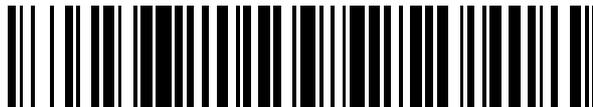


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 558 864**

51 Int. Cl.:

**H04L 12/28** (2006.01)

**H04L 29/06** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **11.02.2004 E 04710206 (6)**

97 Fecha y número de publicación de la concesión europea: **14.10.2015 EP 1593234**

54 Título: **Método y aparato para medir disponibilidad, calidad y rendimiento de flujo**

30 Prioridad:

**11.02.2003 US 364753**

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

**09.02.2016**

73 Titular/es:

**AKAMAI TECHNOLOGIES, INC. (100.0%)  
150 Broadway  
Cambridge MA 02142, US**

72 Inventor/es:

**SITARAMAN, RAMESH K. y  
BARTON, REID W.**

74 Agente/Representante:

**DE ELZABURU MÁRQUEZ, Alberto**

**ES 2 558 864 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**

Método y aparato para medir disponibilidad, calidad y rendimiento de flujo

**Antecedentes de la invención**

Campo técnico

- 5 La presente invención se refiere de manera general a medición del rendimiento de entrega de medios de difusión en forma continua.

Descripción de la técnica relacionada

10 Los medios de difusión en forma continua son un tipo de contenido de Internet que tiene la importante característica de ser capaz de ser reproducido mientras aún está en el proceso de ser descargado. Una máquina cliente puede reproducir el primer paquete del flujo y descomprimir el segundo, mientras que se recibe el tercero. De esta manera, un usuario final puede comenzar a disfrutar el multimedia sin esperar al final de la transmisión. La difusión en forma continua es muy útil para entregar medios debido a que los ficheros de medios tienden a ser grandes, particularmente a medida que aumenta la duración de la programación. Ciertamente, para eventos en directo, el tamaño de fichero, en realidad, es infinito. Para ver un fichero de medios que no se difunde en forma continua, los usuarios deben descargar primero el fichero a un disco duro local – lo cual puede tardar minutos o incluso horas – y luego abrir el fichero con software reproductor que sea compatible con el formato de fichero. Para ver medios de difusión en forma continua, el navegador del usuario abre el software reproductor, el cual almacena temporalmente el fichero durante unos pocos segundos y luego reproduce el fichero mientras que lo descarga simultáneamente. A diferencia de las descargas software, los ficheros de medios de difusión en forma continua no se almacenan localmente en el disco duro de un usuario. Una vez que se usan los bits que representan contenido, el reproductor típicamente los descarta.

15 Los medios de difusión en forma continua normalmente se basan en un software de servidor y cliente propietario. El servidor, el cliente, las herramientas de producción y codificación desarrolladas por un proveedor del software de difusión en forma continua se conocen colectivamente como formato. Los medios de difusión en forma continua codificados en un formato particular se deben servir por un servidor de medios de ese formato y reproducir por un cliente de ese formato. Los clientes de medios de difusión en forma continua a menudo se llaman reproductores y típicamente existen como complementos a navegadores Web. Los tres principales formatos de medios de difusión en forma continua en uso hoy en día son: RealNetworks RealSystem G2, Microsoft Windows Media Technologies (“WMT”) y Apple QuickTime.

20 Un número en constante crecimiento de proveedores de contenidos, empresas y redes de entrega de contenidos utilizan y entregan medios de difusión en forma continua. Como consecuencia, ha llegado a ser importante cuantificar, de una forma científica y medible, lo bien que se entregó un flujo dado, así como idear métricas relevantes que capturen la experiencia del usuario final del flujo. Los reproductores de medios clientes tienen la capacidad de generar estadísticas básicas acerca de los flujos que entregan. Desafortunadamente, la correlación de algunas de estas estadísticas con una experiencia de usuario final de alta calidad es subjetiva y de esta manera extremadamente difícil de medir eficazmente. Por ejemplo, la tasa de tramas, una métrica usada comúnmente notificada por reproductores de medios, puede no ser un indicador robusto de la calidad de flujo debido a que los reproductores pueden interpolar e insertar tramas de calidad variable en un flujo. Como otro ejemplo, algunos reproductores de medios notifican una métrica de ancho de banda que no indica con precisión si se reciben o no los octetos de una forma oportuna de manera que se puedan usar para representación.

25 Hay diversas compañías, tales como Keynote, StreamCheck y otras, que proporcionan soluciones de monitorización de difusión en forma continua parcial o completamente externalizadas. En un sistema ilustrativo, ordenadores de medición de difusión en forma continua (algunas veces llamados “agentes”) conectan a los URL y recaban estadísticas relevantes, tales como el tiempo de almacenamiento temporal inicial, la información de paquetes y la calidad de flujo total. Estos agentes de medición se colocan típicamente estratégicamente alrededor de Internet para representar con precisión la experiencia de un usuario final que conecta a un URL de sitio dado. Los proveedores de tales soluciones de monitorización de rendimiento externalizadas típicamente proporcionan datos de rendimiento en forma de informes en tiempo real e históricos y alarmas basadas en umbral.

30 A modo de antecedentes adicionales, también se conoce en la técnica anterior que la medición de rendimiento de difusión en forma continua es significativamente más difícil que las mediciones de descarga web por diversas razones. A diferencia del rendimiento web, donde basta medir tiempos de descarga de objetos, los flujos pueden mostrar degradación de formas significativamente complejas que necesitan ser cuantificadas y medidas. A diferencia del protocolo HTTP, un protocolo abierto, que mide rendimiento de flujo implica tener en cuenta protocolos de difusión en forma continua propietarios complejos, tales como RTSP (Real) y MMS (Windows Media).

35 Uno puede distinguir tres pasos en poner a disposición contenido de medios de difusión en forma continua al usuario final: los medios se capturan y codifican, los medios se entregan sobre la red al cliente, es decir, al reproductor de medios del usuario final y los medios se representan por el reproductor de medios para ser vistos por el usuario final.

El proceso de entrega de medios de difusión en forma continua sobre la red puede degradar la calidad percibida por el usuario final. Una meta primaria de una red de entrega de contenidos (o "CDN") es asegurar que los medios se entregan a un usuario final solicitante con tan pocas distorsiones como sea posible desde su forma codificada original, maximizando por ello la experiencia del usuario final del flujo. Las causas raíz de degradación de la calidad de flujo son problemas que pueden ocurrir en la red subyacente que transporta los medios desde el servidor al cliente. Por ejemplo, la conectividad de red puede disminuir, lo cual limita la capacidad del cliente para hablar al servidor. El ancho de banda efectivo disponible entre el servidor y el cliente se puede reducir extremadamente. La red puede perder paquetes enviados desde el servidor al cliente. Más sutilmente, la red puede hacer que los paquetes sean reordenados en la ruta al cliente. O, los paquetes se pueden entregar por la red al cliente tarde, en cuyo caso los paquetes no serán útiles en la reproducción de flujo.

Es instructivo examinar cómo el cliente y el servidor reaccionan a estos problemas de red de una manera que degrada la calidad de flujo. Como se señaló anteriormente, típicamente el cliente y el servidor comunican usando un protocolo propietario (tal como RTSP de Real o MMS de Windows Media) para controlar la manera en que los datos se difunden en forma continua desde el servidor al cliente. El protocolo cliente-servidor está diseñado típicamente para superar cantidades moderadas de degradación de red. El reproductor típicamente almacena temporalmente varios segundos de datos delante de lo que está siendo representado actualmente. Este almacenamiento temporal ayuda al reproductor a suavizar cualquier variación en la tasa a la que se están recibiendo los paquetes, incluyendo entrega de paquetes fuera de orden. Además, el almacenamiento temporal proporciona una oportunidad al reproductor para tolerar una cierta cantidad de pérdida de paquetes, debido a que los paquetes perdidos se pueden volver a solicitar desde el servidor y recibir a tiempo para su reproducción. Consecuentemente, se protege eficazmente una degradación moderada en la red desde el usuario.

A medida que la red se degrada más, no obstante, uno podría ver periodos prolongados de reducción en el ancho de banda disponible o una cantidad significativa de paquetes perdidos o un número significativo de paquetes que llegan tarde. El servidor típicamente reacciona a tal situación enviando menos que el ancho de banda codificado completo al cliente, provocando una reproducción adelgazada. El servidor logra este adelgazamiento enviando solamente los octetos "más importantes" y tirando el resto. Por ejemplo, el servidor puede enviar solamente las tramas de vídeo claves y audio y no enviar el resto del contenido. En el caso extremo, el adelgazamiento provoca el denominado efecto de "presentación de diapositivas", donde se muestran tramas aisladas sin interpolación. Si el flujo se codifica a múltiples tasas de bit, el adelgazamiento se efectúa más sin discontinuidad, por ejemplo, mediante el servidor que conmuta a una réplica del flujo codificado a una tasa de bit inferior. Típicamente, el ancho de banda difundido en forma continua desde el servidor al cliente se controla por un bucle de realimentación, donde el cliente estima continuamente el ancho de banda disponible y lo notifica al servidor y el servidor intenta ajustar el flujo al ancho de banda disponible. Una degradación incluso más drástica en la red, por ejemplo, una reducción rápida en el ancho de banda disponible, puede empobrecer el almacenador temporal de datos del reproductor. Cuando el reproductor no puede continuar la reproducción, se congela para rellenar su almacenador temporal antes de reanudar de nuevo, provocando una reproducción interrumpida. El reproductor se diseña típicamente para volver a almacenar temporalmente solamente bajo condiciones de red extremas.

Hay una necesidad en la técnica de proporcionar métodos mejorados para medir calidad y rendimiento de flujo. La presente invención aborda esta necesidad.

El documento GB2355901-A describe un sistema para medir retardos a través de equipamiento en una red de paquetes. En diversas etapas durante, por ejemplo, la generación de una señal de prueba y transmisión de paquetes de flujo de audio, los paquetes maestros se generan por el generador de paquetes en respuesta a desencadenar señales desde el generador de señal. La precisión de la medición se puede mejorar situando la posición de la señal de prueba codificada en los paquetes de flujo de audio.

Pagani E et al, "Distributed bandwidth broker for G05 Multicast Traffic"; IEEE Comp. Soc., EE.UU., Vol CONF. 22, 2 de julio de 2002, p 289-296 XP010595543, ISBN: 978-0-7695-1585-4; trata un protocolo basado en medición distribuida propuesta que realiza funcionalidades de control de admisión para tráfico de multidifusión en redes de diferentes servidores.

### Breve compendio de la invención

Un objeto de la presente invención es proporcionar un agente de medición de difusión en forma continua que se diseña para experimentar, medir y notificar en un flujo de medios como un usuario final real experimentaría el flujo. Preferiblemente, esto se logra teniendo el agente que reside transparentemente dentro de un reproductor de medios en sí mismo de manera que el agente pueda monitorizar los flujos de paquetes de flujo dentro del reproductor a medida que están siendo reproducidos los flujos medidos. En una realización ilustrativa, el agente comprende un módulo de monitorización de rendimiento (PMM), que es un software que reside preferiblemente en una interfaz entre un módulo central existente y un módulo de representación (un "representador") de un reproductor de medios. El agente PMM intercepta preferiblemente cada paquete útil como va desde el núcleo al módulo de representación y, como resultado, el agente es capaz de calcular métricas de calidad acerca de la representación de una forma altamente precisa. El agente funciona "transparentemente" al reproductor de medios presentando el núcleo con una interfaz de programación de aplicaciones (API) que emula la API que el representador presenta normalmente al

núcleo. De esta manera, cuando el núcleo cree que está llamando al representador, el núcleo está llamando realmente al agente PMM, que entonces puede recibir todos los paquetes desde el núcleo y procesarlos. Después de calcular las métricas de rendimiento relevantes usando los paquetes recibidos, el agente PMM llama al representador y desde la perspectiva del representador el agente juega el mismo papel que el núcleo.

- 5 Según un rasgo de la invención, el agente puede calcular una o más métricas de calidad y rendimiento interceptando paquetes útiles que fluyen dentro del reproductor de medios. Generalmente, una estadística dada se genera a partir del paquete útil interceptado. Como se usa en la presente memoria, una estadística dada es alguna función arbitraria que se puede calcular a partir de la secuencia de paquetes útiles que se interceptan. Preferiblemente, un paquete útil es uno que tiene un conjunto dado de propiedades, por ejemplo, no se ha adelgazado por un servidor de difusión en forma continua durante la entrega, no se perdió de una manera irrecuperable durante la entrega y llegó al reproductor a tiempo para ser útil para su reproducción. En una realización particular, el agente usa estos paquetes para calcular un ancho de banda de reproducción real, una medición precisa de la calidad de flujo que se percibe por un usuario final. Otras métricas relacionadas con paquetes y métricas relacionadas con ancho de banda se obtienen también monitorizando el flujo de paquetes a través del módulo de monitorización de rendimiento embebido. Adicionalmente, el agente puede usar funciones de reproducción de medios nativas (por ejemplo, una función de devolución de llamada) para obtener otra información útil tal como tiempo de puesta en marcha de flujo, interrupciones, componentes de tiempo de puesta en marcha y similares.

La monitorización del flujo de paquetes real dentro del reproductor de medios en sí mismo según la presente invención proporciona una nueva forma de evaluar la disponibilidad, calidad y rendimiento de flujo. Debido a todo o sustancialmente todo, el flujo de paquetes útiles a través del PMM, la presente invención proporciona una vista mejorada en la efectividad del proceso de entrega de flujo comparada con técnicas de la técnica anterior. El flujo de paquetes a través del PMM permite la generación de métricas de rendimiento importantes, tales como ancho de banda de reproducción real y estas métricas se pueden complementar con métricas generadas a través de otros medios, por ejemplo, usando funciones de devolución de llamada del reproductor.

Según otro aspecto de la presente invención, la tecnología de agente descrita anteriormente se despliega de una manera ampliamente distribuida para proporcionar mediciones de calidad y rendimiento de flujo desde una pluralidad de ubicaciones. Comparado con las técnicas de la técnica anterior, los datos recabados desde estos agentes pueden notificarse con más precisión sobre con cuánto éxito está siendo entregado un flujo desde un servidor dado a un conjunto de reproductores de medios.

Según un aspecto adicional de la presente invención, un servicio de medición de calidad y rendimiento de flujo gestionado se proporciona por un proveedor de servicios en nombre de los proveedores de contenidos participantes, por ejemplo, entidades que usan una red de entrega de contenidos (CDN) para facilitar entrega de medios de difusión en forma continua. El servicio utiliza un conjunto de agentes de rendimiento tal como se describió anteriormente que se despliegan en un conjunto diverso de redes y ubicaciones geográficas para medir el rendimiento sobre una amplia variedad de condiciones. El proveedor de servicios puede desplegar los agentes de medición en ubicaciones que de otro modo están libres de servidores de difusión en forma continua o cubiertas con tales servidores.

Lo precedente ha perfilado algunos de los rasgos más pertinentes de la presente invención. Estos rasgos se deberían interpretar que son meramente ilustrativos. Se pueden obtener muchos otros resultados beneficiosos aplicando la invención descrita de una manera diferente o modificando la invención como se describirá. Por consiguiente, otros rasgos y una comprensión más completa de la invención se pueden tener refiriéndose a la siguiente Descripción Detallada de la Realización Preferida.

#### Breve descripción de los dibujos

La Figura 1 es un diagrama de bloques altamente simplificado de un mecanismo de entrega de medios de difusión en forma continua;

La Figura 2 es un diagrama simplificado de un reproductor de medios de difusión en forma continua de la técnica anterior;

La Figura 3 es un diagrama de bloques de un sistema de medición de flujo que incluye un conjunto de agentes de monitorización que facilitan un servicio de medición de calidad y rendimiento del flujo gestionado; y

La Figura 4 es un diagrama de bloques del mecanismo agente inventivo.

#### Descripción detallada de la realización preferida

La Figura 1 ilustra un mecanismo de entrega de difusión en forma continua convencional en donde una máquina cliente 100 conecta a un servidor de medios de difusión en forma continua 102 a través de una red 104, tal como Internet, una intranet, una red de entrega de contenidos (CDN) o similares. La máquina cliente es típicamente un ordenador que tiene hardware básico 106, software de sistema operativo 108, un navegador Web 110 y un reproductor de medios de difusión en forma continua 112. En una operación típica, un usuario final lanza su

navegador a un Localizador de Recursos Uniforme (URL) que identifica el flujo de medios deseado. Una vez que el cliente se conecta al servidor de medios 102, el servidor entrega el flujo sobre la red al navegador del cliente, en el que se representa por el reproductor de medios. Como se ilustra en la Figura 2, el reproductor de medios de difusión en forma continua 200 típicamente incluye dos (2) componentes primarios: un núcleo 202 y un representador 204.

5 Se debería señalar que la Figura 2 se supone que proporciona una vista simplificada para ilustrar la tecnología de reproductor conocida. En realidad, un reproductor de medios real, tal como Real y WMS, contendría varios submódulos con una estructura más compleja que implementa las funcionalidades de núcleo y representador descritas anteriormente. Sin embargo, es sencillo correlacionar las ideas perfiladas aquí a reproductores de medios específicos para construir agentes específicos para cada uno de los formatos bien conocidos que incluyen Real, WMS y Quicktime. Por supuesto, la tecnología de agente inventivo no está limitada para uso con estos formatos conocidos, ya que los métodos y técnicas descritos más adelante se pueden usar con cualquier dispositivo de representación de paquetes de flujo.

15 El núcleo 202 es responsable de transportar datos recibidos desde el servidor de medios. Típicamente, el núcleo 202 establece un canal de control para intercambiar información de control con el servidor de medios. Tal información puede incluir peticiones del reproductor para describir, reproducir, hacer una pausa, parar el flujo, realimentar desde el reproductor en el ancho de banda percibido, pérdida de paquetes y otras métricas, información desde el servidor en los medios que está siendo solicitada y así sucesivamente. Como se señaló anteriormente, el protocolo de control entre el núcleo y el servidor de medios es típicamente propietario, por ejemplo, RTSP para Real y MMS para Windows Media. Además del control de canal, el núcleo 202 típicamente establece un canal de datos sobre el cual el flujo de medios real fluye desde el servidor al cliente. El flujo de medios es una serie de paquetes que contienen información de audio, vídeo y texto. Los paquetes se deben recibir de una forma oportuna para que el flujo sea útil. En particular, cada paquete típicamente tiene un sello de tiempo que indica cuándo se debe representar y el núcleo 202 transfiere sobre (al representador 204 para ser representados) aquellos paquetes que llegan de una forma oportuna. El núcleo 202 puede recibir paquetes desde el servidor varios segundos delante del momento en que se deben representar y los paquetes que llegan delante del momento se almacenan en un almacenador temporal (no mostrado). Esto proporciona una oportunidad al núcleo 202 para volver a solicitar paquetes perdidos desde el servidor y para recuperarlos en el momento para que tales paquetes sean representados adecuadamente. El representador 204 recibe paquetes desde el núcleo 202 y representa los datos contenidos en el paquete. El representador contiene varias unidades funcionales, una para cada tipo de medios o puede haber un representador separado para cada tipo. Se representa información de audio a través del dispositivo de sonido en el ordenador del cliente, mientras que la información de vídeo y texto se representa en la pantalla de visualización del ordenador.

35 El agente es una pieza de software capaz de reproducir flujos, medir la disponibilidad y calidad de los flujos y notificar esta información. En una realización representativa, el agente es independiente del núcleo o el representador, aunque este no es un requisito de la invención. En un caso adecuado, la funcionalidad del agente puede ser nativa para el núcleo o para el representador. Un proveedor de reproductor de medios dado puede proporcionar de esta manera el reproductor de medios con la funcionalidad de agente integrada.

40 Un reproductor de medios dado incluye un agente y un conjunto de reproductores de medios con agentes embebidos puede comprender un sistema de monitorización de flujo como se ilustra en la Figura 3. En este ejemplo, el sistema de monitorización de flujo 300 incluye, sin limitación, una circuitería de entrada 302 donde los usuarios pueden establecer pruebas de flujo y empuja estas pruebas fuera a un conjunto de agentes 304a-n desplegados en diversas ubicaciones de red y geográficas; y una circuitería de salida 306 que almacena los datos de rendimiento notificados por los agentes y proporciona herramientas de análisis 308 y una interfaz de usuario (UI) conveniente 310 para ver los resultados de las pruebas. Se accede al uno o más ordenadores que comprenden la circuitería de entrada 302 y la circuitería de salida 306 preferiblemente a través de un canal seguro, aunque esto no es una limitación. Una forma conveniente de implementar estos subsistemas es a través de una aplicación de intranet o extranet, por ejemplo, usando un navegador Web sobre una conexión segura (SSL). En una realización particular, un proveedor de servicios usa la circuitería de entrada 302 para el suministro de las pruebas y el proveedor de servicios y sus clientes usan la circuitería de salida para ver o evaluar los resultados de las pruebas. Preferiblemente, los agentes en sí mismos programan las pruebas, realizan las pruebas y notifican de vuelta los resultados.

55 Como se ve en la Figura 4, un agente ilustrativo 400 tiene dos componentes primarios: una capa de agente 402 y un núcleo de agente 404, cada uno de los cuales se describe ahora a un nivel funcional. La capa de agente 402 como se ilustra es lógicamente distinta del núcleo de agente 404. El núcleo de agente incluye un módulo de monitorización de rendimiento (PMM) 408 que reside a través de la interfaz entre un núcleo 405 y un representador 406 y monitoriza el flujo de paquetes. En esta realización, el núcleo de agente 404 incluye el núcleo y el representador (que, típicamente, son elementos que se proporcionan por el proveedor de reproductor de medios de terceras partes), aunque esto no es un requisito. Alternativamente, el núcleo de agente puede comprender sólo el PMM y el núcleo. Generalmente, la capa de agente 402 proporciona funciones de control para el núcleo de agente (y, en particular, el PMM) e inicia una o más pruebas y la capa de agente 402 recibe los resultados de las pruebas. En una realización ilustrativa, la capa de agente incluye código que proporciona varias funciones, tales como determinar una configuración de pruebas que necesitan ser llevadas a cabo, programación de las pruebas y notificación de vuelta sobre la calidad de flujo, por ejemplo, a la circuitería de salida del sistema de monitorización de flujo. El paso de

lectura de la configuración de prueba se puede lograr leyendo un fichero de configuración presente en la máquina local o poniendo en cola un mecanismo de comunicación en tiempo real. En una realización representativa, el fichero de configuración se empuja desde el sistema de circuitería de entrada a la capa de agente o se extrae bajo demanda por la capa de agente.

- 5 Típicamente, una configuración de pruebas consta de varios conjuntos de pruebas, donde cada conjunto de pruebas se puede realizar independientemente y en paralelo. En una realización representativa, cada conjunto de pruebas tiene un *tiempo de inicio* cuando el agente necesita iniciar las pruebas del conjunto de pruebas, un *tiempo final* cuando el agente necesita parar las pruebas del conjunto de pruebas, una *prioridad* que proporciona información de programación sobre lo importante que es el conjunto de pruebas y una *frecuencia* que determina con qué frecuencia se repite la prueba del conjunto de pruebas. Preferiblemente, cada conjunto de pruebas es una serie de *pruebas* que se ejecutan en una secuencia una después de otra. Preferiblemente, cada prueba se especifica por el *URL* del flujo que necesita ser probado, el *tiempo de puesta en marcha máximo*, que es la cantidad máxima de tiempo que el agente esperará el flujo para iniciar la reproducción antes de declarar un fallo, el *tiempo de flujo máximo*, que es la cantidad de tiempo en el flujo en que el agente reproduce el flujo y el *tiempo de reproducción máximo*, que es la cantidad de tiempo de reloj real máxima que el agente intentará reproducir el flujo. Señalar la distinción sutil entre tiempo de flujo que se define interno al flujo que se está midiendo y tiempo de reproducción, que es el tiempo de reloj real; por ejemplo, el agente podría reproducir el flujo hasta que alcanza el punto de 3 minutos en el fragmento, es decir, el tiempo de flujo transcurrido es 3 minutos, pero podría tardar realmente 4 minutos de tiempo de reloj alcanzar ese punto debido a eventos de realmacenamiento temporal cuando la reproducción se congela, es decir, el tiempo de reproducción transcurrido es 4 minutos. Preferiblemente, el agente detiene la reproducción del fragmento cuando se alcanza o bien el tiempo de flujo máximo o bien el tiempo de reproducción máximo, cualquiera que se alcance a la mayor brevedad.

Como se señaló anteriormente, la capa de agente 402 también realiza la programación de las pruebas. La capa programa las pruebas en los conjuntos de pruebas por las especificaciones y maneja las pruebas en la secuencia adecuada al núcleo de agente 404 para su prueba. El núcleo de agente 404 es responsable de dirigir las pruebas y notificar de vuelta las métricas de calidad de flujo medidas de vuelta a la capa de agente. En particular, la capa registra las métricas de calidad de flujo en un fichero de registro 410 que, por ejemplo, puede contener una línea de registro por prueba. En una realización, los registros se transportan al sistema de circuitería de salida, por ejemplo, a través de correo de registro. Alternativamente, los datos de calidad de flujo también se pueden notificar de vuelta mediante un mecanismo que es más en tiempo real que correo de registro, esto es, un mecanismo de consulta distribuido.

El núcleo de agente 404 coge en el *URL* lo que necesita ser probado, reproduce el flujo (por ejemplo, a través de al menos el núcleo), mide una o más métricas de rendimiento y notifica estas métricas de vuelta a la capa de agente 402. Como se ilustra en la Figura 4, el núcleo incorpora dentro de él (o está asociado de otro modo con) un reproductor de medios (por ejemplo, un reproductor de Windows Media, un reproductor de Real, un reproductor QuickTime) que se usa para reproducir los flujos. En particular, el reproductor de medios es típicamente una caja negra, que es una pieza propietaria de software. Por supuesto, es conocido que “mirar dentro” de una caja negra para extraer información de rendimiento es una tarea desafiante. Según la presente invención, no obstante, el núcleo de agente 404 supera estas deficiencias implementando una interfaz – entre el núcleo y el representador – que se usa para monitorizar el rendimiento de flujo. Como se señaló anteriormente y como se muestra en la Figura 4, el núcleo de agente contiene un módulo de monitorización de rendimiento (PMM) 408 que reside dentro (o a través) de esta interfaz. Como se verá, la calidad del flujo que se representa depende de los paquetes útiles traspasados por el núcleo al representador. Como se usa en la presente memoria, los paquetes útiles se consideran que son aquellos paquetes que no se adelgazaron por el servidor, que no se perdieron de una forma irre recuperable ya que se transmitieron sobre la red desde el servidor al cliente y que llegaron a tiempo para ser útiles para la representación. Preferiblemente, el módulo de monitorización de rendimiento 408 intercepta todos o sustancialmente todos los paquetes útiles a medida que los paquetes van desde el núcleo al representador y es capaz de calcular métricas de calidad acerca de la reproducción.

Con independencia de la arquitectura del reproductor de medios, el módulo de monitorización de rendimiento (PMM) se debería diseñar para “comportarse” como un representador para el núcleo para interceptar eficazmente los paquetes útiles. En particular, el PMM se diseña para proporcionar la misma o sustancialmente la misma interfaz de programación de aplicaciones (API) al núcleo que al representador. De esta manera, cuando el núcleo piensa que está llamando al representador, está llamando realmente a un representador-contenedor 410 que es parte del PMM. El representador-contenedor 410 puede recibir entonces todos los paquetes desde el núcleo. Después de calcular las métricas de rendimiento relevantes usando los paquetes que recibe, el PMM entonces puede completar la transferencia de paquetes llamando al representador real. Desde la perspectiva del representador, el PMM juega el mismo papel que el núcleo.

Según una realización ilustrativa, el PMM calcula una o más métricas de rendimiento interceptando los paquetes útiles que fluyen dentro del reproductor de medios. Una métrica ilustrativa es el ancho de banda de reproducción real, que mide los bits/seg que llegan sin pérdida y a tiempo para su representación. Los reproductores de medios convencionales no proporcionan esta métrica. Como se muestra en la Figura 4, el módulo preferiblemente también

usa una o más funciones de devolución de llamada proporcionadas por el reproductor de medios para calcular métricas adicionales.

Una descripción de cómo el PMM calcula las diversas métricas de rendimiento sigue. En particular, la siguiente es una descripción genérica de la estrategia de cálculo de métrica de rendimiento junto con detalles específicos para implementar la invención con reproductores de medios representativos proporcionados por Real o WMS. Para calcular las métricas, el PMM puede basarse en un kit de desarrollo software (SDK) de reproductor de medios para el formato de difusión en forma continua particular. De esta manera, por ejemplo, una primera realización del agente (para uso con un reproductor de medios Real) se basa en el SDK de RealSystem. Una segunda realización del agente (para uso con un reproductor de medios WMS) usa el SDK de Windows Media Format. En la siguiente discusión, se supone familiaridad con las operaciones del reproductor Real y Windows Media, así como los SDK asociados.

Un marco para medir calidad de flujo

Lo siguiente describe una metodología de medición de calidad de flujo que se puede implementar usando la tecnología de agente. Como se señaló anteriormente, el agente proporciona una vista interior detallada del flujo debido a que preferiblemente se asienta entre el núcleo y el representador del reproductor de medios. Con esta vista, se pueden hacer varias preguntas de primer nivel acerca de cada reproducción que forma la base para las métricas de calidad de flujo: el usuario fue capaz de conectar con el servidor y reproducir el flujo, cuánto tuvo que esperar el usuario antes de que comenzase la reproducción de flujo y qué fracción de los medios fue entregada eficazmente desde el servidor al reproductor para ser utilizable y qué fracción fue adelgazada en el servidor, perdida en tránsito o llegó tarde para ser no utilizable, hubo cualquier interrupción, es decir, realmacenamientos temporales, durante la reproducción. Estas métricas de primer nivel, tomadas juntas (lo cual no es un requisito), proporcionan una imagen clara de cómo se reprodujo el flujo. Además, el uso del agente facilita la medición de varias métricas de segundo nivel o "auxiliares" que proporcionan información adicional acerca de la calidad de flujo. Estas se definirán más adelante.

Las métricas de primer nivel:

Las métricas de primer nivel pueden incluir las siguientes:

- *Tasa de fallos*: Cuando el usuario intenta reproducir un flujo, el flujo puede no ponerse en marcha por cualquier número de razones. El reproductor puede ser incapaz de alcanzar el servidor debido a que falla una búsqueda de nombre de DNS, el servidor está caído, el servidor está demasiado sobrecargado para responder o la conectividad de red entre el cliente y el servidor está caída. Incluso si el cliente es capaz de establecer conectividad con el servidor, el flujo puede no ponerse en marcha debido a que el servidor no puede acceder a los medios o el servidor es incapaz de obtener bastantes datos para el cliente de una forma oportuna para empezar la reproducción. *Capturas de tasa de fallos*, preferiblemente como un porcentaje, la fracción de veces que el usuario es incapaz de reproducir con éxito el flujo.
- *Tiempo de puesta en marcha*: Suponiendo que la reproducción de flujo no falla, el agente mide el *tiempo de puesta en marcha*, que es la cantidad de tiempo que el usuario tiene que esperar para que el flujo inicie la reproducción.
- *Adelgazamiento y Pérdida*: Después de que el flujo comienza a reproducirse, la degradación de rendimiento más común es que el cliente es incapaz de recibir los medios de una forma oportuna desde el servidor. Típicamente, hay tres modos en los cuales se puede perder información: (a) el servidor puede "adelgazar" la información fuera y nunca enviarla al cliente; (b) el servidor puede enviar datos, pero los datos se pierden en ruta y no se pueden recuperar; (c) el servidor puede enviar los datos, pero los datos podrían llegar demasiado tarde para ser usados en el reproductor. Una métrica que mide estas formas de degradación se conoce en la presente memoria como *ancho de banda de reproducción real*; que es una tasa a la que se reciben por el cliente los datos útiles, es decir, datos que llegan de una forma oportuna para ser usados en el reproductor. *Ancho de banda de reproducción ideal* es el ancho de banda de reproducción medido en el cliente cuando los medios se reproducen bajo condiciones ideales sin degradación. El ancho de banda de reproducción real se compara típicamente con el ancho de banda de reproducción ideal para comprender qué parte de los medios se adelgazó, perdió, tardó o de otro modo fue inutilizable por el cliente. Un valor de ancho de banda de reproducción real que es significativamente menor que el valor ideal indica distorsiones visibles en la reproducción, incluyendo presentaciones de diapositivas donde solamente se presentan en la pantalla un subconjunto de las tramas.
- *Interrupciones*: Una forma diferente de degradación ocurre cuando el reproductor es incapaz de continuar con la reproducción, por ejemplo, debido a que su almacenador temporal está vacío. En tal caso, el reproductor hace una pausa en la reproducción y espera que lleguen más datos. Dos métricas importantes para medir interrupciones son *realmacenamientos temporales por minuto*, que es el número medio de incidentes de realmacenamiento temporal en un minuto de reproducción y *tiempo de realmacenamiento*

*temporal por minuto*, que es el tiempo medio gastado realmacenando temporalmente en un minuto de reproducción.

Las métricas auxiliares

Además de las métricas de primer nivel, hay otras métricas de interés que proporcionan información más detallada sobre la calidad del flujo. Ellas incluyen una o más de las siguientes:

- 5
  - *Otras Métricas de Fallos:* Además de la tasa de fallos total, pueden ser de interés tasas de fallos más específicas que detallan dónde y cómo ocurrió el fallo. Las métricas específicas de interés son:
    - 10
      - *Tasa de Fallos de DNS:* falló en la búsqueda de DNS para localizar un servidor.
      - *Tasa de Fallos de Conexión:* falló en la etapa de conexión, por ejemplo, los medios solicitados no estaban presentes.
      - *Tasa de Fallos de Almacenador Temporal:* falló en la puesta en marcha debido a que el almacenamiento temporal inicial no se podría completar, por ejemplo, los datos que no fluyen a una tasa suficiente desde el servidor al cliente.
      - 15
        - *Tasa de Desconexión de Flujo:* falló en el medio de la reproducción, por ejemplo, conectividad perdida con el servidor antes de que pudiera completarse la reproducción.
    - *Componentes de Tiempo de Puesta en Marcha:* Las tres métricas de más adelante se añaden a la métrica de Tiempo de Puesta en Marcha.
      - 20
        - *Tiempo de Búsqueda de DNS:* El tiempo que tarda el cliente en buscar el nombre de dominio en el URL y obtener una dirección IP del servidor de medios que sirve el flujo.
        - *Tiempo de Conexión:* El tiempo que tarda el cliente en completar el inicio de diálogo del protocolo con el servidor y obtener una descripción de los medios.
        - *Tiempo de Almacenamiento Temporal Inicial:* El tiempo que tarda el cliente en recibir suficientes datos desde el servidor para iniciar la reproducción.
      - 25
        - *Métricas de Paquetes:* Mientras que las métricas de ancho de banda, tales como ancho de banda de reproducción, se centran en la tasa a la que se entregan los octetos al cliente, las métricas de paquetes se centran en la tasa a la que se entregan los paquetes al cliente. Las métricas de ancho de banda son generalmente más robustas, debido a que reflejan el contenido de información entregado al cliente con más precisión. Además, los paquetes pueden ser de tamaño variable conteniendo cantidades variables de información, lo que hace el recuento de paquetes menos robusto. Los recuentos de paquetes también pueden variar de reproducción en reproducción para los mismos medios por razones distintas de la calidad, dependiendo de cómo se hace la paquetización en el servidor. Sin embargo, las métricas de paquetes típicamente son de algún interés como una medida de calidad.
          - 30
            - *Tasa de Paquetes Útiles:* Esta es la tasa a la que se entregan paquetes útiles al cliente. Se debería señalar que un paquete típicamente es útil solamente si el paquete llegó al cliente de una forma oportuna, sin ser adelgazado por el servidor o perdido de una forma irrecuperable en el tránsito al cliente o llegó demasiado tarde para ser usado por el reproductor.
            - 35
              - *Tasa de Paquetes Normales:* De los paquetes que se reciben por el núcleo del cliente, algunos paquetes llegan sin que el cliente tenga que volver a solicitarlos desde el servidor. Estos paquetes constituyen la tasa de paquetes normales.
              - 40
                - *Tasa de Paquetes Recuperados:* De los paquetes que se entregan al cliente, algunos de los paquetes se pierden o llegan fuera de orden y se recuperan por el cliente. Estos paquetes constituyen la tasa de paquetes recuperados.
                - *Tasa de Paquetes Irrecuperables:* Estos paquetes se enviaron por el servidor pero se perdieron y nunca podrían ser recuperados por el cliente.
                - 45
                  - Un problema fundamental con las tres métricas de paquetes precedentes es que no capturan lo que nunca fue enviado por el servidor o aquellos paquetes que llegan demasiado tarde para ser útiles. También, estas métricas se miden a nivel de la aplicación de cliente. Puede haber calidad de flujo mala sin embargo no haber paquetes irrecuperables (es decir, perdidos) que se perciben por el cliente. Este es típicamente el caso cuando se difunden en forma continua datos sobre TCP, debido a que TCP proporciona una abstracción (de no pérdida) al cliente de difusión en forma continua que se asienta en la parte superior de él. Las métricas son más significativos cuando se difunde en forma continua sobre UDP, donde el cliente realmente ve la pérdida que ocurre en la
                  - 50

red. Sin embargo, la noción de ancho de banda de reproducción que se describió anteriormente es una medida mucho más robusta y significativa de adelgazamiento y pérdida de información.

- *Otras Métricas de Ancho de Banda:* Además de los anchos de banda de reproducción real e ideal, también pueden ser de interés otros tres anchos de banda, que se describen más adelante.

5           ○ *Ancho de Banda Codificado Máximo y Mínimo:* Para medios de tasa de bit única, hay solamente una codificación y los anchos de banda codificados máximo y mínimo ambos son iguales a la tasa de bit codificada de los medios. En el caso de medios de tasa de bit múltiple, hay múltiples codificaciones intercaladas, cada codificación a una tasa de bit diferente. En este caso, cuando la red se degrada, el servidor puede bajar desde una codificación de tasa de bit más alta a una codificación de tasa de bit más baja, adelgazando eficazmente los medios. Esta bajada se captura registrando tanto el ancho de banda máximo como mínimo de las codificaciones usadas en la reproducción.

10           ○ *Ancho de Banda de Red:* Este es el ancho de banda de los datos recibidos por el reproductor desde el servidor. Señalar que mientras que el ancho de banda de red proporciona alguna idea del rendimiento de flujo, no es una medida precisa. El ancho de banda de red cuenta los paquetes duplicados y retransmitidos multiplicados, aunque no contribuyen a un rendimiento de difusión en forma continua mayor y, de hecho, son signos de degradación de la red. Una situación típica es cuando los paquetes llegan fuera de orden, el cliente pide una retransmisión y recibe dos copias del mismo paquete, aumentando por ello el ancho de banda de red sin aumento en la calidad de flujo. Además, el ancho de banda de red no distingue entre octetos que son útiles y octetos que no lo son. Una situación típica es el servidor que envía octetos al cliente demasiado tarde para ser útiles en la reproducción; estos octetos se incluyen en el ancho de banda de red, pero no contribuyen a un aumento de la calidad de flujo. Finalmente, el servidor irrumpe en una parte inicial de los datos al cliente antes de enviar un flujo estable a la tasa de bit adecuada para el flujo. La ráfaga inicial aumenta el ancho de banda de red inicialmente por encima de la tasa codificada, lo que puede ocultar cualquier deterioro en el ancho de banda y caer en la calidad de flujo más tarde en la prueba. Estos problemas se evitan midiendo el ancho de banda de reproducción real, que indica qué ancho de banda se usó en la representación. Aún, el ancho de banda de red proporciona información general y vale la pena conservarlo como una métrica auxiliar.

30 **Cómo medir las Métricas de primer nivel**

- *Tasa de Fallos:* El PMM usa preferiblemente funciones de devolución de llamada para registrar la etapa en la que ocurrió el fallo, es decir, durante una búsqueda de DNS, durante la fase de conexión, durante la fase de almacenamiento temporal inicial o durante la reproducción. Preferiblemente, se produce un código de error que corresponde a los diferentes modos de fallo. Además, se registra cualquier código de error producido por el reproductor de medios en sí mismo. Las combinaciones de código de error se analizan entonces para producir la *tasa de fallo*.

35           ○ *Detalles para Real:* Cuando ocurre un error, se notifican dos códigos de error, "error" y "real\_error". El campo "error" especifica o bien el tipo de error particular, en el caso de un error "síncrono" en el código de agente, tal como un fallo de búsqueda de DNS o tiempo límite o bien que un error "asíncrono" ocurrió en el código del SDK de RealSystem. Para errores síncronos, se ignora el campo "real\_error". Los fallos notificados por el SDK de RealSystem se capturan implementando el método ErrorOccurred() de la interfaz IRMAErrorSink; el código de error pasado a este método se almacena en el campo "real\_error".

40           El campo "error" describe la etapa en la que ocurrió el fallo. Una vez que se ha abierto la conexión Real, se determina la etapa registrando una interfaz IRMAClientAdviceSink e implementando las devoluciones de llamada OnBuffering(), OnPosLength() y OnStop(). Específicamente, la fase de contacto comienza tan pronto como se abre la conexión; la fase de almacenamiento temporal comienza cuando se invoca la devolución de llamada OnBuffering(); y la fase de reproducción comienza cuando se invoca la devolución de llamada OnPosLength() con un argumento uiPosition positivo.

45           ○ *Detalles para Windows Media:* Cuando ocurre un error, se notifican dos códigos de error por el agente, "error" y "wms\_error". El campo "error" especifica o bien el tipo de error particular, en el caso de un error síncrono en el código de agente tal como un fallo de búsqueda de DNS o tiempo límite o bien que un error "asíncrono" ocurrió en el código del SDK de Windows Media. Para errores en el código de agente, se ignora el campo "wms\_error". Los fallos notificados por el SDK de Windows Media se capturan manejando el valor de estado WMT\_ERROR en el método de OnStatus() de la interfaz IWMMStatusCallback; el código de error pasado a este método se almacena en el campo "wms\_error".

El campo "error" también describe la etapa en la que ocurrió el fallo. Una vez que se ha abierto la conexión WMS, se determina la etapa registrando una interfaz IWMStatusCallback e implementando la devolución de llamada OnStatus(). Específicamente, la fase de contacto comienza en un evento WMT\_CONNECTING; la fase de almacenamiento temporal comienza en un evento WMT\_BUFFERING\_START; y la fase de reproducción comienza en un evento WMT\_STARTED.

- 5
  - 10
  - 15
  - 20
  - 25
  - 30
  - 35
  - 40
  - 45
  - 50
  - 55
- *Tiempo de Puesta en Marcha.* El monitor de rendimiento inicia un reloj al comienzo de la prueba y consulta el estado del reproductor de medios, preferiblemente usando funciones de devolución de llamada de reproductor. Cuando el reproductor de medios transita al estado de reproducción, el tiempo se señala como el tiempo de puesta en marcha.
  - *Detalles para Real.* El tiempo gastado en cada una de las fases de contacto y almacenamiento temporal se mide usando las devoluciones de llamada IRMAClientAdviceSink como se describió anteriormente. El tiempo de conexión es el tiempo entre llamadas a OnContacting() y OnBuffering() y el tiempo de almacenamiento temporal inicial es el tiempo entre llamadas a OnBuffering() y OnPosLength(). El tiempo de puesta en marcha es el tiempo desde el comienzo de la prueba (incluyendo el tiempo de búsqueda de DNS) hasta la primera llamada de OnPosLength().
  - *Detalles para WMS.* El tiempo gastado en cada una de las fases de contacto y almacenamiento temporal se mide usando la interfaz IWMStatusCallback como se describió anteriormente. El tiempo de conexión es el tiempo entre los eventos WMT\_CONNECTING y WMT\_BUFFERING\_START y el tiempo de almacenamiento temporal inicial es el tiempo entre los eventos WMT\_BUFFERING\_START y WMT\_STARTED. El tiempo de puesta en marcha es el tiempo desde el comienzo de la prueba (incluyendo el tiempo de búsqueda de DNS) hasta la primera llamada de WMT\_STARTED.
  - *Adelgazamiento y Pérdida:* Como se señaló anteriormente, el PMM preferiblemente calcula el *ancho de banda de reproducción real* interceptando y examinando los paquetes útiles que se envían desde el núcleo al representador. Se calcula un total de octetos útiles entregados B añadiendo los octetos de los paquetes útiles. Se calcula un tiempo de flujo T usando una función de devolución de llamada para determinar la cantidad del fragmento que se reprodujo con éxito al final de la prueba. El ancho de banda de reproducción real se calcula como B/T. El *ancho de banda de reproducción ideal* es el ancho de banda de reproducción real cuando los medios se difunden en forma continua bajo condiciones ideales sin degradación. Esta cantidad se puede calcular de una de dos maneras. Para flujos en directo, el ancho de banda de reproducción ideal se puede calcular como una función del ancho de banda codificado de los medios con algunos ajustes al hecho de que el ancho de banda de reproducción ideal de un flujo en directo en curso mostrará algunas variaciones con el tiempo. Para flujos bajo demanda, el ancho de banda de reproducción ideal se puede calcular con un 100% de precisión calibrando los medios, es decir, midiendo el ancho de banda de reproducción real cuando los medios se difunden de forma continua bajo condiciones de red ideales sin degradación. En la práctica, este paso de calibración no es completamente necesario; usar el valor observado más grande del ancho de banda de reproducción real sobre varias docenas de pruebas funciona bien. Comparar el ancho de banda de reproducción real con el ancho de banda de reproducción ideal proporciona una buena noción de cuánto adelgazamiento y pérdida ocurrió.
  - *Detalles para Real.* La manera en que el PMM intercepta paquetes útiles que fluyen desde el núcleo al representador es como sigue. Para cada complemento de representador incluido con la aplicación RealPlayer, se proporciona un representador-contenedor que contiene el nombre del complemento original. Cuando se carga el representador-contenedor, se abre el complemento original como una biblioteca dinámica y reenvía las llamadas del método entrantes al representador original. Varios métodos tienen un código adicional para calcular las métricas. Los métodos pertinentes a la medición de ancho de banda de reproducción son OnPacket(), que es llamado cuando el núcleo de RealPlayer envía un paquete de datos al representador y OnTimeSync(), que pide al representador actualizar su tiempo actual.
- La forma en que se cuentan los paquetes difiere ligeramente para flujos en directo y bajo demanda, pero el planteamiento básico es preferiblemente el mismo, es decir, contar los paquetes usados por el representador para un segmento de reproducción dado. Para flujos bajo demanda, el segmento de reproducción se conoce por adelantado (tiempo 0 a tiempo MAX\_STREAM\_TIME), de manera que los paquetes con un sello de tiempo menor que MAX\_STREAM\_TIME se cuentan como útiles. Para flujos en directo, el comienzo del segmento de reproducción no es el tiempo 0, de manera que este planteamiento no funcionará. En su lugar, los paquetes se ponen en cola a medida que se reciben y cuando el método OnTimeSync() es llamado con un tiempo t, todos los paquetes con tiempo menor que t se cuentan como útiles y se eliminan de la cola. (El tipo del flujo se determina por el método IsLive(), un método de la IRMAStreamSource del IRMAStream pasado a StartStream()).
- *Detalles para WMS.* El PMM no requiere representadores-contenedores para obtener los paquetes útiles para WMS. Los paquetes de datos comprimidos se obtienen implementando el método OnStreamSample()

de la interfaz IWMReaderCallbackAdvanced. Un paquete se incluye si su sello de tiempo es menor que MAX\_STREAM\_TIME.

- 5 • *Interrupciones.* El número de eventos de realmacenamiento temporal y la duración de cada evento de realmacenamiento temporal durante la reproducción se registra por el PMM poniendo en cola el estado del reproductor de medios usando funciones de devolución de llamada. Además, se registra el tiempo de reproducción del fragmento, que es el tiempo total que dura la reproducción. El tiempo de reproducción es diferente del tiempo de flujo cuando hay realmacenadores temporales. La métrica *realmacenamientos temporales por minuto* es simplemente el número de eventos de realmacenamiento temporal dividido por el tiempo de reproducción en minutos. El *tiempo de realmacenamiento temporal por minuto* es simplemente la duración total de todos los eventos de realmacenamiento temporal dividido por el tiempo de reproducción.
- 10 • *Detalles para Real.* El tiempo de realmacenamiento temporal se mide entre llamadas para las devoluciones de llamada OnBuffering() y OnPosLength() de la interfaz IRMAClientAdviceSink. El tiempo de flujo se mide por el último tiempo pasado para la devolución de llamada OnPosLength() y el tiempo de reproducción es medido como el tiempo de reloj real desde la primera llamada OnPosLength().
- 15 • *Detalles para WMS.* El tiempo de realmacenamiento temporal se mide entre los eventos WMT\_BUFFERING\_START y WMT\_BUFFERING\_STOP en el método IWMStatusCallback: : OnStatus(). El tiempo de flujo se obtiene a partir del método IWMReaderStreamClock: : GetTime() y el tiempo de reproducción se mide como el tiempo de reloj real desde el primer evento WMT\_STARTED.

#### Cómo medir las Métricas Auxiliares

- 20 • *Otras Tasas de Fallos:* La tasa de fallos de DNS, tasa de fallos de conexión, tasa de fallos de almacenador temporal y tasa de desconexión de flujo todas pueden ser códigos de error calculados sacados por el agente que determina tanto la causa del fallo como el estado de la reproducción cuando ocurrió el fallo.
- 25 • *Componentes del tiempo de Puesta en Marcha:* Para medir el tiempo de búsqueda de DNS, el agente realiza la búsqueda de DNS del nombre de servidor en sí mismo usando gethostbyname() y entonces pasa la dirección IP resultante al núcleo del reproductor como parte del URL. Esto también permite al agente notificar la dirección IP exacta del servidor desde la cual se reproduce el flujo de prueba. Los otros componentes del tiempo de puesta en marcha se miden usando funciones de devolución de llamada como se perfiló anteriormente.
- 30 • *Métricas de Paquetes.* En su mayor parte, estas métricas se obtienen usando una devolución de llamada para las estadísticas del reproductor. La métrica de paquetes útiles es una excepción, que se calcula directamente por el PMM interceptando paquetes útiles.
- *Otras Métricas de Ancho de Banda.* Estas métricas también se calculan usando devoluciones de llamada para las estadísticas del reproductor.

35 Como se señaló anteriormente, el agente de difusión en forma continua se puede implementar como un dispositivo autónomo o se puede implementar un conjunto de tales agentes, preferiblemente de una manera distribuida, para permitir a un proveedor de contenidos o de otros servicios determinar la calidad y rendimiento de flujo desde una pluralidad de ubicaciones diferentes. En tal caso, un agente se implementa en cada ubicación tal y los agentes distribuidos se controlan para cada petición de un flujo dado. A medida que los flujos están siendo entregados desde uno o más servidores de medios a los reproductores de medios que incluyen el software de agente embebido, se recaban los datos. Periódicamente, los datos se transfieren desde los agentes a una ubicación de gestión central, donde se procesan en informes o se ponen a disposición de otra manera para su visión (por ejemplo, a través de un portal seguro), para descarga o para otro análisis. En una realización particular, los agentes comprenden parte de una red de entrega de contenidos (CDN) y se usan para permitir a un proveedor de servicios de CDN (o clientes CDN) determinar la calidad y rendimiento de flujo. Como es bien conocido, una CDN es una colección de servidores de contenidos y mecanismos de control asociados que descargan trabajo de servidores de origen de sitio Web entregando contenido en su nombre a usuarios finales. Una CDN bien gestionada logra esta meta sirviendo medios de difusión en forma continua o algunos o todos los contenidos de páginas Web de un sitio, reduciendo por ello los costes del cliente al tiempo que mejora la experiencia de navegación de un usuario final del sitio. En funcionamiento, la CDN típicamente usa un mecanismo de encaminamiento de petición para localizar un servidor de contenidos de CDN cerca del cliente para servir cada petición dirigida a la CDN, donde la noción de “cerca” se basa, en parte, en la evaluación de los resultados de las pruebas de tráfico de red. Una CDN se puede usar para proporcionar difusión en forma continua en directo o bajo demanda.

Habiendo descrito de esta manera nuestra invención, se expone a continuación lo que ahora reivindicamos.

**REIVINDICACIONES**

1. Un aparato (400), que comprende:

un núcleo (405);

un representador (406); y

5        caracterizado por un primer módulo software situado intermedio del núcleo y del representador, el primer módulo software operable para generar una primera métrica a medida que un flujo de paquetes de datos útiles se pasan entre el núcleo (405) y el representador (406), en donde la primera métrica es una estadística calculada a partir de los paquetes de datos útiles; y

10        en donde cada paquete de datos útil tiene un conjunto de propiedades asociadas con el mismo, en donde el conjunto de propiedades comprende: que el paquete de datos no se ha adelgazado por un servidor, que el paquete de datos no se perdió de una manera irrecuperable durante el transporte y que el paquete de datos llegó al representador a tiempo para ser útil para su representación.

15        2. El aparato que se describe en la Reivindicación 1 que además incluye un segundo módulo software para recibir un conjunto de pruebas que identifican al menos una prueba de calidad de flujo, para programar el conjunto de pruebas y para sacar la primera métrica capturada ejecutando el primer conjunto de código.

3. El aparato que se describe en la Reivindicación 1 en donde se ejecuta el primer módulo software operable para generar una segunda métrica usando una función de devolución de llamada de reproductor.

4. Un sistema de monitorización de flujo (300), que comprende:

20        uno o más ordenadores de monitorización de flujo que incluyen un primer subsistema (302) para suministrar una prueba de calidad de flujo dada; un conjunto de agentes de monitorización (304) colocado en ubicaciones de red dadas para ejecutar la prueba de calidad de flujo dada, cada uno de los agentes de monitorización asociados con un reproductor de medios de difusión en forma continua que tiene un núcleo (405), un representador (406) y caracterizado por cada agente (304) que comprende:

25        un primer conjunto de código ejecutado en asociación con el reproductor de medios de difusión en forma continua para interceptar cada paquete de datos útil de un flujo dado a medida que el paquete de datos útil sale del núcleo (405) y antes de que el paquete de datos útil entre en el representador (406);

30        en donde cada paquete de datos útil tiene un conjunto de propiedades asociadas con el mismo, en donde el conjunto de propiedades comprende: que el paquete de datos no se ha adelgazado por un servidor, que el paquete de datos no se perdió de una manera irrecuperable durante el transporte y que el paquete de datos llegó al representador (406) a tiempo para ser útil para su representación

un segundo conjunto de código para capturar una métrica de calidad de flujo dada a partir de los paquetes de datos interceptados;

35        un tercer conjunto de código para sacar datos asociados con la métrica de calidad de flujo dada; y en donde el uno o más ordenadores de monitorización de flujo que incluyen también un segundo subsistema para recabar los datos a partir del conjunto de agentes de monitorización.

5. Un método de monitorización de flujo que comprende:

entregar un flujo de medios a cada uno de un conjunto de agentes (304), cada agente (304) que tiene un núcleo (405) y un representador (406) el método caracterizado por;

40        en cada agente (304), interceptar paquetes de datos útiles del flujo a medida que esos paquetes de datos atraviesan una interfaz entre el núcleo y el representador;

en donde cada paquete de datos útil tiene un conjunto de propiedades asociadas con el mismo, en donde el conjunto de propiedades comprende: que el paquete de datos no se ha adelgazado por un servidor, que el paquete de datos no se perdió de una manera irrecuperable durante el transporte y que el paquete de datos llegó al representador (406) a tiempo para ser útil para su representación

45        capturar una métrica de calidad de flujo dada a partir de los paquetes de datos útiles interceptados;

sacar datos asociados con la métrica de calidad de flujo dada;

y recabar datos de cada uno de los agentes (304) en un conjunto de uno o más ordenadores de monitorización de flujo.

6. El método de monitorización de flujo que se describe en la Reivindicación 5 que además incluye el paso de: permitir una vista de los datos recabados.

