2009 29th IEEE International Conference on Distributed Computing Systems

Collaboration-Oriented Data Recovery for Mobile Disk Arrays

Tao Xie and Abhinav Sharma
Computer Science Department, San Diego State University, San Diego, USA
xie@cs.sdsu.edu, sharma4@rohan.sdsu.edu

Abstract

Mobile disk arrays, disk arrays located in mobile data centers, are crucial for mobile applications such as disaster recovery. Due to their unusual application domains, mobile disk arrays face several new challenges including harsh operating environments, very limited power supply, and extremely small number of spare disks. Consequently, data reconstruction schemes for mobile disk arrays must be performance-driven, reliability-aware, and energy-efficient. In this paper, we develop a flash assisted data reconstruction strategy called CORE (collaboration-oriented reconstruction) on top of a hybrid disk array architecture, where hard disks and flash disks collaborate to shorten data reconstruction time, alleviate performance degradation during disk recovery. Experimental results demonstrate that CORE noticeably improves the performance and energy-efficiency over existing schemes.

1. Introduction

Mobile data centers [10][16] are an alternative to conventional stationary data centers that are enclosed in buildings. They could be built on self-contained trucks, airplanes, or ships that have onboard generators, UPS, multiple high-capacity servers, and satellite Internet links [16]. For example, an NAAT MED (Mobile Emergency Datacenter) can accommodate up to 100 fully charged laptops, multiple high-performance servers, and a large capacity storage system with multiple Terabytes of data in a 20-25 ft truck [16]. Typical applications for mobile data centers include disaster recovery [6], live video broadcast [15], and homeland security [25], where high mobility and a fast large-volume data processing capability are intrinsically demanded. Apparently, mobile disk arrays are essential in these emergency-oriented applications because they can provide not only huge storage capacities but also high bandwidth. At present, a mobile disk array generally consists of an array of independent small form factor hard disks connected to a host by a storage interface like SAS (Serial-attached SCSI) [10][16].

Mobile disk arrays face several new challenges including harsh operating environments, very limited power supply, and extremely small number of spare disks, which were not experienced by their stationary counterparts before. First of all, mobile disk arrays are more prone to failures than static ones due to their severe application environments. Compared with their static counterparts mobile disk arrays generally operate in a much worse environment, which could result in a higher annual disk replacement rate [21]. Consequently, disk failures in mobile data centers become non-rare events. Next, stationary disk arrays are located in data center buildings where electrical power is guaranteed. Mobile disk arrays, however, only have very limited power supply provided by either gasoline generators or batteries. Therefore, energy-saving becomes more critical for mobile disk arrays because their energy consumption can significantly affect the life time of the entire mobile systems. Finally, very small number of spare disks can be carried in a mobile data center due to its limited space. Thus, mobile disk arrays should be able to gracefully degrade performance after disk failures occur. The new challenges demand a performance-driven, reliability-aware, and energy-efficient data reconstruction algorithm, which is executed in the presence of disk failures.

Traditional data reconstruction algorithms [1][6][8][17][24][27] strived to minimize the reconstruction time and alleviate performance degradation during the recovery process using various approaches including exploiting parallelism [6][8], integrating workload access locality into reconstruction process [24], and employing excess disk capacity [27]. Although they worked well for stationary disk arrays, they are not suitable for mobile data centers for the following two reasons. First, they

1063-6927/09 $25.00 © 2009 IEEE
DOI 10.1109/ICDCS.2009.13

did not consider the fact that mobile disks have much higher failure rates than stationary disks. Consequently, the length of the reconstruction time (or "window of vulnerability") they achieved might not be sufficiently short for mobile data centers, where the probability of a subsequent disk failure during a reconstruction process is not negligible. Second, they normally ignored energy-saving. Moreover, all of these existing data reconstruction algorithms [6][8][24][27] share one common feature: the disk arrays that they targeted on are purely rotating-based hard disk drives. Hard disk drives have some obvious shortcomings like low vibration-resistance, stringent operating environment requirements, energy-inefficiency, and long disk access latency, which make them inadequate for mobile computing environments. Thus, a fundamentally new approach to data reconstruction for mobile data centers needs to be uncovered.

In this paper we develop a collaboration-oriented data reconstruction strategy called CORE (*c*ollaboration-*o*riented *re*construction) on top of a hybrid disk array architecture. The hybrid disk array architecture (Fig. 1) integrates small capacity NAND flash based solid state disks (hereafter, flash disks) into small-factor hard disk drives to form a high-performance and energy-efficient mobile disk array. The basic idea of the CORE strategy is to let a flash disk and its corresponding buddy hard disk help each other during recovery after one of them fails. The collaboration between the flash disk and the hard disk during disk recovery comprises two elements: decreasing the volume of data needs to be rebuilt and sharing workload during recovery. CORE separates the entire data set into three sub-sets: mostly-read, mostly-write, and read-write [20]. Mostly-read data will be sorted in the ascending order in its popularity. And then popular mostly-read data will be reallocated onto the flash disk until it is full. As a result, data is distributed across the hard disk and the flash disk. If the hard disk (or the flash disk) fails, CORE only needs to rebuild the data that was originally allocated in the hard disk (or the flash disk). The decreased amount of data that needs to be reconstructed leads to a shorter reconstruction time compared to rebuilding the whole data set. Further, since each disk accommodates part of arrival requests, the flash disk (or the hard disk) can continue to serve its part when its buddy hard disk (or flash disk) is under reconstruction. Consequently, the interference on the hard disk (or the flash disk) caused by serving normal user requests during recovery is alleviated. Clearly, the workload sharing during disk recovery also results in a shorter data reconstruction time and a more graceful performance degradation during recovery. Shorter reconstruction time in turn reduces the possibility of a second disk failure during disk recovery, and thus, enhances the reliability of the mobile disk array.

The rest of the paper is organized as follows. In the next section we introduce the hybrid mobile disk array architecture. Section 3 presents the CORE strategy. In Section 4 we evaluate performance of CORE using three real-world traces. Section 5 concludes the paper with summary and future directions.

2. The hybrid disk array architecture

Compared with hard disk drives, flash disks exhibit a number of salient advantages that make them ideal building blocks for mobile disk arrays. First of all, they are physically robust with high vibration-tolerance and shock-resistance [2][3][25][26]. Second, they inherently consume much less energy than mechanical mechanism based hard disks [11]. Third, they offer much fast read access times due to lack of moving parts [2][12]. Finally, very recent breakthrough in flash disk technology largely relaxes the three well-known constraints on existing flash disks: small capacity, low throughput, and limited erasure cycles [13][19]. In the meantime, the limitation of erase cycles of flash memory has been significantly escalated from 100,000 times to 1000,000 times [2]. Currently flash disks are much more expensive than hard disk drives in terms of dollars per Gigabyte [18]. Besides, flash disks are constrained by their relatively poor random write performance [12], which is not a concern of hard disks. Thus, it is wise to integrate small capacity flash disks with high capacity hard disk drives to form an affordable, performance-driven, highly robust and energy-efficient hybrid disk storage system for mobile data centers. Thus, we propose a hybrid disk array architecture called FIT (*f*lash-assisted d*i*sk s*t*orage).

The FIT architecture is presented in Fig. 1. Within the FIT architecture, hard disks are organized in some RAID structure such as RAID-5. Similarly, all flash disks are arranged in the same RAID structure as the hard disk array. Both hard disks and flash disks are directly attached to the system bus (see Fig. 1). Further, each flash disk cooperates with a hard disk to compose a buddy-pair through a dedicated high-bandwidth connection. The rationale behind the buddy-pair organization is three-fold. First, two members of a buddy-pair can share workload so that each of them only needs to serve part of user requests. The load sharing is achieved by allocating mostly-read data onto the flash disk while putting mostly-write data onto the hard disk. Obviously, load sharing between a flash disk and a hard disk can noticeably improve

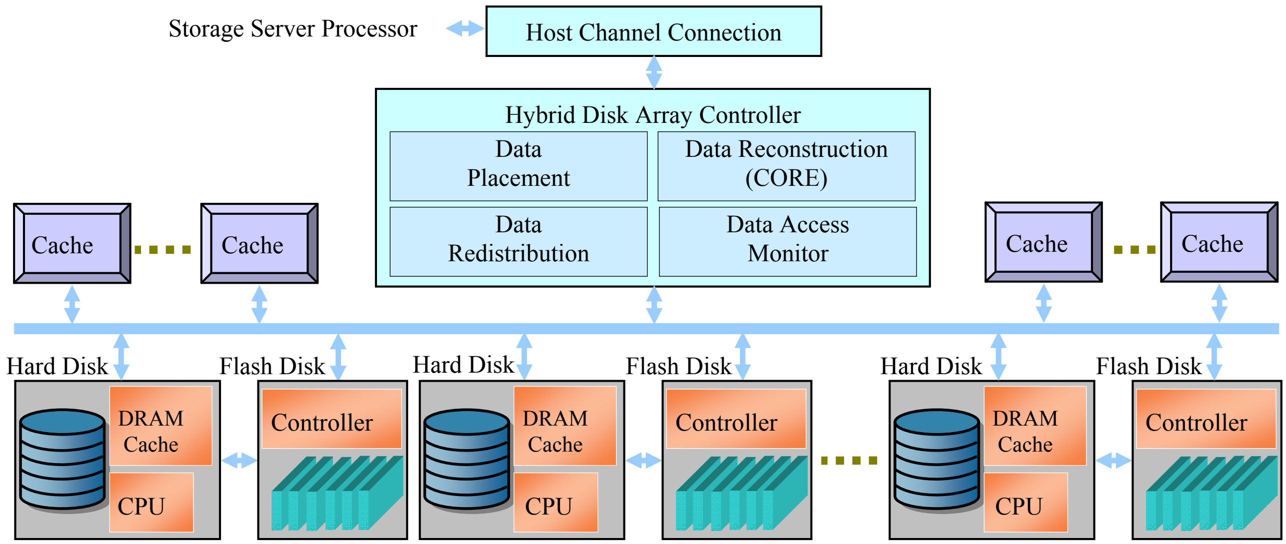

Figure 1. The FIT storage architecture.

performance compared with a traditional pure hard disk based mobile disk array. Second, the two limitations of flash disks can be largely avoided. Since only mostly-read data will be allocated on the flash disk, the concerns on its relatively poor random write performance and limited write/erasure cycles can be safely removed. Third, mutual support during reconstruction in case that one of them fails can be realized. For example, if the hard disk of a buddy-pair fails its buddy flash disk can help the hard disk recovery in two ways: only mostly-write data needs to be rebuilt and the flash disk can continue to serve majority of read requests so that the data reconstruction burden on the replacement hard disk can be alleviated. Similarly, when the flash disk fails the hard disk can help its reconstruction in a similar manner. The mutual support during recovery within a buddy-pair leads to a collaboration-oriented data reconstruction in a hybrid disk array.

Within the hybrid disk array controller, there are four software modules that manage the data across the hybrid disk array and the controller caches (see Fig. 1). The data placement module places all newly arrived data on the hard disk array. The data access monitor dynamically records each disk zone's number of accesses, and then provides the data redistribution module with a data redistribution table. Based on the data redistribution table, the data redistribution module reallocates popular mostly-read onto flash disks. When a flash disk or a hard disk fails, the data reconstruction module (CORE) is launched to provide a fault-tolerant mechanism for the hybrid disk array. Detailed information about data placement and data redistribution on hybrid disk arrays can be found in our previous work [26]. Although the CORE module does need the collaboration from the other three modules, in this research we focus on addressing data reconstruction problem in the context of a hybrid disk array in mobile data centers.

3. The core strategy

We first briefly introduce three collaborating modules of CORE: data access monitor, data placement scheme, and data redistribution algorithm. Next, implementation details of CORE are presented.

3.1. Three assistants of CORE

The CORE strategy needs assistance from data management modules within hybrid disk array controller (see Fig. 1) because the hypothesis of CORE is that data was already appropriately separated between hard disks and flash disks before a disk failure occurs. In fact, dynamic data redistribution between hard disks and flash disks to adapt to the changing data access patterns is another research topic investigated in our previous work [26]. Since the length of reconstruction time is very short compared with disk normal operating time, we assume that data access patterns are stable during this short data reconstruction.

All newly arrived data is allocated across the hard disks in RAID-5 structure by the data placement module. The data access monitor first divides the hard disk array into multiple zones. Each zone contains the same number of blocks with each block being 512 bytes. When the hybrid disk array starts to serve I/O requests, the data access monitor begins to record each zone's popularity in terms of the type and the number of accesses in a table called Data Popularity Table (DPT). The DPT table will be used later by CORE to optimize its reading sequence of reconstruction data from surviving disks. A sample DPT table is given in Table 1, where zone 3 has 109 read accesses, 7 write accesses, and its current location is 1, which represents the flash disk array.

Table 1. A sample DPT table

Zone ID	Reads	Writes	Location
1	278	133	0
2	0	58	0
3	109	7	1
…	…	…	0
m	p	q	0

Based on the statistics obtained from the DPT table, the data access monitor designates each zone as one of the following three categories: mostly-read, mostly-write, and read-write. The data redistribution module sorts all mostly-read zones in terms of number of reads in an ascending order, and then, reallocates popular mostly-read zones onto the flash disk array one by one until it is full. This way the entire data sets are separated between flash disks and hard disks.

3.2. Implementation of CORE

Suppose that one of the hard disks fails in a hybrid disk array with n hard disks and n flash disks. Fig. 2 shows the workflows of the two routines of CORE: Reconstruction Data Grabber and Reconstructed Data Restorer. CORE optimizes the reconstruction workflow by fetching reconstruction data of popular stripe units from the failed disk prior to fetching reconstruction data of unpopular stripe units, a similar idea used in [24]. In fact, the hybrid disk array controller creates n-1 processes called Reconstruction Data Grabber (RDG). Each RDG process associates with one surviving hard disk. Also, a process named Reconstructed Data Restorer (RDR) is launched in the hybrid disk array controller to write the reconstructed data onto the replacement hard disk. The functions of RDG and RDR are similar as those of the DOR algorithm [7] except for the following difference: A RDG process always selects the next most popular “under construction” unit rather than choosing next sequential unit as the DOR algorithm. A more detailed algorithm description of CORE is given as follows. While Fig. 2 only demonstrates how CORE rebuilds data for a failed hard disk, CORE is capable of reconstructing data for flash disks as well. In fact, when a flash disk recovery starts, CORE rebuilds data on a replacement flash disk in the same manner as it does for hard disk recovery.

<u>Reconstruction Data Grabber</u>
1. Sort all zones with Location 0 in DPT into a list *H* in a descending order in their total number of accesses
2. **for** each zone z_i starting from the first one in *H* **do**
3. **Repeat**
4. Issue a low-priority request to read a stripe into a buffer
5. Wait for the read request to complete
6. Submit the unit data to a centralized buffer manager for XOR, or block the process if the buffer is full
7. **Until** (all units of z_i in this disk have been read)
8. **end for**

<u>Reconstructed Data Restorer</u>
1. **Repeat**
2. Request the next full buffer from the buffer manager, blocking itself if none is available
3. Issue a low-priority write to the replacement disk.
4. Wait for the write to complete.
5. **Until** (the failed disk has been reconstructed)

Figure 2. The CORE strategy.

When a hard disk fails, CORE sorts all zones whose location value is 0 in the DPT table (see Table 1) into a list H in a descending order in their total number of accesses. Each of the n-1 RDG processes then starts to fetch stripe units from its associated surviving hard disk according to the zone ID sequence presented in the H list. The RDG process associated grabs the stripe unit into a central buffer located in the hybrid disk array controller. Similarly, the RDG process fetches the strip unit into the center buffer. The central buffer manager XORs stripe units to obtain the original data of the failed stripe unit. On the other hand, the RDR process reads the stripe unit restored in the central buffer, and then, writes it onto the replacement hard disk. The hybrid disk array restores to its normal state after all stripe units on failed disk are rebuilt. The advantage of building popular data before unpopular data is that more user requests can be served during reconstruction. Hence, the performance degradation due to disk recovery can be mitigated and the reconstruction time can be decreased.

4. Performance evaluation

In this section, we present our experimental results for a variety of hybrid disk array configurations. Reconstruction time will be the primary performance metric in this study. We also test mean user response time during reconstruction and energy consumption during reconstruction for all of the three algorithms. Although CORE can reconstruct data for both flash disks and hard disks when disk failure occurs, we only present experimental results of hard disk recovery due to the following two reasons. First, compared with flash disks, hard disks are more prone to failures in an unfriendly mobile computing environment. Thus, hard disk recovery is more frequently observed than flash disk recovery. Second, the two existing data reconstruction algorithm, DOR [7] and PRO [24], were developed for hard disk recovery. In order to make the comparisons between CORE and the two baseline algorithms fair, it is necessary for us to measure experimental results of hard disk recovery. Three real-world traces Financial1, Financial2, and WebSearch3 [22] are used in this simulation study to evaluate the performance of CORE as well as DOR [7] and PRO [24]. Financial1 and Financial2 were collected from requests to OLTP applications at two large financial institutions. WebSearch3 is an I/O trace from a popular search engine [22]. Since the simulation times in our experiments are much shorter compared with the time spans of the traces, we only use the first 8,000 seconds of each trace. In Section 4.1, the experimental settings for the simulations are described. We investigate the impact of flash disk capacity in Section 4.2 and the scalability of a hybrid disk array in Section 4.3.

4.1. Experimental setup

We developed a trace-driven simulator FITSim that models a hybrid disk array, which has one hard disk array and one flash disk array. Each disk array is made up of *m* disks organized in a RAID-5 structure. For hard disk, FITSim uses the parameters of the Seagate Cheetah 15K.4 73.4 GB [4]. For flash disk, it adopts the specifications of the Adtron A25FB-20 Flashpak and the capacity varies from 1 GB to 4 GB with 3 GB as the default value [23]. The main characteristics of the hard disk and the flash disk used by FITSim are shown in Table 2. The number of flash disks is always equal to the number of hard disks and it varies in the range (5, 7, 9, 11) with 7 as the default value. The zone size is set to 10 Mbytes. Performance metrics are:

- *Reconstruction duration*: i.e., reconstruction time, the time in seconds needed for a hybrid disk array recovering from failure mode to normal mode.
- *Mean user response time during reconstruction*: average user response time in milliseconds during recovery process.
- *Energy consumption*: the amount of energy in Joules consumed by reconstruction process.

The DOR algorithm is recognized as the most effective existing reconstruction algorithm as it has been implemented in many real applications [7]. On the other hand, PRO algorithm is one of the latest advances in storage system reconstruction optimization [24]. In order to comprehensively evaluate the CORE strategy, we compare it to DOR and PRO in this section. A brief introduction of the two algorithms is presented below.

(1) DOR (Disk-oriented Reconstruction): In a disk array with n disks, DOR activates n-1 processes associated with n-1 surviving disks to sequentially fetch reconstruction data and then put it in a centralized buffer. Also, one process dedicated for the replacement disk repeatedly transfers reconstructed data from the centralized buffer to the replacement disk. The goal of DOR is to absorb all of the disk array's bandwidth that is not absorbed by users.

(2) PRO (Popularity-based Multi-threaded Reconstruction): PRO reconstructs high-popularity data units of a failed disk, which are the most frequently accessed units in terms of the workload characteristics, prior to reconstructing other units.

Table 2. Simulation parameters

Hard disk	Seagate Cheetah 15K.4	Flash disk	Adtron A25FB-20 Flashpak
Capacity (GB)	73.4	Capacity (GB)	1, 2, 3, 4
Spindle speed (RPM)	15 K	Access time (ms)	0.272
Ave. seek time (ms)	3.5	Seek time	0
Ave. latency (ms)	2.0	Read (Mbytes/sec)	78
Transfer rate (Mbytes/sec)	77	Write (Mbytes/sec)	47
Active power (watts)	17	Read/write power (watts)	3.43
Idle power (watts)	11.9	Idle power (watts)	1.91

4.2. The impact of flash disk capacity

The goal of conducting this experiment is to compare the proposed CORE against two well-known data reconstruction algorithms DOR and PRO, and to understand the impact of flash disk capacity on the performance and energy consumption of the three algorithms. We tested flash disk capacity from 1 GB to 4 GB with disk number changing from 5 to 11.

We observe from Fig. 3 that the performance and energy consumption of both DOR and PRO keep constant in all three traces. This is because neither of them utilizes flash disks. Still, PRO always outperforms DOR in all cases because it optimizes the workflows of reconstruction [24]. In all scenarios,

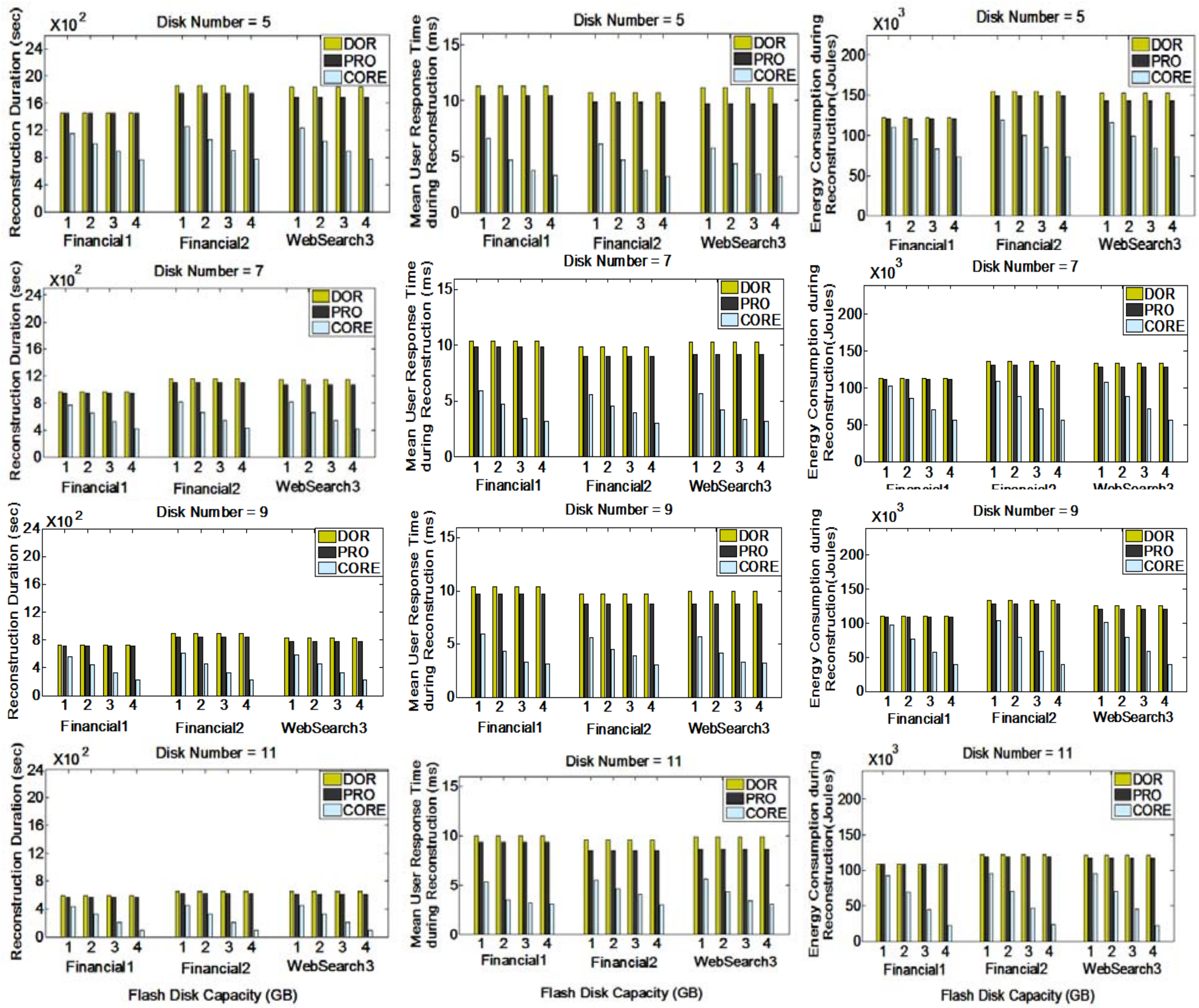

Figure 3. The impact of flash disk capacity.

CORE significantly performs better than DOR and PRO. Especially, when the capacity of flash disks is increasing, the improvement in performance and energy is also noticeably augmented. More specifically, when the hybrid disk array is organized in its default configuration (i.e., 7 hard disks, 7 flash disks, each flash disk 3 GB), compared with DOR, CORE on average reduces reconstruction duration and mean user response time during reconstruction by 50.4% and 65.3%, respectively. In terms of energy consumption, CORE on average saves energy by 43.4%. Compared with PRO, CORE on average shrinks reconstruction duration and mean user response time during reconstruction by 48.2% and 61.9%, respectively. In addition, CORE saves energy on average by 42.5%. Obviously, the performance gain and energy-saving achieved by CORE is at the cost of purchasing several small capacity flash disks. Considering that the prices of flash disks are rapidly decreasing and the substantial improvement obtained in reconstruction performance and energy conservation, we argue that the benefits of CORE outweigh the cost of flash disks. Besides, a largely shrunk window of vulnerability implies a more reliable mobile disk array, which is essential for mobile data centers. One interesting observation from Fig. 3 is that for Financial1 and WebSearch3 traces the improvement in mean user response time during reconstruction is marginal when the flash disk capacity increases from 3 GB to 4 GB. This is because majority of popular mostly-read data has been allocated onto the 3 GB flash disks. Thus, further enlarging flash disk capacity to 4 GB does not make flash disks serve much more read requests during reconstruction.

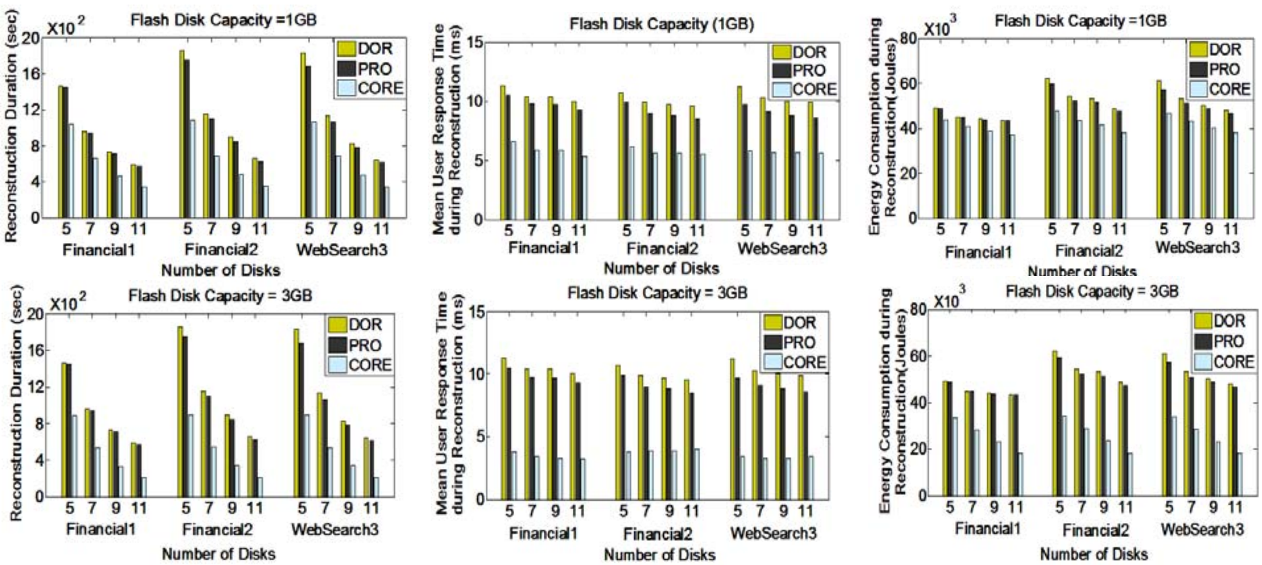

Figure 4. The impact of number of disks.

4.3. Scalability

This experiment is intended to investigate the scalability of the CORE strategy. We scaled the number of disks in the system from 5 to 11. Fig. 4 plots the performance and energy consumption of the three algorithms as functions of the number of disks.

The results show that all algorithms exhibit a good scalability. Specifically, Fig. 4 demonstrates that the three algorithms deliver better performance in both reconstruction duration and energy consumption when the number of disks increases. The results are expected because increasing number of disks leads to less data to be rebuilt on the replacement disk. However, in terms of mean user response time during reconstruction, increasing number of disks does not necessarily result in improvement for CORE. The reason is that the larger number of disks leads to lighter workload for each disk, which in turn implies less access locality can be exploited by CORE. Still, when each flash disk is 3 GB and there are 11 hard disks and flash disks, compared with DOR, CORE on average reduces reconstruction duration and mean user response time during reconstruction by 67.3% and 63.9%, respectively. In terms of energy consumption, CORE on average saves energy by 60.8%. Compared with PRO, CORE on average decreases reconstruction duration and mean user response time during reconstruction by 65.7% and 59.7%, respectively. In addition, CORE saves energy on average by 60.1%.

5. Conclusions

In this paper, we address the issue of data reconstruction in hybrid disk arrays located in mobile data centers with objectives of performance-driven, reliability-awareness, and energy-efficiency. CORE (*c*ollaboration-*o*riented *re*construction), a new data reconstruction strategy, is developed on top of a hybrid disk array architecture called FIT (*f*lash-assisted d*i*sk s*t*orage). CORE exploits the collaboration between hard disks and flash disks during reconstruction to significantly shorten data reconstruction time, alleviate performance degradation during disk recovery, and save energy. There are two main contributions of this paper. First, the idea of integrating flash disks into traditional hard-disk based disk arrays to not only substantially improve data reconstruction performance but also enhance mobile disk array reliability due to a largely shortened window of vulnerability is new. To the best of our knowledge, this research is the first work to address data recovery issue in mobile disk arrays using flash disks. Second, CORE also noticeably enhances energy-efficiency, which is critical for mobile data centers. Moreover, a hybrid disk array can inherently save energy than a hard-disk based disk array because of the use of flash disks. Extensive experiments using real-world traces show that CORE significantly improves the performance in terms of reconstruction duration and mean response time during reconstruction over two baseline data reconstruction schemes.

Acknowledgment

This work is supported by the US National Science Foundation under grants CNS-0834466 and CCF-0742187.

References

[1] G. Alvarez, W. Burkhard, L. Stockmeyer, and F. Cristian, "Declustered disk array architectures with optimal and near-optimal parallelism," *Proc. 25th Int'l Symp. Computer Architecture (ISCA)*, pp. 109-120, 1998.

[2] N. Agrawal, V. Prabhakaran, T. Wobber, J. Davis, M. Manasse, and R. Panigrahy, "Design Tradeoffs for SSD Performance," *USENIX Annual Technical Conference*, pp. 57-70, 2008.

[3] L.P. Chang and T.W. Kuo, "An adaptive striping architecture for flash-memory storage," *Proc. IEEE Real-Time and Embedded Technology Systems of Embedded systems and Applications Symp.*, pp. 187-196, 2002.

[4] Cheetah 15K.4 Mainstream enterprise disc drive storage, Seagate Cheetah ST3146854LW.pdf.

[5] M.E. Gomez and V. Santonja, "Characterizing Temporal Locality in I/O Workload," *Proc. Int'l Symp. Performance Evaluation of Computer and Telecommunication Systems (SPECTS)*, San Diego, USA, July 2002.

[6] S.M. Hawkins, D.C. Yen, and D.C. Chou, "Disaster recovery planning: a strategy for data security," *Information Management & Computer Security*, pp.222-229, 2000.

[7] M. Holland, "On-Line Data Reconstruction in Redundant Disk Arrays," Carnegie Mellon Ph.D. Dissertation CMU-CS-94-164, April 1994.

[8] M. Holland and G. Gibson, "Parity declustering for continuous operation in redundant disk arrays," *Proc. 5th Int'l Conf. Architectural Support for Programming Languages and Operating Systems*, pp. 23 - 35, 1992.

[9] M. Holland, G. Gibson, and D. Siewiorek, "Fast, On-Line Failure Recovery in Redundant Disk Arrays," *Proc. Int'l Symp. Fault-Tolerant Computing*, pp. 422-43, 1993.

[10] InfraStruXure Express: On-demand Mobile Data Center, APC Inc., 2007.

[11] T. Kgil, D. Roberts, and T. Mudge, "Improving NAND Flash Based Disk Caches," *Proc 35th Int'l Symposium on Computer Architecture (ISCA)*, pp. 327-338, 2008.

[12] H. Kim and S. Ahn, "BPLRU: a buffer management scheme for improving random writes in flash storage," *Proc 6th USENIX Conference on File and Storage Technologies (FAST)*, pp. 239-252, 2008.

[13] I. Koltsidas, and S. D. Viglas, "Flashing Up the Storage Layer," *The 34th Int'l Conf. on Very Large Data Bases (VLDB)*, Auckland, New Zealand, August 24-30, 2008.

[14] L.W. Lee, P. Scheuermann and R. Vingralek, "File Assignment in Parallel I/O Systems with Minimal Variance of Service Time," *IEEE Trans. Computers*, Vol. 49, No. 2, pp. 127-140, Feb. 2000.

[15] B.A. Mah, S. Seshan, K. Keeton, R.H. Katz, and D. Ferrari, "Providing network video service to mobile clients," *Proc. 4th Workshop on Workstation on OS*, pp. 48-54, 1993.

[16] Mobile Emergency Datacenter, North American Access Technologies, Inc., 2006.

[17] R. Muntz and J. Lui, "Performance Analysis of Disk Arrays under Failure," *Proc. 16th Int'l Conf. Very Large Data Bases*, pp. 162-173, 1990.

[18] D.A. Patterson and J.L. Hennessy, Computer Organization and Design: The Hardware/Software Interface, ISBN: 978-0-12-374493-7, Morgan Kaufmann, 2008.

[19] News Release: BiTMICRO stirs up the SSD industry with 16 GB 2.5-inch SATA solid state drive, Bitmicro, 2007.

[20] D. Roselli, J.R. Lorch, and T.E. Anderson, "A Comparison of File System Workloads," *Proc. USENIX Technical Conference*, pp. 44-54, June, 2000.

[21] B. Schroeder and G.A. Gibson, "Disk Failures in the Real World: What Does an MTTF of 1,000,000 Hours Mean to You?" *Proc. 5th USENIX conference on File and Storage Technologies*, San Jose, CA, February, 2007.

[22] SPC TRACE FILE FORMAT SPECIFICATION, last modified on June 01, 2007.

[23] Storage Products, A25FB-20-R2spec101507.pdf, 2007.

[24] L. Tian, D. Feng, H. Jiang, K. Zhou, L. Zeng, J. Chen, Z. Wang and Z. Song, "PRO: A Popularity-based Multi-threaded Reconstruction Optimization for RAID-Structured Storage Systems," *Proc. 5th USENIX conference on File and Storage Technologies*, 2007.

[25] P. Weber, A. Premji, T.J. Nohara, and C. Krasnor, "Low-cost radar surveillance of inland waterways for homeland security applications," *Proc. IEEE Radar Conference*, pp. 134-139, April, 2004.

[26] T. Xie and Y. Sun, "PEARL: Performance, Energy, and Reliability Balanced Dynamic Data Redistribution for Next Generation Disk Arrays," *Proc. 16th IEEE MASCOTS*, Baltimore, Maryland, USA, September 8-10, 2008.

[27] Q. Xin, E. Miller, and T.J. Schwarz, "Evaluation of distributed recovery in large-scale storage systems," *Proc. 13th IEEE Int'l Symp. High Performance Distributed Computing (HPDC)*, pp. 172-181, June 2004.