# Opportunities and Challenges to Unify Workload, Power, and Cooling Management in Data Centers

Zhikui Wang, Niraj Tolia, and Cullen Bash
HP Labs, Palo Alto
{zhikui.wang, niraj.tolia, cullen.bash}@hp.com

## ABSTRACT

Independent optimization for workload and power management, and active cooling control have been studied extensively to improve data center energy efficiency. Recently, proposals have started to advocate unified workload, power, and cooling management for further energy savings. In this paper, we study this problem with the objectives of both saving energy and capping power. We present the detailed models derived in our previous work from experiments on an blade enclosure system that can be representative of a data center, discuss the optimization opportunities for coordinated power and cooling management, and the challenges for controller design. We then propose a few design principles and examples for unified workload management, power minimization, and power capping. Our simulation-based evaluation shows that the controllers can cap the total power consumption while maintaining the thermal conditions and improve the overall energy efficiency. We argue that the same opportunities, challenges, and designs are also generally applicable to data center level management.

## 1. INTRODUCTION

Power consumption has become a critical issue in the design and operation of enterprise servers and data centers today [23]. This problem is being further exacerbated by the increase in construction of large data centers for both traditional IT workloads and cloud-based services. In response to this problem, there have been several studies on server and cluster power management [6, 11, 16, 18, 19]. Most of these systems use "compute actuators" such as P-state control, workload migration, load-balancing, and turning machines on or off. More recent proposals also advocate moving workloads across data centers to exploit differences in electricity pricing [17] or operational efficiency [14]

However, server power is only one component of the total power consumed by a data center. The other significant component is facility power consumed by cooling equipment such as fans and computer room air conditioners (CRACs). Studies show that every 1W of power used to operate servers often requires an additional 0.5-1W of power, used by the cooling equipment, to extract the heat at the data center level [8, 15]. The yearly electricity costs for cooling large data centers can thus reach millions of dollars [15]. While there have been a number of proposals to separately optimize for cooling [2, 3, 9, 12, 15], recent work has started exploring the increased benefits possible from unified control of server and cooling resources [1, 20, 21, 24].

Most of the unified control systems mentioned above try to minimize power usage with a minimal impact on workload performance. They currently do no consider fixed power budgets that would require a more careful allocation of power to compute and cooling. At the same time, there has been an increasing amount of interest in the Smart Grid and its impact on IT in general and data centers in particular. The benefits of using a Smart Grid infrastructure includes a reduction in $CO_2$-emitting power plant construction, a more efficient and reliable electricity grid, and better customer pricing for electricity. However, the main actuators used by the Smart Grid include mechanisms such as Time-of-Use (TOU) pricing, Critical Peak Pricing (CPP), Real-Time Pricing (RTP), and Peak Time Rebates. To take advantage of the savings possible, any power management system would require power capping [18] and would introduce power budgets for the entire data center.

The underlying assumption in most unified management systems is that both the compute and facility systems are aware of the other and can therefore make better-informed decisions. These systems generally use a model-based approach to predict the impact of their optimizations on both energy costs and performance of hosted workloads. Our initial investigation of the combination of power capping, power saving, and performance guarantees shows that there are significant challenges to be overcome. The majority of the challenges arise from complicated thermal dynamics. Without careful consideration of this phenomena, systems can oscillate and lead to poor workload performance, violation of power budgets and even thermal safety bounds.

In the rest of this paper, we examine some of the challenges and opportunities to unify workload, power and cooling control including power capping through a discussion of the general problem definition and model development for an exemplar blade enclosure system. We propose design and implementation approaches for robust and effect power capping controllers in a unified control architecture that are general to data center management. We finally evaluate the designs using a simulation-based approach.


FeBID'10, April 13, 2010, Paris, France

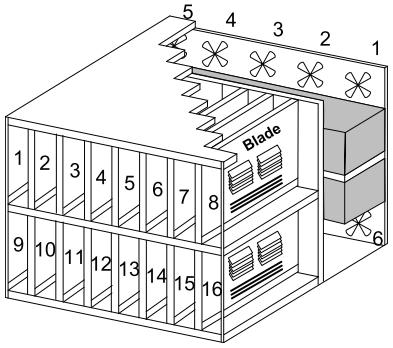


**Figure 1: Enclosure Design**

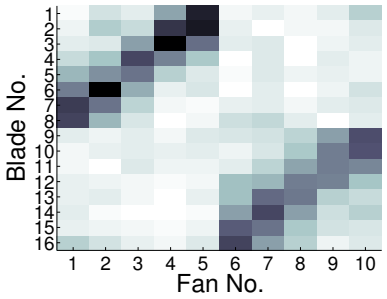


**Figure 2: Blade:Fan Relationship**

# 2. CONSTRAINED OPTIMIZATION

Multiple objectives are to be considered for energy management in data centers, including, but not limited to, power minimization, application performance guarantees, temperature tracking for the safety of electronic components, power capping, and the overall minimization of the total cost of ownership (TCO). Given the time-varying application demand and energy cost, the management problem can be formulated as an optimal control problem with constraints. The object is to minimize the total energy consumption of both the IT components such as servers, networking and storage, denoted as $P_{Servers}$, and the facility components such as CRACs, pumps and chillers, denoted by $P_{Cooling}$:

$$\min \int_0^t \left(P_{Cooling}(\tau) + P_{Servers}(\tau)\right) d\tau. \quad (1)$$

Most of the other objectives can be represented as constraints, three classes of which are considered in this paper. First, the capacity of the servers need to satisfy the resource demand of the applications running on the servers to preserve application performance. We assume that a resource utilization threshold is defined for each server and a utilization under this threshold would guarantee application performance, which means

$$Util_{Servers}(t) \leq Util_{Ref}(t). \quad (2)$$

Second, for safety of the electronic components such as processors, memory DIMMs, and disks, the thermal condition of the servers, usually represented by the server temperatures, are to be maintained below specification-defined thresholds:

$$T_{Servers}(t) \quad \leq \quad T_{Ref} \quad (3)$$

And third, as motived by requirements from the Smart Grid, we assume that the total power consumption needs to be maintained below a budget:

$$P_{Cooling}(t) + P_{Servers}(t) \leq P_{budget}(t). \quad (4)$$

# 3. OPPORTUNITIES AND CHALLENGES

A number of knobs are available to address the problem defined by Equations (1-4). For instance, dynamic voltage and frequency scaling (DVFS), power status (on/sleep/off) tuning, admission control, load balancing, and workload consolidation through virtual machine (VM) migration are used for server power management. On the cooling side, in a raised-floor open environment, the server temperatures can be modulated by the fans inside the servers, the perforated tiles on the floor that are located in front of server racks, blower speeds and supply/return air temperatures of the CRAC units that provide cooling air to thermal zones, chilled water pumps and controls within chillers, and cooling towers that provide chilled water to the whole data center. In this section, we describe the models that relate the knobs to the performance, thermal, and power metrics that are needed for the optimization problem (1-4). The models also provide insights into both the opportunities and the challenges for optimization. In this paper, we do not consider most of the cooling control knobs. Instead, we focus on the fans inside the servers and assume that the intake air temperatures of the racks, which are approximately the same as the ambient temperatures of the servers, are maintained at constant levels that are below a given threshold (e.g., 28C).

## 3.1 Blade enclosure: an exemplar system

We consider a blade enclosure as an exemplar system. Figure 1 illustrates a typical one with a total of sixteen blades in the front, eight on the top and eight on the bottom, cooled by a total of ten fans in the back, five on the top and five on the bottom. The airflow generated by the fans is pulled through the blades towards the back of the enclosure with each fan contributing to the blade-level airflow rate. While a blade enclosure is simpler than a data center, we believe it is still representative for interactions between the power and cooling systems in a data center. We have derived the exact models for such a system through extensive experiments [21, 26]. We briefly discuss the models again to elaborate on the opportunities and challenges in unifying power and cooling management in data centers.

## 3.2 Opportunity for optimization

In data centers, the cold air provided by the CRAC units is shared by multiple racks of severs. Due to the physical layout, the cooling efficiency of each CRAC on each server can be location dependent [3]. Similar effects exist in the blade enclosure, where each fan contributes partially to the airflow across each blade and the air flow through one blade is an aggregate of the flows generated by all of the fans. The per-blade air flow $\dot{V}_j$ was found to be correlated to the fan speed $FS_i$ in the following manner:

$$\dot{V}_j = \sum_i \eta_{ij} \times FS_i, \quad \text{for any blade j.} \quad (5)$$

Figure 2 displays a heatmap illustrating the variation in the derived values for correlation index $\eta_{ij}$. The figure shows that while blade temperatures are most affected by fans closest to them, the degree of influence of each fan as well as the number of fans that can significantly affect a blade shows significant variation between blades.

Sharing of the cooling capacity provides opportunity to

optimize the fan speeds so that the blades are not overcooled. The need for optimization is more obvious if we consider the power consumption of the server fans, notated as $P_F$ in later sections, or the blowers inside CRAC units, which is a cubic function of the rotational speed [10].

As shown earlier [7, 9, 18], the server power ($P_B$) can be defined as,

$$P_{B_j} = g_B * Util_j + P_{B,idle}. \quad \text{for any blade j.} \qquad (6)$$

CPU utilization ($Util$) is a proxy for the effect of active workload management, while the slope $g_B$ and the intercept $P_{B,idle}$ capture the effect of power status tuning.

## 3.3 Coordination between supply and demand

Active cooling control is targeted towards minimizing the cooling power consumption while meeting the cooling requirements of the servers. On the other hand, active workload management and server power control makes it possible to distribute the demand to maximize the overall cooling efficiency. Coordination between the cooling supply and demand is critical to address the problem (1-4), for which thermal models are needed to represent the inter-correlation between the actions.

For the exemplar enclosure system, we developed the dynamic thermal model for CPU temperature, $T_{CPU}$[1], [26] by leveraging heat transfer theory for thermal resistance and energy balance via a lumped capacitance method [13]:

$$C_1 \frac{dT_{CPU,j}}{dt} = \frac{C_2}{R_j}(T_{amb,j} - T_{CPU,j}) + Q_j \qquad (7)$$

where $T_{amb}$ is the ambient temperature. The variable $Q_j$ represents the heat transfered per unit of time between the CPU and the ambient air and can be approximated using CPU power consumption, which is modeled as a linear function of its utilization:

$$P_{CPU_j} = g_{CPU} * Util_j + P_{CPU,idle}, \quad \text{for any blade j.} \quad (8)$$

The variable $R_j$ is the thermal resistance between the CPU temperature sensor and the ambient temperature sensor, which can be approximated as

$$R_j = \frac{C_3}{\dot{V}_j^{n_R}} + C_4, \quad \text{for any blade j,} \qquad (9)$$

where $\dot{V}_j$ is the volumetric air flow rate through blade $j$. The parameter $n_R$ defines the shape of the thermal resistance curve as a function of the air flow rate. It is primarily related to heatsink design and the level of turbulence in the flow. The parameters $C_k$, $k = 1, 2, 3, 4$ are constants related to the fluid and material properties of the air, the CPU package, and the heat sink.

Note that the models (5, 7-9) capture the interaction among the thermal condition of servers, the workloads, local server cooling, and the data center level cooling (represented by the ambient cooling air temperature). From a dynamic control point of view, the CPU temperature is a first-order linear function of both the heat transfered and the ambient temperature. This relation has been widely utilized previously to model electronic component temperatures. However, when the fan speed is actively tuned, the time constant of the first-order system can vary significantly along with the fan speed. This complicated nonlinear and time-varying behavior of the thermal system makes it challenging to design efficient and robust dynamic controllers. In the steady state when the CPU temperature converges, the heat, or the power consumption of the processor, has a direct relation to the fan speeds, or the fan power. In other words, when the total power of the fans and the servers is capped, it is not trivial to allocate the power budget among the fans and servers that can maximize energy efficiency.

[1] $T_{CPU}$ is the dominant temperature sensor in most servers.

# 4. HIERARCHICAL CONTROL

In a typical data center, unified management solutions are needed to coordinate numerous power and cooling control knobs and enforce multiple objectives and constraints. Hierarchical architecture is a natural choice due to two reasons. First, the frequency to tune these knobs ranges from milliseconds to minutes, tens of minutes, or even hours. Second, the objective functions and the constraints are applied to sever components, servers, groups of servers, server racks, groups of racks, and the whole data center. More arguments for a hierarchical architecture can be found in [18].

For the exemplar enclosure system, we created a simulator using the discretized models from those defined in (5-9) and parameters derived through experiments. We implemented a control architecture to address the problem (1-4) where $P_{Cooling}$ and $T_{Server}$ are represented by the fan power $P_F$ and the CPU temperature $T_{CPU}$ respectively. Compared with the proposal in [21], our design takes power capping into consideration. Further, in contrast to [18], we consider cooling management and thermal dynamics in this paper. Again, we believe that the following architecture can be extended to data center level energy management:

- An Efficiency Controller (EC) for each server that modifies CPU P-states according to the resource demand of the applications to maintain the CPU utilization within a given range (e.g., $[70\%, 80\%]$). This local controller minimizes the individual server power consumption while meeting the performance requirement (2). It runs at a sub-second to second granularity.
- A Fan Controller (FC) for the blade enclosure that tunes the fan speeds dynamically to maintain the server temperatures below their threshold (3) while minimizing the cooling power. It runs at second or tens of second granularity.
- A Local Power Capper (LPC) for each server that maintains the power consumption of the server below a given threshold through also P-state tuning. This control runs in sub-second to second granularity.
- A Group Power Capper (GPC) for the enclosure that keeps the total power consumption of the servers and fans below a given threshold. This controller enforces the power budget (4) together with the LPC by setting the power threshold for each server. It runs with a longer time interval than the LPC does.
- A Global Controller (GC) for the enclosure that migrates Virtual Machines to consolidate workload and turns servers on/off when needed. Based on measurement collected in the previous GC interval, the GC tries to minimize the total power consumption (1) for the current GC interval while meeting all the three constraints (2-4). Due to the overhead of VM migration and server power status changes, the GC might run less frequently (10 minutes or longer).

# 5. DESIGN AND IMPLEMENTATION

Unified power and cooling systems exhibit complexity due to the interactions between controllers, tradeoffs between the objectives, and conflicts between the constraints. Instead of detailed algorithms, we discuss a few approaches that we found necessary for robust and efficient control. These are especially relevant when power capping is required.

## 5.1 Prioritization of the constraints

Dynamic power management affects both application performance and server thermal conditions. This implies that enforcing a power budget requires the effective compute capacity of the servers to be reduced. In other words, the performance constraint (2) and the power budget constraint (4) can conflict with each other. In our design, capping power has higher priority than maintaining the application performance. This has been mainly implemented at the server level through the integration of the EC and the LPC. Both controllers target tuning CPU frequencies while the lower of the two outputs that implies lower power is applied to CPU [25].

## 5.2 Differentiation of time scales

Although the main knob for power capping, i.e., P-state tuning, can be done at a higher frequency than the fan speed tuning, the GPC has to work slower than the FC. This is because the GPC needs to cap the total power consumption by both the servers and fans below the total budget. In reality, it requires a fast fan controller to achieve responsive power capping. On the other hand, fan speed cannot be tuned with a very high frequency due to physical and mechanical limitations. Thus, there exists a threshold for the response speed of the system to a power budget setting. Careful consideration therefore has to be taken to define the time scales of the controllers.

## 5.3 Necessity of feedback

Feedback is necessary due to a few reasons. First, the knobs, including P-states, turning servers on/off, and workload migration, are all discrete and the workloads are time-varying. Integration control is needed for the EC, LPC, and GPC so that the utilization target or the power consumption budget can be tracked. Second, feedback is needed to track temperatures due to existence of the dynamic thermal process that are slower when compared to dynamic power tuning, when power consumption can respond almost immediately to actions such as P-state tuning. Below we describe two examples that elucidate the incorporation of feedback into the cooling and power capping controllers.

Feedback is introduced into the FC through a Model Predictive Controller (MPC) with three steps of horizon. More specially, in each control interval, the FC decides the new fan speeds by solving the following problem:

$$\min \textstyle\sum_i P_{F_i} + w||FS(k+1) - FS(k)||^2 \quad (10)$$

$$T_{CPU,j}(k+h) \leq T_{ref}, \quad \text{h} = 1,2,3, \text{for each blade j} \quad (11)$$

$$LB_i \leq FS_i \leq UB_i, \quad \text{for each fan i.} \quad (12)$$

The cost function is a weighted sum of the fan power ($P_{F_i}$) and the size of fan speed changes. The weight $w$ acts as one parameter to tune the responsiveness of the controller. The first set of constraints applies upper bounds to the temperature in three steps, predicted using the same speed of fans. The second set of constraints put lower and upper bounds to the fan speeds, or the cooling capacity.

The case with GPC is more complicated. Note that enforcing the cap, once the total power requirement is above the budget, will affect multiple metrics. When the power budget is reduced, the effective capacity of the servers will also be reduced, resulting in a performance loss and less heat generated by the processor. This will lead to lower CPU temperatures and a correspondingly lower power consumption by the fans. As a result of the reduced fan power consumption, more power will be available for the blades. This in turn will reverse the system dynamic described above. There is thus a positive feedback loop between the cap and the total power consumption. Without careful design of the controller and configuration of the parameters, the system will oscillate, and lead to budget violations, temperatures exceeding the threshold, and possibly higher performance loss. On the other hand, due to the complicated dynamics, it is difficult to predict the effect of power capping and apply proactive control. In our case, we used a PI controller for the budget allocation between the fans and the blades. Based on the error between the total budget and the total power consumption of the fans and the blades, the controller tunes the budget for all the servers, which is then shared by the blades proportional to their demand.

## 5.4 Relaxation of constraints

The MPC fan controller is still complicated given the nonlinear models, cost functions, and more importantly, the conflicting constraints. Both the inputs, $FS$, and the outputs, $T_{CPU}$ are bounded, and it is very possible the problem is infeasible for high server utilizations or high ambient temperatures. In our implementation, we introduced one more weighted term into the cost function, defined as a penalty function when the predicted temperatures go above the threshold. This relaxation simplifies the problem.

The GC tries to minimize the total fan power consumption while maintaining the temperature, budget, and performance constraints from interval to interval. Compared to the FC, the problem in the GC is more challenging to address: a larger space exists for the decision variables on the workload placement onto the servers; the decision variables are 0/1 integers; many more constraints exist; and both the cost function and constraints are nonlinear and discontinuous. Simulated annealing (SA) was therefore chosen to solve the optimization problem. We will not discuss the detailed design but, in our implementation, we relax the budget and performance constraints similar to what is done in the FC. Moreover, the weight for the budget violation penalty is higher than that of the performance loss to preserve the higher priority of power capping. For the temperature constraints, the minimum fan power that can meet the temperature threshold is estimated for each workload placement candidate. Relaxation of the constraints reduce the time to find a feasible workload placement candidate in the *neighbor* function of the SA algorithm and significantly speeds up the optimization process.

# 6. EVALUATION THROUGH SIMULATION

To evaluate the hierarchical architecture and the control algorithms, we ran simulations driven by workload utilization traces collected from a production environment [21]. Note that our previous results for a similar architecture [21]

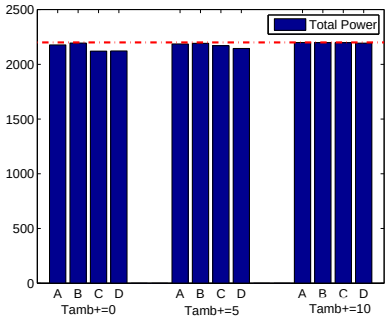


**Figure 3: Total Power Consumption**

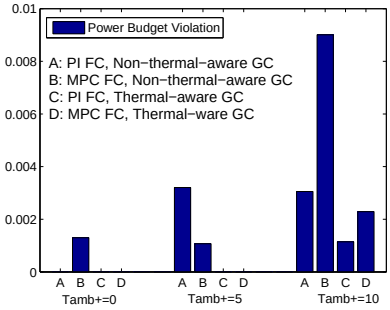


**Figure 4: Power Budget Violation**

did not address the complexities of power capping and therefore had significantly different controller implementations. Through the comparison of statistics in these different experimental scenarios, this paper focuses on the effect of the power capping controller on performance loss and how multiple constraints are met.

Figures 3–7 show the statistics for a series of simulations. $A$, $B$, $C$ and $D$ correspond to combinations of two types of fan controllers and two options for the Global Controller. In cases $A$ and $C$, the fan speeds were tuned through a zonal feedback controller by which the speed of all the fans in each of the two rows was driven by the error between the temperature reference and the highest temperatures of the blades in the corresponding row. The fans were controlled by the MPC controller in cases $B$ and $D$. Both of the two options for the GC controller implemented the Simulated Annealing algorithm and workloads were placed so that the total power could be minimized and capped if necessary. But in cases $C$ and $D$, the minimal fan power (for given workload placement) was derived by solving an optimization problem so that the steady state temperatures could be kept below the threshold. We called this option "Thermal-aware GC" since the cooling efficiency was considered when deciding where to place workloads. In cases $A$ and $B$, no minimal fan power was estimated but the average fan power consumption in the previous GC interval was taken as the estimation. We called this option "Non-thermal-aware GC". For each combination of the controllers, we experimented with three sets of ambient temperatures: those measured in a real data center; those when all the ambient temperatures are 5C higher than the measured ones; and those when they are 10C higher. By doing so we could evaluate how the controllers behave when higher cooling demand is seen. The power capping controllers and constraints were applied in all the cases, with a power budget of $2200W$. The budget was carefully chosen so that it is in the middle of the highest and lowest power consumption of the workloads when there is no capping. Each simulation was run with traces that represent 4 hours of workloads with minimum time granularity of 1 second. All the metrics were collected every second.

A few observations can be made from the results. First, as seen in Figure 3, the mean power consumption was kept below the budget for all 12 cases. Figure 4 shows the percentage of the samples when the total power exceeded the budget by 5%. All of budget violation levels are below 1%, and most of them are below 0.4%.

Capping power consumption can reduce the effective capacity of the servers and lead to performance (throughput in our case) loss. This is shown in Figure 5 as a percentage of the total workload demand over the trace duration. It is obvious that the "Thermal-aware GC" results in a lower performance loss since it optimizes the workload placement in a more power-efficient way. However, the lower performance loss benefits with the "Thermal-aware GC" may not come from only the saving of fan power. Better capacity efficiency could be another reason for this difference. As we can see from Figure 6, even though the performance losses are higher, less fan power is consumed in the $B$ cases than the $C$ cases when the ambient temperature is relative lower. That means the "Thermal-aware GC" can improve capacity utilization as well (resulting in a lower performance loss).

Finally, Figure 7 represents the percentage of temperature samples when they exceeded the $65C$ threshold by 0.5C. In most cases, the temperature was maintained at the reference level, except for case $B$ with the highest ambient temperatures. However, the percentage was still very low, 1.5%, and the largest overshoot (not shown in the figure) was also very small. Further analysis shows that the highest budget and temperature violation levels for case $B$ in Figures 4 and 7 were due to the high ambient temperatures and suboptimal placement of the workloads. In this scenario, even when the fans were run at full speed, the temperatures could not be reduced below the threshold.

# 7. DISCUSSION AND CONCLUSION

Data center level power capping will become an increasingly important tool in the management of global power consumption, particularly as Smart Grid technology is deployed throughout the power delivery infrastructure. However, widespread power capping can negatively impact IT performance and, if not carefully considered, can create conflict with meeting application SLAs. In this paper we have presented key observations and findings from our past and on-going research on unified workload, power, and thermal management for an exemplar server system. Though the results are at the server-level, the general observations on the control and optimization opportunities, challenges, the design principles of the architecture and the controllers, can be extended to the larger data center. For example, the thermal models are critical for proper feedback control at the server-level. At the data center level, air temperature at the server inlet is the key control variable. Like server component temperatures, inlet air temperature is affected by a multiplicity of shared actuators like CRACs and vent tiles. Minimization of global data center power consumption, therefore, will require the development of models that can predict inlet air temperature for given actuator settings and power distribution. Models of this type are currently

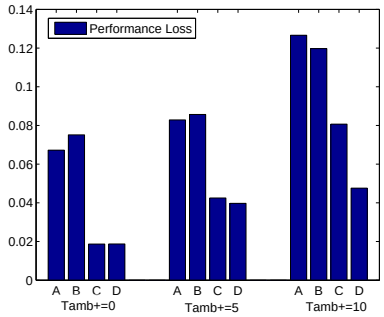


**Figure 5: Performance Loss**

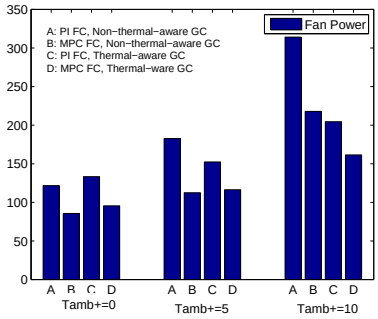


**Figure 6: Fan Power Consumption**

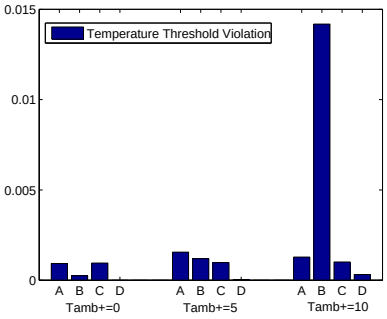


**Figure 7: Temp. Violation**

being developed [4, 5, 22]. As ongoing research, we are investigating extending the framework presented here to that of a large data center using the newly-developed models.

# References


[1] R. Ayoub, S. Sharifi, and T. S. Rosing. Gentlecool: Cooling aware proactive workload scheduling in multi-machine systems. In *Proceedings of DATE 2010*, Dresden, Germany, Mar. 2010.

[2] C. Bash and G. Forman. Cool job allocation: Measuring the power savings of placing jobs at cooling-efficient locations in the data center. In *USENIX Annual Technical Conference*, pages 363–368, Santa Clara, CA, June 2007.

[3] C. E. Bash, C. D. Patel, and R. K. Sharma. Dynamic thermal management of air cooled data centers. In *Proceedings of ITHERM*, pages 445–452, San Diego, CA, May 2006.

[4] T. Breen, E. Walsh, J. Punch, , A. Shah, and C. Bash. From chip to cooling tower data center modeling: Part I influence of server inlet temperature and temperature rise across cabinet. In *Proceedings of ITherm 2010*, Las Vegas, NV, June 2010. In Submission.

[5] T. Breen, E. Walsh, J. Punch, , A. Shah, and C. Bash. From chip to cooling tower data center modeling: Part II – influence of chip temperature control philosophy. In *Proceedings of ITherm 2010*, Las Vegas, NV, June 2010. In Submission.

[6] J. S. Chase, D. C. Anderson, P. N. Thakar, A. M. Vahdat, and R. P. Doyle. Managing energy and server resources in hosting centers. In *Proceedings of the 18th ACM Symposium on Operating Systems Principles*, pages 103–116, Banff, Canada, Oct. 2001.

[7] 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 '07)*, pages 13–23, San Diego, CA, June 2007.

[8] S. Greenberg, E. Mills, B. Tschudi, P. Rumsey, and B. Myatt. Best practices for data centers: Results from benchmarking 22 data centers. In *Proceedings of the 2006 ACEEE Summer Study on Energy Efficiency in Buildings.*, Pacific Grove, CA, Aug. 2006.

[9] T. Heath, A. P. Centeno, P. George, L. Ramos, Y. Jaluria, and R. Bianchini. Mercury and freon: Temperature emulation and management for server systems. In *Proceedings of ASPLOS*, pages 106–116, San Jose, CA, Oct. 2006.

[10] R. Jorgensen, editor. *Fan Engineering.* Buffalo Froge Company, 8th edition, 1983.

[11] C. Lefurgy, X. Wang, and M. Ware. Power capping: A prelude to power shifting. *Cluster Computing*, 11(2):183–195, 2008.

[12] J. Moore, J. Chase, P. Ranganathan, and R. Sharma. Making scheduling "cool": Temperature-aware workload placement in data centers. In *Proceedings of the Annual Conference on USENIX Annual Technical Conference*, pages 61–75, Apr. 2005.

[13] M. N. Özisik. *Heat Transfer: A Basic Approach.* McGraw-Hill Companies, 1985.

[14] C. Patel, R. Sharma, C. Bash, and S. Graupner. Energy aware grid: Global workload placement based on energy efficiency. In *Proceedings of IMECE 2003*, Washington, DC, Nov. 2003.

[15] C. D. Patel, C. E. Bash, R. Sharma, M. Beitelmam, and R. J. Friedrich. Smart cooling of data centers. In *Proceedings of IPACK'03*, Kauai, Hawaii, July 2003.

[16] E. Pinheiro, R. Bianchini, E. V. Carrera, and T. Heath. Dynamic cluster reconfiguration for power and performance. *Compilers and Operating Systems for Low Power*, pages 75–93, 2003.

[17] A. Qureshi, H. Balakrishnan, J. Guttag, B. Maggs, and R. Weber. Cutting the electric bill for internet-scale systems. In *Proceedings of SIGCOMM 2009*, pages 123–134, Barcelona, Spain, Aug. 2009.

[18] R. Raghavendra, P. Ranganathan, V. Talwar, Z. Wang, and X. Zhu. No "power" struggles: Coordinated multi-level power management for the data center. In *Proceedings of ASPLOS*, pages 48–59, Seattle, WA, Mar. 2008.

[19] P. Ranganathan, P. Leech, D. E. Irwin, and J. S. Chase. Ensemble-level power management for dense blade servers. In *33rd International Symposium on Computer Architecture (ISCA 2006)*, pages 66–77, Boston, MA, June 2006.

[20] Q. Tang, S. K. S. Gupta, and G. Varsamopoulos. Thermal-aware task scheduling for data centers through minimizing heat recirculation. In *Proceedings of the 2007 IEEE International Conference on Cluster Computing (CLUSTER '07)*, pages 129–138, Austin, TX, Sept. 2007.

[21] N. Tolia, Z. Wang, P. Ranganathan, C. Bash, M. Marwah, and X. Zhu. Unified power and cooling management in server enclosures. In *Proceedings of the ASME/Pacific Rim Electronic Packaging Technical Conference and Exhibition (InterPACK '09)*, San Francisco, CA, July 2009.

[22] M. Toulouse, G. Doljac, V. Carey, and C. Bash. Exploration of a potential-flow-based compact model of air-flow transport in data centers. In *Proceedings of the 2009 ASME IMECE*, Lake Buena Vista, FL, Nov. 2009.

[23] U.S. Environmental Protection Agency (EPA). Report to congress on server and data center energy efficiency, public law 109-431, Aug. 2007.

[24] L. Wang, A. J. Younge, T. R. Furlani, G. von Laszewski, J. Dayal, and X. He. Towards thermal aware workload scheduling in a data center. In *Proceedings of the 10th International Symposium on Pervasive Systems, Algorithms and Networks (I-SPAN 2009)*, Kao-Hsiung, Taiwan, Dec. 2009.

[25] Z. Wang, X. Zhu, C. McCarthy, P. Ranganathan, and V. Talwar. Feedback control algorithms for power management of servers. In *3rd International Workshop on Feedback Control Implementation and Design in Computing Systems and Networks (FeBID)*, Annapolis, Maryland, June 2008.

[26] Z. Wang, C. Bash, N. Tolia, M. Marwah, X. Zhu, and P. Ranganathan. Optimal fan control for thermal management of servers. In *Proceedings of the ASME/Pacific Rim Electronic Packaging Technical Conference and Exhibition (InterPACK '09)*, San Francisco, CA, July 2009.