



WHITE PAPER

SAN INTERNETWORKING SOLUTIONS

Using Cisco MDS 9000 Series Multilayer Switches

Pressure to reduce and control IT data center storage infrastructure costs continues to drive the migration toward storage area networks (SANs). As more applications are migrated to the SAN, the resiliency of the SAN becomes more critical than ever. As more applications are migrated to the *common SAN*, these applications become susceptible to the faults and potential disruptions of the *common SAN*. In addition, the continual adoption of SAN extension solutions to remote data centers as part of business continuance plans extends not only SAN connectivity, but the potential *fault domain*, as now, any disruptions in a remote facility can affect the central data center.

As IP networks have continued to scale, Layer 2 VLAN technologies and Layer 3 IP routing protocols have helped to constrain fault domains within IP subnet borders while enabling required connectivity across these domains. However, Fibre Channel as a basic Layer 2 protocol does not possess capabilities of fault domain isolation. This capability is needed within the SAN to be able to continually drive cost-saving, consolidated SAN infrastructures and not jeopardize the stability or security of these environments. Cisco Systems® has introduced numerous innovative technologies in the MDS 9000 Series to help build the scalability and stability achieved in traditional data networks into the SAN. New data center technologies such as virtual SANs (VSANs) and Inter-VSAN Routing (IVR), along with integrated SAN extension transport options such as Fibre Channel Interface Protocol (FCIP), bring extensibility and resiliency to Fibre Channel SANs as they have existed in other traditional data networks.

This paper investigates the potential problems associated with Fibre Channel, the single physical fabric/single fault domain model, and the inability to adequately isolate against local and remote network events and failures. The solution to SAN fabric scalability and ability to conquer the effects of a single fault domain are provided in the Cisco solution for building Fibre Channel SANs, using the logical fabric model of VSANs with connectivity and resiliency-enhancement technologies, including IVR and FCIP. Other methods of achieving SAN internetworking are discussed, including Internet Fibre Control Protocol (iFCP), its mechanisms, and its inherent limitations.

SAN INTERNETWORKING

SANs have evolved from smaller networks based on arbitrated loop hubs providing connectivity for small isolated SAN islands to larger SAN networks connecting potentially hundreds of devices. Much like the scalability and resiliency considerations that brought about the evolution of Ethernet networks, relying on a single SAN fabric to interconnect large numbers of devices can jeopardize the resiliency and stability of the SAN. To compound the problem, SAN fabrics that were once contained within a single data center and dedicated to a single application or department may now span countries or even oceans. Larger SAN fabrics have the advantage of centralized manageability and more efficient use of fabric resources (ports and Inter-Switch Links [ISLs], for example) but are quickly approaching the architectural limitation of legacy SAN switches and the basic Fibre Channel protocol which ensures appropriate resiliency as their scale grows.

Like the Ethernet protocol, Fibre Channel is a switched Layer 2 protocol—there is no hierarchical network domain concept that would serve to isolate and localize control protocols and messaging within a given region of the network. Instead, like Ethernet, Fibre Channel maintains a set of fabric control protocols that are fabric-wide in scope, such as zoning and switch-generated state change notifications. Therefore, even Fibre Channel control protocol messaging of local relevance can only be sent fabric-wide and as a result, the effects can be felt fabric-wide. This architectural flaw results in high-availability vulnerability within a fabric, especially with geographically disparate SAN extension scenarios. The high-availability risk comes from potentially disruptive local control protocol events that can result in disruptions that span the full extent of the fabric on both sides of the extension link. An additional concern is the quantity and impact of these events as fabrics grow in size. Many customers are planning for the creation of large-scale fabrics consisting of more than

3000 ports in a single fabric. As these fabrics grow, the amount of fabric control information increases, and its tax on switch CPU and memory resources, which are required to process these fabric events, can compromise fabric stability. Cisco has addressed the need for added CPU processing resources in its MDS 9000 Series switches by offering more than 10 times the CPU processing power of any other director or fabric switch in the market. However, the risk of increased fabric-wide disruption due to increased fabric control protocol traffic still remains.

To solve this problem in the Ethernet domain, Layers 2 and 3 were devised to help bring hierarchy and necessary isolation to larger Ethernet fabrics. VLANs provided an isolating capability at Layer 2, and IP routing protocols such as Open Shortest Path First (OSPF) and Border Gateway Protocol (BGP) provided higher-level purposeful IP connectivity. Ethernet control protocols such as Spanning Tree were isolated to their respective VLANs and would not affect other VLANs in the event of a Spanning Tree event. Unfortunately, the common upper layer protocol in a SAN, the Fibre Channel Protocol (FCP), does not provide the same isolation capabilities as in IP. This is an increasing concern as SANs continue to grow larger and to extend further across various transports, line speeds, and latencies.

Evolution of SAN Islands

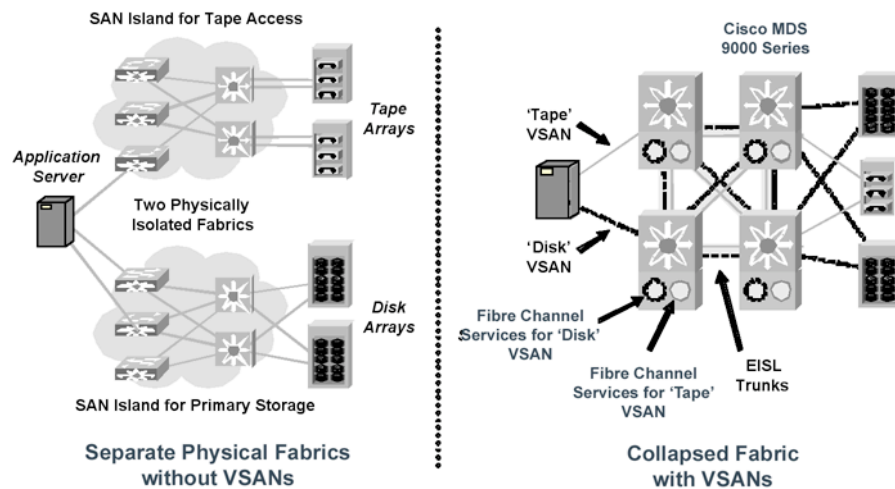
Past SAN design practices dictated the use of many smaller SAN “islands” of connectivity. Often, these SAN islands were deployed on an application-by-application or department-by-department basis to satisfy application security and stability concerns. Many of the stability concerns stemmed from a lack of capability within the Fibre Channel protocol to build domains in a larger fabric, as discussed above. Due to the previous nonexistence of large SAN switches with high port counts and the inability to support a large number of switches in a SAN (again for the above-mentioned Fibre Channel protocol inefficiencies), customers were forced to deploy SANs in smaller configurations. A SAN islands is a SAN that is created using a small number of switches with a lower overall port count for each specific enterprise application or department. The management of many smaller SAN islands has become burdensome. Legacy switches are being replaced with larger director-class switches, enabling higher port density core-edge configurations. However, due to many of the technical and nontechnical reasons for building SAN islands in the first place, the SAN island practice continues. The waste incurred by deploying multiple SAN islands, each with all ports dedicated to a given application and often many ports remaining unused further increases the effective cost of this design approach.

VSANs

To offer a solution to the inefficient practice of SAN islands, Cisco has developed VSAN technology. VSANs overcome the inefficiencies of SAN islands by using a concept proven in Ethernet through the use of VLANs. VSANs map isolated SAN islands onto a common physical infrastructure—without merging the control and management of the independent fabrics. Each port within a Cisco MDS 9000 Series switch is assigned to any created VSAN; up to 256 VSANs can be supported on a given switch within the physical fabric. All of the fabric isolation achieved through the use of SAN islands is also realized with VSAN technology. Within a VSAN-enabled fabric, common ISLs called EISLs (enhanced inter-switch links) are used for multiple VSANs. Traffic from all VSANs is tagged with VSAN membership information and transported over common ISLs while remaining isolated at a logical level. This use of common ISLs across multiple virtual fabrics and applications provides further infrastructure cost savings. VSANs not only provide a Layer 2 isolation capability within a larger physical fabric, but VSAN management can also be centrally controlled. In addition, individual VSAN management can be fully or partially assigned to application administrators by giving designated administrators permission to manage all services and policies associated with a virtual fabric assigned for their respective applications. Using Cisco role-based access control (RBAC) technology along with RBAC VSAN enhancements, individual roles can be defined for administrators that grant selective access to switch configuration functions, which can also be on a per-VSAN basis. For example, a role called *role_1* could be defined that allows an administrator to view the entire physical fabric (all VSANs), but only define and activate zoning and static FCID allocation for *VSAN_10*.

Ports in the physical fabric can be nondisruptively moved from one VSAN to another, yielding real provisioning cost savings over the alternative of having wasted ports confined to a SAN island. Just like with VLANs and Ethernet, EISLs carry Fibre Channel traffic associated with multiple VSANs between switches. EISLs allow the network designer to use the same ISLs for multiple virtual fabrics and to have ultimate control over traffic flow on a per-VSAN basis. In addition, ISL or EISL aggregation, called port channeling, within the Cisco MDS 9000 Series can be used to aggregate bundles of E/ISLs to create larger virtual E/ISLs for higher resiliency and bandwidth (Figure 1).

Figure 1
Example of VSAN Enabled SAN



All control and data traffic generated within a given VSAN is isolated to the VSAN and can only be extended as far as the particular VSAN is extended within the fabric. This design characteristic of VSANs is an advantage for larger physical fabric scalability. Using VSANs, any-to-any physical connectivity can be established within the fabric, but smaller, more manageable, virtual fabrics can be dynamically created as needed throughout the fabric. This VSAN-enabled design practice helps to ensure discreteness and necessary control plane isolation for added fabric resiliency and scalability. Fabric events that occur in a smaller defined virtual fabric only have impact to the extent of the virtual fabric, not the larger physical fabric.

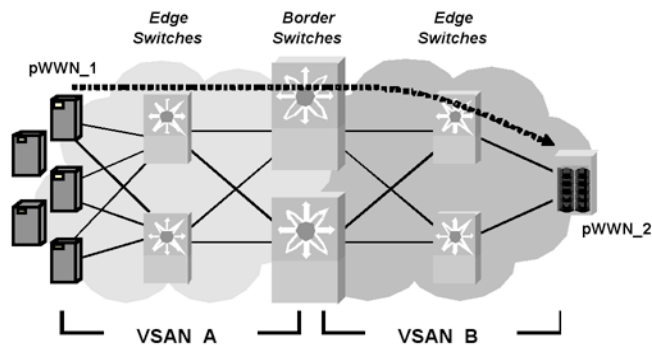
VSANs greatly advance the ability to build larger, more resilient and manageable SAN fabrics and enable the achievement of fabrics in the thousands-of-ports range. However, VSANs can have one limitation based on their design. Because all traffic is isolated within each respective VSAN, each fabric-attached device can only be contained within one VSAN. While this is sufficient for most host bus adaptor (HBA) connections, it may be desirable to share larger storage array ports between multiple applications or departments comprised of multiple VSANs. Also, in the case of extended fabric solutions over longer distances, it may be desirable to not extend a common fabric or VSAN across a larger latency transport service to optimize fabric resiliency. To extend the VSAN solution to meet these desired capabilities, an extension to the VSAN technology called IVR has been developed in Cisco MDS 9000 Series switches.

INTRODUCING IVR

As previously discussed, VSANs enable a unique degree of isolation through the construction of virtual fabrics on top of a common physical fabric. SAN architects may require certain devices to communicate within more than one VSAN, but not at the expense of merging the virtual fabrics for resiliency reasons. Devices such as tape drives or large subsystems are generally candidates to be presented

in more than one VSAN. Cisco IVR technology is used to create a path from a device in one VSAN to potentially multiple devices in different VSANs—without merging the individual VSAN fabrics and hence fault domains. Without the IVR capability, a device such as a large disk array must be physically connected to a unique switch port for each VSAN in which it wants to participate. However, with IVR, the simple procedure of defining an IVR zone containing the devices in various VSANs that require connectivity is all that is required to enable a shared device to communicate across multiple VSANs—again, without merging the participating virtual fabrics (VSANs).

Figure 2



IVR Scenario

The Inter-VSAN Zone

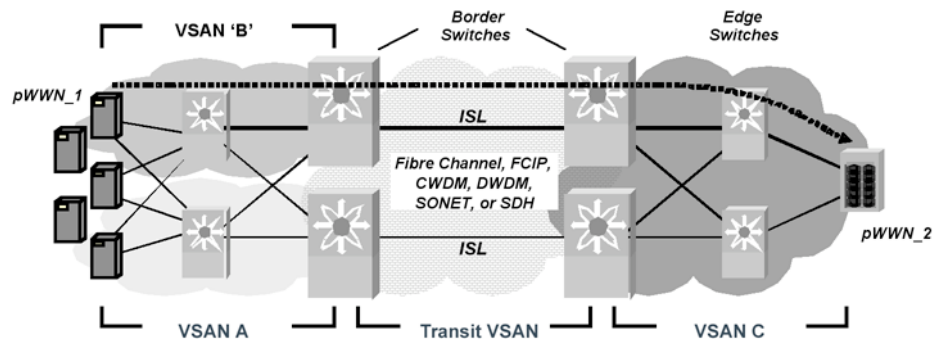
One of the main advantages to IVR technology is its ease of provisioning and manageability. One of the major differentiators between IVR and other approaches, such as those using iFCP, is that the provisioning of IVR is accomplished using a procedure that is well known to any SAN administrator. Using the same zone administration procedures and management tasks familiar to SAN administrators, an IVR instance is created by simply defining a new type of zone called an Inter-VSAN zone (IVZ). An IVZ is defined in the same way as any other zone, but it contains devices from participating edge VSANs requiring cross-VSAN connectivity. The IVR technology automatically determines which routes and name server entries require cross-propagation based on the IVZ definition. Using the familiar zoning mechanism, SAN administrators need not learn any new protocols such as iFCP and Internet Storage Name Service (iSNS) or new management tools to enable and manage the IVR service.

The IVR Border Switch

To accomplish IVR connectivity, at least one switch in the fabric must act as a transport, or IVR border switch, providing a transport path between VSANs. In Figure 2, device *pWWN_1* in *VSAN_1* needs to communicate with device *pWWN_2* in *VSAN_2*. The border switches act as transport between the devices in the separate VSANs and, as is shown in Figure 2, multiple IVR border switches can be used to provide redundant paths. The border switches inject only selective Fabric Shortest Path First (FSPF) routes into their respective local VSANs relating to devices within the defined IVR-enabled VSANs requiring IVR connectivity. Specifically speaking, only FSPF routes to domains (switches) in the defined IVR VSANs containing these devices are cross propagated. In addition, name server database entries containing identification for the communicating devices within defined IVZs are also cross propagated within the defined IVR VSANs. If all devices were contained within a single VSAN, all database information in every switch would be synchronized. With the IVR feature, only a small subset of routing and name server information for the devices requiring IVR communication as defined by the IVZs is cross-propagated—again without merging the participating virtual fabrics (VSANs).

Figure 3

IVR with Transit VSAN Scenario



Another major advantage of the IVR approach to SAN internetworking is that IVR relies on standard FSPF routing for its path selection and path resiliency. Using other internetworking solutions such as iFCP, there is no end-to-end use of the FSPF protocol (due to the conversion to iFCP), so no end-to-end routing intelligence and resiliency. A solution like iFCP strictly relies on host-based multipathing solutions to implement resiliency. However, since IVR propagates selective FSPF routes through the IVR border switches, simple standard FSPF is used to select paths and define redundant paths from an end-to-end fabric perspective. This capability provides a faster, more reliable recovery mechanism due to any path failure in the topology, which cannot be guaranteed by a solution using iFCP.

The IVR Transit VSAN

An IVR configuration option particularly useful for extended SAN disaster recovery solutions is the use of an IVR transit VSAN. Transit VSANs are additional VSANs created to bridge two edge-IVR VSANs, alleviating the need to extend one IVR VSAN to another to set up the IVR service. In a sense, a transit VSAN extends the reach provided by a border switch to reach participating IVR VSANs. As shown in Figure 2, the two edge IVR VSANs that had devices needing to communicate were already connected to common VSAN border switches. In that simple configuration, no transit VSAN is required. In fact, a transit VSAN is always optional, but is recommended to maintain additional isolation in some configurations such as wide-area SAN extension solutions. Figure 3 shows how a transit VSAN can be used to connect devices from three different edge VSANs across a WAN. Note that a transit VSAN could contain devices as well, and multiple transit VSANs could exist between edge VSANs. As is shown in Figure 3, the IVR service is established without the need to extend any of the edge IVR VSANs to a common border switch. In this example, this prevents the need to extend a local IVR VSAN across a higher-latency, long-distance network. Even though selective connectivity is enabled through the IVR function, the three edge VSANs remain completely isolated from a control protocol perspective—and even from a management perspective—using VSAN-based RBAC within the Cisco MDS 9000 Series. Legacy Fibre Channel switches that are not made by Cisco can also participate in any edge IVR-enabled VSAN, can communicate across inter-VSAN boundaries, and can be part of a transit VSAN—but cannot participate as border switches.

In Figure 3, fabric control traffic for IVR-enabled VSAN_A and VSAN_B is terminated locally at the border switches and control traffic for IVR-enabled VSAN_C is terminated locally at remote border switches. The transit VSAN is the only VSAN that spans over the wide-area connection between border switches. Therefore, any fabric events that occur in either of the edge IVR VSANs, whether it be normal control traffic or fault events, remain within the respective edge VSANs and do not traverse nor affect other IVR-enabled VSANs or transit VSANs. In a stable environment, all switches participating in IVR exchange negligible amounts of control traffic. During normal ISL link initialization between two given switches, the switches exchange a small amount of control protocol information. In addition to this information, two switches bordering a transit VSAN also exchange FSPF and name server database information strictly pertaining to

devices that will be IVR connected over the transit VSAN. This exchange adds approximately two Fibre Channel frames worth of traffic for each device requiring IVR connectivity.

Another major advantage of the IVR solution relative to other internetworking approaches such as those using iFCP is that there is no restriction on the transport technology that can be used for the transit VSAN. The IVR service strictly relies on basic Fibre Channel services, and can use any transport option available to Fibre Channel such as Fibre Channel, dense wavelength-division multiplexing/coarse wavelength-division multiplexing (DWDM/CWDM) transport, SONET transport, or FCIP. With the iFCP-based solutions, iFCP is strictly an Ethernet/IP-based protocol and requires an added gateway device.

A further advantage of the IVR solution is that it does not create or require any additional new control protocols. The IVR service strictly uses common mechanisms such as VSANs and zoning, along with some capabilities within the Cisco Fabric Manager.

Fabric Stability and Resiliency

VSANs provide fabric stability by isolating fabric services within each virtual fabric or VSAN. When Cisco MDS 9000 Series switches are enabled for a new VSAN, all participating switches within the new virtual fabric exchange their device, forwarding, and zone databases to help ensure consistency for the new VSAN. The participating switches also build routing topologies based on the VSAN-specific FSPF link-state information that is exchanged. During the virtual fabric initialization process, one switch is elected to become the *principal* switch in the virtual fabric. The principle switch will assign a set of VSAN-specific domain IDs to the other switches used for addressing and routing purposes in the new virtual fabric. If a switch loses contact with the principal switch, the remaining switches begin a new initialization process and elect another principal switch for the VSAN. Switches will again exchange databases, resulting in a momentary exchange of information dependent on the number of switches in the virtual fabric and the number of devices in the VSAN.

When considering extended SAN solutions, long-distance network transport services are only as reliable as detailed by the service level agreements (SLAs) given by the service provider. While connectivity interruption problems with long-distance links are a concern, their occurrence today is quite rare. In the event of a long-distance transport outage, IVR minimizes disruption and the flood of database synchronization traffic. Figure 3 showed an extended SAN solution using IVR and a transit VSAN. The transit VSAN elects its own principal switch that resides on one side of the long-distance link. The individual edge VSANs also elect their own principal switches within their respective VSANs. If the long-distance transport service were to fail, switches in edge IVR VSANs that have devices in IVZs will receive selective RSCN (registered state change notification) that particular switches (domains) on the remote IVR VSANs are no longer available. However, unlike the case where all switches on both sides of the long-distance transport service are contained within a single SAN or VSAN, the edge IVR VSANs connected to the transit VSAN continue to operate without interruption because the principal switches in their respective VSANs are still available. The control plane isolation offered by the IVR configuration and the transit VSAN eliminates the fabric interruption in the edge VSANs resulting from a link failure outside of the edge VSANs themselves.

A significant advantage of the IVR solution over solutions using iFCP is the fact that the IVR solution uses standard Fibre Channel fabric routing and notification services such as RSCNs on a complete end-to-end basis, with no requirement to translate such control information to different control information in another protocol such as iFCP. This significantly raises the availability of the IVR solution over an iFCP-based solution, due to the contiguous and common fabric alerting mechanism of the RSCNs.

Resilient Designs with IVR

VSANs and IVR are based on the same routing principles and mechanisms well-established in Fibre Channel—FSPF-based topology maps. Links between IVR border switches that cross transit VSANs are just like any other Fibre Channel ISLs. They are based on Fibre Channel and can be transported using various options, including CWDM/DWDM, SONET/SDH, or FCIP. Multiple ISLs can be aggregated into port channels for link resiliency and bandwidth capacity up to a maximum of 32 Gbps. If using FCIP as the long-haul transport, the virtual ISLs created from FCIP tunnels can also be aggregated into a maximum 16-Gbps virtual port channel using the fabric-based port channel technology. The administrator does not need to configure any additional components or transport/gateway protocols to enable the IVR service.

Uses for IVR

IVR is a logical extension to the capabilities offered by Cisco VSAN technology. The uses for IVR are outlined below.

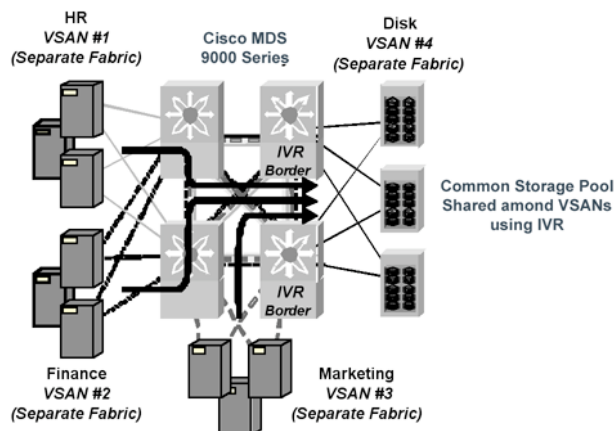
Sharing Access to Storage Devices Across VSAN Boundaries

The resiliency offered by VSANs is due to the replication of fabric services per VSAN along with the explicit tagging of frames belonging to each VSAN. This level of resiliency or isolation mimics the same level achieved through building multiple physically separate SAN islands, using [??] as a more cost-effective model. Prior to the IVR feature, there was no way for traffic to flow across VSAN boundaries, as all isolation mechanisms are enforced in hardware with explicit frame tagging.

IVR is the only mechanism allowed to forward selective traffic across VSAN boundaries. IVR enables cross-VSAN communication without merging the VSAN fabrics. This is an important characteristic, as the separate fault domains are still maintained within each VSAN. Only devices that must communicate to other devices across VSAN boundaries use the IVR technology. Each individual edge VSAN still maintains its own fabric resiliency from a control protocol and management perspective. Since VSANs offer a method of scaling fabrics to high port counts, IVR provides scalable and resilient methods to share connectivity to common storage devices within these larger fabrics without completely merging the entire fabric. For example, one may consider building application-specific or OS-specific VSANs such that all the servers associated with a particular application or OS are contained within a unique VSAN for higher resiliency. Using IVR technology, common large-scale enterprise arrays can be shared among these numerous VSANs—without merging the individual virtual fabrics and, hence, fault domains. Figure 4 shows an example of using IVR to share access to disk arrays among three separate departmental VSANs; all VSANs are still maintained as separate fabrics for manageability and stability.

Figure 4

IVR Example #1 Sharing Target Ports



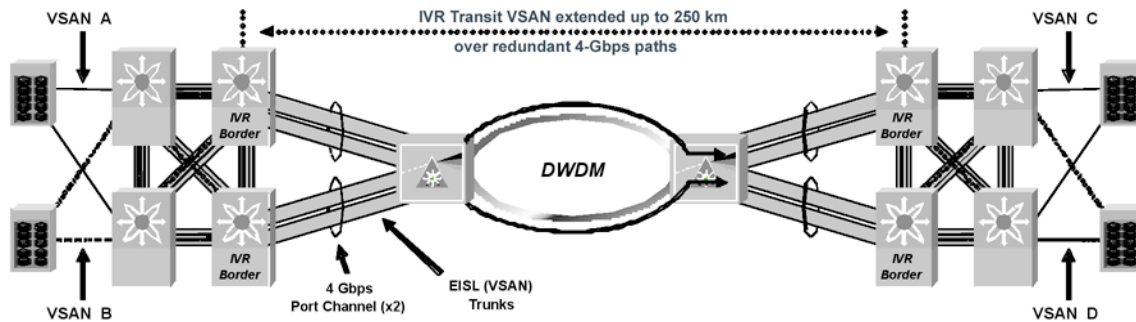
Extended SAN Solutions for Data Replication

IVR enhances the solution provided by VSANs for connections between devices that will only communicate across a SAN extension service and are not used within either local SAN. An example of this is the use of large enterprise storage arrays that are configured to replicate data to a remote data center. Remote replication solutions include such array-based services as EMC SRDF, HDS Truecopy, IBM PPRC, and HP Continuous Access replication services. Most of these arrays require that controller ports be dedicated to the replication service, precluding hosts from using these ports. The simplest solution is to create an additional VSAN that contains only these storage array ports on both ends of the solution used for the replication service, and to permit only that VSAN to traverse the long-distance link. In such a configuration, the amount of control traffic in the designated VSAN would be minimal, as the configuration in the VSAN would remain relatively static and contain very few devices. However, IVR adds an extra layer of resiliency by providing a control protocol boundary between the two ends of the solution, only allowing relevant FSPF routes and name server entries to be exchanged between the two ends. Fabric events, whether intentional or not, remain isolated to the respective edge VSANs and would not propagate over the long-distance link.

However, using IVR, inter-switch traffic can be minimized between SANs connected over long distances. For SAN extension solutions, the use of IVR is also beneficial when the latency between the two endpoints of the SAN extension is higher than within the data center. Only relevant subsets of control protocol information for devices defined as part of IVZs are transported across long-distance links. In addition, separate fabric timers such as ED_TOV can be tuned for the transit VSANs to accommodate the higher transport latency. The fabrics at either end of the SAN extension remain isolated and controlled as separate domains, offering high resiliency of the overall data replication solution. Figure 5 shows an example of IVR combined with DWDM and port channels to build a highly resilient extended SAN solution.

Figure 5

IVR Example #2 SAN Extension with DWDM and Port Channels



IVR Best Practices

The following are best practices recommended to follow when deploying the IVR solution.

Use Transit VSANs for Long-Distance Solutions

Although not completely necessary, IVR transit VSANs provide the optimal amount of resiliency and isolation for an extended SAN solution. These extended SAN solutions can be point-to-point or point-to-multipoint and may involve one or more defined transit VSANs. It is highly recommended to use transit VSANs for these solutions.

Use Zoning, VSANs, and IVR as Necessary

Cisco MDS 9000 Series Multilayer Fibre Channel switches use hardware mechanisms to enforce zoning configurations. In essence, traffic from a given device can only travel to other devices within their defined zones. Zoning provides hardware security and data traffic isolation and is a subset of the isolation qualities offered through VSANs. It is recommended to use VSANs if fabric isolation is necessary at the control protocol and data level such that the isolation provided by separate physical fabrics is to be mimicked. It is recommended to use IVZs if fabric control and data traffic isolation is desired along with enabling selective communication between devices in different VSANs. Please refer to the Cisco white paper “Using VSANs and Zoning in Cisco MDS 9000 Series Multilayer Fibre Channel Switches” for more information.

Create Redundant Inter-VSAN Routes

One of the major advantages of IVR is that it uses standards-based Fibre Channel routing principles and mechanisms. It is therefore recommended to use the same practices to create a highly available inter-VSAN-routed SAN that would be used to create a highly available Fibre Channel SAN. Use redundant inter-VSAN border switches, use port channels in between switches, and use redundant links to devices as would be the common practice in a large-scale SAN fabric.

ALTERNATIVE APPROACHES TO SAN INTERNETWORKING

Cisco offers IVR technology as a natural evolution of the VSAN’s capability. IVR requires no additional protocols, is enabled and managed through the well-known workflow of creating a zone and activating it, and has no negative performance impacts. Any transport that can carry Fibre Channel, such as Fibre Channel, CWDM/DWDM, SONET/SDH, or FCIP, can easily be used with the IVR solution. There are no special control protocols or frame types that are used to support the IVR capability; legacy SAN switches can be used in the solution interoperating with Cisco MDS 9000 Series switches so long as the IVR border switches are Cisco MDS 9000 Series switches. In

addition, there are no restrictions on the protocols that can be carried over the IVR solution, including many array-based, in-band management protocols and other FC-4 protocol types.

Attempts have been made to provide this level of resiliency and isolation—particularly relating to SAN extension solutions that, while interesting in concept, often require learning new protocols and ways of provisioning and managing the solution. These alternative solutions generally require additional dedicated redundant gateway devices that must be added to the solution. The following section looks at one particular alternative approach using iFCP.

iFCP

iFCP was proposed to the Internet Engineering Task Force (IETF) by Nishan Systems (now McDATA). Although iFCP was proposed and accepted by the IETF as a working draft, Nishan Systems is the only vendor implementing iFCP, which has relegated it to be somewhat proprietary. The iFCP solution takes a completely different approach to SAN internetworking than that of Fibre-Channel-based solutions. The original intent of iFCP was to replace a Fibre Channel fabric with an IP network and to provide gateway services such that Fibre Channel end devices could communicate over the IP backbone. The perceived advantage was to use a well-known protocol like IP and the autonomous region isolation capabilities offered by IP routing protocols to add resiliency to the SAN. All Fibre Channel devices in an iFCP SAN communicate among themselves locally using a User Datagram Protocol (UDP)/IP-based iFCP transport, and communicate remotely over a TCP/IP-based iFCP transport. The main idea behind iFCP was to terminate all Fibre Channel traffic as close to the source as possible, such as hosts and disks, and translate that traffic to iFCP.

The motivation for deploying local UDP-based iFCP solutions diminished rapidly, due to the reliability limitations of UDP and the lack of compelling return on investment offered by the Fibre Channel replacement solution. The iFCP solution still required Fibre Channel device attachments along with extra FC-to-IP gateway devices, adding to the exorbitant costs of the solution. With the availability of Small Computer System Interface over IP (iSCSI) solutions today, customers that want to build IP-based SANs can use TCP/IP-based iSCSI to build an IP end-to-end solution and realize the full cost savings offered by an IP SAN. As such, iFCP has been relegated to become a SAN interconnect gateway solution using the TCP/IP variation of iFCP.

As a gateway solution, iFCP must provide a mapping function between local and remote Fibre Channel fabrics across an IP network. The iFCP solution replaces the FC-2 layer of the Fibre Channel stack with an iFCP equivalent. However, the FC-2 layer of the Fibre Channel protocol stack is responsible for addressing and routing, which requires iFCP to modify each frame's addressing information to forward across an iFCP service. In contrast, FCIP provides a completely transparent tunneling of Fibre Channel over a TCP/IP tunnel, and therefore does not manipulate any frame addressing or content. The iFCP solution requires address translation to be performed on each Fibre Channel frame as it traverses the iFCP service, and in some cases, such as with certain Fibre Channel ELS (extended link service) control messages, must also translate parts of the frame payloads.

Another drawback of the iFCP solution is that it must selectively handle FC-4 services and does not directly support all FC-4 services due to the specific mapping required by the iFCP (FC-2) layer. For example, IBM Fiber Connection (FICON) over iFCP is not a supported FC-4 type, but services like FICON over FCIP can be accomplished because FCIP is transparent to the Fibre Channel FC-2 protocol. The iFCP protocol must therefore process various FC-4 frame types, FC-2 frame link service requests, Fibre Channel broadcast traffic, and iFCP control frames in different ways.

The iFCP solution must also emulate all the mandatory Fibre Channel services to Fibre Channel devices and provide coordination between multiple iFCP gateways that must service Fibre Channel end device communication. Some examples of Fibre Channel services include the login server, name server, notification server, alias server, multicast server, and broadcast server. For instance, a device logs in to the fabric through the login server, registers to receive notifications of changes in the fabric through the notification server, and uses the name server to determine the devices it can communicate to. All of these services have to be replicated in the IP world to accommodate this new iFCP protocol. So, to accommodate these services, the iFCP solution was to create yet another protocol.

iSNS for iFCP

iSNS is the mechanism that emulates a subset of Fibre Channel services in iFCP environments. The iSNS service is absolutely required for iFCP to operate, but can be extended to other protocols such as iSCSI. The following services are provided by the iSNS server—device discovery and fabric management, emulation services provided by the Fibre Channel name server and RSCN service, definition and management of discovery domains, and definition and management of *logical fabrics*. Although all these services are provided natively by Fibre Channel in an FCIP solution, iFCP must have at least one iSNS server configured to provide these critical services. In addition, if an iSNS server fails, a delay and traffic disruption is involved while a new iSNS master server is elected. Each local Fibre Channel fabric that is connected over iFCP gateway links must have iSNS servers that can exchange information for devices that are exported across the iFCP network.

One major challenge for iFCP is the fact that the absolute requirement for the iSNS service segregates the fault alerting and recovery mechanisms in the overall fabric. There is no close linkage between the alert and response mechanisms of Fibre Channel and those of iFCP (IP). This is in contrast to an FCIP-based solution, where the alert and response mechanisms of Fibre Channel are common and transparently extended throughout the entire fabric. Using Cisco VSAN technology, VSANs can be used to segregate the larger physical fabric into smaller virtual fabrics, each with their own alert and response mechanisms and scopes. VSANs therefore help limit the impact of fabric alerts and responses within the VSAN in which they occur; selective VSANs can then be extended across an FCIP solution. IVR can be added to the VSAN-based FCIP solution to further reduce the scope of fabric alerts and responses to only those devices requiring IVR connectivity, all while still using the same common Fibre Channel fabric mechanisms.

Another challenge for iFCP comes from the service that iSNS provides. In an edge Fibre Channel fabric reconfiguration, all of the iSNS address translation tables need to be recalculated, involving a consequential loss of every active login session. This major vulnerability can occur simply by someone adding a new switch into a remote fabric causing a reconfigure fabric (RCF) event. These vulnerabilities erode the claimed resiliency of the iFCP solution. There is no better fabric alert and response mechanism than a common end-to-end mechanism, which is provided through the use of FCIP.

Figure 6

iFCP Functional Model

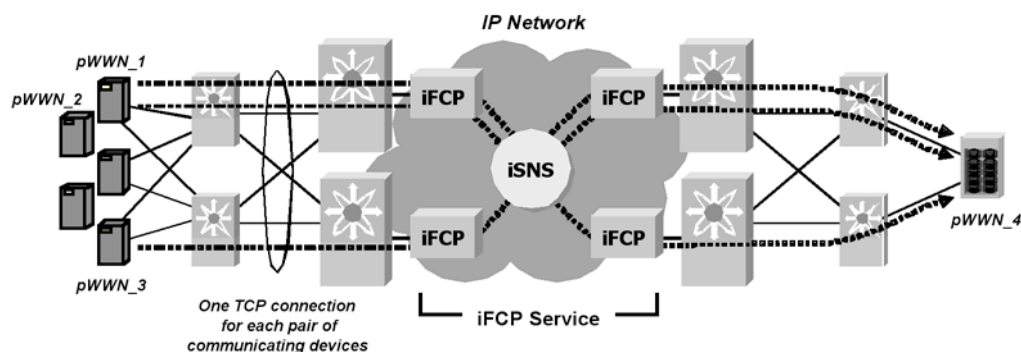
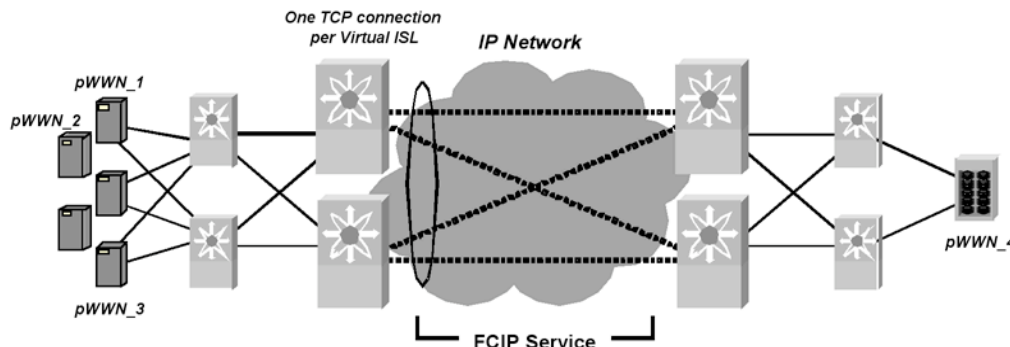


Figure 7
FCIP Functional Model



iFCP and SAN Extension

TCP, which runs over IP, enhances the reliability of IP-based links to be similar to that of Fibre Channel. The gateway function of iFCP runs over TCP/IP to provide this reliability. The transport function provided by iFCP requires a gateway to create a TCP/IP connection for each device pair needing to communicate across the iFCP service. TCP/IP flow control and error recovery must be managed for each connection and therefore each device pair that communicates across the iFCP gateway service.

A handful of devices communicating across the iFCP service does not present an overhead challenge, but requires more processing power and buffering for each additional pair that communicates across the gateway. Each TCP connection that relates to each pair of devices communicating over iFCP requires more memory and CPU to support the multiple TCP windowing and recovery mechanisms. In addition, the failure of one IP transport link can mean the termination of many TCP connections, which will all require simultaneous re-establishment. This re-establishment event, combined with the number of software RSCNs that need to be generated, may overwhelm the iFCP gateway processor and slow the reconvergence.

To contrast this with an FCIP-based solution using IVR where appropriate, FCIP creates a single TCP connection for each Fibre Channel tunnel, or virtual ISL between two switches. The number of devices using the FCIP connection between two Fibre Channel switches does not affect the TCP throughput or fault recovery performance. Using the IVR capability along with FCIP provides the best overall solution, with the required amount of control protocol isolation along with a scalable, high-performance TCP/IP transport.

iFCP Routing

Contrary to much of the documentation that surrounds iFCP, the protocol does not provide any routing function. The iFCP service is a simple gateway function that translates and forwards Fibre Channel traffic to IP target gateways. The routing function for iFCP is provided by the IP routing protocols within the IP transport network. However, because iFCP does not route, it represents a segmentation of the end-to-end fabric topology and routing protocols between Fibre Channel and IP. Fibre-Channel-based FSPF routing has no visibility or path information about the IP network and simply forwards traffic to the configured iFCP gateway. Once the Fibre Channel traffic is translated using iFCP to an IP-based packet, it is forwarded based on IP routing protocols and no longer based on Fibre Channel routing protocols. This loss of visibility prevents the autodiscovery and use of redundant paths across the entire infrastructure from end to end, thereby affecting the resiliency of the solution. As a result, no iFCP network load balancing techniques exist. Compensating for this segmented topology, an iFCP solution must use a host-based multipathing software component to add necessary resiliency. To contrast this with an FCIP-based solution, since FCIP is a transparent tunneling of Fibre Channel across a TCP/IP connection, each FCIP tunnel represents a virtual fabric ISL and participates in FSPF routing topology determination. Therefore, within an FCIP solution, FSPF is a

consistent routing protocol that builds and maintains an entire topology across the Fibre Channel and Ethernet networks. This end-to-end visibility enables FSPF to quickly determine and use redundant paths anywhere in the infrastructure should a network fault occur.

SAN Extension Supporting Services

Nishan Systems has developed both IP-layer compression and *fast-write* features for its iFCP-enabled product line, but Cisco also provides compression and FCIP *write-acceleration* services as part of the embedded feature set within the entire Cisco MDS 9000 Series. Cisco's compression and write-acceleration performance have been proven to outperform those offered as part of the iFCP solution. The added bonus is that the management of these services is also embedded into the Cisco Fabric Manager suite, along with the management of Fibre Channel, iSCSI, FCIP, VSANs, IVR, and other services.

iFCP Challenges

Although iFCP has been in existence as an IETF RFC draft for some time, it has not gained any significant following other than by Nishan Systems (now McDATA). Its lack of acceptance has to do with its complexity and benefit compared with other alternatives that exist in the industry. iFCP initially was targeted to replace the core of a Fibre Channel SAN with IP, but this was not accepted, due to the introduction of iSCSI as a true end-to-end IP SAN technology. Although iFCP has been confined to a SAN extension technology, other, more widely accepted technologies such as FCIP have provided more simplistic and transparent approaches for SAN extension solutions.

iFCP faces not only technical challenges due to its lack of resiliency and compatibility with upper-layer protocols, but it also faces challenges related to its manageability. Administrators must learn new techniques to manage products that implement iFCP and need to manage these in addition to their SANs. iSNS represents another new service that must be understood and managed by system administrators to enable the iFCP service. iFCP-based products need to have a large amount of buffering, because new TCP connections are created between each device pair that communicates across the service. The TCP processing overhead grows rapidly as more devices use the iFCP gateway service. Management and troubleshooting is often difficult with so many connections. Additionally, SAN extensions using iFCP will always be limited to FCP (non-mainframe) applications and must actively translate and manipulate each frame of storage traffic to IP, which may be excessive.

Cisco has introduced several integrated technologies in the Cisco MDS 9000 Series that are simple to deploy and manage, compared to an iFCP solution. VSANs, IVR, and FCIP all use the same set of Fibre Channel routing and alert and response mechanisms to provide a scalable and resilient end-to-end infrastructure.

The following list outlines some of the challenges faced by iFCP-based solutions and a comparison of a Cisco solution involving FCIP, VSANs, and IVR.

1. ***iFCP requires additional gateway devices***—iFCP is only available through the use of redundant pairs of iFCP gateway devices, which are only available from one vendor—Nishan Systems (now McDATA). No integrated solutions exist, making the overall solution unnecessarily complicated. In contrast, VSANs, IVR, and FCIP are all available as an embedded solution in any Cisco MDS 9000 Series multilayer switches, including both directors and fabric switches.
2. ***iFCP requires additional management tools***—iFCP solutions not only involve the iFCP protocol, but the iSNS protocol as well. These solutions must be managed through separate tools available from Nishan Systems. No integration exists with any fabric management tools available today, again making the management solution unnecessarily complicated. In contrast, Cisco provides the Cisco Fabric Manager suite as part of every Cisco MDS 9000 Series switch, integrating the management of Fibre Channel, iSCSI, FCIP, VSANs, IVR, and several other features in one management suite. In addition, third-party management suites, including EMC ECC, IBM Tivoli, Veritas SPC, and HP OV-SAM, continue to integrate increasing capability to manage Cisco MDS 9000 Series products and features directly.

3. ***iFCP requires knowledge of two new protocols***—iFCP involves iFCP and iSNS, which are two new protocols for SANs. Administrators who manage the SAN and iFCP services must be able to troubleshoot these complicated protocols. In contrast, Cisco VSANs and IVR are strictly based on Fibre Channel concepts, including zoning, and are completely integrated into the Cisco management suite. In addition, FCIP is a transparent mapping of a Fibre Channel ISL to a TCP connection and is managed as an ISL from within the Cisco Fabric Manager.
4. ***iFCP has FC-4 type restrictions***—iFCP effectively replaces the FC-2 layer of the Fibre Channel protocol stack. For this reason, iFCP is only compatible with certain FC-4 services that rely on the FC-2 layer service for transport. In contrast, FCIP is a transparent mapping of a Fibre Channel ISL to a TCP connection. Any FC-4 services are easily transported across an FCIP tunnel, including such services as IP-over-FC (IETF RFC 2625) for remote in-band management and FICON services.
5. ***iFCP requires extensive TCP connections***—iFCP requires a new TCP connection for each pair of communicating devices across an iFCP service. Each TCP connection requires additional CPU and memory resources to manage the TCP connection and windowing service, limiting the overall performance as more devices are added to the service. In addition, if a physical link fails, numerous TCP connections may be affected and require re-establishment—consuming even more CPU resources. Each TCP connection requiring re-establishment would also require a series of software RSCNs that must be created and transmitted. In contrast, an FCIP service only requires a single TCP connection per virtual ISL created and is not dependent on the number of devices using the FCIP tunnel. Also, if a physical transport link fails, the representation in FCIP is that of a failed ISL, and a single software RSCN signaling the loss of a domain is generated.
6. ***iFCP Network Address Translation (NAT) requires frame-by-frame manipulation***—iFCP translates addressing from local SAN domains to remote domains and must manipulate each frame. This causes some problems for Fibre Channel management protocol frames such as ELS frames that embed the Fibre Channel Address ID (FCID) address inside the frame. In contrast, an FCIP solution provides a completely transparent tunneling of Fibre Channel ISL traffic. Using VSANs and IVR further limits the amount of control traffic that must traverse the FCIP solution, but still does not manipulate any frames.
7. ***iFCP requires iSNS service to function***—Part of the iFCP solution is to replicate common Fibre Channel fabric services into IP equivalents for use by the iFCP gateways. This is accomplished through the creation of an additional protocol (iSNS). Without the iSNS service, iFCP cannot function. In addition, if an iSNS server fails, a delay and traffic disruption is involved while a new iSNS master server is elected. In contrast, a Cisco solution involving FCIP, VSANs, and IVR only requires the standard set of Fibre Channel services already present in the fabric. No additional network services are required.
8. ***iFCP only uses Ethernet***—Because iFCP must use an Ethernet transport, it cannot provide any domain capability within a Fibre-Channel-only environment. In addition, if other transport protocols are to be used in the solution, a conversion to Ethernet must happen first, and then another conversion to the chosen transport protocol. For example, a SAN extension solution using SONET and iFCP must first translate Fibre Channel traffic to Ethernet-based iFCP and then again to SONET. In contrast, VSANs and IVR can be used either within a Fibre-Channel-only scenario or extended across any transport protocol directly such as SONET or an IP network using the transparent FCIP protocol.
9. ***iFCP segments the routing domain***—iFCP uses the routing services provided by an IP network, providing a gateway function between fabric-based FSPF routing and IP-based routing. However, it does not extend either on an end-to-end basis, which eliminates any load-balancing capability. No end-to-end routing capability results in the inability of the Fibre Channel fabric to dynamically reroute or load balance through other redundant paths in an iFCP-based solution. FSPF within the Fibre Channel fabrics only sees the iFCP gateways and nothing beyond them. In contrast, a solution using Cisco IVR with or without FCIP can take advantage of FSPF routing topologies built from end-to-end visibility of all pertinent domains across Fibre Channel and the IP network. As such, FSPF can build and use complete redundant pathing throughout the end-to-end network, raising the resiliency of the solution. Cisco VSAN technology is used to further limit the routing topology to only those devices requiring access to the SAN extension service.

Table 1. Summarizes the FCIP Protocol, the IVR Technology, and the iFCP Solution.

Feature	FCIP+VSANs	FCIP+IVR+VSANs	IVR+VSANs	iFCP
IP implementation	Fibre Channel ISLs built over transparent TCP/IP connection	Fibre Channel ISLs built over transparent TCP/IP connection	N/A	New protocol to make FC devices IP-visible through another new protocol (iSNS)
Hardware implementation	Embedded into Cisco MDS 9000 Series switches	Embedded into Cisco MDS 9000 Series switches	N/A	Separate gateway devices
TCP connection method	Connection per switch-to-switch FCIP tunnel	Connection per switch-to-switch FCIP tunnel	N/A	Connection per device-to-device pair
Directory services	Basic Fibre Channel name service	Basic Fibre Channel name service	Basic Fibre Channel name service	Totally dependent on iSNS service
Transparent to end Fibre Channel devices	Yes	Yes	Yes	No; restriction on FC-4 types and control protocols
Uses Fibre Channel fabric routing and load balancing	Yes	Yes	Yes	No; segmented routing topology at iFCP gateway; no load balancing
Supports mainframe environments	Yes	Yes	Yes	No
Requires address conversion	No	No	No	Yes; frame-by-frame address conversion
Dependence on IP	Only for FCIP tunnels used to build virtual ISLs	Only for FCIP tunnels used to build virtual ISLs	No	Completely designed around IP-IP version of Fibre Channel services through iSNS
Eliminates propagation of local SAN disruptions	Can be reduced by VSANs	Yes	Yes	Yes
Minimizes control traffic across IP link during initialization	Can be reduced by VSANs	Yes	N/A	Yes
Minimizes control traffic across IP link during steady state	Can be reduced by VSANs	Yes	N/A	Yes
Out-of-band control traffic	None required	None required	None required	iSNS management traffic
Scalability limitations	Amount of control traffic grows as more switches are added to VSANs riding over FCIP	None	None	Scalability limited by TCP buffering and CPU required for each device-to-device connection
Interoperability effects	None	None	None	Substantial
Support for emerging Fibre Channel security standards	Yes	Yes	Yes	No
In-band administration	Yes (via IPFC [SAME AS FCIP?])	Yes (via IPFC[SAME AS FCIP?])	Yes (via IPFC[SAME AS FCIP?])	No
Out-of-band administration	Optional	Optional	Optional	Required
Industry support for IP extension technology	Wide support for FCIP	Wide support for FCIP	N/A	Only Nishan (now McDATA)

**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 Web site at 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 © 2004 Cisco Systems, Inc. All rights reserved. Cisco, Cisco Systems, and the Cisco Systems logo 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. (0403R)
203154_ETMG_5.04