

(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
2018 年 7 月 5 日 (05.07.2018)



(10) 国际公布号
WO 2018/121068 A1

(51) 国际专利分类号:

H04L 12/801 (2013.01)

(21) 国际申请号:

PCT/CN2017/109372

(22) 国际申请日:

2017 年 11 月 3 日 (03.11.2017)

(25) 申请语言:

中文

(26) 公布语言:

中文

(30) 优先权:

201611229376.5 2016年12月27日 (27.12.2016) CN

(71) 申请人: 华为技术有限公司 (HUAWEI TECHNOLOGIES CO., LTD.) [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

(72) 发明人: 袁峰 (YUAN, Feng); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。张弘 (ZHANG, Hong); 中国香港特别行政区九龙清水湾香港科技大学, Hong Kong 999077 (CN)。陈凯 (CHEN, Kai); 中

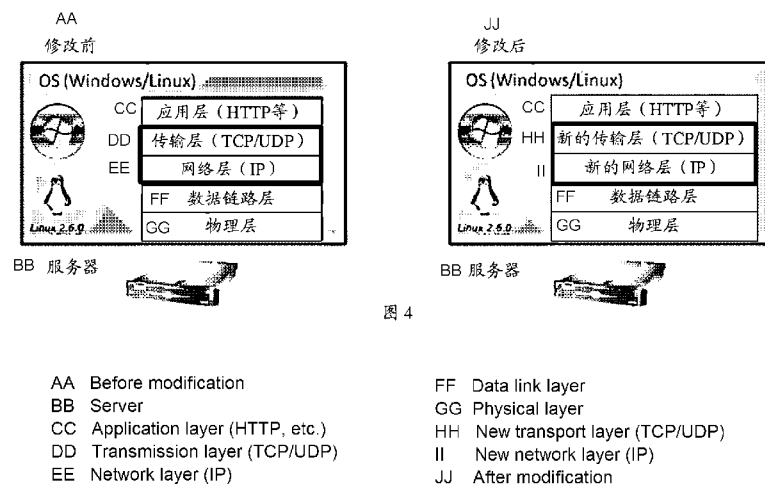
国香港特别行政区九龙清水湾香港科技大学, Hong Kong 999077 (CN)。

(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, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, 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,

(54) Title: METHOD AND DEVICE FOR DETERMINING TRANSMISSION PATH

(54) 发明名称: 确定传输路径的方法和装置



(57) Abstract: Provided in the embodiments of the present application are a method and device for determining a transmission path, the method comprising: determining occurrence of congestion on a current path corresponding to a flow to which a message to be transmitted belongs; determining a target path for the message to be transmitted according to a path congestion information table, and adding the information of the target path to the message to be transmitted so that the message to be transmitted is transmitted according to the target path, the congestion degree of the target path being less than that of the current path, each entry of the path congestion information table comprising a transmission path and congestion information corresponding to the transmission path, and the congestion information being used for indicating the congestion degree of the transmission path. The present application therefore effectively solves the problem of transmission path congestion.

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)。

根据细则4.17的声明:

- 关于申请人有权要求在先申请的优先权(细则4.17(iii))

本国际公布:

- 包括国际检索报告(条约第21条(3))。

(57) 摘要: 本申请实施例提供了一种确定传输路径的方法和装置, 该方法包括: 确定待传输报文所属的流对应的当前路径上发生拥塞; 根据路径拥塞信息表, 为所述待传输报文确定目标路径, 并将所述目标路径的信息添加在所述待传输报文中, 以使得所述待传输报文根据所述目标路径进行传输, 其中, 所述目标路径的拥塞程度比所述当前路径的拥塞程度小, 所述路径拥塞信息表的每个表项包括一条传输路径以及所述传输路径对应的拥塞信息, 所述拥塞信息用于表示所述传输路径的拥塞程度。从而有效地解决传输路径拥塞的问题。

确定传输路径的方法和装置

本申请要求于 2016 年 12 月 27 日提交中国专利局、申请号为 201611229376.5、发明名称为“确定传输路径的方法和装置”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请实施例涉及通信领域，并且更具体地，涉及一种确定传输路径的方法和装置。

背景技术

互联网数据正在以爆炸性的方式增长，中国的新浪微博注册用户数量已经破 3 亿，腾讯的即时通讯工具活跃用户达到 7.1 亿，Facebook 全球用户数量则正逼近 10 亿，仅次于中国和印度的人口数字。仅社交网络这一项所产生的数据就已经非常惊人。据国际数据公司 IDC 发布的报告《Digital Universe Study 2011》显示，全球信息总量每过两年就会增长一倍。仅在 2011 年，全球被创建和被复制的数据总量为 1.8ZB（即 1.8 万亿 GB）。相较 2010 年同期上涨超过 1ZB，到 2020 年这一数值将增长到 35ZB。大数据的出现正迫使企业不断提升自身以数据中心为平台的数据处理能力。

Google、Microsoft 及 Facebook 等巨头在数据中心的网络领域走在业界最前沿，这些业界引领者的多年研究及实践表明，基于克劳斯（英文：Clos）架构的数据中心网络扩展性佳，等价路径多，因此正得到越来越广泛的部署。

Clos 数据中心网络架构尽管有等价路径多的优势，但是在传统的负载均衡（英文：load balancing，简称“LB”）机制下却容易出现严重的路径拥塞。图 1 是一个典型的数据中心 3 层 Clos 网络构架的示意图，每层设备例如可以为交换机、路由器等网络设备，其中，底层设备为架顶设备，或称为叶子设备（英文：leaf device，简称“Leaf”）；中间层设备为汇集设备（英文：aggregation device，简称“Agg”）；顶层设备为核心设备（英文：core device，简称 Core）。如图 1 所示，该 Clos 网络中有四条流（Flow）即 Flow A、Flow B、Flow C 和 Flow D，分别从不同的源设备转发到不同的目的设备。尽管网络中存在不止四条路径可以完成这四条流的转发，但是由于传统的负载均衡机制的缺陷，会导致转发路径出现重叠。如图 1 所示，Flow A 与 Flow B 在中间层设备 Agg 0 处发生了本地冲突（英文：local collision），即，Flow A 与 Flow B 本可以通过走不同路径（例如分别通过路径 Agg 0→Core 0 和路径 Agg 0→Core 1）进行传输，但由于某些控制机制的缺陷等导致 Flow A 与 Flow B 都使用路径 Agg 0→Core 0 进行传输，假设 Flow A 与 Flow B 的流速均为 6G，Agg 0 最大能满足的流速为 10G，当 Flow A 与 Flow B 在 Agg 0 处交汇时，由于 $6G \times 2 = 12G$ 超过了 10G，导致了 Flow A 与 Flow B 在 Agg 0 处发生冲突。图 1 中的 Flow C 与 Flow D 在 Core 2 处发生下游冲突（英文：downstream collision），即，由于 Core 2 到下游目的设备之间只存在一条路径（即 Core 2→Agg 3），从而导致 Flow C 与 Flow D 必然会在 Core 2 处发生冲突。因此，如何通过优化负载均衡机制避免网络中出现局部拥塞导致丢包已成为当下业内的焦点。

发明内容

本申请实施例提供一种确定传输路径的方法和装置，能够有效地解决传输路径 congestion 的问题。

第一方面，提供了一种确定传输路径的方法，该方法包括：

5 确定待传输报文所属的流对应的当前路径上发生 congestion；

根据路径 congestion 信息表，为所述待传输报文确定目标路径，并将所述目标路径的信息添加在所述待传输报文中，以使得所述待传输报文根据所述目标路径进行传输，其中，所述目标路径的 congestion 程度比所述当前路径的 congestion 程度小，所述路径 congestion 信息表的每个表项包括一条传输路径以及所述传输路径对应的 congestion 信息，所述 congestion 信息用于表示所述传
10 输路径的 congestion 程度；

根据所述目标路径，发送所述待传输报文。

因此，本申请实施例中，通过确定传输路径发生 congestion，并在发生路径 congestion 时根据路径 congestion 信息表为待传输报文确定新的路径，能够有效地解决传输路径 congestion 的问题。

15 由于本申请实施例只有在当前路径确实发生 congestion、确实有必要进行路径切换时，才会进行传输路径的切换，避免了 congestion 不匹配等问题，且切换后的路径的 congestion 程度小于该当前路径，避免了长尾效应。

另外，本申请实施例中，交换机、路由器等网络设备上可以不做任何修改，整个路径的决策可以由其他装置例如主机来执行，congestion 信息表可以保存在主机中，从而解决了表项规格受限的问题，能够适用于 3 级 Clos 甚至 N 级 Clos 架构中。其中，N 大于 3。

20 可选地，在第一方面的一种实现方式中，所述确定待传输报文所属的流对应的当前路径上发生 congestion，包括：确定所述流在所述当前路径上传输时被标记显式 congestion 通知 ECN，或者所述当前路径对应的往返时间 RTT 大于时间阈值，或者所述当前路径发生故障

可选地，在第一方面的一种实现方式中，所述 congestion 信息包括以下信息中的至少一种：所述传输路径的平均 ECN 次数、所述传输路径的 RTT、用于表示所述传输路径是否故障
25 的标识、以及所述传输路径上同时存在的流的个数。

相比于现有技术中仅仅依靠一段时间内出端口的利用率来表示路径的 congestion 程度，本申请实施例通过包括平均 ECN 次数、往返时间 RTT、用于表示路径是否故障的标识、以及路径上同时存在的流的个数等参数的 congestion 信息来表征传输路径的 congestion 程度，能够更加准确的对路径的 congestion 程度进行评价，从而更有效地指导路径切换。

30 可选地，在第一方面的一种实现方式中，所述根据路径 congestion 信息表，为所述待传输报文确定目标路径之前，所述方法还包括：确定所述待传输报文所属的流的流速是否大于预设阈值；若所述流的流速大于所述预设阈值，执行所述根据所述路径 congestion 信息表，为所述待传输报文确定目标路径的步骤。

35 由于只在流速超过预设的阈值时才对该待传输报文进行路径切换操作，从而避免了不必要的切换，例如对较小的流进行切片就完全没有必要。

可选地，在第一方面的一种实现方式中，所述根据路径 congestion 信息表，为所述待传输报文确定目标路径，包括：在所述待传输报文和所述当前路径上待传输的其他报文中，若所述待传输报文以预设概率被选择到，则根据所述路径 congestion 信息表，为所述待传输报文确定所述目标路径，以使得所述待传输报文根据所述目标路径进行传输。

这样，当前路径发生拥塞时，在该路径上将要传输的至少一个流中，不是所有流一起进行路径切换，而是根据预设概率选择一部分流进行路径切换，从而降低了切换后再次发生拥塞的概率，提高了报文传输的成功率。

可选地，在第一方面的一种实现方式中，所述方法还包括：确定所述目标路径的 RTT 与所述当前路径的 RTT 之间的时间差；在经过所述时间差的时长后，发送所述待传输报文。

为了避免乱序的发生，主机确定待传输报文对应的当前路径发生拥塞而为该报文所属的流确定了新的路径即目标路径后，主机可以确定该目标路径的 RTT 与该当前路径的 RTT 之间的时间差，从而主机在等于该时间差的时长后，再发送该待传输报文。这样可以避免路径切换导致的报文乱序。例如对报文 1 所述的流进行路径切换没有成功，则需要等待固定时长例如 100ms 后，再对报文 1 所述的流执行路径切换操作。

可选地，在第一方面的一种实现方式中，所述在根据路径拥塞信息表，为所述待传输报文确定目标路径之前，所述方法还包括：在至少一条传输路径中的每条传输路径上发送探测报文；接收针对所述探测报文的响应报文；根据所述响应报文，确定至少一条传输路径中每条传输路径对应的所述拥塞信息；根据所述至少一条传输路径中每条传输路径对应的所述拥塞信息，生成所述路径拥塞信息表。

进一步地，主机还可以周期性地或者实时地，对该路径拥塞信息表进行更新，对路径拥塞信息表的更新例如可以通过有源探测器（英文：active probe）发起定期主动探测，记录平均 ECN 次数、RTT 等信息并保存到路径拥塞信息表中，以主机为例，保存目的地可以为主机本地例如主机内存或硬盘中。同时还可以依据平均 ECN 次数、RTT 等信息为路径划分等级，以对路径切换进行指导。

可选地，在第一方面的一种实现方式中，所述方法还包括：若获取的一条传输路径的拥塞信息中包括表示所述传输路径发生故障的标识，将所述传输路径对应的表项从所述路径拥塞信息表中删除。

也就是说，主机在获取到哪条路径故障后，可以及时将该路径以及该路径的相关信息从该路径拥塞信息表中删除，路径拥塞信息表可以实时地进行该更新操作。在后续的路径选择过程中，就可以在没有故障的路径中进行路径选择。

第二方面，提供了一种确定传输路径的装置，该装置可以用于执行前述第一方面及各种实现方式中所述的确定传输路径的方法中的各个过程。该装置包括确定单元和发送单元，其中，确定单元，用于确定待传输报文所属的流对应的当前路径上发生拥塞；所述确定单元还用于，根据路径拥塞信息表，为所述待传输报文确定目标路径，并将所述目标路径的信息添加在所述待传输报文中，以使得所述待传输报文通过所述目标路径进行传输，其中，所述目标路径的拥塞程度比所述当前路径的拥塞程度小，所述路径拥塞信息表的每个表项包括一条传输路径以及所述传输路径对应的拥塞信息，所述拥塞信息用于表示所述传输路径的拥塞程度；

发送单元，用于根据所述确定单元确定的目标路径，发送所述待传输报文。

第三方面，提供了一种确定传输路径的装置，该装置包括收发器、处理器和存储器。所述存储器存储了程序，所述处理器执行所述程序，以用于执行前述第一方面及各种实现方式中所述的确定传输路径的方法中的各个过程。其中，所述处理器具体用于：确定

待传输报文所属的流对应的当前路径上发生拥塞；根据路径拥塞信息表，为所述待传输报文确定目标路径，并将所述目标路径的信息添加在所述待传输报文中，以使得所述待传输报文根据所述目标路径进行传输，其中，所述目标路径的拥塞程度比所述当前路径的拥塞程度小，所述路径拥塞信息表的每个表项包括一条传输路径以及所述传输路径对应的拥塞信息，所述拥塞信息用于表示所述传输路径的拥塞程度；所述收发器用于，根据所述目标路径，发送所述待传输报文。

第四方面，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程序，所述程序使得上述装置执行上述第一方面及其各种实现方式中的任一种确定传输路径的方法。

基于上述技术方案，本申请实施例通过确定传输路径是否发生拥塞，并在发生路径拥塞时根据路径拥塞信息表为待传输报文确定新的路径，能够有效地解决传输路径拥塞的问题。

由于本申请实施例是在当前路径确实发生拥塞、确实有必要进行路径切换时，才会进行传输路径的切换，避免了拥塞不匹配等问题，且由于路径切换的依据是当前路径上是否发生拥塞，而不是依靠前后报文到达的时间差，因而避免了长尾效应。

并且，本申请实施例中，整个路径的决策并不需要由交换机、路由器等网络设备来执行，拥塞信息表等信息可以保存在主机中，从而解决了网络设备的表项规格有限导致的不能支持大规模网络的问题，能够适用于 3 级 Clos 甚至 N 级 Clos 架构中。

另外，本申请实施例还引入了新的评价指标用来评价路径的拥塞程度，例如平均 ECN 次数、往返时间 RTT、用于表示路径是否故障的标识、以及路径上同时同时存在的流的个数等参数，从而能够更加准确的对路径的拥塞程度进行评价，更有效地指导路径切换。

附图说明

图 1 是现有技术中报文传输的示意性架构图；

图 2 是现有技术中能够实现负载均衡的报文传输的路径示意图。

图 3 是适用本申请实施例提供的确定传输路径的方法的通信系统的示意性架构图。

图 4 是本申请实施例的修改前后的 TCP/IP 协议栈的示意图。

图 5 是本申请实施例的确定传输路径的方法的示意性流程图。

图 6 是本申请实施例的路径切换的示意图。

图 7 是根据本申请实施例的确定传输路径的装置的示意性框图。

图 8 是根据本申请实施例的确定传输路径的装置的结构示意图。

具体实施方式

下面将结合附图，对本申请实施例中的技术方案进行描述。

应理解，本申请实施例中以 Clos 网络架构为例进行描述，但是本申请不限于此，所有等价路径存在的场景下，都可以实现本申请实施例提出的确定传输路径的方法。

现有的针对 2 级 Clos (2 Stage Clos) 提出的负载均衡方案中，每个底层设备，又称为叶子设备 (英文: leaf device, 简称 “Leaf”), 测量所述 Leaf 与整个网络中其他所有 Leaf 之间路径的拥塞程度，并记录在拥塞信息表 (Congestion-To-Leaf) 内，当

收到新的数据流（后面也简称为“流”）时，会选取当前拥塞程度最低的路径转发该流。

这里，2级Clos指有2层架构的Clos网络，其最底层设备一般称为Leaf，最高层设备称为骨干设备（英文：spine device，简称Spine）。同样，3级Clos指具有3层架构的Clos网络，其最底层设备为Leaf，中间层设备为汇聚设备（英文：aggregation device，简称Aggregation或Agg），最高层设备称为Spine（或Core）。N级Clos指有N层架构的Clos。

换句话说，Leaf为2级Clos架构、3级Clos架构、甚至N级Clos架构下最底层的设备，一般为交换机或路由器；Aggregation为3级Clos架构下的中间层设备；Spine为2级Clos架构、3级Clos架构、甚至N级Clos架构下的最高层的设备，一般为交换机或路由器。

流包括可以用特定的五元组（五个关键字段）唯一标识的一组传输控制协议（英文：transmission control protocol，简称“TCP”）/互联网协议（英文：internet protocol，简称“IP”）报文。例如，某用户从源IP地址为202.100.1.2的主机向目的IP地址为100.1.1.2的网页主机发起超文本传输协议（英文：hypertext transport protocol，简称“HTTP”）访问，这是一条TCP流，该流的另外两个五元组字段是源TCP端口20000，目的TCP端口8080，则202.100.1.2+100.1.1.2+TCP+20000+8080可用于惟一标识网络中传输的这条流。

该流中包括多个报文，若前后两个报文到达的时间差大于一个配置的值（例如500微秒）则可将后一个报文作为一个新短流（英文：flowlet）的第一个报文；若前面两个报文达到的时间差小于该配置的值，则这两个报文属于同一个flowlet。

源Leaf在进行传输报文时，首先检测flowlet，如果前后两个报文到达的时间差大于配置的值（例如500微秒），则将后一个报文作为一个新的flowlet的第一个报文，否则认为该两个报文属于同一个flowlet。若当前报文属于新的flowlet，则按下面流程查找出端口：查找路径拥塞程度表（Congestion-To-Leaf表）；根据报文的目的IP地址找到路径拥塞程度表中的目的Leaf；根据目的Leaf找到可转发所述报文的所有出端口；比较可转发所述报文的的所有出端口的拥塞程度，选择其中路径拥塞程度最低的端口转发所述报文；如果所述路径拥塞程度表没有存储所述目的IP地址对应的目的Leaf，或初始时尚未建立起到目的Leaf的拥塞程度信息，则从可供转发的多个出端口中随机选择一个；若当前报文属于上一个flowlet，则查找flowlet表，在表中找到该flowlet对应的出端口转发。

路径拥塞程度表的建立方式分为前向探测和后向反馈两步。前向探测过程主要为：（1）探测报文从源Leaf上转发出去时，源Leaf为该探测报文添加Leaf标识、源Leaf中用于转发该探测报文的出端口的标识、以及表示该出端口对应的路径的拥塞程度的标识等；（2）探测报文从源Leaf转发到Spine时，Spine确定用于继续转发该探测报文的出端口，如果发现该出端口对应的路径的拥塞程度，比从源Leaf接收的探测报文中显示的拥塞程度更高，则Spine更新探测报文里用于记录路径拥塞程度的标识的字段；（3）探测报文到达目的Leaf时，目的Leaf记录探测报文中显示的拥塞程度；（4）目的Leaf转发该探测报文给主机。

后向反馈过程主要为：主机返回针对该探测报文的TCP响应报文给源Leaf；目的

Leaf 查找该 TCP 响应对应的源 Leaf，将之前前向探测过程里记录的拥塞信息写到报文内容里，并执行前述类似转发过程；源 Leaf 得到该响应报文以后，提取其中携带的路径拥塞信息；源 Leaf 用刚刚得到的路径拥塞信息指导下一次转发。

可以看出，现有技术中实现负载均衡的报文传输过程中，主要思想是将一条流（英文：flow）切成多个 flowlet，不同的 flowlet 通过查表走不同的传输路径，从而实现精细化地负载均衡。如果以图 1 为例，假设 Flow A 为带宽为 10KB 的网页浏览应用流速，Flow B 为带宽 10MB 的视频直播流，如果只是按照 flow 来简单调度，例如采用传统的等价多路径（英文：equal-cost multi-path，简称“ECMP”）哈希机制，那么这两条流分别对应两条传输路径，由于其带宽是不同的，因此两条路径的带宽比为 1:1000，这样造成了两条路径上的负载不均衡。当能够通过上述现有技术的方式实现负载均衡时，则 Flow A 和 Flow B 可以散列至两条甚至多条传输路径上，达到负载均衡的效果，从而避免了传输路径的拥塞。

但是，上面现有技术中用于实现负载均衡的报文传输的方法中，存在很多问题，这些问题直接影响报文传输的过程，在某些情况下并不能有效地避免报文传输过程中的路径拥塞问题。

首先，是表项规格的问题。上面描述的方法仅适用于 Leaf-Spine 架构的 2-Stage Clos，原因是拥塞信息表都是记录在 Leaf 上的，每个 Leaf 上的拥塞信息表记录该 Leaf 到所有其他 Leaf 的全部路径，因此表项开销非常大。对于 2-Stage Clos，该表项规格是 $O(K^2)$ ；但在 3-Stage Clos 时，该表项规格是 $O(K^4)$ ！，明显超出了网络设备的承受能力。

其次，是长尾效应。因为 flowlet 的切分是依赖于前后两个报文的时间差大于配置的值，但当某条流带宽较大如千兆比特每秒（Gbps）时，报文的前后时间差非常小，在纳秒（ns）级别，无法切分开。

另外，是拥塞不匹配（英文：Congestion Mismatch）的问题。传统的 ECMP 中一条流严格对应一条路径，而上述方法中是将一个流划分成多个 flowlet，将不同的 flowlet 散列到不同路径，因此一条流同时在两条或多条路径上传输。而在两条拥塞程度不等的路径上进行 flowlet 的切换时，无论当前路径是否真正发生拥塞，都会进行 flowlet 的切换，不仅增加了不必要的切换，甚至还可能导致报文传输的失败。

例如图 2 所示的现有技术中能够实现负载均衡的报文传输的示意图。其中 Flow A 为一个流，flowlet 1 和 flowlet 2 分别为该 Flow A 中的两个 flowlet，L0 和 L1 分别为两个 Leaf，S0 和 S1 分别为两个 Spine。如图 2 所示，Flow A 从 Leaf 0 到 Leaf 1 交换机可以走两条路径，分别是 $L0 \rightarrow S0 \rightarrow L1$ 与 $L0 \rightarrow S1 \rightarrow L1$ 。

Flow A 中的 flowlet 1 首先到达 L0，假设 flowlet 1 的速率为 1G，此时路径 $L0 \rightarrow S0 \rightarrow L1$ 与路径 $L0 \rightarrow S1 \rightarrow L1$ 的拥塞程度都是 0，假设 L0 选择了 $L0 \rightarrow S1 \rightarrow L1$ 路径转发 flowlet 1。

假设路径 $L0 \rightarrow S1 \rightarrow L1$ 的能够支持的最大流速为 10G，L0 发现 $L0 \rightarrow S1 \rightarrow L1$ 这条路径没有发生拥塞，于是增加流速率发送 flowlet 1 之后的 flowlet 2，这时，路径 $L0 \rightarrow S0 \rightarrow L1$ 的路径拥塞程度为 0，而路径 $L0 \rightarrow S1 \rightarrow L1$ 由于转发 flowlet 1，其拥塞程度变为 $1G/10G=10\%$ 。L0 比较后发现路径 $L0 \rightarrow S1 \rightarrow L1$ 的拥塞程度高于路径 $L0 \rightarrow S0 \rightarrow L1$ ，因此，

L0 控制 flowlet 2 通过拥塞程度为 0 的路径 $L0 \rightarrow S0 \rightarrow L1$ 转发出去。

然而，假设 $L0 \rightarrow S0 \rightarrow L1$ 路径的能够支持的最大流速为 5G，而 flowlet 2 的流速为 10G，由于 flowlet 2 的流速超过 5G，从而路径 $L0 \rightarrow S0 \rightarrow L1$ 发生拥塞，导致 flowlet 2 被丢弃。

可以看出，现行协议栈机制的缺陷可能导致拥塞不匹配的问题，即 Flow A 在两条拥塞程度不等的路径上进行的切换操作并不能有效地指导 Flow A 进行传输，用于指导 Flow A 进行路径切换的仅仅是不同路径之间拥塞程度的比较，而不考虑具体路径对 flowlet 2 来说是否真正会发生拥塞。

本申请实施例中，只有在当前路径确实发生拥塞、确实有必要进行路径切换时，才会进行传输路径的切换，解决了拥塞不匹配的问题，且由于路径切换的依据是当前路径上是否发生拥塞，而不是依靠前后报文到达的时间差，因而避免了长尾效应。

本申请实施例通过将拥塞信息表保存在主机中，并且由主机执行路径选择策略，解决了网络设备的表项规格不够的问题，能够适用于 3 级 Clos 甚至 N 级 Clos 架构中。

进一步地，本申请实施例还引入了新的指标用来评价路径的拥塞程度，能够更加准确的对路径的拥塞程度进行评价，从而更有效地指导路径切换。

图 3 是适用本申请实施例提供的确定传输路径的方法的通信系统的示意性架构图。在 Clos 架构中，每个 Leaf 都与所有的 Spine 之间存在传输路径。图 3 中示出了主机 (Host) 10、Host 20、Leaf 30 (用 L 30 表示)、Leaf 60 (用 L 60 表示)、Spine 50 (用 S 50 表示) 和 Spine 40 (用 S 40 表示)。这里的 Leaf 例如为 Leaf 交换机，Leaf 交换机底下连接有主机。在图 3 所示的通信系统中，从 Host 10 到 Host 20 之间存在两条传输路径，即 $Host\ 10 \rightarrow L\ 30 \rightarrow S\ 40 \rightarrow L\ 60 \rightarrow Host\ 20$ 和 $Host\ 10 \rightarrow L\ 30 \rightarrow S\ 50 \rightarrow L\ 60 \rightarrow Host\ 20$ 这两条路径。Host 10 发送报文后，经 L 30 转发到 S 40，若 S 40 到 L 60 的路径发生了拥塞，则 S 40 会给报文标记显式拥塞通知 (英文: explicit congestion notification, 简称“ECN”)，表示该报文所属的流在本设备上经历了拥塞。报文经过 L 60 转发给 Host 20 后，Host 20 看到带有 CE 标记的报文后，会返回显式拥塞通知响应 (英文: explicit congestion notification echo, 简称“ECN-Echo”) 给 Host 10。Host 10 收到 ECN-Echo 后，即感知到该传输路径上发生了拥塞，会降低自己的发送速率，并且发出拥塞窗口降速 (英文: congestion window reduced, 简称“CWR”) 报文告知目的端 Host 20，以告知 Host 20 自己已经执行了降速操作。

应理解，图 3 中仅示出了包括 2 个 Leaf 和 2 个 Spine 的 2 级 Clos 架构，但本申请实施例的方法同样适用于 3 级 Clos 或者更高级别的 Clos 架构中。本申请实施例的确定传输路径的方法也可以称为传输层感知负载均衡 (英文: transport layer aware load balancing, 简称“TLB”)。

还应理解，所有能够通过本申请实施例所述的方法确定传输路径的装置，均应落入本申请的保护范围。下面仅以主机为例进行本申请实施例的描述。

本申请实施例中，在传输路径发生拥塞时可以由主机根据路径拥塞信息表来进行路径的重新选择，也就是说，本申请实施例的确定传输路径的方法可以在主机的操作系统中实现，而无需对网络设备例如交换机、路由器等设备上作任何修改。所述主机的操作系统例如可以是 Windows/Linux 操作系统内核代码的网络处理部分等。本申请实施例的

方法可以应用于但不限于 TCP/IP 协议栈的套接口 (socket) 通信、应用程序编程接口 (英文: application programming interface, 简称 “API”) 接口等与网络交互部分形态的功能模块。本申请实施例的方法还可以应用于虚拟机系统, 例如 VMware、Xen 虚拟机、Openstack 等等云服务操作系统、以及虚拟化操作系统的虚拟机监视器 (Hypervisor) 功能模块部分。

图 4 所示为本申请实施例的修改前后的 TCP/IP 协议栈的示意图。图 4 中, 左图为修改前的 TCP/IP 协议栈的示意图, 右图为修改后的 TCP/IP 协议栈的示意图。其中, 对主机的操作系统中的传输层和网络层进行合适的修改后, 就可以实现本申请实施例的确定传输路径的方法。

由于本申请中路径的选择由主机来执行, 拥塞信息表保存在主机中, 可以不受网络设备的表项规格的限制, 从而能够灵活地适用于 3 级 Clos 架构甚至 N 级 Clos 架构中。

图 5 示出了本申请实施例的确定传输路径的方法 500 的示意性流程图。以该方法由主机执行为例进行描述, 如图 5 所示, 该方法 500 包括:

S510, 主机确定待传输报文所属的流对应的当前路径上发生拥塞。

具体来说, 主机在发送每个流的第一个报文时, 在流-路径表中, 记录所述流和所述主机为所述流选择的转发路径的对应关系。即, 所述流-路径表中的每个表项包括流的标识和所述主机为所述流选择的转发路径的对应关系。

当主机接收到待传输报文时, 确定所述待传输报文所属的流, 然后查找所述流-路径表, 得到所述流对应的转发路径, 所述得到的路径即为所述待传输报文所属的流对应的当前路径。

S520, 主机根据路径拥塞信息表, 为该待传输报文确定目标路径, 并将该目标路径的信息添加在该待传输报文中, 以使得该待传输报文根据该目标路径进行传输, 其中, 该目标路径的拥塞程度比该当前路径的拥塞程度小。

本申请一个实施方式中, 所述主机上还设置有转发路径表。所述转发路径表记录了所述主机连接的 Leaf 到所述网络中其他任意一个 Leaf 的所有路径。例如图 2 中, Host 10 上的转发路径表包括 L30 到 L60 的全部路径: L30→S40→L60; L30→S50→L60。

主机根据路径拥塞信息表, 为该待传输报文确定目标路径包括: 所述主机确定所述待传输报文的地址对应的目的 Leaf, 所述主机根据所述转发路径表确定所述主机连接的 Leaf 到所述目的 Leaf 的所有路径, 并根据所述拥塞信息表, 从所述主机连接的 Leaf 到所述目的 Leaf 的所有路径中选择所述目标路径。

S530, 主机根据该目标路径, 发送该待传输报文

其中, 该路径拥塞信息表的每个表项包括一条传输路径的标识以及所述传输路径对应的拥塞信息, 所述拥塞信息用于表示所述传输路径的拥塞程度。

另外, 该路径拥塞信息表存储在主机中, 该路径拥塞信息表中记录有多条传输路径的标识, 以及多条传输路径各自对应的拥塞信息, 其中, 传输路径的标识例如可以记录为: 目的地 40.0.0.2, 交换机标识 L30、L40、L60, 表示该条传输路径所到达的目的地 (例如另一主机) 的地址为 40.0.0.2, 其中经过的设备依次为 L30、L40 和 L60。

可选地, 一条传输路径的拥塞信息可以包括以下信息中的至少一种: 该传输路径的平均 ECN 次数、该传输路径的往返时间 RTT、用于表示该传输路路径是否故障的标识、

以及该传输路径上同时存在的流的个数。

其中，平均 ECN 次数表示该传输路径上特定时间内传输的流中被标记 ECN 的平均个数，例如，该条传输路径上传输了 100 个报文，其中有 50 个报文在传输过程中都发生了拥塞，那么这 50 个报文都被设置了 ECN 标记，该传输路径的平均 ECN 次数就是
5 50/100=0.5。

具体地说，若主机在确定该待传输报文对应的当前路径上发送拥塞时，可以根据路径拥塞信息表，为该待传输报文确定目标路径，并将该目标路径的信息添加在该待传输报文中，以使得该待传输报文通过该目标路径进行传输，其中，该目标路径的拥塞程度比该当前路径的拥塞程度小。在这里，主机只有在确定了该待传输报文对应的当前路径
10 上真正发生了拥塞时，才会为该待传输报文确定目标路径。并通过确定的目标路径发送该待传输报文。此外，该待传输的报文成为该待传输报文所属的流的一个新的 flowlet 的第一个报文。

应理解，这里的当前路径为该待传输报文当前待使用的路径，即为该待传输报文所属的流中的前一个报文所走的路径。如果待传输报文的前一个报文所走的路径没有发生
15 拥塞，那么主机仍会在这条路径上继续传输后续的报文，这时，该待传输报文的当前路径就为其前一个报文所走的路径。如果前一个报文所走的路径发生了拥塞，则主机会为待传输报文所属的流确定新的路径（即上述的目标路径），该待传输报文就会在目标路径上进行传输，而不在当前路径上进行传输，该待传输的报文成为该待传输报文所属的流的一个新的 flowlet 的第一个报文。

可见，在本申请实施例中，是否为一个流生成新的短流并不是根据前后两个报文到达时间差来确定的，而是根据当前传输路径是否发生拥塞来确定的。如果待传输报文所属的流对应的当前路径上发生拥塞，才会对流进行切片以生成新的短流。也就是说，只有待传输报文所属的流对应的当前路径发生拥塞，才会重新确定该报文的传输路径。

本申请实施例中可以称流传输过程的第一次路径选择为路由（Routing），该流中的
25 新的短流切换至与前一个短流不一样的路径的操作可以称为重路由（Rerouting）。

可选地，在 S510 中主机可以通过待传输报文所述的流是否被标记显式拥塞通知 ECN、传输路径的 RTT 或用于表示传输路径是否故障的标识等来判断当前路径是否发生拥塞。即，主机确定待传输报文所属的流在当前路径进行传输时被设置了 ECN 标记、或者当前路径对应的往返时间 RTT 大于时间阈值、或者当前路径发生故障时，表明当前路
30 径发生拥塞，主机为该待传输报文重新确定目标路径。

也就是说，主机如果发现待传输报文所属的流在当前路径进行传输时被设置了 ECN 标记，那么表明当前路径上发生了拥塞；或者可以设定一个时间阈值，如果网络设备发现从发送报文到接收针对该报文的响应报文之间的时间，即往返时间 RTT 突然变长，大于了预设的时间阈值，那么表明当前路径上发生了拥塞；又或者当前路径发生了故障，
35 那么必然在当前路径上进行传输时会发生拥塞。

另外，在每次发生路径拥塞时，主机都会将拥塞信息记录在路径拥塞信息表中，例如，主机一旦发现在一条路径上进行传输的某条流被设置了 ECN 标记，那么就会更新路径拥塞信息表中该路径的平均 ECN 次数。例如，假设该路径拥塞信息表中记录的该路径的平均 ECN 次数为 50/100=0.5，如果主机发现了之后在该路径上传输的某条流也被设置

了 ECN 标记，那么就会更新路径拥塞信息表中记录该路径的平均 ECN 次数为 $51/100=0.51$ 。

主机这时可以查询路径拥塞信息表，通过该拥塞信息表为该待传输报文确定目标路径，该目标路径的拥塞程度低于当前路径，从而该待传输报文可以切换至更好的路径上进行传输标识主机根据该路径拥塞信息表确定的目标路径，满足以下条件中的至少一种：目标路径的平均 ECN 次数小于当前路径的平均 ECN 次数、目标路径的 RTT 小于当前路径的 RTT、和目标路径上同时存在的流个数小于当前路径上同时存在的流的个数，且该目标路径没有发生故障。

相比于现有技术中仅仅依靠一段时间内出端口的利用率来表示路径的拥塞程度，本申请实施例通过包括平均 ECN 次数、往返时间 RTT、用于表示路径是否故障的标识、以及路径上同时传输的报文个数等参数的拥塞信息来表征传输路径的拥塞程度，能够更加准确的对路径的拥塞程度进行评价，从而更有效地指导路径切换。

可选地，在主机根据路径拥塞信息表，为该待传输报文确定目标路径之前，该方法还包括：主机获至少一条传输路径中每条传输路径对应的拥塞信息；主机根据该至少一条传输路径中每条传输路径对应的拥塞信息，建立或更新该路径拥塞信息表。

具体地说，主机获取可用于报文传输的每条路径，以及每条路径的拥塞信息，从而建立路径拥塞信息表，主机可以在至少一条传输路径中的每条传输路径上发送探测报文，并接收针对所述探测报文的响应报文，从而根据所述响应报文，确定至少一条传输路径中每条传输路径对应的拥塞信息，根据该至少一条传输路径中每条传输路径对应的拥塞信息，生成路径拥塞信息表。

进一步地，在建立该路径拥塞信息表后，主机还可以按照一定的周期，对该路径信息表不断地进行更新。

主机可以通过发送探测报文的方式获取每条路径对应的拥塞信息，通过源路由、X 路径 (Xpath) 或其他类似方式使探测报文在指定的路径上进行传输，从而获取每条路径的拥塞信息，建立起路径拥塞信息表。这里的源路由或 Xpath 是基于源地址进行路由选择的路由策略，可以实现根据多个不同地址，有选择性地将数据包发往不同目的地址的功能。

举例来说，源主机可以按照前面描述的前向探测的方法，依次在所有存在的传输路径上发送探测报文。如果在探测报文传输的过程中，在某个设备处发生了拥塞，那么该设备就会在该探测报文上标记 ECN，该探测报文传输至目的主机时，目的设备就获知该路径上发生了拥塞，在向主机返回针对该探测报文的响应报文时，就会将该路径拥塞的消息携带在响应报文中，从而源主机就知道该路径上发生了拥塞。

又例如，源主机在获取拥塞信息表中每条路径的拥塞信息中的 RTT 时，可以记录发送探测报文的时刻和接收到响应报文的时刻，两个时刻之间的时长即为 RTT，RTT 越长说明该路径的拥塞程度越高。

该路径拥塞信息表可以例如表 1 所示，表 1 中每个表项记录的信息除了包括路径 (path) 标识 (英文: identifier, 简称 “ID”) 和目的设备 (Dest) (即目的 Host)，还包括拥塞信息: 平均 ECN 次数、往返时间 RTT、是否故障链路 (英文: failure) 和传输路径上同时存在的流的数目。

表 1

路径标识	Destination	平均 ECN 次数	RTT	是否故障链路	流的数目
Path A	Destination 1	N_1	T_1	NO	M_1
Path B	Destination 2	N_2	T_2	NO	M_2
Path C	Destination 3	N_3	T_3	NO	M_3
Path D	Destination 4	N_4	T_4	NO	M_4
Path E	Destination 5	N_5	T_5	NO	M_5
Path F	Destination 6	N_6	T_6	NO	M_6
Path G	Destination 7	N_7	T_7	NO	M_7
Path H	Destination 8	N_8	T_8	NO	M_8

假设 Path A、Path B、Path C 和 Path D 用于传输流 1 (flow 1), Path E 和 Path F 用于传输 flow 2, Path G 和 Path H 用于传输 flow 3。流 flow 1、flow 2 和 flow 3 的五元组信息可以存储在表 2 所示的流表 (英文: flow table) 中, 并且表 2 中还包括了每个流可能使用的至少一个传输路径。

表 2

Flow 1	Path A
	Path B
	Path C
	Path D
Flow 2	Path E
	Path F
Flow 3	Path G
	Path H

主机对路径拥塞信息表的更新例如可以通过有源探测器（英文：active probe）发起定期主动探测，记录每条传输路径的平均 ECN 次数、RTT、用于表示传输路径是否故障的标识等信息并保存到该主机的路径拥塞信息表中。主机同时还可以依据这些信息为路径划分等级，以指导路径切换。

应注意，端到端 RTT 大的不一定是差路径（英文：bad path），也可能是主机 OS 处理延时导致的，但是 RTT 小的一定是好路径（英文：good path）；流的平均 ECN 次数小的不一定是好路径，可能是交换机阈值临界或采样不够导致，但是平均 ECN 次数大的一

定是差路径。

这里的 bad path 可以指拥塞程度较高的路径，主要是以每条路径的平均 ECN 次数、RTT 等来作为标准的，比如一条路径上的平均 ECN 次数越大，就说明这条路径越差，假设 ECN 为 $50/100=0.5$ ，表示条传输路径上传输了 100 个报文，其中有 50 个报文在传输过程中都被设置了 ECN 标记，也就意味着这条路径上有 50% 的报文都遇到了拥塞，那显然这条路径相比于平均 ECN 次数为 0.1 的路径来说是 bad path。又例如，若某条路径故障了，那么这条路径肯定是 bad path，需要避开。

进一步，可选地，在主机获取路径的拥塞信息的过程中，若主机获取的多个拥塞信息中包括表示路径发生故障的标识，则主机将该标识对应的传输路径从该路径拥塞信息表中删除。

也就是说主机在获取到哪条路径故障后，可以及时将该路径以及该路径的相关信息从该路径拥塞信息表中删除，路径拥塞信息表可以实时地进行该更新操作。在后续的路径选择过程中，就可以在没有故障的路径中进行路径选择。

因此，本申请实施例中，通过确定传输路径发生拥塞，并在发生路径拥塞时根据路径拥塞信息表为待传输报文确定新的路径，能够有效地解决传输路径拥塞的问题。

本申请实施例只有在当前路径确实发生拥塞、确实有必要进行路径切换时，才会进行传输路径的切换，避免了拥塞不匹配等问题，且由于路径切换的依据是当前路径上是否发生拥塞，而不是依靠前后报文到达的时间差，因而避免了长尾效应。

并且整个路径的决策由主机来执行，拥塞信息表保存在主机中，从而解决了网络设备的表项规格不够的问题，能够适用于 3 级 Clos 甚至 N 级 Clos 架构中。

另外，本申请实施例还引入了新的指标用来评价路径的拥塞程度，能够更加准确的对路径的拥塞程度进行评价，从而更有效地指导路径切换。

可选地，在 S520 中，主机根据路径拥塞信息表，为该待传输报文确定目标路径，包括：在所述待传输报文和所述当前路径上待传输的其他报文中，若所述待传输报文以预设概率被选择到，则主机根据该路径拥塞信息表，为该待传输报文确定目标路径，以使得所述待传输报文根据所述目标路径进行传输。

例如，假设待传输的且属于不同流的报文 1、报文 2 和报文 3 的当前路径均为路径 1，即三个报文都打算在路径 1 上传输，假设路径 1 这时发生了拥塞，则需要分别对报文 1 所属的流 1、报文 2 所属的流 2 和报文 3 所属的流 3 重新选择目标路径进行传输，但是，主机选择新的传输路径时，并不是对流 1、流 2 和流 3 都进行传输路径的切换，而是对三个流中的一个或两个流进行路径切换。举例来说，预设一个概率 $1/3$ ，就表明三个流中只对一个流进行路径切换，例如流 1 的服务器对流 1 进行路径切换，根据路径拥塞信息表为流 1 中的报文 1 确定目标路径作为新的传输路径，而对流 2 和流 3 不进行路径切换，让流 2 中的报文 2 和流 3 中的报文 3 仍在当前路径上进行传输。这样的好处是降低了切换后再次发生拥塞的概率，如果只存在路径 1 和路径 2，三个流同时进行切换的话，流 1、流 2 和流 3 就会都切换至路径 2 上传输，那么报文 1、报文 2 和报文 3 仍有可能发生冲突。该实施例可以适当避免多条流同时切换到另一条相同的路径上而导致另一条路径发生拥塞的情况，有效地将负载均衡到不同的传输路径上，从而降低了切换后再次发生拥塞的概率，提高了报文传输的成功率。

在确定到底对流 1、流 2 和流 3 中的哪条流进行路径切换时，可以是随机选择；也可以产生一个随机数看是不是可以被 3 整除，如果可以被 3 整除就切换路径，否则不切换路径；也可以选择流速最大的一条或两条流进行路径切换等，本申请实施例不做限定。

另外，如果一次切换不成功，可以等待一个短时延（英文：short delay）之后才再次切换，以提高再次切换的成功率。例如主机对报文 1 所属的流 1 进行路径切换没有成功，则需要等待固定时长例如 100ms 后，再对流 1 执行路径切换操作。

可选地，在 S520 中，主机根据路径拥塞信息表，为该待传输报文确定目标路径，还可以包括：若该待传输报文对应的流的流速大于预设阈值，主机根据该路径拥塞信息表，为所述待传输报文确定目标路径。

具体地说，主机选择的对象可以是流速超过一定大小的流，在流的流速超过一定阈值时，主机才对该流进行路径切换的操作。在待传输报文所属的流对应的当前路径上发生拥塞时，主机对流的流速进行感知，例如对 TCP 发送窗口速率计算，无丢包时根据 RTT 来计算，有丢包时根据丢包率和 RTT 并按照 TCP 吞吐公式计算。只有当流的流速超过预设的阈值时才对该待传输报文所属的流进行路径切换操作，否则不进行路径切换，因为对流速较小的流进行切片完全没有必要。从而避免了不必要的路径切换。

可选地，该方法还可以包括：主机确定该目标路径的 RTT 与该当前路径的 RTT 之间的时间差；主机在等于该时间差的时长后，发送该待传输报文。

在进行路径切换后，容易带来的一个问题就是报文乱序。举例来说，Flow A 中依次被分为 flowlet 1、flowlet 2 和 flowlet 3，flowlet 1 和 flowlet 2 在路径 1 上传输，flowlet 3 切换至路径 2 上传输，如果因路径 2 的拥塞程度小于路径 1 导致 flowlet 3 在 flowlet 1 和 flowlet 2 之前到达目的设备，则该 Flow A 中的报文的顺序就由原来的 flowlet 1→flowlet 2→flowlet 3 变为 flowlet 3→flowlet 1→flowlet 2，则发生了乱序。

为了避免乱序的发生，主机确定待传输报文所属的流对应的当前路径发生拥塞而为该报文确定了新的路径即目标路径后，主机可以确定该目标路径的 RTT 与该当前路径的 RTT 之间的时间差，从而主机在经过该时间差的时长后，再发送该待传输报文。这样可以避免路径切换导致的报文乱序。

下面结合图 6，详细地描述本申请实施例的确定传输路径的方法。图 6 为根据本申请实施例的方法进行路径切换的示意图，包括发送端（sender）10、接收端（receiver）20、Leaf 30、Spine 40、Spine 50 和 Leaf 60。如图 6 所示，发送端 10 和接收端 20 之间存在两条传输路径，分别为 Path 1 和 Path 2，传输的流为 Flow A，Flow A 假设被分为两个 flowlet，即 flowlet 1 和 flowlet 2。发送端 10 的主机的传输层可以实现传输层感知负载均衡 TLB，接收端 20 的主机的传输层也可以实现传输层感知负载均衡 TLB，该确定传输路径的方法具体包括：

（1）当有新流需要发送时，主机将该流的五元组记录到流表内。

（2）首次路由操作(Route)：主机针对该新流查找路径拥塞信息表，在除了 bad path 之外的路径中随机选择传输路径，该 bad path 可以由主机按照前述方式即依据平均 ECN 次数、RTT 等信息来确定。如图 6 所示，主机首先为当前 Flow A 中的 flowlet 1 选择 Path 1 进行报文发送。

(3) flowlet 1 在 Path 1 上发送顺利, 主机继续在 Path 1 上发送 Flow A 中的 flowlet 2。

(4) Path 1 发生路径拥塞, 主机确定是否为 Flow A 执行重路由操作。

其中, Path 1 发生路径拥塞可以是: 发送端 Host 10 确定 Flow A 在 Path 1 上传输时被设置了 ECN 标记、Path 1 的平均 ECN 次数大于预设值、Path 1 的 RTT 超过预设阈值、或者 Path 1 发生链路故障。

主机可以是根据 Flow A 的流速确定是否为 flowlet 2 执行重路由操作, 当流速大于预设阈值时, 对 Flow A 进行重路由, 否则跳过选择下一个流是否符合要求; 主机也可以根据预先设定的概率值, 确定是否为 Flow A 执行重路由操作。

(5) 主机确定需要为 Flow A 执行重路由操作时, 根据路径拥塞信息表确定目标路径以用于传输为 Flow A 中的 flowlet 2。

主机查找路径拥塞信息表, 选择目标路径例如这里的 Path 2, 并将 Path 2 的信息携带在 flowlet 2 中, 以使得 flowlet 2 按照 Path 2 进行传输。Path 2 满足以下条件中的至少一个: Path 2 的平均 ECN 次数小于 Path 1 的平均 ECN 次数, Path 2 的 RTT 比 Path 1 的 RTT 短, Path 2 上同时传输的报文个数小于 Path 1 上同时传输的报文个数。

(6) 主机确定 Path 1 的 RTT 与 Path 2 的 RTT 之间的时间差。

主机计算主机 Path 1 的 RTT 与 Path 2 的 RTT 之间的时间差例如为 100us。

(7) 主机将 flowlet 2 在本地缓存 100us 后, 将 flowlet 2 在 Path 2 上进行发送。

这样, 主机在发生路径拥塞时根据路径拥塞信息表为待传输报文确定新的路径, 有效地解决了传输路径拥塞的问题。

应理解, 在本申请的各种实施例中, 上述各过程的序号的大小并不意味着执行顺序的先后, 各过程的执行顺序应以其功能和内在逻辑确定, 而不对本申请实施例的实施过程构成任何限定。

下面将结合图 7, 描述根据本申请实施例的确定传输路径的装置, 方法实施例所描述的技术特征可以适用于以下装置实施例。

图 7 示出了根据本申请实施例的确定传输路径的装置 700。如图 7 所示, 该装置 700 包括确定单元 710 和发送单元 720。其中:

确定单元 710 用于: 确定待传输报文所属的流对应的当前路径上发生拥塞; 根据路径拥塞信息表, 为所述待传输报文确定目标路径, 并将所述目标路径的信息添加在所述待传输报文中, 以使得所述待传输报文通过所述目标路径进行传输, 其中, 所述目标路径的拥塞程度比所述当前路径的拥塞程度小, 所述路径拥塞信息表的每个表项包括一条传输路径以及所述传输路径对应的拥塞信息, 所述拥塞信息用于表示所述传输路径的拥塞程度;

发送单元 720 用于: 根据所述确定单元确定的目标路径, 发送所述待传输报文。

因此, 本申请实施例的确定传输路径的装置, 通过确定传输路径是否发生拥塞, 并在发生路径拥塞时根据路径拥塞信息表为待传输报文确定新的路径, 能够有效地解决传输路径拥塞的问题。

由于本申请实施例只有在当前路径确实发生拥塞、确实有必要进行路径切换时, 才会进行传输路径的切换, 避免了拥塞不匹配等问题, 且切换后的路径的拥塞程度小于该

当前路径，避免了长尾效应。

并且本申请实施例中，整个路径的决策并不需要由交换机、路由器等网络设备来执行，而是通过其他装置例如主机来实现，拥塞信息表等信息也可以保存在主机中，从而解决了表项规格受限的问题，能够适用于 3 级 Clos 甚至 N 级 Clos 架构中。

5 另外，本申请实施例还引入了新的评价指标用来评价路径的拥塞程度，例如平均 ECN 次数、往返时间 RTT、用于表示路径是否故障的标识、以及路径上同时存在的流的个数等参数，从而能够更加准确的对路径的拥塞程度进行评价，更有效地指导路径切换。

10 可选地，所述确定单元 710 具体用于：确定所述流在所述当前路径上传输时被标记显式拥塞通知 ECN，或者所述当前路径对应的往返时间 RTT 大于时间阈值，或者所述当前路径发生故障。

可选地，所述拥塞信息包括以下信息中的至少一种：所述传输路径的平均 ECN 次数、所述传输路径的往返时间 RTT、用于表示所述传输路径是否故障的标识、以及所述传输路径上同时存在的流的个数。

15 可选地，若所述待传输报文对应的流的流速大于预设阈值，根据所述路径拥塞信息表，为所述待传输报文确定目标路径。

可选地，所述确定单元 710 具体用于：在所述待传输报文和所述当前路径上待传输的其他报文中，若所述待传输报文以预设的概率被选择到，则根据所述路径拥塞信息表，为所述待传输报文确定所述目标路径，以使得所述待传输报文根据所述目标路径进行传输。

20 可选地，所述确定单元 710 还用于：确定所述目标路径的 RTT 与所述当前路径的 RTT 之间的时间差；所述发送单元还用于：在经过所述时间差的时长后，发送所述待传输报文。

25 可选地，所述装置还包括接收单元，其中，所述发送单元 720 还用于：在至少一条传输路径中的每条传输路径上发送探测报文；所述接收单元用于，接收针对所述探测报文的响应报文；所述确定单元还用于，根据所述响应报文，确定至少一条传输路径中每条传输路径对应的所述拥塞信息；所述确定单元还用于，根据所述至少一条传输路径中每条传输路径对应的所述拥塞信息，生成所述路径拥塞信息表。

可选地，所述确定单元还用于：若获取的一条传输路径的拥塞信息中包括表示所述传输路径发生故障的标识，将所述传输路径对应的表项从所述路径拥塞信息表中删除。

30 图 8 示出了本申请实施例提供的确定传输路径的装置的结构，处理器 810、收发器 820 和存储器 830，其中，该处理器 810、收发器 820 和存储器 830 之间通过内部连接通路互相通信。该存储器 830 用于存储指令，该处理器 810 用于执行该存储器 830 存储的指令，以控制该收发器 820 接收信号或发送信号。其中：

35 处理器 810 用于：根据路径拥塞信息表，为所述待传输报文确定目标路径，并将所述目标路径的信息添加在所述待传输报文中，以使得所述待传输报文通过所述目标路径进行传输，其中，所述目标路径的拥塞程度比所述当前路径的拥塞程度小，所述路径拥塞信息表的每个表项包括一条传输路径以及所述传输路径对应的拥塞信息，所述拥塞信息用于表示所述传输路径的拥塞程度；

收发器 820 用于：根据所述确定单元确定的目标路径，发送所述待传输报文。

因此，本申请实施例的确定传输路径的装置，通过确定传输路径是否发生拥塞，并在发生路径拥塞时根据路径拥塞信息表为待传输报文确定新的路径，能够有效地解决传输路径拥塞的问题。

5 可选地，所述处理器 810 具体用于：确定所述流在所述当前路径上传输时被标记显式拥塞通知 ECN，或者所述当前路径对应的往返时间 RTT 大于时间阈值，或者所述当前路径发生故障。

可选地，所述拥塞信息包括以下信息中的至少一种：所述传输路径的平均 ECN 次数、所述传输路径的往返时间 RTT、用于表示所述传输路径是否故障的标识、以及所述传输路径上同时存在的流的个数。

10 可选地，若所述待传输报文对应的流的流速大于预设阈值，根据所述路径拥塞信息表，为所述待传输报文确定目标路径。

可选地，所述处理器 810 具体用于：在所述待传输报文和所述当前路径上待传输的其他报文中，若所述待传输报文以预设的概率被选择到，则根据所述路径拥塞信息表，为所述待传输报文确定所述目标路径，以使得所述待传输报文根据所述目标路径进行传
15 输。

可选地，所述处理器 810 还用于：确定所述目标路径的 RTT 与所述当前路径的 RTT 之间的时间差；所述收发器 820 还用于：在经过所述时间差的时长后，发送所述待传输报文。

可选地，所述收发器 820 还用于：在至少一条传输路径中的每条传输路径上发送探测报文；所述收发器 820 还用于，接收针对所述探测报文的响应报文；所述处理器 810
20 还用于，根据所述响应报文，确定至少一条传输路径中每条传输路径对应的所述拥塞信息；所述处理器 810 还用于，根据所述至少一条传输路径中每条传输路径对应的所述拥塞信息，生成所述路径拥塞信息表。

可选地，所述确定单元还用于：若获取的一条传输路径的拥塞信息中包括表示所述
25 传输路径发生故障的标识，将所述传输路径对应的表项从所述路径拥塞信息表中删除。

应理解，在本申请实施例中，该处理器 810 可以是中央处理单元（central processing unit，简称“CPU”），该处理器 810 还可以是其他通用处理器、数字信号处理器（英文：digital signal processor，简称“DSP”）、专用集成电路（英文：application-specific integrated circuit，简称“ASIC”）、现成可编程门阵列（英文：field programmable gate array，简称“FPGA”）或者其他可编程逻辑器件、分立
30 门或者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。

该存储器 830 可以包括只读存储器和随机存取存储器，并向处理器 810 提供指令和数据。存储器 830 的一部分还可以包括非易失性随机存取存储器。例如，存储器 830 还可以存储设备类型的信息。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，上述描述的系统、装置和单元的具体工作过程，可以参考前述方法实施例中的对应
35 过程，在此不再赘述。

在实现过程中，上述方法的各步骤可以通过处理器 810 中的硬件的集成逻辑电路或者软件形式的指令完成。结合本申请实施例所公开的定位方法的步骤可以直接体现为硬

件处理器执行完成，或者用处理器 810 中的硬件及软件模块组合执行完成。软件模块可以位于随机存储器，闪存、只读存储器，可编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介质中。该存储介质位于存储器 830，处理器 810 读取存储器 830 中的信息，结合其硬件完成上述方法的步骤。为避免重复，这里不再详细描述。

5 本领域普通技术人员可以意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，能够以电子硬件、或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行，取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本申请的范围。

10 在本申请所提供的几个实施例中，应该理解到，所揭露的系统、装置和方法，可以通过其它的方式实现。例如，以上所描述的装置实施例仅仅是示意性的，例如，所述单元的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个单元或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口，装置
15 或单元的间接耦合或通信连接，可以是电性，机械或其它的形式。

所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

20 另外，在本申请各个实施例中的各功能单元可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中。

所述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，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，本申请实施例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
25 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，主机，或者网络设备等）执行本申请各个实施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（英文：read-only memory，简称 ROM）、随机存取存储器（英文：random access memory，简称“RAM”）、磁碟或者光盘等各种可以存储程序代码的介质。

30 以上所述，仅为本申请实施例的具体实施方式，但本申请的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本申请揭露的技术范围内，可轻易想到变化或替换，都应涵盖在本申请的保护范围之内。因此，本申请的保护范围应以所述权利要求的保护范围为准。

权利要求

1、一种确定传输路径的方法，其特征在于，所述方法包括：

确定待传输报文所属的流对应的当前路径上发生拥塞；

5 根据路径拥塞信息表，为所述待传输报文确定目标路径，并将所述目标路径的信息添加在所述待传输报文中，以使得所述待传输报文根据所述目标路径进行传输，其中，所述目标路径的拥塞程度比所述当前路径的拥塞程度小，所述路径拥塞信息表的每个表项包括一条传输路径以及所述传输路径对应的拥塞信息，所述拥塞信息用于表示所述传输路径的拥塞程度；

10 根据所述目标路径，发送所述待传输报文。

2、根据权利要求 1 所述的方法，其特征在于，所述确定待传输报文所属的流对应的当前路径上发生拥塞，包括：

确定所述流在所述当前路径上传输时被标记显式拥塞通知 ECN，或者所述当前路径对应的往返时间 RTT 大于时间阈值，或者所述当前路径发生故障。

15 3、根据权利要求 1 或 2 所述的方法，其特征在于，所述拥塞信息包括以下信息中的至少一种：

所述传输路径的平均 ECN 次数、所述传输路径的 RTT、用于表示所述传输路径是否故障的标识、以及所述传输路径上同时存在的流的个数。

4、根据权利要求 1 至 3 中任一项所述的方法，其特征在于，所述根据路径拥塞信息表，为所述待传输报文确定目标路径之前，所述方法还包括：

20 确定所述待传输报文所属的流的流速是否大于预设阈值；

若所述流的流速大于所述预设阈值，执行所述根据所述路径拥塞信息表，为所述待传输报文确定目标路径的步骤。

5、根据权利要求 1 至 4 中任一项所述的方法，其特征在于，所述根据路径拥塞信息表，为所述待传输报文确定目标路径，包括：

25 在所述待传输报文和所述当前路径上待传输的其他报文中，若所述待传输报文以预设概率被选择到，则根据所述路径拥塞信息表，为所述待传输报文确定所述目标路径，以使得所述待传输报文根据所述目标路径进行传输。

6、根据权利要求 1 至 5 中任一项所述的方法，其特征在于，所述方法还包括：

30 确定所述目标路径的 RTT 与所述当前路径的 RTT 之间的时间差；

在经过所述时间差的时长后，发送所述待传输报文。

7、根据权利要求 1 至 6 中任一项所述的方法，其特征在于，所述在根据路径拥塞信息表，为所述待传输报文确定目标路径之前，所述方法还包括：

在至少一条传输路径中的每条传输路径上发送探测报文；

35 接收针对所述探测报文的响应报文；

根据所述响应报文，确定至少一条传输路径中每条传输路径对应的所述拥塞信息；

根据所述至少一条传输路径中每条传输路径对应的所述拥塞信息，生成所述路径拥塞信息表。

8、根据权利要求 7 所述的方法，其特征在于，所述方法还包括：

若获取的一条传输路径的拥塞信息中包括表示所述传输路径发生故障的标识，将所述传输路径对应的表项从所述路径拥塞信息表中删除。

9、一种确定传输路径的装置，其特征在于，所述装置包括：

确定单元，用于确定待传输报文所属的流对应的当前路径上发生拥塞；

5 所述确定单元还用于，根据路径拥塞信息表，为所述待传输报文确定目标路径，并将所述目标路径的信息添加在所述待传输报文中，以使得所述待传输报文通过所述目标路径进行传输，其中，所述目标路径的拥塞程度比所述当前路径的拥塞程度小，所述路径拥塞信息表的每个表项包括一条传输路径以及所述传输路径对应的拥塞信息，所述拥塞信息用于表示所述传输路径的拥塞程度；

10 发送单元，用于根据所述确定单元确定的目标路径，发送所述待传输报文。

10、根据权利要求9所述的装置，其特征在于，所述确定单元具体用于：

确定所述流在所述当前路径上传输时被标记显式拥塞通知 ECN，或者所述当前路径对应的往返时间 RTT 大于时间阈值，或者所述当前路径发生故障。

11、根据权利要求9或10所述的装置，其特征在于，所述拥塞信息包括以下信息
15 中的至少一种：

所述传输路径的平均 ECN 次数、所述传输路径的往返时间 RTT、用于表示所述传输路径是否故障的标识、以及所述传输路径上同时存在的流的个数。

12、根据权利要求9至11中任一项所述的装置，其特征在于，所述确定单元具体用于：

20 若所述待传输报文对应的流的流速大于预设阈值，根据所述路径拥塞信息表，为所述待传输报文确定目标路径。

13、根据权利要求9至12中任一项所述的装置，其特征在于，所述确定单元具体用于：

25 在所述待传输报文和所述当前路径上待传输的其他报文中，若所述待传输报文以预设的概率被选择到，则根据所述路径拥塞信息表，为所述待传输报文确定所述目标路径，以使得所述待传输报文根据所述目标路径进行传输。

14、根据权利要求9至13中任一项所述的装置，其特征在于，所述确定单元还用于：

确定所述目标路径的 RTT 与所述当前路径的 RTT 之间的时间差；

30 所述发送单元还用于：在经过所述时间差的时长后，发送所述待传输报文。

15、根据权利要求9至14中任一项所述的装置，其特征在于，所述装置还包括接收单元，其中，所述发送单元还用于：

在至少一条传输路径中的每条传输路径上发送探测报文；

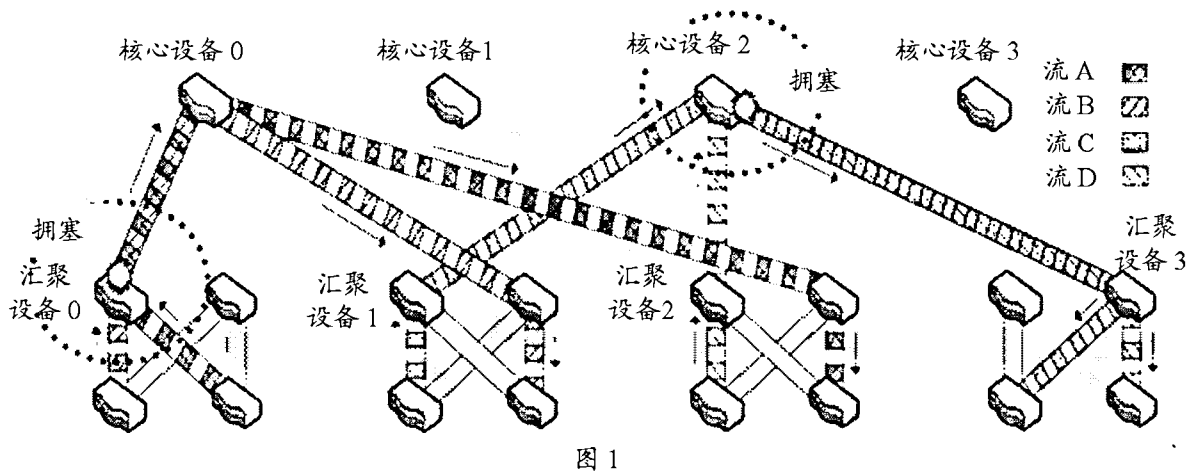
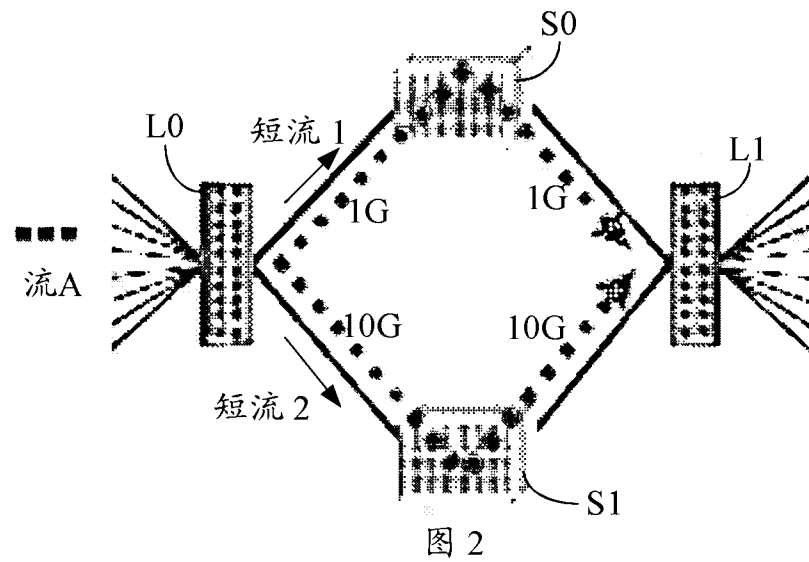
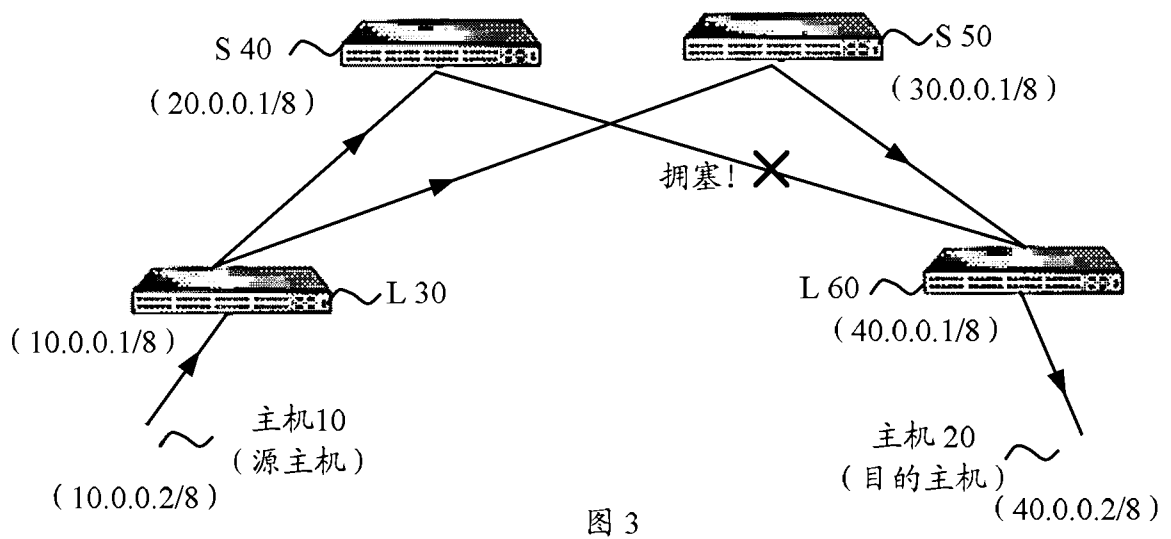
所述接收单元用于，接收针对所述探测报文的响应报文；

35 所述确定单元还用于，根据所述响应报文，确定至少一条传输路径中每条传输路径对应的所述拥塞信息；

所述确定单元还用于，根据所述至少一条传输路径中每条传输路径对应的所述拥塞信息，生成所述路径拥塞信息表。

16、根据权利要求15所述的装置，其特征在于，所述确定单元还用于：

40 若获取的一条传输路径的拥塞信息中包括表示所述传输路径发生故障的标识，将所述传输路径对应的表项从所述路径拥塞信息表中删除。



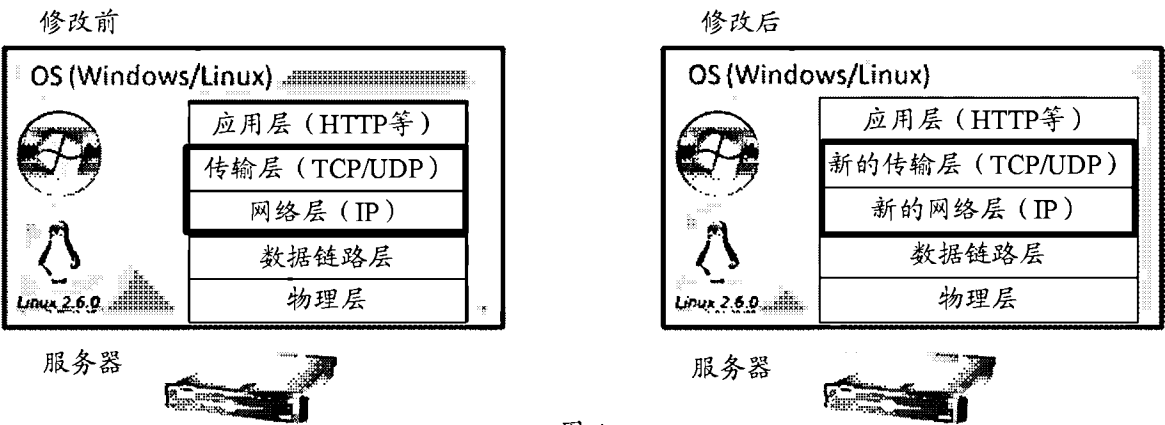


图 4

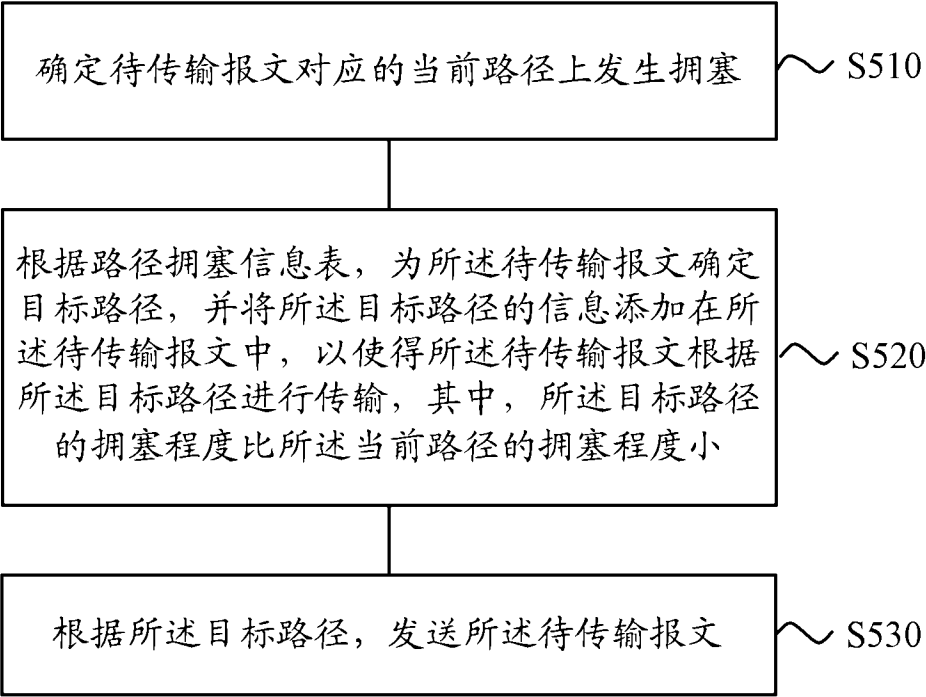


图 5

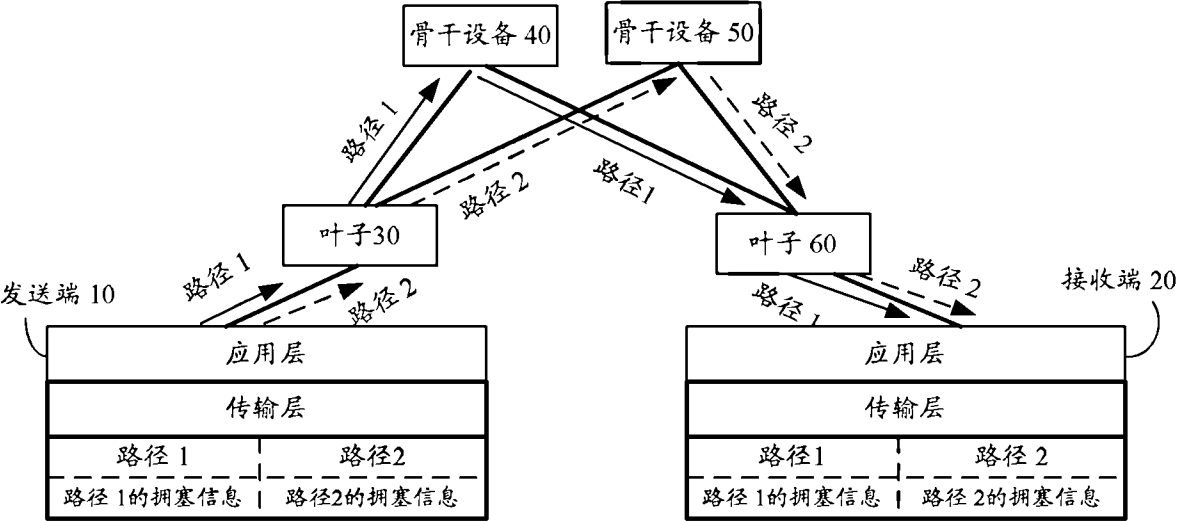


图 6

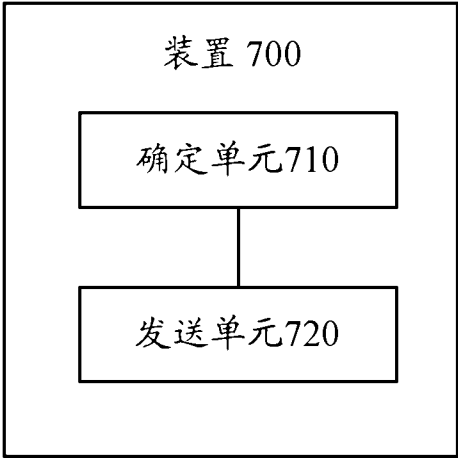


图 7

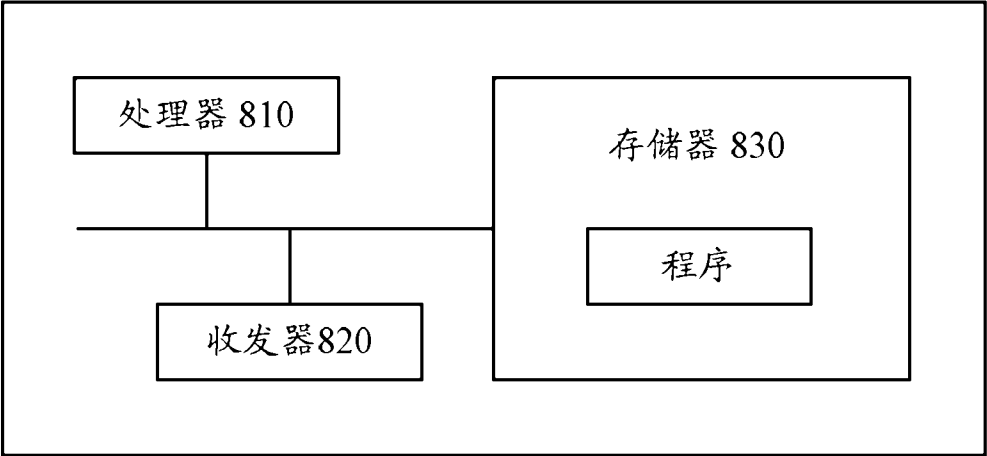


图 8

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/109372

A. CLASSIFICATION OF SUBJECT MATTER

H04L 12/801 (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)

CNKI, CNPAT, WPI, EPODOC: 拥塞, 阻塞, 路径, 路由, 探测, 报文, 流, 短流, 重路由, 选择, 表项, 备用, 流量, 转移, 流速, 阈值, 门限, 时间差, 往返, 时延, 故障, delay, router, ECN, RTT, overlay, network, data, traffic, threshold, time, list, flow, path, link, packet, message, congest, stream

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 103051546 A (INSTITUTE OF COMPUTING TECHNOLOGY, CHINESE ACADEMY OF SCIENCES), 17 April 2013 (17.04.2013), description, paragraphs [0005]-[0061]	1-16
A	CN 106230722 A (SHANDONG COMPUTER SCIENCE CENTER (NATIONAL SUPERCOMPUTER CENTER IN JINAN)), 14 December 2016 (14.12.2016), entire document	1-16
A	CN 102025644 A (HUAWEI TECHNOLOGIES CO., LTD.), 20 April 2011 (20.04.2011), entire document	1-16
A	CN 101335714 A (HUAWEI TECHNOLOGIES CO., LTD.), 31 December 2008 (31.12.2008), entire document	1-16
A	WO 2016124049 A1 (HUAWEI TECHNOLOGIES CO., LTD.), 11 August 2016 (11.08.2016), entire document	1-16

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	“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
“A” document defining the general state of the art which is not considered to be of particular relevance	“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
“E” earlier application or patent but published on or after the international filing date	“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
“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)	“&” document member of the same patent family
“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	

Date of the actual completion of the international search 08 January 2018	Date of mailing of the international search report 01 February 2018
Name and mailing address of the ISA 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 NIU, Xiaojia Telephone No. (86-10) 61648528

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2017/109372

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 103051546 A	17 April 2013	CN 103051546 B	08 July 2015
CN 106230722 A	14 December 2016	None	
CN 102025644 A	20 April 2011	CN 102025644 B	17 October 2012
CN 101335714 A	31 December 2008	None	
WO 2016124049 A1	11 August 2016	US 2017338976 A1	23 November 2017
		EP 3253012 A1	06 December 2017
		CN 105991435 A	05 October 2016

国际检索报告

国际申请号

PCT/CN2017/109372

A. 主题的分类

H04L 12/801(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))

CNKI, CNPAT, WPI, EPODOC; 拥塞, 阻塞, 路径, 路由, 探测, 报文, 流, 短流, 重路由, 选择, 表项, 备用, 流量, 转移, 流速, 阈值, 门限, 时间差, 往返, 时延, 故障, delay, router, ECN, RTT, overlay, network, data, traffic, threshold, time, list, flow, path, link, packet, message, congest, stream

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 103051546 A (中国科学院计算技术研究所) 2013年 4月 17日 (2013 - 04 - 17) 说明书第[0005]-[0061]段	1-16
A	CN 106230722 A (山东省计算中心国家超级计算济南中心) 2016年 12月 14日 (2016 - 12 - 14) 全文	1-16
A	CN 102025644 A (华为技术有限公司) 2011年 4月 20日 (2011 - 04 - 20) 全文	1-16
A	CN 101335714 A (华为技术有限公司) 2008年 12月 31日 (2008 - 12 - 31) 全文	1-16
A	WO 2016124049 A1 (华为技术有限公司) 2016年 8月 11日 (2016 - 08 - 11) 全文	1-16

☐ 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

2018年 1月 8日

国际检索报告邮寄日期

2018年 2月 1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

牛晓佳

电话号码 (86-10)61648528

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2017/109372

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	103051546	A	2013年 4月 17日	CN	103051546	B	2015年 7月 8日
CN	106230722	A	2016年 12月 14日	无			
CN	102025644	A	2011年 4月 20日	CN	102025644	B	2012年 10月 17日
CN	101335714	A	2008年 12月 31日	无			
WO	2016124049	A1	2016年 8月 11日	US	2017338976	A1	2017年 11月 23日
				EP	3253012	A1	2017年 12月 6日
				CN	105991435	A	2016年 10月 5日

表 PCT/ISA/210 (同族专利附件) (2009年7月)