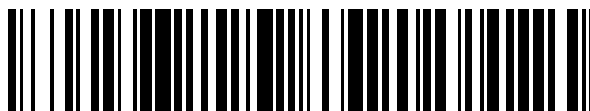


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 791 332**

51 Int. Cl.:

H04L 29/06 (2006.01)

H04N 21/2343 (2011.01)

H04N 21/262 (2011.01)

H04N 21/845 (2011.01)

H04N 21/6547 (2011.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **15.06.2016** **E 16174602 (9)**

97 Fecha y número de publicación de la concesión europea: **29.01.2020** **EP 3107261**

54 Título: **Sistema, método y dispositivos para transmisión de baja latencia**

30 Prioridad:

18.06.2015 IT UB20151488

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

03.11.2020

73 Titular/es:

SKY ITALIA S.R.L. (100.0%)
Via Monte Penice 7
20138 Milano, IT

72 Inventor/es:

BERTOLOTTI, MASSIMO y
SONNATI, FABIO

74 Agente/Representante:

FÚSTER OLAGUIBEL, Gustavo Nicolás

ES 2 791 332 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Sistema, método y dispositivos para transmisión de baja latencia

5 Campo técnico

La presente invención se refiere a un método, tal como se define adicionalmente en la reivindicación 1, a una entidad, tal como se define adicionalmente en la reivindicación 7, y a un sistema, tal como se define adicionalmente en la reivindicación 11, a un programa informático, tal como se define adicionalmente en la reivindicación 12, y a un medio legible por ordenador, tal como se define adicionalmente en la reivindicación 13, para la entrega de contenido.

Antecedentes de la invención

Los protocolos de transmisión en flujo continuo de http (tales como HLS, HDS, SS y DASH) se basan en descomponer un flujo continuo de vídeo en porciones o segmentos gestionados como archivos individuales, que por tanto requieren una determinada cantidad de tiempo para procesarse. Por ejemplo, el procesamiento incluye una compresión realizada de manera progresiva en el contenido a medida que está disponible. Este procedimiento da como resultado la generación de una serie de segmentos que tienen una determinada longitud, que, cuando se completan, pueden descargarse por el usuario para su visualización. Cada segmento está normalmente cerrado, en el sentido de que es una unidad de datos que tiene un principio y un final y sólo puede descargarse una vez que se ha creado totalmente mediante protocolos habituales tales como http 1.0. Por tanto, la presencia de estos segmentos introduce una latencia debida al menos a la duración (y por tanto la longitud) del propio segmento. Reduciendo la longitud del segmento es posible reducir parcialmente la latencia.

Sin embargo, independientemente de los límites posiblemente impuestos sobre la longitud del segmento, en cualquier caso es difícil controlar la latencia global.

También existen técnicas mediante las cuales se hace que un fragmento de un segmento esté disponible para su descarga. Aunque esto contribuye a reducir parcialmente la latencia, no obstante sigue siendo difícil controlar la latencia global.

Las técnicas de transmisión en flujo continuo conocidas también usan los denominados manifiestos, en los que se indican los componentes de un vídeo, por ejemplo los URL de los segmentos en los que se descompuso el vídeo. Cuando los usuarios piden ver contenido tal como un vídeo, reciben en respuesta un manifiesto que contiene, por ejemplo, un URL que les permite descargar uno o más segmentos. Con el fin de reducir la latencia, el manifiesto se envía normalmente lo antes posible en respuesta a la petición del usuario. A pesar de esto, es difícil controlar y limitar la latencia en los sistemas de entrega o distribución de contenido.

El documento WO 2014/145011 A2 se refiere a un sistema y a un método para proporcionar flujo continuo de CTP (flujo continuo de transporte en paquetes comprimidos) como flujo continuo de MPEG a un dispositivo mediante un protocolo de transmisión en flujo continuo en directo. El método comprende generar una lista de reproducción de HLS (transporte de flujo continuo en directo de HTTP) especificando una pluralidad de archivos de medios de HLS antes de generar uno cualquiera de los archivos de HLS, en el que cada uno de la pluralidad de archivos de medios de HLS tiene un nombre único; transmitir la lista de reproducción de HLS a un cliente después de haber generado la lista de reproducción de HLS; realizar una transcodificación del flujo continuo de CPT en al menos un subgrupo de la pluralidad de archivos de medios de HLS; proporcionar al menos uno de los subgrupos de la pluralidad de archivos de medios de HLS al cliente, en el que al menos uno de la pluralidad de archivos de medios de HLS es un archivo de medios pedido por el cliente seleccionado de la lista de reproducción de HLS.

50 Sumario de la invención

El objetivo de la presente invención es resolver al menos algunos de los problemas asociados con las técnicas conocidas y por tanto mejorar las mismas.

La invención se refiere a un método, tal como se define adicionalmente en la reivindicación 1, a una entidad, tal como se define adicionalmente en la reivindicación 7, y a un sistema, tal como se define adicionalmente en la reivindicación 11, a un programa informático, tal como se define adicionalmente en la reivindicación 12, y a un medio legible por ordenador, tal como se define adicionalmente en la reivindicación 13, para entregar contenido a terminales.

60 Lista de figuras

La figura 1 es un diagrama de flujo según una realización de la presente invención;

la figura 2 es un diagrama de bloques de una entidad de suministro de contenido según una realización de la presente invención;

la figura 3 es un diagrama de bloques de un terminal según una realización de la presente invención;

la figura 4 es un diagrama de bloques de un sistema según una realización de la presente invención;

5 la figura 5 es un diagrama a modo de ejemplo para ilustrar el intercambio de mensajes y tiempos entre el terminal y la entidad de suministro;

la figura 6 es un diagrama de bloques de un ordenador capaz de implementar la presente invención según una realización;

10 la figura 7 es un diagrama de tiempo que representa un conjunto de retardos según técnicas conocidas;

la figura 8 es un diagrama de tiempo para ilustrar los retardos según un ejemplo de la presente invención;

15 la figura 9 es un diagrama de bloques que ilustra un ejemplo de una entidad y un terminal según la presente invención;

la figura 10 es un diagrama ilustrativo en el que se representan las operaciones realizadas por el terminal y por la entidad de suministro según un ejemplo de la presente invención.

20 Descripción detallada

Ahora se describirá una primera realización de la presente invención con referencia a la figura 1, que ilustra un diagrama de flujo relacionado con un método para entregar al menos un elemento de contenido por medio de una entidad de suministro.

25 Un ejemplo de una entidad de suministro, que se comentará adicionalmente a continuación, es un servidor capaz de entregar contenido. Contenido significa cualquier información que puede disfrutarse directa o indirectamente (por ejemplo, cuando se necesita decodificación) por el usuario, y por tanto comprende servicios de texto, audio y/o vídeo, datos para ejecutar una aplicación y/o datos relacionados con una aplicación (por ejemplo, para proporcionar un determinado servicio, etc.), así como cualquier combinación de los mismos.

La generación de al menos una primera lista de componentes relacionada con al menos un elemento de contenido se realiza en la etapa S110, en la que se genera la primera lista en un primer instante de tiempo. El al menos un elemento de contenido se refiere a un elemento de contenido que un terminal desea usar, por ejemplo porque se selecciona a partir de una página web de un terminal, mediante una aplicación de un terminal o porque está contenido en una lista en un servidor. La entidad de suministro está configurada para suministrar un elemento de contenido a una pluralidad de terminales. Es decir, todos los terminales incluidos en la pluralidad de terminales pueden comunicarse con la entidad de suministro y pedir dicho contenido a la misma. La lista de componentes relacionada con el contenido puede comprender uno o más componentes del contenido, pero más normalmente proporciona información que permite al terminal acceder a, o descargar, una o más partes de la misma. Por ejemplo, la lista de al menos un componente comprende una lista de al menos un recurso (o de un identificador de un recurso) correspondiente al al menos un componente; el término componente debe entenderse en un sentido amplio, no sólo con referencia a los datos que representan el propio contenido (por ejemplo que representan las tramas de un vídeo) sino también con referencia a un recurso o al identificador del mismo. Recurso significa cualquier elemento físico y/o virtual de un sistema informático en el que el recurso es capaz de entregar al menos una parte del contenido. El recurso puede especificarse por medio de una dirección (por ejemplo IP) del mismo o más habitualmente por medio de un identificador de recurso (por ejemplo URI y/o URL). Si hay varios componentes, habrá una lista de varios recursos o varios identificadores de recursos, correspondiendo cada uno a una parte del contenido.

50 La generación de una segunda lista de componentes relacionada con el mismo contenido también se realiza en la etapa S110, en la que se genera la lista en un segundo instante de tiempo después del primer instante de tiempo. La generación de una lista en un instante de tiempo dado, tal como se ilustrará en varios ejemplos a continuación, puede producirse en tiempos previamente establecidos o tras la generación de un nuevo componente o el recurso del nuevo componente. Se indica que realizar una acción (por parte de una entidad o dispositivo) se refiere a casos (no exhaustivos) en los que la acción se realiza por una unidad interna de la entidad (por ejemplo, un procesador) o por un dispositivo externo (que envía instrucciones correspondientes).

Se indica además que la generación descrita anteriormente comprende hacer que tales listas estén disponibles para la pluralidad de terminales. Más específicamente, y tal como se describe a continuación, todos los terminales incluidos en la pluralidad pueden pedir una lista de componentes relacionada con al menos un elemento de contenido y recibir, en respuesta, una lista generada por la entidad de suministro, en particular la segunda lista, tal como se observará a continuación. Además, se indica que la generación de dichas listas es independiente de una petición del terminal. Además, se hace que la misma lista esté disponible para, y de hecho puede transmitirse a, una multitud de terminales, también opcionalmente al mismo tiempo, tal como puede deducirse a partir de la totalidad de la presente descripción.

65 La recepción, en una entidad de suministro, de una petición a partir de un terminal se realiza en la etapa S120, en la

que la petición se refiere a una petición de una lista de componentes relacionada con al menos un elemento de contenido. La recepción se recibe en un tercer instante de tiempo después del primer instante de tiempo (pudiendo atribuirse la totalidad de los instantes de tiempo primero, segundo y tercero a, o medirse en, la entidad de suministro). Dicho de otro modo, una petición enviada a partir de un terminal se recibe por la entidad de suministro en el tercer instante de tiempo.

El envío de una respuesta que contiene la segunda lista de componentes se realiza en la etapa S130, no antes de un determinado tiempo de retardo, en el que el tiempo de retardo es igual a la diferencia entre el segundo instante de tiempo y el tercer instante de tiempo. Evidentemente, la respuesta debe entenderse en el sentido de una respuesta o contestación a la petición anteriormente mencionada. El tiempo de retardo significa que, aunque la primera lista de componentes ya está disponible en el segundo instante de tiempo (porque se generó en un momento anterior), en vez de responder enviando la primera lista (ya disponible) la entidad de suministro retardará el envío de la respuesta hasta que se haya generado la lista posterior; en ese punto, enviará la lista recién generada, permitiendo así una alineación de la reproducción en el lado de terminal con los instantes en los que se generan las nuevas listas; tal como resultará aún más evidente a partir de los ejemplos, esto permite controlar así como reducir la latencia de manera más eficaz y precisa. Por tanto, la expresión no antes del tiempo de retardo significa que la entidad de suministro espera a que se genere la segunda lista, y después envía, en respuesta a la petición de una lista, una respuesta que contiene la segunda lista. El instante en el que se envía la respuesta con la segunda lista puede coincidir exactamente con el instante en el que se genera, aunque en la práctica hay retardos adicionales debidos al procesamiento o a la preparación de la respuesta así como otros retardos tal como se requiera por la implementación específica. Generar la lista comprende hacer que esté disponible para enviarla al terminal, es decir, puede haberse generado por adelantado pero liberarse (o publicarse) en el tercer instante de tiempo con el fin de poder enviarse al terminal.

En una variante opcional de la primera realización, la segunda lista de componentes comprende al menos un componente que todavía está procesándose. El contenido se divide lógicamente en porciones (correspondientes cada una a un componente), que representan unidades de datos que tienen un comienzo y un final. En particular, un segmento puede considerarse como un archivo o estructura de datos que tiene, precisamente, un comienzo y un final; en este sentido, el segmento es una unidad cerrada. Sin embargo, en la presente variante el componente (o un identificador de un recurso correspondiente) se indica o inserta en la lista aunque esta última aún no se haya cerrado porque todavía está procesándose. Si se hace referencia al componente, por ejemplo, por medio de un URL, entonces será posible acceder al segmento (no cerrado) y empezar a descargar los fragmentos (o porciones) ya procesados; con este propósito, puede ser necesario usar protocolos adecuados para descargar fragmentos de segmentos no cerrados, por ejemplo usando http 1.1 en el modo de codificación en fracciones. Si un protocolo capaz de gestionar únicamente segmentos cerrados (tal como http 1.0) estuviera disponible para el mismo, el terminal tendría que esperar hasta que se haya cerrado el componente o segmento con el fin de poder descargarlo. Se observa que procesamiento significa procesar una señal de inicio que representa el contenido con el fin de hacer que esté disponible en uno o más formatos adecuados para su transmisión. Por tanto, el procesamiento comprende codificar y/o comprimir una señal de inicio, con el fin de obtener segmentos comprimidos de la señal de inicio; en el procesamiento, con el fin de hacer que un segmento abierto esté disponible, es posible procesar fragmentos (en el sentido de porciones) del segmento. Los fragmentos, tal como se ilustrará a continuación, pueden tener un tamaño que es independiente de la unidad de procesamiento del codificador (por ejemplo, pueden ser independientes de la longitud de un GOP en el caso de codificación de MPEG) e independientes de la unidad de transmisión usada por el protocolo de transmisión en flujo continuo (por ejemplo, un segmento).

Según una variante opcional adicional de la presente realización, el terminal determina una banda de recepción basándose en el tamaño de al menos una porción del componente que está procesándose, en el que el tamaño de la porción del componente está determinado o se proporciona al terminal (proporcionado, por ejemplo, por la entidad de suministro o por una entidad distinta que tiene, por ejemplo, la tarea de configurar de manera remota el terminal). La banda puede estimarse, por ejemplo, basándose en el tamaño del fragmento y el tiempo (medido) necesario para descargarlo. Una estimación de la banda de recepción puede resultar útil, por ejemplo, cuando se aplican técnicas de monitorización, diagnóstico y/o comunicación del tipo de tasa de transmisión de bits adaptada (ABR). En este último caso, puede considerarse el ejemplo en el que la lista comprende al menos dos conjuntos (o capas) de componentes, correspondiendo cada conjunto a una determinada resolución de vídeo (y, por tanto, a una determinada tasa de transmisión de bits); el terminal es capaz de seleccionar el componente a partir de un determinado conjunto en vez de otro conjunto basándose en la estimación de banda; por ejemplo, a partir del conjunto correspondiente a la tasa de transmisión de bits que está más cerca de, o inferior a, la banda estimada o determinada, teniendo también en cuenta determinados márgenes. Dado que todavía está procesándose el componente (y, en general, es independiente de otros parámetros asociados con la codificación o transmisión), aún no se conoce su tamaño y por tanto no se comunica por el servidor. Por tanto, según esta variante, la longitud del fragmento se comunica al terminal (o se conoce por este último de alguna otra manera) de tal manera que, midiendo el tiempo necesario para la recepción del mismo, es posible realizar una estimación de la banda. Se indica que estimar la banda también puede resultar útil si la lista de componentes sólo comprende un conjunto relacionado con una única resolución específica (además de los casos de monitorización o diagnóstico): puede considerarse, por ejemplo, el caso en el que el terminal tiene dos o más canales de comunicación disponibles; en este caso, basándose en la banda estimada para un determinado canal, el terminal puede decidir si cambiar o no el canal de comunicación. De hecho, una de las características opcionales pero ventajosas de la transmisión en flujo continuo de HTTP es la capacidad de adaptación de la transmisión en flujo

continuo a fluctuaciones de banda del canal de comunicación cliente-servidor. La banda se mide normalmente por el cliente calculando la razón entre el tamaño de un segmento y el tiempo necesario para descargarlo. Sin embargo, esto requiere normalmente el uso de segmentos cerrados, es decir completos.

5 Medir la banda con los sistemas convencionales no resulta viable si se descargan segmentos (abiertos) de una manera sincronizada con su creación porque el tiempo de descarga es igual a la longitud del segmento aunque la banda disponible sea claramente mayor que la tasa de transmisión de bits del segmento. Por tanto, el método según esta variante opcional prevé que la entidad de suministro gestionará la transmisión del segmento en porciones/fragmentos tal como para permitir la medición de la banda independientemente del uso de segmentos abiertos.

10 Se indica que la porción de componente que está procesándose, o más simplemente el fragmento, es independiente de las unidades lógicas del codificador, cuando están presentes. Por ejemplo, si el codificador produce una serie de unidades lógicas representadas por diferentes GOP, el tamaño del fragmento puede seleccionarse de una manera que será totalmente independiente del GOP y la longitud del mismo. Por tanto, no se necesita que el fragmento contenga necesariamente un número completo de GOP.

15 Esto no descarta que, según otra variante opcional, el fragmento puede diseñarse para contener exactamente un número completo de GOP, sin embargo, puede ser necesario no obstante especificar el tamaño del fragmento si el decodificador no produce GOP que tienen el mismo tamaño. El tamaño o la longitud del fragmento pueden establecerse previamente (por ejemplo, de una vez por todas, o para un determinado intervalo de tiempo), o modificarse de manera dinámica. Además, la longitud del fragmento también puede especificarse en un intervalo incluido dentro del propio fragmento o indicado en un mensaje de señalización independiente, si se desea crear un sistema en el que es posible hacer variar la longitud de cada fragmento o un determinado número de fragmentos según se desee. Debe indicarse que un segmento contiene una porción de vídeo que tiene una determinada duración o longitud, que puede ser fija (por ejemplo 3 segundos) o variar de un segmento a otro segmento. Dicho de otro modo, y según un ejemplo, un segmento contiene información relacionada con 3 segundos de imágenes. Basándose en la complejidad de las imágenes contenidas y los parámetros de compresión/codificación, cada segmento se producirá entonces con un determinado tamaño en bytes. El tamaño de cada segmento puede variar (dada también la misma duración de segmento) según el tipo de codificación seleccionado, los parámetros de la misma o el tipo de vídeo. También puede concebirse establecer determinados parámetros de codificación de tal manera que todos los segmentos tienen la misma duración. Cuando el cliente ha descargado el segmento y lo reproduce, cada segmento puede reproducir por tanto 3 segundos de vídeo.

20 Según una variante opcional adicional de la presente realización, la segunda lista de componentes comprende un componente cuyo procesamiento se ha terminado.

25 En este caso, dado que se ha completado el procesamiento del segmento relacionado con el componente, es posible descargar todo el segmento. Además, de nuevo gracias al hecho de que el segmento está cerrado, se conocerá la longitud y se indicará en el momento de la descarga, de modo que no es estrictamente necesario realizar etapas especiales como en el caso de un segmento abierto.

30 Según otra variante, la segunda lista comprende tanto un componente abierto como un componente cerrado.

35 Según un ejemplo general, cada lista comprende un componente cerrado y un componente abierto, en el que la lista respectiva se genera en el momento en el que se hace que el componente abierto respectivo esté disponible para su acceso o en el momento en el que ha comenzado el procesamiento del mismo.

40 Según una variante opcional adicional de la presente realización, el tiempo de retardo se ignora si el tercer instante de tiempo se encuentra dentro de un intervalo definido por el primer instante de tiempo y un umbral predeterminado. En tal caso, la respuesta comprenderá al menos la primera lista de componentes pero no la segunda lista de componentes. Por ejemplo, si el umbral se establece a un valor de un milisegundo para la longitud de un segmento de 3 segundos, esta variante prevé que, en el caso de peticiones de una lista recibida inmediatamente después (en el ejemplo, no más de un milisegundo después) del momento en el que se generó la última lista, la respuesta ya no se retardará hasta la emisión o generación de la siguiente lista. Por el contrario, se suministrará la lista ya disponible. Por tanto, el umbral en este ejemplo no limitativo de un milisegundo puede ajustarse de modo que las peticiones que llegan lo suficientemente cerca de una lista recién emitida se cumplen sin retardo. Dicho de otro modo, la entidad de suministro puede responder a determinadas peticiones suministrando la lista ya disponible sin retardar la respuesta hasta la generación de la siguiente lista. Las peticiones recibidas después de ese umbral se gestionarán de manera normal, es decir, la respuesta se retardará hasta que se genere la siguiente lista. Esta variante puede usarse, por ejemplo, cuando se desea volver a alinear un terminal que por algún motivo ha perdido una determinada alineación.

45 Las variantes anteriormente descritas pueden combinarse entre sí, tal como resulta evidente para el experto en la técnica. Además, la primera realización se refiere a un método según la presente invención. Sin embargo, todas las consideraciones y variantes anteriores se aplican a dispositivos, tales como la unidad de suministro o el terminal, así como a sistemas que comprenden la unidad de suministro y un terminal, programas de procesador y señales tal como también se ilustrará a continuación. Si se omiten varios detalles por motivos de brevedad, se indica que todas las

observaciones realizadas anteriormente se aplican igualmente a lo que sigue a continuación.

Se indica además que el sistema se ha descrito con referencia a la creación de dos listas en dos instantes de tiempo diferentes. Sin embargo, la invención se aplica igualmente a casos en los que la entidad de suministro genera cualquier número N de manifiestos, cada uno en momentos distintos respectivos.

En la fase de inicialización, es decir en el momento en el que el servidor comienza a procesar el contenido, el sistema puede generar una primera lista que nunca se entregará (porque una petición se cumplirá con el retardo descrito anteriormente) o se entregará sin retardo en respuesta a una petición y únicamente con un componente abierto dentro de la misma. Alternativamente, en la fase de inicialización, el servidor puede crear una lista ficticia vacía para representar la primera lista comentada en la figura 1.

Además, se indica que la invención puede aplicarse por igual independientemente de cómo se estructure una lista y/o de la manera en la que los terminales piden una lista.

De hecho, según un primer ejemplo, el servidor genera listas respectivas en instantes respectivos, conteniendo cada lista un componente recién generado, para generar listas que permiten la recuperación de nuevas porciones de contenido a medida que se generan. El terminal puede pedir de manera repetida una lista a partir del servidor, que responderá con el retardo anteriormente mencionado con el fin de suministrar la siguiente lista (que, tal como se indica, se genera, por ejemplo, en cuanto comienza a procesarse el componente abierto, en cuanto está disponible dentro de la lista el recurso a partir del cual puede descargarse el componente y/o en cuanto puede hacerse que el recurso esté público para el terminal). Por tanto, en este ejemplo, el terminal pide una nueva lista en instantes previamente establecidos o, alternativamente, en cuanto se prepara para completar la descarga de un segmento y/o en cuanto se prepara para reproducir un determinado segmento. Por otro lado, en un segundo ejemplo, puede concebirse construir la lista de tal manera que se especifica el recurso relacionado con un componente N, y después especificar una regla tal como para permitir que el terminal recupere los componentes posteriores (N+i, con $i > 1$) a partir de recursos respectivos que pueden obtenerse basándose en el segmento N y en esa regla. Por ejemplo, de manera convencional, el terminal sabe que los componentes posteriores sólo estarán disponibles en un determinado instante.

En tal caso, el terminal sólo necesitará recibir una lista, siendo capaz de determinar cómo descargar los componentes posteriores. Esto no altera el hecho de que el servidor debe actualizar en cualquier caso la lista en intervalos de tiempo respectivos: de hecho, un terminal B que realiza una petición después que un terminal A comenzará a descargar, como primer segmento para su reproducción, un segmento N+j ($j > 1$) más tarde en el tiempo en comparación con el segmento N descargado por el terminal A como primer segmento. Una vez recibida la lista, según el segundo ejemplo, el terminal A será capaz de determinar dónde (y posiblemente también cuándo) descargar los segmentos N+k ($k > 0$) tras el segmento N; el terminal B sabrá cómo descargar los segmentos N+j+k comenzando a partir del segmento N+j. Por tanto, haciendo referencia al método de figura 1, el terminal A recibirá la lista con un determinado retardo, hasta que se actualice con el segmento N; de manera similar, el terminal B recibirá una respuesta con un retardo que es suficiente para la generación de la segunda lista que contiene el segmento N+j. La regla para determinar los segmentos posteriores puede ser cualquiera: según un ejemplo, se actualiza un identificador de un recurso que comprende un contador cada T segundos o según un patrón de tiempo previamente establecido, etc.

Ahora se describirá una segunda realización con referencia a la figura 2, que ilustra una entidad que suministra al menos un elemento de contenido y que comprende unos medios 210 de generación, unos medios 220 de recepción y unos medios 230 de envío. La entidad puede consistir en un dispositivo o, alternativamente, en un conjunto de dispositivos interconectados entre sí, y puede implementarse, en particular, en hardware, software o una combinación de los mismos, de forma tanto concentrada como distribuida, como los medios incluidos en la misma. Ejemplos de una entidad de suministro son un servidor, una granja de servidores, una solución de software implementada en una plataforma en la nube, etc.

Los medios de configuración están configurados para generar al menos una primera lista de componentes y una segunda lista de componentes, ambas relacionadas con un elemento de contenido deseado por un terminal (o por un usuario que hace funcionar el terminal). La primera lista y la segunda lista son tales como para generarse en dos instantes de tiempo diferenciados, más específicamente, en un primer instante de tiempo en el caso de la primera lista y en un segundo intervalo de tiempo después del primer intervalo de tiempo en el caso de la segunda lista.

Los medios de recepción están configurados para recibir, a partir de un terminal, una petición referente a una lista de componentes relacionada con al menos un elemento de contenido, en los que la petición se recibe en la entidad en un tercer instante de tiempo después del primer instante de tiempo. Tal como se mencionó anteriormente, la lista puede contener los componentes, pero preferiblemente contiene una referencia al contenido o un identificador de un recurso a partir del cual puede obtenerse o descargarse el contenido. Por tanto, los medios 230 de envío están configurados para enviar, en respuesta a la petición, una respuesta que contiene la segunda lista de componentes, sin embargo, no antes de un tiempo de retardo. El tiempo de retardo es igual a la diferencia entre el segundo instante de tiempo y el tercer instante de tiempo, de modo que la entidad de suministro retarda la respuesta a la petición de la lista hasta que está disponible o se genera una nueva lista. Generar la nueva lista comprende diferentes casos tal como se ilustró anteriormente, en particular generar una lista que comprende nuevos componentes o segmentos a

medida que están disponibles o actualizar una lista a partir de la cual el terminal logra establecer cómo obtener diversos componentes empezando a partir de un primer componente actualizado y una regla previamente establecida. La entidad de la presente realización puede estar configurada además para realizar todas las demás operaciones descritas anteriormente con referencia a la primera realización. Ahora se ilustrará una tercera realización con referencia a la figura 3, que representa un terminal 300 para usar al menos un elemento de contenido suministrado por una entidad de suministro.

Los medios 310 de petición están configurados para enviar a la entidad de suministro una petición de al menos un primer elemento de contenido. En cambio, los medios 320 de recepción están configurados para recibir una respuesta a partir de dicha entidad de suministro, en los que la respuesta contiene una lista de componentes relacionada con el contenido deseado. Sin embargo, la respuesta se distingue por el hecho de que se recibe con un determinado retardo basándose en un intervalo de tiempo que transcurre entre un primer instante de tiempo, en el que la petición llega a la entidad de suministro, y un segundo instante de tiempo, en el que se genera la lista en la entidad de suministro (en el que el segundo instante de tiempo coincide con el, o sigue al, primer instante de tiempo). El retardo basado en el intervalo de tiempo indica que también puede haber otros retardos, tales como, por ejemplo, retardos debidos a la transmisión y/o al procesamiento de la señal necesaria para la transmisión. El terminal comprende además unos medios de estimación (no ilustrados) para estimar una banda de recepción basándose en la longitud de al menos una porción del componente que está procesándose, y en el que la longitud de la porción del componente está predeterminada o se proporciona al terminal. En particular, el terminal está configurado para determinar o estimar la banda de recepción del componente basándose en un fragmento del componente y el tiempo (tal como se mide por el terminal) necesario para recibir ese fragmento. La longitud del fragmento, que no se conoce (dado que se refiere a un segmento abierto), la conocerá el terminal, o bien porque está predeterminada o bien porque se comunica al terminal de cualquier manera. La estimación de banda puede resultar útil en diversas circunstancias, tal como se explicará adicionalmente a continuación. Además, todas las consideraciones ya expuestas con referencia al método también se aplican para el terminal.

La figura 4 muestra un sistema según otra realización, en el que el sistema comprende una entidad 420 de suministro tal como se describió anteriormente y un terminal 430, también tal como se describió anteriormente. La entidad 420 y el terminal 430 pueden estar interconectados mediante cualquier red 410 fija y/o móvil.

En otra realización, se proporciona una señal para portar una estructura de datos que comprende un mensaje de respuesta que contiene una lista de componentes relacionada con un elemento de contenido. El mensaje de respuesta se genera en respuesta a un mensaje a partir del terminal que pide una lista de componentes. La respuesta se genera por la entidad de suministro con un retardo basándose en un intervalo de tiempo que transcurre entre un primer instante de tiempo, en el que la petición llega a la entidad de suministro, y un segundo instante de tiempo, en el que se genera la lista en la entidad de suministro.

Según otra realización, se proporciona un programa de procesador configurado para ejecutar, cuando se ejecuta el programa en un ordenador, cualquier etapa o combinación de etapas del método y las variantes del mismo tal como se describió con referencia a la primera realización.

La figura 6 ilustra un diagrama de bloques que muestra a modo de ejemplo un ordenador 600 capaz de ejecutar el programa anteriormente mencionado. En particular, el ordenador 600 comprende una memoria 630 para memorizar las instrucciones del programa y/o datos necesarios para la ejecución del mismo, un procesador 620 para ejecutar las instrucciones y una interfaz 610 de entrada/salida.

Según una realización adicional, se proporciona un medio para soportar un programa de procesador configurado para ejecutar, cuando se ejecuta el programa en un ordenador, una etapa o combinación de etapas según el método descrito en la primera realización. Ejemplos de un medio son una memoria estática y/o dinámica, un disco fijo o cualquier otro medio tal como un CD, DVD, Blu Ray. En el medio también están comprendidos medios capaces de soportar una señal que representa las instrucciones, incluyendo unos medios de transmisión por cable (Ethernet, óptico, etc.) o transmisión inalámbrica (celular, por satélite, transmisión terrestre digital, etc.).

La figura 5 ilustra esquemáticamente el intercambio de mensajes que piden una lista y la respuesta correspondiente según la presente invención. En particular, el eje horizontal en la parte superior muestra los tiempos tal como se miden por la entidad de suministro, mientras que el eje horizontal en la parte inferior muestra los tiempos tal como se miden por el terminal. La entidad de suministro genera una primera lista en el tiempo T1 y una segunda lista después en el tiempo T2. El intervalo de tiempo entre T1 y T2 es un intervalo que puede ser controlable, por ejemplo por la entidad de suministro. De hecho, controlable implica que el intervalo puede ser predefinido como constante o no constante, y variable siempre que permanezca bajo el control de la entidad de suministro. El terminal genera, en el tiempo T3*, una petición de una lista; la petición se recibe por la entidad de suministro en el tiempo T3, después del tiempo T3*, debido al retardo debido a la transmisión (procesamiento y/o propagación). Se indica que el tiempo T3 sigue al T1 y precede a T2. En este punto, la entidad de suministro puede enviar, en respuesta, la lista generada en el tiempo T1. Sin embargo, tal como se explicó anteriormente, la entidad de suministro retarda el envío de la respuesta al menos hasta el tiempo T2, es decir hasta que se genera la nueva segunda lista en el tiempo T2. Por tanto, la lista se recibe en el terminal en el tiempo T2*, que por tanto está alineado con la entidad de suministro con respecto a la generación de la

nueva lista (a menos que haya retardos debidos al procesamiento, transmisión y/o propagación de la señal). El retardo d (referencia 520 en la figura) indica el tiempo en el que un cliente se pone en espera durante la fase de sincronización (por ejemplo, la primera vez que un usuario pide un manifiesto, o cuando un usuario ya no está alineado dentro de determinadas tolerancias y/o cuando pide un manifiesto con el fin de volver a alinearse).

5 Ahora se hará referencia a la figura 10 (indicando que los tiempos no están necesariamente asociados a los de la figura 5), que ilustra un ejemplo particular basándose en las enseñanzas divulgadas anteriormente. En particular, las actividades llevadas a cabo en el servidor con referencia a la generación se muestran esquemáticamente en la parte superior de la figura; mientras que las actividades relacionadas con el procesamiento de los segmentos, de nuevo en el lado de servidor, se muestran en el centro de la figura; finalmente, las actividades en el lado de terminal con referencia a las listas y segmentos se muestran en la parte inferior de la figura.

15 En el ejemplo se supone que cada segmento dura 3 segundos, y que los terminales tardan un segundo en descargarlo. Las listas generadas posteriormente se indican como M0, M1, M2, ...; la generación de las mismas tiene lugar en los instantes T1, T2, T3, respectivamente (en el ejemplo están separados de manera equidistante, pero lo mismo se aplica para intervalos irregulares). Más específicamente, los instantes T1, T2, T3 pueden definirse mediante intervalos regulares o irregulares, en los que tales intervalos son controlables, por ejemplo, por la entidad de suministro.

20 La publicación de la lista es un ejemplo de generación, y de hecho comprende hacer que la lista esté disponible para incluirse en la respuesta a los terminales. En cambio, los segmentos procesados en sucesión se indican como S0, S1, S2 (no se descarta que dos o más segmentos pueden procesarse en paralelo o parcialmente en paralelo según el tipo de vídeo de partida, en particular si no es en directo, y de los recursos de hardware/software disponibles). Una trayectoria lógica que ilustra las etapas necesarias (para el terminal) para obtener el primer segmento necesario para el terminal para comenzar la reproducción del contenido se indica con una línea de rayas y puntos. Además, las partes del diagrama con un fondo gris se refieren a actividades relevantes para el nuevo terminal que desea comenzar a visualizar un determinado contenido.

30 Tal como puede observarse a partir de la figura, en el tiempo T0 el servidor ha preparado la lista M0 (también indicada como manifiesto en la figura) y también ha completado la producción y el procesamiento del segmento S0, mientras que está empezando a producir el segmento S1. En el ejemplo, la generación de la lista M0, la terminación del segmento S0 y el comienzo del procesamiento del segmento S1 pueden tener lugar indistintamente en cualquier instante comprendido en los intervalos representados por el bloque respectivo. Estos detalles se omiten, dado que el experto en la técnica entenderá fácilmente que no se necesita que tales instantes estén necesariamente sincronizados entre sí; pueden variar igualmente según la codificación subyacente. En este punto se supone que un nuevo terminal pide una lista en el instante 15:00:01,05, es decir en la mitad del bloque indicado como 15:00:01. Dado que el servidor ha puesto todos los clientes/terminales en espera, en el sentido de que en esta etapa no se cumple una petición de un manifiesto sino que se pone en espera, precisamente, hasta que servidor determina que la lista M1 está lista (indicando que esto no puede producirse para un manifiesto anteriormente alineado o sincronizado, en el sentido de que pide y obtiene segmentos en los instantes apropiados); la lista M1 está lista cuando se termina completamente la producción del segmento S1 (es decir cuando se cierra el segmento) y/o cuando ha comenzado la producción del segmento S2. Por tanto, la lista M1 incluirá una referencia a S1 y una referencia a la lista abierta S2 (sin embargo, tal como se observa en este caso, la lista M1 también puede comprender únicamente uno de estos). En este punto, el terminal recibe la lista M1 y sabrá dónde descargar tanto el segmento S1 como el segmento S2. Entonces el terminal procederá a descargar el segmento S1, que puede descargarse en su totalidad, al ser un segmento cerrado; en el ejemplo esta operación dura 1 segundo. El terminal también puede decidir en este punto comenzar a reproducirlo, lo cual permitirá tener una memoria intermedia, en el ejemplo, de al menos 3 segundos. Una vez completada la descarga del segmento S1, el terminal puede comenzar a descargar el segundo segmento abierto S2, por ejemplo por medio de codificación de transferencia en fracciones según el protocolo http 1.1, o bien el terminal puede comenzar a descargar fragmentos del segmento abierto S2 a medida que están disponibles en el servidor. La descarga de los fragmentos del segmento S2 puede comenzar en cuanto se ha descargado el segmento cerrado S1 y por tanto mientras todavía está reproduciéndose el segmento S1 (pero también en paralelo, si se permiten dos conexiones físicas o lógicas). En teoría, el segmento S1 puede omitirse, o reducirse o aumentarse el tamaño del mismo según se desee con el fin de reducir o variar el tamaño de memoria intermedia durante la reproducción. Su presencia, además de aumentar o permitir el ajuste de la memoria intermedia, también hace posible estimar la banda con el fin de poder reaccionar (por ejemplo cambiar la capa/resolución, véase anteriormente) en el momento de comenzar a descargar los fragmentos del segmento S2. Asimismo, estimar la banda en los fragmentos del segmento S2 hace posible poder reaccionar con adelanto suficiente, es decir, o bien cambiando la capa dentro de la lista o bien cambiando el canal de comunicación según se necesite. Tal como puede observarse a partir de la figura, en cuanto se ha conectado con el servidor el terminal recibirá la lista con un determinado retardo, pero será capaz de reproducir el contenido con un retardo menor dado que desde ahora en adelante estará alineado con la emisión del manifiesto. Por tanto, esta parte del retardo se vuelve controlable; además, esto significa una reducción en la latencia global tal como se ilustra con referencia a la figura 8.

65 Además, si se conectan varios usuarios, aunque se conecten (como sucede en la práctica) en instantes de tiempo diferentes, recibirán los manifiestos de una manera sustancialmente alineada.

A continuación se ilustran varias diferencias en comparación con técnicas conocidas, tal como se reconoce por los inventores. Con referencia a la figura 7, se toma, por ejemplo, el caso del protocolo de HLS convencional. El codificador adquiere y codifica el vídeo; después de 10 s (retardo D1 en la figura) se crea un archivo .ts y se carga en el servidor de origen (retardos D1 y D2). El cliente (por ejemplo el ordenador de tipo tableta que recibe servicios de OTT, de los cuales Sky Go representa un ejemplo) no puede acceder al segmento de vídeo hasta que este último esté completamente subido en el servidor; por tanto, esto introduce 10 s de latencia (D1), a lo que puede añadirse el tiempo de subida al servidor (D2).

Sin embargo, deben tenerse en cuenta transcurros de tiempo adicionales: (i) la alineación errónea de codificador-cliente (retardo D3), es decir el hecho de que el cliente no pide el segmento exactamente al transcurrir los 10 s sino que también puede pedirlo antes o después (de una manera aleatoria, desde el punto de vista del servidor), y (ii) el tiempo de descargarlo al cliente (retardo D4).

Al final de la cadena, también debe añadirse la actividad de almacenamiento en memoria intermedia realizada por el cliente (es decir, el ordenador de tipo tableta), que, en el caso de usar el protocolo de HLS, es igual a la duración de al menos 2 segmentos (retardo D5).

Puede reducirse la latencia reduciendo el tamaño del segmento (cuanto menor es el segmento que está "procesándose", menos tiempo se necesitará para distribuir el contenido).

Sin embargo, en la técnica tradicional, no se usan segmentos que duren menos de 10 s (como en el ejemplo que acaba de mencionarse). También debe indicarse que una reducción excesiva del tamaño del segmento de vídeo puede reducir la calidad de compresión (codificación). Además, si los segmentos de vídeo tienen una duración reducida, el número de peticiones a partir del cliente al servidor aumentará, de modo que este último se verá sobrecargado con un gran número de peticiones (por ejemplo, si el flujo continuo de vídeo de un partido de fútbol se divide en 1.000 segmentos de 10 s, cada cliente consultará al servidor 1.000 veces; si el mismo flujo continuo de vídeo se divide en 2.000 segmentos de 5 s, cada cliente consultará al servidor 2.000 veces). El uso de técnicas conocidas conduce de hecho a altas latencias. De hecho, mientras que la transmisión de un elemento de contenido (por ejemplo en directo) por satélite tiene un retardo promedio de 5 s, la transmisión del mismo contenido en directo en OTT tiene un retardo promedio de 30 s.

A continuación se ilustra un ejemplo específico que implementa diversos aspectos de la invención, cuyos retardos se ilustrarán entonces con referencia a la figura 8 con el fin de compararlos con los de la figura 7. En particular, según este ejemplo no limitativo de lo que se describió anteriormente, se implementan las siguientes medidas en combinación (aunque pueden aplicarse igualmente cada una de ellas por sí solas o una subcombinación de las mismas):

(A) cuando el cliente pide un elemento de contenido, el servidor suministra un archivo XML, o un denominado "manifiesto" (siendo el manifiesto un ejemplo de una lista descrita anteriormente). El manifiesto contiene la lista de segmentos disponibles (es decir, ya creados y subidos al servidor). Tradicionalmente, el manifiesto indica los últimos segmentos disponibles pero no indica el segmento que está procesándose. En el presente caso, además de los últimos segmentos disponibles, el manifiesto también indica el segmento posterior que está procesándose (es decir, que está creándose en el servidor). Esto significa que el cliente también puede comenzar a descargar el segmento que está procesándose.

(B) La segunda medida usada prevé una "sincronización de cliente". Los clientes piden el archivo de manifiesto en un momento dado, cuando eligen, y normalmente lo obtienen en cualquier momento, independientemente del momento de creación del mismo en el servidor (generando un retardo debido a la falta de sincronía). En el presente caso, el cliente únicamente obtendrá el archivo de manifiesto tras la generación del siguiente manifiesto de modo que, a partir de ese momento en adelante, estará alineado con la generación de los manifiestos y de los segmentos en el servidor, también si el nuevo segmento está disponible después de n segundos. Esto hace posible, también tal como se explicó anteriormente, establecer un control de latencia, y posiblemente reducirla de manera global.

(C) La tercera medida permite al cliente acceder a segmentos en el servidor aunque no se hayan subido completamente al servidor (es decir, si están disponibles 5 s de un segmento de 10 s en el servidor, el cliente ya puede empezar a descargar el segmento antes de que esté disponible la totalidad de los 10 s en el servidor).

(D) La cuarta medida va dirigida a resolver la gestión de la tasa de transmisión de bits adaptativa en el lado de cliente.

El cliente es normalmente capaz de entender la banda disponible, cuando se conoce el tamaño de segmento, y calcular el tiempo de transferencia del segmento, de modo que el cliente es capaz de identificar qué capa de vídeo (véase lo que se ha comentado con respecto a esto) usar en un momento dado. En el presente caso, será difícil calcular la banda disponible porque el cliente no conoce el tamaño de la porción de datos que está transfiriéndose. Con el fin de resolver el problema y ser capaz de gestionar la tasa de transmisión de bits adaptativa en el caso de la transmisión en flujo continuo de segmentos "abiertos" (es decir, aún no subidos completamente en el servidor), tal como se describió anteriormente, el terminal estima la banda basándose en el conocimiento de la longitud o el tamaño de una parte del segmento (fragmento) y/o segmento; el terminal adquiere conocimiento de este tamaño o bien porque está

preestablecido o acordado o bien porque se informa en consecuencia.

Las medidas anteriormente mencionadas también pueden indicarse de la siguiente manera:

- 5 1. Generación del manifiesto “de anticipación”
2. Sincronización del cliente con el momento de emisión del manifiesto
- 10 3. Gestión de los segmentos en codificación en fracciones de HTTP 1.1
4. Gestión del reproductor múltiple de ABR en el lado de cliente.

Con el fin de detallar adicionalmente el ejemplo, se indica lo siguiente con referencia a los puntos (1) y (2), es decir con respecto a la creación de un manifiesto de anticipación (1) y el envío sincronizado del mismo (2).

El manifiesto puede crearse con dos segmentos a los que se hace referencia dentro del mismo, el primero de los cuales es el ya creado, mientras que el segundo es el que todavía está procesándose (“anticipación”), pero que en cualquier caso puede descargarse por los clientes gracias, por ejemplo, a la técnica de codificación en fracciones. Aquí se presenta el ejemplo con más detalle:

- en el instante de tiempo t_0 , cuando termina la creación del segmento $n.º$ 1 y comienza la del segmento $n.º$ 2, se envía un manifiesto que hace referencia tanto al segmento $n.º$ 1 como al segmento $n.º$ 2; el cliente también puede empezar a descargar los primeros fragmentos de segmento $n.º$ 2, de modo que se reduce la latencia debida a D1;

- en el siguiente instante t_1 , igual a t_0 más la duración de un segmento, cuando termina la creación del segmento $n.º$ 2, comienza la del segmento $n.º$ 3; se envía un manifiesto que hace referencia al segmento $n.º$ 2 y al segmento $n.º$ 3.

El ejemplo anterior también prevé la sincronización de la recepción del manifiesto por el cliente, para reducir también D3: se hace que el manifiesto esté disponible para clientes únicamente en el momento en el que se crea por el servidor; si un cliente lo pide en otro instante, el servidor no lo devolverá hasta la siguiente actualización.

Usando estas técnicas y con segmentos de 2 segundos, el retardo parece por tanto limitado:

$$D1 < 0,5 \text{ s, } D2 \text{ despreciable, } D3 \text{ despreciable, } D4+D5 = \sim 3-4 \text{ s.}$$

Por tanto, el retardo es igual a aproximadamente 3,5-4,5 s.

Ahora se comentarán otros detalles con referencia al punto (3) anterior, indicado con referencia al ejemplo, es decir, con referencia al envío en http 1.1 usando codificación de transferencia en fracciones. Parte de la latencia atribuible a las técnicas de transmisión en flujo continuo en directo de HTTP convencionales está asociada a la gestión de los segmentos como archivos indivisibles individuales. Por tanto, en vez de esperar a la formación de un segmento de 10 s, será mejor enviar el segmento inmediatamente a través del servidor y hacer que esté disponible para su descarga aunque el segmento aún no se haya formado completamente. Este sistema permitirá reducir el tiempo de retardo D1 prácticamente hasta 0 (o casi). Es posible obtener este resultado usando el modo de “codificación en fracciones” de Http 1.1 (método previsto por la norma Http 1.1). En este modo, no se especifica el tamaño de la porción de la transferencia de segmentos con Http, pero la transferencia tiene lugar en fracciones “abiertas” (es decir, fragmentos), porciones de datos que pueden ser cualquier fracción del segmento original. Dependiendo del contenedor (o protocolo de transmisión en flujo continuo) usado (fMP4 en el caso de HDS, SS, DASH o .ts como en el caso de HLS, DASH) la división de las porciones de datos tiene lugar basándose en una unidad de transmisión diferente en el caso de contenedores diferentes. Usando este enfoque, se acorta el diagrama de tiempo tal como se ilustra en la figura 8. El tiempo D1 pasa a ser independiente del tamaño de segmento y puede suponerse que se reduce de una manera sencilla hasta aproximadamente medio segundo. El retardo D2 prácticamente desaparece o pasa a ser despreciable. Lo que queda es D3, que de hecho puede pasar a ser más crítico en proporción y requerir una mayor sincronización entre cliente y codificador, y, finalmente, la memoria intermedia, que no puede comprimirse más allá de un determinado grado por los motivos comentados en la sección anterior. Sin embargo, puede reducirse D3 y mantenerse bajo control, gracias a la técnica descrita anteriormente, mediante la cual se envía el manifiesto con un retardo con el fin de alinearlo con el momento de su generación. Se indica que, una vez sincronizado el cliente, se espera que las siguientes veces el cliente pedirá el manifiesto en sincronía (desde un punto de vista de ingeniería casi perfecta) con la creación del mismo en el servidor. Suponiendo fragmentos de 500 ms, puede estimarse que la latencia previa al almacenamiento en memoria intermedia en el lado de cliente será de aproximadamente 2 s. Esto conducirá a una latencia final de aproximadamente 7-8 s. Los retardos D4 y D5 introducidos mediante la descarga de los datos y la memoria intermedia de cliente no se consideran en la solución.

Debe observarse que, en las figuras 7 y 8, se indican los siguientes retardos D1-D5:

D1 creación del segmento

D2 envío del segmento a través del servidor

D3 alineación errónea entre el momento en que se crea el fragmento y el momento en que se pide por el cliente

5

D4 tiempo de descarga

D5 almacenamiento en memoria intermedia en el cliente

10 Una comparación de las figuras 7 y 8 revela que es posible controlar la latencia y asimismo reducirla.

Desde un punto de vista de arquitectura, véase la figura 9, un enfoque de este tipo puede implementarse conectando el codificador al servidor con una conexión de latencia mínima (por ejemplo, RTMP). Por tanto, el servidor recibirá el flujo continuo en tiempo real a partir del codificador y responderá al cliente creando y entregando los segmentos en una porción de datos a la vez: de ese modo se reduce el tiempo D2.

15

Con referencia a la gestión de reproductor múltiple de ABR en el lado de cliente, la transferencia del segmento abierto tiene lugar en pequeños fragmentos (ráfagas) de datos consecutivos de un tamaño conocido para permitir una medición de la banda. No se necesita que el fragmento coincida necesariamente con una fracción, o con un GOP, tal como ya se explicó. Esta solución puede requerir una personalización del reproductor en el lado de cliente.

20

Gracias a las enseñanzas descritas en el presente documento, es posible obtener un sistema para distribuir contenido de audio/vídeo en transmisión en flujo continuo de http con una latencia de transmisión mucho menor de la que se produce actualmente, así como controlar la latencia. Basándose en un prototipo, se encontró que, adoptando segmentos de 2 segundos, una solución basada en las presentes enseñanzas introduce una latencia de aproximadamente 4-5 segundos en comparación con los aproximadamente 11 s de latencia con técnicas tradicionales.

25

Muchas de las realizaciones y los ejemplos se han divulgado con referencia a etapas de métodos o procedimientos. Sin embargo, lo que se ha descrito también puede implementarse en un programa que va a ejecutarse en una entidad de procesamiento (también una distribuida) o una entidad cuyos medios están configurados de manera adecuada o viceversa. Tal como se ilustró anteriormente, la entidad puede implementarse en un único dispositivo, mediante HW/SW o una combinación de estos últimos, o en múltiples unidades o dispositivos interconectados (también en este caso HW, SW o una combinación de los mismos).

30

Además, características tales como la entidad de suministro, medios de generación, medios de recepción, medios de envío y medios de estimación pueden sustituirse, de una manera equivalente, por un elemento de suministro, un generador, un receptor, un transmisor y una unidad de estimación o procesador, respectivamente. Naturalmente, la descripción expuesta anteriormente en el presente documento referente a realizaciones y ejemplos que se aplican a los principios reconocidos por los inventores se proporciona únicamente a modo de ejemplo de estos principios y por tanto no debe entenderse como una limitación del alcance de la invención reivindicada en el presente documento.

35

40

REIVINDICACIONES

1. Método para entregar al menos un elemento de contenido a un terminal de entre una pluralidad de terminales realizado por una entidad de suministro, comprendiendo el método las etapas de:

5 generar (S110) al menos una primera lista de componentes relacionada con dicho al menos un elemento de contenido en un primer instante de tiempo y una segunda lista de componentes relacionada con dicho al menos un elemento de contenido en un segundo instante de tiempo después del primer instante de tiempo, en el que la generación incluye hacer que las listas estén disponibles para dicha pluralidad de terminales y en el que la generación es independiente de cualquier petición a partir de los terminales;

10 recibir (S120) una petición, a partir del terminal, de una lista de componentes relacionada con dicho al menos un elemento de contenido en un tercer instante de tiempo después del primer instante de tiempo;

15 enviar (S130), en respuesta a dicha petición, una respuesta que contiene la segunda lista de componentes no antes de un tiempo de retardo, en el que el tiempo de retardo es igual a la diferencia entre el segundo instante de tiempo y el tercer instante de tiempo.
2. Método según la reivindicación 1, en el que la segunda lista de componentes comprende al menos un componente que todavía está procesándose.
3. Método según una cualquiera de las reivindicaciones anteriores, en el que el terminal determina una banda de recepción basándose en la longitud de al menos una porción del componente que está procesándose, y en el que la longitud de la porción del componente está predeterminada o se proporciona al terminal.
4. Método según una cualquiera de las reivindicaciones anteriores, en el que la segunda lista de componentes comprende un componente cuyo procesamiento se ha completado.
5. Método según una cualquiera de las reivindicaciones anteriores, en el que el terminal determina una banda de recepción basándose en el tamaño de al menos una porción del componente cuyo procesamiento se ha terminado.
6. Método según una cualquiera de las reivindicaciones anteriores, en el que el tiempo de retardo se ignora si el tercer instante de tiempo se encuentra dentro de un intervalo definido por el primer instante de tiempo y un umbral predeterminado, y en el que la respuesta comprende al menos la primera lista de componentes pero no la segunda lista de componentes.
7. Entidad (200) que suministra al menos un elemento de contenido a un terminal de entre una pluralidad de terminales, comprendiendo dicha entidad (200):

40 unos medios (210) de generación configurados para generar al menos una primera lista de componentes relacionada con dicho al menos un elemento de contenido en un primer instante de tiempo y una segunda lista de componentes relacionada con dicho al menos un elemento de contenido en un segundo instante de tiempo después del primer instante de tiempo en la que la generación incluye hacer que las listas estén disponibles para dicha pluralidad de terminales y en la que la generación es independiente de cualquier petición a partir de los terminales;

45 unos medios (220) de recepción configurados para recibir, a partir del terminal, una petición de una lista de componentes relacionada con dicho al menos un elemento de contenido en un tercer instante de tiempo después del primer instante de tiempo;

50 unos medios (230) de envío configurados para enviar, en respuesta a dicha petición, una respuesta que contiene la segunda lista de componentes no antes de un tiempo de retardo,

55 en la que el tiempo de retardo es igual a la diferencia entre el segundo instante de tiempo y el tercer instante de tiempo.
8. Entidad (200) según la reivindicación 7, en la que la segunda lista de componentes comprende al menos un componente que todavía está procesándose.
9. Entidad (200) según la reivindicación 7 u 8, en la que la segunda lista de componentes comprende un componente cuyo procesamiento se ha terminado.
10. Entidad (200) según cualquiera de las reivindicaciones 7 a 9, en la que el tiempo de retardo se ignora si el tercer instante de tiempo se encuentra dentro de un intervalo definido por el primer instante de tiempo y un umbral predeterminado, y en la que la respuesta comprende al menos la primera lista de componentes pero

no la segunda lista de componentes.

- 5
11. Sistema que comprende un terminal y una entidad de suministro según cualquiera de las reivindicaciones 7 a 10.
12. Programa informático configurado para realizar, cuando se ejecuta dicho programa en un ordenador, todas las etapas según una cualquiera de las reivindicaciones de método 1 a 6.
- 10
13. Medio legible por ordenador para soportar un programa de procesador configurado para realizar, cuando se ejecuta dicho programa en un ordenador, todas las etapas según una cualquiera de las reivindicaciones de método 1 a 6.

FIG. 1

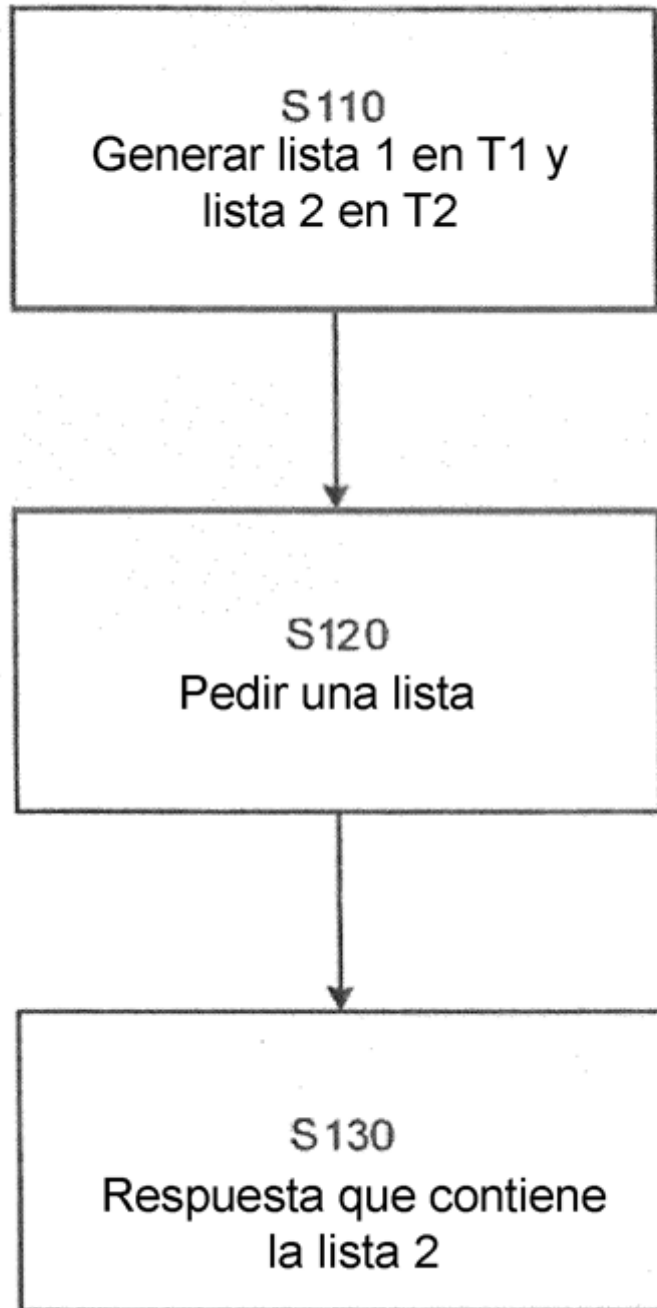


FIG. 2

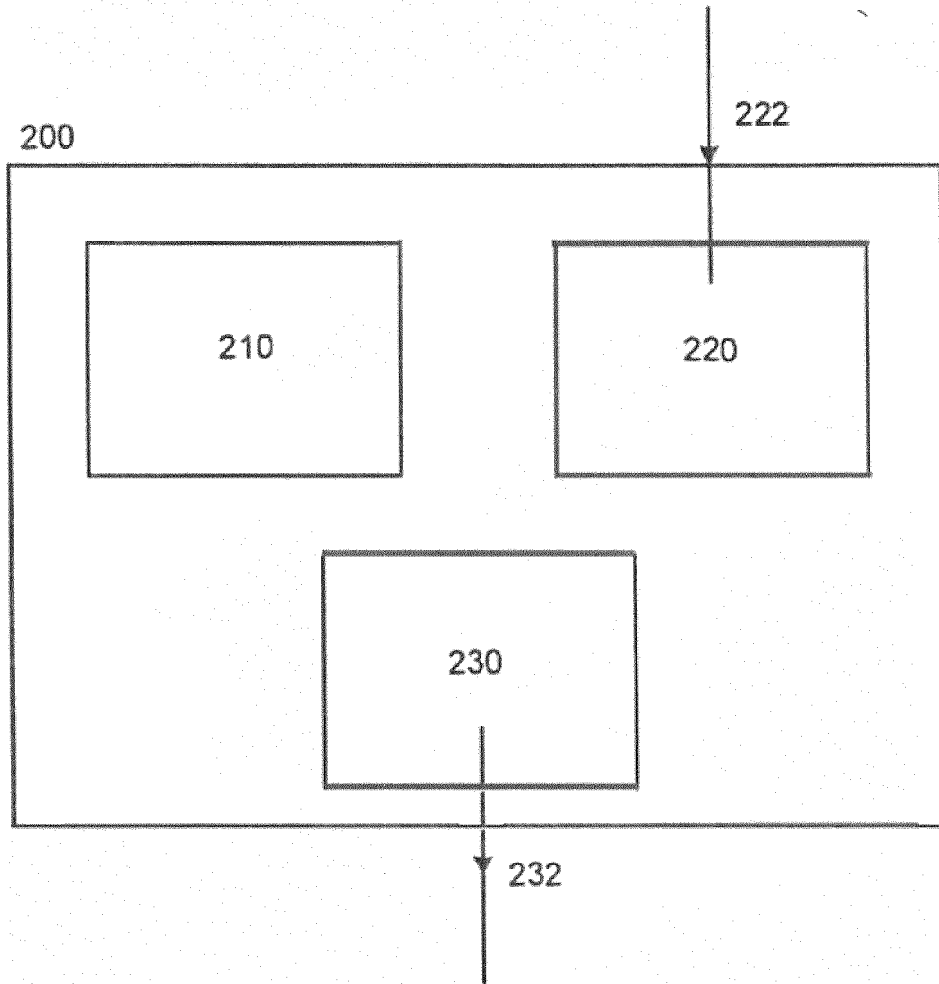


FIG. 3

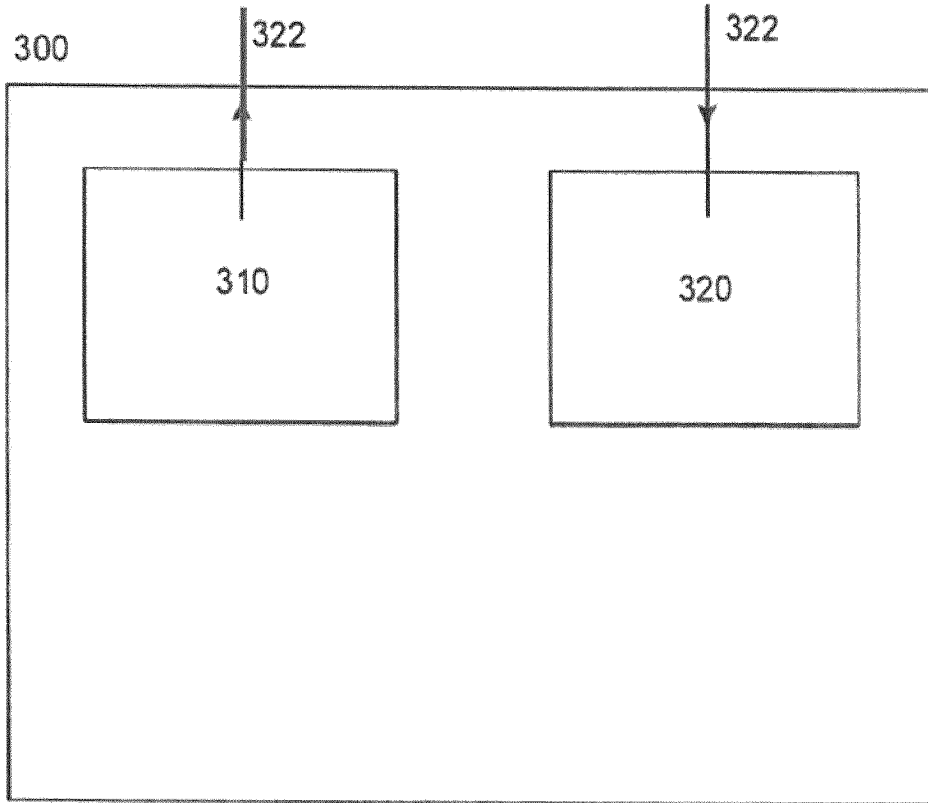
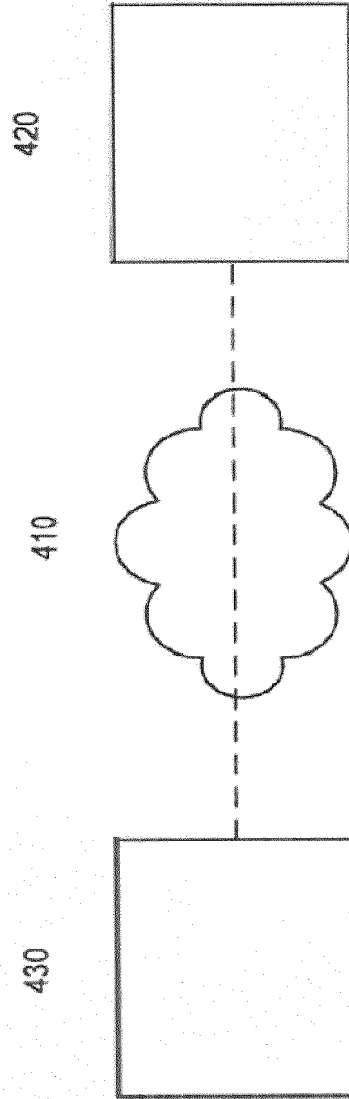


FIG. 4



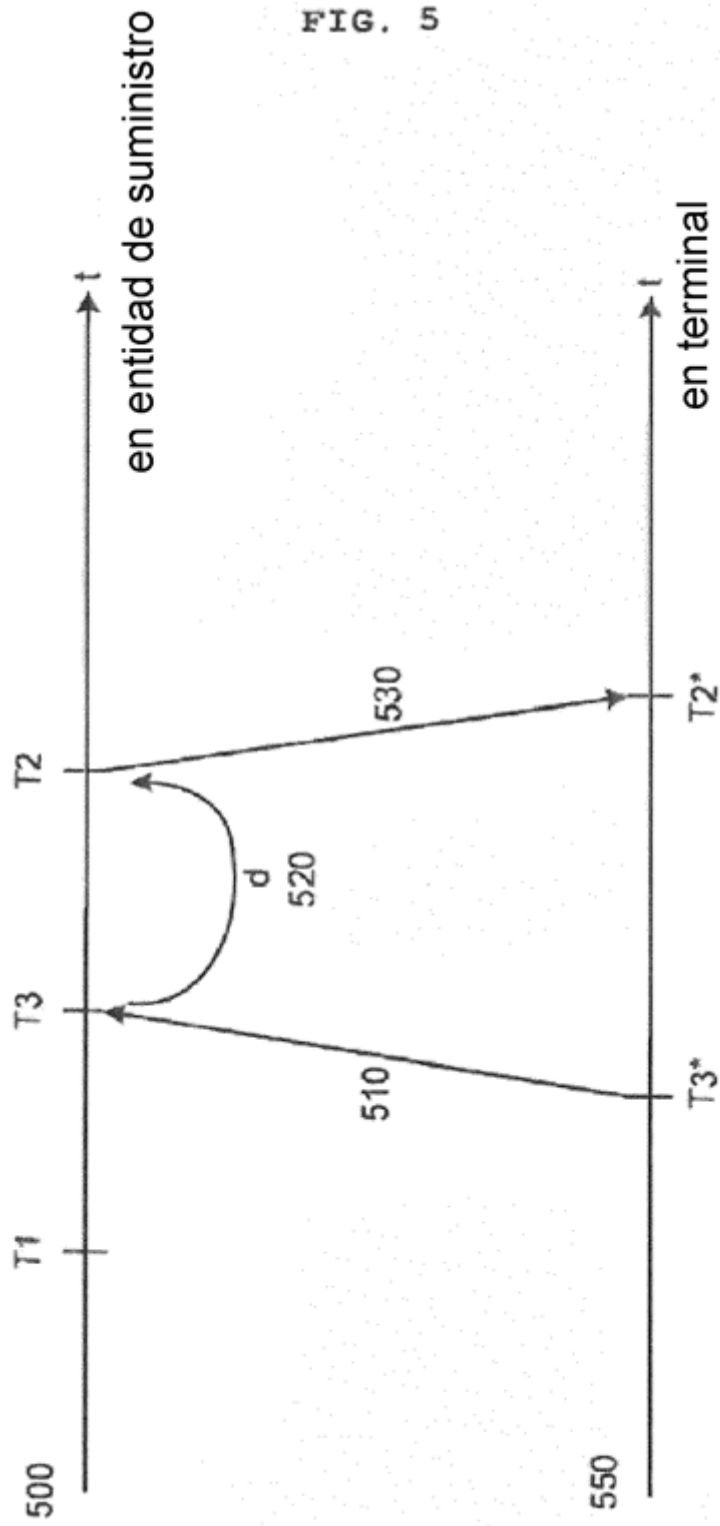


FIG. 6

600

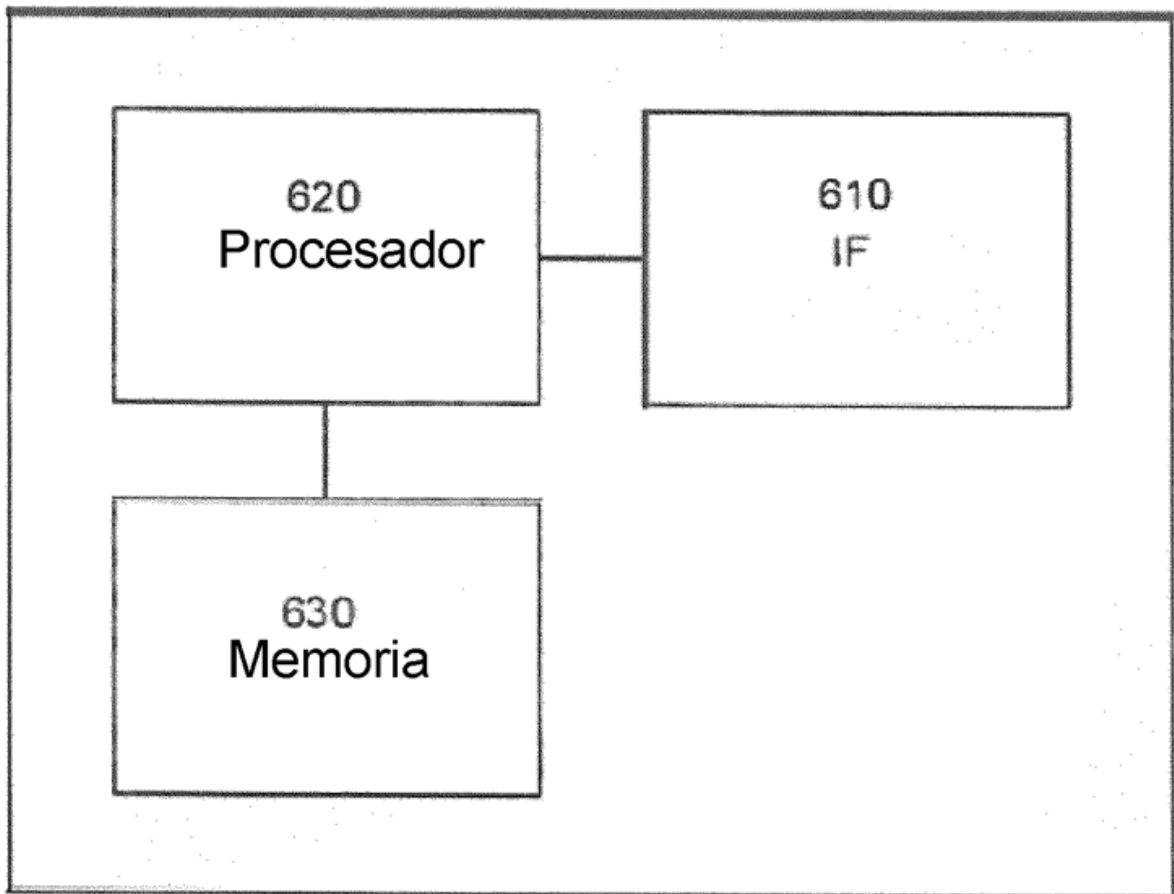


FIG. 7

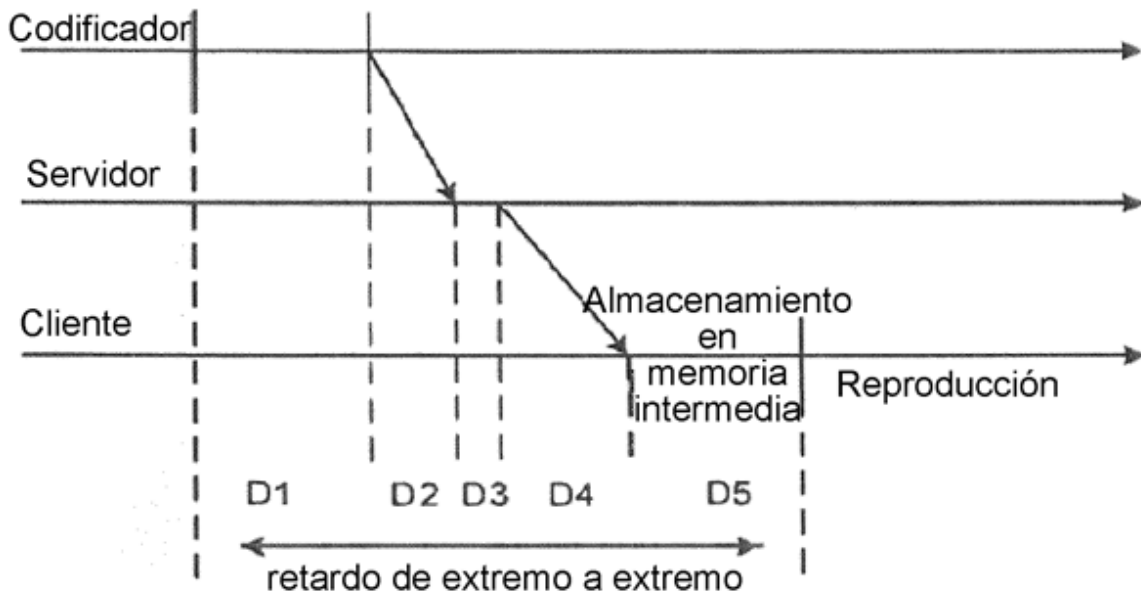


FIG. 8

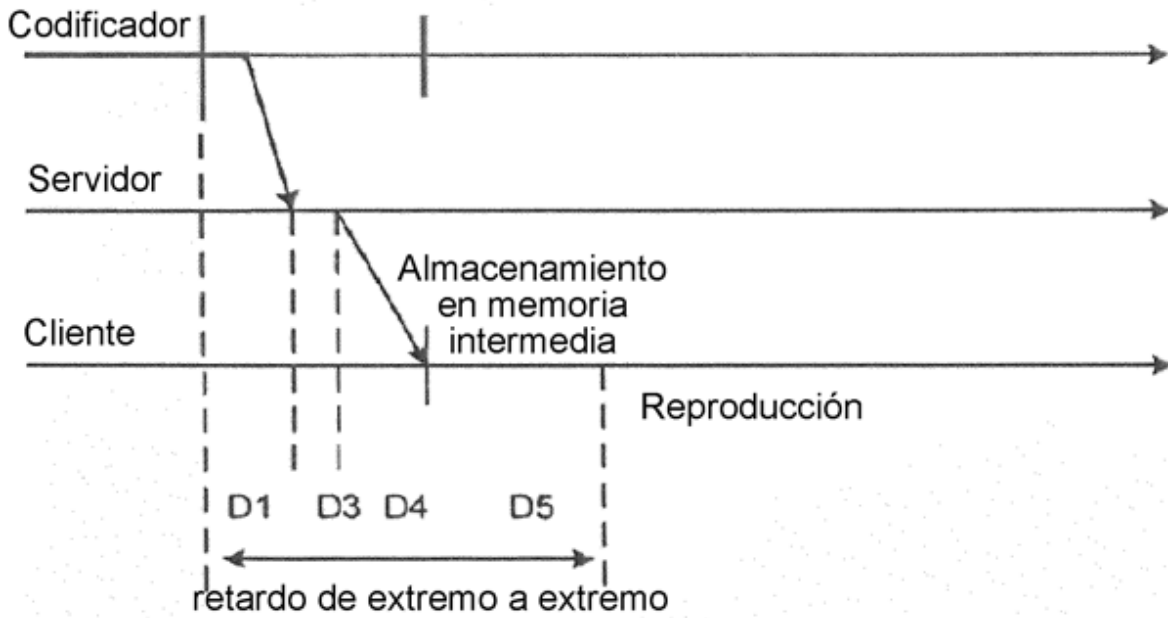


FIG. 9

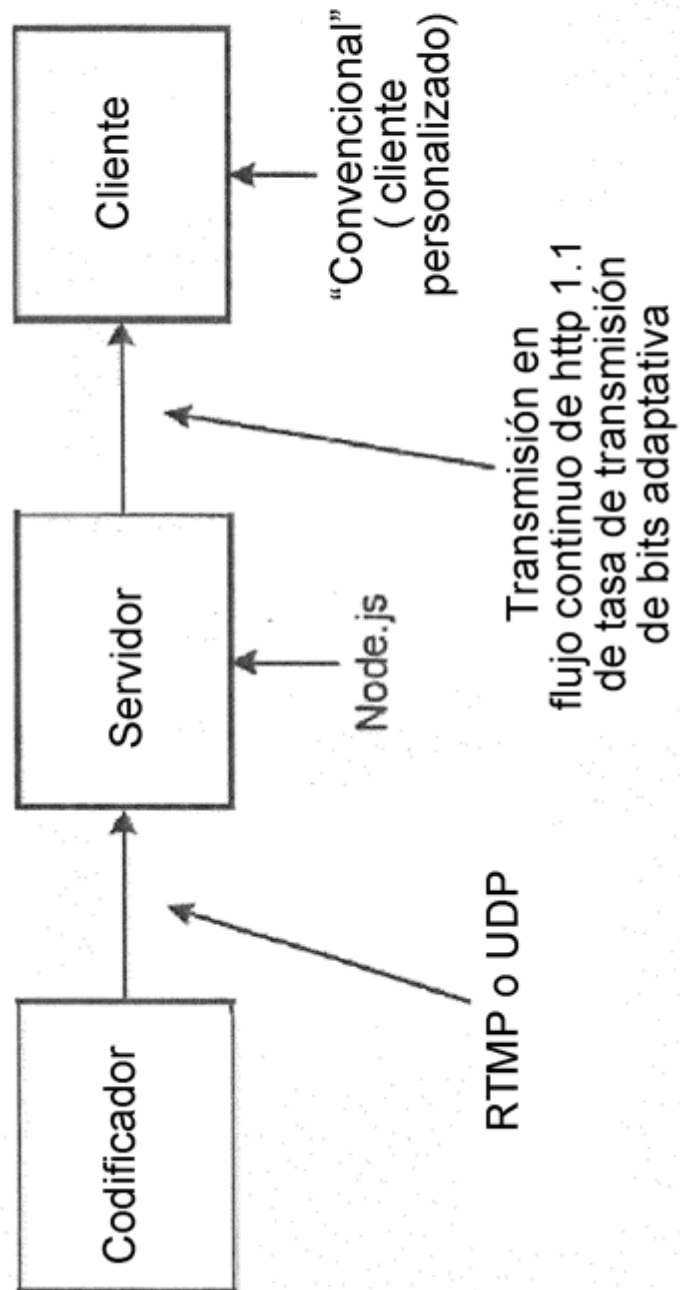


FIG. 10

