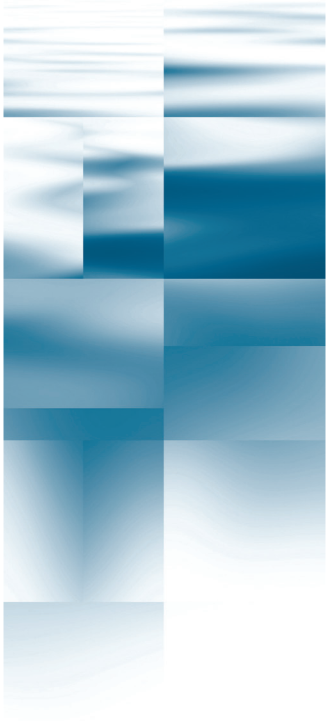
SEPATON®
Intelligent Data Protection



WHITE PAPER

Data Protection for Remote Offices:

Extending the Power of Advanced Data Protection to the Edge of the Enterprise

SEPATON®

Introduction

Widespread disasters caused by earthquakes, hurricanes, tornados, floods, fires and other natural disasters are in the headlines frequently, and each serves as a painful reminder of the need for IT departments to have a comprehensive disaster recovery (DR) strategy that includes redundancies both within and across multiple data centers.

In the corporate data center, most IT departments also have a solution for another important aspect of DR: data protection. Data center data is backed up routinely and replicated to a remote site to enable recovery from any localized disaster. Data that exists outside of the main data center(s), such as in remote offices, however, is often not so well protected.

With today's global and mobile workforce, an increasing amount of data resides in remote offices and even on the laptops, tablets, and smartphones carried by the employees there. Remote offices, large and small, often have local servers running general applications like email and file sharing, as well as specialized ones for order processing and customer support. Gartner estimates that as much as 60% of all enterprise data resides at remote offices, and also reports that 69% of the respondents to their survey are currently unsatisfied with their current remote office backup strategy.

According to IDC: "Historically, remote offices have been treated as standalone islands, without centralized management, policy, or visibility. Limited on-site IT resources, costly WAN bandwidth, and administrative-intensive and failure-prone tape hardware resulted in inadequate and infrequent backup operations, thus compromising operational recovery."

Treating remote offices as islands may have been adequate when the primary cause of data loss was the rare hardware failure, such as a disk crashing. But a significant amount of data loss continues to be caused by less sensational problems, such as a file being accidently overwritten or a database getting corrupted.

Every CIO should ask: Would our DR strategy have enabled us to recover completely from a widespread disaster that destroys both the original data source and the local backup copy? Chances are the answer is: No—at least for some (or maybe most) of

Remote Office Backup Requirements

The data stored in remote offices should be as well-protected as the data stored in the central data center. Specifically, this requires:

- The same SLAs, including frequency, retention and off-site replication
- Data protection that can be managed remotely by data protection experts from the central data center
- Ability to recover files and directories quickly from the most recent backup, while maintaining many additional generations of data for long-term retention.
- The same policies should be implemented remotely and in the central data center
- Ultra-reliable software and hardware, so minimal local IT intervention is required ("set and forget")
- Fast deployment and configuration (in hours, not days)
- Minimal requirements for data center infrastructure (special power, air conditioning, etc).

the data residing in the remote offices. CIOs should also ask: What is the risk of data loss or theft in our remote offices? Chances are, with the traditional practice of storing tapes off-site, the risk is unacceptably high.

This white paper examines the three common strategies for backing up data in remote offices against the same set of stringent criteria used for the enterprise data center, which highlights the importance of data replication, security and centralized management. The next section describes how to achieve a consolidated, enterprise-wide approach to data protection for remote offices using the WAN with disk-based backup systems. The final section highlights the ways the SEPATON S2100-DS3 data protection platform enables enterprises to extend the power of the advanced DeltaScale architecture to the edge of the enterprise – from central data center to remote offices.

Remote Office Backup Alternatives

Before examining the strategies typically used for backing up data in remote offices, it is prudent to review some considerations for distributed data protection. This list is not intended to be a comprehensive set of best practices; rather, it includes only those considerations that constitute the main factors in a remote office backup strategy. In no particular order, these considerations include:

- Applicable regulatory compliance and/or business policies for record retention, as well as for data privacy and security (for both personally-identifiable and business-confidential information)
- The maximum acceptable amount of data loss (usually expressed as a recovery point objective or RPO)
- The maximum time allowed for restoring data from either the most recent backup or an archive (usually expressed as recovery time objective or RTO)
- The ability to periodically test backup and restore procedures
- Provisions for storing backups off-site (while still satisfying the RTO)
- Provisions for backing up critical data stored on desktop PCs and portable devices in the remote offices (the latter possibly as part of an enterprise Mobile Data Management or MDM strategy)
- Comprehensive, centralized management that minimizes or eliminates any dependency on non-technical personnel

While there are numerous different technologies available for protecting and storing remote office data, they distill into three basic alternatives: tape drives, the Wide Area Network (WAN) and specialized disk-based backup systems. Variations of these three, such as removable disks and the cloud, will also be considered. Given its historical dominance in backing up data, tape is covered first.

Backing Up to Tape

The main advantage of tape as a backup medium is its low cost. Tape cartridges are as inexpensive as they are compact, enabling them to be accumulated and stored (off-site, of course) quite cost-effectively. The tape drives are also inexpensive and can even be built into a remote office server by its manufacturer.

The *Information Week Analytics 2011 Backup Survey: New Possibilities for Data Protection* report summarizes the problem at remote offices: “Traditional tape backup systems have long delivered unsatisfactory results in branch offices. These sites are frequently lacking anyone with the technical expertise to change tapes and check backup status. As a result, tapes don’t get changed or sent offsite as scheduled, and backup failures aren’t addressed in a timely fashion.”

IDC summarizes the disadvantages backing up to tape as follows: “The challenges with this traditional approach are many. This process is subject to human error when nontechnical office staff manage backups, rotate tapes, or initiate recovery. The risk of data compromise due to the removable nature of tape cartridges presents risk of loss or theft of sensitive data. The high rates of both tape media and drive failure place backup in jeopardy and scarce technical resources must travel to the remote office to triage and resolve a problem. In locations where server, application, and file system configurations have not been protected, recovery requires a lengthy and complete rebuild of preexisting configurations from scratch… Based on recent IDC research, the most common types of failures associated with data protection are due to bad tape media, partial job completions due to the closing of the backup window, configuration errors, network timeouts, and hardware errors such as tape-drive loading, ejecting, or read/write errors."

The removable nature of tape is both its biggest advantage and its major weakness. Someone at the remote office needs to mount and remove the cartridges. At sites without IT personnel, this daily task falls to someone who already has another full-time job. And when s/he is sick or goes on vacation, someone else needs to fill in. Tape cartridges, once removed, can also be lost or stolen, which raises data privacy concerns and the need for encryption. According to IDC, "For firms that decide to encrypt backup data on tape cartridges, the challenge of managing encryption keys can be significant for remote offices with limited IT staff."

Another problem with tape is its slow performance. As a serial medium, the tape must be wound and rewound to one or more specific points before the backup or recovery can begin. This cannot occur, of course, until each tape's file directory has been located and read—often only to discover that the desired file must be on another tape!

IDC notes one additional problem with tape: "Older data written in a legacy backup application format may need to be restored for legal or business reasons. This can require a firm to maintain the legacy backup product for restores while deploying the new data protection approach for backups on a go forward basis."

Removable and external disk drives are also now used as a backup medium. While more expensive than tape cartridges, disks offer far superior performance. The problems, however, remain largely the same with the burden of the backup continuing to fall on non-technical employees at the remote offices.

The operational challenges with tape make this choice increasingly unsatisfactory for most organizations pursuing best practices in enterprise-wide data protection. Indeed, during many widespread disasters, the backup tapes stored in the trunk of a car, at a home or even in a safe deposit box at a local bank are lost or destroyed.

Backing Up Via the WAN

To ensure that data is preserved and protected from widespread localized disasters, replicating or directly performing backups across the WAN has become an increasingly common practice. The major advantage of WAN-based backups is that they always reside off-site; the wider the WAN the better the data protection in widespread disasters. The major disadvantages are the cost of bandwidth and the relatively poor performance over typical WAN implementations.

The importance of some form of data reduction to improve performance with WAN-based backups is summarized in the *Information Week Analytics 2011 Backup Survey: New Possibilities for Data Protection:* "Without data reduction, most of us couldn't afford sufficient WAN bandwidth to replicate even daily changes. The data reduction rates provided by the combination of deduplication and compression (which reduces the size of data by removing small repetitions) can let IT squeeze remote office data into a much more affordable WAN connection." The same *Information Week Analytics* report also summarizes the challenge of handling this important function in software: "Data deduplication is a compute-intensive function, so media server deduplication performance is limited when compared with the performance of backup appliances."

The throughput performance limitation inherent in the WAN can make it particularly challenging to accommodate shrinking backup windows or to satisfy Recovery Time Objectives during full restores. It also makes it more difficult to periodically test backup and restore procedures at the remote site.

The destination of the data in the WAN can be either a private or public infrastructure. If private, such as at an enterprise data center, the organization enjoys a high level of security and control. If public, such as in cloud-based backup services, the solution can be quite affordable (at least on a small scale), but can raise security concerns. As WAN-based backup scales—with more sites, larger data sets and/or longer retention periods—private solutions become more cost-competitive with public clouds. This creates a third viable alternative for larger organizations.

Backing Up to a Disk-based Backup System

The critical importance of data protection has led to the use of disk-based backup systems designed to improve performance and reduce cumbersome tape media management. These systems normally create virtual tape library (VTL) environments for compatibility with existing physical tape libraries and leading backup applications, such as Symantec NetBackup, IBM Tivoli Storage Manager, EMC Networker and others. As an on-site system, each offers some means of replicating data via the WAN. The best systems also offer sophisticated deduplication engines and other advanced features like rapid recovery and secure erasure.

Disk-based backup systems have numerous advantages, the main one being the highest possible performance. Backups proceed quickly at up to 1.5 Gigabytes/second per processing node, and with an on-site copy (that is also replicated across the WAN, of course), recovery times for single files or entire backup sets are greatly accelerated. The ability to backup files quickly enables organizations to set (and implement) more aggressive RPOs. Built-in deduplication and replication capabilities also improve performance and extend retention timeframes.

The performance advantage of disk-based backup systems are summarized in the *Information Week Analytics 2011 Backup Survey: New Possibilities for Data Protection:* "First, a disk array, VTL or disk-based backup appliance isn't limited to running only as many simultaneous backup jobs as there are drives in the system, the way tape libraries are. Disks don't need to restart and can accept data at any rate up to their limits. The biggest advantage of a backup-to-disk architecture, however, is that since the data is online, backup administrators can satisfy most restore requests in a matter of minutes instead of days. Because restores of one or a few files are significantly more common than full system restores among our survey respondents, this capability may be worth the price of admission."

Remote Office Backup Pain

The causes of pain (and "hidden costs") in remote office backup strategies today are all rooted in the use of tape, which requires manual intervention. IT personnel often find it quite difficult to resolve problems or to explain procedures to remote-office personnel who either cannot understand or are incapable of complying for one reason or another. The remote office staff is equally frustrated, and deeply resents this intrusion into their busy work days.

Then comes that dreaded, albeit rare occasion when a full or partial restore is necessary. Perhaps the tape cannot be located, either because it was lost or was labeled with the incorrect date. Or the files on the tape might have been corrupted or overwritten because it was not handled properly. Or the most recent available backup may not be all that recent, causing a significant loss of data.

For some organizations, though, the ability to completely secure and remotely control backup/restore provisions trumps the performance advantages. These organizations recognize that data is strategic to the organization, and that its protection cannot be a part-time job for employees with other responsibilities. Strict regulations concerning data privacy and retention are also motivating organizations to adopt centralized control.

IDC believes that "… firms will increasingly evaluate and deploy new disk-based data protection, backup, and replication technologies in remote offices to ensure that data can be quickly recovered and managed from a centralized point of control."

With the advent of enterprise-class deduplication, the capital expenditure for disk on a cost-per-gigabyte-stored basis is now comparable to tape. But CapEx is only part of the total cost of ownership in an enterprise-wide backup solution. Ongoing operating expenditures are lower for disk-based systems than they are for both tape and cloud-based solutions, which, as noted above, fail to scale cost-effectively. Over time, the OpEx savings eventually exceed the original CapEx, resulting in a lower total cost of ownership.

"Set and Forget" Data Protection for Remote Offices

The challenge for most organizations endeavoring to achieve enterprise-wide data protection is the available budget and staffing limitations. Indeed, the top data protection challenge identified in the *Information Week Analytics 2011 Backup Survey* was "The cost of backup systems and procedures." Most organizations have other budgetary pressures and priorities, leaving precious little funding for backup. This could explain why most organizations use the cheapest available medium for remote offices: tape. The high "hidden cost" with this approach, however, is in the manual and error-prone procedures.

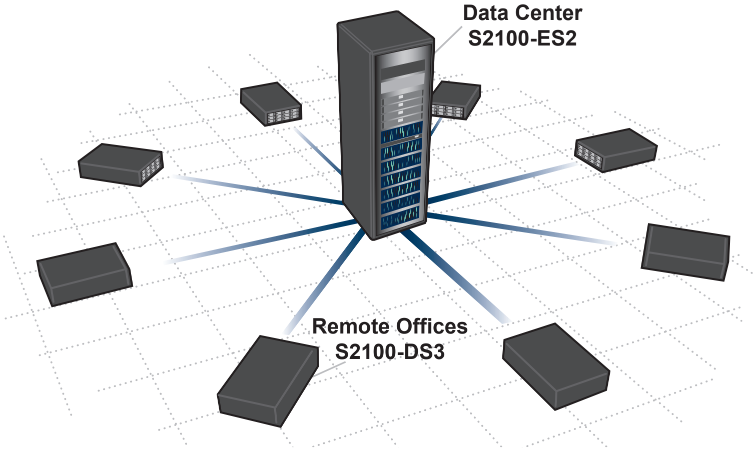
The challenge, therefore, becomes implementing a backup solution that balances capital and operational expenditures (including procedures) in a way that achieves solid protection at an acceptably low total cost of ownership (i.e. one within the available budget) and low level of complexity. In general, the more a system can be completely automated and centrally controlled, the more cost-effective it becomes. This holds true for backup and recovery.

Another unfortunate reality in most backup strategies is that any (potentially costly) flaws are normally only discovered when a problem occurs. Someone in the remote office forgot to change or load a tape, for example, or the required tape is nowhere to be found. Because automated procedures can be verified and tested until perfected, it is possible to make them fool-proof. The result is a "set and forget" approach to data protection.

Achieving centralized control of a fully automated enterprise-wide data protection strategy has two fundamental requirements. The first is the elimination of tape in all remote offices. As shown in the previous section, tape-based backup is error-prone and unreliable owing mostly to its dependency on manual procedures by non-technical personnel (see sidebar on Backup Pain). The second requirement is enterprise-class backup software that scales from the largest data center to the smallest remote office.

Fortunately, both of these requirements are easily satisfied today. Pervasive WAN connectivity and the availability of cost-effective disk-based backup systems make it possible to eliminate tape entirely in all remote offices, regardless of their size. These disk-based systems can also be managed remotely and fully from a central location, such as the primary data center. And enterprise-class backup software, such as Symantec's NetBackup and IBM's Tivoli Storage Manager, all now support remote offices of any size via the WAN and/or disk-based backup systems.

The Hierarchical Hub-and-Spoke Topology



The most efficient topology for automating and centrally controlling enterprise-wide backup and recovery is a hierarchical hub-and-spoke. The central hub exists at the main data center, with another hub at a disaster recovery site or a secondary data center. Below this top level exists some number of intermediate nodes at the larger remote facilities, such as major divisions or regional offices. These nodes become spokes for the central hub(s), and also serve as hubs for some number of smaller remote offices, hence the hierarchy. More layers are possible, but the basic hierarchical topology remains the same.

The amount of data increases with each successive layer in the hierarchy. The intermediate hubs must have sufficient capacity to backup and archive the data housed on-site, as well as the data from all its subordinate spoke sites. To provide adequate protection, all of the intermediate hubs will replicate their data to one another or, more likely, to the central hub(s).

Determining the capacity required in the data protection platforms deployed in the central and intermediate hubs, as well as for any at additional hub layers that might exist in the hierarchy, should be a fairly straightforward exercise. It should also be fairly straightforward to determine whether a remote office requires a disk-based backup system. Factors here involve the size of the daily and weekly backups (taking into account the deduplication effectiveness), the backup window available, the bandwidth of the WAN connection, and the Recovery Time Objective, which must also take into account the WAN bandwidth.

Enterprise-class Data Protection

There are a variety of disk-based backup systems available today, each supporting a different set of capacities and capabilities. Implementing the fully automated, enterprise-wide data protection strategy described here, however, requires a solution that is enterprise-class with:

- Seamless compatibility and proven interoperability with the backup software
- Different size platforms for cost-effective deployment in different facilities
- Scalability to expand capacity as required without a costly upgrade
- Robust deduplication and replication to minimize WAN bandwidth utilization
- Reliability based on the use of RAID and other system-level redundancies
- Rapid-restore of both individual files full data sets.

Protecting data in the smallest remote offices may involve a choice between increasing the WAN bandwidth and installing a disk-based backup system, which can easily replicate its backups in a way that does not interfere with other WAN-based applications. The former has a higher OpEx; the latter has a slightly higher CapEx. For example, a T1 or Digital Subscriber Line link with 1.5 Mbps of symmetric bandwidth (upstream for backup and downstream for restore) can transfer around 500 Megabytes per hour, taking into account protocol overhead and assuming exclusive use of the link during the backup or restore operation. Depending on the volume of deduplicated data involved daily and weekly, this level of throughput may be sufficient for most if not all of the smaller remote offices.

Automating data protection enterprise-wide in this hierarchical hub-and-spoke topology has two significant benefits. The first is that it minimizes and can even eliminate the potential for human error. By using disk-based systems, there are no tapes to load or unload, or to mislabel or misplace. Replication across or up the hierarchy is also automatic, affording constant protection against widespread disasters. The second benefit is its cost-effectiveness. Time is money and mistakes are costly, and both must be factored into the total cost of ownership equation. By eliminating manual procedures and the inevitable errors, there is simply no other approach that is as dependable and affordable as “set and forget.”

Enterprise-wide Data Protection from SEPATON

The goal of enhancing enterprise-wide data protection and disaster protection while minimizing total cost of ownership is the strategy behind SEPATON’s solution. The SEPATON S2100-DS3 data protection platform extends the power of SEPATON’s advanced data protection technology to remote offices at the edge of the enterprise. SEPATON’s advanced deduplication and bandwidth-optimized replication enable enterprises to replicate from the DS3 quickly and easily to (and from) an enterprise-class SEPATON S2100®-ES2 data protection platform in a central data center and/or to a DR site for company wide data protection and disaster protection.

The SEPATON S2100-DS3 brings the power of SEPATON’s advanced DeltaScale architecture to the edge of the enteprise.

The DS3 enables the enterprise to achieve the same level of data protection in remote offices that has long been available in the data center. It has the power to back up more data in less time at rates up to 1.5 Gigabytes/second/node, and to keep more data online longer—all in a secure, high-availability environment. The platform is available starting with usable capacities of either 10 Terabytes or 20 Terabytes, and can scale up to 80 Terabytes; built-in deduplication and compression capabilities increase the capacity by orders of magnitude. The S2100-DS3 is a disk-based system that can appear to backup software applications as a tape library for seamless integration into existing environments.

The DS3 is built to complement SEPATON’s industry-leading S2100-ES2, built on the SEPATON DeltaScale data protection platform. The ES2, which is typically deployed in enterprise data centers and at disaster recovery sites, has a capacity that scales to

1.6 Petabytes or 1600 Terabytes (before deduplication). In addition to its massive, grid scalability, the ES2 also delivers industry-leading backup/recovery performance of up to 43.2 Terabytes/hour.

Both the DS3 and the ES2 feature DeltaStor® deduplication and DeltaRemote® replication. DeltaStor's sophisticated ContentAware™ data deduplication technology dramatically reduces capacity requirements without slowing performance, enabling more data to be kept online longer. DeltaRemote works in conjunction with DeltaStor to reduce replication bandwidth requirements by as much as 97 percent. SEPATON's byte-level delta differencing technology enables more data to be replicated faster than any competing solution—and bi-directionally—between and among sites. Bandwidth throttling makes the most efficient use of the relatively modest data rates of lower cost WAN links, resulting in additional ongoing savings. DeltaRemote remote replication software also helps meet demanding RTOs with industry-leading restore times based on SEPATON's DeltaCache Recovery™ technology, which enables immediate restoration without the reassembly required by many other solutions.

Both the S2100-DS3 and the S2100-ES2 systems maintain full compatibility with popular enterprise-class backup applications to facilitate centralized, enterprise-wide management, while simultaneously overcoming some of their limitations to improve performance.

The SEPATON Advantage

By integrating scalable deduplication with bandwidth-optimized replication, SEPATON affords the industry's best and most effective utilization of WAN resources across the enterprise. The result is robust yet cost-effective data protection as part of a comprehensive disaster recovery strategy.

A critical aspect of enterprise-wide data protection is the ability to consolidate and centrally manage resources, while tracking usage back to the department or remote office consumer of data protection resources. SEPATON systems enable the creation of storage pools within a system and to assign a variety of characteristics to each pool, including disk types, deduplication configurations, replication priorities, backup policies, and/or backup applications to meet the unique needs of different departments or remote offices—from the main data center to the remote sales office. Equally important is the ability to manage all of these storage pools from a single, unified management console and to track usage of capacity at the department or remote office level.

In addition, SEPATON's powerful web-enabled management console enables centralized management of all SEPATON systems throughout the enterprise. Managers can use DeltaView Portal to track efficiency and system status of all key operating functions - backup, deduplication, replication, and restore and report on system utilization to optimize system efficiency.

Conclusion

The increasing amount of data residing in remote offices, including on the mobile devices carried by the employees there, make it prudent for organizations to implement an enterprise-wide backup and recovery strategy. The traditional approach, with inexpensive tape and costly, error-prone procedures, is simply no longer adequate.

Large enterprises have too much remote offices data that is unprotected. The costs and risks of losing that data could be catastrophic to the business. Until now, most organizations were forced to balance this risk against the cost of the equipment and personnel needed to protect this data with physical tape or disk-based backup systems. With the SEPATON S2100-DS3, enterprises now have a fast, convenient, cost-efficient way to deliver the same high levels of data protection throughout the enterprise. These systems deliver the ease-of-use, high performance, deduplication and bandwidth optimized replication needed to make disk-based backup cost-effective in remote offices and departments.

The pervasiveness of WAN connectivity combined with disk-based backup systems purpose-built for remote offices, like the DS3, make it more affordable than ever to implement a “set and forget” backup and recovery strategy that is fully automated and centrally controlled. The result is better data protection across the entire enterprise without the “hidden expenses” associated with tape.

About SEPATON

SEPATON is the recognized leader in enterprise-class, disk-based data protection platforms. SEPATON offers the only data protection platform that delivers the grid scalability, high performance, scalable deduplication, and bandwidth-optimized replication that large enterprises need to address their data protection challenges. The company operates globally and has offices throughout the United States, Europe and Asia and partnerships with leading resellers and OEMs worldwide.