

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 746 922**

51 Int. Cl.:

H04L 29/10 (2006.01)

H04L 29/08 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **26.08.2011 PCT/CN2011/079001**

87 Fecha y número de publicación internacional: **01.03.2012 WO12025061**

96 Fecha de presentación y número de la solicitud europea: **26.08.2011 E 11819436 (4)**

97 Fecha y número de publicación de la concesión europea: **31.07.2019 EP 2569963**

54 Título: **Método y sistema para la optimización de estrato cruzado en redes de transporte-aplicación**

30 Prioridad:

24.08.2011 US 201113216805

26.08.2010 US 377361 P

26.08.2010 US 377352 P

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

09.03.2020

73 Titular/es:

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

72 Inventor/es:

**LEE, YOUNG;
XIA, YANGSONG y
HARES, SUSAN**

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 746 922 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Método y sistema para la optimización de estrato cruzado en redes de transporte-aplicación

Antecedentes

Los operadores de red, también denominados a veces como operadores de telecomunicaciones o proveedores de servicios de comunicaciones, que operan las redes existentes, desean optimizar la utilización de la red para el tráfico pasante, tal como el tráfico de protocolo de internet (IP), a través de una red física, por ejemplo, a través de las capas de red 1 a 5. El tráfico optimizado puede incluir tráfico para servicios de triple reproducción (por ejemplo, Video, Voz y/o Datos) y cualquier tipo de datos masivos. En las redes existentes, los servicios de extremo a extremo se configuran normalmente mediante sistemas de soporte operacional (OSS) o aplicaciones de servicios de gestión de red específicas del proveedor. Los operadores de red han sugerido dos escenarios diferentes para optimizar la utilización de la red y el tráfico: la optimización de servicios de red existentes y la habilitación de nuevos/emergentes servicios de aplicación de red.

El documento WO 02/11459 A1 describe una pasarela parlay que incluye un bus CORBA, una capa interna de las API Parlay, una capa externa de la interfaz de servicios de red para controlar los recursos de red, una interfaz de aplicación y las API Parlay que residen en las plataformas remotas que también albergan aplicaciones que desean acceder a los recursos de red.

El documento US 2010/150161 A1 describe métodos y sistemas para la selección automática de la ruta de transporte para entidades multihogar en el protocolo de control de transmisión de corriente (SCTP) para dirigir la transferencia de datos entre aplicaciones y dispositivos que residen en diferentes ordenadores y dispositivos que utilizan el algoritmo de selección dinámica de ruta para entidades de red multihogar (utilizando SCTP). Cuando la aplicación y el dispositivo solicitan la transferencia a otra aplicación y dispositivo, el algoritmo de selección dinámica de ruta selecciona la ruta más eficiente para la transferencia de datos. La decisión para seleccionar la mejor ruta de red se basa en las condiciones dinámicas de red, tales como la ventana de congestión (CWND), el tiempo de viaje de ida y vuelta (RTT) y similares, y/o la información de aprovisionamiento.

Resumen

En una forma de realización, la descripción incluye un aparato que comprende una pasarela (ACG) de optimización de estrato cruzado (CSO) de aplicación configurada para comunicarse con varios servidores en una capa de aplicación, y una pasarela (NCG) CSO de red acoplada a la ACG por medio de una interfaz red-aplicación (ANI) y configurada para comunicarse con varios nodos de red en varias capas de red por debajo de la capa de aplicación, donde la ANI se configura para intercambiar información de supervisión y configuración entre la capa de aplicación y la capa de red y para habilitar una consulta relacionada con la capa de red y una escalada de eventos desde una capa de aplicación o una consulta relacionada con la capa de aplicación y una escalada de eventos desde la capa de red, junto con la asignación de los recursos de red de aplicación.

En otra forma de realización, la descripción incluye un componente de red que comprende un receptor configurado para recibir una consulta de red de un estrato de aplicación y una respuesta de red de un estrato de red, en donde la consulta de red se recibe desde una pasarela (ACG) de optimización de estrato cruzado (CSO) de aplicación situada en un controlador de aplicación; un controlador de servicio situado en el estrato de red configurado para habilitar la CSO entre el estrato de aplicación y el estrato de red mediante el procesamiento de la respuesta de red para la señalización del estrato de aplicación; y un transmisor configurado para enviar la consulta de red procesada al estrato de red y la respuesta de red al estrato de aplicación por medio de una interfaz servicio-aplicación (ASI) entre el estrato de aplicación y el estrato de red; en donde la ASI se configura para intercambiar información de supervisión y configuración entre el estrato de aplicación y el estrato de red y habilitar un estrato de red con conocimiento de aplicación y un estrato de aplicación con conocimiento de red.

En todavía otra forma de realización, la descripción incluye un método que comprende recibir en una pasarela (NCG) de optimización de estrato cruzado (CSO) de red en un controlador de servicio en un plano de servicio una solicitud de reserva de recursos de una pasarela (ACG) de optimización de estrato cruzado (CSO) de aplicación situada en un controlador de aplicación acoplado a un plano de aplicación por medio de una interfaz servicio-aplicación (ASI) entre un estrato de aplicación y un estrato de red para habilitar una aplicación para un usuario, en donde la ASI se configura para intercambiar información de supervisión y configuración entre el estrato de aplicación y el estrato de red y habilitar un estrato de red con conocimiento de aplicación y un estrato de aplicación con conocimiento de red calculando una ruta para la aplicación, asignando el recurso para la ruta en un plano de red utilizando bases de datos mantenidas por la red, y reenviando una respuesta con el recurso asignado al plano de aplicación por medio de la ASI entre el controlador de servicio en el estrato de red y el controlador de aplicación en el estrato de aplicación.

Estas y otras características se entenderán más claramente a partir de la siguiente descripción detallada tomada en conjunto con los dibujos y las reivindicaciones que se adjuntan.

El alcance de la invención se define mediante las reivindicaciones adjuntas.

Breve descripción de los dibujos

Para una comprensión más detallada de esta descripción, se hace referencia ahora a la siguiente breve descripción, tomada junto con los dibujos y la descripción detallada que se adjuntan, en donde los mismos números de referencia representan las mismas partes.

- La FIG. 1 es un diagrama esquemático de una forma de realización de una arquitectura CSO.
- 5 La FIG. 2 es un diagrama esquemático de otra forma de realización de una arquitectura CSO.
- La FIG. 3 es un diagrama esquemático de otra forma de realización de una arquitectura CSO.
- La FIG. 4 es un diagrama esquemático de otra forma de realización de una arquitectura CSO.
- La FIG. 5 es un diagrama esquemático de la forma de realización de una interacción entre una NCG y un elemento de cálculo de ruta.
- 10 La FIG. 6 es un diagrama esquemático de una forma de realización de una arquitectura CSO multidominio.
- La FIG. 7 es un diagrama esquemático de una forma de realización de una interacción CSO multidominio con un elemento de cálculo de ruta multidominio.
- La FIG. 8 es un diagrama esquemático de otra forma de realización de una arquitectura CSO.
- La FIG. 9 es un diagrama esquemático de una forma de realización de la arquitectura de un controlador de aplicación.
- 15 La FIG. 10 es un diagrama esquemático de una forma de realización de la arquitectura de un controlador de servicio.
- La FIG. 11 es un diagrama esquemático de una forma de realización de una reserva de recursos.
- La FIG. 12 es un diagrama esquemático de una forma de realización de una consulta de recursos.
- La FIG. 13 es un diagrama esquemático de una forma de realización de un equilibrio de carga global con conocimiento de red.
- 20 La FIG. 14 es un diagrama esquemático de una forma de realización de la escalada de eventos de red.
- La FIG. 15 es un diagrama esquemático de una forma de realización de la escalada de eventos de aplicación.
- La FIG. 16 es un diagrama esquemático de una forma de realización de una escalada de degradación de la Calidad de Servicio (QoS).
- 25 La FIG. 17 es un diagrama de flujo de una forma de realización de un método CSO.
- La FIG. 18 es un diagrama esquemático de una forma de realización de una unidad de red.
- La FIG. 19 es un diagrama esquemático de una forma de realización de un sistema informático de propósito general.

Descripción detallada

Se debe entender desde el principio que, aunque más adelante se proporciona una implementación ilustrativa de una o más formas de realización, los sistemas y/o métodos descritos se pueden implementar utilizando cualquier número de técnicas, ya sean conocidas o existentes en la actualidad. La descripción no se debe limitar de ninguna manera a las implementaciones, dibujos y técnicas ilustradas más adelante, incluyendo los diseños e implementaciones de ejemplo ilustradas y descritas en la presente memoria, pero se puede modificar dentro del alcance de las reivindicaciones adjuntas.

30

El aprovisionamiento y la operación de aplicaciones nuevas/emergentes puede implicar la resolución del problema de selección de servidor (SS) en el estrato de aplicación, así como el aprovisionamiento de red en el estrato de red subyacente. El estrato de aplicación puede incluir las aplicaciones y servicios implementados o en ejecución sobre la capa de red, y el estrato de red puede incluir las capas de transporte, red, enlace y física o combinaciones de las mismas. La gestión y coordinación del aprovisionamiento de servicios tanto en el estrato de aplicación como en el estrato de red es diferente de la gestión de servicios tradicionales, tales como el aprovisionamiento de red de servicios de telecomunicaciones de extremo a extremo. En la presente memoria se describe un sistema y métodos para proporcionar un marco de arquitectura para CSO entre el estrato de aplicación y el estrato de red. La CSO puede implicar la optimización integrada de los recursos de aplicación y de red, proporcionando una interfaz para las interacciones e intercambios entre los dos estratos. La CSO también puede incluir la coordinación tanto de los recursos de aplicación como de red. Se puede establecer una interfaz entre el estrato de aplicación y el estrato de

35

40

red para intercambiar información de supervisión y configuración. La CSO se puede lograr independientemente de cualquier optimización posible para las aplicaciones o servicios existentes que se ejecutan en una red.

La CSO puede habilitar nuevos servicios, por ejemplo, utilizando la optimización multidominio y/o multidispositivo. Los nuevos servicios pueden incluir sistemas de distribución de archivos, servicios de video de retransmisión en continuo, servicios de videoconferencia y cálculo en red. Estos servicios pueden utilizar tanto dispositivos móviles como dispositivos fijos. Los sistemas y servicios de distribución de archivos comenzaron por acelerar la descarga de páginas web, tales como las que contienen imágenes, y a continuación se expandieron para incluir suministro de archivos de software, audio y vídeo. Los servicios de retransmisión en continuo se pueden separar en dos tipos, los servicios en directo y los servicios a la carta. También se pueden crear múltiples variantes entre estos dos tipos cuando se incluye la funcionalidad de pausa o reproducción en un servicio de retransmisión en continuo en directo. La retransmisión en continuo en directo puede ser el caso donde el cliente desea recibir la retransmisión en su punto de reproducción actual en lugar de en algún punto de partida preexistente. Los servicios bajo demanda pueden proporcionar desafíos técnicos adicionales. Los proveedores de servicios pueden desear evitar largas demoras en la puesta en marcha del servicio para retener a los clientes, mientras que al mismo tiempo agrupan solicitudes juntas para ahorrar en costes del servidor. La videoconferencia pasa del escenario punto a multipunto de la distribución de contenidos en retransmisión en continuo a una situación multipunto a multipunto. Además, puede haber una restricción adicional de la calidad de servicio (QoS) en estado latente. El cálculo en red puede tener solicitudes de transferencia de archivos considerablemente grandes con dispersión reducida y tamaños de archivo más grandes.

Un problema en las interacciones entre el estrato de aplicación y el estrato de red es la falta de una interfaz estándar abierta que permita una señalización proxy entre los estratos de aplicación y de red. Esto puede limitar el intercambio de información entre estratos, el mecanismo de retroalimentación entre estratos y la asignación y reconfiguración de recursos integrados/sincronizados. Esta falta de coordinación entre los estratos de aplicación y de red puede escalar la posibilidad de desperdicio de recursos, lo que se puede traducir en un mayor coste tanto para la aplicación como para las operaciones de red.

Algunos de los términos utilizados y descritos más adelante con respecto a las características CSO incluyen: perfil de aplicación, recursos de aplicación, plano de control de aplicación, servicio de aplicación, ACG, recursos de red y una NCG. El perfil de aplicación puede comprender las características y solicitudes que el servicio de aplicación puede colocar en la red. Los recursos de aplicación pueden comprender recursos no de red que pueden ser críticos para lograr la funcionalidad del servicio de aplicación. Por ejemplo, los recursos de aplicación pueden incluir cachés, réplicas, servidores específicos de aplicación, contenido, grandes conjuntos de datos y/u otras aplicaciones relacionadas con los recursos. El plano de control de aplicación puede comprender un conjunto de recursos de aplicación que pueden estar geográficamente dispersos y constituir un plano de control con respecto al plano de datos de red. El servicio de aplicación puede ser cualquier aplicación en red ofrecida a una variedad de clientes. La ACG puede ser una entidad CSO en el estrato de aplicación que sea responsable de reunir la carga y utilización de los recursos de aplicación, tomar decisiones de asignación de recursos e interactuar con la NCG. Los recursos de red pueden comprender recursos de cualquier capa 3 o capa inferior, tales como ancho de banda, enlaces, rutas, procesamiento de rutas (por ejemplo, creación, eliminación y administración), bases de datos de red, cálculo de rutas y los protocolos de enrutamiento y señalización para la creación de rutas.

La FIG. 1 ilustra formas de realización de una arquitectura CSO 100. La arquitectura CSO 100 puede comprender un estrato de aplicación 110 y un estrato de red 120. El estrato de aplicación 110 puede comprender varios servidores 112, que se pueden configurar para implementar o ejecutar aplicaciones para usuarios finales o clientes (no mostrados). El estrato de red 120 puede comprender varios nodos de red 122, tales como puentes, enrutadores y/o conmutadores, para el reenvío de datos, por ejemplo, paquetes, asociados con las aplicaciones. Los servidores 112 se pueden situar en un centro de datos y los nodos de red 122 se pueden situar en una red acoplada al centro de datos. Los servidores 112 se pueden comunicar con los nodos de red 122 para habilitar el servicio de las aplicaciones de usuario y el reenvío o transporte de los datos asociados. La CSO se puede implementar 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 aplicación, tales como el cálculo en la nube y otros servicios en la nube, en el estrato de aplicación 110 a los usuarios finales se pueden distribuir geográficamente alrededor del estrato de red 120. Por lo tanto, muchas de las decisiones tomadas en el control y la gestión de los servicios de aplicación, tales como dónde crear otra instancia de servicio o a qué centro de datos se asigna un nuevo cliente, pueden tener un impacto significativo en el estado de la red. Las capacidades y el estado de la red también pueden tener un impacto en el rendimiento de aplicación.

En la actualidad, las decisiones de aplicación se pueden tomar con poca o ninguna información sobre el plano de datos de red utilizado para prestar esos servicios. Por lo tanto, dichas decisiones pueden no ser óptimas tanto para la utilización de los recursos de aplicación como de red y para el logro de los objetivos QoS. La CSO 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, por ejemplo, en el contexto del cálculo en la nube y las redes de centros de datos. Por ejemplo, los objetivos CSO pueden dar soporte a la consulta del estrato de red 110 desde el estrato de aplicación, al aprovisionamiento conjunto entre la aplicación y la red, y/o la reasignación conjunta de recursos en caso de

anomalía tanto en la aplicación como en la red. Los objetivos CSO también pueden proporcionar capacidades de equilibrio de carga global y de red con conocimiento de aplicación y de aplicación con conocimiento de red.

5 Algunos de los objetivos para optimizar las operaciones y/o interacciones entre el estrato de aplicación 110 y el estrato de red 120, por ejemplo, entre los servidores 112 y los nodos de red 122, pueden incluir la mejora de las capacidades de red, la topología, el aprovisionamiento, la supervisión de utilización, la supervisión de fallos o combinaciones de las mismas. Por ejemplo, los objetivos CSO 100 pueden mejorar el intercambio de cualesquiera o ambas capacidades de red o de información sobre demanda/recursos de aplicación, la topología y/o la información relacionada con la ingeniería de tráfico entre las capas (virtualización/abstracción), o ambas. Los objetivos CSO también pueden mejorar la iniciación de la creación de instancias de servicios de aplicación a la red con intercambio de perfiles (aprovisionamiento), el intercambio de información de congestión/fallos aplicación/red (supervisión), o ambos.

10 La FIG. 2 ilustra otra forma de realización de una arquitectura CSO 200 que puede comprender un estrato de aplicación 210 y un estrato de red 220. El estrato de aplicación 210 puede comprender varios servidores 212 y el estrato de red 220 puede comprender varios nodos de red 222, que pueden ser, en esencia, similares a los servidores 112 y a los nodos de red 122, respectivamente. La arquitectura CSO 200 también puede comprender una interfaz CSO que permita mejores interacciones y/o comunicaciones entre los servidores 112 y/o otros componentes (no mostrados) del estrato de aplicación 210 y los nodos de red 122 y/o otros componentes (no mostrados) del estrato de red 220. La interfaz CSO puede ser una interfaz abierta entre los dos estratos y puede habilitar algunas de las características CSO que se describen más adelante. En el estrato de aplicación 210, la interfaz abierta puede permitir la identificación del cliente/cliente de algún tipo, por ejemplo, la dirección de protocolo de internet (IP), los tipos de servidor e identificación, los flujos de datos de aplicación y solicitudes QoS que pueden ser de naturaleza estadística y variar con el tiempo, y/o condiciones de carga y fallo del servidor. En el estrato de red 220, la interfaz abierta puede permitir el intercambio de topología de red, ubicaciones de clientes y servidores dentro de esa topología, capacidades de red y capacidades con respecto a la QoS, el ancho de banda, la información de latencia y/u otras características relacionadas con la red, las condiciones de fallo y carga de la red, o combinaciones de las mismas.

15 La FIG. 3 ilustra otra forma de realización de la arquitectura CSO 300 que puede comprender un estrato de aplicación 310 y un estrato de red 320. El estrato de aplicación 310 puede comprender varios servidores 312 y el estrato de red 320 puede comprender varios nodos de red 322, que pueden ser similares, en esencia, a los servidores 112 y los nodos de red 122, respectivamente. La arquitectura CSO 300 también puede comprender una interfaz CSO que se puede establecer entre una ACG 314 en el estrato de aplicación 310 y una NCG 324 en el estrato de red 320.

20 La ACG 314 se puede configurar para acceder a los datos y procesos relacionados con la aplicación (en el estrato de aplicación 310), comunicarse con la NCG 324 (por medio de la interfaz CSO), y proporcionar abstracción/virtualización de información y limitaciones de acceso a entidades externas (fuera del estrato de aplicación 310) incluyendo las entidades del estrato de red 320. La NCG 324 se puede configurar para acceder a datos relacionados con la red (en el estrato de red 320), comunicarse con la ACG 314 (por medio de la interfaz CSO), comunicarse con procesos de red tales como el control de admisión, la reserva de recursos y/o el procesamiento de conexiones, y proporcionar abstracción/virtualización de información y limitaciones de acceso a entidades externas (fuera del estrato de red 320) incluyendo las entidades del estrato de aplicación 310. Además, la ACG 314 y la NCG 324 se pueden comunicar con los servidores 312 y los nodos de red 322, respectivamente.

25 La FIG. 4 ilustra otra forma de realización de la arquitectura CSO 400 que puede comprender un estrato de aplicación 410 y un estrato de red 420. El estrato de aplicación 410 y el estrato de red 420 también se pueden denominar en la presente memoria como plano de control de aplicación y plano de datos de red, respectivamente. Además, la arquitectura CSO puede incluir uno o más usuarios finales 401 que se pueden comunicar con el estrato de aplicación 410. El estrato o plano de control de aplicación 410 puede comprender una ACG 412 que se puede comunicar con los procesos de aplicación 414 y los datos relacionados con la aplicación 416. La red o el plano de datos 420 puede comprender una NCG 422 que se puede comunicar con los procesos de red 424 y los datos relacionados con la red 426. La ACG 412 y la NCG 422 también se pueden comunicar entre sí por medio de una ANI y un protocolo.

30 El plano de control de aplicación 410 puede ser una red que comprende varios servidores/recursos de aplicación que proporcionan servicios de aplicación, tales como el suministro de contenidos o servicios de vídeo a la carta (VOD), a los usuarios finales 401. En relación con el plano de control de aplicación 410, el plano de datos de red 420 puede ser una red subyacente que transporte tráfico en la unidad de datos en función de su tecnología de transporte. En la arquitectura CSO, cada estrato puede mantener su propia independencia y autonomía. Por ejemplo, si el plano de control de aplicación 410 necesita comunicarse con el plano de control de red 420, cada estrato se puede mantener independiente del otro. Puede haber una relación de confianza establecida entre los dos estratos antes de las comunicaciones y esa relación de confianza se puede verificar por medio de un mecanismo de autorización/autenticación.

- 5 La ANI entre el estrato de aplicación 410 y el estrato de red 420 se puede configurar para permitir la asignación y reasignación/reoptimización de recursos de aplicación-red conjuntos y el aprovisionamiento de recursos de aplicación-red conjuntos. La ANI también puede permitir una consulta del estrato de red desde una capa de aplicación o una aplicación para su aprovisionamiento de servicios y la escalada conjunto de eventos de aplicación-red desde las capas de red a las de aplicación o desde las capas de aplicación a las de red. Además, la ANI puede habilitar una capa de red con conocimiento de la aplicación y una capa de aplicación con conocimiento de la red. Estas propiedades/características de la ANI se describen con más detalle más adelante.
- 10 La ACG 412 puede atender como un proxy al plano de datos de red 420 y a los procesos relacionados con la aplicación, incluyendo el acceso al perfil 401 del usuario final. Algunas de las funcionalidades de la ACG 412 pueden incluir:
- comunicarse con la NCG 422 por medio de un protocolo que pueda permitir solicitudes para:
 - la topología virtual de red/información de ingeniería de tráfico (TE);
 - la estimación de ruta y la reserva de ruta; y
 - el estado e información de los recursos de aplicación (por ejemplo, el servidor).
 - 15 acceder a datos relacionados con la aplicación, tales como:
 - el número máximo de casos simultáneos de la utilización de la aplicación;
 - el almacenamiento máximo asignable;
 - la asignación física o virtual de procesamiento;
 - la memoria, la velocidad de acceso al almacenamiento (disco, memoria de acceso aleatorio (RAM), etc.);
 - 20 la disponibilidad de casos de máquinas virtuales (existentes o creadas) en una ubicación diferente; y
 - si una aplicación se debe ejecutar en múltiples solicitudes físicas y de conmutación por error. comunicándose con los procesos de aplicación; y
 - traducir el perfil de servicio de aplicación/usuario final y crear un perfil de servicio de aplicación "estándar" que se pueda entender por la NCG 422.
 - 25 La NCG 422 puede atender como un proxy al plano de control de aplicación 410 y a los procesos relacionados con la red. Algunas de las funcionalidades de la NCG 422 pueden incluir:
 - comunicarse con la ACG 412 por medio de un protocolo que pueda permitir respuestas a:
 - las solicitudes de aplicación enviadas por la ACG 412, según se describió anteriormente;
 - 30 el acceso a datos relacionados con la red (por ejemplo, la base de gestión de información (MIB)/YANG, la base de datos de estado de enlaces (LSDB), la base de datos TE (TED), etc.);
 - comunicarse con los procesos de red, tales como:
 - el control de admisión, la reserva de recursos;
 - el cálculo de ruta, el aprovisionamiento/configuración de ruta (creación, eliminación y mantenimiento); y
 - la supervisión de red.
 - 35 La gestión de redes de internet emergentes puede utilizar la función netconf para configurar y supervisar los datos. Las MIB basadas en el protocolo simple de administración de red (SNMP) están siendo reemplazadas por MIB de módulos YANG. El nuevo trabajo dentro de la gestión de red emergente de netconf tiene por objeto proporcionar una configuración y supervisión síncrona y sincronizada de toda la red. Si estos servicios están disponibles, entonces la NCG puede usar estos servicios para supervisar y configurar a través de todas las entidades de red a petición de la ACG. La NCG puede requerir la capacidad de tener funciones de configuración en toda la red señaladas a la entidad netconf con la siguiente información:
 - 40 commit-config <transaction #> <time>
 - copy-config < transaction #> < time>
 - edit-config <transaction #> <time>

roll-back-to <transaction #> <time>

roll-forward-to <transaction #> <time>

lock-config <transaction #> <time>

unlock-config <transaction #> <time>

- 5 La NCG puede exigir que dichas funciones y/o información estén disponibles en números basados en transacciones. La NCG también puede requerir la capacidad de tener supervisión de toda la red con la siguiente información:

begin-monitor <transaction-config #> <time>

cease-monitor <transaction #> <time>

modify-monitor < transaction #> < time>

- 10 roll-back-to <transaction #> < time>

roll-forward-to <transaction #> <time>

lock-monitor < transaction #> <time>

unlock-monitor <transaction #> <time>

- 15 De acuerdo con la solicitud NCG para especificar la supervisión de los dispositivos de toda la red en función de un número de transacción, la transacción puede especificar el perfil de la información de supervisión de toda la red. Si en una red existe una gestión de red de internet pre-netconf, tal como las MIB SNMP, la supervisión de red remota (RMON), o la supervisión QoS en tiempo real de aplicación (RAQMON) en función de la solicitud de comentarios (RFC) 3471 del grupo de trabajo de ingeniería de internet (IETF), la cual se incorpora en la presente memoria por referencia, existe en una red, o si existe una mezcla de gestión de internet, entonces el dispositivo NCG puede crear una capa de adaptación para utilizar la mezcla de servicios.

- 20 La gestión de red IP existente también puede permitir un control de admisión en función de la norma. Esta norma se puede basar en una arquitectura de "puntos de ejecución de la norma (PEP)" y "puntos de control de la norma (PCP)" con una herramienta de gestión, según se describe en RFC 3060, RFC 2753, ambas incorporadas en la presente memoria por referencia, y RFC 3471. La CSO puede ampliar el modelo de norma arquitectónica existente.
- 25 Esta arquitectura de norma general se ha adaptado para:

servicios diferenciados (Diff-Serv) dentro de redes IP por medio de servicios de norma abierta común (COPS), según se describe en las RFC 2471, y RFC 3084, RFC 4261, ambas incorporadas en la presente memoria por referencia, o el protocolo de reserva de recursos (RSVP), según se describe en la RFC 2750, que se incorpora en la presente memoria por referencia;

- 30 norma de dispositivos inalámbricos (control y aprovisionamiento de puntos de acceso inalámbricos (CAPWAP));

normas de seguridad (geopriv según se describe en la RFC 4745, que se incorpora en la presente memoria por referencia, seguridad de grupo);

norma de enrutamiento (lenguaje de especificación de normas de enrutamiento (RPSL) según se describe en la RFC 4012, que se incorpora en la presente memoria como referencia);

- 35 elementos de ruta capacitados para normas (PCE);

servicios móviles (multidifusión independiente de protocolo versión 6 (PIMv6)); y

norma de aplicación.

- 40 Además, la arquitectura CSO puede comprender un PCE, que puede ser uno de los bloques de construcción o componentes de la arquitectura CSO. La arquitectura PCE se describe en la RFC 4655 y el Protocolo PCE (PCEP) se describe en la RFC 5440, ambas incorporadas en la presente memoria por referencia. El PCE puede proporcionar el cálculo de la ruta a un cliente denominado en la presente memoria como cliente de cálculo de ruta (PCC). La NCG puede actuar como un PCC del PCE en el contexto de la arquitectura CSO.

- 45 La FIG. 5 ilustra una forma de realización de una interacción 500 entre una NCG 522 y un PCE 530 en el contexto de la estimación de ruta para la arquitectura CSO. Una ACG 512 puede hacer una solicitud de estimación de ruta a la NCG 522, que a su vez puede hacer una solicitud de cálculo de ruta utilizando el PCEP, según se describe en la RFC 5440. Las RFC 5088 y 5089, ambas incorporadas en la presente memoria por referencia, describen cómo descubrir un PCE apropiado desde la perspectiva de la NCG 522. El PCE 530 puede proporcionar rutas candidatas que cumplan con las restricciones específicas que se pueden alimentar originalmente desde la ACG 512, tales como

la conectividad (por ejemplo, punto a punto (P-P), punto a multipunto (P-MP), etc.) y algunos parámetros QoS (por ejemplo, latencia), así como solicitudes de ancho de banda para la conectividad. La ruta calculada por el PCE 530 puede ser una estimación de la ruta de la aplicación en función de los últimos datos de enlace de red y tráfico de nodos, que se pueden conocer como TED. Una vez que se ha encontrado la ruta, a continuación, la NCG 522 puede responder con la ruta resultante a la ACG 512. Si la aplicación requiere la reserva de ancho de banda de la ruta calculada, a continuación, la NCG 522 puede continuar adicionalmente con el proceso de aprovisionamiento de rutas ya sea por medio de un proceso de configuración de gestión de red o por medio de la funcionalidad del plano de control.

La FIG. 6 ilustra una forma de realización de la arquitectura CSO multidominio 600. La arquitectura CSO multidominio 600 puede comprender redes de transporte de acceso y una red de transporte troncal, y se puede ampliar a partir de las arquitecturas CSO anteriores. La arquitectura multidominio 600 puede comprender uno o más usuarios finales 601, un estrato o plano de control de aplicación 610, un estrato o plano de datos de red multidominio 620, que se pueden acoplar y disponer según se muestra en la FIG. 6. El estrato o plano de control de aplicación 610 puede comprender una ACG 612 que se puede comunicar con varias NCG 622 que corresponden a varios dominios en el plano de datos de red multidominio 620 por medio de varios interfaces de comunicación aplicación-red correspondientes.

La arquitectura CSO multidominio 600 se puede utilizar para soportar redes de planos de datos multidominio. La ACG 612 puede funcionar o actuar como el proxy central que interactúa con los usuarios finales 601 y los datos y procesos de aplicación, así como con la NCG 622 en cada dominio N (N es un número entero). La comunicación entre dominios puede hacer que se reutilicen los protocolos multidominio existentes desarrollados en el área de enrutamiento IETF y cualesquiera nuevas solicitudes se pueden introducir en los grupos de trabajo existentes. Por ejemplo, puede ser necesario mantener un identificador de aplicación a través de los dominios de la red (en el plano de datos de red multidominio 620) y en el plano de datos de aplicación 610.

La multitecnología también está implicada y soportada en la arquitectura CSO multidominio 600. Por ejemplo, el Dominio N-1 puede tener tecnología de red diferente del Dominio N. En un caso de este tipo, puede ser necesario proporcionar en cada dominio las funciones adecuadas de translación y adaptación de la información de aplicación original y su solicitud relacionada para garantizar que el perfil de servicio de aplicación se comunique sin fisuras a través de los dominios. Por ejemplo, el Dominio N-1 se puede considerar como una red de acceso en la que residen recursos de consumo y el Dominio N+1 como una red de acceso en la que residen recursos de aplicación (por ejemplo, distribución de vídeo), mientras que el Dominio N se puede considerar como la red troncal/red de agregación que proporciona transporte para los datos de aplicación. Por ejemplo, la red de acceso puede ser una red IP de capa 3 (L3), mientras que la red troncal puede ser una red óptica de capa 1 (L1).

La gestión de red (por ejemplo, netconf/YANG SNMP) para el estrato de red puede cumplir con los límites del dominio administrativo de enrutamiento (AD), según se describe en la RFC 1136, que se incorpora en la presente memoria por referencia. Según indica la RFC 1136, el AD puede comprender múltiples sistemas autónomos si estos sistemas autónomos están bajo el control administrativo de una entidad. Por ejemplo, si un proveedor BIGNet controla el Dominio 1 que tiene 4 sistemas autónomos, entonces una NCG puede operar sobre estos 4 sistemas autónomos. Los sistemas de gestión de normas (por ejemplo, PEP, PCP, etc.) también pueden respetar los límites del AD (RFC 1136). Por otra parte, según se ha mencionado anteriormente y como indica la RFC 1136, el AD puede comprender múltiples sistemas autónomos si estos sistemas autónomos están bajo el control administrativo de una entidad. Por ejemplo, si un proveedor BIGNet existe en el Dominio 1 y tiene 4 sistemas autónomos, entonces la NCG puede operar dentro del alcance de la norma del Dominio 1 del BIGnet.

La FIG. 7 ilustra una forma de realización de una interacción CSO multidominio 700 con PCE multidominio, donde uno o más usuarios finales 701 se pueden comunicar con un plano de control de aplicación 710 que interactúa con un plano de datos de red multidominio 720. El plano de control de aplicación 710 puede comprender una ACG 712 que se puede comunicar con varias NCG 722 correspondientes a varios dominios en el plano de datos de red multidominio 720 por medio de varios interfaces de comunicación de aplicación-red correspondientes. Los componentes del plano de control de aplicación 710 y del plano de datos de red multidominio 720 se pueden configurar, en esencia, de forma similar a los componentes correspondientes descritos anteriormente. Además, el plano de datos de red multidominio 720 puede comprender varios PCE 730 que corresponden a los dominios en el plano de datos de red multidominio 720. Específicamente, cada PCE 730 puede interactuar con el NCG 722 correspondiente, que puede actuar como un PCC, en el dominio correspondiente, por ejemplo, de una manera similar a la interacción 500.

Según se ha descrito anteriormente, cada Dominio de red NCG 722 se puede asociar con ese PCE 730 del Dominio. El recurso de consumo de la aplicación (por ejemplo, el usuario final) puede atravesar múltiples dominios para llegar al origen de la aplicación (por ejemplo, el servidor de vídeo). Por ejemplo, el origen de la aplicación se puede alojar en el Dominio de red N+1, mientras que el recurso de consumo de la aplicación se puede alojar en el Dominio de red N-1. El Dominio de red N puede ser una red de tránsito que conecte el Dominio de red N-1 y N+1. Como tal, se puede calcular una ruta multidominio, por ejemplo, mediante múltiples PCE 730. Una secuencia de dominio se puede determinar mediante la norma. La RFC 5441, que se incorpora en la presente memoria por referencia,

describe cómo se puede calcular una ruta conmutada de la etiqueta TE (LSP) entre dominios de una manera recursiva hacia atrás. La secuencia de dominio se puede conocer antes del cálculo de la ruta.

La FIG. 8 ilustra otra forma de realización de una arquitectura CSO 800, que puede comprender un plano de usuario 801, un estrato de aplicación 810 y un estrato de red 820. El estrato de aplicación 810 puede comprender un plano de aplicación 812 (por ejemplo, en un centro de datos (DC)), y una ACG 814, que se puede comunicar con el plano de aplicación 812 por medio de una interfaz plano-aplicación (API). La ACG 814 en el estrato de aplicación 810 también se puede comunicar con el plano de usuario 801 por medio de una interfaz usuario-aplicación (UAI). El estrato de red 820 puede comprender un plano de servicio 840, un plano de gestión 850, un plano de control 860 y un plano de transporte 870. El plano de transporte 870 puede soportar la tecnología de transporte de la infraestructura de red correspondiente, tal como para el perfil de transporte de conmutación de etiquetas multiprotocolo (MPLS-TP), la red de transporte óptica (ONT) o la red óptica de conmutación de longitud de onda (WSON).

El plano de servicio 840 se puede configurar para permitir las comunicaciones entre el plano de aplicación 812 en el estrato de aplicación 810 y el plano de gestión 850, el plano de control 860 y el plano de transporte 870 en el estrato de red 820, por ejemplo, de una manera optimizada basada en CSO. El plano de servicio 840 se puede comunicar con el plano de aplicación 812 por medio de una interfaz servicio-aplicación (ASI), el plano de gestión 850 por medio de una interfaz plano de servicio-plano de gestión (SMI) y el plano de control 860 por medio de una interfaz plano de control-plano de servicio (SCI). El plano de transporte 870 se podrá comunicar con el plano de gestión 850 y el plano de control 860, y por lo tanto con el plano de servicio 840, por medio de una interfaz control conexión (CCI).

El plano de servicio 840 se puede proporcionar por una parte o entidad (por ejemplo, un proveedor) que puede ser independiente del plano de usuario 801, del estrato de aplicación 810 y del estrato de red 820. Por ejemplo, el estrato de aplicación 810 y el estrato de red 820 se pueden gestionar por diferentes entidades o proveedores, y el plano de servicio 840 se puede gestionar por una tercera parte. El plano de servicio 840 puede comprender una NCG 822 y varias bases de datos de servicio de red 824, que pueden incluir un TED, una base de datos (DB) de enrutamiento de red (NR), una DB de configuración, unas tablas de enrutamiento múltiple (MRT), una MIB y/u otras bases de datos de red. Las bases de datos de servicios de red 824 pueden comprender al menos alguna información que se puede copiar de bases de datos similares en los planos de red. La NCG 822 se puede comunicar con la ACG 814 y, por lo tanto, el plano de aplicación 812, por medio de la ASI, con el plano de gestión 850 por medio de la SMI y con el plano de control 860 por medio de la SCI. La NCG 822 también puede acceder a la información de las bases de datos de servicio de red 824 según sea necesario para permitir el flujo de tráfico y comunicaciones entre los diferentes planos y estratos.

La FIG. 9 ilustra una forma de realización de la arquitectura de un controlador de aplicación 900. La arquitectura del controlador de aplicación 900 puede comprender un controlador de aplicación 910 (que puede incluir la ACG), una UAI 905, una ASI 915 y un plano de servicio 940. El controlador de aplicación 910 se puede situar en un estrato o plano de control de aplicación, por ejemplo, en comunicación con un plano de aplicación. El controlador de aplicación 910 puede comprender varios módulos, motores o entidades, incluyendo un motor de procesamiento de perfiles de usuario 912, un motor de selección de servicio y/o máquina virtual (VM) 914, una ACG 916, y un motor de gestión de recursos de aplicación 918, que se pueden comunicar entre sí.

El motor de procesamiento de perfiles de usuario 912 puede gestionar información relacionada con la autenticación de usuarios, facturación, extracciones de preferencias de usuario y/u otra información relacionada con el usuario final. El motor de procesamiento de perfiles de usuario 912 también se puede comunicar con el usuario final o el plano de usuario por medio de la UAI. El motor de selección de servidor/VM 914 puede asignar uno o más servidores y/o VM para la aplicación del usuario y gestionar las solicitudes de servicio del usuario. La ACG 916 se puede comunicar con el plano de servicio 940 por medio de la ASI. El motor de gestión de recursos de aplicación 918 puede rastrear recursos de aplicación, tales como servidores y VM, y/o otra conectividad con el espacio o estrato de aplicación. Los otros componentes de la arquitectura del controlador de aplicación 900 se pueden configurar de forma similar a los componentes correspondientes descritos anteriormente.

La FIG. 10 ilustra una forma de realización de la arquitectura del controlador de servicio 1000. La arquitectura del controlador de servicio 1000 puede comprender un controlador de servicio 1020, una ASI 1015, una SCI 1065 y una SMI 1055. El controlador de servicio 1020 se puede situar en un estrato o plano de datos de red, por ejemplo, en un plano de servicio. El controlador de servicio 1020 puede comprender varios módulos, motores o entidades, incluyendo un motor de procesamiento de señalización MPLS Generalizado (GMPLS) 1022, varias bases de datos de servicios de red 1024, un motor de estimación de recursos de red 1026, una pasarela CSO de red (NCG) 1028, un motor de mapeo de perfiles 1032, y un motor de abstracción y de virtualización/correlación de recursos de red 1034, que se pueden comunicar entre sí. La NCG 1028 en el controlador de servicio 1020 puede corresponder a la NCG.

El motor de procesamiento de señalización GMPLS 1022 puede formular mensajes y objetos de la interfaz red-usuario (UNI) y comunicarse con el plano de control por medio de la SCI 1065, tal como, para enviar información de la ruta, recibir solicitudes de reserva, recibir anuncios de estado de enlace (LSA) de la primera ruta más corta (OSPF), recibir mensajes de operación, administración y mantenimiento (OAM) del GMPLS y/o intercambiar otra

información relacionada con la ruta. El motor de estimación de recursos de red 1026 se puede corresponder con o puede comprender una entidad PCE o PCE plus (PCE+). La NCG 1028 puede gestionar la autorización de servicios, la norma, la suscripción, el control de admisión a la red y otras funciones según se describió anteriormente. El motor de mapeo de perfiles 1032 puede gestionar la derivación de la ubicación de red, el mapeo de parámetros, por ejemplo, genéricos a OTN, y el mapeo aplicación-conexión. Los otros componentes de la arquitectura del controlador de servicios 1000 se pueden configurar de forma similar a los componentes correspondientes descritos anteriormente.

La FIG. 11 ilustra una forma de realización de una reserva de recursos 1100 que se pueden implementar en función de la arquitectura del controlador de red 1000 (en un estrato o plano de datos de red) utilizando el plano de servicio. La reserva de recursos 1100 se puede enviar desde el estrato de aplicación al estrato de red para reservar recursos de red para una aplicación o servicio de usuario. La reserva de recursos 1100 se puede iniciar en la etapa 1 donde una ACG puede enviar una solicitud de reserva de recursos, por ejemplo, para establecer conectividad o una ruta para habilitar una aplicación o servicio, al plano de servicio. La solicitud se puede enviar a una NCG 1128 (en el controlador de red del plano de servicio) por medio de una ASI 1115. La NCG 1128 puede reenviar la solicitud a un motor de estimación de recursos de red 1126 que puede calcular un recurso, por ejemplo, una ruta, y reenviar la solicitud y/o el recurso calculado a un plano de red por medio de una SCI 1165. A continuación, la solicitud se puede reenviar a un motor de procesamiento de señalización GMPLS 1122, que a su vez puede procesar la solicitud y señalar la respuesta correspondiente por medio de la SCI a un plano de control. En la etapa 2, el plano de control puede gestionar la solicitud, por ejemplo, reservando los recursos para la ruta calculada. En la etapa 3, el plano de control puede reenviar una respuesta por medio de la SCI 1165 a las bases de datos de servicios de red 1124, que se pueden utilizar para implementar la reserva de recursos de acuerdo con la información de red en las bases de datos. La respuesta se puede a continuación reenviar al motor de procesamiento de señalización GMPLS 1122, que a su vez puede procesar la respuesta y señalar la respuesta de acuerdo con la NCG 1128. En la etapa 4, la NCG 1128 puede a continuación devolver la respuesta a la ACG por medio de la ASI 1115. El motor de mapeo de perfiles 1132, el motor de abstracción de recursos de red y correlación de datos 1134 y la SMI 1155 se pueden configurar y funcionar, en esencia, de forma similar al motor de mapeo de perfiles 1032, al motor de abstracción y de virtualización/correlación de recursos de red 1034 y a la SMI 1055. Los componentes anteriores se pueden configurar de forma, en esencia, similar a los componentes correspondientes de la arquitectura del controlador de red 1000.

La FIG. 12 ilustra una forma de realización de una consulta de recursos 1200 que también se puede implementar en función de la arquitectura del controlador de red 1000 (en un estrato o plano de datos de red) utilizando el plano de servicio. La consulta de recursos 1200 se puede enviar desde el estrato de aplicación al estrato de red para consultar si se pueden conceder recursos de red para una aplicación o servicio de usuario. La consulta de recursos 1200 se puede iniciar en la etapa 1 donde una ACG puede enviar una consulta para reserva de recursos, por ejemplo, una consulta sobre si establecer conectividad o una ruta para habilitar una aplicación o servicio, al plano de servicio. La consulta se puede enviar a una NCG 1228 (en el controlador de red del plano de servicio) por medio de una ASI 1215. La NCG 1228 puede reenviar la consulta a un motor de mapeo de perfiles 1232 que puede mapear la información de la consulta a los parámetros correspondientes de la tecnología de red. La consulta se puede enviar a continuación a un motor de estimación de recursos de red 1226 que puede determinar si una ruta o recursos pueden ser calculados, y reenviar la consulta a un motor de abstracción de recursos de red y correlación de datos 1234. En la etapa 2, la abstracción de recursos de red y la correlación de datos 1234 pueden determinar los recursos disponibles para atender a la consulta de recursos y devolver una respuesta al motor de mapeo de perfiles 1232, que puede traducir la información en la respuesta y, posteriormente, reenviar la respuesta a la NCG 1228. La NCG 1228 puede devolver la respuesta a la ACG por medio de la ASI 1215. Los componentes anteriores se pueden configurar de forma, en esencia, similar a los componentes correspondientes de la arquitectura del controlador de red 1000.

La FIG. 13 ilustra una forma de realización de un equilibrado de carga global con conocimiento de red 1300 que se puede implementar de acuerdo con las arquitecturas CSO descritas anteriormente. Inicialmente, un usuario final 1301 puede acceder a una red 1310 y enviar una consulta de servicios de red (NS) a un servidor front-end (FE) 1322, por ejemplo, en un estrato de aplicación, en un primer centro de datos 1320 (DC1). La red 1310 puede comprender varios componentes y recursos, tales como varios nodos o enrutadores 1311 para el reenvío de datos y servicios. La consulta NS se puede enviar para solicitar el acceso a un servidor y/o una ruta en la red 1310 para habilitar una aplicación o proporcionar un servicio para el usuario final 1301. El servidor FE 1322 puede mantener o acceder a información sobre el nivel de utilización del servidor y el nivel de utilización de la red de diferentes enlaces y/o nodos de red 1310. El nivel de utilización del servidor puede determinar las cargas en los diferentes servidores en el mismo centro de datos 1320 y el nivel de utilización de la red puede determinar la utilización de los recursos de red (por ejemplo, la utilización del ancho de banda) de las diferentes rutas y enrutadores 1311 que se conectan a los servidores. Por ejemplo, los niveles de utilización del servidor y de igual forma de la red pueden variar desde muy cargados (HL), medianamente cargados (ML) y ligeramente cargados (LL).

Después de recibir la solicitud NS, el servidor FE 1322 puede comparar los diferentes niveles de utilización del servidor y de la red en el mismo centro de datos 1320, por ejemplo, tal como para cuatro servidores diferentes 1, 2, 3 y 4. El servidor FE 1322 también se puede comunicar con un segundo servidor FE 1322 en un segundo DC 1320

(DC2) para comparar los niveles de utilización del servidor y de la red en DC2 con DC1. Los servidores FE 1322 pueden a continuación determinar las rutas y los enrutadores que se pueden optimizar para proporcionar la aplicación o servicio al usuario final 1301. Las rutas y enrutadores 1311 seleccionados pueden ser un compromiso entre un algoritmo SS que garantice un servidor que comprenda el contenido solicitado y un algoritmo de cálculo de ruta (PC) que garantice una mejor utilización de la red (en términos de recursos o ancho de banda). Las etapas y comunicaciones para el equilibrado de carga global con conocimiento de red 1300 se pueden implementar utilizando la arquitectura CSO. Por ejemplo, las comunicaciones entre los DC 1320 en el estrato de aplicación y el estrato de red se pueden lograr utilizando la consulta de recursos 1200 y/o la reserva de recursos 1100 y sus entidades de red y motores asociados.

5
10
15
20

La FIG. 14 ilustra una forma de realización de una escalada de eventos de red 1400 que se puede implementar en función de la arquitectura CSO. La escalada de eventos de red 1400 se puede implementar entre el estrato de aplicación y el estrato de red para gestionar un evento de red. El evento de red puede ser un fallo de red, una congestión, un evento desencadenado por una condición de red o cualquier otro evento que se pueda producir en el estrato de red y que pueda afectar al estrato de aplicación o al funcionamiento general de la red y, por tanto, afectar a las aplicaciones y servicios de usuario. Cuando se produce un evento de red, la red aplica esquemas de protección/restauración asociados con la conexión de la aplicación. Cuando la protección/restauración del nivel de red no funciona, la red puede escalar hasta el plano de servicio, lo que a su vez escala hasta la aplicación para un posible cambio del origen del recurso. La aplicación puede proporcionar una ubicación de servidor alternativa al plano de servicio. El plano de servicio puede interactuar con el plano de control para encontrar la ruta que puede proporcionar la aplicación al usuario.

25

Por ejemplo, en la etapa 1, un evento o fallo de red se puede producir en un plano de transporte 1470. El evento se puede escalar informando a un plano de control 1460, y posteriormente a un controlador de servicio 1420 en un plano de servicio 1440. El controlador de servicio 1420 puede a continuación escalar el evento a un controlador de aplicación 1410 en comunicaciones con un plano de aplicación 1412. El plano de aplicación 1412 puede a continuación responder al evento y/o informar al plano de usuario 1401. Como tal, en la etapa 2, el controlador de aplicación 1410 puede enviar una nueva solicitud al controlador de servicio 1420. La solicitud puede a continuación ser reenviada al plano de control 1460 y a continuación al plano de transporte 1470, donde la solicitud se puede procesar para gestionar el evento de red. Los componentes anteriores se pueden configurar de forma, en esencia, similar a los componentes correspondientes de la arquitectura del controlador de red 800.

30
35

La FIG. 15 ilustra una forma de realización de una escalada de eventos de aplicación 1500 que se puede implementar en función de la arquitectura CSO. La escalada de eventos de aplicación 1500 se puede implementar entre el estrato de aplicación y el estrato de red para gestionar un evento de aplicación. Cuando se produce un evento a nivel de aplicación (por ejemplo, falla de un servidor, etc.), la ACG puede intentar encontrar servidores alternativos en la misma ubicación del anfitrión. Si los servidores alternativos sólo están disponibles en las ubicaciones remotas, entonces la ACG puede proporcionar dicha información a la NCG para un posible cambio de conectividad para la conexión existente.

40

Por ejemplo, en la etapa 1, un evento o fallo de la aplicación se puede producir en un plano de aplicación 1512. Por lo tanto, un controlador de aplicación 1510 en comunicaciones con el plano de aplicación 1512, donde se produjo el evento, puede escalar el evento informando a un controlador de servicio 1520 en un plano de servicio 1540. A su vez, el controlador de servicio 1520 puede escalar el evento a un plano de control 1560. En la etapa 2, el plano de control 1560 puede informar a un plano de transporte 1570 del evento de aplicación. Los planos de red pueden a continuación gestionar el evento, por ejemplo, estableciendo nuevas rutas y/o asignando nuevos recursos para el plano de aplicación 1512. Los componentes anteriores se pueden configurar de forma, en esencia, similar a los componentes correspondientes de la arquitectura del controlador de red 800.

45
50

La FIG. 16 ilustra una forma de realización de la escalada de degradación QoS 1600 que se puede implementar en función de la arquitectura CSO. La escalada de degradación QoS 1600 se puede implementar entre el estrato de aplicación y el estrato de red para gestionar una degradación QoS, que se puede desencadenar por un evento de red, según se describió anteriormente. Cuando el usuario experimenta degradación de QoS para una aplicación, el usuario puede señalar este evento a la ACG. Si esta degradación se debe a problemas con el servidor, entonces la ACG puede intentar encontrar servidores alternativos en la misma ubicación del anfitrión y cambiar al servidor alternativo para mitigar la degradación. Si la degradación no está relacionada con un servidor o si los servidores alternativos sólo están disponibles en las ubicaciones remotas, entonces la ACG puede proporcionar dicha información a la NCG para un posible cambio de conectividad para la conexión existente.

55
60

Por ejemplo, en la etapa 1, se puede producir una degradación QoS en un plano de usuario 1601. Por lo tanto, un controlador de aplicación 1610 en comunicaciones con el plano de usuario 1601 puede escalar la degradación QoS detectada informando a un controlador de servicio 1620 en un plano de servicio 1640. A su vez, el controlador de servicio 1620 puede escalar la degradación QoS a un plano de control 1660. En la etapa 2, el plano de control 1660 puede informar a un plano de transporte 1670 de la degradación QoS. Los planos de red pueden gestionar a continuación la degradación QoS, tal como estableciendo nuevas rutas y/o asignando nuevos recursos para el plano de aplicación 1612. Los componentes anteriores se pueden configurar de forma, en esencia, similar a los componentes correspondientes de la arquitectura del controlador de red 800.

La FIG. 17 ilustra la forma de realización de un método CSO 1700 que se puede implementar por un plano de servicio, un controlador de servicio (por ejemplo, el controlador de servicio 1020) y/o una NCG (por ejemplo, la NCG 1028). El método 1700 puede comenzar en el bloque 1710, donde se puede recibir una solicitud de reserva de recursos de una ACG para habilitar una aplicación para un usuario. Por ejemplo, el controlador de servicio 1020 o la NCG 1028 pueden recibir de la ACG la solicitud de reserva de recursos por medio de una ASI entre un plano de aplicación y un plano de servicio (o un estrato de aplicación y un estrato de red). En el bloque 1720 se puede calcular una ruta para la aplicación. Por ejemplo, la NCG 1028 puede señalar la solicitud de reserva de recursos, por medio del motor de señalización y procesamiento GMPLS 1122, al motor de estimación de recursos de red 1126 (por ejemplo, un PCE) para calcular la ruta. En el bloque 1730, el recurso se puede asignar a la ruta en el plano de control utilizando bases de datos mantenidas por la red. Por ejemplo, la ruta calculada se puede enviar al plano de control por medio de la SCI para reservar el recurso asignado en función de la información en las bases de datos de servicios de red 1124. En el bloque 1740, se puede reenviar una respuesta con el recurso asignado a la ACG. Por ejemplo, el recurso asignado se puede señalar por medio del motor de señalización y procesamiento GMPLS 1122 a la NCG 1128, que a continuación puede reenviar la respuesta a la ACG. La ACG puede comunicar esta información con el plano de aplicación para establecer la ruta y el recurso para la aplicación del usuario. El método 1700 puede entonces terminar.

La FIG. 18 ilustra una forma de realización de una unidad de red 1800, que puede ser cualquier dispositivo que transporte y procese datos a través de la red. La unidad de red 1800 puede comprender uno o más puertos o unidades de entrada 1810 acopladas a un receptor (Rx) 1812 para recibir señales y tramas/datos de otros componentes de red. La unidad de red 1800 puede comprender una unidad lógica 1820 para determinar a qué componentes de red enviar datos. La unidad lógica 1820 se puede implementar utilizando hardware, software o ambos. La unidad de red 1800 también puede comprender uno o más puertos o unidades de salida 1830 acopladas a un transmisor (Tx) 1832 para transmitir señales y tramas/datos a los otros componentes de red. El receptor 1812, la unidad lógica 1820 y el transmisor 1832 también pueden implementar o soportar la reserva de recursos 1100, la consulta de recursos 1200, el equilibrio de carga global con conocimiento de red 1300, la escalada de eventos de red 1400, la escalada de eventos de aplicación 1500, la escalada de degradación QoS 1600 y/o el método CSO 1700. Los componentes de la unidad de red 1800 se pueden disponer según se muestra en la FIG. 18.

Los componentes de red descritos anteriormente se pueden implementar en cualquier componente de red de propósito general, tal como un ordenador o componente de red con suficiente potencia de procesamiento, recursos de memoria y capacidad de rendimiento de red para gestionar la carga de trabajo necesaria que se le ha asignado. La FIG. 19 ilustra un típico componente de red de propósito general 1900 adecuado para implementar una o más formas de realización de los componentes descritos en la presente memoria. El componente de red 1900 incluye un procesador 1902 (que se puede denominar como unidad de procesador central o CPU) que está en comunicación con dispositivos de memoria, incluyendo el almacenamiento secundario 1904, la memoria de sólo lectura (ROM) 1906, la RAM 1908, los dispositivos de entrada/salida (I/O) 1910 y los dispositivos de conectividad de red 1912. El procesador 1902 se puede implementar como uno o más chips CPU, o puede ser parte de uno o más circuitos integrados de aplicación específica (ASIC).

El almacenamiento secundario 1904 se compone normalmente de una o más unidades de disco o unidades de cinta y se utiliza para el almacenamiento no volátil de datos y como un dispositivo de almacenamiento de datos de sobrecarga si la RAM 1908 no es lo suficientemente grande como para contener todos los datos de trabajo. El almacenamiento secundario 1904 se puede utilizar para almacenar programas que se cargan en la memoria RAM 1908 cuando se seleccionan dichos programas para su ejecución. La ROM 1906 se utiliza para almacenar instrucciones y quizás datos que se leen durante la ejecución del programa. ROM 1906 es un dispositivo de memoria no volátil que normalmente tiene una pequeña capacidad de memoria en relación con la mayor capacidad de memoria del almacenamiento secundario 1904. La RAM 1908 se utiliza para almacenar datos volátiles y quizás para almacenar instrucciones. El acceso tanto a la ROM 1906 como a la RAM 1908 normalmente es más rápido que el acceso al almacenamiento secundario 1904.

Se describe al menos una forma de realización y las variaciones, combinaciones, y/o modificaciones de la(s) forma(s) de realización y/o características de la(s) forma(s) de realización realizadas por un experto en la técnica están dentro del alcance de la descripción. Las formas de realización alternativas que resulten de la combinación, integración u omisión de características de la(s) forma(s) de realización también están dentro del alcance de la descripción. En los casos donde se indiquen expresamente rangos o limitaciones numéricas, se debe entender que dichos rangos o limitaciones expresas incluyen los rangos o limitaciones iterativas de magnitud similar que caen dentro de los rangos o limitaciones expresamente indicadas (por ejemplo, de 1 a 10 incluye, 2, 3, 4, etc.; más de 0,10 incluye 0,11, 0,12, 0,13, etc.). Por ejemplo, siempre que se describe un rango numérico con un límite inferior, R_l , y un límite superior, R_u , se describe específicamente cualquier número que se encuentre dentro del rango. En particular, se describen específicamente los siguientes números dentro del rango: $R = R_l + k * (R_u - R_l)$, en donde k es una variable que varía desde el 1 hasta el 100 por ciento con un incremento del 1 por ciento, es decir, k es el 1 por ciento, el 2 por ciento, el 3 por ciento, el 4 por ciento, el 7 por ciento, ..., el 70 por ciento, el 71 por ciento, el 72 por ciento, ..., el 97 por ciento, el 96 por ciento, el 97 por ciento, el 98 por ciento, el 99 por ciento o el 100 por ciento. Además, cualquier rango numérico definido por dos números R , según se definió en el párrafo anterior, también se describe específicamente. La utilización del término "opcionalmente" con respecto a cualquier elemento de una

reivindicación significa que el elemento es necesario, o alternativamente, que el elemento no es necesario, ya que ambas alternativas están dentro del alcance de la reivindicación. La utilización de términos más amplios, tales como "comprende", "incluye" y "tiene", se deben entender como que proporcionan soporte a términos más ajustados, tales como "consistente en", "consistente esencialmente en" y "compuesto, en esencia, de". Todas y cada una de las
5 reivindicaciones se incorporan como una descripción adicional en la memoria descriptiva y las reivindicaciones son la(s) forma(s) de realización de la presente descripción. La discusión de una referencia en la descripción no es una admisión de que se trate de la técnica anterior, especialmente cualquier referencia que tenga una fecha de publicación posterior a la fecha de prioridad de esta solicitud. Los presentes ejemplos se deben considerar como
10 ilustrativos y no restrictivos, y la intención no es limitarse a los detalles dados en la presente memoria. Por ejemplo, los diversos elementos o componentes se pueden combinar o integrar en otro sistema o determinadas características se pueden omitir o no implementarse.

Además, las técnicas, sistemas, subsistemas y métodos descritos e ilustrados en las diversas formas de realización como distintos o independientes se pueden combinar o integrar con otros sistemas, módulos, técnicas o métodos sin apartarse del alcance de la presente descripción. Otros artículos mostrados o descritos como acoplados o
15 directamente acoplados o en comunicación entre sí pueden estar indirectamente acoplados o comunicándose a través de alguna interfaz, dispositivo o componente intermedio, ya sea eléctrica, mecánicamente o de otro modo.

REIVINDICACIONES

1. Un aparato que comprende:

una pasarela, ACG, (412, 512, 612, 712, 814, 916) de optimización de estrato cruzado, CSO, configurada para comunicarse con varios servidores en un estrato de aplicación; y

5 una pasarela CSO de red, NCG, (422, 522, 622, 722, 822, 1028, 1128, 1228) acoplada a la ACG (412, 512, 612, 712, 814, 916) por medio de una interfaz aplicación-red, ANI, y configurada para comunicarse con varios nodos de red en varias capas de red por debajo del estrato de aplicación,

10 en donde la ANI se configura para intercambiar información de supervisión y configuración entre el estrato de aplicación y un estrato de red para habilitar una consulta relacionada con el estrato de red y una escalada de eventos desde el estrato de aplicación o una consulta relacionada con el estrato de aplicación y una escalada de eventos desde el estrato de red.

2. El aparato de la reivindicación 1, en donde la ANI se configura además para habilitar un estrato de red con conocimiento de aplicación y un estrato de aplicación con conocimiento de red.

15 3. El aparato de la reivindicación 1, en donde la ACG (412, 512, 612, 712, 814, 916) conoce los procesos de aplicación y datos relacionados con la aplicación en el estrato de aplicación, y en donde la NCG (422, 522, 622, 722, 822, 1028, 1128, 1228) conoce los procesos de red y los datos relacionados con la red en el estrato de red.

4. El aparato de la reivindicación 1, en donde la NCG (422, 522, 622, 722, 822, 1028, 1128, 1228) se acopla a un elemento de cálculo de ruta, PCE, (530) en el estrato de red.

20 5. El aparato de la reivindicación 1, en donde varias ACG (412, 512, 612, 712, 814, 916) se configuran para comunicarse con varias NCG (422, 522, 622, 722, 822, 1028, 1128, 1228) que corresponden a varios dominios en el estrato de red por medio de varias ANI correspondientes.

6. El aparato de la reivindicación 5, en donde las NCG (422, 522, 622, 722, 822, 1028, 1128, 1228) se acoplan a varios elementos de cálculo de ruta, PCE, (530) correspondientes en los dominios correspondientes.

25 7. El aparato de la reivindicación 5, en donde los dominios comprenden al menos una de red de acceso, red troncal y red de agregación que se basan en la misma o diferente tecnología y que corresponden a la misma o diferentes capas de red.

8. El aparato de la reivindicación 1, en donde la NCG (422, 522, 622, 722, 822, 1028, 1128, 1228) se configura para operar sobre varios sistemas autónomos que corresponden al mismo dominio administrativo de enrutamiento, AD.

30 9. El aparato de la reivindicación 1, en donde los servidores se distribuyen en varios centros de datos y se configuran para recibir servicios de red, NS, consultas de uno o más usuarios finales en el estrato de aplicación, y en donde las consultas NS se procesan en el estrato de red para comparar diferentes niveles de utilización de servidor y de utilización de red en los centros de datos para seleccionar los servidores y los nodos de red para atender las consultas NS y lograr el equilibrio de carga de red.

10. Un componente de red que comprende:

35 un receptor configurado para recibir una consulta de red de un estrato de aplicación y una respuesta de red desde un estrato de red; en donde la consulta de red se recibe desde una pasarela, ACG, (412, 512, 612, 712, 814, 916) de optimización de estrato cruzado de aplicación, CSO, situada en un controlador de aplicación;

40 un controlador de servicio (1020, 1420, 1520, 1620) situado en el estrato de red configurado para habilitar la optimización de estrato cruzado, CSO, entre el estrato de aplicación y el estrato de red mediante el procesamiento de la respuesta de red para la señalización del estrato de aplicación; y

un transmisor configurado para enviar la consulta de red procesada al estrato de red y la respuesta de red al estrato de aplicación por medio de una interfaz servicio-aplicación, ASI (915, 1015, 1115, 1215) entre el estrato de aplicación y el estrato de red;

45 en donde la ASI (915, 1015, 1115, 1215) se configura para intercambiar información de supervisión y configuración entre el estrato de aplicación y el estrato de red y habilitar un estrato de red con conocimiento de aplicación y un estrato de aplicación con conocimiento de red.

50 11. El componente de red de la reivindicación 10, en donde el controlador de servicio se configura para comunicarse con un controlador de aplicación acoplado a un estrato de aplicación por medio de una interfaz servicio-aplicación, ASI, y en donde el controlador de aplicación se configura para comunicarse con el estrato de aplicación por medio de una interfaz estrato-aplicación, API, y con un plano de usuario por medio de una interfaz aplicación usuario.

12. El componente de red de la reivindicación 11, en donde el controlador de servicio comprende una pasarela, NCG, CSO de red y varias bases de datos de servicios de red que comprenden una base de datos de ingeniería de tráfico, TED, una base de datos DB de enrutamiento de red, NR, una base de datos de configuración, una tabla de enrutamiento múltiple, MRT, una base de información de gestión, MIB, o combinaciones de las mismas.
- 5 13. El componente de red de la reivindicación 11, en donde el controlador de servicio se configura para comunicarse con un plano de gestión por medio de una interfaz plano de gestión-plano de servicio, SMI, y un plano de control por medio de una interfaz plano de control-plano de servicio, SCI, y en donde tanto el plano de gestión como el plano de control se configuran para comunicarse con un plano de transporte por medio de una interfaz control conexión, CCI.
- 10 14. El componente de red de la reivindicación 11, en donde el controlador de servicio comprende una pasarela, NCG, CSO de red, un motor de señalización y procesamiento, un motor de estimación de recursos de red, un motor de mapeo de perfiles, un motor de abstracción, virtualización y correlación de recursos de red y varias bases de datos mantenidas por la red, en donde el motor de señalización y procesamiento se configura para comunicarse con un plano de control por medio de una interfaz plano de control-plano de servicio, SCI, y en donde las bases de datos mantenidas por la red se comunican con el plano de control por medio de la SCI y con un plano de gestión por medio de una interfaz plano de gestión-plano de servicio, SMI.
- 15 15. El componente de red de la reivindicación 11, en donde el controlador de aplicación comprende una pasarela, ACG, CSO de aplicación, un motor de procesamiento de perfiles de usuario, un servidor/máquina virtual, VM, un motor de selección y un motor de gestión de recursos de aplicación, y en donde el motor de procesamiento de perfiles de usuario se comunica con el plano de usuario por medio de una interfaz estrato de aplicación-plano de usuario UAI.
- 20 16. Un método que comprende:
- recibir, en una pasarela, NCG, de optimización de estrato cruzado de red, CSO, en un controlador de servicio (1020, 1420, 1520, 1620) en un plano de servicio (840, 940, 1440, 1540, 1640), una solicitud de reserva de recursos desde una pasarela, ACG, de optimización de estrato cruzado de aplicación, CSO, situada en un controlador de aplicación acoplado a un estrato de aplicación por medio de una interfaz servicio-aplicación, ASI, (915, 1015, 1115, 1215) entre un estrato de aplicación y un estrato de red para habilitar una aplicación para un usuario, en donde la ASI (915, 1015, 1115, 1215) se configura para intercambiar información de supervisión y configuración entre el estrato de aplicación y el estrato de red y habilitar un estrato de red con conocimiento de aplicación y un estrato de aplicación con conocimiento de red;
- 25 30 calcular una ruta para la aplicación;
- asignar el recurso para la ruta en un estrato de red utilizando bases de datos mantenidas por la red; y
- reenviar una respuesta con el recurso asignado al estrato de aplicación por medio de la ASI entre el controlador de servicio en el estrato de red y el controlador de aplicación en el estrato de aplicación.
- 35 17. El método de reivindicación comprende además la comunicación con un servidor externo en un centro de datos remoto, DC, para calcular la ruta y asignar un recurso en función de la utilización del servidor y la utilización de la red.
18. El método de la reivindicación 16, que además comprende:
- escalar un evento de red desde el estrato de red hasta el controlador de aplicación por medio del controlador de servicio, y
- 40 reenviar por medio del controlador de servicio una solicitud desde el controlador de aplicación al estrato de red para gestionar el evento de red.
19. El método de la reivindicación 16, que además comprende:
- escalar un evento de aplicación desde el estrato de aplicación hasta el estrato de red por medio del controlador de aplicación y el controlador de servicio, y
- 45 gestionar el evento de aplicación en el estrato de red.
20. El método de la reivindicación 16, que además comprende:
- escalar una degradación de la calidad de servicio QoS, desde el estrato de aplicación hasta el estrato de red por medio del controlador de aplicación y el controlador de servicio, y
- gestionar la degradación de la calidad de servicio QoS en el estrato de red.

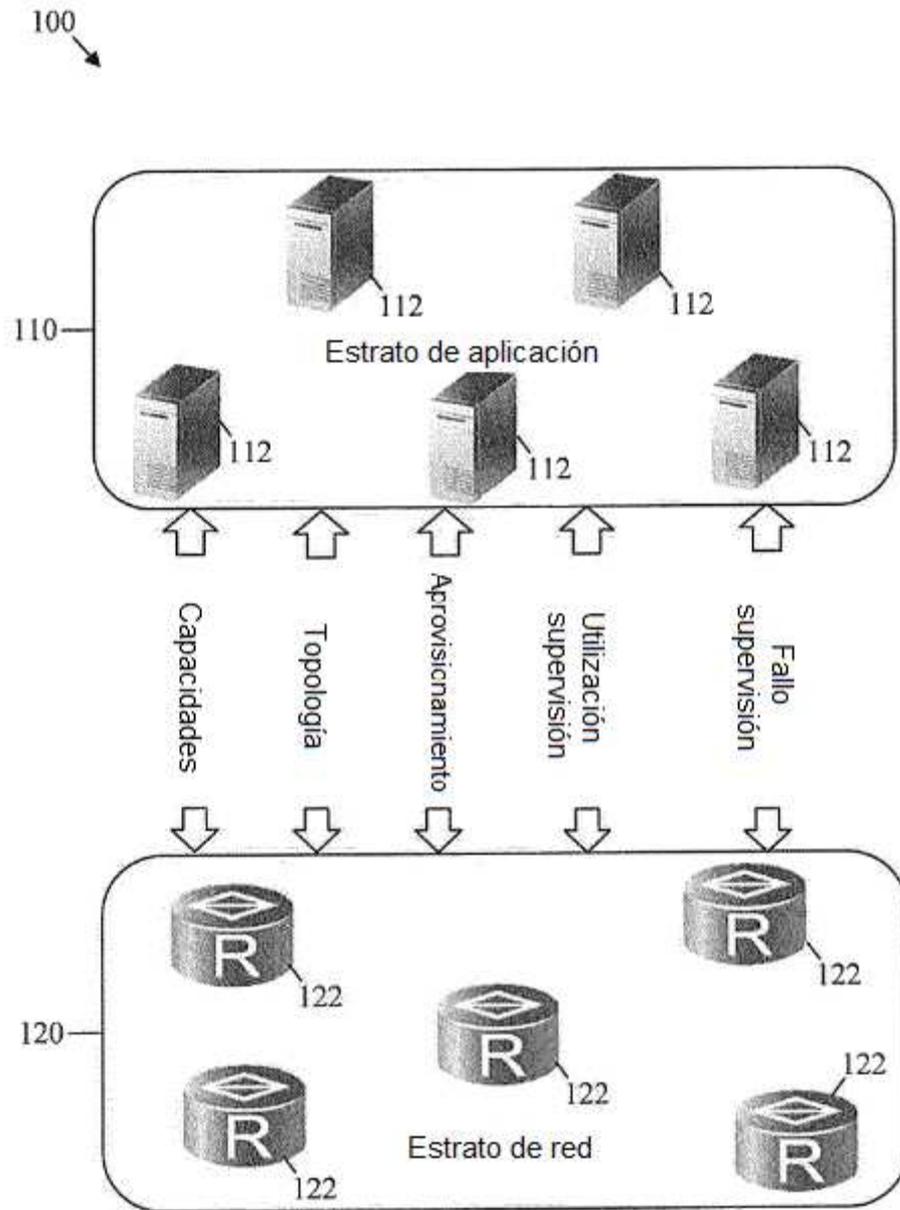


FIG. 1

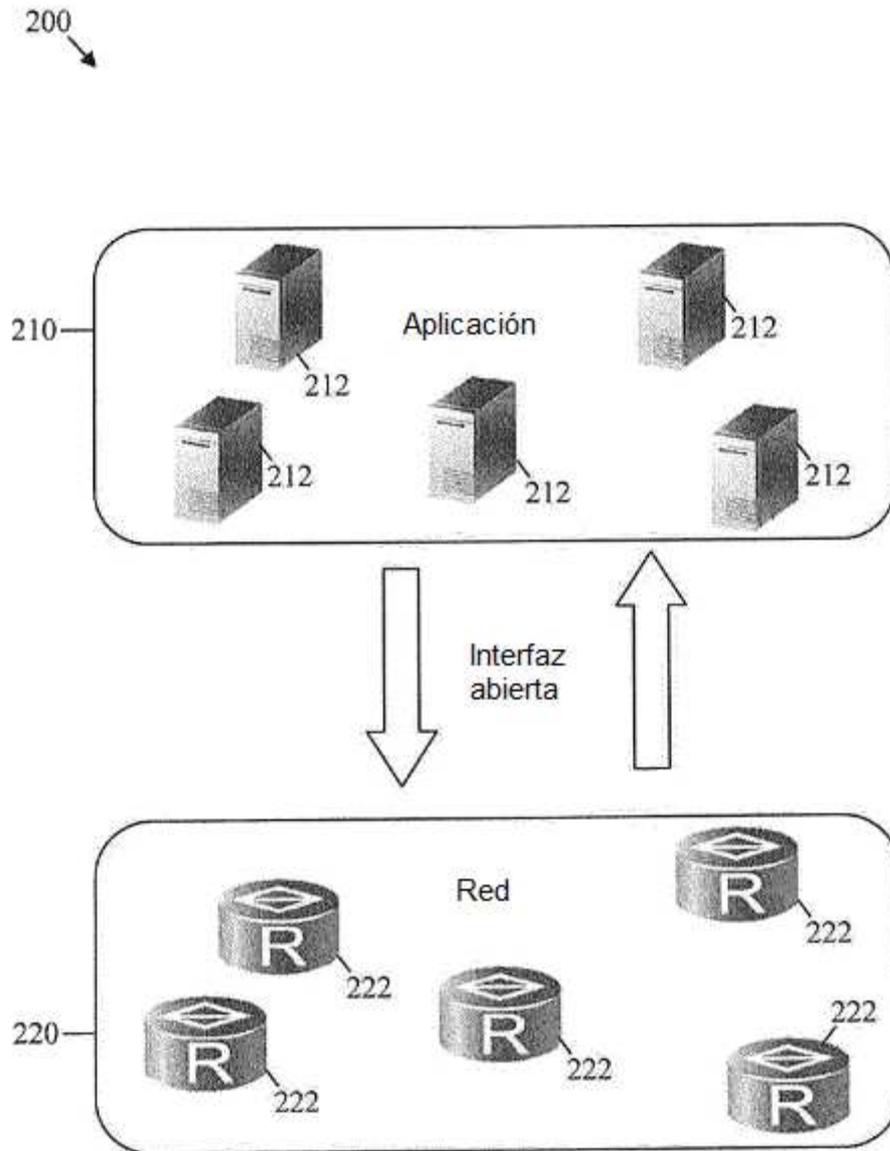


FIG. 2

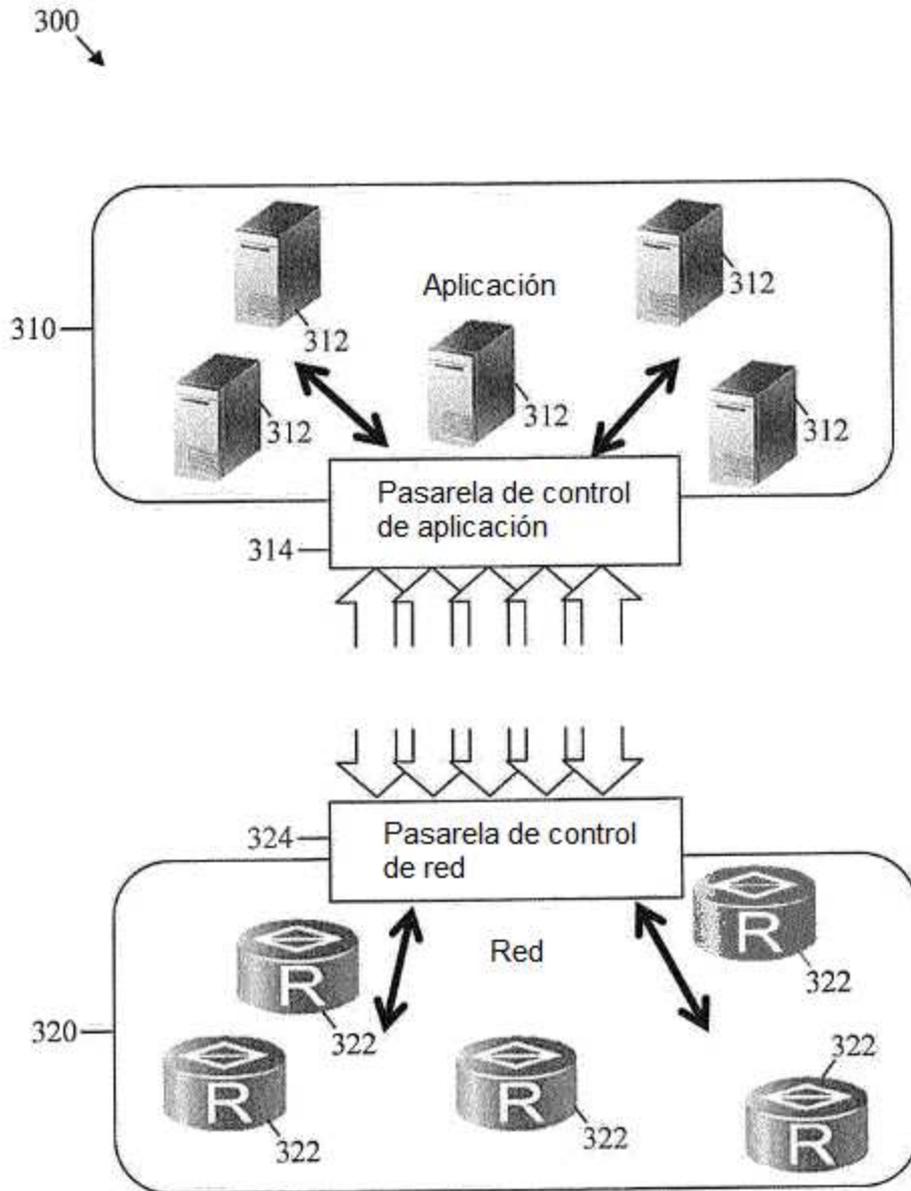


FIG. 3

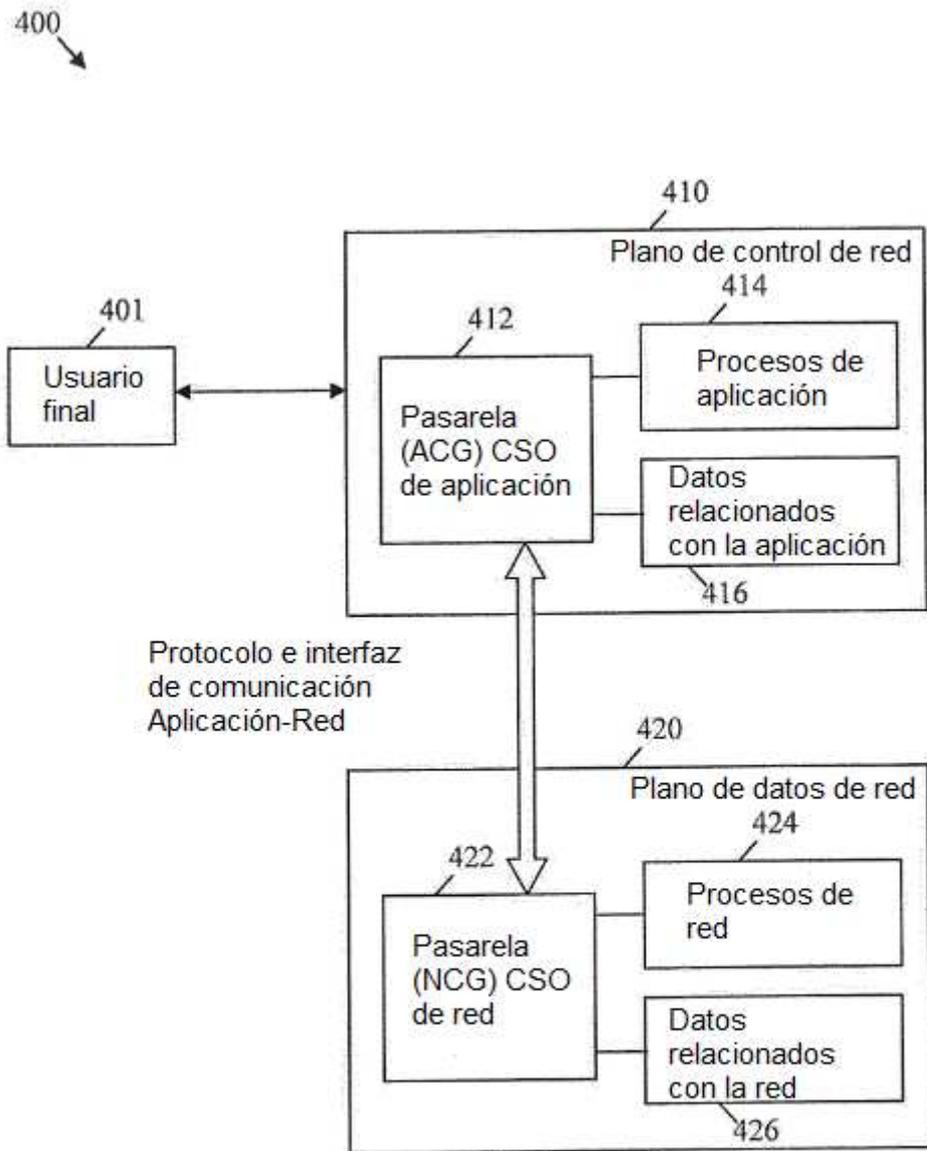


FIG. 4

500

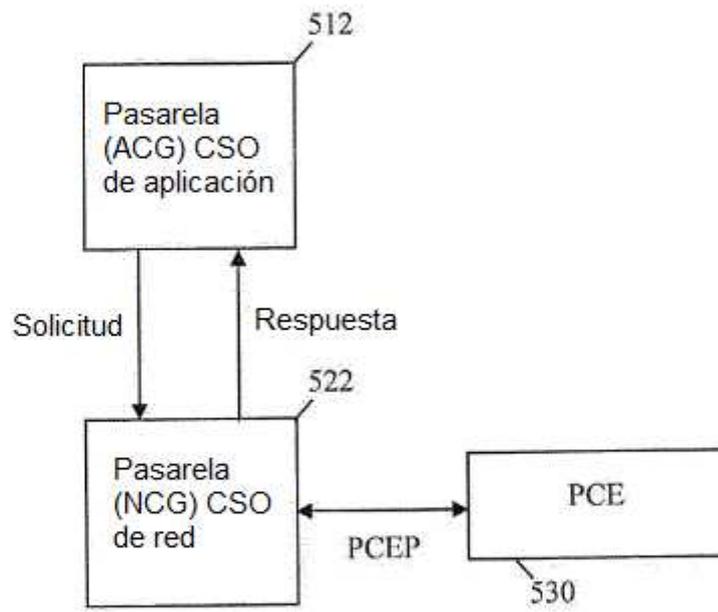


FIG. 5

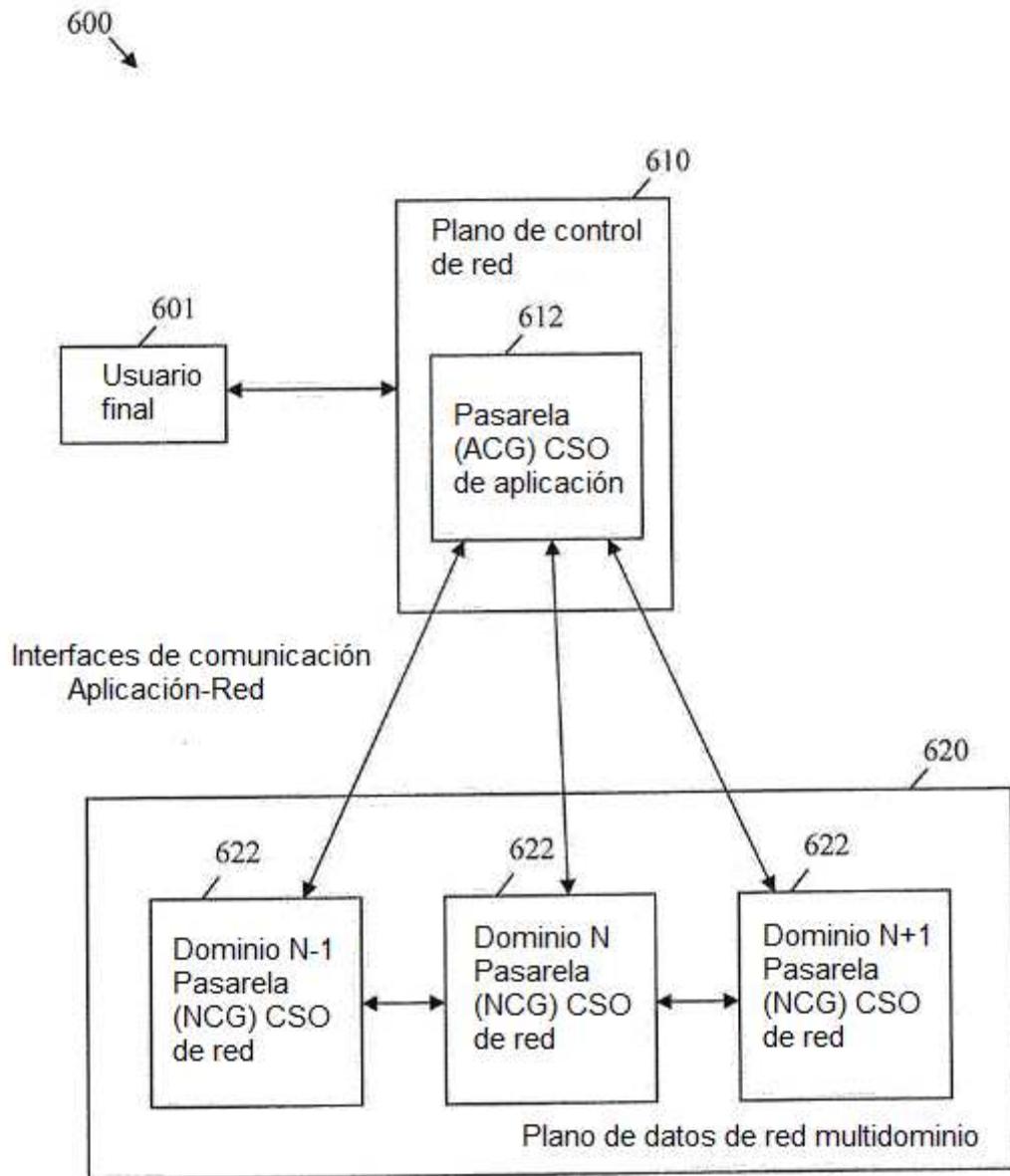


FIG. 6

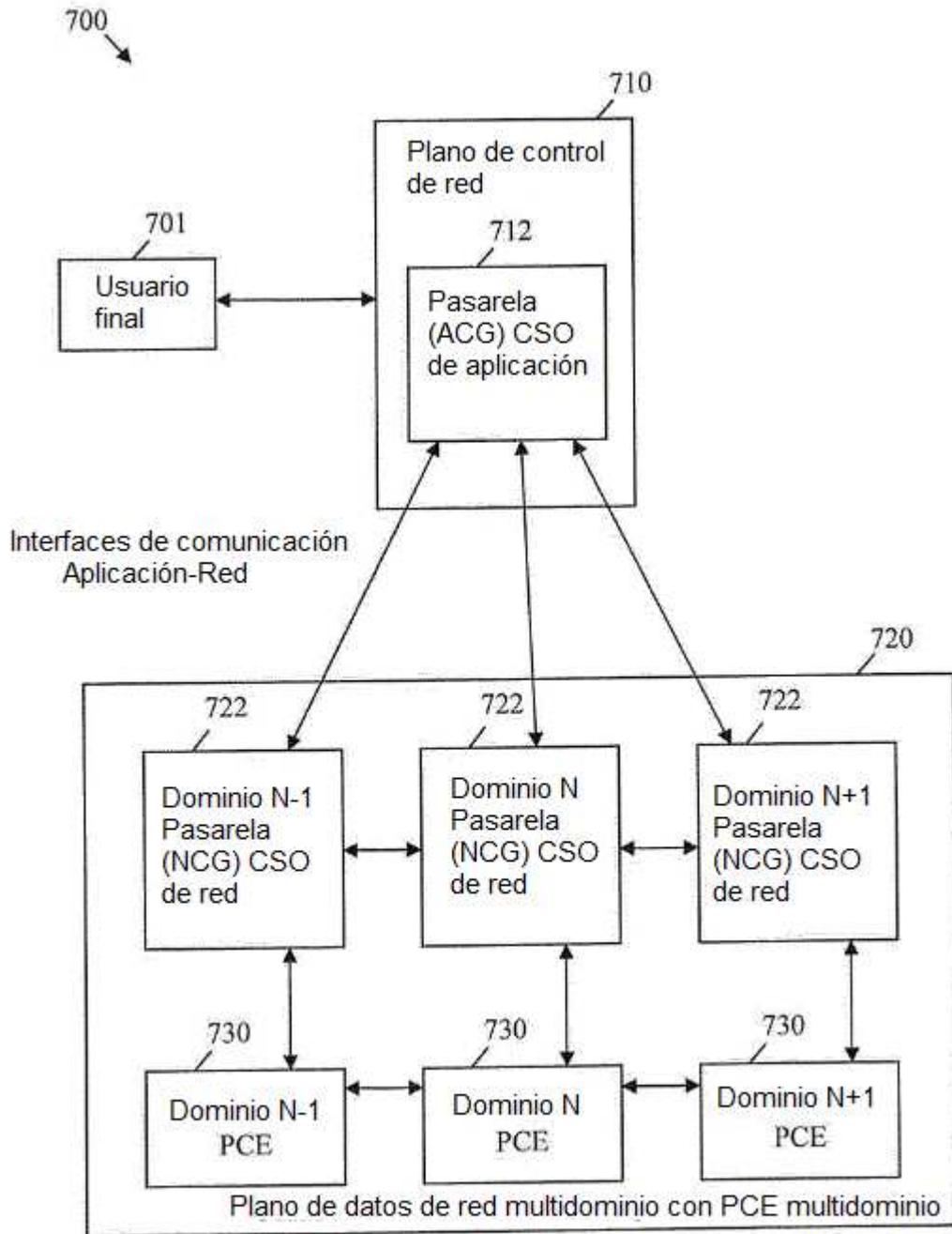


FIG. 7

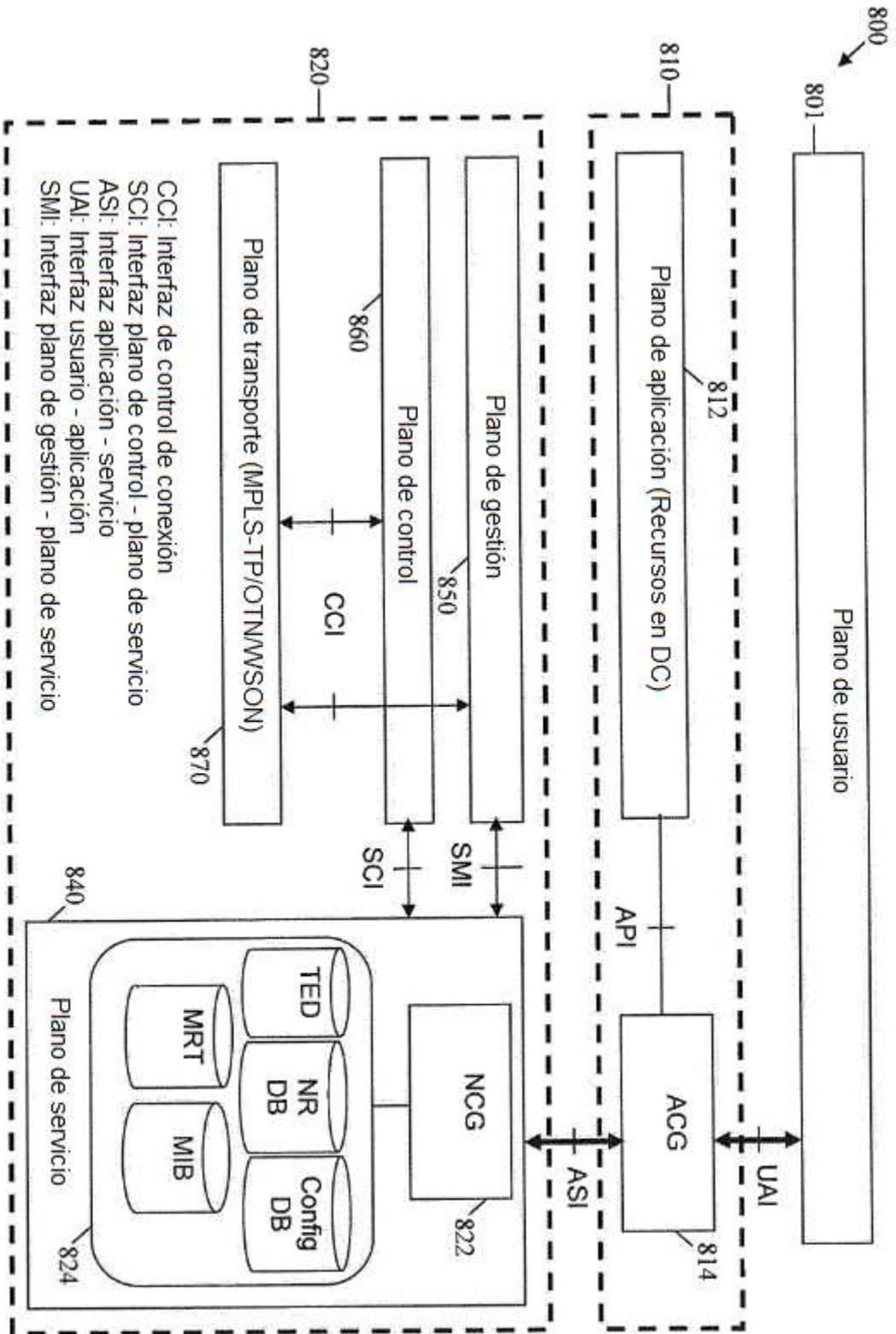


FIG. 8

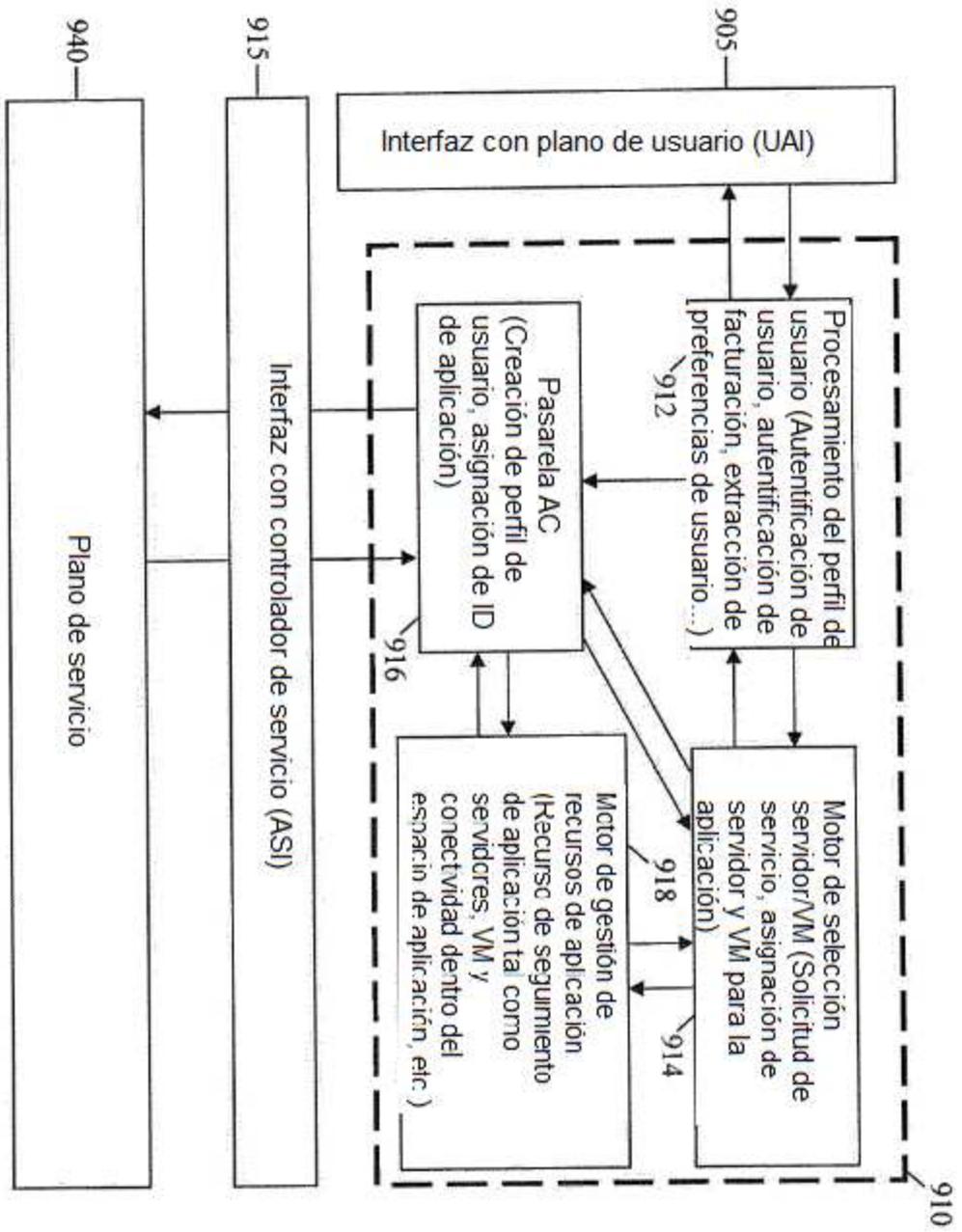


FIG. 9

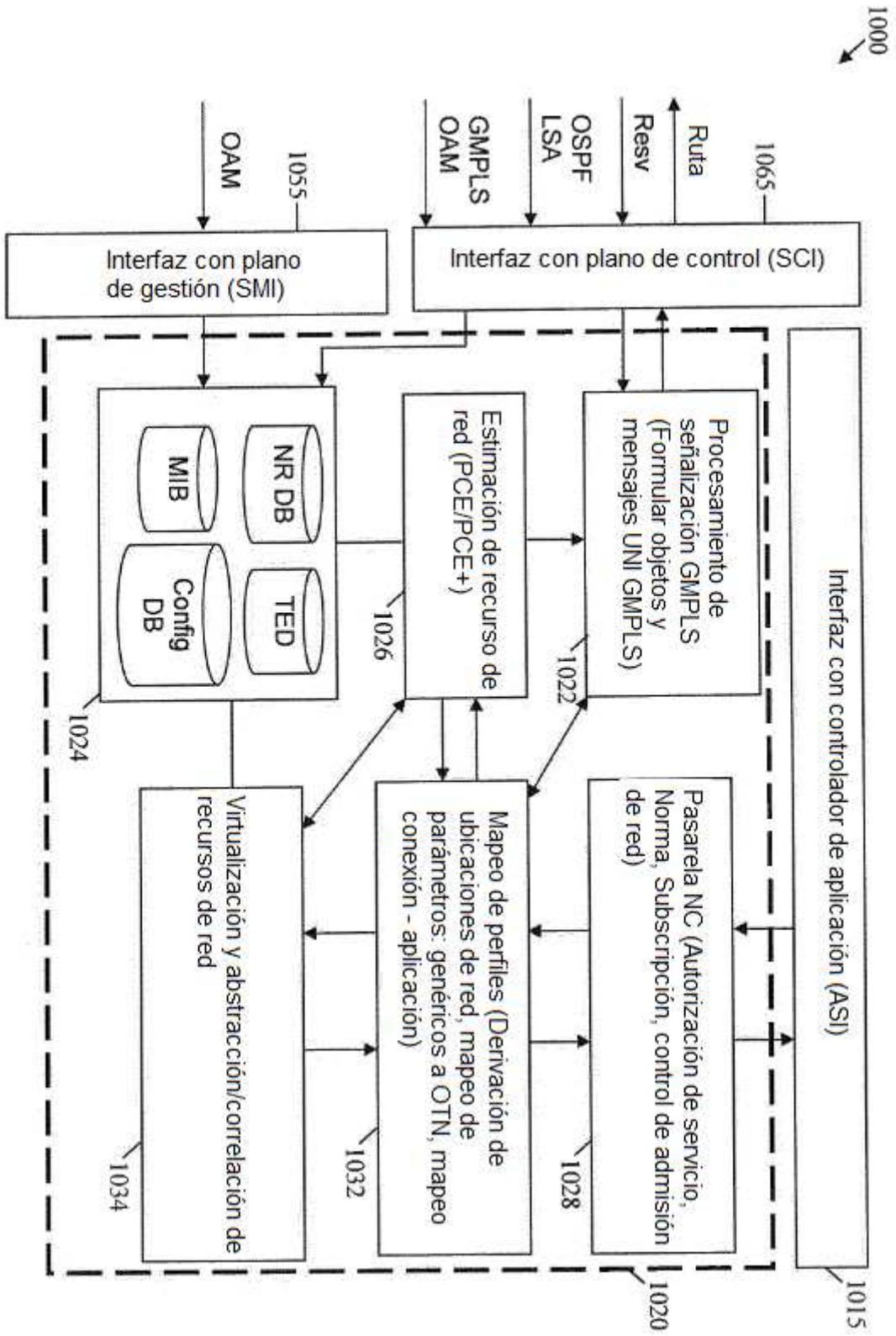


FIG. 10

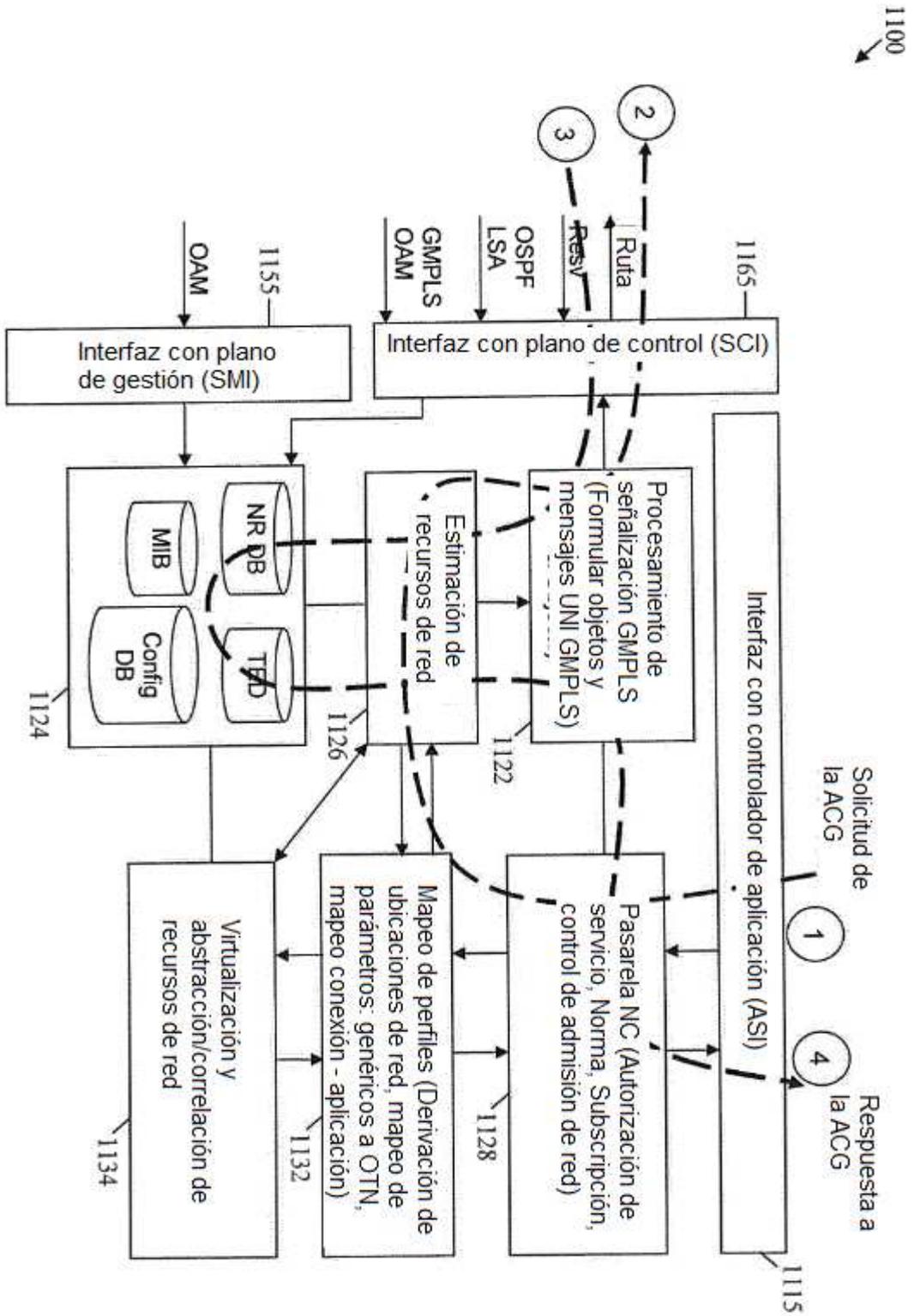


FIG. 11

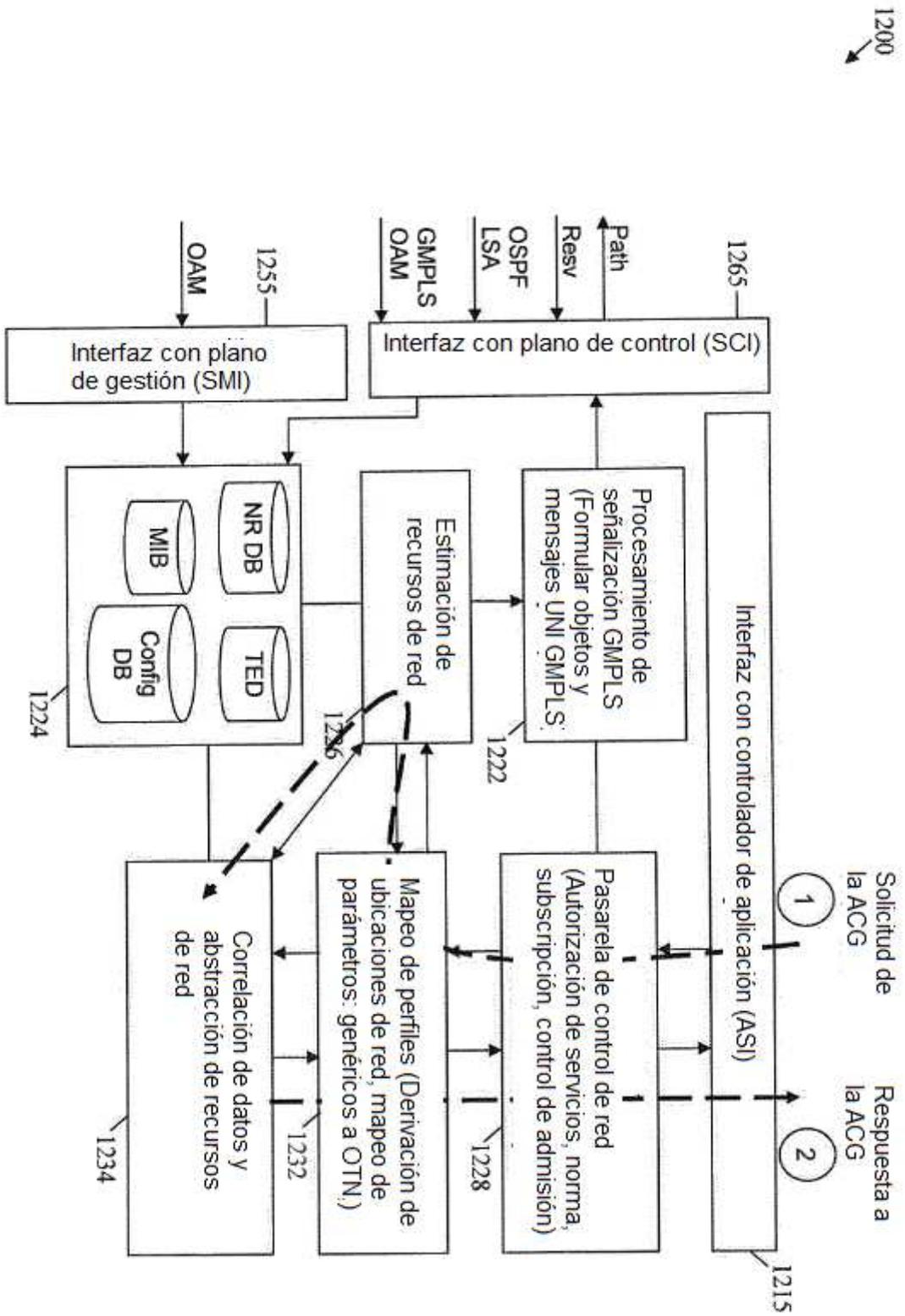


FIG. 12

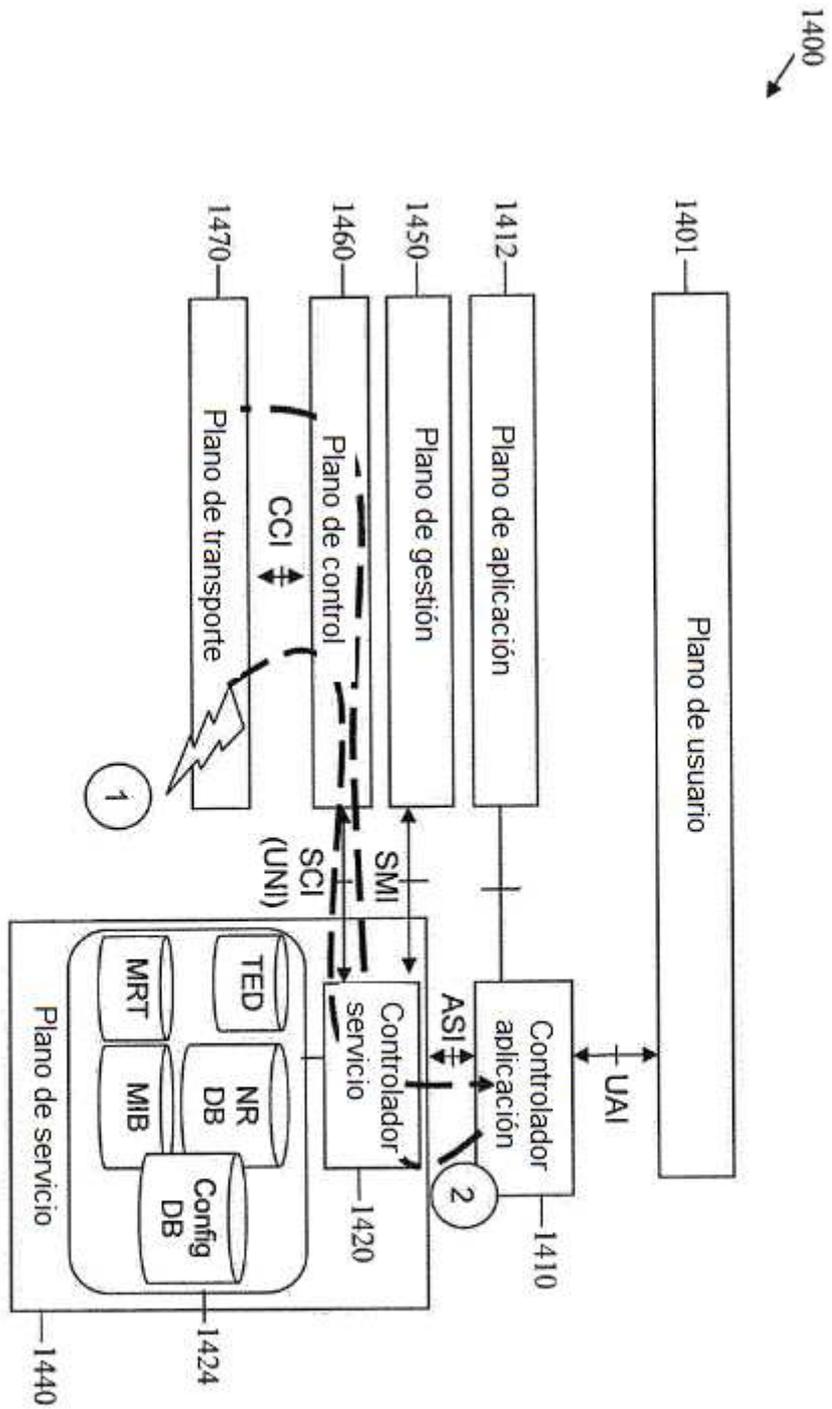


FIG. 14

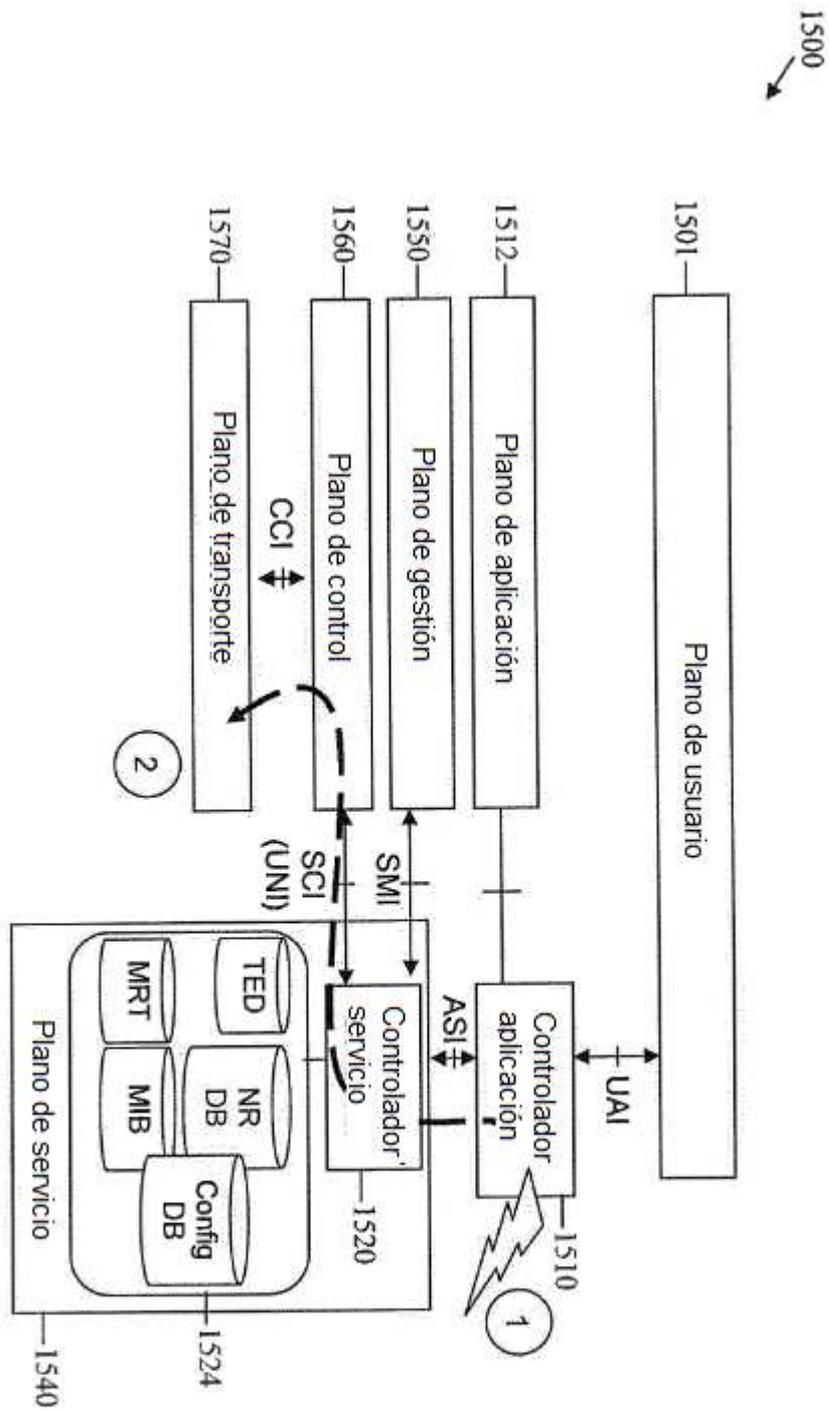


FIG. 15

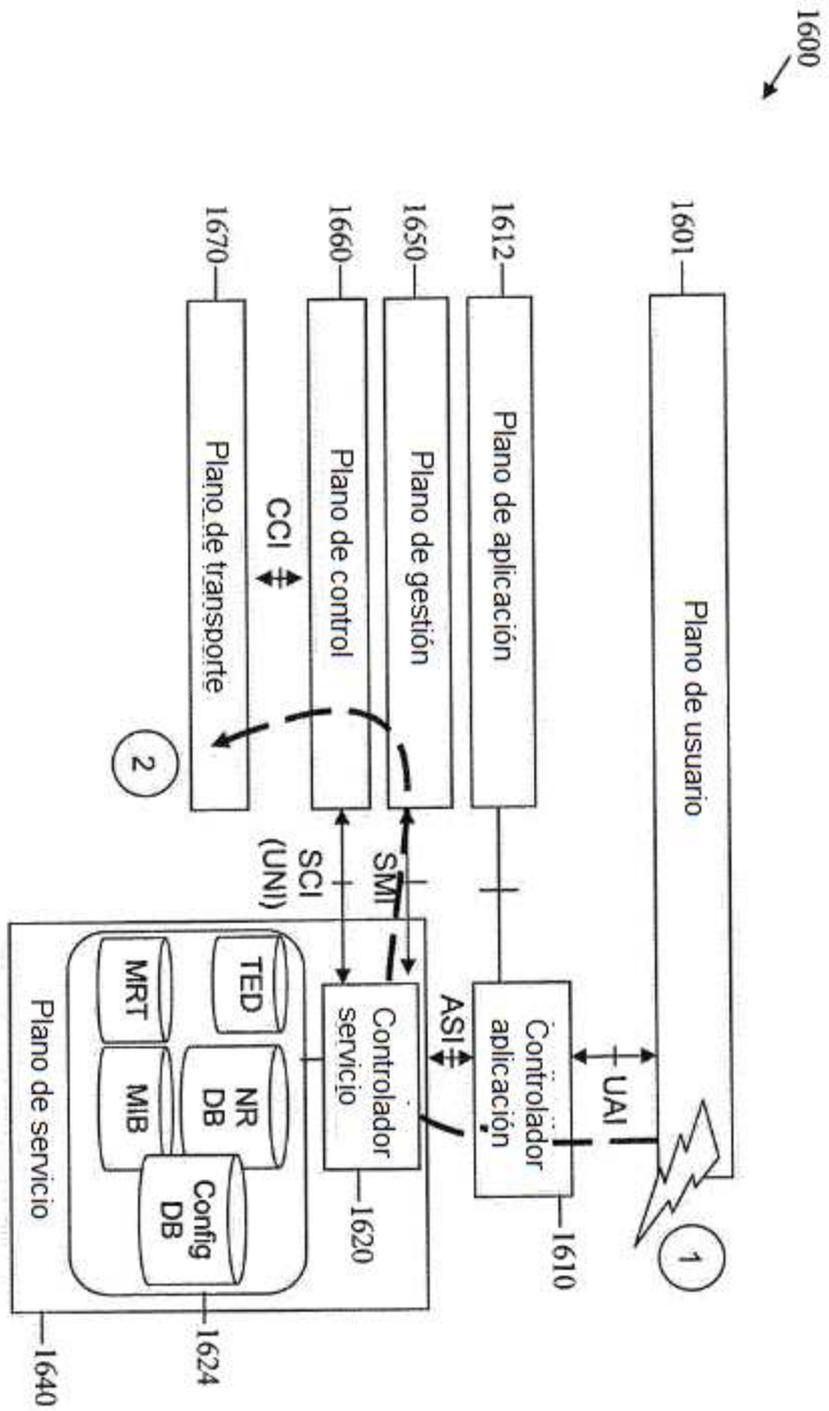


FIG. 16

1700

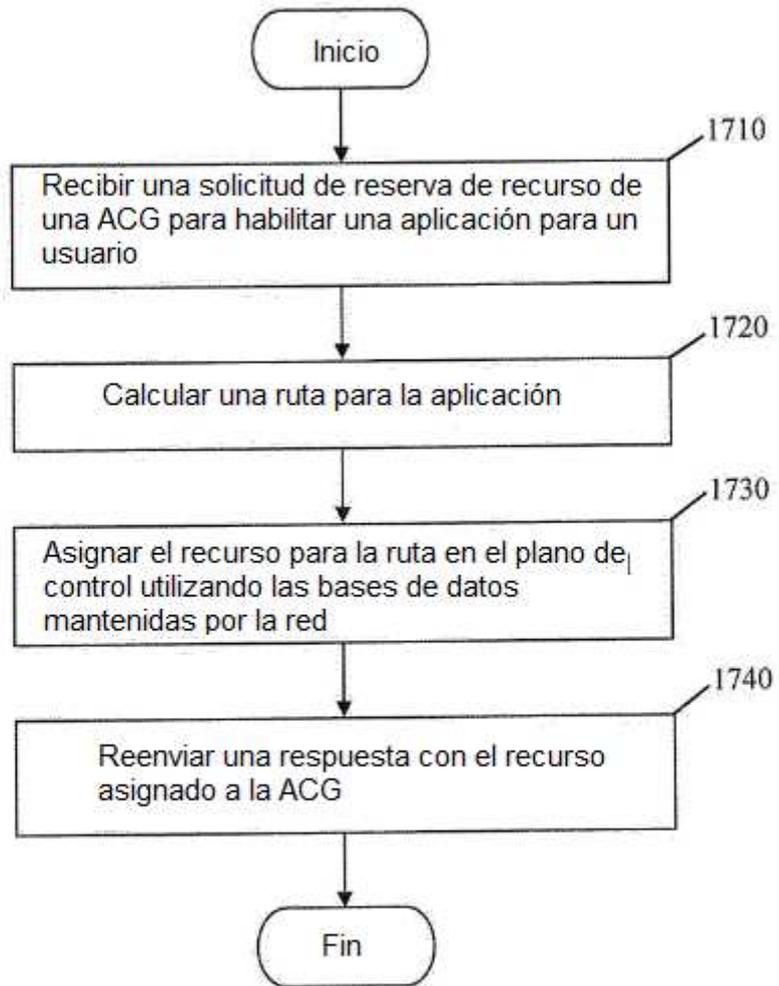


FIG. 17

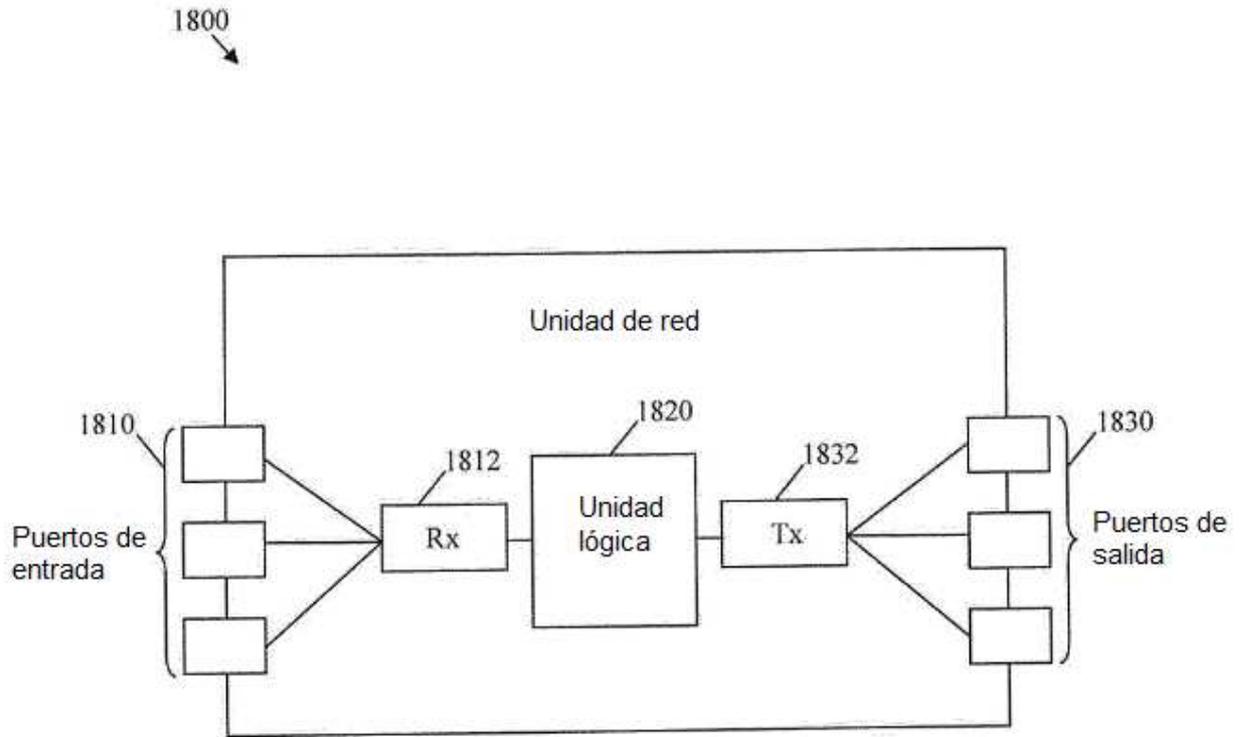


FIG. 18

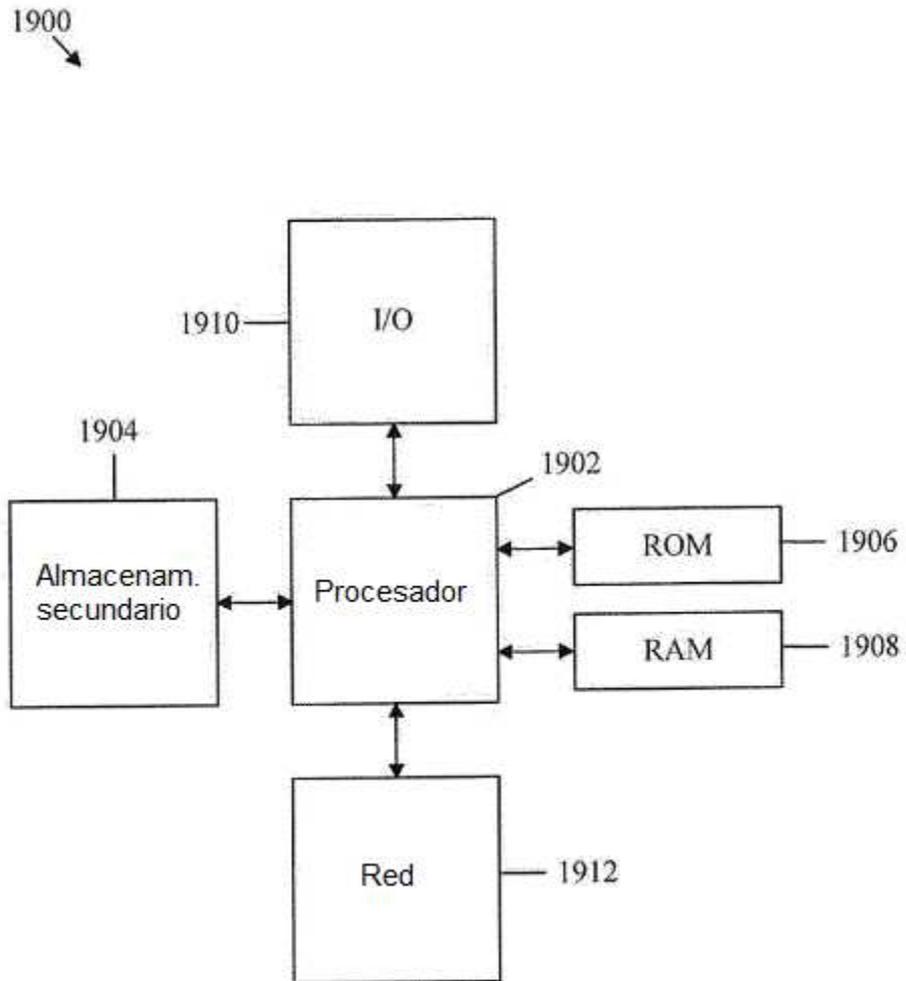


FIG. 19