White Paper

What is Loop Switching?

Tom Clark
Director, Technical Marketing

January, 2000

VIXEL
MAKING THE FIBRE CHANNEL CONNECTION

Introduction

Storage area networks based on Fibre Channel technology have been driving business and information applications for several years. Initially deployed as specific solutions for departmental or business-line needs, SANs have existed as high-performance islands within a sea of slower, traditional SCSI storage configurations. The overwhelming majority of these SAN islands are based on arbitrated loop, a shared 100MBps topology that is suitable for a wide range of application requirements.

Solution providers such as Sun, Compaq and IBM have been shipping tens of thousands of loop-based configurations since 1998. These generally employ servers, Fibre Channel host bus adapters (HBAs), loop hubs, and Fibre Channel storage arrays. Since the number of devices in a given solution is fairly small (usually less than 10), stand-alone arbitrated loop has provided acceptable performance while enabling the flexibility of networking between servers and storage.

For some applications, however, a shared 100MBps transport does not offer sufficient bandwidth for more than a few devices. Streaming video, for example, requires 30 MBps sustained throughput for each video stream. On a single arbitrated loop, only 2-3 workstations could perform concurrent editing before saturating the shared transport.

The obvious solution would appear to be simply attaching servers and storage to a fabric switch. A switch provides a dedicated 100MBps per port, and consequently could sustain multiple concurrent video streams. Several factors may make it difficult to do this. One is cost. A fabric switch is 3 to 4 times as expensive per port than a managed loop hub, and significantly more expensive than an unmanaged hub. So while a fabric switch is the preferred solution in terms of bandwidth, some applications are too cost-sensitive to justify the higher cost per port.

The second, perhaps more persuasive factor working against a fabric switch is the fact that a large installed base of arbitrated loop devices (primarily expensive disk drives) do not support fabric services. In order for a loop device to communicate with a fabric, it must support fabric login, Simple Name Server registration, and so on to make itself known to the fabric switch. This requires sophisticated code on the part of Fibre Channel HBAs and disks which earlier generation products did not support. While it is ironic that a new, high-speed transport such as Fibre Channel should already have a “legacy” base, the previous generation loop products represent a substantial investment. Few customers can afford to forklift upgrade their “old” Fibre Channel devices for new fabric-capable ones.

To address the requirement of switched bandwidth with private arbitrated loop device drivers, Vixel pioneered a loop switching architecture called “Stealth” mode switching. Currently in its third generation of development, Stealth-3 has over two years of field-proven experience linking private loop devices via 100MBps switching for video and other high-bandwidth applications on Vixel’s 4000 series and 8100 series switches. In the discussion below, we will examine the benefits Stealth-3 loop switching brings to arbitrated loop environments as it is implemented in Vixel 7100 and 7200 series switches.

Standard Loops

Arbitrated loop is based on a number of standards that define high speed transport across a shared gigabit media. Analogous to shared Ethernet or token ring, arbitrated loop has a media access method that insures orderly participation by up to 126 nodes on a single transport. Since an arbitrated loop allows only one conversation to occur at a time, all active participants must follow a common loop protocol for gaining access to and then yielding control of the media.

A SCSI transaction launched by a server, for example, would pass down to the server's Fibre Channel host bus adapter and cause the HBA to arbitrate for access to the loop. Upon winning arbitration, the server would open its target (e.g. a disk array) and send or receive data. When the transaction (or this phase of a large transaction) is complete, the conversation is closed and the loop is made available to other participants. Since the transport itself is significantly faster than the ability of storage devices to buffer large amounts of data, a read or write or a large file could require multiple accesses (tenancies) of the loop.

The bandwidth available to an individual device is determined by the number of active contenders on the same loop. Four equally active servers on a loop, for example, would each have 25MBps (100MBps/4) bandwidth. Twenty-five equally active servers would only have 4MBps bandwidth each. Whether this is desirable or not is completely application-dependent. Some SAN applications such as server clustering and storage consolidation may not require high bandwidth, but benefit from Fibre Channel's flexible networking architecture. A standard arbitrated loop may therefore be the best solution if the number of devices balances adequately against the application's average bandwidth requirements.

In addition to bandwidth considerations, the overall distances required may effect loop performance. As longer links are added to the loop (e.g. a 2km optical link between two buildings), the total circumference of the loop expands. This incurs latency and adversely effects overall performance. Depending on application requirements, large loops may not provide adequate response time for data-intensive transactions. A loop with a large circumference may be acceptable for disaster recovery implementations that, for example, only update database records but may be unacceptable for those that require full disk mirroring. Typically, applications that require higher performance over long distances will use fabric switch ports to isolate the long runs.

The vast majority of arbitrated loops in production are private, i.e. the participants only communicate amongst themselves. To go beyond the local loop would require attachment to a fabric switch, as well as device drivers that would allow the participating HBAs and disks to log on to the fabric. A server, for example, could communicate with disks on the local loop segment as before, but having registered with the fabric switch could also communicate with fabric-attached disk or tape devices. This public loop mode combines the benefits of loop with fabric switching and allows a large number of devices to be attached to a single fabric. Unfortunately, the vast majority of SANs deployed over the past few years are private-only and cannot be easily upgraded to fabric support.

Loop Hubs

Normally, loop hubs are passive components of an arbitrated loop. As shown in Figure 1, a loop hub is simply a wiring concentrator that brings the ring configuration of the loop into a more efficient star topology. Aside from port bypass functions, a loop hub's role is to pass the data stream from one port to another. Any advanced management features (e.g. protocol decode in the Vixel 2100 Zoning Managed Hub) must be non-intrusive to the gigabit flow of data from one node to its downstream neighbor. Since the hub is not an active participant in the loop, attached devices do not need any special protocols or procedures to deal with the hub's presence. Unlike a fabric switch, the hub requires no login function, flow control or other services that would require additional logic on the part of its attached HBAs or disks.

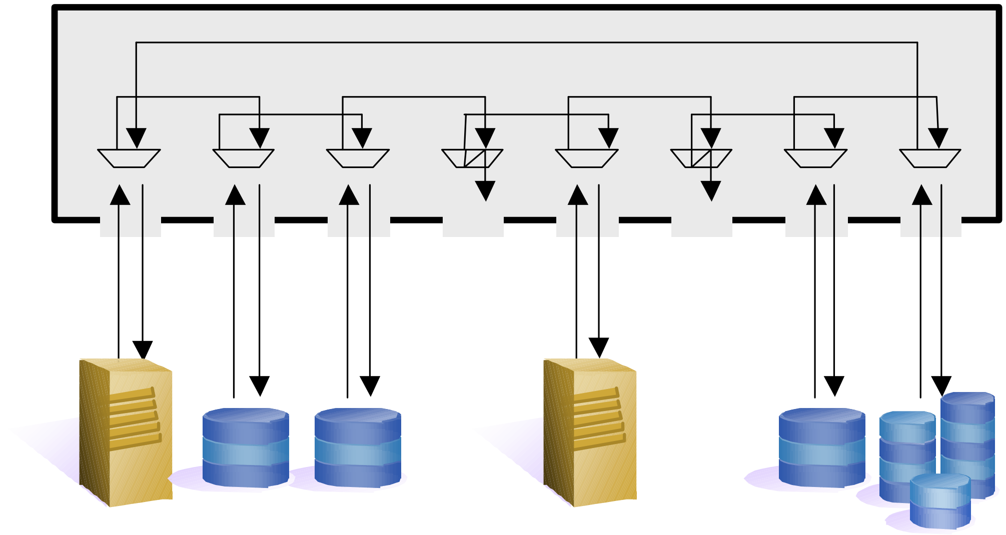

Figure 1 Loop Hub Architecture

This passive aspect of loop hubs provides significant flexibility in designing SANs. Hubs can be attached to fabric switches to support higher populations of public loop devices without consuming switch ports. Hubs can also be used for stand-alone private loops for departmental or application-specific configurations.

The large installed base of loop hubs and attached private loop devices is servicing a wide variety of applications and represents a substantial investment by enterprise networks in SAN technology. Allowing these configurations to benefit from the higher performance of switching requires a hybrid product that can leverage the power of switching while accommodating loop devices that cannot communicate with fabrics. Loop switching thus stands intermediately between loop hubs and fabric switches, and, depending on the vendor's implementation, may provide an upgrade path to full fabric when the customer requires.

Loop Switching

Designing a loop switch is not a trivial task. To function properly, the loop switch must, like a hub, be completely transparent to the attached devices. At the same time, the loop switch must necessarily intervene in loop activity in order to provide switched bandwidth between communicating pairs of devices. Nodes should be able to arbitrate, open their targets, transfer frames and close the transaction as usual, unaware that switch logic is intercepting each transaction. And the loop switch must surreptitiously perform this interception without communicating directly to the nodes or incurring delays.

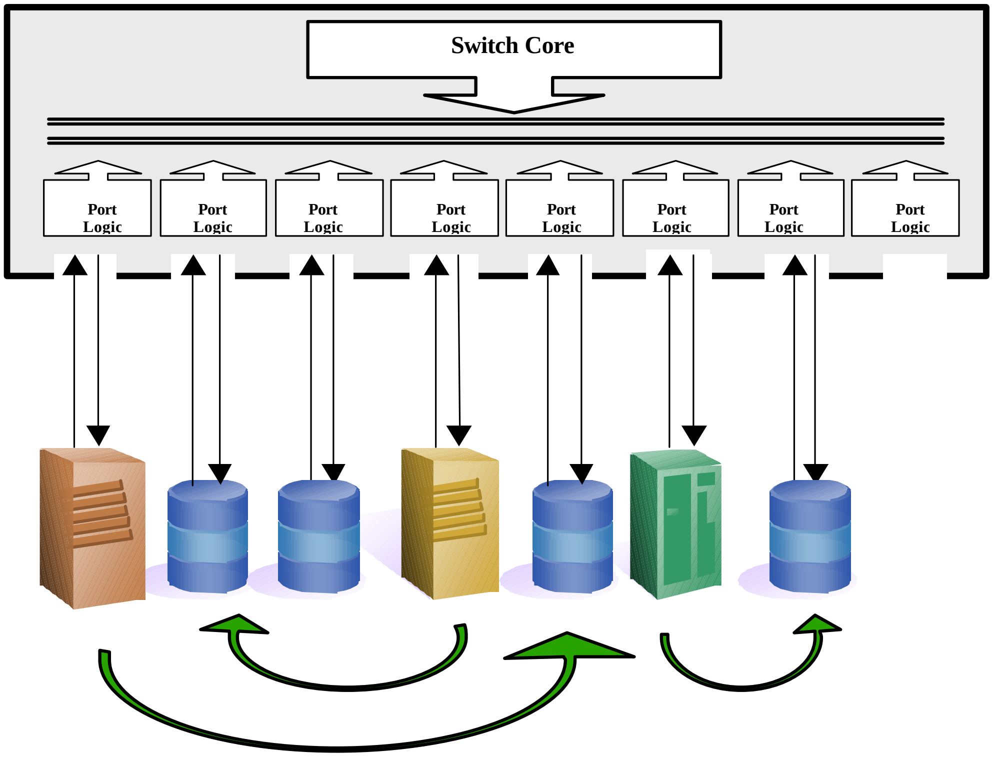

Figure 2 Loop Switch Architecture

On a normal loop hub, a device must arbitrate with other contenders for bandwidth. A server arbitrating for access on a hub port must wait until it has won arbitration, i.e. its ARB must pass through all nodes around the entire loop. In addition, since the loop is shared, only one initiator can possess the loop at a given time. On a loop switch, each port provides a full 100MBps bandwidth. Consequently, although at the protocol level a device must still send out its arbitrate primitive, if it is the only device on a loop switch port it will immediately win arbitration and be able to open its target. This not only accelerates transaction time, but also allows multiple conversations between different initiators and targets to occur simultaneously. As shown in Figure 2, the ability to support multiple, concurrent transactions enables loop devices to enjoy unprecedented performance that would otherwise require fabric support. Vixel's 16 port 7200 loop switch, for example, can sustain 8 concurrent loop transactions for an aggregate bandwidth of 800 MBps.

How does the loop switch intervene in the arbitration process? If a single device is attached to the loop switch port and arbitrates for access, the loop switch will immediately return the device's ARB to indicate that it was won. As far as the device is concerned, the ARB may have passed through multiple nodes before it returned victorious. In reality, the ARB goes no further than the switch port itself and is returned to the node. Two servers arbitrating on two different ports would thus both win arbitration simultaneously. A server would then send out an Open primitive with the address of its intended target. The loop switch must intercept this Open command, identify the target address, and create a switched circuit between the port of the initiator and the port of the target. This establishes a temporary but dedicated 100MBps pipe through the switch matrix between the two ports for frame transfer.

If an initiating node on loop switch port 1 opens a target node on loop switch port 3, a switched circuit will be established between ports 1 and 3. It may happen, however, that an initiator on port 5 also wishes to communicate to the target on port 3. Since the device on port 5 will immediately win arbitration and send out its Open to the target, it has no way of knowing that port 3 is busy communicating with port 1. To avoid aborting port 5's transaction, therefore, it is vital that the loop switch provide sufficient buffering to absorb frames until the device on port 3 is freed. The Vixel 8100 and 7x00 series switches, for example, provide 32 full frame buffers per port, sufficient to avoid frame loss during periods of congestion or concurrent access of target ports.

In addition to enabling multiple, concurrent transactions for private loop devices, Stealth mode loop switching also facilitates long links on a single loop. Since normal loop protocols do not have to pass around the entire loop, the loop can be extended without performance degradation. A 10km run on port 7, for example, would only incur latency for that port alone. Other ports are unaffected. This allows private loop devices to be used for disaster recovery and other long haul applications that would otherwise suffer a significant performance hit.

Loop Addressing

On a standard arbitrated loop, devices are dynamically assigned addresses through a loop initialization procedure. A series of address-selection frames issued during the loop initialization routine insure that every device has an opportunity to select an address and that no addresses are duplicated. The manufacturer of a host bus adapter may have a preferred loop address (e.g., hex '01'), but if that address is already taken by another device, the HBA will have to select something else. Generally, initiators such as host bus adapters tend to claim low addresses (which have higher priority), while targets such as disk arrays will tend to take high addresses. Since the loop hub is transparent to all loop events, it plays no role in the address selection process.

A loop switch may or may not intervene in loop addressing. Stealth-3's auto-configuration mode, for example, simply allows all devices to initialize and select addresses as if they were on a normal hub. Some devices, however, may be problematic during a normal initialization. Although by standards a device is supposed to take an alternative address if its preferred one is unavailable, some devices insist on having a specific address or range of addresses available. Stealth-3 incorporates an Expert Mode to accommodate these insolent nodes. With Expert Mode, the user can assign specific loop addresses to the appropriate ports and thus minimize any problems due to address selection.

Stealth-3 also allows the user to assign a higher initialization priority to a port. As a port is brought up, the devices attached to that port will attempt to take their preferred addresses. By manipulating which ports are allowed to initialize first, Stealth-3 allows favored (or potentially problematic) devices to assume their preferred addresses, while forcing other devices to take alternate address ranges. Potential address conflicts between disk arrays, for example, can be minimized by accommodating the more address-sensitive devices.

LIP Management

Control over LIP propagation is essential for maximizing performance in switched loop environments. Loop initialization is an essential part of life on an arbitrated loop and actually simplifies SAN administration. The loop initialization procedure automates the process of address assignment and insures that every device attached to the loop has a unique identity. This relieves the administrator of any manual supervision of addressing whenever new devices are added to a loop.

Since loop initialization primitives (LIPs) play a useful role in loop behavior, why limit LIP propagation? Although the loop initialization routine may only last a few hundreds of milliseconds, it is nonetheless a momentary disruption to ongoing data transactions. Most business applications tolerate LIPs quite well, recovering, if necessary, at the upper protocol level. But full motion video and some other data streaming applications do not tolerate any transient interruptions. Consequently, controlling which devices see LIPs can provide a means to optimize transactions and significantly reduce disruptions.

In standard arbitrated loop hubs, loop initialization primitives propagate around the entire loop when a new device is attached. Since an ordinary hub does not intervene in loop activity, it has no means to control LIP proliferation. A loop switch, however, is by nature intrusive in loop activity, and so is well-positioned to management the visibility of LIPs from port to port.

Allowing some ports to see LIPs while hiding them from others implies that a lack of re-initialization on some ports may result in duplicate addressing. This could be catastrophic if it was allowed to occur. For example, if two servers had the same loop address, data corruption could easily occur if a target disk sent data to the wrong host. Since Stealth-3 is controlling the allocation of addresses for the entire virtual loop, however, it can insure that even if a port is not LIP'ed, it will not inadvertently acquire an address twin.

Normally, a server would want to be notified if a new disk resource joined the loop. Servers talk to disks. In private arbitrated loop, the routine by which a server discovers a new target is launched by a loop initialization. LIPs let the server know that a topology change has occurred, and, immediately following the loop initialization routine, it can attempt to establish sessions with new disk resources, i.e. the new addresses that appeared following initialization. Stealth-3 gives the user the ability to control this process at the port level. A port can be LIP'ed if storage is added to any other port, or, alternately, screened from LIPs. A video server can thus be protected from disruptions, while other application servers are LIP'ed whenever storage is added to the loop.

Loop Switching and Fabrics

Deploying a loop switch for a specific departmental application offers several benefits. Legacy loop devices can be supported in a switched, high speed environment. Long distances can be accommodated for disaster recovery or campus configurations. And higher populations of loop devices can be supported without impacting overall performance. At some point, however, a customer will want to attach their loop switch to a fabric switch so that public-capable devices can communicate with a fabric. Stealth-3's FabricConnect feature is designed to provide this additional support.

Connecting a standard loop hub to a fabric switch is fairly straight forward. Since the hub is transparent to the loop, it is also transparent to the switch's Fabric Loop (FL) port. The FL port simply sees multiple loop devices which are attempting to login to the fabric. Connecting a loop switch to a fabric is not at all straight forward. Having fooled loop devices into behaving as if they are on a single shared loop segment, the loop switch logic must also make all public-capable loop devices appear to the fabric as a single segment. This requires additional manipulation of the initialization process and control over both private loop transactions and public loop-to-fabric communications.

Additionally, if the fabric switch supports private-to-public translation, the loop switch must make private loop devices visible to the fabric. This presents an additional engineering challenge, since although public loop devices will initiate logins to a fabric which the loop switch can simply pass on, private loop devices do not.

The ability to support loop switching plus fabric transactions is essential for migrating devices in the network and scaling to ever-changing network needs. FabricConnect allows Vixel 7x00 series switches to be used in a wide variety of applications, from stand-alone SAN islands to edge switches for loop devices in extended fabric SANs.

Fabric Upgrades

Although loop switches provide optimum throughput for both private and public loop devices, the customer may at some point need to attach non-loop, direct connect fabric devices. To accommodate this potential requirement, the Vixel 7x00 series loop switches support a firmware key upgrade to full fabric services. This would allow the attachment of both loop and non-loop devices to the switch, with full support for fabric login, Simple Name Server, Registered State Change Notification and other fabric services. Providing a means for the customer to pay for functionality only when it is required extends the usefulness of the 7100 and 7200 switches and maximizes the customer's return on investment.

Summary

Loop switching offers an intermediate step between standard arbitrated loop hubs and full fabric switches. For customers with an installed base of private loop-only devices, loop switching provides higher bandwidth per port and the ability to support long links for campus or disaster recovery applications. Vixel's Stealth-3 implementation for the 7100 and 7200 offers advanced loop switching features that insure loop stability and uptime. Address and LIP management functions allow the administrator to optimize a loop switch configuration to accommodate problematic nodes that would otherwise disrupt loop activity. FabricConnect enables the 7100 and 7200 switches to attach to fabric switches so that both switched loop and fabric transactions can occur. And the ability to upgrade 7100 and 7200 from loop switches to full fabric switches completes the migration path to a full fabric SAN. Combined with Vixel SAN InSite management software, the 7000 series loop switches and Stealth-3 gives customers the power to create highly managed, high performance storage networks.

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