

# Descripción de la Demora en las Redes de Voz de Paquetes

## Contenido

[Introducción](#)

[Flujo básico de la voz](#)

[¿Cómo funciona la compresión de voz?](#)

[Estándares para límites de retardo](#)

[Orígenes de retrasos](#)

[Retardo del codificador \(procesamiento\)](#)

[Retardo de empaquetado](#)

[Demora de serialización](#)

[Retardo de colocación en cola/almacenamiento en memoria intermedia](#)

[Retardo de Switching de la Red](#)

[Eliminar fluctuación de la demora](#)

[Estimación de la Demora Prevista](#)

[Conexión de un solo salto](#)

[Dos Saltos en una Red Pública con un C7200 que Actúa como Switch Tándem](#)

[Conexión de dos saltos a través de una red pública con un switch Tandem PBX](#)

[Conexión de dos saltos en una red privada con un switch tándem PBX](#)

[Efectos de los ciclos de compresión múltiple](#)

[Consideraciones para conexiones con retardo prolongado](#)

[Información Relacionada](#)

## Introducción

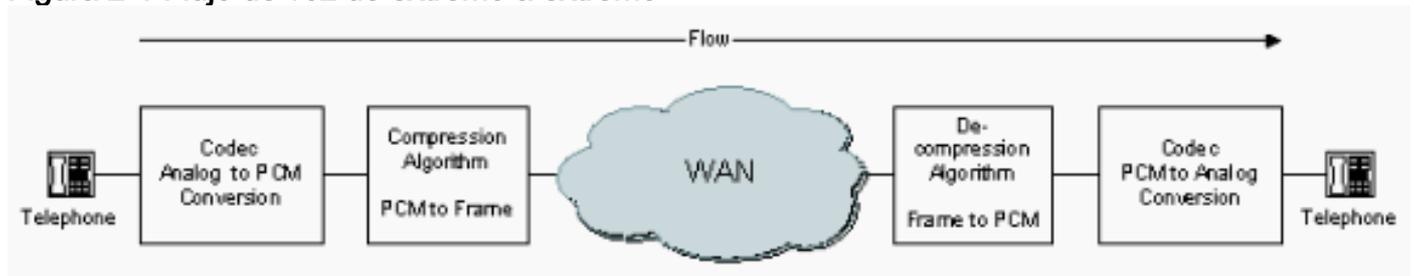
Al diseñar redes que transportan voz en infraestructuras de paquetes, tramas o células, es importante comprender y tener en cuenta los componentes de la demora en la red. Si tiene en cuenta todas las posibles demoras, garantiza un rendimiento global aceptable de la red. La calidad global de voz depende de muchos factores que incluyen el algoritmo de compresión, los errores y la pérdida de tramas, la cancelación del eco y la demora. En este artículo se explican las causas de la demora al usar un router/gateways Cisco sobre redes de paquetes. Aunque los ejemplos estén destinados a la Retransmisión de tramas, los conceptos también son aplicables a las redes de Voz sobre IP (VoIP) y de Voz sobre ATM (VoATM).

## Flujo básico de la voz

En el siguiente diagrama se muestra el flujo de un circuito de voz comprimida. La señal analógica del teléfono se digitaliza en señales de Modulación por impulsos codificados (PCM) mediante el codificador-decodificador (códec). Después, las muestras de PCM son pasadas al algoritmo de compresión, que comprime la voz en un formato de paquete para su transmisión a través de la

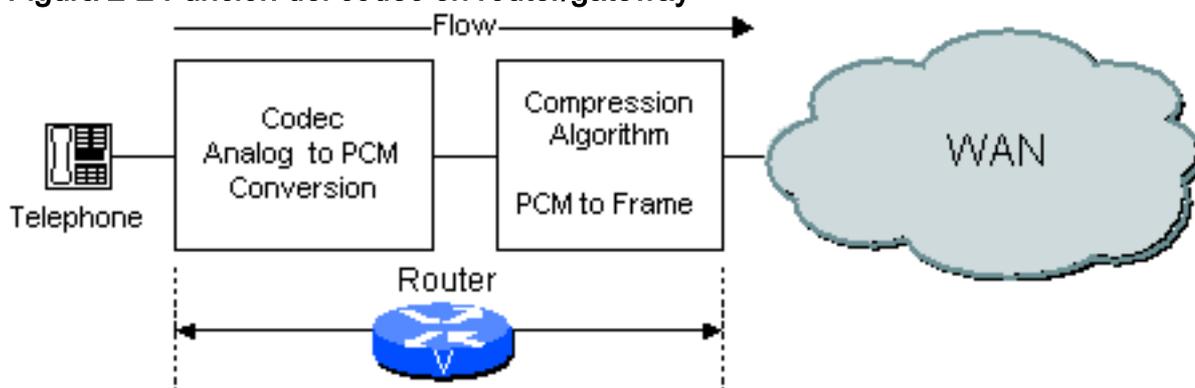
WAN. En el otro extremo de la nube se ejecutan las mismas funciones en orden inverso. El flujo completo se muestra en la Figura 2-1.

Figura 2-1 Flujo de voz de extremo a extremo



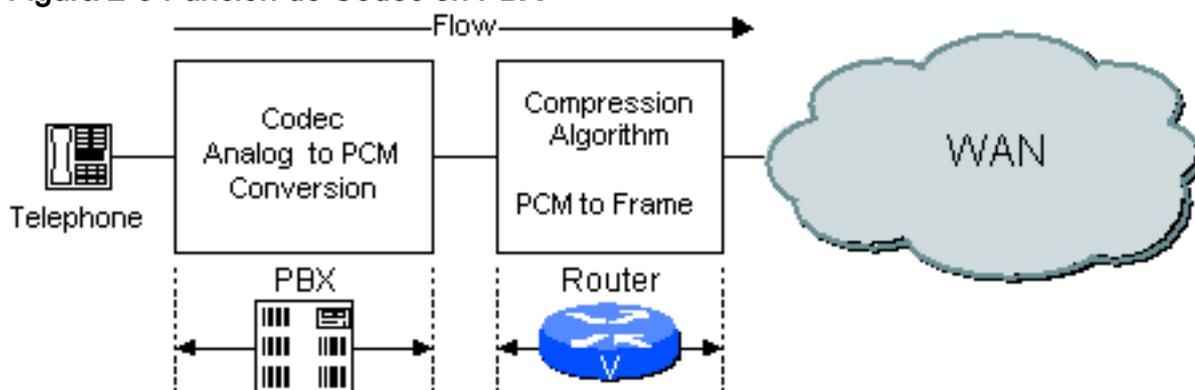
Según la configuración de la red, el router/gateway puede llevar a cabo la función de códec y la de compresión, o solamente una de ellas. Por ejemplo, si se utiliza el sistema de voz análogo, el router/gateway realizará la función CODEC y la función de compresión según se muestra en la figura 2-2.

Figura 2-2 Función del codec en router/gateway



Si se utiliza un PBX digital, el PBX realiza la función de códec y el router procesa las muestras PCM que le pasa el PBX. Se muestra un ejemplo en la Figura 2-3.

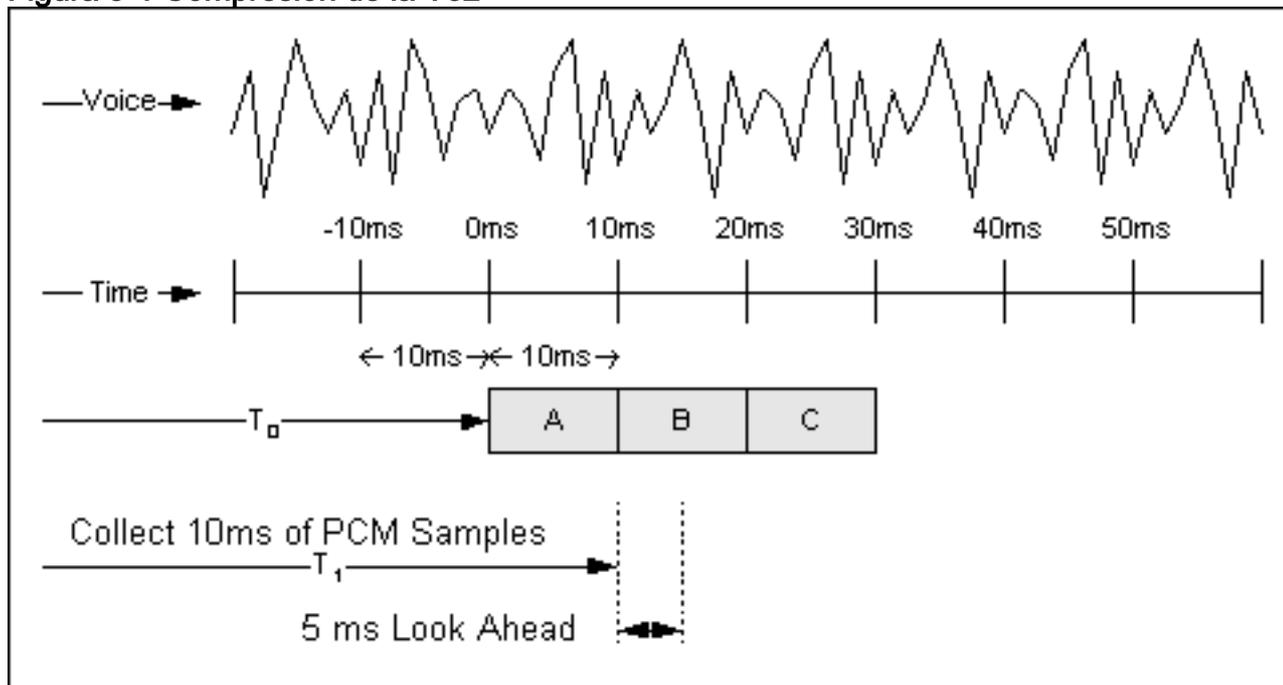
Figura 2-3 Función de Códec en PBX



## ¿Cómo funciona la compresión de voz?

Los algoritmos de compresión de alta complejidad utilizados en el router/gateways Cisco analizan un bloque de muestras PCM suministradas por el códec de voz. La longitud de estos bloques varía según el codificador. Por ejemplo, el tamaño de bloque básico usado por un algoritmo G.279 es 10 ms mientras que el tamaño de bloque básico usado por los algoritmos G.273.1 es 30 ms. La Figura 3-1 muestra un ejemplo del funcionamiento del sistema de compresión G. 729.

Figura 3-1 Compresión de la Voz



La secuencia de voz analógica se digitaliza en muestras de PCM y se entrega al algoritmo de compresión en aumentos de 10 ms. Discutiremos la vista preliminar en Retraso Algorítmico.

## Estándares para límites de retardo

La Unión Internacional de Telecomunicaciones (ITU) tiene en cuenta la demora de red para las aplicaciones de voz en la Recomendación G.114. Esta recomendación define tres bandas de demora unidireccional, tal y como se indica en la Tabla 4.1.

Tabla 4.1 Especificación del retardo

Intervalo en milisegundos	Descripción
0-150	Aceptable para la mayoría de las aplicaciones de usuario.
150-400	Aceptable a condición de que los administradores sean conscientes del tiempo de transmisión y del impacto que tiene en la calidad de transmisión de las aplicaciones de usuario.
Superior a 400	Inaceptable para objetivos de planificación general de la red. Sin embargo, se reconoce que en algunos casos excepcionales se supera este límite.

**Nota:** Estas recomendaciones son para conexiones con eco adecuadamente controladas. Esto implica el uso de canceladores de eco. Se requieren los canceladores de eco cuando el retraso unidireccional excede 25 ms (G.131).

Estas recomendaciones se destinan a las administraciones de telecomunicaciones nacionales. Por ello, son más estrictas que las que se aplican normalmente a las redes de voz privadas. Cuando el diseñador de la red conoce la ubicación y las necesidades comerciales de los usuarios finales, puede ser aceptable una demora mayor. Para redes privadas, una demora de 200 ms es un objetivo razonable y el límite es 250 ms. Todas las redes deben ser diseñadas de forma que se conozca y minimice la máxima demora de conexión de voz esperada.

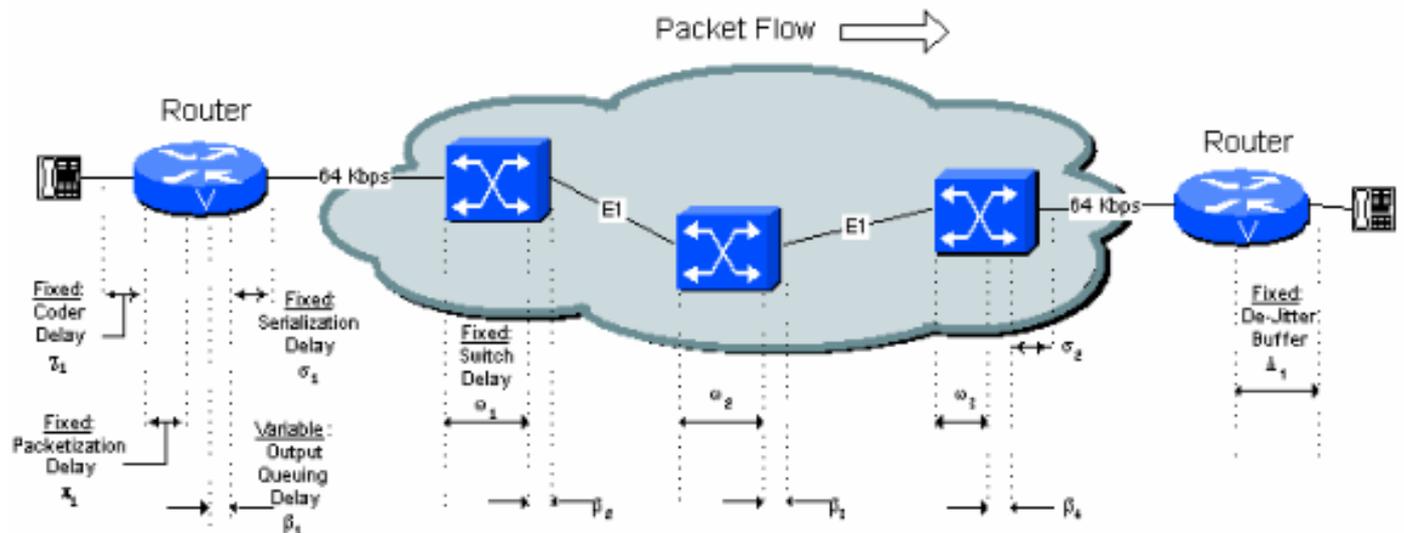
## Orígenes de retrasos

Existen dos tipos distintos de demora: fija y variable.

- Los componentes de la demora fija se añaden directamente a la demora global de la conexión.
- Los retrasos variables surgen de retrasos de almacenamiento en cola en los búfers troncales de egreso en el puerto serial conectado a la WAN. Estos búfers crean retrasos variables, llamados fluctuaciones, a través de la red. Las demoras variables se controlan a través del buffer anti-fluctuaciones y el router/gateway de recepción. El búfer de eliminación de fluctuación se describe en la sección Retraso de eliminación de fluctuación ( $\Delta_n$ ) de este documento.

La figura 5-1 identifica todos los orígenes de retardo fijos y variables de la red. En este documento se describe en detalle cada causa.

Figura 5-1: Causas de la demora



## Retardo del codificador (procesamiento)

La demora del codificador es el tiempo que tarda el procesador de señales digitales (DSP) en comprimir un bloque de muestras de PCM. Esto también se denomina demora de procesamiento ( $\tau_n$ ). La demora varía en función del codificador utilizado y la velocidad del procesador. Por ejemplo, los algoritmos de predicción lineal activada por código algebraica (ACELP) analizan un bloque de 10 ms de muestras de PCM y después las comprimen.

El tiempo de compresión para un proceso de Predicción Lineal Activada por Código Algebraica de Estructura Conjugada (CS-ACELP) varía entre 2,5 ms y 10 ms en función de la carga del procesador DSP. Si se carga el DSP completamente con cuatro canales de voz, la demora del codificador es de 10 ms. Si se carga el DSP solamente con un canal de voz, la demora del

codificador es de 2,5 ms. Para fines de diseño, utilice el tiempo correspondiente al peor caso posible, de 10 ms.

El tiempo de descompresión es aproximadamente el diez por ciento del tiempo de compresión para cada bloque. No obstante, hay que tener en cuenta que el tiempo de descompresión es proporcional al número de muestras por trama debido a la presencia de varias muestras. En consecuencia, el tiempo de descompresión en el peor caso posible para una trama con tres muestras es de 3 x 1 ms, 3 ms. Generalmente, se colocan dos o tres bloques de salida G.729 comprimida en una trama; en cambio una muestra de salida G.723.1 comprimida se envía en una sola trama.

En la Tabla 5.1 se muestran las demoras del codificador correspondientes al mejor y al peor caso posible.

### SRC\_INVALID

Codificador	Velocidad	Muestra de bloque requerido	Mejor caso de retraso del codificador	Peor caso de retardo del decodificador
ADPCM, G.726	32 Kbps	10 ms	2,5 ms	10 ms
CS-ACELP, G.729A	8,0 Kbps	10 ms	2,5 ms	10 ms
MP-MLQ, G.723.1	6,3 Kbps	30 ms	5 ms	20 ms
MP-ACELP, G.723.1	5,3 Kbps	30 ms	5 ms	20 ms

### Demora algorítmica

El algoritmo de compresión se basa en características conocidas de la voz para procesar correctamente el bloque de ejemplo N. El algoritmo debe tener cierto conocimiento sobre qué hay en el bloque N+1 para reproducir de manera precisa el bloque de ejemplo N. Este tiempo que se tarda en mirar el bloque siguiente, que supone en realidad una demora adicional, se denomina demora algorítmica. Esto aumenta de hecho la longitud del bloque de compresión.

Esto sucede repetidamente, ya que el bloque N+1 mira en el bloque N+2, y así sucesivamente. El efecto de red es un agregado de 5 ms al retraso general en el link. Esto significa que el tiempo total requerido para procesar un bloque de información es de 10 ms con un factor de costo operativo constante de 5 ms. Vea la Figura 3-1: Compresión de voz.

- El Retraso algorítmico de los codificadores G.726 es de 0 ms
- El Retraso algorítmico para los codificadores G.729 es de 5 ms.
- El retraso algorítmico para los codificadores G.723.1 es de 7.5 ms.

Para los ejemplos del resto de este documento, suponga una compresión G.729 con un contenido

de 30 ms/30 bytes. Para facilitar el diseño y adoptar un enfoque conservador, en las tablas del resto del documento se supone el peor caso posible para la demora del codificador. La demora del codificador, la demora de descompresión y la demora algorítmica se reúnen en un factor que se denomina demora del codificador.

La ecuación utilizada para generar el parámetro de retardo del codificador localizado es:

**Ecuación 1: Parámetro de Demora del Codificador Concentrada**

<p><b>(Worst Case Compression Time Per Block)</b></p> <p><b>+</b></p> <p><b>(De-Compression Time Per Block)</b></p> <p><b>X (Number of Blocks in Frame)</b></p> <p><b>+</b></p> <p><b>(Algorithmic Delay)</b></p> <hr style="border: 1px solid black; width: 80%; margin: 0 auto;"/> <p><b>= "Lumped" Coder Delay Parameter</b></p>
---

La demora del codificador concentrada para G.729 que se utiliza en el resto de este documento es:

Peor tiempo de compresión de caso por bloque: 10 ms

Tiempo de descompresión por bloque x 3 bloques 3 ms

Retraso algorítmico 5 ms —

Total (χ) 18 ms

**[Retardo de empaquetado](#)**

El retardo de empaquetado ( $T_{pn}$ ) es el tiempo que se tarda en llenar una carga útil de paquete con voz codificada/comprimida. Esta demora es una función del tamaño del bloque de ejemplo requeridos por el codificador de voz y del número de bloques insertados en una sola trama. La demora de generación de paquetes también se puede llamar demora de acumulación, ya que las muestras de voz se acumulan en un buffer antes de ser liberadas.

Como regla general, hay que procurar lograr una demora de generación de paquetes de 30 ms como máximo. En el router/gateways Cisco debe utilizar las cifras de la Tabla 5.2 según el tamaño de contenido configurado:

**Tabla 5 .2: Generación de Paquetes Común**

Codificador		Tamaño de carga útil	Retraso de empaquetado (ms)	Tamaño de carga útil (Bytes)	Retraso de empaquetado (ms)

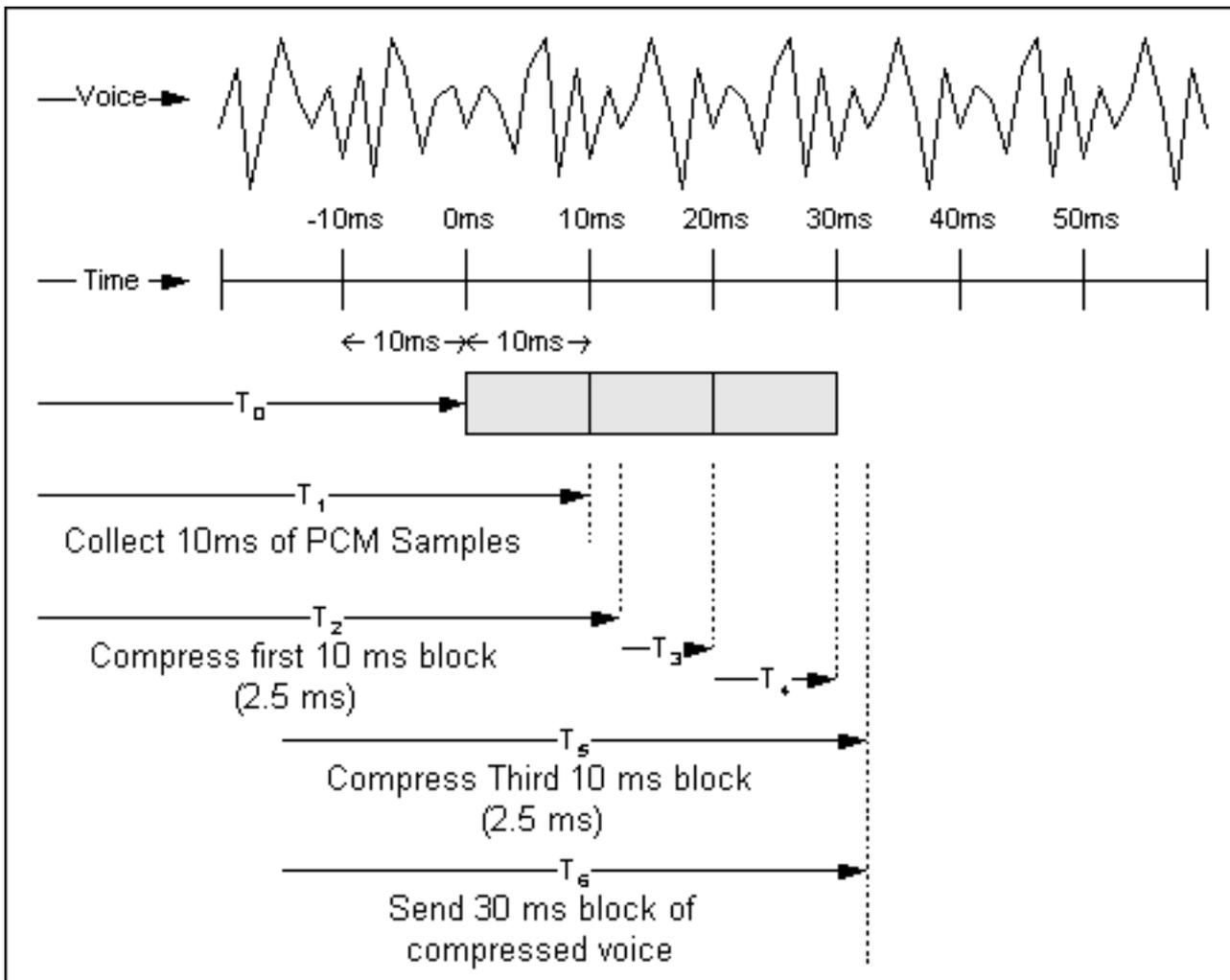
		(Bytes)			
PCM, G.711	64 Kb ps	160	20	240	30
ADPC M, G.726	32 Kb ps	80	20	120	30
CS- ACELP , G.729	8, 0 Kb ps	20	20	30	30
MP- MLQ, G.723. 1	6, 3 Kb ps	24	24	60	48
MP- ACELP , G.723. 1	5, 3 Kb ps	20	30	60	60

Tiene que mantener el equilibrio entre el retraso de empaquetado y la carga de la CPU. Cuanto menor sea el retraso, mayor será la velocidad de la trama y mayor la carga en la CPU. En algunas plataformas más antiguas, los contenidos de 20 ms pueden sobrecargar el CPU principal.

### [Demora de Canalización en el Proceso de Generación de Paquetes](#)

Aunque cada muestra de voz sufre demora algorítmica y demora de generación de paquetes, en realidad los procesos se superponen y esta canalización tiene un efecto neto beneficioso. Considere el ejemplo que se muestra en la Figura 2-1.

**Figura 5-2: Canalización y paquetización**



La línea superior de la figura representa una forma de onda de voz de muestra. La segunda línea es una escala de tiempo en incrementos de 10 ms. En  $T_0$ , el algoritmo CS-ACELP comienza a recolectar muestras PCM del códec. En  $T_1$ , el algoritmo ha recolectado su primer bloque de muestras de 10 ms y comienza a comprimirlo. En  $T_2$ , se comprimió el primer bloque de muestras. En este ejemplo, el tiempo de compresión es de 2,5 ms, como se indica en  $T_2 - T_1$ .

El segundo y el tercer bloque se recolectan en  $T_3$  y  $T_4$ . El tercer bloque se comprime en  $T_5$ . El paquete se ensambla y se envía (se supone que es instantáneo) en  $T_6$ . Debido a la naturaleza de línea de los procesos de compresión y empaquetado, la demora desde el inicio del proceso al envío de la trama de voz es  $T_6 - T_0$ , o aproximadamente 32.5 milésimas de segundo.

Para el ejemplo, este ejemplo se basa en la demora del mejor caso posible. Si se utiliza la demora del peor caso posible, la cifra es de 40 ms, 10 ms de demora del codificador y 30 ms de demora de generación de paquetes.

Observe que en estos ejemplos no se incluye la demora algorítmica.

## [Demora de serialización](#)

El retardo de serialización ( $\sigma_n$ ) es el retardo fijo necesario para conectar una trama de voz o datos a la interfaz de red. Está directamente relacionada con la velocidad de reloj en el trunk. Para velocidades de reloj bajas y tamaño de trama pequeños, el indicador adicional necesario para separar las tramas es significativo.

En la Tabla 5.3 se muestra la demora de serialización requerida para diversos tamaños de trama

y a distintas velocidades de línea. En esta tabla se usa el tamaño total de trama, no el tamaño de contenido, para el cálculo.

**Tabla 5.3: Demora de Serialización en Milisegundos para Diversos Tamaños de Trama**

Tamaño de trama (bytes)	Velocidad de línea (Kbps)										
	19.2	56	64	128	256	384	512	768	1024	1544	2048
38	15.83	5.43	4.75	2.38	1.19	0.79	0.59	0.40	0.30	0.20	0.15
48	20.00	6.86	6.00	3.00	1.50	1.00	0.75	0.50	0.38	0.25	0.19
64	26.67	9.14	8.00	4.00	2.00	1.33	1.00	0.67	0.50	0.33	0.25
128	53.33	18.29	16.00	8.00	4.00	2.67	2.00	1.33	1.00	0.66	0.50
256	106.67	36.57	32.00	16.00	8.00	5.33	4.00	2.67	2.00	1.33	1.00
512	213.33	73.14	64.00	32.00	16.00	10.67	8.00	5.33	4.00	2.66	2.00
1024	426.67	146.29	128.00	64.00	32.00	21.33	16.00	10.67	8.00	5.33	4.00
1500	625.00	214.29	187.50	93.75	46.88	31.25	23.44	15.63	11.72	7.77	5.86
2048	853.33	292.57	256.00	128.00	64.00	42.67	32.00	21.33	16.00	10.66	8.00

En la tabla, en una línea de 64 Kbps, una trama de voz CS-ACELP con una longitud de 38 bytes (37+1 indicador) tiene una demora de serialización de 4,75 ms.

**Nota:** El retardo de serialización para una celda ATM de 53 bytes (T1: 0,275ms, E1: 0,207 ms) es insignificante debido a la elevada velocidad de línea y al reducido tamaño de célula.

### [Retardo de colocación en cola/almacenamiento en memoria intermedia](#)

Una vez creada la carga útil de voz comprimida, se agrega un encabezado y la trama se coloca en cola para su transmisión en la conexión de la red. La voz debe tener prioridad absoluta en el router/gateway. Por lo tanto, una trama de voz debe esperar solamente una trama de datos que ya se reproduce u otras tramas de voz delante de ella. Esencialmente, la trama de voz espera la demora de serialización de cualquier trama anterior en la cola de salida. El retardo de colocación en cola ( $\beta_n$ ) es un retardo variable y es dependiente de la velocidad troncal y del estado de la cola. Hay elementos aleatorios asociados a la demora de envío a cola.

Por ejemplo, suponga que está en una línea de 64 Kbps y que se le envía a la cola detrás de una trama de datos (de 48 bytes) y de una trama de voz (42 bytes). Dada la naturaleza aleatoria de la cantidad de trama de 48 bytes que se ha reproducido, puede suponer con seguridad que, en

promedio, se ha reproducido la mitad de la trama de datos. De acuerdo con los datos de la tabla de serialización, el componente de la trama de datos es de  $6 \text{ ms} * 0,5 \text{ ms} = 3 \text{ ms}$ . Si suma el tiempo de otra trama de voz a continuación en la cola (5,25 ms), el tiempo total de demora de envío a cola es de 8,25 ms.

El modo de caracterizar el retraso en la cola depende del ingeniero de red. En general, al diseñar hay que pensar en el peor escenario posible y ajustar el rendimiento después de instalar la red. Cuantas más líneas de voz estén disponibles para los usuarios, mayor será la probabilidad de que un paquete de voz espere en la cola. A causa de la estructura de prioridades, la trama de voz nunca espera detrás de más de una trama de datos.

## Retardo de Switching de la Red

El Frame Relay público o la red ATM que interconecta las ubicaciones de los extremos es la fuente de las demoras de mayor duración para las conexiones de voz. Los retrasos de switching de red ( $w_n$ ) también son los más difíciles de cuantificar.

Si el equipo Cisco proporciona conectividad de área ancha u otro tipo de red privada, es posible identificar los componentes específicos de retraso. En general, los componentes fijos son de demoras de propagación en los trunks dentro de la red, y las demoras variables son de demoras de envío a cola que temporizan la entrada y salida de tramas en los switches intermedios. Para estimar la demora de propagación se suele utilizar una estimación popular de 10 microsegundos/milla o 6 microseconds/km (G.114). Sin embargo, el equipo de multiplexión intermedio, la red de retroceso, los links de microondas, y otros factores que se encuentran en las redes portadoras crean muchas excepciones.

El otro componente importante del retraso surge de la colocación en cola dentro de la red de área ancha. En una red privada, puede ser posible medir las demoras de envío a cola existentes o estimar el consumo por salto dentro de la red de área extensa.

Las demoras de portador típicas para las conexiones de Frame Relay en E.E.U.U. son fija de 40 ms y variable de 25 ms, por lo que la demora total en el peor caso posible es de 65 ms. Para simplificar, en los ejemplos 6-1, 6-2 y 6-3, se incluye cualquier demora de serialización a baja velocidad en la demora fija de 40 ms.

Éstas son figuras publicadas por los portadores de Frame Relay E.E.U.U., para cubrir dondequiera dondequiera a la cobertura dentro de los Estados Unidos. Cabe esperar que dos ubicaciones geográficamente más cercanas que las del peor caso posible experimenten una demora menor, pero normalmente los portadores documentan solamente el peor caso posible.

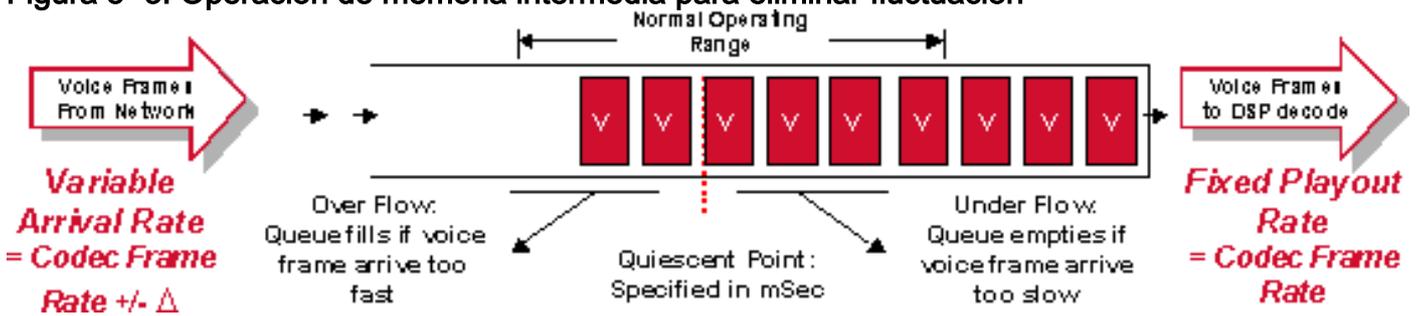
A veces los portadores de Frame Relay ofrecen servicios de gama superior. Estos servicios son generalmente para tráfico de voz o Systems Network Architecture (SNA), en el que la demora de red está garantizada y es inferior a la del servicio estándar. Por ejemplo, una portadora ascendente recientemente anunció tal servicio con un límite de retraso general de 50 ms, en lugar de los 65 ms estándares del servicio.

## Eliminar fluctuación de la demora

Dado que la voz es un servicio con velocidad de bit constante, se debe eliminar la fluctuación de todos los retrasos variables antes de que la señal salga de la red. En los routers/gateways Cisco esto se logra con un búfer de eliminación de fluctuación ( $\Delta_n$ ) en el router/gateway de extremo lejano (receptor). El buffer anti-fluctuaciones transforma la demora variable en una demora fija.

Retiene la primera muestra recibida durante un período de tiempo antes de reproducirla. Este período de espera se conoce como retardo inicial de reproducción.

Figura 5- 3: Operación de memoria intermedia para eliminar fluctuación



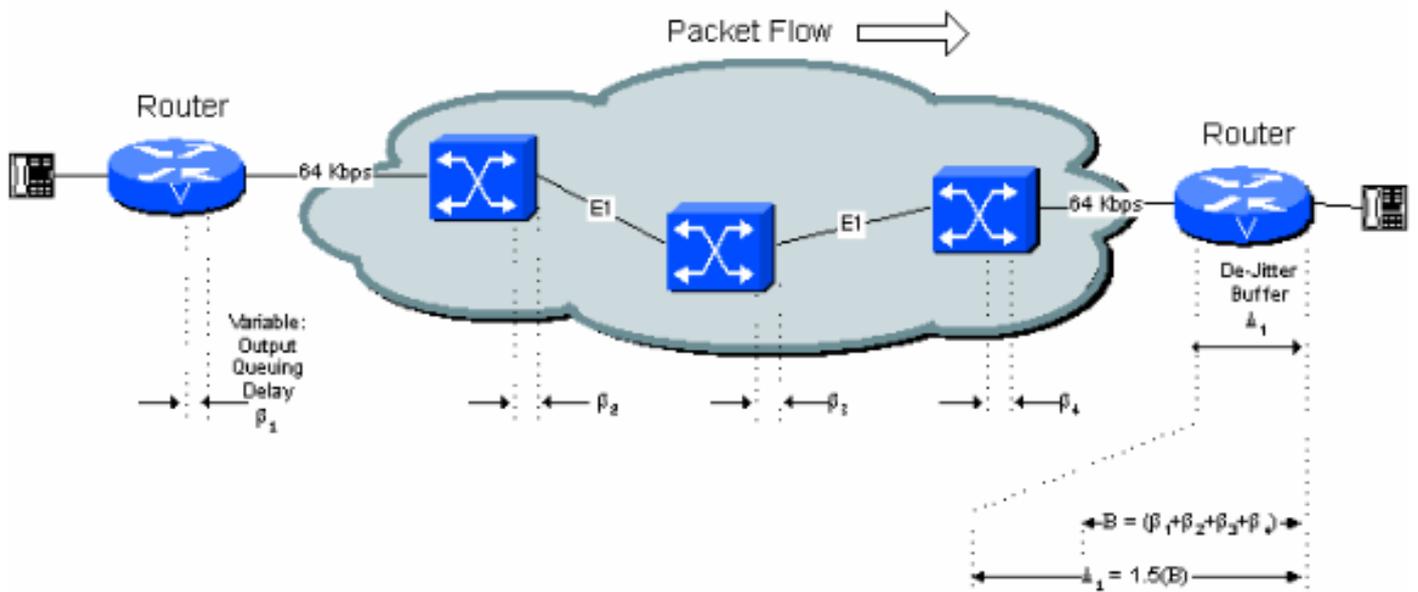
Es fundamental controlar correctamente el buffer anti-fluctuaciones. Si las muestras se retienen durante un tiempo demasiado corto, las variaciones en la demora pueden provocar un subdesbordamiento del buffer y crear interrupciones de voz. Si la muestra se retiene durante demasiado tiempo, el buffer puede desbordarse y los paquetes perdidos también pueden causar interrupciones de voz. Por último, si se retienen los paquetes demasiado tiempo, la demora global de la conexión puede alcanzar niveles inaceptables.

La demora de salida óptima de reproducción inicial para el buffer anti-fluctuaciones es igual a la demora variable total a lo largo de la conexión. Esto se ilustra en la Figura 5-4.

**Nota:** Las memorias intermedias de eliminación de fluctuación pueden ser adaptables, pero el retardo máximo se fija. Cuando se configuran buffers adaptativos, la demora se convierte en variable. Sin embargo, se puede usar la demora máxima como peor caso posible con fines de diseño.

Para obtener más información sobre los buffers adaptativos, refiérase a [Mejoras de la Demora de Reproducción para Voz sobre IP](#).

Figura 5-4: Demora variable y Buffer Anti-fluctuaciones



La demora de la reproducción inicial es configurable. La profundidad máxima del buffer antes de desbordarse suele establecerse en 1,5 ó 2,0 veces este valor.

Si se utiliza la configuración de demora nominal de 40 ms, la primera muestra de voz recibida

cuando el buffer anti-fluctuaciones está vacío se retiene durante 40 ms antes de reproducirse. Esto implica que el siguiente paquete que se reciba de la red puede retrasarse hasta 40 ms (con respecto al primer paquete) sin ninguna pérdida de continuidad de la voz. Si se demora más de 40 ms, el buffer anti-fluctuaciones se vacía y el siguiente paquete recibido se retiene durante 40 ms antes de reproducirse para restablecer el buffer. Esto provoca una interrupción en la voz reproducida durante unos 40 ms.

La contribución real del buffer anti-fluctuaciones a la demora es la demora de reproducción inicial más el tiempo real que se almacenó el primer paquete en la red. El peor caso posible es dos veces la demora inicial del buffer anti-fluctuaciones (se supone que el primer paquete que viaja por la red experimenta solamente una demora mínima de almacenamiento en buffer). En la práctica, tras un número de saltos de switch de red, probablemente no es necesario suponer el peor caso posible. Para los cálculos de los ejemplos del resto de este documento se aumenta la demora de reproducción inicial en un factor de 1,5 para tener en cuenta este efecto.

**Nota:** En el router/gateway receptor hay retraso a través de la función de descompresión. Sin embargo, esto se tiene en cuenta concentrándola con la demora de procesamiento de la compresión, como se describió anteriormente.

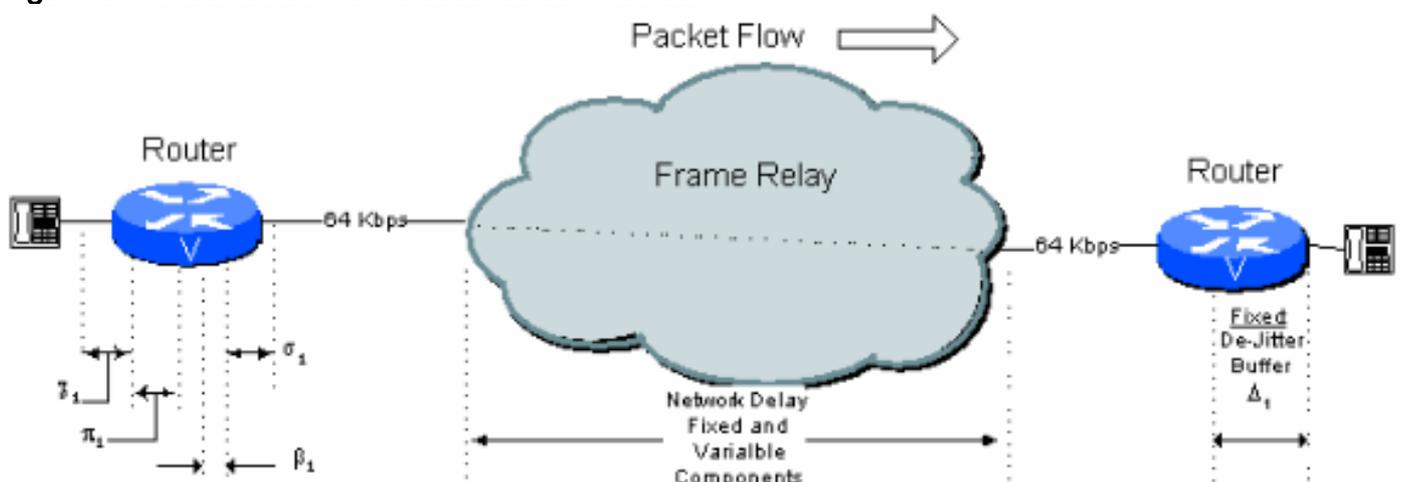
## Estimación de la Demora Prevista

El límite generalmente aceptado para una demora de conexión de voz de buena calidad es de 200 ms unidireccional (o 250 ms como máximo). Cuando las demoras exceden estos valores, se pierde la sincronización de hablantes y oyentes y, a menudo, ambos hablan al mismo tiempo, o ambos esperan que el otro hable. Esta condición se suele denominar superposición de interlocutores. Aunque la calidad global de voz es aceptable, a veces a los usuarios la naturaleza forzada de la conversación les resulta inaceptable. La superposición de interlocutores se puede observar en las llamadas telefónicas internacionales que viajan a través de conexiones vía satélite (la demora de satélite es de unos 500 ms, +/- 250 ms).

En estos ejemplos se ilustran diversas configuraciones de red y las demoras que el diseñador de red debe tener en cuenta.

### Conexión de un solo salto

Figura 6 - 1: Conexión de muestra de salto único



En esta figura se observa que una conexión típica de un salto sobre una conexión pública de

Frame Relay puede tener la demora estimada que se muestra en la Tabla 6.1.

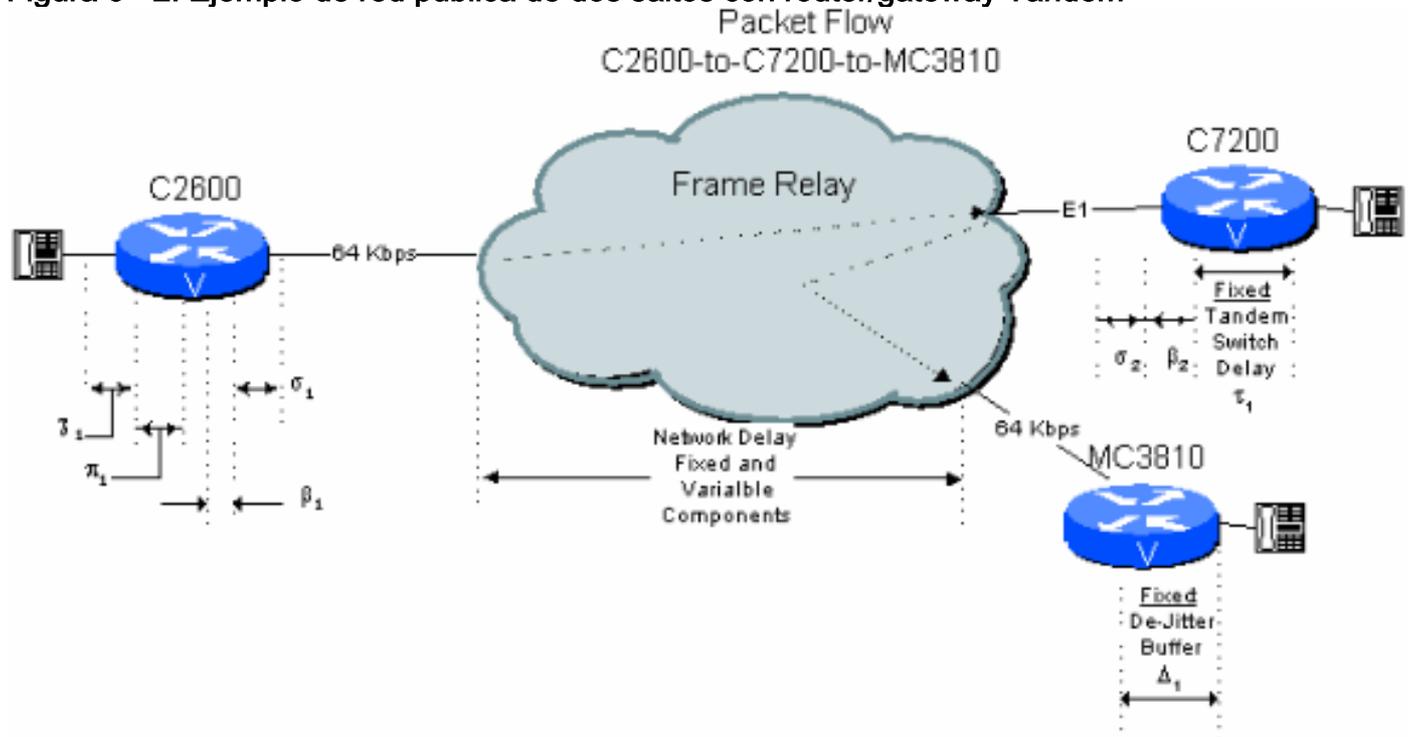
Tabla 6.1: Cálculo de la demora de un salto

Tipo de retraso	Fijo (ms)	Variable (ms)
Retraso del codificador, $\chi_1$	18	
Demora de empaquetado, $\pi_1$	30	
Almacenamiento en cola/Almacenamiento en búfer, $\beta_1$		8
Retraso de serialización (64 kbps), $\sigma_1$	5	
Retraso de red (marco público), $\tau_1$	40	25
Retraso del búfer de eliminación de fluctuación, $\Delta_1$	45	
Totales	138	33

**Nota:** Dado que el retardo de envío a cola y el componente variable del retardo de red ya se contabilizan dentro de los cálculos del búfer de eliminación de fluctuación, el retardo total es efectivamente sólo la suma de todo el Retraso fijo. En este caso la demora total es de 138 ms.

## Dos Saltos en una Red Pública con un C7200 que Actúa como Switch Tándem

Figura 6 - 2: Ejemplo de red pública de dos saltos con router/gateway Tandem



Ahora considere una conexión entre sucursales en una red con topología de estrella donde el C7200 de la sede central realiza la llamada a la sucursal de destino. En este caso la señal permanece en formato comprimido en el C7200 central. Esto permite un ahorro considerable en la estimación de demora con respecto al siguiente ejemplo, una conexión de dos saltos sobre una

red pública con un switch en tándem de PBX.

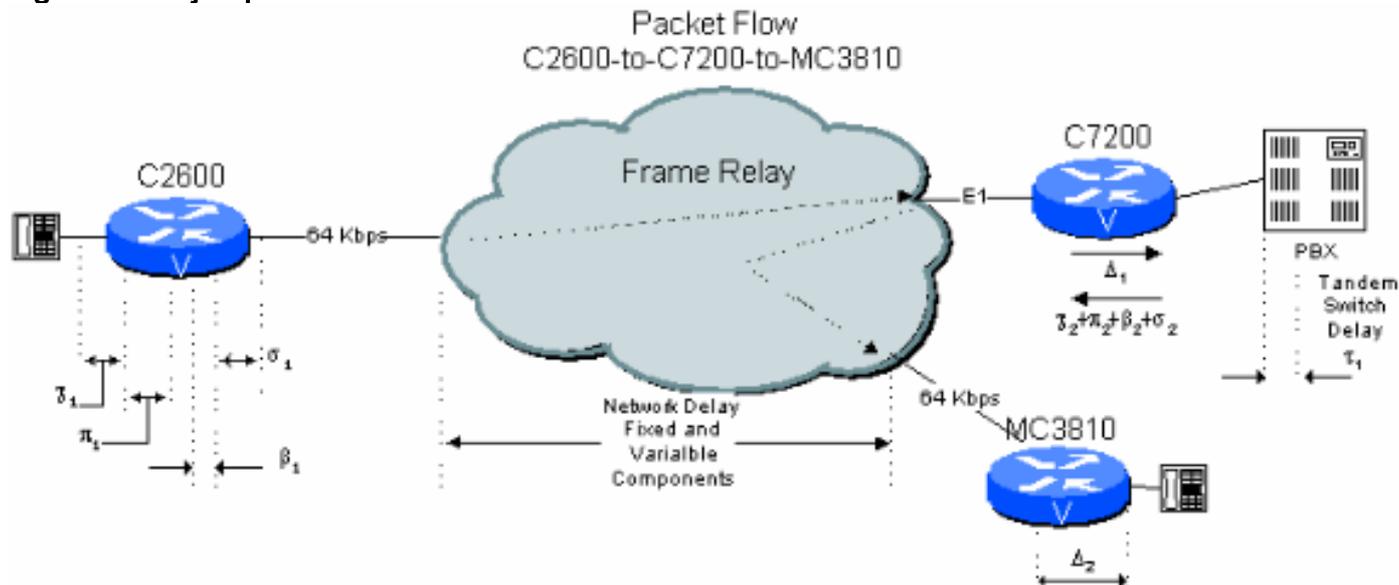
Tabla 6.2: Cálculo de la Demora de Red Pública de Dos Saltos con Tándem de Router/Gateway

Tipo de retraso	Fijo (ms)	Variable (ms)
Retraso del codificador, $\chi_1$	18	
Demora de empaquetado, $\pi_1$	30	
Almacenamiento en cola/Almacenamiento en búfer, $\beta_1$		8
Retraso de serialización (64 kbps), $\sigma_1$	5	
Retraso de red (marco público), $\delta_1$	40	25
Retraso en tándem en MC3810, $\tau_1$	1	
Almacenamiento en cola/Almacenamiento en búfer, $\beta_2$		0,2
Retraso de serialización (2 Mbps), $\sigma_2$	0,1	
Retraso de red (marco público), $\delta_2$	40	25
Retraso del búfer de eliminación de fluctuación, $\Delta_1$	75	
Totales	209.1	58.2

**Nota:** Dado que el retardo de envío a cola y el componente variable del retardo de red ya se contabilizan dentro de los cálculos del búfer de eliminación de fluctuación, el retardo total es efectivamente sólo la suma de todo el Retraso fijo. En este caso la demora total es de 209,1 ms.

### Conexión de dos saltos a través de una red pública con un switch Tandem PBX

Figura 6-3: Ejemplo de Red Pública de Dos saltos con Tándem PBX



Considere una conexión entre sucursales en una red entre sucursal y casa central en la cual el C7200 en el sitio de la casa central transfiere la conexión al PBX de la casa central para la conmutación. Aquí hay que descomprimir la señal de voz y eliminar las fluctuaciones. Después se vuelve a comprimir y se eliminan las fluctuaciones otra vez. Esto da lugar a demoras adicionales con respecto al ejemplo anterior. Además, los dos ciclos de compresión CS-ACELP reducen la calidad de la voz (ver Efectos de los ciclos de compresión múltiple).

**Tabla 6.3: Cálculo de la Demora de Red Pública de Dos Saltos con Tándem PBX**

Tipo de retraso	Fijo (ms)	Variable (ms)
Retraso del codificador, $\chi_1$	18	
Demora de empaquetado, $\pi_1$	30	
Almacenamiento en cola/Almacenamiento en búfer, $\beta_1$		8
Retraso de serialización (64 kbps), $\sigma_1$	5	
Retraso de red (marco público), $\tau_1$	40	25
Retraso del búfer de eliminación de fluctuación, $\Delta_1$		40
Retraso del codificador, $\chi_2$	15	
Retraso de empaquetado, $\pi_2$	30	
Almacenamiento en cola/Almacenamiento en búfer, $\beta_2$		0,1
Retraso de serialización (2 Mbps), $\sigma_2$	0,1	
Retraso de red (marco público), $\tau_2$	40	25
Retraso del búfer de eliminación de fluctuación, $\Delta_2$	40	
Totales	258.1	58.1

**Nota:** Dado que el retardo de envío a cola y el componente variable del retardo de red ya se contabilizan dentro de los cálculos del búfer de eliminación de fluctuación, el retardo total es efectivamente sólo la suma de todo el retardo fijo más el retardo del búfer de eliminación de fluctuación. En este caso la demora total es de 258,1 ms.

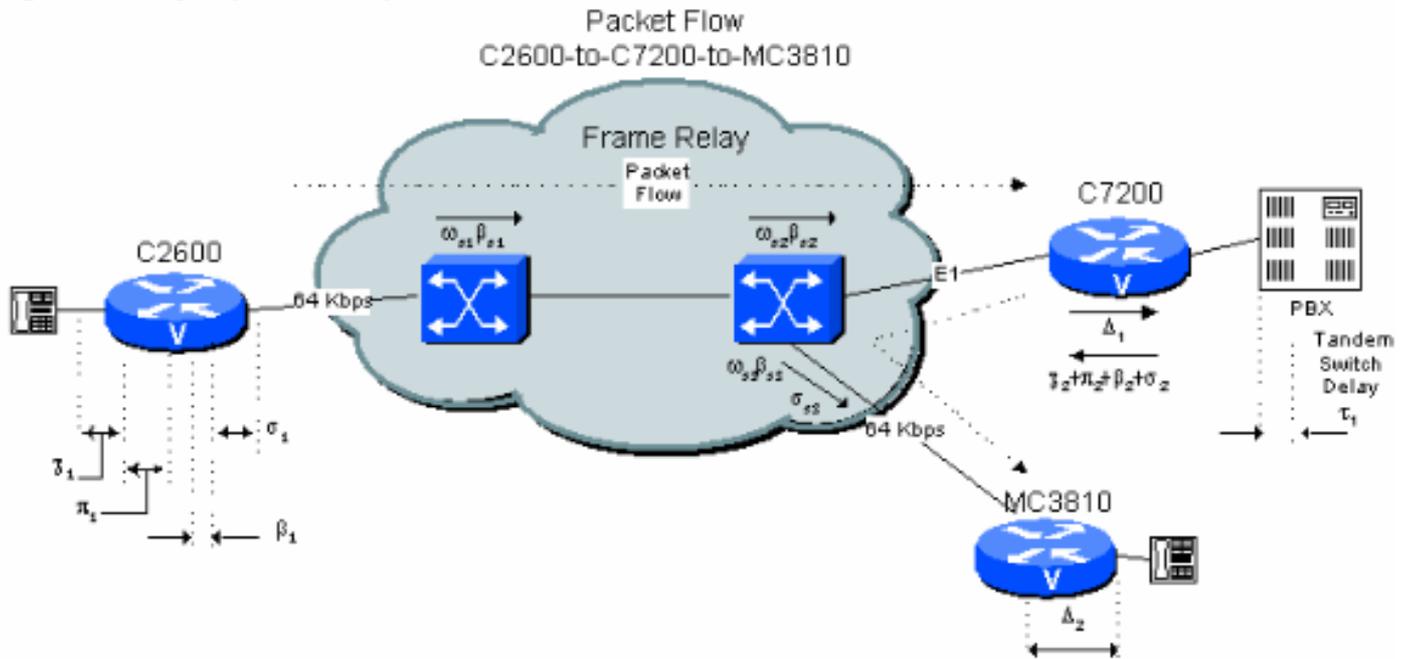
Si utiliza el PBX del sitio central como un switch, aumenta la demora de conexión unidireccional entre 206 ms y 255 ms. Esto se acerca a los límites de la ITU para la demora unidireccional. Este tipo de configuración de red requiere que el ingeniero preste atención al diseño para obtener una demora mínima.

Se supone el peor caso posible para una demora variable (aunque ambos tramos de la red pública no experimenta demoras máximas simultáneamente). Si hace suposiciones más optimistas para las demoras variables, solamente mejorará mínimamente la situación. Sin embargo, con una mejor información sobre las demoras fija y variable en la red Frame Relay del portador, se puede reducir la demora calculada. Se puede esperar que las conexiones locales

(por ejemplo, dentro de un estado) tengan mejores características de retraso, pero las portadoras son resistentes a ofrecer límites de retraso.

### Conexión de dos saltos en una red privada con un switch tándem PBX

Figura 6-4: Ejemplo de red privada de dos saltos con tándem PBX



El ejemplo 4.3 muestra que, con la suposición de las demoras correspondientes al peor caso posible, es muy difícil mantener la demora calculada por debajo de 200 ms cuando una conexión entre sucursales incluye un salto del tándem de PBX en el sitio central con las conexiones de red pública de Frame Relay en cada lado. Sin embargo, si se conocen la topología de red y el tráfico, es posible reducir de manera sustancial la cifra calculada. Esto se debe a que, normalmente, las cifras proporcionadas por los portadores se limitan al peor caso posible de demora de transmisión y envío a cola en un área extensa. Es mucho más fácil establecer límites más razonables en una red privada.

La figura generalmente aceptada para el retraso en la transmisión entre los switches es del orden de 10 microsegundos/milla. En función del equipo, la demora de switch de transición en una red Frame Relay debe ser del orden de 1 ms para demora fija y 5 ms para demora variable de envío a cola. Estas cifras dependen del equipo y el tráfico. Las cifras de demora para switches Cisco WAN MGX son de menos de 1 ms por switch en total si se utilizan trunks E1/T1. Si se supone una distancia de 500 millas, con demora fija de 1 ms y demora variable de 5 ms por cada salto, el cálculo de la demora es el siguiente:

Tabla 6.4: Cálculo de la Demora de Red Privada de Dos Saltos con Tándem PBX

Tipo de retraso	Fijo (ms)	Variable (ms)
Retraso del codificador, $\chi_1$	18	
Demora de empaquetado, $\pi_1$	30	
Almacenamiento en cola/Almacenamiento en búfer, $\beta_1$		8
Retraso de serialización (64 kbps),	5	

$\sigma_1$		
Retraso de Red (Trama Privada), $S_1 + \beta_1 + S_2 + \beta_2$	2	10
Retraso del búfer de eliminación de fluctuación, $\Delta_1$	40	
Retraso del codificador, $\chi_2$	15	
Retraso de empaquetado, $\pi_2$	30	
Almacenamiento en cola/Almacenamiento en búfer, $\beta_2$		0,1
Retraso de serialización (2 Mbps), $\sigma_2$	0,1	
Retraso de red (trama privada), $S_3$ $+ \beta_3$	1	8
Retraso de serialización (64 kbps), $\sigma_3$	5	
Retraso del búfer de eliminación de fluctuación, $\Delta_2$	40	
Retraso por distancia/transmisión (no interrumpido)	5	
Totales	191.1	26.1

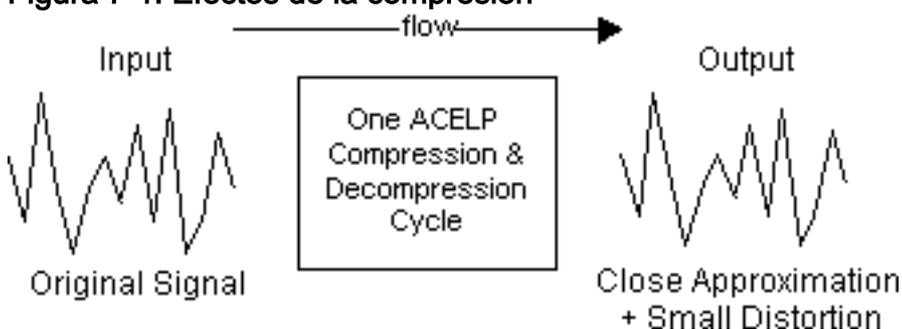
**Nota:** Dado que el retardo de envío a cola y el componente variable del retardo de red ya se contabilizan dentro de los cálculos del búfer de eliminación de fluctuación, el retardo total es sólo la suma de todos los retrasos fijos. En este caso la demora total es de 191,1 ms.

Cuando se funciona sobre una red de Frame Relay privada, es posible hacer una conexión del radio a radio a través del PBX en el sitio del hub y mantenerse en la cifra de 200 ms.

## Efectos de los ciclos de compresión múltiple

Los algoritmos de compresión CS-ACELP no son deterministas. Esto significa que la secuencia de datos de entrada no es exactamente la misma que la secuencia de datos de la salida. Se introducirá una pequeña cantidad de distorsión en cada ciclo de compresión, como se indica en la Figura 7-1.

**Figura 7-1: Efectos de la compresión**



Por lo tanto, múltiples ciclos de compresión CS-ACELP introducen rápidamente niveles

importantes de distorsión. Este efecto adicional de distorsión no es tan pronunciado con algoritmos de modulación de código de impulso diferencial adaptable (ADPCM).

El impacto de esta característica es que, además, de los efectos de la demora, el diseñador de la red debe tener en cuenta el número de ciclos de compresión CS-ACELP en la trayectoria.

La calidad de voz es subjetiva. La mayoría de los usuarios consideran que después de dos ciclos de compresión la calidad de voz sigue siendo adecuada. Un tercer ciclo de compresión suele provocar una degradación significativa, que puede ser inaceptable para algunos usuarios. Como regla general, el diseñador de la red debe limitar a dos el número de ciclos de compresión CS-ACELP en una trayectoria. Si fuera necesario utilizar más ciclos, hay que notificárselo primero al cliente.

En los ejemplos anteriores se muestra que cuando una conexión entre sucursales es conmutada en tándem a través del PBX (con formato PCM) de la sede central, experimenta una demora considerablemente mayor que si fuera conmutada en tándem en el C7200 de la sede central. Está claro que cuando se utiliza el PBX para conmutar, hay dos ciclos de compresión CS-ACELP en la trayectoria, en lugar de un ciclo, como cuando se conmutan las tramas de voz en el C7200 central. La calidad de voz es mejor con el ejemplo conmutado en C7200 (4.2), aunque pueda haber otras razones, como la administración de planes de llamadas, que puede requerir incluir el PBX en la trayectoria.

Si una conexión entre sucursales se hace con un PBX central, y de la segunda sucursal se extiende la llamada sobre la red de voz pública y después termina en una red de teléfono celular, hay tres ciclos de compresión CS-ACELP en la trayectoria, así como una demora considerablemente más alta. En este escenario, la calidad se ve afectada de forma significativa. En este caso el diseñador de la red también debe considerar la trayectoria en el peor caso posible y decidir si es aceptable teniendo en cuenta la red de usuarios, las expectativas y los requisitos comerciales.

## [Consideraciones para conexiones con retardo prolongado](#)

Es relativamente fácil diseñar las redes de voz por paquetes que superen el límite de demora unidireccional de 150 ms generalmente aceptado por la ITU.

Al diseñar redes de voz por paquetes, el ingeniero debe considerar con qué frecuencia se utilizará la conexión, cuáles serán las necesidades de los usuarios, y qué tipo de actividad comercial está implicado. No es infrecuente que tales conexiones sean aceptables en circunstancias concretas.

Si las conexiones de Frame Relay no atraviesan una distancia grande, es muy probable que el comportamiento de la red en cuanto a la demora sea mejor que el mostrado en los ejemplos.

Si la demora total experimentada por las conexiones de router/gateway en tándem aumenta demasiado, una alternativa común es configurar circuitos virtuales permanentes (PVC) adicionales directamente entre los MC3810 terminales. Esto añade un costo recurrente a la red, ya que los portadores suelen cobrar por PVC, pero en algunos casos puede ser necesario.

## [Información Relacionada](#)

- [Unión Internacional de Telecomunicaciones \(ITU\)](#)
- [Soporte de tecnología de voz](#)

- [Soporte de Productos de Voice and Unified Communications](#)
- [Troubleshooting de Cisco IP Telephony](#)
- [Soporte Técnico - Cisco Systems](#)