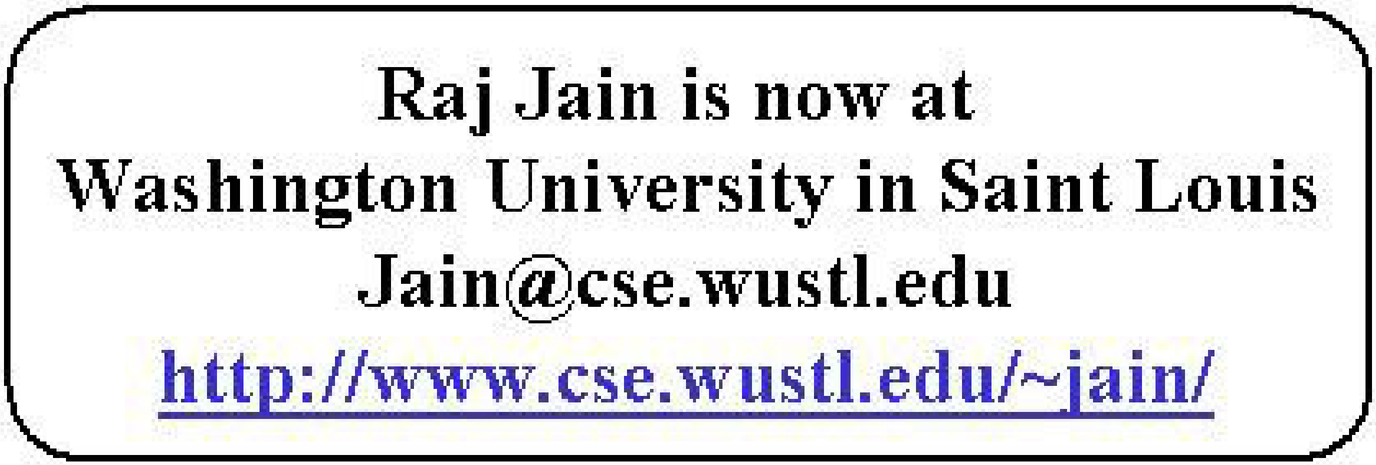
Next Generation Data Networking

Raj Jain

The ks

( )35

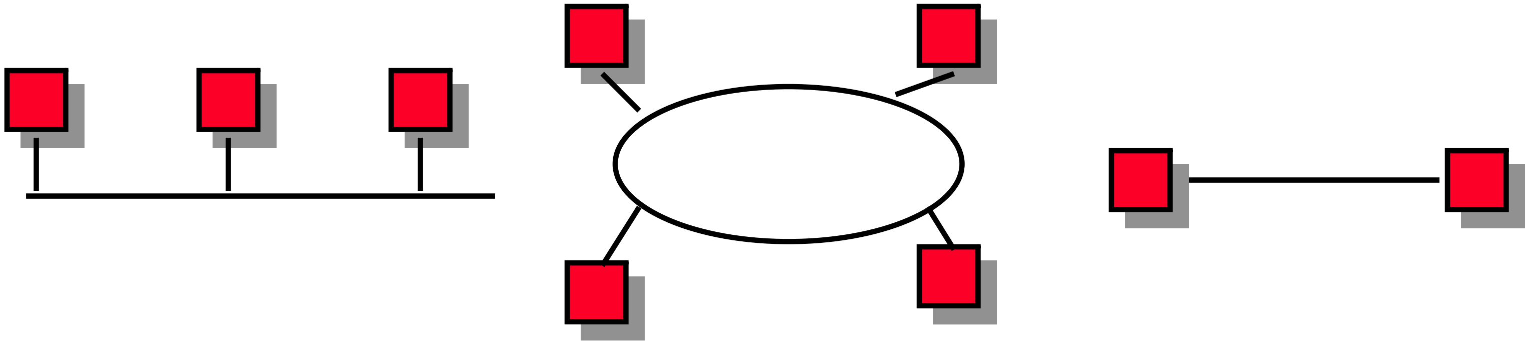


Raj Jain is now at
Washington University in Saint Louis
Jain@cse.wustl.edu
http://www.cse.wustl.edu/~jain/

Overview

- 100 Mbps Ethernet
- Gigabit Ethernet
- 10 G Ethernet
- Resilient Packet Rings
- Next Generation SONET: VCAT, GFP, LCAS
- Frame Relay

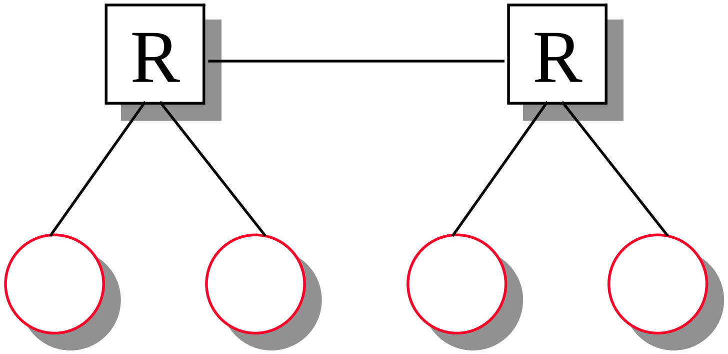
Distance-B/W Principle



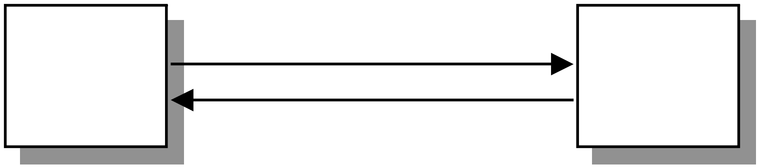
- Efficiency = Max throughput/Media bandwidth
- Efficiency is a non-increasing function of α
  α = Propagation delay /Transmission time
  = (Distance/Speed of light)/(Transmission size/Bits/sec)
  = Distance×Bits/sec/(Speed of light)(Transmission size)
- Bit rate-distance-transmission size tradeoff.
- 100 Mb/s ⇒ Change distance or frame size

Ethernet vs Fast Ethernet

	Ethernet	Fast Ethernet
Speed	10 Mbps	100 Mbps
MAC	CSMA/CD	CSMA/CD
Network diameter	2.5 km	205 m
Topology	Bus, star	Star
Cable	Coax, UTP, Fiber	UTP, Fiber
Standard	802.3	802.3u
Cost	X	2X



Full-Duplex Ethernet



- Uses point-to-point links between TWO nodes
- Full-duplex bi-directional transmission
- Transmit any time
- Many vendors are shipping switch/bridge/NICs with full duplex
- No collisions ⇒ 50+ Km on fiber.
- Between servers and switches or between switches

1 GbE: Key Design Decisions

- P802.3z ⇒ Update to 802.3
  Compatible with 802.3 frame format, services, management
- 1000 Mb vs. 800 Mb Vs 622 Mbps
  Single data rate
- LAN distances only
- No Full-duplex only ⇒ Shared Mode
  Both hub and switch based networks
- Same min and max frame size as 10/100 Mbps
  ⇒ Changes to CSMA/CD protocol
  Transmit longer if short packets

1000Base-X

- 1000Base-LX: 1300-nm laser transceivers
  - 2 to 550 m on 62.5-μm or 50-μm multimode, 2 to 5000 m on 10-μm single-mode
- 1000Base-SX: 850-nm laser transceivers
  - 2 to 275 m on 62.5-μm, 2 to 550 m on 50-μm. Both multimode.
- 1000Base-CX: Short-haul copper jumpers
  - 25 m 2-pair shielded twinax cable in a single room or rack.

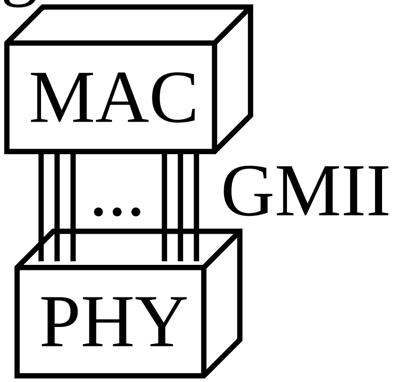
    Uses **8b/10b** coding ⇒ 1.25 GBaud/s line rate
- *1000Base-ZX: Long haul lasers to 70 km (not Std)*

1000Base-T

- 100 m on 4-pair Cat-5 UTP
  ⇒ Network diameter of 200 m
- Applications: Server farms, High-performance workgroup, Network computers
- Supports CSMA/CD (Half-duplex):
  Carrier Extension, Frame Bursting
- 250 Mbps/pair full-duplex DSP based PHY
  ⇒ Requires new 5-level (PAM-5) signaling
  with 4-D 8-state Trellis code FEC
- FEC coded symbols.
  Octet data to 4 quinary (5-level) symbols and back,
  e.g., 001001010 = {0, -2, 0, -1}

1000BASE-T (Cont)

- Inside PHY, before coding, the data is scrambled using $x^{33}+x^{20}+1$ in one direction and $x^{33}+x^{13}+1$ self-synchronizing scrambler in the other direction
- Automatically detects and corrects pair-swapping, incorrect polarity, differential delay variations across pairs
- Autonegotiation ⇒ Compatibility with 100Base-T
- Complies with Gigabit Media Independent Interface
- 802.3ab-1999



10 GbE: Key Design Decisions

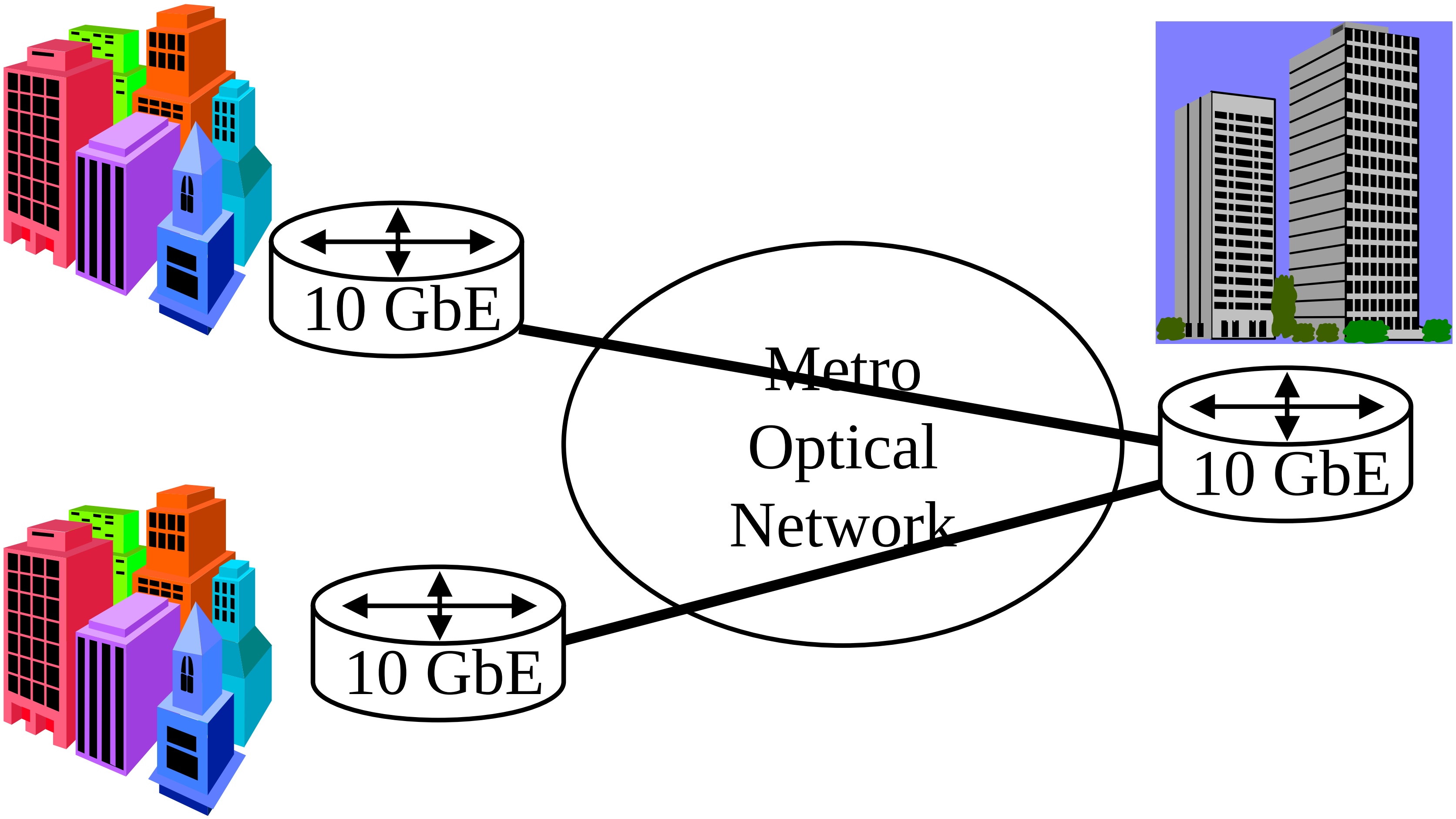
- P802.3ae ⇒ Update to 802.3
  Compatible with 802.3 frame format, services, management
- 10 Gbps vs. 9.5 Gbps. **Both** rates.
- LAN and MAN distances
- Full-duplex only ⇒ No Shared Mode
  Only switch based networks. No Hubs.
- Same min and max frame size as 10/100/1000 Mbps
  Point-to-point ⇒ No CSMA/CD protocol
- 10.000 Gbps at MAC interface
  ⇒ Flow Control between MAC and PHY

10 GbE PMD Types

PMD	Description	MMF	SMF
10GBASE-R:			
10GBASE-SR	850nm Serial LAN	300 m	N/A
10GBASE-LR	1310nm Serial LAN	N/A	10 km
10GBASE-ER	1550nm Serial LAN	N/A	40 km
10GBASE-X:			
10GBASE-LX4	1310nm WWDM LAN	300 m	10 km
10GBASE-W:			
10GBASE-SW	850nm Serial WAN	300 m	N/A
10GBASE-LW	1310nm Serial WAN	N/A	10 km
10GBASE-EW	1550nm Serial WAN	N/A	40 km
10GBASE-LW4	1310nm WWDM WAN	300 m	10 km

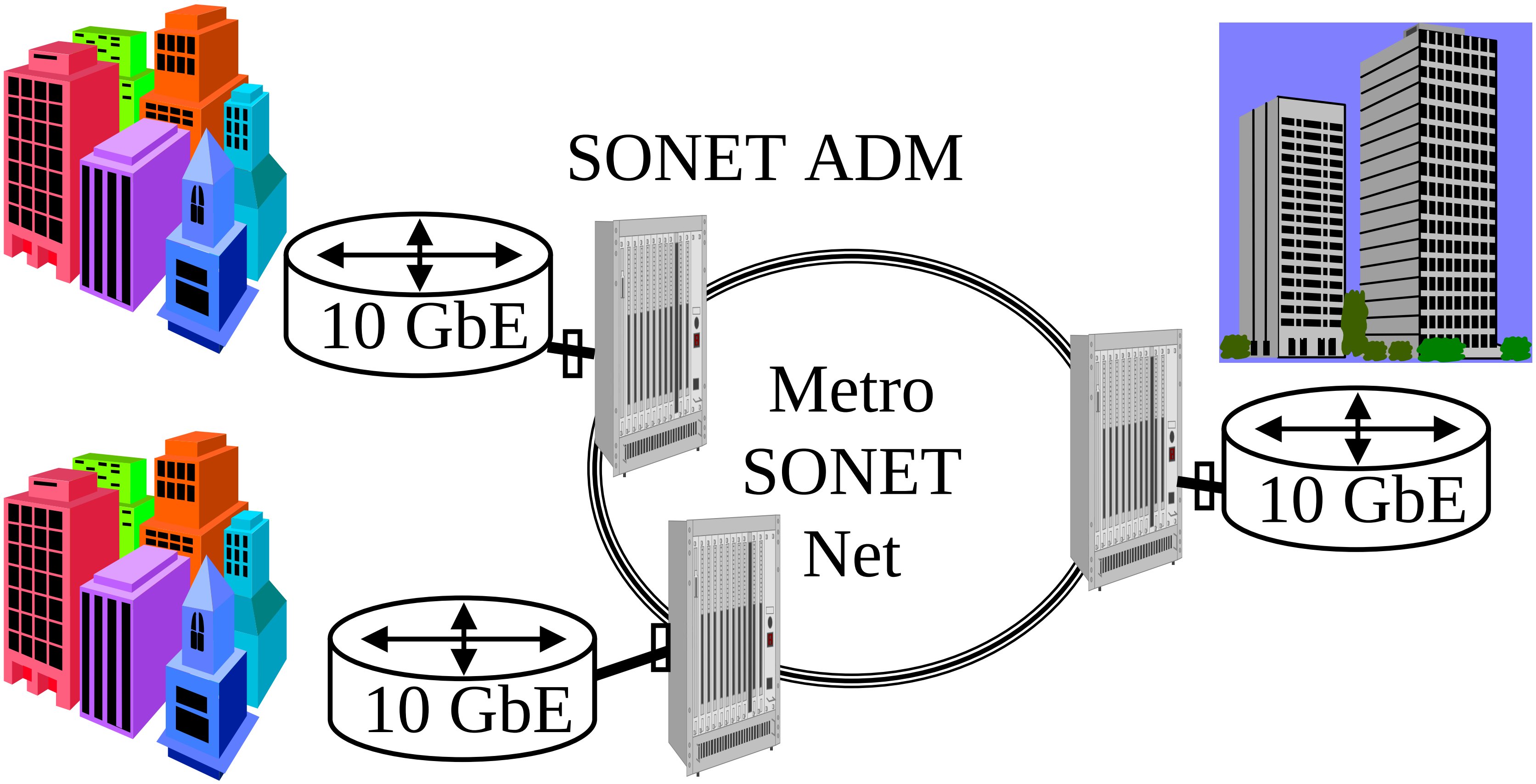
- S = Short Wave, L=Long Wave, E=Extra Long Wave
- R = Regular reach (64b/66b), W=WAN (64b/66b + SONET Encapsulation), X = 8b/10b
- 4 = 4 λ's

10 GbE over Dark Fiber



- Need only LAN PMD up to 40 km.
  No SONET overhead. No protection.

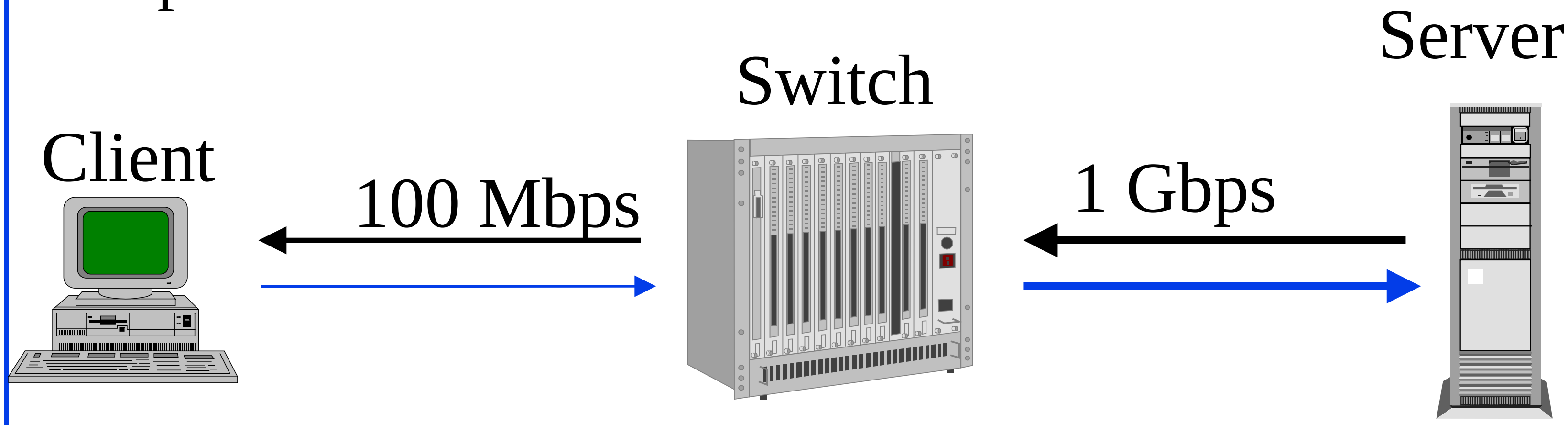
10 GbE over SONET/SDH



- Using WAN PMD.
  Legacy SONET. Protection via rings.
  ELTE = Ethernet Line Terminating Equipment

802.3x Full-Duplex Flow Control

- Pause frame with pause time sent to multicast address 01-80-C2-00-00-01 not forwarded by bridges
- Autonegotiation updated to include a “flow-control capable” bit



1. Client disk full

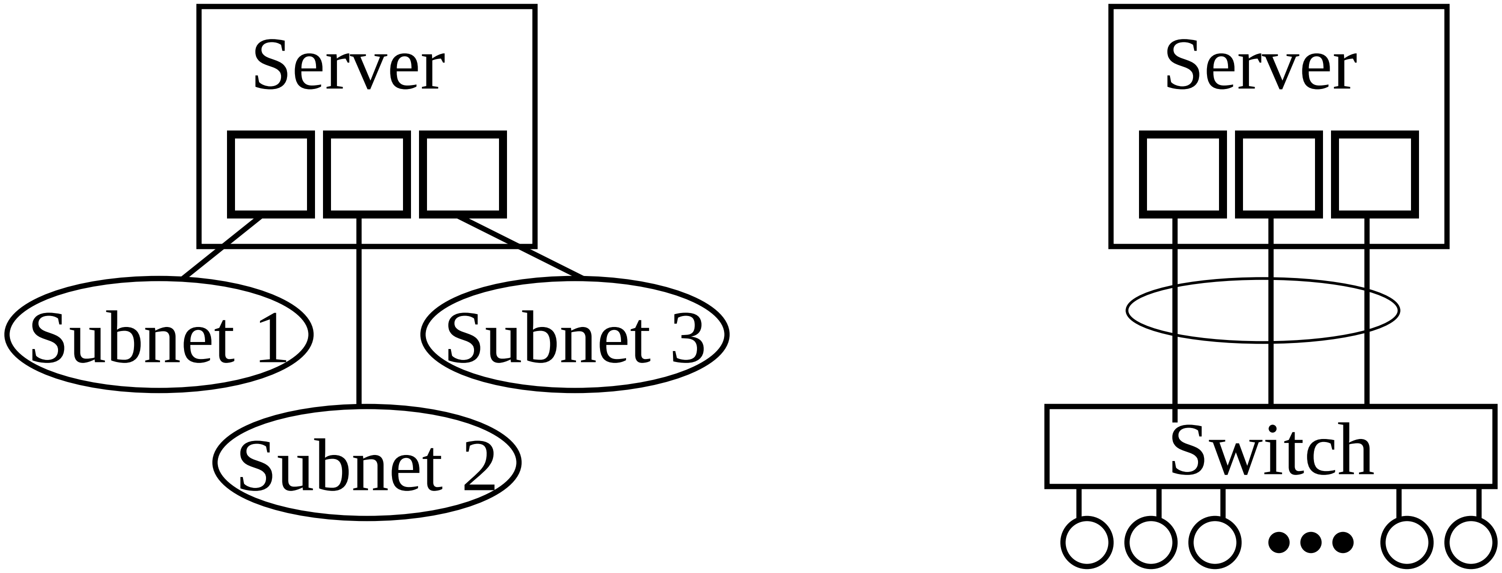
2. Client sends Pause

3. Switch buffers fill up

4. Switch sends Pause

5. Server Stops

802.3ad Link Aggregation



- ❑ Multi-Link Trunking (MLT) allows n parallel links to act as one link ⇒ Server needs only one IP address.
- ❑ For redundancy and incremental bandwidth
- ❑ Cost < nX
- ❑ Ideal up to 4 links. Approved March 2000.

Jumbo Frames

- Maximum Ethernet Frame Size = 1518 bytes or 1522 bytes (with VLAN Tags)
- Frame size too small at Gbps and higher speed
- 9kB implemented by Alteon WebSystems
- 9k-16kB being talked about in the industry
- Is not an IEEE standard
- Ref: http://www.nwfusion.com/newsletters/lans/0614lan1.html

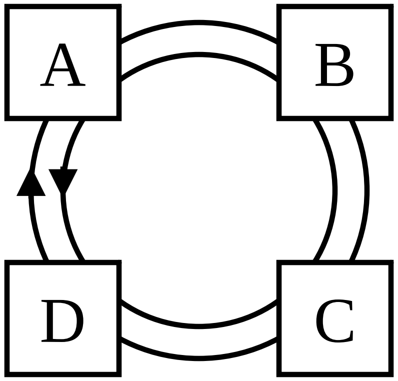
Future Possibilities

- 40 Gbps
- 100 Gbps:
  - 16λ×6.25 Gbps
  - 8λ × 12.5 Gbps
  - 4λ × 12.5 using PAM-5
- 160 Gbps
- 1 Tbps:
  - 12 fibers with 16λ × 6.25 Gbps
  - 12 fibers with 8λ × 12.5 Gbps
- 70% of 802.3ae members voted to start 40G in 2002

Feature	SONET	Ethernet
Payload Rates	51M, 155M, 622M, 2.4G, 9.5G	10M, 100M, 1G, 10G
Payload Rate Granularity	Fixed	√Any
Bursty Payload	No	√Yes
Payload Count	One	√Multiple
Protection	√Ring	Mesh
OAM&P	√Yes	No
Synchronous Traffic	√Yes	No
Restoration	√50 ms	Minutes
Cost	High	√Low
Used in	Telecom	Enterprise

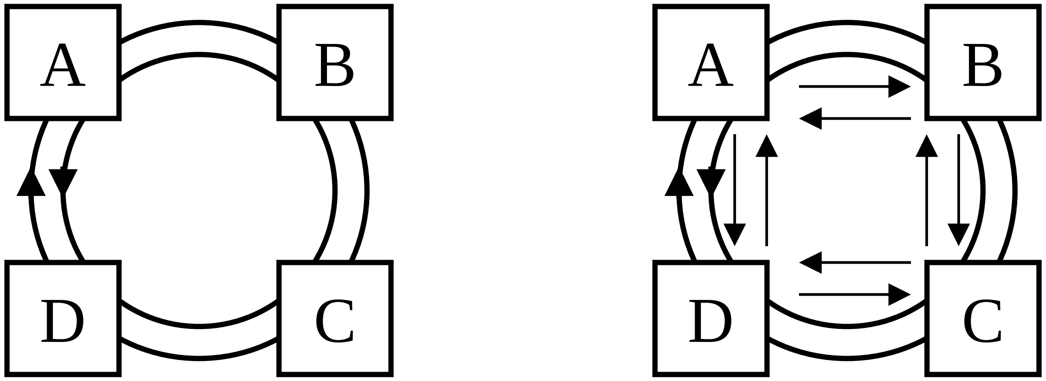
Feature	SONET	Ethernet	Remedy
Payload Rates	51M, 155M, 622M, 2.4G, 9.5G	10M, 100M, 1G, 10G	10GE at 9.5G
Payload Rate Granularity	Fixed	√Any	Virtual Concatenation
Bursty Payload	No	√Yes	Link Capacity Adjustment Scheme
Payload Count	One	√Multiple	Packet GFP
Protection	√Ring	Mesh	Resilient Packet Ring (RPR)
OAM&P	√Yes	No	In RPR
Synchronous Traffic	√Yes	No	MPLS + RPR
Restoration	√50 ms	Minutes	Rapid Spanning Tree
Cost	High	√Low	Converging
Used in	Telecom	Enterprise	

RPR: Key Features



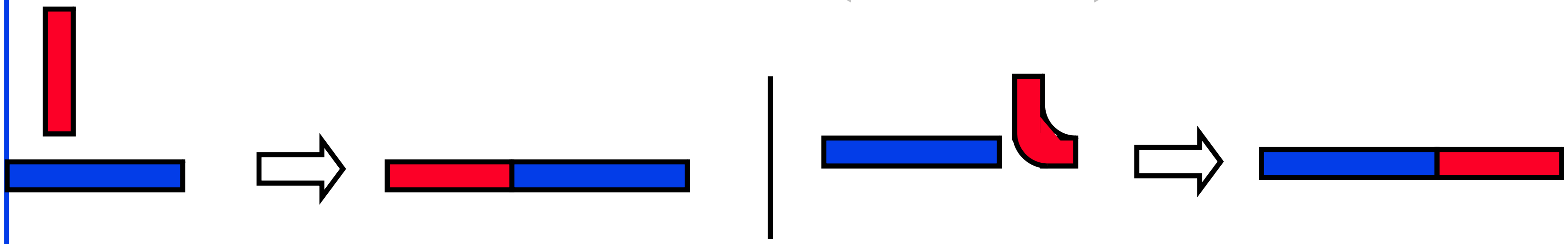
- Dual Ring topology
- Supports broadcast and multicast
- Packet based ⇒ Continuous bandwidth granularity
- Max 256 nodes per ring
- MAN distances: Several hundred kilometers.
- Gbps speeds: Up to 10 Gbps

RPR Features (Cont)



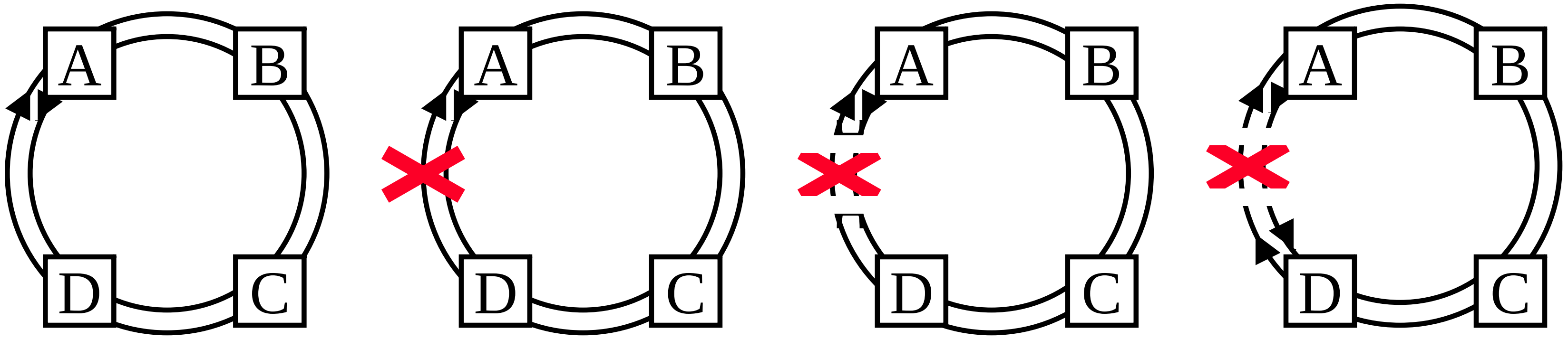
- Both rings are used (unlike SONET)
- Normal transmission on the shortest path
- Destination stripping ⇒ Spatial reuse
  Multicast packets are source stripped
- Five Classes of traffic: Reserved, High-Priority, Medium Priority, Low Priority, Control

RPR (Cont)



- Buffer Insertion Ring: Absolute but non-preemptive priority to pass-through traffic
- Cut-through of transit packets optional.
- Bandwidth management: Unused bandwidth is advertised so that others can use it
- Fairness Algorithm for fair and efficient bandwidth use
- Physical Layer Independent: GbE/10GE or SONET with GFP or PoS

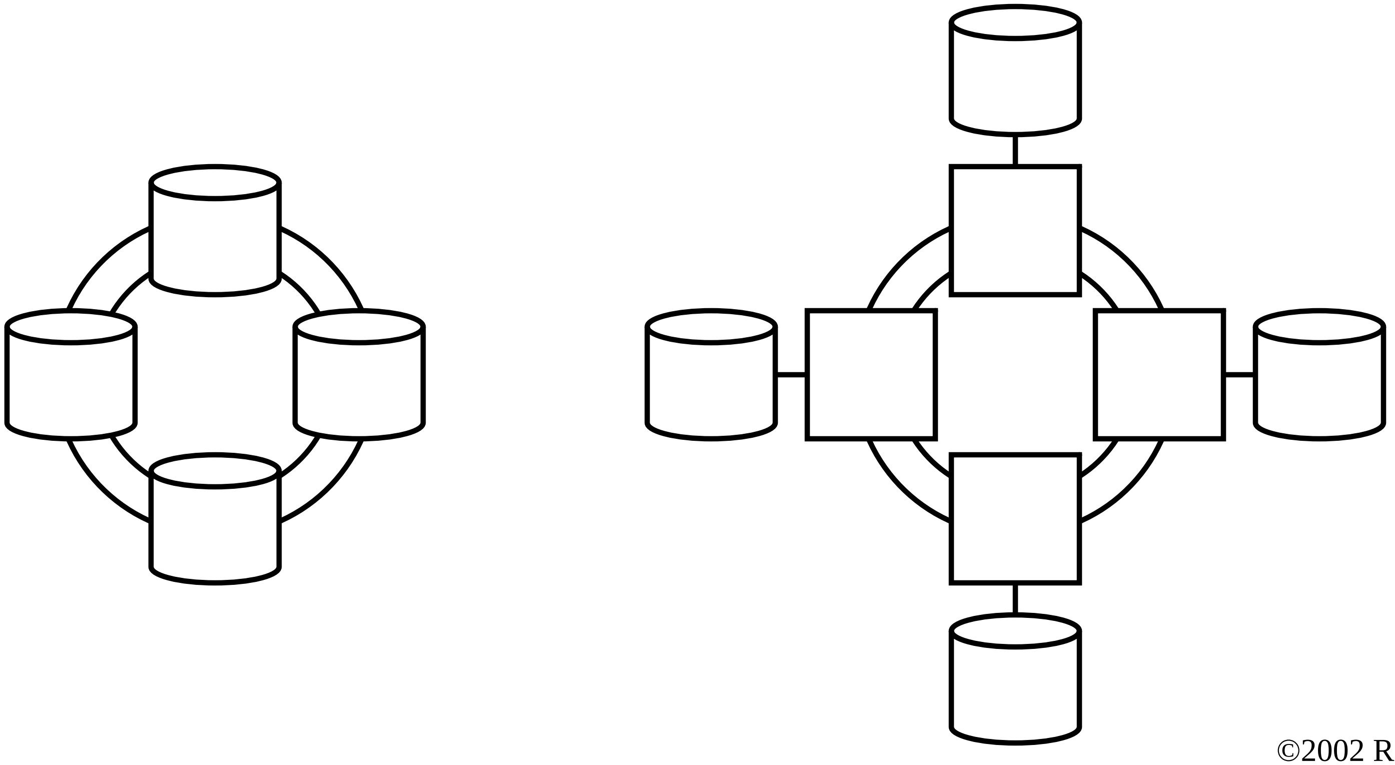
RPR Protection Mechanisms



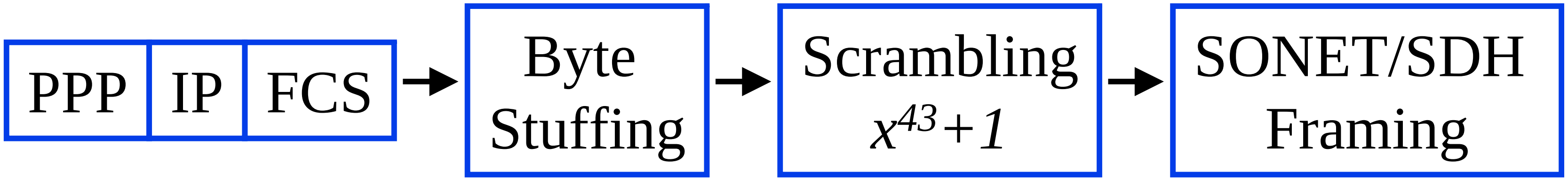
1. **Wrapping**: Stations adjacent to failure wrap. After re-org, packets sent on shortest path. Multicast packets are sent on **one** ring with TTL=Total number of stations.
2. **Source Steering**: Failure detecting station sends a Protection Request message to every station. Sources select appropriate ringlet to reach their destination. Multicast packets are sent on **both** rings with TTL=Total number of stations

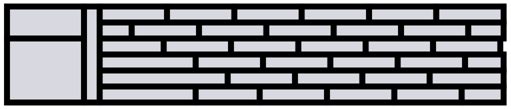
RPR Issues

- Ring vs Mesh (Atrica)
- Router Feature vs Dedicated RPR Node (Cisco, Redback, Riverstone vs Luminous)



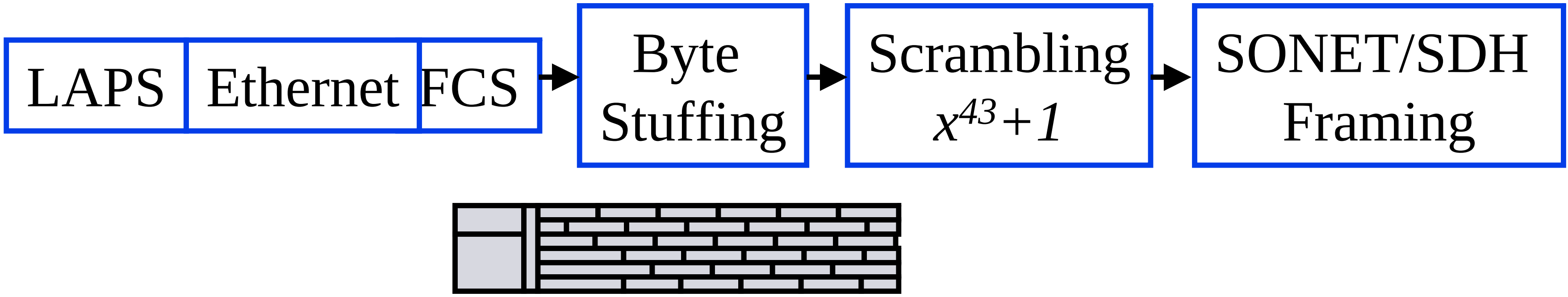
Packet over SONET (PoS)


PPP IP FCS → Byte Stuffing → Scrambling $x^{43}+1$ → SONET/SDH Framing



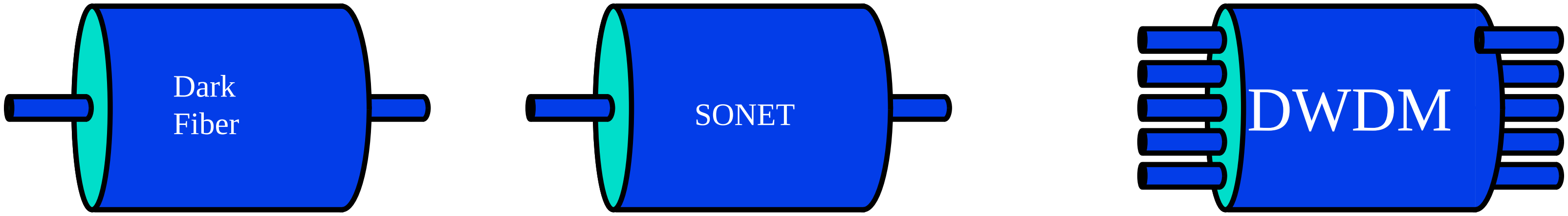
- ❑ PoS = IP over PPP over SONET
- ❑ Byte stuffing to avoid “Frame delimiter” in data
- ❑ Scrambling to avoid all zeros or all ones in SONET payload
- ❑ Path Signal Label C2 = 2210 ⇒ PPP w scambling
  20710 ⇒ PPP w/o scambling
- ❑ Ref: RFC 2615

X.86: Ethernet over SDH (EoS)



LAPS | Ethernet | FCS → Byte Stuffing → Scrambling $x^{43}+1$ → SONET/SDH Framing

- Link Access Procedure for SDH (LAPS)
  Like PPP but a different variant of HDLC



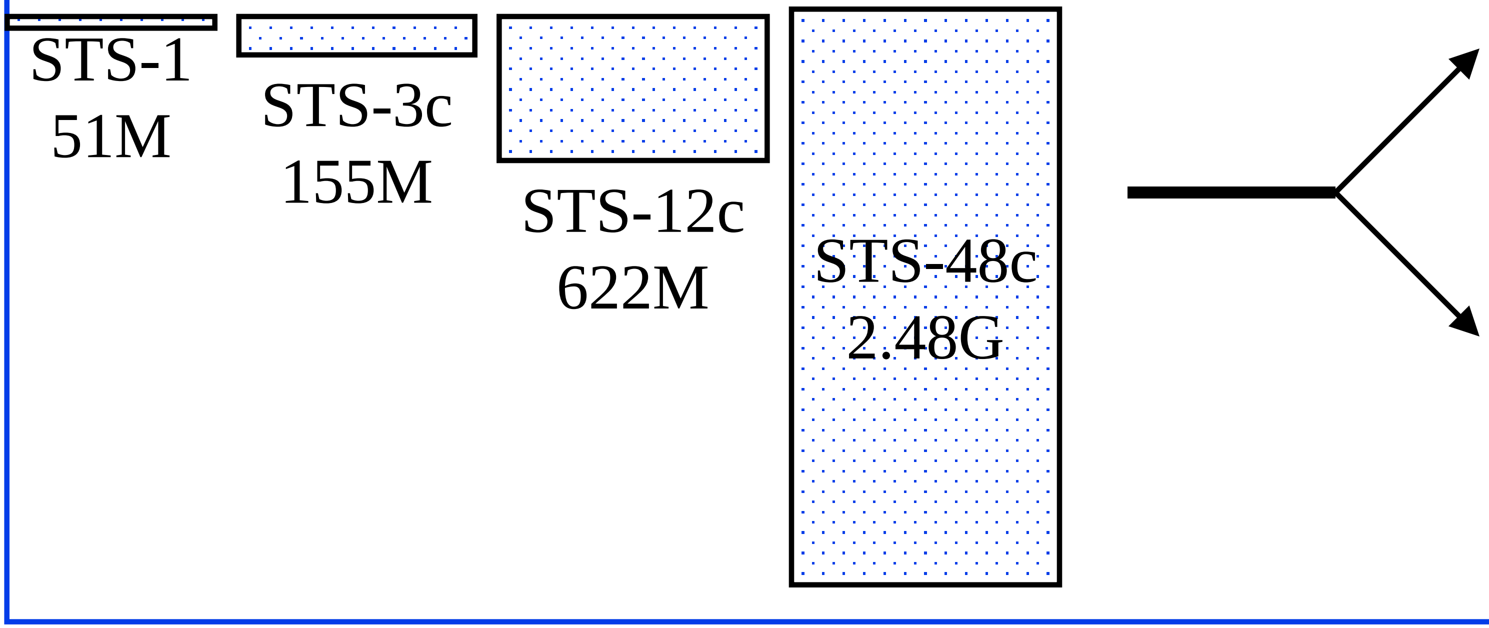
Ethernet over Dark Fiber

Ethernet over SONET

Ethernet over DWDM

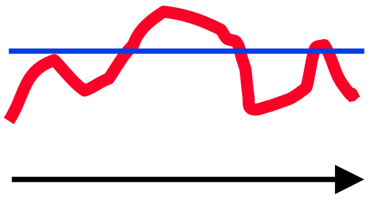
Data over SONET: Problems

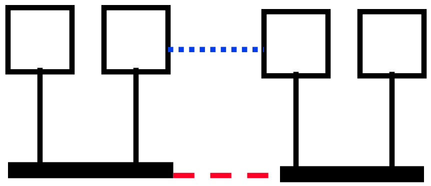
1. Rates highly discrete: In units of STS-3c's. Can't do STS-2c.
2. Entire payload on one path. No splitting, no multipath.
3. Size mismatch: 10 Mbps over 51.84, 100 Mbps over 155 Mbps, 1 Gbps over 1.24 Gbps

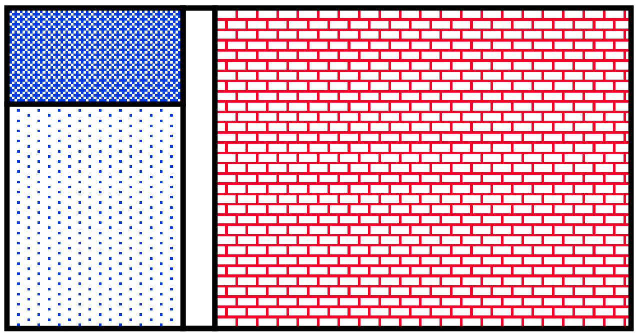


SONET Problems (Cont)

4. Data is bursty (Dynamic). SONET is fixed (static).
5. Inefficient Transparent Connections:
   1 GE = 1.25 Gbps at PHY layer ⇒ Needs OC-48c
6. Only one type of payload per stream: TDM, ATM, FDDI, Packets, Ethernet, Fiber Channel



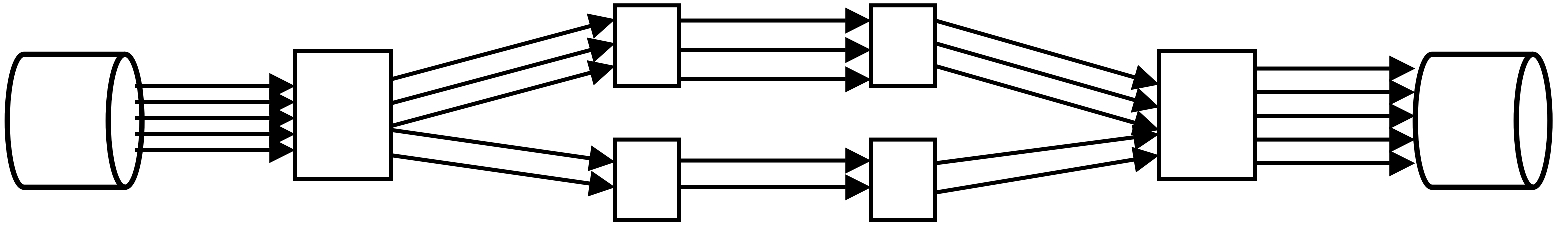




Data over SONET: Solutions

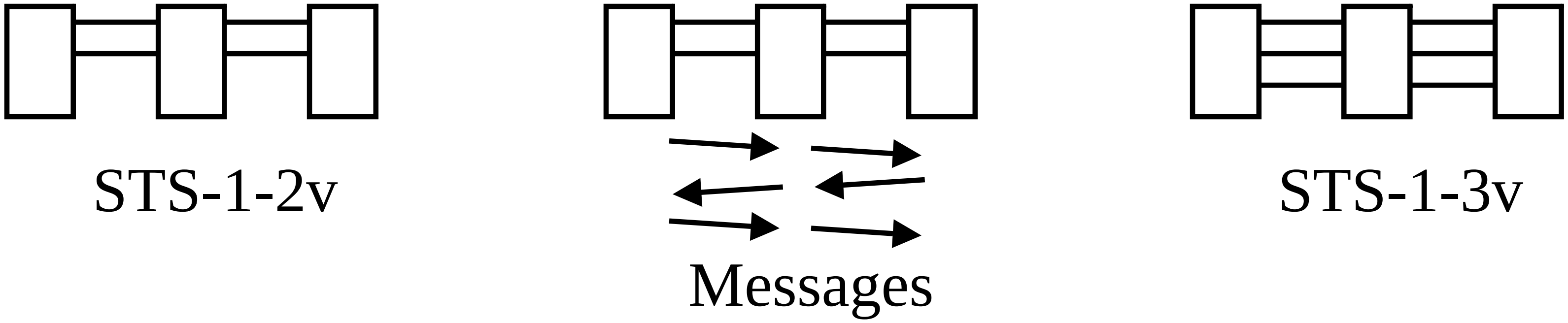
- Virtual Concatenation: n-STS-1's over multiple paths
  1. A channel can be n×STS-1 or nxT1 for any n
  2. Different STS-1's can follow different path
  3. Size match: 10 Mbps over 7 T1,
     100 Mbps over 2 STS-1, 1 Gbps over 21 STS-1
- LCAS: Link Capacity Adjustment Scheme
  4. Can dynamically change number of STS-1's
- GFP: Generic Framing Procedure
  5. Efficient Transparent Connections:
  6. Allows multiple type of payload per stream

SONET Virtual Concatenation



- VCAT: Bandwidth in increments of VT1.5 or STS-1
- For example: 10 Mbps Ethernet in 7 T1's = VT1.5-7v
  100 Mbps Ethernet in 2 OC-1 = STS-1-2v,
  1GE in 7 STS-3c = STS-3c-7v
- The concatenated channels can travel different paths
  ⇒ Need buffering at the ends to equalize delay
- All channels are administered together.
  Common processing only at end-points.

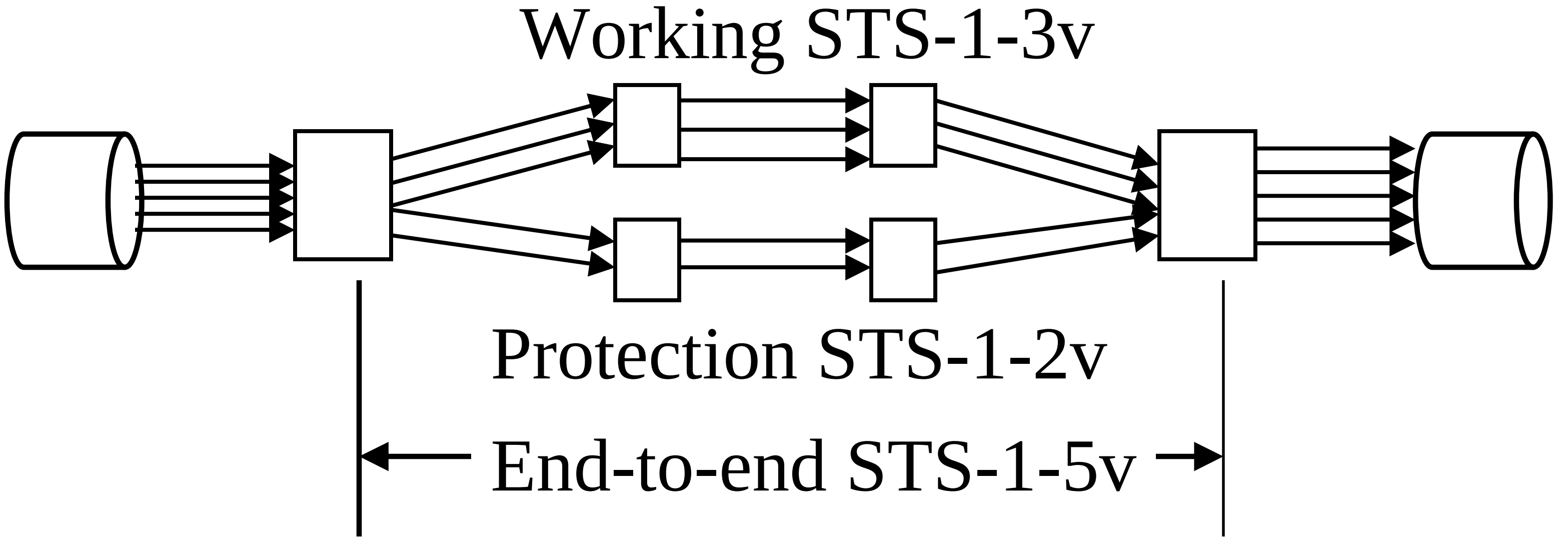
SONET LCAS



- Link Capacity Adjustment Scheme for Virtual Concatenation
- Allows hitless addition or deletion of channels from virtually concatenated SONET/SDH connections
- Control messages are exchanged between end-points to accomplish the change

LCAS (Cont)

- Provides enhanced reliability. If some channels fail, the remaining channels can be recombined to produce a lower speed stream



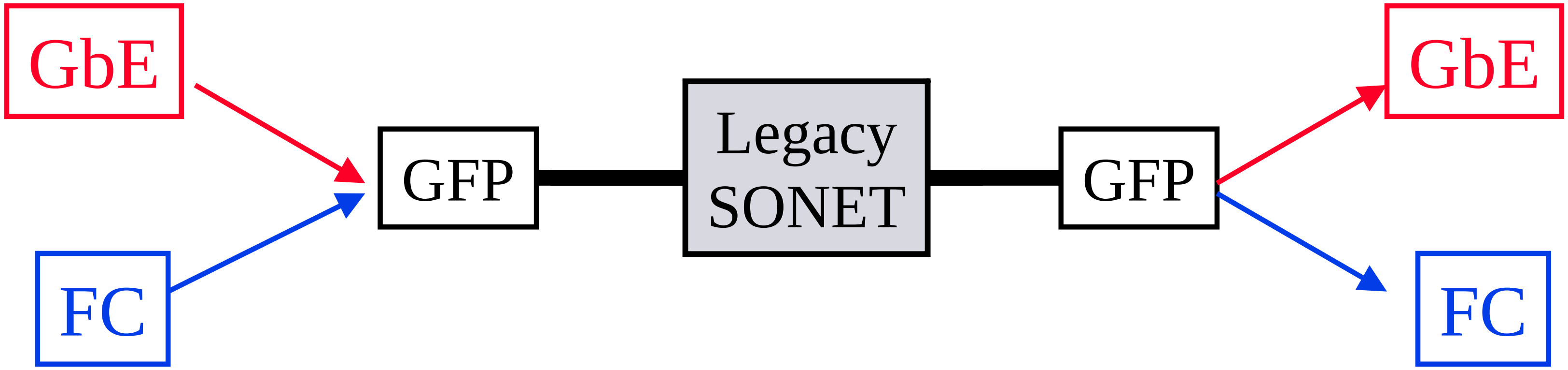
Working STS-1-3v

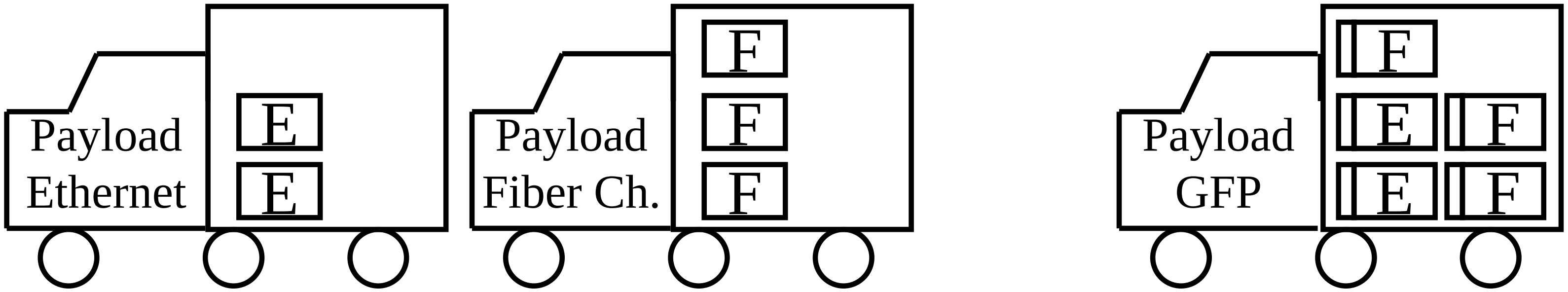
Protection STS-1-2v

← End-to-end STS-1-5v →

Generic Framing Procedure (GFP)

❑ Allows multiple payload types to be aggregated in one SONET path and delivered separately at destination

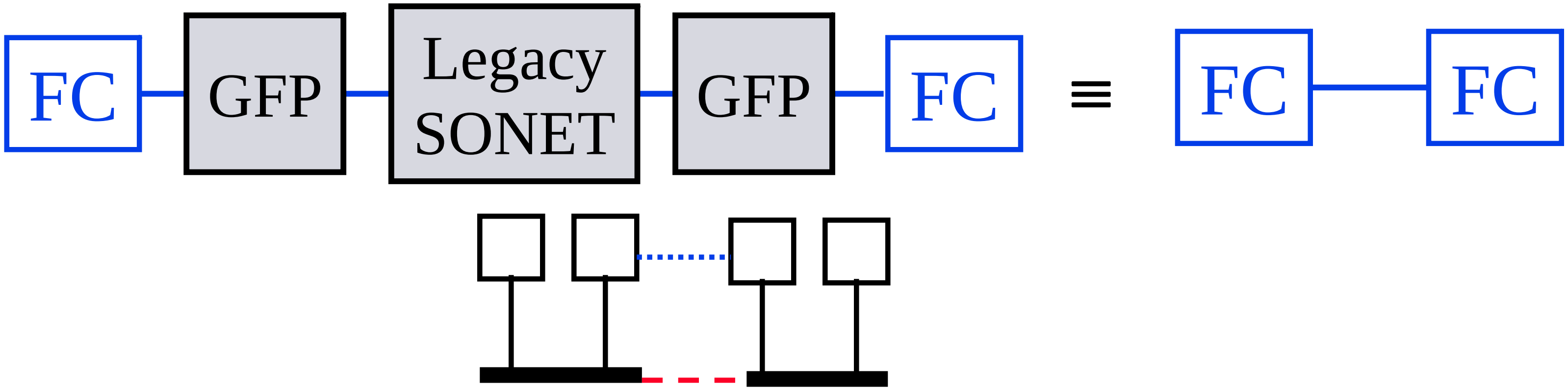




Legacy SONET/SDH   NextGen SONET/SDH

Transparent GFP

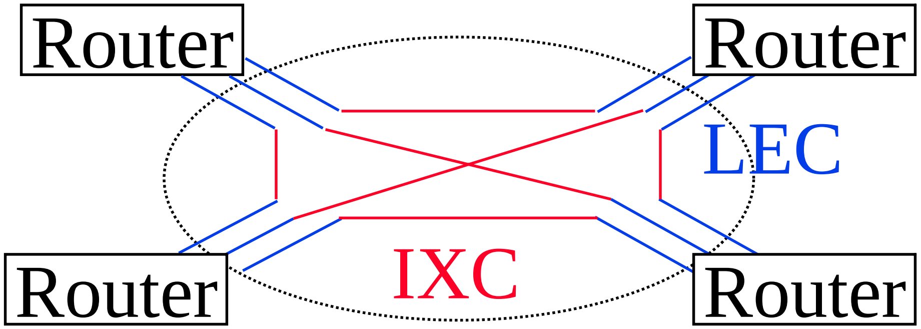
- Allows LAN/SAN PHY extension over SONET links Control codes carried as if it were a dark fiber.



- Problem: 8b/10b results in 1.25 Gb stream for 1 GbE
- Solution: Compress 80 PHY bits to 65 bits
  ⇒ 1.02 Gbps SONET payload per GbE

Problems with Leased Lines

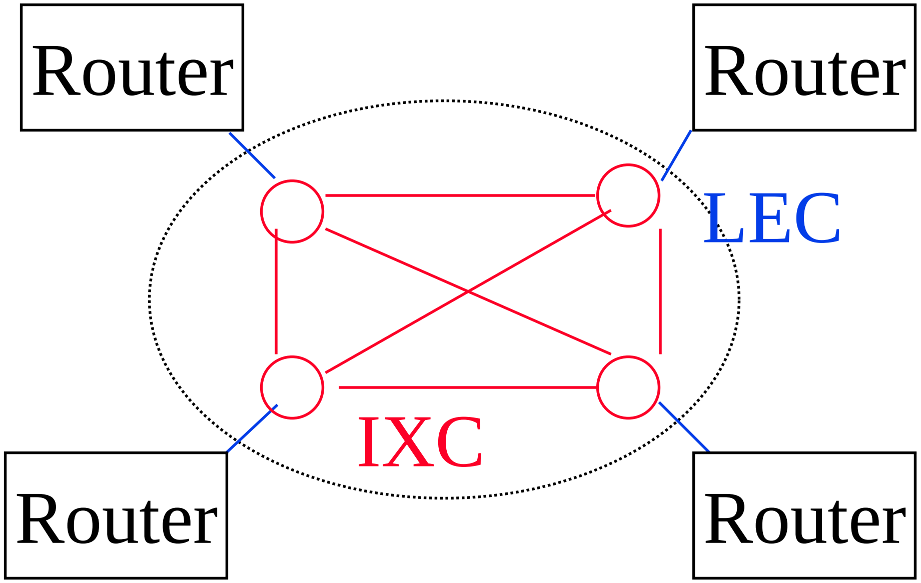
- Multiple logical links ⇒ Multiple connections
- Four nodes ⇒ 12 ports,
  12 local exchange carrier (LEC) access lines,
  6 inter-exchange carrier (IXC) connections
- One more node ⇒ 8 more ports, 8 more LEC lines, 4 more IXC circuits



©2002 Raj Jain

Solution: Frame Relay

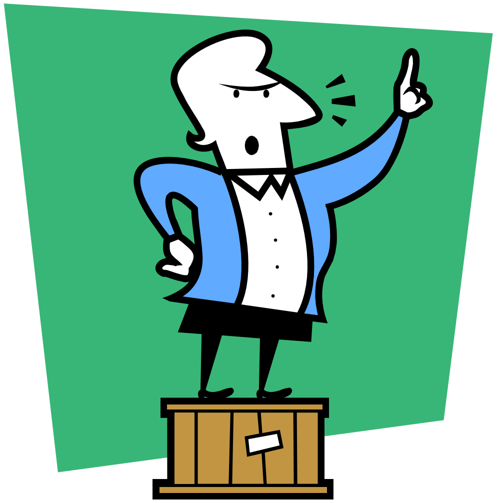
- Four nodes: 4 ports, 4 LEC access lines, 6 IXC circuits
- One more node: 1 more port, 1 more access line, 4 more IXC circuits



Frame Relay: Key Features

- X.25 simplified
  No flow and error control
- Out-of-band signaling
- Congestion control added
  ⇒ Higher speed possible.
  X.25 suitable to 200 kbps. Frame relay to 2.048 Mbps.
- Allows bursting:
  Committed Information Rate,
  Committed Burst Size and Excess Burst Size
  Extra frames are marked "Discard Eligible"

Summary



- Gigabit Ethernet runs at 1000 Mbps
- 10 GbE for full duplex LAN and WAN links
- 1000 Mbps and 9,584.640 Mbps
- RPR will make it more suitable for Metro

Summary (Cont)

- Virtual concatenation allows a carrier to use any arbitrary number of STS-1's or T1's for a given connection. These STS-1's can take different paths.
- LCAS allows the number of STS-1's to be dynamically changed
- Frame-based GFP allows multiple packet types to share a connection
- Transparent GFP allows 8b/10 coded LANs/SANs to use PHY layer connectivity at lower bandwidth.

Homework 5

True or False?

T F

❑ ❑ Full-duplex Ethernet devices do not use CSMA/CD.

❑ ❑ Gigabit Ethernet standard covers metropolitan distances.

❑ ❑ Gigabit Ethernet uses CSMA/CD

❑ ❑ 10 G Ethernet uses CSMA/CD.

❑ ❑ 1000BASE-CX and 1000BASE-T use UTP-5.

❑ ❑ 10GBASE-LW4 uses 4 wavelengths in 1310nm band for WAN distances.

❑ ❑ Link aggregation allows multiple links to be combined for reliability.

❑ ❑ Next Generation of Ethernet is expected to be 100 Gbps.

❑ ❑ RPR provides 1+1 protection

❑ ❑ Source steering consists of sources selecting the ringlet for transmission.

❑ ❑ Virtual Concatenation allows multiple types of payloads to share a SONET connection.

❑ ❑ LCAS allows the data rate of SONET connections to be changed on demand.

Marks = Correct Answers _____ - Incorrect Answers_____ = ______