

# CoolIT: Coordinating Facility and IT Management for Efficient Datacenters \*

*Ripal Nathuji<sup>1</sup>, Ankit Somani<sup>2</sup>, Karsten Schwan<sup>1</sup>, and Yogendra Joshi<sup>2</sup>*  
<sup>1</sup>*Center for Experimental Research in Computer Systems (CERCS)*  
<sup>2</sup>*Consortium for Energy Efficient Thermal Management (CEETHERM)*  
*Georgia Institute of Technology*  
*Atlanta, GA 30032*  
*rnathuji@ece.gatech.edu, asomani3@mail.gatech.edu,*  
*schwan@cc.gatech.edu, yogendra.joshi@me.gatech.edu*

## Abstract

The expanding computational needs of today’s businesses and organizations have led to a significant increase in the number and size of datacenter facilities. While technological advances in cooling solutions and server densities are enabling the performance levels required by applications, the energy costs of these computational and heat management solutions may soon outpace the capital expense of purchasing them. To address the associated issues of cost, scalability, and sustainability, prior work has created management solutions that enable efficient operating conditions, including techniques for power aware workload management in the IT space, and at the facility level, new cooling and thermal management methods. This paper articulates the opportunities derived from the ability to coordinate across these two management domains, by describing CoolIT, an ongoing collaborative effort between Computer Science (CERCS) and Mechanical Engineering (CEETHERM) researchers to improve overall datacenter efficiency.

## 1 Introduction

Modern datacenters are achieving significant advances in computational capabilities. Improved densities enabled by small form factor blade platforms and multicore processor architectures have fueled this trend, but they are also instigating a significant problem concerning the sustainability of these environments. In the U.S., datacenter facilities consume approximately 2% of all electricity, with an estimated growth rate of 12% per year [1]. Unmitigated, these statistics underscore fundamental potential bottlenecks of energy costs and provisioning. To alleviate such restrictions, it is clear that datacenters must incorporate aggressive energy efficient technologies to enable “green” computing.

Recent surveys have identified that compute hardware alone can consume from 33% to 75% of datacenter power [5]. In addition, the cooling (HVAC), power delivery, and other infrastructure subsystems required to support the IT equipment can constitute up to 50% of power [1]. These characteristics emphasize two promising avenues for improving datacenter efficiency: (1) reducing the active power signature of compute resources when executing applications, while simultaneously meeting service level agreements (SLAs); and (2) minimizing the energy costs of the HVAC systems used to cool server hardware, by balancing cooling capacity with the dynamic heat generation of workloads.

The research described in this paper demonstrates the need for going beyond implementing these two design goals independently and instead adopting an approach that coordinates across them. The CoolIT project is an ongoing effort to realize such coordination, undertaken jointly by Computer Science researchers and Mechanical Engineers at Georgia Tech. The preliminary learnings presented here utilize a novel cooling management framework, AILM [19], that can be driven with online measurements to manage the cooling system and enable coordination by determining constraints for IT workload placement policies. Initial results attained with AILM highlight the interesting tradeoffs that can be exploited by a coordinated management solution illustrating the benefits of coordination.

## 2 Coordinated Management

In the IT space, workloads are managed using VM migration capabilities to consolidate applications amongst servers in a manner that can reduce the power footprint of resources while still meeting performance requirements. In the facilities cooling domain, a similarly useful mechanism that is being increasingly integrated into modern CRAC units is the ability to change the air velocity at which the system pushes cool air into the datacenter

---

\*Funding for this research was provided in part by the National Science Foundation and by OSISoft Corporation.

room [9]. A given air flow rate dictates the amount of heat that can be dissipated from servers, and therefore determines whether thermal constraints are violated. By judiciously reducing air velocity based upon load, heat can be dissipated while minimizing the power profile of the cooling infrastructure.

Management decisions made to maximize the efficiency of IT power usage can conflict with load allocations that allow for optimal efficiency of cooling components. For example, a heterogeneity-aware VM allocation policy may decide to migrate workloads to a set of servers that are optimal in terms of energy usage [14], but happen to be physically placed in a localized area of the datacenter facility. The heat generated by these resources may create a hot spot that requires the HVAC to utilize a higher air flow rate than what could have been achieved with an alternate allocation. Therefore, independent decisions made in these two domains are not likely to achieve a balance in which overall datacenter efficiency is maximized. Instead, coordination between the IT-level and facility-level management decisions is necessary to attain balanced operating points.

The goal of CoolIT is to allow management components that drive IT and cooling systems to behave in a synergistic manner, by coordinating their actions. In the remainder of this section, we briefly present the VirtualPower-based IT management infrastructure we have been developing within the Computer Science research group. We then proceed to describe the thermal-driven AILM tool used for cooling management. We conclude by discussing how these components can interact for coordinated management.

## 2.1 VirtualPower-Based IT Management

Modern datacenters routinely use virtualization [2, 20], often supported by CPU hardware [17] extensions, to attain basic benefits like fault and performance isolation, and to perform power friendly operations such as resource consolidation via VM migration [4]. The VirtualPower software architecture developed in our research allows IT level management policies to reduce the power consumption of hardware resources while at the same time, meeting the performance requirements of workloads.

In the VirtualPower system, IT management policies distributed between local (PM-L) and global (PM-G) components are driven by VM application modifications to virtualized ACPI states, termed VirtualPower Management (VPM) states. Actuations on these states are communicated via a VPM channel abstraction to IT policies. The policies then use this information to exercise a rich set of VPM mechanisms, such as setting physical operating points, scheduling resources to VMs within a plat-

form, and migrating applications across servers. Further details regarding this architecture and its implementation can be found in our previous work [15], and approaches towards enhancing the flexibility and generality of the system are discussed in [11].

In addition to the components outlined, we have further extended management abstractions to aid in setting and adhering to power budgets in a Quality of Service (QoS) aware manner. By using sets of VPM Tokens [16], IT policies can strictly allocate power when necessary limits must be imposed due to power delivery or thermal management considerations. As we will describe in this paper, these budgeting capabilities can also help enable coordination between the VirtualPower IT management framework, and cooling based management trade-offs provided by the AILM approach.

## 2.2 AILM Based CRAC Management

To develop efficient cooling systems in datacenters, we have developed a tool that can be used to determine the load capacities maintainable within a particular CRAC air velocity. The ‘ambient intelligence-based load management’ (AILM) approach is designed to manage a datacenter from a cooling-centric perspective, by determining within a given air velocity, the heat load limits at each server that must be observed to prevent thermal violations.

The guiding principle of the AILM algorithm is the temperature field linearity concept when the flow conditions remain the same inside the room, as the temperature variations are not high enough to cause a density variation. Thermal radiation effects are also negligible. Thus, a change in the volumetric heat generation of server  $i$  (for the given facility) present in the room contributes towards the change in the inlet temperature of server  $j$  (for the given facility). As a result, a datacenter can be calibrated based on how much a unit change in volumetric heat generation at server  $i$  can alter the inlet temperature of server  $j$ . An overview of the algorithm based upon this premise is provided in Figure 1.

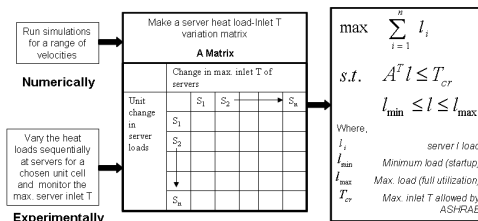


Figure 1: AILM Approach.

In order to use the AILM tool, it must be calibrated

to a datacenter configuration. As the figure shows, the tool can use either experimental or simulated data. The results in this paper are based upon a simulation-driven AILM using the Fluent computational fluid dynamics tool. For a given set of velocities of the CRAC unit, a baseline is obtained based upon the minimum power dissipation the servers can achieve. For this condition, the maximum temperature at the inlet of each server is noted. Next, for a server, the power dissipated is increased by a unit amount and the system is left to reach steady state without changing the CRAC velocity. The maximum temperatures at the server inlets are noted again, and the difference from the baseline case is recorded. This difference, when calculated for all of the servers, provides an estimate of how sensitive server  $j$ 's inlet T is with respect to server  $i$ 's heat load, for a given CRAC velocity. This process is repeated sequentially for all of the servers. The outcome is a  $n \times n$  matrix of values, for a CRAC velocity and for  $n$  servers. By using this method for each reasonable CRAC velocity, the output is the maximum power dissipation a datacenter's cooling system can handle. To calculate this and the respective power dissipation for each server, we optimize the server loads within the constraints of the maximum and minimum loads of servers and the maximum critical inlet server temperature (as specified by ASHRAE). Once the above calibration is performed, AILM effectively provides for each velocity  $V_x$  a total possible maximum load capacity, as well as a load vector that maps a power budget to each server. As we will discuss next, this interface information can be used by the IT management infrastructure to manage the datacenter in a coordinated fashion.

### 2.3 The CoolIT Approach

Having described the VirtualPower and AILM driven management components that are designed to improve IT and facility CRAC efficiencies, respectively, we now describe the CoolIT vision towards coordinating the two systems. As Figure 2 shows, the IT management infrastructure strives to allocate and manage VMs in a manner that minimizes the power consumption of server resources. At the same time, we would like to use the AILM thermal assessment capabilities to determine whether implementing certain power limits on servers would allow for the load capacity being used, but would also provide for a reduced operating point for the CRAC. This results in a cyclical coordination relationship, as illustrated in Figure 2. In the remainder of this paper, we utilize the ability to drive AILM with simulation-based data to determine what types of tradeoffs and dependencies the coordination feedback loop depicted above should try to address.

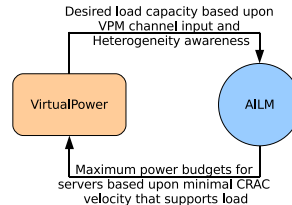


Figure 2: CoolIT Coordinated Management.

### 3 Evaluation: Initial Results with AILM

Consider an 8 rack configuration with rows of four racks each and one cold aisle as shown in Figure 3. The rack labels identify the type of rack (e.g., composed of type 'A' or 'B' hardware), assuming a heterogeneous equipment configuration. With a naming convention that assumes that racks across the cold aisle are of the same type, the setup depicted in Figure 3 would be an 'ABAB' configuration.

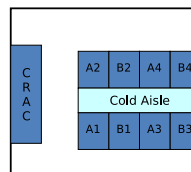
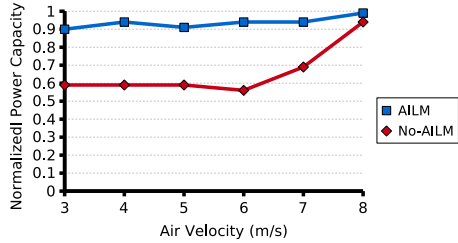


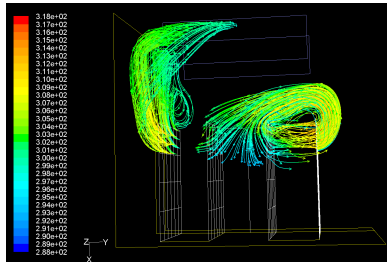
Figure 3: Datacenter Configuration.

Starting with a homogeneous datacenter configuration, we use AILM to determine the overall load capacity the datacenter can support for a given CRAC air velocity. Figure 4(a) provides the results normalized to the maximum possible load, comparing an 'AILM' case to a 'No-AILM' case. The 'AILM' case assumes that load is allocated based upon thermal effects captured by AILM to allow for intelligent load management within a given CRAC velocity. In the 'No-AILM' case, load is distributed using a simple load balancing scheme where each server is provided equal workload up until the point that the thermal constraints are violated somewhere, thereby constituting a completely thermally unaware solution.

As apparent from the figure, with 'AILM' knowledge, across all air velocities, the IT system can operate to within 10% of full load. In the 'No-AILM' case, however, there are limits of up to 40% for lower air velocities. This difference can be attributed to the fact that the AILM model is able to capture air recirculation effects like those illustrated in Figure 4(b). In our model, these effects can cause significant hotspots at a small set of servers. By allowing an IT manager to be aware of these



(a) Maximum Load Capacity



(b) Recirculation Effects Captured by AILM

Figure 4: Datacenter Capacity Benefits with AILM Recirculation Awareness.

bottlenecks, close to maximum load can be supported by simply shutting these systems down. Therefore, in this example, with AILM we can actually utilize a low operating point for the CRAC as long as we adhere to the set of per server power budgets provided by AILM.

Given the importance of heterogeneity-awareness in datacenter management [7, 14], we next consider two types of servers whose performance and power characteristics vary based upon actual measurements of our Intel P4- vs. Core-based hardware [15]. The P4 hardware consumes more power, while the Core hardware is more power efficient and can improve performance within a smaller power profile. Figure 5 contains the resulting power and performance capacities using AILM, where the datacenter configuration is ‘ABAB’ (type A racks are based upon the P4 hardware).

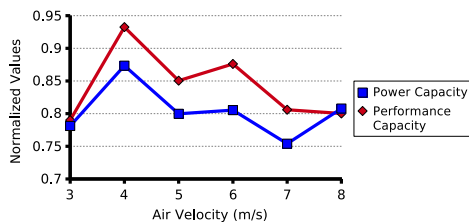


Figure 5: Power and Performance Capacity Analysis for Heterogeneous Configurations.

The figure depicts two interesting trends. First, neither

power nor performance capacity increase asymptotically with air velocity. This can be attributed to the recirculation effects found in the homogeneous case. Further, at higher air flow rates, changes in air flow patterns occur that actually create hot spots that don’t exist at lower air velocities. In other words, despite the expenditure of increased power on cooling, at higher flow rates, there is no or negative benefit to datacenter performance capacity. This counterintuitive result underscores the need for utilizing a thermal aware management component like AILM with the IT management domain. A second interesting outcome is the change in performance and power capacity between 7 and 8 m/s. Here, the power capacity (cost) increases, but performance actually decreases slightly. This is because more power can be consumed, but it is within the less efficient type ‘A’ racks and at the expense of the type ‘B’ racks. Therefore, the aggregate work that can be performed is slightly reduced. This effect shows that a purely thermally aware solution is not appropriate, either. Instead, it should rely on IT management support to understand workload performance and use constraints from thermal characteristics to guide its decision. This effect is magnified in our final results.

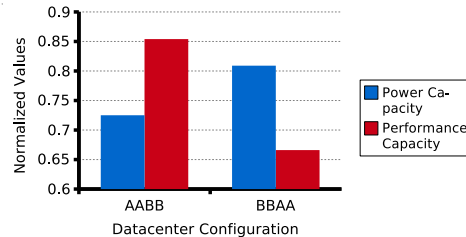


Figure 6: Datacenter Configuration Tradeoffs with Heterogeneous Configurations.

Figure 6 presents the power and performance capacity at a particular air flow velocity (7 m/s), but with two different datacenter configurations. Here, the maximum load capacity that can be supported is higher with the ‘BBAA’ configuration, by 8%. Without IT workload awareness, a physical configuration based upon this AILM result would dictate utilizing such a setup. However, as we observe from the associated performance capacity, the ‘AABB’ configuration can achieve a maximum performance that is 18% higher. These simulation-based results again show that neither IT nor cooling governor alone should dictate management decisions, but, instead, coordination should be applied.

## 4 Related Work

There has been significant recent work focused on power management of compute resources. Methods have been

developed to utilize capabilities such as processor voltage/frequency scaling for reduced power profiles of processors and platforms [8]. Storage resources have also provided a strong opportunity to reduce power and thermal usage in enterprise systems [21]. The importance and benefits of being able to manage heterogeneous compute resources in the IT space have been documented from low level processor management to multi-platform management [10, 14]. The CoolIT approach strives to leverage these contributions, while coordinating them with thermal based management decisions.

At the datacenter level, power consumption can be reduced by turning servers off and bringing them online based on demand [3]. Other proposed datacenter management approaches have considered temperature-aware workload placement [13]. Emulation tools that estimate the thermal implications of power management can aid in the offline design of management policies as well [6]. Compared to such tools, and the heuristic based thermal prediction approach proposed by Moore et al. [12], AILM uses a temperature linearity concept when the flow field remains the same, and applies a simplex based optimization framework. This allows it to be used for online management that leverages the use of coordination, a beneficial control paradigm for computing systems [15, 11, 18], for improving efficiencies in modern datacenters.

## 5 Conclusions and Future Work

The principal argument advanced in this paper is that additional efficiencies in datacenter power consumption can be attained by coordinating the operation of the IT and the cooling management subsystems. Benefits attained from such coordination can be substantial, as demonstrated by simulation-based results presented in this paper. For future work, our team of Mechanical Engineering and Computer Science researchers is constructing a fully instrumented and manageable datacenter facility in which coordinated IT and cooling management will be performed. At the same time, we have already completed several initial steps toward that end. First, we have developed IT management solutions based on the Xen virtualization infrastructures that have been shown effective for power management at a datacenter level [16]. Further, we have recently enhanced these infrastructures to be able to interface with other management subsystems, including those used for cooling management. Second, the AILM tool used in this paper is equally capable of operating with measurements produced by thermal sensors, and it operates at real-time speeds (in contrast to full-scale fluid flow modeling tools like Fluent). From these facts, it is apparent that coordinated operation across these management domains is

feasible. Potential benefits from such coordination are shown by the simulation results presented in this paper.

## References

- [1] Report to congress on server and data center energy efficiency. Technical report, U.S. Environmental Protection Agency, Energy Star Program, August 2007.
- [2] P. Barham, B. Dragovic, K. Fraser, S. Hand, T. Harris, A. Ho, R. Neugebauer, I. Pratt, and A. Warfield. Xen and the art of virtualization. In *Proceedings of the ACM Symposium on Operating Systems Principles (SOSP)*, 2003.
- [3] J. Chase, D. Anderson, P. Thakar, A. Vahdat, and R. Doyle. Managing energy and server resources in hosting centers. In *Proceedings of the 18th Symposium on Operating Systems Principles (SOSP)*, 2001.
- [4] 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 ACM/USENIX Symposium on Networked Systems Design and Implementation (NSDI)*, May 2005.
- [5] 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 ACEEE Summer Study on Energy Efficiency in Buildings*, 2006.
- [6] T. Heath, A. P. Centeno, P. George, L. Ramos, Y. Jaluria, and R. Bianchini. Mercury and freon: Temperature emulation and management in server systems. In *Proceedings of the International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, October 2006.
- [7] T. Heath, B. Diniz, E. V. Carrera, W. Meira Jr., and R. Bianchini. Energy conservation in heterogeneous server clusters. In *Proceedings of the 10th Symposium on Principles and Practice of Parallel Programming (PPoPP)*, 2005.
- [8] C. Isci, G. Contreras, and M. Martonosi. Live, runtime phase monitoring and prediction on real systems with application to dynamic power management. In *Proceedings of the 39th International Symposium on Microarchitecture (MICRO-39)*, December 2006.
- [9] Y. Joshi, E. Samadiani, and F. Mistree. Reduced modeling based robust thermal design of energy efficient data centers. In *Proceedings of the Eighteenth International Symposium on Transport Phenomena*, August 2007.
- [10] R. Kumar, D. Tullsen, P. Ranganathan, N. Jouppi, and K. Farkas. Singleisa heterogeneous multi-core architectures for multithreaded workload performance. In *Proceedings of the International Symposium on Computer Architecture (ISCA)*, June 2004.
- [11] S. Kumar, V. Talwar, P. Ranganathan, R. Nathuji, and K. Schwan. M-channels and m-brokers: Coordinated management in virtualized systems. In *Proceedings of the Workshop on Managed Many-Core Systems (MMCS)*, June 2008.
- [12] J. Moore, J. Chase, and P. Ranganathan. Weatherman: Automated, online, and predictive thermal mapping and management for data centers. In *Proceedings of the IEEE International Conference on Autonomic Computing (ICAC)*, June 2006.
- [13] J. Moore, J. Chase, P. Ranganathan, and R. Sharma. Making scheduling cool: Temperature-aware workload placement in data centers. In *Proceedings of the USENIX Annual Technical Conference*, June 2005.
- [14] R. Nathuji, C. Isci, and E. Gorbato. Exploiting platform heterogeneity for power efficient data centers. In *Proceedings of the IEEE International Conference on Autonomic Computing (ICAC)*, June 2007.
- [15] R. Nathuji and K. Schwan. Virtualpower: Coordinated power management in virtualized enterprise systems. In *Proceedings of the 21st ACM Symposium on Operating Systems Principles (SOSP)*, October 2007.
- [16] R. Nathuji and K. Schwan. Vpm tokens: Virtual machine-aware power budgeting in datacenters. In *Proceedings of the ACM/IEEE International Symposium on High Performance Distributed Computing (HPDC)*, June 2008.
- [17] G. Neiger, A. Santoni, F. Leung, D. Rodgers, and R. Uhlig. Intel virtualization technology: Hardware support for efficient processor virtualization. In *Intel Technology Journal* (<http://www.intel.com/technology/itj/2006/v10i3/>), August 2006.
- [18] R. Raghavendra, P. Ranganathan, V. Talwar, Z. Wang, and X. Zhu. No power struggles: A unified multi-level power management architecture for the data center. In *Proceedings of the International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, March 2008.
- [19] A. Somani and Y. Joshi. Ambient intelligence based load management. USPTO Provisional Patent, GTRC ID No. 4524, 2008.
- [20] J. Sugerman, G. Venkitachalam, and B.-H. Lim. Virtualizing i/o devices on vmware workstation's hosted virtual machine monitor. In *Proceedings of the USENIX Annual Technical Conference*, 2001.
- [21] Q. Zhu, Z. Chen, L. Tan, Y. Zhou, K. Keeton, and J. Wilkes. Hibernator: Helping disk arrays sleep through the winter. In *Proceedings of the 20th ACM Symposium on Operating Systems Principles (SOSP)*, October 2005.