



## Solution Overview

# Designing High-Performance EMC File Systems Using Cisco MDS iSCSI

**Storage administrators in network attached storage (NAS) environments often must determine how to scale storage and increase utilization without increasing the number of NAS file servers and making management more difficult. When these environments grow from hundreds to thousands of servers accessing terabytes of data, the NAS appliance often becomes the bottleneck and impedes scalability. To scale to thousands of servers, administrators often deploy multiple NAS appliances, each hosting a copy of the data. This approach complicates management and reduces scalability of these appliances, and wastes raw storage space for multiple copies of the same data.**

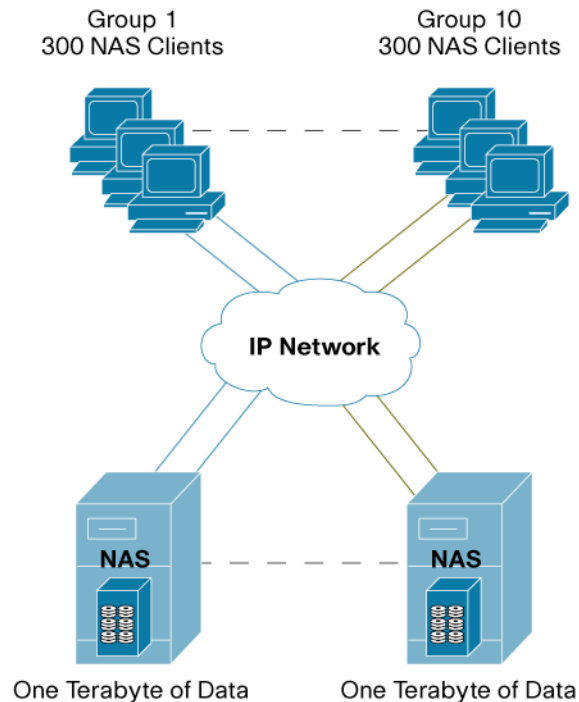
An Internet Small Computer System Interface (iSCSI)-based solution involving EMC Celerra Multi-Path File System—iSCSI (MPFSi) with the Cisco® MDS 9000 Family platform provides ease of management, scalability to thousands of servers, and increased storage utilization. Celerra MPFSi provides the capability to split data locking and block data movement into two paths, where all meta-data locking is processed by the Celerra NAS appliance and all data I/O flows through a storage area network (SAN). The highly available integrated iSCSI features on the Cisco MDS 9000 platform, such as Virtual Router Redundancy Protocol (VRRP) and iSCSI Proxy Initiator, allow the Celerra MPFSi solution to scale for block-level access through iSCSI. This is a lower-cost alternative to the NAS-based deployment mentioned earlier.

## EMC CELERRA MPFSI OVERVIEW

EMC Celerra MPFSi answers the question: Should I choose NAS for file sharing or SAN for the best application performance? With MPFSi and EMC's industry-leading family of file servers, applications can now benefit from both NAS file sharing and the performance and scalability of block access through SAN. By using existing IP network infrastructure and providing easy management, MPFSi with Cisco MDS 9000 offers a cost-effective solution that reduces the complexity of managing multiple storage assets.

Current solutions that have applications accessing NAS appliances provide processing for clients' access to the storage that is only available on that NAS system. As the number of servers increases, the NAS system becomes the bottleneck and limits scaling of the design. Another NAS system can be built to support more clients but the data on all the NAS systems must be duplicated. This results in greater management requirements because of increasing numbers of NAS systems and multiple copies of the same data, which is an inefficient use of storage. Figure 1 depicts a typical NAS environment that needs to support 3000 servers. The number of clients accessing a single NAS appliance may vary. In the following scenario, 300 clients is an average limit on these NAS appliances. If the number of supported clients accessing the NAS appliance is lower, the management and data storage cost would increase.

**Figure 1.** NAS Appliance Solution for 3000 Servers



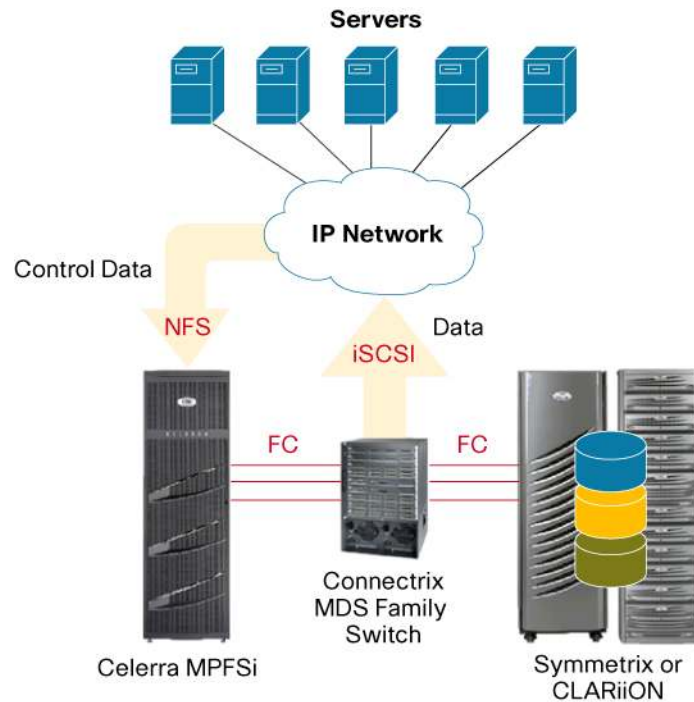
**Note:** To support 3000 servers, this solution needs a total of 10 terabytes of raw storage (excluding any mirrored local disks or any other features that require more raw storage), where only 1 terabyte is actual user or application data. Instead of managing one NAS appliance, this solution requires management of 10 NAS appliances, 10 copies of data, and making sure that the data across the NAS appliances is consistent.

### EMC Celerra MPFSi Architecture

Celerra MPFSi clients communicate over the IP network to share metadata and control information using Common Internet File System (CIFS) and network file server (NFS) protocols (Figure 2). When the MPFSi agent receives a data request from an application, it uses Celerra-provided metadata to access the storage array directly using the low latency and high performance of iSCSI. The following are architectural benefits with MPFSi:

- Celerra systems can support 10 times more clients for demanding applications because they do not have to process data transfers.
- Because Celerra systems are not in the data path, bandwidth to applications can be scaled by simply adding Symmetrix or Clariion arrays to the MPFSi environment.
- Single files can be read and written by clients simultaneously because of a sophisticated MPFSi range lock manager.
- A single client can use the Celerra MPFSi system to access multiple file systems, and back-end arrays.

**Figure 2.** EMC Celerra MPFSi Architecture



### EMC Celerra MPFSi Components

The following components make up the Celerra MPFSi topology with the Cisco MDS 9000 Family, including hardware and client/server software components:

- EMC Celerra or NSX Gateway running DART NAS code
- Multiple EMC Symmetrix or Clariion back-end storage arrays
- Cisco MDS 9200 or MDS 9500 Series Switches with:
  - 4-port IP storage services (IPS) module, 8-port IPS module, or multiprotocol services (MPS) line card
- Clients, application servers, or grid servers sharing files over the LAN to the Celerra system
- MPFSi agent and iSCSI initiator loaded on each client/grid node that wants accelerated performance (MPFSi does not have to be loaded on clients that want standard NFS file access)

### ISCSI ON THE CISCO MDS 9000 PLATFORM

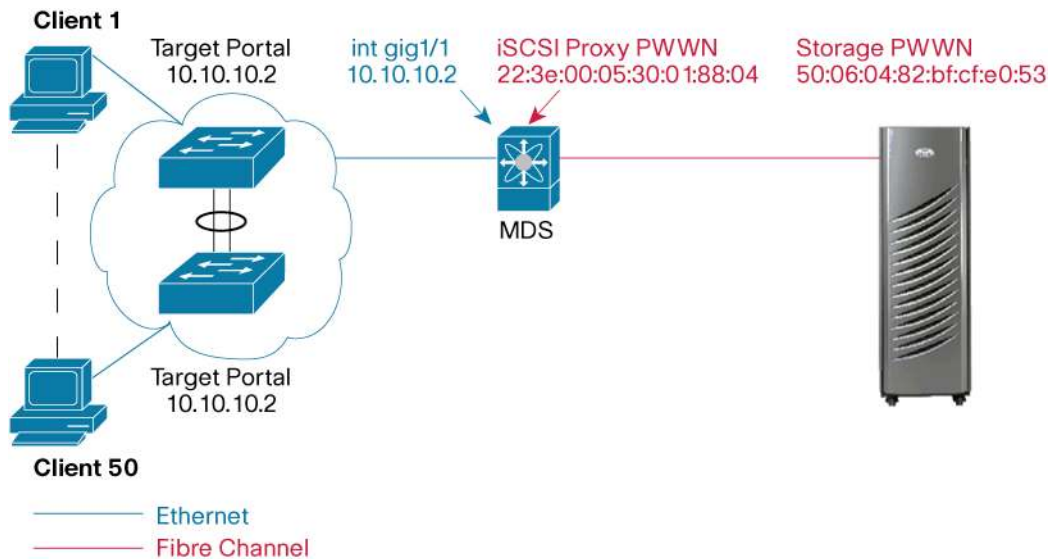
The Cisco MDS 9000 Family of multilayer directors and fabric switches provides enterprises with the ability to build large-scale Fibre Channel SANs and extend these SANs to low and mid-range servers. Cisco MDS 9000 Family directors and switches offer advanced features such as virtual SAN (VSAN) and a fabric management tool (the Fabric Manager) supporting Role-Based Access Control (RBAC), making these switches an ideal choice in deploying large-scale solutions. By using the iSCSI protocol, enterprises can take advantage of Ethernet and IP technologies to further extend their storage environments and realize the cost savings derived from storage consolidation. With the Cisco IPS switching module providing Gigabit Ethernet ports for iSCSI services, enterprises can extend the SAN to other low to mid-level servers with the iSCSI protocol. Customers may use their existing IP infrastructure along with their in-house IP expertise to consolidate and optimize enterprise storage.

Two key features of the Cisco MDS iSCSI solution with EMC Celerra MPFSi are iSCSI Proxy Initiator and Virtual Router Redundancy Protocol (VRRP). These features provide simplified management for large-scale iSCSI implementations and redundancy in case of Gigabit Ethernet port failures.

### Simplified Provisioning and Management

Solutions such as Celerra MPFSi have many clients accessing the same storage, which may cause access control lists (ACLs) on the storage array to reach their limit of supported number of hosts. With the iSCSI Proxy Initiator feature on the Cisco MDS 9000 platform, many iSCSI initiators can now impersonate as a single Fibre Channel initiator to the storage array. This dramatically reduces the configuration needed on the storage array itself and simplifies management for storage access. In addition, the Cisco MDS 9000 platform can perform iSCSI client authentication and can define mapping and masking of storage logical unit numbers (LUNs) to iSCSI clients, creating an additional level of security for storage access. See Figure 3.

**Figure 3.** iSCSI Simplified Management with Cisco MDS iSCSI Proxy Initiator Feature

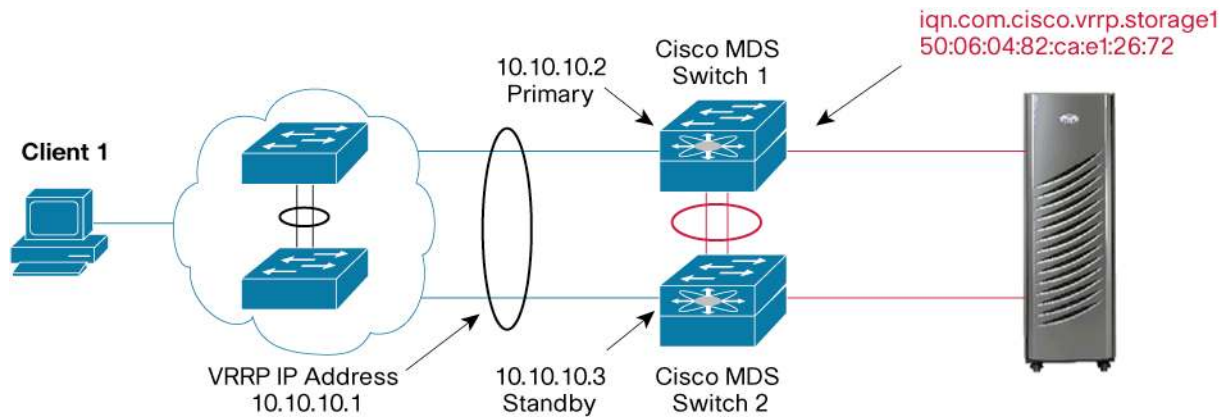


**Note:** Instead of creating 50 zones for each server access to the storage Port World Wide Name (PWWN); the iSCSI Proxy Initiator feature will simplify the management. In Figure 3, a single zone of the iSCSI Proxy PWWN and the storage PWWN is all that is needed for the servers to access to the storage. Any LUN masking done on the storage array will be simplified as well.

### Virtual Router Redundancy Protocol

Cisco MDS 9000 Family switches support RFC 2338, which describes Virtual Router Redundancy Protocol (VRRP). VRRP is a feature performed by the IPS or MPS module, and guarantees high availability for iSCSI by providing a redundant alternative path. This is done by exposing storage through a separate physical Gigabit Ethernet interface on the same switch or on a different physical switch. In case of a Gigabit Ethernet port failure, iSCSI sessions will resume across the standby Gigabit Ethernet port. This VRRP feature will only work within the same physical fabric and does not work across physical fabrics. Figure 4 depicts the configuration of the iSCSI VRRP feature.

**Figure 4.** Cisco MDS iSCSI VRRP Feature



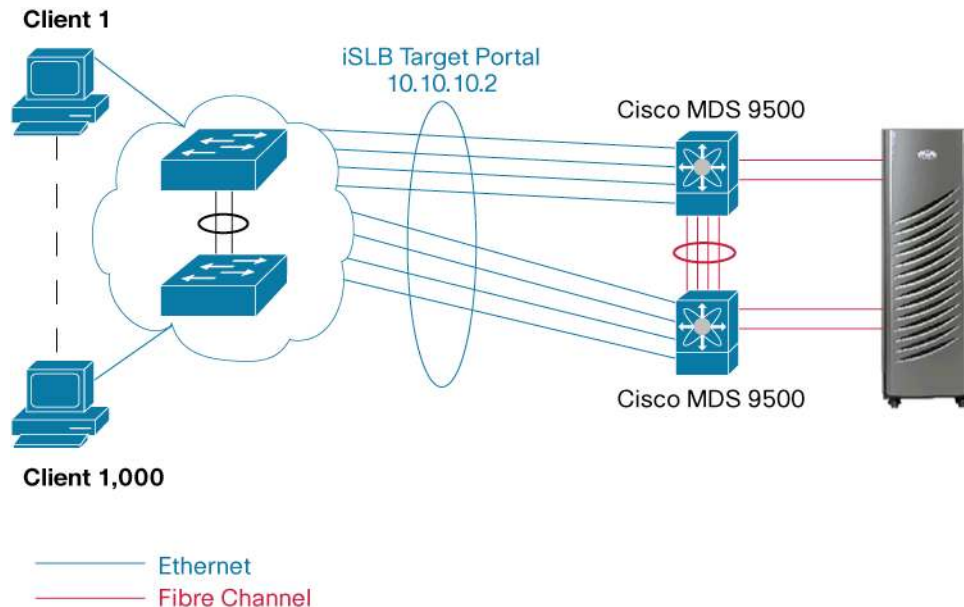
**Note:** In case of a failed Gigabit Ethernet link on Cisco MDS Switch 1, host I/O will be briefly interrupted and will resume once the Gigabit Ethernet link on Cisco MDS Switch 2 assumes the VRRP primary role and the host re-establishes its iSCSI session. This minimizes application downtime.

#### Advanced Simplified iSCSI Management Through iSCSI Server Load Balancing

With the iSCSI Proxy Initiator feature, consolidating servers to a single Gigabit Ethernet port simplifies management. The new iSCSI Server Load Balancing (iSLB) feature in Cisco MDS 9000 SAN-OS Software Release 3.0 provides consolidation of Gigabit Ethernet ports and further simplifies configuration. iSLB allows multiple Gigabit Ethernet interfaces, spanning across different line cards or different Cisco MDS 9000 switches within the same physical fabric, to act as a single iSCSI target portal. Hundreds or thousands of servers need to be configured with just one target portal IP address for iSCSI storage access. The iSLB feature on the Cisco MDS 9000 automatically assigns each server to a specific Gigabit Ethernet port. iSLB provides load balancing of iSCSI initiators within a physical fabric as well as redundancy capability using VRRP.

**Note:** iSLB provides load balancing based on the number of initiators per Gigabit Ethernet port and not based on the number of sessions or the amount of traffic. Each iSCSI initiator is assigned exactly one Gigabit Ethernet port.

**Figure 5.** iSCSI Simplified Management with Cisco MDS iSLB Feature



**Note:** Instead of grouping a smaller number of servers to a single Gigabit Ethernet port IP address, hundreds or thousands of servers can be configured to point to a single iSCSI target portal IP address. This simplifies management on the server side and also extends the same functionality as the iSCSI Proxy Initiator feature, which simplifies management for storage access to large numbers of clients. In the example of a single Cisco MDS 9509 chassis with two 16-port line cards for storage access, five 8-port IPS modules for iSLB access, and with a ratio of 25:1 server per Gigabit Ethernet interface, iSLB can consolidate and manage up to 1000 servers for the Cisco MDS 9509. So in Figure 5, iSLB can manage up to 2000 servers in a single fabric.

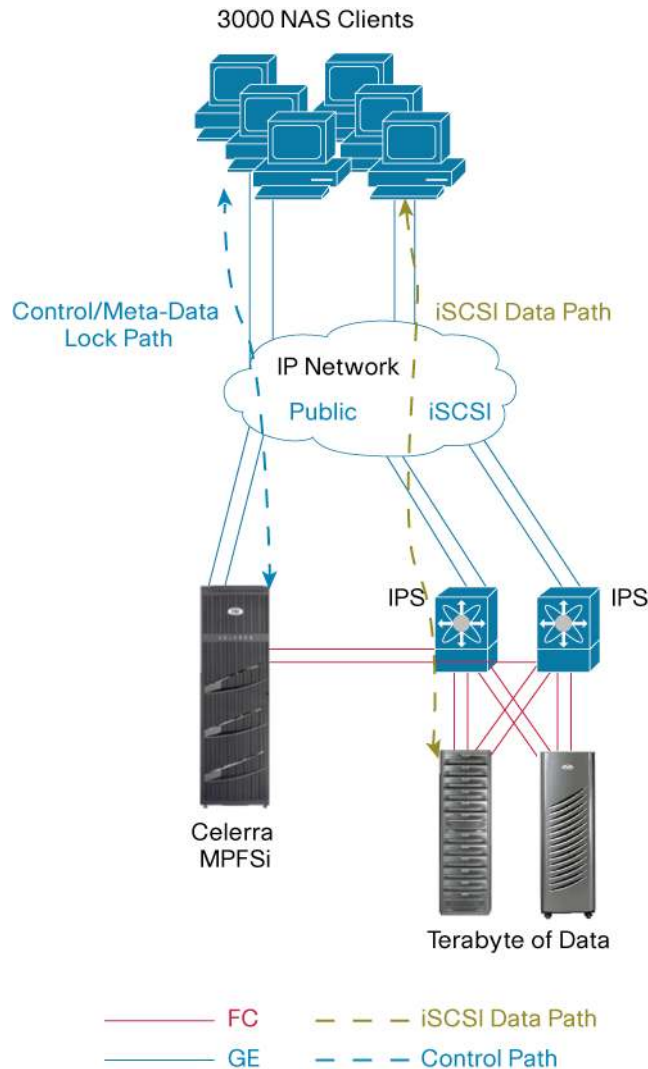
### CISCO MDS ISCSI AND EMC CELERRA MPFSI SOLUTION

Prior to the introduction of Celerra MPFSi, EMC Celerra MPFS had the capability to function with the split-path architecture but only allowed block-level access through Fibre Channel. This provided for a scalable and high-performance solution. The limiting factor to this solution was the cost of providing a Fibre Channel host bus adapter (HBA) for each server and a Fibre Channel port on the SAN switch. As the number of servers grew to hundreds and then thousands, this solution became cost-prohibitive.

With EMC Celerra MPFSi providing the multi-path architecture and allowing block-level I/O data flow to go through iSCSI, you have a cost-effective solution that scales to thousands of NAS clients. With the Cisco MDS 9000 providing advanced iSCSI features for redundancy with VRRP and simplified management with iSCSI Proxy Initiator, the Cisco MDS iSCSI and EMC MPFSi can provide enterprise environments with a scalable architecture. See Figure 6.

The recommended fan-out ratio per Gigabit Ethernet port ranges from 10 to 50 servers, depending on throughput requirements and the EMC support matrix. Instead of multiple copies of data residing on multiple NAS enclosures, this solution provides a single copy of the data in a scalable and high-performance SAN.

**Figure 6.** Cisco MDS iSCSI and EMC Celerra MPFSi Solution



**Note:** This solution provides both server scalability and efficient storage utilization, compared to other NAS appliance solutions. With Cisco MDS 9000 being the only platform to provide scalability and high availability with EMC Celerra MPFSi, iSCSI environments can now scale from hundreds to thousands of servers.

## SUMMARY

With the growth of NAS deployments, management, scalability, and efficient use of storage becomes more difficult. Together, Cisco MDS iSCSI and EMC Celerra MPFSi provide a cost-effective solution that allows NAS environments to scale, reduce the number of managed systems, and increase storage utilization. With EMC Celerra MPFSi providing the multi-path architecture and intelligent functionality to improve performance in the NAS environment, and Cisco MDS iSCSI providing the availability, scalability, and security, NAS environments now can scale from hundreds to thousands of clients. The strong partner relationship of Cisco Systems® and EMC assures customers of the quality of solutions that Cisco and EMC customers have come to expect.

**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.  
168 Robinson Road  
#28-01 Capital Tower  
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 Website 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 • Cyprus  
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

Copyright © 2006 Cisco Systems, Inc. All rights reserved. CCSP, CCVP, the Cisco Square Bridge logo, Follow Me Browsing, and StackWise are trademarks of Cisco Systems, Inc.; Changing the Way We Work, Live, Play, and Learn, and iQuick Study are service marks of Cisco Systems, Inc.; and Access Registrar, Aironet, BPX, Catalyst, CCDA, CCDP, CCIE, CCIP, CCNA, CCNP, Cisco, the Cisco Certified Internetwork Expert logo, Cisco IOS, Cisco Press, Cisco Systems, Cisco Systems Capital, the Cisco Systems logo, Cisco Unity, Enterprise/Solver, EtherChannel, EtherFast, EtherSwitch, Fast Step, FormShare, GigaDrive, GigaStack, HomeLink, Internet Quotient, IOS, IP/TV, iQ Expertise, the iQ logo, iQ Net Readiness Scorecard, LightStream, Linksys, MeetingPlace, MGX, the Networkers logo, Networking Academy, Network Registrar, Packet, PIX, Post-Routing, Pre-Routing, ProConnect, RateMUX, ScriptShare, SlideCast, SMARTnet, The Fastest Way to Increase Your Internet Quotient, and TransPath are registered trademarks of Cisco Systems, Inc. and/or its affiliates in the United States and certain other countries.

All other trademarks mentioned in this document or Website 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. (0601R)

