

QoS: Transporte de Voz

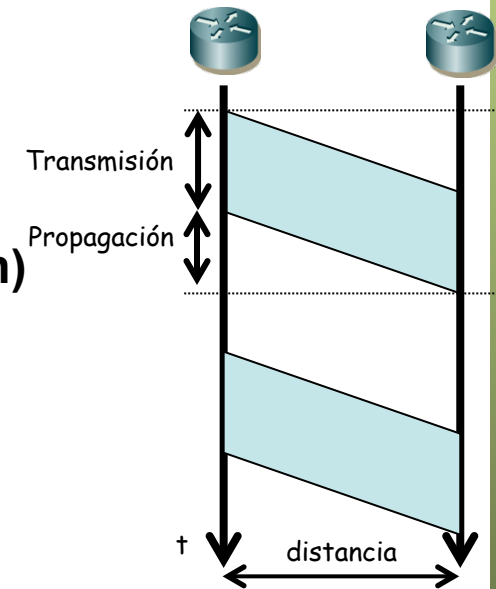
Area de Ingeniería Telemática
<http://www.tlm.unavarra.es>

Grado en Ingeniería en Tecnologías de
Telecomunicación, 3º

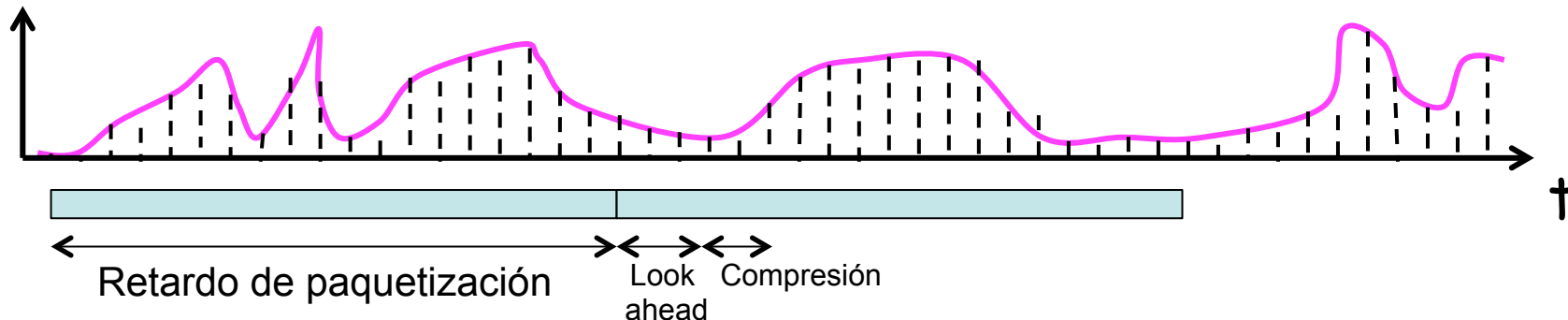
Parámetros en el transporte de voz

Componentes del retardo

- Retardo de paquetización
- Retardo de procesamiento del codificador
- *Algorithmic Delay (look ahead)*
- Retardo de serialización (tiempo de transmisión)
- Retardo de propagación
- Tiempo de procesamiento/conmutación
- *De-jitter delay*
- Retardo en cola
- Transcodificación



Ear-to-mouth delay (D)	R factor	Objective MOS
$D < 150$ ms	80–89	5
$150 \text{ ms} < D < 250$ ms	70–79	4
$250 \text{ ms} < D < 325$ ms	60–69	3
$325 \text{ ms} < D < 425$ ms	50–59	2
$D > 425$ ms	90–100	1



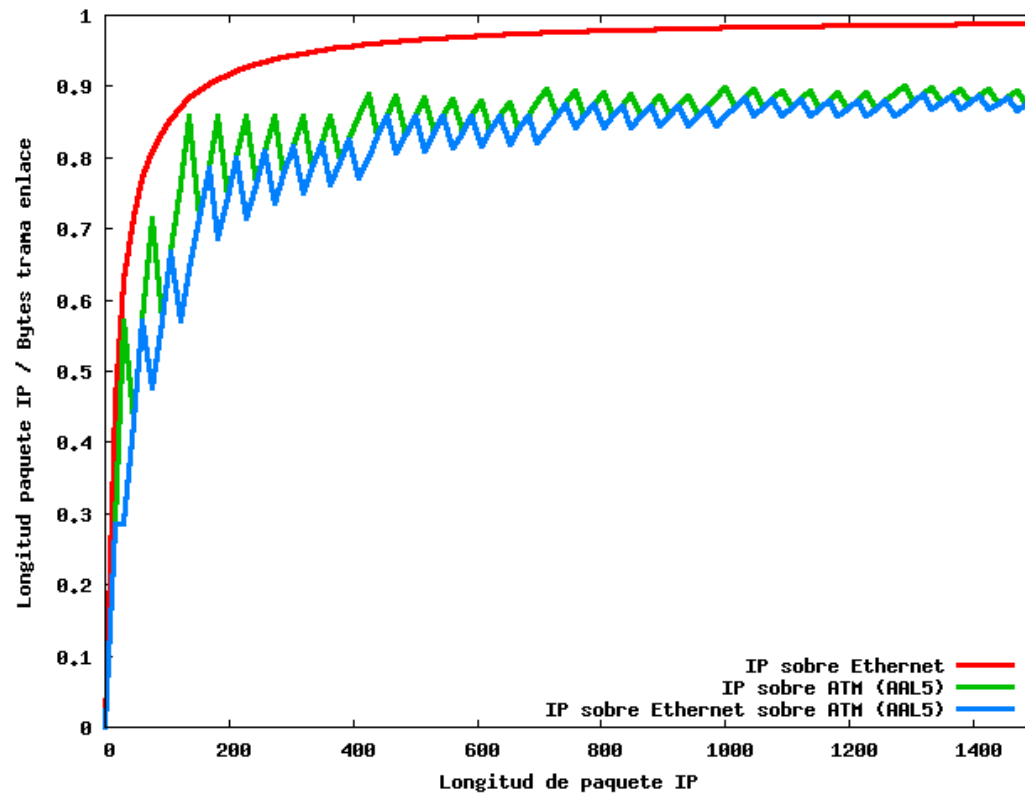
Throughput

- Generalmente los codecs producen un flujo a bitrate constante
- Esto puede no ser así si se emplea supresión de silencios (VAD, *Voice Activation Detection*)
 - Una conversación suele contener aproximadamente un 50% de silencios
 - VAD reduce el ancho de banda medio pero no el de pico
- La capacidad suele estar dimensionada para soportar la tasa de pico
- Esto no quita para que se haga sobresuscripción



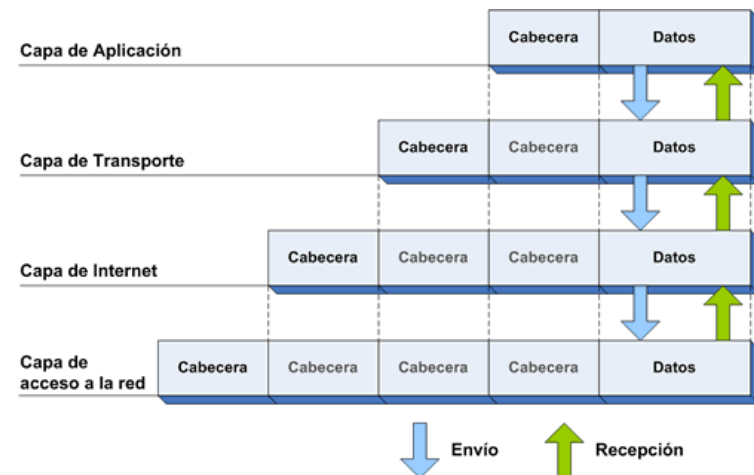
Throughput y encapsulado

- A la hora de asegurar un throughput hay que tener en cuenta que:
 - El servicio genera un bitrate a nivel de aplicación
 - Habrá al menos encapsulado IP (más transporte UDP?)
 - El encapsulado de nivel de enlace va a depender de cómo se haga el transporte de red



Throughput y encapsulado

- Hay que tener en cuenta en qué nivel se está asegurando el throughput
- Esto afecta a cualquier servicio
- Si se asegura un bitrate o pkts/s a nivel IP hay que tener en cuenta que al enviar se añade la cabecera de nivel de enlace
- Ejemplo:
 - Flujo A paquetes de 100 bytes, flujo B paquetes de 1000 bytes (a nivel IP)
 - Se asegura un reparto del 50:50 % de la capacidad del enlace a nivel IP
 - Enlace Ethernet
 - En un cierto periodo se enviarán 10 paquetes de A por cada paquete de B
 - Se han enviado $10 \times (100 + 18)$ bytes de A y $1000 + 18$ bytes de B
 - Eso son 1180 bytes de A y 1018 de B
 - Eso es un 54:46 %

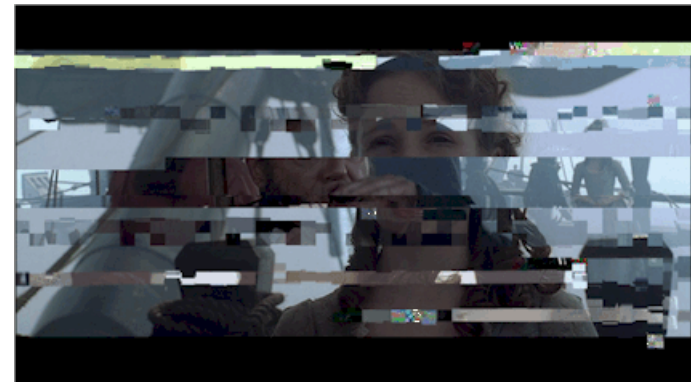


Encapsulado: Ejemplo

- Cada paquete suele llevar unos 20-30ms de muestras
- En el cálculo del BW hay que tener en cuenta la encapsulación
 - X bytes de *payload* (muestras de voz)
 - +12 de cabecera RTP
 - + 8 de cabecera UDP
 - +20 de cabecera IP (mínima sin opciones)
 - + Y bytes de cabecera de enlace
- Ejemplo: G.711 (64 Kbps, Con paquetes cada 20ms, 50 pps)
 - 8 muestras/ms, 1 byte/muestra, 20 ms/paquete \Rightarrow 160 bytes/paquete
 - $160+12+8+20 = 200$ bytes de paquete IP \Rightarrow 10.000 Bps (80 Kbps)
 - Enlace PPP (+6 Bytes) \Rightarrow 206 bytes/trama \Rightarrow 82.4 Kbps
 - o enlace Ethernet (+18 Bytes) \Rightarrow 218 bytes/trama \Rightarrow 87.2 Kbps
 - o enlace Frame Relay (+4 Bytes) \Rightarrow 204 bytes/trama \Rightarrow 81.6 Kbps
 - o ATM/AAL5-LLC/SNAP \Rightarrow 5 celdas = 265 bytes/paquete \Rightarrow 106 Kbps
- Ejemplo: G.729a (8 Kbps, con paquetes cada 20ms, 50 pps)
 - PPP = 26.4 Kbps, Ethernet = 29.6 Kbps, FR = 25.6 Kbps, ATM = 42.2 Kbps

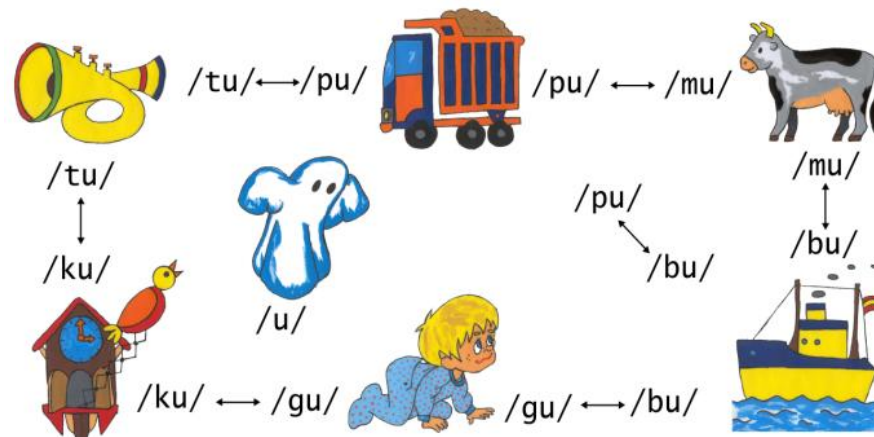
Pérdidas

- *Packet Loss Concealment* (PLC)
- Permite enmascarar el efecto de pérdida de paquetes de VoIP
- En codecs tipo G.711 se repite la última muestra
 - Se basa en que la onda cambia despacio
 - Se puede cubrir así hasta en torno a 20ms de muestras
 - La paquetización en el codec determina cuántas muestras hay en un paquete
 - Si se crean los paquetes conteniendo 20ms de muestras entonces dos o más pérdidas consecutivas degradan la calidad
 - Paquetes más grandes reducen la sobrecarga de cabeceras y por lo tanto el ancho de banda consumido
 - Sin embargo, si los paquetes contienen más de 20ms de muestras, puede que con PLC no se puedan mitigar las pérdidas



Pérdidas

- Codecs *frame-based* (G.729 y G.723) usan técnicas más sofisticadas, cubriendo pérdidas de hasta 30-40ms si no son fonemas cortos
- Se puede recuperar una pérdida pero mejor diseñar la red para pérdidas cercanas a 0 para el tráfico de voz



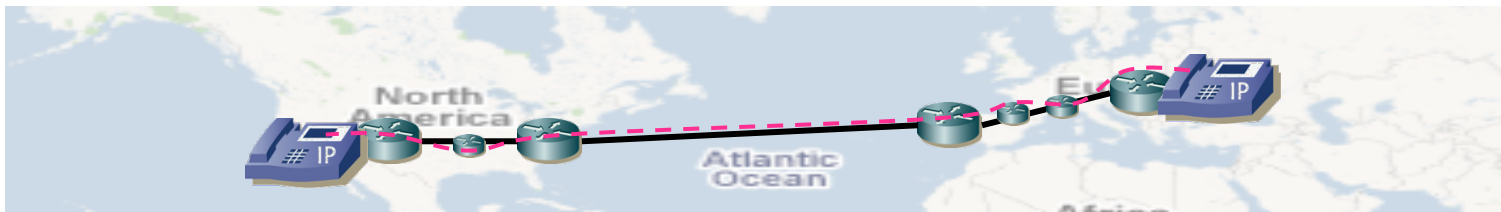
Codecs

ITU-T Codec	Codec type	Maximum codec delay (ms) (a1 d)	Bitrate (bps)	Packetization interval (ms) (b)	pps	Payload size (bytes)	IP pkt size (bytes) ¹	IP bps
G.711	PCM	0.375	64 000	10	100	80	120	96 000
G.711	PCM	0.375	64 000	20	50	160	200	80 000
G.711	PCM	0.375	64 000	30	33.33	240	280	74 659
G.723.1	ACELP	97.5	5 300	30	33.33	20	60	15 998
G.723.1	ACELP	97.5	5 300	15	16.67	40	80	10 669
G.726.16	ADPCM	0.375	16 000	10	100	20	60	48 000
G.726.16	ADPCM	0.375	16 000	20	50	40	80	32 000
G.726.16	ADPCM	0.375	16 000	30	33.33	60	100	26 664
G.726.24	ADPCM	0.375	24 000	10	100	30	70	56 000
G.726.24	ADPCM	0.375	24 000	10	50	60	100	40 000
G.726.24	ADPCM	0.375	24 000	10	33.33	90	130	34 663
G.726.32	ADPCM	0.375	32 000	10	100	40	80	64 000
G.726.32	ADPCM	0.375	32 000	20	50	80	120	48 000
G.726.32	ADPCM	0.375	32 000	30	33.33	120	160	42 662
G.726.40	ADPCM	0.375	40 000	10	100	50	90	72 000
G.726.40	ADPCM	0.375	40 000	20	50	100	140	56 000
G.726.40	ADPCM	0.375	40 000	30	33.33	150	190	50 662
G.728	LD-CELP	1.875	16 000	10	100	20	60	48 000
G.728	LD-CELP	1.875	16 000	20	50	40	80	32 000
G.728	LD-CELP	1.875	16 000	30	33.33	60	100	26 664
G.729A	CS-ACELP	35	8 000	10	100	10	50	40 000
G.729A	CS-ACELP	35	8 000	20	50	20	60	24 000
G.729A	CS-ACELP	35	8 000	30	33.33	30	70	18 665

Cálculo de retardos: Ejemplo

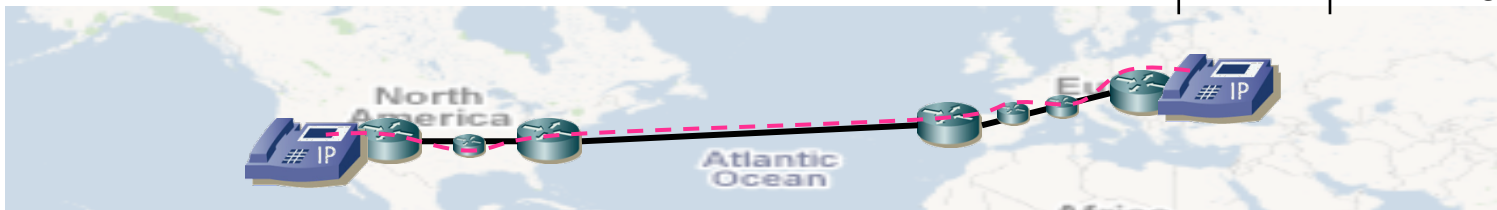
Ejemplo de retardo VoIP

- Llamada transatlántica
- Estimación de caso peor
- Haciendo prioritario el tráfico de voz
- Solo 1 llamada de voz es tráfico prioritario



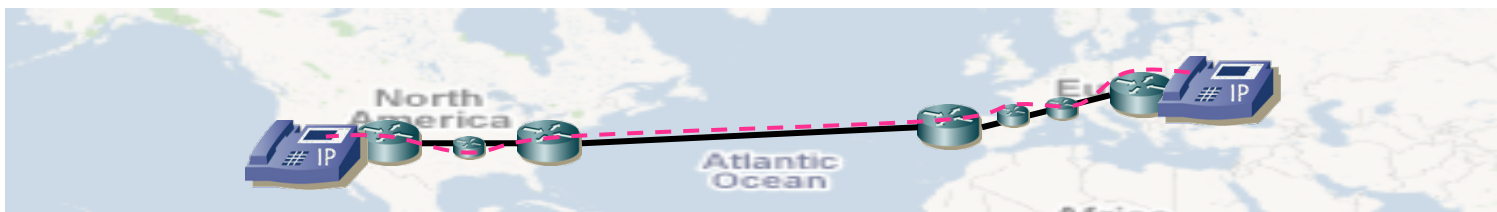
Ejemplo: max. end-to-end delay

1 sola llamada de voz, tráfico prioritario	mseg	Quedan
Delay Budget	150	
Retardo de codificación	10	140
Paquetes por ejemplo de 206 Bytes (G.711 sobre Ethernet, 87,2 Kbps)		
Enlaces de acceso upstream de 512 Kbps, tiempo de transmisión aprox.	3,2	136,8
Caso peor encuentra un paquete de tam. MTU empezando a transmitirse	23,7	113,1
En el camino enlaces a 100 Mbps		
Por ej. otros 9 routers en europa + internacional (ignorando sw. capa2) $10 \times (Ttx + Ttx_max_pkt) = 10 \times (0,016 + 0,121)$	1,4	111,7
Propagación intra-europea aprox. (según trayecto, ej. 3.000Km)	15	96,7
Propagación transatlántica aprox. (según cable)	45	51,7
Por ejemplo 9 conmutadores en USA + internacional $10 \times (Ttx + Ttx_max_pkt) = 10 \times (0,016 + 0,121)$	1,4	50,3
Propagación en EE.UU., por ejemplo 2.000Km	10	40,3
Enlace acceso a 6 Mbps, downstream, transmisión + pkt máximo aprox. $Ttx + Ttx_max_pkt = 0,3 + 2$	2,3	38
Remanente		38



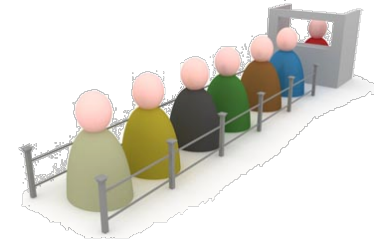
Ejemplo: retardo

- Solo hemos considerado *store-and-forward* en los routers IP
- Pero seguramente habrá en capa 2 si el transporte son tecnologías de conmutación de paquetes



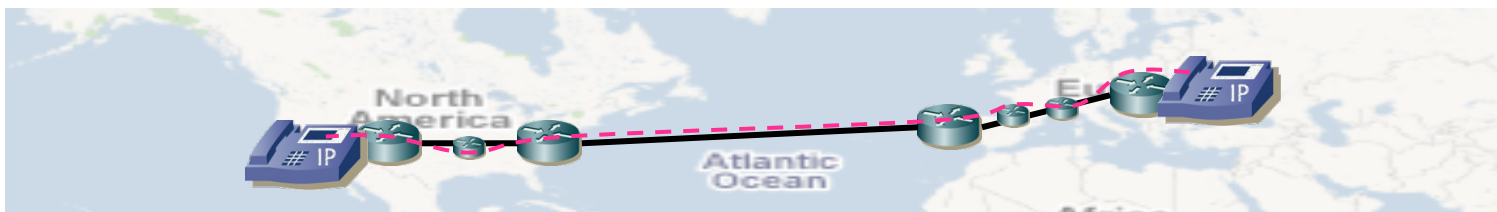
Ejemplo: retardo

- Total: 112ms, pero es caso peor
- Mínimo: 80ms (sin otro tráfico)
- La diferencia entre máximo y mínimo: unos 30ms
- La componente variable viene del retardo en cola
- Con una sola llamada y prioridad alta es el retardo de encontrar un paquete delante (en cada salto)



Ejemplo: Jitter?

- El retardo variable viene del tamaño de esos paquetes delante
- Se podría calcular un retardo medio con tamaños medios de paquete
- El retardo en cola es lo que contaríamos en el *de-jitter buffer*
- Con retardo mínimo de 80ms quedan 70ms para este buffer
- Los retardos fijos pueden dejar poco margen de maniobra al jitter y por lo tanto exigir SLAs estrictos



Ejemplo: 2 llamadas en acceso

- $2 \times 87,2 \text{ Kbps} = 174,4 \text{ Kbps} < 512 \text{ Kbps}$
- Caso peor:
 - Llega y hay un paquete de datos enviándose
 - Y mientras espera llega uno de voz de la otra llamada
 - Ese último paquete espera al de datos y al de voz
- Incrementa en el tiempo de transmisión del paquete de voz
- Paquete pequeño así que retardo pequeño
- Salvo que la velocidad sea baja (el acceso a 512Kbps)



Ejemplo: N Llamadas

- N llamadas pero solo en los troncales (no en el acceso)
- No siguen el mismo camino
- En cada salto troncal hay que tener en cuenta en el caso peor el retardo de N paquetes de alta prioridad + 1 de baja
- Ejemplo:
 - 100Mbps, digamos 20Mbps para voz, a 87,2Kbps/llamada son unas 230 llamadas
 - En un salto a 100Mbps en caso peor $230 \times 0,016 + 0,121 = 3,801\text{ms}$
 - Sigue siendo la peor componente el paquete grande en el acceso de baja velocidad (23ms a 512Kbps)
 - La siguiente serían las llamadas de voz en los enlace de baja capacidad



Máximo retardo de propagación

- Hemos estimado 60ms de propagación
 - Circunferencia (aprox) de la Tierra: 40.000Km => máx. distancia: 20.000Km
 - 20.000Km a 200.000Km/s = 100ms
 - Caso peor de distancia pero caso mejor de línea recta
 - Deja poco margen
 - Adaptar las expectativas
- Si se cambia de codec en el camino se incrementa el retardo en transcodificar
- ¿ Y si queremos encriptación (VPN) ? Añade más retardo de codificación/decodificación y en cabeceras



Transporte de voz

Transporte de voz

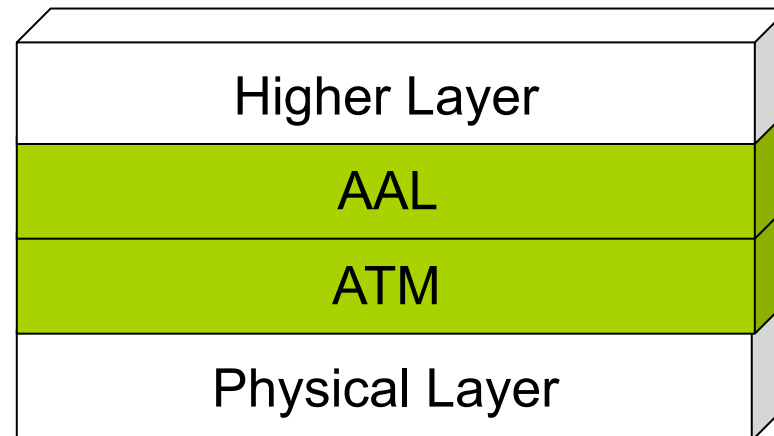
- Sobre red de conmutación de paquetes
- Múltiples formas según la tecnología de red
- Y cada una puede ofrecer diferentes mecanismos

Transporte de voz

- VoATM: CES con AAL1 estructurado o no, trunking con AAL2
- VoIP y DiffServ
- Voz y MPLS: TDMoMPLS, VoMPLS con AAL2
- Voz y Carrier Ethernet: CESoETH

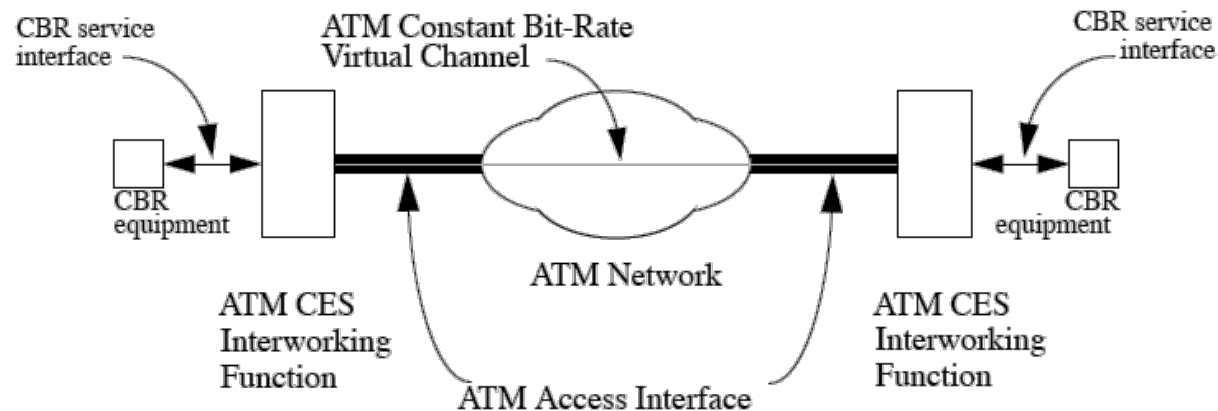
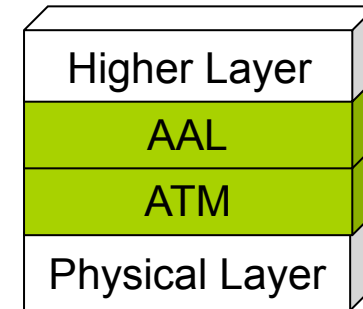
Voice over ATM (VoATM)

- No ha logrado llegar hasta el escritorio
- Modelos de transporte de voz
 - Voice trunking: tunneling del tráfico de voz. Adecuado para interconectar PBXs
 - Voice switching: la red ATM interpreta la señalización y encamina la llamada



ATM: Circuit Emulation Services

- CES
- ITU-T I.363.1
- Emplea AAL1 para transportar un flujo CBR
- En el PVC es importante la CDV además del PCR
- *Unstructured AAL1*: transporte de DS1/E1, DS3/E3
- *Structured AAL1*
 - Transporte de DS1/E1, permite no enviar los DS0 no utilizados (entrada structured E1/T1 Nx64)
 - Puede repartir los DS0 entre varios destinos
- Simple



DB-CES

- Dynamic Bandwidth CES
- Reconoce la señalización (ej: on-hook, off-hook) CCS o CAS (entonces debe ser structured)
- Envía celdas solo cuando hay llamada establecida en un DS0

ATM: Trunking con AAL2

- ITU-T I.363.2
- Transporte de voz comprimida con detección/supresión de silencios y eliminación de canales inactivos
- Múltiples canales de voz en un circuito
- VCCs VBR
- Modo *Non-switched trunking*
 - Cada canal de voz siempre en el mismo canal AAL2 en el mismo VCC
 - No procesa señalización
- Modo *Switched trunking*
 - Procesa señalización
 - El extremo selecciona el VCC y canal AAL2 en que colocar el canal de voz

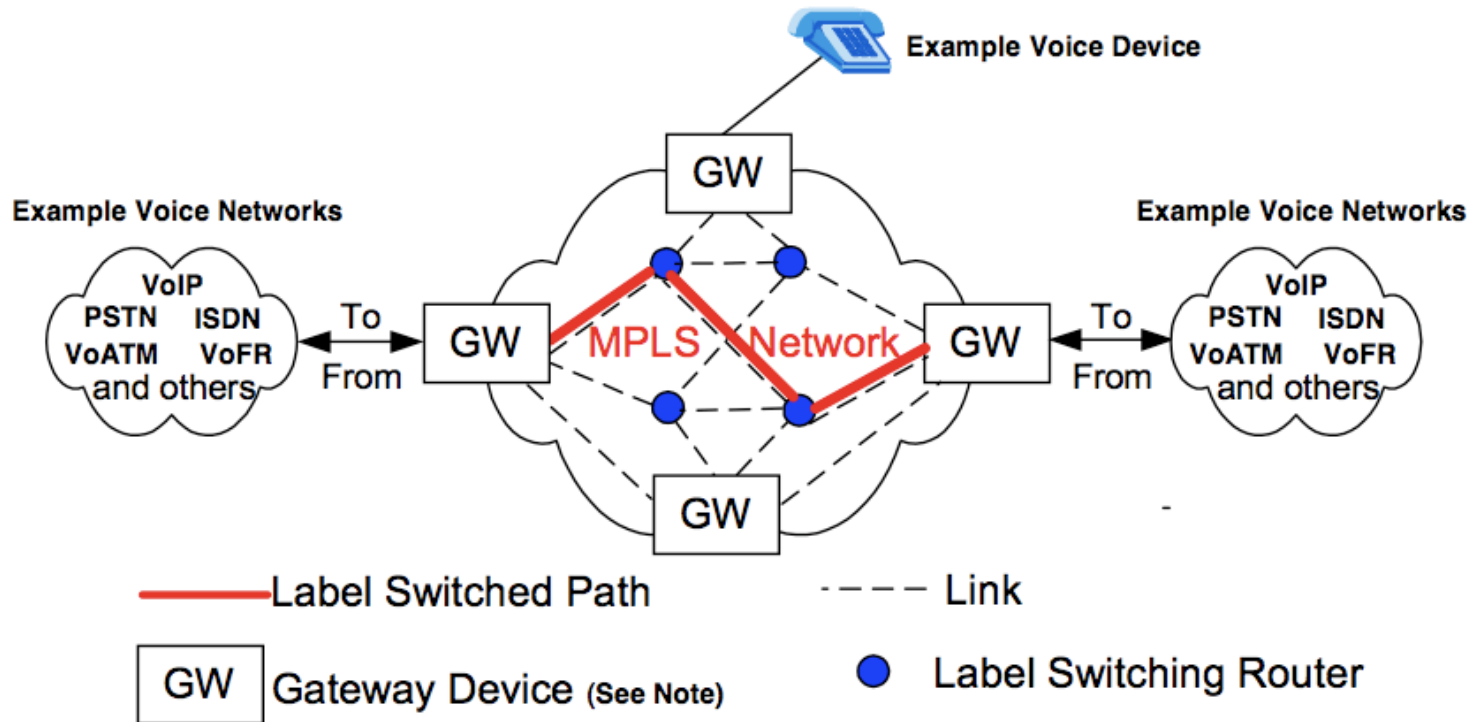
VoIP y DiffServ

- Clasificación en función de interfaz, puertos TCP/UDP, IP Precedence, DSCP, direcciones IP, etc
- Expedited Forwarding (EF)
- Marcado de EF DSCP 101110
- Planificador con prioridad (PQ, LLQ)
- Para la señalización:
 - DSCP CS5 (RFC 4594) o CS3 (Cisco)
 - En torno a 150 bps garantizados por llamada
 - No usar mecanismos de AQM
- (...)

VoIP y DiffServ

- Limitar la cantidad de voz con CAC (si es caso limitar local al nodo con policer)
- Local CAC
 - El propio gateway determina si tiene suficientes recursos
 - Si tiene suficiente memoria o DSP para soportar la llamada
- Network CAC
 - Validar que la red tiene suficientes recursos
 - Retardo, pérdidas y jitter
 - Si no hay camino con calidad entonces usar la PSTN (*PSTN fallback*)
- RSVP

VoMPLS

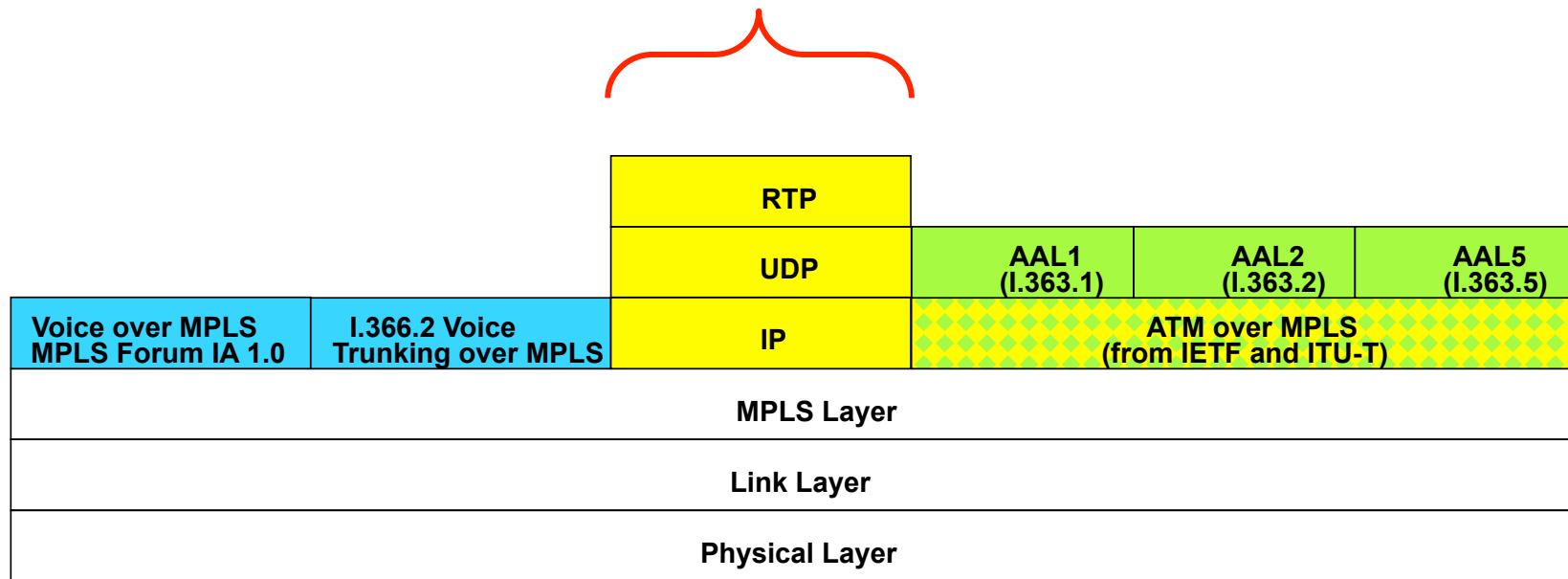


VoMPLS

ITU-T Y.1414 “Interfuncionamiento de los servicios vocales y las redes con conmutación por etiquetas multiprotocolo”

Encapsulación de audio codificado en paquetes MPLS:

- Voz sobre IP sobre MPLS (simple transporte de IP)
- (...)



Legend:

Alliance

IETF

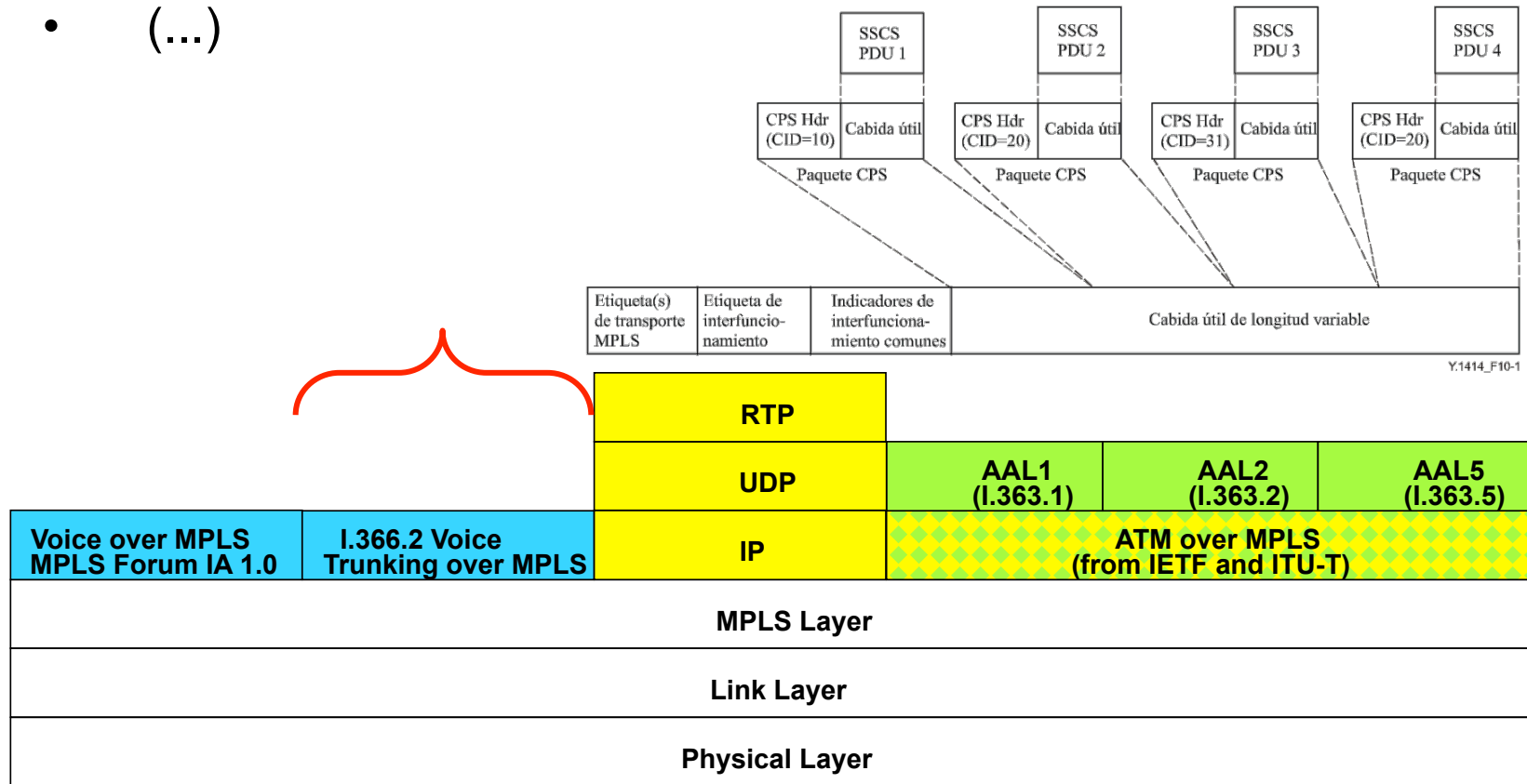
ITU-T

VoMPLS

ITU-T Y.1414 “Interfuncionamiento de los servicios vocales y las redes con conmutación por etiquetas multiprotocolo”

Encapsulación de audio codificado en paquetes MPLS:

- Voz sobre MPLS usando AAL2 (ITU-T I.366.2)
- (...)



Legend:

Alliance

IETF

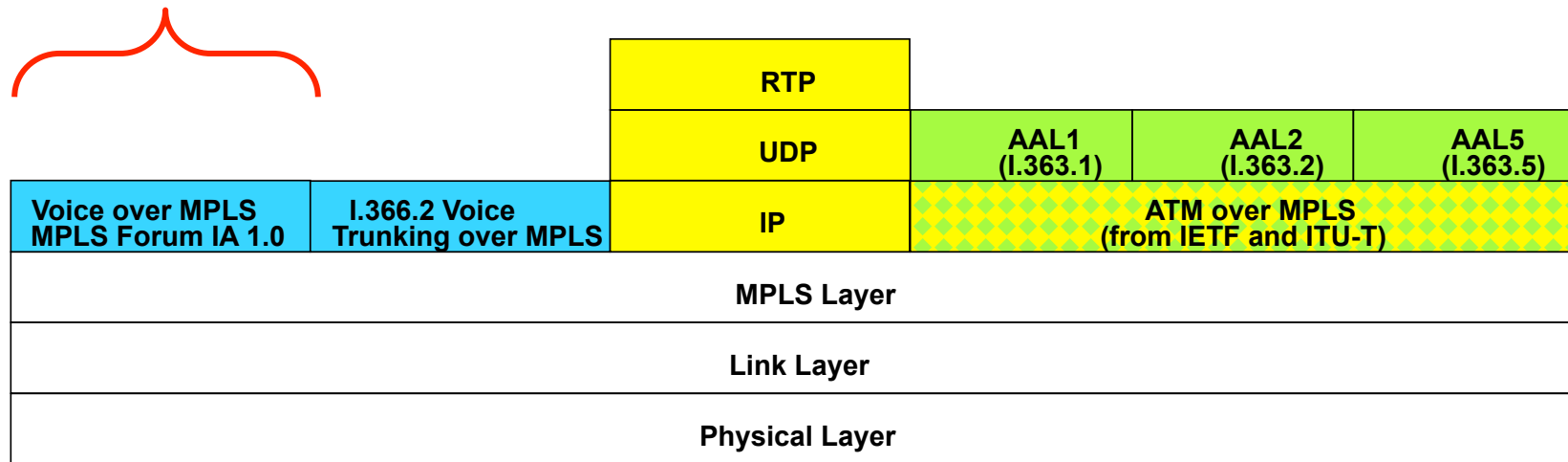
ITU-T

VoMPLS

ITU-T Y.1414 “Interfuncionamiento de los servicios vocales y las redes con conmutación por etiquetas multiprotocolo”

Encapsulación de audio codificado en paquetes MPLS:

- MPLS Forum: “Voice over MPLS - Bearer Transport Implementation Agreement 1.0”
 - Comprimida o sin comprimir
 - Supresión de silencios
 - Transporte de señalización



Legend:

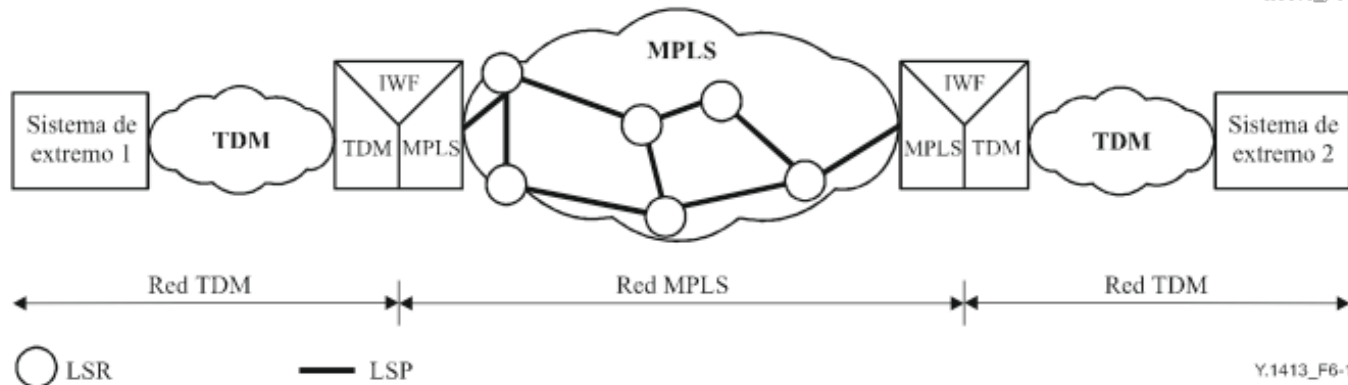
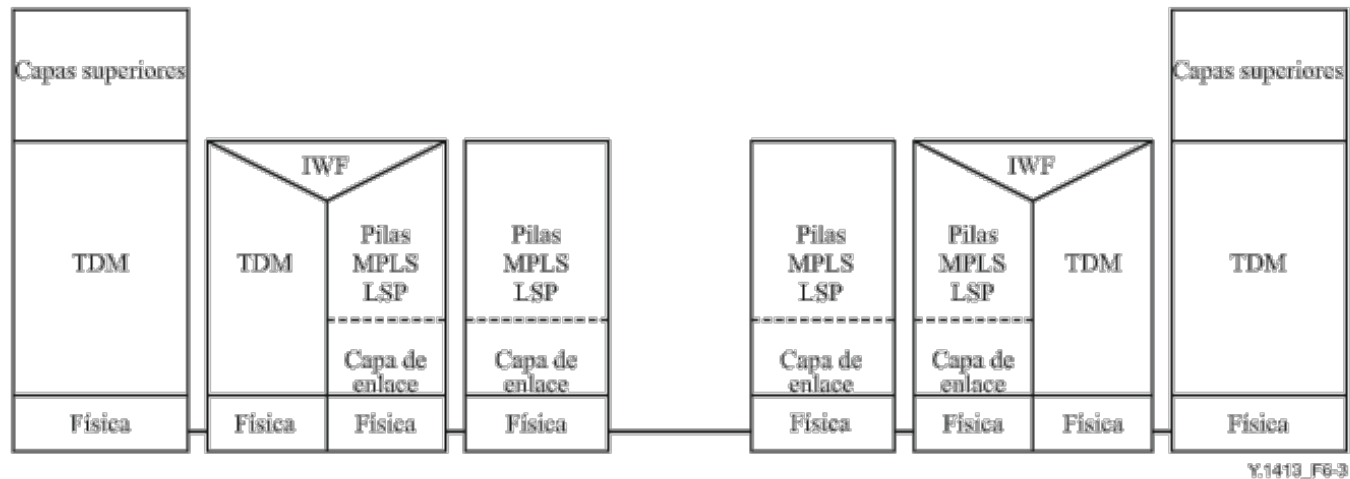
Alliance

IETF

ITU-T

TDMoMPLS

- ITU-T Y.1413 “TDM-MPLS network interworking -- User plane interworking”
- TDM hasta T3/E3
- Temporización de señal externa o recuperada por métodos adaptativos
- Varias conexiones TDM pueden ir en el mismo LSP
- BW en el LSP (bidireccional) debe ser suficiente para todas

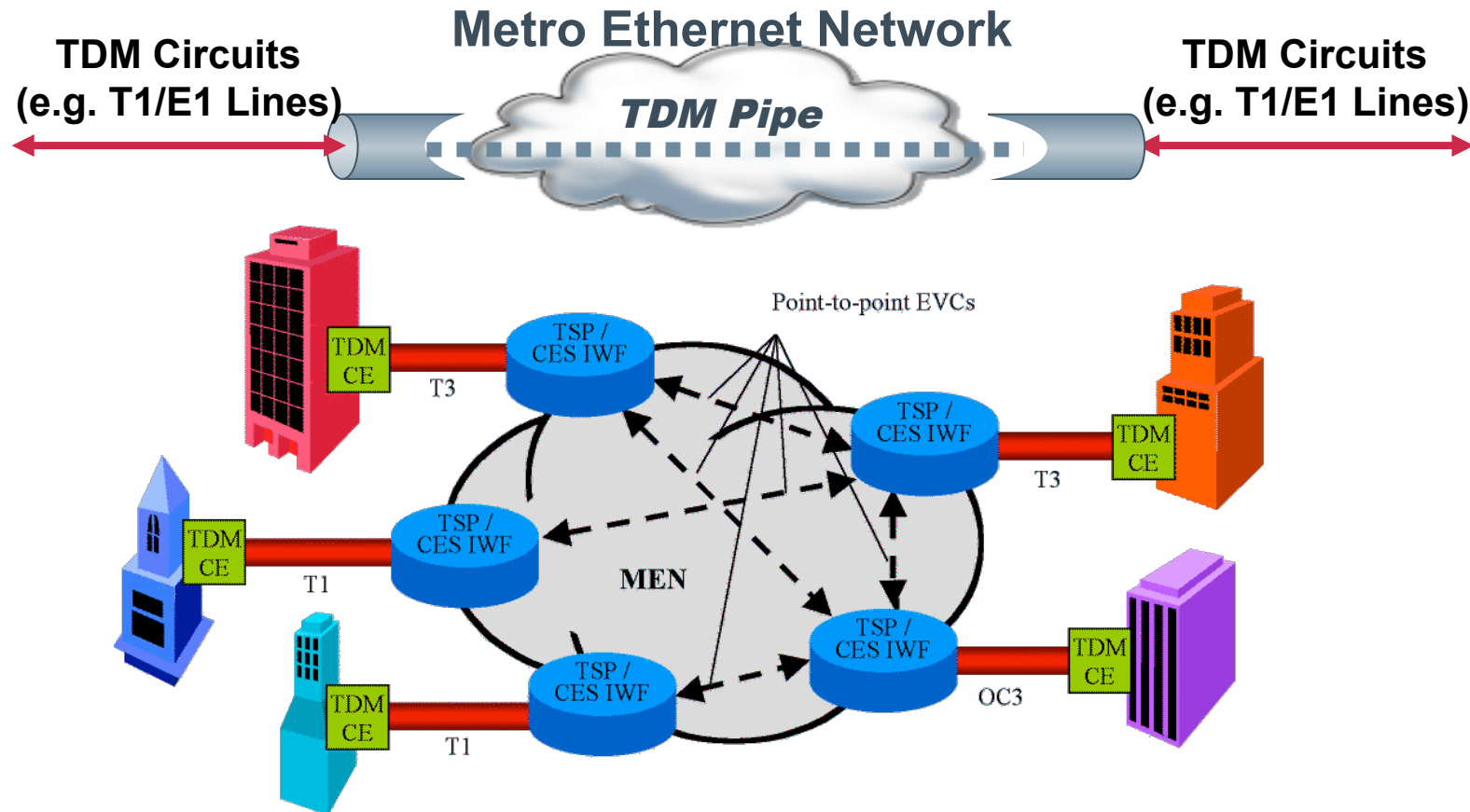


TDMoMPLS

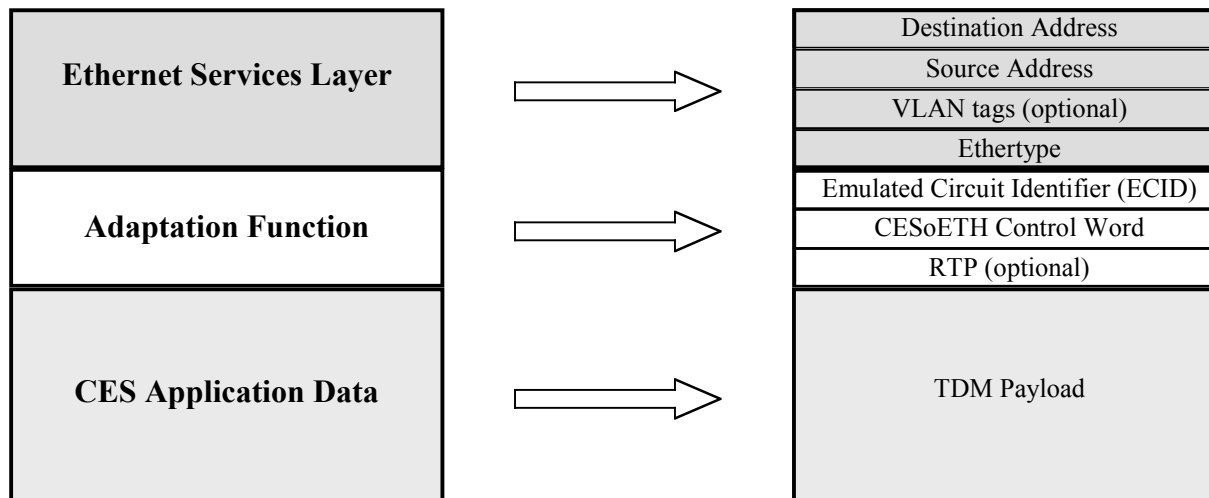
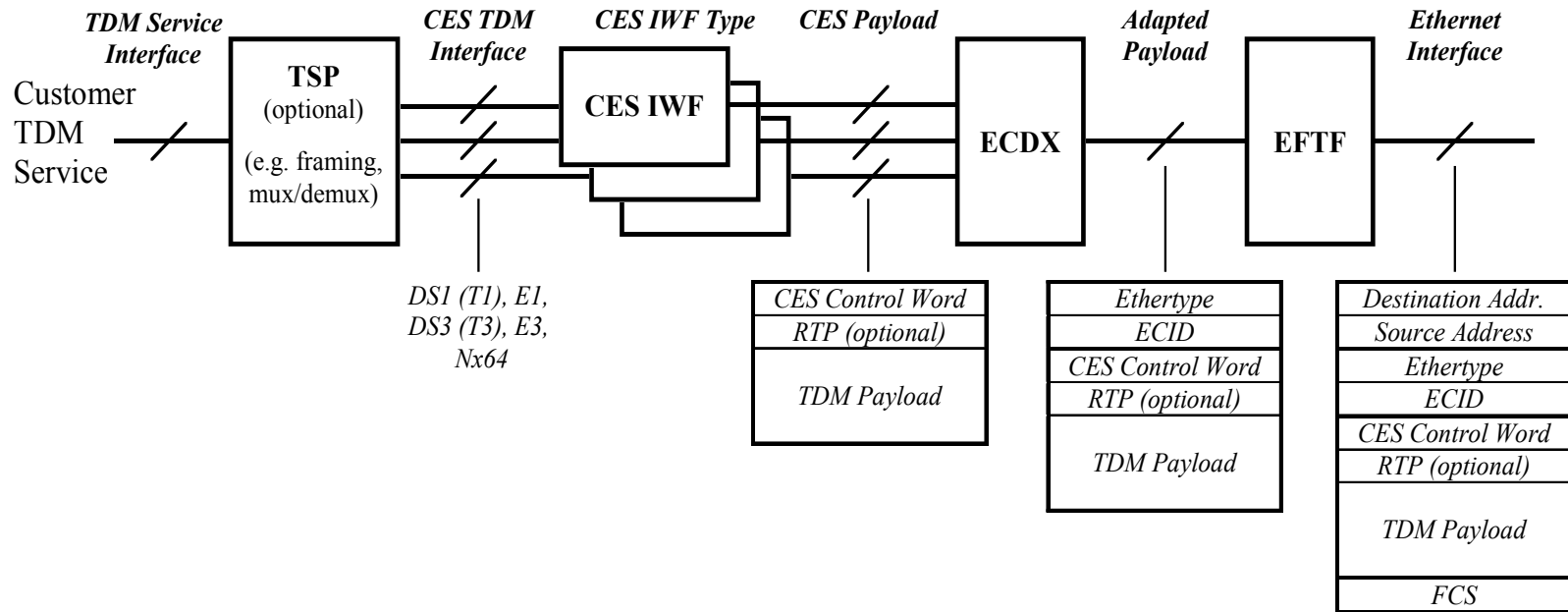
- Si emplea DiffServ entonces usará el PHB EF
- Si emplea IntServ entonces se usará el *Guaranteed Service* con reserva de BW adecuada
- Transporte agnóstico con respecto a la estructura
 - Ignora la estructura TDM
 - El número de bytes por paquete es configurable
- Transporte atento a la estructura
 - Puede usar AAL1

CESoETH

- Circuit Emulation Service (CES). TDM Line (T-Line) Service
- Transporte de circuitos TDM por la MEN
- Tanto PDH (Nx64, T1/E1, T3/E3) como SONET/SDH (STS-1, STM-1, STS-3, STM-3, STM-4, etc.)



CESoETH



Resumen

- El encapsulado consume capacidad
- El encapsulado afecta en cómo se reparte la capacidad si se piensa en flujos de alto nivel
- Se pueden mitigar pérdidas de hasta 20ms (1pkt)
- En llamadas internacionales el retardo de propagación es una componente importante
- Con enlaces de baja capacidad lo es también el retardo de serialización en ellos
- Voz sobre WAN
 - Emulación de circuito (transporte de flujo TDM)
 - Transporte de llamadas con atención a señalización

