



(12)发明专利申请

(10)申请公布号 CN 110545240 A
(43)申请公布日 2019.12.06

(21)申请号 201910711301.8

(22)申请日 2019.08.02

(71)申请人 新华三大数据技术有限公司
地址 450000 河南省郑州市高新技术产业
开发区杜英街166号总部大观B18号楼

(72)发明人 郭威

(74)专利代理机构 北京三聚阳光知识产权代理
有限公司 11250
代理人 张琳琳

(51) Int. Cl.
H04L 12/723(2013.01)
H04L 12/709(2013.01)

权利要求书2页 说明书13页 附图7页

(54)发明名称

基于分布式聚合系统的标签转发表的建立及报文转发方法

(57)摘要

本公开涉及网络通信技术领域,具体涉及基于分布式聚合系统的标签转发表的建立方法及报文转发方法,其中标签转发表的建立方法包括获取本地的第一标签转发表;其中,所述第一标签转发表用于表示下一跳节点与标签的映射关系;获取对端为下一跳节点分配的入标签;其中,对端与本地属于同一分布式聚合系统;从第一标签转发表中提取本地为下一跳节点分配的出标签;基于入标签以及出标签,建立第二标签转发表。通过获取对端为下一跳节点分配的入标签,即本地存储有对端为下一跳节点分配的入标签,当本地所获取到的报文中携带有该入标签时,本地通过在标签转发表中能够查找到该入标签,实现报文的正常转发,从而解决了标签转发与DRNI网络不兼容的问题。



1. 一种基于分布式聚合系统的标签转发表的建立方法,其特征在于,包括:
 - 获取本地的第一标签转发表;其中,所述第一标签转发表用于表示下一跳节点与出标签的映射关系;
 - 获取对端为所述下一跳节点分配的入标签;其中,所述对端与本地属于同一所述分布式聚合系统;
 - 从所述第一标签转发表中提取本地为所述下一跳节点分配的出标签;
 - 基于所述入标签以及所述出标签,建立第二标签转发表。
2. 根据权利要求1所述的方法,其特征在于,所述基于所述入标签以及所述出标签,建立第二标签转发表,包括:
 - 从所述第一标签转发表中提取本地对应于所述下一跳节点的出方向接口;
 - 建立所述入标签、所述出标签、所述出方向接口与所述下一跳节点的映射关系,以得到所述第二标签转发表。
3. 根据权利要求1所述的方法,其特征在于,所述获取对端为所述下一跳节点分配的入标签,包括:
 - 向所述对端发送标签转发表请求消息;其中,所述标签转发表请求消息用于请求所述对端对应于所述下一跳节点的第三标签转发表;
 - 接收所述对端发送的所述第三标签转发表;
 - 提取所述第三标签转发表中的入标签,以得到所述对端为所述下一跳节点分配的入标签。
4. 根据权利要求1-3中任一项所述的方法,其特征在于,所述获取本地对应于下一跳节点的第一标签转发表的步骤之前,还包括:
 - 与所述对端交互Keepalive报文,以确定本地与所述对端之间的所述分布式聚合系统的邻居关系。
5. 一种基于分布式聚合系统的报文转发方法,其特征在于,包括:
 - 获取待转发报文;其中,所述待转发报文中携带有待匹配标签,所述待匹配标签用于确定与其对应的出标签,所述出标签用于表示本地的下一跳节点;
 - 在标签转发表中查找所述待匹配标签,以确定所述下一跳节点;其中,所述标签转发表是根据权利要求1-4中任一项所述的标签转发表的建立方法所建立的;
 - 向所述下一跳节点转发所述待转发报文。
6. 根据权利要求5所述的方法,其特征在于,所述在标签转发表中查找所述待匹配标签,以确定所述待转发报文的下一跳节点,包括:
 - 判断在所述第一标签转发表以及所述第二标签转发表中是否均存在与所述待匹配标签相同的标签;
 - 当在所述第一标签转发表以及所述第二标签转发表中均存在与所述待匹配标签相同的标签时,判断所述待转发报文是否为预设物理口所接收到的;
 - 当所述待转发报文为预设物理口所接收到的时,在所述第二标签转发表中查找所述待匹配标签,以确定所述待转发报文的下一跳节点。
7. 根据权利要求6所述的方法,其特征在于,所述在标签转发表中查找所述待匹配标签,以确定所述待转发报文的下一跳节点,还包括:

当所述待转发报文不是所述预设物理口所接收到的时,在所述第一标签转发表中查找所述待匹配标签,以确定所述待转发报文的下一跳节点。

8.一种基于分布式聚合系统的标签转发表的建立装置,其特征在于,包括:

第一获取模块,用于获取本地的第一标签转发表;其中,所述第一标签转发表用于表示下一跳节点与出标签的映射关系;

第二获取模块,用于获取对端为所述下一跳节点分配的入标签;其中,所述对端与本地属于同一所述分布式聚合系统;

出标签提取模块,用于从所述第一标签转发表中提取本地为所述下一跳节点分配的出标签;

建立模块,基于所述入标签以及所述出标签,建立第二标签转发表。

9.一种基于分布式聚合系统的报文转发装置,其特征在于,包括:

第三获取模块,用于获取待转发报文;其中,所述待转发报文中携带有待匹配标签,所述待匹配标签用于确定与其对应的出标签,所述出标签用于表示本地的下一跳节点;

查找模块,用于在标签转发表中查找所述待匹配标签,以确定所述下一跳节点;其中,所述标签转发表是根据权利要求1-4中任一项所述的标签转发表的建立方法所建立的;

报文转发模块,用于向所述下一跳节点转发所述待转发报文。

10.一种网络设备,其特征在于,包括:

存储器和处理器,所述存储器和所述处理器之间互相通信连接,所述存储器中存储有计算机指令,所述处理器通过执行所述计算机指令,从而执行权利要求1-4中任一项所述的基于分布式聚合系统的标签转发表的建立方法,或权利要求5-7中任一项所述的基于分布式聚合系统的报文转发方法。

11.一种计算机可读存储介质,其特征在于,所述计算机可读存储介质存储有计算机指令,所述计算机指令用于使所述计算机执行权利要求1-4中任一项所述的基于分布式聚合系统的标签转发表的建立方法,或权利要求5-7中任一项所述的基于分布式聚合系统的报文转发方法。

基于分布式聚合系统的标签转发表的建立及报文转发方法

技术领域

[0001] 本公开涉及网络通信技术领域,具体涉及基于分布式聚合系统的标签转发表的建立方法及报文转发方法。

背景技术

[0002] 在数据报文传输过程中,为保证数据报文的正常的转发,即当某一台网络设备故障时,数据报文可以快速切换到另一台网络设备中进行转发,通常采用分布式弹性网络互连(Distributed Resilient Network Interconnect,以下简称为DRNI)构建网络系统。其中DRNI是一种跨设备链路聚合技术,是将两台物理设备在聚合层面虚拟层一台设备来实现跨设备链路聚合,从而提供设备级冗余保护和流量负载分担。

[0003] 图1示出了DRNI转发场景,对于设备C和设备D而言,中间的设备A和设备B为虚拟网关,存在主备关系。其中,设备A和设备B构成分布式聚合系统(Distributed Relay System,简称为DR系统),设备A以及设备B为分布式聚合系统中的分布式聚合设备(又可称之为DR设备)。为了让图1中的L1-L4任意一个作为主的链路故障时,数据报文可以走IPL链路继续转发,最终通常会设置:设备A上,L1和L2均为主,L3和L4均为备份,而二层聚合口若需要三层转发,需要绑定到VLAN中。其中,设备A、设备B、L1、L3以及设备A和设备B之间的内部控制链路(Intra-Portal Link,简称为IPL)在一个VLAN,例如VLAN100;设备A、设备B、L2、L4以及IPL在另一个VLAN,例如VLAN200;其中,IPL同时允许VLAN100和VLAN200通过,最终IPL均能转发两侧传来的报文。

[0004] 然而,如果基于图1所示的网络使用多协议标签交换(Multiprotocol Label Switching,简称为MPLS)进行报文转发,即在MPLS进行报文转发时使用了标签分发协议(Label Distribution Protocol,简称为LDP),需要在各台设备的VLAN 100、VLAN 200中使能LDP配置,最终对于每个设备来说,一个接口可能会对应两个LDP邻居。请参见图1,例如设备C对应的LDP邻居有设备B和设备A,出接口都是VLAN100。若使用MPLS转发,组网的标签分配情况可能如下:设备A为设备D分配的入标签为1000,设备B为设备D分配的入标签为2000,那么数据报文在从主机A转发至主机B的过程中,设备C接受上述两个标签,如果认为两者均激活,但由于出接口是一个,所以出去带什么标签可能都不合适。

发明内容

[0005] 有鉴于此,本公开实施例提供了一种基于分布是聚合系统的标签转发表的建立及报文转发方法,以解决标签转发与DRNI网络不兼容的问题。

[0006] 根据第一方面,本公开实施例提供了一种基于分布式聚合系统的标签转发表的建立方法,包括:

[0007] 获取本地的第一标签转发表;其中,所述第一标签转发表用于表示所述下一跳节点与出标签的映射关系;

[0008] 获取对端为所述下一跳节点分配的入标签;其中,所述对端与本地属于同一所述

分布式聚合系统；

[0009] 从所述第一标签转发表中提取本地为所述下一跳节点分配的出标签；

[0010] 基于所述入标签以及所述出标签，建立所述第二标签转发表。

[0011] 本公开实施例提供的基于分布式聚合系统的标签转发表的建立方法，通过获取对端为下一跳节点分配的入标签，即本地存储有对端为下一跳节点分配的入标签，当本地所获取到的报文中携带有该入标签时，本地通过在标签转发表中能够查找到该入标签，实现报文的正常转发，从而解决了标签转发与DRNI网络不兼容的问题。

[0012] 结合第一方面，在第一方面第一实施方式中，所述基于所述入标签以及所述出标签，建立所述第二标签转发表，还包括：

[0013] 从所述第一标签转发表中提取本地对应于所述下一跳节点的出方向接口；

[0014] 建立所述入标签、所述出标签、所述出方向接口与所述下一跳节点的映射关系，以得到所述第二标签转发表。

[0015] 本公开实施例提供的基于分布式聚合系统的标签转发表的建立方法，由于在网络系统中，本地的下一跳节点可能有多个，而在同一时刻报文只能从唯一的出方向接口转发至下一跳节点，因此，本地还可以从第一标签转发表中提取本地对应于下一跳节点的出方向接口，以唯一确定出下一跳节点。

[0016] 结合第一方面，在第一方面第二实施方式中，所述获取对端为所述下一跳节点分配的入标签，包括：

[0017] 向所述对端发送标签转发表请求消息；其中，所述标签转发表请求消息用于请求所述对端对应于所述下一跳节点的第三标签转发表；

[0018] 接收所述对端发送的所述第三标签转发表；

[0019] 提取所述第三标签转发表中的入标签，以得到所述对端为所述下一跳节点分配的入标签。

[0020] 结合第一方面、第一方面第一实施方式至第一方面第二实施方式中任一项，在第一方面第三实施方式中，所述获取本地对应于下一跳节点的第一标签转发表的步骤之前，还包括：

[0021] 与所述对端交互Keplive报文，以确定本地与所述对端之间的所述分布式聚合系统的邻居关系。

[0022] 本公开实施例提供的基于分布式聚合系统的标签转发表的建立方法，通过交互Keplive报文进行分布式聚合系统的邻居关系的确定，且利用Keplive报文对同一分布式聚合系统中的设备进行故障检测，能够保证报文的正常转发。

[0023] 根据第二方面，本公开实施例还提供了一种基于分布式聚合系统的报文转发方法，包括：

[0024] 获取待转发报文；其中，所述待转发报文中携带有待匹配标签，所述待匹配标签用于确定与其对应的出标签，所述出标签用于表示本地的下一跳节点；

[0025] 在标签转发表中查找所述待匹配标签，以确定所述下一跳节点；其中，所述标签转发表是根据本公开第一方面，或第一方面任一项实施方式中所述的标签转发表的建立方法所建立的；

[0026] 向所述下一跳节点转发所述待转发报文。

[0027] 本公开实施例提供的基于分布式聚合系统的报文转发方法,由于本地存储有对端为下一跳节点分配的入标签,当本地所获取到的报文中携带有该入标签时,本地通过在标签转发表中能够查找到该入标签,实现报文的正常转发,从而解决了标签转发与DRNI网络不兼容的问题。

[0028] 结合第二方面,在第二方面第一实施方式中,所述在标签转发表中查找所述待匹配标签,以确定所述待转发报文的下一跳节点,包括:

[0029] 判断在所述第一标签转发表以及所述第二标签转发表中是否均存在与所述待匹配标签相同的标签;

[0030] 当在所述第一标签转发表以及所述第二标签转发表中均存在与所述待匹配标签相同的标签时,判断所述待转发报文是否为预设物理口所接收到的;

[0031] 当所述待转发报文为预设物理口所接收到的时,在所述第二标签转发表中查找所述待匹配标签,以确定所述待转发报文的下一跳节点。

[0032] 本公开实施例提供的基于分布式聚合系统的报文转发方法,通过设置预设物理口,当在第一标签转发表与第二标签转发表中均存在与待匹配标签相同的标签,且待转发报文中所携带的待匹配标签在两个标签转发表中均存在时,优先在第二标签转发表中进行查找,若此时仍在第一标签转发中查找待匹配标签,则会出现转发会不符合用户预期的情形。这是由于,比如图中设备A还接了其他组网凑巧用了这个标签(待转发标签),并且这个标签是分给再其他设备的,那么再从其他设备收到的这个标签,会转给其他组网。如果按这个转,流量就转到其他组网了,所以需要在第二标签转发表中进行查找,而且可以支持关联到从某个分配的邻居的入接口,从而提高了报文转发的准确性。

[0033] 结合第二方面第一实施方式,在第二方面第二实施方式中,所述在标签转发表中查找所述待匹配标签,以确定所述待转发报文的下一跳节点,还包括:

[0034] 当所述待转发报文不是所述预设物理口所接收到的时,在所述第一标签转发表中查找所述待匹配标签,以确定所述待转发报文的下一跳节点。

[0035] 根据第三方面,本公开实施例还提供了一种基于分布式聚合系统的标签转发表的建立装置,包括:

[0036] 第一获取模块,用于获取本地的第一标签转发表;其中,所述第一标签转发表用于表示所述下一跳节点与标签的映射关系;

[0037] 第二获取模块,用于获取对端为所述下一跳节点分配的入标签;其中,所述对端与本地属于同一所述分布式聚合系统;

[0038] 标签转发表建立模块,用于根据所述入标签以及所述第一标签转发表,建立第二标签转发表;

[0039] 出标签提取模块,用于从所述第一标签转发表中提取本地为所述下一跳节点分配的出标签;

[0040] 建立模块,基于所述入标签以及所述出标签,建立所述第二标签转发表。

[0041] 本公开实施例提供的基于分布式聚合系统的标签转发表的建立装置,通过获取对端为下一跳节点分配的入标签,即本地存储有对端为下一跳节点分配的入标签,当本地所获取到的报文中携带有该入标签时,本地通过在标签转发表中能够查找到该入标签,实现报文的正常转发,从而解决了标签转发与DRNI网络不兼容的问题。

[0042] 根据第四方面,本公开实施例还提供了一种基于分布式聚合系统的报文转发装置,包括:

[0043] 第三获取模块,用于获取待转发报文;其中,所述待转发报文中携带有待匹配标签,所述待匹配标签用于确定与其对应的出标签,所述出标签用于表示本地的下一跳节点;

[0044] 查找模块,用于在标签转发表中查找所述待匹配标签,以确定所述下一跳节点;其中,所述标签转发表是根据本公开第一方面,或第一方面任一项实施方式中所述的标签转发表的建立方法所建立的;

[0045] 报文转发模块,用于向所述下一跳节点转发所述待转发报文。

[0046] 本公开实施例提供的基于分布式聚合系统的报文转发装置,由于本地存储有对端为下一跳节点分配的入标签,当本地所获取到的报文中携带有该入标签时,本地通过在标签转发表中能够查找到该入标签,实现报文的正常转发,从而解决了标签转发与DRNI网络不兼容的问题。

[0047] 根据第五方面,本公开实施例还提供了一种网络设备,包括:

[0048] 存储器和处理器,所述存储器和所述处理器之间互相通信连接,所述存储器中存储有计算机指令,所述处理器通过执行所述计算机指令,从而执行本公开第一方面,或第一方面任一项实施方式中所述的基于分布式聚合系统的标签转发表的建立方法,或本公开第二方面,或第二方面任一项实施方式中所述的基于分布式聚合系统的报文转发方法。

[0049] 根据第六方面,本公开实施例还提供了一种计算机可读存储介质,所述计算机可读存储介质存储有计算机指令,所述计算机指令用于使所述计算机执行本公开第一方面,或第一方面任一项实施方式中所述的基于分布式聚合系统的标签转发表的建立方法,或本公开第二方面,或第二方面任一项实施方式中所述的基于分布式聚合系统的报文转发方法。

附图说明

[0050] 为了更清楚地说明本公开具体实施方式或现有技术中的技术方案,下面将对具体实施方式或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图是本公开的一些实施方式,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0051] 图1是根据本公开实施例的DRNI转发场景;

[0052] 图2是根据本公开实施例的基于分布式聚合系统的标签转发表的建立方法流程图;

[0053] 图3是根据本公开实施例的基于分布式聚合系统的标签转发表的建立方法流程图;

[0054] 图4是根据本公开实施例的基于分布式聚合系统的标签转发表的建立方法流程图;

[0055] 图5是根据本公开实施例的基于分布式聚合系统的报文转发方法流程图;

[0056] 图6是根据本公开实施例的基于分布式聚合系统的报文转发方法流程图;

[0057] 图7是根据本公开实施例的DRNI转发场景;

[0058] 图8是根据本公开实施例的基于分布式聚合系统的标签转发表的建立装置的结构

框图；

[0059] 图9是根据本公开实施例的基于分布式聚合系统的标签转发表的建立装置的部分结构框图；

[0060] 图10是根据本公开实施例的基于分布式聚合系统的标签转发表的建立装置的部分结构框图；

[0061] 图11是根据本公开实施例的基于分布式聚合系统的报文转发装置的结构框图；

[0062] 图12是根据本公开实施例的基于分布式聚合系统的报文转发装置的部分结构框图；

[0063] 图13是本公开实施例提供的网络设备的硬件结构示意图。

具体实施方式

[0064] 为使本公开实施例的目的、技术方案和优点更加清楚，下面将结合本公开实施例中的附图，对本公开实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例是本公开一部分实施例，而不是全部的实施例。基于本公开中的实施例，本领域技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本公开保护的范围。

[0065] 如图1所示，是本公开的实施例的一种应用场景示意图。其中，设备A与设备B属于同一个DRNI系统，且设备A与设备B之间建立LDP邻居。设备A与设备B分别为设备D分配标签，例如，设备A为设备D分配的入标签为1000，设备B为设备D分配的入标签为2000。若需要数据报文从主机A转发至主机B，设备C在接收到报文之后，为报文添加标签。但是由于与设备C对应于两个LDP邻居，分别为设备A以及设备B，而设备C只能为报文添加一个入标签，若设备C为报文添加的入标签为1000，而此时由于设备A故障，报文会经过设备A与设备B之间的IPL转发至设备B，但是设备B在解析报文时，得到的入标签为1000，通过查找自身的标签转发表，并未发现标签转发表中存在该标签，就会将该报文丢弃，从而导致数据报文的丢失。因此，本申请发明人发现导致DRNI网络与标签转发不兼容的原因在于DRNI系统中的DR设备仅存储有本地为下一跳节点分配的入标签，而并不知晓对端为下一跳节点分配的入标签。

[0066] 其中，需要说明的是，本公开实施例提供的基于分布式聚合系统的标签转发表的建立方法，对具体组网形式并不做任何限定，只需保证组网中具有DRNI网络。具体标签转发表的建立是针对DRNI网络中的设备进行描述的，对于复杂组网若带多层标签，则主要是针对这种情况的最外层标签处理。

[0067] 基于此，根据本公开实施例，提供了一种基于分布式聚合系统的标签转发表的建立方法实施例，需要说明的是，在附图的流程图示出的步骤可以在诸如一组计算机可执行指令的计算机系统中执行，并且，虽然在流程图中示出了逻辑顺序，但是在某些情况下，可以以不同于此处的顺序执行所示出或描述的步骤。

[0068] 在本实施例中提供了一种基于分布式聚合系统的标签转发表的建立方法，可用于网络设备，例如路由器等等，图2是根据本公开实施例的基于分布式聚合系统的标签转发表的建立方法的流程图，如图2所示，该流程包括如下步骤：

[0069] S11，获取本地的第一标签转发表。

[0070] 其中，所述第一标签转发表用于表示所述下一跳节点与出标签的映射关系。

[0071] 所述的标签转发表与IP网络中的转发信息库 (Forwarding Information Base, 简

称为FIB)类似,其中,在多协议标签交换系统(Multiprotocol Label Switching,简称为MPLS)中,标签交换路由器(Label Switching Router,简称为LSR)接收到带标签的报文后,通过查找标签转发表获取对应的标签操作类型、出标签值、下一跳等等,以确定如何转发报文。即,标签转发表用于确定报文转发的下一跳节点。

[0072] 具体地,在MPLS网络中,标签交换路径是通过标签转发表体现的,其中,标签交换路径(Label Switching Path,简称为LSP)的建立过程实际就是将转发等价类(Forwarding Equivalence Class,简称为FEC)和标签进行绑定,在LSR上建立标签转发表的过程。以动态建立LSP为例,下游LSR根据目的地址划分FEC,为特定FEC分配标签,并将FEC与标签绑定关系告知上游LSR,上游LSR根据该绑定关系建立标签转发表。

[0073] 本地的第一标签转发表用于表示本地的下一跳节点与出标签的映射关系,本地在获取到报文之后,利用该第一标签转发表即可实现报文的转发。

[0074] S12,获取对端为下一跳节点分配的入标签。

[0075] 其中,所述对端与本地属于同一分布式聚合系统。

[0076] 具体地,由于本地与对端的标签分配是独立进行的,本地并不知晓对端为下一跳节点分配的入标签。因此,本地通过获取对端为下一跳节点分配的入标签,在获取到带该入标签的报文时,可以实现报文的转发。其中,所述的下一跳节点为本地与对端共同的下一跳节点,或者可以理解为,当本地与对端属于同一DRNI时,该DRNI系统的下一跳节点。

[0077] 例如,请参见图1,设备A与设备B属于同一DRNI,本实施例中所述的本地以设备B为例,设备A以及设备B分别独立地为设备D分配标签值,本地(设备B)获取到对端(设备A)为设备D分配的入标签。若设备A为设备D分配的入标签为1000,设备B为设备D分配的入标签为2000,那么此时设备B获取到的入标签为1000。

[0078] S13,从第一标签转发表中提取本地为下一跳节点分配的出标签。。

[0079] 网络设备在获取到对端为下一跳节点分配的入标签之后,可以利用入标签以及第一标签转发表,建立对应于所述入标签的转发表项。在本地第一标签转发表中,对应于下一跳节点分配有入标签以及出标签。其中,出标签用于确定下一跳节点。具体地,本地从第一标签转发表中提取出标签,即对于本地而言,不论入标签是对端为下一跳节点分配的入标签,还是本地为下一跳节点分配的入标签,其对应的出标签均是本地为下一跳节点分配的出标签。

[0080] S14,基于入标签以及出标签,建立第二标签转发表。

[0081] 本地可以利用S12所获取到的入标签以及S13提取出的标签建立对应关系,即,本地获取到的入标签与提取出的出标签,建立第二标签转发表。在下文中,将对该步骤进行详细描述。

[0082] 本实施例提供的基于分布式聚合系统的标签转发表的建立方法,通过获取对端为下一跳节点分配的入标签,即本地存储有对端为下一跳节点分配的入标签,当本地所获取到的报文中携带有该入标签时,本地通过在标签转发表中能够查找到该入标签,实现报文的正常转发,从而解决了标签转发与DRNI网络不兼容的问题。

[0083] 在本实施例中还提供了一种基于分布式聚合系统的标签转发表的建立方法,可用于网络设备,例如路由器等等,图3是根据本公开实施例的基于分布式聚合系统的标签转发表的建立方法的流程图,如图3所示,该流程包括如下步骤:

- [0084] S21,获取本地的第一标签转发表。
- [0085] 其中,所述第一标签转发表用于表示所述下一跳节点与出标签的映射关系。
- [0086] 详细请参见图2所示实施例的S11,在此不再赘述。
- [0087] S22,获取对端为下一跳节点分配的入标签。
- [0088] 其中,所述对端与本地属于同一分布式聚合系统。
- [0089] 详细请参见图2所示实施例的S12,在此不再赘述。
- [0090] S23,从第一标签转发表中提取本地为下一跳节点分配的出标签。
- [0091] 详细请参见图2所示实施例的S13,在此不再赘述。
- [0092] S24,基于入标签以及出标签,建立第二标签转发表。
- [0093] 本地可以利用S22所获取到的入标签以及S23提取出的标签建立对应关系,即,本地获取到的入标签与提取出的出标签,建立第二标签转发表。
- [0094] 具体地,S24还可以包括以下步骤:
- [0095] S241,从第一标签转发表中提取本地对应于下一跳节点的出方向接口。
- [0096] 由于在网络系统中,本地的下一跳节点可能有多个,而在同一时刻报文只能从唯一的出方向接口转发至下一跳节点,因此,本地还可以从第一标签转发表中提取本地对应于下一跳节点的出方向接口,以唯一确定出下一跳节点。
- [0097] S242,建立入标签、出标签、出方向接口与下一跳节点的映射关系,以得到第二标签转发表。
- [0098] 本地即可利用S22获取到的入标签、S23提取出的出标签、S241中提取出的出方向接口,建立与下一跳节点的映射关系,从而得到第二标签转发表。
- [0099] 在上述步骤之后,本地中存储的标签转发表包括第一标签转发表以及第二标签转发表,其中第一标签转发表为本地对应于下一跳节点的标签转发表,第二标签转发表为基于对端为下一跳节点分配的入标签建立的标签转发表。
- [0100] 在本实施例中还提供了一种基于分布式聚合系统的标签转发表的建立方法,可用于网络设备,例如路由器等等,图4是根据本公开实施例的基于分布式聚合系统的标签转发表的建立方法的流程图,如图4所示,该流程包括如下步骤:
- [0101] S31,与对端交互Keypalive报文,以确定本地与对端之间的分布式聚合系统的邻居关系。
- [0102] 在同一DRNI网络中的两个设备之间通过Keypalive链路检测邻居状态,即,通过DRNI之间的Keypalive报文,确定两个设备之间存在了DRNI的邻居关系。例如,请参见图1,设备A与设备B之间通过交互Keypalive报文,确定相互之间的邻居关系。
- [0103] S32,向对端发送标签转发表请求消息。
- [0104] 其中,所述标签转发表请求消息用于请求对端对应于下一跳节点的第三标签转发表。
- [0105] 本地与对端在确定了相互之间的邻居关系之后,本地与对端之间也建立标签分发协议(Label Distribution Protocol,简称为LDP)的邻居关系。本地与对端建立邻居关系之后,本地向对端发送标签转发表请求消息,用于请求获取对端对应于下一跳节点的第三标签转发表。
- [0106] S33,接收对端发送的第三标签转发表。

[0107] 对端在接收到S32中的标签转发表请求消息之后,将第三标签转发表发送给本地;即,本地接收对端发送的第三标签转发表。

[0108] S34,提取第三标签转发表中的入标签,以得到对端为下一跳节点分配的入标签。

[0109] 本地接收到第三标签转发表之后,从第三标签转发表中提取相应的入标签,从而即可得到对端为下一跳节点分配的入标签。

[0110] S35,获取本地的第一标签转发表。

[0111] 其中,所述第一标签转发表用于表示所述下一跳节点与出标签的映射关系。

[0112] 详细请参见图3所示实施例的S21,在此不再赘述。

[0113] S36,获取对端为下一跳节点分配的入标签。

[0114] 其中,所述对端与本地属于同一分布式聚合系统。

[0115] 详细请参见图3所示实施例的S22,在此不再赘述。

[0116] S37,从第一标签转发表中提取本地为下一跳节点分配的出标签。

[0117] 详细请参见图3所示实施例的S23,在此不再赘述。

[0118] S38,基于入标签以及出标签,建立第二标签转发表。

[0119] 详细请参见图3所示实施例的S24,在此不再赘述。

[0120] 本实施例提供的基于分布式聚合系统的标签转发表的建立方法,通过交互Keepalive报文进行分布式聚合系统的邻居关系的确定,且利用Keepalive报文对同一分布式聚合系统中的设备进行故障检测,能够保证报文的正常转发。

[0121] 作为本实施例的一个具体应用实例,请参见图1,设备A与设备B之间通过Keepalive报文建立了分布式聚合系统的邻居关系;设备A与设备B之间也建立了LDP邻居。若报文从主机A到主机B转发,路由学习是通过设备D来发布主机B的主机路由,假设为4.4.4.4/32。由于设备D也会和设备A、设备B建立LDP邻居,设备A以及设备B会分配为4.4.4.4/32作为FEC分配一个入标签,设备A为设备D分配的入标签为1000,设备B为设备D分配的入标签为2000。在设备B上的第一标签转发表FEC为4.4.4.4/32,对应的入标签为2000,出标签为4000;在设备A上的第一标签转发表FEC为4.4.4.4/32,对应的入标签为1000,出标签为3000。设备A与设备B之间建立了LDP邻居,设备A以及设备B会将FEC、入标签信息扩散给对端邻居。以设备A为参照,设备A收到设备B发的LDP消息,会将FEC:4.4.4.4/32,label:2000记录到第二标签转发表。即,设备A上第一标签转发表,FEC:4.4.4.4/32的入标签1000,出标签3000;第二标签转发表,FEC:4.4.4.4/32的入标签2000,出标签3000;且两张标签转发表的出方向接口均为第一标签转发表中对应于FEC:4.4.4.4/32的出方向接口。

[0122] 根据本公开实施例,提供了一种基于分布式聚合系统的报文转发方法实施例,需要说明的是,在附图的流程图示出的步骤可以在诸如一组计算机可执行指令的计算机系统中执行,并且,虽然在流程图中示出了逻辑顺序,但是在某些情况下,可以以不同于此处的顺序执行所示出或描述的步骤。

[0123] 在本实施例中提供了一种基于分布式聚合系统的报文转发方法,可用于网络设备,例如路由器等等,图5是根据本公开实施例的基于分布式聚合系统的报文转发方法的流程图,如图5所示,该流程包括如下步骤:

[0124] S41,获取待转发报文。

[0125] 其中,所述待转发报文中携带有待匹配标签,所述待匹配标签用于确定与其对应

的出标签,所述出标签用于表示本地的下一跳节点。

[0126] 例如,请参见图1,设备A与设备B形成负载分担,若设备A为主设备,设备B为从设备。当设备A正常工作时,设备B处于待机状态,设备A获取待转发报文,进行报文转发;当设备A故障时,待转发报文通过设备A与设备B之间的内部控制链路(Intra-Portal Link,简称为IPL)传输至设备B,设备B对该待转发报文进行报文转发。

[0127] S42,在标签转发表中查找待匹配标签,以确定下一跳节点。

[0128] 其中,所述标签转发表是根据上述任一项实施例所述的基于分布式聚合系统的标签转发表的建立方法所建立的。

[0129] 网络设备在获取到待转发报文时,对待转发报文进行解析,以得到待匹配标签。在标签转发表中进行待匹配标签的查找,以确定出下一跳节点。

[0130] 其中,属于同一DRNI网络的两个设备中的标签转发表包括第一标签转发表以及第二标签转发表,其中第一标签转发表为本地对应于下一跳节点的标签转发表,第二标签转发表为基于对端为下一跳节点分配的入标签建立的标签转发表。

[0131] S43,向下一跳节点转发待转发报文。

[0132] 本地在确定出下一跳节点之后,将待转发报文转发至下一跳节点中,以实现报文的转发。

[0133] 本实施例提供的基于分布式聚合系统的报文转发方法,由于本地存储有对端为下一跳节点分配的入标签,当本地所获取到的报文中携带有该入标签时,本地通过在标签转发表中能够查找到该入标签,实现报文的正常转发,从而解决了标签转发与DRNI网络不兼容的问题。

[0134] 在本实施例中提供了一种基于分布式聚合系统的报文转发方法,可用于网络设备,例如路由器等等,图6是根据本公开实施例的基于分布式聚合系统的报文转发方法的流程图,如图6所示,该流程包括如下步骤:

[0135] S51,获取待转发报文。

[0136] 其中,所述待转发报文中携带有待匹配标签,所述待匹配标签用于确定与其对应的出标签,所述出标签用于表示本地的下一跳节点。

[0137] 详细请参见图5所示实施例的S41,在此不再赘述。

[0138] S52,在标签转发表中查找待匹配标签,以确定下一跳节点。

[0139] 其中,所述标签转发表是根据上述任一项实施例所述的基于分布式聚合系统的标签转发表的建立方法所建立的。

[0140] 因为第二转发表的入标签是根据对端的情况填充的,对端和本端的标签申请是独立的,可能出现重复,那么就考虑第一标签转发表与第二标签转发表存在相同的标签的情况。具体地,包括以下步骤:

[0141] S521,判断在第一标签转发表以及第二标签转发表中是否均存在与待匹配标签相同的标签。

[0142] 本地解析待转发报文,并得到待匹配标签之后,首先确定该待匹配标签是否同时属于第一标签转发表以及第二标签转发表。

[0143] 当在第一标签转发表以及第二标签转发表中均存在与待匹配标签相同的标签时,执行S522;否则,确定出下一跳节点。

[0144] S522,判断待转发报文是否为预设物理口所接收到的。

[0145] 由于属于同一个DRNI网络的两个设备对外体现为一个虚拟口,而这个虚拟口实际可以关联多个物理口,但是物理口是唯一的。例如,请参见图7,设备A与设备B属于同一个DRNI网络,该网络实际关联的物理口有GE1/0/1,以及GE1/0/2。对于整个DRNI网络而言,待转发报文进来的逻辑口会对应两个物理口中的一个,而具体设置这两个物理口中的哪一个作为预设物理口是基于这个组网的特征决定的。若设备B接收到待转发报文,将预设物理口指定为GE1/0/1则是明确该待转发报文从L3收到的。即,指定预设物理口的目的在于,确定待转发报文是通过哪个链路接收到的。

[0146] 当待转发报文为预设物理口所接收到的时,执行S523;否则,执行S524。

[0147] S523,在第二标签转发表中查找待匹配标签,以确定待转发报文的下一跳节点。

[0148] 本地确定出在第一标签转发表以及第二标签转发表中同时存在与待匹配标签相同的标签,且待转发报文为预设物理口所接收到的时,本地在第二标签转发表中查找待匹配标签。

[0149] 若此时仍在第一标签转发中查找待匹配标签,则会出现转发会不符合用户预期的情形。这是由于,比如图中设备A还接了其他组网凑巧用了这个标签(待转发标签),并且这个标签是分给再其他设备的,那么再从其他设备收到的这个标签,会转给其他组网。如果按这个转,流量就转到其他组网了,所以需要在第二标签转发表中进行查找,而且可以支持关联到从某个分配的邻居的入接口。

[0150] S524,在第一标签转发表中查找待匹配标签,以确定待转发报文的下一跳节点。

[0151] 本地确定出在第一标签转发表以及第二标签转发表中同时存在与待匹配标签相同的标签,但是待转发报文不是预设物理口发送时,本地在第一标签转发表中查找待匹配标签。

[0152] S53,向下一跳节点转发待转发报文。

[0153] 详细请参见图5所示实施例所述的S43,在此不再赘述。

[0154] 作为本实施例的一个具体应用实例,当第一标签转发表和第二标签转发表出现了相同的入标签,则在收到标签报文时,对入方向进行核对,通过配置指定的预设物理口方式,指定某个物理接口在接受到相同标签值的报文时会直接在第二标签转发表查找报文。需要说明的是,虽然标签转发表中的标签转发表项是通过FEC进行区分的,但是在查找时不能根据FEC来核对,因为对L3VPN/L2VPN组网,报文不一定会带FEC的地址,但是标签会带。后续若在第二标签表查找到对应的标签转发表,则根据前面已经填充的出标签,和出方向信息,将报文转给下一台设备。

[0155] 请参见图7,设备A上第一标签转发表,FEC:4.4.4.4/32的入标签1000出标签3000,第二标签转发表FEC:4.4.4.4/32的入标签2000出标签3000,两张标签转发表出方向接口均为GE1/0/2。设备A接收到的待转发报文中的待匹配标签同时属于第一标签转发表以及第二标签转发表,因此,为了避免标签2000和第一标签转发表的标签值冲突,设备A指定GE1/0/1口为重复标签优先进第二标签转发表进行转发,这样后续流量转发,若设备C从GE1/0/1发出的待转发报文中携带的标签为2000,则A判断走第二标签转发表。

[0156] 在本实施例中还提供了一种基于分布式聚合系统的标签转发表的建立装置,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。

[0157] 本实施例提供一种基于分布式聚合系统的标签转发表的建立装置,如图8所示,包括:

[0158] 第一获取模块61,用于获取本地的第一标签转发表;其中,所述第一标签转发表用于表示所述下一跳节点与标签的映射关系。

[0159] 第二获取模块62,用于获取对端为所述下一跳节点分配的入标签;其中,所述对端与本地属于同一所述分布式聚合系统。

[0160] 出标签提取模块63,用于从所述第一标签转发表中提取本地为所述下一跳节点分配的出标签。

[0161] 建立模块64,基于所述入标签以及所述出标签,建立所述第二标签转发表。

[0162] 本实施例提供的基于分布式聚合系统的标签转发表的建立装置,通过获取对端为下一跳节点分配的入标签,即本地存储有对端为下一跳节点分配的入标签,当本地所获取到的报文中携带有该入标签时,本地通过在标签转发表中能够查找到该入标签,实现报文的正常转发,从而解决了标签转发与DRNI网络不兼容的问题。

[0163] 作为本实施例的一种可选实施方式,如图9所示,所述建立模块64,包括:

[0164] 提取单元641,用于从所述第一标签转发表中提取本地对应于所述下一跳节点的出方向接口。

[0165] 第二标签转发表建立单元642,用建立所述入标签、所述出标签、所述出方向接口与所述下一跳节点的映射关系,以得到所述第二标签转发表。

[0166] 在本实施例的另一些可选实施方式中,如图10所示,所述基于分布式聚合系统的标签转发表的建立装置,还包括:

[0167] 标签请求模块65,用于向所述对端发送标签转发表请求消息;其中,所述标签转发表请求消息用于请求所述对端对应于所述下一跳节点的第三标签转发表。

[0168] 接收模块66,用于接收所述对端发送的所述第三标签转发表。

[0169] 入标签提取模块67,用于提取所述第三标签转发表中的入标签,以得到所述对端为所述下一跳节点分配的入标签。

[0170] 作为本实施例的一种可选实施方式,如图10所示,所述基于分布式聚合系统的标签转发表的建立装置,还包括:

[0171] 报文交互模块68,用于与所述对端交互Keepalive报文,以确定本地与所述对端之间的所述分布式聚合系统的邻居关系。

[0172] 在本实施例中还提供了一种基于分布式聚合系统的报文转发装置,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。

[0173] 本实施例提供一种基于分布式聚合系统的报文转发装置,如图11所示,包括:

[0174] 第三获取模块71,用于获取待转发报文;其中,所述待转发报文中携带有待匹配标签,所述待匹配标签用于确定与其对应的出标签,所述出标签用于表示本地的下一跳节点。

[0175] 查找模块72,用于在标签转发表中查找所述待匹配标签,以确定所述下一跳节点;

其中,所述标签转发表是根据上述任一项所述的标签转发表的建立方法所建立的。

[0176] 报文转发模块73,用于向所述下一跳节点转发所述待转发报文。

[0177] 本实施例提供的基于分布式聚合系统的报文转发装置,由于本地存储有对端为下一跳节点分配的入标签,当本地所获取到的报文中携带有该入标签时,本地通过在标签转发表中能够查找到该入标签,实现报文的正常转发,从而解决了标签转发与DRNI网络不兼容的问题。

[0178] 在本实施例的一些可选实施方式中,如图12所示,所述查找模块72,包括:

[0179] 第一判断单元721,用于判断在所述第一标签转发表以及所述第二标签转发表中是否均存在与所述待匹配标签相同的标签。

[0180] 第二判断单元722,用于当在所述第一标签转发表以及所述第二标签转发表中均存在与所述待匹配标签相同的标签时,判断所述待转发报文是否为预设物理口所接收到的。

[0181] 第一查找单元723,用于当所述待转发报文为预设物理口所接收到的时,在所述第二标签转发表中查找所述待匹配标签,以确定所述待转发报文的下一跳节点。

[0182] 作为本实施例的一种可选实施方式,如图12所示,所述查找模块72,还包括:

[0183] 第二查找单元724,用于当所述待转发报文不是所述预设物理口所接收到的时,在所述第一标签转发表中查找所述待匹配标签。

[0184] 上述实施例中的基于分布式聚合系统的标签转发表的建立装置以及基于分布式聚合系统的报文转发装置是以功能单元的形式来呈现,这里的单元是指ASIC电路,执行一个或多个软件或固定程序的处理器和存储器,和/或其他可以提供上述功能的器件。

[0185] 上述各个模块的更进一步的功能描述与上述对应实施例相同,在此不再赘述。

[0186] 本公开实施例还提供一种网络设备,具有上述图8-10中任一项所示的基于分布式聚合系统的标签转发表的建立装置,或具有上述图11-12任一项所示的基于分布式聚合系统的报文转发装置。

[0187] 请参阅图13,图13是本公开可选实施例提供的一种网络设备的结构示意图,如图13所示,该网络设备可以包括:至少一个处理器801,例如CPU(Central Processing Unit,中央处理器),至少一个通信接口803,存储器804,至少一个通信总线802。其中,通信总线802用于实现这些组件之间的连接通信。其中,通信接口803可以包括显示屏(Display)、键盘(Keyboard),可选通信接口803还可以包括标准的有线接口、无线接口。存储器804可以是高速RAM存储器(Random Access Memory,易挥发性随机存取存储器),也可以是非不稳定的存储器(non-volatile memory),例如至少一个磁盘存储器。存储器804可选的还可以是至少一个位于远离前述处理器801的存储装置。其中处理器801可以结合图8-10中任一项,或图11-12中任一项所描述的装置,存储器804中存储应用程序,且处理器801调用存储器804中存储的程序代码,以用于执行上述任一方法步骤。

[0188] 其中,通信总线802可以是外设部件互连标准(peripheral component interconnect,简称PCI)总线或扩展工业标准结构(extended industry standard architecture,简称EISA)总线等。通信总线802可以分为地址总线、数据总线、控制总线等。为便于表示,图13中仅用一条粗线表示,但并不表示仅有一根总线或一种类型的总线。

[0189] 其中,存储器804可以包括易失性存储器(英文:volatile memory),例如随机存取

存储器(英文:random-access memory,缩写:RAM);存储器也可以包括非易失性存储器(英文:non-volatile memory),例如快闪存储器(英文:flash memory),硬盘(英文:hard disk drive,缩写:HDD)或固态硬盘(英文:solid-state drive,缩写:SSD);存储器804还可以包括上述种类的存储器的组合。

[0190] 其中,处理器801可以是中央处理器(英文:centeral processing unit,缩写:CPU),网络处理器(英文:network processor,缩写:NP)或者CPU和NP的组合。

[0191] 其中,处理器801还可以进一步包括硬件芯片。上述硬件芯片可以是专用集成电路(英文:application-specific integrated circuit,缩写:ASIC),可编程逻辑器件(英文:programmable logic device,缩写:PLD)或其组合。上述PLD可以是复杂可编程逻辑器件(英文:complex programmable logic device,缩写:CPLD),现场可编程逻辑门阵列(英文:field-programmable gate array,缩写:FPGA),通用阵列逻辑(英文:generic array logic,缩写:GAL)或其任意组合。

[0192] 可选地,存储器804还用于存储程序指令。处理器801可以调用程序指令,实现如本申请图1-3实施例中所示的基于分布式聚合系统的标签转发表的建立方法,或本申请图4-5实施例中所示的基于分布式聚合系统的报文转发方法。

[0193] 本公开实施例还提供了一种非暂态计算机存储介质,所述计算机存储介质存储有计算机可执行指令,该计算机可执行指令可执行上述任意方法实施例中的基于分布式聚合系统的标签转发表的建立方法,或基于分布式聚合系统的报文转发方法。其中,所述存储介质可为磁碟、光盘、只读存储记忆体(Read-Only Memory,ROM)、随机存储记忆体(Random Access Memory,RAM)、快闪存储器(Flash Memory)、硬盘(Hard Disk Drive,缩写:HDD)或固态硬盘(Solid-State Drive,SSD)等;所述存储介质还可以包括上述种类的存储器的组合。

[0194] 虽然结合附图描述了本公开的实施例,但是本领域技术人员可以在不脱离本公开的精神和范围的情况下做出各种修改和变型,这样的修改和变型均落入由所附权利要求所限定的范围之内。

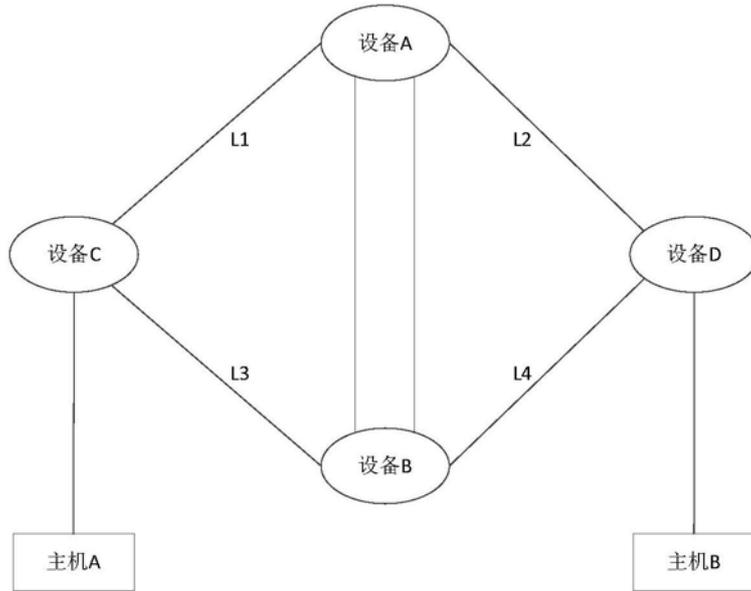


图1



图2

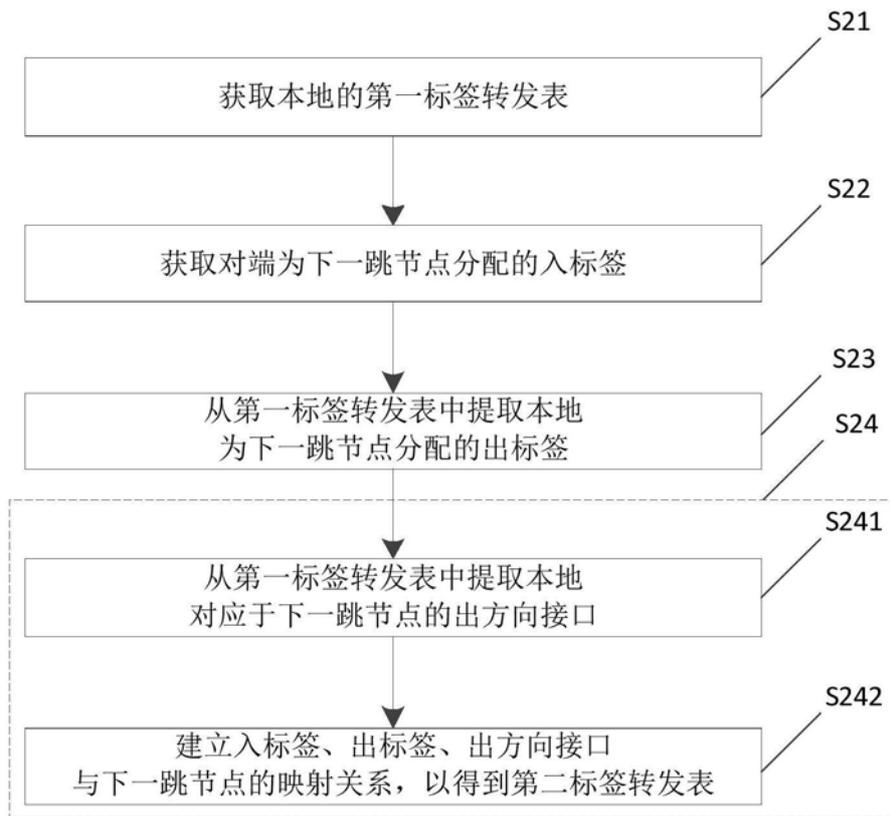


图3

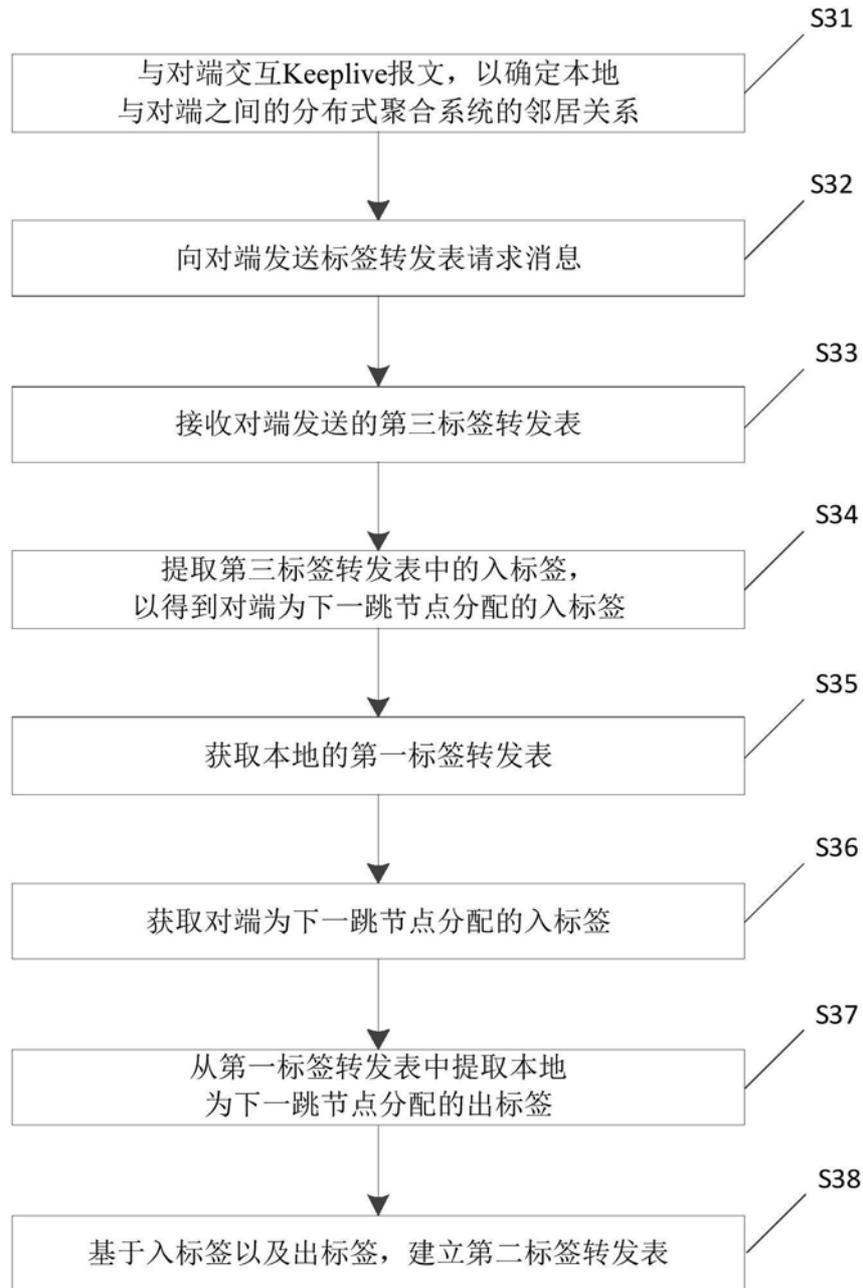


图4

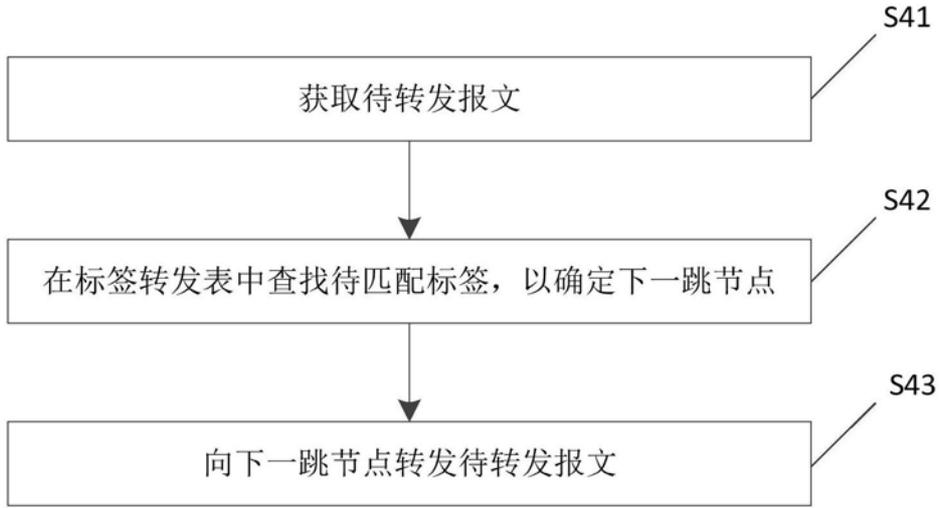


图5

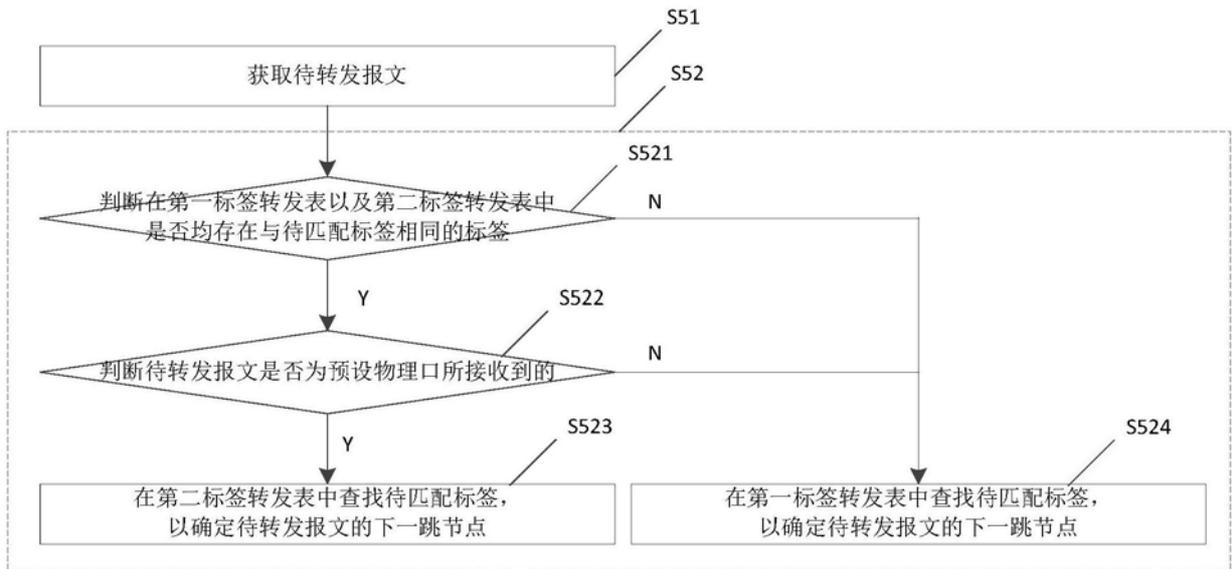


图6

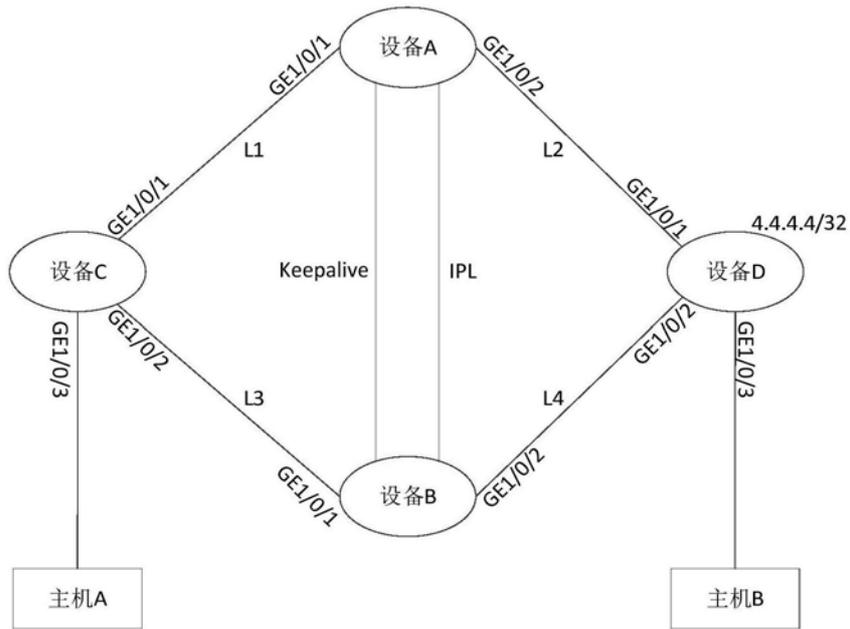


图7

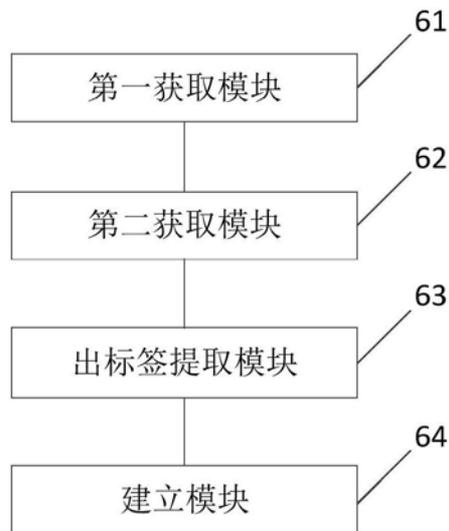


图8

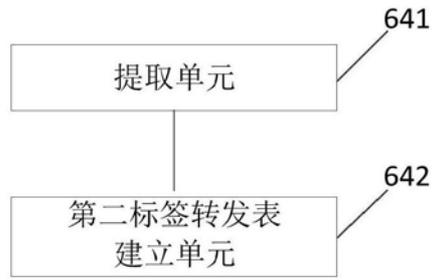


图9

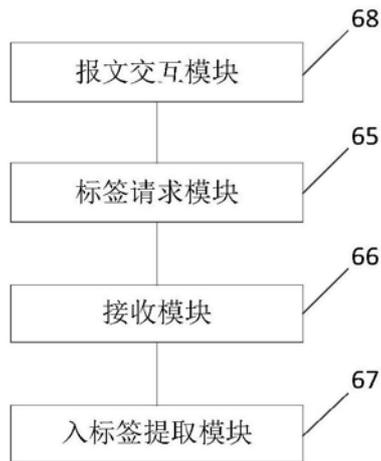


图10

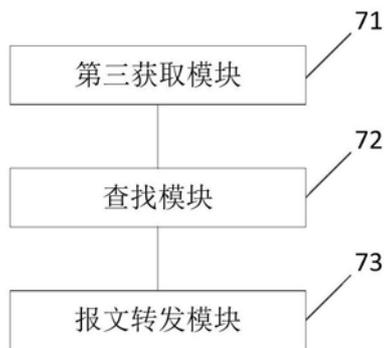


图11

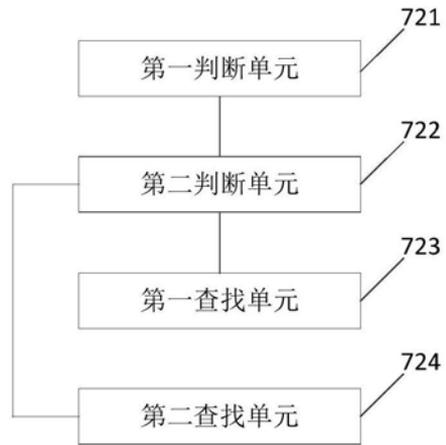


图12

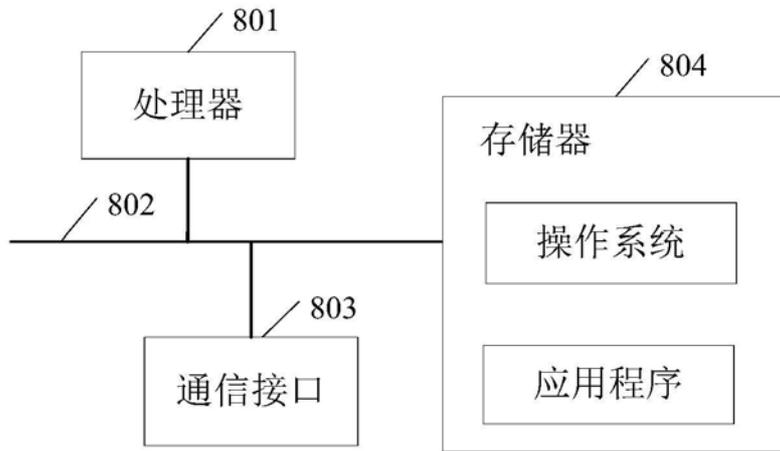


图13