

(19) 世界知识产权组织

国际局

(43) 国际公布日

2014年6月19日 (19.06.2014)



(10) 国际公布号

WO 2014/090020 A1

(51) 国际专利分类号:

H04L 29/08 (2006.01)

(21) 国际申请号:

PCT/CN2013/084519

(22) 国际申请日:

2013年9月27日 (27.09.2013)

(25) 申请语言:

中文

(26) 公布语言:

中文

(30) 优先权:

201210543591.8 2012年12月14日 (14.12.2012) CN

(71) 申请人: 中兴通讯股份有限公司 (ZTE CORPORATION) [CN/CN]; 中国广东省深圳市南山区高新技术产业园科技南路中兴通讯大厦, Guangdong 518057 (CN)。

(72) 发明人: 胡永生 (HU, Yongsheng); 中国广东省深圳市南山区高新技术产业园科技南路中兴通讯大厦, Guangdong 518057 (CN)。

(74) 代理人: 北京派特恩知识产权代理事务所(普通合伙) (CHINA PAT INTELLECTUAL PROPERTY OFFICE); 中国北京市海淀区海淀南路21号中关村知识产权大厦B座2层, Beijing 100080 (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, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, 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, 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)。

本国国际公布:

— 包括国际检索报告(条约第21条(3))。

(54) Title: DHT-BASED CONTROL NETWORK IMPLEMENTATION METHOD AND SYSTEM, AND NETWORK CONTROLLER

(54) 发明名称: 一种基于 DHT 的控制网络实现方法、系统和网络控制器

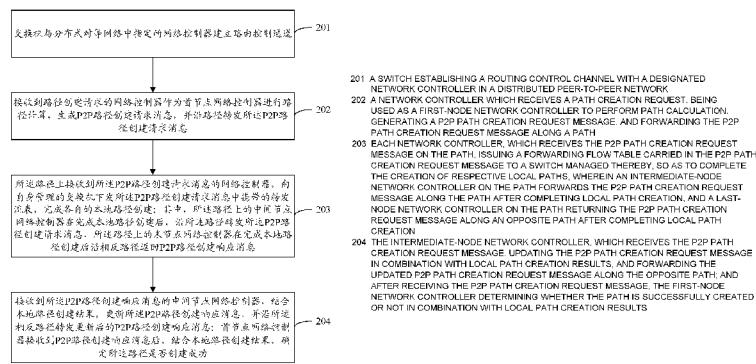


图 2 / FIG. 2

(57) Abstract: Disclosed are a DHT-based control network implementation method and system. The method comprises: a switch establishing a routing control channel with a designated network controller in a distributed peer-to-peer network; a first-node network controller generating a P2P path creation request message, and forwarding same along a path until reaching a last-node network controller; each of the network controllers on the path issuing a forwarding flow table carried in the P2P path creation request message to a switch managed thereby according to the P2P path creation request message, so as to complete the creation of respective local paths; and according to the received P2P path creation responses and in combination with local path creation results, the first-node network controller determining whether paths are successfully created or not. Also disclosed is a network controller. By means of the present invention, the network of network controllers is self-organized and switches are automatically accessed, and meanwhile, calculation and issue of network paths are easy to manage.

(57) 摘要:

[见续页]



本发明公开了一种基于 DHT 的控制网络实现方法和系统，包括：交换机与分布式对等网络中指定的网络控制器建立路由控制通道；首节点网络控制器生成 P2P 路径创建请求消息，并沿路径转发，直到末节点网络控制器；路径上的每个网络控制器根据 P2P 路径创建请求消息，向自身管理的交换机下发 P2P 路径创建请求消息中携带的转发流表，完成各自的本地路径创建；首节点网络控制器根据接收到的 P2P 路径创建响应并结合本地路径创建结果确定路径是否创建成功。本发明还公开了一种网络控制器。通过本发明使得网络控制器网络实现自组织、交换机自动接入，同时使得网络路径的计算与下发容易管理。

一种基于 DHT 的控制网络实现方法、系统和网络控制器

技术领域

本发明涉及计算机网络领域，尤其是涉及一种基于分布式哈希（DHT，Distributed Hash Table）的控制网络实现方法、系统和网络控制器。

5 背景技术

美国斯坦福大学于 2008 年提出了开放流表控制（OpenFlow）协议，该协议采用转发/控制分离架构，外置控制面实体采用 OpenFlow 协议控制转发面设备，实现各种转发逻辑。而转发面设备主要功能就是根据 OpenFlow 网络控制器下发的流表执行受控转发，其行为是标准化的：收到一条报文，
10 取出其头部 L2/L3/L4 相关字段值，以其作为关键字查找流表，匹配到一个表项后，根据表项内容中的指令集对报文字段进行变换，完毕后根据指示转发到某一逻辑或物理端口。此协议进一步演进，成为软件定义网络（SDN，Software Defined Network）技术的基础，即可以在控制面采用软件编程实现各种复杂的网络应用，而转发面设备无需任何改变，由于控制面采用通用
15 服务器+通用操作系统，并且可以使用通用的软件编程工具，也可以使用 Python（一种面向对象、直译式计算机程序设计语言）这样的脚本编程语言实现，这使得新的网络协议的支持变得非常简单，而且新技术部署周期大大缩短。

一个基本的 OpenFlow 网络包括网络控制器和交换机两个主要的控制
20 实体。通常地，在网络中部署一个集中的网络控制器，负责整个网络域中所有交换机的管理域控制。然而，单个集中式的网络控制器会成为网络规模化的瓶颈，如与网络控制器距离较远的交换机建立流的时延增加，单个网络控制器处理交换路径请求的吞吐量受限，端到端的路径带宽可控性差

等。为此，学术界提出分布式网络网络控制器方案，如超级流管理(HyperFlow)、ONIX等，这些方案一般是在网络上部署多个网络控制器共同管理与控制SDN。

公开号为US20110261722的美国专利申请公开的技术方案中：将路由路径信息放在流的首个数据报文头部中，路径上的交换机在进行报文转发时根据报文头部的路由路径信息建立转发流表。对于这种大规模的网络，交换机和网络控制器之间的连接选择与维护、路径的计算与下发，经常涉及多个交换机、多个网络控制器，大大增加了网络维护的成本。

由此可见，在现有SDN的分布式网络控制器方案中，路径建立过程中经常涉及多个网络控制器、多个交换机的流表下发，复杂的交互过程增加了事务维护的复杂性，且不利于管理面和转发面的状态统一，大大增加了网络的维护成本。

发明内容

有鉴于此，本发明实施例的主要目的在于提供一种基于DHT的控制网络实现方法和系统，使得网络控制器网络实现自组织、交换机自动接入，同时使得网络路径的计算与下发容易管理。

为达到上述目的，本发明实施例的技术方案是这样实现的：

本发明实施例提供了一种基于DHT的控制网络实现方法，所述方法包括：

20 交换机与分布式对等网络中指定的网络控制器建立路由控制通道；

接收到路径创建请求的网络控制器作为首节点网络控制器进行路径计算，生成P2P路径创建请求消息，并沿路径转发所述P2P路径创建请求消息；

25 所述路径上接收到所述P2P路径创建请求消息的网络控制器，向自身管理的交换机下发所述P2P路径创建请求消息中携带的转发流表，完成各

自的本地路径创建；其中，所述路径上的中间节点网络控制器在完成本地路径创建后，沿所述路径转发所述 P2P 路径创建请求消息，所述路径上的末节点网络控制器在完成本地路径创建后沿相反路径返回 P2P 路径创建响应消息；

5 接收到所述 P2P 路径创建响应消息的中间节点网络控制器，结合本地路径创建结果，更新所述 P2P 路径创建响应消息，并沿所述相反路径转发更新后的 P2P 路径创建响应消息；首节点网络控制器接收到 P2P 路径创建响应消息后，结合本地路径创建结果，确定所述路径是否创建成功。

10 优选地，所述交换机与分布式对等网络中指定的网络控制器建立路由控制通道，包括：

交换机与引导服务器直接返回的指定的网络控制器列表中的网络控制器建立路由控制通道；或者，引导服务器将交换机引导到任意的网络控制器，交换机根据所述任意的网络控制器返回的信息重定向到指定的网络控制器、并与所述指定的网络控制器建立路由控制通道。

15 优选地，所述路由控制通道为 OpenFlow 协议通道、SNMP 通道或 DIAMETER 通道。

优选地，所述路径计算得到路径上网络控制器的顺序和路径上交换机的顺序。

20 优选地，所述 P2P 路径创建请求消息的目的节点为所述路径上的末节点网络控制器；

所述 P2P 路径创建请求消息的路由路径头部由所述路径上除了末节点网络控制器之外的其它网络控制器的标识组成；

所述 P2P 路径创建请求消息的消息体由所述路径上向各个交换机下发的转发流表组成。

25 优选地，所述其它网络控制器的标识按所述路径上的所述其它网络控

制器的顺序排列；

所述向各个交换机下发的转发流表按照所述路径上的交换机的顺序排列。

优选地，所述首节点网络控制器生成 P2P 路径创建请求消息后，所述方法还包括：所述首节点网络控制器向自身管理的交换机下发所述 P2P 路径创建请求消息中携带的转发流表。

优选地，所述路径上的末节点网络控制器在完成本地路径创建后沿相反路径返回 P2P 路径创建响应消息，为：

若本地路径创建成功，所述末节点网络控制器返回成功的 P2P 路径创建响应消息；若本地路径创建失败，所述末节点网络控制器返回失败的 P2P 路径创建响应消息。

优选地，所述接收到所述 P2P 路径创建响应消息的中间节点网络控制器，结合本地路径创建结果，更新所述 P2P 路径创建响应消息，并沿所述相反路径转发更新后的 P2P 路径创建响应消息，为：

若本地路径创建成功、且接收到成功的 P2P 路径创建响应消息，则沿所述相反路径转发更新后的、成功的 P2P 路径创建响应消息；

若本地路径创建成功、且接收到失败的 P2P 路径创建响应消息，则沿所述相反路径转发更新后的、失败的 P2P 路径创建响应消息；

若本地路径创建失败，且接收到失败或成功的 P2P 路径创建响应消息，
20 则沿所述相反路径转发更新后的、失败的 P2P 路径创建响应消息。

优选地，所述首节点网络控制器接收到 P2P 路径创建响应消息后，结合本地路径创建结果，确定所述路径是否创建成功，为：

所述首节点网络控制器接收到成功的 P2P 路径创建响应消息、且本地路径创建成功时，确认所述路径创建成功；

所述首节点网络控制器接收到失败的 P2P 路径创建响应消息或者本地
25

路径创建失败时，确认所述路径创建失败，并重新进行路径计算。

优选地，所述路径上的网络控制器接收到 P2P 路径创建响应消息后，所述方法还包括：若接收到为失败的 P2P 路径创建响应消息或者本地路径创建失败时，网络控制器对本地已下发的转发流表进行回滚操作。

5 本发明实施例还提供了一种基于 DHT 的控制网络实现系统，所述系统包括：交换机和网络控制器，其中：

所述交换机，配置为与分布式对等网络中指定的网络控制器建立路由控制通道；

所述网络控制器，配置为接收到路径创建请求时作为首节点网络控制器进行路径计算，生成 P2P 路径创建请求消息，并沿路径转发所述 P2P 路径创建请求消息；还配置为接收到所述 P2P 路径创建请求消息时，向自身管理的交换机下发所述 P2P 路径创建请求消息中携带的转发流表，完成本地路径创建；还配置为作为所述路径上的中间节点网络控制器在完成本地路径创建后，沿所述路径转发所述 P2P 路径创建请求消息；还配置为，作为所述路径上的末节点网络控制器在完成本地路径创建后沿相反路径返回 P2P 路径创建响应消息；还配置为作为路径上的中间节点网络控制器接收到所述 P2P 路径创建响应消息时，结合本地路径创建结果，更新所述 P2P 路径创建响应消息，并沿所述相反路径转发更新后的 P2P 路径创建响应消息；还配置为作为首节点网络控制器接收到 P2P 路径创建响应消息后，结合本地路径创建结果，确定所述路径是否创建成功。

优选地，该系统还包括引导服务器，配置为向所述交换机指定的网络控制器列表；或者，将所述交换机引导到任意的网络控制器；

相应地，所述交换机，还配置为与所述引导服务器直接返回的指定的网络控制器列表中的网络控制器建立路由控制通道；或者，根据所述任意的网络控制器返回的信息重定向到指定的网络控制器、并与所述指定的网

络控制器建立路由控制通道。

优选地，所述网络控制器，还配置为作为首节点网络控制器生成 P2P 路径创建请求消息后，向自身管理的交换机下发所述 P2P 路径创建请求消息中携带的转发流表。

5 优选地，所述网络控制器，还配置为作为末节点网络控制器，若本地路径创建成功，沿相反路径返回成功的 P2P 路径创建响应消息；若本地路径创建失败，沿相反路径返回失败的 P2P 路径创建响应消息。

10 优选地，所述网络控制器，还配置为作为中间节点网络控制器接收到所述 P2P 路径创建响应消息时，若本地路径创建成功、且接收到成功的 P2P 路径创建响应消息，则沿所述相反路径转发更新后的、成功的 P2P 路径创建响应消息；若本地路径创建成功、且接收到失败的 P2P 路径创建响应消息，则沿所述相反路径转发更新后的、失败的 P2P 路径创建响应消息；若本地路径创建失败，且接收到失败或成功的 P2P 路径创建响应消息，则沿所述相反路径转发更新后的、失败的 P2P 路径创建响应消息。

15 优选地，所述网络控制器，还配置为作为首节点网络控制器接收到 P2P 路径创建响应消息后，若接收到成功的 P2P 路径创建响应消息、且本地路径创建成功时，确认所述路径创建成功；若接收到失败的 P2P 路径创建响应消息或者本地路径创建失败时，确认所述路径创建失败，并重新进行路径计算。

20 优选地，所述网络控制器，还配置为在接收到 P2P 路径创建响应消息后，若接收到为失败的 P2P 路径创建响应消息或者本地路径创建失败时，对本地已下发的转发流表进行回滚操作。

本发明实施例还提供了一种网络控制器，包括：第一路由控制协议模块、分布式处理模块、路由控制与下发模块和分布式路由协议模块，其中：

25 所述第一路由控制协议模块，配置为负责网络控制器与交换机之间路

由控制通道的建立以及网络控制器与交换机之间信息的交互；

所述分布式处理模块，配置为根据所述第一路由控制与下发模块提供的路径信息生成 P2P 路径创建请求消息，提供给所述分布式路由协议模块进行转发；还配置为根据所述路由控制与下发模块提供的本地路径创建结果生成 P2P 路径创建响应消息，提供给分布式路由协议模块进行转发；还配置为接收所述分布式路由协议模块提供的 P2P 路径创建请求消息，并将其中携带的转发流表提供给所述路由控制与下发模块；还配置为接收所述分布式路由协议模块提供的 P2P 路径创建响应消息，并根据所述路由控制与下发模块提供的本地路径创建结果对接收的 P2P 路径创建响应消息进行更新后，提供给所述分布式路由协议模块进行转发；还配置为根据路由控制与下发模块提供的本地路径创建结果和所述分布式路由协议模块提供的 P2P 路径创建响应消息，确定路径是否创建成功；

路由控制与下发模块，配置为在首节点网络控制器接收到路径创建请求时进行路径计算，得到路径信息，并提供给所述分布式处理模块；还配置为将所述分布式处理模块提供的转发流表通过所述第一路由控制协议模块下发给对应的交换机，完成本地路径创建，并将本地路径创建结果提供给所述分布式处理模块；

所述分布式路由协议模块，配置为沿路径转发或接收 P2P 路径创建请求消息；还配置为相反路径转发或接收所述 P2P 路径创建响应消息；还配置为将接收到的 P2P 路径创建请求消息或 P2P 路径创建响应消息提供给所述分布式处理模块。

优选地，所述分布式处理模块，还配置为确定路径创建失败时，通知所述路由控制与下发模块执行转发流表回滚操作；

所述路由控制与下发模块，还配置为通过所述路由控制协议模块，对 25 下发的转发流表执行回滚操作。

本发明实施例基于 DHT 的控制网络实现方法和系统：交换机与分布式对等网络中指定的网络控制器建立路由控制通道；接收到路径创建请求的网络控制器作为首节点网络控制器进行路径计算，生成 P2P 路径创建请求消息，并沿路径转发 P2P 路径创建请求消息；路径上接收到 P2P 路径创建请求消息的网络控制器，向自身管理的交换机下发 P2P 路径创建请求消息中携带的转发流表，完成各自的本地路径创建；其中，路径上的中间节点网络控制器在完成本地路径创建后，沿路径转发 P2P 路径创建请求消息，路径上的末节点网络控制器在完成本地路径创建后沿相反路径返回 P2P 路径创建响应消息；接收到 P2P 路径创建响应消息的中间节点网络控制器，结合本地路径创建结果，更新 P2P 路径创建响应消息，并沿相反路径转发更新后的 P2P 路径创建响应消息；首节点网络控制器接收到 P2P 路径创建响应消息后，结合本地路径创建结果，确定路径是否创建成功，使得网络路径的计算与下发容易管理。另外，交换机与分布式对等网络中指定的网络控制器建立路由控制通道，使得网络控制器网络实现自组织、交换机自动接入。

附图说明

图 1 为本发明实施例提供的基于 DHT 的 SDN 网络系统架构示意图；

图 2 为本发明实施例提供的基于 DHT 的控制网络实现方法；

图 3 为本发明实施例提供的交换机与指定的网络控制器建立路由控制通道的示意图；

图 4 为本发明实施例提供的网络控制器进行路径计算与转发的流程；

图 5 为本发明实施例提供的网络控制器处理 P2P 路径创建请求消息的流程图；

图 6 为本发明实施例提供的网络控制器处理 P2P 路径创建请求消息的流程图；

图 7 为本发明实施例中网络控制器的结构示意图；

图 8 为本发明实施例中交换机的结构示意图。

具体实施方式

下面结合附图和具体实施例对本发明的技术方案进一步详细阐述。

5 本发明实施例基于 DHT 的控制网络实现方案基于的网络架构如图 1 所示。图 1 是本发明实施例提供的基于 DHT 的 SDN 系统架构示意图，SDN 包括：交换机网络 101、分布式对等网络 103 和引导服务器 105，其中：

10 交换机网络 101，由多个物理上互连的交换机 102 组成，交换机 102 之间通过物理链路进行互联，每个交换机 102 根据 MAC 地址或 DPID 等信息通过哈希（HASH）算法、MID5 算法等生成固定长度的交换机标识（ID）。交换机 102 在上电时通过引导服务器 105 与指定的网络控制器 104 之间建立可靠路由控制通道，进行相关的安全验证、基本信息上报、交换机配置等操作。交换机 102 完全受网络控制器 104 控制，对于收到的数据报文，交换机 102 根据网络控制器 104 下发的转发流表进行处理，对于查找本地流表无命中数据报文，交换机 102 上报给网络控制器 104 进行处理。
15

20 分布式对等网络 103，是由至少一个网络控制器 104 组成的分布式 DHT 网络，如单跳（One-Hop）DHT、Pastry 网络等。网络控制器 104 之间基于分布式路由协议（如资源定位与发现（REsource LOcation And Discovery，RELOAD）协议）进行通信。其中，每个网络控制器 104 根据 IP 地址等信息通过哈希（HASH）算法、MID5 算法等生成固定长度的网络控制器标识（ID）。需要指出的是：网络控制器标识和交换机标识均为同一固定长度的字符串，例如，对交换机的 DPID 或 MAC 地址进行 HASH 或 MID5 运算，得到 128 位的字符串作为标识；对网络控制器的 IP 地址进行 HASH 或 MID5 运算，得到 128 位的字符串作为标识。每个网络控制器 104 负责管理一个或多个交换机 102（即一个或多个交换机 102 分别与一个网络控制器 104 建
25

立路由控制通道)，两者之间对应关系的建立可以通过分布式算法来确定相关的原则，例如，两者的标识距离最近的交换机 102 和网络控制器 104 建立路由控制通道，或者，两者的标识距离大于预定的阈值，且距离最小时，对应的交换机 102 和网络控制器 104 建立路由控制通道。当然，网络控制器 5 104 和交换机 102 之间对应关系的建立所依据的原则并不限于上述两种方式，只要通过分布式算法确定的均可应用。每个网络控制器 104 对所管理的交换机 102 上报的数据报文进行路由计算，并通过分布式路由协议向所管理的交换机 102 下发转发流表以完成数据交换。

引导服务器 105，负责引导交换机 102 与指定的网络控制器 104 建立路 10 由控制通道。引导服务器 105 可以是普通的 DNS 服务器，负责将交换机 102 引导到任意的网络控制器 104，交换机 102 再通过该网络控制器 104 返回的信息重定向到指定的网络控制器 104；引导服务器 105 还可以与分布式对等网络 103 进行交互，直接将分布式算法确定的网络控制器 104 通知给交换机 102，交换机 102 直接与该网络控制器 104 进行路由控制通道建立。

15 本发明实施例提供的基于 DHT 的控制网络实现方法如图 2 所示，包括如下步骤：

步骤 201，交换机与分布式对等网络中指定的网络控制器建立路由控制通道。

步骤 202，接收到路径创建请求的网络控制器作为首节点网络控制器进 20 行路径计算，生成 P2P 路径创建请求消息，并沿路径转发 P2P 路径创建请求消息。

步骤 203，路径上接收到 P2P 路径创建请求消息的网络控制器，向自身 25 管理的交换机下发 P2P 路径创建请求消息中携带的转发流表，完成各自的本地路径创建；其中，路径上的中间节点网络控制器在完成本地路径创建后，沿路径转发 P2P 路径创建请求消息，路径上的末节点网络控制器在完

成本地路径创建后沿相反路径返回 P2P 路径创建响应消息。

步骤 204，接收到 P2P 路径创建响应消息的中间节点网络控制器，结合本地路径创建结果，更新 P2P 路径创建响应消息，并沿相反路径转发更新后的 P2P 路径创建响应消息；首节点网络控制器接收到 P2P 路径创建响应
5 消息后，结合本地路径创建结果，确定路径是否创建成功。

下面结合附图及具体实施例对本发明上述的方案进行详细说明。

步骤 201 的实现方式包括：交换机与引导服务器直接返回的网络控制器列表中的网络控制器建立路由控制通道；或者，引导服务器将交换机引导到任意的网络控制器，交换机根据所述任意的网络控制器返回的信息重定向到指定的网络控制器、并与该指定的网络控制器建立路由控制通道。
10

图 3 是本发明实施例提供的交换机与指定的网络控制器建立路由控制通道的示意图。其中，图 3 (a) 描述的是在交换机加入网络过程中，交换机与引导服务器直接返回的网络控制器列表中的网络控制器建立路由控制通道的流程；图 3 (b) 描述的是引导服务器执行普通的 DNS 服务器功能，
15 负责将交换机引导到任意的网络控制器，交换机根据任意的网络控制器返回的信息重定向到指定的网络控制器、并与该指定的网络控制器建立路由控制通道的流程。

优选地，本发明实施例所述的路由控制通道可以为 OpenFlow 协议通道、简单网络管理协议（SNMP，Simple Network Management Protocol）通道或 DIAMETER 通道。
20

如图 3 (a) 所示，交换机加入网络的具体步骤描述如下：

步骤 301a、交换机向引导服务器请求与网络控制器建立连接。

步骤 302a、引导服务器根据分布式算法选择负责管理该交换机的一个网络控制器，并将该网络控制器的信息返回给交换机。

25 这里，引导服务器还可根据分布式算法选择负责管理该交换机的一个

或多个备份网络控制器，并将备份网络控制器的信息也返回给交换机。

优选地，网络控制器的信息可以以列表的形式发送给交换机。

该步骤中，选择负责管理该交换机的网络控制器即为上述的：依据网络控制器和交换机之间对应关系的建立原则，选择可与当前交换机建立对
5 应关系的网络控制器。

步骤 303a)、交换机根据引导服务器返回的网络控制器列表，与网络控制器列表中的一个或多个网络控制器建立路由控制通道。

引导服务器在返回网络控制器列表时，一般会将根据分布式算法确定的最合适
10 的网络控制器排列在第一位，将其它备份网络控制器依次向后排列。优选地，交换机可以按照排列顺序优先与排列最靠前的一个网络控制器建立路
制通道。如果与排列在第一位的网络控制器不能建立路由控制通道，则尝试与第二位的网络控制器建立路由控制通道，依次类推。

优选地，根据分布式算法，交换机还可以与列表中的多个网络控制器建立路
制通道，具体如何从列表中选择，通过分布式算法确定。

15 步骤 304a)、网络控制器与交换机建立路由控制通道，将交换机信息及
交换机的网络邻居拓扑关系在分布式对等网络中进行更新。

如图 3 (b) 所示，交换机加入网络的具体步骤描述如下：

步骤 301b、交换机向引导服务器请求与网络控制器建立连接。

步骤 302b、引导服务器执行 DNS 重定向功能，返回一个网络控制器列

20 表 1 给交换机。

步骤 303b、交换机从网络控制器列表 1 中任意选择一个网络控制器 Ci
发送 OpenFlow 协议通道建立请求。

步骤 304b、网络控制器 Ci 根据分布式算法选择负责管理该交换机的一
个或多个网络控制器，生成网络控制器列表 2。（该步骤选择网络控制器的
25 实现与步骤 202a 相同）

步骤 305b、网络控制器 Ci 向交换机返回重定向响应，指示交换机与网络控制器列表 2 中的网络控制器建立 OpenFlow 协议通道。

步骤 306b、交换机根据重定向响应，与网络控制器列表 2 中的一个或多个网络控制器建立 OpenFlow 协议通道。该实施例中假设交换机与网络控制器列表 2 中网络控制器 Cj 建立 OpenFlow 协议通道。
5

步骤 307b、网络控制器 Cj 与交换机建立 OpenFlow 协议通道，将交换机信息及交换机的网络邻居拓扑关系在分布式对等网络中进行更新。

上述步骤 202 的实现方式包括：通过路径计算可得到以下的路径信息：
路径上网络控制器的顺序和路径上交换机的顺序。进行路径计算的网络控制器（首节点网络控制器）生成 P2P 路径创建请求消息，该消息的目的节点为路径上的最后一个网络控制器，即末节点网络控制器；该消息的路由路径头部由路径上除了末节点网络控制器之外的其它网络控制器的标识组成，优选地，其它网络控制器的标识按路径上的其它网络控制器的顺序排列；该消息的消息体由路径上向各个交换机下发的转发流表组成，优选地，
10 向各个交换机下发的转发流表按照路径上的交换机的顺序排列。在转发时，
路径头部由路径上除了末节点网络控制器之外的其它网络控制器的标识组成，优选地，其它网络控制器的标识按路径上的其它网络控制器的顺序排列；该消息的消息体由路径上向各个交换机下发的转发流表组成，优选地，
15 向各个交换机下发的转发流表按照路径上的交换机的顺序排列。在转发时，
按照 P2P 路径创建请求消息的路由路径头部，向下一个网络控制器转发 P2P
路径创建请求消息。

图 4 是本发明实施例提供的网络控制器进行路径计算与转发的流程。
路径计算节点 C1 根据网络拓扑信息和主机位置信息计算得到路由路径，根
据分布式算法获取路径上各交换机标识和各网络控制器标识生成 P2P 路径
20 创建请求消息，进行路由路径的转发。具体步骤描述如下：

步骤 401、网络控制器 C1 收到路径 {src, dst} 创建请求，计算路径 {src, dst}
上的交换机顺序和网络控制器顺序，假设交换机顺序为 {S1, S2, S3, S4, S6,
S7, S9}，网络控制器顺序为 {C1, C2, C3 }；

步骤 402、网络控制器 C1 根据分布式算法获取路径 {src, dst} 上的交换

机标识序列 {H(S1), H(S1), ..., H(S9)}, 并获取路径 {src, dst} 上的网络控制器标识序列 {Ctl_ID1, Ctl_ID2, Ctl_ID3};

步骤 403、网络控制器 C1 生成 P2P 路径创建请求消息，该消息的路由路径头部为 {Ctl_ID1, Ctl_ID2}; 该消息的目的节点为 Ctl_ID3; 该消息的消息体为路径 {src, dst} 上向各个交换机下发的转发流表 {S1:FlowEntry list, S2:FlowEntry list, ..., S9:FlowEntry list};

步骤 404、网络控制器 C1 根据路由路径头部 {Ctl_ID1, Ctl_ID2}, 向网络控制器 C2 转发 P2P 路径创建请求消息；网络控制器 C2 根据该消息的目的节点向网络控制器 C3 转发 P2P 路径创建请求消息。

上述步骤 203 的实现方式为：路径上接收到 P2P 路径创建请求消息的网络控制器首选向自身管理的交换机下发 P2P 路径创建请求消息中携带的转发流表，完成本地路径创建；然后，该网络控制器根据 P2P 路径创建请求消息的目的节点判断本节点是否为末节点网络控制器，如果否，按照路由路径头部将 P2P 路径创建请求消息转发给下一个网络控制器；如果是，即末节点网络控制器沿相反路径向上一个网络控制器返回 P2P 路径创建响应消息。

承接图 4 所示的实施例，图 5 是本发明实施例提供的网络控制器处理 P2P 路径创建请求消息的流程图。网络控制器收到 P2P 路径创建请求消息后，对于需要本节点负责的转发流表下发，并根据向下一节点继续转发 P2P 路径创建请求消息。具体步骤描述如下：

步骤 501，网络控制器 C2 收到网络控制器 C1 转发的 P2P 路径创建请求消息；

步骤 502，网络控制器 C2 首先向所负责的交换机下发对应的转发流表，如网络控制器 C2 向交换机 S2 下发转发流表 S2:FlowEntry list, 向 S7 下发转发流表 S7:FlowEntry list, 完成本地路径创建；

步骤 503，完成本地路径创建后，网络控制器 C2 根据 P2P 路径创建请求消息的目的节点判断本节点是否为路径{src, dst}上的末节点网络控制器；

步骤 504，由于网络控制器 C2 不是路径{src, dst}上的末节点网络控制器，因此，网络控制器 C2 根据目的节点向网络控制器 C3 转发 P2P 路径创
5 建请求消息；

步骤 505，网络控制器 C3 首先向所负责的交换机下发对应的转发流表，如网络控制器 C3 向交换机 S3 下发转发流表 S3:FlowEntry list，向交换机 S4 下发转发流表 S4:FlowEntry list，完成本地路径创建；

步骤 506，完成本地路径创建后，网络控制器 C3 根据 P2P 路径创建请
10 求消息的目的节点判断本节点是否为路径{src, dst}上的末节点网络控制器；

步骤 507，由于网络控制器 C3 是路径{src, dst}上的末节点网络控制器，因此，网络控制器 C3 向上一个节点、即网络控制器 C2 返回 P2P 路径创建响应消息。

进一步地，在上述步骤 203 的实现方式中：

对于路径上的末节点网络控制器，在完成本地路径创建后，若本地路
15 径创建成功，则向上一个网络控制器返回成功的 P2P 路径创建响应消息；如果本地路径创建失败，向上一个网络控制器返回失败的 P2P 路径创建响
应消息。

对于步骤 204 的实现：对路径上的中间节点网络控制器，接收到 P2P
20 路径创建响应消息时，根据本地路径创建结果，若本地路径创建成功、且接收到了成功的 P2P 路径创建响应消息，则沿相反路径转发更新后的成功的 P2P 路径创建响应消息；若本地路径创建成功、且接收到失败的 P2P 路径创建响应消息，则沿相反路径转发更新后的、失败的 P2P 路径创建响应消息；若本地路径创建失败，且接收到失败或成功的 P2P 路径创建响应消息，则沿相反路径转发更新后的、失败的 P2P 路径创建响应消息。
25

对于首节点网络控制器，接收到成功的 P2P 路径创建响应消息、且本地路径创建成功时，确认路径创建成功；接收到失败的 P2P 路径创建响应消息或者本地路径创建失败时，确认路径创建失败，并重新进行路径计算。

承接图 4、图 5 的实施例，图 6 是本发明实施例提供的网络控制器处理 5 P2P 路径创建请求消息的流程图。具体步骤描述如下：

步骤 601，网络控制器 C2 收到路径 {src, dst} 上的末节点网络控制器 C3 发送的失败的 P2P 路径创建响应消息；

步骤 602，因为网络控制器 C2 接收到的路径创建响应消息为失败的 P2P 路径创建响应消息，且网络控制器 C2 不是生成 P2P 路径创建请求消息的节点，所以，不管网络控制器 C2 本节点的路径创建是否成功，网络控制器 10 C2 都向网络控制器 C1 返回失败的 P2P 路径创建响应消息，同时，网络控制器 C2 对本地已经下发的转发流表进行回滚操作；

步骤 603，网络控制器 C1 接收到失败的 P2P 路径创建响应消息，因为 15 网络控制器 C1 为生成 P2P 路径创建请求消息的节点，所以网络控制器 C1 确定路径创建失败，则需要重新进行路径计算，同时，网络控制器 C1 对本地已经下发的转发流表进行回滚操作。

步骤 604，网络控制器 C2 收到路径 {src, dst} 上的末节点网络控制器 C3 发送的成功的 P2P 路径创建响应消息；

步骤 605，网络控制器 C2 不是生成 P2P 路径创建请求消息的节点，网 20 络控制器 C2 判断本地路径创建是否成功，若成功，执行步骤 606；若失败，执行步骤 607。

步骤 606，网络控制器 C2 向上一个网络控制器 C1 返回成功的 P2P 路径创建响应消息，后续的流程包括：

步骤 6061，网络控制器 C1 是生成 P2P 路径创建请求消息的节点，在 25 接收到成功的 P2P 路径创建响应消息时，判断本地路径创建是否成功，若

成功，执行步骤 6062，路径创建成功；若失败，执行步骤 6063，路径创建失败，需要重新进行路径计算，同时，对本地已经下发的转发流表进行回滚操作。

步骤 607，网络控制器 C2 向上一个网络控制器 C1 返回失败的 P2P 路径创建响应消息，同时，对本地已经下发的转发流表进行回滚操作，后续的流程包括：

步骤 6071，因为网络控制器 C1 是生成 P2P 路径创建请求消息的节点即首节点，在接收到失败的 P2P 路径创建响应消息时，不论本地路径创建是否成功，均确定为路径创建失败，需要重新进行路径计算，同时对本地已经下发的转发流表进行回滚操作。

为了实现上述方法，图 1 中所示各个部分执行的操作为：

交换机，配置为与分布式对等网络中指定的网络控制器建立路由控制通道；

网络控制器，配置为接收到路径创建请求时作为首节点网络控制器进行路径计算，生成 P2P 路径创建请求消息，并沿路径转发 P2P 路径创建请求消息；还配置为接收到 P2P 路径创建请求消息时，向自身管理的交换机下发 P2P 路径创建请求消息中携带的转发流表，完成本地路径创建；还配置为作为路径上的中间节点网络控制器在完成本地路径创建后，沿路径转发 P2P 路径创建请求消息；还配置为，作为路径上的末节点网络控制器在完成本地路径创建后沿相反路径返回 P2P 路径创建响应消息；还配置为作为路径上的中间节点网络控制器接收到 P2P 路径创建响应消息时，结合本地路径创建结果，更新 P2P 路径创建响应消息，并沿相反路径转发更新后的 P2P 路径创建响应消息；还配置为作为首节点网络控制器接收到 P2P 路径创建响应消息后，结合本地路径创建结果，确定路径是否创建成功。

引导服务器，配置为向交换机返回指定的网络控制器列表；或者，将

交换机引导到任意的网络控制器；

相应地，交换机，还配置为与引导服务器直接返回的指定的网络控制器列表中的网络控制器建立路由控制通道；或者，根据任意的网络控制器返回的信息重定向到指定的网络控制器、并与该指定的网络控制器建立路
5 由控制通道。

网络控制器，还配置为作为首节点网络控制器生成 P2P 路径创建请求消息后，向自身管理的交换机下发 P2P 路径创建请求消息中携带的转发流表。

网络控制器，还配置为作为末节点网络控制器，若本地路径创建成功，
10 沿相反路径返回成功的 P2P 路径创建响应消息；若本地路径创建失败，沿相反路径返回失败的 P2P 路径创建响应消息。

网络控制器，还配置为作为中间节点网络控制器接收到 P2P 路径创建响应消息时，若本地路径创建成功、且接收到成功的 P2P 路径创建响应消息，则沿相反路径转发更新后的、成功的 P2P 路径创建响应消息；若本地
15 路径创建成功、且接收到失败的 P2P 路径创建响应消息，则沿相反路径转发更新后的、失败的 P2P 路径创建响应消息；若本地路径创建失败，且接收到失败或成功的 P2P 路径创建响应消息，则沿相反路径转发更新后的、失败的 P2P 路径创建响应消息。

网络控制器，还配置为作为首节点网络控制器接收到 P2P 路径创建响应消息后，若接收到成功的 P2P 路径创建响应消息、且本地路径创建成功时，确认路径创建成功；若接收到失败的 P2P 路径创建响应消息或者本地
20 路径创建失败时，确认路径创建失败，并重新进行路径计算。

网络控制器，还配置为在接收到 P2P 路径创建响应消息后，若接收到失败的 P2P 路径创建响应消息或者本地路径创建失败时，对本地已下发的
25 转发流表进行回滚操作。

具体地，本发明实施例中网络控制器的结构如图 7 所示，主要功能模块包括第一路由控制协议模块 701、路由控制与下发模块 702、分布式处理模块 703、分布式路由协议模块 704 和信息存储模块 705，其中：

第一路由控制协议模块 701，主要负责网络控制器与交换机之间的路由控制通道的建立与维护，以及网络控制器和交换机之间的信息传递。例如：在处理交换机的路由控制通道的建立请求时，若交换机标识与网络控制器标识的对应关系不符合分布式算法时，则通过路由控制与下发模块 702 向分布式处理模块 703 查询管理所述交换机的网络控制器列表，并通过重定向响应将该网络控制器列表返回给交换机。网络控制器和交换机之间的交互信息包括：通信安全验证、交换机信息上报、交换机基本配置功能、交换机数据报文上报、交换机流表的下发与管理等。

路由控制与下发模块 702，负责执行路径计算和转发流表的下发、回滚等操作。对网络拓扑变化事件或用户数据的转发请求，路由控制与下发模块 702 根据信息存储模块 705 中的网络拓扑和/或主机位置信息，进行路径计算，并与分布式处理模块 703 交互生成 P2P 路径创建请求消息。对来自分布式处理模块 703 的 P2P 路径创建请求消息，路由控制与下发模块 702 向本节点负责的交换机通过第一路由控制协议模块 701 下发相关转发流表。对来自分布式处理模块 703 的 P2P 路径创建响应消息，若为失败，路由控制与下发模块 702 向本节点负责的交换机通过第一路由控制协议模块 701 删除下发的转发流表、即回滚操作。具体地：路由控制与下发模块 702 在首节点网络控制器接收到路径创建请求时进行路径计算，得到路径信息，并提供给分布式处理模块 703；还配置为将分布式处理模块 703 提供的转发流表通过第一路由控制协议模块 701 下发给对应的交换机，完成本地路径创建，并将本地路径创建结果提供给分布式处理模块 703。

分布式处理模块 703，负责分布式对等网络的业务处理，如 P2P 路径创

建请求消息的处理。当进行路径创建时，根据路由控制与下发模块 702 提供的路径信息，生成 P2P 路径创建请求消息；处理 P2P 路径创建请求消息时，将转发流表提供给路由控制与下发模块 702 进行下发；处理 P2P 路径创建响应消息时，对于失败响应，通知路由控制与下发模块 702 将对下发的转发流表执行回滚操作。具体地：分布式处理模块 703 根据路由控制与下发模块 702 提供的路径信息生成 P2P 路径创建请求消息，提供给分布式路由协议模块 704 进行转发；还配置为根据路由控制与下发模块 702 提供的本地路径创建结果生成 P2P 路径创建响应消息，提供给分布式路由协议模块 704 进行转发；还配置为接收分布式路由协议模块 704 提供的 P2P 路径创建请求消息，并将其中携带的转发流表提供给路由控制与下发模块 703；还配置为接收分布式路由协议模块 704 提供的 P2P 路径创建响应消息，并根据路由控制与下发模块 702 提供的本地路径创建结果对接收的 P2P 路径创建响应消息进行更新后，提供给分布式路由协议模块 704 进行转发；还配置为根据路由控制与下发模块 702 提供的本地路径创建结果和分布式路由协议模块 704 提供的 P2P 路径创建响应消息，确定路径是否创建成功。

分布式路由协议模块 704，负责与其它网络控制器进行消息交互，包括分布式路由表的维护、P2P 路径创建请求消息和 P2P 路径创建响应消息的路由等。

信息存储模块 705，负责网络控制器上相关数据的存储，主要包括本节点管理的交换机状态及对应的转发流表，以及根据分布式算法本节点负责管理的主机位置数据、全局网络拓扑数据的存储等。

实际应用时，第一路由控制协议模块、路由控制与下发模块、分布式处理模块、以及分布式路由协议模块可由网络控制器中的中央处理器 (CPU, Central Processing Unit)、数字信号处理器 (DSP, Digital Singnal Processor) 或可编程逻辑阵列 (FPGA, Field – Programmable Gate Array)

实现；信息存储模块可由网络控制器中的存储器实现。

具体地，本发明实施例中交换机的结构如图 8 所示，主要功能模块包括第二路由控制协议模块 801、交换机控制模块 802 和报文转发功能模块 803。其中，

5 第二路由控制协议模块 801，负责网络控制器与交换机之间的路由控制通道的建立与维护，以及网络控制器和交换机之间的信息传递。在向网络控制器发送路由控制通道的建立请求时，对于网络控制器返回的重定向响应，可以向新的网络控制器重新发起路由控制通道建立请求。网络控制器和交换机之间的交互信息包括：通信安全验证、交换机信息上报、交换机基本配置功能、交换机数据报文上报、交换机流表的下发与管理等。

10 交换机控制模块 802，负责将相关的转发面数据报文通过第二路由控制协议模块 801 上报给网络控制器，将网络控制器下发的转发流表映射为报文转发功能模块 803 的硬件要求的格式，下发到报文转发功能模块 803。

15 报文转发功能模块 803，负责维护硬件上的转发流表，并对数据报文进行匹配处理与转发。

实际应用时，第二路由控制协议模块、交换机控制模块以及报文转发功能模块可由交换机中的 CPU、DSP 或 FPGA 实现。

以上所述，仅为本发明的较佳实施例而已，并非用于限定本发明的保护范围。

权利要求书

1、一种基于 DHT 的控制网络实现方法，所述方法包括：

交换机与分布式对等网络中指定的网络控制器建立路由控制通道；

接收到路径创建请求的网络控制器作为首节点网络控制器进行路径计

5 算，生成 P2P 路径创建请求消息，并沿路径转发所述 P2P 路径创建请求消息；

所述路径上接收到所述 P2P 路径创建请求消息的网络控制器，向自身
管理的交换机下发所述 P2P 路径创建请求消息中携带的转发流表，完成各
自的本地路径创建；其中，所述路径上的中间节点网络控制器在完成本地
10 路径创建后，沿所述路径转发所述 P2P 路径创建请求消息，所述路径上的
末节点网络控制器在完成本地路径创建后沿相反路径返回 P2P 路径创建响
应消息；

接收到所述 P2P 路径创建响应消息的中间节点网络控制器，结合本地
路径创建结果，更新所述 P2P 路径创建响应消息，并沿所述相反路径转发
15 更新后的 P2P 路径创建响应消息；首节点网络控制器接收到 P2P 路径创建
响应消息后，结合本地路径创建结果，确定所述路径是否创建成功。

2、根据权利要求 1 所述基于 DHT 的控制网络实现方法，其中，所述
交换机与分布式对等网络中指定的网络控制器建立路由控制通道，包括：

20 交换机与引导服务器直接返回的指定的网络控制器列表中的网络控
制器建立路由控制通道；或者，引导服务器将交换机引导到任意的网络控
制器，交换机根据所述任意的网络控制器返回的信息重定向到指定的网络控
制器、并与所述指定的网络控制器建立路由控制通道。

3、根据权利要求 2 所述基于 DHT 的控制网络实现方法，其中，所述
路由控制通道为 OpenFlow 协议通道、SNMP 通道或 DIAMETER 通道。

25 4、根据权利要求 1 所述基于 DHT 的控制网络实现方法，其中，所述

路径计算得到路径上网络控制器的顺序和路径上交换机的顺序。

5、根据权利要求 4 所述基于 DHT 的控制网络实现方法，其中，

所述 P2P 路径创建请求消息的目的节点为所述路径上的末节点网络控制器；

5 所述 P2P 路径创建请求消息的路由路径头部由所述路径上除了末节点网络控制器之外的其它网络控制器的标识组成；

所述 P2P 路径创建请求消息的消息体由所述路径上向各个交换机下发的转发流表组成。

6、根据权利要求 5 所述基于 DHT 的控制网络实现方法，其中，

10 所述其它网络控制器的标识按所述路径上的所述其它网络控制器的顺序排列；

所述向各个交换机下发的转发流表按照所述路径上的交换机的顺序排列。

7、根据权利要求 1 至 6 任一所述基于 DHT 的控制网络实现方法，其

15 中，所述首节点网络控制器生成 P2P 路径创建请求消息后，所述方法还包括：所述首节点网络控制器向自身管理的交换机下发所述 P2P 路径创建请求消息中携带的转发流表。

8、根据权利要求 7 所述基于 DHT 的控制网络实现方法，其中，所述

路径上的末节点网络控制器在完成本地路径创建后沿相反路径返回 P2P 路径创建响应消息，为：

若本地路径创建成功，所述末节点网络控制器返回成功的 P2P 路径创建响应消息；若本地路径创建失败，所述末节点网络控制器返回失败的 P2P 路径创建响应消息。

9、根据权利要求 8 所述基于 DHT 的控制网络实现方法，其中，所述

25 接收到所述 P2P 路径创建响应消息的中间节点网络控制器，结合本地路径

创建结果，更新所述 P2P 路径创建响应消息，并沿所述相反路径转发更新后的 P2P 路径创建响应消息，为：

若本地路径创建成功、且接收到成功的 P2P 路径创建响应消息，则沿所述相反路径转发更新后的、成功的 P2P 路径创建响应消息；

5 若本地路径创建成功、且接收到失败的 P2P 路径创建响应消息，则沿所述相反路径转发更新后的、失败的 P2P 路径创建响应消息；

若本地路径创建失败，且接收到失败或成功的 P2P 路径创建响应消息，则沿所述相反路径转发更新后的、失败的 P2P 路径创建响应消息。

10 10、根据权利要求 9 所述基于 DHT 的控制网络实现方法，其中，所述首节点网络控制器接收到 P2P 路径创建响应消息后，结合本地路径创建结果，确定所述路径是否创建成功，为：

所述首节点网络控制器接收到成功的 P2P 路径创建响应消息、且本地路径创建成功时，确认所述路径创建成功；

15 所述首节点网络控制器接收到失败的 P2P 路径创建响应消息或者本地路径创建失败时，确认所述路径创建失败，并重新进行路径计算。

11、根据权利要求 10 所述基于 DHT 的控制网络实现方法，其中，所述路径上的网络控制器接收到 P2P 路径创建响应消息后，所述方法还包括：若接收到为失败的 P2P 路径创建响应消息或者本地路径创建失败时，网络控制器对本地已下发的转发流表进行回滚操作。

20 12、一种基于 DHT 的控制网络实现系统，所述系统包括：交换机和网络控制器，其中：

所述交换机，配置为与分布式对等网络中指定的网络控制器建立路由控制通道；

25 所述网络控制器，配置为接收到路径创建请求时作为首节点网络控制器进行路径计算，生成 P2P 路径创建请求消息，并沿路径转发所述 P2P 路

径创建请求消息；还配置为接收到所述 P2P 路径创建请求消息时，向自身管理的交换机下发所述 P2P 路径创建请求消息中携带的转发流表，完成本地路径创建；还配置为作为所述路径上的中间节点网络控制器在完成本地路径创建后，沿所述路径转发所述 P2P 路径创建请求消息；还配置为作为所述路径上的末节点网络控制器在完成本地路径创建后沿相反路径返回 P2P 路径创建响应消息；还配置为作为路径上的中间节点网络控制器接收到所述 P2P 路径创建响应消息时，结合本地路径创建结果，更新所述 P2P 路径创建响应消息，并沿所述相反路径转发更新后的 P2P 路径创建响应消息；还配置为作为首节点网络控制器接收到 P2P 路径创建响应消息后，结合本地路径创建结果，确定所述路径是否创建成功。

13、根据权利要求 12 所述基于 DHT 的控制网络实现系统，其中，所述系统还包括引导服务器，配置为向所述交换机返回指定的网络控制器列表；或者，将所述交换机引导到任意的网络控制器；

相应地，所述交换机，还配置为与所述引导服务器直接返回的指定的网络控制器列表中的网络控制器建立路由控制通道；或者，根据所述任意的网络控制器返回的信息重定向到指定的网络控制器、并与所述指定的网络控制器建立路由控制通道。

14、根据权利要求 12 或 13 所述基于 DHT 的控制网络实现系统，其中，所述网络控制器，还配置为作为首节点网络控制器生成 P2P 路径创建请求消息后，向自身管理的交换机下发所述 P2P 路径创建请求消息中携带的转发流表。

15、根据权利要求 14 所述基于 DHT 的控制网络实现系统，其中，所述网络控制器，还配置为作为末节点网络控制器，若本地路径创建成功，沿相反路径返回成功的 P2P 路径创建响应消息；若本地路径创建失败，沿相反路径返回失败的 P2P 路径创建响应消息。

16、根据权利要求 15 所述基于 DHT 的控制网络实现系统，其中，
所述网络控制器，还配置为作为中间节点网络控制器接收到所述 P2P
路径创建响应消息时，若本地路径创建成功、且接收到成功的 P2P 路径创
建响应消息，则沿所述相反路径转发更新后的、成功的 P2P 路径创建响应
5 消息；若本地路径创建成功、且接收到失败的 P2P 路径创建响应消息，则
沿所述相反路径转发更新后的、失败的 P2P 路径创建响应消息；若本地路
径创建失败，且接收到失败或成功的 P2P 路径创建响应消息，则沿所述相
反路径转发更新后的、失败的 P2P 路径创建响应消息。

17、根据权利要求 16 所述基于 DHT 的控制网络实现系统，其中，
10 所述网络控制器，还配置为作为首节点网络控制器接收到 P2P 路径创
建响应消息后，若接收到成功的 P2P 路径创建响应消息、且本地路径创建
成功时，确认所述路径创建成功；若接收到失败的 P2P 路径创建响应消息
或者本地路径创建失败时，确认所述路径创建失败，并重新进行路径计算。

18、根据权利要求 17 所述基于 DHT 的控制网络实现系统，其中，
15 所述网络控制器，还配置为在接收到 P2P 路径创建响应消息后，若接
收到为失败的 P2P 路径创建响应消息或者本地路径创建失败时，对本地已
下发的转发流表进行回滚操作。

19、一种网络控制器，包括：第一路由控制协议模块、分布式处理模
块、路由控制与下发模块和分布式路由协议模块，其中：

20 所述第一路由控制协议模块，配置为负责网络控制器与交换机之间路
由控制通道的建立以及网络控制器与交换机之间信息的交互；

25 所述分布式处理模块，配置为根据所述第一路由控制与下发模块提供
的路径信息生成 P2P 路径创建请求消息，提供给所述分布式路由协议模块
进行转发；还配置为根据所述路由控制与下发模块提供的本地路径创建结
果生成 P2P 路径创建响应消息，提供给分布式路由协议模块进行转发；还

配置为接收所述分布式路由协议模块提供的 P2P 路径创建请求消息，并将其中携带的转发流表提供给所述路由控制与下发模块；还配置为接收所述分布式路由协议模块提供的 P2P 路径创建响应消息，并根据所述路由控制与下发模块提供的本地路径创建结果对接收的 P2P 路径创建响应消息进行
5 更新后，提供给所述分布式路由协议模块进行转发；还配置为根据路由控制与下发模块提供的本地路径创建结果和所述分布式路由协议模块提供的 P2P 路径创建响应消息，确定路径是否创建成功；

路由控制与下发模块，配置为在首节点网络控制器接收到路径创建请求时进行路径计算，得到路径信息，并提供给所述分布式处理模块；还配置为将所述分布式处理模块提供的转发流表通过所述第一路由控制协议模块下发给对应的交换机，完成本地路径创建，并将本地路径创建结果提供给所述分布式处理模块；
10

所述分布式路由协议模块，配置为沿路径转发或接收 P2P 路径创建请求消息；还配置为相反路径转发或接收所述 P2P 路径创建响应消息；还配置为将接收到的 P2P 路径创建请求消息或 P2P 路径创建响应消息提供给所述分布式处理模块。
15

20、根据权利要求 19 所述网络控制器，其中，

所述分布式处理模块，还配置为确定路径创建失败时，通知所述路由控制与下发模块执行转发流表回滚操作；

所述路由控制与下发模块，还配置为通过所述路由控制协议模块，对
20 下发的转发流表执行回滚操作。

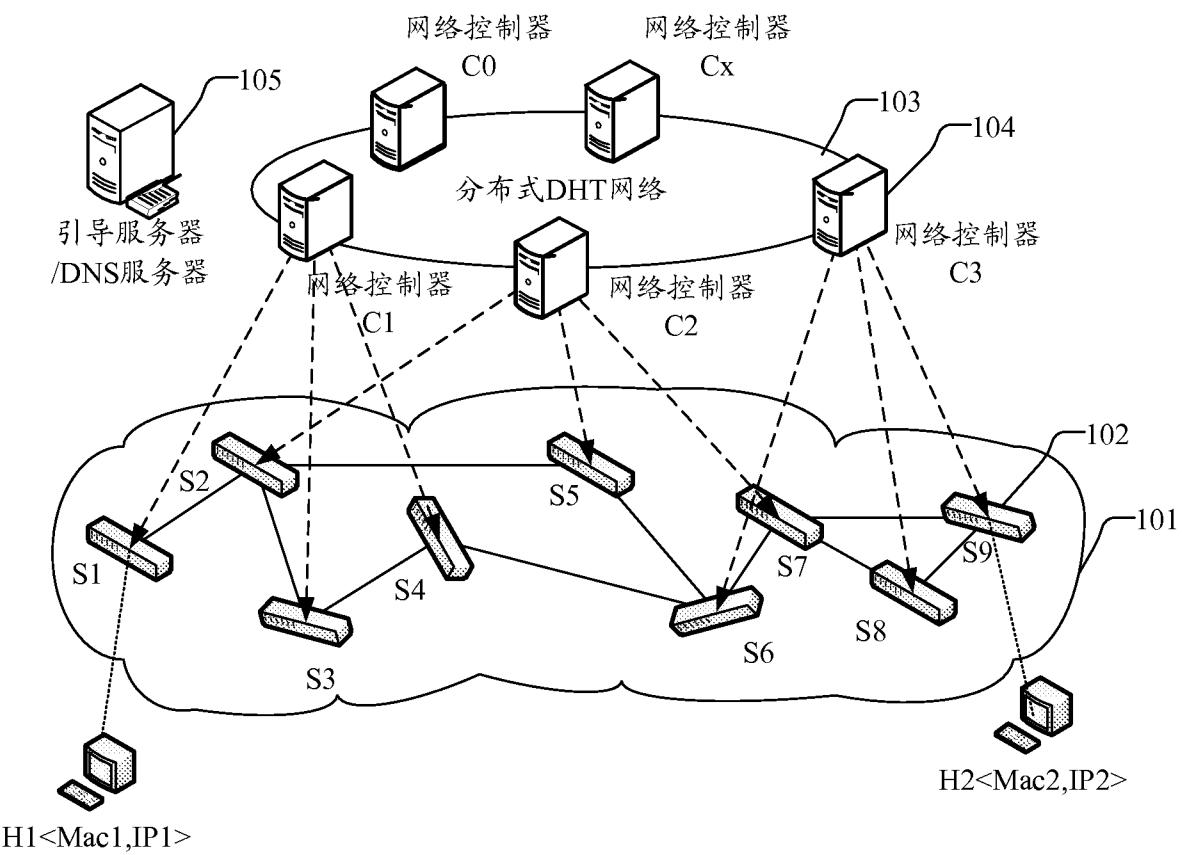


图 1

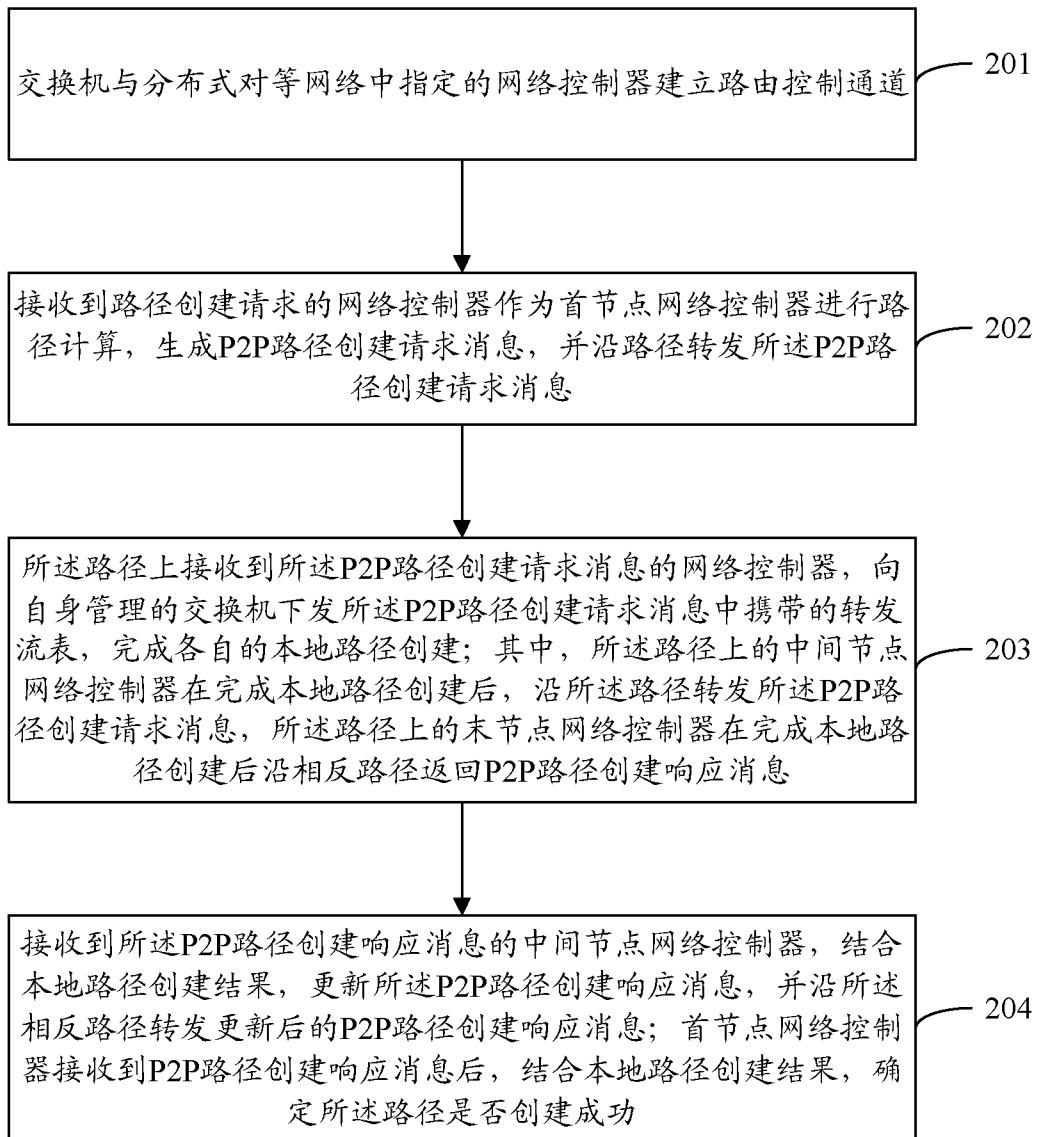
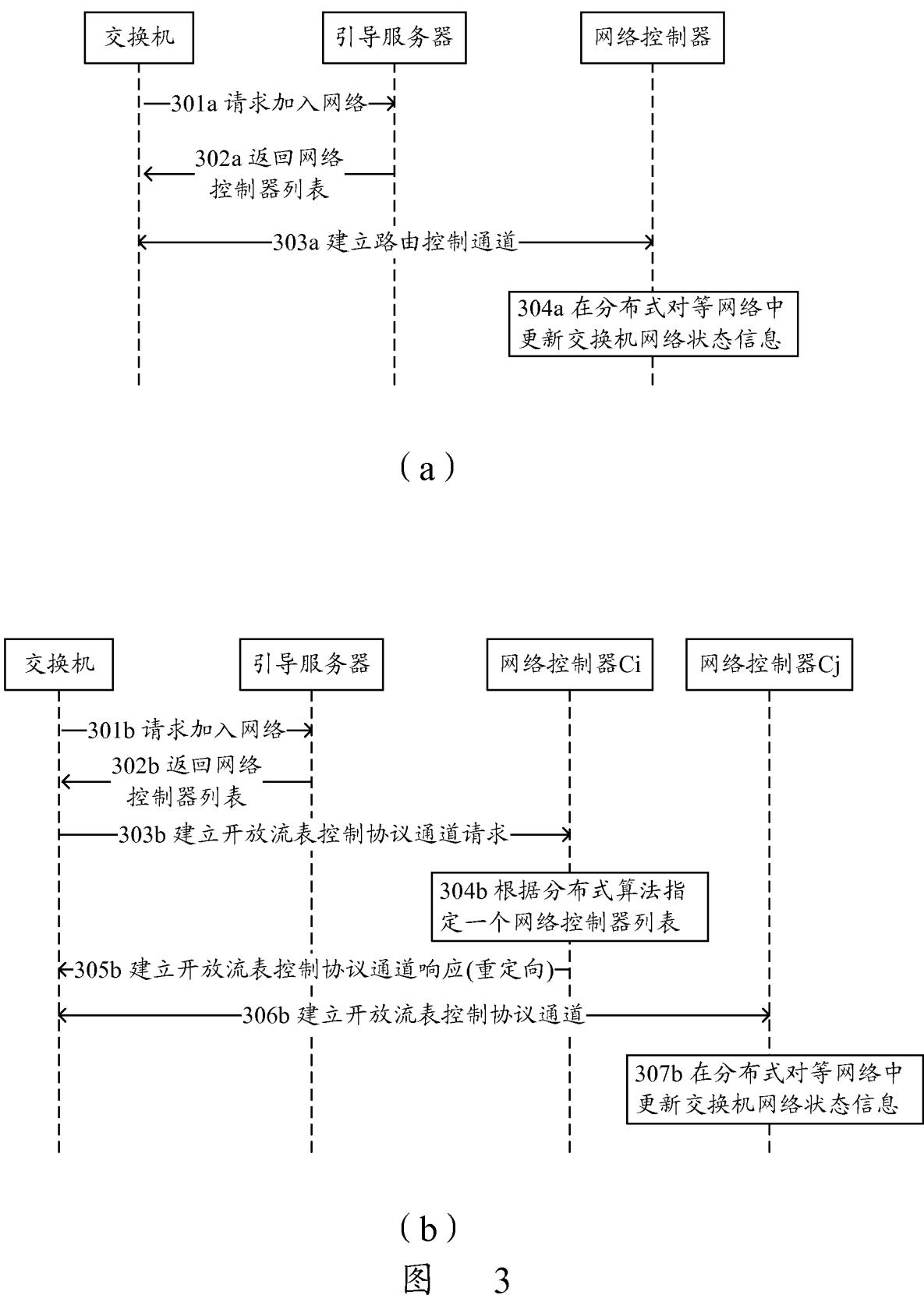


图 2



(b)

图 3

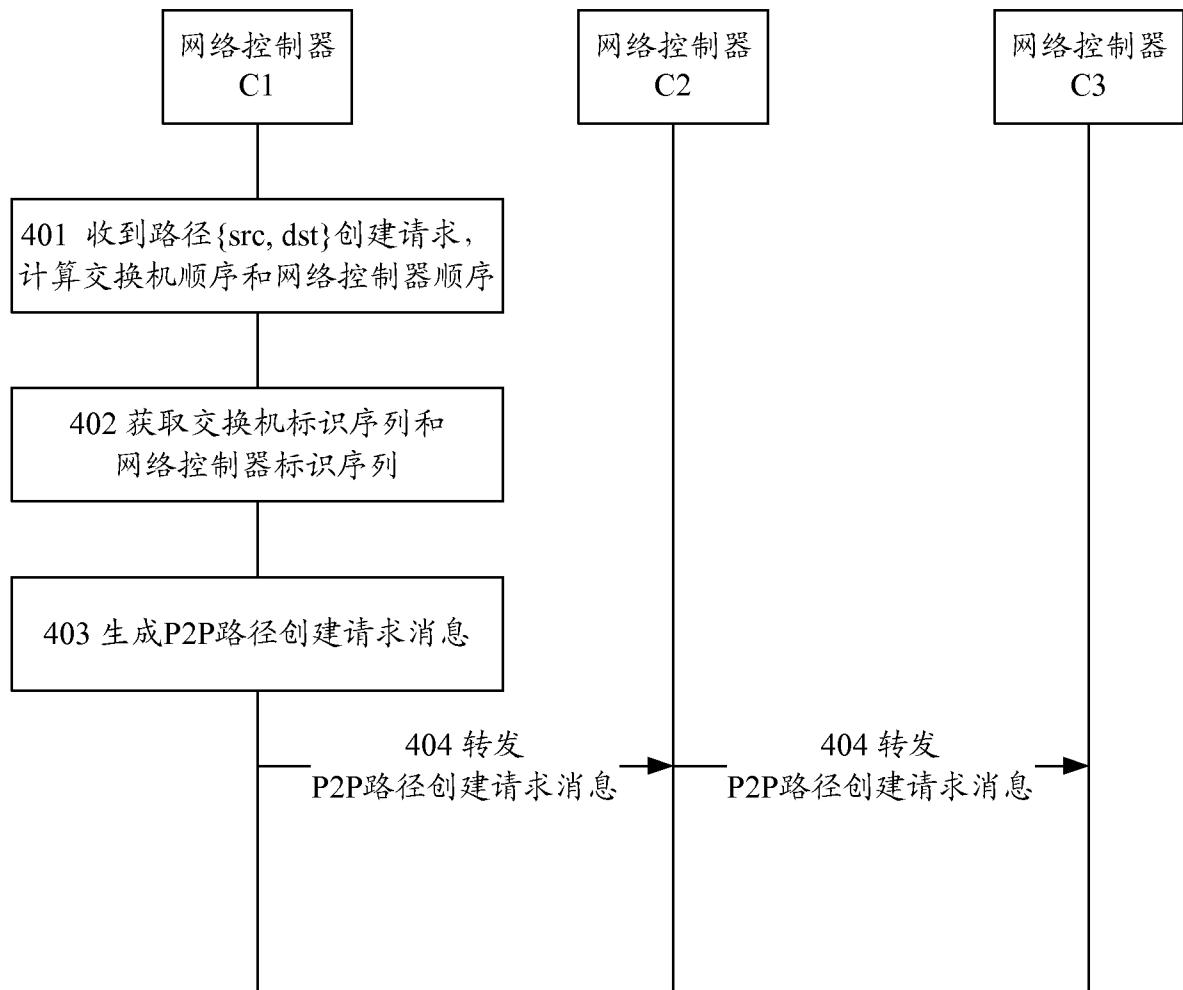


图 4

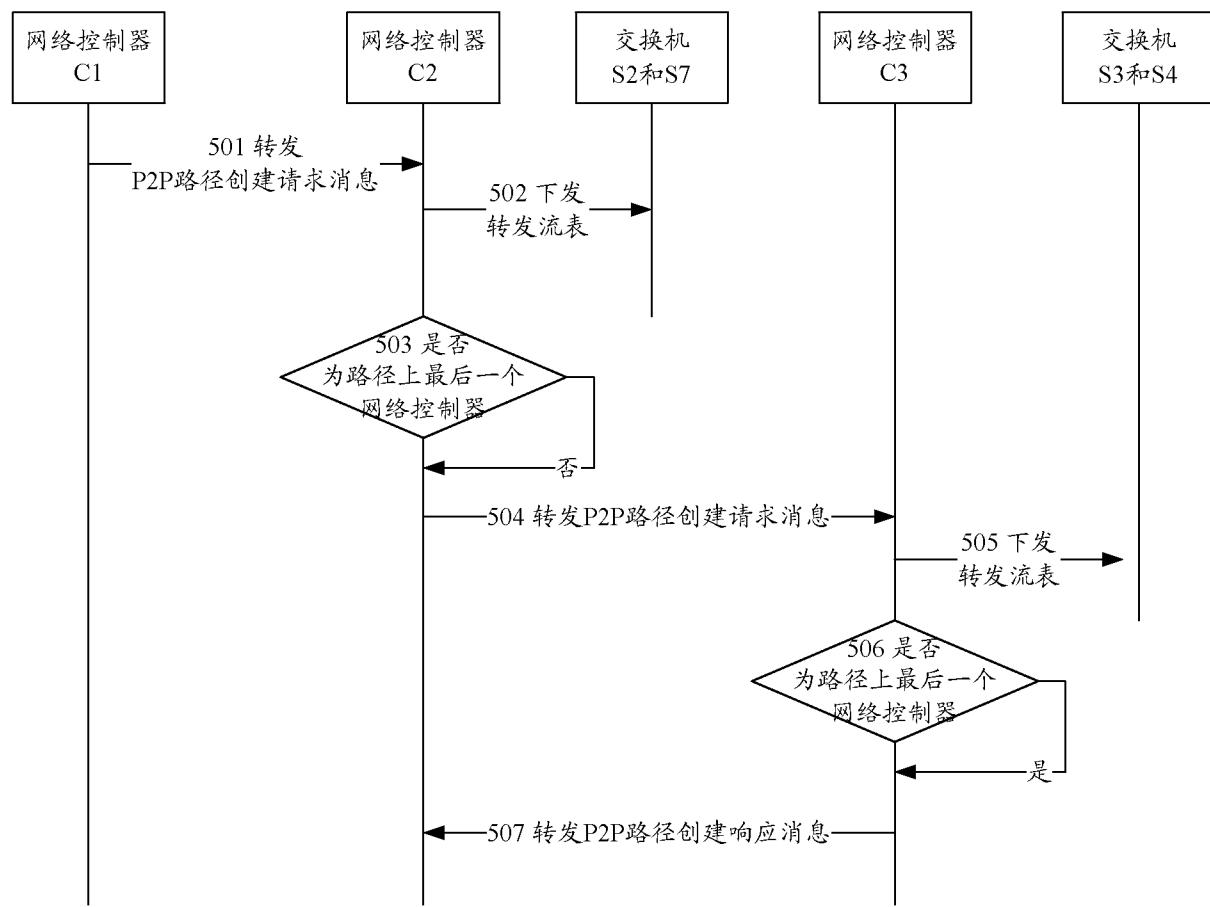


图 5

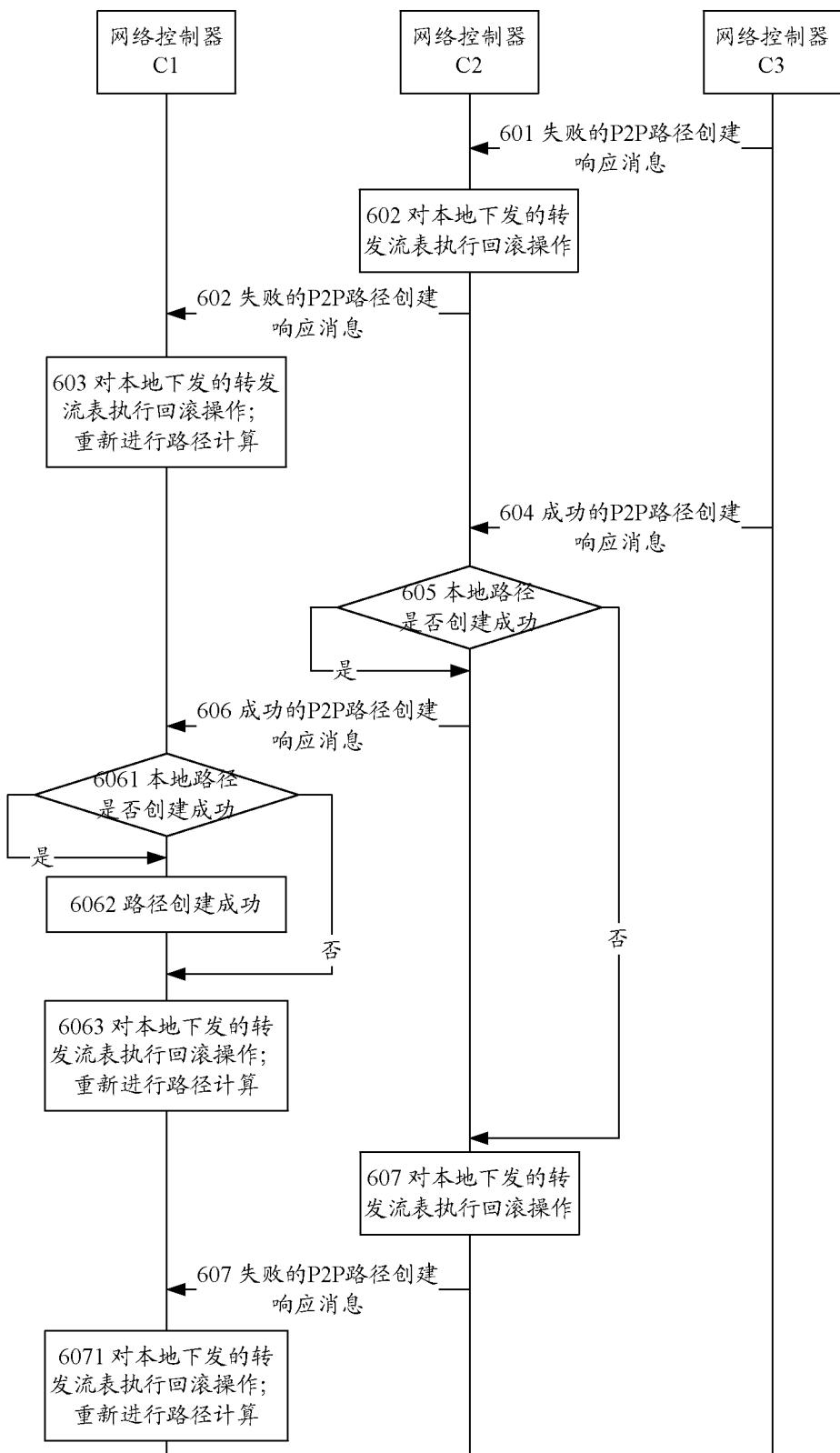


图 6

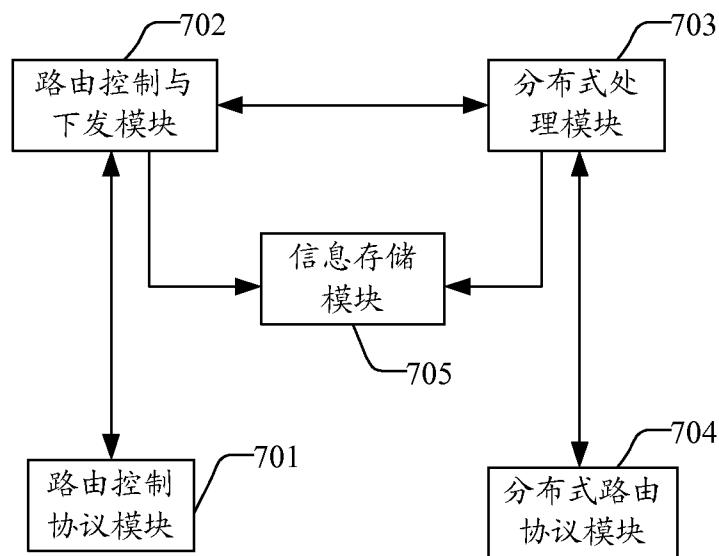


图 7

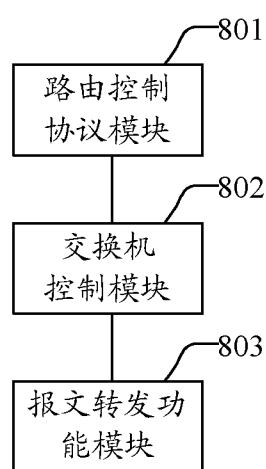


图 8

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2013/084519

A. CLASSIFICATION OF SUBJECT MATTER

H04L 29/08 (2006.01) 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)

IPC: H04W; H04Q; H04L; H04B

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)

WPI, EPODOC, CNPAT, CNKI: hash table, distribut+, DHT, path, traffic? channel, route, establish, build, set w up, message, request, response

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 103051539 A (ZTE CORP.), 17 April 2013 (17.04.2013), the whole document	1-20
A	WO 2012022751 A1 (ALCATEL LUCENT), 23 February 2012 (23.02.2012), description, page 9, line 3 to page 11, line 9	1-20
A	US 2011261722 A1 (NEC CORPORATION), 27 October 2011 (27.10.2011), the whole document	1-20
A	CN 102200768 A (OLYMPUS CORP.), 28 September 2011 (28.09.2011), the whole document	1-20
A	CN 102594689 A (ZTE CORP.), 18 July 2012 (18.07.2012), the whole document	1-20
A	CN 101039275 A (HUAWEI TECHNOLOGIES CO., LTD.), 19 September 2007 (19.09.2007), the whole document	1-20

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
10 November 2013 (10.11.2013)

Date of mailing of the international search report
02 January 2014 (02.01.2014)

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
JIANG, Li
Telephone No.: (86-10) 62413330

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/CN2013/084519

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 103051539 A	17.04.2013	None	
WO 2012022751 A1	23.02.2012	EP 2421225 A1	22.02.2012
		KR 20130050358 A	15.05.2013
		US 2013198332 A1	01.08.2013
		CN 103081445 A	01.05.2013
		JP 2013537765 A	03.10.2013
US 2011261722 A1	27.10.2011	WO 2011030889 A1	17.03.2011
		EP 2479938 A1	25.07.2012
		CN 102498694 A	13.06.2012
		JPWO 2011030889 A1	07.02.2013
CN 102200768 A	28.09.2011	US 2011238959 A1	29.09.2011
CN 102594689 A	18.07.2012	JP 2011203867 A	13.10.2011
CN 101039275 A	19.09.2007	WO 2013123846 A1	29.08.2013
		CN 100583831 C	20.01.2010

国际检索报告

国际申请号 PCT/CN2013/084519

A. 主题的分类

H04L29/08 (2006.01) 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)

IPC: H04W;H04Q;H04L;H04B

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))

WPI, EPODOC, CNPAT, CNKI: 分布式, 哈希表, 路径, 通道, 通路, 路由, 建立, 创建, 消息, 响应, 请求, distribut+, DHT, path, traffic, channel, route, establish, build, set w up, message, request, response

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
PX	CN103051539A (中兴通讯股份有限公司) 17.4 月 2013 (17.04.2013) 全文	1-20
A	WO 2012022751 A1 (ALCATEL LUCENT) 23.2 月 2012 (23.02.2012) 说明书第 9 页第 3 行至第 11 页第 9 行	1-20
A	US 2011261722 A1 (NEC CORPORATION) 27.10 月 2011 (27.10.2011) 全文	1-20
A	CN 102200768 A (奥林巴斯株式会社) 28.9 月 2011 (28.09.2011) 全文	1-20
A	CN 102594689 A (中兴通讯股份有限公司) 18.7 月 2012 (18.07.2012) 全文	1-20
A	CN 101039275 A (华为技术有限公司) 19.9 月 2007 (19.09.2007) 全文	1-20

 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

“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

“E” 在国际申请日的当天或之后公布的在先申请或专利

“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

“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)

“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

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“&” 同族专利的文件

“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

国际检索实际完成的日期
10.11 月 2013 (10.11.2013)国际检索报告邮寄日期
02.1 月 2014 (02.01.2014)ISA/CN 的名称和邮寄地址:
中华人民共和国国家知识产权局
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
蒋莉
电话号码: (86-10) **62413330**

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2013/084519

检索报告中引用的专利文件	公布日期	同族专利	公布日期
CN103051539A	17.04.2013	无	
WO2012022751A1	23.02.2012	EP2421225A1	22.02.2012
		KR20130050358A	15.05.2013
		US2013198332A1	01.08.2013
		CN103081445A	01.05.2013
		JP2013537765A	03.10.2013
		WO2011030889A1	17.03.2011
US2011261722A1	27.10.2011	EP2479938A1	25.07.2012
		CN102498694A	13.06.2012
		JPWO2011030889A1	07.02.2013
		US2011238959A1	29.09.2011
CN102200768A	28.09.2011	JP2011203867A	13.10.2011
		WO2013123846A1	29.08.2013
CN102594689A	18.07.2012	CN100583831C	20.01.2010
CN101039275A	19.09.2007		