

(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9月27日 (27.09.2018)



(10) 国际公布号
WO 2018/171529 A1

- (51) 国际专利分类号:
H04L 12/46 (2006.01) *H04L 29/12* (2006.01)
H04L 12/741 (2013.01)
- (21) 国际申请号: PCT/CN2018/079340
- (22) 国际申请日: 2018年3月16日 (16.03.2018)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201710165894.3 2017年3月20日 (20.03.2017) 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: 张平平 (ZHANG, Pingping); 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)。 陈志伟 (CHEN, Zhiwei); 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)。 夏迎春 (XIA, Yingchun); 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)。 孙军欢 (SUN, Junhuan); 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有限公司 (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, 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,

(54) Title: METHOD, DEVICE AND COMPUTER STORAGE MEDIUM FOR IMPLEMENTING DOUBLE CONTROL PLANE

(54) 发明名称: 一种实现双控制平面的方法、装置、计算机存储介质

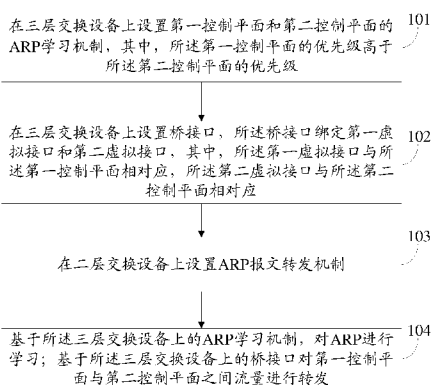


图 1

- 101 CONFIGURE AN ADDRESS RESOLUTION PROTOCOL (ARP) LEARNING MECHANISM OF A FIRST CONTROL PLANE AND A SECOND CONTROL PLANE ON A THREE-LAYER SWITCHING DEVICE, WHEREIN THE PRIORITY OF THE FIRST CONTROL PLANE IS HIGHER THAN THAT OF THE SECOND CONTROL PLANE
- 102 CONFIGURE A BRIDGE INTERFACE ON THE THREE-LAYER SWITCHING DEVICE, THE BRIDGE INTERFACE BINDING A FIRST VIRTUAL INTERFACE AND A SECOND VIRTUAL INTERFACE, WHEREIN THE FIRST VIRTUAL INTERFACE CORRESPONDS TO THE FIRST CONTROL PLANE, AND THE SECOND VIRTUAL INTERFACE CORRESPONDS TO THE SECOND CONTROL PLANE
- 103 CONFIGURE AN ARP MESSAGE FORWARDING MECHANISM ON A TWO-LAYER SWITCHING DEVICE
- 104 CARRY OUT LEARNING FOR AN ARP ON THE BASIS OF THE ARP LEARNING MECHANISM ON THE THREE-LAYER SWITCHING DEVICE; FORWARD TRAFFIC BETWEEN THE FIRST CONTROL PLANE AND THE SECOND CONTROL PLANE ON THE BASIS OF THE BRIDGE INTERFACE ON THE THREE-LAYER SWITCHING DEVICE

(57) Abstract: Disclosed in the present application are a method, device and computer storage medium for implementing a double control plane, the method comprising: configuring an address resolution protocol (ARP) learning mechanism of a first control plane and a second control plane on a three-layer switching device, wherein the priority of the first control plane is higher than that of the second control plane; configuring a bridge interface on the three-layer switching device, the bridge interface binding a first virtual interface and a second virtual interface, wherein the first virtual interface corresponds to the first control plane, and the second virtual interface corresponds to the second control plane; configuring an ARP message forwarding mechanism on a two-layer switching device; carrying out learning for an ARP on the basis of the ARP learning mechanism on the three-layer switching device; forwarding traffic between the first control plane and the second control plane on the basis of the bridge interface on the three-layer switching device; and forwarding an ARP message on the basis of the ARP message forwarding mechanism on the two-layer switching device.

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))。

(57) 摘要: 本申请公开了一种实现双控制平面的方法、装置、计算机存储介质, 所述方法包括: 在三层交换设备上设置第一控制平面和第二控制平面的地址解析协议ARP学习机制, 其中, 所述第一控制平面的优先级高于所述第二控制平面的优先级; 在三层交换设备上设置桥接口, 所述桥接口绑定第一虚拟接口和第二虚拟接口, 其中, 所述第一虚拟接口与所述第一控制平面相对应, 所述第二虚拟接口与所述第二控制平面相对应; 在二层交换设备上设置ARP报文转发机制; 基于所述三层交换设备上的ARP学习机制, 对ARP进行学习; 基于所述三层交换设备上的桥接口对第一控制平面与第二控制平面之间流量进行转发; 基于所述二层交换设备上ARP报文转发机制, 对ARP报文进行转发。

一种实现双控制平面的方法、装置、计算机存储介质

相关申请的交叉引用

本申请基于申请号为 201710165894.3、申请日为 2017 年 03 月 20 日的中国专利申请提出，并要求该中国专利申请的优先权，该中国专利申请的全部内容在此引入本申请作为参考。

技术领域

本申请涉及以太网数据中心领域中的数据交换技术，尤其涉及一种实现双控制平面的方法、装置、计算机存储介质。

背景技术

10 软件定义网络（SDN，Software Defined Network），是Emulex网络一种新型网络创新架构，是网络虚拟化的一种实现方式，其核心技术OpenFlow通过将网络设备控制面与数据面分离开来，从而实现了网络流量的灵活控制，使网络作为管道变得更加智能。

传统 IT 架构中的网络，根据业务需求部署上线以后，如果业务需求发生变动，重新修改相应网络设备（路由器、交换机、防火墙）上的配置是一件非常繁琐的事情。在互联网/移动互联网瞬息万变的业务环境下，网络的高稳定与高性能还不足以满足业务需求，灵活性和敏捷性反而更为关键。SDN 所做的是将网络设备上的控制权分离出来，由集中的控制器管理，无须依赖底层网络设备（路由器、交换机、防火墙），屏蔽了来自底层网络设备的差异。而控制权是完全开放的，用户可以自定义任何想实现的网络路由和传输规则策略，从而更加灵活和智能。

近年来由于数据中心的网络虚拟化要求，SDN 技术迅速崛起，SDN 为

网络设计带来了极大便利，SDN 时代的网络，不再是以设备为中心，而是以应用为中心，应用驱动网络变革。这就需要很多深度定制的工作，云计算网络尤其如此。

5 虽然 SDN 在网络设备组网的应用前景非常明朗，它所带来的技术革命也被广泛认可，但是目前全 SDN 网络的应用存在极高的风险，因为 SDN 网络完全受控于器控制，一旦控制器发生故障或者掉线会造成整个网络的瘫痪。

发明内容

10 为解决上述技术问题，本申请实施例提供了一种实现双控制平面的方法、装置、计算机存储介质。

本申请实施例提供的实现双控制平面的方法，包括：

在三层交换设备上设置第一控制平面和第二控制平面的地址解析协议（ARP，Address Resolution Protocol）学习机制，其中，所述第一控制平面的优先级高于所述第二控制平面的优先级；

15 在三层交换设备上设置桥接口，所述桥接口绑定第一虚拟接口和第二虚拟接口，其中，所述第一虚拟接口与所述第一控制平面对应，所述第二虚拟接口与所述第二控制平面对应；

在二层交换设备上设置 ARP 报文转发机制；

20 基于所述三层交换设备上的 ARP 学习机制，对 ARP 进行学习；基于所述三层交换设备上的桥接口对第一控制平面与第二控制平面之间流量进行转发；

基于所述二层交换设备上 ARP 报文转发机制，对 ARP 报文进行转发。

本申请实施例中，所述第一控制平面的 ARP 学习机制为自学习机制，所述第二控制平面的 ARP 学习机制为控制器下发机制。

25 本申请实施例中，所述基于所述三层交换设备上的 ARP 学习机制，对

ARP 进行学习，包括：

当所述三层交换设备接收到 ARP 报文时，查询本地路由表里是否存在所述 ARP 报文对应的 ARP 地址；

5 如果所述本地路由表里不存在所述 ARP 地址，则将所述 ARP 地址添加至所述本地路由表里，并且将所述 ARP 地址对应的表项标记为自学习；

如果所述本地路由表里存在所述 ARP 地址，且所述 ARP 地址对应的 MAC 地址与出端口对应的 MAC 地址一致时，将所述 ARP 地址对应的表项标记为自学习并对所述 ARP 地址对应的表项进行更新时间的刷新；

10 如果所述本地路由表里存在所述 ARP 地址，且所述 ARP 地址对应的 MAC 地址与出端口对应的 MAC 地址不一致时，依据所述 ARP 报文对应的 ARP 地址对所述本地路由表里中对应的表项进行覆盖，并且将所述 ARP 地址对应的表项标记为自学习。

本申请实施例中，所述基于所述三层交换设备上的 ARP 学习机制，对 ARP 进行学习，包括：

15 当所述三层交换设备从控制器接收到 ARP 流表添加消息时，查询本地路由表里是否存在所述 ARP 流表添加消息对应的 ARP 地址；

如果所述本地路由表里不存在所述 ARP 地址，则将所述 ARP 地址添加至所述本地路由表里，并且将所述 ARP 地址对应的表项标记为控制器下发；

20 如果所述本地路由表里存在所述 ARP 地址，且所述 ARP 地址对应的 MAC 地址与出端口对应的 MAC 地址一致时，则保持所述本地路由表不变；

如果所述本地路由表里存在所述 ARP 地址，且所述 ARP 地址对应的 MAC 地址与出端口对应的 MAC 地址不一致时，如果所述本地路由表里对应的表项标记为自学习，则保持所述本地路由表不变，如果所述本地路由表里对应的表项标记为控制器下发，则依据所述 ARP 流表添加消息对应的
25 ARP 地址对所述本地路由表里中对应的表项进行覆盖。

本申请实施例中，所述第一虚拟接口为 IP 协议栈虚拟接口，所述第二虚拟接口为 SDN 协议栈虚拟接口，其中，

所述三层交换设备通过所述 IP 协议栈虚拟接口将 SDN 网络的路由发布至出口路由器；

5 所述三层交换设备通过所述 SDN 协议栈虚拟接口接收控制器下发的流表。

本申请实施例中，所述三层交换设备包括：SDN 协议栈模块、IP 协议栈模块、收发包模块；

10 所述基于所述三层交换设备上的桥接口对第一控制平面与第二控制平面之间流量进行转发，包括：

SDN 协议栈模块接收所述控制器发送的 ARP 请求，所述 SDN 协议栈模块在所述 ARP 请求上打上 SDN 协议栈虚拟接口的标记，并发送给收发包模块，通过所述收发包模块将所述 SDN 协议栈虚拟接口的标记替换为 IP 协议栈虚拟接口的标记，并发送给 IP 协议栈模块；

15 所述 IP 协议栈模块对 ARP 进行学习，在 ARP 应答中打上 IP 协议栈虚拟接口的标记，并发送给收发包模块，通过所述收发包模块将所述 IP 协议栈虚拟接口的标记替换为 SDN 协议栈虚拟接口的标记，并发送给 SDN 协议栈模块；

20 所述 SDN 协议栈模块将所述 ARP 应答发送给控制器，触发所述控制器根据所述 ARP 应答进行 ARP 学习。

本申请实施例中，所述三层交换设备包括：SDN 协议栈模块、IP 协议栈模块、收发包模块；

所述基于所述三层交换设备上的桥接口对第一控制平面与第二控制平面之间流量进行转发，包括：

25 针对南向流量的三层报文，当所述三层报文达到所述桥接口后没有查

到路由时，将所述报文通过网段路由发送给 IP 协议栈模块；所述 IP 协议栈模块根据所述报文的目的地 IP 进行路由查询，如果查询到 IP 协议栈虚拟接口的出口路由，则所述 IP 协议栈模块在所述报文上打上 IP 协议栈虚拟接口的标记；如果未查询到 IP 协议栈虚拟接口的出口路由，则所述 IP 协议栈模块
5 触发自学习机制的 APR 学习；

所述 IP 协议栈模块将打上 IP 协议栈虚拟接口的标记的报文发送给收发包模块，通过所述收发包模块将所述 IP 协议栈虚拟接口的标记替换为 SDN 协议栈虚拟接口的标记，并发送给 SDN 协议栈模块；

所述 SDN 协议栈模块对所述报文进行封装后，发送给所述控制器，触
10 发所述控制器进行 ARP 学习；

所述 SDN 协议栈模块接收所述控制器下发的流表。

本申请实施例中，所述基于所述二层交换设备上 ARP 报文转发机制，对 ARP 报文进行转发，包括：

所述二层交换设备接收到服务器发送的 ARP 报文时，将所述 ARP 报文
15 发送给所述三层交换设备进行 ARP 学习，以及将所述 ARP 报文发送给控制器进行 ARP 学习。

本申请另一实施例提供的实现双控制平面的方法，包括：

在 SDN 域配置好出口路由器的默认路由后，控制器主动触发所述默认路由的下一跳 ARP 学习；

20 报文达到三层交换设备的桥接口后没有查到路由时，通过报文触发所述控制器进行 ARP 学习；所述控制器将流表下发给所述三层交换设备。

本申请实施例中，所述在 SDN 域配置好出口路由器的默认路由后，控制器主动触发所述默认路由的下一跳 ARP 学习，包括：

SDN 域配置好出口路由器的默认路由后，控制器发送 ARP 请求给所述
25 三层交换设备的 SDN 协议栈模块，其中，所述 SDN 协议栈模块在所述 ARP

请求上打上 SDN 协议栈虚拟接口的标记，并发送给收发包模块，通过所述收发包模块将所述 SDN 协议栈虚拟接口的标记替换为 IP 协议栈虚拟接口的标记，并发送给 IP 协议栈模块；所述 IP 协议栈模块对 ARP 进行学习，在 ARP 应答中打上 IP 协议栈虚拟接口的标记，并发送给收发包模块，通过所述收发包模块将所述 IP 协议栈虚拟接口的标记替换为 SDN 协议栈虚拟接口的标记，并发送给 SDN 协议栈模块；

控制器接收所述 SDN 协议栈模块发送的 ARP 应答，根据所述 ARP 应答进行 ARP 学习。

本申请实施例中，所述报文达到三层交换设备的桥接口后没有查到路由时，通过报文触发所述控制器进行 ARP 学习，包括：

当三层报文达到所述桥接口后没有查到路由时，所述控制器接收 SDN 协议栈模块发送的封装后的报文，其中，所述封装后的报文通过以下方式生成：当三层报文达到所述桥接口后没有查到路由时，将所述报文通过网段路由发送给 IP 协议栈模块；所述 IP 协议栈模块根据所述报文的 IP 进行路由查询，如果查询到 IP 协议栈虚拟接口的出口路由，则所述 IP 协议栈模块在所述报文上打上 IP 协议栈虚拟接口的标记；所述 IP 协议栈模块将打上 IP 协议栈虚拟接口的标记的报文发送给收发包模块，通过所述收发包模块将所述 IP 协议栈虚拟接口的标记替换为 SDN 协议栈虚拟接口的标记，并发送给 SDN 协议栈模块；所述 SDN 协议栈模块对所述报文进行封装，得到封装后的报文；

所述控制器接收到所述封装后的报文后，进行 ARP 学习并更新流表；

所述控制器将更新的流表发送至三层交换设备中的 SDN 协议栈模块。

本申请实施例中，所述方法还包括：

所述控制器周期性地比较本地路由表和交换设备上的路由表，当所述本地路由表中的表项与交换设备上的路由表中的表项不一致时，将所述交

交换机上的表项更新至本地。

本申请实施例提供的实现双控制平面的装置，包括：

第一配置单元，配置为在三层交换设备上设置第一控制平面和第二控制平面的 ARP 学习机制，其中，所述第一控制平面的优先级高于所述第二控制平面的优先级；在三层交换设备上设置桥接口，所述桥接口绑定第一
5 虚拟接口和第二虚拟接口，其中，所述第一虚拟接口与所述第一控制平面相对应，所述第二虚拟接口与所述第二控制平面相对应；

第二配置单元，配置为在二层交换设备上设置 ARP 报文转发机制；

第一处理单元，配置为基于所述三层交换设备上的 ARP 学习机制，对
10 ARP 进行学习；基于所述三层交换设备上的桥接口对第一控制平面与第二控制平面之间流量进行转发；

第二处理单元，配置为基于所述二层交换设备上 ARP 报文转发机制，对 ARP 报文进行转发。

本申请实施例中，所述第一控制平面的 ARP 学习机制为自学习机制，
15 所述第二控制平面的 ARP 学习机制为控制器下发机制。

本申请实施例中，所述第一处理单元包括：

第一学习子单元，配置为当所述三层交换设备接收到 ARP 报文时，查询本地路由表里是否存在所述 ARP 报文对应的 ARP 地址；如果所述本地路由表里不存在所述 ARP 地址，则将所述 ARP 地址添加至所述本地路由表里，
20 并且将所述 ARP 地址对应的表项标记为自学习；如果所述本地路由表里存在所述 ARP 地址，且所述 ARP 地址对应的 MAC 地址与出端口对应的 MAC 地址一致时，将所述 ARP 地址对应的表项标记为自学习并对所述 ARP 地址对应的表项进行更新时间的刷新；如果所述本地路由表里存在所述 ARP 地址，且所述 ARP 地址对应的 MAC 地址与出端口对应的 MAC 地址不一致
25 时，依据所述 ARP 报文对应的 ARP 地址对所述本地路由表里中对应的表项

进行覆盖，并且将所述 ARP 地址对应的表项标记为自学习。

本申请实施例中，所述第一处理单元包括：

第二学习子单元，配置为当所述三层交换设备从控制器接收到 ARP 流表添加消息时，查询本地路由表里是否存在所述 ARP 流表添加消息对应的
5 ARP 地址；如果所述本地路由表里不存在所述 ARP 地址，则将所述 ARP 地址添加至所述本地路由表里，并且将所述 ARP 地址对应的表项标记为控制器下发；如果所述本地路由表里存在所述 ARP 地址，且所述 ARP 地址对应的 MAC 地址与出端口对应的 MAC 地址一致时，则保持所述本地路由表不变；如果所述本地路由表里存在所述 ARP 地址，且所述 ARP 地址对应的
10 MAC 地址与出端口对应的 MAC 地址不一致时，如果所述本地路由表里对应的表项标记为自学习，则保持所述本地路由表不变，如果所述本地路由表里对应的表项标记为控制器下发，则依据所述 ARP 流表添加消息对应的 ARP 地址对所述本地路由表里中对应的表项进行覆盖。

本申请实施例中，所述第一虚拟接口为 IP 协议栈虚拟接口，所述第二
15 虚拟接口为 SDN 协议栈虚拟接口；其中，

所述三层交换设备通过所述 IP 协议栈虚拟接口将 SDN 网络的路由发布至出口路由器；

所述三层交换设备通过所述 SDN 协议栈虚拟接口接收控制器下发的流表。

20 本申请实施例中，所述三层交换设备包括：SDN 协议栈模块、IP 协议栈模块、收发包模块；其中，

所述 SDN 协议栈模块接收所述控制器发送的 ARP 请求，所述 SDN 协议栈模块在所述 ARP 请求上打上 SDN 协议栈虚拟接口的标记，并发送给收发包模块，通过所述收发包模块将所述 SDN 协议栈虚拟接口的标记替换
25 为 IP 协议栈虚拟接口的标记，并发送给 IP 协议栈模块；

所述 IP 协议栈模块对 ARP 进行学习, 在 ARP 应答中打上 IP 协议栈虚拟接口的标记, 并发送给收发包模块, 通过所述收发包模块将所述 IP 协议栈虚拟接口的标记替换为 SDN 协议栈虚拟接口的标记, 并发送给 SDN 协议栈模块;

5 所述 SDN 协议栈模块将所述 ARP 应答发送给控制器, 触发所述控制器根据所述 ARP 应答进行 ARP 学习。

本申请实施例中, 所述三层交换设备包括: SDN 协议栈模块、IP 协议栈模块、收发包模块; 其中,

针对南向流量的三层报文, 当所述三层报文达到所述桥接口后没有查到路由时, 将所述报文通过网段路由发送给 IP 协议栈模块; 所述 IP 协议栈模块根据所述报文的 IP 进行路由查询, 如果查询到 IP 协议栈虚拟接口的出口路由, 则所述 IP 协议栈模块在所述报文上打上 IP 协议栈虚拟接口的标记; 如果未查询到 IP 协议栈虚拟接口的出口路由, 则所述 IP 协议栈模块触发自学习机制的 APR 学习;

15 所述 IP 协议栈模块将打上 IP 协议栈虚拟接口的标记的报文发送给收发包模块, 通过所述收发包模块将所述 IP 协议栈虚拟接口的标记替换为 SDN 协议栈虚拟接口的标记, 并发送给 SDN 协议栈模块;

所述 SDN 协议栈模块对所述报文进行封装后, 发送给所述控制器, 触发所述控制器进行 ARP 学习;

20 所述 SDN 协议栈模块接收所述控制器下发的流表。

本申请实施例中, 所述第二处理单元, 具体用于当所述二层交换设备接收到服务器发送的 ARP 报文时, 将所述 ARP 报文发送给所述三层交换设备进行 ARP 学习, 以及将所述 ARP 报文发送给控制器进行 ARP 学习。

本申请另一实施例提供的实现双控制平面的装置, 包括:

25 第一学习单元, 配置为在 SDN 域配置好出口路由器的默认路由后, 控

制器主动触发所述默认路由的下一跳 ARP 学习；

第二学习单元，配置为报文达到三层交换设备的桥接口后没有查到路由时，通过报文触发所述控制器进行 ARP 学习；

流表下发单元，配置为将流表下发给所述三层交换设备。

5 本申请实施例中，所述第一学习单元，具体用于：SDN 域配置好出口路由器的默认路由后，发送 ARP 请求给所述三层交换设备的 SDN 协议栈模块，其中，所述 SDN 协议栈模块在所述 ARP 请求上打上 SDN 协议栈虚拟接口的标记，并发送给收发包模块，通过所述收发包模块将所述 SDN 协议栈虚拟接口的标记替换为 IP 协议栈虚拟接口的标记，并发送给 IP 协议栈
10 模块；所述 IP 协议栈模块对 ARP 进行学习，在 ARP 应答中打上 IP 协议栈虚拟接口的标记，并发送给收发包模块，通过所述收发包模块将所述 IP 协议栈虚拟接口的标记替换为 SDN 协议栈虚拟接口的标记，并发送给 SDN 协议栈模块；接收所述 SDN 协议栈模块发送的 ARP 应答，根据所述 ARP 应答进行 ARP 学习。

15 本申请实施例中，所述第二学习单元，具体用于：当三层报文达到所述桥接口后没有查到路由时，所述控制器接收 SDN 协议栈模块发送的封装后的报文，其中，所述封装后的报文通过以下方式生成：当三层报文达到所述桥接口后没有查到路由时，将所述报文通过网段路由发送给 IP 协议栈模块；所述 IP 协议栈模块根据所述报文的 IP 进行路由查询，如果查询
20 到 IP 协议栈虚拟接口的出口路由，则所述 IP 协议栈模块在所述报文上打上 IP 协议栈虚拟接口的标记；所述 IP 协议栈模块将打上 IP 协议栈虚拟接口的标记的报文发送给收发包模块，通过所述收发包模块将所述 IP 协议栈虚拟接口的标记替换为 SDN 协议栈虚拟接口的标记，并发送给 SDN 协议栈模块；所述 SDN 协议栈模块对所述报文进行封装，得到封装后的报文；接
25 收到所述封装后的报文后，进行 ARP 学习并更新流表。

本申请实施例中，所述装置还包括：第三学习单元，配置为周期性地比较本地路由表和交换设备上的路由表，当所述本地路由表中的表项与交换设备上的路由表中的表项不一致时，将所述交换机上的表项更新至本地。

本申请实施例还提供一种计算机存储介质，该计算机存储介质存储有
5 计算机程序，该计算机程序配置为执行上述实现双控制平面的方法。

本申请实施例的技术方案中，在三层交换设备上设置第一控制平面和第二控制平面的 ARP 学习机制，其中，所述第一控制平面的优先级高于所述第二控制平面的优先级；在三层交换设备上设置桥接口，所述桥接口绑定第一虚拟接口和第二虚拟接口，其中，所述第一虚拟接口与所述第一控制平面相对应，所述第二虚拟接口与所述第二控制平面相对应；在二层交换设备上设置 ARP 报文转发机制；基于所述三层交换设备上的 ARP 学习机制，对 ARP 进行学习；基于所述三层交换设备上的桥接口对第一控制平面与第二控制平面之间流量进行转发；基于所述二层交换设备上 ARP 报文转发机制，对 ARP 报文进行转发。采用本申请实施例的技术方案，利用 SDN
10 网络带来的便利的同时还增加了整个网络的稳定性，当控制器发生故障或者掉线时整个网络还能够依据传统控制平面保持路由的顺利进行，在数据中心领域中有极高的实用性。

附图说明

附图以示例而非限制的方式大体示出了本文中所讨论的各个实施例。

20 图 1 为本申请实施例的实现双控制平面的方法的流程示意图一；

图 2 为本申请实施例的实现双控制平面的方法的流程示意图二；

图 3 为本申请实施例的实现双控制平面的方法的流程示意图三；

图 4 为本申请实施例的网络组网的拓扑图；

图 5 为本申请实施例的网络设备的结构示意图一；

25 图 6 为本申请实施例的网络设备的结构示意图二；

图 7 为本申请实施例的桥接口的工作流程图；

图 8 为本申请实施例的控制器触发学习下一跳 ARP 的流程图；

图 9 为本申请实施例的流量触发控制器学习 ARP 的流程图；

图 10 为本申请实施例的实现双控制平面的装置的结构组成示意图一；

5 图 11 为本申请实施例的实现双控制平面的装置的结构组成示意图二。

具体实施方式

为了能够更加详尽地了解本申请实施例的特点与技术内容，下面结合附图对本申请实施例的实现进行详细阐述，所附附图仅供参考说明之用，并非用来限定本申请实施例。

10 本申请实施例给出一种 SDN 控制平面和传统控制平面混跑的方案，即交换设备即可以运行在传统控制平面，也可以运行在 SDN 控制平面。由于 SDN 控制平面和传统控制平面共存，所以要解决 SDN 控制平面和传统控制平面双大脑听谁指挥的问题，即表项下发优先级的问题；由于 SDN 控制平面和传统控制平面共存，数据中心内部网络与外部路由器之间的接口的流
15 量有的需要走传统控制平面有的要走 SDN 控制平面，如何处理控制平面之间的流量也是本申请实施例解决的问题。传统控制平面需要跑各种路由协议学习动态路由，以及学习下一跳路由；SDN 控制平面需要报文能触发控制器学习精确流表。本申请实施例将从双控制平面表项优先级控制、SDN 网络与传统网络桥接、南向流量和北向流量的处理以及接入设备 ARP 处理
20 四个方面系统阐述本申请实施例的技术方案，以下为本申请实施例的具体实施细节。

图 1 为本申请实施例的实现双控制平面的方法的流程示意图一，本示例中的实现双控制平面的方法应用于交换设备侧，如图 1 所示，所述方法包括以下步骤：

25 步骤 101: 在三层交换设备上设置第一控制平面和第二控制平面的 ARP

学习机制，其中，所述第一控制平面的优先级高于所述第二控制平面的优先级。

预先，对三层交换设备进行三层虚拟可扩展局域网（VXLAN，Virtual eXtensible LAN）的配置。

5 如图 4 所示的网络拓扑中，数据中心内部包括多组网络设备结构，例如，A/B 设备、C/D 设备、ServerA/ServerB；E/F 设备、G/H 设备、ServerC/ServerD。其中，A、B、E、F 设备为三层交换设备，而 C、D、G、H 为二层交换设备。本申请实施例以 A/B 设备、C/D 设备、ServerA/ServerB，为例进行解释说明，E/F 设备、G/H 设备、ServerC/ServerD 同理。

10 如图 4 所示的网络拓扑中，A/B 设备与 C/D 设备之间走 VXLAN 隧道，A/B 设备在传统控制平面配置 L3 VXLAN 相关配置，C/D 设备在传统控制平面配置 L2 VXLAN 相关配置。

如图 4 所示的网络拓扑中，A/B 设备与控制器相连，此外，C/D 设备也与控制器相连。

15 本申请实施例中，所述第一控制平面的 ARP 学习机制为自学习机制，所述第二控制平面的 ARP 学习机制为控制器下发机制。

如图 4 所示的网络拓扑中，对于网络设备 A/B 的 VXLAN ARP 的学习同时支持传统自学习以及控制器下发两个控制面，且自学习的优先级始终最高。

20 步骤 102：在三层交换设备上设置桥接口，所述桥接口绑定第一虚拟接口和第二虚拟接口，其中，所述第一虚拟接口与所述第一控制平面对应，所述第二虚拟接口与所述第二控制平面对应。

步骤 103：在二层交换设备上设置 ARP 报文转发机制。

25 步骤 104：基于所述三层交换设备上的 ARP 学习机制，对 ARP 进行学习；基于所述三层交换设备上的桥接口对第一控制平面与第二控制平面之

间流量进行转发。

本申请实施例中，所述基于所述三层交换设备上的 ARP 学习机制，对 ARP 进行学习，包括：

5 当所述三层交换设备接收到 ARP 报文时，查询本地路由表里是否存在所述 ARP 报文对应的 ARP 地址；

如果所述本地路由表里不存在所述 ARP 地址，则将所述 ARP 地址添加至所述本地路由表里，并且将所述 ARP 地址对应的表项标记为自学习；

10 如果所述本地路由表里存在所述 ARP 地址，且所述 ARP 地址对应的 MAC 地址与出端口对应的 MAC 地址一致时，将所述 ARP 地址对应的表项标记为自学习并对所述 ARP 地址对应的表项进行更新时间的刷新；

如果所述本地路由表里存在所述 ARP 地址，且所述 ARP 地址对应的 MAC 地址与出端口对应的 MAC 地址不一致时，依据所述 ARP 报文对应的 ARP 地址对所述本地路由表里中对应的表项进行覆盖，并且将所述 ARP 地址对应的表项标记为自学习。

15 本申请实施例中，所述基于所述三层交换设备上的 ARP 学习机制，对 ARP 进行学习，包括：

当所述三层交换设备从控制器接收到 ARP 流表添加消息时，查询本地路由表里是否存在所述 ARP 流表添加消息对应的 ARP 地址；

20 如果所述本地路由表里不存在所述 ARP 地址，则将所述 ARP 地址添加至所述本地路由表里，并且将所述 ARP 地址对应的表项标记为控制器下发；

如果所述本地路由表里存在所述 ARP 地址，且所述 ARP 地址对应的 MAC 地址与出端口对应的 MAC 地址一致时，则保持所述本地路由表不变；

25 如果所述本地路由表里存在所述 ARP 地址，且所述 ARP 地址对应的 MAC 地址与出端口对应的 MAC 地址不一致时，如果所述本地路由表里对应的表项标记为自学习，则保持所述本地路由表不变，如果所述本地路由

表里对应的表项标记为控制器下发，则依据所述 ARP 流表添加消息对应的 ARP 地址对所述本地路由表里中对应的表项进行覆盖。

上述方案中，提出了**双控制平面表项优先级的控制**，具体地：

如图 4 所示的网络拓扑中，传统自学习（对应自学习机制）具体包括：

5 当网络设备 A/B 从业务端口收到 VXLAN ARP 报文时，会按照传统自学习流程对该 ARP 报文进行动态学习，也即自学习。1) 若本地路由表里没有该 ARP 地址，则直接添加表项，FLAG 标记为动态学习或自学习；2) 若本地已存在该 ARP 地址，且 MAC 和出端口信息一致，FLAG 标记为动态学习修改 ARP 的状态为动态学习或自学习，3) 如果 MAC 和出端口不一致覆盖原有表项，FLAG 标记为动态学习或自学习。

如图 4 所示的网络拓扑中，控制器下发 ARP 具体包括：当网络设备 A/B 从控制器收到 ARP 流表添加消息时，1) 若本地路由表中没有该 ARP 地址，则直接添加表项，FLAG 标记为控制器下发；2) 若本地已存在该 ARP 地址，且 MAC 和出端口信息一致，则不做任何处理。3) 若本地已存在该
15 ARP 地址，但 MAC 和出端口信息不一致，若原有的 FLAG 标记为动态学习的，不做任何处理，若原有的 FLAG 标记是控制器下发的，则覆盖原有的表项。

传统控制平面学习到的 VXLAN ARP 的优先级高，控制器对本身保存的 ARP 和交换机上的 ARP 表周期进行对比来保障两者之间的同步；若控制
20 器有而交换机上没有则下发给交换机，如果两者都有，但是出口不一致，则控制器取回交换机上的覆盖本地的。

通过上述方法当控制器异常时流量会按照传统控制平面学习的表项转发，当传统平面没有学习到相关表项时流量会按照控制器下发的流表转发流量，当虚拟机迁移场景发生时，控制器上的表项与传统控制平面学习到的
25 的表项数据不同步时，控制器会取回传统平面学到的表项覆盖本地实现同

步。

基于图 1 所示的方法，本申请实施例的方案还可以包括：

所述第一虚拟接口为 IP 协议栈虚拟接口，所述第二虚拟接口为 SDN 协议栈虚拟接口；其中，

5 所述三层交换设备通过所述 IP 协议栈虚拟接口将 SDN 网络的路由发布至出口路由器；

所述三层交换设备通过所述 SDN 协议栈虚拟接口接收控制器下发的流表。

在一实施方式中，所述三层交换设备包括：SDN 协议栈模块、IP 协议栈模块、收发包模块；SDN 协议栈模块接收所述控制器发送的 ARP 请求，所述 SDN 协议栈模块在所述 ARP 请求上打上 SDN 协议栈虚拟接口的标记，并发送给收发包模块，通过所述收发包模块将所述 SDN 协议栈虚拟接口的标记替换为 IP 协议栈虚拟接口的标记，并发送给 IP 协议栈模块；

15 所述 IP 协议栈模块对 ARP 进行学习，在 ARP 应答中打上 IP 协议栈虚拟接口的标记，并发送给收发包模块，通过所述收发包模块将所述 IP 协议栈虚拟接口的标记替换为 SDN 协议栈虚拟接口的标记，并发送给 SDN 协议栈模块；

所述 SDN 协议栈模块将所述 ARP 应答发送给控制器，触发所述控制器根据所述 ARP 应答进行 ARP 学习。

20 在另一实施方式中，针对南向流量的三层报文，当所述三层报文达到所述桥接口后没有查到路由时，将所述报文通过网段路由发送给 IP 协议栈模块；所述 IP 协议栈模块根据所述报文的目 IP 进行路由查询，如果查询到 IP 协议栈虚拟接口的出口路由，则所述 IP 协议栈模块在所述报文上打上 IP 协议栈虚拟接口的标记；如果未查询到 IP 协议栈虚拟接口的出口路由，
25 则所述 IP 协议栈模块触发自学习机制的 APR 学习；

所述 IP 协议栈模块将打上 IP 协议栈虚拟接口的标记的报文发送给收发包模块，通过所述收发包模块将所述 IP 协议栈虚拟接口的标记替换为 SDN 协议栈虚拟接口的标记，并发送给 SDN 协议栈模块；

所述 SDN 协议栈模块对所述报文进行封装后，发送给所述控制器，触
5 发所述控制器进行 ARP 学习；

所述 SDN 协议栈模块接收所述控制器下发的流表。

上述方案中，提出了内部网络与外部网络的桥接、南向流量和北向流量的处理，具体地：

1) 内部网络与外部网络的桥接

10 由于采用传统控制平面和 SDN 控制平面共存的模式，这就对于内部网络和外部网络桥接的端口的报文收发提出了挑战，因为需要采用一种合理的方式让报文上送到对的协议栈（传统 IP 协议栈与 SDN 协议栈）。

对于 SDN 域和传统域之间的流量的桥接采用桥接口的方式，该接口即可以发挥传统控制平面现有路由协议的特点将内部 SDN 网络的路由发布到
15 上行路由器（出口路由器）从而将外部发往数据中心（DC，Data Center）内的流量引下来，同时又能响应控制器的指令承担由 SDN 内部发往外部流量的流表出口。桥接口是普通物理口和 SDN 端口概念的延伸，普通物理口对于接受到的协议报文只能上送到传统 IPSTACK 进程，SDN 端口对于接收到协议报文只能上送到 SDN 相关协议进程。对于传统控制平面和 SDN 控
20 制平面共存的场景，这里提出桥接口的概念，即通过这个接口即能发布传统路由协议，又能受控制器的控制作为 SDN 域到域外的接口。

由于交换机的两个控制平面（传统协议栈控制平面和 SDN 控制平面）属于不同的进程，对于桥接口会绑定两个虚拟接口，分别称为传统协议栈虚拟接口以及 SDN 控制平面虚拟接口，我们在下文中分别称为
25 IPSTACK_VP 以及 SDN_VP。这两个虚拟接口与桥接口的物理口绑定，绑

定关系在用户配置桥接端口后下发。交换机底层不需要保存虚拟出口路由，只需要保存实际转发使用的路由，但要实现传统 IP 协议栈进程与 SDN 协议进程之间的报文上送通道，即要根据报文下发（来自传统 IP 协议栈进程与 SDN 协议进程）信息中的虚拟出口标记（IPSTACK_VP 或者 SDN_VP），决定将报文上送给相应的进程。此外，对于南向流量的转发控制器会给传统控制平面 IP 协议栈下发一个网段路由出口指向 IPSTACK_VP，该网段是虚拟机所在网络的网段。这样当南向数据报文没有匹配到精确的流表时，会走该网段路由上送到控制器，控制器会下发精确流表的表项。桥接口的实现流程可参见附图 7 所述。

10 通过引入桥接口的方案，采用虚拟接口的方式在不改动现有交换机传统控制平面数据转发以及协议报文上送流程的情况下，实现传统平面和 SDN 平面数据报文和协议报文互不干扰。

2) 南向流量和北向流量的处理

对于北向流量，SDN 域配置默认出口路由后，控制器会主动触发默认路由下一跳 ARP 学习。控制器发送 ARP 请求给交换机的 SDN 协议栈，SDN 协议栈会将此报文带上 SDN_VP 的出口标记，然后这个报文会送个 IP 协议栈，IP 协议栈负责在传统的控制平面学到 ARP 后，再给交换机 SDN 协议栈发送 ARP 回复报文并打上 IPSTACK_VP 的标记，最后 SDN 控制器收到 ARP 应答后，学习到 ARP，下发默认路由的相关流表。这里实现两个协议栈的交换采用收发包模块换回的方式实现，即两个协议栈不直接交换信息通过收发包模块中转，这样的好处是不需要改动现有交换机协议栈收发包的处理模式。

对于南向流量最关键的处理是报文到达桥接口后没有查到路由要触发控制学习主机 ARP，然后下发给交换机下发流表。交换机在收到三层报文后如果没有路由会走网段路由上送给传统 IP 协议栈，IP 协议栈如果能匹配

IPSTACK_VP，就会打上 IPSTACK_VP 的标记，然后 SDN 协议栈就会收到这个报文，SDN 协议栈负责封装 SDN 的协议报文（例如 OPEN_FLOW）后送 SDN 控制器。SDN 控制器收到后，触发 ARP 学习，下发内网主机路由。

5 通过上述方法由南到北的流量在没有下一跳的情况下控制器可以在传统控制平面上学到下一跳的出口信息然后下发外网方向路由；由北向南进入数据中心内部的流量在没有匹配到精确流表的情况下数据报文通过上传到传统控制平面的协议栈然后再桥接到 SDN 相关协议栈并触发控制器下发精确流表。

10 步骤 105：基于所述二层交换设备上 ARP 报文转发机制，对 ARP 报文进行转发。

本申请实施例中，所述基于所述二层交换设备上 ARP 报文转发机制，对 ARP 报文进行转发，包括：

15 所述二层交换设备接收到服务器发送的 ARP 报文时，将所述 ARP 报文发送给所述三层交换设备进行 ARP 学习，以及将所述 ARP 报文发送给控制器进行 ARP 学习。

图 2 为本申请实施例的实现双控制平面的方法的流程示意图二，本实施例中的实现双控制平面的方法应用于控制器侧，如图 2 所示，所述方法包括以下步骤：

20 步骤 201：在 SDN 域配置好出口路由器的默认路由后，控制器主动触发所述默认路由的下一跳 ARP 学习。

具体地，SDN 域配置好出口路由器的默认路由后，控制器发送 ARP 请求给所述三层交换设备的 SDN 协议栈模块，其中，所述 SDN 协议栈模块在所述 ARP 请求上打上 SDN 协议栈虚拟接口的标记，并发送给收发包模块，通过所述收发包模块将所述 SDN 协议栈虚拟接口的标记替换为 IP 协议栈虚拟接口的标记，并发送给 IP 协议栈模块；所述 IP 协议栈模块对 ARP

25

进行学习，在 ARP 应答中打上 IP 协议栈虚拟接口的标记，并发送给收发包模块，通过所述收发包模块将所述 IP 协议栈虚拟接口的标记替换为 SDN 协议栈虚拟接口的标记，并发送给 SDN 协议栈模块；控制器接收所述 SDN 协议栈模块发送的 ARP 应答，根据所述 ARP 应答进行 ARP 学习。

5 上述方案中，对于北向流量，SDN 域配置默认出口路由后，控制器会主动触发默认路由下一跳 ARP 学习。控制器发送 ARP 请求给交换机的 SDN 协议栈，SDN 协议栈会将此报文带上 SDN_VP 的出口标记，然后这个报文会送个 IP 协议栈，IP 协议栈负责在传统的控制平面学到 ARP 后，再给交换机 SDN 协议栈发送 ARP 回复报文并打上 IPSTACK_VP 的标记，最后 SDN
10 控制器收到 ARP 应答后，学习到 ARP，下发默认路由的相关流表。这里实现两个协议栈的交换采用收发包模块换回的方式实现，即两个协议栈不直接交换信息通过收发包模块中转，这样的好处是不需要改动现有交换机协议栈收发包的处理模式。

步骤 202: 报文达到三层交换设备的桥接口后没有查到路由时，通过报
15 文触发所述控制器进行 ARP 学习；所述控制器将流表下发给所述三层交换设备。

具体地，当三层报文达到所述桥接口后没有查到路由时，所述控制器接收 SDN 协议栈模块发送的封装后的报文，其中，所述封装后的报文通过
20 以下方式生成：当三层报文达到所述桥接口后没有查到路由时，将所述报文通过网段路由发送给 IP 协议栈模块；所述 IP 协议栈模块根据所述报文的
目的 IP 进行路由查询，如果查询到 IP 协议栈虚拟接口的出口路由，则所述 IP 协议栈模块在所述报文上打上 IP 协议栈虚拟接口的标记；所述 IP 协议栈模块将打上 IP 协议栈虚拟接口的标记的报文发送给收发包模块，通过所
25 述收发包模块将所述 IP 协议栈虚拟接口的标记替换为 SDN 协议栈虚拟接口的
标记，并发送给 SDN 协议栈模块；所述 SDN 协议栈模块对所述报文进

行封装，得到封装后的报文；所述控制器接收到所述封装后的报文后，进行 ARP 学习并更新流表；所述控制器将更新的流表发送至三层交换设备中的 SDN 协议栈模块。

上述方案中，对于南向流量最关键的处理是报文到达桥接口后没有查
5 到路由要触发控制学习主机 ARP，然后下发给交换机下发流表。交换机在收到三层报文后如果没有路由会走网段路由上送给传统 IP 协议栈，IP 协议栈如果能匹配 IPSTACK_VP，就会打上 IPSTACK_VP 的标记，然后 SDN 协议栈就会收到这个报文，SDN 协议栈负责封 SDN 的协议报文（例如 OPEN_FLOW）后送 SDN 控制器。SDN 控制器收到后，触发 ARP 学习，
10 下发内网主机路由。

本申请实施例中，由南到北的流量在没有下一跳的情况下控制器可以在传统控制平面上学到下一跳的出口信息然后下发外网方向路由；由北向南进入数据中心内部的流量在没有匹配到精确流表的情况下数据报文通过
15 上传到传统控制平面的协议栈然后再桥接到 SDN 相关协议栈并触发控制器下发精确流表。

本申请实施例中，所述控制器周期性地比较本地路由表和交换设备上的路由表，当所述本地路由表中的表项与交换设备上的路由表中的表项不一致时，将所述交换机上的表项更新至本地。

图 3 为本申请实施例的实现双控制平面的方法的流程示意图三，本示
20 例中的实现双控制平面的方法应用于交换设备侧，如图 3 所示，所述方法包括以下步骤：

步骤 301：二层交换设备从业务接口接收 ARP 报文。

本申请实施例中，首先对所述二层交换设备进行二层 VXLAN 的配置；将所述二层交换设备连接至控制器，具体可参照图 4 所示的网络拓扑图。

25 步骤 302：所述二层交换设备将所述 ARP 报文发送给三层交换设备时，

同时将所述 ARP 报文发送给控制器。

本申请实施例中，二层交换设备（即与 SERVER 连接的网络设备）对下行口（与其他网络设备连接）和上行口（与 SERVER 连接的口）做不同的区分。对上行口，对 ARP 报文需要支持在普通传统控制平面转发（本地
5 VLAN 内的 LOCAL 转发以及进入 VXLAN 隧道终端（VTEP, VXLAN Tunnel End Point）隧道转发）的同时，还要对该 ARP 报文复制一份给控制器，对下行口，对 ARP 报文是遵循普通传统控制面的转发，无需复制和上送控制器。

通过上述方案，接入网络设备将 SERVER 发送的 ARP 报文给上游网络
10 设备使上行网络设备在传统平面内学习 ARP 的同时还上送到控制器，这样在控制器和传统平面都能学习到 ARP，从而实现双备份。

下面结合具体场景对本申请实施例的方案做进一步详细描述，以下以 OPENFLOW 协议类型的控制器为例进行说明。

图 4 为本申请的技术背景，服务器 A、B、C、D 分别单网卡接入网络
15 设备 C、D、G、H，与网络设备 C、D、G、H 的接口为 SDN 实例口。网络设备 C 与网络设备 D 做堆叠配置，网络设备 C 负载分担连接网络设备 A 和网络设备 B，网络设备 D 负载分担连接网络设备 A 和网络设备 B。传统控制平面上配置网络设备 C、D 与网络设备 A、B 之间的 VXLAN 隧道，C/D 做为 L2 VTEP 接入，A 和 B 做 L3 VXLAN 网关。受控制器控制的端口
20 分别是网络设备 C 以及网络设备 D 的 SDN 实例口以及网络设备 A 和网络设备 B 的桥接口。桥接口上启动路由协议，控制器可以将内部 SDN 网络的路由发布到外网路由器（出口路由器）上。由北向的流量在网络设备 C 和 D 上走 VXLAN 的 L2 接入，报文上隧道后进入网络设备 A/B，在此设备上解封装后查路由发送给外部路由器。由南到北的流量在网络设备 A/B 上
25 查路由走 L3 VXLAN 封装到隧道，到达网络 C/D 后 VXLAN 解封装后走 L2

转发到 ServerA/ServerB。

图 5 为本申请所述的网络设备的主要部件图，主要包括数据配置模块、协议栈模块、收发包模块、转发控制模块。其中，协议栈模块分为 SDN 相关协议栈（例如 OPENFLOW 协议栈）和传统控制平面协议栈（例如 IPSTACK），SDN 相关协议栈负责将收到的报文封装成 PACKET_IN 报文发送给控制器或者解析控制器发送给交换机的协议报文，然后给网络设备下发控制器的指令；传统控制平面协议栈负责 IP 协议栈的统一处理。这里由于采用双平面控制，桥接口即有传统业务流量又有 SDN 流量，所以传统协议栈和 SDN 相关协议栈之间存在一些交互。交互是借助传统协议栈虚接口以及 SDN 控制平面虚拟接口即 IPSTACK_VP 以及 SDN_VP 实现的。桥接口的物理端口和这两个虚拟接口的绑定关系在用户配置混合端口后下发的。配置模块负责提供用户层配置接口，并且将该配置下发给转发控制模块，转发控制模块负责记录该信息，并且向交换芯片下发 SDN 域的转发动作，本申请的系统框图如图 6 所示。

15

实施例一：网关设备双控制平面北向流量转发处理

如所述附图 4 所示，网络设备 C/D 为二层接入设备，网络设备 A/B 为三层网关设备。SERVER 在向外部网络发送流量之前，由于内部 IP（DIP）不是本网段的 IP，所以要请求网关的目的 MAC 信息。

对于北向流量（从数据中心内部网络到外部网络方向），SERVER 发出的报文的的目的 MAC 为三层网关的 MAC，即网络设备 A/B 的 MAC，目的 IP 为外网 IP。流量在接入网络设备上（即如附图 4 所示的网络设备 C/D）查找二层精确流表（MAC 表），如果查到走 VXLAN 隧道，网络设备 C/D 将报文封装 VXLAN 报文头然后传递给网络设备 A/B，报文在网络设备 A/B 上查找控制器下发的默认流表，出口指向外部网络。

25

在网络设备 A/B 上下发的默认流表由 SDN 域控制器下发，配置默认路

由后，控制器会主动触发下一跳 ARP 学习，整个学习的过程是从 SDN 控制平面向传统控制平面获取 ARP。控制器学习默认路由下一跳 ARP 学习的过程如附图 8 所示，这个过程分为六步，分别对应图中的 1、2、3、4、5、6:

5 (1) SDN 域配置默认出口路由，控制器下发 ARP 请求报文给 SDN 相关协议栈。

(2) SDN 相关协议栈将 ARP 请求报文（带 SDN_VP 出口标记）下发给交换机收发包模块。

10 (3) 交换机收发包模块收到 SDN 协议栈下发的报文，如果带 SDN_VP 出口标记，就会根据映射关系将该虚拟端口转成 IPSTACK_VP,然后上送传统 IP 协议栈平面。

(4) 传统 IP 协议栈收到 ARP 请求，会回复 ARP 应答报文。传统 IP 协议栈下发 ARP 应答报文（带 IPSTACK_VP 出口标记）给收发包模块。

15 (5) 收发包模块收到传统 IP 协议栈下发的报文包，如果带 IPSTACK_VP 出口标记，就会根据映射关系将改虚拟端口转成 SDN_VP 上送 SDN 协议栈，由 SDN 协议栈封装 PACKET_IN 头部后送 SDN 控制器。

(6) SDN 控制器收到 ARP 应答后，学习到 ARP，下发内网主机路由。

20 控制器下发了外部路由后，当 SERVER 发出的流量到达附图 4 所示的网络设备 C/D 后，在 C/D 设备封装 VXLAN 头进入 VXLAN 隧道指向所示网络设备 A/B，到达网络设备 A/B 后进行 VXLAN 解封装，然后查询控制器下发的默认路由进入外网。

实施例二：网关设备双控制器平面南向流量处理

25 传统面通过桥接口上开启的路由协议，将内网网段路由发布给连接的外网路由器。将发往 SDN 内网的流量引下来。

流量到达如附图所示的网络设备 A/B 时，如果能查到精确表项（主机

路由), 则直接按照该表项的出口以及封装信息进入 VXLAN 隧道, 将流量发给所述网络设备 C/D, 网络设备 C/D 将 VXLAN 报文解封装后, 发给 SERVER.

对于没有匹配主机路由的流量, 网络设备 A/B 需要将流量上送到传统 IP 协议栈平面, 然后通过传统协议栈平面通知触发 SDN 协议平面触发控制器学习 SDN 域内的主机流表, 如附图 9 过程如下:

在网络设备 A/B 的桥接口上收到的数据包, 转发模块如果查硬件路由表发现无主机路由, 查询虚机所属网段的网段路由上送 CPU.

(1) 收发包模块收到 CPU 发送的数据报文后, 将该数据报文上送到传统控制器平面的 IP 协议栈。

(2) 在传统控制平面的 IP 协议栈, 用报文的 DIP 查询路由, 如果匹配到 IPSTACK_VP 出口路由, 则传统控制平面的 IP 协议栈进程下发数据包 (带 IPSTACK_VP 出口标记) 给收发包模块。否则传统控制平面的 IP 协议栈进程触发传统面的 ARP 学习。

(3) 收发包模块收到传统控制平面的 IP 协议栈进程下发的数据包, 如果带 IPSTACK_VP 出口标记, 则收发包模块根据虚拟端口的映射关系将其转换成相应的 SDN_VP 然后发给 SDN 协议栈进程,

(4) 然后由其封装 pkt in 头部后送 SDN 控制器。

(5) SDN 控制器收到后, 触发 ARP 学习, 发送给 SDN 协议栈。

(6) SDN 相关协议栈将表项下发给转发模块。

上述过程完成后, 相同的报文由外网发往内网时, 报文进入网络设备 A/B 后, 就会查找精确表项 (主机路由) 封装 VXLAN 报文头进入隧道, 然后经由所述网络设备 C/D 解封装发给 SERVER.

25 实施例三: 接入网络设备 ARP 的学习处理

如图 4 所示, 在网络设备 C/D 的传统的控制平面上配置 L2 VXLAN 隧

道并且配置 VLAN 与 VNI 的映射关系。所述网络设备 C/D 为二层 VXLAN 接入设备，所述网络设备 A/B 为三层 VXLAN 网关设备。

当 SERVER 要向外部网络发送流量前，因为 DIP 不是本网段所以需要发送 ARP 请求解决网关 MAC。当网络设备 C/D 接收到 Server 发送过来的
5 ARP 请求时，对 ARP 报文封装 VXLAN 报文进入 VTEP 隧道，从而将这个封装了 VXLAN 报文头的 ARP 报文上送到了网络设备 A/B，然后网络设备 A/B 在传统的控制平面学习这个 VXLAN 报文的 ARP。

网络设备 C/D 在将该 ARP 报文引入 VTEP 隧道的同时还要将这个 ARP 报文复制一份上送到 SDN 相关协议栈进程，然后由该进程将报文按照
10 openflow 协议封装并上送给控制器，这样控制器就学到了这个 ARP。

通过上述方法，接入网络设备将 SERVER 发送的 ARP 报文给上游网络设备使上行网络设备在传统平面内学习 ARP 的同时还上送到控制器，这样在控制器和传统平面都能学习到 ARP，从而实现双备份，此外当虚拟机迁移时控制器也能在第一时间感知到虚机的变化。

15 对与下行口，即由网络设备 A/B 发给网络设备 C/D 的 ARP 报文遵循普通传统控制面的转发，无需复制和上送控制器。

实施例四：优先级控制

网络设备 A/B 上 VXLAN 的 ARP 的学习同时支持传统自学习以及控制器
20 器下发两个控制面，且自学习的优先级始终最高。如附图一所示，当网络设备 A/B 从业务端口收到 ARP 报文时，会按照传统自学习流程对该 ARP 进行动态学习，具体流程如下：

a) 若本地路由表里没有该 ARP 地址，则直接添加表项，FLAG 标记是动态学习。

25 b) 若本地已存在该 ARP 地址，且 MAC 和出端口信息一致，则：

i 若原有 FLAG 标记是动态学习的，则进行 timeout 的刷新；

ii 若原有 FLAG 标记为控制器下发的，则将表项的 FLAG 修改为动态学习。

c) 若本地已存在该 ARP 地址，但 MAC 或者出端口不一致，则覆盖原有表项，且将 FLAG 标记设置为动态学习。

5 当交换机从控制收到 ARP 流表添加消息时，也需要进行 ARP 的学习，具体过程如下：

a) 若本地路由表中没有该 ARP 地址，则直接添加表项，FLAG 标记为控制器下发；

10 b) 若本地已存在该 ARP 地址，且 MAC 和出端口信息一致，则不做任何处理。

c) 若本地已存在该 ARP 地址，但是 MAC 和出端口信息不一致，则

i 若原有的 FLAG 标记为动态学习的，不做任何处理；

ii 若原有的 FLAG 标记是控制器下发的，则覆盖原有的表项。

15 控制器对本身保存的 ARP 和交换机上的 ARP 表周期进行对比来保障两者之间的同步；若控制器有而交换机上没有则下发给交换机，如果两者都有，但是出口不一致，则控制器取回交换机上的覆盖本地的。

另外，网络设备 C/D 上 VXLAN 的 MAC 只受传统平面控制，通过自主学习得到，并不受控制器的控制。

20 图 10 为本申请实施例的实现双控制平面的装置的结构组成示意图一，如图 10 所示，所述装置包括：

25 第一配置单元 1001，配置为在三层交换设备上设置第一控制平面和第二控制平面的 ARP 学习机制，其中，所述第一控制平面的优先级高于所述第二控制平面的优先级；在三层交换设备上设置桥接口，所述桥接口绑定第一虚拟接口和第二虚拟接口，其中，所述第一虚拟接口与所述第一控制平面相对应，所述第二虚拟接口与所述第二控制平面相对应；

第二配置单元 1002, 配置为在二层交换设备上设置 ARP 报文转发机制;

第一处理单元 1003, 配置为基于所述三层交换设备上的 ARP 学习机制, 对 ARP 进行学习; 基于所述三层交换设备上的桥接口对第一控制平面与第二控制平面之间流量进行转发;

5 第二处理单元 1004, 配置为基于所述二层交换设备上 ARP 报文转发机制, 对 ARP 报文进行转发。

本申请实施例中, 所述第一控制平面的 ARP 学习机制为自学习机制, 所述第二控制平面的 ARP 学习机制为控制器下发机制。

本申请实施例中, 所述第一处理单元 1003 包括:

10 第一学习子单元, 配置为当所述三层交换设备接收到 ARP 报文时, 查询本地路由表里是否存在所述 ARP 报文对应的 ARP 地址; 如果所述本地路由表里不存在所述 ARP 地址, 则将所述 ARP 地址添加至所述本地路由表里, 并且将所述 ARP 地址对应的表项标记为自学习; 如果所述本地路由表里存在所述 ARP 地址, 且所述 ARP 地址对应的 MAC 地址与出端口对应的 MAC
15 地址一致时, 将所述 ARP 地址对应的表项标记为自学习并对所述 ARP 地址对应的表项进行更新时间的刷新; 如果所述本地路由表里存在所述 ARP 地址, 且所述 ARP 地址对应的 MAC 地址与出端口对应的 MAC 地址不一致时, 依据所述 ARP 报文对应的 ARP 地址对所述本地路由表里中对应的表项进行覆盖, 并且将所述 ARP 地址对应的表项标记为自学习。

20 本申请实施例中, 所述第一处理单元 1003 包括:

第二学习子单元, 配置为当所述三层交换设备从控制器接收到 ARP 流表添加消息时, 查询本地路由表里是否存在所述 ARP 流表添加消息对应的 ARP 地址; 如果所述本地路由表里不存在所述 ARP 地址, 则将所述 ARP 地址添加至所述本地路由表里, 并且将所述 ARP 地址对应的表项标记为控
25 制器下发; 如果所述本地路由表里存在所述 ARP 地址, 且所述 ARP 地址对

应的 MAC 地址与出端口对应的 MAC 地址一致时，则保持所述本地路由表不变；如果所述本地路由表里存在所述 ARP 地址，且所述 ARP 地址对应的 MAC 地址与出端口对应的 MAC 地址不一致时，如果所述本地路由表里对应的表项标记为自学习，则保持所述本地路由表不变，如果所述本地路由表里对应的表项标记为控制器下发，则依据所述 ARP 流表添加消息对应的 ARP 地址对所述本地路由表里中对应的表项进行覆盖。

本申请实施例中，所述第一虚拟接口为 IP 协议栈虚拟接口，所述第二虚拟接口为 SDN 协议栈虚拟接口；其中，

所述三层交换设备通过所述 IP 协议栈虚拟接口将 SDN 网络的路由发布至出口路由器；

所述三层交换设备通过所述 SDN 协议栈虚拟接口接收控制器下发的流表。

本申请实施例中，所述三层交换设备包括：SDN 协议栈模块、IP 协议栈模块、收发包模块；其中，

所述 SDN 协议栈模块接收所述控制器发送的 ARP 请求，所述 SDN 协议栈模块在所述 ARP 请求上打上 SDN 协议栈虚拟接口的标记，并发送给收发包模块，通过所述收发包模块将所述 SDN 协议栈虚拟接口的标记替换为 IP 协议栈虚拟接口的标记，并发送给 IP 协议栈模块；

所述 IP 协议栈模块对 ARP 进行学习，在 ARP 应答中打上 IP 协议栈虚拟接口的标记，并发送给收发包模块，通过所述收发包模块将所述 IP 协议栈虚拟接口的标记替换为 SDN 协议栈虚拟接口的标记，并发送给 SDN 协议栈模块；

所述 SDN 协议栈模块将所述 ARP 应答发送给控制器，触发所述控制器根据所述 ARP 应答进行 ARP 学习。

本申请实施例中，所述三层交换设备包括：SDN 协议栈模块、IP 协议

栈模块、收发包模块；其中，

针对南向流量的三层报文，当所述三层报文达到所述桥接口后没有查到路由时，将所述报文通过网段路由发送给 IP 协议栈模块；所述 IP 协议栈模块根据所述报文的目的地 IP 进行路由查询，如果查询到 IP 协议栈虚拟接口的出口路由，则所述 IP 协议栈模块在所述报文上打上 IP 协议栈虚拟接口的标记；如果未查询到 IP 协议栈虚拟接口的出口路由，则所述 IP 协议栈模块触发自学习机制的 APR 学习；

所述 IP 协议栈模块将打上 IP 协议栈虚拟接口的标记的报文发送给收发包模块，通过所述收发包模块将所述 IP 协议栈虚拟接口的标记替换为 SDN 协议栈虚拟接口的标记，并发送给 SDN 协议栈模块；

所述 SDN 协议栈模块对所述报文进行封装后，发送给所述控制器，触发所述控制器进行 ARP 学习；

所述 SDN 协议栈模块接收所述控制器下发的流表。

本申请实施例中，所述第二处理单元 1004，具体用于当所述二层交换设备接收到服务器发送的 ARP 报文时，将所述 ARP 报文发送给所述三层交换设备进行 ARP 学习，以及将所述 ARP 报文发送给控制器进行 ARP 学习。

在实际应用中，所述实现双控制平面的装置中的各个单元所实现的功能，均可由位于实现双控制平面的装置中的中央处理器（CPU，Central Processing Unit）、或微处理器（MPU，Micro Processor Unit）、或数字信号处理器（DSP，Digital Signal Processor）、或现场可编程门阵列（FPGA，Field Programmable Gate Array）等实现。

图 11 为本申请实施例的实现双控制平面的装置的结构组成示意图二，如图 11 所示，所述装置包括：

第一学习单元 1101，配置为在 SDN 域配置好出口路由器的默认路由后，控制器主动触发所述默认路由的下一跳 ARP 学习；

第二学习单元 1102，配置为报文达到三层交换设备的桥接口后没有查到路由时，通过报文触发所述控制器进行 ARP 学习；

流表下发单元 1103，配置为将流表下发给所述三层交换设备。

本申请实施例中，所述第一学习单元 1101，具体用于：SDN 域配置好出口路由器的默认路由后，发送 ARP 请求给所述三层交换设备的 SDN 协议栈模块，其中，所述 SDN 协议栈模块在所述 ARP 请求上打上 SDN 协议栈虚拟接口的标记，并发送给收发包模块，通过所述收发包模块将所述 SDN 协议栈虚拟接口的标记替换为 IP 协议栈虚拟接口的标记，并发送给 IP 协议栈模块；所述 IP 协议栈模块对 ARP 进行学习，在 ARP 应答中打上 IP 协议栈虚拟接口的标记，并发送给收发包模块，通过所述收发包模块将所述 IP 协议栈虚拟接口的标记替换为 SDN 协议栈虚拟接口的标记，并发送给 SDN 协议栈模块；接收所述 SDN 协议栈模块发送的 ARP 应答，根据所述 ARP 应答进行 ARP 学习。

本申请实施例中，所述第二学习单元 1102，具体用于：当三层报文达到所述桥接口后没有查到路由时，所述控制器接收 SDN 协议栈模块发送的封装后的报文，其中，所述封装后的报文通过以下方式生成：当三层报文达到所述桥接口后没有查到路由时，将所述报文通过网段路由发送给 IP 协议栈模块；所述 IP 协议栈模块根据所述报文的目 IP 进行路由查询，如果查询到 IP 协议栈虚拟接口的出口路由，则所述 IP 协议栈模块在所述报文上打上 IP 协议栈虚拟接口的标记；所述 IP 协议栈模块将打上 IP 协议栈虚拟接口的标记的报文发送给收发包模块，通过所述收发包模块将所述 IP 协议栈虚拟接口的标记替换为 SDN 协议栈虚拟接口的标记，并发送给 SDN 协议栈模块；所述 SDN 协议栈模块对所述报文进行封装，得到封装后的报文；接收到所述封装后的报文后，进行 ARP 学习并更新流表。

本申请实施例中，所述装置还包括第二学习单元，配置为周期性地比

较本地路由表和交换设备上的路由表，当所述本地路由表中的表项与交换设备上的路由表中的表项不一致时，将所述交换机上的表项更新至本地。

在实际应用中，所述实现双控制平面的装置中的各个单元所实现的功能，均可由位于实现双控制平面的装置中的 CPU、或 MPU、或 DSP、或
5 FPGA 等实现。

本领域内的技术人员应明白，本申请的实施例可提供为方法、系统、或计算机程序产品。因此，本申请可采用硬件实施例、软件实施例、或结合软件和硬件方面的实施例的形式。而且，本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质（包括但不限于磁盘
10 存储器和光学存储器等）上实施的计算机程序产品的形式。

本申请是参照根据本申请实施例的方法、设备（系统）、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、
15 嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器，使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理
20 设备以特定方式工作的计算机可读存储器中，使得存储在该计算机可读存储器中的指令产生包括指令装置的制品，该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上，使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机
25 实现的实现，从而在计算机或其他可编程设备上执行的指令提供用于实现

在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

相应地，本发明实施例还提供一种计算机存储介质，其中存储有计算机程序，该计算机程序配置为执行本发明实施例的实现双控制平面的方法。

5 以上所述，仅为本申请的较佳实施例而已，并非用于限定本申请的保护范围。

工业实用性

本申请实施例的技术方案，在三层交换设备上设置第一控制平面和第二控制平面的 ARP 学习机制，其中，所述第一控制平面的优先级高于所述
10 第二控制平面的优先级；在三层交换设备上设置桥接口，所述桥接口绑定第一虚拟接口和第二虚拟接口，其中，所述第一虚拟接口与所述第一控制平面相对应，所述第二虚拟接口与所述第二控制平面相对应；在二层交换设备上设置 ARP 报文转发机制；基于所述三层交换设备上的 ARP 学习机制，对 ARP 进行学习；基于所述三层交换设备上的桥接口对第一控制平面与第
15 二控制平面之间流量进行转发；基于所述二层交换设备上 ARP 报文转发机制，对 ARP 报文进行转发。采用本申请实施例的技术方案，利用 SDN 网络带来的便利的同时还增加了整个网络的稳定性，当控制器发生故障或者掉线时整个网络还能够依据传统控制平面保持路由的顺利进行，在数据中心领域中有极高的实用性。

20

权利要求书

1、一种实现双控制平面的方法，所述方法包括：

在三层交换设备上设置第一控制平面和第二控制平面的地址解析协议 ARP 学习机制，其中，所述第一控制平面的优先级高于所述第二控制
5 平面的优先级；

在三层交换设备上设置桥接口，所述桥接口绑定第一虚拟接口和第二虚拟接口，其中，所述第一虚拟接口与所述第一控制平面对应，所述
第二虚拟接口与所述第二控制平面对应；

在二层交换设备上设置 ARP 报文转发机制；

10 基于所述三层交换设备上的 ARP 学习机制，对 ARP 进行学习；基于所述三层交换设备上的桥接口对第一控制平面与第二控制平面之间流量进行转发；

基于所述二层交换设备上 ARP 报文转发机制，对 ARP 报文进行转发。

2、根据权利要求 1 所述的方法，其中，所述第一控制平面的 ARP
15 学习机制为自学习机制，所述第二控制平面的 ARP 学习机制为控制器下发机制。

3、根据权利要求 2 所述的方法，其中，所述基于所述三层交换设备上的 ARP 学习机制，对 ARP 进行学习，包括：

20 当所述三层交换设备接收到 ARP 报文时，查询本地路由表里是否存在所述 ARP 报文对应的 ARP 地址；

如果所述本地路由表里不存在所述 ARP 地址，则将所述 ARP 地址添加至所述本地路由表里，并且将所述 ARP 地址对应的表项标记为自学习；

如果所述本地路由表里存在所述 ARP 地址，且所述 ARP 地址对应的 MAC 地址与出端口对应的 MAC 地址一致时，将所述 ARP 地址对应的表
25 项标记为自学习并对所述 ARP 地址对应的表项进行更新时间的刷新；

如果所述本地路由表里存在所述 ARP 地址,且所述 ARP 地址对应的 MAC 地址与出端口对应的 MAC 地址不一致时,依据所述 ARP 报文对应的 ARP 地址对所述本地路由表里中对应的表项进行覆盖,并且将所述 ARP 地址对应的表项标记为自学习。

5 4、根据权利要求 2 所述的方法,其中,所述基于所述三层交换设备上的 ARP 学习机制,对 ARP 进行学习,包括:

当所述三层交换设备从控制器接收到 ARP 流表添加消息时,查询本地路由表里是否存在所述 ARP 流表添加消息对应的 ARP 地址;

10 如果所述本地路由表里不存在所述 ARP 地址,则将所述 ARP 地址添加至所述本地路由表里,并且将所述 ARP 地址对应的表项标记为控制器下发;

如果所述本地路由表里存在所述 ARP 地址,且所述 ARP 地址对应的 MAC 地址与出端口对应的 MAC 地址一致时,则保持所述本地路由表不变;

15 如果所述本地路由表里存在所述 ARP 地址,且所述 ARP 地址对应的 MAC 地址与出端口对应的 MAC 地址不一致时,如果所述本地路由表里对应的表项标记为自学习,则保持所述本地路由表不变,如果所述本地路由表里对应的表项标记为控制器下发,则依据所述 ARP 流表添加消息对应的 ARP 地址对所述本地路由表里中对应的表项进行覆盖。

20 5、根据权利要求 1 所述的方法,其中,所述第一虚拟接口为 IP 协议栈虚拟接口,所述第二虚拟接口为软件定义网络 SDN 协议栈虚拟接口,其中,

所述三层交换设备通过所述 IP 协议栈虚拟接口将 SDN 网络的路由发布至出口路由器;

25 所述三层交换设备通过所述 SDN 协议栈虚拟接口接收控制器下发的

流表。

6、根据权利要求 5 所述的方法，其中，所述三层交换设备包括：SDN 协议栈模块、IP 协议栈模块、收发包模块；

所述基于所述三层交换设备上的桥接口对第一控制平面与第二控制
5 平面之间流量进行转发，包括：

SDN 协议栈模块接收所述控制器发送的 ARP 请求，所述 SDN 协议栈模块在所述 ARP 请求上打上 SDN 协议栈虚拟接口的标记，并发送给收发包模块，通过所述收发包模块将所述 SDN 协议栈虚拟接口的标记替换为 IP 协议栈虚拟接口的标记，并发送给 IP 协议栈模块；

10 所述 IP 协议栈模块对 ARP 进行学习，在 ARP 应答中打上 IP 协议栈虚拟接口的标记，并发送给收发包模块，通过所述收发包模块将所述 IP 协议栈虚拟接口的标记替换为 SDN 协议栈虚拟接口的标记，并发送给 SDN 协议栈模块；

所述 SDN 协议栈模块将所述 ARP 应答发送给控制器，触发所述控制
15 器根据所述 ARP 应答进行 ARP 学习。

7、根据权利要求 5 所述的方法，其中，所述三层交换设备包括：SDN 协议栈模块、IP 协议栈模块、收发包模块；

所述基于所述三层交换设备上的桥接口对第一控制平面与第二控制平面之间流量进行转发，包括：

20 针对南向流量的三层报文，当所述三层报文达到所述桥接口后没有查到路由时，将所述报文通过网段路由发送给 IP 协议栈模块；所述 IP 协议栈模块根据所述报文的的目的 IP 进行路由查询，如果查询到 IP 协议栈虚拟接口的出口路由，则所述 IP 协议栈模块在所述报文上打上 IP 协议栈虚拟接口的标记；如果未查询到 IP 协议栈虚拟接口的出口路由，则所述 IP
25 协议栈模块触发自学习机制的 APR 学习；

所述 IP 协议栈模块将打上 IP 协议栈虚拟接口的标记的报文发送给收发包模块，通过所述收发包模块将所述 IP 协议栈虚拟接口的标记替换为 SDN 协议栈虚拟接口的标记，并发送给 SDN 协议栈模块；

所述 SDN 协议栈模块对所述报文进行封装后，发送给所述控制器，
5 触发所述控制器进行 ARP 学习；

所述 SDN 协议栈模块接收所述控制器下发的流表。

8、根据权利要求 1 所述的方法，其中，所述基于所述二层交换设备上 ARP 报文转发机制，对 ARP 报文进行转发，包括：

所述二层交换设备接收到服务器发送的 ARP 报文时，将所述 ARP
10 报文发送给所述三层交换设备进行 ARP 学习，以及将所述 ARP 报文发送给控制器进行 ARP 学习。

9、一种实现双控制平面的方法，所述方法包括：

在 SDN 域配置好出口路由器的默认路由后，控制器主动触发所述默认路由的下一跳 ARP 学习；

15 报文达到三层交换设备的桥接口后没有查到路由时，通过报文触发所述控制器进行 ARP 学习；所述控制器将流表下发给所述三层交换设备。

10、根据权利要求 9 所述的方法，其中，所述在 SDN 域配置好出口路由器的默认路由后，控制器主动触发所述默认路由的下一跳 ARP 学习，
包括：

20 SDN 域配置好出口路由器的默认路由后，控制器发送 ARP 请求给所述三层交换设备的 SDN 协议栈模块，其中，所述 SDN 协议栈模块在所述 ARP 请求上打上 SDN 协议栈虚拟接口的标记，并发送给收发包模块，通过所述收发包模块将所述 SDN 协议栈虚拟接口的标记替换为 IP 协议栈虚拟接口的标记，并发送给 IP 协议栈模块；所述 IP 协议栈模块对 ARP
25 进行学习，在 ARP 应答中打上 IP 协议栈虚拟接口的标记，并发送给收发

包模块，通过所述收发包模块将所述 IP 协议栈虚拟接口的标记替换为 SDN 协议栈虚拟接口的标记，并发送给 SDN 协议栈模块；

控制器接收所述 SDN 协议栈模块发送的 ARP 应答，根据所述 ARP 应答进行 ARP 学习。

- 5 11、根据权利要求 9 所述的方法，其中，所述报文达到三层交换设备的桥接口后没有查到路由时，通过报文触发所述控制器进行 ARP 学习，包括：

当三层报文达到所述桥接口后没有查到路由时，所述控制器接收 SDN 协议栈模块发送的封装后的报文，其中，所述封装后的报文通过以下方式生成：当三层报文达到所述桥接口后没有查到路由时，将所述报
10 文通过网段路由发送给 IP 协议栈模块；所述 IP 协议栈模块根据所述报文的
目的 IP 进行路由查询，如果查询到 IP 协议栈虚拟接口的出口路由，则
所述 IP 协议栈模块在所述报文上打上 IP 协议栈虚拟接口的标记；所述 IP
协议栈模块将打上 IP 协议栈虚拟接口的标记的报文发送给收发包模块，
15 通过所述收发包模块将所述 IP 协议栈虚拟接口的标记替换为 SDN 协议栈
虚拟接口的标记，并发送给 SDN 协议栈模块；所述 SDN 协议栈模块对
所述报文进行封装，得到封装后的报文；

所述控制器接收到所述封装后的报文后，进行 ARP 学习并更新流表；

所述控制器将更新的流表发送至三层交换设备中的 SDN 协议栈模
20 块。

- 12、根据权利要求 9 所述的方法，其中，所述方法还包括：

所述控制器周期性地比较本地路由表和交换设备上的路由表，当所
述本地路由表中的表项与交换设备上的路由表中的表项不一致时，将所
述交换机上的表项更新至本地。

- 25 13、一种实现双控制平面的装置，所述装置包括：

第一配置单元，配置为在三层交换设备上设置第一控制平面和第二控制平面的 ARP 学习机制，其中，所述第一控制平面的优先级高于所述第二控制平面的优先级；在三层交换设备上设置桥接口，所述桥接口绑定第一虚拟接口和第二虚拟接口，其中，所述第一虚拟接口与所述第一控制平面相对应，所述第二虚拟接口与所述第二控制平面相对应；

第二配置单元，配置为在二层交换设备上设置 ARP 报文转发机制；

第一处理单元，配置为基于所述三层交换设备上的 ARP 学习机制，对 ARP 进行学习；基于所述三层交换设备上的桥接口对第一控制平面与第二控制平面之间流量进行转发；

第二处理单元，配置为基于所述二层交换设备上 ARP 报文转发机制，对 ARP 报文进行转发。

14、根据权利要求 13 所述的装置，其中，所述第一控制平面的 ARP 学习机制为自学习机制，所述第二控制平面的 ARP 学习机制为控制器下发机制。

15、根据权利要求 14 所述的装置，其中，所述第一处理单元包括：

第一学习子单元，配置为当所述三层交换设备接收到 ARP 报文时，查询本地路由表里是否存在所述 ARP 报文对应的 ARP 地址；如果所述本地路由表里不存在所述 ARP 地址，则将所述 ARP 地址添加至所述本地路由表里，并且将所述 ARP 地址对应的表项标记为自学习；如果所述本地路由表里存在所述 ARP 地址，且所述 ARP 地址对应的 MAC 地址与出端口对应的 MAC 地址一致时，将所述 ARP 地址对应的表项标记为自学习并对所述 ARP 地址对应的表项进行更新时间的刷新；如果所述本地路由表里存在所述 ARP 地址，且所述 ARP 地址对应的 MAC 地址与出端口对应的 MAC 地址不一致时，依据所述 ARP 报文对应的 ARP 地址对所述本地路由表里中对应的表项进行覆盖，并且将所述 ARP 地址对应的表项标

记为自学习。

16、根据权利要求 14 所述的装置，其中，所述第一处理单元包括：

第二学习子单元，配置为当所述三层交换设备从控制器接收到 ARP 流表添加消息时，查询本地路由表里是否存在所述 ARP 流表添加消息对
5 应的 ARP 地址；如果所述本地路由表里不存在所述 ARP 地址，则将所述 ARP 地址添加至所述本地路由表里，并且将所述 ARP 地址对应的表项标记为控制器下发；如果所述本地路由表里存在所述 ARP 地址，且所述 ARP 地址对应的 MAC 地址与出端口对应的 MAC 地址一致时，则保持所述本地路由表不变；如果所述本地路由表里存在所述 ARP 地址，且所述
10 ARP 地址对应的 MAC 地址与出端口对应的 MAC 地址不一致时，如果所述本地路由表里对应的表项标记为自学习，则保持所述本地路由表不变，如果所述本地路由表里对应的表项标记为控制器下发，则依据所述 ARP 流表添加消息对应的 ARP 地址对所述本地路由表里中对应的表项进行覆盖。

15 17、根据权利要求 13 所述的装置，其中，所述第一虚拟接口为 IP 协议栈虚拟接口，所述第二虚拟接口为 SDN 协议栈虚拟接口；其中，

所述三层交换设备通过所述 IP 协议栈虚拟接口将 SDN 网络的路由发布至出口路由器；

所述三层交换设备通过所述 SDN 协议栈虚拟接口接收控制器下发的
20 流表。

18、根据权利要求 17 所述的装置，其中，所述三层交换设备包括：SDN 协议栈模块、IP 协议栈模块、收发包模块；其中，

所述 SDN 协议栈模块接收所述控制器发送的 ARP 请求，所述 SDN 协议栈模块在所述 ARP 请求上打上 SDN 协议栈虚拟接口的标记，并发送
25 送给收发包模块，通过所述收发包模块将所述 SDN 协议栈虚拟接口的标

记替换为 IP 协议栈虚拟接口的标记，并发送给 IP 协议栈模块；

所述 IP 协议栈模块对 ARP 进行学习，在 ARP 应答中打上 IP 协议栈虚拟接口的标记，并发送给收发包模块，通过所述收发包模块将所述 IP 协议栈虚拟接口的标记替换为 SDN 协议栈虚拟接口的标记，并发送给
5 SDN 协议栈模块；

所述 SDN 协议栈模块将所述 ARP 应答发送给控制器，触发所述控制器根据所述 ARP 应答进行 ARP 学习。

19、根据权利要求 17 所述的装置，其中，所述三层交换设备包括：SDN 协议栈模块、IP 协议栈模块、收发包模块；其中，

10 针对南向流量的三层报文，当所述三层报文达到所述桥接口后没有查到路由时，将所述报文通过网段路由发送给 IP 协议栈模块；所述 IP 协议栈模块根据所述报文的目 IP 进行路由查询，如果查询到 IP 协议栈虚拟接口的出口路由，则所述 IP 协议栈模块在所述报文上打上 IP 协议栈虚拟接口的标记；如果未查询到 IP 协议栈虚拟接口的出口路由，则所述 IP
15 协议栈模块触发自学习机制的 APR 学习；

所述 IP 协议栈模块将打上 IP 协议栈虚拟接口的标记的报文发送给收发包模块，通过所述收发包模块将所述 IP 协议栈虚拟接口的标记替换为 SDN 协议栈虚拟接口的标记，并发送给 SDN 协议栈模块；

所述 SDN 协议栈模块对所述报文进行封装后，发送给所述控制器，
20 触发所述控制器进行 ARP 学习；

所述 SDN 协议栈模块接收所述控制器下发的流表。

20、根据权利要求 13 所述的装置，其中，所述第二处理单元，具体用于当所述二层交换设备接收到服务器发送的 ARP 报文时，将所述 ARP 报文发送给所述三层交换设备进行 ARP 学习，以及将所述 ARP 报文发送
25 给控制器进行 ARP 学习。

21、一种实现双控制平面的装置，所述装置包括：

第一学习单元，配置为在 SDN 域配置好出口路由器的默认路由后，
控制器主动触发所述默认路由的下一跳 ARP 学习；

第二学习单元，配置为报文达到三层交换设备的桥接口后没有查到
5 路由时，通过报文触发所述控制器进行 ARP 学习；

流表下发单元，配置为将流表下发给所述三层交换设备。

22、根据权利要求 21 所述的装置，其中，所述第一学习单元，具体
用于：SDN 域配置好出口路由器的默认路由后，发送 ARP 请求给所述三
层交换设备的 SDN 协议栈模块，其中，所述 SDN 协议栈模块在所述 ARP
10 请求上打上 SDN 协议栈虚拟接口的标记，并发送给收发包模块，通过所
述收发包模块将所述 SDN 协议栈虚拟接口的标记替换为 IP 协议栈虚拟接
口的标记，并发送给 IP 协议栈模块；所述 IP 协议栈模块对 ARP 进行学
习，在 ARP 应答中打上 IP 协议栈虚拟接口的标记，并发送给收发包模块，
通过所述收发包模块将所述 IP 协议栈虚拟接口的标记替换为 SDN 协议栈
15 虚拟接口的标记，并发送给 SDN 协议栈模块；接收所述 SDN 协议栈模
块发送的 ARP 应答，根据所述 ARP 应答进行 ARP 学习。

23、根据权利要求 21 所述的装置，其中，所述第二学习单元，具体
用于：当三层报文达到所述桥接口后没有查到路由时，所述控制器接收
SDN 协议栈模块发送的封装后的报文，其中，所述封装后的报文通过以
20 下方式生成：当三层报文达到所述桥接口后没有查到路由时，将所述报
文通过网段路由发送给 IP 协议栈模块；所述 IP 协议栈模块根据所述报
文的目的 IP 进行路由查询，如果查询到 IP 协议栈虚拟接口的出口路由，则
所述 IP 协议栈模块在所述报文上打上 IP 协议栈虚拟接口的标记；所述 IP
协议栈模块将打上 IP 协议栈虚拟接口的标记的报文发送给收发包模块，
25 通过所述收发包模块将所述 IP 协议栈虚拟接口的标记替换为 SDN 协议栈

虚拟接口的标记，并发送给 SDN 协议栈模块；所述 SDN 协议栈模块对所述报文进行封装，得到封装后的报文；接收到所述封装后的报文后，进行 ARP 学习并更新流表。

24、根据权利要求 21 所述的装置，其中，所述装置还包括：第三学习单元，配置为周期性地比较本地路由表和交换设备上的路由表，当所述本地路由表中的表项与交换设备上的路由表中的表项不一致时，将所述交换机上的表项更新至本地。

25、一种计算机存储介质，所述计算机存储介质中存储有计算机可执行指令，该计算机可执行指令配置为执行权利要求 1-8 任一项所述的实现双控制平面的方法，或者权利要求 9-12 任一项所述的实现双控制平面的方法。

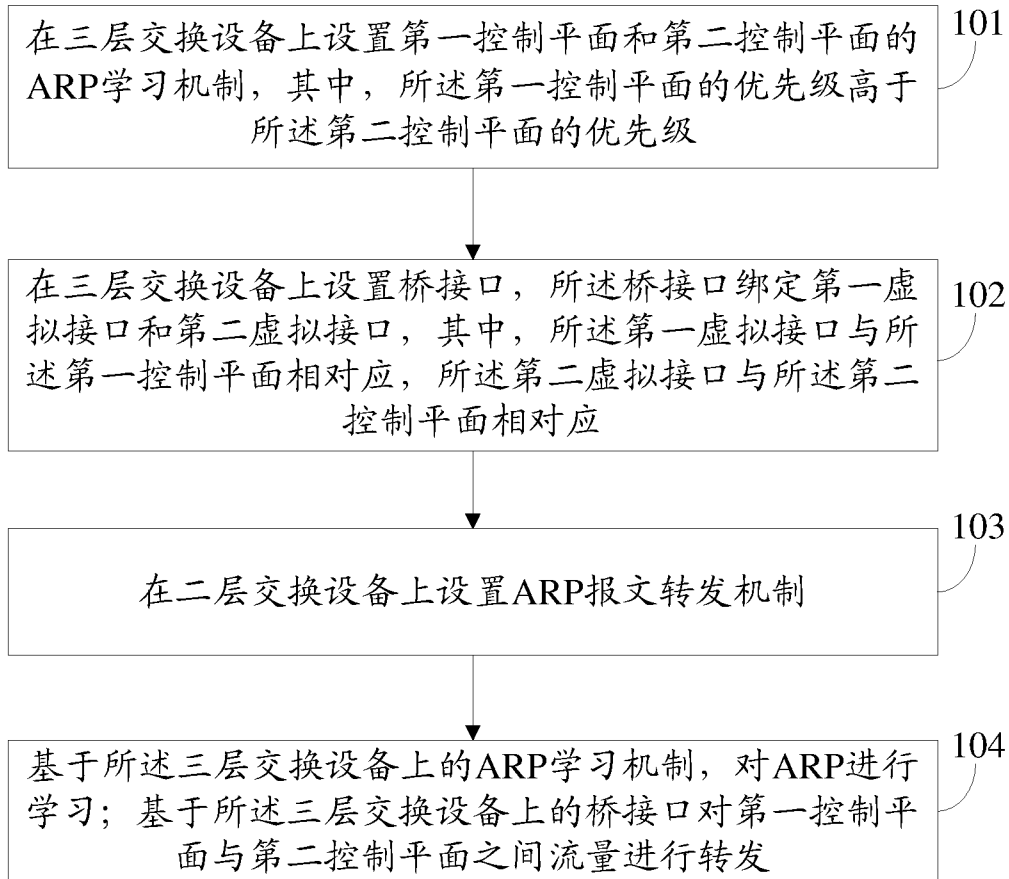


图 1

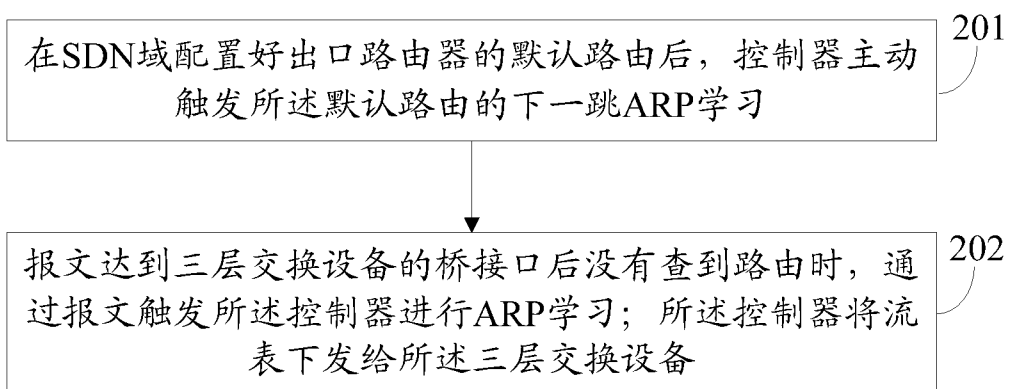


图 2

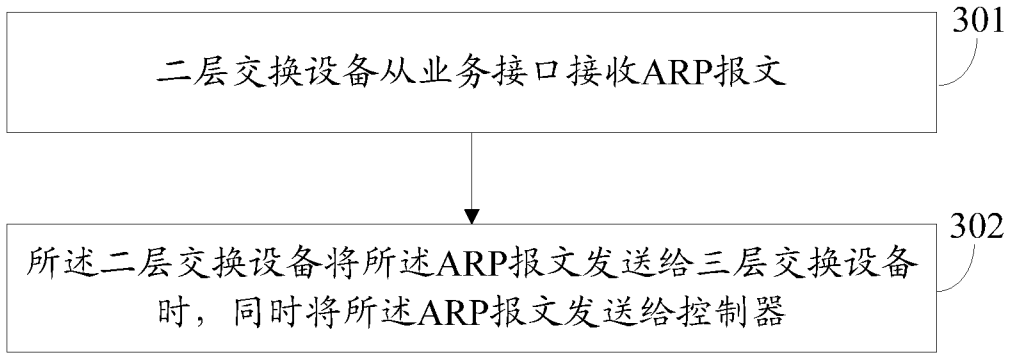


图 3

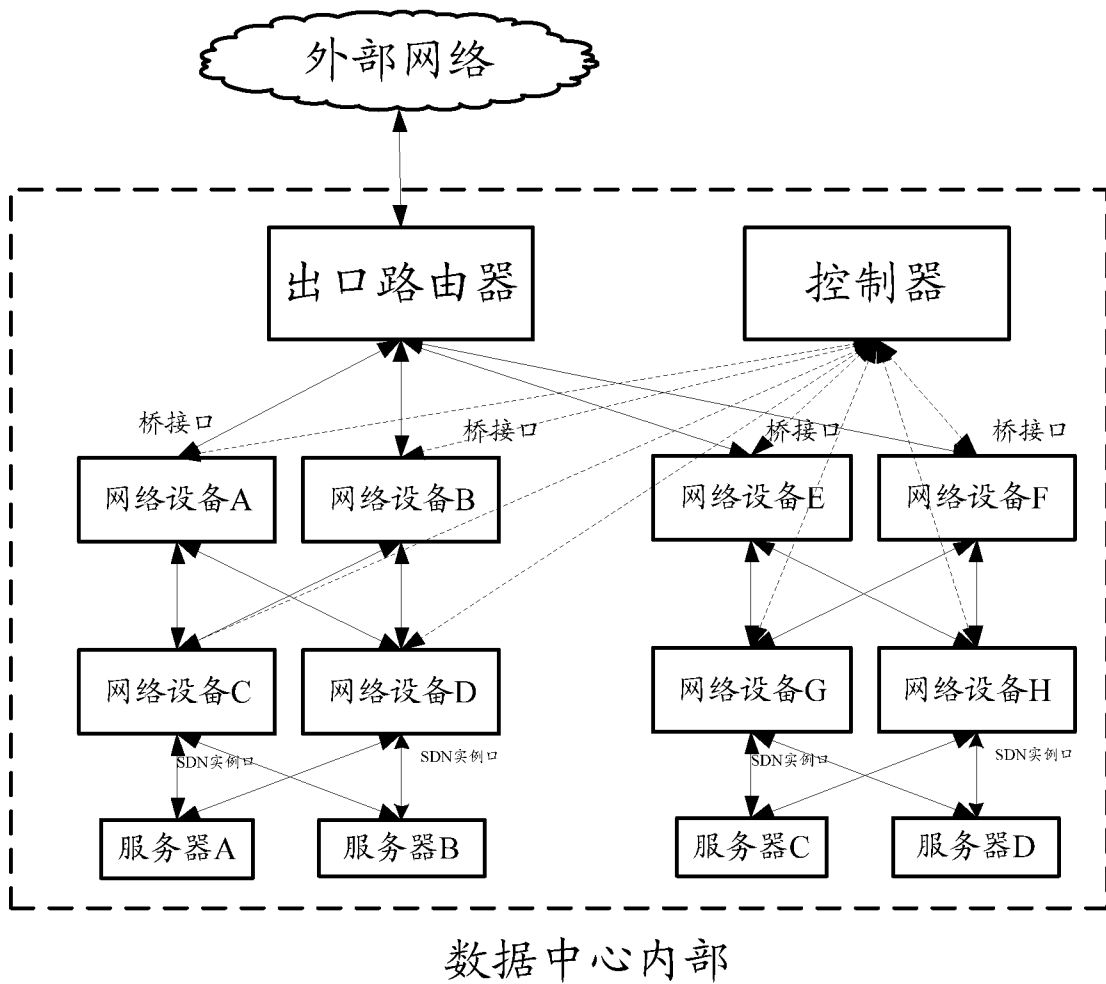
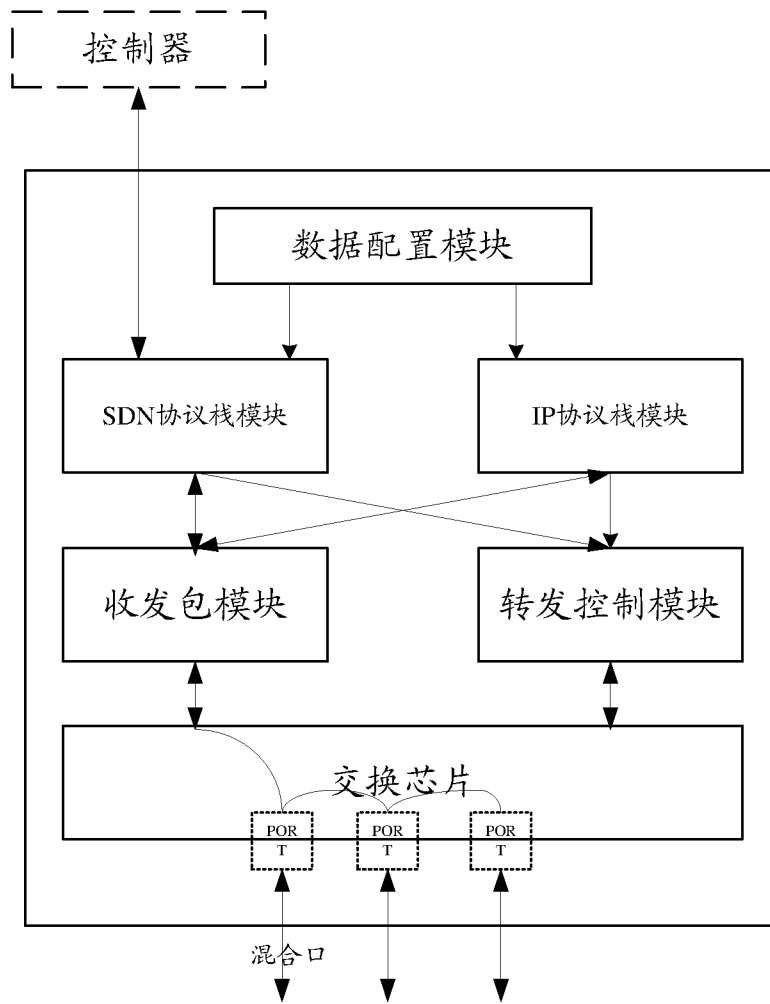
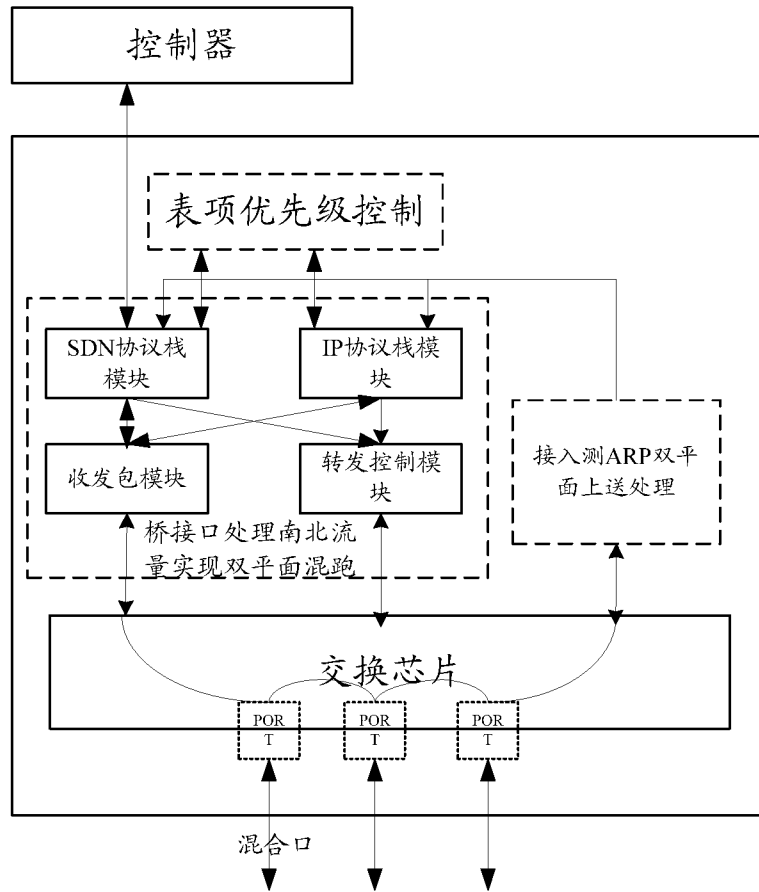


图 4



SDN交换机

图 5



系统框图

图 6

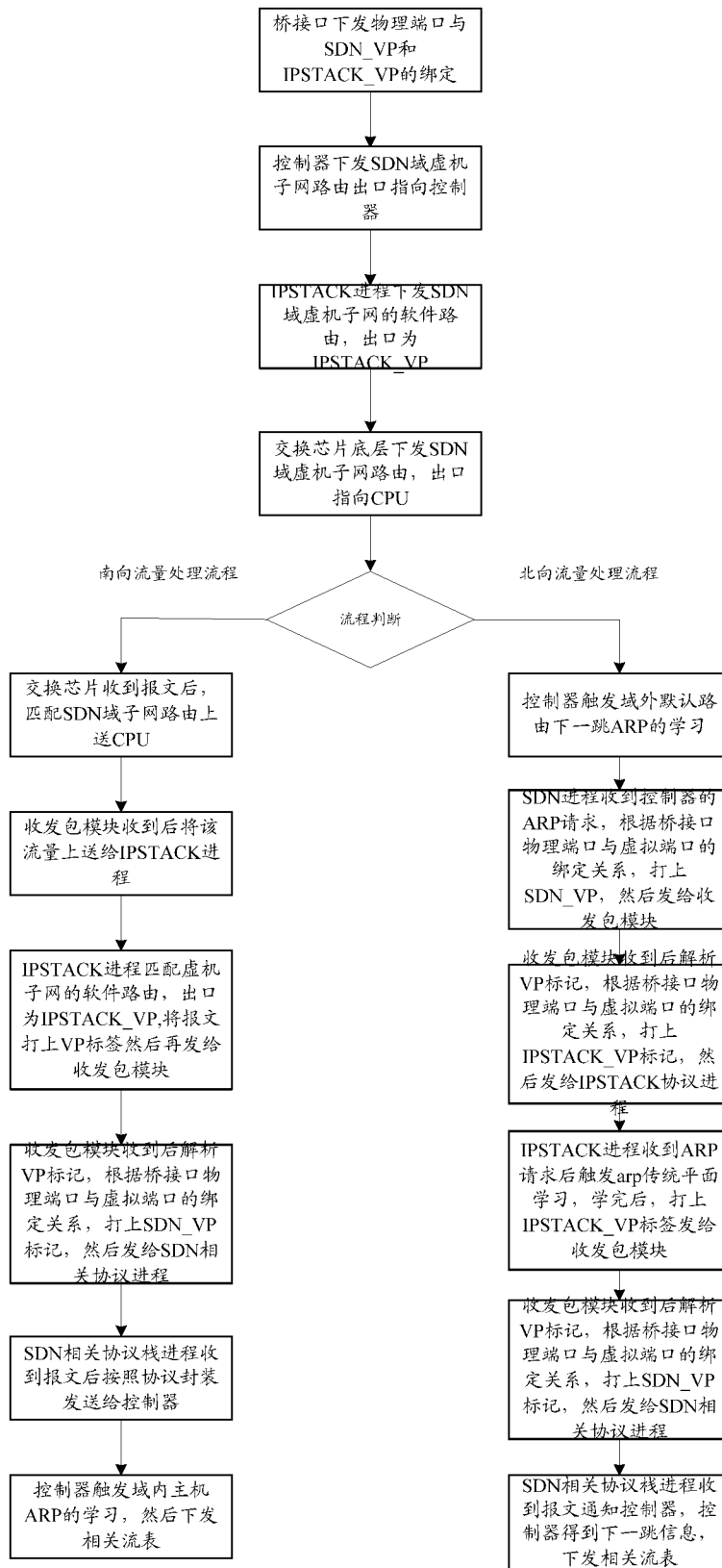
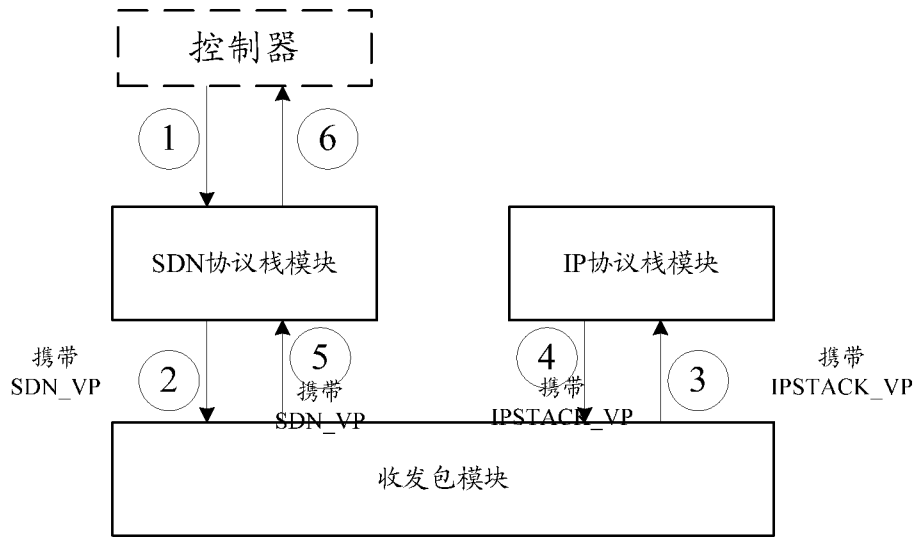
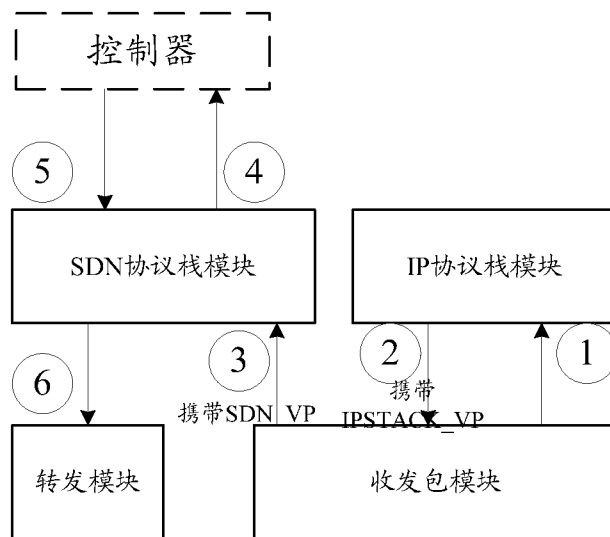


图 7



控制器触发学习外网路由下一跳流程

图 8



流量触发控制器学习 ARP

图 9

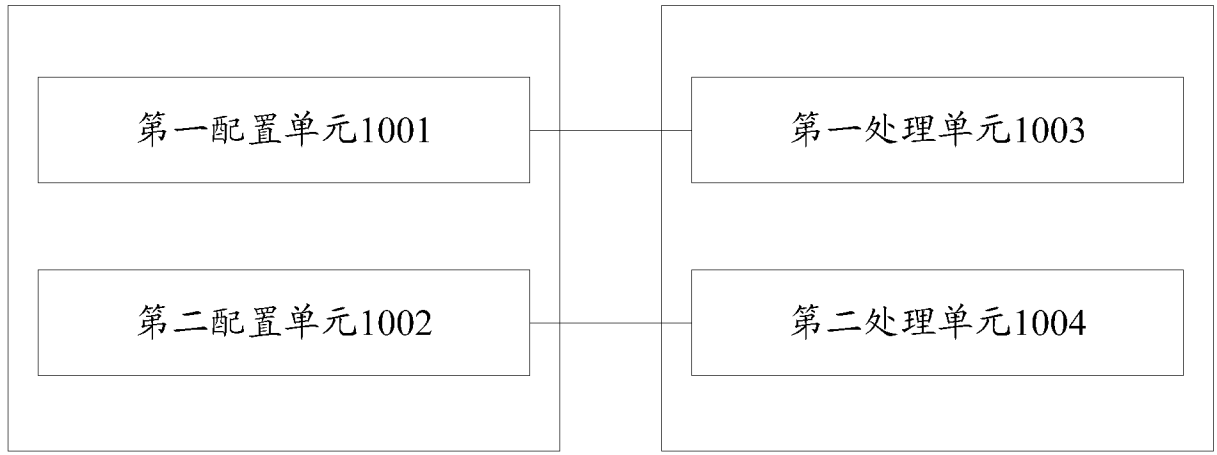


图 10

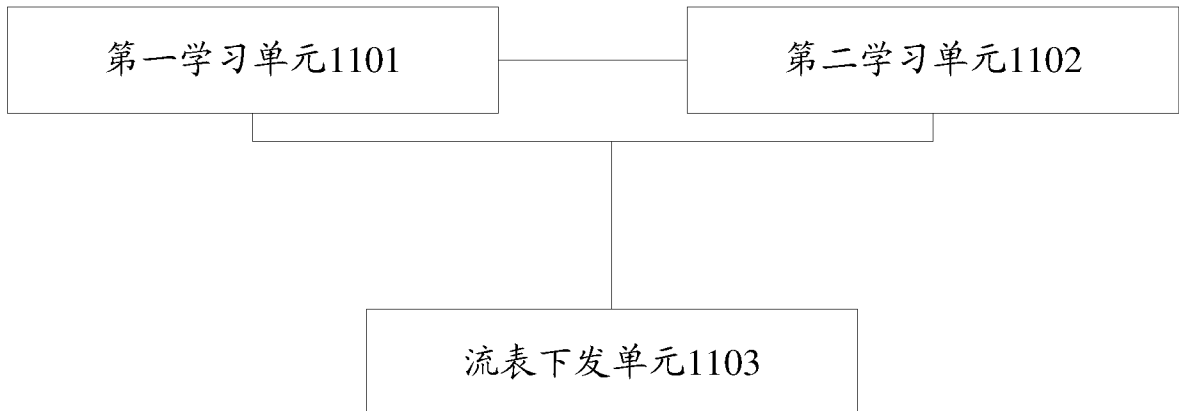


图 11

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2018/079340

A. CLASSIFICATION OF SUBJECT MATTER

H04L 12/46 (2006.01) i; H04L 12/741 (2013.01) i; H04L 29/12 (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)

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)

CNABS; CNTXT; VEN; USTXT; CNKI; IEEE: 三层, 二层, 交换, 网关, 软件定义网络, 第一, 第二, 双, 两, 控制平面, 控制器, 地址解析协议, 优先级, 自, 学习, 虚, 桥, 接口, 流表, 报文, SDN, ARP, layer, three, two, 3, 2, switch, gateway, Software Defined Network, first, second, double, two, control, plane, panel, controller, address resolution protocol, priority, self, learn+, virtual, bridge, interface, flow, chart, table, packet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 106210157 A (CHINA TELECOM CORPORATION LIMITED), 07 December 2016 (07.12.2016), see description, paragraphs [0100]-[0136] and [0194]	9-12, 21-25
A	CN 106105117 A (HUAWEI TECHNOLOGIES CO., LTD.), 09 November 2016 (09.11.2016), see entire document	1-25
A	WO 2016166403 A1 (CORIANT OY), 20 October 2016 (20.10.2016), see entire document	1-25
A	US 2016226817 A1 (ELECTRONICS & TELECOMMUNICATIONS RES INST), 04 August 2016 (04.08.2016), see entire document	1-25

Further documents are listed in the continuation of Box C.

See patent family annex.

<p>* Special categories of cited documents:</p> <p>“A” document defining the general state of the art which is not considered to be of particular relevance</p> <p>“E” earlier application or patent but published on or after the international filing date</p> <p>“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)</p> <p>“O” document referring to an oral disclosure, use, exhibition or other means</p> <p>“P” document published prior to the international filing date but later than the priority date claimed</p>	<p>“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</p> <p>“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</p> <p>“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</p> <p>“&” document member of the same patent family</p>
---	---

Date of the actual completion of the international search

01 June 2018

Date of mailing of the international search report

08 June 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

JIA, Yu

Telephone No. 86-(010)-62411258

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2018/079340

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 106210157 A	07 December 2016	None	
CN 106105117 A	09 November 2016	WO 2015096758 A1	02 July 2015
		EP 3053312 A4	02 November 2016
		EP 3053312 A1	10 August 2016
		US 2015188837 A1	02 July 2015
WO 2016166403 A1	20 October 2016	CN 107534567 A	02 January 2018
		EP 3284225 A1	21 February 2018
		US 2018102968 A1	12 April 2018
US 2016226817 A1	04 August 2016	KR 20160095554 A	11 August 2016

国际检索报告

国际申请号

PCT/CN2018/079340

<p>A. 主题的分类</p> <p>H04L 12/46(2006.01)i; H04L 12/741(2013.01)i; H04L 29/12(2006.01)i</p> <p>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</p>																	
<p>B. 检索领域</p> <p>检索的最低限度文献(标明分类系统和分类号)</p> <p>H04L</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>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))</p> <p>CNABS; CNTXT; VEN; USTXT; CNKI; IEEE; 三层, 二层, 交换, 网关, 软件定义网络, 第一, 第二, 双, 两, 控制平面, 控制器, 地址解析协议, 优先级, 自, 学习, 虚, 桥, 接口, 流表, 报文, SDN, ARP, layer, three, two, 3, 2, switch, gateway, Software Defined Network, first, second, double, two, control, plane, panel, controller, address resolution protocol, priority, self, learn+, virtual, bridge, interface, flow, chart, table, packet</p>																	
<p>C. 相关文件</p> <table border="1"> <thead> <tr> <th>类型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>CN 106210157 A (中国电信股份有限公司) 2016年 12月 7日 (2016 - 12 - 07) 参见说明书第 [0100]-[0136]、[0194]段</td> <td>9-12, 21-25</td> </tr> <tr> <td>A</td> <td>CN 106105117 A (华为技术有限公司) 2016年 11月 9日 (2016 - 11 - 09) 参见全文</td> <td>1-25</td> </tr> <tr> <td>A</td> <td>WO 2016166403 A1 (CORIANT OY) 2016年 10月 20日 (2016 - 10 - 20) 参见全文</td> <td>1-25</td> </tr> <tr> <td>A</td> <td>US 2016226817 A1 (ELECTRONICS & TELECOMMUNICATIONS RES INST) 2016年 8月 4日 (2016 - 08 - 04) 参见全文</td> <td>1-25</td> </tr> </tbody> </table> <p><input type="checkbox"/> 其余文件在C栏的续页中列出。 <input checked="" type="checkbox"/> 见同族专利附件。</p> <p>* 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</p>			类型*	引用文件, 必要时, 指明相关段落	相关的权利要求	X	CN 106210157 A (中国电信股份有限公司) 2016年 12月 7日 (2016 - 12 - 07) 参见说明书第 [0100]-[0136]、[0194]段	9-12, 21-25	A	CN 106105117 A (华为技术有限公司) 2016年 11月 9日 (2016 - 11 - 09) 参见全文	1-25	A	WO 2016166403 A1 (CORIANT OY) 2016年 10月 20日 (2016 - 10 - 20) 参见全文	1-25	A	US 2016226817 A1 (ELECTRONICS & TELECOMMUNICATIONS RES INST) 2016年 8月 4日 (2016 - 08 - 04) 参见全文	1-25
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求															
X	CN 106210157 A (中国电信股份有限公司) 2016年 12月 7日 (2016 - 12 - 07) 参见说明书第 [0100]-[0136]、[0194]段	9-12, 21-25															
A	CN 106105117 A (华为技术有限公司) 2016年 11月 9日 (2016 - 11 - 09) 参见全文	1-25															
A	WO 2016166403 A1 (CORIANT OY) 2016年 10月 20日 (2016 - 10 - 20) 参见全文	1-25															
A	US 2016226817 A1 (ELECTRONICS & TELECOMMUNICATIONS RES INST) 2016年 8月 4日 (2016 - 08 - 04) 参见全文	1-25															
国际检索实际完成的日期	国际检索报告邮寄日期																
2018年 6月 1日	2018年 6月 8日																
ISA/CN的名称和邮寄地址	受权官员																
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	贾煜																
传真号 (86-10)62019451	电话号码 86-(010)-62411258																

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/079340

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	106210157	A	2016年 12月 7日	无			
CN	106105117	A	2016年 11月 9日	WO	2015096758	A1	2015年 7月 2日
				EP	3053312	A4	2016年 11月 2日
				EP	3053312	A1	2016年 8月 10日
				US	2015188837	A1	2015年 7月 2日
WO	2016166403	A1	2016年 10月 20日	CN	107534567	A	2018年 1月 2日
				EP	3284225	A1	2018年 2月 21日
				US	2018102968	A1	2018年 4月 12日
US	2016226817	A1	2016年 8月 4日	KR	20160095554	A	2016年 8月 11日

表 PCT/ISA/210 (同族专利附件) (2015年1月)