International Journal of Engineering Research-Online
A Peer Reviewed International Journal
Articles available online http://www.ijoer.in

Vol.3., Issue.2, 2015

REVIEW ARTICLE

ISSN: 2321-7758

REDUCING ENERGY CONSUMPTION OF DISK STORAGE IN PARALLEL DISK SYSTEMS USING ENERGY SAVING SCHEMES

JAYAPRIYA.J[1] S.NANDHAKUMAR[2]

[1]PG Scholar, Department of Computer Science and Engineering, Dhanalakshmi Srinivasan Engineering College

[2]Professor, Department of Computer Science and Engineering, Dhanalakshmi Srinivasan Engineering College

Article Received: 18/03/2015 | Article Revised on:31/03/2015 | Article Accepted on:03/04/2015

ABSTRACT

The Popular Disk Concentration (PDC) technique and the Massive Array of Idle Disks (MAID) technique are two effective energy conservation schemes for parallel disk systems. The goal of PDC and MAID is to skew I/O load toward a few disks so that other disks can be transitioned to low power states to conserve energy. To study reliability impacts of energy-saving techniques on parallel disk systems, develop a mathematical modeling framework called MINT. It makes use of data access patterns as input parameters to estimate each disk's utilization and power-state transitions. Then, derives each disk's reliability in terms of annual failure rate from the disk's utilization, age, operating temperature, and power-state transition frequency. Next, calculates the reliability of PDC and MAID parallel disk systems in accordance with the annual failure rate of each disk in the systems. Finally, use real-world trace to validate out MINT model.

Keywords : Parallel disk system, energy conservation, reliability, MAID, PDC, load balancing

1. INTRODUCTION

Parallel disk systems are of great value to large scale parallel computers, because it provides high I/o storage capacity . In the past decades, parallel disk systems have increasingly become popular for data-intensive applications running on massively parallel computing platforms. Parallel disk systems comprised of arrays of independent disks are usually cost-effective, since the parallel disk systems can be built from low-cost commodity hardware components.

Recent studies indicate that the energy cost and carbon footprint of parallel disk systems and storage services has become exorbitant. More specifically, storage devices account for approximately 27 percent of the overall energy consumption in a data centre. When it comes to Web proxies, disk energy consumption may account for up to 77 percent .Current utilization and technological trends of parallel disk systems result in unacceptable economical and environmental consequences . To address this problem, a broad spectrum of energy-saving techniques were energy conservation techniques include software-directed power management strategies , dynamic power management (DPM) schemes , data redundancy

technique, workload skew and multi-speed settings. Prior findings show that existing energy conservation techniques in disk drives can deliver significant energy savings in large-scale storage systems. Although few energy-saving schemes such as cache-based energy saving approaches may have marginal impacts on disk reliability, many energy conservation techniques like dynamic power management and workload skew techniques inevitably have adverse impacts on parallel disk systems . For example, the DPM technique reduce energy consumption in disks by the virtue of frequent disk spin-downs and spin-ups, which in turn can shorten disk lifetime . Unlike DPM, workload-skew techniques, for example, MAID ,PDC , BUD , and PARAID move frequently accessed data sets to a subset of disks arrays acting as work-horses, thereby keeping other disks in standby mode to save energy.

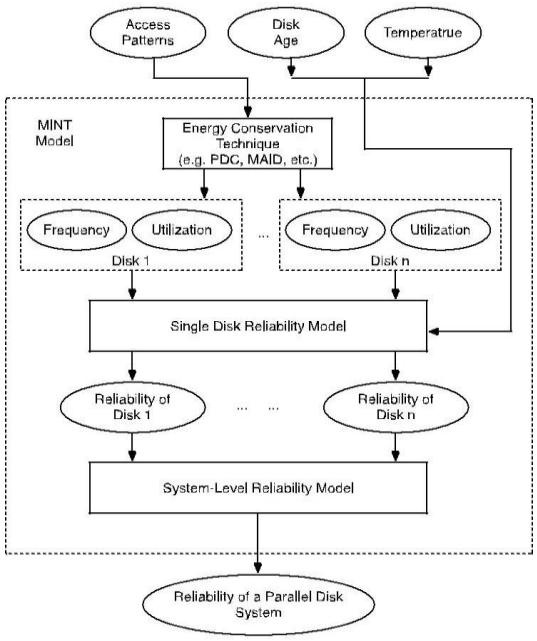

Fig 1.1 Architecture

II. Related Works

The modeling process starts by capturing the behaviors of parallel disk systems coupled with power management optimization policies. First, it makes use of data access patterns as input parameters, which are used to estimate each disk's utilization and power-state transition frequency. Then, derive each disk's reliability in terms of annual failure rate from the disk's utilization, operating temperature as well as power-state transition frequency. These three parameters are key reliability-affecting factors in addition to disk ages. Finally, calculate the reliability of the parallel disk system in accordance with the annual failure rate of each disk in the system. Fig. 1.1 depicts the framework of the MINT reliability model for parallel disk systems with energy conservation schemes. MINT is composed of a single-disk reliability modeling module, a system-level reliability modeling module, and four reliability-affecting factors - disk age, temperature, disk state transition frequency (hereinafter referred to as frequency) and utilization. Many energy-saving schemes inherently affect reliability-related factors like disk utilization and transition frequency. Given an energy optimization mechanism, MINT first transfers data access patterns into frequency and utilization - the two reliability-affecting factors. The single-disk reliability model can derive individual disk's annual failure rate from utilization, power-state transition frequency, age, and temperature. Reliabilities of all the disks in a parallel disk system are used as input to the system-level reliability modeling module that is responsible of estimating the annual failure rate of parallel disk systems. There are several reliability-related factors, among which we consider four factors in MINT. It does not, how-ever, necessarily imply that disk utilization, age, temperature, and power-state transitions are the only parameters affecting disk reliability. Other factors that may have impacts on disk reliability include handling, humidity, voltage variation, vintage, duty cycle, and altitude . must pay attention to disks that are more than one year old, because the infant mortality phenomena is out the scope of this study. The reliability models presented in this paper are focused on read-intensive I/O activities, because a wide range of applications are read-intensive in nature. These applications include, but not limited to, web applications (e.g., Gmail and Facebook), video streaming servers (e.g., Youtube, Hulu), and search engines (e.g., Google and Bing).

Disk utilization can be characterized as the fraction of active time of a disk drive out of its total powered-on-time . Different from RAID systems in which each disk only handles partial data of a file , each single disk in our disk array model stores entire files. Therefore, the access patterns of all disks in a disk array are unlikely to be identical. Thus, the disks in a MAID system or a PDC system tend to have

various utilizations, thereby leading to different failure rates rather than an identical one. In our single disk reliability model, the impacts of disk utilization on reliability is good way of providing a baseline characterization of disk annual failure rate or AFR. Using field failure data collected by Google, one can investigate the impact of utilization on AFR across the different age groups. For example, Pinheiro et al. studied AFR value of multiple disk groups with different ages, focusing on the impact on disk utilization on AFR. Disk utilization are categorized by Pinheiro et al. in three levels—low, medium, and high. Fig. 2 shows AFRs of seven disk groups, representing disks whose ages are 3 months, 6 months, 1 year, 2 years, 3 years, 4 years, and 5 years under the three utilization levels Since the single-disk reliability model in MINT needs a baseline AFR value derived from disk utilization, we make use of the polynomial curve-fitting technique to extrapolate the baseline AFR a single disk from Google's field failure date. Extrapolating AFRs from the field failure data is important, because such an extrapolation approach allows us to estimate failure rate of a disk in accordance to the disk's utilization.

In what follows, we give an example of a failure-rate model for three-year old disks. Failure-rate models of disks that are not three-year old (e.g., one-year, or five-year old) can be built in the same manner. For instance, given a three-year old disk, the AFR value under 30 percent utilization is even higher than AFR under 80 percent utilization.

It is reasonable to use MINT to compare the reliability performance of different energy-efficient storage systems, because the reliability models of the MAID and PDC storage systems use the same experimental data. It is challenging to validate the accuracy of the MINT modeling framework, since it is unable to watch MAID and PDC running for a couple of decades. One way to address this problem is to maintain and monitor a large number of MAID and PDC systems for a short period of time (e.g., 5 to 10 years). If one can watch the MAID and PDC systems over their entire service life, failure-rate data will be collected to validate reliability models. Even if it can test MAID and PDC with 100 disks for five years, the sample size is still considered small.

To address this validation problem, verify MINT using the combination of the following two validation techniques , which are practical approaches to verification and validation of models.

Reliability impacts of power management on disks. Recent studies show that both power management and workload skew schemes inherently impose adverse reliability impacts on disk systems . For example, the power management schemes are likely to result in a huge number of disk spin-downs and spin-ups that can significantly reduce the lifespan of hard disks.

The workload skew techniques dynamically migrates frequently accessed data to a subset of disks which inherently have higher risk of breaking down than other disks usually being kept standby. Disks storing popular data tend to have high failure rates due to extremely unbalanced workload. Thus, the popular data disks have a strong likelihood to become reliability bottleneck. The design of MINT is orthogonal to the aforementioned energy saving studies, because MINT is focused on reliability impacts of the power management and workload skew schemes in parallel disks. Model validation is a process of improving levels of confidence. Major approaches to validating models include historical methods and extreme condition tests. It has a trace-driven simulator using the Berkeley Web Trace as a reference model with which the MINT model is compared. The Web trace is used to drive the simulator, because if focuses on read-intensive applications .

III. Scope and Outline of the Paper

The contribution of this paper is 1) to present a new reliability modeling framework for energy-efficient parallel disk systems and 2) to improve energy-efficient parallel disks by alternating disks storing hot data with disks holding cold data. In particular, the main contributions of this study include are summarized as follows:

Implement a generic mathematical approach—called MINT (Mathematical Reliability Models for Energy- efficient Parallel Disk System) to modeling reliability of energy-efficient parallel disks coupled with power management optimization policies. We built two reliability models for the two well known energy saving schemes-PDC and MAID. Analyze intriguing the impacts of PDC and MAID on the annual failure rates of parallel disk systems.

Impacts of Disk Utilization

When the average file access rate increases,

the utilizations of PDC, MAID-1, and MAID-2 increase accordingly. Compared with the utilizations of MAID-1 and MAID-2, the utilization of PDC is a whole lot more sensitive to the file access rate. The utilization of PDC is significantly higher than those of MAID-1 and MAID-2. For example, when the average file access rate is 5 *10^5 no./month, the utilizations of PDC, MAID-1, and MAID-2 are approaching to 90, 48, and 40 per-cent, respectively. PDC has high utilization, because disks in PDC spend noticeable amount of time in migrating data among the disks.

Impacts of Temperature on Disks

Temperature is often considered as the most important environmental factor affecting disk reliability. Field failure data of disks in a Google data center shows that in most cases when temperatures are higher than 35¯C, increasing temperatures lead to an increase in disk annual failure rates ,the failure rates decreases when temperature increases . Growing evidence shows that disk reliability should reflect disk drives operating under environmental conditions like temperature . Since temperature apparently affect disk reliability, the temperature can be considered as a multiplier to baseline failure rates where environmental factors are integrated .

Power-State Transition Frequency

To conserve energy in single disks, power management policies turn idle disks from the active state into standby. The disk power-state transition frequency (or frequency for short) is often measured as the number of power-state transitions (i.e., from active to standby or vice versa) per month. The reliability of an individual disk is affected by power-state transitions and; therefore, the increase in failure rate .

Single Disk Reliability Rate

Initially compute a baseline AFR as a function of disk utilization. Then use temperature factor as a multiplier to the baseline AFR. Finally, we add a power-state transition frequency adder to the baseline value of AFR. Hence, the failure rate R of an individual disk can be expressed as

$$F ¼ a \_ F_{base} \_ t þ b \_ F_{freq}$$

where F_{base} is the baseline failure rate derived from disk uti-lization , t is the temperature factor , F_{freq} is the power-state transition frequency adder to the base AFR and a and b are two coefficients for the value of reliability F . If reliability F is more sensitive to frequency than to utilization and temperature, then b must be greater than a. Otherwise, b is smaller than a. In either case, a and b can be readily set in accordance with F's sensitivities to utilization, temperature, and frequency.

Annual Failure Rate

AFR of PDC keeps increasing from 5.6 to 8.3 percent when the file access rate is larger than 150,due to data migrations. Unlike PDC's AFR, the AFRs of MAID-1 and MAID-2 continue decreasing from 6.3 to 5.8 percent with the increasing file access rate. This declining trend might be explained by two reasons. First, increasing the file access rates reduces the number of power-state transitions. Second, the range of the disk utilization is close to 40 percent, which is in the declining part of the curve.

Configuration of the MAID and PDC systems

Energy-saving Scheme	Number of Disks	File Access Rate (No. per month)	File Size (KB)
PDC	20 data (20 in total)	0~10^6	300
MAID-1	15 data+5 cache (20 in total)	0~10^6	300
MAID-2	20 data+5 cache (25 in total)	0~10^6	300

The main difference between MAID and PDC is that MAID makes data replicas on cache disks, whereas PDC lays data out across disk arrays without generating any replicas. If one of the cache disks fails in MAID, files residing in the failed cache disks can be found in the corresponding data disks. In contrast, any failed disk in PDC can inevitably lead to data loss. Although PDC tends to have lower reliability than MAID, PDC does not need to trade disk capacity for improved energy efficiency and I/O performance.

IV. Conclusion

In recognition that there is a lack of models designed to evaluate reliability of energy-efficient disk systems, in that a new modeling framework called MINT to measure the reliability of parallel disk systems equipped with reliability-affecting energy conservation techniques. Initially the model is developed to study the impacts of disk utilization and power-state transition frequency on reliability of each disk in a parallel disk system. And the reliability of an individual disk is derived from its utilization, age, temperature, and power-state transitions. Finally, the MINT framework is applied to investigate the reliability of parallel disks coupled

with the MAID technique and the PDC technique. MINT has the following advantages:

MINT captures the behaviors of PDC and MAID in terms of data movement and migration as well as power-state transitions.

MINT seamlessly integrates multiple reliability-affecting factors into a coherent form.

MINT can be used to evaluate the system-level reliability of an energy-efficient parallel disk system.

The validation results indicate that MINT exhibits a similar trend as that of the simulator driven by a real-world trace.

REFERENCES

[1] "The now Trace Collection Project," http://tracehost.cs.berkeley. edu/web/.

[2] "The Distributed-Parallel Storage System (Dpss) Home Pages," http://www-didc.lbl.gov/DPSS/, June 2004.

[3] K. Bellam, A. Manzanares, X. Ruan, X. Qin, and Y.-M. Yang, "Improving Reliability and Energy Efficiency of Disk Systems via Utilization Control," Proc. IEEE Symp. Computers and Comm., 2008.

[4] R.E. Brown and J.R. Ochoa, "Distribution System Reliability: Default Data and Model Validation," IEEE Trans. Power Systems, vol. 13, no. 2, pp. 704-709, May 1998.

[5] W.A. Burkhard and J. Menon, "Disk Array Storage System Reliability," Proc. 23rd Int'l Symp. Fault-Tolerant Computing, 432-441, 1993.

[6] Enrique V. Carrera, Eduardo Pinheiro, and Ricardo Bianchini, "Conserving Disk Energy in Network Servers," Proc. 17th Ann. Int'l Conf. Supercomputing (ICS '03), pp. 86-97, 2003.

[7] D. Colarelli and D. Grunwald, "Massive Arrays of Idle Disks for Storage Archives," Proc. ACM/IEEE Conf. Supercomputing, pp. 1-11, 2002.

[8] F. Douglis, P. Krishnan, and B. Marsh, "Thwarting the Power-Hungry Disk," Proc. USENIX Winter Technical Conf., pp. 23-23, 1994.

pp. J.G. Elerath, "Specifying Reliability in the Disk Drive Industry: No More MTBF's," Proc. Ann. Reliability and Maintainability Symp., 194-199, 2000.

[9] J.G. Elerath and M. Pecht, "Enhanced Reliability Modeling of Raid Storage Systems," Proc. IEEE/IFIP Int'l Conf. Dependable Systems and Networks, 2007.

[10] S. Gurumurthi, A. Sivasubramaniam, M. Kandemir, and H. Franke, "DRPM: Dynamic Speed Control for Power Management in Server Class Disks," Proc. 30th Ann. Int'l Symp. Computer Archi-tecture, pp. 169-179, June 2003.

[11] D.P. Helmbold, D.E. Long, T.L. Sconyers, and B. Sherrod, "Adaptive Disk Spin—Down for Mobile Computers," Mobile Net-works and Applications, vol. 5, no. 4, pp. 285-297, 2000.

[12] G.F. Hughes and J.F. Murray, "Reliability and Security of Raid Storage Systems and d2d Archives using Sata Disk Drives," ACM Trans. Storage, vol. 1, no. 1, pp. 95-107, Feb. 2005.

[13] X.-F. Jiang, M.I. Alghamdi, J. Zhang, M. Al Assaf, X.-J. Ruan, T. Muzaffar, and X. Qin, "Thermal Modeling and Analysis of Stor-age Systems," Proc. IEEE 31st Int'l Performance Computing and Comm. Conf., 2012.

[14] S. Jin and A. Bestavros, "Gismo: A Generator of Internet Streaming Media Objects and Workloads," ACM SIGMETRICS Performance Evaluation Rev., vol. 29, Nov. 2001.