

# QoS: Transporte de Voz

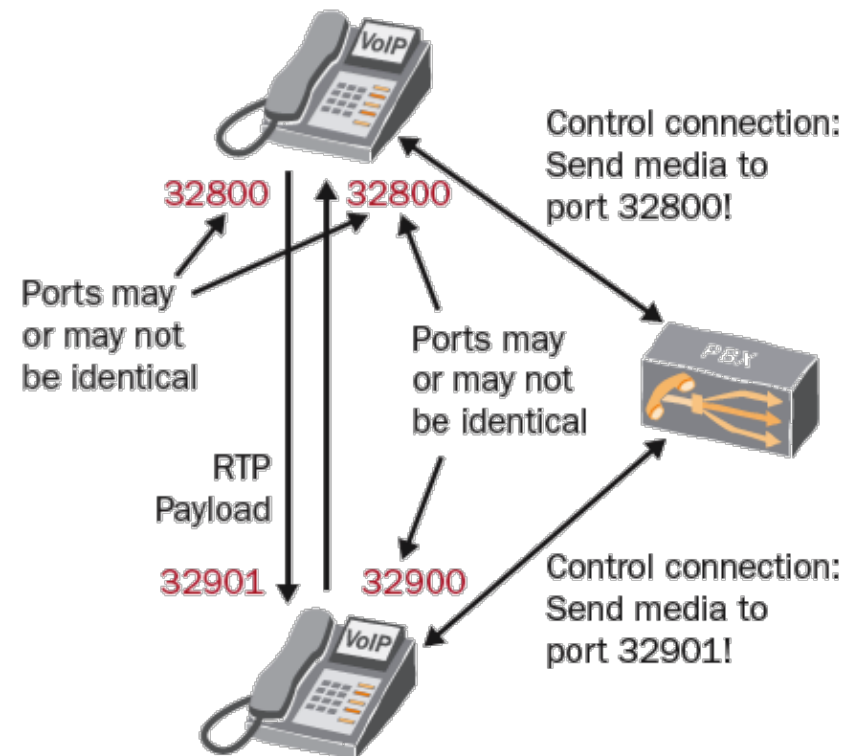
Area de Ingeniería Telemática  
<http://www.tlm.unavarra.es>

Grado en Ingeniería en Tecnologías de  
Telecomunicación, 3º

# Implementación de VoIP

# Esquema básico de flujos en VoIP

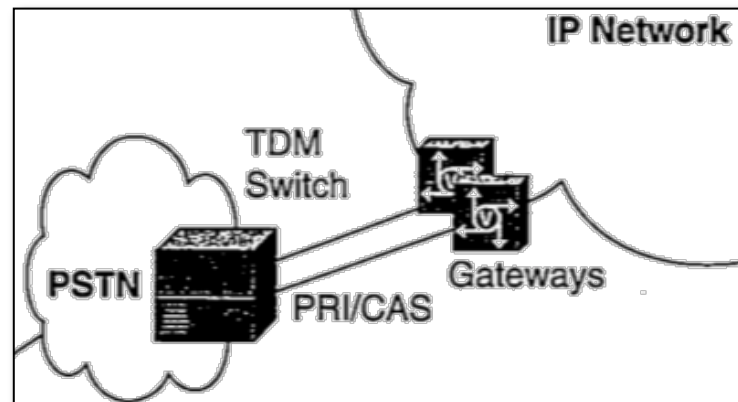
- Dos tipos de flujos
  - Voz, generalmente directa entre los peers (RTP)
  - Señalización, entre peers o con servidores (SIP, H.323, MGCP...)
- Diferentes requisitos de calidad



# Terminología PSTN-VoIP

## ***(Media) Gateway***

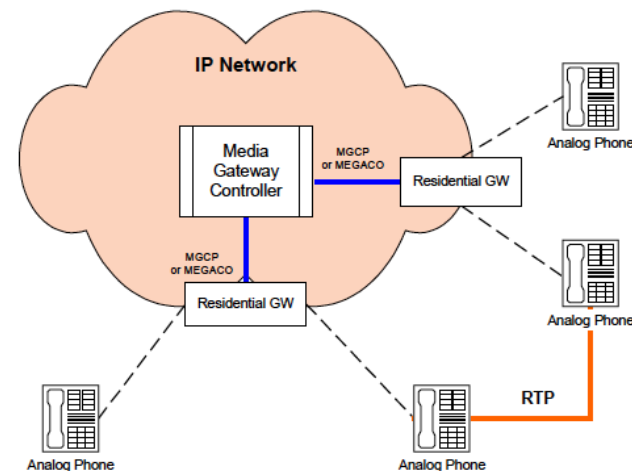
- En cualquiera: H.323, SIP, MGCP, Megaco
- Terminación de llamadas entre un medio y otro
- “Traduce” voz y también la señalización
- Generalmente entre la PSTN y la red de datos
- O puede ser entre dos partes de la red con diferentes requisitos
  - *Transcoding* (cambio de codificador)
  - Diferente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



# Terminología PSTN-VoIP

## Media (Gateway) Controller

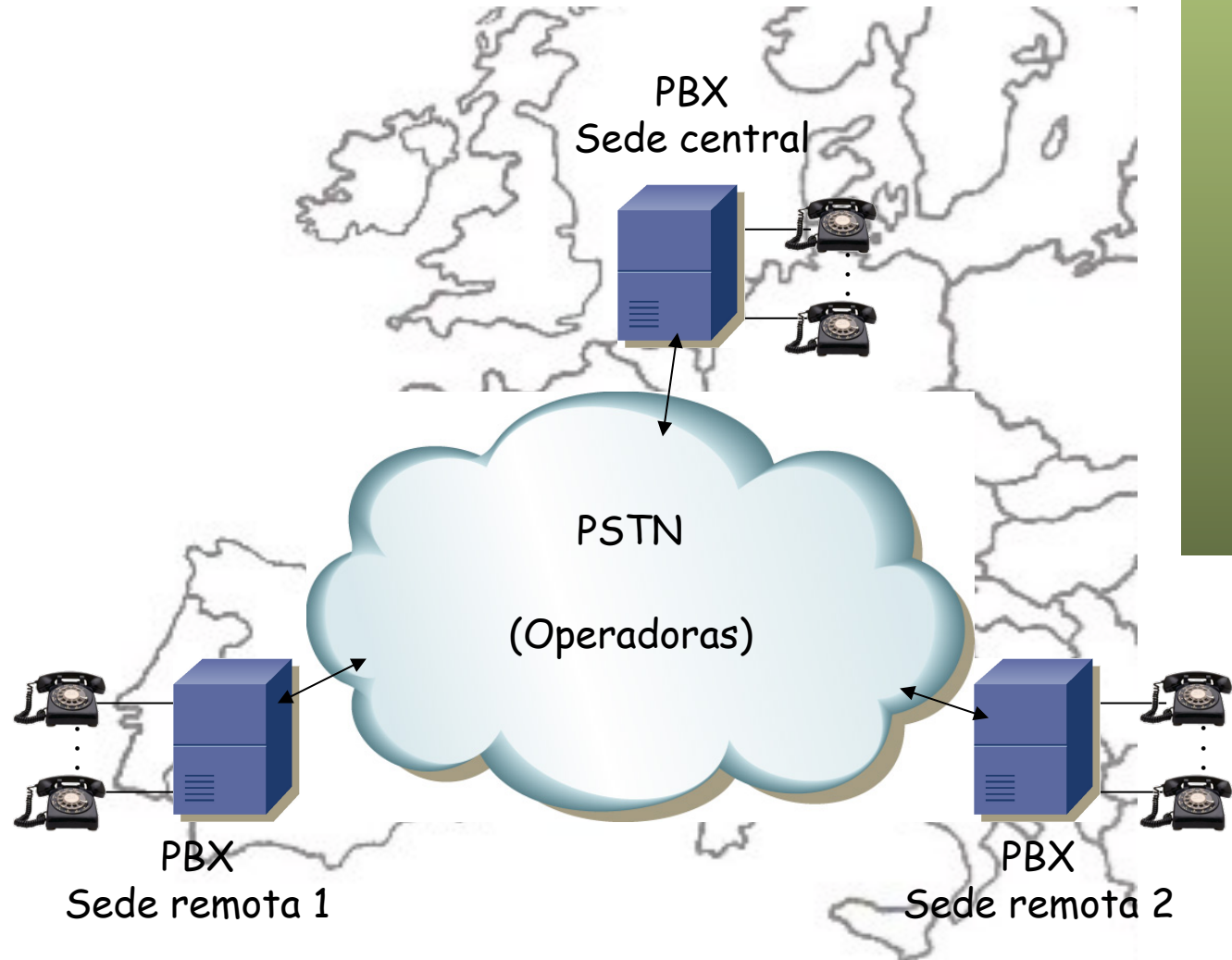
- Controla Media Gateways para proveer llamadas extremo a extremo
- Registro de llamadas, autenticación, autorización, encaminamiento, facturación, gestión de recursos ...
- Traducción de direcciones (de nº telef., URL, e-mail, etc a dirección IP)
- Cada MGC controla una *zona*
- Media Gateway Controller 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...



# Voz en escenarios privados

# Voz entre sedes

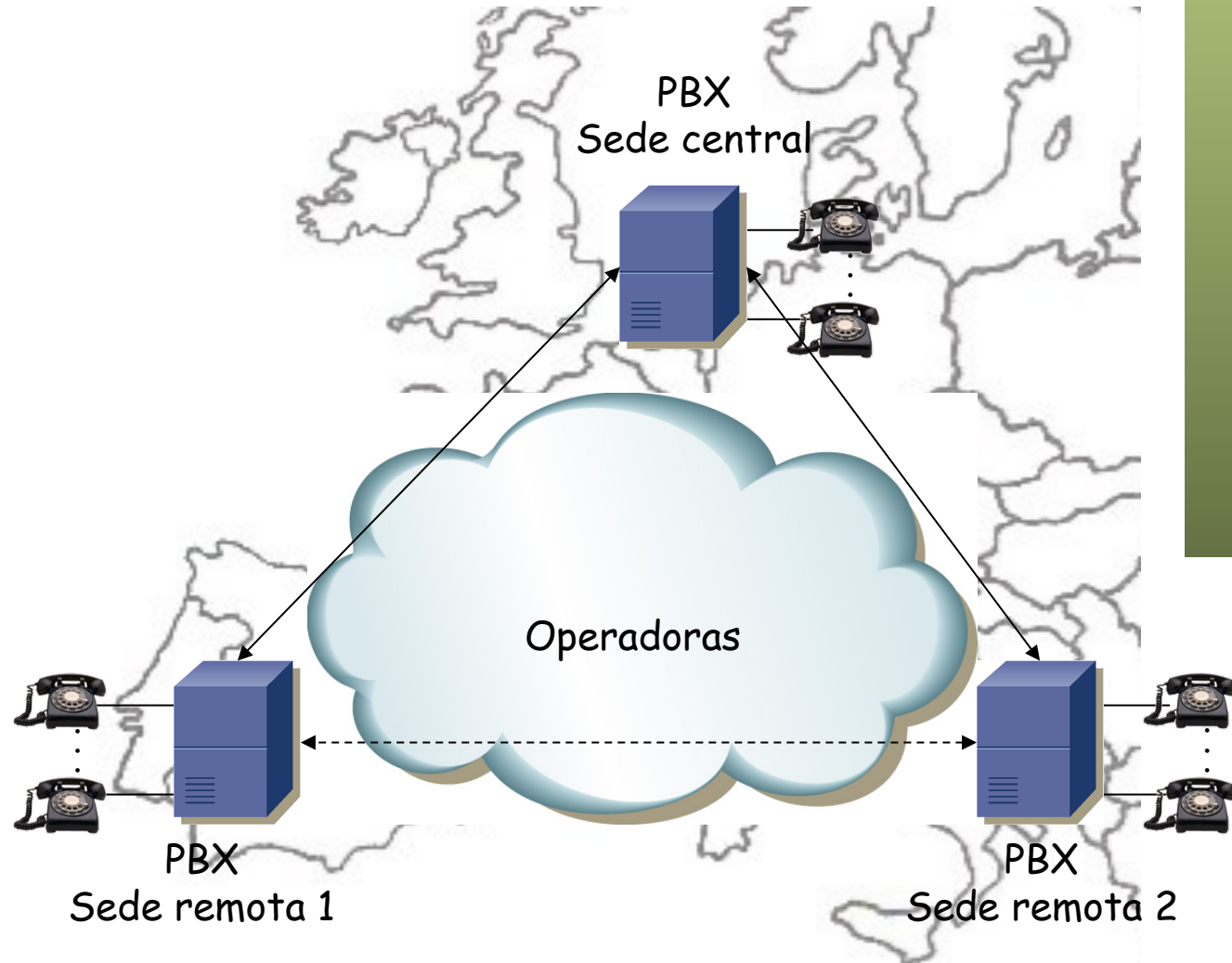
Mediante llamadas por la red pública



PBX = *Private Branch eXchange*

# Voz entre sedes

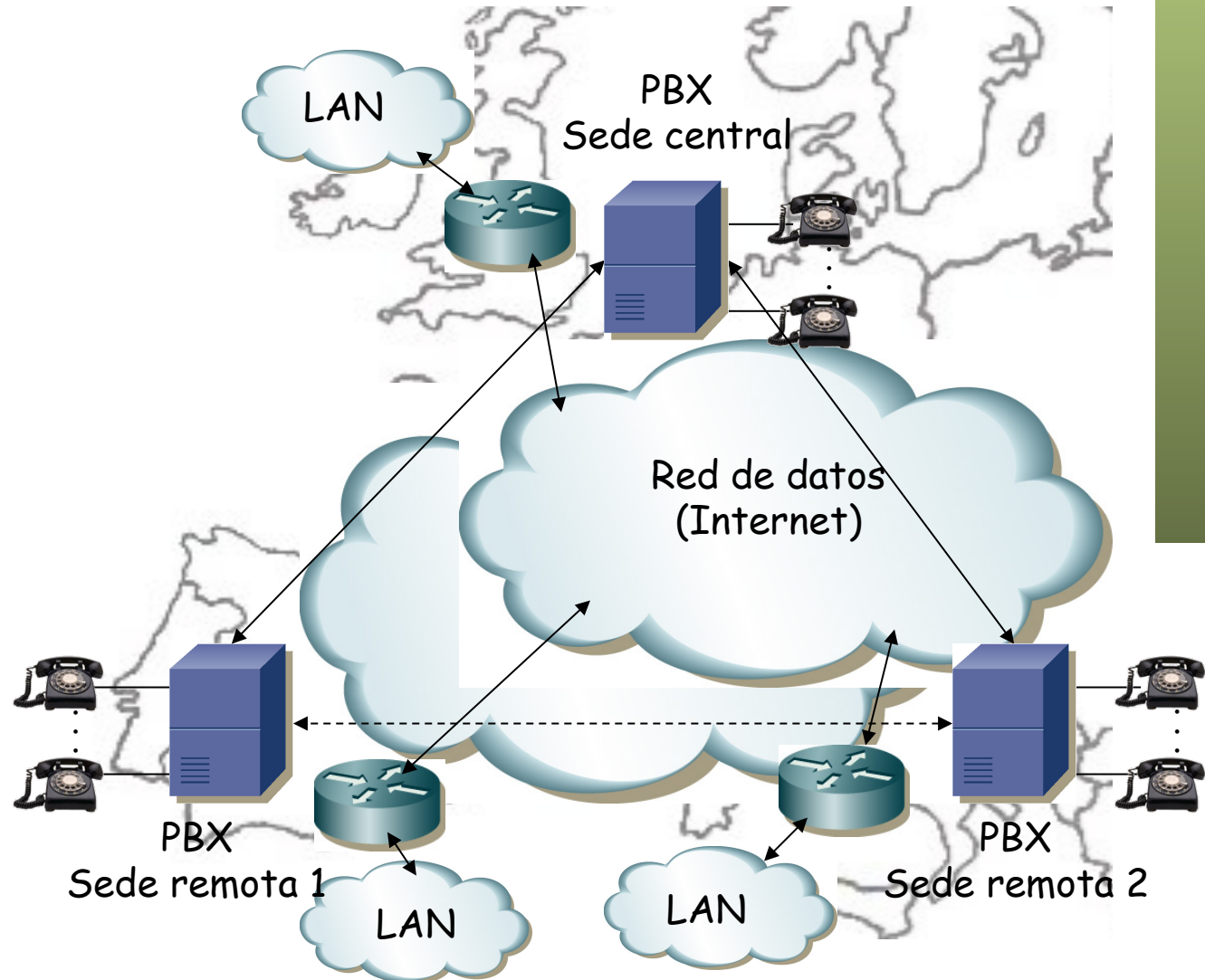
## Enlaces dedicados (malla o hub)





# Voz + datos

Probablemente tenga enlaces de datos simultáneamente



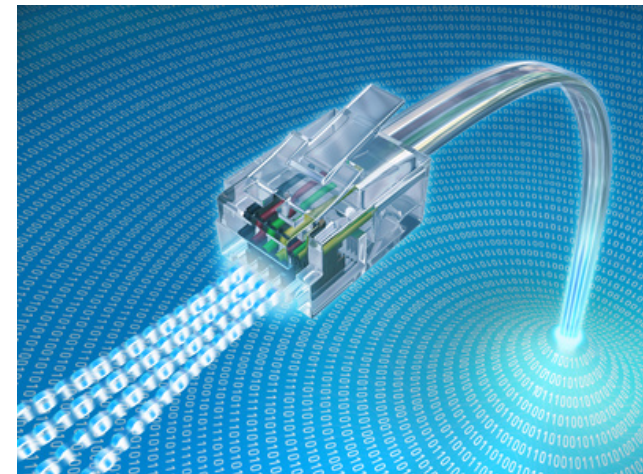
# ¿Por qué dejar de usar TDM?

- Utilizar la misma infraestructura de datos: reduce CAPEX y OPEX
- Negocio:
  - Añadir más servicios al cliente
  - Telcos añaden datos, ISPs añaden voz
- (...)



# ¿Por qué dejar de usar TDM?

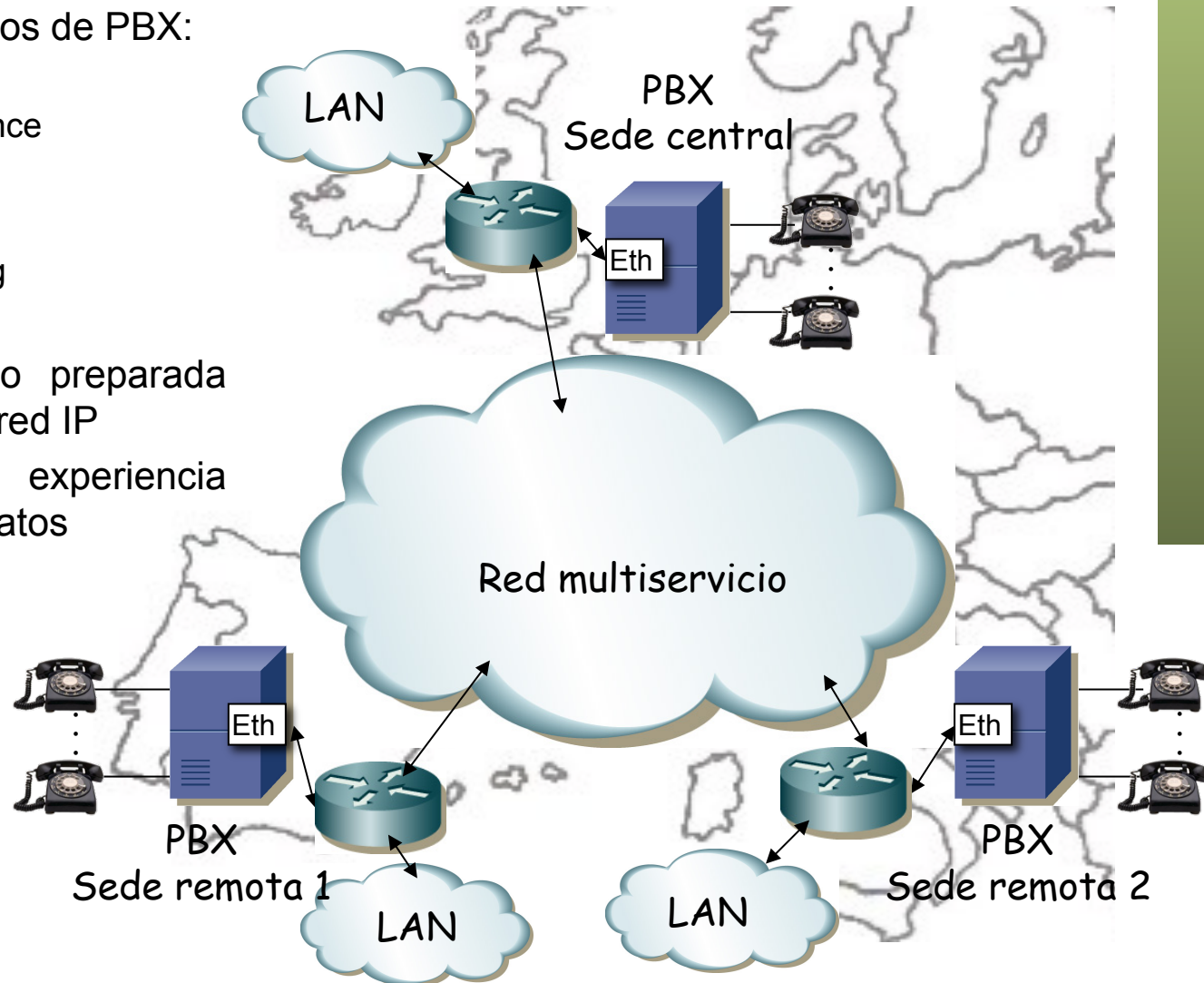
- Aumentar la cantidad de llamadas que se pueden cursar por un enlace
  - *Voice compression*
    - vs los 64 kbps PCM
    - Cuidado, reduce la calidad
  - *Silence supresion*
    - *VAD = Voice Activity Detection*
    - Habla tiene en torno a un 40-50% de actividad frente al tiempo total
  - *Statistical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



# Convergencia

## PBXs con interfaces IP (Ethernet)

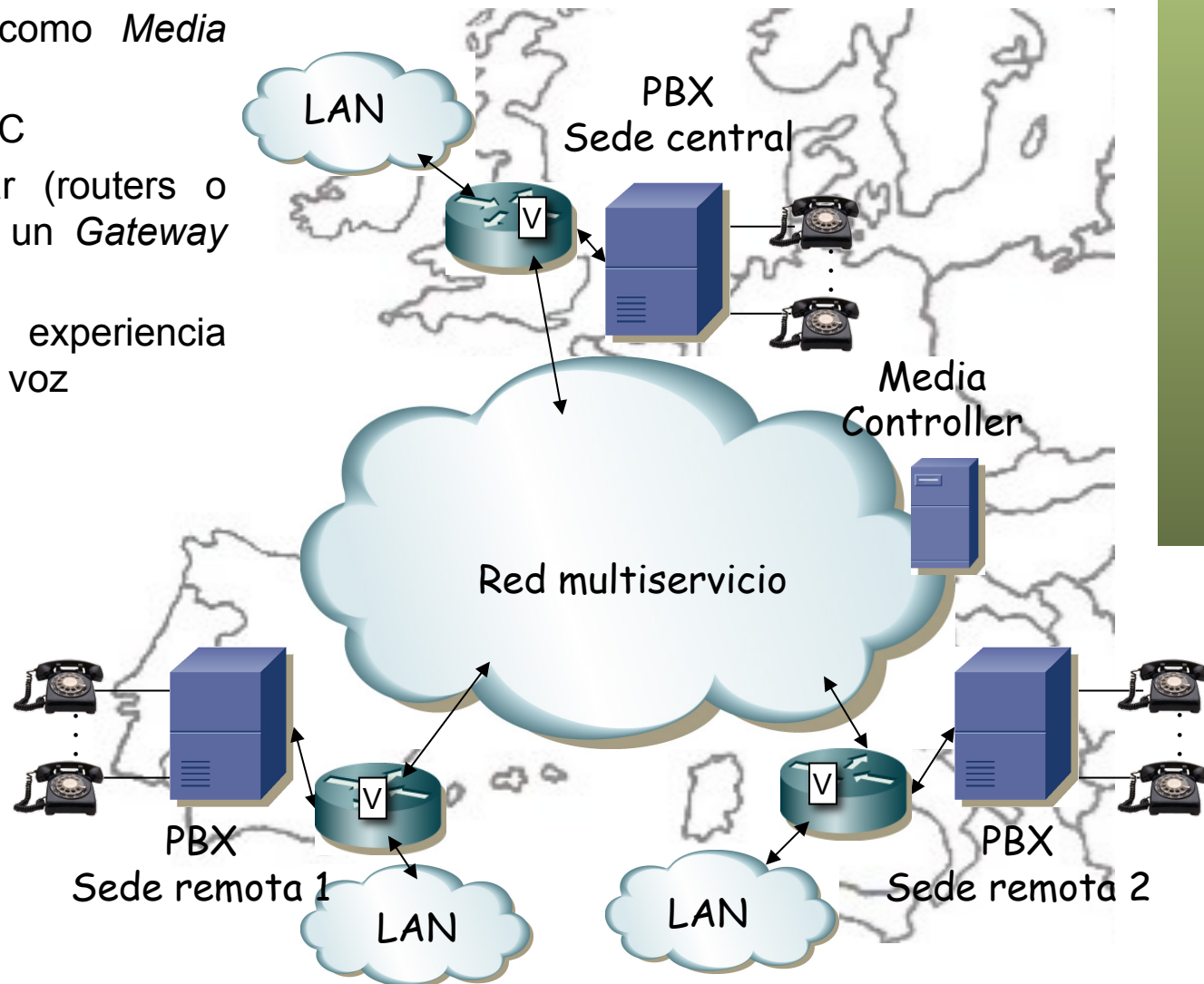
- Reutiliza servicios de PBX:
  - Call transfer
  - Call conference
  - Paging
  - Bridging
  - Group calling
  - Etc.
- Señalización no preparada para delays en red IP
- Fabricante con experiencia en voz, no en datos



# Convergencia

## PBXs trunk TDM y conversión en router

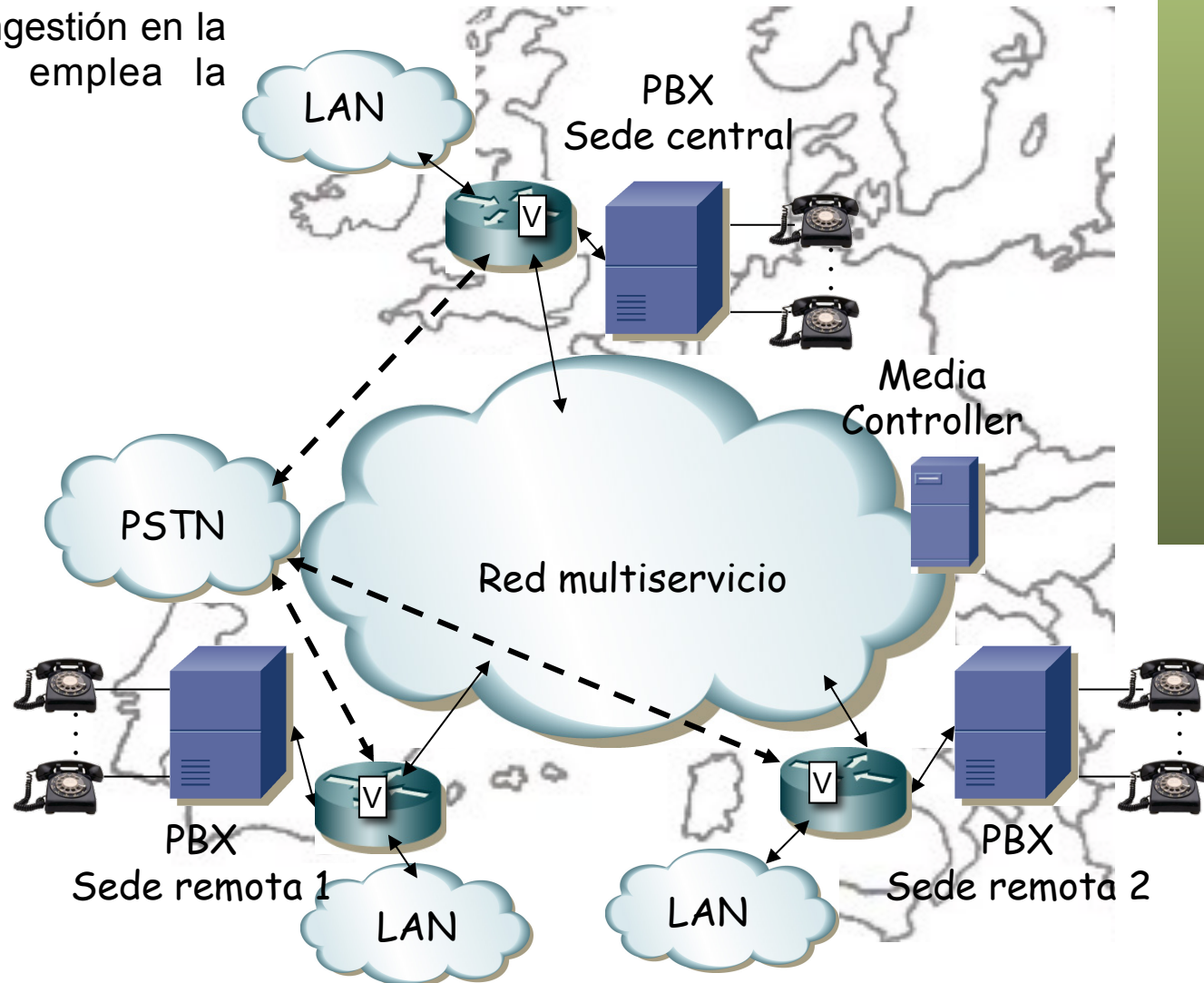
- Router actúa como *Media Gateway*
- Puede ser un PC
- En algún lugar (routers o externo) habrá un *Gateway Controller*
- Fabricante con experiencia en datos, no en voz



# Convergencia

## Multi-Point Switched Gateway

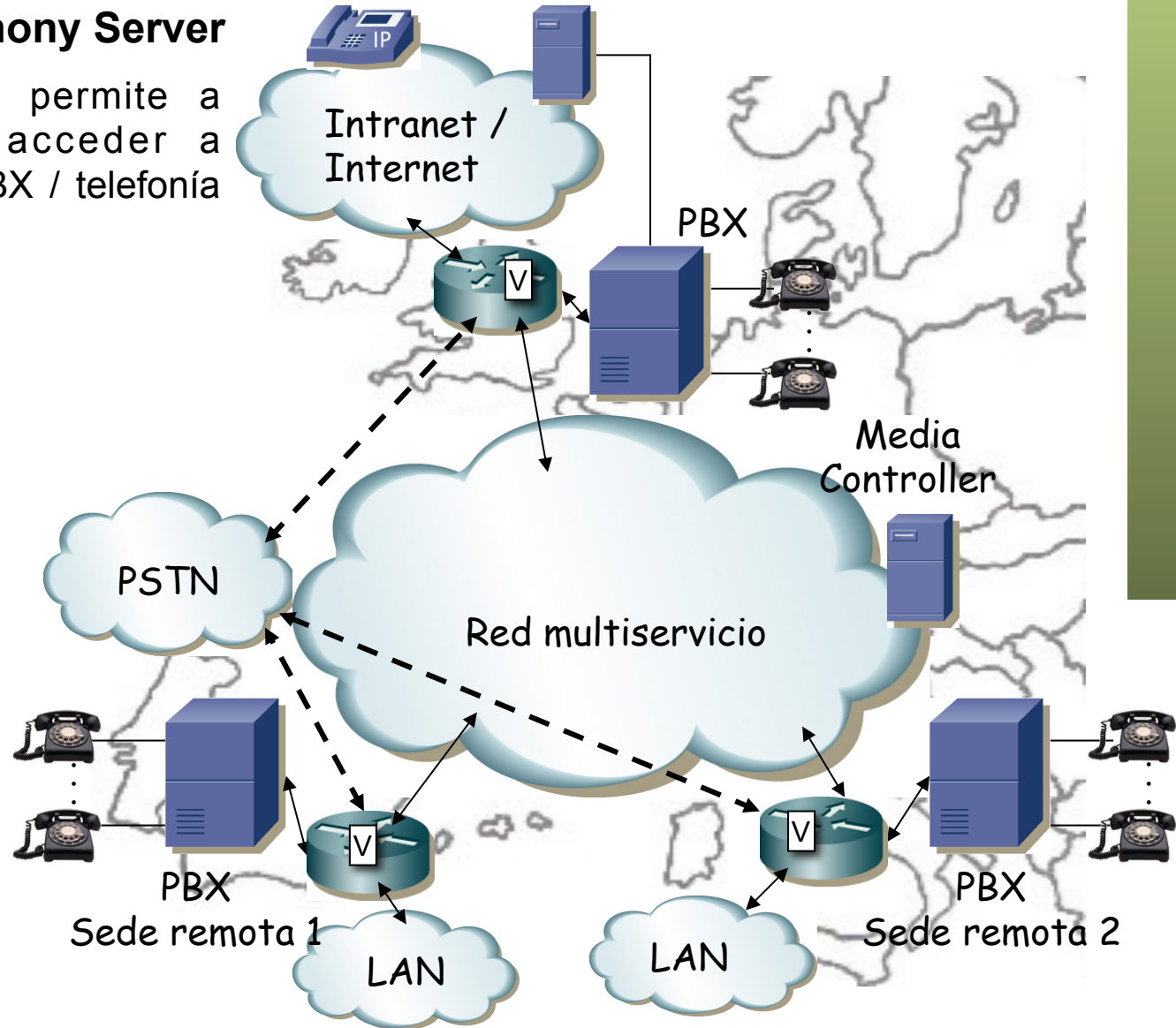
- En caso de congestión en la red de datos emplea la PSTN



# Convergencia

## (Remote) Telephony Server

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



# 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
  - El servicio ofrecido por la red
  - La calidad del decodificador
- Pueden ser métricas objetivas o subjetivas
- (...)



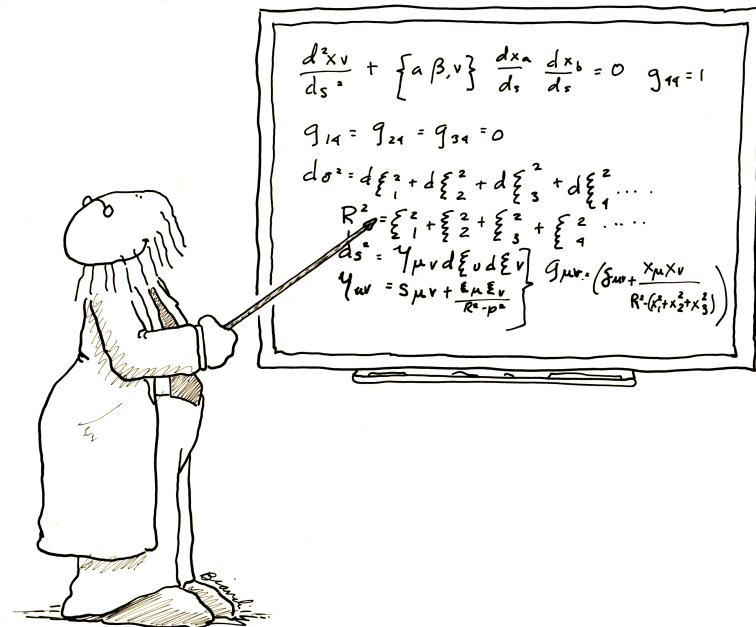
# QoE

- 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
- (...)



# QoE

- 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
- En entornos privados es razonable un límite de 200-250 ms
- ¿Más de eso? (...)

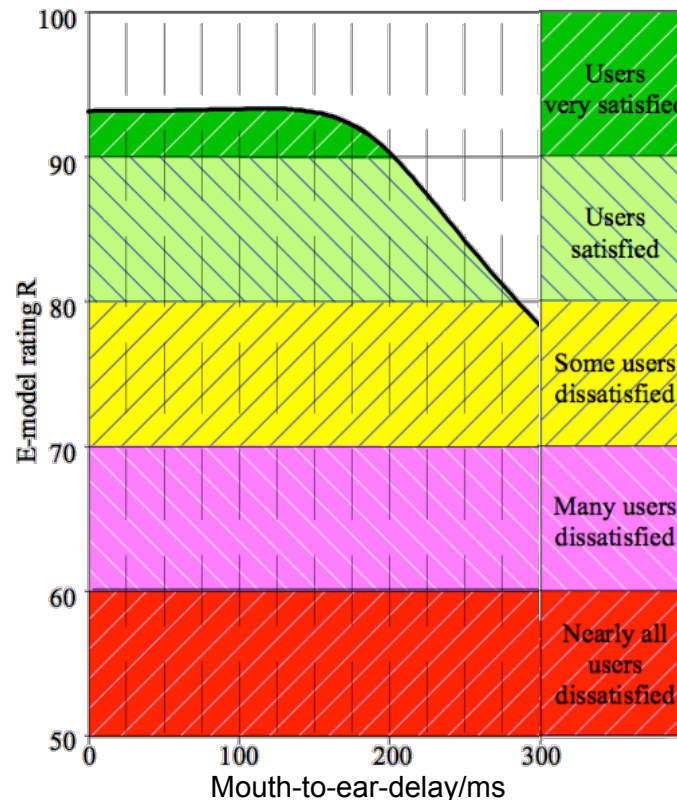


Figure 1/G.114 – Determination of the effects of absolute delay by the E-model

# Retardo end-to-end

- Baja el MOS
- >400ms es inaceptable
- ¿Escenarios con alto retardo?

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

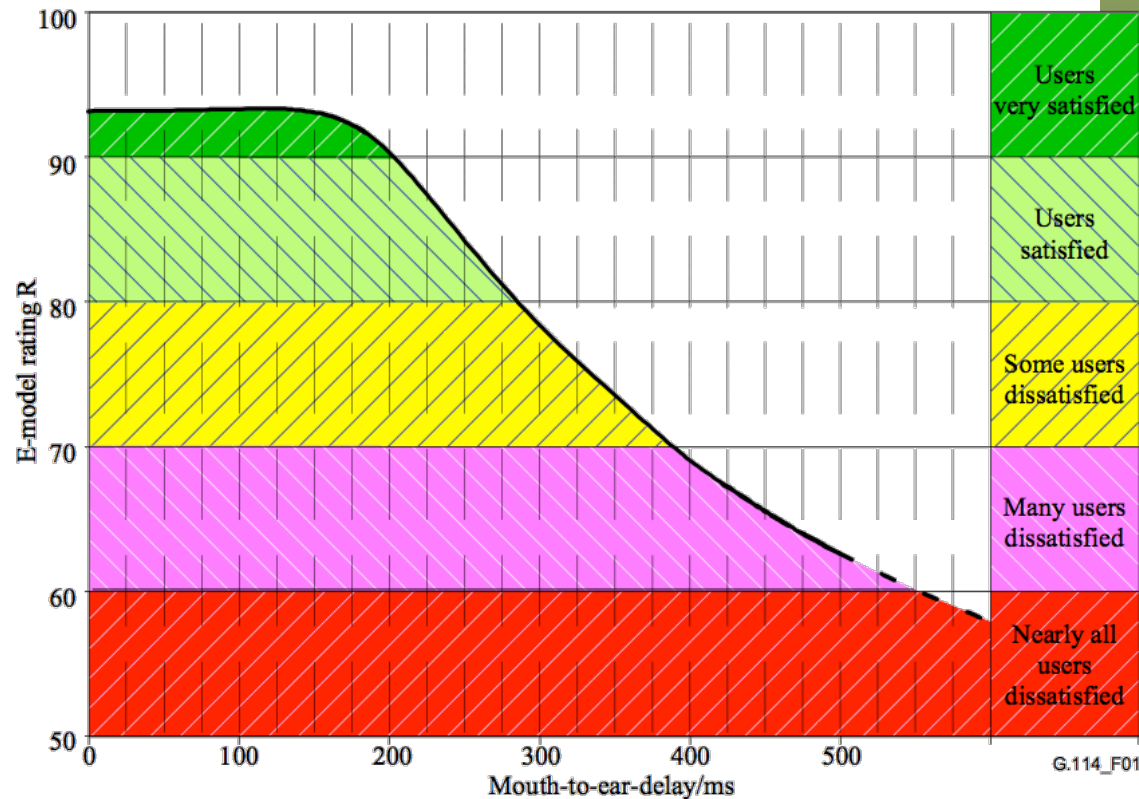


Figure 1/G.114 – Determination of the effects of absolute delay by the E-model

# Retardo end-to-end

- Órbita geosíncrona:  $36.000\text{km} / 300.000 \text{ km/s} = 120\text{ms}$
- Hay que subir al satélite y volver a bajar así que  $2 \times 120\text{ms} = 240\text{ms}$
- ¡ Sólo de propagación ?

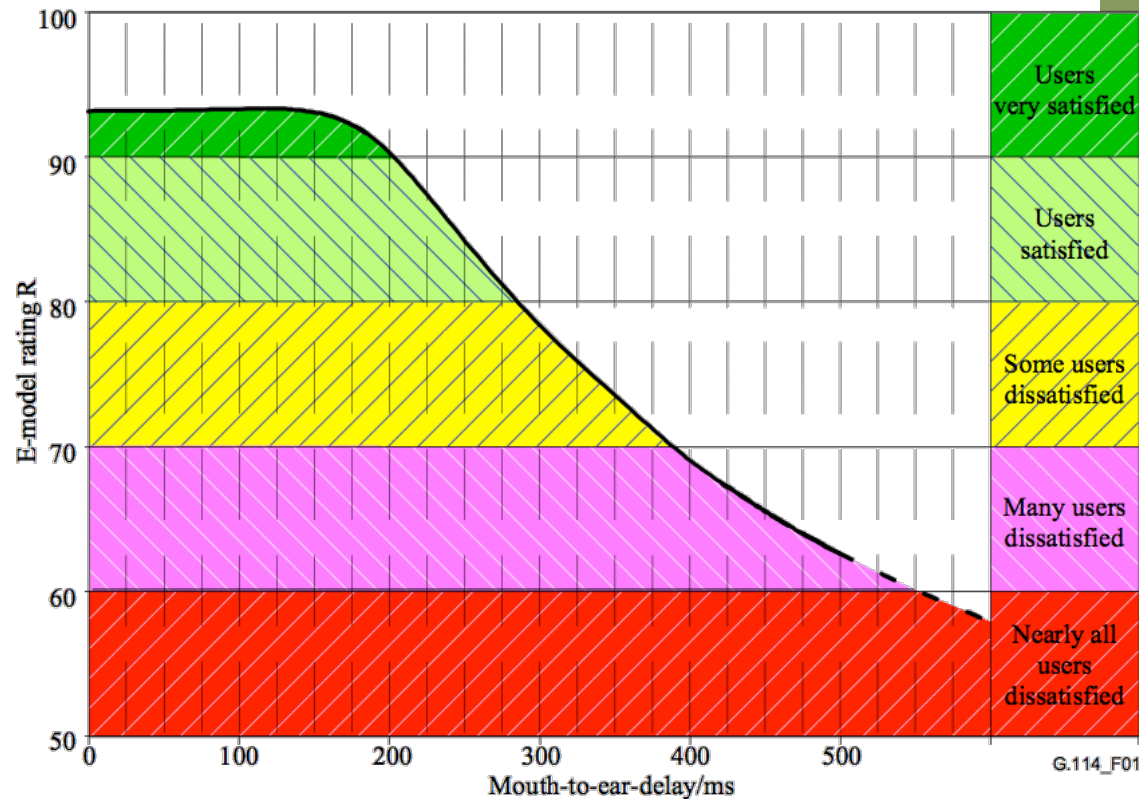
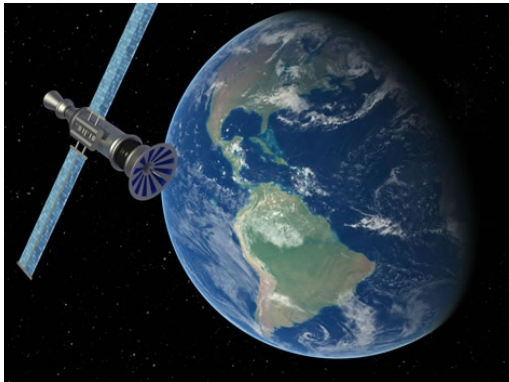
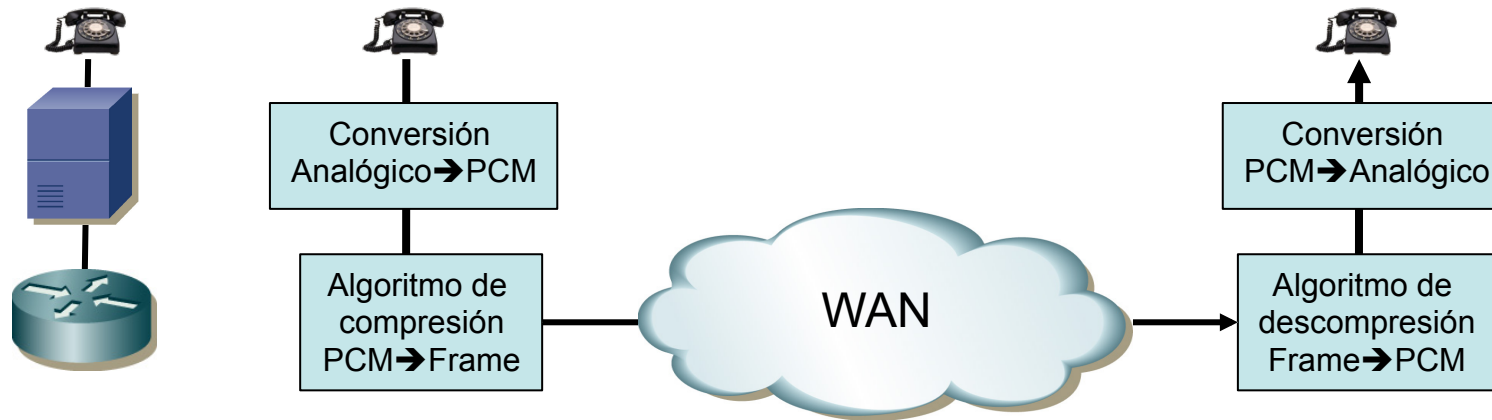


Figure 1/G.114 – Determination of the effects of absolute delay by the E-model

# Codificación

# Flujo extremo a extremo

- Flujo a través de una WAN

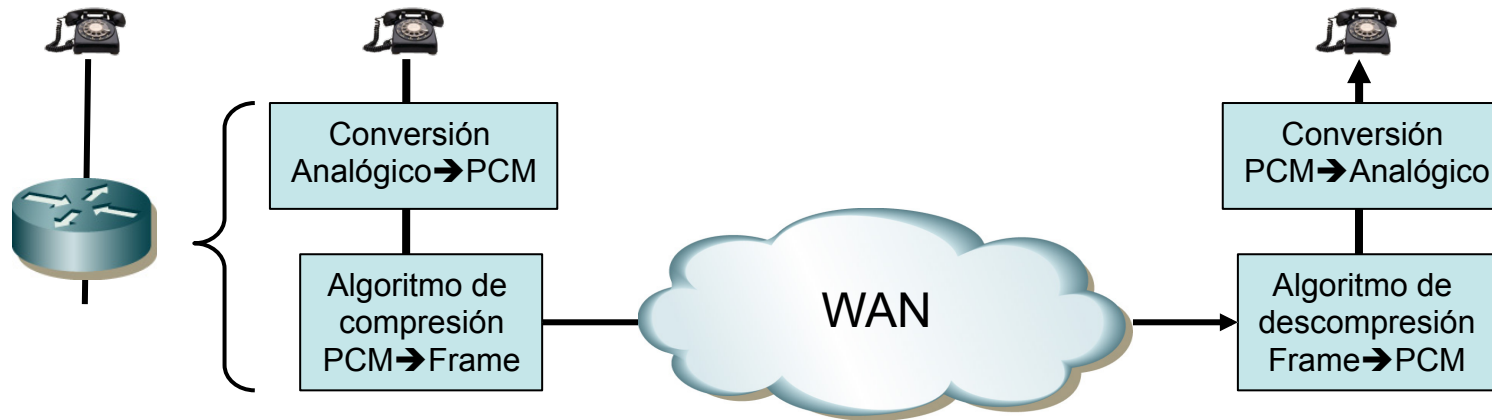


- Una PBX digital puede ser quien hace la digitalización
- Un router entonces suele cubrir la función de compresión y paquetización
- Podría la PBX integrar la funcionalidad del router
- (...)



# 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
- Varían en complejidad, BW requerido y calidad de sonido ofrecida
- Los más complejos ofrecen mayor calidad percibida y menor bitrate pero con mayores tiempos de procesamiento

## Ejemplos:

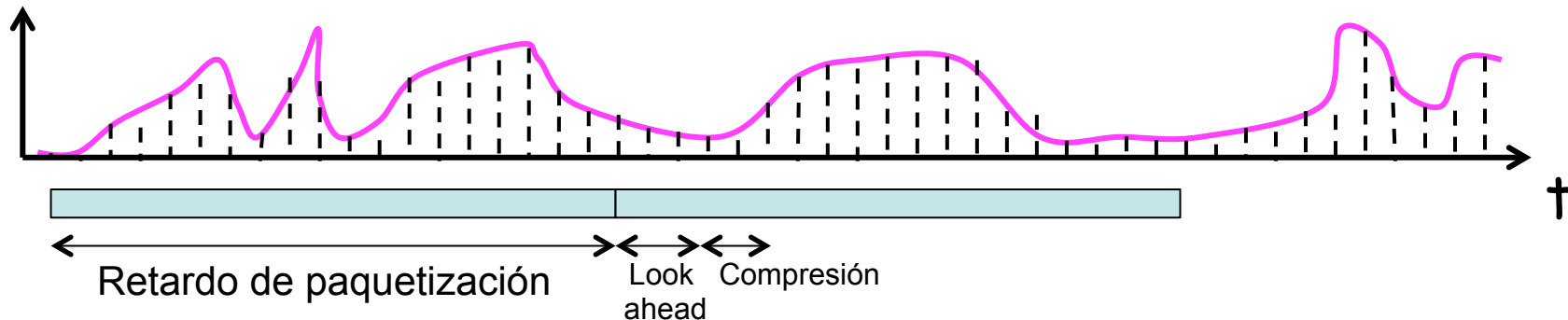
- G.711 basado en PCM (Pulse Code Modulation), simplemente muestreando a intervalos regulares
- G.726, ADPCM (Adaptive Differential PCM), usa predicción de la siguiente muestra y cuantiza eso
- G.723 y G.729 trabajan con bloques de muestras (*frame-based*) para las que aplican técnicas de compresión (ACELP = Algebraic Code Excited Linear Prediction)

The screenshot shows a software interface with three tabs: 'Captura', 'Programación', and 'Configuración'. The 'Configuración' tab is active. Under the heading 'Formato audio:', there are four settings: 'Codec audio:' set to 'MP2', 'Bitrate' set to '128 kbps', 'Frec. de muestreo:' set to '44100 Hz', and 'Canales audio:' set to 'Stereo'. Below this, under the heading 'Parámetros del dispositivo:', there are two settings: 'Entrada vídeo:' set to 'Ninguno' and 'Estándar vídeo:' set to 'Ninguno'.

# Retardos

# Componentes del retardo

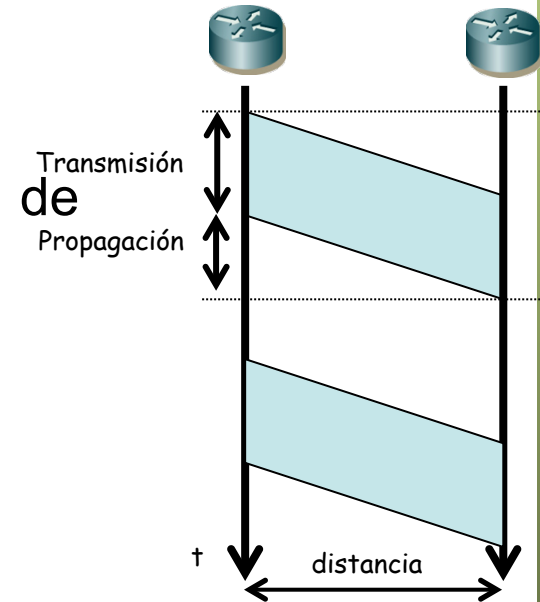
- **Algorithmic Delay (look ahead)**
  - Los algoritmos de compresión suelen necesitar conocer muestras siguientes a las del bloque a comprimir
  - Eso implica que hay que esperar a que se generen
  - Para G.726 es de 0 ms, para G.729 de 5 ms, para G.723.1 de 7.5 ms



# Retardos constantes

## Retardo de serialización

- Tiempo de transmisión
  - Solo se mejora aumentando la velocidad de transmisión
  - Despreciable por encima de 100Mbps
- (...)

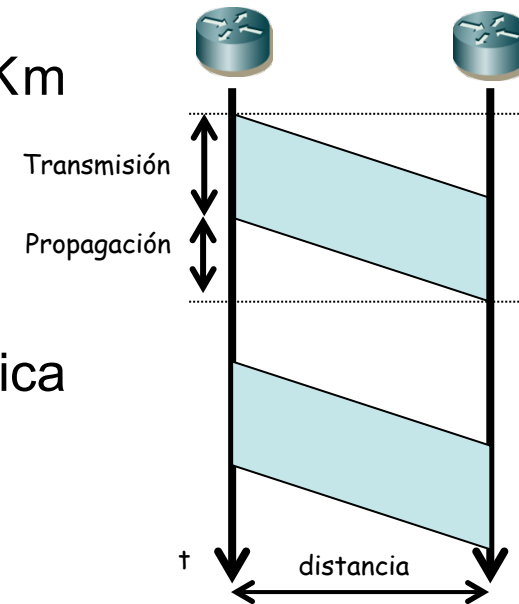


# Retardos constantes

## 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)

(...)



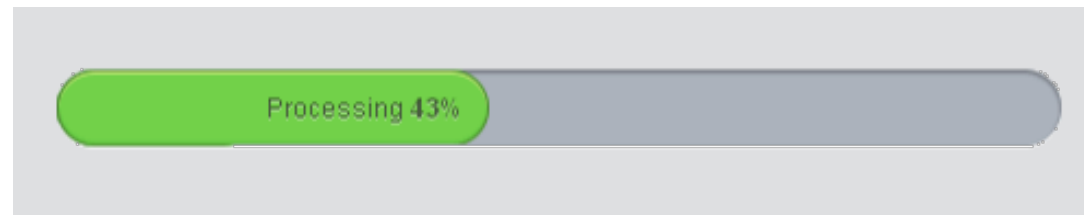
D	R
$D < 1000 \text{ km}$	$1.5 \times D$
$1000 \text{ km} \leq D \leq 1200 \text{ km}$	1500 km
$D > 1200 \text{ km}$	$1.25 \times D$

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

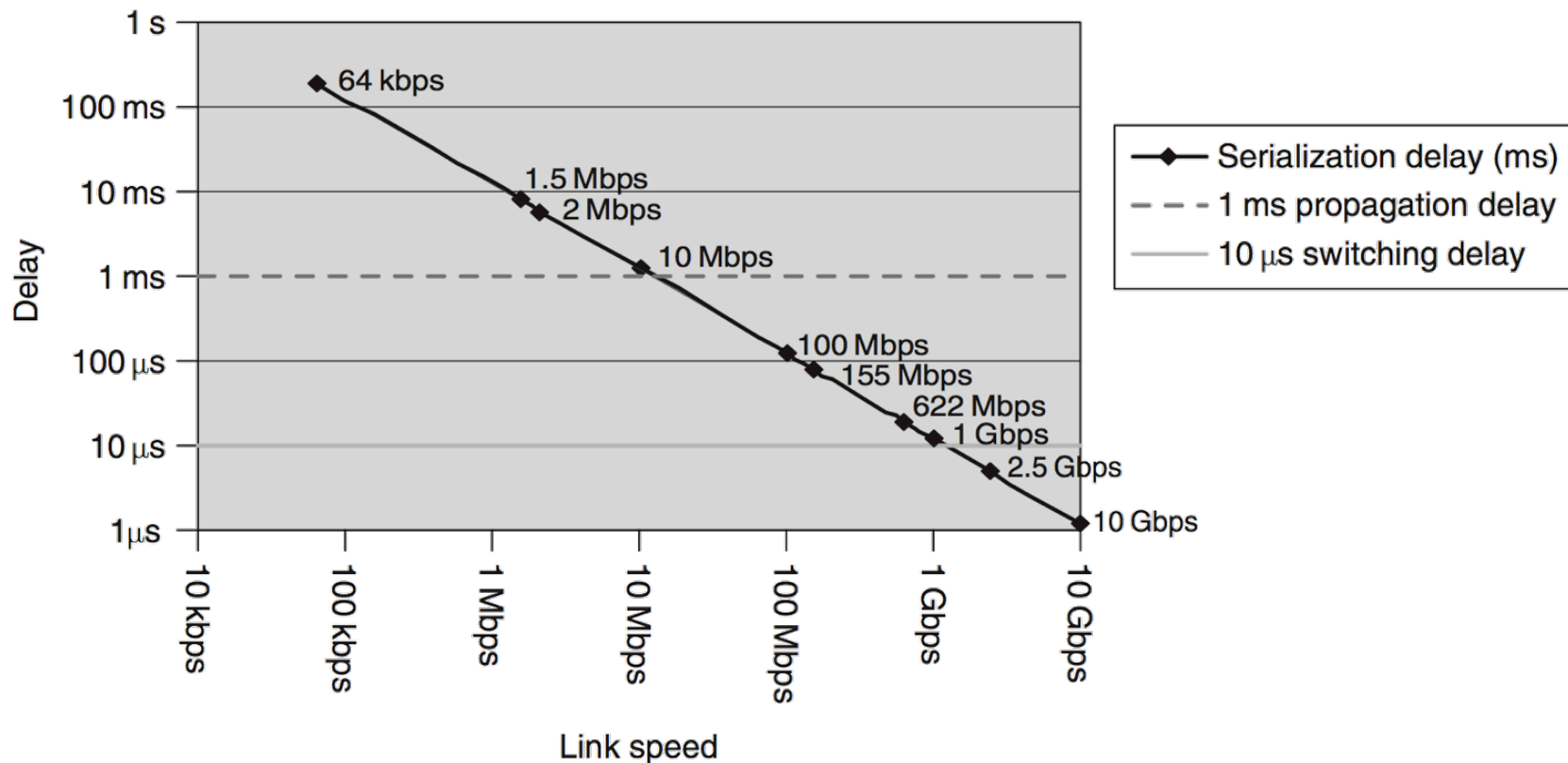
## Tiempo de procesamiento/conmutación

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



# Retardos constantes

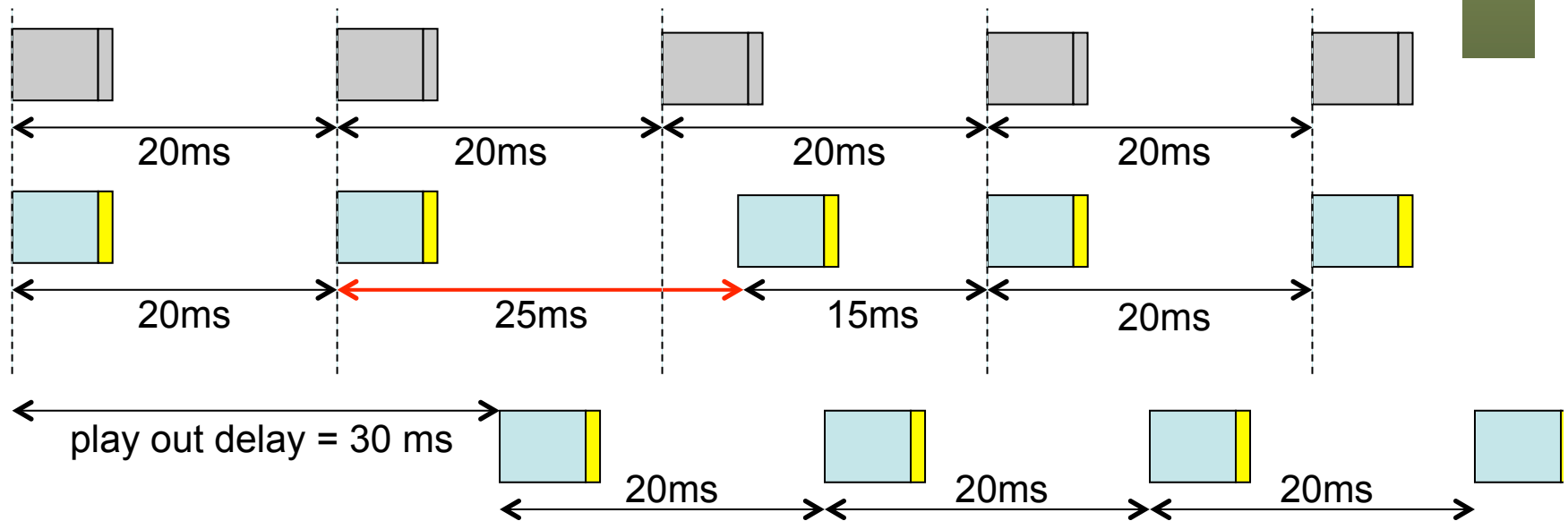
- Ejemplo comparativo
  - Retardo de serialización, de propagación y de conmutación
  - Paquete de 1500 bytes
  - Unos 200Km de fibra : 1ms de propagación





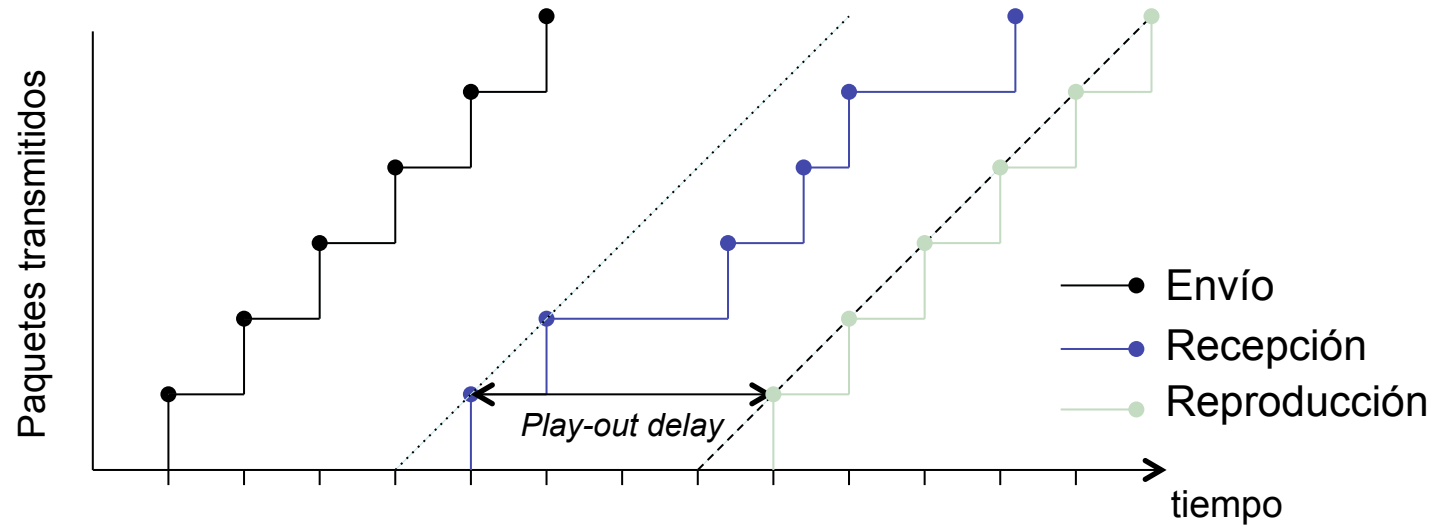
# De-jitter delay

- “De-jitter buffer” o “play-out buffer”
- Introducen un *play out delay*
- Variación en el retardo → retardo constante en el receptor
- Obliga a la red a tener menor retardo
- Si el buffer es demasiado pequeño obliga a la red a garantizar un jitter bajo

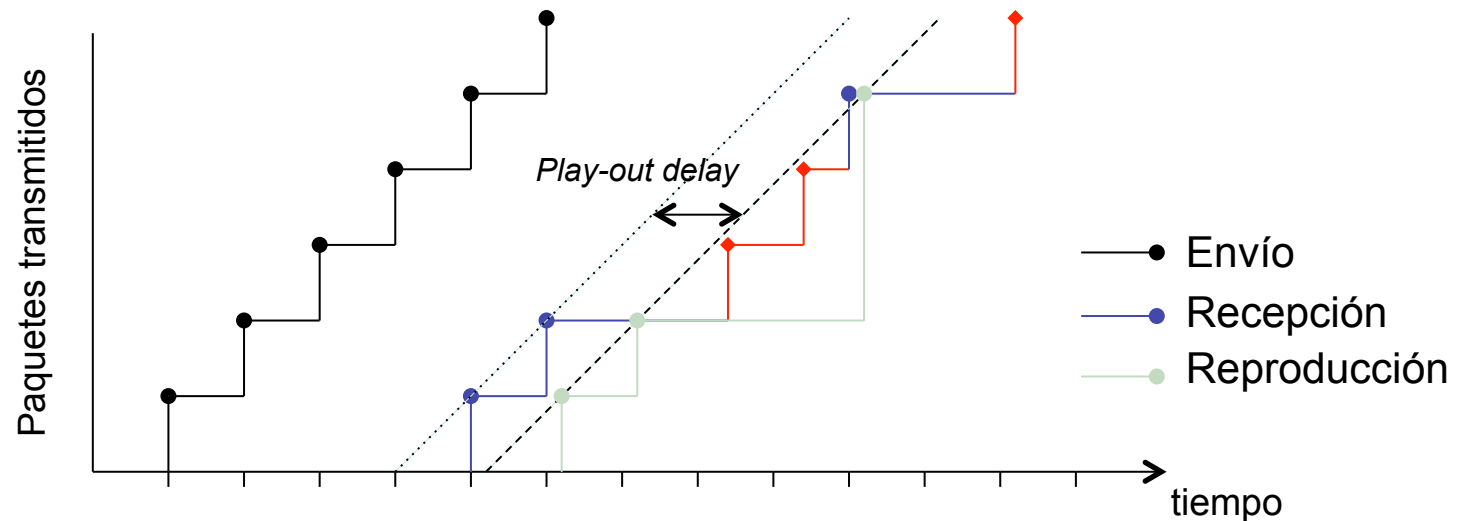


# De-jitter delay

- Ejemplo de buffer muy grande:

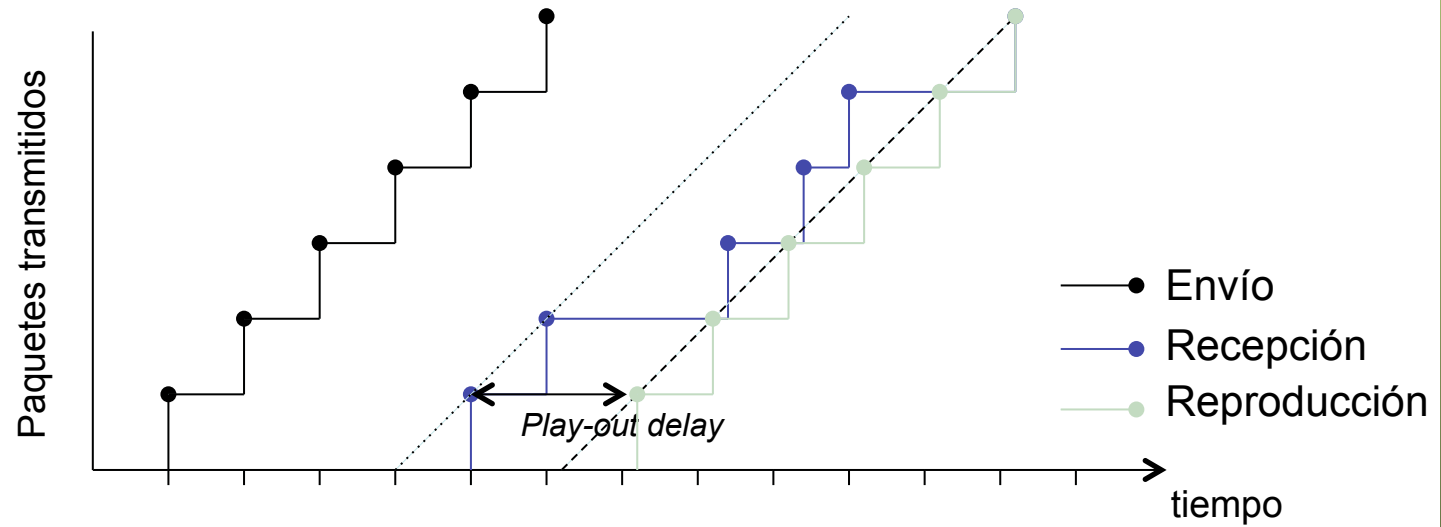


- Ejemplo de buffer demasiado **pequeño**:



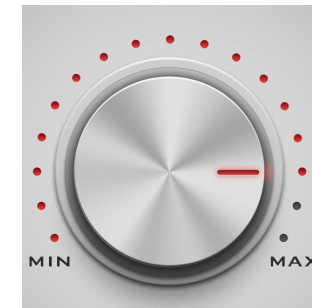
# De-jitter delay

- Ejemplo de buffer óptimo:



# De-jitter delay

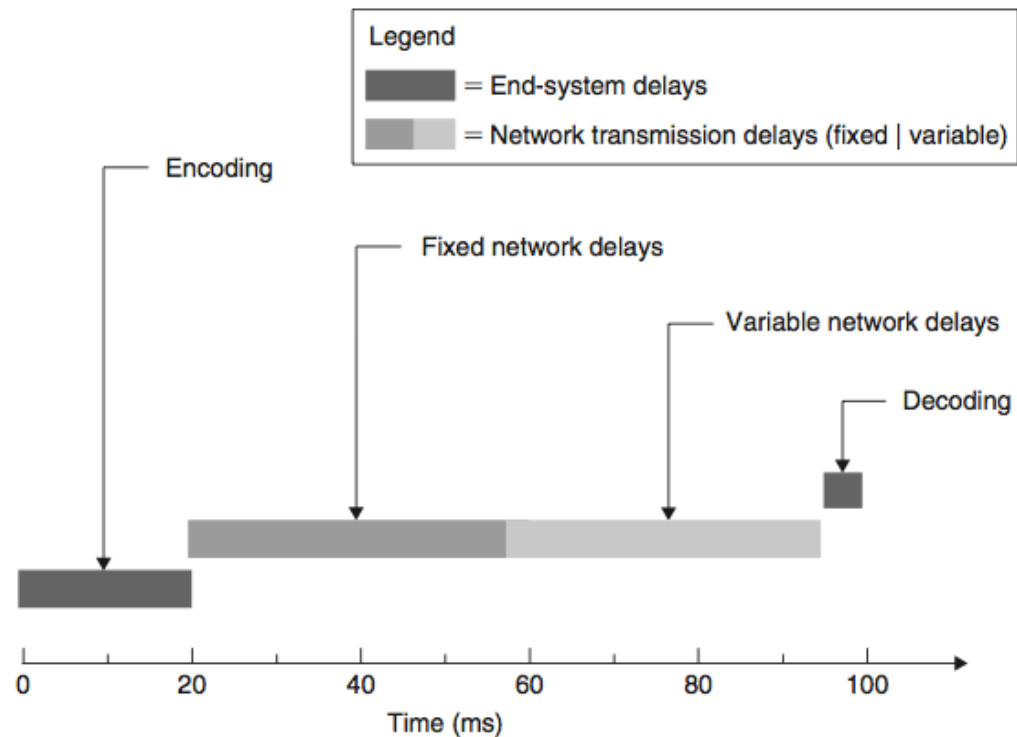
- Valor óptimo: el del retardo variable extremo a extremo
- Valores de *play out delay* en torno a los 40 ms
- Se emplea buffer adaptativo (con un valor máximo)
- Si se vacía el buffer entonces la próxima llegada se retiene durante el *play out delay* para reiniciar buffer
- Si se produce un underflow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



# Retardos variables

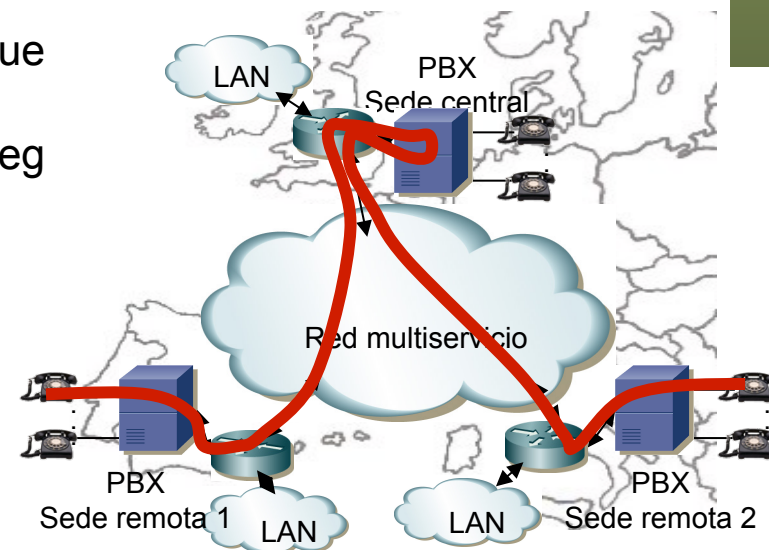
## Retardo en cola (scheduling delay)

- 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