Agenda

- What is a SAN?
- Fibre Channel Overview
- SAN Implementation

A SAN is

a high speed network that connects computer systems and storage elements and allows movement of data between computer systems and storage elements

But it takes more than a hardware interconnect..

A SAN must unconditionally consist of two layers: a storage plumbing layer that must carry device commands, and a value− added software layer that exploits the plumbing layer to bring value.

You should use this SAN definition to separate marketing hype from true added value.

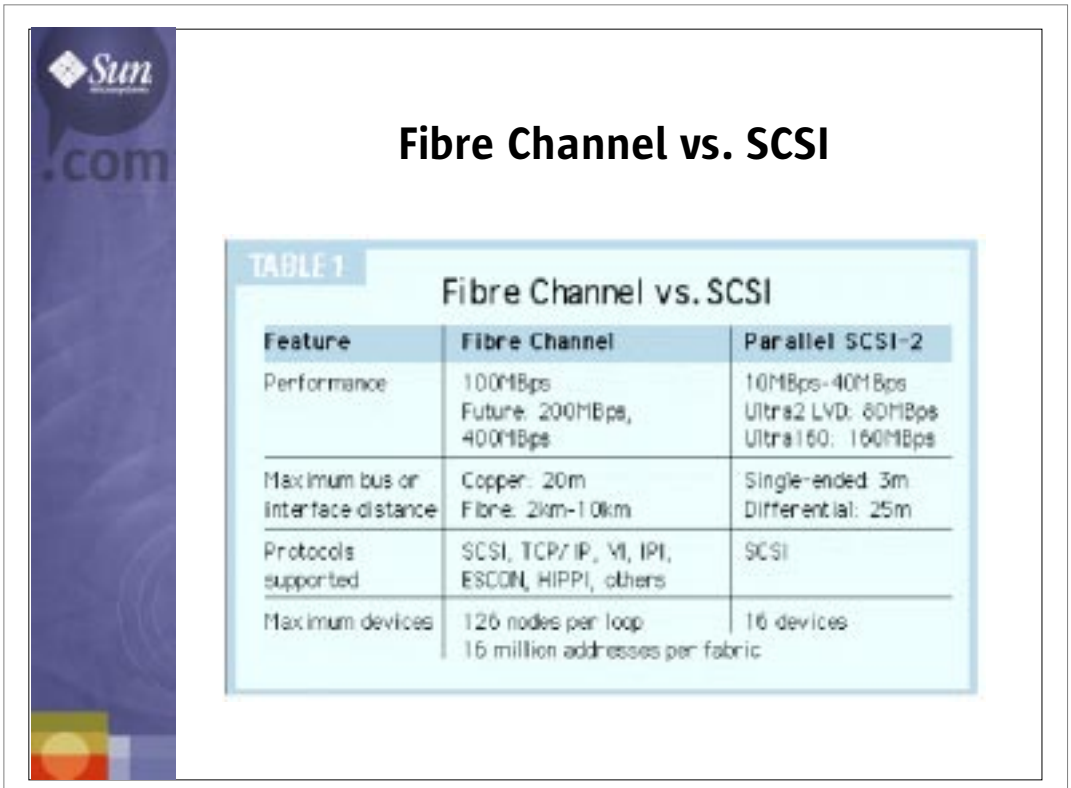
So what's happening today?

SAN’s are being installed but;

- No one owns the entire solution.
- They are not hetrogenious (yet).
- They are a long way from the "ideal" solution.
- Need new breed of hardware & software

Why SAN's are not there yet.

- To build a SAN, you need:
  - Hardware
    - HBAs
    - Device drivers
    - FC Storage
    - Switches (Ideal), Hubs
  - Software
    - SAN Infrastructure management software
    - Storage Virtualization/Pooling
- To build a fabric SAN, all these components must be fabric capable



Fibre Channel is high−speed serial protocol that runs over fiber optic or copper cables. Fibre channel benefits include: greater throughput (100MB/s, with future support up to 400MB/s), longer distances support than SCSI (up to 10km), easier connectivity options (hubs and switches), and multiple topology support (point−to−point, arbitrated loop, and fabric).

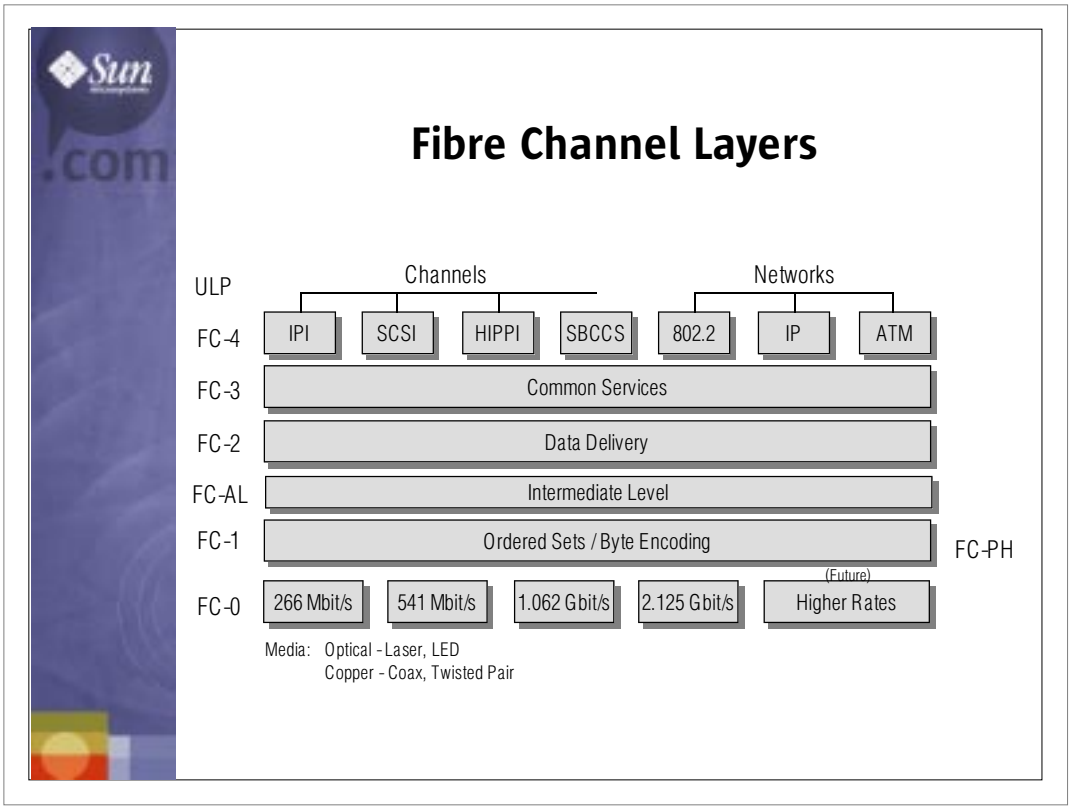
The table above highlights some of the key differences between SCSI and Fibre Channel.

Although SANs can be deployed with technologies such as SCSI and ESCON, the most common technology for SAN deployment is Fibre Channel. From this point forward, SANs built with FC will be referred to as FC/SANs.

Some of the reasons for FC's popularity in building SANs are:

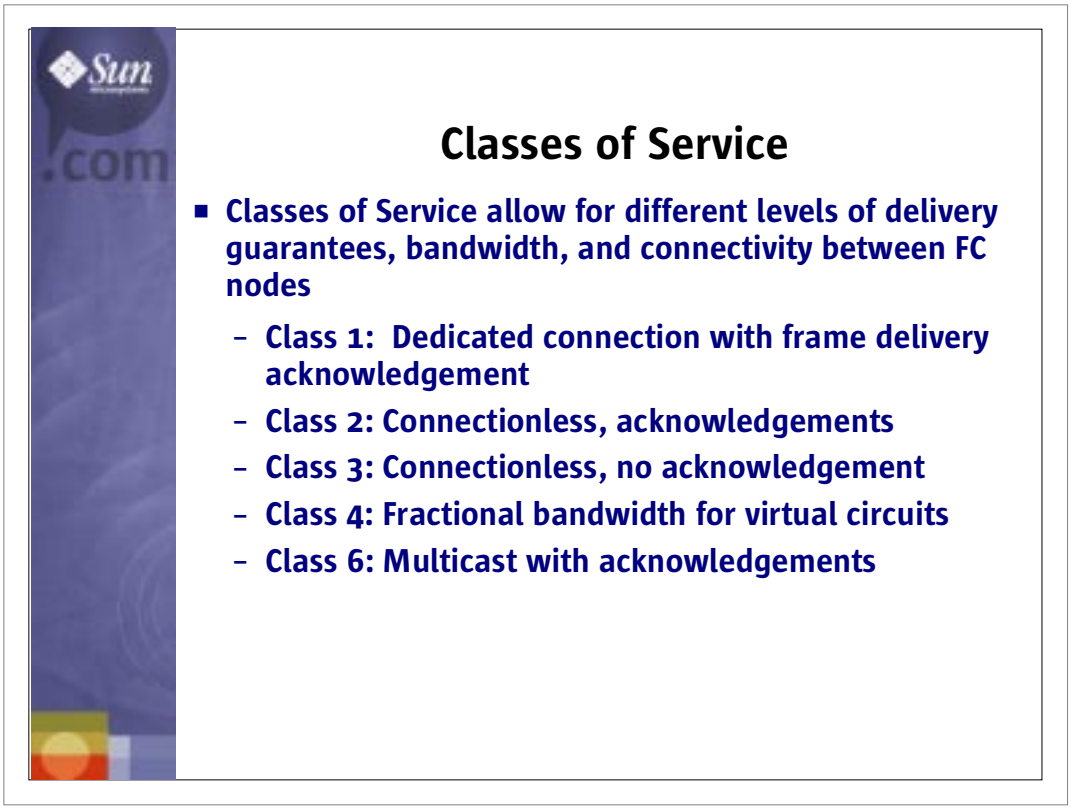
- Support for multiple topologies
- Supports upper layer protocols
- Performance

These reasons will be discussed in further details in the next few slides.



Fibre Channel Layers: The FC standard defines a multi−layered architecture for transporting data across a FC network. These cooperative layers define common networking principles such as access methods, addressing, packetizing and routing of data, and upper layer protocol support.

Fibre channel also supports several upper layer protocols. This concept is analogous to ethernet or token ring protocols used in the networking industry. Ethernet and token ring are lower level network protocols that support different upper layer protocols like TCP/IP and IPX. Similarly, fibre channel is a low level protocol that supports both networking protocols like TCP/IP and storage protocols such as SCSI and ESCON. This means that theoretically one wiring plant could support both networking and storage protocols.

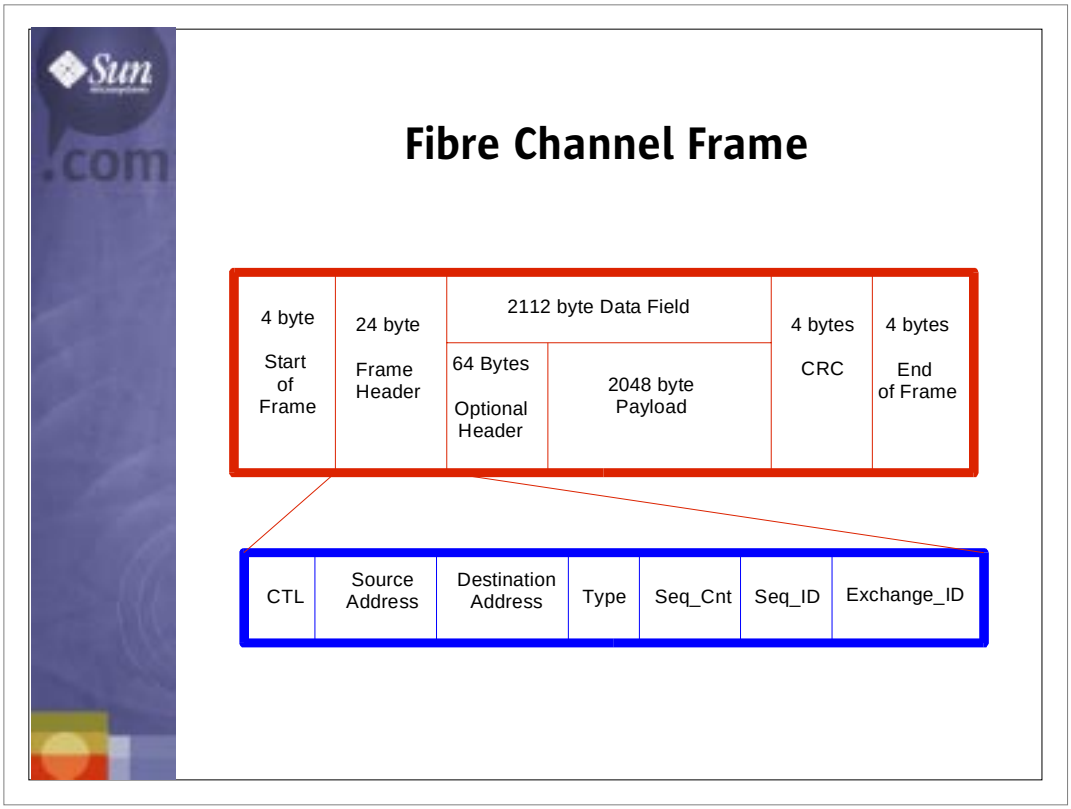


Fibre channel has several classes of service that allow for different levels of delivery guarantees, bandwidth, and connectivity between FC nodes. Although the FC specification defines multiple service levels, classes 2 and 3 are the most common levels used in FC products today. Below has some examples of Sun products and their respective classes of service that were used.

- Class1: Example− Ancor's 1st Generation FC switches
- Class2: Example − StorEdge SSA
- Class3: Example − StorEdge A5x00, A3500FC

Some customers are concerned with using Class 3 products in mission−critical environments instead of Class 2. Basically, Class 3 provides the benefits of better performance and less delivery overhead compared to the inherent reliability in Class 2 FC delivery. In well constructed FC environments (i.e.− Loops, Fabrics) Class 3 can be acceptable in mission−critical environments because in well constructed FC environments (i.e.− FC−AL, Fabrics) reliability will not be a huge issue. However in the cases that there is a frame delivery problem (i.e.− a dropped FC frame),the upper layer protocols (ie.− SCSI protocol) can resolve the problem.

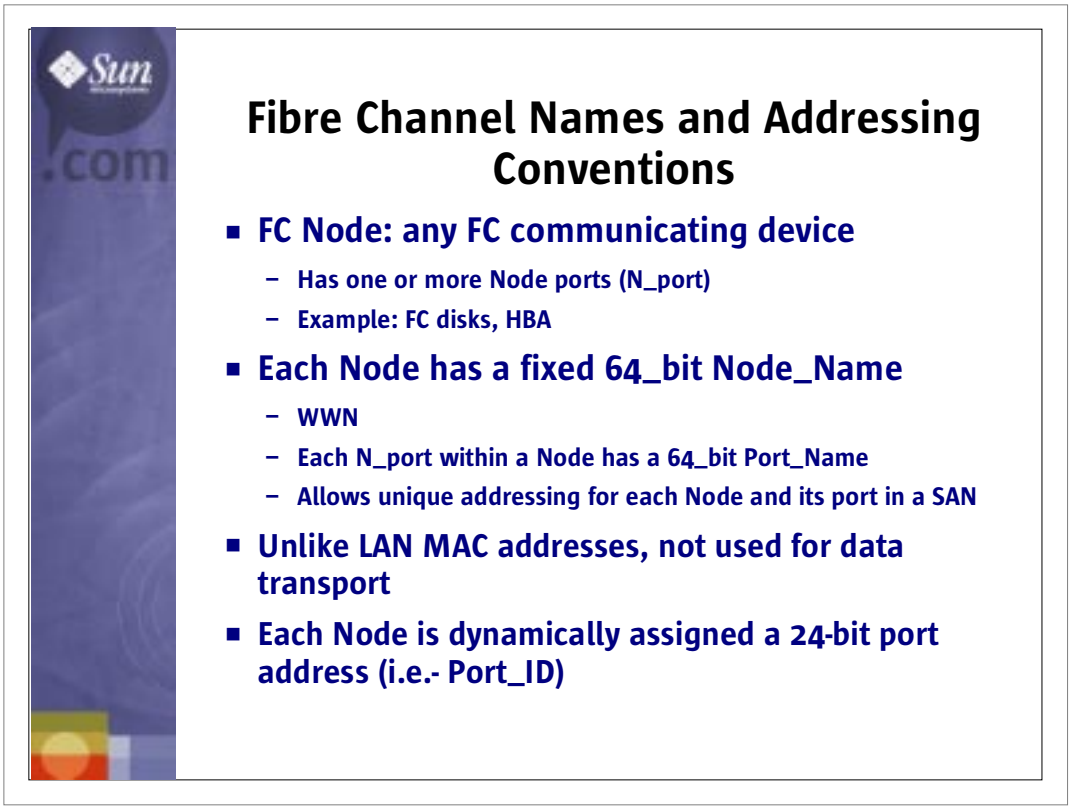
Note: There is a slide on the layers of FC available in the backup slides section



In order to move data, the upper layer protocols (i.e.− SCSI) organizes the data in packets called FC frames. This is a diagram of a FC frame. Each FC frame is 2148 bytes and has the following main components.

- Start of Frame delimiter: Marks the start of a FC frame as well as the class service. The SOF also denotes whether the frame is the start or part of a series of FC frames
- Frame Header: Discussed below
- Optional header
- Payload for data
- CRC: 32−bit Cyclic Redundancy Check
- End of Frame delimiter: Marks whether the FC frame isjust one of or the end of a FC series.

The frame header contains subarea information. Particularly, the subarea contains the 24−bit source (S_ID) and destination (D_ID) addresses. Where the 24−bit address comes from will be discussed on the next slide however note that the location within the FC frame and the size of S_ID and D_ID addresses allow the FC switch to make quick routing decisions without reading the entire 2148 byte frame.

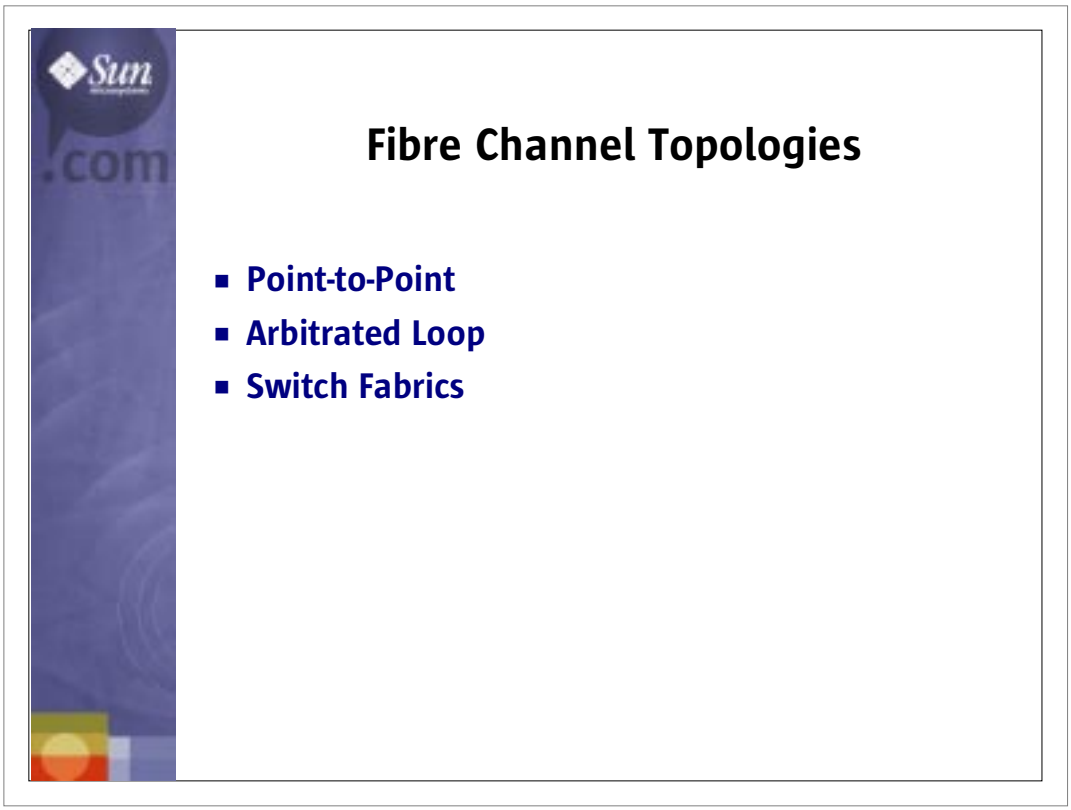


FC Node is any fibre channel communicating device. A FC node can be the initiator, responder, or both for any given communication device. Each FC node has three addresses assigned to it: a 64−bit unique node address called the world wide number (WWN), a 64−bit port address assigned to each port on the FC node (Port_Name), and a 24−bit dynamic address(N_port ID). The combination of the WWN and Port_Name address allows for unique addressing for FC nodes and their associated ports even in large complex SAN environments. Depending on the FC topology, the 24−bit address is dynamically assigned to the FC node during either an arbitrated loop initialization (LIP) or fabric login process(FLOGIN) to a switch.As mentioned earlier, this allows for quick switch routing and eliminates any requirements for address administration in a SAN. Examples of FC nodes include:

− Disks, FC Controllers, switches, FC HBAs, FC tape drives

Note: A simple hub or hub port is not an example of a FC node.

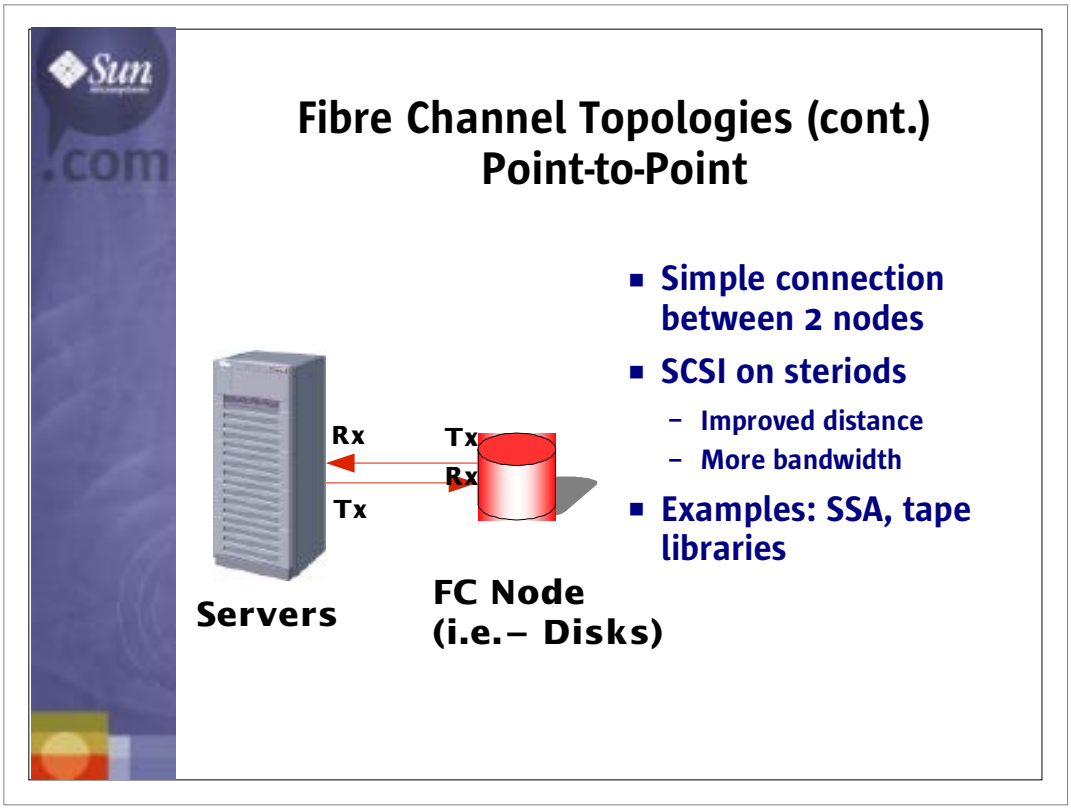
If the speaker is familiar with Sun's FC disks used in the A5x00 and E3500, they can further explain these different addresses. Note that the FC disks have a WWN assigned on drive(i.e.− this looks like a serial number). However, when using the "luxadm dump_map" command, the reader will noticed that the FC disks will have some C21xxxxx and C22xxxxx numbers. These represent the 64−bit port numbers for the A and B ports of the drive. Also displayed are the AL_PA numbers. AL_PA stands for Arbitrated loop physical address. These are the 24−bit dynamically assigned addresses. In reality, these are 8−bit addresses because the A5x00 is a private loop device. Private loop will be discussed further in the switch section



Fibre supports multiple topologies:

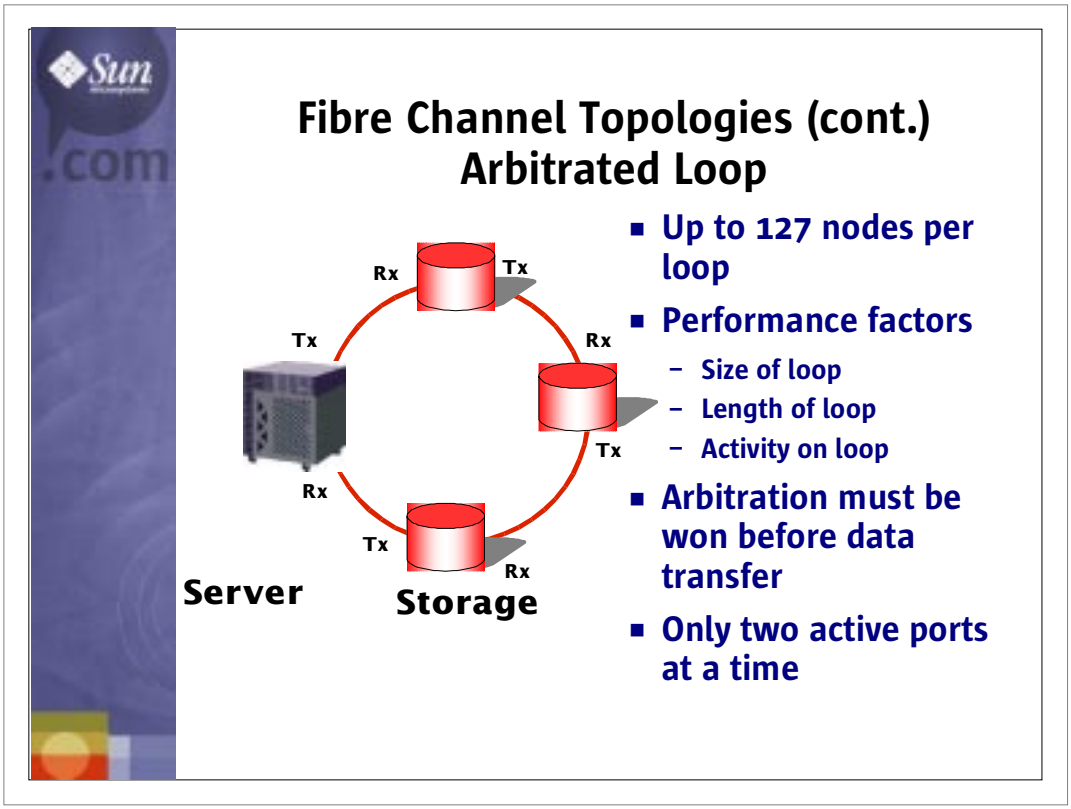
− Point−to−point
− Arbitrated Loop
− Switched Fabric

Let’s take a closer look at these topologies.



Point−to−Point is a FC topology with a simple connection between two FC nodes. A good example of this FC topology can be seen by looking at the configuration of our servers connected to the older generation StorEdge SSA storage device. In a p2p topology, the transmit lead of one FC node (i.e.− server) is connected to the receive lead of another FC node (i.e.− disk) and vice versa. Note that even P2P configurations have a procedure to dynamically assign the 24−bit address (Port_ID) that was discussed earlier.

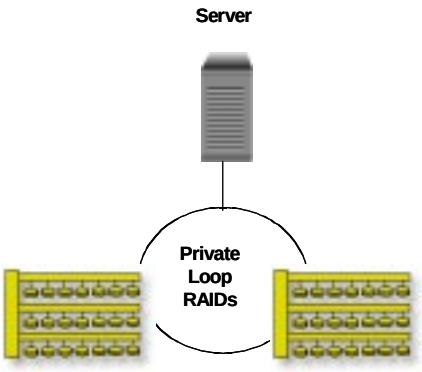
Compared to SCSI, the main benefits of a point−to−point FC topology are faster bandwidth and support for longer connectivity distances.



Fibre Channel Arbitrated Loop (FC−AL) is a shared 100MB/s topology that supports up to 126 FC nodes and 1 fabric device. FC−AL is very similar to network topologies like Token Ring and FDDI. Before sending data on the loop, a FC node must win arbitration to access the loop. Once the FC node (initiator) wins arbitration, the node can communicate with the target FC node via a point−to−point connection. To ensure that each FC node has the opportunity to communicate, FC−AL uses a fairness algorithm. The fairness algorithm allows all FC nodes the chance to communicate before a node can send data again. There is overhead due to the combination of the arbitration process and the fairness algorithm. Other factors that effect the performance of FC−AL are:

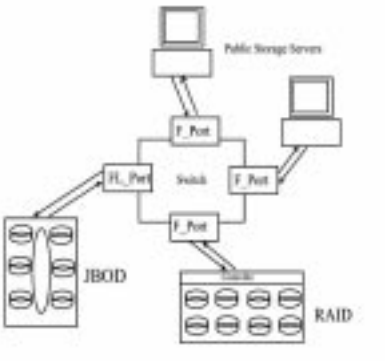
- Number of devices per loop
- Length of the FC loop
- Activity on the FC loop (i.e.− adding/removing FC nodes)

Private Loop

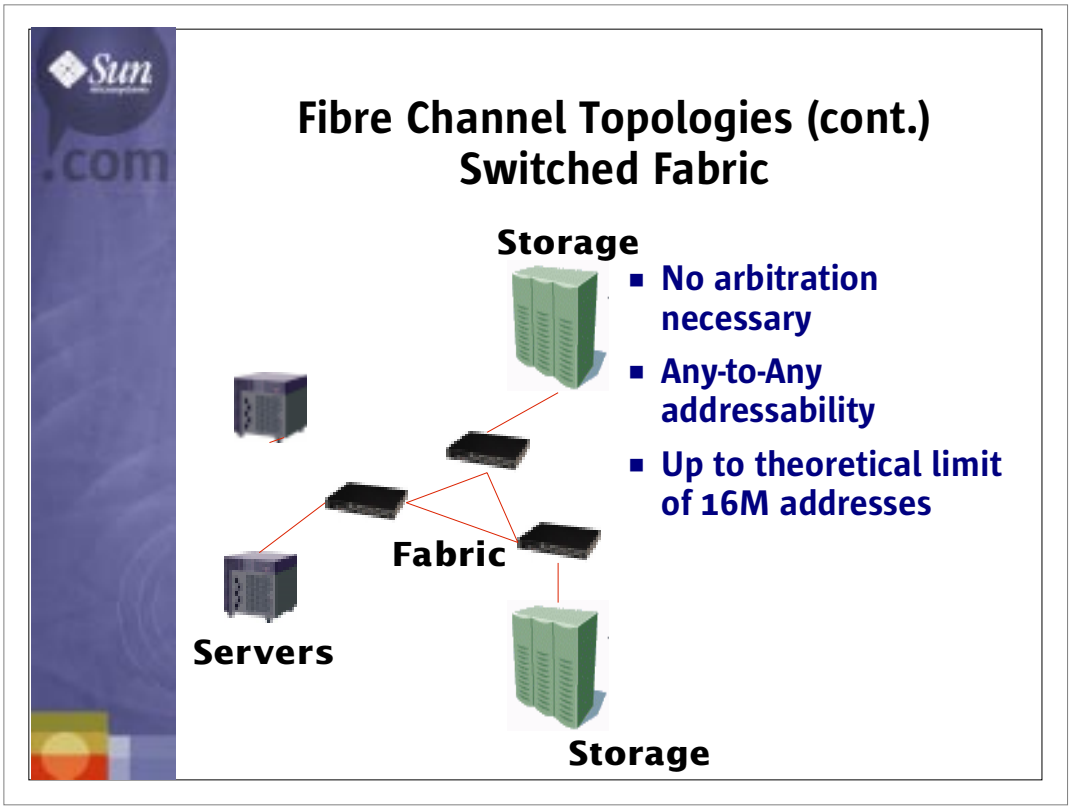


- Limited fibre channel addressing (126 devices)
  - 8-Bit addressing
  - 8B/10B encoding algorithm
- Does not support fabric login
- Generally used for small SAN configurations

Public Loop



- Full fibre channel address capability (i.e.- 24-bits)
- 16M nodes possible
- Can communicate with any other public device on the fabric
- Supports fabric login devices



Fabric is a FC network topology that uses one or more FC switches to interconnect multiple servers and storage devices. Fabric networks provide full 100MB/s bandwidth per switch port. Additionally, since switched fabrics use a 24−bit address, the fabric can theoretically have up to 16 million addresses (i.e.− 2^{24}). In order for switched fabrics to exists, the SAN needs these components:

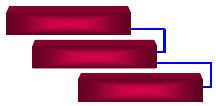
- Fabric capable storage
- Fabric capable HBAs and device drivers
- FC switch

Also, switched fabrics don't have arbitration overhead that is used in FC−AL. Switched fabrics allow any−to−any connections with routing logic embedded in the FC switch. Most FC switches use a cut−through, non−blocking switch mechanism. Cut−through means that the switch provides routing logic by examining the port's destination address located in the FC frame header. This allows the switch to make decision by reading only the first few bytes of the FC frame rather than the entire 2148 byte frame. Non−blocking means that the switch doesn't block or suspend other port−to−port conversations while some other conversation is occurring within the switch.

Switch Topologies

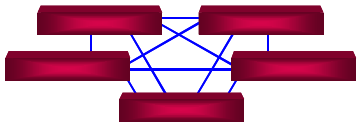
Cascade

- Low cost expansion for small installations
- Switches "daisy-chained" one after another

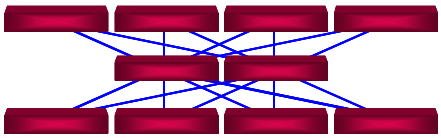


Mesh

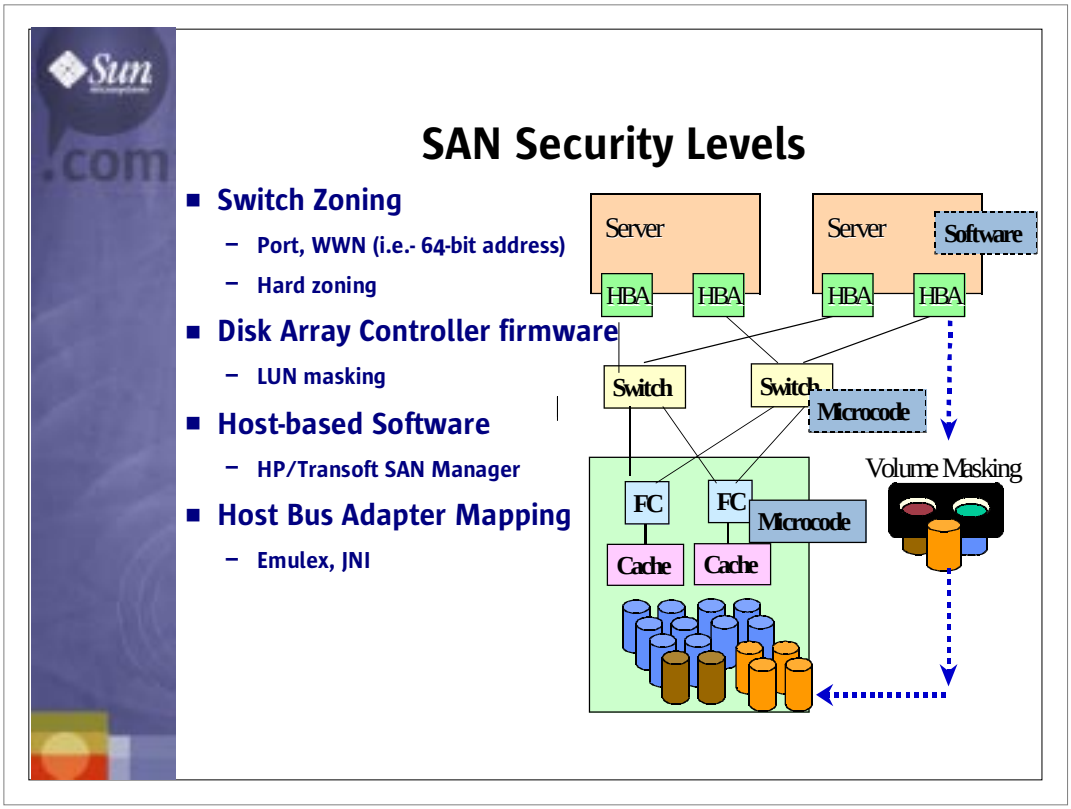
- Higher performance and more resilience
- Each switch connects to every other switch



Multistage



- Highest performance and resilience
- Ideal for larger installations and/or tape drives requiring data streams



Security is one of the primary issues with deploying SANs today. By default, SANs allow any server connected to the SAN, to see any other storage subsystem or tape library. There is nothing inherent in FC to act as a "traffic cop" to prevent one server from writing on another server's disks. There are several methods to manage data security within SAN environments and these methods are discussed in the slide above. These levels of security are called LUN masking. FC vendors have various marketing names for LUN masking

LUN Masking is a security mechanism within SAN environments that grants or blocks access to a specific LUN. The types of LUN masking and brief descriptions are listed below:

Controller−based Masking: LUN isolation on an FC−AL or fabric connection via the disk subsystems controller. Vendors such as Compaq, EMC, and MTI used this method of data security. Since each FC HBA has an associated WWN, a disk controller can use this WWN to grant or deny access to specific disks. Controller−based LUN masking provides the highest degree of granularity compared to the other methods.

Host−based Masking: LUN isolation using host−based software installed on the server. In order to be effective, the software needs to be running on all SAN−attached servers and clients. Example of this approach are HP's Transoft SAN Mgr software.

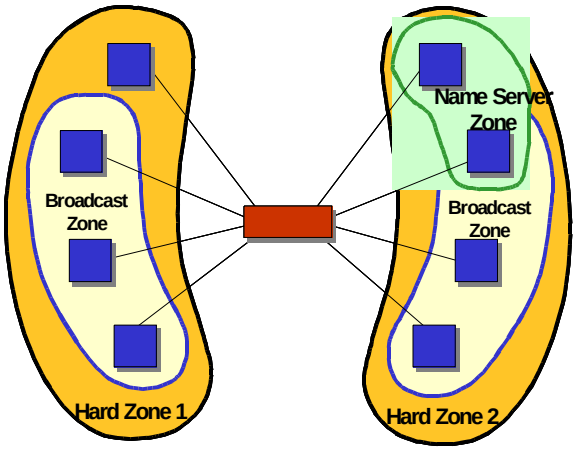
SAN or Switched based masking: Granting or denying access to a volume by masking at the switch component level. The switch cracks the FC packets to figure out whether a HBA can have access to LUN. This procedure requires VERY powerful and fast switches and therefor. SAN−based masking is not as popular as some of the other LUN masking methods available on the market. This approach maps at the device (switch and subsystem) level and therefore does not provide the same level of granularity compared to controller−based LUN masking.

Host−Bus Adapter is LUN security management via HBA device drivers. This approach is not ideal for large heterogeneous SAN environments. HBA manufacturers such as Emulex provide this feature today.

Remember!

- **SANs must consists of two Tiers**
  - **Plumbing (i.e.- hardware connectivity)**
  - **Software (i.e.-remote mirrors, zoning, disk pooling/virtualization**

SANMaster Zoning

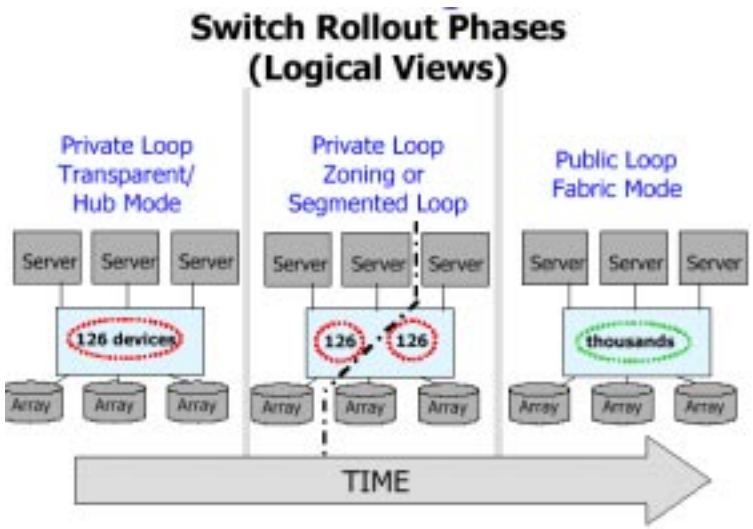


Although Hard Zones can be divided using Name Server and broadcast zones, there is **zero** possibility of communication across Hard Zone boundaries.

StorEdge SANMaster Switch Management

- Configuration/Zoning Management
- Port management
- Chassis Management
- Performance Management
- Topology Management
- Diagnostics
- Planned Integration under Component Manager

SAN Timeline



Further references for those that want to know more about FC and SANs.

The first 2 listings are highly recommended reading and can be found at any good bookstore or via Amazon and FatBrain.

The FC Consultant book series are more in−depth books about FC specifications.

SNIA and the FC Website are good websites to follow the developments of NAS, SANs and FC technology.