



(12)发明专利申请

(10)申请公布号 CN 106685826 A

(43)申请公布日 2017. 05. 17

(21)申请号 201611094388.1

(22)申请日 2016.12.01

(71)申请人 瑞斯康达科技发展股份有限公司
地址 100085 北京市海淀区上地六街28号
院2号楼

(72)发明人 刘文丽

(74)专利代理机构 北京同达信恒知识产权代理
有限公司 11291
代理人 黄志华

(51) Int. Cl.

H04L 12/741(2013.01)

H04L 12/931(2013.01)

H04L 12/933(2013.01)

H04L 12/935(2013.01)

H04L 29/06(2006.01)

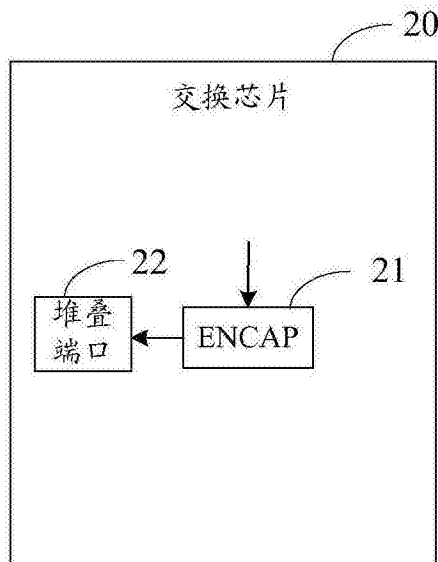
权利要求书3页 说明书11页 附图5页

(54)发明名称

交换机堆叠系统、从设备、交换芯片及处理
协议报文方法

(57)摘要

本发明实施例公开了一种交换机堆叠系统、从设备、交换芯片及处理协议报文方法。该方案中,单独利用交换芯片中的ENCAP模块这一硬件模块,专门对待主设备CPU处理的协议报文进行封装,与现有技术中转发至从设备的CPU进行封装相比,不仅封装过程所需时间更短,而且简化了转发过程,只利用交换芯片进行硬件转发速度更快,这样极大的提升了堆叠系统协议收敛时间,对协议报文的处理效率提高了,减少了协议倒换过程中的丢包数量。另外,还在RCPU头中一次性封装了协议报文原始上送主设备的CPU的DCB信息,能够做到从设备端接收的协议报文中上送给主设备的CPU的信息与交换机单机下协议报文中上送给单机的CPU的信息是一致的。



1. 一种交换芯片,应用于交换机堆叠系统的从设备中,其特征在于,包括封装ENCAP模块和堆叠端口;

所述ENCAP模块,用于:

对接收到的待所述堆叠系统中的主设备中央处理器CPU处理的协议报文,封装远端处理器RCPU头;其中,所述RCPU头中包括二层以太网协议L2头和协议报文原始上送主设备CPU的数据中心桥接DCB信息;其中,所述L2头包括内部使用的目的媒体访问控制MAC地址,和所述从设备CPU的MAC地址;所述DCB信息至少包括入端口信息、协议报文上送原因、时间戳、虚拟局域网VLAN信息;

根据L2转发表以及所述内部使用的目的MAC地址、VLAN信息,查找到所述主设备的交换芯片及主设备的交换芯片的CPU口的信息;其中,所述L2转发表中至少包括所述内部使用的目的MAC地址,VLAN信息,所述主设备的交换芯片及所述主设备的交换芯片的CPU口的信息的对应关系;

根据模块/端口MOD/PORT转发表以及查找到的所述主设备的交换芯片的信息,查找到与所述主设备的交换芯片CPU口通信的堆叠端口信息,并将封装后的所述协议报文转发至查找到的堆叠端口信息所指示的堆叠端口;所述MOD/PORT转发表中至少包括所述主设备的交换芯片的信息与堆叠端口信息的对应关系;

所述堆叠端口,用于接收到所述ENCAP模块封装的所述协议报文后,为所述协议报文封装携带所述主设备的交换芯片的CPU口的信息的HIGIG头,并发送给所述主设备的交换芯片。

2. 根据权利要求1所述的交换芯片,其特征在于,所述交换芯片还包括中央处理器管理接口控制器CMIC模块和远端中央处理器处理检测RCPU CHECK模块;

所述CMIC模块,用于在RCPU CHECK模块使能的状态下,将接收到的协议报文转发至RCPU CHECK模块;在RCPU CHECK模块未使能的状态下,将接收到的协议报文转发至所述从设备的CPU进行封装;

所述RCPU CHECK模块,用于判断接收到的协议报文是否为待所述主设备CPU处理的协议报文;如果是,将接收到的协议报文转发至所述ENCAP模块;否则,将接收到的协议报文返回所述CMIC模块,由所述CMIC模块转发至所述从设备的CPU处理。

3. 根据权利要求2所述的交换芯片,其特征在于,所述RCPU CHECK模块,具体用于:

判断协议报文的的上送原因是否属于预存的上送给所述主设备的CPU的原因;如果是,则确定接收到的协议报文为待主设备处理的协议报文。

4. 根据权利要求1所述的交换芯片,其特征在于,所述DCB信息中还包括协议报文中送的内部优先级信息。

5. 一种交换机堆叠系统中的从设备,其特征在于,所述从设备包括如权利要求1~4任一项所述的交换芯片。

6. 一种如权利要求5所述的从设备处理协议报文的方法,其特征在于,包括:

交换芯片的ENCAP模块对接收到的待堆叠系统中的主设备中央处理器CPU处理的协议报文,封装远端处理器RCPU头;其中,所述RCPU头中包括二层以太网协议L2头和协议报文原始上送主设备CPU的数据中心桥接DCB信息;其中,所述L2头包括内部使用的目的媒体访问控制MAC地址,和所述从设备CPU的MAC地址;所述DCB信息至少包括入端口信息、协议报文中

送原因、时间戳、虚拟局域网VLAN信息；

所述ENCAP模块根据L2转发表以及所述内部使用的目的MAC地址、VLAN信息，查找到所述主设备的交换芯片及主设备的交换芯片的CPU口的信息；其中，所述L2转发表中至少包括所述内部使用的目的MAC地址，VLAN信息，所述主设备的交换芯片及所述主设备的交换芯片的CPU口的信息的对应关系；

所述ENCAP模块根据模块/端口MOD/PORT转发表以及查找到的所述主设备的交换芯片的信息，查找到与所述主设备的交换芯片CPU口通信的堆叠端口信息，并将封装后的所述协议报文转发至查找到的堆叠端口信息所指示的堆叠端口；所述MOD/PORT转发表中至少包括所述主设备的交换芯片的信息与堆叠端口信息的对应关系；

查找到的堆叠端口接收到所述ENCAP模块封装的所述协议报文后，为所述协议报文封装携带所述主设备的交换芯片的CPU口的信息的HIGIG头，并发送给所述主设备的交换芯片。

7. 根据权利要求6所述的方法，其特征在于，交换芯片的ENCAP模块对接收到的待堆叠系统中的主设备CPU处理的协议报文，封装RCPU头之前，该方法还包括：

交换芯片的转发逻辑模块判断接收到的报文是否为协议报文；如果是协议报文，将该协议报文转发至交换芯片的中央处理器管理接口控制器CMIC模块；

所述CMIC模块在交换芯片的远端中央处理器处理检测RCPU CHECK模块使能的状态下，将接收到的协议报文转发至RCPU CHECK模块；在RCPU CHECK模块未使能的状态下，将接收到的协议报文转发至所述从设备的CPU进行封装；

所述RCPU CHECK模块判断接收到的协议报文是否为待所述主设备CPU处理的协议报文；如果是，将接收到的所述协议报文转发至所述ENCAP模块；否则，将接收到的协议报文返回所述CMIC模块，由所述CMIC模块转发至所述从设备的CPU处理。

8. 根据权利要求7所述的方法，其特征在于，所述RCPU CHECK模块判断接收到的协议报文是否为待主设备处理的协议报文，包括：

所述RCPU CHECK模块判断协议报文的发送原因是否属于预存的发送给所述主设备的CPU的原因；如果是，则确定所述协议报文为待主设备处理的协议报文。

9. 根据权利要求8所述的方法，其特征在于，如果所述转发逻辑模块判断所述从设备的端口接收到的报文是协议报文，该方法还包括：

所述转发逻辑模块分配所述协议报文的内部优先级信息，以便主设备的CPU根据协议报文的内部优先级信息对所述协议报文进行处理。

10. 一种交换机堆叠系统，其特征在于，包括主设备和如权利要求5所述的从设备。

11. 一种如权利要求10所述的交换机堆叠系统中处理协议报文的方法，其特征在于，包括：

从设备中，交换芯片的ENCAP模块对接收到的待堆叠系统中的主设备中央处理器CPU处理的协议报文，封装远端处理器RCPU头；其中，所述RCPU头中包括二层以太网协议L2头和协议报文原始发送主设备CPU的数据中心桥接DCB信息；其中，所述L2头包括内部使用的目的媒体访问控制MAC地址，和所述从设备CPU的MAC地址；所述DCB信息至少包括入端口信息、协议报文的发送原因、时间戳、虚拟局域网VLAN信息；

所述ENCAP模块根据L2转发表以及所述内部使用的目的MAC地址、VLAN信息，查找到所

述主设备的交换芯片及主设备的交换芯片的CPU口的信息；其中，所述L2转发表中至少包括所述内部使用的目的MAC地址，VLAN信息，所述主设备的交换芯片及所述主设备的交换芯片的CPU口的信息的对应关系；

所述ENCAP模块根据模块/端口MOD/PORT转发表以及查找到的所述主设备的交换芯片的信息，查找到与所述主设备的交换芯片CPU口通信的堆叠端口信息，并将封装后的所述协议报文转发至查找到的堆叠端口信息所指示的堆叠端口；所述MOD/PORT转发表中至少包括所述主设备的交换芯片的信息与堆叠端口信息的对应关系；

查找到的堆叠端口接收到所述ENCAP模块封装的所述协议报文后，为所述协议报文封装携带所述主设备的交换芯片的CPU口的信息的HIGIG头，并发送给所述主设备的交换芯片；

主设备中，交换芯片接收所述从设备发送的协议报文，并转发至主设备的CPU。

12. 根据权利要求11所述的方法，其特征在于，从设备中，交换芯片的ENCAP模块对接收到的待所述堆叠系统中的主设备CPU处理的协议报文，封装RCPU头之前，该方法还包括：

所述从设备中，交换芯片的逻辑转发模块通过从设备的入端口接收报文；

所述转发逻辑模块判断接收到的报文是否为协议报文；如果是协议报文，将该协议报文转发至从设备的交换芯片的中央处理器管理接口控制器CMIC模块；

所述CMIC模块在从设备的交换芯片的远端中央处理器处理检测RCPU CHECK模块使能的状态下，将接收到的协议报文转发至RCPU CHECK模块；在RCPU CHECK模块未使能的状态下，将接收到的协议报文转发至所述从设备的CPU进行封装；

所述RCPU CHECK模块判断接收到的协议报文是否为待所述主设备CPU处理的协议报文；如果是，将接收到的所述协议报文转发至所述ENCAP模块；否则，将接收到的协议报文返回所述CMIC模块，由所述CMIC模块转发至所述从设备的CPU处理。

交换机堆叠系统、从设备、交换芯片及处理协议报文方法

技术领域

[0001] 本发明涉及网络通信技术领域,尤其涉及一种交换机堆叠系统、从设备、交换芯片及处理协议报文方法。

背景技术

[0002] 随着信息技术的飞速发展,网络规模急剧扩大,以前网络规划的交换机接口数量不足以满足网络应用需求,这时交换机堆叠技术就应运而生了。交换机堆叠技术是用堆叠线缆将几台盒式交换机连接在一起,用于增加交换的端口密度和带宽,实现交换机高速互联和统一管理,从逻辑上来说,它们属于同一个设备,只要登录到其中任何一台交换机,就可对堆叠中的所有交换机进行管理和配置,非常方便用户使用,就像管理单机设备一样。堆叠系统通常由一个主设备和多个从设备组成,每个设备包括中央处理器(Central Processing Unit,CPU)和交换芯片(SWITCH)。主设备和从设备之间采用高速端口连接,其中主设备承担整个堆叠系统的管理以及通信协议的运行。当堆叠系统开启某种协议时,例如链路汇聚控制协议(Link Aggregation Control Protocol,LACP),各设备的端口收到LACP协议时,都要送到主设备的CPU处理,目前来说一般堆叠系统中从设备的端口接收协议报文后,协议报文的数流如图1所示,协议报文依次经过:从设备01的端口、从设备01的交换芯片02、从设备的CPU03、从设备的交换芯片02、主设备04的交换芯片05、主设备的CPU06。相应的步骤具体如下:

[0003] 步骤1、从设备的端口接收到报文后,转发至从设备的交换芯片。

[0004] 步骤2、经从设备的交换芯片判断该报文为协议报文,说明需要由主设备的CPU进行协议处理;从设备的交换芯片将协议报文通过存储器直接访问(Direct Memory Access,DMA)机制发送给从设备的CPU处理;相应的,从设备的CPU收到协议报文后,将协议报文进行封装,需要封装从设备的媒体访问控制(Media Access Control,MAC)为源MAC,主设备的MAC为目的MAC,默认虚拟局域网(Virtual Local Area Network,VLAN),堆叠系统内部的以太网协议类型,这样就封装生成了一个新的二层以太网协议(L2)头后,然后重新计算循环冗余校验(Cyclic Redundancy Check,CRC),并设置HIGIG头的目的端口为主设备的交换芯片的CPU口;其中,HIGIG是一种串行总线互联方案。

[0005] 步骤3、从设备的CPU再将封装后的协议报文发回给从设备的交换芯片。

[0006] 步骤4、从设备的交换芯片接收到从设备的CPU封装后的协议报文后,根据报文的HIGIG头中的目的端口,查找交换芯片的转发逻辑,将报文转发到与主设备通过堆叠线连接的堆叠端口,通过该堆叠端口将报文送到堆叠系统的主设备的交换芯片。

[0007] 步骤5、主设备的交换芯片收到该协议报文后,同样通过报文的HIGIG头中的目的端口,查找交换芯片的转发逻辑再转发给主设备的CPU;相应的,主设备的CPU接收到该协议报文后,对该协议报文进行解封装,然后再送给协议模块处理。

[0008] 从上述描述可以看出,从设备的端口接收协议报文后将其发送到主设备的CPU的流程繁多,而且需要从设备的CPU对协议报文进行软件封装操作,会封装源MAC,目的MAC,

VLAN,以太网协议类型信息,封装完成后再发回给从设备的交换芯片,由从设备的交换芯片再转发给主设备的交换芯片,再到主设备的CPU去处理。该方案中,由从设备的CPU封装报文的方式速率比较慢,这样会导致主设备的CPU接收到协议报文的延迟很大,协议报文处理不及时,对一些对性能要求很高的协议达不到性能要求,尤其是在协议报文比较多的情况下,从设备的CPU软件处理不过来,还会产生丢包,导致协议运行不正常。

发明内容

[0009] 本发明实施例的目的是提供一种交换机堆叠系统、从设备、交换芯片及处理协议报文方法,用于解决由从设备的CPU对协议报文进行封装的方式处理速率较慢的问题。

[0010] 本发明实施例的目的是通过以下技术方案实现的:

[0011] 一种交换芯片,应用于交换机堆叠系统的从设备中,包括封装ENCAP模块和堆叠端口;

[0012] 所述ENCAP模块,用于:

[0013] 对接收到的待所述堆叠系统中的主设备中央处理器CPU处理的协议报文,封装远端处理器RCPU头;其中,所述RCPU头中包括二层以太网协议L2头和协议报文原始上送主设备CPU的数据中心桥接DCB信息;其中,所述L2头包括内部使用的目的媒体访问控制MAC地址,和所述从设备CPU的MAC地址;所述DCB信息至少包括入端口信息、协议报文中送原因、时间戳、虚拟局域网VLAN信息;

[0014] 根据L2转发表以及所述内部使用的目的MAC地址、VLAN信息,查找到所述主设备的交换芯片及主设备的交换芯片的CPU口的信息;其中,所述L2转发表中至少包括所述内部使用的目的MAC地址,VLAN信息,所述主设备的交换芯片及所述主设备的交换芯片的CPU口的信息的对应关系;

[0015] 根据模块/端口MOD/PORT转发表以及查找到的所述主设备的交换芯片的信息,查找到与所述主设备的交换芯片CPU口通信的堆叠端口信息,并将封装后的所述协议报文转发至查找到的堆叠端口信息所指示的堆叠端口;所述MOD/PORT转发表中至少包括所述主设备的交换芯片的信息与堆叠端口信息的对应关系;

[0016] 所述堆叠端口,用于接收到所述ENCAP模块封装的所述协议报文后,为所述协议报文封装携带所述主设备的交换芯片的CPU口的信息的HIGIG头,并发送给所述主设备的交换芯片。

[0017] 较佳地,所述交换芯片还包括中央处理器管理接口控制器CMIC模块和远端中央处理器处理检测RCPU CHECK模块;

[0018] 所述CMIC模块,用于在RCPU CHECK模块使能的状态下,将接收到的协议报文转发至RCPU CHECK模块;在RCPU CHECK模块未使能的状态下,将接收到的协议报文转发至所述从设备的CPU进行封装;

[0019] 所述RCPU CHECK模块,用于判断接收到的协议报文是否为待所述主设备CPU处理的协议报文;如果是,将接收到的协议报文转发至所述ENCAP模块;否则,将接收到的协议报文返回所述CMIC模块,由所述CMIC模块转发至所述从设备的CPU处理。

[0020] 较佳地,所述RCPU CHECK模块,具体用于:

[0021] 判断协议报文中送原因是否属于预存的上送给所述主设备的CPU的原因;如果

是,则确定接收到的协议报文为待主设备处理的协议报文。

[0022] 较佳地,所述DCB信息中还包包括协议报文中上送的内部优先级信息。

[0023] 一种交换机堆叠系统中的从设备,所述从设备包括如以上任一项所述的交换芯片。

[0024] 一种如以上所述的从设备处理协议报文的方法,包括:

[0025] 交换芯片的ENCAP模块对接收到的待堆叠系统中的主设备中央处理器CPU处理的协议报文,封装远端处理器RCPU头;其中,所述RCPU头中包括二层以太网协议L2头和协议报文原始上送主设备CPU的数据中心桥接DCB信息;其中,所述L2头包括内部使用的目的媒体访问控制MAC地址,和所述从设备CPU的MAC地址;所述DCB信息至少包括入端口信息、协议报文中上送原因、时间戳、虚拟局域网VLAN信息;

[0026] 所述ENCAP模块根据L2转发表以及所述内部使用的目的MAC地址、VLAN信息,查找到所述主设备的交换芯片及主设备的交换芯片的CPU口的信息;其中,所述L2转发表中至少包括所述内部使用的目的MAC地址,VLAN信息,所述主设备的交换芯片及所述主设备的交换芯片的CPU口的信息的对应关系;

[0027] 所述ENCAP模块根据模块/端口MOD/PORT转发表以及查找到的所述主设备的交换芯片的信息,查找到与所述主设备的交换芯片CPU口通信的堆叠端口信息,并将封装后的所述协议报文转发至查找到的堆叠端口信息所指示的堆叠端口;所述MOD/PORT转发表中至少包括所述主设备的交换芯片的信息与堆叠端口信息的对应关系;

[0028] 查找到的堆叠端口接收到所述ENCAP模块封装的所述协议报文后,为所述协议报文封装携带所述主设备的交换芯片的CPU口的信息的HIGIG头,并发送给所述主设备的交换芯片。

[0029] 较佳地,交换芯片的ENCAP模块对接收到的待堆叠系统中的主设备CPU处理的协议报文,封装RCPU头之前,该方法还包括:

[0030] 交换芯片的转发逻辑模块判断接收到的报文是否为协议报文;如果是协议报文,将该协议报文转发至交换芯片的中央处理器管理接口控制器CMIC模块;

[0031] 所述CMIC模块在交换芯片的远端中央处理器处理检测RCPU CHECK模块使能的状态下,将接收到的协议报文转发至RCPU CHECK模块;在RCPU CHECK模块未使能的状态下,将接收到的协议报文转发至所述从设备的CPU进行封装;

[0032] 所述RCPU CHECK模块判断接收到的协议报文是否为待所述主设备CPU处理的协议报文;如果是,将接收到的所述协议报文转发至所述ENCAP模块;否则,将接收到的协议报文返回所述CMIC模块,由所述CMIC模块转发至所述从设备的CPU处理。

[0033] 较佳地,所述RCPU CHECK模块判断接收到的协议报文是否为待主设备处理的协议报文,包括:

[0034] 所述RCPU CHECK模块判断协议报文中上送原因是否属于预存的上送给所述主设备的CPU的原因;如果是,则确定所述协议报文为待主设备处理的协议报文。

[0035] 较佳地,如果所述转发逻辑模块判断所述从设备的端口接收到的报文是协议报文,该方法还包括:

[0036] 所述转发逻辑模块分配所述协议报文中上送的内部优先级信息,以便主设备的CPU根据协议报文中上送的内部优先级信息对所述协议报文进行处理。

- [0037] 一种交换机堆叠系统,包括主设备和如以上所述的从设备。
- [0038] 一种如以上所述的交换机堆叠系统中处理协议报文的方法,包括:
- [0039] 从设备中,交换芯片的ENCAP模块对接收到的待堆叠系统中的主设备中央处理器CPU处理的协议报文,封装远端处理器RCPU头;其中,所述RCPU头中包括二层以太网协议L2头和协议报文原始上送主设备CPU的数据中心桥接DCB信息;其中,所述L2头包括内部使用的目的媒体访问控制MAC地址,和所述从设备CPU的MAC地址;所述DCB信息至少包括入端口信息、协议报文上送原因、时间戳、虚拟局域网VLAN信息;
- [0040] 所述ENCAP模块根据L2转发表以及所述内部使用的目的MAC地址、VLAN信息,查找到所述主设备的交换芯片及主设备的交换芯片的CPU口的信息;其中,所述L2转发表中至少包括所述内部使用的目的MAC地址,VLAN信息,所述主设备的交换芯片及所述主设备的交换芯片的CPU口的信息的对应关系;
- [0041] 所述ENCAP模块根据模块/端口MOD/PORT转发表以及查找到的所述主设备的交换芯片的信息,查找到与所述主设备的交换芯片CPU口通信的堆叠端口信息,并将封装后的所述协议报文转发至查找到的堆叠端口信息所指示的堆叠端口;所述MOD/PORT转发表中至少包括所述主设备的交换芯片的信息与堆叠端口信息的对应关系;
- [0042] 查找到的堆叠端口接收到所述ENCAP模块封装的所述协议报文后,为所述协议报文封装携带所述主设备的交换芯片的CPU口的信息的HIGIG头,并发送给所述主设备的交换芯片;
- [0043] 主设备中,交换芯片接收所述从设备发送的协议报文,并转发至主设备的CPU。
- [0044] 较佳地,从设备中,交换芯片的ENCAP模块对接收到的待所述堆叠系统中的主设备CPU处理的协议报文,封装RCPU头之前,该方法还包括:
- [0045] 所述从设备中,交换芯片的逻辑转发模块通过从设备的入端口接收报文;
- [0046] 所述转发逻辑模块判断接收到的报文是否为协议报文;如果是协议报文,将该协议报文转发至从设备的交换芯片的中央处理器管理接口控制器CMIC模块;
- [0047] 所述CMIC模块在从设备的交换芯片的远端中央处理器处理检测RCPU CHECK模块使能的状态下,将接收到的协议报文转发至RCPU CHECK模块;在RCPU CHECK模块未使能的状态下,将接收到的协议报文转发至所述从设备的CPU进行封装;
- [0048] 所述RCPU CHECK模块判断接收到的协议报文是否为待所述主设备CPU处理的协议报文;如果是,将接收到的所述协议报文转发至所述ENCAP模块;否则,将接收到的协议报文返回所述CMIC模块,由所述CMIC模块转发至所述从设备的CPU处理。
- [0049] 本发明实施例的有益效果如下:
- [0050] 本发明实施例提供的交换机堆叠系统、从设备、交换芯片及处理协议报文方法中,单独利用交换芯片中的ENCAP模块这一硬件模块,专门对待主设备CPU处理的协议报文进行封装,与现有技术中转发至从设备的CPU进行封装相比,不仅封装过程所需时间更短,而且简化了转发过程,只利用交换芯片进行硬件转发速度更快,这样极大的提升了堆叠系统协议收敛时间,对协议报文的处理效率提高了,减少了协议倒换过程中的丢包数量。另外,还在RCPU头中一次性封装了协议报文原始上送主设备的CPU的DCB信息,可以保证将协议报文原始上送的信息保存下来,发送给主设备的CPU,而不会被丢失,能够做到从设备端接收的协议报文上送给主设备的CPU的信息与交换机单机下协议报文上送给单机的CPU的信息是

一致的。

附图说明

- [0051] 图1为现有技术中的一种交换机堆叠系统中的协议报文的数据流示意图；
- [0052] 图2为本发明实施例提供的一种交换芯片的结构示意图；
- [0053] 图3为本发明实施例提供的另一种交换芯片的结构示意图；
- [0054] 图4为本发明实施例提供的又一种交换芯片的结构示意图；
- [0055] 图5为本发明实施例提供的一种交换机堆叠系统中的从设备处理协议报文的方法流程图；
- [0056] 图6为本发明实施例提供的一种交换机堆叠系统中处理协议报文的方法流程图；
- [0057] 图7为本发明实施例提供的一种交换机堆叠系统的结构示意图。

具体实施方式

[0058] 下面结合附图和实施例对本发明提供的一种交换机堆叠系统、从设备、交换芯片及处理协议报文方法进行更详细地说明。

[0059] 如图2所示,本发明实施例提供一种交换芯片20,应用于交换机堆叠系统的从设备中,包括封装(ENCAP)模块21和堆叠端口22;

[0060] ENCAP模块21,用于:

[0061] 对接收到的待堆叠系统中的主设备CPU处理的协议报文,封装远端处理器(Remote CPU,RCPU)头;其中,RCPU头中包括二层以太网协议L2头和协议报文原始上送主设备CPU的数据中心桥接(Data Center Bridge,DCB)信息;其中,L2头包括内部使用的目的MAC地址,和从设备CPU的MAC地址;DCB信息至少包括入端口信息、协议报文上送原因、时间戳、虚拟局域网VLAN信息;

[0062] 根据L2转发表以及内部使用的目的MAC地址、VLAN信息,查找到主设备的交换芯片及主设备的交换芯片的CPU口的信息;其中,L2转发表中至少包括内部使用的目的MAC地址,VLAN信息,主设备的交换芯片及主设备的交换芯片的CPU口的信息的对应关系;

[0063] 根据模块/端口(MOD/PORT)转发表以及查找到的主设备的交换芯片的信息,查找到与主设备的交换芯片CPU口通信的堆叠端口信息,并将封装后的协议报文转发至查找到的堆叠端口信息所指示的堆叠端口;MOD/PORT转发表中至少包括主设备的交换芯片的信息与堆叠端口信息的对应关系;

[0064] 堆叠端口22,用于接收到ENCAP模块封装的协议报文后,为协议报文封装携带主设备的交换芯片的CPU口的信息的HIGIG头,并发送给主设备的交换芯片。

[0065] 其中,RCPU是对于从设备来说的,除本从设备的CPU以外的其它设备的CPU。

[0066] 其中,L2头中从设备的MAC是源MAC;内部(即交换机堆叠系统内部)使用的目的MAC,可以是预留的MAC;VLAN为默认设置,例如可以为VLAN1;以太网协议类型为交换机堆叠系统内部使用的协议类型。

[0067] 其中,L2转发表和MOD/PORT转发表是预先设置的;L2转发表存储不同类型的报文的源MAC、目的MAC、VLAN信息与目的端口信息(由目的端口所在模块信息及端口信息组成)的对应关系;MOD/PORT转发表存储目的端口所在模块信息与本地的端口信息的对应关系。

配协议报文上送的内部优先级信息。

[0080] 基于同样的发明构思,本发明实施例还提供一种交换机堆叠系统中的从设备,从设备包括如以上任意实施例所述的交换芯片。

[0081] 基于同样的发明构思,本发明实施例还提供一种如以上所述的从设备处理协议报文的方法,如图5所示,包括:

[0082] 步骤510、交换芯片的ENCAP模块对接收到的待堆叠系统中的主设备CPU处理的协议报文,封装RCPU头;其中,RCPU头中包括L2头和协议报文原始上送主设备CPU的DCB信息;其中,L2头包括内部使用的目的MAC地址,和从设备CPU的MAC地址;DCB信息至少包括入端口信息、协议报文上送原因、时间戳、VLAN信息;

[0083] 步骤520、ENCAP模块根据L2转发表以及内部使用的目的MAC地址、VLAN信息,查找主设备的交换芯片及主设备的交换芯片的CPU口的信息;其中,L2转发表中至少包括内部使用的目的MAC地址,VLAN信息,主设备的交换芯片及主设备的交换芯片的CPU口的信息的对应关系;

[0084] 步骤530、ENCAP模块根据MOD/PORT转发表以及查找到的主设备的交换芯片的信息,查找与主设备的交换芯片CPU口通信的堆叠端口信息,并将封装后的协议报文转发至查找到的堆叠端口信息所指示的堆叠端口;MOD/PORT转发表中至少包括主设备的交换芯片的信息与堆叠端口信息的对应关系;

[0085] 步骤540、查找到的堆叠端口接收到ENCAP模块封装的协议报文后,为协议报文封装携带主设备的交换芯片的CPU口的信息的HIGIG头,并发送给主设备的交换芯片。

[0086] 本发明实施例中,单独利用交换芯片中的ENCAP模块这一硬件模块,专门对待主设备CPU处理的协议报文进行封装,与现有技术中转发至从设备的CPU进行封装相比,不仅封装过程所需时间更短,而且简化了转发过程,只利用交换芯片进行硬件转发速度更快,这样极大的提升了堆叠系统协议收敛时间,对协议报文的处理效率提高了,减少了协议倒换过程中的丢包数量。另外,还在RCPU头中一次性封装了协议报文原始上送主设备的CPU的DCB信息,可以保证将协议报文原始上送的信息保存下来,发送给主设备的CPU,而不会被丢失,能够做到从设备端接收的协议报文上送给主设备的CPU的信息与交换机单机下协议报文中上送给单机的CPU的信息是一致的。

[0087] 较佳地,交换芯片的ENCAP模块对接收到的待堆叠系统中的主设备CPU处理的协议报文,封装RCPU头之前,本发明实施例提供的方法还包括:

[0088] 交换芯片的转发逻辑模块判断接收到的报文是否为协议报文;如果是协议报文,将该协议报文转发至交换芯片的CMIC模块;

[0089] CMIC模块在交换芯片的RCPU CHECK模块使能的状态下,将接收到的协议报文转发至RCPU CHECK模块;在RCPU CHECK模块未使能的状态下,将接收到的协议报文转发至从设备的CPU进行封装;

[0090] RCPUCHECK模块判断接收到的协议报文是否为待主设备CPU处理的协议报文;如果是,将接收到的协议报文转发至ENCAP模块;否则,将接收到的协议报文返回CMIC模块,由CMIC模块转发至从设备的CPU处理。

[0091] 较佳地,RCPU CHECK模块判断接收到的协议报文是否为待主设备处理的协议报文,具体实现方式可以是:RCPU CHECK模块判断协议报文的的上送原因是否属于预存的上送

给主设备的CPU的原因;如果是,则确定协议报文为待主设备处理的协议报文。

[0092] 较佳地,如果转发逻辑模块判断从设备的端口接收到的报文是协议报文,本发明实施例提供的处理方法还包括:

[0093] 转发逻辑模块分配协议报文中上送的入内部优先级信息,以便主设备的CPU根据协议报文中上送的内部优先级信息对协议报文进行处理。

[0094] 基于同样的发明构思,本发明实施例还提供一种交换机堆叠系统,包括主设备和如以上任意实施例所述的从设备。

[0095] 基于同样的发明构思,本发明实施例还提供一种如以上实施例所述的交换机堆叠系统中处理协议报文的方法,如图6所示,该方法包括:

[0096] 步骤610、从设备中,交换芯片的ENCAP模块对接收到的待堆叠系统中的主设备CPU处理的协议报文,封装RCPU头;其中,RCPU头中包括L2头和协议报文原始上送主设备CPU的DCB信息;其中,L2头包括内部使用的目的MAC地址,和从设备CPU的MAC地址;DCB信息至少包括入端口信息、协议报文中上送原因、时间戳、VLAN信息;

[0097] 步骤620、ENCAP模块根据L2转发表以及内部使用的目的MAC地址、VLAN信息,查找主设备的交换芯片及主设备的交换芯片的CPU口的信息;其中,L2转发表中至少包括内部使用的目的MAC地址,VLAN信息,主设备的交换芯片及主设备的交换芯片的CPU口的信息的对应关系;

[0098] 步骤630、ENCAP模块根据MOD/PORT转发表以及查找到的主设备的交换芯片的信息,查找与主设备的交换芯片CPU口通信的堆叠端口信息,并将封装后的协议报文转发至查找到的堆叠端口信息所指示的堆叠端口;MOD/PORT转发表中至少包括主设备的交换芯片的信息与堆叠端口信息的对应关系;

[0099] 步骤640、查找到的堆叠端口接收到ENCAP模块封装的协议报文后,为协议报文封装携带主设备的交换芯片的CPU口的信息的HIGIG头,并发送给主设备的交换芯片;

[0100] 步骤650、主设备中,交换芯片接收从设备发送的协议报文,并转发至主设备的CPU。

[0101] 较佳地,从设备中,交换芯片的ENCAP模块对接收到的待堆叠系统中的主设备CPU处理的协议报文,封装RCPU头之前,本发明实施例提供的方法还包括:

[0102] 从设备中,交换芯片的逻辑转发模块通过从设备的入端口接收报文;

[0103] 转发逻辑模块判断接收到的报文是否为协议报文;如果是协议报文,将该协议报文转发至从设备的交换芯片的CMIC模块;

[0104] CMIC模块在从设备的交换芯片的RCPU CHECK模块使能的状态下,将接收到的协议报文转发至RCPU CHECK模块;在RCPU CHECK模块未使能的状态下,将接收到的协议报文转发至从设备的CPU进行封装;

[0105] RCPUCHECK模块判断接收到的协议报文是否为待主设备CPU处理的协议报文;如果是,将接收到的协议报文转发至ENCAP模块;否则,将接收到的协议报文返回CMIC模块,由CMIC模块转发至从设备的CPU处理。

[0106] 下面结合具体的应用场景对本发明实施例提供的一种交换机堆叠系统、从设备、交换芯片及处理协议报文方法进行更加详细地描述。

[0107] 本实施例的交换机堆叠系统中,如图7所示,包括主设备30和从设备40,主设备包

括CPU50和交换芯片60,从设备40包括交换芯片20和CPU70。

[0108] 本实施例中的交换芯片的硬件结构如图4所示,包括:入端口26、转发逻辑模块25,CMIC模块23,RCPU CHECK模块24,ENCAP模块21、堆叠端口22。

[0109] 基于此,本实施例中对协议报文的处理过程具体如下:

[0110] 步骤一、从入端口26接收到报文后,发送给转发逻辑模块25。

[0111] 步骤二、转发逻辑模块25判断该报文是否为协议报文,如果是协议报文,转发逻辑模块25在协议报文对应的DCB信息中写入内部优先级信息,并将协议报文转发至CMIC模块23。

[0112] 步骤三、CMIC模块23将协议报文转发至RCPU CHECK模块24。

[0113] 其中,RCPU CHECK模块24的上述检测功能在从设备初始化时已经使能。

[0114] 步骤四、RCPU CHECK模块24判断接收到的协议报文是否为待主设备CPU处理的协议报文;如果是,将接收到的协议报文转发至ENCAP模块;否则,将接收到的协议报文返回CMIC模块,由CMIC模块转发至从设备的CPU处理。

[0115] 实施中,可以预先设置交换芯片的RCPU CHECK模块的原因寄存器,寄存器的值为协议报文中上送给主设备的CPU的原因,至少包括通过协议报文寄存器上送的原因,通过CAP匹配规则上送的原因,通过MAC表设置上送的原因,如果符合设置的这些原因,就需要进入ENCAP模块自动封装,以便转发到主设备的CPU。

[0116] 步骤五、ENCAP模块21为协议报文封装RCPU头;其中,RCPU头中包括L2头和协议报文原始上送主设备CPU的DCB信息;其中,L2头包括内部使用的目的MAC地址,和从设备CPU的MAC地址;DCB信息至少包括入端口信息、协议报文中送的内部优先级信息、协议报文中送原因、时间戳、VLAN信息。

[0117] 实施中,L2头中,VLAN为默认VLAN 1,以太网协议类型为内部使用的协议类型。

[0118] 实施中,ENCAP模块将该协议报文中送的内部优先级信息单独封装到RCPU头中的CPU_TC字段中。该字段占用2字节宽度,该CPU_TC字段位于封装报文的第28字节(不算VLAN TAG的情况下是第24个字节)。

[0119] 步骤六、ENCAP模块21根据L2转发表以及内部使用的目的MAC地址、VLAN信息,查找主设备的交换芯片及主设备的交换芯片的CPU口的信息;其中,L2转发表中至少包括内部使用的目的MAC地址,VLAN信息,主设备的交换芯片及主设备的交换芯片的CPU口的信息的对应关系。

[0120] 步骤七、ENCAP模块21根据MOD/PORT转发表以及查找到的主设备的交换芯片的信息,查找与主设备的交换芯片CPU口通信的堆叠端口信息,并将封装后的协议报文转发至查找到的堆叠端口信息所指示的堆叠端口;MOD/PORT转发表中至少包括主设备的交换芯片的信息与堆叠端口信息的对应关系。

[0121] 步骤八、查找到的堆叠端口22接收到ENCAP模块封装的协议报文后,为协议报文封装携带主设备的交换芯片的CPU口的信息的HIGIG头,并发送给主设备的交换芯片60。

[0122] 步骤九、主设备的交换芯片60通过与从设备连接的堆叠端口接收到封装后的协议报文后,该协议报文带有HIGIG头,主设备的交换芯片查找该HIGIG头发现协议报文的目的是交换芯片的CPU口,则直接将协议报文转发给交换芯片的CPU口。

[0123] 步骤十、与交换芯片的CPU口连接的主设备的CPU50接收到封装后的协议报文时,

解析RCPU的头,得到上述L2头和DCB信息,就可以获得协议报文上送给主设备的CPU的原因,接收协议报文的入端口,内部优先级信息等等。

[0124] 由于在协议报文的RCPU头中封装有内部优先级信息,在主设备交换芯片中需要创建一个新的内部配置访问接口(Ingress Content Aware Processor, ICAP)模板表项用来匹配协议报文封装后的协议字段,以及设置用户自定义(User File Directory, UDF),设置匹配的字段为报文偏移开始是24字节,字节宽度是2个字节,用于匹配封装后报文的RCPU头中的CPU_TC内部优先级值字段。这样通过匹配这两个字段就可以标识该类协议报文以及某个内部优先级值的报文了,然后针对这类报文进行下一步的特殊处理。

[0125] 针对上面标识的某个内部优先级的报文,设置ICAP相关动作,动作为:在主设备的CPU的内部优先级修改上送CPU的服务等级(Class of Service, COS)队列,修改内部优先级的值为匹配的CPU_TC字段(内部优先级值)。这样就能保证上送从设备的CPU的优先级COS队列和协议报文通过转发后到主设备的CPU的优先级COS队列一致,从而保证高优先级的报文,处于高优先级队列,优先被主设备的CPU处理。

[0126] 本实施例中,协议报文的数据流的走向如图7的箭头所示,经过从设备的交换芯片、主设备的交换芯片、主设备的CPU,与现有技术中的图1所示的协议报文的数据流相比,节省了经过从设备的CPU的流程,对协议报文的处理速率更快。

[0127] 本发明实施例提供的交换机堆叠系统、从设备、交换芯片及处理协议报文方法中,单独利用交换芯片中的ENCAP模块这一硬件模块,专门对待主设备CPU处理的协议报文进行封装,与现有技术中转发至从设备的CPU进行封装相比,不仅封装过程所需时间更短,而且简化了转发过程,只利用交换芯片进行硬件转发速度更快,这样极大的提升了堆叠系统协议收敛时间,对协议报文的处理效率提高了,减少了协议倒换过程中的丢包数量。另外,还在RCPU头中一次性封装了协议报文原始上送主设备的CPU的DCB信息,可以保证将协议报文原始上送的信息保存下来,发送给主设备的CPU,而不会被丢失,能够做到从设备端接收的协议报文上送给主设备的CPU的信息与交换机单机下协议报文上送给单机的CPU的信息是一致的。

[0128] 本领域内的技术人员应明白,本发明的实施例可提供为方法、系统、或计算机程序产品。因此,本发明可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且,本发明可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[0129] 本发明是参照根据本发明实施例的方法、设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器,使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[0130] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令装置的制品,该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或

多个方框中指定的功能。

[0131] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上,使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0132] 尽管已描述了本发明的优选实施例,但本领域内的技术人员一旦得知了基本创造性概念,则可对这些实施例作出另外的变更和修改。所以,所附权利要求意欲解释为包括优选实施例以及落入本发明范围的所有变更和修改。

[0133] 显然,本领域的技术人员可以对本发明进行各种改动和变型而不脱离本发明的精神和范围。这样,倘若本发明的这些修改和变型属于本发明权利要求及其等同技术的范围之内,则本发明也意图包含这些改动和变型在内。

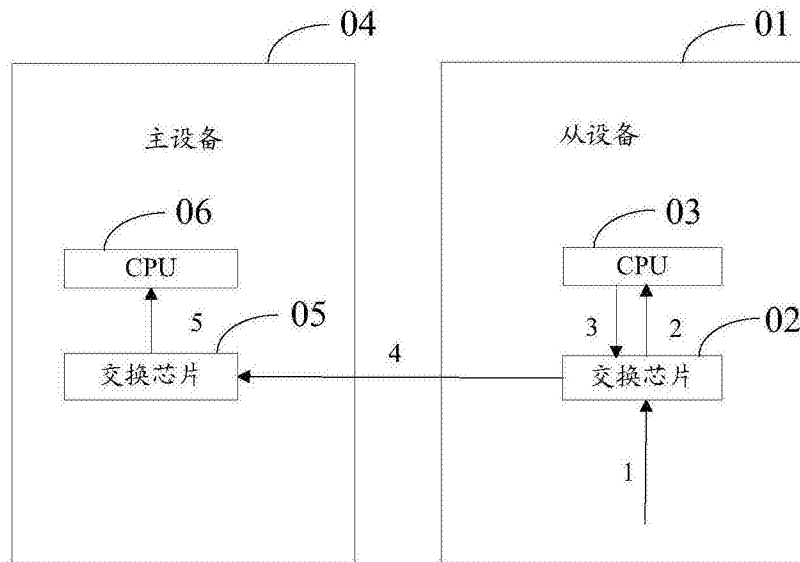


图1

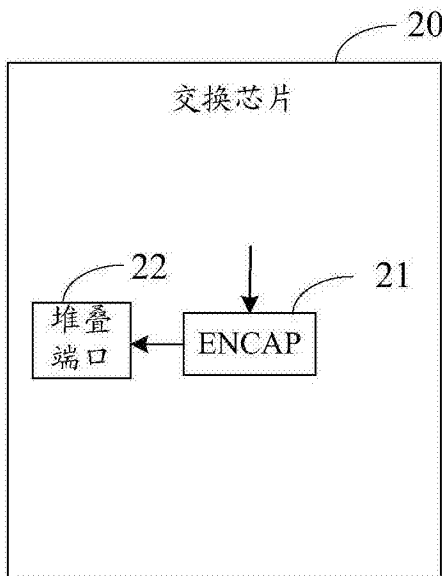


图2

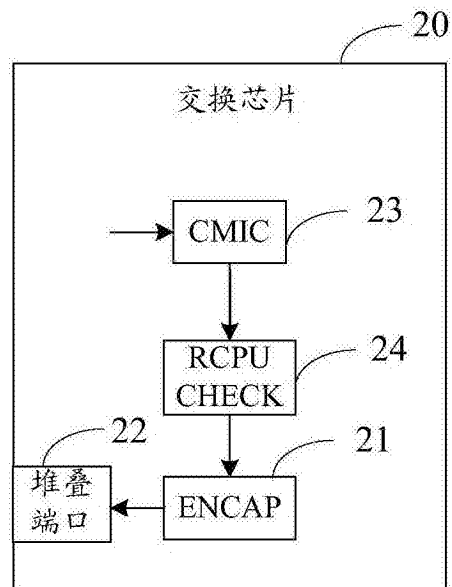


图3

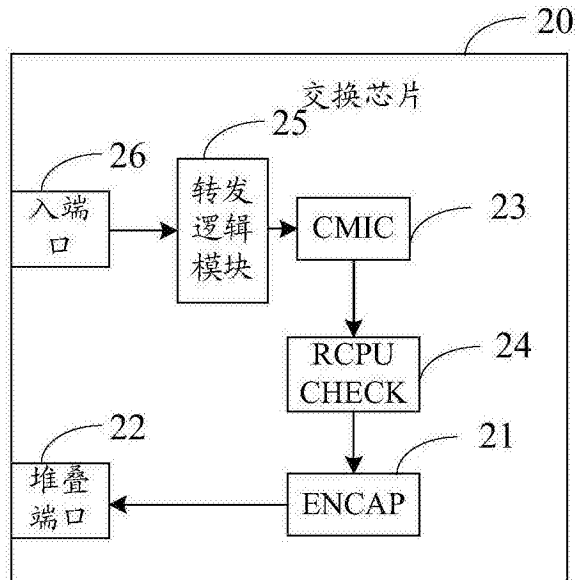


图4

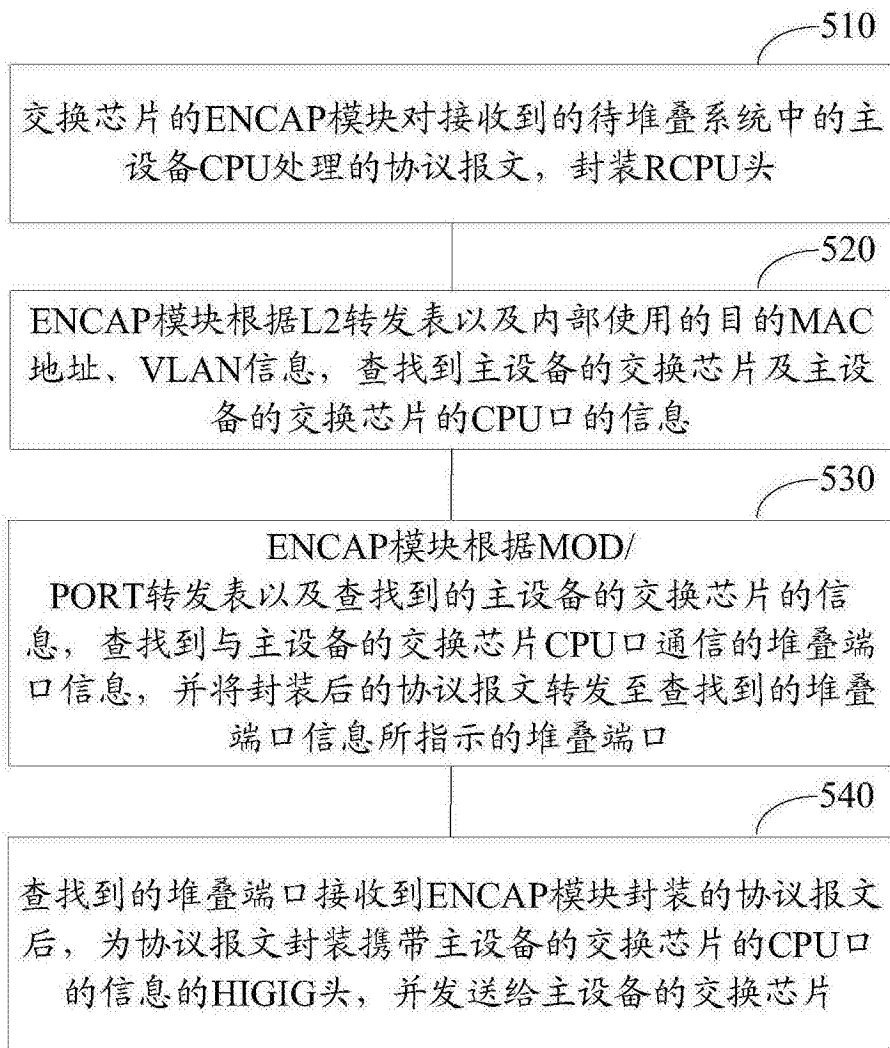


图5

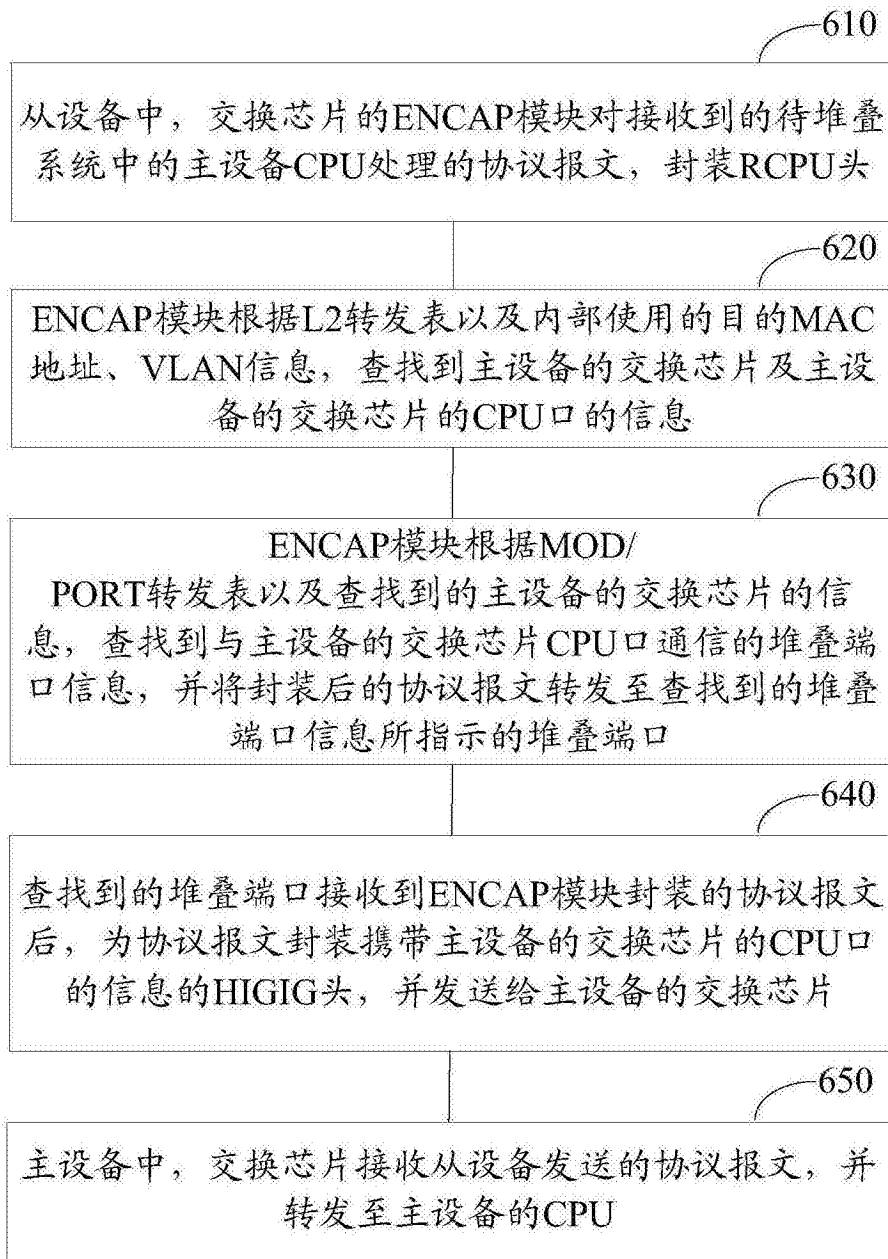


图6

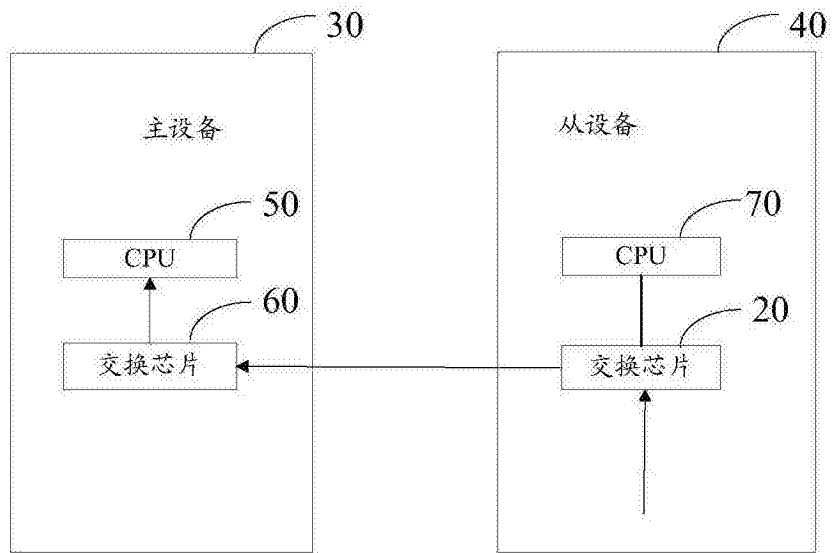


图7