Storage Enclosure Reliability:

Analysis of Various Enclosure Configurations With Implications on Data Reliability and resulting Backup applicability

White Paper

By James Wiebe, CEO, WiebeTech LLC

Abstract: This paper analyzes the failure rates of various types of external storage enclosures, so that users can understand why certain types of enclosures are more reliable than other types. After analysis, recommendations are made as to types of enclosures which are more favorable for selection as data backup devices. As a result, users may purchase enclosures with a greater understanding of what can go wrong, and why. Users will also be able understand why certain types of enclosures are suitable as part of a backup strategy, and others are not.

I am grateful for the assistance of Bill Head in the editing of this paper.

I am also grateful for editing assistance from another individual who shall remain anonymous.

Send your comments to: james@wiebetech.com

1. Introduction

Hard drives are the long term memory of your computer system. Virtually every computer uses an internal hard drive as a repository for the operating system, applications, and user data. Hard drives are also used outside of a computer, usually within storage enclosures. *External storage enclosures* containing hard drives can be used for all of the same reasons as internal hard drives (e.g. storing the operating system, saving programs, and saving user data) but enclosures have the benefit of expandability and portability, and as a result are also commonly used for data transport and data backup.

The failure of a hard drive within or attached to your computer system (via an enclosure) is an eventual certainty. Consider the following statement:

> "If you use computers long enough and often enough, you will eventually have to deal with a failed drive. It really is only a matter of time and luck. When this occurs with a drive that is reasonably new, you'll want to have it repaired or replaced under warranty. Unfortunately, most people couldn't tell you anything about their hard disk's warranty other than its length--and in some cases they later find out that they were mistaken about even that! The vast majority of hard disk users don't read the "fine print", and many manufacturers don't exactly make it easy to find even if you *want* to read it."
>
> *– Charles M. Kozierok, author, The PC Guide.*

Definitions #1: Here's a quick timeout to explain some nomenclature: **Hard drives** are devices with rotating magnetic media, which are made in varying physical dimensions, varying logical capacities and varying interface standards. Hard drives are commonly found inside computer systems and are also commonly found within **enclosures**.

A **storage device** refers to the collective idea of either a *hard drive* or an *enclosure*, which contains one or more hard drives. For instance, if you say that your *storage device* has gone bad, I'll ask you: 'Are you referring to the one within your computer (which might also be called a *hard drive*) or are you referring to your *enclosure* (which happens to contain one or more *hard drives*)?'

Hereafter, we'll substitute the single shorthand word "enclosure" when we are referring to any external storage device.

When the failure of an external storage device occurs, it's important to distinguish whether the enclosure has failed, or a hard drive inside it. If only the enclosure fails, it's a simple matter of moving the hard drive(s) to another enclosure or repairing the enclosure, after which the hard drive(s) contained therein will once again be functional. This is a critical concept for discussing the overall reliability of hard drives, which shall be covered in later chapters.

Hard Drive *External Storage Enclosure* *Enclosures have hard drives inside*

Figure 1: Differentiating Hard Drives from Enclosures

Enclosures are made in many different configurations. The very simplest type allows a user to attach a single hard drive within the enclosure to a computer via a cable. More advanced configurations allow users to attach several hard drives, all contained within one enclosure. Certain kinds of enclosures are very reliable, while others offer very high performance but are relatively unreliable in comparison, and would make poor choices as a backup system. The key to making a confident purchase decision is to gain a basic understanding of the type of technology used within the enclosure, along with an understanding of the kind of performance and reliability compromises that you are willing to make.

Consider the value of the data stored on your hard drive or in your enclosure. If you are like most people, you have many files that you could not stand to lose. This is true for all storage devices, but especially for enclosures, because of their common usage for data transport and for data backup.

I've learned that people place significant value on the files that are placed on their storage devices, yet often have a poor understanding of the reliability of those storage devices. Most individuals also lack any plan or useful strategy for backing up their important data. As a result, significant heartache and financial loss occur when a user's hard drive is damaged or ceases operation.

The purpose of this White Paper is to explain the risks and failure modes of various kinds of storage devices, with a specific emphasis on enclosures, and as a result encourage users to more thoughtfully select enclosures based on a realistic analysis of failure probabilities and recoverability scenarios. Since the eventual failure of a hard drive is a statistical certainty, users will benefit tremendously from an understanding of why certain kinds of enclosures are more reliable than others, and how this greater reliability offsets other features such as ease of reparability or performance.

2. Hard Drives described in a little greater detail

Remembering our definitions, a *hard drive* is what is inside virtually every computer system and every *enclosure*. Since hard drives are the basic element of our storage strategy, we need to spend a little time understanding what a hard drive really is.

Virtually all personal computers ship with at least one internal hard drive. Today's desktop computers usually use a 3.5-inch *SATA* hard drive. (3.5-inch refers to the size of the magnetic platter hidden within the hard drive. It is not an external dimension. 3.5-inch hard drives are actually 4 inches wide, 1 inch high, and about 5.7 inches long!)

Here is a picture of a typical 3.5-inch hard drive, with its cover removed. The platter is clearly visible, along with the head-actuator mechanism:

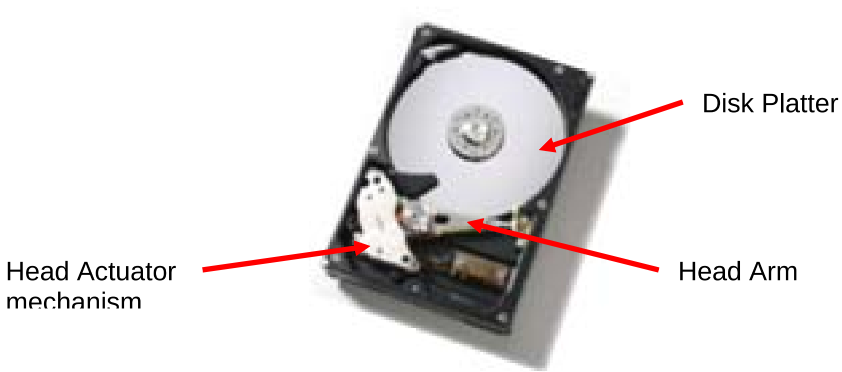

Figure 2: A common 3.5-inch hard drive, with cover removed

Prior to 2006, many computer companies used 3.5-inch *IDE* hard drives as their internal hard drive, but 3.5-inch *SATA* hard drives have really taken over the market in the last year. The technology issues and history between *IDE* and *SATA* drives are somewhat important to this paper, and have been very slightly covered in two of my other white papers (specifically: FireWire Evolution and FireWire Evolution Redux, which may be found on the WiebeTech website under /media/whitepapers.)

Here are some differences between *IDE* and *SATA* drives which are worth noting:

- *SATA* drives feature transfer rates up to 3.0 Gb/sec, which sounds very fast, but is never realized as a practical real world sustained transfer rate.
- *IDE* drive transfer rates have a transfer rate of about half the speed of *SATA*, at best. However, in the real world, using real benchmarks, comparing one type of drive against another, the *SATA* advantage over *IDE* is much more slight.
- *SATA* drives are easier to attach and dock.

- *IDE* drives are limited to two per cable attachment. *SATA* drives are normally limited to one per cable attachment, unless a technology called Port Multiplying is used.

> Definitions #2: There's a lot of meaningless alphabet soup in the world of hard drives and enclosures. You really don't need to know what all these things stand for, and the exact abbreviations are easily Googled, so I will use a new kind of definition: an **IDE** hard drive is an old-fashioned hard drive which attaches to its host using an ugly or inflexible ribbon cable. A **SATA** hard drive attaches using a flexible, thin cable. Both types of drives are exactly the same size, but *SATA* is newer, faster, and easier to use. *IDE hard drives* will be around for many more years, but their unit volume has fallen off in comparison to the much newer *SATA hard drives*.

Figure 3: Comparing the interfaces of IDE and SATA hard drives

Notebook hard drives use a 2.5-inch platter, and are commonly used in… you guessed it: notebook computers. Notebook hard drives are available with ATAPI interfaces as well as SATA interfaces. ATAPI may be thought of as analogous to IDE; (big ugly ribbon cables) whereas as SATA is just plain. But in the world of notebook hard drives, the marketplace has not moved to SATA nearly as rapidly as it has in the 3.5-inch hard drive market.

There are other sizes of hard drives available, such as the 1.0-inch Hitachi MicroDrive, which is used inside Apple's discontinued mini iPods and many high end digital cameras. Toshiba has introduced a drive with 0.85-inch media, and its remarkably small form factor lends itself well to usage in PDAs and cellphones. The drive is similar in size to a big postage stamp. Apple's full-sized iPods use a 1.8-inch hard drive; a few notebook computers use 1.8-inch hard drives as well.

Hard drives with rotating magnetic media are at the core of all common high-speed storage devices. (If you'd like to discuss esoteric non-rotating solid state media with me, write me at james@wiebetech.com. If you'd like to discuss high-speed streaming-tape backup methodologies… you are reading the wrong white paper).

Hard drives are found inside all enclosures and also inside all personal computers. They come in all shapes and sizes, with several different interface types, and with widely varying capacities. The one thing all hard drives have in common is that they will eventually break down.

It is extremely important to understand that hard drives are very reliable devices, but they do fail with statistical certainty. The impact of statistical failure rates for a given individual hard drive will become clearer as we explore the varying types and reliability risks of enclosures in the next two chapters.

3. Hard Drive failure modes

A few years ago, a customer called me up to discuss a profoundly sad situation. He'd spent several weeks in a South American country, taking irreplaceable photos. All of the photos were stored on a *hard drive* within a WiebeTech *enclosure*. After his return, he'd used the enclosure without difficulty for a few weeks but had then suffered a breakdown of the hard drive within the enclosure. The data was irretrievable. We were eager to replace the defective hard drive, but this clearly would not solve the customer's massive problem of retrieving the data off of the defective hard drive. (Our warranty specifically disclaims all liability in respect to data recovery. Rational competitors have similar clauses within their warranties.)

The cost of data recovery is very high, and can range into the neighborhood of $4,000 for a single hard drive. (Unfortunately, there is no certainty that the recovery effort will be successful.) Several companies are in the business of providing this service, and do a quick and professional job at that task. An example of such a company is DriveSavers (800-440-1904, www.drivesavers.com).

When a hard drive fails, DriveSavers is tasked with a very difficult problem. The drive may require disassembly, repair, and reassembly of the hard drive in a clean room, followed thereafter by painstaking sector-by-sector rebuilding of lost or corrupted files. It's easy to understand why data recovery can be so expensive (and unpredictable)—it's technically demanding work, often done by hand, and there are no guarantees. It requires the use of multiple highly demanding skill sets, including clean-room technique, electronic repair, forensic sleuthing of damaged digital file structures… etc.

Hard drives can break for a variety of reasons, a few of which are:

a) Failure of drive logic / motor driver electronics. This is perhaps the easiest to repair. Just replace the affected board; and voila! The hard drive comes back to life.

b) Failure of the spindle motor. Requires disassembly and reassembly of the hard drive in a clean room.

c) Failure of the head actuator assembly. Requires disassembly and reassembly of the hard drive in a clean room; may also cause a head crash.

d) Head crash. The most dreaded failure mode. The extremely small magnetic head assembly contacts the surface of the disk, causing a portion of the media to flake or abrade off, causing data destruction on the magnetic media. The loss of magnetic media is not necessarily fatal to the data, for technical reasons involving complex error protection protocols recorded on the magnetic platters contained within the hard drive. However, it's still a really bad event.

e) Software failure within the host computer or within the enclosure's firmware. This one's a little awkward to explain. Basically, it's a polite way of saying that some sort of software problem fried your data. Historically, this has occurred frequently enough to gain notice and notoriety. This kind of abhorrent system behavior can result in the total destruction of the file structure on the disk. While not irreversible, it sure can be a pain. Worse yet, the destruction occurs through no fault of hardware or the user. An example of this kind of failure was data destruction that occurred in rare instances when Apple updated an early version of OS X and users had external FireWire 800 drives using a particular chipset with a specific firmware version. It is possible for a computer to 'crash' in such a way that it wipes out the directory or file allocation table (FAT) of a hard drive. This is a statistically remote failure mode, but it has happened in the past and may happen again at some point in the future.

   To really prove that a directory has been corrupted requires substantial skills on the part of the user. While it is common to hear that 'the directory of a drive has been corrupted', so many erratic behaviors are explained with this catch all phrase that a certain amount of circumspect caution is appropriate. The drive may simply be broken in a way that simulates a corrupted directory.

f) Mishandling. Physical shock is never a good thing, especially when irreplaceable data and miniscule tolerances are mixed together. A common result is gross failure in the drive mechanism (such as a head crash, resulting in destruction of the media surface and all data contained therein).

Some of these failure modes are easily repairable, assuming access to spare parts, a clean room, and skilled personnel. (With slight sarcasm in the statement: 'easily repairable') The head-crash failure mode is particularly difficult, because of the hardware failure within the head / media structure compounded by the data destruction on the magnetic disk media.

Sometimes it is possible to recover corrupted data files on a hard drive using recovery utility software, such as ProSoft Engineering's excellent Data Rescue II (877-477-6763, www.prosofteng.com). WiebeTech has found that this recovery software is effective for many of our Macintosh customers. It is capable of examining the data out on the disk, and 'restitching' the destroyed data structure back together.

In the case of the customer with the failed drive and the irreplaceable photos, we recommended that he use the services of an independent data-recovery company. We sent our replacement-warranty hard drive to that company, along with the user's failed hard drive. They were able to completely restore his data, but at significant financial expense to the customer.

When failure occurs, the only rational strategy is to have prepared for it by having a backup available.

Prior to the failure of the hard drive, it is wise for the user to understand the various types of enclosures which are currently in use, along with their associated risk of failure. Certain enclosures are highly tolerant of hard drive failure, and will allow users to easily recover and even rebuild data on the fly. Other enclosures are especially intolerant of hard drive failure.

What follows is discussion various types of enclosures, along with explanations of why failures rates vary dramatically between them.

4. Types of enclosures

There are several popular enclosures designed for use with personal computers. I categorize them as follows:

A. Enclosures which utilize one fixed hard drive. Most products in this class attach to the computer via a FireWire, USB, or eSATA cable. All are commonly used for general processing and for data backup. This type of enclosure contains either a nonremovable 3.5-inch hard drive or a nonremovable 2.5-inch hard drive. The enclosures which are based on 3.5-inch hard drives are called ‘desktop’ drives, while the enclosures based on 2.5-inch hard drives are often referred to as ‘pocket’ drives and are more portable.

	
WiebeTech’s FireWire 800 / 400 / USB2 Desktop Drive, which uses a fixed internal 3.5-inch hard drive	*WiebeTech’s FireWire 800 / 400 / USB2 Pocket Drive, which uses a fixed internal 2.5-inch hard drive*

Figure 4: Examples of enclosures containing a single, fixed drive

B. Enclosures which utilize one removable hard drive. WiebeTech and others market external enclosures that feature a removable hard drive in a slide-out tray. This has significant advantages for backup (it’s easy to rotate a changing schedule of hard drives into the enclosure, for instance, on a daily basis) and also allows easy replacement of the hard drive in the event of failure.

Some of these enclosures are designed to allow attachment of *IDE* or *SATA* drives to FireWire or USB ports; while others are designed to allow attachment of SATA drives to eSATA ports. All are easy to use.

	
WiebeTech's TrayDock, which utilizes either an IDE or a SATA hard drive. This enclosure has both FireWire and USB ports on the rear. The hard drive slides out the front.	*WiebeTech's SilverSATA I, which utilizes a removable SATA drive. The enclosure has both eSATA and USB ports on the rear. It's easy to see how the hard drive slides out the front of this unit.*

Figure 5: Examples of enclosures containing a single, easily-removable drive

C. Enclosures which utilize multiple fixed hard drives in a non-redundant configuration. This type of external storage device is built using two or more hard drives, and contains an internal RAID controller which causes the multiple hard drives to appear on the desktop as a single volume.

Definitions #3: For the purposes of this paper, a **volume** is what appears either as a drive letter (in the case of a PC) or on the user's desktop (in the case of a Mac) when an enclosure is properly attached to a computer system. **RAID** is a common method of attaching several hard drives to a computer system in such a way that they all appear as a single volume. Two common types of RAID are striping and mirroring.

Striping (aka RAID level 0) is a RAID methodology which allows multiple hard drives to be accessed simultaneously, thus preserving the total capacity of the hard drives while increasing performance. The opposite of *striping* is **mirroring** (aka RAID level 1) which causes two identical sets of data to be written to and read from two hard drives. As a result, reliability is increased while total capacity of the hard drives is decreased by a factor of two. A **JBOD** is the opposite of *RAID*: Just a Bunch Of Disks. Really! In summary, *RAID striping* increases performance and preserves total capacity, while *RAID mirroring* increases reliability and causes total capacity to be cut in two, while a *JBOD* is all the hard drives without any *RAID* hardware or software.

LaCie (503-844-4500, www.lacie.com) has been dominant in this product area, marketing these drives as the Big Disk, Bigger Disk, etc. The lack of redundancy within this type of enclosure puts data at peril in the event of the loss of any one drive. A typical drive within this product category is the 2TB Bigger Disk Extreme, which is, to my best guess, actually composed of four 500GB drives.

This type of enclosure typically uses a *RAID* controller in a striped or a spanned configuration. (Spanning is like striping in that multiple drives are seen as one logical volume, but without the performance benefit. Spanned drives are written to and read from sequentially rather than simultaneously.) The host computer sees these enclosures as one great big drive, while the RAID controller contained within the enclosure itself is busy cutting and pasting data to and from the various hard drives. (Remember, there are multiple hard drives within the enclosure, even though the enclosure appears as one volume on the user's desktop.) This type of enclosure may be an effective choice for high-speed data manipulation, such as for video editing, but is a poor choice for backup, for reasons which will be more carefully explained in a later chapter. It is wise to use this type of product only when you have an effective and well-administered backup strategy for the data contained within the enclosure.

D. Enclosures using multiple removable hard drives in a non-redundant configuration. WiebeTech and others have produced a variety of enclosures, featuring multiple bays, removable drives, and individual drive interfaces for each bay. Examples of this kind of product include WiebeTech's TrayDock eSATA, which is designed for each individual hard drive to attach directly to a computer's individual eSATA port:

Figure 6: WiebeTech's TrayDock eSATA, which utilizes removable SATA hard drives with individual eSATA ports on the rear for each bay. 2 bay and 4 bay models are shown.

These types of enclosures are often referred to as JBODs—remember the definition? ("Just a Bunch Of Disks".) The failure of any one hard drive or controller will usually not cause the other hard drives within the same enclosure to fail, as each hard drive is viewed as an independent volume by the host computer system.

These drives can be RAIDed together using operating system or third-party software utilities. This is easily done in Windows XP Professional or in Mac OS X. After RAIDing is performed, the individual drives are combined into a single, larger volume, courtesy of the operating system RAID software.

The key advantage of this type of external storage system is the ability to use the different disk drives independently (JBOD), and to repair individual drives easily in the event of a failure. When used in a RAID 0 configuration created by software, however, the failure of any one drive will cause the entire RAID to fail. Physical repair of the product is easy (just replace the drive) although the recovery of the data within the RAID may be difficult to impossible. It is a poor idea to use this type of enclosure as a striped RAID in the absence of an effective backup strategy.

Another similar enclosure is WiebeTech's SilverSATA V storage system. It features multiple drives within one enclosure. A key feature enhancement is that all drives are connected via a single external eSATA connector. From the standpoint of the system, it is viewed as five individual hard drives; all of these hard drives appear as separate volumes on the desktop, even though all connections to all five hard drives occur over a single cable. The drives may be RAIDed using aforementioned operating system utilities or third-party software utilities or even through the use of a hardware-RAID host card (an expansion card for a computer that has built-in RAID management). Even with all of these RAIDing options, many users prefer to use the product as a hot-swappable JBOD. The core technology which enables much of the SilverSATA V feature set is buried within the SATA 2 specification.   On a practical note, SilverSATA V shares individual tray specifications with its line mates, specifically the SilverSATA I (previously noted in paragraph c., above) and SilverSATA II (noted in section G, below). Trays may be swapped between these compatible units, although this is not allowable if the trays are part of a striped RAID. (Removing any one tray from a striped RAID kills the data set for that entire RAID volume, making everything nonrecoverable, and that's bad.)

E. Enclosures which utilize multiple removable hard drives in a redundant configuration, specifically using RAID 5 technology.

Definitions #4: We've covered RAID 0 (striping) and RAID 1 (mirroring) earlier in this paper. If that wasn't complicated enough, RAID technology also allows for some other configurations which have attractive benefits, in particular RAID 5. When a group of three or more hard drives are configured as a RAID 5, the data is spread across all of the hard drives in such a way that the failure of any single hard drive within the RAID 5 will not cause data loss! This amazing feat is performed through the magic of mathematics, so that the resulting system is fault-tolerant, without being totally redundant. As a result, RAID 5 technology can be very large in capacity, very reliable, and relatively fast.

For a very, very quick explanation of RAID 5, consider the following mathematical equation: 3+4 = 7. If the '3' is stored on one drive, while the '4' is stored on a second drive, and the '7' is stored on a third drive, the loss of any one drive still allows us to reconstruct the original equation exactly. 3 + X = 7; X + 4 = 7; 3 + 4 = X. No matter where the data is missing, we have enough information to correctly solve the equation. Once again, we can move on…

WiebeTech sells a product (RT5) which, when properly configured and formatted, provides that the failure of any one drive within the enclosure will not cause the RAID data set to be lost. A new drive may be inserted in place of the failing hard drive, and the RT5 will automatically rebuild itself. The rebuilding may occur while the device is in operation.

Figure 7: WiebeTech's RT-5 and RT5x2 transportable RAIDs

This class of product is easily repairable. In the event that any individual hard drive fails, the offending drive may be replaced and the data set rebuilt without loss of data. The product also has the additional attribute of directly supporting RAID 5 with a built in hardware-based controller. The built-in controller is less susceptible to malware than software based solutions, and relieves the burden of RAID calculations from the computer's CPU. No operating system utility is necessary, and operation of the RAID is 100% transparent to the user and the host computer. As a result, the product will redundantly spread data across multiple drives, so that the failure of any one hard drive will not cause data loss.

The mathematics behind this ‘self-repairing’ feature push the reliability of this device up by one (or more) orders of magnitude. As a result, people and organizations with serious data safety issues must seriously consider this class of product for backup solutions.

The math and exact reliability information will be explained more carefully later in the paper.

F. Enclosures which utilize dual removable hard drives in an automatic mirroring configuration. WiebeTech sells two different products (RAIDTech and SilverSATA II) each of which feature dual hard drives. These products are designed to automatically mirror the contents of one hard drive to the other. To the host computer, these products appear as a single volume. Within the products, a dedicated RAID controller is continuously writing all data to both drives, identically. The failure of any one hard drive will not cause a loss of data.

Figure 8: WiebeTech’s RAIDTech and SilverSATA II transportable RAIDs

As a result, the reliability improvement benefit is similar to the RAID 5 devices which were previously discussed; perhaps mirrored RAID enclosures have even slightly higher reliability than RAID 5. They are also relatively inexpensive. When everything’s considered (performance, reparability, reliability, cost) this class of product is perhaps the finest choice of backup for individuals and small businesses. Their characteristics make them suitable for many applications and for high-reliability general purpose backup. A comprehensive backup strategy should still involve periodic backups to another drive, however. While a RAID 1 (mirror) offers excellent protection against a mechanical drive failure, it cannot prevent user error. If files are accidentally deleted, both hard drives are simultaneously affected.

The primary difference between the RAIDTech and SilverSATA II is that the RAIDTech is designed to connect via FireWire or USB to the host computer, while the SilverSATA II is designed to connect via USB or eSATA. Both are excellent choices, and the decision to purchase is often based on what type of host port is desired. Both of these products may utilize SATA hard drives; while the RAIDTech has the additional capability of being compatible with IDE and SATA hard drives.

At WiebeTech, we “eat our dogfood” by using a RAIDTech connected to one of our servers. Critical system information (such as our accounting data files) is contained within the RAIDTech. The failure of any one hard drive within the RAIDTech will not cause the RAIDTech (or the server) to fail, and if an individual drive failure should occur, we can rebuild the backup hard drive in real time. In addition to this relatively simple storage strategy, we make daily backups of critical files which are taken offsite.

5. Reliability Risks within enclosures

One of the most misunderstood aspects of hard drives is their reliability. Opinions of a particular drive's reliability are clouded by the manufacturer's MTBF (Mean Time Before Failure) information, brand reputation, application, damage events, host operating system, drive cooling… the list goes on. What is lacking is information as to how a hard drive will perform for a given user in their real world experience, over a reasonably long period of time. Manufacturers are loath to give out real world statistical information.

I'd like to break that pattern. Over the course of the last year, WiebeTech has shipped a large number of our ToughTech 800 enclosures. I'll cut straight to the chase, and let you know a closely guarded corporate secret: WiebeTech's ToughTech 800 storage device has a RMA ('Return Material Authorization') rate of 3.2%. Now you know: if you buy the ToughTech 800 from us, there is approximately a one-in-31 chance that you'll return it sometime in the next year.

But even this simple statistic is quite misleading. It certainly doesn't cover the actual failure rate of the product, because we allow returns for a variety of reasons: we have a brief period of time in which consumers may return products for any reason or no reason whatsoever; people change their minds; computers break, and sometimes the wrong failure culprit is identified; etc. We test returned products and often discover no problem whatsoever with the product. In that case, perhaps the user tried to run the product with an unsupported older operating system—we just don't know what the problem was. And once in a while, the product is actually broken. What really causes us grief is when a product is returned because the user didn't like the corporate color pattern on the cardboard box, and the product is unopened.

What is clear to us is that enclosures fail for a variety of reasons and with different failure rates, depending on internal 'architecture' of the enclosure. Here are many of the significant enclosure architecture issues to consider, which in turn affect the reliability of the enclosure:

a) How many hard drives are within the device? Every additional hard drive has the potential of decreasing the reliability of the product—we've seen this specifically within our own product line and heard anecdotal accounts regarding competitor's products as well. It only makes sense: if one hard drive has a failure rate of X% (where X is unknown… but we'll get into that shortly…) then an enclosure with two hard drives has a failure rate of 2 * X%. This assumes that the failure of any one hard drive causes the failure of the entire enclosure.

   As a consequence of this question, it is important for the user to understand how many hard drives are inside the enclosure which is being purchased. As of this writing, any enclosure with a capacity of more than 500GB is internally composed of two or more hard drives, which have been RAIDed together via striping or spanning. The fact that a RAID hardware component is within the enclosure may or may not be disclosed by the manufacturer.

b) Are failing hard drives easily replaced? If an enclosure has only one hard drive within it, this is rarely an issue. The entire unit is easily returned for service. If the enclosure has two or more hard drives, then this issue gains significance. The best enclosures allow each drive to be removed individually via a slide-out tray. This helps with maintenance issues and also with backup issues, as drives are easily removed from the enclosure for offsite backup purposes.

c) Is RAID 0 (stripe) or a drive-spanning architecture employed? It's important to remember that the objective of a RAID 0 is to maintain total capacity of multiple drives and improve performance, but at the expense of reliability. RAID 0 is a great choice for performance users; it is a terrible choice as a backup drive. If you are using a RAID 0–based enclosure as your backup volume, you should re-evaluate your backup strategy, because RAID 0 is less reliable than other choices.

d) Is RAID 1 (mirror) architecture employed? RAID 1 will cause the storage enclosure to make two copies of data. As a result, RAID 1 devices are good candidates for inclusion in your backup strategy.

e) Is RAID 5 (automatic error-correcting) architecture employed? RAID 5 allows systems with three or more drives to be configured so that the failure of any one hard drive will not cause the loss of the data set. The best RAID 5 enclosures allow hot swapping of a failed drive and rebuilding of the data set 'on the fly', so that the data redundancy of RAID 5 is regained. RAID 5 devices will continue to work as long as the failure is limited to a single hard drive. A failure of two or more hard drives (or the failure of the RAID controller) may still cause the loss of data. (Emerging technologies such as RAID 6 allow two drives to fail; the cost is greater overhead on the extra storage space required to create this redundancy).

f) What cooling methodology is used? Most enclosures employing a single hard drive are fanless, so they are very quiet. For anything more than one hard drive, we recommend an active cooling system—in other words, a fan. All WiebeTech products with two or more hard drives have fans in them.

g) What is the company's reputation for support and reliability? This is self-explanatory.

6. Statistical estimates of hard drive failures within enclosures

Users want to know two things: How likely is an enclosure to fail? If it does fail, how likely is total data loss? (For instance, if an enclosure fails, but the hard drive within it is undamaged, repairing the enclosure will cause complete and immediate recovery of all data.)

Let's start by recapping our hard drive configurations. As has been explained, users are likely to encounter the following configurations of enclosures:

- **a) Enclosures which utilize one fixed hard drive.**
- **b) Enclosures which utilize one removable hard drive.**
- **c) Enclosures which utilize multiple fixed hard drives in a non-redundant configuration**.
- **d) Enclosures using multiple removable hard drives in a non-redundant configuration**.
- **e) Enclosures which utilize multiple removable hard drives in a redundant configuration, specifically using RAID 5 technology**.
- **f) Enclosures which utilize dual removable hard drives in an automatic mirroring configuration.**

In order to make everything make sense, I have to briefly delve into some mathematical analysis of failure rates.

If math isn't your thing, skip this chapter and go straight to the summary (Chapter 7). What I am doing in this chapter is showing that certain types of enclosures are very reliable, and others are very unreliable. All of conclusions are summarized in the next chapter.

I will make some broad conservative estimates of reliability and probability of data recovery in the event of failure. The assumptions are as follows:

a) **Hard Drive Failure Rate** – A hard drive has a probability of failure of 2.0% per year. This is my gut estimate; it is not scientifically created or based off of any available information other than my years of experience in the storage industry and observing failure patterns. This is not the failure rate of the enclosure; it is the failure rate of the hard drive which is within the enclosure. If the reader delves into the published MTBF (Mean Time Between Failure) figures for various hard drives, he will discover little correlation between the published MTBF and the real world failure rate. This is because my failure rate (derived, in the best Malcolm Gladwell fashion, from his excellent book *Blink:* in my gut) includes real-world factors such as heat, shock, power supply spikes, all of which are carefully excluded from more rigorous MTBF calculations.

b) **Enclosure Failure Rate** – An enclosure itself, exclusive of the hard drives which it contains, has a failure probability of 1.0% to 4.0% per year per enclosure. These numbers are not exact; they are estimates and are subject to variables beyond the scope of this paper. (We have to make some simple assumptions as a starting place.) For enclosures which are designed to contain one hard drive, we will assume 1.0% per year. For enclosures designed to contain two hard drives, we will assume 2.0% per year. For even larger units (such as a 5 drive RAID) we will assume either 3.0 or 4.0% per year. Remember, this rate is for the enclosure, not for the drives contained therein.

   By splitting the failure rates of enclosures from the failures of the hard drives contained therein, we are able to create mathematical formulas which analyze the total failure risk of the entire sum of the enclosure and all of the hard drives which are inside of it.

c) **Hard Drive Failure Data Loss Assumption** – We will make another important assumption, which is that the failure of any hard drive causes unrecoverable data loss, unless that hard drive is part of a RAID with mirroring or RAID 5 technology. This is an assumption with a built-in flaw: a failure of a hard drive within a JBOD causes the failure of only that one hard drive, not the rest of the hard drives within the JBOD. For the purpose of this discussion, we will overlook this glaring analytical flaw. If this isn't quite clear, perhaps you should move on to the next chapter.

d) **Enclosure Data Loss Assumption** – The failure of an enclosure rarely causes data loss, unless the hard drives within it are also somehow damaged. Once the enclosure is repaired, the data is recoverable. For the purposes of the analysis, we will generalize to assume that enclosure breakdowns never cause unrecoverable hard drive failures. An extreme example to the contrary is the enclosure which catches fire and destroys the hard drive contained within. Everything's gone; there's no hope for the data (except for DriveSavers); and it was the fault of the enclosure.

Using these assumptions, let's try to create a probability equation which describes failure of an enclosure with one fixed hard drive:

Total Failure Probability = (Number of Hard Drives) * (Hard drive Failure Rate) + (Storage Enclosure Failure Rate), or in shorthand:

Equation 1:

TFP = (N) * HDFR + SEFR, where N = number of drives

(In other words, we've covered the box and everything inside it in one easy failure rate equation.)

Substituting real numbers, we proceed as follows for our example:

Equation 2:

Total Failure Probability = (1) * (2.0%) + (1.0%) = 3%

As a result, we conclude that an enclosure with one hard drive has an annual Total Failure Probability of 3%. Further analyzing this data, Unrecoverable Data Loss (UDL) will happen annually in 2.0% of all failures, while Recoverable Data Loss (RDL) will occur annually in 1.0% of all failures. This is because we deem the failure of the hard drive as unrecoverable (I'd call this the "DriveSavers" scenario) while failure of the enclosure which surrounds it is usually recoverable. Here's some more shorthand for these two equations, as it relates to this sample scenario:

Equation 3: UDL = 2 %
Equation 4: RDL = 1 %

Let's apply this same kind of thinking across the storage types. You will notice that certain types of enclosures have common failure probabilities with others. However, the ease of repair becomes a significant factor, and that will be covered in our concluding section later in this paper.

a) **Enclosures which utilize one fixed hard drive**

Equation 5: TFP = (1) * 2% + 1% = 3%
Equation 6: UDL = 2%
Equation 7: RDL = 1%

b) **Enclosures which utilize one removable hard drive**

Equation 8: TFP = (1) * 2% + 1% = 3%
Equation 9: UDL = 2%
Equation 10: RDL = 1%

c) **Enclosures which utilize multiple fixed hard drives in a non-redundant configuration**

c.1) For a dual drive configuration, the equation is as follows:

Equation 11: TFP = (2) * 2% + 2% = 6%
Equation 12: UDL = 4%
Equation 13: RDL = 2%

(The UDL figure assumes that either hard drive might fail; but the data loss is confined to just the hard drive that does fail.)

c.2) For a quad drive configuration, the equation is as follows:

Equation 14: TFP = (4) * 2% + 4% = 12%
Equation 15: UDL = 8%
Equation 16: RDL = 4%

(The UDL figure assumes that any of the 4 hard drives might fail; but the data loss is confined to just the hard drive that does fail.)

c.3) This math may be extended to enclosures which contain LOTS of hard drives. Consider an enclosure with 8 drives:

Equation 17: TFP = (8) * 2% + 5% = 23%
Equation 18: UDL = 16%
Equation 19: RDL = 5%

(The UDL figure assumes that any hard drive might fail; but the data loss is confined to just the hard drive that does fail.)

d) **Enclosures using multiple removable hard drives in a non-redundant configuration**

d.1) For a dual hard drive configuration, the equation is as follows:

Equation 20: TFP = (2) * 2% + 2% = 6%
Equation 21: UDL = 4%
Equation 22: RDL = 2%

(The UDL figure assumes that either hard drive might fail; but the data loss is confined to just the hard drive that does fail.)

d.2) For a quad hard drive configuration, the equation is as follows:

Equation 23: TFP = (4) * 2% + 4% = 12%
Equation 24: UDL = 8%
Equation 25: RDL = 4%

(The UDL figure assumes that any of the 4 hard drives might fail; but the data loss is confined to just the hard drive that does fail.)

d.3) This math may be extended to enclosures which contain LOTS of hard drives. Consider an enclosure with 8 drives:

Equation 26: TFP = (8) * 2% + 5% = 23%
Equation 27: UDL = 16%
Equation 28: RDL = 5%

(The UDL figure assumes that any of the eight hard drives might fail; but the data loss is confined to just the hard drive that does fail.)

e) **Enclosures which utilize multiple removable hard drives in a redundant configuration, specifically using RAID 5 technology**

e.1) For a configuration with 5 hard drives, **in the absence of RAID 5**, the equation would be as follows:

Equation 29: TFP = (5) * 2% + 5% = 15%
Equation 30: UDL = 10%
Equation 31: RDL = 5%

(The UDL figure assumes that any of the 5 hard drives might fail; and the data loss would affect all 5 hard drives, because they are joined as a common set by the RAID controller.)

e.2) For a configuration with 5 hard drives **with functional RAID 5**, the equation changes significantly, simply because the RAID 5 controller guarantees that the loss of data of any one hard drive does not cause UDL. However, the failure of two hard drives would cause UDL. I'll skip the math involved in determining this failure pattern, and simply state that UDL = 0.4%. (We won't even analyze the probability of 3 hard drives failing… It's a real number but very low.) For this scenario (5 drives, functional RAID 5 configuration, 2 drives causing UDL, here's the equation:

Equation 32: TFP = 0.4% + 5% = 5.4%
Equation 33: UDL = 0.4%
Equation 34: RDL =5%

In this configuration, UDL has been pushed below 1%!!! A functional RAID 5 implementation is approximately 5 times more reliable than an enclosure which contains a single hard drive.

f) **Enclosures which utilize dual removable hard drives in an automatic mirroring configuration**

f.1) In the absence of functional mirroring, the equation is as follows:

Equation 35: TFP = (2) * 2% + 2% = 6%
Equation 36: UDL = 4%
Equation 37: RDL = 2%

(The UDL figure assumes that either hard drive might fail; but the data loss is confined to just the drive that does fail.)

f.2) However, with functional mirroring, the likelihood of simultaneous hard drive failure is reduced to less than .1%, using some mathematical magic. Here are the resulting equations:

Equation 38: TFP = 0.1% + 2% = 2.1%
Equation 39: UDL = 0.1%
Equation 40: RDL = 2%

This has the best UDL rate of any configuration analyzed. It also has the lowest TFP of any configuration. Combined with low acquisition costs, dual drive RAIDs in a mirrored configuration are the finest choice for general purpose backup.

7. Conclusion

Let's summarize the previous chapter's findings, especially for those who skipped it:

Enclosures and the hard drives contained therein will experience failures which may be analyzed using some simple mathematics. Depending on the type of enclosure, the failure rate will vary according to the number of hard drives, the complexity of the enclosure, and the type of RAID level employed (which might be none [JBOD], RAID 0, RAID 1, or RAID 5).

Here is a summary table of the failure rates. Please use these probabilities as relative indicators of failure probability, not as absolute predictions:

Enclosure Type	Total Failure Probability	Unrecoverable Data Loss	Recoverable Data Loss
a) one fixed drive	3%	2%	1%
b) one removable drive	3%	2%	1%
c) multiple fixed drives, non-redundant			
2-drive configuration	6%	4%	2%
4-drive configuration	12&	8%	4%
d) multiple removable drives, non-redundant			
2-drive configuration	6%	4%	2%
4-drive configuration	12%	8%	4%
8-drive configuration	23%	16%	5%
e) multiple removable drives, redundant configuration (specifically RAID 5)			
five drives without RAID 5	15%	10%	5%
five drives with RAID 5	5.4%	0.4%	5%
f) dual removable drives in an automatic-mirroring configuration			
with mirroring turned off	6%	4%	2%
with mirroring turned on	2.1%	0.1%	2%

Please note that enclosure types (**a**) and (**b**) carry identical failure rates. The only difference is that (**b**) has a removable tray. This means that (**b**) is easier to service in the event that the hard drive fails. It also means that (**b**) is an ideal solution to swap one or more hard drives out of the same enclosure.

Also note that (**c**) and (**d**) carry identical failure rates. Once again, the difference is that (**d**) has removable trays, while (**c**) carries all drives inside a fixed enclosure which makes user access difficult or impossible. A strong implication is that (**c**) will require relatively frequent repair at the factory; while (**d**) is more easily serviceable at the hard drive level due to the pop-out trays. WiebeTech does not sell any enclosures in category (**c**).

If avoidance of Unrecoverable Data Loss is a high priority, (**e**) and (**f**) are the clear choices. (**e**) is more expensive (because it has five hard drives within it, along with RAID 5 technology); while (**f**) is the overall reliability champion.

Selecting an Enclosure

Here are the questions you should answer when selecting an enclosure:

- Is this enclosure to be used for backup? If so, only consider those devices with low rates of UDL. This eliminates everything but mirrored and RAID 5 units. If the backup is for casual use, also consider using single-drive systems, especially those with removable trays, because their UDL is 2%.

  Specifically exclude enclosures which contain multiple fixed non-removable internal drives (type **c**) because of the likelihood of failure and the difficulty of repair (requires factory return of the entire unit, also causes data loss of the entire data).

- If the primary purpose of the enclosure is high performance (using a RAID 0 configuration), ensure that you couple this strategy with a secondary enclosure designed to provide an adequate backup. In other words, use one enclosure for performance and another enclosure (of a different, more reliable configuration) for backup.

- Be aware that all enclosures bear the risk of breakdowns. When it comes to the safety of your data, always have a 'Plan B'.

- Remain aware that Total Failure Probabilities for various types of enclosures range between 2.1% and 23%, while Unrecoverable Data Loss ranges between 0.1% and 16% -- covering two orders of magnitude on the user's odds of suffering irretrievable data loss! This makes the process of recognizing what is and is not a reliable enclosure critically important. Some enclosures are far more likely to fail. Other enclosures are designed to work well and provide a very low statistical failure rate.

- Selection of an enclosure is useless if the data isn't backed up! Always have a plan for backing up your data, and work the plan. This is the subject of an entirely different white paper, which hopefully I'll get to sometime later this year.

While it is clear that this information is helpful for selecting an enclosure for any application, our top priority within this paper is to help users select enclosures which are great for backup purposes. At the time of this writing, my top picks for personal or small-business data backup is WiebeTech's RAIDTech (RT8) or SilverSATA II, or the RT5 / RT5x2 if higher capacities are required.

WiebeTech LLC was founded in July 2000 by James Wiebe and has focused on marketing high performance, highly portable storage solutions for consumer and government markets. www.wiebetech.com