J Sign Process Syst (2013) 72:181–196
DOI 10.1007/s11265-013-0787-6

Thermal Modeling of Hybrid Storage Clusters

Xunfei Jiang · Maen M. Al Assaf · Ji Zhang · Mohammed I. Alghamdi · Xiaojun Ruan · Tausif Muzaffar · Xiao Qin

Received: 2 November 2012 / Revised: 2 May 2013 / Accepted: 29 May 2013 / Published online: 28 June 2013
© Springer Science+Business Media New York 2013

Abstract There is a lack of thermal models for storage clusters; most existing thermal models do not take into account the utilization of hard drives (HDDs) and solid state disks (SSDs). To address this problem, we build a thermal model for hybrid storage clusters that are comprised of HDDs and SSDs. We start this study by generating the thermal profiles of hard drives and solid state disks. The profiling results show that both HDDs and SSDs have profound impacts on temperatures of storage nodes in a cluster. Next, we build two types of hybrid storage clusters, namely, inter-node and intra-node hybrid storage clusters. We develop a model to estimate the cooling cost of a storage cluster equipped with hybrid storage nodes. The thermal model is validated against data acquired by temperature sensors. Experimental results show that, compared to the HDD-first strategy, the SSD-first strategy is an efficient approach to minimize negative thermal impacts of hybrid storage clusters.

Keywords Thermal · Model · Hybrid · Storage · Cluster

X. Jiang (✉) · J. Zhang · T. Muzaffar · X. Qin
Department of Computer Science and Software Engineering, Auburn University, Auburn, AL 36849-5347, USA
e-mail: xzj0009@auburn.edu

J. Zhang
e-mail: jzz0014@auburn.edu

T. Muzaffar
e-mail: tausifm@auburn.edu

X. Qin
e-mail: xqin@auburn.edu

M. M. Al Assaf
King Abdullah II School for Information Technology, The University of Jordan, Amman, Jordan
e-mail: m_alassaf@ju.edu.jo

M. I. Alghamdi
Department of Computer Science, Al-Baha University, Al-Baha City, Kingdom of Saudi Arabia
e-mail: mialmushilah@bu.edu.sa

X. Ruan
Department of Computer Science, West Chester University of Pennsylvania, West Chester, PA 19383, USA
e-mail: xruan@wcupa.edu

1 Introduction

The cooling cost of a data center is significantly impacted by thermal management strategies; traditional thermal models developed for thermal management mechanisms do not take into account the utilization of hard drives (HDDs) and solid state disks (SSDs). In this study, we address the thermal impacts of disks on hybrid storage clusters containing both HDDs and SSDs. We investigate two types of hybrid storage clusters, namely, inter-node and intra-node hybrid storage clusters. We show how to apply our new thermal model to predict the cooling cost of the two types of hybrid storage clusters.

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

1. the ever-increasing cooling and energy costs of large-scale storage clusters,
2. the impact of hybrid storage on thermal management of data centers,
3. the importance of reducing thermal monitoring cost,

4. the capability of estimating the cooling cost of a data center, and
5. the lack of study on the impacts of hard drives and solid state disks on outlet temperatures of storage nodes in a cluster.

With the increase of energy consumption and cooling costs of large-scale storage clusters, there is an urgent need for data center designers to address the energy efficiency issues [20]. Evidence shows that the energy cost of a data center for four years can exceed the cost of building a new data center. Conventional approaches of saving energy cost for data centers include improving the energy efficiency of computing facilities as well as cooling systems.

Cooling costs contribute a large portion of the total energy cost of data centers [6, 20]. For instance, the power and cooling cost to support the IT equipments take more than half of the total energy cost of a data center [6]. Previous studies demonstrate that energy efficiency could be enhanced by reducing the energy dissipation in cooling systems [27, 39]. Reducing outlet temperatures or optimizing air recirculation can improve energy efficiency [36]. Moreover, many load balancing strategies were proposed to gain good temperature distribution [27, 39]. Recent studies show that reducing outlet temperatures of servers in a data center could save up to 40 % energy consumption [27]. Lowering outlet temperatures of storage nodes not only conserves cooling cost, but it also improves the reliability and lifetime of disks [30, 41].

A handful of studies have focused on modelling the energy consumption of storage clusters in the past years. For example, an energy model is introduced to estimate the power consumption of storage nodes running under specific workloads [7]. Unfortunately, thermal models of storage clusters are still in their infancy. Little attention has been paid to the thermal impact of disks, including HDDs and SSDs, on the energy efficiency of cooling systems in data centers.

Deploying temperature sensors on storage nodes of a cluster is a usual method to monitor the storage cluster's temperature. For each data node, one needs to apply at least two sensors to obtain the inlet and outlet temperatures. If temperatures of interior devices of the node need to be monitored, additional sensors must be set up. Although this traditional approach is practical for measuring temperatures of small-scale storage clusters, it becomes a sophisticated solution when a storage cluster has thousands of nodes. It is extremely expensive to set up a huge number of sensors in a large-scale storage cluster; deploying sensors also leads to extra energy cost. Thermal models are a promising alternative to obtain temperatures of storage clusters.

Building a data center is a huge investment for enterprises. Estimating the energy costs, which include cooling cost and power cost, offers an important guideline in the designing phase. Simulations and thermal models help data center designers make good decisions on thermal management during the planning phase.

A variety of factors impact the outlet temperatures of storage nodes. A study shows that inlet temperatures and CPU utilization affect the outlet temperatures of data nodes [36]. In a second study, a temperature model was proposed using a data node's historical temperature data and airflow of a data center [25]. When it comes to the thermal behavior of disks, Kim et al. investigate the relationship between disk seek time, inter-seek time, and disk temperatures [23]. They also observed that the number and size of flatters in a disk affect its temperatures. In enterprise-level Tera-data centers, a single node is capable of supporting more than 100 disks [28]. The temperature of these large number of disks within a data node plays a crucial role in impacting the outlet temperatures of the node. However, there is the lack of studies on the impact of hard drives and solid state disks on outlet temperatures of storage nodes in a cluster.

Contributions We make the following three contributions. First, we generate thermal profiles of HDDs and SSDs. The profiling results are obtained by running CPU-intensive workloads imposed by stress [4] and I/O-intensive workloads imposed by Postmark [22]. When a storage node is running under various load conditions, we monitor the node's CPU temperatures and disk temperatures as well as the inlet and outlet temperatures. Then, we build a thermal model to estimate inlet/outlet temperature differences using inlet temperatures, CPU temperatures, and disk temperatures. Second, we develop a model that can be used in combination with the coefficient of performance model (or COP for short) to derive the cooling cost of hybrid storage clusters from supply temperatures of computer room air conditioning units or CRAC. Third, to demonstrate the usage of the model, we make use of this model to investigate the impact of HDDs and SSDs on the cooling cost of storage clusters.

Organization The rest of this paper is organized as follows. The next section presents prior studies and related research issues. Section 3 describes the thermal impacts of hard drive disks and solid state disks on storage nodes. In Section 4, we develop a model used to estimate the cooling cost of hybrid storage clusters. In Section 5, we discuss the impacts of HDD-first strategy and SSD-first strategy on cooling costs of two hybrid storage clusters. Finally, Section 6 concludes the paper.

2 Relative Work

2.1 Energy-Efficient Data Centers

The rapid increase of energy consumption has brought great attention to energy efficiency of data centers [11, 12]. Koomey's study conducted in 2000 shows that the total energy cost of data centers is approximately 1.2 % of the energy consumption in the U.S. [24]. Various energy-efficient techniques have been proposed to lower the energy consumption of data centers. For instance, Guerra et al.'s preliminary analysis on energy saving techniques indicates that significant energy reduction in data centers can potentially be achieved by appropriately choosing energy saving techniques for storage systems [18]. A measurement and management technology (MMT) is applied to build energy-efficient data centers [8]. This MMT model incorporates the real temperature data measured by temperature sensors, thereby providing a run-time analysis of energy consumption. With the analytical data, data centers can be operated in an optimal schedule in terms of energy consumption.

Some previous studies focused on reducing the power consumption of storage nodes to save energy. For example, Colarelli and Grunwald proposed an energy-efficient storage called MAID by setting a subset of disks as cache disks. MAID keeps cache disks active while allowing other disks to stay in the low-power state for a long time period [15]. Pinheiro and Bianchini developed the popular data concentration technique or PDC that migrates frequently accessed data to a subset of disks, which are kept in the active state [29]. In doing so, PDC can place a large number of disks in the low-power mode to save energy.

2.2 Thermal-Aware Resource Management Strategies

Much attention has been paid to managing workloads according to thermal distribution. A few thermal-aware resource management strategies were proposed to balance workload based on temperature distributions in a data center. For example, a thermal-load-balancing framework was developed to reduce energy consumption and improve equipment reliability [34]. In this framework, local and regional policies are used to dynamically distribute the workload in order to gain a uniform temperature distribution in data centers. Moore and Chase proposed an approach to controlling heat generation in data nodes through temperature-aware workload placement [27]. Vasic et al. introduced a thermal-dynamic model and a temperature control strategy coupled with the model [39]. The strategy combines the air flow control and the thermal-aware scheduling. Simulations with synthetic and real workload traces show that the strategy delievers good performance in keeping temperatures of a data center under a threshold. Air recirculation can be minimized to decrease the energy cost of data centers. A study shows that balancing the workload within a data center to minimize the air recirculation could gain the most significant energy saving compared with other existing strategies [37].

Inlet temperatures of data nodes have significant impacts on the cooling cost of a data center. Tang et al. proposed MPIT-TA, a task assignment policy, to minimize peak inlet temperatures in order to reduce cooling energy consumption [38]. The simulation results show that their policy could save cooling costs by at least 20 %.

Temperature-aware workload balancing strategies are mostly focused on computing resources and CPU utilization [31, 32]. It has become a traditional wisdom to save energy costs by keeping CPU temperatures under a certain threshold through the dynamic CPU voltage/frequency scaling technique.

Ayoub et al. proposed a thermal management strategy in the memory subsystem to save energy by allocating workloads to a few memory units and powering down the rest of the memory [9]. This strategy gains an improvement of energy savings by 43 % and reduces performance overhead by 85 %.

El-Sayed et al. collected a large amount of thermal data from several data centers to study the impacts of temperature on hardware reliability, and they provided recommendations for temperature management which can save energy and limit the negative impacts on system reliability and performance at the same time [17].

Thermal-aware resource management techniques for processors and memories have been widely studied. However, the impact of disks, including HDD and SSD, on thermal management has not been fully explored. For large-scale data centers, each data node may contain a large number (e.g., up to 100) of disks [28]. And disk utilizations would be extremely high when I/O intensive tasks are handled by these disks. Hence, appropriately managing I/O workload can potentially reduce the cooling cost in data centers.

2.3 Disk Energy Consumption and Temperature Models

Eibeck and Cohen proposed a thermal model to predict transient temperatures of IBM 5-1/4-in fixed disk drives [16]. Tan et al. built a three-dimensional model to evaluate transient temperatures during frequent seeking [35]. 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 [19]. The model calculates the heat generated from the following components: internal drive air, spindle motor, the

base and cover of a disk, the voice-coil motor, and disk arms. Kim et al. inferred the relationship between seek times and disk temperatures [23]. 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. Another research studied the thermal behaviour of disks and CPUs, and generate an outlet temperature model under a certain workload [21].

In this paper, we comprehensively evaluate the impact of CPU and disk temperatures on the inlet and outlet temperatures of a data node.

2.4 Solid State Disk (SSD)

Solid state disks or SSDs are an emerging storage technology, thanks to SSDs' high I/O performance and energy efficiency. There is a good potential to widely apply SSDs in large-scale cluster storage systems. SSDs are more expensive than traditional hard drives. To improve both performance and energy efficiency, one may employ hybrid SSD devices to build large storage systems. Recently, Chang proposed an SSD-based hybrid storage system that combines MLC flash-based and SLC flash-based SSDs [13]. The experimental results show that compared with MLC-flash-based SSD storage, the hybrid system can gain significant improvements in terms of throughput and energy savings.

Apart from hybrid SSDs, hybrid storage systems that combine HDDs and SSDs have been proposed to make a good tradeoff between performance and cost. For example, Chen et al. designed a hybrid storage system – Hystor – in which hot data is stored in SSDs to optimize system I/O performance [14]. All data accesses are periodically recorded and analyzed by a *monitor* module. When any data becomes hot, it will be moved to an SSD to reduce data access time. Wu et al. developed a hybrid page/block architecture along with an advanced replacement policy called BPAC to exploit both temporal and spatial locality [40]. Mao et al. proposed a hybrid parity-based disk array architecture (HPDA), where SSDs and HDDs are integrated in a RAID system to improve the performance and reliability of the RAID [26]. Balakrishnan et al. proposed Diff-RAID, a parity-based redundancy solution that unevenly distributes and balances the parity across SSDs to improve the reliability of storage systems [10]. Schall et al. investigated the performance and energy efficiency of SSDs and HDDs in I/O-intensive database applications [33]. Although hybrid storage systems can offer good performance and reliability, less attention has been paid to the thermal characteristics of hybrid storage devices that have significant impacts on the energy costs of cooling systems in future data centers.

3 Thermal Impacts of Disk I/O

To characterize the thermal behavior of nodes in a cluster storage system, we start this study by investigating CPU and disk temperatures as well as inlet/outlet temperatures of a storage node. In this section, we first describe the testbed used in our experiments. Next, we show noticeable impacts of both CPU and disk temperatures on the outlet temperatures of a storage node. Finally, we conduct experiments to explore the thermal characteristics of HDDs and SSDs in a storage node.

3.1 Testbed

We use a Linux server as a storage node in our testbed. The storage node is equipped with an Intel(R) Celeron(R) 2.2 GHz processor, 1.0 GB main memory, and a 160 GB SATA disk. The configuration parameters of the node are summarized in Table 1.

The inlet and outlet temperatures are monitored by temperature sensors, and the data is collected by MiniGoose II [3]; for disk temperature, we collect temperature data from both the inner-disk sensor and temperature sensor applied outside of the disk. The CPU temperature is examined by the use of lm-sensor [2]. What's more, a temperature sensor is applied at the outlet of the air conditioner to monitor the supply temperature.

In the following experiments, the air-conditioner temperature is set to 23.2 °C. And we use T_{diff} to represent the discrepancy between inlet and outlet temperature.

3.2 Impacts of CPU/Disk Temperatures on Outlet Temperatures

Outlet temperatures of storage nodes in a cluster storage are affected by various factors, including CPU and disk temperatures, motherboard temperatures, and inlet temperatures. Although impacts of CPU temperatures on computing nodes has been explored in prior studies [31, 32, 38], and the thermal impact of memory has also been studied [9], thermal impacts of disks on storage nodes are an open issue. Therefore, we conduct the following experiments to characterize the impacts of disks and processors on the outlet temperatures of nodes in a storage cluster.

Table 1 Testbed configurations.

Hardware	Software
1× Intel Celeron CPU 2.2 GHz	Ubuntu 10.04
1× 1.0 GBytes of RAM	Linux kernel 2.6.32
1× WD 160 GBytes Sata disk (WD1600AAJS-75M0A0 [5])	lm-sensors

3.2.1 Impacts of CPU Temperatures

In the first group of tests, we measure the inlet and outlet temperatures of an idle node by placing disks outside of a chassis and keeping the processor in the idle state for a hour. In this test case, we observe that the CPU temperature remains around 39 °C and the difference between inlet and outlet temperatures (T_{diff}) is 1.74 °C in average. And we also observe that, the supply temperature is 8 °C lower than the inlet temperature. This test condition is regarded as a baseline because the outlet temperature is independent of the disk placed outside the chassis.

In the second group of experiments, we perform stress tests on processors by keeping CPU utilization very high and placing the disk outside its chassis. This usage condition represents data-intensive computing environments. Results plotted in Fig. 1 show that the CPU temperature changes from 39 °C to 57 °C during the stress tests. It also illustrates that when CPU arrives at its peak temperature 57 °C, the discrepancy between the inlet and outlet temperatures (T_{diff}) is 3.38 °C, which is 1.64 °C (or 94 %) higher than the one observed in the baseline case. In this case, the CPU temperature is considered as a driving force behind the increment in the outlet temperature of the tested storage node.

3.2.2 Impacts of Disk Temperatures

In the third group of experiments, we move the disk from outside of its chassis to inside, and we observe that the disk temperature reduces. This is because the fan inside the chassis cools down the disk. A comparison of disk temperatures outside of the chassis and in the chassis are shown in Fig. 2. When the disk is placed outside of chassis, the disk temperature is around 30 °C detected by the outer disk sensor, and 39 °C by the inner-disk sensor. After moving the disk inside its chassis, the disk temperature is 27 °C detected by the outer disk sensor and 33 °C by the inner-disk sensor. From this, we can see a big difference between the temperature

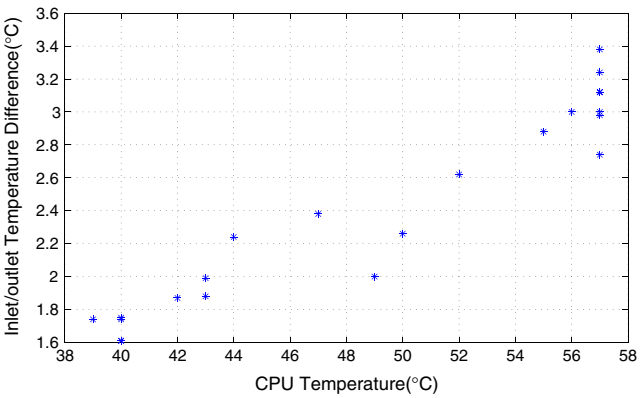


Figure 1 The impact of CPU temperature on outlet temperature.

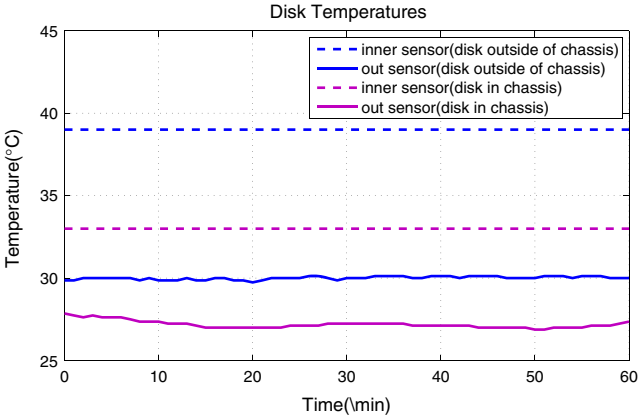


Figure 2 The comparison of disk temperature get by inner-disk sensor and the outside sensor.

inside the disk and outside of the disk. Since most SSDs do not have an inner sensor, in the rest of this paper, we use the outside sensor to detect disk temperature. The discrepancy of inlet and outlet temperatures monitored by outside sensors when CPU and the disk are both in idle state is 2.01 °C on average, which is 0.27 °C higher than while the disk is placed outside of its chassis. Since CPU temperature is 40 °C when it is idle, we can conclude that this 0.27 °C increment comes from the disk.

In the fourth group of experiments, we run CPU-intensive applications when the disk is inside its chassis. Figure 3 plots the difference between the inlet and outlet temperatures under various CPU temperatures. We observe that when CPU is 59 °C, T_{Diff} is 3.63 °C.

In the fifth group of experiments, we keep the disk inside its chassis and run I/O-intensive tasks on the storage node. This test case represents usage conditions of cluster storage systems supporting data-intensive applications. Since the processor is heavily loaded, the CPU temperature does not dramatically change during the course of task executions. Figure 4 shows that the increasing disk temperature

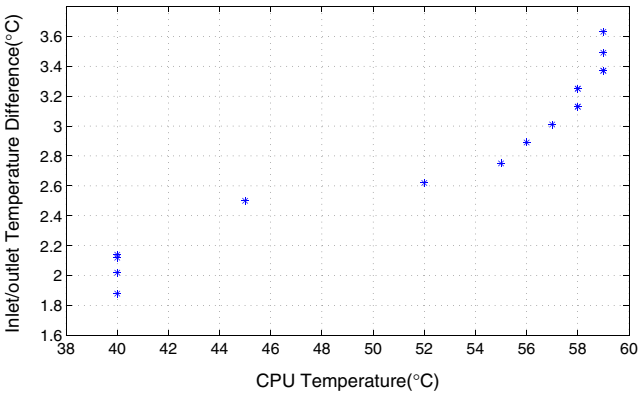


Figure 3 The impact of CPU temperature on outlet temperature while disk in chassis.

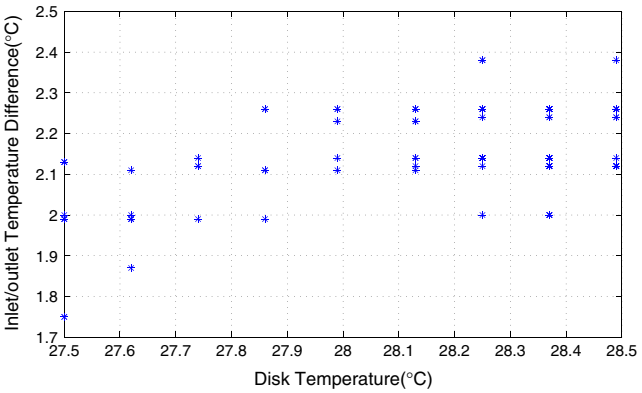


Figure 4 Impact of disk temperature on outlet temperature.

gives rise to the increment in the outlet temperature of the node. For example, when the disk reaches its peak temperature of 28.5 °C, the outlet temperature is increased by approximately 0.25 °C. Compared with the processor temperature, the disk temperature changes mildly when the disk is transitioning from the idle to active state.

From the preliminary results, we observe that one disk may have a slight impact on the outlet temperature; however, if multiple disks are applied in a data node, the impact may be different. Thus we conduct another group of experiments to study the impact of disk temperature on outlet temperature. The testbed used in these experiments is equipped with 4 Intel(R) Xeon 2.4 GHz CPU, 2.0 GBytes RAM, and we applied from one to four disks on this testing node. In our testbed, the four homogeneous disks are placed within a disk array enclosure where there is a cooling fan. A task with buffering setting enabled issues 2000 transactions to each disk, where disk utilization is driven to as high as 100 %. Unlike disks, the CPU remains idle in each experiment.

As shown in Fig. 5, when the disks are in idle state, the initial differences between inlet and outlet temperature are 2.4 °C for one disk, 2.8 °C for two disks, 2.9 °C for three disks, and 3.4 °C for four disks. Compared with the experiment that has only one disk deployed, the experiment with four disks result in one more degree at the inlet/outlet difference. After running busy for about 20 minutes, the inlet and outlet temperature differences keeps around a specific value steadily. One disk result in a difference of 2.5 °C, two disks result in 2.8 °C, and three disks result in 3.2 °C. While four disks are applied in a single data node, the difference could be increased to around 3.7 °C.

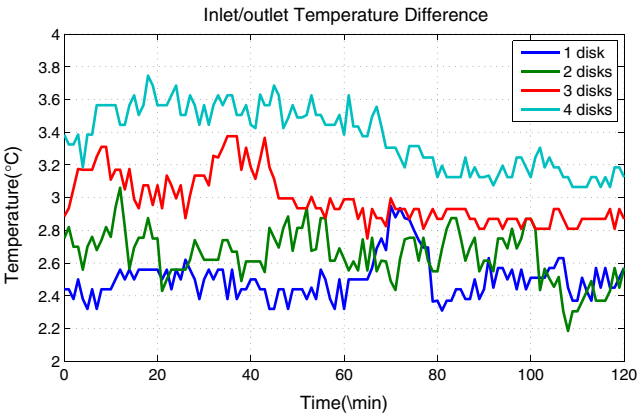


Figure 5 Inlet/outlet temperature difference with different number of disks.

Our finding shows that the inlet/outlet temperature difference of each disk is around 0.3 °C, which is an increase of 10 % since the disk's power state is transitioned from idle to active. Recall that an increment of CPU temperature for 20 °C makes the outlet temperature increase less than 2 °C (see Section 3.2.1); on the other hand, 1 °C increase in disk temperature makes the outlet temperature go up by about 0.5 °C. Compared with CPU, disks have a more significant impact on outlet temperatures. The thermal impact of disks is a major contributor to the outlet temperatures. When more disks are installed in a data node, the impact of disks become more pronounced.

Unlike CPUs and memories, extra disks can be readily installed in a data node. Nowadays, enterprise data centers contain a huge number of data nodes, which are capable of supporting more than 100 disks. When an excessive number of disks are deployed in a single storage node, the discrepancy between inlet and outlet temperatures will be enlarged.

3.2.3 Consider CPU and Disk Temperatures Together

In this group, we show how CPU and disk together impact the outlet temperature. We run two tasks together to keep the CPU and the disk both in active state with full utilization. For the CPU, a stress test running for 30 minutes is assigned. And for the disk, we assign a task with 2000 transactions, which will result in the disk running busy for about 33 minutes. The discrepancy between inlet and outlet temperature based on the CPU temperature and disk temperature is shown in Fig. 6. We observe that, when CPU and disk both

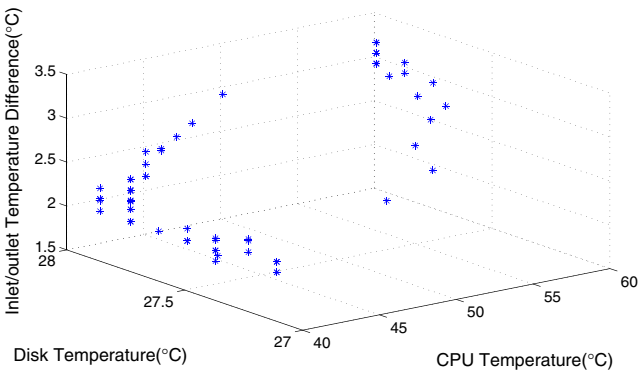


Figure 6 Impact of CPU and disk on inlet/outlet temperature difference.

keep in idle state, with the CPU temperature being 40 °C and disk temperature being 27.5 °C, the difference between inlet and outlet temperature is only 2 °C. While CPU and disk both reach their steady temperature in the active state, the difference goes up to 3.36 °C.

The aforementioned experiments indicate that outlet temperatures of a storage node heavily rely on both CPU and disk temperatures, which in turn depend on usage conditions of storage clusters. When the CPU and/or disk temperatures go up due to increased workloads, the node's outlet temperature increases accordingly. More importantly, increasing the number of disks in a storage node can further exacerbate the thermal impacts of disks on the outlet temperature of the storage node. It is worth noting that in most real-world cluster storage systems, each storage node contains multiple disks. Thus, we conclude that like processors, disks have a profound impact on outlet temperatures of storage nodes in clusters.

3.3 Thermal Characteristics of Disks

To study the thermal characteristics of HDDs and SSDs, we choose a Western Digit disk (WD1600AAJS [5]) and Intel SSD (SSDSA2M080G2GC [1]). The specifications of these two disks are shown in Table 2. The Intel SSD has a faster sequential read rate than the Western Digit HDD, but slower sequential write rate. And it also consumes less energy than the Western Digit HDD both in idle and active states.

Throughout the rest of this section, the following four features are measured to study disk thermal characteristics in the context of cluster storage systems.

1. Steady Temperature: The temperature of a disk that stays in a steady state.
2. Temperature Increment: The difference between an initial temperature and a steady temperature when a disk is active.
3. Heat-up Time: A time interval during which a disk is heating up from its initial temperature to a steady temperature when the disk is active.
4. Cool-down Time: A time interval during which a disk is cooling down from a steady temperature to the disk's initial temperature.

Table 2 Disk specification.

	WD1600AAJS	Intel SSD
Capacity (GB)	160	80
Sequential read (MB/s)	93.5	250
Sequential write (MB/s)	93.5	70
Power (Idle)	8.75 W	75 mW
Power (Active)	9.5W	150 mW

3.3.1 Different Transactions

To study how HDD and SSD would have an impact on outlet temperature, we use postmark to generate three tasks, whose configurations are shown in Table 3. The file numbers are set to 100, and file sizes are set from 1.E+6 to 1.E+8 Byte. The only difference between these three tasks are the number of transactions. All the other parameters are using postmark's default values. We run these three tasks separately on HDD and SSD while they are in idle state with their temperatures cooled down to initial temperatures.

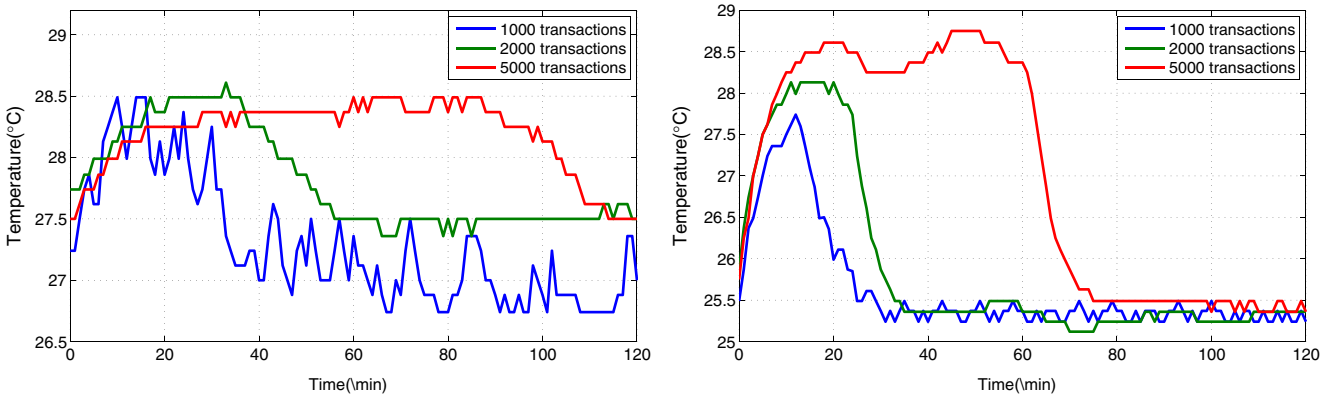
These three tasks address the Western Digital HDD running with disk utilization of 100 %. HDD's temperatures of running these tasks are shown in Fig. 7a. We observe that even with 5000 transactions assigned to the disk, the peak disk temperature is the same as running 2000 transactions. Thus, we can conclude that 2000 transactions is enough to drive the disk temperature to reach the peak temperature. And it takes about 30 minutes for the disk to heat up to the peak temperature or cool down from the peak temperature to its initial temperature.

The experiment results of running these three tasks on Intel SSD are shown in Fig. 7b. We observe that for Intel SSD, its steady temperature in idle state is around 25.75 °C, and its temperature goes up very fast when it is running busy. While running 1000 or 2000 transactions on Intel SSD, the peak temperature is not the same as running 5000 transactions. And for running 5000 transactions, the Intel SSD's temperature heats up to 28.75 °C, and then keeps steady around that value. Compared with its initial steady temperature, there is a temperature increment of 3.0 °C. Thus, when considering the thermal characters of Intel SSD, it would be better to run 5000 transactions to make sure that the Intel SSD has heat up to its steady temperature in busy state. The Intel SSD's heat-up stage is 20 minutes, and cool-down stage is a little shorter than 20 minutes. Both of them are shorter than the time for HDD.

A comparison of these two disks running 5000 transactions in the chassis is shown in Fig. 8. Intel SSD results in a higher steady temperature while working very hard, and its execution time is shorter than the execution time of HDD. We summarized the execution time and heat-up, cool-down time of these two disks (see Fig. 9) to make a better comparison. We observe that Western Digital HDD costs

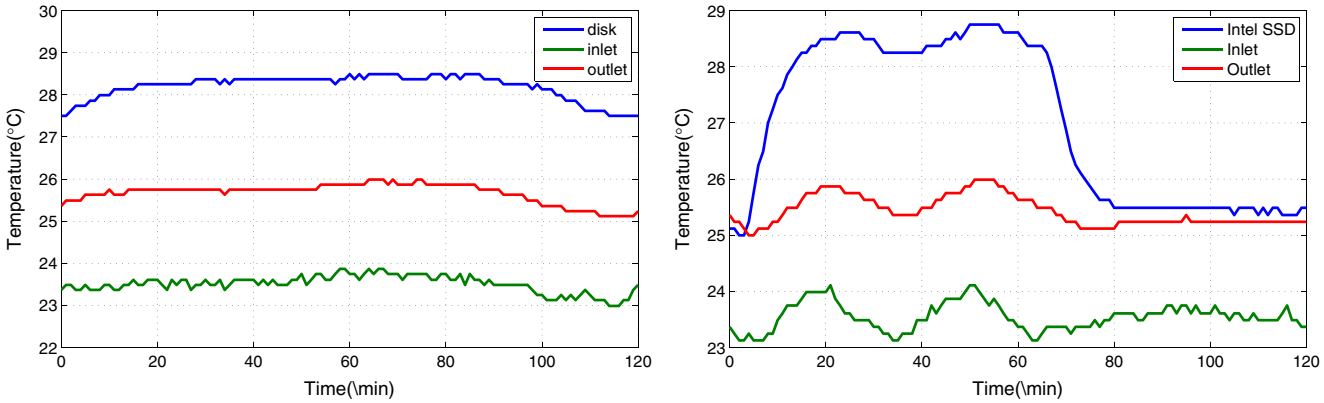
Table 3 Task configurations.

	Task1	Task2	Task3
File number	100	100	100
Transactions	1,000	2,000	5,000
File size (Byte)	1.E+6 ∼ 1.E+8	1.E+6 ∼ 1.E+8	1.E+6 ∼ 1.E+8



(a) Western Digital HDD (b) Intel SSD

Figure 7 Disk temperature of running different tasks.



(a) Western Digital HDD (b) Intel SDD

Figure 8 Thermal characteristics of running 5000 transactions.

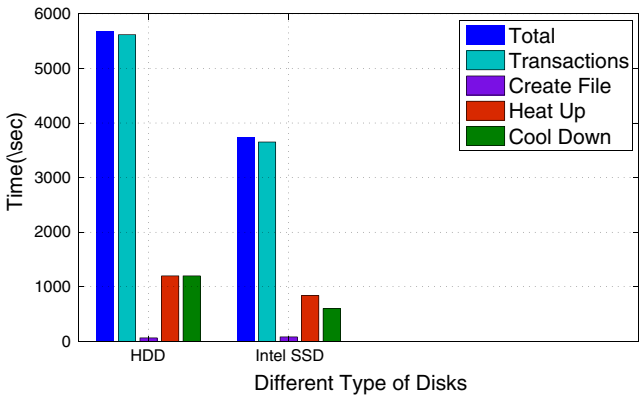


Figure 9 Time comparison of two disks running 5000 transactions.

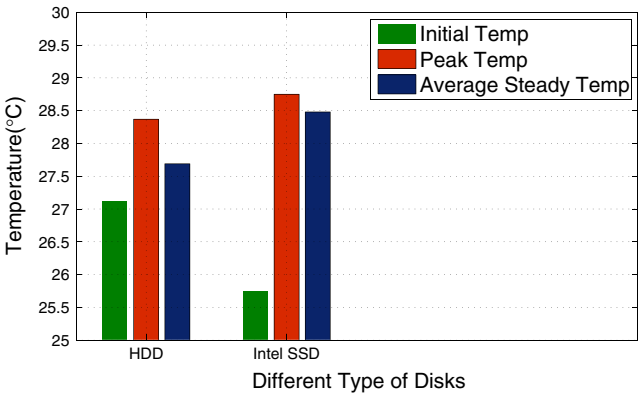


Figure 10 Temperature comparison of two disks running 5000 transactions.

about 42 % more time than Intel SSD to finish the task, which dues to SSD's significant fast read rate. And the HDD needs more time to heat-up or cool-down than Intel SSD. Figure 10 show the comparison of temperature data for these two disks. The HDD's initial temperature is about 1.5 °C higher than that of Intel SSD. However, its peak temperature and steady temperature in active state is less than Intel SSD. From all of the above, we conclude that Intel SSD is more sensitive to the disk activity, and it heats up and cools down faster than the Western Digital HDD.

3.3.2 Different Utilization

With postmark's default setting for buffering(buffer is enabled that buffered stdio function calls should be used instead of the lower level raw system calls [22]), disk would work very hard (with the disk utilization at 100 %) to finish the task as soon as possible. In order to simulate different disk utilization, we need to set postmark buffering to false. And we also found setting various write block sizes without buffering will result in different disk utilizations, while setting various read block sizes without buffering still results in 100 % disk utilization. To characterise how disk utilization impacts disk temperature, we set the buffering to false and set different write block sizes in the following part. Four experiments are conducted on each disk to study the characteristics of disks under different disk utilizations.

In the experiments running on HDD, we design tasks whose task configuration are the same as Task2 shown in Table 3 except for buffering are set to false and different write block sizes are used. In the experiments running on Intel SSD, we design tasks whose task configuration are the same as Task3 shown in Table 3 except for buffering setting and write block size. Write block size on these two group of experiments are both set to 16, 32, 64, 128 Bytes respectively.

Without buffering, the disk utilizations are different while setting different write block sizes. A comparison of the disk utilizations of HDD and Intel SSD is shown in Fig. 11. The average disk utilizations for the experiments running on HDD are 14.24 %, 28.91 %, 53.49 %, 80.57 % while setting write block size to 16 Byte, 32 Byte, 64 Byte, and 128 Byte. And for Intel SSD, the disk utilizations are 11.00 %, 30.57 %, 52.90 %, and 78.20 %, respectively. Disk utilizations of these two disks are very close when they are set to the same write block size without buffering. And we observe that higher write block size leads to higher average disk utilization for both disks.

For HDD, when write block size is 16 Bytes, the executing time is 8481 seconds; while the write block size is 32 Bytes, the executing time is 4760 seconds; when write block size is set to 64 Bytes, the running time of the task is 2973 seconds; and setting write block size to 128 Bytes results

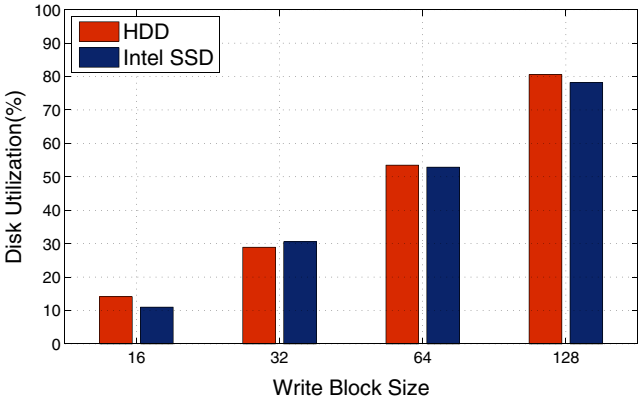


Figure 11 Disk utilizations under different write block size.

in a task executing time of 2313 seconds. We could draw a conclusion that larger write block size(/higher disk utilization) would result in shorter execution time. For SSD, it is also the same that larger write block size results in shorter execution time.

For the four experiments with different write block sizes, the initial temperature(/steady temperature in idle state) of HDD is about 28 °C. While for Intel SSD, its initial temperature is 25.75 °C. Under different disk utilization, the highest temperature that disk stay steadily is different. Peak disk temperature of these experiments could be summarized as Fig. 12. From this figure, we could observe that big write block size results in high peak disk temperature. Thus, we could have a conclusion that disk utilization has a positive impact on disk temperature.

With linear regression, equations used to estimate the disks' temperature increment could be generated. A comparison of the estimated values and real measurements of Western Digital HDD in the heat up stage is shown in Fig. 13. A comparison of the estimated values and real measurements of Intel SSD is shown in Fig. 14.

The curves marked as "$w16$", "$w32$", "$w64$", and "$w128$" in Figs. 13 and 14 show the disk temperature

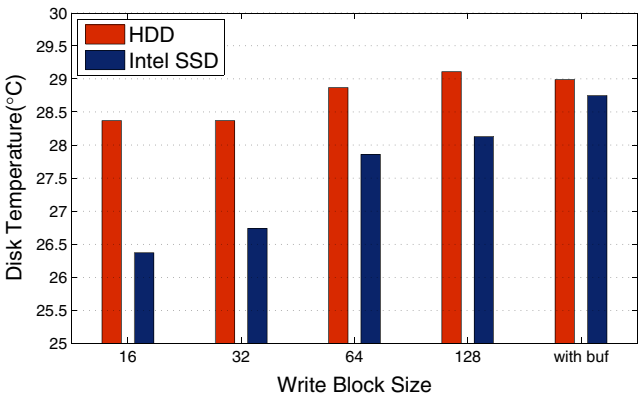


Figure 12 Peak disk temperature under different write block size.

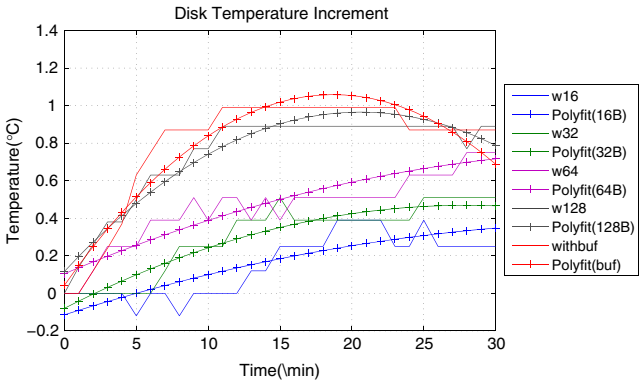


Figure 13 Western digital HDD's temperature increments produced during the heat-up stage. The $w16$, $w32$, $w64$, and $w128$ curves are the cases where buffering is disabled and the write block size is set to 16, 32, 64, and 128 Bytes, respectively. The $Polyfit(16B)$, $Polyfit(32B)$, $Polyfit(64B)$, and $Polyfit(128B)$ curves are the modeling results for the $w16$, $w32$, $w64$, and $w128$ cases. Buffering is enabled in the $withbuf$ case; the $Polyfit(buf)$ curve reveals modeling results of the $withbuf$ case.

increments produced during the heat-up stage when the buffering feature of Postmark is disabled and the write block size is set to 16, 32, 64, and 128 Bytes, respectively. The "$Polyfit(16B)$", "$Polyfit(32B)$", "$Polyfit(64B)$", and "$Polyfit(128B)$" curves plot the estimated disk temperatures generated by the linear model for the "$w16$", "$w32$", "$w64$", and "$w128$" cases. The "$withbuf$" curve plotted in Figs. 13 and 14 depicts the disk temperature increments produced during the heat-up stage when buffering is enabled in Postmark. The "$Polyfit(buf)$" curve reveals the corresponding estimated temperature produced by the model for the "$withbuf$" case.

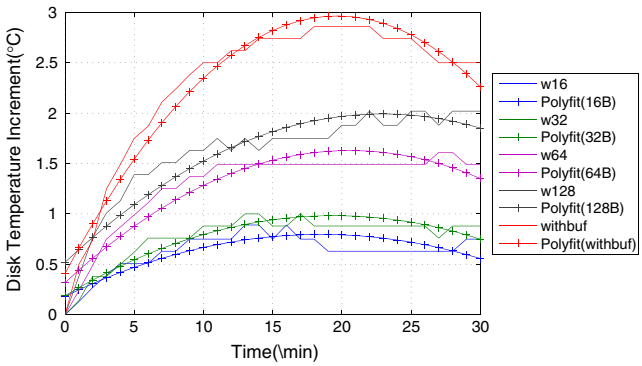


Figure 14 Intel SSD's temperature increments produced during the heat-up stage. The $w16$, $w32$, $w64$, and $w128$ curves are the cases where buffering is disabled and the write block size is set to 16, 32, 64, and 128 Bytes, respectively. The $Polyfit(16B)$, $Polyfit(32B)$, $Polyfit(64B)$, and $Polyfit(128B)$ curves are the modeling results for the $w16$, $w32$, $w64$, and $w128$ cases. Buffering is enabled in the $withbuf$ case; the $Polyfit(buf)$ curve reveals modeling results of the $withbuf$ case.

We observe from Figs. 13 and 14 that both the HDD and SSD scenarios share a similar trend in the sense that a large write block size leads to high disk utilization, which in turn gives rise to high disk temperature. The results confirm that our models are very accurate. For example, the average precision error of the HDD linear models is 0.47 %; the average precision error of the SSD linear models is 0.40 %.

According to these preliminary experiment results, we could conclude that disk utilization and the time the disk stays in active state are factors that could impact the disk temperature. We use a simple model to present the disk temperature:

$$T_{disk}(t) = f(T_i, U, t), \tag{1}$$

where T_i is the initial temperature of the disk, U is the disk utilization and t is the time that the disk is running under a specific utilization.

4 Thermal Models of Hybrid Storage Clusters

It is a challenge to model energy consumption of cooling systems for cluster storage systems in data centers. The cooling cost for cluster storage systems depends not only on cooling settings (e.g., inlet temperatures and cooling equipment placement), but also on heat dissipation of computing facilities. Processors and disks are two major heat contributors in storage nodes of clusters. In this section, we develop a model that aims to estimate outlet temperatures of storage nodes by considering the thermal impacts of processors and disks. To predict cooling costs, our model can be used in combination with a coefficient of performance model (or COP for short) that derives cooling costs from supply temperatures of computer room air conditioning units (or CRAC).

4.1 Framework

Figure 15 displays a conceptual framework of our thermal model, which consists of two components, namely, the inlet/outlet-temperature model and the COP model. The inlet/outlet-temperature model builds up the relationship between inlet and outlet temperatures via profiling analysis. In other words, given an outlet temperature, our model estimates inlet temperatures under certain CPU and I/O 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 storage node in clusters and (2) the ability to estimate cooling costs for cluster storage systems in data centers.

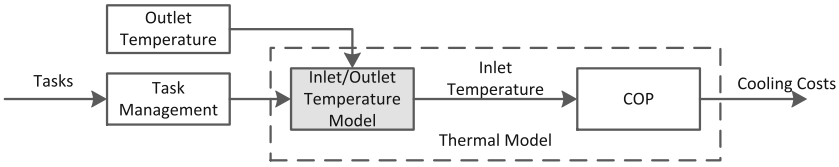


Figure 15 Framework of proposed solution.

4.1.1 The COP Model

The energy cost caused by a storage node can be attributed to energy consumption of the node and its cooling cost. We use the coefficient of performance model (or the COP model) proposed in [27] to calculate the cooling cost.

Figure 16 plots COP values that increase with the increasing supply temperature of CRAC. Equation (2) below defines the COP curve plotted in Fig. 16. A large COP value indicates a high energy efficiency in terms of cooling costs.

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

COP is defined as the ratio of heat removed from a data center to its cooling system's energy cost for heat removal. Let T be the supply temperature of CRAC. The cooling cost P_{AC} is directly proportional to the power consumption P_C of computing facilities; P_{AC} is inversely proportional to the COP value $COP(T)$. Thus, cooling power consumption P_{AC} can be derived from P_C and $COP(T)$ as (3):

$$P_{AC} = \frac{P_C}{COP(T)} \quad (3)$$

4.2 Thermal Models of Disks

In light of the thermal characteristics of disks discovered in this study (see Section 3.3), we classify disk temperatures into four phases: idle-temperature phase, heat-up-temperature phase, active-temperature phase, and cool-down-temperature phase. In this subsection, we build a temperature model by considering the characteristics of HDDs and SSDs within these four phases. To make the model sample, we consider the disks only when they are idle or fully used(with disk utilizations of 100 %). In our models, t refers to the time interval (measured in minutes), during which a disk stays in a specific temperature stage.

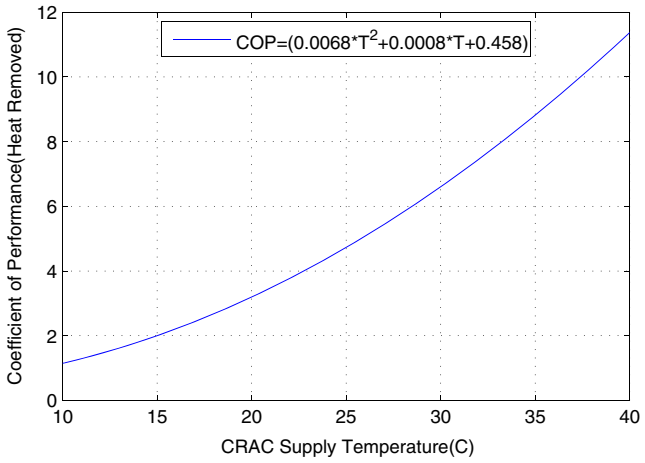


Figure 16 Coefficient of the performance curve for the chilled-water CRAC units in a utility data center at the HP Labs [27].

4.2.1 Hard Disk Drives (HDDs)

When the tested hard disk is residing in the chassis, the disk temperature remains constant during both the idle and active phases. The idle and active temperatures are 27.5 °C and 28.4 °C, respectively.

In the heat-up phase, the disk temperature gradually goes up until it reaches the maximum temperature of the active phase. The above observation motivates us to formally define disk temperature in the heat-up phase as (4):

$$T_{HDDHeat}(t) = -0.0008 * t^2 + 0.0513 * t + 27.5552 \quad (4)$$

The precision error of this heat-up model is 0.16 %.

After completing the tasks, the disk returns to the idle state. Consequently, the disk temperature drops down to the initial temperature. We formally describe this cool-down process from the perspective of temperature as (5):

$$T_{HDDCool}(t) = -0.0008 * t^2 - 0.0130 * t + 28.4154, \quad (5)$$

where t represents a time interval during which the disk temperature returns back to its initial value. The precision error of this cool-down model is 0.16 %. After the cool-down process, the disk remains in the idle state.

Incorporating the aforementioned four temperature phases, we model temperatures of hard disk drives as (6):

$$T_{HDD}(t) = \begin{cases} 27.5 & \text{(if the disk is idle)} \\ T_{HDDHeat}(t) & \text{(if the disk is heated up)} \\ 28.4 & \text{(if the disk is active)} \\ T_{HDDCool}(t) & \text{(if the disk is cooled down)} \end{cases} \quad (6)$$

4.2.2 Solid State Disks (SSDs)

Recall that the temperature of the tested SSD in the idle and active state are 25.75 °C and 28.75 °C, respectively. During

the heat-up phase, the SSD's temperature can be expressed as (7):

$$T_{SSDHeat}(t) = -0.0066 * t^2 + 0.2597 * t + 26.1594 \quad (7)$$

The precision error of this heat-up model is 0.5 %.

The SSD's temperature during the course of cooling down can be calculated as (8):

$$T_{SSDCool}(t) = -0.0027 * t^2 + 0.0085 * t + 28.7495 \quad (8)$$

The precision error of this cool-down model is 0.14 %.

Now we can derive the temperature of an SSD from $T_{SSDHeat}(t)$ and $T_{SSDCool}(t)$ as (9):

$$T_{SSD}(t) = \begin{cases} 25.75 & \text{(if the disk is idle)} \\ T_{SSDHeat}(t) & \text{(if the disk is heated up)} \\ 28.75 & \text{(if the disk is active)} \\ T_{SSDCool}(t) & \text{(if the disk is cooled down)} \end{cases} \quad (9)$$

It is worth noting that although current disk temperature is not considered in our model, the disk temperatures in the future can be predicted based on current disk temperatures and given workloads. In our model, the workloads during a given period can be divided into independent sub-workloads, in which disks are either active or idle. We assume that the disk temperature would be $temp$ at time t_{n-1}; the disk would complete the next workload w_n from t_{n-1} to t_n. If reversely calling the functions in our module, we can get t_0, which is the time when the disk virtually processes a workload, and its temperature reaches $temp$ at time t_{n-1}. Then, if the disk starts working on w_n from t_{n-1} to t_n, this case is exactly the same as the one that a long task would be processed from t_0 to t_n. Therefore, regardless of how long the disk has been active, our module works as if w_n is the first workload that starts at t_0 from the temperature's perspective. Since the disk temperature consistently increases during the heat-up stage, time t_0 for a hard drive (HDD) and a solid state disk (SSD) can be computed as:

$$t_0 = \begin{cases} t_{n-1} - T_{HDD}^{-1}(temp) & \text{(ifthediskisanHDD)} \\ t_{n-1} - T_{SSD}^{-1}(temp) & \text{(ifthediskisanSSD)} \end{cases} \quad (10)$$

After the workload w_n is handled, the disk temperature at t_n for the case of HDD and SSD can be expressed as:

$$\begin{cases} T_{HDD}(t_n - t_0) & \text{(if the disk is an HDD)} \\ T_{SSD}(t_n - t_0) & \text{(if the disk is an SSD)} \end{cases} \quad (11)$$

If the disk is sitting idle in the next sub-period, t_0 can be determined as if the disk were returning to the cooling-down phase at t_0. Therefore, given a workload set, one can compute the disk temperature after the entire workload set has been processed by recursively applying the above functions.

4.3 Thermal Model of a Storage Node

Now we focus on a thermal model that can be used to derive outlet temperatures of a storage node from inlet, CPU, and disk temperatures. In this model, we denote T_{outlet} and T_{inlet} as outlet and inlet temperatures of a storage node; let T_{disk} and T_{CPU} be disk and CPU temperatures. T_{diff} represents the discrepancy between T_{outlet} and T_{inlet}. Thus, we have $T_{outlet} = T_{inlet} + T_{diff}$.

According to our earlier experimental results (see Section 3.2), we can express the inlet/outlet temperature difference T_{diff} as:

$$T_{diff} = 0.074 * T_{CPU} + 0.3036 * T_{disk} - 9.3483 \quad (12)$$

To verify our model, we apply this model to the experiment in Section 3.2.3 where CPU and disk are both active(with disk in chassis). A comparison of estimate T_{diff} values generated by our model with the real measurements of T_{diff} is shown in Fig. 17. Our model precision error is 8.59 %.

5 Hybrid Storage Clusters

After developing a thermal model for a single disk, we are in position to investigate thermal behaviors of hybrid disks in the context of cluster storage systems, each of which is comprised of a number of storage nodes. Thanks to good I/O performance offered by SSDs, future cluster storage systems are likely to be powered by a large number of hybrid disks containing both HDDs and SSDs. In this section, we pay attention to the thermal behaviors of two types of hybrid storage clusters. We show that data placement is an efficient approach to minimize negative thermal impacts of a hybrid storage cluster for high-performance clusters.

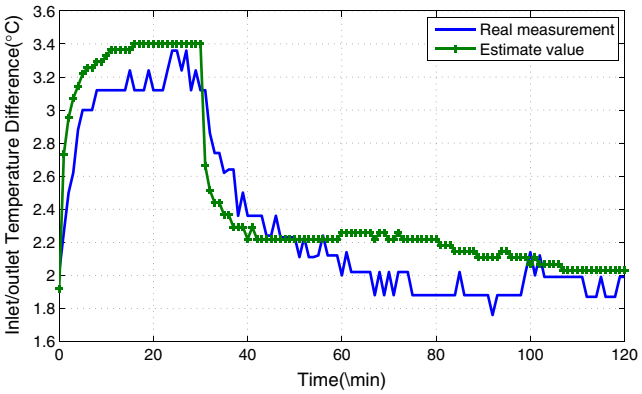


Figure 17 A comparison of our model with real measurements.

5.1 System Configuration of Hybrid Storage

In this part of study, we build two types of hybrid cluster storage systems, namely, inter-node and intra-node hybrid cluster storage systems (see Fig. 18). In an *inter-node hybrid cluster storage system*, there are two types of storage nodes – SSD-enabled nodes and HDD-enabled nodes. All disks in an SSD-enabled node are solid state disks, whereas all disks in an HDD-enabled node are hard drives. In an *intra-node hybrid cluster storage system*, each node contains both solid state disks and hard drives. Intra-node hybrid cluster storage systems are homogeneous systems in the sense that all the nodes share an identical configuration. In contrast, inter-node hybrid systems are heterogeneous systems because some nodes are equipped with SSDs while others are comprised of HDDs.

5.2 Case Studies

We investigate HDD-first and SSD-first data placement strategies, in which data would be distributed to either HDDs or SSDs. By using the HDD-first strategy, one of the HDDs will be randomly selected if both HDDs and SSDs are available; while the SSD-first strategy will choose SSDs at first. In our evaluation, the inter-node hybrid storage cluster is comprised of 128 SSD-enabled nodes and 128 HDD-enabled nodes. The intra-node hybrid storage cluster has 256 nodes. We make use of Postmark to resemble 128 I/O-intensive tasks, in each of which 1,000 files are created and 5,000 I/O requests are issued. We set the outlet temperatures of nodes to 40 °C.

5.2.1 Inter-Node Hybrid Storage Cluster

In an inter-node hybrid storage cluster (see Fig. 18a), the I/O tasks will be evenly issued to the HDD-enabled nodes by the HDD-first strategy. In this case, the requests can be completed within 88 minutes based on our preliminary experiments. According to the HDD temperature model shown in (6), the working HDD temperature increases to 28.40 °C. The temperature of another HDD in the node remains 27.50 °C. The temperatures of both SSDs residing in SSD-enabled nodes remain unchanged (i.e., 25.75 °C). We define the average value of two disk temperatures as the disk temperature of a storage node. The discrepancy between inlet and outlet temperatures of HDD-enabled nodes is $T_{diff}(27.95) = 2.10$ °C; the discrepancy between inlet and outlet temperatures of SSD-enabled nodes is $T_{diff}(25.75) = 1.43$ °C. Therefore, if the inlet temperatures of HDD-enabled and SSD-enabled nodes are 37.90 °C and 38.57 °C respectively, we could get the same outlet temperature of 40 °C. Since our preliminary experiments show that there is about 8 °C difference between the inlet temperature and the air-conditioner supply temperature, the air-conditioner supply temperatures should be set to 29.9 °C for HDD-enabled nodes and 30.57 °C for SSD-enabled nodes in order to gain the same outlet temperature of 40 °C.

The power consumptions of a HDD-enabled and SSD-enabled node are 66.25 W and 48.9 W in idle state. The COP model (see Fig. 16 in Section 4) indicates that the COP values of HDD-enabled and SSD-enabled nodes are 6.56 and 6.84. Let's consider the power consumption of this inter-node cluster. The mechanical power consumptions are 353,760 J for a HDD-enabled node and 258,129 J for an SSD-enabled node. Using the COP values, we estimate that the cooling costs with respect to HDD-enabled and SSD-enabled nodes are 53,917 J and 37,362 J. Therefore, the total energy consumption incurred by the inter-node hybrid storage cluster and its cooling system is 90,064,864 J.

By using the SSD-first strategy, the I/O requests will be evenly handled by SSD-enabled nodes. In this case, the requests can be finished within 62 minutes based on preliminary results. The temperature of the active SSD is 28.75 °C, whereas the other SSD and HDDs remain at 25.75 °C and 27.50 °C. At HDD-enabled nodes, the difference between inlet and outlet temperatures is 1.96 °C; such temperature difference at SSD-enabled nodes is 1.88 °C. Thus, The inlet temperatures of HDD-enabled and SSD-enabled nodes are nearly 38.04 °C and 38.12 °C. And the supply temperatures are 30.04 °C for HDD-enabled nodes and 30.12 °C for SSD-enabled nodes. Using the same method, we could calculate the total power consumption of this case is 63,139,305 J. The SDD-first strategy could save 42.64 % power consumption than the HDD-first strategy in the Inter-node Hybrid Storage Cluster.

5.2.2 Intra-Node Hybrid Storage Cluster

In an intra-node hybrid storage cluster, the I/O requests will be processed by HDDs in 128 nodes under the HDD-first

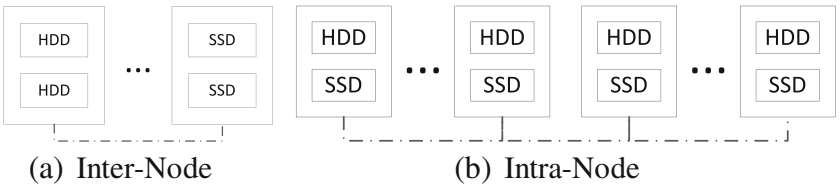


Figure 18 Two types of hybrid cluster storage systems.

strategy. The other 128 nodes will remain idle. If the SSD-first strategy is applied, the only difference from the HDD-first case is that the I/O requests will be executed on SSDs rather than HDDs.

Due to the space limitation, we do not present the intermediate results that can be calculated in a similar way. The total energy consumption is 90,022,885 J under the HDD-first strategy, and 63,137,638 J under the SSD-first strategy. The SSD-first strategy reduces the energy consumption by 42.58 %.

We observe that the total energy consumption of the HDD-first strategy on an inter-node hybrid cluster is the maximum one, and using the SSD-first strategy on intra-node hybrid cluster results in the minimum total power consumption. In the same hybrid architecture, the SSD-first strategy will save more power than the HDD-first strategy. We conclude that keeping SSD active in the intra-node hybrid storage cluster can achieve the best energy efficiency.

6 Conclusion

Cooling costs of large-scale storage clusters in data centers have been increasing in the past decade; therefore, thermal management of storage clusters must be urgently addressed. Recent studies show that cooling costs contribute a significant portion of the operational costs of data centers. Thermal management techniques are applied to reduce the energy consumption in cooling systems for storage clusters, thereby significantly improving the energy efficiency of data centers.

Thermal models play a key role in thermal management; however, there is a lack of thermal models for storage clusters. Most existing thermal models do not take into account the utilization of hard drives and solid state disks. In this paper, we proposed a thermal model to investigate thermal impacts of hybrid storage on clusters. We started this study by focusing on the thermal behavior of hard drives and solid state disks. Our model can be applied to estimate the cooling cost of a storage cluster equipped with hybrid storage nodes. We built two types of hybrid storage clusters, namely, inter-node and intra-node hybrid storage clusters. We show that, compared with the HDD-first strategy, the SSD-first strategy is an efficient approach to minimize negative thermal impacts of hybrid storage clusters for cluster computing.

Our thermal model offers the following two benefits. First, the model makes it possible to reduce thermal monitoring cost. Thermal management of hybrid storage clusters helps cut the cooling cost and energy consumption. Second, our thermal model enables data center designers to make intelligent decisions on thermal management during the design phase of hybrid storage clusters.

Acknowledgments 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. Intel ssd sa2m080g2gc. http://download.intel.com/newsroom/kits/ssd/pdfs/X25-M_34nm_DataSheet.pdf.
2. lm-sensors. http://www.lm-sensors.org/.
3. Minigooseii. http://www.itwatchdogs.com/datasheets/MiniGoose_II_User_Manual_v1_05.pdf.
4. stress-1.0.1. http://weather.ou.edu/apw/projects/stress/.
5. Wd1600aajs specification. http://www.wdc.com/wdproducts/library/SpecSheet/ENG/2879-.701277.pdf.
6. U.S. Environmental Protection Agency (2007). *Report to congress on server and data center energy efficiency.* Technical report.
7. Allalouf, M., Arbitman, Y., Factor, M., Kat, R.I., Meth, K., Naor, D. (2009). Storage modeling for power estimation. In *Proceedings of SYSTOR 2009: the Israeli experimental systems conference, SYSTOR '09* (pp. 3:1–3:10). New York. ACM.
8. Bieswanger, H.F.H.A., & Wehle, H.-D. (2012). *Energy efficient data center*. Technical Report 1.
9. Ayoub, R.Z., Indukuri, K.R., Rosing, T.S. (2010). Energy efficient proactive thermal management in memory subsystem. In *Proceedings of the 16th ACM/IEEE international symposium on low power electronics and design, ISLPED '10* (pp. 195–200). New York: ACM.
10. Balakrishnan, M., Kadav, A., Prabhakaran, V., Malkhi, D. (2010). Differential raid: rethinking raid for ssd reliability. *Transactions Storage*, *6*(2), 4:1–4:22.
11. Barroso, L.A., & Hölzle, U. (2007). The case for energy-proportional computing. *Computer*, *40*(12), 33–37.
12. Brown, D.J., & Reams, C. (2010). Toward energy-efficient computing. *Communications of the ACM*, *53*(3), 50–58.
13. Chang, L.-P. (2008). Hybrid solid-state disks: Combining heterogeneous nand flash in large ssds. In *Proceedings of the 2008 Asia and South Pacific design automation conference, ASP-DAC '08* (pp. 428–433). Los Alamitos: IEEE Computer Society Press.
14. Chen, F., Koufaty, D.A., Zhang, X. (2011). Hystor: making the best use of solid state drives in high performance storage systems. In *Proceedings of the international conference on supercomputing, ICS '11* (pp. 22–32). New York: ACM.
15. Colarelli, D., & Grunwald, D. (2002). Massive arrays of idle disks for storage archives. In *Proceedings of the 2002 ACM/IEEE conference on supercomputing, supercomputing '02* (pp. 1–11). Los Alamitos: IEEE Computer Society Press.
16. Eibeck, P.A., & Cohen, D.J. (1988). Modeling thermal characteristics of a fixed disk drive. *IEEE Transactions on Components, Hybrids, and Manufacturing Technology*, *11*(4), 566–570.
17. El-Sayed, N., Stefanovici, I.A., Amvrosiadis, G., Hwang, A.A., Schroeder, B. (2012). Temperature management in data centers: why some (might) like it hot. *SIGMETRICS Performance Evaluation Review*, *40*(1), 163–174.
18. Guerra, J., Belluomini, W., Glider, J., Gupta, K., Pucha, H. (2010). Energy proportionality for storage: impact and feasibility. *SIGOPS Operations Systematics Review*, *44*(1), 35–39.

19. Gurumurthi, S., Sivasubramaniam, A., Natarajan, V.K. (2005). Disk drive roadmap from the thermal perspective: a case for dynamic thermal management. *SIGARCH Computer Architecture News*, *33*(2), 38–49.
20. Datacenter Dynamics (2011). Global data center energy demand forecasting. http://www.dcd-intelligence.com/Census-2013/Key-Findings-2011-12. Technical report.
21. Jiang, X., Alghamdi, M.I., Zhang, J., Assaf, M.A., Ruan, X., Muzaffar, T., Qin, X. (2012). Thermal modeling and analysis of storage systems. In *Performance computing and communications conference (IPCCC) 2012 IEEE 31st international* (pp. 31–40).
22. Katcher, J. (1997). Postmark: a new file system benchmark. *System*, *3022*, 1–8.
23. Kim, Y., Gurumurthi, S., Sivasubramaniam, A. (2006). Understanding the performance-temperature interactions in disk i/o of server workloads. In *High-performance computer architecture, 2006. The twelfth international symposium on* (pp. 176–186).
24. Koomey, J.G. (2007). *Estimating total power consumption by servers in the U.S. and the world.* Technical report, Lawrence Derkley National Laboratory.
25. Li, L., Liang, C.-J.M., Liu, J., Nath, S., Terzis, A., Faloutsos, C. (2011). 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* (pp. 1370–1378). New York: ACM.
26. Mao, B., Jiang, H., Wu, S., Tian, L., Feng, D., Chen, J., Zeng, L. (2012). Hpda: a hybrid parity-based disk array for enhanced performance and reliability. *Trans Storage*, *8*(1), 4:1–4:20.
27. Moore, J., Chase, J., Ranganathan, P., Sharma, R. (2005). Making scheduling "cool": temperature-aware workload placement in data centers. In *Proceedings of the annual conference on USENIX annual technical conference, ATEC '05* (pp. 5–5). Berkeley: USENIX Association.
28. Pavlo, A., Paulson, E., Rasin, A., Abadi, D.J., DeWitt, D.J., Madden, S., Stonebraker, M. (2009). A comparison of approaches to large-scale data analysis. In *Proceedings of the 2009 ACM SIGMOD international conference on management of data, SIGMOD '09* (pp. 165–178). New York: ACM.
29. Pinheiro, E., & Bianchini, R. (2004). Energy conservation techniques for disk array-based servers. In *Proceedings of the 18th annual international conference on supercomputing, ICS '04* (pp. 68–78). New York: ACM.
30. Pinheiro, E., Weber, W.-D., Barroso, L.A. (2007). Failure trends in a large disk drive population. In *Proceedings of the 5th USENIX conference on file and storage technologies* (pp. 2–2). Berkeley: USENIX Association.
31. Sarood, O., Gupta, A., Kale, L.V. (2011). Temperature aware load balancing for parallel applications: preliminary work. In *Parallel and distributed processing workshops and phd forum (IPDPSW), 2011 IEEE international symposium on* (pp. 796–803).
32. Sarood, O., & Kale, L.V. (2011). A 'cool' load balancer for parallel applications. In *Proceedings of 2011 international conference for high performance computing, networking, storage and analysis, SC '11* (pp. 21:1–21:11). New York: ACM.
33. Schall, D., Hudlet, V., Härder, T. (2010). Enhancing energy efficiency of database applications using ssds. In *Proceedings of the third C* conference on computer science and software engineering, C3S2E '10* (pp. 1–9). New York: ACM.
34. Sharma, R.K., Bash, C.E., Patel, C.D., Friedrich, R.J., Chase, J.S. (2005). Balance of power: dynamic thermal management for internet data centers. *IEEE Internet Computing*, *9*(1), 42–49.
35. Tan, C.P.H., Yang, J.P., Mou, J.Q., Ong, E.H. (2009). Three dimensional finite element model for transient temperature prediction in hard disk drive. In *Magnetic recording conference, 2009. APMRC'09, Asia-Pacific* (pp. 1–2).
36. Tang, Q., Gupta, S., Varsamopoulos, G. (2007). Thermal-aware task scheduling for data centers through minimizing heat recirculation. In *Cluster computing, 2007 IEEE international conference on* (pp. 129–138).
37. Tang, Q., Gupta, S., Varsamopoulos, G. (2007). Thermal-aware task scheduling for data centers through minimizing heat recirculation. In *Cluster computing, 2007 IEEE international conference on* (pp. 129–138).
38. Tang, Q., Gupta, S.K.S., Varsamopoulos, G. (2008). Energy-efficient thermal-aware task scheduling for homogeneous high-performance computing data centers: a cyber-physical approach. *IEEE Trans Parallel Distribution Systems*, *19*(11), 1458–1472.
39. Vasic, N., Scherer, T., Schott, W. (2010). Thermal-aware workload scheduling for energy efficient data centers. In *Proceedings of the 7th international conference on autonomic computing, ICAC '10* (pp. 169–174). New York: ACM.
40. Wu, G., He, X., Eckart, B. (2012). An adaptive write buffer management scheme for flash-based ssds. *Trans Storage*, *8*(1), 1:1–1:24.
41. Xie, T., & Sun, Y. (2011). Understanding the relationship between energy conservation and reliability in parallel disk arrays. *Journal of Parallel Distribution Computation*, *71*, 198–210.



Xunfei Jiang is a PhD student in Computer Science and Software Engineering, Auburn University. She received the BS. and M.S. degrees in Computer Science from Huazhong University of Science and Technology (HUST), China, in 2004 and 2007. Then she joined Digital Video Networks Co., Ltd. in 2007 and Cisco Systems (Shanghai) Video Technology Co., Ltd in 2010. Her research interests include storage systems.



Maen M. Al Assaf received BS degree in computer science from Applied Science University, Amman, Jordan in 2006 and MA degree in Computer and Network Security from DePaul University (Jordan Campus) in 2008 and the Ph.D. degree in Computer Science from Auburn University, AL in 2011. Currently, he is an assistant professor of Computer Science in King Abdullah II School for Information Technology - University of Jordan, Amman, Jordan. His research focus is distributed operating systems. He is a member of IEEE, ACM, and ASIST.



Ji Zhang is a Ph.D student in the Department of Computer Science and Software Engineering at Auburn University. He received the BS. and M.S. in Computer Science from HUST, China in 2004 and 2007. He worked as a software engineer in Huawei Technologies from 2007 to 2010. His research interests include I/O-intensive computation, parallel and distributed file systems and geographic information systems.



Mohammed I. Alghamdi received the B.S. degree in Computer Science from King Saud University, Riyadh, Saudi Arabia in 1997. He received the M.S. degrees in Software Engineering and Information Technology Management from Colorado Technical University, Denver, Colorado, 2003. He received the Ph.D. degree in Computer Science from New Mexico Institute of Mining and Technology. Currently, he is an Assistant Professor with the Department of Computer Science, Al-Baha University, Kingdom of Saudi Arabia.



Xiaojun Ruan is an Assistant Professor in the Department of Computer Science, West Chester University of Pennsylvania. He received B.S. degree in Computer Science and Technology from Shandong University, Jinan, China, in 2005 and his Ph.D. degree in the Department of Computer Science and Software Engineering, at Auburn University 2011 advised by Dr. Xiao Qin. His research interests include Energy Efficient Storage Systems, parallel and distributed systems, Computer Architecture, Operating Systems, HDD and SSD technologies, and Computer Security.



Tausif Muzaffar received his BS in Electrical Computer Engineering in 2012 at Auburn University. He is currently pursing a MS degree in Software Engineering from Auburn University. He is currently a member of STARS as an Pre-K technical mentor. His research interests include Artificial Intelligence, Thermal Computing, and Large Data Mapping.



Xiao Qin received the BS. and M.S. degrees in computer science from the HUST, China, and the Ph.D. degree in computer science from the University of Nebraska-Lincoln, Lincoln, in 1992, 1999, and 2004, respectively. Currently, he is an Associate Professor with the Department of Computer Science and Software Engineering, Auburn University. His research interests include parallel and distributed systems, storage systems, fault tolerance, real-time systems, and performance evaluation. He received the U.S. NSF Computing Processes and Artifacts Award and the NSF Computer System Research Award in 2007 and the NSF CAREER Award in 2009. He is a senior member of the IEEE.