(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国 际 局

(43) 国际公布日
2016 年 12 月 29 日 (29.12.2016)

WIPO | PCT

(10) 国际公布号
WO 2016/206620 A1

(51) 国际专利分类号:
H04L 12/721 (2013.01) *H04L 12/861* (2013.01)
H04L 12/741 (2013.01)

(21) 国际申请号: PCT/CN2016/086989

(22) 国际申请日: 2016 年 6 月 24 日 (24.06.2016)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201510363082.0 2015 年 6 月 26 日 (26.06.2015) CN

(71) 申请人: 杭州华三通信技术有限公司 (HANGZHOU H3C TECHNOLOGIES CO., LTD.) [CN/CN]; 中国浙江省杭州市滨江区长河路 466 号, Zhejiang 310052 (CN)。

(72) 发明人: 宋玉兵 (SONG, Yubing); 中国北京市海淀区上地信息产业基地创业路 2 号东方电子大厦 730 室, Beijing 100085 (CN)。

(74) 代理人: 北京博思佳知识产权代理有限公司 (BEIJING BESTIPR INTELLECTUAL PROPERTY LAW CORPORATION); 中国北京市海淀区上地三街 9 号嘉华大厦 B 座 409, Beijing 100085 (CN)。

(81) 指定国 (除另有指明，要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国 (除另有指明，要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:
— 包括国际检索报告(条约第 21 条(3))。

(54) Title: MESSAGE FORWARDING

(54) 发明名称 : 报文转发

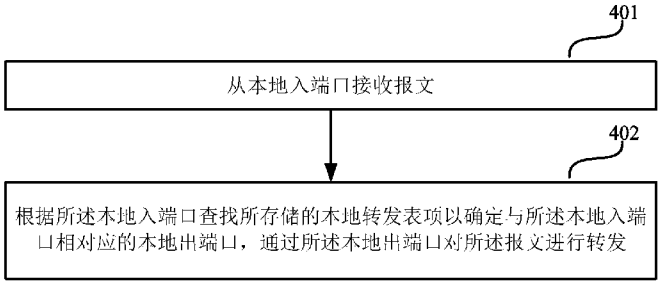


图 4

401 Receiving messages from a local inlet port
402 Searching a stored local forwarding table item according to the local inlet port in order to determine a local outlet port corresponding to the local inlet port, and forwarding messages via the local outlet port

(57) Abstract: In the example, a message forwarding method comprises: a port extender (PE) device of a PE stacking system receives messages from a local inlet port; the PE device searches a local forwarding table item according to the local inlet port in order to determine a local outlet port corresponding to the local inlet port; the PE device forwards messages via the local outlet port.

(57) 摘要: 在示例中，一种报文转发方法包括：端口扩展(PE)堆叠系统中的 PE 设备从本地入端口接收报文；所述 PE 设备根据所述本地入端口查找本地转发表项，以确定与所述本地入端口相对应的本地出端口；所述 PE 设备经由所述本地出端口对所述报文进行转发。

WO 2016/206620 A1

报文转发

背景技术

[01] 为了满足数据中心的需求，提出了堆叠组网技术。例如，CB（control bridge，控制桥）设备之间可通过堆叠链路连接形成CB堆叠系统。PE（port extender，端口扩展）设备可连接CB设备并作为CB设备的远程接口设备提供端口扩展功能。PE设备接收到的报文可上送至CB设备，由CB设备通过查找转发表来转发所述报文。为了节省PE设备向CB设备的上行链路，可以将多个独立的PE设备通过堆叠链路连接形成PE堆叠系统。报文被该PE堆叠系统的入口PE设备接收后，经过PE堆叠系统内的报文转发以被上送至CB设备，然后该报文可以被CB设备处理并下发至PE堆叠系统。再由PE堆叠系统将CB设备下发的报文转发至目的设备。

[02] PE堆叠系统可基于报文中插入的标签（tag）进行报文转发。比如，支持tag封装的同厂家同型号的PE设备可进行堆叠。

附图说明

[03] 图1示出了根据本公开的一示例的一种报文转发的示意图；

[04] 图2示出了根据本公开的一示例的报文转发方法应用的组网结构；

[05] 图3示出了根据本公开的一示例的PE设备的逻辑功能结构；

[06] 图4示出了根据本公开的一示例的一种报文转发方法的流程图；

[07] 图5示出了根据本公开的另一示例的一种报文转发方法的流程图；

[08] 图6示出了根据本公开的一示例的表项转换的流程图；

[09] 图7示出了根据本公开的一示例的报文转发路径的示意图；

[10] 图8示出了根据本公开的一示例的报文转发路径的示意图；

[11] 图 9 示出了根据本公开的一示例的一种表项转换的示意图;

[12] 图 10 示出了根据本公开的一示例的另一种表项转换的示意图;

[13] 图 11 示出了根据本公开的一示例的端口确定的流程图;

[14] 图 12 示出了根据本公开的一示例的 PE 堆叠系统间转发报文的示意图;

[15] 图 13 示出了根据本公开的一示例的 PE 堆叠系统间转发报文的流程图;

[16] 图 14 示出了根据本公开的一示例的 PE 堆叠系统内转发报文的示意图;

[17] 图 15 示出了根据本公开的一示例的 PE 堆叠系统内转发报文的流程图;

[18] 图 16 示出了根据本公开的一示例的环形 PE 堆叠系统的示意图;

[19] 图 17 示出了根据本公开的一示例的组播报文转发的示意图;

[20] 图 18 示出了根据本公开的一示例的组播报文的下行转发的示意图;

[21] 图 19 示出了根据本公开的一示例的组播报文的下行转发表项;

[22] 图 20 示出了根据本公开的一示例的组播报文的下行转发的流程图;

[23] 图 21 示出了根据本公开的一示例的 PE 设备的示意性硬件结构框图;

[24] 图 22 示出了根据本公开的一示例的一种报文转发控制逻辑的功能模块框图;

[25] 图 23 示出了根据本公开的另一示例的一种报文转发控制逻辑的功能模块框图。

具体实施方式

[26] 本公开示例的报文转发方法可以应用于 PE 堆叠系统。例如，通过该报文转发方法，PE 堆叠系统可将源设备发送的报文通过上行链路转发至 CB 设备，或者将 CB 设备发出的报文通过下行链路转发至目的设备。并且，使用该报文转发方法在 PE 堆叠系统的各个 PE 设备之间转发报文时，该报文不用进行 tag 封装（例如，常用的 Higig 封装）即可被转发。

[27] 图 1 示例了一种报文转发的示意图。如图 1 所示，PC 1、PC 2 和 PC 3 可以作为发送报文的源设备或者接收报文的目的设备，例如，PC 1 可向 PC 3 发送报文。在报文发送的过程中，可由图 1 中示例的 PE 堆叠系统 1000 和 CB 堆叠系统 2000 进行转发，也可以是单个 CB 设备连接 PE 堆叠系统或 CB 堆叠系统连接单个 PE 设备进行转发。如图 1 的带箭头虚线所示意（该带箭头虚线只是用于表示报文的大致转发路径，并不限定实际经过的转发设备），在 PC 1 向 PC 3 发送报文时，PC 1 可将报文发送至 PE 堆叠系统 1000 中的 PE 12。接着，可由 PE 堆叠系统 1000 中的 PE 11 将报文上送至 CB 堆叠系统 2000 中的 CB 10，以由 CB 堆叠系统 2000 中的 CB 10 通过查找转发表来确定报文的输出端口。在这种情况下，PE 设备可相当于 CB 设备的扩展端口，报文查表转发处理仍然在 CB 设备执行。然后，CB 堆叠系统 2000 中的 CB 20 将报文经由通过查找转发表确定的输出端口发送到 PE 堆叠系统 1000，PE 堆叠系统 1000 经由 PE 14、13 将报文发送至目的设备 PC 3。

[28] 在下文中，根据本公开示例的报文转发方法，将描述在上送 CB 设备或下发 PC 的报文转发中，报文如何在 PE 堆叠系统内部进行转发。

[29] 首先，结合图 2，简单说明在该报文转发方法的描述中涉及到的组网结构。如图 2 所示，示例了两个 PE 堆叠系统，包括：由 PE 11、PE 12、PE 13 和 PE 14 组成的第一 PE 堆叠系统 1000，以及由 PE 15 和 PE 16 组成

的第二 PE 堆叠系统 3000。

[30] 其中，PE 堆叠系统可由多个 PE 设备连接组成，以第一 PE 堆叠系统 1000 为例，各个 PE 设备之间的连接关系可以参见图 2 所示。每个 PE 设备都包括多个端口，每个端口可以被分配一个端口号。其中，用于 PE 设备之间连接的端口称为“堆叠口”（stacking port），属于 PE 设备但是不是用于 PE 设备之间连接的端口称为“非堆叠口”。“堆叠口”和“非堆叠口”都是该 PE 设备的本地端口（local port）。针对某个 PE 设备，其他 PE 设备相对其都可以称为“远端 PE 设备”，远端 PE 设备的端口称为“远端端口”。以图 2 中的 PE 12 为例，该 PE 12 包括端口 P1、S1 和 S2。其中，端口 P1 是非堆叠口，用于连接 PC 1；端口 S1 和 S2 都是堆叠口，端口 S1 用于连接 PE 11，端口 S2 用于连接 PE 13。PE 13 可称为 PE 12 的远端 PE 设备，PE 13 上的端口 S1 相当于 PE 12 的远端端口。

[31] 此外，PE 设备的非堆叠口可以用于连接 PC 或者 CB 设备，例如图 2 的示例，PE 12 的端口 P1 连接 PC 1，PE 13 的端口 P1 连接 PC 2，PE 15 的端口 P1 连接 PC 3。并且，在以 802.1BR 协议为基础的 CB-PE 纵向堆叠系统中，PE设备的端口可被分配扩展通道标识 PCID（Port Extended Channel ID），作为端口标识符。比如，PE 12 的用于连接 PC 1 的端口 P1 的 PCID 是“100”，PE 13 的用于连接 PC 2 的端口 P1 的 PCID 是“101”，而 PE 15 的用于连接 PC 3 的端口 P1 的 PCID 可以为“200”。在报文从 PE 设备的非堆叠口被接收后，该 PCID 可以被设置在报文中的电子标签（E-tag）的 E-CID 字段。这里的 E-tag 是根据 802.1BR 标准中定义的 CB 设备与 PE 设备之间交互报文时给报文插入的 tag，用以实现集中式的转发模型。例如，如果报文中携带 E-CID “100”，则可以根据该报文的来源端口判断该报文是从 PC 1 发出或者要发送至 PC 1。例如，如果报文是从 PE 设备的非堆叠口接收的报文，则通过该报文携带的 E-CID “100”可以判断该报文是从 PC 1 发出。而如果报文是从 PE 设备的堆叠口接收的报文，则确定该堆叠

口是否有 ACL 重定向规则（下文中将进行描述）。如果该堆叠口有 ACL 重定向规则，则通过该报文携带的 E-CID “100”可以判断该报文从 PC 1 上行发出到 CB 设备。而如果该堆叠口没有 ACL 重定向规则，则可以判断该报文是要发送给 PC 1。PE 设备还可以通过非堆叠口连接 CB 设备，例如，PE 11 通过其端口 P1 连接 CB 堆叠系统中的 CB 10 的端口 P1，PE 14 通过端口 P1 连接 CB 堆叠系统中的 CB 20 的端口 P1，并且通过端口 P2 连接 CB 堆叠系统中的 CB 20 的端口 P2，其他原理类似。

[32] 每一个 PE 设备都可以按照控制平面和转发平面分离的方式进行报文转发。如图 3 所示的 PE 设备的逻辑功能结构，可以在控制平面 31 上生成转发报文依据的“转发表”，可以在转发平面 32 根据该“转发表”中的转发表项进行报文转发。PE 堆叠系统的各个 PE 设备中通常都会存储相同的转发表以转发报文，比如，为了发送单播报文（上行或下行），第一 PE 堆叠系统 1000 中的 PE 11-PE 14 存储一份完全相同的单播转发表。然而，在本公开示例中，PE 堆叠系统中的每个 PE 设备都可进行表项转换，以将上述完全相同的转发表转换为各自的包括本地转发表项的本地转发表。不同的 PE 设备生成的本地转发表项可以是不同的，并各自根据自己的本地转发表项进行报文转发。

[33] 图 4 示例了单个 PE 设备执行的报文转发方法，包括以下步骤。

[34] 在步骤 401，从本地入端口接收报文。

[35] 在步骤 402，根据所述本地入端口查找所存储的本地转发表项以确定与所述本地入端口相对应的本地出端口，并通过所述本地出端口对所述报文进行转发。

[36] 上述的图 4 描述了根据本公开的报文转发方法，PE 堆叠系统中的各个 PE 设备是如何转发报文的。并且，由上述流程可以看到，PE 设备是根据自身存储的本地转发表项进行报文转发。PE 设备可以根据接收报文的本

地入端口查找本地转发表项，以确定与本地入端口对应的本地出端口，并通过该本地出端口来转发报文。以图 2 中的 PE 12 为例，当从 PC 1 向 PC 2 沿顺时针方向转发报文时，如表 1 所示，PE 12 根据本地转发表项将从本地入端口 P1 接收的报文从本地出端口 S1 送出。

表 1 PE 12 的本地转发表

位置	动作	物理出端口
P1	端口重定向	S1

[37] 可见，在该方法中，PE 设备都是依据本地转发表进行报文转发。这种方式使得即使不对报文进行 tag 封装，PE 设备仍然可以根据本地转发表项转发报文，从而摆脱对 tag 封装的依赖，提高 PE 堆叠系统中 PE 设备选择的灵活性。

[38] 在另一个例子中，PE 设备在根据本地转发表转发报文时，如果接收该报文的本地入端口是非堆叠口，那么可以按照图 4 示例的方式，根据本地入端口查找本地转发表来转发报文。而如果接收该报文的本地入端口是堆叠口，那么为了使得对报文的转发更加准确，可以使用根据本地入端口和扩展通道标识（E-CID）进行报文转发的方式。

[39] 例如，可以参见图 5 所示的流程，在步骤 501 中，当 PE 设备接收报文的本地入端口为边缘非堆叠口时，比如可以是图 2 中的 PE 11 和 PE 12 的 P1 端口，那么 PE 设备可以在步骤 502 中在报文中增加扩展通道标识 E-CID。这样，如步骤 503 中，当携带 E-CID 的报文由 PE 堆叠系统中的某个 PE 设备的作为本地入端口的堆叠口进入该 PE 设备时，该 PE 设备可以根据该本地入端口和报文中的扩展通道标识查找存储的本地转发表，以确定与所述本地入端口和 E-CID 对应的本地出端口，并通过该本地出端口对报文进行转发，参见步骤 504 所示。

[40] 此外，上述的本地转发表，可以是各个 PE 设备针对堆叠转发表中的堆叠转发表项进行表项转换得到，该“堆叠转发表”即上述提到的 PE 堆叠系统中的各 PE 设备上存储的完全相同的一份转发表。通常该堆叠转发表项中可以包括指示目的端口的目的端口信息（即报文由 PE 堆叠系统的哪个端口出去），该目的端口位于 PE 堆叠系统中的某个 PE 设备上，那么对于该 PE 设备来说，该目的端口是其本地端口，而对于其他 PE 设备来说，该目的端口是远端端口。本步骤中，每个 PE 设备都要对该堆叠转发表项进行解析，以转换成用于本 PE 设备转发报文依据的本地转发表项，而本地转发表项中包括的是 PE 设备的本地端口。

[41] PE 设备可以按照图 6 所示的流程进行表项转换得到本地转发表项，图 6 所示的流程包括以下步骤。

[42] 在步骤 601，根据 PE 堆叠系统的堆叠拓扑信息，按照预设规则计算从堆叠转发表项中的入端口到目的端口的报文转发路径。

[43] 在步骤 602，确定所述报文转发路径经过的本地端口以生成本地转发表项。

[44] 结合图 6 和图 7 的示例，来说明表项转换的原理。假设“堆叠转发表项”中指定了入端口为 PE 12 的端口 P1，并指定了出端口为 PE 14 的端口 P2。在进行表项转换时，PE 设备可以根据该入端口、出端口，并结合堆叠拓扑信息，计算从入端口到达出端口所经过的路径。该堆叠拓扑信息指 PE 堆叠系统中各个 PE 设备之间的连接关系（例如，PE 12 通过堆叠口 S1 连接 PE 11 的堆叠口 S2），其对于每个 PE 设备是已知的。在此基础上，各 PE 设备使用相同的内部选路算法，计算从堆叠转发表项中的入端口到出端口的报文转发路径。由于使用的选路算法相同，各个 PE 设备得到的路径是相类似的。例如图 7 所示，以 PE12 的 P1 端口为入端口，PE14 的 P2 端口为出端口时，路径为：PE 12 的端口 P1——PE 12 的端口 S1——PE 11 的端口 S2——PE 11 的端口 S1——PE 14 的端口 S2——PE 14 的端口 P2。

[45] 以 PE 11 为例，按照上述计算得到的路径，PE 11 可以获知，报文是从其端口 S2 进入 PE 11，并从端口 S1 出 PE 11，即报文转发路径上经过的本地端口是 S2 和 S1。据此，PE 11 生成的本地转发表项中可以包括入端口 S2 和出端口 S1。比如，表项中包括的表项下发位置是入端口 S2，并且报文转发的本地出端口是S1。PE 11向入端口S2下发用于重定向报文的表项，该表项用于指示入端口 S2“在接收到满足诸如匹配 E-CID 等的条件的报文时，将该报文由出端口 S1 发出”。

[46] 不论是报文的上行转发或者下行转发，表项转换的原理是相同的，都是通过路径计算得到报文转发路径上经过的本地端口，以生成包含该本地端口的本地转发表项，这样各个 PE 设备就能根据各自的本地转发表项进行报文转发。

[47] 在封装 tag 的方式中，tag 中封装了报文的源地址和目的地址等信息，各个 PE 设备只能通过解析报文中的 tag 获取目的地址等信息并进行转发。如果取消了 tag 封装，PE 设备将无法转发报文。相对的，本公开示例中，即使不封装 tag，各个 PE 设备能够根据“本地转发表项”转发报文，相当于各 PE 设备只负责报文在自身设备的转发即可，通过各个 PE 设备的相继转发最终将报文送出 PE 堆叠系统。

[48] 由于取消了 tag 封装，各 PE 设备之间可直接采用 802.1BR 协议进行报文传递，也极大地扩展了 PE 设备选取的灵活性。例如，不同芯片厂家之间的 PE 设备可进行混合堆叠，或者同芯片厂家不同芯片之间进行混合堆叠，即使是不支持 tag 封装的设备也能参与 PE 堆叠，PE 堆叠系统的组网更加灵活，也能节省组网成本。

[49] 如下，分别以单播上行、单播下行、组播下行等的情况为例，结合图 2 所示的组网结构来详细描述本公开示例的报文转发方法中 PE 设备如何进行表项转换。

[50] 单播上行：单播报文的上行转发的情况下，PE 堆叠系统要将 PC 发送的报文向 CB 设备转发。例如，若 PC 1 向 PC 3 发送报文，PC 1 可以先将报文发送至第一 PE 堆叠系统 1000，由第一 PE 堆叠系统 1000 将报文上送至 CB 堆叠系统 2000 中的 CB 设备。

[51] 如下的表 2，示例了第一 PE 堆叠系统 1000 中的 PE 11-PE 14 上都存储的一份相同的堆叠转发表。以该表中的第一条表项为例，该表项指定了在 PE 12 的端口 P1 下发重定向规则，以及指明报文的目的出端口是聚合口（Trunk Port），包括 PE 11 的端口 P1，以及 PE 14 的端口 P1 和 P2。也就是说，当 PE 12 的端口 P1 接收到匹配的报文时，将报文发往 PE 11 的端口 P1 或者 PE 14 的端口 P1 和 P2 以送出 PE 堆叠系统。

表 2 堆叠转发表

位置	动作	物理出端口
PE 12/P1	端口重定向	Trunk Port (PE 11/P1, PE 14/P1, PE 14/P2)
PE 13/P1	端口重定向	Trunk Port (PE 11/P1, PE 14/P1, PE 14/P2)

[52] 以该表 2 中的第一条表项为例，描述各个 PE 设备是如何将堆叠转发表项转换为各自的本地转发表项。PE 11 在控制平面上根据报文的入端口、出端口以及 PE 堆叠系统的堆叠拓扑信息，通过诸如最短路径选路算法等的内部选路算法，计算出从报文的入端口到出端口的路径。例如，计算出的从 PE 12 的端口 P1 到物理出端口 “Trunk Port (PE 11/P1, PE 14/P1, PE 14/P2)” 所经过的路径参见图 8 所示。PE 11 可以确定该路径上经过的本地端口包括：用于接收报文的本地入端口（即堆叠口 S2）、以及用于发出报文的本地出端口（即本地端口 P1、以及堆叠口 S1）。

[53] 根据上述得到的本地端口，PE 11 可以获知重定向规则的下发位置为接收报文的堆叠口 S2，以及报文的目的端口为作为聚合口（Trunk Port）

的本地端口 P1、以及堆叠口 S1。这是由于宏观上 PE 堆叠系统与 CB 设备之间可能存在多条聚合链路，为了使各条聚合链路均匀分担流量，目的端可以为包括多个端口的聚合口。其中，报文可以通过聚合口中的任一端口发出。例如，可以通过哈希算法根据报文的信息动态选择聚合口中的一个端口来发出报文，由此实现流量负载分担。由此得到 PE 11 的本地转发表如下：

表 3 PE 11 的本地转发表

位置	动作	物理出端口
S2	ACL 重定向 （匹配 E-CID 100）	Trunk (P1, S1)

[54] 如表 3 所示，PE 11 根据该本地转发表，向入端口 S2（命令下发位置）下发重定向规则，指示报文的本地出端口是“Trunk (P1, S1)”。其中，该重定向规则是 ACL（Access Control List，访问控制列表）重定向规则，因为规则下发到的本地入端口 S2 是一个用于 PE 设备间连接的堆叠口。堆叠口的特点是经过的业务流可能来自多个不同的设备，例如，经过堆叠口 S2 的业务流不仅有来自 PC 1 的流量，还有来自其他设备的流量，比如 PC 2 发出的报文经 PE 13、PE 12 到达 PE 11。因此堆叠口的报文匹配包括两方面，即要分别对接收报文的本地端口的端口号以及报文中的 E-tag 中携带的 E-CID（扩展通道标识）字段进行匹配。根据表 3 所示，该报文要满足“由 PE 11 的端口 S2 进入，且报文的 E-tag 中携带的 E-CID 字段是 100（即该报文是 PE 12 的 PCID 为 100 的端口接收的报文）”。PE 11 可根据表 3 的表项，在其本地的堆叠口 S2 下发 ACL 重定向规则，如果 PE 11 接收的报文满足匹配入端口 S2 和报文的 E-CID 100 的条件，就将报文送往本地端口 P1 以及 S1。

[55] 上述是以 PE 11 为例，说明了 PE 11 如何将表 2 中的表项“PE 12/P1，端口重定向，Trunk Port (PE 11/P1,PE 14/P1,PE 14/P2)”转换为表 3 所示的 PE 11 的本地转发表项，并且 PE 11 根据该表项向本地端口 S2 下发了 ACL 重定向规则，用于指示报文在 PE 11 的转发。其他 PE 设备（例如，PE 12、PE 13、PE 14）对上述表 2 中的表项的转换方式与 PE 11 相同，不再详述，参见图 9 所示的表项转换的对应关系，图 9 示例了各个 PE 设备分别转换堆叠转发表项后得到的对应的本地转发表项。

[56] 其中，需要说明的是，PE 设备的非堆叠口上下发的是“端口重定向规则”，“端口重定向规则”相比于“ACL 重定向规则”来说，只需要报文匹配入端口即可。例如，以 PE 12 为例，根据图 8 所示的报文转发路径，报文从 PE 12 的本地端口 P1 进入，从 PE 12 的本地端口 S1 发出，那么 PE 12 的命令下发位置是 P1，该 P1 是非堆叠口，采用端口重定向规则即可，只要报文是从端口 P1 进入，对应的出端口是 S1，参见图 9 中 PE 12 上对应堆叠转发表项第一条的本地转发表项。此外，对于 PE 13，由于报文的转发路径不会经过 PE 13，因此 PE 13 上不会有转换的本地转发表项。

[57] 单播下行: 在 PE 堆叠系统将 PC 发送的报文上送至 CB 设备后，CB 设备经过查表还会将报文发回 PE 堆叠系统，由 PE 堆叠系统将报文最终发往目的设备（例如 PC 3），因此，单播下行即 PE 堆叠系统将 CB 设备发送的报文向 PC 转发。

[58] 图 10 示例了在单播下行转发报文时，将堆叠转发表项转换为各个 PE 上对应的本地转发表项，表项转换的原理与单播上行时类似，不再详述。在单播下行时的表项转换，根据 E-CID 属性值所指示的位置以及 PE 堆叠系统的最短路径算法，共同确定下行转发报文时经过的 PE 设备和端口。

[59] 参见图 10，以堆叠转发表中的第二条表项（101，PE 13/P1）的转换为例，结合图 2 所示的组网结构，来描述表项的转换。例如，E-CID 101 所指示的位置为 PE 13 的端口 P1，如果 PE 11 要寻址到 E-CID 101 所指示

的位置，有两条路径可选择，PE 11-PE 14-PE 13，或者，PE 11-PE 12-PE 13，但是选路算法将选择一条路。假设各个 PE 设备的选路算法都会选择顺时针方向的路径，即 PE 11-PE 14-PE 13，那么报文在 PE 11 的出端口为端口 S1，在 PE 14 的出端口为端口 S1，在 PE 13 上的出端口是端口 P1。而 PE 12 去往 E-CID 101 所指示位置的最短路径只能是经过 PE 12 的端口 S2 直接到 PE 13，因此报文在 PE 12 的出端口是 PE 12/S2。可见，在单播下行的表项转换时，每个 PE 设备将分别计算从堆叠转发表项中的入端口到出端口的路径，确定该路径经过的本地出端口，当接收到报文时，将报文由该本地出端口发出。

[60] 此外，在进行表项转换时，如果堆叠转发表项中的目的端口为聚合口，那么在确定报文转发路径经过的本地端口时，可以按照图 11 的流程确定本地端口。

[61] 在步骤 1101，分别计算从所述堆叠转发表项中的入端口到达所述聚合口中的各个端口的路径，并根据所述路径确定：用于接收报文的本地入端口、作为部分所述聚合口的第一本地端口、以及到达位于远端 PE 设备的部分所述聚合口时经过的第二本地端口。

[62] 例如，仍然结合图 8 的示例，堆叠转发表项中的目的端口即出端口是聚合口（trunk port），那么 PE 11 在确定本地出端口时，计算从所述堆叠转发表项中的入端口到达聚合口中的各个端口的路径，确定的第一本地端口是 PE 11 的 P1 端口，确定的到达位于远端 PE 设备即 PE 14 的部分聚合口时经过的第二本地端口是 PE 11 的 S1。而对于 PE 14 来说，确定的第一本地端口是 PE 14 的 P1 和 P2，而第二本地端口是 PE 14 的 S2。

[63] 在步骤 1102，在确定所述第二本地端口与接收报文的本地入端口不同时，将所述第一本地端口和第二本地端口作为本地出端口；在确定所述第二本地端口与所述本地入端口相同时，将所述第一本地端口作为本地出端口。

[64] 例如，对于 PE 11 来说，报文是从端口 S2 进入 PE 11，即 PE 11 的本地入端口是 S2，而第二本地端口是 S1，可见，第二本地端口与本地入端口不同，那么 PE 11 可以将端口 P1 和 S1 作为本地出端口。而对于 PE 14 来说，报文是从 PE 14 的端口 S2 进入，即 PE 14 的本地入端口是 S2，而 PE 14 的第二本地端口也是 S2，那么第二本地端口与本地入端口相同，PE 14 可以将端口 P1 和 P2 作为本地出端口，而不将端口 S2 作为本地出端口。

[65] 在上述描述单播上行和单播下行的“表项转换”的基础上，如下将结合图 9 和图 10 中的各个 PE 设备进行表项转换后得到的本地转发表项，说明 PE 堆叠系统根据该本地转发表项进行单播报文转发的过程，并且分别以 PE 堆叠系统间的单播报文转发、和 PE 堆叠系统内的单播报文转发为例。

[66] 如图 12 的示例，图 12 中用带箭头的虚点线表示报文转发的路径，并用圆圈中的数字表示该路径上的各个转发处理步骤。此外，由于图 12 示例的是 PC 1 发送报文至 PC 3 的 PE 堆叠系统间的报文转发，因此还需要依据 PE 15 上的本地转发表。

表 3 PE 15 的本地转发表

E-CID	物理出端口
200	PE 15/P1

[67] 图 13 示出了图 12 中的在 PE 堆叠系统间转发报文的步骤，可包括如下步骤 1001-1005。

[68] 在步骤 1001，PC 1 发送以太报文至第一 PE 堆叠系统 1000 中的 PE 12。

[69] 其中，报文中携带目的设备 PC 3 的 MAC 地址，PC 1 发送的该以太报文从 PE 12 的端口 P1 进入第一 PE 堆叠系统 1000。

[70] 在步骤 1002，PE 12 根据本地存储的端口重定向规则，将接收的报文从端口 S1 发往 PE 11。

[71] 其中，PE 12 将接收到的 PC 1 发送的报文，封装成不具有 Hig tag 的 802.1BR 报文，并在报文的 E-CID 字段填入为 PE 12 的端口 P1 分配的 PCID 100，在报文中的 ingress E-CID 字段填 0；或者根据具体的芯片，也可以是在 E-CID 字段填 0，在 ingress E-CID 字段填 100。

[72] 并且，如上面示例描述的，PE 12 会在其本地端口 P1 下发端口重定向规则，只要报文匹配入端口（由端口 P1 进），就将报文由堆叠口 S1 发出，参见图 9 中的 PE 12 上的第一条表项。则 PE 12 根据端口重定向规则将报文从堆叠口 S1 发出。

[73] 在步骤 1003，PE 11 根据 ACL 重定向规则，将报文从端口 P1 送出，到达 CB 10。

[74] 其中，PE 11 根据其本地转发表项，已经在本地的堆叠口 S2 下发了 ACL 重定向规则，只要报文从端口 S2 进入 PE 11，并且报文中的 E-tag 中的 E-CID 是 100，则将报文从 Trunk (P1，S1)送出。本步骤中，PE 11 收到 PE 12 发来的 802.1BR 报文，确定其命中了 ACL 重定向规则，且本地出端口是 trunk 口，则 PE 11 从 trunk 口中随机选择一个端口送出报文即可（即从聚合口的其中一个端口发出报文即可），假设本示例 PE 11 将报文从本地端口 P1 送出，向上到达 CB 10。

[75] 在步骤 1004，CB 10 将通过查表来确定报文的出端口为端口 P2，以通过该端口 P2 将报文发出到第二 PE 堆叠系统 3000 中的 PE 15。

[76] 其中，CB 10 对于接收到的报文，根据目的 MAC（Destination media access control, DMAC）地址与虚拟局域网（VLAN）结合的方式查表得到目的 E-CID 是 200，以及对应的本地出端口是 P2，则将 E-CID 的属性值 200 填入报文的 E-tag 的 E-CID 字段，替换原 E-CID100，ingress E-CID 字

段仍旧填 0，并将报文从端口 P2 发出。报文从 CB 10 的端口 P2 发出后，将到达第二 PE 堆叠系统 3000 中的 PE 15，从 PE 15 的端口 P2 进入第二 PE 堆叠系统 3000。

[77] 在步骤 1005，PE 15 根据本地转发表项，将报文从端口 P1 发出，到达 PC 3。

[78] 其中，PE 15 根据上述表 3 所示的本地转发表项，获知 E-CID 200 的报文对应本地出端口是 PE 15/P1，则将报文剥去 E-tag 后从 P1 端口发出。

[79] 最终 PC 3 接收到该报文，PC 1 发出的报文经过第一 PE 堆叠系统 1000、CB 堆叠系统 2000、第二 PE 堆叠系统 3000，到达了目的设备 PC 3，并且通过图 13 所示流程可以看到，在该报文转发的过程中，PE 堆叠系统中的各个 PE 设备，都是根据本地转发表项进行报文转发的。

[80] 图 14 示例了 PE 堆叠系统内转发报文，即报文转发依据同一个 PE 堆叠系统的路径，例如 PC 1 发送报文至 PC 2。并且，PE 堆叠系统中的各个 PE 设备仍然依据图 9 和图 10 中转换得到的本地转发表项进行转发。图 15 示出了该 PE 堆叠系统内转发报文的流程，该流程可包括如下步骤。

[81] 在步骤 1201，PC 1 发送以太报文至 PE 12。

[82] 其中，PC 1 发送的该以太报文中携带目的设备 PC 2 的 MAC 地址，并从 PE 12 的端口 P1 进入第一 PE 堆叠系统 1000。

[83] 在步骤 1202，PE 12 根据端口重定向规则，将报文从端口 S1 发往 PE 11。

[84] 其中，PE 12 将接收到的 PC 1 发送的报文，封装成不具有 Hig tag 的 802.1BR 报文，并在报文的 E-CID 字段填入端口标识符 100，报文中的 ingress E-CID 字段填 0。

[85] 并且，如上面示例描述的，PE 12 会在其本地端口 P1 下发端口重定向规则，只要报文匹配入端口（由端口 P1 进），就将报文由堆叠口 S1 发

出，参见图 9 中的 PE 12 上的第一条表项。则 PE 12 根据端口重定向规则将报文从端口 S1 发出。

[86] 在步骤 1203，PE 11 根据 ACL 重定向规则，将报文从端口 S1 送出。

[87] 其中，PE 11 根据其本地转发表项，已经在本地的堆叠口 S2 下发了 ACL 重定向规则，只要报文从端口 S2 进入 PE 11，并且报文中的 E-tag 中的 E-CID 是 100，则将报文从 Trunk (P1，S1)送出。

[88] 本步骤中，PE 11 收到 PE 12 发来的 802.1BR 报文，确定其命中了 ACL 重定向规则，且本地出端口是 trunk 口，则 PE 11 从 trunk 口中随机选择一个端口送出报文即可（即从聚合口的其中一个端口发出报文即可），假设本示例 PE 11 将报文从本地端口 S1 送出到达 PE 14。

[89] 在步骤 1204，PE 14 将报文从其本地端口 P1 发往 CB 20。

[90] 其中，PE 14 在通过堆叠口 S2 接收到报文后，也要进行 ACL 重定向，确定报文的本地出端口是 Trunk（P1，P2），则 PE 14 随机选择端口 P1 发出报文，报文到达 CB 20。

[91] 在步骤 1205，CB 20 将通过查表确定报文的出端口为端口 P2，以通过端口 P2 将报文发出到 PE 14。

[92] 其中，CB 20 对于接收到的报文，根据 DMAC+VLAN 的方式查表得到目的 E-CID 是 101，以及对应的本地出端口为 trunk 口，假设 CB 20 将 E-CID 101 填入报文的 E-tag 的 E-CID 字段，替换原 E-CID 100，并随机选择端口 P2 发出报文，到达第一 PE 堆叠系统 1000 的 PE 14。其中，为了实现源过滤，即在发送源过滤由该发送源自身发出的报文（比如不能再接收自身发送的广播报文），可以在用于携带源通道信息的 Ingress E-CID 字段中填入为源端口分配的 PCID 的值，例如在 ingress E-CID 字段中填入 100。

[93] 在步骤 1206，PE 14 将报文从端口 S1 发出，到达 PE 13。

[94] 其中，结合图 10 中的 PE 14 的转发表项，PE 14 将 E-CID 是 101 的

报文从端口 S1 发出，然后报文到达 PE 13。

[95] 在步骤 1207，PE 13 将报文的 E-tag 剥去后从端口 P1 发出到 PC 2。

[96] 其中，结合图 10 中的 PE 13 的转发表项，PE 13 将 E-CID 是 101 的报文从端口 P1 发出，然后报文到达目的设备 PC 2。

[97] 当报文是多播报文（或广播报文）时，将报文向 CB 设备发送的上行转发过程，与单播上行转发的原理类似，不再赘述，但是在下行转发时，多播报文转发与单播报文转发有不同之处。

[98] 以环形堆叠情况下的组播报文转发为例，在组播报文进入 PE 堆叠系统的入口 PE 设备后，入口 PE 设备会将该报文沿顺时针和逆时针方向的本地出端口都发出，其余的 PE 设备为了避免报文的传输形成环路，遵照如下的堆叠口转发原则：传输路径是最短路径的报文允许被转发，传输路径非最短路径的报文不允许被转发，等价路径只选其中一条。

[99] 例如，参见图 16 的示例，PE 14、PE 13、PE 12 和 PE 11 形成环形 PE 堆叠系统，组播报文从 PE 14 进入该 PE 堆叠系统，假设报文的组播地址中包括多个目的 PC，其中的两个目的 PC 是 PC 1 和 PC 2。PE 14 将报文分别从其本地端口 S7 和 S8 沿顺时针和逆时针方向发送。根据上述的堆叠口转发原则，当 PE 12 接收到来自端口 S3 和 S4 的报文时，确定端口 S3 和 S4 到入口 PE 14 的距离相同（都是两跳），PE 12 可以选择转发其中一个端口的报文，另一个端口的报文被阻塞，以避免重复发送。例如，PE 12 可以将逆时针方向从端口 S3 来的报文转发给 PC 1，但是对于顺时针方向通过端口 S4 接收的报文，PE 12 阻塞不继续发送，相当于顺时针方向在 PE 12 的端口 S4 被阻塞。同理，对于 PE 13 来说，由于通过端口 S6 接收的报文的传输路径是最短路径，将该报文转发给 PC 2，但是沿逆时针方向通过端口 S5 接收的报文的传输路径，不是距离 PE 14 的最短路径，因此通过 PE 13 的端口 S5 接收到的报文被阻塞。

[100] 与图 16 示例的处理相对应的，图 17 示例了本公开示例的组播报文的转发方式，即，报文并不是在接收该报文的 PE 设备的入端口被阻塞，而是直接在发出该报文的 PE 设备的出端口被阻塞。例如，在图 16 中，PE 12 将报文从堆叠口 S4 逆时针发送至 PE 13，PE 13 在接收该报文的堆叠口 S5 阻塞报文，不继续发送；而在图 17 中，PE 12 不再将报文从堆叠口 S4 发出，直接在其本地端口阻塞报文。同理，在图 16 中，针对顺时针方向从 PE 13 的堆叠口 S5 发送至 PE 12 的堆叠口 S4 的报文，该报文在 PE 12 的堆叠口 S4 被阻塞；而在图 17 中，该报文直接在 PE 13 的本地堆叠口 S5 被阻塞，不会发送至 PE 12。而对于 PE 从自身的哪个端口阻塞报文，可以根据堆叠组播的阻塞算法确定。

[101] 综上，PE 设备在根据报文转发路径生成组播转发表时，还要根据堆叠组播的阻塞算法生成用于阻断环路流量的本地阻塞端口，并且从所述本地转发表项中的本地出端口中删除所述本地阻塞端口。

[102] 结合图 18 所示，示例了组播下行的报文转发，假设 PC 1 是源设备，PC 2、PC 3 和 PC 4 是目的设备，并且假设组 E-CID 是 5000。图 19 还示例了各个 PE 设备上的组播下行转发表项，结合图 20 的流程，描述 PE 堆叠系统的 PE 设备根据图 19 的表项下行转发组播报文的流程。

[103] 在步骤 1701，CB 10 将报文从端口 P1 和 P2 发出。

[104] 其中，假设 CB 10 查找本地转发表项确定要经过出端口 P1 和 P2 来到达该组播报文的目的设备，CB 10 在报文的 E-CID 字段中填入 5000。报文从 CB 10 的出端口 P1 发出到达第一 PE 堆叠系统 1000，PE 11 作为入口 PE 设备；报文从 CB 10 的出端口 P2 发出到达第二 PE 堆叠系统 3000，PE 15 作为入口 PE 设备。

[105] 在步骤 1702，PE 11 收到报文，将该报文从本地端口 S1 和 S2 发出。

[106] 在步骤 1703，PE 12 收到报文，根据图 19 所示的下行本地转发表项，

将报文从本地端口 P1 和 S2 发出；其中，报文从端口 P1 发出后到达其中一个组播目的设备 PC 4。

[107] 在步骤 1704，PE 13 收到报文，根据图 19 所示的下行本地转发表项，将报文剥去 E-tag 后从本地端口 P1 送出，报文到达另一个组播目的设备 PC 2。

[108] 需要说明的是，PE 13 已经根据堆叠组播的阻塞算法确定了本地阻塞端口是堆叠口 S2，因此，PE 13 接收到 PE 12 发来的报文后，不会再继续通过堆叠口 S2 往 PE 14 发送，报文在逆时针方向上在 PE 13 的堆叠口 S2 处被阻塞。图 19 的表项中的出端口也没有堆叠口 S2。

[109] 在步骤 1705，PE 14 接收到报文，查找本地转发表项发现没有目的端口，不进行处理。

[110] 其中，PE 14 接收到的是 PE 11 顺时针发送的报文，并且，PE 14 也通过堆叠组播的阻塞算法确定了本地阻塞端口是堆叠口 S1，因此在图 19 的 PE 14 表项中的出端口不包括堆叠口 S1，则 PE 14 将报文在顺时针方向上在本地端口 S1 处阻塞，不继续发往 PE 13。

[111] 在步骤 1706，PE 15 将报文从本地端口 P1 发出，到达又一个组播目的设备 PC 3。

[112] 此外，当 PE 堆叠系统的堆叠拓扑发生变化，或者 PE 设备的端口发生变化（例如，PE 设备与 CB 设备连接的端口，或者 PE 设备之间连接的堆叠口）时，各个 PE 设备将获取更新后的堆叠拓扑信息，并根据更新后的堆叠拓扑信息重新计算本地端口，更新用于报文转发的本地转发表项。另外，本公开示例中的堆叠口不是 802.1BR 协议定义的 cascade port/upstream port/extended port 中的任何一种，而是单独的另外一类端口，无方向性。

[113] 图 21 提供了一种 PE 设备，该 PE 设备可以包括处理器

(processor)1801、通信接口(Communications Interface)1802、存储介质 1803 和通信总线 1804。处理器 1801、通信接口 1802、存储介质 1803 可通过通信总线 1804 完成相互间的通信。通信接口 1802 用于与网元通信，比如与其他 PE 设备通信。处理器 1801 用于调用执行存储介质 1803 中存储的机器可执行指令 1805。处理器 1801 可以是一个中央处理器（CPU）。存储介质 1803 例如可以是非易失性存储器（non-volatile memory）。

[114] 所述的机器可执行指令 1805 可以对应于报文转发控制逻辑，并从逻辑上可以划分为图 22 所示的功能模块，包括报文接收模块 2201 和转发处理模块 2202。PE 设备通过处理器执行该机器可执行指令，可以进行上述的报文转发。其中，报文接收模块 2201 可从本地入端口接收报文。转发处理模块 2202 可根据所述本地入端口查找存储的本地转发表项来确定与所述本地入端口对应的本地出端口，以通过所述本地出端口对报文进行转发。

[115] 进一步的，当所述本地入端口为边缘非堆叠口时，所述报文接收模块 2201 还可在从本地入端口接收报文之后，在所述报文中增加扩展通道标识。所述转发处理模块 2202 可当所述本地入端口为堆叠口时，根据本地入端口和报文中的扩展通道标识查找存储的本地转发表项，来确定与所述本地入端口和扩展通道标识对应的本地出端口，以通过所述本地出端口对所述报文进行转发。

[116] 图 23 示例了另一种报文转发控制逻辑的功能模块框图。与图 22 相比，该报文转发控制逻辑还可以包括路径计算模块 2203、表项生成模块 2204。其中，路径计算模块 2203 可根据所述 PE 堆叠系统的堆叠拓扑信息，按照预设规则计算从堆叠转发表项中的入端口到目的端口的报文转发路径。表项生成模块 2204 可确定所述报文转发路径经过的本地端口，并根据所述本地端口生成所述本地转发表项。

[117] 进一步的，路径计算模块 2203 可根据报文进入所述 PE 堆叠系统的入端口以及所述堆叠转发表项中的目的端口，计算将所述报文由所述入端

口转发至目的端口的报文转发路径。所述表项生成模块 2204，在确定所述报文转发路径经过的本地端口时，可包括：获得所述报文转发路径经过的用于接收报文的本地入端口、以及用于发出报文的本地出端口。

[118] 进一步的，表项生成模块 2204 可以包括：端口计算单元 2301、端口选择单元 2302。端口计算单元 2301 可在堆叠转发表项中的目的端口为聚合口时，分别计算从所述堆叠转发表项中的入端口到达所述聚合口中的各个端口的路径，并根据所述路径确定用于接收报文的本地入端口、作为部分聚合口的第一本地端口、以及到达位于远端 PE 设备的部分所述聚合口时经过的第二本地端口；端口选择单元 2302 可在确定所述第二本地端口与所述本地入端口不同时，将所述第一本地端口和第二本地端口作为本地出端口；在确定所述第二本地端口与所述本地入端口相同时，将所述第一本地端口作为本地出端口。

[119] 进一步的，表项生成模块 2204 还可以包括阻塞算法单元 2303。阻塞算法单元 2303 可在所述报文是多播报文时，根据堆叠组播的阻塞算法生成用于阻断环路流量的本地阻塞端口，并且从所述本地转发表项中的本地出端口中删除所述本地阻塞端口。

[120] 通过上述结构的 PE 设备，可以根据本地转发表项进行报文转发，并且还可以将堆叠转发表项转换为本地转换表项，使得 PE 设备可以不再依赖于 tag 封装，提高了 PE 堆叠系统中设备选择的灵活性。

[121] 所属领域的技术人员可以清楚地了解到，为描述的方便和简洁，上述描述的各个模块的具体工作过程，可以参考前述方法示例中的对应过程，在此不再赘述。

[122] 上述的报文转发控制逻辑的功能如果以软件功能单元的形式实现并作为独立的产品销售或使用时，可以存储在一个计算机可读取存储介质中。基于这样的理解，本公开的技术方案本质上或者说对现有技术做出贡献的

部分或者该技术方案的部分可以以软件产品的形式体现出来，该计算机软件产品存储在一个存储介质中，包括若干指令用以使得一台计算机设备（可以是个人计算机，服务器，或者网络设备等）执行本发明各个示例所述方法的全部或部分步骤。而前述的存储介质包括：U 盘、移动硬盘、只读存储器（ROM，Read-Only Memory）、随机存取存储器（RAM，Random Access Memory）、磁碟或者光盘等各种可以存储程序代码的介质。

[123] 以上所述示例仅为示例性的，并不用以限制本公开，凡在本公开的精神和原则之内，所做的任何修改、等同替换、改进等，均应包含在本公开保护的范围之内。

权利要求书

1. 一种报文转发方法，包括:

端口扩展（PE）堆叠系统中的PE设备从本地入端口接收报文;

所述 PE 设备根据所述本地入端口查找本地转发表项，以确定与所述本地入端口相对应的本地出端口;

所述PE设备经由所述本地出端口对所述报文进行转发。

2. 根据权利要求 1 所述的方法，还包括:

当所述本地入端口为非堆叠口时，所述 PE 设备在所述报文中增加扩展通道标识。

3. 根据权利要求 2 所述的方法，其中，当所述本地入端口为堆叠口时，所述PE设备根据所述本地入端口查找所述本地转发表项包括:

所述 PE 设备根据所述本地入端口和所述报文中的扩展通道标识查找所述本地转发表项，以确定与所述本地入端口和所述扩展通道标识对应的本地出端口。

4. 根据权利要求 1 所述的方法，在所述 PE 设备从所述本地入端口接收所述报文之前，还包括:

所述PE设备根据所述PE堆叠系统的堆叠拓扑信息，按照预设规则进行计算，以确定从堆叠转发表项中的入端口到目的端口的报文转发路径，并基于所述报文转发路径经过的本地端口生成所述本地转发表项。

5. 根据权利要求 4 所述的方法，其中，

所述PE设备根据所述PE堆叠系统的堆叠拓扑信息，按照预设规则进行计算，包括:

所述PE设备根据报文进入所述PE堆叠系统的入端口以及所述堆叠转发表项中的目的端口，计算将所述报文由所述入端口转发至所述目的端口的报文转发路径;

所述 PE 设备基于所述报文转发路径经过的本地端口生成所述本地转

发表项，包括:

所述 PE 设备确定所述报文转发路径经过的用于接收报文的本地入端口以及用于发出报文的本地出端口，并根据所确定的本地入端口和本地出端口生成所述本地转发表项。

6. 根据权利要求 4 所述的方法，其中，当所述堆叠转发表项中的目的端口为聚合口时，

所述 PE 设备根据所述 PE 堆叠系统的堆叠拓扑信息，按照预设规则进行计算，包括:

所述 PE 设备分别计算从所述堆叠转发表项中的入端口到达所述聚合口中的各个端口的路径，以及

所述 PE 设备基于所述报文转发路径经过的本地端口生成所述本地转发表项，包括:

所述 PE 设备根据所述路径确定用于接收报文的本地入端口、作为部分所述聚合口的第一本地端口、以及到达位于远端 PE 设备的部分所述聚合口时经过的第二本地端口;

在所述第二本地端口与所述本地入端口不同时，所述 PE 设备将所述第一本地端口和第二本地端口作为本地出端口;

在所述第二本地端口与所述本地入端口相同时，所述 PE 设备将所述第一本地端口作为本地出端口;

所述 PE 设备根据所确定的本地入端口和本地出端口生成所述本地转发表项。

7. 根据权利要求 4 所述的方法，其中，在所述 PE 设备基于所述报文转发路径经过的本地端口生成所述本地转发表项之后，还包括:

若所述报文是多播报文，则所述 PE 设备根据堆叠组播的阻塞算法生成用于阻断环路流量的本地阻塞端口，并且从所述本地转发表项中的本地出端口中删除所述本地阻塞端口。

8. 一种端口扩展（PE）设备，包括处理器，通过读取并执行机器可读存

储介质上所存储的与报文转发控制逻辑对应的机器可执行指令，所述机器可执行指令促使所述处理器：

从本地入端口接收报文；

根据所述本地入端口查找本地转发表项，以确定与所述本地入端口相对应的本地出端口；

经由所述本地出端口对所述报文进行转发。

9. 根据权利要求 8 所述的 PE 设备，其中，所述机器可执行指令还促使所述处理器：

当所述本地入端口为边缘非堆叠口时，在所述报文中增加扩展通道标识。

10. 根据权利要求 9 所述的 PE 设备，其中，当所述本地入端口为堆叠口时，所述机器可执行指令还促使所述处理器：

根据所述本地入端口和所述报文中的扩展通道标识查找所述本地转发表项，以确定与所述本地入端口和所述扩展通道标识对应的本地出端口。

11. 根据权利要求 8 所述的 PE 设备，在从所述本地入端口接收所述报文之前，所述机器可执行指令还促使所述处理器：

根据所述 PE 设备所在的 PE 堆叠系统的堆叠拓扑信息，按照预设规则进行计算，以确定从堆叠转发表项中的入端口到目的端口的报文转发路径，并基于所述报文转发路径经过的本地端口生成所述本地转发表项。

12. 根据权利要求 11 所述的 PE 设备，其中，

在根据所述 PE 堆叠系统的堆叠拓扑信息，按照预设规则进行计算时，所述机器可执行指令还促使所述处理器：

根据报文进入所述 PE 堆叠系统的入端口以及所述堆叠转发表项中的目的端口，计算将所述报文由所述入端口转发至所述目的端口的报文转发路径；

基于所述报文转发路径经过的本地端口生成所述本地转发表项时，所述机器可执行指令还促使所述处理器：

确定所述报文转发路径经过的用于接收报文的本地入端口以及用于发出报文的本地出端口，并根据所确定的本地入端口和本地出端口生成所述本地

转发表项。

13. 根据权利要求 11 所述的 PE 设备，其中，当所述堆叠转发表项中的目的端口为聚合口时，

在根据所述 PE 堆叠系统的堆叠拓扑信息，按照预设规则进行计算时，所述机器可执行指令还促使所述处理器：

分别计算从所述堆叠转发表项中的入端口到达所述聚合口中的各个端口的路径，

在基于所述报文转发路径经过的本地端口生成所述本地转发表项时，所述机器可执行指令还促使所述处理器：

根据所述路径确定用于接收报文的本地入端口、作为部分所述聚合口的第一本地端口以及到达位于远端 PE 设备的部分所述聚合口时经过的第二本地端口；

在所述第二本地端口与所述本地入端口不同时，将所述第一本地端口和第二本地端口作为本地出端口；

在所述第二本地端口与所述本地入端口相同时，将所述第一本地端口作为本地出端口；

根据所确定的本地入端口和本地出端口生成所述本地转发表项。

14. 根据权利要求 11 所述的 PE 设备，其中，在基于所述报文转发路径经过的本地端口生成所述本地转发表项之后，所述机器可执行指令还促使所述处理器：

若所述报文是多播报文，则根据堆叠组播的阻塞算法生成用于阻断环路流量的本地阻塞端口，并且从所述本地转发表项中的本地出端口中删除所述本地阻塞端口。

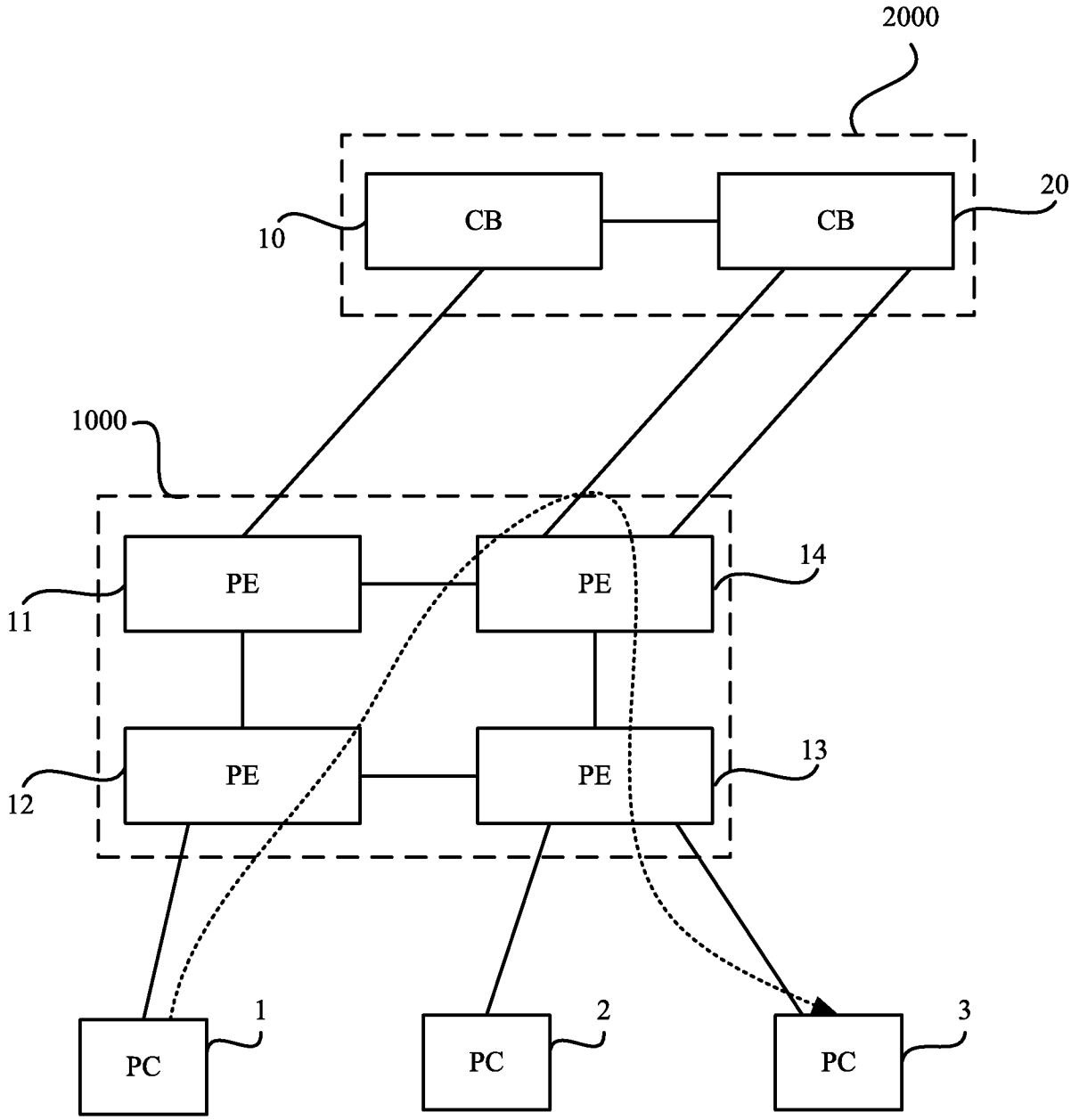


图 1

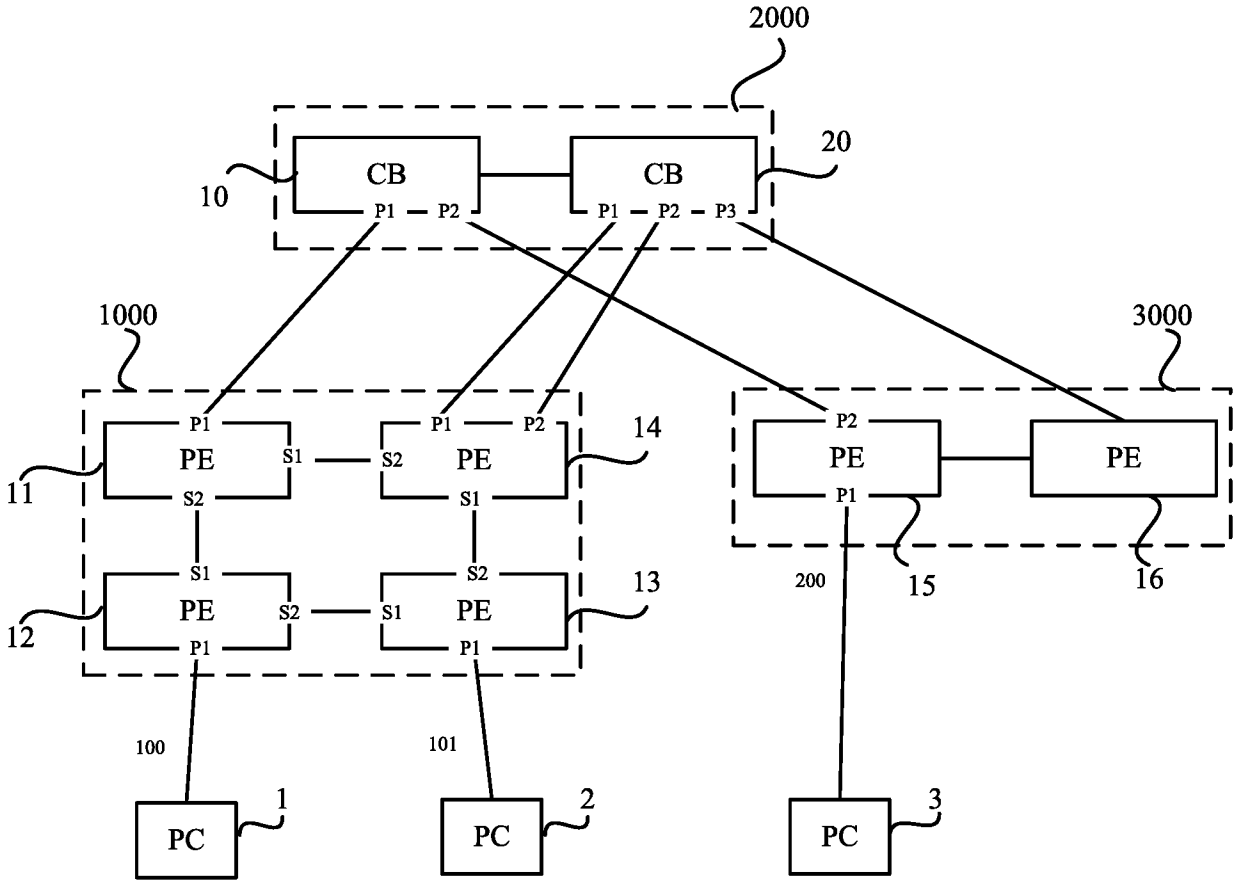

2000
CB
CB
20
10
P1
P2
P1
P2
P3
1000
3000
P1
PE
S1
S2
PE
14
11
P1
P2
PE
PE
S2
S1
S1
S2
PE
S2
S1
PE
13
12
P1
P1
200
15
16
100
101
PC
1
PC
2
PC
3

图 2

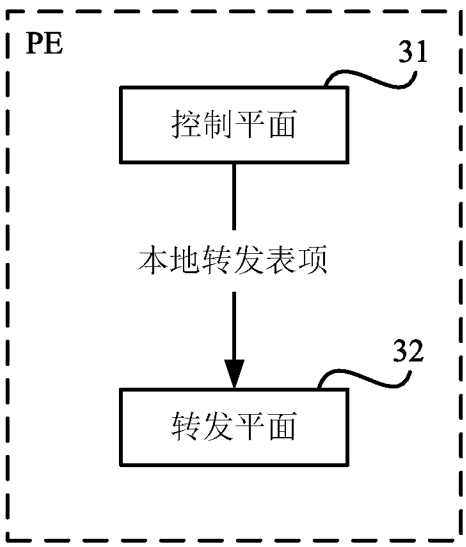

PE
31
控制平面
本地转发表项
32
转发平面

图 3

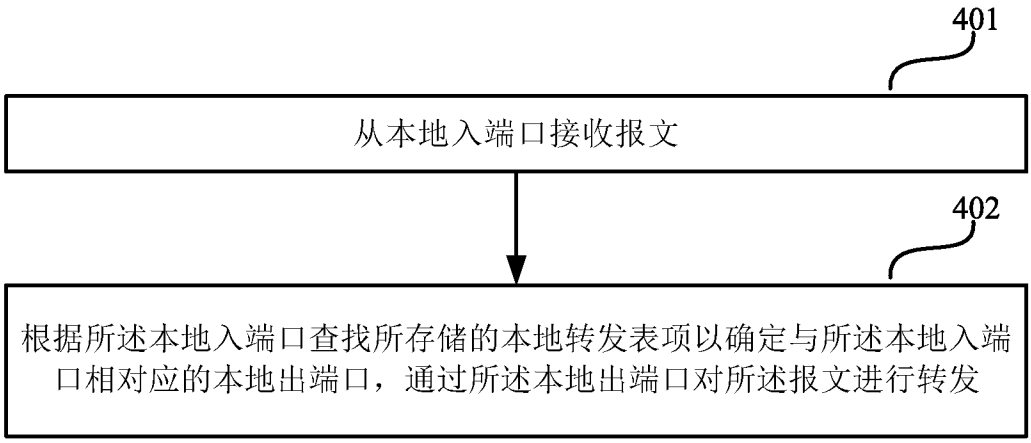

401
从本地入端口接收报文
402
根据所述本地入端口查找所存储的本地转发表项以确定与所述本地入端口相对应的本地出端口，通过所述本地出端口对所述报文进行转发

图 4

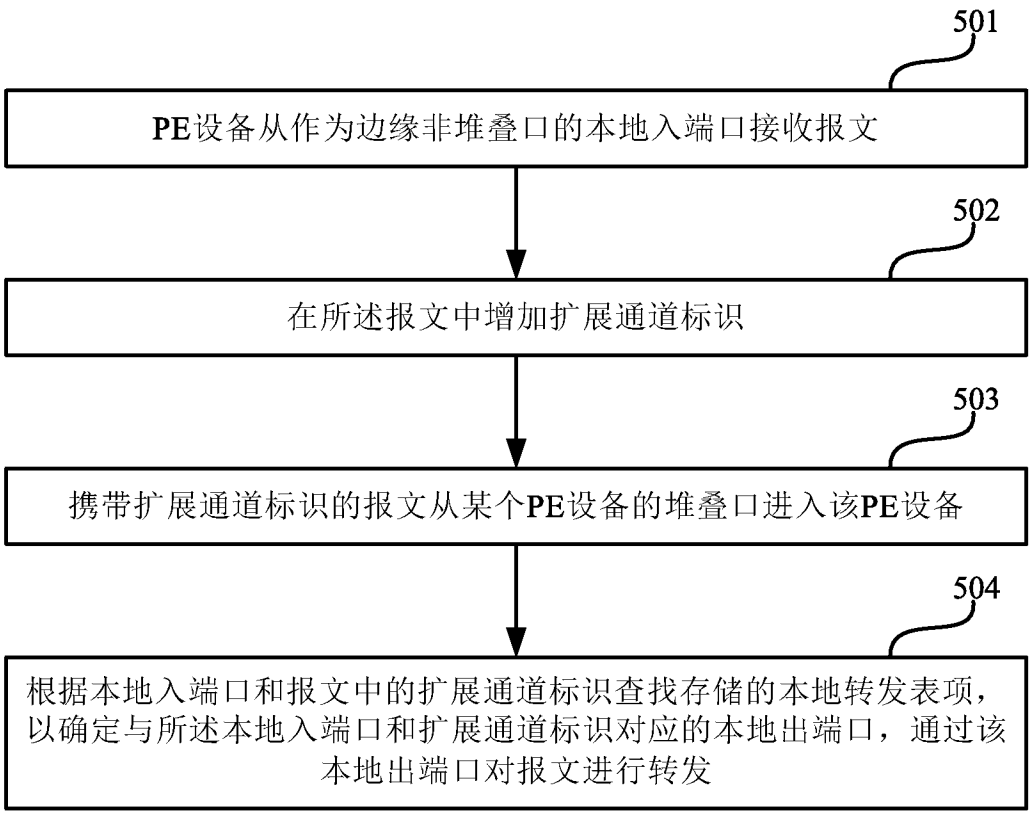

501
PE设备从作为边缘非堆叠口的本地入端口接收报文
502
在所述报文中增加扩展通道标识
503
携带扩展通道标识的报文从某个PE设备的堆叠口进入该PE设备
504
根据本地入端口和报文中的扩展通道标识查找存储的本地转发表项，以确定与所述本地入端口和扩展通道标识对应的本地出端口，通过该本地出端口对报文进行转发

图 5

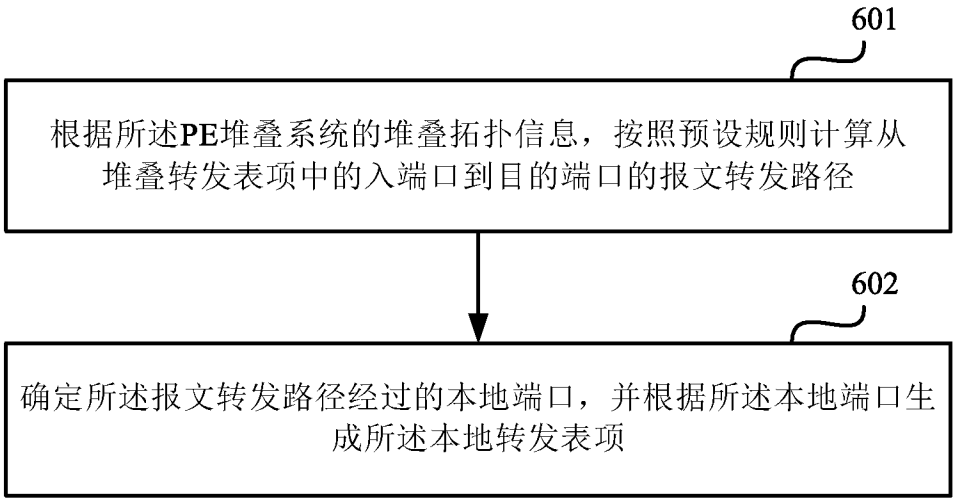

601
根据所述PE堆叠系统的堆叠拓扑信息，按照预设规则计算从堆叠转发表项中的入端口到目的端口的报文转发路径
602
确定所述报文转发路径经过的本地端口，并根据所述本地端口生成所述本地转发表项

图 6

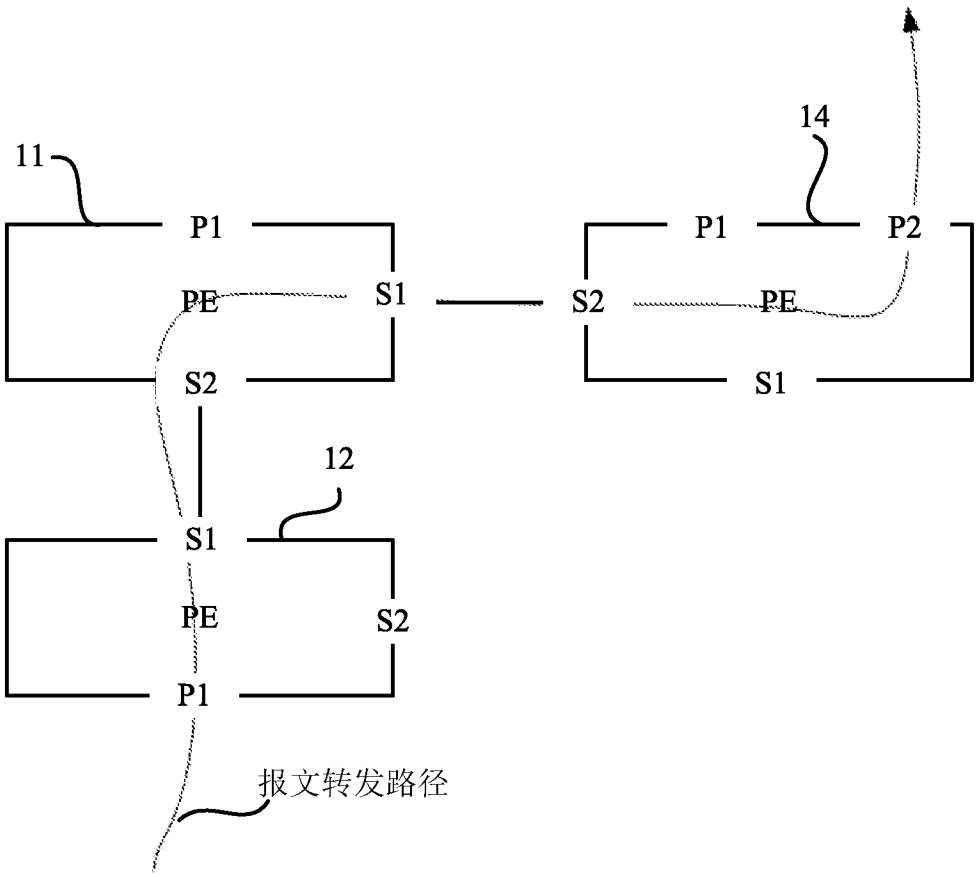

11
P1
PE
S1
S2
14
P1
P2
S2
PE
S1
12
S1
PE
S2
P1
报文转发路径

图 7

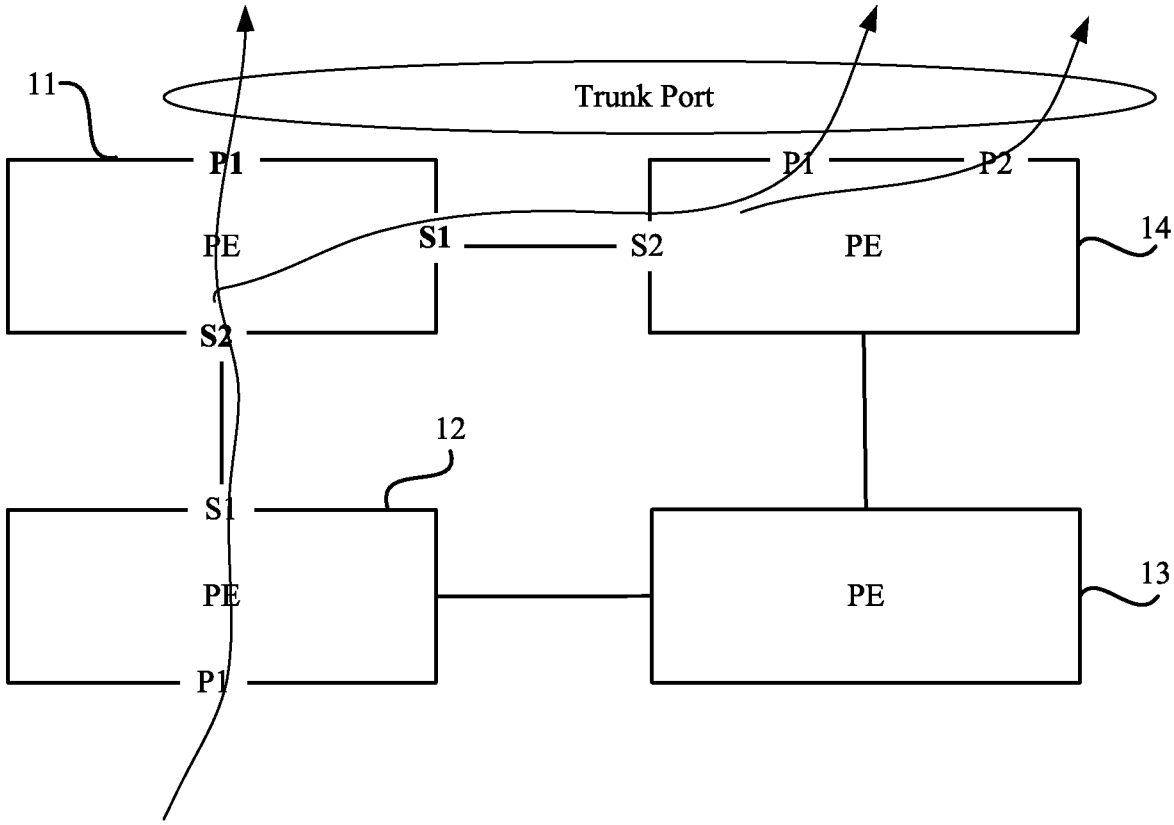

11
Trunk Port
P1
PE
S1
S2
P1
P2
S2
PE
14
12
S1
PE
P1
PE
13

图 8

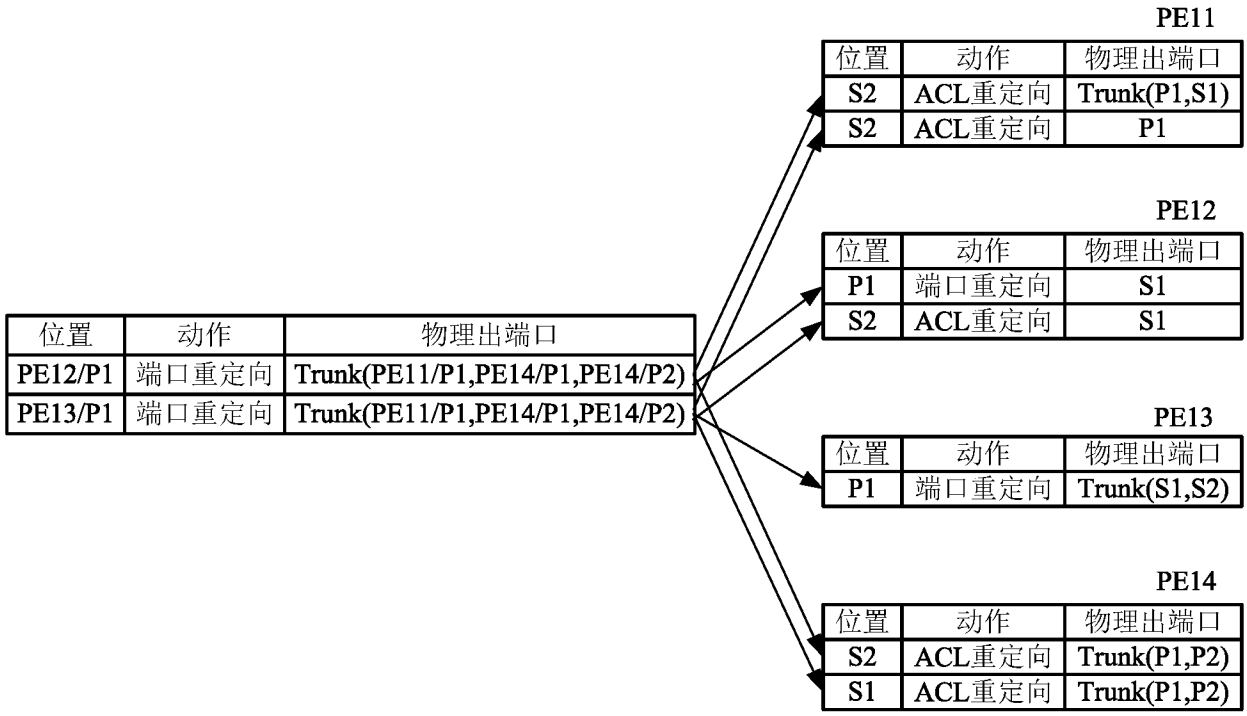

PE11
位置 动作 物理出端口
S2 ACL重定向 Trunk(P1,S1)
S2 ACL重定向 P1
PE12
位置 动作 物理出端口
P1 端口重定向 S1
S2 ACL重定向 S1
位置 动作 物理出端口
PE12/P1 端口重定向 Trunk(PE11/P1,PE14/P1,PE14/P2)
PE13/P1 端口重定向 Trunk(PE11/P1,PE14/P1,PE14/P2)
PE13
位置 动作 物理出端口
P1 端口重定向 Trunk(S1,S2)
PE14
位置 动作 物理出端口
S2 ACL重定向 Trunk(P1,P2)
S1 ACL重定向 Trunk(P1,P2)

图 9

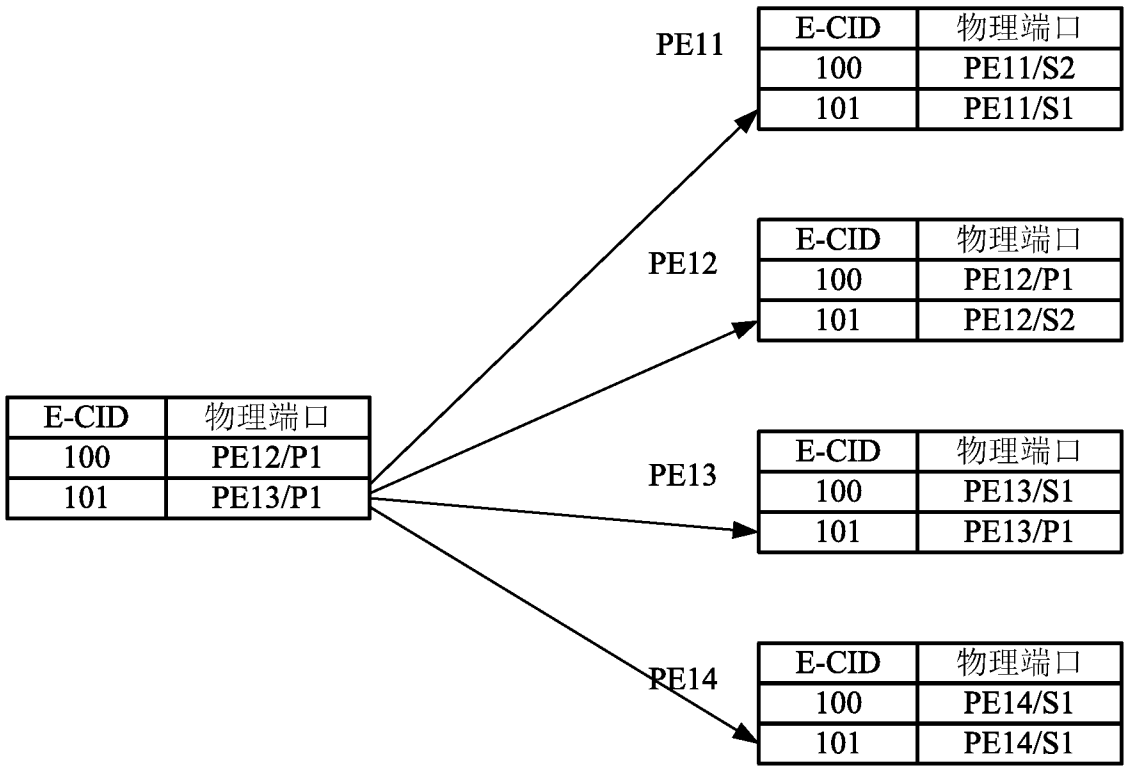

E-CID
物理端口
100
PE12/P1
101
PE13/P1
PE11
E-CID
物理端口
100
PE11/S2
101
PE11/S1
PE12
E-CID
物理端口
100
PE12/P1
101
PE12/S2
PE13
E-CID
物理端口
100
PE13/S1
101
PE13/P1
PE14
E-CID
物理端口
100
PE14/S1
101
PE14/S1

图 10

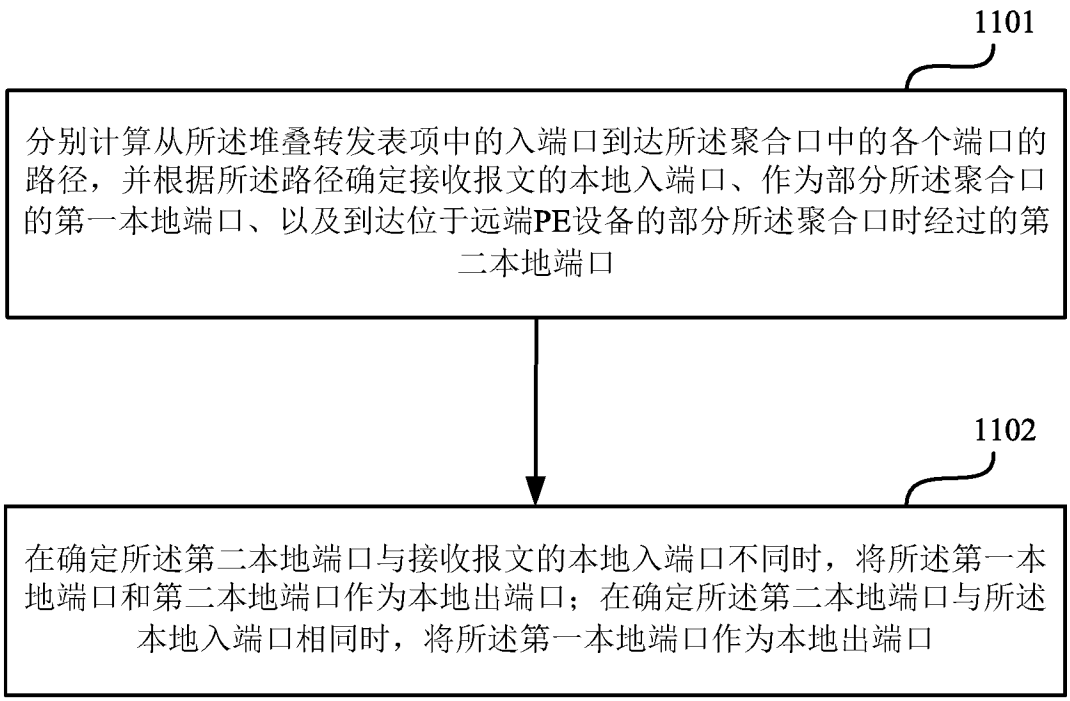

1101
分别计算从所述堆叠转发表项中的入端口到达所述聚合口中的各个端口的路径，并根据所述路径确定接收报文的本地入端口、作为部分所述聚合口的第一本地端口、以及到达位于远端PE设备的部分所述聚合口时经过的第二本地端口
1102
在确定所述第二本地端口与接收报文的本地入端口不同时，将所述第一本地端口和第二本地端口作为本地出端口；在确定所述第二本地端口与所述本地入端口相同时，将所述第一本地端口作为本地出端口

图 11

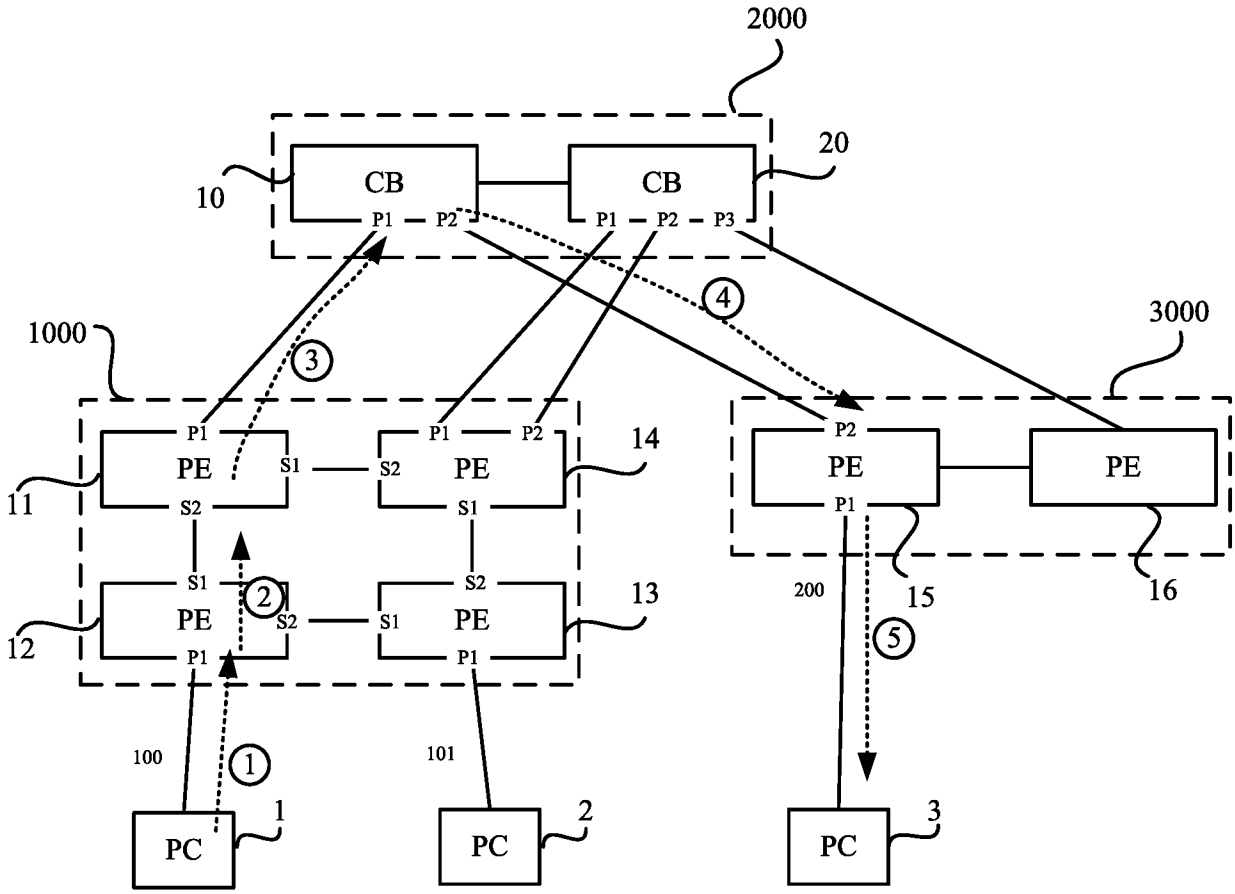

2000
20
10
CB
CB
P1
P2
P1
P2
P3
4
3000
1000
3
P1
P1
P2
P2
PE
S1
S2
PE
PE
14
11
S2
S1
P1
S1
S2
S2
S1
PE
2
PE
13
12
P1
P1
100
101
200
15
16
5
1
PC
PC
PC
1
2
3

图 12

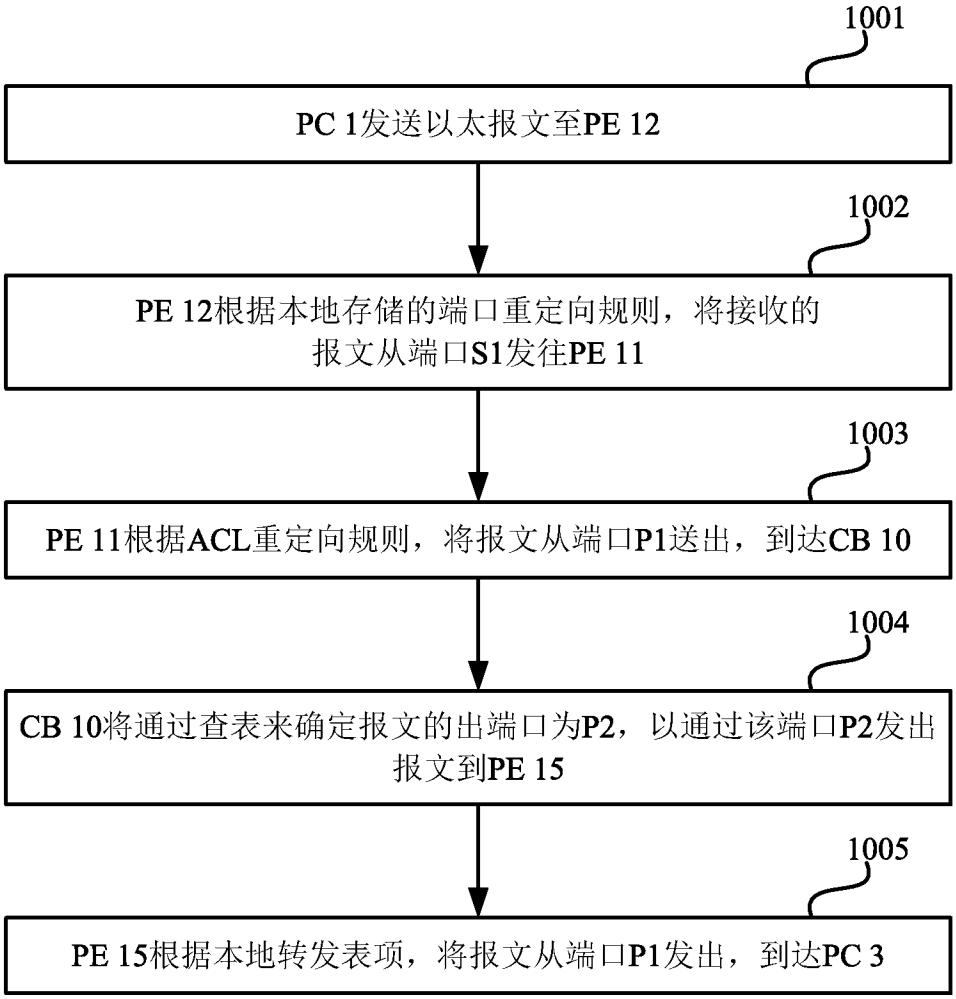

1001
PC 1发送以太报文至PE 12
1002
PE 12根据本地存储的端口重定向规则，将接收的报文从端口S1发往PE 11
1003
PE 11根据ACL重定向规则，将报文从端口P1送出，到达CB 10
1004
CB 10将通过查表来确定报文的出端口为P2，以通过该端口P2发出报文到PE 15
1005
PE 15根据本地转发表项，将报文从端口P1发出，到达PC 3

图 13

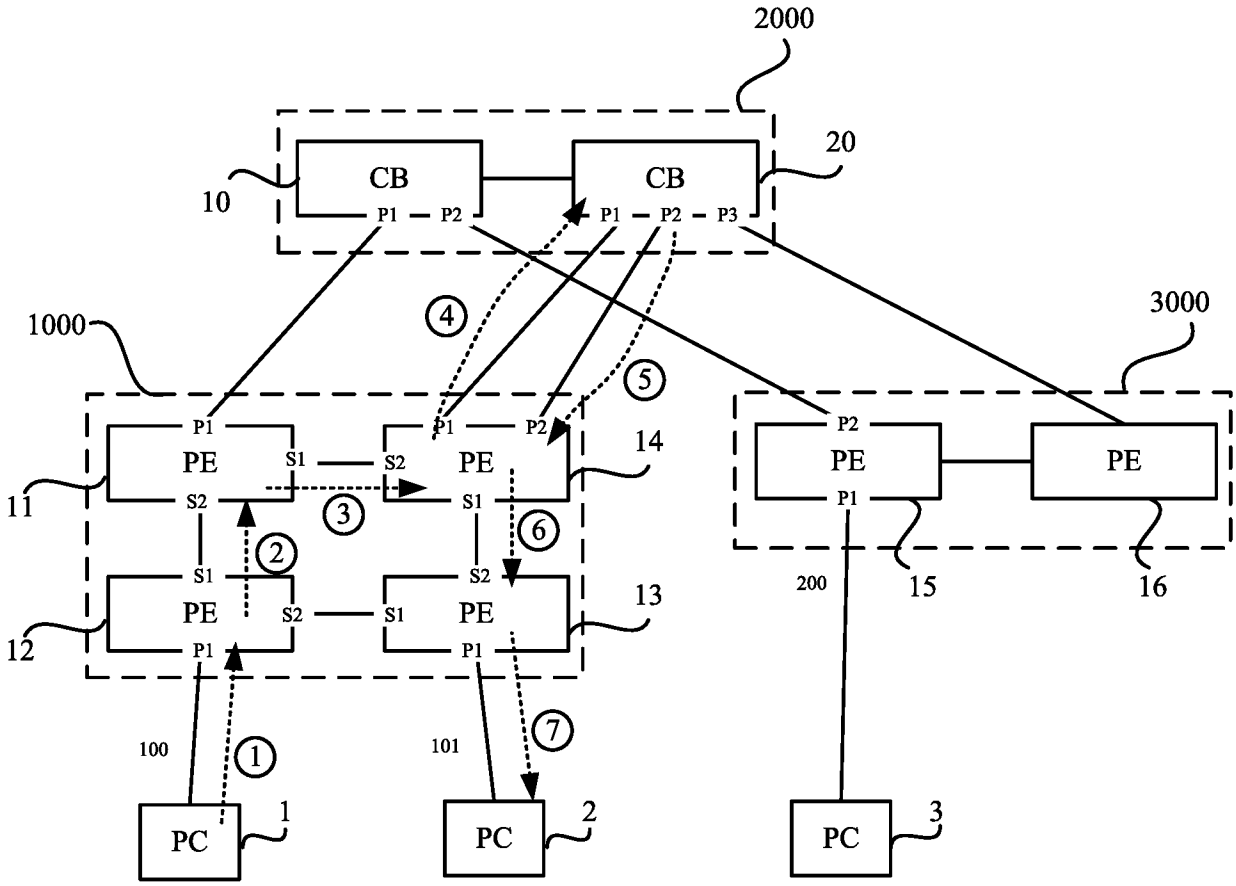

2000
CB
CB
10
20
P1
P2
P1
P2
P3
1000
3000
4
5
P1
P1
P2
P2
PE
S1
S2
PE
PE
14
11
S2
S1
3
6
S1
S2
2
PE
S2
S1
PE
13
12
P1
P1
200
15
16
100
101
7
1
PC
PC
PC
1
2
3

图 14

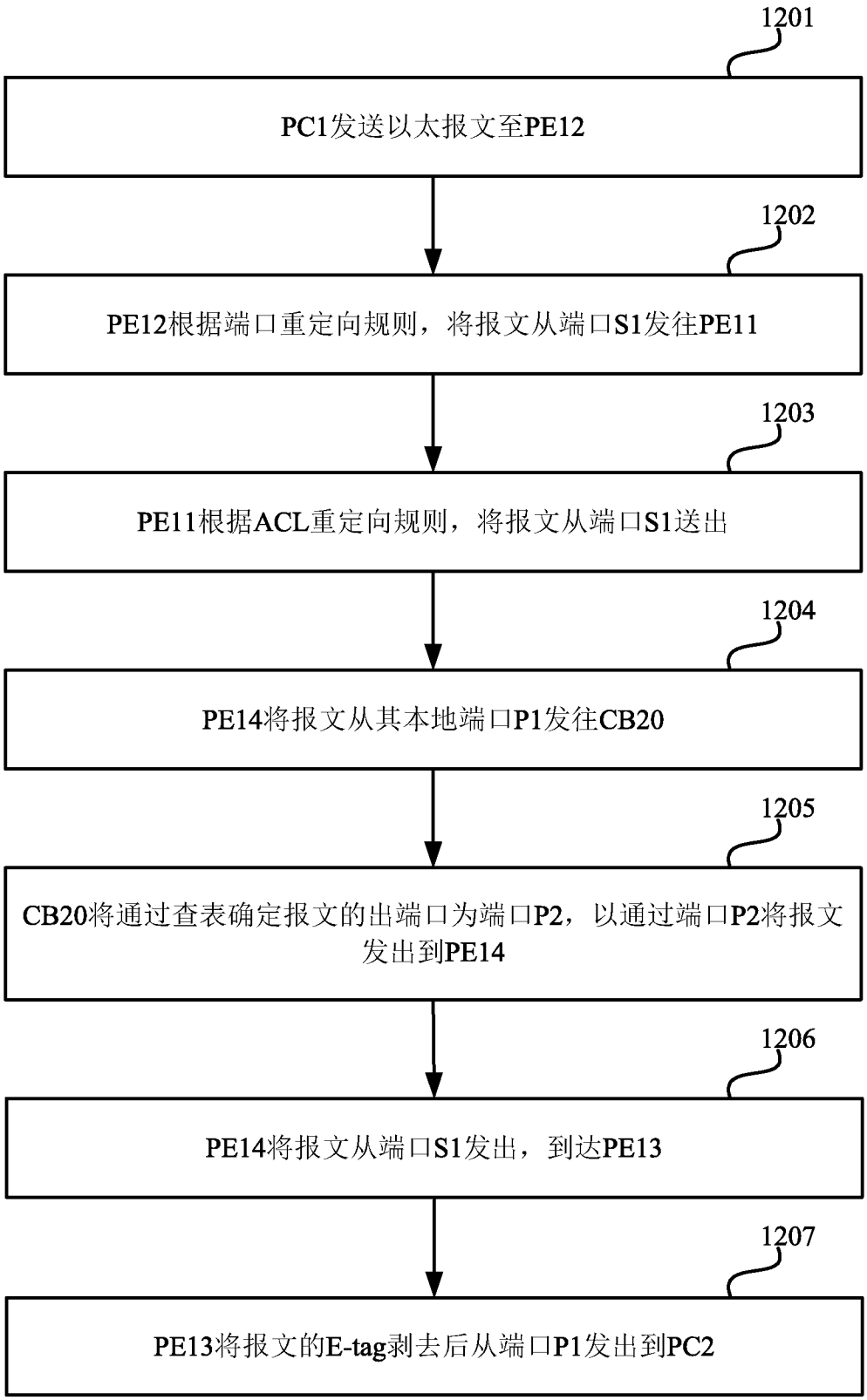

1201
PC1发送以太报文至PE12
1202
PE12根据端口重定向规则，将报文从端口S1发往PE11
1203
PE11根据ACL重定向规则，将报文从端口S1送出
1204
PE14将报文从其本地端口P1发往CB20
1205
CB20将通过查表确定报文的出端口为端口P2，以通过端口P2将报文发出到PE14
1206
PE14将报文从端口S1发出，到达PE13
1207
PE13将报文的E-tag剥去后从端口P1发出到PC2

图 15

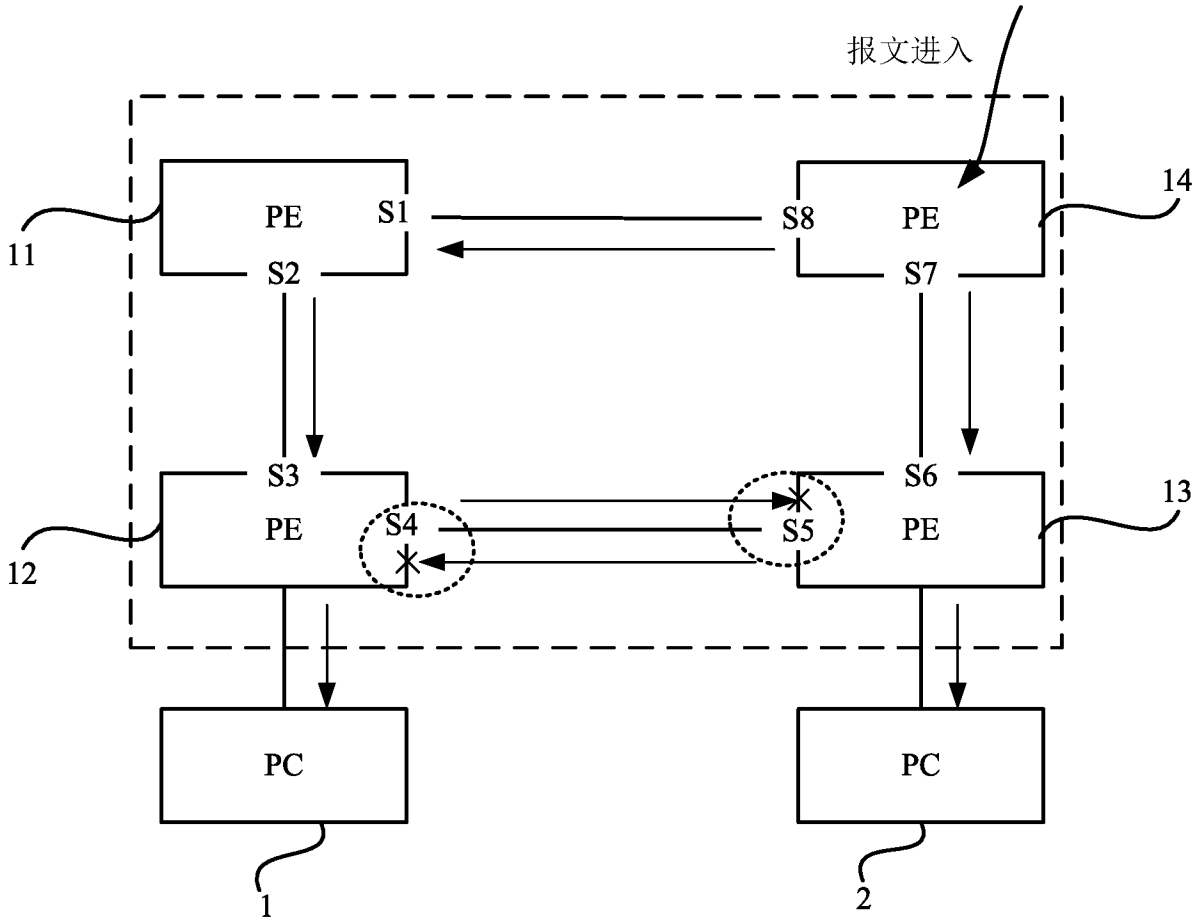

报文进入
PE
S1
S2
S8
PE
S7
11
14
S3
PE
S4
S5
S6
PE
12
13
PC
PC
1
2

图 16

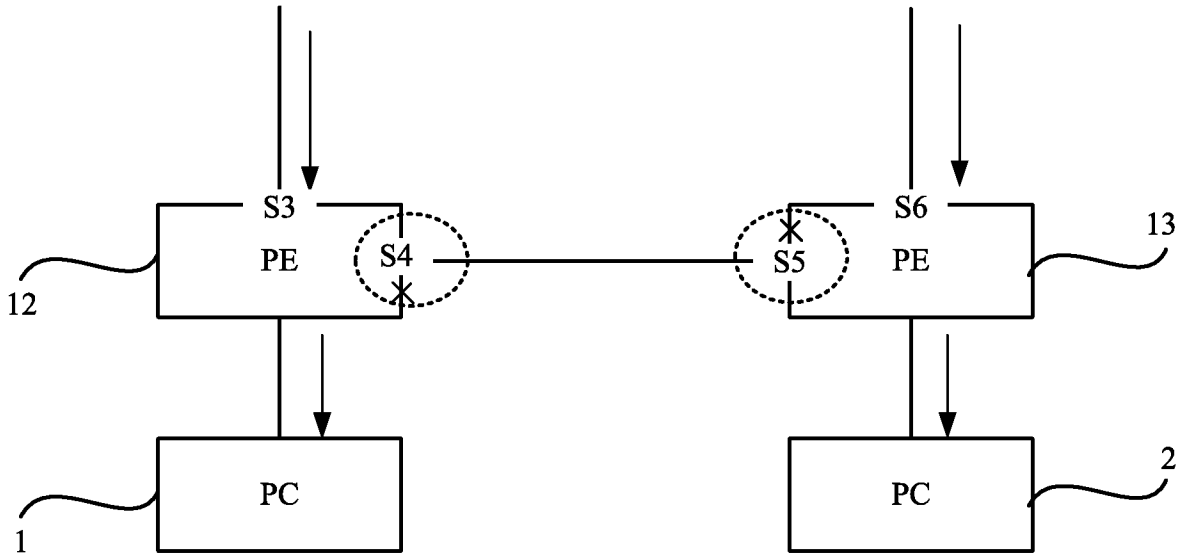

S3
PE
S4
S5
S6
PE
12
13
PC
PC
1
2

图 17

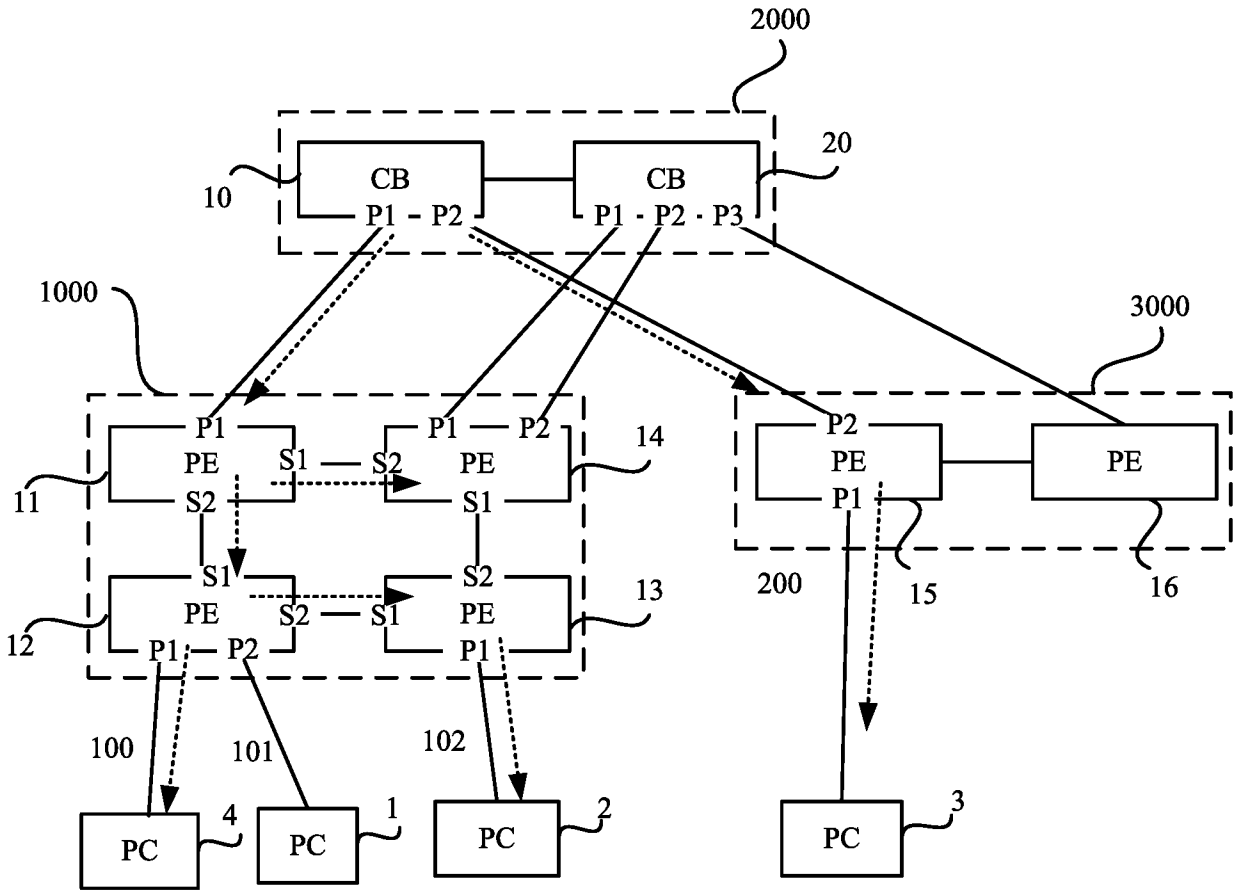


图 18

E-CID	物理端口
5000	S1/S2

PE11

E-CID	物理端口
5000	P1/S2

PE12

E-CID	物理端口
5000	P1

PE13

E-CID	物理端口
5000	

PE14

E-CID	物理端口
5000	P1

PE15

图 19

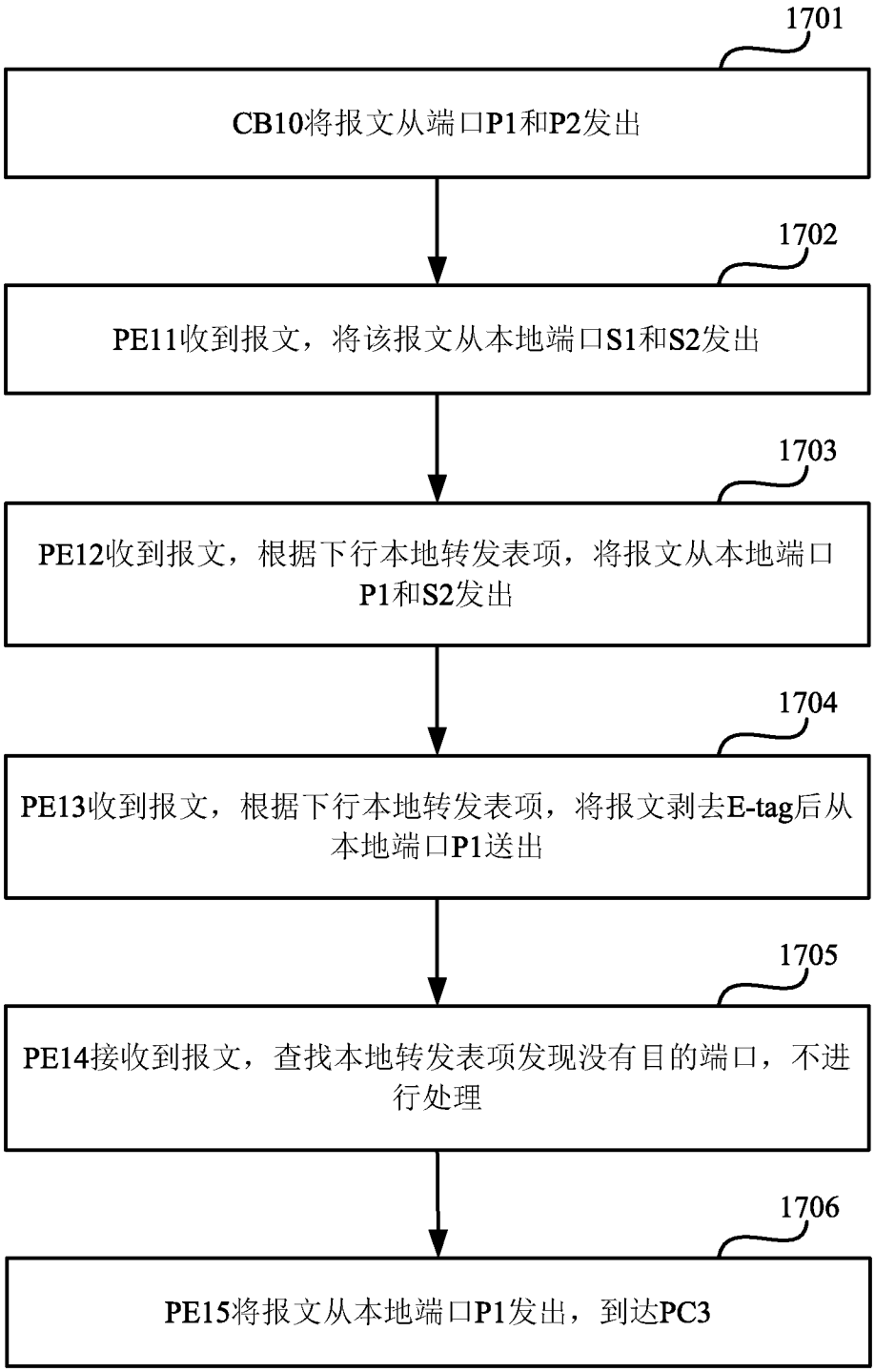

1701
CB10将报文从端口P1和P2发出
1702
PE11收到报文，将该报文从本地端口S1和S2发出
1703
PE12收到报文，根据下行本地转发表项，将报文从本地端口P1和S2发出
1704
PE13收到报文，根据下行本地转发表项，将报文剥去E-tag后从本地端口P1送出
1705
PE14接收到报文，查找本地转发表项发现没有目的端口，不进行处理
1706
PE15将报文从本地端口P1发出，到达PC3

图 20

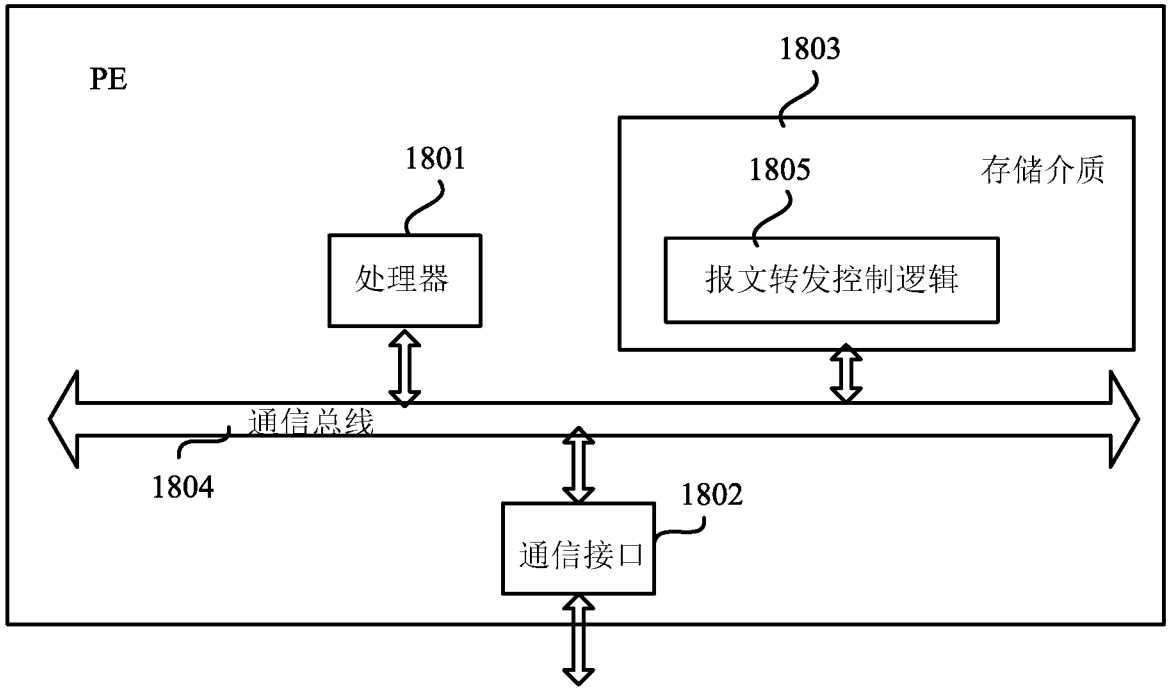

PE
1803
1801
1805
存储介质
处理器
报文转发控制逻辑
通信总线
1804
1802
通信接口

图 21

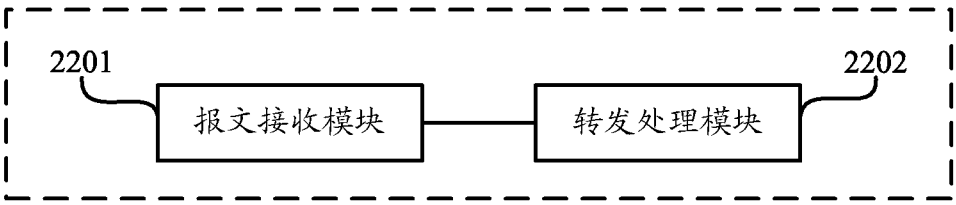

2201
2202
报文接收模块
转发处理模块

图 22

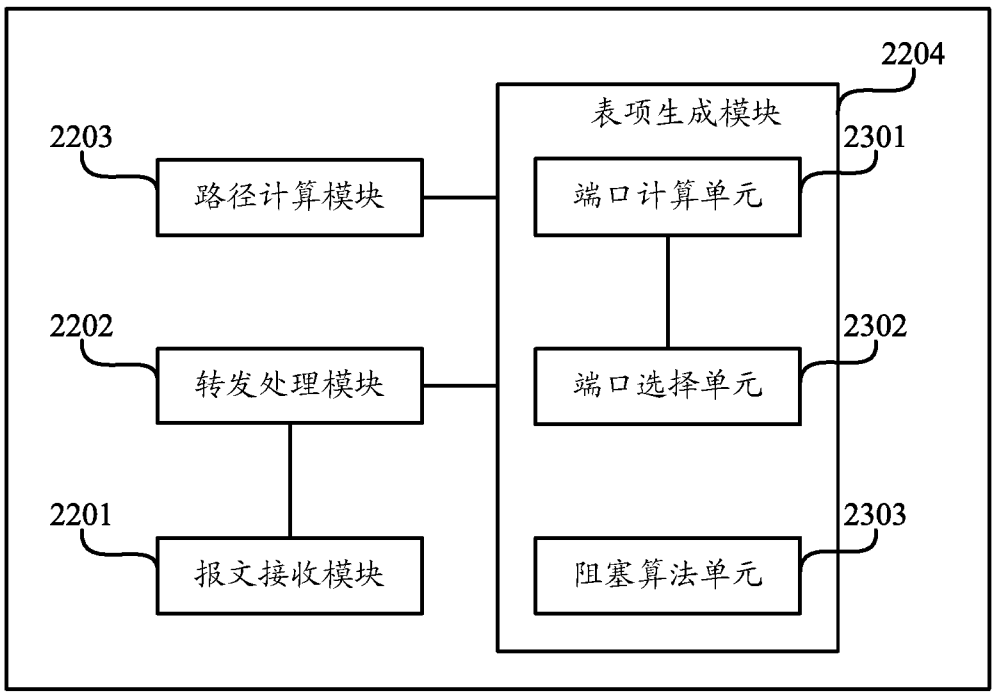

2204
表项生成模块
2203
2301
路径计算模块
端口计算单元
2202
2302
转发处理模块
端口选择单元
2201
2303
报文接收模块
阻塞算法单元

图 23

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2016/086989

A. CLASSIFICATION OF SUBJECT MATTER

H04L 12/721 (2013.01) i; H04L 12/741 (2013.01) i; H04L 12/861 (2013.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

H04L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNTXT, CNABS, CNKI: port extension, stack, local, forwarding table, port, PE

VEN: port extender, stack, local, forward, table, port, PE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 102904803 A (H3C TECHNOLOGIES CO., LIMITED), 30 January 2013 (30.01.2013), description, paragraphs 2-6 and 20-34	1-14
X	US 2014269710 A1 (DELL PRODUCTS LP.), 18 September 2014 (18.09.2014), description, paragraphs 47-52, and figure 2	1-14
X	CN 104639441 A (H3C TECHNOLOGIES CO., LIMITED), 20 May 2015 (20.05.2015), description, paragraphs 20-57	1-14
X	CN 103516609 A (H3C TECHNOLOGIES CO., LIMITED), 15 January 2014 (15.01.2014), description, paragraphs 33-54	1-14

☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
18 August 2016 (18.08.2016)	**29 August 2016 (29.08.2016)**
Name and mailing address of the ISA/CN: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451	Authorized officer **SU, Qin** Telephone No.: (86-10) **62089136**

Form PCT/ISA/210 (second sheet) (July 2009)

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2016/086989

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 102904803 A	30 January 2013	None	
US 2014269710 A1	18 September 2014	US 9294396 B2	22 March 2016
CN 104639441 A	20 May 2015	None	
CN 103516609 A	15 January 2014	None	

国际检索报告

国际申请号
PCT/CN2016/086989

A. 主题的分类

H04L 12/721(2013.01)i; H04L 12/741(2013.01)i; H04L 12/861(2013.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

H04L

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称，和使用的检索词（如使用）)

CNTXT, CNABS, CNKI：端口扩展，堆叠，本地，转发表，端口，PE; VEN: port extender, stack, local, forward, table, port, PE

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
X	CN 102904803 A (杭州华三通信技术有限公司) 2013年 1月 30日 (2013-01-30) 说明书第2-6，20-34段	1-14
X	US 2014269710 A1 (戴尔产品有限公司) 2014年 9月 18日 (2014-09-18) 说明书第47-52段，图2	1-14
X	CN 104639441 A (杭州华三通信技术有限公司) 2015年 5月 20日 (2015-05-20) 说明书第20-57段	1-14
X	CN 103516609 A (杭州华三通信技术有限公司) 2014年 1月 15日 (2014-01-15) 说明书第33-54段	1-14

☐ 其余文件在C栏的续页中列出。

☑ 见同族专利附件。

* 引用文件的具体类型：

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件，或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件（如具体说明的）

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布，与申请不相抵触，但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件，单独考虑该文件，认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件，当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时，要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期	国际检索报告邮寄日期
2016年 8月 18日	2016年 8月 29日
ISA/CN的名称和邮寄地址 中华人民共和国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451	受权官员 苏琴 电话号码 (86-10)62089136

表 PCT/ISA/210 (第2页) (2009年7月)

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2016/086989

检索报告引用的专利文件	公布日（年/月/日）	同族专利	公布日（年/月/日）
CN 102904803 A	2013年 1月 30日	无	
US 2014269710 A1	2014年 9月 18日	US 9294396 B2	2016年 3月 22日
CN 104639441 A	2015年 5月 20日	无	
CN 103516609 A	2014年 1月 15日	无	