The road to optical networking

www.science.uva.nl/~delaat

www.science.uva.nl/research/air

Cees de Laat

University of Amsterdam

With an intermezzo of

Erik Radius

SURFnet

λ λ λ λ λ λ λ λ λ λ λ λ λ λ λ λ λ λ λ λ λ λ λ λ λ

Programme

- Why optical networking and IP
- Reference models
- Standardization bodies
- Physical layer
- ITU signaling
- IP addressing, Networking Layer
- IP-optical protocols
- Open issues, current work

So, what's up doc

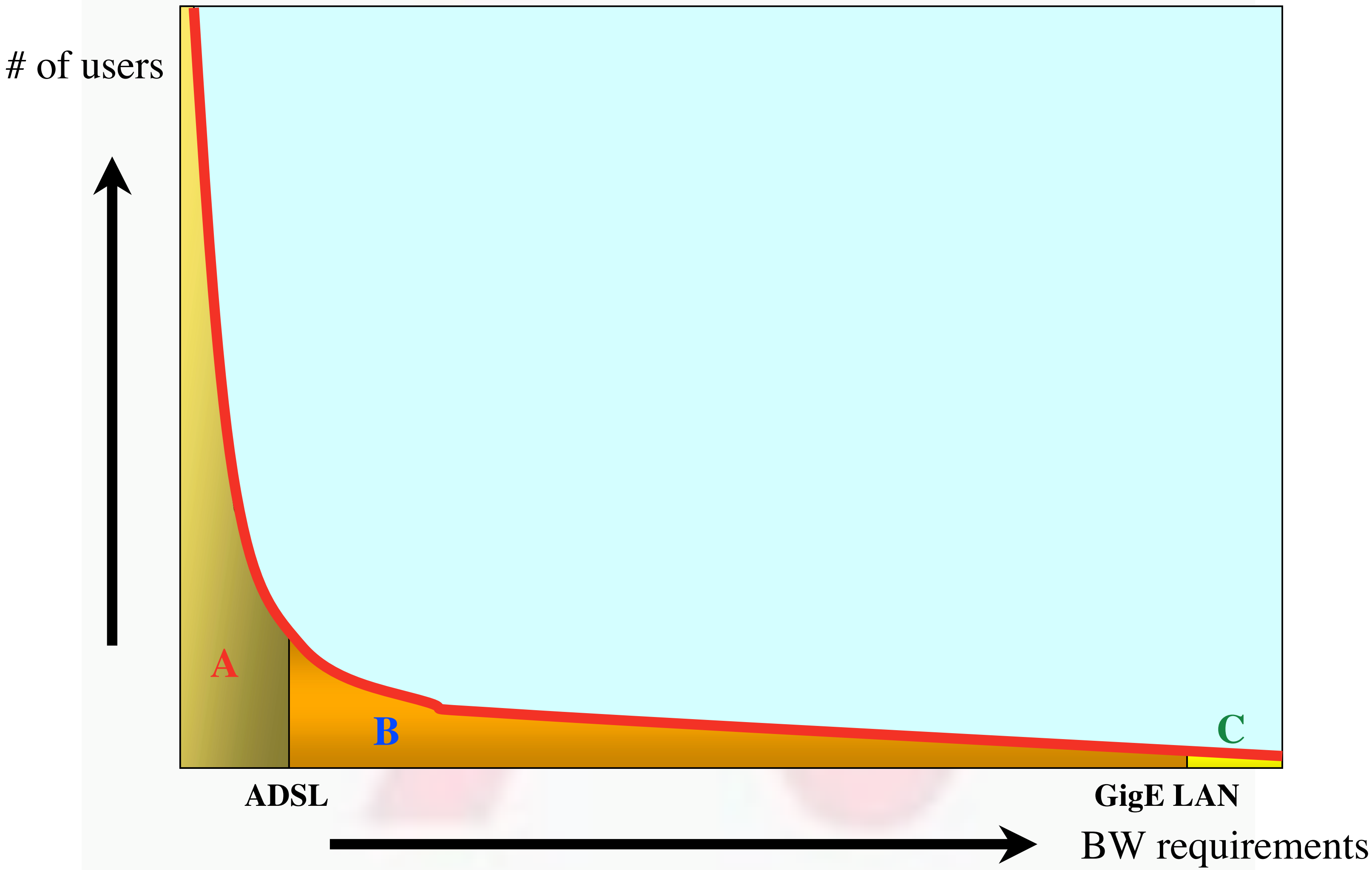
Suppose:

- **Optical components get cheaper and cheaper**
- **Dark (well, dark?) fibers abundant**
- **Number of available λ/user -> ∞**
- **Speeds of 10, 100, 1000 Gbit/s make electrical domain packet handling physically difficult**

Then:

- **λ provisioning for grid applications becomes feasible**
- **Long term view ---> full optical**

Know the user



A -> Lightweight users, browsing, mailing, home use

B -> Business applications, multicast, streaming

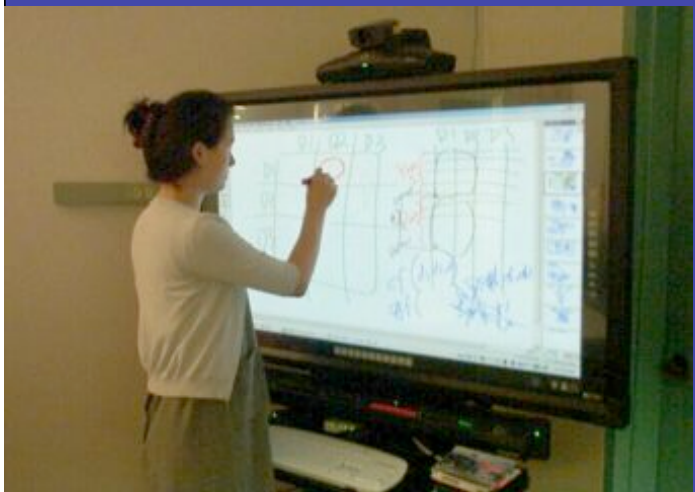
C -> Special scientific applications, computing, data grids, virtual-presence

Integrating Distributed Collaborative Visualization

CAVE

AccessGrid

ImmersaDesk

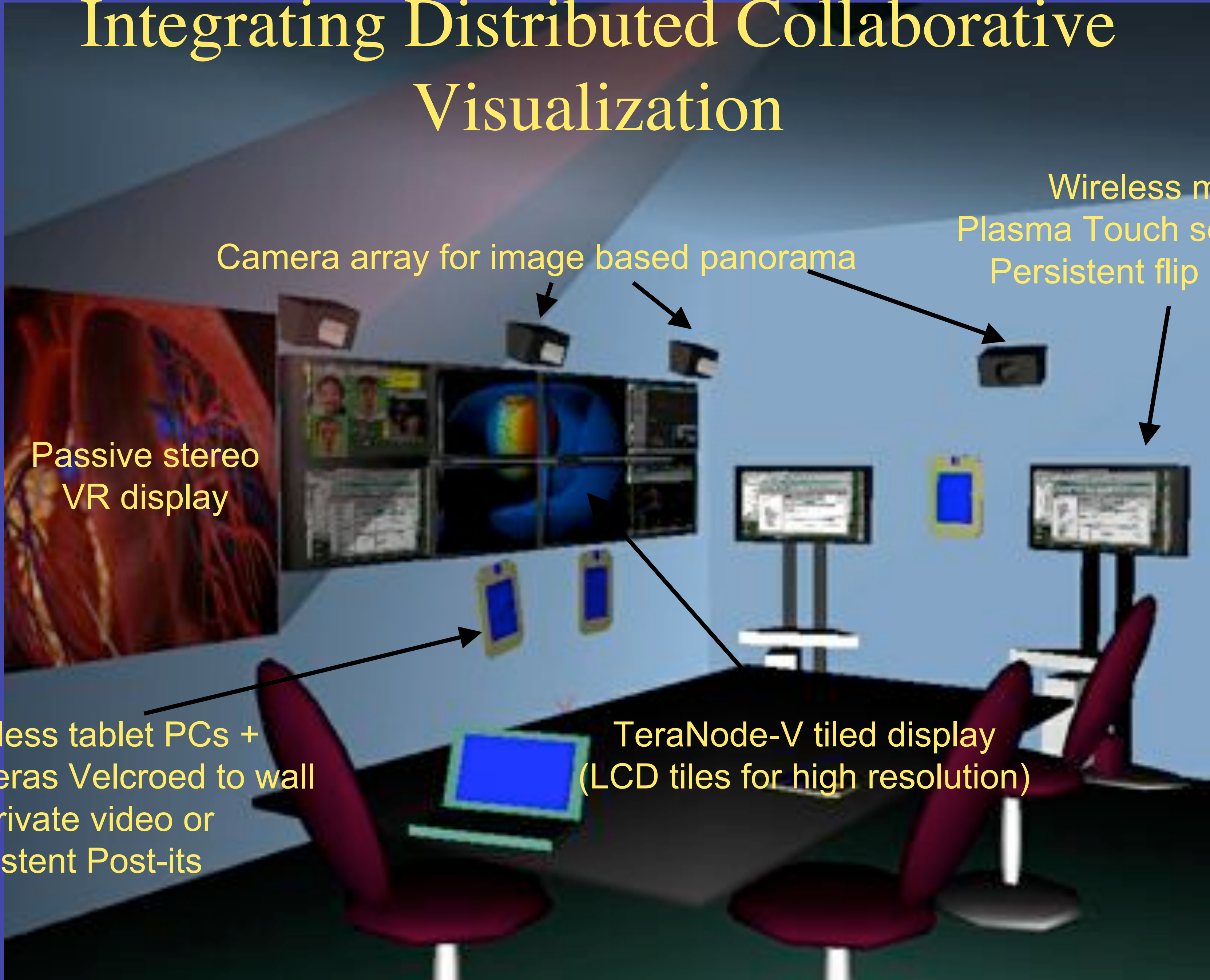


Plasma Touch Screen

AGAVE: Passive Stereo Wall

PDAs, Tablet PCs, Laptops

Integrating Distributed Collaborative Visualization

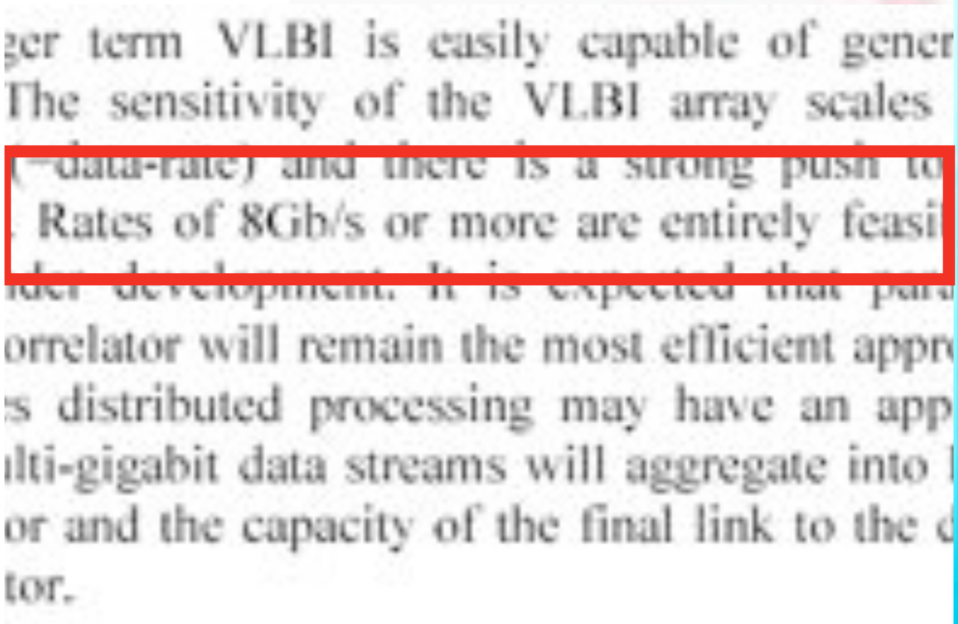


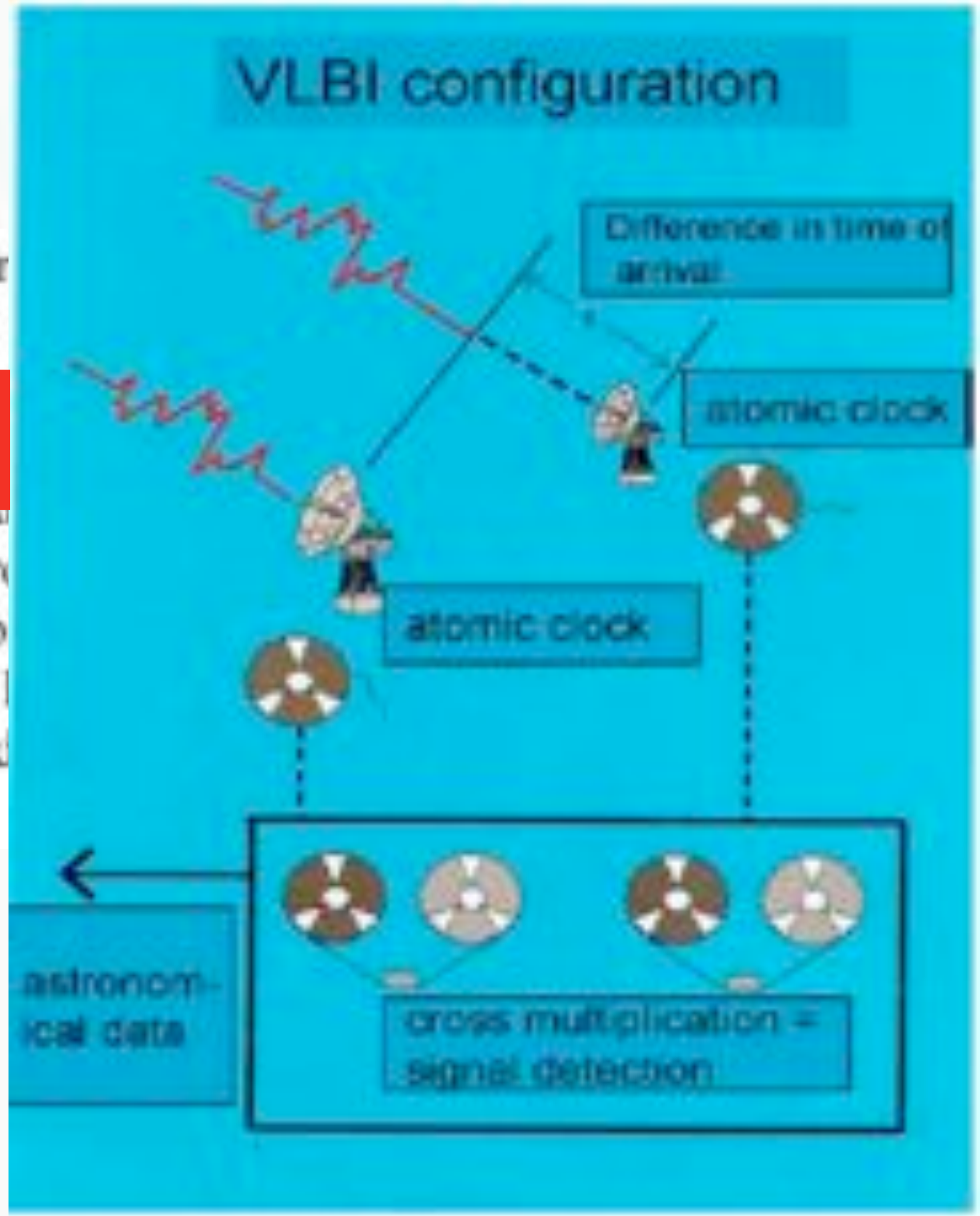
The pilot test described here aims to create a single baseline, real-time, radio interferometer between Jodrell Bank in the UK and Westerbork in the north of The Netherlands. The JIVE data processor in Dwingeloo, close to Westerbork, will be used to correlate the data.

Jodrell Bank Lovell Telescope - UK

Westerbork Synthesis Radio Telescope - Netherlands

VLBI


ger term VLBI is easily capable of gener
The sensitivity of the VLBI array scales
(=data-rate) and there is a strong push to
. Rates of 8Gb/s or more are entirely feasi
der development. It is expected that par
orrelator will remain the most efficient appr
s distributed processing may have an app
lti-gigabit data streams will aggregate into
or and the capacity of the final link to the c
tor.



Why optical networking and IP ?

- Well established IP world of applications
- Provides worldwide addressing scheme, DNS, URL's, Routing, etc.
- Optical networks are supposed to bring speed
- Only Lambda's may look like a telephone system
- How to marry both worlds ?

Standardization bodies

- ISO = International Standards Organisation
  - OSI (Open Systems Interconnect) 7 layer model
- ITU = International Telecommunications Union (www.itu.org)
- OIF - Optical Internet Forum
- IEEE = The Institute of Electrical and Electronics Engineers, Inc. (www.ieee.org)
- IETF = Internet Engineering Task Force (www.ietf.org)
  - ISOC =Internet Society
  - IESG = Internet Engineering Steering Group
  - IAB = Internet Architecture Board
  - IANA = Internet Assigned Numbers Authority -> ICANN
  - ICANN = Internet Corporation for Assigned Names and Numbers
  - IRTF = Internet Research Task Force
  - standards (IETF RFC's), see ftp.ietf.net
  - Internet Protocol (IP, TCP/IP, UDP)

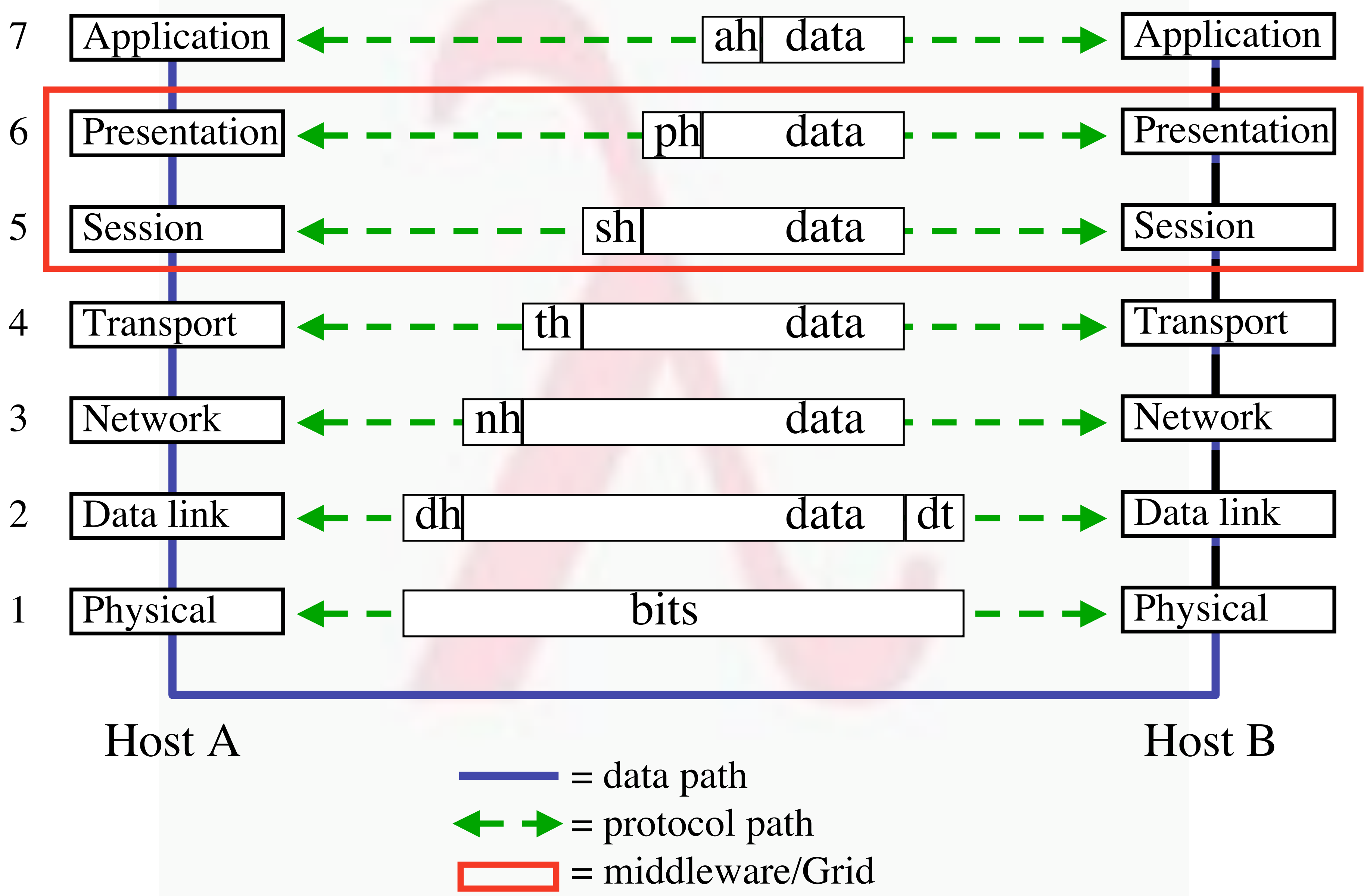
Functionality of Layered models

- Layer model
  - new layer where new level of abstraction is needed
  - each layer does well defined function
  - function of each layer toward international standards
  - layer boundaries chosen to minimize information flow across interfaces
  - number of layers: enough that distinct functions need not be thrown together in one layer out of necessity, and small enough that architecture does not become unwieldy

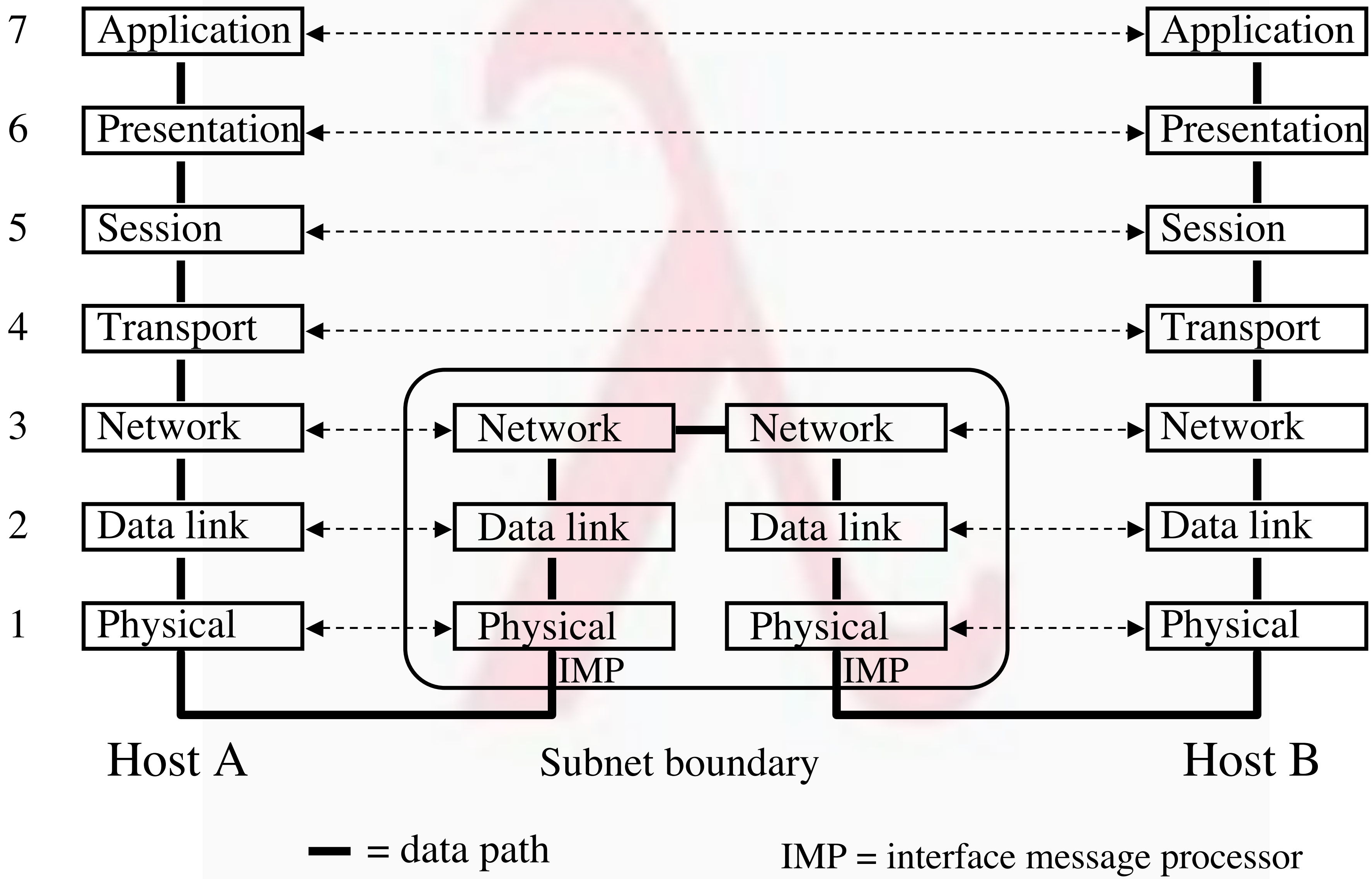
OSI model

- OSI 7 layers
  1 - The Physical Layer
  2 - The Data Link Layer
  3 - The Network Layer
  4 - The Transport Layer
  5 - The Session Layer
  6 - The Presentation Layer
  7 - The Application Layer
- Layers 5 and 6 are almost empty, nowadays usually taken together with the application layer.

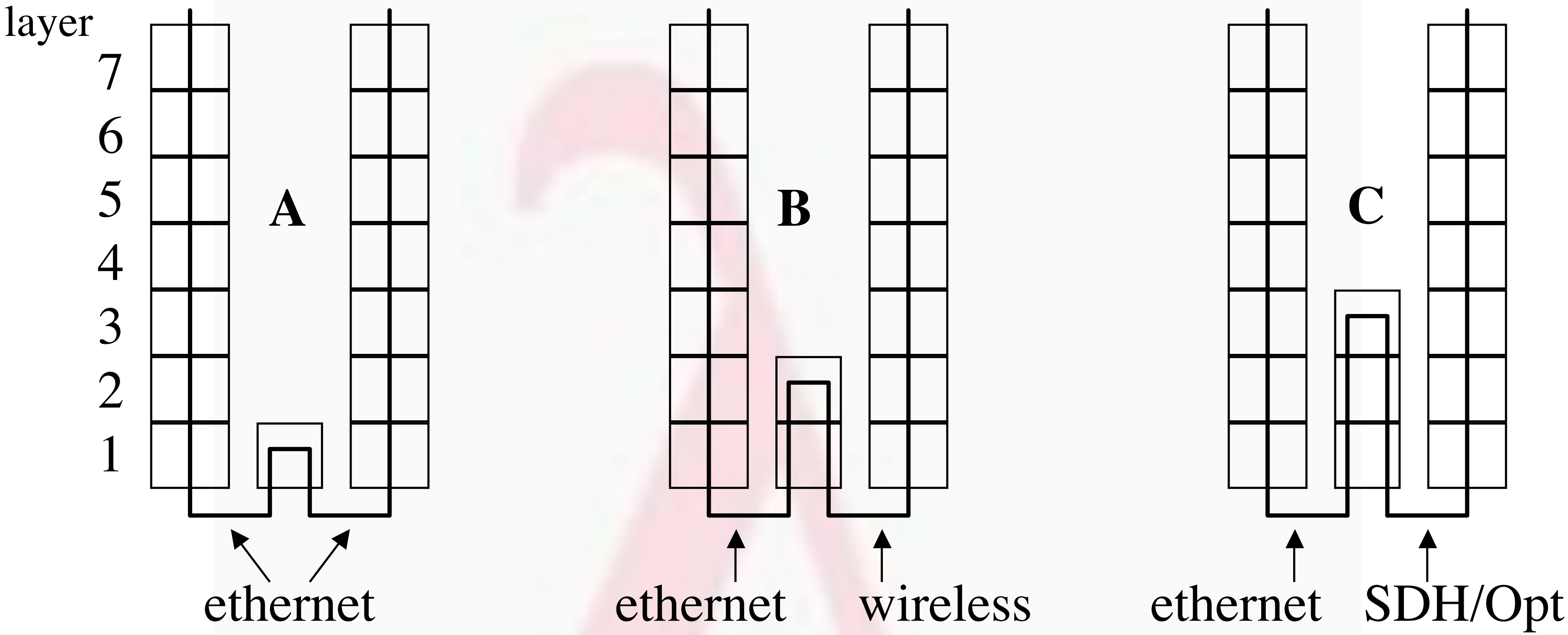
The OSI Reference Model



The OSI Reference Model



Repeater, bridge, switch, router



A: Repeater

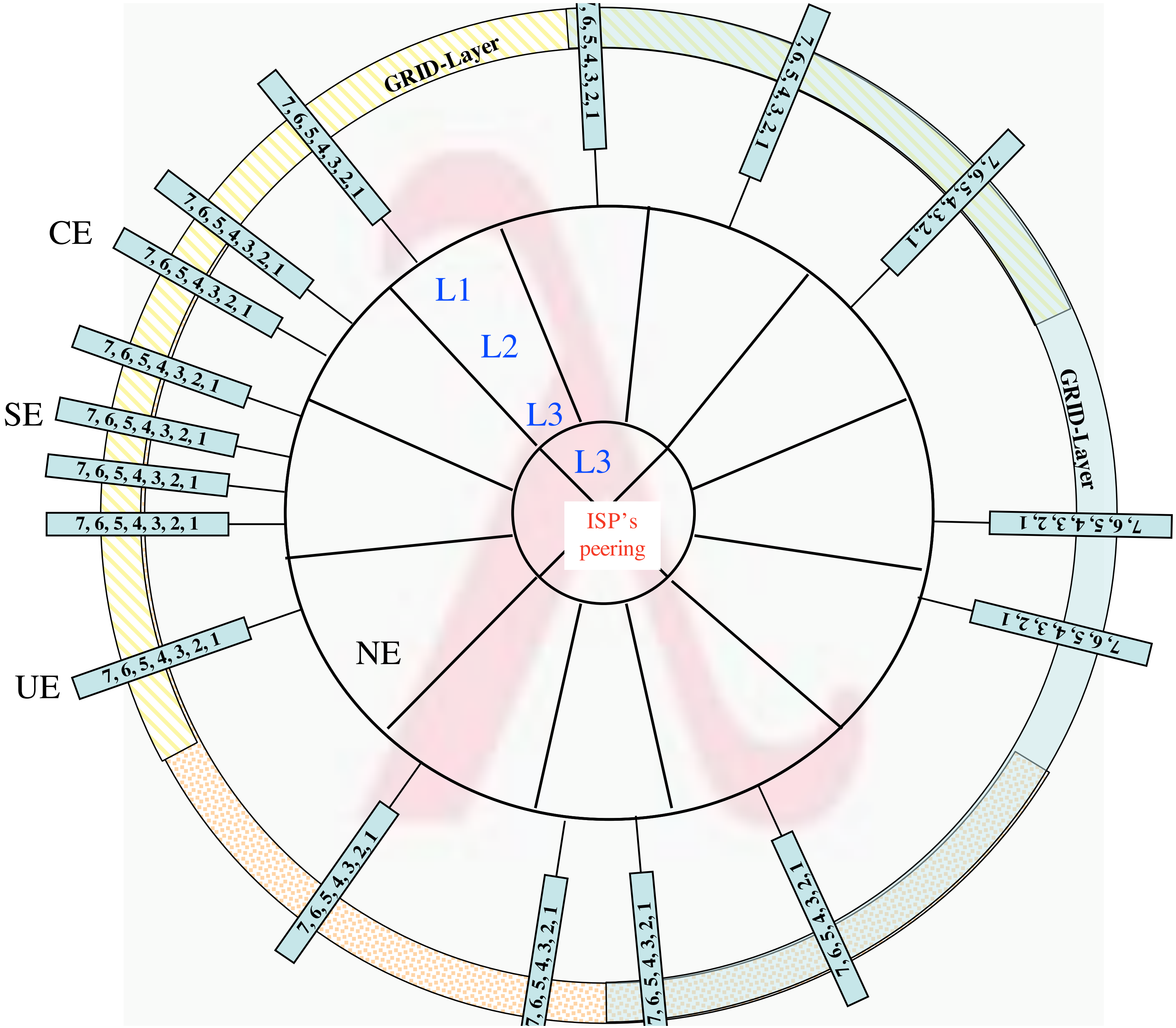
- transfers bits, makes two nets look like one, cable length

B: Bridge, switch

- connect two different data link layers, selective forwarding

C: Router/gateway

- protocol converter, connect network layers, sub-netting, logical map of internet



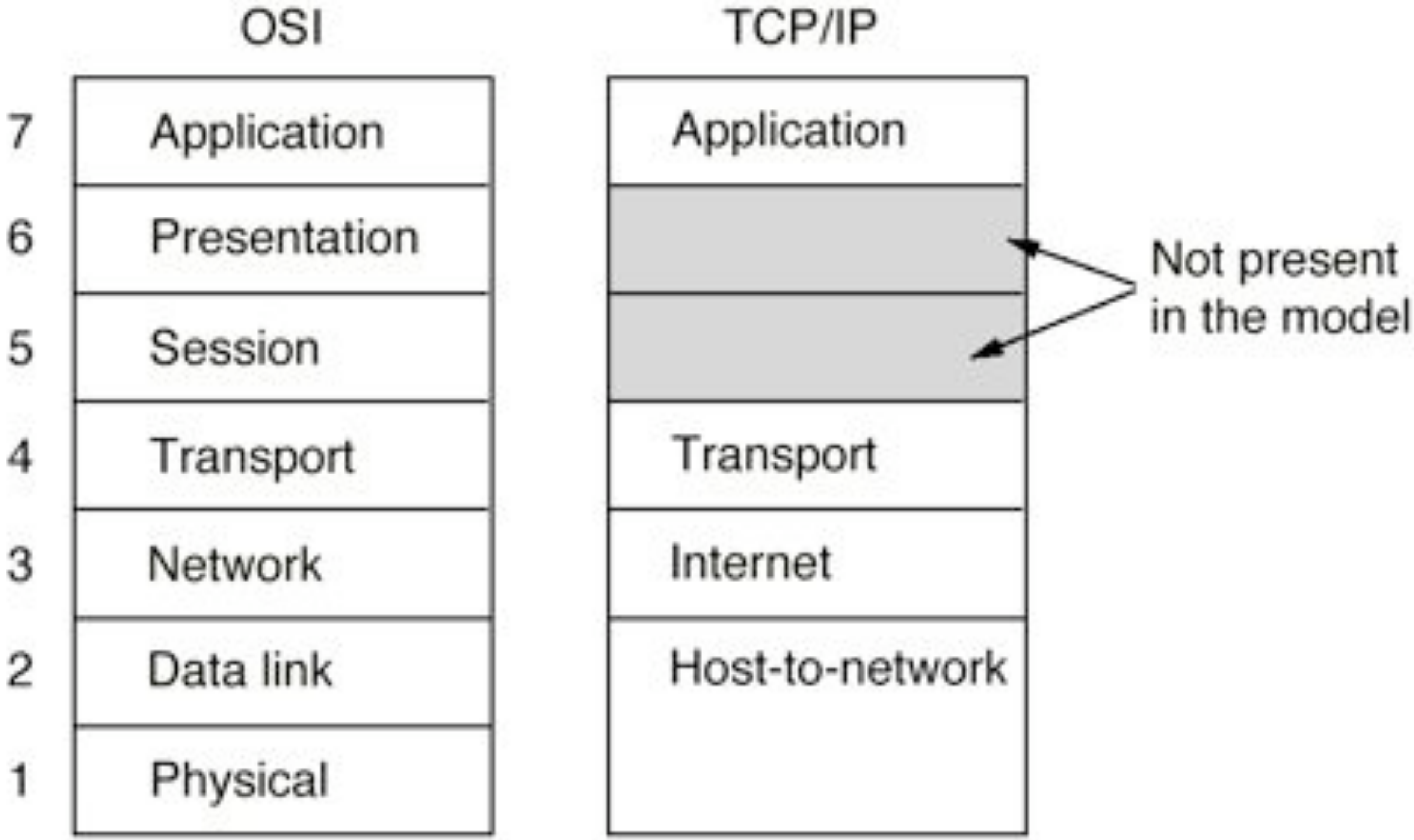
GRID-Layer
GRID-Layer
CE
SE
UE
NE
L1
L2
L3
L3
ISP's peering
7, 6, 5, 4, 3, 2, 1

Connection less versus connection oriented

- Connection less
  - postal office
  - mail
  - internet (IP)
  - datagram delivery
- Connection oriented
  - telephone system
  - 3 phases: establish, use, release
  - order preserved
  - file transfer
  - waste of resources
  - TCP
- role can change in each layer

Example of connection less communication

TCP/IP reference model



DataLink Layer

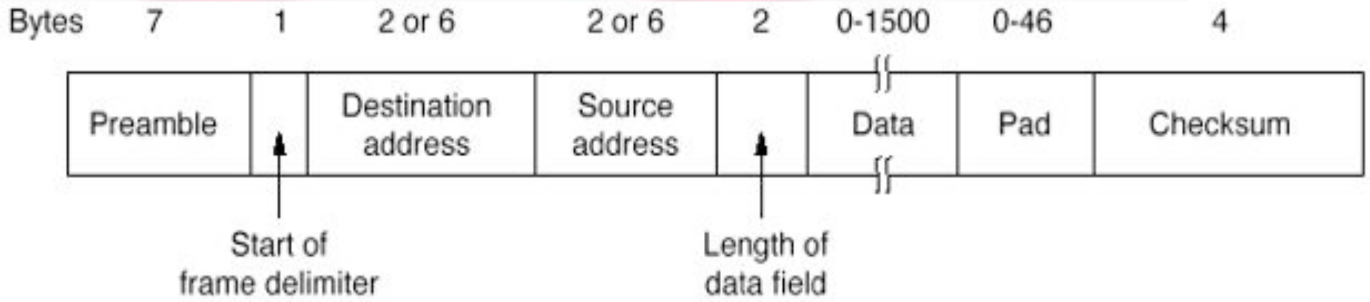
Functions:

- **Point to point (or point to multipoint)**
  - Addressing on media (mac addresses)
- **Framing**
  - Bit streams get structure
- **Error control**
  - Error detection and correction
- **Flow control**
  - fast computers and slow clients

00110100100011100101001010110110001010001010010101101100010

IEEE 802.3 => Ethernet

- 1-persistent CSMA/CD with exponential backoff
- over coax , fiber, utp
- topologies: linear, Spine, Tree, Segmented
- repeaters to amplify signals
- Manchester Encoding: 1 = high-low, 0 = low-high
- 10 Mbit/s -> 2500 meters, 51.2 usec. 64 bytes
- 100 Mbit/s -> 250 meter, 5.12 usec. 64 bytes, IEEE 802.3u
- 1 Gbit/s -> 250 meter, 5.12 usec, 512 bytes, IEEE 802.3z
- 10 Gbit/s -> unlimited length, full duplex only, 64 bytes, IEEE 802.3ae
- 48 bits unique addresses
- high order bit: 0 = ordinary address, 1 = group address, all 1's = broadcast
- next bit: local / global addresses
- Frame format:



Sub-IP Area

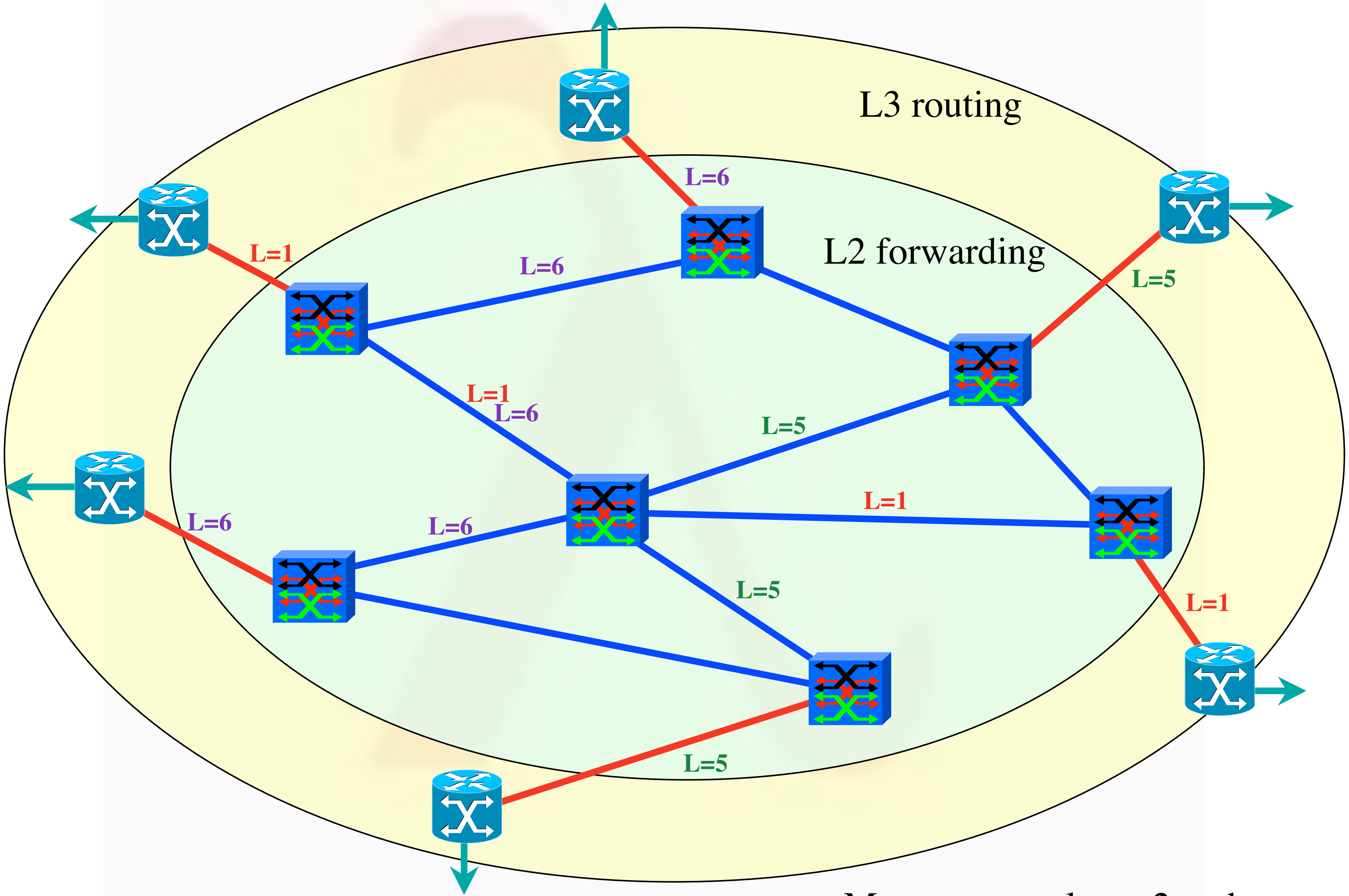
- Area Director(s): Scott Bradner <sob@harvard.edu>
- Bert Wijnen <bwijnen@lucent.com>

- Working Groups:
- ccamp Common Control and Measurement Plane
- gsmp General Switch Management Protocol
- ipo IP over Optical
- iporpr IP over Resilient Packet Rings
- mpls Multiprotocol Label Switching
- ppvpn Provider Provisioned Virtual Private Networks
- tewg Internet Traffic Engineering

Signaling

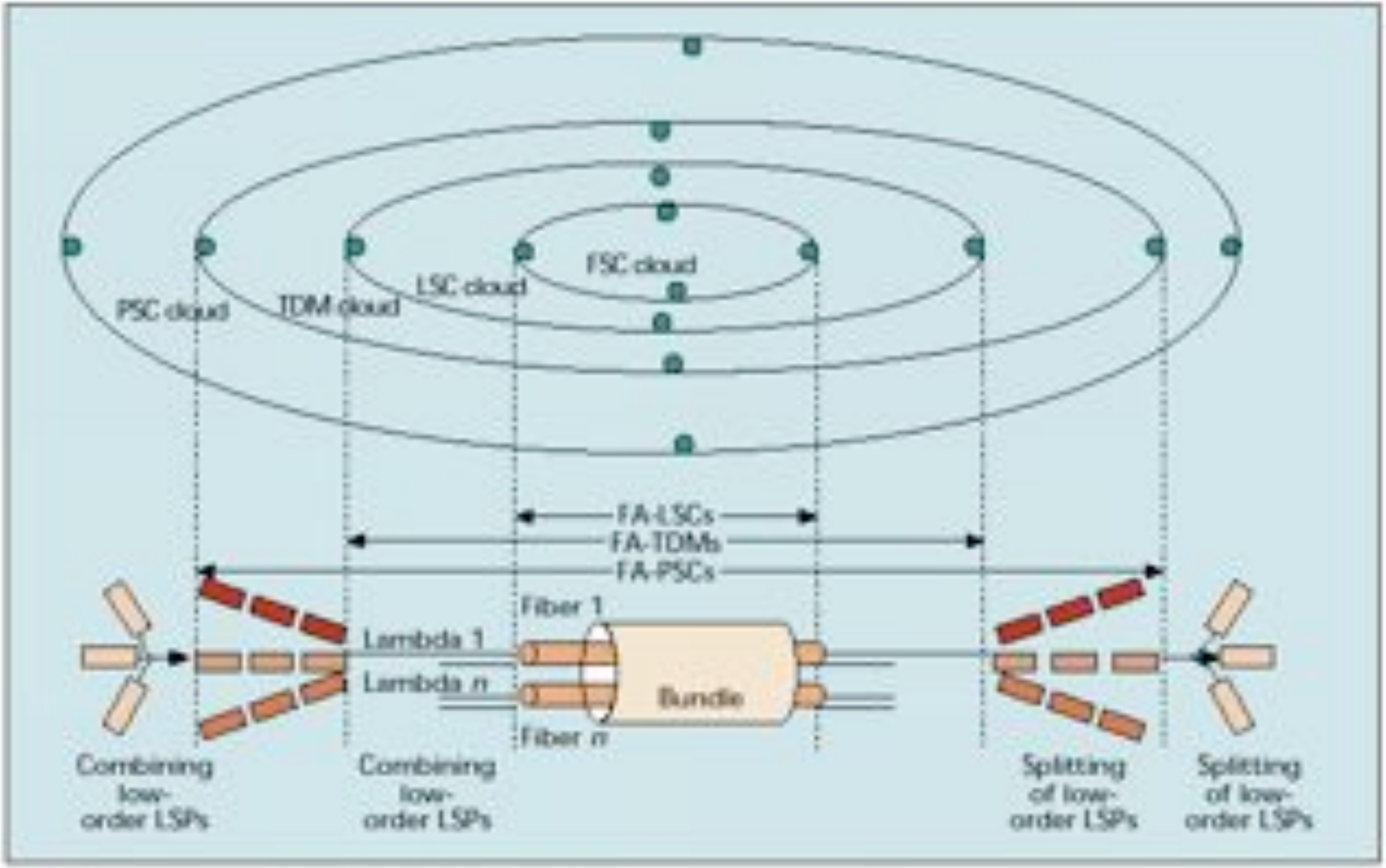
- GMPLS
- OBG
- RSVP
- UNI vs NNI vs peer to peer
- Single versus multi domain

(G)MPLS



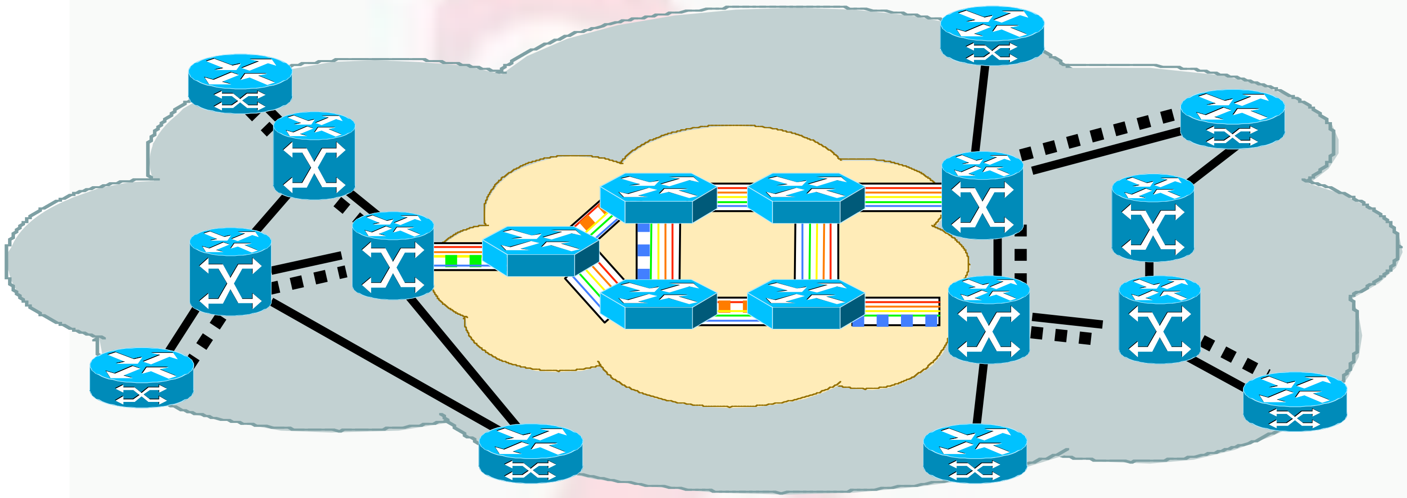
Map routes to layer 2 path
LDP = label distribution protocol

GMPLS LSP Hierarchy



Source: Turner, et al, IEEE Communications, Feb. 2001

Generalized MPLS (GMPLS)



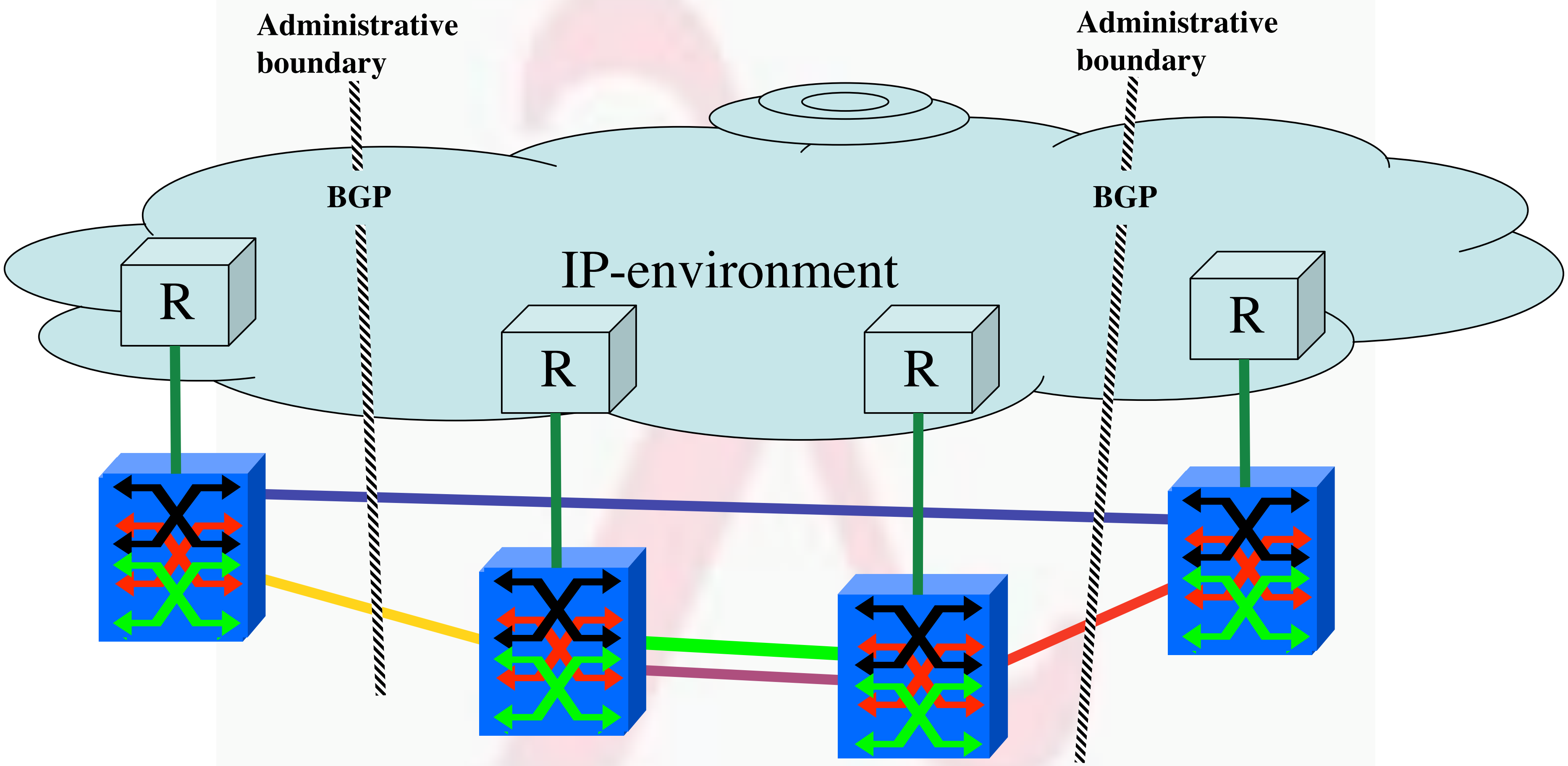
- Reduces the multiple layers into a single, integrated, control layer
- Extends MPLS control plane to address optical layer constraints and attributes
- Leverages IP layer management simplicity and distributed intelligence
- Provides sophisticated traffic engineering capabilities for resource management and control

Control Plane Protocols Standards Summary

Function	MPλS/GMPLS	O-UNI	G.ASON
Routing Protocol	IGP TE extensions	N/A	N/A
Signaling	RSVP/CR-LDP extensions	RSVP/CR-LDP extensions	Out-of-band client UNI
Link Management, verification, neighbor discovery, etc	LMP	LMP	Central Control, IP/ATM/ SONET clients
Model	Peer/Overlay	Overlay to Peer	Overlay
Standards Body	Peer/IETF	OIF	ITU-T

Drafts as of January 2001

Multi domain IP controlled



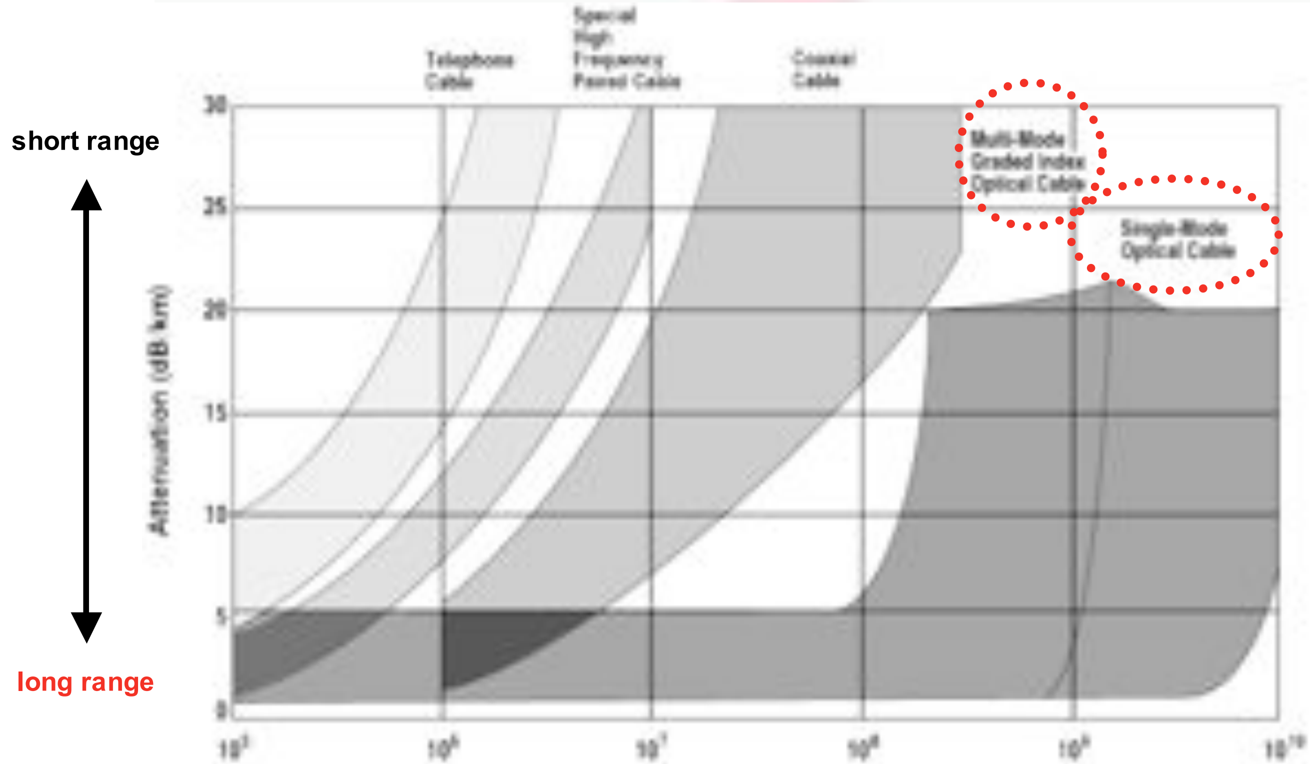
Optical technologies for IP networks

an intermezzo

Erik Radius, SURFnet

Why fibers for data transport?

- Fibers are medium of choice for
  - high bitrate signal transport (1-1000 gigabit/s)
  - over large distances (2 km … trans-Atlantic)



Source: Corning Corp.

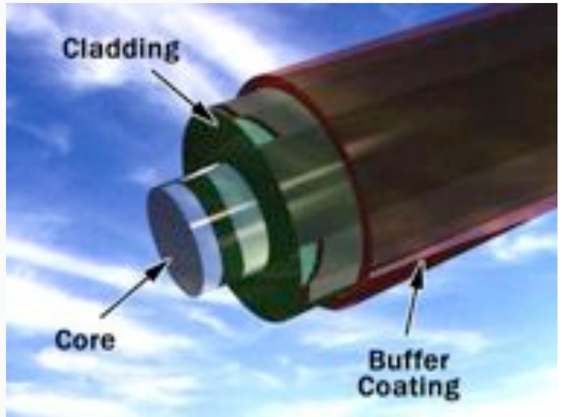
SURFnet

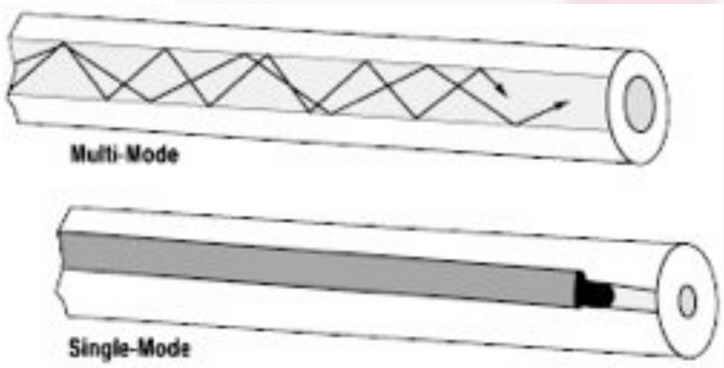
Fiber technology

- Fiber = glass core (9µm) with glass cladding (125µm)

- Low attenuation due to total internal reflection of light

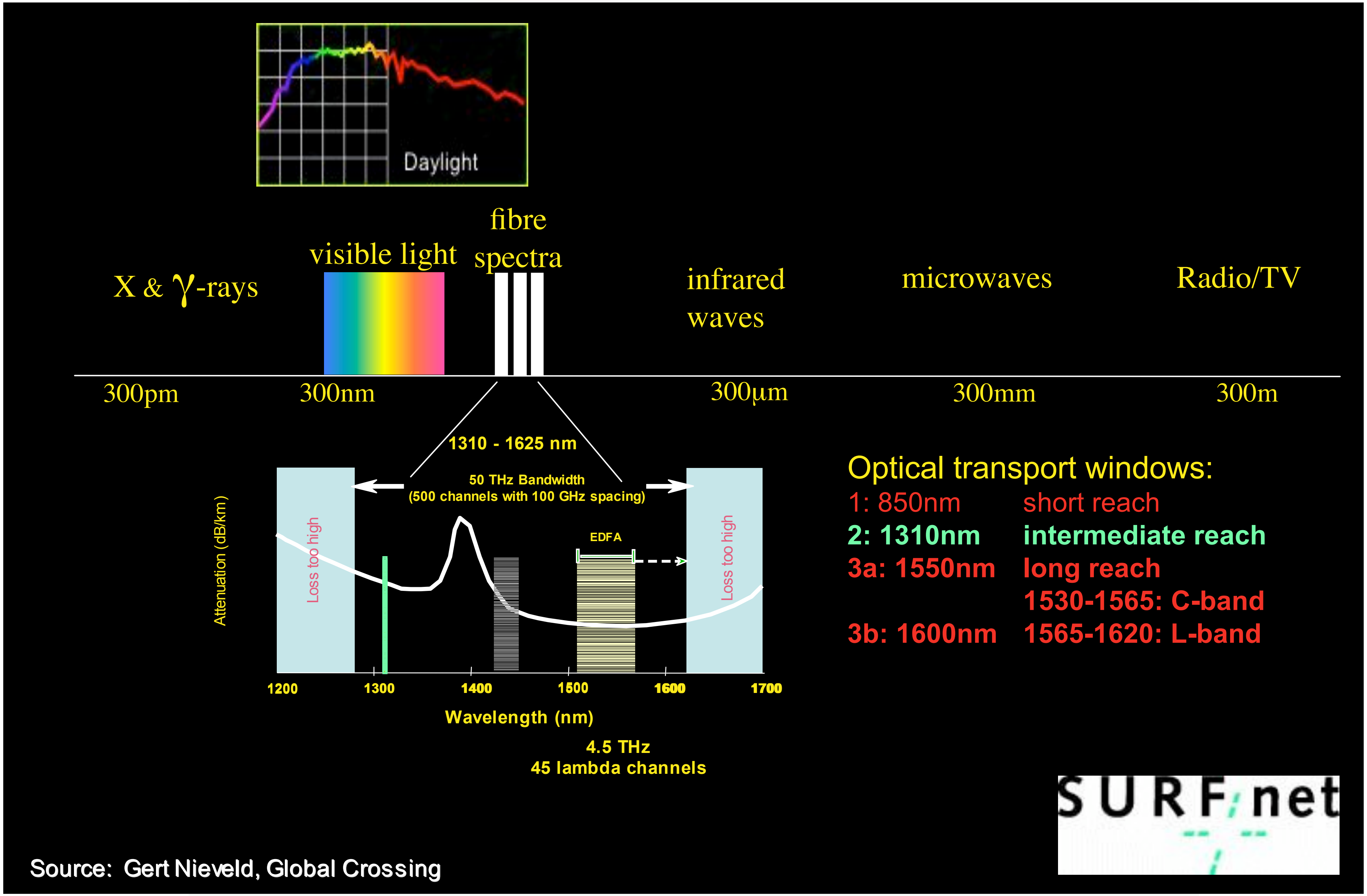
- Fiber types:
  - Multi-Mode
  - Single-Mode





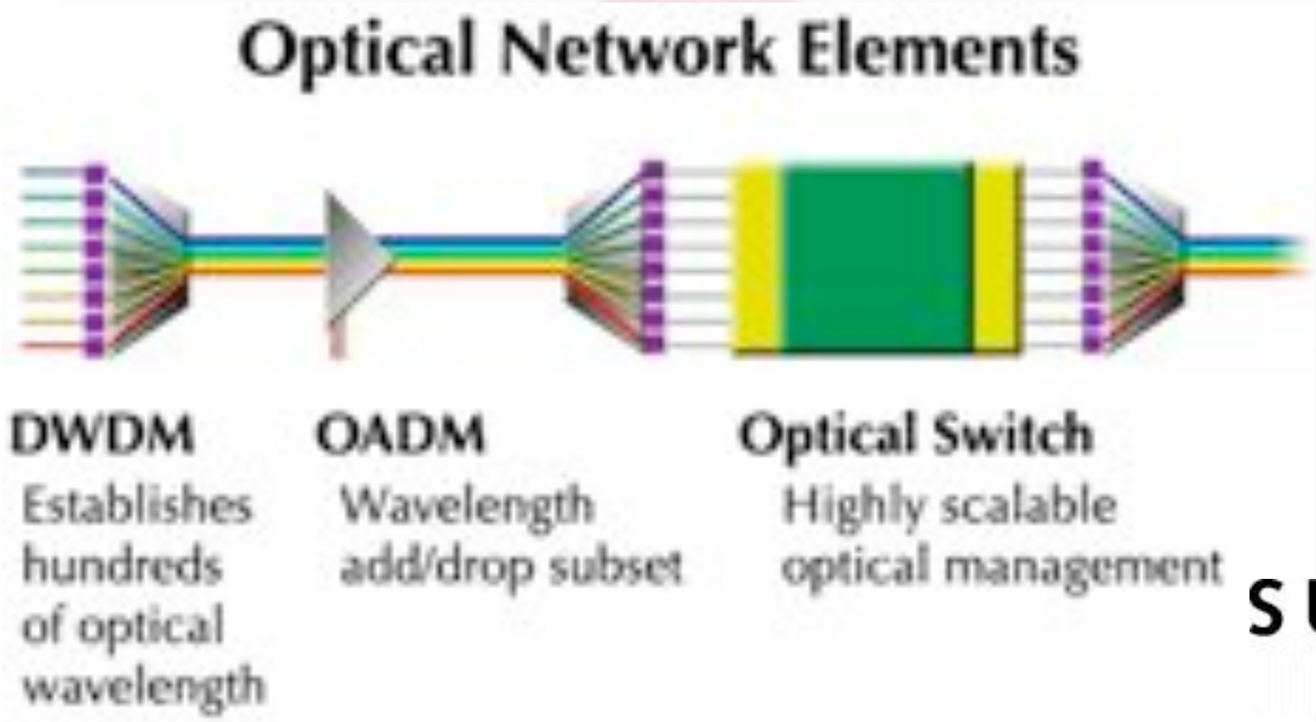
SURFnet

Optical bandwidth: room to grow



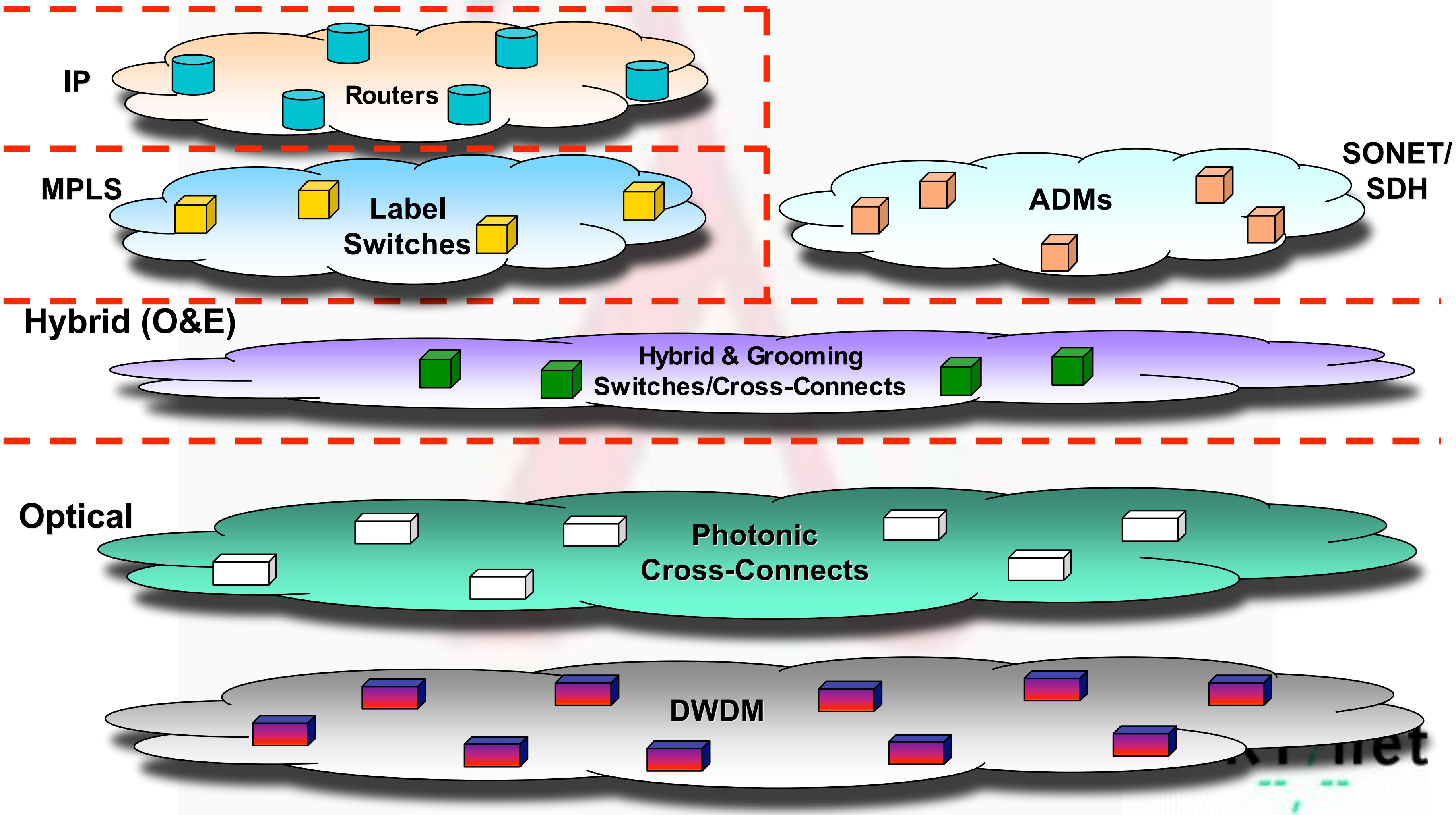
Lambda networking

- WDM: Wavelength Division Multiplexing:
  - multiple colors (*lambdas*) on a single fiber
- OADM: Add/drop traffic in optical domain
- OXC: Optical Cross-Connect

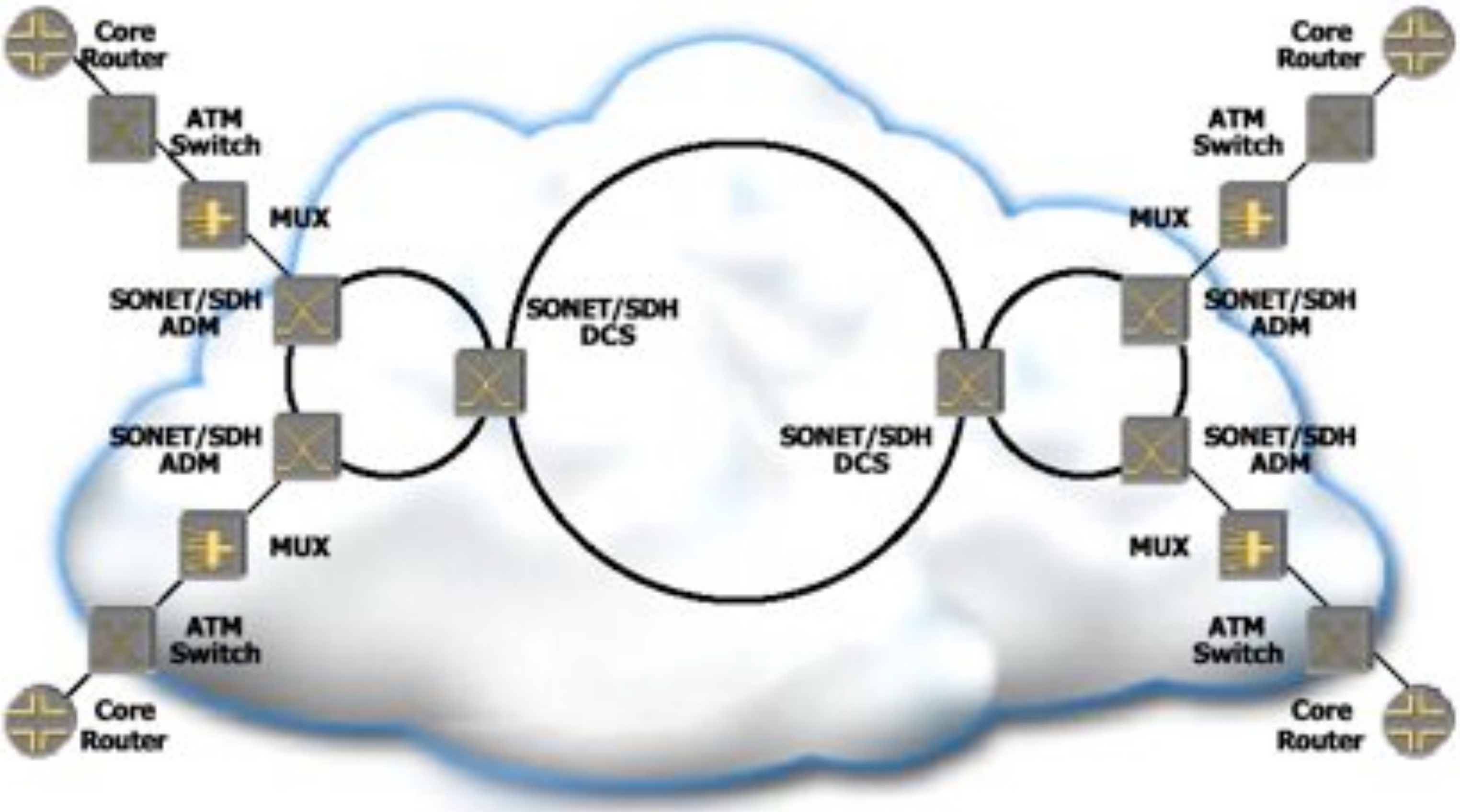


SURF net

Optical networking tomorrow



Typical IP Backbone (Late 1990's)



- **Data piggybacked over traditional voice/TDM transport**

Juniper NETWORKS

Source: Jean-Marc Uzé, Juniper Networks

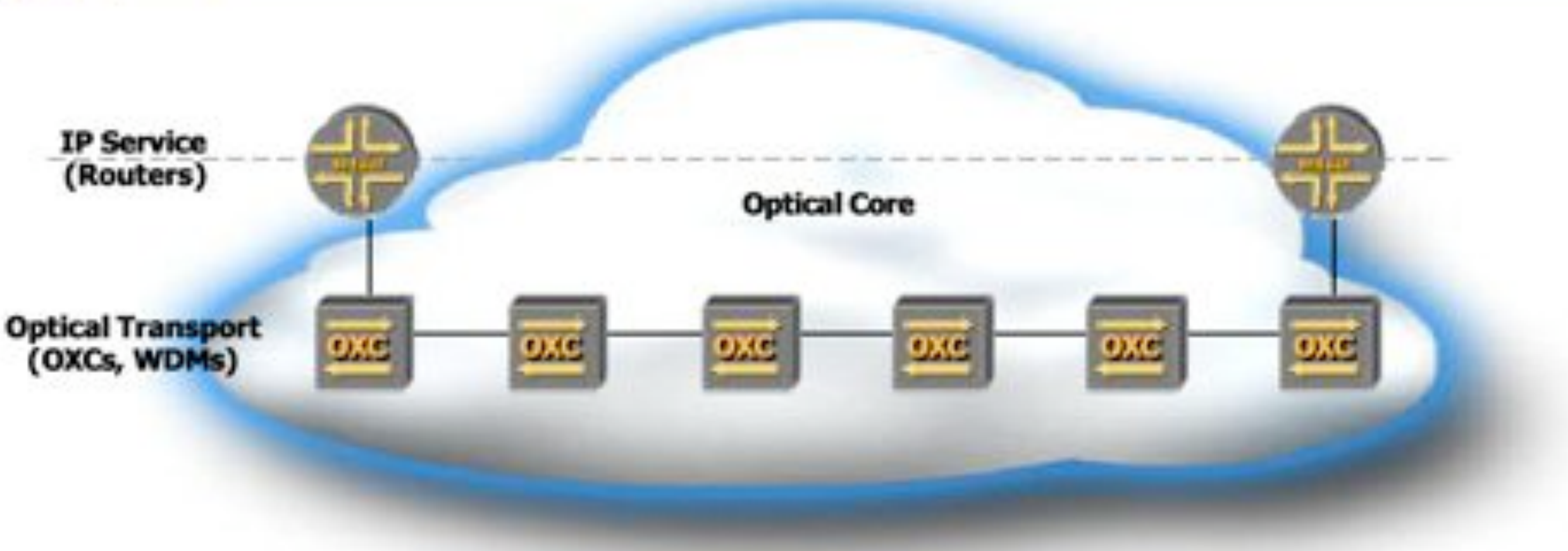
Why So Many Layers?

- **Router**
  - **Packet switching**
  - **Services (Diffserv, filtering, Multicast, VPN...)`**
  - **Statistical multiplexing gain**
  - **Any-to-any connections**
  - **Restoration (several seconds)**
- **ATM/Frame switches**
  - **Hardware forwarding**
  - **Traffic engineering**
  - **Restoration (sub-second)**
- *Result*
  - *More vendor integration*
  - *Multiple NM Systems*
  - *Increased capital and operational costs*
- **MUX**
  - **Speed match router/ switch interfaces to transmission network**
- **SONET/SDH**
  - **Time division multiplexing (TDM)**
  - **Fault isolation**
  - **Restoration (50mSeconds)**
- **DWDM**
  - **Raw bandwidth**
  - **Defer new construction**

Juniper NETWORKS

Source: Jean-Marc Uzé, Juniper Networks

A Two-Layer Network

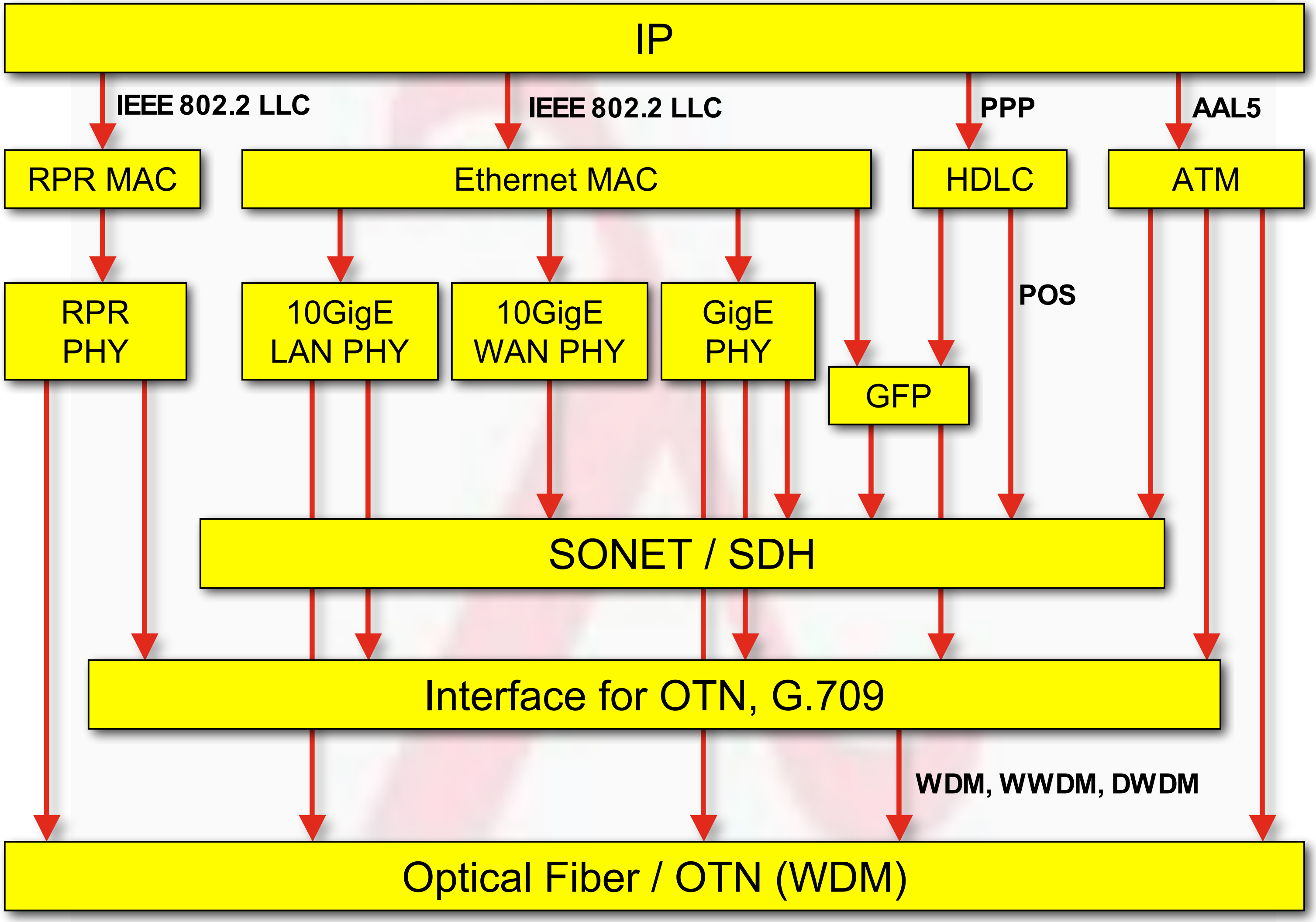


- Reduce capital cost, complexity
- Reduce administrative and operational cost
- Dynamically build optical circuits between routers
- What of functionality of "missing layers"?

Juniper NETWORKS

Source: Jean-Marc Uzé, Juniper Networks

IP networking over Optics



RPR = Resilient Packet Ring, IEEE 802.17

HDLC = High-Level Data-Link Control

POS = Packet over SONET/SDH

GFP = Generic Framing Procedure (ANSI T1 X1-driven standard)

OTN = Optical Transport Network

WDM = Wavelength Division Multiplexing

WWDM = Wide WDM

DWDM = Dense WDM

Optical technologies for IP networks

end of intermezzo

Erik Radius, SURFnet

SURFnet

Optical networking, 3 scenarios

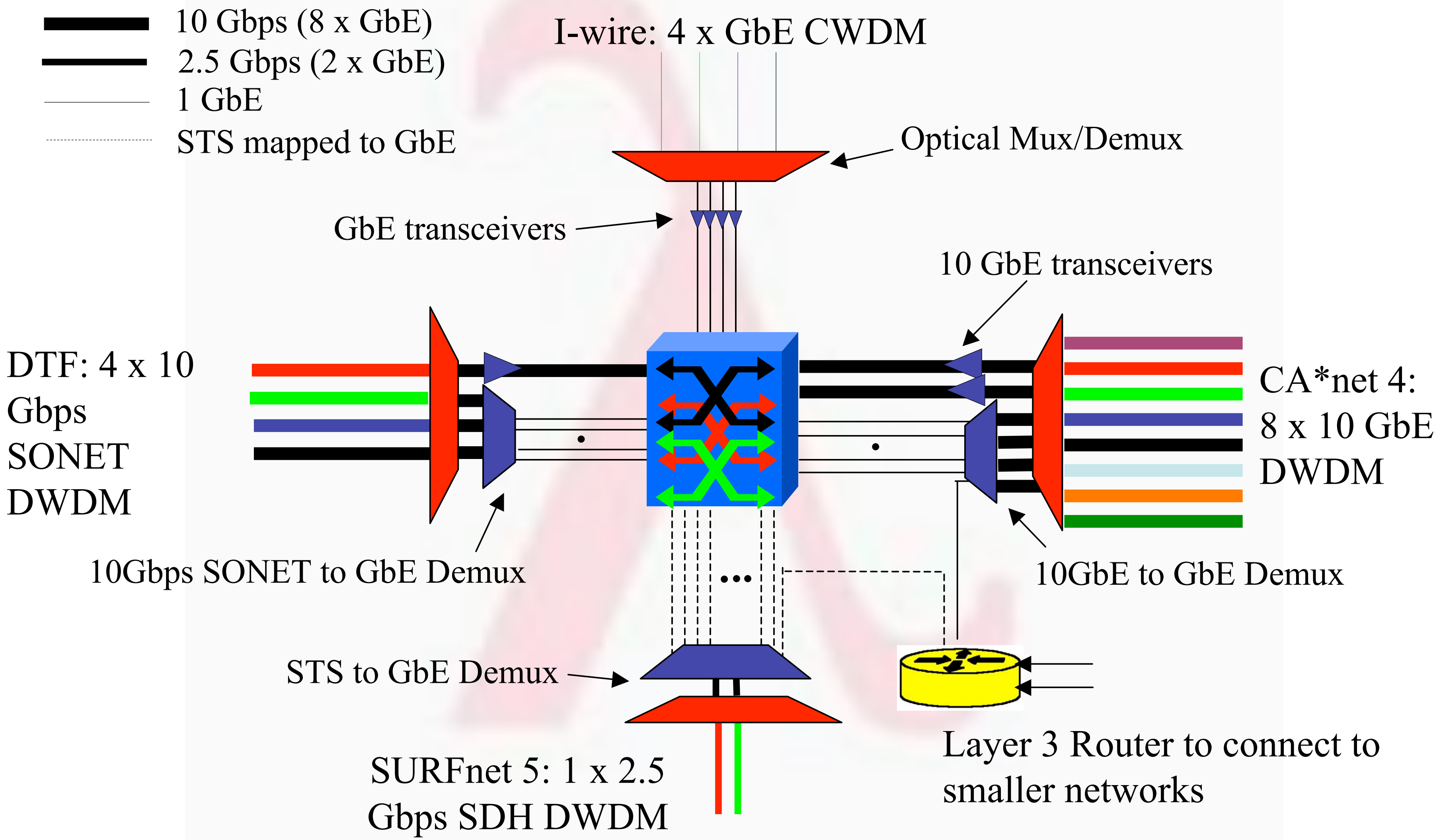
- **Lambdas for internal ISP bandwidth provisioning**
  - **An ISP uses a lambda switching network to make better use of its (suppliers) dark fibers and to provision to the POP's. In this case the optical network is just within one domain and as such is a relatively simple case.**
- **Lambda switching as peering point technology**
  - **In this use case a layer 1 Internet exchange is build. ISP's peer by instantiating lambdas to each other. Is a N*(N-1) and multi domain management problem.**
- **Lambda switching as grid application bandwidth provisioning**
  - **This is by far the most difficult since it needs UNI and NNI protocols to provision the optical paths through different domains.**

Current technology + (re)definition

- Current (to me) available technology consists of SONET/SDH switches
- DWDM+switching coming up
- Starlight uses for the time being VLAN's on Ethernet switches to connect [exactly] two ports
- So redefine a λ as:

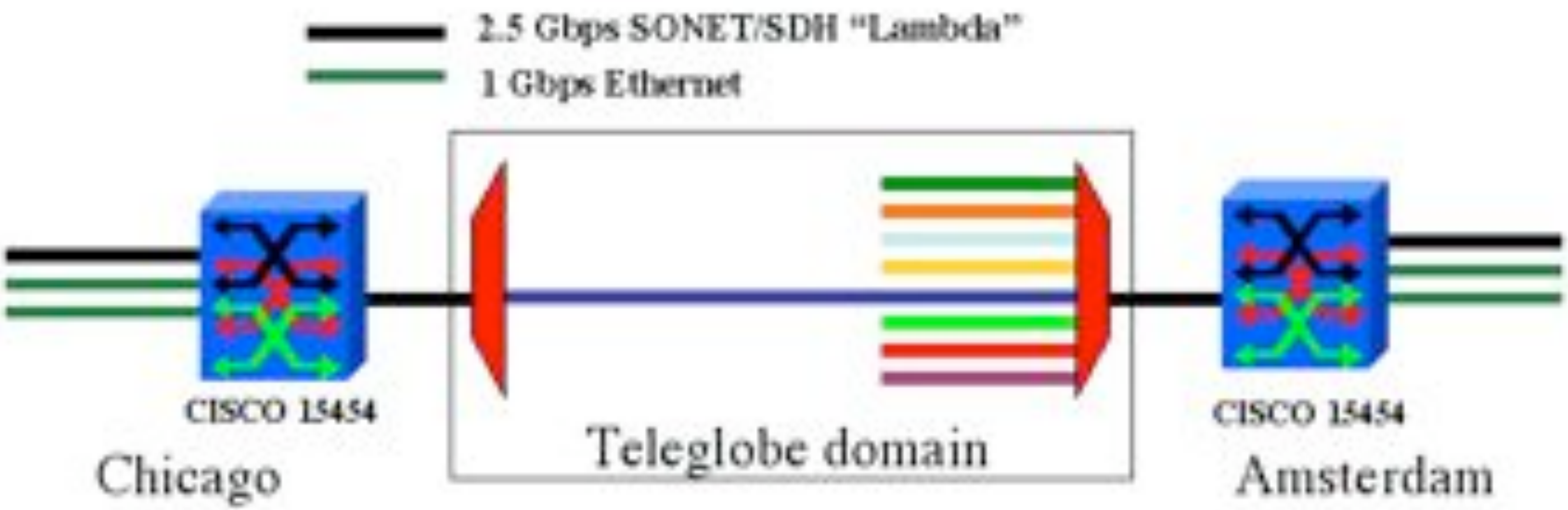
"a λ is a pipe where you can inspect packets as they enter and when they exit, but principally not when in transit. In transit one only deals with the parameters of the pipe: number, color, bandwidth"

Possible STAR LIGHT configuration

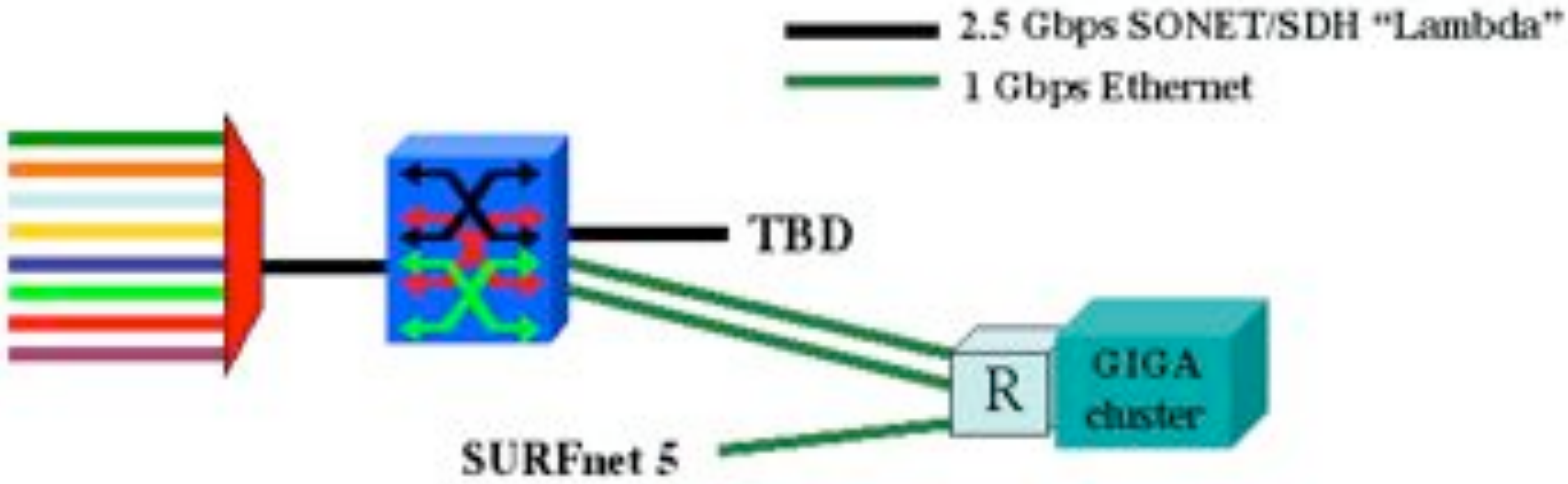


Courtesy Bill St.Arnaud

Basic phase 1

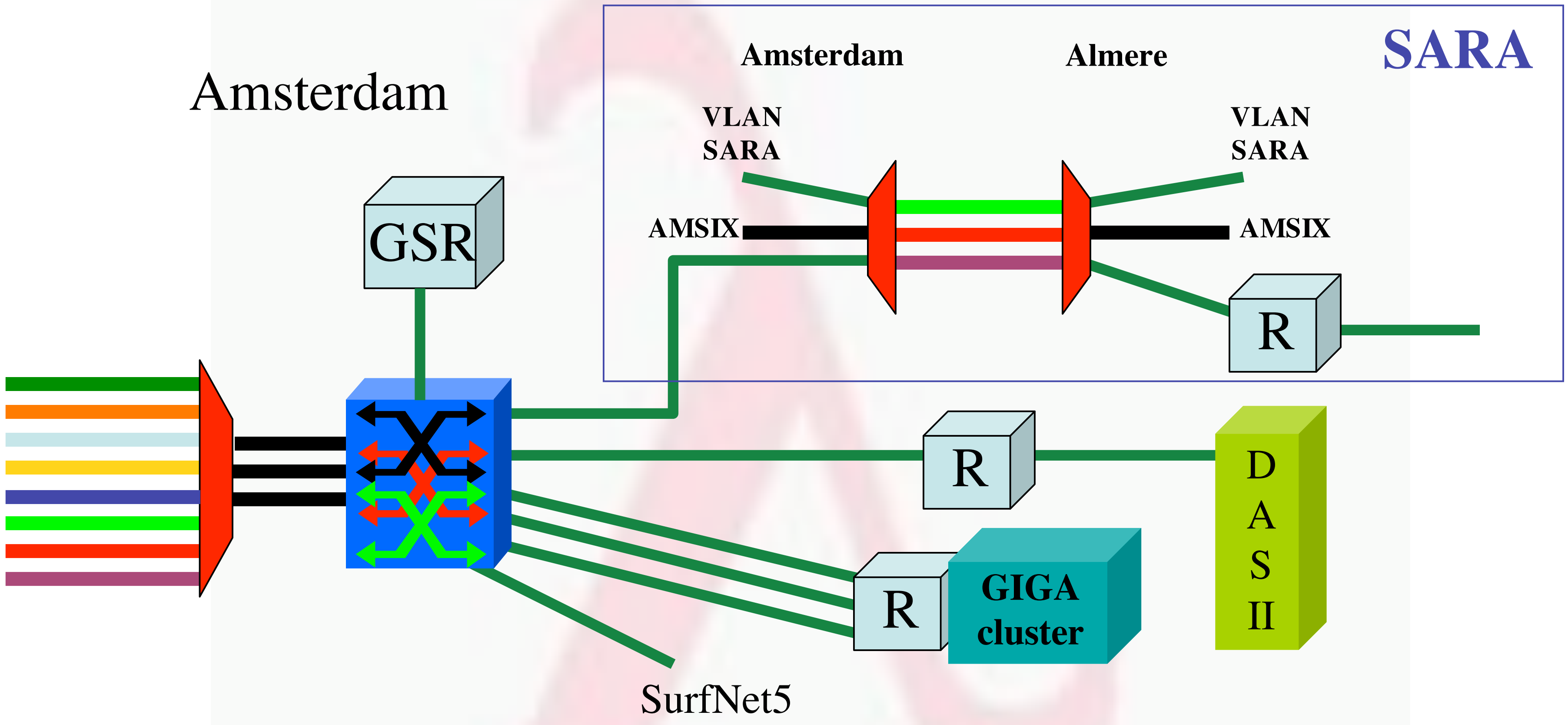


Amsterdam 1st phase



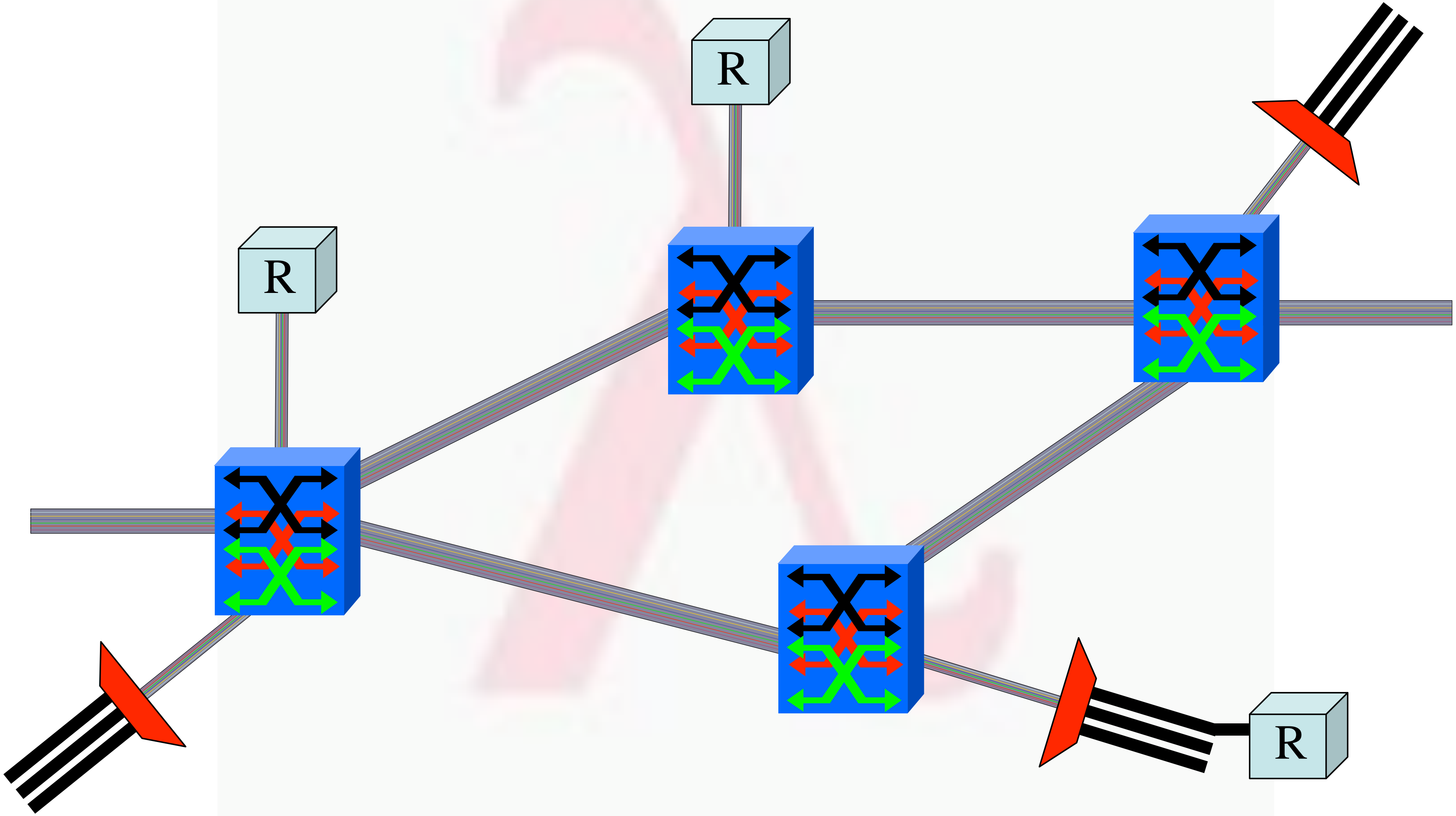
2.5 Gbps SONET/SDH "Lambda"

10/100/1000 Mbps Ethernet

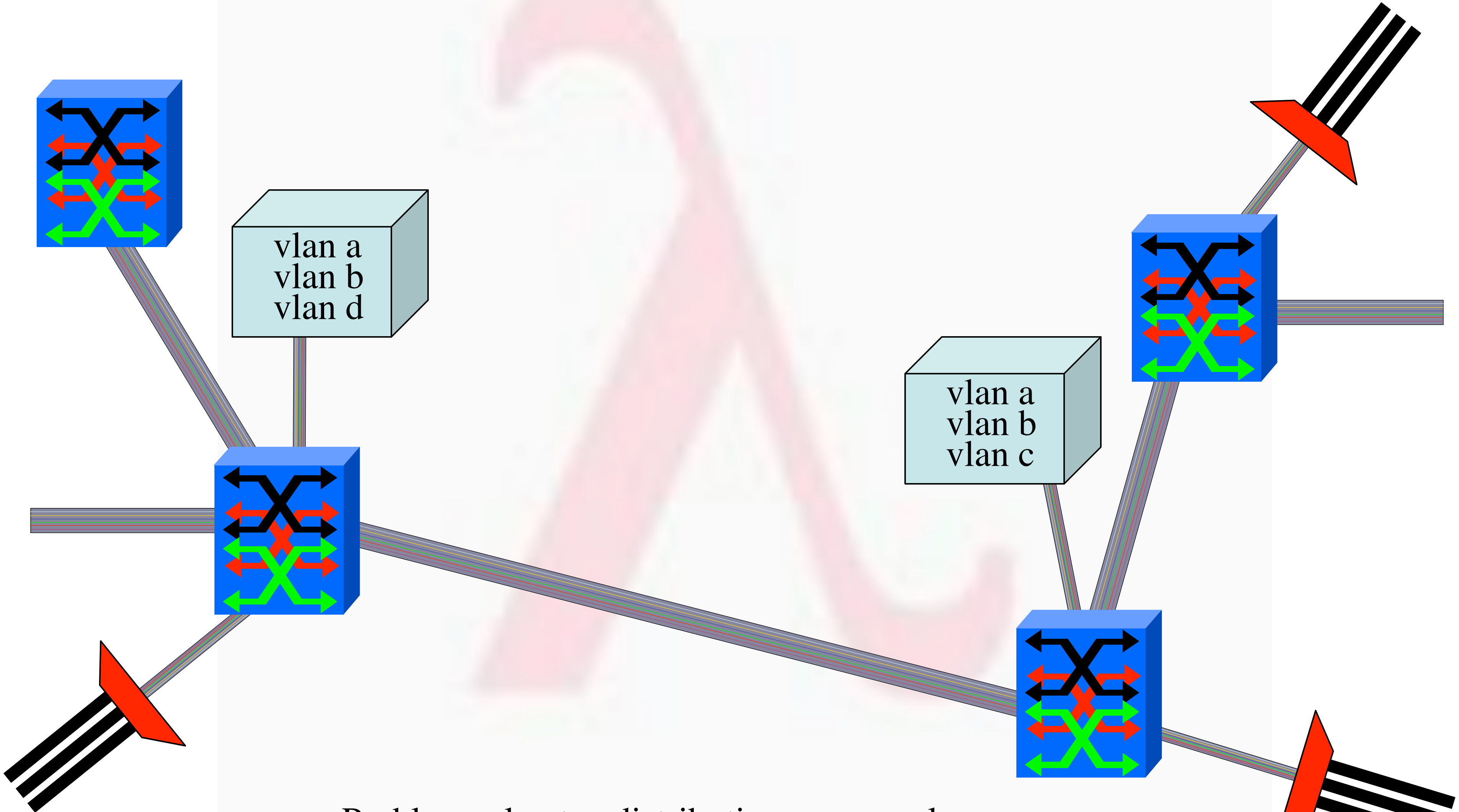


Amsterdam 2nd phase

Other architectures - L1 - 3

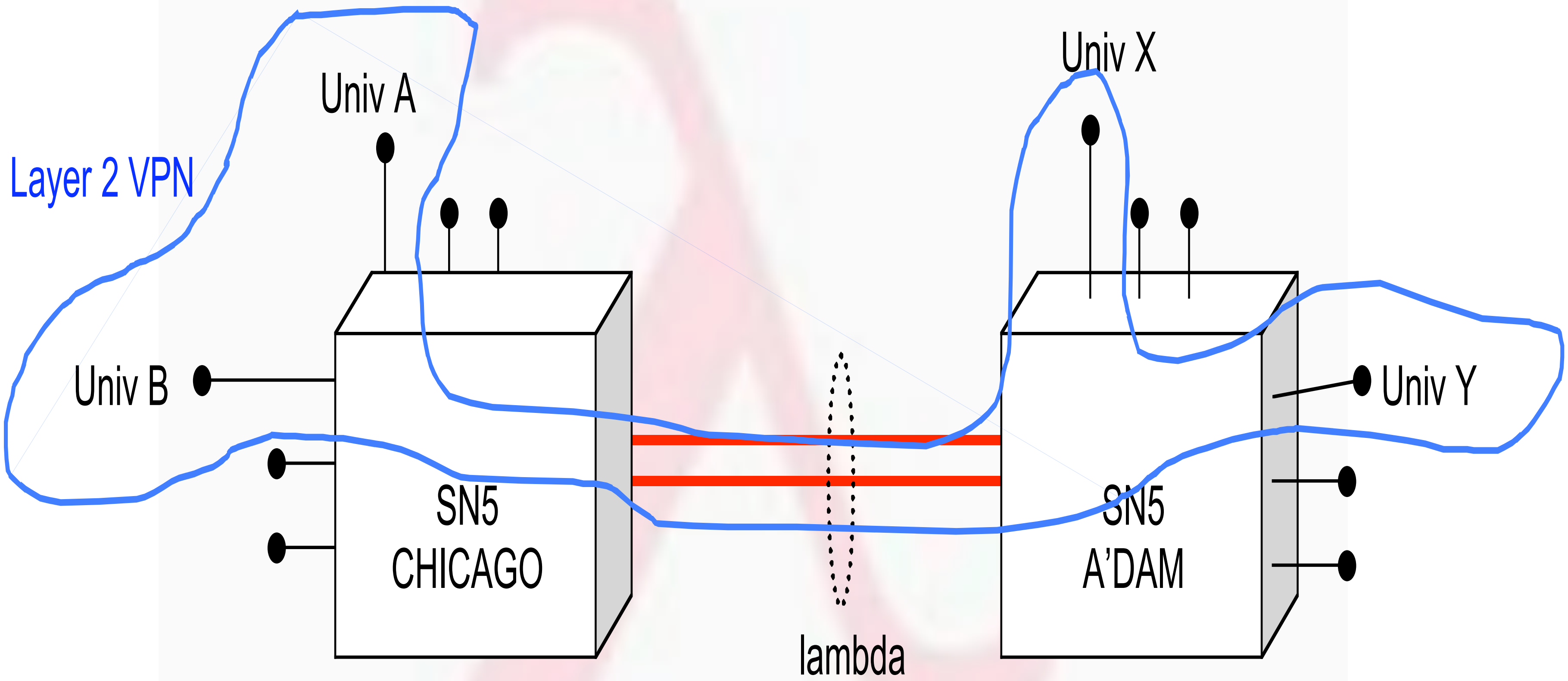


Other architectures - Distributed virtual IEX'es

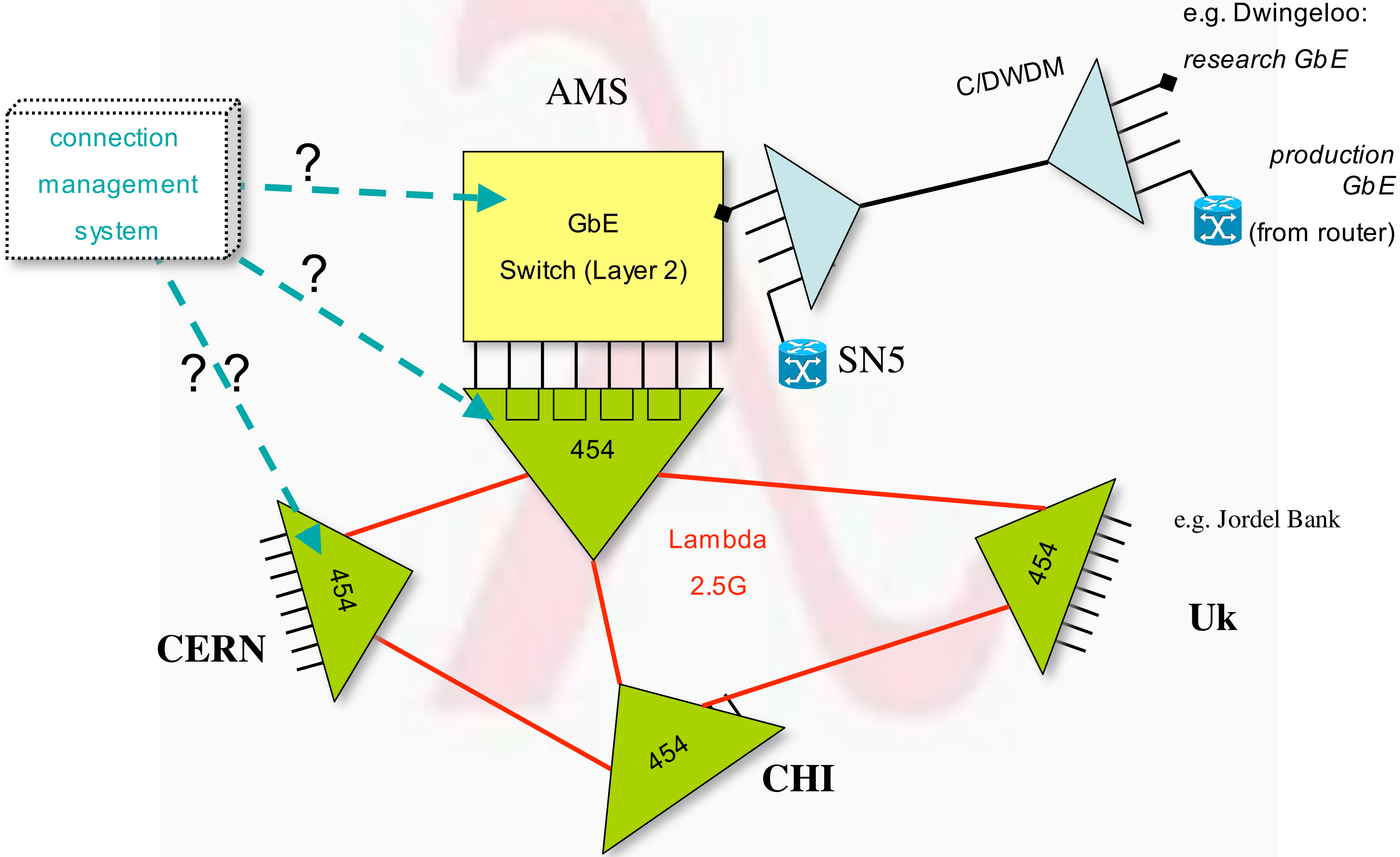


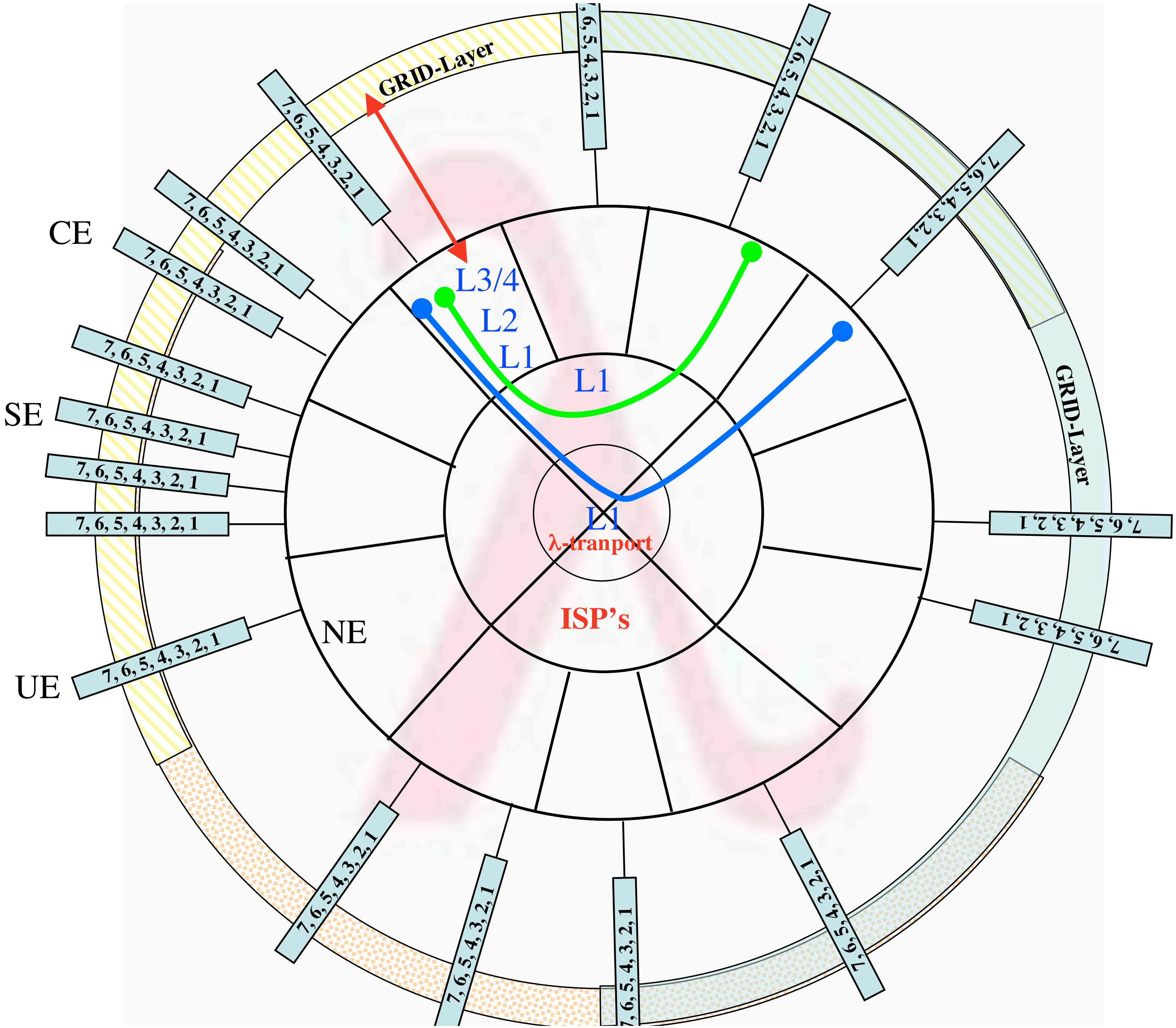
Problem: vlan tag distribution ==> gmpls

Distributed L2

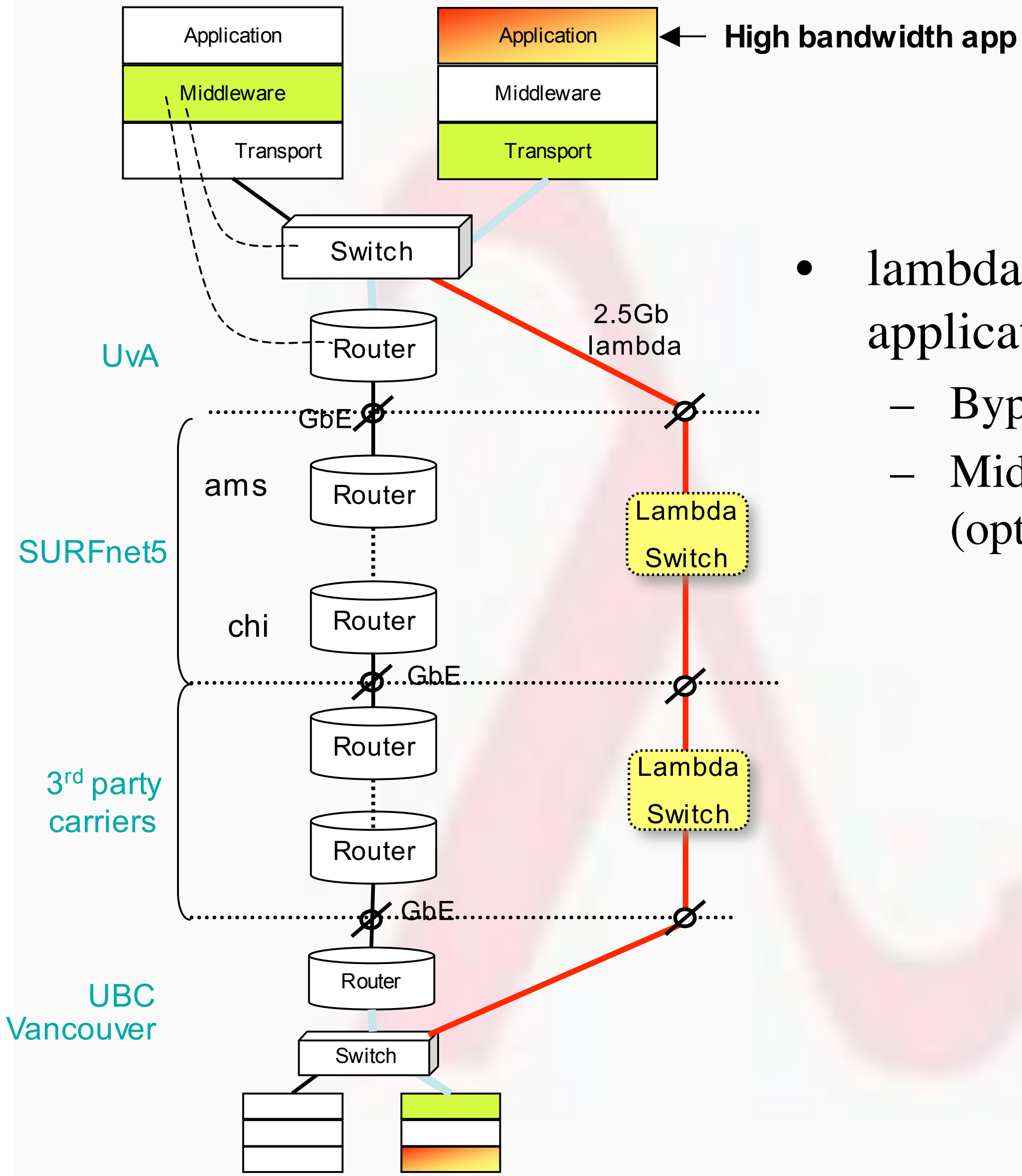


Lambda/GbE exchange





GRID-Layer
GRID-Layer
CE
SE
UE
NE
L3/4
L2
L1
L1
L1
λ-tranport
ISP's
7, 6, 5, 4, 3, 2, 1



- lambda for high bandwidth applications
  - Bypass of production network
  - Middleware may request (optical) pipe

research on λ's

- how to get traffic in and out of lambdas
- how to map load on the network to a map of lambdas
- how to deal with lambdas at peering points
- how to deal with provisioning when more administrative domains are involved
- how to do fine grain near real time grid application level lambda provisioning

Research with λ's

- High speed TCP (high rtt and BW)
- Routing stability
- Routing responsibility
- Extremely multihomed Networks
- Roles, organizational issues
- SLA's
- Models (Connection less versus oriented)
- Discreet versus continuous in time

The End

- This space is intentionally left blank