

Towards Service Differentiation on the Internet

from

**“New Internet and Networking Technologies for Grids and
High-Performance Computing”**,
tutorial given at IEEE HOTI 2006, Stanford, California
August 25th, 2006

C. Pham
University of Pau, France
LIUPPA laboratory

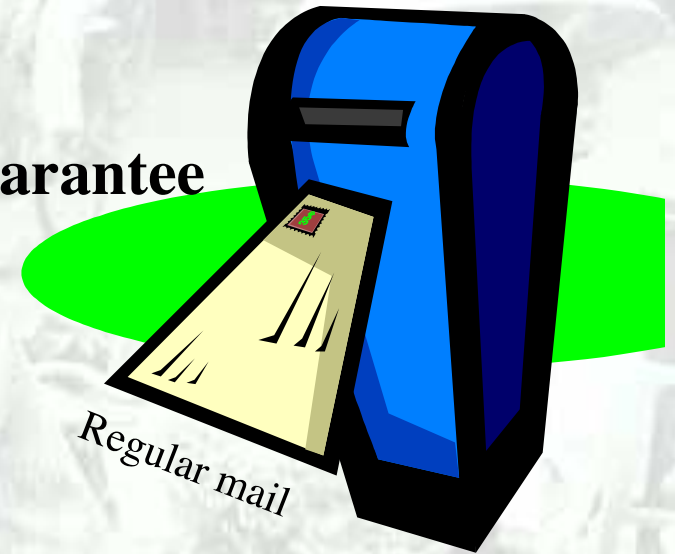
Revisiting the *same service* for all paradigm

NEW
CHAPTER



No delivery guarantee

INTERNET



Enhancing the best-effort service



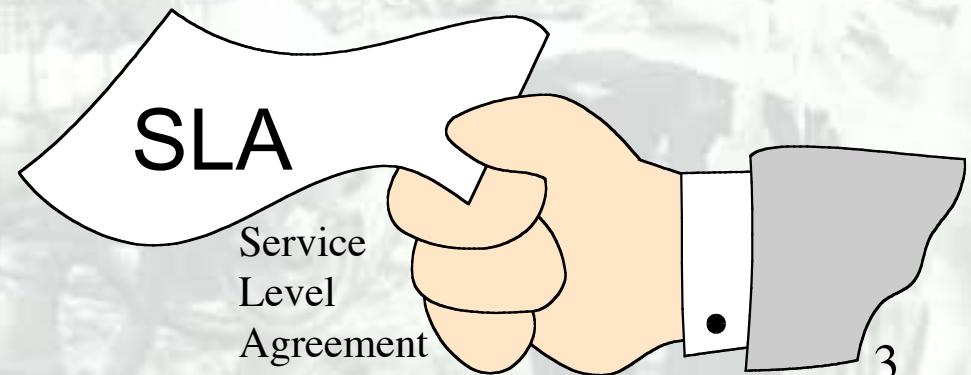
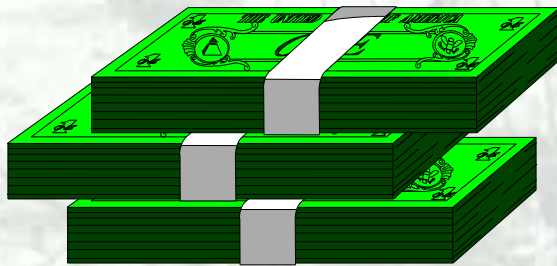
Introduce
Service Differentiation



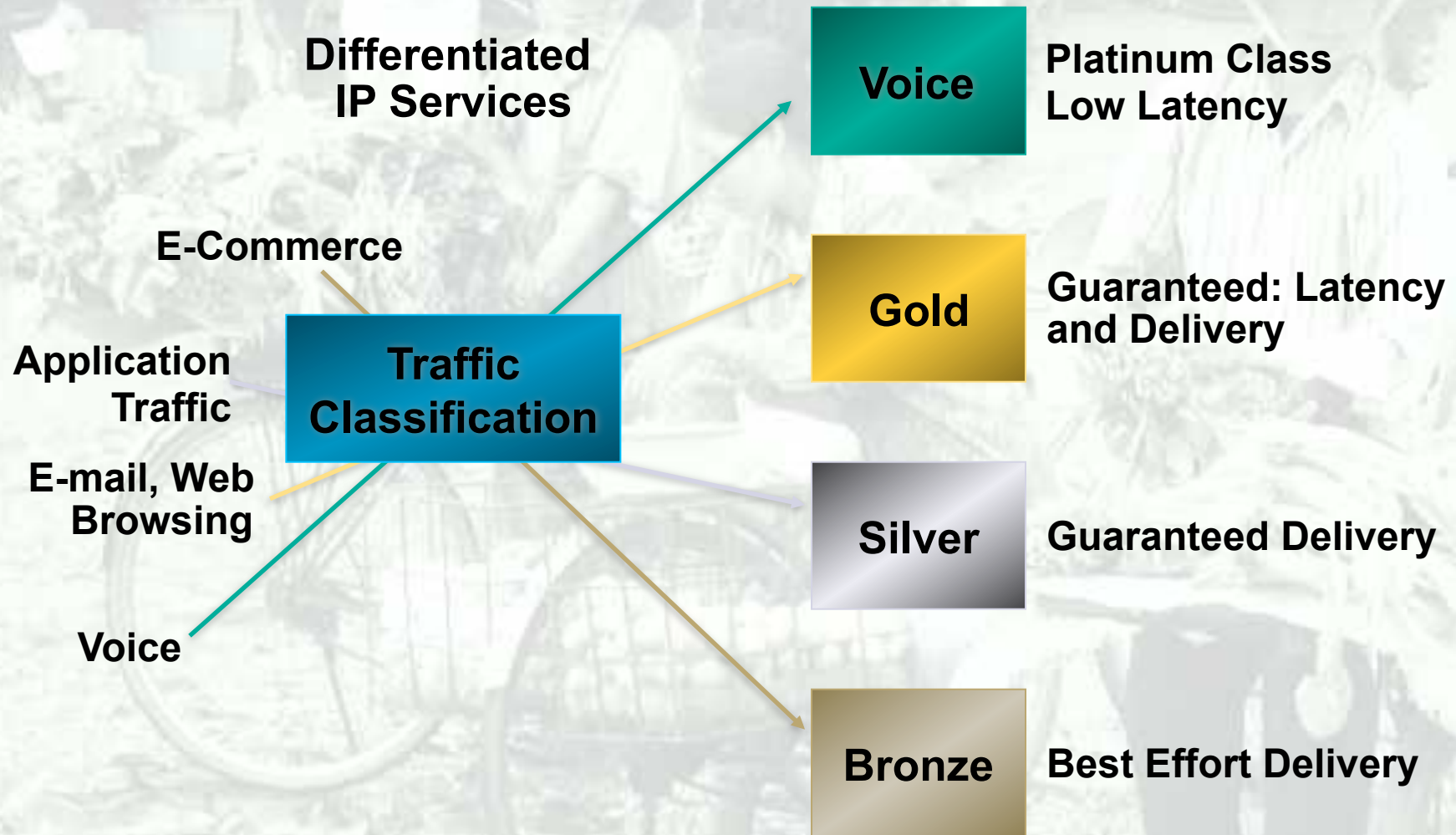
Service Differentiation

The real question is to choose which packets shall be dropped. The first definition of differential service is something like "not mine."
-- Christian Huitema

- ❑ Differentiated services provide a way to specify the relative priority of packets
- ❑ Some data is more important than other
- ❑ People who pay for better service get it!



Divide traffic into classes



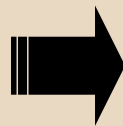
Design Goals/Challenges

- ❑ Ability to charge differently for different services
- ❑ No per flow state or per flow signaling
- ❑ All policy decisions made at network boundaries
 - ❑ Boundary routers implement policy decisions by tagging packets with appropriate priority tag
- ❑ Traffic policing at network boundaries
- ❑ Deploy incrementally: build simple system at first, expand if needed in future

IP implementation: DiffServ

RFC 2475

No per flow state in the core

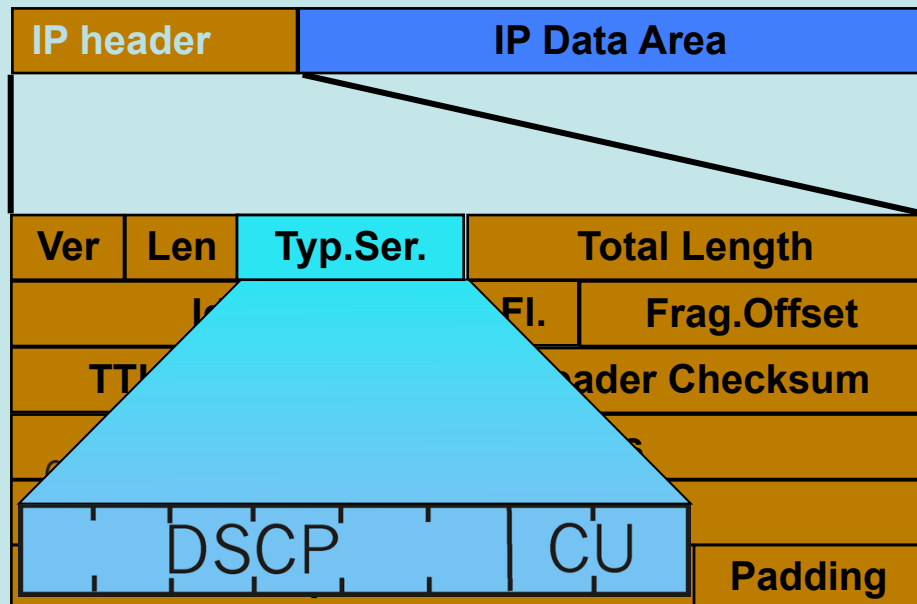
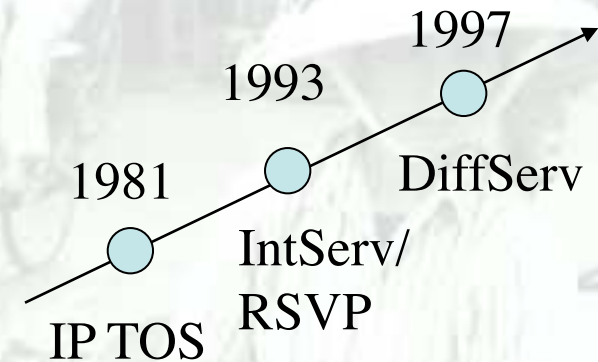


~~Flow 1
Flow 2
Flow 3
Flow 4
...~~

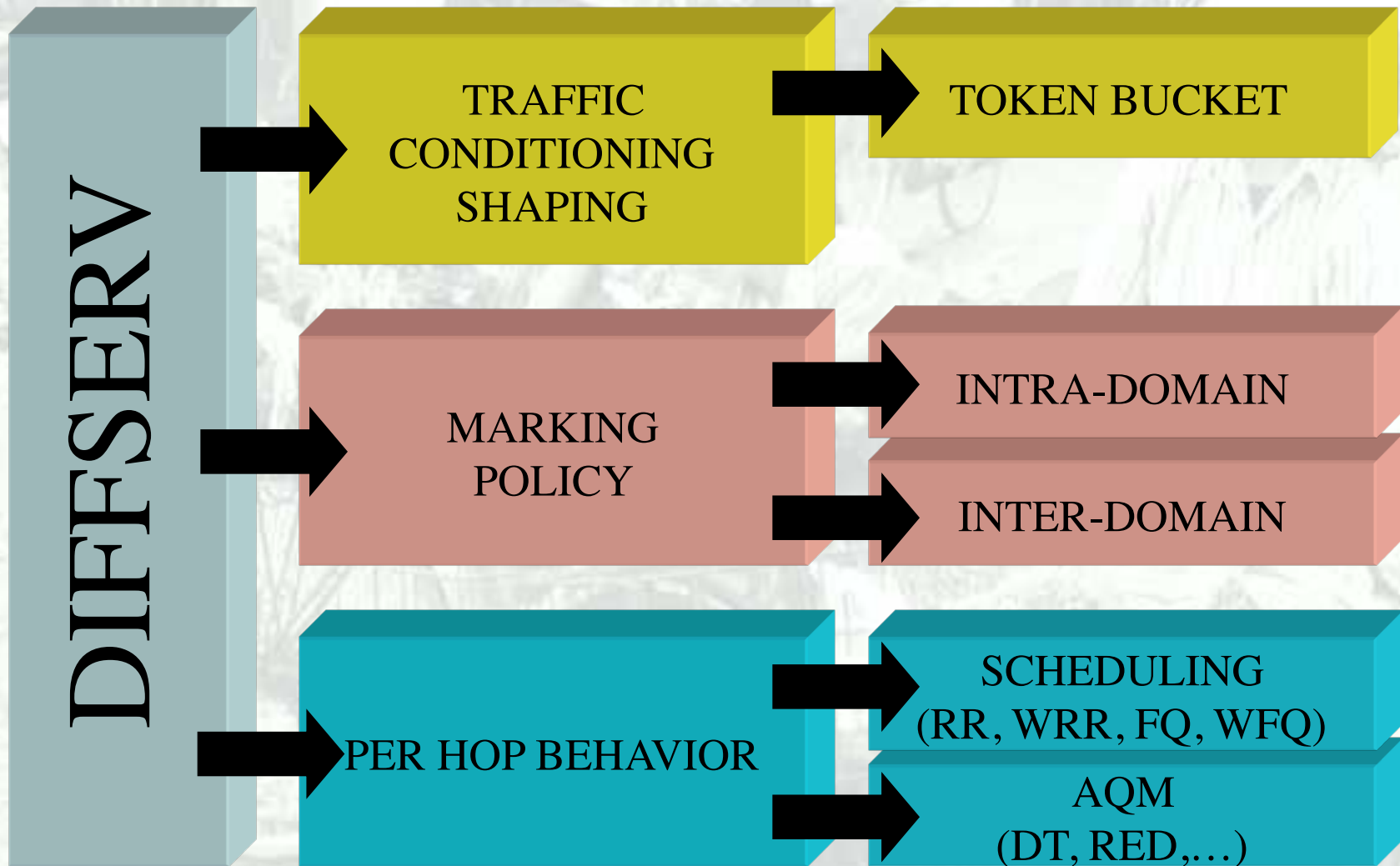
10Gbps=2.4Mpps
with 512-byte packets

**Stateful approaches
scalable
at gigabit rates**

6 bits used for Differentiated Service Code Point (DSCP) and determine PHB that the packet will receive

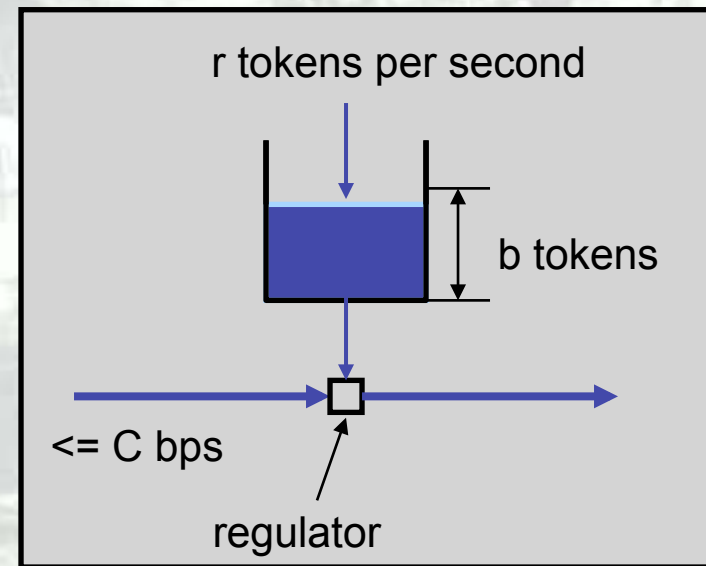
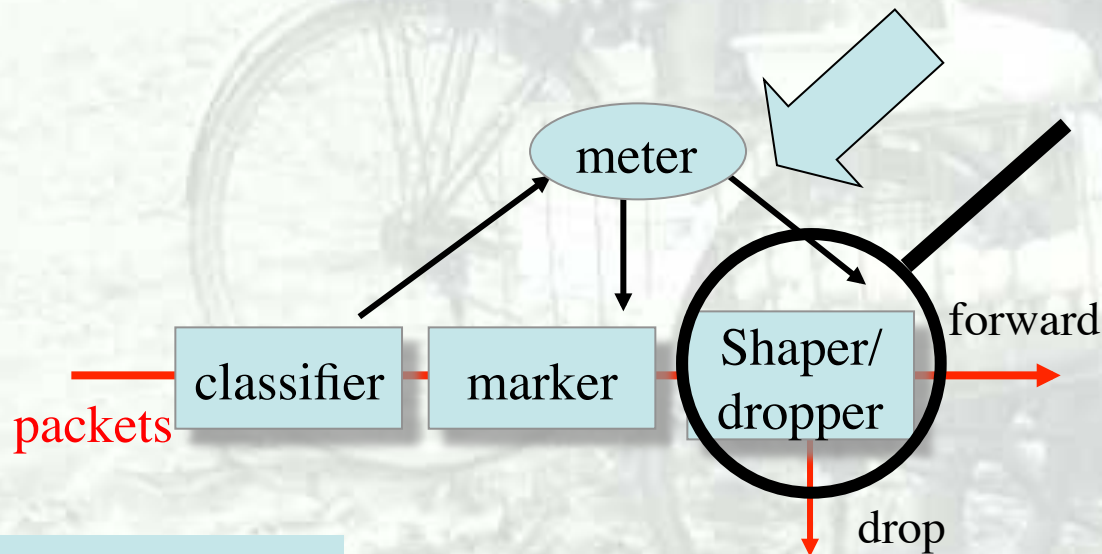
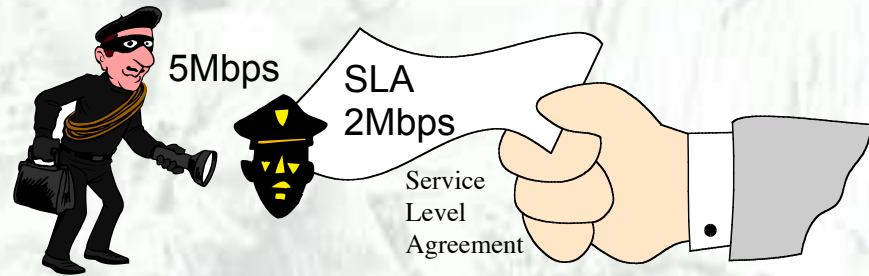


DiffServ building blocks



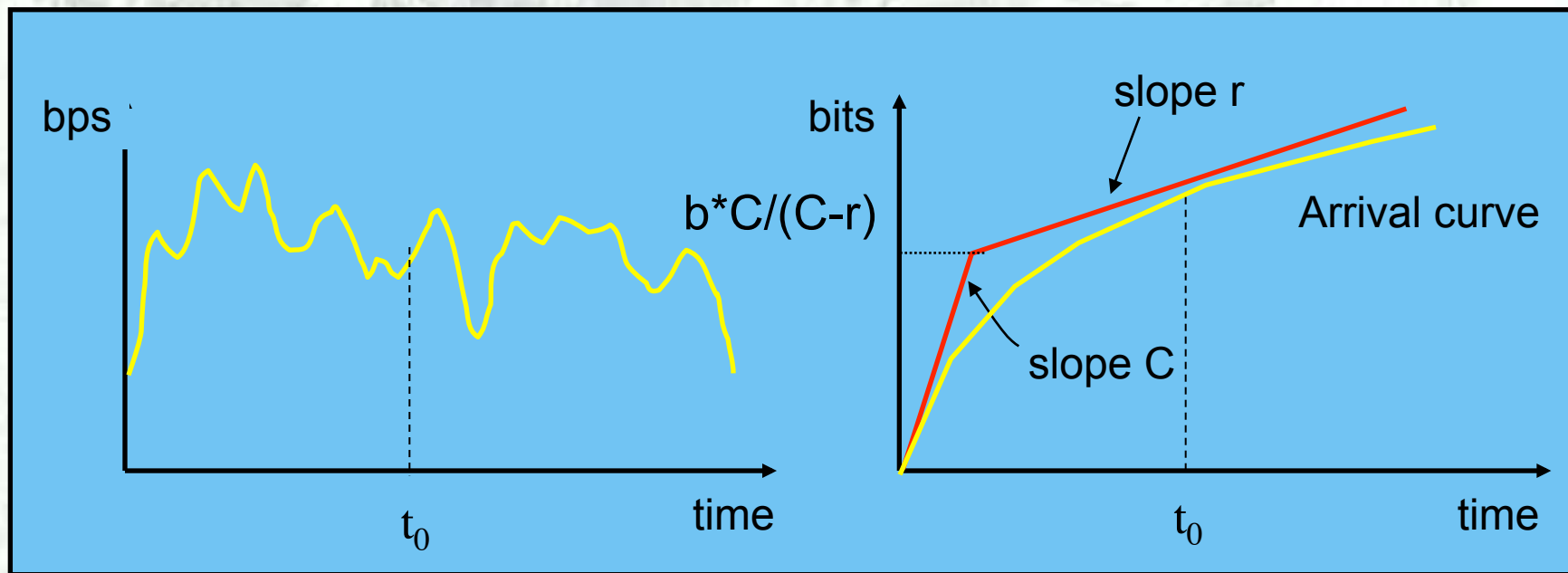
Traffic Conditioning

- User declares traffic profile (eg, rate and burst size); traffic is metered and shaped if non-conforming

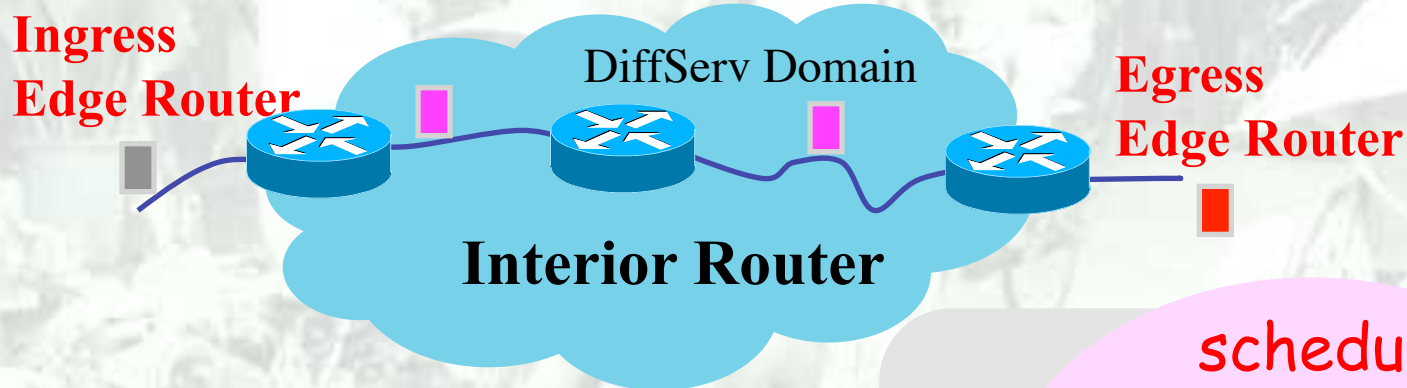


Token Bucket for traffic characterization

- Given b =bucket size, C =link capacity and r =token generation rate



Differentiated Architecture

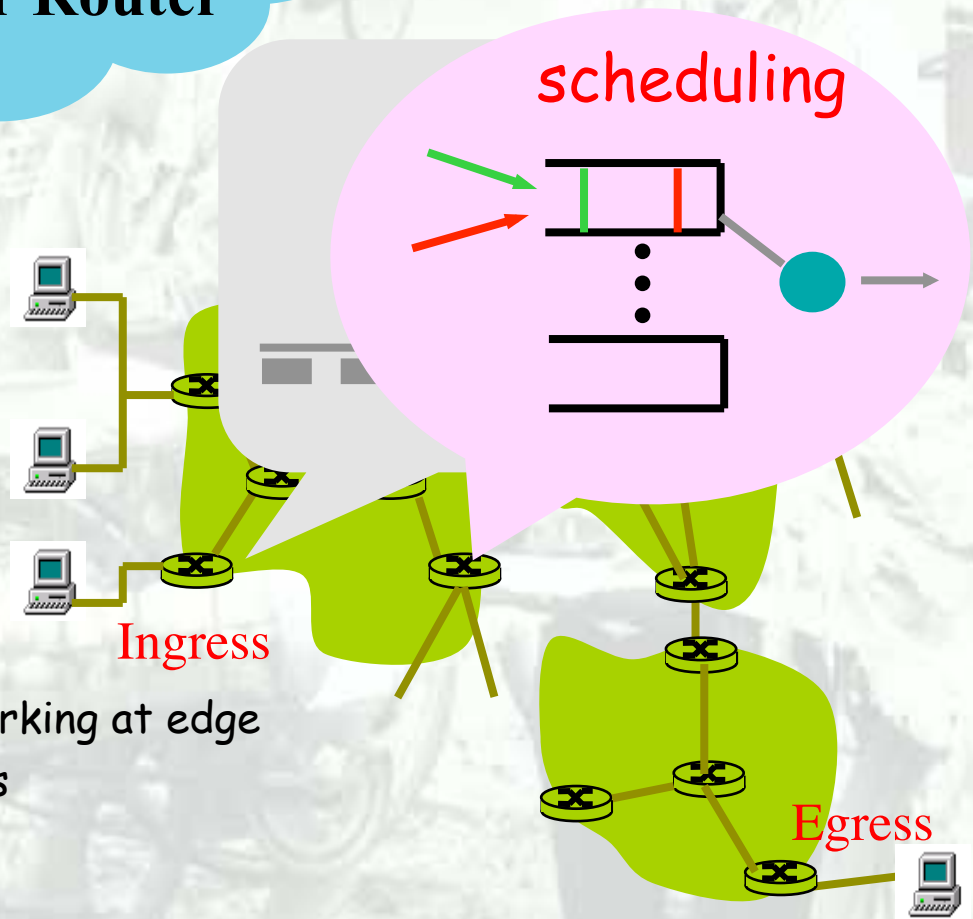


Marking:

per-flow traffic management
marks packets as in-profile and out-profile

Per-Hop-Behavior (PHB):

per class traffic management
buffering and scheduling based on marking at edge
preference given to in-profile packets



Pre-defined PHB

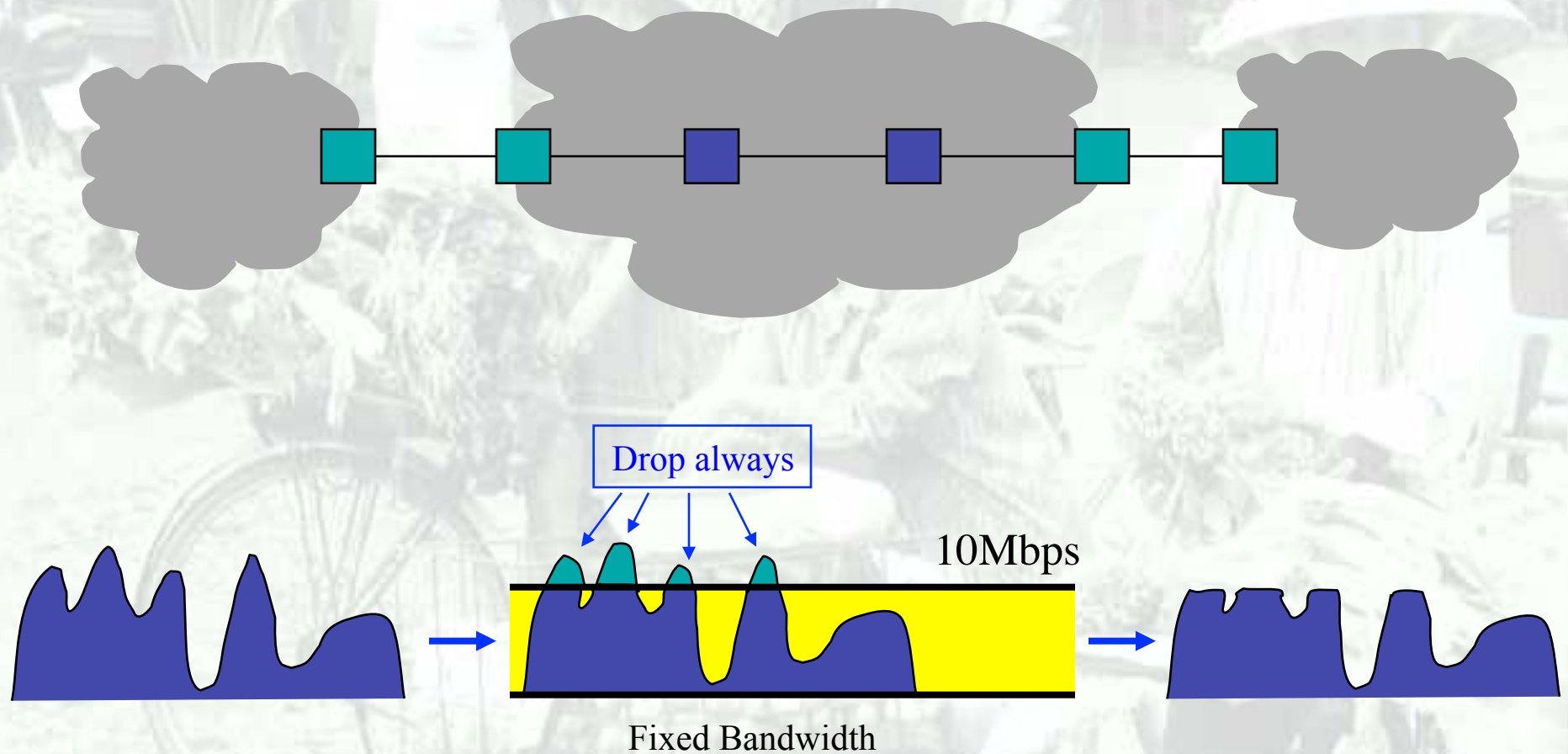
❑ Expedited Forwarding (EF, premium):

- ❑ departure rate of packets from a class equals or exceeds a specified rate (logical link with a minimum guaranteed rate)
- ❑ Emulates leased-line behavior

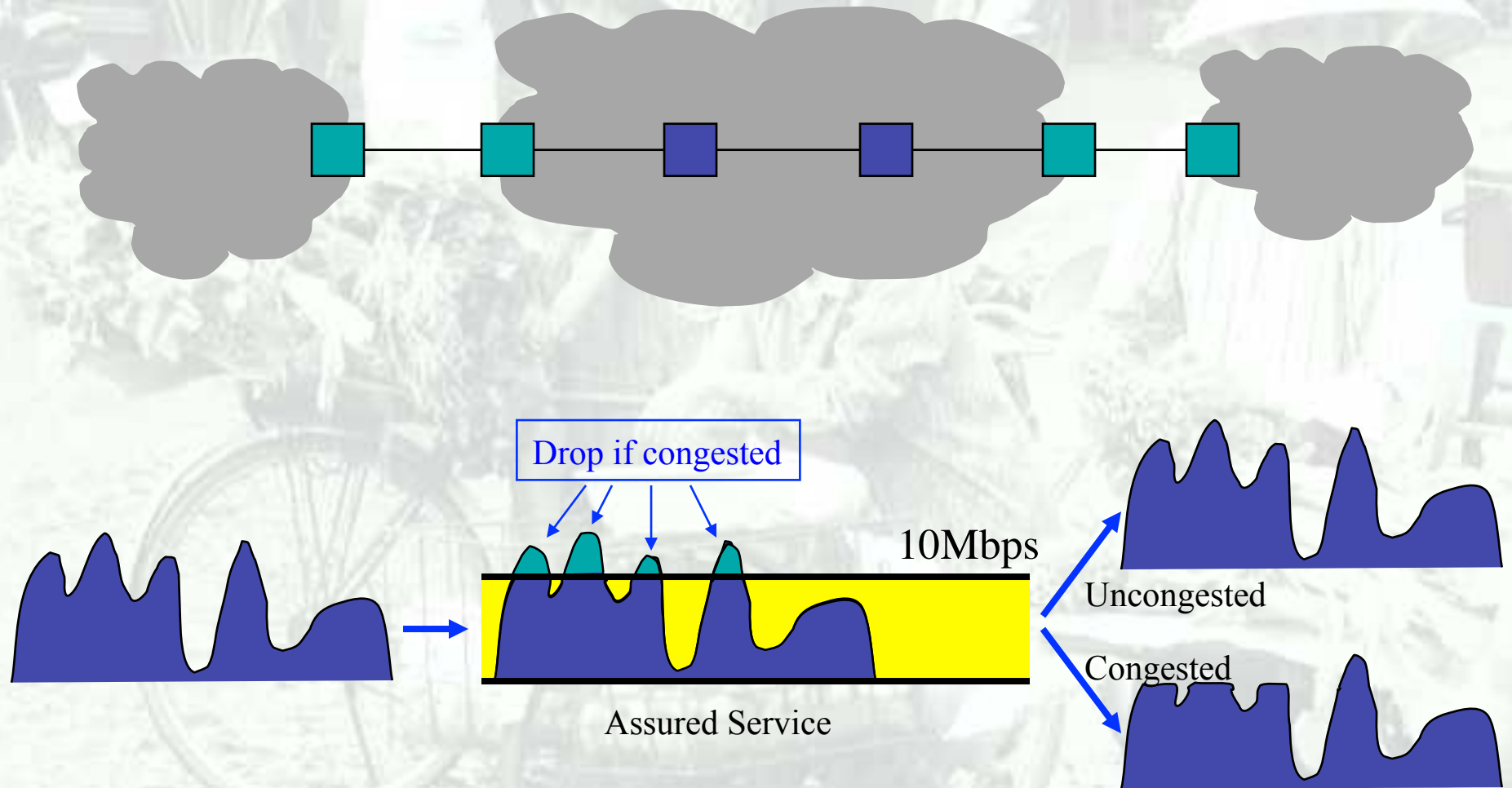
❑ Assured Forwarding (AF):

- ❑ 4 classes, each guaranteed a minimum amount of bandwidth and buffering; each with three drop preference partitions
- ❑ Emulates frame-relay behavior

Premium Service Example

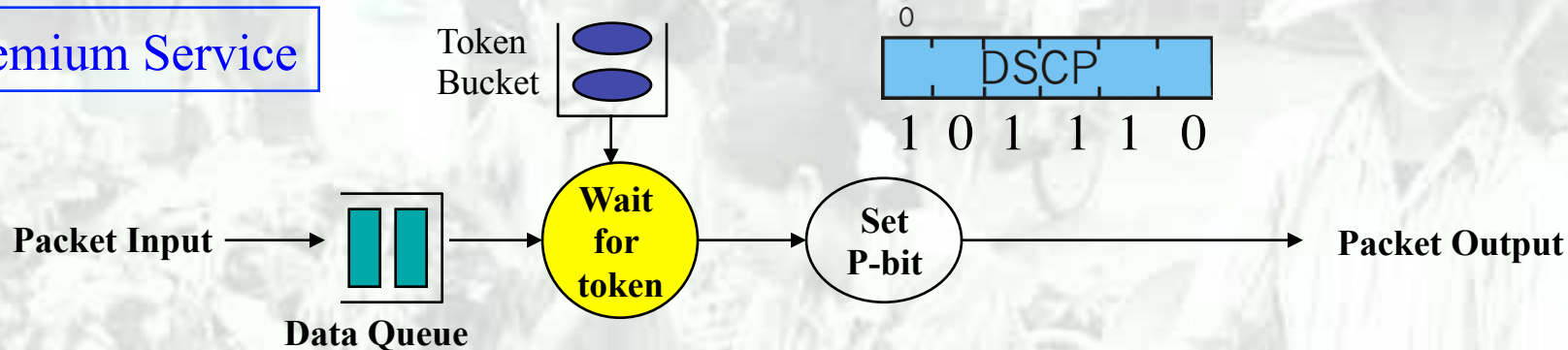


Assured Service Example

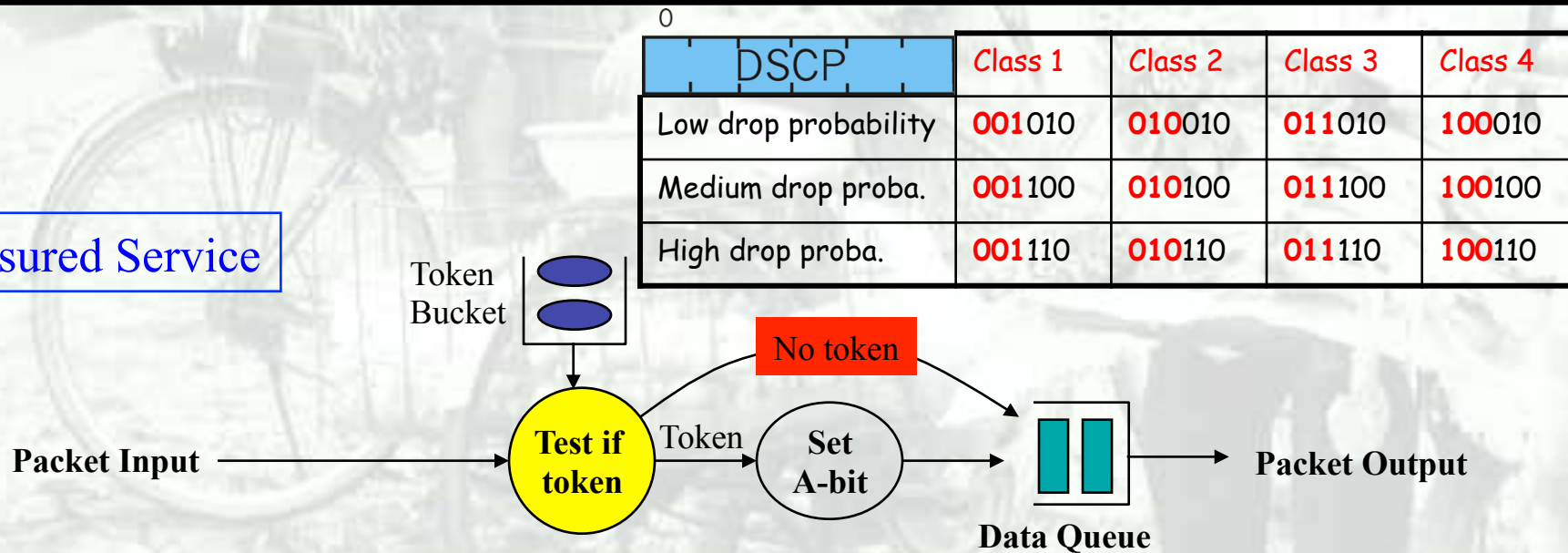


Border Router Functionality

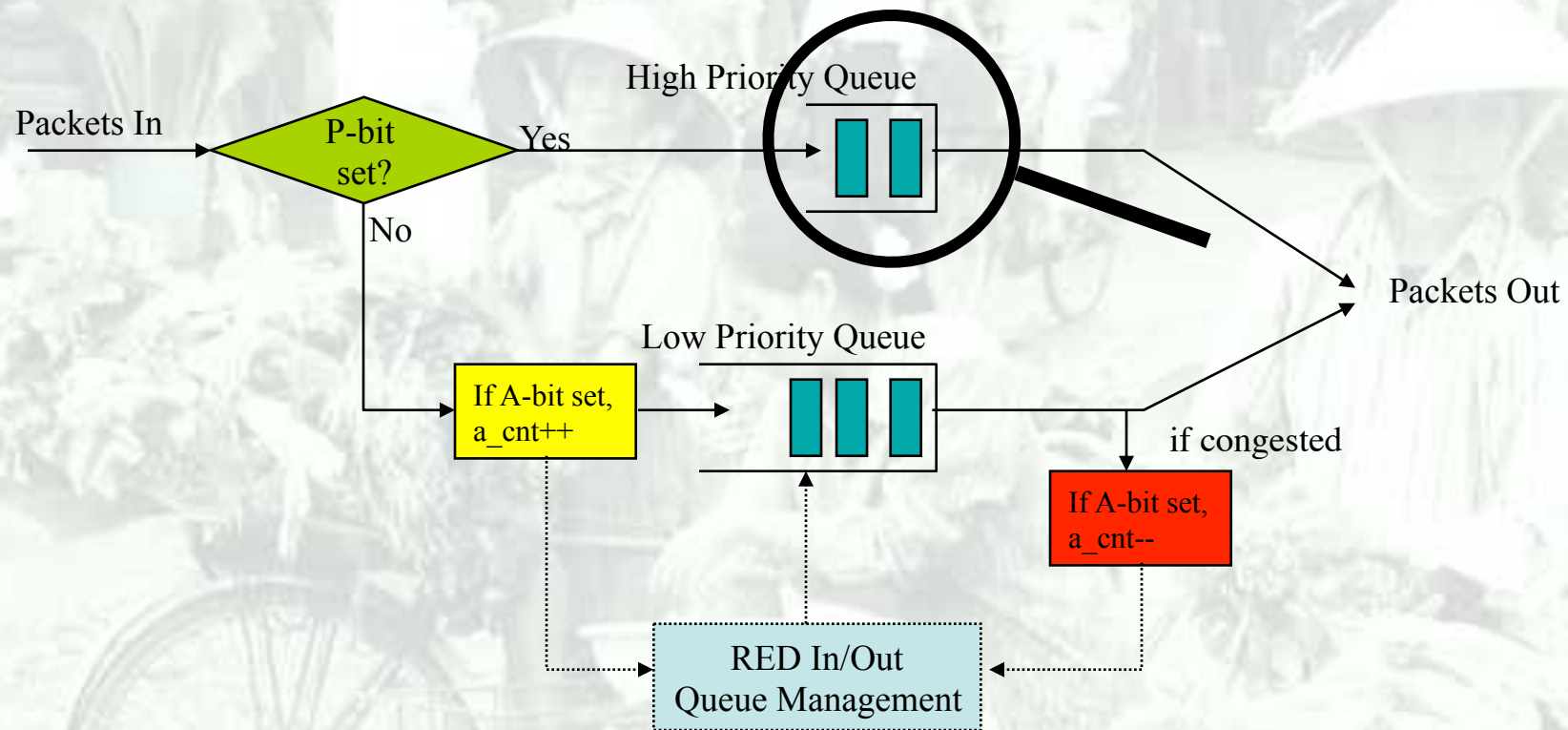
Premium Service



Assured Service



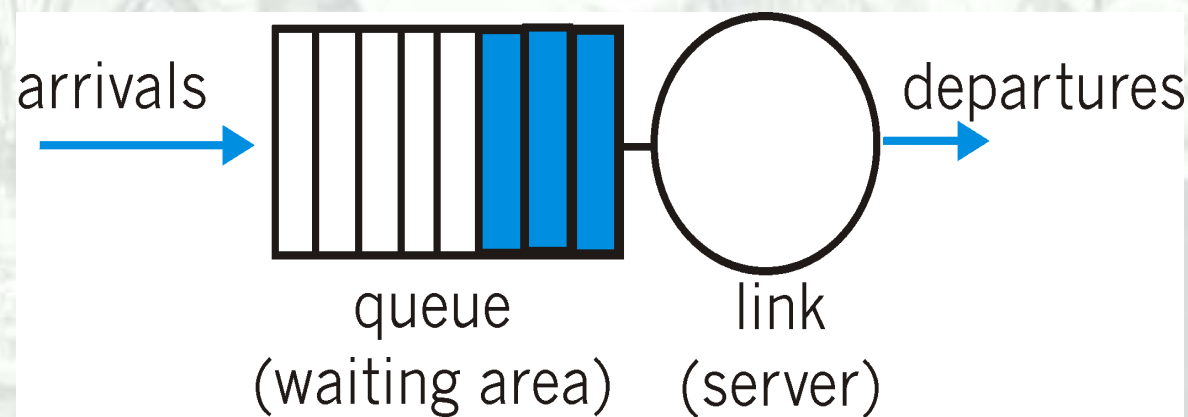
Internal Router Functionality



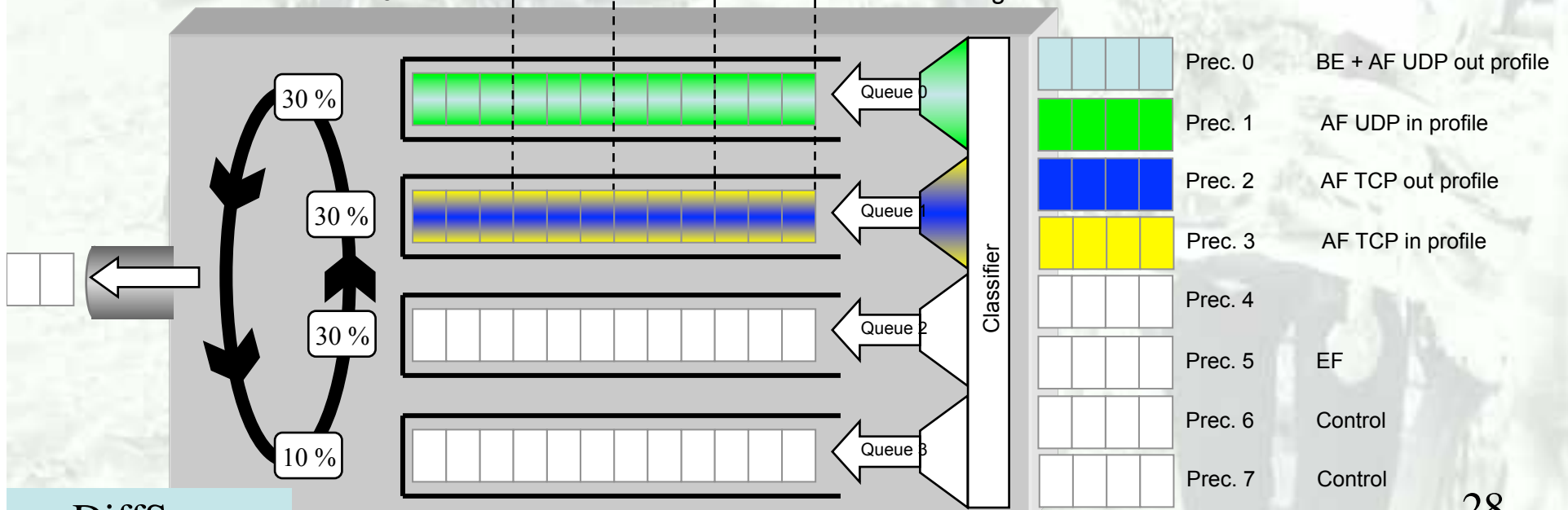
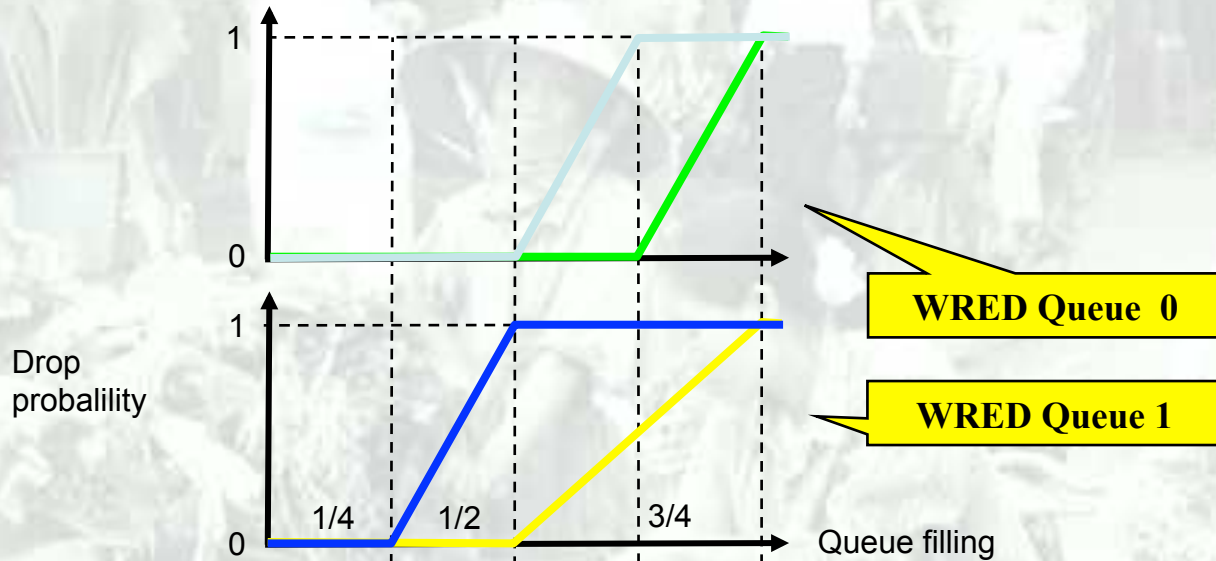
A DSCP codes aggregates, not individual flows
No state in the core
Should scale to millions of flows

Scheduling

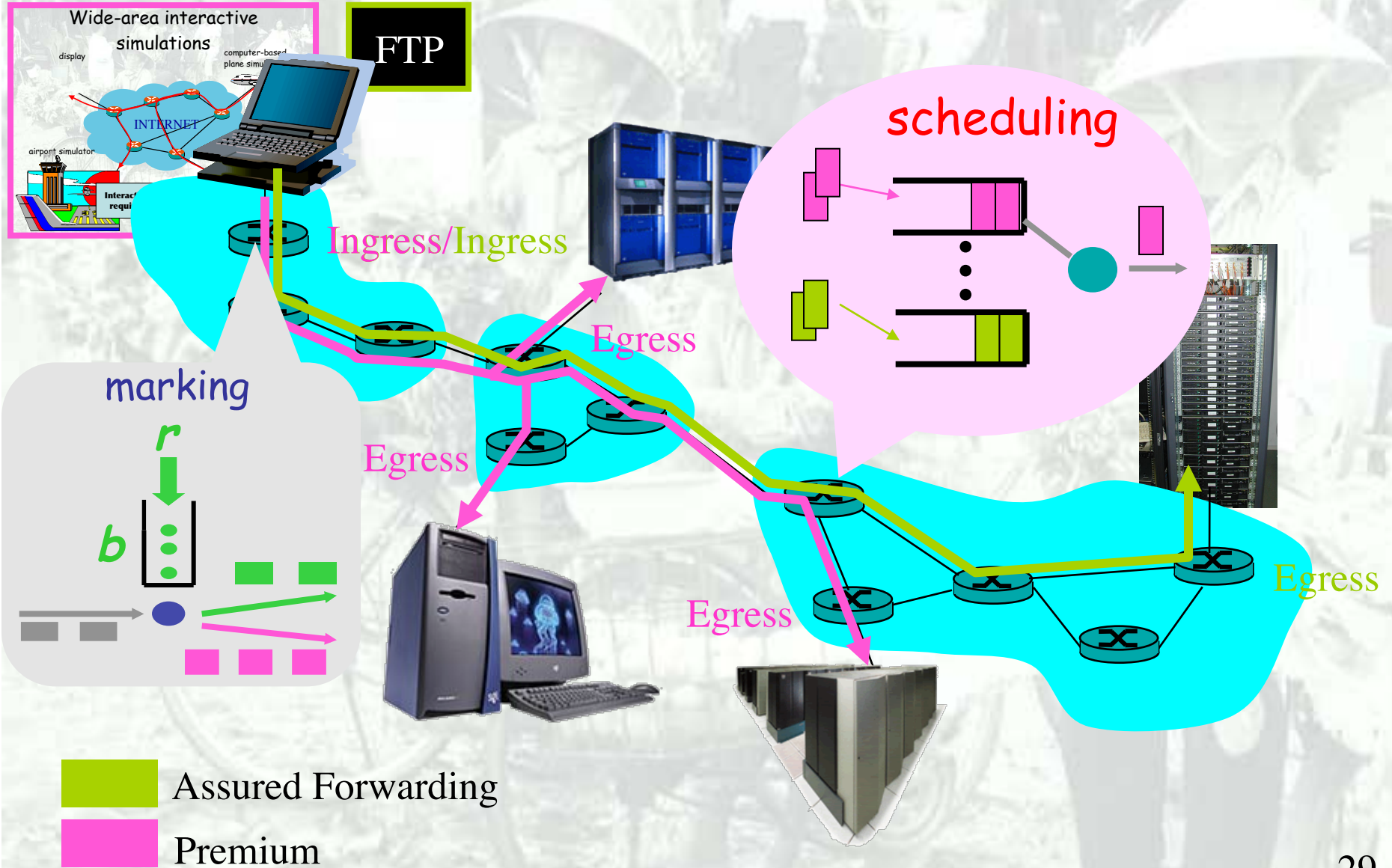
- ❑ DiffServ PHB relies mainly on scheduling
 - ❑ choose the next packet for transmission
 - ❑ FIFO: in order of arrival to the queue; packets that arrive to a full buffer are either discarded, or a discard policy is defined.
 - ❑ More complex policies: FCFS, PRIORITY, EDD...



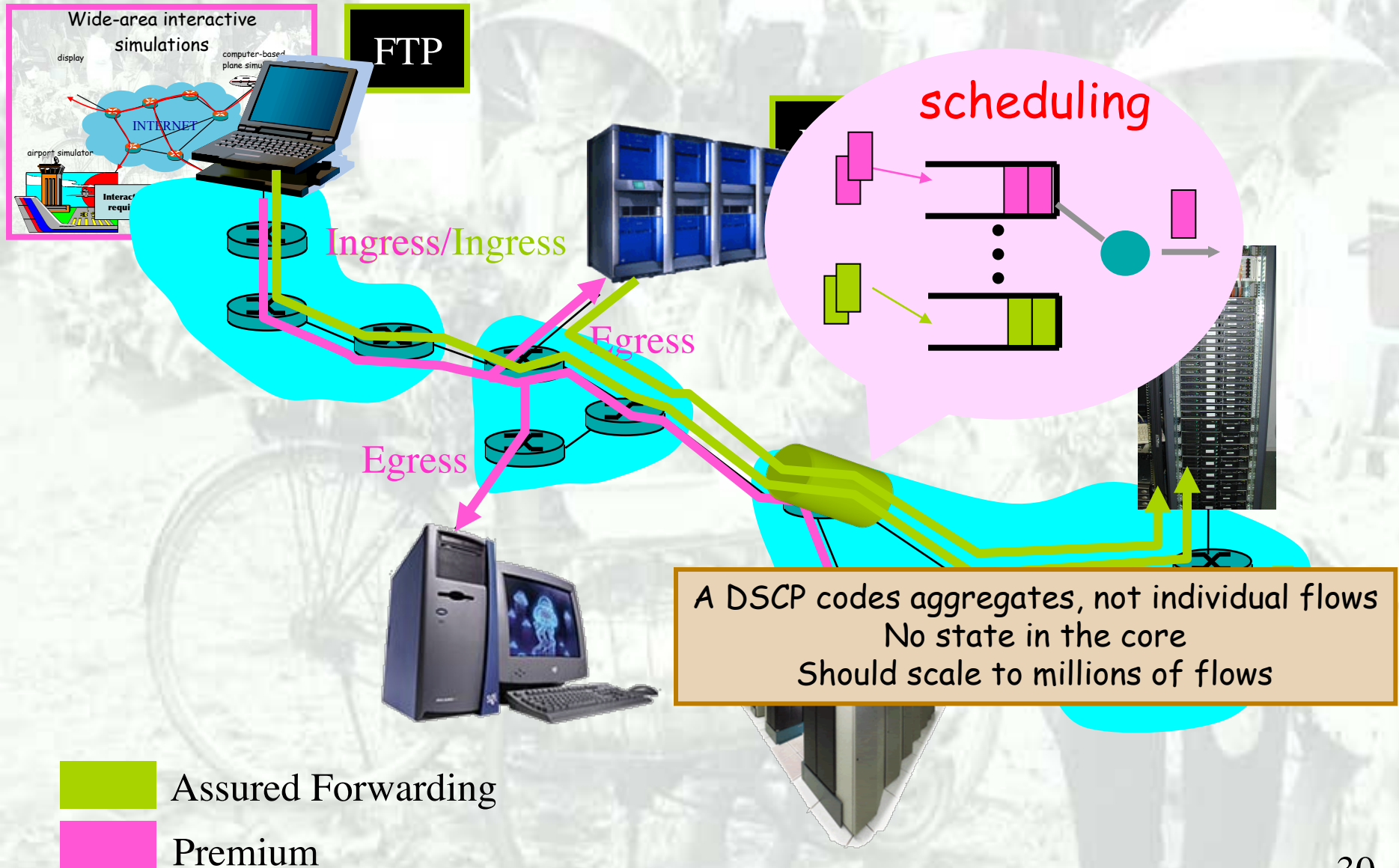
Putting it together!



DiffServ for grids

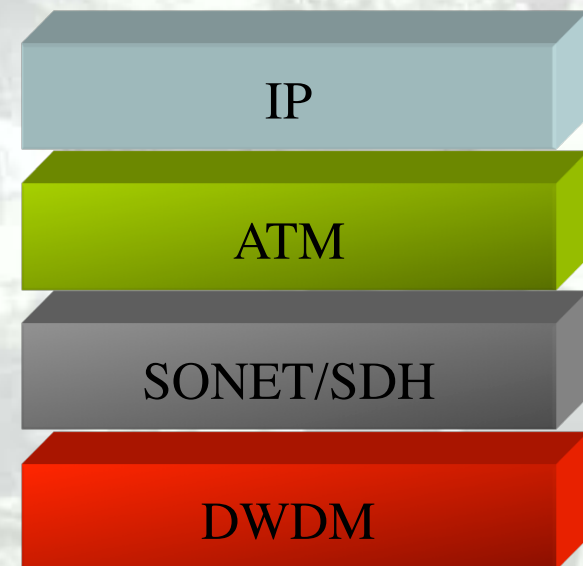


DiffServ for grids (con't)

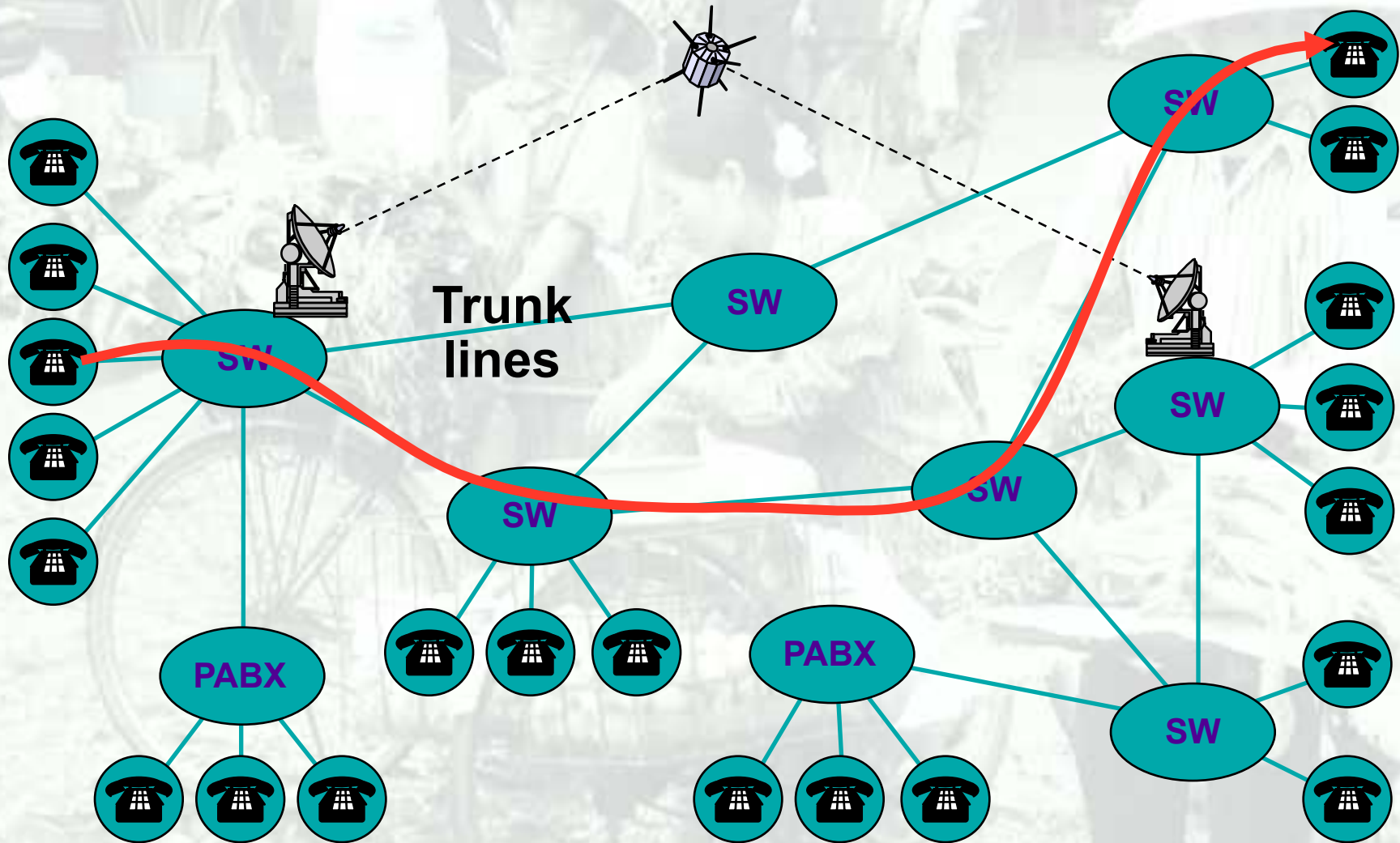


Bandwidth provisioning

- ❑ DWDM-based optical fibers have made bandwidth very cheap in the backbone
- ❑ On the other hand, dynamic provisioning is difficult because of the complexity of the network control plane:
 - ❑ Distinct technologies
 - ❑ Many protocols layers
 - ❑ Many control software



The telephone circuit view



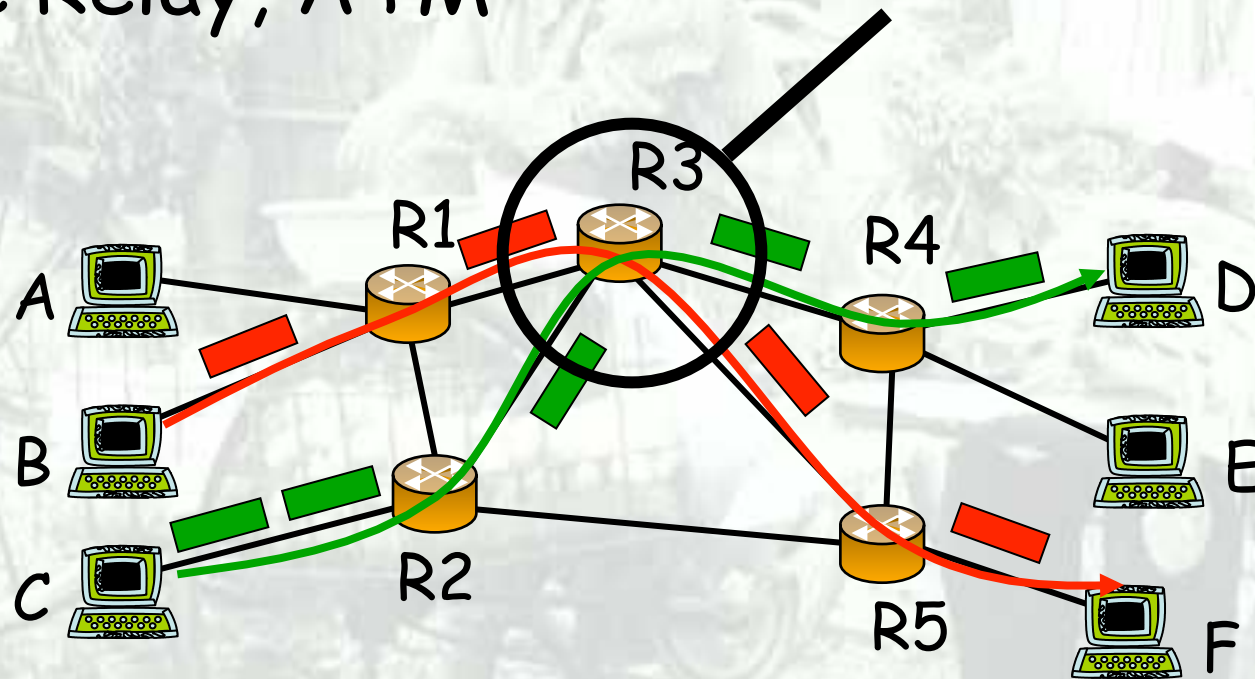
Advantages of circuits

- ❑ Provides the same path for information of the same connection: less out-of-order delivery
- ❑ Easier provisioning/reservation of network's resources: planning and management features

Back to virtual circuits

- Virtual circuit refers to a connection oriented network/link layer: e.g. X.25, Frame Relay, ATM

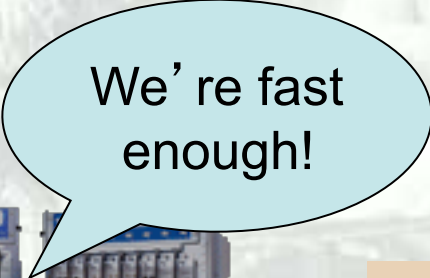
Virtual Circuit Switching: a path is defined for each connection



But IP is connectionless!

Why virtual circuit?

- Initially to speed up router's forwarding tasks: X.25, Frame Relay, ATM.



We're fast enough!



Now: Virtual circuits for traffic engineering!

Virtual circuits in IP networks

- ❑ Multi-Protocol Label Switching

- ❑ Fast: use label switching → LSR



- ❑ Multi-Protocol: above link layer, below network layer

- ❑ Facilitate traffic engineering



PPP Header(Packet over SONET/SDH)

PPP Header

MPLS Header

Layer 3 Header

Ethernet

Ethernet Hdr

MPLS Header

Layer 3 Header

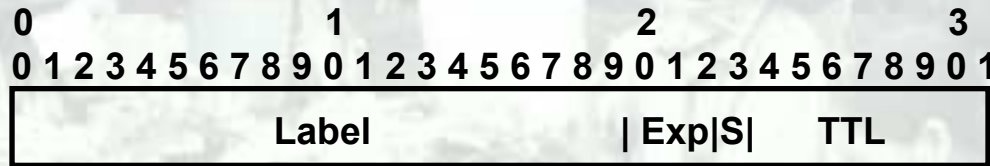
Frame Relay

FR Hdr

MPLS Header

Layer 3 Header

Label structure



Label = 20 bits

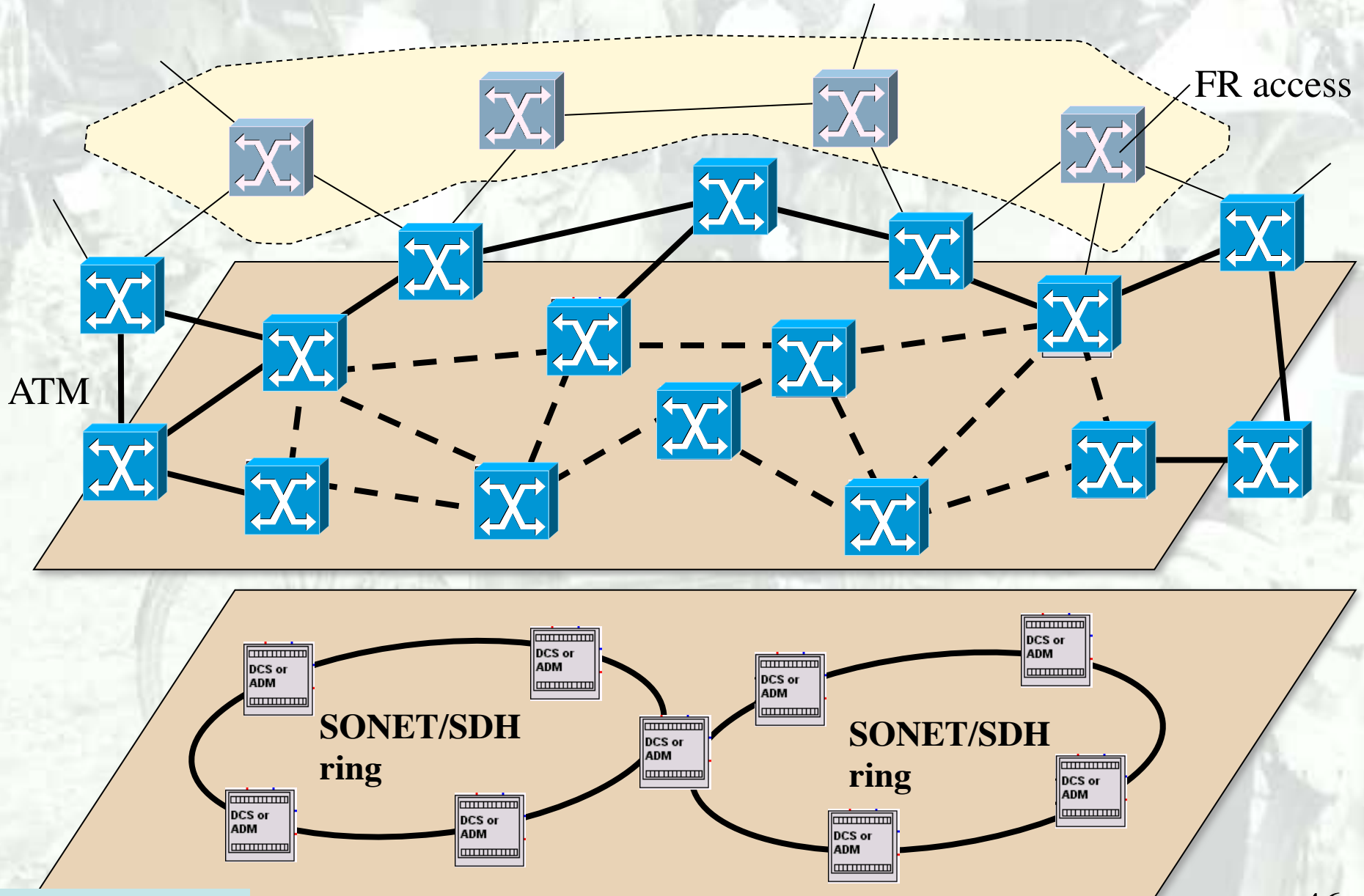
Exp = Experimental, 3 bits

S = Bottom of stack, 1bit

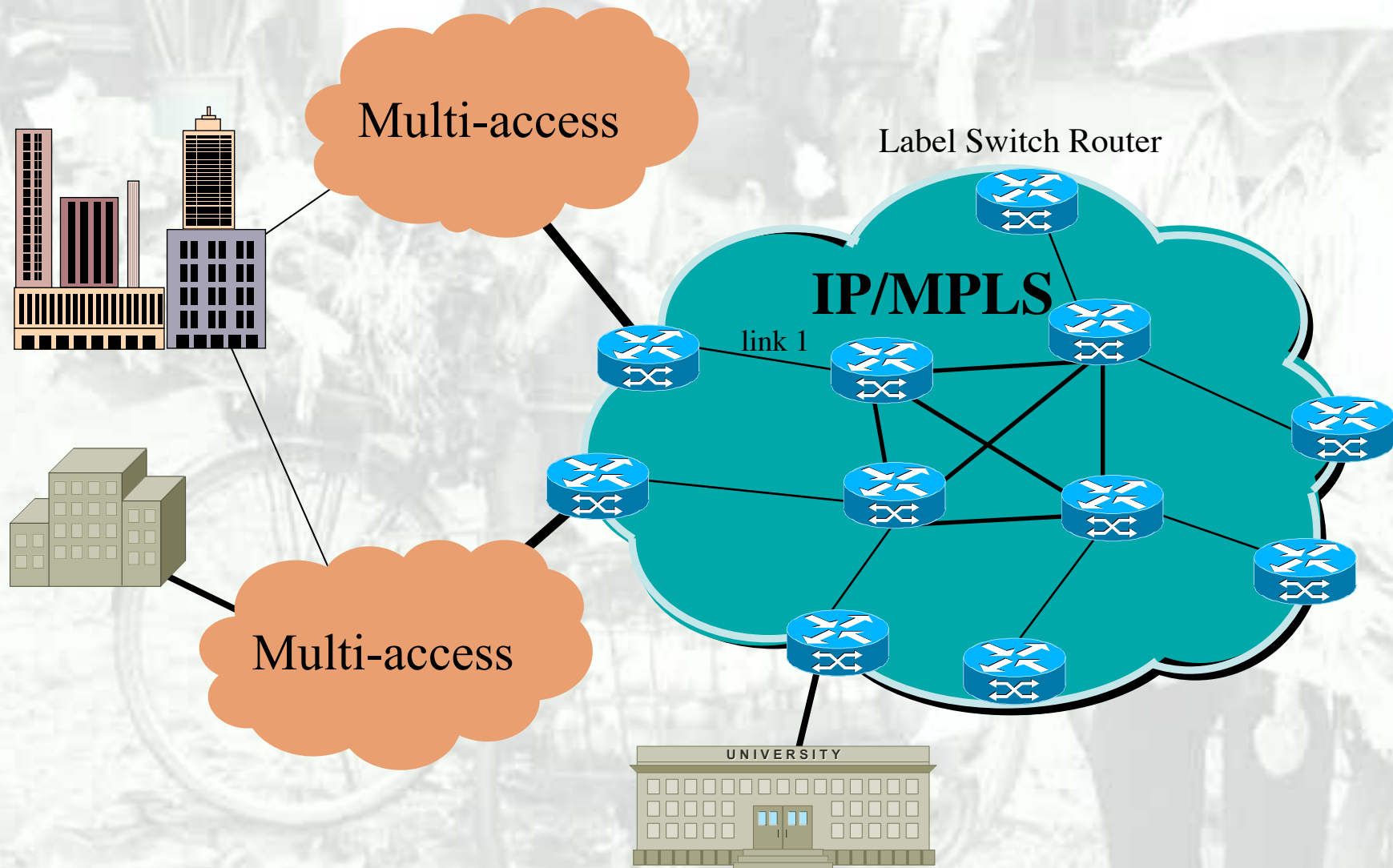
TTL = Time to live, 8 bits

- ❑ More than one label is allowed -> Label Stack
- ❑ MPLS LSRs always forward packets based on the value of the label at the top of the stack

From multilayer networks...



...to IP/MPLS networks

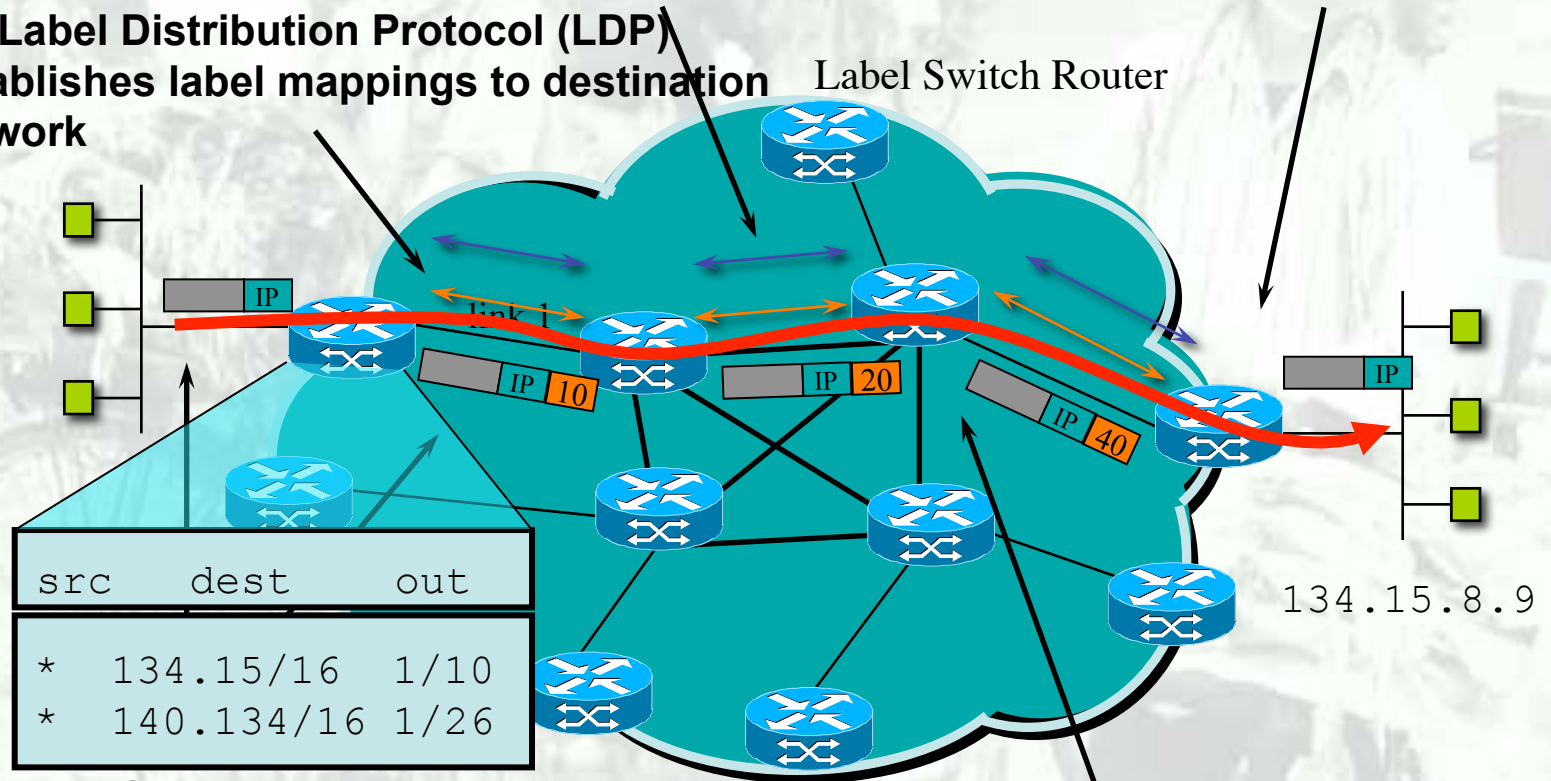


MPLS operation

1a. Routing protocols (e.g. OSPF-TE, IS-IS-TE) exchange reachability to destination networks

1b. Label Distribution Protocol (LDP) establishes label mappings to destination network

4. LSR at egress removes label and delivers packet

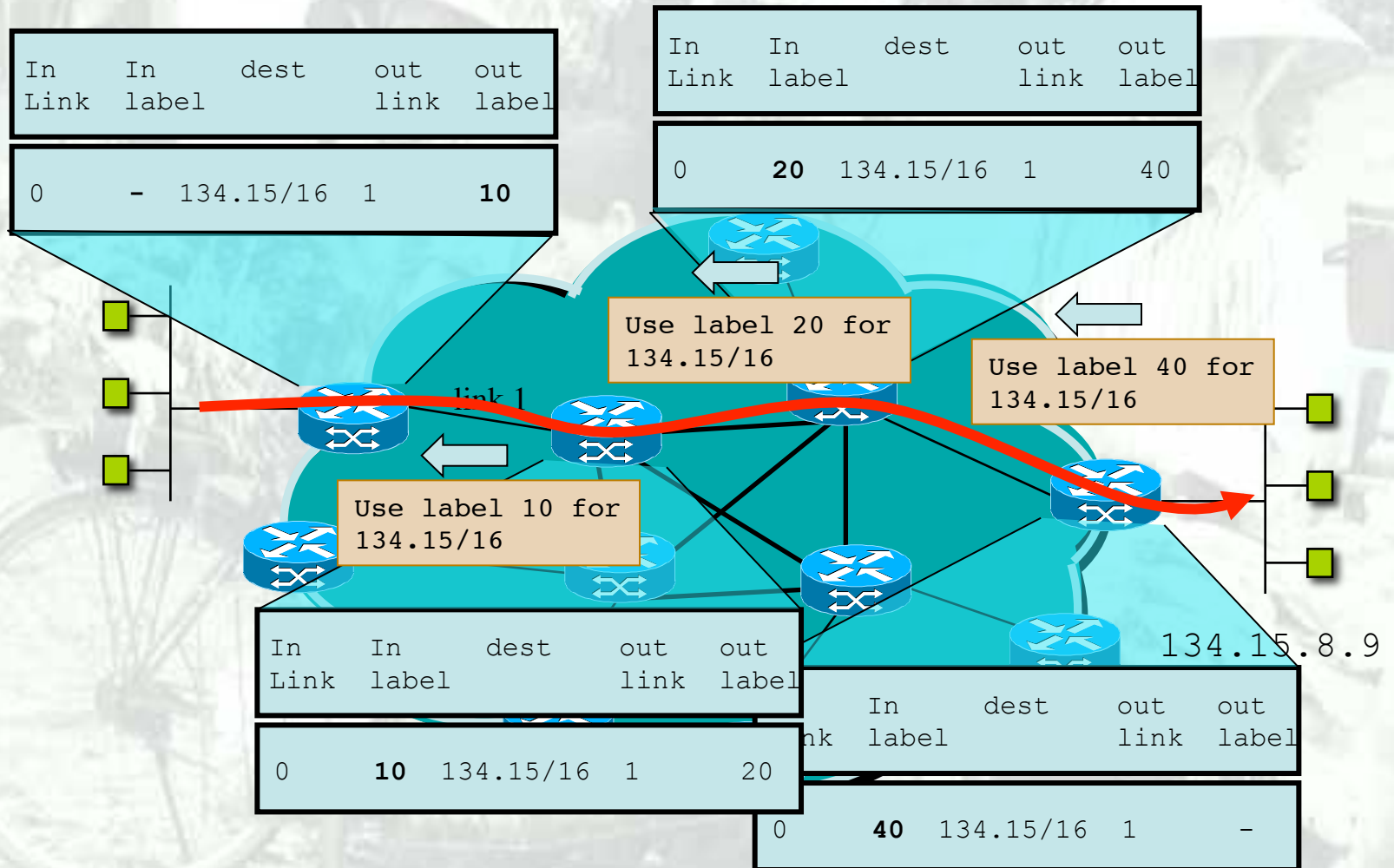


2. Ingress LSR receives packet and "label"s packets

Source Yi Lin, modified C. Pham

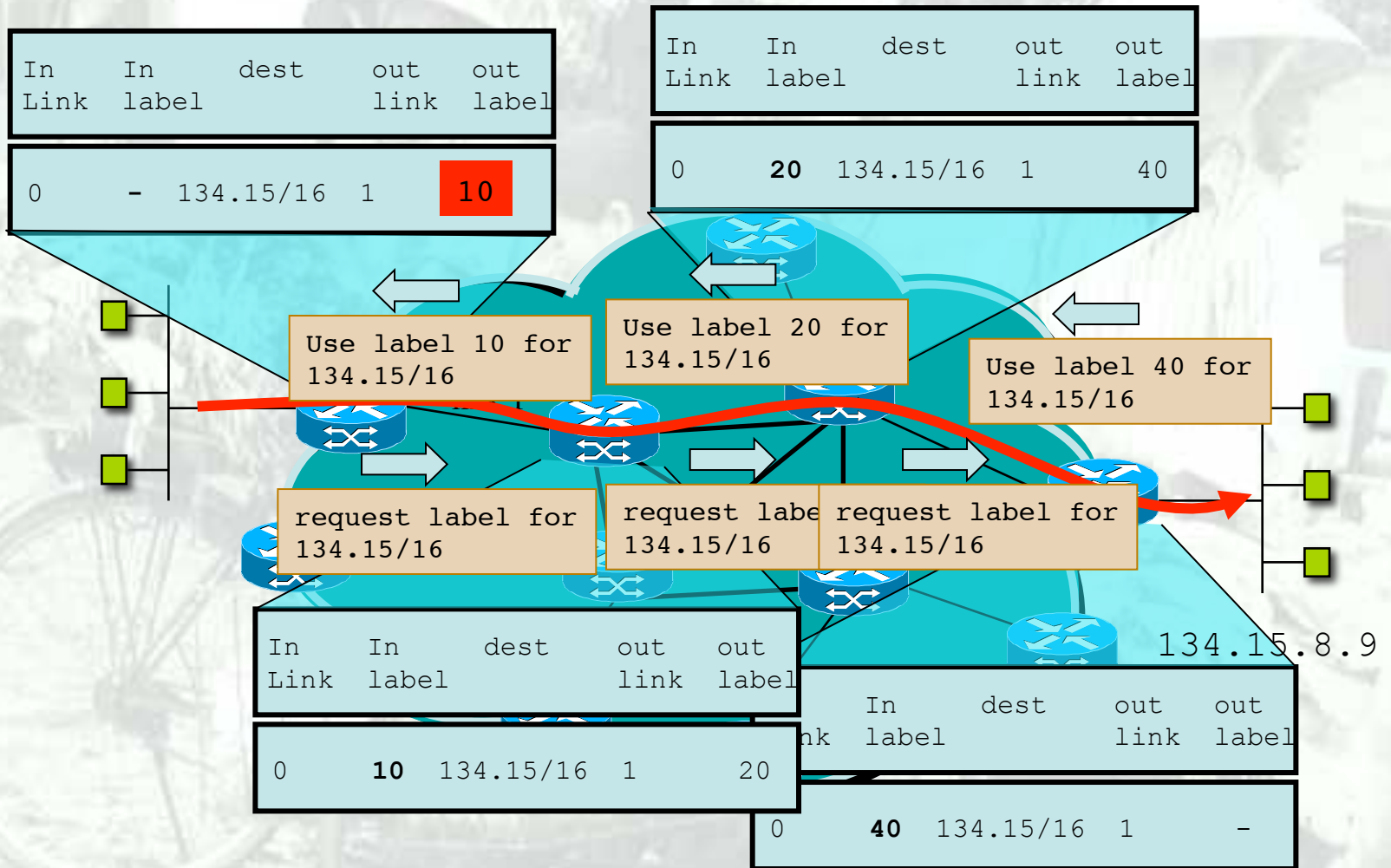
3. LSR forwards packets using label switching

Label Distribution



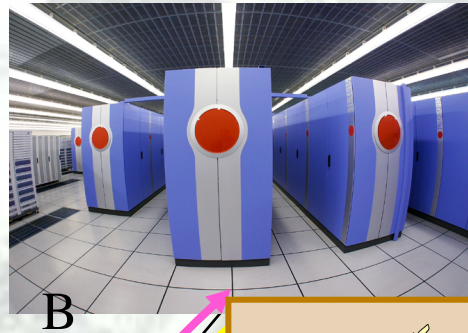
Unsolicited downstream label distribution

Label Distribution (con't)



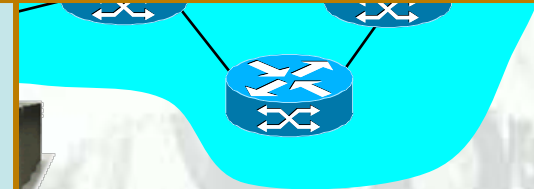
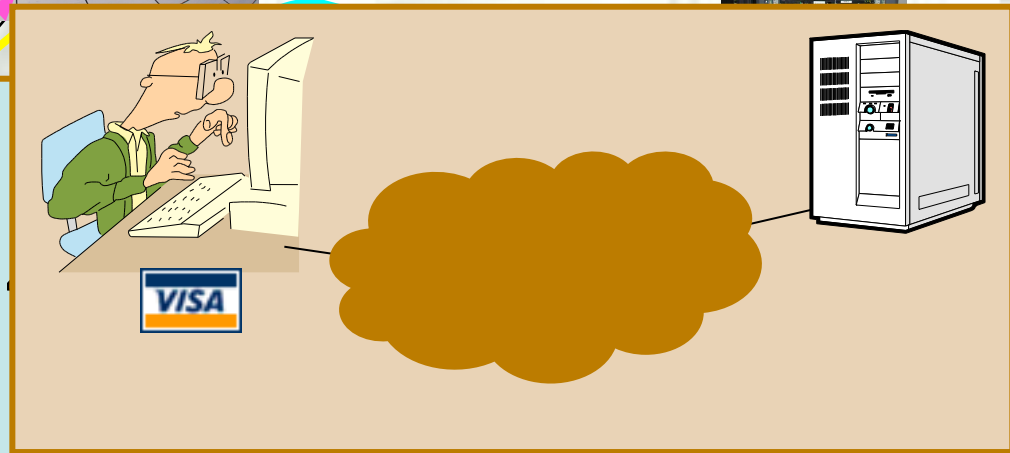
On-demand downstream label distribution

Dynamic circuits for grids

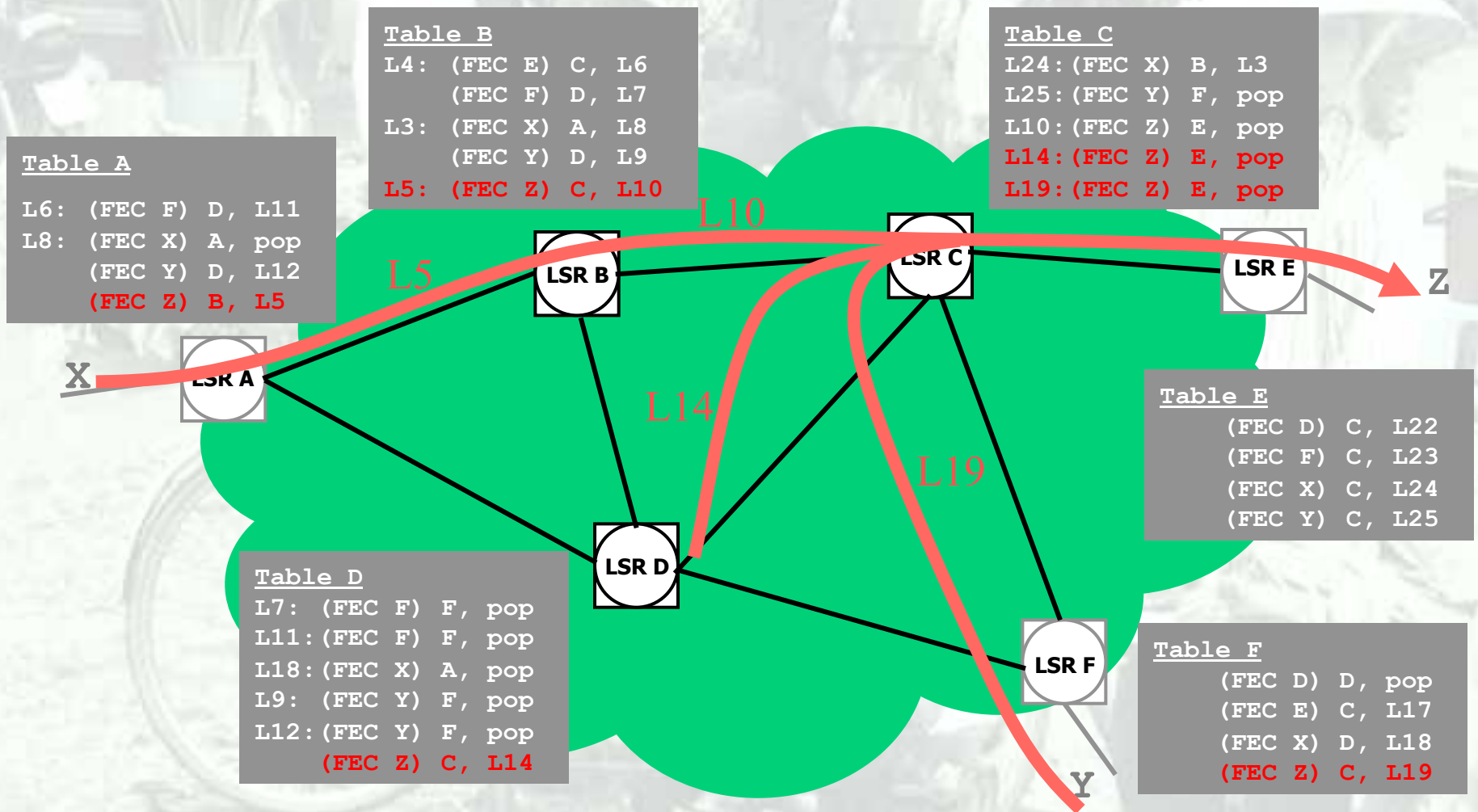


I need 2.5 Gbps between:

- A & B
- B & C
- D & C
- E & A

A cartoon character with a large nose is talking on a red phone. He is wearing a grey suit and a blue tie. A speech bubble is coming from him, containing the text above.

Forwarding Equivalent Class: high-level forwarding criteria



Forwarding Equivalent Class

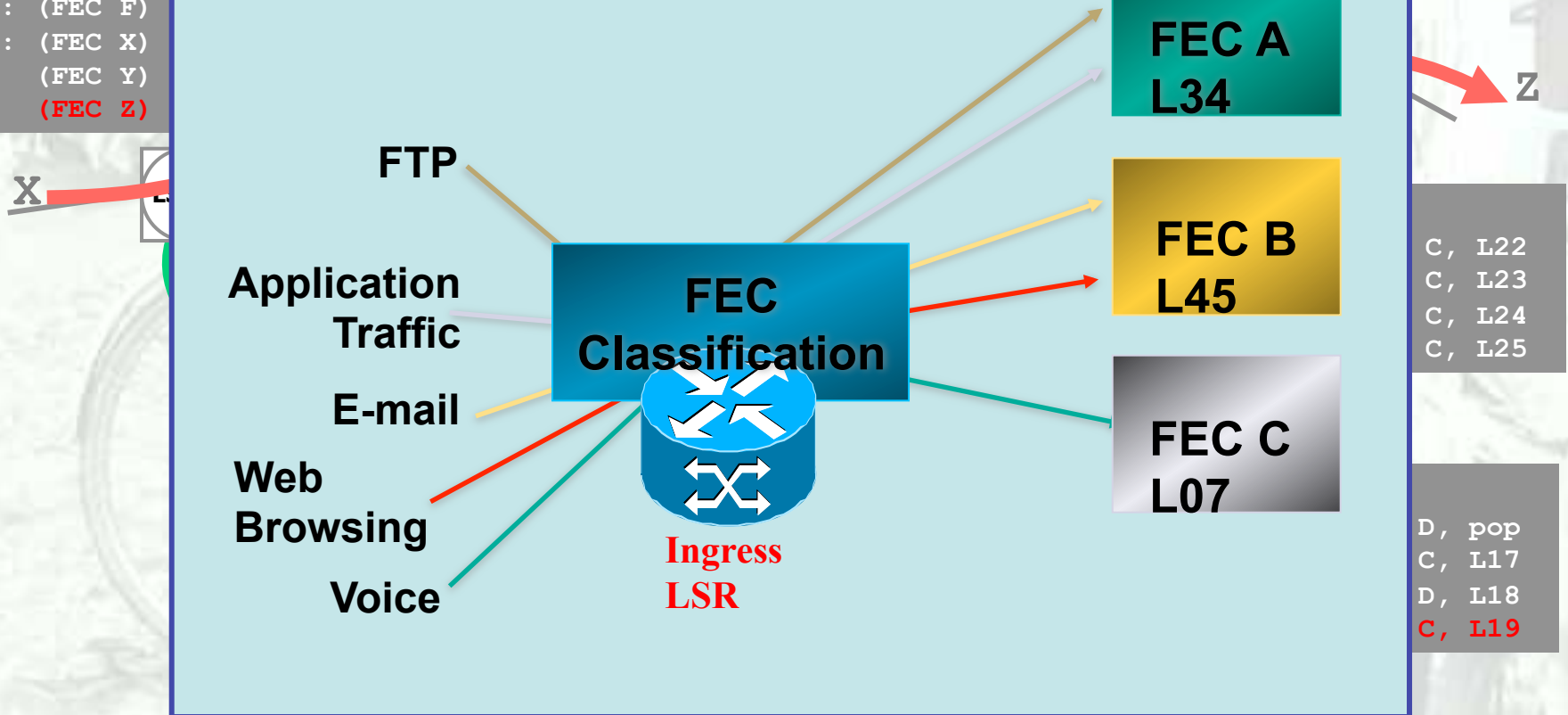
A FEC aggregates a number of individual flows with the same characteristics: IP prefix, router ID, delay or bandwidth constraints...

) B, L3
) F, pop

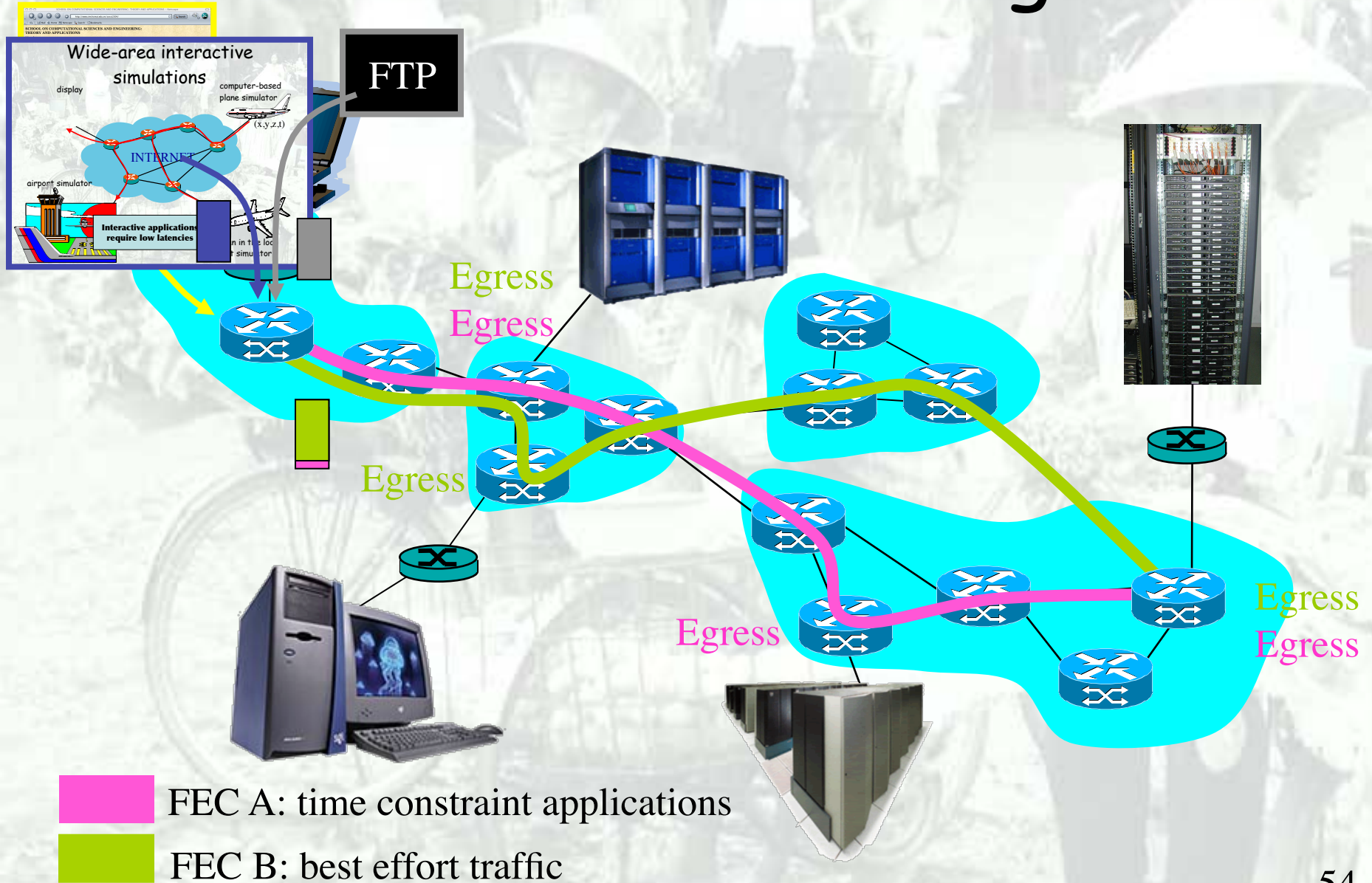
Table A

L6: (FEC F)
L8: (FEC X)
(FEC Y)
(FEC Z)

One possible utilization of FEC



MPLS FEC for the grid



Label & FEC

- ❑ Independent LSP control
 - ❑ An LSR binds a label to a FEC, whether or not the LSR has received a label from the next-hop for the FEC
 - ❑ The LSR then advertises the label to its neighbor

- ❑ Ordered LSP control
 - ❑ An LSR only binds and advertises a label for a particular FEC if:
 - it is the egress LSR for that FEC or
 - it has already received a label binding from its next-hop

Label Distribution Protocols

- ❑ LDP

- Maps unicast IP destinations into labels

- ❑ RSVP-TE, CR-LDP

- Used in traffic engineering

- ❑ BGP

- External labels (VPN)

- ❑ PIM

- For multicast states label mapping

MPLS for resiliency

MPLS FastReroute

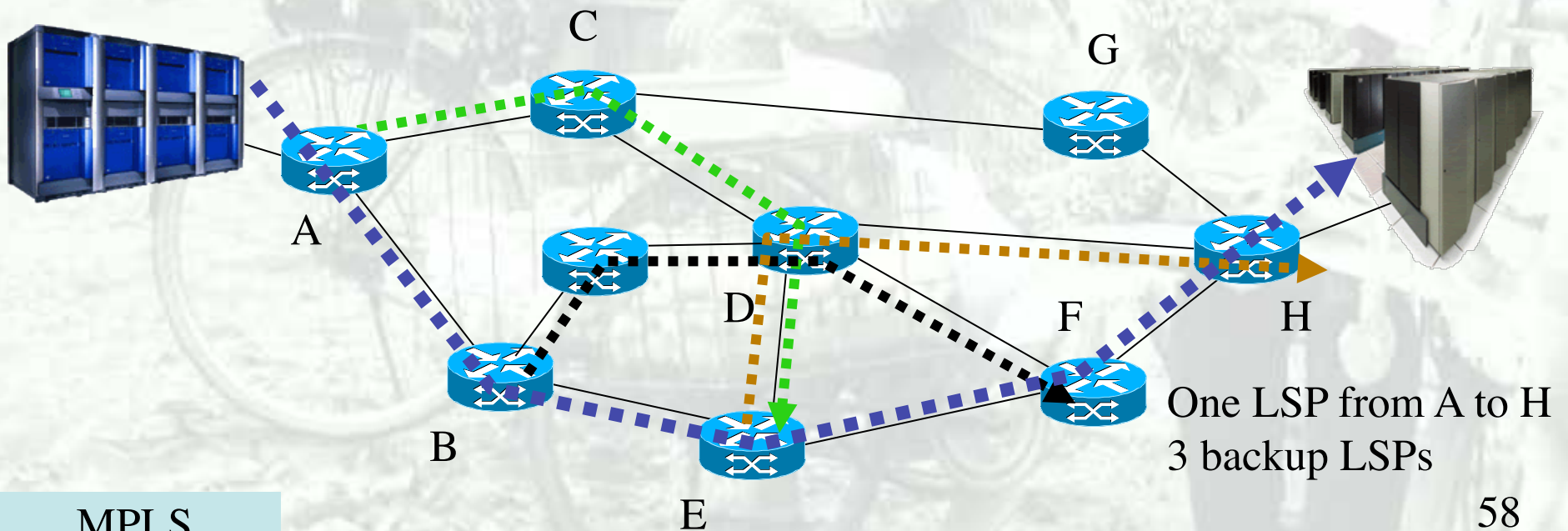
- ❑ Intended to provide SONET/SDH-like healing capabilities
- ❑ Selects an alternate route in tenth of ms, provides path protection
- ❑ Traditional routing protocols need minutes to converge!
- ❑ FastReroute is performed by maintaining backup LSPs

MPLS for resiliency, con't

Backup LSPs

- One-to-one

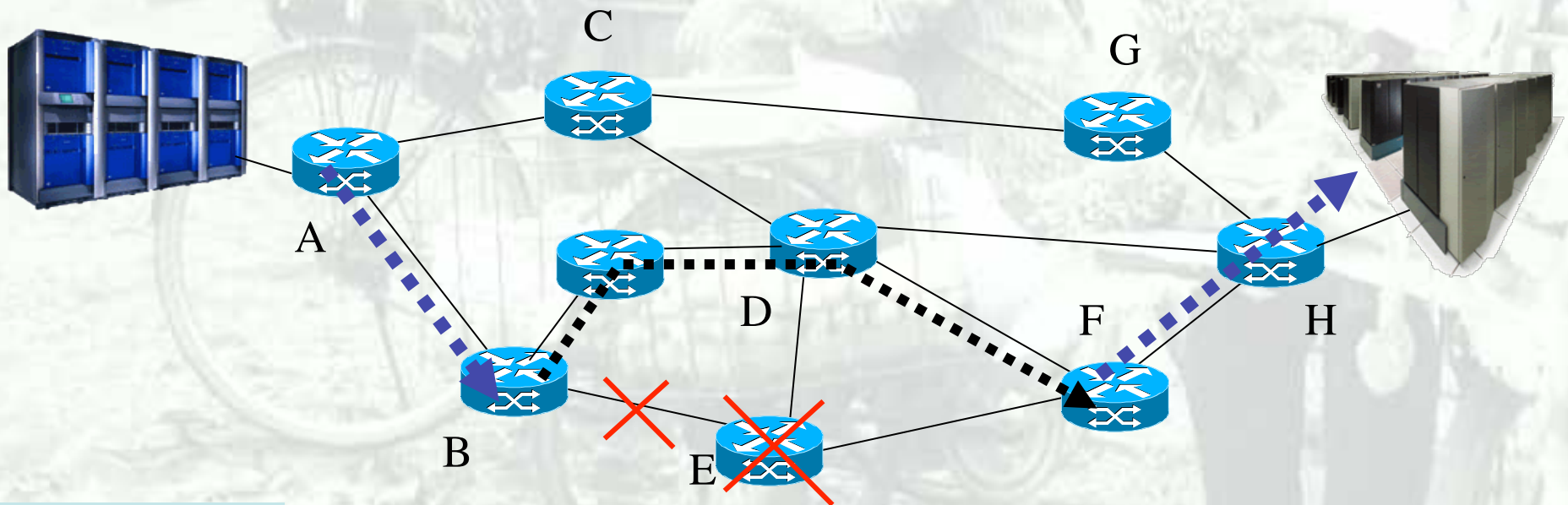
- Many-to-one: more efficient but needs more configurations



MPLS for resiliency, con't

Recovery on failures

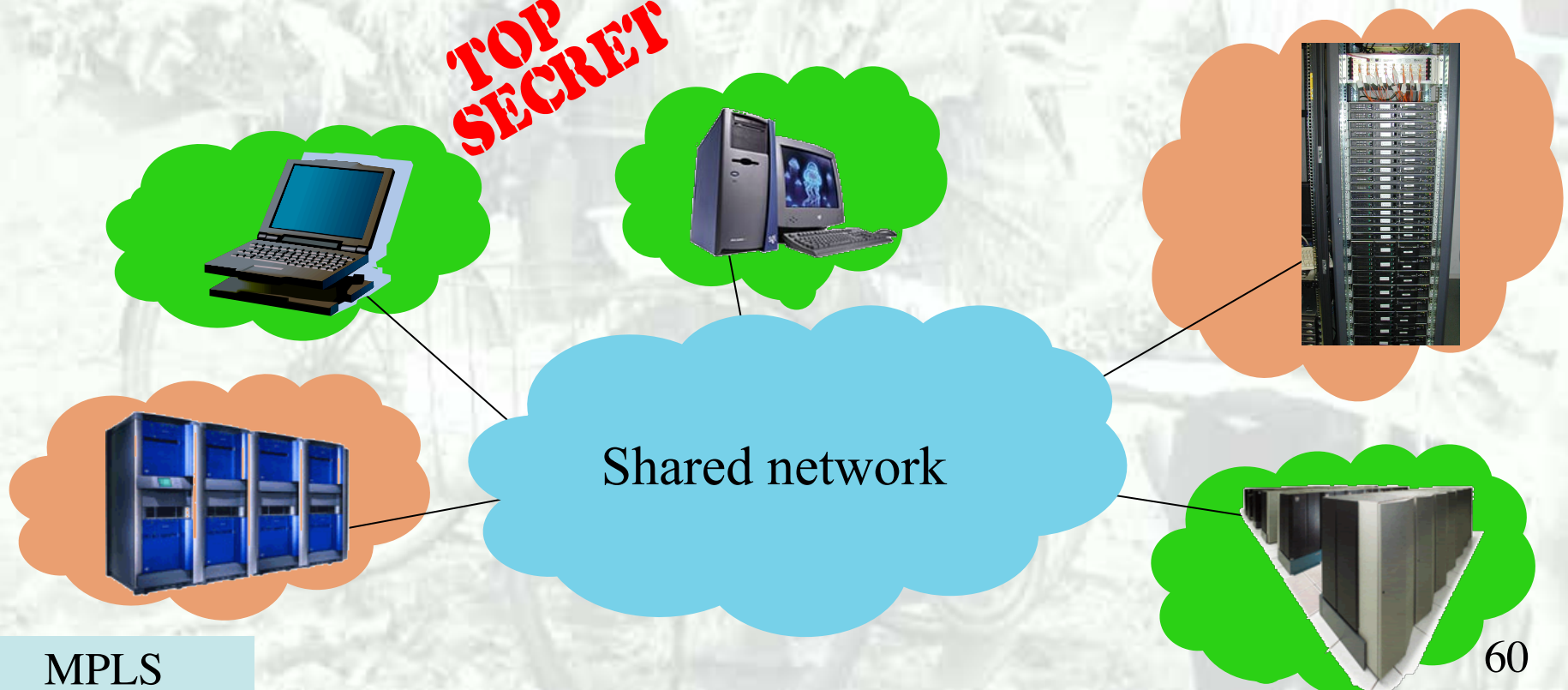
- ❑ Suppose E or link B-E is down...
- ❑ B uses detour around E with backup LSP



MPLS for VPN (Virtual Private Networks)

- ❑ **Virtual Private Networks:** build a secure, confidential communication on a public network infrastructure using routing, encryption technologies and controlled accesses

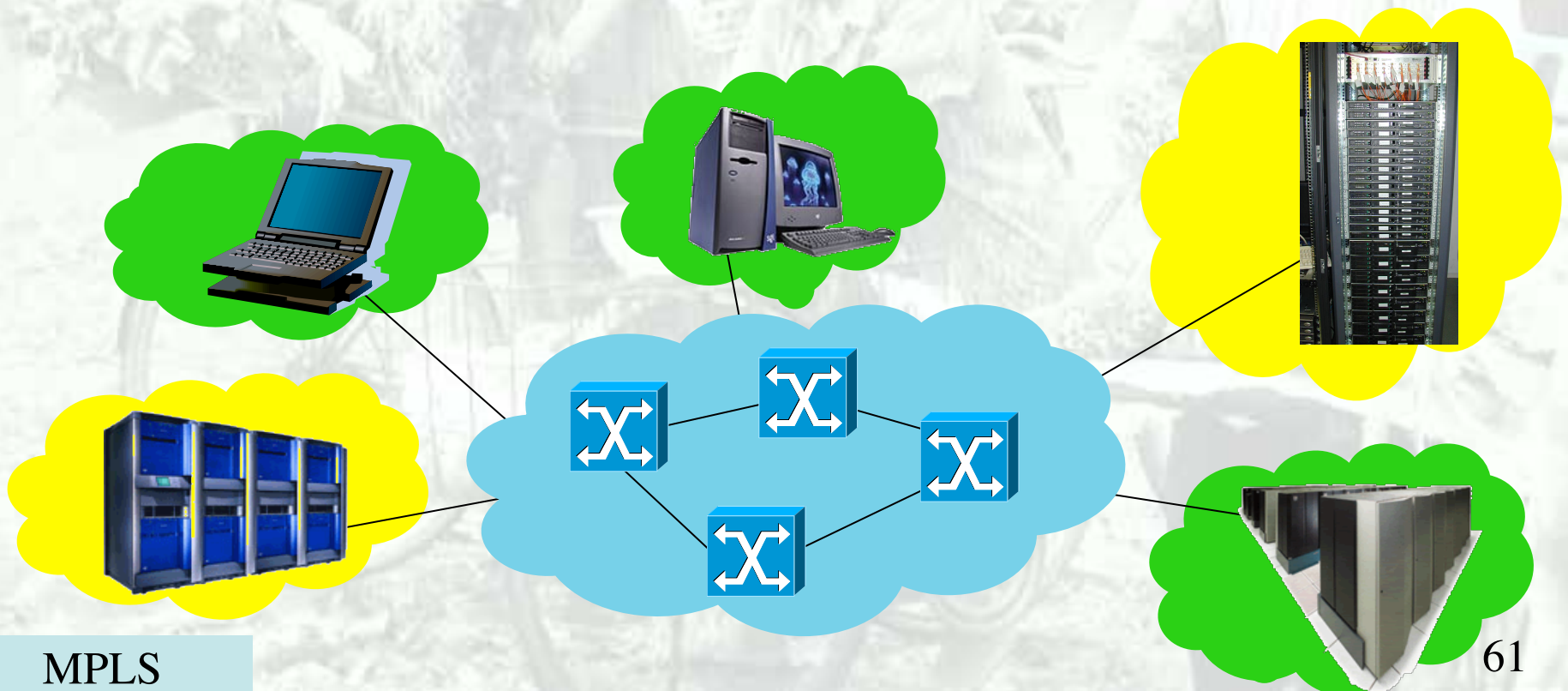
TOP SECRET



MPLS for VPN, con't

The traditional way of VPN

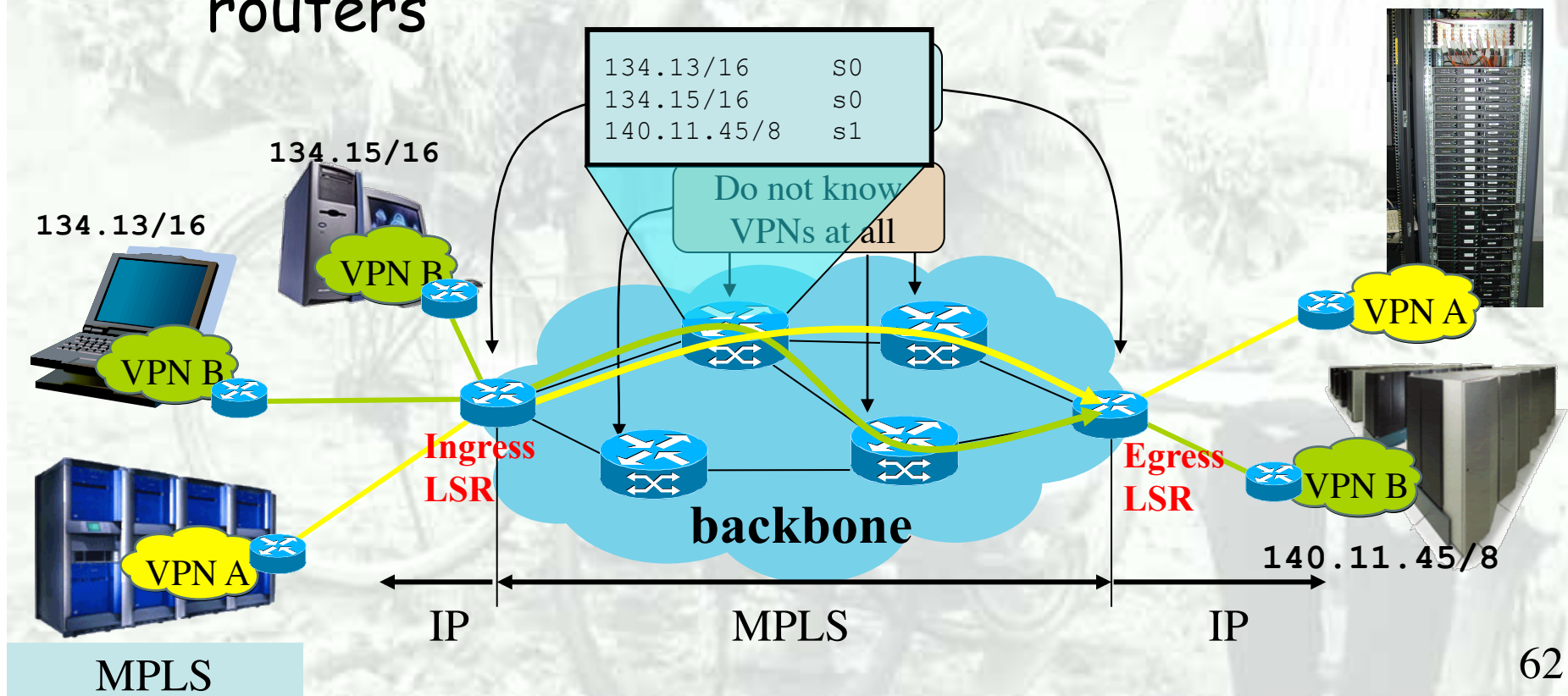
- ❑ Uses leased lines, Frame Relay/ATM infrastructures...



MPLS for VPN, con't

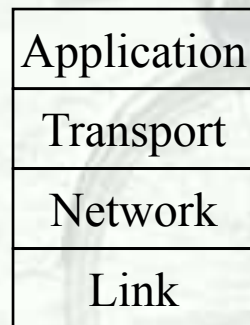
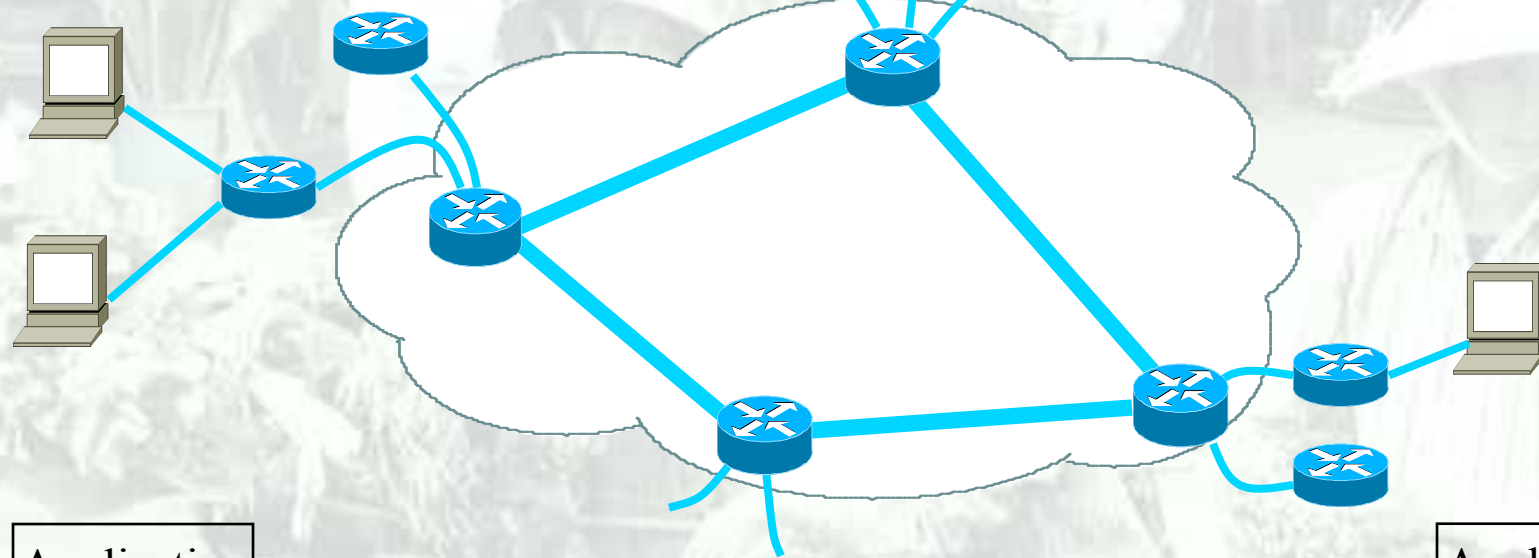
VPN over IP/MPLS

- ❑ IP/MPLS replace dedicated networks
- ❑ MPLS reduces VPN complexity by reducing routing information needed at provider's routers

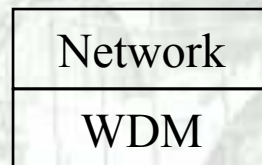


MPLS for optical networks

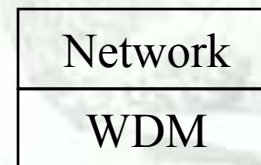
Before MPLS



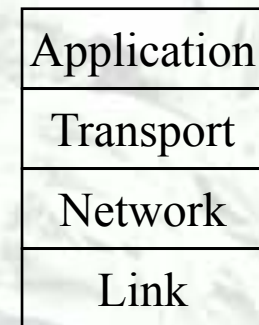
Terminals



IP router



IP router

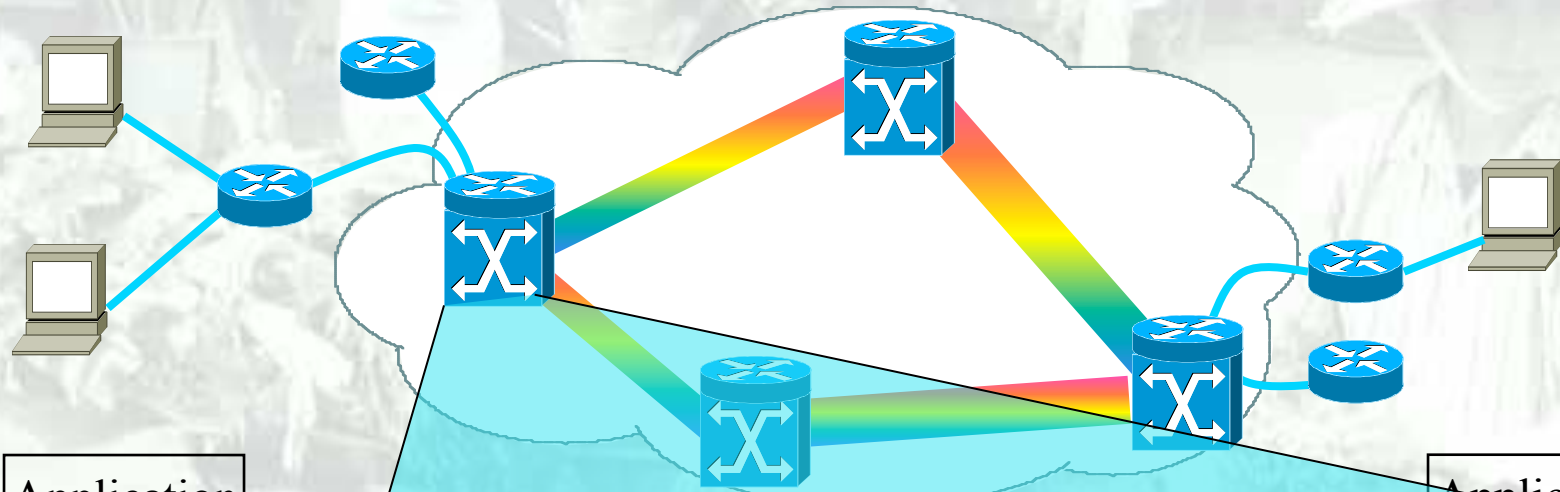


Terminals

Source J. Wang, B. Mukherjee, B. Yoo

MPLS for ON, con't

$MP\lambda S = MPLS + \lambda$ lightpath

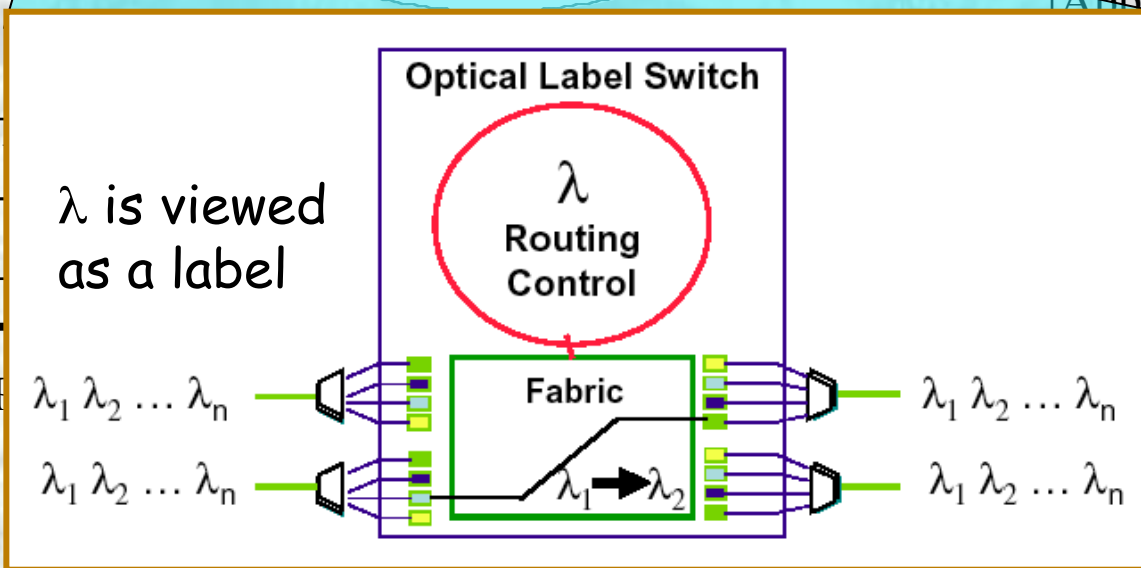


Application
Transport
Network
Link

Application
Transport
Network
Link

Terminals

MI



inals

MPLS for ON, con't

GMPLS

- ❑ GMPLS stands for “Generalized Multi-Protocol Label Switching”
- ❑ Extends the concept of MPLS beyond data networks to address legacy transport networks
- ❑ Reduce OPEX cost for operators
- ❑ A suite of protocols that provides a common set of control functions for disparate transport technologies (IP, ATM, SONET/SDH, DWDM)
- ❑ Hot issue at IETF!

MPLS for ON, con't

GMPLS control plane

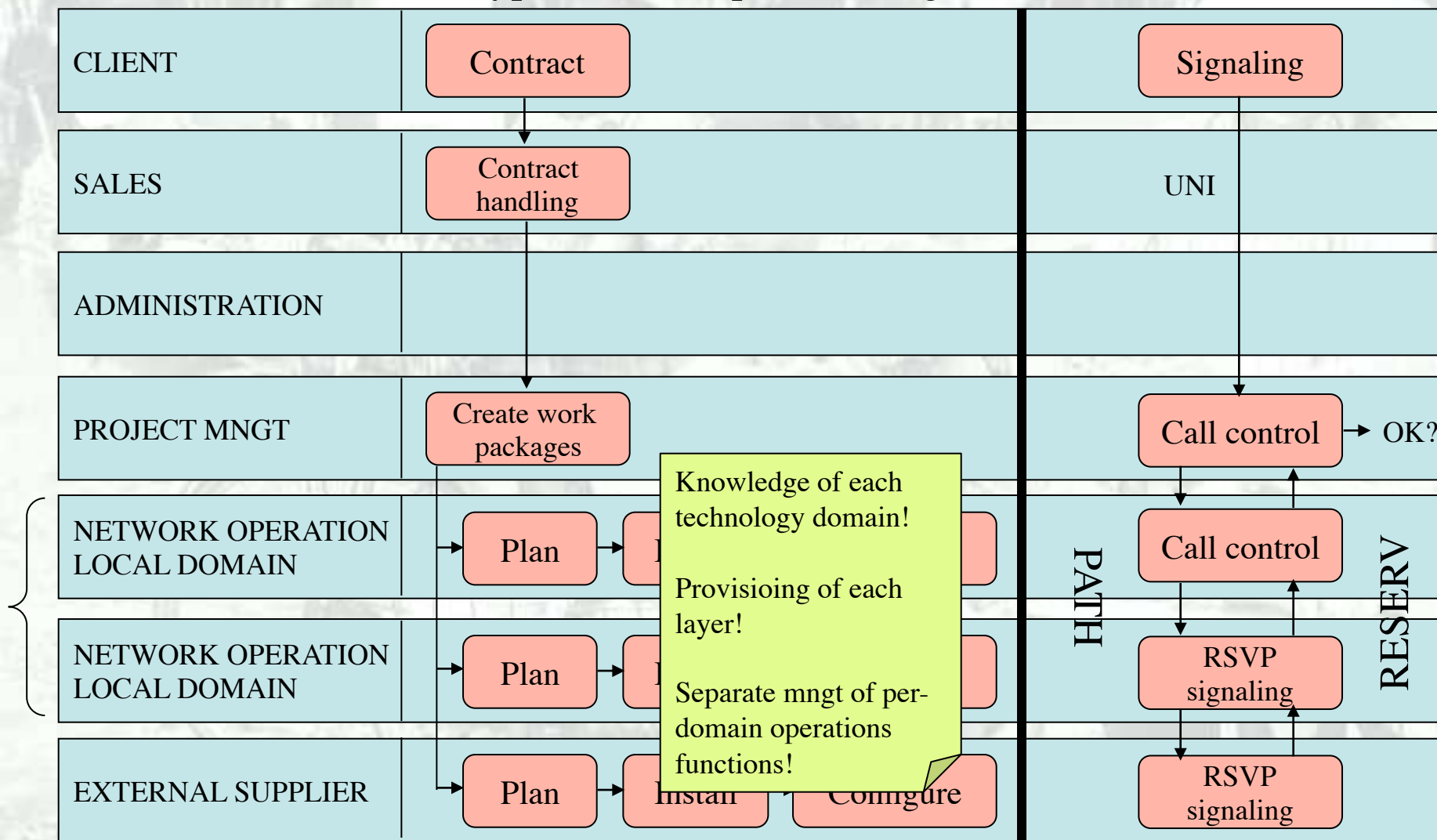
LINK MANAGEMENT: Link Management Protocol (LMP)	<ul style="list-style-type: none">-Neighbor discovery-Maintain control channel connectivity-Verify data link connectivity-Correlate link property information-Suppress downstream alarms-Localize link failures
ROUTING: Open Shortest Path First-Traffic Engineering (OSPF-TE)	<ul style="list-style-type: none">-Distribute TE link information-Advertise nodes in the network and create topology-Calculate constrained shortest path (CSPF)-Routing information for control and data plane
SIGNALING: Resource ReserVation Protocol-Traffic Engineering (RSVP-TE)	<ul style="list-style-type: none">-Signals setup/teardown/refresh of paths with QoS requirements (e.g., circuit size)-Uses control channel to setup an optical LSP-Supports refresh reduction-Supports Explicit Route Object (ERO) and Record Route Object (RRO)

Source S. Kinoshita, R. Rabbat, APNOMS 2005

Ex: Service Provisioning

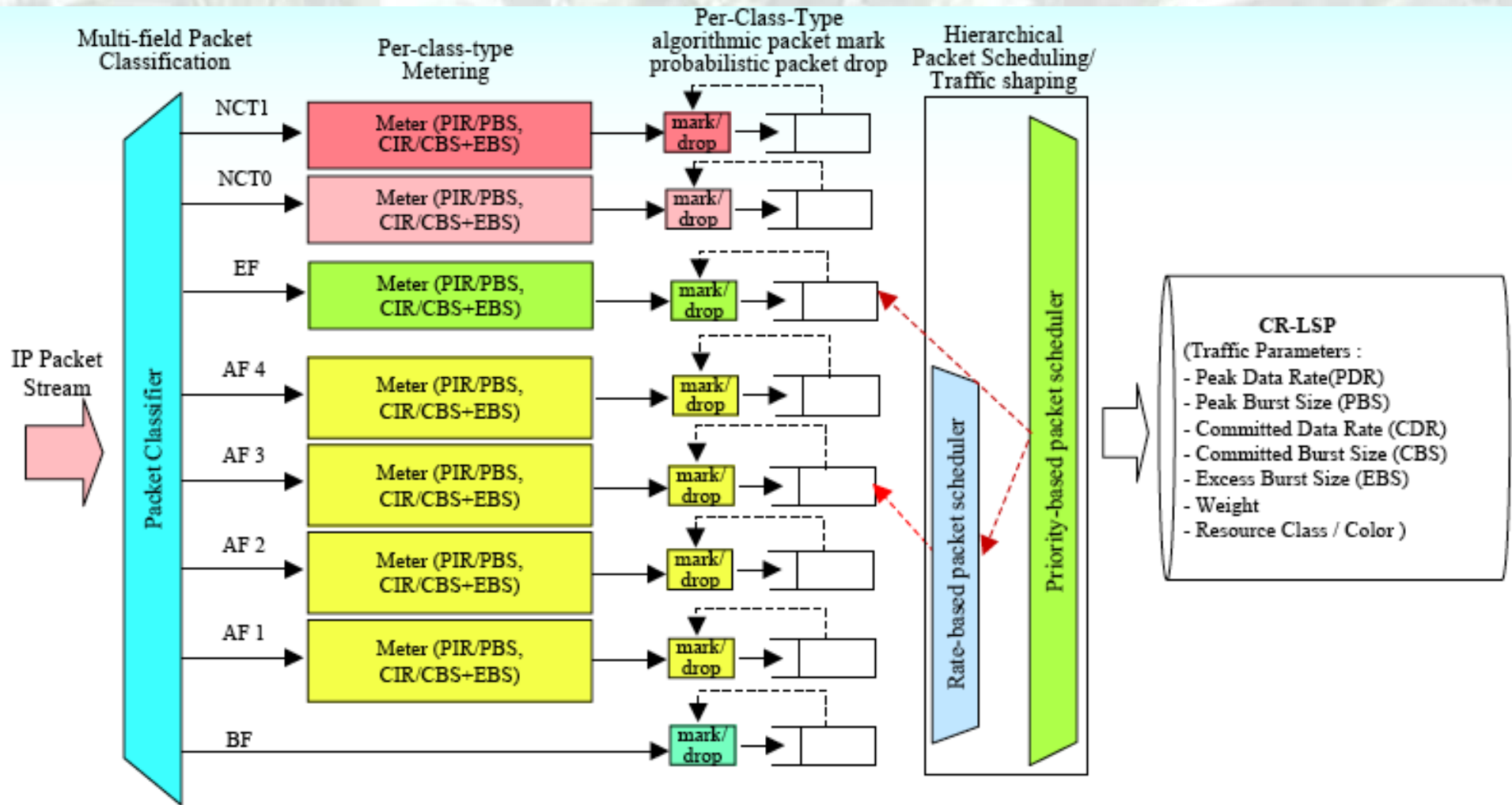
Typical service provisioning

With GMPLS

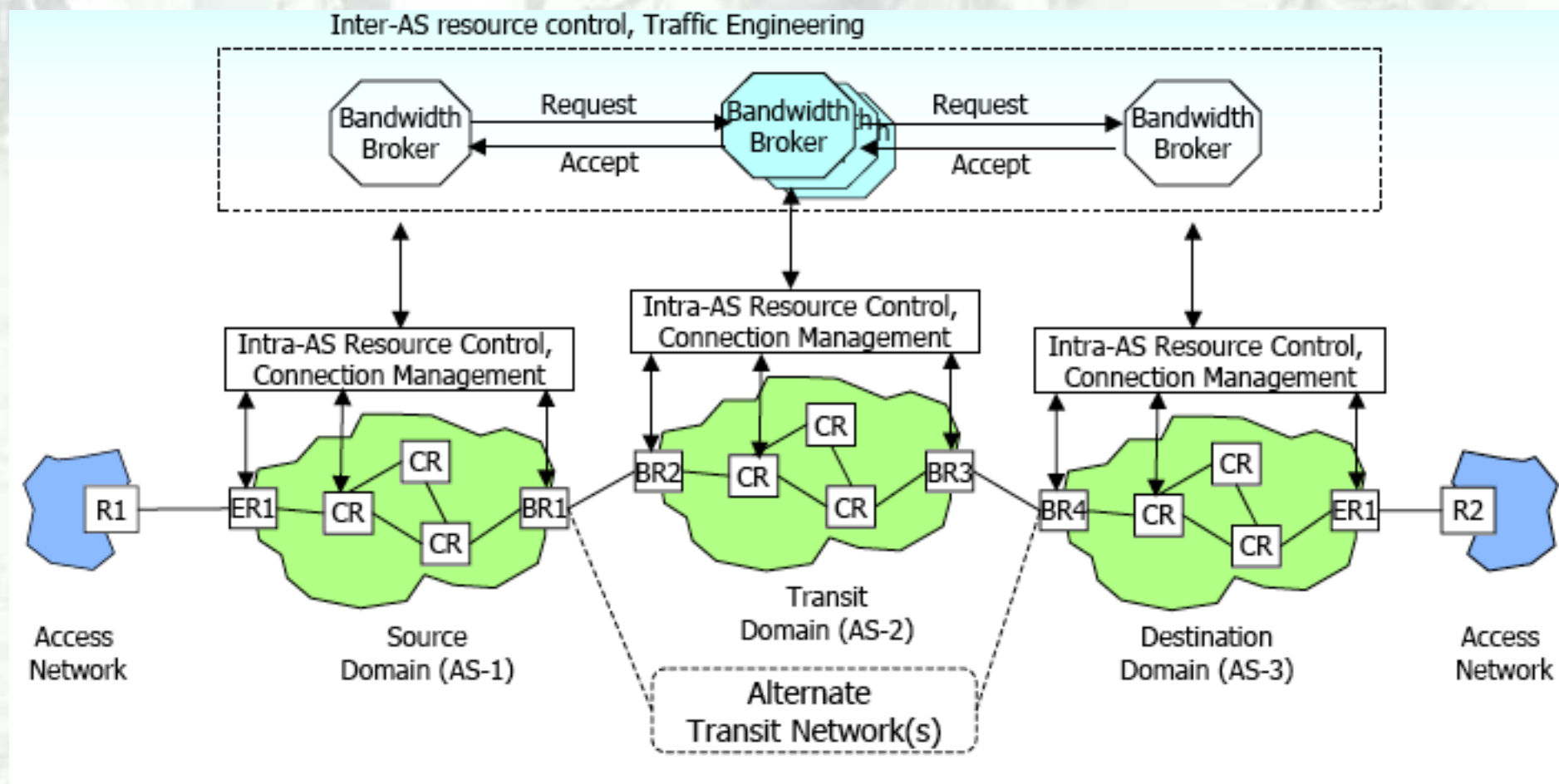


DiffServ over (G)MPLS

map DiffServ class on MPLS FEC



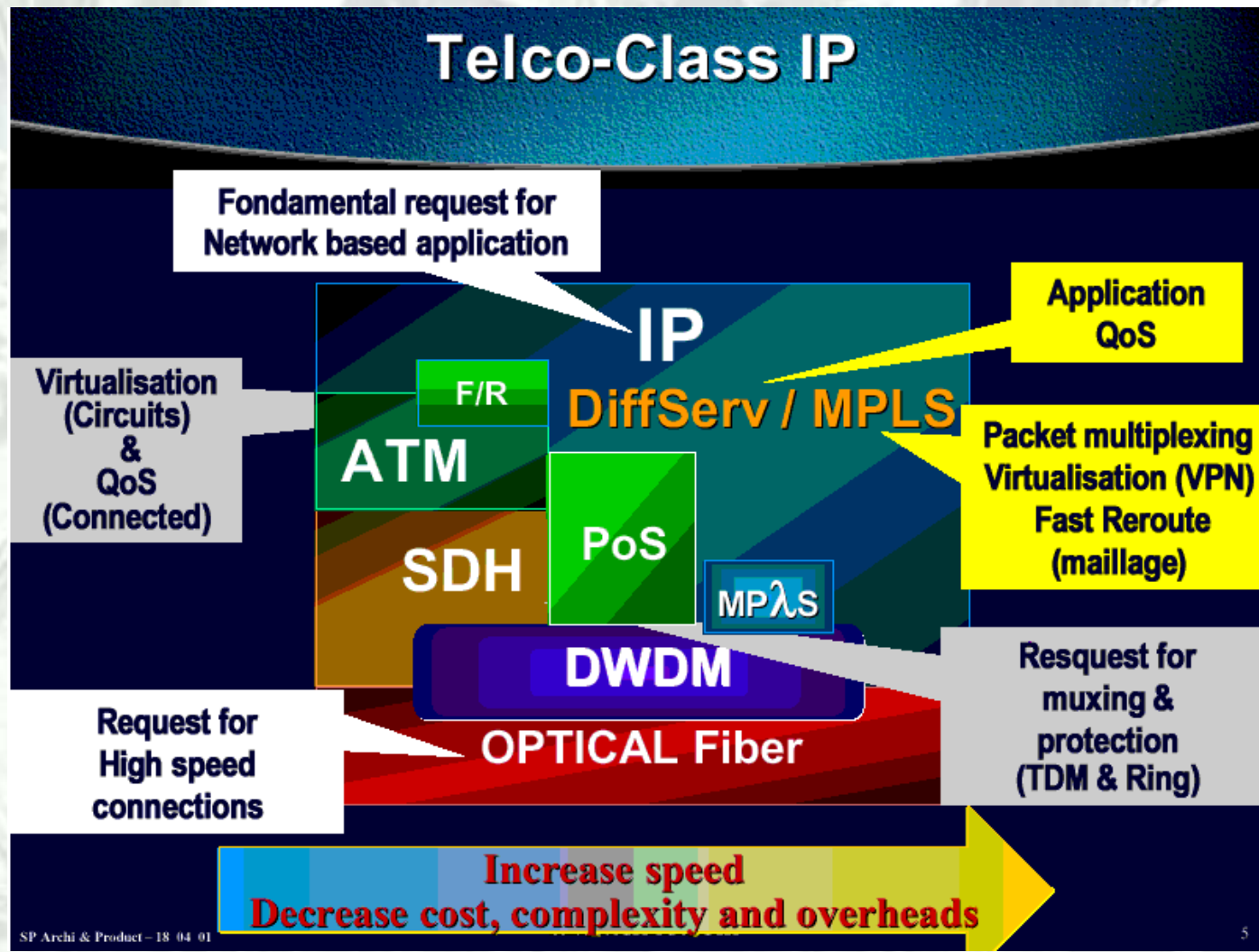
Some words on inter-domain



Summary

Towards IP/(G)MPLS/DWDM

From cisco

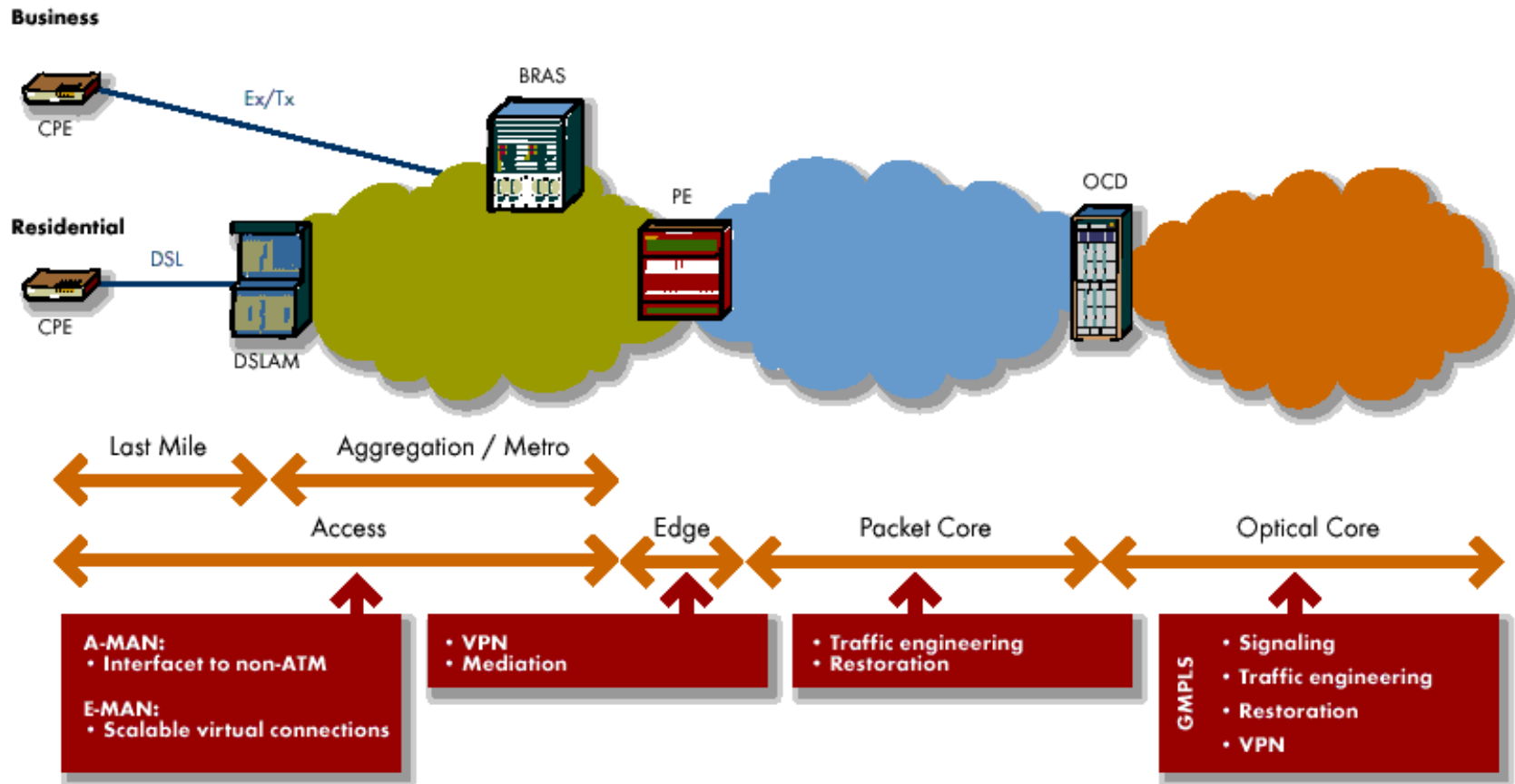


Summary

Technology scope

Fig. 1 New MPLS applications and application areas

Source Alcatel



A-MAN: ATM Metropolitan Area Network
BRAS: Broadband Remote Access Server
CPE: Customer Premises Equipment

DSLAM: Digital Subscriber Line Access Multiplexer
Ex/Tx: E1/T1 or E3/T3
OCD: Optical Core Device
PE: Provider Edge

Want to know more?

- ❑ GMPLS: IEEE Comm. Mag., Vol. 43(7), July 2005
- ❑ Optical Control Plane for the Grid Community: IEEE Comm. Mag., Vol. 44(3), March 2006.
- ❑ “Optical Transport Systems/Networks” by S. Kinoshita & R. Rabbat, APNOMS 2005. <http://www.apnoms.org/2005/tutorial/Tutorial%202.pdf>
- ❑ « Inter-domain Traffic Engineering for QoS-guaranteed DiffServ Provisioning », Young-Tak Kim, APNOMS 2005.
<http://www.apnoms.org/2005/tutorial/Tutorial%203.pdf>
- ❑ See Tutorial IV of HOTI 2006: Dynamic Optimal Networks for Grid Computing

End of part 1, go to part 2

