LAN, SAN, MAN, WAN: Making an Intelligent Choice for your Storage

Silvano Gai

Cisco/Andiamo Fellow

© 2002, Cisco Systems, Inc. All rights reserved.

Agenda

- Storage architectures
  - DAS: Direct Attached Storage
  - NAS: Network Attached Storage
  - SAN: Storage Area Network
- Network Architectures
  - Ethernet
  - FC
  - Sonet
  - DWDM
- Networked Storage
  - Comparison
  - Congestion Control
  - Scaling issues

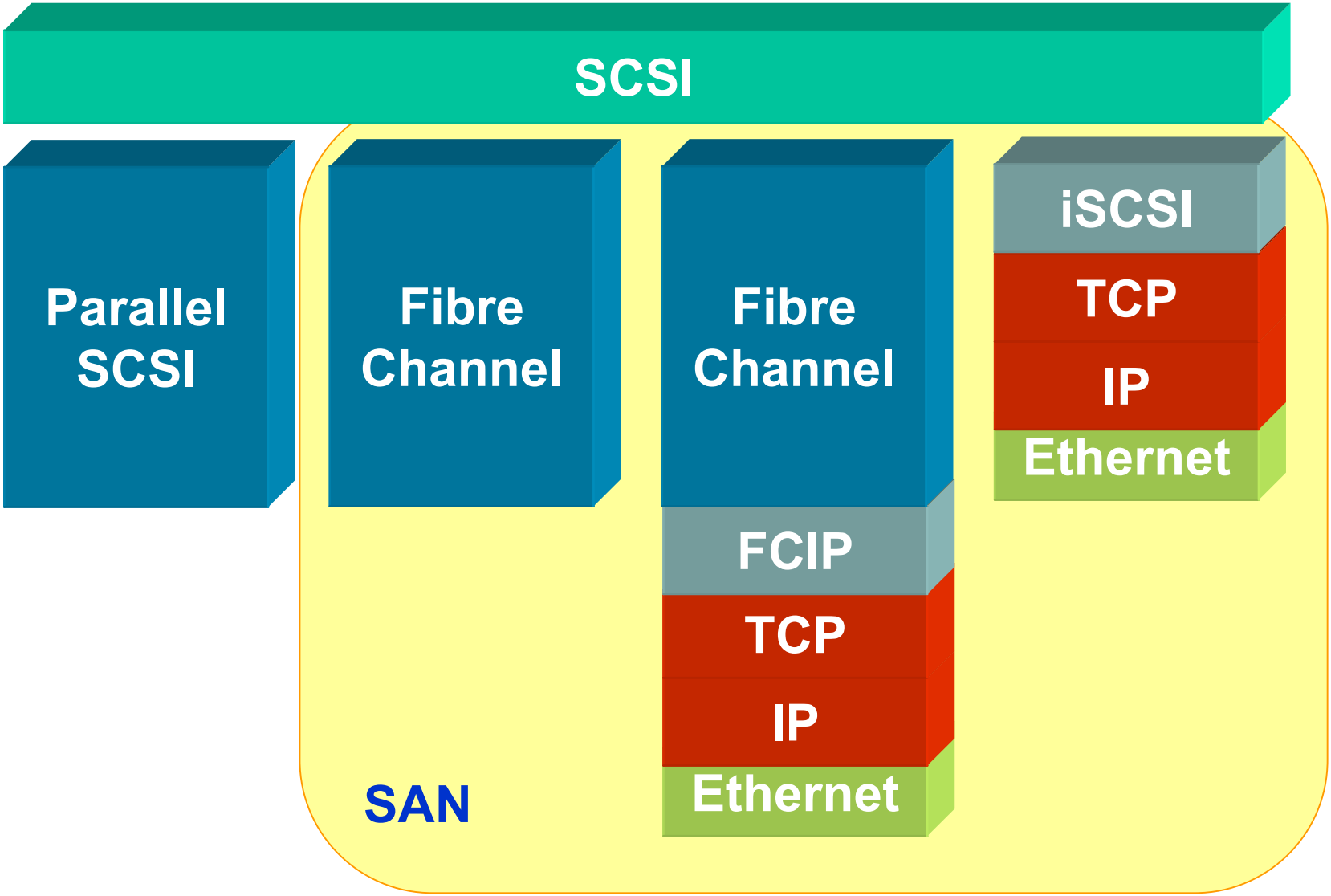
© 2002, Cisco Systems, Inc. All rights reserved.

Networked Storage

- **NAS (Network Attached Storage)**
  - Storage accessible at the file system level through:
    - NFS
    - CIFS/SMB
  - IP/Ethernet network
  - Main application: Engineering
- **SAN (Storage Area Network)**
  - Storage accessible at the block level through SCSI
  - Fibre Channel or IP/Ethernet networks
  - Main Application: Database
  - The topic of this tutorial

© 2002, Cisco Systems, Inc. All rights reserved.

SCSI History



© 2002, Cisco Systems, Inc. All rights reserved.

Storage issues

- SCSI has a lot of baggage from the past
  - It assumes the old bus based architecture
  - It is not efficient in recovering from packet loss

    Not an issue in bus architecture
  - Drivers are still based on old SCSI standards and they have been retrofitted with the “network”
- Applications are designed to cope with the above
  - Pipeline is hardly used
- Applications need to commit to stable storage
  - When you send Status(OK) you own the data and you cannot lose it

© 2002, Cisco Systems, Inc. All rights reserved.

Storage Latency

- Storage response time is:
  - Few milliseconds for disks
  - Sub-millisecond for caches
- Latency budget for SAN should be less than storage response time
  - Speed of light on Fiber is 200 Km/ms

© 2002, Cisco Systems, Inc. All rights reserved.

3 possible technology + 1

- There are 3 possible technology for SAN:
  - Ethernet
  - FC
  - Sonet
- Plus one
  - DWDM

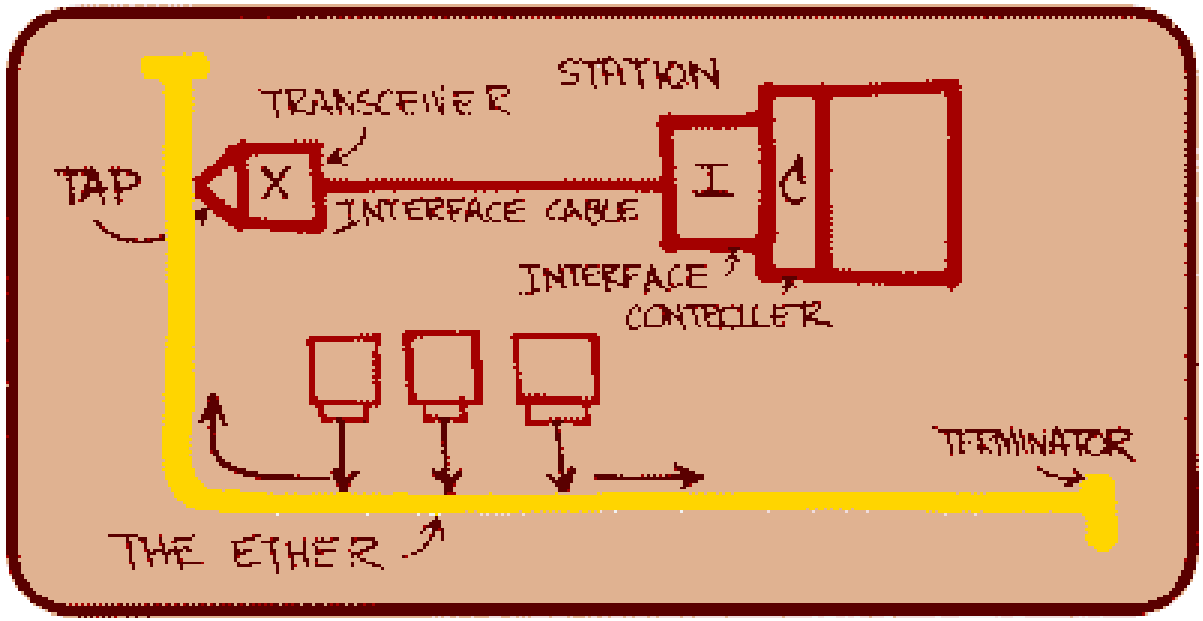
© 2002, Cisco Systems, Inc. All rights reserved.

An Historical perspective

- Metcalfe in 1976 presents Ethernet to the National Computer Conference
  - 1980 Digital, Intel and Xerox had released a de facto standard for a 10 Mbps
  - in 1991 10Mbps on UTP
  - In 1995 100Mbps
  - In 1998-1999 1Gps
  - In 2002 10Gb/s Ethernet
- Fibre Channel initial development in 1988
  - In 1994, the first Fibre Channel standard was approved (FC-PH)
  - In 1995 1 Gb/s based products are deployed
  - In 2003 10Gb/s Fibre Channel
- Sonet is developed in 1985 by Bellcore
  - In 1988 first ITU standard (G.707)
  - In 2000 10 Gb/s OC-192

© 2002, Cisco Systems, Inc. All rights reserved.

Ethernet: the origin



"The diagram ... was drawn by Dr. Robert M. Metcalfe in 1976 to present Ethernet ... to the National Computer Conference in June of that year.

© 2002, Cisco Systems, Inc. All rights reserved.

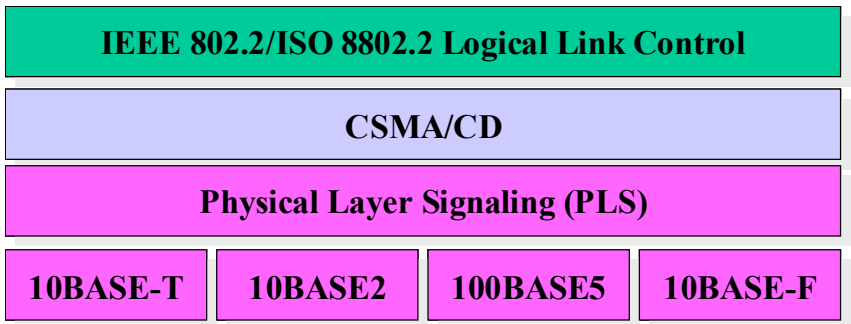
Ethernet: characteristics

- Simple
- Single MAC design
- Broad range of speeds
  - From 10 Mbps to 10 Gbps
- High volumes/Low costs
- Only survivor
- No guaranteed delivery
  - +/- of loosing frames

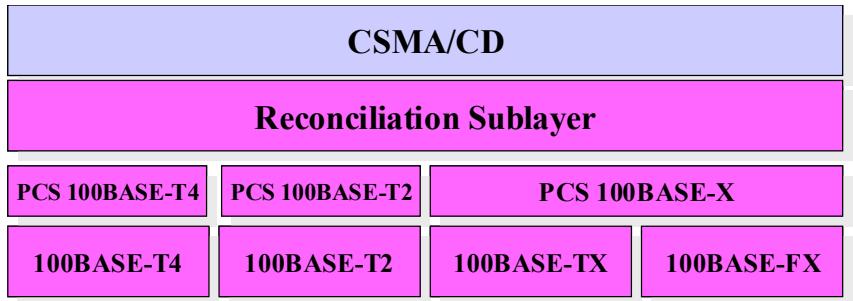
© 2002, Cisco Systems, Inc. All rights reserved.

Ethernet: the standards

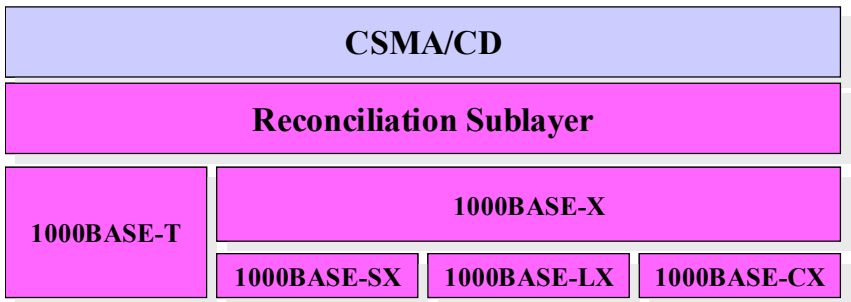
10Mbps



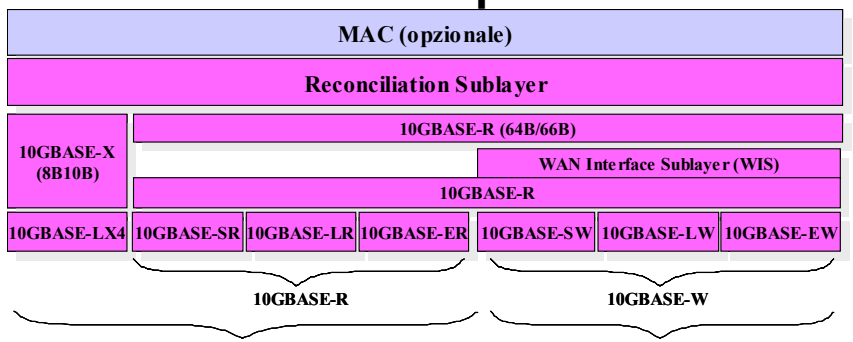
100Mbps



1 Gbps



10 Gbps



© 2002, Cisco Systems, Inc. All rights reserved.

Ethernet: the evolution

- Ethernet kills all other LANs
  - Token Ring, FDDI, etc. (except 802.11)
- IP kills all other network architectures
  - IPX, NetBeui, Decnet, AppleTalk
- Ethernet and IP get married ☺
- Everything over IP implies
  … everything over Ethernet

© 2002, Cisco Systems, Inc. All rights reserved.

Fibre Channel: the origin

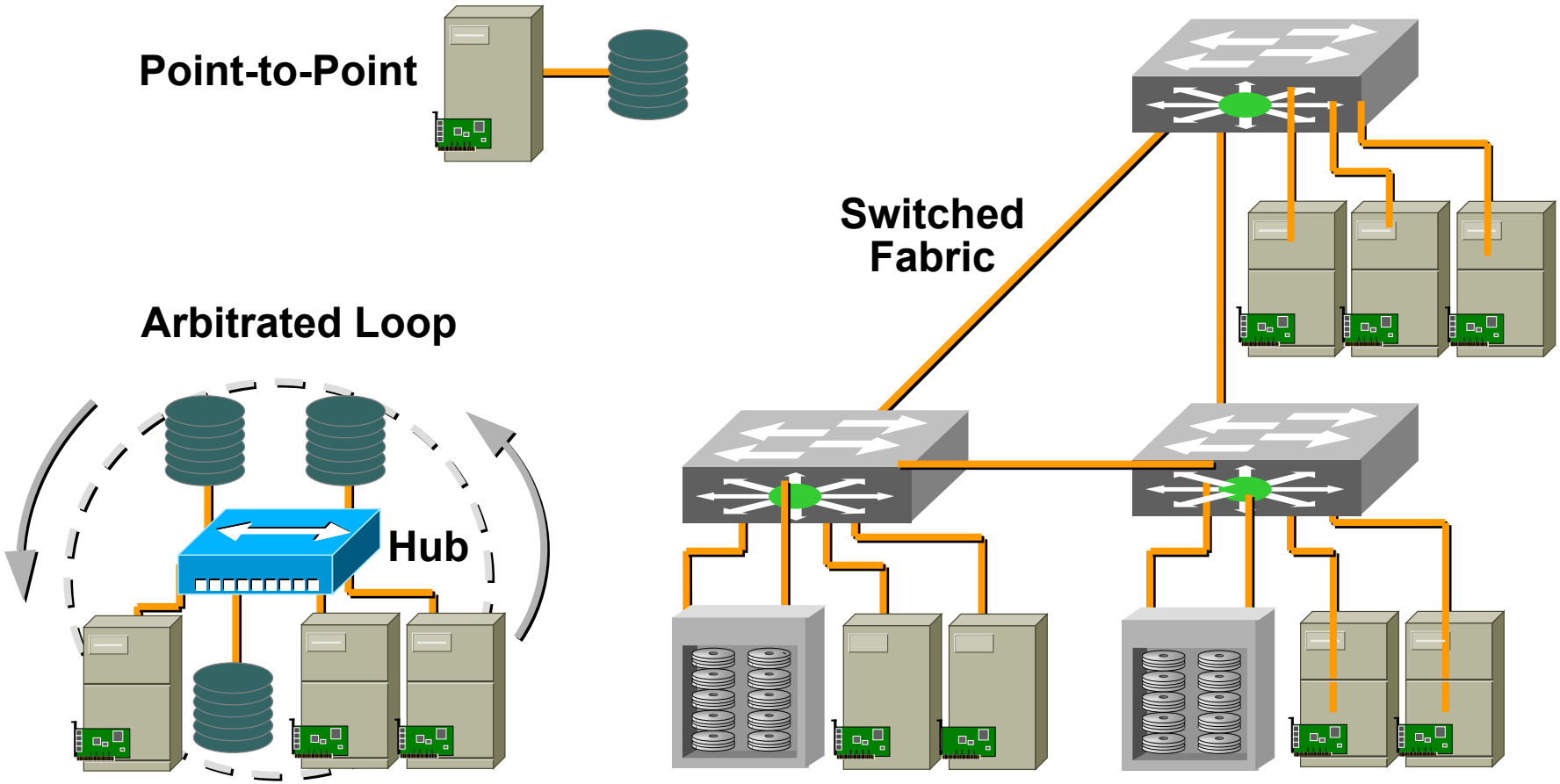
- Why
  - SCSI needed to get out of the parallel bus
- When
  - 1988 – 1995

    In 1995 Ethernet 100 Mb/s
  - 1 Gb/s in HW without loosing frames
  - Ad Hoc network
  - NIH syndrome
  - IETF was “basic Internet”

    January 1994, RFC 1577 “Classical IP and ARP over ATM”

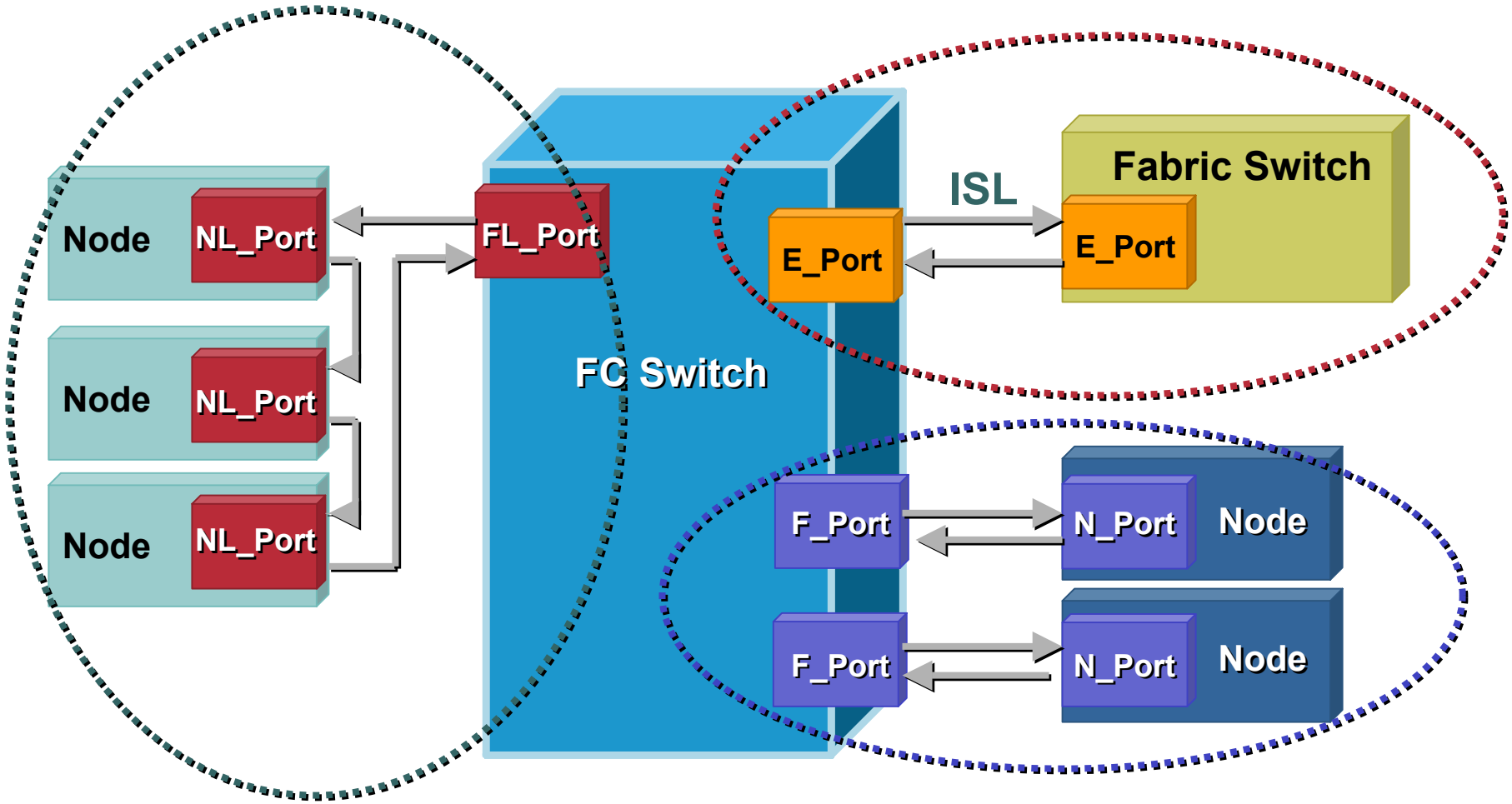
© 2002, Cisco Systems, Inc. All rights reserved.

Fibre Channel Topologies



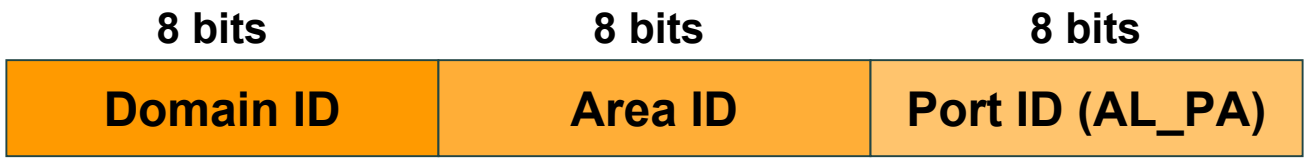
© 2002, Cisco Systems, Inc. All rights reserved.

FC Port Types



© 2002, Cisco Systems, Inc. All rights reserved.

Fibre Channel ID Format



- Domain ID
  - Identifies the switch
- Area ID
  - Identifies different loops connected to the same switch
- Port ID (or AL_PA)
  - Identifies the port on the switch (for N_Ports) or the specific node on the loop (for NL_Ports)

© 2002, Cisco Systems, Inc. All rights reserved.

FC: the evolution

- Stagnated on
  - Poor protocol design
  - Poor Interoperability
- Low volumes (nature of the beast)
- Resurrects few years ago on storage needs
  - Today it is the totality of the SAN market
- Improved interoperability
  - FC-PI, FC-FS, FC-MI, FC-DA, FC-SW3, FC-GS4
- Added
  - 4 Gbps
  - 10 Gbps

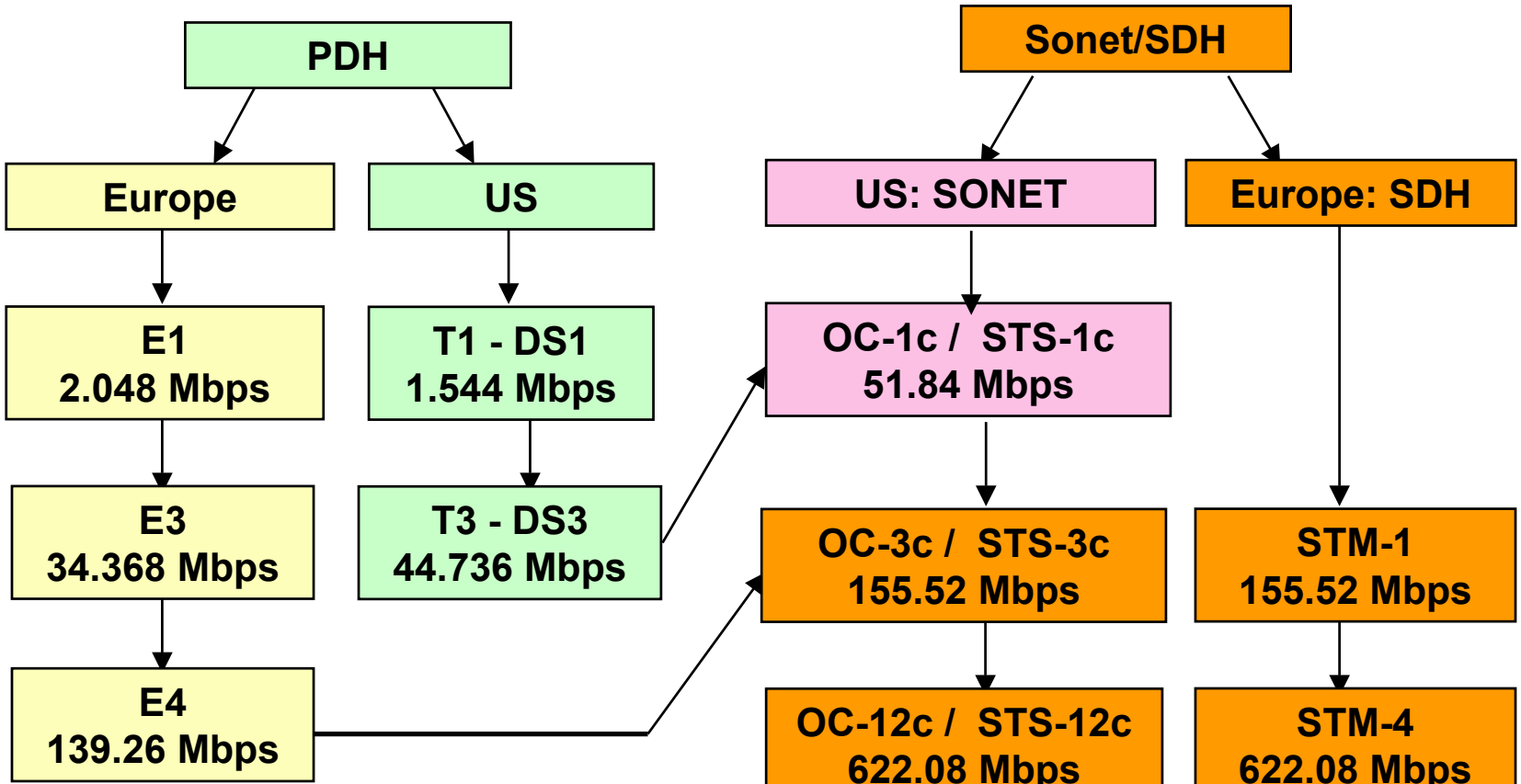
© 2002, Cisco Systems, Inc. All rights reserved.

Sonet/SDH: the Origin

- Telco flavor
  - Isochronous traffic
  - High Availability/Resiliancy
  - Distance
  - NEBS compliant
- Higher level protocols may see Sonet has a synchronous point-to-point link without loss

© 2002, Cisco Systems, Inc. All rights reserved.

PDH & Sonet/SDH



PDH: Plesiochronous Digital Hierarchy
SDH: Synchronous Digital Hierarchy

© 2002, Cisco Systems, Inc. All rights reserved.

Sonet: the evolution

- Grows on Telco money
  - Enabling Technology for ATM backbone, later dead
  - Now used for IP over WAN
- Gains some momentum
- Widespread adoption of fiber increases the momentum
- First to reach 10Gb/s
- Popular at OC-3 (155 Mbps) and OC-12 (622 Mbps)

© 2002, Cisco Systems, Inc. All rights reserved.

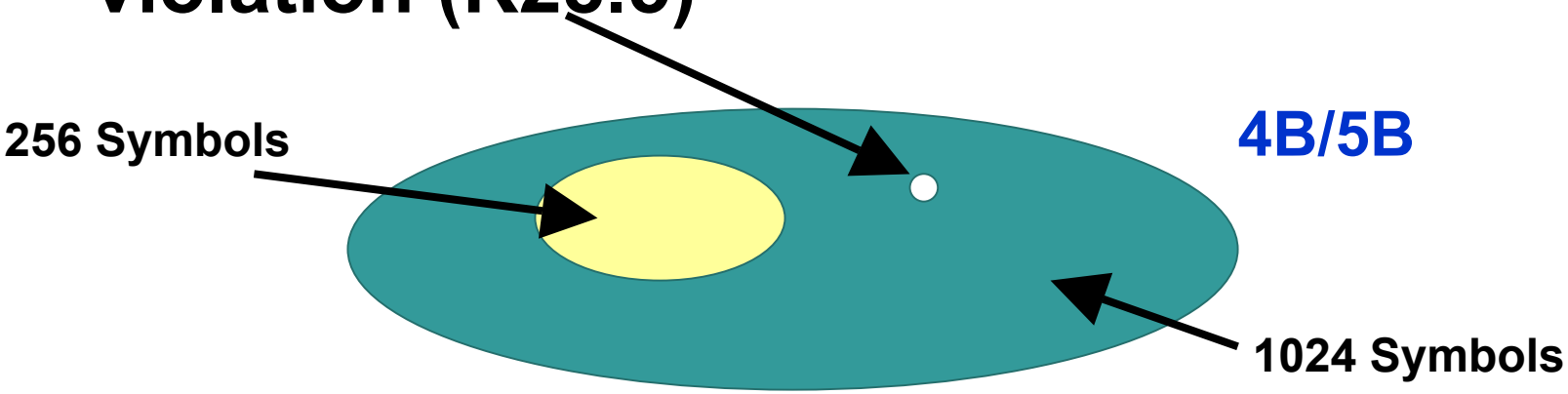
GFP/G.7041

- Generic Framing Procedure
- Frame-Mapped GFP:
  - Ethernet
  - PPP
- Transparent GFP:
  - Fibre Channel
  - Ficon
  - Escon
  - Transparent Gb Ethernet

© 2002, Cisco Systems, Inc. All rights reserved.

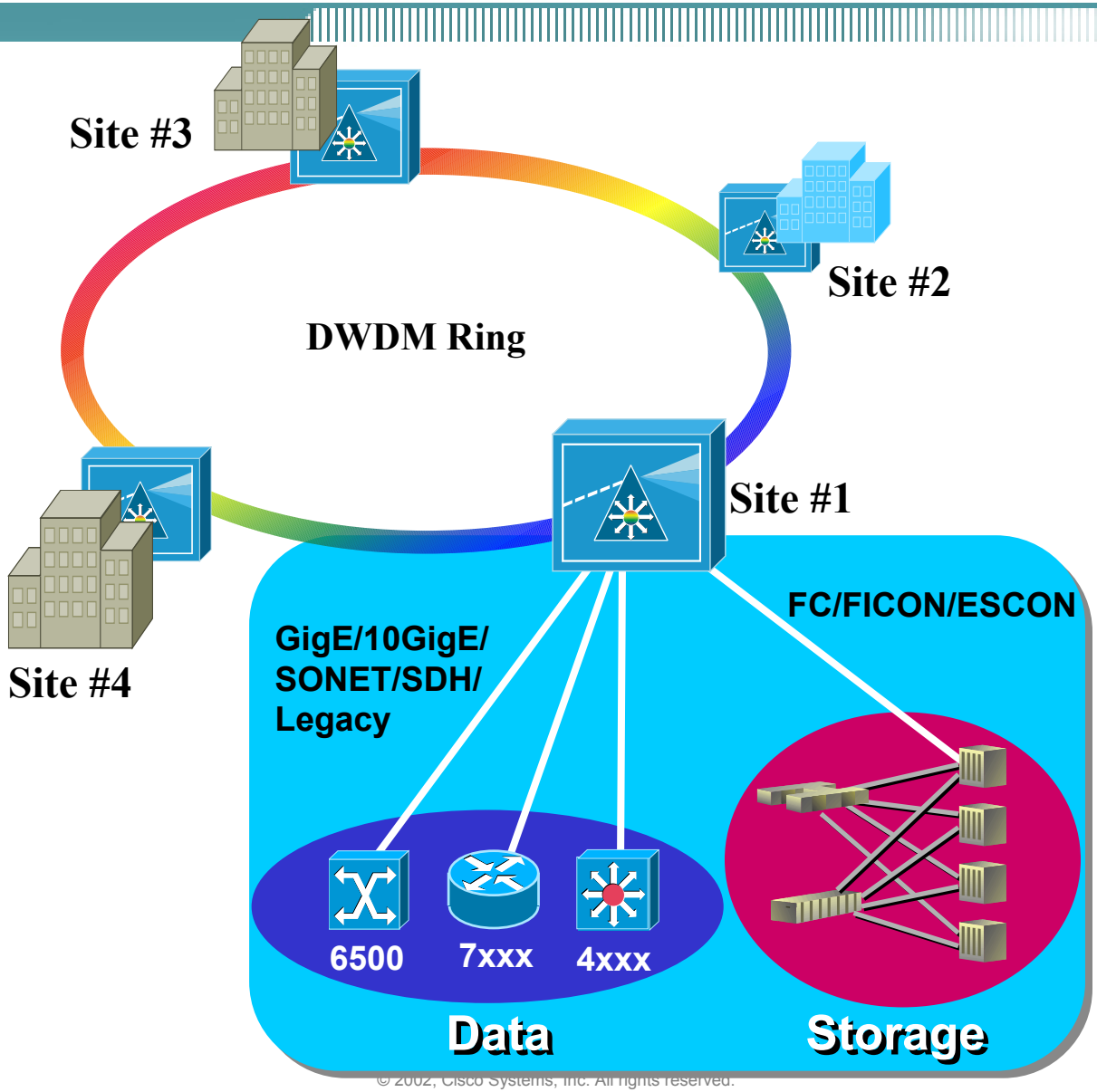
Why transparent GFP

- FC uses:
  - Frames
  - Ordered Sets (e.g. Idle, R_RDY)
- Ordered sets are special transmission words (4 bytes), the first byte is a code violation (K28.5)



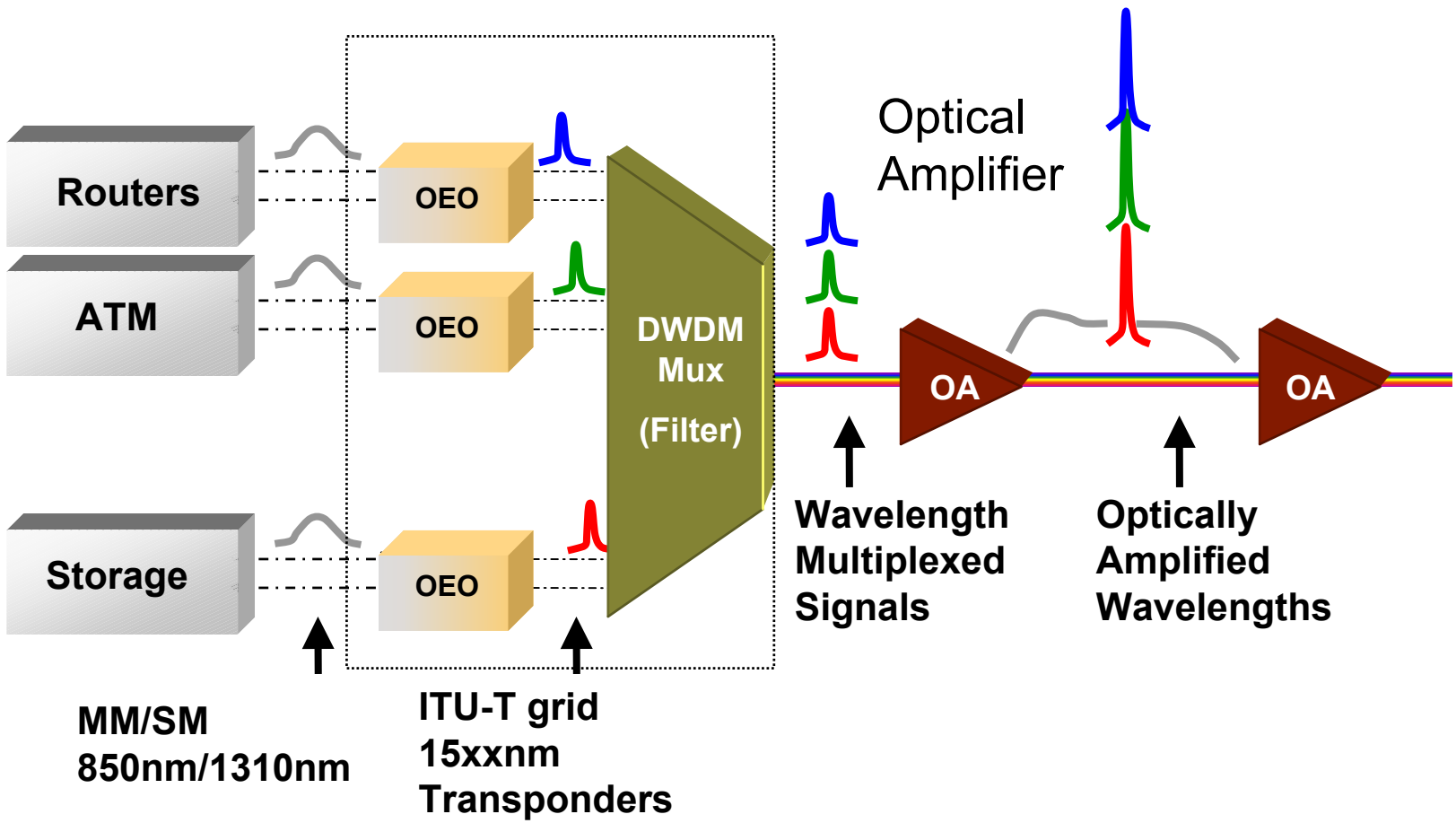
© 2002, Cisco Systems, Inc. All rights reserved.

DWDM/CWDM



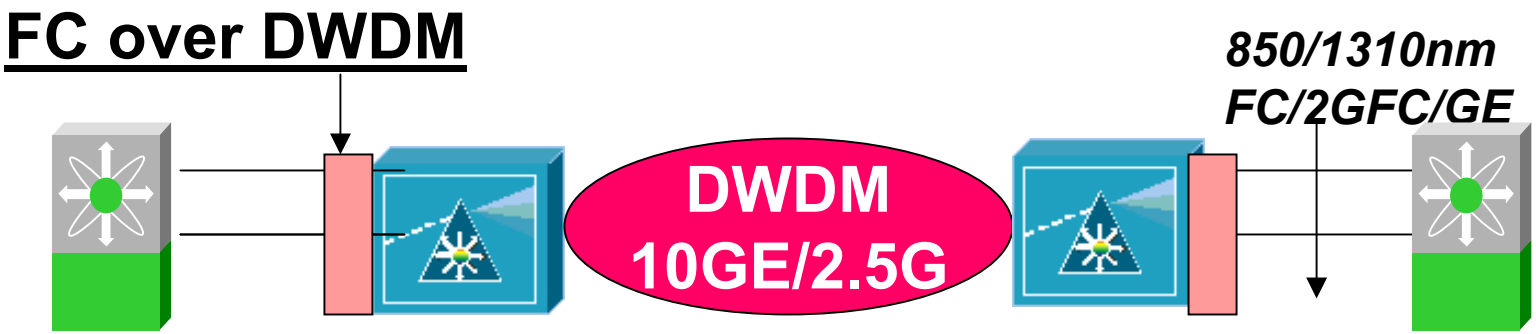
© 2002, Cisco Systems, Inc. All rights reserved.

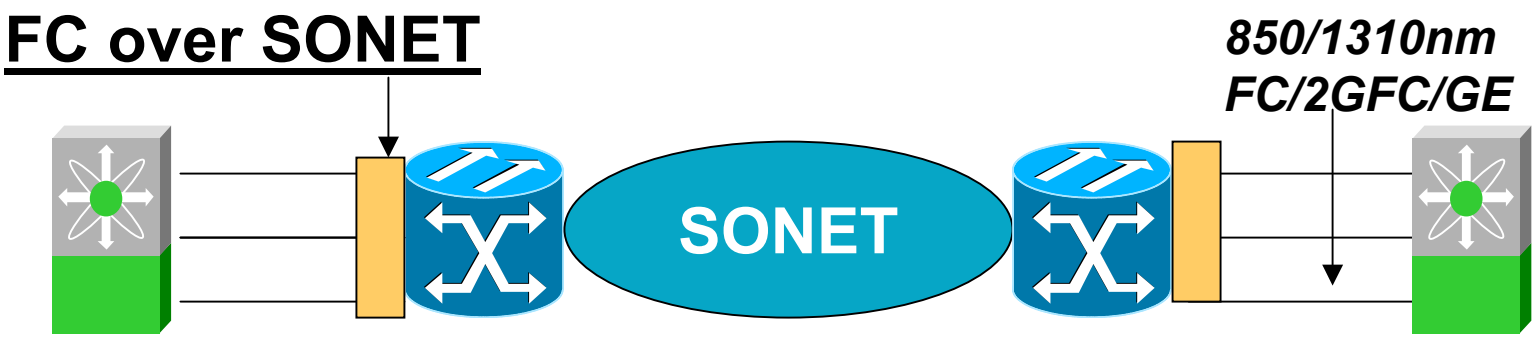
DWDM Principles



© 2002, Cisco Systems, Inc. All rights reserved.

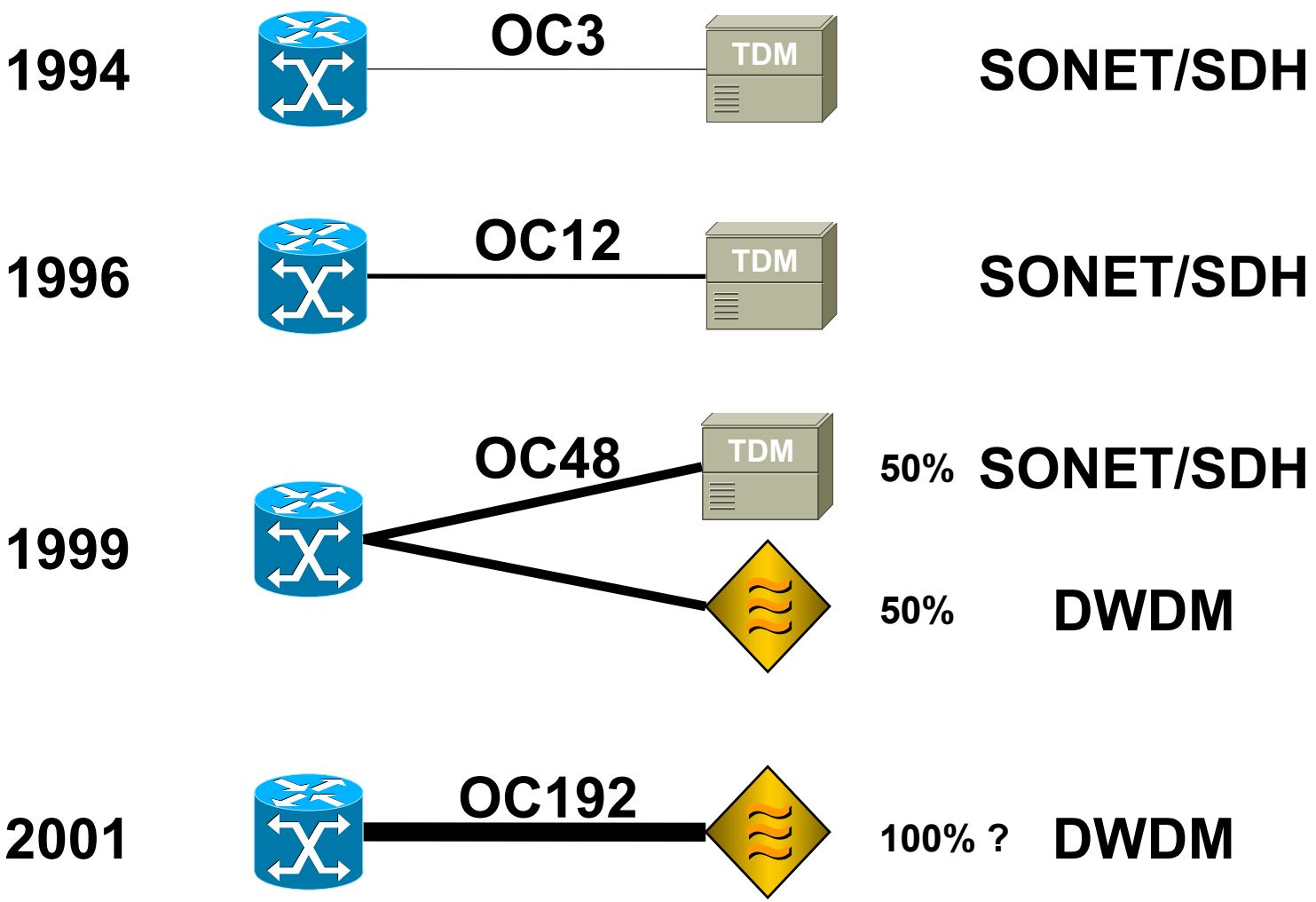
FC over Optical Today's Solutions





© 2002, Cisco Systems, Inc. All rights reserved.

Storage Interconnections



© 2002, Cisco Systems, Inc. All rights reserved.

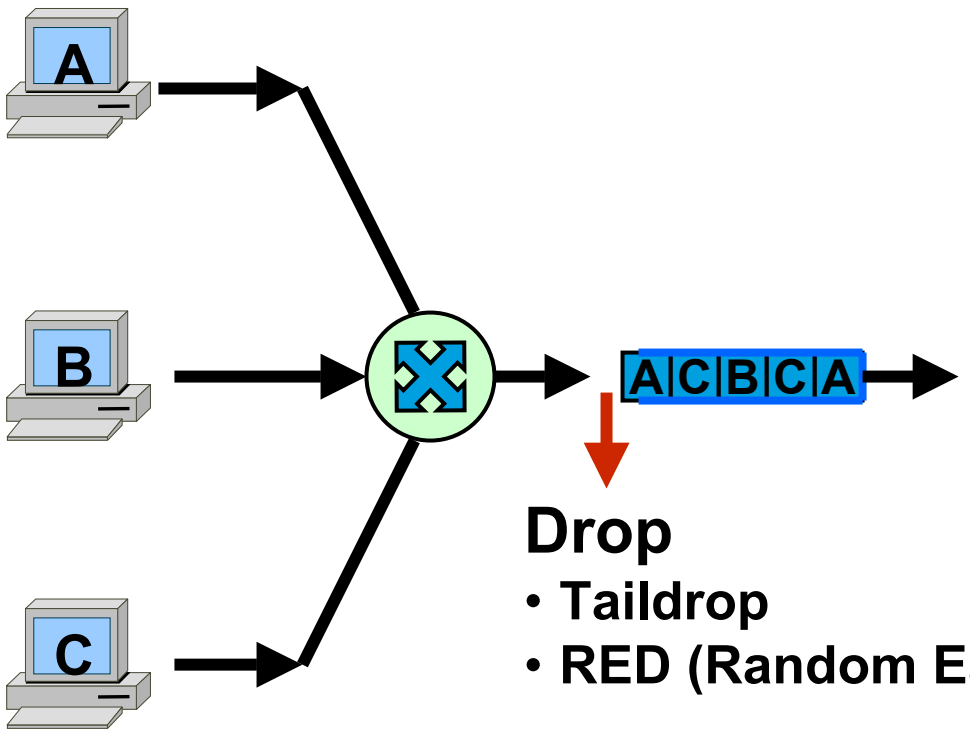
Packet Loss

- In IP/Ethernet
  - It's part of the game!
  - Used by TCP/IP to handle congestions
- In SCSI/Fibre Channel
  - Will throw you out of the market!
  - There is no congestion control!

© 2002, Cisco Systems, Inc. All rights reserved.

Why frames get dropped

- Not for transmission errors …
- Nor for collisions …
- … but for queue overflow due to congestion



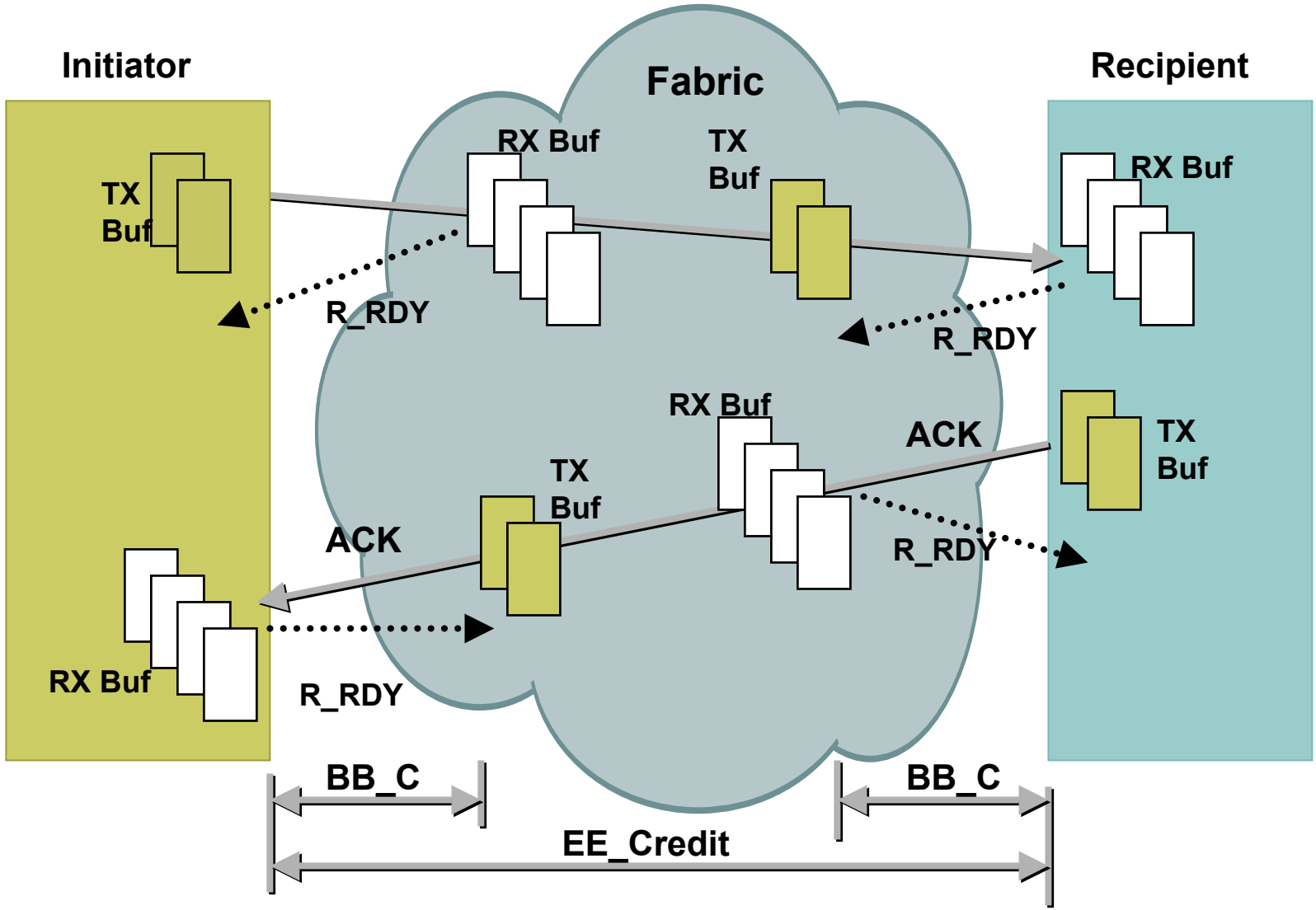
© 2002, Cisco Systems, Inc. All rights reserved.

To Drop or NOT TO Drop?

- *Queue in front of the link*
  - *Dropping or crediting?*
- No drop
  - FC native, or FC over Sonet/DWDM
  - No TCP
  - Credits
- Drop
  - TCP needed to recover
  - SCSI over TCP/IP

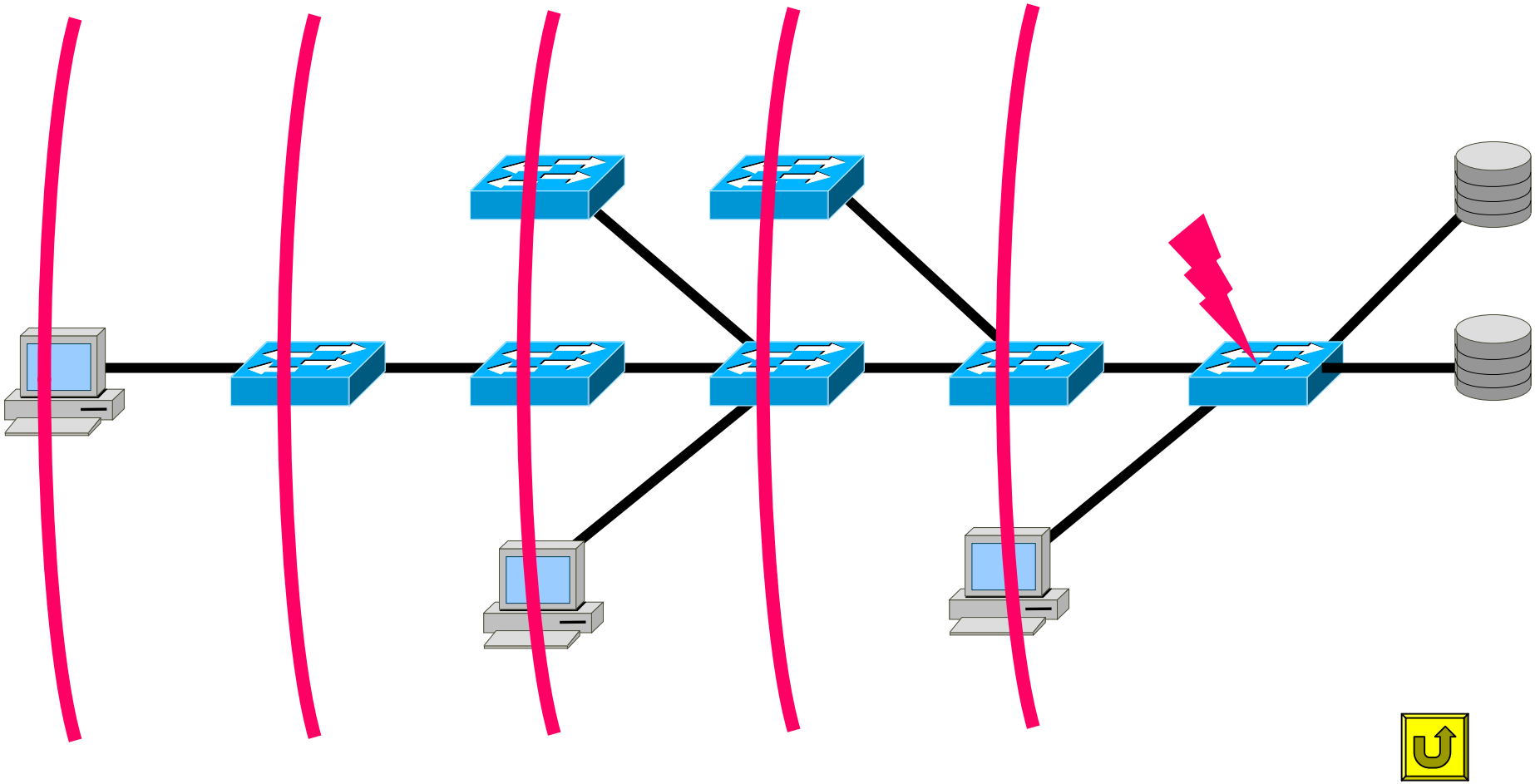
© 2002, Cisco Systems, Inc. All rights reserved.

Flow Control and Credits



© 2002, Cisco Systems, Inc. All rights reserved.

B2B Congestion Control



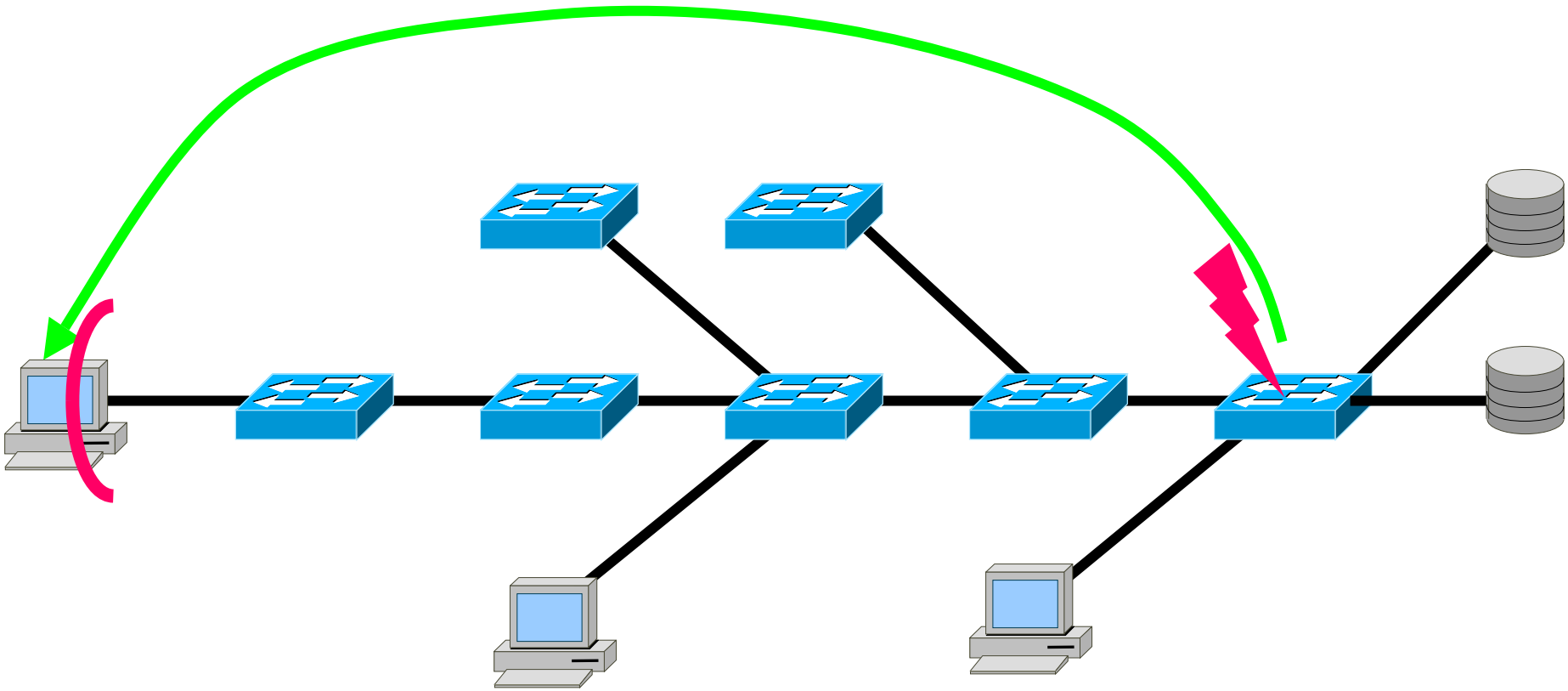
© 2002, Cisco Systems, Inc. All rights reserved.

Congestion

- TCP reacts to congestion differently from FC
  - It scales to the Internet
  - Van Jacobson taught us about windows
  - Congestion is signaled by packet loss
  - TCP slows down in the presence of congestion

© 2002, Cisco Systems, Inc. All rights reserved.

TCP Congestion Control



© 2002, Cisco Systems, Inc. All rights reserved.

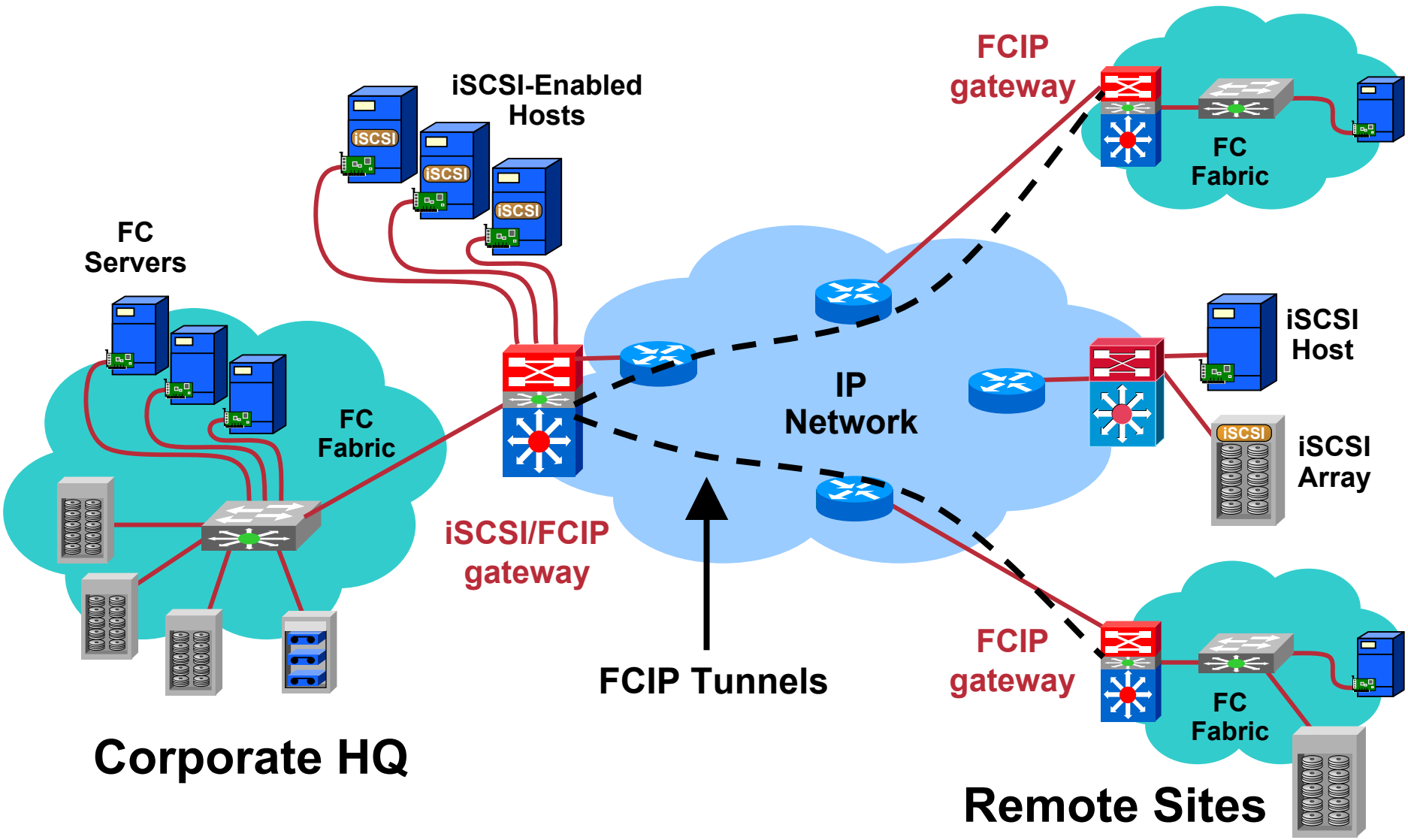
SCSI over TCP

- SCSI over TCP provides solution to carry storage traffic over Intranet/Internet
- Uses TCP, a reliable transport for delivery
- Can be used for local data center and long haul applications
- Two primary protocols:
  - iSCSI – IP-SCSI - used to transport SCSI CDBs and data within TCP/IP connections

  - FCIP – Fibre-Channel-over-IP – used to transport Fibre Channel frames within TCP/IP connections

© 2002, Cisco Systems, Inc. All rights reserved.

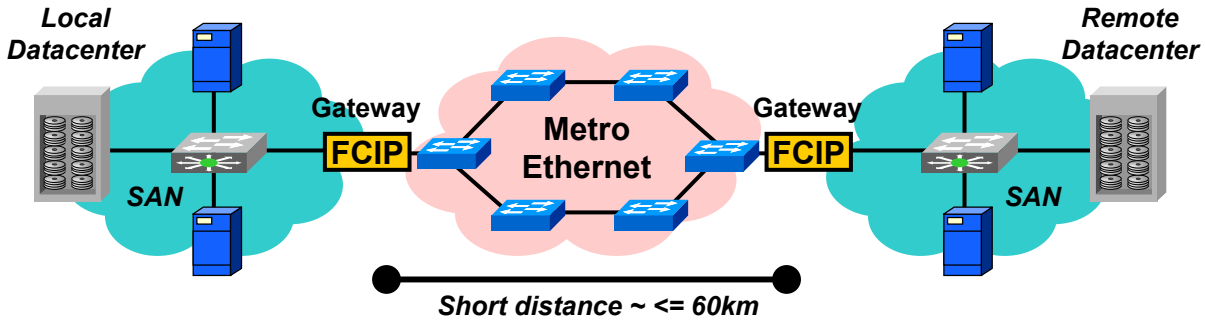
Example of iSCSI/FCIP Environment



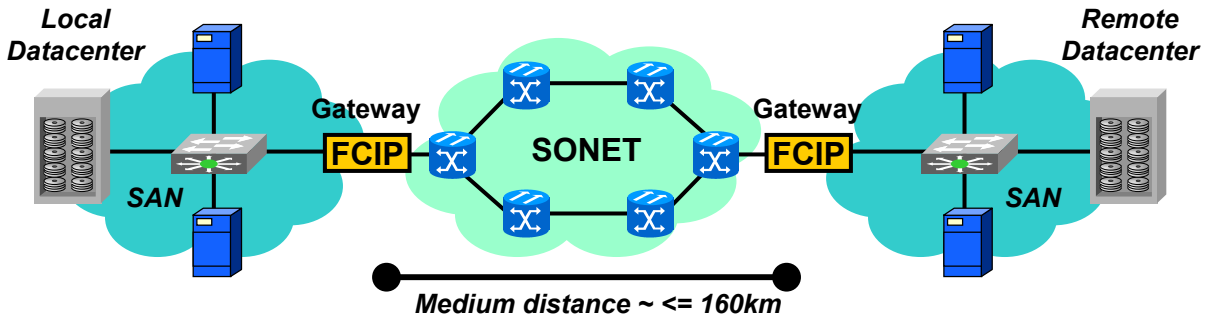
© 2002, Cisco Systems, Inc. All rights reserved.

Potential FCIP Environments

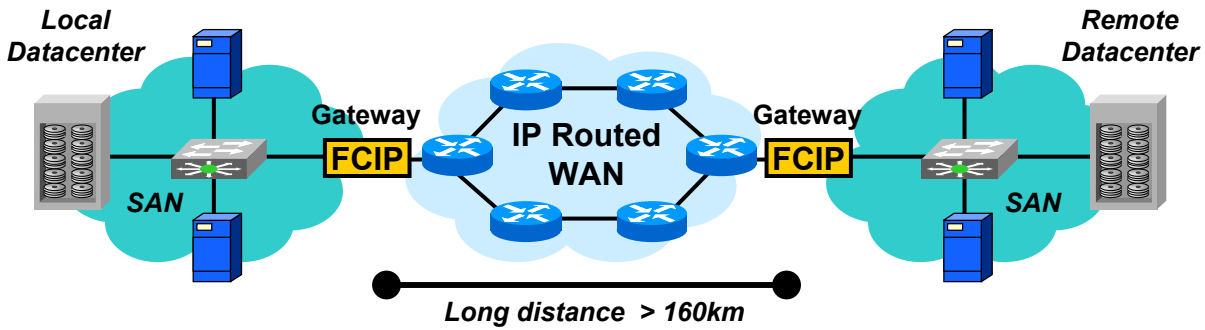
- Near wire-rate (1Gbps)
- Relatively low latency
- Mainly asynchronous
- Suitable for some synchronous apps



- Typical OC3 / OC12
- Relatively low latency
- Mainly asynchronous
- Suitable for some synchronous apps



- Low speed (T1 – DS3)
- Higher latency
- Longer distance
- Mainly asynchronous



© 2002, Cisco Systems, Inc. All rights reserved.

TOEs

- It is difficult to implement TCP in HW
  - At 10Gb/s TCP is tough !!!
- The few TOEs that work are aliens in the OS
- Overall performance is required
  - True Zero Copy
- RDMA
  - Significant OS and application changes
  - Never took off
- At same performance/same efficiency, same cost of HBAs

© 2002, Cisco Systems, Inc. All rights reserved.

Comparing IP with FC

- FC is limited
  - Size
  - Congestion
- … while IP is not, … or is it?

© 2002, Cisco Systems, Inc. All rights reserved.

Size limitation

- 239 switches in 100 KM radius
  - 256 available Domain IDs minus 17 reserved
  - Is it a limitation?
- Above 100 KM, light is slow
  - 100 KM = 200KM round trip, 1 ms
- Asynchronous operation
  - To avoid delay issues

© 2002, Cisco Systems, Inc. All rights reserved.

Congestion

- TCP reacts to congestion differently from FC
  - … but slow down implies

    Reduced throughput

    Increased latency

© 2002, Cisco Systems, Inc. All rights reserved.

Ethernet/IP vs Fibre Channel

- Neither of them guarantees low latency and high throughput in the presence of congestion
- Should we rethink the solution and add traffic engineering concepts?
  - The telephone network has used it with success
  - The IETF has had some success with MPLS
- Traffic engineering for Storage?

© 2002, Cisco Systems, Inc. All rights reserved.

© 2002, Cisco Systems, Inc. All rights reserved.