Thermal Modeling and Analysis of Storage Systems

Xunfei Jiang*, Mohammed I. Alghamdi†, Ji Zhang*, Maen Al Assaf‡, Xiaojun Ruan§, Tausif Muzaffar*, and Xiao Qin*

*Department of Computer Science and Software Engineering, Auburn University, Auburn, AL 36849-5347
{xzj0009, jzz0014, tausifm, xqin}@auburn.edu
†Department of Computer Science, Al-Baha University, Al-Baha City, Kingdom of Saudi Arabia, mialmushilah@bu.edu.sa
‡King Abdullah II School for Information Technology, The University of Jordan, Amman, Jordan, m_alassaf@ju.edu.jo
§Department of Computer Science, West Chester University of Pennsylvania, West Chester, PA 19383, xruan@wcupa.edu

***Abstract*—Recognizing that power and cooling cost for data centers are increasing, we address in this study the thermal impact of storage systems. In the first phase of this work, we generate the thermal profile of a storage server containing three hard disks. The profiling results show that disks have comparable thermal impacts as processing and networking elements to overall storage node temperature. We develop a thermal model to estimate the outlet temperature of a storage server based on processor and disk utilizations. The thermal model is validated against data acquired by an infrared thermometer as well as build-in temperature sensors on disks. Next, we apply the thermal model to investigate the thermal impact of workload management on storage systems. Our study suggests that disk-aware thermal management techniques have significant impacts on reducing cooling cost of storage systems. We further show that this work can be extended to analysis the cooling cost of data centers with massive storage capacity.**

***Keywords*-Thermal; Model; Storage System;**

I. Introduction

Thermal management techniques for storage systems can significantly impact the cooling costs of data centers; traditional thermal models for data centers do not take into account disk utilizations. In this paper, we address the thermal impact of hard disks by developing a thermal model for storage systems. We show how to apply the thermal model to estimate the outlet temperature of a storage server based on processor and disk utilizations. With the thermal model in place, we investigate the thermal impact of workload management on storage systems.

Motivations. Our proposed thermal model is indispensable for next-generation storage systems because of the following five factors:

1) the ever-increasing cooling and energy costs of large-scale storage systems,
2) the impact of storage systems' temperature on cooling costs of data centers,
3) the growing importance of reducing thermal monitoring cost,
4) the capability of estimating the cooling cost of a data center during its planning phase, and
5) the lack of study on the impacts of disk utilizations and temperatures on outlet temperatures of a data node.

With ever-increasing energy consumption and cooling costs of large-scale storage systems, data center designers need to urgently address the energy efficiency issues [11]. The electricity cost of maintaining a data center for four years may be equivalent to the cost of building a new data center. Traditional approaches to saving energy cost for data centers is to improve the energy efficiency of servers and storage systems as well as cooling systems.

Growing evidence shows that cooling costs contribute a significant portion of the operational cost of data centers [5][11]. For example, the power and cooling infrastructure supporting IT equipment can consume up to 50% of energy in a data center [5]. Prior research shows reducing the energy dissipation in cooling systems can effectively improve the energy efficiency of data centers [17][26]. For instance, energy cost of cooling systems in data centers can be saved by reducing the outlet temperatures of servers or optimizing the air recirculation [23]. A handful of workload placement strategies are proposed to balance temperature distribution through workload management [17] [26]. Experimental results obtained by Moore *et al.* show that energy consumption of a data center can be saved up to 40% by setting a low outlet temperature of data nodes [17]. Reducing the temperature of hard disks in storage systems can not only conserve the energy consumption in cooling systems, but also enhance the reliability and lifetime of the storage systems [19][28].

The energy models of storage systems have been investigated in the past years. For example, Allalouf *et al.* proposed a model to estimate the power consumption of storage nodes running under certain workload conditions [6]. However, thermal models of storage systems are still in its infancy. Little attention has been paid to the impact of disk temperatures on the energy efficiency of cool systems in data centers.

Setting up temperature sensors in data nodes of a storage system is a common way of monitoring the system's temperature. Given a single data node, one can apply at

least two sensors to monitor inlet and outlet temperatures of the node. In case detailed interior temperatures of the data node need to be measured, additional sensors must be deployed. Although this approach is practical to measure temperatures of small-scale storage systems, it becomes an infeasible solution when a storage system consists of hundreds of thousands nodes. It is prohibitively expensive to acquire and set up a huge number of sensors in a large-scale data center; deploying sensors can lead to extra energy cost. Thermal models are a promising alternative to monitoring temperatures of storage systems.

A data center is a large investment for many companies. A great deal of planning is a must to ensure a high return on investment. Cooling and power are two important considerations to be addressed during the planning process of data centers. Thus, accurately estimating the cooling and energy cost of a data center is a key guideline during the planning phase. Thermal models and simulators can used to help data center designers to make critical decisions on thermal management during the design phase.

A variety of factors contribute to the outlet temperatures of storage systems. Tang *et al.* show how the outlet temperature of a data node is affected by its inlet temperature and CPU utilization [23]. Li *et al.* propose a model to forecast temperature of a data node using historical temperatures and air flow measurements [16]. Kim *et al.* investigate the relationship between seek times and disk temperatures [14]; Kim's study demonstrates how platters affect disk temperatures. In a modern storage system, a data node is comprised of up to more than 100 hard disks [18]. The temperatures of these disks play a crucial role in affecting the the data node's temperature. Unfortunately, there is a lack of study on the impacts of disk utilizations and temperatures on outlet temperatures of a data node.

Contributions. The goal of this study is to build a thermal model to estimate the outlet temperature of a storage server (a.k.a., data node) based on processor and disk utilizations. We make the following three contributions. First, we generate the thermal profile of a storage server containing multiple hard disks. The profiling results are obtained by running I/O intensive workloads imposed by Postmark [13]. When the disks are running under various load scenarios, we monitor disk temperatures as well as the inlet and outlet temperatures of the data node. Second, we build a thermal model to estimate inlet/outlet temperature differences using inlet temperatures, and workloads. The model also is able to derive outlet temperatures from CPU and disk utilizations. Third, to demonstrate the usage of the model, we make use of this model to investigate the impact of disk temperatures on the cooling cost of storage systems.

Organization. The rest of this paper is organized as follows. The next section presents prior studies and related research issues. Section III describes four preliminary experiments and observations. In Section IV, we develop a thermal model for storage systems. We validate the thermal model against real-world measurement acquired by a thermal meter as well as build-in temperature sensors on disks. In Section V, we discuss the impact of data placement on cooling cost. Finally, Section VI concludes the paper.

II. Relative work

A. Energy-Efficient Data Centers

Increasing attention has been paid to energy efficiency of data centers [8][7]. A study conducted by Koomey in 2000 shows that the total energy consumption in data centers is approximately 1.2% of U.S. energy consumption [15]. A reason behind the striking energy consumption in data centers is the rapid growth of computing and storage capacity in recent years.

Researchers have proposed a number of energy-saving approaches to reduce energy costs of data centers. For example, Bieswanger *et. al* developed measurement and management technologies (MMT, for short) for an energy-efficient data centers [12]. The MMT model integrates real measurements by deploying sensors in data centers, thereby providing run-time analysis of energy consumption. Based on these analytical data, data centers can be operated in an optimal schedule in terms of energy consumption.

Greenberg *et. al* benchmarked 22 data centers and observed that annual energy cost per square foot of a typical data center is more than 15 times of an office building [9]. Greenberg *et. al* also examined a set of best practices, including air management, optimizing the size of data centers, utilizing free cooling by using chilled water and the like. The data collected from these practices indicate that energy savings in data centers can be potentially achieved.

Verma *et. al* proposed an approach called sample-replicate-consolidate mapping or SRCMap to enable energy proportionality for dynamic I/O workload [27]. SRCMap activates a minimal number of physical volumes, in which a selected subset of data in other volumes is duplicated. When serving I/O requests, SRCMap redirects the requests to replicas on active volumes in order to keep other volumes in the sleep mode as long as possible. The experimental results show that SRCMap is able to effectively reduce the power consumption of enterprise storage systems.

B. Thermal-aware Resource Management Strategies

Energy efficiency is an important issue in both data center planning and maintenance. A handful of studies focus on optimizing power consumption of cooling systems - a major contributor to the power cost of data centers. In these studies, thermal-aware resource management strategies were proposed to balance the temperature distribution among data nodes in data centers.

Sharma *et al.* developed a $thermal\text{-}load\text{-}balancing$ framework by applying local and regional policies for dynamical workload distribution. The simulation results

show that uniform temperature distribution, promoted by an asymmetric workload placement, is able to reduce energy consumption and improve equipment reliability. Tang *et al.* highlighted recirculation process in data centers [24]. They proposed a task scheduling algorithm, XInt, for homogeneous data centers, thereby minimizing recirculation costs by balancing the workload within the data center.

Tang *et al.* observed that cooling costs significantly depend on peak inlet temperatures [25]. In order to achieve the lowest cooling power, Tang *et al.* designed a task assignment policy, MPIT-TA, that minimizes the peak inlet temperature. They simulated a small-scale data center; the results show that at MPIT-TA offers at least 20% of cooling energy savings.

Temperature-aware loading balancing strategies are investigated in [20] [21]. Taking into account energy conservation, the strategies limit CPU temperatures to a customized threshold. If the CPU temperatures exceed the threshold, the CPU's voltage and frequency will be dynamically adjusted with execution time penalty.

Thermal-aware resource management techniques for processors have been extensively studied. However, the impact of disks on thermal management has not been fully explored. To provide large data capacity, each data node may contain a number of disks. Under I/O-intensive workload, disk utilization is extremely high. Hence, appropriately managing I/O workload can potentially reduce the cooling cost in data centers.

C. Disk Energy Consumption and Temperature Models

Tan *et al.* built a three-dimensional model to evaluate transient temperatures during frequent seeking [22]. However, the impact of workload on disk temperatures has been overlooked in the past years.

Gurumurthi *et al.* constructed an integrated disk drive model used to investigate the thermal behavior of a hard disk [10]. The model calculates the heat generated from the following components: internal drive air, spindle motor, the base and cover of disk, the voice-coil motor, and disk arms. Kim *et al.* built the relationship between seek times and disk temperatures [14]. They also studied the thermal behaviors of disks by varying platter types and number of platters. It is worth noting that the above studies ignore the impact of disk temperatures on cooling systems. In this paper, we comprehensively evaluate the impact of CPU and disk temperatures on the inlet and outlet temperatures of a data node.

III. Thermal Impacts of Disk I/O

A. Testbed

To characterize the impacts of CPU and disks on the inlet/outlet temperatures of a data node, we conduct a number of experiments on a Linux server, in which CPU temperatures are detected by software lm-sensors [2] and disk temperatures are measured by hddtemp [1]. The inlet and outlet temperature are acquired by an infrared thermometer.

The testbed used in these experiments is equipped with four Intel(R) Xeon 2.4 GHz CPU, 2.0 GBytes RAM, and three 160 GBytes SATA disks deployed in a disk array. The configuration parameters are summarized in Table I.

Table I Testbed Configurations

Hardware	Software
4 × Intel(R) Xeon 2.4 GHz CPU X3430 1 × 2.0 GBytes of RAM 3 × WD 160 GBytes Sata disk (WD1600AAJS-75M0A0 [4])	Ubuntu 10.04 Linux kernel 2.6.32

B. Impact of CPU and Disks on Inlet/Outlet Temperatures

Outlet temperatures of a node are determined by various factors, including CPU and disk temperatures, motherboard temperatures, and inlet temperatures. The CPU factor has been addressed in prior studies (see, for example, [25] [20] [21]). Unfortunately, the thermal impact of disk I/O on data nodes remains an open issue. To investigate the relationship between CPU/disks and the inlet/outlet temperatures, we conduct four groups experiments, in which a combination of high (100%) and low (0%) utilizations of CPU and disks are considered. The configuration details are shown in Table II. In these experiments, CPU and I/O workloads are generated by stress [3] and postmark [13], respectively. The power consumption of the testbed are measured by a power meter. The temperatures of the four cores and three disks in the testbed are presented in the rest of this section.

Table II Experiment Configuration

Experiments	**Utilization(%)**		**Power (W)**
	CPU	**Disk**	
1	0	0	73
2	100	0	135
3	0	100	85
4	100	100	142

1) Low CPU and Low Disk Utilization: in the first experiment, we place both CPU and disks in the idle mode. Fig. 1 shows that the node's inlet temperature varies slowly from 24.8 °C to 30.6 °C, which leads the outlet temperature to vary accordingly. When the outlet temperature goes up, the inlet temperature also increase due to heat recirculation. On average, the difference between the inlet and outlet temperatures is 3.8654 °C, ranging anywhere between 3.2 °C and 5.0 °C. In this case, the discrepancy between inlet and outlet temperatures can be expressed as a constant. Thus, we have:

$$T_{diff1}(t) = 3.8654 \quad (1)$$

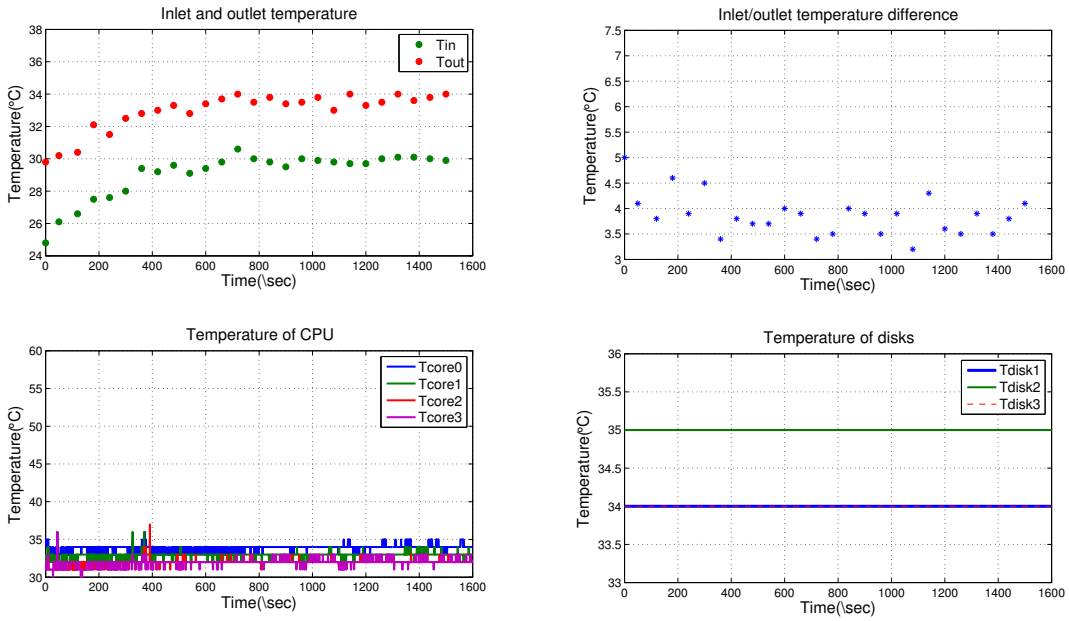


Figure 1: Temperature Evaluation under the Low CPU and Low Disk Utilizations.

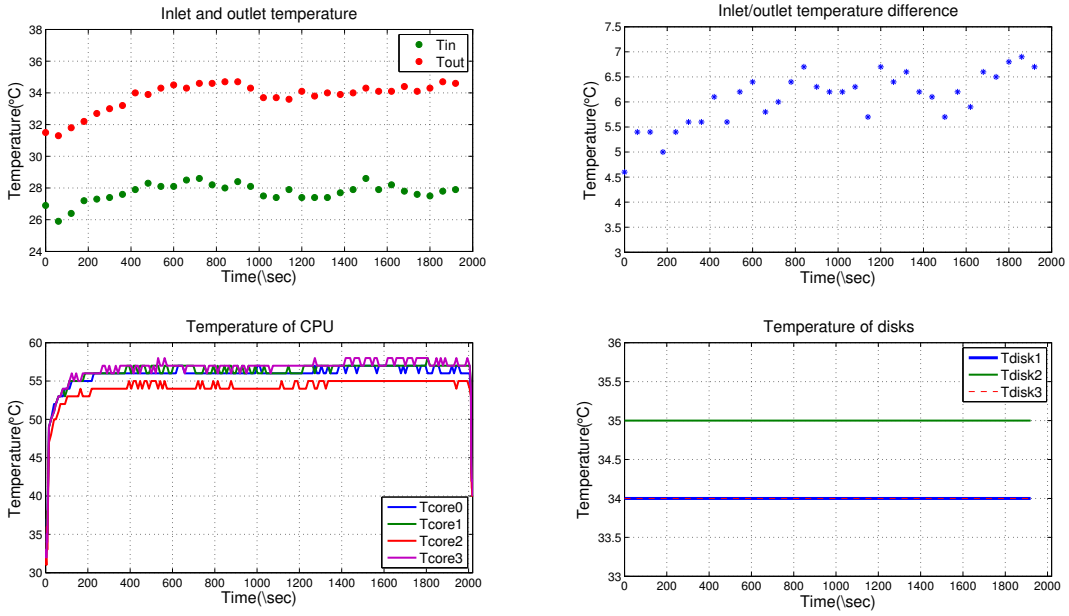


Figure 2: Temperature Evaluation under the High CPU and Low Disk Utilizations.

2) High CPU and Low Disk Utilizations: in the second experiment, we keep CPU extremely busy (i.e., CPU utilization approaches to 100%) while placing disks in the idle mode. Fig. 2 shows that the CPU temperature goes up fast; it increases 20 °C in 4 minutes. On the other hand, the disk temperatures do not change much. The difference between the inlet and outlet temperatures increases slowly from 4.6 °C to 6.6 °C in the first 600 seconds, and then maintain at a constant value in the next 1200 seconds. We denote inlet and outlet temperature difference as T_{diff2}, where t refers to the time at which the data node has run under 100% CPU and 0% disk utilizations. Thus, we have:

$$T_{diff2}(t) = \begin{cases} 0.0023 * t + 4.8818, & if\ t \leq 600 \\ 6.2692, & if\ t > 600 \end{cases} \quad (2)$$

3) Low CPU and High Disk Utilizations: in the third experiment, we keep a low CPU utilization while increasing disk utilization up to approximately 100%. We run three

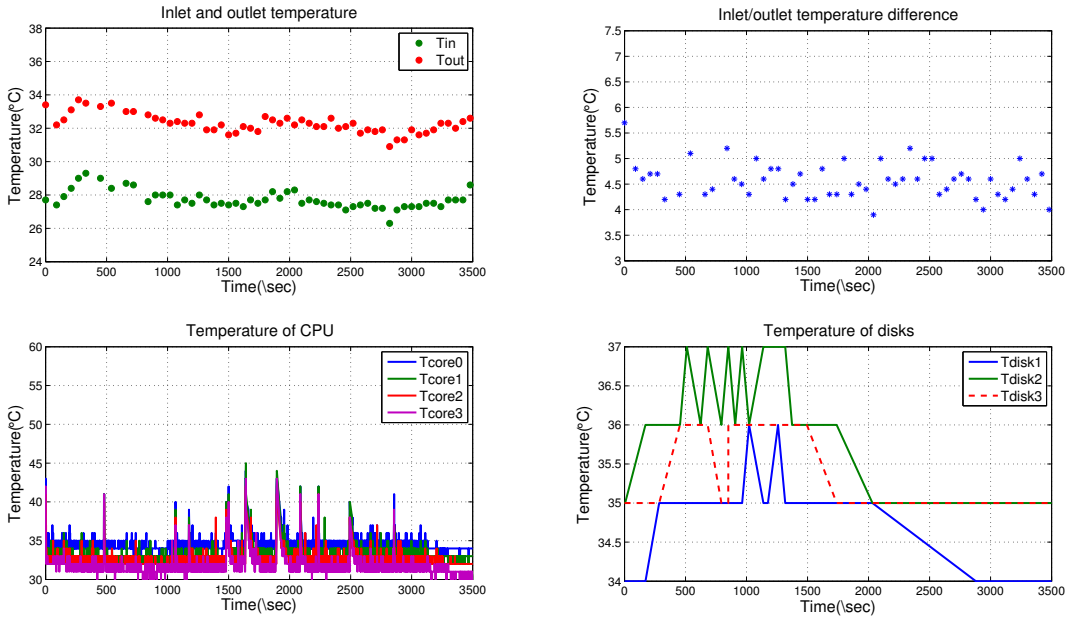


Figure 3: Temperature Evaluation under the Low CPU and High Disk Utilizations.

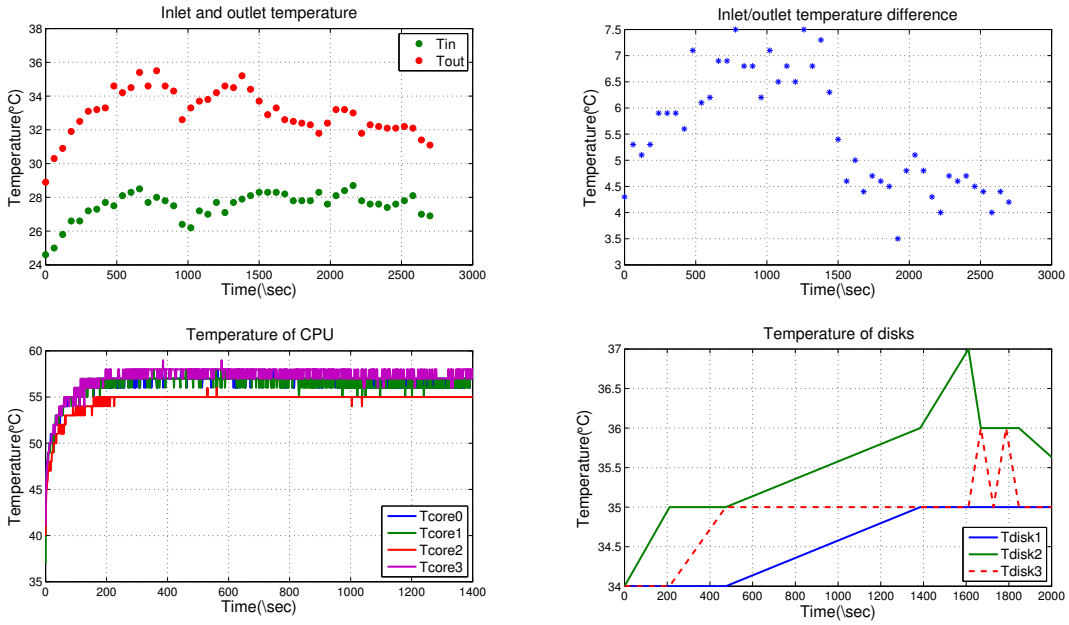


Figure 4: Temperature Evaluation under the High CPU and High Disk Utilizations.

tasks, each of which imposes I/O-intensive load on the disk. We observe from Fig. 3 that CPU temperature frequently fluctuates between 31 °C and 35 °C , because the three I/O-intensive tasks require the CPU resource to issues I/O requests. Nevertheless, the CPU utilization remains fairly low. After completing the tasks, CPU returns to the idle status and its temperature decreases to the normal value. In this case, the thermal impact of CPU is negligible. In contrast, disk temperatures slowly increase at the rate of around 2 °C per 1000 seconds. The difference between inlet and outlet temperature can be expressed by (3).

$$T_{diff3}(t) = \begin{cases} 0.0001 * t + 4.6086, & if\ t \leq 1000 \\ 4.7086, & if\ t > 1000 \end{cases} \quad (3)$$

4) High CPU and High Disk Utilization: in the final experiment, we push both CPU and disks utilizations up to 100%. We observe that the CPU temperature increases 20 °C at the beginning and goes back to the original

value after 1500 seconds when CPU-intensive tasks are completed. Therefore, we focus on the data collected before 1500 seconds. The inlet and outlet temperature difference falls in the range from 4.3 °C to 7.5 °C . In the first 660 seconds, the temperature difference increase very fast and then do not fluctuate much. Thus, we conclude from the experiment that CPU and disks significantly affect outlet temperatures, and the discrepancy between inlet and outlet temperature can expressed as (4).

$$T_{diff4}(t) = \begin{cases} 0.0014 * t + 5.3720, & if\ t \leq 660 \\ 6.8923, & if\ t > 660 \end{cases} \quad (4)$$

Fig. 4 also shows that the average cold-start time for the three disks is more than 1200 seconds, much larger than the cold-start time of CPU (i.e., CPU cold-start time is 100 seconds).

IV. Thermal Models

It is extremely challenging to model the energy consumption relationship between computing and cooling systems. The cooling cost depends not only on cooling setting (*e.g.*, inlet temperatures and cooling equipment placement), but also on heat dissipated by computing facilities. CPU and disks are two major types of components and heat contributors in data nodes. In this section, we develop a thermal model that aim to estimate outlet temperatures by considering the impacts of CPU and disks. Moreover, by combining a coefficient of performance (COP, for short) model that predicts cooling costs by CRAC supply temperature [17], our model can be used to predict the impact of CPUs and disks on cooling cost.

A. Framework

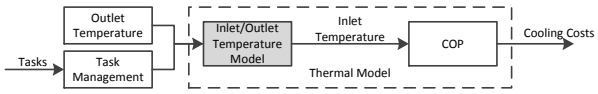


Figure 5: Framework of proposed solution

Fig. 5 displays our thermal-modeling framework, which consists of two components, namely, inlet/outlet-temperature model and COP. The inlet/outlet-temperature model builds up the relationship between inlet and outlet temperatures by profiling analysis. In addition, given an outlet temperature, our model estimates inlet temperatures under certain workloads. The COP model computes cooling costs by taking into account inlet temperatures offered by the inlet/outlet-temperature model. The main contributions of this framework are: (1) a thermal model that characterizes the relationship between inlet and outlet temperatures of a data node and (2) cooling cost estimation for data center designers.

B. An Inlet/Outlet Temperature Model

Considering CPU and disk utilizations, we classify workloads of a node into four basic types (i.e., see Section III-B for a combination of high and low utilizations of CPU and disks). During any time period, the workload of a node can be decomposed into a number of sub-period, in which the node runs under one of the four basic types. Thus, in each sub-period, the discrepancy between inlet and outlet temperatures is modeled by incorporating the four basis workload types.

$$T_{diff}(t) = \begin{cases} T_{diff1}(t), & if\ U_{CPU} = 0 \quad, U_{disk} = 0 \\ T_{diff2}(t), & if\ U_{CPU} = 100, U_{disk} = 0 \\ T_{diff3}(t), & if\ U_{CPU} = 0 \quad, U_{disk} = 100 \\ T_{diff4}(t), & if\ U_{CPU} = 100, U_{disk} = 100 \end{cases} \quad (5)$$

Given workloads and a number of sub-period $\mathfrak{T} = \{t_1, ...t_n\}$, we derive the outlet temperature from (1)-(4) as:

$$T_{diff}(\mathfrak{T}) = \frac{\sum_{i=1}^{n} T_{diff}(t_i)}{|\mathfrak{T}|} \quad (6)$$

C. The COP Model

The energy cost of a node is contributed by the energy consumption of the node and the cooling cost. We use COP (i.e., the Coefficient Of Performance model), described in [17], to calculate the cooling cost.

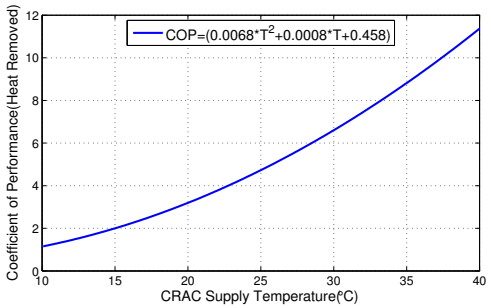


Figure 6: Coefficient of the performance curve for the chilled-water CRAC units at the HP Labs Utility Data Center [17]

Fig. 6 plots COP values that increase with the supply temperature of CRAC. A large COP value indicates a high energy efficiency.

$$COP(T) = 0.0068 * T^2 + 0.0008 * T + 0.458 \quad (7)$$

In 7, COP is defined as the ratio of heat removed to the energy cost of the cooling system for heat removal. T refers to the supply temperature of CRAC. The cooling cost is inversely proportional to the COP value.

$$P_{AC} = \frac{P_C}{COP(T)} \quad (8)$$

D. Case Studies

In order to demonstrate the application of our thermal model, we conduct three case studies, representing three typical access patterns of applications. We use the same testbed (see Section III) to perform the case studies. We keep all the three disks busy in the high-disk utilization cases. Let us consider the following access patterns (see Fig. 7) in our case studies:

- Pattern 1: In the $Computing\ After\ Reading$ pattern, applications first load data from disks, then process the loaded data using CPU resources.
- Pattern 2: In the $Computing\ Then\ Writing$ pattern, applications perform CPU-intensive computation first, followed by write-intensive activities to output data to disks.
- Pattern 3: In the $Computing\ and\ Reading/Writing\ in\ Parallel$ pattern, applications concurrently impose both CPU-intensive and I/O-intensive load to the node.

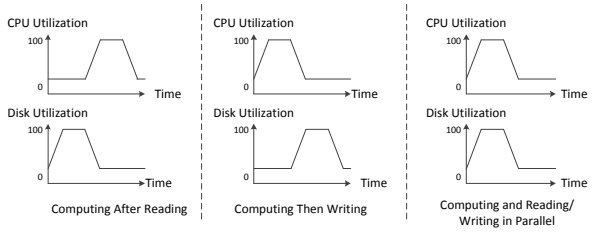


Figure 7: Three typical access patterns

Since the cold-start phase of disks is longer than that of CPU, we consider two scenarios in each case study. The first scenario represents cases where that the execution time of I/O tasks is smaller than the cold start phase. In this scenario, the cold-start issue significantly affects outlet temperatures. The second scenario represents case where the execution time of I/O tasks is much longer than the cold-start time. In the second scenario, the cold-start issue becomes negligible. In the case studies, P_C is the node's power consumption.

1) Impact of the Cold-Start Phase: We set the execution time of both CPU- and I/O-intensive tasks to 10 minutes, which is smaller than the cold start phase of disks. During the period of 10 minutes, the difference between inlet and outlet temperatures under the four basic workload types are:

$$T_{diff1}(600) = 3.8654\ (°\text{C})$$
$$T_{diff2}(600) = 6.2618\ (°\text{C})$$
$$T_{diff3}(600) = 4.6686\ (°\text{C})$$
$$T_{diff4}(600) = 6.2120\ (°\text{C})$$

After processing the CPU- and I/O-intensive tasks for 20 minutes in each case study, we evaluate the differences between inlet and outlet temperatures as follows.

Access Pattern 1. Disks are kept in the busy status in the first phase; $T_{diff3}(600)$ denotes the inlet/outlet-temperature difference. The increase of difference between inlet and outlet temperatures is $T_{diff3}(600)$-$T_{diff1}(0)$, which is 0.8032 °C. Since the cold-start time for disks are longer than 10 minutes, the disk temperature remains unchanged in the second phase. In this case, if the increase of the inlet/outlet-temperature difference in the first phase is considered as the increase in the inlet temperature for the second phase, and then this increment should be accumulated to the second phase. Therefore, the overall inlet/outlet-temperature difference can be derived as:

$$\begin{aligned} &T_{pattern1}(1200) \\ &= \frac{T_{diff3}(600) + T_{diff3}(600) - T_{diff1}(0) + T_{diff2}(600)}{2} \\ &= 5.8668\ (°\text{C}) \end{aligned}$$

Access Pattern 2. We obtain an average difference between inlet and outlet temperatures (i.e., 5.4652 °C) after running the test for 20 minutes. $T_{diff2}(600)$ is the temperature increment in the first phase, in which CPU is busy. Then, in the second phase, the CPU temperature falls down to the normal value in the first 10 seconds; the CPU temperatures in the second phase can be considered as a constant. The inlet/outlet temperature difference in the second phase can be calculated by $T_{diff3}(600)$. The average difference of inlet/outlet temperature is described below:

$$\begin{aligned} T_{pattern2}(1200) &= \frac{T_{diff2}(600) + T_{diff3}(600)}{2} \\ &= 5.4652\ (°\text{C}) \end{aligned}$$

Access Pattern 3. the inlet and outlet temperature difference increases from 3.8654 °C to 6.2120 °C in the first phase. In the second phase, the CPU temperature drops down quickly; whereas the disk temperature slowly decreases. The increasing and decreasing rates of disk temperature are slow; no difference is observed in a 10-minute period. Hence, we use T_{diff4} and T_{diff1} to calculate $T_{pattern3}(1200)$ as:

$$\begin{aligned} T_{pattern3}(1200) &= \frac{T_{diff4}(600) + T_{diff1}(600)}{2} \\ &= 5.0387\ (°\text{C}) \end{aligned}$$

Theoretically, cooling costs under these three patterns can be reflected by the inlet-outlet-temperature difference. To precisely evaluate cooling costs, we use the COP model that takes inlet temperatures as an input and produces cooling energy consumption. The inlet temperatures in the case studies are calculated in the way that identical outlet temperatures will be produced after the CPU- and I/O-intensive tasks are executed. For example, the inlet temperatures under the aforementioned access patterns are 24.1 °C, 24.5 °C and 25.0 °C with outlet temperature being 30 °C.

According to the COP model, the COP values of these access patterns are:

$$COP_{pattern1} = COP\ (24.1) = 4.4268$$
$$COP_{pattern2} = COP\ (24.5) = 4.5593$$
$$COP_{pattern3} = COP\ (25.0) = 4.728$$

Given power (see Table Table II) of the node, we derive the energy dissipation as:

$$P_{POWER1} = 135 * 600 + 85 * 600 = 132,000(J)$$
$$P_{POWER2} = 135 * 600 + 85 * 600 = 132,000(J)$$
$$P_{POWER3} = 142 * 600 + 73 * 600 = 129,000(J)$$

The cooling costs calculated by the COP model are:

$$P_{AC1} = \frac{P_{POWER1}}{COP_{pattern1}} = 29,818(J)$$
$$P_{AC2} = \frac{P_{POWER2}}{COP_{pattern2}} = 28,952(J)$$
$$P_{AC3} = \frac{P_{POWER3}}{COP_{pattern3}} = 27,284(J)$$

From the above analysis, access patten 3 saves the cooling cost of patterns 1 and 2 by 2,534 J and 1,668 J, respectively. The total energy cost, including computing and cooling energy consumption, are shown below:

$$P_{TOTAL1} = P_{POWER1} + P_{AC1} = 161,818(J)$$
$$P_{TOTAL2} = P_{POWER2} + P_{AC2} = 160,952(J)$$
$$P_{TOTAL3} = P_{POWER3} + P_{AC3} = 156,284(J)$$

We observe that access pattern 3 leads to the lowest energy. Pattern 3 makes it possible to increase CRAC temperature to lower cooling cost. This observation motivates us to propose an thermal-aware workload management that minimizes the total energy consumption by data placement optimization (see Section. V).

To validate the accuracy of the model, we manually measure the inlet and outlet temperatures of the node by using an infrared thermometer. We collect 20 temperature samples in each case study. We compare inlet-outlet-temperature differences obtained from our model against the real-world measurement. Table III shows that the precision-errors of our model for the three case studies are 2.28 %, 3.74%, and 4.84%, respectively. The precision is calculated by dividing an average difference between real measurement and simulation results by real measurement.

Table III Thermal Model Validation

	Case Study 1	Case Study 2	Case Study 3
Precision Error (%)	2.28	3.74	4.84

2) Negligible Cold-Start Phase is Insignificant: if the execution time of CPU- and I/O-intensive tasks are sufficiently long, impact of the cold-start phase becomes negligible. Now, we extend the model to consider cases where the cold-start phase can be ignored. We set the execution time of the tasks to be 60 minutes (totally 120 minutes), T_{diff} of the basic workload types are given below:

$$T_{diff1}(3600) = 3.8654(°C)$$
$$T_{diff2}(3600) = 6.2692(°C)$$
$$T_{diff3}(3600) = 4.7086(°C)$$
$$T_{diff4}(3600) = 6.8923(°C)$$

The average inlet-outlet-temperature differences under the three access patterns are:

$$T_{pattern1}(7200) = 5.4889(°C)$$
$$T_{pattern2}(7200) = 5.4889(°C)$$
$$T_{pattern3}(7200) = 5.3789(°C)$$

We can obtain the total energy costs of these cases as:

$$P_{TOTAL1} = 1,610,000(J)$$
$$P_{TOTAL2} = 1,610,000(J)$$
$$P_{TOTAL3} = 1,570,000(J)$$

The results show that compared with patterns 1 and 2, pattern 3 offer 40,000 J savings in energy.

V. Data Placement Strategies

A. Thermal Impacts of Data Placement

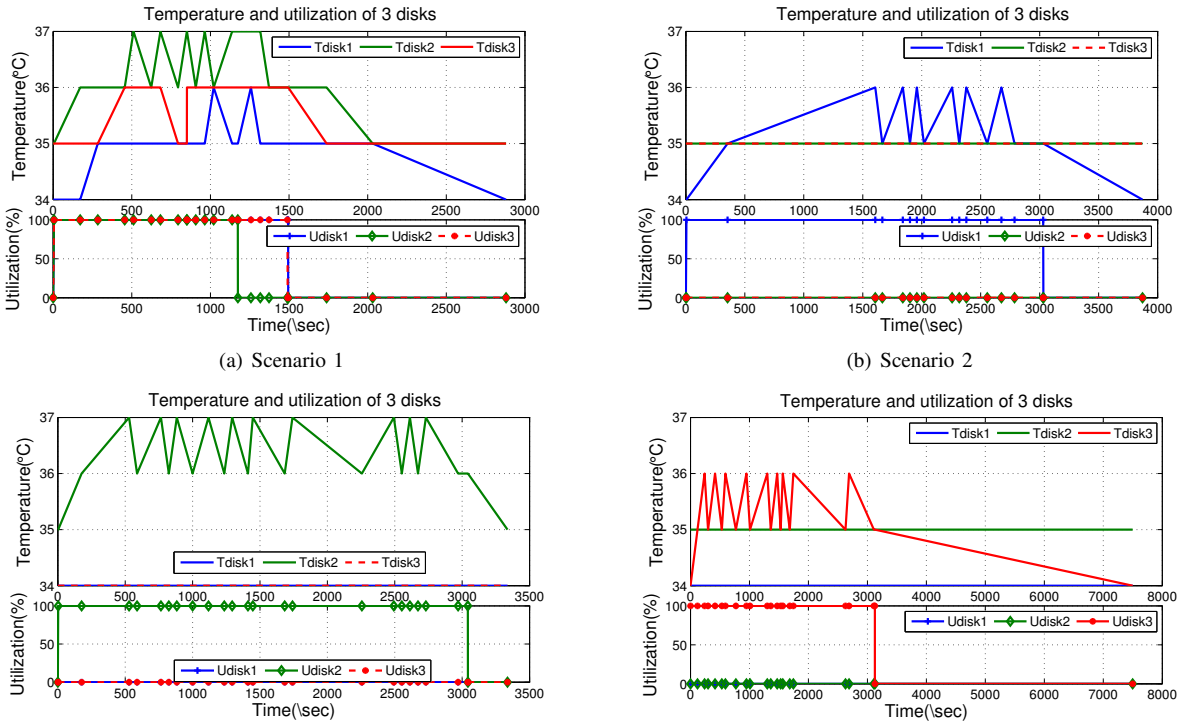
In Section III, we evaluate the thermal impacts of a data node equipped with three disks. It is worth noting that disk configurations may vary greatly among nodes. More importantly, our results show that workloads and disk configurations affect heat dissipation in disks.

In this section, we show that given CPU and I/O loads, workload distribution significantly affects thermal performance of data nodes. In this data placement study, we use the the same testbed described in Section III. We use postmark to initially create 100 files, the size of which ranges from 1 to 100 MBytes. Three postmark tasks issue 1,000 requests to the disks. Four scenarios are investigated in this group of experiments. In the first scenarios, the three tasks are accessing the three disks respectively. In the other three scenarios, the three tasks are sharing a single disk.

Fig. 8 plots the disk utilization and temperature of the four scenarios examined in the three-disk case. In scenario 1, it takes 1,500 seconds to complete all the I/O requests. Fig. 8(a) shows that the temperatures of disk 1 and 2 increase by 2 °C; and the temperature of disk 3 increases by 1 °C. When the three tasks are sharing one disk, the disk temperature increases by 1 °C , whereas temperatures of the other two disks remain unchanged. Thus, we conclude that sharing a disk among the three tasks can maintain low disk temperatures at the cost of increased I/O processing time (e.g., from 1500 to 3,000 seconds).

B. Thermal-Aware Data Placement

The experimental results shown in the previous subsection indicate that outlet temperatures affected by disks vary greatly among cases. Nevertheless, a number of scenarios have not been evaluated. For example, one possible scenario



(a) Scenario 1

(b) Scenario 2

(c) Scenario 3

(d) Scenario 4

Figure 8: Thermal Impacts of Data Placement in the Three-Disk Case.

might be the load of three disks are high, low, and idle, respectively. To provide large storage capacity, one may increase the number of disks in each data node. Manually measuring all possible scenarios is time-consuming and impractical. A promising solution is to use real measurements collected in the simple disk configurations, and to model the thermal characteristics of complicated scenarios.

Our results suggest that disk temperatures significantly affect the outlet temperatures of a node. Disk temperatures in turn depends on I/O activities and disk configurations. These observations motivate us to study thermal-aware data placement strategies, which aim to migrate data among disks in order to minimize the cooling costs.

VI. Conclusion

Energy efficiency and thermal management of storage systems must be urgently addressed, because energy consumption and cooling costs of large-scale storage systems in data centers have been increasing in the past decade. Recent studies show that cooling costs contribute a significant portion of the operational cost of data centers. Thermal management techniques are applied to reduce the energy consumption in cooling systems, thereby significantly improving the energy efficiency of data centers. Thermal models play a key role in thermal management; however, traditional thermal models for data centers do not take into account disk utilizations. In this study, we develop a thermal model to investigate thermal impacts of hard disks on storage systems. We show how to apply the thermal model to estimate the outlet temperature of a storage server based on processor and disk utilizations. In addition, we study the impact of data placement on the temperature of disks which affect the outlet temperature and cooling cost.

Our thermal model offers the following two benefits. First, the model makes it possible to reduce thermal monitoring cost. Thermal management of hard disks in storage systems helps to cut cooling cost and boost system reliability. Monitoring temperatures is a key issue in thermal management techniques; however, it is prohibitively expensive to acquire and set up a huge number of sensors in a large-scale data center. Our model is an alternative to monitoring temperatures of storage systems. Second, our thermal model enables data center designers to make intelligent decisions on thermal management during the design phase.

Acknowledgment

This research was supported by the U.S. National Science Foundation under Grants CCF-0845257 (CAREER), CNS-0917137 (CSR), CNS-0757778 (CSR), CCF-0742187 (CPA), CNS-0831502 (CyberTrust), CNS-0855251 (CRI), OCI-0753305 (CI-TEAM), DUE-0837341 (CCLI), and DUE-0830831 (SFS). Mohammed Alghamdi's research was supported by AL-Baha University.

References

[1] hddtemp. http://manpages.ubuntu.com/manpages/natty/man8/hddtemp.8.html.

[2] lm-sensors. http://www.lm-sensors.org/.

[3] stress. http://www.unixref.com/manPages/stress.html.

[4] Wd1600aajs specification. http://www.wdc.com/wdproducts/library/SpecSheet/ENG/2879-701277.pdf.

[5] U.S. Environmental Protection Agency. Report to congress on server and data center energy efficiency. Technical report, August 2007.

[6] Miriam Allalouf, Yuriy Arbitman, Michael Factor, Ronen I. Kat, Kalman Meth, and Dalit Naor. Storage modeling for power estimation. In *Proceedings of SYSTOR 2009: The Israeli Experimental Systems Conference*, SYSTOR '09, pages 3:1–3:10, New York, NY, USA, 2009.

[7] Luiz André Barroso and Urs Hölzle. The case for energy-proportional computing. *Computer*, 40(12):33–37, December 2007.

[8] David J. Brown and Charles Reams. Toward energy-efficient computing. *Commun. ACM*, 53(3):50–58, March 2010.

[9] Steve Greenberg, Evan Mills, Bill Tschudi, Peter Rumsey, and Bruce Myatt. Best Practices for Data Centers: Lessons Learned from Benchmarking 22 Data Centers. 2006.

[10] Sudhanva Gurumurthi, Anand Sivasubramaniam, and Vivek K. Natarajan. Disk drive roadmap from the thermal perspective: A case for dynamic thermal management. *SIGARCH Comput. Archit. News*, 33(2):38–49, May 2005.

[11] http://www.datacenterdynamics.com/research/energy-demand 2011-12. Global data center energy demand forecasting. Technical report, institution, 2011.

[12] A. Hylick, R. Sohan, A. Rice, and B. Jones. Energy efficient data center. In *it - Information Technology*, volume 54, pages 17–23, 2012.

[13] Jeffrey Katcher. Postmark: A new file system benchmark. *System*, (3022):1–8, 1997.

[14] Youngjae Kim, S. Gurumurthi, and A. Sivasubramaniam. Understanding the performance-temperature interactions in disk i/o of server workloads. In *High-Performance Computer Architecture, 2006. The Twelfth International Symposium on*, pages 176 –186, feb. 2006.

[15] Jonathan G. Koomey. Estimating total power consumption by servers in the U.S. and the world. Technical report, Lawrence Derkley National Laboratory, February 2007.

[16] Lei Li, Chieh-Jan Mike Liang, Jie Liu, Suman Nath, Andreas Terzis, and Christos Faloutsos. Thermocast: a cyber-physical forecasting model for datacenters. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '11, pages 1370–1378, New York, NY, USA, 2011. ACM.

[17] Justin Moore, Jeff Chase, Parthasarathy Ranganathan, and Ratnesh Sharma. Making scheduling "cool": temperature-aware workload placement in data centers. In *Proceedings of the annual conference on USENIX Annual Technical Conference*, ATEC '05, pages 5–5, Berkeley, CA, USA, 2005. USENIX Association.

[18] Andrew Pavlo, Erik Paulson, Alexander Rasin, Daniel J. Abadi, David J. DeWitt, Samuel Madden, and Michael Stonebraker. A comparison of approaches to large-scale data analysis. In *Proceedings of the 2009 ACM SIGMOD International Conference on Management of data*, SIGMOD '09, pages 165–178, New York, NY, USA, 2009. ACM.

[19] Eduardo Pinheiro, Wolf-Dietrich Weber, and Luiz André Barroso. Failure trends in a large disk drive population. In *Proceedings of the 5th USENIX conference on File and Storage Technologies*, pages 2–2, Berkeley, CA, USA, 2007. USENIX Association.

[20] O. Sarood, A. Gupta, and L.V. Kale. Temperature aware load balancing for parallel applications: Preliminary work. In *Parallel and Distributed Processing Workshops and Phd Forum (IPDPSW), 2011 IEEE International Symposium on*, pages 796 –803, may 2011.

[21] Osman Sarood and Laxmikant V. Kale. A 'cool' load balancer for parallel applications. In *Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis*, SC '11, pages 21:1–21:11, New York, NY, USA, 2011. ACM.

[22] C.P.H. Tan, J.P. Yang, J.Q. Mou, and E.H. Ong. Three dimensional finite element model for transient temperature prediction in hard disk drive. In *Magnetic Recording Conference, 2009. APMRC '09. Asia-Pacific*, pages 1–2, jan. 2009.

[23] Q. Tang, S. Gupta, and G. Varsamopoulos. Thermal-aware task scheduling for data centers through minimizing heat recirculation. In *Cluster Computing, 2007 IEEE International Conference on*, pages 129 –138, sept. 2007.

[24] Q. Tang, S. Gupta, and G. Varsamopoulos. Thermal-aware task scheduling for data centers through minimizing heat recirculation. In *Cluster Computing, 2007 IEEE International Conference on*, pages 129 –138, sept. 2007.

[25] Qinghui Tang, Sandeep Kumar S. Gupta, and Georgios Varsamopoulos. Energy-efficient thermal-aware task scheduling for homogeneous high-performance computing data centers: A cyber-physical approach. *IEEE Trans. Parallel Distrib. Syst.*, 19(11):1458–1472, November 2008.

[26] Nedeljko Vasic, Thomas Scherer, and Wolfgang Schott. Thermal-aware workload scheduling for energy efficient data centers. In *Proceedings of the 7th international conference on Autonomic computing*, ICAC '10, pages 169–174, New York, NY, USA, 2010. ACM.

[27] Akshat Verma, Ricardo Koller, Luis Useche, and Raju Rangaswami. Srcmap: energy proportional storage using dynamic consolidation. In *Proceedings of the 8th USENIX conference on File and storage technologies*, FAST'10, pages 20–20, Berkeley, CA, USA, 2010. USENIX Association.

[28] Tao Xie and Yao Sun. Understanding the relationship between energy conservation and reliability in parallel disk arrays. *J. Parallel Distrib. Comput.*, 71:198–210, February 2011.