

Reliability and Security of D2D Backup Storage

Systems using SATA Drives

Dr. Gordon F. Hughes

Associate Director
Center for Magnetic Recording Research
University of California, San Diego
gfhughes@ucsd.edu

Report 2004-01

March 2004

The Information Storage Industry Center

Graduate School of International Relations and Pacific Studies
University of California
9500 Gilman Drive
La Jolla, CA 92093-0519
<http://isic.ucsd.edu>

Copyright © 2004 Gordon F. Hughes



University of California, San Diego

Funding for the Information Storage Industry Center
is provided by the Alfred P. Sloan Foundation

Abstract

Magnetic tape has long been used to backup computer information, customarily during dedicated, low use, 'overnight windows' when tape backup programs could be run without interfering with user applications. Over time these windows have shrunk to near non-existence due to the globalization of business through the Internet and the World Wide Web. While tape system speeds have accelerated and tape capacities have increased, they have not keep pace with the demand for shorter backup windows, quicker restoration requirements, and the rapidly escalating volume of disk drive data being backed up. At the same time, the cost of PC disk drives ("ATA" computer interface) has dropped to be competitive with tape, and disk-to-disk (D2D) backup has become popular, particularly using the new Serial ATA (SATA) PC drives. A D2D system can run at the full speed of disk, and can use the higher capacity of PC disk drives (up to 400 GBytes today), because backup is primarily serial data storage not needing the high random access speed of the enterprise storage systems being backed up (which get high random access speed by running many smaller capacity disks in parallel). The current interest in SATA D2D backup is of concern because these are PC drives sold with narrow profit margins that limit drive reliability.

This paper addresses the fact that the reliability of SATA PC drives is inherently lower than enterprise-class drives (SCSI and Fiber Channel "SCSI/FC"). Methods are proposed for SATA storage system designers to achieve high system level reliability by requiring appropriate SATA drive reliability testing, by increased RAID redundancy, and by system management of drive failure warnings. A method is proposed for maintaining user data security in removable D2D archival drives.

Reliability comparison of Magnetic Tape to SATA and SCSI disk drives.

Tape is designed for archival backup, and has data preservation standards (store reels vertically, control temperature and humidity), which includes periodic exercise (run tapes end-to-end periodically and check for read errors). This avoids and/or tests for several problems, like dropouts, tape wrap sticking, and data print through. Tape archival reliability is a highly mature technology, proven over the last half-century of archival tape use. There simply is no justification for assuming that disk drive archival reliability is comparable. Unlike storage device performance specs, reliability claims are easy to make and hard to prove.

Disk storage has not historically had an archival role, and disks normally are designed for only a five year service life. This includes stored data lifetime as well as electronic and mechanical reliability. Disk magnetic media is designed to retain data against thermal decay for five years (with 100% margin, i.e. ten year nominal MTBF). Many PC drives are designed for daytime office use, not for 24x7. The market lifetime of PC drive products is closer to six months than five years, so long term field reliability data isn't available. Some ATA drives are only partially flaw marked during manufacturing final test, because necessary test hours are not available given the manufacturing test floor space and sales profit margins. Enterprise class SCSI/FC drives are designed and tested for high reliability under heavy service duty, and fully tested and flaw marked in their manufacturing process, making their cost higher. Their product life is longer, and field reliability failures are monitored for design correction and improvement. For SATA drives to have the same reliability at SCSI/FC drives (requiring similar design considerations and testing cycles), they would also have to carry nearly the same price, because the same basic technology is used in all drives. The differences are in engineering reliability design, test, and performance margins. For example, SCSI/FC drives have lower bit areal density than ATA and SATA, to provide more margin against read errors.

What new failure modes can be expected in archival disk drives? Drive failure analysis is a mature engineering discipline with few mysteries, but drives do have dozens of failure modes which all need to be considered. Archival disks will probably be stored removed from storage systems and unpowered, perhaps in remote sites for disaster protection (like tape). Load-on-the-fly drives may avoid long term stiction issues that contact stop-start drives would have (their heads sit on disks when not spinning). Spin motor bearing issues also suggest power-off archival storage. Disc lube migration and contamination suggest on-off power cycles be minimized. Leaving drives spinning also has failure exposure to "fly stiction," if the heads are left at one track position for long periods.

There is also the thermal decay issue: modern drives are designed for five-year data life against thermal decay (ten years is the typical nominal design, for 2X MTBF margin). The size of each bit at 60-100 GBytes/platter is so small that simple Boltzman kT thermal energy at room temperature slowly disorders bit "0" vs. "1" magnetization states, turning stored data bits into magnetic noise. It's a hard physics-based limitation and the subject of major technology conferences.

Disk drive data security is controlled when physically inside storage systems. Removed backup or archival (or discarded) drives raise new security issues. A third of after market used drives contain unerased user data, and RAID striping does not necessarily avoid the security risk. ATA drives have standard security commands which should be used to satisfy data security concerns.

Here is a D2D SATA drive reliability and Security proposal for discussion:

SATA Storage System Study

- The SATA storage industry (system and drive makers) should undertake a systems study of design factors for backup and archival disc systems, perhaps coordinated by storagenetworking.org (SNUG). System designs should account for Pareto lists of major drive failure modes.
- Life cycle cost will be an important factor to study, because storage system administration costs far exceed drive hardware costs, even at SCSI/FC drive prices. Lowered drive reliability could impose a cost burden exceeding the lower SATA drive hardware costs.
- A new class of “Enterprise SATA” drives might emerge at a higher price due to the factors discussed below, but still allowing the higher capacity per drive of ATA and only slightly higher cost.

SATA drive design requirements:

- Require that SATA drive models undergo standard SCSI/FC reliability tests, including design verification testing, reliability demonstration testing, and design maturity testing. The reliability test levels may be balanced against the drive sales price, but reliability claims must be verified.
- Require drive manufacturer final test to include flaw marking of all drive data location, with a specified number of data writes and reads. Drive customers may choose to do part of the flaw marking in storage systems via drive internal test commands, to lower the price of mature SATA drives.

SATA Storage System design

- RAID designs should insure full system performance with one drive per stripe failed (double parity).
- SATA storage server software should periodically read the SMART failure prediction warning flags from drives. As a minimum, a warning drive’s data should be copied onto a hot spare. Additionally, system designers should consider reading drive SMART attribute data directly, and implementing the CMRR "Smarter SMART" algorithm in server software, which can greatly improve SMART warning accuracy.
- Drives with marginal SMART performance should be copied onto spare drives. If the drive subsequently fails (as it self-predicted it would), the system reconstruction overhead and performance burden could be avoided.
- Archival drive storage systems should store system access control data on the drives themselves, so they will work properly when remounted in a system, such as a RAID. Example: Access Control Lists with world-wide-names of allowed users (in the SCSI drive specs).
- Drives to be dismounted for archival storage should be locked by the storage server software with a secure 32-byte password, for user data security (a standard ATA password command).
- Drives pulled from systems should be Secure Erased for user data security (a standard ATA command)

Archived drives:

- Each drive should be mounted in a drive tester at least once a year, to run drive data integrity self tests, to run SMART drive internal failure prediction tests, and to have drive data exchanged with another drive to avoid the thermal decay problem. The testers should be able to unlock and relock drives, and able to mount a full RAID stripe for reconstructing the data from a failed drive.