

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 810 859**

51 Int. Cl.:

H04N 21/2662 (2011.01)

H04N 21/2343 (2011.01)

H04N 21/845 (2011.01)

H04N 21/472 (2011.01)

H04N 21/462 (2011.01)

G06F 15/16 (2006.01)

H04L 29/06 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **31.12.2016 PCT/US2016/069655**

87 Fecha y número de publicación internacional: **06.07.2017 WO17117590**

96 Fecha de presentación y número de la solicitud europea: **31.12.2016 E 16882795 (4)**

97 Fecha y número de publicación de la concesión europea: **06.05.2020 EP 3398340**

54 Título: **Maximización de la calidad de servicio para la difusión en continuo de vídeos que se adaptan a la QOS a través de la configuración dinámica de la tasas de flujo en la capa de aplicaciones**

30 Prioridad:

31.12.2015 US 201514986551

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

09.03.2021

73 Titular/es:

**HUGHES NETWORK SYSTEMS, LLC (100.0%)
11717 Exploration Lane
Germantown, MD 20876, US**

72 Inventor/es:

**SU, CHI-JIUN;
JAIN, KAUSTUBH y
HONG, SE, GI**

74 Agente/Representante:

ARPE FERNÁNDEZ, Manuel

ES 2 810 859 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

Maximización de la calidad de servicio para la difusión en continuo de vídeos que se adaptan a la QOS a través de la configuración dinámica de las tasas de flujo en la capa de aplicaciones

5

ANTECEDENTES

[0001] Las capacidades globales de los satélites de banda ancha están aumentando exponencialmente, y tales aumentos de capacidad presentan desafíos singulares en el sistema terrestre asociado y los diseños de red. El objetivo de los diseñadores de sistemas, operadores de sistemas y proveedores de servicios es apoyar y proporcionar servicios eficientes, robustos, confiables y flexibles, en un entorno de red de ancho de banda compartido, utilizando sistemas satelitales de alta capacidad.

10

[0002] Según estudios recientes de tráfico de Internet, el tráfico de difusión en continuo de media (por ejemplo, difusión en continuo de video) constituye más del 50% del ancho de banda del enlace de retransmisión desde servidores a dispositivos de cliente. Además, la tendencia avanza a medida que más y más proveedores de contenido comienzan a ofrecer servicios de difusión en continuo de media (por ejemplo, video). Por ejemplo, adiciones recientes incluyen HBO, CBS y otros proveedores de red y servicios de difusión en continuo de contenido. Cuando un usuario mira o consume un video, si el video se almacena en un dispositivo de almacenamiento local (por ejemplo, una memoria caché local), el contenido de difusión en continuo de video se proporciona directamente desde la ubicación de almacenamiento local. Alternativamente, cuando el contenido de video no reside en un dispositivo de almacenamiento local, el contenido de difusión en continuo se proporciona a través de una red de comunicaciones de área amplia o WAN (por ejemplo, Internet) desde un servidor de contenido remoto. Cuando el video se proporciona al usuario o dispositivo/aplicación cliente a través de difusión en continuo de video adaptativa, el dispositivo/aplicación cliente del usuario (por ejemplo, una aplicación de reproductor de video que se ejecuta en un ordenador personal del cliente u otro dispositivo de media) selecciona una tasa de reproducción y recupera segmentos de video de la tasa de reproducción respectiva desde el servidor de contenido a través de un protocolo de solicitud/respuesta. Además, tales dispositivos/aplicaciones de reproducción de cliente típicamente almacenan una cierta cantidad de contenido para proporcionar el contenido desde la memoria intermedia local a una tasa constante (por lo tanto, no teniendo que depender de una tasa de entrega constante a través de la WAN). Además, cuando el contenido de difusión en continuo se proporciona desde un servidor de contenido remoto a través de una WAN, generalmente hay varios saltos de enlaces de comunicaciones entre el dispositivo cliente y el servidor de contenido remoto. Dichos enlaces de comunicaciones son compartidos por múltiples usuarios, donde los niveles de congestión sobre esos enlaces cambian dinámicamente con el tiempo, y el ancho de banda disponible para un usuario en particular también cambia dinámicamente con el tiempo. Además, si uno o más de los enlaces de comunicaciones se realizan a través de un medio de transmisión de radio, tal como enlaces celulares o satelitales, las condiciones del enlace de radio, tales como distancia, meteorología e interferencias, también pueden cambiar dinámicamente, lo que impone un mayor impacto en el ancho de banda disponible. En consecuencia, cuando cualquiera de los enlaces a través de los cuales se proporciona el contenido de difusión en continuo al dispositivo cliente disminuye por debajo de ciertos niveles, la tasa de reproducción seleccionada puede no ser compatible de manera constante durante la reproducción de video. Como resultado, la memoria intermedia de reproducción puede quedarse sin datos de video y la reproducción de video puede detenerse u oscilar o saltar. Es posible que el reproductor tenga que seleccionar una tasa más baja y volver a almacenar el contenido de video para reanudar una reproducción continua uniforme. Cuando la condición del enlace fluctúa, la tasa de reproducción también puede alternar entre dos tasas. Para resolver estos problemas, actualmente se proporcionan enfoques de configuración de tráfico y vigilancia para aplicaciones de difusión en continuo de media.

15

20

25

30

35

40

45

[0003] La conformación de tasa se lleva a cabo generalmente en la capa de IP por flujo de IP. Es una tecnología asentada compatible con dispositivos de red comercial y con el sistema operativo Linux. Si bien la configuración de la tasa en la capa IP se puede configurar para las direcciones de avance y retroceso, la tasa configurada puede ser diferente de la tasa de reproducción de video objetivo, especialmente cuando el tiempo de ida y vuelta (RTT) es uno de los factores dominantes, lo que determina el flujo de capa de aplicación. Además, el dispositivo de conformación de tasa generalmente se coloca en la red ISP y, por lo tanto, está demasiado lejos del dispositivo de reproducción del cliente para medir con precisión el RTT, especialmente cuando la mayor contribución al RTT se realiza mediante el enlace de última milla a las instalaciones/dispositivos cliente. Además, la configuración de tasa en la capa IP puede no ser reactiva en respuesta a las condiciones cambiantes de la red. Como resultado de ello, puede hacer que la tasa de reproducción sea más baja de lo que debería ser.

50

55

[0004] La conformación de la tasa se puede realizar en la capa TCP por flujo TCP para aplicaciones TCP. Se puede establecer dinámicamente un límite superior en la ventana de congestión TCP del remitente para limitar la tasa, lo que se conoce como "bloqueo de la ventana de congestión". Dado que está configurado en el remitente TCP, el dispositivo de configuración de tasa puede ubicarse en el servidor de contenido o en un punto medio entre un reproductor y un servidor de contenido. El llevar a cabo bloqueos de ventana de congestión TCP en un punto medio, es un enfoque solo se puede realizar cuando se despliega un proxy TCP. Un proxy TCP divide una conexión TCP de extremo a extremo en dos o tres conexiones TCP con TCP optimizado para cada salto. El rendimiento de TCP es directamente proporcional a la ventana de congestión del remitente e inversamente proporcional a RTT. La ventana de congestión TCP se establece en términos del tamaño máximo de segmento TCP, que normalmente es de 1460 bytes. Sin embargo, a un RTT bajo entre el RSE y el reproductor, este enfoque ofrece solo una granularidad de

60

65

control baja para la configuración de la tasa. El enfoque tampoco tiene en cuenta el RTT real entre el reproductor y el servidor de contenido, que puede ser un contribuyente importante al rendimiento de la aplicación.

[0005] Por ejemplo, ciertos fabricantes actuales de equipos de red incluyen procesos de conformación de tráfico y vigilancia en equipos de red (por ejemplo, conmutadores y enrutadores). Consúltese, por ejemplo, Cisco, "Policing and Shaping Overview" Guía de configuración de soluciones IOS QOS, entrega 12.2 www.cisco.com/c/en/us/td/docs/ios/12_2/qos/configuration/guide/fqos_c/qcf-polsh.pdf. Dichos enfoques emplean la configuración de tasa de flujo por capa o por interfaz del Protocolo de Internet (IP) utilizando mecanismos *token bucket*. Tres parámetros básicos de dichos enfoques *token bucket* son la tasa de información comprometida (CIR), que es la tasa media formada, el tamaño de ráfaga comprometida (BC), que es la cantidad de bytes que se pueden enviar durante un intervalo de tiempo de medición y un intervalo de tiempo de medición. Sin embargo, estos enfoques tienen en cuenta el tiempo de ida y vuelta (RTT), no proporcionan la configuración del flujo de capa de aplicación y, por lo general, se despliegan dentro de la red del proveedor de servicios de Internet (ISP).

[0006] Como ejemplo adicional, el sistema operativo Linux, por defecto, proporciona funciones de control de tráfico tales como qdisc, filtro y clase, para realizar la configuración de tasa por flujo de tráfico. Véase, por ejemplo, Doru Gabriel Balan, Dan Alin Potorac, "Extended Linux HTB Queuing Discipline Implementations," Primera Conferencia Internacional sobre Tecnologías Digitales en Red, IEEE, páginas 122-126, 28 de julio de 2009. Sin embargo, el enfoque de Linux tampoco tiene en cuenta el tiempo de ida y vuelta (RTT) y no proporciona la configuración de la tasa de capa de aplicación.

[0007] Se propuso la conformación de la tasa de QoS para ayudar a un reproductor de aplicación cliente a seleccionar una resolución adecuada en condiciones de red fluctuantes en la red de evolución a largo plazo (LTE). Ver, por ejemplo, Hyunwoo Nam, Kyung Hwa Kim, Bong Ho Kim, Doru Calin, Henning Schulzrinne, "Towards dynamic QoS-aware over-the-top video streaming" 15º Simposio Internacional IEEE en World of Wireless, Mobile and Multimedia Networks, páginas 1-9, 19-jun-2014. Con este enfoque, la configuración de tasa establece un ancho de banda máximo permitido por flujo en un eNodeB, que se encuentra antes de la última milla dentro de la red LTE. Sin embargo, este enfoque no considera el tiempo de ida y vuelta (RTT) entre una aplicación cliente (por ejemplo, un reproductor de video) y el servidor de contenido, y no configura el flujo de capa de aplicación.

[0008] La limitación de tasa en un servidor de contenido se ha propuesto para reducir la cola y la pérdida de paquetes para YouTube. Ver, 15 por ejemplo, Monia Ghobadi, Yuchung Cheng, Ankur Jain, Matt Mathis, "Trickle: rate limit Youtube video streaming", Conferencia técnica anual de USENIX, Asociación USENIX, Berkeley, CA, EE. UU., páginas 191-196, 2012. Con la finalidad de eliminar grandes ráfagas causadas por los mecanismos de estimulación de aplicaciones existentes, el enfoque de YouTube establece dinámicamente una ventana de congestión máxima del remitente TCP y limita estrictamente el tamaño máximo de las ráfagas. Sin embargo, este enfoque tiene deficiencias tales como una baja granularidad en la capacidad de control de la tasa, y la configuración de la tasa se emplea remotamente desde el dispositivo/aplicación cliente y, por lo tanto, no configura el flujo de capa de aplicación.

[0009] La técnica anterior US 2013/097309 A1 presenta un caché proxy de video de flujo HTTP con adaptación de tasa controlada y reducción de la tasa de reproducción.

[0010] La técnica anterior US 2015/026309 A1 se ocupa del control de difusión en continuo adaptable para un dispositivo de red que supervisa las condiciones de la sesión de medios para una sesión de medios con un cliente de media servido por la red y aplica la política sobre la sesión de medios en función de las condiciones de la sesión de medios. Se determina una calidad de medios, objetivo para la sesión de medios y modifica la sesión de medios en respuesta a una diferencia entre la calidad de media estimada actual y la calidad de media objetivo.

[0011] Por lo tanto, lo que se necesita es un enfoque para llevar a cabo de manera efectiva y precisa la conformación del flujo de capa de aplicación en la difusión en continuo de video adaptable para proporcionar niveles maximizados de Calidad de Servicio (QoS) al dispositivo/aplicaciones cliente del usuario (por ejemplo, reproductor de video) .

ALGUNOS EJEMPLOS DE REALIZACIÓN

[0012] La presente invención aborda ventajosamente los requisitos y necesidades anteriores, así como otros, proporcionando un aparato (según la reivindicación 3) y un procedimiento (según la reivindicación 1) que facilitan el rendimiento efectivo y preciso de la conformación del flujo de capa de aplicación en difusión en continuo de video adaptativa para proporcionar niveles maximizados de calidad de servicio (QoS) al dispositivo/aplicaciones cliente del usuario (por ejemplo, reproductor de video)

[0013] Se puede emplear un elemento de configuración de tasa (RSE) para configurar el flujo de aplicación de manera que el reproductor seleccione una tasa de reproducción que se pueda mantener de manera constante y se pueda proporcionar de manera continua la mejor QoS para el consumo/reproducción de usuario.

[0014] Esta invención presenta un enfoque novedoso para realizar de manera efectiva y precisa la conformación del flujo de capa de aplicación en una difusión en continuo de video adaptativa para proporcionar a los usuarios la mejor QoS. Puede emplearse en un sistema descrito en la descripción de la invención [1] para servir el video a una tasa binaria particular que se almacena en la memoria caché. También se puede usar configurar la tasa de reproducción de video que pueda mantenerse durante toda la reproducción sin nuevo almacenamiento intermedio, cuando el enlace de comunicación entre el cliente y el servidor de contenido no puede proporcionar el ancho de banda requerido de manera consistente cuando el video no se sirve desde la caché sino desde el contenido servidor.

[0015] La invención propone un elemento de configuración de tasa (RSE) que mide dinámicamente el flujo de capa aplicación visto por un reproductor de video que estima la tasa de reproducción actual del reproductor. El flujo de

capa aplicación no solo está determinado por el ancho de banda de un servidor de contenido a un reproductor (ancho de banda de enlace directo) sino también por el ancho de banda de un reproductor a un servidor de contenido (ancho de banda de enlace de retorno) y por el tiempo de ida y vuelta (RTT) desde el momento en que se envía una solicitud de contenido hasta el momento en que el reproductor recibe la respuesta. El flujo de capa aplicación es uno de los parámetros clave utilizados por el reproductor de difusión en continuo de video adaptativo para seleccionar una tasa de reproducción de video particular. El RSE propuesto en esta invención se encuentra en la premisa de cliente. Da forma al flujo de la aplicación para que la tasa binaria seleccionada por los reproductores de difusión en continuo de video adaptativos coincida con la del video almacenado en la caché o con las características de los enlaces de comunicación entre el reproductor y el servidor de contenido cuando el video no se sirve desde la caché si no desde el servidor de contenido.

[0016] El RSE supervisa las transferencias de solicitudes HTTP para segmentos de video desde un reproductor hacia un servidor de contenido y las respuestas HTTP con segmentos de video desde un servidor de contenido hacia un reproductor. Mantiene una nota de lo siguiente: el momento en que recibe una petición HTTP de un jugador antes de reenviarlo al servidor de contenido, el momento en que recibe una respuesta HTTP correspondiente del servidor de contenido antes de reenviarlo al cliente y la longitud del contenido de El segmento de video contenido en la respuesta HTTP. Luego calcula el flujo HTTP de un segmento de video. Además, calcula la estimación de tasa de reproducción actual del reproductor utilizando la longitud del contenido de los segmentos y la duración del segmento de video. A partir de las estadísticas de las mediciones de flujo HTTP anteriores y el requisito de tasa de reproducción de video configurado, toma una decisión sobre si se requiere o no un ajuste de tasa de reproducción. Si la tasa debe ser menor, agregará un retardo apropiado antes de transmitir la respuesta HTTP al reproductor para que flujo HTTP vista por el reproductor se reduzca adecuadamente. De lo contrario, la respuesta HTTP se enviará al reproductor inmediatamente.

[0017] Las técnicas se extienden luego al caso en el que la difusión en continuo de video adaptativa sea servida a través de HTTPS.

[0018] De acuerdo con las realizaciones de ejemplo, se proporciona un procedimiento para la conformación efectiva y precisa del flujo de capa aplicación en la difusión en continuo de video adaptativa para proporcionar niveles maximizados de Calidad de Servicio (QoS) al dispositivo/aplicaciones cliente del usuario. Un dispositivo de configuración de tasas de flujo de capa aplicación recibe un mensaje de petición de contenido de un dispositivo cliente respectivo, el mensaje de petición de contenido solicita un segmento de contenido actual de una secuencia de segmentos de contenido de un archivo de datos de transmisión para una sesión de datos de transmisión del dispositivo cliente, y se registra un momento de recepción del mensaje de petición de contenido y la petición de contenido se reenvía a un servidor de contenido remoto. El dispositivo de conformación de la tasas de flujo de capa aplicación recibe un mensaje de respuesta de contenido transmitido por el servidor de contenido en respuesta a la petición de contenido retransmitida, incluyendo la respuesta de contenido el segmento de contenido actual solicitado, y se registra un momento de recepción del mensaje de respuesta de contenido, se registra una longitud de contenido, se determina un tamaño de segmento del segmento actual y se determina una tasas de flujo de capa aplicación (ALT) actual para el segmento actual en función del momento de recepción del mensaje de petición de contenido, el momento de recepción del mensaje de respuesta de contenido y el tamaño del segmento. El dispositivo de configuración de la tasas de flujo de capa aplicación determina si la tasa de ALT actual es mayor que la tasa de ALT objetivo requerida para soportar una tasa de reproducción (PBR) deseada para la sesión de transmisión de datos del dispositivo cliente. Cuando se determina que la tasa de ALT actual es mayor que la tasa ALT objetivo, el mensaje de respuesta de contenido se mantiene durante un tiempo de retención determinado para reducir la tasa de ALT actual a la tasa de ALT objetivo, y la respuesta de contenido se proporciona al dispositivo cliente tras la finalización del tiempo de espera. Cuando se determina que la tasa de ALT actual no es mayor que la tasa de ALT objetivo, la respuesta de contenido se proporciona rápidamente al dispositivo cliente.

[0019] Según otro ejemplo de realización del procedimiento, como parte de la etapa de recepción del mensaje de respuesta de contenido, se determina uno de entre un índice de reproducción medido y estimado (PBR) asociado con el segmento de contenido actual de la sesión de difusión en continuo de datos del dispositivo del cliente. Después de la etapa de recepción del mensaje de respuesta de contenido, se determina si la PBR medida o estimada es mayor que la PBR deseada, el procedimiento pasa a la etapa de determinar si la tasa de ALT actual es mayor que la tasa de ALT objetivo. Cuando se determina que la PBR medida o estimada no es mayor que la PBR deseada, el procedimiento comprende además proporcionar rápidamente la respuesta de contenido al dispositivo cliente, y la etapa de determinar si la tasa de ALT actual es mayor que la tasa de ALT objetivo, junto con las etapas asociadas con el resultado de esta determinación, no se realizan.

[0020] De acuerdo con otras realizaciones ejemplares, se proporciona un dispositivo para configuración efectiva y precisa de la transmisión capa aplicación en la difusión en continuo de video adaptable para proporcionar al dispositivo/aplicaciones cliente del usuario niveles maximizados de Calidad de Servicio (QoS). El dispositivo comprende una interfaz de dispositivo de cliente, una interfaz de terminal de comunicaciones de red y un elemento de configuración de tasa. La interfaz del dispositivo cliente es utilizable para recibir un mensaje de petición de contenido desde un dispositivo cliente respectivo, el mensaje de petición de contenido solicita un segmento de contenido actual de una secuencia de segmentos de contenido de un archivo de transmisión de datos para una sesión de transmisión de datos del dispositivo cliente. La interfaz del terminal de comunicaciones de red es utilizable para retransmitir la petición de contenido a un servidor de contenido remoto y para recibir un mensaje de respuesta de contenido transmitido por el servidor de contenido en respuesta a la petición de contenido retransmitida, la respuesta de contenido incluye el segmento de contenido actual solicitado. El elemento de configuración de tasa es

utilizable para registrar una hora de recepción del mensaje de petición de contenido, para registrar una hora de recepción del mensaje de respuesta de contenido, para registrar una longitud de contenido y un tamaño de segmento del segmento actual, y para determinar una tasa de flujo de capa aplicación actual (ALT) para el segmento actual a partir del tiempo de recepción del mensaje de petición de contenido, el tiempo de recepción del mensaje de respuesta de contenido y el tamaño del segmento. El elemento de conformación de tasa es utilizable para determinar si la tasa de ALT actual es mayor que una tasa de ALT objetivo requerida para soportar una tasa de reproducción deseada (PBR) para la sesión de difusión en continuo de datos del dispositivo cliente. Cuando se determina que la tasa de ALT actual es mayor que la tasa de ALT objetivo, el elemento de configuración de tasa puede seguir funcionando para retener el mensaje de respuesta de contenido durante un tiempo de retención determinado para reducir la tasa de ALT actual a la tasa de ALT objetivo y liberar la respuesta de contenido al finalizar el tiempo de retención, y aún pudiendo funcionar además la interfaz del dispositivo cliente para proporcionar la respuesta de contenido al dispositivo del cliente una vez que el elemento de configuración de tasa libera la respuesta. Cuando se determina que la tasa ALT actual no es mayor que la tasa ALT objetivo, el elemento de configuración de tasa se puede hacer funcionar aún además para liberar rápidamente la respuesta de contenido, y la interfaz del dispositivo cliente se puede hacer funcionar aún además para proporcionar la respuesta de contenido al dispositivo cliente una vez la respuesta es liberada por el elemento de configuración de tasa.

[0021] De acuerdo con una realización ejemplar adicional del dispositivo, el elemento de configuración de tasa es utilizable además para determinar una de entre una tasa de reproducción medida y estimada (PBR) asociada con el segmento de contenido actual de la sesión de difusión en continuo de datos del dispositivo cliente. Por lo tanto, antes de realizar la determinación de si la tasa ALT actual es mayor que una tasa ALT objetivo, el elemento de conformación de tasa es además utilizable para determinar si la PBR medida o estimada es mayor que la PBR deseada. Cuando se determina que la PBR medida o estimada es mayor que la PBR deseada, el elemento de conformación de la tasa luego puede utilizarse para proceder a determinar si la tasa ALT actual es mayor que una tasa ALT objetivo. Cuando se determina que la PBR medida o estimada no es mayor que la PBR deseada, el elemento de conformación de tasa se puede utilizar aún adicionalmente para liberar rápidamente la respuesta de contenido, y la interfaz del dispositivo del cliente se puede utilizar aún además para proporcionar la respuesta de contenido al dispositivo del cliente una vez el elemento de conformación de tasa libera la respuesta, y el elemento de conformación de tasa no se utiliza para determinar si la tasa de ALT actual es mayor que una tasa de ALT objetivo, ni para las funciones asociadas con el resultado de esa determinación.

[0022] Dichas realizaciones de la presente invención facilitan la conformación dinámica del flujo de aplicación de una manera efectiva y precisa de modo que la tasa de reproducción de video seleccionada por el reproductor de difusión en continuo de video adaptativa coincida con la del video almacenado en la caché o con las características de enlaces de comunicación entre el reproductor y el servidor de contenido cuando el video no es servido desde la memoria caché sino desde el servidor de contenido. Dichas realizaciones de la presente invención facilitan aún más las operaciones transparentes sin requerir ninguna modificación en los servidores de contenido, los reproductores y el protocolo utilizado entre ellos.

[0023] De acuerdo con los enfoques de las realizaciones de ejemplo, se controla el dispositivo de reproducción cliente para elegir una tasa de reproducción requerida para proporcionar la mejor QoE (Calidad de Experiencia) a un usuario con las siguientes propiedades: (1) el elemento de configuración de tasa está ubicado cerca a los respectivos dispositivos de reproducción del cliente, lo que proporciona una estimación más precisa del flujo de capa aplicación tal como la ven los dispositivos de reproducción; (2) la tasa se configura en la capa de aplicación, lo que proporciona el impacto más directo en las decisiones de adaptación de tasa deseadas de los dispositivos de reproducción; (3) la configuración se realiza por sesión de aplicación, y no por flujo; y (4) el flujo de capa de aplicación es medida se mide y configurada de forma continua para proporcionar respuestas dinámicas a los cambios en las condiciones del enlace, y para no sobre-configurar una tasa de reproducción más baja.

[0024] Además, el flujo de capa de aplicación en la difusión en continuo de video adaptativa, de acuerdo con las realizaciones ejemplares, puede ser significativamente diferente del flujo de capa IP ya que depende de los siguientes factores: (1) Retardo de ida y vuelta (RTT) entre un reproductor y un servidor de contenido; (2) La aplicación de difusión en continuo de video adaptativa utiliza el protocolo de solicitud/respuesta (HTTP) para obtener cada segmento de video/audio; (3) la transmisión de TCP también depende de RTT; (4) La tasa disminuye a medida que aumenta RTT; (5) Ancho de banda del enlace de retorno (el enlace de retorno es un enlace de comunicación de un reproductor a un servidor de contenido); (6) Ancho de banda del enlace directo (el enlace directo es un enlace de comunicación desde un servidor de contenido a un reproductor); (7) Parámetros TCP tales como tamaño de ventana y tamaño máximo de segmento (MSS); (8) Interacción entre caída de paquetes TCP/IP y retardo y algoritmo de control de congestión TCP. La configuración de la tasa de capa IP puede controlar el enlace directo y el ancho de banda del enlace de retorno solamente.

[0025] Aún otros aspectos, características y ventajas de la presente invención resultarán fácilmente evidentes a partir de la siguiente descripción detallada, mediante la simple ilustración de una serie de realizaciones e ejecuciones particulares, incluyendo el mejor modo contemplado para llevar a cabo la presente invención. La presente invención también es capaz de otras y diferentes realizaciones, y sus diversos detalles pueden modificarse en varios aspectos obvios, todo ello sin apartarse del espíritu y el alcance de la presente invención. En consecuencia, los dibujos y la descripción deben considerarse de naturaleza ilustrativa y no restrictiva.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

[0026] Las formas de realización de ejemplo de la presente invención se ilustran a modo de ejemplo, y no de manera limitativa, en las figuras de los dibujos adjuntos, en los que los números de referencia similares se refieren a elementos similares, y en los que:

- 5 - La figura 1 ilustra un diagrama de bloques de un sistema de comunicaciones por satélite para proporcionar servicios de difusión en continuo de medios adaptativo, de acuerdo con realizaciones de ejemplo;
- La figura 2, ilustra un diagrama de bloques de un sistema de comunicaciones por satélite adicional para proporcionar servicios difusión en continuo de medios adaptativos, de acuerdo con realizaciones de ejemplares;
- La figura 3, ilustra los segmentos de video de una pluralidad de archivos de audio y video a tasa binaria de presentación o reproducción (PBR) diferente, de acuerdo con realizaciones de ejemplo;
- 10 - La figura 4 ilustra un diagrama de bloques de un dispositivo/aplicación de reproducción cliente, que emplea adaptación de tasa, en un sistema adaptativo de difusión en continuo de medios, de acuerdo con realizaciones de ejemplo;
- La figura 5 ilustra una secuencia de segmentos de video descargados por el dispositivo/aplicación de reproducción cliente de la figura 4, para una sesión de reproducción de video que emplea difusión en continuo de video adaptativa, de acuerdo con realizaciones de ejemplo;
- 15 - La figura 6 ilustra un sistema adaptativo de difusión en continuo de medios que emplea un elemento de conformación de tasa (RSE), de acuerdo con realizaciones de ejemplo;
- La figura 7 ilustra la señalización general de un sistema adaptativo de difusión en continuo de datos o medios, y el proceso de configuración de tasa de capa de aplicación del RSE, de acuerdo con realizaciones de ejemplo; y
- 20 - La figura 8 ilustra un diagrama de flujo que representa un algoritmo de conformación de capa aplicación realizado por el RSE, de acuerdo con realizaciones de ejemplo;
- La figura 9 ilustra un gráfico que representa un ejemplo de la aplicación de ALT y PBR a los segmentos de un archivo de video, de acuerdo con realizaciones de ejemplo;
- 25 - La figura 10 ilustra un diagrama de flujo que representa un algoritmo de conformación de capa de aplicación realizado por el RSE, sin la utilización de una estimación de tasa de reproducción actual ($PBR_{Estimate}$), de acuerdo con realizaciones de ejemplo;
- La figura 11 ilustra un diagrama de flujo que representa un algoritmo para estimación de la longitud del contenido de un segmento de video basada en los paquetes de capa de registro TLS recibidos, de acuerdo con realizaciones de ejemplo;
- 30 - La figura 12 ilustra un gráfico que representa una distribución de tamaño de segmento de ejemplo para un archivo de audio de una tasa de reproducción específica, y para varios archivos de video, cada uno de una tasa de reproducción específica (PBR), de acuerdo con realizaciones de ejemplo;
- La figura 13 ilustra un diagrama de flujo que representa un algoritmo para estimación de la tasa de reproducción actual, de acuerdo con realizaciones de ejemplo;
- 35 - La figura 14 ilustra un diagrama de bloques de un conjunto de placas de circuitos que ejecutan aspectos de enfoques de almacenamiento en caché oportunistas, de acuerdo con realizaciones de ejemplo; y
- La figura 15 ilustra un diagrama de bloques de un sistema de ordenador que ejecuta aspectos de enfoques de almacenamiento en caché oportunistas, de acuerdo con realizaciones de ejemplo.

40 DESCRIPCIÓN DETALLADA

[0027] Se proporcionan sistemas y procedimientos que facilitan el rendimiento efectivo y preciso de la configuración de transmisión de capa de aplicación en la difusión en continuo de video adaptable para proporcionar niveles maximizados de calidad de servicio (QoS) al dispositivo/aplicaciones cliente del usuario (por ejemplo, reproductor de video). En la siguiente descripción, para fines explicativos, se exponen numerosos detalles específicos con el fin de proporcionar una comprensión exhaustiva de la presente invención. La presente invención no pretende estar limitada por medio de las realizaciones descritas, y varias modificaciones serán fácilmente aparentes. Resultará evidente que la invención se puede practicar sin los detalles específicos de la siguiente descripción y/o con disposiciones equivalentes. Además, se pueden mostrar estructuras y dispositivos bien conocidos en forma de diagrama de bloques para evitar oscurecer innecesariamente la invención. Además, las aplicaciones específicas discutidas aquí se proporcionan solo como ejemplos representativos, y los principios descritos aquí pueden aplicarse a otras realizaciones y aplicaciones sin apartarse del alcance general de la presente invención.

[0028] La configuración de la tasa para la difusión en continuo de video adaptativa es diferente de la configuración de la tasa en general, ya que la aplicación de difusión en continuo de video adaptativa funciona de manera única. Un reproductor de difusión en continuo de video adaptativo selecciona una tasa de reproducción a partir del flujo de capa de aplicación que depende del tiempo de ida y vuelta entre un reproductor y un servidor de contenido y el ancho de banda del enlace de retorno, además del ancho de banda del enlace directo. Además, hay un conjunto discreto bien definido de tasas de reproducción disponibles para que un reproductor las seleccione para la difusión en continuo adaptativa de video, y siendo dinámica la adaptación de la tasa: el reproductor mide continuamente el flujo de capa de aplicación para ajustar dinámicamente la tasa de reproducción. La configuración de la tasa sin tener en cuenta los hechos anteriores puede llevar a un reproductor a seleccionar una tasa más baja de manera innecesaria.

[0029] Existen dos factores de diseño principales en la configuración de la tasa para difusión en continuo adaptativa de video: la ubicación de las funciones del elemento de configuración de la tasa dentro de la pila de protocolos de Internet y la ubicación física del elemento de configuración de la tasa (RSE). Por ejemplo, la configuración de tasa de RSE se puede realizar en diferentes capas de la pila de protocolos de Internet: IP, TCP y capas de aplicación.

Además, el RSE se puede ubicar en varias ubicaciones de red, tal como en el servidor de contenido, dentro de una red de proveedor de servicios de Internet (ISP) y dentro de la red local del cliente. El RSE en la red de las instalaciones del cliente proporciona la ubicación más precisa para determinar y controlar el flujo de capa de aplicación como lo ve el dispositivo de reproducción cliente. El flujo de capa de aplicación está determinado en conjunto por la *First Mile* (proveedor de contenido para ISP), la *Middle Mile* (red de ISP) y la *Last Mile* (ISP según las premisas del cliente).

[0030] La figura 1 ilustra un diagrama de bloques de un sistema de comunicaciones por satélite para proporcionar servicios de difusión adaptativa en continuo de medios, de acuerdo con realizaciones de ejemplo de la presente invención. El sistema de comunicaciones por satélite 130 incluye un satélite 132 que admite comunicaciones entre múltiples terminales de satélite (ST) 134 (1)a -134(1)n y 134 (2)a - 134(2)n, una serie de pasarelas (GW) 138a - 138n, y un centro de operaciones de red (NOC) 142. Los ST, GW y NOC transmiten y reciben señales a través de las antenas 136(1)a-136(1)n y 136(2)a-136(2)n, 146a-146n, y 156, respectivamente. De acuerdo con diferentes realizaciones, el NOC 142 puede residir en una ubicación separada accesible a través de un canal de satélite separado o puede residir dentro de una ubicación GW, y alternativamente, el NOC 142 puede comprender un sistema distribuido que se distribuye entre una pluralidad de sitios para propósitos de distribución procesamiento, escalabilidad, etc. Además, como se muestra, en un sistema de haz puntual, el satélite se comunica con las diversas infraestructuras terrestres (ST, GW y NOC) a través de una pluralidad de haces puntuales de enlace ascendente y descendente 110a, 110b, 110c, 110d, 110e. A modo de ejemplo, el NOC y cada GW pueden ser atendidos por un par de haz puntual de enlace ascendente y descendente dedicado, y cada uno de los grupos de ST (por ejemplo, el grupo de ST 134(1)a-134(2)n y el grupo de los ST 134(2)a-134(2)n) puede ser atendido como grupo por un par de haz puntual dedicado de enlace ascendente y descendente - una arquitectura de ancho de banda compartida, donde los ST comparten el respectivo ancho de banda de enlace ascendente y descendente de un haz puntual en los mismos. Cada haz de enlace ascendente y haz de enlace descendente cubre un área geográfica respectiva en la Tierra que está configurada de acuerdo con los diseños de antena respectivos de las antenas de enlace ascendente y descendente del satélite. Además, el haz de enlace ascendente de un área de cobertura particular no necesita ser contiguo con un haz de enlace descendente respectivo de esa área, sino que, por ejemplo, el área de cobertura de un haz de enlace ascendente particular puede dividirse y cubrirse con múltiples haces puntuales de enlace descendente. Además, los haces puntuales de enlace ascendente y descendente también pueden ser escalables para el ajuste dinámico de los planes de capacidad en función de una distribución de demanda de capacidad geográfica que cambia con el tiempo.

[0031] El NOC 142 realiza las funciones de plano de gestión del sistema 130, mientras que las GW 138a-138n realizan las funciones de plano de datos del sistema 130. Por ejemplo, el NOC 142 realiza funciones tales como gestión y configuración de red, descargas de software (p. ej., a los ST 134a-134n), supervisión de estado, funciones estadísticas (p. ej., recopilación, agregación e informes), funciones de seguridad (p. ej., generación, gestión y distribución de claves), registro y autenticación de ST, y gestión de diversidad de GW. El NOC 142 se comunica con cada GW a través del satélite 132, o a través de una red segura de comunicaciones privadas 152 (por ejemplo, un túnel IPsec a través de un enlace dedicado o una red privada virtual (VPN) o un túnel IPsec a través de una red pública, como Internet) Debe observarse que, según un ejemplo de realización, los enfoques de clasificación de tráfico de las realizaciones de la presente invención abordan la clasificación del tráfico de datos que fluye a través de un punto o nodo de agregación. Además, cada GW y el NOC tienen conectividad a una o más redes de comunicaciones públicas, como Internet o una PSTN.

[0032] Según otro ejemplo de realización, cada una de las GW 138a-138n incluye una o más pasarelas IP (IPGW), por lo que las funciones del plano de datos se dividen entre una GW y sus respectivas IPGW. Por ejemplo, GW 138a incluye IPGW 148a(1)-148a(n) y GW 138n incluye IPGW 148n(1)-148n(n). Una GW puede realizar funciones tales como codificación y modulación de salida de capa física y capa de enlace (p. ej., codificación y modulación adaptativa DVB-S2), manejo de ruta de capa de enlace y capa física (p. ej., IPOS), asignación de ancho de banda en ruta y equilibrado de carga, priorización de salida, aceleración web y compresión HTTP, control de flujo, cifrado, cambios de redundancia y aplicación de políticas de restricción de tráfico. Considerando que, la IPGW puede realizar funciones tales como compresión de datos, mejoras de rendimiento de TCP (por ejemplo, *proxies* de mejora de rendimiento de TCP, como falsificación de TCP), funciones de calidad de servicio (por ejemplo, clasificación, priorización, diferenciación, detección temprana aleatoria (RED), TCP/Control de flujo UDP), control de uso de ancho de banda, equilibrio dinámico de carga y enrutamiento. Además, una GW y una IPGW respectiva se pueden colocar con el NOC 142. Los ST 134a-134n proporcionan conectividad a uno o más hosts 144a-144n y/o enrutadores 154a-154n, respectivamente. A modo de ejemplo, el sistema de comunicaciones por satélite 130 puede funcionar como un sistema de transpondedor *bent-pipe*, donde el satélite opera esencialmente como un repetidor o transpondedor *bent-pipe*. En un sistema transpondedor *bent-pipe* de una realización de ejemplo, el satélite 132 funciona como un repetidor o *bent-pipe*, por lo que las comunicaciones hacia y desde los ST 134a-134n se transmiten a través del satélite 132 hacia y desde las IPGW respectivas asociadas con ST particulares. Además, en un sistema de haz puntual, cualquier haz puntual funciona como un *bent-pipe* hacia una región geográfica cubierta por el haz. Por ejemplo, cada haz puntual funciona como un canal de comunicaciones *bent-pipe* hacia y desde los ST y/o IPGW (s) dentro de la región geográfica cubierta por el haz. En consecuencia, las transmisiones de señal al satélite provienen de un ST y están destinadas a una pasarela asociada, o de una pasarela y están destinadas a un ST asociado. De acuerdo con una realización, se distribuyen varias GW/IPGW a través de la región geográfica cubierta por todos los haces puntuales del satélite 132, donde, en un haz en el que se ubican una GW (y sus respectivas IPGW), solo una GW (y ningún ST) ocupa esa haz. Además, cada IPGW puede servir como un nodo de agregación para una multitud de nodos remotos o ST. El número total de GW/IPGW y la distribución geográfica de las GW/IPGW

dependen de varios factores, como la capacidad total del satélite dedicado al tráfico de datos, la carga del tráfico geográfico del sistema (por ejemplo, en función de las densidades de población y la distribución geográfica de los ST), ubicaciones de los centros de datos terrestres disponibles (por ejemplo, troncales de datos terrestres para el acceso a redes dedicadas públicas y privadas). Por ejemplo, el contenido (p. ej., contenido de difusión en continuo de video, tal como distribución en continuo de una película) que se origina en un servidor de contenido (no mostrado) se puede proporcionar a la GW 138a y a la IPGW 148a (1). Entonces (a través de un haz de difusión o haz puntual del satélite 132) el contenido puede ser difundido posteriormente por la pasarela 138a a los terminales 134a-134n dentro de un haz de emisión respectivo del satélite, o por multidifusión a un subconjunto de los terminales 134a-134n a través de un haz puntual del satélite. Además, si bien el contenido puede dirigirse a uno o más terminales específicos (por ejemplo, que solicitaron el contenido), otros terminales dentro del haz de recepción pueden almacenar en caché el contenido de manera oportunista para el caso de que los usuarios de uno o más de los restantes terminales soliciten posteriormente el contenido.

[0034] La figura 2 ilustra un diagrama de bloques de un sistema de comunicaciones por satélite adicional para proporcionar servicios de difusión en continuo de medios adaptativos, de acuerdo con realizaciones de ejemplo de la presente invención. El sistema 200 comprende un satélite 232, varios terminales de cliente/usuario de satélite ST201a, ST201b, ST201c, ST201d, ST201e, ST201f, ..., una pasarela 238, una red de comunicaciones 250 (tal como Internet) y un servidor web 241 y un servidor de contenido 242. A modo de ejemplo, el servidor web 241 puede ser un servidor web empresarial, tal como un servidor host de Internet Netflix o un servidor host de Internet Apple. A modo de ejemplo adicional, el servidor de contenido puede ser un servidor de contenido general accesible directamente a través de Internet, o un servidor de contenido empresarial específico, tal como un servidor de contenido de Netflix utilizado para el almacenamiento de contenido de Netflix y accesible a través de un servidor host de Internet de Netflix. Como sería evidente para un experto en la materia, el sistema de comunicaciones 200 no necesita limitarse a una sola pasarela, una red de comunicaciones, un servidor web y un servidor de contenido. Por ejemplo, como se representa en la figura 1, una serie de pasarelas y pasarelas IP pueden servir respectivamente a grupos del universo de terminales en el sistema. Además, se puede proporcionar contenido a una o más pasarelas, a través de la red de comunicaciones, desde una multitud de servidores web y servidores de contenido (con uno o más servidores de contenido proporcionados por cada uno de varios proveedores de contenido). Además, el contenido de diferentes proveedores de contenido puede proporcionarse a través de diferentes redes de entrega; por ejemplo, una gran red corporativa puede proporcionar contenido a una multitud de sitios corporativos y terminales de usuarios corporativos a través de Internet (por ejemplo, a través de una VPN) y/o un servicio privado red corporativa de área amplia (WAN). Además, también sería evidente para un experto en la materia que el sistema no necesita limitarse a un número máximo específico de terminales de satélite 201: el número de terminales se regirá por factores de diseño del sistema asociados, tales como la capacidad del satélite, el número de haces de satélite, el número de pasarelas y pasarelas IP, la gestión de capacidad y los sistemas de gestión de terminal/cliente de la red, etc. Además, como se describió anteriormente, con respecto a los haces de enlace ascendente y enlace descendente del satélite, la figura 2 representa un haz de enlace ascendente 210 y dos haces de enlace descendente 220a y 220b. Todos los ST 201a-201f reciben servicio del único haz de enlace ascendente 210, mientras que los ST 201a-201b reciben servicio del haz de enlace descendente 220a y los ST 201c-201f reciben servicio del haz de enlace descendente 220b. El satélite 232 proporciona además una capacidad de transmisión a través del haz de enlace descendente 230, que cubriría un área geográfica relativamente grande (por ejemplo, Estado Unidos continental) que contiene múltiples haces individuales de enlace ascendente y enlace descendente sobre esa área de cobertura.

[0035] Además, como se representa, en dicho sistema de comunicaciones de banda ancha, que comprende una o más pasarelas de satélite, uno o más satélites, y una pluralidad de terminales de satélite cliente/usuario, las pasarelas se comunican con servidores de contenido (por ejemplo, servidores web y de aplicaciones) a través de la red de comunicaciones 250 (por ejemplo, Internet), y los terminales cliente transmiten solicitudes y reciben respuestas de contenido desde las pasarelas a través de canales o haces del satélite. Como se utiliza en este documento, un enlace directo se refiere a una ruta o canal de comunicaciones desde una pasarela a los terminales, y un enlace de retorno se refiere a una ruta o canal de comunicaciones desde un terminal a una pasarela. Generalmente, todos los terminales dentro del área de cobertura de un haz de satélite de enlace directo pueden recibir los datos transmitidos por una pasarela a través de ese enlace directo. El contenido generalmente se entrega desde un servidor de contenido a un dispositivo cliente a través de un protocolo de solicitud/respuesta. A modo de ejemplo, el Protocolo de transferencia de hipertexto (HTTP) es un ejemplo común de dicho protocolo de solicitud/respuesta. Cuando un cliente quiere consumir contenido, el cliente emite una solicitud de contenido a un servidor de contenido. La solicitud es recibida por el terminal de satélite y el terminal transmite la solicitud a la pasarela que da servicio al terminal a través del canal de enlace de retorno respectivo. La pasarela luego transmite la solicitud al servidor de contenido. Cuando el contenido recibe la petición de contenido, adquiere o recupera el contenido y lo vuelve a difundir en continuo a la pasarela a través de la red de comunicaciones. La pasarela luego retransmite la respuesta de contenido al terminal/cliente solicitante a través del canal de enlace directo. En el caso de que la difusión de respuesta de contenido de pasarela por la pasarela pueda ser recibida por todos los terminales en el haz del satélite. El terminal luego proporciona la respuesta de contenido al cliente. Durante la entrega de contenido, tanto las solicitudes de contenido como las respuestas de contenido pasan por la pasarela y el terminal entre un cliente y un servidor de contenido.

[0036] A modo de ejemplo, un usuario de un terminal de cliente particular (por ejemplo, el terminal de cliente 221a) puede solicitar un archivo de contenido específico, por ejemplo, un archivo de película a través de la cuenta de Netflix del usuario. Como resultado, la aplicación cliente del navegador web del terminal 221a retransmitirá la

petición, dirigida al servidor host de Netflix (por ejemplo, el servidor web 241), al ST 201a. El ST 201a primero determina si el contenido solicitado ya está almacenado en su caché local y, de ser así, el ST proporciona el contenido al terminal cliente respectivo directamente desde su efectivo. Si el contenido no se restaura en la memoria caché local del ST 201a, el ST vuelve a empaquetar o encapsular la petición con una dirección de origen del ST 201a y una dirección de destino de la pasarela respectiva que sirve al ST (por ejemplo, la pasarela 238), mientras mantiene la dirección de origen original del terminal cliente y la dirección de destino del servidor web dentro del paquete encapsulado, y transmite el mensaje por el satélite 232 a la pasarela 238. La pasarela recibe el mensaje transmitido, lo desencapsula para obtener el original la dirección de destino prevista (la del servidor web 241, en este caso el servidor host de Internet Netflix), vuelve a encapsular la petición con una dirección de origen de la pasarela 238 y una dirección de destino del servidor web 241 (mientras se mantiene la fuente original dirección del terminal del cliente y la dirección de destino del servidor web dentro del paquete encapsulado), y transmite el mensaje a través de la red de comunicaciones 250 al servidor de web. En respuesta, el servidor web recupera el contenido de la película desde el servidor de contenido 242, y transmite el contenido, a través de la red de comunicaciones 250, hasta la pasarela 238 (que se indicó al servidor web como fuente inmediata del mensaje de petición). Por ejemplo, el servidor web encapsula los datos de contenido transmitidos con el servidor web como la dirección de origen inmediata y la pasarela como la dirección de destino inmediata, mientras que incluye al servidor web como la dirección fuente original y el terminal de cliente 221a como la dirección de destino final con los paquetes de datos encapsulados, y transmite los paquetes a la pasarela 238 a través de la red de comunicaciones 250. Alternativamente, el servidor web puede controlar el servidor de contenido 242 para procesar y transmitir los datos de contenido directamente hasta la pasarela a través de la red de comunicaciones 250.

[0037] La pasarela desencapsula los paquetes para obtener la dirección de destino prevista (la del terminal cliente 221a) y resuelve esa dirección tal como la maneja el ST 201a, vuelve a encapsular los paquetes con una dirección de origen de la pasarela 238 y una dirección de destino del ST 201a (mientras se mantiene la dirección fuente original del servidor web 241 y la dirección de destino del terminal cliente 221a dentro del paquete encapsulado), y transmite los paquetes al satélite 232. Al recibir los paquetes, el respectivo satélite transpondedor transmite los paquetes a través del haz/canal de enlace descendente 220a correspondiente para su recepción por el ST 201a. El ST 201a recibe los paquetes de datos de contenido, desencapsula dichos paquetes para determinar la resolución de dirección apropiada y resuelve la dirección de destino como el terminal cliente 221a, y retransmite los paquetes al terminal de destino. Como reconocería un experto en la materia, dicho sistema según las realizaciones de ejemplo no se limitaría a ningún protocolo de comunicación específico, sino que podría emplear cualquiera de los diferentes formatos o protocolos de comunicación conocidos para el intercambio de mensajes y contenido a través de los distintos enlaces de la red. Por ejemplo, en la capa de Internet, el Protocolo de Internet estandarizado (IP) puede aplicarse para transmitir datagramas a través de los límites de la red, entregando paquetes desde el host de origen al host de destino únicamente en función de las direcciones IP de los encabezados de los paquetes. Además, en la capa de transporte, se puede emplear uno cualquiera de varios protocolos conocidos, incluidos el Protocolo de control de transmisión (TCP), el Protocolo de datagramas de usuario (UDP), etc. De manera similar, se pueden aplicar varios protocolos bien conocidos en la aplicación y el enlace capas.

[0038] El contenido de video y audio está codificado en diferentes formatos tales como VC-1, H.264, HEVC, WMA, ACC, DD, etc. Para cada esquema de codificación, un único archivo de contenido de video será codificado en varios archivos diferentes, cada uno correspondiente a diferentes resoluciones a diferentes tasas binarias respectivas. Por ejemplo, los diferentes archivos de contenido de video codificado pueden variar desde 235 kbps hasta 16 Mbps y más. En general, para un archivo de video de contenido particular, habrá varios archivos de video con diferentes tasas binarias para cada formato de codificación de video y un par de archivos de audio de diferentes tasas binarias para cada formato de codificación de audio diferente. Por lo general, habrá aproximadamente ocho archivos de video con diferente tasa binaria para cada esquema de codificación y un par de archivos de audio para diferentes esquemas de codificación. Los archivos de audio y video codificados generalmente también están protegidos por esquemas de gestión de derechos digitales (DRM) que dan como resultado diferentes formatos DRM tales como WMDRM, PlayReady, Widevine, etc. Además, cada archivo de video y audio está dividido virtual o físicamente en segmentos de determinada duración. A modo de ejemplo, la duración de segmento puede variar desde un par de segundos hasta decenas de segundos por segmento. Existen varias técnicas de segmentación como ASF multiplexada, ASF sin multiplexar, M2TS multiplexada, M2TS sin multiplexar, FMP4 sin multiplexar, etc. La segmentación y el empaquetado en diferentes formatos también, se conoce como formateado en diferentes "envases". La combinación de esquemas de codificación de video y audio, procedimientos DRM y tipo de envases se conoce como perfil. Un video generalmente se procesa en varios perfiles y los perfiles se crean para servir a diferentes plataformas tales como iDevice, Android, Roku, Xbox, PlayStation, etc. Para mayor seguridad, generalmente se usa una clave de cifrado única común para cifrar un archivo de video/audio, y la clave suele ser la misma para todos los usuarios. La clave se cifra con una clave de sesión y se entrega a diferentes usuarios. Como resultado, los archivos de video/audio a una tasa binaria particular de un perfil son idénticos para todos los usuarios. En aplicaciones prácticas, varios servicios de difusión en continuo de video emplean el enfoque anterior de codificar y segmentar archivos de contenido de audio y video.

[0039] La figura 3 ilustra los segmentos de video de una pluralidad de diferentes archivos "P" de presentación o tasa binaria de reproducción (PBR) (PBR (1), PBR (2), ..., PBR (j), ..., PBR (P)), y un archivo de audio "Am" con una PBR (a), para un perfil particular del archivo de contenido de video (m), de acuerdo con realizaciones de ejemplo de la presente invención. Un perfil se determina mediante una combinación del esquema de codificación de video, el esquema de codificación de audio, el esquema de gestión de derechos digitales (DRM) y el formato de envase. En el ejemplo de la fig. 3, el archivo de contenido de video (m) comprende archivos de video "P" (cada uno de una tasa

binaria PBR respectiva) y un archivo de audio. Cada archivo de video y el archivo de audio se componen de "n" segmentos, cada uno con "S" segundos de duración. La nomenclatura Seg. (x, y, z) representa el segmento de video z-ésimo del archivo de video de tasa binaria "y" del archivo de contenido de video "x". Por ejemplo, el segmento Seg. (m, 1,1) representa el primer segmento del archivo de video de PBR (1) del m-ésimo archivo de contenido de video (m), Seg. (m, 1, k) representa el segmento k-ésimo del archivo de video de PBR (1) del archivo de contenido de video (m), y así sucesivamente. Del mismo modo, el segmento Seg. (m, j, 1) representa el primer segmento del archivo de video de PBR (j) del archivo de contenido de video (m), Seg. (m, j, k) representa el segmento k-ésimo del archivo de video de PBR (j) del archivo de contenido de video (m), y así sucesivamente. Además, el segmento Seg. (Am, 1,1) representa el primer segmento del archivo de audio "Am" en PBR (a) para el archivo de video (m), Seg. (am, 1, k) representa el segmento k-ésimo del archivo de audio en PBR (a) para el archivo de video (m), y así sucesivamente. Generalmente, en la práctica, para la presentación o reproducción de un archivo de contenido de video en particular, el dispositivo/aplicación cliente solo descarga un subconjunto de segmentos de video respectivos, ya que no todos los segmentos son necesarios para una reproducción.

[0040] La figura 4 ilustra un diagrama de bloques de un dispositivo/aplicación de reproducción de cliente, que emplea adaptación de tasa, en un sistema adaptativo de difusión en continuo de medios, de acuerdo con realizaciones de ejemplo de la presente invención. La figura 5 ilustra una secuencia de segmentos de video descargados por el dispositivo/aplicación de reproducción cliente de la figura 4, para una sesión de reproducción de video que emplea difusión en continuo de video adaptativa, donde la línea discontinua refleja la secuencia de descarga, de acuerdo con realizaciones de ejemplo de la presente invención. El dispositivo/aplicación de reproducción cliente 410 comprende un elemento de adaptación de tasa 401, un receptor / transmisor 403 y una memoria intermedia de reproducción 405. El elemento de adaptación de tasa 401 comprende un controlador de tasa 411, una unidad de cálculo de transmisión de aplicación 413 y un supervisor de memoria intermedia de reproducción 415. La secuencia de descarga de los segmentos depende de la interacción entre el algoritmo de adaptación de tasa de dispositivo/aplicación de reproducción y el estado de los diversos enlaces de la ruta de descarga a través de la red (por ejemplo, congestión de enlace y margen de enlace o tasas de flujo, - tal como margen de enlace o tasas de flujo de un enlace de comunicaciones por satélite en función de las condiciones climáticas). Además, la secuencia o patrón de descarga puede no ser predecible ya que un usuario puede detener, reiniciar, adelantar y/o revertir la reproducción de manera arbitraria en cualquier momento. La difusión en continuo de video generalmente comienza con la señalización de control entre el dispositivo de reproducción/aplicación 410 y los servidores de control del proveedor de contenido (por ejemplo, el servidor de contenido 420 de la figura 4). A modo de ejemplo, una sesión de difusión en continuo de video adaptativa puede emplear protocolos HTTP y HTTPS, tales como HTTPS para la señalización de control y HTTP para la descarga de video y audio. Alternativamente, HTTPS también se puede usar para descarga de video y audio. Dicha señalización de control puede incluir varias señales de petición y respuesta (por ejemplo, petición HTTP GET y mensajes de respuesta asociados) y transmisiones de comunicación (por ejemplo, comunicación TCP/IP) para establecer la sesión e identificar el archivo de contenido de video deseado y otras funciones de control, como acceso de seguridad y controles de autorización de usuario. Una vez que se ha completado la señalización de control, comienza la descarga de video y audio y el dispositivo/aplicación de reproducción comienza el almacenamiento en memoria intermedia de los segmentos de video y audio y la reproducción del archivo de contenido de video.

[0041] Con referencia adicional a la figura 4 y la figura 5, como ejemplo, el dispositivo/aplicación de reproducción 410 descarga típicamente los primeros segmentos de cada uno de los archivos de video PBR y archivos de audio (como se indica por la línea discontinua a través del primer segmento de cada uno de los archivos de video PBR de la figura 5). El controlador de tasa 411 generalmente comienza a la tasa binaria más baja, controlando el transmisor/receptor 403 para solicitar segmentos adicionales del archivo de video con PBR menor (como lo indica la línea discontinua a través del segundo segmento [Seg. (m, 1, 2)] del archivo de video PBR (1) de la figura 5). El algoritmo de adaptación de tasa 401 mide o estima continuamente el ancho de banda disponible y supervisa su estado de memoria intermedia de reproducción, a través del elemento de cálculo de transmisión de aplicación 413 y el supervisor de memoria intermedia de reproducción 415, respectivamente. Por ejemplo, la transmisión de la aplicación se puede calcular para cada segmento descargado dividiendo el tamaño del segmento por el tiempo de descarga respectivo, y el ancho de banda disponible se puede estimar a partir de las estadísticas de rendimiento de la aplicación. El algoritmo de adaptación de tasa 401 determinará entonces si se puede mantener una PBR mayor en función de la estimación del ancho de banda y el estado de la memoria intermedia de reproducción. En el caso de que la determinación sea positiva, el algoritmo de adaptación de tasa controlará el transmisor/receptor 403 para comenzar a descargar segmentos adicionales del archivo de video de PBR mayor y cambiará a la tasa de reproducción mayor (como lo indica la línea discontinua a través del segmento k-ésimo [Seg. (m, 2, k)] del archivo de video PBR (2) de la figura 5). En general, después de una activación inicial, cuando la condición de red entre el cliente del usuario y el servidor de contenido es estable, la reproducción generalmente se estabiliza a una tasa binaria mayor sostenible y permanece allí hasta el final del video (como lo indica la línea discontinua a través del segmento (k+1)-ésimo [Seg. (m, j, k + 1)] del archivo de video PBR (j) de la figura 5). Aunque aparentemente es estable, el algoritmo de adaptación de tasa 401 puede continuar midiendo/estimando el ancho de banda disponible y supervisando el estado de la memoria intermedia de reproducción. En consecuencia, si las condiciones de la red disminuyen durante la reproducción, el algoritmo de adaptación de tasa puede detectar una reducción del ancho de banda disponible y una reducción continua resultante de segmentos de video almacenados en memoria intermedia. En este caso, el algoritmo de adaptación de tasa puede determinar que la PBR actual ya no puede mantenerse, y en consecuencia controlará el transmisor/receptor 403 para comenzar a descargar segmentos adicionales del archivo

de video de PBR menor y cambiar a la tasa de reproducción más baja (como se indica por la línea discontinua a través del segmento enésimo [Seg. (m, 2, n)] del archivo de video PBR (2) de la figura 5).

[0042] La figura 6 ilustra un sistema adaptativo de difusión en continuo de medios que emplea un elemento de conformación de tasa (RSE), de acuerdo con realizaciones de ejemplo de la presente invención. Como se muestra en la figura, el equipo local del usuario/cliente (el equipo local de usuario (CPE)), ubicado localmente dentro de las instalaciones cliente del usuario, incluye el dispositivo/aplicación de reproducción cliente (que puede ser igual o diferente que el dispositivo/aplicación de reproducción cliente 410 de la figura 4), un elemento de configuración de tasa (RSE) 620, y el equipo del proveedor de datos y aplicaciones local 630, según se requiera para la prestación de servicios de comunicaciones suscritos o adquiridos por el usuario o cliente del cliente, y proporcionados por un servicio proveedor a través de la red de proveedores de servicios 640. Dicho equipo de proveedor de servicios puede incluir módems y enrutadores, tales como módems DSL y/o por cable, nodos de fibra óptica o módems para comunicaciones de fibra óptica y/o terminales de satélites para servicios de comunicaciones por satélite. Utilizando los servicios suscritos de la red de proveedores de servicios 640, el cliente o comprador puede, a su vez, acceder a varias fuentes de datos y aplicaciones, por ejemplo, a través de una red de comunicaciones de área extensa 650 (tal como Internet). Además, tales fuentes de datos y aplicaciones pueden incluir servicios de difusión en continuo de datos (tales como difusión en continuo de medios) a través de un servidor de contenido remoto 660.

[0043] La figura 7 ilustra la señalización general de un sistema adaptativo de difusión en continuo de datos o medios, y el proceso de configuración de tasa de capa de aplicación del RSE, de acuerdo con realizaciones de ejemplo de la presente invención. En el contexto de una aplicación de difusión en continuo de video HTTP, el dispositivo/aplicación de reproducción del cliente emite una petición HTTP GET para el contenido multimedia/video deseado. El RSE reenvía la petición HTTP GET al servidor de contenido respectivo. En respuesta a la petición GET, el servidor de contenido responde con un mensaje de respuesta HTTP GET. El RSE recibe el mensaje de respuesta y, de acuerdo con realizaciones ejemplares, dicho RSE realiza ciertas funciones relacionadas con el servicio de difusión en continuo de datos (por ejemplo, difusión en continuo de video), que incluyen (1) medición dinámica del flujo de capa aplicación, (2) estimación dinámica de tasa de reproducción actual del respectivo dispositivo/aplicación de reproducción de cliente 610, y (3) conformación dinámica del flujo de capa de aplicación vista por el dispositivo/aplicación de reproducción cliente, para controlar el dispositivo/aplicación de reproducción cliente para seleccionar una tasa de reproducción deseada. A este respecto, el RSE se coloca en un punto estratégico tan cerca del dispositivo/aplicación de reproducción cliente como sea razonablemente práctico para que dicho RSE pueda aproximar con precisión el flujo de capa de aplicación vista desde el dispositivo/aplicación de reproducción cliente. El RSE luego retransmite la respuesta HTTP GET al dispositivo/aplicación de reproducción, donde el RSE puede retener la respuesta HTTP para aumentar el tiempo de ida y vuelta experimentado por el dispositivo/aplicación de reproducción en caso de que el RSE determine que el flujo de capa de aplicación necesita reducirse para lograr la tasa de reproducción deseada. Como sería evidente para un experto en la materia, el RSE puede funcionar como un dispositivo proxy entre el dispositivo/aplicación de reproducción cliente y el servidor de contenido, y en consecuencia puede encapsular o volver a encapsular los mensajes respectivos con las direcciones IP de origen y destino apropiadas para que el elemento de configuración de tasa reciba los mensajes de respuesta a través de su propia dirección IP y pueda hacer coincidir la dirección IP de origen (por ejemplo, la dirección del servidor de contenido respectivo) con la dirección IP apropiada para el dispositivo/aplicación de reproducción cliente. Alternativamente, el RSE puede funcionar como una entidad de transferencia que utiliza la dirección IP del dispositivo/aplicación de reproducción cliente como su propia dirección IP para comunicaciones con el servidor de contenido remoto final.

[0044] La figura 8 ilustra un diagrama de flujo que representa un algoritmo de conformación de capa de aplicación realizado por el RSE, de acuerdo con realizaciones de ejemplo de la presente invención. En una primera etapa, el RSE recibe una petición HTTP, desde un dispositivo/aplicación de reproducción cliente, para un segmento de video actual de un archivo de video (S801). Una vez que se recibe la petición HTTP, el RSE retransmite la petición al servidor de contenido 660 (por ejemplo, a través del equipo del proveedor de servicios 630, la red del proveedor de servicios 640 y WAN 650), registra el tiempo de recepción de la solicitud y obtiene el nombre y longitud del archivo de video a partir de URL de petición HTTP y los campos de encabezado (S803). El RSE luego espera una respuesta HTTP del servidor de contenido correspondiente a la petición HTTP para el segmento de video actual. Una vez que el RSE recibe una respuesta HTTP (S805), confirma que la respuesta HTTP recibida corresponde a la petición HTTP para el segmento actual (S807). Si, por alguna razón, la respuesta HTTP no corresponde a la petición HTTP, el RSE continúa esperando para recibir la respuesta HTTP correspondiente. Una vez que se confirma la respuesta HTTP correspondiente, el RSE registra el tiempo de recepción de la respuesta, que se determina como el tiempo de recepción del último paquete de respuesta (S809). El RSE obtiene la longitud del contenido de video del campo de la cabecera de respuesta HTTP y/o el tamaño del propio cuerpo de respuesta HTTP (S809). El RSE calcula además el flujo de capa de aplicación (ALT) y efectúa una estimación de la tasa de reproducción actual (S809). El RSE determina si la tasa de reproducción actual es menor o igual a la tasa de reproducción deseada (S811). En caso que el RSE determine que la tasa de reproducción actual es inferior o igual a la tasa de reproducción deseada, retransmite o reenvía rápidamente la respuesta HTTP para el segmento actual al dispositivo de reproducción del cliente para almacenamiento en memoria intermedia y reproducción (S813).

[0045] Alternativamente, cuando el RSE determina que la tasa de reproducción actual es mayor que la tasa de reproducción deseada, el RSE determina que la tasa de reproducción debe reducirse. De acuerdo con una realización, el RSE controla el dispositivo de reproducción del cliente para disminuir la tasa de reproducción disminuyendo la ALT. Cuando la ALT se reduce en una cantidad requerida, el dispositivo/aplicación de reproducción del cliente determinará que la tasa de rendimiento actual no es suficiente para mantener la tasa de reproducción y responde bajando la tasa de reproducción a la tasa deseada (que sería compatible con la ALT, como fue bajada por

el RSE). A modo de ejemplo, al determinar que debe reducirse la tasa de reproducción, el RSE realiza primero una etapa opcional para determinar si el segmento actual debe filtrarse para no ser considerado para la ALT (S815). El RSE puede determinar excluir la consideración de algunos segmentos cuando el tamaño del segmento es significativamente más pequeño que un tamaño de segmento promedio para la tasa de reproducción. En el caso de que el RSE realice dicha etapa opcional y determine que el segmento actual debe filtrarse, libera la respuesta HTTP para el segmento actual al dispositivo de reproducción del cliente para almacenamiento en memoria intermedia y reproducción (S817). Alternativamente, en el caso de que el RSE realice la etapa opcional y determine que debe considerarse el segmento actual, el RSE determina si la ALT medida correspondiente es mayor que la ALT requerida para la tasa de reproducción deseada (S819). En el caso de que el RSE determine que la ALT medida es menor o igual que la ALT requerida, libera la respuesta HTTP para el segmento actual al dispositivo de reproducción del cliente para almacenamiento en memoria intermedia y reproducción (S821). Alternativamente, en el caso de que el RSE determine que la ALT medida es mayor que la ALT requerida, mantiene la respuesta HTTP respectiva durante un período de tiempo requerido para reducir el flujo de capa de aplicación a un nivel correspondiente a la tasa de reproducción deseada (S823). Una vez que se alcanza el período de tiempo requerido, el RSE libera la respuesta HTTP para el segmento actual al dispositivo de reproducción del cliente para almacenamiento en memoria intermedia y reproducción (S825). Después de cualquiera de las etapas donde el RSE libera la respuesta HTTP para el segmento actual al dispositivo de reproducción del cliente (Etapas, S813, S817, S821, S825), el RSE determina si se ha recibido el último segmento del archivo de video, en otro caso, entonces el RSE espera la petición HTTP GET para el siguiente segmento de archivo de video (S810) y el proceso continúa para los segmentos de archivo de video subsiguientes y, en caso afirmativo, el proceso de reproducción finaliza.

[0046] Más específicamente, el algoritmo de conformación de capa de aplicación de tales realizaciones funciona basándose en ciertos parámetros o variables. A modo de ejemplo, el algoritmo funciona basándose de ciertos parámetros, incluido un conjunto de tasas de reproducción del dispositivo/aplicación de reproducción cliente; a los efectos de la presente, el conjunto de tasas de reproducción se denominará como $\{PBR(i), i = 1, 2, \dots, P\}$, donde $PBR(i)$ refleja la tasa de reproducción promedio i , P refleja el número total de tasas de reproducción en el conjunto, y $PBR(i) < PBR(i+1)$, con $i = 1, 2, \dots, P - 1$. Un conjunto de ejemplos de tasas de reproducción puede ser $\{0,235, 0,375, 0,56, 0,75, 1,05, 1,75, 2,35, 3,0\}$ en megabits por segundo (Mbps). Además, dichos parámetros pueden incluir un tamaño de segmento (en segundos) para cada una de las tasas de reproducción, donde el tamaño del segmento sería proporcionado por las especificaciones para la tasa de reproducción respectiva del dispositivo/aplicación de reproducción cliente. A modo de ejemplo adicional, el algoritmo funciona basándose en de ciertos parámetros configurables, incluida una tasa de reproducción deseada para cada archivo de contenido de video y sesión de reproducción; para los fines de la presente, cada tasa de reproducción se denominará $PBR_{Desired}$. Además, dichos parámetros configurables pueden incluir un margen para el flujo de capa de aplicación requerida para admitir un porcentaje de tasa de reproducción específico (un porcentaje de almacenamiento intermedio por encima de la tasa de reproducción como tasa de reproducción es un valor promedio). Por ejemplo, para un nivel de transmisión de capa aplicación, $ALT_{Required}$, requerida para admitir una tasa de reproducción deseada específica, $PBR_{Desired}$, y un margen del 40%, $ALT_{Required} = 1,4 * PBR_{Desired}$. PBR es una tasa promedio, donde una tasa de reproducción instantánea puede ser significativamente mayor que la tasa promedio. La ALT requerida para soportar una tasa de reproducción deseada específica, por lo tanto, debe tener un margen de maniobra significativo para acomodar las fluctuaciones de tasa instantáneas y evitar el nuevo almacenamiento intermedio en el dispositivo de reproducción. Como corolario, el margen no puede ser tan grande como para llevar la PBR a la siguiente tasa de reproducción más alta.

[0047] Con respecto a la estimación de la tasa de reproducción llevada a cabo como parte del algoritmo de configuración de capa aplicación, mientras que el RSE recibe segmentos de audio/video, puede no saber realmente si un segmento pertenece a audio o video o la tasa de reproducción particular del archivo de video al que pertenece un segmento. Sin embargo, según las realizaciones de ejemplo, el RSE puede estimar con precisión la tasa de reproducción actual de un dispositivo/aplicación de reproducción a partir de cierta información de metadatos que puede estar disponible en la URL de petición HTTP y los campos de cabecera. A modo de ejemplo, puede estar presente una cadena de caracteres identificable efímera o una ID de archivo de video, que representa un archivo de video/audio en particular correspondiente a una tasa de reproducción respectiva; para los fines de la presente, la ID del archivo de video se denominará "*VideoFileID(i)*" donde $i = 1, 2, \dots, P$. Además, una cadena de caracteres identificable efímera o una ID de video, que representa un contenido o título de video en particular, puede estar presente - para los fines de la presente, la ID del video se denotará como "*VideoID*". En algunos casos, por motivos de seguridad, los proveedores de contenido intentan ocultar intencionalmente *VideoFileID* y *VideoID* para evitar la piratería o el robo de videos, y también pueden cambiar los parámetros de una a otra sesión (incluso para un dispositivo cliente común). En tales casos, sin embargo, el RSE aún puede usar los parámetros *VideoFileID* y *VideoID*, ya que estos parámetros son de naturaleza efímera y permanecen constantes para todos los segmentos dentro de una sesión de reproducción. En otras palabras, dentro de una sesión particular, los parámetros *VideoFileID* y *VideoID* se pueden utilizar para identificar el título del contenido de video y el archivo PBR correspondiente, ya que permanecen constantes para todos los segmentos de la sesión. De acuerdo con una realización, por lo tanto, los segmentos de video de un archivo o título de contenido de video en particular se pueden identificar mediante el parámetro *VideoID*, y los segmentos pertenecientes a un título de contenido de video común se pueden clasificar en grupos de archivos a partir del parámetro *VideoFileID*, donde cada uno grupo corresponde a un archivo de video de una PBR respectiva. Luego, dado que todos los segmentos de un archivo de video PBR común en un sistema adaptativo de difusión en continuo de video o medios tienen la misma duración, la PBR medida por segmento en cada grupo de archivos se puede calcular de la siguiente manera:

$$PBR_{Measured} = \frac{\text{Tamaño segmento}}{S}$$

donde $PBR_{Measured}$ es la tasa de reproducción medida para un segmento y S es la duración del segmento en segundos. Además, una estimación precisa de la PBR ($PBR_{Estimate}$) para un grupo de archivos de video en particular puede determinarse como el promedio estadístico de los valores de PBR medidos ($PBR_{Measured}$) para los segmentos recibidos del grupo de archivos de video.

[0048] Con respecto a la medición del flujo de capa aplicación ($ALT_{Measured}$) para un segmento de video, realizado como parte del algoritmo de conformación de capa de aplicación, según una realización, la ALT medida se puede calcular de la siguiente manera:

$$ALT_{Measured} = \frac{\text{Tamaño segmento}}{RTT_{Measured}}$$

donde $RTT_{Measured}$ es el tiempo de ida y vuelta (RTT) para el segmento, medido desde el momento en que el RSE recibe una petición HTTP GET para un segmento actual hasta el momento en que el RSE recibe el último paquete de la respuesta HTTP correspondiente. RTT se compone del retardo del enlace de retorno más el retardo del enlace directo. El retardo del enlace de retorno es el tiempo desde el momento en que el RSE recibe la petición HTTP GET para el segmento actual y envía la petición HTTP al servidor de contenido respectivo hasta el momento en que el servidor de contenido recibe dicha petición HTTP. El retardo del enlace directo es el tiempo desde el momento en que el servidor de contenido envía la respuesta HTTP correspondiente hasta el momento en que el RSE recibe la respuesta HTTP correspondiente.

[0049] En otras palabras, el retardo del enlace de retorno y directo consiste en el tiempo de propagación de la petición del RSE al servidor de contenido y el tiempo de propagación de la respuesta del servidor de contenido al RSE, el tiempo de transmisión de la petición del RSE al servidor de contenido y el tiempo de transmisión de la respuesta del servidor de contenido al RSE, y cualquier retardo en la cola de la petición a través de los enlaces de la trayectoria desde el RSE al servidor de contenido y de la respuesta a través de los enlaces de la trayectoria desde el contenido servidor a la RSE. Cuando los retardos de propagación y en la cola son pequeños, la ALT viene determinada por el retardo de la transmisión (por ejemplo, especialmente por el retardo de la transmisión directa ya que el tamaño de la respuesta HTTP sería significativamente mayor que el tamaño de la petición HTTP. El retardo de la transmisión directa se debe principalmente al ancho de banda disponible en el enlace directo. En este caso, ALT puede configurarse controlando la tasa del enlace directo. Cuando los retardos de propagación y cola son significativos, como en los enlaces de comunicaciones por satélite de banda ancha (por ejemplo, en un sistema de satélite geosíncrono, donde el retardo de propagación en un sentido es de aproximadamente 300 ms) no es suficiente para conformar precisa y efectiva a la ALT.

[0050] Según las realizaciones de ejemplo, por tanto, el algoritmo de configuración de tasa mide dinámicamente la tasa de reproducción (PBR) por segmento, calcula una estimación actual de tasa de reproducción ($PBR_{Estimate}$) y compara la PBR estimada con la PBR deseada $PBR_{Desired}$ para determinar cuándo se requiere una reducción de la ALT. De acuerdo con una de tales realizaciones, en el caso de que el RSE determine que se debe reducir la ALT, el RSE puede realizar primero una etapa opcional para determinar si algunos de los segmentos (y qué segmentos) deben filtrarse para no ser considerados en el ajuste de ALT (por ejemplo, algunos segmentos pueden quedar fuera de consideración cuando el tamaño del segmento es significativamente menor que el tamaño promedio del segmento para una tasa de reproducción). El RSE luego compara la $ALT_{Measured}$ para el segmento actual con la $ALT_{Required}$ para la PBR deseada. Si $ALT_{Measured} > ALT_{Required}$, entonces el RSE agrega un retardo ($Delay_{Add}$) al RTT medido ($RTT_{Measured}$) (por ejemplo, mantiene el segmento de respuesta HTTP durante un período de tiempo requerido) tal que $ALT_{Measured} = ALT_{Required}$. De acuerdo con una de tales realizaciones, el retardo a agregar puede calcularse de la siguiente manera:

$$Delay_{Add} = RTT_{Measured} * \left(\frac{ALT_{Measured}}{ALT_{Required}} - 1 \right)$$

El RSE luego libera el segmento de video de respuesta actual al dispositivo de reproducción del cliente después de mantener el segmento durante el período de tiempo adicional $Delay_{Add}$.

[0051] La figura 9 ilustra un gráfico que representa un ejemplo de la aplicación de las ALT y PBR a los segmentos de un archivo de video, de acuerdo con las realizaciones de ejemplo de la presente invención, donde $PBR_{Desired} = 1.750.000$ bps, el margen es del 35% y, en consecuencia, $ALT_{Required} = 2.360.000$ bps. Aunque los valores de la PBR por segmento varían, el promedio estadístico de la PBR por segmento ($PBR_{Estimate}$) proporciona una estimación precisa de 1.750.000 bps. Los valores de ALT están limitados en $ALT_{Required}$.

[0052] La utilización de la estimación de la tasa de reproducción para la conformación de capa de aplicación en un sistema adaptativo de difusión en continuo de video o medios, como se describió anteriormente, proporciona un control preciso y efectivo de la tasa de reproducción del dispositivo/aplicación de reproducción cliente respectivo. Sin embargo, el uso del algoritmo utilizando una estimación de la tasa de reproducción requiere el conocimiento de las duraciones de los segmentos respectivos y la extracción de información de metadatos de archivos de video efímeros, tales como la identificación del archivo de video y el contenido del video o la identificación del título de las

URL de HTTP y campos cabecera HTTP. La ID del archivo de video se utiliza para ordenar los segmentos de video en diferentes grupos de archivos de video correspondientes a diferentes tasas de reproducción respectivas.

[0053] Sin embargo, según otras realizaciones de ejemplo, el algoritmo de conformación de capa aplicación de tales realizaciones puede emplearse sin utilizar una estimación actual de la tasa de reproducción ($PBR_{Estimate}$). La figura 10 ilustra un diagrama de flujo que representa un algoritmo de conformación de capa aplicación realizado por el RSE, sin utilizar una estimación de tasa de reproducción actual ($PBR_{Estimate}$) de acuerdo con realizaciones de ejemplo de la presente invención. El flujo del algoritmo de la figura 10 refleja el de la figura 8 (como se describió anteriormente), excepto que (1) en la etapa S1003, el RSE retransmite la petición al servidor de contenido y registra el tiempo de recepción de la petición HTTP (el RSE tampoco registra el nombre del archivo de contenido y la longitud del contenido, como en S803 de la figura 8), (2) en la etapa S1009, se registra el tiempo de recepción de la respuesta HTTP y se registra la longitud del contenido, y se calcula la ALT (no se calcula la estimación de la tasa de reproducción actual, como en S809 de la figura 8), (3) las etapas S811 y S813 no se realizan, y en su lugar, la etapa S1009 es seguida por la etapa S815. Sin utilizar la estimación de la tasa de reproducción, sin embargo, el margen para la PBR deseada debe establecerse con mayor precisión.

[0054] En la actualidad, los principales proveedores de difusión en continuo de video utilizan el protocolo HTTP para la difusión en continuo de video y protegen los archivos de video y audio a través de procesos de gestión de derechos digitales (DRM) (por ejemplo, encriptado). En caso que un proveedor de contenido de video utilizara HTTPS para la difusión en continuo de contenido de medios digitales, pueden emplearse algoritmos de acuerdo con realizaciones de ejemplo adicionales de la presente invención para la configuración de la tasa de capa de aplicación para controlar las tasas de reproducción deseadas en los sistemas de difusión en continuo de medios. Cuando el contenido de difusión en continuo de video se transmite a través de HTTPS, la información de las URL HTTP y los campos de cabecera en los mensajes de petición y respuesta HTTP ya no es visible para el RSE. Aunque la ID del archivo de video ya no está disponible con HTTPS, sin embargo, la tasa de reproducción actual se puede aún estimar, por ejemplo, a partir de la distribución de los tamaños de segmento de video. Además, el campo de longitud de contenido en la respuesta HTTP ya no sería accesible por el RSE con HTTPS, por lo que la longitud del contenido puede calcularse, por ejemplo, a partir de una estimación del tamaño de la respuesta HTTP.

[0055] Según una realización, los tamaños de segmento respectivos pueden estimarse a partir de una serie de factores. A modo de ejemplo, dado que los segmentos de video con una PBR dada tienen duraciones en segundos "S", la mayoría de peticiones HTTP se emiten aproximadamente cada S segundos y, como resultado de ello, existiría un intervalo medible entre las peticiones HTTP. Además, el campo de longitud en un protocolo de seguridad de capa de transporte (TLS) (HTTPS consta de TLS aplicado sobre HTTP) puede usarse para identificar segmentos de datos de aplicación de capa de registro TLS. La cabecera de la capa de registro TLS no está encriptada y se transmite sin cifrar en las versiones actuales del protocolo TLS. Además, la longitud de la carga útil del paquete IP también se puede utilizar para agrupar segmentos de datos de aplicación de la capa de registro TLS en segmentos de video. Un enlace en la ruta de transmisión entre el servidor de contenido y el dispositivo de reproducción del cliente tiene una limitación en el tamaño de un paquete IP, que se conoce como Unidad de Transmisión Máxima (MTU). Usualmente, la MTU en un paquete IP es de 1500 bytes, y un paquete IP de un tamaño menor que la MTU, recibido después de una secuencia de paquetes de tamaños iguales a la MTU, puede indicar el final de un segmento de video.

[0056] A modo de ejemplo adicional, la unidad de transmisión en la capa de registro TLS es un "bloque de datos de aplicación", que normalmente tiene 214 bytes o 16 KB de longitud. Un segmento de video se dividirá en uno o más bloques de datos de aplicación según el tamaño del segmento (por ejemplo, si es mayor de 16 KB). Cada bloque de datos de aplicación está precedido por una cabecera de capa de registro TLS, que generalmente incluye un campo "Tipo de contenido" y un campo "Longitud". El campo tipo de contenido indica si el respectivo bloque de datos de la aplicación contiene "Datos de la aplicación" en el registro TLS y la longitud de campo indica la longitud del respectivo bloque de datos de aplicación. De acuerdo con esta realización, el RSE puede, por lo tanto, rastrear bloques de datos de aplicación de la capa de registro TLS en una secuencia de paquetes IP recibida desde el servidor de contenido. La longitud del contenido de un segmento se puede construir a partir de los tamaños de los bloques de datos de aplicación de la capa de registro TLS.

[0057] La figura 11 ilustra un diagrama de flujo que representa un algoritmo para estimar la longitud del contenido de un segmento de video a partir de los paquetes de capa de registro TLS recibidos, de acuerdo con realizaciones de ejemplo de la presente invención. En una primera etapa, el RSE recibe un paquete IP con una cabecera de capa de registro TLS para un bloque de datos de aplicación respectivo (S1101). El RSE luego determina si el paquete IP contiene un final del bloque de datos de aplicación actual (S1103). El final de un bloque de datos de aplicación de la capa de registro TLS se puede determinar verificando la suma acumulada de las cargas útiles de TCP en la secuencia de paquetes IP con el valor del campo de longitud en la cabecera de la capa de registro TLS respectiva, asegurándose de no contar dos veces paquetes TCP duplicados retransmitidos por el protocolo TCP. En el caso de que el RSE determine que el paquete IP no contiene el final del bloque de datos de la aplicación actual, el RSE espera recibir el siguiente paquete IP del respectivo bloque de datos de la aplicación de capa de registro TLS (S1105) y, una vez recibido, el algoritmo regresa para determinar si el siguiente paquete IP contiene el final del actual bloque de datos de aplicación de capa de registro TLS (S1103). Alternativamente, en el caso de que el RSE determine que el paquete IP contiene el final del bloque de datos de la aplicación de la capa de registro TLS actual, el RSE determina si el bloque de datos de la aplicación actual contiene el final de un segmento de audio/video actual (S1107). El final de un segmento puede determinarse por un intervalo de tiempo entre dos bloques de datos de aplicación de capa de registro TLS. En caso que el RSE determine que el bloque de datos de la aplicación actual no contiene el final del segmento de audio/video actual, el RSE espera recibir el paquete IP con la cabecera de la capa de registro TLS para el siguiente bloque de datos de la aplicación (S1101). Alternativamente, cuando el RSE

determina que el bloque de datos de la aplicación actual contiene el final del segmento de audio/video actual, el RSE determina que se ha recibido un segmento de audio/video completo y calcula el tamaño del segmento de audio/video respectivo (S1109).

[0058] **¡Error! Fuente de referencia no encontrada.** Muestra un gráfico que muestra una distribución de tamaño de segmento de ejemplo para un archivo de audio con una tasa de reproducción específica, y para varios archivos de video, cada uno con una tasa de reproducción específica (PBR), de acuerdo con realizaciones de ejemplo de la presente invención – las tasas de reproducción archivo de video representado son 0,56 Mbps, 0,75 Mbps, 1,05 Mbps, 1,75 Mbps, 2,35 Mbps y 3,0 Mbps. Según una realización, para excluir valores atípicos, puede considerarse una distribución de corte con un límite inferior (lb) y un límite superior (ub) PBR, basado en los límites inferior y superior correspondientes para los tamaños de segmento respectivos, como sigue:

[PBR_{lb}(i), PBR_{ub}(i)] para cada PBR (i) donde PBR (i) es una tasa de reproducción promedio, i, y PBR_{lb}(i) y PBR_{ub}(i) son los puntos de corte inferior y superior, respectivamente. Un ejemplo del punto de corte se puede obtener de la siguiente manera:

$$PBR_{lb}(i) = (1 - CutOff) * PBR(i)$$

$$PBR_{ub}(i) = (1 + CutOff) * PBR(i)$$

donde *CutOff* está comprendido entre 0 y 1. A modo de ejemplo, el valor de *CutOff* se puede establecer en 0,5. Además, de acuerdo con las realizaciones de ejemplo, el valor de *CutOff* se puede establecer de manera diferente para que diferentes archivos PBR tengan en general valores diferentes de límites inferiores y superiores.

[0059] La figura 13 ilustra un diagrama de flujo que representa un algoritmo para estimar la tasa de reproducción actual, de acuerdo con realizaciones de ejemplo de la presente invención. En una primera etapa, se omite un período de inicio para la sesión de difusión en continuo de video adaptativa (S1201); por ejemplo, un ejemplo de período de inicio puede ser tan corto como 100 segundos y puede ser tan largo como 10 minutos. El período de inicio debe determinarse como un período de tiempo suficiente para que la tasa de reproducción de la sesión se estabilice. Después del período de inicio, se recopilan muestras "N" de segmentos de video/audio (S1203). Las "N" se clasifican en grupos o contenedores de archivos de video con diferentes tasas de reproducción respectivas (S1205), según los límites superior e inferior respectivos, [PBR_{lb}(i), PBR_{ub}(i)], para la tasa de reproducción respectiva, PBR (i). Téngase en cuenta que algunas muestras pueden caer en más de un contenedor. La tasa de reproducción se estima como la tasa asociada con el contenedor que tiene el número máximo de muestras (S1207). Si "N" es suficientemente grande, la tasa de reproducción estimada puede representar la tasa de reproducción actual del dispositivo/aplicación de reproducción cliente.

[0060] Según la presente realización, el cálculo del flujo de capa de aplicación (ALT) requiere un RTT para una pareja petición/respuesta HTTP. Para determinar el RTT, es necesario asociar una petición HTTP y una respuesta HTTP correspondiente. Aunque la asociación en HTTPS puede no ser tan obvia como la de HTTP, la asociación aún puede estimarse. A modo de ejemplo, cuando hay un intervalo notable entre dos peticiones HTTPS consecutivas (que pueden representar el caso habitual), el RTT se puede calcular como se describió anteriormente. De manera alternativa, cuando las peticiones HTTP llegan relativamente próximas entre sí, el servidor de contenido debe entregar la respuesta en el orden de recepción de las peticiones HTTP en la canalización HTTP/1.1. Incluso en HTTP/2 o SPDY, la respuesta debe entregarse en el orden de recepción de las peticiones HTTP, ya que todas las respuestas tienen la misma prioridad porque los segmentos de video se reproducen secuencialmente por el dispositivo de reproducción cliente. Por tanto, un grupo de peticiones puede asociarse con un grupo respectivo de respuestas correspondientes, cuando las solicitudes están próximas entre sí y las respuestas pueden estar correspondientemente próximas entre si o incluso pueden superponerse. La agregación de parejas petición/respuesta se puede usar de la siguiente manera para calcular el flujo de capa de aplicación:

$$RTT_{Measured} = \text{Tiempo recepción RSE última respuesta} - \text{Tiempo recepción RSE primera respuesta}$$

$$\text{Tamaño-Segmento} = \text{Suma longitud contenido de todas las respuestas a correspondientes a un respectivo grupo de peticiones}$$

[0061] A la vista de lo anterior, de acuerdo con tales realizaciones de ejemplo para la difusión en continuo de video a través del protocolo HTTPS, se pueden emplear dos enfoques para la configuración del flujo de capa aplicación para la difusión en continuo de video sobre HTTPS. Según una realización, el algoritmo de conformación de la transmisión de aplicación es similar al algoritmo representado en la figura 8, a partir de la determinación asociada de la tasa de reproducción estimada y la determinación del flujo de capa de aplicación respectiva, descrita anteriormente con respecto a la figura 8. De acuerdo con una realización adicional, el algoritmo de conformación de la transmisión de aplicación, sin utilizar una estimación de la tasa de reproducción, es similar al algoritmo representado en la figura 10, basándose en la determinación del flujo de capa de aplicación respectiva, descrita anteriormente con respecto a la figura 8 y la figura 10)

[0062] La figura 14 ilustra un diagrama de bloques de un conjunto de placas de circuitos 1400 que ejecutan aspectos de enfoques para la conformación de la tasa de tráfico de capa de aplicación para difusión en continuo de medios adaptativa, de acuerdo con realizaciones de ejemplo de la presente invención. El conjunto de placas de circuitos 1400 incluye, por ejemplo, componentes de procesador y memoria incorporados en uno o más paquetes físicos. A modo de ejemplo, un paquete físico incluye una disposición de uno o más materiales, componentes y/o cableado en

un ensamblaje estructural (por ejemplo, una placa base o placa de circuito impreso) para proporcionar una o más características tales como resistencia física, conservación de tamaño y/o limitación de interacción eléctrica.

[0063] En una realización, el conjunto de placas de circuitos 1400 incluye un mecanismo de comunicación tal como un bus 1401 para pasar información entre los componentes del conjunto de placas de circuitos. Un procesador 1403 está conectado al bus 1401 para ejecutar instrucciones/programas y procesar información almacenada en, por ejemplo, una memoria 1405. El procesador puede incluir uno o más núcleos de procesamiento con cada núcleo configurado para funcionar independientemente. Un procesador de núcleos múltiples permite el procesamiento múltiple dentro de un solo paquete físico, como dos, cuatro, ocho o más núcleos de procesamiento. Alternativamente o además, el procesador puede incluir uno o más microprocesadores configurados en tándem a través del bus para permitir la ejecución independiente de instrucciones, canalización y subprocesamiento múltiple. El procesador también puede ir acompañado de uno o más componentes especializados para realizar ciertas funciones y tareas de procesamiento, tales como uno o más procesadores de señal digital (PSD) 1407, y/o uno o más circuitos integrados específicos de la aplicación (ASIC) 1409. Un PSD normalmente está configurado para procesar señales en tiempo real (por ejemplo, sonido o vídeo) en tiempo real independientemente del procesador. Del mismo modo, los ASIC se pueden configurar para realizar funciones especializadas que no son fácilmente realizadas por un procesador de propósito general. Otros componentes especializados para ayudar a realizar las funciones inventivas descritas en el presente documento pueden incluir uno o más disposiciones de puertas programables in situ (FPGA) (no se muestran), uno o más controladores (no mostrados), o uno o más de otras placas de circuitos de ordenador para propósitos especiales.

[0064] El procesador 1403 y los componentes que lo acompañan tienen conectividad con la memoria 1405 a través del bus 1401. La memoria puede incluir tanto memoria dinámica (por ejemplo, RAM) como memoria estática (por ejemplo, ROM) para almacenar instrucciones ejecutables que, cuando son ejecutadas por el procesador y/o el PSD 1407 y/o los ASIC 1409, llevan a cabo el proceso de realizaciones de ejemplo como se describen aquí. La memoria también puede almacenar los datos asociados o generados por la ejecución del proceso.

[0065] Además, la funcionalidad de las realizaciones de ejemplo de la presente invención puede proporcionarla el conjunto de placas de circuitos 1400, en respuesta al procesador 1403 que ejecuta una disposición de instrucciones de programa contenidas en la memoria 1405. La ejecución de las instrucciones de programa contenidas en la memoria provoca al procesador para realizar las etapas del proceso y generar los resultados descritos en este documento, o equivalentes de los mismos. También se pueden emplear uno o más procesadores en una disposición de procesamiento múltiple para ejecutar las instrucciones del programa. En realizaciones alternativas, la circuitería cableada puede usarse en lugar de o en combinación con instrucciones de software para ejecutar las realizaciones de ejemplo. Por lo tanto, las realizaciones de la presente invención no se limitan a ninguna combinación específica de circuitos de hardware y software.

[0066] Además, como se apreciará, un módulo o componente (como se menciona en el presente documento) puede estar compuesto por componentes de software, que se almacenan en una memoria u otro medio de almacenamiento legible por ordenador, y ejecutados por uno o más procesadores o UCP de los respectivos dispositivos. Sin embargo, como también se apreciará, que un módulo puede estar compuesto alternativamente por componentes de hardware o componentes de firmware, o una combinación de componentes de hardware, firmware y/o software. Además, con respecto a las diversas realizaciones de ejemplo descritas en el presente documento, mientras que ciertas funciones se describen como realizadas por ciertos componentes o módulos (o combinaciones de los mismos), tales descripciones se proporcionan como ejemplos y, por lo tanto, no pretenden ser limitantes. En consecuencia, cualquiera de estas funciones puede ser imaginada como realizada por otros componentes o módulos (o combinaciones de los mismos), sin apartarse del espíritu y el alcance general de la presente invención. Además, los procedimientos, procesos y enfoques descritos en este documento pueden implementarse en el procesador utilizando circuitos de procesamiento que pueden comprender uno o más microprocesadores, circuitos integrados específicos de la aplicación (ASIC), matrices de puertas programables en campo (FPGA) u otros dispositivos operables para ser configurados o programados para implementar los sistemas y / o procedimientos descritos aquí. Para la ejecución en tales dispositivos que son accionables para ejecutar instrucciones de software, los diagramas de flujo y los procedimientos descritos en este documento pueden implementarse en instrucciones de procesador almacenadas en un medio legible por ordenador, tal como software ejecutable almacenado en una memoria de almacenamiento de ordenador.

[0067] La figura 15 ilustra un diagrama de bloques de un sistema de ordenador que implementa aspectos de enfoques para la configuración de la tasa de tráfico de capa de aplicación para difusión en continuo de medios adaptativa, de acuerdo con realizaciones de ejemplo de la presente invención. El sistema de ordenador 1500 incluye un bus 1501 u otro mecanismo de comunicación para comunicar información, y un procesador 1503 acoplado al bus para procesar información. El sistema de ordenador también incluye la memoria principal 1505, como una memoria de acceso aleatorio (RAM) u otro dispositivo de almacenamiento dinámico, acoplado al bus para almacenar información e instrucciones para que el procesador las ejecute. La memoria principal también se puede utilizar para almacenar variables temporales u otra información intermedia durante la ejecución de las instrucciones que debe ejecutar el procesador. El sistema de ordenador incluye además una memoria de solo lectura (ROM) 1507 u otro dispositivo de almacenamiento estático acoplado al bus para almacenar información estática e instrucciones para el procesador. Un dispositivo de almacenamiento 1509, como un disco magnético o un disco óptico, se acopla adicionalmente al bus para almacenar información e instrucciones.

[0068] Según una realización de la invención, el sistema de ordenador 1500 proporciona enfoques dinámicos y flexibles para la conformación de la tasa de tráfico de capa de aplicación para la difusión en continuo de medios adaptativa en respuesta al procesador 1503 que ejecuta una disposición de instrucciones contenidas en la memoria

principal 1505. Dichas instrucciones pueden leerse en la memoria principal desde otro medio legible por ordenador, tal como el dispositivo de almacenamiento 1509. La ejecución de la disposición de las instrucciones contenidas en la memoria principal hace que el procesador realice las etapas del proceso descritas en este documento. También se pueden emplear uno o más procesadores en una disposición de procesamiento múltiple para ejecutar las instrucciones contenidas en la memoria principal. En realizaciones alternativas, la circuitería cableada se usa en lugar de o en combinación con instrucciones de software para ejecutar la realización de la presente invención. Por lo tanto, las realizaciones de la presente invención no se limitan a ninguna combinación específica de circuitos de hardware y software.

[0069] El sistema de ordenador 1500 también incluye una interfaz de comunicación 1517 acoplada al bus 1501. A modo de ejemplo, la interfaz de comunicación proporciona un acoplamiento de comunicación de datos bidireccional a un enlace de red 1519 conectado a una red local 1521. La interfaz de comunicación, por ejemplo, puede ser una tarjeta o módem de línea de abonado digital (DSL), una tarjeta de red digital de servicios integrados (ISDN), un módem de cable u otro módem para proporcionar una conexión de comunicación de datos a un tipo correspondiente de línea telefónica. Otro ejemplo, de interfaz de comunicación puede ser una tarjeta de red de área local (LAN) (por ejemplo, para Ethernet™ o una red de modo de transferencia asincrónica (ATM)) para proporcionar una conexión de comunicación de datos a una LAN compatible, o un módem óptico configurado para proporcionar comunicaciones con un enlace de red de fibra óptica. Los enlaces inalámbricos también se pueden ejecutar. Además, la interfaz de comunicación, por ejemplo, incluye dispositivos de interfaz periféricos, tales como una interfaz Universal Serial Bus (USB), una interfaz PCMCIA (Asociación Internacional de Tarjeta de Memoria de Ordenador Personal), etc.

[0070] El enlace de red 1519 típicamente proporciona comunicación de datos a través de una o más redes a otros dispositivos de datos. Por ejemplo, el enlace de red proporciona una conexión a través de la red local 1521 a un ordenador principal 1523, que tiene conectividad a una red 1525, tal como una red privada de área amplia (WAN) o una WAN pública (por ejemplo, Internet), o equipo de datos operado por el proveedor de servicios. El sistema de ordenador 1500 envía mensajes y recibe datos, incluido el código del programa, a través de la red o redes, a través del enlace de red 419 y la interfaz de comunicación 1517. En el ejemplo de Internet, un servidor (no mostrado) podría transmitir el código solicitado o el contenido perteneciente a un programa o servicio de aplicación para ejecutar una realización de la presente invención a través de la red 1525. El procesador 1503 ejecuta el código transmitido mientras se recibe y/o almacena el código en un dispositivo de almacenamiento u otro almacenamiento no volátil para su posterior ejecución.

[0071] Además, la terminología que se refiere a medios legibles por ordenador o medios informáticos o similares como se usa en este documento se refiere a cualquier medio que participe en proporcionar instrucciones al procesador de un ordenador o módulo o componente de procesador para la ejecución. Tal medio puede tomar muchas formas, incluyendo pero no limitado a medios no volátiles no transitorios y medios volátiles. Los medios no volátiles incluyen, por ejemplo, medios de disco óptico, medios de disco magnético o medios de disco eléctrico (por ejemplo, disco de estado sólido o SDD). Los medios volátiles incluyen memoria dinámica, como memoria de acceso aleatorio o RAM. Las formas comunes de medios legibles por ordenador incluyen, por ejemplo, disquete o disco flexible, disco duro, cinta magnética, cualquier otro medio magnético, CD ROM, CDRW, DVD, cualquier otro medio óptico, memoria de acceso aleatorio (RAM), solo lectura programable memoria (PROM), PROM borrable, EPROM flash, cualquier otro chip o cartucho de memoria, o cualquier otro medio desde el cual un ordenador pueda leer datos.

[0072] Si bien las realizaciones de ejemplo de la presente invención pueden proporcionar diversas ejecuciones (por ejemplo, incluyendo componentes de hardware, firmware y/o software) y, a menos que se indique lo contrario, todas las funciones son realizadas por una UCP o un procesador que ejecuta el código de programa ejecutable por ordenador almacenados en una memoria no transitoria o en un medio de almacenamiento legible por ordenador, los diversos componentes pueden implementarse en diferentes configuraciones de hardware, firmware, software y/o una combinación de los mismos. Salvo que se describa lo contrario en el presente documento, los diversos componentes mostrados en forma resumida o en bloque en las figuras son bien conocidos individualmente y su construcción y funcionamiento internos no son críticos para la fabricación o el uso de esta invención o para una descripción del mejor modo de realización la misma.

[0073] En la descripción anterior, se han descrito diversas realizaciones con referencia a los dibujos adjuntos. Sin embargo, será evidente que pueden realizarse diversas modificaciones sin apartarse del alcance de la invención como se establece en las siguientes reivindicaciones. Por consiguiente, la descripción y los dibujos deben considerarse en un sentido ilustrativo más que restrictivo.

REIVINDICACIONES

1. Procedimiento para difusión en continuo de datos de contenido que comprende:

5 recibir, mediante un dispositivo de configuración de tasa de flujo de capa de aplicación (620), un mensaje de petición de contenido actual de un dispositivo cliente respectivo (610), el mensaje de petición de contenido actual que solicita un segmento de contenido actual de una secuencia de segmentos de contenido de un archivo de datos de difusión en continuo para una sesión de difusión en continuo de datos del dispositivo cliente y reenviar la petición de contenido actual a un servidor de contenido remoto (660);
 10 recibir, mediante el dispositivo de configuración de tasa de flujo de capa de aplicación, un mensaje de respuesta de contenido actual transmitido por el servidor de contenido (660) en respuesta a la petición de contenido actual transmitida, incluyendo la respuesta de contenido actual el segmento de contenido actual solicitado, determinando un tamaño de segmento actual del segmento de contenido actual;
 determinar, una tasa ALT objetivo requerida para soportar una tasa de reproducción deseada, PBR, para la sesión de difusión en continuo de datos del dispositivo cliente;
 15 caracterizado porque recibir (S801) el mensaje de petición por el dispositivo de configuración de tasa de flujo de capa de aplicación, comprende además el registro de un tiempo de recepción del mensaje de petición de contenido actual;
 recibir (S805) el mensaje de respuesta por el dispositivo de configuración de tasa de flujo de capa de aplicación, comprende además el registro de un tiempo de recepción del mensaje de respuesta de contenido actual;
 20 determinar (S809) la tasa de flujo actual de capa de aplicación, ALT, para el segmento actual basándose en el tiempo de recepción del mensaje de petición de contenido, el tiempo de recepción del mensaje de respuesta de contenido y el tamaño del segmento;
 determinar (S819) mediante el dispositivo de configuración de tasa de flujo de capa de aplicación, si la tasa ALT actual es mayor que la tasa ALT objetivo determinada;
 25 cuando se determina que la tasa ALT actual es mayor que la tasa ALT objetivo, el procedimiento comprende además retener (S823) el mensaje de respuesta de contenido durante un tiempo de retención determinado para reducir la tasa ALT actual a la tasa ALT objetivo y proporcionar (S825) la respuesta de contenido al dispositivo de cliente al finalizar el tiempo de retención; donde cuando el tamaño del segmento es menor que un tamaño de segmento promedio para la tasa de reproducción (S815), esos segmentos se excluyen de consideración,
 30 comprendiendo el procedimiento adicionalmente liberar rápidamente (S817) el segmento de contenido actual al dispositivo del cliente; y cuando se determina que una tasa ALT actual no es mayor que la tasa ALT objetivo, dicho procedimiento comprende además proporcionar rápidamente (S821) el segmento de contenido actual al dispositivo cliente.

35 2. Procedimiento según la reivindicación 1:

en el que, como parte de la etapa de recepción del mensaje de respuesta de contenido actual, el procedimiento comprende además determinar una PBR (tasa de reproducción) actual de la sesión de difusión en continuo de datos del dispositivo cliente; y
 en el que, después de la etapa de recepción del mensaje de respuesta de contenido actual, el procedimiento
 40 comprende además determinar si la PBR actual es mayor que la PBR deseada; y
 cuando se determina que la PBR actual es mayor que la PBR deseada, el procedimiento pasa a la etapa de determinar si la tasa ALT actual es mayor que la tasa ALT objetivo, y
 cuando se determina que la PBR actual no es mayor que la PBR deseada, el procedimiento comprende además proporcionar rápidamente el segmento de contenido actual al dispositivo cliente, y la etapa de determinar si la tasa
 45 ALT actual es mayor que la tasa ALT objetivo, junto con las etapas asociadas con el resultado de esta determinación, no se llevan a cabo.

3. Dispositivo de configuración de la tasa de flujo de la capa de aplicación que comprende:

50 una interfaz de dispositivo cliente;
 una interfaz de terminal de comunicaciones de red; y un procesador de configuración de tasa;
 en el que la interfaz del dispositivo cliente es utilizable para recibir un mensaje de petición de contenido actual desde un dispositivo cliente respectivo, solicitando el mensaje de petición de contenido actual un segmento de contenido actual de una secuencia de segmentos de contenido de un archivo de difusión en continuo de datos para una sesión de difusión en continuo de datos del dispositivo cliente,
 55 en donde la interfaz del terminal de comunicaciones de red es utilizable para retransmitir la petición de contenido actual a un servidor de contenido remoto (660), y para recibir un mensaje de respuesta de contenido actual transmitido por el servidor de contenido (660) en respuesta a la petición de contenido actual retransmitida, incluyendo la actual respuesta de contenido, el segmento de contenido actual solicitado,
 en el que el procesador de conformación de tasa (620) es utilizable para determinar una tasa flujo de capa de
 60 aplicación objetivo, ALT, requerida para soportar una tasa de reproducción deseada, PBR, para la sesión de transmisión de datos en continuo del dispositivo cliente, y para determinar un tamaño de segmento actual del segmento actual recibido;
 caracterizado porque el procesador de conformación de tasa (620) es utilizable además para registrar un tiempo de recepción del mensaje de petición de contenido actual, para registrar un tiempo de recepción del mensaje de
 65 respuesta de contenido actual y para determinar una tasa ALT actual para el actual segmento a partir del tiempo de

recepción del mensaje de petición de contenido actual, el tiempo de recepción del mensaje de respuesta de contenido actual y el tamaño del segmento actual;

el procesador de configuración de tasa también es utilizable para determinar si la tasa de ALT actual es mayor que la tasa de ALT objetivo determinada; y

5 cuando se determina (S819) que la tasa de ALT actual es mayor que la tasa de ALT objetivo, el procesador de conformación de tasa se utiliza además para retener (S823) el segmento de contenido actual durante un tiempo de retención determinado para disminuir la tasa ALT actual hasta la tasa ALT objetivo y para liberar (S825) el segmento de contenido actual al finalizar el tiempo de retención, y siendo utilizable además la interfaz de dispositivo de cliente para proporcionar el segmento de contenido actual al dispositivo del cliente una vez que el procesador de configuración de tasa libera la respuesta,

10 en el que cuando el tamaño del segmento es más pequeño que un tamaño de segmento promedio de la tasa de reproducción (S815), esos segmentos se excluyen entonces de consideración por el procesador de configuración de tasa, y la interfaz del dispositivo del cliente se utiliza además para liberar (S817) el segmento de contenido actual al dispositivo cliente;

15 cuando se determina (S819) que la tasa de ALT actual no es mayor que la tasa de ALT objetivo, el procesador de conformación de la tasa es además utilizado para liberar rápidamente (S821) el segmento de contenido actual, mientras que la interfaz del dispositivo del cliente es utilizado además para proporcionar el segmento de contenido actual al dispositivo del cliente una vez que la respuesta es liberada por el procesador de configuración de tasa.

20 4. Dispositivo de configuración de tasa de flujo de capa de aplicación de acuerdo con la reivindicación 3: en el que, el procesador de configuración de tasa se utiliza además para determinar una PBR actual asociada con el segmento de contenido actual de la sesión de difusión de datos en continuo del dispositivo de cliente;

en el que antes de llevar a cabo la determinación de si la tasa de ALT actual es mayor que una tasa de ALT objetivo, el procesador de conformación de tasa (620) puede funcionar adicionalmente para determinar si la PBR actual es mayor que la PBR deseada; y

25 en el que, cuando se determina que la PBR actual es mayor que la PBR deseada, el procesador de conformación de tasa (620) puede entonces funcionar para proceder a determinar si la tasa ALT actual es mayor que una tasa ALT objetivo, y

en el que, cuando se determina que la PBR actual no es mayor que la PBR deseada, el procesador de conformación de tasa (620) puede funcionar adicionalmente para liberar rápidamente el segmento de contenido actual, mientras que la interfaz del dispositivo de cliente puede funcionar además para proporcionar el segmento de contenido actual al dispositivo cliente una vez que el procesador de configuración de tasa (620) libera la respuesta, y en el que la determinación de si la tasa ALT actual es mayor que una tasa ALT objetivo junto con las funciones asociadas con el resultado de esa determinación, no se llevan a cabo.

35 5. Procedimiento de acuerdo con la reivindicación 1: en el que, antes de la etapa para determinar si la tasa de ALT actual es mayor que la tasa de ALT objetivo requerida para soportar la PBR deseada para la sesión de difusión en continuo de datos del dispositivo cliente, el procedimiento comprende además determinar si el segmento de contenido actual debe ser filtrado;

40 en el que, cuando se determina que el segmento de contenido actual debe filtrarse, el procedimiento comprende además proporcionar rápidamente el segmento de contenido actual al dispositivo cliente, mientras la etapa de determinar si la tasa de ALT actual es mayor que la tasa de ALT objetivo, junto con las etapas asociadas con el resultado de esta determinación, no se llevan a cabo; y

45 en el que, cuando se determina que el segmento de contenido actual no debe filtrarse, el procedimiento pasa a la etapa para determinar si la tasa de ALT actual es mayor que la tasa de ALT objetivo requerida para soportar la PBR deseada para la sesión de difusión en continuo de datos del dispositivo cliente.

50 6. Procedimiento de acuerdo con la reivindicación 5, en el que la determinación de si el segmento de contenido actual debe filtrarse está basada en el tamaño del segmento del segmento de contenido actual.

7. Procedimiento de acuerdo con la reivindicación 2, en el que la determinación de la PBR actual comprende: determinar una PBR medida para el segmento de contenido actual; y determinar la PBR actual como un promedio estadístico de la PBR medida para el segmento de contenido actual y la PBR medida para segmentos de contenido recibidos previamente del archivo de datos de difusión en continuo.

55 8. Procedimiento de acuerdo con la reivindicación 7, en el que la PBR medida ($PBR_{Measured}$) para el segmento de contenido actual se determina como:

$$PBR_{Measured} = \frac{\text{Tamaño segmento actual}}{\text{duración temporal del segmento actual}}$$

60 9. Procedimiento según la reivindicación 1, en el que la ALT actual ($ALT_{Measured}$) para el segmento de contenido actual se determina como:

$$ALT_{Measured} = \frac{\text{Tamaño segmento actual}}{RRT_{Measured}}$$

en donde $RTT_{Measured}$ es un tiempo de ida y vuelta medido para el segmento de contenido actual determinado a partir del tiempo de recepción del mensaje de petición de contenido actual y el tiempo de recepción del mensaje de respuesta de contenido actual.

- 5 10. Procedimiento de acuerdo con la reivindicación 9, en el que el tiempo de retención ($Delay_{Add}$) para el segmento de contenido actual se determina como:

$$Delay_{Add} = RTT_{Measured} * \left(\frac{ALT_{Measured}}{ALT_{Target}} - 1 \right),$$

- 10 en donde ALT_{Target} es la tasa ALT objetivo requerida para soportar una tasa de reproducción deseada (PBR) para la sesión de difusión en continuo de datos del dispositivo cliente.

11. Dispositivo de conformación de tasa de flujo de la capa de aplicación de acuerdo con la reivindicación 3: en el que, antes de realizar la determinación de si la tasa de ALT actual es mayor que la tasa de ALT objetivo requerida para soportar la PBR deseada para la sesión de difusión en continuo de datos del dispositivo cliente, el procesador de configuración de tasa está configurado para determinar si el segmento de contenido actual debe ser filtrado;

- 15 donde, en un caso en el que el procesador de configuración de tasa determina que el segmento de contenido actual debe ser filtrado, el procesador de configuración de tasa puede funcionar para liberar rápidamente el segmento de contenido actual, mientras que la interfaz del dispositivo cliente puede funcionar para proporcionar el segmento de contenido actual al dispositivo cliente una vez que el procesador de configuración de tasa (620) libera la respuesta; y
20 donde, en un caso en el que el procesador de configuración de tasa determina que el segmento de contenido actual no debe ser filtrado, el procesador de configuración de tasa puede funcionar para proceder con la determinación de si la tasa ALT actual es mayor que la tasa ALT objetivo requerida para soportar la PBR deseada para la sesión de
25 difusión en continuo de datos del dispositivo cliente.

12. Dispositivo de conformación de la tasa de flujo de capa de aplicación de acuerdo con la reivindicación 11, en el que la determinación de si el segmento de contenido actual debe filtrarse se basa en el tamaño de segmento del segmento de contenido actual.

- 30 13. Dispositivo de conformación de la tasa de flujo de capa de aplicación según la reivindicación 4, en el que la determinación de la PBR actual comprende:
determinar una PBR medida para el segmento de contenido actual; y
35 determinar la PBR actual como un promedio estadístico de la PBR medida para el segmento de contenido actual y la PBR medida para segmentos de contenido recibidos previamente del archivo de datos de difusión en continuo.

14. Dispositivo de conformación de tasa de flujo de capa de aplicación de acuerdo con la reivindicación 13, en el que la PBR medida ($PBR_{Measured}$) para el segmento de contenido actual se determina como:

40
$$PBR_{Measured} = \frac{\text{Tamaño segmento actual}}{\text{duración temporal del segmento actual}}$$

15. Dispositivo de conformación de tasa de flujo de capa de aplicación según la reivindicación 3, en el que la ALT actual ($ALT_{Measured}$) para el segmento de contenido actual se determina como:

45
$$ALT_{Measured} = \frac{\text{Tamaño segmento actual}}{RTT_{Measured}}$$

en donde $RTT_{Measured}$ es un tiempo de ida y vuelta medido para el segmento de contenido actual determinado a partir del tiempo de recepción del mensaje de petición de contenido actual y el tiempo de recepción del mensaje de respuesta de contenido actual.

- 50 16. Dispositivo de conformación de tasa de flujo de capa de aplicación de acuerdo con la reivindicación 15, en el que el tiempo de retención ($Delay_{Add}$) para el segmento de contenido actual se determina como:

55
$$Delay_{Add} = RTT_{Measured} * \left(\frac{ALT_{Measured}}{ALT_{Target}} - 1 \right),$$

en donde ALT_{Target} es la tasa de ALT objetivo requerida para soportar una tasa de reproducción deseada, PBR, para la sesión de difusión en continuo de datos del dispositivo cliente.

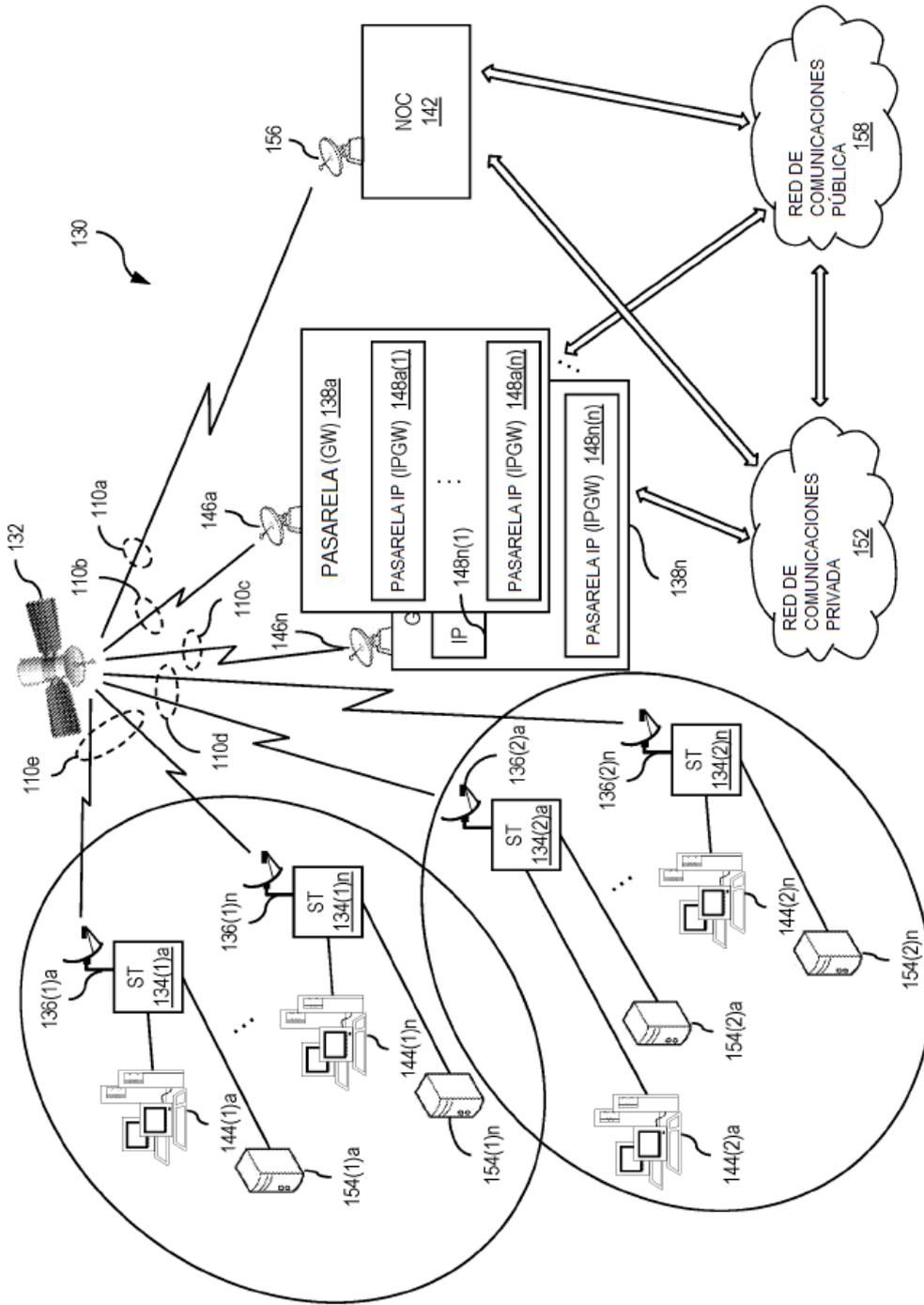


FIG. 1

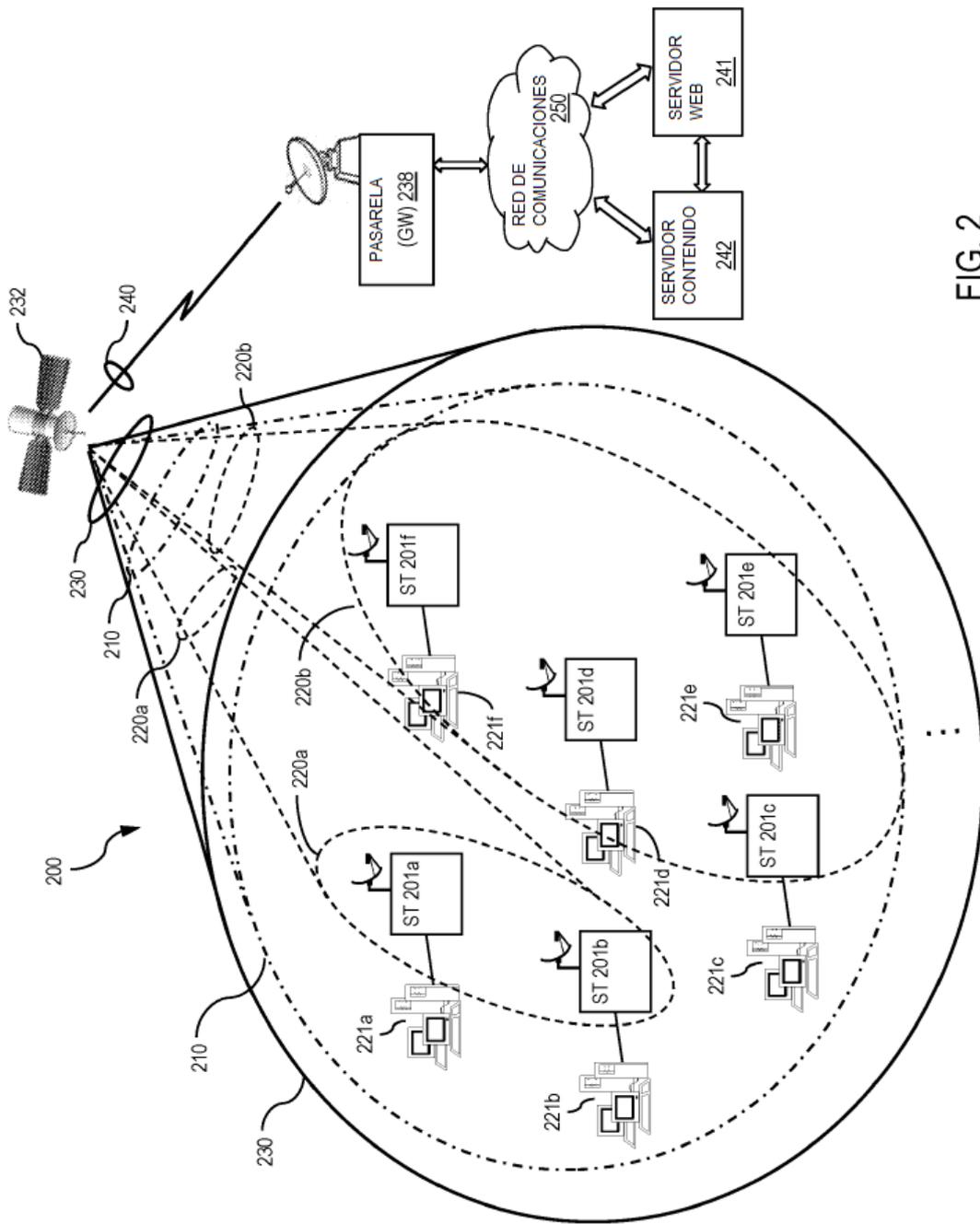


FIG. 2

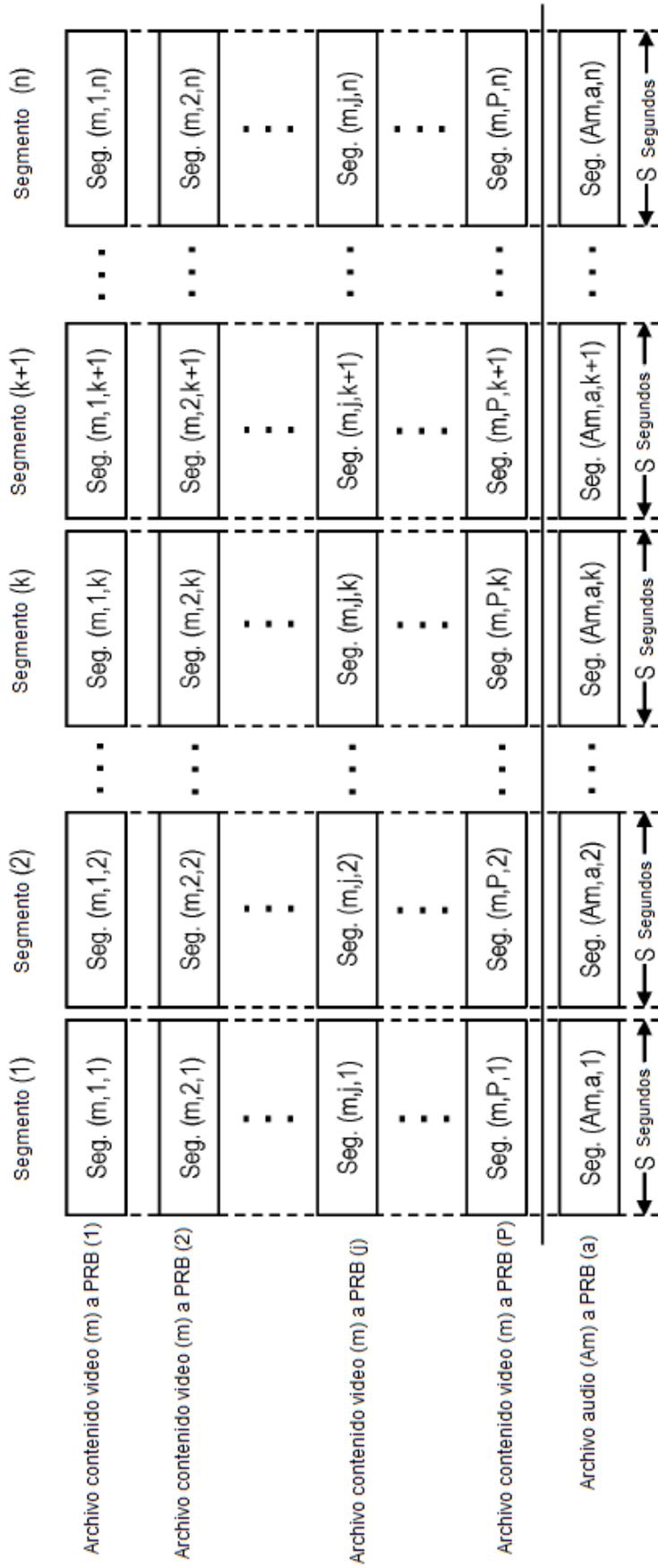


FIG. 3

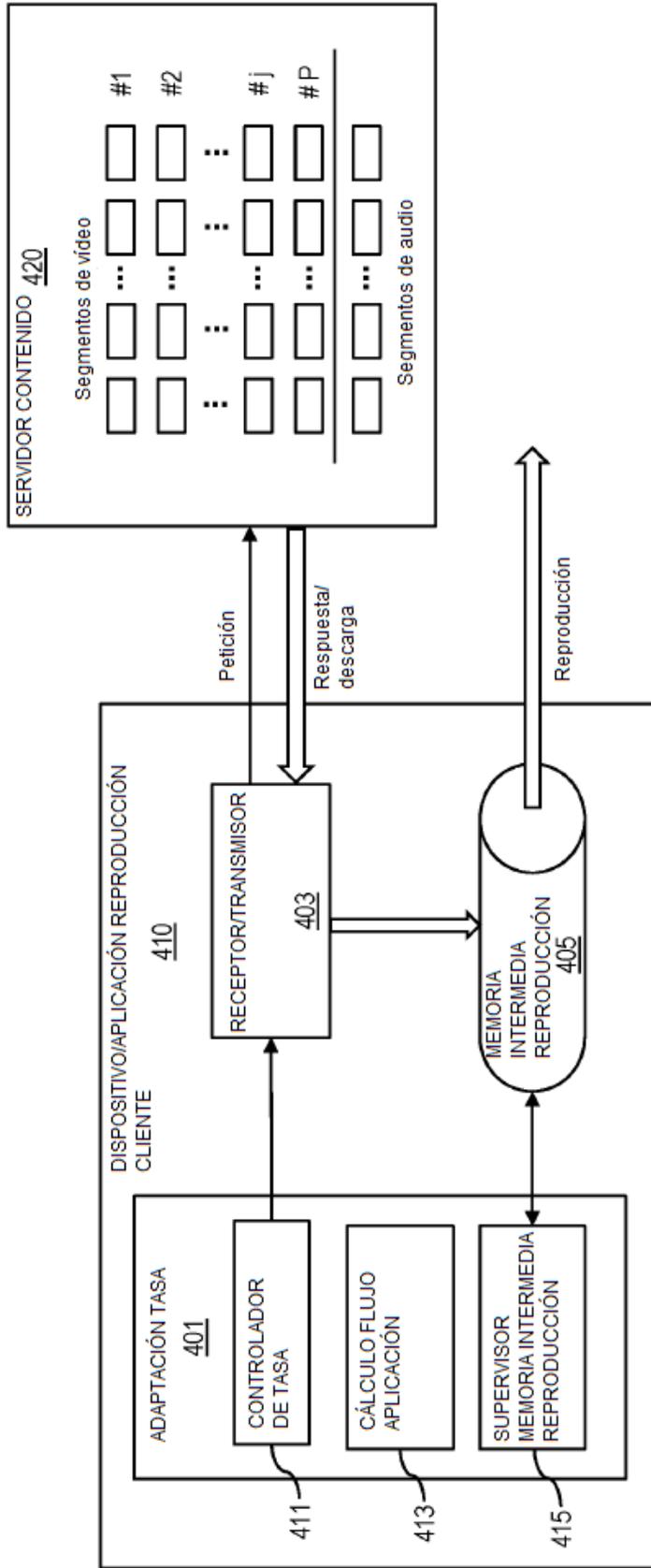


FIG. 4

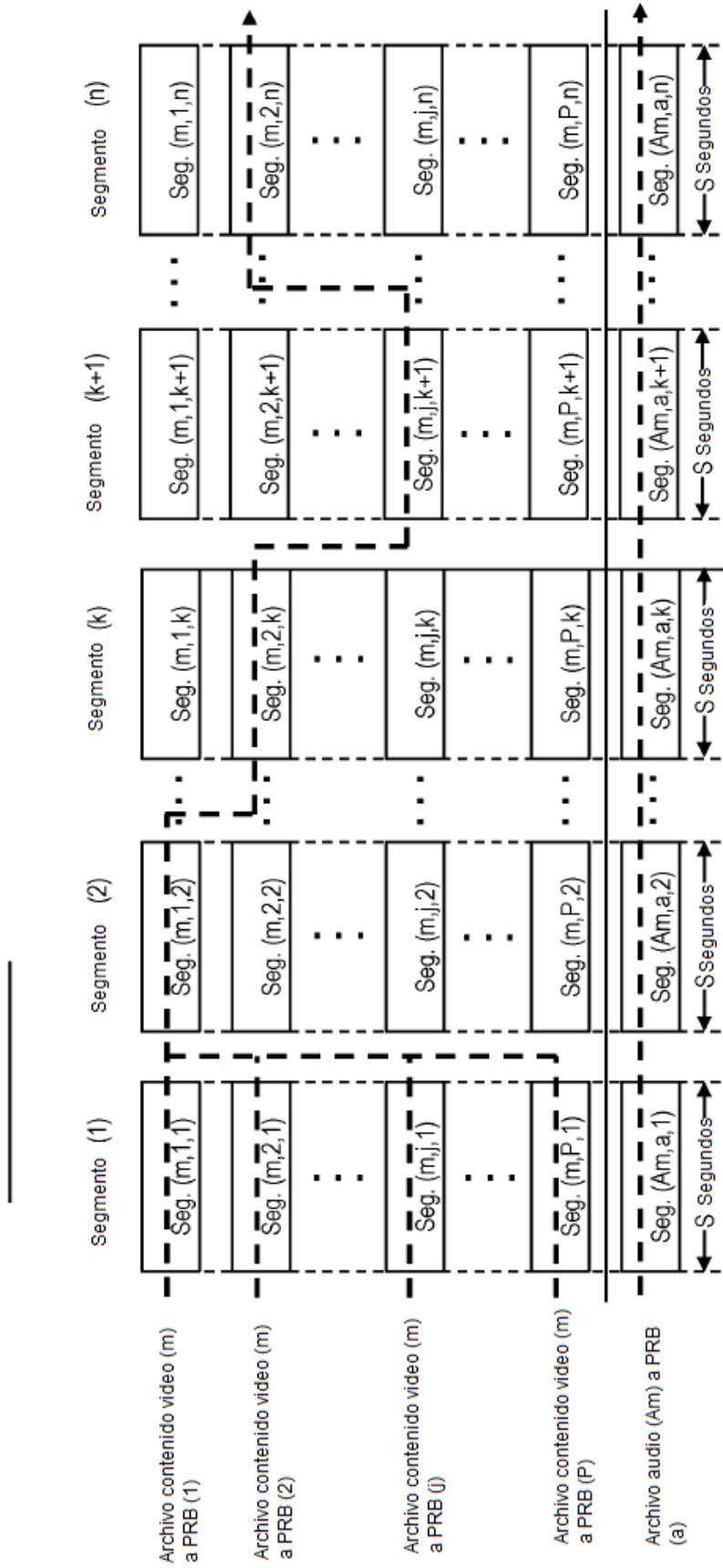


FIG. 5

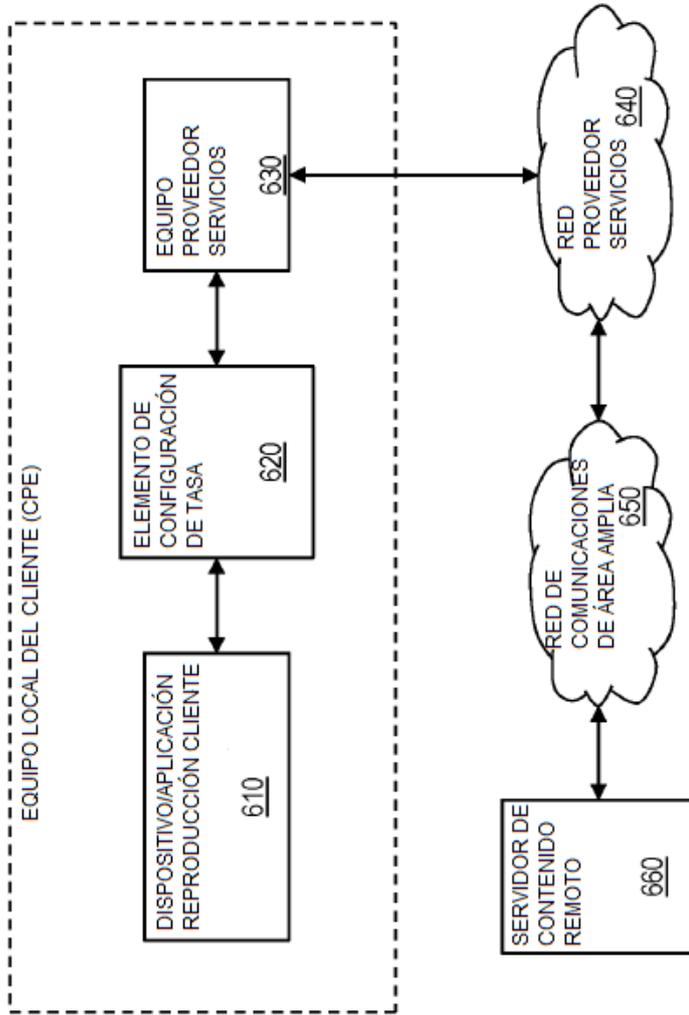


FIG. 6

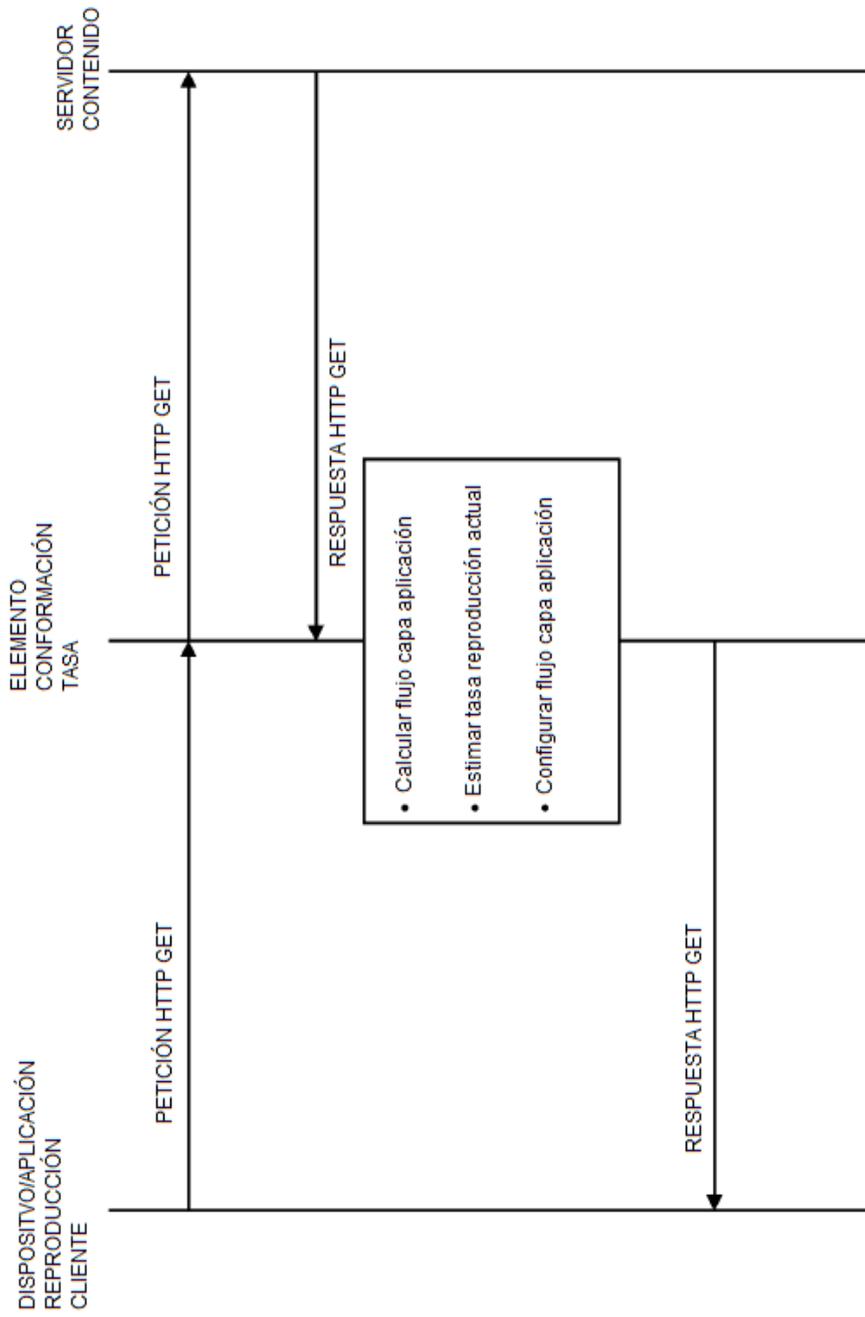


FIG. 7

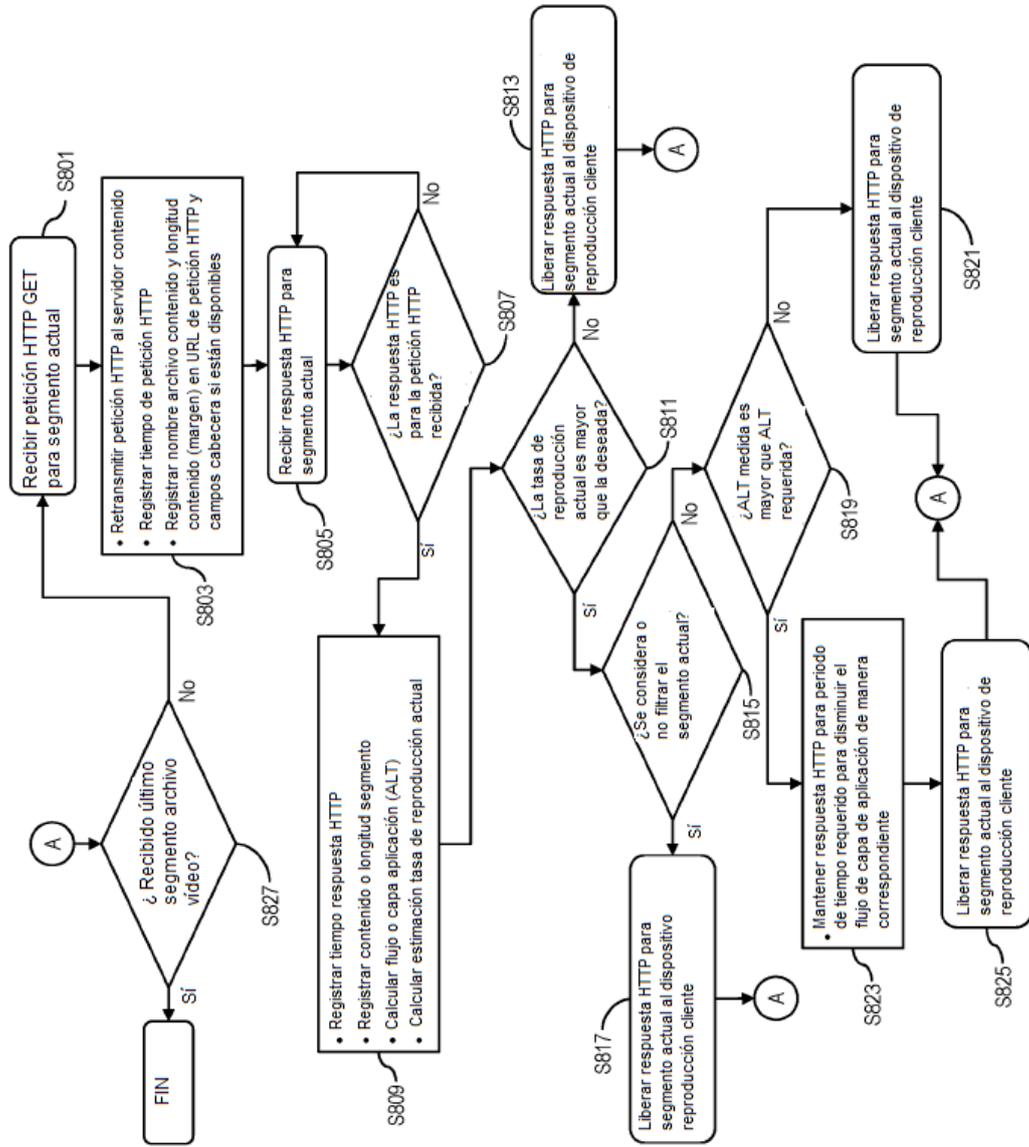


FIG. 8

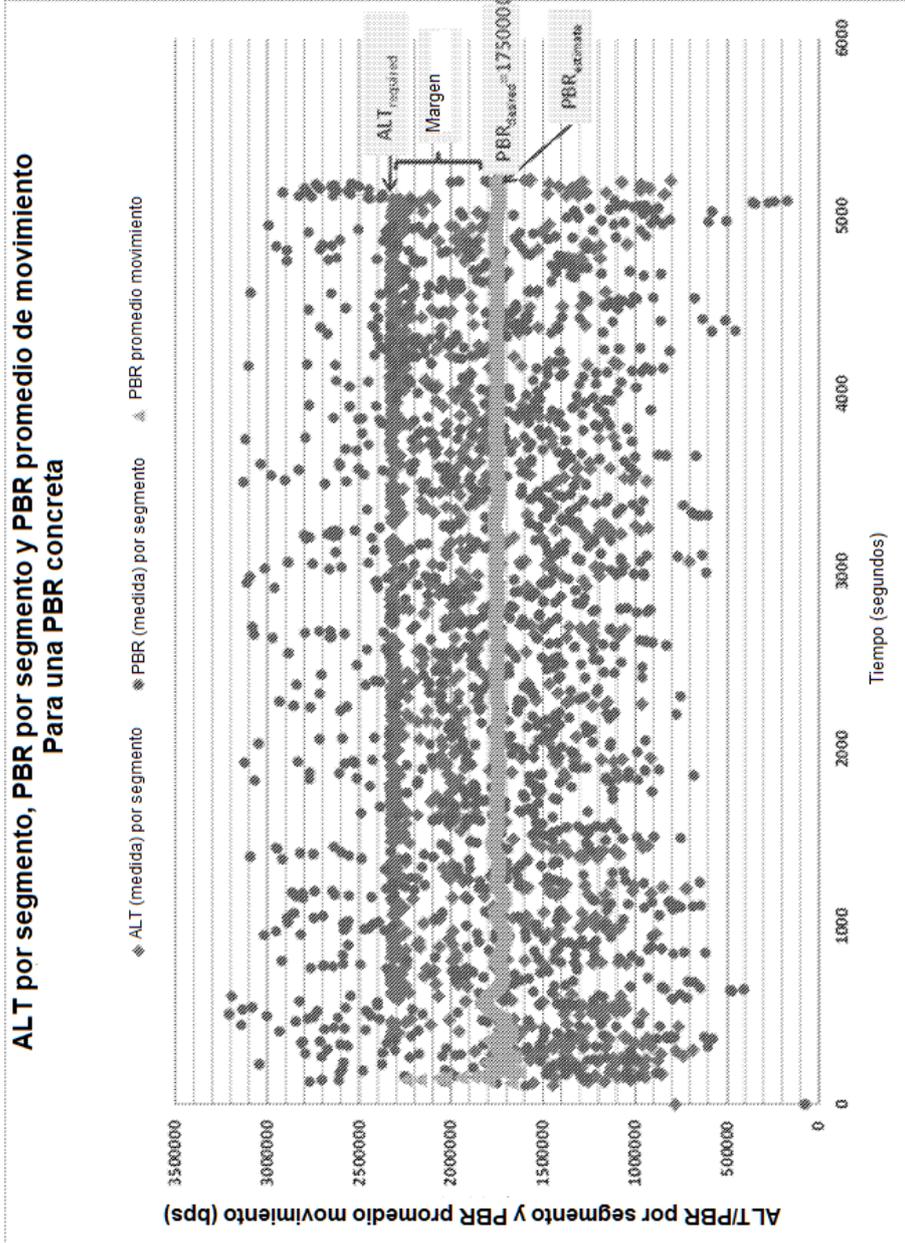


FIG. 9

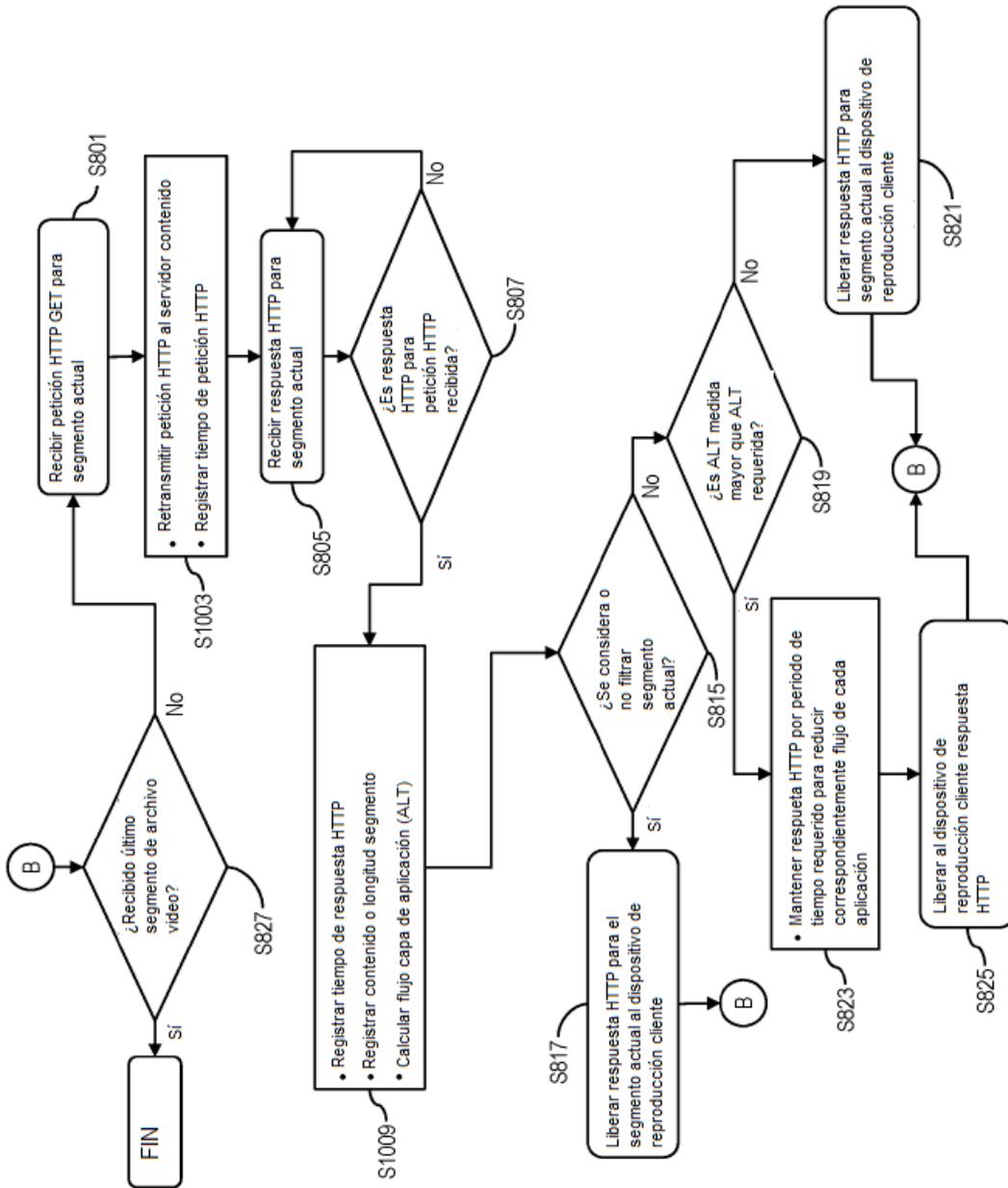


FIG. 10

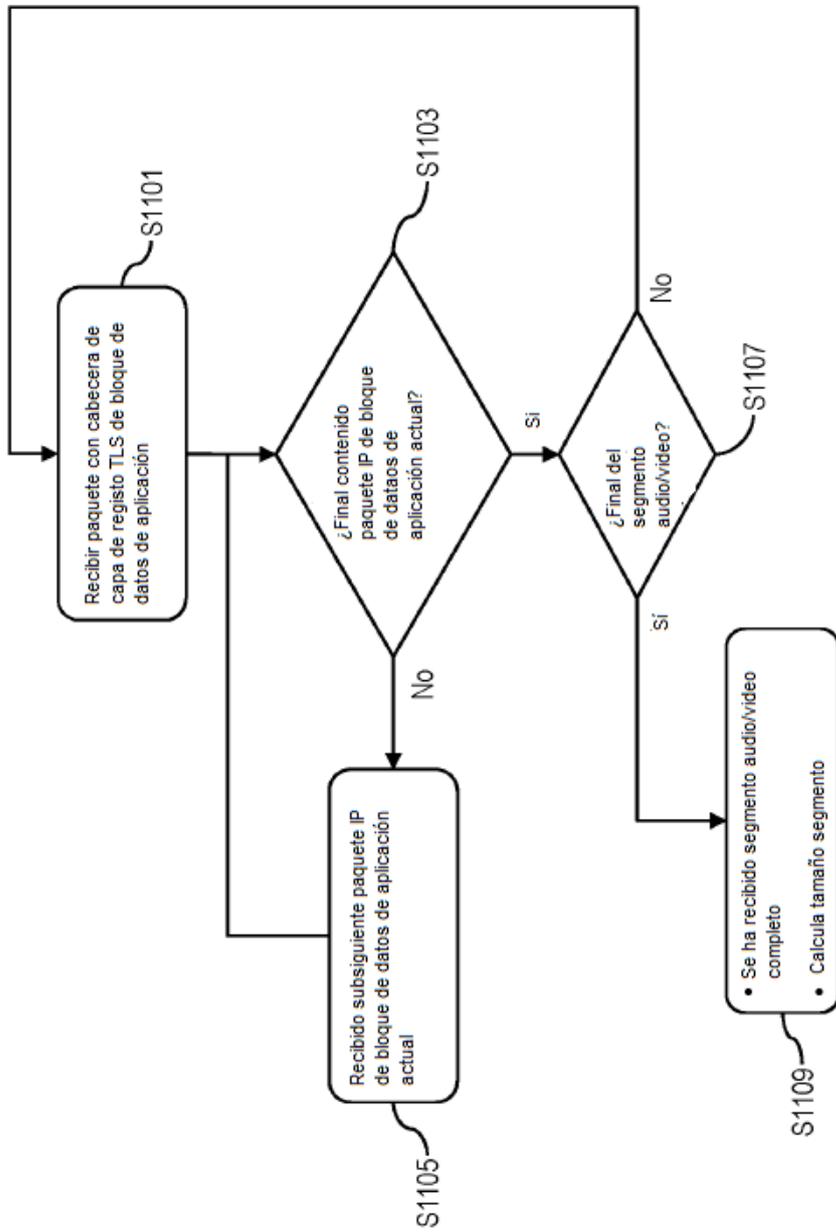


FIG. 11

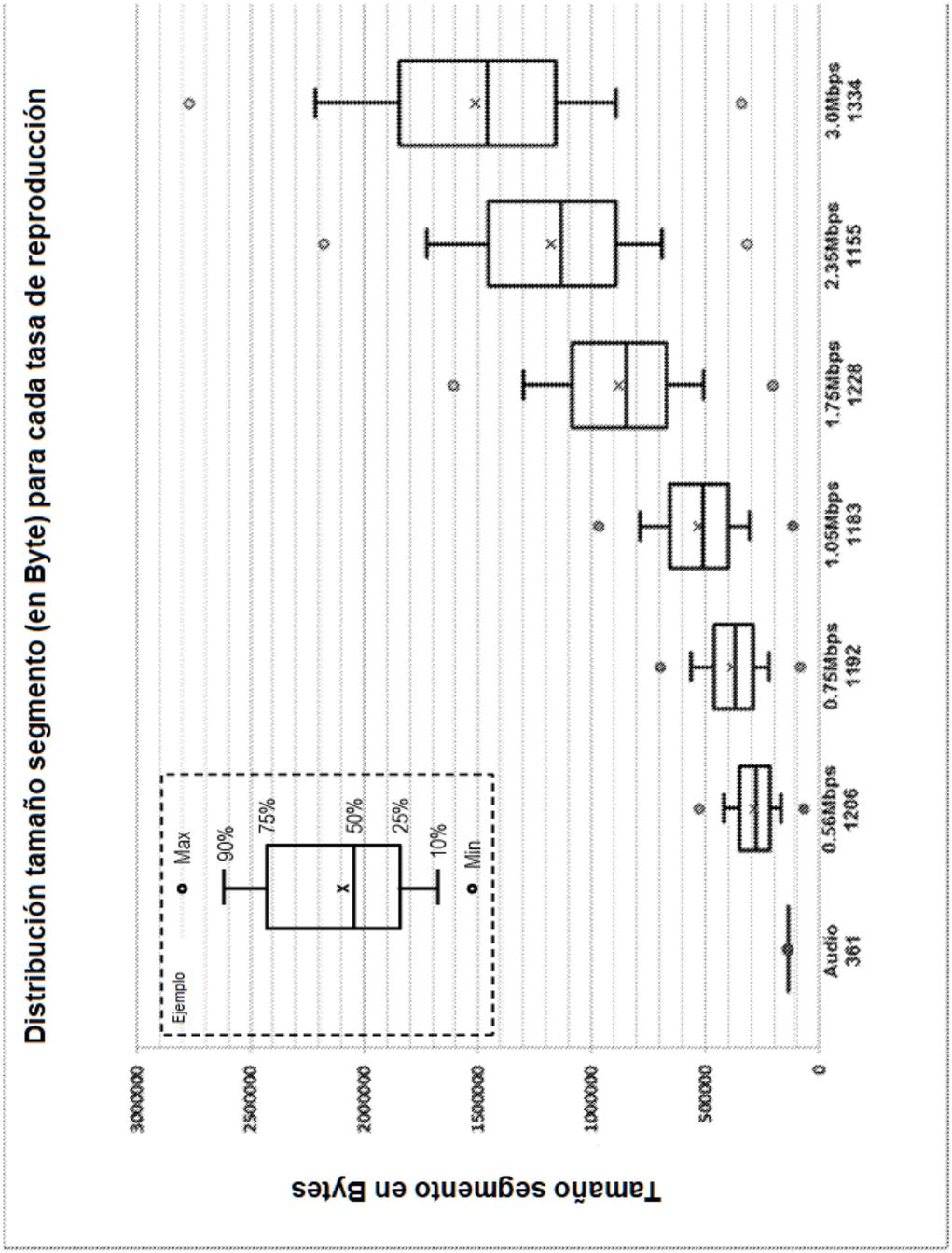


FIG. 12

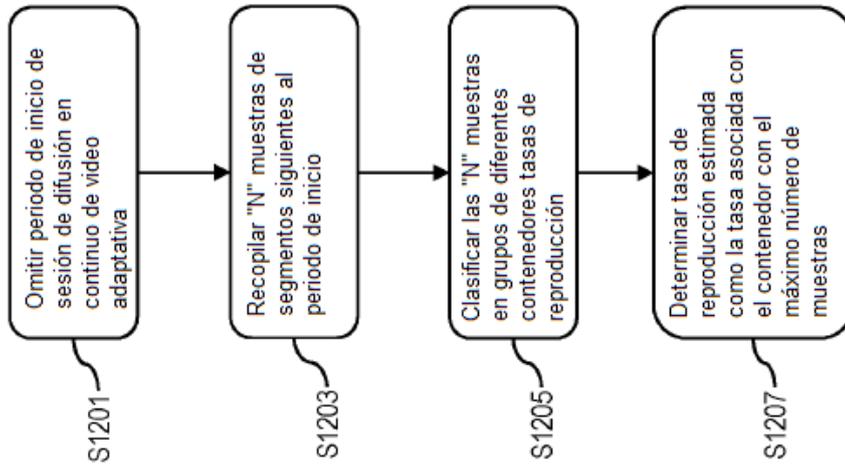
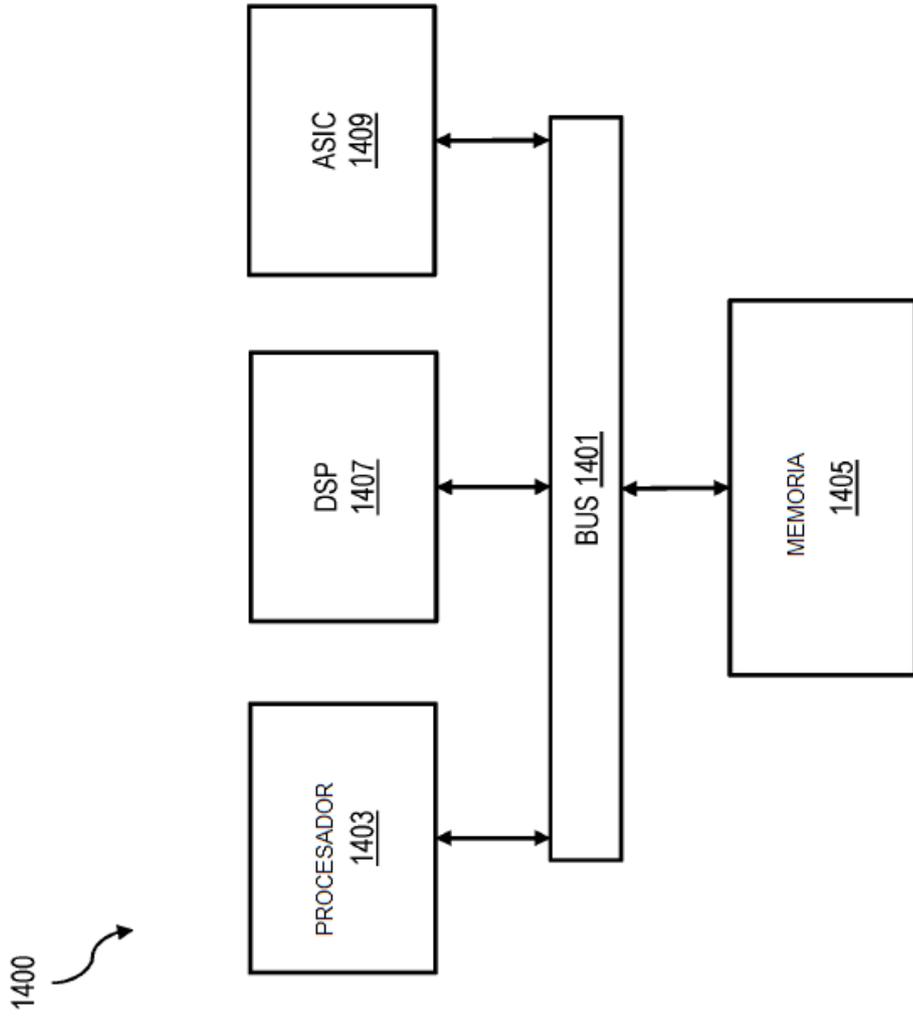


FIG. 13

FIG. 14



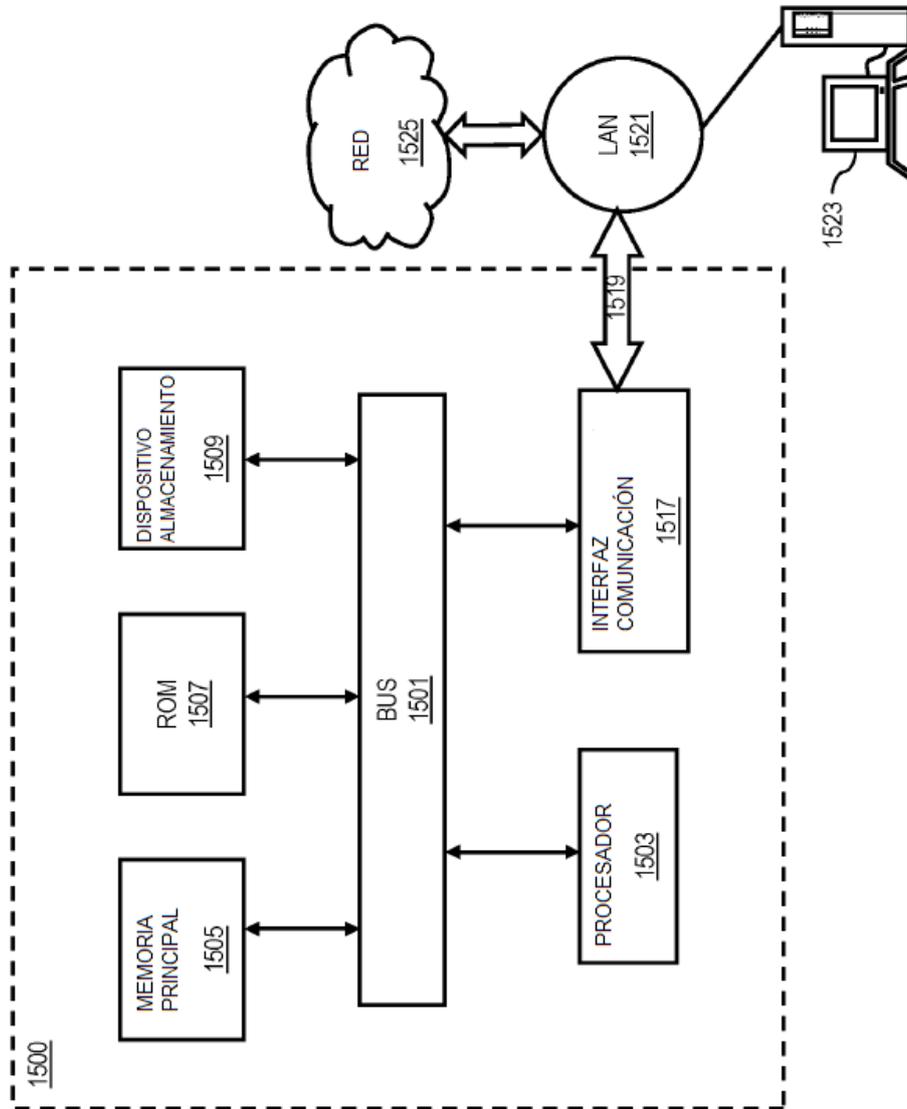


FIG. 15

REFERENCIAS CITADAS EN LA DESCRIPCIÓN

5 La lista de referencias citada por el solicitante lo es solamente para utilidad del lector, no formando parte de los documentos de patente europeos. Aún cuando las referencias han sido cuidadosamente recopiladas, no pueden excluirse errores u omisiones y la OEP rechaza toda responsabilidad a este respecto.

Documentos de patente citado en la descripción

- US 2013097309 A1 [0009]
- US 2015026309 A1 [0010]

10 **Bibliografía no de patentes citada en la descripción**

- CISCO. Policing and Shaping Overview. *IOS QoS Solutions Configuration Guide, Rel. 12.2*, www.cisco.com/c/en/us/td/docs/ios/12_2/qos/configuration/guide/fqos_c/qcfcplsh.pdf [0005]
- Extended Linux HTB Queuing Discipline Implementations. **DORU GABRIEL BALAN ; DAN ALIN POTORAC**. First International Conference on Networked Digital Technologies. IEEE, 28 July 2009, 122-126 [0006]
- Towards dynamic QoS-aware over-the-top video streaming. **HYUNWOO NAM ; KYUNG HWA KIM ; BONG HO KIM ; DORU CALIN ; HENNING SCHULZRINNE**. IEEE 15th International Symposium on a World of Wireless. Mobile and Multimedia Networks, 19 June 2014, 1-9 [0007]
- Trickle: rate limiting YouTube video streaming. **MONIA GHOBADI ; YUCHUNG CHENG ; ANKUR JAIN ; MATT MATHIS**. USENIX Annual Technical Conference. USENIX Association, 2012, 191-196 [0008]