Publication: Continuity Insights; Date: Oct 1, 2011; Section: Information Security; Page: 12

Information Security

Data Loss, Encryption & Backups

Ron LaPedis, Managing Director, Seacliff Partners International, LLC

The word "loss" has one meaning with physical items, but can have multiple meanings when talking about digital assets. When you say, "I lost my iPhone," I know that you don't have it. Maybe you just misplaced it, or maybe it was stolen (did you leave it in a taxi?).

Based on the last time that you saw your iPhone, you probably have a good idea whether or not it was actually stolen from you. If you knew you had it after you arrived home but couldn't find it an hour later, the chances are high that it was not stolen.

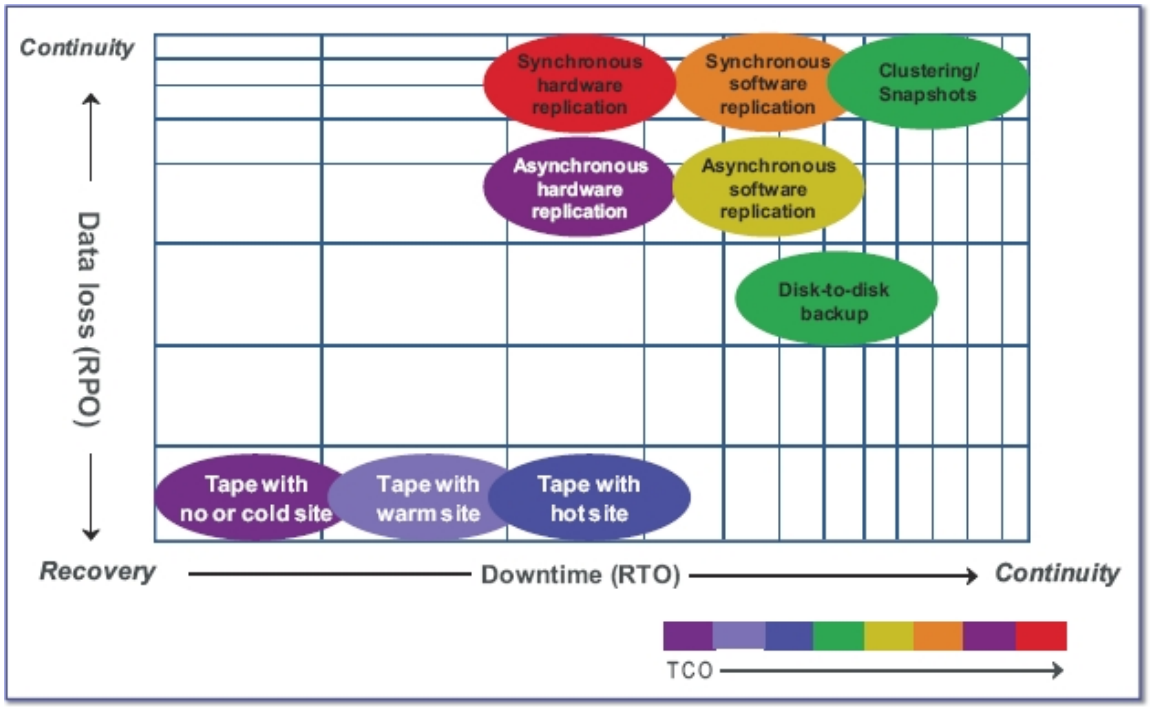
On the other hand, if you left it on the table while you picked up a double grande latte with low-fat milk and extra syrup, maybe someone did take it.

But what do you mean (and what do others think you mean) when you say, "I lost my customer information?" Do you mean that the actual data was corrupted, or that it just isn't on the disk anywhere? Maybe the power went out and it's only lost for some period of time. Or did disaster strike and destroy the equipment or the entire building?

Perhaps you mean that it was stolen by a hacker or otherwise disclosed. Could it have been purposely deleted, perhaps maliciously, in violation of company policy or government regulations? If it was corrupted, destroyed or it simply disappeared, do you have a backup? How old is the backup and how much data could you be missing if the backup is too old?

No matter what "data loss" means to you, planning ahead can help keep you covered. Let's assume that the data just isn't on the disk or has been corrupted. It's important to work with your business units ahead of time to determine their data recovery time objective (RTO) – how long can they do without the data – and recovery point objective (RPO) – how "fresh" the data is when you get it back.

Once armed with the business unit's requirements, you can select from one or more of the many different technologies offered by vendors to help protect data. Here are just a few of your choices (figure 1):



- High availability storage clustering, either local or remote, that offers instantaneous RTO with no data loss.
- File system snapshots that offer an RTO of seconds and RPO of zero to seconds.
- Synchronous data replication with an RTO of seconds to minutes and RPO of zero.
- Asynchronous data replication with an RTO of seconds to minutes and RPO of seconds.
- Disk-to-disk backup with an RTO and RPO of minutes to hours.
- Tape Backup with an RTO and RPO of hours to days.
- The use of one or more cloud providers for storage, replication or backup, with RTO and RPO consisting of whatever the vendors can provide.

These technologies can be combined as required to meet your RTO and RPO, with one or more as Plan A, some as Plan B and others as Plan C. For example, you might want to implement storage clustering with replicated snapshots as Plan A, and disk-to-disk or disk-to-tape backup as Plan B.

You might be asking why backup is needed if you have implemented replication. Isn't a replicate a backup? The answer is no, and I'll discuss why this is the case later in the article.

Preventing Data Disclosure

There are several techniques you can use to mitigate data disclosure and theft. Encrypting data is one way to prevent data disclosure, but it is not a panacea. Encryption alone cannot protect your data without other processes and controls, such as protection of the encryption keys, access control list (ACL) enforcement and separation of duties.

Several companies sell data leakage prevention (DLP) solutions that protect data by enforcing rules on how data can and cannot be accessed or distributed. For example, DLP can allow specific files to be viewed but not be printed or emailed, while other files can be emailed within the company but not to external addresses.

Most of these solutions require you to first classify your data, although some of them try to classify data automatically by scanning files for keywords such as "internal use only" or "confidential." Whether the classification is manual or automatic, you need to specify the rules to be followed for each classification and authorized user.

Encryption Backgrounder

Companies that sell encryption products will have you believe that your data is safe once it is run through their product. In fact, many rules and regulations state that your company is protected from punishment as long as the data is encrypted.

California SB 1386, one of the earliest data breach notification acts, simply states, "'personal information' means an individual's first name or first initial and last name in combination with any one or more of the following data elements, when either the name or the data elements are not encrypted." Note that there is no specification of accepted encryption algorithms, minimum key lengths or key management.

While I am not a lawyer, it seems to me that this law assumes that whatever is in the hands of the attacker is encrypted. Since data that is protected by full disk encryption (FDE) or self-encrypting drives (SED) is transparently decrypted when the drive is active, the exemption based on the use of encryption probably only applies if the actual drive was stolen and not if the data was taken by hacking into a running computer.

Self-Encrypting Drives

SED encryption is done within the drive hardware itself. Once an administrator unlocks the drive, access to the data is completely transparent to the software accessing it. To make the data permanently inaccessible, the administrator changes the drive's encryption key, cryptographically shredding everything on the drive.

Normally, the data on SEDs is only protected when the computer is completely powered down. That is, your confidential information is not protected if the computer is sleeping or hibernating (the technical term is "S3 Standby Mode"). If the computer is lost or stolen while in these modes, a thief might be able to access the data.

There is new aftermarket software that eliminates this threat by locking the SED when the computer is put into sleep mode. When the device resumes from sleep, the user is prompted to re-enter their credentials, once again unlocking the drive.

Best Practices

Unless you are running software that locks a SED on sleep or hibernate, company policy should require that computers carrying confidential information be completely shut down when the employee is not using it.

This functionality occurs immediately, with no productivity loss to the user. Recovery from a hibernated state is also supported, without the "blue screens of death" that can be caused by encrypted drives.

In directly attached redundant array of independent disks (RAID) configurations or in large storage systems, such as those from EMC, HP and NetApp, multiple disk drives can be aggregated into a volume or share. If SEDs are in use, all of the disks in the aggregate must be unlocked for the volume or share to be accessible.

Full Disk Encryption

FDE can be implemented within the operating system or storage system file system and is normally implemented on a volume or network share (also called a mapped drive). Data is automatically encrypted when it is written to the volume or share, and automatically decrypted when it is read off of the volume or share. Depending on how FDE is implemented, your data might face the same risks as data on an SED, or it may be decrypted only after a user is re-authenticated when the computer is woken up from sleep or hibernation.

Even if a volume or network share is made up of multiple disk drives, only one encryption key is used to unlock it.

Higher Level Encryption

There are many places where data encryption can be performed (figure 2):

1. Application
2. Encryption library bound to the application
3. Database engine
4. Disk driver
5. Storage host adaptor
6. Encryption switch or appliance
7. Storage array
8. FDE or SED

The advantages and drawbacks of these encryption points are addressed in an earlier article.

No matter where the encryption is done, a key manager will both protect and keep track of the keys required to encrypt and decrypt your information. The fact that financial information in the United States must be stored for seven years means that you will have a lot of keys to maintain.

There are several personal encryption devices on the market that allow users to encrypt files and folders under a unique key. Files encrypted with these devices cannot be decrypted unless the device is connected to the user's computer and the user has authenticated to it. To gain access to information protected in this manner, an attacker would need to steal the files, the encryption device and the owner's password.

Why Encryption Alone Is Not Enough

The most secure encryption system in the world cannot protect your data if a hacker can get to it as an authorized user. This is why compartmentalized data and separation of duties is so important within your organization.

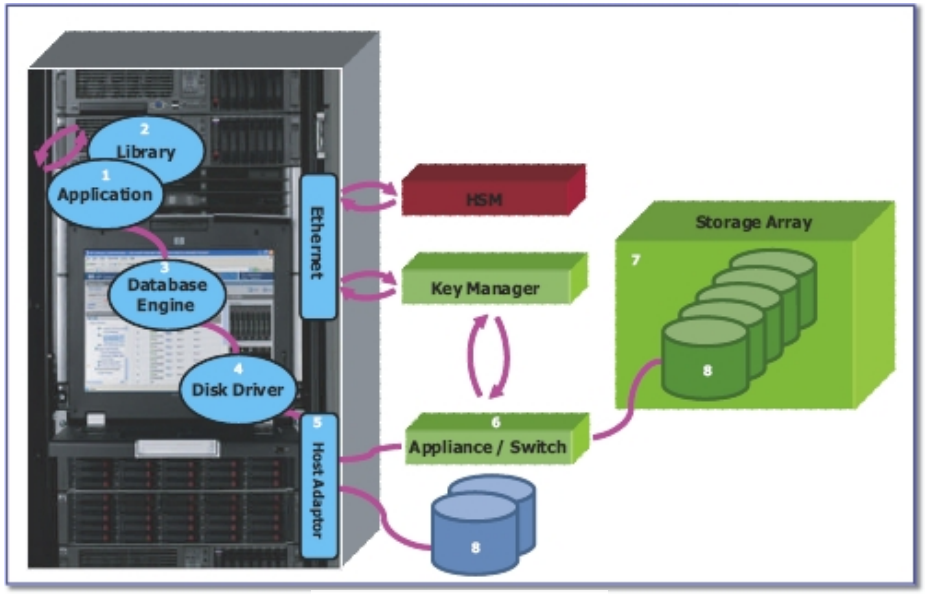


Figure 2: Encryption Injection Points

When your data is properly compartmentalized, access is restricted solely to the information that an employee needs to do their job and no more. In last year's WikiLeaks case, Private Manning had access to thousands of documents that had nothing to do with his job, making it easy for him to download and burn them to a CD-ROM.

Best Practices

Proper separation of duties includes the following:

- The root or superuser account should be locked away and never used except in case of a dire emergency. The root account should be prevented from logging on to any other account without using a password or changing the password of any other user.
- One or more security administrator accounts should be created, each with specific, limited authority. For example:
  - Administrator I, who can create, roll and alter security logs, should have no other system access.
  - Administrator II, who can modify security subsystem settings, should not be able to access the logs or modify user settings.
  - Administrator III, who can modify user security settings, should not be able to alter the security subsystem or access the logs.
- System, backup and application accounts should all be separated:
  - The administrator that backs up and restores files and databases should not be able to read and write them.
  - The administrator who can start and stop applications and databases should not be able to access them.
- Applications and databases should be owned by "frozen" accounts. These accounts help prevent anyone from logging on to this account where they might have full access to an application or database.

By combining separation of duties and encryption, along with securing files and databases to specific users and computers, you can create an environment that provides comprehensive defense for your critical applications and data. These settings will make access to applications and data much more difficult for an attacker because they won't be able to login to the accounts that own the applications and databases.

Why You Need Backups

To save money on IT costs, you might be tempted to forgo offline backup to disk or tape and rely on the data that is replicated to a remote site for disaster recovery purposes as your backup. Understand that a replicate cannot be your backup or archive.

Firstly, any changes that are made to your primary database are sent to the replicate in real time or near real time. If you are required by regulatory or e-discovery requirements to archive data, this is not acceptable since old data can be deleted. Think of an email system in a government or financial system which is required to archive deleted messages for a minimum period of time. When an email is deleted from the primary, it is also deleted from the replicate.

Even if you are not required to maintain an archival copy of specific information, such as an online catalog, employee contact information or website, a replicate still is not safe as a backup, and here is why: If your infrastructure is attacked, or if a human or application makes an error, your primary data could be corrupted or deleted. Depending on the speed of your replication solution, you might not have time to stop it so that the corruption or deletion isn't replicated. And if you can't stop the replication, then the replicate is ruined as well. If you don't have a pre-corruption or pre-deletion backup, your data is gone forever.

So, how often should you take backups of your live data? That depends on your RPO. Your loss-of-site Plan A might be your replicated data, while your data corruption Plan A and your loss-of-site Plan B might be an offline or disk-to-disk backup copy. Again, you need to work with your business units to determine their requirements.

Application Consistency

In today's always-on infrastructure, data is always in transition and specific steps must be taken to ensure that a backup is "application consistent." That is, if an application transaction consists of multiple steps, the backup must either capture all of those steps or none at all.

Many companies sell backup solutions that use agents or other technology to ensure that online backups provide application or transactional consistency when the database is recovered.

If you are required to archive data for seven or more years, you may need to backup the soft-

Best Practices

It really doesn't matter whether you use hard drives, tape, CD-ROMs, DVDs or WORM (write once, ready many) to take backups as long as your backups are:

- Electronically marked "read only" so that they cannot accidentally be overwritten.
- Transactionally consistent across your entire application.
- Actually usable in the event of a problem (do you run restore tests?).
- Encrypted to prevent unauthorized access.
- Secured from loss or damage.

ware and encryption keys that are required to access the backups, since later software revisions may not be able to read older formats. In some cases, you may need to maintain previous generations of backup hardware, such as tape drives, that can read the storage media.

The ideal method and frequency of backups, and the number of generations (previous backup sets) retained is not easy to determine. One common backup strategy is to take a full backup once a week, where you make a copy of all of your files, and daily partial backups, where you copy only changed files. Assuming that you only take backups on weekdays, you will have five backup sets at the end of the week: The one full backup and four partial backups. The following week you again take a full backup, which becomes the next generation. When you run out of disk space or tape, or can no longer afford to store older tapes, erase the disks or tapes that make up the oldest generation.

Best Practices

When formulating the frequency of backups and the number of generations retained, consider:

- **E-discovery: If you are required to turn over e-mail messages, you need to keep a searchable archive that includes messages that have been deleted from users' mailboxes.**
- **Recovery from corrupt data: When you discover that you have corrupt data, how many generations back will you need to go to recover it? What if the last two, three or four generations have the corrupted data on them?**
- **Recovery from deleted data: When you discover that you have missing data, how many generations back will you need to go to recover it? And will you find that you just erased the last generation of disks or tapes that contained the missing data?**
- **Retention regulations.**

The Cloud

Moving to the Cloud does not mean that you can let someone else worry about data loss. In fact, the Cloud presents many more risks that you might not even think about.

When was the last time you read a software license agreement? The service level agreement (SLA) from your Cloud provider may have caveats against both corruption or destruction of data, and data theft. This means that it is up to you to protect your own data from disclosure or destruction. In addition, you need to assess the risk to your company if your Cloud provider goes out of business, is raided by law enforcement, or can no longer be accessed because of an accidental or deliberate disconnect between your company and wherever the Cloud provider has put your data.

Data Unavailability, Corruption & Archive

When assessing Cloud providers, don't just study their SLAs – check historical data for as far back as you can to develop downtime statistics. If the provider cannot meet your RTO and RPO, select a different provider. If none of them can meet your requirements, think about using more than one at the same time. There are rumors that the new Apple iCloud infrastructure uses both Azure and Amazon so that if one goes down, the data will be available from the other.

Questions for cloud providers include:

- If data becomes corrupted or is deleted, can the provider get it back for you?
- How granular is the provider's recovery, and what RTO and RPO will they guarantee?
- Can you go back a day, a week and a month?
- Will the provider archive data for you? If they do, how long will it take to get it back when you need it?
- Does the provider lock your data away so that you cannot download it in order to move to another provider or bring it back in house?

Data Disclosure

Do you trust your Cloud provider to protect your information from disclosure? What if your provider receives a subpoena for your data? Several vendors sell data encryption appliances that can integrate with your existing user infrastructure and encrypt your data before it is transferred to the Cloud. These same appliances can be placed within your Cloud provider's data center to give you control over your data, including the encryption keys, while preventing unauthorized access to your information.

Summary

Whether in house, or in a private or public Cloud, protecting your data against loss means protecting it against a spectrum of risks. If you are worried about corruption or deletion of data, then you should be thinking about replication and backup. If you are worried about your data falling into the wrong hands, then you should be thinking about data encryption solutions, along with access controls and separation of duties. Last but not least, if you are worried about regulatory compliance or e-discovery, then you might also have a requirement to store your data in a provably read-only form. CI