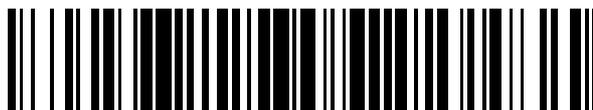


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 760 875**

51 Int. Cl.:

H04W 48/00 (2009.01)

H04W 28/08 (2009.01)

H04W 8/18 (2009.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **30.04.2015 PCT/US2015/028496**

87 Fecha y número de publicación internacional: **05.11.2015 WO15168394**

96 Fecha de presentación y número de la solicitud europea: **30.04.2015 E 15785224 (5)**

97 Fecha y número de publicación de la concesión europea: **11.09.2019 EP 3138307**

54 Título: **Optimización de extensión de capacidad en una red móvil**

30 Prioridad:

30.04.2014 US 201461986462 P

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

18.05.2020

73 Titular/es:

**AFFIRMED NETWORKS, INC. (100.0%)
35 Nagog Park, 1st Floor
Acton, MA 01720, US**

72 Inventor/es:

TO, PHILIP

74 Agente/Representante:

SÁEZ MAESO, Ana

ES 2 760 875 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

Optimización de extensión de capacidad en una red móvil

5 Referencia cruzada a solicitudes relacionadas

Campo técnico

10 Las realizaciones de la presente invención se refieren en general a métodos y aparatos informatizados para optimizar la extensión de capacidad en una red móvil.

Antecedentes

15 Los métodos tradicionales de aprovisionamiento de recursos en una red móvil han incluido agregar infraestructura física adicional cuando un recurso está a capacidad. El equipo físico está diseñado para tener relaciones de capacidad fija. Una vez que se agota una dimensión particular (por ejemplo, rendimiento, actividades de señalización, capacidad de sesión), un operador de red móvil no tiene más remedio que instalar más equipo a pesar de que todas las otras dimensiones pueden estar subutilizadas. Esto lleva a mayores gastos de capital y operativos.

20 El documento WO2012/175140 se refiere a la selección de puerta de enlace para el equilibrio de carga en una red.

El documento US 2014/0023029 se refiere a un dispositivo de puerta de enlace y a un método de selección de puerta de enlace.

25 "Network Functions Virtualisation (NFV); Virtualisation Requirements; NFV", vol. ISG_NFV, No. 1.1.1, 30 de octubre de 2013 (2013-10-30), páginas 1-17, XP055224747 es la especificación de requisitos de un marco de virtualización de funciones de red.

Resumen de la invención

30 En algunas realizaciones, se divulgan sistemas y métodos para optimizar la capacidad del equipo de red en redes móviles. En algunas realizaciones, un dispositivo informático recibe una identificación de usuario y un atributo de usuario, la identificación de usuario correspondiente a una característica del usuario de la red móvil, el atributo de usuario correspondiente a al menos una característica del uso de la red móvil por parte del usuario de la red móvil. En algunas realizaciones, el dispositivo informático genera una predicción de uso basada en la identificación del usuario y el atributo del usuario, la predicción de uso incluye información correspondiente al uso futuro anticipado de datos del usuario de la red móvil, el uso futuro anticipado de la red móvil correspondiente a al menos un recurso móvil. En algunas realizaciones, el dispositivo informático transmite la predicción de uso a una puerta de enlace de servicio (SGW) de modo que la SGW enruta al usuario de la red móvil a una de una puerta de enlace de red de paquetes de datos (PGW) heredada y una PGW de virtualización de funciones de red (NFV) basada en la predicción de uso, el PGW heredado que incluye una capacidad fija para al menos un recurso móvil y el PGF NFV que incluye una capacidad configurable para al menos un recurso móvil.

45 En algunas realizaciones, la al menos una característica del uso de la red móvil del usuario de la red móvil incluye la cantidad de uso anterior de la red móvil, un tiempo correlacionado con el uso de la red móvil, la ubicación de un dispositivo móvil correspondiente al usuario móvil, cantidad de tiempo dedicado a la itinerancia por el dispositivo móvil, marca y modelo del dispositivo móvil, aplicación instalada en el dispositivo móvil, sistema operativo y versión de firmware del dispositivo móvil, plan de suscripción, cuota restante e información demográfica. En algunas realizaciones, la al menos una característica del usuario de la red móvil incluye una ID de dispositivo móvil o un número de teléfono. En algunas realizaciones, en las que recibir el atributo de usuario comprende además recibir el atributo de usuario de al menos uno de un Servidor de abonado doméstico (HSS), una Entidad de gestión de movilidad (MME), un sistema de facturación y una puerta de enlace de Evolución de la arquitectura del sistema (SAE). En algunas realizaciones, el recurso móvil comprende al menos una de las actividades de señalización, rendimiento, ocupación de sesión, cifrado y transcodificación.

55 Estas y otras capacidades de la materia divulgada se entenderán más completamente después de una revisión de las siguientes figuras, descripción detallada y reivindicaciones. Debe entenderse que la fraseología y la terminología empleadas en este documento tienen fines descriptivos y no deben considerarse como limitantes.

60 Breve descripción de las figuras

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

65

La figura 1 es un diagrama que muestra las demandas en las que incurre un usuario de la red móvil en una red móvil, de acuerdo con algunas realizaciones de la presente divulgación.

La figura 2 es un diagrama que ilustra un método tradicional de extensión utilizando equipos heredados.

La figura 3 es un diagrama que ilustra el método de extensión utilizando equipos heredados y equipos basados en NFV, de acuerdo con algunas realizaciones de la presente divulgación.

La figura 4 es un diagrama del sistema que ilustra a un usuario en relación con una red móvil, de acuerdo con algunas realizaciones de la presente divulgación.

La figura 5 es un diagrama que ilustra un motor de predicción de uso, de acuerdo con algunas realizaciones de la presente divulgación.

La figura 6 es un diagrama del sistema de una red móvil, de acuerdo con algunas realizaciones de la presente divulgación.

La figura 7 es un diagrama del sistema que ilustra la optimización de la capacidad en una red móvil, de acuerdo con algunas realizaciones de la presente invención.

Descripción detallada

Una red móvil puede incluir usuarios móviles con características de uso muy diferentes. Por ejemplo, algunos usuarios móviles tienen muchos datos y consumen grandes cantidades de datos, lo que aumenta la cantidad total de rendimiento de datos que la red necesita soportar. Otros usuarios pueden hacer mucha señalización y hacer muchas conexiones (por ejemplo, utilizando aplicaciones móviles “habladoras” que envían actualizaciones con frecuencia) pero transfieren solo una pequeña cantidad de datos. Aunque solo usan una pequeña cantidad de datos, aumentan la cantidad de procesamiento de señalización requerida por la red. Sin embargo, algunos otros usuarios pueden estar relativamente inactivos tanto en el rendimiento de datos como en las dimensiones de señalización (por ejemplo, lectores de medidores de potencia en red), pero hay muchos de ellos y solo firmarlos en la red ocupa mucha capacidad de sesión. Para acomodar a todos los tipos de usuarios, un operador de red necesita desplegar suficiente equipo de red para cubrir el peor de los casos de todas estas dimensiones (por ejemplo, rendimiento, actividades de señalización, capacidad de sesión y quizás otras dimensiones). Dado que el equipo de red heredado está diseñado con una relación de capacidad fija (que admite X número de usuarios, cantidad Y de señalización y cantidad Z de rendimiento de datos), cubrir en el peor de los casos una dimensión que conducirá a la subutilización de los demás. Por ejemplo, una plataforma heredada implementada en la red puede estar alcanzando el 100% de la capacidad de la sesión, pero utilizando solo el 20% de la capacidad de rendimiento. Aunque todavía hay un exceso de capacidad de rendimiento, es necesario instalar nuevos equipos para aumentar el número de usuarios admitidos. Esto aumenta los costes operativos y de capital.

Anteriormente, una base de usuarios se segrega en aplicaciones separadas. Por ejemplo, los consumidores habituales están separados de los dispositivos de máquina a máquina. Los dispositivos se pueden clasificar de manera que los dispositivos tengan demandas similares. Los dispositivos con demandas similares pueden asignarse a equipos con diferentes características de rendimiento, que a menudo provienen de un fabricante diferente. Incluso con este enfoque, el problema de optimizar la capacidad no se resuelve debido al menos a lo siguiente: (1) la amplia categorización de usuarios no garantiza que los usuarios dentro de un grupo tengan demandas de uso similares. El equipo que sirve a una categoría de usuarios todavía puede ser subutilizado en algunas dimensiones; (2) los gastos de capital y operativos aumentarán, ya que los operadores ahora necesitan tratar potencialmente con equipos de múltiples proveedores, que pueden o no interactuar bien; (3) cuando las demandas de diferentes grupos de usuarios cambian con el tiempo, el operador necesitaría repartir los usuarios y reasignar los recursos de la red que pueden llevar mucho tiempo y ser costosos.

Las realizaciones preferidas de la presente divulgación incluyen el uso de la virtualización de funciones de red (NFV) en plataformas con diferentes capacidades y características de coste para manejar las demandas introducidas por diferentes tipos de usuarios. Configuradas de manera diferente (tanto en términos de hardware como de software), diferentes plataformas basadas en NFV pueden tener diferentes fortalezas y debilidades. Por ejemplo, se puede diseñar una plataforma basada en NFV para que pueda albergar a muchos usuarios (por ejemplo, mediante el uso de servidores con mucha memoria), pero con un rendimiento limitado y capacidad de señalización. Se puede diseñar otra plataforma basada en NFV para que pueda procesar una gran cantidad de rendimiento (por ejemplo, mediante el uso de tarjetas adaptadoras de red especializadas). Sin embargo, se puede diseñar otra plataforma basada en NFV para que pueda manejar muchas señales (por ejemplo, con una CPU de alta potencia). Estas plataformas basadas en NFV con diferentes características se pueden agrupar en una red para satisfacer las diferentes demandas introducidas por los diferentes tipos de usuarios. Para maximizar la efectividad y minimizar los costes, los usuarios de diferentes características son dirigidos a los servidores con la potencia correspondiente para que cada servidor sea mejor utilizado. De esta manera, la fortaleza del legado y las plataformas basadas en NFV complementan las debilidades

del otro. Por lo tanto, el equipo de red se puede utilizar mejor, lo que conduce a un menor coste general de capital y operación.

Las realizaciones preferidas de la presente divulgación incluyen una función para clasificar y dirigir a los usuarios móviles o suscriptores a diferentes equipos de red en función de las características de uso pasadas y futuras previstas para que coincidan con las características de capacidad del equipo de red. No es necesario segregar a los usuarios en diferentes grupos (por ejemplo, separar a los usuarios en diferentes nombres de puntos de acceso o APN). La red puede parecer perfecta para los usuarios finales y, por lo tanto, hay pocos cambios en la experiencia del usuario. Los operadores pueden hacer uso de NFV para implementar plataformas con diferentes características de coste y rendimiento. NFV es adecuado para tales aplicaciones, ya que permite que la misma función de red se ejecute en diferentes plataformas de hardware. Estas plataformas de hardware van desde chasis de servidores blade altamente sofisticados hasta cajas de servidores de bajo costo, que proporcionan un rendimiento y una capacidad diferentes.

En algunos casos, se puede obtener capacidad adicional construyendo un servidor con hardware especializado, como un chip para cifrado de hardware, para soportar ciertos grupos de usuarios.

Las realizaciones preferidas de la presente divulgación se pueden usar como una solución de campo verde (por ejemplo, una nueva red compuesta solo por plataformas basadas en NFV), o para complementar una red heredada existente que se está quedando sin capacidad. En el último caso, los servidores NFV pueden diseñarse para aliviar específicamente los cuellos de botella del equipo heredado y hacer que la utilización de todas las dimensiones de rendimiento sea más equilibrada. A continuación, se describen técnicas para determinar el punto de estrangulamiento de las plataformas existentes, construir plataformas basadas en NFV para aliviar estos puntos de estrangulamiento, predecir e identificar las características de uso de los usuarios y dirigirlos a las plataformas basadas en NFV que puede manejar mejor las demandas.

Las realizaciones preferidas de la presente divulgación hacen uso de plataformas basadas en virtualización de funciones de red con diferentes capacidades y características de coste para complementar el equipo heredado existente que tiene relaciones de capacidad fija. Las plataformas heredadas y basadas en NFV pueden funcionar sin problemas como una sola red. En algunas realizaciones, las plataformas basadas en NFV están diseñadas para complementar la debilidad de la plataforma heredada, de modo que cuando trabajan juntas, reduce la posibilidad de sobrecargar una determinada dimensión de capacidad y, en general, todos los nodos de la red se utilizarán mejor.

En algunas realizaciones, para hacer el mejor uso de las diferentes capacidades de las plataformas heredadas y basadas en NFV, cuando un usuario intenta acceder a la red, las características de uso del usuario se predicen en función de una serie de factores que incluyen su patrón de uso pasado. Luego, el usuario se dirige a los nodos de red que mejor pueden manejar las demandas del usuario.

La figura 1 es un diagrama que muestra las demandas en las que incurre un usuario de la red móvil en una red móvil, de acuerdo con algunas realizaciones de la presente divulgación. La figura 1 muestra un dispositivo 101 móvil, actividades 102 de señalización, rendimiento 103, ocupación 104 de sesión y otras dimensiones 105.

Como se muestra en la Figura 1, un usuario de la red 101 móvil incurre en demandas en la red móvil en varias dimensiones diferentes. El usuario genera actividades 102 de señalización cuando se registra y se da de baja de la red, cuando se encuentra en itinerancia alrededor de la red, etc. El usuario impone demandas en la dimensión 103 de rendimiento cuando navega por páginas web o envía actualizaciones de estado. El usuario también ocupa uno o más espacios 104 de sesión cuando está conectado a la red. Finalmente, existen demandas en otras dimensiones 105, por ejemplo, si el usuario requiere servicios de transcodificación de imagen/video o cifrado. No todos los usuarios se comportan de la misma manera. Un usuario pesado de datos consume muchos datos y aumenta la demanda en la dimensión de rendimiento. Otros usuarios podrían estar haciendo mucha señalización y hacer muchas conexiones (por ejemplo, utilizando aplicaciones "habladoras" que envían muchas actualizaciones) pero transfiriendo solo una pequeña cantidad de datos. Algunos usuarios pueden aumentar las demandas en la dimensión de señalización en la red. Sin embargo, algunos otros usuarios pueden estar relativamente inactivos tanto en el rendimiento de datos como en las dimensiones de señalización (por ejemplo, lectores de medidores de potencia en red), pero hay muchos de ellos y solo mantenerlos conectados a la red requiere mucha capacidad de sesión.

Los operadores de red a menudo tienen que instalar más equipos de red para manejar las demandas agregadas en las diferentes dimensiones descritas anteriormente. Dado que el equipo de red heredado está diseñado para admitir una relación de capacidad fija (que admite X número de usuarios, cantidad Y de señalización y cantidad Z de rendimiento de datos), cubrir el peor de los casos de una dimensión a menudo conduce a la subutilización de los demás. Por ejemplo, un nodo de red puede alcanzar el 100% de la capacidad de la sesión, pero utilizar solo el 50% del rendimiento. Aunque todavía hay un exceso de capacidad de rendimiento, se instalan nuevos equipos heredados para aumentar el número de usuarios admitidos. La instalación de nuevos equipos heredados puede aumentar los costes de capital y operativos.

La figura 2 es un diagrama que ilustra un método tradicional de extensión utilizando equipos heredados. La figura 2 muestra un operador de red que alcanza la capacidad máxima de sesión con una primera plataforma 201 heredada,

la capacidad de una primera plataforma 202 heredada después de una extensión 210 2x y la capacidad de una segunda plataforma 203 heredada después de una extensión 210 2x.

Como se muestra en 201, un operador de red alcanza la capacidad máxima en una primera plataforma heredada. La primera plataforma heredada tiene un rendimiento máximo de 100 unidades y un máximo de 100 sesiones. Se usan 50 de las 100 unidades de rendimiento, mientras que se usan 100 de las 100 sesiones. Cuando el operador de la red anticipa una duplicación de la demanda (por ejemplo, 100 unidades de rendimiento y 200 sesiones), el operador de la red tiene que encontrar una manera de aumentar la capacidad. Para duplicar la capacidad 210, el operador de red instala una segunda plataforma heredada. En algunas realizaciones, un operador puede determinar el uso de la capacidad de una plataforma monitoreando los niveles de uso pico de un dispositivo (por ejemplo, monitoreando el uso durante las horas ocupadas). Una plataforma puede especificar el valor máximo para cada dimensión (por ejemplo, 10 millones de sesiones, 50 Gbps de rendimiento en un límite de CPU como el 80%). Por ejemplo, para determinar la cantidad de uso de la sesión, un operador puede usar un contador de estadísticas para ver cuántas sesiones se usan durante una hora ocupada. Como otro ejemplo, un operador puede determinar una cantidad de rendimiento midiendo el rendimiento en un límite de CPU específico durante una hora ocupada. Un operador puede determinar la capacidad midiendo una cantidad de CPU durante uso. Tanto en la primera plataforma 202 heredada como en la segunda plataforma 203 heredada, se usan 50 de las 100 unidades de rendimiento, mientras que se usan 100 de las 100 sesiones. Después de la extensión, la dimensión de la sesión sigue bloqueando ambas plataformas heredadas. La relación de capacidad en la plataforma heredada expandida (por ejemplo, igual capacidad para sesión y rendimiento) no coincide con la demanda de los usuarios (por ejemplo, muchas sesiones, pero no tanto rendimiento).

En contraste, la realización preferida de la presente invención requiere comprender la causa del cuello de botella de las plataformas heredadas existentes, el patrón de uso presente y futuro de los usuarios, y construir plataformas basadas en NFV para complementar las plataformas heredadas de modo que se pueden utilizar mejor todas las dimensiones de capacidad.

La figura 3 es un diagrama que ilustra el método de extensión utilizando equipos heredados y equipos basados en NFV, de acuerdo con algunas realizaciones de la presente divulgación. La figura 3 muestra un operador de red que alcanza una capacidad de sesión máxima con una primera plataforma 201 heredada, capacidad de una primera plataforma 302 heredada después de una extensión 310 2x, y capacidad de una segunda plataforma 303 basada en NFV después de una extensión 310 2x. Mientras que la figura 3 ilustra la extensión en dos dimensiones (por ejemplo, sesión y rendimiento), se puede aplicar una técnica similar a cualquier cantidad de dimensiones.

Como se describió anteriormente, un operador ha alcanzado la capacidad en una primera plataforma heredada con 50 de 100 unidades de rendimiento utilizadas y 100 de 100 sesiones utilizadas 201. Cuando el operador de red anticipa una duplicación de la demanda (por ejemplo, 100 unidades de rendimiento y 200 sesiones), el operador de red duplica la capacidad 210 instalando una plataforma 303 basada en NFV. Como se muestra después de la extensión 310, la combinación de la primera plataforma 302 heredada y la plataforma 303 basada en NFV toma en consideración los patrones de uso presentes y futuros de los usuarios. Por ejemplo, si el 20% de los usuarios está utilizando el 80% del rendimiento, significa que de los 100 usuarios:

- 20 usuarios pesados están utilizando 40 unidades de rendimiento; y
- 80 usuarios livianos utilizan 10 unidades de rendimiento.

Cuando la demanda se duplica, hay un total de 200 usuarios, de los cuales

- 40 usuarios pesados usan 80 unidades de rendimiento; y
- 160 usuarios livianos utilizan 20 unidades de rendimiento.

Una plataforma basada en NFV puede construirse para soportar 200 usuarios, pero solo soporta 40 unidades de rendimiento, probablemente a una fracción del coste en comparación con la plataforma heredada. Esto se puede hacer debido a la naturaleza flexible de las soluciones NFV: se puede construir una plataforma con mucha memoria para admitir más sesiones, pero solo una CPU moderadamente potente para el procesamiento de rendimiento para reducir los costes. Los 160 usuarios livianos pueden dirigirse a la plataforma 303 basada en NFV, mientras que los usuarios pesados pueden dirigirse a la plataforma 302 heredada. Si la plataforma heredada cuesta \$ 1M y la plataforma NFV cuesta \$ 0.2M, entonces el coste por duplicar la capacidad sería :

- \$ 2M si solo se utilizan plataformas heredadas (por ejemplo, como se muestra en la figura 2); y
- \$ 1.2M si se agrega una plataforma basada en NFV a la plataforma heredada (por ejemplo, como se muestra en la figura 3).

El uso de plataformas basadas en NFV puede ahorrar \$ 0.8M o el 40% del coste de usar solo plataformas heredadas. Como se describe en las Figs. 2 y 3, la plataforma heredada tiene una alta capacidad de rendimiento, pero no suficiente

capacidad de sesión. La plataforma basada en NFV complementa la plataforma heredada al ofrecer una alta capacidad de sesión, pero ofreciendo una capacidad de bajo rendimiento para mantener bajos los costes. Hay muchas formas diferentes de construir la plataforma basada en NFV para complementar las plataformas heredadas. Los operadores pueden decidir sobre las compensaciones de coste y rendimiento de diferentes componentes como memoria, CPU u otros chips especializados, así como también cómo cambiarán las demandas futuras. El uso de una plataforma basada en NFV permite a un operador analizar las demandas de diferentes usuarios, construir plataformas basadas en NFV con capacidades que complementan las plataformas heredadas y dirigir a los usuarios de manera apropiada a las diferentes plataformas para hacer el mejor uso de la capacidad en todas las plataformas.

En algunas realizaciones, los sistemas y métodos descritos en este documento dirigen y clasifican a los usuarios en base a demandas pasadas y predichas. La predicción de las demandas de capacidad del usuario puede ayudar a equilibrar el uso de la capacidad tanto en la plataforma heredada como en las plataformas basadas en NFV.

La figura 4 es un diagrama del sistema que ilustra a un usuario en relación con una red móvil, de acuerdo con algunas realizaciones de la presente divulgación. La figura 4 muestra el usuario 401 de la red móvil, el clasificador 402, el motor 403 de predicción de uso, la plataforma 404 de red heredada, la plataforma 405 basada en NFV y la red 406 móvil.

El usuario 401 de red móvil puede incluir suscriptores de red móvil que acceden a la red 406 móvil a través de uno o más dispositivos de red móvil (por ejemplo, teléfonos inteligentes, ordenadores portátiles, tabletas). Como se describe con más detalle a continuación, la red 406 móvil comprende una pluralidad de dispositivos de red. Brevemente, el dispositivo de red en la red 406 móvil puede enrutar y analizar el tráfico de usuarios.

Cuando un usuario 401 inicia sesión en la red 406, un clasificador 402 consulta el motor 403 de predicción de uso para predecir un patrón de uso de recursos. El clasificador 402 es un componente que toma información del usuario y su equipo (por ejemplo, El identificador del equipo móvil), consulta el motor 403 de predicción de uso y toma la decisión sobre qué plataforma colocar al usuario en la red 406 móvil. Clasificador 402 puede implementarse como un componente separado o como parte de un determinado dispositivo de red (por ejemplo, en el equilibrador de carga) en la red móvil. El motor 403 de predicción de uso, que se describe con más detalle a continuación, es un componente que toma la identificación del usuario y otros atributos relacionados con el usuario, y predice el uso futuro de los recursos de red del usuario. Basado en un resultado del motor 403 de predicción de uso, se dirige al usuario 401 para que sea atendido por la plataforma 404 de red heredada o una plataforma 405 basada en NFV. Como se describió anteriormente, el clasificador también recibe una entrada de las plataformas heredadas y basadas en NFV correspondientes a sus nivel de capacidad disponible y sus capacidades (por ejemplo, cifrado, transcodificación de vídeo).

En algunas realizaciones, un usuario puede ser dirigido a una plataforma heredada o una plataforma basada en NFV en función de las características del usuario o de la plataforma. Por ejemplo, un usuario puede ser dirigido cuando el usuario se une a una red (por ejemplo, cuando el usuario enciende el teléfono por la mañana). Además, los usuarios existentes también pueden migrarse activamente de un sistema a otro si la carga del sistema existente alcanza un cierto umbral o si las características del usuario cambian significativamente.

La figura 5 es un diagrama que ilustra un motor de predicción de uso, de acuerdo con algunas realizaciones de la presente divulgación. La figura 5 muestra la identificación 501 de usuario, el motor 502 de predicción de uso, la predicción 503 de uso, el patrón y tendencia 504 de uso pasado, la información 505 temporal, la ubicación 506 del usuario, el patrón 507 de movilidad pasado, la marca y el modelo del dispositivo 508 móvil, las aplicaciones 509 instaladas, el sistema operativo (OS) y la versión de firmware 510, plan 511 de suscripción, cuota 512 restante e información 513 demográfica .

El motor 502 de predicción de uso recibe la identificación 501 de usuario y los atributos 504-513 de usuario. Como se describe con más detalle a continuación, el motor 502 de predicción de usuario predice las demandas de uso futuro del usuario 503 en función de las entradas. La identificación 501 de usuario corresponde a la información sobre el dispositivo móvil de un usuario (por ejemplo, Identidad Internacional de Equipo Móvil (IMEI)). Los atributos de usuario se pueden recopilar de varios componentes en una red móvil, como se describe con más detalle en la figura 6.

Los atributos 504-513 de usuario incluyen, pero no se limitan a:

(1) El patrón de uso pasado y la tendencia del usuario 504: un usuario con gran cantidad de datos es probable que tenga muchos datos en el futuro.

(2) Hora del día, día de la semana y fecha del año 505: la información temporal proporciona pistas sobre qué servicio utiliza el usuario en el dispositivo móvil. La ocurrencia de cualquier evento masivo (por ejemplo, Super Bowl) también puede ser útil para predecir el patrón de uso del usuario.

(3) La ubicación 506 del usuario: similar a la información temporal, la información de ubicación geográfica puede ser útil para predecir el patrón de uso. Por ejemplo, si el usuario se encuentra en la ciudad donde hay más sitios de células

de menor tamaño, es probable que el usuario experimente una mayor cantidad de eventos de transferencia a medida que avanza y retrocede entre los sitios de células. Mientras que, si el usuario se encuentra en un área suburbana, es probable que el sitio de una celda cubra un área más grande y la posibilidad de transferencias sea menor.

5 (4) El patrón 507 de movilidad pasado: un usuario que deambula mucho en el pasado probablemente lo hará en el futuro.

10 (5) Marca y modelo del dispositivo 508 móvil: a veces, tipos específicos de dispositivos móviles tienen usos de recursos muy diferentes. Por ejemplo, un usuario con un teléfono con pantalla táctil utilizará más servicios de datos en comparación con el usuario de un teléfono con función sin soporte de pantalla táctil.

(6) Aplicaciones 509 móviles instaladas: algunas aplicaciones móviles son más “comunicativas” que otras y desencadenan muchas más conexiones.

15 (7) Sistema operativo y versión de firmware del dispositivo 510 móvil: las demandas pueden ser diferentes con diferentes versiones del sistema operativo. Por ejemplo, la aplicación de mensajería en Apple iOS 8 admite voz y video además de texto. Lo más probable es que se traduzca en un mayor uso de rendimiento.

20 (8) Plan de suscripción del usuario 511- Por ejemplo, un usuario con un límite de datos bajo usará menos datos que uno con un límite de datos grande.

(9) La cuota restante para el período 512 de facturación actual: Por ejemplo, es probable que un usuario con una cuota restante baja esté más restringido en el uso del ancho de banda que uno con una cuota restante abundante.

25 (10) El perfil demográfico del usuario 513: Por ejemplo, es probable que el comportamiento de uso sea drásticamente diferente entre un usuario adolescente y un usuario adulto. Es probable que un usuario adolescente consuma más datos a través de sus actividades sociales, mientras que un usuario adulto puede usar más llamadas de voz que datos en sus actividades diarias.

30 La figura 6 es un diagrama del sistema de una red móvil, de acuerdo con algunas realizaciones de la presente divulgación. La figura 6 muestra el Servidor de Abonado Doméstico (HSS) 601, Entidad de Gestión de Movilidad (MME) 602, sistema 603 de facturación, eNodeB 604, Puerta 605 de enlace de Evolución de Arquitectura de Sistema (SAE) y Servidor 606 Analítico. Todos los elementos que se muestran en la figura 6 puede ser heredado o virtual.

35 En algunas realizaciones, algunos operadores pueden tener un Servidor 606 de Análisis para recopilar y analizar estadísticas de uso sobre los usuarios. Esta información se puede alimentar directamente al motor de predicción. En otras realizaciones, el motor de predicción contiene capacidades analíticas del servidor 606 de análisis, y los dos componentes se integran en una unidad.

40 El Servidor 601 de Abonado Doméstico (HSS) contiene información sobre la movilidad del usuario. La información de movilidad se puede alimentar periódicamente a un servidor 606 de análisis para calcular el patrón de movilidad pasado del usuario.

45 La Entidad 602 de Gestión de Movilidad (MME) rastrea la ubicación actual de un dispositivo y puede enviar información de ubicación al servidor 606 de análisis para su posterior procesamiento.

El sistema de facturación contiene un plan de suscripción, cuotas restantes y otra información relacionada con la facturación del usuario. La información de facturación se puede alimentar al servidor 606 de análisis para determinar la tendencia de uso.

50 Las puertas 605 de enlace SAE pueden examinar todo el tráfico hacia y desde el usuario. Mediante el uso de técnicas de inspección profunda de paquetes (DPI), la información de uso se puede extraer del tráfico de datos, incluida la marca y el modelo del dispositivo, las aplicaciones instaladas y más utilizadas, el sistema operativo y la versión de firmware, etc. En algunas realizaciones, los datos DPI se introducen en el servidor 606 de análisis para un análisis posterior antes de ser utilizado por el motor de predicción.

55 En algunas realizaciones, la tendencia de uso de un usuario cambia lentamente con el tiempo. Cuando las tendencias de uso cambian lentamente, el motor de predicción de uso no necesita actualizar su predicción para un usuario en tiempo real. Por ejemplo, la predicción para un usuario en particular se puede actualizar una vez por semana, y para diferentes usuarios se puede usar un intervalo diferente. En algunas realizaciones, las tendencias de uso cambian más rápidamente. Por ejemplo, cuando ocurren ciertos eventos, la predicción puede activarse para actualizarse bajo demanda. Por ejemplo, la predicción puede actualizarse inmediatamente si el usuario cambió a otro plan de suscripción o a un nuevo teléfono.

65 Cuando un nuevo suscriptor se une a la red, no habrá mucho historial de uso para construir predicciones. Inicialmente, los nuevos usuarios pueden ser tratados como un usuario “promedio” con un rendimiento promedio y cargas de

señalización. Alternativamente, se pueden hacer predicciones basadas en la cantidad limitada de información disponible. Por ejemplo, si el nuevo suscriptor es un adolescente, es probable que tenga más aplicaciones habladoras como Facebook, Instagram o Snapchat, lo que generará más carga de señalización. Si, por el contrario, el nuevo suscriptor es una cuenta comercial que se ha inscrito para la prueba, es probable que sea un usuario de datos más pesado. La frecuencia de actualización de la predicción para nuevos usuarios puede ser mayor, de modo que la predicción puede converger rápidamente en función de los factores recién adquiridos. En esta etapa, el usuario puede ponerse en el sistema heredado o en el sistema NFV. Una vez que se clasifica al usuario, se puede mover entre los sistemas heredados y NFV para un uso óptimo de los recursos de la red.

La figura 7 es un diagrama del sistema que ilustra la optimización de la capacidad en una red móvil, de acuerdo con algunas realizaciones de la presente invención. La figura 7 muestra el motor 701 de predicción, Puerta 702 de Enlace de Servicio (SGW), Puerta 703 de Enlace de Red de Datos de Paquetes Heredado (PGW) y PGW 704 basado en NFV.

Cuando un suscriptor enciende el teléfono, el teléfono intentará establecer una sesión con la red móvil. La solicitud eventualmente se enviará a SGW 701 y el SGW 701 selecciona un PGW 703 704 para albergar la sesión del usuario. Uno de los nodos de PGW incluye el equipo 703 heredado y el otro nodo de PGW incluye una plataforma 704 basada en NFV. Normalmente, SGW selecciona un PGW basado únicamente en el nombre del punto de acceso (APN). El APN identifica la red de paquetes de datos (PDN) con la que un usuario de datos móviles desea comunicarse y se le asigna a un usuario cuando activa su plan de suscripción. En realizaciones preferidas, SGW consulta el motor de predicción para determinar el mejor lugar para albergar la sesión del usuario en función de las características del usuario. Por ejemplo, el motor de clasificación/predicción puede proporcionar una API basada en el protocolo de acceso a objetos simple (SOAP) o transferencia de estado representativo (REST) en el que SGW puede llamar para obtener una decisión sobre dónde configurar la sesión. Una vez que el SGW decida configurar la sesión en, por ejemplo, el PGW basado en NFV, toda la señalización futura y el tráfico de datos relacionados con el suscriptor serán manejados por el PGW seleccionado.

El tema descrito en este documento puede implementarse en circuitos electrónicos digitales, o en software de ordenador, firmware o hardware, incluidos los medios estructurales divulgados en esta especificación y sus equivalentes estructurales, o en combinaciones de ellos. El tema descrito en este documento puede implementarse como uno o más productos de programas de ordenador, como uno o más programas de ordenador incorporados de manera tangible en un soporte de información (por ejemplo, en un dispositivo de almacenamiento legible por máquina), o incorporados en una señal propagada, para ejecución por o para controlar el funcionamiento de un aparato de procesamiento de datos (por ejemplo, un procesador programable, un ordenador o varios ordenadores). Un programa de ordenador (también conocido como programa, software, aplicación de software o código) se puede escribir en cualquier forma de lenguaje de programación, incluidos los lenguajes compilados o interpretados, y se puede implementar en cualquier forma, incluso como un programa independiente o como un módulo, componente, subrutina u otra unidad adecuada para su uso en un entorno informático. Un programa de ordenador no necesariamente corresponde a un archivo. Un programa puede almacenarse en una parte de un archivo que contiene otros programas o datos, en un único archivo dedicado al programa en cuestión o en múltiples archivos coordinados (por ejemplo, archivos que almacenan uno o más módulos, subprogramas o partes de código). Un programa de ordenador puede implementarse para ejecutarse en un ordenador o en varios ordenadores en un sitio o distribuirse en múltiples sitios e interconectarse mediante una red de comunicación.

Los procesos y flujos lógicos descritos en esta especificación, que incluyen los pasos del método de la materia objeto aquí descrita, pueden ser realizados por uno o más procesadores programables que ejecutan uno o más programas de ordenador para realizar funciones de la materia descrita aquí operando en datos de entrada y generación de salida. Los procesos y los flujos lógicos también pueden realizarse y el aparato de la materia descrita en este documento puede implementarse como un circuito lógico de propósito especial, por ejemplo, un FPGA (matriz de compuerta programable de campo) o un ASIC (circuito integrado específico de la aplicación).

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

Para proporcionar interacción con un usuario, la materia objeto descrita en este documento puede implementarse en un ordenador que tiene un dispositivo de visualización, por ejemplo, un monitor CRT (tubo de rayos catódicos) o LCD

5 (pantalla de cristal líquido), para mostrar información al usuario y un teclado y un dispositivo señalador (por ejemplo, un mouse o una bola de seguimiento), mediante los cuales el usuario puede proporcionar entrada al ordenador. También se pueden utilizar otros tipos de dispositivos para proporcionar interacción con un usuario. Por ejemplo, la retroalimentación proporcionada al usuario puede ser cualquier forma de retroalimentación sensorial (por ejemplo, retroalimentación visual, retroalimentación auditiva o retroalimentación táctil), y la entrada del usuario puede recibirse en cualquier forma, que incluye entrada acústica, del habla o táctil.

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

20 Debe entenderse que el tema revelado no está limitado en su aplicación a los detalles de construcción y a las disposiciones de los componentes establecidos en la siguiente descripción o ilustrados en los dibujos. La materia objeto divulgada es capaz de otras realizaciones y de ser practicado y llevado a cabo de varias maneras. Además, debe entenderse que la fraseología y la terminología empleadas en este documento tienen fines descriptivos y no deben considerarse limitantes.

25 Como tal, los expertos en la materia apreciarán que la concepción, en la que se basa esta divulgación, puede utilizarse fácilmente como base para el diseño de otras estructuras, métodos y sistemas para llevar a cabo los diversos propósitos de la materia objeto divulgada. Es importante, por lo tanto, que se considere que las reivindicaciones incluyen construcciones equivalentes en la medida en que no se aparten del alcance de la materia divulgada.

30 Aunque la materia divulgada se ha descrito e ilustrado en las realizaciones ejemplares anteriores, se entiende que la presente divulgación se ha hecho solo a modo de ejemplo, y que pueden producirse numerosos cambios en los detalles de implementación de la materia objeto divulgada sin apartarse del alcance de la misma, que está limitada solo por las siguientes reivindicaciones.

REIVINDICACIONES

1. Un método informatizado para optimizar la capacidad de los equipos de red en redes móviles, el método informatizado comprende:

5 recibir, mediante un dispositivo (402) informático, una predicción (503) de uso basada en una identificación (501) de usuario y un atributo de usuario, la identificación (501) de usuario correspondiente a una característica de un usuario de la red móvil, el atributo del usuario correspondiente a al menos una característica del uso de la red móvil por parte del usuario de la red móvil, la predicción del uso incluye información correspondiente al uso futuro anticipado de datos del usuario de la red móvil, el uso futuro anticipado de datos de la red móvil correspondiente a al menos un recurso móvil;

15 recibir, por el dispositivo (402) informático, un nivel de capacidad de al menos uno de una puerta de enlace de red de paquetes de datos heredada, PGW, (703) y una virtualización de funciones de red, NFV, PGW (704), la PGW heredada (703) incluyendo una capacidad fija para al menos un recurso móvil y el NFV PGW (704) que incluye una capacidad configurable para al menos un recurso móvil;

20 calcular, mediante el dispositivo (402) informático, una decisión de enrutamiento basada en una comparación de la predicción de uso y el nivel de capacidad, la decisión de enrutamiento asociada con dirigir una puerta de enlace de servicio, SGW, (702) para enrutar al usuario de la red móvil a por lo menos uno de los PGW (703) heredados y el PGW NFV (704); y

transmitir, mediante el dispositivo informático, la decisión de enrutamiento al SGW (702).

25 2. El método informatizado de la reivindicación 1, en el que por lo menos una característica del uso de la red móvil del usuario de la red móvil incluye la cantidad de uso (504) anterior de la red móvil, un tiempo correlacionado con el uso (505) de la red móvil, ubicación de un dispositivo móvil que corresponde al usuario (506) móvil, la cantidad de tiempo dedicado a la itinerancia por el dispositivo (507) móvil, la marca y el modelo del dispositivo (508) móvil, la aplicación instalada en el dispositivo (509) móvil, el sistema operativo y el firmware versión del dispositivo (510) móvil, plan (511) de suscripción, cuota (512) restante e información (513) demográfica.

30 3. El método informatizado de la reivindicación 1, en el que la al menos una característica del usuario de la red móvil incluye una ID de dispositivo móvil o un número de teléfono.

35 4. El método informatizado de la reivindicación 1, en el que recibir el atributo de usuario comprende además recibir el atributo de usuario de al menos uno de un Servidor de abonado doméstico, HSS, (601), Entidad de Gestión de Movilidad, MME, (602), un sistema (603) de facturación, y una puerta (605) de enlace de Evolución de Arquitectura de Sistema, SAE.

40 5. El método informatizado de la reivindicación 1, en el que el recurso móvil comprende al menos una de las actividades de señalización, rendimiento, ocupación de sesión, cifrado y transcodificación.

6. Un sistema para optimizar la capacidad del equipo de red en redes móviles, el sistema comprende:

45 un procesador; y

una memoria acoplada al procesador que incluye instrucciones legibles por ordenador que, cuando son ejecutadas por un procesador, hacen que el procesador:

50 reciba una predicción de uso basada en una identificación (501) de usuario y un atributo de usuario, la identificación (501) de usuario correspondiente a una característica de un usuario de la red móvil, el atributo del usuario correspondiente a al menos una característica del uso de la red móvil por el usuario de la red móvil, la predicción del uso que incluye información correspondiente al uso futuro anticipado de datos del usuario de la red móvil, el uso de datos de red móvil futura anticipada correspondiente a al menos un recurso móvil;

55 reciba un nivel de capacidad de al menos uno de una puerta de enlace de red de paquetes de datos heredada, PGW, (703) y una virtualización de funciones de red, NFV, PGW (704), la PGW (703) heredada que incluye una capacidad fija para al menos un recurso móvil y el NFV PGW (704) que incluye una capacidad configurable para al menos un recurso móvil;

60 calcule una decisión de enrutamiento basada en una comparación de la predicción de uso y el nivel de capacidad, la decisión de enrutamiento asociada con dirigir una puerta de enlace de servicio, SGW, (702) para enrutar al usuario de la red móvil al menos a una de las PGW heredadas (703) y el NFV PGW (704); y

65 transmita la decisión de enrutamiento al SGW (702).

- 5 7. El sistema de la reivindicación 6, en el que la al menos una característica del uso de la red móvil del usuario de la red móvil incluye la cantidad de uso previo de la red (504) móvil, un tiempo correlacionado con el uso (505) de la red móvil, la ubicación de un dispositivo móvil correspondiente al usuario (506) móvil, cantidad de tiempo de itinerancia del dispositivo (507) móvil, marca y modelo del dispositivo (508) móvil, aplicación instalada en el dispositivo (509) móvil, sistema operativo y versión de firmware del dispositivo (510) móvil, plan (511) de suscripción, cuota (512) restante e información (513) demográfica.
- 10 8. El sistema de la reivindicación 6, en el que la al menos una característica del usuario de la red móvil incluye una ID de dispositivo móvil o un número de teléfono.
- 15 9. El sistema de la reivindicación 6, en el que el procesador recibe además el atributo de usuario de al menos uno de un Servidor de abonado doméstico, HSS (601), Entidad de gestión de movilidad, MME, (602), un sistema (603) de facturación, y una puerta (604) de enlace de Evolución de Arquitectura de Sistema, SAE, puerta de enlace.
10. El sistema de la reivindicación 6, en el que el recurso móvil comprende al menos una de las actividades de señalización, rendimiento, ocupación de sesión, cifrado y transcodificación.

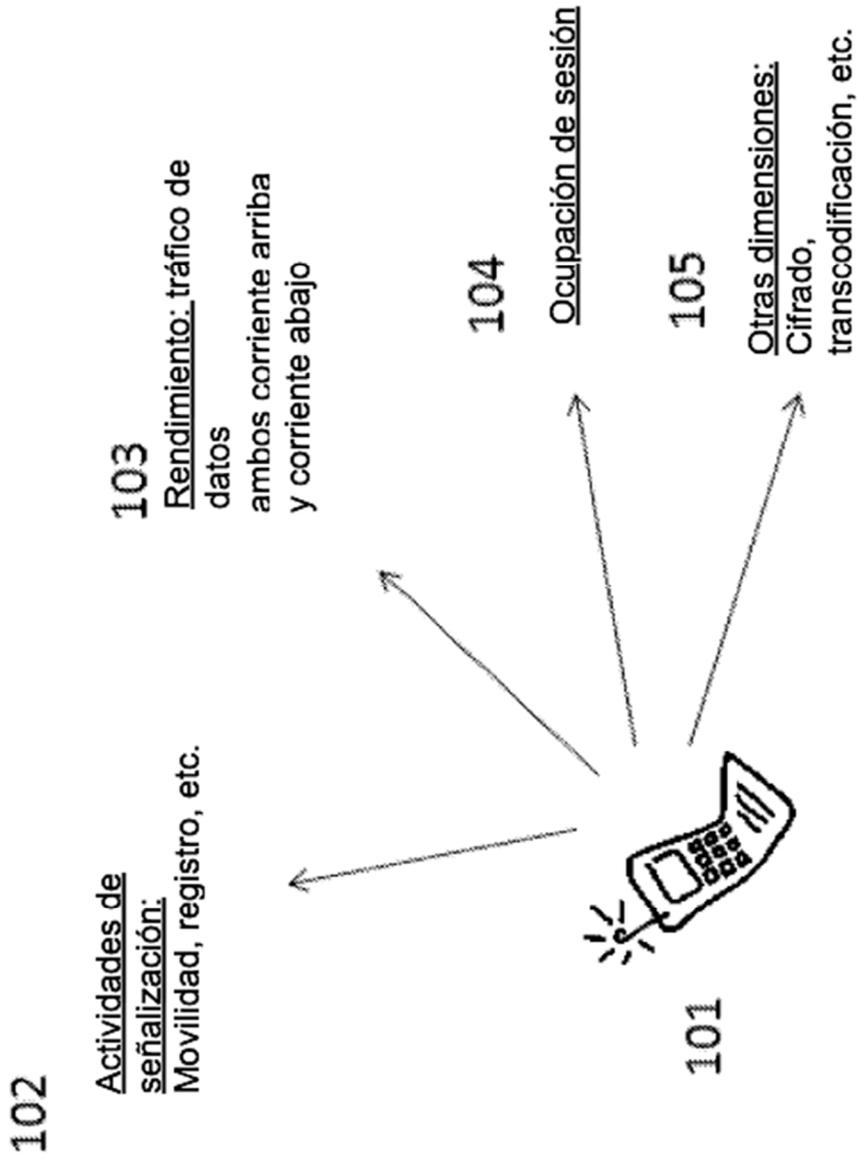


FIG. 1

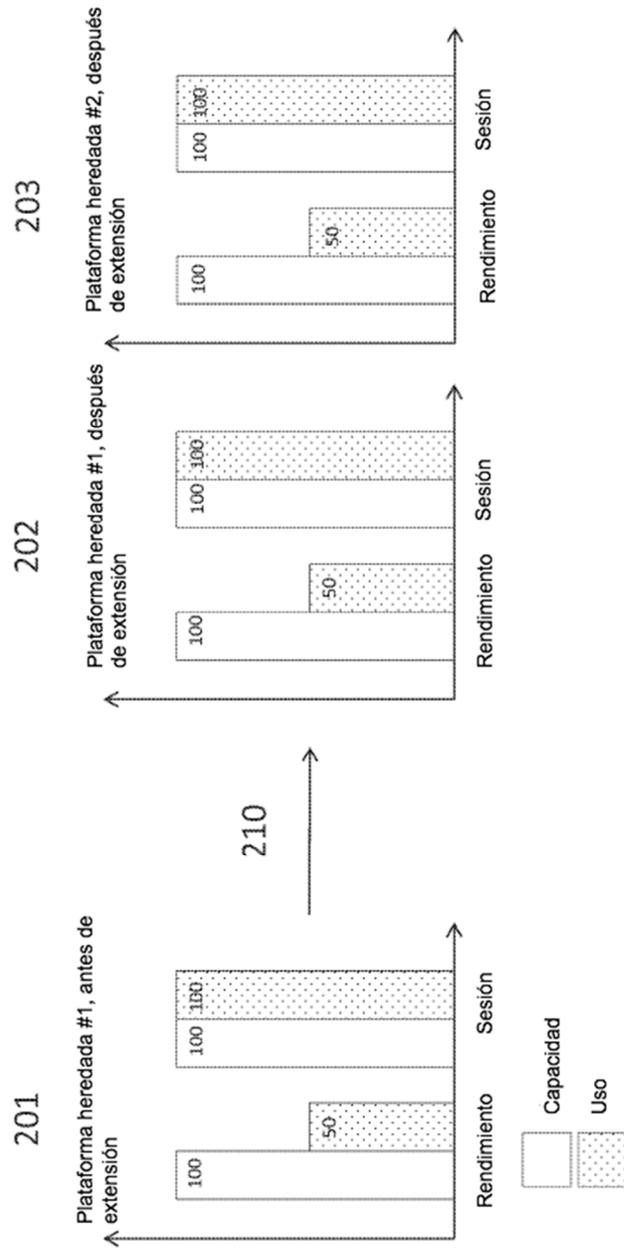


FIG. 2 (Técnica anterior)

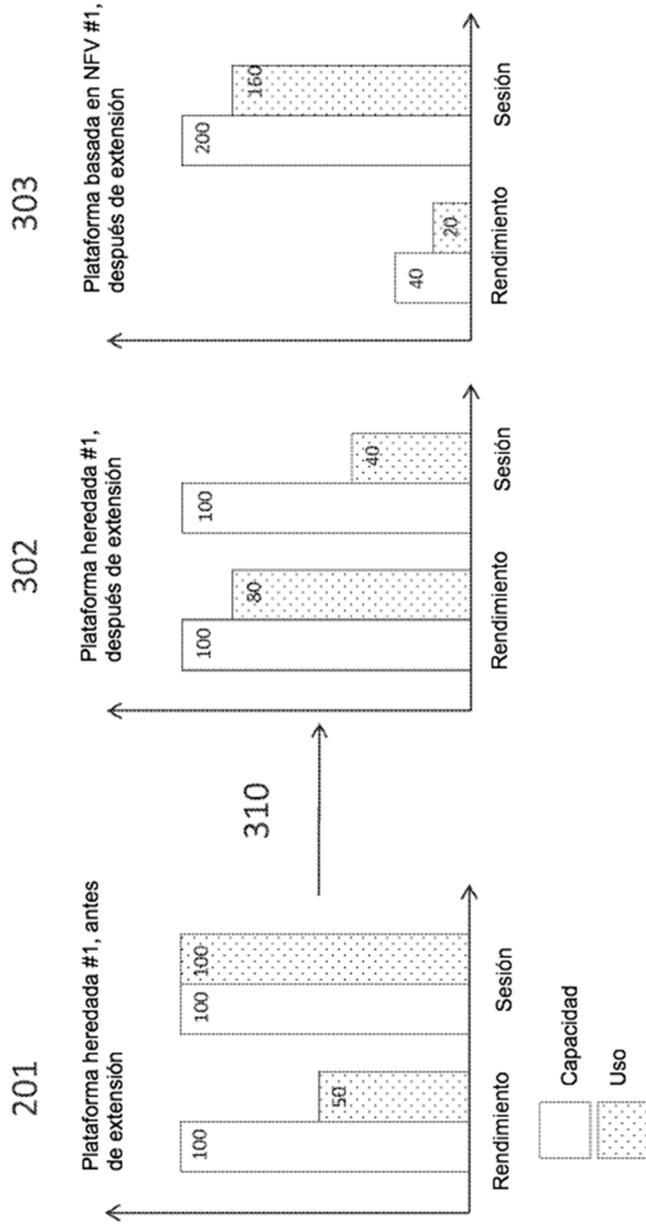


FIG. 3

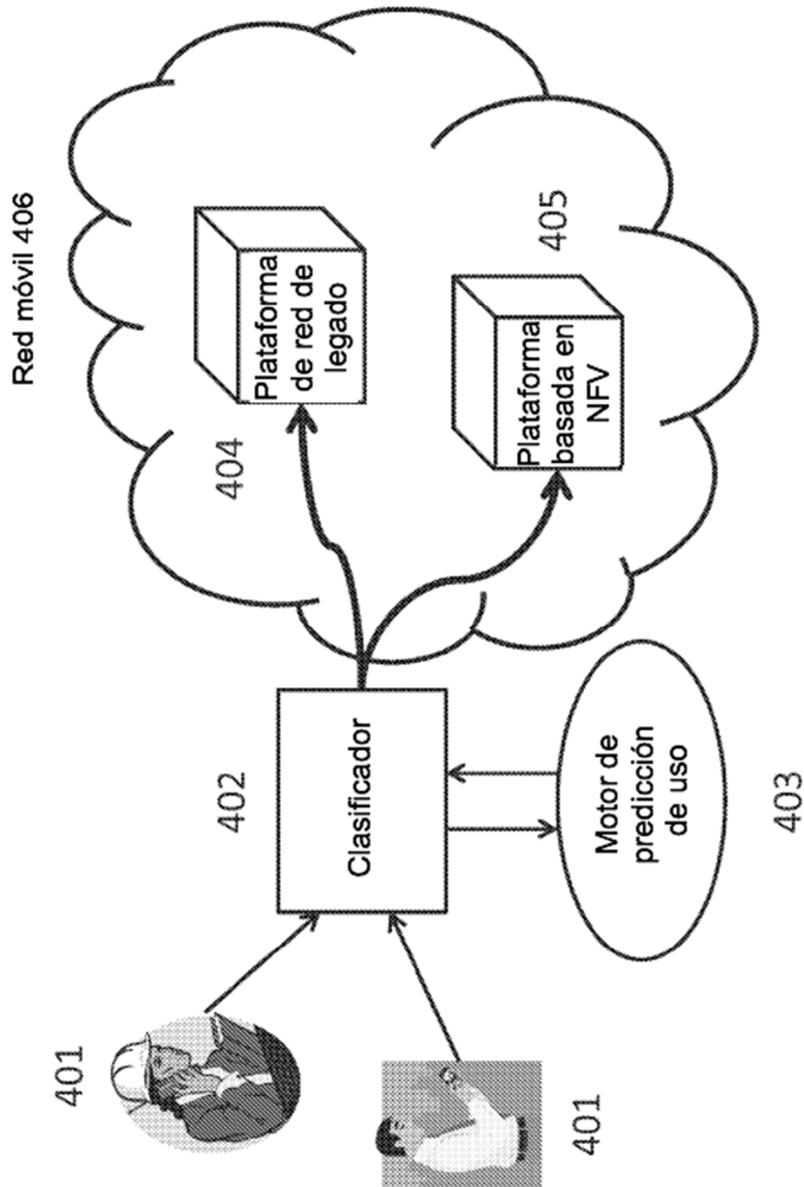


FIG. 4

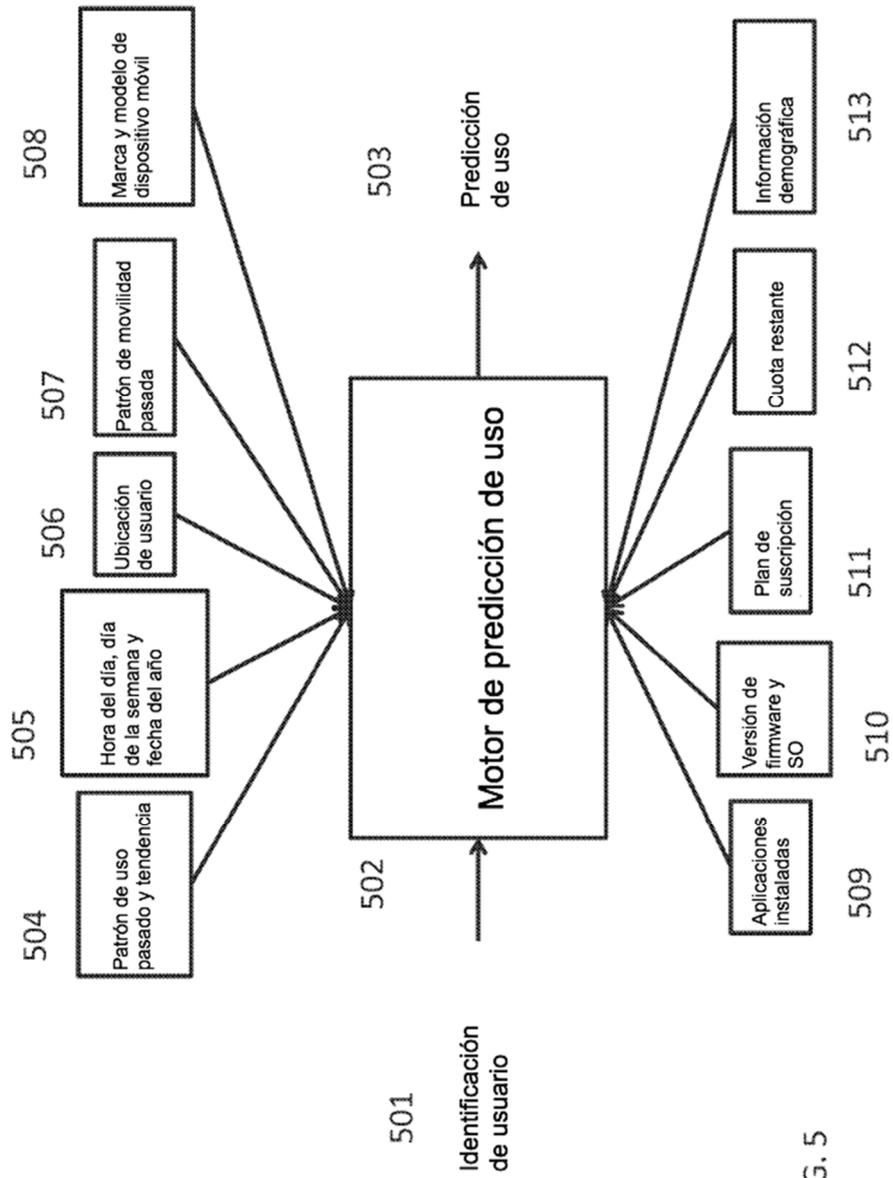


FIG. 5

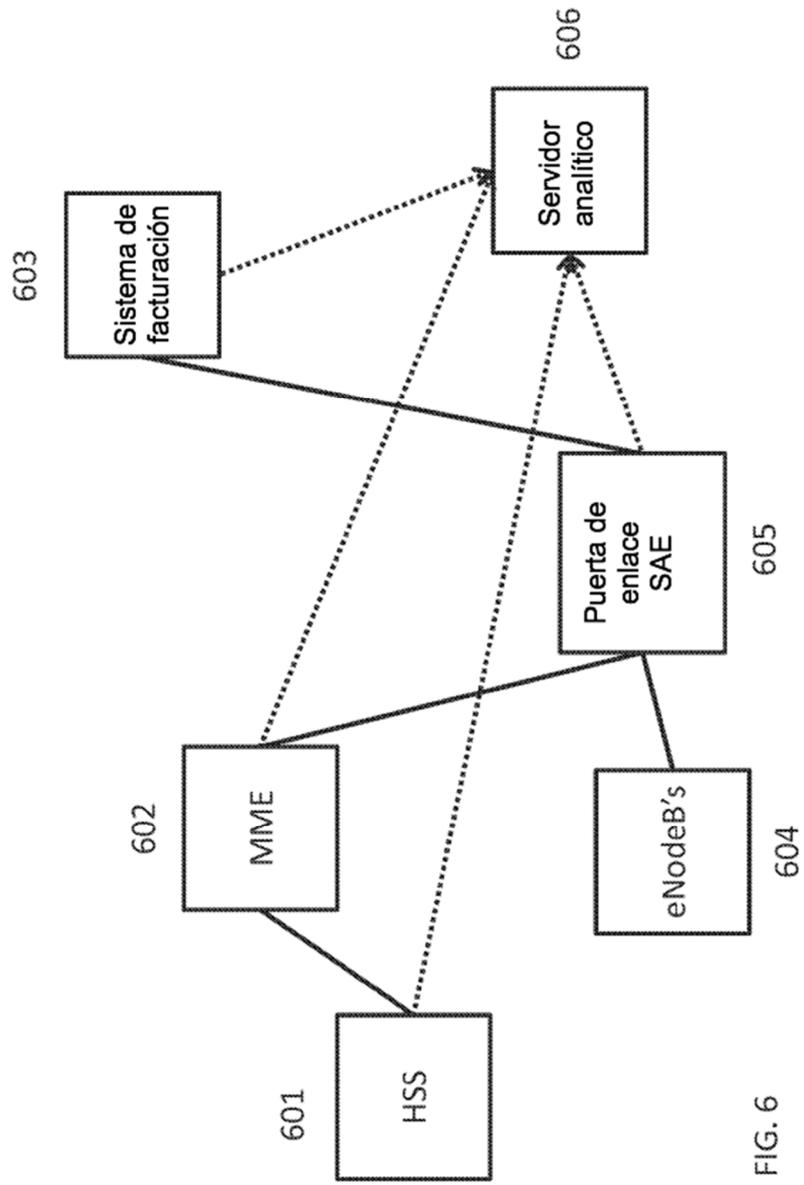


FIG. 6

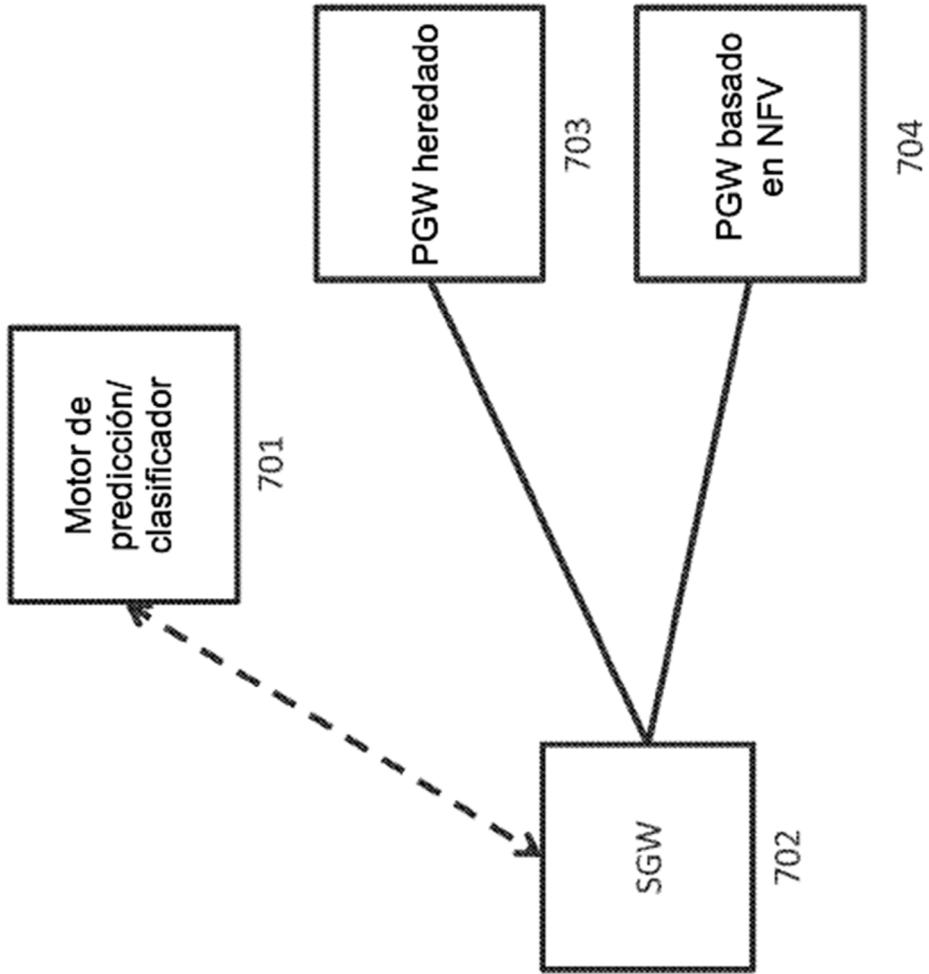


FIG. 7