UPC

Facultat d' informàtica de Barcelona

Capítol 3. Xarxes troncals

3.1 Commutació de trames

3.2 Commutació de cel·les

3.3 Commutació d'etiquetes

3.4 Carrier Ethernet

3.5 Control de la congestió

3.1 Commutació de trames

LAPF Core

Flag	Address	Information	FCS	Flag
<---1----> octet	<-------2-4------>	<--------------Variable--------------->	<--------2------->	<---1-->

(a) Frame format

8	7	6	5	4	3	2	1
Upper DLCI						C/R	EA 0
Lower DLCI				FECN	BECN	DE	EA 1

(b) Address field - 2 octets (default)

8	7	6	5	4	3	2	1
Upper DLCI						C/R	EA 0
DLCI				FECN	BECN	DE	EA 0
Lower DLCI or DL-CORE control						D/C	EA 1

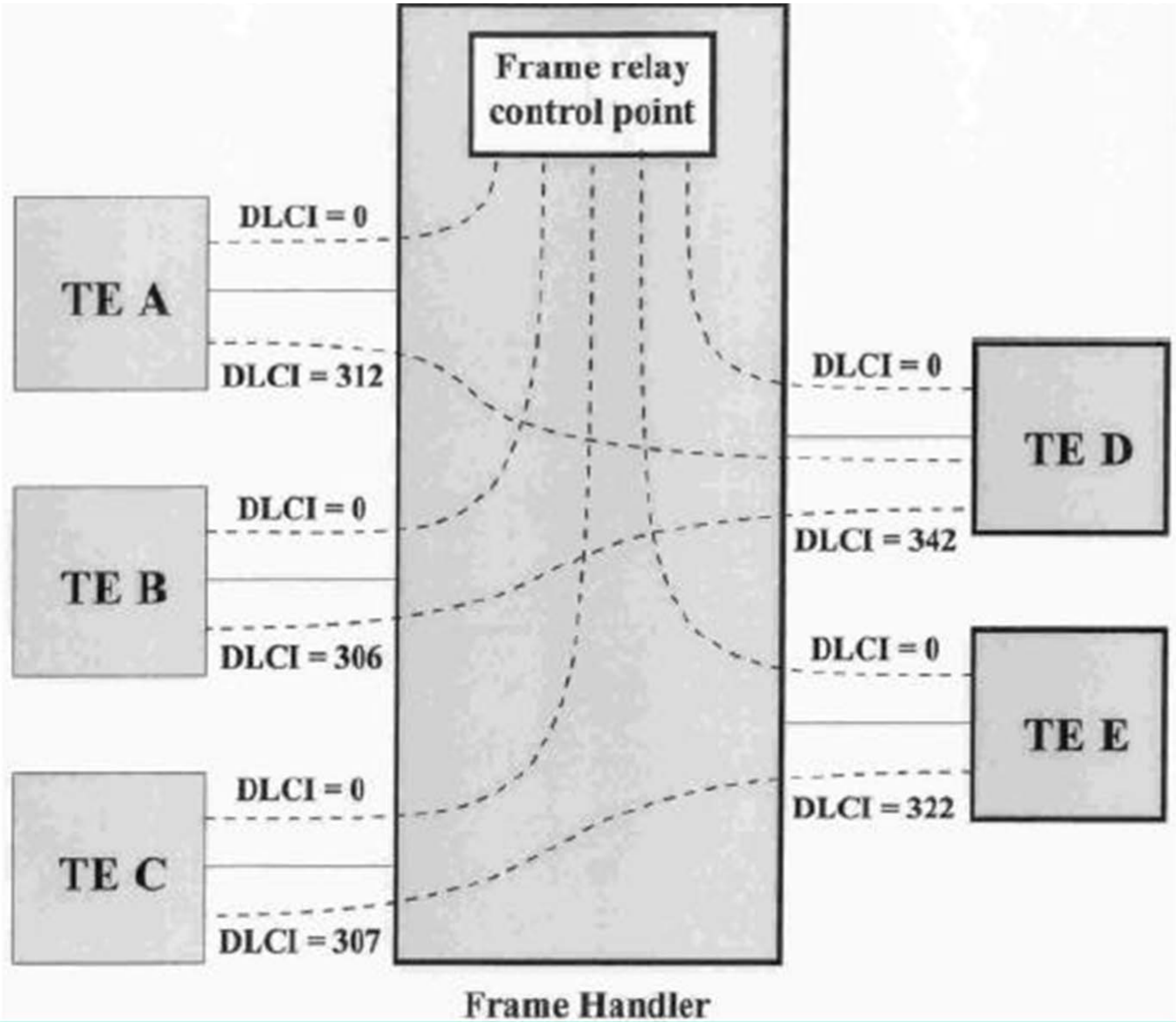
(c) Address field - 3 octets

8	7	6	5	4	3	2	1
Upper DLCI						C/R	EA 0
DLCI				FECN	BECN	DE	EA 0
DLCI							EA 0
Lower DLCI or DL-CORE control						D/C	EA 1

(d) Address field - 4 octets

EA	Address field extension bit
C/R	Command/response bit
FECN	Forward explicit congestion notification
BECN	Backward explicit congestion notification
DLCI	Data link connection identifier
D/C	DLCI or DL-CORE control indicator
DE	Discard eligibility

Data Link Connection

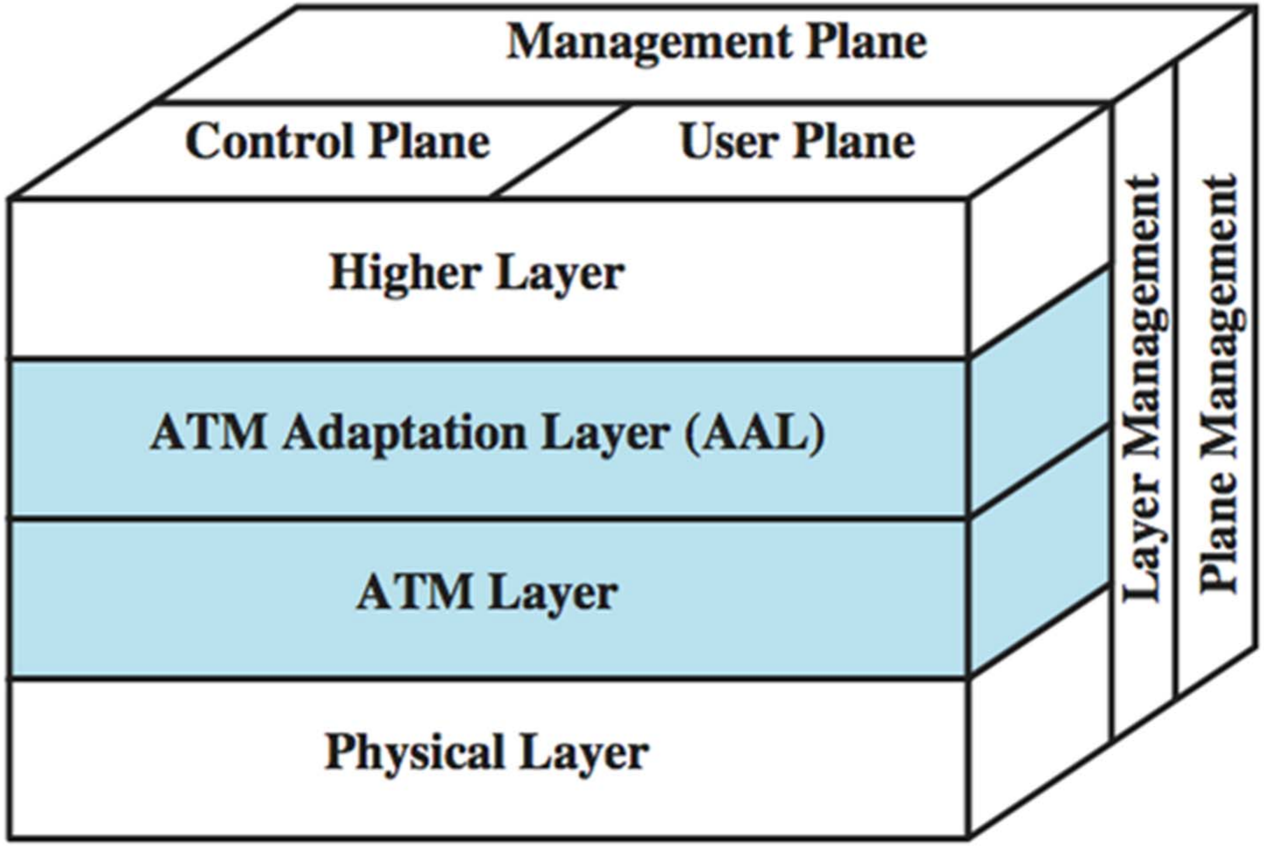


3.2 Commutació de cel·les

The Role of Asynchronous Transfer Mode (ATM)

- ATM uses packets called cells
- cells are small and fixed-length
- connection-oriented
- performance of a circuit-switching network and the flexibility and efficiency of a packet-switching network
- supports data, voice, video
- transmission based on priority and QoS

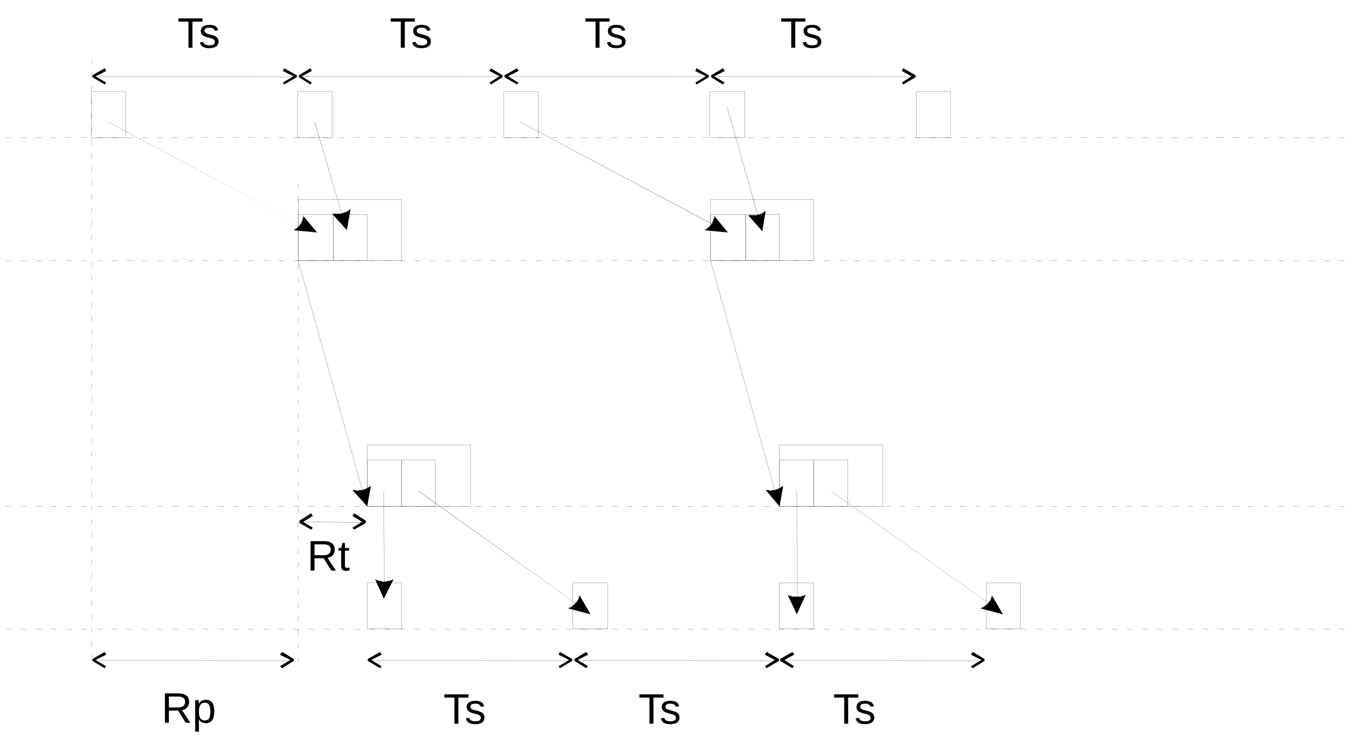
Protocol Architecture



Delay in ATM networks

- End to end delay R = Rp + Rt
- Rp (packet delay) = $48x8 / V_{ts}$
- Rt (transfer delay) = Tt + Tp + W
  - *Tt (transmission time)* Σt_t $(t_t = 53x8 / V_{tn})$
  - *Tp (propagation time)* Σt_p $(t_p = d / V_p)$
  - *W (queue waiting time)* Σw $(w = nxt_t)$

Packet delay



ATM Logical Connections

virtual channel connections (VCC)

- analogous to virtual circuit in X.25

basic unit of switching between two end users

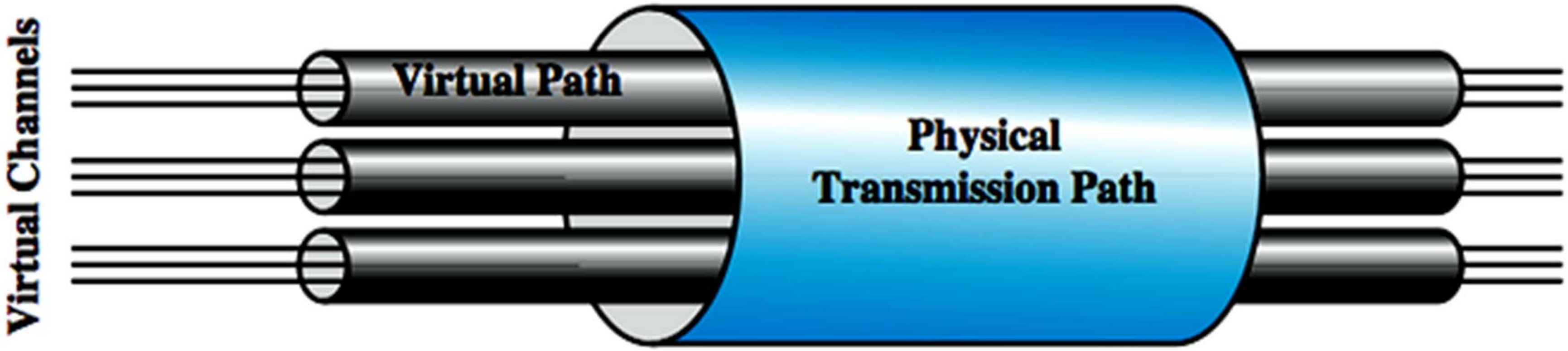
- variable rate
- full duplex
- fixed size cells

VCCs also used for

- user-network exchange (control signaling)
- network-network exchange (network management and routing)

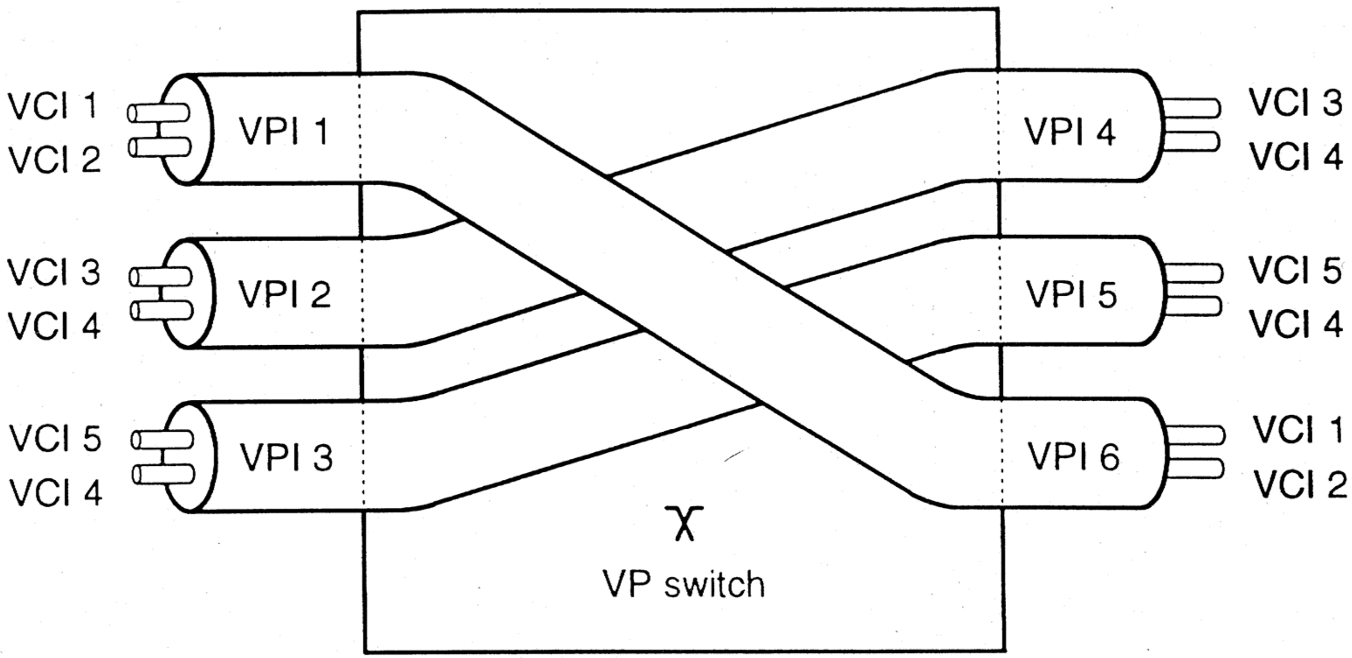
ATM Virtual Path Connection

- virtual path connection (VPC)
  - bundle of VCC with same end points



Virtual Channel and Virtual Path

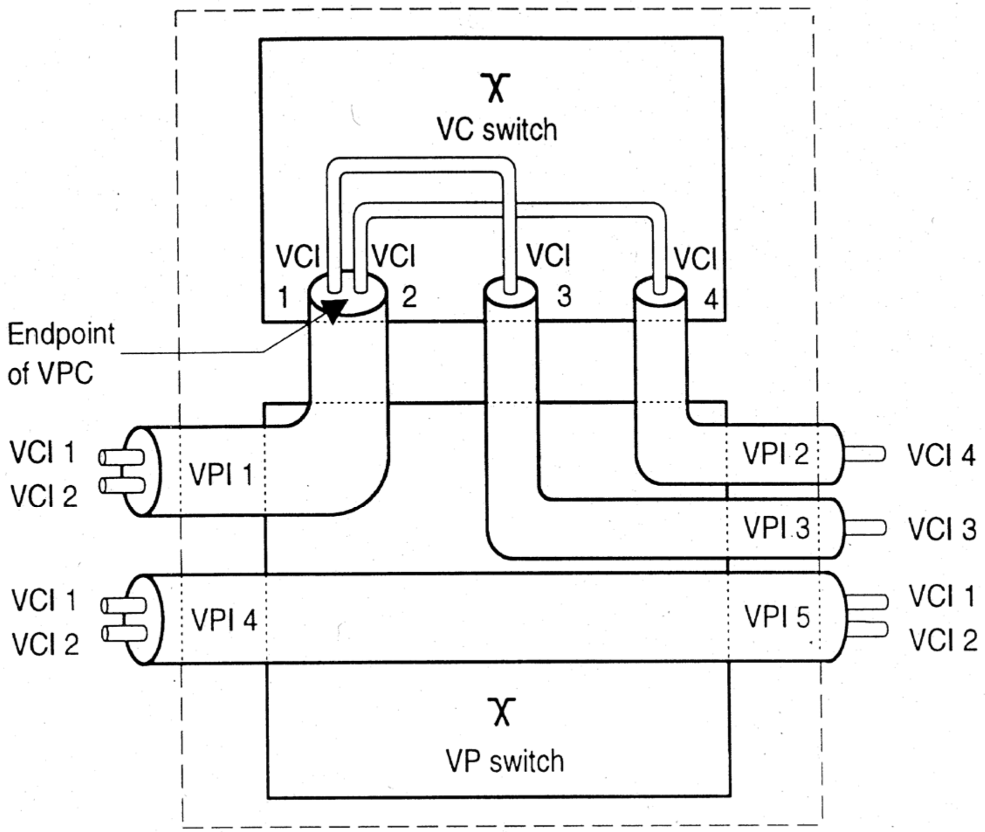
- Virtual Path switch



VCI Virtual channel identifier
VP Virtual path
VPI Virtual path identifier

Virtual Channel and Virtual Path

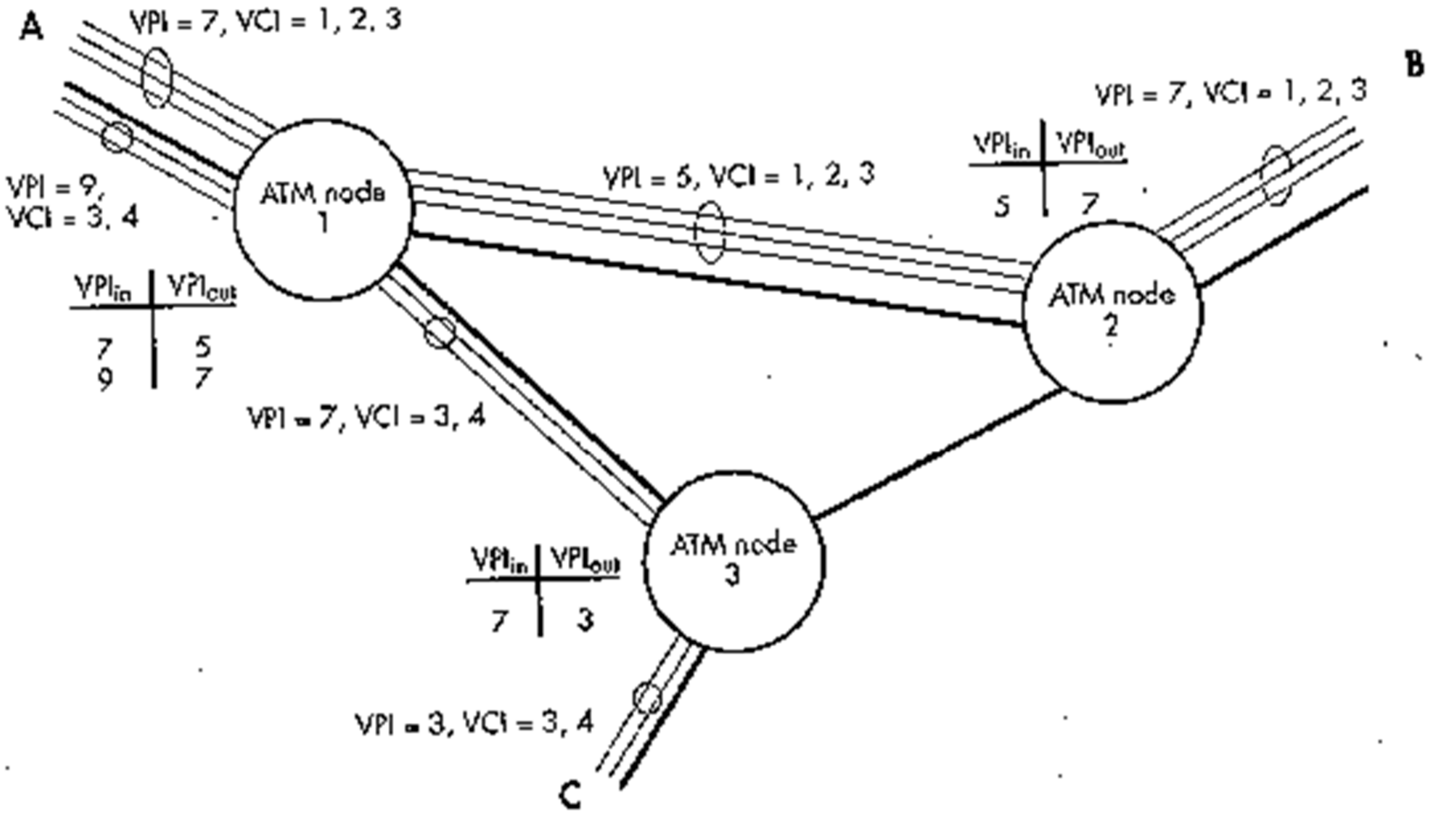
- Virtual Channel switch



VC	Virtual channel	VP	Virtual path
VCI	Virtual channel identifier	VPC	Virtual path connection
		VPI	Virtual path identifier

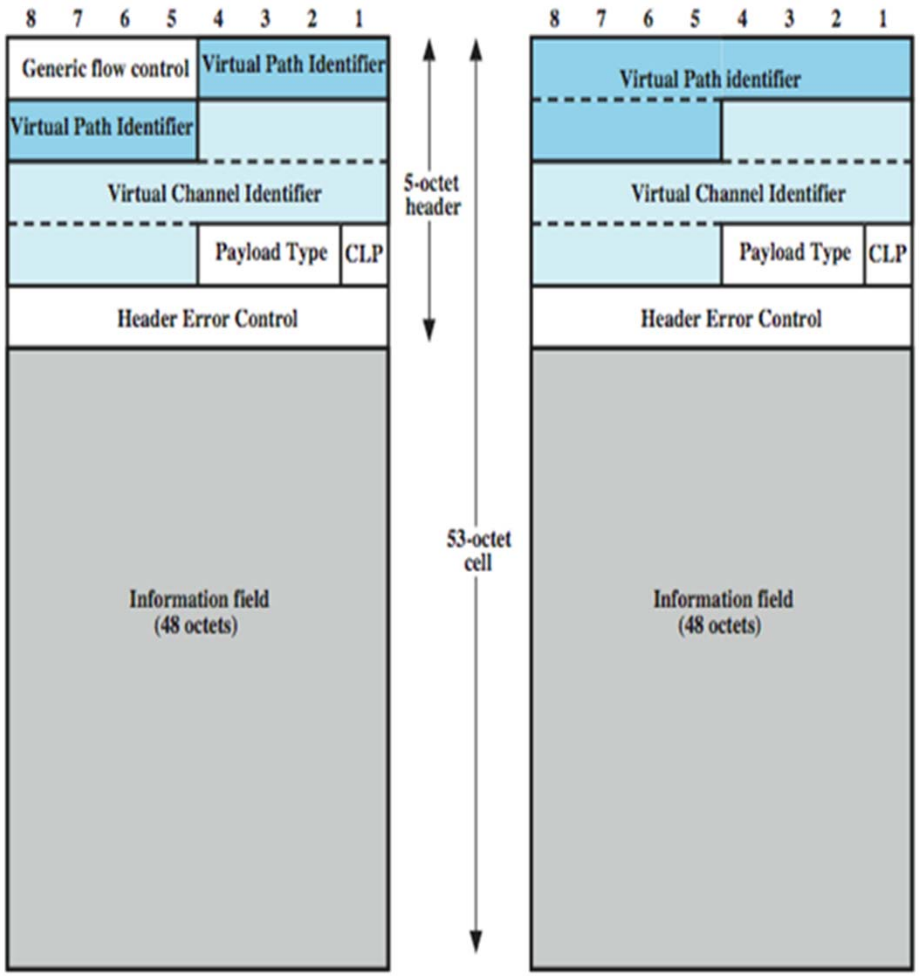
Virtual Channel and Virtual Path

- VPI and VCI relation



VPI Use in an ATM Network

ATM Cells



(a) User-network interface

(b) Network-network interface

ATM Header Fields

- generic flow control
- virtual path identifier
- virtual channel identifier
- payload type
- cell loss priority
- header error control

ATM Cells Format

- Pre-assigned values of the Cell Header at the physical layer

Cell type	Octet 1	Octet 2	Octet 3	Octet 4
IDLE cells	00000000	00000000	00000000	00000001
Physical Layer OAM	00000000	00000000	00000000	00001001
Reserved for use by Physical Layer	PPP0000	00000000	00000000	0000PPP1

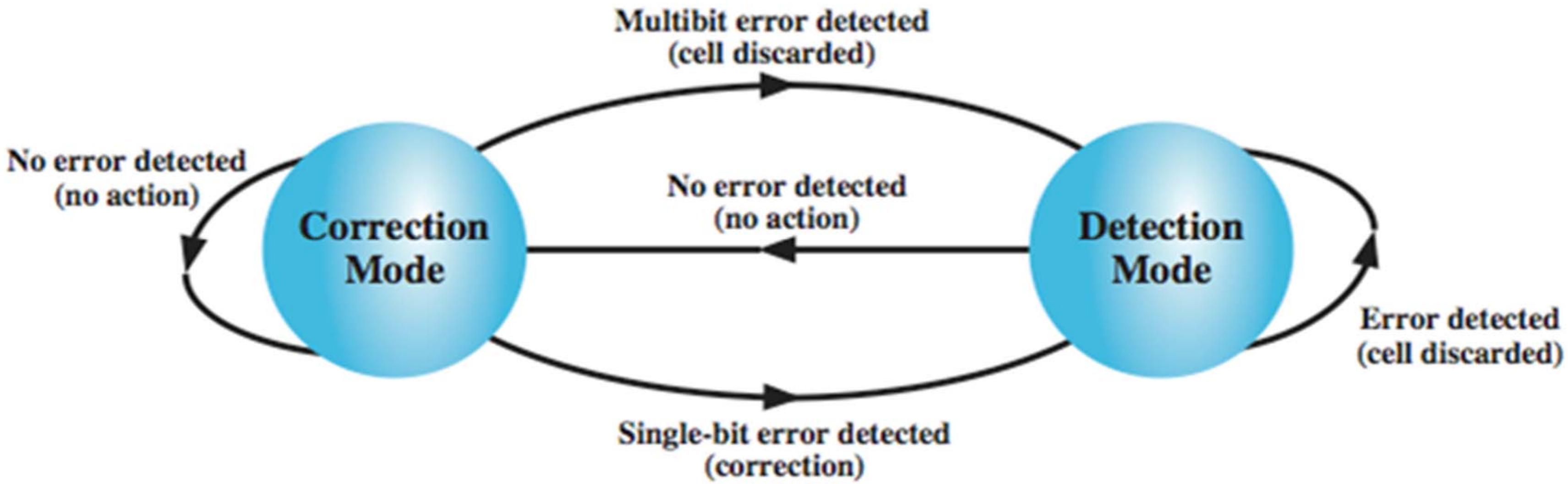
P : Bit is available for use by the PHY layer

ATM Cells Format

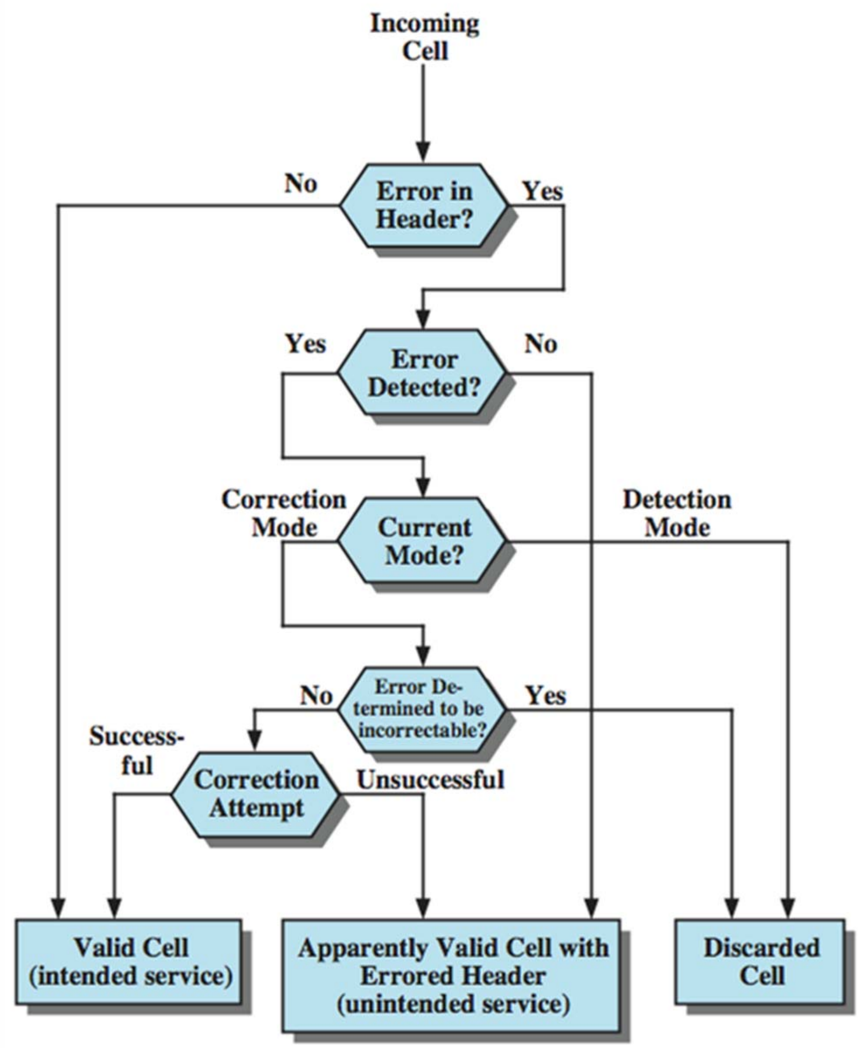
- Payload Type indicator

Payload type	Meaning
000	User data cell, no congestion, cell type 0
001	User data cell, no congestion, cell type 1
010	User data cell, congestion experienced, cell type 0
011	User data cell, congestion experienced, cell type 1
100	Maintenance information between adjacent switches
101	Maintenance information between source and destination switches
110	Resource Management cell (used for ABR congestion control)
111	Reserved for future function

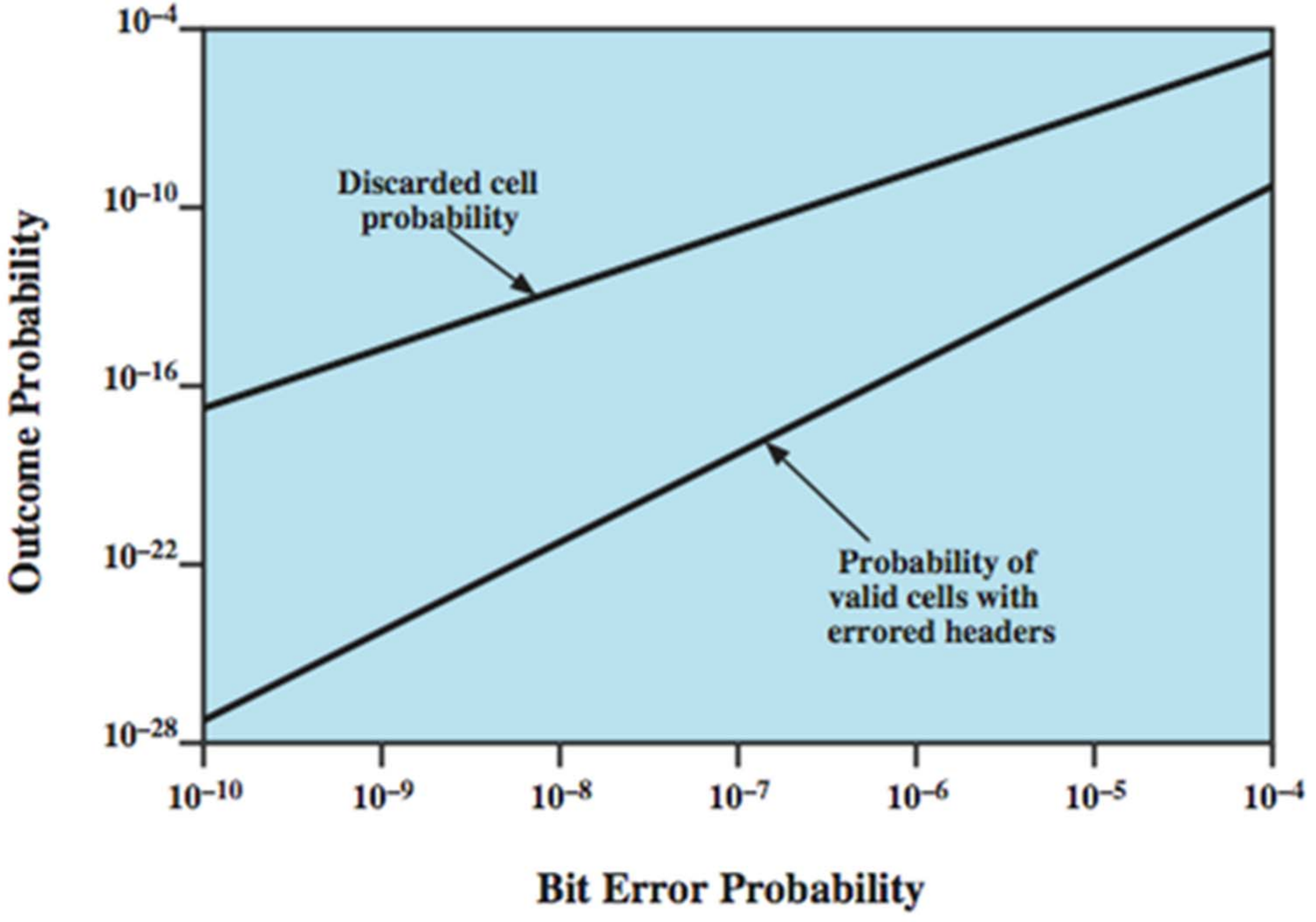
Header Error Control



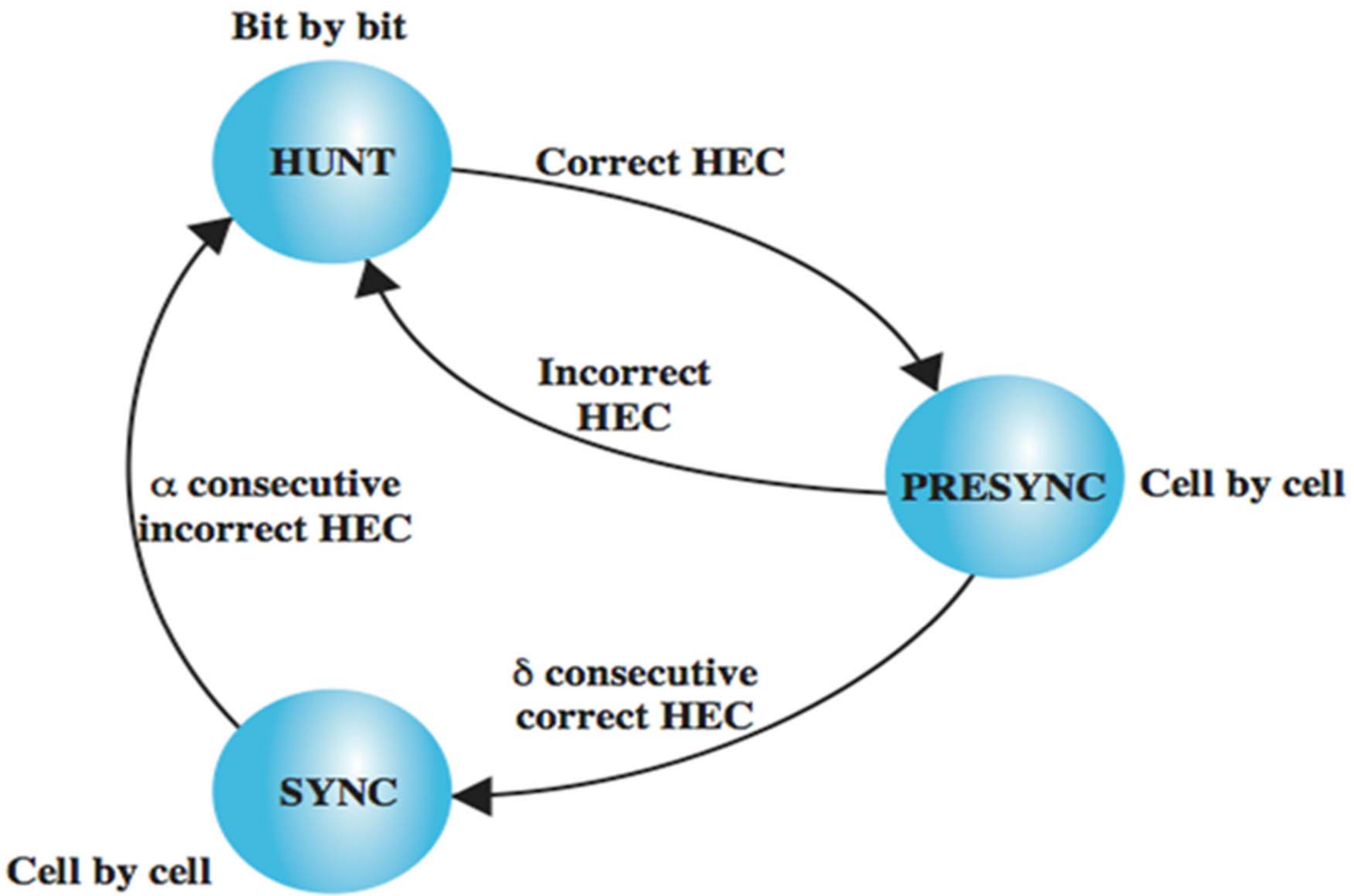
Effect of Error Cell Header



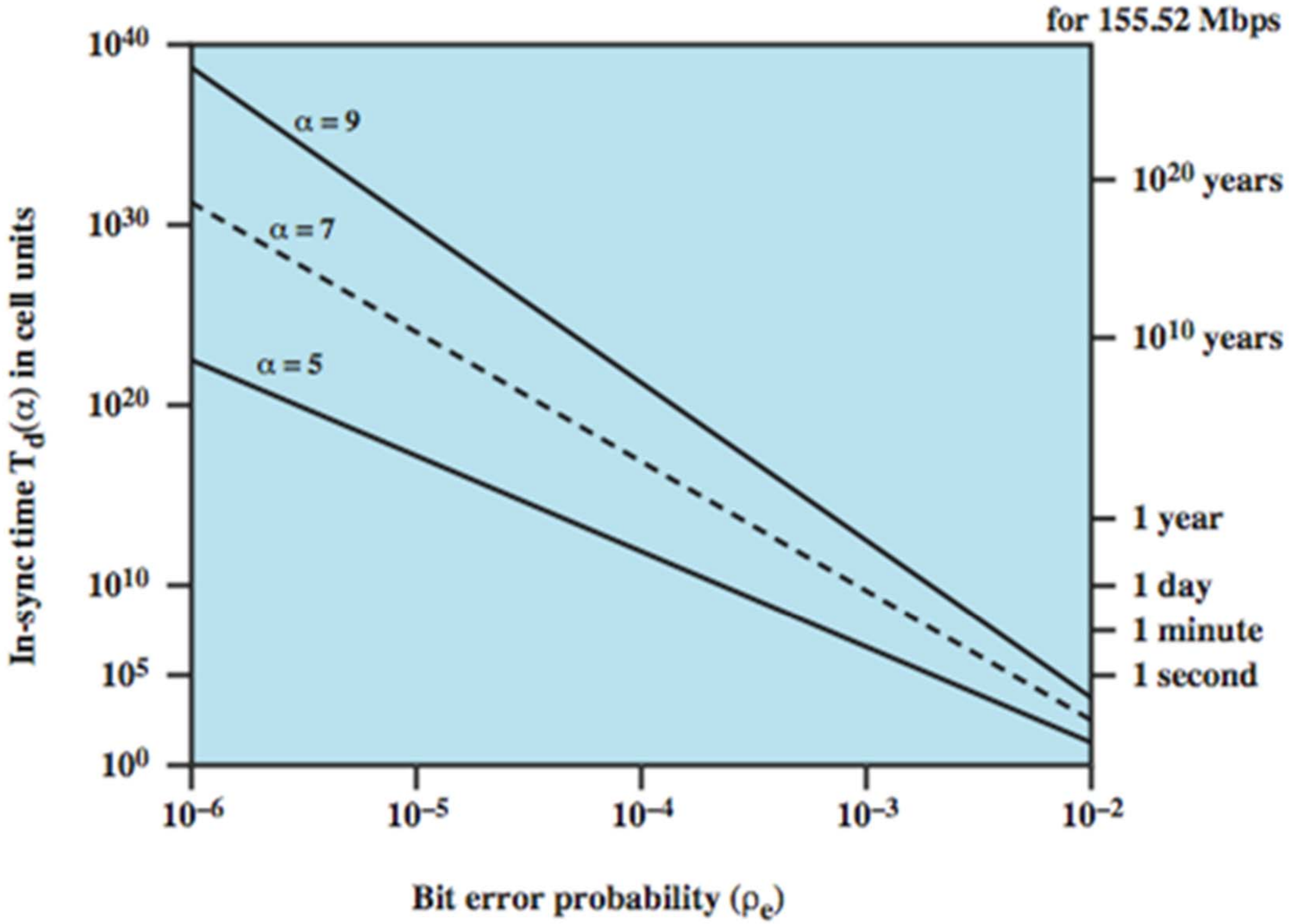
Impact of Random Bit Errors on HEC Performance



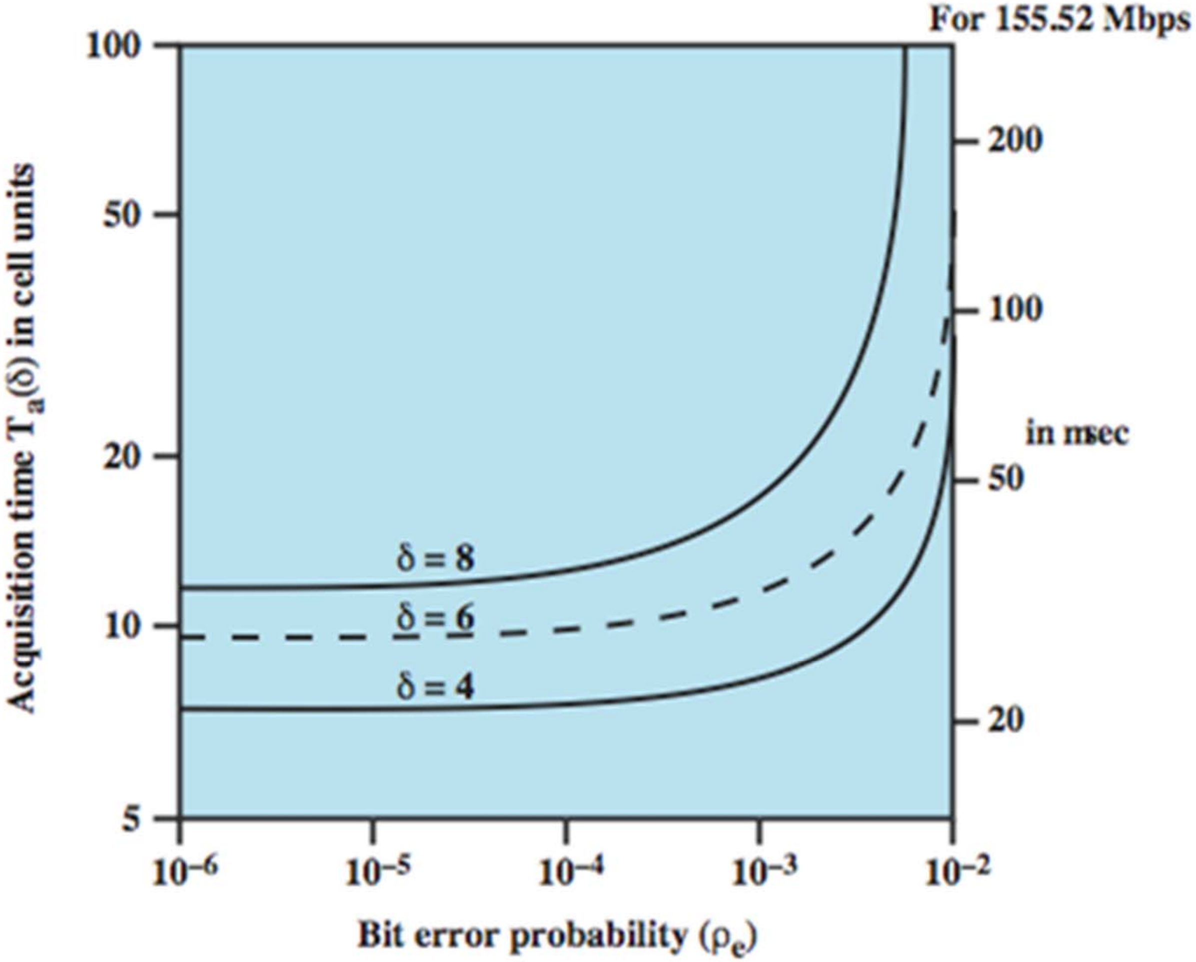
Cell Delineation State Diagram



Impact of Random Bit Errors on Cell Delineation Performance



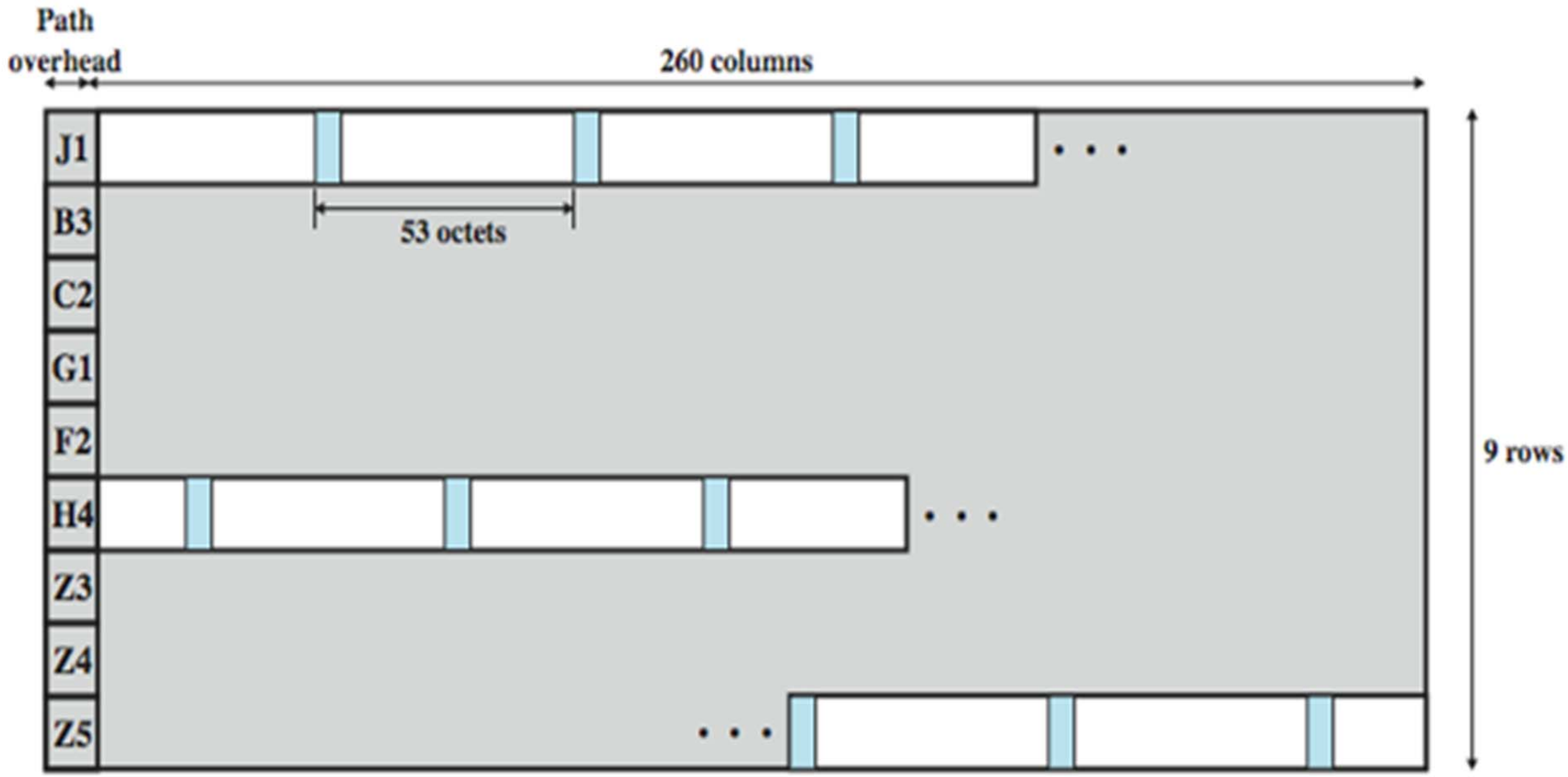
Acquisition Time vs. Bit Error Rate



SDH Based Physical Layer

- imposes structure on ATM stream
  - eg. for 155.52Mbps
  - use STM-1 (STS-3) frame
- can carry ATM and STM payloads
- specific connections can be circuit switched using SDH channel
- SDH multiplexing techniques can combine several ATM streams

STM-1 Payload for SDH-Based ATM Cell Transmission

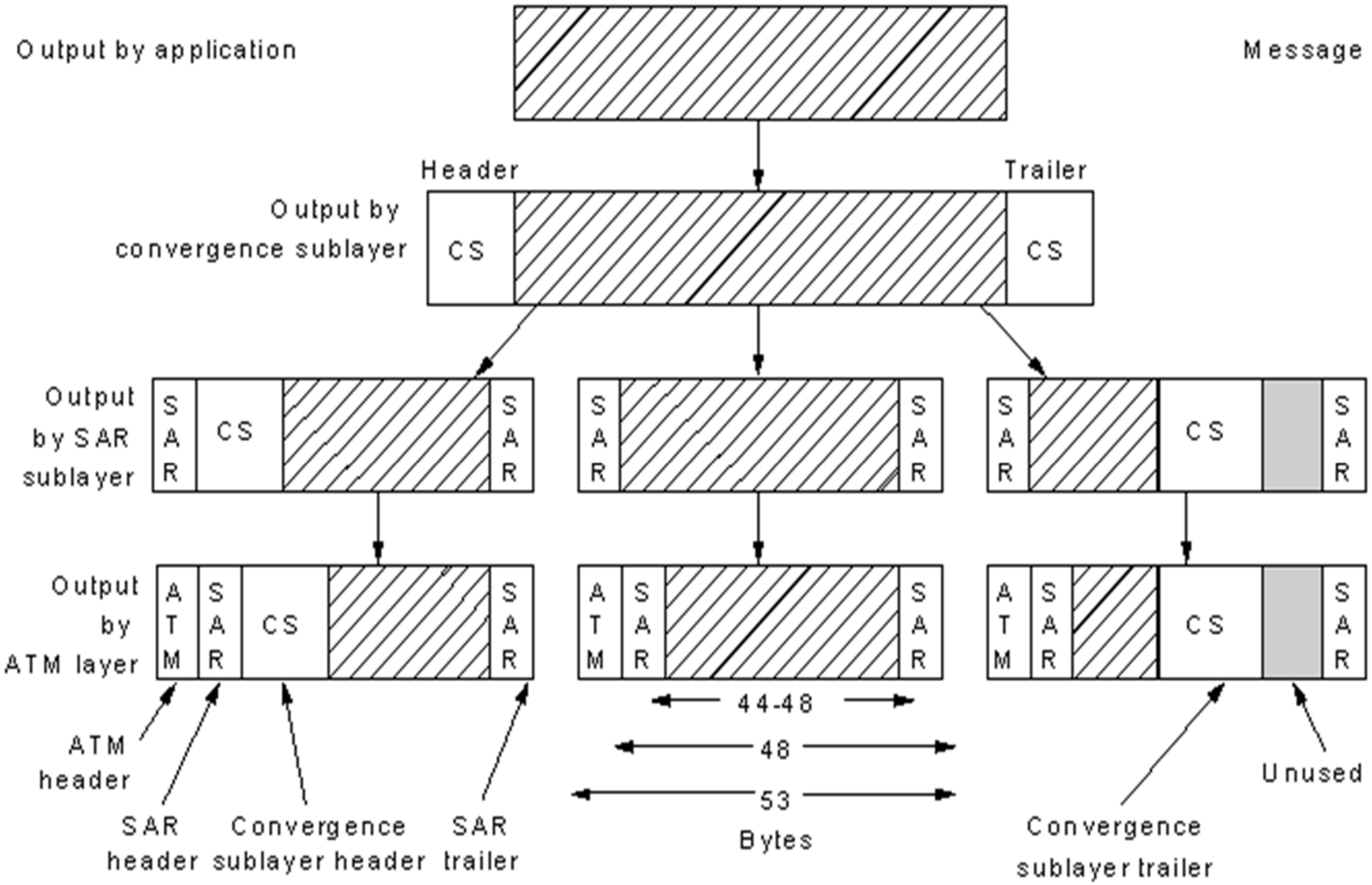


ATM Adaptation Layer

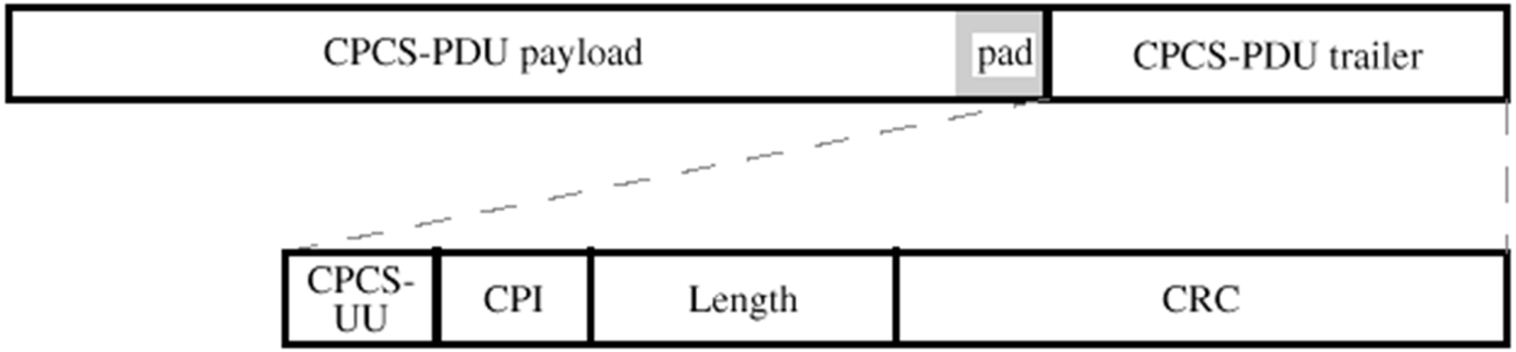
- Service Classification for AAL

	Class A	Class B	Class C	Class D
Timing relation between source and destination	Required		Not required	
Bit rate	Constant	Variable		
Connection mode	Connection-oriented			Connectionless
AAL Protocol	Type 1	Type 2	Type 3/4	Type 5

Segmentation and Reassembly

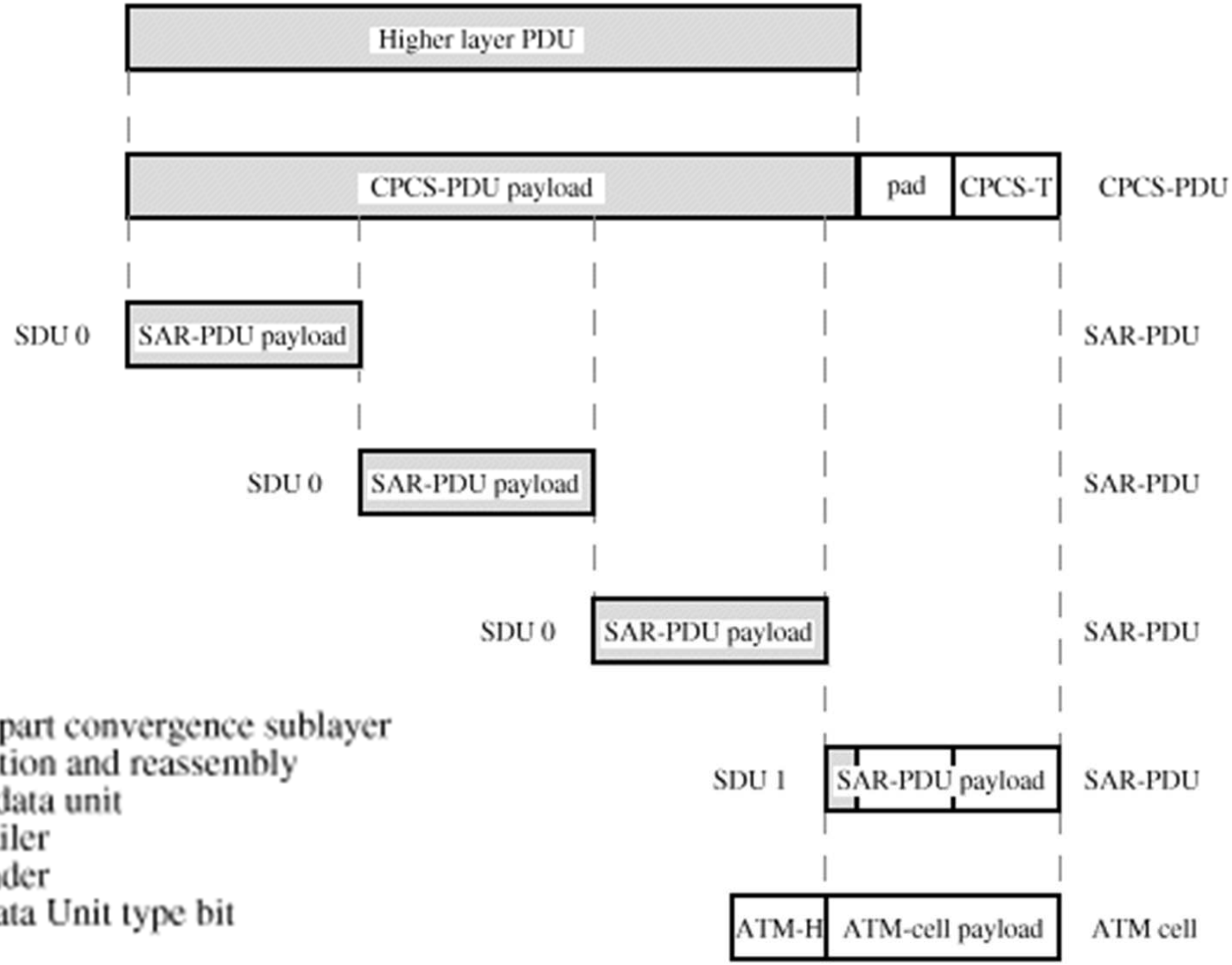


AAL5 CPCS-PDU



CPCS-UU = CPCS user-to-user indication (1 octet)
CPI = common part indicator (1 octet)
Length = length of CPCS-PDU payload (2 octets)
CRC = cyclic redundancy check (4 octets)

AAL5 transmission



CPCS = common part convergence sublayer
SAR = segmentation and reassembly
PDU = protocol data unit
CPCS-T = CPCS trailer
ATM-H = ATM header
SDU = Service Data Unit type bit

ATM Service Categories

Real time - limit amount/variation of delay

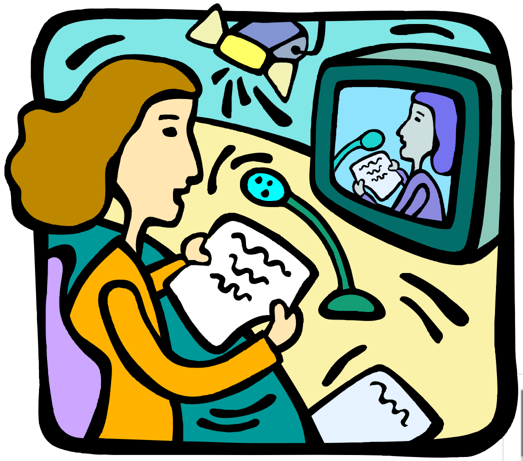
- Constant bit rate (CBR)
- Real time variable bit rate (rt-VBR)

Non-real time - for bursty traffic

- Non-real time variable bit rate (nrt-VBR)
- Available bit rate (ABR)
- Unspecified bit rate (UBR)
- Guaranteed frame rate (GFR)

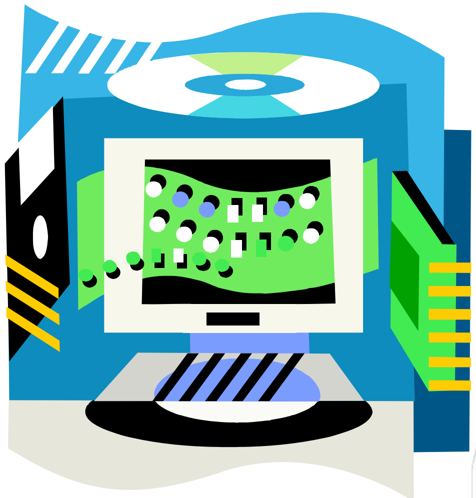
Constant Bit Rate (CBR)

- fixed data rate continuously available
- tight upper bound on delay
- uncompressed audio and video
  - video conferencing
  - interactive audio
  - A/V distribution and retrieval



Real-Time Variable Bit Rate (rt-VBR)

- for time sensitive applications
  - tightly constrained delay and delay variation
- rt-VBR applications transmit data at a rate that varies with time
- characterized as bursty
- allow more flexibility than CBR



Non-Real-Time Variable Bit Rate (nrt-VBR)

- used for data transfers with critical response time
  - airline reservations, banking transactions
- end system specifies:
  - a peak cell rate
  - a sustainable or average cell rate
  - measure of how bursty or clumped cells can be

Unspecified Bit Rate (UBR)

- may be additional capacity over and above that used by CBR and VBR traffic
  - not all resources dedicated to CBR/VBR traffic
  - unused cells due to bursty nature of VBR
- for application that can tolerate some cell loss or variable delays
  - eg. TCP based traffic
- cells forwarded on FIFO basis
- best effort service
- examples:
  - text/data/image transfer
  - telecommuting

3.3 Commutació d’etiquetes

Multiprotocol Label Switching (MPLS)

- MPLS is a set of IETF specifications for including routing and traffic engineering information in packets
- comprises a number of interrelated protocols - - MPLS protocol suite
  - is used to ensure that all packets in a particular flow take the same route over a backbone
  - deployed by many telecommunication companies and service providers
  - delivers QoS required to support real-time voice and video and SLAs that guarantee bandwidth

Role of MPLS

- efficient technique for forwarding and routing packets
- designed with IP networks in mind
  - can be used with any link-level protocol
- fixed-length label encapsulates an IP packet or a data link frame
- MPLS label contains all information needed to perform routing, delivery, Qos, and traffic management functions
- is connection oriented

Traffic Engineering

- ability to define routes dynamically, plan resource commitments on the basis of known demand, and optimize network utilization
- effective use can substantially increase usable network capacity
- ATM provided strong traffic engineering capabilities prior to MPLS
- with basic IP there is a primitive form

MPLS:

- **is aware of flows with QoS requirements**
- **possible to set up routes on the basis of flows**
- **paths can be rerouted intelligently**

MPLS Operation

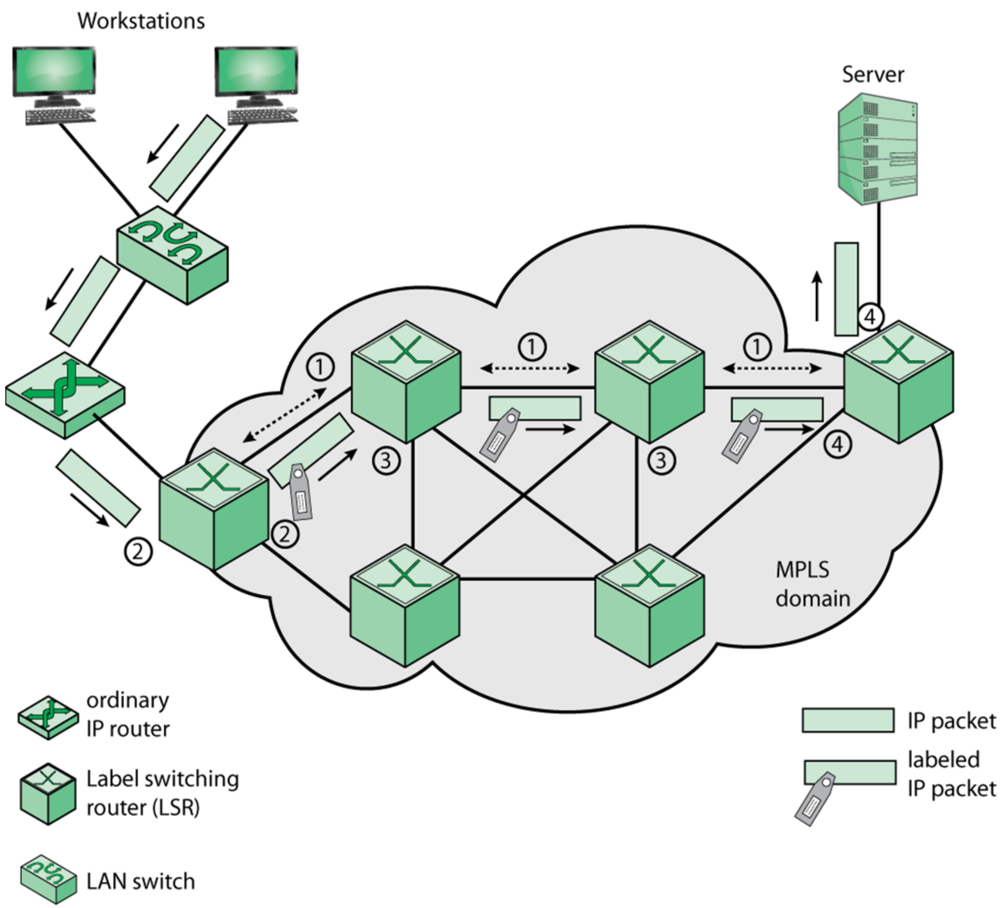


Figure 21.1 MPLS Operation

MPLS Packet Forwarding

FEC	out i'face	out label
a	1	19
b	1	20
...	...	...

in label	in i'face	out i'face	out label
19	2	0	24
13	3	0	25
20	2	1	22

in label	in i'face	out i'face	out label
24	1	0	—
25	1	0	—
...	...	...	...

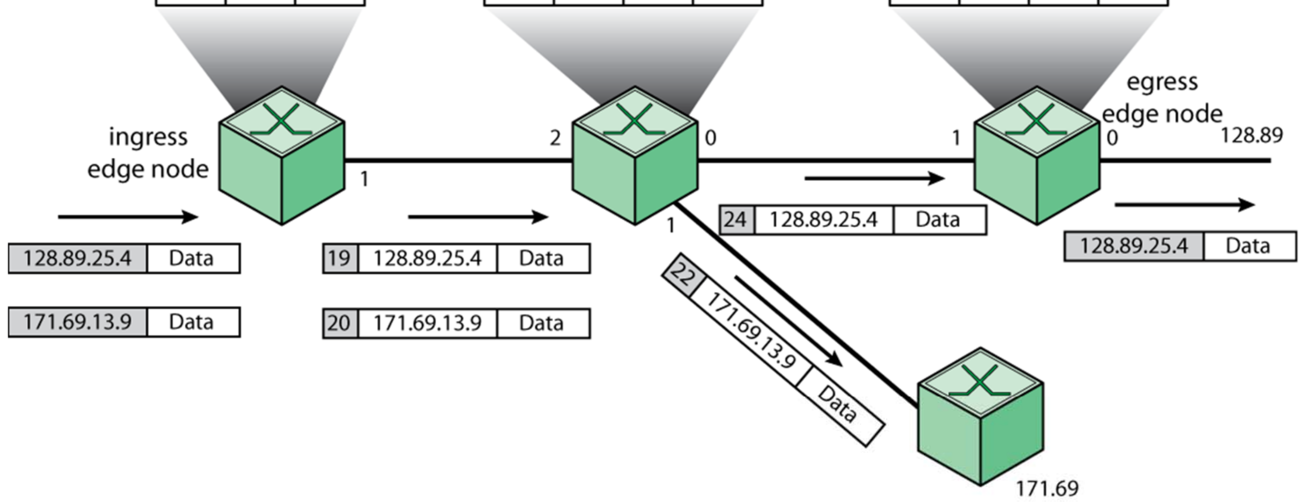


Figure 21.2 MPLS Packet Forwarding

LSP Creation and Packet Forwarding

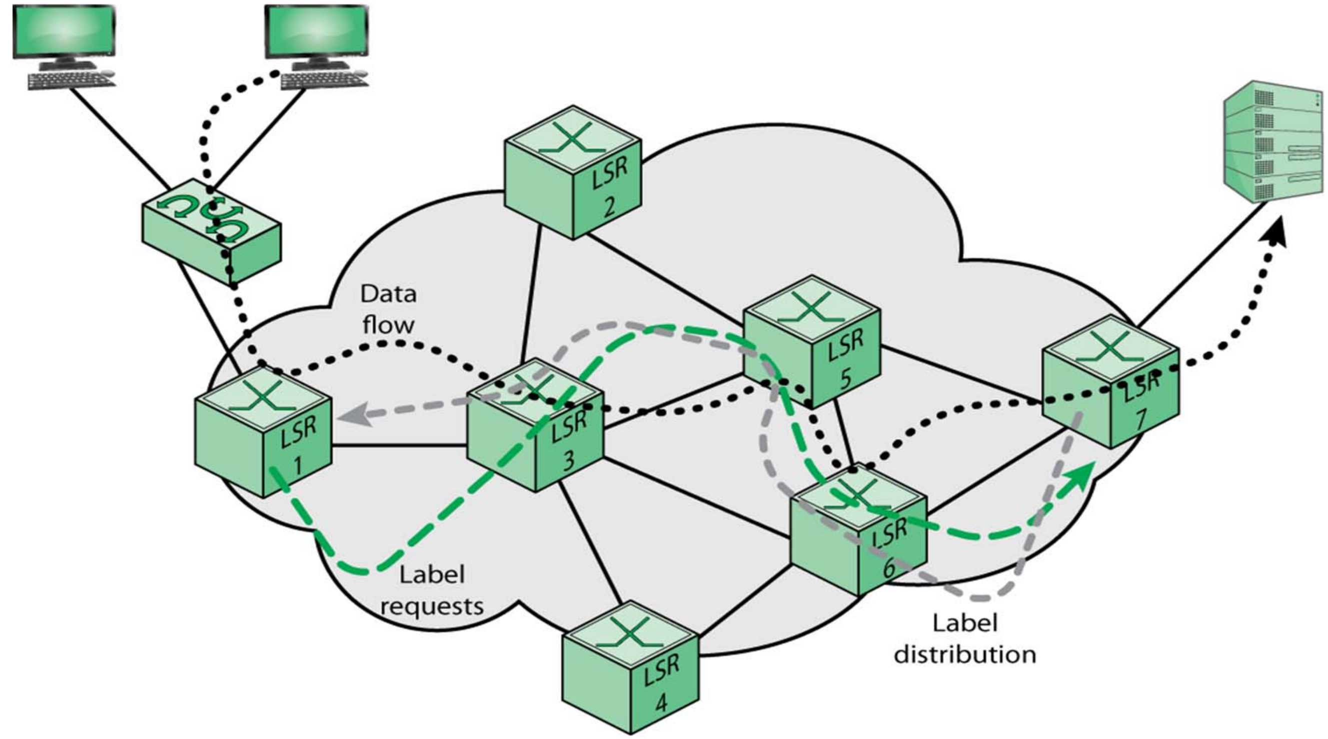


Figure 21.3 LSP Creation and Packet Forwarding through an MPLS Domain

Label Stacking

- one of the most powerful features of MPLS
  - processing is always based on the top label
  - at any LSR a label may be removed or added
- allows creation of tunnels
  - tunnel refers to traffic routing being determined by labels
- provides considerable flexibility
- unlimited stacking

UNLIMITED

STACKING

Label Format

- defined in RFC 3032
- 32-bit field consisting of:
  - Label value
  - Traffic class (TC)
  - S
  - Time to live (TTL)

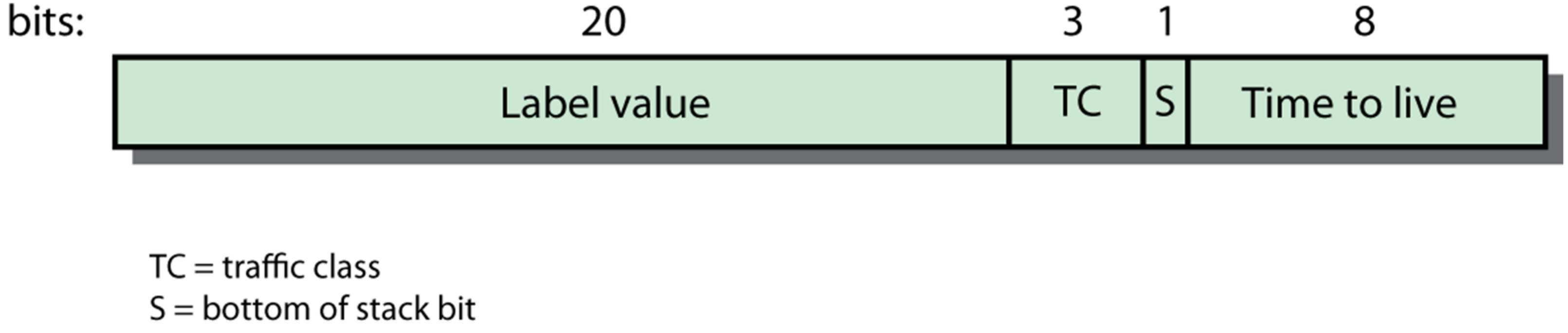


Figure 21.4 MPLS Label Format

Time to Live Field (TTL)

- key field in the IP packet header
- decremented at each router and packet is dropped if the count falls to zero
  - *done to avoid looping*
  - *having the packet remain too long in the Internet due to faulty routing*
- included in the label so that the TTL function is still supported

Label Placement

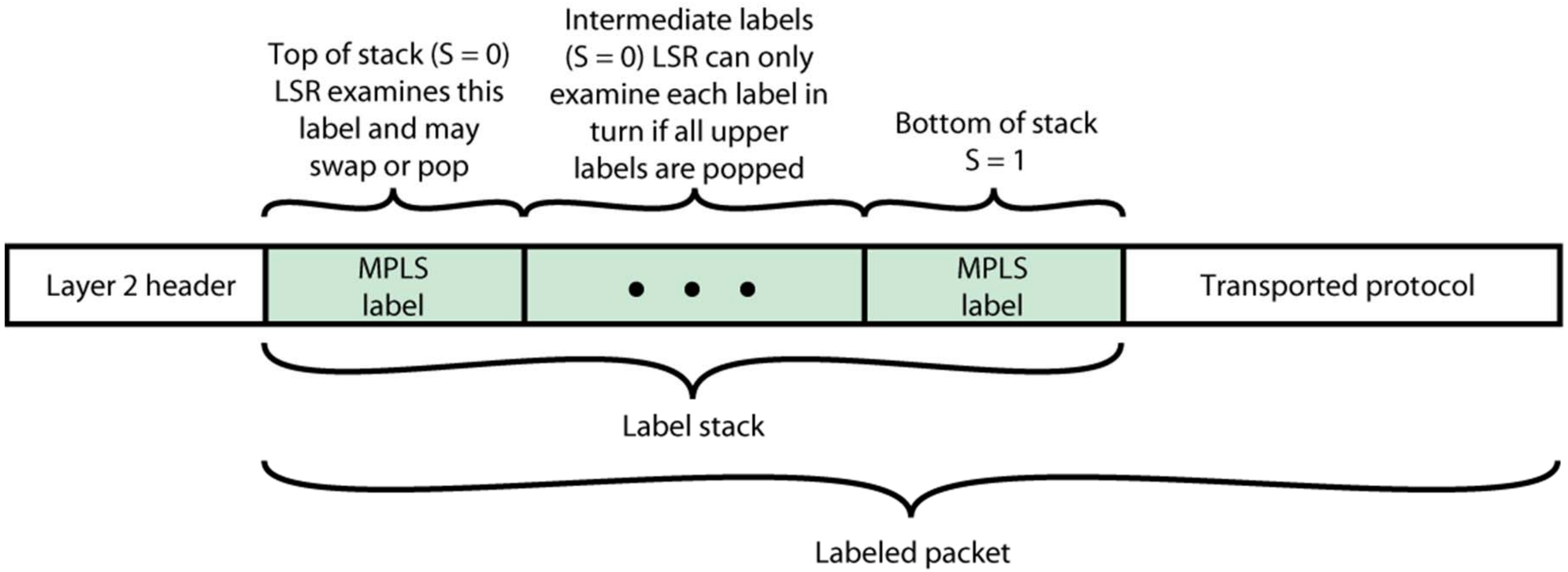


Figure 21.5 Encapsulation for Labeled Packet

Label Stack

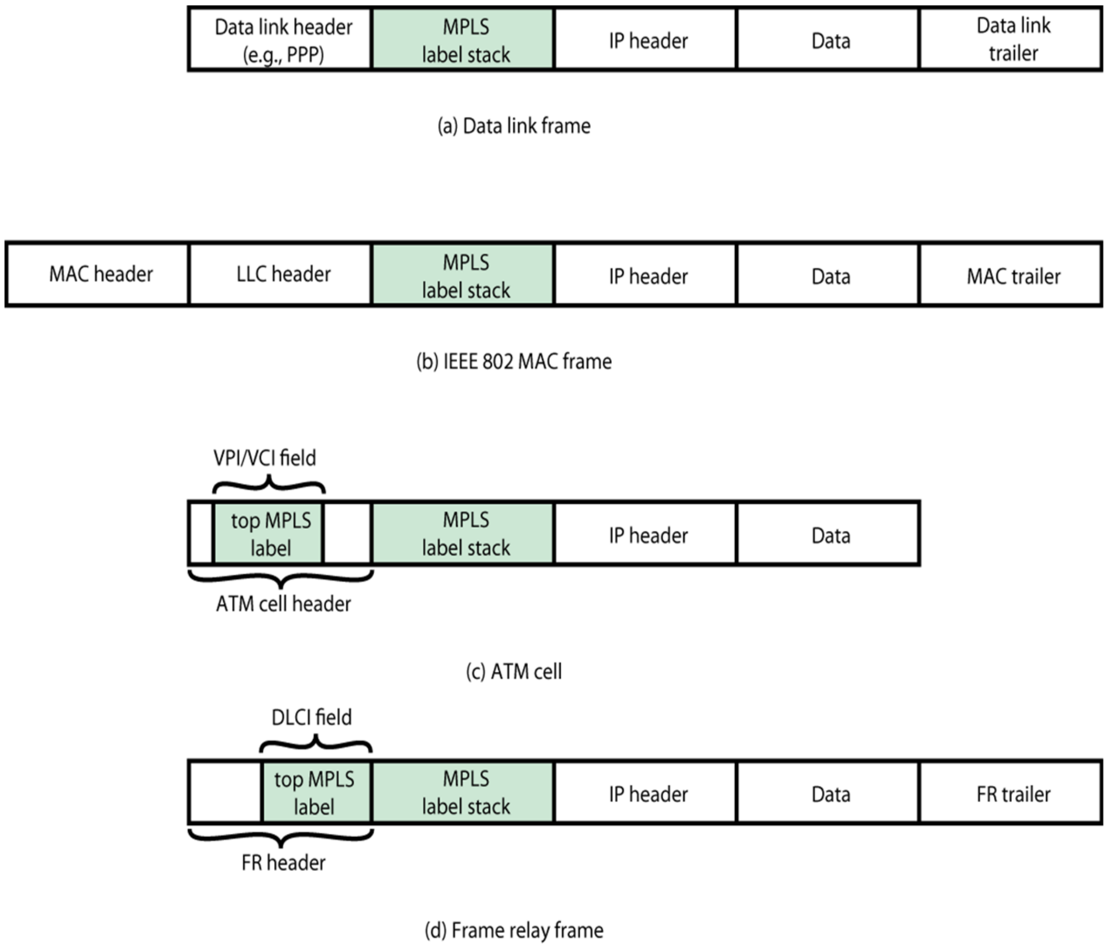
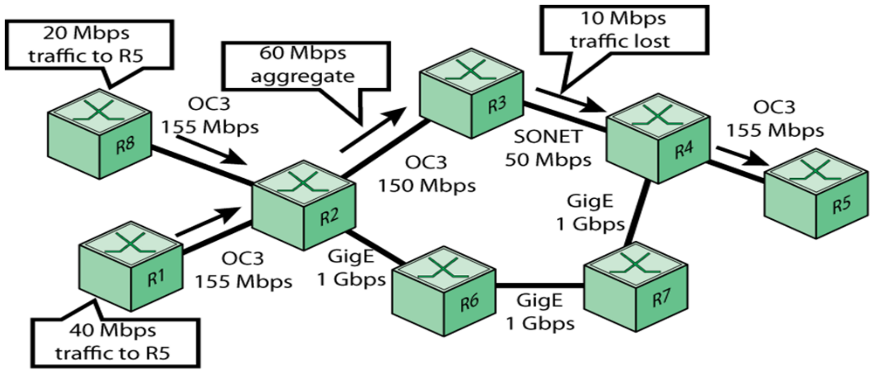
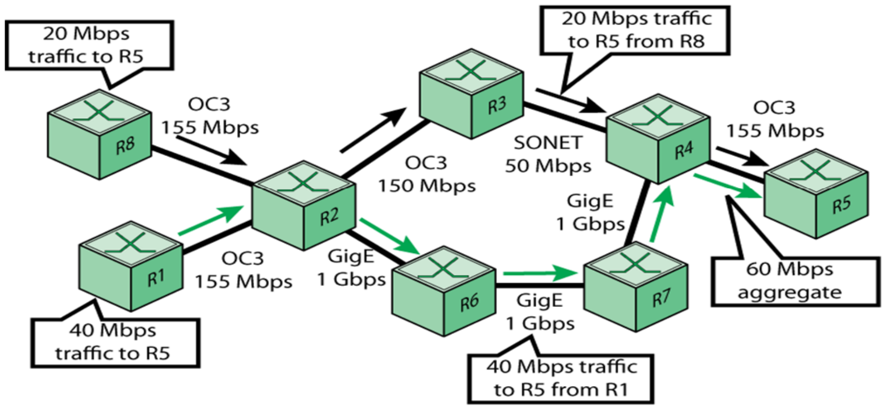


Figure 21.6 Position of MPLS Label Stack

Example of Traffic Engineering



(a) A shortest-path solution



(b) A traffic-engineered solution

Figure 21.9 Traffic Engineering Example

RSVP – TE Operation

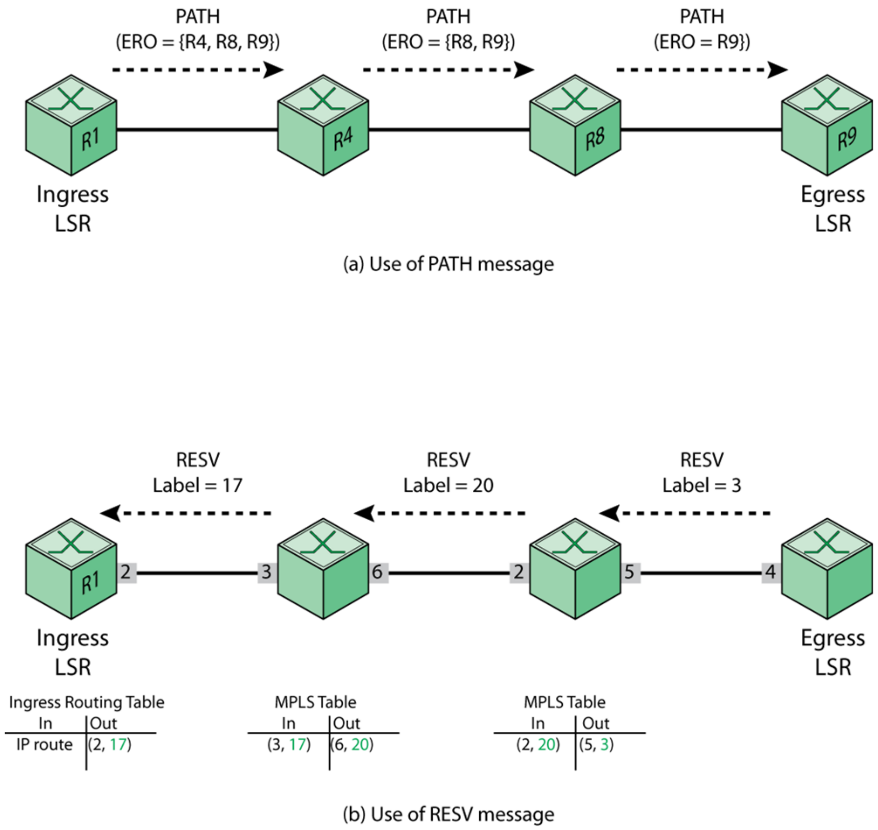


Figure 21.11 RSVP-TE Operation

Layer 2 VPN Concepts

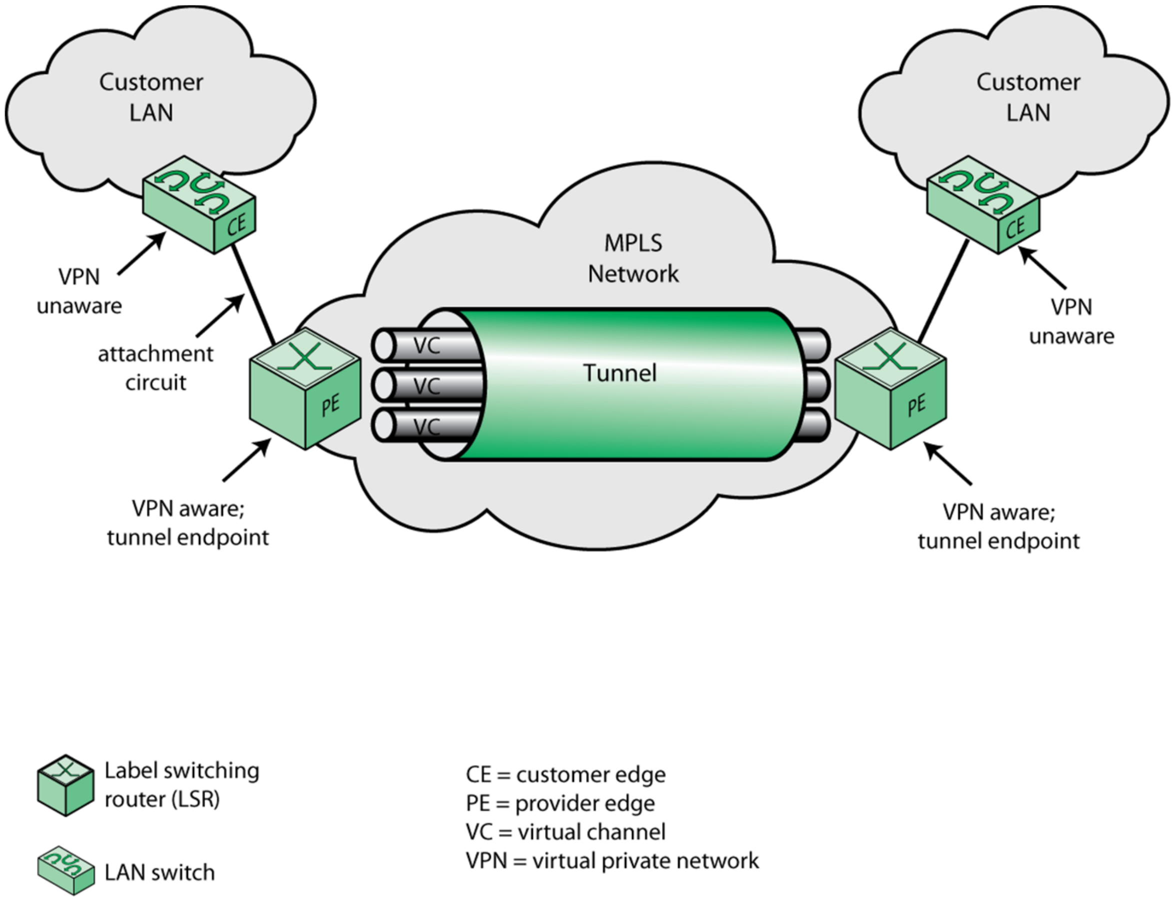


Figure 21.12 Layer 2 VPN Concepts

C16 p4

2.4 Carrier Ethernet

Entenem per carrier ethernet, la capacitat per transortar trames ethernet.

A ivell 1 puc fer servir PDH, SDH o ethernet
A nivell físic, un sistema ethernet transmet trames ethernet.
els altres 2 (PDH i SDH) com tenen canals, per sobre d'aquests canals poden portar trames ethernet
si la linia física és ethernet només podm transemtre ethernet
si és una atra podem transportar ATM frame-rlay pdh sdh ... (que poden fer de carrier ethernet)

Cada vegada s'utilitzen cables de major capacitat

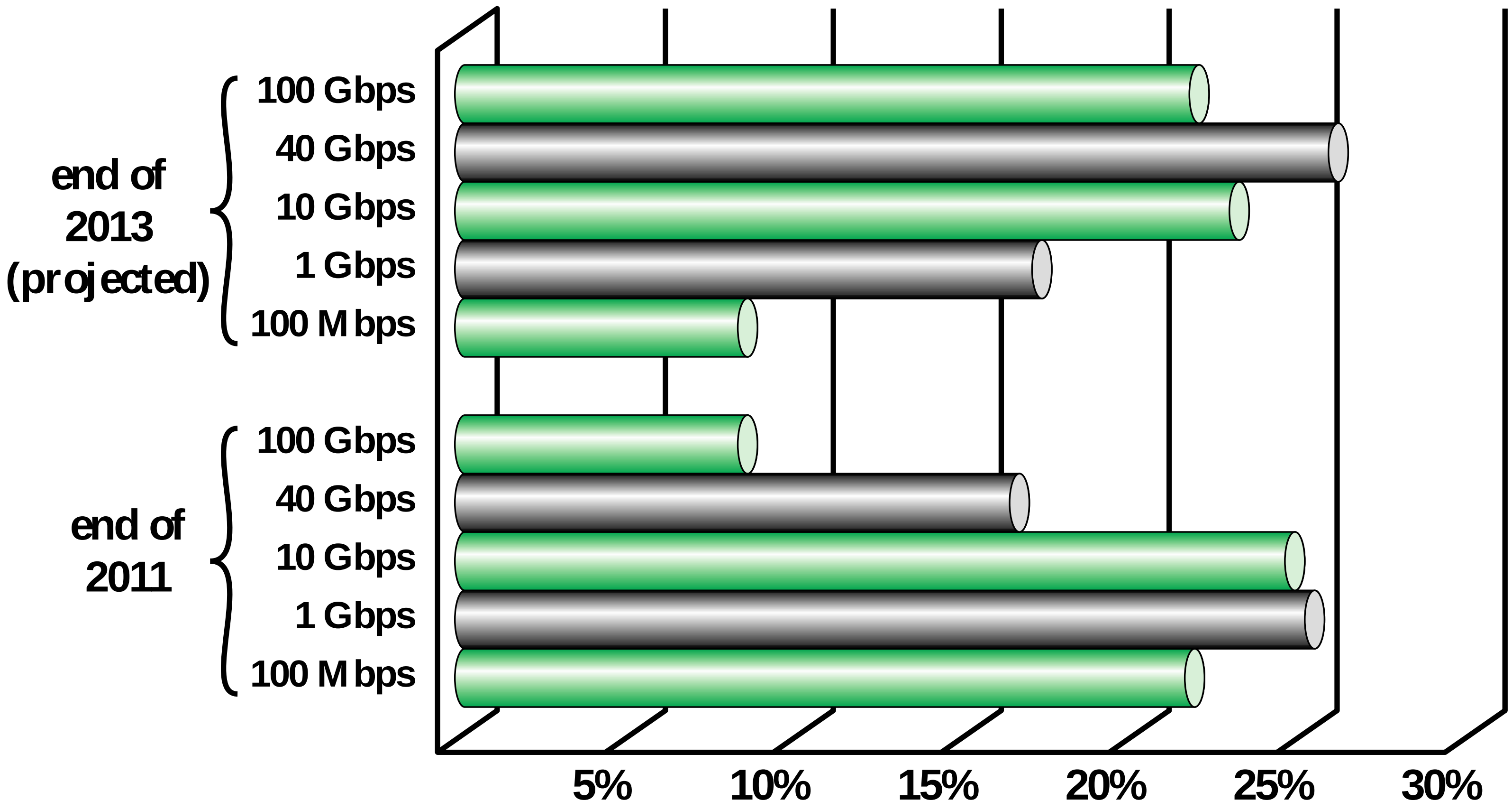


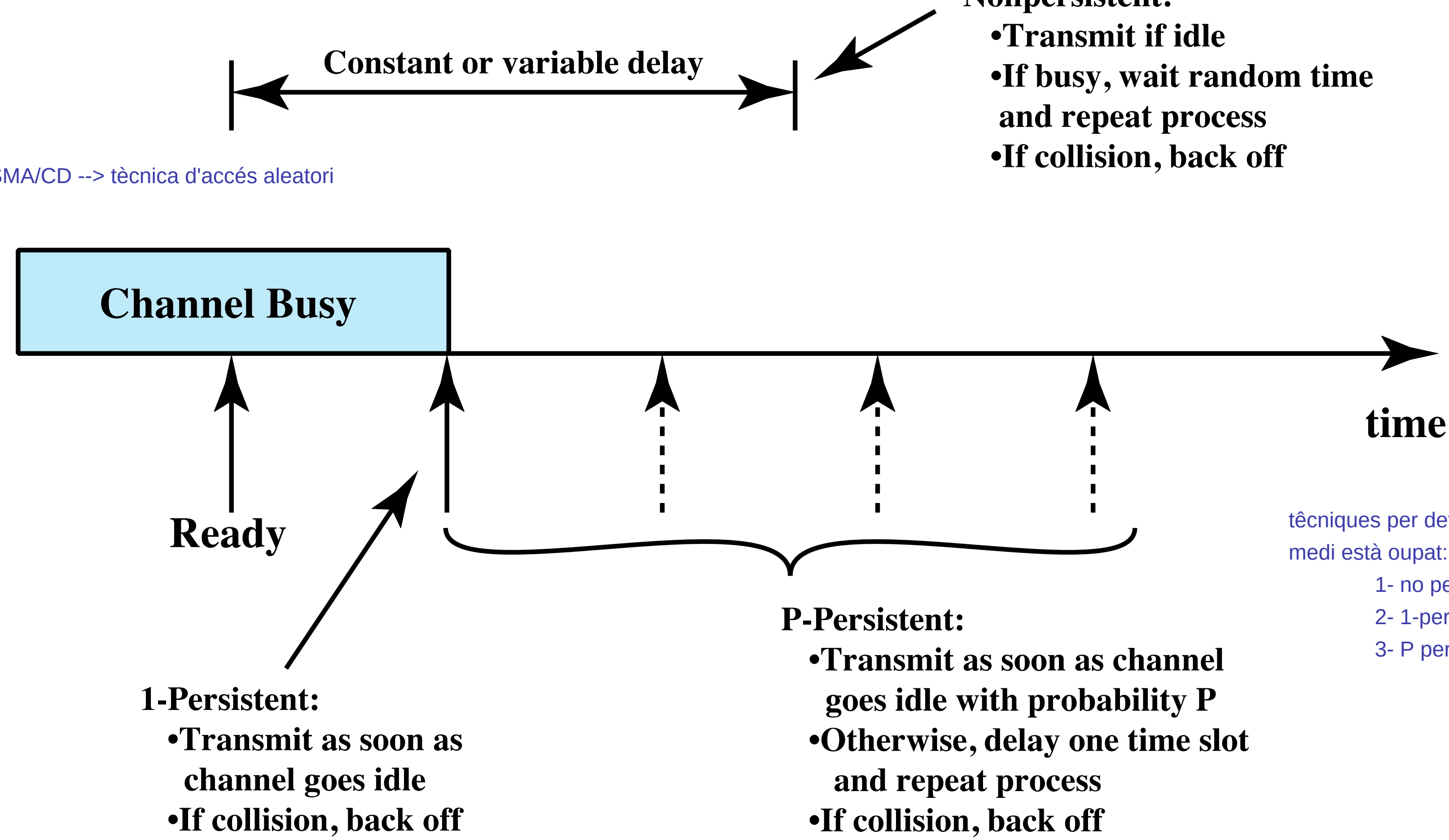
Figure 12.1 Data Center Study—Percentage of Ethernet Links by Speed

Ethernet té 2 entors de funcionament (el hub i el switch)

-hub: tots els temianals de la xarxa comparteixen el medi de transmissió

C16 - p4-

CSMA/CD --> tècnica d'accés aleatori



têcniques per determinar què fa una estació si el medi està oupat:

1- no persistent

2- 1-persistent

3- P persistent

CSMA/CD --> el terminal quan vol transemtre, escolta la línia, i si veu que ningu està transmetent, transmet ell.

pot passar que 2 vegin el mitja lliure i transmetin alhora.

el que es fa és esperar a que arribi l'ACK del que s'ha transmes, si no arriba, s'assumeix colisió i es torna a transmetre.

Figure 12.2 CSMA Persistence and Backoff

P-Persistent CSMA

- A compromise to try and reduce collisions and idle time

"1 persistent voldria dir que quan no detectes col·lisió, transmetries sempre, aixo no és bona idea"

- P-persistent CSMA rules:
  1. *If medium is idle, transmit with probability p, and delay one time unit with probability (1–p)*

     per time unit es sol agafar el temps de propagació de punta a punta de la xarxa
  2. *If medium is busy, listen until idle and repeat step 1*
  3. *If transmission is delayed one time unit, repeat step 1*
- Issue of choosing effective value of *p* to avoid instability under heavy load

tamany minim da la trama > 2*tau

tau = temps de propagacio de la xarxa

C16 - p10

per treballar amb ethernet tindrà que ser mab distàncies petites.
per treballar amb distàncies grans, podem treure el protocol, o fer comuncacions punt-a-punt (excepte si hi ha commutadors ethernet, perquè no van amb CSMA/CD)

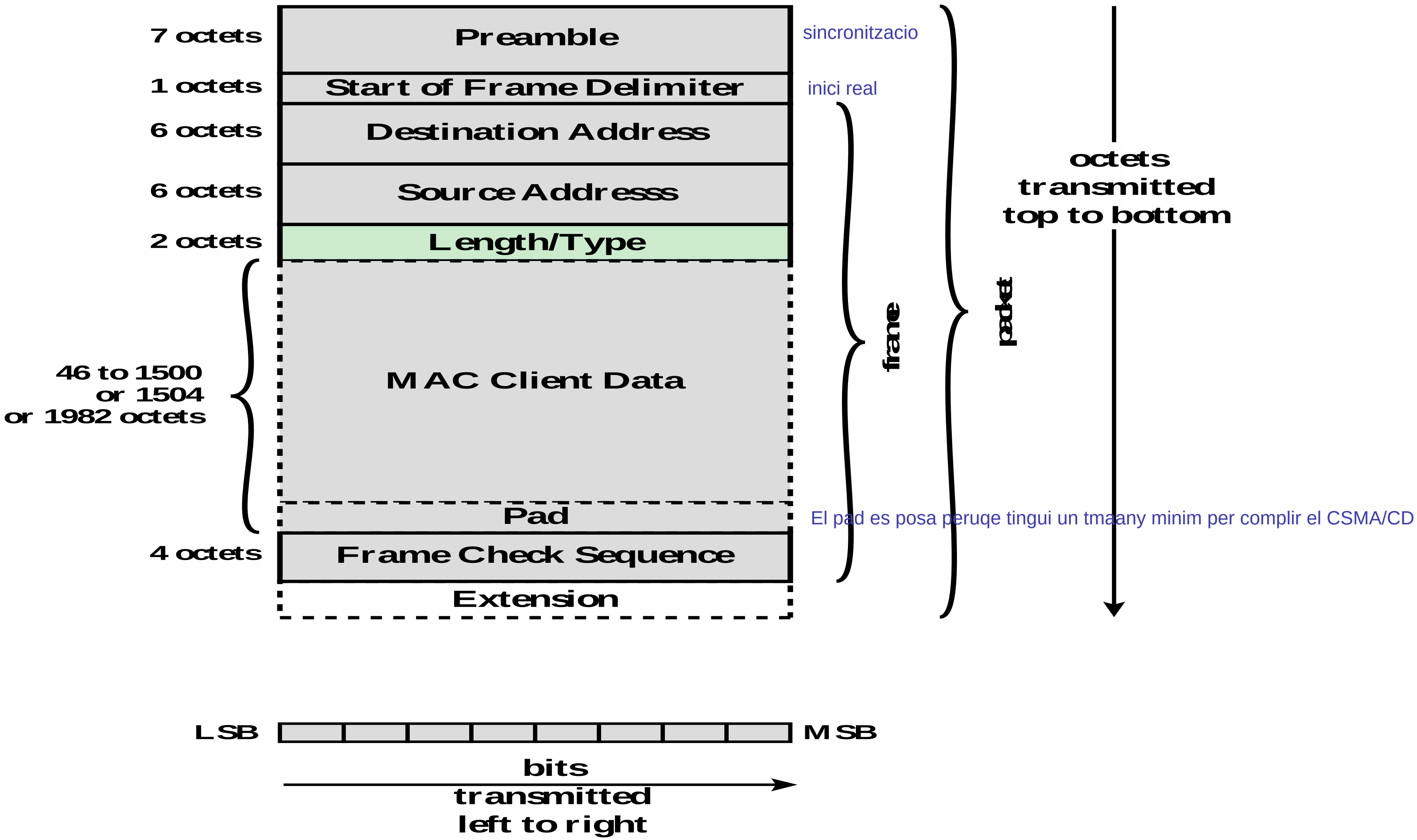


Figure 12.4 IEEE 802.3 MAC Frame Format

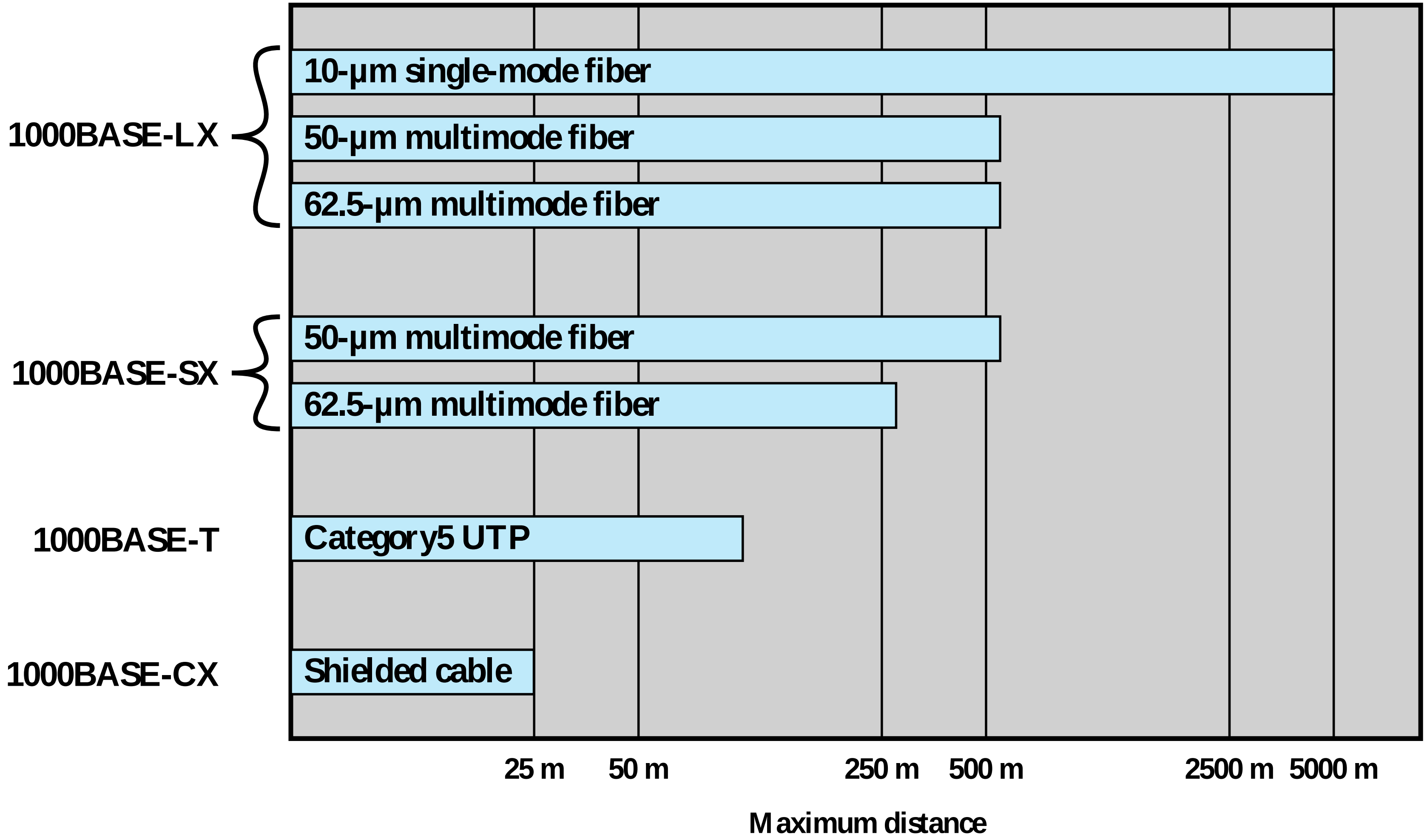


Figure 12.5 Gigabit Ethernet Medium Options (log scale)

configuració de ethernet amb switches ethernet (switck --> n hi ha CSMA/CD)

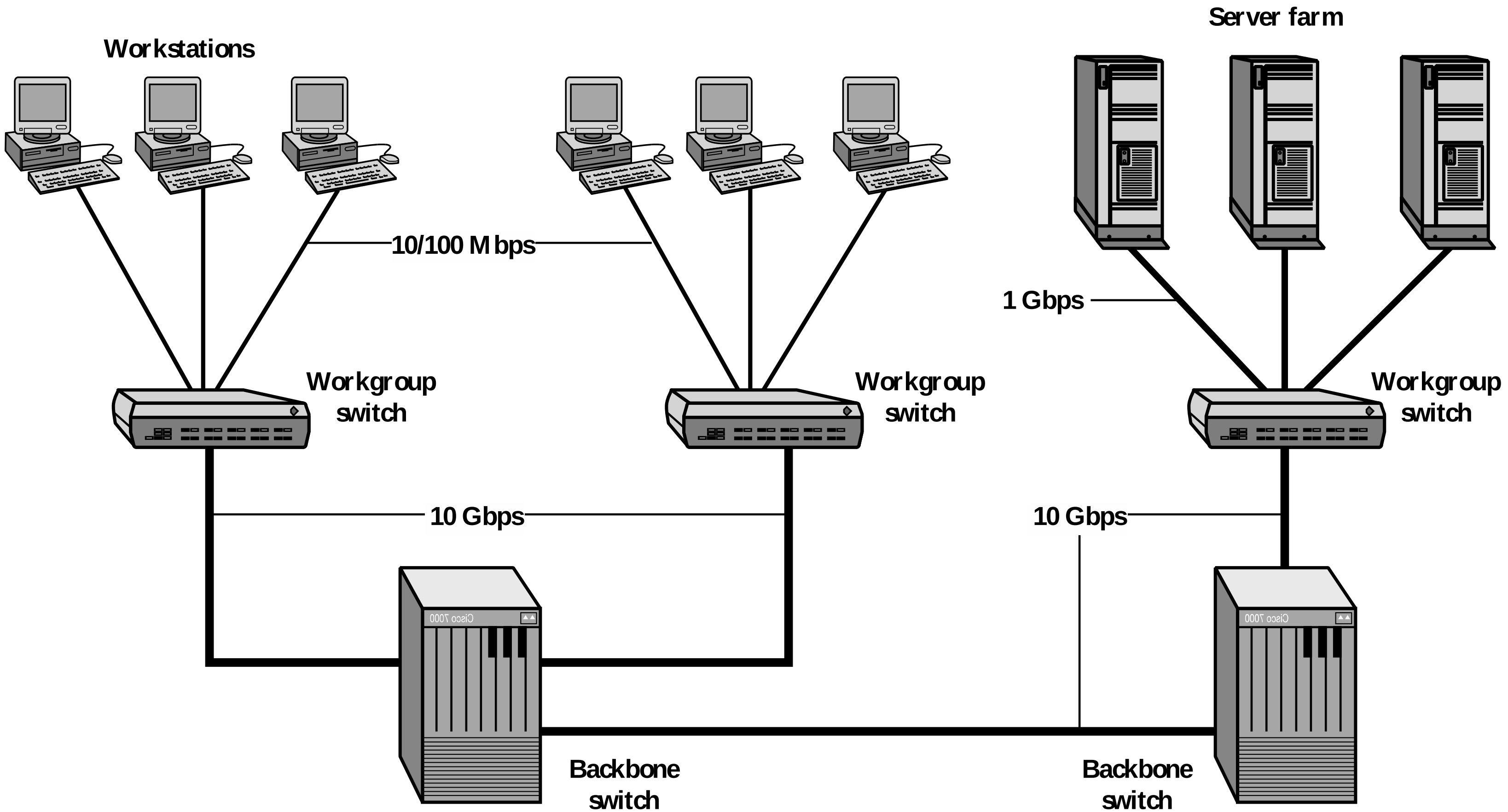


Figure 12.6 Example 10 Gigabit Ethernet Configuration

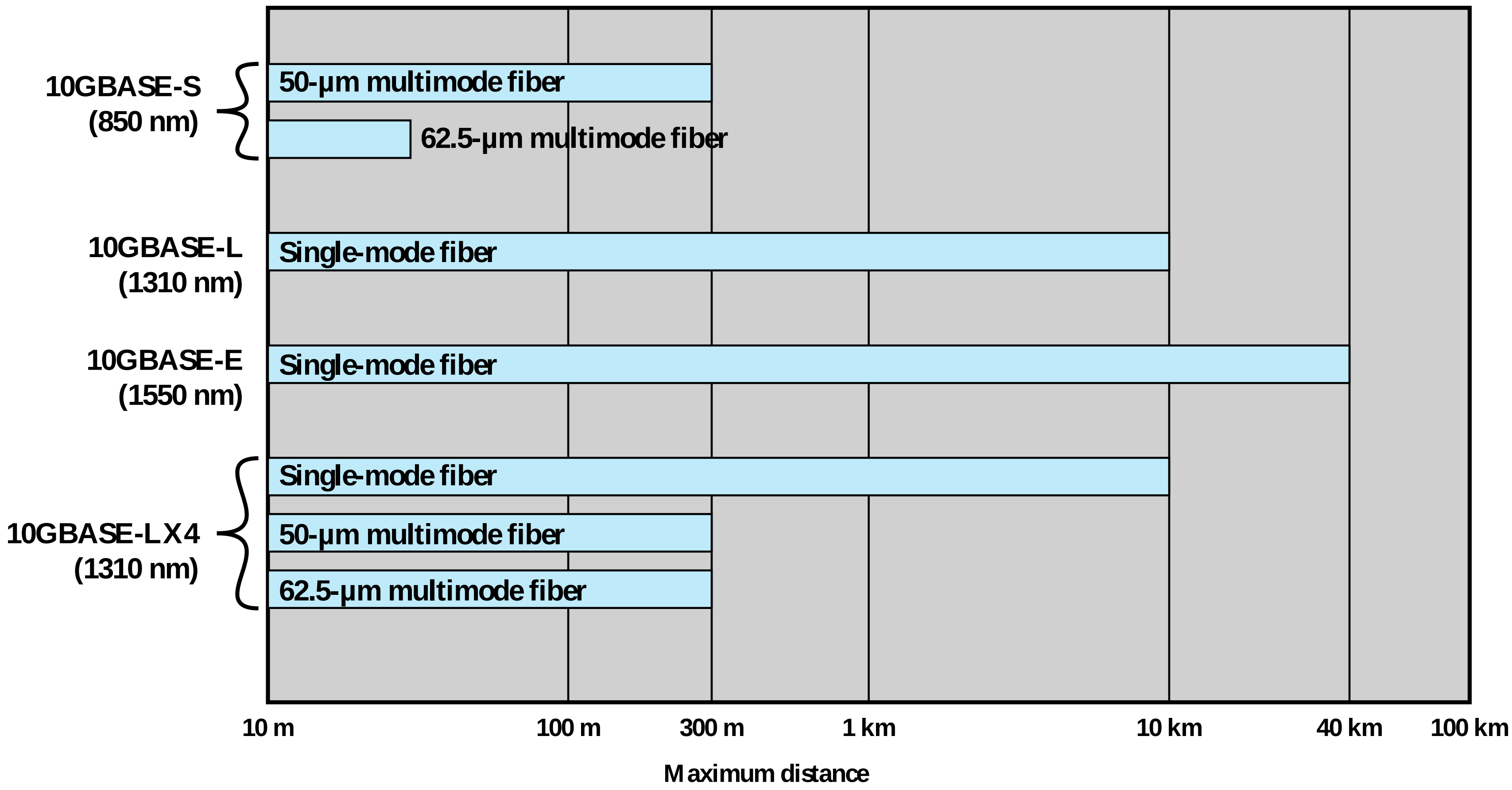


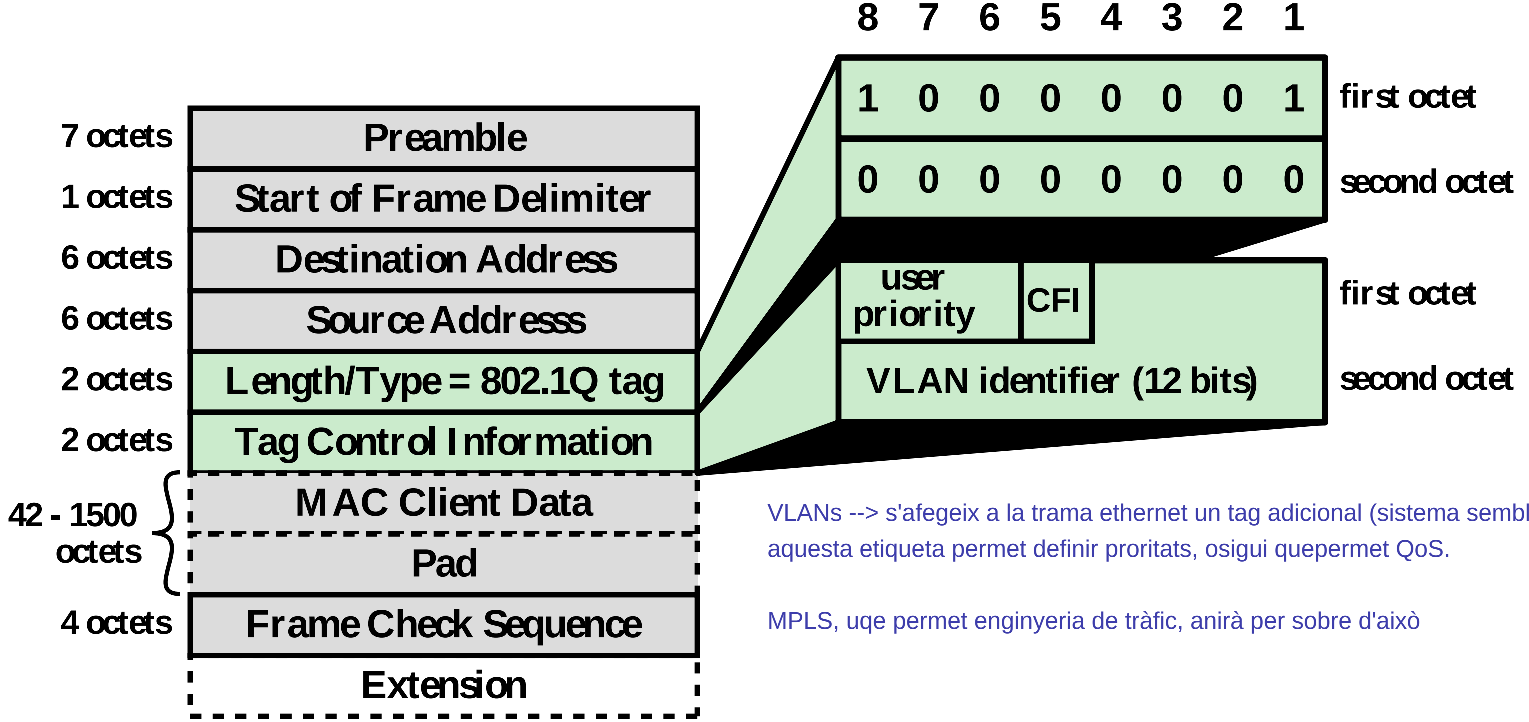
Figure 12.7 10-Gbps Ethernet Distance Options (log scale)

Media Options for 40-Gbps and 100-Gbps Ethernet

	40 Gbps	100 Gbps
1m backplane	40GBASE-KR4	
10 m copper	40GBASE-CR4	1000GBASE-CR10
100 m multimode fiber	40GBASE-SR4	1000GBASE-SR10
10 km single mode fiber	40GBASE-LR4	1000GBASE-LR4
40 km single mode fiber		1000GBASE-ER4

Naming nomenclature:

Copper: K = backplane; C = cable assembly
Optical: S = short reach (100m); L - long reach (10 km); E = extended long reach (40 km)
Coding scheme: R = 64B/66B block coding
Final number: number of lanes (copper wires or fiber wavelengths)



VLANs --> s'afegeix a la trama ethernet un tag adicional (sistema semblant a MPLS) aquesta etiqueta permet definir proritats, osigui quepermet QoS.

MPLS, uqe permet enginyeria de tràfic, anirà per sobre d'això

CFI = Canonical Format Indicator
VLAN = virtual local area network

Figure 12.10 Tagged IEEE 802.3 MAC Frame Format

Xarxa basada completament en ethernet

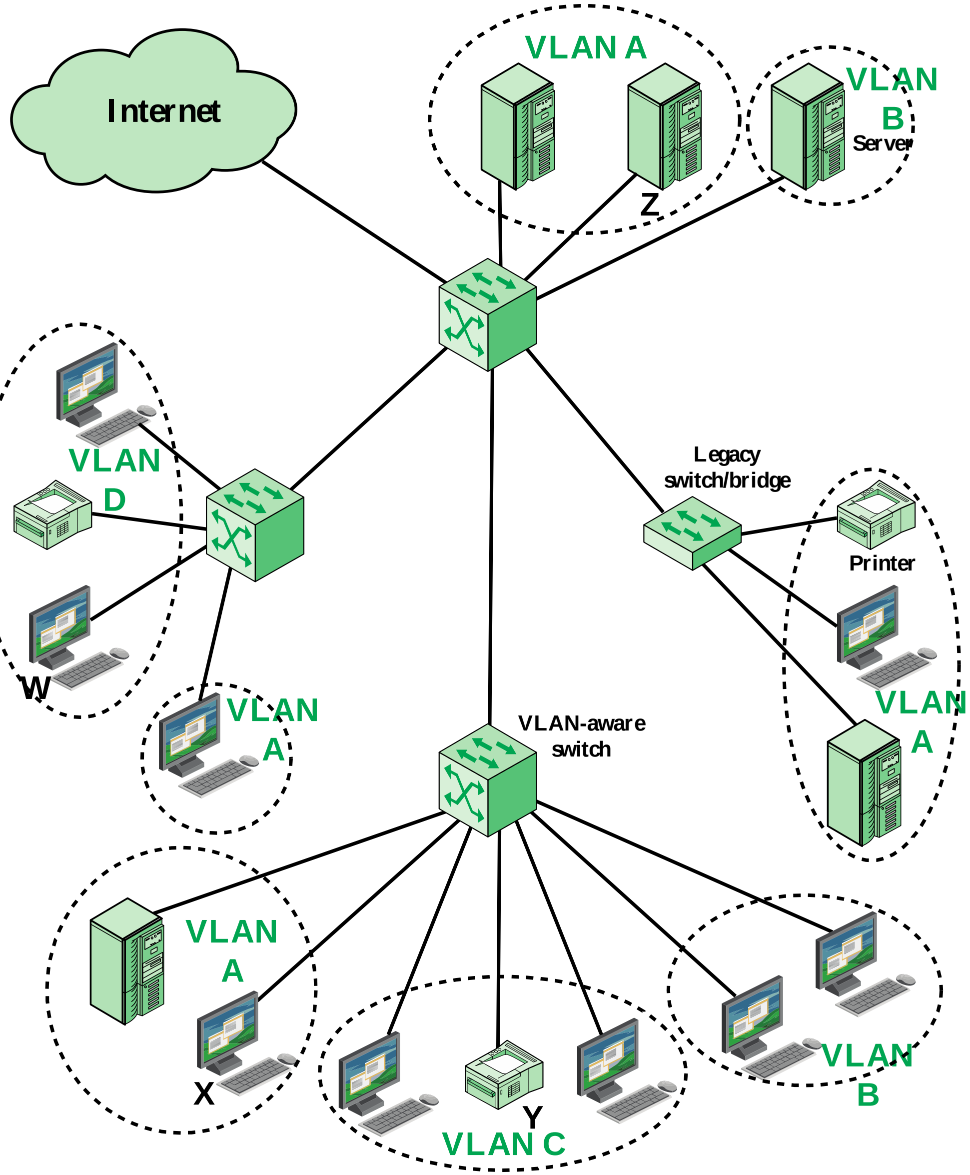


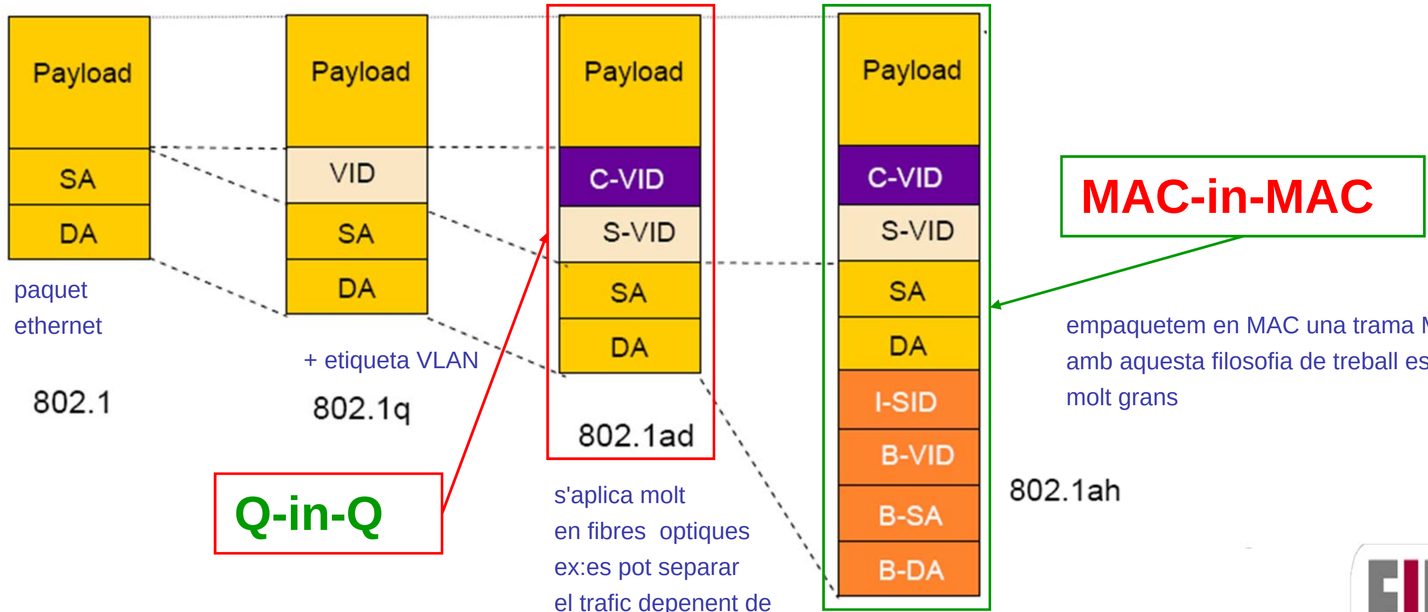
Figure 12.11 A VLAN Configuration

Stacking de etiquetes de VLANs (similar al de MPLS)

The way to PBB-TE/PBT

- **IEEE has developed a number of standards providing enhancements to the original Ethernet standards**

  - **802.1Q**: Virtual LAN
  - **802.1ad**: Provider Bridging
  - **802.1ah**: Provider Backbone Bridging
  - **802.1ag**: Connectivity Fault Management (OAM)



paquet ethernet

+ etiqueta VLAN

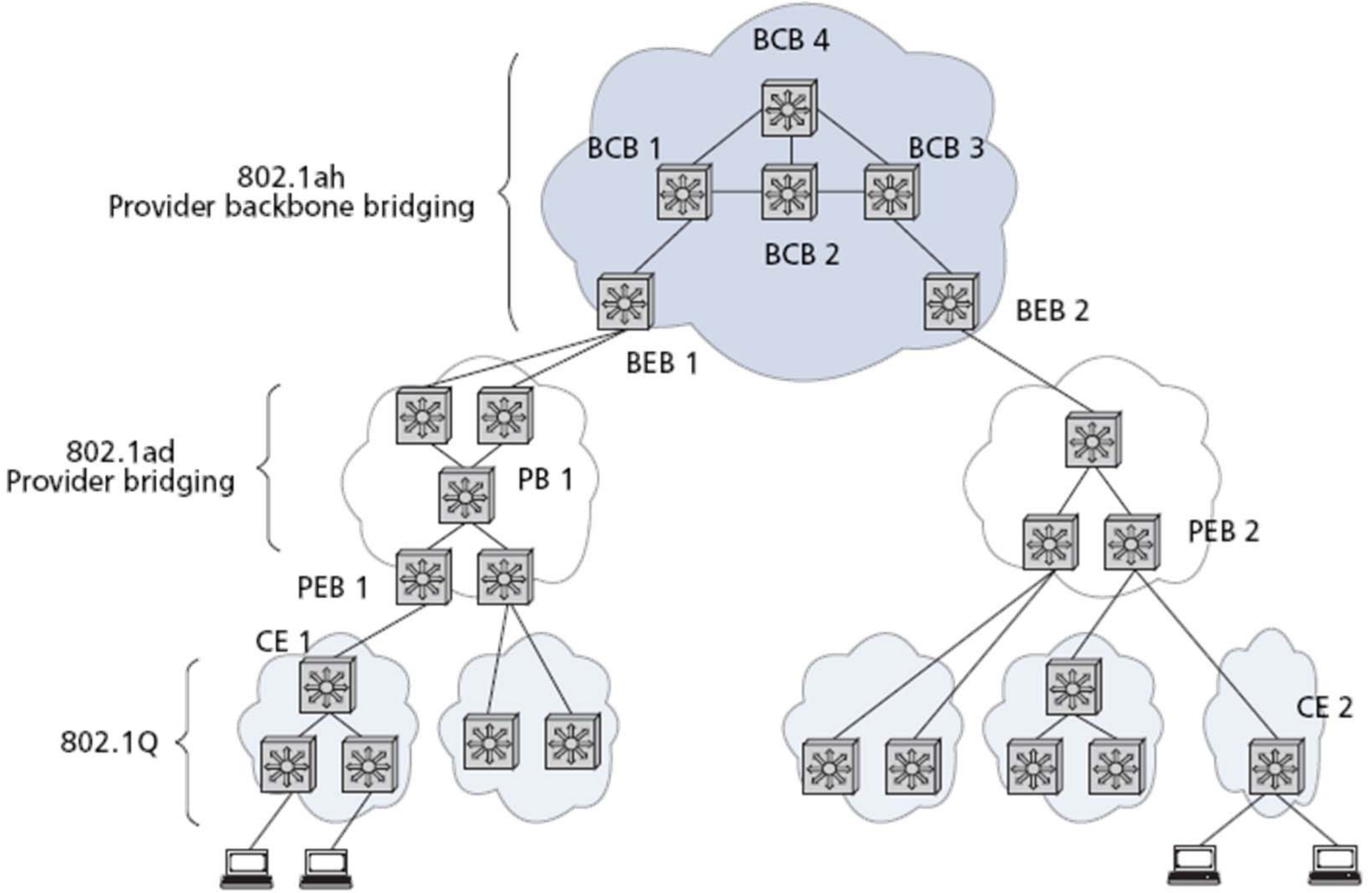
s'aplica molt en fibres optiques ex:es pot separar el trafic depenent de l'operador

empaquetem en MAC una trama MAC amb aquesta filosofia de treball es poden cobrir xarxes molt grans

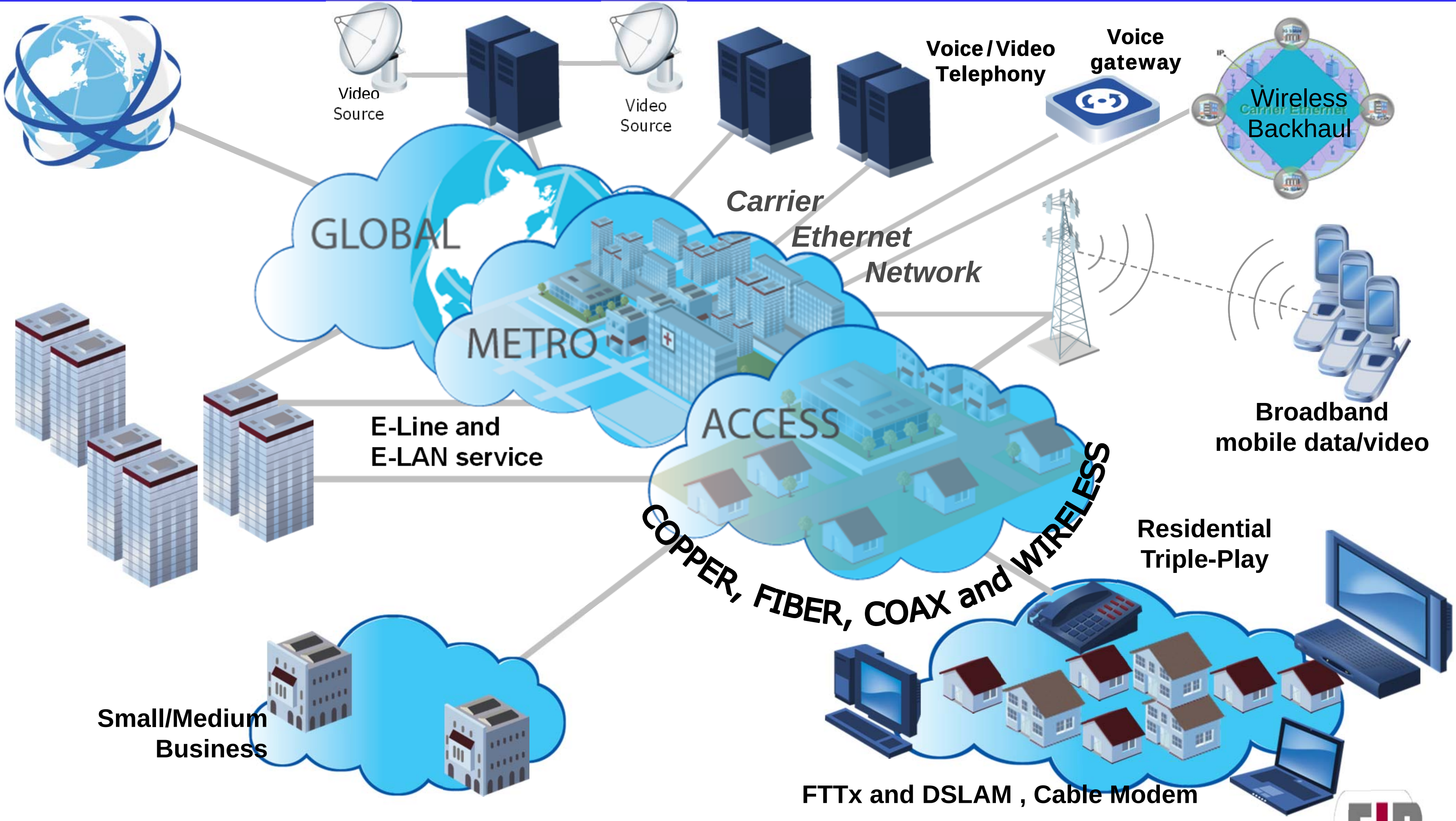
The way to PBB-TE/PBT

El problema a mesura que escala una xarxa, és que les taules de MAC i de routing cada cop són més ocmplexes.

per simplificar això, es divideix la xarxa en nivells



Carrier Ethernet Scope



2.5 Control de la congestió

CAP13

"estem parlant de xarxes de commutació de paquets"

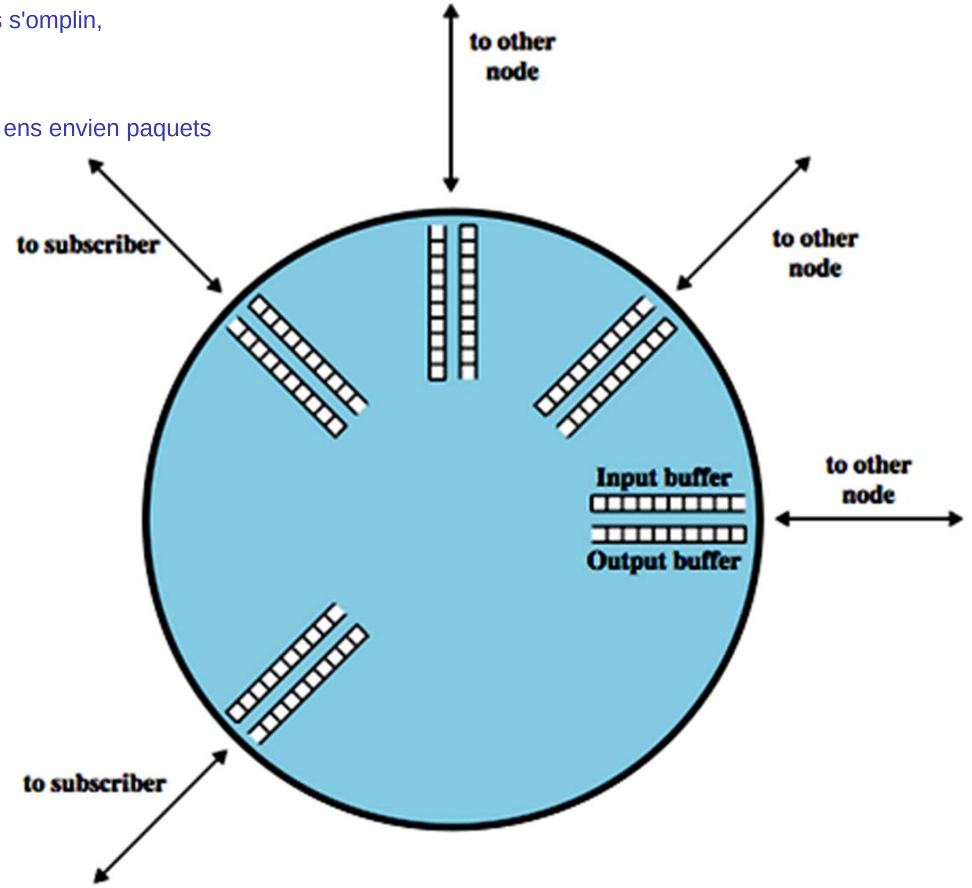
ATM

C13 - p3

Queues at a Node

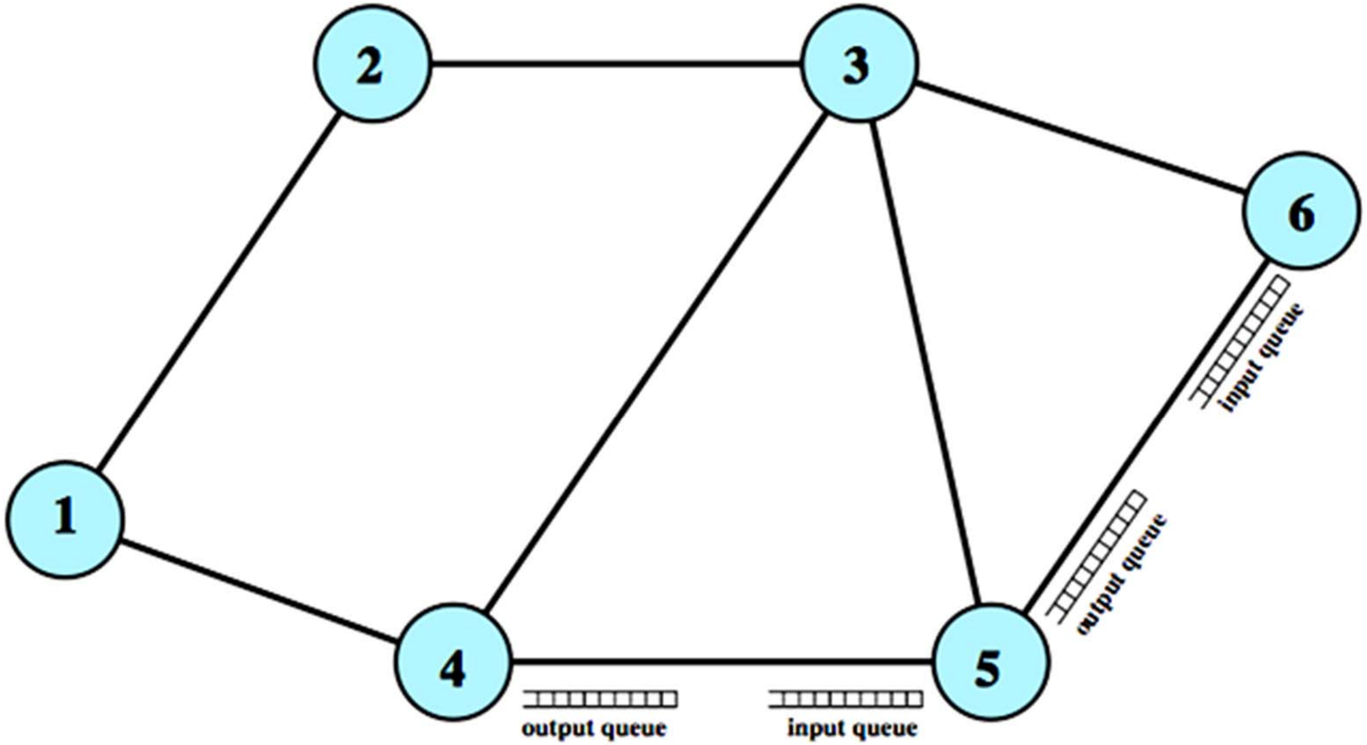
cues (buffers) en un commutador de paquets.
a cada port h ha un buffer d'entrada i un de sotid

es pot arribar al punt que els buffers s'omplin,
llavors podem tenim dues opcions:
1- descartar paquets
2- controlar el ritme al que els veins ens envien paquets



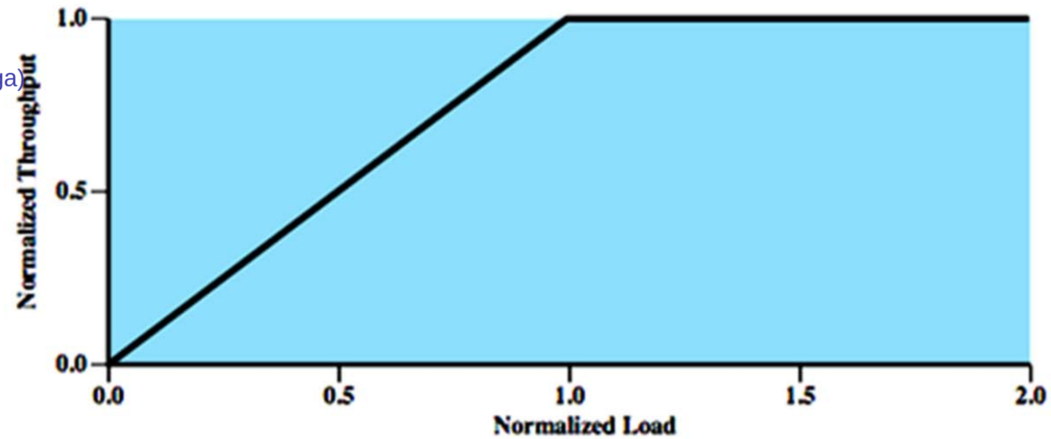
Interaction of Queues

El problema és que cada node ha de gestionar les seves pròpies cues.I la cngestió en un punt de la xarxa es pot propagar facilment



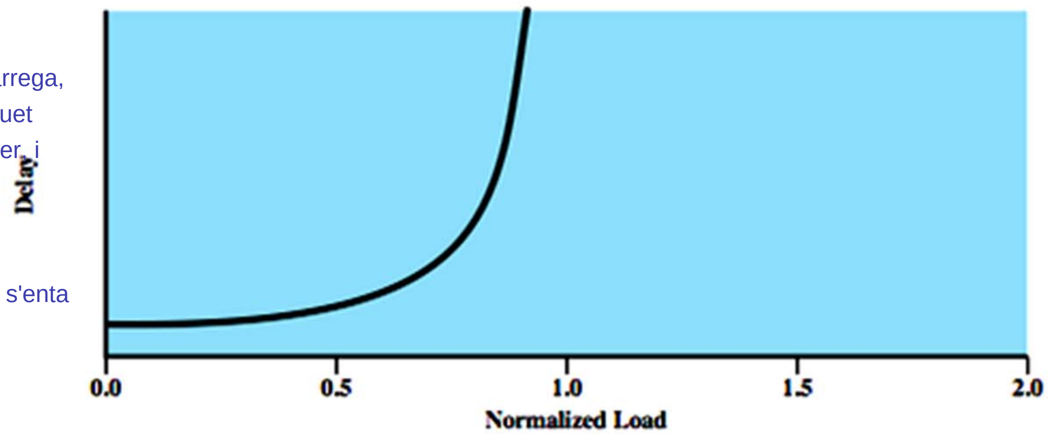
Ideal Network Utilization

no pot passar de 1 perque ja és el cas ideal, que el throughput (la trnasmissio util) sigui tot el trafic de la xarxa (càrrega)



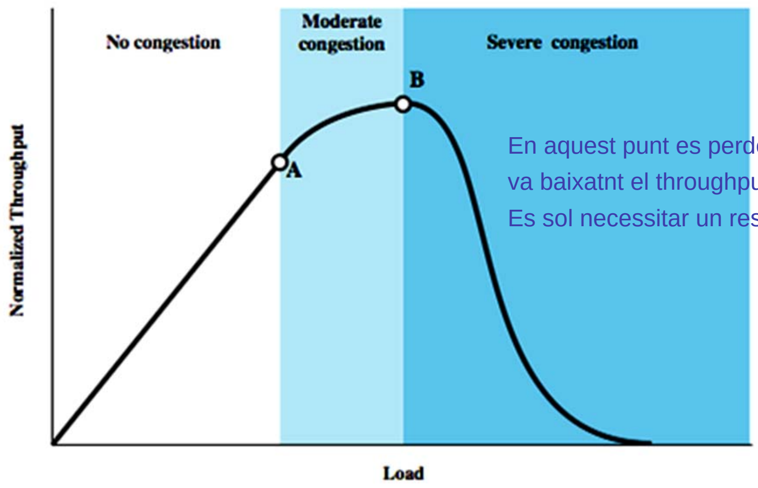
"el throughput no té perquè coincidir amb la càrrega
el througput és el tràfic desitjat (transmissió efectiva), i la càrega és el trafic real
ex: si envies un paquet, es perd i el tens que tornar a reenviar, has enviat 2 paquets (càrrega), però e ltràfic efectiu ha sigut e ld'un paquet
"

El delay es manté constant
(el delay mai és 0 per poc que sigui la càrrega, perquè en una xarxa de paquets, un paquet sempre entra en un buffer i surt d'un buffer, i això ja és un retràs)
quan s'incrementa la càrrega cada cop augmenta el delay
en càrrega 1 el temps és infinit ja que no s'enta en cap cua

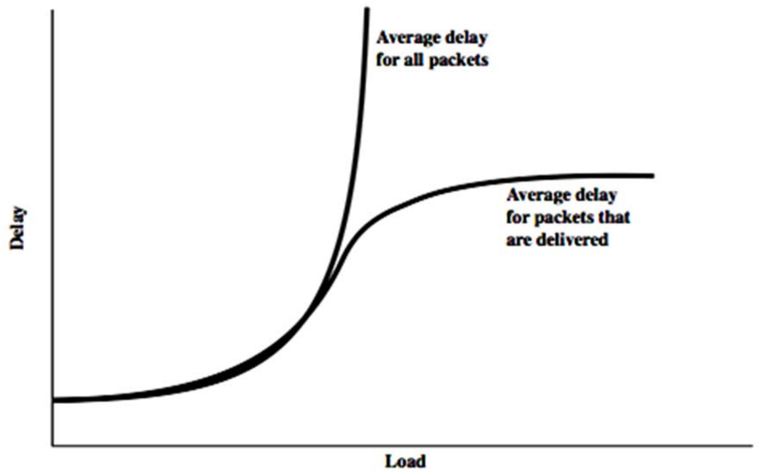


Effects of Congestion - No Control

aquest és el model real, l'anterior era teòric

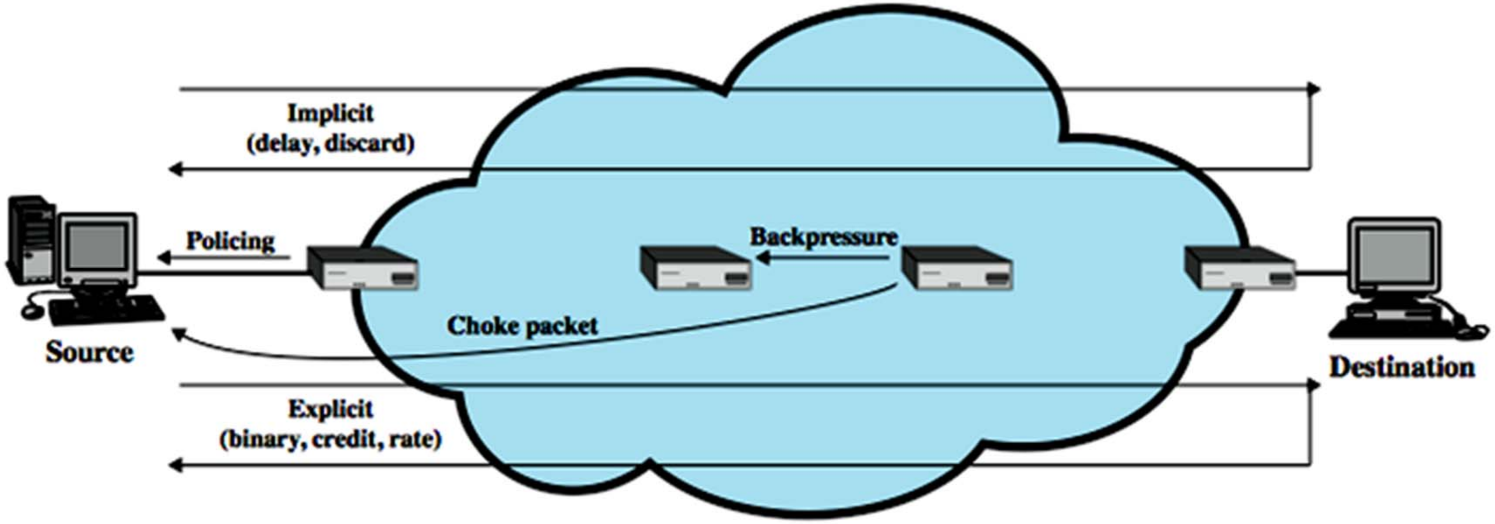


En aquest punt es perden tants paquts que la majoria del tràfic són paquets repetits, va baixatnt el throughput
Es sol necessitar un reset de la xarxa



Mechanisms for Congestion Control

tot això són mètodes que s'expliquen al llibre



Gestió del tràfic, les tècniques de l'apartat anterior són molt genèriques, per situacions més cncretes es poden fer servir altres tècniques i politiques de descart, com les seguents:

Traffic Management

C13 p11

fairness

- provide equal treatment of various flows

tenir una cua per connexio per exemple.

quality of service

- different treatment for different connections

reservations

- traffic contract between user and network
- excess traffic discarded or handled on a best-effort basis

C13 p13

Token bucket

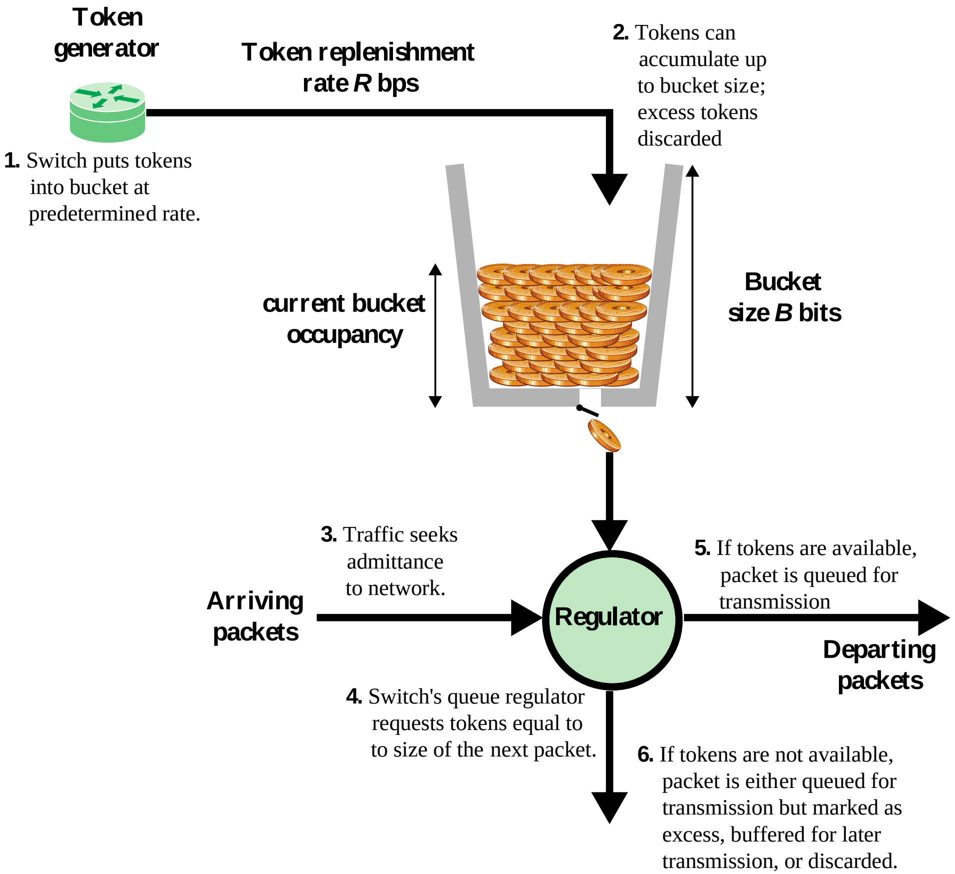


Figure 20.6 Token Bucket Scheme

Leacky bucket

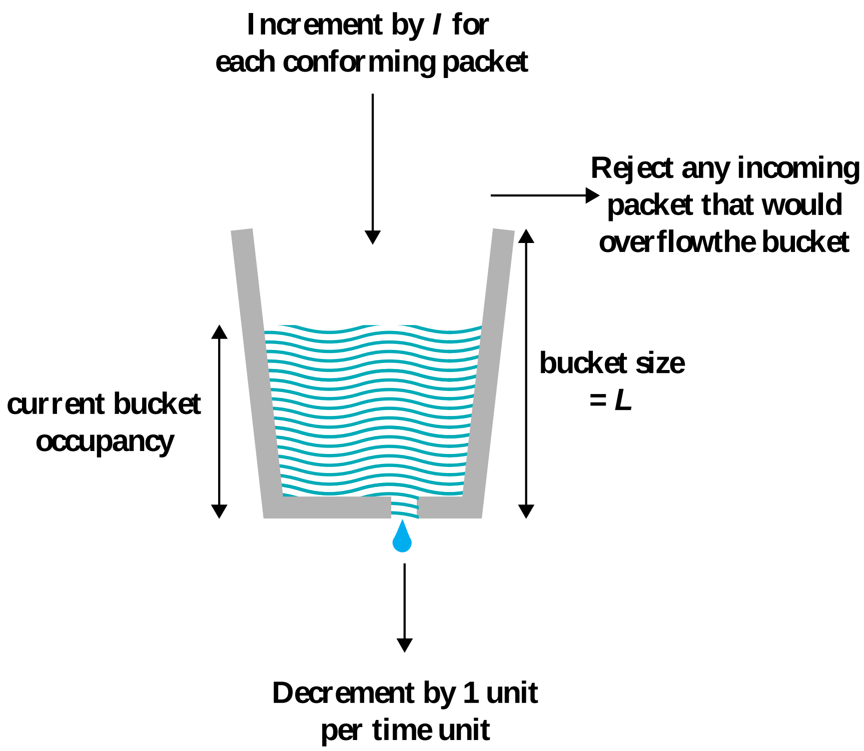
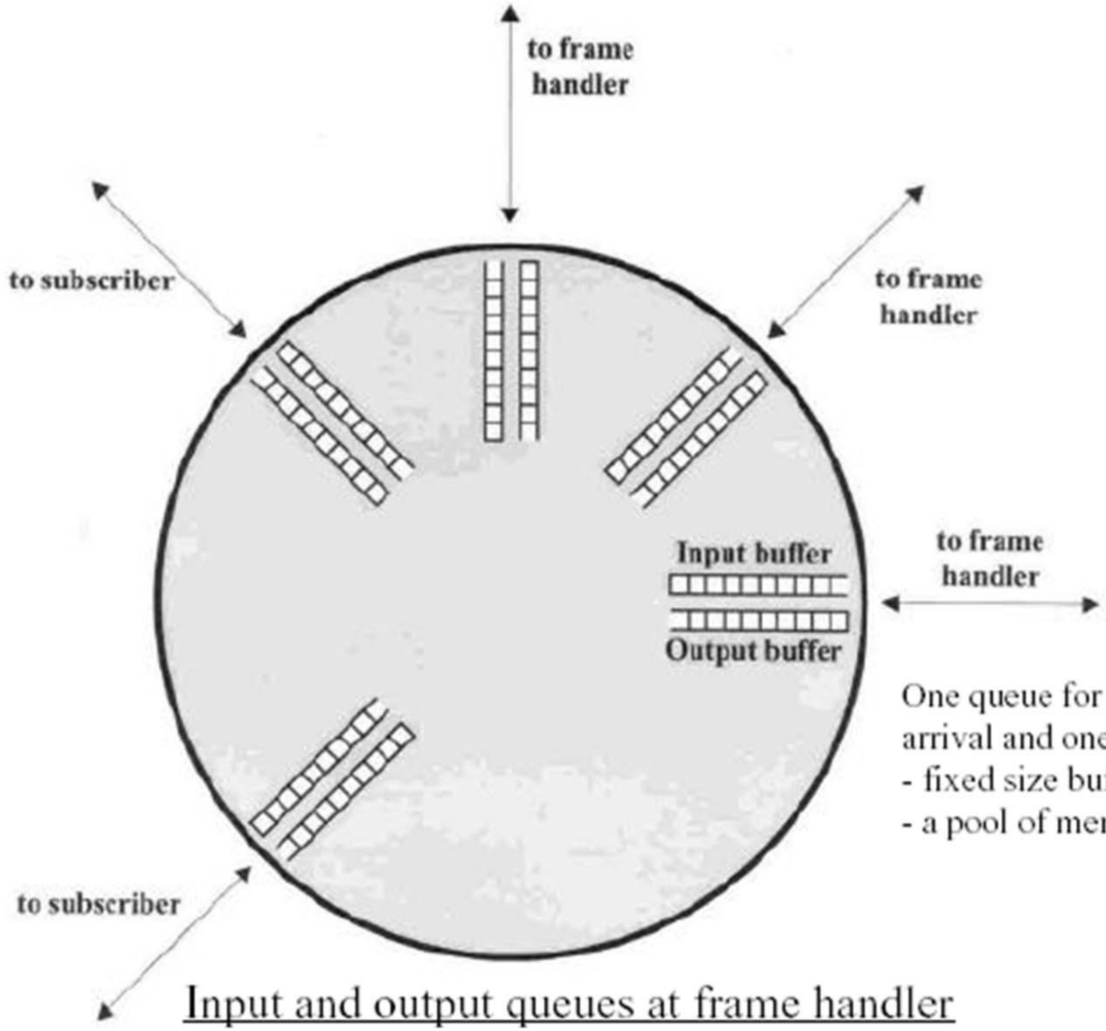


Figure 20.7 Leaky Bucket Algorithm

Exemple: FR Congestion



One queue for frame arrival and one for departure
- fixed size buffers
- a pool of memory

Input and output queues at frame handler

FR congestion control

Technique	Type	Function	Key Elements
Discard control	Discard strategy	Provides guidance to network concerning which frames to discard	DE bit
Backward explicit Congestion Notification	Congestion avoidance	Provides guidance to end systems about congestion in network	BECN bit or CLLM message
Forward explicit Congestion Notification	Congestion avoidance	Provides guidance to end systems about congestion in network	FECN bit
Implicit congestion notification	Congestion recovery	End system infers congestion from frame loss	Sequence numbers in higher-layer PDU

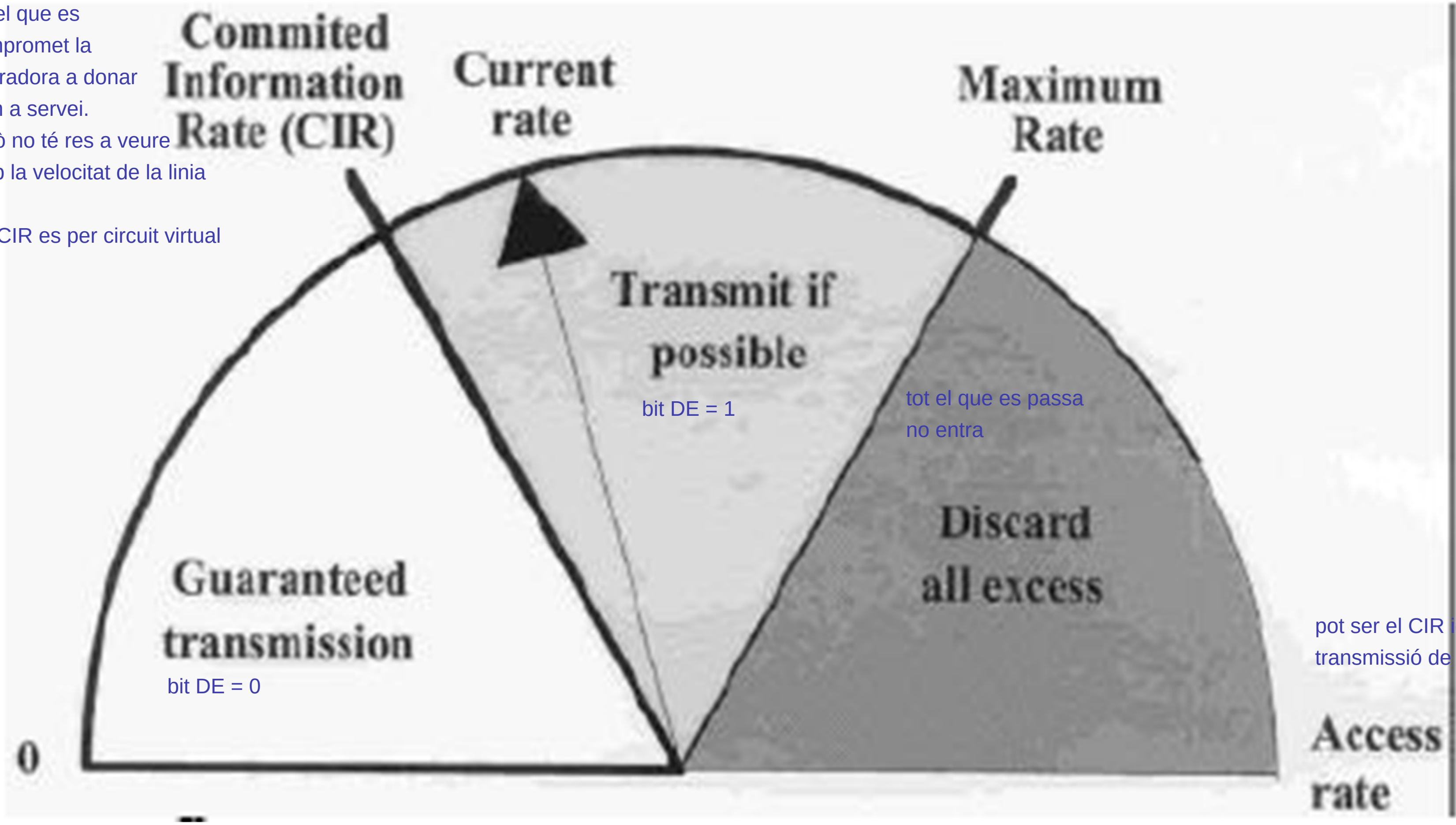
Sempre s'arriba a un punt en el que si la xarxa està saturada, s'han de descartat paquets. Un cop està un buffer ple i arriba un paquet, o es descarta el nou, o algun dels que hi havia per fer-li lloc
El CIR et dona una velocitat, el que estigui fora d'aquests b/s es susceptible a ser descartat.

C13 - p14 'gestió de la velocitat del trènist'

Discard Control

És el que es compromet la operadora a donar com a servei. però no té res a veure amb la velocitat de la linia

*El CIR es per circuit virtual



pot ser el CIR > Acces rate? no! seria impossible
----
pot ser la suma dels CIRs dels circuits més gran que la velocitat de la línia? no!
----
però en el cas que sabem els circuits que hi haurà treballant alhora si que pot ser que la suma de tots sigui superior
----

pot ser el CIR igual a la velocitat de transmissió de la linia? si

CIR =0 --> és best-effort

"EX: quan en ADSL et diuen 1 mb de pujada, no és el CIR"

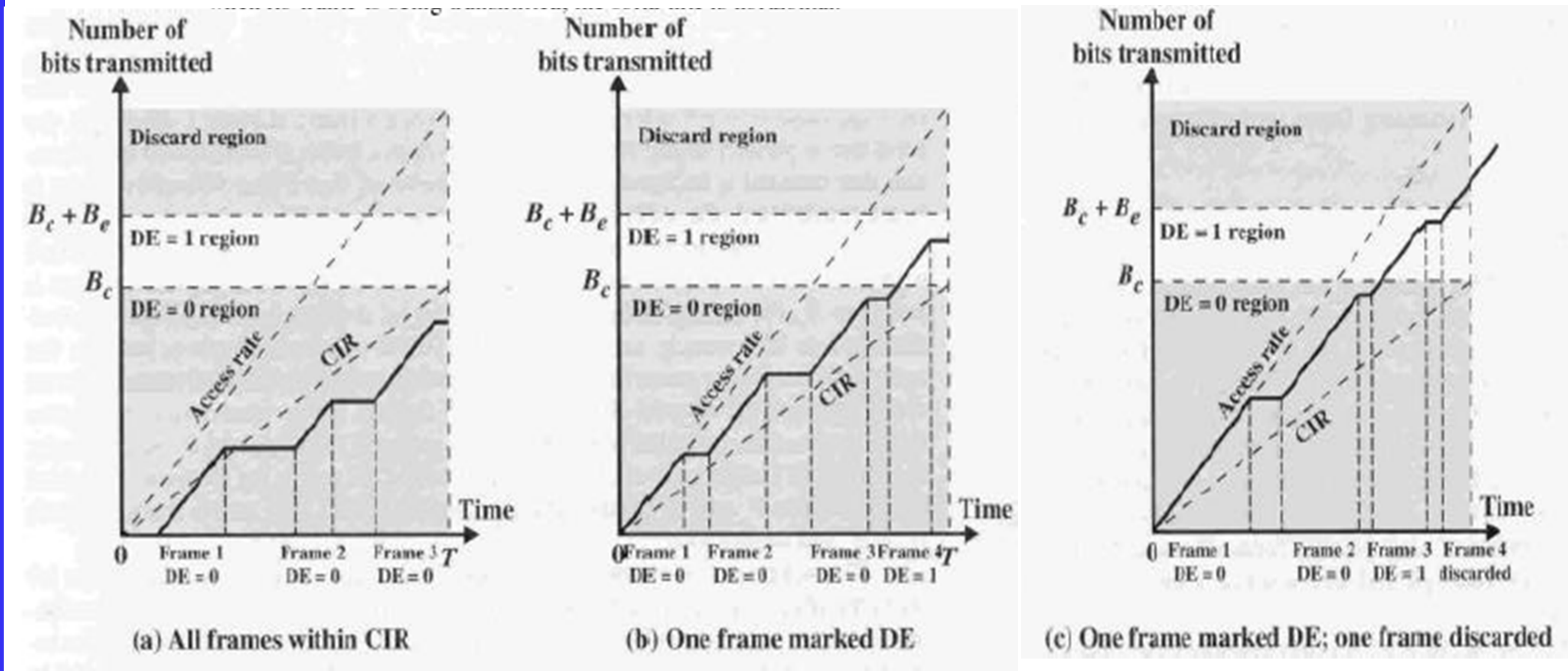
Bc mida de rafaga confirmada: dades de quantitat maxima que la xarxa accepta transferir, en un interval T (aquestes dades poden anar en una o en varies trames)

Be mida de rafaga excedida: quantitat de dades per sobre de Bc, que la xarxa intentarà transferir, en condicions normals en un interval T.

T = Bc / CIR

C13 p 15

Congestion control



Leaky Bucket