

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 906 171**

51 Int. Cl.:

**G06F 9/50** (2006.01)

**H04L 12/24** (2006.01)

**H04L 29/08** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **19.09.2016 PCT/IB2016/055592**

87 Fecha y número de publicación internacional: **11.05.2017 WO17077406**

96 Fecha de presentación y número de la solicitud europea: **19.09.2016 E 16779189 (6)**

97 Fecha y número de publicación de la concesión europea: **03.11.2021 EP 3371696**

54 Título: **Sistema y métodos para la colocación de funciones de servicio inteligente y escala automática basado en aprendizaje automático**

30 Prioridad:

**02.11.2015 US 201514930546**

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

**13.04.2022**

73 Titular/es:

**TELEFONAKTIEBOLAGET LM ERICSSON (PUBL)  
(100.0%)  
164 83 Stockholm, SE**

72 Inventor/es:

**HALPERN, JOEL;  
SHIRAZIPOUR, MERAL;  
XIA, MING;  
MAHKONEN, HEIKKI y  
MANGHIRMALANI, RAVI**

74 Agente/Representante:

**ELZABURU, S.L.P**

ES 2 906 171 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

Sistema y métodos para la colocación de funciones de servicio inteligente y escala automática basado en aprendizaje automático

### Referencia cruzada a solicitudes relacionadas

- 5 Esta solicitud reivindica la prioridad de la Solicitud de Estados Unidos N.º 14/930.546, presentada el 2 de noviembre de 2015.

### Campo

Las realizaciones de la invención se refieren al campo de las redes de paquetes y, más específicamente, a la optimización del uso de recursos en una red utilizando técnicas de aprendizaje automático.

### 10 Antecedentes

15 El encadenamiento de servicios diferencia el reenvío de flujos de tráfico a través de un conjunto ordenado definido por políticas de cajas intermedias (también denominados comúnmente servicios, servicios en línea, aparatos, funciones de red (o funciones de red virtual en el caso de la virtualización de funciones de red (NFV)), o funciones de servicio (SF)). Los ejemplos de SF incluyen cortafuegos, filtros de contenido, sistemas de detección de intrusos (IDS), inspección profunda de paquetes (DPI), traducción de direcciones de red (NAT), cachés de contenido, equilibradores de carga, aceleradores de red de área amplia (WAN), transcodificadores multimedia y aplicaciones de registro/medición/carga/carga avanzada.

20 El encadenamiento de servicios requiere un proceso de clasificación para ceder el tráfico a la cadena de servicios correcta, seguido de un reenvío/encaminamiento diferenciado del flujo de tráfico a través del conjunto correcto de SF (p. ej., una cadena de función de servicio (SFC)). El Grupo de Trabajo de Ingeniería de Internet (IETF) está desarrollando protocolos que permitirán formas más eficientes de implementar las SFC. El IETF también está trabajando en la definición de un encabezado de servicio de red (NSH) que se agrega a los paquetes para ayudar con el encadenamiento de servicios. A continuación, los reenviadores de funciones de servicio (SFF) crearán las rutas de funciones de servicio (SFP) en forma de superposición. Dicho encadenamiento de servicios es aplicable tanto a las  
25 funciones de red física como a las funciones de red virtual (vNF).

30 La medición de la red proporciona al operador de la red datos que le permiten caracterizar el estado de la red, las demandas de tráfico, el consumo real de los recursos de la red y el rendimiento de la red experimentado por los usuarios finales. El operador de red puede utilizar esta información para optimizar el uso de los recursos de red en la red. Los paradigmas de red, tales como la interconexión en red definida por software (SDN) y la virtualización de funciones de red (NFV), pueden utilizarse para lograr una visión holística de los recursos de la red y la nube.

35 La cantidad de recursos de red y de nube (p. ej., recursos de la unidad central de procesamiento (CPU), memoria, ancho de banda, almacenamiento) que utiliza una SFC puede variar según las SF específicas (p. ej., las vNF) incluidos en la SFC, la cantidad de tráfico en la red, y otros factores. También, la cantidad de recursos requeridos por una SFC puede cambiar con el tiempo a medida que cambian las circunstancias. Las soluciones existentes para el aprovisionamiento de recursos SFC se basan en reglas y políticas estáticas. Las reglas y políticas estáticas generalmente sobreaprovisionan recursos (lo que lleva a un uso ineficiente de los recursos) o sub-aprovisionan recursos (lo que lleva a una degradación del rendimiento).

La patente europea EP 2 757 474 A2 (MITEL NETWORKS CORP [CA]) 23 de julio de 2014 (23-07-2014) es la técnica anterior relevante.

### 40 Compendio

La invención está definida por las reivindicaciones independientes 1, 5, 9, 10, 11 y 15.

### Breve descripción de los dibujos

La invención puede entenderse mejor haciendo referencia a la siguiente descripción y los dibujos adjuntos que se utilizan para ilustrar las realizaciones de la invención. En los dibujos:

- 45 La figura 1 es un diagrama de bloques que ilustra una red en la que puede implementarse la optimización de recursos de las cadenas de funciones de servicio (SFC), según algunas realizaciones.

La figura 2 es un diagrama de flujo que ilustra un proceso para optimizar el uso de recursos de una nueva SFC que se va a aprovisionar en la red, según algunas realizaciones.

- 50 La figura 3 es un diagrama que ilustra las entradas y salidas de un sistema de aprendizaje automático de uso de recursos, según algunas realizaciones.

La figura 4 es un diagrama que ilustra las etapas internas de un sistema de aprendizaje automático de uso de recursos, según algunas realizaciones.

La figura 5 es un diagrama de flujo que ilustra un proceso para optimizar el uso de recursos de las SFC existentes en la red, según algunas realizaciones.

- 5 La figura 6 es un diagrama que ilustra los componentes de una función de red virtual (vNF), según algunas realizaciones.

La figura 7 es un diagrama que ilustra las entradas y salidas de un sistema de aprendizaje automático de escala automática, según algunas realizaciones.

- 10 La figura 8 es un diagrama que ilustra el interior de un sistema de aprendizaje automático de escala automática, según algunas realizaciones.

La figura 9A ilustra la conectividad entre dispositivos de red (ND) dentro de una red ilustrativa, así como tres implementaciones ilustrativas de los ND, según algunas realizaciones.

La figura 9B ilustra una forma ilustrativa de implementar un dispositivo de red de fin especial según algunas realizaciones.

- 15 La figura 9C ilustra diversas formas ilustrativas en las que los elementos de red virtual (VNE) pueden acoplarse según algunas realizaciones.

La figura 9D ilustra una red con un solo elemento de red (NE) en cada uno de los ND, y dentro de este enfoque directo contrasta un enfoque distribuido tradicional (utilizado comúnmente por los encaminadores tradicionales) con un enfoque centralizado para mantener la accesibilidad e información de reenvío (también denominado control de red), según algunas realizaciones.

- 20

La figura 9E ilustra el caso sencillo en el que cada uno de los ND implementa un solo NE, pero un plano de control centralizado ha abstraído múltiples NE en diferentes ND en (para representar) un solo NE en una de las red o redes virtuales, según algunas realizaciones.

La figura 9F ilustra un caso en el que se implementan múltiples VNE en diferentes ND y están acoplados entre sí, y en el que un plano de control centralizado ha abstraído estos múltiples VNE de modo que aparecen como un solo VNE dentro de una de las redes virtuales, según algunos realizaciones.

- 25

La figura 10 ilustra un dispositivo de plano de control de fin general con software de plano de control centralizado (CCP), según algunas realizaciones.

### Descripción de las realizaciones

- 30 La siguiente descripción describe métodos y aparato para optimizar el uso de recursos de las cadenas de funciones de servicio (SFC) en una red utilizando técnicas de aprendizaje automático. En la siguiente descripción, se exponen numerosos detalles específicos, tales como implementaciones lógicas, códigos de operación, medios para especificar operandos, implementaciones de partición/compartición/duplicación de recursos, tipos e interrelaciones de los componentes del sistema y opciones de partición/integración lógica para proporcionar un entendimiento minucioso de la presente invención. Sin embargo, un experto en la técnica apreciará que la invención puede ponerse en práctica sin tales detalles específicos. En otros casos, las estructuras de control, los circuitos de nivel de puertas y las secuencias de instrucciones de software completas no se han mostrado en detalle para no oscurecer la invención. Los expertos en la técnica, con las descripciones incluidas, podrán implementar la funcionalidad apropiada sin experimentación indebida.

- 40 Las referencias en la memoria descriptiva a "una realización", "una realización", "una realización ilustrativa", etc., indican que la realización descrita puede incluir un rasgo, estructura o característica particular, pero es posible que cada realización no incluya necesariamente el rasgo, estructura o característica particular. Además, tales expresiones no se refieren necesariamente a la misma realización. Además, cuando se describe un rasgo, estructura o característica particular en relación con una realización, se afirma que está dentro del conocimiento de un experto en la técnica afectar tal rasgo, estructura o característica en relación con otras realizaciones ya se haya descrito explícitamente o no.

- 45

El texto entre paréntesis y los bloques de línea discontinua (p. ej., de línea discontinua grande, de línea discontinua pequeña, de punto y guion y de puntos) pueden utilizarse en la presente memoria para ilustrar operaciones opcionales que agregan características adicionales a las realizaciones de la invención. Sin embargo, dicha notación no debe interpretarse como que estas son las únicas opciones u operaciones opcionales, y/o que los bloques con bordes de línea continua no son opcionales en ciertas realizaciones de la invención.

- 50

En la siguiente descripción y reivindicaciones, pueden utilizarse los términos "acoplado" y "conectado", junto con sus derivados. Debe entenderse que estos términos no pretenden ser sinónimos entre sí. "Acoplado" se utiliza para indicar

que dos o más elementos, que pueden o no estar en contacto físico o eléctrico directo entre sí, cooperan o interactúan entre sí. "Conectado" se utiliza para indicar el establecimiento de comunicación entre dos o más elementos que se acoplan entre sí.

5 Un dispositivo electrónico (p. ej., un dispositivo informático) almacena y transmite (internamente y/o con otros dispositivos electrónicos a través de una red) código (que se compone de instrucciones de software y que a veces se denomina código de programa informático o programa informático) y/o datos que utilizan medios legibles por máquina (también llamados medios legibles por ordenador), tales como medios de almacenamiento legibles por máquina (p. ej., discos magnéticos, discos ópticos, memoria de solo lectura (ROM), dispositivos de memoria flash, memoria de cambio de fase) y medios de transmisión legibles por máquina (también llamados portadores) (p. ej., señales eléctricas, 10 ópticas, de radio, acústicas u otras formas de propagación, tales como ondas portadoras, señales infrarrojas). Por lo tanto, un dispositivo electrónico (p. ej., un ordenador) incluye hardware y software, tales como un conjunto de uno o más procesadores acoplados a uno o más medios de almacenamiento legibles por máquina para almacenar código para su ejecución en el conjunto de procesadores y/o para almacenar datos. Por ejemplo, un dispositivo electrónico puede incluir una memoria no volátil que contiene el código, ya que la memoria no volátil puede hacer persistir el código/datos incluso cuando el dispositivo electrónico está apagado (cuando se desconecta la alimentación), y, mientras el dispositivo electrónico está encendido, parte del código que se va a ejecutar por el procesador o procesadores de ese dispositivo electrónico normalmente se copia de la memoria no volátil más lenta a la memoria volátil (p. ej., memoria dinámica de acceso aleatorio (DRAM), memoria estática de acceso aleatorio (SRAM)) de ese dispositivo electrónico. Los dispositivos electrónicos típicos también incluyen un conjunto o una o más interfaces de red físicas para establecer conexiones de red (para transmitir y/o recibir código y/o datos utilizando señales de propagación) con otros dispositivos electrónicos. Una o más partes de una realización de la invención pueden implementarse utilizando diferentes combinaciones de software, firmware y/o hardware.

Un dispositivo de red (ND) es un dispositivo electrónico que interconecta de manera comunicativa otros dispositivos electrónicos en la red (p. ej., otros dispositivos de red, dispositivos de usuario final). Algunos dispositivos de red son "dispositivos de red de servicios múltiples" que brindan soporte para múltiples funciones de interconexión en red (p. ej., encaminamiento, conexión en puente, conmutación, agregación de capa 2, control de borde de sesión, calidad de servicio y/o gestión de suscriptores) y/o brindan soporte para múltiples servicios de aplicaciones (p. ej., datos, voz y video).

Las soluciones existentes para el aprovisionamiento de recursos de la cadena de funciones de servicio (SFC) se basan en reglas y políticas estáticas. Las reglas y políticas estáticas generalmente sobreaprovisionan recursos (lo que lleva a un uso ineficiente de los recursos) o sub-aprovisionan recursos (lo que lleva a una degradación del rendimiento). Las realizaciones introducen sistemas de aprendizaje automático que pueden aprender las tendencias de consumo de recursos de las SFC. Los sistemas de aprendizaje automático pueden proporcionar una determinación más precisa de la cantidad de recursos que necesitará una SFC y pueden predecir cuándo se debe ajustar la cantidad de recursos aprovisionados para una SFC. Las realizaciones utilizan información obtenida de sistemas de aprendizaje automático para optimizar el uso de recursos de las SFC.

La figura 1 es un diagrama de bloques que ilustra una red en la que puede implementarse la optimización de recursos de las SFC, según algunas realizaciones. En la realización ilustrada, la red 100 incluye un dispositivo informático 110, un sistema de monitorización y gestión de recursos 120 y un plano de reenvío 140. El plano de envío 140 incluye dispositivos de red (ND) 150A-D.

El sistema de monitorización y gestión de recursos 120 proporciona una plataforma unificada para configurar, gestionar y monitorizar recursos en la red. Como se ilustra, el sistema de monitorización y gestión de recursos 120 incluye un controlador de SDN 125 y un gestor de la nube 130. El controlador de SDN 125 puede controlar el encaminamiento del tráfico en el plano de reenvío 140 configurando el comportamiento de reenvío de los dispositivos de red 150A-D del plano de reenvío. En una realización, el controlador de SDN 125 configura los dispositivos de red 150A-D utilizando un protocolo de comunicaciones del plano de control tal como OpenFlow (p. ej., OpenFlow 1.3). El gestor de la nube 130 puede instanciar y gestionar recursos virtuales en la red 100. Por ejemplo, el gestor de la nube 130 puede instanciar y gestionar máquinas virtuales (VM), funciones de red virtual (vNF) (p. ej., la vNF 155B y la vNF 155C), conmutadores virtuales y otros tipos de recursos virtuales en la red 100. En una realización, el sistema de monitorización y gestión de recursos 120 puede incluir otras entidades/unidades funcionales que pueden facilitar la configuración, gestión y monitorización de recursos en la red, tales como un orquestador de virtualización de funciones de red (NFVO) u otra entidad de gestión y organización (MANO).

El sistema de monitorización y gestión de recursos 120 puede aprovechar el controlador de SDN 125 y el gestor de la nube 130 para proporcionar el encadenamiento de servicios en la red 100. El encadenamiento de servicios hace referencia al reenvío diferenciado de flujos de tráfico a través de un conjunto de dispositivos intermedios definidos por políticas (también denominados servicios, servicios en línea, aparatos, funciones de red, vNF (p. ej., la vNF 155B y la vNF 155C) en el caso de virtualización de funciones de red (NFV), o funciones de servicio (SF) como término general). Los ejemplos de SF incluyen cortafuegos, filtros de contenido, sistemas de detección de intrusos (IDS), inspección profunda de paquetes (DPI), traducción de direcciones de red (NAT), cachés de contenido, equilibradores de carga, aceleradores de red de área amplia (WAN), transcodificadores multimedia y aplicaciones de registro/medición/carga/carga avanzada. El encadenamiento de servicios normalmente implica un proceso de

clasificación para asignar paquetes a la cadena de servicios adecuada, seguido del reenvío/encaminamiento diferenciado del flujo de tráfico a través del conjunto adecuado de SF (es decir, una cadena de función de servicio (SFC)).

5 Una SFC define una secuencia ordenada de SF que debería atravesar un flujo de tráfico. El sistema de monitorización y gestión de recursos 120 puede aprovisionar una SFC en la red desplegando y/o configurando las SF especificadas por la SFC y dirigiendo el flujo de tráfico a través de las SF especificadas por la SFC. Por ejemplo, una SFC puede especificar que el flujo de tráfico debería atravesar una función de servicio de inspección profunda de paquetes seguida de una función de servicio de sistema de detección de intrusos. El sistema de monitorización y gestión de recursos 120 puede aprovisionar esta SFC iniciando el despliegue de la vNF 155B en el dispositivo de red 150B para realizar el servicio de inspección profunda de paquetes e iniciar el despliegue de la vNF 155C en el dispositivo de red 150C para realizar el servicio del sistema de detección de intrusos (p. ej., a través del gestor de la nube 130). Si ya se han implementado todas las vNF 155 necesarias para la SFC (p. ej., la vNF 155B y 155C), entonces no es necesario que el sistema de monitorización y gestión de recursos 120 inicie el despliegue de las vNF 155. El sistema de monitorización y gestión de recursos 120 puede dirigir, a continuación, el flujo de tráfico (p. ej., a través del controlador de SDN 125) a través de la vNF 155B y la vNF 155C de modo que el flujo de tráfico se procese en primer lugar por la función de servicio de inspección profunda de paquetes y, a continuación, se procese por la función de servicio de detección de intrusos. De esta forma, el sistema de monitorización y gestión de recursos 120 puede aprovisionar una SFC en la red 100.

20 El sistema de monitorización y gestión de recursos 120 es capaz de monitorizar las SFC aprovisionadas en la red 100. Por ejemplo, el sistema de monitorización y gestión de recursos 120 puede recopilar información con respecto a las funciones de servicio (p. ej., la vNF 155) y las SFC aprovisionadas en la red. El sistema de monitorización y gestión de recursos 120 también puede recopilar información con respecto a la utilización de recursos de las vNF 155 (incluyendo cuántas instancias se despliegan y su utilización de unidad de procesamiento central (CPU)/memoria/ancho de banda). El sistema de monitorización y gestión de recursos 120 también puede recopilar información sobre la disponibilidad de recursos informáticos y recursos de red para el encadenamiento de servicios. Además, el sistema de monitorización y gestión de recursos 120 puede recopilar información con respecto al tráfico de suscriptores, que incluye cuánto tráfico de suscriptores se dirige a cada SFC. El sistema de monitorización y gestión de recursos 120 también puede recopilar otras métricas de red y de la nube en tiempo real relacionadas con las SFC aprovisionadas en la red, tales como métricas de rendimiento de tráfico y métricas de calidad de experiencia (QoE).

30 Como se ilustra, el dispositivo informático 110 está acoplado al sistema de monitorización y gestión de recursos 120 y tiene un componente de optimización de recursos de SFC 115 instalado en el mismo. Como se describirá con más detalle en la presente memoria a continuación, el dispositivo informático 110 puede ejecutar el componente de optimización de recursos de SFC 115 para optimizar el uso de recursos de las SFC. Aunque el componente de optimización de recursos de SFC 115 se muestra implementado por el dispositivo informático 110, en otras realizaciones, algunas o toda la funcionalidad del componente de optimización de recursos de SFC 115 puede implementarse mediante una entidad diferente (p. ej., el sistema de monitorización y gestión de recursos 120). El componente de optimización de recursos de SFC 115 se basa en un sistema de aprendizaje automático (ML) de uso de recursos 160 y un conjunto de sistemas de aprendizaje automático (ML) de escala automática 170 para optimizar el uso de recursos de las SFC. El sistema de aprendizaje automático de uso de recursos 160 utiliza técnicas de aprendizaje automático para determinar la cantidad de recursos para aprovisionar a una SFC antes de que se aprovisionen las SFC en la red 100. El conjunto de sistemas de aprendizaje automático de escala automática 170 utiliza técnicas de aprendizaje automático para determinar cuándo se debe ajustar la cantidad de recursos aprovisionados para una SFC (y cuánto debería ajustarse) una vez que se aprovisiona la SFC en la red 100.

45 Cuando el componente de optimización de recursos de SFC 115 recibe una solicitud para aprovisionar una nueva SFC (o actualizar una SFC existente) en la red 100, el componente de optimización de recursos de SFC 115 puede basarse en el sistema de aprendizaje automático de uso de recursos 160 para determinar la cantidad de recursos a aprovisionar para la SFC. El sistema de aprendizaje automático de uso de recursos 160 utiliza técnicas de aprendizaje automático para determinar la cantidad sugerida de recursos para aprovisionar a la SFC. El componente de optimización de recursos de SFC 115 puede obtener esta información del sistema de aprendizaje automático de uso de recursos 160 y proporcionarla a un componente asignador de recursos (no mostrado) que determinará un esquema de asignación de recursos óptimo que asigna la cantidad de recursos para la SFC según lo sugerido por el sistema de aprendizaje automático de uso de recursos 160 (lo mejor posible). El componente asignador de recursos puede implementarse como parte del componente de optimización de recursos de SFC 115 o separado del componente de optimización de recursos de SFC 115. El esquema de asignación de recursos puede proporcionarse al sistema de monitorización y gestión de recursos 120 y el sistema de monitorización y gestión de recursos 120 puede aprovisionar la SFC en la red según el esquema de asignación de recursos. Como se ilustra, el sistema de aprendizaje automático de uso de recursos 160 se implementa como parte del componente de optimización de recursos de SFC 115. Sin embargo, en otras realizaciones, el sistema de aprendizaje automático de uso de recursos 160 puede implementarse por separado del componente de optimización de recursos de SFC 115 y, en algunos casos, externo al dispositivo informático 110.

60 Una vez que se ha aprovisionado una SFC en la red 100, el componente de optimización de recursos de SFC 115 puede basarse en un conjunto de sistemas de aprendizaje automático de escala automática 170 para predecir automáticamente cuándo es necesario ajustar la cantidad de recursos aprovisionados para la SFC (p. ej., aumentar o

reducir la escala) antes de la necesidad. La escala automática hacia arriba o hacia abajo de los recursos se denomina generalmente en la presente memoria como escala automática. En una realización, cada vNF 155 implementada en la red está asociada con un sistema de aprendizaje automático de escala automática 170. Por ejemplo, como se ilustra, la vNF 155B está asociada con el sistema de aprendizaje automático de escala automática 170B y la vNF 155C está asociada con el sistema de aprendizaje automático de escala automática 170C. El sistema de aprendizaje automático de escala automática 170 para una vNF 155 puede predecir cuándo debe ajustarse la cantidad de recursos provisionados para la vNF 155. Cuando un sistema de aprendizaje automático de escala automática 170 asociado con una vNF 155 predice que la cantidad de recursos provisionados para la vNF 155 debe ajustarse, el sistema de aprendizaje automático de escala automática 170 envía una solicitud de escala automática al componente de optimización de recursos de SFC 115. La solicitud de escala automática puede solicitar un ajuste en la cantidad de recursos provisionados para la vNF 155. Por ejemplo, la solicitud de escala automática puede indicar que la cantidad de recursos provisionados para la vNF 155 debe ajustarse hacia arriba o hacia abajo y en qué medida. El sistema de aprendizaje automático de escala automática 170 para una vNF 155 puede monitorizar constantemente la vNF 155 para determinar cuándo se necesita escala automática y enviar solicitudes de escala automática al componente de optimización de recursos de SFC 115 cuando se necesita escala automática. Como se ilustra, los sistemas de aprendizaje automático de escala automática 170 se implementan en el mismo dispositivo de red 150 que sus respectivas vNF 155. Sin embargo, en otras realizaciones, los sistemas de aprendizaje automático de escala automática 170 pueden implementarse en un servidor separado o en un centro de datos de rendimiento optimizado (POD). Desde el punto de vista de la arquitectura, todos los sistemas de aprendizaje automático de escala automática 170 pueden residir en la misma ubicación (p. ej., colocalizados con el sistema de monitorización y gestión de recursos 120 o con el dispositivo informático 110). Sin embargo, en general, el rendimiento puede mejorarse si los sistemas de aprendizaje automático de escala automática 170 se ubican más cerca de sus respectivas vNF 155. En una realización, el sistema de monitorización y gestión de recursos 120 o el dispositivo informático 110 pueden iniciar la implementación de un nuevo sistema de aprendizaje automático de escala automática 170 para una vNF 155 (p. ej., cuando se implementa la vNF 155).

El componente de optimización de recursos de SFC 115 puede obtener solicitudes de escala automática de uno o más de los sistemas de aprendizaje automático de escala automática 170 y proporcionar las solicitudes de escala automática a un componente de reasignación de recursos (no mostrado) que determinará un esquema de reasignación de recursos óptimo que cumpla con las solicitudes de escala automática (lo mejor posible). El componente de reasignación de recursos puede implementarse como parte del componente de optimización de recursos de SFC 115 o separado del componente de optimización de recursos de SFC 115. El esquema de reasignación de recursos puede implicar realizar un equilibrio de carga, migrar o implementar nuevas vNF 155, o incluso desasignar o apagar las vNF 155. El esquema de reasignación de recursos puede proporcionarse al sistema de monitorización y gestión de recursos 120 y el sistema de monitorización y gestión de recursos 120 puede reasignar los recursos según el esquema de reasignación de recursos. De esta forma, las realizaciones optimizan el uso de recursos para las SFC utilizando técnicas de aprendizaje automático.

La figura 2 es un diagrama de flujo que ilustra un proceso para optimizar el uso de recursos de una nueva SFC que se va a aprovisionar en la red, según algunas realizaciones. Las operaciones en el diagrama de flujo pueden implementarse mediante un dispositivo informático 110 (p. ej., el componente de optimización de recursos de SFC 115). Las operaciones en este y otros diagramas de flujo se describirán con referencia a las realizaciones ilustrativas de las otras figuras. Sin embargo, debe entenderse que las operaciones de los diagramas de flujo pueden realizarse por realizaciones de la invención distintas de las analizadas con referencia a las otras figuras, y las realizaciones de la invención analizadas con referencia a estas otras figuras pueden realizar operaciones diferentes a las analizadas con referencia a los diagramas de flujo.

En una realización, el proceso se inicia cuando el dispositivo informático 110 recibe una solicitud para aprovisionar una SFC (bloque 210). En una realización, la solicitud puede realizarla un operador de red o un cliente empresarial.

El dispositivo informático 110 proporciona una descripción de la SFC como una entrada a un sistema de aprendizaje automático de uso de recursos 160 (bloque 220). El sistema de aprendizaje automático de uso de recursos 160 puede predecir una cantidad de recursos que la SFC utilizará basándose en el aprendizaje (utilizando el aprendizaje automático) del uso real de recursos de las SFC anteriores provisionadas en la red 100. El sistema de aprendizaje automático de uso de recursos 160 recibe como entrada la descripción de la SFC (bloque 230). En una realización, la descripción de la SFC incluye una indicación de los puntos finales de la SFC y una lista de funciones de servicio incluidas en la SFC. En una realización, la descripción de la SFC incluye las reglas de clasificación utilizadas para asignar tráfico a la SFC (p. ej., la intención de la SFC). La descripción de la SFC puede incluir cualquier otra característica o propiedad de la SFC. El sistema de aprendizaje automático de uso de recursos 160 utiliza, a continuación, técnicas de aprendizaje automático para determinar una cantidad sugerida de recursos para aprovisionar a la SFC basándose en la entrada (es decir, la descripción de la SFC) (bloque 240). El sistema de aprendizaje automático de uso de recursos 160 a continuación emite la cantidad sugerida de recursos para aprovisionar a la SFC (bloque 250). En una realización, la cantidad sugerida de recursos para aprovisionar a la SFC incluye una cantidad sugerida de recursos de CPU (p. ej., número de núcleos) para aprovisionar a la SFC, una cantidad sugerida de memoria para aprovisionar a la SFC, una cantidad sugerida de ancho de banda para aprovisionar a la SFC, una cantidad sugerida de recursos de almacenamiento para aprovisionar a la SFC, o una cantidad sugerida para cualquier otro tipo de red/recurso informático para aprovisionar a la SFC. En una realización, la cantidad sugerida de recursos

para aprovisionar a la SFC incluye una cantidad sugerida de recursos para aprovisionar por función de servicio incluida en la SFC.

5 El dispositivo informático 110 obtiene la cantidad sugerida de recursos para aprovisionar a la SFC del sistema de aprendizaje automático de uso de recursos 160 (bloque 260). En algunos casos, la entidad que realizó la solicitud para aprovisionar a la SFC (p. ej., el operador de red o el cliente empresarial) puede especificar la demanda de recursos de la SFC en su solicitud. En una realización, el dispositivo informático 110 anula la demanda de recursos especificada en la solicitud con la cantidad sugerida de recursos obtenidos del sistema de aprendizaje automático de uso de recursos 160.

10 El dispositivo informático 110 a continuación proporciona la cantidad sugerida de recursos para aprovisionar a la SFC a un componente asignador de recursos (bloque 270). El componente asignador de recursos puede utilizar cualquier algoritmo o estrategia adecuada para determinar un esquema óptimo de asignación de recursos que asigne la cantidad de recursos sugerida por el sistema del sistema de aprendizaje automático de uso de recursos 160 para la SFC. En una realización, el componente asignador de recursos tiene acceso a información con respecto a las condiciones de la red/nube (p. ej., a través del sistema de monitorización y gestión de recursos 120) y utiliza esta información para  
 15 determinar un esquema de asignación de recursos óptimo que especifica o sugiere en qué parte de la red 100 deberían asignarse los recursos. El componente asignador de recursos puede considerar el estado actual de la red y la infraestructura, la información sobre las vNF 155 actualmente en funcionamiento, las políticas del operador sobre la asignación de recursos y otros tipos de información al determinar el esquema óptimo de asignación de recursos. El esquema de asignación de recursos puede reutilizar las vNF 155 existentes en la red o implementar nuevas vNF 155.  
 20 En una realización, el componente asignador de recursos puede aplicar otras políticas y validaciones al esquema de asignación de recursos. El esquema de asignación de recursos puede proporcionarse al sistema de monitorización y gestión de recursos 120 y el sistema de monitorización y gestión de recursos 120 puede aprovisionar la SFC en la red según el esquema de asignación de recursos.

25 La figura 3 es un diagrama que ilustra las entradas y salidas de un sistema de aprendizaje automático de uso de recursos, según algunas realizaciones.

El sistema de aprendizaje automático de uso de recursos 160 se entrena (utilizando técnicas de aprendizaje automático) con el uso de recursos de SFC anteriores aprovisionadas en la red 100 (datos 310). Se dice que el sistema de aprendizaje automático de uso de recursos 160 está "en línea" ya que se entrena utilizando datos en vivo. El sistema de aprendizaje automático de uso de recursos 160 puede entrenarse utilizando aprendizaje automático para  
 30 aprender cómo difiere el uso de recursos de diferentes SFC dependiendo de los puntos finales de las SFC, el tipo de tráfico transportado por las SFC, el tiempo, la ubicación de las SFC y otra información relacionada con las SFC y/u otra información contextual. Esto puede hacerse observando y aprendiendo (utilizando el aprendizaje automático) el comportamiento de uso de recursos de las SFC anteriores aprovisionadas en la red 100. Basándose en este entrenamiento, el sistema de aprendizaje automático de uso de recursos 160 puede determinar la cantidad de recursos que necesitará una SFC dada. Por ejemplo, si el sistema de aprendizaje automático de uso de recursos 160 ha observado previamente que una SFC en particular usó una cierta cantidad de recursos, entonces el sistema de aprendizaje automático de uso de recursos 160 puede determinar que cualquier SFC nueva que sea similar a esa SFC en particular utilizará una cantidad de recursos similar. Esta determinación puede ser una función compleja dependiendo de los puntos finales, los tipos de tráfico, el tiempo (p. ej., tráfico producido en diferentes momentos del día/semana/año), el tipo de funciones de red incluidas en las SFC y su comportamiento de rendimiento con respecto  
 40 a tipos de tráfico dados y otros parámetros.

La entrada al sistema de aprendizaje automático de uso de recursos 160 es una descripción de una SFC (datos 320). En una realización, la descripción de la SFC puede incluir información de origen/destino (es decir, puntos finales) de la SFC y una demanda de recursos de la SFC. A continuación se proporciona una descripción ilustrativa de una SFC:

```

45 {"chain1": {
    "order": ["DPI", "IDS"],
    "resources": {
        "CPU": [
            {"SF": "DPI", "value": 2},
50     {"SF": "IDS", "value": 4}
        ]
        "MEM": [
            {"SF": "DPI", "value": 16},
            {"SF": "IDS", "value": 8}
    ]
}
    
```

```

    ]
    "BW": 50
    },
    "EP src": "IP@EricssonSJ",
5    "EPdst": "IP@exitGW"
    }
}

```

La descripción ilustrativa de la SFC específica a Ericsson SJ como el origen de la SFC e Internet como el destino de la SFC. La descripción también especifica una cadena (es decir, "cadena 1") de funciones de servicio que consta de una función de servicio de inspección profunda de paquetes seguida de una función de servicio del sistema de detección de intrusos. La descripción también especifica la cantidad de CPU y memoria solicitada para cada una de las funciones de servicio, así como la cantidad de ancho de banda solicitado. En este ejemplo, se solicitan 2 CPU virtuales y 16 GB de memoria para la función de servicio de inspección profunda de paquetes y 4 CPU virtuales y 8 GB de memoria para la función de servicio del sistema de detección de intrusos. El ancho de banda solicitado es de 50 Gbps.

La cantidad de recursos solicitados para la SFC (p. ej., como se especifica en la descripción de la SFC) puede ser diferente de la cantidad de recursos que la SFC realmente consumirá. Por lo tanto, el aprovisionamiento de la cantidad de recursos que se especifica en la solicitud puede dar como resultado un aprovisionamiento insuficiente o excesivo de recursos para la SFC. El sistema de aprendizaje automático de uso de recursos 160 puede utilizar técnicas de aprendizaje automático para determinar con mayor precisión la cantidad de recursos que consumirá la SFC basándose en su entrenamiento/aprendizaje previo. El sistema de aprendizaje automático de uso de recursos 160 puede generar esta cantidad como la cantidad sugerida de recursos para aprovisionar a la SFC (datos 330). Por ejemplo, el sistema de aprendizaje automático de uso de recursos 160 puede sugerir que la función de servicio de inspección profunda de paquetes debería aprovisionarse con 3 CPU virtuales (en lugar de las 2 CPU virtuales solicitadas inicialmente) y que la función de servicio del sistema de detección de intrusos debería aprovisionarse con 16 GB de memoria (en lugar de los 8 GB de memoria solicitados inicialmente). Además, el sistema de aprendizaje automático de uso de recursos 160 puede sugerir que se aprovisionen 55 Gbps de ancho de banda para la SFC (en lugar de los 50 Gbps solicitados inicialmente).

En una realización, el sistema de aprendizaje automático de uso de recursos 160 incluye un mecanismo de realimentación de errores para ayudar al sistema de aprendizaje automático de uso de recursos 160 a aprender de cualquier resultado inexacto que produzca. El mecanismo de realimentación de errores compara la salida del sistema de aprendizaje automático de uso de recursos 160 con el uso real de recursos de la SFC, y realimenta la diferencia como un error al sistema de aprendizaje automático de uso de recursos 160. El sistema de aprendizaje automático de uso de recursos 160 puede utilizar la información de error para tomar mejores decisiones para futuras solicitudes. En una realización, el mecanismo de realimentación de errores verifica que el error no haya sido causado por factores externos tales como fallos de software/hardware.

La figura 4 es un diagrama que ilustra las etapas internas de un sistema de aprendizaje automático de uso de recursos, según algunas realizaciones.

El sistema de aprendizaje automático de uso de recursos 160 toma un vector de características 410 como entrada. El vector de características 410 puede incluir información que describe una SFC (p. ej., los datos 320 en la figura 3). Como se ilustra, el sistema de aprendizaje automático de uso de recursos 160 incluye dos etapas. La primera etapa agrupa la entrada, por ejemplo, basándose en la hora/día/año, origen/destino de la SFC u otras características/propiedades de la SFC descritas en el vector de características 410 y/u otra información contextual. La agrupación puede clasificar el vector de características 410 en una clase (p. ej., de clase 1 a clase X). La segunda etapa incluye una caja de aprendizaje automático 450 para cada clase. Un subconjunto del vector de características 410 (y la información contextual, p. ej., información de hora/día/año) y el error de la realimentación de corrección anterior se dirige a la caja de aprendizaje automático 450 apropiada en la segunda etapa, según lo determinado por la agrupación en la primera etapa. En una realización, el subconjunto del vector de características 410 puede excluir las características del vector de características original 410 que se utilizaron para la agrupación. En una realización, el subconjunto del vector de características 410 puede incluir características de ingeniería basadas en las características sin procesar en el vector de características 410. En una realización, el subconjunto del vector de características 410 puede incluir un intervalo de confianza de la agrupación realizada en la primera etapa. En la segunda etapa, la caja de aprendizaje automático 450 (que se seleccionó por la agrupación en la primera etapa) utiliza técnicas de aprendizaje automático para determinar la cantidad sugerida de recursos para aprovisionar a la SFC basándose en el subconjunto del vector de características 410 y la información contextual (p. ej., información de la hora/día/año) que recibió como entrada. Además, la caja de aprendizaje automático 450 puede utilizar la realimentación de error para aprender de sus errores. La caja de aprendizaje automático 450 a continuación genera

un vector de características ajustado 460 con valores ajustados (p. ej., ajustados con la cantidad sugerida de recursos para aprovisionar a la SFC según lo determinado por un caja de aprendizaje automático 450 en la segunda etapa).

5 La agrupación puede ser útil porque la cantidad de recursos que consume una SFC puede depender de la hora del día/semana/año, los puntos finales y otros factores. La agrupación puede proporcionar resultados más precisos al agrupar las SFC que tienen características/propiedades similares en una clase y proporcionar un subsistema de aprendizaje automático especializado e independiente para cada clase. Las clases pueden formarse basándose en la hora del día, el día del calendario, por puntos finales y por tipos de cadena. Los tipos de cadena podrían formarse basándose en las vNF de cómputo alto o las vNF de memoria alta o ambas. El número de clases puede configurarse, según sea necesario para lograr el nivel de precisión deseado.

10 Pueden utilizarse diversas técnicas de aprendizaje automático para implementar las etapas internas del sistema de aprendizaje automático de uso de recursos 160. Por ejemplo, la primera etapa puede implementarse utilizando un árbol de decisión y la segunda etapa puede implementarse utilizando una red neuronal (p. ej., una red neuronal bayesiana).

15 La figura 5 es un diagrama de flujo que ilustra un proceso para optimizar el uso de recursos de las SFC existentes en la red, según algunas realizaciones. Las operaciones en el diagrama de flujo pueden implementarse mediante un dispositivo informático 110 (p. ej., el componente de optimización de recursos de SFC 115).

El proceso se basa en la información proporcionada por un sistema de aprendizaje automático de escala automática 170 asociado con una vNF 155. El sistema de aprendizaje automático de escala automática 170 para la vNF 155 puede predecir cuándo se debe ajustar la cantidad de recursos aprovisionados para la vNF 155 y en qué medida. El sistema de aprendizaje automático de escala automática 170 para la vNF 155 se entrena en línea utilizando el aprendizaje automático para predecir cuándo se necesita la escala automática para la vNF 155. El sistema de aprendizaje automático de escala automática 170 recibe como entrada una cantidad de recursos actualmente utilizados por la vNF 155 y una cantidad de recursos actualmente disponibles para la vNF 155 (bloque 510). El sistema de aprendizaje automático de escala automática 170 utiliza el aprendizaje automático para determinar un ajuste sugerido a una cantidad de recursos aprovisionados para la vNF 155 basándose en la cantidad de recursos utilizados actualmente por la vNF 155 y la cantidad de recursos actualmente disponibles para la vNF 155 (bloque 520). En una realización, el sistema de aprendizaje automático de escala automática 170 también recibe información con respecto a las SFC que están utilizando actualmente la vNF como entrada y determina el ajuste sugerido a la cantidad de recursos aprovisionados para la vNF basándose en la información con respecto a las SFC que actualmente utilizan la vNF. El sistema de aprendizaje automático de escala automática 170 a continuación emite el ajuste sugerido a la cantidad de recursos aprovisionados para la vNF 155 (bloque 530). En una realización, el ajuste sugerido a la cantidad de recursos aprovisionados para la vNF 155 incluye un ajuste sugerido a una cantidad de recursos de CPU (p. ej., número de núcleos) aprovisionados para la vNF 155, un ajuste sugerido a la cantidad de memoria aprovisionada para la vNF 155, un ajuste sugerido a una cantidad de ancho de banda aprovisionado para la vNF 155, un ajuste sugerido a una cantidad de recursos de almacenamiento aprovisionados para la vNF 155, o un ajuste sugerido a cualquier otro tipo de recurso de red/informático aprovisionado para la vNF 155. En una realización, la vNF 155 incluye múltiples componentes de vNF. En una realización de este tipo, el ajuste sugerido puede incluir un ajuste sugerido a una cantidad de recursos aprovisionados por cada componente vNF de la vNF 155. En una realización, el ajuste sugerido incluye una bandera de urgencia. La bandera de urgencia indica si la necesidad de ajuste de recursos es urgente o no. En una realización, la bandera de urgencia puede ir acompañada de un valor que indica una cantidad máxima de tiempo en la que debe tener lugar el ajuste de recursos.

El dispositivo informático 110 obtiene, del sistema de aprendizaje automático de escala automática 170 asociado con la vNF 155, el ajuste sugerido a la cantidad de recursos aprovisionados para la vNF 155 (bloque 540) y proporciona el ajuste sugerido a un componente de reasignación de recursos (bloque 550). El componente de reasignación de recursos puede utilizar cualquier algoritmo o estrategia adecuada para determinar un esquema óptimo de reasignación de recursos que cumpla con los ajustes de recursos sugeridos. En una realización, el componente de reasignación de recursos tiene acceso a información con respecto a las condiciones de la red/nube (p. ej., a través del sistema de monitorización y gestión de recursos 120) y utiliza esta información para determinar un esquema de reasignación de recursos óptimo que especifica o sugiere en qué parte de la red 100 deberían asignarse y/o desasignarse los recursos. El componente de reasignación de recursos puede considerar el estado actual de la red y la infraestructura, la información sobre otras vNF 155 actualmente en funcionamiento, las políticas del operador sobre la asignación de recursos y otros tipos de información al determinar el esquema óptimo de reasignación de recursos. El esquema de reasignación preferiblemente debería minimizar el impacto en las SFC y servicios existentes, y minimizar la cantidad de migraciones de VM. El esquema de reasignación de recursos puede proporcionarse al sistema de monitorización y gestión de recursos 120 y un sistema de monitorización y gestión de recursos 120 puede reasignar los recursos según el esquema de reasignación de recursos.

La figura 6 es un diagrama que ilustra los componentes de una función de red virtual, según algunas realizaciones.

Como se ilustra, la vNF 155 incluye N componentes de vNF (componente 1 de vNF, componente 2 de vNF,..., componente N de vNF). Cada componente de vNF proporciona un subconjunto de la funcionalidad de vNF. Los componentes de vNF pueden comunicarse entre sí para realizar la funcionalidad de la vNF 155. El flujo de tráfico entre

los componentes de vNF puede denominarse en la presente memoria tráfico entre componentes de vNF o tráfico entre componentes. El tráfico que fluye a través de toda la vNF 155 se denomina en la presente memoria tráfico este-oeste. En una realización, la vNF 155 puede estar asociada con un sistema de aprendizaje automático de escala automática 170. El sistema de aprendizaje automático de escala automática 170 puede monitorizar cada uno de los componentes de vNF y utilizar técnicas de aprendizaje automático para determinar cuándo se necesita la escala automática para la vNF 155. La vNF 155 se ilustra con una línea de discontinua para indicar que los diversos componentes de la vNF 155 (p. ej., los componentes 1-N de vNF y el sistema de aprendizaje automático de escala automática 170) no necesitan estar coubicados o implementarse por el mismo dispositivo. En algunas realizaciones, los diversos componentes de la vNF 155 pueden implementarse de manera distributiva a través de la red 100.

La figura 7 es un diagrama que ilustra las entradas y salidas de un sistema de aprendizaje automático de escala automática, según algunas realizaciones.

El sistema de aprendizaje automático de escala automática 160 para una vNF 155 puede entrenarse utilizando técnicas de aprendizaje automático para saber cuándo cambiará el uso de recursos de la vNF 155. Esto puede hacerse observando y aprendiendo (utilizando aprendizaje automático) el comportamiento de uso de recursos de la vNF 155 y cómo cambia el uso de recursos de la vNF 155 dependiendo de las SFC que atraviesan la vNF, el tiempo, la ubicación de la vNF y otra información con respecto a la vNF y/u otra información contextual. La entrada al sistema de aprendizaje automático de escala automática 170 para una vNF 155 incluye la cantidad de recursos actualmente utilizados por la vNF 155 (datos 710) y la cantidad de recursos actualmente disponibles para la vNF 155 (datos 720). Por ejemplo, la cantidad de recursos utilizados actualmente por la vNF 155 puede incluir el uso de la CPU y el uso de la memoria por componente vNF de la vNF 155, así como el uso del ancho de banda de la vNF 155 (tanto el ancho de banda entre componentes como el ancho de banda de este-oeste utilizado por la vNF 155). La cantidad de recursos actualmente disponibles para la vNF 155 puede incluir la cantidad de recursos de CPU y memoria disponibles por componente vNF de la vNF 155, así como la cantidad de ancho de banda disponible para la vNF 155 (tanto el ancho de banda entre componentes como el ancho de banda de este-oeste disponible para la vNF 155). En una realización, la entrada al sistema de aprendizaje automático de escala automática 170 incluye información sobre las SFC que utilizan la vNF 155 (datos 730). El sistema de aprendizaje automático de escala automática 170 puede utilizar esta información para aprender la diferencia en el consumo de recursos cuando se une una nueva SFC a la carga de tráfico. En una realización, la entrada al sistema de aprendizaje automático de escala automática 170 incluye información de hora y fecha u otra información contextual. Las entradas al sistema de aprendizaje automático de escala automática 170 se proporcionan a modo de ejemplo y no de limitación. Las entradas al sistema de aprendizaje automático de escala automática 170 pueden incluir cualquier tipo de información que ayude al sistema de aprendizaje automático de escala automática 170 a determinar cuándo se necesita la escala automática y cuánta escala automática se necesita. Por ejemplo, las entradas también pueden incluir entradas externas de expertos en el dominio y efectos de decisiones de escala automática anteriores.

El sistema de aprendizaje automático de escala automática 170 utiliza técnicas de aprendizaje automático para determinar un ajuste sugerido a la cantidad de recursos provisionados para la vNF 155 basándose en la entrada. La salida del sistema de aprendizaje automático de escala automática 170 es el ajuste sugerido a la cantidad de recursos provisionados para la vNF 155 (datos 740). El ajuste sugerido puede sugerir que la cantidad de recursos provisionados para la vNF 155 debería aumentarse o reducirse, y también proporcionar una cantidad sugerida por la cual debería ajustarse la cantidad de recursos. En una realización, el ajuste sugerido puede sugerir un ajuste para cada componente vNF de la vNF 155. Por ejemplo, el ajuste sugerido puede sugerir aumentar o reducir la cantidad de recursos de CPU y de memoria provisionados para cada componente vNF de la vNF 155. El ajuste sugerido también puede sugerir aumentar o reducir la cantidad de ancho de banda entre componentes y/o el ancho de banda este-oeste provisionado para la vNF 155. En una realización, el ajuste sugerido incluye una bandera de urgencia que indica si la necesidad de ajuste es urgente. Por ejemplo, el sistema de aprendizaje automático de escala automática 170 puede establecer la bandera de urgencia en un valor afirmativo si la vNF 155 está muy cerca de utilizar todos los recursos provisionados para la vNF 155. En una realización, la bandera de urgencia puede ir acompañada de un valor que indica una cantidad máxima de tiempo en la que debe tener lugar el ajuste de recursos.

La figura 8 es un diagrama que ilustra el interior de un sistema de aprendizaje automático de escala automática, según algunas realizaciones.

Internamente, el sistema de aprendizaje automático de escala automática 170 para una vNF 155 realiza un seguimiento de los componentes de vNF de la vNF 155 y las conexiones entre componentes (las interconexiones de los componentes de vNF). El sistema de aprendizaje automático de escala automática 170 monitoriza y predice la utilización de recursos por componente de vNF y también monitoriza y predice el uso/rendimiento del ancho de banda de vNF, así como el ancho de banda este-oeste que pasa a través de la vNF 155. Basándose en esta información, el sistema de aprendizaje automático de escala automática 170 puede generar un ajuste de escala automática por componente de vNF.

Como se ilustra, el sistema de aprendizaje automático de escala automática 170 incluye una caja de aprendizaje automático para cada componente de vNF. La caja de aprendizaje automático para un componente de vNF recibe como entrada la utilización de recursos del componente de vNF, información de hora/fecha e información de SFC (de las SFC que utilizan el componente de vNF). Basándose en esta entrada, la caja de aprendizaje automático utiliza

técnicas de aprendizaje automático para determinar la cantidad de aumento/disminución de la escala automática para el componente de vNF y una bandera de urgencia que indica si la necesidad de la escala automática es urgente o no. La caja de aprendizaje automático proporciona esta información como salida. La salida de cada caja de aprendizaje automático se proporciona como una salida del sistema de aprendizaje automático de escala automática 170.

5 Las cajas de aprendizaje automático pueden utilizar cualquier técnica de aprendizaje automático adecuada. En una realización, las cajas de aprendizaje automático pueden utilizar un sistema de umbral de múltiples etapas tal como el que se describe en la Solicitud de Patente de Estados Unidos N.º 14/689.040 presentada el 16 de abril de 2015 para monitorizar la utilización/disponibilidad de recursos y predecir la necesidad de escala automática. La idea es monitorizar constantemente la utilización de recursos (p. ej., CPU, memoria y ancho de banda) y observar cuándo la  
10 tasa de cambio de la utilización de recursos varía con una tendencia específica (velocidad, valor). Pueden utilizarse otros tipos de técnicas para predecir la tasa de cambio en la utilización de recursos. Por ejemplo, pueden utilizarse los filtros de Kalman o las redes neuronales para predecir la tasa de cambio en la utilización de recursos.

En una realización, el sistema de aprendizaje automático de escala automática 170 aprende el efecto de diversas SFC en sus diversos componentes de vNF en diversos momentos (p. ej., hora del día, día de la semana, mes, etc.). Esta  
15 información puede permitir que el sistema de aprendizaje automático de escala automática 170 proporcione predicciones más precisas para la escala automática.

Los sistemas de aprendizaje automático descritos en la presente memoria pueden utilizarse de otras formas para optimizar el uso de recursos de las SFC. En una realización, el sistema de aprendizaje automático de uso de recursos  
20 160 puede servir como un sistema de recomendación. Por ejemplo, si un usuario solicita que se aprovisione una cierta cantidad de recursos para una SFC que es diferente de la cantidad real de recursos que consumirá la SFC (según lo determinado por el sistema de aprendizaje automático de uso de recursos 160), el sistema de aprendizaje automático de uso de recursos 160 puede proporcionar la cantidad sugerida de recursos para aprovisionar a la SFC al usuario y preguntar si el usuario desea aceptar la cantidad sugerida o continuar con la cantidad solicitada originalmente.

El aprendizaje realizado por los sistemas de aprendizaje automático puede importarse a una herramienta de gestión y planificación de red (p. ej., un simulador) y puede utilizarse para predecir el dimensionamiento requerido para los  
25 recursos de red (p. ej., dónde se necesitan más memoria y recursos de CPU). En una realización, los sistemas de aprendizaje automático pueden entrenarse previamente en un simulador antes de implementarse en una red activa. Los administradores de red pueden obtener un gráfico que muestra el historial de consumo de recursos y la proyección con algún margen de error, y estos datos pueden utilizarse para entrenar previamente los sistemas de aprendizaje  
30 automático.

La figura 9A ilustra la conectividad entre dispositivos de red (ND) dentro de una red ilustrativa, así como tres implementaciones ilustrativas de los ND, según algunas realizaciones. La figura 9A muestra los ND 900A-H y su  
35 conectividad por medio de líneas entre A-B, B-C, C-D, D-E, E-F, F-G y A-G, así como entre H y cada uno de A, C, D y G. Estos ND son dispositivos físicos, y la conectividad entre estos ND puede ser inalámbrica o por cable (a menudo denominada enlace). Una línea adicional que se extiende desde los ND 900A, E y F ilustra que estos ND actúan como puntos de entrada y salida para la red (y, por lo tanto, estos ND a veces se denominan ND perimetrales; mientras que los otros ND pueden llamarse ND centrales).

Dos de las implementaciones ilustrativas del ND en la figura 9A son: 1) un dispositivo de red de fin especial 902 que utiliza circuitos integrados específicos de la aplicación (ASIC) personalizados y un sistema operativo (SO) propietario;  
40 y 2) un dispositivo de red de fin general 904 que utiliza procesadores estándar disponibles en el mercado (COTS) y un SO estándar.

El dispositivo de red de fin especial 902 incluye hardware de red 910 que comprende recurso o recursos informáticos 912 (que normalmente incluyen un conjunto de uno o más procesadores), recurso o recursos de reenvío 914 (que normalmente incluyen uno o más ASIC y/o procesadores de red), e interfaces de red (NI) físicas 916 (a veces denominadas puertos físicos), así como medios de almacenamiento legibles por máquina no transitorios 918 que  
45 tienen almacenado en los mismos software de red 920. Una NI física es hardware en un ND a través del cual se realiza una conexión de red (p. ej., de forma inalámbrica a través de un controlador de interfaz de red inalámbrica (WNIC) o a través de la conexión de un cable a un puerto físico conectado a un controlador de interfaz de red (NIC)), tal como las mostradas por la conectividad entre los ND 900A-H. Durante la operación, el software de interconexión en red 920 puede ejecutarse por el hardware de interconexión en red 910 para instanciar un conjunto de una o más instancia o instancias de software de interconexión en red 922. Cada una de la instancia o instancias de software de red 922, y esa parte del hardware de interconexión en red 910 que ejecuta esa instancia de software de red (ya sea hardware especializado a esa instancia de software de interconexión en red y/o segmentos de tiempo de hardware compartidos temporalmente por esa instancia de software de interconexión en red con otras de la instancia o instancias del software  
50 de interconexión en red 922), forman un elemento de red virtual separado 930A-R. Cada uno del elemento o elementos de red virtual (VNE) 930A-R incluye un módulo de configuración y comunicación de control 932A-R (a veces denominado módulo de control local o módulo de comunicación de control) y tabla o tablas de reenvío 934A-R, de manera que un elemento de red virtual dado (p. ej., 930A) incluye el módulo de configuración y comunicación de control (p. ej., 932A), un conjunto de una o más tabla o tablas de reenvío (p. ej., 934A) y esa parte del hardware de  
55 interconexión en red 910 que ejecuta el elemento de red virtual (p. ej., 930A).

A menudo se considera física y/o lógicamente que el dispositivo de red de fin especial 902 incluye: 1) un plano de control de ND 924 (a veces denominado plano de control) que comprende el recurso o recursos informáticos 912 que ejecutan el módulo o módulos de comunicación y configuración de control 932A-R; y 2) un plano de reenvío de ND 926 (a veces denominado plano de reenvío, plano de datos o plano de medios) que comprende el recurso o recursos de reenvío 914 que utilizan la tabla o tablas de reenvío 934A-R y los NI físicos 916. A modo de ejemplo, cuando el ND es un encaminador (o está implementando la funcionalidad de encaminamiento), el plano de control de ND 924 (el recurso o recursos informáticos 912 que ejecutan el módulo o módulos de comunicación y configuración de control 932A-R) normalmente es responsable de participar en el control de cómo se enrutarán los datos (p. ej., paquetes) (p. ej., el siguiente salto para los datos y el NI físico saliente para esos datos) y almacenar esa información de encaminamiento en la tabla o tablas de reenvío 934A-R, y el plano de reenvío de ND 926 es responsable de recibir esos datos en los NI físicos 916 y reenviar esos datos a los apropiados de los NI físicos 916 basándose en la tabla o tablas de reenvío 934A-R.

La figura 9B ilustra una forma ilustrativa de implementar el dispositivo de red de fin especial 902 según algunas realizaciones. La figura 9B muestra un dispositivo de red de fin especial que incluye las tarjetas 938 (normalmente conectables en caliente). Mientras que en algunas realizaciones las tarjetas 938 son de dos tipos (una o más que funcionan como el plano de reenvío ND 926 (a veces denominadas tarjetas de línea) y una o más que funcionan para implementar el plano de control de ND 924 (a veces denominadas tarjetas de control)), las realizaciones alternativas pueden combinar la funcionalidad en una sola tarjeta y/o incluir tipos de tarjetas adicionales (p. ej., un tipo adicional de tarjeta se denomina tarjeta de servicio, tarjeta de recursos o tarjeta de múltiples aplicaciones). Una tarjeta de servicio puede proporcionar procesamiento especializado (p. ej., servicios de capa 4 a capa 7 (p. ej., cortafuegos, seguridad de protocolo de Internet (IPsec), capa de conexión segura (SSL)/seguridad de capa de transporte (TLS), sistema de detección de intrusos (IDS), comunicación entre pares (P2P), controlador perimetral de sesión de voz sobre IP (VoIP), pasarelas inalámbricas móviles (nodo de soporte de servicio general de paquetes de radio (GPRS) de pasarela (GGSN), pasarela de núcleo de paquete evolucionado (EPC))). A modo de ejemplo, puede utilizarse una tarjeta de servicio para terminar los túneles IPsec y ejecutar los algoritmos de autenticación y cifrado correspondientes. Estas tarjetas se acoplan entre sí a través de uno o más mecanismos de interconexión ilustrados como el panel de conexiones 936 (p. ej., una primera malla completa que acopla las tarjetas de línea y una segunda malla completa que acopla todas las tarjetas).

Volviendo a la figura 9A, el dispositivo de red de fin general 904 incluye el hardware 940 que comprende un conjunto de uno o más procesadores 942 (que a menudo son procesadores COTS) y el controlador o controladores de interfaz de red 944 (los NIC; también conocidos como tarjetas de interfaz de red) (que incluyen las NI físicas 946), así como medios de almacenamiento legibles por máquina no transitorios 948 que tienen almacenado en los mismos el software 950. Durante la operación, el procesador o procesadores 942 ejecutan el software 950 para instanciar uno o más conjuntos de una o más aplicaciones 964A-R. Si bien una realización no implementa la virtualización, las realizaciones alternativas pueden utilizar diferentes formas de virtualización, representadas por una capa de virtualización 954 y contenedores de software 962A-R. Por ejemplo, una realización alternativa de este tipo implementa la virtualización a nivel del sistema operativo, en cuyo caso la capa de virtualización 954 representa el núcleo de un sistema operativo (o una cuña que se ejecuta en un sistema operativo base) que permite la creación de múltiples contenedores de software 962A-R que pueden utilizarse cada uno para ejecutar uno de los conjuntos de aplicaciones 964A-R. En esta realización, los múltiples contenedores de software 962A-R (también llamados motores de virtualización, servidores privados virtuales o cárceles) son cada uno una instancia de espacio de usuario (normalmente un espacio de memoria virtual); estas instancias de espacio de usuario están separadas entre sí y separadas del espacio del núcleo en el que se ejecuta el sistema operativo; el conjunto de aplicaciones que se ejecutan en un espacio de usuario dado, a menos que se permita explícitamente, no puede acceder a la memoria de los otros procesos. Otra realización alternativa de este tipo implementa la virtualización completa, en cuyo caso: 1) la capa de virtualización 954 representa un hipervisor (a veces denominado monitor de máquina virtual (VMM)) o un hipervisor que se ejecuta sobre un sistema operativo anfitrión; y 2) cada uno de los contenedores de software 962A-R representa una forma estrechamente aislada de contenedor de software denominada máquina virtual que es ejecutada por el hipervisor y puede incluir un sistema operativo invitado. Una máquina virtual es una implementación de software de una máquina física que ejecuta programas como si se estuvieran ejecutando en una máquina física no virtualizada; y las aplicaciones generalmente no saben que se están ejecutando en una máquina virtual en lugar de ejecutarse en un dispositivo electrónico anfitrión de tipo "metal desnudo", aunque algunos sistemas proporcionan para-virtualización que permite que un sistema operativo o una aplicación sean conscientes de la presencia de virtualización para propósitos de optimización.

La creación de instancias de uno o más conjuntos de una o más aplicaciones 964A-R, así como la capa de virtualización 954 y los contenedores de software 962A-R, si se implementan, se denominan colectivamente instancia o instancias de software 952. Cada conjunto de aplicaciones 964A-R, el correspondiente contenedor de software 962A-R si está implementado, y la parte del hardware 940 que los ejecuta (ya sea hardware especializado a esa ejecución y/o segmentos de tiempo de hardware compartidos temporalmente por los contenedores de software 962A-R), forma uno o más elementos de red virtual separados 960A-R.

El elemento o elementos de red virtual 960A-R realizan una funcionalidad similar al elemento o elementos de red virtual 930A-R, p. ej., similar al módulo o módulos de configuración y comunicación de control 932A y a la tabla o tablas de reenvío 934A (esta virtualización del hardware 940 a veces se denomina virtualización de funciones de red (NFV)). Por lo tanto, NFV puede utilizarse para consolidar muchos tipos de equipos de red en hardware de servidor de alto

- volumen estándar de la industria, conmutadores físicos y almacenamiento físico, que podrían ubicarse en centros de datos, ND y equipos en las instalaciones del cliente (CPE). Sin embargo, diferentes realizaciones de la invención pueden implementar uno o más de los contenedores de software 962A-R de manera diferente. Por ejemplo, mientras que las realizaciones de la invención se ilustran correspondiendo cada contenedor de software 962A-R a un VNE 960A-R, las realizaciones alternativas pueden implementar esta correspondencia a un nivel de granularidad más preciso (p. ej., las máquinas virtuales de tarjeta de línea virtualizan tarjetas de línea, la máquina virtual de tarjeta de control virtualiza tarjetas de control, etc.); debe entenderse que las técnicas descritas en la presente memoria con referencia a una correspondencia de los contenedores de software 962A-R con los VNE también se aplican a realizaciones en las que se utiliza un nivel de granularidad más preciso.
- 5
- 10 En determinadas realizaciones, la capa de virtualización 954 incluye un conmutador virtual que proporciona servicios de reenvío similares a los de un conmutador Ethernet físico. Específicamente, este conmutador virtual reenvía el tráfico entre los contenedores de software 962A-R y el o los NIC 944, así como, opcionalmente, entre los contenedores de software 962A-R; Además, este conmutador virtual puede imponer el aislamiento de la red entre los VNE 960A-R que, por política, no pueden comunicarse entre sí (p. ej., respetando las redes de área local virtual (VLAN)).
- 15 El tercer ejemplo de implementación de ND en la figura 9A es un dispositivo de red híbrido 906, que incluye procesadores ASIC personalizados/SO propietario y procesadores de COTS/SO estándar en un solo ND o una sola tarjeta dentro de un ND. En ciertas realizaciones de un dispositivo de red híbrido de este tipo, una VM de plataforma (es decir, una VM que implementa la funcionalidad del dispositivo de red de fin especial 902) podría proporcionar para-virtualización al hardware de red presente en el dispositivo de red híbrido 906.
- 20 Independientemente de las implementaciones ilustrativas anteriores de un ND, cuando se está considerando uno solo de los múltiples VNE implementados por un ND (p. ej., solo uno de los VNE es parte de una red virtual dada) o donde solo se está implementando un único VNE actualmente por un ND, el término abreviado elemento de red (NE) se utiliza a veces para referirse a ese VNE. Además, en todas las implementaciones ilustrativas anteriores, cada uno de los VNE (p. ej., el o los VNE 930A-R, el o los VNE 960A-R, y aquellos en el dispositivo de red híbrido 906) recibe datos
- 25 en las NI físicas (p. ej., 916, 946) y reenvía esos datos a las correspondientes de las NI físicas (p. ej., 916, 946). Por ejemplo, una VNE que implementa la funcionalidad de un encaminador IP reenvía paquetes IP basándose en algo de la información del encabezado IP en el paquete IP; donde la información del encabezado IP incluye la dirección IP de origen, la dirección IP de destino, el puerto de origen, el puerto de destino (donde "puerto de origen" y "puerto de destino" se refieren en la presente memoria a puertos de protocolo, en contraposición a los puertos físicos de un ND),
- 30 protocolo de transporte (p. ej., valores de protocolo de datagramas de usuario (UDP), protocolo de control de transmisión (TCP) y servicios diferenciados (DSCP).
- La figura 9C ilustra diversas formas ilustrativas en las que los VNE pueden acoplarse según algunas realizaciones. La figura 9C muestra los VNE 970A.1-970A.P (y, opcionalmente, los VNE 970A.Q-970A.R) implementados en el ND 900A y el VNE 970H. 1 en ND 900H. En la figura 9C, los VNE 970A. 1-P están separados entre sí en el sentido de que
- 35 pueden recibir paquetes desde fuera del ND 900A y reenviar paquetes fuera del ND 900A; VNE 970A. 1 está acoplado con el VNE 970H. 1, y así comunican paquetes entre sus respectivos ND; VNE 970A.2-970A.3 puede opcionalmente reenviar paquetes entre ellos mismos sin reenviarlos fuera del ND 900A; y el VNE 970A.P puede ser opcionalmente el primero en una cadena de VNE que incluye el VNE 970A.Q seguido del VNE 970A.R (esto a veces se denomina encadenamiento de servicio dinámico, donde cada uno de los VNE en la serie de VNE proporciona un servicio
- 40 diferente, p. ej., uno o más servicios de red de capa 4-7). Mientras que la figura 9C ilustra diversas relaciones ilustrativas entre los VNE, las realizaciones alternativas pueden soportar otras relaciones (p. ej., más/menos VNE, más/menos cadenas de servicios dinámicos, múltiples cadenas de servicios dinámicos diferentes con algunos VNE comunes y algunos VNE diferentes).
- Los ND de la figura 9A, por ejemplo, pueden formar parte de Internet o de una red privada; y otros dispositivos electrónicos (no mostrados; tales como dispositivos de usuario final que incluyen estaciones de trabajo, ordenadores
- 45 portátiles, portátiles, tabletas, agendas electrónicas, teléfonos móviles, teléfonos inteligentes, teléfonos-tabletas, teléfonos multimedia, teléfonos con protocolo de voz sobre Internet (VOIP), terminales, reproductores multimedia portátiles, unidades GPS, dispositivos portátiles, sistemas de juegos, decodificadores de salón, electrodomésticos habilitados para Internet) pueden acoplarse a la red (directamente o a través de otras redes, tales como las redes de acceso) para comunicarse a través de la red (p. ej., Internet o redes privadas virtuales (VPN) superpuestas en (p. ej., tunelizadas a través de) Internet) entre sí (directamente o a través de servidores) y/o acceder a contenidos y/o servicios. Tal contenido y/o servicios generalmente se proporcionan por uno o más servidores (no mostrados) que pertenecen a un proveedor de servicios/contenido o uno o más dispositivos de usuario final (no mostrados) que participan en un servicio entre iguales (P2P), y pueden incluir, p. ej., páginas web públicas (p. ej., contenido gratuito,
- 55 escaparates, servicios de búsqueda), páginas web privadas (p. ej., páginas web a las que se accede con nombre de usuario/contraseña que proporcionan servicios de correo electrónico) y/o redes corporativas a través de VPN. Por ejemplo, los dispositivos de usuario final pueden acoplarse (p. ej., a través de un equipo en las instalaciones del cliente acoplado a una red de acceso (alámbrica o inalámbrica)) a ND perimetrales, que están acoplados (p. ej., a través de uno o más ND centrales) a otros ND perimetrales, que están acoplados a dispositivos electrónicos que actúan como servidores. Sin embargo, a través de la virtualización de cómputo y almacenamiento, uno o más de los dispositivos electrónicos que funcionan como los ND en la figura 9A también pueden albergar uno o más de dichos servidores (p. ej., en el caso del dispositivo de red de fin general 904, uno o más de los contenedores de software 962A-R pueden
- 60

funcionar como servidores; lo mismo sería cierto para el dispositivo de red híbrido 906; en el caso del dispositivo de red de fin especial 902, uno o más de dichos servidores también podrían ejecutarse en una capa de virtualización ejecutada por el o los recursos informáticos 912); en cuyo caso se dice que los servidores están coubicados con los VNE de ese ND.

5 Una red virtual es una abstracción lógica de una red física (tal como la de la figura 9A) que proporciona servicios de red (p. ej., servicios L2 y/o L3). Una red virtual puede implementarse como una red superpuesta (a veces denominada superposición de virtualización de red) que proporciona servicios de red (p. ej., servicios de capa 2 (L2, capa de enlace de datos) y/o capa 3 (L3, capa de red)) sobre una red subyacente (p. ej., una red L3, tal como una red de Protocolo de Internet (IP) que utiliza túneles (p. ej., encapsulación de encaminamiento genérico (GRE), protocolo de túnel de  
10 capa 2 (L2TP), IPSec) para crear la red superpuesta).

Un perímetro de virtualización de red (NVE) se ubica en el perímetro de la red subyacente y participa en la implementación de la virtualización de red; el lado de la NVE que mira hacia la red utiliza la red subyacente para tunelizar tramas hacia y desde otros NVE; el lado que mira hacia fuera del NVE envía y recibe datos hacia y desde sistemas fuera de la red. Una instancia de red virtual (VNI) es una instancia específica de una red virtual en un NVE (p. ej., un NE/VNE en un ND, una parte de un NE/VNE en un ND donde ese NE/VNE se divide en múltiples VNE a través de emulación); pueden instanciarse una o más VNI en un NVE (p. ej., como diferentes VNE en un ND). Un punto de acceso virtual (VAP) es un punto de conexión lógica en el NVE para conectar sistemas externos a una red virtual; un VAP puede ser puertos físicos o virtuales identificados a través de identificadores de interfaz lógica (p. ej., una ID de VLAN).

20 Ejemplos de servicios de red incluyen: 1) un servicio de emulación de LAN Ethernet (un servicio multipunto basado en Ethernet similar a un servicio de conmutación de etiquetas multiprotocolo (MPLS) o Ethernet VPN (EVPN) del grupo de trabajo de ingeniería de Internet (IETF)) en el que los sistemas externos están interconectados a través de la red mediante un entorno de LAN a través de la red subyacente (p. ej., un NVE proporciona VNI (instancias de conmutación virtual) de L2 separadas para diferentes redes virtuales de este tipo, y encapsulación de tunelización L3 (p. ej.,  
25 IP/MPLS) a través de la red subyacente); y 2) un servicio de reenvío de IP virtualizado (similar a IETF IP VPN (p. ej., protocolo de pasarela perimetral (BGP)/MPLS IPVPN) desde una perspectiva de definición de servicio) en el que los sistemas externos están interconectados a través de la red mediante un entorno L3 sobre la red subyacente (p. ej., un NVE proporciona las VNI (instancias de encaminamiento y reenvío) de L3 independientes para diferentes redes virtuales de este tipo, y encapsulación de tunelización L3 (p. ej., IP/MPLS) a través de la red subyacente)). Los servicios de red también pueden incluir capacidades de calidad de servicio (p. ej., marcado de clasificación de tráfico, acondicionamiento y planificación del tráfico), capacidades de seguridad (p. ej., filtros para proteger las instalaciones del cliente de ataques originados en la red, para evitar anuncios de ruta mal formados) y capacidades de gestión (p. ej.,  
30 detección completa y procesamiento).

La figura 9D ilustra una red con un solo elemento de red en cada uno de los ND de la figura 9A, y dentro de este enfoque directo contrasta un enfoque distribuido tradicional (utilizado comúnmente por los encaminadores tradicionales) con un enfoque centralizado para mantener la accesibilidad e información de reenvío (también denominado control de red), según algunas realizaciones. Específicamente, la figura 9D ilustra elementos de red (NE) 970A-H con la misma conectividad que los ND 900A-H de la figura 9A.

La figura 9D ilustra que el enfoque distribuido 972 distribuye la responsabilidad de generar la información de accesibilidad y reenvío a través de los NE 970A-H; en otras palabras, se distribuye el proceso de descubrimiento de vecinos y descubrimiento de topología.

Por ejemplo, cuando se utiliza el dispositivo de red de fin especial 902, el módulo o módulos de configuración y comunicación de control 932A-R del plano de control de ND 924 normalmente incluyen un módulo de información de accesibilidad y reenvío para implementar uno o más protocolos de encaminamiento (p. ej., un protocolo de pasarela exterior, tal como el protocolo de pasarela perimetral (BGP), el protocolo o protocolos de pasarela interior (IGP) (p. ej., abrir primero la ruta más corta (OSPF), sistema intermedio a sistema intermedio (IS-IS), protocolo de información de encaminamiento (RIP)), protocolo de distribución de etiquetas (LDP), protocolo de reserva de recursos (RSVP), así como RSVP-Ingeniería de tráfico (TE): Extensiones a RSVP para túneles LSP, conmutación de etiquetas multiprotocolo generalizada (GMPLS), señalización RSVP-TE que se comunica con otros NE para intercambiar rutas y a continuación selecciona esas rutas basándose en una o más métricas de encaminamiento. Por lo tanto, los NE 970A-H (p. ej., el recurso o recursos informáticos 912 que ejecutan el módulo o módulos de comunicación y configuración de control de control 932A-R) desempeñan su responsabilidad de participar en el control de cómo se enrutarán los datos (p. ej., los paquetes) (p. ej., el siguiente salto para los datos y la NI física saliente para esos datos) determinando distributivamente la accesibilidad dentro de la red y calculando su respectiva información de reenvío.  
45 Las rutas y adyacencias se almacenan en una o más estructuras de encaminamiento (p. ej., la base de información de encaminamiento (RIB), base de información de etiquetas (LIB), una o más estructuras de adyacencia) en el plano de control de ND 924. El plano de control de ND 924 programa el plano de reenvío de ND 926 con información (p. ej., información de ruta y adyacencia) basándose en la estructura o estructuras de encaminamiento. Por ejemplo, el plano de control ND 924 programa la información de ruta y adyacencia en una o más tablas de reenvío 934A-R (p. ej., base de información de reenvío (FIB), base de información de reenvío de etiquetas (LFIB) y una o más estructuras de adyacencia) en el plano de reenvío de ND 926. Para el reenvío de capa 2, el ND puede almacenar una o más tablas  
50  
55  
60

punteo que se utilizan para reenviar datos basándose en la información de capa 2 en esos datos. Si bien el ejemplo anterior utiliza el dispositivo de red de fin especial 902, el mismo enfoque distribuido 972 puede implementarse en el dispositivo de red de fin general 904 y en el dispositivo de red híbrido 906.

La figura 9D ilustra un enfoque centralizado 974 (también conocido como interconexión en red definida por software (SDN)) que desacopla el sistema que toma decisiones sobre dónde se envía el tráfico desde los sistemas subyacentes que reenvían el tráfico al destino seleccionado. El enfoque centralizado ilustrado 974 tiene la responsabilidad de generar información de accesibilidad y reenvío en un plano de control centralizado 976 (a veces denominado módulo de control SDN, controlador, controlador de red, controlador OpenFlow, controlador de SDN, nodo del plano de control, autoridad de virtualización de red o entidad de control de gestión) y, por lo tanto, el proceso de descubrimiento de vecinos y descubrimiento de topología está centralizado. El plano de control centralizado 976 tiene una interfaz de límite sur 982 con un plano de datos 980 (a veces denominado capa de infraestructura, plano de reenvío de red o plano de reenvío (que no debería confundirse con un plano de reenvío de ND)) que incluye los NE 970A-H (a veces denominados conmutadores, elementos de reenvío, elementos del plano de datos o nodos). El plano de control centralizado 976 incluye un controlador de red 978, que incluye un módulo de información de reenvío y accesibilidad centralizada 979 que determina la accesibilidad dentro de la red y distribuye la información de reenvío a los NE 970A-H del plano de datos 980 a través de la interfaz de límite sur 982 (que puede utilizar el protocolo OpenFlow). Por lo tanto, la inteligencia de la red está centralizada en el plano de control centralizado 976 que se ejecuta en dispositivos electrónicos que normalmente están separados de los ND. En una realización, la capa de aplicación 986 puede incluir un componente de optimización de recursos de SFC 981 que, cuando lo ejecuta un dispositivo informático, hace que el dispositivo informático realice operaciones de una o más realizaciones descritas anteriormente en la presente memoria. En una realización, algunas de las operaciones de una o más realizaciones también pueden realizarse en el plano de control centralizado 976.

Por ejemplo, cuando el dispositivo de red de fin especial 902 se utiliza en el plano de datos 980, cada uno del módulo o módulos configuración y comunicación de control 932A-R del plano de control ND 924 normalmente incluye un agente de control que proporciona el lado del VNE de la interfaz de límite sur 982. En este caso, el plano de control de ND 924 (el recurso o recursos informáticos 912 que ejecutan el módulo o módulos de configuración y comunicación de control 932A-R) realiza su responsabilidad de participar en el control de cómo se encaminarán los datos (p. ej., los paquetes) (p. e., el siguiente salto para los datos y la NI física saliente para esos datos) a través del agente de control que se comunica con el plano de control centralizado 976 para recibir la información de reenvío (y en algunos casos, la información de accesibilidad) del módulo de información de reenvío y accesibilidad centralizada 979 (debe entenderse que, en algunas realizaciones, el módulo o módulos de configuración y comunicación de control 932A-R, además de comunicarse con el plano de control centralizado 976, también pueden desempeñar algún papel en la determinación de la accesibilidad y/o el cálculo de la información de reenvío. aunque menos que en el caso de un enfoque distribuido; generalmente se considera que tales realizaciones caen bajo el enfoque centralizado 974, pero también pueden considerarse un enfoque híbrido).

Si bien el ejemplo anterior utiliza el dispositivo de red de fin especial 902, el mismo enfoque centralizado 974 puede implementarse con el dispositivo de red de fin general 904 (p. ej., cada uno de los VNE 960A-R realiza su responsabilidad de controlar cómo los datos (p. ej., paquetes) se van a encaminar (p. ej., el siguiente salto para los datos y la NI física saliente para esos datos) comunicándose con el plano de control centralizado 976 para recibir la información de reenvío (y, en algunos casos, la información de accesibilidad) del módulo de información de reenvío y accesibilidad centralizada 979; debe entenderse que, en algunas realizaciones, los VNE 960A-R, además de comunicarse con el plano de control centralizado 976, también pueden desempeñar algún papel en la determinación de la accesibilidad y/o el cálculo de la información de reenvío, aunque menos que en el caso de un enfoque distribuido) y el dispositivo de red híbrido 906. De hecho, el uso de técnicas de SDN puede mejorar las técnicas de NFV que se utilizan normalmente en las implementaciones del dispositivo de red de fin general 904 o del dispositivo de red híbrido 906, ya que NFV es capaz de admitir SDN al proporcionar una infraestructura en la que puede ejecutarse el software de SDN, y NFV y SDN tienen como objetivo hacer uso de hardware de servidor de servicios básicos y conmutadores físicos.

La figura 9D también muestra que el plano de control centralizado 976 tiene una interfaz de límite norte 984 a una capa de aplicación 986, en la que reside la aplicación o aplicaciones 988. El plano de control centralizado 976 tiene la capacidad de formar redes virtuales 992 (a veces denominado plano de reenvío lógico, servicios de red o redes superpuestas (siendo los NE 970A-H del plano de datos 980 la red subyacente)) para la aplicación o aplicaciones 988. Por lo tanto, el plano de control centralizado 976 mantiene una vista global de todos los ND y NE/VNE configurados, y correlaciona las redes virtuales con los ND subyacentes de manera eficiente (incluyendo el mantenimiento de estas asignaciones a medida que la red física cambia a través de fallo, adición o eliminación de hardware (ND, enlace o componente de ND)).

Mientras que la figura 9D muestra el enfoque distribuido 972 separado del enfoque centralizado 974, el esfuerzo del control de la red puede distribuirse de manera diferente o los dos pueden combinarse en ciertas realizaciones de la invención. Por ejemplo: 1) las realizaciones generalmente pueden utilizar el enfoque centralizado (SDN) 974, pero tienen ciertas funciones delegadas a los NE (p. ej., puede utilizarse el enfoque distribuido para implementar uno o más de monitorización de fallos, monitorización de rendimiento, conmutación de protección y primitivas para descubrimiento de vecinos y/o topología); o 2) las realizaciones de la invención pueden realizar el descubrimiento de vecinos y el

descubrimiento de topología a través del plano de control centralizado y los protocolos distribuidos, y los resultados se comparan para generar excepciones donde no concuerdan. Generalmente se considera que tales realizaciones caen bajo el enfoque centralizado 974, pero también pueden considerarse un enfoque híbrido.

5 Si bien la figura 9D ilustra el caso sencillo en el que cada uno de los ND 900A-H implementa un solo NE 970A-H, debe entenderse que los enfoques de control de red descritos con referencia a la figura 9D también funcionan para redes donde uno o más de los ND 900A-H implementan múltiples VNE (p. ej., VNE 930A-R, VNE 960A-R, los del dispositivo de red híbrido 906). Como alternativa o además, el controlador de red 978 también puede emular la implementación de múltiples VNE en un solo ND. Específicamente, en lugar de (o además de) implementar múltiples VNE en un solo ND, el controlador de red 978 puede presentar la implementación de un VNE/NE en un solo ND como múltiples VNE  
10 en las redes virtuales 992 (todas en la misma de la red o redes virtuales 992, cada una en una diferente de la red o redes virtuales 992, o alguna combinación). Por ejemplo, el controlador de red 978 puede hacer que un ND implemente un solo VNE (un NE) en la red subyacente y luego divida lógicamente los recursos de ese NE dentro del plano de control centralizado 976 para presentar diferentes VNE en la red o redes virtuales 992 (donde estos diferentes VNE en las redes superpuestas comparten los recursos de la implementación del VNE/NE único en el ND en la red subyacente).  
15

Por otra parte, las figuras 9E y 9F ilustran respectivamente abstracciones ilustrativas de NE y VNE que el controlador de red 978 puede presentar como parte de diferentes de las redes virtuales 992. La figura 9E ilustra el caso sencillo en el que cada uno de los ND 900A-H implementa un solo NE 970A-H (véase la figura 9D), pero el plano de control centralizado 976 ha abstraído múltiples de los NE en diferentes ND (los NE 970A-C y G-H) en (para representar) un único NE 970I en una de la red o redes virtuales 992 de la figura 9D, según algunas realizaciones. La figura 9E muestra que en esta red virtual, el NE 970I está acoplado al NE 970D y al 970F, que ambos aún están acoplados al NE 970E.  
20

La figura 9F ilustra un caso en el que se implementan múltiples VNE (VNE 970A.1 y VNE 970H.1) en diferentes ND (ND 900A y ND 900H) y se acoplan entre sí, y en el que el plano de control centralizado 976 ha abstraído estos múltiples VNE de manera que aparecen como un solo VNE 970T dentro de una de las redes virtuales 992 de la figura 9D, según algunas realizaciones. Por lo tanto, la abstracción de un NE o VNE puede abarcar múltiples ND.  
25

Si bien algunas realizaciones implementan el plano de control centralizado 976 como una sola entidad (p. ej., una sola instancia de software que se ejecuta en un solo dispositivo electrónico), las realizaciones alternativas pueden distribuir la funcionalidad entre múltiples entidades con fines de redundancia y/o escalabilidad (p. ej., múltiples instancias de software que se ejecuta en diferentes dispositivos electrónicos).  
30

De manera similar a las implementaciones de dispositivos de red, el dispositivo o dispositivos electrónicos que ejecutan el plano de control centralizado 976 y, por lo tanto, el controlador de red 978, incluyendo el módulo de información de reenvío y accesibilidad centralizada 979, pueden implementarse de diversas formas (p. ej., un dispositivo de fin especial, un dispositivo de fin general (p. ej., COTS) o un dispositivo híbrido). Este dispositivo o dispositivos electrónicos incluirían de manera similar un recurso o recursos informáticos, un conjunto o uno o más NIC físicos y un medio de almacenamiento legible por máquina no transitorio que tiene almacenado en el mismo el software del plano de control centralizado. Por ejemplo, la figura 10 ilustra un dispositivo de plano de control de fin general 1004 que incluye hardware 1040 que comprende un conjunto de uno o más procesador o procesadores 1042 (que a menudo son procesadores COTS) y controlador o controladores de interfaz de red 1044 (NIC); también conocidos como tarjetas de interfaz de red) (que incluyen NI físicas 1046), así como medios de almacenamiento legibles por máquina no transitorios 1048 que tienen almacenados en los mismos software de plano de control centralizado (CCP) 1050 y un componente de optimización de recursos de SFC 1051.  
35  
40

En las realizaciones que utilizan la virtualización informática, el procesador o procesadores 1042 normalmente ejecutan software para instanciar una capa de virtualización 1054 y un contenedor o contenedores de software 1062A-R (p. ej., con virtualización a nivel de sistema operativo, la capa de virtualización 1054 representa el núcleo de un sistema operativo (o una cuña que se ejecuta en un sistema operativo base) que permite la creación de múltiples contenedores de software 1062A-R (que representan instancias de espacio de usuario separadas y también llamados motores de virtualización, servidores privados virtuales o cárceles) que pueden utilizarse para ejecutar un conjunto de una o más aplicaciones; con virtualización completa, la capa de virtualización 1054 representa un hipervisor (a veces denominado monitor de máquina virtual (VMM)) o un hipervisor que se ejecuta sobre un sistema operativo anfitrión, y cada uno de los contenedores de software 1062A-R representa una forma estrechamente aislada de contenedor de software denominado una máquina virtual que ejecuta por el hipervisor y puede incluir un sistema operativo invitado; con para-virtualización, un sistema operativo o aplicación que se ejecuta con una máquina virtual puede ser consciente de la presencia de virtualización para fines de optimización). De nuevo, en las realizaciones en las que se utiliza la virtualización informática, durante la operación se ejecuta una instancia del software CCP 1050 (ilustrado como instancia CCP 1076A) dentro del contenedor de software 1062A en la capa de virtualización 1054. En las realizaciones en las que no se utiliza la virtualización informática, la instancia de CCP 1076A en la parte superior de un sistema operativo anfitrión se ejecuta en el dispositivo de plano de control de fin general de tipo "metal desnudo" 1004. La creación de instancias de la instancia de CCP 1076A, así como la capa de virtualización 1054 y los contenedores de software 1062A-R, si se implementan, se denominan colectivamente instancias o instancias de software 1052.  
45  
50  
55  
60

- En algunas realizaciones, la instancia de CCP 1076A incluye una instancia de controlador de red 1078. La instancia de controlador de red 1078 incluye una instancia de módulo de información de reenvío y accesibilidad centralizada 1079 (que es una capa de soporte intermedio que proporciona el contexto del controlador de red 978 al sistema operativo y que se comunica con los diversos NE), y una capa de aplicación de CCP 1080 (a veces denominada como una capa de aplicación) sobre la capa de soporte intermedio (que proporciona la inteligencia requerida para diversas operaciones de red, tales como protocolos, conocimiento de la situación de la red e interfaces de usuario). En un nivel más abstracto, esta capa de aplicación de CCP 1080 dentro del plano de control centralizado 976 funciona con una vista o vistas de red virtual (vista o vistas lógicas de la red) y la capa de soporte intermedio proporciona la conversión de las redes virtuales a la vista física.
- El componente de optimización de recursos de SFC 1051 puede ejecutarse mediante el hardware 1040 para realizar operaciones de una o más realizaciones de la presente invención como parte de las instancias de software 1052.
- El plano de control centralizado 976 transmite mensajes relevantes al plano de datos 980 basándose en los cálculos de la capa de aplicación de CCP 1080 y la correlación de capa de soporte intermedio para cada flujo. Un flujo puede definirse como un conjunto de paquetes cuyos encabezados coinciden con un patrón de bits dado; en este sentido, el reenvío de IP tradicional también es un reenvío basado en flujo en el que los flujos están definidos por la dirección IP de destino, por ejemplo; sin embargo, en otras implementaciones, el patrón de bits dado utilizado para una definición de flujo puede incluir más campos (p. ej., 10 o más) en los encabezados de paquete. Diferentes ND/NE/VNE del plano de datos 980 pueden recibir mensajes diferentes y, por lo tanto, información de reenvío diferente. El plano de datos 980 procesa estos mensajes y programa la información de flujo apropiada y las acciones correspondientes en las tablas de reenvío (a veces denominadas tablas de flujo) de los NE/VNE apropiados y, a continuación, los NE/VNE correlacionan los paquetes entrantes con los flujos representados en las tablas de reenvío y reenvían paquetes basándose en las coincidencias en las tablas de reenvío.
- Estándares tales como OpenFlow definen los protocolos utilizados para los mensajes, así como un modelo para procesar los paquetes. El modelo para procesar paquetes incluye el análisis de encabezados, la clasificación de paquetes y la toma de decisiones de reenvío. El análisis de encabezados describe cómo interpretar un paquete basándose en un conjunto bien conocido de protocolos. Se utilizan algunos campos de protocolo para crear una estructura (o clave) de coincidencia que se utilizará en la clasificación de paquetes (p. ej., un primer campo de clave podría ser una dirección de control de acceso al medio (MAC) de origen y un segundo campo de clave podría ser una dirección MAC de destino).
- La clasificación de paquetes implica ejecutar una búsqueda en la memoria para clasificar el paquete determinando qué entrada (también conocida como entrada de tabla de reenvío o entrada de flujo) en las tablas de reenvío coincide mejor con el paquete basándose en la estructura, o clave, de coincidencia de las entradas de tabla de reenvío. Es posible que muchos flujos representados en las entradas de tabla de reenvío puedan corresponder/coincidir con un paquete; en este caso, el sistema normalmente está configurado para determinar una entrada de tabla de reenvío de las muchas según un esquema definido (p. ej., seleccionando una primera entrada de tabla de reenvío que coincida). Las entradas de tabla de reenvío incluyen tanto un conjunto específico de criterios de coincidencia (un conjunto de valores o comodines, o una indicación de qué partes de un paquete deberían compararse con un valor/valores/comodines en particular, según lo definido por las capacidades de coincidencia - para campos específicos en el encabezado del paquete, o para algún otro contenido del paquete), como un conjunto de una o más acciones para que el plano de datos lleve a cabo al recibir un paquete coincidente. Por ejemplo, una acción puede ser insertar un encabezado en el paquete, para el paquete que utiliza un puerto en particular, inundar el paquete o simplemente descartar el paquete. Por lo tanto, una entrada de tabla de reenvío para paquetes IPv4/IPv6 con un puerto de destino de protocolo de control de transmisión (TCP) particular podría contener una acción que especifique que estos paquetes deberían descartarse.
- Se produce la toma de decisiones de reenvío y la realización de acciones, basándose en la entrada de tabla de reenvío identificada durante la clasificación del paquete, ejecutando el conjunto de acciones identificadas en la entrada de tabla de reenvío coincidente en el paquete.
- Sin embargo, cuando un paquete desconocido (por ejemplo, un "paquete perdido" o una "coincidencia perdida" como se utiliza en el lenguaje de OpenFlow) llega al plano de datos 980, el paquete (o un subconjunto del encabezado y contenido del paquete) normalmente se reenvía al plano de control centralizado 976. El plano de control centralizado 976 programará, a continuación, las entradas de tabla de reenvío en el plano de datos 980 para acomodar los paquetes que pertenecen al flujo del paquete desconocido. Una vez que una entrada específica de la tabla de reenvío se ha programado en el plano de datos 980 por el plano de control centralizado 976, el siguiente paquete con credenciales coincidentes coincidirá con esa entrada de tabla de reenvío y llevará a cabo el conjunto de acciones asociadas con esa entrada coincidente.
- Una interfaz de red (NI) puede ser física o virtual; y en el contexto de IP, una dirección de interfaz es una dirección IP asignada a una NI, ya sea una NI física o una NI virtual. Una NI virtual puede asociarse con una NI física, con otra interfaz virtual o ser independiente (p. ej., una interfaz de bucle invertido, una interfaz de protocolo punto a punto). Una NI (física o virtual) puede ser numerada (NI con una dirección IP) o no numerada (NI sin una dirección IP). Una interfaz de bucle invertido (y su dirección de bucle invertido) es un tipo específico de NI virtual (y dirección IP) de un NE/VNE

(físico o virtual) utilizado a menudo para fines de gestión; en la que dicha dirección IP se denomina dirección de bucle invertido nodal. La dirección o direcciones IP asignadas a la o las NI de un ND se denominan direcciones IP de ese ND; a un nivel más granular, la dirección o direcciones IP asignadas a la o las NI asignadas a un NE/VNE implementado en un ND pueden denominarse direcciones IP de ese NE/VNE.

5 Algunas partes de las descripciones detalladas anteriores se han presentado en términos de algoritmos y representaciones simbólicas de transacciones en bits de datos dentro de una memoria informática. Estas descripciones y representaciones algorítmicas son las formas utilizadas por los expertos en la técnica del procesamiento de datos para transmitir de manera más eficiente la esencia de su trabajo a otros expertos en la técnica. Un algoritmo se concibe en este punto, y generalmente, como una secuencia autoconsistente de transacciones que  
10 conducen a un resultado deseado. Las transacciones son aquellas que requieren manipulaciones físicas de cantidades físicas. Generalmente, aunque no necesariamente, estas cantidades toman la forma de señales eléctricas o magnéticas que se pueden almacenar, transferir, combinar, comparar y manipular de otra manera. A veces ha resultado conveniente, principalmente por razones de uso común, referirse a estas señales como bits, valores, elementos, símbolos, caracteres, términos, números o similares.

15 Sin embargo, debe tenerse en cuenta que todos estos términos y otros similares deben asociarse con las cantidades físicas apropiadas y son simplemente etiquetas convenientes aplicadas a estas cantidades. A menos que se indique específicamente lo contrario como es evidente a partir de la descripción anterior, se aprecia que a lo largo de la descripción, las descripciones que utilizan términos tales como "procesamiento" o "informático" o "cálculo" o "determinación" o "visualización" o similares, se refieren a la acción y procesos de un sistema informático, o dispositivo  
20 informático electrónico similar, que manipula y transforma datos representados como cantidades físicas (electrónicas) dentro de los registros y memorias del sistema informático a otros datos representados de manera similar como cantidades físicas dentro de las memorias o registros del sistema informático u otros dichos dispositivos de almacenamiento, transmisión o visualización de información.

25 Los algoritmos y visualizaciones que se presentan en la presente memoria no están inherentemente relacionados con ningún ordenador u otro aparato particular. Pueden utilizarse diversos sistemas de fin general con programas según las enseñanzas en la presente memoria, o puede resultar conveniente construir un aparato más especializado para realizar las transacciones de método requeridas. La estructura requerida para una variedad de estos sistemas aparecerá a partir de la descripción anterior. Además, las realizaciones de la presente invención no se describen con referencia a ningún lenguaje de programación particular. Se apreciará que pueden utilizarse una variedad de lenguajes  
30 de programación para implementar las enseñanzas de las realizaciones de la invención como se describe en la presente memoria.

Una realización de la invención puede ser un artículo de fabricación en el que un medio legible por máquina no transitorio (tal como una memoria microelectrónica) ha almacenado en el mismo instrucciones que programan uno o  
35 más componentes de procesamiento de datos (denominados genéricamente en este punto como "procesador") para realizar las operaciones descritas anteriormente. En otras realizaciones, algunas de estas operaciones podrían realizarse por componentes de hardware específicos que contienen lógica cableada (p. ej., bloques de filtro digital especializados y máquinas de estado). Esas operaciones podrían realizarse alternativamente por cualquier combinación de componentes de procesamiento de datos programados y componentes de circuitos cableados fijos.

40 En la memoria descriptiva anterior, se han descrito realizaciones de la invención con referencia a realizaciones ilustrativas específicas de la misma. Será evidente que pueden realizarse diversas modificaciones sin alejarse del alcance más amplio de la invención tal como se expone en las siguientes reivindicaciones. Por consiguiente, la memoria descriptiva y los dibujos deben considerarse en un sentido ilustrativo en lugar de en un sentido restrictivo.

45 A lo largo de la descripción, se han presentado realizaciones de la presente invención a través de diagramas de flujo. Se apreciará que el orden de las transacciones y las transacciones descritas en estos diagramas de flujo se conciben para fines ilustrativos y no para ser una limitación de la presente invención. Un experto en la técnica reconocerá que pueden hacerse variaciones en los diagramas de flujo sin alejarse del alcance más amplio de la invención tal como se expone en las siguientes reivindicaciones.

**REIVINDICACIONES**

1. Un método, por un sistema que comprende un dispositivo informático y un sistema de aprendizaje automático, ML, de escala automática asociado con una función de red virtual, vNF, para optimizar el uso de recursos de cadenas de funciones de servicio, SFC, en una red que utiliza aprendizaje automático, el dispositivo informático acoplado a un sistema de monitorización y gestión de recursos, comprendiendo el método:
- 5 obtener (540), por el dispositivo informático, del sistema de aprendizaje automático (ML) de escala automática asociado con la función de red virtual (vNF), un ajuste sugerido a una cantidad de recursos aprovisionados para la vNF, en donde el sistema de ML de escala automática se entrena en línea utilizando aprendizaje automático para predecir una cantidad de recursos a utilizar por la vNF, y en donde el sistema de ML de escala automática recibe (510) como entrada:
- 10 una cantidad de recursos utilizados actualmente por la vNF, una cantidad de recursos disponibles actualmente para la vNF e información con respecto a las SFC que utilizan actualmente la vNF, en donde la información con respecto a las SFC indica los puntos finales de cada SFC y una demanda de recursos de cada SFC, determina (520) utilizando aprendizaje automático:
- 15 el ajuste sugerido a la cantidad de recursos aprovisionados para la vNF basándose en la cantidad de recursos utilizados actualmente por la vNF, la cantidad de recursos disponibles actualmente para la vNF y la información con respecto a las SFC que están utilizando actualmente la vNF y los emite (530) el ajuste sugerido a la cantidad de recursos aprovisionados para la vNF; y proporcionar (550), por el dispositivo informático, el ajuste sugerido a la cantidad de recursos aprovisionados para la vNF a un componente de reasignación de recursos.
- 20 2. El método de la reivindicación 1, en donde la vNF incluye una pluralidad de componentes de vNF, y en donde el ajuste sugerido obtenido del sistema de ML de escala automática incluye un ajuste sugerido a una cantidad de recursos aprovisionados por componente vNF de la vNF.
3. El método de la reivindicación 1, en donde el ajuste sugerido obtenido del sistema de ML de escala automática y proporcionado al componente de reasignación de recursos incluye una bandera de urgencia, en donde la bandera de urgencia indica si la necesidad de ajuste de recursos es urgente.
- 25 4. El método de la reivindicación 1, en donde el ajuste sugerido obtenido del sistema de ML de escala automática indica uno cualquiera de un ajuste sugerido a una cantidad de recursos de unidad central de procesamiento (CPU) aprovisionados para la vNF, un ajuste sugerido a una cantidad de memoria aprovisionada para la vNF y un ajuste sugerido a una cantidad de ancho de banda aprovisionado para la vNF.
- 30 5. Un método, por un sistema que comprende un dispositivo informático y un sistema de aprendizaje automático, ML, de uso de recursos, para optimizar el uso de recursos de las cadenas de funciones de servicio, SFC, en una red que utiliza aprendizaje automático, estando el dispositivo informático acoplado a un sistema de monitorización y gestión de recursos, comprendiendo el método:
- 35 recibir (210), por el dispositivo informático, una solicitud para aprovisionar una SFC;
- proporcionar (220), por el dispositivo informático, una descripción de la SFC como una entrada a un sistema de aprendizaje automático (ML) de uso de recursos, en donde la descripción de la SFC indica los puntos finales de la SFC y una demanda de recursos de la SFC, en donde el sistema de ML de uso de recursos se ha entrenado previamente utilizando aprendizaje automático para aprender una cantidad de recursos utilizados por SFC anteriores aprovisionadas en la red, y en donde el sistema de ML de uso de recursos recibe (230) como entrada
- 40 la descripción de la SFC, determina (240) utilizando aprendizaje automático una cantidad sugerida de recursos para aprovisionar a la SFC basándose en la entrada, y emite (250) la cantidad sugerida de recursos para aprovisionar a la SFC;
- obtener (260), por el dispositivo informático del sistema de ML de uso de recursos, la cantidad sugerida de recursos para aprovisionar a la SFC; y
- 45 proporcionar (270), por el dispositivo informático, la cantidad sugerida de recursos para aprovisionar a la SFC a un componente asignador de recursos.
6. El método de la reivindicación 5, en donde la cantidad sugerida de recursos para aprovisionar a la SFC incluye una cualquiera de una cantidad sugerida de recursos de unidad central de procesamiento (CPU) para aprovisionar a la SFC, una cantidad sugerida de memoria para aprovisionar a la SFC y una cantidad sugerida de ancho de banda para aprovisionar a la SFC.
- 50 7. El método de la reivindicación 5, en donde la descripción de la SFC incluye una indicación de los puntos finales de la SFC y una lista de funciones de servicio incluidas en la SFC.
8. El método de la reivindicación 5, en donde la cantidad sugerida de recursos para proporcionar a la SFC incluye una cantidad sugerida de recursos para aprovisionar por función de servicio incluida en la SFC.
- 55 9. Un sistema que comprende un dispositivo informático (1004) y un sistema de aprendizaje automático, ML, de escala automática asociado con una función de red virtual, vNF, adaptado para optimizar el uso de recursos de cadenas de funciones de servicio, SFC, en una red que utiliza aprendizaje automático, el dispositivo informático que se acoplará a un sistema de monitorización y gestión de recursos, comprendiendo el dispositivo informático:

un medio de almacenamiento legible por máquina no transitorio (1048) que tiene almacenado en el mismo un componente de optimización de recursos de SFC (1051); y

un conjunto de uno o más procesadores (1042) acoplados al medio de almacenamiento legible por máquina no transitorio, el conjunto de uno o más procesadores configurado para ejecutar el componente de optimización de recursos de SFC, el componente de optimización de recursos de SFC configurado para obtener, del sistema de ML de escala automática asociado con la función de red virtual (vNF), un ajuste sugerido a una cantidad de recursos provisionados para la vNF, en donde el sistema de ML de escala automática está configurado para entrenarse en línea utilizando aprendizaje automático para predecir una cantidad de recursos a utilizar por la vNF, y en donde el sistema de ML de escala automática está configurado para recibir como entrada:

una cantidad de recursos utilizados actualmente por la vNF, una cantidad de recursos disponibles actualmente para la vNF e información con respecto a las SFC que están actualizando actualmente la vNF, en donde la información con respecto a las SFC indica los puntos finales de cada SFC y una demanda de recursos de cada SFC, determinar utilizando el aprendizaje automático:

el ajuste sugerido a la cantidad de recursos provisionados para la vNF basándose en la cantidad de recursos utilizados actualmente por la vNF, la cantidad de recursos disponibles actualmente para la vNF y la información con respecto a las SFC que están utilizando actualmente la vNF, y emitir el ajuste sugerido a la cantidad de recursos provisionados para la vNF, el componente de optimización de recursos de SFC configurado además para proporcionar el ajuste sugerido a la cantidad de recursos provisionados para la vNF a un componente de reasignación de recursos.

10. Un sistema que comprende un dispositivo informático (1004) y un sistema de aprendizaje automático, ML, de uso de recursos adaptado para optimizar el uso de recursos de cadenas de funciones de servicio, SFC, en una red que utiliza aprendizaje automático, el dispositivo informático que se acoplará a un sistema de monitorización y gestión de recursos, comprendiendo el dispositivo informático:

un medio de almacenamiento legible por máquina no transitorio (1048) que tiene almacenado en el mismo un componente de optimización de recursos de SFC (1051); y

un conjunto de uno o más procesadores (1042) acoplados al medio de almacenamiento legible por máquina no transitorio, el conjunto de uno o más procesadores configurado para ejecutar el componente de optimización de recursos de SFC, el componente de optimización de recursos de SFC configurado para recibir una solicitud para aprovisionar una SFC, proporcionar una descripción de la SFC como una entrada para el sistema de aprendizaje automático (ML) de uso de recursos, en donde el sistema de ML de uso de recursos está configurado para haberse entrenado previamente utilizando aprendizaje automático para aprender una cantidad de recursos utilizados por SFC anteriores provisionadas en la red, y en donde el sistema de ML de uso de recursos está configurado para recibir como entrada la descripción de la SFC, en donde la descripción de la SFC indica los puntos finales de la SFC y una demanda de recursos de la SFC, determinar utilizando aprendizaje automático una cantidad sugerida de recursos para aprovisionar a la SFC basándose en la entrada, y emitir la cantidad sugerida de recursos para aprovisionar la SFC, el componente de optimización de recursos de SFC configurado además para obtener, del sistema de ML de uso de recursos, la cantidad sugerida de recursos para aprovisionar a la SFC y proporcionar la cantidad sugerida de recursos para aprovisionar a la SFC a un componente asignador de recursos.

11. Un medio de almacenamiento legible por máquina no transitorio que tiene un código de ordenador almacenado en el mismo, que cuando se ejecuta por un conjunto de uno o más procesadores de un sistema que comprende un dispositivo informático y un sistema de aprendizaje automático, ML, de escala automática asociado con una función de red virtual, vNF, provoca que el sistema realice operaciones para optimizar el uso de recursos de cadenas de función de servicio, SFC, en una red que utiliza aprendizaje automático, el dispositivo informático que se acoplará a un sistema de monitorización y gestión de recursos, comprendiendo las operaciones:

obtener (540), por el dispositivo informático del sistema de aprendizaje automático (ML) de escala automática asociado con la función de red virtual (vNF), un ajuste sugerido a una cantidad de recursos provisionados para la vNF, en donde el sistema de ML de escala automática se entrena en línea utilizando aprendizaje automático para predecir una cantidad de recursos a utilizar por la vNF, y en donde el sistema de ML de escala automática recibe (510) como entrada:

una cantidad de recursos utilizados actualmente por la vNF, una cantidad de recursos disponibles actualmente para la vNF e información con respecto a las SFC que utilizan actualmente la vNF, en donde la información con respecto a las SFC indica los puntos finales de cada SFC y una demanda de recursos de cada SFC, determina (520) utilizando aprendizaje automático:

el ajuste sugerido a la cantidad de recursos provisionados para la vNF basándose en la cantidad de recursos utilizados actualmente por la vNF, la cantidad de recursos disponibles actualmente para la vNF y la información con respecto a las SFC que están utilizando actualmente la vNF y los emite (530) el ajuste sugerido a la cantidad de recursos provisionados para la vNF; y proporcionar (550), por el dispositivo informático, el ajuste sugerido a la cantidad de recursos provisionados para la vNF a un componente de reasignación de recursos.

12. El medio de almacenamiento legible por máquina no transitorio de la reivindicación 11, en donde la vNF incluye una pluralidad de componentes de vNF, y en donde el ajuste sugerido obtenido del sistema de ML de escala automática incluye un ajuste sugerido a una cantidad de recursos provisionados por componente vNF de la vNF.

13. El medio de almacenamiento legible por máquina no transitorio de la reivindicación 11, en donde el ajuste sugerido obtenido del sistema de ML de escala automática y proporcionado al componente de reasignación de recursos incluye

una bandera de urgencia, en donde la bandera de urgencia indica si la necesidad de un ajuste de recursos es urgente.

5 14. El medio de almacenamiento legible por máquina no transitorio de la reivindicación 11, en donde el ajuste sugerido obtenido del sistema de ML de escala automática indica uno cualquiera de un ajuste sugerido a una cantidad de recursos de unidad central de procesamiento (CPU) aprovisionados para la vNF, un ajuste sugerido a una cantidad de memoria aprovisionada para la vNF y un ajuste sugerido a una cantidad de ancho de banda aprovisionado para la vNF.

10 15. Un medio de almacenamiento legible por máquina no transitorio que tiene un código de ordenador almacenado en el mismo, que cuando se ejecuta por un conjunto de uno o más procesadores de un sistema que comprende un dispositivo informático y un sistema de aprendizaje automático, ML, de uso de recursos, provoca que el sistema realice operaciones para optimizar el uso de recursos de cadenas de funciones de servicio, SFC, en una red que utiliza aprendizaje automático, el dispositivo informático que se acoplará a un sistema de monitorización y gestión de recursos, comprendiendo las operaciones:

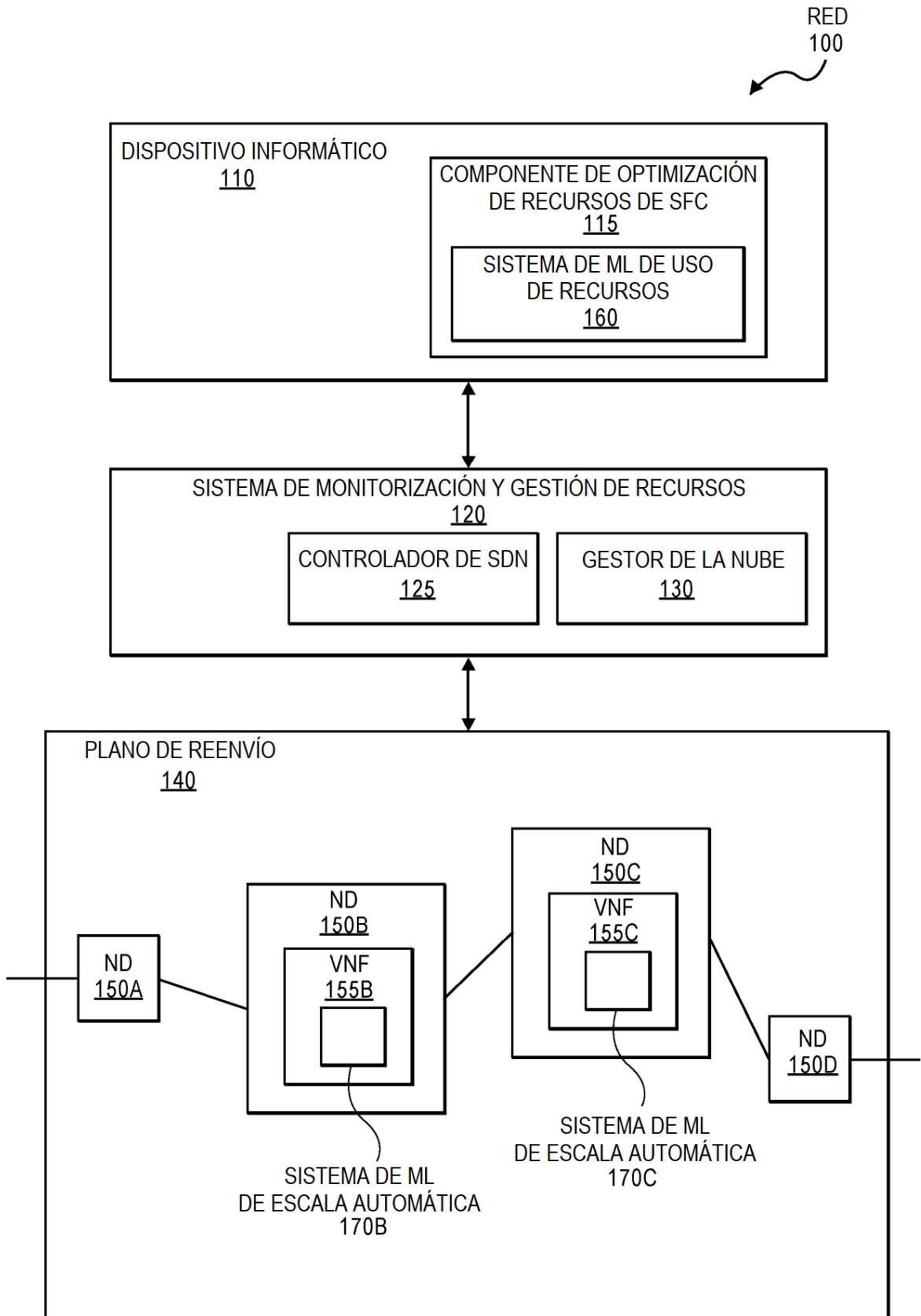
15 recibir (210), por el dispositivo informático, una solicitud para aprovisionar una SFC;  
 proporcionar (220), por el dispositivo informático, una descripción de la SFC como una entrada al sistema de aprendizaje automático (ML) de uso de recursos, en donde el sistema de ML de uso de recursos se ha entrenado previamente utilizando aprendizaje automático para aprender una cantidad de recursos utilizados por SFC anteriores aprovisionadas en la red, y en donde el sistema de ML de uso de recursos recibe (230) como entrada la descripción de la SFC, en donde la descripción de la SFC indica los puntos finales de la SFC y una demanda de recursos de la SFC determina (240) utilizando aprendizaje automático una cantidad sugerida de recursos para aprovisionar a la SFC basándose en la entrada, y emite (250) la cantidad sugerida de recursos para aprovisionar a la SFC;  
 20 obtener (260), por el dispositivo informático del sistema de ML de uso de recursos, la cantidad sugerida de recursos para aprovisionar a la SFC; y  
 proporcionar (270), por el dispositivo informático, la cantidad sugerida de recursos para aprovisionar a la SFC a un  
 25 componente asignador de recursos.

16. El medio de almacenamiento legible por máquina no transitorio de la reivindicación 15, en donde la cantidad sugerida de recursos para aprovisionar a la SFC incluye una cualquiera de una cantidad sugerida de recursos de unidad central de procesamiento (CPU) para aprovisionar a la SFC, una cantidad sugerida de memoria para aprovisionar a la SFC y una cantidad sugerida de ancho de banda para aprovisionar a la SFC.

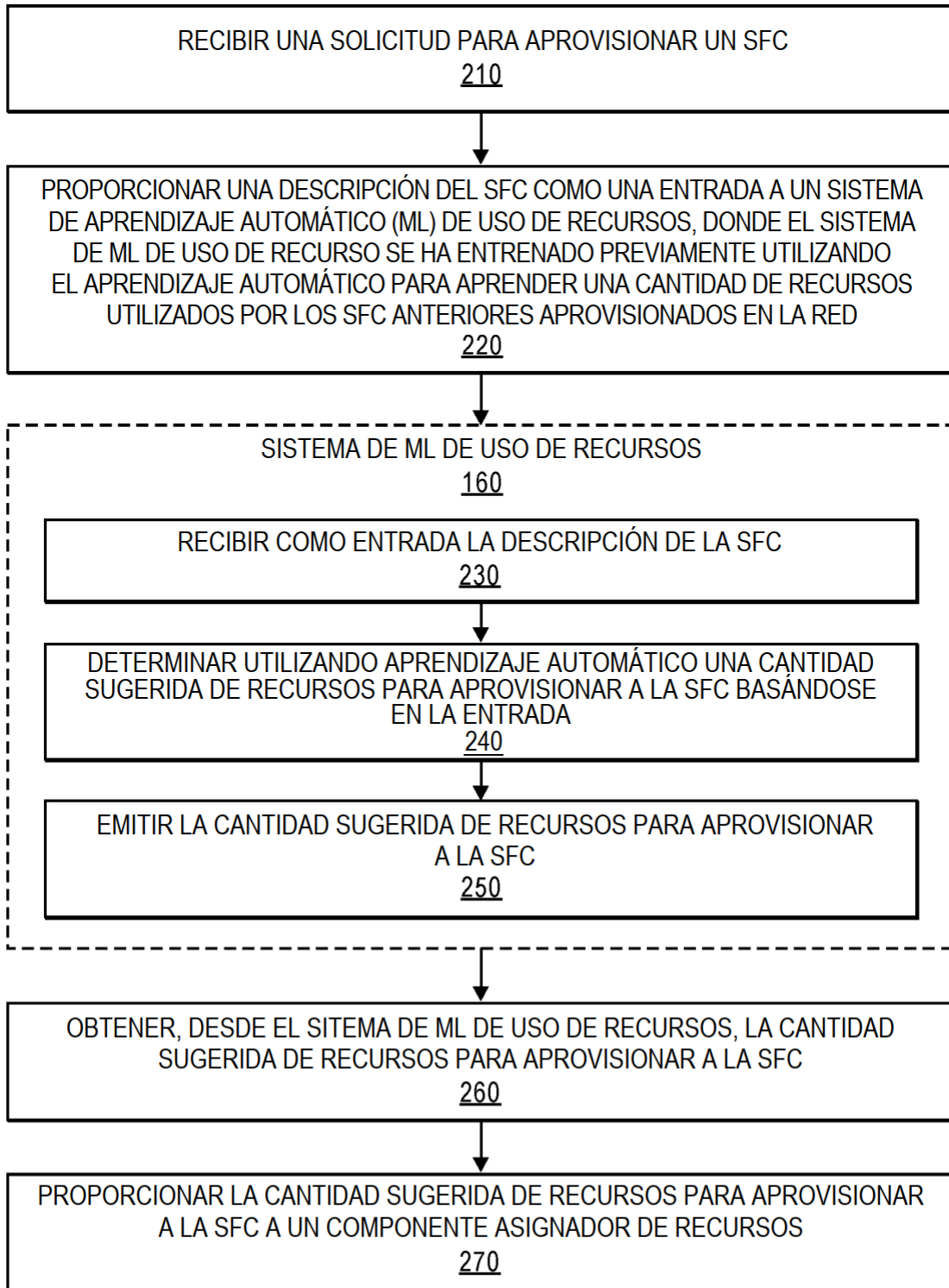
30 17. El medio de almacenamiento legible por máquina no transitorio de la reivindicación 15, en donde la descripción de la SFC incluye una indicación de los puntos finales de la SFC y una lista de funciones de servicio incluidas en la SFC.

18. El medio de almacenamiento legible por máquina no transitorio de la reivindicación 15, en donde la cantidad sugerida de recursos para proporcionar a la SFC incluye una cantidad sugerida de recursos para proporcionar por función de servicio incluida en la SFC.

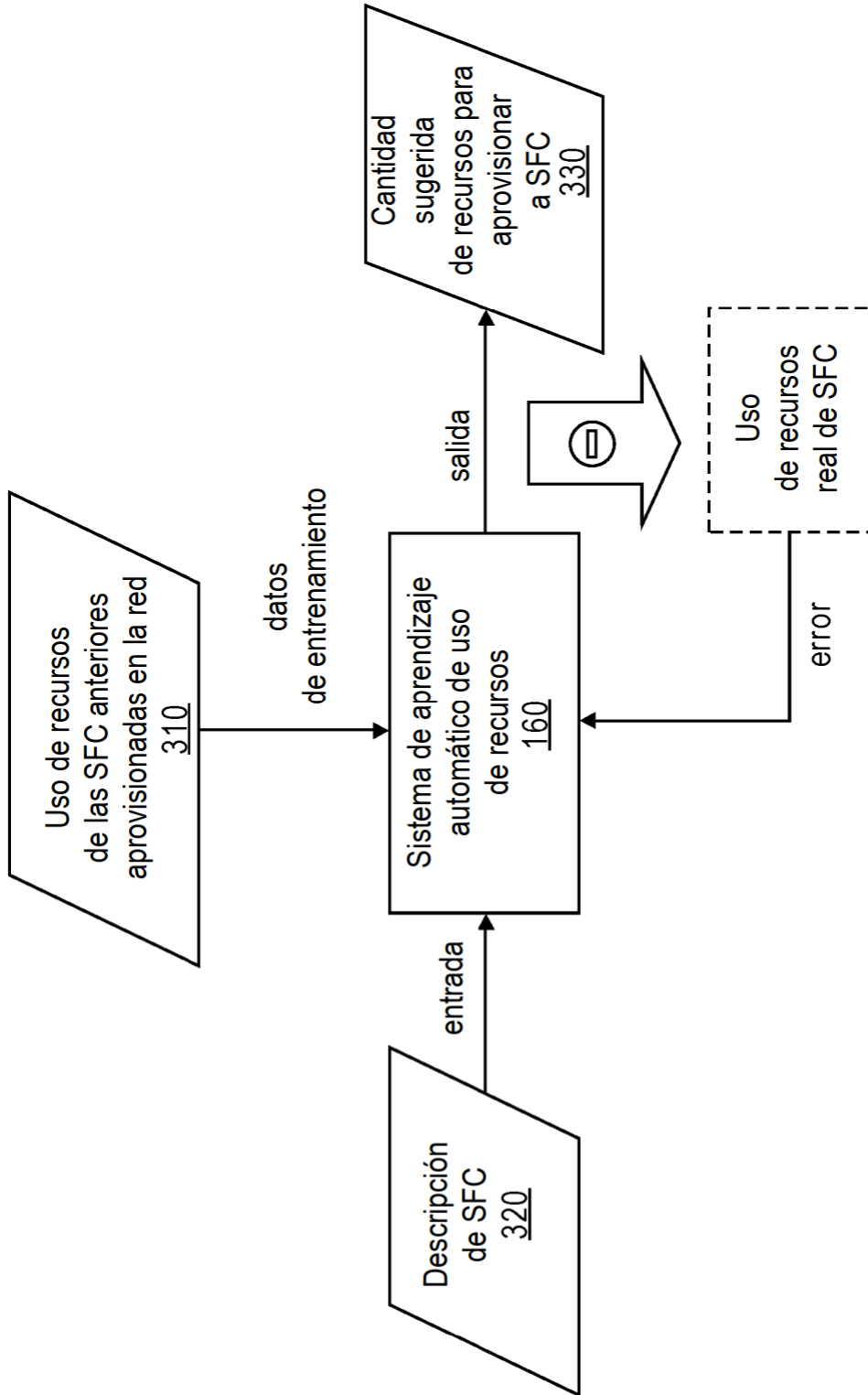
35



**FIG. 1**



**FIG. 2**



**FIG. 3**

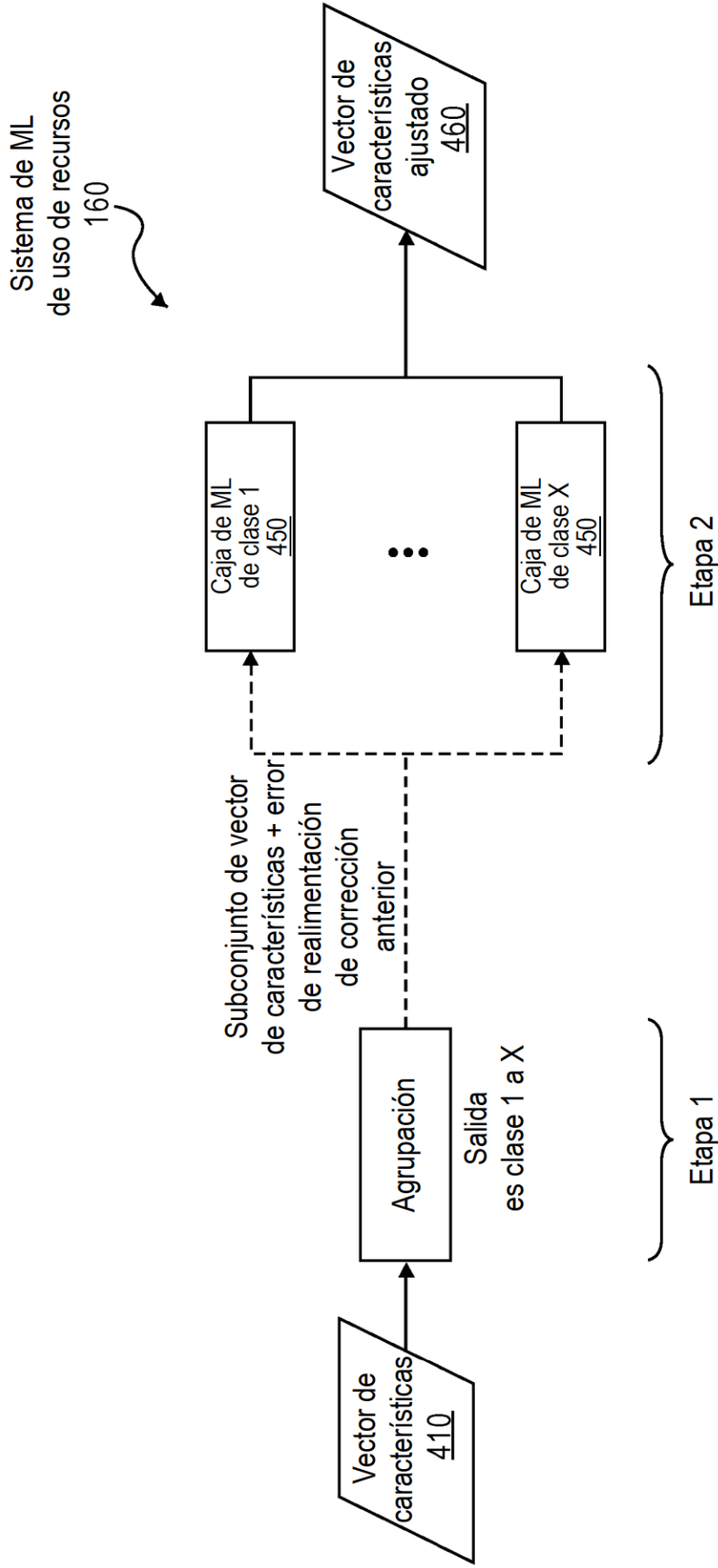
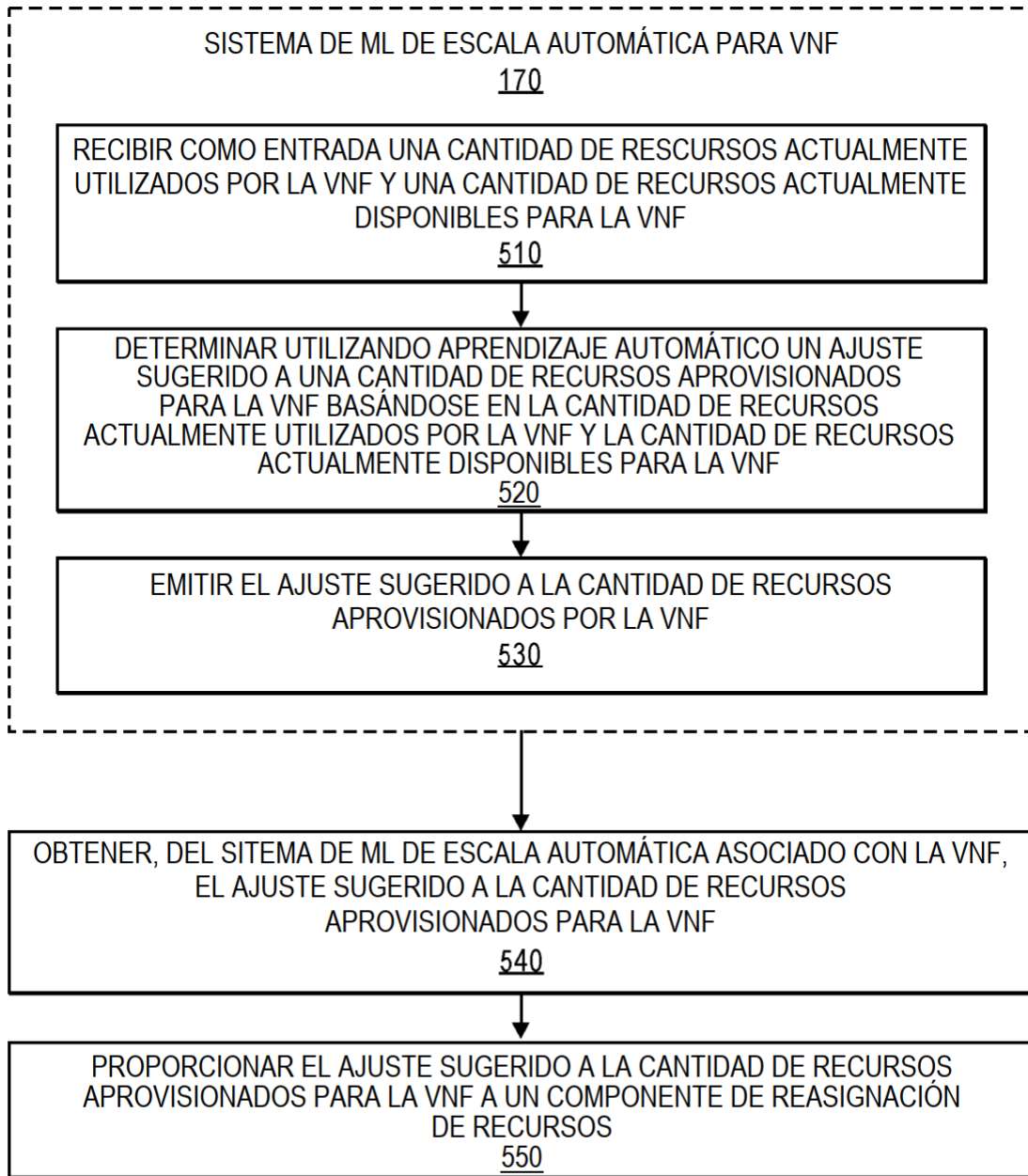
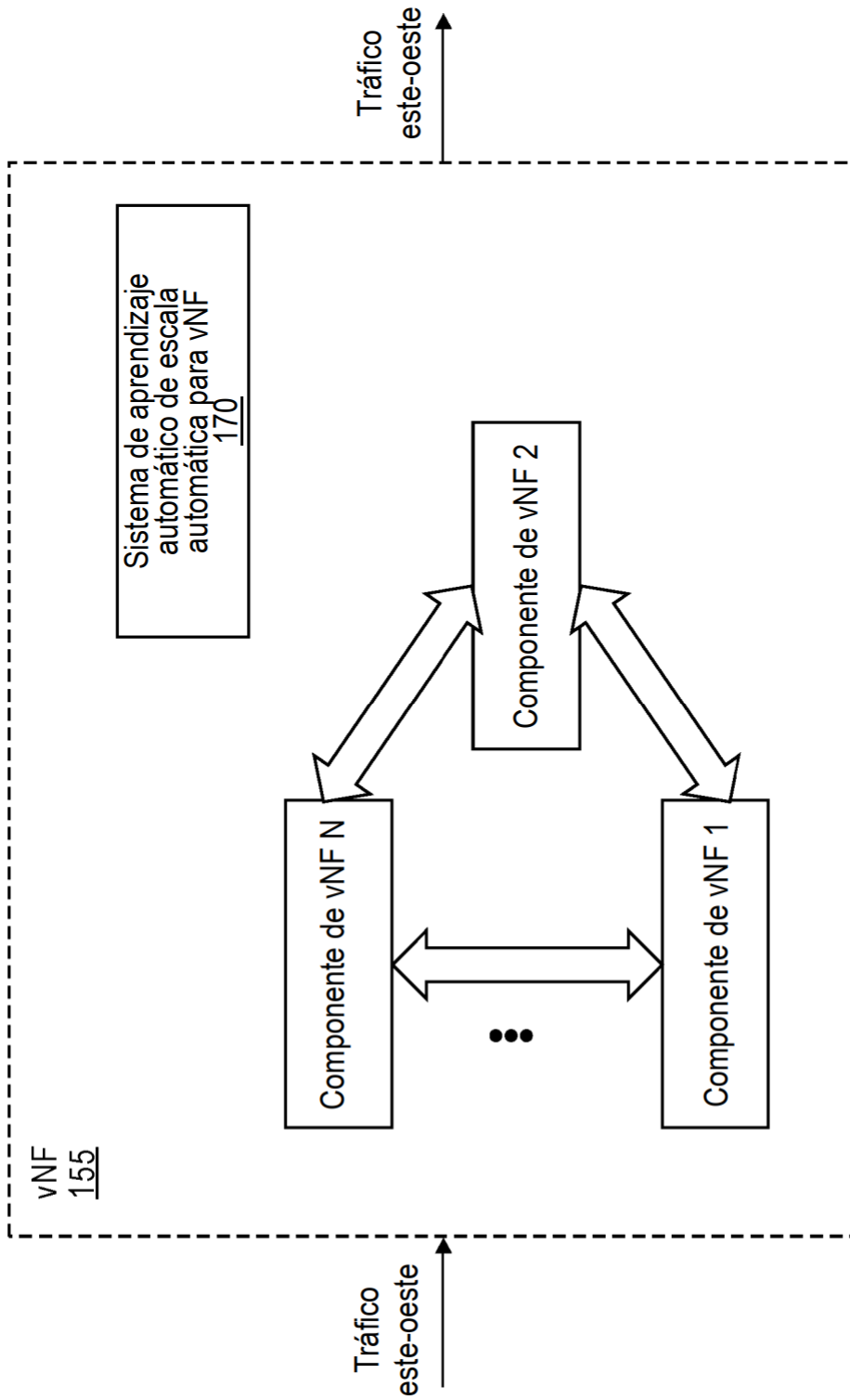


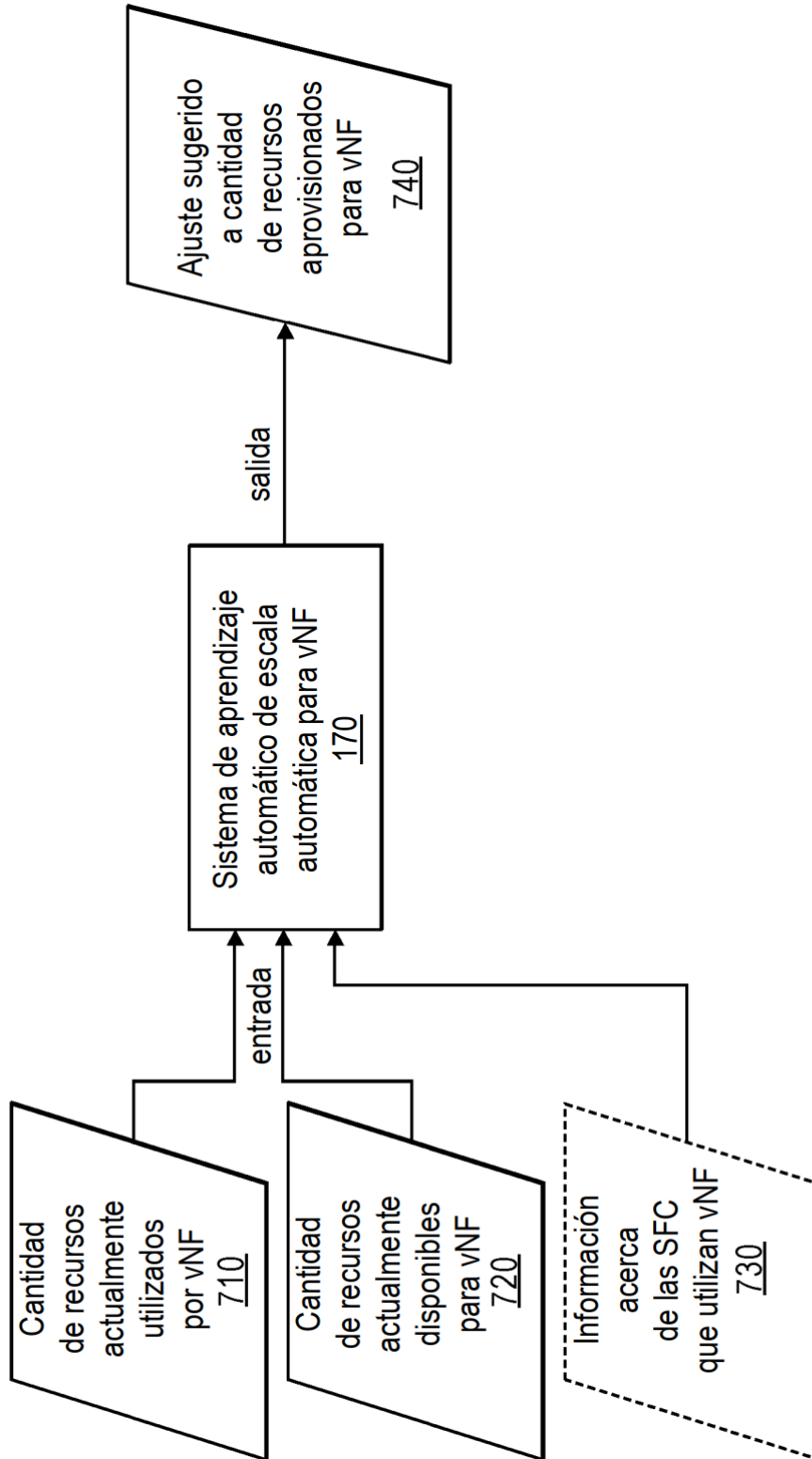
FIG. 4



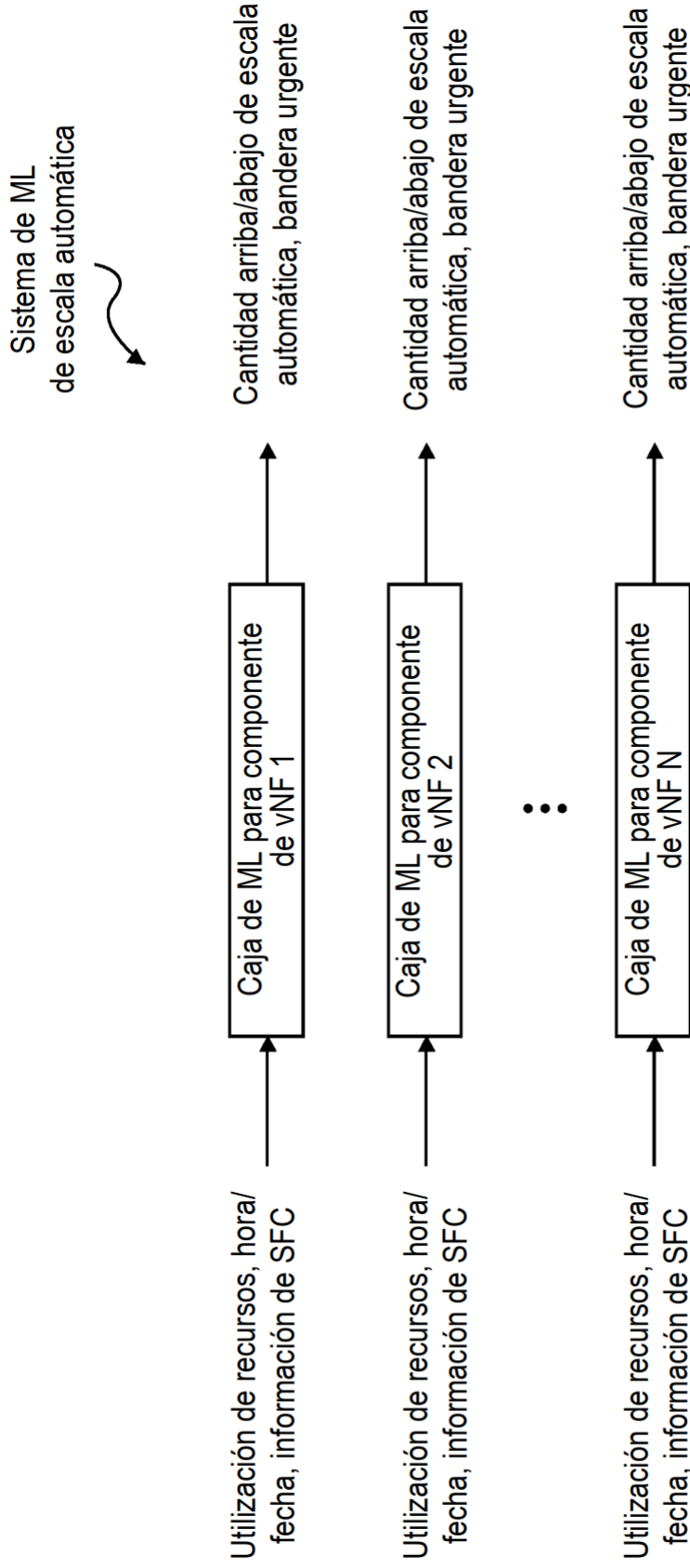
**FIG. 5**



**FIG. 6**



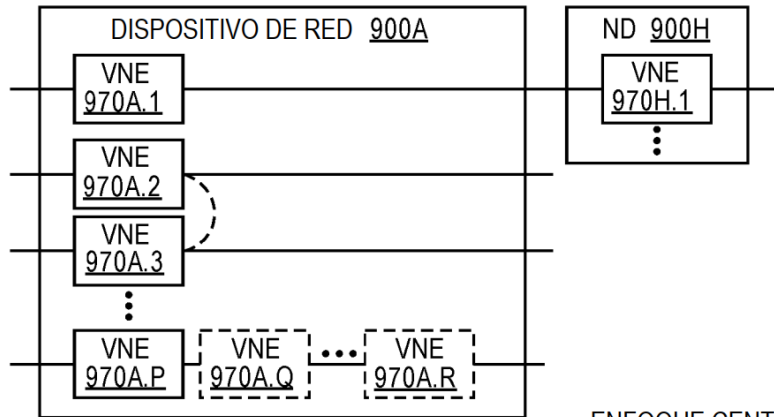
**FIG. 7**



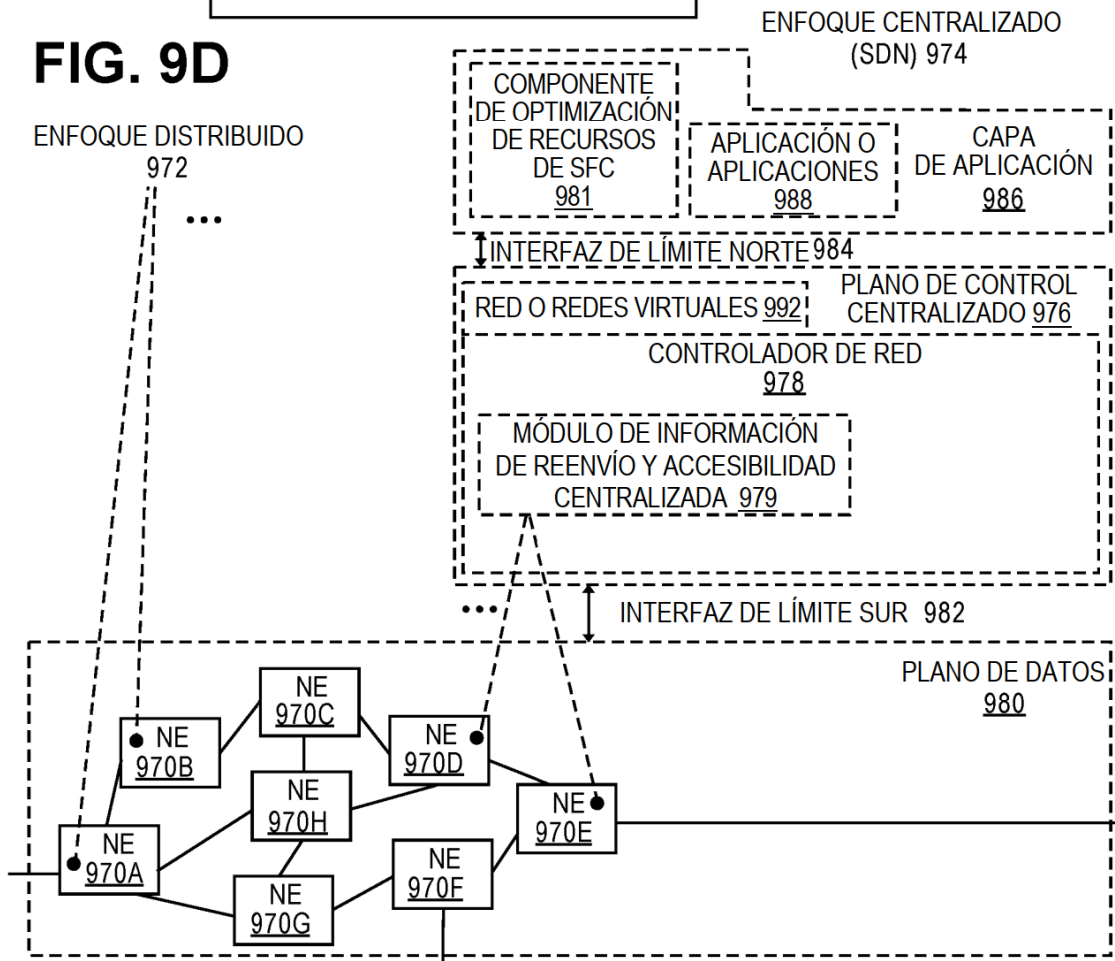
**FIG. 8**



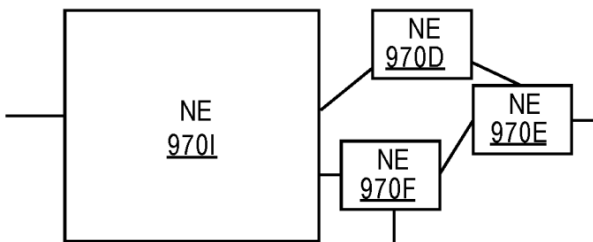
**FIG. 9C**



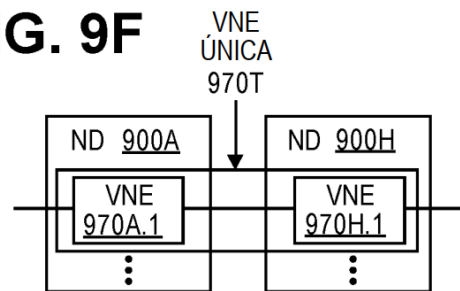
**FIG. 9D**



**FIG. 9E**



**FIG. 9F**



**FIG. 10**

