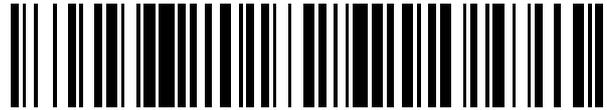


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 539 276**

51 Int. Cl.:

H04L 12/803 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **16.05.2011 E 11781406 (1)**

97 Fecha y número de publicación de la concesión europea: **06.05.2015 EP 2569907**

54 Título: **Métodos, sistemas y medios legibles por ordenador para balanceo de carga sin estado de flujos de tráfico de red**

30 Prioridad:

14.05.2010 US 334724 P
24.05.2010 US 786152

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
29.06.2015

73 Titular/es:

EXTREME NETWORKS, INC. (100.0%)
3585 Monroe Street
Santa Clara, CA 95051, US

72 Inventor/es:

GROSSER, DONALD B., JR. y
RICHARDSON, HOOD L., JR.

74 Agente/Representante:

DE ELZABURU MÁRQUEZ, Alberto

ES 2 539 276 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

Métodos, sistemas y medios legibles por ordenador para balanceo de carga sin estado de flujos de tráfico de red

Campo técnico

5 La materia objeto descrita en la presente memoria se refiere a balanceo de carga. Más particularmente, la materia objeto descrita en la presente memoria se refiere a métodos, sistemas y medios legibles por ordenador para balanceo de carga sin estado de flujos de tráfico de red.

Antecedentes

10 El balanceo de carga (también referido en la presente memoria como balanceo de carga de servidor (SLB)) es un método de distribución de carga de trabajo a través de una serie de dispositivos (por ejemplo, servidores) a fin de aumentar la productividad. Por ejemplo, se pueden usar múltiples servidores para proporcionar servicios para nodos clientes, tales como ordenadores, teléfonos móviles, etc., en una red. Las implementaciones actuales de SLB generalmente caen en 2 categorías principales: con estado y sin estado. Las implementaciones con estado pueden hacer el seguimiento y grabar el estado (por ejemplo, protocolos, puertos, direcciones, hora del día, números de secuencias, tamaños de paquetes, duración de flujo, etc.) acerca de cada flujo de red (por ejemplo, un flujo de paquetes relacionados) y, por tanto, puede proporcionar los esquemas de balanceo de carga más precisos, granulares y ricos en características. En particular, las implementaciones con estado pueden proporcionar filtrado de contenidos y conmutación consciente del contenido. Por ejemplo, un balanceador de carga con estado puede bloquear que ciertos flujos (por ejemplo, en base a identificadores de protocolo o números de puertos) sean balanceados de carga y un balanceador de carga con estado puede reconocer ciertos eventos (por ejemplo, que una sesión está establecida, que un tiempo de espera ha ocurrido, que una sesión ha terminado, etc.) o puede reconocer contenido (por ejemplo, ficheros que se transfieren) en flujos y, usando esta información, puede realizar las acciones adecuadas (por ejemplo, bloquear un flujo a un servidor particular para una aplicación, sesión o servicio dado). Aunque las implementaciones con estado pueden proporcionar diversas características, las implementaciones con estado generalmente son caras en recursos y pueden sufrir problemas de escalabilidad y rendimiento.

25 Por el contrario, las implementaciones sin estado generalmente son baratas en recursos. Las implementaciones sin estado típicamente distribuyen el tráfico usando troceado basado en paquetes sin almacenar el estado o la información acerca de cada flujo de red. Por ejemplo, se puede introducir información de campo de paquete (por ejemplo, dirección de protocolo de Internet (IP) de origen y dirección IP de destino) en un esquema de troceado para determinar qué servidor recibe cada paquete. Dado que los valores de troceado generalmente son recursos baratos de calcular, estas implementaciones sin estado generalmente son más rápidas y más fáciles de implementar. Aunque las implementaciones sin estado pueden proporcionar mejor rendimiento (por ejemplo, mayor caudal de paquetes) que las implementaciones con estado, las implementaciones sin estado carecen de características y tienen limitaciones. Por ejemplo, las implementaciones sin estado pueden carecer de chivato o conocimiento de sesión. Adicionalmente, la persistencia de flujo en implementaciones sin estado puede ser generalmente menos fiable. Es decir, las implementaciones sin estado generalmente pueden ser menos consistentes en la distribución de paquetes de un flujo de tráfico de red al mismo servidor. Además, los balanceadores de carga sin estado convencionales requieren esquemas de direccionamiento restrictivos que pueden limitar la granularidad en la asignación de cargas de trabajo de balanceo de carga.

40 La publicación de la solicitud de patente de Estado Unidos US 2009/030773 describe un método de conexión en red de una pluralidad de servidores juntos dentro de un centro de datos. El método según esta descripción incluye el paso de direccionamiento de un paquete de datos para entrega a un servidor de destino proporcionando la dirección del servidor de destino como una dirección plana. El método además incluye los pasos de obtener información de encaminamiento para encaminar el paquete al servidor de destino.

45 Por consiguiente, existe una necesidad de métodos, sistemas y medios legibles por ordenador mejorados para balanceo de carga sin estado de flujos de tráfico de red.

Compendio

50 Según un aspecto, la materia objeto descrita en la presente memoria incluye un método para realizar balanceo de carga sin estado de flujos de tráfico de red. El método ocurre en un dispositivo de reenvío de paquetes de capa 3 y de conmutación de capa 2. El método incluye responder a peticiones de protocolo de resolución de direcciones (ARP) de los clientes, las peticiones ARP que incluyen una dirección IP virtual (VIP) compartida por el dispositivo y una pluralidad de servidores acoplados al dispositivo, con la dirección de control de acceso al medio (MAC) del dispositivo. El método también incluye recibir, desde los clientes, paquetes dirigidos a la dirección VIP y que tienen la dirección MAC del dispositivo. El método además incluye la compartición de carga de los paquetes entre los servidores usando una operación de reenvío de capa 3 que parece a los clientes como una operación de conmutación de capa 2.

55 Según otro aspecto, la materia objeto descrita en la presente memoria incluye un dispositivo de reenvío de paquetes de capa 3 y de conmutación de capa 2 para realizar balanceo de carga sin estado de flujos de tráfico de red. El

5 dispositivo de reenvío de paquetes de capa 3 y de conmutación de capa 2 incluye una función intermediaria de protocolo de resolución de direcciones (ARP) para responder a las peticiones ARP desde clientes, las peticiones ARP que incluyen una dirección IP virtual (VIP) compartida por el dispositivo y una pluralidad de servidores acoplados al dispositivo, con la dirección de control de acceso al medio (MAC) del dispositivo. El dispositivo de reenvío de paquetes de capa 3 y de conmutación de capa 2 también incluye un módulo de balanceo de carga para recibir, desde los clientes, paquetes dirigidos a la dirección VIP y que tienen la dirección MAC del dispositivo y para compartición de carga de los paquetes entre los servidores usando una operación de reenvío de capa 3 que parece a los clientes como una operación de conmutación de capa 2.

10 La materia objeto descrita en la presente memoria para balanceo de carga sin estado de flujo de tráfico de red se puede implementar en hardware, software, microprograma o cualquier combinación de los mismos. Por tanto, los términos "función" o "módulo" como se usan en la presente memoria se refieren a hardware, software y/o microprograma para implementar la característica que se describe. En una implementación ejemplar, la materia objeto descrita en la presente memoria se puede implementar usando un medio legible por ordenador no transitorio que tiene almacenadas en el mismo instrucciones ejecutables por ordenador que cuando se ejecutan por el procesador de un ordenador controlan el ordenador para realizar los pasos. Los medios legibles por ordenador ejemplares adecuados para implementar la materia objeto descrita en la presente memoria incluyen medios legibles por ordenador no transitorios, tales como dispositivos de memoria de disco, dispositivos de memoria de circuito integrado, dispositivos de lógica programable y circuitos integrados de aplicaciones específicas. Además, un medio legible por ordenador que implementa la materia objeto descrita en la presente memoria se puede situar en un dispositivo o plataforma de cálculo única o se puede distribuir a través de múltiples dispositivos o plataformas de cálculo.

Breve descripción de los dibujos

La Figura 1 es un diagrama de red de un sistema ejemplar para balanceo de carga sin estado según una realización de la materia objeto descrita en la presente memoria;

25 La Figura 2 es un diagrama de una tabla de reenvío de capa 3 y estructura de datos ECMP según una realización de la materia objeto descrita en la presente memoria;

La Figura 3 es un diagrama de flujo de mensajes que ilustra mensajes ejemplares asociados con la realización de balanceo de carga sin estado para flujos de tráfico de red;

La Figura 4 es un diagrama de bloques de una arquitectura interna ejemplar de un dispositivo de reenvío de capa 2 con funcionalidad de capa 3 para realizar un balanceo de carga sin estado descrito en la presente memoria; y

30 La Figura 5 es un diagrama de flujo de un proceso ejemplar para realizar un balanceo de carga sin estado de flujos de tráfico de red según una realización de la materia objeto descrita en la presente memoria.

Descripción detallada

35 Como se usa en la presente memoria, balanceo de carga (también referido en la presente memoria como balanceo de carga de servidor (SLB)) se refiere a compartición, balanceo, división o de otro modo dividir carga o tráfico de red entre uno o más dispositivos (por ejemplo, servidores). Como se usa en la presente memoria, carga de red, tráfico de red, carga o tráfico se refieren a paquetes que viajan en, a o desde una red. Como se usa en la presente memoria, los flujos de tráfico de red, flujos de tráfico, flujos de red, flujos y similares se refieren a uno o más paquetes relacionados que viajan en, a o desde una red (por ejemplo, paquetes en una sesión, flujo unidireccional de paquetes desde o a un origen o destino o paquetes desde o a una entidad particular). Como se usa en la presente memoria, nodo de red, nodo cliente y similares se refieren a dispositivos para comunicar con uno o más servidores. Como se usa en la presente memoria, servidores se refieren a cualquier dispositivo para proporcionar servicios o mensajes de procesamiento de los clientes. Como se usa en la presente memoria, dominio de capa 2 o dominio de difusión de capa 2 se refiere a nodos capaces de recibir comunicaciones usando direcciones de capa 2 (por ejemplo, direcciones de control de acceso al medio (MAC)). Por ejemplo, un dominio de capa 2 puede incluir nodos dentro de la misma red física o red de área local (LAN), por ejemplo, nodos que usan una misma máscara de subred y/o nodos dentro de la misma LAN virtual (VLAN).

La Figura 1 es un diagrama de red de un sistema ejemplar para balanceo de carga sin estado. Con referencia a la Figura 1, la red **100** incluye una pluralidad de dispositivos conectados juntos donde grupos de los dispositivos están en la misma VLAN y subred IP. Por ejemplo, la red **100** puede incluir dispositivos de red o clientes **106**, un dispositivo de reenvío de paquetes de capa 3 y de conmutación de capa 2 o dispositivo L3PF/L2S **102** y servidores **104**. Los clientes **106** y servidores **104** se pueden conectar a uno o más puertos del dispositivo L3PF/L2S **102**.

En la realización mostrada en la Figura 1, los clientes **106** y servidores **104** están agrupados en dos dominios de capa 2. El dispositivo L3PF/L2S **102** está agrupado en ambos dominios, ya que el dispositivo L3PF/L2S **102** es capaz de recibir paquetes desde todos los nodos en la Figura 1. En esta realización, un dominio de capa 2 incluye nodos que están dentro de la misma subred y la misma VLAN. En una segunda realización, un dominio de capa 2 puede incluir nodos que están dentro de la misma LAN física, por ejemplo, una red representada por nodos con una

máscara de subred común o nodos conectados juntos o bien directamente o bien a través de un único salto (por ejemplo, un dispositivo L3PF/L2S **102**). En una tercera realización, cada dominio de capa 2 puede incluir nodos que están dentro de la misma VLAN pero son parte de diferentes LAN físicas.

En la Figura 1, el dominio de capa 2 **108** incluye un grupo de nodos cliente y servidor que tienen direcciones IP dentro de un intervalo de direcciones IP versión 4 (v4) de 192.154.234.01-192.154.234.254. Los nodos del dominio **108** están asociados con la VLAN 1. El dominio de capa 2 **108** es un grupo de cliente y servidor que tiene direcciones dentro de un intervalo de direcciones IP v4 de 192.128.10.01-192.128.10.254. Los nodos de dominio **208** están asociados con la VLAN 2. En la realización mostrada en la Figura 1, cada subred está asociada con una VLAN (o identificador de VLAN) única. En una segunda realización, múltiples subredes se pueden asociar con una VLAN. En una tercera realización, una subred se puede asociar con múltiples VLAN.

Las VLAN permiten a diferentes redes de área local físicas comunicar entre sí usando conmutadores de capa 2, más que encaminadores de capa 3. Los conmutadores de capa 2 se pueden configurar para controlar tráfico en base a identificadores VLAN. Por ejemplo, las VLAN se pueden implementar insertando un identificador VLAN en una trama de capa 2 y/o se pueden implementar en base a puertos en el dispositivo L3PF/L2S **102**. Es decir, cuando el tráfico de capa 2 llega al dispositivo L3PF/L2S **102** en un puerto particular y/o tiene un identificador VLAN particular, si no está presente una entrada de base de datos de reenvío de capa 2 en la tabla de reenvío para el tráfico de capa 2, el tráfico de capa 2 se inunda solamente en los puertos del dispositivo L3PF/L2S **102** asociados con el mismo identificador VLAN y/o puerto de entrada. De este modo, las VLAN se usan para controlar la distribución e inundación de tráfico de capa 2. El dispositivo L3PF/L2S **102** proporciona reenvío de paquetes entre el cliente **106** y los servidores **104**. El dispositivo L3PF/L2S **102** puede proporcionar uno o más puertos (a los que se refiere en la presente memoria como puertos de enfrentamiento de red, puertos de red o puertos cliente) para comunicar con los clientes **106** y uno o más puertos (a los que se refiere en la presente memoria como puertos de enfrentamiento de red o puertos de servidor) para comunicar con servidores **104**. El dispositivo L3PF/L2S **102** también proporciona una funcionalidad de balanceo de carga para flujos de tráfico de red. En una realización, el dispositivo L3PF/L2S **102** se puede configurar para usar direccionamiento IP virtual y protocolo de resolución de direcciones (ARP) intermediario en la provisión de SLB. Adicionalmente, el dispositivo L3PF/L2S **102** se puede configurar para usar una funcionalidad de capa 3, tal como hardware o módulos de encaminamiento ECMP, para balancear carga de flujos de tráfico de red dentro de un dominio de capa 2.

En la realización mostrada en la Figura 1, tres clientes **106** del dominio **108** conectan con un primer puerto de red en el dispositivo L3PF/L2S **102**. Como se muestra, los clientes **106** del dominio **108** y el primer puerto de red están asociados con la VLAN 1. Tres clientes **106** del dominio **208** conectan a un segundo puerto de red en el dispositivo L3PF/L2S **102**. Como se muestra, los clientes **106** del dominio **208** y el segundo puerto de red están asociados con la VLAN 2. Los servidores **104** conectan a puertos del servidor en el dispositivo L3PF/L2S **102**. En algunas realizaciones, se pueden conectar múltiples servidores **104** a un puerto único. En la realización mostrada en la Figura 1, cada servidor **104** está conectado a un puerto separado. Los servidores **104** y los puertos de servidor también pueden estar asociados con las VLAN. En una realización, los servidores **104** son miembros de la misma VLAN. En una segunda realización, los servidores **104** son miembros de múltiples VLAN. Como se muestra, los servidores **104** del dominio **108** y sus puertos de servidor relacionados están asociados con la VLAN 1. Los servidores **104** del dominio **208** y sus puertos de servidor relacionados están asociados con la VLAN 2.

Los servidores **104** se pueden organizar en grupos de balanceo de carga **110**. Por ejemplo, un operador de red puede configurar los puertos del servidor para balanceo de carga del tráfico de red asociado con un nodo o grupos de nodos particulares (por ejemplo, nodos dentro de un dominio de capa 2) a través de servidores unidos. En una realización, los servidores **104** pueden estar asociados con la misma VLAN. Para permitir mayor granularidad en la asignación de tareas de balanceo de carga, cada grupo de LB **110** se puede asociar con diferentes direcciones IP virtuales (VIP). Cada dirección VIP asociada con un grupo de LB **110** puede estar compartida por sus miembros (por ejemplo, servidores). Por ejemplo, el dispositivo L3PF/L2S **102** puede configurar dos grupos de LB **110** para balancear carga de los flujos de tráfico de red asociados con la VLAN 1. Un grupo de LB **110** puede estar asociado a una dirección VIP de 192.154.234.2. Un segundo grupo de LB **110** puede estar asociado con una dirección VIP de 192.128.10.2. Cada servidor **104** de cada grupo de LB respectivo **110** puede usar una interfaz de bucle invertido configurada para recibir paquetes dirigidos a una dirección VIP particular (por ejemplo, los servidores **104** en el primer grupo de LB **110** pueden usar 192.154.234.2 y servidores en el segundo grupo de LB **110** pueden usar 192.128.10.2). El dispositivo L3PF/L2S **102** puede usar un mecanismo de selección para balancear carga de los paquetes dirigidos a una dirección VIP. Por ejemplo, el dispositivo L3PF/L2S **102** puede determinar un grupo de LB **110** en base a una dirección VIP dentro de un paquete y puede determinar, usando un esquema de troceado de paquetes, qué servidor **104** del grupo de LB **110** va a recibir el paquete.

En una realización donde los servidores **104** son miembros de múltiples VLAN, el dispositivo L3PF/L2S **102** puede realizar compartición de carga de una forma por VLAN. La compartición de carga de una forma por VLAN puede incluir compartir la carga de los paquetes dirigidos a una dirección VIP entre los servidores que son miembros de una VLAN particular. En una realización, el dispositivo L3PF/L2S **102** puede configurar grupos de LB **110** con diferentes direcciones VIP para diferentes VLAN. Por ejemplo, el dispositivo L3PF/L2S **102** puede configurar dos grupos de LB **110** para balanceo de carga de flujos de tráfico de red. Un grupo de LB **110** puede estar asociado con

una VLAN 1 y una dirección VIP de 192.154.234.2. Un segundo grupo de LB **110** puede estar asociado con una VLAN 2 y una dirección VIP de 192.128.10.2. En este ejemplo, el dispositivo L3PF/L2S **102** puede cargar los paquetes compartidos dirigidos a la dirección VIP 192.154.234.2 desde clientes en la VLAN 1 al primer grupo de LB **110** y compartir la carga de los paquetes dirigidos a una dirección VIP 192.128.10.2 desde clientes en la VLAN 2 al segundo grupo de LB **110**. En otra realización, el dispositivo L3PF/L2S **102** puede configurar grupos de LB **110** con una o más de las mismas direcciones VIP para diferentes VLAN. Por ejemplo, el dispositivo L3PF/L2S **102** puede compartir la carga de los paquetes desde un cliente en la VLAN 1 al primer grupo de LB **110** (u otro(s) grupo(s) de LB que comparte(n) la misma VLAN como clientes) con independencia de qué dirección VIP se usa. En una realización adicional, el dispositivo L3PF/L2S **102** puede informar a los clientes (por ejemplo, usando un mensaje de error u otros mensajes) si los paquetes se dirigen a una dirección VIP asociada con una VLAN diferente.

En una realización, cada servidor **104** puede estar asociado con su propia dirección MAC única y una dirección VIP compartida por servidores **104** y el conmutador **102**. Los servidores **104** también pueden tener sus propias direcciones IP únicas e identificadores VLAN. En una realización, el dispositivo L3PF/L2S **102** puede reenviar los paquetes dirigidos a la dirección IP única del servidor **104** sin realizar balanceo de carga. Por ejemplo, si los paquetes se dirigen a un servidor usando su dirección IP única y dirección MAC única (por ejemplo, una dirección IP que no es una dirección VIP y una dirección MAC que no es la dirección MAC del dispositivo L3PF/L2S), el dispositivo L3PF/L2S **102** puede usar una operación de reenvío de capa 2 para determinar información de reenvío adecuada para reenviar los paquetes al servidor direccionado.

En la realización mostrada en la Figura 1, el Grupo de LB 1 **110** incluye dos servidores **104** conectados al dispositivo L3PF/L2S **102** para procesar tráfico de red del dominio L2 2 **108**. Como se muestra en el Grupo de LB 1 **110**, el servidor 1 **104** tiene una dirección IP v4 única de 192.154.234.5 y una dirección MAC de 10-F3-27-51-22-7A y el servidor 2 **104** tiene una dirección IP v4 única de 192.154.234.12 y una dirección MAC de 02-11-34-4E-4B-47. El grupo de LB 2 **110** incluye dos servidores **104** conectados al dispositivo L3PF/L2S **102** para procesar tráfico de red del dominio de L2 1 **108**. Como se muestra, en el Grupo de LB 2 **110**, el servidor 3 **104** tiene una dirección IP v4 única de 192.128.10.6 y una dirección MAC de 00-53-23-C0-FE-FF y el servidor 4 **104** tiene una dirección IP v4 única de 192.128.10.4 y una dirección MAC de 00-23-13-D0-3B-FF.

Se aprecia que las direcciones IP v4, los identificadores VLAN y las direcciones MAC que se muestran en la Figura 1 representan posibles formas de conexión de información y que se pueden usar otras formas. Por ejemplo, otra o adicional información de conexión, tal como diferentes versiones IP (por ejemplo, IP v6) o información adicional de una o más capas OSI (por ejemplo, número de puerto UDP/TCP, información de capa de aplicaciones), se puede usar en balanceo de carga sin estado como se describe en la presente memoria.

La Figura 2 es un diagrama de una tabla de reenvío ECMP de capa 3 y estructura de datos ECMP ejemplares para realizar balanceo de carga sin estado según una realización de la materia objeto descrita en la presente memoria. Según la RFC 2991 y 2992, que se incorporan por referencia en la presente memoria en sus totalidades, ECMP es una implementación de encaminamiento usada por dispositivos de reenvío de capa 3 o encaminadores para encaminar paquetes usando múltiples rutas o siguientes saltos. De este modo, el encaminamiento ECMP se usa convencionalmente para encaminar paquetes (por ejemplo, a través de redes entre dominios) a un mismo destino (o red de destino) usando múltiples rutas de igual coste.

Según un aspecto de la presente materia objeto, un dispositivo de reenvío de capa 2 es capaz de usar una funcionalidad de capa 3, tal como tablas de reenvío de capa 3 (también referidas en la presente memoria como bases de datos de reenvío (FDB) de capa 3) y hardware de encaminamiento ECMP, para implementar operaciones de conmutación de capa 2. En una realización, un dispositivo de reenvío de capa 2 usa una funcionalidad de encaminamiento ECMP para implementar balanceo de carga dentro de un dominio de capa 2, tal como destinos dentro de una misma subred y una misma VLAN. Para implementar balanceo de carga usando una funcionalidad de encaminamiento ECMP, un dispositivo de reenvío de capa 2 puede incluir una o más estructuras de datos de reenvío de capa 3 con información de grupo de LB e información de reenvío relacionada.

Con referencia a la Figura 2, grupos de balanceo de carga e información de reenvío relacionada (por ejemplo, información del siguiente salto) se pueden almacenar en una o más tablas de búsqueda de encaminamiento. La tabla de rutas **200** representa una estructura de datos (por ejemplo, una tabla de reenvío) para mantener las asociaciones entre direcciones VIP y grupos de LB. En una realización, los grupos de LB se refieren como grupos de ECMP. Por ejemplo, los servidores de un grupo de balanceo de carga pueden ser siguientes saltos de igual coste para un paquete dirigido a una dirección VIP asociada con el grupo de balanceo de carga. La tabla de troceado de siguiente salto (NHHT) de grupo de ECMP **202** representa una estructura de datos (por ejemplo, una tabla de reenvío que usa valores de troceado para realizar búsquedas) para mantener una información de reenvío asociada con los servidores de un grupo de LB o ECMP.

Se apreciará que la tabla de rutas **200** y la NHHT **202** son una de muchas formas posibles para representar balanceo de carga e información de reenvío relacionada. Además se apreciará que son posibles y se contemplan variaciones en las estructuras, formas y uso de datos. Por ejemplo, se pueden usar múltiples tablas de rutas **200**, tal como usar una tabla de ruta **200** para cada dominio de capa 2. Adicionalmente, se pueden usar múltiples NHHT **202** para representar uno o más grupos de LB. Por ejemplo, la NHHT **202** puede incluir grupos de LB de una o más

redes, VLAN, etc. Además, las estructuras de datos (por ejemplo, la tabla de rutas **200** y la NHHT **202**) pueden incluir información adicional, se pueden enlazar o combinar (por ejemplo, información en tabla de rutas **200** y la NHHT **202** puede estar en una estructura de datos única, tal como una tabla o base de datos de reenvío de capa 3) y puede estar en diversas ubicaciones (por ejemplo, en un módulo de I/O en un dispositivo L3PF/L2S o en un módulo de balanceo de carga (LB) separado). La tabla de rutas **200** y la NHHT **202** pueden utilizar los mismos o separados recursos hardware, tales como una primera memoria y una segunda memoria (por ejemplo, una memoria de acceso aleatorio (RAM)).

En la realización mostrada en la Figura 2, la tabla de rutas **200** y la NHHT **202** incluyen datos de balanceo de carga asociados con un dominio de capa 2, tal como una LAN o VLAN. Por ejemplo, la tabla de rutas **200** y la NHHT **202** pueden incluir información acerca de grupos de LB asociados con una VLAN particular. En una tabla de rutas **200**, cada entrada de tabla **204** incluye una dirección VIP y un grupo de LB que está asociado con la dirección VIP. Por ejemplo, la tabla de rutas **200** muestra una entrada **204** que tiene un valor de dirección VIP de "192.154.234.2" y un valor de grupo de ECMP de "LB 1". Los grupos de ECMP pueden estar asociados con una o más direcciones VIP. En una realización, una dirección VIP está asociada con un grupo de ECMP único. En una segunda realización, una dirección VIP puede estar asociada con múltiples grupos de ECMP. Por ejemplo, una dirección VIP puede estar asociada con dos grupos de ECMP. Dependiendo de parámetros adicionales (por ejemplo, dirección IP de origen, tipo de mensaje, tipo de carga útil, número de puerto recibido, etc.), algunos flujos de tráfico se pueden manejar por el primer grupo de ECMP mientras que otros flujos de tráfico se pueden manejar por el segundo grupo de ECMP. En este ejemplo, un dispositivo L3PF/L2S puede usar políticas para determinar qué parámetros adicionales se usan en hacer determinaciones de grupo de ECMP.

Las direcciones VIP y/u otros parámetros se pueden usar en una función de búsqueda de una o más estructuras de datos. En la realización mostrada en la Figura 2, las direcciones VIP se pueden usar como un valor de búsqueda para la tabla de rutas **200**. Usando una dirección VIP como un valor de búsqueda para la tabla de rutas **200**, se puede determinar, seleccionar o proporcionar un grupo de LB y/o información de reenvío relacionada. Por ejemplo, la tabla de rutas **200** puede incluir enlaces, tales como punteros de memoria o referencias, a datos o estructuras de datos adicionales (por ejemplo, un campo de grupo de ECMP en la tabla de rutas **200** puede incluir referencias o punteros a una o más estructuras de datos (por ejemplo, la NHHT **202**) que contienen información de reenvío. En una realización, la tabla de rutas **200** incluye entradas de 32 bit (barra 32) (por ejemplo, direcciones IP o VIP) que apuntan a entradas ECMP (por ejemplo, entrada de NHHT **206**). En un segundo ejemplo, la tabla de rutas **200** puede incluir información de reenvío (por ejemplo, cada entrada de tabla de rutas **200** puede incluir información del siguiente salto como la NHHT **202**). En un tercer ejemplo, la tabla de rutas **200** puede proporcionar información, tal como un valor de clave, para uso con otras estructuras de datos, tales como la NHHT **202**.

En la Figura 2, la NHHT **202** proporciona información de reenvío para reenviar paquetes a servidores del grupo de LB (es decir, el grupo de ECMP) "LB 1". La información de reenvío incluye información del siguiente salto para reenviar paquetes hacia un destino (por ejemplo, un servidor). En una realización, la información de reenvío puede incluir una dirección MAC de destino y un puerto de salida (por ejemplo, un puerto para transmitir paquetes desde el dispositivo L3PF/L2S **102**). También se puede incluir información relacionada adicional. Por ejemplo, también se pueden incluir en la NHHT **202** direcciones VIP, identificadores VLAN, información relacionada con la subred (por ejemplo, máscaras de subred), listas de puertos.

La información de reenvío en la NHHT **202** se puede indexar para búsqueda. Por ejemplo, las entradas **206** de la NHHT **202** puede incluir valores de índice únicos para seleccionar información del siguiente salto. En una realización, un dispositivo L3PF/L2S puede seleccionar una entrada **206** (también referida en la presente memoria como una celda) que contiene información de reenvío (por ejemplo, información del siguiente salto) comparando un valor de troceado calculado con valores de índice asociados con las entradas **206** en la NHHT **202**. Los valores de troceado se pueden producir introduciendo información relacionada con el paquete en una o más funciones de troceado. Como se usa en la presente memoria, una función de troceado se refiere a una función matemática que convierte una cantidad de datos de tamaño potencialmente variable (por ejemplo, uno o más parámetros o valores de campo de un paquete) en un conjunto de datos más pequeño (por ejemplo, enteros) que se pueden usar como índices en una formación u otra estructura de datos (por ejemplo, la NHHT **202**). Una implementación de troceado puede incluir una o más funciones de troceado. En una realización, un valor de troceado se calcula con una implementación de troceado que usa una o más variables asociadas con un paquete. Las variables pueden incluir diversos campos en el paquete, tales como una dirección IP de origen, una dirección IP de destino, información de capa 3, información de capa 2, información de capa 4; información SIP, puerto de origen de capa 4, puerto de destino de capa 4, información de puerto de protocolo de control de transporte (TCP), información de puerto de protocolo de datagrama de usuario (UDP) y uno o más parámetros de campo tunelizados.

Según un aspecto de la presente materia objeto, se pueden usar valores de troceado en la determinación o selección de un servidor para recibir o procesar un flujo de red dado (es decir, paquetes relacionados). Por ejemplo, una función de troceado puede usar unas direcciones de origen y de destino e información de puerto de TCP o UDP de un paquete en el cálculo de un valor de troceado. Comparando los valores de troceado con valores de índice en la NHHT **202**, se selecciona una celda que contiene información de reenvío asociada con un servidor. Se apreciará que una función de troceado basada en paquete generalmente debería seleccionar el mismo servidor para cada

paquete en un flujo de tráfico de red, proporcionando de este modo cada flujo de tráfico de red a un mismo servidor.

En la realización mostrada en la Figura 2, la NHHT **202** incluye entradas **206** que contienen información relacionada con servidores de balanceo de carga, tales como direcciones VIP e información del siguiente salto. Cada entrada **206** tiene un valor de índice, un puerto de salida, un identificador de VLAN de salida, una dirección MAC y una dirección VIP. Como se muestra en la NHHT **202**, cada entrada **206** tiene un valor de índice único. Cada entrada también tiene un mismo valor de dirección VIP de "192.154.234.2" que indica que cada servidor en el grupo de ECMP "LB 1" puede recibir paquetes dirigidos a esta dirección VIP. Cada entrada **206** también tiene un valor de identificador de VLAN de salida de "VLAN 1" que indica que cada siguiente salto (por ejemplo, un servidor de grupo de ECMP) está dentro de la misma VLAN. En una realización, las entradas **206** que tienen la misma dirección MAC también incluirán los mismos valores para la VLAN de salida y el puerto de salida. En una segunda realización, las entradas **206** que tienen una misma dirección MAC pueden tener varios valores para puertos de salida o VLAN de salida. Por ejemplo, un dispositivo L3PF/L2S se puede configurar para reenviar diferentes tipos de tráfico a través de diferentes interfaces de un servidor. En algunas realizaciones, múltiples direcciones MAC pueden usar un mismo puerto de salida o VLAN de salida. Por ejemplo, un servidor puede tener múltiples direcciones MAC. Cada dirección MAC puede ser para una interfaz diferente o para tráfico diferente.

La NHHT **202** puede incluir múltiples celdas o entradas **206**. En una realización, se pueden asociar uno o más servidores con múltiples entradas **206**. Por ejemplo, los servidores se pueden ponderar para manejar una mayor o menor porción de tráfico de red que otros servidores de un grupo de LB. Por tanto, un servidor se puede asociar con más celdas (es decir, entradas **206**) si el servidor va a manejar una mayor porción de tráfico de red o el servidor se puede asociar con menos celdas si el servidor va a manejar una menor porción de tráfico de red. En otras palabras, cuantas más entradas **206** en la NHHT **202** asociadas con un servidor generalmente se reenviará al servidor mayor tráfico de red (por ejemplo, flujos).

La determinación de las ponderaciones asociadas con un servidor puede incluir usar varias métricas o variables. En una realización, la determinación de las ponderaciones se puede basar en cálculo y métricas de red, tales como métricas relacionadas con ancho de banda y recursos de cálculo, tales como velocidad de procesamiento de la CPU y memoria. Las determinaciones de ponderación pueden ser dinámicas o estáticas. Por ejemplo, una ponderación asociada con un servidor se puede determinar en un tiempo inicial en base a la velocidad de procesamiento de la CPU y la memoria total (por ejemplo, RAM) instalada en el servidor. En un segundo ejemplo, las ponderaciones se pueden determinar periódica o dinámicamente como un cambio de disponibilidad de recursos (por ejemplo, las ponderaciones pueden cambiar si un servidor experimenta problemas o si el enlace de red al servidor cae). Dado que las ponderaciones pueden influir en el número de entradas **206** en la NHHT **202**, pueden cambiar las entradas **206** asociadas con un servidor. En una realización, un dispositivo L3PF/L2S puede retardar o esperar a ajustar la ponderación de servidores de manera que se mantenga la persistencia de flujo (es decir, de manera que los paquetes relacionados no se envíen a diferentes servidores debido al cambio en las asociaciones de las entradas **206** y valor de troceado). En una realización, un dispositivo L3PF/L2S puede determinar si y cuándo ajustar las ponderaciones de entradas **206**. Por ejemplo, un dispositivo L3PF/L2S puede ajustar ponderaciones para todos los servidores cuando un nuevo servidor se pone "en línea" (es decir, se configura para balanceo de carga) o cuando un servidor llega a estar indisponible para mitigar el número de flujos interrumpidos.

Como se puede ver a partir de la Figura 2, no se muestran partes de la NHHT **202**. En particular, se usan elipses verticales para representar entradas con un intervalo de valores de índice entre valores de índice de entradas **206** adyacentes. En la NHHT **202**, un intervalo de índices **208** puede incluir elipses verticales que están entre una entrada con un valor de índice de "0" y una entrada con un valor de índice de "100". En este ejemplo, las elipses verticales representan 99 entradas cada una con un valor de índice único inclusive entre 1 y 99.

Cada intervalo **208** se puede asociar con un siguiente salto o destino (por ejemplo, servidor). Por tanto, las entradas **206** de un intervalo de índice dado **208** puede incluir información redundante (es decir, duplicada). En una realización, los intervalos de índices **208** corresponden a ponderaciones de servidor. Por ejemplo, si la NHHT **202** tiene 256 celdas o entradas **206** y cada entrada tiene un valor de índice único (por ejemplo, entre 0 hasta 255). Un intervalo **208** que incluye 101 entradas **206**, cada una que tiene la misma información de siguiente salto y un valor de índice único (por ejemplo, inclusive entre 0 hasta 100), generalmente puede representar un destino (por ejemplo, servidor) para recibir 101/256 o por encima del 39% del tráfico de red dirigido a un grupo de LB o dirección VIP dada.

Aunque los intervalos **208** en la Figura 2 se representan por las entradas **206** que tienen la misma información de reenvío como se muestra, se pueden usar otras formas para representar los intervalos **208** o duplicar el reenvío o información del siguiente salto. En una realización, un dispositivo L3PF/L2S puede almacenar un siguiente salto único o información de reenvío una vez (por ejemplo, en una entrada **206** única en la NHHT **202**). Entradas adicionales que están asociadas con información de reenvío almacenada pueden incluir un puntero de memoria o referencia a esa información, permitiendo por ello a un dispositivo de reenvío usar eficientemente los recursos de memoria y evitar almacenar datos idénticos múltiples veces.

La Figura 3 es un diagrama de flujo de mensajes que ilustra mensajes ejemplares asociados con realizar balanceo de carga sin estado para flujos de tráfico de red. Con referencia a la Figura 3, en la línea 1 del diagrama de flujo de

mensajes, el cliente 1 **106** formula un mensaje de petición ARP para aprender una dirección MAC asociada con una dirección VIP (representada por “VIP” en la Figura 3) y difunde la petición a nodos dentro de su dominio de capa 2 (por ejemplo, a nodos dentro de la misma subred y VLAN). En la línea 2, el dispositivo L3PF/L2S **102** recibe la petición ARP. En una realización, el dispositivo L3PF/L2S **102** está configurado para actuar como un intermediario ARP para una o más direcciones VIP. Por ejemplo, el dispositivo L3PF/L2S **102** o el software dentro del mismo (por ejemplo, un sistema operativo (OS)), puede usar el intermediario de ARP para responder a peticiones ARP para una dirección VIP. En la línea 3, el dispositivo L3PF/L2S **102** responde a la petición ARP con una respuesta ARP que contiene una dirección MAC representada por “X” que es la dirección MAC del dispositivo L3PF/L2S y envía la respuesta ARP al cliente 1 **106**. En la línea 4, el cliente 1 **106** formula un paquete dirigido a la dirección MAC del dispositivo L3PF/L2S **102** en un campo de dirección MAC de destino y envía el paquete al dispositivo L3PF/L2S **102**. Como se muestra en la Figura 3, el paquete incluye unas direcciones MAC de origen y de destino, representadas por “S” y “X” respectivamente. Adicionalmente, el paquete incluye un identificador VLAN representado por “1”. Aunque no se representa en la Figura 3, el paquete también puede incluir información adicional, tal como información de capa 3 y capa 4. Por ejemplo, el paquete puede incluir un campo de dirección IP de origen que tiene la dirección IP del cliente 1 **106** (representada por “A.C”) y un campo de dirección IP de destino que tiene la dirección VIP (representada por “VIP”).

En la línea 5, se recibe el paquete en el dispositivo L3PF/L2S **102**. El dispositivo L3PF/L2S **102** puede determinar cómo manejar el paquete. En una realización, el dispositivo L3PF/L2S **102** o el software dentro del mismo, puede usar desencadenadores o condiciones para anular o abstenerse de balancear la carga de los paquetes. Por ejemplo, el dispositivo L3PF/L2S **102** puede usar listas de control de acceso (ACL). Si la información en los paquetes relacionados coincide con algunos criterios de ACL, tales como número de puerto de TCP o UDP, el dispositivo L3PF/L2S **102** puede reenviar el flujo a un servidor particular **104** o manejar los paquetes de una forma particular (por ejemplo, descartar paquetes sospechosos o no autenticados). En un segundo ejemplo, el dispositivo L3PF/L2S **102** puede abstenerse de balancear la carga de los paquetes si los paquetes no se dirigen a una dirección VIP (o si el paquete no se dirige a la dirección MAC del dispositivo L3PF/L2S **102**). Si no va a ser balanceada la carga del paquete, el dispositivo L3PF/L2S **102** puede reenviar el paquete al puerto correspondiente a la dirección de destino en la tabla de reenvío del dispositivo L3PF/L2S.

Si va a ser balanceada la carga del paquete (por ejemplo, la dirección MAC de destino del paquete es la misma que la dirección MAC del dispositivo L3PF/L2S **102** o una dirección IP de destino del paquete es una dirección VIP), el dispositivo L3PF/L2S **102** puede usar una funcionalidad de capa 3, tal como tablas de encaminamiento ECMP y hardware relacionado, en la realización de la compartición de carga o una porción dentro de las mismas. En una realización, el hardware de encaminamiento ECMP se puede usar en la implementación de un algoritmo de selección para determinar un servidor **104** para recibir el paquete. Para determinar un servidor adecuado **104**, el dispositivo L3PF/L2S **102** puede examinar el paquete para una dirección VIP. Si se encuentra una dirección VIP, la dirección VIP se puede usar para determinar un grupo de servidores para realizar balanceo de carga. Por ejemplo, se puede usar una dirección VIP como un valor de búsqueda en una estructura de datos de capa 3 (por ejemplo, una tabla de rutas **200**, como se ilustra en la Figura 2) que contiene asociaciones entre direcciones VIP y grupos de ECMP (es decir, grupos de LB).

Los valores de troceado basados en información de paquete pueden determinar un servidor **104** e información de reenvío relacionada para reenviar el paquete al servidor **104**. En una realización, se pueden usar funciones de troceado (por ejemplo, funciones o esquemas de troceado ECMP que se describen en la RFC 2991) en el cálculo de un mismo valor de troceado para paquetes relacionados. Por ejemplo, un valor de troceado se puede calcular introduciendo campos particulares de paquetes (por ejemplo, una dirección de origen y de destino). El valor de troceado se puede comparar con valores de índice de una estructura de datos (por ejemplo, una NHHT) con información de reenvío. Si se encuentra una coincidencia de índice (por ejemplo, un valor de troceado coincide con el índice de una celda asociada con el servidor 1 **104**), se puede usar la información de reenvío en la entrada. Como se expuso anteriormente, el dispositivo L3PF/L2S **102** generalmente debería calcular un mismo valor de troceado para paquetes relacionados, proporcionando por ello un flujo de tráfico de red a un mismo servidor **104** sin necesidad de almacenar información de estado para los flujos.

En la línea 6, el dispositivo L3PF/L2S **102** modifica el paquete para incluir información de reenvío asociada con el servidor 1 **104** y reenvía el paquete al servidor 1 **104**. Por ejemplo, el dispositivo L3PF/L2S **102** puede usar información de reenvío del servidor 1 **104** desde una NHHT. El dispositivo L3PF/L2S **102** puede sustituir un campo de dirección MAC de destino del paquete con la dirección MAC única del servidor 1 **104**. El dispositivo L3PF/L2S **102** también puede sustituir un campo de dirección MAC de origen del paquete con la dirección MAC del dispositivo L3PF/L2S **102**. Adicionalmente, un campo de VLAN de salida del paquete se puede sustituir con el identificador de VLAN asociado con el servidor 1 **104**. En la Figura 3, se apreciará que el identificador VLAN se sustituye con un mismo valor dado que el paquete está siendo reenviado dentro del mismo dominio.

Como se muestra en la Figura 3, el paquete incluye una dirección MAC de destino (representada por “Y”) asociada con el servidor 1 **104** y una dirección MAC de origen (representada por “X”) del dispositivo L3PF/L2S **102**. En la línea 7, el servidor 1 **104** recibe el paquete y procesa el paquete en consecuencia. En la línea 8, el servidor 1 **104** formula un paquete en respuesta al paquete recibido y lo transmite hacia el cliente 1 **106** (por ejemplo, el paquete se puede

transmitir hacia el cliente 1 **106** a través del puerto en el que se recibió en el servidor 1 **104**). En una realización, el paquete de respuesta puede incluir un campo de dirección MAC de origen que tiene la dirección MAC única del servidor 1 **104**. El cliente **106** entonces puede dirigir paquetes relacionados adicionales (por ejemplo, paquetes en el mismo flujo de tráfico de red) al servidor usando la dirección MAC del servidor. El dispositivo L3PF/L2S **102** puede realizar reenvío de servidor directo (por ejemplo, reenviando un paquete a un servidor solicitado por el cliente **104** usando una base de datos de reenvío de capa 2 sin usar un algoritmo de balanceo de carga) si un paquete se dirige a una dirección MAC única de un servidor **104**. En una segunda realización, el paquete de respuesta puede incluir la dirección MAC del dispositivo L3PF/L2S **102** en el campo de dirección MAC de origen.

En la realización mostrada en la Figura 3, el paquete de respuesta incluye unas direcciones de origen y de destino, representadas por "Y" y "S" respectivamente. Adicionalmente, el paquete de respuesta incluye un identificador VLAN representado por un "1". Aunque no se representa en la Figura 3, el paquete también puede incluir información adicional, tal como información de capa 3 y capa 4. Por ejemplo, el paquete puede incluir un campo de dirección IP de origen con la dirección VIP (representada por "VIP") y un campo de dirección IP de destino con la dirección del cliente 1 **106** (representado por "A.C").

La línea 8, el cliente 1 **106** formula un mensaje de petición ARP para aprender una dirección MAC asociada con una dirección IP del servidor 2 **104** (representado por "A.B") y difunde la petición a los nodos dentro de su dominio de capa 2 (por ejemplo, su LAN física o VLAN). El dispositivo L3PF/L2S **102** un recibe la petición ARP y reenvía el paquete hacia el servidor 2 **104**. Aunque el dispositivo L3PF/L2S **102** tiene una funcionalidad para realizar un ARP intermediario para direcciones de capa 3, tales como direcciones VIP, el dispositivo L3PF/L2S **102** puede reenviar peticiones ARP de direcciones de capa 3 que no están configuradas para intermediario ARP sin enviar una respuesta. En otras palabras, las peticiones ARP a todas las direcciones no VIP, tales como la dirección IP del servidor 2 **104**, se pueden pasar a través del dispositivo L3PF/L2S en la capa 2 hacia su destino. Por ejemplo, el dispositivo L3PF/L2S **102** puede reenviar una petición ARP hacia un destino usando una tabla de reenvío de capa 2 o puede difundir la petición ARP a un dominio de capa 2 si no se ha aprendido el destino.

La Figura 4 es un diagrama de bloques de un dispositivo de reenvío ejemplar adecuado para balanceo de carga sin estado según una realización de la materia objeto descrita en la presente memoria. Se apreciará que aunque el dispositivo de reenvío de paquetes de capa 3 y de conmutación de capa 2 mostrado en la Figura 4 representa una implementación posible de un dispositivo de reenvío adecuado para balanceo de carga sin estado, la materia objeto descrita en la presente memoria se puede implementar usando otros dispositivos de reenvío. Un ejemplo de un dispositivo de reenvío de paquetes disponible comercialmente adecuado para usar con realizaciones de la materia objeto descrita la presente memoria es el conmutador de la serie Summit[®] X450 disponible en Extreme Networks de Santa Clara, California.

Con referencia la Figura 4, el dispositivo L3PF/L2S **102** puede realizar balanceo de carga sin estado usando módulos implementados en hardware, microprograma y/o software. Por ejemplo, el dispositivo L3PF/L2S **102** puede incluir una pluralidad de módulos de entrada/salida (I/O) **400** para enviar y recibir paquetes. Los módulos de I/O **400** cada uno pueden incluir uno o más puertos de I/O **402** para conectar a una red externa. En una realización, cada módulo de I/O **400** puede incluir un motor de búsqueda **404**, una estructura de datos de listas de puertos (y/o un siguiente salto) **406** y una o más bases de datos de reenvío de capa 3 **408**. La estructura de datos de listas de puertos **406** incluye identificadores de puertos para reenviar paquetes hacia un destino. En una realización, la información de siguiente salto (por ejemplo, como se muestra en la NHHT **202** en la Figura 2) asociada con servidores de balanceo de carga se puede incluir con información de lista de puertos. La base de datos de reenvío de capa 3 **408** puede incluir direcciones VIP y datos de grupo de ECMP asociados. En una realización, la base de datos de reenvío de capa 3 **408** también puede incluir información del siguiente salto (por ejemplo, como se muestra en la NHHT **202** en la Figura 2). El motor de búsqueda **404** puede realizar búsquedas en la base de datos **408** en base a las direcciones VIP para determinar una coincidencia de grupo de ECMP en la estructura de datos del siguiente salto **406**.

El módulo de balanceo de carga (LB) **410** puede proporcionar una funcionalidad para balancear la carga de los paquetes entre un grupo de servidores. En una realización, el módulo de LB **410** puede incluir un módulo de encaminamiento ECMP para uso en balanceo de carga sin estado. El módulo de LB **410** puede determinar un grupo de LB (por ejemplo, un grupo de ECMP) para un paquete dado. Adicionalmente, el módulo de LB **410** puede determinar qué servidor en el grupo de LB para reenviar el paquete y la funcionalidad para reenviar el paquete usando la información del siguiente salto asociada. En particular el módulo **410** puede incluir una implementación de troceado para calcular valores de troceado usados en la selección de servidores para balanceo de carga de flujos de tráfico de red. El módulo **410** también puede incluir una funcionalidad para modificar los paquetes con información del siguiente salto y reenviar los paquetes hacia los servidores adecuados (por ejemplo, a través de puertos de salida que se describen en la estructura de datos **406**). En una implementación alternativa, el módulo de LB **410** puede incluir uno o más módulos o componentes de capa 3, por ejemplo, el módulo **410** puede incluir componentes o módulos, tales como tablas de búsqueda y estructuras de datos, que se representa separadamente en la Figura 4.

Aunque se representa en la Figura 4 un único motor de búsqueda **404** para realizar búsquedas de reenvío de capa 3, la materia objeto descrita en la presente memoria no está limitada a tal implementación. Por ejemplo, se pueden incluir tablas adicionales, motores de búsqueda y otros componentes. En particular, puede estar presente una base

de datos de reenvío de capa 2 o módulos relacionados para realizar operaciones de conmutación de capa 2 normales (es decir, sin balanceo de carga).

La matriz de conmutación **412** conmuta los paquetes entre los módulos de I/O **400** y el módulo de gestión de conmutador **414**. El modo de gestión de conmutador **414** puede incluir un motor de aprendizaje software de capa 3 **416**, una base de datos de reenvío de capa 3 maestra **418**, un motor de búsqueda software **420** y una función de ARP intermediaria **422**. El módulo de gestión de conmutador **414** y sus componentes dentro del mismo (por ejemplo, el motor de aprendizaje software **416**, una base de datos de reenvío de capa 3 maestra **418**, un motor de búsqueda software **420** y una función de ARP intermediaria **422**), se pueden almacenar en memoria **424** y ejecutar por la CPU **426**. El modo de gestión de conmutador **414** puede incluir software (por ejemplo, un sistema operativo) para realizar comprobaciones de salud de los servidores y componentes y otra funcionalidad para permitir balanceo de carga sin estado en la capa 2 a través de módulos de capa 3 (por ejemplo, un módulo de LB).

El motor de aprendizaje **416** puede incluir software de protocolo de encaminamiento para intercambiar información de protocolo de encaminamiento con otros nodos. Por lo tanto, el motor de aprendizaje **416** puede añadir entradas aprendidas a la FDB software maestra **418**. El motor de aprendizaje **416** u otro módulo, puede usar uno o más protocolos (por ejemplo, protocolo de mensajes de control de Internet (ICMP), ARP, etc.) o puede intentar conectar a aplicaciones o servicios en servidores para propósitos de comprobación de salud. Por ejemplo, usando un protocolo de control de trayecto tal como ICMP (por ejemplo, para enviar un comando ping), el dispositivo L3PF/L2S **102** puede determinar si los servidores están disponibles o indisponibles para propósitos de balanceo de carga. En base a esta determinación, el dispositivo L3PF/L2S **102** puede poner los servidores “en línea” o “fuera de línea”.

La base de datos de reenvío de capa 3 maestra **418** puede incluir una copia de todas las entradas en las bases de datos de reenvío de capa 3 basadas en hardware **408** mantenidas por los módulos de I/O **400** así como cualquier entrada adicional aprendida usando protocolos de encaminamiento de capa 3 implementados por software **416**. El motor de búsqueda software **420** puede realizar búsquedas en la base de datos de reenvío de capa 3 maestra **418** para paquetes recibidos por los módulos de I/O **400** que no se pueden reenviar usando los datos de reenvío de capa 3 mantenidos localmente por cada módulo de I/O (procesamiento de “camino lento”). Como se mencionó anteriormente, es deseable limitar la cantidad de procesamiento de “camino lento” para paquetes realizado por el motor de búsqueda **420** conservando recursos hardware de listas de puertos **406**.

La función ARP intermediaria **422** puede proporcionar funcionalidad para responder a peticiones ARP desde clientes, las peticiones ARP que incluyen una dirección IP virtual (VIP) compartida por el dispositivo y una pluralidad de servidores acoplados al dispositivo, con la dirección de control de acceso al medio (MAC) del dispositivo. Por ejemplo, la función ARP intermediaria **422** puede incluir una funcionalidad para examinar cada petición ARP recibida para una dirección VIP y responder con su propia dirección MAC (es decir, la dirección MAC del dispositivo L3PF/L2S **102**) si las peticiones ARP incluyen una dirección VIP.

La Figura 5 es un diagrama de flujo de un proceso ejemplar para realizar balanceo de carga sin estado de flujos de tráfico de red según una realización de la materia objeto descrita en la presente memoria. El proceso ejemplar puede ocurrir en un dispositivo de reenvío de paquetes de capa 3 y de conmutación de capa 2 (por ejemplo, un conmutador de capa 2 que incluye un módulo de LB). Con referencia la Figura 5, en el paso **500**, el dispositivo de reenvío de paquetes de capa 3 y de conmutación de capa 2 responde a peticiones de protocolo de resolución de direcciones (ARP) desde clientes, las peticiones ARP que incluyen una dirección IP virtual (VIP) compartida por el dispositivo y una pluralidad de servidores acoplados al dispositivo, con la dirección de control de acceso al medio (MAC) del dispositivo. Por ejemplo, los nodos cliente pueden difundir una petición ARP para una dirección VIP. El dispositivo de reenvío de paquetes de capa 3 y de conmutación de capa 2 se puede configurar para reconocimiento de peticiones ARP para un VIP con respuestas que contienen su dirección MAC.

En el paso **502**, el dispositivo de reenvío de paquetes de capa 3 y de conmutación de capa 2 recibe, desde los clientes, los paquetes dirigidos a la dirección VIP y que tienen la dirección MAC del dispositivo. Por ejemplo, los nodos cliente dentro de la misma subred como una dirección VIP y asociados con la misma VLAN pueden formular paquetes dirigidos a la dirección VIP que incluye la dirección MAC del dispositivo de reenvío como la dirección MAC de destino del paquete. Los nodos cliente pueden haber recibido la dirección MAC del dispositivo L3PF/L2S en respuesta a enviar una petición ARP para la dirección VIP.

En el paso **504**, el dispositivo de reenvío de paquetes de capa 3 y de comunicación de comunicación de capa 2 comparte la carga de los paquetes entre los servidores usando una operación de reenvío de capa 3 que parece a los clientes como una operación de conmutación de capa 2. En una realización, una operación de reenvío de capa 3 puede parecer a los clientes o la red como una operación de conmutación de capa 2 debido a que los paquetes, desde los nodos cliente, que van a ser compartidos de carga están siendo enviados a nodos de servidor que están o parecen estar en el mismo dominio de capa 2 que los nodos cliente. Por ejemplo, los paquetes que se dirigen a y desde los nodos dentro de una misma LAN física y/o virtual (por ejemplo, nodos de origen y de destino dentro de la misma subred y la misma VLAN) pueden parecer que están conmutados en capa 2 a nivel de red o cliente.

Se apreciará que aunque la entrega de paquetes dentro de un dominio de capa 2 parece como una operación de conmutador de capa 2 a los clientes, el dispositivo de reenvío de paquetes de capa 3 y de conmutación de capa 2 se

5 puede configurar para usar una funcionalidad de reenvío de capa 3 para propósitos de balanceo de carga. Por ejemplo, el dispositivo de reenvío de paquetes de capa 3 y de conmutación de capa 2 puede usar un módulo de LB para transmitir paquetes a servidores de balanceo de carga asociados con una dirección VIP. En una realización, realizar una operación de conmutación de capa 2 en el paquete incluye realizar, usando la dirección VIP, una búsqueda en una o más bases de datos de reenvío de capa 3 para determinar un grupo de balanceo de carga para seleccionar información del siguiente salto. Después de realizar un troceado en el paquete (por ejemplo, introduciendo uno o más campos de paquete en una función de troceado y sacando un valor de troceado), el valor de troceado se usa en la determinación o selección de un servidor del grupo de balanceo de carga para recibir el paquete. El paquete se puede modificar con información del siguiente salto asociada con el servidor seleccionado y el paquete se puede reenviar hacia el destino para procesamiento. Se apreciará que los paquetes relacionados (por ejemplo, paquetes en el mismo flujo de tráfico de red) deberían generar el mismo valor de troceado y, por tanto, se deberían reenviar al mismo servidor. De este modo, el reenvío de paquetes del mismo flujo de tráfico de red al mismo servidor se puede realizar sin almacenar información de estado para los flujos y el dispositivo de reenvío de paquetes de capa 3 y de conmutación de capa 2 puede usar una funcionalidad de reenvío de capa 3 (por ejemplo, un módulo de encaminamiento ECMP) para balanceo de carga sin estado de los flujos de tráfico de red.

10 Se entenderá que diversos detalles de la materia objeto descrita en el momento presente se pueden cambiar sin apartarse del alcance de la materia objeto descrita en el momento presente. Adicionalmente, la descripción precedente es para el propósito de ilustración solamente y no para el propósito de limitación.

REIVINDICACIONES

1. Un método para realizar balanceo de carga sin estado de flujos de tráfico de red, el método que comprende:
 - en un dispositivo de reenvío de paquetes de capa 3 y de conmutación de capa 2 (102):
 - 5 responder a peticiones de protocolo de resolución de direcciones (ARP) desde clientes (106), las peticiones ARP que incluyen una dirección IP virtual (VIN) compartida por el dispositivo (102) y una pluralidad de servidores (104) acoplados al dispositivo (102), con la dirección de control de acceso al medio (MAC) del dispositivo;
 - recibir, desde los clientes (106), paquetes dirigidos a la dirección VIP y que tienen la dirección MAC del dispositivo; y
 - 10 compartir la carga de los paquetes entre los servidores (104) usando una operación de reenvío de capa 3 que parece a los clientes como una operación de conmutación de capa 2.
2. El método de la reivindicación 1 en donde la compartición de carga de los paquetes usando una operación de reenvío de capa 3 comprende:
 - 15 determinar un grupo de multitrayecto de igual coste (ECMP) que corresponde a la dirección VIP y compartir la carga de los paquetes entre los servidores en el grupo de ECMP.
3. El método de la reivindicación 1 en donde los servidores (104) son miembros de diferentes redes de área local virtual (VLAN) y en donde el dispositivo de reenvío de paquetes de capa 3 y de conmutación de capa 2 (102) está configurado para realizar la compartición de carga de una forma por VLAN usando diferentes direcciones VIP para diferentes VLAN.
- 20 4. El método de la reivindicación 1 que comprende reenviar paquetes del mismo flujo de tráfico de red al mismo servidor sin almacenar información de estado para los flujos.
5. El método de la reivindicación 1 en donde el dispositivo de reenvío de paquetes de capa 3 y de conmutación de capa 2 (102) está configurado para comprobar activamente el estado de los servidores (104) y actualizar la información de compartición de carga en base al estado.
- 25 6. El método de la reivindicación 2 en donde la compartición de carga de los paquetes usando una operación de reenvío de capa 3 comprende acceder a una tabla de encaminamiento de multitrayecto de igual coste (ECMP) (200) usando parámetros en los paquetes para realizar la compartición de carga.
7. El método de la reivindicación 1 en donde el dispositivo de reenvío de paquetes IP usa listas de control de acceso (ACL) en la determinación de si balancear la carga de los flujos de tráfico de red.
- 30 8. Un dispositivo de reenvío de paquetes de capa 3 y de conmutación de capa 2 (102) para realizar balanceo de carga sin estado, el dispositivo que comprende:
 - 35 una función de protocolo de resolución de direcciones (ARP) intermediaria (422) para responder a peticiones ARP desde clientes (106), las peticiones ARP que incluyen una dirección IP virtual (VIP) compartida por el dispositivo (102) y una pluralidad de servidores (104) acoplados al dispositivo (102), con la dirección de control de acceso al medio (MAC) del dispositivo (102); y
 - un módulo de balanceo de carga (410) para recibir, desde los clientes (106), paquetes dirigidos a la dirección VIP y que tienen la dirección MAC del dispositivo y para compartir la carga de los paquetes entre los servidores (104) usando una operación de reenvío de capa 3 que parece a los clientes (106) como una operación de conmutación de capa 2.
- 40 9. El dispositivo (102) según la reivindicación 8, en donde el módulo de balanceo de carga (410) está configurado para encaminar los paquetes a un mismo destino usando múltiples rutas de igual coste.
10. El dispositivo (102) de la reivindicación 8 en donde los servidores (104) son miembros de diferentes redes de área local virtuales (VLAN) y en donde el módulo de balanceo de carga (410) está configurado para realizar la compartición de carga de una forma por VLAN usando diferentes direcciones VIP para diferentes VLAN.
- 45 11. El dispositivo (102) de la reivindicación 8 en donde el módulo de balanceo de carga (410) está configurado para reenviar los paquetes del mismo flujo de tráfico de red al mismo servidor sin almacenar información de estado para los flujos.
12. El dispositivo (102) de la reivindicación 8 en donde el dispositivo (102) está configurado para comprobar activamente el estado de los servidores (104) y actualizar la información de compartición de carga en base al estado.
- 50

13. El dispositivo (102) de la reivindicación 8 en donde el dispositivo está configurado para reenviar los paquetes dirigidos a un servidor individual sin realizar compartición de carga.

5 14. El dispositivo (102) de la reivindicación 9 en donde el módulo de balanceo de carga (410) está configurado para acceder a una tabla de encaminamiento multitrayecto de igual coste (ECMP) usando parámetros en los paquetes para realizar la compartición de carga.

15. El dispositivo (102) de la reivindicación 8 que comprende:

listas de control de accesos (ACL) para uso en la determinación de si balancear la carga de los flujos de tráfico de red.

10 16. Un medio legible por ordenador no transitorio que contiene un programa de ordenador que incluye instrucciones ejecutables por ordenador que cuando se ejecutan por el procesador de un ordenador realizan los pasos que comprenden:

responder a una petición de protocolo de resolución de direcciones (ARP) desde clientes (106), la petición ARP que incluye una dirección IP virtual (VIP) compartida por el dispositivo (102) y una pluralidad de servidores (104) acoplados al dispositivo (102), con la dirección de control de acceso al medio (MAC) del dispositivo;

15 recibir, desde los clientes (106), paquetes dirigidos a la dirección VIP y que tienen la dirección MAC del dispositivo; y

compartir la carga de los paquetes entre los servidores usando una operación de reenvío de capa 3 que parece a los clientes (106) como una operación de conmutación de capa 2.

20

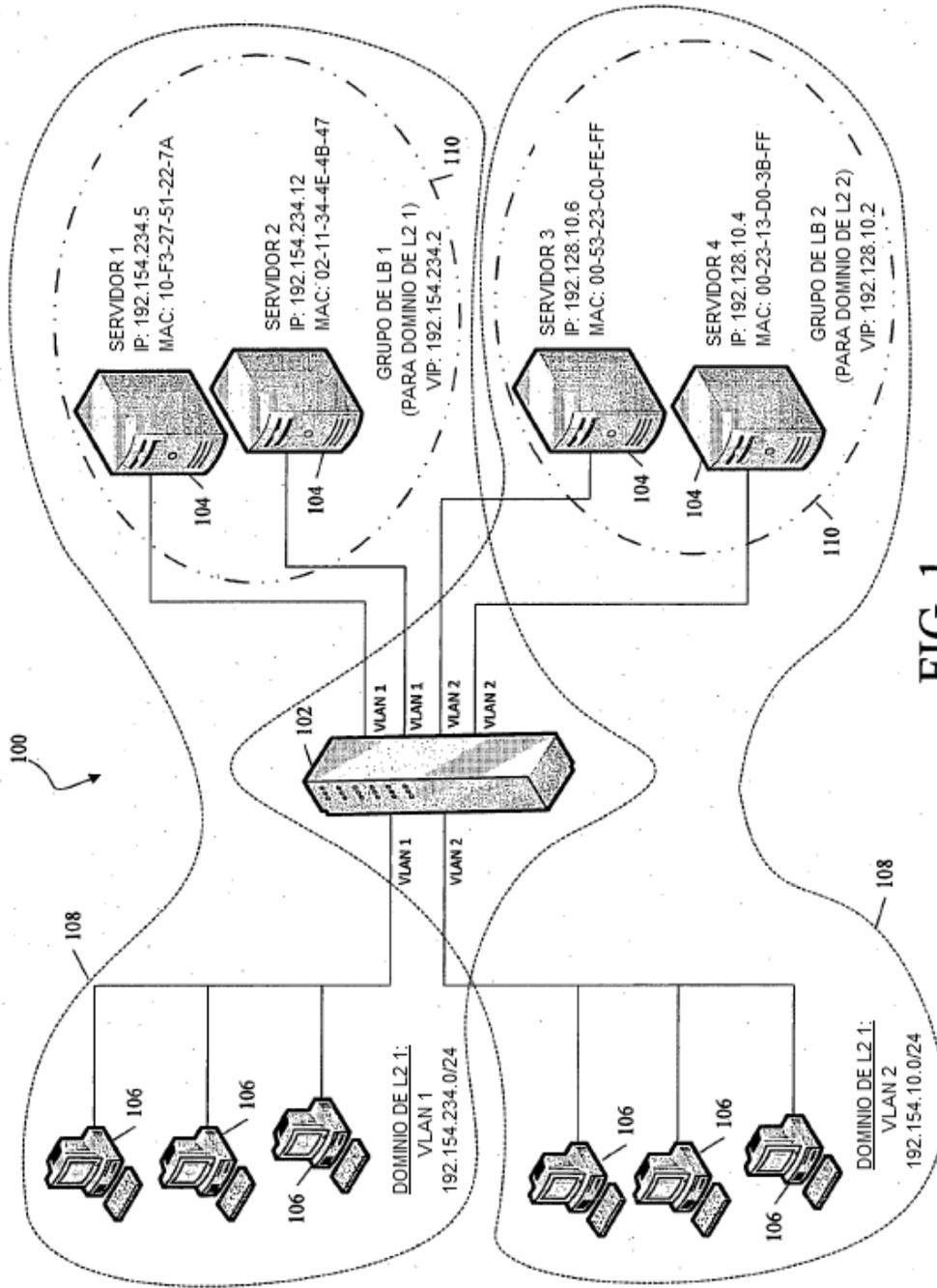


FIG. 1

| TABLA DE RUTAS | |
|--------------------|---------------|
| DIRECCIÓN VIP | GRUPO DE ECMP |
| 204 192.154.234.2 | LB 1 |
| 204 192.128.10.2 | LB 2 |
| 204 192.154.123.2 | LB 3 |
| 204 192.154.123.12 | LB 4 |
| 204 192.214.42.3 | LB 5 |
| 204 192.214.345.6 | LB 6 |
| 204 153.34.32.12 | LB 7 |
| 204 42.33.145.6 | LB 8 |

200

| TABLA DE TROCEADO DE GRUPO DE LB 1 | | | | |
|------------------------------------|--------|------|-------------------|---------------|
| ÍNDICE | PUERTO | VLAN | MAC | DIRECCIÓN VIP |
| 206 0 | 42 | 1 | 02-11-34-4E-4B-47 | 192.154.234.2 |
| 208 ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 206 100 | 42 | 1 | 02-11-34-4E-4B-47 | 192.154.234.2 |
| 206 101 | 26 | 1 | 10-F3-27-51-22-7A | 192.154.234.2 |
| 208 ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 206 190 | 26 | 1 | 10-F3-27-51-22-7A | 192.154.234.2 |
| 206 191 | 44 | 1 | 02-11-34-4E-4B-47 | 192.154.234.2 |
| 208 ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 206 214 | 44 | 1 | 02-11-34-4E-4B-47 | 192.154.234.2 |
| 206 215 | 28 | 1 | 10-F3-27-51-22-7A | 192.154.234.2 |
| 208 ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 206 255 | 28 | 1 | 10-F3-27-51-22-7A | 192.154.234.2 |

202

FIG. 2

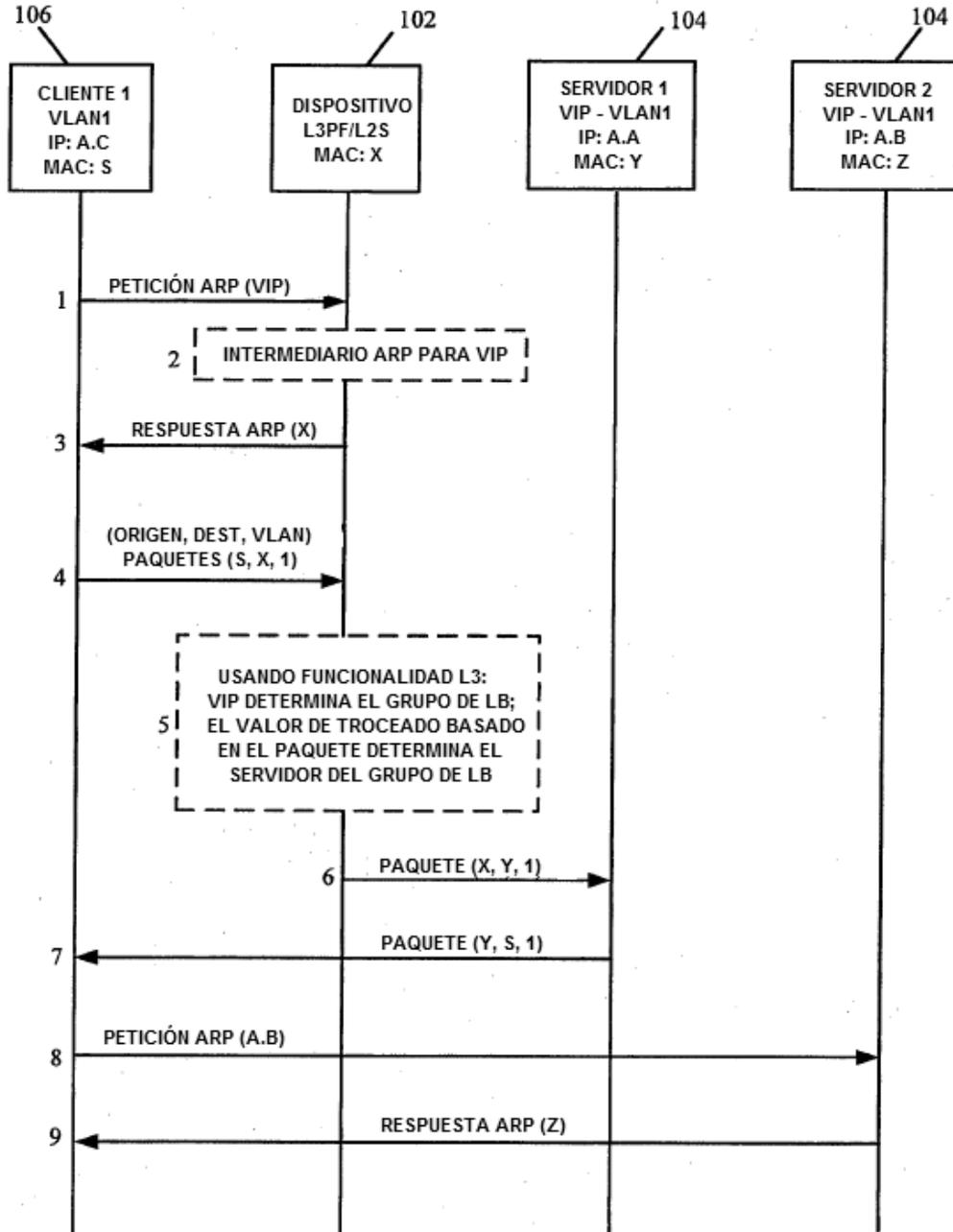


FIG. 3

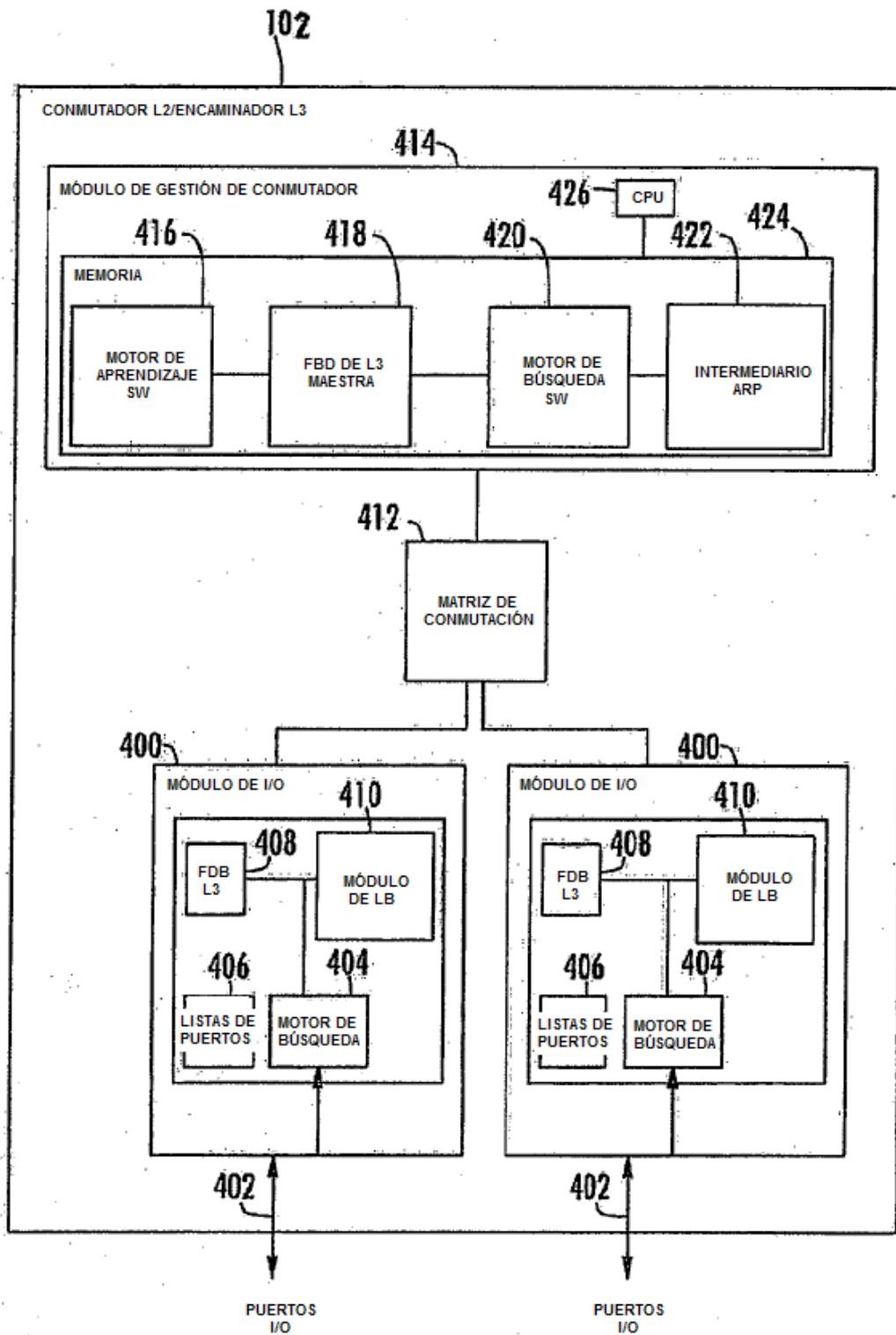


FIG. 4

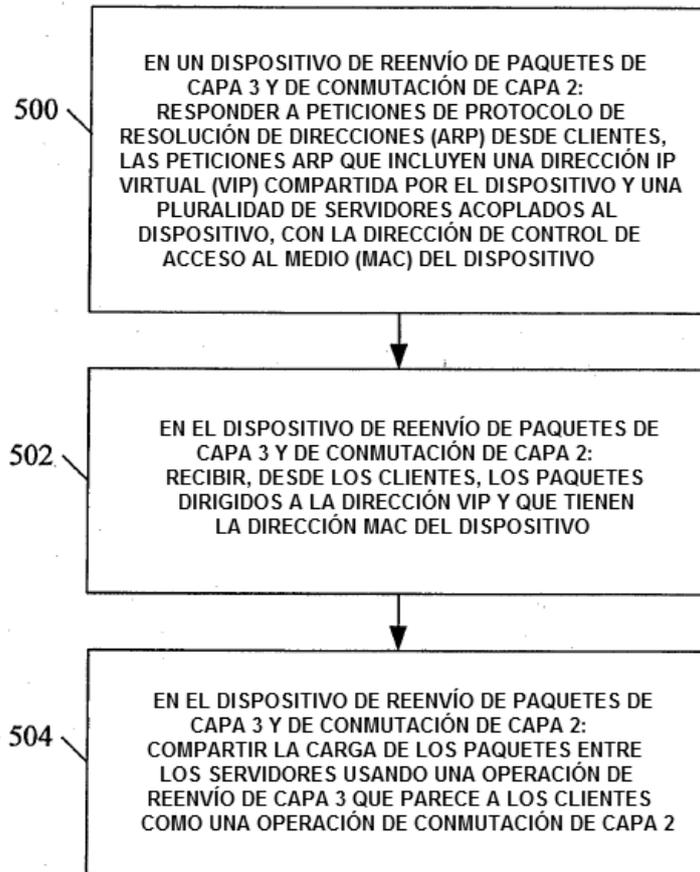


FIG. 5