

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 807 927**

51 Int. Cl.:

H04L 1/00 (2006.01)

H04L 12/801 (2013.01)

H04W 28/02 (2009.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **23.12.2016 PCT/US2016/068485**

87 Fecha y número de publicación internacional: **06.07.2017 WO17117040**

96 Fecha de presentación y número de la solicitud europea: **23.12.2016 E 16882458 (9)**

97 Fecha y número de publicación de la concesión europea: **01.07.2020 EP 3398276**

54 Título: **Control adaptativo de sobrecarga de pares en redes móviles**

30 Prioridad:

31.12.2015 US 201562273734 P

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

24.02.2021

73 Titular/es:

**MICROSOFT TECHNOLOGY LICENSING, LLC
(100.0%)
One Microsoft Way
Redmond, WA 98052, US**

72 Inventor/es:

**LOGAN, JAMES, J.;
LIBBY, MARK;
VELDANDA, NARSI;
PALAKKAL, RAJESH y
BHOJNAGARWALA, SIDDHARTH**

74 Agente/Representante:

SÁEZ MAESO, Ana

ES 2 807 927 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

Control adaptativo de sobrecarga de pares en redes móviles

5 Referencia cruzada a solicitudes relacionadas

Campo técnico

10 La presente descripción se refiere al control adaptativo de las tasas de transacción. Más específicamente, la presente descripción se refiere a sistemas y métodos de control adaptativo de sobrecarga de pares en redes móviles.

Antecedentes de la invención

15 En las redes móviles, un nodo puede sobrecargarse cuando un nodo par envía más transacciones de las que el nodo puede manejar. Tradicionalmente, para no colapsar la red, los operadores móviles generalmente configuran tasas de transacción máximas muy conservadoras/artificiales en cada nodo que interactúa con el servidor. La configuración manual puede plantear muchos desafíos, incluida la reconfiguración ante un cambio en la red.

20 El documento US2010/0274893 describe métodos y aparatos para detectar una sobrecarga limitada de servidores enfocados en una red.

Resumen

25 Un aspecto de los sistemas y métodos descritos en la presente descripción pueden incluir técnicas para ajustar adaptativamente una tasa de transacción en un nodo de red móvil. En algunos aspectos, los sistemas y métodos incluyen recibir, en un primer nodo cliente de la red móvil, parámetros de configuración que incluyen al menos uno de: una tasa de transacción inicial, la tasa de transacción inicial asociada con un número inicial de transacciones por segundo que se comunican por el primer nodo cliente de la red móvil a un nodo servidor a través de una red en un momento inicial, una
30 aumento en la tasa de transacción inicial, una tasa de reducción, la tasa de reducción asociada con un segundo número de transacciones por segundo indicativo de una disminución en la tasa de transacción inicial, una tasa mínima, una tasa máxima, un período de tiempo de evaluación, un umbral de tiempo de incremento de respuesta asociado con un tiempo de respuesta indicativo de aumento de la tasa de transacción inicial, un umbral de tiempo de reducción de respuesta asociado con un tiempo de respuesta indicativo de disminución de la tasa de transacción inicial, un porcentaje del umbral de incremento de respuestas asociado con el umbral de tiempo de incremento de respuesta, y un porcentaje del umbral de reducción de respuestas asociado con el umbral de tiempo de reducción de respuesta. En algunos aspectos, los sistemas y métodos incluyen recibir, en el primer nodo cliente de la red móvil, respuestas asociadas con las comunicaciones con el nodo servidor durante un primer período de tiempo asociado con el período de tiempo de evaluación. En algunos aspectos, los sistemas y métodos incluyen determinar, en el primer nodo cliente de la red móvil,
40 durante el primer período de tiempo, un porcentaje de las respuestas recibidas que es menor que el umbral de tiempo de incremento de respuesta y un porcentaje de las respuestas recibidas que es mayor que el umbral de tiempo de reducción de respuesta. En algunos aspectos, los sistemas y métodos incluyen aumentar, en el primer nodo cliente de la red móvil, la tasa de transacción inicial por la tasa de incremento para formar una tasa incrementada cuando el porcentaje de las respuestas recibidas que es menor que el umbral de tiempo de incremento de respuesta es mayor que el porcentaje del umbral de incremento de respuestas, y la tasa incrementada es menor o igual a la tasa máxima. En algunos aspectos, los sistemas y métodos incluyen disminuir, en el primer nodo cliente de la red móvil, la tasa de transacción inicial por la tasa de reducción para formar una tasa disminuida cuando: el porcentaje de las respuestas recibidas que es mayor que el umbral de tiempo de reducción de respuesta es mayor que el porcentaje del umbral de respuestas de reducción, y la tasa disminuida es mayor o igual que la tasa mínima.

50 En algunas modalidades, la velocidad aumentada está asociada con al menos una de una actualización del nodo servidor, un cambio en una condición de la red y la eliminación de un segundo nodo cliente de la red móvil en comunicación con el nodo servidor.

55 En algunas modalidades, la actualización al nodo servidor comprende al menos uno de: agregar al menos una CPU al nodo servidor; agregar memoria al nodo servidor; y agregar al menos una máquina virtual al nodo servidor.

60 En algunas modalidades, la velocidad reducida está asociada con al menos una de una degradación al nodo servidor, un cambio en una condición de la red y una adición de un segundo nodo cliente de la red móvil en comunicación con el nodo servidor.

65 En algunas modalidades, la degradación al nodo servidor comprende agregar al menos uno de: eliminar al menos una CPU al nodo servidor; eliminar memoria al nodo servidor; y eliminar al menos una máquina virtual al nodo servidor.

En algunas modalidades, el primer nodo cliente de la red móvil comprende al menos una puerta de enlace de servicio (SGW), puerta de enlace de red de paquetes de datos (PGW), nodo de soporte de puerta de enlace GPRS (GGSN), entidad de gestión de movilidad (MME), Nodo de Soporte del Servicio GPRS (SGSN), y Controlador de la Red de Radio (RNC). En algunas modalidades, el nodo servidor comprende al menos una función de políticas y cobros (PCRF), sistema de cobro en línea (OCS), servidor de autenticación, autorización y contabilización (AAA), servidor de abonado doméstico (HSS), función de pasarela de cobro (CGF), sistema de cobro fuera de línea (OFCS).

Estas y otras capacidades de la materia descrita se comprenderán más completamente después de una revisión de las siguientes figuras, descripción detallada y reivindicaciones. Debe entenderse que la fraseología y terminología que se emplea en la presente descripción tienen el propósito de la descripción y no deben considerarse como limitantes.

Breve descripción de las figuras

Varios objetivos, características y ventajas de la materia descrita pueden apreciarse más plenamente con referencia a la siguiente descripción detallada de la materia descrita cuando se considera en relación con los siguientes dibujos, en los que los números de referencia similares identifican elementos similares.

la Figura 1 es un diagrama del sistema que muestra diferentes nodos en un sistema en red operado inalámbricamente 100, de acuerdo con algunas modalidades de la presente descripción.

La Figura 2 es un diagrama del sistema que muestra nodos clientes y un nodo servidor en un sistema en red, de acuerdo con algunas modalidades de la presente descripción.

La Figura 3 es un diagrama del sistema que muestra un módulo de tasa de transacción máxima provista por el operador en un nodo cliente.

La Figura 4 es un diagrama del sistema que muestra un módulo de tasa de transacción adaptativa en un nodo cliente, de acuerdo con algunas modalidades de la presente descripción.

La Figura 5 es un método informatizado para ajustar dinámicamente una tasa óptima con base en el tiempo de respuesta de la transacción, de acuerdo con algunas modalidades de la presente descripción.

Descripción detallada

En las redes de operadores móviles, puede ser importante no sobrecargar los nodos pares (por ejemplo, función de políticas y cobros (PCRF), sistema de cobro en línea (OCS), servidor de autenticación, autorización y contabilización (AAA), servidor de abonado doméstico (HSS), función de la pasarela de cobro (CGF), sistema de cobro fuera de línea (OFCS)). La sobrecarga de un nodo par puede ocurrir al enviar más transacciones de las que puede manejar un nodo, desbordando así a un nodo par. A veces, un nodo par puede sobrecargarse catastróficamente si el nodo par es crítico para la operación (por ejemplo, HSS/PCRF) de la red.

En las redes puede ser difícil predecir por adelantado la tasa de transacción máxima óptima a la que un Nodo (Nodo-A) puede enviar a su Nodo par (Nodo-X) porque la tasa a la que el Nodo-X maneja las transacciones depende de varios factores. La mayoría de estos factores son de naturaleza muy dinámica y siguen cambiando en una red en vivo, por ejemplo: el soporte físico del Nodo-X (#CPU, Velocidad de CPU, Memoria, Caché, E/S, etc.), número de otros Nodos (Nodo-B, Nodo-C, etc.) que se conectan a él más la velocidad a la que estos otros nodos envían transacciones (dinámico), el número de instrucciones de CPU que toma el software para procesar cada transacción (dinámico), la congestión de red (dinámico). Con la configuración manual de los nodos de la red, existen al menos los siguientes problemas:

1. El operador necesita determinar y aprovisionar de alguna manera la tasa de transacción máxima a la que cada cliente puede enviar al servidor.
2. Cada vez que se agrega un nuevo cliente (por ejemplo, para la expansión de capacidad) al servidor, luego el operador debe determinar la nueva tasa de transacción máxima que cada cliente (Nodo-A y Nodo-B) puede enviar al servidor (Nodo-X) y aprovisionar en consecuencia.
3. Cada vez que el servidor se actualiza con nuevo soporte físico, el operador debe determinar nuevamente la nueva tasa máxima de transacción y la provisión en cada cliente.
4. En la red del operador, si los clientes pueden comunicarse con varios servidores (primario/secundario o para equilibrar la carga), entonces, para cada cliente-servidor, esta tasa de transacción máxima necesita aprovisionarse y necesita ajustarse cada vez que se agrega un nuevo cliente o se cambia el servidor, la condición de la red cambia.
5. Para evitar la interrupción de la red, el operador se ve obligado a usar un número conservador y esto afecta innecesariamente la experiencia del usuario final (Caída de llamadas con la red ocupada) y esto puede llevar a perder clientes frente a sus competidores.

Existe la necesidad de una solución para determinar cómo cada nodo cliente o elemento de red cliente (por ejemplo, puerta de enlace de servicio (SGW), puerta de enlace de red de paquetes de datos (PGW), nodo de soporte de puerta de enlace GPRS (GGSN), entidad de gestión de movilidad (MME), Nodo de Soporte del Servicio GPRS (SGSN) y el Controlador de la Red de Radio (RNC) pueden determinar dinámicamente la tasa de transacción máxima permitida para enviar a cada uno de sus nodos pares (también se denominan en la presente descripción como nodos servidores o elementos de la red servidor) (por ejemplo, PCRF, OCS, AAA, HSS, CGF, OFCS). El iniciador de una transacción se denomina en la presente descripción como un elemento de red cliente y el receptor de una transacción se denomina en la presente descripción como elemento de red servidor. Son posibles muchas combinaciones de pares de servidor-cliente, incluido un elemento central de paquete que funciona como un nodo servidor para otro elemento central de paquete.

La Figura 1 es un diagrama del sistema que muestra diferentes nodos en un sistema en red operado inalámbricamente 100, de acuerdo con algunas modalidades de la presente descripción. El sistema 100 incluye múltiples instancias de equipo de usuario (UE) 102, Red de Acceso Terrestre Universal Evolucionada (E-UTRAN) Nodo B (eNodeB) 104, Evolución de Paquetes (EPC) Entidad de Gestión de Movilidad (MME) 106, Servidor de Abonado Doméstico (HSS) 108, Puerta de Enlace de Servicio (SGW) 110, Puerta de Enlace de Red de Paquetes de Datos (PGW) 112, Función de Políticas y Cobros (PCRF) 114, Sistema de Cobro en Línea (OCS) 116, Sistema de Cobro Fuera de Línea (OFCS) 118, Servidor de Autenticación, Autorización y Contabilización (AAA) 120, Red de Paquetes de Datos 122, Centro de Monitoreo de Cumplimiento de la Ley (LEMF) 124, Nodo de Soporte del Servicio GPRS (SGSN) 126, Nodo de Soporte de Puerta de Enlace GPRS (GGSN) 128.

El UE 102 está en comunicación con el eNodeB 104 cuando está en un área de cobertura de red 4G y el UE 102 está en comunicación con el Nodo B (no se muestra) cuando está en un área de cobertura de red GSM 2G/3G. El UE 102 puede ser cualquier dispositivo capaz de comunicarse con la red 100. Por ejemplo, el UE 102 puede ser un ordenador portátil, un ordenador personal, un teléfono celular, incluido un asistente digital personal (PDA), un teléfono inteligente, un dispositivo de máquina a máquina (M2M) y un dispositivo de Internet de las Cosas (IOT). En las redes de operadores inalámbricos Teir-1 puede haber cientos de millones de UE 102 que se conectan a la red 100.

El eNodeB 104 es un elemento de red en la arquitectura E-UTRAN que se comunica con múltiples UE 102 a través de la interfaz aérea, se comunica con una MME 106 para mensajes de ruta de control y se comunica con SGW 110 para mensajes de ruta de usuario. Cada eNodeB proporciona servicio inalámbrico a una parte de un área de cobertura geográfica y tiene una capacidad limitada. En las redes de operadores inalámbricos Teir-1 puede haber cientos de miles de eNodeBs para cubrir todo el país y satisfacer las necesidades de capacidad en un área densamente poblada.

La MME 106 es uno de los elementos clave de la red del plano de control en una arquitectura E-UTRAN la cual es responsable de autenticar el UE 102 con el HSS 108. La MME 106 maneja los eventos de configuración de sesión, desmontaje de sesión, movilidad, paginación y administración de sesión. La MME también selecciona los elementos de red SGW 110, PGW 112 para alojar la sesión del UE 102. En las redes de operadores inalámbricos Teir-1 puede haber decenas de miles de estos nodos.

El HSS 108 es un elemento de red con una base de datos central que contiene información relacionada con el usuario y la suscripción. La MME 106 se comunica con el HSS 108 para la autenticación del usuario y para otra información de suscripción.

La SGW 110 es un elemento de red que enruta y reenvía paquetes de datos de usuario, mientras que también actúa como el ancla de movilidad para un plano de usuario durante las transferencias entre eNodeB. La SGW 110 se comunica con la PGW 112 hacia el lado central para el control y el tráfico de datos y en el lado de acceso se comunica con la MME 106 para el control y el eNodeB 104 para el tráfico de datos. En las redes de operadores inalámbricos Teir-1 podría haber miles de estos nodos.

La PGW 112 es un elemento de red que proporciona conectividad al UE 102 con redes externas de paquetes de datos 122. Un UE 102 puede tener conectividad simultánea con más de una red de paquetes de datos 122. La PGW 112 realiza la aplicación de la política con base en una política que se recibe desde PCRF 114, cobro en línea con OCS 114, cobro fuera de línea con OFCS 116, interceptación legal con LEMF 124, autenticación, autorización y contabilidad con AAA 120 =. Cada PGW 112 puede anclar millones de UEs 102 y en las redes de operadores inalámbricos Teir-1 podría haber miles de estos nodos.

En una red de operador hay múltiples nodos (elementos de red) y cada nodo se comunica con múltiples instancias de otros nodos para procesar transacciones. Esta descripción proporciona soluciones para al menos dos problemas:

1. Cuando un nodo cliente (por ejemplo, PGW 112) envía transacciones al nodo servidor (por ejemplo, PCRF 114), puede ser útil determinar, en algunas modalidades, la tasa de transacción óptima a la que el nodo cliente puede enviar transacciones al nodo servidor (y más allá) de manera que el nodo servidor no se sobrecargue. Determinar la tasa de transacción óptima puede ser un desafío, especialmente en un entorno dinámico donde múltiples clientes dinámicos (por ejemplo, clientes con tasas de transacciones variables) se comunican con un solo nodo servidor. Además, determinar una tasa de transacción óptima puede ser un desafío porque cada nodo servidor puede manejar diferentes tasas con base en el soporte físico que ejecuta el servidor. Por ejemplo, cuando un servidor ejecuta en un mundo virtualizado (por ejemplo: nube), los recursos del soporte físico se asignan y liberan dinámicamente, y las tasas de transacción del cliente deben ajustarse en consecuencia.

2. Cuando el nodo Cliente (por ejemplo, PGW 112) puede enviar una transacción a cualquier nodo Servidor en un grupo de nodos Servidores (PCRF1, PCRF2, etc. 114), puede ser útil, en algunas modalidades, establecer los criterios que usa un nodo Cliente para seleccionar un nodo Servidor en el grupo de servidores de manera que las transacciones se equilibren en la carga en este grupo de nodos Servidores. Cada nodo Servidor puede ejecutarse con diferentes recursos de soporte físico y estos recursos pueden asignarse o desasignarse dinámicamente (en VM o en el entorno de la nube) y, por lo tanto, los números de capacidad/rendimiento del nodo Servidor son diferentes y pueden ser dinámicos. Además,

5 cada nodo Servidor puede estar comunicándose con un número diferente de nodos Clientes y un nodo Cliente en particular no necesariamente conoce la cantidad de transacciones que otros nodos Clientes envían a un nodo Servidor. Por ejemplo, cuando hay varios servidores en el grupo de servidores y cada uno de estos servidores puede manejar diferentes tasas de transacción, al enviar una transacción, el cliente elige un servidor de este grupo de servidores. En algunas modalidades, los clientes usan un algoritmo de turno rotativo para enviar transacciones a través de múltiples servidores, pero cuando estos servidores tienen diferentes potencias de procesamiento, entonces los clientes pueden necesitar usar un algoritmo de turno rotativo ponderado. Sin embargo, cuando un servidor agrega o elimina dinámicamente potencia de procesamiento, un operador puede modificar la configuración del lado del cliente. Como se describió en la presente descripción, puede usarse un tiempo de respuesta de transacción de un servidor para seleccionar un servidor para la próxima transacción. Si un servidor responde rápidamente, entonces está cargado ligeramente y puede procesar más transacciones.

15 La Figura 2 es un diagrama del sistema que muestra nodos clientes y un nodo servidor en un sistema en red, de acuerdo con algunas modalidades de la presente descripción. La Figura 2 muestra el Nodo-A Cliente 202, el Nodo-B Cliente 204, el Nodo-N Cliente 206, el Nodo-X Servidor 208, las transacciones entre el Nodo-A Cliente y el Nodo-X Servidor 210, las transacciones entre el Nodo-B Cliente y el Nodo-X Servidor 212, y transacciones entre el Nodo-N Cliente y el Nodo-X Servidor 214.

20 El Nodo-A Cliente 202, el Nodo-B Cliente 204 y el Nodo-N cliente 206 son clientes del Nodo-X Servidor 208. En algunas modalidades, un nodo cliente solicita un servicio del nodo servidor y el nodo servidor proporciona el servicio. Por ejemplo, en la Figura 1, varias instancias de PGW 112 se comunican con PCRF 114, cada PGW 112 actúa como un Nodo Cliente y PCRF 114 actúa como un Nodo Servidor. Mediante el uso de este canal de comunicación, la PGW 112 recupera la información de la Política UE 102 de PCRF 114 y aplica la política. Otro ejemplo son varias instancias de PGW 112 que se comunican con el OCS 116 donde cada PGW 112 actúa como un Nodo Cliente, donde el OCS 116 actúa como un Nodo Servidor. Mediante el uso de este canal de comunicación, la PGW 112 recupera los créditos de uso en línea del UE 102 del nodo servidor OCS 116.

30 Cada uno de los Nodo-A Cliente 202, Nodo-B Cliente 204 y Nodo-N Cliente 206 puede enviar una tasa diferente de transacciones 210, 212 y 214 al servidor. Por ejemplo, el Nodo-A Cliente 202 puede manejar un área del centro poblada densa al generar una tasa de transacción más alta (transacciones/segundo) 210 al Nodo-X Servidor 208, el Nodo-B Cliente 204 puede manejar un área poblada moderada al generar una tasa de transacción moderada 212 al Nodo-X Servidor 208 y Nodo-C Cliente 206 pueden manejar áreas poco pobladas al generar una menor tasa de transacción 214 al Nodo-X Servidor 208. Cada uno de los Nodos Clientes 202, 204 y 206 puede enviar primero una solicitud al Nodo Servidor 208 para determinar una tasa máxima permitida para enviar al servidor. Si todos los Nodos Clientes 202, 204 y 206 envían simultáneamente transacciones al Nodo Servidor 208 a una tasa alta, el Nodo Servidor 208 puede verse desbordado. Por ejemplo, varios nodos PGW que se comunican con un solo servidor PCRF pueden desbordar al servidor PCRF si todos le envían transacciones simultáneamente a una tasa alta. Esto puede suceder durante el período de horas ocupadas o durante la actualización de software en estos nodos clientes o en otros eventos en los que muchos UE intentan conectarse a la red al mismo tiempo.

40 En algunas modalidades, un nodo servidor puede ser un nodo remoto en una Virtualización de Funciones de Red (NFV). Un nodo remoto puede aumentar o disminuir dinámicamente su capacidad de procesamiento automáticamente sin notificar a los nodos emisores. Durante estos eventos, los nodos que envían al nodo remoto pueden configurarse de tal manera que pueden desbordar o disminuir el nodo remoto. Por ejemplo, considere un nodo virtualizado remoto que ha determinado que necesita capacidad de procesamiento adicional y solicitó y recibió la capacidad adicional de la infraestructura virtualizada. En su nuevo estado, el nodo remoto ahora es capaz de procesar un mayor número de transacciones de lo que era. Los nodos emisores desconocen la capacidad adicional y pueden estar estrangulando innecesariamente sus transacciones hacia el nodo remoto, de esta manera resulta en una reducción en el servicio. Adicionalmente, considere el caso en el que un nodo virtualizado remoto ha determinado que no necesita la capacidad que se le asignó y libera los recursos virtuales de vuelta a la infraestructura virtualizada. En este caso, sin una solución adaptativa de sobrecarga de pares, los otros nodos en la red pueden configurarse de tal manera que desborden al nodo remoto que ha reducido su consumo de recursos.

55 La Figura 3 es un diagrama del sistema que muestra un módulo de tasa de transacción máxima provista por el operador en un nodo cliente. La Figura 3 muestra el Nodo-A Cliente 202, el Nodo-B Cliente 204, el Nodo-X Servidor, la tasa de transacción entre el Nodo-A Cliente y el Nodo-X Servidor 310 y el Módulo de Tasa de Transacción Máxima Estática 320. En este modelo, el operador configura por adelantado tasas de transacción máximas estáticas muy conservadoras en el lado del cliente.

60 Típicamente, las tasas de transacción máximas estáticas se configuran en el lado del cliente para controlar la tasa a la cual puede enviarse al servidor. Para controlar la tasa de transacción máxima entre el Nodo-A Cliente 202 y el Nodo-X Servidor 208, los operadores móviles deben determinar y proporcionar manualmente una tasa de transacción máxima mediante el uso de un módulo de tasa de transacción máxima estática 320. Para no colapsar la red, los operadores móviles generalmente configuran tasas de transacción máximas muy conservadoras/artificiales en cada nodo de cliente (por ejemplo, Nodo-A Cliente). Los operadores móviles también tienen que configurar la tasa de transacción máxima cada vez que cambia una condición en la red. Por ejemplo, cada vez que se agrega un nuevo cliente (por ejemplo, Nodo-B

204) (por ejemplo, para la expansión de capacidad) al servidor, entonces el operador necesita determinar la nueva tasa de transacción máxima de cada cliente (Nodo-A 202 y Nodo-B 204) puede enviar al servidor (Nodo-X 208) y aprovisionar en consecuencia. Como otro ejemplo, cada vez que el servidor se actualiza con nuevo soporte físico, el operador debe determinar nuevamente la nueva tasa de transacción máxima y la provisión en cada cliente. Como un ejemplo adicional, si los clientes pueden comunicarse con varios servidores (primario/secundario o para equilibrar la carga), entonces, para cada cliente-servidor, la tasa de transacción máxima necesita aprovisionarse y necesita ajustarse cada vez que se agrega un nuevo cliente, un servidor es cambiado, o las condiciones de la red cambian. Para evitar la interrupción de la red, un operador se ve obligado a usar una tasa de transacción máxima conservadora, que afecta innecesariamente la experiencia del usuario final (por ejemplo, desconectando llamadas con la red ocupada) y podría llevar a la pérdida de clientes frente a sus competidores.

En lugar de usar una tasa de transacción máxima estática y conservadora, las modalidades de la presente descripción describen la adaptación de la tasa de transacción máxima con base en varios factores, que incluyen, entre otros: tiempo de respuesta de la transacción, umbrales de incrementos de la tasa y cierta histéresis.

La Figura 4 es un diagrama del sistema que muestra un módulo de tasa de transacción adaptativa en un nodo cliente, de acuerdo con algunas modalidades de la presente descripción. La Figura 4 muestra el Nodo-A Cliente 202, el Nodo-B Cliente 204, el Nodo-X Servidor 208, la tasa de transacción máxima entre el Nodo-A Cliente y el Nodo-X Servidor 410 y el módulo de tasa de transacción adaptativa 420.

El Nodo-A Cliente 202 incluye un módulo de tasa de transacción adaptativa 420. El módulo de tasa de transacción adaptativa 420 mediante el uso de un algoritmo adaptativo puede determinar automáticamente una tasa de transacción óptima mediante el uso del tiempo de respuesta de la transacción. En algunas modalidades, el módulo de tasa de transacción adaptativa 420 puede determinar automáticamente una tasa de transacción óptima mediante el uso de al menos una parte de los siguientes atributos provistos por el operador:

1. tasa inicial
2. tasa máxima
3. tasa mínima
4. período de evaluación
5. tasa de incremento
6. umbral de tiempo de respuesta de transacción de incremento
7. porcentaje de respuestas que se reciben dentro del umbral de incremento
8. tasa de reducción
9. umbral de tiempo de respuesta de transacción de reducción
10. porcentaje de respuestas que excede el umbral de reducción

En algunas modalidades, un algoritmo puede comenzar a una tasa inicial y el operador puede establecer la tasa inicial a un valor razonable en el cual el Servidor pueda manejar las transacciones. En algunas modalidades, la tasa máxima puede referirse a una tasa que un algoritmo adaptativo deja de aumentar más allá y una tasa mínima puede referirse a una tasa que un algoritmo adaptativo deja de disminuir a continuación. Tanto la tasa máxima como la tasa mínima pueden configurarse por el operador.

En algunas modalidades, el período de tiempo de evaluación se refiere a un período de tiempo de espera antes de cambiar la tasa. En algunas modalidades, la tasa de incremento se refiere a una tasa incremental, el algoritmo adaptativo aumentará una tasa de transacción y una tasa de reducción se refiere a una tasa de disminución, el algoritmo adaptativo disminuirá una tasa de transacción. En algunas modalidades, el umbral de tiempo de respuesta de transacción de incremento se refiere a un umbral de tiempo de respuesta, o respuestas que se reciben del servidor dentro de un cierto período de tiempo, para contar las respuestas en un depósito de incremento y un umbral de tiempo de respuesta de transacción de reducción se refiere al umbral de tiempo de respuesta para contar las respuestas en el depósito de reducción. El depósito de incremento se refiere a un número de respuestas de transacción que se reciben dentro de un umbral de tiempo de respuesta de transacción de incremento. El depósito de reducción se refiere a un número de respuestas de transacción que se reciben más allá del umbral de tiempo de respuesta de transacción de reducción. En algunas modalidades, el porcentaje de respuestas que se reciben dentro del umbral de incremento incluye un porcentaje de respuestas que deben recibirse durante un período de tiempo de evaluación dentro del umbral de tiempo de respuesta de transacción de incremento para aumentar la tasa por una tasa de incremento y un porcentaje de respuestas que se reciben dentro del umbral de reducción se refiere al porcentaje de respuestas que deben recibirse durante un período de tiempo de evaluación, dentro del umbral de tiempo de respuesta de transacción de reducción para disminuir la tasa en una tasa de reducción.

Incluso si no se establece una tasa inicial tan precisa, durante un período de tiempo, el módulo de tasa de transacción adaptativa 420 puede determinar automáticamente la tasa de transacción óptima y operar a esa tasa. La tasa de transacción óptima puede variar con el tiempo. La posición del módulo de tasa de transacción adaptativa 420 en un nodo cliente permite que el módulo de tasa de transacción adaptativa 420 se adapte a las variaciones en el rendimiento de procesamiento del nodo remoto. El módulo de tasa de transacción adaptativa 420 puede desplegarse en cualquier número o combinaciones de nodos cliente dentro de una red. Particularmente en los casos donde el nodo remoto es una Función de Red Virtual (VNF), puede haber casos en los que el VNF alcanzó un umbral y solicitó capacidad de procesamiento

adicional de la infraestructura virtualizada. Esta capacidad adicional puede tener la forma de una CPU virtual adicional, Memoria virtual o máquinas virtuales adicionales. Durante estos tiempos, la capacidad del nodo remoto puede cambiar. Al finalizar un evento de escalamiento horizontal del nodo remoto donde se agregan máquinas virtuales adicionales al VNF, el módulo de tasa de transacción adaptativa 420 proporciona que las tasas de transacción aumenten proporcionalmente.

La Figura 5 es un método informatizado 500 para ajustar dinámicamente una tasa óptima con base en el tiempo de respuesta de la transacción, de acuerdo con algunas modalidades de la presente descripción. En la etapa 502, recibe los parámetros de configuración. En la etapa 504, establezca la tasa óptima a la tasa inicial configurada y envíe la transacción a la tasa inicial. En la etapa 506, para el período de tiempo de evaluación configurado, mida el porcentaje de respuestas con un tiempo de respuesta menor que el valor de umbral de tiempo de respuesta de transacción de incremento configurado y también mida el porcentaje de respuestas con el tiempo de respuesta mayor que el valor umbral de tiempo de respuesta de transacción de reducción. En la etapa 508, verifique si se calculó el "porcentaje de respuestas que se reciben dentro del umbral de tiempo de respuesta de transacción de incremento" > el porcentaje de respuestas que se reciben dentro del umbral de incremento. En la etapa 510, si la respuesta a la etapa 508 es "Sí" y la tasa de entrada \geq tasa óptima, entonces aumente la tasa óptima por la tasa de incremento si $(\text{tasa óptima} + \text{tasa de incremento}) < \text{tasa máxima}$ de lo contrario, establezca la tasa óptima con la tasa máxima. En la etapa 512, si la respuesta a la etapa 508 es "No", verifique ahora si se calculó el "porcentaje de respuestas que se reciben que excede el umbral de tiempo de respuesta de transacción de reducción" > el porcentaje de respuesta que excede el umbral de reducción. En la etapa 514, si la respuesta a la etapa 512 es "Sí", disminuya la tasa óptima mediante la tasa de reducción si $(\text{tasa óptima} - \text{tasa de reducción}) > \text{tasa mínima}$, de lo contrario, establezca la tasa óptima con la tasa mínima. En la etapa 516, limite el número de transacciones a la nueva tasa óptima y repita las etapas 506 a 512 y ajuste continuamente la tasa óptima.

Con referencia a la etapa 502, los parámetros de configuración pueden incluir los parámetros como se discutió anteriormente. Brevemente, los parámetros de configuración pueden incluir lo siguiente:

1. tasa inicial,
2. tasa máxima
3. tasa mínima
4. período de evaluación
5. tasa de incremento
6. umbral de tiempo de respuesta de transacción de incremento
7. porcentaje de respuestas que se reciben dentro del umbral de incremento
8. tasa de reducción
9. umbral de tiempo de respuesta de transacción de reducción
10. porcentaje de respuestas que excede el umbral de reducción

Con referencia a la etapa 504, el algoritmo adaptativo asociado con el módulo de tasa de transacción adaptativa 420 comienza a una tasa óptima con tasa inicial provista por el operador y limita la tasa de transacción saliente a esta tasa óptima. Si la tasa de transacción entrante es más alta que la tasa de transacción saliente, entonces las transacciones en exceso pueden descartarse de manera inteligente (por ejemplo, descartar la transacción de menor prioridad) o almacenar temporalmente al asumir que la tasa óptima aumenta con el tiempo.

Con referencia a la etapa 506, para el período de tiempo de evaluación, el módulo de tasa de transacción adaptativa 420 mide el tiempo de respuesta para cada transacción y mide el porcentaje de respuestas de transacción que se reciben dentro del umbral de tiempo de respuesta de transacción de incremento. El módulo de tasa de transacción adaptativa 420 también mide un porcentaje de respuestas de transacción que se reciben que exceden el umbral de tiempo de respuesta de transacción de reducción.

Con referencia a la etapa 508, el módulo de tasa de transacción adaptativa 420 verifica si el porcentaje de respuestas de transacción que se reciben dentro del período de tiempo de respuesta de transacción de incremento es mayor que el porcentaje de respuestas que se reciben dentro del umbral de incremento.

Con referencia a la etapa 510, si la respuesta a la etapa 508 es "Sí" y si la tasa entrante es mayor o igual a la tasa actual, entonces la tasa óptima operativa actual se incrementa en la cantidad de tasa de incremento siempre que la tasa nueva sea menor que la tasa máxima; de lo contrario, se establece en la tasa máxima permitida.

Con referencia a la etapa 512, si la respuesta a la etapa 508 es "No", verifique si el "porcentaje de respuestas de transacción que se reciben que excede el período de tiempo de respuesta de transacción de reducción" es mayor que el porcentaje de respuestas que excede el umbral de reducción.

Con referencia a la etapa 514, si la respuesta a la etapa 512 es "Sí", entonces la tasa óptima operativa actual se reduce en una cantidad de tasa de reducción siempre que la nueva tasa óptima sea mayor que la tasa mínima; de lo contrario, establezca la tasa óptima en tasa mínima.

Con referencia a la etapa 516, el algoritmo usa la nueva tasa óptima para limitar el procesamiento de las transacciones entrantes.

Las etapas 506 a la 516 se repiten cada período de tiempo de evaluación y ajusta dinámicamente la tasa óptima.

5 Con referencia a las Figuras 5, si un nodo servidor está ligeramente cargado, procesa las transacciones y responde a una velocidad más rápida. El nodo cliente puede aumentar la tasa de transacción en respuesta, siempre que el nodo cliente tenga más transacciones que ofrecer que la tasa óptima actual. Por otro lado, si el nodo servidor se está sobrecargando, entonces lleva más tiempo procesar y responder a las transacciones. Esto activa el nodo Cliente para reducir la tasa de transacción en respuesta. Así es como el sistema se ajusta y opera automáticamente en una tasa de transacción óptima.

- 10 Por ejemplo, suponga que un operador configuró:
- tasa inicial = 1000,
 - tasa máxima = 20 000,
 - tasa mínima = 100
 - período de tiempo de evaluación = 5 segundos
- 15
- tasa de incremento = 100,
 - umbral de tiempo de respuesta de transacción de incremento = 1000 milisegundos
 - porcentaje de respuestas que se recibe dentro del umbral de incremento $d = 98$,
 - tasa de reducción = 200,
 - umbral de tiempo de respuesta de transacción de reducción = 1500 milisegundos,
- 20
- porcentaje de respuestas que excede el umbral de reducción = 5,

25 Supongamos que el nodo Servidor puede manejar ~5000 transacciones por segundo (envía respuestas dentro de 1000 milisegundos) y con la configuración anterior, el cliente comienza en 1000 transacciones por segundo (tasa inicial) y si el 98 % (porcentaje de respuestas que se recibe dentro del umbral de incremento) de las respuestas se encuentran dentro de los 1000 milisegundos (umbral de tiempo de respuesta de transacción de incremento) durante los próximos 5 segundos (período de tiempo de evaluación), luego la tasa se incrementa en 100 adicionales (tasa de incremento) y este proceso se repite cada 5 segundos (período de tiempo de evaluación) y el aumento de la tasa aumenta a 5000 transacciones por segundo en 200 segundos. La tasa se aumenta hasta que el 98 % de las respuestas se encuentran dentro de 1000 milisegundos. Más tarde, suponga que el operador agrega más capacidad (escala de VM) en el lado del servidor y ahora,

30 por ejemplo, puede manejar 6000 transacciones por segundo, luego el cliente comienza a ver que más del 98 % de las respuestas llegan dentro de 1000 milisegundos y, por lo tanto, comienza a aumentar la tasa nuevamente y alcanza 6000 transacciones por segundo en los próximos 50 segundos y opera a esa velocidad. Ahora, suponga que el operador reduce la capacidad del servidor (por ejemplo: escala de VM o falla parcial del servidor) y debido a esto, ahora solo puede manejar 5000 transacciones por segundo. El cliente comienza a ver respuestas de transacciones más largas y si las respuestas de más del 5 % (porcentaje de respuestas que excede el umbral de reducción en el último segundo) son mayores de 1500 milisegundos, entonces la tasa se reduce en 200 (tasa de reducción) y dentro de 25 segundos la tasa baja a 5000 transacciones/segundo. Y si el servidor está completamente inactivo, en otros 125 segundos la velocidad baja a la tasa mínima (100 transacciones por segundo).

40 En las modalidades de la presente descripción, si alguna condición operativa de la red/servidor cambia, entonces el cliente aumentará o disminuirá automáticamente la velocidad y funcionará a la velocidad óptima. El algoritmo anterior es una función de aceleración y desaceleración lineal simple con base en el tiempo de respuesta, pero si el operador prefiere diferentes funciones de aceleración y desaceleración (por ejemplo: logarítmica o exponencial), esto puede mejorarse fácilmente.

45 En algunas modalidades de la presente descripción, se describen sistemas y métodos para ajustar adaptativamente una tasa de transacción en un nodo de la red móvil. En algunas modalidades, los sistemas y métodos descritos en la presente descripción incluyen recibir, en un primer nodo cliente de la red móvil, parámetros de configuración asociados con al menos uno de una tasa de transacción inicial, una tasa de incremento, una tasa de reducción, un período de tiempo de evaluación, un umbral de tiempo de incremento de respuesta, un umbral de tiempo de reducción de respuesta, un porcentaje de respuestas que se reciben dentro del umbral de incremento y un porcentaje de respuestas que se reciben dentro del umbral de reducción. En algunas modalidades, los sistemas y métodos descritos en la presente descripción incluyen recibir, en el primer nodo cliente de la red móvil, respuestas asociadas con comunicaciones con un nodo servidor.

50 En algunas modalidades, los sistemas y métodos descritos en la presente descripción incluyen determinar, en el primer nodo cliente de la red móvil, un porcentaje de respuestas que se reciben que es menor que el umbral de tiempo de incremento de respuesta y un porcentaje de respuestas que se reciben que es mayor que el umbral de tiempo de reducción de respuesta. En algunas modalidades, los sistemas y métodos descritos en la presente descripción incluyen aumentar, en el primer nodo cliente de la red móvil, la tasa de transacción inicial por la tasa de incremento cuando el porcentaje de respuestas que se reciben que es menor que el umbral de tiempo de incremento de respuesta es mayor que el porcentaje de respuestas que se reciben dentro del umbral de incremento durante el período de tiempo de evaluación. En algunas modalidades, los sistemas y métodos descritos en la presente descripción incluyen disminuir, en el primer nodo cliente de la red móvil, la tasa de transacción inicial por la tasa de reducción cuando el porcentaje de respuestas que se reciben que es mayor que el umbral de tiempo de reducción de respuesta es mayor que el porcentaje de respuestas que se reciben dentro del umbral de reducción durante el período de tiempo de evaluación.

65

La materia descrita en la presente descripción puede implementarse en circuitos electrónicos digitales, o en programa informático, microprograma o soporte físico, incluidos los medios estructurales descritos en esta descripción y sus equivalentes estructurales, o en combinaciones de ellos. La materia descrita en la presente descripción puede implementarse como uno o más productos de programas informáticos, como uno o más programas informáticos incorporados de manera tangible en un soporte de información (por ejemplo, en un dispositivo de almacenamiento legible por máquina), o incorporados en una señal propagada, para su ejecución por o para controlar el funcionamiento de un aparato de procesamiento de datos (por ejemplo, un procesador programable, un ordenador o varios ordenadores). Un programa informático (también conocido como un programa, software, aplicación de software, o código) puede escribirse en cualquier forma de lenguaje de programación, que incluye lenguajes compilados o interpretados, y puede implementarse en cualquier forma, que incluye un programa independiente o un módulo, componente, subrutina, u otra unidad adecuada para su uso en un ambiente de ordenadores. Un programa informático no necesariamente corresponde a un archivo. Un programa puede almacenarse en una parte de un archivo que contiene otros programas o datos, en un único archivo dedicado al programa en cuestión o en múltiples archivos coordinados (por ejemplo, archivos que almacenan uno o más módulos, subprogramas o partes de código). Un programa informático puede implementarse para ejecutarse en un ordenador o en múltiples ordenadores en un sitio o distribuidos a través de múltiples sitios e interconectados por una red de comunicación.

Los procesos y flujos lógicos descritos en esta descripción, incluidas las etapas del método de la materia descrita en la presente descripción, pueden ser realizados por uno o más procesadores programables que ejecutan uno o más programas informáticos para realizar funciones de la materia descrita en la presente descripción al operar con datos de entrada y generando salida. Los procesos y flujos lógicos también pueden realizarse, y el aparato de la materia descrita en la presente descripción pueden implementarse como un circuito lógico de propósito especial, por ejemplo, una FPGA (matriz de compuerta programable de campo) o un ASIC (circuito integrado específico de aplicación).

Los procesadores adecuados para la ejecución de un programa informático incluyen, a manera de ejemplo, microprocesadores de propósito general y especial, y uno o más procesadores de cualquier tipo de ordenador digital. Generalmente, un procesador recibirá instrucciones y datos de una memoria de solo lectura o una memoria de acceso aleatorio o ambas. Los elementos esenciales de un ordenador son un procesador para ejecutar instrucciones y uno o más dispositivos de memoria para almacenar instrucciones y datos. En general, un ordenador también incluirá, o estará operativamente acoplado para recibir datos o transferir datos a, o ambos, uno o más dispositivos de almacenamiento masivo para almacenar datos, por ejemplo, discos magnéticos, discos magnetoópticos o discos ópticos. Los soportes de información adecuados para incorporar instrucciones y datos de programas informáticos incluyen todas las formas de memoria no volátil, incluidos, a manera de ejemplo, dispositivos de memoria de semiconductores (por ejemplo, EPROM, EEPROM y dispositivos de memoria flash); discos magnéticos (por ejemplo, discos duros internos o discos extraíbles); discos magnetoópticos; y discos ópticos (por ejemplo, discos CD y DVD). El procesador y la memoria pueden complementarse o incorporarse en un circuito lógico de propósito especial.

Para proporcionar una interacción con un usuario, la materia descrita en la presente descripción puede implementarse en un ordenador que tenga un dispositivo de visualización, por ejemplo, un monitor CRT (tubo de rayos catódicos) o LCD (pantalla de cristal líquido), para mostrar información al usuario y un teclado y un dispositivo señalador (por ejemplo, un ratón o una bola de seguimiento), mediante el cual el usuario puede proporcionar información al ordenador. También se pueden utilizar otros tipos de dispositivos para proporcionar la interacción con un usuario. Por ejemplo, la retroalimentación proporcionada al usuario puede ser cualquier forma de retroalimentación sensorial (por ejemplo, retroalimentación visual, retroalimentación auditiva o retroalimentación táctil), y la entrada del usuario puede recibirse en cualquier forma, incluyendo entrada acústica, del habla o táctil.

La materia descrita en la presente descripción puede implementarse en un sistema informático que incluye un componente del lado del servidor (por ejemplo, un servidor de datos), un componente intermedio (por ejemplo, un servidor de aplicaciones) o un componente del lado del usuario (por ejemplo, un ordenador cliente que tiene una interfaz gráfica de usuario o un navegador web a través del cual un usuario puede interactuar con una implementación de la materia descrita en la presente descripción), o cualquier combinación de dichos componentes del lado del servidor, intermedios y del lado del usuario. Los componentes del sistema pueden estar interconectados por cualquier forma o medio de comunicación de datos digitales, por ejemplo, una red de comunicación. Los ejemplos de redes de comunicación incluyen una red de área local ("LAN") y una red de área amplia ("WAN"), por ejemplo, Internet.

Debe entenderse que la materia descrita no se limita en cuanto a su aplicación a los detalles de la construcción y a las disposiciones de los componentes expuestos en la siguiente descripción o ilustrados en los dibujos. La materia descrita es capaz de otras modalidades y de realizarse o llevarse a la práctica de varias maneras. Además, debe entenderse que la fraseología y terminología empleadas en la presente descripción tienen el propósito de la descripción y no deben considerarse como limitantes.

Como tal, los expertos en la técnica apreciarán que la concepción sobre la que se basa esta descripción, puede usarse fácilmente como una base para el diseño de otras estructuras, métodos y sistemas para llevar a cabo los diversos propósitos de la materia descrita.

Aunque la materia descrita se ha descrito e ilustrado en las modalidades ilustrativas anteriores, se entiende que la presente descripción se ha realizado solo a manera de ejemplo, y que pueden realizarse numerosos cambios en los detalles de implementación de la materia descrita sin apartarse del alcance de la materia descrita, que está limitado solo por las reivindicaciones que siguen.

5

REIVINDICACIONES

1. Un método informatizado (500) para ajustar adaptativamente una tasa de transacción en un nodo de la red móvil (202), el método informatizado (500) que comprende:
- 5 recibir (502), en un primer nodo cliente de la red móvil (202), parámetros de configuración que incluyen al menos uno de:
- una tasa de transacción inicial, la tasa de transacción inicial asociada con un número inicial de transacciones por segundo que se comunica por el primer nodo cliente de la red móvil (202) a un nodo servidor (208) a través de una red en un momento inicial,
- 10 una tasa de incremento, la tasa de incremento asociada con un primer número de transacciones por segundo que indica un aumento en la tasa de transacción inicial,
- una tasa de reducción, la tasa de reducción asociada con un segundo número de transacciones por segundo que indica una disminución en la tasa de transacción inicial,
- 15 una tasa mínima,
- una tasa máxima,
- un período de evaluación,
- un umbral de tiempo de incremento de respuesta asociado con un tiempo de respuesta indicativo de aumentar la tasa de transacción inicial,
- 20 un umbral de tiempo de reducción de respuesta asociado con un tiempo de respuesta indicativo de disminución de la tasa de transacción inicial,
- un porcentaje del umbral de incremento de respuestas asociado con el umbral de tiempo de incremento de respuesta, y
- un porcentaje de umbral de reducción de respuestas asociado con el umbral de tiempo de reducción de respuesta; y el método caracterizado por:
- 25 recibir, en el primer nodo cliente de la red móvil (202), respuestas asociadas con las comunicaciones con el nodo servidor (208) durante un primer período de tiempo asociado con el período de tiempo de evaluación;
- al determinar (506), en el primer nodo cliente de la red móvil (202), durante el primer período de tiempo, un porcentaje de las respuestas que se reciben que es menor que el umbral de tiempo de incremento de respuesta y un porcentaje de las respuestas que se reciben que es mayor que el umbral de tiempo de reducción de respuesta;
- 30 al aumentar (510), en el primer nodo cliente de la red móvil (202), la tasa de transacción inicial por la tasa de incremento para formar una tasa incrementada cuando:
- el porcentaje de respuestas que se reciben que es menor que (508) el umbral de tiempo de incremento de respuesta que es mayor que el porcentaje del umbral de incremento de respuestas, y
- 35 la tasa incrementada es menor o igual a la tasa máxima; y
- al disminuir (514), en el primer nodo cliente de la red móvil (202), la tasa de transacción inicial por la tasa de reducción para formar una tasa disminuida cuando:
- 40 el porcentaje de las respuestas que se reciben que es mayor que (512) el umbral de tiempo de reducción de respuesta es mayor que el porcentaje del umbral de reducción de respuestas, y
- la tasa disminuida es mayor o igual a la tasa mínima.
2. El método (500) de acuerdo con la reivindicación 1, en donde la tasa incrementada está asociada con al menos una de una actualización del nodo servidor (208), un cambio en una condición de la red y la eliminación de un segundo nodo cliente de la red móvil (204) en comunicación con el nodo servidor.
- 45 3. El método (500) de acuerdo con la reivindicación 2, en donde la actualización al nodo servidor (208) comprende al menos uno de:
- agregar al menos una CPU al nodo servidor;
- agregar memoria al nodo servidor; y
- 50 agregar al menos una máquina virtual al nodo servidor.
4. El método (500) de acuerdo con la reivindicación 1, en donde la tasa disminuida está asociada con al menos una de una degradación al nodo servidor (208), un cambio en una condición de la red y una adición de un segundo nodo cliente de la red móvil (204) en comunicación con el nodo servidor.
- 55 5. El método (500) de acuerdo con la reivindicación 4, en donde la degradación al nodo servidor (208) comprende agregar al menos uno de:
- eliminar al menos una CPU al nodo servidor;
- eliminar memoria al nodo servidor; y
- 60 eliminar al menos una máquina virtual al nodo servidor.
6. El método (500) de acuerdo con la reivindicación 1, en donde el primer nodo cliente de la red móvil (202) comprende al menos uno de la puerta de enlace de servicio (110), SGW, puerta de enlace de red de paquetes de datos (112), PGW, nodo soporte de puerta de enlace GPRS, GGSN, entidad de gestión de movilidad (106), MME, que sirve al nodo de soporte GPRS, SGSN y Controlador de la Red de Radio, RNC.
- 65

7. El método de acuerdo con la reivindicación 1, en donde el nodo servidor (208) comprende al menos una de la función de políticas y cobros (114), PCRF, sistema de cobro en línea (116), OCS, servidor de autenticación, autorización y contabilización (120), AAA, servidor de abonado doméstico, HSS, función de pasarela de cobro, CGF, sistema de cobro fuera de línea, (118), OFCS.
- 5
8. Un dispositivo informático para ajustar adaptativamente una tasa de transacción en un nodo de la red móvil (202), siendo el dispositivo informático un primer nodo cliente de la red móvil (202), el primer nodo cliente de la red móvil que comprende:
- 10 una memoria; y
un procesador en comunicación con la memoria y que se configura para ejecutar un módulo (420) que se almacena en memoria que se configura para hacer que el procesador:
- 15 reciba los parámetros de configuración que incluyen al menos uno de:
una tasa de transacción inicial, la tasa de transacción inicial asociada con un número inicial de transacciones por segundo que se comunican por el primer nodo cliente de la red móvil (202) a un nodo servidor (208) a través de una red en un momento inicial,
una tasa de incremento, la tasa de incremento asociada con un primer número de transacciones por segundo que indica un aumento en la tasa de transacción inicial,
una tasa de reducción, la tasa de reducción asociada con un segundo número de transacciones por segundo que indica una disminución en la tasa de transacción inicial,
20 una tasa mínima,
una tasa máxima,
un período de evaluación,
un umbral de tiempo de incremento de respuesta asociado con un tiempo de respuesta indicativo de aumentar la tasa de transacción inicial,
25 un umbral de tiempo de reducción de respuesta asociado con un tiempo de respuesta indicativo de disminución de la tasa de transacción inicial,
un porcentaje del umbral de incremento de respuestas asociado con el umbral de tiempo de incremento de respuesta, y
un porcentaje del umbral de reducción de respuestas asociado con el umbral de tiempo de reducción de respuesta;
30 y caracterizado porque el módulo se configura además para hacer que el procesador:
reciba respuestas asociadas con las comunicaciones con el nodo servidor (208) durante un primer período de tiempo asociado con el período de tiempo de evaluación;
determine durante el primer periodo de tiempo, un porcentaje de las respuestas recibidas que es menor que el umbral de tiempo de incremento de respuesta y un porcentaje de las respuestas recibidas que es mayor que el umbral de tiempo de reducción de respuesta;
35 aumente la tasa de transacción inicial por la tasa de incremento para formar una tasa incrementada cuando:
el porcentaje de respuestas recibidas que es menor que el umbral de tiempo de incremento de respuesta que es mayor que el porcentaje del umbral de incremento de respuestas, y
la tasa incrementada es menor o igual a la tasa máxima; y
40 disminuya la tasa de transacción inicial por la tasa de reducción para formar una tasa reducida cuando:
el porcentaje de las respuestas recibidas que es mayor que el umbral de tiempo de reducción de respuesta es mayor que el porcentaje del umbral de reducción de respuestas, y
la tasa disminuida es mayor o igual a la tasa mínima.
- 45 9. El dispositivo informático de acuerdo con la reivindicación 8, en donde la tasa incrementada está asociada con al menos una de una actualización del nodo servidor (208), un cambio en una condición de la red y la eliminación de un segundo nodo cliente de la red móvil (204) en comunicación con el nodo servidor (208).
- 50 10. El dispositivo informático de acuerdo con la reivindicación 9, en donde la actualización al nodo servidor (208) comprende al menos uno de:
agregar al menos una CPU al nodo servidor (208);
agregar memoria al nodo servidor; y
agregar al menos una máquina virtual al nodo servidor (208).
- 55 11. El dispositivo informático de acuerdo con la reivindicación 8, en donde la tasa disminuida está asociada con al menos una de una degradación al nodo servidor (208), un cambio en una condición de la red y una adición de un segundo nodo cliente de la red móvil (204) en comunicación con el nodo servidor (208).
- 60 12. El dispositivo informático de acuerdo con la reivindicación 11, en donde la degradación al nodo servidor (208) comprende agregar al menos uno de:
eliminar al menos una CPU al nodo servidor (208);
eliminar memoria al nodo servidor (208); y
eliminar al menos una máquina virtual al nodo servidor (208).
- 65 13. El dispositivo informático de acuerdo con la reivindicación 8, en donde el primer nodo cliente de la red móvil (202) comprende al menos uno de la puerta de enlace de servicio (110), SGW, puerta de enlace de red de paquetes de

datos (112), PGW, nodo soporte de puerta de enlace GPRS, GGSN, entidad de gestión de movilidad (106), MME, que sirve al Nodo de Soporte del Servicio GPRS, SGSN y Controlador de la Red de Radio, RNC.

- 5 14. El dispositivo informático de acuerdo con la reivindicación 8, en donde el nodo servidor (208) comprende al menos una función de políticas y cobros (114), PCRF, sistema de cobro en línea (116), OCS, servidor de autenticación, autorización y contabilización, 120, AAA, servidor de abonado doméstico, HSS, función de pasarela de cobro, CGF, sistema de cobro fuera de línea (118), OFCS.
- 10 15. Un producto de programa informático que comprende un código de programa informático que, cuando lo ejecuta un procesador de un primer nodo cliente de la red móvil, hace que el procesador implemente el método de cualquiera de las reivindicaciones 1 a la 7.

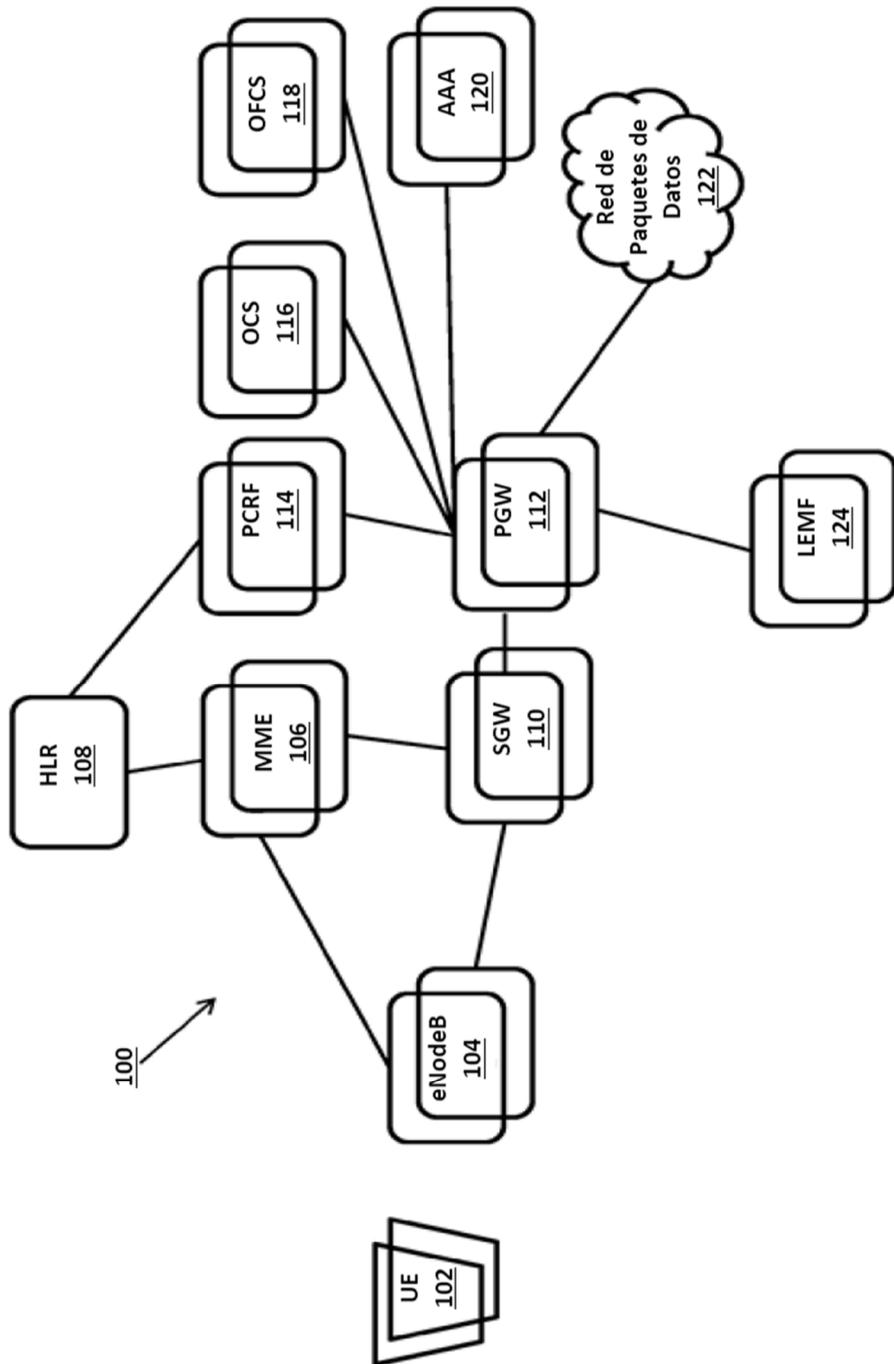


Figura 1

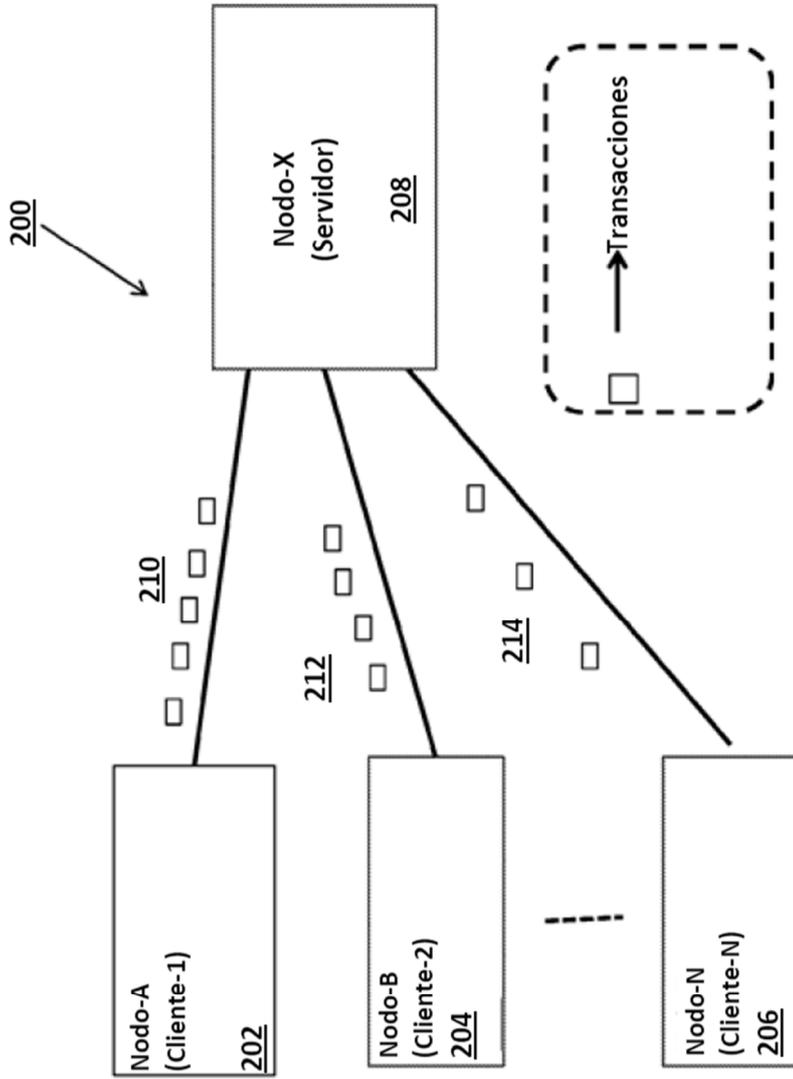


Figura 2

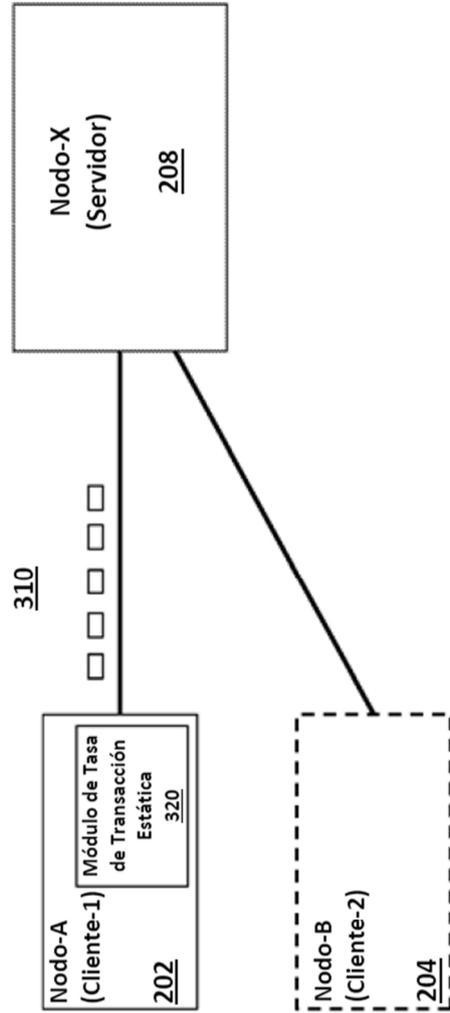


Figura 3 (Técnica Anterior)

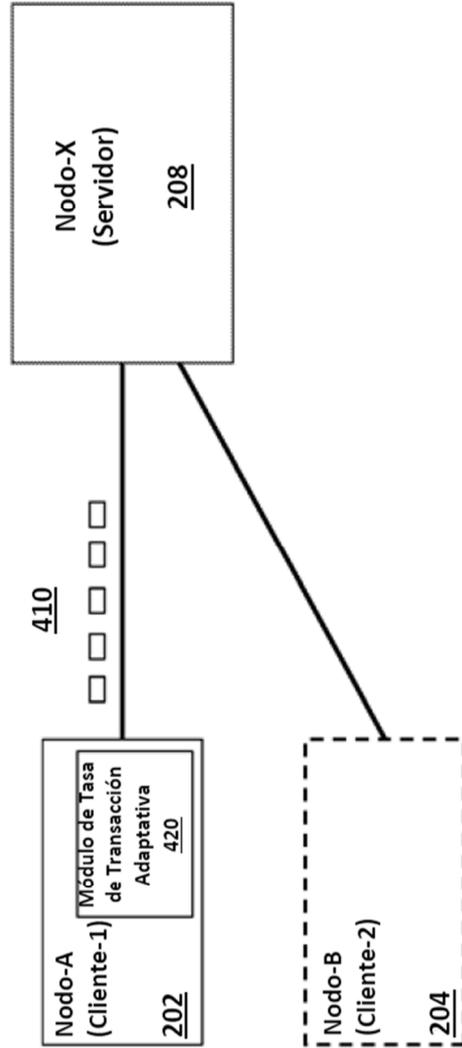


Figura 4

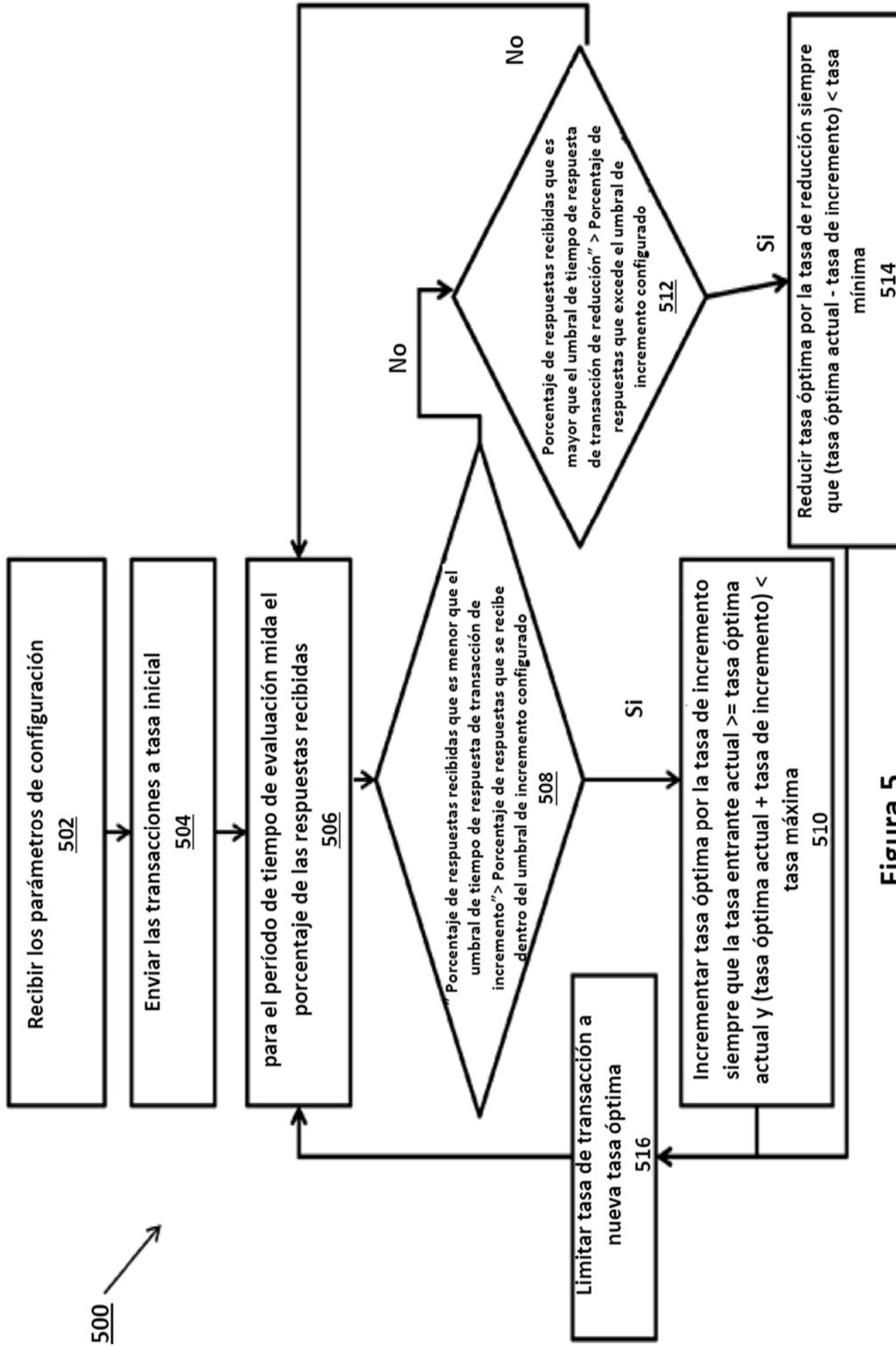


Figura 5