

(19)



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS
ESPAÑA



(11) Número de publicación: **2 665 571**

(51) Int. Cl.:
H04L 29/08 (2006.01)

(12)

TRADUCCIÓN DE PATENTE EUROPEA

T3

(86) Fecha de presentación y número de la solicitud internacional: **15.06.2012 PCT/US2012/042768**

(87) Fecha y número de publicación internacional: **20.12.2012 WO12174444**

(96) Fecha de presentación y número de la solicitud europea: **15.06.2012 E 12729798 (4)**

(97) Fecha y número de publicación de la concesión europea: **10.01.2018 EP 2712480**

(54) Título: **Control de servicio en nube informática y arquitectura de gestión extendida para conectarse al estrato de red**

(30) Prioridad:

17.06.2011 US 201161498337 P

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

26.04.2018

(73) Titular/es:

**HUAWEI TECHNOLOGIES CO., LTD. (100.0%)
Huawei Administration Building, Bantian
Longgang District,
Shenzhen, Guangdong, 518129, CN**

(72) Inventor/es:

LEE, YOUNG

(74) Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 665 571 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 de servicio en nube informática y arquitectura de gestión extendida para conectarse al estrato de red

5 ANTECEDENTES DE LA INVENCIÓN

Los operadores de redes, también a veces referidos como operadores de telecomunicaciones o proveedores de servicios de comunicaciones, que explotan las redes existentes, desean optimizar la utilización de la red para transmitir tráfico, tal como tráfico de Protocolo Internet (IP), a través de la parte física de la red, p.ej., a través de las capas 1 a 5 de red. El tráfico optimizado puede incluir el tráfico para servicios de reproducción triples (p.ej., Vídeo, Voz y/o Datos) y cualquier tipo de datos masivos. En las redes existentes, los servicios de extremo a extremo se suelen establecer por sistemas del tipo Sistemas de Soportes Operativos (OSS) o aplicaciones del servicio de gestión de específicas del proveedor. Los operadores de redes han recomendado dos diferentes escenarios operativos para optimizar la utilización de red el tráfico: optimización de los servicios de redes existentes y permitir nuevos/emergentes servicios de aplicaciones de redes.

El documento US 7636764 publicado con fecha 22-12-2009 da a conocer un método para seleccionar un nodo de red en una nube informática para el almacenamiento de datos.

20 SUMARIO DE LA INVENCIÓN

El alcance de la invención se define por las reivindicaciones adjuntas. La invención se refiere a un sistema según la reivindicación 1.

25 BREVE DESCRIPCIÓN DE LOS DIBUJOS

Para un entendimiento más completo de esta idea inventiva, se hace referencia, a continuación, a la siguiente breve descripción, tomada en relación con los dibujos adjuntos y una descripción detallada, en donde las referencias numéricas similares representan elementos similares

30 La Figura 1 es un diagrama esquemático de una forma de realización de una arquitectura de Optimización de Estrato Cruzado (CSO).

La Figura 2 es un diagrama esquemático de otra forma de realización de una arquitectura de CSO.

35 La Figura 3 es un diagrama esquemático de otra forma de realización de una arquitectura CSO.

La Figura 4 es un diagrama esquemático de otra forma de realización de una arquitectura CSO.

40 La Figura 5 es un diagrama esquemático de otra forma de realización de una arquitectura CSO.

La Figura 6 es un diagrama esquemático de una forma de realización de un mapa de topología de red, a modo de ejemplo.

45 La Figura 7 ilustra una forma de realización de un protocolo de demanda de recursos de pasarela CSCG.

La Figura 8 ilustra una forma de realización de un protocolo de reserva de pasarela CSCG.

50 La Figura 9 es un diagrama de flujo de una forma de realización de un método de demanda de recursos de una pasarela CSCG.

La Figura 10 es un diagrama de flujo de una forma de realización de un método de demanda de reserva de una pasarela CSCG.

55 La Figura 11 es un diagrama esquemático de una forma de realización de una unidad de red.

La Figura 12 es un diagrama esquemático de una forma de realización de un sistema informático de uso general.

60 DESCRIPCIÓN DETALLADA DE LA INVENCIÓN

Debe entenderse, desde el principio, que aunque una puesta en práctica ilustrativa de una o más formas de realización se proporcionan a continuación, los sistemas y/o métodos dados a conocer pueden ponerse en práctica utilizando cualquier número de técnicas, sean actualmente conocidas o en existencia. La idea inventiva no debe limitarse, en forma alguna, a las puestas en práctica ilustrativas, dibujos y técnicas ilustradas a continuación, incluyendo los diseños a modo de ejemplo y las puestas en práctica ilustradas y aquí descritas, sino que pueden modificarse dentro del alcance de las reivindicaciones adjuntas.

El aprovisionamiento y la operación de aplicaciones nuevas/emergentes, tales como la infraestructura de servicio de la denominada nube informática pueden implicar el aprovisionamiento de recursos de procesamiento y/o espacio de memorización sobre múltiples componentes de red (p.ej., servidores). Sobre la base de las necesidades cambiantes de un usuario, los recursos/datos pueden ser objeto de migración entre servidores. La red puede dividirse en un estrato de aplicación y un estrato de red. El estrato de aplicación puede incluir las aplicaciones y servicios puestos en práctica o utilizados a través de la capa de red, y el estrato de red puede incluir el transporte, la red, el enlace y las capas físicas o sus combinaciones. Los servicios que operan en el estrato de aplicación pueden estar constituidos por recursos de servidores disponibles, pero no pueden referirse a recursos y topología de la red. Los servicios que operan en el estrato de red pueden tratarse de recursos y topología de la red, pero no como recursos de servidores. Los servicios que operan en uno u otro estrato pueden ser incapaces de seleccionar un servidor de destino óptimo para la migración de datos/recursos debido a la carencia de información completa, dando lugar a un problema en la selección del servidor (SS). Habida cuenta de la separación estricta entre los estratos, el aprovisionamiento del servicio de manipulación y coordinación tanto en el estrato de aplicación como en el estrato de red es diferente de los servicios tradicionales de manipulación, tales como aprovisionamiento de red de servicios de telecomunicaciones del tipo extremo a extremo.

La idea inventiva consiste en un sistema y método para proporcionar una Optimización de Estrato Cruzado (CSO) para el problema de servicio de servidor SS. Una Pasarela de Control de Servicio de Nube Informática (CSCG) en el estrato de aplicación puede mantener datos de recursos de aplicaciones para varios servidores, que pueden estar situados en varios centros de datos (DCs). Los datos de recursos de aplicaciones pueden comprender la utilización de una memoria de acceso aleatorio (RAM) de un servidor, consumo de energía, utilización de unidad central de procesamiento (CPU), etc. Un origen puede demandar que la pasarela CSCG realice una migración de datos. La pasarela CSCG puede utilizar los datos de recursos de aplicaciones para generar una lista de servidores que pueden ser destinos adecuados para la migración de datos. La pasarela CSCG puede transmitir la lista de destinos a una pasarela de control de red (NCG) en el estrato de red por intermedio de una interfaz de estrato cruzado. La pasarela CSCG puede transmitir también una o más necesidades de recursos de red a la pasarela NCG para su uso en una optimización adicional. Los requisitos de recursos de red pueden comprender: latencia máxima, ancho de banda mínimo/máximo, etc. La pasarela NCG puede utilizar componentes de red, tales como un elemento de cálculo de ruta (PCE), para calcular rutas de red óptimas entre el origen y cada destino así como datos de recursos de red asociados para cada ruta. La pasarela NCG puede filtrar destinos con rutas o enlaces de rutas con datos de recursos de red que no satisfacen las necesidades de recursos de red desde la pasarela CSCG. La pasarela NCG puede enviar un mapa de recursos de red de destinos a la pasarela CSCG junto con datos de recursos de red de ruta asociados para el servidor SS. Además, el mapa de recursos de red puede incluir todos los enlaces en cada ruta y los datos de recursos de red para cada ruta, cada enlace de cada ruta o la topología de red completa. La pasarela CSCG puede utilizar la información procedente de la pasarela NCG para seleccionar un servidor y/o gestionar la migración de datos desde el origen al servidor seleccionado. En otra forma de realización, la pasarela NCG puede recibir la información procedente de la pasarela CSCG, determinar la ruta óptima, reservar la ruta e informar a la pasarela CSCG de la ruta reservada. La pasarela CSCG puede utilizar la información para gestionar la migración de datos al servidor seleccionado por NGW.

Algunos de los términos aquí utilizados y descritos a continuación con respecto a las características de CSO incluyen: recursos de aplicación, servicios de aplicación, pasarela CSCG, recursos de red, una pasarela NCG, un estrato de red y estrato de aplicación. Los recursos de aplicaciones pueden comprender recursos no de red que pueden ser críticos para conseguir la funcionalidad del servicio de aplicación. A modo de ejemplo, los recursos de aplicaciones pueden incluir recursos de cálculo informático y recursos de contenidos tales como memorias caché, espejos, servidores específicos de aplicaciones, máquinas virtuales, memoria, espacio de disco, grandes conjuntos de datos, datos de vídeo, datos de audio, bases de datos y/o otras aplicaciones relacionadas con los recursos. El servicio de aplicación puede ser cualquier aplicación conectada en red ofrecida a una diversidad de clientes. La pasarela CSCG puede ser una entidad de CSO en el estrato de aplicación que es responsable de reunir la carga de recursos de aplicaciones y su utilización, tomar decisiones respecto a la asignación de recursos e interaccionar con la pasarela NCG. La pasarela CSCG puede implantarse en práctica en un procesador de un elemento de red tal como un servidor o una entidad de control de red. La pasarela CSCG puede situarse en un centro de datos y puede ponerse en práctica, o no, en el mismo servidor que una pasarela NCG. Los recursos de red pueden comprender recursos de cualquier capa 3 o capa inferior, tal como ancho de banda, enlaces, rutas, procesamiento de rutas (p.ej., creación, supresión y gestión), bases de datos de redes, cálculo de ruta y los protocolos de enrutamiento y señalización para crear rutas. La pasarela NCG puede ser una entidad de CSO en el estrato de red que es responsable de interaccionar con la pasarela CSCG, iniciar la función de demanda de servicio para la entidad de red de transporte responsable para las funciones de aprovisionamiento, configuración, estimación/cálculo de ruta y otras funciones de gestión/control de redes. La pasarela NCG puede ponerse en práctica en un procesador de un elemento de red tal como un servidor o una entidad de control de red. El estrato de red puede comprender componentes y/o funciones que operan en o bajo la capa de red en el modelo de Interconexión de Sistemas Abiertos (OSI) de siete capas, tal como una función de Comutación de Niveles Multiprotocolo (MPLS), Jerarquía Digital Síncrona (SDH), Red de Transporte Óptico (OTN), Multiplexación por División de Longitud de Onda (WDM). El estrato de aplicación puede comprender componentes y/o funciones que operan en o por encima de la capa de transporte de modelo OSI de siete capas.

La Figura 1 es un diagrama esquemático de una forma de realización de una arquitectura de CSO 100. La arquitectura de CSO 100 puede comprender un estrato de aplicación 110 y un estrato de red 120. El estrato de red 110 puede comprender una pluralidad de servidores 112, que pueden configurarse para poner en práctica o ejecutar aplicaciones para usuarios finales o clientes (no ilustrados). El estrato de red 120 puede comprender una pluralidad de nodos de red 122, tales como puentes, enrutadores y/o conmutadores, para reenviar datos, p.ej., paquetes asociados con las aplicaciones. Los servidores 112 pueden estar situados en un centro de datos y los nodos de redes 122 pueden estar situados en una red acoplada al centro de datos. Los servidores 112 pueden comunicarse con los nodos de red 122 para permitir el servicio a las aplicaciones de usuarios y el reenvío o transporte de los datos asociados. La arquitectura de CSO 100 puede ponerse en práctica para optimizar las diferentes operaciones de los servidores 112 y los nodos de red 122.

En una forma de realización, los centros de datos utilizados para proporcionar servicios de aplicaciones, tales como cálculo de nube informática y otros servicios de la nube informática, en el estrato de aplicación 110 para los usuarios finales pueden distribuirse geográficamente alrededor del estrato de red 120. De este modo, numerosas decisiones tomadas en el control y gestión de servicios de aplicaciones, tales como dónde instanciar otra instancia de servicio o a qué centro de datos se asigna un nuevo cliente, pueden tener un impacto importante sobre el estado operativo de la red. Las capacidades y el estado de la red pueden tener también un impacto sobre el rendimiento de las aplicaciones.

Actualmente, las decisiones de aplicaciones pueden tomarse con poca o ninguna información respecto a la red subyacente utilizada para prestar dichos servicios. Por consiguiente, dichas decisiones pueden ser sub-óptimas desde la utilización de recursos de redes y aplicaciones y desde el logro de los objetivos de Calidad de Servicio (QoS). Una arquitectura de CSO, tal como la arquitectura de CSO 100 puede proporcionar un método y un sistema para coordinar la asignación de recursos entre el estrato de aplicación 110 y el estrato de red 120, p.ej., en el contexto de cálculo de la nube informática y redes de centro de datos. A modo de ejemplo, la arquitectura de CSO 100 puede soportar la demanda de un estrato de red 110 respecto a aplicación, aprovisionamiento conjunto entre aplicación y red y/o reasignación conjunta de recursos sobre anomalía operativa en la aplicación y la red a la vez. La arquitectura de CSO 100 puede proporcionar también una red orientada a la aplicación y una aplicación orientada a la red y una capacidad de equilibrio de carga global.

Algunos de los objetivos para optimizar las operaciones y/o interacciones entre el estrato de aplicación 110 y el estrato de red 120, p.ej., entre los servidores 112 y los nodos de red 122, pueden incluir la mejora de las capacidades de red, topología, aprovisionamiento, control de utilización, control de anomalías o sus combinaciones.

A modo de ejemplo, la arquitectura de CSO 100 puede mejorar el intercambio de capacidades de redes o demanda de aplicación/información de recursos, topología y/o información relacionada con la denominada ingeniería de tráfico entre las capas (virtualización/abstracción) o ambas funciones a la vez. La arquitectura de CSO 100 puede mejorar también la iniciación de la instanciación del servicio de aplicación para la red con intercambio de perfiles (aprovisionamiento), intercambio de información de aplicación/congestión de la red/fallos (supervisión) o ambas a la vez.

La Figura 2 es un diagrama esquemático de otra forma de realización de una arquitectura de CSO 200. La arquitectura de CSO 200 puede comprender un estrato de aplicación 210 y un estrato de red 220. El estrato de aplicación 210 puede comprender una pluralidad de servidores 212 y el estrato de red 220 puede comprender una pluralidad de nodos de red 222, que pueden ser prácticamente similares a los servidores 112 y los nodos de red 122, respectivamente. La arquitectura de CSO 200 puede comprender, además, una interfaz de CSO que permite interacciones y/o comunicaciones entre los servidores 212 y/o otros componentes (no ilustrados) del estrato de aplicación 210 y los nodos de red 222 y/o otros componentes (no ilustrados) del estrato de red 220. La interfaz de CSO puede ser una interfaz abierta entre los dos estratos y puede permitir algunas de las características de CSO descritas a continuación. En el estrato de aplicación 210, la interfaz abierta puede permitir la identificación de cliente/consumidor de algún tipo, p.ej., dirección de Protocolo Internet (IP), tipos de servidores e identificación, flujos de datos de aplicaciones y requisitos de calidad de servicio (QoS) que pueden ser estadísticos por naturaleza y variar en el transcurso del tiempo y/o condiciones de carga y fallos operativos de servidores. En el estrato de red 220, la interfaz abierta puede permitir el intercambio de topología de red, localizaciones de clientes y servidores dentro de esa topología, capacidades de red y capacidades con respecto a la calidad de servicio QoS, ancho de banda, información de latencia y/o otras características relacionadas con la red, condiciones de carga y fallos operativos de la red o sus combinaciones.

La Figura 3 es un diagrama esquemático de otra forma de realización de una arquitectura de CSO 300. La arquitectura de CSO 300 puede comprender un estrato de aplicación 310 y un estrato de red 320. El estrato de aplicación 310 puede comprender una pluralidad de servidores 312 y el estrato de red 320 puede comprender una pluralidad de nodos de red 322, que pueden ser prácticamente similares a los servidores 112 y los nodos de red 122, respectivamente. La arquitectura de CSO 300 puede comprender, además, un interfaz de CSO que puede establecerse entre una pasarela CSCG 314 en el estrato de aplicación 310 y una pasarela NCG 324 en el estrato de red 320.

La pasarela CSCG 314 puede configurarse para acceder a procesos y datos relacionados con aplicaciones (en el estrato de aplicación 310), comunicarse con la pasarela NCG 324 (por intermedio de la interfaz de CSO) y proporcionar abstracción/virtualización informativas y limitaciones de acceso a entidades externas (fuera del estrato de aplicación 310) incluyendo las entidades del estrato de red 320. La pasarela NCG 324 puede configurarse para

5 acceder a datos relacionados con la red (en el estrato de red 320), comunicarse con la pasarela CSCG 314 (por intermedio de la interfaz CSO), comunicarse con los procesos de redes tales como control de admisión, reserva de recursos y/o procesamiento de conexiones y proporcionar abstracción/virtualización informativa y limitaciones de acceso a entidades externas (fuera del estrato de red 320) incluyendo las entidades del estrato de aplicación 310. Además, la pasarela CSCG 314 y la pasarela NCG 324 pueden comunicarse con los servidores 312 y los nodos de

10 red 322, respectivamente.

La Figura 4 es un diagrama esquemático de otra forma de realización de una arquitectura de CSO 400. La arquitectura de CSO 400 puede comprender un plano de usuario 401, un estrato de aplicación 410 y un estrato de red 420. El estrato de aplicación 410 puede comprender un plano de aplicación 412 (p.ej., en un centro de datos (DC)), y una pasarela CSCG 414, que puede comunicarse con el plano de aplicación 412 por intermedio de una interfaz de plano de aplicación (API). La pasarela CSCG 414 en el estrato de aplicación 410 puede comunicarse también con el plano del usuario 401 por intermedio de una interfaz de aplicación-usuario (UAI). El estrato de red 420 puede comprender un plano de servicio 440, un plano de gestión 450, un plano de control 460 y un plano de transporte 470. El plano de transporte 470 puede soportar la tecnología de transporte de la infraestructura de red correspondiente, tal como para Perfil de Transporte-Conmutación Etiquetas Multiprotocolo (MPLS-TP), Red de Transporte Óptico (ONT) o Red Óptica Conmutada de Longitud de Onda (WSON).

El plano de servicio 440 puede configurarse para permitir las comunicaciones entre el plano de aplicación 412 en el estrato de aplicación 410 y el plano de gestión 450, el plano de control 460, y el plano de transporte 470 en el estrato de red 420, p.ej., en una manera optimizada basada en CSO. El plano de servicio 440 puede comunicarse con el plano de aplicación 412 por intermedio de una interfaz de aplicación-servicio (ASI), el plano de gestión 450 por intermedio de una interfaz de plano de gestión-servicio (SMI) y el plano de control 460 por intermedio de una interfaz de plano de control-servicio (SCI). El plano de transporte 470 puede comunicarse con el plano de gestión 450 y el plano de control 460 y de este modo, el plano de servicio 440 por intermedio de una interfaz de control de conexión (CCI).

El plano de servicio 440 puede proporcionarse por una parte o entidad (p.ej., un proveedor) que puede ser independiente del plano del usuario 401, el estrato de aplicación 410 y el estrato de red 420. A modo de ejemplo, el estrato de aplicación 410 y el estrato de red 420 pueden gestionarse por diferentes entidades o proveedores, y el plano de servicio 440 puede gestionarse por una tercera parte. El plano de servicio 440 puede comprender una pasarela NCG 422, y una pluralidad de bases de datos de servicio de red 424, que pueden comprender una Base de Datos de Ingeniería de Tráfico (TED), una Base de Datos (DB) de Enrutamiento de Red (NR), una base de datos Config DB, una Tabla de Enrutamiento Múltiple (MRT), una Base de Información de Gestión (MIB) y/o otras bases de datos de conexiones de redes. Las bases de datos de servicio de redes 424 pueden comprender al menos alguna información que puede copiarse a partir de bases de datos similares en los planos de red. La pasarela NCG 422 puede comunicarse con la pasarela CSCG 414, y de este modo, el plano de aplicación 412, por intermedio de ASI, el plano de gestión 450 por intermedio de la interfaz SMI y el plano de control 460 por intermedio de la interfaz SCI. La pasarela NCG 422 puede acceder también a la información en las bases de datos de servicio de red 424 cuando sea necesario para permitir el flujo de tráfico y comunicaciones entre los diferentes planos y estratos. Formas de realización adicionales de las arquitecturas de CSO pueden incluirse en la publicación de patente de Estados Unidos 2012/0054346 por Y. Lee et al, titulada "Método y Sistema para Optimización de Estratos Cruzados en Redes de Transporte-Aplicaciones", que se incorpora aquí por referencia.

50 La Figura 5 ilustra otra forma de realización de una arquitectura de CSO 500. La arquitectura de CSO 500 puede comprender prácticamente los mismos componentes que las arquitecturas de CSO 100-400 en una configuración diferente. La arquitectura de CSO 500 puede comprender una pasarela CSCG 514 en comunicación con servidores 511-513 y la pasarela NCG 524. La pasarela NCG 524 puede estar en comunicación con PCE 530 y un elemento de red (NE) 540. La pasarela CSCG 514 y los servidores 511-513 pueden situarse en el estrato de aplicación y la pasarela NCG 524, el elemento PCE 530 y el elemento NE 540 puede situarse en el estrato de red.

55 La pasarela CSCG 514 puede obtener datos de recursos de aplicaciones a partir de los servidores 511-513 sobre una base periódica y/o sobre una base según se necesite, y puede mantener dichos datos de recursos de aplicaciones localmente para uso en el servidor SS. La pasarela CSCG 514 puede comunicarse con los servidores 511-513 utilizando el Protocolo de Transferencia de Hipertexto (HTTP). La pasarela CSCG 514 puede enviar demás de estimación y/o reserva de ruta a la pasarela NCG 524 utilizando el protocolo de Optimización de Tráfico de Capa de Aplicación (ALTO) según se establece en el documento de Internet Engineering Task Force (IETF) draft-ietf-ALTO-protocol-11, que se incorpora aquí por referencia. La pasarela NCG 524 puede realizar, a su vez, demandas de cálculo de ruta de PCE 530 utilizando el protocolo de comunicaciones PCE (PCEP), según se describe en el documento de IETF Demanda de Comentario (RFC) 5440, que se incorpora aquí por referencia. Las RFCs 5088 y 5089, que se incorporan aquí ambas por referencia, describen cómo descubrir un PCE adecuado desde la perspectiva de la pasarela NCG 524 específica. El PCE 530 puede proporcionar rutas candidatos que cumplan las

limitaciones de recursos de redes específicas tal como conectividad (p.ej., punto a punto P-P), punto a multipunto (P-MP), etc.) y algunos parámetros de QoS (p.ej., latencia) así como requisitos de ancho de banda para la conectividad. Las rutas calculadas por el PCE 530 pueden ser una estimación de las rutas desde la aplicación basadas en el más reciente enlace de red y datos de tráfico de nodos, que pueden memorizarse por el PCE 530 en una Base de Datos de Ingeniería de Tráfico (TED).

- 5 El PCE 530 puede responder con las rutas candidatos y datos de recursos de red relacionados. Una vez que se hayan encontrado las rutas, entonces la pasarela NCG 524 puede responder con las rutas resultantes y los datos de recursos de red a la pasarela CSCG 514. Dichas rutas pueden enviarse a la pasarela CSCG 514 en el formato de origen-destino o como una lista de enlaces de ruta. Si la aplicación requiere una reserva de ancho de banda de una ruta calculada, en tal caso, la pasarela NCG 522 puede proseguir, además, 10 con el proceso de aprovisionamiento de rutas bien sea mediante un proceso de configuración de gestión de red, bien sea mediante una funcionalidad del plano de control. El proceso de aprovisionamiento puede iniciarse mediante la comunicación con el NE 540. NE 540 puede ser un commutador virtual, un enrutador de red, un controlador del plano de control o entidad similar.

- 15 Dependiendo de la forma de realización, la pasarela CSCG 514 puede seleccionar y transmitir cualquiera o la totalidad de las direcciones de servidores 511-513 y cualesquier restricciones de recursos de redes deseadas a la pasarela NCG 524 para el cálculo de rutas y la pasarela NCG 524 puede responder con rutas que cumplan los requisitos de la pasarela CSCG 514, permitiendo a la pasarela CSCG 514 realizar SS. Como alternativa, la pasarela CSCG 514 puede transmitir algunos o la totalidad de los datos de recursos de aplicaciones conocidos para la pasarela CSCG 514, junto con necesidades de recursos de aplicaciones y/o redes, hacia la pasarela NCG 524, lo 20 que permite a NCG 524 realizar SS e informar a la pasarela CSCG 514 de los resultados obtenidos.

25 La Figura 6 es un diagrama esquemático de una forma de realización de un mapa de topología de red, a modo de ejemplo 600. El mapa de topología de red 600 se presenta simplemente para proporcionar datos ejemplo para uso en el método descrito a continuación. El mapa de topología de red 600 puede comprender un Origen 601 conectado al Destino 621-623 por intermedio de los elementos de red NEs 610-613. Las conexiones entre los componentes del mapa de topología de red 600 pueden ser enlaces de red. El origen 601 puede comprender un servidor y puede requerir una migración de datos a un servidor en los Destinos 621-623. El origen 601 puede conectarse al elemento de red NE 610. El elemento NE 610 puede conectarse a NE 611-613. NE 613 puede conectarse a NE 610-612 y los destinos 30 622-623, NE 612 puede conectarse a los destinos 621-622.

35 La Figura 7 ilustra una forma de realización de un protocolo de demanda de pasarela CSCG 700. Una pasarela CSGC puede demandar 701 datos de recursos de aplicación a partir de una pluralidad de servidores en el estrato de aplicación. Cada servidor puede responder 702 con los datos de recursos de aplicación locales del servidor. Los datos de recursos de aplicación pueden comprender la utilización de memoria RAM del servidor, consumo de energía, utilización de CPU, etc. La demanda-respuesta 701-702 puede realizarse utilizando funciones de adquisición/registro de HTTP. La pasarela CSGC puede memorizar los datos de recursos de aplicación para su uso posterior. En un momento posterior, un origen (p.ej., Origen 601) puede enviar una demanda de migración de datos 710 a la pasarela CSCG. Como alternativa, la demanda de migración de datos 710 puede enviarse a la pasarela CSCG 40 desde otro componente de red en representación del origen. La pasarela CSCG puede comparar los más recientes datos de recursos de aplicación con una necesidad de recursos de aplicación y determinar los destinos que son capaces de recibir los datos del origen y realizar las tareas relacionadas (p.ej., Destinos 621-623). La pasarela CSCG puede determinar una necesidad de recursos de red para rutas de redes entre el origen los destinos. La pasarela CSCG puede enviar una demanda de recursos de red 711 a la pasarela NCG. La demanda de recursos de red 711 puede comprender la dirección de red del Origen 601, la necesidad de recursos de red y una lista de direcciones de destino que puede comprender destinos potenciales de migración de datos (p.ej., Destinos 621-623). La demanda de recursos de red 711 puede ser de tipo sumario para indicar que la pasarela CSGW desea recibir información de recursos de red sobre cada enlace de cada ruta o del tipo gráfico para indicar que el CSGW 45 desea recibir información de recursos de red sobre cada enlace de cada ruta. La pasarela NCG puede transmitir una demanda de cálculo de ruta 712 a un PCE, demandar una ruta óptima entre cada par de origen-destino e información de recursos de red sobre cada ruta óptima. El PCE puede transmitir una respuesta de PCE 713 con las rutas y la información de recursos de red relacionada. La pasarela NCG puede recibir la respuesta 713 y rechazar cualquier destino con una ruta o enlace de ruta con información de recursos de red que no satisfagan la necesidad de recursos de red de la CSCG. La pasarela NCG puede crear un mapa de recursos de red que comprende el 50 origen, los destinos, rutas de red, y/o los enlaces que se relacionan con la demanda de recursos de la pasarela CSCG 711 y que satisfacen la necesidad de recursos de red de la CSCG. El mapa de recursos de red puede comprender también la información de recursos de red asociada con cada ruta, cada enlace de ruta o la topología de red completa. El mapa de recursos de red puede comprender información de recursos de red en forma no abstracta o abstracta. La forma no abstracta puede comprender información de recursos de red real, mientras que la forma abstracta puede comprender información de recursos de red en una forma lógica/abstracta con datos no esenciales o duplicativos extraídos o combinados para reducir la complejidad. Si la demanda de recursos 711 fuera de tipo sumario, el mapa de recursos de red puede comprender información del nivel de ruta, y si la demanda de recursos 711 fuera de tipo gráfico, el mapa de recursos de red puede comprender información del nivel del enlace. La pasarela NCG 55 puede enviar una respuesta 715 a la pasarela CSCG con el mapa de recursos de red. La CSCG puede seleccionar un servidor de destino a partir del mapa de recursos de red sobre la base de la información de recursos de red incluida. La pasarela CSCG puede enviar entonces una respuesta 716 a la demanda de migración 60 65

de datos 710. La respuesta 716 puede comprender el servidor de destino seleccionado. El origen puede transferir 717 datos al servidor de destino según lo necesite.

La Figura 8 ilustra una forma de realización de un protocolo de reserva de pasarela CSCG 800. Una pasarela CSGC 5 puede demandar 801 datos de recursos de aplicación a partir de una pluralidad de servidores en el estrato de aplicación. Los servidores pueden responder 802 con sus datos de recursos de aplicación locales. Un origen o entidad similar puede enviar una demanda de migración de datos 810 a la pasarela CSCG. La pasarela CSCG 10 puede enviar una demanda de reserva 811 a la pasarela NCG. La demanda de reserva 811 puede comprender prácticamente la misma información que la demanda de recursos 711, pero puede incluir datos que indican que la pasarela NCG 15 debe completar la selección del servidor de destino y reservar una ruta de migración de datos adecuada. La pasarela NCG puede enviar una demanda de cálculo de ruta 812 a un PCE y recibir una respuesta de cálculo de ruta 813 en prácticamente la misma manera que 712 y 713 respectivamente. La pasarela NCG puede seleccionar un destino utilizando la información de recursos de ruta procedente de PCE y las necesidades de recursos de red procedentes de la pasarela CSCG. Como alternativa, la demanda de reserva 811 puede 20 comprender, también datos de recursos de aplicación y/o necesidades de recursos de aplicaciones y la pasarela NCG puede seleccionar un destino sobre la base de la información de recursos de aplicación y la información de recursos de red. Una vez que se seleccione una ruta, la pasarela NCG puede enviar una demanda de reserva de ruta 814 a un elemento de red NE capaz de reservar la ruta, a modo de ejemplo, el elemento NE 540 y/o el extremo de cabecera de la ruta o el nodo final. El elemento de red NE 25 puede reservar la ruta y enviar una confirmación 815 a la pasarela NCG. La pasarela NCG puede enviar una confirmación 816 a la pasarela CSCG indicando que ruta y/o la información de recursos de red de las rutas. La pasarela CSCG puede enviar una respuesta 817 al origen indicando el destino y la ruta. El origen puede iniciar la migración de los datos 818 por intermedio del elemento de red NE o por intermedio de un nodo de final de cabeza de ruta. El elemento de red NE puede transferir entonces 819 los datos a lo largo de la ruta hacia el destino.

La Figura 9 es un diagrama de flujo de una forma de realización de un método de demanda de recursos de pasarela CSCG 900, que puede utilizar en el protocolo de demanda de CSCG 700. Según se describió con anterioridad, la pasarela CSCG 30 puede obtener 910 datos de recursos de aplicación a partir de los servidores de centros de datos. Un origen puede iniciar entonces 920 una transferencia de migración de datos mediante la señalización de la CSCG. La pasarela CSCG 35 puede transmitir entonces una demanda de recursos de red 930 a la pasarela NCG con un origen (p.ej., Origen 601), la lista de direcciones de destino que comprende potenciales servidores de destino (p.ej., Destinos 621-623) y las necesidades de recursos de red. La pasarela NCG puede enviar demás de cálculos de ruta 940 a un PCE para cada par de origen-destino. El PCE puede responder 950 a la pasarela NCG con rutas para cada par de origen-destino.

La pasarela NCG 35 puede crear 960 un mapa de recursos de red rechazando cualquier destino con una ruta o enlaces individuales con información de recursos de red que no satisfagan las necesidades de recursos de red de la pasarela CSCG. A modo de ejemplo, la pasarela CSCG puede demandar que cada ruta tenga una latencia máxima de 20 y una disponibilidad de ancho de banda mínima de 5 en unidades adecuadas. Las rutas recibidas 950 desde el PCE 40 pueden asociarse con la información de recursos de red siguiente:

Ruta	Ancho de banda	Latencia
Origen 601-Destino 621	10	20
Origen 601-Destino 622	1	10
Origen 601-Destino 623	5	30

Sobre la base de los datos precedentes, la pasarela NCG 45 puede rechazar o filtrar para selección, el Destino 622 puesto que la ruta Origen 601-Destino 622 tiene una disponibilidad de ancho de banda de uno, lo que es insuficiente para satisfacer la necesidad de ancho de banda mínima de cinco. La pasarela NCG 50 puede rechazar también el Destino 623 puesto que la ruta Origen 601-Destino 623 tiene una latencia de 30, que excede y no cumple el requisito de latencia máxima de 30. Si la demanda de recursos 930 fue de tipo sumario, la pasarela NCG 55 puede crear 960 un mapa de recursos de red que comprende la información de nivel de ruta que incluye el Origen 601, los destinos originalmente recibidos por la pasarela NCG que no fueron seleccionados, en este caso, el Destino 621 y la información de recursos de red asociada con la ruta o rutas aceptadas, en este caso una disponibilidad de ancho de banda de 10 y una latencia de 20. La pasarela NCG 60 puede filtrar los destinos sobre la base de los recursos acumulativos de la ruta, los recursos de cada enlace en la ruta o ambos a la vez. Si la demanda de recursos 930 era del tipo gráfico, la pasarela NCG puede crear un mapa de recursos de red que comprenda cada enlace de cada ruta que no fue filtrada para selección y la información de recursos de red asociada con cada enlace aceptado. A modo de ejemplo, la ruta óptima entre el Origen 601 y el Destino 621 puede pasar a través de los elementos de red NE 610 y NE 612. El enlace de Origen 601-NE 610 puede tener una disponibilidad de ancho de banda de 15 y una latencia de 7, el enlace de NE 610-NE 612 puede tener una disponibilidad de ancho de banda de 10 y una latencia de 8 y el enlace de NE 612-Destino 621 puede tener una disponibilidad de ancho de banda de 14 y una latencia de 5, en cuyo caso la ruta de Origen 601-Destino 621 tiene una disponibilidad de ancho de banda mínima de extremo a extremo de 10 y una latencia de 20. Si la demanda de recursos 930 era del tipo gráfico, la pasarela NCG 65 puede

crear 960 un mapa de recursos de red con la información siguiente:

Enlace	Ancho de banda	Latencia
Origen 601-NE 610	15	7
NE 610-NE 612	10	8
NE 612-Destino 621	14	5

5 La pasarela NCG puede enviar una respuesta 970 a la pasarela CSCG que comprende el mapa de recursos de red con la dirección de destino, rutas y/o enlaces de rutas con los datos de recursos de red asociados. La pasarela CSCG puede facilitar entonces 980 una migración de datos entre el Origen 601 y el Destino 621.

10 La Figura 10 es un diagrama de flujo de una forma de realización de un método de demanda de reserva de CSCG 1000 que puede utilizarse en el protocolo de reserva de CSCG 800. Según se describió con anterioridad, la pasarela CSCG puede obtener 1010 datos de recursos de aplicación a partir de servidores de centros de datos. Un origen puede iniciar entonces 1020 una migración de datos. La pasarela CSCG puede transmitir 1030 una demanda de reserva a la pasarela NCG. La demanda de reserva puede comprender la dirección origen, la lista de direcciones de destino y las necesidades de recursos de red según se describió con anterioridad. La pasarela NCG puede enviar 1040 una demanda de cálculo de ruta a un PCE, y el PCE puede enviar 1050 rutas de origen-destino calculadas e información de recursos de red relacionada a la pasarela NCG. La pasarela NCG puede seleccionar un destino y reservar 1060 la ruta de red asociada mediante su comunicación con un elemento de red NE. A la recepción de la confirmación de que se ha reservado la ruta de red, la pasarela NCG puede enviar una confirmación 1070 a la pasarela CSCG con la ruta reservada.

15 20 La Figura 11 ilustra una forma de realización de una unidad de red 1100, que puede ser cualquier dispositivo que transporta y procesa datos a través de la red, y puede comprender un elemento de red NE en las redes 100-600. La unidad de red 1100 puede comprender uno o más puertos de entrada o unidades 1110 acopladas a un receptor (Rx) 1112 para recibir señales y tramas/datos desde otros componentes de la red. La unidad de red 1100 puede comprender una unidad lógica 1120 para determinar a qué componentes de red enviar datos. La unidad lógica 1120 puede ponerse en práctica utilizando hardware, software o ambos a la vez. La unidad de red 1100 puede comprender también uno o más puertos o unidades de salida 1130 que están acopladas a un transmisor (Tx) 1132 para transmitir señales y tramas/datos a los otros componentes de la red. El receptor 1112, la unidad lógica 1120 y el transmisor 1132 pueden poner en práctica o soportar también el protocolo de demanda de recursos 700, el protocolo de reserva 800, el método de demanda de recursos 900 y el método de demanda de reserva 1000. Los componentes de la unidad de red 1100 pueden disponerse según se ilustra en la Figura 11.

25 30 Los componentes de red anteriormente descritos pueden ponerse en práctica en cualquier componente de red de uso general, tal como un ordenador o componente de red con potencia de procesamiento suficiente, recursos de memoria y capacidad de rendimiento de la red para gestionar la carga de trabajo necesaria que se le aplica. La Figura 12 ilustra un componente de red de uso general típico 1200 adecuado para poner en práctica una o más formas de realización de los componentes aquí dados a conocer. El componente de red 1200 incluye un procesador 1202 (que puede referirse como una unidad central de procesador o CPU) que está en comunicación con dispositivos de memoria incluyendo la memoria secundaria 1204, la memoria de solamente lectura (ROM) 1206, una memoria RAM 1208, dispositivos de entrada/salida (I/O) 1210 y dispositivos de conectividad de red 1212. El procesador 1202 puede ponerse en práctica como uno o más circuitos integrados de CPU o puede ser parte de uno o más circuitos integrados específicos de la aplicación (ASICs).

35 40 45 La memoria secundaria 1204 suele estar constituida por unas o más unidades de disco o unidades de cinta y se utiliza para la memorización no volátil de datos y como un dispositivo de memorización de datos de desbordamiento de capacidad si la memoria RAM 1208 no es suficientemente amplia para mantener todos los datos de trabajo. La memoria secundaria 1204 puede utilizarse para memorizar programas que se cargan en la memoria RAM 1208 cuando dichos programas se seleccionan para su ejecución. La memoria ROM 1206 se utiliza para memorizar instrucciones y quizás datos que sean objeto de lectura durante la ejecución del programa. La memoria ROM 1206 es un dispositivo de memoria no volátil que suele tener una pequeña capacidad de memoria relativa a la mayor capacidad de memoria de la memoria secundaria 1204. La memoria RAM 1208 se utiliza para memorizar datos volátiles y quizás, para memorizar instrucciones. El acceso a ambas memorias ROM 1206 y RAM 1208 suele ser más rápido que a la memoria secundaria 1204. El componente de red de uso general 1200 puede configurarse para poner en práctica o soportar el protocolo de demanda de recursos 700, el protocolo de reserva 800, el método de demanda de recursos 900 y/o el método de demanda de reserva 1000.

50 55 60 Al menos una forma de realización se da a conocer y las variaciones, combinaciones y/o modificaciones de las formas de realización y/o características de las formas de realización realizadas por un experto en esta técnica están dentro del alcance de la idea inventiva. Formas de realización alternativas que resulten de las funciones de combinar, integrar y/o omitir características de las formas de realización están también dentro del alcance de la idea inventiva. En donde se establecen expresamente limitaciones o márgenes numéricos, dichas limitaciones o

márgenes expresos deben entenderse que incluyen márgenes iterativos o limitaciones de magnitud similar que caen dentro de las limitaciones o márgenes que se indican expresamente (p.ej., desde aproximadamente 1 aproximadamente 10 incluye 2, 3, 4, etc.; mayor que 0,10 incluye 0,11, 0,12, 0,13, etc.). A modo de ejemplo, siempre que se da a conocer un margen numérico con un límite inferior R_l y un límite superior R_u cualquier número que cae dentro del margen se da a conocer concretamente. En particular, los siguientes números dentro del margen se dan a conocer específicamente: $R = R_l + k * (R_u - R_l)$ en donde k es una variable cuyo margen es desde el 1 por ciento al 100 por ciento con un incremento del 1 por ciento, esto es, k es 1 por ciento, 2 por ciento, 3 por ciento, 4 por ciento, 7 por ciento, ..., 70 por ciento, 71 por ciento, 72 por ciento, ..., 97 por ciento, 96 por ciento, 97 por ciento, 98 por ciento, 99 por ciento o 100 por ciento. Además, cualquier margen numérico definido por dos números R según se definió con anterioridad se da a conocer también de forma específica. A no ser que se indique de otro modo, el término "aproximadamente" significa $\pm 10\%$ del número que le sigue. El uso del término "opcionalmente" con respecto a cualquier elemento de una reivindicación significa que el elemento es requerido, o de forma alternativa, el elemento no es requerido, estando ambas alternativas dentro del alcance de la reivindicación. El uso de términos más amplios tales como comprende, incluye y teniendo, debe entenderse que proporcionan apoyo a términos menos amplios tales como constituido por, constituido esencialmente por y comprendido prácticamente por. En consecuencia, el alcance de protección no está limitado por la descripción anteriormente establecida sino que está definido por las reivindicaciones que siguen, incluyendo ese alcance todos los equivalentes del contenido de las reivindicaciones. Todas y cada una de las reivindicaciones se incorpora como elemento adicional dado a conocer en la especificación y las reivindicaciones son formas de realización de la presente invención. El examen de una referencia en la idea inventiva no es una admisión de que se trate de una técnica anterior, en particular, cualquier referencia que tenga una fecha de publicación después de la fecha de prioridad de esta solicitud de patente. La idea inventiva de todas las patentes, solicitudes de patentes y publicaciones citadas en la invención se incorporan aquí por referencia, en la medida en que proporcionan datos a modo de ejemplo, procedimientos y otros detalles que son complementarios para la idea inventiva.

Los presentes ejemplos han de considerarse como ilustrativos y no restrictivos, y la intención no está limitada a los detalles aquí proporcionados. A modo de ejemplo, los diversos elementos o componentes pueden combinarse o integrarse en otro sistema o algunas características pueden omitirse o no ponerse en práctica.

30

REIVINDICACIONES**1. Un sistema que comprende:**

- 5 una pasarela de control de servicio de la denominada nube informática, CSCG (314, 414), situada en un estrato de aplicación (310, 410), y una pasarela de control de red, NCG (324, 422), situada en un estrato de red (320, 420),
 en donde la pasarela CSCG (314, 414) está configurada para acoplarse a la pasarela NCG (324, 422), y
- 10 en donde la pasarela CSCG (314, 414) está configurada para transmitir una dirección de origen, una lista de direcciones de destino y una primera necesidad de recursos de red a la pasarela NCG (324, 422), en donde la pasarela CSCG (314, 414) está configurada además para recibir un mapa de recursos de red desde la pasarela NCG (324, 422), donde el mapa de recursos de red comprende un valor de recursos de red de una ruta de red entre la dirección de origen y una dirección de destino contenida en la lista de direcciones de destino, en donde la pasarela CSCG (314, 414) está configurada además para obtener datos de recursos de aplicación desde una pluralidad de servidores, correspondientes a la lista de direcciones de destino, y en donde la pasarela CSCG (314, 414) está configurada además para seleccionar un servidor de destino para una demanda de aplicación basada en un análisis de los datos de recursos de aplicación y el mapa de recursos de red, mediante lo cual la demanda de aplicación es una demanda para la migración de datos desde la dirección de origen a uno de la pluralidad de servidores y;
- 15 20 en donde la lista de direcciones de destino comprende potenciales servidores de destino; en donde si un primer valor de recursos de red de una ruta de red entre una dirección de origen y una dirección de destino satisface la primera necesidad de recursos de red, en ese caso el mapa de recursos de red comprende la dirección de destino correspondiente al servidor de destino seleccionado por la pasarela CSCG;
- 25 25 en donde los datos de recursos de aplicación comprenden la utilización de una memoria de acceso aleatorio, RAM, consumo de energía, utilización de unidad central de procesamiento, CPU, de la pluralidad de servidores (312);
- 30 30 en donde la pasarela NCG (324, 422) está configurada para recibir una comunicación de la pasarela CSCG (314, 414) que comprende la dirección de origen, la lista de direcciones de destino, que comprende una pluralidad de direcciones de destino, y una necesidad de recursos de red, en donde la pasarela NCG (324, 422) está configurada además para transmitir una demanda de cálculo de ruta a un elemento de cálculo de ruta, PCE, situado en el estrato de red (320, 420), y en donde la demanda de cálculo de ruta comprende la dirección de origen y la pluralidad de direcciones de destino, en donde la pasarela NCG (324, 422) está configurada además para recibir una respuesta de cálculo de ruta desde el elemento PCE, en donde la respuesta de cálculo de ruta comprende una pluralidad de rutas de red, en donde cada ruta de red está entre la dirección de origen y cada una de las direcciones de destino,
- 35 40 y en donde la respuesta de cálculo de ruta comprende además un valor de recursos de red para cada ruta de red o un valor de recursos para cada enlace de una topología de la red.
- 40 45 2. El sistema según la reivindicación 1, en donde la pasarela CSCG (314, 414) está configurada además para obtener los datos de recursos de aplicación desde una pluralidad de servidores, y en donde la pasarela CSCG está configurada además para llenar la lista de direcciones de destino con direcciones de red de servidores (312) con datos de recursos de aplicación que satisfacen una necesidad de recursos de aplicación.
- 45 3. El sistema según la reivindicación 1, en donde la pasarela CSCG (314, 414) está configurada además para transmitir una segunda necesidad de recursos de red, y en donde la segunda necesidad de recursos de red se transmite como parte de una comunicación que comprende la primera necesidad de recursos de red.
- 50 4. El sistema según la reivindicación 1, en donde el mapa de recursos de red comprende un valor de recursos de red de cada enlace en una ruta de red entre una dirección de origen y una dirección de destino contenida en la lista de direcciones de destino.

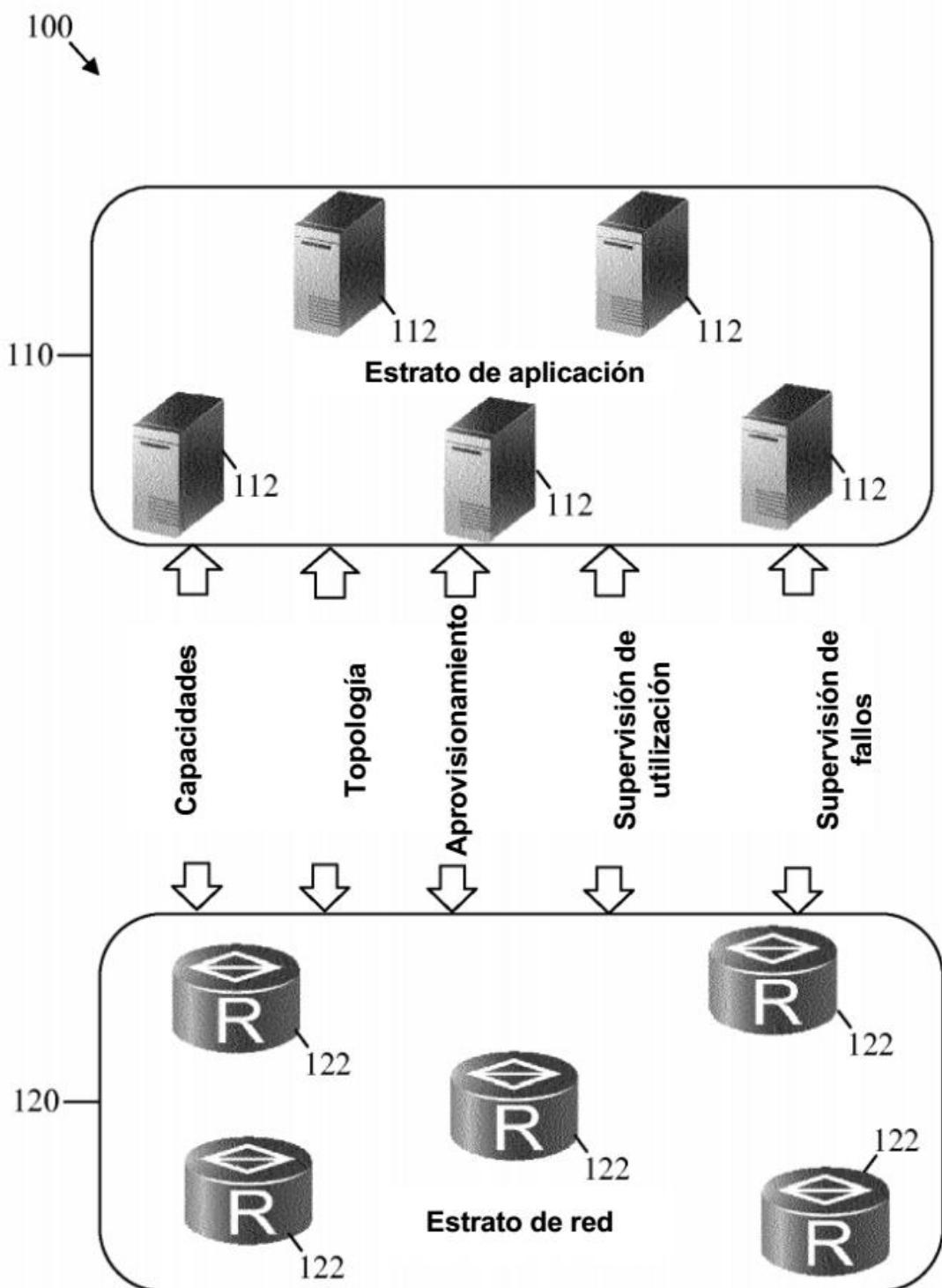


FIG. 1

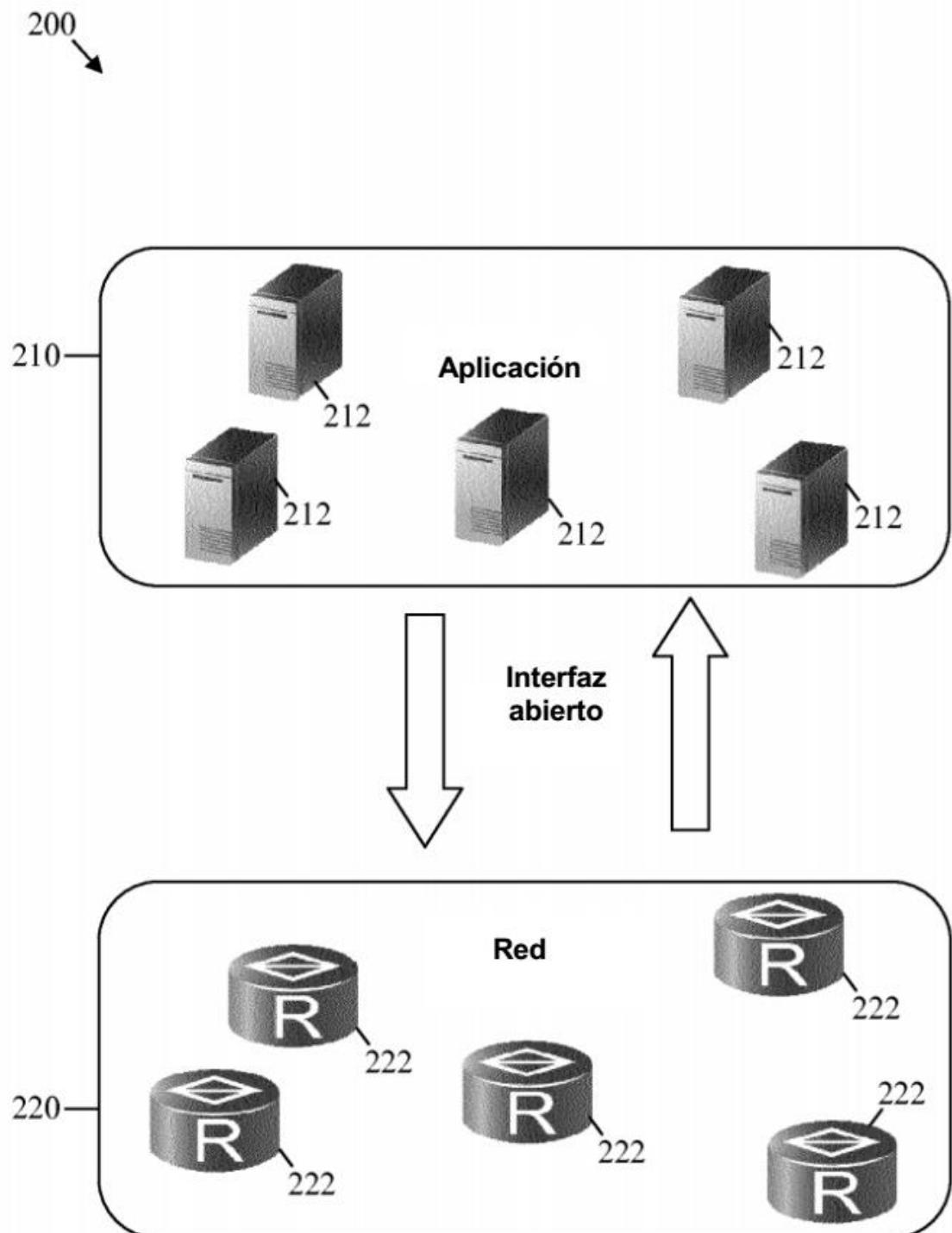


FIG. 2

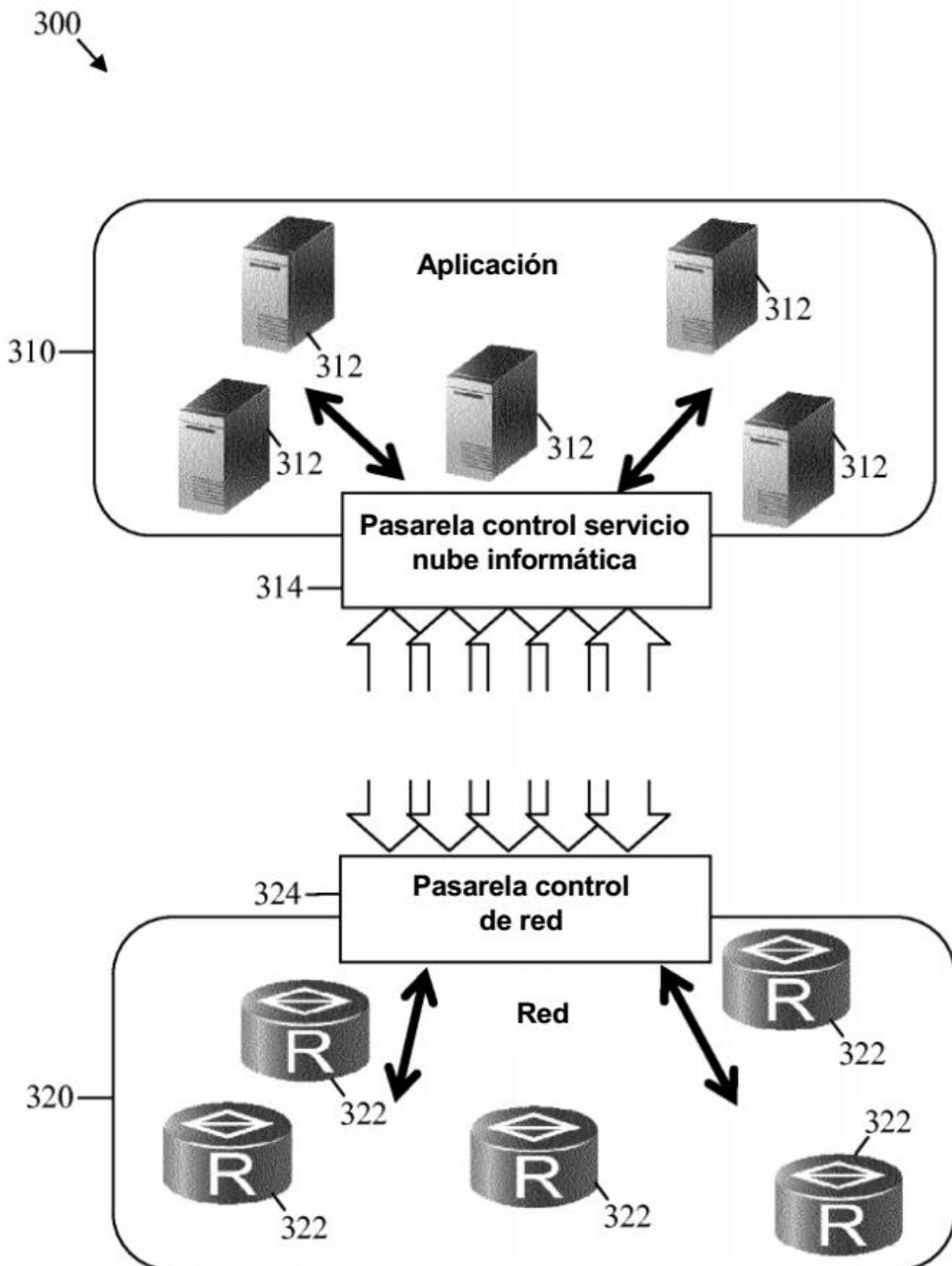
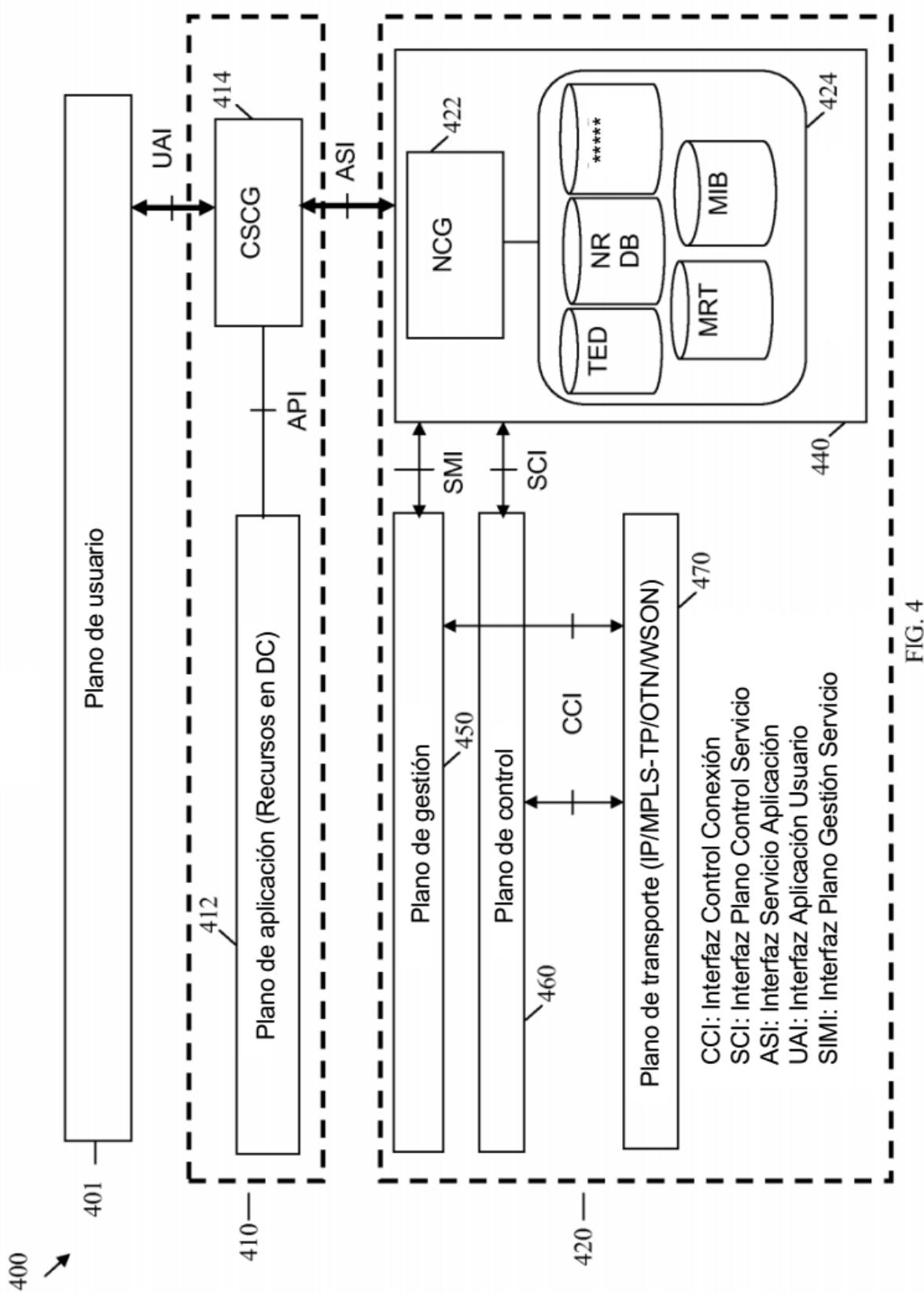


FIG. 3



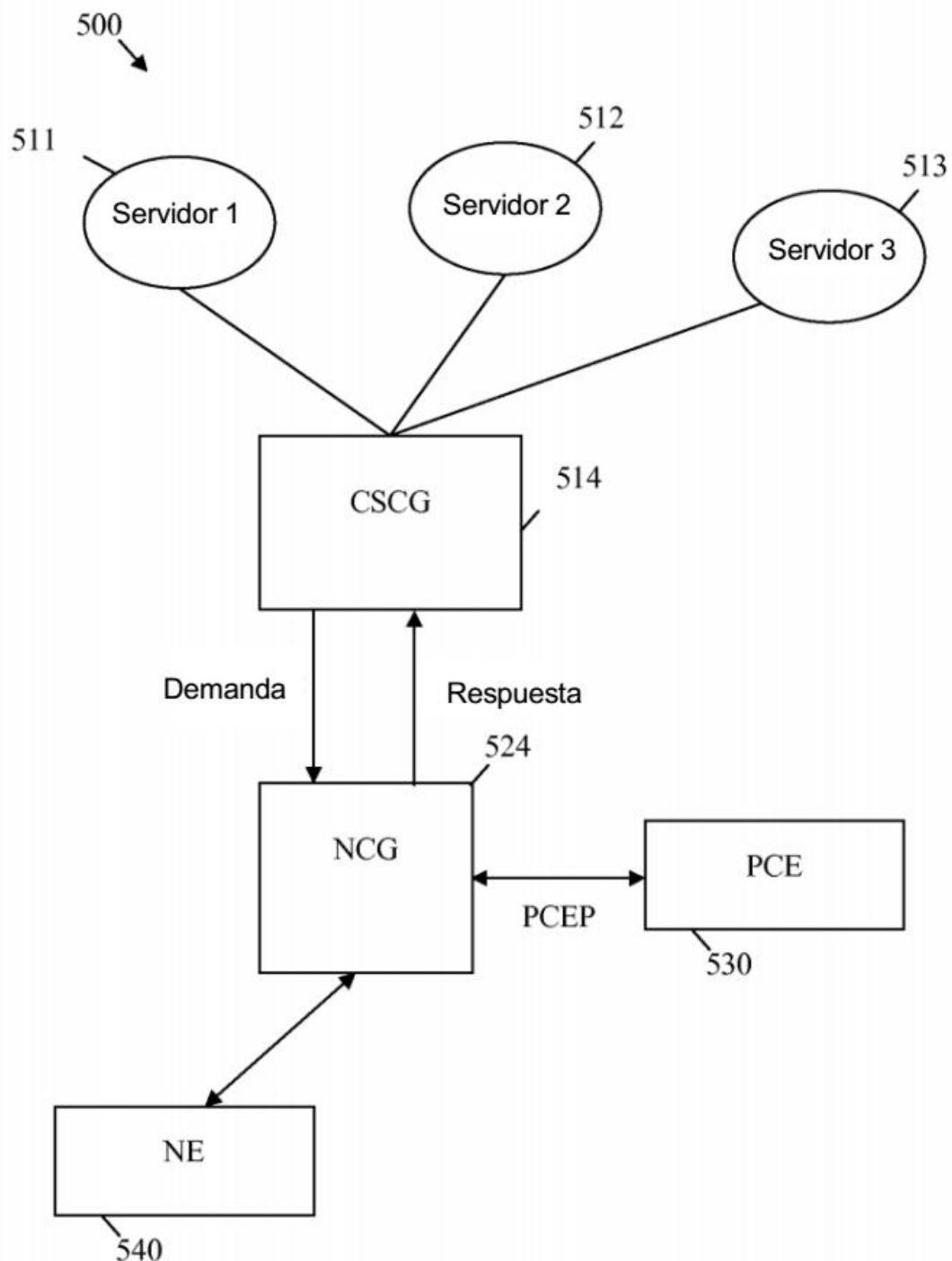


FIG. 5

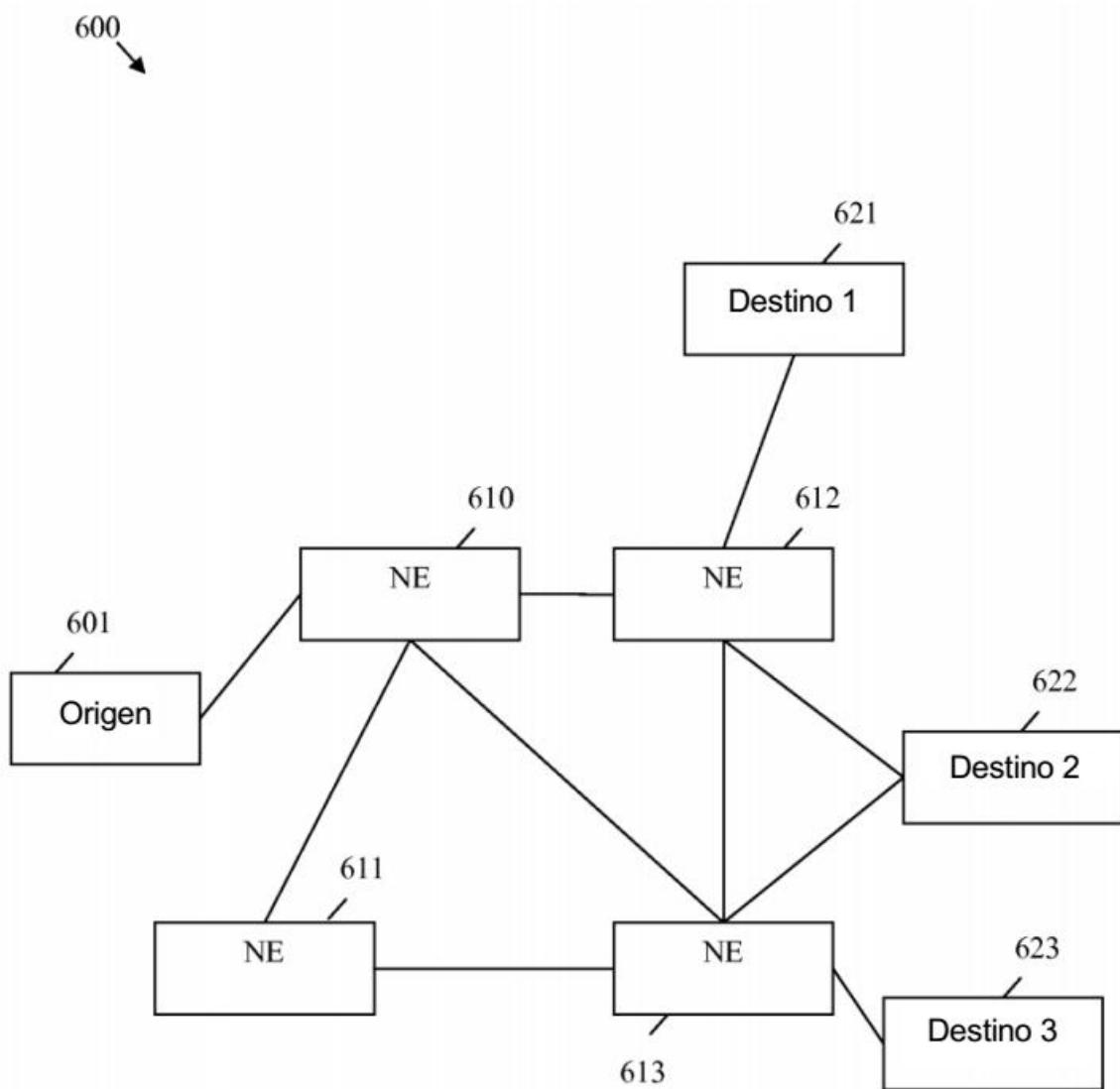


FIG. 6

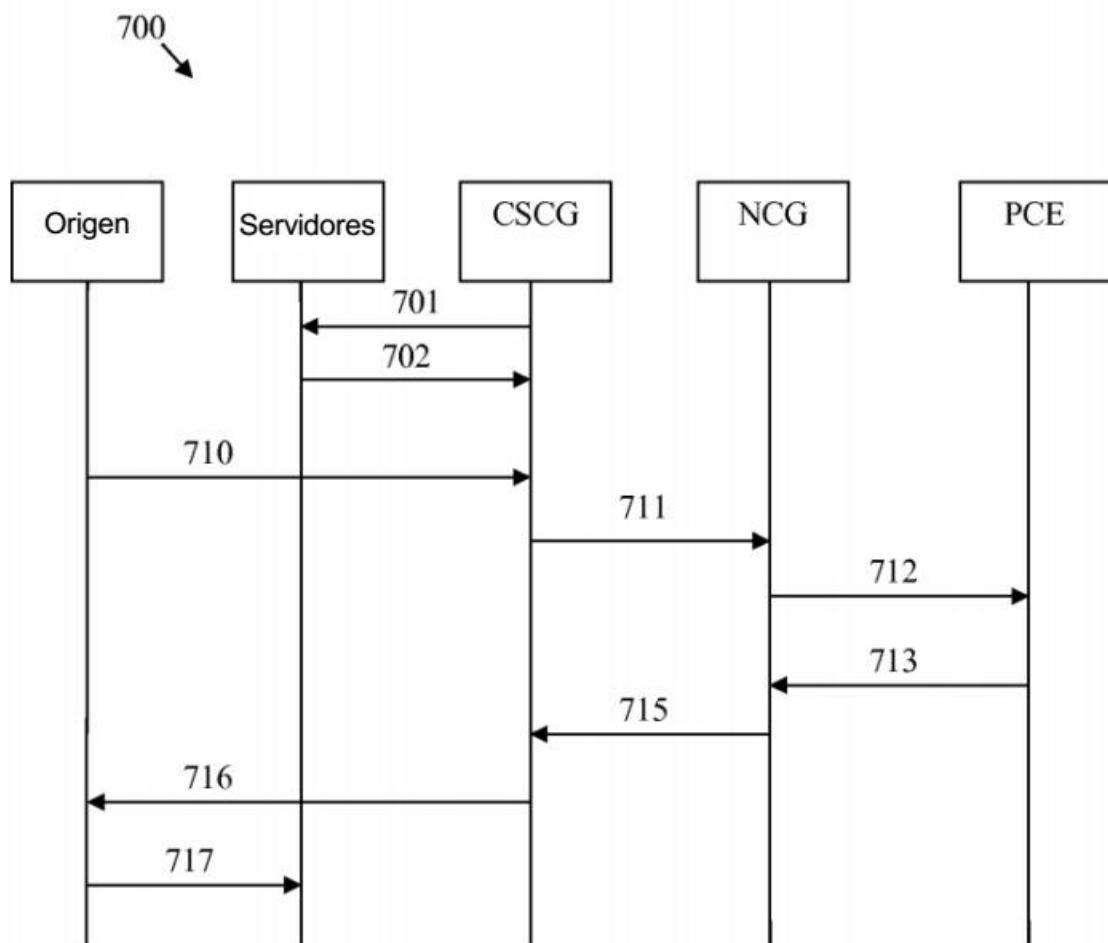


FIG. 7

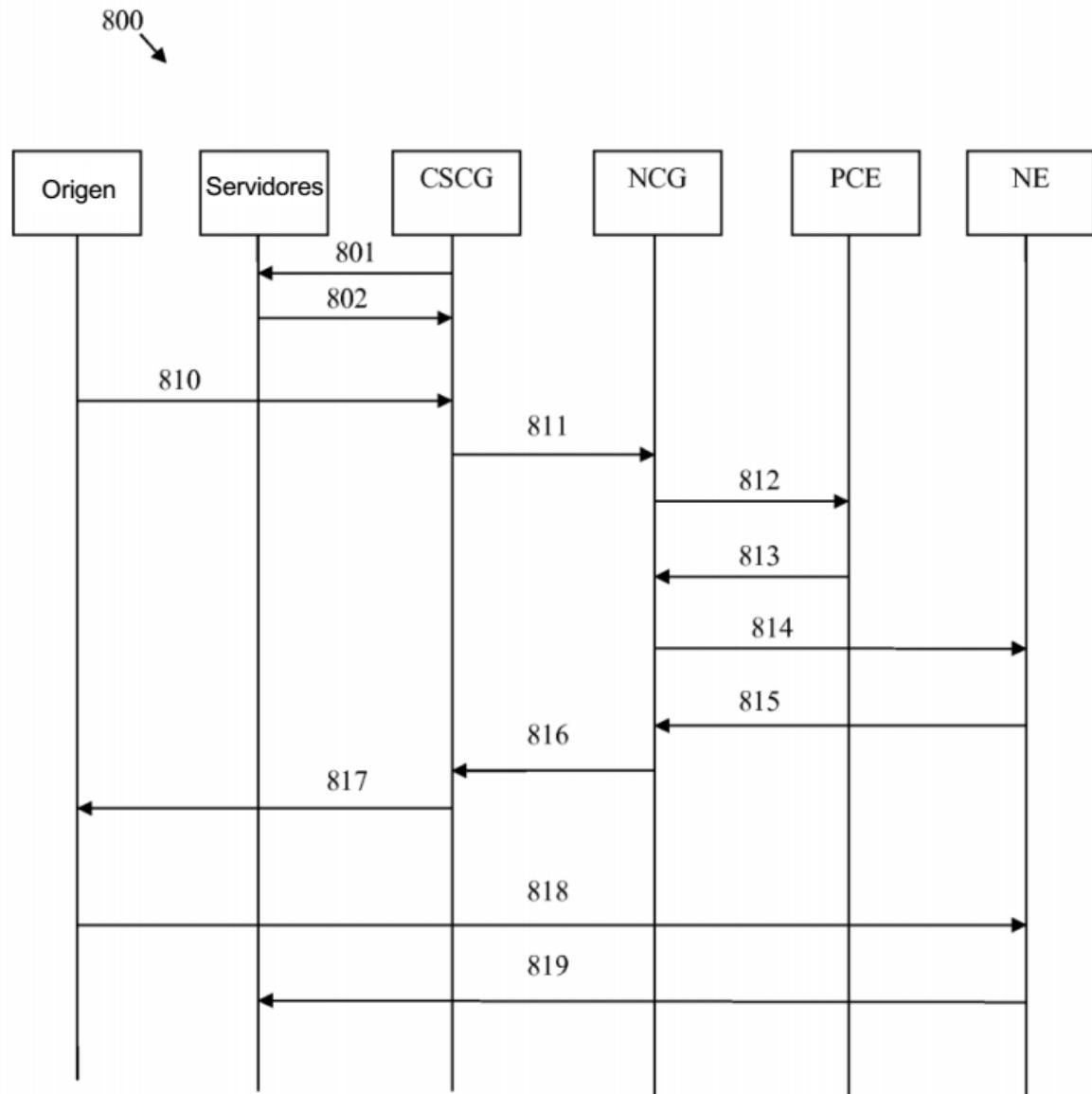


FIG. 8

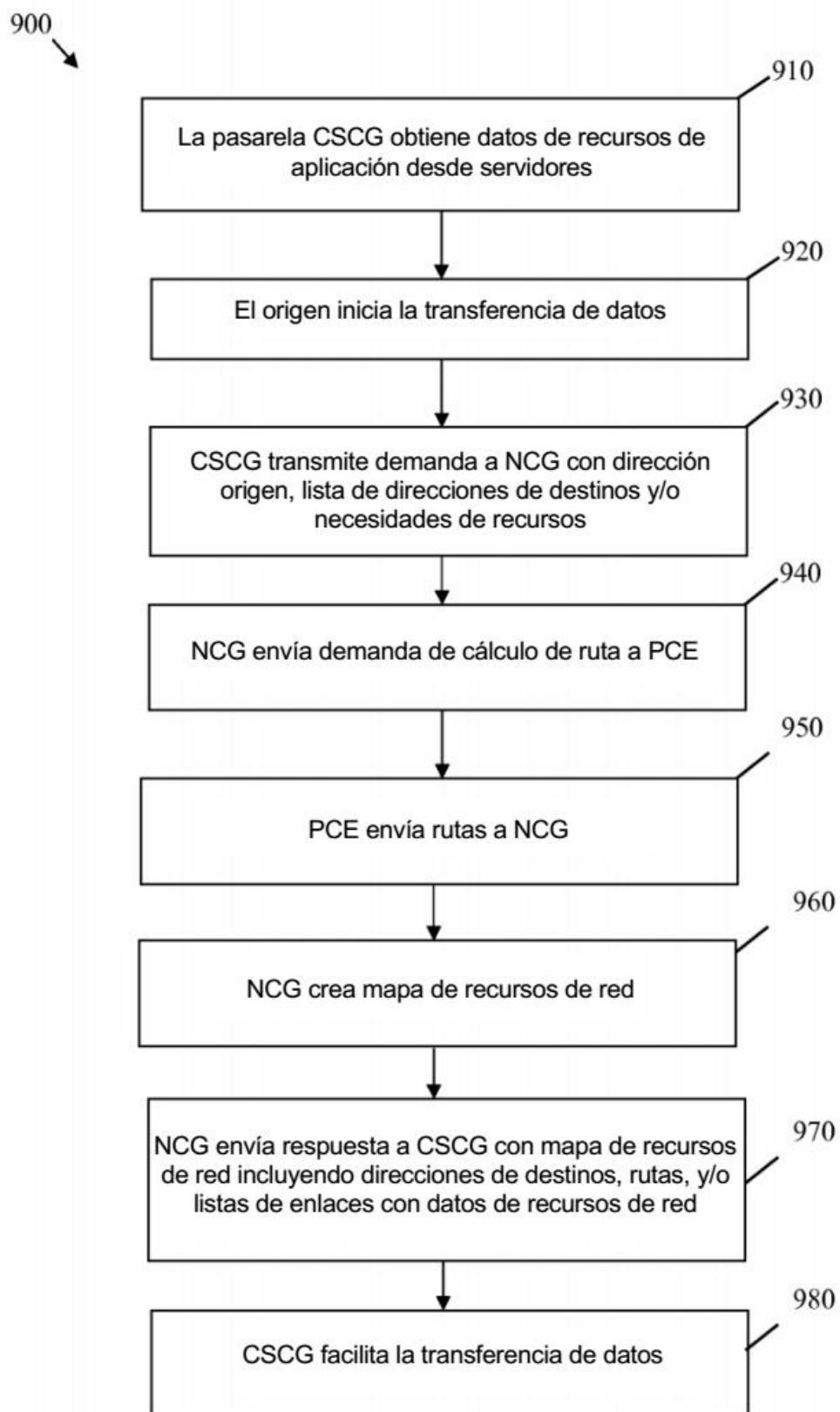


FIG. 9

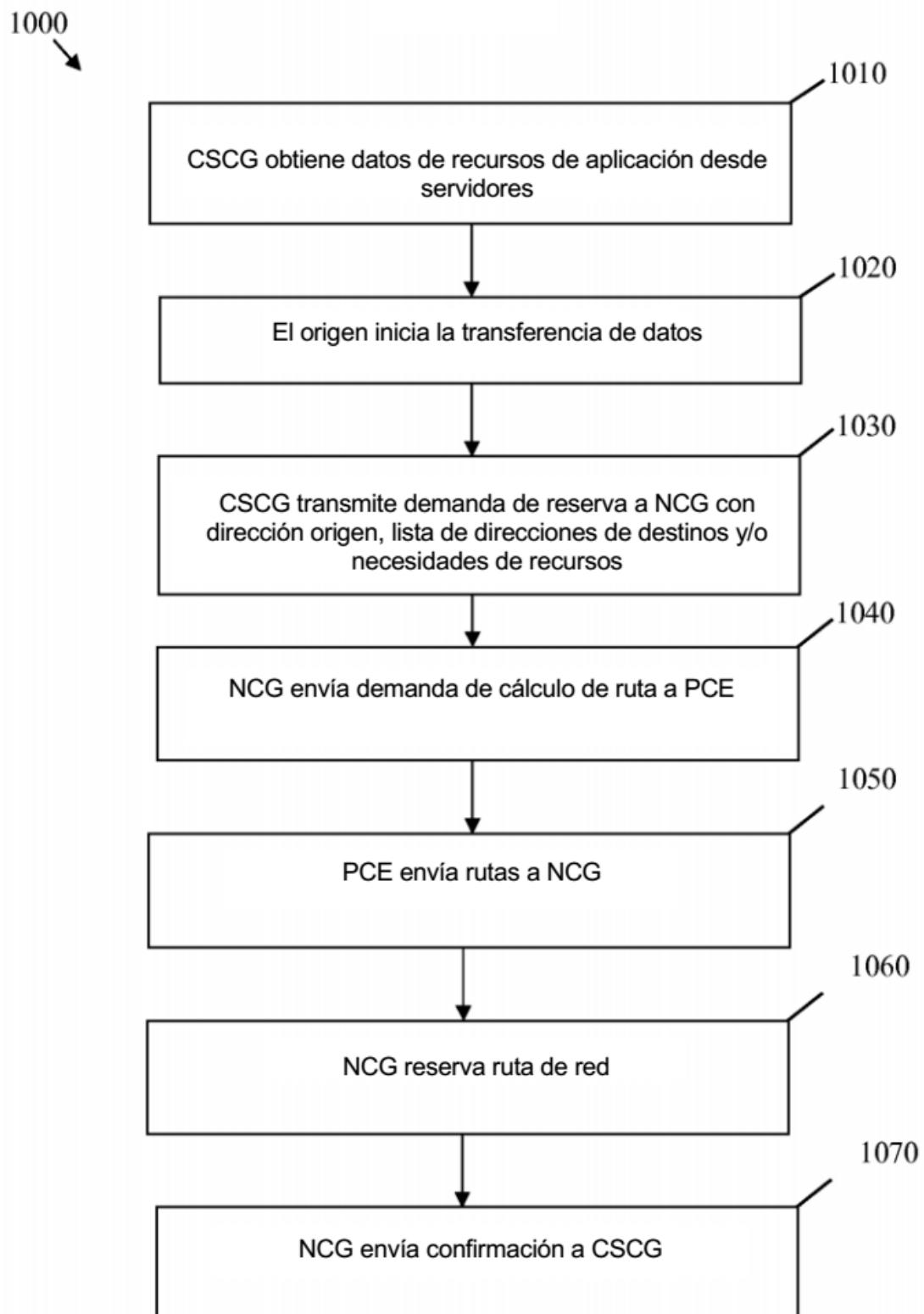


FIG. 10

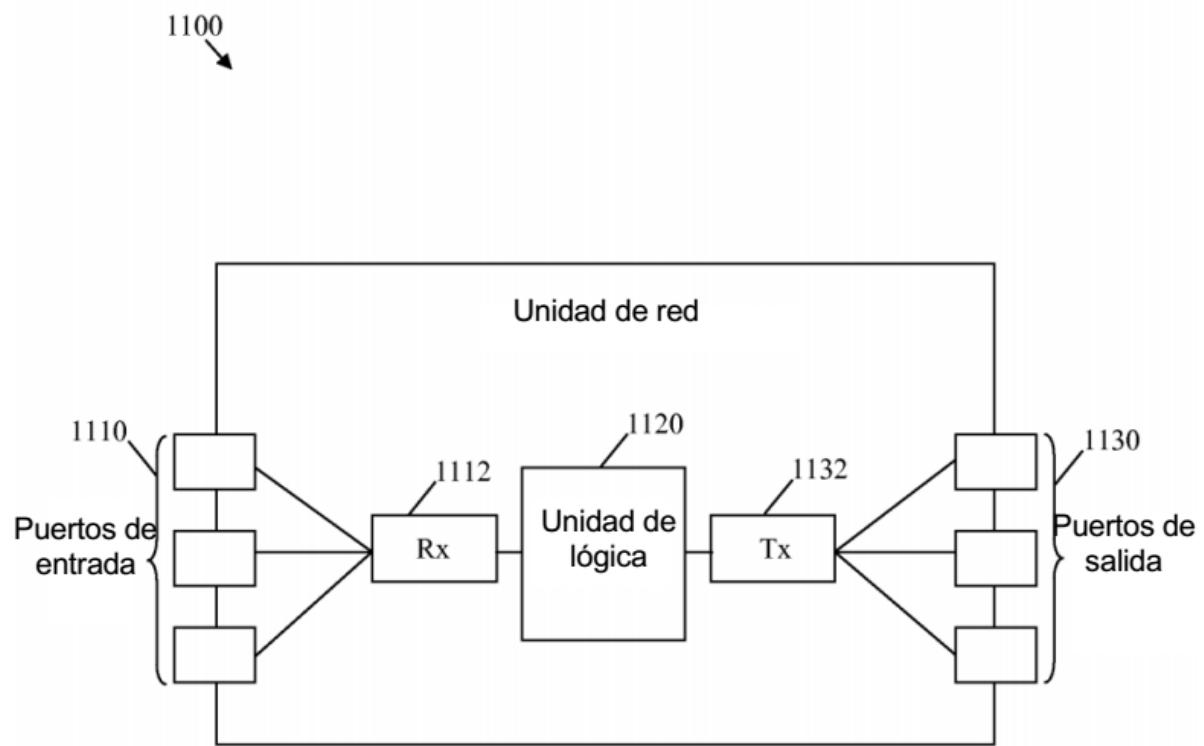


FIG. 11

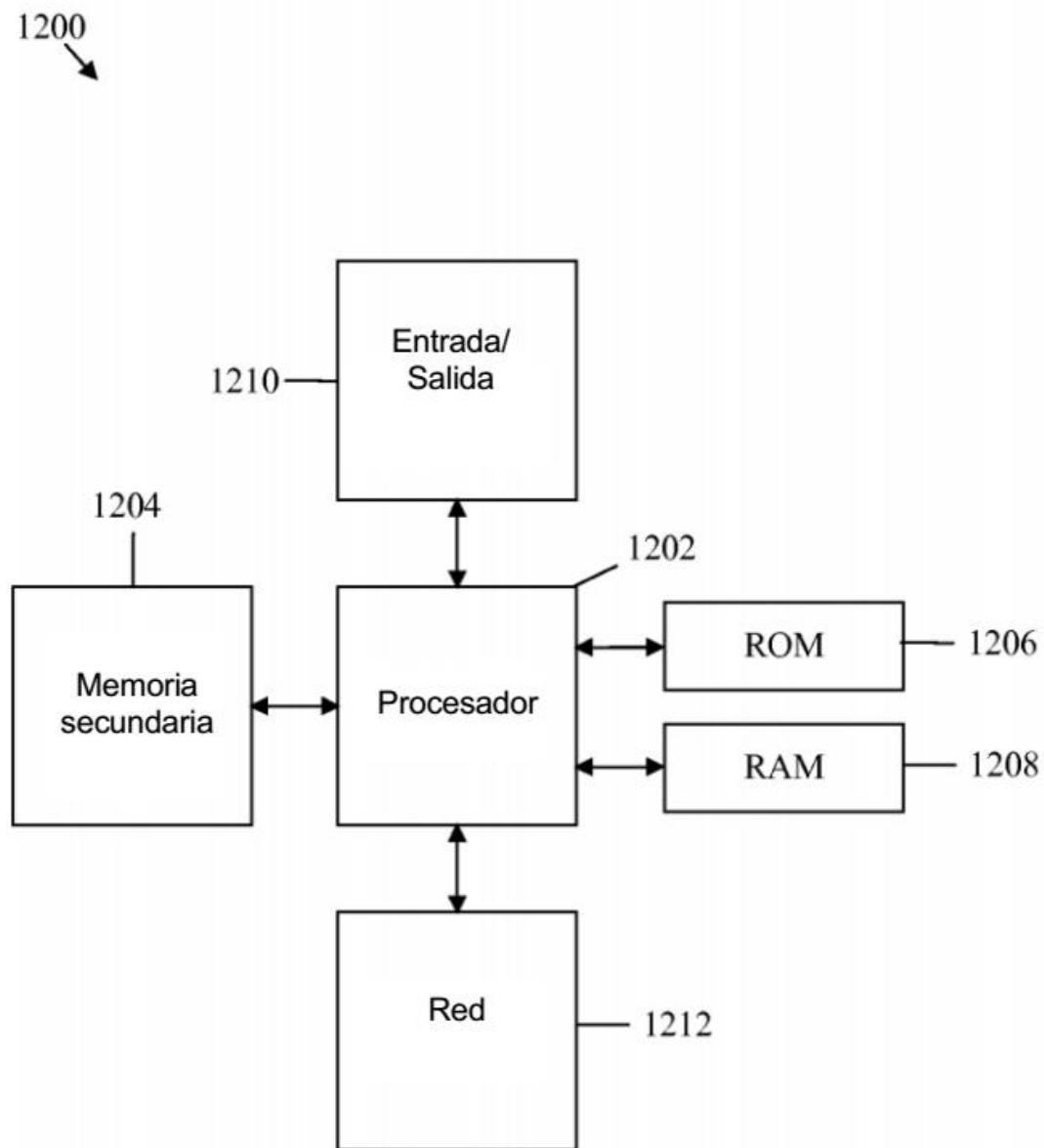


FIG. 12