White Paper

Fabric Services

Tom Clark
Director, Technical Marketing

April 2000

VIXEL

MAKING THE FIBRE CHANNEL CONNECTION

Introduction

A fabric is one or more Fibre Channel switches connected together as a network. Typically, fabrics are used to build storage area networks (SANs), although peer-to-peer IP networks can also be constructed. Fabrics may include arbitrated loop hubs in addition to switches so that switched bandwidth for high speed access can be combined with shared bandwidth for less demanding traffic. Analogous to Ethernet network designs, a fabric can be extended by connecting additional switches and hubs to accommodate hundreds or thousands of devices.

In Fibre Channel SANs, fabric switches are often compared to loop hubs and loop switches. How do they differ? Loop hubs provide a shared 100 MBps bandwidth for the attached devices. As more devices are added to the hub (or cascade of hubs), the bandwidth available for each active device (i.e. servers) declines. Loop switches provide 100 MBps bandwidth per port, and so do not suffer performance degradation as more active devices are added. In both loop hub and loop switch architectures, however, the hub or switch is transparent to the end devices. End devices communicate with each other and do not establish sessions with the intervening hub or switch. Therefore, at the protocol level, the end devices do not need to be fabric-aware.

Early market Fibre Channel products did not support fabric services such as fabric logon (FLOGI), Simple Name Server (SNS) registration, or Registered State Change Notification (RSCN). Before mid-1999, most Fibre Channel Host Bus Adapters, Fibre Channel disks, and FC-SCSI routers only supported stand-alone arbitrated loop or point-to-point connections. Because these configurations could not be directly connected to larger switched networks, they are referred to as *private* configurations. For a device to be *public*, it must be able to make its presence known to the fabric, perform login, SNS, RSCN and other functions that allow network-wide communications. Currently, most Fibre Channel HBAs, disk controllers and routers support fabric services. The legacy, private-only devices can be accommodated with a special feature most fabric switches provide, generically called *translational mode*.

Fabric Devices

Host Bus Adapters, disk arrays, and FC-SCSI routers that support fabric services have special microcode and device drivers to provide fabric login, SNS registration, and RSCN. If the fabric-aware device is attached to a fabric switch, it will execute each of these functions to become part of the public network. What happens if fabric-aware devices are connected to a stand-alone loop hub? Since their attempt at fabric login will fail, they default to private loop behavior and communicate happily amongst themselves. If the hub is at some point connected to a fabric switch, the devices will reinitialize. Their renewed attempt to perform fabric login will now succeed, and they will be able to communicate across the fabric to other public devices as well as their peers on the same loop hub. Fabric-aware devices, fabric switches and arbitrated loop hubs are therefore completely compatible, although this fact is often obscured by vendors of fabric switches.

Why Fabric Switches?

Since Fibre Channel products may be used in a variety of network solutions, the choice of fabric switches over loop hubs or private loop switches should be based on actual application requirements. A small server cluster with 4 servers, 2 disk arrays, and a tape backup subsystem, for example, may not require switched bandwidth or fabric services. If each server only requires 25 MBps or less bandwidth, an economical loop hub would be adequate. Scaling to larger configurations, a private loop switch could provide a full 100 MBps bandwidth to each server or disk array without requiring fabric support on the end devices. In either case, the advanced services offered by a fabric switch may offer little value compared to the higher per-port cost fabric switches command.

The address space for arbitrated loop (shared or switched), however, only provides support for 126 devices (plus one fabric attachment). Hubs or loop switches alone are not sufficient to build large enterprise SANs. As a customer's storage network grows from ten to several hundred devices, fabric switches are required to support the larger population. To accommodate this growth strategy, a switch may offer two levels of capabilities: private loop switch only, with a fabric upgrade when required. The Vixel 7000 series switches, for example, provide loop switching support by default and a firmware key upgrade to full fabric services as an option. This allows the customer to more cost-effectively implement advanced fabric services on an as-needed basis.

If an enterprise SAN must support several hundred (or several thousand) devices, this still does not mean that the entire infrastructure must be built with fabric switches alone. It depends on the applications such a network must support. How much bandwidth is required by each application? How many servers must be co-located in the same pool? What distances between servers, storage and tape must be supported? From a network design standpoint, these issues can help determine where shared segments and switched segments should be deployed.

Unfortunately, the popularity of Ethernet switching in enterprise networks has encouraged a perceived need for switching for all SAN applications as well. There is one substantial difference between shared Ethernet and shared Fibre Channel. A shared Fibre Channel loop segment offers ten times the bandwidth (1 gigabit per second) of a typical shared Ethernet segment (100 megabits per second). While a tape subsystem requiring 10-15 MBps may stress the capabilities of a shared Ethernet (or even switched Ethernet) environment, *multiple* tape streams can be supported on a shared Fibre Channel segment. And since fabric-aware devices can be supported on loop hubs which are in turn attached to fabric switches, it is possible to design very large SANs that rationally and cost-effectively deploy bandwidth for specific application requirements.

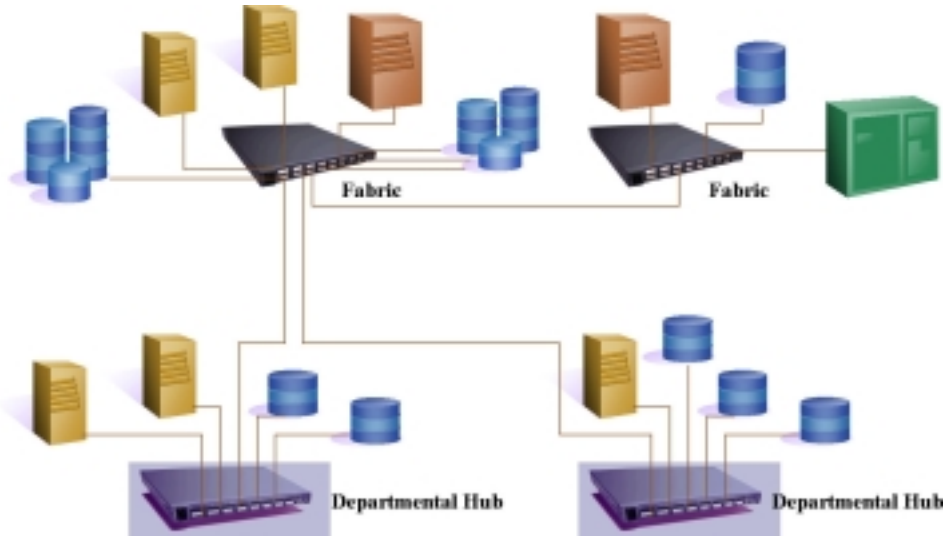

Figure 1 Fabric-aware servers and disks communicating across a shared loop and fabric switch network

As shown in Figure 1, fabric switches can be connected to form a high speed SAN backbone and to deliver 100 MBps bandwidth to traffic-intensive servers and disks. Less demanding applications can be satisfied with downstream loop hub segments. At the departmental level, the majority of traffic may remain on the local loop, with only occasional access across the fabric (e.g., for tape backup to a central resource). And with fabric-capable devices deployed throughout the SAN, any server can access any disk or tape subsystem. For large enterprise networks, this more flexible design can be implemented at one-third the cost of a switch-only solution.

Some switch-only vendors have attempted to eliminate the role of loop hubs altogether by offering variants of their products as private loop switches. For any loop requirement, they suggest installing their own private loop switch. This would be acceptable (although more expensive) if their products offered the flexibility of loop hubs in fabric environments. They do not. To expand the number of ports on a loop segment, for example, hubs can be simply cascaded together. Some loop switches do not support this, or can only be cascaded up to 2 switches. Likewise, attaching a loop hub to a fabric switch is straightforward and requires no upgrade. Some loop switches must be upgraded to full fabric capability before they can be attached to a fabric switch. This adds significantly to the cost of the original product. And while multiple loop hubs can be attached to a fabric switch to provide separate loop segments, some vendors restrict the number of their own loop switches that can be connected to their own fabric switches. Of course, few of these limitations and hidden costs are detailed in the respective vendors' data sheets.

While a SAN design should focus on actual application requirements, fabric switches and fabric-aware devices are essential once the threshold of 126 nodes is reached. At that point, fabric switches are required to join both switched and shared segments, and fabric support on HBAs and disk arrays is required so that all server and disk resources are

reachable across the fabric. As discussed below, there may be applications that benefit from advanced fabric services even if the number of servers and arrays is fairly low.

Fabric Services

Support for fabrics has two basic components: fabric functionality on the Host Bus Adapters, disk controllers, FC-SCSI routers and other peripherals, and fabric functionality on the switch itself. As an end node is connected to a fabric switch, it must first establish sessions on the switch in order to participate in the public network. Since a variety of products may be attached to a central switch resource, interoperability is essential for stable operation. The procedures for fabric conversations are specified in a number of Fibre Channel standards. Standards define expected behavior of the end nodes and the switch resource, but the onus falls on each vendor to insure both standards-compliance and compatibility. Working through interoperability issues for fabric SANs is a central concern for the ANSI working groups, FCIA (Fibre Channel Industry Association), and the SNIA (Storage Networking Industry Association).

Fabric Login

Fibre Channel architecture employs a dynamic addressing scheme that allows the topology itself to manage address assignment. This eliminates the need for manual address management and facilitates scaling very large SANs with minimal administrative overhead. In arbitrated loop, for example, 126 possible addresses are provided for a single loop. An end device acquires one of these one-byte addresses through a loop initialization process which insures that each device has a unique identity on the loop. During initialization, the end devices themselves contend for favored addresses. The loop hub plays no role in the process. As devices are added or removed from the loop, subsequent loop initializations may result in a change in a device's address. This is acceptable because the mapping between Fibre Channel transport addresses and upper layer SCSI addressing is maintained by the device drivers on each end node. Any change in transport addressing is automatically monitored by the topology itself.

The address space for Fibre Channel fabrics is a three-byte field. After subtracting reserved addresses, this allows for approximately 15 _ million unique addresses for end nodes. Since this is somewhat more than the 126 addresses allowed for private loop, a fabric cannot rely on the end nodes themselves to resolve the address selection process. The role of address manager is therefore assume by the fabric switch.

When a fabric-capable device is connected to a switch port, it attempts to log in (FLOGI) to a well-known address, "FFFFFE". The fabric-capable device identifies itself with the address "000000", indicating that it needs to acquire a fabric address. The switch responds to the login request with an acknowledgement that contains a unique three-byte address, and it is this address that the end node will use thereafter.
In addition to address acquisition, the login process may also be used to negotiate service parameters between the end node and the switch, such as the number of frames that can be sent before an acknowledgment is required. The Vixel 7000 series fabric switches, for example, provide 16 receive and 16 transmit buffers per port, which provides highly efficient frame transfer once login is complete.

Simple Name Server

In a private arbitrated loop, initiators (typically servers) discover targets (typically disk arrays) by one of two methods. If all devices on the loop support positional mapping during the initialization process, an initiator may simply check to see what addresses have been entered into the map and then attempt login directly to each address. A port login (PLOGI) sent to each address allows the initiator to establish a session with each disk target. If any loop devices do not support positional mapping, that phase of loop initialization will not be performed. In that case, the initiator may send PLOGI's to each of the 126 possible loop addresses and wait to see who responds. Walking the address space may seem inefficient, but occurs fairly quickly at gigabit speeds.

Walking the address space of 15 _ million possible addresses, however, is tiresome, even at Fibre Channel speeds. Every change in the topology would require polling millions of addresses. Fabric switches rationalize the discovery of targets by initiators by providing a Simple Name Server (SNS) function. The SNS is a small database that contains, among other information, the capabilities of each fabric-attached device, its fabric-assigned address, and its manufacturer-assigned 64-bit World Wide Name.

After a fabric-capable disk as logged onto the fabric, it will send a PLOGI to another well-known address, "FFFFFC". It registers with the Simple Name Server by delivering its vital statistics, including class of service parameters, World Wide Name, fabric address, port type, and, most useful for initiators, the upper layer protocols it supports. Since this information is centrally maintained by the switch, an initiator can simply query the switch's SNS to discover which fabric devices support SCSI-3 protocol (Figure 2). The initiator can then send PLOGIs to each potential target and establish SCSI sessions for storage transactions.

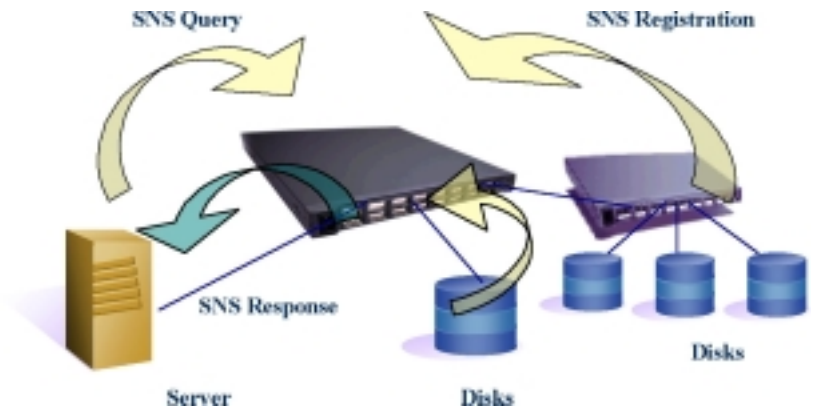

Figure 2 Using the Simple Name Server to expedite discovery of targets by initiators

The Vixel 7000 series fabric services include enhanced support for SNS entries and queries, including the ability to register devices that may not specify the upper layer protocols they support. Certain disk controllers, for example, may register their address and class of service information with the SNS, but fail to indicate their support of SCSI-3 protocol. This would normally prevent an initiator from discovering a disk resource. The 7000 series switch microcode monitors such behavior, and will probe the reluctant device with port and process logins to discover whether it is a SCSI target. If the probe is

successful, the switch will complete the device's SNS entry so that a SNS query from an initiator can find all potential targets. Since probing may be disruptive to some devices, this feature can be enabled or disabled on a port by port basis.

Registered State Change Notification
An initiator may query the SNS to quickly discover potential targets, but it needs another mechanism to discover whether targets have disappeared from or reentered the fabric. The fabric switch's Registered State Change Notification (RSCN) provides this function. Once an initiator has logged onto the fabric, queried the SNS to locate targets, and has established sessions to the desired disk resources, it can register with the fabric to be notified if there is any change in the topology. RSCN provides a function similar to loop initialization in private loop environments since it prompts all initiators to verify that their targets are still available. The Vixel 7000 series fabric switches support RSCN so that servers can be proactively notified if any changes in disk availability occurs.

Fabric Address Notification
Another proactive tool for stable SAN operation is defined by Fabric Address Notification, or FAN. The FAN function allows end devices to verify that their ongoing transactions and operating parameters are still valid, particularly after an initialization has occurred on a fabric loop segment. The Vixel 7000 series FAN implementation allows devices such as tape backup subsystems to resume ongoing backups instead of simply aborting the current job due to temporary disruption.

Broadcast Server
In order to support IP (Internet Protocol) over Fibre Channel, the fabric must facilitate address resolution between Fibre Channel addresses and the upper level IP addresses. For example, an IP-capable device may issue a network-wide query using Address Resolution Protocol (ARP). The responses to this query will give the initiator addressing information that will allow it to build an ARP table of IP to Fibre Channel addresses required for communications. The Vixel 7000 series fabric switches enable such network-wide queries through a Broadcast Server function. A broadcast query from one port will be sent to all other ports, thus allowing the query to propagate to the appropriate devices. Zoning can be used to limit the scope of the broadcast within a single switch.

Other Fabric Functions
Additional fabric-specific functions include Principle Switch and Domain Address Manager. The Principle Switch selection process allows one of the switches in a complex fabric to be the root switch. Without such a procedure, there would be no central management of switch addresses and it might be possible for conflicting switch addresses to occur. Once the Principle Switch procedure has selected the root switch, the Domain Address Manager insures that every switch in a fabric is assigned a unique domain name. Analogous to fabric address assignment to end nodes by a single switch, the Vixel 7000 series fabric implementation of Principle Switch and Domain Address Manager allow the Fibre Channel transport to dynamically configure itself and manage addresses without manual intervention. This helps reduce administrative costs for the SAN and simplifies deployment as new departments and applications are added to the fabric.

Port Auto Detection

Fibre Channel standards define multiple switch port types, including fabric port (F_Port) for direct-attached devices, fabric loop port (FL_Port) for arbitrated loop segments, and expansion port (E_Port) for switch-to-switch connections. While some switch hardware designs require port definitions to be manually configured, the Vixel 7000 series fabric switches provide auto-configuration of ports. Devices on each port are auto detected, and the appropriate port protocol is enabled. For example, if a JBOD (Just a Bunch of Disks) is attached to a Vixel 7200 port, port auto detection would enable arbitrated loop protocol (FL_Port) for that port.

Enhanced Features for Private Loop Devices

The large installed base of private loop-only disk arrays and FC-SCSI routers has prompted the development of new switch features to accommodate legacy devices. As the pioneer of private loop switching, Vixel has innovated proven solutions for both private loop-only and mixed fabric / private loop environments. The Vixel 7000 series switches allow customers to maximize their investment in early market Fibre Channel products while migrating to full fabric functionality.

Public Private Communication
Since disk arrays represent a substantial investment by the customer, it is highly desirable if private loop-only arrays can be brought into a fabric configuration. Vixel 7000 series switches make these resources available to fabric-aware initiators through public to private communication. As discussed above, a private loop device has a single-byte address, known as a Arbitrated Loop Physical Address, or AL_PA. Fabric devices, however, use a three-byte address. Consequently, communication between public and private devices requires a means to spoof or proxy three-byte fabric addressing for one-byte private targets.

Another prerequisite for public to private communications is device discovery. Since public initiators rely on the SNS for a listing of SCSI-capable targets, the switch must be able to insert an entry for each private loop device into the SNS. The Vixel 7000 series fabric switches accommodate both of these requirements. When a private device is discovered, the switch proxies the upper two bytes required for fabric addressing and inserts an entry into the Simple Name Server. When a fabric-capable host queries the SNS, it discovers a series of public targets with which it can establish sessions, unaware that some may be private loop-only devices. Using public to private communication, terabytes of data can migrate from private loops to public fabrics without disruption or additional expense.

Stealth-3 Private Loop Switching
Now in its third generation of engineering development, Vixel's Stealth Mode switching has proven to be the most stable private loop switching implementation in the market. Stealth-3 support on the Vixel 7000 series switches offer exclusive enhancements for switched loop environments, including automatic configuration, Expert Mode, and full cascades of up to 126 devices.

In hub-based arbitrated loop configurations, the loop devices rely on loop initialization to acquire unique AL_PAs. Subtle differences in engineering implementations make some loop devices better citizens than others. Stealth-3's auto configuration allows all well-behaved devices to acquire a unique address, while Expert Mode gives the administrator full control over address assignment for problematic devices. Once address are assigned by auto or Expert Mode means, the Vixel 7000 series switches in Stealth-3 Mode can provide full 100 MBps bandwidth for each port.

Additionally, Stealth-3 provides all 126 possible private loop addresses, unlike other vendor implementations that restrict the number of devices that can be attached. While some vendors only allow two loop switches in a single configuration, Vixel's Stealth-3 supports up to 16 switches in a single virtual loop. This allows very large private loop configurations to be built for customers who do not require fabric services. By using 16-port Vixel 7200s and 8-port Vixel 7100s, a full 126 node switched loop SAN is now possible.

Enhanced Features SAN Management

Vixel's fabric services are engineered to minimize administrative overhead and keep the total cost of storage network solutions down. Fabric support alone, however, cannot maximize utilization of a SAN or assist the administrator in properly sizing SAN

requirements over time. To fully leverage the SAN, advanced management is required. Vixel's SAN InSite 2000 meets this need by providing comprehensive management of both Vixel and third party products, auto discovery and topology mapping of the SAN, and advanced performance monitoring to track bandwidth utilization. As shown in Figure 3, Vixel SAN InSite 2000 provides at a glance status of the SAN (or multiple SAN segments) with ready identification of fabric switch and end device connectivity.

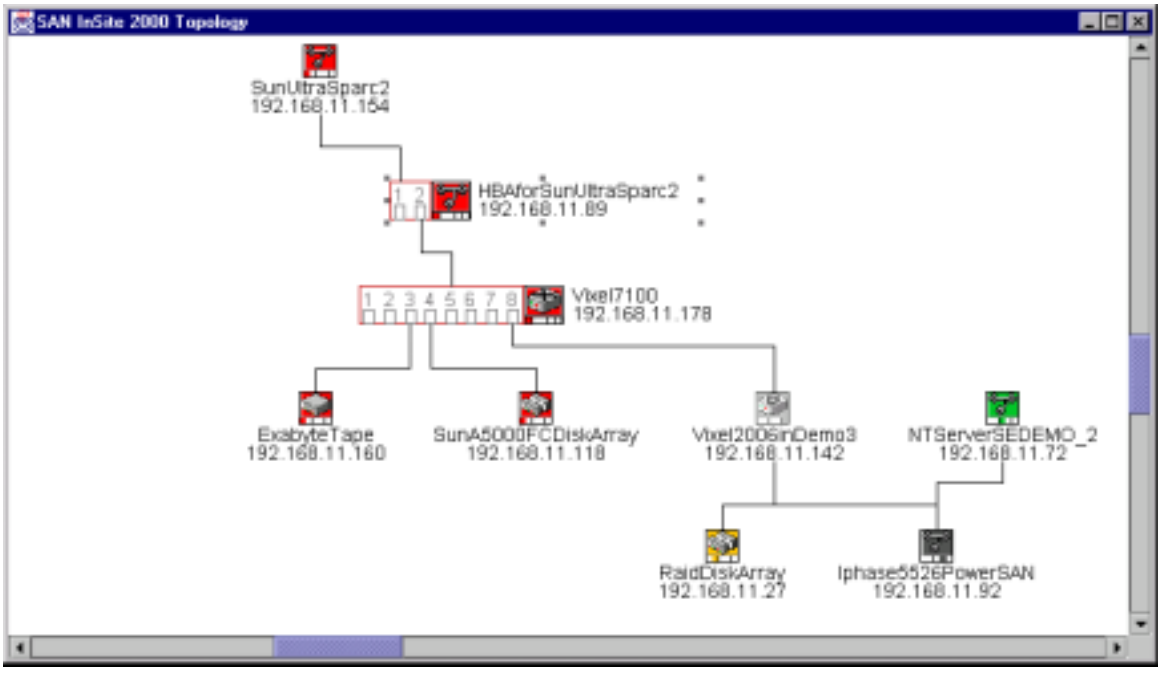

Figure 3 Vixel SAN InSite 2000's topology mapping of the fabric, including HBAs and disks

Summary

Enterprise SANs require fabric switches for large populations of servers and storage and for the advanced features fabrics offer. Depending on application requirements, extended fabrics can be constructed with a combination of fabric switches and downstream loop hubs. This enables the most efficient use of switched and shared gigabit segments. The Vixel 2100 Zoned Managed Hub and Vixel 7000 series switches are fully interoperable for these enterprise requirements.

Fabric login, Simple Name Server, and RSCN are the most common fabric services and simplify address management, device discovery and notification of changes in the topology. Fabric Address Notification facilitates applications such as streaming tape backup, while the Broadcast Server is required for IP address resolution over Fibre Channel. To minimize manual administration of the SAN, the Principle Switch and Domain Address Manager allow a large fabric to be self-defining.

Vixel's support of advanced fabric services is complemented by support for legacy, private loop-only devices. Public to private communication preserves the customer's investment in expensive disk resources and makes them available to fabric-enabled hosts. Stealth-3's auto configuration and Expert Mode provide flexible options for accommodating both private loop hosts and private loop targets.

To manage both fabric and fabric / private environments, SAN InSite 2000 completes Vixel's SAN solution. As the only interconnect management platform to manage multi-vendor configurations, SAN InSite 2000 can help the administrator determine where fabric functionality can be most productively applied. This total SAN solution lowers the customer's total cost of ownership and streamlines installation of both fabric switches and managed hubs for enterprise applications.

About the author:

Tom Clark is Technical Marketing Director for Vixel Corporation and the author of *Designing Storage Area Networks: A Practical Reference for Implementing Fibre Channel SANs*, Addison Wesley Longman

For more information on the above subject, contact your reseller:

Unylogix Technologies Inc.
Tel.: (514) 253-5200
email: info@unylogix.com
web site: www.unylogix.com