

QoS: Transporte de Voz

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

Grado en Ingeniería en Tecnologías de
Telecomunicación, 3º

Temas de teoría

1. Introducción
2. QoS
3. Encaminamiento dinámico en redes IP
4. Tecnologías móviles
5. Otros temas

Objetivos

- Ser capaz de comprender diferentes arquitecturas de red y servicio para telefonía
- Ser capaz de desglosar las componentes del retardo extremo a extremo para el servicio de voz digital

Voz digital en la PSTN: Elementos

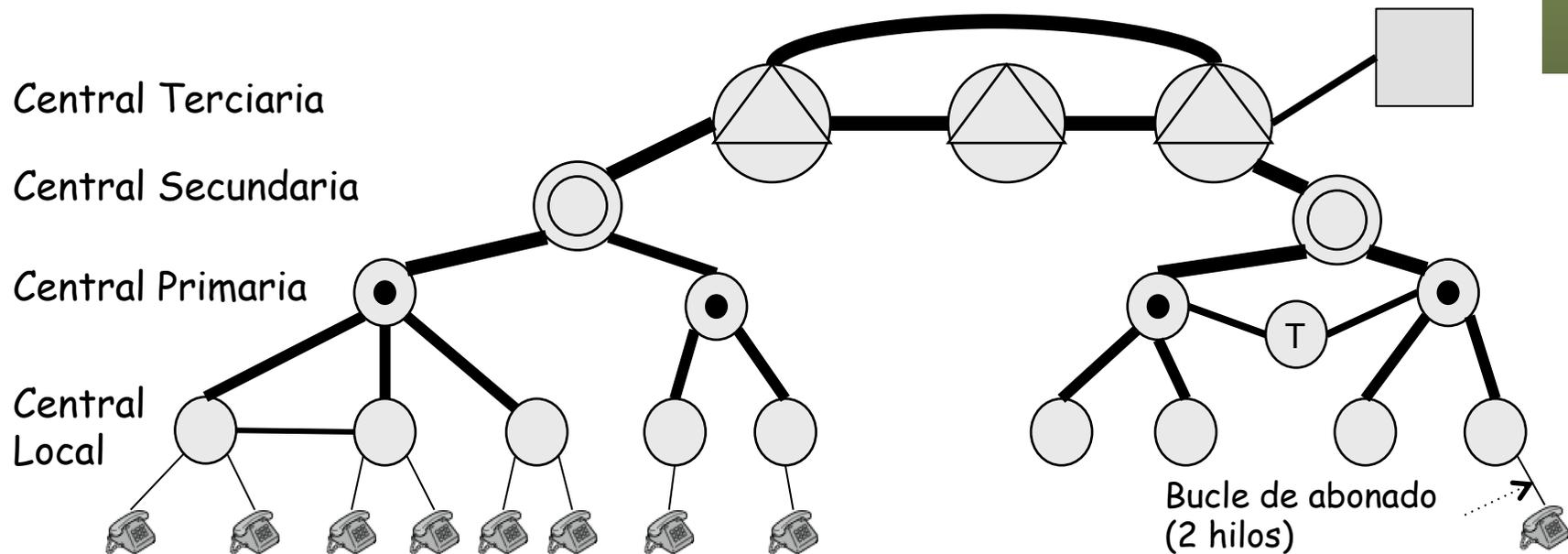
Red jerárquica

Es una red Jerárquica

- **Centrales locales** (“Central terminal”, “Central urbana”):
- **Centrales primarias**
- **Centrales secundarias**
- **Centrales terciarias**
- **Centrales Internacionales**

Red complementaria

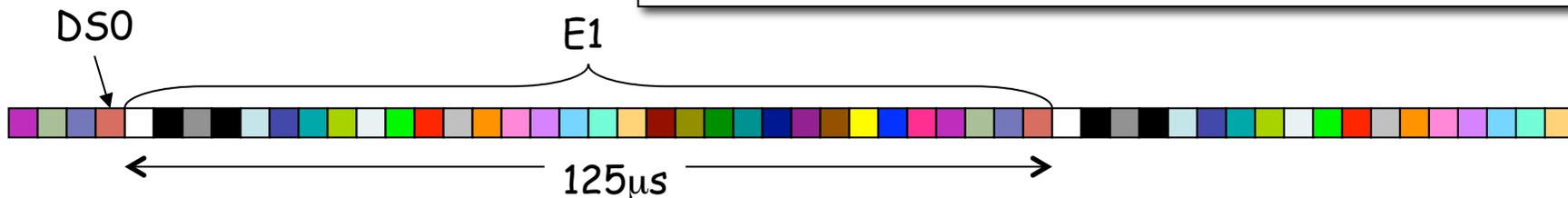
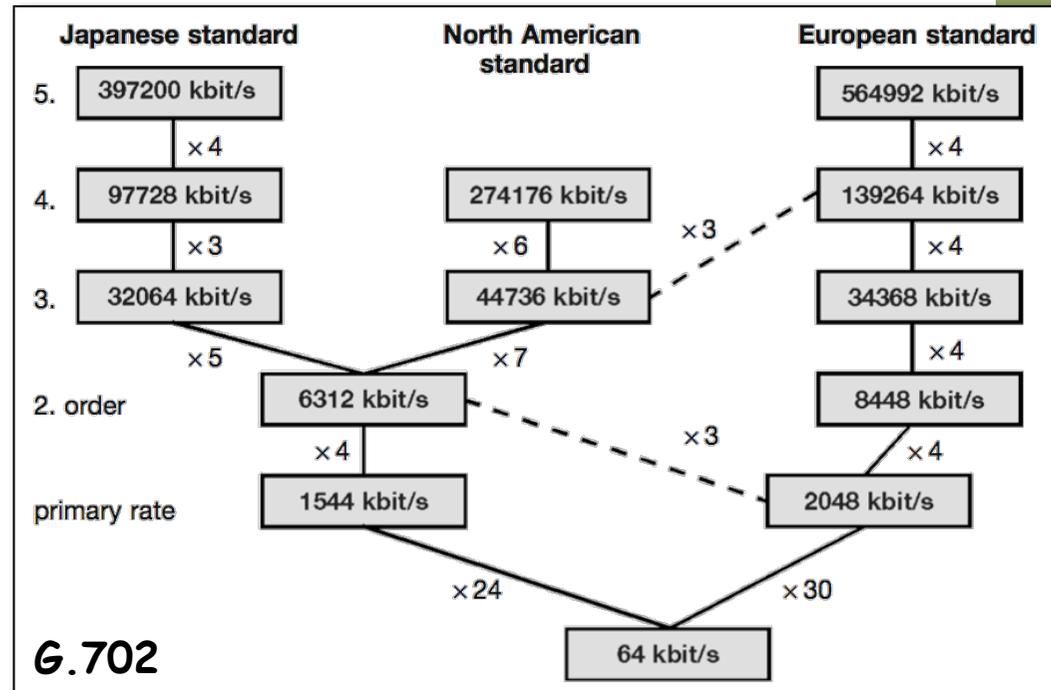
- Secciones directas entre centrales con suficiente tráfico entre ellas
- Centrales Tándem: centrales de tránsito sin abonados



PDH (Plesiochronous Digital Hierarchy)

Multiplexación TDM

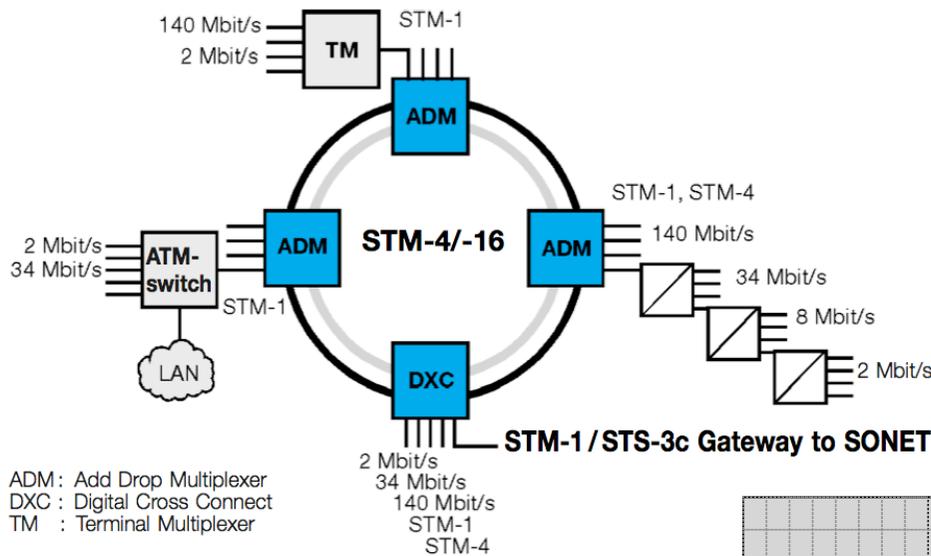
- E1 (2048Kbps) = 32xE0
- E2 = 4xE1, E3 = 4xE2, E4 = 4xE3
- T1 (DS1, 1.54Mbps) = 24xDS0
- T2 = 4xT1, T3 = 7xT2
- G.701-703



SONET/SDH

- SDH se diseñó para transportar señales de 1.5, 2, 6, 34, 45 y 140 Mbps
- Conmutación de circuitos
- Red de transporte, de gestión y de señalización

SDH	OC Level	Line Rate (Mbps)
	OC-1	51.84
STM-1	OC-3	155.52
STM-4	OC-12	622.08
STM-16	OC-48	2488.32
STM-64	OC-192	9953.28
STM-256	OC-768	39813.12

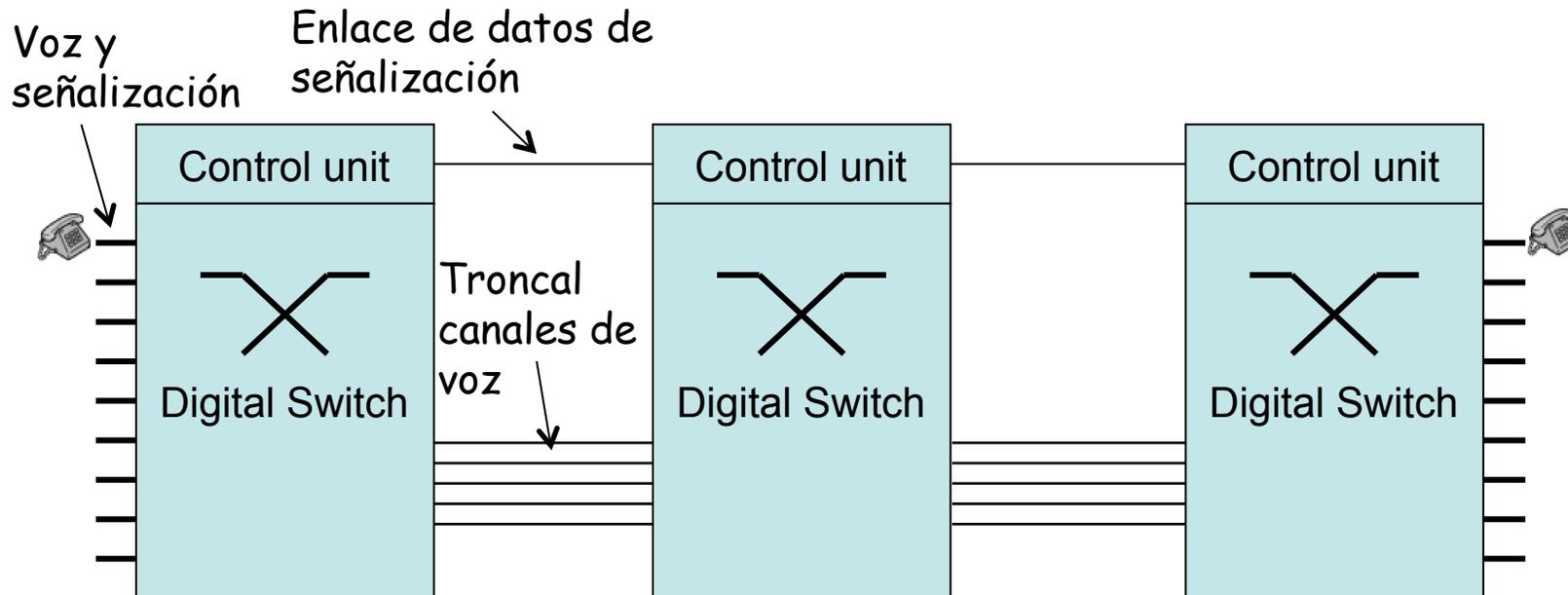


Señalización

Señalización en canal

Señalización por canal común (CCS = Common Channel Signaling)

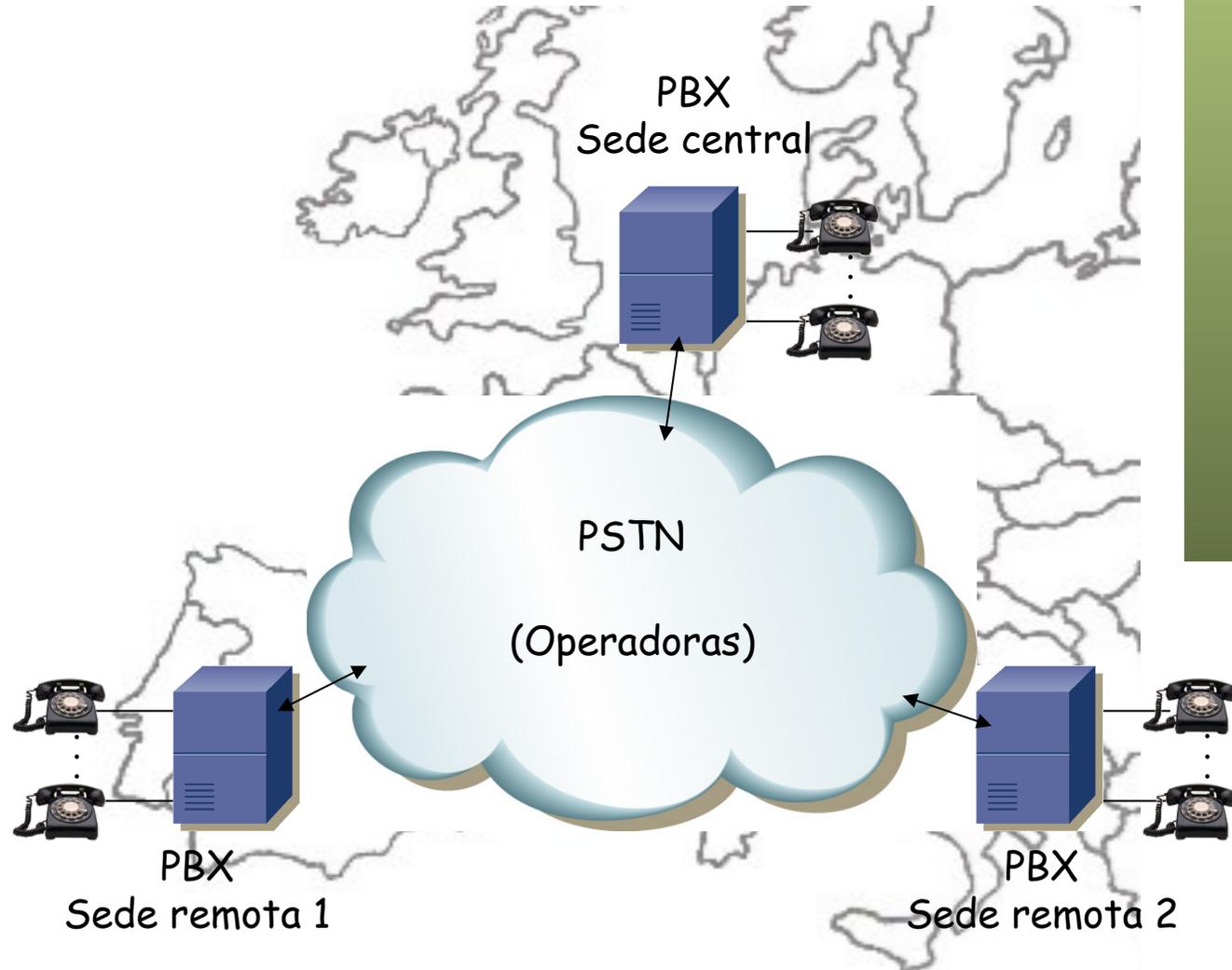
- CCITT Signaling System No. 7 (SS7)
- Forma una WAN que enlaza las lógicas de control de los switches de la PSTN
- En acceso a ella, dos modos:
 - *Fully associated signaling*: un enlace de señalización por cada enlace de voz (ej: canal 16 en E1)
 - *Quasi associated signaling*: señalización por otro enlace que el de tráfico entre los dos switches
- IETF SIGTRAN: SS7 over IP



Voz en escenarios privados

Voz entre sedes

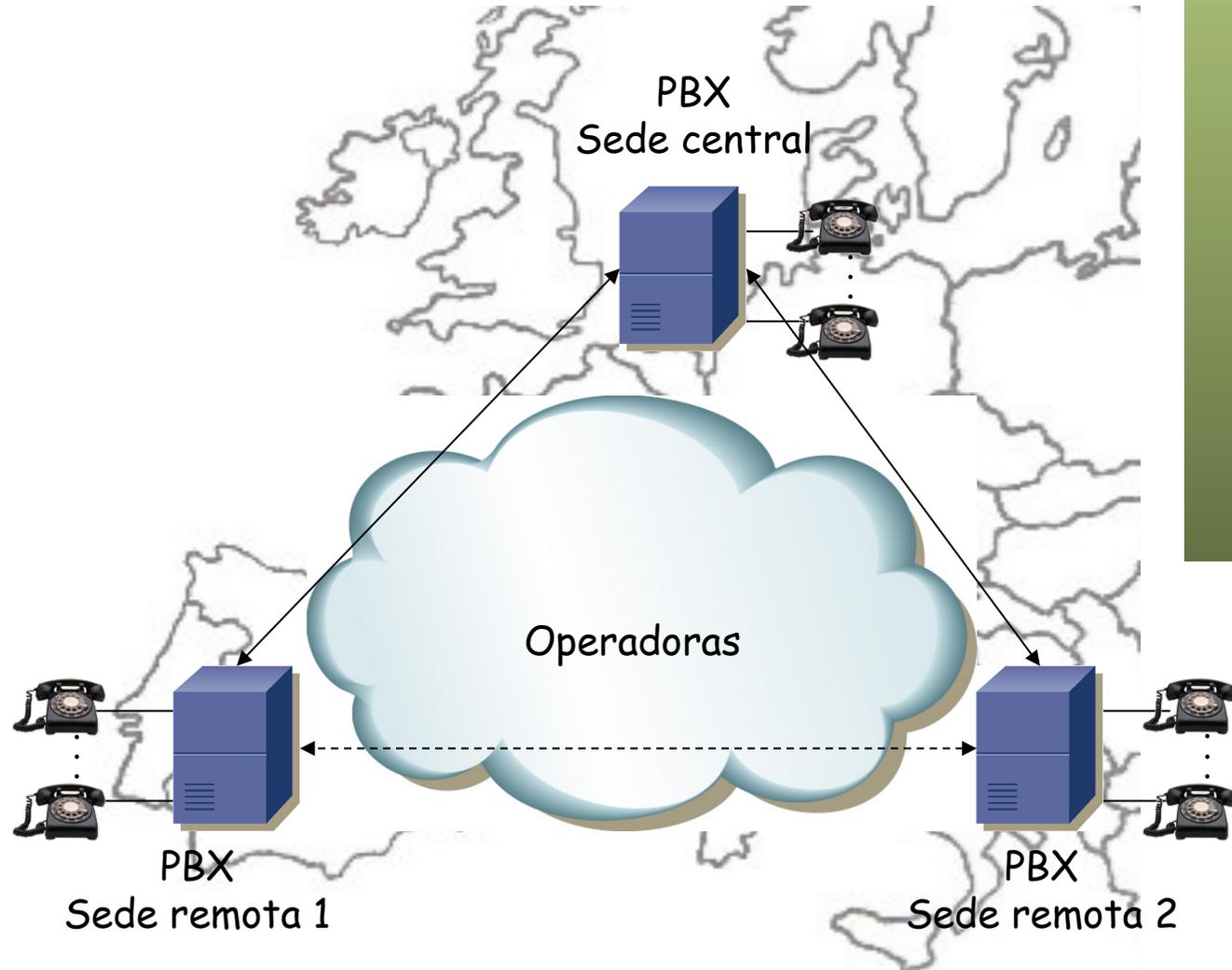
Mediante llamadas por la red pública



PBX = *Private Branch eXchange*

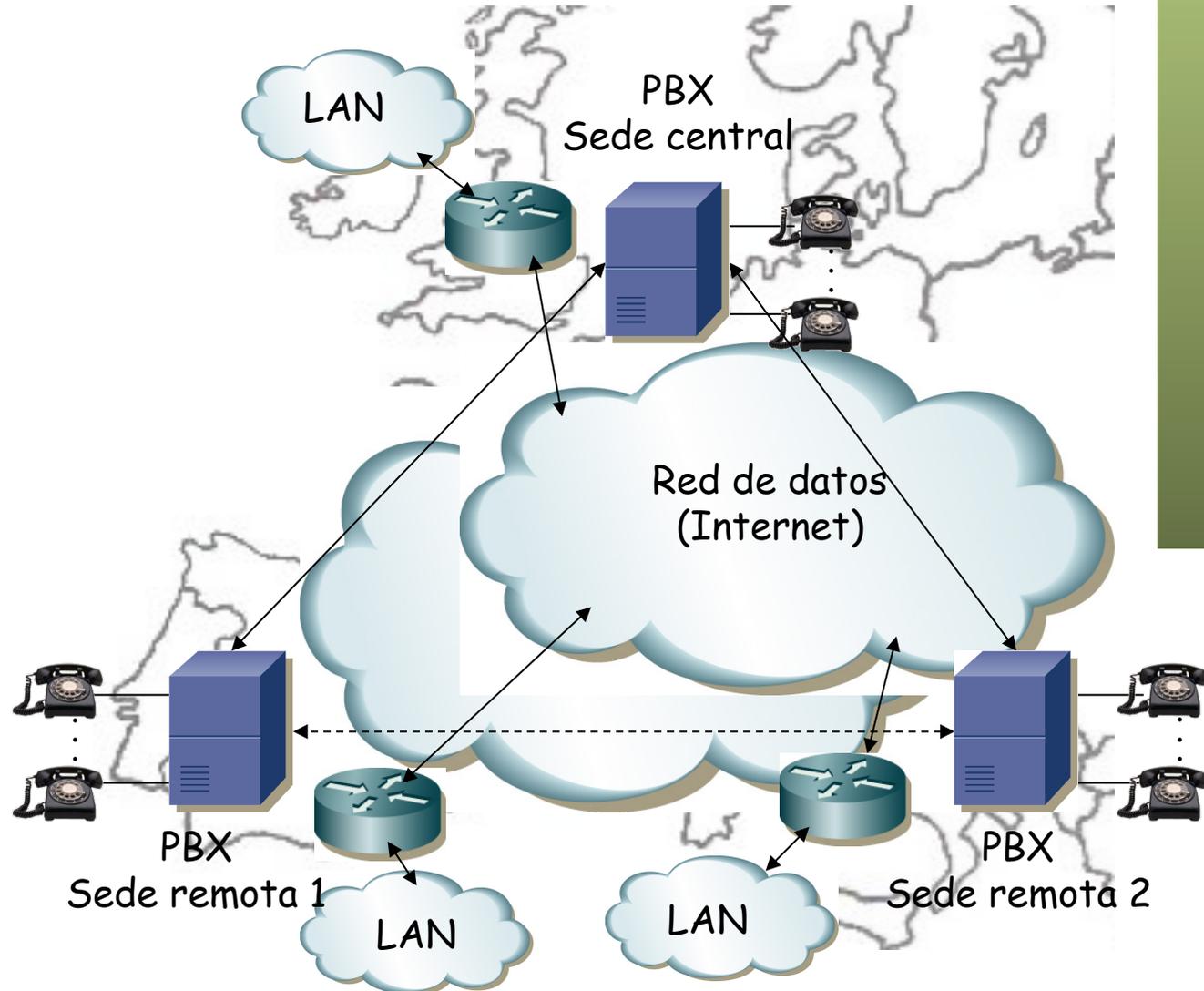
Voz entre sedes

Enlaces dedicados (malla o hub)



Voz + datos

Probablemente tenga enlaces de datos simultáneamente



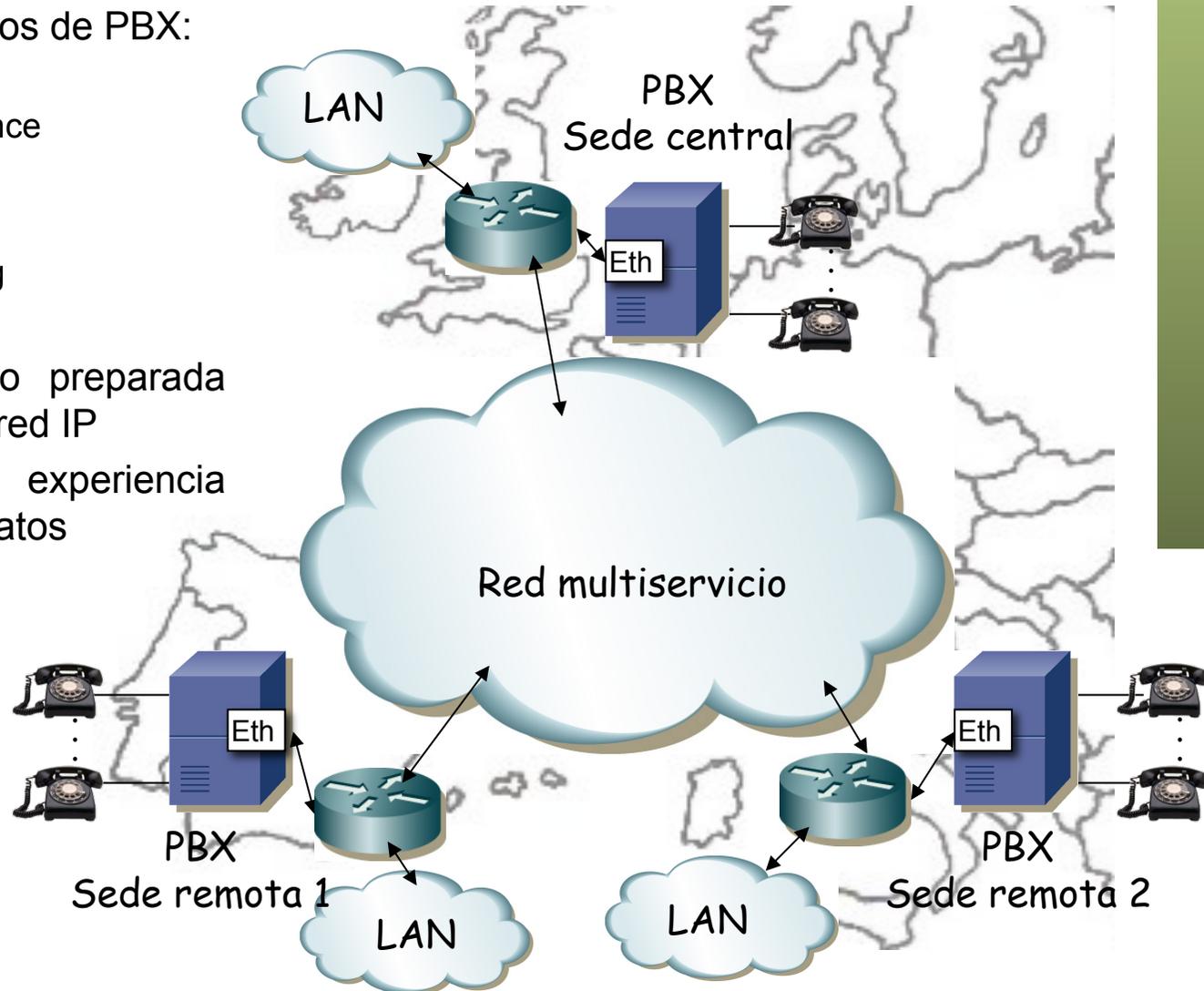
¿Por qué dejar de usar TDM?

- Utilizar la misma infraestructura de datos: reduce CAPEX y OPEX
- Negocio:
 - Añadir más servicios al cliente
 - Telcos añaden datos, ISPs añaden voz
- Aumentar la cantidad de llamadas que se pueden cursar por un enlace
 - *Voice compression*
 - vs los 64 kbps PCM
 - Cuidado, reduce la calidad
 - *Silence supresion*
 - *VAD = Voice Activity Detection*
 - Habla tiene en torno a un 40-50% de actividad frente al tiempo total
 - *Statistical gain*
- Más sencillo incluir nuevos servicios de valor añadido
- Hacer escalabilidad más sencilla
- Simplificar enrutamiento alternativo

Convergencia

PBXs con interfaces IP (Ethernet)

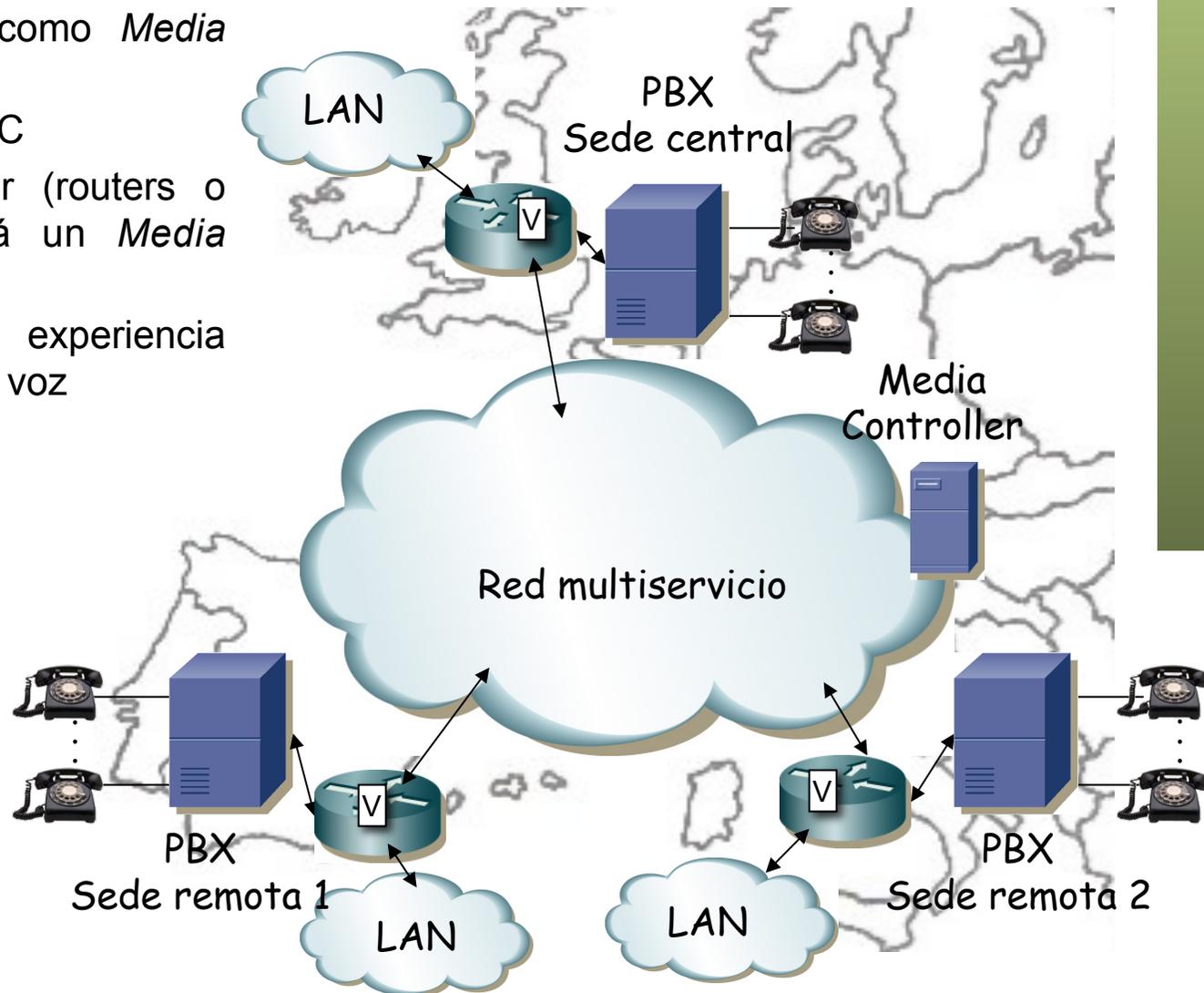
- Reutiliza servicios de PBX:
 - Call transfer
 - Call conference
 - Paging
 - Bridging
 - Group calling
 - Etc.
- Señalización no preparada para delays en red IP
- Fabricante con experiencia en voz, no en datos



Convergencia

PBXs trunk TDM y conversión en router

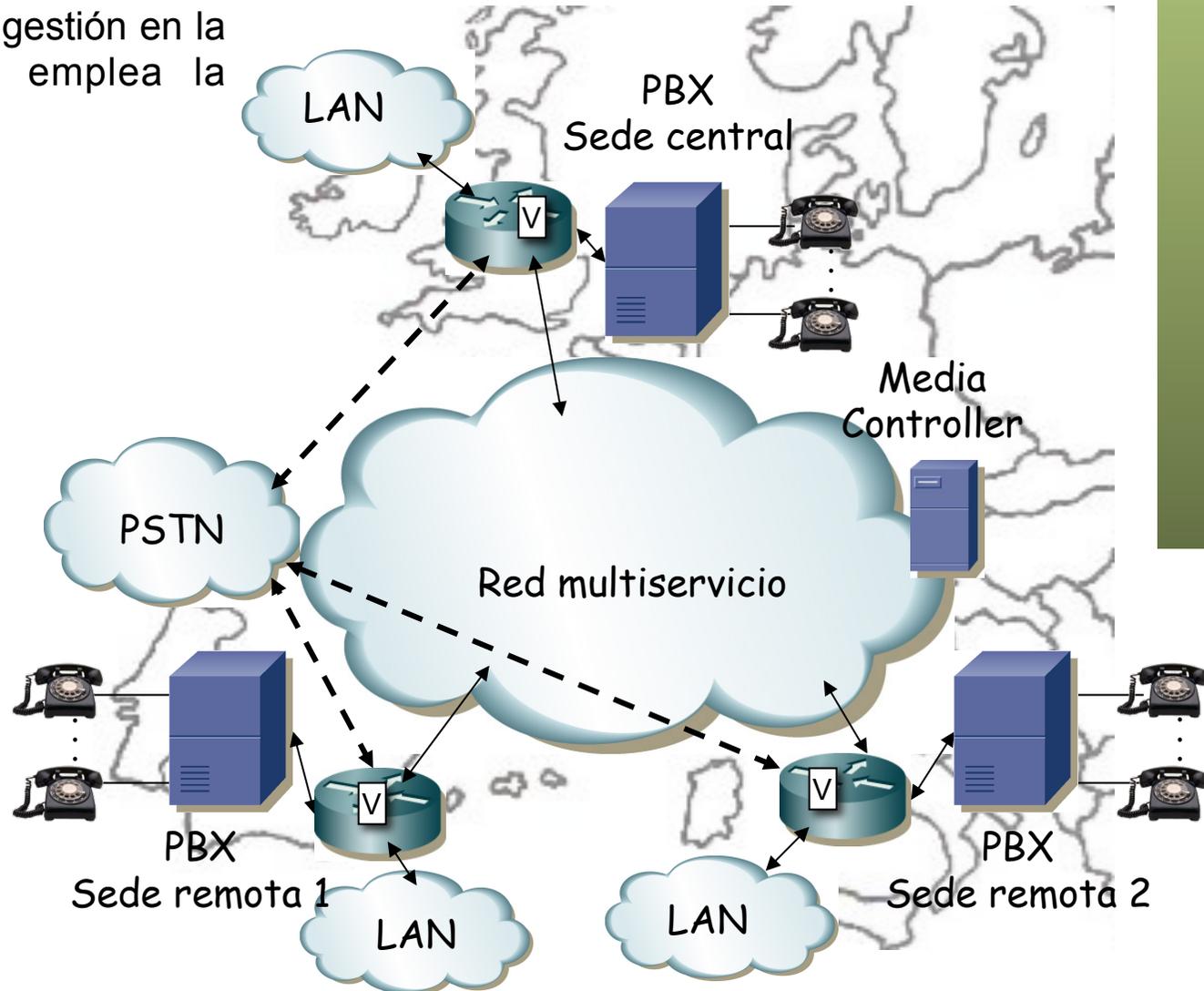
- Router actúa como *Media Gateway*
- Puede ser un PC
- En algún lugar (routers o externo) habrá un *Media Controller*
- Fabricante con experiencia en datos, no en voz



Convergencia

Multi-Point Switched Gateway

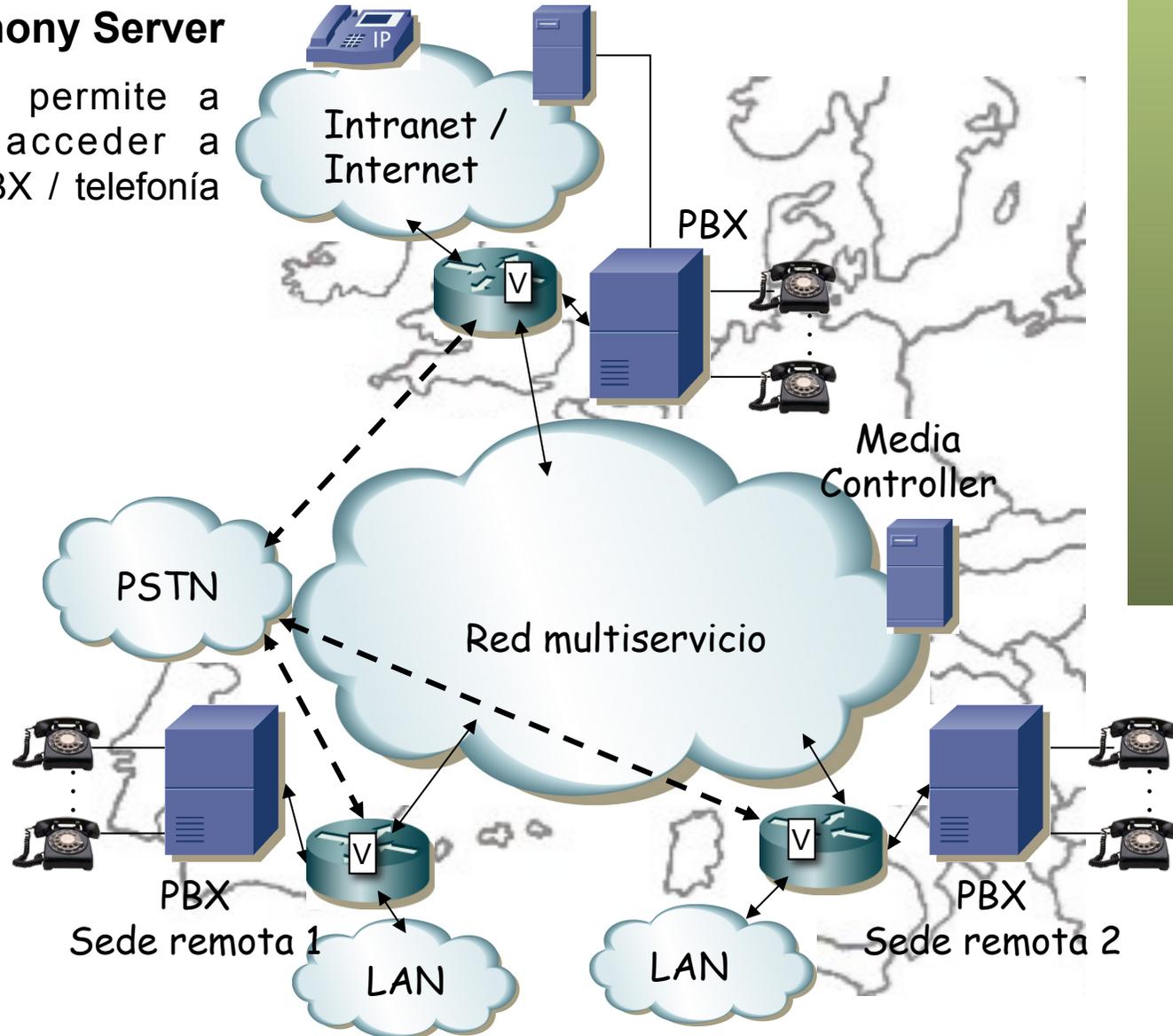
- En caso de congestión en la red de datos emplea la PSTN



Convergencia

(Remote) Telephony Server

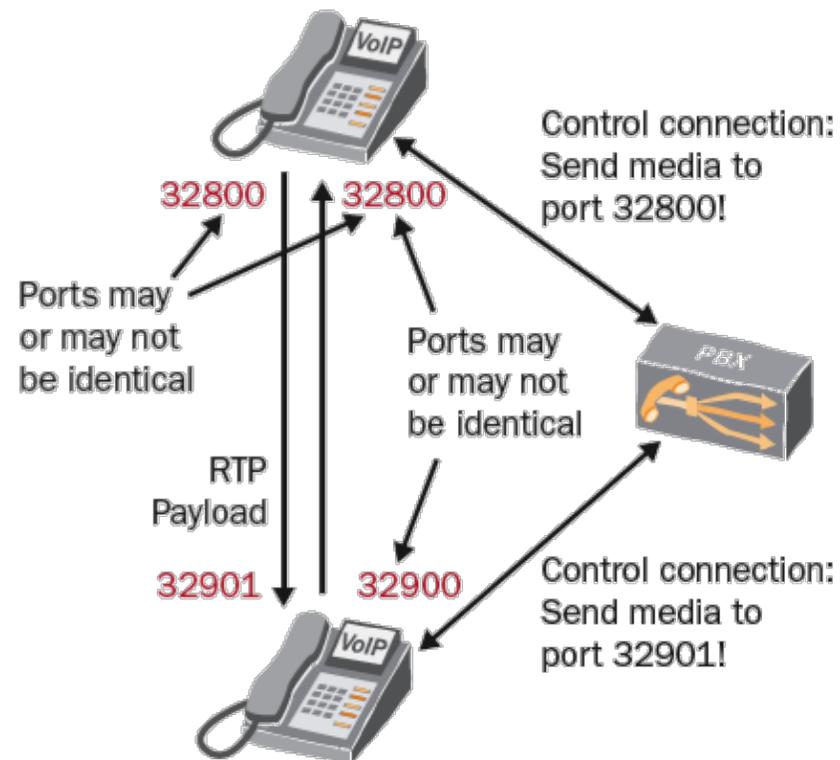
- Gateway que permite a teléfono IP acceder a servicios de PBX / telefonía tradicional



Implementación de VoIP

Esquema básico de flujos en VoIP

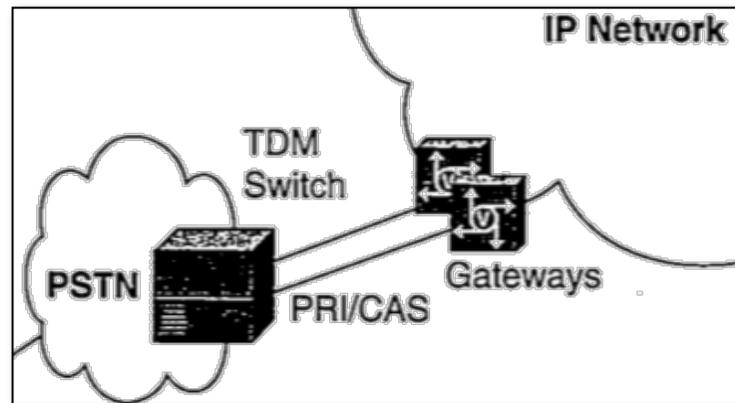
- Dos tipos de flujos
 - Voz, generalmente directa entre los peers (RTP)
 - Señalización, entre peers o con servidores (SIP, H.323, MGCP...)
- Diferentes requisitos de calidad



Terminología

(Media) Gateway

- En cualquiera: H.323, SIP, MGCP, Megaco
- Terminación de llamadas entre un medio y otro
- “Traduce” voz y también la señalización
- Generalmente entre la PSTN y la red de datos
- O puede ser entre dos partes de la red con diferentes requisitos
 - *Transcoding* (cambio de codificador)
 - Diferente señalización (entre SIP y H.323)
- *Residential Gateway, Access Gateway, Business Gateway, Trunking Gateway, Signaling Gateway*



Terminología

Media (Gateway) Controller

- Controla Media Gateways para proveer llamadas extremo a extremo
- Registro de llamadas, autenticación, autorización, encaminamiento, facturación, gestión de recursos ...
- Traducción de direcciones (de nº telef., URL, e-mail, etc a dirección IP)
- Cada MGC controla una *zona*
- Media Gateway Controller en Megaco/H.248.1
- A.k.a. Call Agent en MGCP, a.k.a Gatekeeper en H.323
- Media Server, Telephony Server, Call Manager, Virtual Switch, Softswitch...

Multipoint Control Unit (MCU)

- Para llamadas/videoconf con más de 2 terminales/gateways participantes (a.k.a. bridge)

VoIP, QoS y QoE

QoE

- *Quality of Experience*
- Intenta medir la percepción que tiene el usuario
- Para VoIP o vídeo depende de la calidad del codificador, del servicio ofrecido por la red y de la calidad del decodificador
- Pueden ser métricas objetivas o subjetivas
- Métricas subjetivas para voz:
 - Basadas en la opinión de usuarios
 - Mean Opinion Score (MOS) da una medida numérica 1-5
 - MOS definido en ITU-T P.800
 - POTS tiene un MOS de 4.3 y la telefonía móvil entre 2.9 y 4.1
- Métricas objetivas para voz:
 - Perceptual Evaluation of Speech Quality (PESQ)
 - ITU-T P.862
 - Requiere la señal original y la recibida para predecir el MOS
 - *E model*
 - ITU-T G.107 (factor R)
 - Tiene en cuenta el bitrate, pérdidas, ruido, eco, etc

Retardo end-to-end

- ITU-T G.114 “One-way transmission time”
- Por debajo de 150 ms la mayoría de las aplicaciones experimentan interactividad transparente
- La calidad de servicio público exige un máximo de 150 ms
- Retardos superiores a 400 ms son inaceptables
- Enlace por satélite geoestacionario añade unos 260 ms
- En entornos privados es razonable un límite de 200-250 ms

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

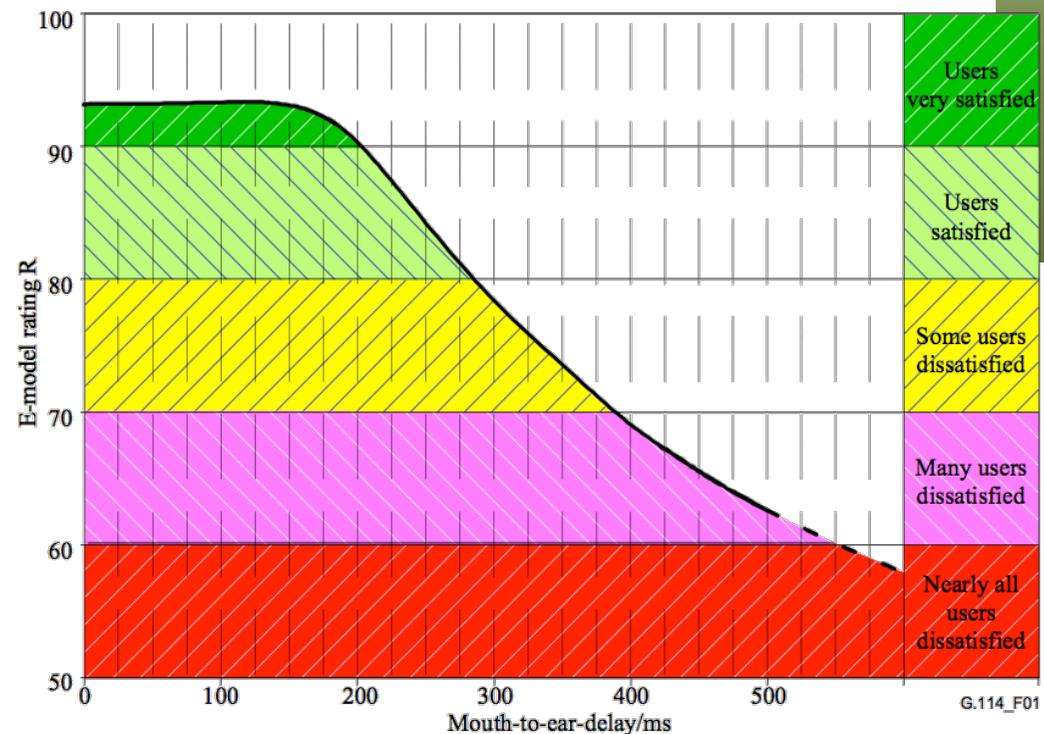


Figure 1/G.114 – Determination of the effects of absolute delay by the E-model

Flujo extremo a extremo

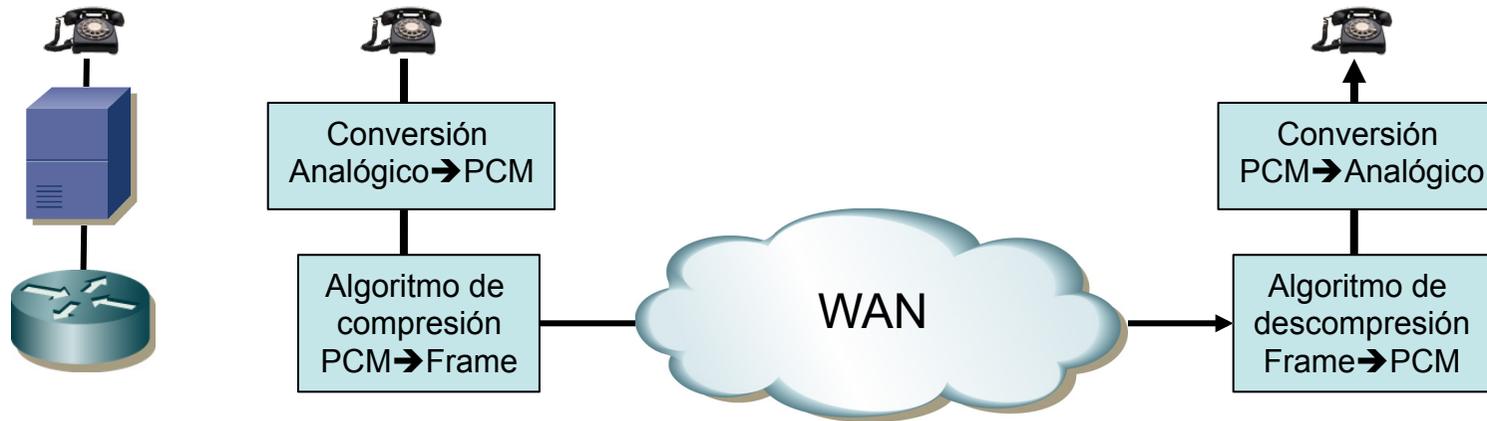
- Flujo a través de una WAN



- (...)

Flujo extremo a extremo

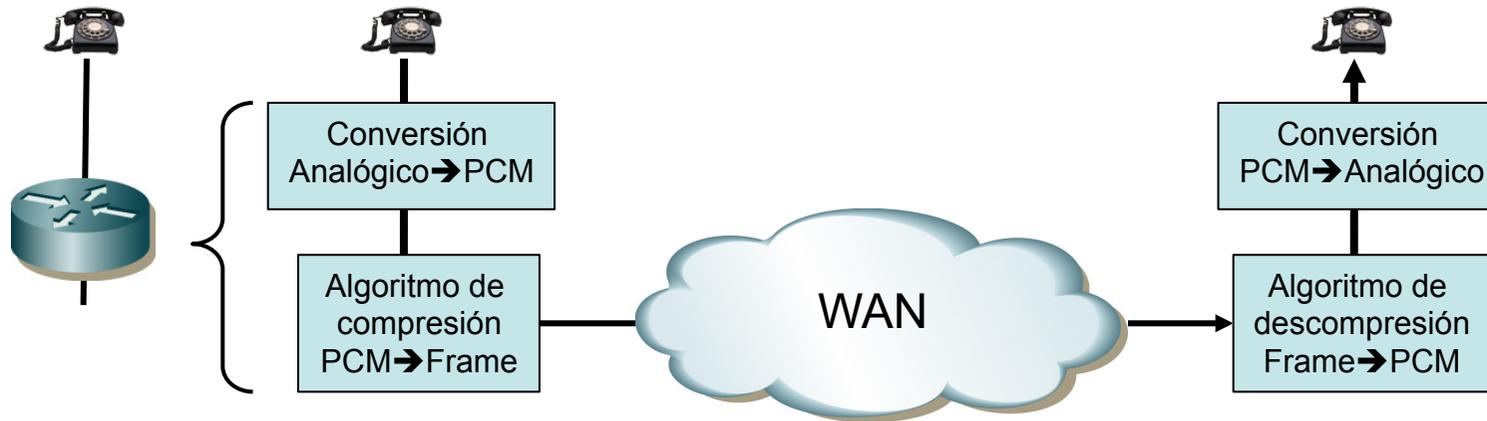
- Flujo a través de una WAN



- Una PBX digital puede ser quien hace la digitalización
- Un router entonces suele cubrir la función de compresión y paquetización
- Podría la PBX integrar la funcionalidad del router
- (...)

Flujo extremo a extremo

- Flujo a través de una WAN



- O puede el router implementar ambas posibilidades

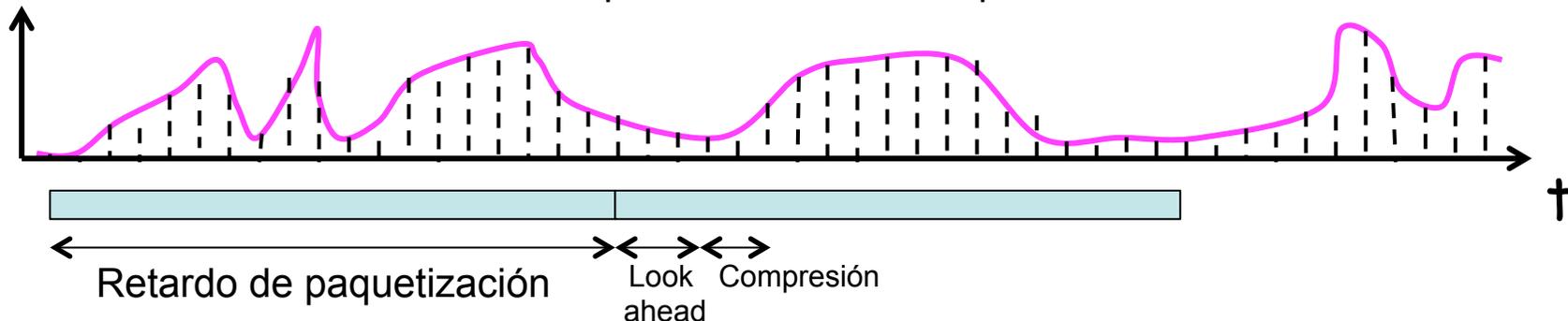
Codecs

- El codec es crítico en el efecto que el SLA tenga sobre la calidad
- Los codecs varían en complejidad, en BW requerido y en la calidad de sonido ofrecida
- G.711 basado en PCM (Pulse Code Modulation), simplemente muestreando a intervalos regulares
- G.726, ADPCM (Adaptive Differential PCM), usa predicción de la siguiente muestra y cuantiza eso
- G.723 y G.729 trabajan con bloques de muestras (*frame-based*) para las que aplican técnicas de compresión (ACELP = Algebraic Code Excited Linear Prediction)
- Los más complejos ofrecen mayor calidad percibida y menor bitrate pero con mayores tiempos de procesado

Captura	Programación	Configuración
Formato audio:		
Codec audio:	MP2	
Bitrate	128 kbps	
Frec. de muestreo:	44100 Hz	
Canales audio:	Stereo	
Parámetros del dispositivo:		
Entrada vídeo:	Ninguno	
Estándar vídeo:	Ninguno	

Componentes del retardo

- **Retardo de paquetización**
 - El tiempo necesario para acumular las muestras
 - Depende del tamaño en muestras que se busque
 - En general no se excede de 30 ms de muestras (240 bytes PCM)
 - A menor tamaño mayor tasa de llegadas de paquetes de voz
- **Retardo de procesamiento del codificador**
 - Tiempo que lleva al DSP comprimir las muestras
 - También afecta en descompresión (menor, se suelen juntar al calcular)
 - Depende del coder y de su implementación (fabricante, en torno a 2-20ms)
- **Algorithmic Delay (look ahead)**
 - Los algoritmos de compresión suelen necesitar conocer muestras siguientes a las del bloque a comprimir
 - Eso implica que hay que esperar a que se generen
 - Para G.726 es de 0 ms, para G.729 de 5 ms, para G.723.1 de 7.5 ms



Retardos constantes

Retardo de serialización (tiempo de transmisión)

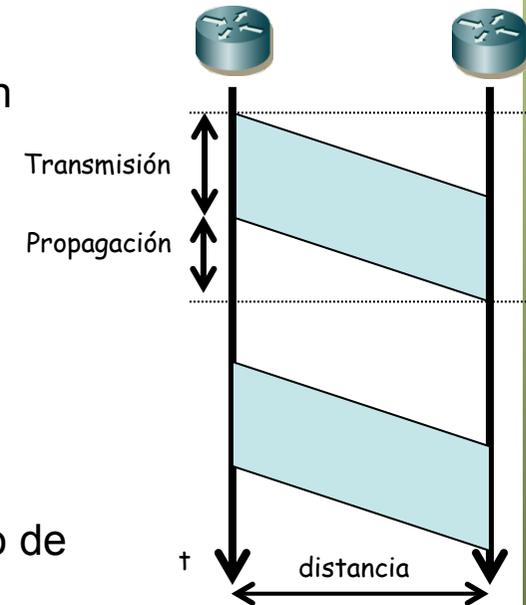
- Solo se mejora aumentando la velocidad de transmisión
- Despreciable por encima de 100Mbps

Retardo de propagación

- Coaxial terrestre, radio: $4\mu\text{s}$ cada Km (250.000Km/s)
- Fibra: $5\mu\text{s}$ cada Km (200.000Km/s)
- Coaxial submarino: $6\mu\text{s/Km}$
- Siendo D la distancia en línea recta geográfica (a vuelo de pájaro)
- Los enlaces no siguen una línea recta
- ITU-T G.826 hace una estimación (R)

Tiempo de procesamiento/conmutación

- En función del hardware
- Típicamente $10\text{-}20\mu\text{s}$
- Router software $2\text{-}3\text{ms}$

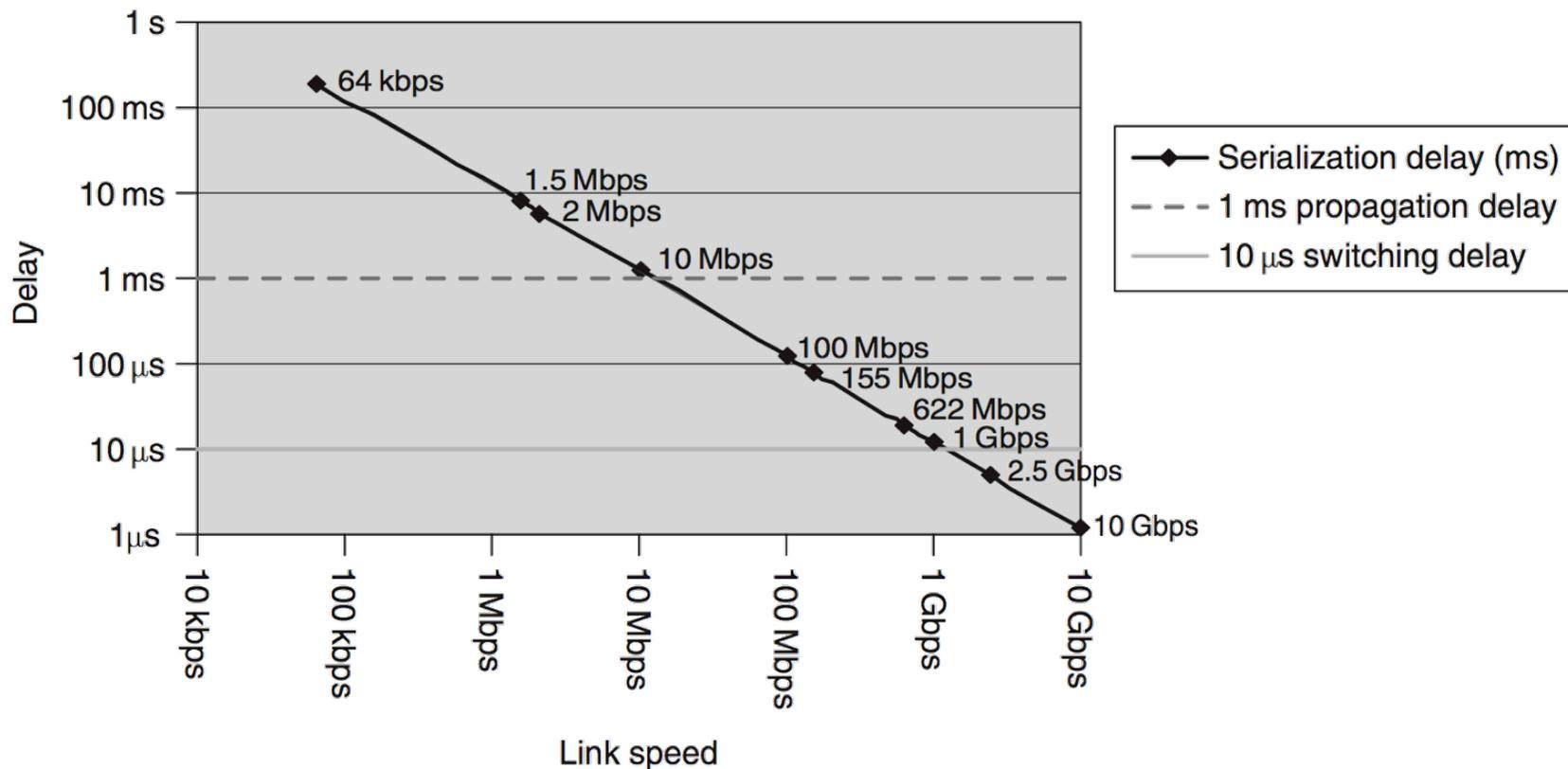


D	R
$D < 1000 \text{ km}$	$1.5 \times D$
$1000 \text{ km} \leq D \leq 1200 \text{ km}$	1500 km
$D > 1200 \text{ km}$	$1.25 \times D$

G.826 "End-to-end error performance parameters and objectives for international, constant bit-rate digital paths and connections"
 G.114 "One-way transmission time"

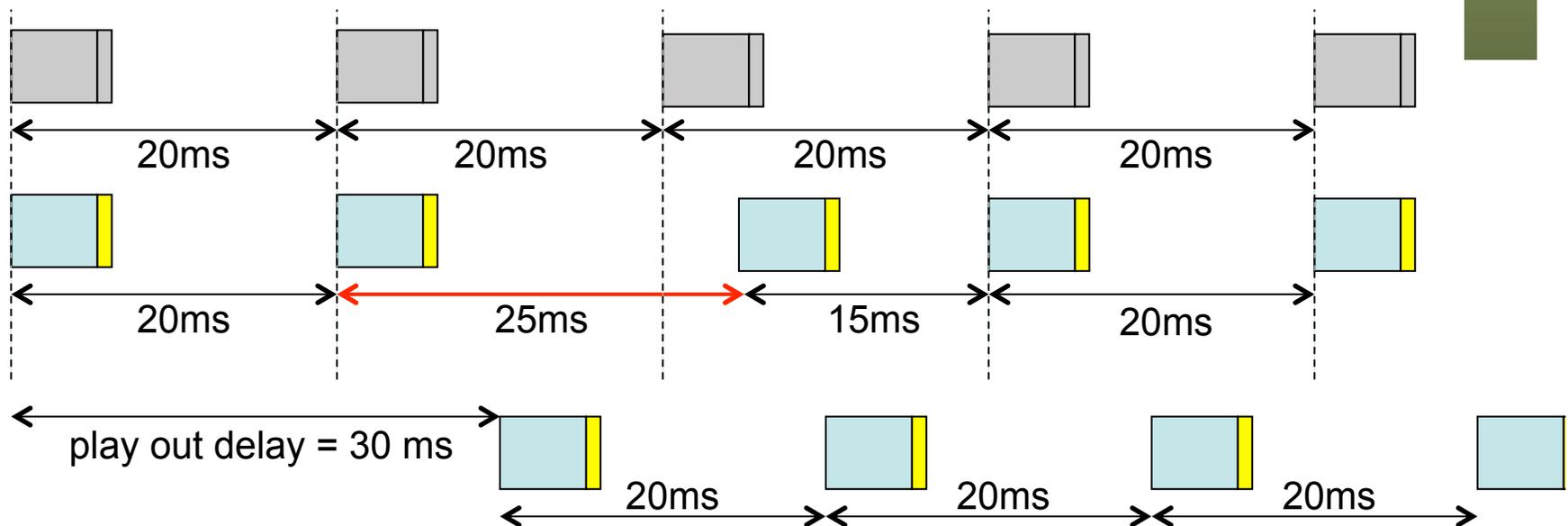
Retardos constantes

- Ejemplo comparativo
 - Retardo de serialización, de propagación y de conmutación
 - Paquete de 1500 bytes
 - Unos 200Km de fibra : 1ms de propagación



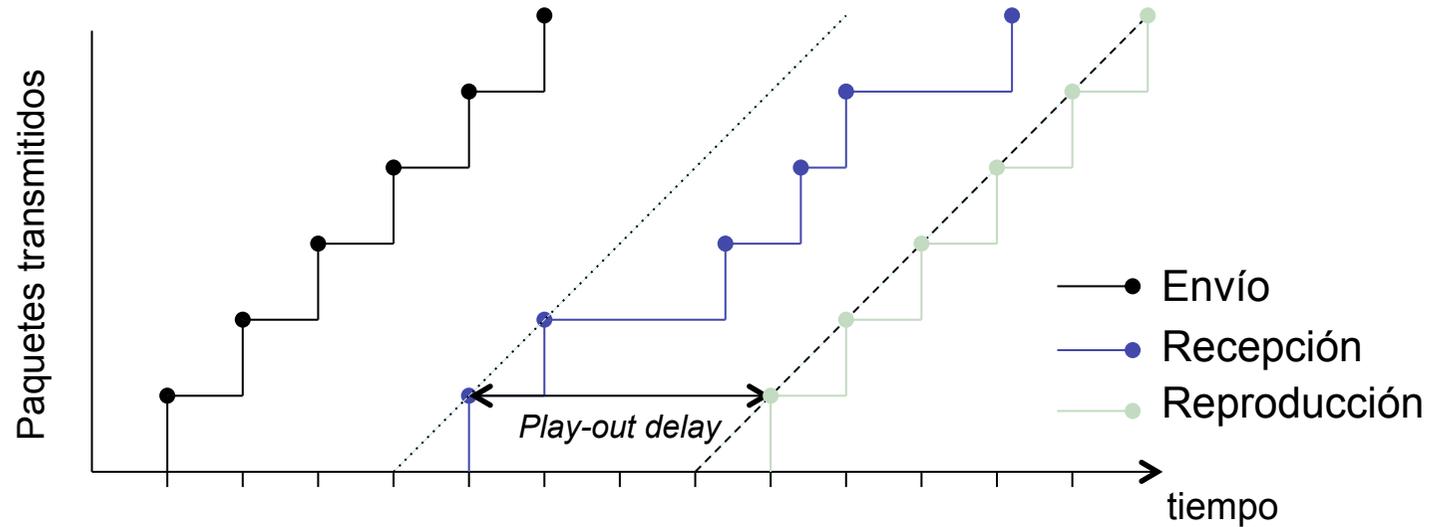
De-jitter delay

- VoIP es sensible ante el Jitter (a.k.a. Delay-Jitter, a.k.a. Packet Delay Variation)
- Se emplean *de-jitter buffers* (*play-out buffers*)
- Introducen un *play out delay*
- Se convierte la variación en el retardo en un retardo constante en el receptor
- Si el retardo añadido es demasiado grande obliga a la red a tener bajos retardos
- Si el retardo añadido es demasiado pequeño obliga a la red a garantizar un jitter bajo (o podrá haber *buffer underflows*)
- Buffer físico en torno al doble de lo necesario para el *play out delay*

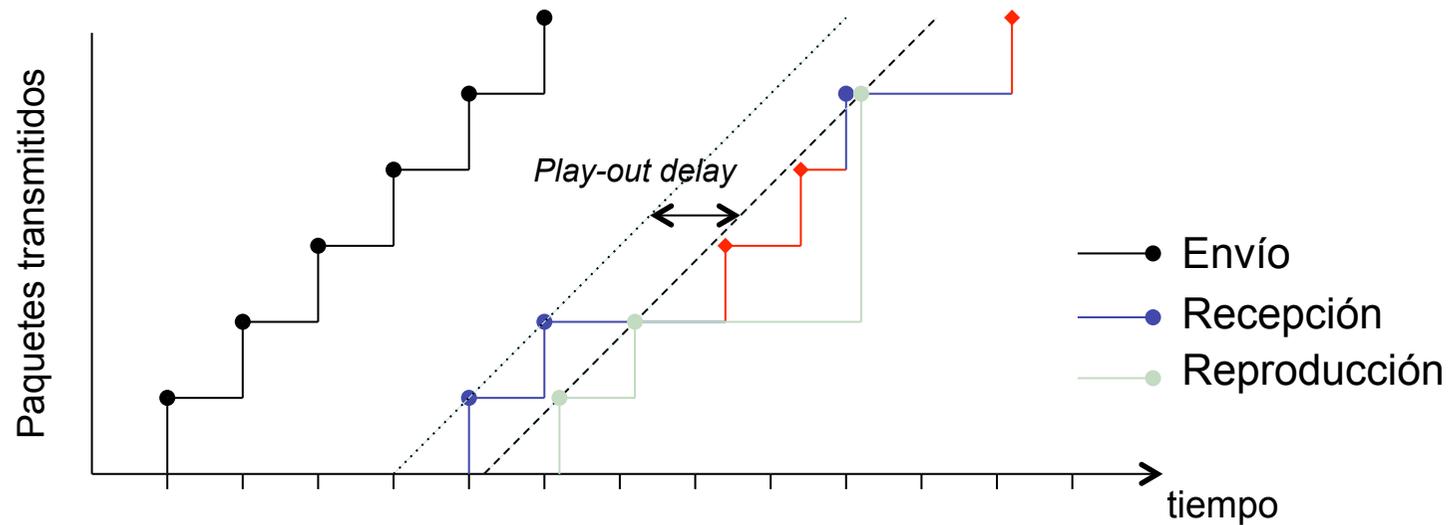


De-jitter delay

- Ejemplo de buffer muy grande:

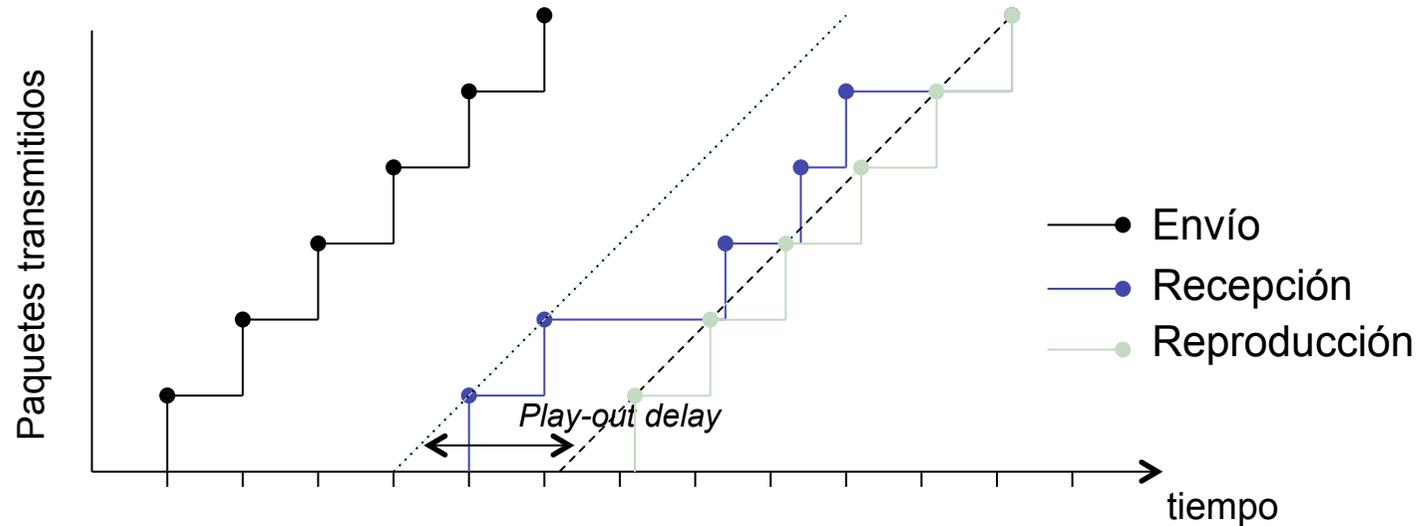


- Ejemplo de buffer demasiado pequeño:



De-jitter delay

- Ejemplo de buffer óptimo:

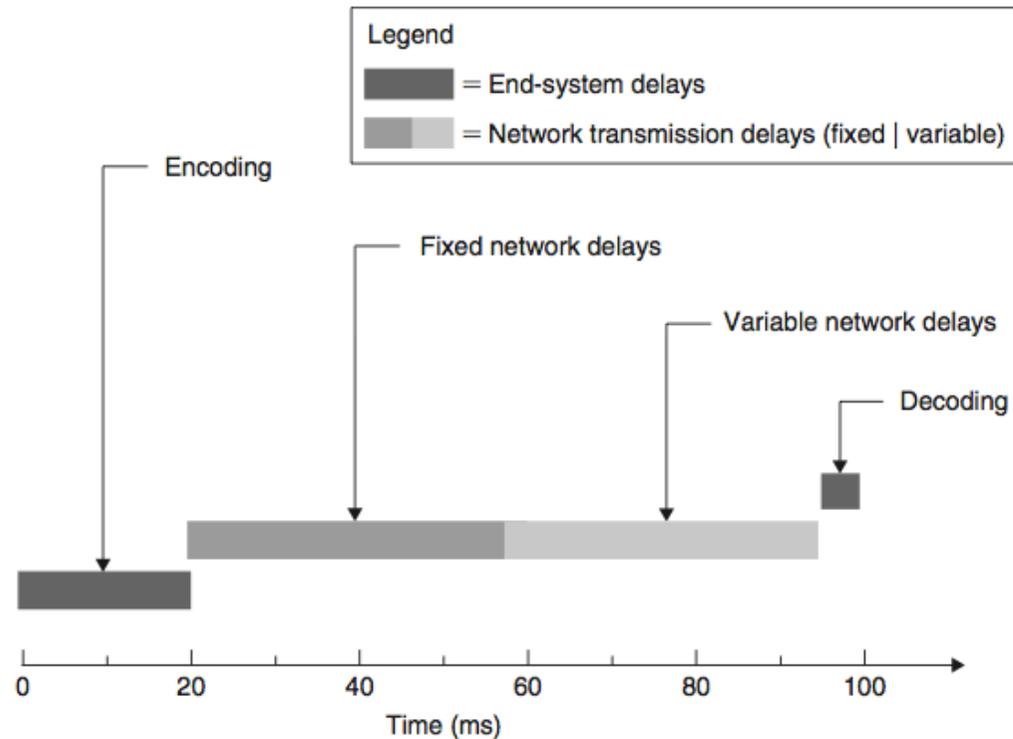


- El valor óptimo es el del retardo variable extremo a extremo
- Si se vacía el buffer entonces la próxima llegada se retiene durante el *play out delay* para reiniciar buffer
- Valores de *play out delay* en torno a los 40 ms
- Se emplea buffer adaptativo (con un valor máximo)
- Ajusta el tamaño del buffer dinámicamente
- Si se produce un underflow lo intenta “cubrir” (*packet loss concealment*) y aumentan el *play-out delay*)
- Si pueden reducir el buffer lo hace de forma lenta

Retardos variables

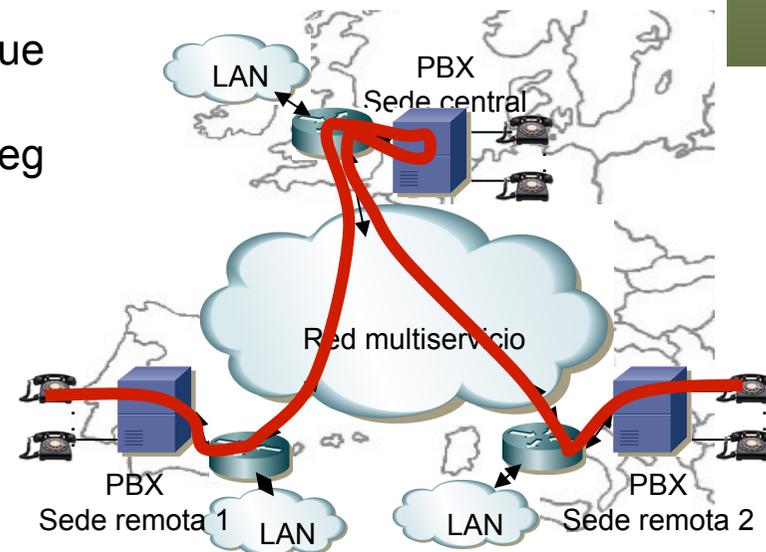
Retardo en cola (scheduling delay)

- Tiempo entre que el paquete se dirige al interfaz de salida y que empieza a transmitirse
- Depende de la carga
- Depende del planificador



Otros retardos

- La llamada puede pasar por recodificadores, con lo que implica un nuevo tiempo de codificación
 - Sedes remotas podrían conmutar llamadas en sede central
 - Conmutación en PBX digital
 - Requiere convertir a PCM para pasar a la PBX (de-jitter, decodec)
 - Requiere volver a comprimir para enviar a la segunda sede
 - Mayores retardos de procesado
 - Más de dos compresiones CS-ACELP degrada la calidad de voz
- PDD = Post Dial Delay
 - Entre marcar el último número y que suene el otro teléfono
 - 1-2 seg para llamadas nacionales, 4+ seg para internacionales



Resumen

- Voz sobre red de datos debe transportar audio y señalización
- Debe coexistir con voz sobre PSTN o redes privadas de circuitos
- Diferentes arquitecturas de servidores y agentes según protocolos para VoIP
- El retardo máximo one-way debería ser de 150ms pero en entornos corporativos se puede exceder
- Las comunicaciones por satélite imponen un retardo alto imposible de reducir
- Retardos constantes: paquetización, procesado del codificador, look ahead, serialización, propagación, conmutación
- Retardos variables: retardo en cola
- El retardo variable implica mayor retardo pues se compensa con buffers
- La señalización también añade un retardo al servicio, no a la comunicación