



## Crossroads' Strongbox Simplifies Data Protection

Analyst: David Reine

### Management Summary

Some things simply get better with age. For example, *collectibles*, such as coins and stamps, become more valuable as they grow older. Collectors put them into albums, and then put the albums into vaults to secure them and protect them from the elements. If they can be kept in a pristine condition, they become more valuable. There are some items that continue to be used long after their projected useful life. The *Craftsman* screwdriver set from Sears that is guaranteed for life and simply won't break or, perhaps, the item that you wish would break down (so that you could justify buying a new one) just keeps on going. Then there are the items that we save in the attic because we cannot bring ourselves to part with them, Grandpa's old chair or Grandma's old cutlery. Then we turn on the TV to one of the home and garden networks on cable TV and discover that what we have in the attic is not just a memory, it is a valuable antique. We all know the saying, "One man's junk is another man's treasure". Well, in today's world, what we have stored away in our attics and basements literally might be worth a fortune. Do you know what you have in your attic? Do you know what you have stuck in the back of a closet? Unfortunately, most of us do not keep an accurate inventory of our possessions and run the risk of throwing away a valuable antique in tomorrow's trash.

We have similar stories in the data center where, despite all attempts to kill it off, the mainframe continues to thrive. No one is going to say that a mainframe is the cheapest alternative on an acquisition basis; however, it continues to be a good economic value on a total cost of ownership (TCO) basis for many enterprise uses. Another technology that continues to live on is magnetic tape. Despite decades of negative rumors, tape remains the most cost efficient means to store and preserve information in the data center, if not *THE* most efficient, for the long-term on a TCO basis. After 60 years, tape continues to be a primary target for Tier-2 backup and recovery data in many enterprises, even though disk is often used in rapid recovery scenarios. Now, however, with an increased emphasis on the preservation of data for longer periods in order to meet industry and government regulations, tape is being positioned as the most efficient and reliable target for Tier-3 *archive* data, referred to as *WORR* data, or *Write Once, Read Rarely*, as well. With the introduction of LTO-5, with its higher capacity enterprise drives, along with the *Linear Tape File System* (LTFS) technology, tape has resumed its role in terms of the lowest cost medium with the highest capacity, and now with the capability to search by metadata. LTFS, however, has added something new to the open systems tape arena: a means to treat tape as a file system and use metadata stored on tape to assist all users in the search for specific data on the volume.

LTFS can enable the IT staff to utilize tape as an on-line, active archive for Tier-3 data, enabling all enterprises to address the requirement for low cost and highly-reliable, long-term data preservation. This is especially true for the broadcast and entertainment media. One company that is responding rapidly to this requirement is Crossroads Systems. Crossroads has introduced a new tape-based archiving capability that leverages LTFS and internal cache to provide a reliable, long-term, low-cost, high-performance data preservation solution for all enterprises. This product, *StrongBox*, uses an open format to relieve tape of any dependency on proprietary applications to store and retrieve data. To learn more about StrongBox, please read on.

### IN THIS ISSUE

➤ Data Storage in the Enterprise .....	2
➤ Data Center Requirements.....	3
➤ Crossroads StrongBox .....	3
➤ Conclusion .....	5

## Data Storage in the Enterprise

The proliferation of application servers and storage devices throughout the IT infrastructure has had a dramatic effect upon the budget for both enterprise and SMB data center managers. The TCO of these infrastructure components has been out of control due to inefficient servers and the compounding increase in the amount of data being retained. These have created a complex infrastructure that is difficult to manage. Open systems server inefficiency has been resolved by consolidating and virtualizing servers with advanced multi-core x86 microprocessors from Intel and AMD. More-efficient processing power, via these designs, has been the key to a more efficient deployment of server applications, limiting the amount of server sprawl, simplifying the infrastructure, and lowering the TCO of the environment. **Now is the time to address storage efficiency.**

Every enterprise is keeping information on hand for longer periods of time, as the value of that data continues to grow. Data retention policies are being based upon both data value and legal requirements to meet regulatory compliance (such as to address eDiscovery (legal) inquiries), as well as reference activity. Some data files, such as medical images and digital videos for the entertainment industry, may never be deleted, as they may be needed or useful later.

Enterprises are looking for a more effective solution for data preservation and to consolidate historical data into on-line *active archives*<sup>1</sup> to simplify the requirement for the searching and analyzing of historical business data<sup>2</sup>. In an era of economic downturns, reducing the costs of the enterprise data protection and retention environment is becoming even more important, as enterprises either delay or reduce data center budgets. Reducing this outlay will mean more budget availability for storage growth and the development of additional mission-critical applications.

Many enterprises continue to see a growth rate of 40% to 50% annually in the amount of Tier-2 and Tier-3 data being generated by mission- and business-critical applications with Tier-3 data described often as an archive level

for stable (likely never changing), stagnant, or “stale” data that has infrequent access. This is reflected in the IT budget by the demand for more high-performance Fibre Channel (FC) and SAS arrays for Tier-1 storage, SATA disk and high-capacity tape for Tier-2 backup and recovery, and Tier-3 capacity for the long-term preservation of WORM archive data. The growth rate for secondary storage has been even greater than the overall rate, often exceeding 50%.

Let me ask you a question: **How many times is a single version of a file or attachment stored?** If the answer to this is “too many”, then you can understand why the acquisition of more SATA disks and enterprise tape products to be used for *data protection* and *data retention* is out of control. Storage capacity for Tier-2 and Tier-3 data has grown from gigabytes to petabytes, with exabytes of storage predicted for many in 2012. **The challenge of addressing the data protection and retention of this growth, and future growth, and simplifying a complex storage infrastructure, must be addressed, now!**

The requirement for *immediate* recovery of data to replace corrupted or deleted data will never go away. We need to retain some level of Disk-to-Disk (D2D) data protection in order to satisfy this need and maintain the 24x7x365 availability of enterprise data. **However, in many cases, back up to tape in a disk-to-tape (D2T) or disk-to-disk-to-tape (D2D2T) hybrid environment will help control the spiraling costs while still providing a sufficient response.** In order to continue to lower costs in this category, the enterprise needs to deploy higher-capacity cartridges, with higher data transfer rates, and engage in better management of those cartridges<sup>3</sup>.

Despite the additional costs associated with the acquisition and deployment of an adequate secondary storage capability, it is far less than the cost of recovering from a single natural disaster without adequate preparation, or preparing for, and possibly losing, lawsuits for failure to provide an audit trail of embarrassing emails or financial records. The best practices for every enterprise now include the requirements for a highly-efficient data preservation system to help to reduce the TCO for data protection and reten-

<sup>1</sup> See **The Clipper Group Navigator** entitled *Simplified, Online Access to Archived Data – Turning History into an Active Archive* dated June 11, 2010, and available at <http://www.clipper.com/research/TCG2010029.pdf>.

<sup>2</sup> For example, the U.S. healthcare industry is under pressure to retain and protect patient health records accurately, and for a very long time, yet be able to share them safely and promptly.

<sup>3</sup> See the issue of **Clipper Notes** dated December 20, 2010, entitled *In Search of the Long-Term Archiving Solution – Tape Delivers Significant TCO Advantages over Disk*, and available at <http://www.clipper.com/research/TCG2010054.pdf>.

tion. In addition, smart (or active) archiving is needed to lower the cost of writing, searching, and reading, all of this historical data.

### Data Center Requirements

With all of this growth, the data center has a growing need for more efficient management of all tiers of storage in order to lower the costs of data retention and administration. The IT staff must reduce infrastructure complexity with the implementation of a tiered storage environment, perhaps consolidating and virtualizing storage to improve efficiency, and facilitate an archiving environment. Besides compression and data deduplication to limit the number of copies of any given record, *what else must the IT staff do in order to eliminate complexity?*

*Specifically, what does the modern data center need to do to satisfy its data protection and data retention requirements?* If we take a step back, we can see a primary need for an efficient repository to handle all Tier-2 and Tier-3 data for email, database, and all of the enterprise's unstructured data in the form of file storage. More than half of a typical company's unstructured data files are inactive; e.g., they have not been accessed in over 90 days. For whatever reasons, this data must be retained and should be accessible readily. Unstructured data is fast becoming THE reigning storage hog in the data center.

It needs to deliver the most responsible utilization of system resources possible, for rapidly expanding storage. This digital archive must meet the needs of an enterprise-class data center, including an improvement in productivity for the IT personnel. This includes faster access to data through high-performance reads and writes, greater data mobility, reliability, independence from hardware and software constraints, and the tiering of data to the right place at the right time. Moreover, it should be self-healing and, of course, it must be affordable!

This archive must be able to communicate with the various data center servers via a network-attached connection. It needs to break the paradigm of storage having to be proprietary and complex, inflexible, and expensive. In order to achieve the proper level of interoperability, the platform should be a non-proprietary, open solution that is easy to use with a simplified deployment strategy. In order to satisfy long-term business continuity requirements, the solution needs to meet industry standards for RAS compliance, i.e., reliability, availability, and serviceability. From an economic standpoint, the archive should

have a low acquisition cost, and an even lower TCO, including maintenance and administration, in comparison with alternative solutions. In the end, the best solution is the one that meets the requirements at the lowest TCO per unit of data stored at this archival tier. Most importantly, the platform must be highly scalable in terms of processing power, throughput, and the number of files supported, in order to protect the investment that the enterprise is making.

*What features are important to the staff in product selection? What functionality is critical to the successful operation of the data center?* First of all, the archive needs to have a policy-based data management, with granular policies. It must be plug-n-play in a CIFS/NFS environment, in order to enable ease-of-use in an open systems environment. It needs to be able to write files directly to the lowest cost media and read them independently of operating system or application, in order to remove the complexity often associated with proprietary data protection applications. Protection is mandatory, including encryption, replication, and monitoring to meet the business continuity needs of the enterprise.

One company that has been paying attention to the data protection and data retention needs of the enterprise data center is Crossroads Systems in Austin, Texas. With over a decade in data protection and a member of the *Active Archive Alliance*, their goal is to proactively protect and reliably recover data whenever required. Crossroads has been a leader with solutions such as *SPHiNX*, a primary repository for data center backups, including replication and data encryption services, and their *ReadVerify Appliance (RVA)*, which provides a non-intrusive method for monitoring, validating, and reporting on the performance, utilization, and health of the physical tape library environment. Now, Crossroads has launched *StrongBox*, a new tape-based archive appliance to serve as an online archive for Tier-3 data, providing low-cost and highly-reliable data preservation.

### Crossroads StrongBox

StrongBox is an on-line, file-based archive, combining disk with high-capacity LTF5 tape storage for fast, yet cost-effective, long-term archival of Tier-3/unstructured data. It behaves like a standard network file storage system (i.e., NAS) and is compatible with any file-based system architecture. It takes advantage of multi-core CPUs to improve performance and efficiency. In traditional tape environments, accessing

### Exhibit 1 — Key Concepts for Strongbox

- **The Ingest Buffer** – As files are delivered to the system (ingested), they are first written to disk. The files remain on disk until all data movement policies have been applied. Files always are moved to tape. The number of tape copies made of the file, either one or two, is specified by policy. A replication policy may also need to be applied. Upon completion of all data movement policies for a given file, the file's data storage will be reduced to only that required for the Read Cache. The ingest buffer is sized at 1TB from the factory.
- **Read Cache** – After a file is moved to tape, the first portion of the file is retained on disk. When the file is read from the Archive Node, the initial reads are satisfied from disk, allowing time for the tape to be loaded, after which reads are done from the tape. Currently, Read Cache size is fixed at 512K/file from the factory.
- **File Cache** – Files read from tape are kept on disk on the assumption that there will be repeated reads of the same file. The size of the File Cache is a system configuration option and set at 1TB from the factory. Files are aged out of the File Cache, using a least-recently-used algorithm.
- **Read Buffer** – Files being read from tape are read in at full tape speed and stored in memory to be delivered to the client at the speed of the client's requests (over the 10 Gbit link). This frees up the tape drive for other use sooner than would be the case if the tape were only read at the speed of the client's requests. The Read Buffer is sized at 60 GBytes, which will handle multiple parallel read operations, allowing tape operations to occur even if the network link is saturated. The File Cache (disk) will be written to, in parallel, as a secondary process; however, the client data path takes priority and will not be delayed (due to potential disk performance limitations).

Source: Crossroad Systems

files required the user (or application) to read the whole record to disk and then transfer the data to the host. StrongBox enables files to be accessed directly from the tape media.

StrongBox leverages *LTO-5*<sup>4</sup> and *LTFS*, the *Linear Tape File System*, to meet the online digital archiving need for rapid access to low-cost, reliable data preservation for all data centers, both enterprise and SMB. LTFS provides a direct, intuitive, graphical access to data stored on LTO-5 cartridges. There is no need for additional tape management software to access the data. LTFS works with LTO-5 to provide ease-of-use and portability for open systems tape storage. The IT staff can run any application designed for disk files against tape data without concern that the data is actually on tape. While StrongBox currently supports LTO-5 drives, LTFS, and a wide variety of tape libraries from both HP and IBM<sup>5</sup>, it should work with most tape libraries that use these standards.

With caching and performance optimization, StrongBox addresses data access with fast file reads from any qualified tape archive, via “drag

and drop”, along with policy-based data management. StrongBox presents itself to the user as familiar hard disk storage, but combines hard disk for fast storage and retrieval with high-capacity tape for low-cost, long-term storage. Early versions have been sent to enterprises with a voracious appetite for storage and a need for reliable, long-term data archives for unstructured data. Crossroads has focused on healthcare, video broadcasting, finance, manufacturing, and cloud service providers, as well as government and other industries. Early reports from customers have been positive, according to Crossroads.

StrongBox resides in a non-proprietary platform providing data portability and random file access. It is self-healing and has unlimited scalability for seamless capacity expansion and performance tuning for added disk and/or tape drives. It provides high performance and continuous online access. StrongBox employs a policy-based data management with Plug-n-Play CIFS/NFS file shares in order to provide a persistent network view. It uses standard file properties and actions, has a web-based GUI for easy-to-use remote management, and is easy to install. The GUI is simple and intuitive, with a secure web interface. Policies for number of copies, encryption, and replication can be established by volume, with data being compressed when written to disk. The system management

<sup>4</sup> See [The Clipper Group Navigator](#) dated January 29, 2010, entitled *LTO Program Announces Next Gen Tape – LTO-5 Raises the Bar for Tier-3 Storage*, available at <http://www.clipper.com/research/TCG2010002.pdf>.

<sup>5</sup> HP *MSL2024*, *MSL2048*, and *ESLG3*, and IBM *TS3100*, *TS3200*, and *TS3500*.

provides alerts and notifications, both current and historical, including the monitoring and reporting of a wide variety of system parameters, using an integrated RVA for real-time monitoring of tape drives, media, and the library. Strongbox issues alerts for system, drive, and media errors, and does a full tape read/verify, per data center policy. It enables the sorting and filtering of the data.

LTFS write operations are optimized to enable large data transfers, while the load/unload time is minimized. LTFS read operations are also optimized, enabling random file access with a standard NAS mount with no application modifications required. Strongbox provides a persistent file view for all files on tapes loaded into the library. It has a read cache for initial file storage, along with CIFS/NFS timeout management. It has a file cache for file access retry optimization with pre-fetch enablement and direct tape read to eliminate the need for staging back to disk. This enables Strongbox to take advantage of tape's data transfer superiority over most disk solutions and enables the data center to operate their tape drives at full rate, maximizing resource utilization. (See Exhibit 1, at the top of the previous page, to understand the key concepts that Crossroads has used to implement StrongBox.)

Strongbox maintains solid data integrity with support for WORM<sup>6</sup> (to effect erasure control and preservation of the original data), file hash comparisons, and protection from silent data corruption using *T10-PI*<sup>7</sup>. In terms of data protection, Strongbox has a dual-copy policy, along with export for offsite storage, in case of a disaster. It can also replicate both the data and the metadata to a secondary StrongBox archive. (See Exhibit 2, at the top of the next column, for the system reliability characteristics.)

In terms of security, Strongbox has role-based access with secure access to the Web. It has *Active Directory* integration and strong password management. It uses a file hash comparison for data integrity and uses a granular user access log.

Crossroads offers two models of StrongBox, as shown in Exhibit 3 on the next page.

<sup>6</sup> Write Once, Read Many.

<sup>7</sup> *T10 Protection Information* allows a checksum to be transmitted from the application for the drive. The implementation of T10-PI helps to ensure that data is validated as it moves through the data path from the application to the adapter, and to storage, enabling a seamless, end-to-end integrity.

## Exhibit 2 — System Reliability Specs

- Automatic database snapshots
- Database change logs
- Configuration backup
- Redundant, hot swappable power supplies
- RAID 6 + spare for all data storage
- Replication of data and metadata to secondary StrongBox archive
- Protection from silent data corruption

Source: Crossroad Systems

- **A 1U T1 entry model** comes with a single quad-core Intel processor with dual-port GigE and dual-port 6Gb SAS connectivity. It supports up to 200 million files with a maximum file transfer rate of 160MB/second with up to 5.5TB of usable internal disk cache<sup>8</sup>. The T1 model supports up to 72TB of native capacity, along with two 1Gb Ethernet ports for server connectivity, an Ethernet port for management interface, a 1GigE port for replication, and another for user interface.
- **A 3U T3 model** comes in two configurations, one supporting up to 500 million files and the other for up to five billion files. Both have a maximum file transfer rate of 600MB/second and options for multiple GigE, 10GigE, Fibre Channel, and SAS communication ports. The first uses a single quad-core Intel processor (upgradable to two processors), up to 14TB of usable disk, and solid-state disk for disaster recovery. It also comes standard with quad-port GigE and quad-port 6Gb SAS connectivity and an option for 12TB of additional internal storage. The second comes with two quad-core processors, two front-end options, which could be two quad-port GigE cards or one quad-port GigE and one 10GigE. This model also comes with two back-end options for tape library and external storage, which are quad-port 6Gb SAS and quad-port 8Gb FC.

## Conclusion

No one can predict a disaster; however, you have to be prepared for a disaster. Natural disasters, such as hurricanes, build up and are tracked for weeks, enabling all in their path to get prepared. Earthquakes appear out of nowhere; you must be prepared in advance for survival. With the tenth anniversary of 9/11 still fresh in our minds, we all know the destruction that man can create.

<sup>8</sup> After the application of RAID-5.

**Exhibit 3 — StrongBox Models**



**Comes With**

- Single quad core processor
- Dual port GigE
- Dual port 6Gb SAS
- 5.5 TB internal storage, after RAID
- SSD for DR



**< Than 500 Million Files**

**Comes With**

- Single quad core processor \*
- Quad port GigE \*
- Quad port 6Gb SAS \*
- 14 TB internal storage, after RAID
- SSD for DR

\*Defaults

**Options**

- Second processor
- Network Connection (2 slots)
  - 4 \* GigE (1 slot each, choose 1 or 2)
  - 2 \* 10GigE (1 slot, only one allowed)
- Library and/or external RAID connection
  - Quad-port 8 Gb FC
  - Quad-port 6 Gig SAS
- 12 TB Internal storage, after RAID

**> Than 500 Million Files**

**Comes With**

- Dual quad core processor
- 2 Front side Card Options
  - Quad Port GigE
  - Dual Port 10 GigE
    - Only One is Allowed
- 2 Backside Card Options - External RAID and/or Tape Library
  - Quad Port 6 Gig SAS
  - Quad Port 8 Gb FC
- SSD for VPD

Source: Crossroad Systems

Enterprises must also be prepared with an adequate infrastructure for corporate survival. For many, this means keeping backup or disaster recovery copies of our data in a remote, and safe, place. Tape provides the data center with the portability that data protection and data retention requires. It is the best low-cost technology and the standard for long-term data repositories. It is the ideal medium for data archives.

The Crossroads Strongbox represents a paradigm shift in the use of tape that is cost effective and energy efficient, lowering the TCO of the storage infrastructure for the enterprise. It separates tape from its legacy relationship with backup applications, such as *NetBackup* and *Networker*, among others. By using LTO-5 and LTFS, Strongbox has the capacity, scalability, and flexibility required for use as an active archive. If you have a need for a Tier-3 data repository, now is the time to investigate the value of Strongbox.



### ***About The Clipper Group, Inc.***

**The Clipper Group, Inc.**, is an independent consulting firm specializing in acquisition decisions and strategic advice regarding complex, enterprise-class information technologies. Our team of industry professionals averages more than 25 years of real-world experience. A team of staff consultants augments our capabilities, with significant experience across a broad spectrum of applications and environments.

- ***The Clipper Group can be reached at 781-235-0085 and found on the web at [www.clipper.com](http://www.clipper.com).***

### ***About the Author***

**David Reine is a Senior Contributing Analyst for The Clipper Group.** Mr. Reine specializes in enterprise servers, storage, and software, strategic business solutions, and trends in open systems architectures. In 2002, he joined The Clipper Group after three decades in server and storage product marketing and program management for Groupe Bull, Zenith Data Systems, and Honeywell Information Systems. Mr. Reine earned a Bachelor of Arts degree from Tufts University, and an MBA from Northeastern University.

- ***Reach David Reine via e-mail at [dave.reine@clipper.com](mailto:dave.reine@clipper.com) or at 781-235-0085 Ext. 123. (Please dial "123" when you hear the automated attendant.)***

### ***Regarding Trademarks and Service Marks***

**The Clipper Group Navigator, The Clipper Group Explorer, The Clipper Group Observer, The Clipper Group Captain's Log, The Clipper Group Voyager, Clipper Notes,** and "*clipper.com*" are trademarks of The Clipper Group, Inc., and the clipper ship drawings, "*Navigating Information Technology Horizons*", and "*teraproductivity*" are service marks of The Clipper Group, Inc. The Clipper Group, Inc., reserves all rights regarding its trademarks and service marks. All other trademarks, etc., belong to their respective owners.

### ***Disclosures***

Officers and/or employees of The Clipper Group may own as individuals, directly or indirectly, shares in one or more companies discussed in this bulletin. Company policy prohibits any officer or employee from holding more than one percent of the outstanding shares of any company covered by The Clipper Group. The Clipper Group, Inc., has no such equity holdings.

After publication of a bulletin on *clipper.com*, The Clipper Group offers all vendors and users the opportunity to license its publications for a fee, since linking to Clipper's web pages, posting of Clipper documents on other's websites, and printing of hard-copy reprints is not allowed without payment of related fee(s). Less than half of our publications are licensed in this way. In addition, analysts regularly receive briefings from many vendors. Occasionally, Clipper analysts' travel and/or lodging expenses and/or conference fees have been subsidized by a vendor, in order to participate in briefings. The Clipper Group does not charge any professional fees to participate in these information-gathering events. In addition, some vendors sometime provide binders, USB drives containing presentations, and other conference-related paraphernalia to Clipper's analysts.

### ***Regarding the Information in this Issue***

The Clipper Group believes the information included in this report to be accurate. Data has been received from a variety of sources, which we believe to be reliable, including manufacturers, distributors, or users of the products discussed herein. The Clipper Group, Inc., cannot be held responsible for any consequential damages resulting from the application of information or opinions contained in this report.