

November 19, 2007

## **RE: The Hidden Downsides to Internet Based Backup.**

By Darren McBride



***Summary: There are some hidden disadvantages to internet based backup which are often not understood until the system is already in place and the real problems arise. This document tries to relate a few of those pitfalls.***

**Introduction.** Backing up business data over the Internet to a data vaulting facility has become increasingly popular over the last five years. While such services work well for some businesses, there are a number of problems and pitfalls to be aware of. Internet based backup services often fail to deliver the performance, features, and security they advertise. The problems can be mitigated somewhat by using a local backup appliance that stores the backs up locally and then spools the data off-site as time permits. This at least allows a first backup to complete locally before attempting to send the backup through the internet to a backup data vault (which can be an order of magnitude slower). However, even this improved (and more expensive) design can have problems. Just a couple of the common pitfalls are that typical Internet connections move data at half the rated raw data bit rate (they have 50% overhead) and while most internet backup strategies only attempt to send data which has changed, modern data sets (i.e. Exchange and SQL) change more on a daily basis than many people believe.

This whitepaper outlines these and other real world problems associated with internet based backup services in an effort to educate the customer and allow them to consider all the pros and cons of Internet based backup versus local backup.

**The Need to Seed.** Most experienced Internet backup providers will acknowledge the need to “seed” the initial backup by doing a full local backup with a portable hard drive (such as High-Rely’s USB or eSATA drives) and sending this seed backup via Fed-Ex to the data center. This is

because many full data sets take too long to backup over most internet connections. For example, 100 GB of data takes 9 days to upload over a T1 line assuming the T1 is used for nothing else 24 hours per day and has the typical 50% protocol overhead (see “Hidden cost of Bandwidth” below for more on this).

**The small daily change fallacy.** Once you “seed” your backup, you must next be able to copy the changed data over the Internet on a daily basis. The argument is usually made that “change data” at a block level (little data chunks) is relatively small on a day to day basis. With modern databases, this is often an incorrect assumption. For example, a Microsoft Exchange server stores the bulk of its mail in files ending in .stm and .edb. It is not unusual for these files to be many Gigabytes in size, even in a small environment. An important characteristic of this type of data is that receiving even one email can cause over 50% of the data store to change at the block level (due to reindexing and ongoing data maintenance). Similarly, SQL or other databases can change radically at the block level with relatively minor database changes. For this reason, the assumption that change data will be only a few megabytes per day is often quite wrong. This is significant because many internet connection speed requirement calculations are based on this small change assumption.

**The cost of storage & service.** Internet backup services must pay for hard drive space like everyone else. To make a profit they must mark their costs up. Using a backup service is like renting your home versus buying it. The convenience of letting someone else deal with the off-site storage is a trade off with the (typically) higher cost. Many small organizations discover that the monthly recurring cost of a backup service and additional bandwidth requirements is

too much of a financial burden and revert to local backup solutions like High-Rely removable drives. Generally speaking, local backup costs less than Internet service based backup. It is also true that the more data that must be backed up, the larger the cost savings.

**Underestimating the amount of data.** Because Internet backup services are slower than local backup, are expensive, and are priced “per Gigabyte” end users are forced to be selective in what to back up. When estimating the cost of an Internet backup service, realize that data will grow over time and that leaving out certain types of data to save money defeats the purpose of having the backup in the first place. A good rule of thumb is that if it takes more than a day to recreate any of the data, then it's data that should be backed up.

**Lack of Redundancy.** Many Internet backup services do not have any “peering” arrangements. This means they do not make extra copies of your data and store them with a peer (or at another location) to protect against the case where something in their own business fails. Higher costs are generally incurred if you contract with Internet backup providers to duplicate your data to a second data center. By contrast, most local backup schemes employ multiple media. As long as a reasonable rotation and off-site procedure is carried out, several copies of the data will exist, with at least one in a separate location.

**The Hidden cost of bandwidth.** When pricing Internet based backup services, remember that there are often additional “hidden” costs...that of the Internet bandwidth. Many people assume that they will be able to use their existing connection to the internet. However, bandwidth needs for backup are often far greater than expected. If you are unable to transport the changed data in 12 hours (i.e. at night when the office is closed), then the internet link must be upgraded. Here are some common reasons why bandwidth needs are under-calculated that may require you to upgrade your Internet connection:

- Many connections are “asymmetric,” meaning they have much slower upload speed compared to the download speed.

For example, a typical DSL line might have 384Kbps up and 1.544Mbps down. This speed “feels” great when surfing the web, but that is deceiving because it is based primarily on the download speed. It is the upload speed that matters for Internet backup.

- Discrepancies between the rate advertised by your provider and reality of what you actually get. Customers usually won't get the advertised rate in the real world. To test the speed, make sure no one is using your connection and then test your actual speed at one of these sites (you may need to install Java or flash for the test). Many other speed tests exist on the Internet and several should be run to get a baseline speed.

<http://www.dslreports.com/speedtest>

<http://www.numion.com/YourSpeed3/Run.php?QuickStart=MaxSpeed>

- The various transport protocols combine to a **typical 50% overhead**. This means that for you to move one Gigabyte of data you should take the bit rate of your connection and cut it in half for the purposes of determining how long it will take to move this data. Use the link below to an online calculator that will estimate how long a particular size of data will take to transport (we recommend using 50% for the overhead number).

<http://www.numion.com/Calculators/Time.html>

- During the day the connection is used for business operations. Avoid doing calculations that include the hours of daytime operations if you do not want to impact your daily internet speed by using the connection for backup operations.

**The uncertainty of data security.** Although most internet backup services provide data encryption, many customers feel uncomfortable with their mission critical data being sent over the Internet to a relatively unknown data center. Can the data be decrypted? Who will have access to

the data? Is the data center highly secure and using best practices for security? All of these should be considered before sending data offsite.

**No support for Active Directory or open files.**

Some Internet backup services are not designed to handle “enterprise” features such as backing up Active Directory, Microsoft Exchange, SQL databases, or other open files. These services are “always on,” which presents a challenge to backup software because files that are “open” are often locked and cannot be accessed. The result is that some of the most critical files on the server won’t get backed up. Online services that properly handle these special files exist but are more expensive. It is possible that a backup service successfully backs up 99.9% of customer data yet is unable to do a “bare metal” restore. When a server must be replaced and there is no bare metal capability, the Windows operating system, service packs, licensing, printer setups, user passwords, OS configuration, and many mission critical applications must be manually restored. This results in 2 to 7 days of down time for the typical small office.

**Lack of support for hardware re-targeting.** If a server fails catastrophically, the normal course of action is to replace it with a new server. With Windows server operating systems, it is often impossible to do a full restore of the old server onto a new hardware platform. Doing so usually results in a non-bootable server. This necessitates the same painstaking process described in the section above about Active directory and open files.... reinstalling and reconfiguring Windows and then selectively restoring the data only. By contrast, some local backup systems (including imaging products like Symantec’s BESystem Recovery, Storage Craft’s ShadowProtect, or Acronis True Image) allow re-targeting to foreign hardware quickly and easily.

**Who is Responsible?** Many services leave the day to day management of backup in the hands of the customer. Because backup jobs tend to fail over time for a variety of reasons, It is important that services monitor and notify the customer when backups are not occurring. Many services will

send out an automated email but have no human interaction and no way to verify the client has seen the email. Most services will charge extra fees to do a restore or send out a hard drive in an emergency situation. All of these suggest that ultimately the responsibility for good backups still rests with the customer. This is fine as long as the service does not imply that everything is automatic and no user interaction is required. As much time should be spent monitoring results as when the backup solution is local.

**Achieving a timely restore.** Restoring single files and folders using Internet backup service software is generally easy and quick. However, due to many of the issues we have already discussed, doing a “ground up” restore over the Internet is all but impossible. The first issue is the need to load windows and the restore software so that a replacement server can even access the data stored at the data center.. Once this has been accomplished, consider this question: “If doing the initial full backup over the internet was too time consuming, is doing a full restore over the internet to a downed server even practical?” Chances are you will need to request a portable drive be created and sent out, meaning at least a 1 to 2 day delay to gain access to critical data (at additional cost).

**Conclusion.** Consider carefully what the hidden costs associated with Internet backup as well as the advantages and disadvantages. The whole purpose of a backup is to be able to recover quickly and easily when problems occur. Internet based backup solutions can provide this functionality if enough bandwidth is available and if cost is not a major consideration. For many customers, local removable hard drive based backup solutions may make more sense.