Disk Drive Failure Factors

Jeremy Langston
School of Electrical and Computer Engineering
Tennessee Technological University
Cookeville, Tennessee 38505
Email: jwlangston21@tntech.edu

***Abstract*—Modern disk drives have become highly complex with increasing demands on high throughput, high reliability, high capacity, and low cost - most of which are typically mutually exclusive. This paper takes a look at the reliability aspect of disk drives, and, more specifically, factors that cause a drive to fail. Failures fall into three main categories: hardware, software, and process. Many of the frequent factors are given, focusing on data thrashing, disk age, and hot spots.**

I. Introduction

Disks fail. Nearly all system administrators have seen this many times. Why do they fail? Hard disk drives are highly complex whose design and operation crosses over many disciplines. The main aspects of a drive are the physical drives themselves, how they operate on the software level, and the human interaction or process views. There are many components to a disk system, including the hard disk itself, disk controller, connections, and I/O system bus, all of which affect overall reliability of a hard disk. While most of the disk system is electrical and computer oriented, a close interaction is made with the mechanical side in rotation of a disk's platters and movement of the heads.

A high level view of disk reliability comes in on the process level. Here, process refers to the procedures, configurations, and external factors used and performed on a system - and typically done by a human. This is included as procedures and configurations will always be prone to error due to human interaction. A. Brown et al discusses the need for including this factor in reliability analyses and benchmarking [1]. Process factors range from misconfigurations to natural disasters.

The main focus of this paper is on software level factors. Software factors are distinct from hardware in that they are logical, or virtual. This essentially means software factors are data-oriented. However, in most cases they directly affect hardware either through bit errors or degradation of components. Several of the main software factors are discussed. A more in-depth analysis is given toward how locality of the data affects reliability, as well as age and hot spots. These analyses are formed from previous work from either experimental or real world data. Merely focusing on what affects reliability does not lead to a complete analysis. The impact of the failure both for performance and uptime should be counted, whether it is recoverable or not, and the frequency at which a factor occurs that most critically affects a system.

Between the three main aspects of disk failures, physical factors rank the lowest in occurrence. Disk reliability has come a long way. Enterprise level disks (e.g. SCSI) hover around 1,500,000 hours as an mean time till failure, MTTF. However, the overall physical reliability must take into account that, at least in an enterprise setting, many disks are used in a single system, driving down the MTTF drastically. Software failures rank second and process oriented failures rank first, both on frequency and overall impact.

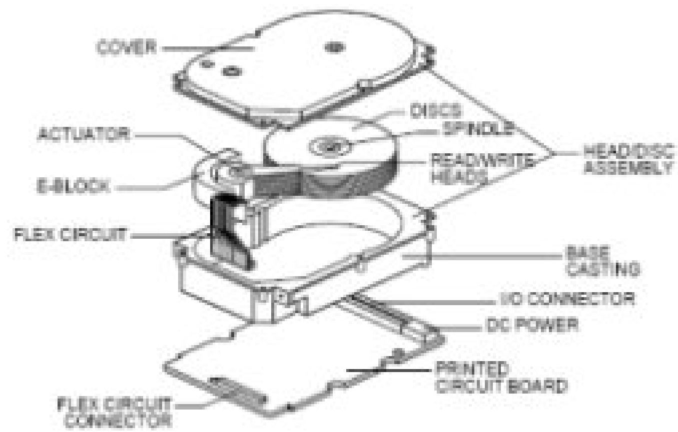

Fig. 1. Typical Disk Drive Components, from D. Anderson.

The remainder of this paper is broken up into the following sections. Sections 2 and 3 overview the many types of physical and process factors involved in disk failures. The focus of this paper is Section 4, depicting the types of software factors involved; in particular, locality of data, hot spots, and age. Section 5 concludes the paper.

II. Physical Factors

A hard disk has many different physical components designed by multiple disciplines. These aspects vary from electrical, computational, mechanical, mathematical, industrial, and business. Take into account a typical disk. Such a disk has the following components and subsystems, most seen in Figure 1: spindle, discs (platters), lubricants, actuator, heads, circuitry, controller, power supply, and casing [15]. From there, the disk is connected to an I/O bus via an I/O interface to a controller.

In brief, the disk reads and writes data via heads electromagnetically. The actuator arm move the heads to different sectors on the platters. A thin film of lubrication is deposited by the manufacturer on the platters to allow the head to move more readily without grinding into the surface of the platter.

Initially the head is at rest at a home position off of the platter (less possible damage due to vibration/shock), or in the center of the platter (more speed as the head only has to move half the distance at maximum compared to off platter positions). As the spindle builds speed to allow for reading/writing of the attached platters, a cushion of air is formed between the platters and the heads. This air cushion is referred to as the head flying height. A higher height leads to errors in writes and reads whereas a low height can lead to the head hitting (and crashing) into the platter, causing phyisical damage to both the head and the platter. It is easy to see this through an analogy given by [9]: "if the size of the flying head slider were equated to the wingspan of a Boesing 747, the 747 would be flying less than one inch above the ground at 600 mph". The nominal head flying height is between 22 and 23 nm [14]. As listed in Table I, the flying height is affected by humidity, temperature, and air pressure/altitude.This characteristic is further affected by design errors, manufacturing defects, age, and the model/manufacturer/vintage (these factors are much more broad, however). Humidity and air pressure also affect the magnetic characteristics of the drive, but at the extremes.

Head and platter life is further affected by power failure and dirty environments. The effect of power failure can be seen by what disks do when turned off. The actuator arm has a magnetic latching circuit that pulls the head to the side for avoiding vibration and shock affects. If the head is currently in contact with the platter, the head will scrape across the platter. Power cycling, the act of turning off and then back on, a disk has been found to negatively affect reliability [6]. Also, even a speck of dust can cause complete failure of a head, scratching the surface while the spindle is spinning.

Disk drives have many electronics components associated with them. Not only internally, but also external. Internally there are heads, servo loops and controllers for the actuator, read/write electronics, spindle motor, cache, and interfaces. Externally are connectors and cables, controllers, and system busses [9]. All are subject to interference from electromagnetism and radiation [3] and problems from heat. Symptoms range from bit errors and transport errors, to total disk failure.

In the case of a multiple disk system, the configuration plays a role. RAID configurations allow for both performance and reliability improvements, with different versions striking different balances between the two. Drive mirroring is the most straightforward approach to increasing overall data reliability. Here, the data is saved in multiple locations in anticipation of one of the disks failing. For a single mirror of the data, only one failure can be tolerated. If both the main and mirror disks fail, the data is considered lost, excepting some procedure to recover this data in some other fashion. A second configuration is automated/manual backup to a separate location. The backup may be a full backup or a checkpointing variation in which changes to the original backup are saved.

It should be noted that there are two main types of disks mass produced: SCSI and ATA. In [15], the commonalities and, more importantly, the differences in these two disk types are discussed. SCSI drives are often referred to as enterprise level drives as they are built for higher reliability and throughput needed for enterprise applications (e.g. patient database for a group of hospitals). ATA drives, personal, are typically budget, commodity drives that sacrifice reliability and throughput for cost. The SCSI drives implement many mechanisms to ensure data integrity that aren't necessary for common user applications. The majority of these mechanisms are implemented in hardware.

Factor	**Characteristic**
Humidity	Head flying height
Temperature	Viscosity of lubricants, head flying height
Electromagnetic and Radiation Interference	Bit errors
Vibration/Shock (internal and external)	Head crashes and platter damage
Manufacturing Defects	All physical aspects of the drive and I/O system
Power Failure	Lead to corrupt data, head crashes, and platter damage
Dirty Environment	Head crashes and platter damage
Damaged Cable	Transport errors
Damaged Controller	Transport errors
Age	Physical breakdown of materials (e.g. lubricant)
Design Errors	All physical aspects of the drive and I/O system
Air Pressure/Altitude	Head flying height
Configuration	Redundancy or reduced MTTF (e.g. RAID)
Type of Drive	Different requirements (e.g. SCSI vs. ATA)
Number of Platters/Heads	Decreased disk MTTF due to higher component count, seek time
Manufacturer/Vintage Drive Model	Different requirements, materials, firmware, etc.
Power Cycles	Spindle and actuator wearout

TABLE I
PHYSICAL FACTORS

III. PROCESS

Installation, management, and removal fall under the category of process factors. These are the configurations and procedures performed and devised, typically, by a technician or engineer, and are prone to human error. A list of generalized failure factors is given in Table II. Human factors play a major role in the reliability of a disk [1]. Brown et al reports that human error can account for roughly half of all outages in mid to high range environments. It is difficult to account for human error as each person is different, each with certain backgrounds. A system run by an expert in the field will be much more reliable than a novice running the same system. This can be further compounded by the ability of a technician to learn from mistakes at different rates. Some people catch on to a concept much quicker than others. Also, systems have varying levels of required knowledge and experience. A simplistic example is a system using a command line interface versus a system using a graphical user interface. This can affect the time it takes to configure or repair some aspect of the system.

Process factors can be further broadened upon to include external factors impeding on the reliability of a disk. External factors include power outages due to brownouts, powerline damage, etc. A power outage can cause direct damage to a disk by causing the head to slap into the platter and scrap across to its home position. Not only that, without a backup power source, data currently in transport will be affected and corrupted. Attack on the data also plays a part. The disk system's suceptibility to destruction, theft, denial of service, and so on directly affects a system's reliability. Indirect factors include budget cuts, organizational failure, or loss of context. Many of these process factors are a concern for long-term data storage [5].

The ability of a disk system to support hot swapping increases overall disk data reliability. Hot swapping is the act of installing and removing components such as disks without bringing down the system and on varying levels of serviceability. While this does not change the individual disk reliability, it does increase the proactive reliability aspect of the system. As [4] points out, manufacturers heavily market hot swapping. This is because a significant portion of downtime is related to maintenance and maintenance can be expedited through hot swapping.

Obviously, process factors are far reaching and numerous. It would be impossible to enumerate each individual factor that may cause a decrease in reliability due to the many factors at play. For this reason, this area is left for further research.

Factor	**Characteristic**
Misconfiguration	Human error in setup of system
Planned Outages	Upgrades, side-grades, reconfiguration
Maintenance	Pre-emptive adjustments to a system (e.g. disk scrubbing)
Cascading Failure	Failover to a system causing a subsequent failure
Mistaken Erasure	Human error causes data to be lost
Outdated Media, Formats, Applications, and Systems	No way to respond to data outages
Loss of Context	In long term storage, data has no meaning (e.g. meta data is destroyed)
Attack	Destruction, censorship, modification, denial-of-service, theft
Budget	Lack of funds to respond to a failure
Organizational Failure	Supporting sponsor, administrators, vendors, or service providers deny support

TABLE II
PROCESS FACTORS

IV. SOFTWARE FACTORS

Software, or logical, factors are typically data oriented, but are expanded to take into account the usage of a disk. There is a often a strong correlation between software and hardware factors. In cases such as high seek counts, this number directly affects the underlying hardware. This is covered further below. However, this is not always the case. Disk device drivers influence both system stability and data integrity. Drivers also account for an overwhelming amount of bugs in operating system code. A study [12] was done on 21 different versions of the Linux kernel. They found that not only did the millions of lines of driver code increase, but the percentage of driver code increased (to 70% of all code). After error testing, their results revealed that driver code has a relative error rate that is up to 10 times that of the remaining kernel code.

Many of the main factors at this level are given in Table III. A recurring error here is a latent sector error. Latent sectors errors are bit fluctuations that change the meaning of the stored data, but aren't immediately apparent. Consider a data warehousing application in which the data is stored and very rarely, if ever, accessed. If a error occurred in the data, it would not be seen until that particular data was read. Such is the case in the failure analysis given in [2] at the Internet Archive. Bairavasundaram [7] gives a comprehensive analysis of these errors. Using a disk sample size of 1.53 million over a 32 month period, they found that there were over 53,000 disks with latent sector errors (3.45%).

Factor	**Characteristic**
Locality of Data/ Thrashing	Amount of movement the actuator arm must make
Seek/Access Count	Over-utilization
Illegal Usage	Bus error, address error, etc.
Error Handling	Recovery or avoidance of errors
Size	Latent sector errors
Age	All subsystems degrade, bit rot, sector errors
Stop-Start Count	Disk fatigue (spindle)
Duty Hours	Disk fatigue (all mechanical subsystems)
Hot Spots	Increased risk of usage errors, decreased risk of latent sector errors
Drivers	Illegal/faulty usage of drive

TABLE III
SOFTWARE FACTORS

A. *Locality of Data*

Consider the activity of reading a couple of books located in different places, such as at the library and at home. If you were to read a page from the book at the library, stop, walk home, read a page from a book at home, stop, walk back to the library and continue for the entirety of the books, not only will you have spent unnecessary time in transit, but you will be quite tired. This is analogous to the locality of data on a disk. For each separate access to data on a disk, the actuator must move the head and must wait for the spindle to put the

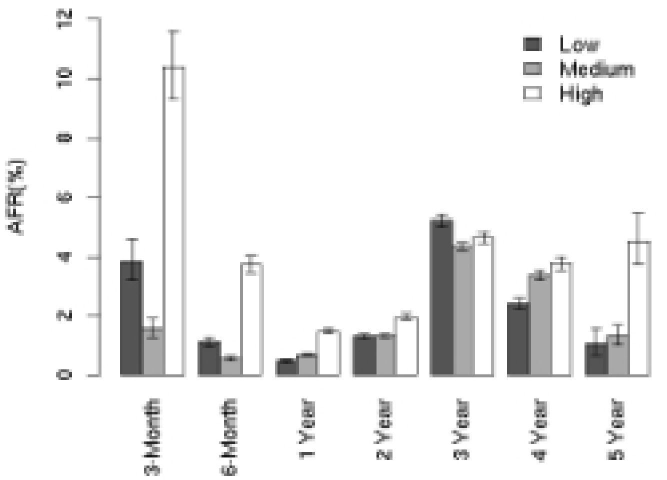

Fig. 2. Utilization AFR findings from E. Pinheiro.

correct sector under the head. Continually switching between two or more locations on a disk is referred to as disk thrashing.

The question is the impact thrashing has on the disk overall. Clearly, performance will suffer. Early disks had average seek times over 25ms, which is the average time it takes to move the head to the designated track and sector and begin reading [15]. While these seek times have been driven down considerably (less than 5ms), constant thrashing will incur this seek time to occur very frequently, thus causing frequent delays. The three biggest ways to reduce this is to group related data, add a level of cache, and rewrite applications. Cache works by moving a certain amount of data from a disk to a faster memory component, such as RAM. However, if the memory locations are not spatially or temporarly local, cache will not help. In this case, movement of the data or rewriting the offending application is in order.

Reliability is definitely affected by disk thrashing. However, the extent at which it is affected has different views. The common expectation is that thrashing, and thus utilization, is strongly correlated with failures. However, a report given by Pinheiro et al [6] finds that this is not necessarily the case. Using a pool of over 100,000 disks at Google, Inc., Pinheiro groups the disks by age and then determines the annualized failure rate caused by utilization. As seen in Figure 2, only very new and old disks were strongly affected by utilization. They gave an explanation that follows the survival of the fittest theory.

B. Hot Spots

The notion of a hot spot is used to describe a sector or portion of a disk that is frequently accessed. Hot spots are common in most applications with vast amounts of data that are rarely used, but some with a high access count. An example of this is in a data archive in which some file gains great attention. There is also the case for operating system swap files. Swap files are basically blocks of data that are frequently or recently used, but put to disk from main memory in order to load new data to main memory. In some OS's, the swap file may be located on the outer track of a platter. Outer sectors/tracks are the fastest as they are spinning the fastest (in relation to the head at a constant spindle speed) and the sectors are packed closer. In this case, the outer tracks would be accessed any time the swap file is used.

On its own, hots spots do not affect performance. Reading the same location repetitively does not increase throughput. However, as discussed previously, disks often use some form of caching. This is the application caching was most meant for. Caching allows for faster access to frequently used data by copying data from a slower medium (in this case, a disk) to a faster medium (solid state memory such as RAM). There are also many levels of cache. All modern hard disk drives implement some amount of onboard cache memory to make the most use of rotational cycles taking into account spatial locality of data. (Data in neighboring sectors are often related.) After onboard disk cache is main memory, then the levels of processor cache.

Since a sector is frequently being accessed, disk data is affected, both positively and negatively. Ignoring the effects of caching, many accesses, reads and writes, to a location increase the chance an error will occur due to failure or illegal usage. Consider the probabilities in having a driver error as discussed above. Each access has a certain probability. Continual usage of the driver will continually increase the probability an error will occur due to a disk driver. However, hot spots are very resilient to latent sector errors by nature. Since a latent error is one that is not immediately visible due to a low frequency of accesses, hot spots drive down the frequency of these errors.

C. Age

The age of a disk has many connotations. Age can be defined by the time since manufacture and the time spent in-service, or duty hours. Each meaning is different, but related. Time since manufacture directly deals with the hardware. Mechanical aspects of the disk degrade with time, such as lubrication decay and electromagnetic bit rot. An easy way to differentiate this aspect is to compare two drives originally loaded with identical content, but one is brand new and the other is 100 years old - will they contain the same data? In optical discs, CD-R, DVD-R, etc., manufacturers list a life expectancy of 100 to 200 years, but may dip below 20 years [16]. In-service time is different in that the disk is actively used. The more revolutions a spindle, for example, spins, the better the probability that either the servo motor becomes damaged. The line between hardware and software, and even procees, is blurred beyond recognition as it crosses all aspects and is not discussed further.

The performance of a disk is relatively unchanged by age. Research performed for this paper has not uncovered any solid evidence of a correlation between disk age and performance. However, reliability is quite a different matter.

Users have the tendency to think drive failure and drive age are linearly related. In reality, this relationship is more complex. Numerous papers [6], [11], [7] note the occurrence of a bathtub curve in age versus disk failures. The two ends of the bathtub curve denote relatively new disks and older disk have higher failure rates. Older disks are expected to have a higher failure rate as previously discussed. Higher failure rates in new disks seem counter-intuitive: if a disk is brand new, it should be as close to perfect as it ever will be. And this is true, disks do not spontaneously grow more adept to age-related failures. The bathtub curve represents average failure rates. A disk is not counted both as failing early in its life and not failing later. On average, disks have a higher chance of failing within the first 3 to 6 months than they do from 6 to 18 months. This is known as the infant mortality effect. Disk infant mortality is due to manufacturing errors [11]. The manufacturers implement a burn-in process to proactively catch the defective disks before making them available. Due to business concerns, the burn-in time is limited for SCSI drives, and even moreso for ATA budget drives. The infant mortality effect is of concern for systems in which uptime is very important. This led to the International Disk Drive Equipment and Materials Association proposing a more detailed view of MTBF (mean time before failure). Vendors and manufacturers are encouraged to display the MTBFs for their drives from 0-3 months, 3-6 months, 6-12 months, and the remainder of their disk life.

V. Conclusion

This paper has outlined the many aspects of hard disk failure factors, giving many examples of common factors. The focus was software-related disk failures, more specifically disk thrashing, hot spots, and age. Only thorough understanding of failures and their causes will lead to fixing the source. Proper understanding may also lead to the ability to benchmark for reliability and predict failures. SMART is an attempt at failure prediction, but falls too short to be relied on completely (reports of SMART missing more than 36% of failed drives can be found in [6]). A roadblock in front of enhancing disk reliability is in defining failure. End users have a different definition of failure compared to manufacturers, making failure analysis a difficult and ambiguous task.

References

[1] A. Brown, C. Chung, D. Patterson, "Including the Human Factor in Dependability Benchmarks", University of California at Berkeley, 2002.

[2] T. Schwarz, et al, "Disk Failure Investigations at the Internet Archive", whitepaper, Internet Archive, San Francisco, CA, Feb. 2007.

[3] E. Vargas, "High Availability Fundamentals", Sun BluePrints Online, Enterprise Engineering, Sun Microsystems, Inc., 2000.

[4] J. Zhu, J. Mauro, and I. Pramanick, "R-Cubed (R^3): Rate, Robustness, and Recovery - An Availability Benchmark Framework", Sun Microsystems, Inc., 2002.

[5] M. Baker, K. Keeton, S. Martin, "Why Traditional Storage Systems Don't Help Us Save Stuff Forever," *Proc. of 1st IEEE Workshop on Hot Topics in Sys. Dependability*, HP Laboratories Palo Alto, June, 2005.

[6] E. Pinheiro, W. Weber, L. Barroso, "Failure Trends in a Large Disk Drive Population," *Proc. of the 5th USENIX Conf. On File and Storage Tech.*, Google, Inc., Mountain View, CA, 2007.

[7] L. Bairavasundaram, et al, "An Analysis of Latent Sector Errors in Disk Drives," *Proc. of SIGMETRICS'07*, University of Wisconsin-Madison and NetApp, 2007.

[8] G. Hughes, et al, "Improved Disk Drive Failure Warnings," *IEEE Trans. on Reliability*, UCSD, La Jolla, CA, 2002.

[9] R. Kaseta, "Ruggedized Disk Drives for Commercial Airborne Computer Systems," Miltope Corporation.

[10] C. Hillman, "Reliability of Hard Disk Drives (HDD)," whitepaper, CALCE and University of Maryland, College Park, MD, 2003.

[11] Q. Xin, T. Schwarz, E. Miller, "Disk Infant Mortality in Large Storage Systems," *Proc. of 13th Annual IEEE Intl. Symp. on Modeling, Analysis, and Sim. of Comp. and Tele. Sys.*, Atlanta, GA, Sept. 2005.

[12] A. Chou, J. Yang, B. Chelf, S. Hallem, and D. Engler, "An Empirical Study of Operating System Errors," *In Proceedings of the 18th ACM Symposium on Operating System Principles (SOSP '01)*, Oct. 2001.

[13] B. Schroeder, G. Gibson, "Disk failures in the real world: What does an MTTF of 1,000,000 hours mean to you?," *Proc. 5th USENIX Conf. on File and Storage Tech.*, San Jose, CA, Feb. 2007.

[14] A. Khurshudov, P. Ivett, "Head-disk contact detection in the hard-disk drives," IBM Storage System Division, San Jose, CA, 2003.

[15] D. Anderson, J. Dykes, E. Riedel, "More than an interface - SCSI vs. ATA," *Proc. of the 2nd Annual Conf. on File and Storage Tech.*, March 2003.

[16] F. Byers. (2003) *Care and Handling of CDs and DVDs: A Guide for Librarians and Archivists* [Online]. Available: http://www.clir.org/pubs/reports/pub121/contents.html.