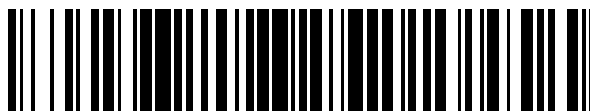


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 660 219**

51 Int. Cl.:

H04L 29/08 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **18.07.2014 E 14177582 (5)**

97 Fecha y número de publicación de la concesión europea: **06.12.2017 EP 2827561**

54 Título: **Reducción de potencia adaptativa controlada por servidor para protección de sobrecarga usando contadores de errores internos**

30 Prioridad:

18.07.2013 US 201361847876 P

16.07.2014 US 201414333038

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

21.03.2018

73 Titular/es:

**SYNCHRONOSS TECHNOLOGIES, INC. (100.0%)
200 Crossing Boulevard
Bridgewater, NJ 08807, US**

72 Inventor/es:

BYRNE, EOIN

74 Agente/Representante:

VALLEJO LÓPEZ, Juan Pedro

ES 2 660 219 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

Reducción de potencia adaptativa controlada por servidor para protección de sobrecarga usando contadores de errores internos

5

Solicitudes relacionadas

Esta solicitud reivindica el beneficio de prioridad bajo 35 U.S.C. sección 119(e) de la Solicitud de Patente Provisional de Estados Unidos con N.º de serie 61/847.876, en trámite con la presente, presentada el 18 de julio de 2013, titulada "Reducción de potencia adaptativa controlada por servidor para protección de sobrecarga usando contadores de errores internos".

10

Campo de la invención

La presente invención se refiere a una protección de sobrecarga de servidor. Más particularmente, la presente invención se refiere a una reducción de potencia adaptativa controlada por servidor para protección de sobrecarga usando contadores de errores internos.

15

Antecedentes de la invención

20

Durante periodos ocupados, tales como actualizaciones de producto o periodos de tráfico elevado activados por eventos externos, los servidores de HTTP pueden verse sometidos a cargas que tienen mucho acceso de sus cargas de operación pretendidas. Las soluciones existentes para sobrecargas esperan que los clientes resuelvan el problema controlando con qué frecuencia los clientes reintentan. Sin embargo, estas soluciones existentes se basan en comportamiento de manejo de errores complejo creado en cada implementación de los clientes y fuerzan decisiones de cuándo debería intentarse un reintento por un respectivo cliente.

25

Por ejemplo, un producto de respaldo, que permite el respaldo de imágenes de teléfonos móviles, ha existido en el mercado durante algún tiempo y tiene una gran base de instalación de millones de usuarios. Si ese producto tuviera que actualizarse para soportar también el respaldo de otros ficheros, tales como ficheros de audio y vídeos, entonces existe un peligro real de una tormenta de tráfico, donde la actualización se envía a millones de usuarios a través de un periodo de tiempo muy corto. Cada usuario iniciaría respaldos que contienen todos los ficheros de audio y de vídeo existentes, lo que representa meses, o incluso años, de carga de usuario normal. Esto podría dar como resultado una sobrecarga en el servidor, que se le está solicitado que procese un retraso acumulado de un año de duración de trabajo para cada usuario a través de un periodo de tiempo muy corto. Otros escenarios implican grandes números de personas que reaccionan a un evento externo, tal como un desastre natural o similar, que pueden producir enormes picos de tráfico muy superiores a la carga normal que el servidor está dimensionado a manejar.

30

35

Una solución de la técnica anterior es conocida como reducción de potencia exponencial de cliente. En este caso, cada cliente recibe un error, espera una cantidad de tiempo preconfigurado y reintenta. Si esa solicitud encuentra un error, el cliente esperará una cantidad más larga de tiempo antes de volver a reintentar de nuevo. Esto continuará hasta que se realice una cantidad preconfigurada de intentos aumentando el tiempo entre cada intento exponencialmente. Este enfoque se basa en que los clientes se comportan de la manera correcta. El servidor se encuentra aún bajo carga significativa a medida que todos los clientes aumentan sus tiempos de reducción de potencia con fallos tempranos. Debido a que la reducción de potencia exponencial en cada cliente es la misma, el servidor podría verse afectado por múltiples ondas de intentos. Si los contactos iniciales de todos los clientes son a un tiempo similar, entonces todos los intentos posteriores serán a aproximadamente el mismo tiempo también. Esto aumentará la sobrecarga del servidor a medida que el servidor gasta su tiempo tratando errores en lugar de tratando solicitudes y haciendo un buen trabajo.

40

45

50

Otra solución de la técnica anterior es conocida como una reducción de potencia dictada por servidor. Este otro enfoque común es usar una característica de protocolo que posibilita que el servidor ordene a los clientes reintentar después de un tiempo constante pero configurable después de cada error. Esto generalmente se comporta peor que la reducción de potencia exponencial de cliente puesto que, sin el componente exponencial, el servidor se ve afectado constantemente por ondas de solicitudes a intervalos cortos fijados. El documento US 2005/102393 se refiere a técnicas para equilibrado de carga adaptativo de acuerdo con unas condiciones de operación del servidor.

55

El documento US 2012/030326 se refiere a un método de dar servicio a múltiples solicitudes desde el mismo cliente dependiendo del número de intentos de la solicitud realizada por el dispositivo de cliente.

60

Breve resumen de la invención

La invención está definida por un medio legible por ordenador no transitorio de acuerdo con la reivindicación 1 y un dispositivo de acuerdo con la reivindicación 12. Se exponen realizaciones adicionales en las reivindicaciones dependientes. Las realizaciones de la presente invención se refieren a reducción de potencia adaptativa controlada

65

5 por servidor para protección de sobrecarga. El servidor controla un periodo de reducción de potencia para cada solicitud, que indica un tiempo de reintento de cuándo debería reenviarse una solicitud al servidor. Este enfoque de reducción de potencia se basa en el servidor puesto que el servidor tiene información mucho más precisa disponible sobre la que tomar decisiones de reducción de potencia. El servidor cambia el tiempo de reintento basándose en cómo de ocupado está y su capacidad para manejar la carga actual y/o sus sistemas dependientes aguas abajo. Este enfoque de reducción de potencia aumenta la estabilidad del servidor durante una carga muy alta, tal como cuando un servicio se enciende en primer lugar y recibe niveles de tráfico promedio mucho más altos de los niveles de los clientes que se comportan bien, distribuyendo la carga durante un periodo de tiempo más largo. El servidor puede convertir un pico de tráfico en una carga constante, que es más fácil y más eficaz de manejar para el servidor.

10 En un aspecto, se proporciona un medio legible por ordenador no transitorio. El medio legible por ordenador no transitorio almacena instrucciones que, cuando se ejecutan por un dispositivo informático, provocan que el dispositivo informático realice un método. El método incluye alojar al menos un servicio, acoplar comunicativamente con un primer dispositivo de usuario final, recibir una solicitud desde el primer dispositivo de usuario final para el al menos un servicio, controlar un periodo de reducción de potencia del primer dispositivo de usuario final determinando un tiempo de reintento que es específico para la solicitud desde el primer dispositivo de usuario final, y retransmitir el tiempo de reintento al primer dispositivo de usuario final.

15 En algunas realizaciones, el tiempo de reintento está basado al menos en una función de una tasa de error interna, en la que la tasa de error interna se observa a través de un periodo de tiempo. En algunas realizaciones, la tasa de error interna está asociada con un número de solicitudes que se han rechazado en el periodo de tiempo. En algunas realizaciones, la tasa de error interna se observa en una base por servicio.

20 En algunas realizaciones, el tiempo de reintento está basado en una función de una tasa de error observada desde sistemas aguas abajo. En algunas realizaciones, el tiempo de reintento está basado en una función de un número de eventos aguas abajo pendientes.

25 En algunas realizaciones, el tiempo de reintento está basado en un acceso de prioridad asociado con un usuario del primer dispositivo de usuario final.

30 En algunas realizaciones, el método también incluye recibir, después de que haya pasado el tiempo de reintento, la solicitud para el al menos un servicio reenviado desde el primer dispositivo de usuario final. Si el servidor puede manejar la solicitud reenviada, entonces se procesa la solicitud reenviada. Si el servidor no puede manejar la solicitud reenviada, entonces se repite la etapa de control de un periodo de reducción de potencia y la etapa de retransmisión del tiempo de reintento.

35 En algunas realizaciones, el método también incluye recibir una solicitud para el al menos un servicio desde un segundo dispositivo de usuario final a sustancialmente el mismo tiempo que se recibe la solicitud para el al menos un servicio desde el primer dispositivo de usuario final, en el que un tiempo de reintento determinado para la solicitud desde el segundo dispositivo de usuario final es diferente del tiempo de reintento determinado para la solicitud desde el primer dispositivo de usuario final.

40 En algunas realizaciones, el método también incluye recibir una solicitud para el al menos un servicio desde un segundo dispositivo de usuario final después de recibir la solicitud para el al menos un servicio desde el primer dispositivo de usuario final, en el que un tiempo de reintento determinado para la solicitud desde el segundo dispositivo de usuario final es más corto que el tiempo de reintento determinado para la solicitud desde el primer dispositivo de usuario final.

45 En algunas realizaciones, el método también incluye recibir una solicitud para el al menos un servicio desde un segundo dispositivo de usuario final después de recibir la solicitud para el al menos un servicio desde el primer dispositivo de usuario final, en el que un tiempo de reintento determinado para la solicitud desde el segundo dispositivo de usuario final es más largo que el tiempo de reintento determinado para la solicitud desde el primer dispositivo de usuario final.

50 En otro aspecto, se proporciona un medio legible por ordenador no transitorio. El medio legible por ordenador no transitorio almacena instrucciones que, cuando se ejecutan por un dispositivo informático, provocan que el dispositivo informático realice un método. El método incluye recibir una pluralidad de solicitudes desde dispositivos de usuario finales que están acoplados comunicativamente con el dispositivo informático y, basándose en una función de una tasa de error interna, determinar un tiempo de reintento para un primer subconjunto de los dispositivos de usuario final.

55 En algunas realizaciones, la tasa interna se observa en una base por servicio.

60 En algunas realizaciones, el tiempo de reintento se ajusta a las sobrecargas y recuperaciones del dispositivo informático.

65

El método también incluye informar al primer subconjunto de los dispositivos de usuario final del tiempo de reintento, y procesar las correspondientes solicitudes desde un segundo subconjunto de los dispositivos de usuario final.

5 En algunas realizaciones, las correspondientes solicitudes desde el primer subconjunto de los dispositivos de usuario final y las correspondientes solicitudes desde el segundo subconjunto de los dispositivos de usuario final son para el mismo servicio.

10 En algunas realizaciones, las correspondientes solicitudes desde un tercer subconjunto de los dispositivos de usuario final son para un servicio que es diferente de un servicio que está solicitando el primer subconjunto de los dispositivos de usuario final, en el que el método incluye adicionalmente procesar las correspondientes solicitudes desde el tercer subconjunto de los dispositivos de usuario final antes de procesar las correspondientes solicitudes desde el primer subconjunto de los dispositivos de usuario final.

15 En algunas realizaciones, el método también incluye convertir un pico de tráfico en una carga constante.

20 En otro aspecto más, se proporciona un dispositivo informático. El dispositivo informático incluye una carga de sistema durante un pico de tráfico, una interfaz de red para acoplar comunicativamente con al menos un dispositivo de usuario final para recibir una solicitud, y un medio legible por ordenador no transitorio que almacena instrucciones. Las instrucciones implementan un contador que cuenta un número de errores que han tenido lugar en un periodo de tiempo, y un módulo de reducción de potencia adaptativa controlada por servidor que ajusta un tiempo de reintento basándose en una tasa de error a través del periodo de tiempo. El tiempo de reintento se retransmite normalmente a al menos un dispositivo de usuario final de manera que la carga de sistema se expande con el tiempo.

25 En algunas realizaciones, la interfaz de red recibe la solicitud reenviada desde el al menos un usuario final después de que haya pasado el tiempo de reintento.

30 En algunas realizaciones, el tiempo de reintento calculado en un primer punto en el tiempo es más largo que el tiempo de reintento calculado en un segundo punto en el tiempo posterior al primer punto en el tiempo. Como alternativa, el tiempo de reintento calculado en un primer punto en el tiempo es más corto que el tiempo de reintento calculado un segundo punto en tiempo posterior al primer punto en el tiempo.

35 En algunas realizaciones, la tasa de error se observa a través de todos los servicios alojados por el dispositivo informático. Como alternativa, la tasa de error se observa en una base por servicio.

En algunas realizaciones, el tiempo de reintento está basado en un acceso de prioridad asociado con un usuario del al menos un dispositivo de usuario final.

40 En algunas realizaciones, el módulo de reducción de potencia adaptativa controlada por servidor influencia sobre cómo se comportan los dispositivos de usuario final que están acoplados comunicativamente con el dispositivo informático, en el que cada influencia es diferente.

Breve descripción de los dibujos

45 Lo anterior será evidente a partir de la siguiente descripción más particular de las realizaciones de ejemplo de la invención, como se ilustra en los dibujos adjuntos en los que caracteres de referencia similares hacen referencia a las mismas partes a través de todas las diferentes vistas. Los dibujos no están necesariamente a escala, debe hacerse en su lugar énfasis que ilustran realizaciones de la presente invención.

50 La Figura 1 ilustra un sistema a modo de ejemplo de acuerdo con una realización de la presente invención.

La Figura 2 ilustra un diagrama de bloques de un dispositivo informático a modo de ejemplo de acuerdo con una realización de la presente invención.

La Figura 3 ilustra un método a modo de ejemplo de acuerdo con una realización de la presente invención.

55 La Figura 4 ilustra otro método a modo de ejemplo más de acuerdo con una realización de la presente invención.

Descripción detallada de la invención

60 En la siguiente descripción, se exponen numerosos detalles para fines de explicación. Sin embargo, un experto en la materia se dará cuenta de que la invención puede ponerse en práctica sin el uso de estos detalles específicos. Por lo tanto, la presente invención no se pretende que esté limitada a las realizaciones mostradas sino que esté acorde con el alcance más amplio coherente con los principios y características descritos en el presente documento.

65 Las realizaciones de la presente invención se refieren a reducción de potencia adaptativa controlada por servidor para protección de sobrecarga. El servidor controla un periodo de reducción de potencia para cada solicitud, que indica un tiempo de reintento de cuándo debería reenviarse una solicitud al servidor. Este enfoque de reducción de potencia se basa en el servidor puesto que el servidor tiene información mucho más precisa disponible sobre la que

tomar decisiones de reducción de potencia. El servidor cambia el tiempo de reintento basándose en cómo de ocupado se encuentra y su capacidad para manejar la carga actual y/o sus sistemas dependientes aguas abajo. Este enfoque de reducción de potencia aumenta la estabilidad de servidor durante una carga muy alta, tal como cuando un servicio se enciende en primer lugar y recibe niveles de tráfico promedio mucho más altos de los niveles de los clientes que se comportan bien, distribuyendo la carga durante un periodo de tiempo más largo. El servidor puede convertir un pico de tráfico en una carga constante, que es más fácil y más eficaz de manejar para el servidor.

La Figura 1 ilustra un sistema 100 a modo de ejemplo de acuerdo con una realización de la presente invención. El sistema 100 normalmente incluye una red 105, tal como Internet, y un servidor o servidores 110 que están comunicativamente acoplados con la red 105. El servidor 110 está configurado para proporcionar al menos un servicio a los usuarios. El servidor 110 puede ser un servidor de respaldo, un servidor de aplicaciones, un servidor web, un servidor de noticias o similares. El servidor puede estar comunicativamente acoplado con uno o más repositorios 115 para almacenar y/o recuperar datos. En algunas realizaciones, el uno o más repositorios 115 pueden almacenar información de abonado y restaurar datos de abonados del al menos un servicio. Otros tipos de datos pueden almacenarse en el uno o más repositorios 115.

El servidor 110 normalmente incluye un contador y un módulo de reducción de potencia adaptativa controlada por servidor, que puede implementarse en software, hardware o una combinación de los mismos. En resumen, el contador cuenta un número de errores que han tenido lugar en un periodo de tiempo, y el módulo de reducción de potencia adaptativa controlada por servidor ajusta un tiempo de reintento basándose en una función de una tasa de error interna a través de ese periodo de tiempo. La tasa de error normalmente está asociada con un número de solicitudes que se han rechazado por el servidor 110 en ese periodo de tiempo. La tasa de error interna puede observarse en una base por servicio. Como alternativa, la tasa de error interna puede observarse a través de todos los servicios alojados por o en el servidor 110. En algunas realizaciones, el servidor también ajusta el tiempo de reintento basándose en una función una tasa de error observada desde sistemas aguas abajo y/o basándose en una función de un número de eventos aguas abajo pendientes.

El sistema 100 también incluye al menos un dispositivo de usuario final 120. Cada dispositivo de usuario final 120 normalmente pertenece a o es usado por un usuario para solicitar el al menos un servicio alojado por o en el servidor 110. En algunas realizaciones, cada usuario tiene una cuenta que permite a un respectivo usuario suscribirse o acceder a al menos un servicio. En algunas realizaciones, la cuenta permite al abonado establecer sus preferencias, tales como frecuencia de respaldo y notificaciones. El abonado normalmente puede acceder a la cuenta mediante una página web o un programa cliente instalado en el dispositivo de usuario final 120.

Como se ha explicado en alguna otra parte, el módulo de reducción de potencia adaptativa controlada por servidor influencia sobre cómo se comportan los dispositivos de usuario final 120 con respecto a cuánto dura la reducción de potencia y cuándo reenviar las solicitudes al servidor 110. En algunas realizaciones, cada influencia es diferente para cada solicitud o para un grupo de solicitudes. En algunas realizaciones, el servidor 110 comunica con un dispositivo de usuario final 120 mediante el programa cliente instalado en el mismo. Cuando cada dispositivo de usuario final 120 recibe instrucciones (por ejemplo, tiempo de reintento) desde el servidor 110, el dispositivo de usuario final 120 normalmente cumple con las instrucciones.

Suponiendo que el servidor 110 no puede manejar o satisfacer una solicitud desde el dispositivo de usuario final 120, el dispositivo de usuario final 120 recibirá un mensaje de error y el servidor 110 determinará un tiempo de reintento para esa solicitud. La determinación es normalmente basándose en cómo de ocupado se encuentra el servidor 110 y su capacidad para mejorar la carga del sistema actual y/o sus sistemas dependientes aguas abajo. Como se ha explicado anteriormente, el servidor 110 establece el tiempo de reintento basándose en la función de la tasa de error interna. El servidor 110 aumenta automáticamente el tiempo de reintento cuando el servidor 110 está sobrecargado y reduce automáticamente el tiempo de reintento cuando se recupera. Esto permite un tiempo de reintento altamente adaptativo que se aumenta de manera natural cuando el servidor está ocupado, permitiendo que los picos grandes se distribuyan con el tiempo.

Por ejemplo, el tiempo de reintento calculado en un primer punto en el tiempo es más largo que el tiempo de reintento calculado en un segundo punto en el tiempo posterior al primer punto en el tiempo. Para otro ejemplo, el tiempo de reintento calculado en un primer punto en el tiempo es más corto que el tiempo de reintento calculado en un segundo punto en el tiempo posterior al primer punto en el tiempo.

En algunas realizaciones, el tiempo de reintento puede ser basándose en una función de una tasa de error observada desde sistemas aguas abajo y/o basándose en una función de un número de eventos aguas abajo pendientes. Por ejemplo, durante una carga de fichero al servidor 110, es posible que los datos estén llegando más rápido en el servidor 110 que lo que pueden escribirse en el repositorio 115. El servidor 110 puede interpretar cualquier error desde el repositorio 115 o tiempos de cola largos que se crean debido a que el repositorio 115 no puede ejecutar lo suficientemente rápido para el procesamiento de todas las solicitudes, y puede usar esta información para ajustar el tiempo de reintento para aliviar la presión. El servidor 110, por lo tanto, puede usar el módulo de reducción de potencia adaptativa para proteger otros servidores y/o servicios en el sistema 100.

En algunas realizaciones, el tiempo de reintento puede ser basándose en acceso de prioridad asociado con un dispositivo de usuario final o con un usuario del dispositivo de usuario final. Por ejemplo, el servidor 110 recibe una solicitud del Usuario A y una solicitud del Usuario B a sustancialmente el mismo tiempo o en la misma franja de tiempo. Al usuario A se le proporciona un tiempo de reintento más corto que al Usuario B puesto que el Usuario A tiene una prioridad superior que el usuario B. El acceso de prioridad puede ser basándose en un nivel de servicio de suscripción del usuario, un tipo de dispositivo de usuario final, o similares.

Después de que el servidor 110 determina el tiempo de reintento, el tiempo de reintento se retransmite al dispositivo de usuario final 120 desde el servidor 110. El tiempo de reintento puede comunicarse con el mensaje de error al dispositivo de usuario final 120. El dispositivo de usuario final 120 debe respetar o cumplir con lo que ha comunicado el servidor 110 (por ejemplo, instrucciones con respecto al periodo de reducción de potencia) y reintenta su solicitud después de que el tiempo de reintento finalice o en un periodo de gracia después de que el tiempo de reintento finalice. Si el servidor 110 puede manejar esta solicitud posterior, entonces el servidor 110 procesará la solicitud posterior. De otra manera, el servidor 110 determinará otro tiempo de reintento más e informará al dispositivo de usuario final 120 del nuevo tiempo de reintento puesto que el servidor 110 no puede manejar de nuevo esta solicitud posterior.

La Figura 2 ilustra un diagrama de bloques de un dispositivo informático 200 a modo de ejemplo de acuerdo con una realización de la presente invención. El dispositivo informático 200 puede usarse para obtener, almacenar en caché, almacenar, calcular, buscar, transferir, comunicar y/o visualizar información. El servidor 110 y/o el dispositivo de usuario final 120 de la Figura 1 pueden configurarse de manera similar como el dispositivo informático 200.

En general, una estructura de hardware adecuada para implementar el dispositivo informático 200 incluye una interfaz de red 202, una memoria 204, procesador o procesadores 206, dispositivo o dispositivos de E/S 208, un bus 210 y un dispositivo de almacenamiento 212. La elección del procesador 206 no es crítica siempre que se elija un procesador adecuado con suficiente velocidad. En algunas realizaciones, el dispositivo informático 200 incluye una pluralidad de procesadores 206. La memoria 204 puede ser cualquier memoria informática convencional conocida en la técnica. El dispositivo de almacenamiento 212 puede incluir un disco duro, CDROM, CDRW, DVD, DVDRW, tarjeta de memoria flash, RAM, ROM, EPROM, EEPROM o cualquier otro dispositivo de almacenamiento. El dispositivo informático 200 puede incluir una o más interfaces de red 202. Un ejemplo de una interfaz de red incluye una tarjeta de red conectada a una Ethernet u otro tipo de LAN. El dispositivo o dispositivos de E/S 208 pueden incluir uno o más de lo siguiente: teclado, ratón, monitor, visor, impresora, módem, pantalla táctil, interfaz de botón y otros dispositivos. La aplicación o aplicaciones de reducción de potencia adaptativa controlada por servidor 216 es probable que se almacenen en el dispositivo de almacenamiento 212 y en la memoria 204 y se procesen por el procesador 206. Más o menos componentes mostrados en la Figura 2 pueden incluirse en el dispositivo informático 200. En algunas realizaciones, se incluye el hardware de reducción de potencia adaptativa controlada por servidor 214. Aunque el dispositivo informático 200 en la Figura 2 incluye las aplicaciones 216 y el hardware 214 para implementar el enfoque de reducción de potencia adaptativa controlada por servidor, el enfoque de reducción de potencia adaptativa controlada por servidor puede implementarse en un dispositivo informático en hardware, firmware, software o cualquier combinación de los mismos. Por ejemplo, en algunas realizaciones, se programa el software de reducción de potencia adaptativa controlada por servidor 216 en una memoria y se ejecuta usando un procesador. En otro ejemplo, en algunas realizaciones, el hardware de reducción de potencia adaptativa controlada por servidor 214 es lógica de hardware programado que incluye puertas específicamente diseñadas para implementar el método.

En algunas realizaciones, la aplicación o aplicaciones de reducción de potencia adaptativa controlada por servidor 216 incluyen varias aplicaciones y/o módulo o módulos. En algunas realizaciones, los módulos incluyen uno o más sub-módulos también.

El dispositivo informático 200 puede ser un servidor o un dispositivo de usuario final. Dispositivos de usuario final a modo de ejemplo incluyen, pero sin limitación, una tableta, un teléfono móvil, un teléfono inteligente, un ordenador de sobremesa, un ordenador portátil, un ordenador portable, o cualquier dispositivo informático adecuado tal como dispositivos de fin especial, incluyendo decodificadores de salón y consolas de automóvil.

La Figura 3 ilustra un método 300 a modo de ejemplo de acuerdo con una realización de la presente invención. El método 300 se realiza normalmente por el servidor 110 de la Figura 1 cuando el servidor 110 o el repositorio 115 de la Figura 1 están sobrecargados. En una etapa 305, al menos un servicio se aloja por o en el servidor. Un servicio a modo de ejemplo es un servicio de respaldo o un servicio de noticias. En una etapa 310, un primer dispositivo de usuario final está acoplado comunicativamente entre los mismos y envía una solicitud para el al menos un servicio. En una etapa 315, la solicitud se recibe desde el primer dispositivo de usuario final para el al menos un servicio. En una etapa 320, se controla un periodo de reducción de potencia del primer dispositivo de usuario final. En particular, se determina un tiempo de reintento que es específico para la solicitud desde el primer dispositivo de usuario final. El tiempo de reintento está basado en un estado interno del servidor, tal como un porcentaje de utilización de servidor (por ejemplo, procesador, memoria, disco, red, solicitudes pendientes, etc.). Como alternativa o además de, el tiempo de reintento puede basarse una función de una tasa de error de sistemas aguas abajo y/o basándose en una función de un número de eventos aguas abajo pendientes. Como alternativa o además de, el tiempo de reintento

5 puede ser basándose en un acceso de prioridad asociado con el primer dispositivo de usuario final o un usuario del primer dispositivo de usuario final. Como alternativa o además de, el tiempo de reintento puede ser basándose en el tipo del servicio que se esté solicitando. En una etapa 325, el tiempo de reintento se retransmite al primer dispositivo de usuario final. Normalmente, el primer dispositivo de usuario final reduce la potencia de la duración del tiempo de reintento y vuelve a enviar la solicitud al final del tiempo de reintento.

10 El primer dispositivo de usuario final normalmente respeta la instrucción o instrucciones desde el servidor y vuelve a enviar la solicitud en el tiempo ordenado. Si el servidor puede manejar esta solicitud posterior, que se envía al final del periodo de reducción de potencia, a continuación el servidor procesará la solicitud posterior. De otra manera, se repiten las etapas 320 y 325. En otras palabras, el servidor controla el periodo de reducción de potencia del primer dispositivo de usuario final determinando un nuevo tiempo de reintento, y retransmite el nuevo tiempo de reintento al primer dispositivo de usuario final.

15 Una solicitud para el al menos un servicio desde un segundo dispositivo de usuario final puede recibirse a sustancialmente el mismo tiempo que se recibe la solicitud para el al menos un servicio desde el primer dispositivo de usuario final. En algunas realizaciones, un tiempo de reintento determinado para la solicitud desde el segundo usuario final es diferente del tiempo de reintento determinado para la solicitud desde el primer dispositivo de usuario final.

20 De manera similar, la solicitud para el al menos un servicio desde el segundo dispositivo de usuario final puede recibirse después de que se recibe la solicitud para el al menos un servicio desde el primer dispositivo de usuario final. En algunas realizaciones, un tiempo de reintento determinado para la solicitud desde el segundo dispositivo de usuario final es más corto que el tiempo de reintento determinado para la solicitud desde el primer dispositivo de usuario final. Como alternativa, el tiempo de reintento determinado para la solicitud desde el segundo dispositivo de usuario final es más largo que el tiempo de reintento determinado para la solicitud desde el primer dispositivo de usuario final.

30 La Figura 4 ilustra otro método 400 a modo de ejemplo más de acuerdo con una realización de la presente invención. El método 400 se realiza normalmente por el servidor 110 de la Figura 1. En una etapa 405, se recibe una pluralidad de solicitudes desde dispositivos de usuario final que están acoplados comunicativamente con el servidor. En una etapa 410, basándose en una función de una tasa de error interna, se determina un tiempo de reintento para un primer subconjunto de los dispositivos de usuario final. En una etapa 415, se informa al primer subconjunto de los dispositivos de usuario final del tiempo de reintento. En una etapa 420, se procesan las correspondientes solicitudes desde un segundo subconjunto de los dispositivos de usuario final. En algunas realizaciones, las correspondientes solicitudes desde el primer subconjunto de los dispositivos de usuario final y las correspondientes solicitudes desde el segundo subconjunto de los dispositivos de usuario final son para el mismo servicio. En algunas realizaciones, el segundo subconjunto de los dispositivos de usuario final tiene una prioridad superior que el primer subconjunto de los dispositivos de usuario final.

40 En algunas realizaciones, las correspondientes solicitudes desde un tercer subconjunto de los dispositivos de usuario final son para un servicio que es diferente del servicio del primer subconjunto de los dispositivos de usuario final que está solicitando. Puesto que la tasa de error interna se observa en una base por servicio, como en algunas realizaciones, las correspondientes solicitudes desde el tercer subconjunto de los dispositivos de usuario final pueden procesarse antes de procesar las correspondientes solicitudes desde el primer subconjunto de los dispositivos de usuario final.

50 En algunas realizaciones, el servicio proporcionado por el servidor 110 es un servicio de respaldo. Normalmente, una sesión de respaldo es transparente para el abonado del servicio de respaldo. El respaldo de datos del dispositivo de usuario final del abonado al servidor 110 es automático y tiene lugar en segundo plano. Suponiendo que el abonado recibe una notificación con respecto al estado del respaldo, tal como "Respaldo al 33 %." Pero, el servidor 110 a continuación se vuelve ocupado y el respaldo se paraliza. Sin embargo, el servicio de respaldo se reanuda después del periodo de reducción de potencia. La notificación del respaldo se actualiza tan pronto como el respaldo se reanuda. Debería entenderse que las notificaciones sobre dispositivos de usuario final son específicas de la aplicación y pueden incluir el tiempo de reintento, por ejemplo "El servicio actualmente no está disponible. Se reintentará en 10 minutos". El dispositivo de usuario final reenvía automáticamente la solicitud de servicio después de que se haya terminado el periodo de reducción de potencia.

60 Un experto en la materia se dará cuenta de que existen también otros usos y ventajas. Aunque se ha descrito la invención con referencia a numerosos detalles específicos, un experto en la materia reconocerá que la invención puede realizarse en otras formas específicas. Por lo tanto, un experto en la materia entenderá que la invención no se ha de limitar por los detalles ilustrativos anteriores, sino en su lugar se ha de definir por las reivindicaciones adjuntas.

REIVINDICACIONES

1. Un medio legible por ordenador no transitorio que almacena instrucciones que, cuando son ejecutadas por un dispositivo informático, provocan que el dispositivo informático realice un método, comprendiendo el método:
- 5 alojar al menos un servicio;
 acoplar comunicativamente a un primer dispositivo de usuario final;
 recibir una solicitud desde el primer dispositivo de usuario final para el al menos un servicio;
 controlar un periodo de reducción de potencia del primer dispositivo de usuario final determinando un tiempo de
10 reintento que es específico para la solicitud desde el primer dispositivo de usuario final, en donde el tiempo de
 reintento está basado en acceso de prioridad asociado a un usuario del primer dispositivo de usuario final; y
 retransmitir el tiempo de reintento al primer dispositivo de usuario final.
- 15 2. El medio legible por ordenador no transitorio de la reivindicación 1, en el que el tiempo de reintento está basado al
 menos en una función de una tasa de error interna del dispositivo informático, en donde la tasa de error interna se
 observa a través de un periodo de tiempo en donde la tasa de error interna del dispositivo informático está asociada
 a un número de solicitudes que han sido rechazadas dentro del periodo de tiempo.
- 20 3. El medio legible por ordenador no transitorio de la reivindicación 2, en el que la tasa de error interna del
 dispositivo informático se observa sobre una base por servicio.
4. El medio legible por ordenador no transitorio de la reivindicación 1, en el que el tiempo de reintento está basado
 en una función de un número de eventos aguas abajo pendientes.
- 25 5. El medio legible por ordenador no transitorio de la reivindicación 1, en donde el método incluye adicionalmente
 recibir, después de que haya pasado el tiempo de reintento, la solicitud para el al menos un servicio reenviado desde
 el primer dispositivo de usuario final.
- 30 6. El medio legible por ordenador no transitorio de la reivindicación 5, en donde el método incluye adicionalmente
 procesar la solicitud reenviada.
7. El medio legible por ordenador no transitorio de la reivindicación 5, en donde el método incluye adicionalmente
 repetir la etapa de control de un periodo de reducción de potencia y la etapa de retransmisión del tiempo de
35 reintento.
8. El medio legible por ordenador no transitorio de la reivindicación 1, en donde el método incluye adicionalmente
 recibir una solicitud para el al menos un servicio desde un segundo dispositivo de usuario final al mismo tiempo que
 se recibe la solicitud para el al menos un servicio desde el primer dispositivo de usuario final, en donde un tiempo de
40 reintento determinado para la solicitud desde el segundo dispositivo de usuario final es diferente del tiempo de
 reintento determinado para la solicitud desde el primer dispositivo de usuario final.
9. El medio legible por ordenador no transitorio de la reivindicación 1, en donde el método incluye adicionalmente
 recibir una solicitud para el al menos un servicio desde un segundo dispositivo de usuario final después de recibir la
 solicitud para el al menos un servicio desde el primer dispositivo de usuario final, en donde un tiempo de reintento
45 determinado para la solicitud desde el segundo dispositivo de usuario final es más corto que el tiempo de reintento
 determinado para la solicitud desde el primer dispositivo de usuario final.
10. El medio legible por ordenador no transitorio de la reivindicación 1, en donde el método incluye adicionalmente
 recibir una solicitud para el al menos un servicio desde un segundo dispositivo de usuario final después de recibir la
50 solicitud para el al menos un servicio desde el primer dispositivo de usuario final, en donde un tiempo de reintento
 determinado para la solicitud desde el segundo dispositivo de usuario final es más largo que el tiempo de reintento
 determinado para la solicitud desde el primer dispositivo de usuario final.
- 55 11. Un dispositivo informático que comprende:
 una interfaz de red configurada para acoplar comunicativamente a al menos un dispositivo de usuario final para
 recibir una solicitud; y
 un medio legible por ordenador no transitorio que almacena instrucciones para implementar:
- 60 un contador que cuenta un número de errores que han tenido lugar en un periodo de tiempo;
 en donde los errores son rechazos de solicitudes de dispositivo de usuario por el dispositivo informático; y
 un módulo de reducción de potencia adaptativa controlada por servidor que está configurado para ajustar un
 tiempo de reintento basándose en una tasa de error a través del periodo de tiempo y también basándose en
 un acceso de prioridad asociado a un usuario del al menos un dispositivo de usuario final, en donde el tiempo
65 de reintento se reenvía a al menos un dispositivo de usuario final.

12. El dispositivo informático de la reivindicación 11, en el que los tiempos de reintento consecutivamente retransmitidos son diferentes.

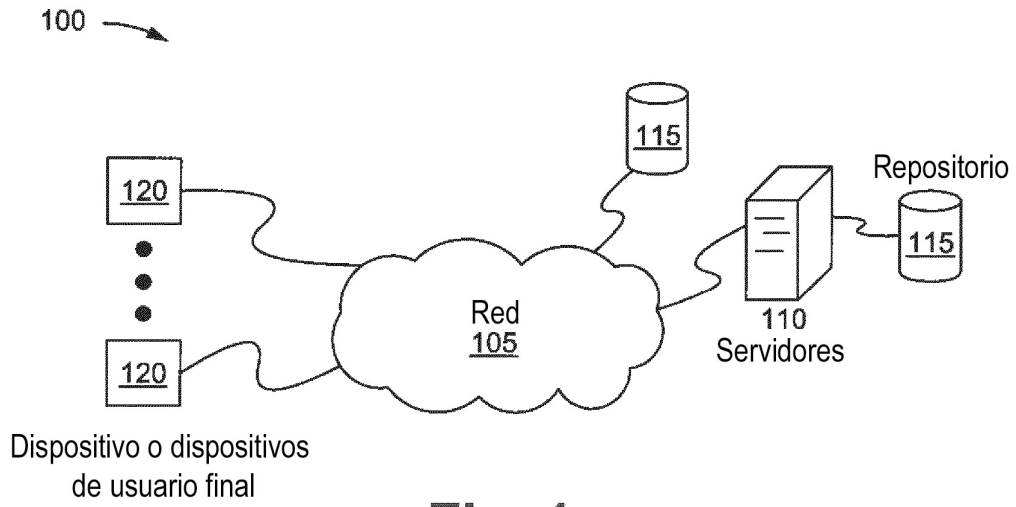


Fig. 1

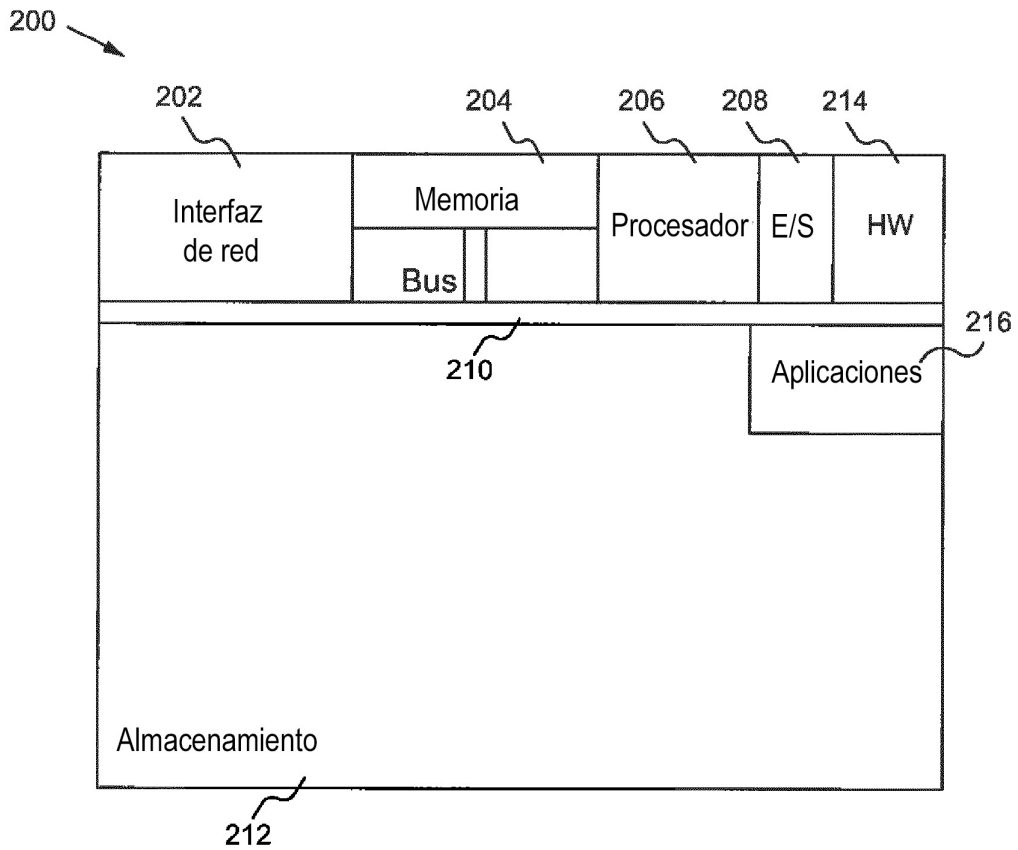


Fig. 2

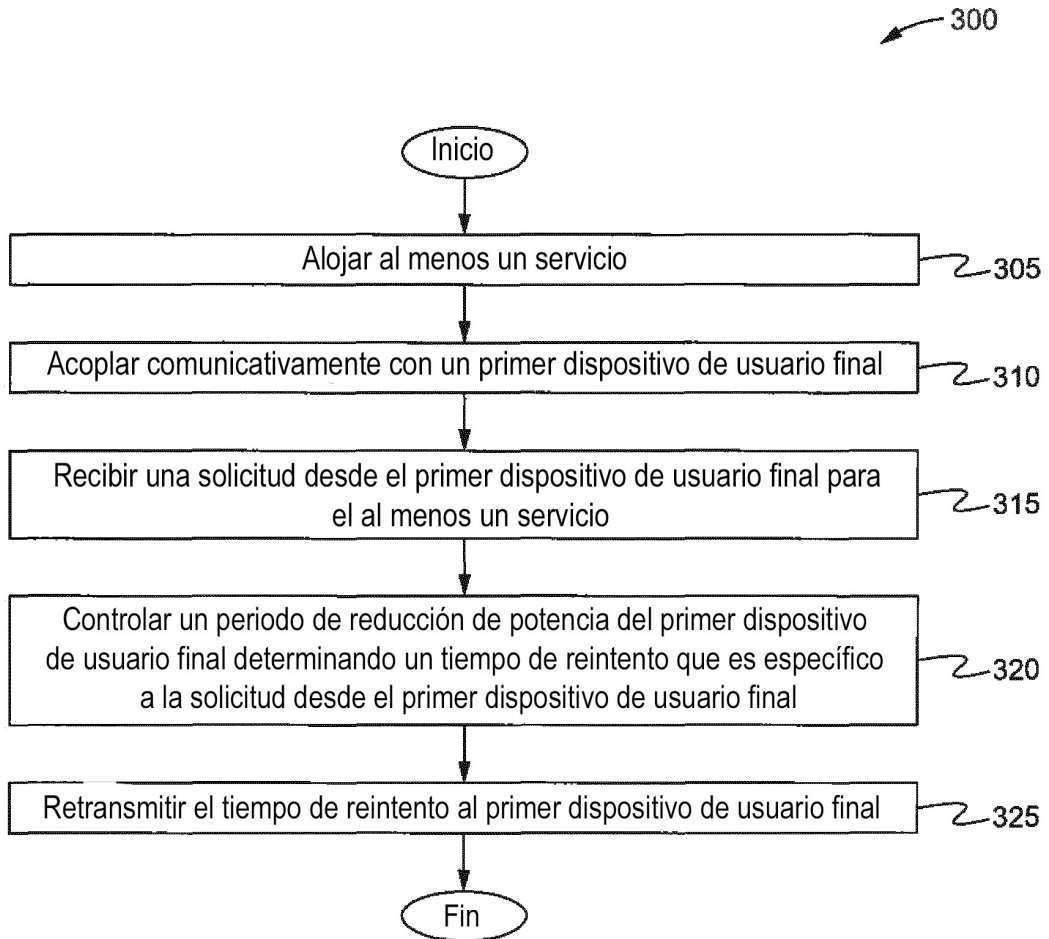


Fig. 3

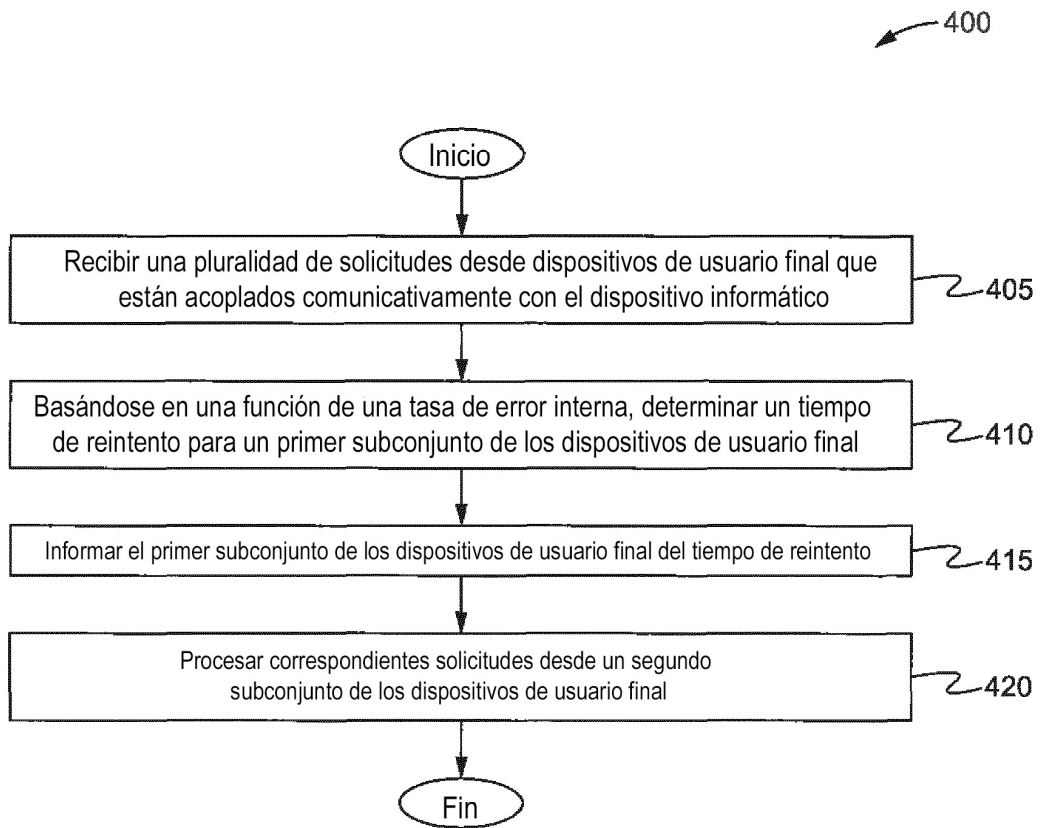


Fig. 4