(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 773 032**

(51) Int. CI.:

H04L 12/935 (2013.01)
G06F 13/38 (2006.01)
G06F 13/40 (2006.01)
G06F 13/42 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **13.01.2017 PCT/CN2017/071081**

(87) Fecha y número de publicación internacional: **20.07.2017 WO17121376**

(96) Fecha de presentación y número de la solicitud europea: **13.01.2017 E 17738189 (4)**

(97) Fecha y número de publicación de la concesión europea: **13.11.2019 EP 3276899**

(54) Título: **Dispositivo de conmutación, sistema de interconexión de componentes periféricos rápida y procedimiento de inicialización del mismo**

(30) Prioridad:

13.01.2016 CN 201610022697

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
09.07.2020

(73) Titular/es:

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

(72) Inventor/es:

FANG, HONGCAN

(74) Agente/Representante:

ELZABURU, S.L.P

DESCRIPCIÓN

Dispositivo de conmutación, sistema de interconexión de componentes periféricos rápida y procedimiento de inicialización del mismo

Campo técnico

Esta aplicación se refiere al campo de las comunicaciones y, más específicamente, a un dispositivo de conmutación, un sistema de interconexión de componentes periféricos rápida y un procedimiento para inicializar un sistema de interconexión de componentes periféricos rápida.

Antecedentes

Una tecnología de virtualización de entrada/salida de raíz única (en inglés, Single-Root Input/Output Virtualization – SR-IOV) propuesta por el Grupo de Interés Especial de Interconexión de Componentes Periféricos (en inglés, Peripheral Component Interconnect Special Interest Group - PCI-SIG) se aplica a una situación en la que un dispositivo de entrada/salida (Entrada/Salida - E/S) es compartido por múltiples procesos en un host. La tecnología SR-IOV virtualiza, para un dispositivo PCI rápido (en inglés, PCI express - PCIe), múltiples funciones virtuales (en inglés, Virtual Function - VF) para un host de capa superior a invocar.

Una tecnología de entrada/salida de raíz múltiple (en inglés, Multi-Root Input/Output Virtualization - MR-IOV) se aplica a una situación en la que un dispositivo de E/S es compartido por múltiples hosts. La tecnología virtualiza múltiples jerarquías virtuales (en inglés, Virtual Hierarchies - VH) a ser invocadas por múltiples hosts, y a las múltiples jerarquías virtuales se puede hacer referencia como múltiples dominios PCI. Sin embargo, la tecnología modifica protocolos de capas, salvo una capa física y una capa de enlace de datos en un protocolo PCIe. Por lo tanto, múltiples nodos, por ejemplo, un nodo raíz (en inglés, Root Node - RN) que soporta un MR-IOV, un conmutador (Conmutador) o un dispositivo PCIe, en una estructura de topología, deben ejecutar una adaptación. La tecnología MR-IOV requiere una adaptación de la cadena industrial, lo que genera una compatibilidad pobre. Actualmente, no hay ningún producto que soporte el estándar.

En la técnica anterior, se proporciona una solución para que múltiples hosts compartan un dispositivo de E/S mediante el uso de un dispositivo SR-IOV. La FIG. 1 muestra una implementación típica en la técnica anterior. Un sistema 100 incluye: N hosts (desde un host 1 a un host N) 110, un conmutador PCIe 120, M dispositivos de E/S (desde un dispositivo de E/S 1 a un dispositivo de E/S M) 130, una CPU de administración externa 140 y una memoria 150, donde $1 \leq M \leq N$. La CPU de administración 140 es responsable de: ejecutar la enumeración de dispositivos y descubrir un lado de dispositivo de E/S y ejecutar la configuración y administración de dispositivos. El conmutador PCIe 120 incluye un módulo de enlace no transparente (en inglés, Non-Transparent Link – NT-L), múltiples puertos PCIe aguas arriba (en inglés, Upstream Port - UP), un módulo EP no transparente global (en inglés, Non-Transparent - NT) y múltiples puertos PCIe aguas abajo virtuales (en inglés, Downstream Port - DP). El módulo EP NT global se configura para: administrar un registro y mapear una dirección de memoria, e implementar una función virtual no transparente (en inglés, Non-Transparent Virtual - NT-V). El sistema 100 puede incluir múltiples jerarquías virtuales (en inglés, Virtual Hierarchy - VH).

En la solución técnica, debe disponerse adicionalmente una CPU de administración, debe disponerse externamente una memoria que corresponda a la CPU de administración y debe proporcionarse una interfaz de administración PCIe separada. Por tanto, la solución técnica no es adecuada para un diseño de producto. Además, en el sistema 100 que se muestra en la FIG. 1, solo una CPU de administración se dispone para administrar todo el sistema y hay solo un enlace de administración. Una vez que el enlace de administración presenta una excepción, toda una red se rompe y ya no puede usarse. En consecuencia, la estabilidad y la confiabilidad son relativamente pobres.

El documento de los EE.UU. 2012/179804 se refiere a la administración de un conmutador de interconexión de componentes periféricos PCI por sus siglas en inglés. En particular, describe que el servidor de administración incluye un módulo de administración de dispositivo PCI 200 que administra los cambios en la configuración de los conmutadores PCIe.

El documento EP 2549716 se refiere a un dispositivo de administración de red y un procedimiento para administrar una red. En particular, describe un dispositivo de administración de red que es diferente del conmutador y adquiere información de identificación del nodo que se conecta al conmutador.

Compendio

Las realizaciones de la presente invención proporcionan un dispositivo de conmutación para inicializar un sistema PCIe, a fin de lograr una estabilidad y una confiabilidad relativamente altas. Según un primer aspecto, se proporciona un dispositivo de conmutación, incluyendo: múltiples puertos PCIe aguas arriba, configurados para conectarse al, al menos un, host; al menos un puerto PCIe aguas abajo, configurado para conectarse al, al menos un, dispositivo de E/S; y un aparato de procesamiento interno, conectado al, al menos un, puerto PCIe aguas abajo mediante el uso de una línea de conexión interna del dispositivo de conmutación, donde el aparato de procesamiento interno se configura para:

transmitir un paquete de lectura/escritura de configuración para el al menos un puerto PCIe aguas abajo mediante el uso de una línea de conexión interna;

recibir un paquete de respuesta de lectura/escritura de configuración por parte del al menos un puerto PCIe aguas abajo mediante el uso de la línea de conexión interna, donde el paquete de respuesta de lectura/escritura de configuración transporta una identificación de completador; y

determinar, en función de la identificación de completador transportada en el paquete de respuesta de lectura/escritura de configuración, que el dispositivo de conmutación está conectado a un dispositivo de E/S cuya identificación es la identificación de completador.

El aparato de procesamiento interno se puede configurar para enumerar un dispositivo físico (es decir, un dispositivo real) conectado al dispositivo de conmutación. Opcionalmente, el aparato de procesamiento interno puede enviar múltiples paquetes de lectura/escritura de configuración al, al menos un, puerto PCIe aguas abajo, donde los múltiples paquetes de lectura/escritura de configuración transportan diferentes identificaciones de completador. El al menos un puerto PCIe aguas abajo se configura para: reenviar los múltiples paquetes de lectura/escritura de configuración a un dispositivo físico conectado al, al menos un, puerto PCIe aguas abajo, y al recibir al menos un paquete de respuesta de lectura/escritura de configuración enviado por el dispositivo físico conectado, reenviar el al menos un paquete de respuesta de lectura/escritura de configuración al aparato de procesamiento interno, donde el al menos un paquete de respuesta de lectura/escritura de configuración corresponde a al menos uno de los múltiples paquetes de lectura/escritura de configuración en una correspondencia de uno a uno, y cada paquete de respuesta de lectura/escritura de configuración transporta una identificación de completador en un paquete de lectura/escritura de configuración correspondiente al paquete de respuesta de lectura/escritura de configuración. El aparato de procesamiento interno puede recibir el al menos un paquete de respuesta de lectura/escritura de configuración reenviado por el al menos un puerto PCIe aguas abajo y determinar, en función de la identificación de completador transportada en cada uno de los al menos un paquete de respuesta de lectura/escritura de configuración, el dispositivo físico conectado al, al menos un, puerto PCIe aguas abajo (o el dispositivo de conmutación). El dispositivo físico puede incluir un dispositivo de E/S, o a demás puede incluir un conmutador de PCIe y/u otro aparato.

Por lo tanto, el dispositivo de conmutación según esta realización de la presente invención incluye los múltiples puertos PCIe aguas arriba configurados para conectarse al, al menos un, host; el al menos un puerto PCIe aguas abajo configurado para conectarse al, al menos un, dispositivo de E/S; y el aparato de procesamiento interno. El aparato de procesamiento se conecta al, al menos un, puerto PCIe aguas abajo mediante el uso de una línea de conexión interna y se configura para: transmitir un paquete de lectura/escritura de configuración al, al menos un, puerto PCIe aguas abajo mediante el uso de la línea de conexión interna; recibir un paquete de respuesta de lectura/escritura de configuración transmitido mediante el al menos un puerto PCIe aguas abajo mediante el uso de una línea de conexión interna; y determinar, en función de una identificación de completador transportada en el paquete de respuesta de lectura/escritura de configuración, que el dispositivo de conmutación está conectado a un dispositivo de E/S cuya identificación (por ejemplo, una BDF) es la identificación de completador. De esta manera, cuando un dispositivo de conmutación falla, solo el al menos un host y el al menos un dispositivo de E/S que están conectados al dispositivo de conmutación resultan afectados, y otro dispositivo de conmutación, y un host y un dispositivo de E/S que están conectados al otro dispositivo de conmutación en un sistema no son afectados. Por lo tanto, en comparación con una CPU de administración externa en la técnica anterior, es posible mejorar tanto la estabilidad como la confiabilidad del sistema.

En una posible primera implementación del primer aspecto, si el al menos un puerto PCIe aguas abajo se conecta a un dispositivo de E/S de función única, el dispositivo de conmutación se configura específicamente para transmitir de manera transparente información entre el dispositivo de E/S de función única y un host que corresponde al dispositivo de E/S de función única.

En este caso, una función del dispositivo de E/S de función única solo puede ser usada por uno de los al menos un host, es decir que el dispositivo de E/S de función única corresponde al host que usa la función del dispositivo de E/S de función única.

En referencia a las posibles implementaciones anteriores, en una segunda implementación posible del primer aspecto, el dispositivo de conmutación además incluye: al menos un módulo de dispositivo terminal de espejo, configurado para almacenar el contenido de configuración PCIe del al menos un dispositivo de E/S conectado al, al menos un, puerto PCIe aguas abajo; un módulo de mapeo, configurado para implementar el mapeo entre un dominio PCIe correspondiente al, al menos un, host y un dominio PCIe correspondiente al, al menos un, dispositivo de E/S; y al menos un módulo de dispositivo terminal virtual, configurado para virtualizar una función del al menos un dispositivo de E/S conectado al, al menos un, puerto PCIe aguas abajo, de modo tal que la función sea usada por el al menos un host.

Opcionalmente, un extremo del al menos un módulo de dispositivo terminal de espejo se conecta al, al menos un, puerto PCIe aguas abajo, un extremo del al menos un módulo de dispositivo terminal virtual se conecta a los múltiples puertos PCIe aguas arriba, y el módulo de mapeo se conecta de manera separada a otro extremo del al menos un módulo de dispositivo terminal de espejo y otro extremo del al menos un módulo de dispositivo terminal virtual.

Opcionalmente, el módulo de mapeo puede configurarse específicamente para implementar el mapeo entre el al menos un módulo de dispositivo terminal de espejo y el al menos un módulo de dispositivo terminal virtual, por ejemplo, identificaciones y/o direcciones.

En referencia a la posible implementación anterior, en una tercera implementación posible del primer aspecto, un primer puerto PCIe aguas abajo en el al menos un puerto PCIe aguas abajo no presenta espacio de configuración PCIe, y cada uno de los múltiples puertos PCIe aguas arriba presenta un espacio de configuración PCIe.

En referencia a la posible implementación anterior, en una cuarta implementación posible del primer aspecto, el dispositivo de conmutación además incluye al menos un espacio de configuración PCIe que corresponde a cada uno de los al menos un puerto PCIe aguas abajo, y el al menos un módulo de dispositivo terminal virtual se conecta específicamente a los múltiples puertos PCIe aguas arriba mediante el uso del al menos un espacio de configuración PCIe correspondiente al, al menos un, puerto PCIe aguas abajo.

En referencia a las posibles implementaciones anteriores, en una quinta posible implementación del primer aspecto, el módulo de mapeo almacena: una primera tabla de mapeo, usada para almacenar una relación de mapeo desde una identificación en el dominio PCIe correspondiente al, al menos un, host hasta una identificación en el dominio PCIe correspondiente al, al menos un, dispositivo de E/S; una segunda tabla de mapeo, usada para almacenar una relación de mapeo desde la identificación en el dominio PCIe correspondiente al, al menos un, dispositivo de E/S hasta la identificación en el dominio PCIe correspondiente al, al menos un, host.

Opcionalmente, la primera tabla de mapeo se usa para almacenar una relación de mapeo desde una identificación del al menos un módulo de dispositivo terminal virtual hasta una identificación del al menos un módulo de dispositivo terminal de espejo. La segunda tabla de mapeo se usa específicamente para almacenar una relación de mapeo desde la identificación del al menos un módulo de dispositivo terminal de espejo hasta la identificación del al menos un módulo de dispositivo terminal virtual. Opcionalmente, el módulo de mapeo se configura específicamente para implementar el mapeo entre la identificación del al menos un módulo de dispositivo terminal virtual y la identificación del al menos un módulo de dispositivo terminal de espejo en función de la primera y la segunda tabla de mapeo.

En referencia a las posibles implementaciones anteriores, en una sexta implementación posible del primer aspecto, un primer módulo de dispositivo terminal de espejo en el al menos un módulo de dispositivo terminal de espejo es específicamente una tercera tabla de mapeo, y la tercera tabla de mapeo se usa para almacenar un registro de dirección base (BAR, por sus siglas en inglés) y un tamaño de BAR de una función virtual de un primer dispositivo de E/S en el al menos un dispositivo de E/S, donde el primer módulo de dispositivo terminal de espejo se configura para almacenar el contenido de la configuración PCIe del primer dispositivo de E/S.

Opcionalmente, el módulo de mapeo se configura específicamente para implementar el mapeo desde una dirección del al menos un módulo de dispositivo terminal virtual a una dirección del al menos un módulo de dispositivo terminal de espejo en función de la tercera y la primera tabla de mapeo.

En referencia a las posibles implementaciones anteriores, en una séptima implementación posible del primer aspecto, un primer módulo de dispositivo terminal virtual en el al menos un módulo de dispositivo terminal virtual se configura específicamente para virtualizar una función física de un segundo dispositivo de E/S en el al menos un dispositivo de E/S, de modo tal que la función física sea usada por un primer host en el al menos un host, donde un lector de función física del segundo dispositivo de E/S es cargado por un procesador del primer host.

En referencia a las posibles implementaciones anteriores, en una octava implementación posible del primer aspecto, un segundo módulo de dispositivo terminal virtual en el al menos un módulo de dispositivo terminal virtual se configura específicamente para virtualizar una función virtual de un tercer dispositivo de E/S en el al menos un dispositivo de E/S, de modo tal que la función virtual sea usada por un segundo host en el al menos un host, donde un lector de función física del tercer dispositivo de E/S es cargado por un BMC de administración en el al menos un host.

En referencia a las posibles implementaciones anteriores, en una novena implementación posible del primer aspecto, un tercer módulo de dispositivo terminal virtual en el al menos un módulo de dispositivo terminal virtual se configura específicamente para virtualizar una función física y una función virtual de un cuarto dispositivo de E/S en el al menos un dispositivo de E/S, de modo tal que la función física y la función virtual sean usadas por un tercer host en el al menos un host, donde un lector de función física del cuarto dispositivo de E/S es cargado por el BMC de administración en el al menos un host, y un lector de función física del tercer módulo de dispositivo terminal virtual es cargado por un procesador del tercer host.

En referencia a las posibles implementaciones anteriores, en una décima implementación posible del primer aspecto, el dispositivo de conmutación es específicamente un dispositivo de conmutación de host y un dispositivo de conmutación de E/S, y el dispositivo de conmutación de host y el dispositivo de conmutación de E/S están conectados mediante el uso de una interfaz Ethernet, donde el dispositivo de conmutación de host incluye múltiples puertos PCIe aguas arriba; y el dispositivo de conmutación de E/S incluye el aparato de procesamiento interno, el al menos un módulo de dispositivo terminal virtual, el módulo de mapeo, el al menos un módulo de dispositivo terminal de espejo y el al menos un puerto PCIe aguas abajo.

En referencia a las posibles implementaciones anteriores, en una onceava implementación posible, un aparato de procesamiento interno se configura para: recibir una instrucción de inicialización enviada por el BMC de administración en el al menos un host; y transmitir, en función de la instrucción de inicialización, un paquete de lectura/escritura de configuración al, al menos un, puerto PCIe aguas abajo mediante el uso de la línea de conexión interna; y el aparato de procesamiento interno además se configura para: después de que un dispositivo de E/S conectado al dispositivo de conmutación sea determinado, informar, al BMC de administración, los datos sobre el dispositivo de E/S conectado al dispositivo de conmutación, donde la información incluye información de identificación.

Opcionalmente, el aparato de procesamiento interno puede conectarse a los múltiples puertos PCIe aguas arriba mediante el uso de la línea de conexión interna, y el aparato de procesamiento interno puede recibir, específicamente, mediante el uso de los múltiples puertos PCIe aguas arriba, la instrucción de inicialización enviada por el BMC de administración. El aparato de procesamiento interno puede transmitir el paquete de lectura/escritura de configuración al, al menos un, puerto PCIe aguas abajo en función de la instrucción de inicialización, en función de un paquete de respuesta de lectura/escritura de configuración recibido, la información sobre el dispositivo de E/S conectado al dispositivo de conmutación, e informar, al BMC de administración, mediante el uso de los múltiples puertos PCIe aguas arriba, la información sobre el dispositivo de E/S conectado al dispositivo de conmutación, donde la información puede incluir información de identificación (por ejemplo, una BDF) o puede incluso contener una estructura de topología (por ejemplo, un árbol de estructura PCIe) o similares.

En referencia a las posibles implementaciones anteriores, en una doceava implementación posible del primer aspecto, el aparato de procesamiento interno además se configura para: recibir información de administración de configuración enviada por el BMC de administración en el al menos un host; y configurar los múltiples puertos PCIe aguas arriba y el al menos un puerto PCIe aguas abajo en función de la información de administración de configuración.

En referencia a las posibles implementaciones anteriores, en una treceava implementación posible del primer aspecto, el aparato de procesamiento interno además se configura para procesar un evento de excepción y un evento de intercambio en caliente.

Breve descripción de los dibujos

La FIG. 1 es un diagrama de bloque esquemático de un sistema PCIe en la técnica anterior;

la FIG. 2 es un diagrama de bloque esquemático de un sistema PCIe según una realización de la presente invención;

la FIG. 3 es otro diagrama de bloque esquemático de un sistema PCIe según una realización de la presente invención;

la FIG. 4 es otro diagrama de bloque esquemático de un sistema PCIe según una realización de la presente invención;

la FIG. 5 es otro diagrama de bloque esquemático de un sistema PCIe según una realización de la presente invención;

la FIG. 6 es otro diagrama de bloque esquemático de un sistema PCIe según una realización de la presente invención;

la FIG. 7 es otro diagrama de bloque esquemático de un sistema PCIe según una realización de la presente invención;

la FIG. 8 es otro diagrama de bloque esquemático de un sistema PCIe según una realización de la presente invención;

la FIG. 9 es otro diagrama de bloque esquemático de un sistema PCIe según una realización de la presente invención; y

la FIG. 10 es un diagrama de flujo esquemático de un procedimiento para inicializar un sistema PCIe según una realización de la presente invención.

Descripción de las realizaciones

La invención realizada se describe en las reivindicaciones independientes adjuntas. En el conjunto de reivindicaciones dependientes, se describen realizaciones adicionales.

Lo siguiente describe las soluciones técnicas en las realizaciones de la presente invención en referencia a los dibujos adjuntos en las realizaciones de la presente invención.

La FIG. 2 es un diagrama arquitectónico de un sistema PCIe 200 según una realización de la presente invención. El sistema PCIe 200 puede configurarse para implementar el uso compartido de un dispositivo de E/S mediante múltiples procesos en un host o implementar el uso compartido de un dispositivo de E/S por parte de múltiples hosts.

Como se muestra en la FIG. 2, el sistema PCIe 200 incluye: N_1 hosts 210, un dispositivo de conmutación 220 y M_1 dispositivos de E/S 230, donde $N_1 \geq 1$ y $M_1 \geq 1$. Al dispositivo de E/S también se puede hacer referencia como el dispositivo terminal (en inglés, Endpoint - EP).

La FIG. 2 muestra, a través del uso de un ejemplo, que el sistema PCIe 200 incluye solo un dispositivo de conmutación y el dispositivo de conmutación está conectado a al menos un host y al menos un dispositivo de E/S. Sin embargo,

debe entenderse que el sistema PCIe 200 puede incluir, de manera alternativa, múltiples dispositivos de conmutación, y cada dispositivo de conmutación puede estar conectado a al menos un host y al menos un dispositivo de E/S. Cada dispositivo de conmutación y al menos un host y al menos un dispositivo de E/S que están conectados al dispositivo de conmutación pueden considerarse como una unidad de red. De manera correspondiente, el sistema PCIe 200 puede incluir una o más unidades de red. Sin embargo, esto no se limita a esta realización de la presente invención.

Opcionalmente, en esta realización de la presente invención, el host puede ser específicamente un dispositivo como un servidor o un ordenador personal. De manera alternativa, el host puede ser una máquina virtual. Esto no se limita a esta realización de la presente invención. Opcionalmente, algunos o todos los N_1 hosts pueden incluir un mando de administración de la placa base (en inglés, Baseboard Management Controller - BMC), es decir que los N_1 hosts pueden incluir uno o más BMC. Al menos uno del uno o más BMC puede presentar una función de administración de configuración, y se configura específicamente para administrar y controlar la unidad de red, incluyendo el al menos un host, el dispositivo de conmutación y el al menos un dispositivo de E/S. Por cuestiones de facilidad de descripción, en la descripción a continuación, a un BMC que presenta una función de administración de configuración se hace referencia como BMC de administración. Opcionalmente, el BMC de administración se puede configurar para implementar una función de administración de configuración que posee una CPU de administración en la técnica anterior. Sin embargo, a diferencia de la CPU de administración de la técnica anterior, el BMC de administración puede responsabilizarse solo de la administración de configuración de un dispositivo de conmutación, y al menos un host y al menos un dispositivo de E/S que se conectan al dispositivo de conmutación en una unidad de red a la que pertenece el BMC de administración, pero no puede ser responsable de la administración de configuración de los hosts, los dispositivos de E/S, y los dispositivos de conmutación en otras unidades de red que pueden existir en una red. Esto no se limita a esta realización de la presente invención.

Opcionalmente, el dispositivo de E/S puede ser específicamente un dispositivo tal como un adaptador de bus de host (en inglés, Host Bus Adapter - HBA) o un disco de estado sólido. (en inglés, Solid State disk - SSD). Esto no se limita a esta realización de la presente invención.

En esta realización de la presente invención, el dispositivo de conmutación 220 puede incluir N_2 puertos PCIe aguas arriba 221, M_2 puertos PCIe aguas abajo 222 y un aparato de procesamiento interno 223, donde $N_2 > 1$ y $M_2 \geq 1$. Los N_2 puertos PCIe aguas arriba 221 se configuran para conectarse a los N_1 hosts 210. Los M_2 puertos PCIe aguas abajo 222 se configuran para conectarse a los M_1 dispositivos de E/S 230. El aparato de procesamiento interno 223 puede conectarse a los M_2 puertos PCIe aguas abajo 222 mediante el uso de una línea de conexión interna del dispositivo de conmutación, y puede conectarse a los N_2 puertos PCIe aguas arriba 221 mediante el uso de la línea de conexión interna del dispositivo de conmutación. Específicamente, el dispositivo de conmutación 220 puede conectarse a los N_1 hosts 210 mediante el uso de los N_2 puertos PCIe aguas arriba 221. Si $N_2 = N_1$, es decir, una cantidad de los puertos PCIe aguas arriba del dispositivo de conmutación es igual a la cantidad de hosts conectados al dispositivo de conmutación, los N_2 puertos PCIe aguas arriba pueden conectarse a los N_1 hosts en una correspondencia de uno a uno. Si $N_2 > N_1$, es decir, una cantidad de los puertos PCIe aguas arriba del dispositivo de conmutación es superior a una cantidad de los hosts conectados al dispositivo de conmutación, N_1 de los N_2 puertos PCIe aguas arriba puede conectarse a los N_1 hosts en una correspondencia de uno a uno, y los puertos PCIe aguas arriba restantes (N_2-N_1) pueden permanecer inactivos o establecerse como puertos PCIe aguas abajo. Sin embargo, esto no se limita a esta realización de la presente invención.

De manera similar, el dispositivo de conmutación 220 puede conectarse a los M_1 dispositivos de E/S 230 mediante el uso de los M_2 puertos PCIe aguas abajo 222. Si $M_2 = M_1$, es decir, una cantidad de los puertos PCIe aguas abajo del dispositivo de conmutación es igual a la cantidad de dispositivos de E/S conectados al dispositivo de conmutación, los M_2 puertos PCIe aguas abajo pueden conectarse a los M_1 dispositivos de E/S en una correspondencia de uno a uno. Si $M_2 > M_1$, es decir, una cantidad de los puertos PCIe aguas abajo del dispositivo de conmutación es superior a una cantidad de los dispositivos de E/S conectados al dispositivo de conmutación, el M_1 de los M_2 puertos PCIe aguas abajo puede conectarse a los M_1 dispositivos de E/S en una correspondencia de uno a uno, y los puertos PCIe aguas abajo restantes (M_2-M_1) pueden permanecer inactivos o establecerse como puertos PCIe aguas arriba. Si $M_2 < M_1$, es decir, una cantidad de los puertos PCIe aguas abajo del dispositivo de conmutación es inferior a la cantidad de dispositivos de E/S conectados al dispositivo de conmutación, el sistema PCIe 200 puede incluir además uno o más conmutadores PCIe, y algunos o todos los M_2 puertos PCIe aguas abajo pueden conectarse a los M_1 dispositivos de E/S mediante el uso del conmutador PCIe. De esta manera, la cantidad de los dispositivos de E/S conectados al dispositivo de conmutación puede aumentar, y el rendimiento de conmutación no se reduce. Sin embargo, esto no se limita a esta realización de la presente invención.

Opcionalmente, el sistema PCIe 200 incluye además el conmutador PCIe, y los M_2 puertos PCIe aguas abajo pueden conectarse a los M_1 dispositivos de E/S mediante el uso del conmutador PCIe.

Además, el dispositivo de conmutación 220 además incluye el aparato de procesamiento interno 223. El aparato de procesamiento interno 223 puede ser específicamente un módulo de procesamiento, por ejemplo, un procesador y/o un circuito de procesamiento de hardware. Opcionalmente, el aparato de procesamiento interno puede implementarse mediante un procesador y un circuito de hardware. De manera alternativa, el aparato de procesamiento interno puede implementarse solo mediante un circuito de hardware. Esto no se limita a esta realización de la presente invención.

En esta realización de la presente invención, el aparato de procesamiento interno puede configurarse para enumerar un dispositivo físico (es decir, un dispositivo real, incluso sin módulo funcional alguno) conectado al dispositivo de conmutación, por ejemplo, determinar un dispositivo físico conectado al dispositivo de conmutación mediante el uso de los M_2 puertos PCIe aguas abajo 222, o puede, adicionalmente, obtener información, por ejemplo, una identificación (una BDF) o una estructura de topología, de múltiples dispositivos físicos conectados al dispositivo de conmutación y puede, además, configurar un árbol de estructura de dominio PCIe en función la información sobre los múltiples dispositivos físicos. El dispositivo físico además puede incluir el dispositivo de E/S y/o el conmutador PCIe. En este caso, el aparato de procesamiento interno puede considerarse como un puerto raíz virtual (en inglés, Root Port - RP) o un complejo raíz (en inglés, Root Complex - RC). Sin embargo, esto no se limita a esta realización de la presente invención.

Opcionalmente, el aparato de procesamiento interno 223 puede ejecutar activamente la enumeración de dispositivos, por ejemplo, ejecutar una enumeración de dispositivo en una etapa de inicialización del sistema. Alternativamente, el aparato de procesamiento interno 223 puede efectuar, al recibir una instrucción de otro dispositivo, la enumeración de dispositivos en función de la instrucción. Por ejemplo, el aparato de procesamiento interno 223 puede recibir una instrucción de inicialización o una instrucción de enumeración que se envía mediante el BMC de administración, y ejecutar la enumeración de dispositivos en función de la instrucción de inicialización o la instrucción de enumeración. Sin embargo, esto no se limita a esta realización de la presente invención.

Específicamente, al ejecutar la enumeración de dispositivos, el aparato de procesamiento interno 223 puede enviar múltiples paquetes de lectura/escritura de configuración al, al menos un, puerto PCIe aguas abajo 222. El paquete de lectura/escritura de configuración transporta una identificación de solicitud (en inglés, Request Identification - RID) y una identificación de completador (en inglés, Completer Identification - CID), donde un valor de la RID puede establecerse en una BDF del aparato de procesamiento interno 223, y un valor de la CID puede enumerarse en secuencia, comenzando desde un valor inicial. El aparato de procesamiento interno 223 puede recibir un paquete de respuesta de lectura/escritura de configuración correspondiente a al menos uno de múltiples paquetes de lectura/escritura de configuración. El paquete de respuesta de lectura/escritura de configuración transporta una RID y una CID en el paquete de lectura/escritura de configuración correspondiente. De esta manera, el aparato de procesamiento interno 223 determina, mediante el reconocimiento del paquete de respuesta de lectura/escritura de configuración (por ejemplo, mediante el reconocimiento de la CID en el paquete de respuesta de lectura/escritura de configuración), si el dispositivo de conmutación 220 está conectado a un dispositivo físico cuyo número de BDF es la CID en el paquete de respuesta de lectura/escritura de configuración Sin embargo, esto no se limita a esta realización de la presente invención.

En una realización opcional, si el aparato de procesamiento interno 223 es implementado por un procesador y un circuito de hardware, donde el procesador y el circuito de hardware pueden conectarse mediante el uso de un bus interno, el procesador se puede configurar para: recibir una instrucción de inicialización (o una instrucción de enumeración) desde el BMC de administración; generar, en función de la instrucción de inicialización, información sobre un paquete de enumeración, la cual incluye datos que necesitan los paquetes de lectura/escritura de configuración de tipo 0 (Tipo0) y tipo 1 (Tipo1); y producir la información generada al bus interno. El circuito de hardware puede recibir, mediante el uso del bus interno, la información generada por el procesador, encapsular la información como un paquete de capa de transacciones (en inglés, Transaction Layer Packet - TLP) y enviar el TLP al, al menos un, puerto PCIe aguas abajo 222 mediante el uso de un circuito o módulo de siguiente nivel conectado al aparato de procesamiento interno 223. Además, el circuito de hardware puede configurarse adicionalmente para reconocer un paquete de respuesta de lectura/escritura de configuración al recibir el paquete de respuesta de lectura/escritura de configuración enviado por el al menos un puerto PCIe aguas abajo 222. Por ejemplo, el circuito de hardware puede recibir un paquete de compleción (en inglés, Completion - CPL)/(en inglés, Completion Data - CPLD) transmitido por el al menos un puerto PCIe aguas abajo 222 y reconocer una ID en el paquete CPL/CPLD recibido. Por medio del disparo de interrupción del circuito de hardware, el paquete CPL/CPLD puede transmitirse al procesador del aparato de procesamiento interno 223 mediante el uso del bus interno. Específicamente, el circuito de hardware puede desensamblar el paquete CPL/CPLD en un formato de datos que cumpla con una secuencia de tiempo del bus interno y enviar el paquete CPL/CPLD al bus interno. El procesador del aparato de procesamiento interno 223 puede configurarse adicionalmente para: analizar la información del paquete recibido, almacenar la información del paquete y generar un árbol de estructura de un dominio PCIe en función de la información del paquete. Después de haber completado la enumeración del dispositivo, el procesador puede transmitir, además, la información sobre el árbol de estructura PCIe generado al BMC de administración mediante el uso de los múltiples puertos PCIe aguas arriba 221. Sin embargo, esto no se limita a esta realización de la presente invención.

Por lo tanto, el sistema PCIe provisto en esta realización de la presente invención incluye al menos un host, un dispositivo de conmutación y el al menos un dispositivo de E/S. El dispositivo de conmutación incluye los múltiples puertos PCIe aguas arriba configurados para conectarse al, al menos un, host; el al menos un puerto PCIe aguas abajo configurado para conectarse al, al menos un, dispositivo de E/S; y el aparato de procesamiento interno. El aparato de procesamiento se conecta al, al menos un, puerto PCIe aguas abajo mediante el uso de un bus interno y se configura para: transmitir un paquete de lectura/escritura de configuración al, al menos un, puerto PCIe aguas abajo mediante el uso de la línea de conexión interna; recibir un paquete de respuesta de lectura/escritura de configuración transmitido mediante el al menos un puerto PCIe aguas abajo mediante el uso de una línea de conexión interna; y determinar, en función de una identificación de completador transportada en el paquete de respuesta de

lectura/escritura de configuración, que el dispositivo de conmutación está conectado a un dispositivo de E/S cuya identificación es la identificación de completador. De esta manera, cuando un dispositivo de conmutación en el sistema PCIe falla, solo el al menos un host y el al menos un dispositivo de E/S que están conectados al dispositivo de conmutación puede resultar afectados, y otro dispositivo de conmutación, y un host y un dispositivo de E/S que están conectados al otro dispositivo de conmutación en el sistema PCIe no son afectados. Por lo tanto, en comparación con una CPU de administración externa en la técnica anterior, es posible mejorar tanto la estabilidad como la confiabilidad del sistema PCIe.

Opcionalmente, el aparato de procesamiento interno 223 puede presentar además al menos una de las siguientes funciones: procesamiento de un evento de intercambio en caliente, procesamiento de eventos de excepción o configuración de componentes (por ejemplo, configuración de registros).

El aparato de procesamiento interno 223 se puede configurar para procesar un evento de intercambio en caliente. Específicamente, el aparato de procesamiento interno 223 puede detectar un evento de intercambio en caliente accionado por el BMC de administración y/o el hardware, y ejecutar un procedimiento de procesamiento de intercambio en caliente correspondiente. El aparato de procesamiento interno 223 puede configurarse además para procesar un evento de excepción. Específicamente, el aparato de procesamiento interno 223 puede procesar una excepción de hardware, asentar un registro de evento de excepción e informar el registro de evento de excepción. Además, el aparato de procesamiento interno 223 se puede configurar adicionalmente para configurar un componente. Específicamente, el aparato de procesamiento interno 223 puede configurar un registro en una etapa de inicialización o modificar una configuración del registro. El aparato de procesamiento interno 223 puede obtener información necesaria para la configuración de componentes en un procedimiento de enumeración de dispositivos, o puede recibir información de administración de configuración enviada por el BMC de administración, configurar el componente en función de la información de administración de configuración y similares. Esto no se limita a esta realización de la presente invención.

En una realización opcional, el aparato de procesamiento interno 223 es implementado por el procesador y el circuito de hardware, al implementar la función de procesamiento de evento de excepción, el circuito de hardware en el aparato de procesamiento interno 223 puede capturar la información de excepción relacionada con una función física del dispositivo de E/S y accionar el procesador mediante el uso de la información de interrupción configurada por el procesador en el aparato de procesamiento interno 223. El procesador del aparato de procesamiento interno 223 puede reconocer un tipo de excepción. Si este último es específicamente corregible o es un error no fatal que no puede corregirse, el procesador puede ejecutar un autoprocesamiento y producir solo una alarma sin informar. Si el tipo de excepción es específicamente un error fatal que no puede corregirse, el procesador puede romper un enlace del dispositivo de E/S e informar el enlace roto del dispositivo de E/S al BMC de administración. Después de reconocer el dispositivo de E/S cuyo enlace está roto, el BMC de administración puede desinstalar un lector correspondiente al dispositivo de E/S. Además, opcionalmente, el BMC de administración puede notificar un módulo de administración de administración de plataformas (en inglés, Shelf Management Module - SMM) sobre la actualización de la información de red. El SMM administra y controla a un host correspondiente al dispositivo de E/S que presenta una excepción, por ejemplo, ejecuta la desinstalación del lector o restablece el procesamiento de inicialización. Esto no se limita a esta realización de la presente invención.

En esta realización de la presente invención, el aparato de procesamiento interno 223 puede configurarse para: recibir un paquete de datos enviado por los M_1 dispositivos de E/S 230 mediante el uso de los M_2 puertos PCIe aguas abajo 222, procesar el paquete de datos recibido, por ejemplo, convertir una secuencia de tiempo, y enviar el paquete de datos procesado a los N_2 puertos PCIe aguas arriba 221 mediante el uso de la línea de conexión interna. Además, el aparato de procesamiento interno 223 puede configurarse adicionalmente para: recibir un paquete de datos enviado por los N_1 hosts 210 mediante el uso de los N_2 puertos PCIe aguas arriba 221, procesar el paquete de datos recibido, por ejemplo, convertir una secuencia de tiempo, y enviar el paquete de datos procesado a los M_2 puertos PCIe aguas abajo 222 mediante el uso de la línea de conexión interna.

En una realización opcional, si el aparato de procesamiento interno 223 es implementado por el procesador y el circuito de hardware, al recibir los datos enviados por el host, el procesador en el aparato de procesamiento interno puede analizar una secuencia de tiempo del bus de una interfaz del BMC de administración y producir los datos en una secuencia de tiempo del bus interno. Al recibir los datos reenviados mediante un circuito o módulo de siguiente nivel, el procesador puede analizar la secuencia de tiempo del bus interno y producir los datos para un bus de la interfaz del BMC de administración. Sin embargo, esto no se limita a esta realización de la presente invención.

Opcionalmente, en esta realización de la presente invención, el BMC de administración puede informar una información de red (o información de configuración de red) al SMM mediante el uso de un puerto de red. El SMM se puede configurar para que sea responsable de administrar toda una red. Un usuario puede establecer, en función de un requerimiento, una conexión de mapeo entre el host y el dispositivo de E/S mediante el uso de una interfaz de interacción humano-ordenador del SMM. En este caso, opcionalmente, el BMC de administración puede configurarse adicionalmente para: recibir una información de requerimiento de configuración enviada por el SMM y determinar la información de administración de configuración en función de la información de requerimiento de configuración. Sin embargo, esto no se limita a esta realización de la presente invención.

Opcionalmente, en esta realización de la presente invención, cada uno de los N_2 puertos PCIe aguas arriba 221 presenta un espacio de configuración PCIe, y ninguno de los M_2 puertos PCIe aguas abajo 222 presenta espacio de configuración PCIe alguno.

En esta realización de la presente invención, el puerto PCIe aguas arriba 221 en el dispositivo de conmutación 220 presenta el espacio de configuración PCIe y es un puerto PCIe estándar (es decir, un puente P2P aguas arriba estándar). El puerto PCIe aguas abajo 222 en el dispositivo de conmutación 220 no presenta ningún espacio de configuración PCIe y no es un puerto PCIe estándar (es decir, no es un puente P2P aguas abajo estándar). Opcionalmente, el dispositivo de conmutación 220 podría no almacenar, adicionalmente, los espacios de configuración PCIe de los M_2 puertos PCIe aguas abajo 222, es decir que no hay ningún espacio de configuración PCIe del puerto PCIe aguas abajo 222 en el dispositivo de conmutación 220. De manera alternativa, el dispositivo de conmutación 220 puede almacenar, adicionalmente, un espacio de configuración de cada uno de los M_2 puertos PCIe aguas abajo 222, es decir que el espacio de configuración PCIe del puerto PCIe aguas abajo 222 se separa del puerto PCIe aguas abajo 222 en el dispositivo de conmutación 220. De manera alternativa, el dispositivo de conmutación 220 puede almacenar adicionalmente espacios de configuración PCIe de algunos de los M_2 puertos PCIe aguas abajo 222 pero no almacenar adicionalmente espacios de configuración PCIe de algunos otros puertos PCIe aguas abajo 222. Esto no se limita a esta realización de la presente invención.

Opcionalmente, en esta realización de la presente invención, el puerto PCIe aguas abajo 222 puede corresponder al, al menos un, espacio de configuración PCIe, y el al menos un espacio de configuración PCIe correspondiente al puerto PCIe aguas abajo 222 y el puerto PCIe aguas abajo 222 pueden disponerse individualmente en el dispositivo de conmutación 220, es decir, el puerto PCIe aguas abajo 222 puede estar separado del al menos un espacio de configuración PCIe correspondiente al puerto PCIe aguas abajo 222. Además, opcionalmente, el al menos un espacio de configuración PCIe correspondiente al puerto PCIe aguas abajo 222 puede pertenecer al dominio PCIe correspondiente a los N_1 hosts 210. De esta manera, cuando el dispositivo de conmutación 220 almacena múltiples espacios de configuración PCIe correspondientes a un puerto PCIe aguas abajo 222, el puerto PCIe aguas abajo 222 puede usarse como puertos múltiples a ser usados por los hosts 210, a fin de mejorar la utilización de recursos del sistema.

Debe entenderse que, para la descripción en la realización anterior, se usa un ejemplo en el que todos los N_2 puertos PCIe aguas arriba 221 en el dispositivo de conmutación 220 presentan el espacio de configuración PCIe y todos los M_2 puertos PCIe aguas abajo 222 no presentan espacio de configuración PCIe alguno. Opcionalmente, algunos de los M_2 puertos PCIe aguas abajo 222 pueden presentar un espacio de configuración PCIe y algunos otros puertos PCIe aguas abajo 222 podrían no presentar un espacio de configuración PCIe. De manera alternativa, todos los M_2 puertos PCIe aguas abajo 222 presentan un espacio de configuración PCIe. En otra realización opcional, algunos o todos los N_2 puertos PCIe aguas arriba 221 pueden no presentar espacio de configuración PCIe alguno. Esto no se limita a esta realización de la presente invención.

Opcionalmente, como se muestra en la FIG. 3, el dispositivo de conmutación 220 además incluye: N_3 módulos de dispositivo terminal virtual (abreviado, vEP, por sus siglas en inglés) 224, un módulo de mapeo (MAP, por sus siglas en inglés) 225 y M_3 módulos de dispositivo terminal de espejo (abreviado, mEP, por sus siglas en inglés) 226.

Los N_3 vEP 224 pueden conectarse a los N_2 puertos PCIe aguas arriba 221, y se configuran para virtualizar funciones de los M_1 dispositivos de E/S 230 conectados a los M_2 puertos PCIe aguas abajo 222, de modo tal que las funciones sean usadas por los N_1 hosts 221 conectados a los N_2 puertos PCIe aguas arriba 221, donde $N_3 \geq 1$.

Los M_3 mEP 226 pueden conectarse a los M_2 puertos PCIe aguas abajo 222, y se configuran para almacenar contenido de configuración PCIe de los M_1 dispositivos de E/S 230 conectados a los M_2 puertos PCIe aguas abajo 222, donde $M_3 \geq 1$.

El módulo de mapeo 225 puede conectarse por separado a los N_3 vEP 224 y los M_3 mEP 226, y se configura para implementar el mapeo entre un dominio PCIe correspondiente a los N_1 hosts 210 y un dominio PCIe correspondiente a los M_1 dispositivos de E/S 230.

En esta realización de la presente invención, opcionalmente, el dispositivo de conmutación 220 puede incluir además uno o más mEP 226, el módulo de mapeo 225 y uno o más vEP 224. Un valor de N_3 puede determinarse por una cantidad de funciones que deben o pueden ser usadas por los N_1 hosts 210, y N_3 puede ser igual a M_3 o no. Por ejemplo, tanto N_3 y M_3 pueden ser iguales a una cantidad total de funciones (por ejemplo, una cantidad total de VF) que poseen los M_1 dispositivos de E/S 230. Esto no se limita a esta realización de la presente invención.

Específicamente, el mEP 226 puede ser un reflejo de un EP real y se configura para almacenar el contenido de configuración del dispositivo de E/S (es decir, el dispositivo terminal) 230. El vEP 224 puede configurarse para virtualizar una función física y/o una función virtual del dispositivo terminal 230, y puede ser específicamente un espacio de configuración PCIe correspondiente a la función física de la función virtual del dispositivo terminal 230. El módulo de mapeo 225 puede configurarse para implementar el mapeo entre identificaciones y/o direcciones del dominio PCIe correspondiente a los N_1 hosts 210 y el dominio PCIe correspondiente al, al menos un, dispositivo de E/S 230, para determinar un módulo (o componente) al que se envía la información y/o los datos recibidos. En un ejemplo opcional,

el vEP 224 puede pertenecer al dominio PCIe correspondiente al host 210 y el mEP 226 puede pertenecer al dominio PCIe correspondiente al dispositivo de E/S 230. Correspondientemente, el módulo de mapeo 225 puede configurarse específicamente para implementar un mapeo entre una identificación y/o una dirección de los M_3 mEP 226 y una identificación y/o dirección de los N_3 vEP 224 (es decir, ejecutar un mapeo desde los vEP a los mEP o ejecutar un mapeo desde los mEP a los vEP), a fin de ejecutar un procesamiento de reenvío en un paquete de datos transmitido entre los N_1 hosts 210 y los M_1 dispositivos de E/S 230. Sin embargo, esto no se limita a esta realización de la presente invención.

Opcionalmente, los N_2 puertos PCIe aguas arriba 221, los N_3 vEP 224, el módulo de mapeo 225, los M_3 mEP 226 y los M_2 puertos PCIe aguas abajo 222 pueden conectarse en secuencia. Específicamente, un extremo del vEP 224 puede conectarse al puerto PCIe aguas arriba 221 y otro extremo puede conectarse al módulo de mapeo 225. Los N_3 vEP 224 pueden conectarse directa o indirectamente a algunos o todos los N_2 puertos PCIe aguas arriba 221. Un extremo del mEP 226 puede conectarse al puerto PCIe aguas abajo 222 y otro extremo puede conectarse al módulo de mapeo 225. Los M_3 mEP 226 pueden conectarse directa o indirectamente a algunos o todos los M_2 puertos PCIe aguas abajo 222. Sin embargo; esto no se limita a esta realización de la presente invención.

Si los espacios de configuración PCIe de los M_2 puertos PCIe aguas abajo 222 no se configuran adicionalmente en el dispositivo de conmutación 220, los N_3 vEP 224 pueden conectarse directamente a algunos o todos los N_2 puertos PCIe aguas arriba 221. Opcionalmente, si el dispositivo de conmutación 220 además almacena, de manera adicional, un espacio de configuración PCIe correspondiente a uno o más puertos PCIe aguas abajo 222, por ejemplo, como se muestra en la FIG. 3, el dispositivo de conmutación 220 puede almacenar N_4 espacios de configuración de puerto PCIe aguas abajo (abreviado, DP_CFG, por sus siglas en inglés) 227, cada puerto PCIe aguas abajo 222 puede corresponder a cero, uno o más piezas del DP_CFG 227. En este caso, los N_3 vEP 224 pueden estar conectados a los N_2 puertos PCIe aguas arriba 221 mediante el uso de las N_4 piezas del DP_CFG 227. De este modo, el puerto PCIe aguas arriba 221 se conecta al vEP 224 mediante el uso del espacio de configuración PCIe que corresponde al puerto PCIe aguas abajo. 222, a fin de aumentar una cantidad de vEP que puede ser usada por cada host, mejorando así el rendimiento del sistema.

Además, en esta realización de la presente invención, una función virtual se presenta al host, de modo tal que el sistema PCIe soporte cualquier versión de Windows para usar la función virtual en el host.

En una realización opcional, el módulo de mapeo 225 almacena:

una primera tabla de mapeo, usada para almacenar una relación de mapeo desde una identificación en el dominio PCIe correspondiente a los N_1 hosts 210 hasta una identificación en el dominio PCIe correspondiente a los M_1 dispositivos de E/S 230; y

una segunda tabla de mapeo, usada para almacenar una relación de mapeo desde la identificación en el dominio PCIe correspondiente a los M_1 dispositivos de E/S 230 hasta la identificación en el dominio PCIe correspondiente a los N_1 hosts 210.

Opcionalmente, la primera tabla de mapeo puede usarse específicamente para almacenar una relación de mapeo desde una identificación de los N_3 vEP 224 hasta una identificación de los M_3 mEP 226, por ejemplo, una relación de mapeo desde un número de RDF de al menos un vEP 224 para un número de RDF de al menos un mEP 226. La segunda tabla de mapeo puede usarse específicamente para almacenar una relación de mapeo desde la identificación de los M_3 mEP 226 hasta la identificación de los N_3 vEP 224. Sin embargo, esto no se limita a esta realización de la presente invención.

Opcionalmente, en esta realización de la presente invención, el módulo de mapeo 225 puede implementarse mediante un registro y un circuito de procesamiento de hardware. En este caso, la primera tabla de mapeo y la segunda tabla de mapeo pueden ser almacenadas individualmente por el registro. Además, opcionalmente, las identificaciones en la primera y la segunda tabla de mapeo pueden estar representadas por las BDF. De manera alternativa, una función del dispositivo de E/S 230 o una función del vEP 224 pueden volverse a enumerar, por ejemplo, con una secuencia que empieza en 1, a fin de obtener un índice de cada función, y el índice se almacena en las tablas de mapeo en una forma de índice de función, para ahorrar espacio de almacenamiento.

La Tabla 1 y la Tabla 2, respectivamente, muestran posibles implementaciones de la primera tabla de mapeo y la segunda tabla de mapeo. Como se muestra en la Tabla 1, la primera tabla de mapeo puede incluir cuatro columnas. Un valor en la columna de "índice de función" indica un índice de una función en el dominio PCIe correspondiente al dispositivo de E/S 230. El índice se puede obtener mediante la numeración de las BDF de todas las funciones, y puede ocupar uno o varios bits. Un valor en una columna de "habilitar indicación" indica si existe una función correspondiente al índice de función. Un número de bus mEP y un número de función mEP pueden indicar respectivamente un número de bus y un número de un mEP 226 correspondiente al Índice de función. El número de bus mEP y el número de función mEP se combinan en un número de BDF del mEP 226. El número de BDF del -mEP 226 se puede asociar a un número de BDF de un dispositivo de E/S 230 correspondiente al mEP 226. Una cantidad de filas incluida en la primera tabla de mapeo no se limita a esta realización de la presente invención. Por ejemplo, la primera tabla de mapeo puede incluir 1024 entradas (entrada). Sin embargo, esto no se limita a esta realización de la presente

invención.

Tabla 1: Posible implementación de la Primera tabla de mapeo

Índice de función	Habilitar indicación (un bit)	Número del bus mEP (ocho bits)	Número de función mEP

Como se muestra en la Tabla 2, la segunda tabla de mapeo puede incluir cinco columnas. Un valor en una columna de "índice de función" indica un índice (por ejemplo, un índice de función del vEP 224) de una función en el dominio PCIe (que puede ser un módulo funcional virtual y/o un dispositivo físico descubierto por el host 210 por medio de la enumeración) correspondiente al host 210. El índice se puede obtener mediante la numeración de las BDF de todas las funciones y puede ocupar uno o varios bits. Un valor en una columna de "habilitar indicación" indica si existe una función correspondiente al índice de función. Un número de bus de eVF y un número de función de dispositivo de eVF indican, respectivamente, un número de bus y un número de función de dispositivo de una eVF 224 correspondiente al índice de función. El número de bus de eVF y el número de función del dispositivo de eVF se combinan en una BDF de la eVF 224. Un valor en una columna de "indicación PF" se usa para indicar si la función correspondiente al índice de función es una VF o un PE. Una cantidad de filas incluidas en la segunda tabla de mapeo no se limita a esta realización de la presente invención. Por ejemplo, la segunda tabla de mapeo puede incluir 1024 entradas (entrada). Sin embargo, esto no se limita a esta realización de la presente invención.

Tabla 2: Posible implementación de la Segunda tabla de mapeo

Índice de función	Habilitar indicación (un bit)	Número del bus de eVF (ocho bits)	Número de función del dispositivo de eVF (ocho bits)	Indicación PF

De este modo, la relación de mapeo entre la identificación en el dominio PCIe correspondiente al host 210 y la identificación en el dominio PCIe correspondiente al dispositivo de E/S 230 se almacena de una manera de tabla de mapeo. En comparación con la técnica anterior, esta solución puede reducir el espacio de almacenamiento que ocupa el módulo de mapeo 225 y la complejidad, así como también ahorrar un recurso de almacenamiento del sistema. Además, se enumeran las BDF de todas las funciones y los números se almacenan en la tabla de mapeo. En comparación con el almacenamiento de una BDF de 16 bits, el espacio de almacenamiento que ocupa el módulo de mapeo 225 puede reducirse adicionalmente.

Opcionalmente, en esta realización de la presente invención, el contenido de la configuración PCIe del dispositivo terminal 230, almacenado en el mEP 226, puede ser específicamente un espacio de configuración PCIe del dispositivo terminal 230 o puede ser un contenido de configuración parcial en un espacio de configuración PCIe del dispositivo terminal 230. Por ejemplo, todos los M_3 mEP 226 pueden ser específicamente tablas de mapeo. De manera alternativa, algunos de los M_3 mEP 226 pueden ser específicamente tablas de mapeo. Esto no se limita a esta realización de la presente invención.

En una realización opcional, un primer mEP en los M_3 mEP 226 es específicamente una tercera tabla de mapeo. La tercera tabla de mapeo se usa para almacenar una dirección de registro de dirección de base (en inglés, Base Address Register - BAR) y un tamaño de BAR de una función virtual de un primer dispositivo de E/S en los M_1 dispositivos de E/S. El primer mEP está configurado para almacenar contenido de configuración del primer dispositivo de E/S.

La Tabla 3 muestra una posible implementación de una tercera tabla de mapeo. La tercera tabla de mapeo incluye un índice de mEP y una dirección BAR y un tamaño de BAR que corresponden al índice de mEP. Opcionalmente, cada mEP puede incluir seis BAR. Sin embargo; esto no se limita a esta realización de la presente invención.

Tabla 3: Posible implementación de la Tercera tabla de mapeo

Índice de mEP	BAR0 de VF (32 bits)	…	BAR5 de VF (32 bits)

En este caso, el módulo de mapeo 225 puede implementar el mapeo desde las direcciones en el dominio PCIe del

host 210 a las direcciones en el dominio PCIe del dispositivo de E/S 230 en referencia a una tercera y la primera tabla de mapeo. Sin embargo, esto no se limita a esta realización de la presente invención.

De este modo, la información de dirección de la función virtual del dispositivo de E/S se almacena en forma de tabla de mapeo. En comparación con la técnica anterior, esta solución puede reducir el espacio de almacenamiento que ocupa el mEP y ahorrar un recurso de almacenamiento del sistema.

Debe entenderse que los ejemplos de la Tabla 1 a la Tabla 3 pretenden ayudar a un experto en la materia a entender mejor esta realización de la presente invención, pero no limitan el alcance de esta realización de la presente invención. Aparentemente, un experto en la materia puede llevar a cabo varias modificaciones o cambios en función de los ejemplos proporcionados de la Tabla 1 a la Tabla 3. Las modificaciones o los cambios también se enmarcan en el alcance de esta realización de la presente invención.

Opcionalmente, en esta realización de la presente invención, la primera tabla de mapeo, la segunda tabla de mapeo o la tercera tabla de mapeo pueden configurarse en conjunto mediante el aparato de procesamiento interno 223 en el dispositivo de conmutación 220 y el host 210. Por ejemplo, un aparato de procesamiento interno 223 puede obtener, por medio de una enumeración de dispositivos o de otra manera, información sobre un dispositivo físico conectado al dispositivo de conmutación 220 mediante el uso de los M_1 puertos PCIe aguas abajo 222 y configurar la tabla de mapeo en función de la información obtenida, por ejemplo, almacenar la información sobre el dispositivo físico en la tabla de mapeo. El host 210 puede almacenar, en la tabla de mapeo, la información de configuración del dominio PCIe correspondiente al host 210, obtenida en un proceso de enumeración de dispositivos (es decir, descubriendo un dispositivo físico y/o un módulo funcional conectado al host 210) o de otra manera. Sin embargo, esto no se limita a esta realización de la presente invención.

Además, los N_3 vEP 224, el módulo de mapeo 225 y los M_3 mEP 226 pueden implementarse mediante un registro o por medio de un registro y un circuito de hardware, y son simples de usar sin implementar un árbol de estructura PCIe, mediante el uso de un código de software. Además, se evita un proceso de almacenamiento y reenvío en un proceso de implementación de software mediante el uso de una operación doble de escritura de configuración de un dispositivo virtual y un dispositivo real.

En esta realización de la presente invención, el sistema PCIe 200 puede soportar múltiples modos de operación. Los múltiples modos de operación pueden incluir al menos uno de los módulos siguientes: un modo directo de VP, un modo compartido de VF, un modo compartido de PF o un modo de transmisión transparente de EP. El modo directo de VF y el modo compartido de VF puede aplicarse a un dispositivo de E/S que soporte la SR-IOV; el modo de compartido de PF puede aplicarse a un dispositivo de E/S que soporta funciones múltiples; el modo de transmisión transparente de EP puede aplicarse a un dispositivo de E/S de función única.

Específicamente, en el modo directo de VF, los N_3 vEP 224 se configuran para virtualizar funciones virtuales de los M_1 dispositivos de E/S 230. Por ejemplo, como se muestra en la FIG. 4, un sistema PCIe puede incluir: dos hosts: un host 0 y un host 1; un conmutador PCIe; dos dispositivos terminales: un EP 0 y un EP 1; y un dispositivo de conmutación. El dispositivo de conmutación incluye: dos puertos PCIe aguas arriba respectivamente conectados al host 0 y al host 1, dos puertos PCIe aguas abajo conectados respectivamente al EP 0 y al EP 1 mediante el uso del conmutador PCIe, espacios de configuración PCIe DP_CFG 0 y DP_CFG 1 respectivamente, correspondientes a los puertos PCIe aguas abajo, un aparato de procesamiento interno, un módulo de mapeo, un vEP 0 y un vEP 1. El vEP 0 se conecta a un puerto PCIe aguas arriba mediante el uso de DP_CFG 0 y el vEP 1 se conecta al otro puerto PCIe aguas arriba mediante el uso de DP_CFG 1. Tanto el EP 0 como el EP 1 presentan una PF: una PF 0; y una VF: una VF 0 correspondiente a la PF 0. En este caso, el vEP 0 puede configurarse para virtualizar la VF 0 del EP 0, de modo tal que la VF 0 sea usada por el host 0. Específicamente, el vEP 0 puede ser una combinación de espacios de configuración PCIe de la PF 0 y la VF 0 del EP 0. El vEP 1 puede configurarse para virtualizar la VF 0 del EP 0, de modo tal que la VF 0 sea usada por el host 1. Específicamente, el vEP 1 puede ser una combinación de espacios de configuración PCIe de la PF 0 y la VF 0 del EP 1. Sin embargo, esto no se limita a esta realización de la presente invención.

Opcionalmente, en el modo directo de VF, los lectores de PF de los M_1 dispositivos de E/S 230 pueden cargarse a través del BMC de administración y los lectores de VF de los M_1 dispositivos de E/S 230 pueden cargarse a través de los procesadores de los N_1 hosts 210. Específicamente, cada host puede cargar un lector de VF de un dispositivo de E/S, virtualizada por el vEP que es enumerado por el host.

Sin embargo; esto no se limita a esta realización de la presente invención.

En el modo compartido de VF, los N_3 vEP 224 se configuran para virtualizar funciones físicas y virtuales de los M_1 dispositivos de E/S 230. En este caso, en las funciones virtualizadas por el vEP 224, solo la VF puede usarse como un puerto de servicio por parte del host 210, y la PF no puede usarse como tal por el host 210. Opcionalmente, el vEP 224 puede ser específicamente un espacio de configuración PCIe de una PF virtualizada. El espacio de configuración PCIe puede incluir un espacio de capacidad SR-IOV. Sin embargo, esto no se limita a esta realización de la presente invención. Por ejemplo, como se muestra en la FIG. 5, el sistema PCIe puede incluir: dos 'hosts: un host 0 y un host 1; un conmutador PCIe; un dispositivo terminal: un EP 0; y un dispositivo de conmutación. El dispositivo de

conmutación incluye: dos puertos PCIe aguas arriba respectivamente conectados al host 0 y al host 1, un puerto PCIe aguas abajo conectado al EP 0 mediante el uso del conmutador PCIe, dos espacios de configuración PCIe DP_CFG 0 y DP_CFG 1 correspondientes al puerto PCIe aguas abajo, un aparato de procesamiento interno, un módulo de mapeo, un vEP 0 y un vEP 1. El EP 0 presenta una función física (en inglés, Physical Function - PF): una PF 0 y dos funciones virtuales (en inglés, Virtual Functions - VF): una VF 0 y una VF 1 correspondientes a la PF 0. En este caso, el vEP 0 y el vEP 1 pueden ser específicamente un espacio de configuración PCIe de la PF 0 que transporta un espacio de capacidad SR-IOV, a fin de virtualizar respectivamente la VF 0 y la VF 1. Opcionalmente, en el modo compartido de VF, los lectores de PF de los M_1 dispositivos de E/S 230 pueden cargarse a través del BMC de administración. A diferencia del modo directo de VF, los procesadores de los N_1 hosts 210 además necesitan cargar lectores de PF correspondientes a los N_3 vEP 224 para habilitar una VF virtualizada por los N_3 vEP 224, de modo tal que los N_1 hosts 210 puedan usar normalmente la VF.

En el modo compartido de PF, los N_3 vEP 224 se pueden configurar para virtualizar funciones físicas de los M_1 dispositivos de E/S 230. En este caso, opcionalmente, los M_1 dispositivos de E/S 230 pueden presentar solo una PF y ninguna VF, es decir, la PF en los M_1 dispositivos de E/S no presentan ninguna VF correspondiente a la PF. Opcionalmente, el vEP 224 puede ser específicamente un espacio de configuración PCIe de una PF virtualizada por el vEP 224 y el espacio de configuración PCIe puede no incluir un espacio de capacidad SR-IOV. Por ejemplo, como se muestra en la FIG. 6, el sistema PCIe puede incluir: dos hosts: un host 0 y un host 1; un conmutador PCIe; un dispositivo terminal: un EP 0; y un dispositivo de conmutación. El dispositivo de conmutación incluye: dos puertos PCIe aguas arriba respectivamente conectados al host 0 y al host 1, un puerto PCIe aguas abajo conectado al EP 0 mediante el uso del conmutador PCIe, dos espacios de configuración PCIe DP_CFG 0 y DP_CFG 1 correspondientes al puerto PCIe aguas abajo, un aparato de procesamiento interno, un módulo de mapeo, un vEP 0 y un vEP 1. El EP 0 presenta dos PF: una PF 0 y una PF 1, y no presenta ninguna VF. En este caso, el vEP 0 puede configurarse para virtualizar la PF 0, de modo tal que la PF 0 es usada por el host 1. El vEP 1 puede configurarse para virtualizar la PF 1, de modo tal que la PF 1 sea usada por el host 2. Sin embargo, esto no se limita a esta realización de la presente invención. Opcionalmente, en el modo compartido de PF, los lectores de PF de los M_1 dispositivos de E/S 230 pueden cargarse a través de los procesadores de los N_1 hosts 210. Específicamente, un procesador de cada host 210 puede cargar un lector de un PF virtualizada por un vEP enumerada por el procesador.

En el modo de transmisión transparente EP, el dispositivo de conmutación 220 puede transmitir de manera transparente la información transmitida entre los M_1 dispositivos de E/S 230 y los N_1 hosts 210 que están conectados al dispositivo de conmutación 220. En este caso, los M_1 dispositivos de E/S 230 pueden ser un dispositivo de función única, es decir, presentar solo una PF y ninguna VF. Por ejemplo, como se muestra en la FIG. 7, un sistema PCIe puede incluir: dos hosts: un host 0 y un host 1; un conmutador PCIe; tres dispositivos terminales: un EP 0, un EP 1 y un EP 2; y un dispositivo de conmutación. El dispositivo de conmutación incluye: dos puertos PCIe aguas arriba respectivamente conectados al host 0 y el host 1, un puerto PCIe aguas abajo conectado al EP 0 y el EP 1 mediante el uso del conmutador PCIe, y un puerto PCIe aguas abajo directamente conectado al EP 2. Tanto el EP 0, como el EP 1 y el EP 2 solo presentan una PF: una PF 0. El dispositivo de conmutación puede no almacenar ningún espacio de configuración PCIe de cualquier puerto PCIe aguas abajo. En este caso, el host 0 puede usar las PF del EP 0 y el EP 1. El host 1 puede usar una PF del EP 2. Los módulos AU en el dispositivo de conmutación ejecutan una transmisión transparente y no pueden implementar una función de compartir un mismo EP entre hosts múltiples. Opcionalmente, en el modo de transmisión transparente de EP, no es necesario cargar un lector de VF y un lector de PF.

Opcionalmente, el sistema PCIe 200 puede operar siempre en un modo particular de los múltiples modos de operación anteriores, o puede conmutarse entre los múltiples modos de operación anteriores. Opcionalmente, el sistema PCIe 200 puede soportar incluso otro modo de operación. Esto no se limita a esta realización de la presente invención.

Debe entenderse que los ejemplos de la FIG. 4 a la FIG. 7 pretenden ayudar a un experto en la materia a entender mejor esta realización de la presente invención, pero no limitan el alcance de esta realización de la presente invención. Aparentemente, un experto en la materia puede llevar a cabo varias modificaciones o cambios en función de los ejemplos proporcionados de la FIG. 4 a la FIG. 7. Las modificaciones o los cambios también se enmarcan en el alcance de esta realización de la presente invención.

Además, debe entenderse que todos los ejemplos se describen mediante el uso de un ejemplo en el que el sistema PCIe incluye un dispositivo de conmutación. Opcionalmente, el sistema PCIe puede incluir múltiples dispositivos de conmutación. Como se muestra en la FIG. 8, el sistema PCIe 200 incluye dos dispositivos de conmutación 220: un dispositivo de conmutación 1 y un dispositivo de conmutación 2. Cada dispositivo de conmutación puede conectarse a dos hosts y a dos dispositivos de E/S. Específicamente, el dispositivo de conmutación 1 puede estar conectado por separado a un host 0, a un host 1, un ES 0 y un ES 1. El dispositivo de conmutación 2 se conecta por separado a un host 2, a un host 3, un ES 2 y un ES 3. De este modo, si un dispositivo de conmutación presenta una excepción, por ejemplo, el dispositivo de conmutación 1 conectado al ES 0 y el ES 1 presenta una excepción, un procesamiento de excepción, por ejemplo, una operación de restablecimiento o una interrupción de servicio relacionada pueden ejecutarse en función de una situación real. Si el dispositivo de conmutación 2 conectado al ES 2 y el ES 3 se ejecuta normalmente, el host 2 y el host 3 aún podrán ejecutarse de manera normal, mejorando así la confiabilidad y la estabilidad del sistema.

Además, debe entenderse que todas las realizaciones anteriores se describen mediante el uso de un ejemplo en el

que el dispositivo de conmutación 220 es específicamente un dispositivo físico. Opcionalmente, el dispositivo de conmutación 220 puede ser específicamente múltiples dispositivos físicos. Por ejemplo, como se muestra en la FIG. 8 y la FIG. 9, el dispositivo de conmutación 220 puede estar en una forma de chip AB, es decir, el dispositivo de conmutación 220 puede ser específicamente un dispositivo de conmutación 220a y un dispositivo de conmutación de E/S 220b. El dispositivo de conmutación de host 220a y el dispositivo de conmutación de E/S 220b puede conectarse mediante el uso de una interfaz Ethernet o una interfaz conmutada de otro tipo. En una realización opcional, el dispositivo de conmutación de host 220a puede incluir: N_2 puertos PCIe aguas arriba 221 y al menos una primera interfaz conmutada que se configura para conectar el dispositivo de conmutación de E/S 220b. El dispositivo de conmutación de E/S 220b puede incluir: al menos una segunda interfaz conmutada para conectarse al dispositivo de conmutación de host 220a, un aparato de procesamiento interno 223 y M_2 puertos PCIe aguas abajo 222. Opcionalmente, como se muestra en la FIG. 8, la al menos una interfaz conmutada y la al menos una segunda interfaz conmutada pueden conectarse directamente. De manera alternativa, como se muestra en la FIG. 9, la al menos una primera interfaz conmutada y la al menos una segunda interfaz conmutada pueden conectarse mediante el uso de un conmutador Ethernet o un dispositivo de conmutación interna de otro tipo. El dispositivo de conmutación interna puede proporcionar múltiples interfaces conmutadas configuradas para conectar el dispositivo de conmutación de host y el dispositivo de conmutación de E/S. En este caso, la al menos una primera interfaz conmutada puede configurarse para conectarse al dispositivo de conmutación interna, y la al menos una segunda interfaz conmutada puede configurarse para conectarse al dispositivo de conmutación interna. Esto no se limita a esta realización de la presente invención.

En otra realización opcional, el aparato de procesamiento interno 223 puede incluir un aparato de procesamiento (es decir, un primer aparato de procesamiento) ubicado en el dispositivo de conmutación del host 220a y un aparato de procesamiento (es decir, un segundo aparato de procesamiento) ubicado en el dispositivo de conmutación de E/S 220b. En este caso, el primer aparato de procesamiento interno puede conectarse de manera separada a los múltiples puertos PCIe aguas arriba y la al menos una primera interfaz conmutada mediante el uso de una línea de conexión interna del dispositivo de conmutación de host. El segundo aparato de procesamiento interno puede conectarse de manera separada al, al menos un, puerto PCIe aguas abajo y la al menos una segunda interfaz conmutada mediante el uso de una línea de conexión interna del dispositivo de conmutación de E/S. Sin embargo, esto no se limita a esta realización de la presente invención.

En este caso, opcionalmente, el primer aparato de procesamiento interno se configura para: recibir, mediante el uso de los múltiples puertos PCIe aguas arriba, un primer paquete de datos desde el al menos un host; procesar el primer paquete de datos para obtener un primer paquete de datos en el que se ejecuta el primer procesamiento; y transmitir al dispositivo de conmutación interna, mediante el uso de la al menos una primera interfaz conmutada, el primer paquete de datos en el que se ejecuta el primer procesamiento.

El segundo aparato de procesamiento interno se configura para: recibir, mediante el uso de la al menos una segunda interfaz conmutada, el primer paquete de datos en el que se ejecuta el primer procesamiento que es transmitido por el dispositivo de conmutación interna; procesar el primer paquete de datos en el que se ejecuta el primer procesamiento para obtener un primer paquete de datos en el que se ejecuta el segundo procesamiento; y transmitir al, al menos un, dispositivo de E/S, mediante el uso del al menos un puerto PCIe aguas abajo, el primer paquete de datos en el que se ejecuta el segundo procesamiento.

Opcionalmente, el segundo aparato de procesamiento interno se configura adicionalmente para: recibir, mediante el uso del al menos un puerto PCIe aguas abajo, un segundo paquete de datos desde el al menos un dispositivo de E/S; procesar el segundo paquete de datos para obtener un segundo paquete de datos en el que se ejecuta el primer procesamiento; y transmitir al dispositivo de conmutación interna, mediante el uso de la al menos una segunda interfaz conmutada, el segundo paquete de datos en el que se ejecuta el primer procesamiento.

El primer aparato de procesamiento interno se configura además para: recibir, mediante el uso de la al menos una primera interfaz conmutada, el segundo paquete de datos en el que se ejecuta el primer procesamiento que es transmitido por el dispositivo de conmutación interna; procesar el segundo paquete de datos en el que se ejecuta el primer procesamiento para obtener un segundo paquete de datos en el que se ejecuta el segundo procesamiento; y transmitir al, al menos un, host, mediante el uso de los múltiples puertos PCIe aguas arriba, el segundo paquete de datos en el que se ejecuta el segundo procesamiento.

Opcionalmente, el dispositivo de conmutación de E/S 220b puede incluir además los N_3 vEP 224, un módulo de mapeo 225 y los M_3 mEP 226. El dispositivo de conmutación del host 220a puede incluir además N_4 piezas de DP_CFG 227 y los N_3 vEP 224.

Como se muestra en la FIG. 8, un dispositivo de conmutación 1 puede incluir un dispositivo de conmutación de host 1 y un dispositivo de conmutación de E/S 1, y un dispositivo de conmutación 2 puede incluir un dispositivo de conmutación de host 2 y un dispositivo de conmutación de E/S 2. El host, los dispositivos de conmutación y los dispositivos de conmutación de E/S se conectan mediante el uso de una interfaz Ethernet (por ejemplo, una interfaz de control de acceso al medio (en inglés, Media Access Control - MAC)). En este caso, si el dispositivo de conmutación de host 1 conectado a un host 0 y un host 1 presenta una excepción, el dispositivo de conmutación 1 puede reemplazarse en función de una situación real, pero un host 2 y un host 3 aún pueden continuar accediendo a un

dispositivo de E/S y el servicio puede continuar siendo operado. Si el dispositivo de conmutación de E/S 1 conectado a un ES 0 y un ES 1 presenta una excepción, porque el dispositivo de conmutación de E/S 2 conectado a un ES 2 y un ES 3 es normal, no solo el host 2 y el host 3 pueden usar el ES 2 y el ES 3 mediante el uso del dispositivo de conmutación de host 2, un conmutador Ethernet y el dispositivo de conmutación de E/S 2, sino que también el host 0 y el host 1 pueden usar el ES 2 y el ES 3 mediante el uso de un dispositivo de conmutación de host 1, el conmutador Ethernet y el dispositivo de conmutación de E/S 2, de modo tal que los cuatro hosts pueden continuar operando el servicio, mejorando así de manera adicional la confiabilidad y la estabilidad del sistema.

Específicamente, en esta realización de la presente invención, el dispositivo de conmutación de host puede incluir los múltiples puertos PCIe aguas arriba configurados para conectarse al, al menos un, host. El dispositivo de conmutación de E/S puede incluir el aparato de procesamiento interno y el al menos un puerto PCIe aguas abajo configurado para conectarse al, al menos un, dispositivo de E/S.

En otra realización opcional, si el dispositivo de conmutación incluye al menos un mEP, un módulo de mapeo y al menos un vEP, el al menos un mEP, el módulo de mapeo y el al menos un vEP pueden implementarse solo en el dispositivo de conmutación de E/S. De manera alternativa, como se muestra en la FIG. 9, el dispositivo de conmutación de host puede incluir al menos un vEP y el al menos un vEP en el dispositivo de conmutación de host es el mismo que el al menos un vEP en el dispositivo de conmutación de E/S en una correspondencia de uno a uno. Sin embargo, esto no se limita a esta realización de la presente invención.

Debe entenderse que "el dispositivo de conmutación de host" y "el dispositivo de conmutación de E/S" son meramente nombres para distinguir diferentes dispositivos de conmutación. De manera alternativa, es posible hacer referencia al "dispositivo de conmutación de host" como el primer dispositivo de conmutación, y al "dispositivo de conmutación de E/S" como el segundo dispositivo de conmutación. Los nombres no deben constituir limitación alguna del alcance de protección de las realizaciones de la presente invención.

Por lo tanto, el sistema PCIe provisto en esta realización de la presente invención incluye al menos un host, un dispositivo de conmutación y el al menos un dispositivo de E/S. El dispositivo de conmutación incluye los múltiples puertos PCIe aguas arriba configurados para conectarse al, al menos un, host; el al menos un puerto PCIe aguas abajo configurado para conectarse al, al menos un, dispositivo de E/S; y el aparato de procesamiento interno. El aparato de procesamiento se conecta al, al menos un, puerto PCIe aguas abajo mediante el uso de una línea de conexión interna y se configura para: transmitir un paquete de lectura/escritura de configuración al, al menos un, puerto PCIe aguas abajo mediante el uso de la línea de conexión interna; recibir un paquete de respuesta de lectura/escritura de configuración transmitido mediante el al menos un puerto PCIe aguas abajo mediante el uso de una línea de conexión interna; y determinar, en función de una identificación de completador transportada en el paquete de respuesta de lectura/escritura de configuración, que el dispositivo de conmutación está conectado a un dispositivo de E/S cuya identificación es la identificación de completador. De esta manera, cuando un dispositivo de conmutación en el sistema PCIe falla, solo el al menos un host y el al menos un dispositivo de E/S que están conectados al dispositivo de conmutación resultan afectados, y otro dispositivo de conmutación, y un host y un dispositivo de E/S que están conectados al otro dispositivo de conmutación en el sistema no son afectados. Por lo tanto, en comparación con una CPU de administración externa en la técnica anterior, es posible mejorar tanto la estabilidad como la confiabilidad del sistema.

Además, una tecnología PCIe convencional en la técnica anterior se aplica principalmente a una placa, y presenta relativamente muchas limitaciones, por ejemplo, problemas tales como una cantidad limitada de puertos PCIe de un host, espacio limitado o expansión pobre. Un dispositivo de conmutación se dispone como dos dispositivos físicos, y un dominio PCIe es transportado en la Ethernet para su transmisión. De esta manera, es posible derribar las limitaciones de espacio y distancia, se puede configurar de manera flexible una topología PCIe, puede implementarse flexiblemente un dispositivo de E/S y se soluciona un problema de la técnica anterior de que el dominio PCIe no podía implementarse de manera remota.

Una realización de la presente invención proporciona adicionalmente un dispositivo de conmutación. El dispositivo de conmutación puede ser como se describe en la realización del sistema anterior. Por cuestiones de concisión, los detalles no se describen en esta invención nuevamente.

El sistema PCIe y el dispositivo de conmutación proporcionado en las realizaciones de la presente invención se describieron en detalle anteriormente en referencia a la FIG. 2 a la FIG. 9. A continuación, se describe en detalle un procedimiento de operación del sistema PCIe.

La FIG. 10 muestra un procedimiento 300 para inicializar un sistema PCIe según una realización de la presente invención. El procedimiento de inicialización 300 puede aplicarse al sistema PCIe en la realización anterior. S310: El BMC de administración envía una instrucción de inicialización al aparato de procesamiento interno en el dispositivo de conmutación.

Después de encenderse, el BMC de administración puede enviar la instrucción de inicialización al aparato de procesamiento. La instrucción de inicialización puede usarse para instruir al aparato de procesamiento que ejecute una operación, por ejemplo, que ejecute la enumeración y/o ejecute la configuración de inicialización en un módulo (o

un componente) en el dispositivo de conmutación, en relación con la inicialización del sistema.

S320: Después de recibir la instrucción de inicialización, el aparato de procesamiento del dispositivo de conmutación enumera, en función de la instrucción de inicialización, un dispositivo físico conectado al dispositivo de conmutación mediante el uso del al menos un puerto PCIe aguas abajo del dispositivo de conmutación, donde el dispositivo físico incluye el al menos un dispositivo de E/S.

En esta realización de la presente invención, el BMC de administración puede asignar una identificación, por ejemplo, asignar un identificador de dispositivo, al aparato de procesamiento interno en el dispositivo de conmutación. El identificador del dispositivo puede incluir tres campos: un número de bus, un número de dispositivo y un número de función. El campo del número de bus indica un número de un bus en el que se ubica el dispositivo, el campo del número de dispositivo indica un número del dispositivo y el campo del número de función indica un número de una función que el dispositivo posee. Habitualmente, también es posible hacer referencia al identificador de dispositivo como un número de función de dispositivo de bus (en inglés, Bus Device Function - BDF). Opcionalmente, el identificador de dispositivo asignado por el BMC de administración al aparato de procesamiento puede ser único en el sistema. Por ejemplo, en un sistema X86, el BMC de administración puede asignar 0.0.7 como el identificador de dispositivo del aparato de procesamiento. Sin embargo, esto no se limita a esta realización de la presente invención.

Al ejecutar una enumeración de dispositivos, el aparato de procesamiento puede detectar específicamente un dispositivo real conectado al dispositivo de conmutación mediante el uso del al menos un puerto PCIe aguas abajo. Específicamente, el aparato de procesamiento interno puede enviar múltiples TLP al, al menos un, puerto PCIe aguas abajo mediante el uso de una línea de conexión interna. El TLP puede ser específicamente un paquete de lectura/escritura de configuración, un valor de una RID transportada en el TLP puede establecerse en una BDF del aparato de procesamiento y una CID puede enumerarse en secuencia. Al recibir el paquete de lectura/escritura de configuración, un puerto PCIe aguas abajo puede reenviar la solicitud de lectura/escritura de configuración recibida a un dispositivo físico conectado al puerto PCIe aguas abajo. El dispositivo físico puede enviar, en función de la solicitud de lectura/escritura de configuración recibida, un paquete de respuesta de lectura/escritura de configuración, un paquete CPL o un paquete CPLD, al puerto PCIe aguas abajo conectado al dispositivo físico. El paquete de respuesta de lectura/escritura de configuración puede transportar la RID y la CID que están en el paquete de lectura/escritura de configuración correspondiente. De esta manera, al recibir un paquete de respuesta de lectura/escritura de configuración mediante el uso de un puerto PCIe aguas abajo, el aparato de procesamiento puede determinar que el dispositivo de conmutación se conecta a un dispositivo físico cuyo número de BDF es una CID en el paquete de respuesta de lectura/escritura de configuración, y puede obtener, además, información sobre el dispositivo físico que es transportada en el paquete de respuesta de lectura/escritura de configuración. Sin embargo, esto no se limita a esta realización de la presente invención.

S330: El dispositivo de conmutación envía, al BMC de administración, un resultado de la enumeración de dispositivos del aparato de procesamiento interno.

Específicamente, el aparato de procesamiento interno puede informar el resultado de la enumeración del dispositivo al BMC de administración mediante el uso de los múltiples puertos PCIe aguas arriba. El resultado de la enumeración puede incluir el dispositivo físico conectado al dispositivo de conmutación mediante el uso del al menos un puerto PCIe aguas abajo y una estructura de topología del dispositivo físico, o puede incluir un árbol de estructura PCIe. Esto no se limita a esta realización de la presente invención.

S340: Cada uno de los al menos un host enumera un dispositivo conectado a cada host.

Un procesador de cada uno de los al menos un host puede enumerar un dispositivo conectado al procesador. El dispositivo puede incluir un dispositivo virtual (es decir, un módulo funcional) y/o un dispositivo físico, por ejemplo, un puerto PCIe aguas arriba que es del dispositivo de conmutación y que se conecta al host, y al menos un vEP. La enumeración puede terminar en el vEP. Sin embargo; esto no se limita a esta realización de la presente invención.

Específicamente, un procesador de un host puede enviar un paquete de lectura de configuración, donde el paquete de lectura de configuración termina en el vEP del dispositivo de conmutación. Después de recibir el paquete de lectura de configuración, el dispositivo de conmutación puede regresar un paquete de respuesta de lectura de configuración al host mediante el uso de un puerto PCIe aguas arriba. Después de recibir el paquete de respuesta de lectura de configuración, el procesador del host puede enviar un paquete de escritura de configuración al vEP. Específicamente, el paquete de escritura de configuración puede usarse para acceder a un registro del vEP. Opcionalmente, cada paquete de escritura de configuración puede acceder al espacio de 4 KB del registro cada vez. El paquete de escritura de configuración aún termina en el vEP del dispositivo de conmutación. Después de recibir el paquete de escritura de configuración, el vEP del dispositivo de conmutación puede reenviar, mediante el uso de un puerto PCIe aguas abajo, el paquete de escritura de configuración a un dispositivo de E/S correspondiente al vEP (es decir que el vEP se configura para virtualizar una función del dispositivo de E/S). Después de recibir el paquete de escritura de configuración, el dispositivo de E/S correspondiente al vEP puede regresar un paquete de respuesta de escritura de configuración al dispositivo de conmutación mediante el uso de un puerto PCIe aguas abajo, donde el paquete de respuesta de escritura de configuración puede transportar un número de BDF del dispositivo de E/S. Opcionalmente, después de recibir el paquete de respuesta de escritura de configuración, el módulo de mapeo del dispositivo de

conmutación puede buscar una segunda tabla de mapeo en función una CID transportada en el paquete de respuesta de escritura de configuración, para obtener un número de BDF del vEP correspondiente al paquete de respuesta de escritura de configuración, y puede reemplazar a la CID en el paquete de respuesta de escritura de configuración con el número de BDF del vEP correspondiente a la CID. Después, el módulo de mapeo del dispositivo de conmutación puede determinar, en función de una RID transportada en el paquete de respuesta de escritura de configuración, para enviar el paquete de respuesta de escritura de configuración a un puerto PCIe aguas arriba del dispositivo de conmutación y regresar el paquete de respuesta de escritura de configuración al host mediante el uso del puerto PCIe aguas arriba.

Opcionalmente, el al menos un host y/o el dispositivo de conmutación pueden escribir, al vEP y/o el mEP, por medio de un mecanismo de autoaprendizaje, información (por ejemplo, el número de BDF) obtenida en el procedimiento anterior de la ejecución de la enumeración. Esto no se limita a esta realización de la presente invención.

S350; el al menos un host carga un lector de VF y un lector de PF del al menos un dispositivo de E/S.

Específicamente, el lector de VF puede cargarse a través de un procesador del host, y el lector de PF puede cargarse a través del procesador del host o el BMC de administración en función de los diferentes modos de operación.

Específicamente, en un modo directo de VF y en un modo compartido de VF compartido, el lector de PF del dispositivo de E/S puede cargarse a través del BMC de administración. Además, en el modo compartido de VF, el procesador del host puede cargar, además, un lector de PF del vEP, es decir que un procesador de cada host puede cargar un lector de PF de un vEP enumerado por el procesador del host. En un modo PF compartido, el lector de PF del dispositivo de E/S puede cargarse a través del procesador del host, es decir que un procesador de cada host puede cargar un lector de PF (o un lector de PF de una PF virtualizada por un vEP) de un dispositivo de E/S correspondiente a un vEP enumerado por el procesador del host. Sin embargo, esto no se limita a esta realización de la presente invención.

En otra realización opcional, después de la etapa S330, el procedimiento 300 puede incluir además la siguiente etapa.

El aparato de procesamiento interno en el dispositivo de conmutación configura un módulo funcional en el dispositivo de conmutación. Específicamente, el módulo funcional incluye: el al menos un vEP el módulo de mapeo, y el al menos un mEP. En este caso, el aparato de procesamiento puede establecer una relación de mapeo entre el al menos un mEP y el al menos un vEP, es decir, establecer una relación de mapeo entre una función del al menos un dispositivo de E/S y una función que se virtualiza a través del vEP y que puede ser usada por el al menos un host. El aparato de procesamiento puede almacenar, específicamente, en el módulo funcional del dispositivo de conmutación, información sobre un dispositivo físico conectado al dispositivo de conmutación. Por ejemplo, el aparato de procesamiento puede configurar la primera tabla de mapeo, la segunda tabla de mapeo y la tercera tabla de mapeo en la realización anterior, en función de la información del dispositivo obtenida. La información sobre el dispositivo físico se puede obtener en un procedimiento de ejecutar una enumeración de dispositivo mediante el aparato de procesamiento. Por ejemplo, la información, por ejemplo, un número de BDF, sobre cada dispositivo físico, se obtiene a partir de un paquete de respuesta de lectura/escritura de configuración. De manera alternativa, la información sobre el dispositivo físico puede ser entregada por el BMC de administración. De manera correspondiente y opcional, antes de que el aparato de procesamiento configure el módulo funcional en el dispositivo de conmutación, el procedimiento 300 puede incluir además: El BMC de administración envía información de administración de configuración al aparato de procesamiento. La información de administración de configuración puede incluir información, por ejemplo, un modo de operación actualmente usado (por ejemplo, cualquiera de los siguientes modos de operación), información de configuración (por ejemplo, una cantidad de espacios de configuración PCIe que corresponden a los puertos PCIe aguas abajo, una cantidad de vEP bajo cada espacio de configuración PCIe, una relación de conexión entre el puerto PCIe aguas abajo y el dispositivo de E/S o una dirección MAC de cada puerto PCIe) del puerto PCIe aguas abajo y el puerto PCIe aguas arriba, o información sobre un espacio de configuración PCIe y el dispositivo de E/S, necesaria, para configurar el módulo funcional del dispositivo de conmutación. Sin embargo, esto no se limita a esta realización de la presente invención.

Después de efectuar el procedimiento de inicialización, el sistema puede operar un servicio. En este caso, para un paquete de escritura de configuración en un paquete de configuración enviado por el host, el dispositivo de conmutación (por ejemplo, el módulo de mapeo del dispositivo de conmutación) necesita modificar una CID en el paquete de escritura de configuración. Para un paquete de lectura/escritura Mem entregado por el host, debe modificarse una dirección del paquete de lectura/escritura Mem (por ejemplo, el reemplazo de la dirección base se ejecuta en direcciones que corresponden a diferentes dominios, y una compensación permanece sin cambios). Para un paquete de compleción CPLD/CPL enviado por el host, solo se necesita modificar una CID. Para un paquete de lectura/escritura Mem enviado por el dispositivo de E/S, puede modificarse una RID del paquete de lectura/escritura del Mem, y no es necesario modificar ninguna dirección. Para un paquete de compleción CPLD/CPL enviado por el dispositivo de E/S, solo debe modificarse una CID del paquete de compleción CPLD/CPL. Sin embargo, esto no se limita a esta realización de la presente invención.

Debe entenderse que, para el procedimiento de inicialización, se debe hacer referencia a la descripción específica de la realización del sistema anterior. Por cuestiones de concisión, los detalles no se describen en esta invención nuevamente. Además, la descripción de cada realización de esta solicitud enfatiza las diferencias entre la realización

y otra realización, y las similitudes entre la realización y la otra realización pueden usarse para una referencia mutua.

Debe entenderse que los números de secuencia de los anteriores procesos no significan un orden de ejecución. El orden de ejecución de los procesos debe determinarse en función de las funciones y la lógica interna de los procesos, y no debe constituir ninguna limitación a un procedimiento de implementación de esta realización de la presente invención.

Además, una realización de la presente invención proporciona adicionalmente un sistema de conmutación. El sistema de conmutación incluye un primer dispositivo de conmutación y un segundo dispositivo de conmutación. El primer dispositivo de conmutación y el segundo dispositivo de conmutación están conectados mediante el uso de una red.

El primer dispositivo de conmutación puede incluir: múltiples puertos PCIe aguas arriba, configurados para conectarse a al menos un host.

El segundo dispositivo de conmutación puede incluir: al menos un puerto PCIe aguas abajo, configurado para conectarse a al menos un dispositivo.

Opcionalmente, el sistema de conmutación puede incluir además un dispositivo de conmutación de red ubicado en la red. El dispositivo de conmutación de red puede presentar múltiples interfaces conmutadas, configuradas para conectar el primer dispositivo de conmutación y el segundo dispositivo de conmutación.

En este caso, opcionalmente, el primer dispositivo de conmutación además incluye: al menos una primera interfaz conmutada, configurada para conectarse al dispositivo de conmutación de red. El segundo dispositivo de conmutación además incluye: al menos una segunda interfaz conmutada, configurada para conectarse al dispositivo de conmutación de red.

Opcionalmente, el dispositivo de conmutación de red puede ser específicamente un conmutador Ethernet mostrado en la FIG. 8. Sin embargo, esto no se limita a esta realización de la presente invención.

Opcionalmente, el primer dispositivo de conmutación puede ser específicamente el dispositivo de conmutación de host y el segundo dispositivo de conmutación puede ser específicamente el dispositivo de conmutación de E/S. Como se muestra en la FIG. 8 y la FIG. 9, el dispositivo de conmutación de host 220a y el dispositivo de conmutación de E/S 220b puede conectarse mediante el uso de una interfaz Ethernet o una interfaz conmutada de otro tipo. En una realización opcional, el dispositivo de conmutación 220a puede incluir N_2 puertos PCIe aguas arriba 221 y al menos una primera interfaz conmutada que se configura para conectarse al dispositivo de conmutación de E/S 220b. El dispositivo de conmutación de E/S 220b puede incluir: al menos una segunda interfaz conmutada configurada para conectarse al dispositivo de conmutación de host 220a, un aparato de procesamiento interno 223 y M_2 puertos PCIe aguas abajo 222.

Opcionalmente, como se muestra en la FIG. 9, la al menos una interfaz conmutada y la al menos una segunda interfaz conmutada pueden conectarse directamente. De manera alternativa, como se muestra en la FIG. 8, la al menos una primera interfaz conmutada y la al menos una segunda interfaz conmutada pueden conectarse mediante el uso de un conmutador Ethernet o un dispositivo de conmutación de red de otro tipo. El dispositivo de conmutación de red puede proporcionar múltiples interfaces conmutadas configuradas para conectar el dispositivo de conmutación de host y el dispositivo de conmutación de E/S. En este caso, la al menos una primera interfaz conmutada puede configurarse para conectarse al dispositivo de conmutación de red, y la al menos una segunda interfaz conmutada puede configurarse para conectarse al dispositivo de conmutación de red. Esto no se limita a esta realización de la presente invención.

En una realización opcional, el segundo dispositivo de conmutación puede configurarse para: recibir un primer paquete de datos transmitido por el primer dispositivo de conmutación mediante el uso de la red, procesar el primer paquete de datos para obtener un segundo paquete de datos que cumpla con el protocolo de PCIe y transmitir el segundo paquete de datos a un dispositivo de E/S meta del segundo paquete de datos.

Opcionalmente, el segundo dispositivo de conmutación puede configurarse para convertir el primer paquete de datos al segundo paquete de datos, cumpliendo con el protocolo PCIe. Opcionalmente, la conversión puede incluir un procesamiento de análisis, un procesamiento de conversión de formato de paquete, un procesamiento de conversión de secuencia de tiempo y similares. Sin embargo, esto no se limita a esta realización de la presente invención.

Opcionalmente, el primer paquete de datos puede ser específicamente un paquete de datos que cumple con un protocolo de red de la red. En una realización opcional, como se muestra en la FIG. 8, la red puede ser la Ethernet. En este caso, el segundo dispositivo de conmutación puede configurarse específicamente para convertir el primer paquete de datos, cumpliendo con un protocolo de Ethernet al segundo paquete de datos, cumpliendo con el protocolo PCIe. Sin embargo, esto no se limita a esta realización de la presente invención.

Opcionalmente, el segundo dispositivo de conmutación puede configurarse adicionalmente para: determinar, en función de la información adquirida en un paquete de datos recibido o en función de un puerto configurado para recibir el paquete de datos, un dispositivo meta del paquete de datos y enviar el paquete de datos en función del dispositivo

meta del paquete de datos. Por ejemplo, el segundo dispositivo de conmutación puede determinar además, en función de la información transportada en el primer paquete de datos, un dispositivo de E/S meta del primer paquete de datos y enviar, al dispositivo de E/S meta del primer paquete de datos, el segundo paquete de datos que se obtiene mediante el procesamiento del primer paquete de datos. Sin embargo, esto no se limita a esta realización de la presente invención.

Opcionalmente, el primer dispositivo de conmutación puede recibir, mediante el uso de la al menos una primera interfaz conmutada, un paquete de datos de enlace ascendente transmitido por el dispositivo de conmutación de red, y puede transmitir un paquete de datos de enlace descendente al dispositivo de conmutación de red usando la al menos una primera interfaz conmutada. Por ejemplo, el primer dispositivo de conmutación puede transmitir el primer paquete de datos al segundo dispositivo de conmutación mediante el uso de la al menos una primera interfaz conmutada. Sin embargo, esto no se limita a esta realización de la presente invención.

Opcionalmente, el segundo dispositivo de conmutación puede recibir, mediante el uso de la al menos una segunda interfaz conmutada, un paquete de datos de enlace descendente transmitido por el dispositivo de conmutación de red, y puede transmitir un paquete de datos de enlace ascendente al dispositivo de conmutación de red usando la al menos una segunda interfaz conmutada. Por ejemplo, el segundo dispositivo de conmutación puede recibir, usando la al menos una segunda interfaz conmutada, el primer paquete de datos transmitido por el primer dispositivo de conmutación. Sin embargo, esto no se limita a esta realización de la presente invención.

Opcionalmente, el segundo dispositivo de conmutación incluye además: un primer aparato de procesamiento interno, conectado al, al menos un, puerto PCIe aguas abajo mediante el uso de una línea de conexión interna del segundo dispositivo de conmutación. Opcionalmente, el hecho de que el segundo dispositivo de conmutación se configura para procesar el primer paquete de datos para obtener un segundo paquete de datos, cumpliendo con un protocolo, PCIe incluye: el primer aparato de procesamiento interno se configura para procesar el primer paquete de datos, para obtener el segundo paquete de datos, cumpliendo con el protocolo PCIe.

Opcionalmente, el segundo dispositivo de conmutación puede configurarse además para procesar un paquete de datos de enlace ascendente.

Opcionalmente, un primer aparato de procesamiento en el segundo dispositivo de conmutación se configura para: recibir un tercer paquete de datos desde el al menos un dispositivo de E/S, procesar el tercer paquete de datos para obtener un cuarto paquete de datos, cumpliendo con el protocolo de red de la red, y transmitir el cuarto paquete de datos al primer dispositivo de conmutación usando la red.

Opcionalmente, el primer aparato de procesamiento puede ejecutar el procesamiento de conversión en un primer paquete de datos, cumpliendo con el protocolo PCIe, para obtener un cuarto paquete de datos que cumple con el protocolo de red de la red.

Opcionalmente, el primer dispositivo de conmutación puede procesar un paquete datos de enlace ascendente y/o de enlace descendente recibido. En otra realización opcional, el aparato de procesamiento interno 223 puede incluir un aparato de procesamiento (es decir, un primer aparato de procesamiento) ubicado en el dispositivo de conmutación del host 220a y un aparato de procesamiento (es decir, un segundo aparato de procesamiento) ubicado en el dispositivo de conmutación de E/S 220b. En este caso, el primer aparato de procesamiento interno puede conectarse de manera separada a los múltiples puertos PCIe aguas arriba y la al menos una primera interfaz conmutada mediante el uso de una línea de conexión interna del dispositivo de conmutación de host. El segundo aparato de procesamiento interno puede conectarse de manera separada al, al menos un, puerto PCIe aguas abajo y la al menos una segunda interfaz conmutada mediante el uso de una línea de conexión interna del dispositivo de conmutación de E/S. Sin embargo, esto no se limita a esta realización de la presente invención.

Opcionalmente, el primer dispositivo de conmutación incluye además: un segundo aparato de procesamiento interno, conectado a los múltiples puertos PCIe aguas arriba mediante el uso de una línea de conexión interna del primer dispositivo de conmutación.

Opcionalmente, el segundo aparato de procesamiento interno puede configurarse para procesar un paquete de datos recibido de enlace ascendente/descendente.

En una realización opcional, el segundo aparato de procesamiento interno se configura para: recibir un quinto paquete de datos desde el al menos un host, procesar el quinto paquete de datos para obtener el primer paquete de datos cumpliendo con el protocolo de red de la red y transmitir el primer paquete de datos al segundo dispositivo de conmutación por medio del uso de la red.

Opcionalmente, el quinto paquete de datos puede cumplir con el protocolo PCIe. El segundo aparato de procesamiento puede configurarse específicamente para convertir el paquete de datos al primer paquete de datos, donde el primer paquete de datos cumple con el protocolo de red de la red.

Opcionalmente; el segundo aparato de procesamiento puede determinar además, en función de la información transportada en el primer paquete de datos y/o el puerto PCIe aguas arriba que recibe el quinto paquete de datos, una

primera interfaz conmutada correspondiente al quinto paquete de datos y transmitir el primer paquete de datos al segundo dispositivo de conmutación mediante el uso de la primera interfaz conmutada correspondiente. Sin embargo, esto no se limita a esta realización de la presente invención.

En otra realización opcional, el segundo aparato de procesamiento interno se configura para: recibir, mediante el uso de la red, el cuarto paquete de datos transmitido por el primer dispositivo de conmutación, procesar el cuarto paquete de datos para obtener el sexto paquete de datos cumpliendo con el protocolo PCIe y transmitir el sexto paquete de datos al host meta del sexto paquete de datos.

Opcionalmente, el segundo aparato de procesamiento puede recibir, mediante el uso de la red, un cuarto paquete de datos transmitido por el segundo dispositivo de conmutación, donde el cuarto paquete de datos puede ser un paquete de datos que cumple con un protocolo de red de la red. El segundo aparato de procesamiento puede ejecutar el procesamiento de conversión en el cuarto paquete de datos para obtener el sexto paquete de datos cumpliendo con el protocolo PCIe.

Opcionalmente, el segundo aparato de procesamiento puede configurarse para: determinar, en función de la información transportada en un paquete de datos recibido o en función de un puerto configurado para recibir el paquete de datos, un dispositivo meta del paquete de datos y enviar el paquete de datos en función del dispositivo meta del paquete de datos. Por ejemplo, el segundo aparato de procesamiento puede determinar, en función de la información transportada en el cuarto paquete de datos, un host meta del cuarto paquete de datos y transmitir, al host meta del cuarto paquete de datos, el sexto paquete de datos que se obtiene mediante el procesamiento del cuarto paquete de datos. Sin embargo, esto no se limita a esta realización de la presente invención.

Opcionalmente, el primer dispositivo de conmutación además incluye: al menos un módulo de dispositivo terminal virtual, conectado a múltiples puertos PCIe aguas arriba, y configurado para virtualizar una función del al menos un dispositivo de E/S, de modo tal que la función se usada por el al menos un host.

El segundo dispositivo de conmutación además incluye:

al menos un módulo de dispositivo terminal de espejo, conectado al, al menos un, puerto PCIe aguas abajo, y configurado para almacenar el contenido de configuración PCIe del al menos un dispositivo de E/S; y

un módulo de mapeo, configurado para implementar el mapeo entre un dominio PCIe correspondiente al, al menos un, host y un dominio PCIe correspondiente al, al menos un, dispositivo de E/S.

Opcionalmente, ninguno de los al menos un puerto PCIe aguas abajo presenta espacio de configuración PCIe alguno, y cada uno de los múltiples puertos PCIe aguas arriba presenta un espacio de configuración PCIe.

Opcionalmente, el primer dispositivo de conmutación además incluye al menos un espacio de configuración PCIe correspondiente al, al menos un, puerto PCIe aguas abajo. El al menos un módulo de dispositivo terminal virtual se conecta específicamente a los múltiples puertos PCIe aguas arriba mediante el uso del al menos un espacio de configuración PCIe correspondiente al, al menos un, puerto PCIe aguas abajo.

Opcionalmente, el al menos un espacio de configuración PCIe correspondiente al, al menos un, puerto PCIe aguas abajo pueden ser específicamente las N_4 piezas del DP_CFG 227. Sin embargo, esto no se limita a esta realización de la presente invención. Opcionalmente, cada uno del al menos un puerto PCIe aguas abajo puede corresponder a cero, uno o más espacios de configuración PCIe. Esto no se limita a esta realización de la presente invención.

Opcionalmente, el módulo de mapeo almacena:

una primera tabla de mapeo, usada para almacenar una relación de mapeo desde una identificación en el dominio PCIe correspondiente al, al menos un, host hasta una identificación en el dominio PCIe correspondiente al, al menos un, dispositivo de E/S; y

una segunda tabla de mapeo, usada para almacenar una relación de mapeo desde la identificación en el dominio PCIe correspondiente al, al menos un, dispositivo de E/S hasta la identificación en el dominio PCIe correspondiente al, al menos un, host.

Opcionalmente, un primer módulo de dispositivo terminal de espejo en el al menos un módulo de dispositivo terminal de espejo es específicamente una tercera tabla de mapeo, y la tercera tabla de mapeo se usa para almacenar un registro de dirección base (BAR, por sus siglas en inglés) y un tamaño de BAR de una función virtual de un primer dispositivo de E/S en el al menos un dispositivo de E/S, donde el primer módulo de dispositivo terminal de espejo se configura para almacenar el contenido de la configuración PCIe del primer dispositivo de E/S.

Opcionalmente, un primer módulo de dispositivo terminal virtual en el al menos un módulo de dispositivo terminal virtual se configura específicamente para virtualizar una función física de un segundo dispositivo de E/S en el al menos un dispositivo de E/S, de modo tal que la función física sea usada por un primer host en el al menos un host, donde un lector de función física del segundo dispositivo de E/S es cargado por un procesador del primer host.

Opcionalmente, un segundo módulo de dispositivo terminal virtual en el al menos un módulo de dispositivo terminal virtual se configura específicamente para virtualizar una función virtual de un tercer dispositivo de E/S en el al menos un dispositivo de E/S, de modo tal que la función virtual sea usada por un segundo host en el al menos un host, donde un lector de función física del tercer dispositivo de E/S es cargado por un mando de administración de la placa madre BMC de administración en el al menos un host.

Opcionalmente, un tercer módulo de dispositivo terminal virtual en el al menos un módulo de dispositivo terminal virtual se configura específicamente para virtualizar una función física y una función virtual de un cuarto dispositivo de E/S en el al menos un dispositivo de E/S, de modo tal que la función física y la función virtual sean usadas por un tercer host en el al menos un host, donde un lector de función física del cuarto dispositivo de E/S es cargado por el BMC de administración en el al menos un host, y un lector de función física del tercer módulo de dispositivo terminal virtual es cargado por un procesador del tercer host.

Debe entenderse que el al menos un módulo de dispositivo terminal virtual pueden ser específicamente los N_3 vEP 224, el al menos un módulo de dispositivo terminal de espejo pueden ser específicamente los M_3 mEP 226 y el módulo de mapeo puede ser específicamente el módulo de mapeo 225. Por lo tanto, para las implementaciones específicas del al menos un módulo de dispositivo terminal virtual, el al menos un módulo de dispositivo terminal de espejo y el módulo de mapeo, consulte la descripción anterior. Por cuestiones de concisión, los detalles no se describen en esta invención nuevamente.

Opcionalmente, el primer dispositivo de conmutación puede ser específicamente el dispositivo de conmutación de host y el segundo dispositivo de conmutación puede ser específicamente el dispositivo de conmutación de E/S. Para las implementaciones específicas del primer dispositivo de conmutación y el segundo dispositivo de conmutación, consulte la descripción anterior. Por cuestiones de concisión, los detalles no se describen en esta invención nuevamente. Debe entenderse que, "el dispositivo de conmutación de host", "el dispositivo de conmutación de E/S", “el primer dispositivo de conmutación” y "el segundo dispositivo de conmutación" en esta invención son meramente nombres para distinguir los diferentes dispositivos de conmutación. Los nombres no deben constituir limitación alguna del alcance de protección de las realizaciones de la presente invención.

Una realización de la presente invención además proporciona un sistema PCIe, incluyendo: al menos un host, el sistema de conmutación en la realización anterior y al menos un dispositivo de E/S.

Debe entenderse que, en las realizaciones de la presente invención, el término "múltiple" puede indicar al menos dos o tres, y que el término "y/o" describe solo una relación de asociación para describir objetos asociados y representa que pueden existir tres relaciones. Por ejemplo, A y/o B pueden representar los siguientes tres casos: Solo A existe, ambos A y B existen y solo B existe. Además, el carácter "/" en esta memoria descriptiva indica una relación de "o" entre los objetos asociados.

Una persona con conocimientos ordinarios en la materia puede ser consciente de que, en combinación con los ejemplos descritos en las realizaciones descritas en esta especificación, las etapas y las unidades de los procedimientos pueden implementarse mediante hardware electrónico, software de ordenador o una combinación de los mismos. Para describir claramente la intercambiabilidad entre el hardware y el software, lo anterior ha descrito de manera general las etapas y composiciones de cada realización en función de las funciones. El hecho de que las funciones sean ejecutadas por hardware o software depende de las aplicaciones particulares y las condiciones de restricción de diseño de las soluciones técnicas. Una persona con un conocimiento ordinario en la materia puede usar procedimientos diferentes para implementar las funciones descritas para cada aplicación particular, pero no debe considerarse que la implementación va más allá del alcance de esta solicitud.

Una persona experta en la materia debe entender claramente que, para los fines de una descripción conveniente y breve, para los procesos de trabajo detallados del sistema, el aparato y la unidad anteriores, puede hacerse referencia a procesos correspondientes en las realizaciones del procedimiento anteriores y los detalles no se describen nuevamente en esta invención.

En las varias realizaciones proporcionadas en esta solicitud, debe entenderse que el sistema, el aparato y el procedimiento descritos pueden implementarse de otras maneras. Por ejemplo, la realización del aparato descrito es meramente un ejemplo. Por ejemplo, la división de unidad es meramente una división de función lógica y puede ser otra división en la implementación real. Por ejemplo, pueden combinarse múltiples unidades o componentes o bien estas pueden integrarse en otro sistema, o algunas características pueden ignorarse o no ejecutarse. Además, el o los acoplamientos mutuos mostrados o analizados, los acoplamientos directos o las conexiones de comunicación pueden implementarse a través de algunas interfaces, acoplamientos indirectos o conexiones de comunicación entre los aparatos o unidades, o conexiones eléctricas, mecánicas o conexiones de otras formas.

Las unidades descritas como partes separadas pueden ser o no partes separadas, y las partes mostradas como unidades puede ser o no unidades físicas, pueden ubicarse en una posición o estar distribuidas en múltiples unidades de red. Una parte o todas las unidades pueden seleccionarse en función de las necesidades reales, a fin de lograr las metas de las soluciones en las realizaciones de la presente invención.

Además, las unidades funcionales en las realizaciones de esta solicitud pueden integrarse en una unidad de

procesamiento, o cada una de las unidades puede existir físicamente sola, o dos o más unidades pueden integrarse en una unidad. La unidad integrada puede implementarse en la forma de hardware, o puede implementarse en una forma de una unidad funcional de software.

Cuando la unidad integrada se implementa en la forma de una unidad funcional de software y se vende o usa como un producto independiente, la unidad integrada puede almacenarse en un medio de almacenamiento legible por ordenador. En base a dicho entendimiento, las soluciones técnicas de la presente solicitud esencialmente, o la parte que contribuye a la técnica anterior, o todo o parte de las soluciones técnicas puede implementarse en la forma de un producto de software. El producto de software de ordenador se almacena en un medio de almacenamiento e incluye varias instrucciones para instruir a un dispositivo de ordenador (que puede ser un ordenador personal, un servidor, un dispositivo de red o similares) para que efectúe todas o parte de las etapas de los procedimientos descritos en las realizaciones de esta solicitud. El medio de almacenamiento anterior incluye cualquier medio que pueda almacenar un código de programa, como una unidad Flash USB, un disco duro extraíble, una memoria solo de lectura (en inglés, Read-Only Memory - ROM), una memoria de acceso aleatorio (en inglés, Random Access Memory - RAM), un disco magnético o un disco óptico.

Las descripciones anteriores son meramente implementaciones específicas de esta solicitud, pero no pretenden limitar el alcance de protección de esta solicitud. Cualquier modificación o reemplazo que deduzca de inmediato un experto en la materia, dentro del alcance técnico descrito en esta solicitud, debe encuadrarse dentro del alcance de protección de esta solicitud. Por lo tanto, el alcance de protección de la presente solicitud debe estar sujeto al alcance de protección de las reivindicaciones.

REIVINDICACIONES

1. Un dispositivo de conmutación (220) que comprende:

múltiples puertos de Interconexión de Componentes Periféricos rápidos, PCIe, aguas arriba (221), configurados para conectar el al menos un host (210);

al menos un puerto PCIe aguas abajo (222), configurado para conectarse a al menos un dispositivo de entrada/salida, E/S, (230); y

un aparato de procesamiento interno (223) conectado al al menos un puerto PCIe aguas abajo (221) mediante el uso de una línea de conexión interna del dispositivo de conmutación (220), caracterizado por que el aparato de procesamiento interno (223) está configurado para:

transmitir un paquete de lectura/escritura de configuración para el al menos un puerto PCIe aguas abajo (222) mediante el uso de una línea de conexión interna;

recibir un paquete de respuesta de lectura/escritura de configuración por parte del al menos un puerto PCIe aguas abajo (222) mediante el uso de la línea de conexión interna, donde el paquete de respuesta de lectura/escritura de configuración transporta una identificación de completador; y

determinar, en función de la identificación de completador transportada en el paquete de respuesta de lectura/escritura de configuración, que el dispositivo de conmutación (220) está conectado a un dispositivo de E/S cuya identificación es la identificación de completador.

2. El dispositivo de conmutación (220) según la reivindicación 1, donde el dispositivo de conmutación (220) comprende además:

al menos un módulo de dispositivo terminal de espejo (226), conectado al, al menos un, puerto PCIe aguas abajo (222), y configurado para almacenar el contenido de configuración PCIe del al menos un dispositivo de E/S (230) conectado al, al menos un, puerto PCIe aguas abajo (222);

al menos un módulo de dispositivo terminal virtual (224), conectado a los múltiples puertos PCIe aguas arriba (221) y configurado para virtualizar una función del al menos un dispositivo de E/S (230) conectado al, al menos un, puerto PCIe aguas abajo (222), de modo que la función sea usada por el al menos un host (210) conectado a los múltiples puertos PCIe aguas arriba (221); y

un módulo de mapeo (225), conectado por separado al, al menos un, módulo de dispositivo terminal de espejo (226) y el al menos un módulo de dispositivo terminal virtual (224), y configurado para implementar el mapeo entre un dominio PCIe correspondiente al, al menos un, host (210) y el dominio PCIe correspondiente al, al menos un, dispositivo de E/S (230).

3. El dispositivo de conmutación (220) según la reivindicación 2, en donde ninguno del al menos un puerto PCIe aguas abajo (222) presenta espacio de configuración PCIe alguno y cada uno de los múltiples puertos PCIe aguas arriba (221) tiene espacio de configuración PCIe.

4. El dispositivo de conmutación (220) según la reivindicación 3, en donde el dispositivo de conmutación (220) comprende además al menos un espacio de configuración PCIe que corresponde a cada uno de los al menos un puerto PCIe aguas abajo (222), y el al menos un módulo de dispositivo terminal virtual (224) se conecta específicamente a los múltiples puertos PCIe aguas arriba (221) mediante el uso del al menos un espacio de configuración PCIe correspondiente al, al menos un, puerto PCIe aguas abajo (222).

5. El dispositivo de conmutación (220) según una cualquiera de las reivindicaciones 2 a 4, en donde el módulo de mapeo (225) almacena:

una primera tabla de mapeo, usada para almacenar una relación de mapeo desde una identificación en el dominio PCIe correspondiente al, al menos un, host (210) hasta una identificación en el dominio PCIe correspondiente al, al menos un, dispositivo de E/S (230); y

una segunda tabla de mapeo, usada para almacenar una relación de mapeo desde la identificación en el dominio PCIe correspondiente al, al menos un, dispositivo de E/S (230) hasta la identificación en el dominio PCIe correspondiente al, al menos un, host (210).

6. El dispositivo de conmutación (220) según una cualquiera de las reivindicaciones 2 a 5, en donde un primer módulo de dispositivo terminal de espejo en el al menos un módulo de dispositivo terminal de espejo (224) es específicamente una tercera tabla de mapeo, y la tercera tabla de mapeo se usa para almacenar un registro de dirección base BAR y un tamaño de BAR de una función virtual de un primer dispositivo de E/S (230) en el al menos un dispositivo de E/S (230), en donde el primer módulo de dispositivo terminal de espejo se configura para almacenar el contenido de la configuración PCIe del primer dispositivo de E/S (230).

7. El dispositivo de conmutación (220) según una cualquiera de las reivindicaciones 2 a 6, en donde un primer módulo de dispositivo terminal virtual en el al menos un módulo de dispositivo terminal virtual (224) se configura específicamente para virtualizar una función física de un segundo dispositivo de E/S (230) en el al menos un dispositivo de E/S (230), de modo que la función física sea usada por un primer host (210) en el al menos un host (210), en donde un lector de función física del segundo dispositivo de E/S (230) es cargado por un procesador del primer host (210).

8. El dispositivo de conmutación (220) según una cualquiera de las reivindicaciones 2 a 7, en donde un segundo módulo de dispositivo terminal virtual en el al menos un módulo de dispositivo terminal virtual (224) se configura específicamente para virtualizar una función virtual de un tercer dispositivo de E/S (230) en el al menos un dispositivo de E/S (230), de modo que la función virtual sea usada por un segundo host (210) en el al menos un host (210), en donde un lector de función física del tercer dispositivo de E/S (230) es cargado por un mando de administración de la placa madre BMC de administración en el al menos un host (210).

9. El dispositivo de conmutación (220) según una cualquiera de las reivindicaciones 2 a 8, en donde un tercer módulo de dispositivo terminal virtual en el al menos un módulo de dispositivo terminal virtual (224) se configura específicamente para virtualizar una función física y una función virtual de un cuarto dispositivo de E/S (230) en el al menos un dispositivo de E/S (230), de modo que la función física y la función virtual sean usadas por un tercer host (210) en el al menos un host (210), en donde un lector de función física del cuarto dispositivo de E/S (230) es cargado por el BMC de administración en el al menos un host (210), y un lector de función física del tercer módulo de dispositivo terminal virtual es cargado por un procesador del tercer host (210).

10. El dispositivo de conmutación (220) según una cualquiera de las reivindicaciones 2 a 9, en donde el dispositivo de conmutación (220) es específicamente un dispositivo de conmutación de host (220a) y un dispositivo de conmutación de E/S (220b), y el dispositivo de conmutación de host (220a) y el dispositivo de conmutación de E/S (220b) están conectados mediante el uso de una interfaz Ethernet, en donde

el dispositivo de conmutación de host (220a) comprende los múltiples puertos PCIe aguas arriba (221); y

el dispositivo de conmutación de E/S (220b) comprende el aparato de procesamiento interno (223), el al menos un módulo de dispositivo terminal virtual (224), el módulo de mapeo (225), el al menos un módulo de dispositivo terminal de espejo (226) y el al menos un puerto PCIe aguas abajo (222).

11. El dispositivo de conmutación (220) según una cualquiera de las reivindicaciones 1 a 10, en donde el aparato de procesamiento interno (223) se configura para:

recibir una instrucción de inicialización enviada por el BMC de administración en el al menos un host (210); y

transmitir, en función de la instrucción de inicialización, un paquete de lectura/escritura de configuración al, al menos un, puerto PCIe aguas abajo mediante el uso de la línea de conexión interna; y

el aparato de procesamiento interno (223) además se configura para: después de determinar que un dispositivo de E/S (230) ha sido conectado al dispositivo de conmutación (220), informar, al BMC de administración, la información sobre el dispositivo de E/S (230) conectado al dispositivo de conmutación (220), en donde la información comprende la información de identificación.

12. El dispositivo de conmutación (220) según una cualquiera de las reivindicaciones 1 a 11, en donde el aparato de procesamiento interno (223) se configura adicionalmente para:

recibir la información de administración de configuración enviada por el BMC de administración en el al menos un host (210); y

configurar los múltiples puertos PCIe aguas arriba (221) y el al menos un puerto PCIe aguas abajo (222) en función de la información de administración de configuración.

13. El dispositivo de conmutación (220) según una cualquiera de las reivindicaciones 1 a 12, en donde el aparato de procesamiento interno (223) se configura adicionalmente para procesar un evento de excepción y un evento de intercambio en caliente.

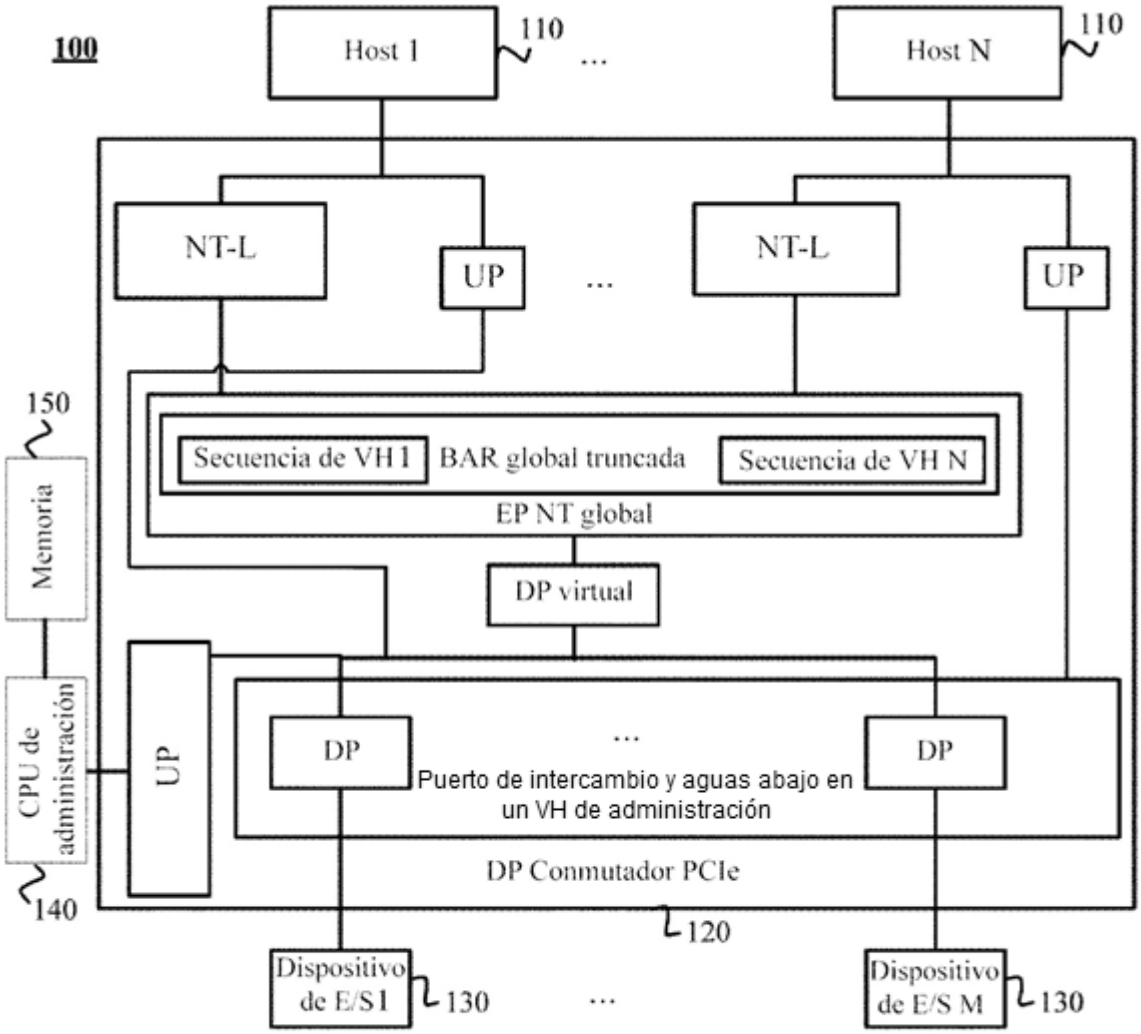
100
Host 1
110
...
Host N
110
NT-L
UP
...
NT-L
UP
150
Memoria
Secuencia de VH1
BAR global truncada
Secuencia de VH N
EP NT global
DP virtual
CPU de administración
UP
DP
...
DP
Puerto de intercambio y aguas abajo en un VH de administración
DP Conmutador PCIe
140
120
Dispositivo de E/S1
130
...
Dispositivo de E/S M
130

FIG. 1

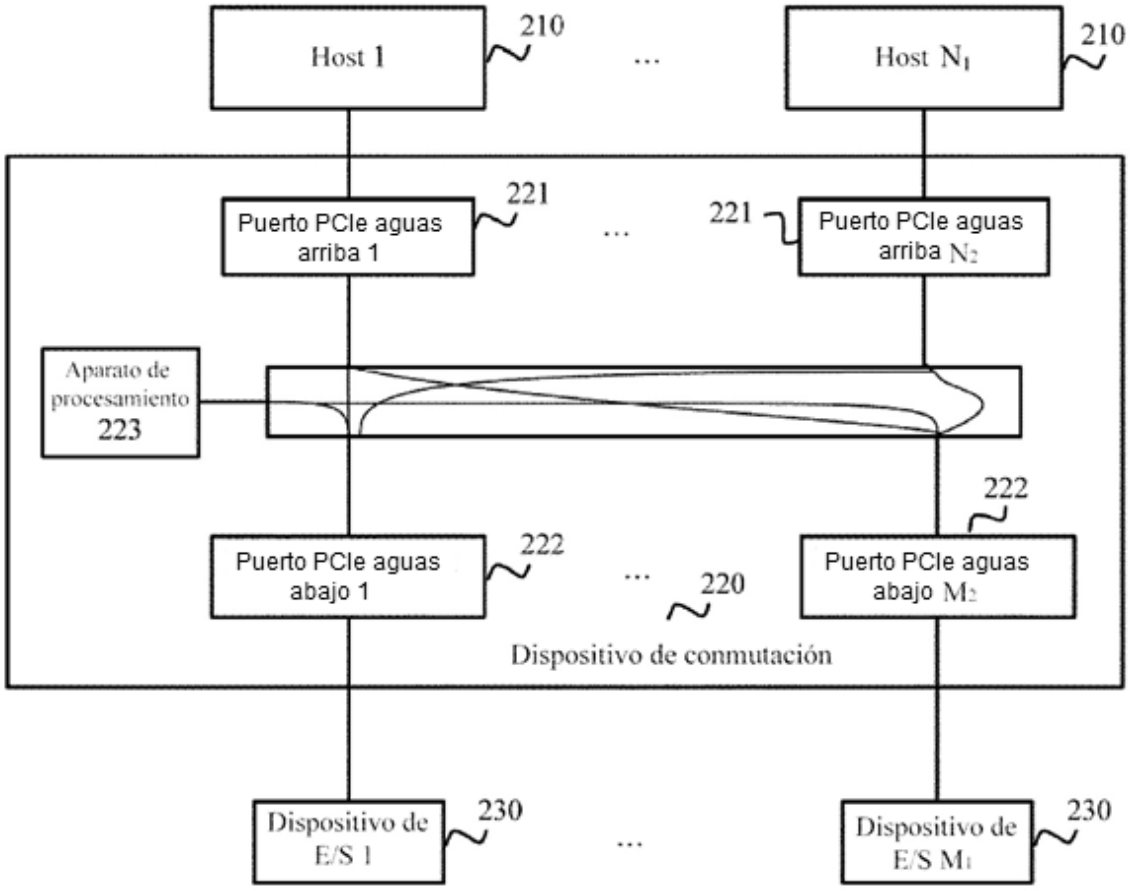
Host 1
210
...
Host N_1
210
Puerto PCIe aguas arriba 1
221
...
221
Puerto PCIe aguas arriba N_2
Aparato de procesamiento 223
222
Puerto PCIe aguas abajo 1
222
...
220
Puerto PCIe aguas abajo M_2
Dispositivo de conmutación
Dispositivo de E/S 1
230
...
Dispositivo de E/S M_1
230

FIG. 2

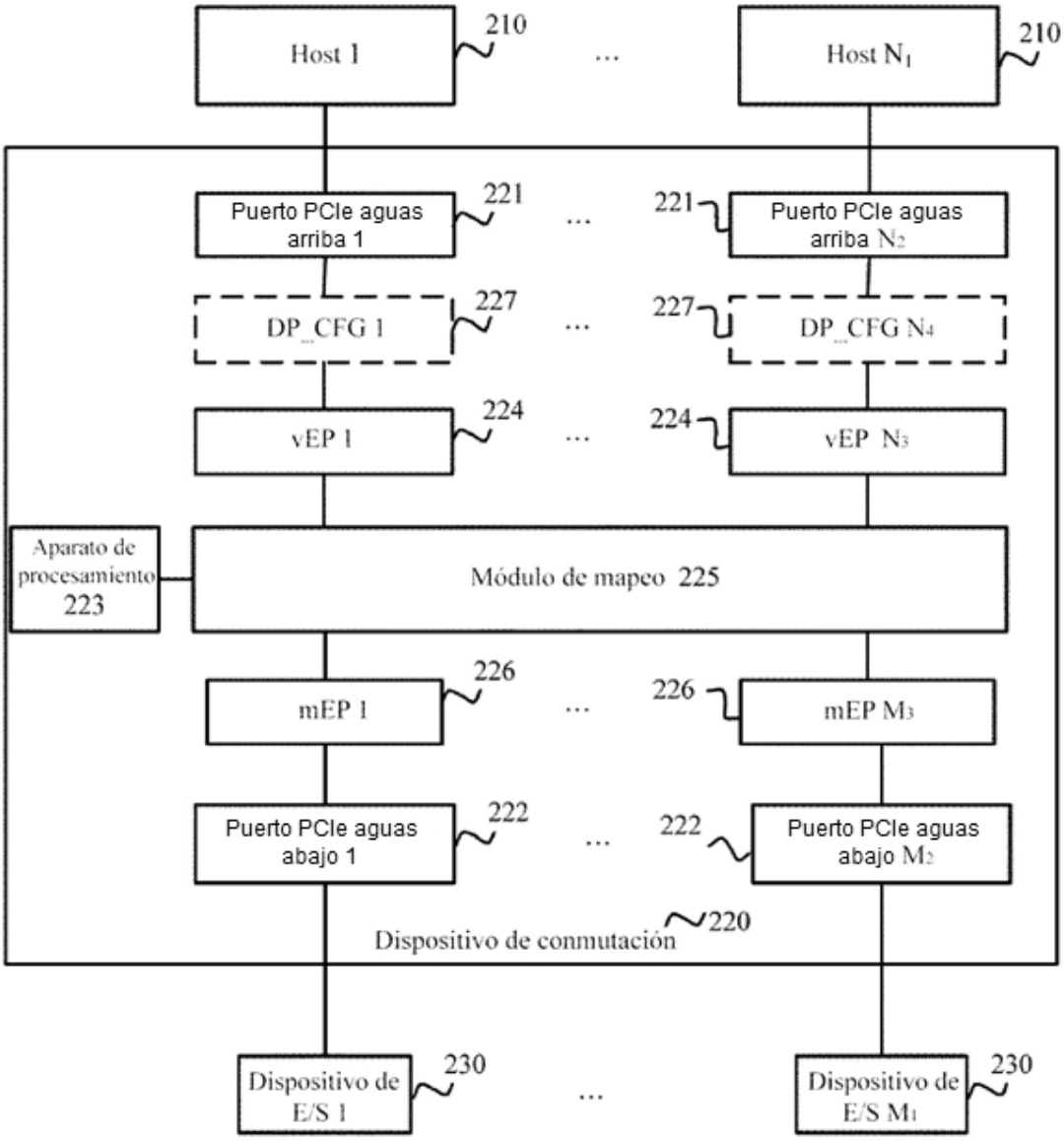
Host 1
210
Host N1
210
Puerto PCIe aguas arriba 1
221
Puerto PCIe aguas arriba N2
221
DP_CFG 1
227
DP_CFG N4
227
vEP 1
224
vEP N3
224
Aparato de procesamiento 223
Módulo de mapeo 225
mEP 1
226
mEP M3
226
Puerto PCIe aguas abajo 1
222
Puerto PCIe aguas abajo M2
222
Dispositivo de conmutación
220
Dispositivo de E/S 1
230
Dispositivo de E/S M1
230

FIG. 3

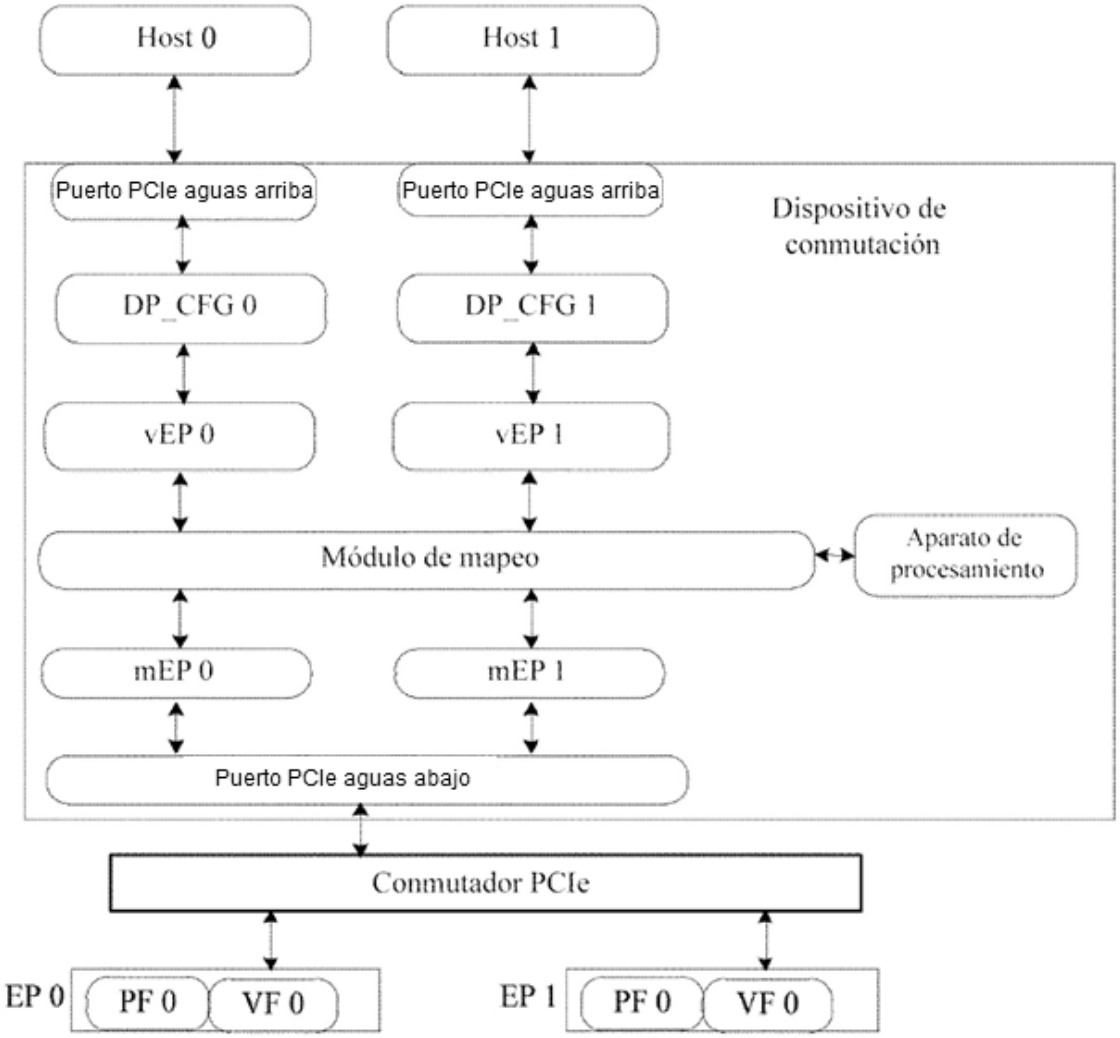
Host 0
Host 1
Puerto PCIe aguas arriba
Puerto PCIe aguas arriba
Dispositivo de conmutación
DP_CFG 0
DP_CFG 1
vEP 0
vEP 1
Módulo de mapeo
Aparato de procesamiento
mEP 0
mEP 1
Puerto PCIe aguas abajo
Conmutador PCIe
EP 0
PF 0
VF 0
EP 1
PF 0
VF 0

FIG. 4

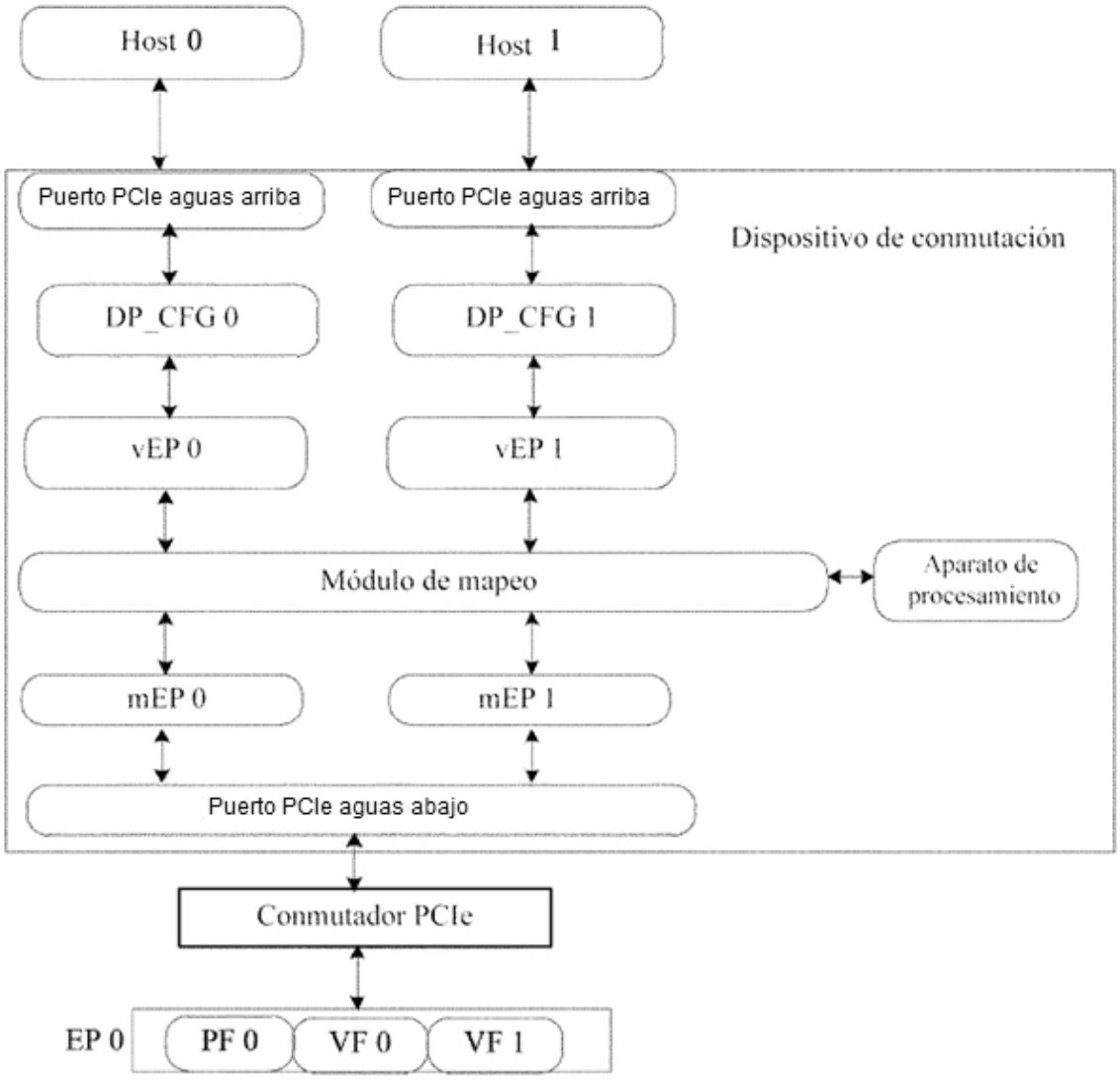
Host 0
Host 1
Puerto PCIe aguas arriba
Puerto PCIe aguas arriba
Dispositivo de conmutación
DP_CFG 0
DP_CFG 1
vEP 0
vEP 1
Módulo de mapeo
Aparato de procesamiento
mEP 0
mEP 1
Puerto PCIe aguas abajo
Conmutador PCIe
EP 0
PF 0
VF 0
VF 1

FIG. 5

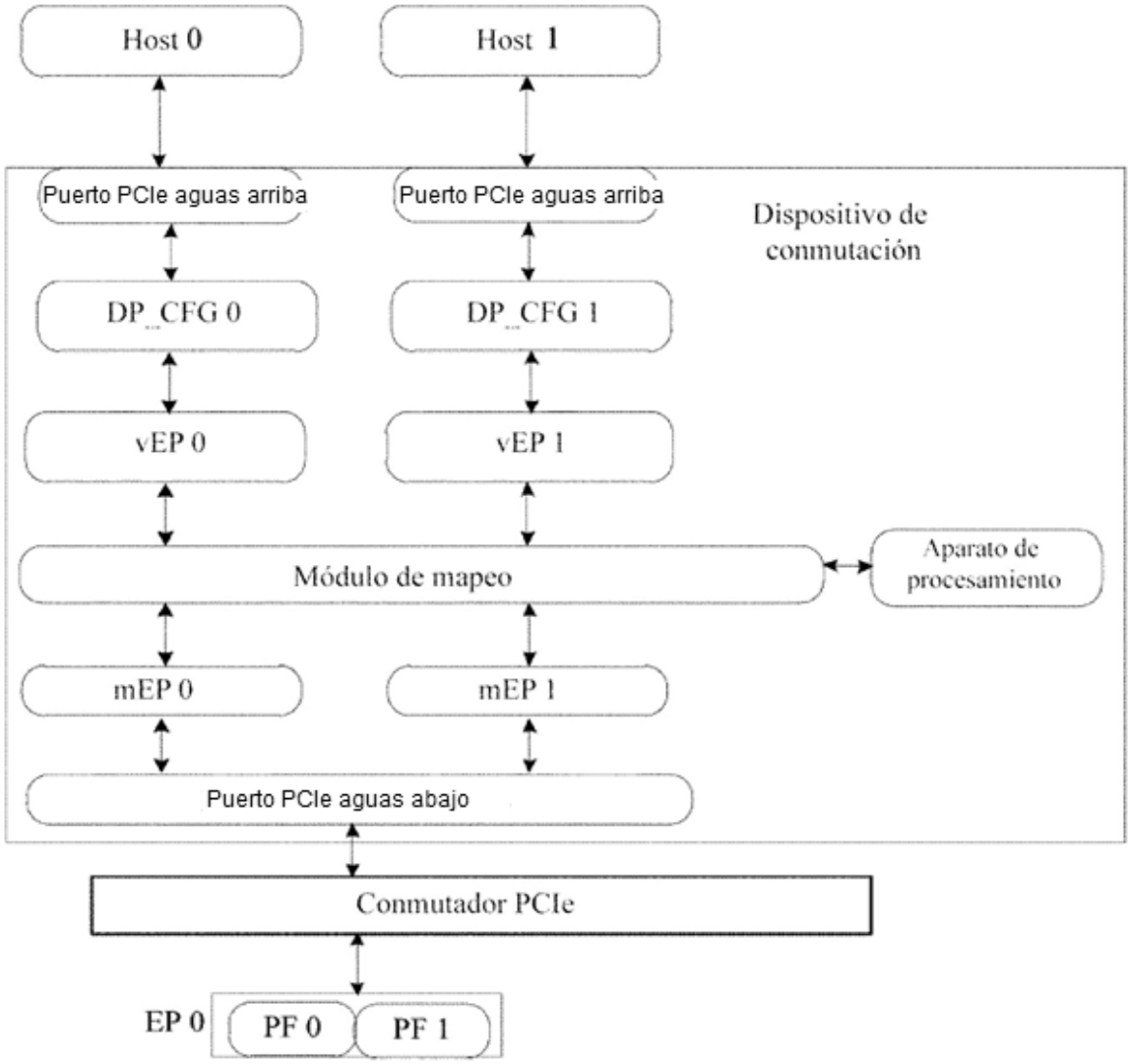
Host 0
Host 1
Puerto PCIe aguas arriba
Puerto PCIe aguas arriba
Dispositivo de conmutación
DP_CFG 0
DP_CFG 1
vEP 0
vEP 1
Módulo de mapeo
Aparato de procesamiento
mEP 0
mEP 1
Puerto PCIe aguas abajo
Conmutador PCIe
EP 0
PF 0
PF 1

FIG. 6

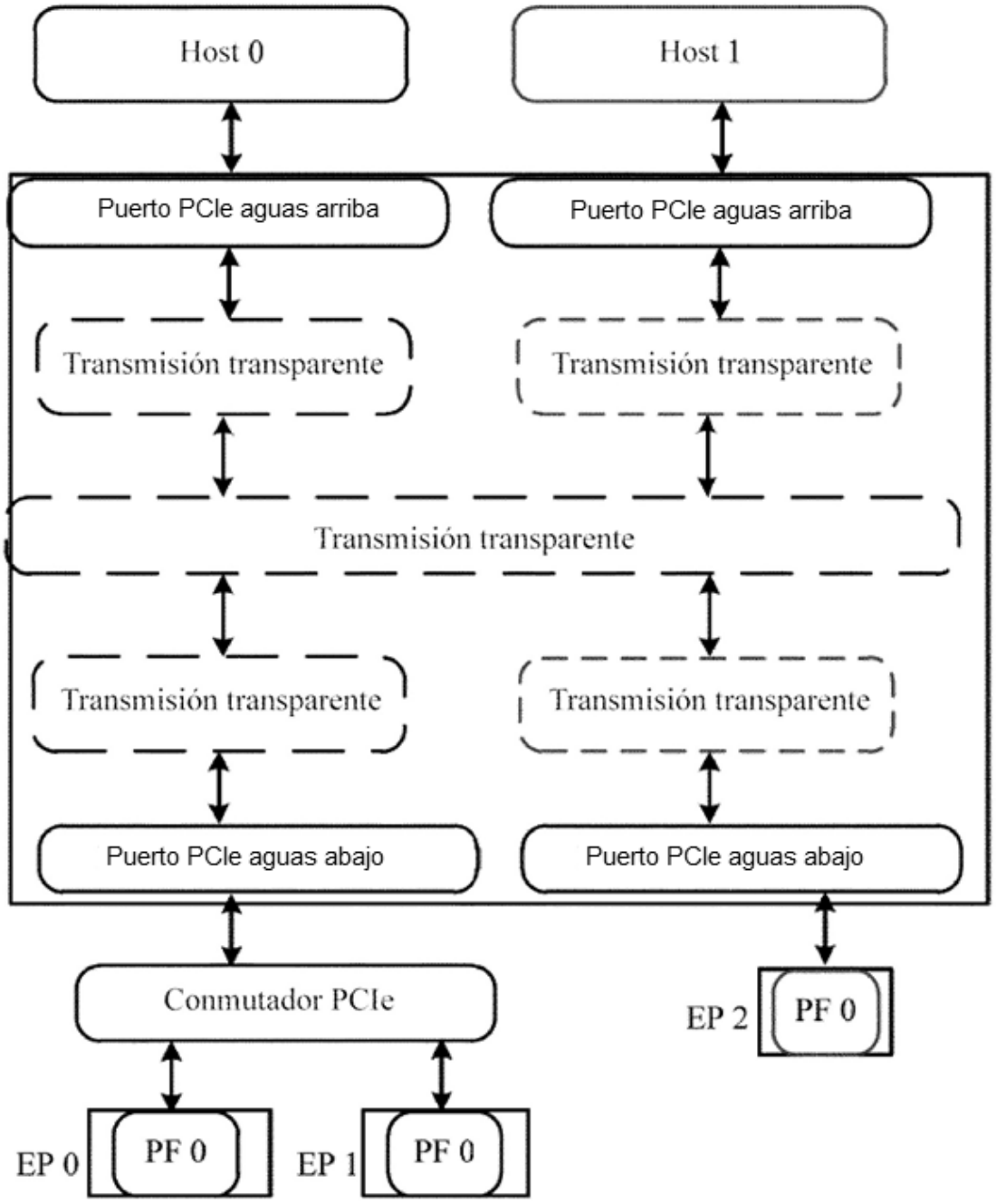
Host 0
Host 1
Puerto PCIe aguas arriba
Puerto PCIe aguas arriba
Transmisión transparente
Transmisión transparente
Transmisión transparente
Transmisión transparente
Transmisión transparente
Puerto PCIe aguas abajo
Puerto PCIe aguas abajo
Conmutador PCIe
EP 2
PF 0
EP 0
PF 0
EP 1
PF 0

FIG. 7

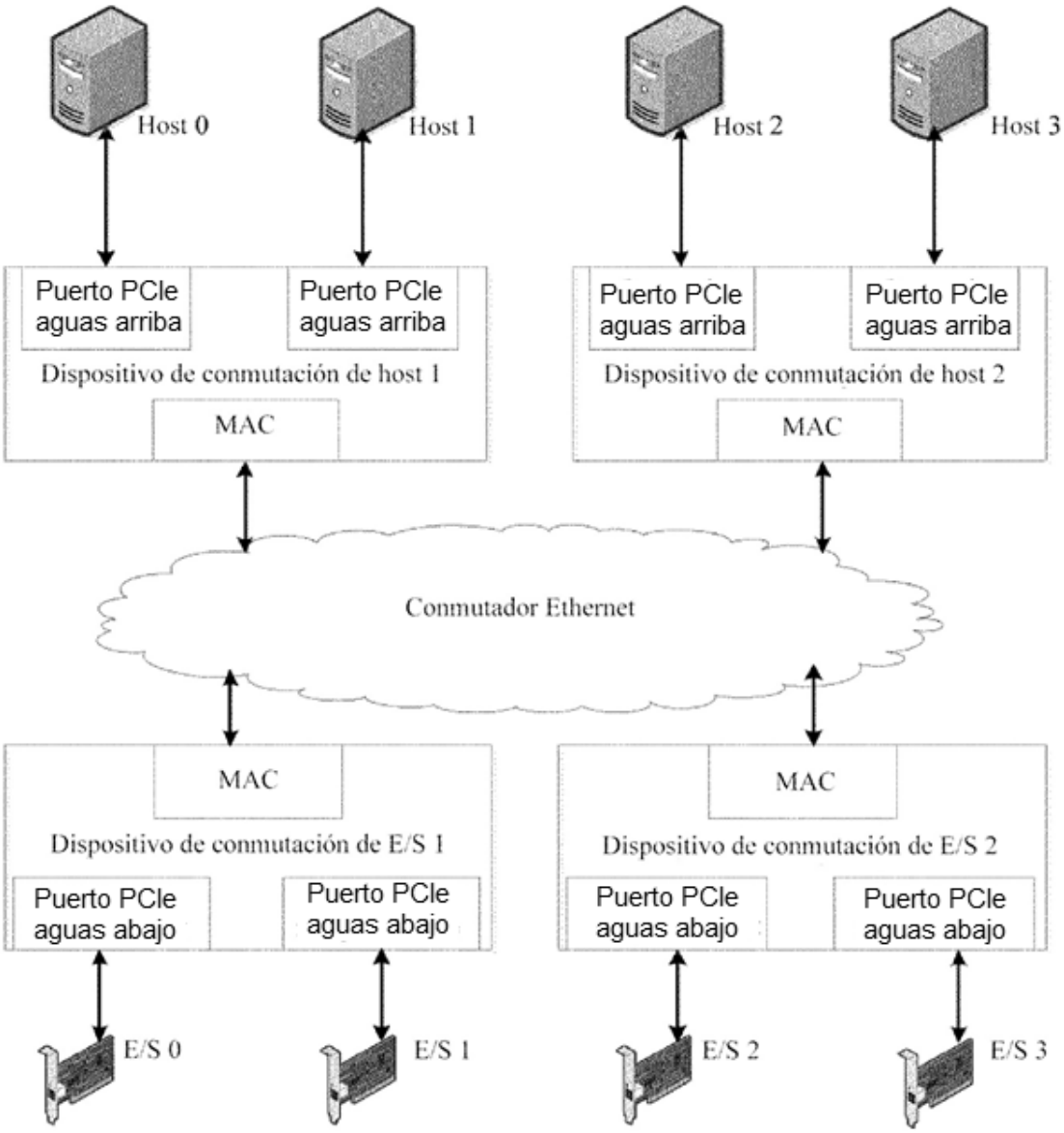
Host 0
Host 1
Host 2
Host 3
Puerto PCIe aguas arriba
Puerto PCIe aguas arriba
Puerto PCIe aguas arriba
Puerto PCIe aguas arriba
Dispositivo de conmutación de host 1
Dispositivo de conmutación de host 2
MAC
MAC
Conmutador Ethernet
MAC
MAC
Dispositivo de conmutación de E/S 1
Dispositivo de conmutación de E/S 2
Puerto PCIe aguas abajo
Puerto PCIe aguas abajo
Puerto PCIe aguas abajo
Puerto PCIe aguas abajo
E/S 0
E/S 1
E/S 2
E/S 3

FIG. 8

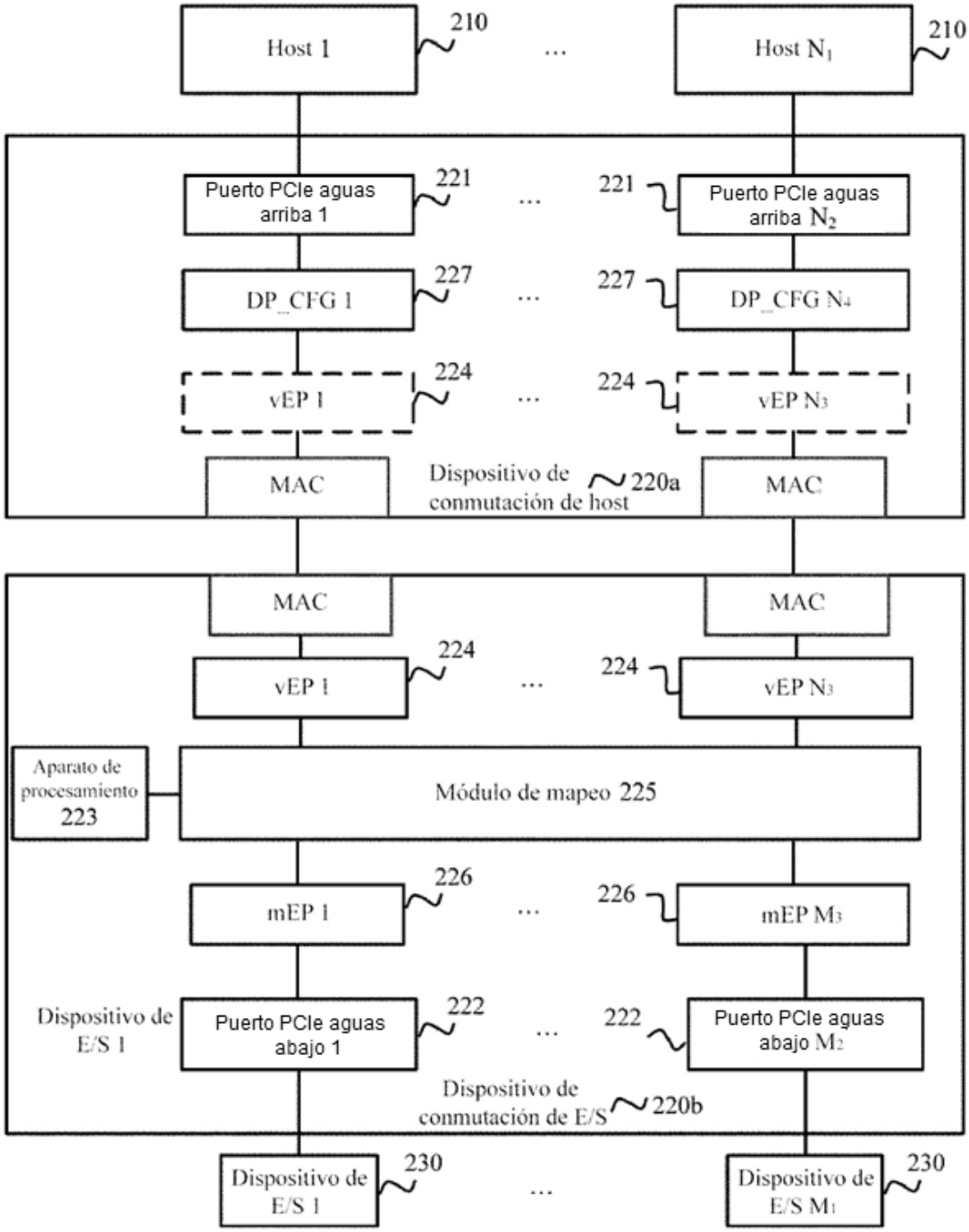
Host 1
210
Host N_1
210
Puerto PCIe aguas arriba 1
221
Puerto PCIe aguas arriba N_2
221
DP_CFG 1
227
DP_CFG N_4
227
vEP 1
224
vEP N_3
224
MAC
Dispositivo de conmutación de host
220a
MAC
MAC
MAC
vEP 1
224
vEP N_3
224
Aparato de procesamiento 223
Módulo de mapeo 225
mEP 1
226
mEP M_3
226
Dispositivo de E/S 1
Puerto PCIe aguas abajo 1
222
Puerto PCIe aguas abajo M_2
222
Dispositivo de conmutación de E/S
220b
Dispositivo de E/S 1
230
Dispositivo de E/S M_1
230

FIG. 9

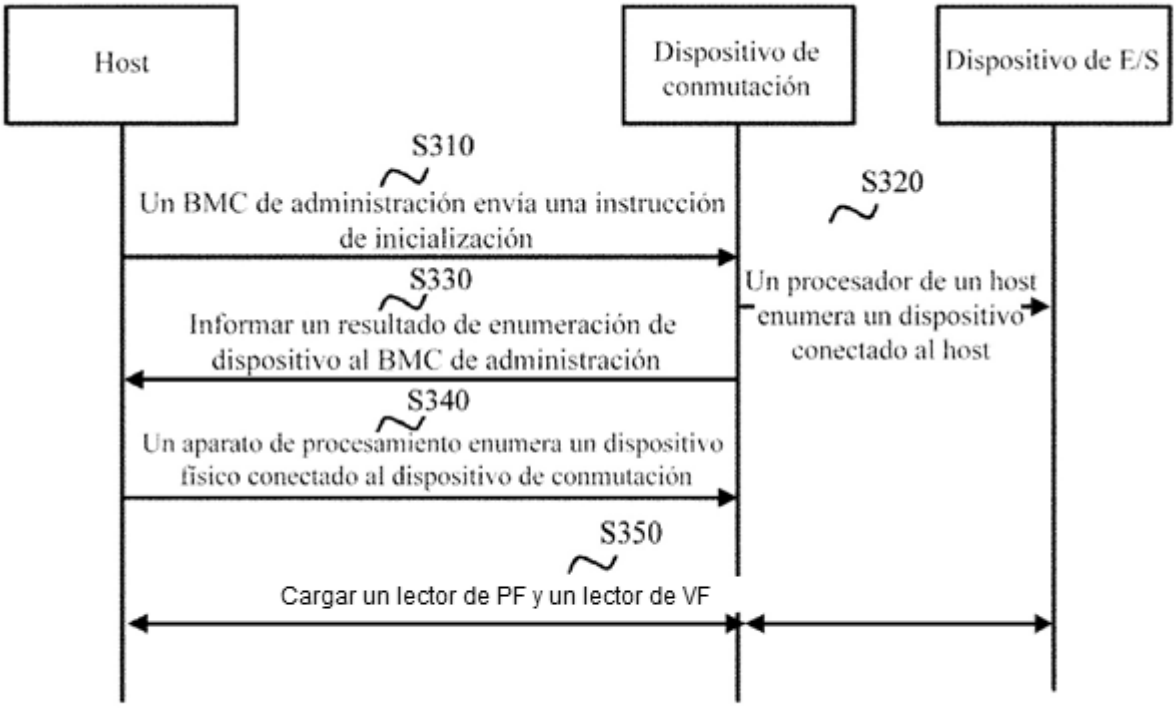
Host
Dispositivo de conmutación
Dispositivo de E/S
S310
Un BMC de administración envía una instrucción de inicialización
S320
Un procesador de un host enumera un dispositivo conectado al host
S330
Informar un resultado de enumeración de dispositivo al BMC de administración
S340
Un aparato de procesamiento enumera un dispositivo físico conectado al dispositivo de conmutación
S350
Cargar un lector de PF y un lector de VF

FIG. 10