

MicroScan De-Duplication Minimizes Replication Bandwidth and Storage Requirements

Abstract: MicroScan™ is an integral part of the FalconStor Software suite of products including Continuous Data Protection (CDP), Continuous Data Replication (CDR), FileSafe™ host backup software, and DiskSafe™ host-based CDP software. FalconStor CDP and CDR with MicroScan technology can dramatically reduce your cost structure for data protection. This document explains how MicroScan works to minimize duplicate data and how this can save considerable bandwidth, storage resources, and overhead.

MicroScan™ is a patented data de-duplication mechanism within the FalconStor® IPStor® platform and related products that eliminates any exaggerated block-level changes due to inefficiencies at the application and file system layer to minimize the amount of data transferred. As a result, only real changes at the granularity or disk-sector level (512-byte) are transferred. This helps reduce bandwidth and associated storage costs for disaster recovery (DR), or any time data is replicated from one source to another.

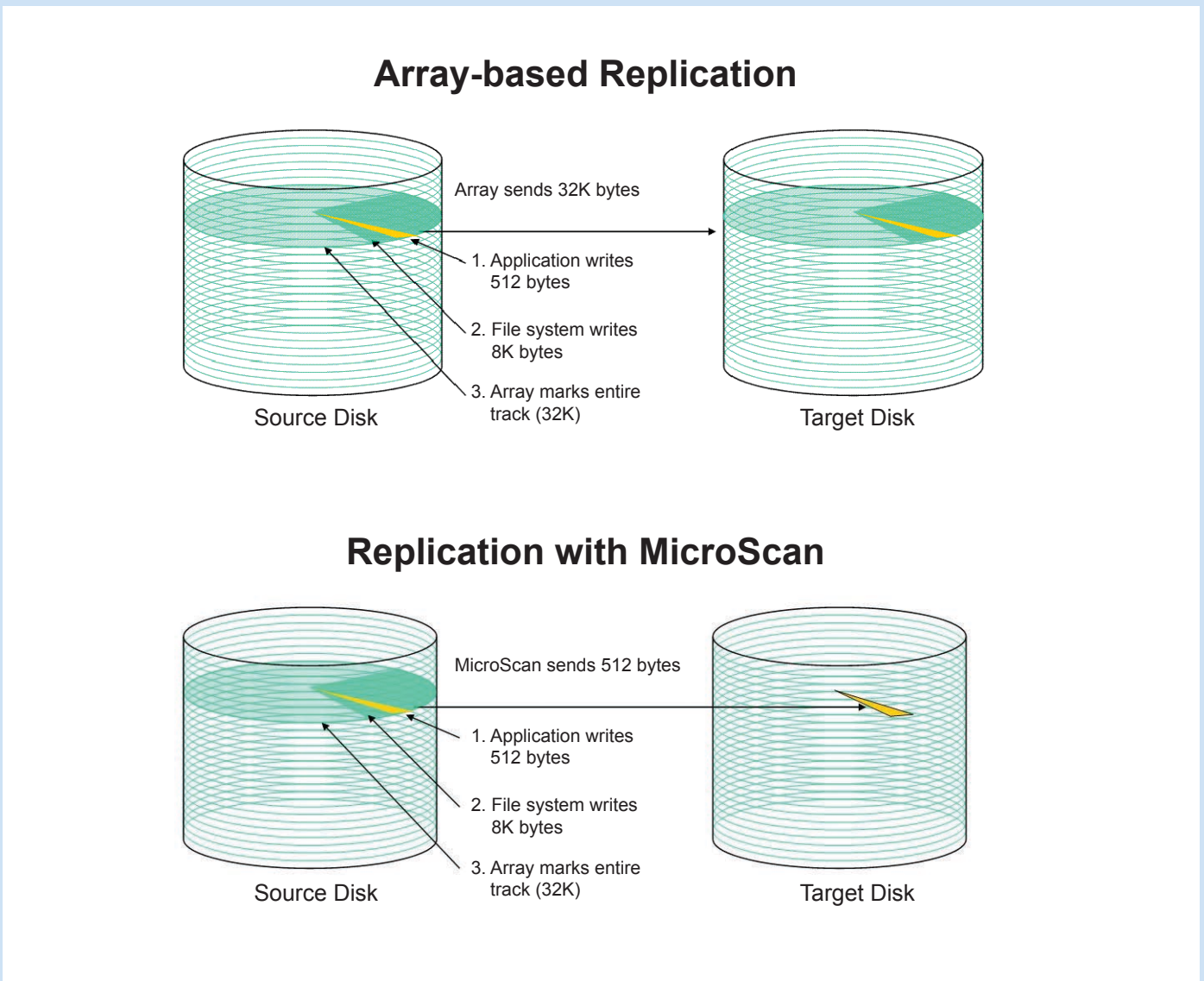
MicroScan is built upon a hash-based “Reverse Challenge-Respond” algorithm which examines the differences between source and target data at the sector level prior to transmission over a network. MicroScan is vendor-agnostic, allowing it to work with any standard storage array. It works at the disk-sector level, and can be applied to any data source.

If the source and the destination disks are connected by a high-speed Fibre Channel (FC) link, simple mirroring without MicroScan may be adequate. However, in a long-distance, high latency, and low bandwidth scenario, MicroScan can be used to substantially reduce the amount of data transmitted by eliminating any redundancy down to the 512 byte sector level.

MicroScan effectively eliminates the data multiplication factor introduced during low writing granularity, where a change of even one bit will lead to a whole cluster of several kilobytes to be overwritten, with mostly redundant data or “white space”.

An example of this effect can be seen in Figure 1, which shows the differences between traditional data replication updates using a storage array-based solution versus MicroScan replication.

Figure 1: Track-based versus sector-based replication



Typical storage array-based solutions track changes track-by-track. This means that when a file system writes 8K of data, it will send an entire track. In contrast, MicroScan scans the 8K of “new” data, concludes that only 512 bytes (1 sector) has changed, and only sends 512 bytes over the network. Note that all open system disks are formatted to a 512-byte sector standard, which is the smallest possible write to a disk drive.

By comparing the replication granularity provided by MicroScan with other replication solutions, we can see where dramatic bandwidth savings are possible:

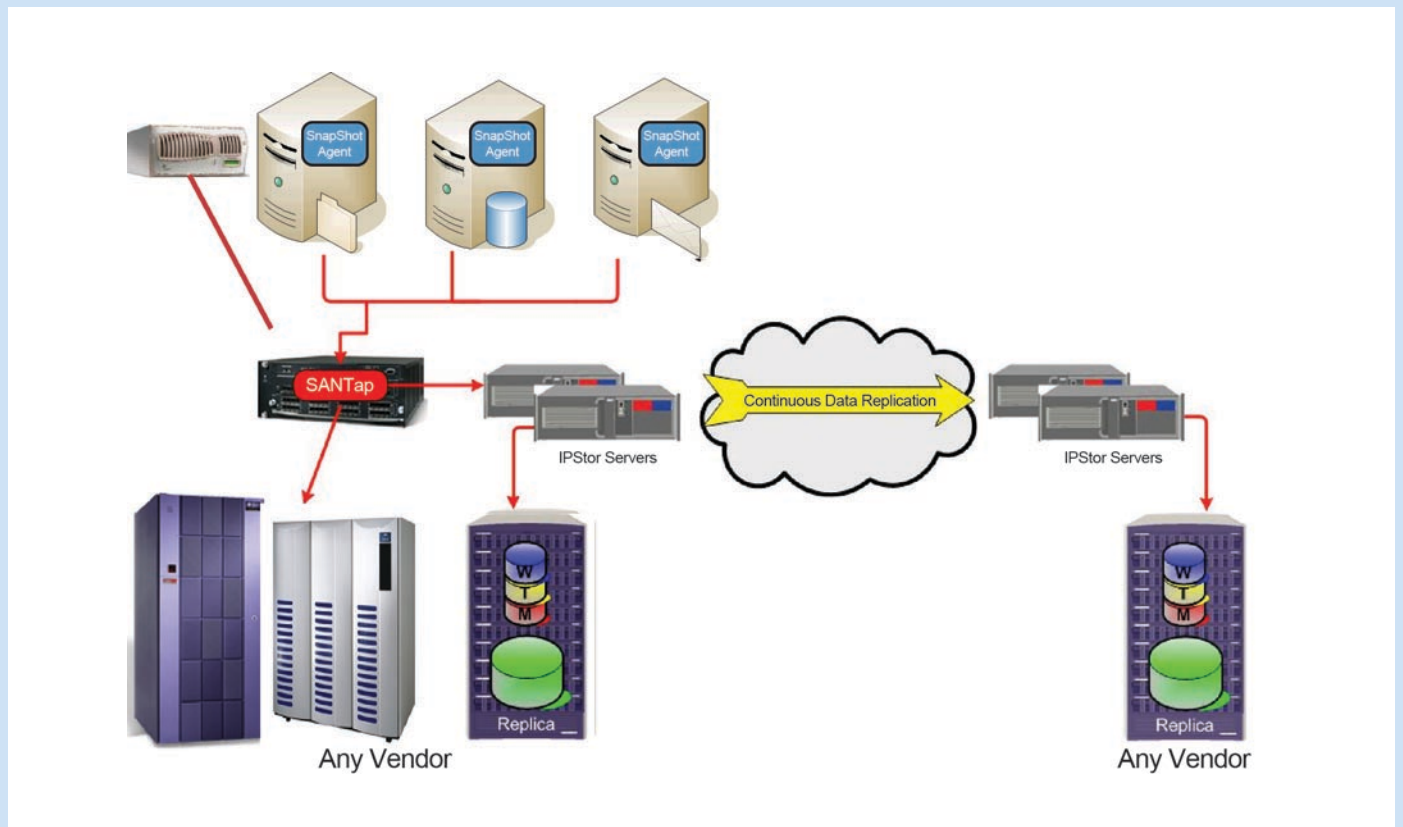
Minimum amount of data sent over a network link for a single 512-byte write by the application:

- Vendor A = 32KB (Array-based)
- Vendor B = 16KB (Appliance-based)
- Vendor C = 8KB (Storage-appliance-based)
- FalconStor Replication with MicroScan: 512 Bytes

Therefore, FalconStor Software with MicroScan is:

- 64 times more efficient than Vendor A
- 32 time more efficient than Vendor B
- 16 times more efficient than Vendor C

Figure 2: Storage-agnostic replication with extreme efficiency



The FalconStor delta replication solution with MicroScan is certified for solutions such as Cisco SANTap (Figure 2) or the IPStor storage virtualization platform, and can be implemented with them to provide a cost-effective solution across all storage platforms.

Customers can also leverage FalconStor® CDP in conjunction with MicroScan based replication to provide a complete solution for data protection.

When performing offsite replication, organizations can use MicroScan along with the FalconStor® TimeMark® Snapshot option to minimize snapshot storage requirements at the DR site. If 16 times saving of bandwidth can be realized using MicroScan, then the amount of snapshot storage is also reduced by 16 times.

MicroScan Case Study: Large Law Firm

In order to understand the impact MicroScan has in a real-world enterprise environment, we worked with one of our customers on a long-term study of data change rates, and the effect on data replication and storage. The study was done across their four major data centers in the US in Houston, New York, Washington DC, and Atlanta. Over an 83-day period, the measured data change rate based at the file system level across all four data centers was well over 100 terabytes (101.710TB to be exact).

Figure 3 shows the data changes that would have been replicated and stored at the DR site for the company’s 90-day retention requirements, and also shows the impact of MicroScan. Over the

83-day period, approximately 1.2TB of delta changes were created per day, with MicroScan technology removing over 1TB of redundant data on a daily basis. This translates to an effective savings of 84% WAN bandwidth and storage requirements.

Figure 3: MicroScan case study

Legal services firm benchmark, November–January 2006

| Site | Real Delta GB | Filtered by MicroScan | Actual GB Transferred |
|-----------------|---------------|-----------------------|-----------------------|
| Houston | 4,869 | 4,017 | 852 |
| NYC | 16,176 | 13,409 | 2,767 |
| WDC | 7,289 | 5,581 | 1,708 |
| ATL | 73,376 | 62,527 | 10,849 |
| Total | 101,710 | 85,534 | 16,176 |
| # Days | 83 | | |
| Average per day | 1,225 | 1,031 | 195 |

| | FalconStor | Vendor A | Vendor B | Vendor C |
|--|------------|----------|----------|----------|
| Min. storage unit to track changes on disk | 512 byte | 4 KB | 16 KB | 32 KB |
| Relative efficiency | 64 | 8 | 2 | 1 |

84% bandwidth and disk savings of 195GB vs 1.2TB per day!

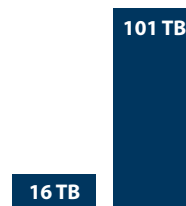
To put that into perspective, we can see from Table 1 that MicroScan was able to save this customer \$600,000 USD in network bandwidth costs in their first year alone. We can also see that since FalconStor data replication with MicroScan is applicable at the fabric level, and is therefore storage-agnostic, the customer also eliminated the need for array-based replication licenses, and was able to leverage lower cost modular SATA storage at their DR location, enabling a true tiered storage solution.

Table 1: MicroScan Network ROI

| | |
|---|---------------|
| Data | 1.2TB |
| WAN link | T3 |
| Normal replication time | 68 hours |
| With MicroScan | 10.7 hours |
| WAN needed without MicroScan to replicate in 10.7 hours | 2 x OC3 |
| MicroScan annual WAN cost savings | \$600,000 USD |

16TB vs. 101TB required at DR site for storing 83 days worth of recovery points on disk:

Tier 1 storage @ 40K per TB fully burdened cost x 101TB = 4.04 million
 Tier 2 storage @ 20K per TB fully burdened cost x 16TB = 320 K



Storage capacity savings: 630%

Instead of having to buy another 101TB of Tier 1 monolithic storage for the DR site, the customer was able to use lower cost Tier 2 modular storage, for a total savings of \$3.7 million USD (based on the fully burdened costs indicated above).

Summary

MicroScan is an integral part of FalconStor technology. FalconStor CDP and CDR with MicroScan technology can dramatically reduce cost structures for data protection. MicroScan enables TimeMark snapshots to be extremely efficient, introducing further cost reductions while improving overall recovery point objectives (RPO) and recovery time objectives (RTO) for your company.

For more information, visit www.falconstor.com or contact your local FalconStor representative.

Corporate Headquarters
 USA
 +1 631 777 5188
sales@falconstor.com

European Headquarters
 France
 +33 1 39 23 95 50
infoeurope@falconstor.com

Asia-Pacific Headquarters
 Taiwan
 +866 4 2259 1868
infoasia@falconstor.com



Information in this document is provided "AS IS" without warranty of any kind, and is subject to change without notice by FalconStor, which assumes no responsibility for any errors or claims herein. Copyright © 2007 FalconStor Software. All Rights Reserved. FalconStor Software, FalconStor, TimeMark, MicroScan, and IPStor are trademarks or registered trademarks of FalconStor Software, Inc. in the United States and other countries. All other company and product names contained herein are or may be trademarks of the respective holder. WP071121.