Proactive Disk Failure Management and Data Protection for Highly Available Storage Systems

Song Huang [1] Song Fu [1] Weisong Shi [2] Devesh D Tiwari [3]
[1]University of North Texas, [2]Wayne State University, [3]Northeastern University
songhuang@my.unt.edu, song.fu@unt.edu, weisong@wayne.edu, tiwari@northeastern.edu

Introduction and Objectives

Disk failures that cause data loss and system malfunctions have severe impacts on the performance and availability of data centers. The objective of this work is to provide proactive disk failure prediction method:

- Categorize the failures into different groups, characterize and understand their features.
- Pinpoint the possible causes of each type of failures and provide remediations.
- Describe the degradation processes of disk failures and predict the upcoming failures.

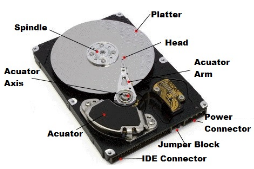


Figure : Disk components

Disk Fault Modes

Each component of the disk can fail and the failures show various manifestations. The failures can be:

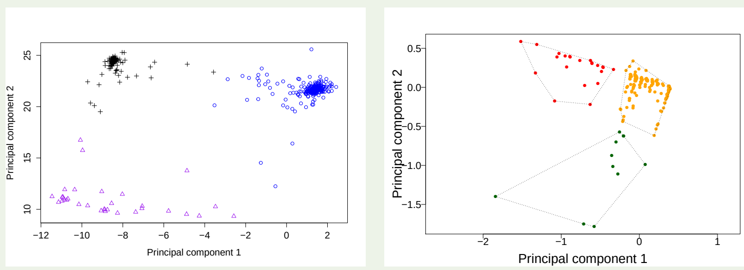
- **Physical failure**, e.g., head crash, motor failure, and media defects.
- **Logical failure**, e.g., corrupted file and file system error.

Failure can also be differenticated by extents of severeness and failing time windows:

- **Immediate and total failure**, e.g., broken head and stiction, strong vibrations.
- **Progressive failure**, e.g., bad sectors and wearing failures

Each kind of failures have their feature and possible causes. We try to pinpoint the causes and discuss the remedy actions for different kinds of failures.

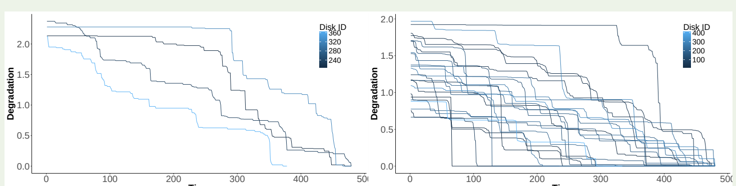
Disk Failure Categorization



(a) Failure clusters of all replaced drives. Logical failures(circles); read/write head failures(crosses); bad sector failures(triangles)
(b) Subgroups in logical failures. SUT failures (red); HFW failures (green); others logical failures (orange)

Figure : Failure clustering. Apply K-means clustering algorithm on the data of failed drives to separate failures into different groups.

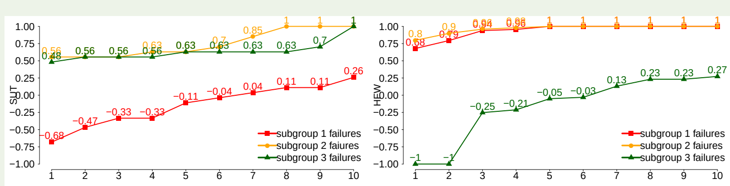
Degradation Processes of Disk Failures



(a) Degradation of Read/Write head failures. (b) Degradation of bad sector failures.

Figure : Degradation of different kinds of failures. Use Euclidean distance method to calculate distances to the failure points.

Two Types of Logical Failures



(a) SUT deciles of different subgroup of logical failures. (b) HFW deciles of different subgroup of logical failures.

Figure : Deciles of logical failures. Analyze the features of subgroups of failues by calcualating the deciles.

Important Results

After categorizing the failed drives into different groups, we find out the features of different kinds of failures:

- RUE and RSC are highly correlated with read/write head failures and bad sector failures respectively.
- We define the degradation status by calculating the Euclidean distance betwen the real time status to the failed point.
- SUT faults are manifested by longer spin up time, which are caused by different power management modes.
- HFW faults are demonstrated by the read/write head flying out of normal range, which result from the accumulated lubricant between platters and heads.

Conclusion

We proposed a proactive approach to discover disk failure categoriies, pinpoint the failure causes and model the disk health degradation processes. This enable us to predict disk failures of different types more accurately and efficiently, and apply cost-effective disk recovery and data rescue techniques to assure highly dependable storage systems.

Acknowledgements

This work was supported in part by NSF under grant CCF-1563750.

References

[1] Song Huang, Song Fu, Quan Zhang, and Weisong Shi. Characterizing disk failures with quantified disk degradation signatures: An early experience. In *IISWC, 2015*, pages 150–159. IEEE, 2015.

[2] Ao Ma, Rachel Traylor, Fred Douglis, Mark Chamness, Guanlin Lu, Darren Sawyer, Surendar Chandra, and Windsor Hsu. Raidshield: characterizing, monitoring, and proactively protecting against disk failures. *ACM TOS*, 11(4):17, 2015.

Proactive Disk Failure Management and Data Protection for Highly Available Storage Systems

Extended Abstract

Song Huang and Song Fu
Department of Computer Science and Engineering
University of North Texas
SongHuang@my.unt.edu,song.fu@unt.edu

Weisong Shi
Department of Computer Science
Wayne State University
weisong@wayne.edu

Devesh D Tiwari
Department of Electrical and Computer Engineering
Northeastern University
tiwari@northeastern.edu

ABSTRACT

Disk failures that cause data loss and system malfunctions have severe impacts on the performance and availability of data centers. Existing data protection practices are mostly reactive, i.e., disk rebuilds after disks fail, and expensive. In this paper, we present a proactive method to improve storage reliability by tracking disk health degradation and failure prediction. This is realized by achieving a deep understanding of disk failures, their characteristics and degradation processes. We leverage machine learning technologies to identify failures with distinct manifestations and further determine their types. Then we analyze each failure type, pinpointing possible causes, proposing remediations and modeling the degradation processes of disks. Our experimental results using SMART (Self-Monitoring, Analysis and Reporting Technology) data collected from a large-scale, production data center show that logical failure is a significant contributor to disk replacements and it can be mitigated by changing power management policy and data access patterns. We can also model the degradation of disk health by using only a limited number of SMART attributes.

1 INTRODUCTION

Storage systems become increasingly larger with the booming of big data and extreme-scale computing applications. Although hard drives are reliable in general, they are believed to be the most commonly replaced hardware components [3, 4]. It is reported that 78% of all hardware replacements were for hard drives in the data centers of Microsoft [4]. Moreover, with the increased capacity of single drives and an entire system, block and sector level failures, such as latent sector errors [1], cannot be ignored anymore.

The common practice to handle disk failures is reactive, that is disk rebuild in RAID systems. However, disk rebuild is expensive. It takes several days to rebuild an 8 TB disk drive. Data is not available on the failed disk and the entire RAID suffers from severe performance degradation. Data loss happens when more drives fail during a rebuild.

To address these issues, in this paper, we present a proactive method to improve storage reliability by tracking disk health degradation and failure prediction. Specifically, we analyze manifestations of disk failures in a production data center and explore data mining techniques combined with statistical analysis methods to discover different categories of disk failures and their distinctive properties. For logical failures which account for a significant portion of overall disk failures, we pinpoint the possible causes and propose remediations to reduce their occurrences. We use similarity measures to quantify the degradation process of each failure category. We conduct performance evaluation on a SMART dataset collected from a production data center. Experimental results show that the degradation of disk health can be modeled by using a small number of SMART attributes, and changing power management policy and data access patterns can mitigate logical failures and thus reduce disk replacements in a data center.

2 FAULT MODES AND SMART DATA

Disk drive is a complex system composed of a large number of magnetic, mechanical, and electronic components. Each of these components can fail and their failures show various manifestations with different extents of severeness. Disk failures can be physical (e.g., head crash and motor failure) or logical (e.g., corrupted file and file system error); immediate and total (e.g., broken head and stiction) or progressive (e.g., bad sectors and wearing failure). For example, media defects and handling damage can cause a failure due to excessive bad sectors. Head crashes or broken heads can lead to a large number of read/write errors. Motor or bearing failures are potentially caused by handling damages. Bad servo positioning can also lead to failures. In addition, corrupted files and human errors can cause damages to disk's file structure or software leading to logical failures.

Our SMART dataset has been collected from a production data center, which consists of more than 23,000 enterprise-class disk drives, over eight weeks. In total, 433 failed drives and 22,962 working drives are recorded in the dataset.

3 METHODOLOGY AND EXPERIMENTAL RESULTS

In our previous work [2], we used data clustering techniques, including K-means clustering and support vector clustering (SVC), to characterize manifestations of all failed disks. Three distinct groups have been identified, as shown in Fig. 1 and the corresponding types were determined. In this work, we focus on invesgating the logical failures, which account for 59.6% of total number failures, by pinpointing their possible causes and provide remedy solutions. Meanwhile, we improve the degradation models of physical failures based on our previous results.

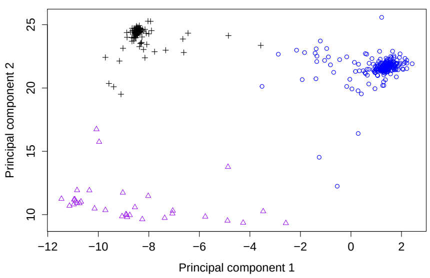
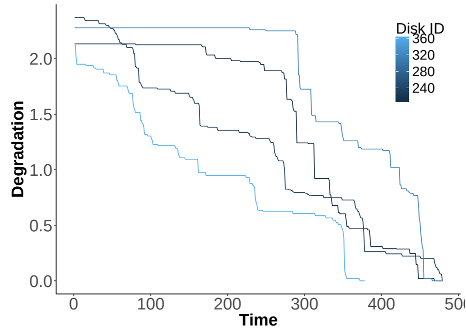
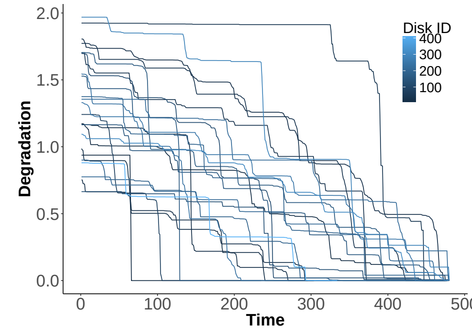


Figure 1: Groups of failure manifestations of all replaced drives. Logical failures (blue circles): 59.6%, read/write head failures (black crosses): 32.8% and bad sector failures (purple triangles): 7.6%.



(a) Degradation curves of read/write head failures.



(b) Degradation curves of bad sector failures.

Figure 2: Degradation processes of read/write head failures and bad sector failures.

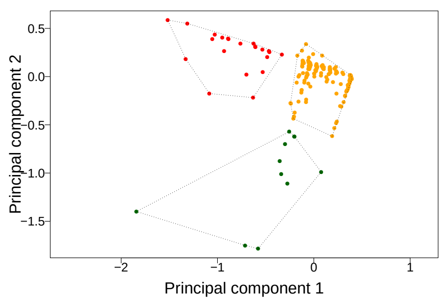
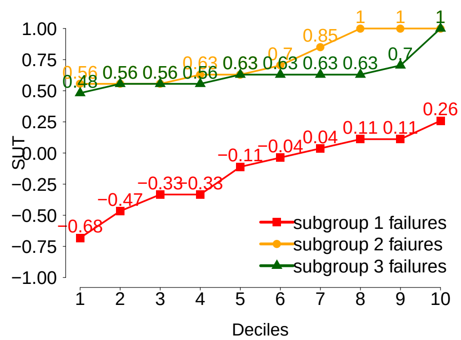
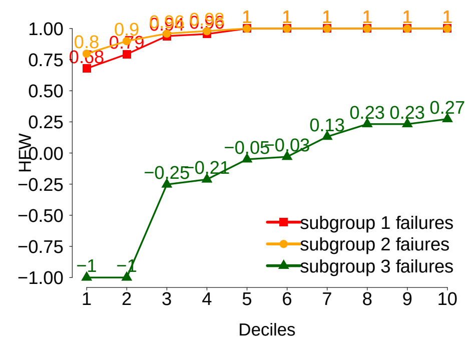


Figure 3: Sub groups of logical failures



(a) Decile distributions of Spin Up Time (SUT).



(b) Decile distributions of High Fly Write (HFW).

Figure 4: Decile distributions of logical failures on SUT and HFW.

3.1 Remediations for Logical Disk Failure

To pinpoint the causes of logical failures, we further leverage data clustering only on the 258 logical failures identified from the previous work. Three sub-groups are produced by both K-means and SVC, as shown in Fig 3. To analyze their difference, we calculate the deciles of the three sub-groups for all collected SMART attributes. Fig. 4a and 4b show the distributions of Spin Up Time (SUT) and High Fly Write (HFW). Drives from Subgroup 1 take longer time to spin platters to full speed, while read/write heads of drives from Subgroup 3 are higher than normal compared with the other two subgroups. After consulting with disk drive experts, we find that low-power disk modes can significantly affect the spin up time and accessing the same track for a long time can accumulate lubricant between platters and heads. As remediations, we can change the power management policy and data access patterns, which could reduce 33% logical failures, that is 18% total disk replacements.

3.2 Disk Degradation Process

For drives having physical faults, i.e., read/write head failures and bad sector failures as shown in Fig. 1, we aim to understand how they change from being healthy to having more errors and finally becoming failed. We compare the distributions of key SMART attributes from both failed drives and working drives to identify the starting points of disk degradations. Six SMART attributes are selected to calculate the dissimilarity (we use Euclidean distance and Mahalanobis distance) of each SMART record with the failure record for every drive with physical faults. Fig. 2b and 2a show the dissimilarity curves of drives which have the full degradation process in these two failure types. We explore these results to model disk degradations.

4 CONCLUSIONS

The contributions of this work are significant. The proposed approach provides a systematic way to discover disk failure categories, pinpoint failure causes and model disk health degradation automatically, which enables us to predict disk failures of different types more accurately, and apply cost-effective disk recovery and data rescue techniques to assure storage systems are highly available.

REFERENCES

[1] Lakshmi N. Bairavasundaram, Garth R. Goodson, Shankar Pasupathy, and Jiri Schindler. An Analysis of Latent Sector Errors in Disk Drives. In *SIGMETRICS, 2007*.

[2] Song Huang, Song Fu, Quan Zhang, and Weisong Shi. Characterizing disk failures with quantified disk degradation signatures: An early experience. In *IISWC, 2015*.

[3] Ao Ma, Fred Douglis, Guanlin Lu, Darren Sawyer, Surendar Chandra, and Windsor Hsu. RAIDShield: Characterizing, Monitoring, and Proactively Protecting Against Disk Failures. In *FAST, 2015*.

[4] Kashi Venkatesh Vishwanath and Nachiappan Nagappan. Characterizing Cloud Computing Hardware Reliability. In *SOCC, 2010*.