(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 107634905 A
(43)申请公布日 2018.01.26

(21)申请号 201610571283.4

(22)申请日 2016.07.19

(71)申请人 南京中兴新软件有限责任公司
地址 210012 江苏省南京市雨花台区紫荆花路68号

(72)发明人 高月

(74)专利代理机构 北京康信知识产权代理有限责任公司 11240
代理人 江舟 董文倩

(51)Int.Cl.
H04L 12/707(2013.01)
H04L 12/741(2013.01)
H04L 12/775(2013.01)

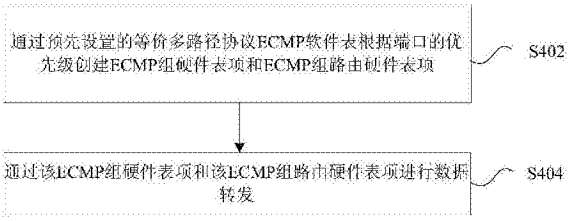
权利要求书3页 说明书10页 附图5页

(54)发明名称

堆叠系统中数据转发方法及装置

(57)摘要

本发明提供了一种堆叠系统中数据转发方法及装置，其中，该方法包括：通过预先设置的等价多路径协议ECMP软件表根据端口的优先级创建ECMP组硬件表项和ECMP组路由硬件表项，其中，堆叠系统中本地端口的优先级大于非本地端口的优先级，通过该ECMP组硬件表项和该ECMP组路由硬件表项进行数据转发，由于优先通过主设备转发，解决了相关技术中在本地端口还可以使用的情况下由于主设备将数据经由堆叠链路传输到备设备进行转发导致数据传输效率低的问题，提高了数据转发的效率。



CN 107634905 A

1.一种堆叠系统中数据转发方法,其特征在于,包括:

通过预先设置的等价多路径协议ECMP软件表根据端口的优先级创建ECMP组硬件表项和ECMP组路由硬件表项,其中,堆叠系统中本地端口的优先级大于非本地端口的优先级,所述ECMP软件表包括所有目的地址相同的等价路由条目和ECMP组硬件表项的索引,所述等价路由条目包括以下信息:下一跳IP地址、出接口和有效标志,所述ECMP组硬件表项是由多条等价路由条目组成的等价路由条目组,所述ECMP组路由硬件表项包括目的IP地址、所述ECMP组硬件表项的索引,其中,所述有效标志用于表示等价路由条目是否有效;

通过所述ECMP组硬件表项和所述ECMP组路由硬件表项进行数据转发。

2.根据权利要求1所述的方法,其特征在于,通过所述ECMP组硬件表项和所述ECMP组路由硬件表项进行数据转发包括:

根据目的IP地址从所述ECMP组路由硬件表查找出与所述目的IP地址匹配的ECMP组硬件表项的索引,根据匹配到的ECMP组硬件表项的索引查找ECMP组硬件表项;

根据所述ECMP组硬件表项中的等价路由条目的信息进行数据转发。

3.根据权利要求2所述的方法,其特征在于,通过预先设置的ECMP软件表根据端口的优先级创建ECMP组硬件表项和ECMP组路由硬件表项包括:

在多条等价路由条目的出接口包括本地端口和非本地端口的情况下,优先选择所述出接口为本地端口的有效等价路由条目创建所述ECMP组硬件表项和所述ECMP组路由硬件表项;

在出接口为本地端口的所有有效等价路由条目均被删除或变为无效的情况下,删除所述ECMP组硬件表项和所述ECMP组路由硬件表项中的所有记录,选择所述出接口为非本地端口的有效等价路由条目添加到所述ECMP组硬件表项和所述ECMP组路由硬件表项中;

在出接口为本地端口的已删除或已失效的等价路由条目重新被添加并有效或恢复为有效的情况下,删除所述ECMP组硬件表项和所述ECMP组路由硬件表项中的所有记录,将所述重新被添加并有效或恢复为有效的出接口为本地端口的有效等价路由条目添加到所述ECMP组硬件表项和所述ECMP组路由硬件表项中。

4.根据权利要求3所述的方法,其特征在于,所述方法还包括:

对所述ECMP组硬件表项和所述ECMP组路由硬件表项的成员进行更新,包括:

向所述ECMP组硬件表项和所述ECMP组路由硬件表项中添加一条等价路由条目;

从所述ECMP组硬件表项和所述ECMP组路由硬件表项中删除一条等价路由条目。

5.根据权利要求4所述的方法,其特征在于,向所述ECMP组硬件表项和所述ECMP组路由硬件表项中添加一条等价路由条目条目包括:

在待添加的等价路由条目为有效路由的情况下,判断所述ECMP组硬件表项和所述ECMP组路由硬件表项中记录的对应出接口;

在所述对应出接口全部为本地端口的情况下,对所述待添加的等价路由条目的出接口进行判断,如果待添加的等价路由条目的出接口为本地端口,将所述待添加的等价路由条目添加到所述ECMP组硬件表项和所述ECMP组路由硬件表项中;

在对应出接口全部为非本地端口的情况下,对所述待添加的等价路由条目的出接口进行判断,如果所述待添加的等价路由条目的出接口为本地端口,删除所述ECMP组硬件表项和所述ECMP组路由硬件表项中的所有记录,将所述待添加的等价路由条目添加到所述ECMP

组硬件表项和所述ECMP组路由硬件表项中。

6.根据权利要求4所述的方法，其特征在于，从所述ECMP组硬件表项和所述ECMP组路由硬件表项中删除一条等价路由条目包括：

在待删除的等价路由条目为有效路由的情况下，对所述ECMP组硬件表项和所述ECMP组路由硬件表项中记录的对应出接口进行判断：

在所述对应出接口全部为本地端口的情况下，对所述待删除的等价路由条目的出接口进行判断，如果所述待删除的等价路由条目的出接口为本地端口，从所述ECMP组硬件表项和所述ECMP组路由硬件表项中删除所述待删除的等价路由条目，如果所述ECMP组硬件表项和所述ECMP组路由硬件表项中的记录已删空，根据所述ECMP组硬件表项和所述ECMP组路由硬件表项中出接口为非本地端口的所有有效记录向所述ECMP组硬件表项和所述ECMP组路由硬件表项中添加记录；

在所述对应出接口全部为非本地端口的情况下，删除所述ECMP组硬件表项和所述ECMP组路由硬件表项中所述待删除的等价路由条目。

7.根据权利要求1至6中任一项所述的方法，其特征在于，所述方法还包括：

删除创建的所述ECMP软件表。

8.一种堆叠系统中数据转发装置，其特征在于，所述装置包括：

创建模块，用于通过预先设置的等价多路径协议ECMP软件表根据端口的优先级创建ECMP组硬件表项和ECMP组路由硬件表项，其中，堆叠系统中本地端口的优先级大于非本地端口的优先级，所述ECMP软件表包括所有目的地址相同的等价路由条目和ECMP组硬件表项的索引，所述等价路由条目包括以下信息：下一跳IP地址、出接口和有效标志，所述ECMP组硬件表项是由多条等价路由条目组成的等价路由条目组，所述ECMP组路由硬件表项包括目的IP地址、所述ECMP组硬件表项的索引，其中，所述有效标志用于表示等价路由条目是否有效；

数据转发模块，用于通过所述ECMP组硬件表项和所述ECMP组路由硬件表项进行数据转发。

9.根据权利要求8所述的装置，其特征在于，所述数据转发模块包括：

查找单元，用于根据目的IP地址从所述ECMP组路由硬件表查找出与所述目的IP地址匹配的ECMP组硬件表项的索引，根据匹配到的ECMP组硬件表项的索引查找ECMP组硬件表项；

数据转发单元，用于根据所述ECMP组硬件表项中的等价路由条目的信息进行数据转发。

10.根据权利要求9所述的装置，其特征在于，所述创建模块包括：

第一创建单元，用于在多条等价路由条目的出接口包括本地端口和非本地端口的情况下，优先选择所述出接口为本地端口的有效等价路由条目创建所述ECMP组硬件表项和所述ECMP组路由硬件表项；

第二创建单元，用于在出接口为本地端口的所有有效等价路由条目均被删除或变为无效的情况下，删除所述ECMP组硬件表项和所述ECMP组路由硬件表项中的所有记录，选择所述出接口为非本地端口的有效等价路由条目添加到所述ECMP组硬件表项和所述ECMP组路由硬件表项中；

第三创建单元，用于在出接口为本地端口的已删除或已失效的等价路由条目重新被添

加并有效或恢复为有效的情况下，删除所述ECMP组硬件表项和所述ECMP组路由硬件表项中的所有记录，将所述重新被添加并有效或恢复为有效的出接口为本地端口的有效等价路由条目添加到所述ECMP组硬件表项和所述ECMP组路由硬件表项中。

堆叠系统中数据转发方法及装置

技术领域

[0001] 本发明涉及通信领域，具体而言，涉及一种堆叠系统中数据转发方法及装置。

背景技术

[0002] 随着网络规模需求的不断增长，尤其是数据中心的规模化应用，使单一网络设备所能提供的物理端口数量不能满足实际网络的部署需求。为此，各类以太网交换设备开始采用多芯片堆叠方式来扩充端口数量。堆叠是一种网络系统虚拟化技术，支持将多台可以单独运行的交换设备组合成单一的逻辑交换设备，即几个物理的成员交换设备相当于一个虚拟的逻辑交换设备，成员交换设备之间的连接称为堆叠链路，成员交换设备之间连接的端口称为堆叠端口，彼此之间通过拓扑发现协议发现对方，并通过一定的机制选择一台主设备和若干台备设备，其中主设备承担整个系统的管理及协议的运行。堆叠技术增加了交换设备的端口数量，即堆叠系统的总端口数是堆叠系统中所有成员设备端口数的总和，以及实现了交换设备之间高速互连和统一管理，从逻辑上来说，它们属于同一个交换设备，只要登录到其中一台交换设备上，就可以对堆叠中的所有交换设备进行管理和配置，方便使用。图1是根据相关技术中堆叠系统的示意图，如图1所示，可以将成员交换设备1和成员交换设备2通过堆叠端口连接构成一个虚拟的交换设备，即堆叠系统(Virtual Switch Cluster，简称为VSC)。

[0003] 在单设备环境中，在通过多条不同链路到达同一目的地址的网络环境中，等价多路径协议(Equal-Cost Multipath Routing protocol，简称为ECMP)可以同时使用多条链路并进行负载分担，这不仅增加了传输带宽，并且实现无延时、无丢包地备份失效链路的数据传输。图2是根据相关技术中的单设备多路径负载分担的示意图，如图2所示，在交换设备1上，存在四条等价路由到达目的网络，基于此，可通过该四条等价路由对访问目的网络的流量进行负载分担，增加了传输带宽，并且，该四条等价路由互为备份，以防止其中的路由出现故障时，其他路由替代该故障路由继续转发流量。可以看出，ECMP最大的特点是实现了等价多路径情况下，多路径负载均衡和链路备份。

[0004] 与单设备环境类似，在堆叠系统中，也可以通过等价多路径协议实现多路径负载均衡和链路备份。但是，在单设备环境中，等价多路径是在所有ECMP成员路径出端口间进行流量分担，而堆叠系统中设备间堆叠链路的带宽十分有限，等价多路径的成员路径出端口可能既包含有本地端口(主设备上的端口)也包含有非本地端口(备设备上的端口)，图3是根据相关技术中堆叠系统中多路径负载分担的示意图，如图3所示，交换设备1与交换设备2组成堆叠系统，在堆叠系统上，存在三条等价路由到达目的网络，路径1和路径2的出接口为本地端口，路径3和路径4的出接口为非本地端口。在这种跨设备等价多路径的堆叠系统中，一般情况下，在主设备的管理下，当堆叠系统接收到数据时，会从ECMP成员路径中选择路径进行数据转发，如果，选择的路径为非本地端口的链路，此时主设备便会将数据经由堆叠链路传输到备设备之后再由备设备进行转发，从而导致数据传输效率低。

[0005] 因此，相关技术中在本地端口还可以使用的情况下由于主设备将数据经由堆叠链

路传输到备设备进行转发导致数据传输效率低的问题，尚未提出解决方案。

发明内容

[0006] 本发明实施例提供了一种堆叠系统中数据转发方法及装置，以至少解决相关技术中在本地端口还可以使用的情况下由于主设备将数据经由堆叠链路传输到备设备进行转发导致数据传输效率低的问题。

[0007] 根据本发明的一个实施例，提供了一种堆叠系统中数据转发方法，包括：通过预先设置的等价多路径协议ECMP软件表根据端口的优先级创建ECMP组硬件表项和ECMP组路由硬件表项，其中，堆叠系统中本地端口的优先级大于非本地端口的优先级，所述ECMP软件表包括所有目的地址相同的等价路由条目和ECMP组硬件表项的索引，所述等价路由条目包括以下信息：下一跳IP地址、出接口和有效标志，所述ECMP组硬件表项是由多条等价路由条目组成的等价路由条目组，所述ECMP组路由硬件表项包括目的IP地址、所述ECMP组硬件表项的索引，其中，所述有效标志用于表示等价路由条目是否有效；通过所述ECMP组硬件表项和所述ECMP组路由硬件表项进行数据转发。

[0008] 优选地，通过所述ECMP组硬件表项和所述ECMP组路由硬件表项进行数据转发包括：根据目的IP地址从所述ECMP组路由硬件表查找出与所述目的IP地址匹配的ECMP组硬件表项的索引，根据匹配到的ECMP组硬件表项的索引查找ECMP组硬件表项；根据所述ECMP组硬件表项中的等价路由条目的信息进行数据转发。

[0009] 优选地，通过预先设置的ECMP软件表根据端口的优先级创建ECMP组硬件表项和ECMP组路由硬件表项包括：在多条等价路由条目的出接口包括本地端口和非本地端口的情况下，优先选择所述出接口为本地端口的有效等价路由条目创建所述ECMP组硬件表项和所述ECMP组路由硬件表项；在出接口为本地端口的所有有效等价路由条目均被删除或变为无效的情况下，删除所述ECMP组硬件表项和所述ECMP组路由硬件表项中的所有记录，选择所述出接口为非本地端口的有效等价路由条目添加到所述ECMP组硬件表项和所述ECMP组路由硬件表项中；在出接口为本地端口的已删除或已失效的等价路由条目重新被添加并有效或恢复为有效的情况下，删除所述ECMP组硬件表项和所述ECMP组路由硬件表项中的所有记录，将所述重新被添加并有效或恢复为有效的出接口为本地端口的有效等价路由条目添加到所述ECMP组硬件表项和所述ECMP组路由硬件表项中。

[0010] 优选地，所述方法还包括：对所述ECMP组硬件表项和所述ECMP组路由硬件表项的成员进行更新，包括：向所述ECMP组硬件表项和所述ECMP组路由硬件表项中添加一条等价路由条目；从所述ECMP组硬件表项和所述ECMP组路由硬件表项中删除一条等价路由条目。

[0011] 优选地，向所述ECMP组硬件表项和所述ECMP组路由硬件表项中添加一条等价路由条目条目包括：在待添加的等价路由条目为有效路由的情况下，判断所述ECMP组硬件表项和所述ECMP组路由硬件表项中记录的对应出接口；在所述对应出接口全部为本地端口的情况下，对所述待添加的等价路由条目的出接口进行判断，如果待添加的等价路由条目的出接口为本地端口，将所述待添加的等价路由条目添加到所述ECMP组硬件表项和所述ECMP组路由硬件表项中；在对应出接口全部为非本地端口的情况下，对所述待添加的等价路由条目的出接口进行判断，如果所述待添加的等价路由条目的出接口为本地端口，删除所述ECMP组硬件表项和所述ECMP组路由硬件表项中的所有记录，将所述待添加的等价路由条目

添加到所述ECMP组硬件表项和所述ECMP组路由硬件表项中。

[0012] 优选地，从所述ECMP组硬件表项和所述ECMP组路由硬件表项中删除一条等价路由条目包括：在待删除的等价路由条目为有效路由的情况下，对所述ECMP组硬件表项和所述ECMP组路由硬件表项中记录的对应出接口进行判断：在所述对应出接口全部为本地端口的情况下，对所述待删除的等价路由条目的出接口进行判断，如果所述待删除的等价路由条目的出接口为本地端口，从所述ECMP组硬件表项和所述ECMP组路由硬件表项中删除所述待删除的等价路由条目，如果所述ECMP组硬件表项和所述ECMP组路由硬件表项中的记录已删空，根据所述ECMP组硬件表项和所述ECMP组路由硬件表项中出接口为非本地端口的所有有效记录向所述ECMP组硬件表项和所述ECMP组路由硬件表项中添加记录；在所述对应出接口全部为非本地端口的情况下，删除所述ECMP组硬件表项和所述ECMP组路由硬件表项中所述待删除的等价路由条目。

[0013] 优选地，所述方法还包括：删除创建的所述ECMP软件表。

[0014] 根据本发明的另一实施例，还提供了一种堆叠系统中数据转发装置，所述装置包括：

[0015] 创建模块，用于通过预先设置的等价多路径协议ECMP软件表根据端口的优先级创建ECMP组硬件表项和ECMP组路由硬件表项，其中，堆叠系统中本地端口的优先级大于非本地端口的优先级，所述ECMP软件表包括所有目的地址相同的等价路由条目和ECMP组硬件表项的索引，所述等价路由条目包括以下信息：下一跳IP地址、出接口和有效标志，所述ECMP组硬件表项是由多条等价路由条目组成的等价路由条目组，所述ECMP组路由硬件表项包括目的IP地址、所述ECMP组硬件表项的索引，其中，所述有效标志用于表示等价路由条目是否有效；

[0016] 数据转发模块，用于通过所述ECMP组硬件表项和所述ECMP组路由硬件表项进行数据转发。

[0017] 优选地，所述数据转发模块包括：

[0018] 查找单元，用于根据目的IP地址从所述ECMP组路由硬件表查找出与所述目的IP地址匹配的ECMP组硬件表项的索引，根据匹配到的ECMP组硬件表项的索引查找ECMP组硬件表项；

[0019] 数据转发单元，用于根据所述ECMP组硬件表项中的等价路由条目的信息进行数据转发。

[0020] 优选地，所述创建模块包括：

[0021] 第一创建单元，用于在多条等价路由条目的出接口包括本地端口和非本地端口的情况下，优先选择所述出接口为本地端口的有效等价路由条目创建所述ECMP组硬件表项和所述ECMP组路由硬件表项；

[0022] 第二创建单元，用于在出接口为本地端口的所有有效等价路由条目均被删除或变为无效的情况下，删除所述ECMP组硬件表项和所述ECMP组路由硬件表项中的所有记录，选择所述出接口为非本地端口的有效等价路由条目添加到所述ECMP组硬件表项和所述ECMP组路由硬件表项中；

[0023] 第三创建单元，用于在出接口为本地端口的已删除或已失效的等价路由条目重新被添加并有效或恢复为有效的情况下，删除所述ECMP组硬件表项和所述ECMP组路由硬件表

项中的所有记录，将所述重新被添加并有效或恢复为有效的出接口为本地端口的有效等价路由条目添加到所述ECMP组硬件表项和所述ECMP组路由硬件表项中。

[0024] 通过本发明，通过预先设置的等价多路径协议ECMP软件表根据端口的优先级创建ECMP组硬件表项和ECMP组路由硬件表项，其中，堆叠系统中本地端口的优先级大于非本地端口的优先级，通过所述ECMP组硬件表项和所述ECMP组路由硬件表项进行数据转发，由于优先通过主设备转发，解决了相关技术中在本地端口还可以使用的情况下由于主设备将数据经由堆叠链路传输到备设备进行转发导致数据传输效率低的问题，提高了数据转发的效率。

附图说明

[0025] 此处所说明的附图用来提供对本发明的进一步理解，构成本申请的一部分，本发明的示意性实施例及其说明用于解释本发明，并不构成对本发明的不当限定。在附图中：

[0026] 图1是根据相关技术中堆叠系统的示意图；

[0027] 图2是根据相关技术中的单设备多路径负载分担的示意图；

[0028] 图3是根据相关技术中堆叠系统中多路径负载分担的示意图；

[0029] 图4是根据本发明实施例的堆叠系统中数据转发方法的流程图；

[0030] 图5是根据本发明实施例中ECMP组创建的流程图；

[0031] 图6是根据本发明实施例中ECMP组成员添加的流程图；

[0032] 图7是根据本发明实施例中ECMP组成员删除的流程图；

[0033] 图8是根据本发明实施例中本地端口和非本地端口共存时的等价多路径策略的示意图；

[0034] 图9是本发明实施例中本地端口链路全部down掉时的等价多路径策略的示意图；

[0035] 图10是根据本发明实施例的堆叠系统中数据转发装置的框图；

[0036] 图11是根据本发明优选实施例的堆叠系统中数据转发装置的框图一；

[0037] 图12是根据本发明优选实施例的堆叠系统中数据转发装置的框图二。

具体实施方式

[0038] 下文中将参考附图并结合实施例来详细说明本发明。需要说明的是，在不冲突的情况下，本申请中的实施例及实施例中的特征可以相互组合。

[0039] 需要说明的是，本发明的说明书和权利要求书及上述附图中的术语“第一”、“第二”等是用于区别类似的对象，而不必用于描述特定的顺序或先后次序。

[0040] 实施例1

[0041] 本发明实施例提供了一种堆叠系统中等价多路径的方法，用以减少堆叠系统中跨设备等价多路径进行数据转发时对堆叠端口造成的带宽压力。该方法应用于包括本端设备和对端设备的网络环境中，该本端设备是由两台或两台以上通过堆叠端口相连的堆叠成员设备所组成的堆叠系统，该本端设备与该对端设备之间存在多条等价路径，该方法的实现如下：在本端堆叠系统中，用软件表记录下ECMP相关信息，称为ECMP软件表，主要包括：去往同一目的地址所包含的所有等价路由条目，该等价路由条目包含下一跳IP地址、出接口和有效标志等信息，ECMP组硬件表项的索引，该ECMP组硬件表项包含该所有等价路由条目经

过本发明方法选出的其中的若干条等价路由条目组成的等价路由组，该等价路由组中包含的等价路由条目的数目。该有效标志表示相应的等价路由条目是否有效，等价路由条目可能出于多种原因而失效，包括等价路由条目的出接口down掉，连接到对端的链路的故障(例如，线路中断)，对端端口down掉等。

[0042] 根据该ECMP相关信息创建该ECMP软件表。对该ECMP组中包含的每一条等价路由条目进行判断，找出所有有效的等价路由条目。在该所有有效的等价路由条目中，对每一条等价路由条目的出接口进行判断，则，或者为本地端口，或者为非本地端口，有以下三种情况：

[0043] 全部为本地端口：在该所有有效的等价路由条目中，用所有有效等价路由条目创建ECMP组硬件表项以及ECMP组路由硬件表项，该ECMP组路由硬件表项包含该目的IP地址、该ECMP组硬件表项的索引等信息。更新该ECMP软件表及其他相关软件表。

[0044] 全部为非本地端口：在该所有有效的等价路由条目中，用所有有效路由条目创建ECMP组硬件表项以及ECMP组路由表项。更新该ECMP软件表及其他相关软件表。

[0045] 部分为本地端口，部分为非本地端口：在该所有有效的等价路由条目中，用其中具有本地端口的所有有效路由条目创建ECMP组硬件表项以及ECMP组路由表项。更新该ECMP软件表及其他相关软件表。

[0046] 在本实施例中提供了一种堆叠系统中数据转发方法，图4是根据本发明实施例的堆叠系统中数据转发方法的流程图，如图4所示，该流程包括如下步骤：

[0047] 步骤S402，通过预先设置的等价多路径协议ECMP软件表根据端口的优先级创建ECMP组硬件表项和ECMP组路由硬件表项，其中，堆叠系统中本地端口的优先级大于非本地端口的优先级，该ECMP软件表包括所有目的地址相同的等价路由条目和ECMP组硬件表项的索引，该等价路由条目包括以下信息：下一跳IP地址、出接口和有效标志，该ECMP组硬件表项是由多条等价路由条目组成的等价路由条目组，该ECMP组路由硬件表项包括目的IP地址、该ECMP组硬件表项的索引，其中，该有效标志用于表示等价路由条目是否有效；

[0048] 步骤S404，通过该ECMP组硬件表项和该ECMP组路由硬件表项进行数据转发。

[0049] 通过上述步骤，通过预先设置的等价多路径协议ECMP软件表根据端口的优先级创建ECMP组硬件表项和ECMP组路由硬件表项，其中，堆叠系统中本地端口的优先级大于非本地端口的优先级，通过所述ECMP组硬件表项和所述ECMP组路由硬件表项进行数据转发，由于优先通过主设备转发，解决了相关技术中在本地端口还可以使用的情况下由于主设备将数据经由堆叠链路传输到备设备进行转发导致数据传输效率低的问题，提高了数据转发的效率。

[0050] 优选地，通过该ECMP组硬件表项和该ECMP组路由硬件表项进行数据转发可以包括：根据目的IP地址从该ECMP组路由硬件表查找出与该目的IP地址匹配的ECMP组硬件表项的索引，根据匹配到的ECMP组硬件表项的索引查找ECMP组硬件表项；根据该ECMP组硬件表项中的等价路由条目的信息进行数据转发。

[0051] 优选地，通过预先设置的ECMP软件表根据端口的优先级创建ECMP组硬件表项和ECMP组路由硬件表项可以包括：在多条等价路由条目的出接口包括本地端口和非本地端口的情况下，优先选择该出接口为本地端口的有效等价路由条目创建该ECMP组硬件表项和该ECMP组路由硬件表项；在出接口为本地端口的所有有效等价路由条目均被删除或变为无效的情况下，删除该ECMP组硬件表项和该ECMP组路由硬件表项中的所有记录，选择该出接口

为非本地端口的有效等价路由条目添加到该ECMP组硬件表项和该ECMP组路由硬件表项中；在出接口为本地端口的已删除或已失效的等价路由条目重新被添加并有效或恢复为有效的情况下，删除该ECMP组硬件表项和该ECMP组路由硬件表项中的所有记录，将该重新被添加并有效或恢复为有效的出接口为本地端口的有效等价路由条目添加到该ECMP组硬件表项和该ECMP组路由硬件表项中。

[0052] 优选地，还可以对该ECMP组硬件表项和该ECMP组路由硬件表项的成员进行更新，具体包括：向该ECMP组硬件表项和该ECMP组路由硬件表项中添加一条等价路由条目；从该ECMP组硬件表项和该ECMP组路由硬件表项中删除一条等价路由条目。

[0053] 优选地，向该ECMP组硬件表项和该ECMP组路由硬件表项中添加一条等价路由条目条目包括：在待添加的等价路由条目为有效路由的情况下，判断该ECMP组硬件表项和该ECMP组路由硬件表项中记录的对应出接口；在该对应出接口全部为本地端口的情况下，对该待添加的等价路由条目的出接口进行判断，如果待添加的等价路由条目的出接口为本地端口，将该待添加的等价路由条目添加到该ECMP组硬件表项和该ECMP组路由硬件表项中；在对应出接口全部为非本地端口的情况下，对该待添加的等价路由条目的出接口进行判断，如果该待添加的等价路由条目的出接口为本地端口，删除该ECMP组硬件表项和该ECMP组路由硬件表项中的所有记录，将该待添加的等价路由条目添加到该ECMP组硬件表项和该ECMP组路由硬件表项中。

[0054] 对于ECMP组成员的添加，即向该ECMP组中添加一条等价路由条目。具体的，根据该添加的一条等价路由条目向该ECMP软件表中添加一条记录。对该添加的一条等价路由条目进行判断，如果该添加的一条路由条目是有效的路由条目，则对该ECMP组硬件表项中的记录对应的出接口进行判断，则有两种情况，或者全部为本地端口，或者全部为非本地端口：全部为本地端口：对该添加的一条有效的等价路由条目的出接口进行判断，如果出接口为本地端口，则根据该条有效的等价路由条目向该ECMP组硬件表项添加一条记录。全部为非本地端口：对该添加的一条有效的等价路由条目的出接口进行判断，如果出接口为本地端口，则删除该ECMP组硬件表项中的所有记录，否则不执行删除。根据该添加的一条有效的等价路由条目向该ECMP组硬件表项添加一条记录。

[0055] 优选地，从该ECMP组硬件表项和该ECMP组路由硬件表项中删除一条等价路由条目包括：在待删除的等价路由条目为有效路由的情况下，对该ECMP组硬件表项和该ECMP组路由硬件表项中记录的对应出接口进行判断：在该对应出接口全部为本地端口的情况下，对该待删除的等价路由条目的出接口进行判断，如果该待删除的等价路由条目的出接口为本地端口，从该ECMP组硬件表项和该ECMP组路由硬件表项中删除该待删除的等价路由条目，如果该ECMP组硬件表项和该ECMP组路由硬件表项中的记录已删空，根据该ECMP组硬件表项和该ECMP组路由硬件表项中出接口为非本地端口的所有有效记录向该ECMP组硬件表项和该ECMP组路由硬件表项中添加记录；在该对应出接口全部为非本地端口的情况下，删除该ECMP组硬件表项和该ECMP组路由硬件表项中该待删除的等价路由条目。

[0056] 对于ECMP组成员的删除，即向该ECMP组中删除一条等价路由条目。具体的，根据该删除的一条等价路由条目向该ECMP软件表中删除一条记录。对该删除的一条等价路由条目进行判断，如果该删除的一条等价路由条目是有效的路由条目，则对该ECMP组硬件表项中的记录对应的出接口进行判断，则有两种情况，或者全部为本地端口，或者全部为非本地端

口:全部为本地端口:对该删除的一条有效的路由条目的出接口进行判断,如果出接口为本地端口,则根据该条有效的等价路由条目向该ECMP组硬件表项删除一条记录。如果该ECMP组硬件表项中的记录已删空,则根据该ECMP软件表中的,出接口是非本地端口的所有有效记录向该ECMP组硬件表项中添加记录。全部为非本地端口:根据该删除的一条有效的路由条目向该ECMP组硬件表项删除一条记录。

[0057] 优选地,该方法还包括:删除创建的该ECMP软件表。

[0058] 本发明实施例提供的一种堆叠系统中等价多路径的方法,该方法应用于包括本端设备和对端设备的网络中,该本端设备是由两台以上通过堆叠端口相连的堆叠成员设备所组成的堆叠系统,该本端设备与该对端设备之间存在多条等价路径。在本端堆叠系统中,在通过有效的多条等价路由转发数据时,采用本地优先的原则,即在多条等价路由的出接口既包含本地端口又包含非本地端口的情况下,优先选择其中出接口是本地端口的有效等价路由创建ECMP组硬件表项,并通过该ECMP组进行数据的转发。其次,只有在其中出接口是本地端口的所有有效等价路由都被删除或者变为无效,删除该ECMP组硬件表项中的所有记录,这时才会选择其中出接口是非本地端口的有效等价路由添加到ECMP组硬件表项中,并通过该ECMP组进行数据的转发。另外,当其中出接口是本地端口的已删除或者已失效的等价路由重新被添加并有效或者恢复为有效时,则删除该ECMP组硬件表项中的所有记录,用该重新被添加并有效或者恢复为有效的出接口是本地端口的有效等价路由添加到该ECMP组硬件表项中,并通过该ECMP组进行数据的转发。可以大大减少堆叠系统中跨设备等价多路径数据转发对堆叠端口造成的带宽压力。

[0059] 基于堆叠系统的等价多路径方法中,在通过有效的多条等价路由转发数据时,采用本地优先的原则,即在多条等价路由的出接口既包含本地端口又包含非本地端口的情况下,优先选择其中出接口是本地端口的有效等价路由创建ECMP组硬件表项,并通过该ECMP组进行数据的转发。其次,只有在其中出接口是本地端口的所有有效等价路由都被删除或者变为无效,删除该ECMP组硬件表项中的所有记录,这时才会选择其中出接口是非本地端口的有效等价路由添加到ECMP组硬件表项中,并通过该ECMP组进行数据的转发。另外,当其中出接口是本地端口的已删除或者已失效的等价路由重新被添加并有效或者恢复为有效时,则删除该ECMP组硬件表项中的所有记录,用该重新被添加并有效或者恢复为有效的出接口是本地端口的有效等价路由添加到该ECMP组硬件表项中,并通过该ECMP组进行数据的转发。

[0060] 图5是根据本发明实施例中ECMP组创建的流程图,如图5所示,ECMP组创建的流程,该流程包括以下步骤:

[0061] 步骤S502:创建ECMP软件表,记录ECMP组所包含的所有等价路由条目,其中既包含有效的等价路由条目又包含无效的等价路由条目。

[0062] 步骤S504:找出其中所有有效的等价路由条目。

[0063] 步骤S506:在所有有效的等价路由条目中,出接口既有本地端口又有非本地端口。根据本地优先的原则,从中优先选出出接口是本地端口的有效等价路由条目,并用所选出的有效等价路由条目创建ECMP组硬件表项以及ECMP组路由表项。

[0064] 图6是根据本发明实施例中ECMP组成员添加的流程图,如图6所示,ECMP组成员添加的流程,需要添加的一条等价路由条目是有效的,其出接口为本地端口,且ECMP组硬件表

项中的记录的出接口都是非本地的。该流程包括以下步骤：

[0065] 步骤S602：根据需要添加的一条等价路由条目向ECMP软件表中添加一条记录。

[0066] 步骤S604：根据本地优先的原则，删除ECMP组硬件表项中所有记录，根据需要添加的一条有效的等价路由条目向ECMP组硬件表项添加一条记录。

[0067] 图7是根据本发明实施例中ECMP组成员删除的流程图，如图7所示，ECMP组成员删除的流程，需要删除的一条等价路由条目是有效的，其出接口为本地端口，且相应的ECMP组硬件表项中的记录只有一条。该流程包括以下步骤：

[0068] 步骤S702：根据需要删除的一条等价路由条目向ECMP软件表中删除一条记录。

[0069] 步骤S704：根据需要删除的该条有效的等价路由条目向ECMP组硬件表项删除一条记录，此时ECMP组硬件表项已被删空，根据本地优先的原则，在ECMP软件表中找出出接口为非本地端口的所有有效记录向ECMP组硬件表项添加记录。

[0070] 下面结合具体实施例对本发明做进一步详细说明。本实施例中，交换设备1(主设备)和交换设备2(备设备)通过一条堆叠链路组成堆叠系统，去往目的网络存在四条有效的等价路径，其中路径1和路径2的出接口为本地端口(主设备端口)，路径3和路径4的出接口为非本地端口(备设备端口)。

[0071] 图8是根据本发明实施例中本地端口和非本地端口共存时的等价多路径策略的示意图，如图8所示，根据本地优先的原则，优先选择出接口为本地端口的路径1和路径2创建ECMP组硬件表项，并通过该ECMP组进行流量分担，最终到达目的网络。

[0072] 图9是本发明实施例中本地端口链路全部down掉时的等价多路径策略的示意图，如图9所示，根据本地优先的原则，由于路径1和路径2全部down掉，所以从ECMP组硬件表项中删除相应的记录，选择出接口为非本地端口的路径3和路径4添加到ECMP组硬件表项中，并通过该ECMP组进行流量分担，最终到达目的网络。

[0073] 本发明实施例本着本地优先的原则，优先地选择出接口为本地端口的等价路径创建ECMP组，大大减少了跨设备的流量传输，减少了对堆叠端口的带宽压力。

[0074] 实施例2

[0075] 在本实施例中还提供了一种路径建立装置，该装置用于实现上述实施例及优选实施方式，已经进行过说明的不再赘述。如以下所使用的，术语“模块”可以实现预定功能的软件和/或硬件的组合。尽管以下实施例所描述的装置较佳地以软件来实现，但是硬件，或者软件和硬件的组合的实现也是可能并被构想的。

[0076] 本发明实施例还提供了一种堆叠系统中数据转发装置，图10是根据本发明实施例的堆叠系统中数据转发装置的框图，如图10所示，该装置包括：

[0077] 创建模块102，用于通过预先设置的等价多路径协议ECMP软件表根据端口的优先级创建ECMP组硬件表项和ECMP组路由硬件表项，其中，堆叠系统中本地端口的优先级大于非本地端口的优先级，该ECMP软件表包括所有目的地址相同的等价路由条目和ECMP组硬件表项的索引，该等价路由条目包括以下信息：下一跳IP地址、出接口和有效标志，该ECMP组硬件表项是由多条等价路由条目组成的等价路由条目组，该ECMP组路由硬件表项包括目的IP地址、该ECMP组硬件表项的索引，其中，该有效标志用于表示等价路由条目是否有效；

[0078] 数据转发模块104，用于通过该ECMP组硬件表项和该ECMP组路由硬件表项进行数据转发。

[0079] 图11是根据本发明优选实施例的堆叠系统中数据转发装置的框图一，如图11所示，数据转发模块104包括：

[0080] 查找单元112，用于根据目的IP地址从该ECMP组路由硬件表查找出与该目的IP地址匹配的ECMP组硬件表项的索引，根据匹配到的ECMP组硬件表项的索引查找ECMP组硬件表项；

[0081] 数据转发单元114，用于根据该ECMP组硬件表项中的等价路由条目的信息进行数据转发。

[0082] 图12是根据本发明优选实施例的堆叠系统中数据转发装置的框图二，如图12所示，创建模块102包括：

[0083] 第一创建单元122，用于在多条等价路由条目的出接口包括本地端口和非本地端口的情况下，优先选择该出接口为本地端口的有效等价路由条目创建该ECMP组硬件表项和该ECMP组路由硬件表项；

[0084] 第二创建单元124，用于在出接口为本地端口的所有有效等价路由条目均被删除或变为无效的情况下，删除该ECMP组硬件表项和该ECMP组路由硬件表项中的所有记录，选择该出接口为非本地端口的有效等价路由条目添加到该ECMP组硬件表项和该ECMP组路由硬件表项中；

[0085] 第三创建单元126，用于在出接口为本地端口的已删除或已失效的等价路由条目重新被添加并有效或恢复为有效的情况下，删除该ECMP组硬件表项和该ECMP组路由硬件表项中的所有记录，将该重新被添加并有效或恢复为有效的出接口为本地端口的有效等价路由条目添加到该ECMP组硬件表项和该ECMP组路由硬件表项中。

[0086] 需要说明的是，上述各个模块是可以通过软件或硬件来实现的，对于后者，可以通过以下方式实现，但不限于此：上述模块均位于同一处理器中；或者，上述各个模块以任意组合的形式分别位于不同的处理器中。

[0087] 实施例3

[0088] 本发明的实施例还提供了一种存储介质。可选地，在本实施例中，上述存储介质可以被设置为存储用于执行以下步骤的程序代码：

[0089] 步骤S1，通过预先设置的等价多路径协议ECMP软件表根据端口的优先级创建ECMP组硬件表项和ECMP组路由硬件表项，其中，堆叠系统中本地端口的优先级大于非本地端口的优先级，该ECMP软件表包括所有目的地址相同的等价路由条目和ECMP组硬件表项的索引，该等价路由条目包括以下信息：下一跳IP地址、出接口和有效标志，该ECMP组硬件表项是由多条等价路由条目组成的等价路由条目组，该ECMP组路由硬件表项包括目的IP地址、该ECMP组硬件表项的索引，其中，该有效标志用于表示等价路由条目是否有效；

[0090] 步骤S2，通过该ECMP组硬件表项和该ECMP组路由硬件表项进行数据转发。

[0091] 可选地，存储介质还被设置为存储用于执行以下步骤的程序代码：根据目的IP地址从该ECMP组路由硬件表查找出与该目的IP地址匹配的ECMP组硬件表项的索引，根据匹配到的ECMP组硬件表项的索引查找ECMP组硬件表项；根据该ECMP组硬件表项中的等价路由条目的信息进行数据转发。

[0092] 可选地，存储介质还被设置为存储用于执行以下步骤的程序代码：

[0093] 对该ECMP组硬件表项和该ECMP组路由硬件表项的成员进行更新，包括：向该ECMP

组硬件表项和该ECMP组路由硬件表项中添加一条等价路由条目;从该ECMP组硬件表项和该ECMP组路由硬件表项中删除一条等价路由条目。

[0094] 可选地,在本实施例中,上述存储介质可以包括但不限于:U盘、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、移动硬盘、磁碟或者光盘等各种可以存储程序代码的介质。

[0095] 可选地,本实施例中的具体示例可以参考上述实施例及可选实施方式中所描述的示例,本实施例在此不再赘述。

[0096] 显然,本领域的技术人员应该明白,上述的本发明的各模块或各步骤可以用通用的计算装置来实现,它们可以集中在单个的计算装置上,或者分布在多个计算装置所组成的网络上,可选地,它们可以用计算装置可执行的程序代码来实现,从而,可以将它们存储在存储装置中由计算装置来执行,并且在某些情况下,可以以不同于此处的顺序执行所示出或描述的步骤,或者将它们分别制作成各个集成电路模块,或者将它们中的多个模块或步骤制作成单个集成电路模块来实现。这样,本发明不限制于任何特定的硬件和软件结合。

[0097] 以上所述仅为本发明的优选实施例而已,并不用于限制本发明,对于本领域的技术人员来说,本发明可以有各种更改和变化。凡在本发明的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。

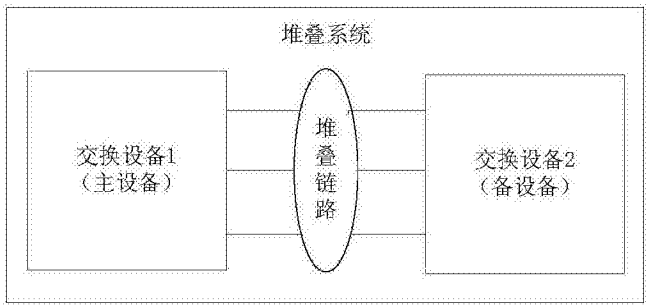

堆叠系统
交换设备1
（主设备）
堆
叠
链
路
交换设备2
（备设备）

图1

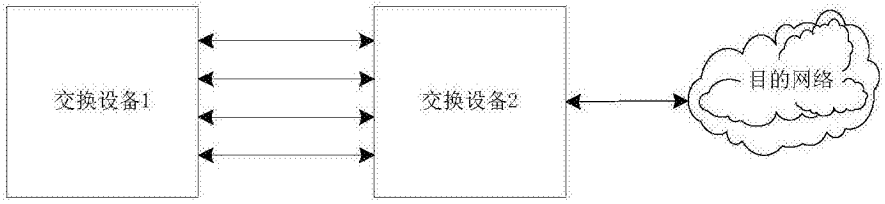

交换设备1
交换设备2
目的网络

图2

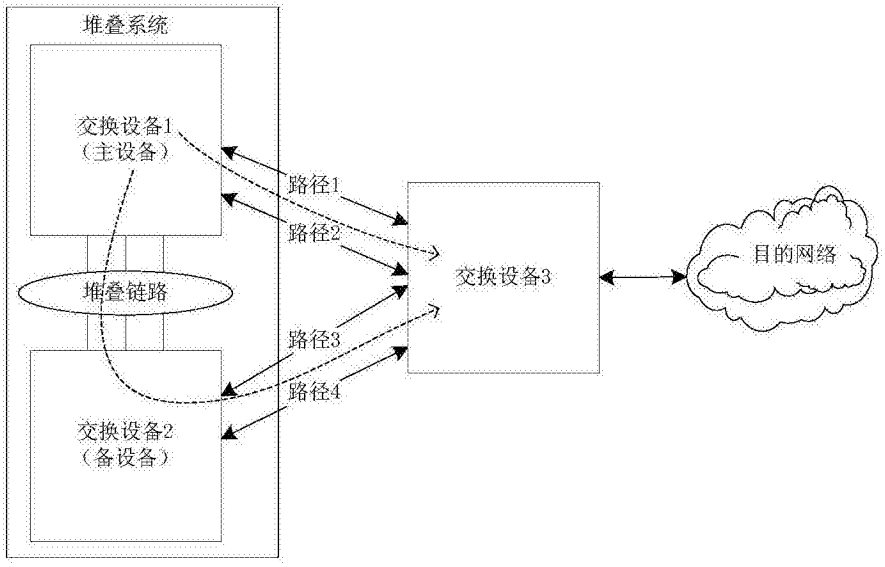

堆叠系统
交换设备1
（主设备）
路径1
路径2
堆叠链路
交换设备3
目的网络
路径3
路径4
交换设备2
（备设备）

图3

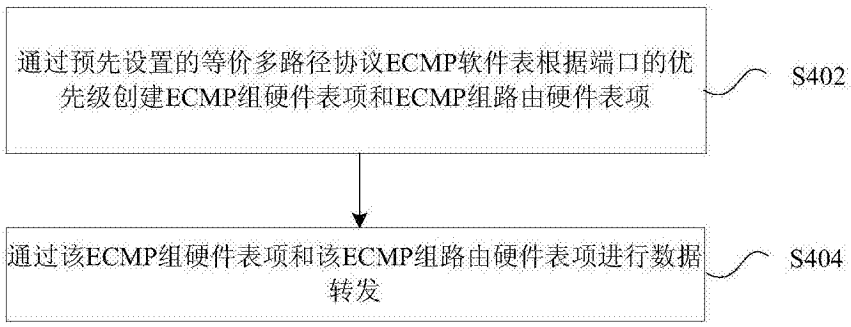

通过预先设置的等价多路径协议ECMP软件表根据端口的优先级创建ECMP组硬件表项和ECMP组路由硬件表项
S402
通过该ECMP组硬件表项和该ECMP组路由硬件表项进行数据转发
S404

图4

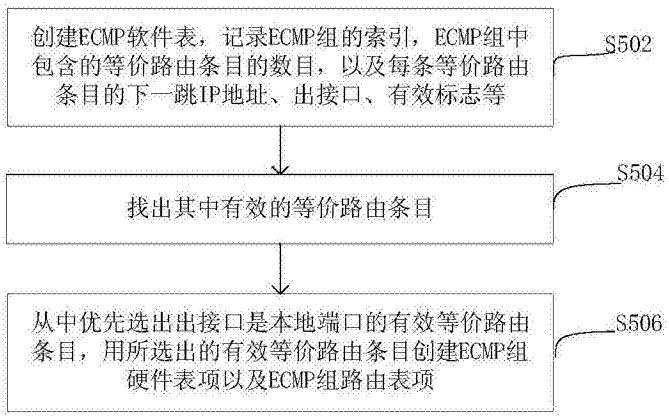

创建ECMP软件表，记录ECMP组的索引，ECMP组中包含的等价路由条目的数目，以及每条等价路由条目的下一跳IP地址、出接口、有效标志等
S502
找出其中有效的等价路由条目
S504
从中优先选出出接口是本地端口的有效等价路由条目，用所选出的有效等价路由条目创建ECMP组硬件表项以及ECMP组路由表项
S506

图5

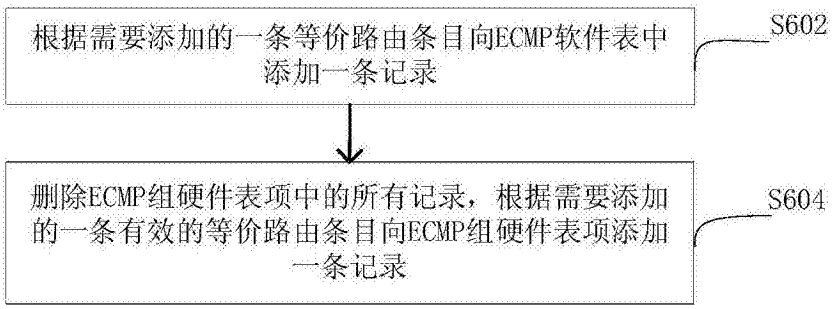

根据需要添加的一条等价路由条目向ECMP软件表中添加一条记录
S602
删除ECMP组硬件表项中的所有记录，根据需要添加的一条有效的等价路由条目向ECMP组硬件表项添加一条记录
S604

图6

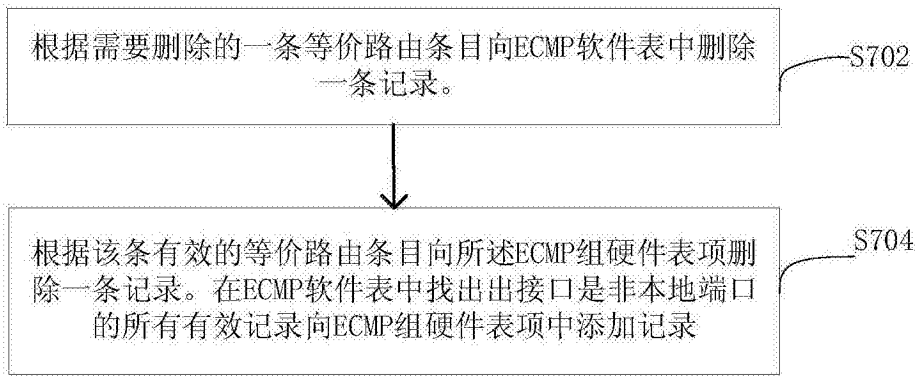

根据需要删除的一条等价路由条目向ECMP软件表中删除一条记录。
S702
根据该条有效的等价路由条目向所述ECMP组硬件表项删除一条记录。在ECMP软件表中找出出接口是非本地端口的所有有效记录向ECMP组硬件表项中添加记录
S704

图7

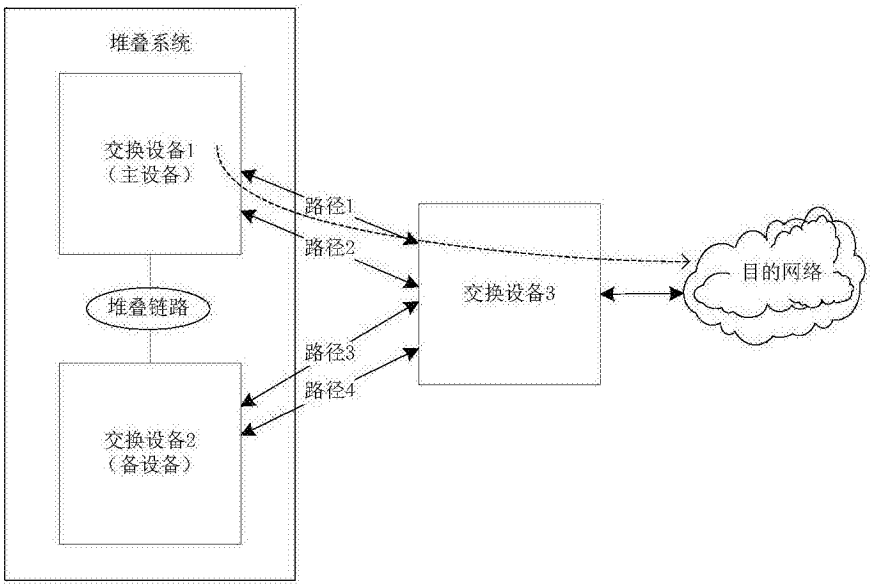

堆叠系统
交换设备1
（主设备）
堆叠链路
交换设备2
（备设备）
路径1
路径2
路径3
路径4
交换设备3
目的网络

图8

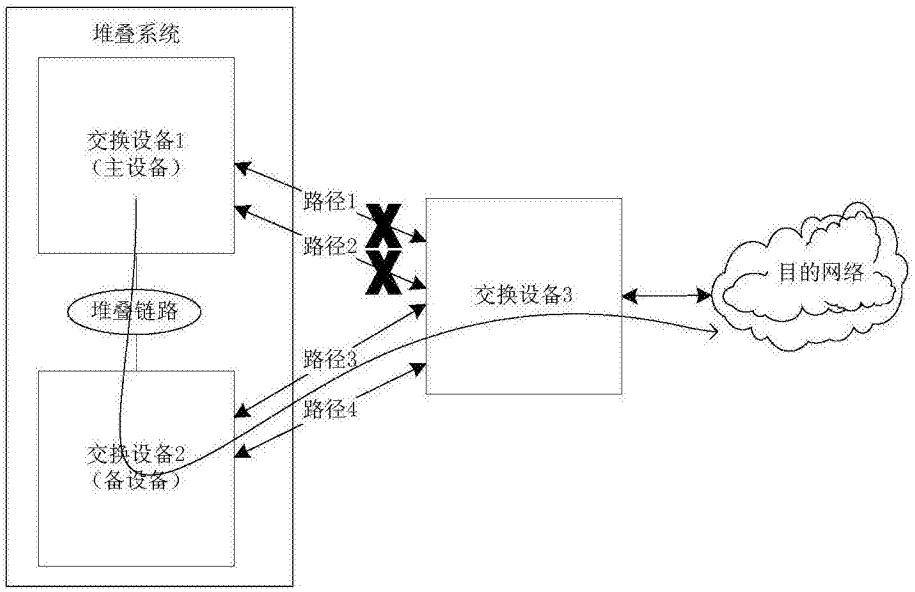

堆叠系统
交换设备1
（主设备）
堆叠链路
交换设备2
（备设备）
路径1
路径2
路径3
路径4
交换设备3
目的网络

图9


创建模块102
数据转发模块104

图10


创建模块102
数据转发模块104
查找单元112
数据转发单元114

图11

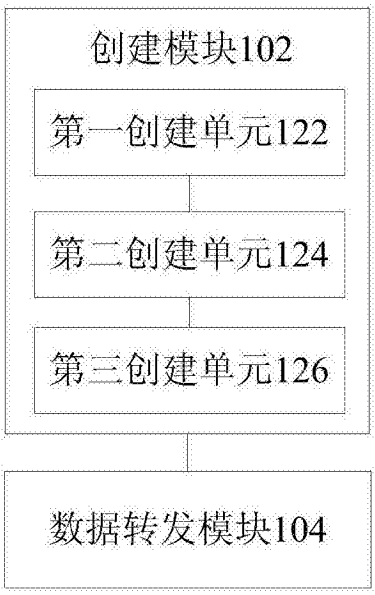


图12