BACKUPS

M. E. Kabay, PhD, CISSP-ISSMP

Professor, Computer Information Systems

Norwich University, Northfield VT

Contents

1 Introduction

Nothing is perfect. Equipment breaks, people make mistakes, and data become corrupted or disappear. Everyone and every system needs a well-thought-out backup policy. In addition to making backups, data processing personnel must also consider requirements for archival storage and retrieval of data copies. Backups also apply to personnel, equipment, and electrical power, but this chapter deals exclusively with data backups; for other applications of redundancy, see Chapters 15 and 31.

1.1 Definitions

Backups are copies of data. Normally, backups are stored on a different medium from the original data. In particular, a copy of a file on the same disk as the original is an acceptable backup only for a short time; the *.bak, *.bk!, *.wbk, and *.sav files are examples of such limited-use backups. However, a copy on a separate disk loses value as a backup as the original file is modified; typically, backups are taken on a schedule that balances the costs and inconvenience of the process with the probable cost of reconstituting data that were modified after each backup. Finally, deletion of a working file converts a copy from a backup into an original. Naïve users, in particular, may not understand this relationship; some of them mistakenly believe that once they have a "backup" on a diskette, they can delete the original file safely. However, if original files are to be deleted (e.g., when a disk volume is to be formatted), there must be double backups of all required data to ensure safety should there be any storage or retrieval problems on a particular backup medium.

Throughout this series, the following abbreviations are used to denote data storage capacities:

* KB = kilobyte = 1024 bytes (characters)

* MB = megabyte = 1024 KB = 1,048,576 bytes

* GB = gigabyte = 1024 MB = 1,073,741,824 bytes

* TB = terabyte = 1024 GB = 1,099,511,627,776 bytes.

Incidentally, according to a 1998 proposal from the International Electrotechnical Commission (IEC), the prefixes above should be KiB (kibibytes), MiB (mebibytes), GiB (gibibytes), and TiB (tebibytes) (see http://physics.nist.gov/cuu/Units/binary.html) to distinguish them from the powers-of-ten notations using kilo (10^3), mega (10^6), giga (10^9) and tera (10^{12}), but this suggestion has not yet been widely accepted by the technical community.

1.2 Need

Backups are used for many purposes:

- * First and foremost, to provide is valid information in case of data corruption or data loss;
- * To satisfy legal requirements for access to his storable data; e.g., for audit purposes;

- * In forensic examination of data to recognize and characterize a crime and to identify suspects;

- * For statistical purposes in research;

- * To satisfy requirements of due care and diligence in safeguarding corporate assets;

- * To meet unforeseen requirements.

2 Mechanisms

Because data change at different rates in different applications, backups may be useful when made at frequencies ranging from milliseconds to years.

2.1 Parallel Processing

The ultimate backup strategy is to do everything twice at the same time. Systems such as Tandem and Stratus use redundant components at every level of processing; for example, they use arrays of processors, dual I/O buses, multiple banks of random-access memory and duplicate disk storage devices to permit immediate recovery should any computations go awry. Redundant systems use sophisticated communications between processors to assure identity of results. If any computational components fail, processing can continue uninterrupted while the defective components are replaced.

2.2 Hierarchical Storage Systems

Large computer systems with TB of data typically use a *hierarchical storage system* to place often-used data on fast, relatively expensive disks while migrating less-used data to less expensive, somewhat slower storage media. However, users need have no knowledge of or involvement in such migration; all files are listed by the file system and can be accessed without special commands. Because the tape cartridges are stored in dense arrays (usually cylindrical for minimum mean distance among tapes, hence the name *silo*) with total capacities in the hundreds of TB per silo, fast-moving robotic arms can locate and load the right tape within seconds. The user may experience a brief delay of a few seconds as data are copied from tape cartridges back onto hard disk, but otherwise there is no problem for the users. This system provides a degree of backup simply because data are not erased from tape when they are copied from tape to disk, nor are data removed from disk when they are appended to tape; this data remanence provides a degree of backup (albeit a temporary one) because of the duplication of data.

2.3 Disk Mirroring

At a less sophisticated level, it is possible to duplicate disk operations so that disk failures cause limited or no damage to critical data.

Redundant arrays of Independent Disks (RAID) were described in the late 1980s and have become a practical approach to providing fault tolerant mass storage. The falling price of disk storage (1 Mb of hard disk cost about $200 in 1980 versus about $0.02 in 2001) has allowed inexpensive disks to be combined into highly reliable units that contain different levels of redundancy among the components for applications with increasing requirements for full-time availability. The disk architecture involves special measures for ensuring that every sector of the disk can be checked for validity at every input and output (I/O) operation. If the primary copy of a file shows data corruption, the secondary file is used and the system automatically makes corrections to resynchronize the primary file. From the user's point of view, there is no interruption in I/O and no error.

2.3.1 Workstation and PC Mirroring

Software solutions can also provide automatic copying of data onto separate disks. For example,

* SureSync software can mirror files on Windows NT and Windows 2000;

* UnixWare Optional Services include Disk Mirroring software for SCO UNIX;

* McAfee's Safe & Sound product (among others) allows PC users to duplicate specific files, folders and volumes onto separate disks in real time.

Early users of software-based disk mirroring suffered from slower responses when updating their primary files because the system had to complete output operations to the mirror file before releasing the application for further activity. However, today's software uses part of system memory as a buffer to prevent performance degradation. Secondary (mirror) files are nonetheless rarely more than a few milliseconds behind the current status of the primary file.

3 Logging and Software

3.1 Logging

If real-time access to perfect data is not essential, a well-established approach to high-availability backups is to keep a log file of all changes to critical files. _Roll-forward_ recovery requires

* Backups that are synchronized with log files to provide an agreed-upon starting point;

* Markers in the log files to indicate completed sequences of operations (called _transactions_) that can be recovered;

* Recovery software that can read the log files and re-do all the changes to the data, leaving out incomplete transactions.

An alternative to roll-forward recovery is _roll-backward_ recovery, in which diagnostic software scans log files and identifies only the incomplete transactions and then returns the data files to a consistent state.

3.2 Backup Software

All operating systems have utilities for making backups. However, sometimes the utilities that are included with the installation sets are limited in functionality; for example, they may not provide the full flexibility required to produce backups on different kinds of removable media. Generally, manufacturers of removable media include specialized backup software suitable for use with their own products.

When evaluating backup software, users will want to check for the following particularly minimum requirements:

* The software should allow complete control over which files are backed up.

* Users should be able to obtain a report on exactly which files were successfully backed up and detailed explanations of why certain files could not be backed up.

* Data compression should be available.

* Backups must be able to _span_ multiple volumes of removable media (i.e., a backup must not be limited to the space available on a single volume).

* If free space is available, it should be possible to put more than one backup on a single volume.

* The backup software must be able to verify the readability of all backups as part of the backup process.

* It should not be easy to create backup volumes that have the same name.

* The _restore_ function should allow selective retrieval of individual files and folders or directories.

* The destination of restored data should be controllable by the user.

* During the restore process, the user should be able to determine whether to overwrite files that are currently in place; the overwriting should be controllable both with file-by-file confirmation dialogs and globally (without further dialog).

* Restore operations should be configurable to respect the read-only attribute on files or to override that attribute globally or selectively.

4 Removable Media

4.1 Removable Media

The density of data storage on removable media has increased thousands of times in the last quarter century. For example, in the 1970s, an eight-inch word-processing diskette could store up to 128 KB. In contrast, at the time of writing (September 2001), a removable disk cartridge three inches in diameter stored 20 GB, provided data transfer rates of 15 MB per second, and cost around $100. Moore's Law states that computer equipment power and capacity doubles every 12 to 18 months for a given cost; this relationship definitely applies to mass storage and backup media.

4.1.1 Diskettes

Because of the growing size of application files (e.g., an empty document created with MS-Word 2000 can take 24 KB – 1.6% of a 1.44 MB 3.5" diskette), old-style diskettes are no longer practical for any but the simplest manual backups. In addition, floppy disk drives are so slow that users revolt against requirements to do backups using this medium.

The modern equivalents of floppy disks are in fact hard disks, but they are almost the same size as 3.5" floppy disks despite carrying hundreds or thousands of times more data. For example, IOMEGA Corporation < http://www.iomega.com > is the leading provider of the widely-used ZIP diskette-like storage media with 100 MB and 250 MB capacities. Many PCs and servers being sold today (2001) include ZIP drives as well as or instead of 3.5" floppy drives. In addition, add-on drives are available as portable or in-system units with a variety of interfaces (SCSI, parallel port, and USB).

4.1.2 Large-Capacity Hard Disk Cartridges

IOMEGA also makes JAZ drives in 1 GB and 2 GB capacities. Their Peerless units provide cartridges of 10 GB or 20 GB and drives with a high-speed Firewire interface. Their DataSafe product, intended for servers, has capacities of 160 GB or 320 GB per unit.

4.1.3 Optical Storage

Many systems are now using optical storage for backups. Compact-Disc Read-Write (CD-RW) disks are the most widely-used format; each disk can hold approximately 700 MB of data and costs only a few dollars. The read/write drives cost a few hundred dollars in 2001. In addition, large numbers of CDs are easily handled using "jukeboxes" that apply robotics to access specific disks from collections of hundreds or thousands on demand.

4.1.4 Tape Cartridge Systems

The old 9-track, reel-to-reel 6250 bpi systems used in the 1970s and 1980s held several hundred MB. Today's pocket-sized tape cartridges hold GBs. For example, the industry leader in this field, StorageTek (< http://www.storagetek.com > makes individual tape drives with uncompressed capacities of 20 GB, 60 GB, and 110 GB; compression typically doubles, triples or quadruples these capacities, depending on the nature of the data. Data seek can take 40 seconds; data transfer rates for such systems are typically in the range of 10-15 MB/second. Cartridges have mean-time-between-failure (MTBF) of 250,000 hours with 100% duty cycles and can tolerate 1,000,000 tape passes. All such systems have streaming I/O using about 10 MB of RAM buffer to prevent interruption of the read/write operations from and to the tapes and thus keep the tape moving smoothly to maximize data transfer rates.

In conjunction with automated tape library systems holding many cartridges and capable of automatically switching to the next cartridge, these systems are ideal for backing up servers and mainframes with TB of data. Small library systems keep 10-20 cartridges in position for immediate access (approximately 9 seconds for an exchange). These libraries have approximately 2,000,000 mean exchanges between failures with MTBF of around 360,000 hours at 100% duty cycle.

The largest library systems (e.g., the StorageTek L700) can have up to 678 cartridges loaded at once, up to 20 drives for concurrent access, and total capacities of up to 149 TB with compression. Total throughput can exceed 2 TB/hour.

5 Labeling Media

Regardless of the size of one's backup, every storage device (from diskettes to tape cartridges) should be clearly and unambiguously labeled. At a minimum, one should see the date, system from which data were copied, a description of the data, and the name of the person responsible for the storage medium. Most devices allow electronic labeling; diskettes or removable hard disks used on Windows systems, for example, can be labeled with up to 11 letters, numbers or the underscore character. Larger-capacity media such as cartridges used for UNIX and mainframe systems have extensive electronic labeling available. On some systems, it is possible to request specific storage media and have the system automatically refuse the wrong media if they are mounted in error. Tape library systems typically use optical bar codes that are automatically generated by the backup software and affixed to each cartridge for unique identification. Magnetic tapes and cartridges have electronic labels written onto the start of each unit with specifics that are particular to each operating system and tape-handling software.

5.1 Minimum Requirements

Giving someone a blank storage medium or one with a flimsily attached label is a bad practice that leads to confusion and error. Sticky notes, for example, are not a good way of labeling diskettes and removable disks: if they are taken off, they can get lost; if they are left on, they can jam the disk drives. There are many types of labels for storage media, including printable sheets suitable for laser or inkjet printers and using adhesive that allows removal of the labels without leaving a sticky residue. At the very least, an exterior label should include the following information:

* Date the volume was created (e.g., "2001-09-08")

* Originator (e.g., "Bob R. Jones, Accounting Dept")

* Description of the contents (e.g., "Engineering Accounting Data for 2000")

* Application program (e.g., "Quicken v2002")

* Operating system (e.g., Windows 98).

5.2 README Files

Storing files with _canonical names_ (names with a fixed structure) on the media themselves is also useful. An example of a canonical file much used on installation disks is "READ_ME.TXT" An organization can mandate the following files as minimum standards for its storage media:

* ORIGIN.mmm (where _mmm_ represents a sequence number for unique identification of the storage set) indicating the originating system (e.g., "Accounting Workstation number 3875-3" or "Bob Whitmore's SPARC in Engineering");

* DATE.mmm showing the date (preferably in year-month-day sequence) on which the storage volume was created; e.g., "2001-09-08."

* SET.mmm to describe exactly which volumes are part of a particular set; contents could include "SET 123; VOL 444, VOL 445, VOL 446;"

* INDEX.mmm, an index file listing all the files on all the volumes of that particular storage set; e.g., "SET 123; VOL 444 FIL F1, F2; VOL 445 FIL F3, F4; VOL 446 FIL F5."

* VOLUME.nnn (where _nnn_ represents a sequence number for unique identification of the medium) that contains an explanation such as "VOL 444 SET 123 NUMBER 1 OF 3."

* FILES.nnn which lists all the files on that particular volume of that particular storage set; for example, contents could include "SET 123, VOL 444, Files F1, F2, F3".

Such labeling is best handled by an application program; many backup programs automatically generate similar files.

5.3 Indexing

Backup volumes need a mechanism for identifying the data stored on each medium. Equally important is the capacity to locate the storage media where particular files are stored – otherwise, one would have to search serially through multiple media to locate specific data. Although not all backup products for personal computers include such functionality, many do; server and mainframe utilities routinely include automatic indexing and retrieval. These systems allow the user to specify file names and dates (using wild-card characters to signify ranges) and display a menu of options from which the user can select the appropriate files for recovery.

6 Open Files

All backup systems have trouble with files that are currently in use by processes that have opened them with write access (i.e., which may be adding or changing data within the files). The danger in copying such files is that they may be in an _inconsistent state_ when the backup software copies their data. For example, a multi-phase transaction may have updated some records in a detail file but the corresponding master records may not yet have been posted to disk. Copying the data before the transaction completes will store a corrupt version of the files and lead to problems when they are later restored to disk.

Backup software usually generates a list of everything backed up and of all the files _not_ backed up; for the latter, there is usually an explanation or a code showing the reason for the failure. Operators must always verify that all required files have been backed up and must take corrective action if files have been omitted.

Some high-speed, high-capacity backup software packages provide a buffer mechanism to allow high-availability systems to continue processing while backups are in progress. In these systems, files are frozen in a consistent state so that backup can proceed and all changes are stored in buffers on disk for later entry into the production databases. However, even this approach cannot obviate the need for a minimum period of quiescence so that the databases can reach a consistent state. In addition, it is impossible for full functionality to continue if changes are being held back from the databases until a backup is complete; all _dependent_ transactions (those depending on the previously-changed values of records) must also be held up until the files are unlocked.

7 Fundamental Types of Backups

Backups can include different amounts and kinds of data, as described in the following list:

- * _*Full*_ backups store a copy of everything that resides on the mass storage of a specific system. To restore a group of files from a full backup, the operator mounts the appropriate volume of the backup set and restores the file in a single operation.

- * _*Differential*_ backups store all the data that have changed since a specific date or event; typically, a differential backup stores everything that has changed since the last full backup. The number of volumes of differential backups can increase with each additional backup. To restore a group of files from a differential backup, the operator needs to locate the latest differential set and also the full backup upon which it is based to ensure that all files are restored.

- * _*Incremental*_ backups are a more limited type of differential backup that typically stores everything that has changed since the previous full or incremental backup. As long as multiple backup sets can be put on a single volume, the incremental backup requires fewer volumes than a normal differential backup for a given period. To restore a set of files from incremental backups, the operator may have to mount volumes from all the incremental sets plus the full backup upon which they are based.

- * _*Delta*_ backups store only the portions of files that have been modified since the last full or delta backup; delta backups are a rarely-used type more akin to logging than to normal backups. Delta backups use the fewest backup volumes of all the methods listed; however, to restore data using delta backups, the operator must use special-purpose application programs and mount volumes from all the delta sets plus the full backup upon which they are based.

Another aspect of backups is whether they include all the data on a system or only the data particular to specific application programs or groups of users:

- * _*System*_ backups copy everything on a system;

- * _*Application*_ backups copy the data needed to restore operations for particular software systems.

In addition to these terms, one often hears operators and users refer to _*daily*_ and _*partial*_ backups. These terms are ambiguous and should be defined in writing when setting up procedures.

- In the next article, we'll look at backup strategies for mainframes and servers.

8 Types of Systems to Back Up

8.1 Mainframes

Large production systems using mainframes or networks of servers routinely do full system backups every day because of the importance of rapid recovery in case of data loss. Using high-capacity tape libraries with multiple drives and immediate access to tape cartridges, these systems are capable of data throughput of up to 2 TB/hour (see section 41.2.6.4). Typically, all backups are performed automatically during the period of lowest system utilization. Because of the problems caused by concurrent access, mainframe operations usually reserve a time every day during which users are not permitted to access production applications. A typical approach sends a series of real-time messages to all open sessions announcing, “Full Backup in xx minutes; please log off now.” Operations staff have been known to phone offending users who are still logged on to the network when backups are supposed to start. To prevent unattended sessions from interfering with backups (as well as to reduce risks from unauthorized use of open sessions), most systems configure a timeout after a certain period of inactivity (typically 10 minutes). If users have left their sessions online despite the automatic logoff, mechanisms such as forced logoffs can be implemented to prevent user processes from continuing to hold production files open.

In addition to system backups, mainframe operations may be instructed to take more frequent backups of high-utilization application systems. Mission-critical transaction-processing systems, for example, may have several incremental or delta backups performed throughout the day. Transaction log files may be considered so important that they are also copied to backup media as soon as the files are closed (i.e., when a log file reaches its maximum size and a new log file is initiated for the application programs).

8.2 Servers

Managers of networks with many servers have the same options as mainframe operations staff, but in addition they have increased flexibility because of the decentralized, distributed nature of the computing environment. Many network architectures segregate specific application systems or groups of users to specific servers; therefore, it is easy to schedule system backups at times convenient for different groups of users. In addition to flexible system backups, the distributed aspect of such networks facilitates application backups.

8.3 Workstations

Individual workstations pose special challenges for backup. Although software and backup media are readily available for all operating systems, the human factor interferes with reliable backup. Users are typically not focused on their computing infrastructure; taking care of backups is not a high priority for busy professionals. Even technically-trained users who ought to know better sometimes skip their daily backups; many novice or technically-unskilled workers do not even understand the concept of backups.

If the workstations are connected to a network, there are automated centralized backup software utilities that can protect all the users' files; however, with user disk drives in the many GB of storage (at the time of writing, new PCs were being sold with 30 GB of disk as an unremarkable capacity) and the popularity of large files such as pictures and videos, storing the new data (let alone the full system) for hundreds of workstations can consume TB of backup media and saturate limited bandwidths (it takes a minimum of 291 hours to transfer 1 TB over a communications channel running at 1 MB/sec). There are also privacy issues in such centralized backup if users fail to encrypt their hard disks.

8.4 Portable Computers

Portable or laptop computers are sometimes the only computer a user owns or is assigned; in other cases, the portable computer is an adjunct to a desktop computer. Laptop computers that are the primary system must be treated like workstations. Portables that are used as adjuncts – for example, when traveling – can be backed up separately or they can be synchronized with the corresponding desktop system.

Synchronization software (e.g., the well-known LapLink product) offers a number of options to meet user needs; e.g.,

- * A variety of hardwired connection methods, including cables between serial ports, parallel ports, SCSI ports and USB ports.

- * Remote access protocols allowing users to reach their computer workstations via modem or through TCP/IP connections via the Internet to ensure synchronization or file access.

- * Cloning, which duplicates the selected file structure of a source computer onto the target computer; cloning deletes files from the target which are not found on the source.

- * Filtering, which prevents specific files or types of files from being transferred between computers;
- * Synchronization, in which all changes on the source computer(s) are replicated onto the target computer(s). One-way synchronization updates the target only; two-way synchronization makes changes to both the target and the source computers.
- * Compression and decompression routines to increase throughput during transfers and synchronizations.
- * Data comparison functions to update only those portions of files which are different on source and target; for large files, this feature raises effective throughput by orders of magnitude.
- * Security provisions to prevent unauthorized remote access to user computers.
- * Log files to record events during file transfers and synchronization.

In addition to making it easier to leave the office with all the right files on one's hard disk, synchronization of portable computers has the additional benefit of creating a backup of the source computer's files. My practice, for example, is to make a daily backup of my desktop system onto two separate disks (a ZIP 100 MB and a JAZ 2 GB) and synchronize my portable computer every morning before I head off to the University. Then in the evening I synchronize my main computer to keep everything in synch (no, not the rock band).

8.5 Handheld Computers

Another area that is often overlooked is handheld computers (Palm, Psion, Handspring Visor, RIM Blackberry, daVinci, Helio, HP200LX, Newton/eMate, Rex, Zaurus, Smart Phones, Smart Pagers. PocketMail). These PDAs often contain critically important information for their users, yet not everyone realizes the value of making regular backups. Luckily, synchronizing a PDA with a workstation also makes a backup on the workstation's disk. Security managers would do well to circulate an occasional reminder to users to synchronize or backup their PDAs to prevent data loss should they lose, step on, sit on, or soak their accessory brain. Some PDA docking cradles have a prominent button which allows instant activation of synchronization, which takes only a minute or two.

9 Retention and Rotation

9.1 Backup Archives, Maintenance and Retention

Having created a backup set, what should one do with it? The first thing to do is to test the readability of the data. Modern backup software automatically verifies the readability of backups; this functionality must not be turned off. When preparing for any operation that destroys or may destroy the original data, one should make two independent backups of critical data; it is unlikely that exactly the same error will occur in both copies of the backup. Such dangerous activities include partitioning disk drives, physical repair of systems, moving disk drives from one slot or system to another, and installation of new versions of the operating system.

9.2 Retention Policies

One of the obvious reasons to make backup copies is to recover from damage to files; however, there are also legal and business requirements for data storage and retention. For example, certain jurisdictions require seven years of business data to be available for audits by regulatory or taxation agencies. The corporate legal staff may advise retention of certain data for even longer periods as support for claims of patent rights or if litigation is envisaged. In all cases, the combination of business and legal requirements necessitates consultation outside the data processing department; decisions on data retention policies must involve more than technical resources.

The probability that a backup will be useful declines with time. The backup from yesterday is more likely to be needed than the same kind of backup from last week or last month. On the other hand, each backup contains copies of files which were changed in the period covered by that backup but which may have been deleted since the backup was made. Data center policies on retention vary because of perceived needs and experience as well as in response to the business and legal demands mentioned in the first paragraph of this section. The following gives a sample policy to illustrate some of the possibilities in creating retention policies:

* Keep daily backups for one month.

* Keep the end-of-week backups for three months.

* Keep the end-of-month backups for 5 years.

* Keep the end-of-year backups for 10 years.

9.3 Rotation

Re-using backup volumes makes economic and functional sense. In general, when planning a backup strategy, different types of backups may be kept for different lengths of time, as suggested in section 41.4.1. To ensure even wear on media, volumes should be labeled with the date on which they are returned to a storage area of available media and used in order of recovery (first in, first out). Whenever possible, newer media should be reserved for backup volumes destined for longer retention. An expiry date should be stamped on all tapes when they are acquired so that operations staff will know when to discard out-dated media.

10 Media Degradation

10.1 Media Longevity and Technology Changes

For short-term storage, there is no problem ensuring that stored information will be usable. Even if a software upgrade changes file formats, the previous versions are usually readable. In a year, technological changes such as new storage formats will not make older formats unreadable.

Over the medium term, up to five years, difficulties of compatibility do increase, although not catastrophically. There are certainly plenty of five-year old systems still in use, and it is unlikely that this level of technological inertia will be seriously reduced in the future.

Over the longer term, however, there are serious problems to overcome in maintaining the availability of electronic records. Over the last ten to twenty years, certain forms of storage have become essentially unusable.

As an example, AES was a powerful force in the dedicated word-processor market in the 1970s; eight-inch disks held dozens or hundreds of pages of text and could be read in almost any office in North America. By the late 1980s, AES had succumbed to word-processing packages running on general-purpose computers; by 1990, the last company supporting AES equipment closed its doors. Today, it would be extremely difficult to recover data from AES diskettes.

The problems of obsolescence include data degradation, software incompatibilities and hardware incompatibilities.

10.2 Media Degradation

Magnetic media degrade over time. Over a period of a few years, thermal disruption of magnetic domains gradually blurs the boundaries of the magnetized areas, making it harder for I/O devices to distinguish between the domains representing ones and those representing zeroes. These problems affect tapes, diskettes and magnetic disks and cause increasing parity errors. Specialized equipment and software can compensate for these errors and recover most of the data on such old media.

Tape media suffer from an additional source of degradation: the metal oxide becomes friable and begins to flake off the Mylar backing. Such losses are unrecoverable. They occur within a few years in media stored under inadequate environmental controls and within five to ten years for properly-maintained media. Regular regeneration by copying the data before the underlying medium disintegrates prevents data loss.

Optical disks, which use laser beams to etch bubbles in the substrate, are much more stable than magnetic media. Current estimates are that CD-ROMs, CD-RW and DVD disks will remain readable in theory for at least a decade and probably longer. However, they will remain readable in practice if and only if future optical-storage systems include backward compatibility.

11 Technological Change

11.1.1 Software Changes

Software incompatibilities include the application software and the operating system.

The data may be readable, but will they be usable? Manufacturers provide backward compatibility, but there are limits. For example, MS-Word 2000 can convert files from earlier versions of Word—but only back to version 4 for Windows. Over time, application programs evolve and drop support of the earliest data formats. Database programs, E-mail, spreadsheets—all of today's and tomorrow's versions may have trouble interpreting data files correctly.

In any case, all conversion raises the possibility of data loss since new formats are not necessarily supersets of old formats. For example, in 1972, RUNOFF text files on mainframe systems included instructions to pause a daisy-wheel impact printer so the operator could change daisy wheels—but there was no requirement to document the desired daisy wheel. The operator made the choice. What would document conversion do with that instruction?

Even operating systems evolve. Programs intended for Windows 3.11 of the early 1990s do not necessarily function on Windows ME in the year 2000. And the operating systems of yesteryear do not necessarily even run on today's hardware.

Finally, even hardware eventually becomes impossible to maintain. It would be extremely difficult to retrieve and interpret data from word-processing equipment from even twenty years ago. No one outside museums or hobbyists can read an 800 bpi 9-track ¾-inch magnetic tape from a 1980 HP3000 Series III minicomputer. Over time, even such parameters as electrical power attributes may change, making obsolete equipment difficult to run even if they can be located.

The most robust method developed to date for long-term storage of data is COM (Computer Output to Microfilm). Documents are printed to microfilm, appearing exactly as if they had been printed to paper and then micro-photographed. Storage densities are high, storage costs are low, and in the worst case, the images can be read with a source of light and a simple lens.

In the next article in this series, we'll look at temporarily storing backups onsite.

12 Onsite Storage

It is obviously foolish to keep backups in a place where they are subject to the same risks of destruction as the computer systems they are intended to protect. However, unless one uses a remote location for making backups through telecommunications channels, all backups must spend at least some time in the same location as the systems that are being backed up.

At a minimum, backup policies should stipulate that backups are to be removed to a secure, relatively distant location as soon as possible after completion. Temporary onsite storage areas that may be suitable for holding backups until they can be moved offsite include specialized fire-resistant media storage cabinets or safes, secure media-storage rooms in the data center, a location on a different floor of a multi-floor building, or an appropriate location in a different building of a campus. What is _*not*_ acceptable is to store backup volumes in a cabinet right next to the computer that was backed up. Even worse is the unfortunate habit of leaving backup volumes in a disorganized heap on top of the computer from which the data were copied.

In a small office, backups should be kept in a fire-resistant safe if possible while waiting to take the media somewhere else.

12.1 Environmental Protection

Magnetic and optical media can be damaged by dust, mould, condensation, freezing, and excessive heat. All locations considered for storage of backup media should conform to the media manufacturer's environmental tolerances; typical values are 40-60% humidity and temperatures of ~50-75 F (~10-25 C). In addition, magnetic media should not be stacked horizontally in piles; the housings of these devices are not built to withstand much pressure, so large stacks can cause potentially damaging contact between the protective shell and the data storage surface. Electromagnetic pulses and magnetic fields are also harmful to magnetic backup media; keep mobile phones (both wireless and cellular) away from magnetic media. If the organization uses degaussers to render data more difficult to read before discarding data these devices should never be allowed into an area where magnetic media are in use or stored.

In the next article in this series, we'll start looking at options for offsite storage of backups.

13 Homes, Safes and Banks

It is normal to store backups away from the computers and buildings where the primary copies of the backed-up data reside.

13.1 Care During Transport

When sending backup media out for storage, operations staff should use lockable carrying cases designed for specific media. If external firms specializing in data storage pick up media, their staff will usually supply such cases as part of the contract. If media are being transported by corporate staff, it is essential to explain the dangers of leaving such materials in a car: in summertime the cars can get so hot that they melt the media, whereas in winter they can get so cold that the media instantly attract harmful water condensation when they are brought inside. In any case, leaving valuable data in an automobile is asking for theft.

13.2 Homes

The obvious but dangerous choice for people in small offices is to send backup media to the homes of trusted employees. This practice is a very bad idea:

Although the employee may be trustworthy, members of that person's family may not be so. Especially where teenaged and younger children are present, keeping an organization's backups in a private home poses serious security risks.

Environmental conditions in homes may be incompatible with safe long-term storage of media. For example, depending on the cleaning practices of the household, storing backups in a cardboard box under the bed may expose the media to dust, insects (e.g., bed-mites), cats, dogs, pet rodents and damage from vacuum cleaners.

In addition, temperature and humidity controls may be inadequate for safe storage of magnetic media.

Homeowner's insurance policies are unlikely to cover loss of an employer's property and will surely not cover consequential losses resulting from damage to crucial backup volumes.

Legal requirements for demonstrable chain of custody of corporate documentation on backup volumes will not be met if the media are left lying around a private home where unknown persons may have access to them.

13.3 Safes

There are no fire-proof safes, only fire-resistant safes. Safes are available with different degrees of guaranteed resistance to specific temperatures commonly found in ordinary fires (those not involving arson and flame accelerants). Sturdy small (one or two cubic foot) safes are available for use in small offices or homes and can withstand the relatively short time required to burn a house or small building down. They can withstand a fall through one or two floors without breaking open. However, for use in taller buildings, only more expensive and better-built safes are appropriate to protect valuable data.

13.4 Data Vaults

Most enterprises will benefit from contracting with professional, full-time operations specializing in maintaining archives of backup media. Some of the key features to look for in evaluating such facilities include

* Storage facilities made of concrete and steel construction to reduce risk of fire.

* No storage of paper documents in the same building as magnetic or optical media storage.

* Full air-conditioning including humidity, temperature and dust controls throughout the storage vaults.

* Fire sensors and fire retardant technology, preferably without the use of water.

* Full time security monitoring including motion detectors, guards, and tightly controlled access.

* Uniformed, bonded personnel.

* Full time, 24 x 7 x 365 data pickup and delivery services.

* Efficient communications with procedures for authenticating requests for changes in the lists of client personnel authorized to access archives.

* Evidence of sound business planning and stability.

References from customers similar in size and complexity to the enquiring enterprise will help manager make a wise choice among alternative suppliers.

13.5 Online Backups

An alternative to making one’s own backup copies is to pay a third party to make automatic backups via high-speed telecommunications channels and to store the data on behalf of customers. Some of the firms involved in these services move data to magnetic or optical backup volumes, but others use RAID (ganged disks) for instant access to the latest backups.

Additional features to look for when evaluating online backup facilities:

* Compatibility of backup software with computing platform, operating system and application programs.

* Availability of different backup options (full, differential, incremental, delta).

* Handling of open files.

* Availability and costs of sufficient bandwidth to support desired data backup rates.

* Encryption for data during transmission and when stored at service facility.

* Strong access controls to limit access to stored data to authorized personnel.

* Physical security at the storage site and other criteria similar to those for data vaults.

In the next article in this series, we'll start looking at disposal of discarded backup media.

14 Scavenging

Before throwing out backup media containing unencrypted sensitive information, operations and security staff should ensure that the media are unreadable. This section looks at the problem of data scavenging and then recommends methods for preventing such unauthorized data recovery.

Computer crime specialists have described unauthorized access to information left on discarded media as scavenging, browsing, and Dumpster-diving (from the trademarked name of metal bins often used to collect garbage outside office buildings).

Scavenging is probably the third most important method of computer crime; the first two are data diddling and unauthorized use of computer services.

Scavenging can take place within an enterprise; for example, there have been documented cases of criminals who arrange for requests to read *scratch tapes* (tapes that are used for temporary storage of data) before they read them. These people were prospecting for tidbits of data left by previous users. Operations policies should not allow scratch tapes or other media to contain confidential data; all scratch media (including backup media that are being returned to the free list) should be erased them before they are put on the media rack.

Before deciding to toss potentially valuable documents or backup media into the garbage can, managers should realize that in the United States, discarded garbage is not considered private property under the law, according to a U.S. Supreme Court ruling. Anything that is thrown out is fair game for warrantless searches or inspection by anyone who can gain access to the garbage without violating laws against physical trespass. Readers in other jurisdictions should obtain legal advice on the applicable statutes.

Under these circumstances, the only reasonable protection against data theft is to make the garbage unreadable.

15 Destroying Data and Media

Most people know that when a file is erased or purged from a magnetic disk, operating systems usually leave the information entirely or largely intact but remove the pointers from the directory. For example, DOS and Windows obliterates the first character of an erased file with a random character and removes its entries from the FAT (file allocation table). _*Unerase*_ utilities search the disk or diskette and reconstruct the chain of *extents* (areas of contiguous storage), usually with human intervention to verify that the data are still good.

Multiuser operating systems remove pointers from the disk directory and returns all sectors in a purged file to a disk free-space map, but the data in the original extents persist unless specific measures are taken to obliterate them.

Formatting a disk actually zeroes diskettes and hard disks; however, even formatting and overwriting data on magnetic media may not make the data unreadable to the most sophisticated equipment. Since information on magnetic tapes and disks resides in the difference in intensity between highly-magnetized areas (1s) and less-magnetized areas (0s), writing the same thing (0s or 1s) in all areas of the obliteration merely reduces the signal-to-noise ratio. That is, the residual magnetic fields still vary in more or less the original pattern—they're just less easily distinguished. Using highly sensitive readers, a magnetic tape or disk that has been zeroed will still yield much of the original information. Degaussers (portable electromagnets) are of limited use in making data unreadable.

One way of destroying data on magnetic media is to overwrite several passes of random patterns. The random patterns make it far more difficult to extract useful information from the discarded disks and tapes. Military-grade erasure programs use seven passes to obliterate data remanence.

Another solution is physical destruction of magnetic or optical backup media before they are discarded. For end-user departments, operations and security can provide identifiable secure collection receptacles (typically black) throughout the enterprise. Discarded media can be erased or destroyed by appropriate staff on a regular schedule.

Hard disks, tapes, optical disks and floppy disks can be cut into pieces, melted with oxy-acetylene torches, crushed in large compactors and incinerated (although proper incineration requires specialized equipment to prevent atmospheric release of toxic byproducts). Some commercial companies specialize in secure destruction of sensitive records and can provide bonded pickup services or mobile destruction units that move from enterprise to enterprise on a regular schedule and handle paper as well as magnetic and optical media.

16 Costs

All data center managers should be able to answer questions about the costs of the backups being made on their systems. In calculating costs, the following factors should be included in a simple spreadsheet (the example uses tapes as the example of backup medium):

- Tape costs
    - Tapes/backup
    - Purchase cost/tape
    - Storage cost/tape (including cost of tape rack and cost of owning or renting floor space)
    - Tape cost/backup (total cost * number of tapes)
    - Backup cycles saved (current week, end of week, end-of-week set, month-end set, year-end set
    - Cost of all tapes (sets * total cost)
- Time costs
    - Hour/tape for backup
    - Hour/backup total
    - Cost/hour for operator (salary + benefits)/hour
    - System cost/month (purchase, finance, maintenance, floor space, electricity, air conditioning, insurance, system management services, software licenses and maintenance)
    - Days used per month

    - Hour/day availability
    - Cost/hour system
    - Time cost/hour backup
- Total cost/backup (tapes + time)
- Annualized costs
    - Backups/year
    - Total backup hours/year
    - Time cost/year
- Total investment/year (tape costs + time costs)

This detailed information will be invaluable in answering management questions about backup policies and their rationale.

17 Optimizing Frequency

Suppose a manager asks the security and operations staff the following questions:

- "If backups are so important that you do a daily full backup, why don't you do a full backup twice a day?"
- "If taking a daily full backup is good enough for you, why don't you save money by doing a full backup only every other day?"

To answer such questions, managers must be able to adjust the frequency of backups to the perceived risk. One of the ways of approaching a rational allocation of resources when faced with random threats is to calculate the _*expected value*_ of a strategy. The expected value is the average gain (if it's a positive quantity) or loss (if it's negative) that participants will incur in a process that involves random events. When this technique applies to losses over an entire year, it is called the _*annualized loss expectancy*_. This approach is used by insurance companies to balance the costs of premiums against the disbursements to customers.

For backups, the principle is summarized by the following equation:

$$\mathbf{E(x) = P(u)*C(u) - P(n)*C(n)}$$

where

- * x is some particular strategy such as doing a daily full backup

- * E(x) is the expected value or cost of the strategy

- * P(u) is the probability of having to use the backup within a single day; e.g., 1 chance in a 1000 or 0.001

- * C(u) is the money saved by not having to redo all the work that would otherwise be lost if there were no backup; e.g., the cost of paying for reconstruction of the previous day's data (e.g., $9,000) + avoidance of lost business, wasted salary and other expenses during 3 hours of downtime during reconstruction (e.g., $30,000) for a savings of $39,000 per incident when the backups are available

- * P(n) is the probability of not having to use the backups at all in given day = 1 - P(y) = 0.999

- * C(n) is the cost of doing and storing a daily backup that won't be used (e.g., $50).

Then the expected value of doing a single daily full backup using the figures used in the examples above is

E(x) = (0.001 * $39,000) - (0.999 * $50) = $39 - $49.95 = -$10.95.

In other words, the daily full backup has an average cost of about $11 per day when the likelihood of its use is factored into the calculations. This is equivalent to a self-insurance strategy to prevent larger disasters by investing money in preventive mechanisms and measures for rapid and less expensive recovery than possible without the backups.

If one adjusts the frequency of backups, the calculated loss expectancy can be forced to zero or even to a positive number; however, no self insurer makes a profit from loss avoidance measures. Nonetheless, adjusting the frequency and costs of backup strategies using the suggested factors and calculation of loss expectancies can help a data center manager to answer questions from management about backup strategies in a rational manner.

Unfortunately, no one can actually estimate precisely how much a disaster costs nor compute precise probabilities of having to use backups for recovery.

In many organizations, the volume of changes follows a seasonal pattern. For example, 80% of all orders taken might come in two two-month periods spaced half a year apart. Registration for colleges occurs mostly in the autumn, with another bulge in January. Boat sales and ski sales follow seasonal variations. Despite this obvious variability, many organizations follow the same backup schedule regardless of date. It makes sense to adjust the frequency of backups to the volatility of data: operations can schedule more frequent backups when there are lots of changes and fewer when the data are relatively stable.

18 For Further Reading

Desai, A. (2000). *SQL Server 2000 Backup & Recovery.* McGraw-Hill Professional Publishing. ISBN 0-072-13027-X. 698 pp.

Farkas, D. F. (2000). Backups for Beginners. *PCWorld.com* < http://www.pcworld.com/resource/printable/article/0,aid,15593,00.asp >

Indiana University (2000). Unix System Administration Independent Learning (USAIL). Backups. < http://uwsg.iu.edu/usail/library/backups.html >

Kozierok, C. M. (2001). The PC Guide: Backups and Disaster Recovery. < http://www.pcguide.com/care/bu/ >

McMains, J. R. (1998). Windows NT Backup & Recovery. Osborne McGraw-Hill. ISBN 0-078-82363-3. 474 pp.

Molina, Joe (2001). The RAB Guide to Non-Stop Data Access. < http://www.raid-advisory.com/rabguide.html >

Preston, W. C. & G. Estabrook (1999). *UNIX Backup and Recovery*. O'Reilly & Associates. ISBN 1-565-92642-0. 707 pp.

Velpuri, R. & A. Adkoli (1998). *Oracle8 Backup and Recovery Handbook*. McGraw-Hill Professional Publishing. ISBN 0-078-82389-7. 608 pp.

Winegarden, J. (2000). Linux backups HOWTO. < http://www-jerry.oit.duke.edu/linux/bluedevil/HOWTO/backups_howto.html >