

## Comprensión de la latencia End-to-end para aplicaciones de vídeo en red



### ¿Por qué preocuparse?

La latencia ha sido una preocupación de los diseñadores de sistemas desde la primera comunicación electrónica - la invención del teléfono. Hoy en día, es necesario evaluar los efectos de la latencia en cualquier comunicación de bucle cerrado. Para el vídeo, la latencia es sin duda un factor clave en el desempeño de los sistemas de videoconferencia y telepresencia. Para los diseñadores de aplicaciones, se ha establecido como latencia aceptable de transmisión de extremo a extremo (de la cámara a la pantalla) por debajo de 200 milisegundos (2 décimas de segundo). Por encima de 200 milisegundos la atención de los participantes se desvía y experimentan fatiga después de un corto periodo de tiempo. Para otras aplicaciones en las que las acciones del espectador a distancia tienen algún impacto (como la toma de decisiones en misiones críticas, en comunicaciones de vídeo, robótica y control remoto), o los diseñadores simplemente desean que el vídeo tenga mínima latencia para reducir al mínimo los errores imputables a la comunicación, se han determinado diferentes métricas. A veces, se requiere una comprensión de la latencia, porque se necesita realizar una entrega sincronizada de múltiples streams independientes, como el vídeo estéreo o varios displays que muestren la misma información de audio y vídeo. Este documento examina los elementos que contribuyen a la latencia extremo a extremo en redes de vídeo.

### ¿Cómo de rápido es rápido?

El rendimiento por encima de métricas de 200 milisegundos en las comunicaciones visuales es de particular interés. Fue un punto de referencia establecido por los clientes de HaiVision para el servicio de vídeo en red, enfocado a la educación a distancia (como Verizon y SBC). Si la entrega de vídeo extremo a extremo entre las clases se realiza en un tiempo superior a 200 milisegundos, se determina que los esfuerzos para mantener la atención se ven gravemente afectados. Teniendo en cuenta este punto de referencia y el hecho de que casi todas las redes WAN presentan una latencia de transmisión del orden de 35 a 40 milisegundos, al principio la tecnología Haivision fue diseñada para lograr resultados consistentes 150-160 milisegundos - con lo que la latencia acumulada de codificar-enviar-decodificar quedase cerca de 200 milisegundos en el peor de los casos. Hay que tener en cuenta que la palabra consistente en términos de rendimiento determinista es crítica para muchos diseños de sistemas. Haivision se involucró en la cirugía a distancia (telecirugía) desde el principio, y su tecnología fue seleccionada debido a que entrega con la menor latencia. Algunos de los experimentos midieron la eficacia del funcionamiento de un robot cirujano, controlado desde una ubicación remota por vídeo en red, como guía de referencia. Tenga en cuenta que el cirujano se basa en la coordinación ojo-mano, con la mano y los ojos separados por la latencia. A medida que la latencia crece, el cirujano debe trabajar con más cautela, y a un ritmo mucho más lento. Sorprendentemente, aunque éste es un caso donde siempre es mejor una menor latencia, la efectividad del cirujano a partir de un punto empezó a disminuir de forma exponencial, 200 milisegundos. Otro ejemplo son los eventos deportivos, donde los aficionados estarán molestos si las pantallas van más de 200 milisegundos retrasadas con respecto al sonido del campo.



### Elementos que contribuyen

Hay una serie de componentes que influyen en la ecuación total de la latencia de vídeo. Estos componentes incluyen: la latencia de la cámara, la latencia del codificador, la topología de la red y distancia de transmisión, la latencia del decodificador y la latencia de la pantalla. Todos estos factores deben ser considerados cuando se trata de minimizar la latencia total del sistema.

### Cámaras

Las cámaras de alta definición con una resolución de 720p y 60 frames por segundo suelen añadir una latencia total de menos de 5 ms. En cámaras de alta definición a menudo se entrelaza el video (1080i por ejemplo),



**Videoconferencia**



**Broadcast**



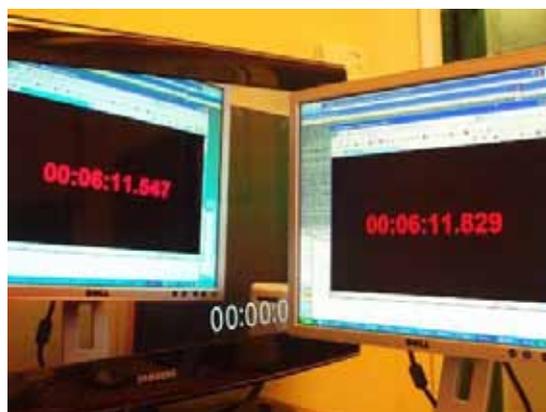
**Médicas**

pero esto no aumenta de manera significativa la latencia total. Cuando si puede aumentar la latencia total de las cámaras es cuando se utiliza firewire o una salida similar que realiza un poco de compresión de vídeo. Típicamente, sin embargo, la propia cámara es sólo un pequeño componente en la ecuación de latencia total. Tenga en cuenta que algunas cámaras de consumo pueden añadir una latencia visible.

### Codificadores

Los codificadores y las redes de larga distancia suelen ser los principales contribuyentes a la latencia de extremo a extremo. En la industria de hoy en día existen diferentes codificadores y la selección del correcto puede reducir considerablemente la latencia total de extremo a extremo de la solución. Los codificadores que se utilizan normalmente en el mundo de la televisión, hoy por hoy, suman entre 300 y 4000 ms de latencia al vídeo (¡más de cuatro segundos!). Estos codificadores están diseñados para proporcionar la más alta calidad de vídeo posible, y así toman tiempo adicional para completar sus algoritmos de compresión multi-paso. Por este motivo, dichos codificadores generalmente no se consideran soluciones donde la latencia es un factor importante. Además, los codificadores de difusión tienden a ser la opción más costosa.

Hoy en día, algunas empresas eligen códecs de videoconferencia (codificador / decodificador de sistemas) o codificadores de vídeo de tipo profesional- ambos con una latencia típica de entre 180 y 400 ms. Representan una mejora con respecto a los codificadores de radiodifusión tradicionales, sin embargo, estas opciones tienen inconvenientes importantes. Los codificadores de videoconferencia están diseñados para: 1) Obtener un bajo costo, 2) La utilización a través de cualquier conexión de red, y 3) La interoperabilidad con cualquier sistema de conferencias. Con el fin de alcanzar dichos objetivos se sacrifica la calidad de vídeo adaptándose al “mínimo común denominador.” Los códecs de videoconferencia también están limitados en cuanto a la interoperabilidad con la infraestructura de transmisión de elementos comunes como: reproductores software, decodificadores y servidores de streaming. La mayoría de los codificadores de tipo profesional suelen ser compatibles con soluciones streaming estándar, pero no están típicamente diseñados para funcionamiento 24 / 7 o la entrega de vídeo compatible.



Haivision aplica la tecnología de compresión H.264 de alto rendimiento en el Mako, Makito (HD), y Barracuda (SD). Los codificadores no sólo brindan la más baja latencia total de codificación, siempre entre los 50 y 55 ms, sino que también proporcionan una calidad de imagen muy alta e interoperan con soluciones de streaming estándar. Estos codificadores también se benefician de la herencia de las telecomunicaciones, los dispositivos HaiVision se diseñan para uso continuado 24 / 7. Es importante entender algunos de los procesos involucrados durante la codificación del audio y vídeo, ya que pueden mejorar el rendimiento general del sistema, pero también pueden aumentar la latencia interna del codificador.



Mako



Makito y Barracuda

Estas características, que son seleccionables por el usuario en los codificadores HaiVision (a veces en función del stream), incluyen:

- Desentrelazado que puede añadir latencia 10-15 ms, y es necesaria cuando se trabaja con 1080i o fuentes 480i/576i y se especifica la compatibilidad con reproductores software o Flash.
- Escalado (si fuese necesario) -que puede añadir hasta 5 ms.
- Cifrado (para el vídeo de seguridad) que puede agregar 20 a 30 ms de latencia de codificación.
- Configuración del tráfico, para conseguir un flujo constante de velocidad de bits, que añade 30 a 100 ms de latencia.
- Corrección de errores que puede añadir de 10 a 30 ms de latencia, pero permite que el sistema pueda manejar hasta un 5% de pérdida de paquetes.

Los números presentados son sólo de referencia y pueden variar dependiendo de la complejidad del stream, el ancho de banda aplicada, y la resolución de vídeo. Los codificadores Haivision se caracterizan por ofrecer un nivel de rendimiento de base, independientemente de estos factores (es decir, codifica con 55 ms, independientemente del ancho de banda o la resolución seleccionados), pero las latencias adicionales mencionadas anteriormente pueden variar.

### Topología de red

Otro factor, en la ecuación de la latencia, es la distancia de la red de datos y la topología de la misma que se utiliza para entregar el vídeo. La red introduce latencia en los datos de video cada vez que debe pasar por un router (saltos). Además, la variación en el path de acceso y la calidad del enrutamiento pueden agregar pérdida de paquetes y jitter a la ecuación de latencia. Por último, la distancia de transmisión puede añadir latencia significativa.

En una red pequeña (en la que el codificador y decodificador están separados por un sólo switch) el jitter, la pérdida de paquetes y la latencia de la transmisión serán efectivamente cero. En redes de tamaño considerable, debe asegurarse que la red está bien dimensionada. Además, una LAN compleja debe estar protegida contra la pérdida de paquetes moderada.

En la entrega de vídeo a distancia se debe tener en cuenta, tanto la pérdida de paquetes, como el jitter, para comprender la latencia de la transmisión.

La pérdida de paquetes puede ser tratada de tres formas, en primer lugar, simplemente se ignoran, algo que no aumenta la latencia, pero disminuye la calidad del vídeo emitido. Una pérdida pequeña de paquetes (<0,1%) puede presentarse como muescas en el vídeo. En segundo lugar, se puede considerar la utilización de una solución que maneje la pérdida de paquetes volviendo a enviar los paquetes perdidos (como TCP / IP) y un almacenamiento en búfer con el fin de mantener la calidad, pero este método puede aumentar significativamente la latencia total de la solución. Esto es típico del vídeo en Internet mediante descarga progresiva - la mayor parte de lo que se ve a través de Internet. En tercer lugar, se puede acomodar la pérdida de paquetes mediante la aplicación de corrección de errores. FEC calcula e inserta la información redundante en el stream para que pueda ser reconstruido si no llegan algunos de los paquetes. Esto añade una pequeña cantidad de latencia y ancho de banda, pero ofrece una mejora de rendimiento espectacular a través de redes congestionadas o mal diseñadas.

Jitter es la variabilidad en el tiempo de entrega de los paquetes que fuerza al decodificador a recoger los paquetes y volver a pedirlos antes de hacer su función. El jitter puede tener un efecto exponencial en la latencia y puede convertirse en un problema importante si no se toma en consideración. Normalmente en una red bien dimensionada no es una consideración importante y por lo general añade menos de 5 ms de latencia a la ecuación. En una red con un alto jitter sin embargo, el aumento de la latencia puede ser de 300 ms o más.

También hay elementos intermedios ubicados en la red que también pueden tener un impacto en la latencia. Estos dispositivos incluyen servidores de streaming que replican y a veces vuelven a empaquetar el stream de vídeo y así puedan llegar a añadir hasta 50 ms de latencia. Además, los dispositivos, tales como Transmux VF, pueden interceptar los flujos y aplicar funciones tales como la corrección de errores y el cifrado de los streams, que potencialmente pueden llegar a añadir hasta 50 ms de latencia si la manipulación y adaptación del stream es necesaria.

La distancia física entre los puntos finales también debe ser considerada. A grandes distancias la limitación de la velocidad de la luz se convierte en un factor importante a considerar. Por cada 300 kilómetros que el vídeo debe viajar añade un milisegundo de latencia en el sistema. Esto significa que la conexión puede tener más de 67 ms de latencia añadida, simplemente por el tiempo necesario para viajar entre un punto y sus antípodas.

En el párrafo anterior se define que una red estatal puede contribuir con 35-40 ms de latencia, se podría suponer: 15 ms de latencia de transmisión, 5 ms para dar cabida a las cuestiones relacionadas con el jitter, y hasta 20 ms debido a los equipos (router) correspondientes.

## Decodificadores

Hay muchas maneras diferentes para decodificar vídeo en red una vez ha llegado a su destino 1) decodificador hardware dedicado, 2) en un set top box standard, o 3) mediante un cliente de software para visualizar el vídeo en un PC.

Decodificadores hardware dedicados ofrecen la menor latencia posible. El decodificador Mako de Haivision, muy enfocado al motor de codificación, puede recibir paquetes de vídeo y entregarlos en banda base en menos de 10 ms y el decodificador Makito en 70ms, siempre que no tenga que esperar a la llegada de paquetes, debido al jitter. Los decodificadores hardware se han diseñado desde el punto de vista del rendimiento para comunicaciones punto a punto, como la telepresencia.

Decodificador	Retraso	Nota
Mako	~ 10 ms	Más rápido decodificador H.264
Makito decodificador	~ 15 ms	Decodificador H.264 de alta definición
Sencore decodificador	530 + ms	Broadcast - salida HD-SDI

Los Set Top Boxes fueron diseñados inicialmente para dar cabida al cable digital y satélite para uso doméstico. Como tales, pueden proporcionar interfaces de usuario sencillas y están disponibles a un costo muy bajo. Estos son efectivamente decodificadores MPEG base. La mayoría de los vendedores establecidos han ampliado su tecnología para proporcionar interfaces IP haciendo estos dispositivos ideales para despliegues empresariales, del estilo de IPTV. Muchos diseñadores están diseñando STB de bajo coste para sustituir a reproductores digital signage con el fin de distribuir la información de la señalización digital en todo el campus, debido a las dificultades económicas actuales. Dado que los STB se basan en la tecnología comúnmente disponible de bajo costo ASIC y trabajan con una amplia variedad de formatos, son de rendimiento más bajo que un hardware dedicado. Tenga en cuenta que esto es sólo en lo que se refiere a la latencia, la calidad de imagen es la misma.

Decodificador	Retraso	Nota
Stingray STB	180 + ms	Depende del formato, pero es consistente
Enseo STB	430 + ms	Depende del formato y contenido
Amino STB	730 + ms	Por lo general el rendimiento es menos consistente

Los decodificadores software presentan el coste final más bajo y la ventaja de la accesibilidad. Por lo general son gratuitos y permiten ver vídeo en directo en un PC. Eliminan la necesidad de comprar un hardware para ver el video, y a menudo proporcionan una interfaz gráfica de usuario con selección de canales y guía de programación electrónica. Además, por lo general permiten la personalización deseada por la empresa. La latencia de decodificación depende de la configuración de CPU/GPU/RAM del PC.

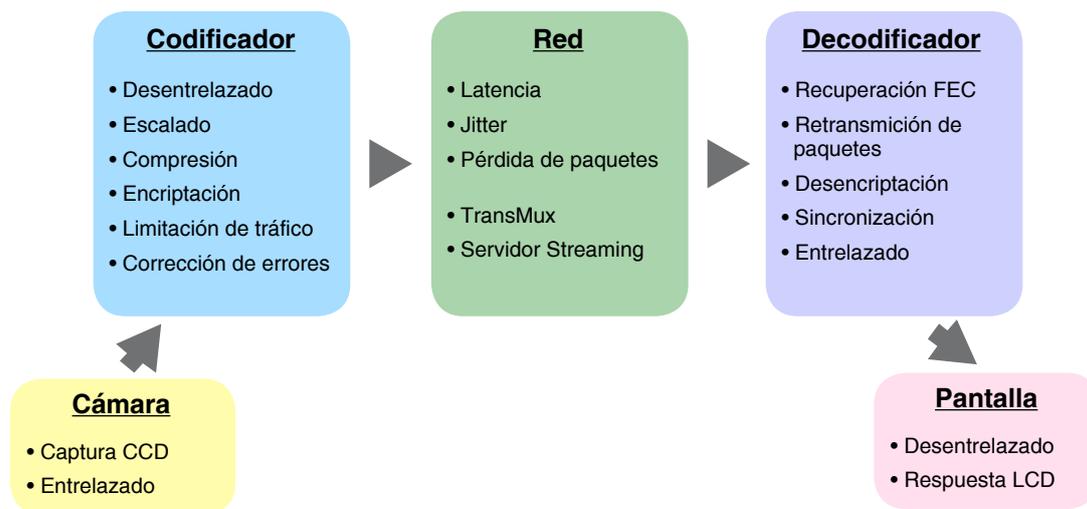
Decodificador	Retraso	Nota
haiPLAY	80 ms	Muy rápido, pero depende de la CPU
InStream	280 + ms	Depende de la CPU, pero puede ser sincronizado con los STB
VideoLAN VLC	350 ms	De código abierto
Wowza / Flash	630 ms	La latencia es para SD, el vídeo HD tiene una mayor latencia

Otros factores también pueden afectar a la latencia de decodificación. En el Mako Haivision, decodificador Makito, STB Stingray o un reproductor InStream, el uso de las siguientes características adicionales puede incrementar la latencia:

- Re-entrelazado añade aproximadamente 30 ms, pero es necesario para 1080i o vídeo 480i/576i con el fin de evitar los artefactos digitales.
- Escalado que puede añadir hasta 5 ms de latencia, aunque no suele ser un factor de peso en el STB.
- Descifrado añade aproximadamente 10 ms utilizando un algoritmo AES de 128 bits.
- El buffer para corregir el jitter puede agregar fácilmente 30 ms a 200 ms si este comienza a ser grande.
- Corrección FEC añade entre 10 y 30 ms de latencia.

## Dispositivos de visualización

Los dispositivos de pantalla a menudo agregan poco o nada de latencia significativa a la ecuación. Hay varias excepciones a ésta, sin embargo, los diseñadores deben tenerla en cuenta al considerar una solución. CRT estándar añade latencia imperceptible, pero tiene otras consecuencias negativas obvias. Las tecnologías de LCD y Plasma suelen añadir tan sólo 4 ms de latencia a la ecuación, pero se conoce que pueden añadir hasta 120 ms de retraso al mostrar entrelazado de alta definición, así pues, se debe tener cuidado en el momento de decidir qué dispositivo de salida utilizar como parte de la solución.



## Sincronización - Múltiples flujos o Muestra

Dos casos de uso están a la vanguardia cuando se habla de la sincronización de varias secuencias de vídeo. El primero es cuando varias pantallas muestran la misma información. En una implementación de señalización digital en varios monitores, de forma que se vea el mismo contenido en distintos monitores a la vez, se debe tener cuidado para seleccionar STBs que actúen en concordancia. Debido a que los STBs son sintonizados a la misma fuente y probablemente en el mismo segmento de red, la única variable que podría causar la falta de sincronización es la capacidad de múltiples STB para decodificar con la misma latencia. El STB Stingray está diseñado para proporcionar una operación consistente en múltiples unidades. Algunos STB de bajo coste presentan variaciones significativas en el rendimiento arruinando los objetivos del diseñador.

Otro desafío al que se enfrentan a veces es el requisito de sincronizar STB con reproductores software. Un gran ejemplo es la entrega de contenido de televisión para un entorno de comercio financiero. Diseñadores podrían dar cabida a pantallas de panel plano (por supuesto silenciado) con audio entregado al reproductor de escritorio. HaiVision que permiten la sincronización de este tipo entre el STB Stingray y el reproductor software InStream.

El tema final es la sincronización de múltiples secuencias en vivo. Esto es necesario en algunas implementaciones de telepresencia (sala de varias cámaras) y para la distribución de vídeo estéreo (como los utilizados en los robots quirúrgicos para la percepción de profundidad). Haivision no proporciona actualmente ninguna tecnología específica para dar cabida a estos casos de uso fuera de la coherencia absoluta determinista de nuestros codificadores. Dos cadenas de codificadores / decodificadores uno al lado del otro exhiben idénticas características de latencia. La única variable que hará que los streams no se sincronicen es la diferencias de tráfico en la red que deben recorrer ambos. Si ambos flujos están atravesando la misma red tal variación será mínima (<3.1 milisegundos).

## Los mejores escenarios extremo a extremo

Teniendo en cuenta todos los factores que contribuyen a la latencia, los diseñadores necesitan planificar en consecuencia si la latencia o la sincronización son factores de decisión. Al minimizar la red y tráfico añadido, considerando una LAN de alta calidad y cámaras y displays profesionales, las siguientes latencias extremo a extremo son posibles:

En-d to-End (de extremo a extremo)	Retraso / latencia
Mako a Mako	70 milisegundos
Makito a Makito Decoder	70 milisegundos
Makito a haiPLAY	150 milisegundos
Makito a Stingray	250 milisegundos
Makito a InStream	350 milisegundos
Makito a Enseio	500 milisegundos
Makito a Sencore	600 milisegundos
Makito a Wowza / Flash	700 + + milisegundos
Makito de Amino	800 milisegundos

Mediante el uso de los codificadores y decodificadores Haivision las empresas son capaces de cumplir con latencias de vídeo en red que, a día de hoy, son incomparables en el mercado. Estos componentes permiten a las empresas llevar a cabo reuniones de telepresencia sin la fatiga del usuario, ofrecer vídeo en directo en las pantallas de un estadio sin apenas retraso en la acción en vivo, y permiten a los cirujanos realizar telecirugía.