is the temperature of the parts of the servers, and T_f is the ambient temperature of the CDC. If the air volume of the CRAC fan is increased, the CPUs of the servers can be cooled. This situation was confirmed with a real machine. A server (RX200 S7) of 30 1U types was installed, and it had two CPUs [E5-2650, 2.0 GHz, thermal design power (TDP) 95 W] [7]. In this paper, the CPU temperature refers to the temperature measured by the Digital Temperature Sensor (DTS). The Platform Environment Control Interface (PECI) temperature of this CPU is 89°C. That is, if the CPU temperature reaches 89°C, it begins to lower the clock frequency with the PECI. The Intel Power Thermal Utility (PTU) was used for the load to the CPU. Figure 2 shows the CPU temperature map in the server rack. The power consumption of the entire rack and fan rotation speed of CRAC is shown at the top of the rack map in Figure 2. On the left, the numbers 15-42 indicate the location of the server in the rack. Setting the fan of the CDC for minimum rotation is shown in Figure 2 (a), and setting the rotation of the fan to 4360 rpm is shown in Figure 2 (b). The CPU temperature becomes low when the rotation of the CDC fan increases to 4360 rpm. It can be easily seen that the CPU temperature color of the entire rack changed. The power consumption of the server has decreased from 8.8 kW to 8.6 kW. As the rotation of the CRAC fan increased, the rotation of the server fan decreased, which led to a decrease in power consumption. The rotation of the server fan had slowed because the CPU temperature decreased in this experiment. The CPU temperature decreased from 73°C to 66°C for the U20 server, and the power consumption decreased from 328 W to 320 W. The rotation speed of the U20 server fan decreased from 7920 rpm to 6120 rpm. In addition, if the server fan can be slowed down further to maintain the CPU temperature, the power consumption of the server can be reduced. This is one advantage of this architecture. The CPU temperature at U20 is 66°C owing to the excess cooling. Our cooling algorithm aims to keep the CPU temperature constant with both a large CDC fan and a small server fan. The electric power of CRAC and the electric power of the built-in fan are compared and controlled so that the total electric power consumption can be reduced.

Figure 3 (a)-(b) shows the details of the proposed

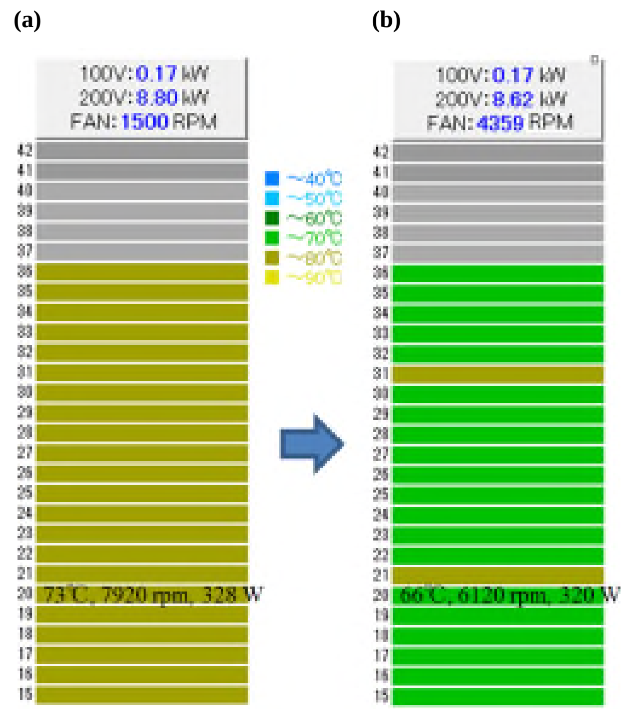

Figure 2. CPU Temperature map in server rack. CPU temperature, server fan speed, and power consumption of server at U20 are shown. Total power consumption and CDC fan speed are shown at top of rack map. (a) CDC fan speed of 1500 rpm. (b) CDC fan speed of 4360 rpm.

 ISBN: 978-1-61208-406-0

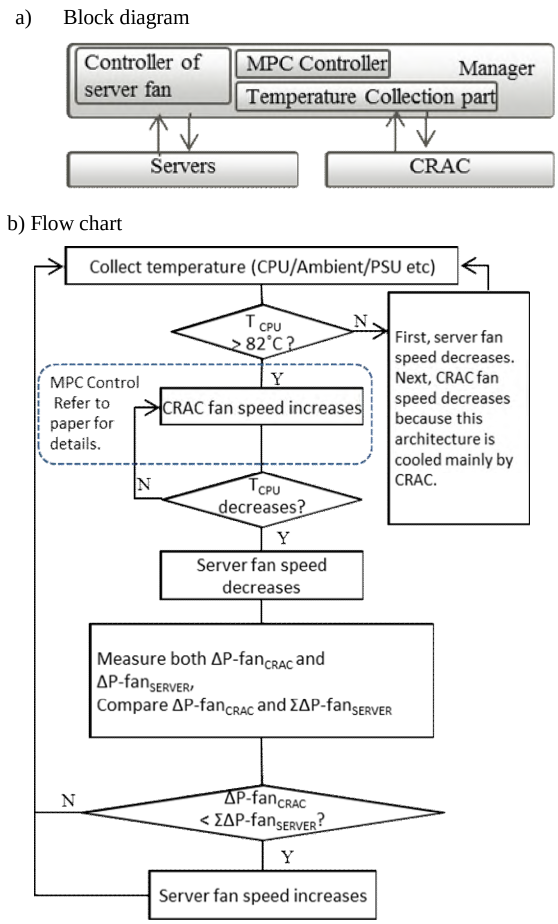

Figure 3. Block diagram of the proposed architecture.

architecture. The manager collected the temperatures of the CPUs and power consumption from the servers with IPMI. Of course, the manager collected the ambient temperature and other component temperatures from the BMC with IPMI. The server fan speed was controlled by the manager so that it was not based on the load of the server but was set to keep the CPU temperature at a fixed temperature. We called this method the server power-saving system (SEPOSS). The manager controlled the server fan speed with SEPOSS, and the CRAC fans were controlled by the MPC method [6]. The manager performed power-saving control by both SEPOSS and MPC and kept the CPU temperature constant at any load. In this architecture, the CPU temperature was regulated at 82°C. Figure 3(b) shows the processing of the case where the temperature of the CPU (T_{CPU}) is near 82°C (regulated temperature). The CRAC fan speed increases with MPC control. If T_{CPU} decreases, the manager decreases the server fan speed. The increment of electric power to increase the speed of the CRAC fan is assumed to be ΔP-fan$_{CRAC}$. The increment of electric power to increase the speed of the server fan is assumed to be ΔP-fan$_{server}$. The manager compares ΔP-fan$_{CRAC}$ and Σ ΔP-fan$_{server}$. The manager chooses the fan that does not have much power consumption.

III. Experiments with Actual Equipment

A. *Results of regulating CPU temperature control*

The CPU temperature control that we propose was experimented using an actual CDC [8]. Figure 4 shows the layout of our CDC. In this experiment, 26 servers were installed in rack 7 and rack 5, and the facilities used CRAC_A and CRAC_B. To adjust the maximum temperature of the CPUs of the servers installed in rack 7 to 82°C, CRAC_A was controlled by MPC. Similarly, CRAC_B was controlled at the maximum temperature of the CPUs of rack 5. The heaters were installed in the rack so that the quantity of heat per set of racks could be set to about 18 kW. Moreover, the partition was set up at the center of the CDC. The temperature of the Supply air (SA) of CRAC was 20°C.

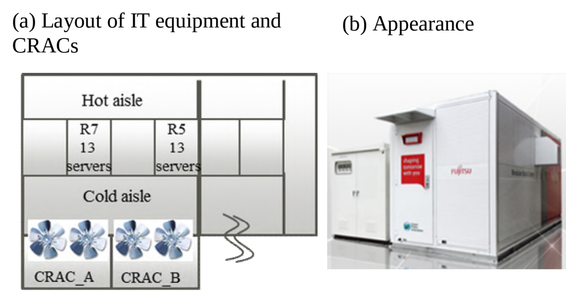

Figure 4. The CDC used for experiment.

The CPU load given to a server changed every 30 min in this order: 50%, 100%, idle (0%), and 80%. The highest temperature of the CPUs in rack 5 is shown in Figure 5 (a). The red line is the standard control, and the blue line is the regulated control. In standard control, the temperature of the CPU with loads of 50%, 100%, 0%, and 80% was 73°C, 78°C, 42°C, and 76°C, respectively. On the other hand, in regulated control, the CPU temperature was almost regulated to 82°C. When CPU loading was switched from 50% to 100%, some overshooting was observed. The temperature of the CPU was measured every 10 seconds, and overshooting was observed for 20 seconds. The air flow volume of the CRAC_A fan is shown in Figure 5 (b). In regulated control, when the loads of the server were 100% and 80%, the air flow volume increased. The comparison of the total power consumption of the standard control and cooperative control is shown in Figure 6. The regulated control reduces the power consumption by about 30% compared with the standard control. This is the maximum case, and it changes by the air specification of the CRAC fan. For instance, when the CRAC fan is changed from 10,000 m^3/hour to 6000 m^3/hour, the reduction rate is thought to be 10% or more. However, it is clear that the proposed method is able to reduce the power consumption. Therefore, the effectiveness of the proposed method is confirmed.

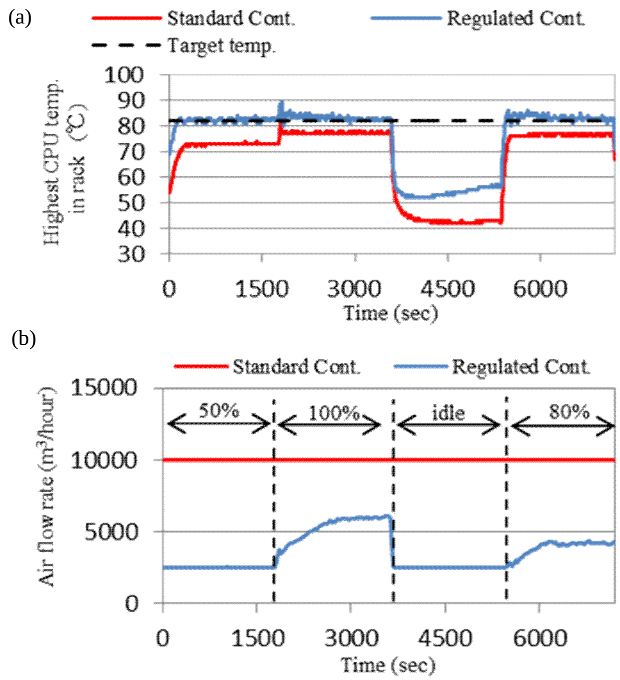

Figure 5. Comparison of standard control and regulated control. (a) Maximum CPU temperature at various loads. (b) Air flow volume of CRAC at various loads.

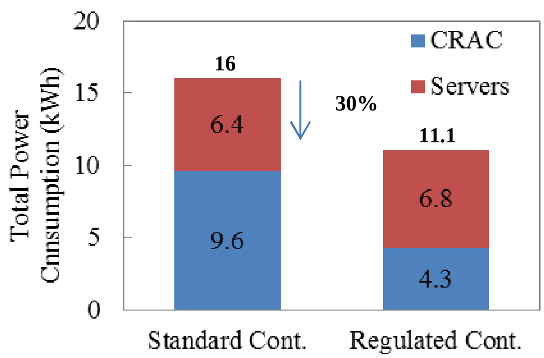

Figure 6. Comparison results of total power consumption

B. *Operation by class of A3 environment*

The American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE) shows the guidelines of data centers in TC9.9 [9]. In these guidelines, the equipment environment specification of 5°C-35°C is defined as class 3 (A3). In the environment of 35°C, the power consumption of the CRAC decreases compared with the environment of 20°C. On the other hand, a standard server begins to raise the server fan speed when the ambient temperature increases, and the server fan rotates at the highest speed at 35°C. ASHREA reported on page 13 of ASHRAE TC 9.9 [9], "if inlet temperature increases to 35°C, the IT equipment power could increase in the range of 7 to 20% compared to operating at 15°C." That is, even if it sets the environment temperature to a high temperature, the result that improves the power consumption of a server may be caused instead by the power consumption of the CRAC decreasing. In particular, in the regulated control that we propose, the effect of SEPOSS control is that it can perform power-saving in such high temperature environments. The temperature of supply air (SA) in the CDC was changed from 17°C to 35°C, and the standard control was compared with SEPOSS control. The CPU load was not given to the server. Figure 7(a) shows the relation between the environment temperature and power consumption in the CDC. Moreover, the relation of the environment temperature and the server fan speed is shown in Figure 7(b). In standard control, power consumption went up as environmental temperature became high, but the increase in power consumption was suppressed in SEPOSS control. This is based on the inhibiting effect of the server fan speed. In this case, a reduction in power consumption of 16% was confirmed by SEPOSS control. SEPOSS control supervised the temperature of the parts inside the server and did not control fans in accordance with the intake air temperature. We confirmed that the temperature of each part in the server was below the regulated value and also under SEPOSS control. The temperature in the server with an environmental temperature of 35°C is shown in Table I. It does not become a problem because it is below the alarm temperature though the temperature of the CPU when SEPOSS is controlled is about 4°C (CPU1) higher than usual. The part whose alarm

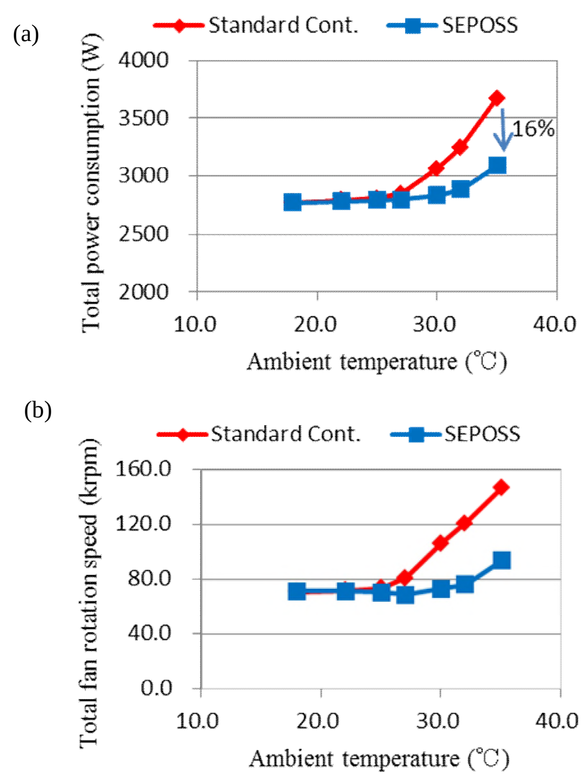

Figure 7. Comparison results of SEPOSS control under high ambient temperature. (a) Total power consumption. (b) Total server fan speed.

temperature is the lowest in the server is the PSU inlet. It is understood that the temperature of the PSU inlet is below the alarm temperature when SEPOSS is controlled.

TABLE I. INTERNAL TEMPERATURE OF SERVER AND ALARM TEMPERATURE OF SEPOSS CONTROL AND STANDARD CONTROL.

[℃]

	Ambient	CPU1	CPU2	Memory	PSU inlet	PSU	System board
Standard	34.6	46.4	47.0	40.9	43.0	62.3	48.2
SEPOSS	34.3	51.0	50.4	42.6	41.4	62.3	51.3
Alarm	37	88	88	78	52	90	75

IV. RELATED WORK

Researchers have investigated the optimization of the temperature control of CRAC in the datacenter. In 2012, T. Hayashi et al. mentioned energy saving systems [10]. This paper presents the energy-saving potential of coordinated control based on experiments. The results indicate that additional energy can be saved by controlling the rotation of fans and temperature settings among multiple CRAC units in response to heat generation due to operation of ICT equipment. In 2011, Wei Huang et al. mentioned power optimization techniques for servers and DCs [11]. They demonstrated a run-time optimization technique that reduces the aggregate server fan power and processor leakage power of a server system. In 2014, S. Yeo et al. mentioned a system where CPU power capping and CRAC control were combined [12]. They proposed a technique which trades off between performances and reliably.

V. CONCLUSION AND FUTURE WORK

In this paper, we have proposed a power-aware control architecture for the server and CRAC. The CRAC fan and the server fan were controlled by a manager server to bring the temperature below a constant temperature (for example, the CPU temperature) with the proposed architecture. The CRAC fan was controlled by the MPC method, and the server fan was controlled by the SEPOSS method. Our proposed architecture was examined in an actual CDC. As a result, a power saving of 30% was confirmed compared with the standard control case. We would like to fine-tune the MPC and the SEPOSS control in the future. To reduce temperature overshooting during regulated controlling, we need to improve it first. In this experiment, the temperature was controlled from the management server by IPMI commands. The problem of this method is that the number of the controlled servers is limited. When the number of controlled servers exceeds 1000, a delay is expected to be caused in the control due to the management server. We want to give only the target temperature from the management server and improve the control directly by programming the BMC. This method has been actually used with various workloads though the server load remained the same as that of this study.

REFERENCES

[1] J. G. Koomey, “Estimating total power consumption by servers in the U.S. and the world”, Final report, Lawrence Berkeley National Laboratory, Palo Alto, CA, February 15, 2007.
http://hightech.lbl.gov/documents/DATA_CENTERS/svrpwrusecompletefinal.pdf [Retrieved: 4-2015]

[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] H. Endo, H. Kodama, H. Fukuda, T. Sugimoto, T. Horie, and M. Kondo “Cooperative control architecture of fan-less servers and fresh-air cooling in container servers for low power operation”, Proceedings of the Workshop on Power-Aware Computing and Systems (HotPower 2013). Nov. 2013. doi: 10.1145/2525526.2525844.

[4] Fujitsu, “PRIMERGY RX200 S7 SERVER, Upgrade and Maintenance Manual”, pp152, 2013.
http://manuals.ts.fujitsu.com/file/10630/rx200s7-umm-en.pdf [Retrieved: 4-2015]

[5] J. M. Maciejowski: “Predictive Control with Constraints” Pearson Education, 2002. ISBN: 9780201398236

[6] M. Ogawa et al., “Cooling control based on model predictive control using temperature information of IT equipment modular date center utilizing fresh-air” , 13th International Conference on Control, Automation and Systems(ICCAS 2013), pp. 1815-1820. Oct. 2013, doi: 10.1109/ICCAS.2013.6704235.

[7] Intel, “Intel Xeon Processor E5-1600/E5-2600/E5-4600 v2 Product Families”, p. 12, March 2014.
http://www.intel.com/content/dam/www/public/us/en/documents/datasheets/xeon-e5-v2-datasheet-vol-1.pdf [Retrieved: 4-2015]

[8] Fujitsu, Datacenter Product Modular Data Center, http://jp.fujitsu.com/platform/server/container/ [Retrieved: 4-2015]

[9] White paper prepared by ASHRAE Technical Committee (TC)9.9, “2011 Thermal Guidelines for Data Processing Environments -Expanded Data Center Classes and Usage Guidance”, p. 13, 2011.
http://ecoinfo.cnrs.fr/IMG/pdf/ashrae_2011_thermal_guidelines_data_center.pdf [Retrieved: 4-2015]

[10] T. Hayashi, T. Tominaga, K. Saigo, and P. Gemma, “Minimum data set for controlling data center equipment for energy saving management” Power and Energy Society General Meeting, 2012 IEEE, July 2012, doi: 10.1109/PESGM.2012.6344974.

[11] Wei Huang, et al., “TAPO: Thermal-Aware Power Optimization Techniques for Servers and Data Centers”, Green Computing Conference and Workshops (IGCC), July 2011, doi: 10.1109/IGCC.2011.6008610.

[12] S. Yeo, M. M. Hossain, Jen-Cheng Huang, and Hsien-Hsin S. Lee, “ATAC: Ambient Temperature-Aware Capping for Power Efficent Datacenters”, Proceedings of the ACM Symposium on Clud Cpmputing 2014 (SoCC ’14), Nov. 2014, doi: 10.1145/2670979.2670996.