



(12)发明专利

(10)授权公告号 CN 103703723 B

(45)授权公告日 2017.03.29

(21)申请号 201280038641.5

(22)申请日 2012.07.25

(65)同一申请的已公布的文献号
申请公布号 CN 103703723 A

(43)申请公布日 2014.04.02

(30)优先权数据
13/196717 2011.08.02 US

(85)PCT国际申请进入国家阶段日
2014.02.07

(86)PCT国际申请的申请数据
PCT/IB2012/053807 2012.07.25

(87)PCT国际申请的公布数据
W02013/017998 EN 2013.02.07

(73)专利权人 瑞典爱立信有限公司
地址 瑞典斯德哥尔摩

(72)发明人 K.耶达瓦利 R.米什拉

(74)专利代理机构 中国专利代理(香港)有限公司
72001

代理人 叶晓勇 汤春龙

(51)Int.Cl.
H04L 12/24(2006.01)
H04L 12/70(2006.01)

(56)对比文件
CN 1947390 A,2007.04.11,
US 2010077067 A1,2010.03.25,
MARTIN CASADO.《Rethinking Enterprise
Network Control》.《IEEE/ACM TRANSACTIONS
ON NETWORKING》.2009,第1270-1282页.
Nick Mckeown等.《OpenFlow: Enabling
Innovation in Campus Networks》.《URL:
[http://www.openflow.org/documents/
openflow-wp-latest.pdf](http://www.openflow.org/documents/openflow-wp-latest.pdf)》.2008,全文.

审查员 陈晓伟

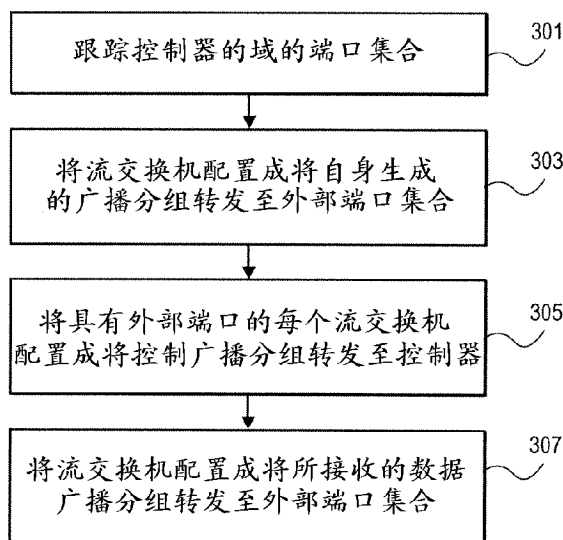
权利要求书2页 说明书9页 附图8页

(54)发明名称

分体式架构网络中的分组广播机制

(57)摘要

一种方法和系统在作为分体式架构网络中的域的控制器起作用的网络节点中实现。域包括多个流交换机,其中多个流交换机实现分体式架构网络的数据平面,并且控制器实现分体式架构网络的远离数据平面的控制平面。该方法和系统将多个流交换机配置成高效处置分体式架构网络的域中的每种类型的广播分组,而没有用广播分组对域泛洪,由此减少在域内转发的广播分组的数量。



1. 一种在作为分体式架构网络中的域的控制起作用的网络节点中实现的方法,所述域包括多个流交换机,其中所述多个流交换机实现所述分体式架构网络的数据平面,并且所述控制器实现所述分体式架构网络的远离所述数据平面的控制平面,所述方法将所述多个流交换机配置成高效处置所述分体式架构网络的所述域中的每种类型的广播分组,而没有用所述广播分组对所述域泛洪,由此减少在所述域内转发的广播分组的数量,所述方法包括以下步骤:

跟踪所述域的外部端口集合,其中每个外部端口将所述多个流交换机之一连接到所述域之外的装置;

将所述域中的所述多个流交换机中的每个流交换机配置成将自身生成的广播分组转发至具有所述域的所述外部端口集合中的至少一个外部端口的每个流交换机,而没有用所述自身生成的广播分组对所述域泛洪;

将所述多个流交换机中的具有所述域的所述外部端口集合中的至少一个外部端口的每个流交换机配置成将所接收的控制广播分组转发至所述域的所述控制器,而没有用所接收的控制广播分组对所述域泛洪;以及

将所述域中的所述多个流交换机配置成将所接收的数据广播分组转发至所述域的所述外部端口集合,而没有用所接收的数据广播分组对所述域泛洪。

2. 如权利要求1所述的方法,其中跟踪所述外部端口集合进一步包括以下步骤:标识所述域中的每个流交换机的每个活动端口。

3. 如权利要求2所述的方法,其中跟踪所述外部端口集合进一步包括以下步骤:标识每个活动端口是否是内部端口,所述内部端口将所述域中的第一流交换机连接到所述域中的第二流交换机。

4. 如权利要求3所述的方法,其中跟踪所述外部端口集合进一步包括以下步骤:通过比较所述域的所述内部端口与所述域的所述活动端口来标识所述域的每个外部端口,每个外部端口是所述域的活动非内部端口。

5. 如权利要求3所述的方法,其中标识所述内部端口进一步包括以下步骤:使用邻居发现协议标识内部端口。

6. 如权利要求1所述的方法,进一步包括以下步骤:

从所述域外的网络节点接收邻居发现协议NDP消息;以及

比较所述NDP消息的源与受控流交换机集合。

7. 如权利要求6所述的方法,进一步包括以下步骤:向外部端口列表添加通过其接收来自外部源的NDP消息的流交换机的入口端口。

8. 如权利要求1所述的方法,进一步包括以下步骤:由所述控制器将所接收的控制分组转发至具有所述外部端口集合中的外部端口的每个流交换机,每个流交换机通过每个外部端口转发所述控制分组。

9. 如权利要求1所述的方法,进一步包括以下步骤:

计算从具有外部端口的每个流交换机到具有外部端口的每个其它流交换机的流集合;以及

对所述流集合的路径中的每个流交换机编程以根据所述流集合转发所接收的数据广播分组。

10. 一种作为分体式架构网络中的域的控制起作用的网络节点,所述域包括多个流交换机,其中所述多个流交换机实现所述分体式架构网络的数据平面,并且所述控制器实现所述分体式架构网络的远离所述数据平面的控制平面,所述网络节点将所述多个流交换机配置成高效处置所述分体式架构网络中的每种类型的广播分组,而没有用广播分组对所述域泛洪,由此减少在所述域内转发的广播分组的数量,所述网络节点包括:

外部端口跟踪模块,跟踪所述域的外部端口集合,其中每个外部端口将所述多个流交换机之一连接到所述域之外的装置;

广播配置模块,在通信上耦合到所述外部端口跟踪模块,所述广播配置模块配置成将所述域中的所述多个流交换机中的每个流交换机配置成将自身生成的广播分组转发至具有所述域的所述外部端口集合中的至少一个外部端口的每个流交换机,而没有用所述自身生成的广播分组对所述域泛洪;所述广播配置模块配置成将所述多个流交换机中具有来自所述域的所述外部端口集合中的至少一个外部端口的每个流交换机配置成将所接收的控制广播分组转发至所述域的所述控制器,而没有对所述域泛洪,并且配置成将所述域的所述多个流交换机配置成将所接收的数据广播分组转发至所述域的所述外部端口集合,而没有用所接收的数据广播分组对所述域泛洪。

11. 如权利要求10所述的网络节点,进一步包括:广播处理模块,在通信上耦合到所述广播配置模块,所述广播处理模块配置成将所接收的广播分组转发至具有所述外部端口集合的所述流交换机。

12. 如权利要求10所述的网络节点,其中所述外部端口跟踪模块配置成标识所述域中的每个流交换机的每个活动端口。

13. 如权利要求12所述的网络节点,其中所述外部端口跟踪模块配置成标识每个活动端口是否是内部端口,所述内部端口将所述域中的第一流交换机连接到所述域中的第二流交换机。

14. 如权利要求13所述的网络节点,其中所述外部端口跟踪模块配置成通过比较所述域的所述内部端口与所述域的所述活动端口来标识所述域的每个外部端口,每个外部端口是所述域的活动非内部端口。

15. 如权利要求13所述的网络节点,其中所述外部端口跟踪模块配置成使用邻居发现协议标识内部端口。

16. 如权利要求11所述的网络节点,其中所述外部端口跟踪模块配置成从所述域外的网络节点接收邻居发现协议NDP消息,并且比较所述NDP消息的源与受控流交换机集合。

17. 如权利要求16所述的网络节点,其中所述外部端口跟踪模块配置成向外部端口列表添加通过其接收来自外部的源的NDP消息的流交换机的入口端口。

18. 如权利要求11所述的网络节点,其中所述广播处理模块配置成由所述控制器将所接收的控制分组转发至具有所述外部端口集合中的外部端口的每个流交换机,每个流交换机通过每个外部端口转发所述控制分组。

19. 如权利要求10所述的网络节点,其中所述广播配置模块配置成计算从具有外部端口的每个流交换机到具有外部端口的每个其它交换机的流集合,并配置成对所述流集合中的每个流交换机编程以根据所述流集合转发所接收的数据广播分组。

分体式架构网络中的分组广播机制

技术领域

[0001] 本发明的实施例涉及分体式架构网络(split architecture network)内的分组广播。具体地说,本发明的实施例涉及用于当处置要广播的分组时通过标识域的外部端口并对域的交换机中的适当流编程来优化该域中资源的使用的方法和系统。

背景技术

[0002] 分体式架构网络是网络架构的备选范例。分体式架构网络像传统网络架构一样由多个节点(诸如路由器或交换机)组成。然而,在此范例中,网络的控制平面功能性驻留在单个中央控制器处,该控制器可以在网络的单个节点中,如图1中所图示的。这与传统网络架构形成对比,在传统网络架构中控制平面功能性驻留在网络中的每个节点上。在分体式架构网络中,网络节点被剥夺了它们的控制平面功能性,并作为“流交换机”提供了基本数据平面功能性。如本文所使用的,“域”是流交换机在单个控制器控制下的分体式架构网络的区域。

[0003] 分体式架构网络的示例是基于OpenFlow的分体式架构网络。在OpenFlow分体式架构网络中,控制器和数据平面中的流交换机使用OpenFlow协议通信。OpenFlow协议提供了用于控制器的消息基础设施,以用适当流对数据平面中的流交换机编程以便进行分组转发和处理。控制器也可通过查询每个网络节点来获得网络的完整网络状态信息。如本文所使用的,“OpenFlow域”是流交换机在单个逻辑OpenFlow控制器控制下的OpenFlow分体式架构网络的区域。此单个逻辑OpenFlow控制器可实现为负责单个交换机域的控制器的集合或集群。

[0004] 根据OpenFlow协议规范,控制器使用以下原语对流交换机编程以便在数据平面中进行分组转发和处理:规则,其定义分组报头字段(诸如以太网报头字段、多协议标记交换(MPLS)报头字段和因特网协议(IP)报头字段)上的匹配;动作,它们与规则匹配关联;以及匹配规则的流的统计的收集。流交换机对入局分组应用控制器编程的规则,并在匹配的流上执行关联的动作,同时收集有关流的统计。图1图示了控制器与流交换机之间的这个基本OpenFlow协议接口。

[0005] 分体式架构网络的主要优点之一是,它们允许控制平面和数据平面中的独立优化。控制器的组件、功能或应用可独立于底层网络而优化,并且类似地,底层网络流交换机能力可独立于控制器组件、功能和应用而优化。

[0006] 有许多情形需要网络中的节点发送出广播分组。这些情形可被分成两类:自身生成的广播分组和所接收的广播分组。节点可自身生成用于特定协议应用的广播分组。自身生成的广播分组的重要示例是ARP REQUEST分组。地址解析协议(ARP)用于解析给定其IP地址的网络节点(路由器或交换机)的媒体接入控制(MAC)地址。为了查询网络节点的MAC地址,请求方发送出ARP REQUEST消息。此消息含有查询IP地址,并被广播给网络中的所有装置。IP地址匹配REQUEST消息中IP地址的装置用其MAC地址应答。

[0007] 节点还可接收需要被广播的分组。网络节点需要广播它在所有活动端口(包含控

制器端口)上接收的广播分组。如果网络节点用虚拟局域网(VLAN)配置,则它应该在VLAN内的所有活动端口上广播分组。由于转发分组的智能驻留在OpenFlow域中的控制器中,因此控制器需要对用于广播分组的底层流交换机适当编程。此配置的基本实现是将每个流交换机的转发表设置成在流交换机的所有活动端口上盲目转发广播分组。

发明内容

[0008] 一种方法在作为分体式架构网络中的域的控制器的网络节点中实现。该域包括多个流交换机,其中多个流交换机实现分体式架构网络的数据平面,并且控制器实现分体式架构网络的远离数据平面的控制平面。该方法将多个流交换机配置成高效处置分体式架构网络的域中的每种类型的广播分组,而没有用广播分组对域泛洪(flooding),由此减少在域内转发的广播分组数量。该方法包括以下步骤:跟踪该域的外部端口集合,其中每个外部端口将多个流交换机之一连接到该域外的装置;将该域中的多个流交换机中的每个流交换机配置成将自身生成的广播分组转发至具有该域的外部端口集合中的至少一个外部端口的每个流交换机,而没有用自身生成的广播分组对该域泛洪;将多个流交换机中具有来自该域的外部端口集合中的至少一个外部端口的每个流交换机配置成将所接收的控制广播分组转发至该域的控制器,而没有用所接收的控制广播分组对该域泛洪;以及将该域中的多个流交换机配置成将所接收的数据广播分组转发至该域的外部端口集合,而没有用所接收的数据广播分组对该域泛洪。

[0009] 一种网络节点作为分体式架构网络中的域的控制器的作用。该域包括多个流交换机,其中多个流交换机实现分体式架构网络的数据平面,并且该控制器实现分体式架构网络的远离数据平面的控制平面。该网络节点将多个流交换机配置成高效处置分体式架构网络中的每种类型的广播分组,而没有用广播分组对该域泛洪,由此减少在该域内转发的广播分组数量。该网络节点包括:外部端口跟踪模块,用于跟踪该域的外部端口集合,其中每个外部端口将多个流交换机之一连接到该域外的装置;广播配置模块,在通信上耦合到外部端口跟踪模块,该广播配置模块配置成将该域中的多个流交换机中的每个流交换机配置成将自身生成的广播分组转发至具有该域的外部端口集合中的至少一个外部端口的每个流交换机,而没有用自身生成的广播分组对该域泛洪。该广播配置模块配置成将多个流交换机中具有来自该域的外部端口集合中的至少一个外部端口的每个流交换机配置成将所接收的控制广播分组转发至该域的控制器,而没有对该域泛洪,并配置成将该域的多个流交换机配置成将所接收的数据广播分组转发至该域的外部端口集合,而没有用所接收的数据广播分组对该域泛洪。

附图说明

[0010] 本发明在附图的各图中以示例的方式而不以限制的方式图示,附图中相似的附图标记指示类似的元件。应该注意到,此公开中对“一”或“一个”实施例的不同提及不一定针对同一实施例,并且这样的提及意味着至少一个。另外,当结合实施例描述具体特征、结构或特性时,认为结合其它实施例实现这样的特征、结构或特性是在本领域技术人员知识范围内,而不管是否明确描述。

[0011] 图1是简单分体式架构OpenFlow网络的示例配置的一个实施例的图解。

[0012] 图2是分体式架构网络中的控制平面中的控制器和数据平面中的流交换机的一个实施例的图解。

[0013] 图3是用于配置用于高效处置广播分组的分体式架构网络的过程的一个实施例的流程图。

[0014] 图4是跟踪外部端口的过程的一个实施例的流程图。

[0015] 图5是用于处置自身生成的广播分组的过程的一个实施例的流程图。

[0016] 图6是控制器正在处理自身生成的分组的示例分体式架构网络的图解。

[0017] 图7是用于配置和处理所接收的广播分组的过程的一个实施例的图解。

[0018] 图8是处理所接收的控制广播分组的控制器和域的一个实施例的图解。

[0019] 图9是处理所接收的数据广播分组的控制器和域的一个实施例的图解。

[0020] 图10是用于隐式外部端口标识的过程的一个实施例的流程图。

[0021] 图11是使用邻居发现协议的隐式外部端口确定的一个实施例的图解。

[0022] 图12是使用交换机级发现协议的隐式外部端口确定的一个实施例的图解。

[0023] 图13是用于外部端口标识的显式过程的一个实施例的流程图。

具体实施方式

[0024] 在以下描述中,阐述了许多特定细节。然而,要理解,本发明的实施例可以在没有这些特定细节的情况下实施。在其它实例中,众所周知的电路、结构和技术未详细示出,以免模糊了对此描述的理解。然而,本领域技术人员将认识到,可在没有这样的特定细节的情况下实施本发明。本领域普通技术人员用所包含的描述将能够实现适当功能性而无需过多实验。

[0025] 流程图的操作将参考图2、6、8、9、11和12的示范实施例进行描述。然而,应该理解,图3-5、7、10和13中流程图的操作可由不同于参考图2、6、8、9、11和12所讨论的实施例的本发明的实施例执行,并且参考图2、6、8、9、11和12所讨论的实施例可执行与参考图3-5、7、10和13的流程图所讨论的操作不同的操作。

[0026] 图中所示的技术可使用在一个或多个电子装置(例如终端站、网络元件、服务器或类似电子装置)上存储和执行的代码和数据实现。这样的电子装置使用诸如非暂态机器可读或计算机可读存储介质(例如磁盘、光盘、随机存取存储器、只读存储器、闪存装置和相变存储器)等非暂态机器可读或计算机可读介质存储和传递(在内部和/或通过网络与其它电子装置)代码和数据。此外,这样的电子装置通常包含耦合到诸如一个或多个存储装置、用户输入/输出装置(例如键盘、触摸屏和/或显示器)和网络连接等一个或多个其它组件的一个或多个处理器的集合。处理器的集合和其它组件的耦合通常通过一个或多个总线和桥(也称为总线控制器)。存储装置表示一个或多个非暂态机器可读或计算机可读存储介质以及非暂态机器可读或计算机可读通信介质。由此,给定电子装置的存储装置通常存储代码和/或数据以便在那个电子装置的一个或多个处理器的集合上执行。当然,本发明实施例的一个或多个部分可使用软件、固件和/或硬件的不同组合来实现。

[0027] 如本文所使用的,网络元件或网络节点(例如路由器、交换机、桥或类似的连网装置)是一件连网设备,其包含在通信上互连网络上的其它设备(例如其它网络元件或节点、终端站或类似的连网装置)的硬件和软件。一些网络元件或网络节点是为多个连网功能(例

如路由、桥接、交换、层2聚集、会话边界控制、多播和/或订户管理) 提供支持和/或为多个应用服务(例如数据收集) 提供支持的“多服务网络元件”或“多服务网络节点”。术语网络元件和网络节点在本文中可交替使用。

[0028] 本发明的实施例提供了用于避免现有技术的缺点的方法和系统。处置广播分组转发的简单实现是在网络的每个节点中的所有活动端口上盲目转发广播分组。然而,这是低效的,因为使用此实现转发的许多分组不需要到达所有预计接收方。对于自身生成的和所接收的广播分组两种情况,广播分组仅需要在域的所有外部端口上发送出。外部端口被定义为连接到不在域的控制器的网络节点或装置的域(例如OpenFlow域)的端口。

[0029] 现有分体式架构网络(包含OpenFlow分体式架构网络)未提供用于向域外发送出自身生成的广播分组的高效机制。对于所接收的广播分组,流交换机由控制器预先编程以在域中的所有交换机上的所有活动端口上泛洪广播分组。这种简单机制在转发广播分组的网络资源的不必要使用方面是非常低效的。例如,如果OpenFlow域网络具有 m 个链路和 n 个外部端口,则理想上,所接收的广播分组应该仅在 n 个外部端口上发送出。使用当前技术,它将至少在所有 m 个链路和 n 个外部端口上发送。所有交换机都转发用于在所有 m 个链路上泛洪分组的资源导致网络资源的低效使用。在典型网络中, m 可以是大于 n 的数量级,在此情况下,资源利用低效程度可能更高。

[0030] 本发明的实施例克服了现有技术的这些缺点。本发明的实施例包含提供广播自身生成的广播分组的高效方式以及用于处置所接收的广播分组情况的优化解决方案的分组广播方法和系统。本发明的实施例提供广播以下多种情况下的分组的机制:自身生成的广播分组、所接收的控制广播分组以及所接收的数据广播分组。这些机制比上面描述的盲目转发的标准实现更加高效得多。

[0031] 例如,使用当前技术,任何广播分组都将在域内的所有交换机上的所有端口上泛洪,导致宝贵的网络资源的低效使用。本发明的实施例提供了用于广播分组处置的机制,其仅按所需比例使用网络资源,从而导致在网络资源上有重大节省。

[0032] 在一个实施例中,控制器将域的流交换机配置成将广播分组直接转发至相关外部端口,而没有浪费域中其它流交换机的资源。控制器标识外部端口,并对底层流交换机中的适当流编程来进行分组转发。

[0033] 图2是具有高效广播分组处置的分体式架构网络中的控制器和流交换机的一个实施例的图解。控制器103和流交换机133分别由网络节点101、131实现。这些网络节点可以是任何类型的连网元件,包含路由器、交换机以及类似装置。网络节点101、131可通过任何类型的网络(包含局域网(LAN)、诸如因特网等广域网(WAN))通信,或通过类似网络通信。网络可由有线或无线装置的任何组合构成。控制器103可与任何数量的流交换机133通信,并被指定控制它们。为了简洁起见,图示了单个流交换机。

[0034] 在一个实施例中,控制器103和流交换机133分别是OpenFlow控制器和OpenFlow交换机。OpenFlow是在驻留在www.openflow.org的OpenFlow交换规范中定义的系统 and 协议。在其它实施例中,利用实现协议或系统的其它类似分体式架构来分开控制平面功能性和数据平面功能性。

[0035] 在一个实施例中,控制器103可由应用、分布式网络中间件119和操作系统121的集合构成。分布式网络中间件119提供了支持控制器与其它网络节点之间的通信的功能和数

据结构的集合。操作系统121管理网络节点101的资源,诸如处理器和存储器资源。可利用提供这些服务的任何分布式网络中间件和操作系统。

[0036] 控制器的应用或组件可包含外部端口跟踪模块105、广播配置模块107、广播处理模块109、边界网关协议(BGP)模块111、开放最短路径优先(OSPF)模块113、邻居发现协议(NDP)模块115以及类似组件。BGP模块111保持指明自治系统之间的网络可达性的前缀表。OSPF模块113计算跨网络拓扑的最短路径。NDP模块115可向其邻居广告网络节点的能力和特性以实现网络节点之间的通信。类似地,NDP模块115通过在域之间的连接链路上从这些邻居接收广告来发现邻居。在另一实施例中,交换机实现NDP并向控制器的NDP模块115报告所发现的邻居和链路。NDP的示例是CISCO SYSTEMS公司的CISCO发现协议(CDP)和链路层发现协议(LLDP)。

[0037] 外部端口跟踪模块105创建并保持控制器的域的所有外部端口的列表。外部端口跟踪模块105的功能是确定域的外部端口集合。外部端口跟踪模块105结合其它模块(诸如NDP模块)起作用来与流交换机通信以标识它们的外部端口。此过程相对于图11-13更详细描述。外部端口跟踪模块105也可分开地或者作为跟踪外部端口的过程的一部分地跟踪内部端口和所有活动端口。

[0038] 广播配置模块107管理域中流交换机的配置以正确处置所接收的和自身生成的广播分组。配置流交换机的过程相对于图3-7更详细描述。广播处理模块109处置由控制器103接收的或自身生成的广播分组。控制器103可生成要广播的分组,或者可接收由其域中的流交换机转发的广播分组。广播处理模块109可结合其它模块(诸如BGP模块)起作用来处理所接收的广播分组。广播分组的处理关于本文下面的另外的细节更详细描述。

[0039] 操作系统121和分布式网络中间件119可通过提供在其上可传送控制平面信令协议的安全通信信道127来便于控制器103与流交换机133之间的通信。在一个实施例中,控制器103和流交换机133使用OpenFlow协议通信。

[0040] 流交换机133由网络节点131提供。网络节点131和流交换机133可连接到任何数量的其它网络节点或类似装置。这些连接中的每个连接都通过分开的通信端口,本文简称为“端口”。这些端口135可以是与同一域内的其它网络节点或与该域外的网络节点或装置的连接。与域内的网络节点连接的端口135在本文称为“内部端口”。与域外的节点连接的端口135在本文称为“外部端口”。

[0041] 流交换机133可接收、生成和/或转发数据分组,包含广播分组。转发表137或类似的数据结构定义确定流交换机如何处置每个数据分组的规则集合。转发表137可以是包含用于标识各种不同类型的数据分组并对它们采取措施的规则集合的OpenFlow转发表或类似类型的转发表。控制器103将这个转发表137配置成例如通过建立用于处置广播分组的规则来管理流交换机133的转发行为。

[0042] 图3是用于配置用于高效处置广播分组的分体式架构网络的过程的一个实施例的流程图。这个过程由控制器的组件执行以将其域中的流交换机集合配置成高效处置各种类型的广播分组。控制器跟踪分体式架构网络内的其域的外部端口集合(框301)。如本文所使用的,“集合”是指任何整数个项目,包含一个项目。外部端口可在任何类型数据结构(诸如列表、表或类似数据结构)中被跟踪。标识外部端口的过程在本文下面相对于图11-13更详细讨论。

[0043] 过程继续将域中具有外部端口的所有流交换机配置成将自身生成的广播分组转发至已经被标识的外部端口的集合中的每个外部端口(框303)。该配置可使用控制平面协议执行以配置用于域中的每个流交换机的转发规则。自身生成的广播分组是由域内的流交换机产生的广播分组。这个自身生成的广播分组(例如ARP消息)仅需要被发送到域外的网络节点,这是因为流交换机共享控制器,该控制器已经有权访问配置域内的流交换机所需的信息。因此,用这些自身生成的消息对域的内部端口泛洪是资源的浪费。这些自身生成的广播消息在控制器处产生,该控制器基于其所保持的外部端口列表将该消息直接发送到域的每个外部端口。

[0044] 该过程将域中具有外部端口的每个流交换机配置成将所接收的控制广播分组转发至控制器(框305)。具有外部端口的流交换机从域外部的装置接收所接收的控制广播分组。所接收的控制广播分组内的信息仅与控制器相关。因而,流交换机配置成将这个信息发送到控制器,代替用这个接收的控制广播分组对内部端口泛洪,这个接收的控制广播分组不是域中的其它流交换机需要的。所接收的控制广播分组的示例是从控制器的域的外部的装置接收的ARP消息。控制器然后将类似于自身生成的分组地在所有外部端口上转发这个接收的控制广播分组。控制器还可处理已经由流交换机转发的所接收的控制广播分组的信息。

[0045] 控制器将其域中的流交换机配置成将所接收的数据广播分组转发至域的外部端口集合(框307)。这些数据广播分组不需要控制平面处理,并且不需要被转发到控制器。控制器配置外部端口的所有集合之间的流的集合,使得在外部端口上接收的每个数据广播分组都跨域转发到每个其它外部端口。使用所建立的流执行这个处理避免了在域内的所有内部端口上盲目转发所接收的数据广播分组,该盲目转发浪费资源。流确保所接收的数据广播分组到达每个外部端口,而没有使用泛洪。流还可建立成确保每个流交换机都接收数据广播分组,而没有使用泛洪。流实现这些分组的转发,使得控制器将数据分组引导到具有最短路径或类似配置的路径集合的每个外部端口。

[0046] 图4是跟踪外部端口的过程的一个实施例的流程图。这个过程由控制器执行。控制器使用显式过程或隐式过程来确定外部端口。显式过程和隐式过程相对于图11-13更详细讨论。该过程首先标识控制器的域内的流交换机的所有活动端口(框401)。可选地,控制器可标识活动端口列表内的内部端口(框403)。取决于分体式架构网络和控制平面协议的实现,控制器可能需要确定内部端口以推断外部端口集合。在其它实施例中,控制器可直接确定外部端口集合。控制器确定外部端口集合并使用这个将流交换机配置成高效处置广播分组(框405)。

[0047] 图5是用于处置自身生成的广播分组的过程的一个实施例的流程图。这个过程依赖于上面相对于图4所描述的和下面相对于图11-13进一步描述的外部端口的跟踪(框501)。自身生成的分组由控制器产生(框503)。自身生成的分组的示例是ARP消息。控制器使用外部端口列表将转发分组引导到具有外部端口的流交换机(框505)。控制器使用每个流交换机建立的流或安全信道转发自身生成的分组。

[0048] 图6是控制器正在处理自身生成的分组的示例分体式架构网络的图解。域用虚线603标识。虚线603上或其内的所有流交换机都是控制器601的域的一部分。外部端口用圆圈605标记。自身生成的分组被从控制器转发至具有外部端口605的每个流交换机。黑箭头实

线607示出了自身生成的分组的转发。点线609示出了所建立的控制通信信道或安全信道以允许控制器601与所有流交换机之间的控制协议通信。自身生成的分组可通过这些控制通信信道或通过其它类似路径转发至外部端口605。自身生成的分组不在域的内部端口上泛洪,由此减少了正确转发自身生成的分组所需的资源。

[0049] 图7是用于配置和处理所接收的广播分组的过程的一个实施例的图解。所接收的广播分组是通过控制器的域中的流交换机的外部端口接收的广播分组。因此,从域外部的网络节点或装置接收所接收的广播分组。

[0050] 这个过程也依赖于外部端口列表的可用性,并且假定这个外部端口列表正在如上面相对于图4和本文下面相对于图11-13所描述的那样创建和保持(框701)。控制器执行这个过程,或引导这个过程的执行。控制器对其域中具有外部端口的每个流交换机编程以建立从那个流交换机到控制器的流(框703)。这个流被设计成将这些控制分组的转发限制于仅将它们从接收来自外部装置的这些消息的流交换机转发到控制器。所接收的控制分组不在域的内部端口上广播,由此节省了分体式架构网络中的资源。

[0051] 在一个实施例中,具有外部端口的流交换机被编程为分析入局分组以在目的地MAC地址上查找指示入局分组是广播分组的匹配。例如,可利用MAC寻址的FF:FF:FF:FF:FF:FF标识广播分组。对于每个检测到的广播分组,检查入局分组的以太类型以确定该分组是否是控制分组。流交换机被配置有转发具有指示广播分组的目的地和指示控制分组的以太类型的任何分组的动作。例如,可接收ARP请求分组,该分组具有FF:FF:FF:FF:FF:FF MAC地址和指示控制分组的以太类型。满足这些标准的那些分组然后被编程为使用流交换机内的规则或类似可配置动作被转发到控制器。

[0052] 控制器然后计算从域中具有外部端口的每个流交换机到具有外部支持的每个其它流交换机的流集合(框705)。流可以是跨域的外部端口之间的最短路径或类似路径。控制器可使用其域的拓扑的其知识计算这些路径。

[0053] 在计算流集合之后,控制器对沿流的每个流交换机编程,以朝与该流关联的外部端口转发数据广播分组(框707)。在其它实施例中,流也可被建立成确保将数据广播分组高效分发到域中的所有流交换机,而没有使域内的数据广播分组泛洪。流可以是单播路径或多播路径,或者其任何组合。这些路径可以是通过最短路径算法(诸如Dijkstra算法或类似过程)所确定的最短路径。在一个实施例中,流交换机被编程为分析所接收的数据广播分组以确定它们是否是广播分组(例如在FF:FF:FF:FF:FF:FF的MAC地址上匹配)。还分析分组的以太类型以确定它们是控制分组还是数据分组。如果所接收的数据分组是数据广播分组,则流交换机配置成沿关联的流的集合转发它。这避免了这些分组的盲目转发,并且避免了向控制器转发这些分组。

[0054] 在流交换机被正确配置之后,广播分组的处理开始。在处理广播分组时控制器的角色包含处理从流交换机接收的控制广播分组,并且然后将它们转发到具有外部端口的每个流交换机以在域的每个外部端口上转发(框709)。

[0055] 图8是处理所接收的控制广播分组的控制器和域的一个实施例的图解。在外部端口805接收控制广播分组803。具有外部端口805的流交换机已经被配置成将所接收的控制广播分组转发807至控制器801。流交换机未用所接收的控制广播分组对域泛洪。

[0056] 控制器801按需要处理分组,并将控制广播分组809转发至具有外部端口的每个流

交换机。这些流交换机中的每个流交换机然后都在关联的外部端口上转发控制广播分组。这确保了所接收的控制广播分组的正确分发,同时避免了不必要地用所接收的控制广播分组对域的内部端口和链路泛洪。

[0057] 图9是处理所接收的数据广播分组的控制器和域的一个实施例的图解。在外部端口905接收数据广播分组903。具有外部端口905的流交换机已经配置成沿着由控制器配置的流集合转发907所接收的数据广播分组。流将所接收的数据广播分组的转发引导到每个其它外部端口909。与其它外部端口909关联的流交换机接收数据广播分组,并在外部端口上转发分组。因而,所接收的数据广播分组通过域转发至每个外部端口,而没有对内部链路泛洪,并且没有向控制器转发分组。

[0058] 图10是用于隐式外部端口标识的过程的一个实施例的流程图。在此过程中,通过从活动端口列表中去除内部端口来标识外部端口。在一个实施例中,可利用OpenFlow协议来获得域中所有流交换机的所有活动端口的列表。在其它实施例中,利用链路级或交换机级发现协议来确定域中的活动端口和内部端口。

[0059] 该过程通过控制器查询域中的所有流交换机而发起,以生成每个流交换机的活动端口列表(框1001)。在一个实施例中,查询通过使用OpenFlow协议来执行。当从域中的每个流交换机接收到对查询的回答时,活动端口列表被编译成域的活动端口列表(框1003)。一旦活动端口列表已经被编译,控制器就使用NDP(诸如LLDP)来确定哪些活动端口是域的内部端口(框1005)。当接收到内部端口列表时,控制器将它们编译成域的内部端口列表(框1007)。

[0060] 然后可以通过比较域的内部端口列表与域的活动端口列表来确定来自域的外部端口(框1009)。未被列为内部端口的所有活动端口都被推断为外部端口。外部端口列表然后可被保持用于广播分组处置。这可以是具有活动端口列表的正在进行的过程,内部端口列表和外部端口列表被连续或定期更新或重新计算。

[0061] 图11是使用邻居发现协议的隐式外部端口确定的一个实施例的图解。控制器在其每个端口上发送出NDP消息1103A。这个NDP消息1103A由接收网络元件在其端口1103B上转发。NDP消息1103C然后返回到控制器1101,由此使控制器1101能够确定NDP消息1103B的路由是覆盖域的内部链路的内部路由。一旦所有内部链路都被确定,就能从这个信息以及所有活动端口的列表推断外部端口。

[0062] 图12是使用交换机级发现协议的隐式外部端口确定的一个实施例的图解。控制器1201与域中的每个流交换机1203A、1203C通信以发起在流交换机之间的交换机级发现协议消息1203B的发送。流交换机然后向控制器报告它们的链路信息,使得控制器能确定链路和端口是内部的或者是外部的。

[0063] 图13是用于外部端口标识的显式过程的一个实施例的流程图。在这个实施例中,使用诸如链路级发现协议(LLDP)等邻居发现协议(NDP)或类似协议来标识外部端口。这个实施例要求与域的流交换机和控制器通信的外部装置支持与控制器相同的NDP。当从不是域一部分的装置接收到NDP消息时,认出外部端口。消息的入口端口被记录为外部端口。

[0064] 可响应于在流交换机从另一网络节点接收到NDP消息而发起外部端口列表保持的这个过程(框1301)。这个NDP消息是由流交换机转发至控制器的控制消息。控制器检查所接收的NDP消息以确定消息源(框1303)。该源可根据NDP消息中的源地址来确定。源网络节点

与控制器的域中的已知网络节点的列表相比较(框1305)。如果源网络节点不在控制器的域内,则通过其接收NDP消息的流交换机的端口被添加到域的外部端口的列表(框1307)。当NDP消息被转发或响应于控制器的查询被转发到向控制器转发NDP消息的流交换机时,可提供这个端口的标识。

[0065] 因而,提供了用于优化分体式架构网络中广播分组的处置的方法、系统和设备。要理解到,以上描述意图是说明性的而非约束性的。在阅读并理解了以上描述后,许多其它实施例对于本领域技术人员将是显而易见的。因此,本发明的范围应该根据所附权利要求书连同赋予这样的权利要求书权利的等效方案的整个范围来确定。

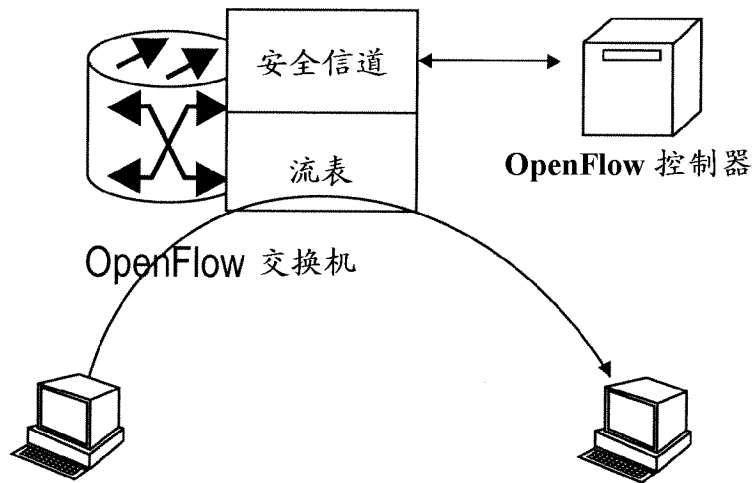


图 1

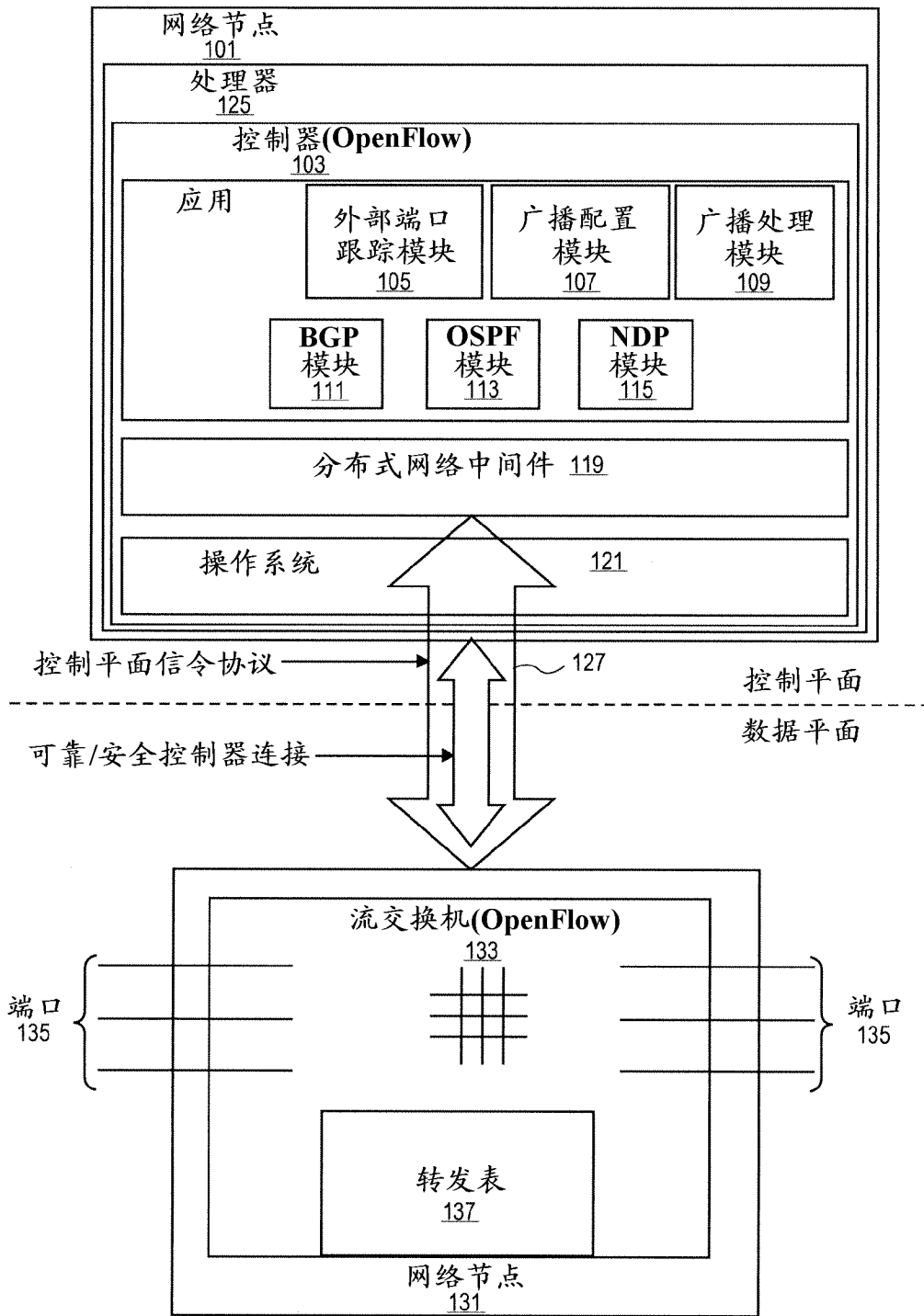


图 2

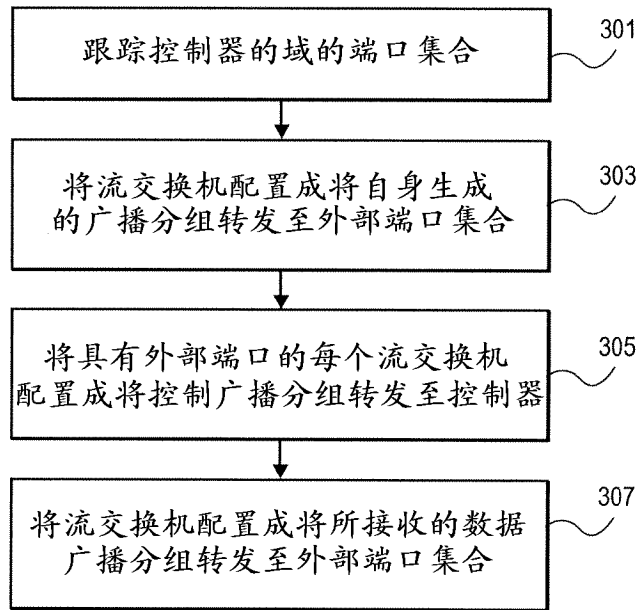


图 3

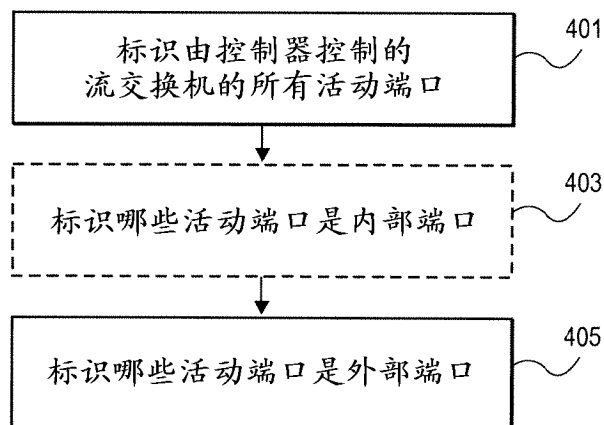


图 4

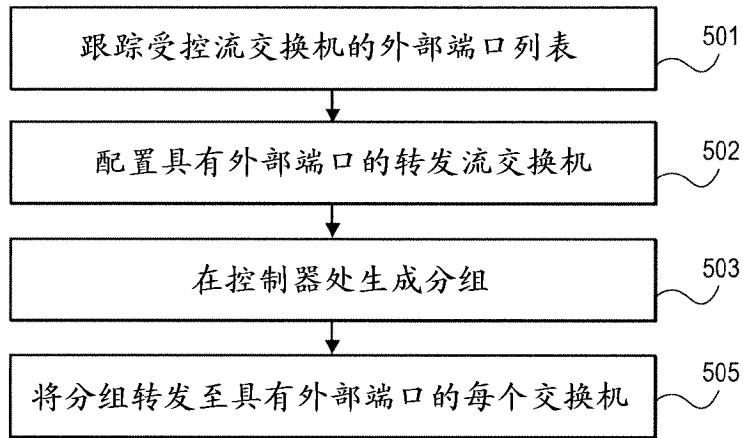


图 5

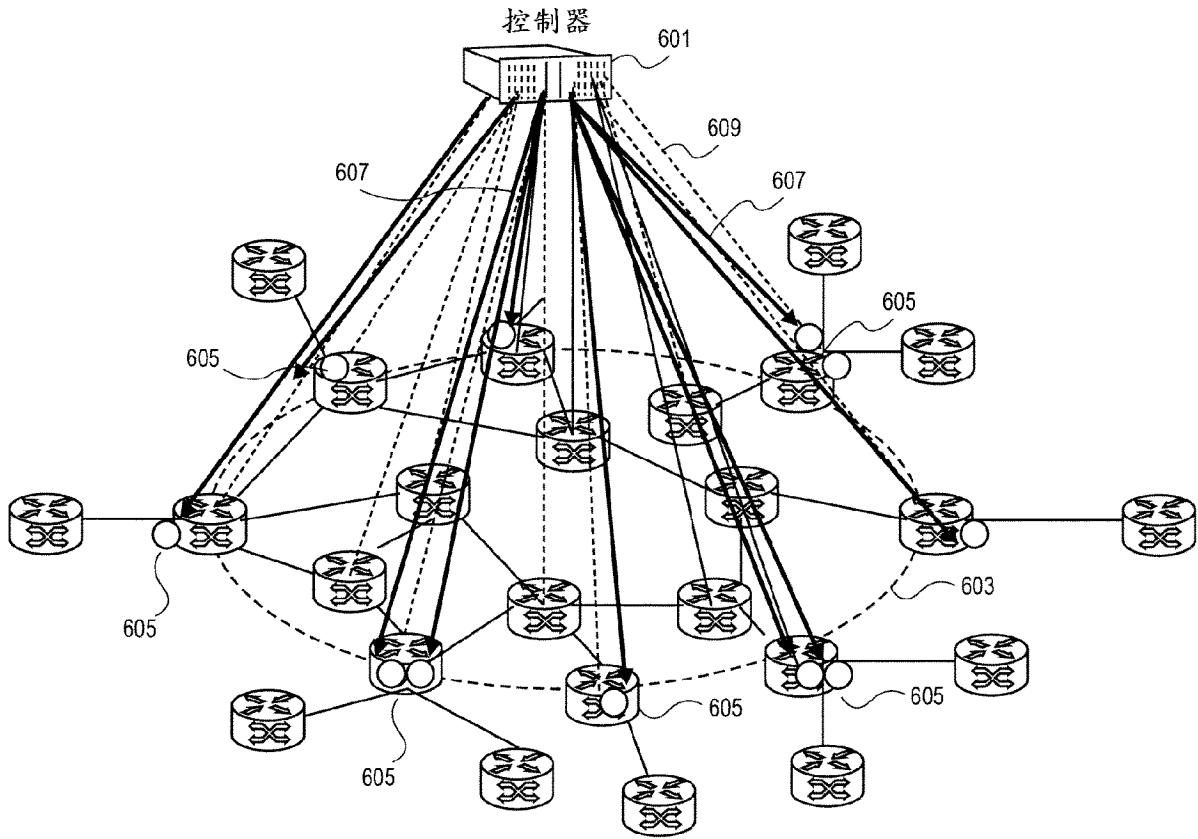


图 6

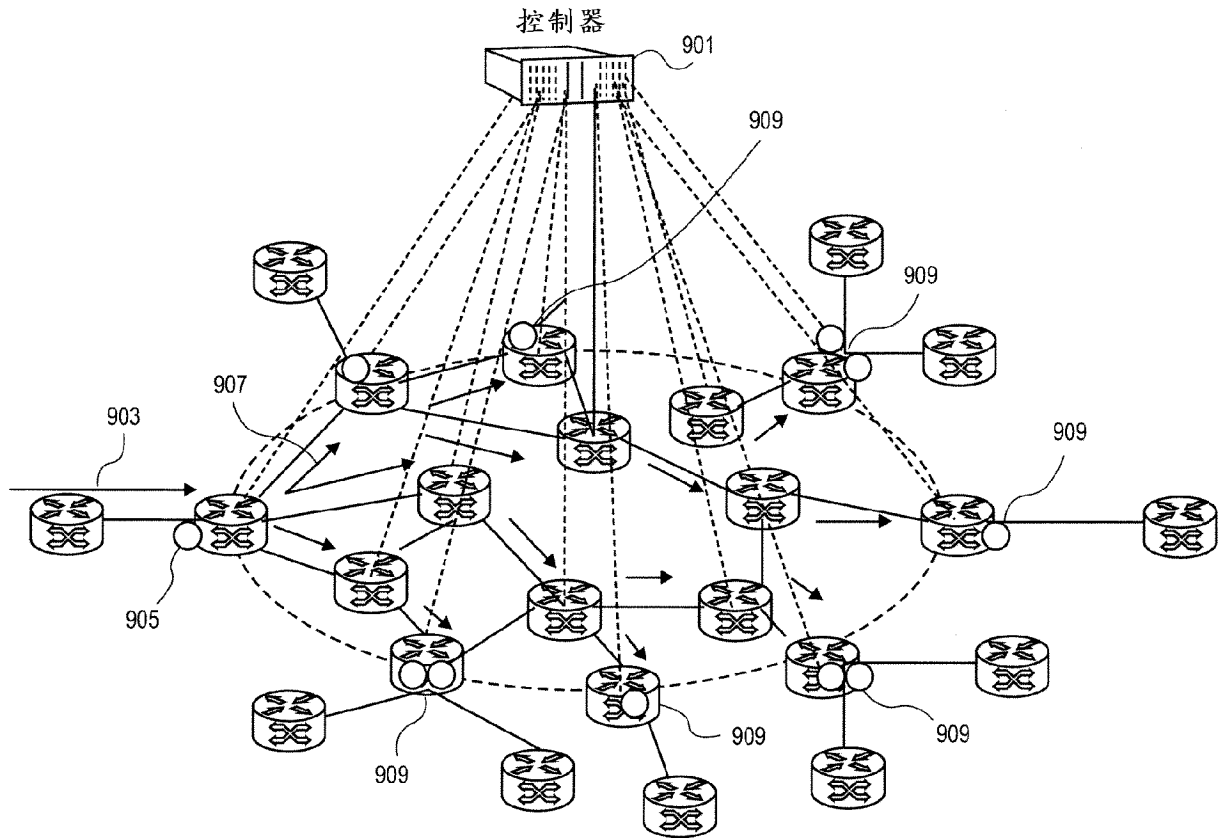


图 9

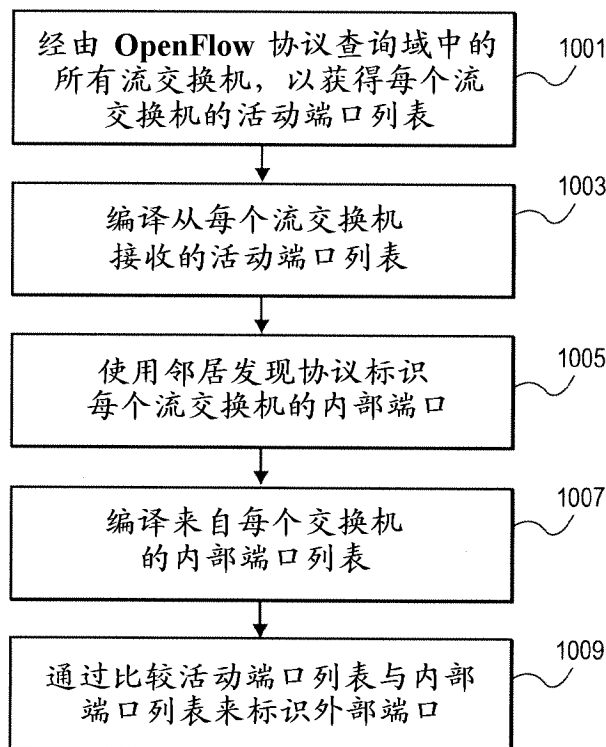


图 10

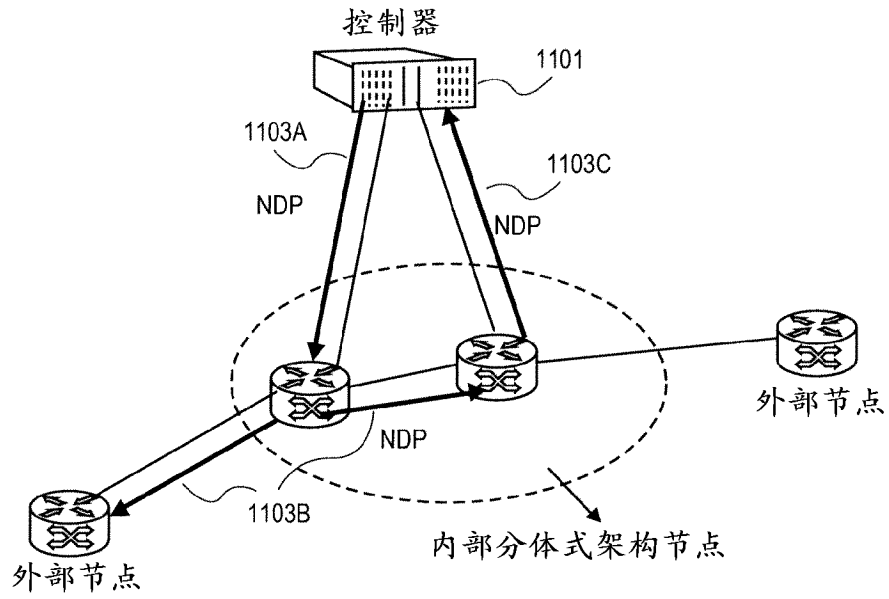


图 11

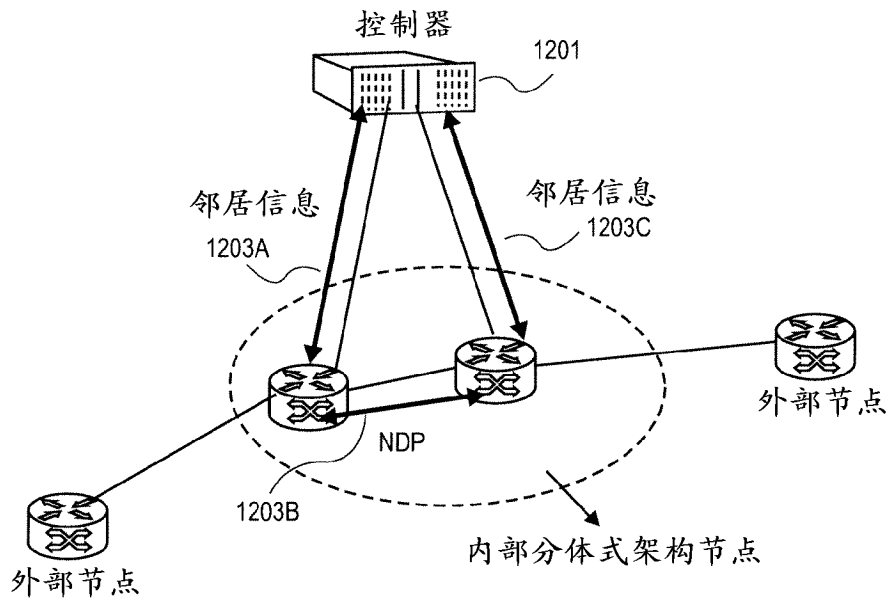


图 12

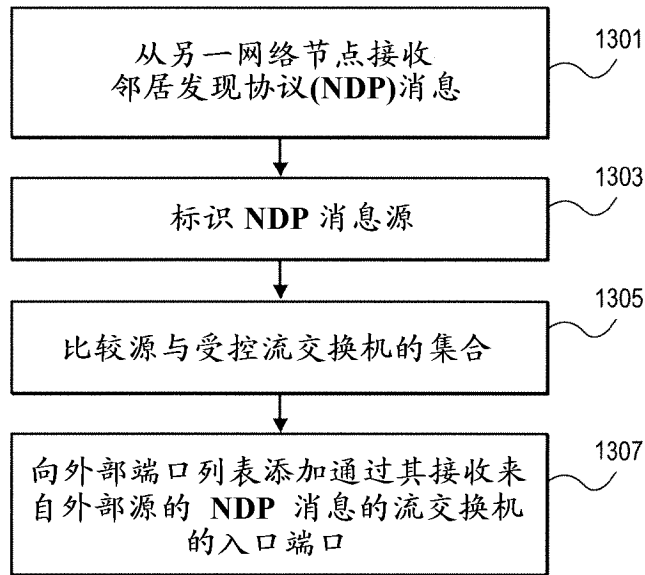


图 13