

# Storage Networking over a Metro Network

## Drivers for Transporting Storage over Metro Networks

Storage-area networks (SANs) are rapidly gaining acceptance as they promise to increase efficiencies and reduce capital and operational costs (CapEx and OpEx). At the same time, recent events have increased the awareness of the need for disaster recovery and business continuance plans. When put together, these trends create a need for interconnectivity between disparate storage “islands,” and encourage transport solutions that are specifically geared towards carrying SAN protocols over metro transport networks.

The main motivations for this trend are:

### 1. Regulatory constraints [1]:

- *Financial services*—Regulation that requires broker and dealer institutions to recover critical functions the same business day a disruption occurs. It also mandates backup data centers on separate grids, and recommends 200–300 km (120–180 miles) separation between primary and backup sites.
- *Health Care*—The U.S. Health Insurance Portability and Accountability Act (HIPAA) regulation addresses payors, providers, and clearinghouses (insurance organizations), and describes security policies and procedures meant to ensure secure access, transmission, and retention of personal health information.

- *Life sciences and pharmaceutical industries*—Regulation that addresses various types of companies involved in the industry, including biotechnology, medical equipment, and food and beverage manufacturers. These rules describe the types of records that must be retained, in addition to discussing the use of electronic systems and records in place of paper or manual systems.
  - *Government*—U.S. Department of Defense (DoD) regulation that addresses all agencies within the DoD and certifies which applications or technology solutions an agency may implement to manage records.
- ### 2. The cost of downtime—Studies [2]
- show that each hour of downtime costs millions of dollars for businesses. For example, an hour of downtime costs an average of US\$6.5 million for financial brokerage firms, and US\$2.6 million for credit card authorization. This gives rise to two approaches:
- *Business continuity plan*—A plan to recreate lost portions of the business, including critical processes, data, technology infrastructure, staffing, workload transfer, and partner and customer communication.
  - *Disaster recovery plan*—This is a subset of business continuity; a plan to recover critical technology and applications from an event that results in total loss of the computing environment for an extended period.



3. Connecting storage islands for streamlined IT infrastructure:

- SAN centralization increases storage efficiency as resources are pooled together instead of being overused in some locations and underused in others.
- *SAN centralization also reduces OpEx*—More terabytes per second can be managed by an administrator when the storage system is consolidated because of the existence of a simplified, central point of control for monitoring, backup, replication, and provisioning. For example, [3] indicates that the total cost of ownership (TCO) of a SAN network over three years is \$0.38 per megabyte of user data, while the TCO for traditional non-networked storage (direct-attached storage [DAS]) is \$0.84 per megabyte.

Now that we've outlined the motivation behind storage networking, the rest of the paper will follow a structured approach, starting from the storage application level and moving to the storage protocol and its requirements. This paper then examines how these protocols can be carried over the transport layer, before pointing out how the Cisco SL-Series fits into the solution.

#### Applications and Protocols

The motivations listed above give rise to the extension of SANs between different locations, either connecting servers to storage systems, servers to other servers, or storage systems to other storage systems. The main storage applications that facilitate a transparent operation of a SAN over multiple locations can be divided into two types:

1. *Replication (or mirroring) of storage between a primary site and a backup site*—This is achieved by EMC's Symmetrix Remote Data Facility (SRDF) running on EMC Symetrix storage arrays and a midrange solution called Mirrorview, which runs on EMC's Clarion storage arrays, IBM Peer-to-Peer Remote Copy (PPRC) and Extended Remote Copy (XRC), Hitachi Data Systems (HDS) TrueCopy, and Compaq's Data Replication Manager (DRM).
2. *Zero-downtime server architectures*—These are represented mainly by IBM's Geographically Dispersed Parallel Sysplex (GDPS). They consist of several separate functions and protocols beyond the basic storage replication function, such as server-to-server communications, multisystem coordination and data sharing among clusters, and timer links that provide the clock synchronization between various components. See [4] for details.

The different SAN protocols that are used for the implementation of the above connectivity are:

1. *Fibre Channel*—A set of open standards developed by ANSI for a serial I/O protocol designed for reliably connecting various storage system components through a host of higher-level protocols used for server-to-storage connectivity, data replication, and other applications.
2. *Enterprise System Connection (ESCON)*—A 200-Mbps unidirectional serial protocol used to dynamically connect mainframes with their various control units (and also serves as an element of GDPS).
3. *Fiber Connection (FICON)* is the next-generation bidirectional channel protocol used to connect mainframes directly with control units or ESCON aggregation switches. FICON is compatible with Fibre Channel from a transport perspective (thus the discussion applies to FICON as well).
4. Other types of protocols that are not nearly as prevalent as the ones above (and thus will not be the focus of this paper). One such protocol is internet Small Computer System Interface (iSCSI), which carries SCSI blocks directly over TCP/IP, allowing connectivity of a computer to a storage device over an IP network. Other important protocols are part of the above-mentioned GDPS architecture, and include InterSystem Coupling (ISC), External Time Reference (ETR), and Control Local Oscillator (CLO).



This paper focuses on the transport of Fibre Channel/FICON, as this is currently the fastest-growing protocol for SAN. (ESCON transport will be addressed, in the future, via a separate white paper.)

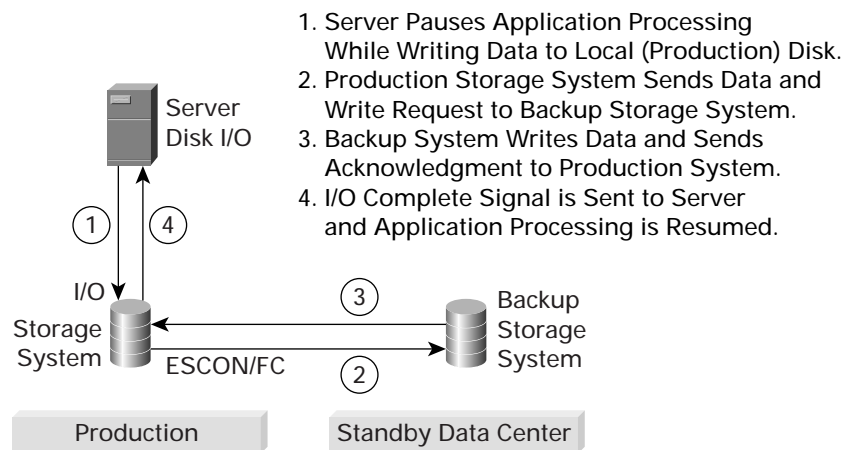
One of the main factors affecting the transport of Fibre Channel over a network is latency. The allowable latency depends on the type of storage application, which typically falls into one of two buckets:

1. Synchronous replication
2. Asynchronous replication

In synchronous mode, data between primary and secondary is copied, validated, and committed all at the same time. Synchronous applications require having an acknowledgment back for every sequence of a “write” operation. The reliability is very high, but the cost to that reliability is that applications might be in a waiting state while the secondary disk is “syncing up.”

The write operation can take up to several microseconds, depending upon the distance and remote disk processing speed. The application performance and throughput degrades as distance increases because round-trip latency increases. Figure 1 illustrates the stages of operation of synchronous data replication.

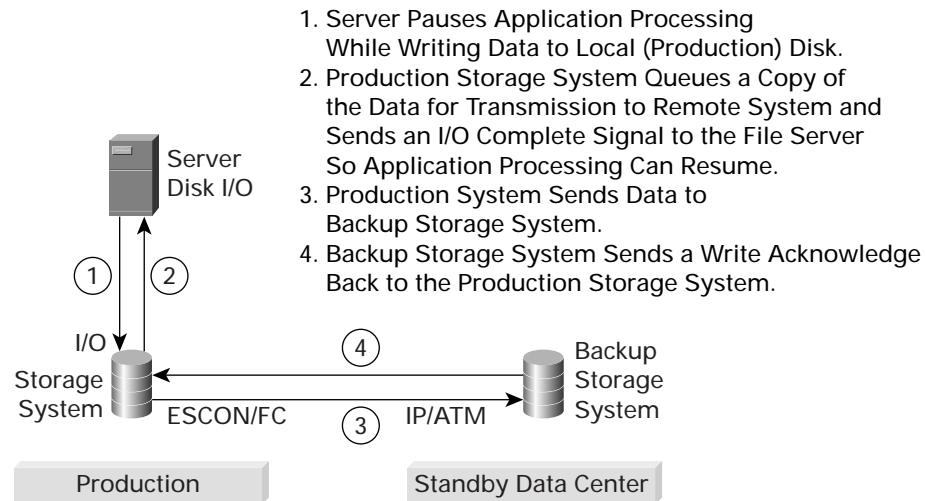
Figure 1  
Synchronous Data Replication



Asynchronous applications are not usually sensitive to round-trip delays. In asynchronous mode, data between primary and secondary is copied, but the commits are done separately between primary and secondary systems. Over extended distances, application performance has less impact than in synchronous mode. Figure 2 illustrates the operation of asynchronous data replication.



Figure 2  
Asynchronous Data Replication



While most data replication schemes fit into these two categories, there are other applications that do not fit. For example, GDPS cannot be classified as synchronous or asynchronous, and EMC has a hybrid mode of SRDF that normally runs asynchronously and switches to synchronous mode to slow the application down if the backlog is too significant.

### Options for Storage Transport over the Metro Network

Several mechanisms exist for transport, and their applicability depends on the following factors:

1. *Application sensitivity to delay*—In particular, synchronous vs. asynchronous replication (see above).
2. *Distance*—The propagation delay affects the feasibility of applications that are delay-sensitive.
3. *Service availability at a customer site*—Does the service provider offer leased dark fiber, a SONET/SDH private line, a “direct” Fibre Channel over SONET/SDH service, a Gigabit Ethernet “private line” service, or a switched IP service.
4. *Required bandwidth*—The higher the bandwidth demand, the more attractive it becomes to transport Fibre Channel over lower layers in the transport hierarchy.

The different transport alternatives are:

1. *Fibre Channel over leased fiber*—Fibre Channel is carried over a separate network, which is not part of the larger carrier network.
2. *Fibre Channel over wavelength*—Fibre Channel is carried over a separate wavelength as part of a metro dense or coarse wavelength-division multiplexing (DWDM/CWDM) network, along with other services on different wavelengths. It could still be electrically multiplexed using muxponder devices.
3. *Fibre Channel over SONET/SDH (FCoS)*—Fibre Channel is carried over a DS-3, an STS-n, or even a direct Fibre Channel service, potentially using virtual concatenation (VCAT) for greater transport efficiency. The service is then switched and multiplexed through a SONET/SDH network.



4. *Fibre Channel over Gigabit Ethernet*—This is a variant of Fibre Channel over IP (FCIP), where the IP network is a point-to-point Gigabit Ethernet service between a pair of customer sites.
5. *Fibre Channel over switched IP*—This is another variant of FCIP, in which the service is carried over a switched IP network and thus does not use a dedicated transport service offering.

These different means of carrying Fibre Channel are shown in Table 1.

Table 1 Fibre Channel Transport Alternatives

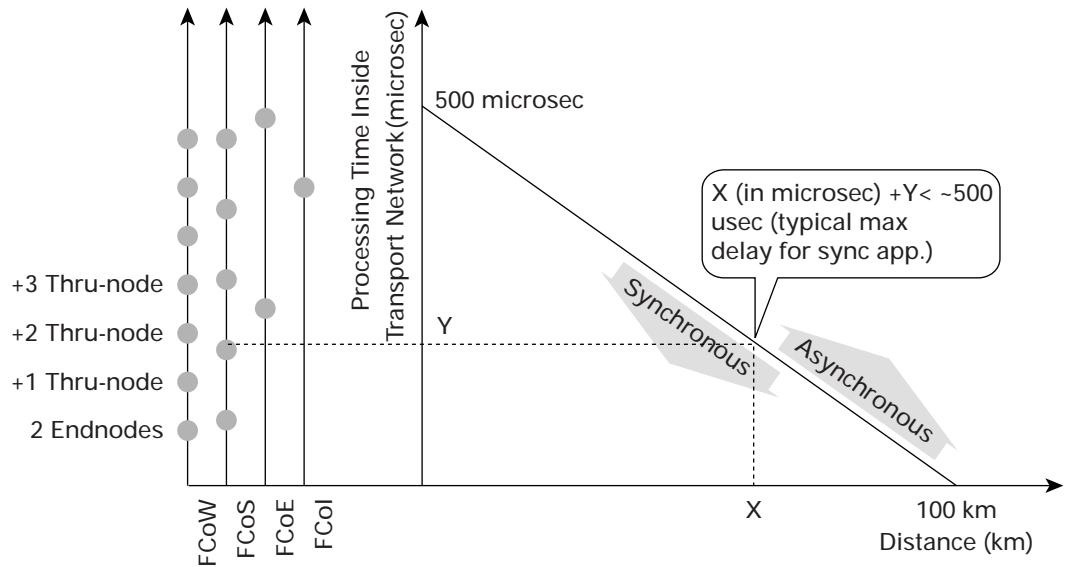
Fibre Channel over...	Lease What?	Protocol	Network	Latency/ Application	Distance/ Bandwidth
<b>Fiber</b>	Dark fiber	Native Fibre Channel, or Fibre Channel over enterprise-owned WDM	Point-to-point optical loop	Propagation delay only (synchronous/asynchronous)	80 km Full Fibre Channel speed
<b>Wavelength</b>	Wavelength	Fibre Channel over GFP in SONET/SDH payload over DWDM	Point-to-Point/ Ring Muxponding of multiple Fibre Channels over a wavelength is also possible	Low extra latency – (synchronous/asynchronous)	10's–100's km Full Fibre Channel speed
<b>SONET/SDH</b>	DS-3, OC-n, or VCAT private line or direct Fibre Channel service	GFP-mapped Fibre Channel in SONET/SDH payload	A metro/regional transport network, including rings and spurs	Medium extra latency (synchronous/asynchronous)	Unlimited distance in some cases Fibre Channel speed or sub rate
<b>Ethernet</b>	GbE service either full-rate or sub rate	FCIP or iSCSI over GbE; GbE is GFP mapped in SONET/SDH payload	A metro/regional transport network including rings and spurs	Higher extra latency (synchronous/asynchronous)	Longer delay limits distance in some applications Fibre Channel speed or sub rate
<b>Switched IP</b>	Data connection with given quality of service (QoS) service-level agreement (SLA)	FCIP or iSCSI	Any network, including metro and core	High latency; may not be predictable (asynchronous only)	Depends on IP buffering, QoS Typically low sub rate

The choice of transport options depends on the available services that can be offered by the service provider to the customer, and the price points. The feasibility of a certain solution depends on the allowable delay for the application and the delay introduced by a given transport technology (Figure 3). Figure 3 considers the maximal allowable delay for a synchronous application (approximately 500 microsec.), and shows how it is split between processing delays at the nodes and propagation delay along the fiber. The processing delay depends on the number of transport nodes the signal must travel through (where each node is represented as a blue dot on the Y axis), and also on the adaptation of Fibre Channel over the transport layer (as represented by the red dots). Figure 3 assumes different delays for



different technologies—for example, the adaptation for FCoS typically takes 25 microsec while the processing at a through node takes 10 microsec. Thus, the more processing that is needed, the less of the delay budget is left for fiber propagation delay.

Figure 3  
Applicability of Fibre Channel Transport Alternatives Due to Latency (One-Way Is Considered)



Two techniques have been introduced to overcome the round-trip delay limitation:

1. *Large buffer-to-buffer credits*—Fibre Channel has a built-in flow control mechanism based on buffer-to-buffer credits. This mechanism is sensitive to round-trip delay, since such a delay implies that the acknowledgments from the receiver of Fibre Channel frames are delayed, and consequently delay the ability to transmit frames at the source and reduce the effective throughput. This phenomenon is called “drooping.” A system that supports a large number of buffer-to-buffer credits can send more frames before it has to wait for acknowledgments. As a result, drooping only occurs for very long connections.
2. *Spoofing of acknowledgments*—If the transport gear creates Fibre Channel acknowledgments instead of waiting for the SAN gear at the remote end to generate them, the Fibre Channel protocol is streamlined and becomes much less sensitive to distance. This requires buffering capability at the transport gear to store the Fibre Channel frames that are “in flight,” in case the remote SAN gear cannot accept them due to congestion.

Neither technique totally eliminates the impact of propagation delay, as the higher-level protocol (HP’s DRM protocol, for example) still needs to respond before and after the Fibre Channel data transfer.

### An End-to-End Storage Networking Solution

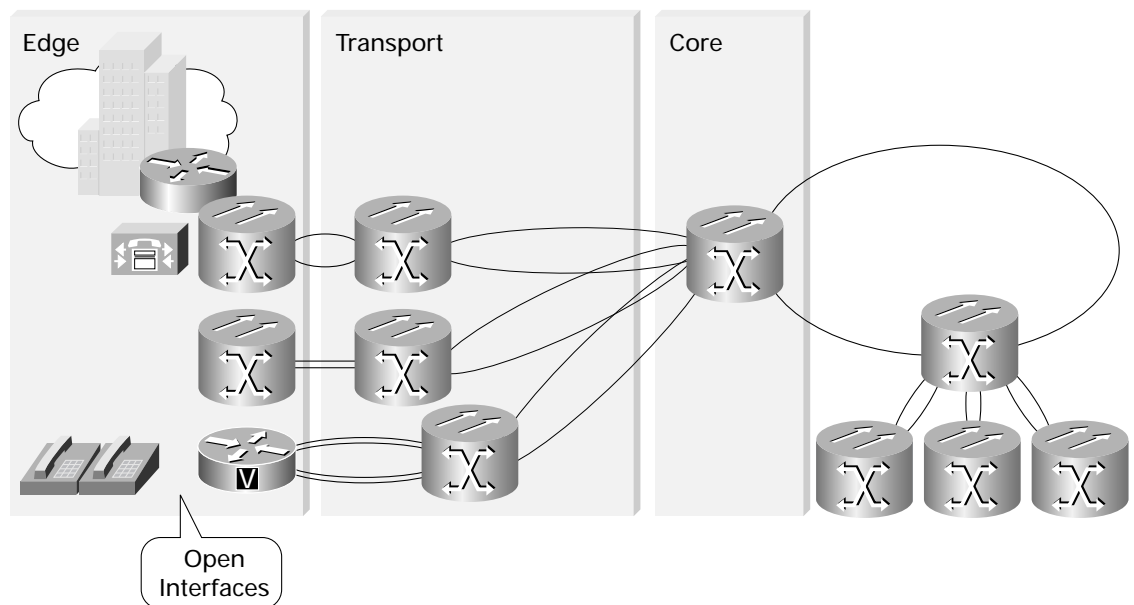
We now turn to how Cisco Systems addresses storage networking in general and how the Cisco SL-Series fits into the overall solution space.



As shown in Figure 4, a typical solution is based on four elements:

1. *Edge*: Fibre Channel switches at the customer premise
2. *Transport*: Variety of transport vehicles over metro networks
3. *Core*: Variety of switching solutions in the central office or CO (MSSP, Cisco Catalyst® Family)
4. *Open interfaces*: Certification with storage vendors (such as IBM and EMC) and telecommunications bodies (NEBS and OSMINE, for example)

Figure 4  
The End-to-End Storage Networking Solution



Cisco provides an overall storage transport solution, while most other vendors typically provide one piece of the complex puzzle—either a pure transport solution, a Fibre Channel switch, or a Fibre Channel extension device.

Recognizing that there is no single best solution for this problem, Cisco addresses the application space with various equipment options, as well as a comprehensive network management solution across the different platforms, as shown in Table 2.



Table 2 Coverage of the Fibre Channel Transport Space

Fibre Channel over...	Edge	Transport	Core	Open Interfaces
Fiber	Cisco MDS 9000, Brocade, McData	Fibre Channel directly over fiber or using enterprise-owned Cisco ONS 15530/540 technology	N/A	IBM, EMC, HDS, Compaq  NEBS, OSMINE
Wavelength		Cisco ONS 15454 plus SL-Series, with integrated ITU interfaces feeding into the multiservice transport platform (MSTP) optical add/drop multiplexer (OADM)	Cisco ONS 15454 MSTP	
SONET/SDH		Cisco ONS 15454 w/ DS-3 or OC-n interfaces or with SL-Series card (depends on type of offered service)	Cisco ONS 15454 and 15600	
Ethernet		Cisco ONS 15454+G-Series or ML-Series	Cisco ONS 15454 and 15600	
Switched IP		Cisco ONS 15454+G-Series or ML-Series or IP over DS-3	Cisco ONS 15454 and Cisco core switches and routers	

The main value of the Cisco approach is that it is possible to guarantee interoperability across the network in such a demanding environment—especially in the case of synchronous replication. For example, many Gigabit Ethernet alternatives do not support low-latency adaptation for the large frames which are generated by some SAN applications (called “Jumbo frames”), while the Cisco G-Series and ML-Series cards were designed with such applications in mind.

The Cisco solution has some additional unique features:

- A single-box Fibre Channel transport solution, allowing service providers to offer native Fibre Channel services on their existing MSPP platform or offer an integrated managed service, while many of the competitors require a two-box solution (Fibre Channel extender + SONET/SDH ADM).
- Carrier-class robustness through the Cisco ONS 15454 MSPP and MSTP transport family (as well as NEBS, OSMINE certification).
- The Cisco MDS 9000 Series of Fibre Channel switches represents the next step in the evolution of Fibre Channel switches. While many of its advantages are beyond the scope of this paper [6], a few benefits are:
  - *Virtual SAN (VSAN)*—The ability to segregate SAN islands of different parts of the organization (analogous to the concept of VLAN for the data network).
  - *Large buffer-to-buffer credits*—The ability to provide high-effective throughput for longer distances.



## Introducing the SL-Series

The Cisco SL-Series card (Cisco ONS 15454 Fibre Channel-MR-4) is a single-slot card with four client ports, each supporting 1.0625 or 2.125 Gbps Fibre Channel/FICON. It uses pluggable GBIC optical modules for the client interfaces, enabling greater user flexibility and better pay-as-you-grow cost. The payload from a client interface is mapped directly to SONET/SDH payload via GFP-T encapsulation. This payload is then cross-connected to the system's optical trunk interfaces (up to OC-192) for transport, along with other services, to other network elements.

The new card fills the Fibre Channel over Wavelength and Fibre Channel over SONET gaps in the transport category of the solution space (see highlighted cells in Table 2). Not only does this allow Cisco to provide 100 percent coverage of the Fibre Channel transport space, but it also provides end-to-end coverage of data center and enterprise storage networking solutions over metropolitan, regional, and wide-area networks.

The card plugs into the existing Cisco ONS 15454 chassis that many carriers have already deployed and is managed through the same management umbrella. Therefore, its introduction does not pose a major CapEx and OpEx investment, but, rather, an evolutionary extension of services that the carrier can offer. This is in keeping with Cisco's Optical Networking Group (ONG) vision of "evolution, not revolution."

The SL-Series protects the investment of the customer in other ways as well:

- It does not require upgrade of costly components of the MSPP, such as the switch matrix of the network element.
- The integration into the MSPP ensures that the card will be manageable and upgradable through the same infrastructure as the rest of the platform, streamlining OpEx. For example, each software load supports transport and data capabilities, eliminating unnecessary guesswork for ordering, installation, and upgrades.
- It is designed with future enhancements in mind, such as subrate capability and VCAT. These enhancements not only optimize the use of the SONET/SDH bandwidth, but they also allow the service provider to offer Fibre Channel services at increments of 50 Mbps.
- It will support data compression in the future, allowing further optimization of bandwidth in the transport layer, thereby enabling more services to be supported over the same infrastructure.
- Furthermore, with the future introduction of distance extension functioning via R\_RDY spoofing, the Cisco SL-Series will serve as an integrated Fibre Channel extension device, obviating the need for external SAN extension devices.

Following the tradition of the Cisco ONS 14545 MSPP, supporting TDM, Ethernet, and now storage, the Cisco SL-Series leads the industry in bit-rate and density for FCoS:

1. It supports 1- and 2-G Fibre Channel with low-latency GFP-T mapping, allowing customers to grow beyond 1-G Fibre Channel.
2. It supports the industry's highest Fibre Channel density over protected SONET/SDH transport network in a single network element—16 line-rate Fibre Channel on a single shelf over fully protected transport network such as 4F-BLSR OC-192, and Dual 2F-BLSR/UPSR OC-192.

Note: Our closest competitor(s) can also have 16 line-rate Fibre Channel port on the same single shelf, but for protected transport it can only transport 8 line-rate Fibre Channel from a single shelf, due to the more limited support for OC-192 line interfaces, namely 2F-BLSR OC-192 and one UPSR OC-192.



## Summary

Storage networking is growing due to regulatory as well as cost-related reasons [5]. The ability to efficiently transport Fibre Channel over metro and regional networks is a primary enabler for storage networking. The introduction of the Cisco SL-Series allows carriers to offer an integrated SAN extension solution over SONET/SDH (FCoS) using the ubiquitous Cisco ONS 15454 MSPP platform, thereby streamlining capital and operational expenses compared to a nonintegrated solution. Cisco is uniquely positioned in offering an end-to-end Fibre Channel networking solution, combining industry-leading Fibre Channel switches and Fibre Channel transport.

To summarize, the main features of the SL-Series are:

1. *Flexible tributary and line rate*—1- and 2-Gbps interfaces on the same card.
2. Low-latency Fibre Channel adaptation over SONET/SDH by means of transparent GFP mapping.
3. *Compact design*—Up to four Fibre Channel pluggable (GBIC) interfaces in a single-width card slot (two of which are active today) and up to eight SL-Series cards per shelf assembly.
4. *Flexible optical protection options*—UPSR/SNCP, 2F- and 4F-BLSR/MS-SPRing, PPMN, and unprotected (0+1).
5. *Network architecture flexibility*—Ring, multiple interconnected rings, linear ADM with an optional DWDM layer in the same network element.

This platform will support a smooth evolution path for future support of increased density (four Fibre Channel interfaces per card), distance extension, subrating in increments of 50 Mbps using VCAT, and data compression. Based on what is currently known about competitive offerings, Cisco expects that this platform should ultimately prove to be best-in-class for both regulated and nonregulated customers and applications.

## References

- [1] “*Compliance: The effect on information management and the storage industry*”; ESG Impact Report; May 2003
- [2] “*Business Continuity When Disaster Strikes*”; Fibre Channel Industry Association; Search <http://www.fibrechannel.org> for “*When Disaster Strikes*”
- [3] “*The Storage Report—Customer Perspectives & Industry Evolution*”; McKinsey & Co. and Merrill Lynch; June 2001
- [4] *IBM @ server zSeries Connectivity Handbook*; Search <http://www.ibm.com/> for “*zSeries Connectivity Handbook*”
- [5] “*Cisco Storage Networking Solution—The Strategic and Financial Justification of SANs White Paper*”; Search <http://www.cisco.com> for “*justification of SAN*”
- [6] “*Cisco MDS 9000 Family Overview*”; Search <http://www.cisco.com> for “*MDS 9000 Family Overview*”

## CISCO SYSTEMS



Corporate Headquarters  
Cisco Systems, Inc.  
170 West Tasman Drive  
San Jose, CA 95134-1706  
USA  
www.cisco.com  
Tel: 408 526-4000  
800 553-NETS (6387)  
Fax: 408 526-4100

European Headquarters  
Cisco Systems International BV  
Haarlerbergpark  
Haarlerbergweg 13-19  
1101 CH Amsterdam  
The Netherlands  
www-europe.cisco.com  
Tel: 31 0 20 357 1000  
Fax: 31 0 20 357 1100

Americas Headquarters  
Cisco Systems, Inc.  
170 West Tasman Drive  
San Jose, CA 95134-1706  
USA  
www.cisco.com  
Tel: 408 526-7660  
Fax: 408 527-0883

Asia Pacific Headquarters  
Cisco Systems, Inc.  
Capital Tower  
168 Robinson Road  
#22-01 to #29-01  
Singapore 068912  
www.cisco.com  
Tel: +65 6317 7777  
Fax: +65 6317 7799

Cisco Systems has more than 200 offices in the following countries and regions. Addresses, phone numbers, and fax numbers are listed on the  
**Cisco Web site at [www.cisco.com/go/offices](http://www.cisco.com/go/offices)**

Argentina • Australia • Austria • Belgium • Brazil • Bulgaria • Canada • Chile • China PRC • Colombia • Costa Rica • Croatia  
Czech Republic • Denmark • Dubai, UAE • Finland • France • Germany • Greece • Hong Kong SAR • Hungary • India • Indonesia • Ireland  
Israel • Italy • Japan • Korea • Luxembourg • Malaysia • Mexico • The Netherlands • New Zealand • Norway • Peru • Philippines • Poland  
Portugal • Puerto Rico • Romania • Russia • Saudi Arabia • Scotland • Singapore • Slovakia • Slovenia • South Africa • Spain • Sweden  
Switzerland • Taiwan • Thailand • Turkey • Ukraine • United Kingdom • United States • Venezuela • Vietnam • Zimbabwe

All contents are Copyright © 1992–2003 Cisco Systems, Inc. All rights reserved. Catalyst, Cisco, Cisco Systems, and the Cisco Systems logo are registered trademarks of Cisco Systems, Inc. and/or its affiliates in the U.S. and certain other countries.

All other trademarks mentioned in this document or Web site are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company.  
(0304R) CCG/DH/LW5009 0803