



(12)发明专利

(10)授权公告号 CN 103718519 B

(45)授权公告日 2016.08.17

(21)申请号 201280039289.7

(22)申请日 2012.07.26

(30)优先权数据

13/208251 2011.08.11 US

(85)PCT国际申请进入国家阶段日

2014.02.11

(86)PCT国际申请的申请数据

PCT/IB2012/053833 2012.07.26

(87)PCT国际申请的公布数据

W02013/021304 EN 2013.02.14

(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

N.贝赫什蒂-扎瓦雷赫 张颖

(74)专利代理机构 中国专利代理(香港)有限公司 72001

代理人 杨美灵 汤春龙

(51)Int.Cl.

H04L 12/721(2013.01)

H04L 12/751(2013.01)

(56)对比文件

CN 1783841 A,2006.06.07,

CN 101188568 A,2008.05.28,

US 2010/0322255 A1,2010.12.23,

审查员 刘静微

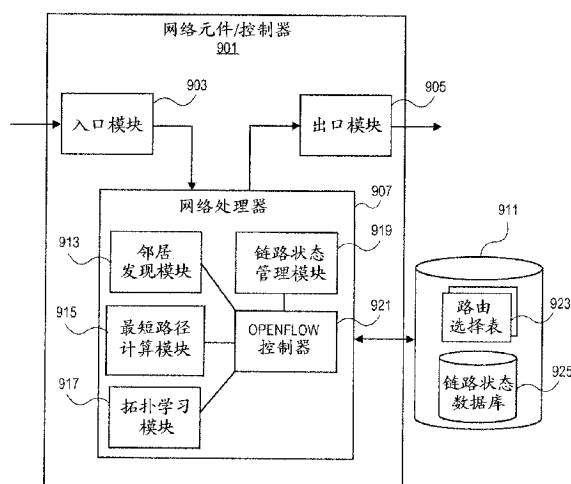
权利要求书3页 说明书10页 附图12页

(54)发明名称

实现分离式架构网络中的OSPF

(57)摘要

在用作分离式架构网络的多个区域之一的多个控制器之一的网络元件中实现一种方法。控制器为分离式架构网络的区域提供控制平面,其中控制器远离多个交换机(其为分离式架构网络的区域提供数据平面)。通过将有限区域内链路成本数据提供给分离式架构网络的其它区域的其它控制器以及包括分离式架构网络的网络的传统路由器,控制器促进跨分离式架构网络的多个区域的优化路由选择。有限区域内链路成本数据提供控制器的区域的各可能最短路径遍历的成本,而无需提供所有内部链路成本数据。



1. 一种在用作分离式架构网络的多个区域之一的多个控制器之一的网络元件中实现的方法,所述控制器为所述分离式架构网络的所述区域提供控制平面,其中所述控制器远离为分离式架构网络的所述区域提供数据平面的多个交换机,所述控制器通过将有限区域内链路成本数据提供给所述分离式架构网络的其它区域的其它控制器以及包括所述分离式架构网络的网络的传统路由器,来促进跨所述分离式架构网络的所述多个区域的优化路由选择,所述有限区域内链路成本数据提供穿越所述控制器的所述区域的各可能最短路径的成本,而无需提供所有内部链路成本数据,所述方法包括下列步骤:

学习所述分离式架构网络中的所述控制器的所述区域的拓扑,所述拓扑包括所述控制器的所述区域中的各边界交换机,其中所述控制器的所述区域中的各边界交换机具有至少一个外部端口,所述至少一个外部端口将所述控制器的所述区域链接到所述分离式架构网络的另一个区域或者所述网络中的传统路由器之一;

计算所述控制器的所述区域中的各边界交换机对之间的最短路径;

在所述控制器的路由选择表中存储各边界交换机对之间的各最短路径的成本;

使用问候协议来识别所述分离式架构网络中的各相邻控制器或者所述网络中的相邻传统路由器,其中各相邻控制器控制通过所述控制器的所述区域的至少一个外部端口可访问的所述分离式架构网络的另一个区域中的交换机;

与各相邻控制器和相邻传统路由器交换链路状态数据库,所述链路状态数据库包括各边界交换机对之间的各最短路径的所述成本;

以所述控制器作为树的根来计算所述网络的最短路径树;以及

更新所述控制器的所述区域的交换机中的转发表,以实现按照所述最短路径树的转发。

2. 如权利要求1所述的方法,其中,识别各相邻控制器的步骤还包括下列步骤:

在所述控制器的所述区域的外端口上发送问候分组;以及

通过所述控制器的所述区域的至少一个外部端口从各相邻控制器接收问候分组。

3. 如权利要求1所述的方法,其中,交换链路状态数据的步骤还包括向各相邻控制器发送链路状态通告的步骤,所述链路状态通告包括到所述控制器的各外部链路的成本。

4. 如权利要求1所述的方法,其中,交换链路状态数据的步骤还包括向各相邻控制器发送链路状态通告的步骤,所述链路状态通告包括所述区域中的各边界交换机对的最短路径的所述成本。

5. 如权利要求1所述的方法,还包括向各相邻控制器通告更新链路状态数据的步骤。

6. 如权利要求5所述的方法,其中,通告所述更新链路状态数据的步骤还包括向各相邻控制器发送具有更新链路状态数据的链路状态通告消息的步骤。

7. 如权利要求1所述的方法,其中,更新转发表的步骤还包括使用OpenFlow协议来更新所述控制器的所述区域中的各交换机的转发表的步骤。

8. 如权利要求1所述的方法,还包括在所述控制器初始化开放最短路径优先(OSPF)协议数据结构的步骤。

9. 一种用作分离式架构网络的多个区域之一的多个控制器之一的网络元件,所述控制器为所述分离式架构网络的所述区域提供控制平面,其中所述控制器远离为分离式架构网络的所述区域提供数据平面的多个交换机,所述控制器通过将有限区域内链路成本数据提

供给所述分离式架构网络的其它区域的其它控制器以及包括所述分离式架构网络的网络的传统路由器,来促进跨所述分离式架构网络的所述多个区域的优化路由选择,所述有限区域内链路成本数据提供穿越所述控制器的所述区域的各可能最短路径的成本,而无需提供所有内部链路成本数据,所述网络元件包括:

入口模块,配置成通过网络来接收数据;

出口模块,配置成通过所述网络来传送数据;

网络处理器,耦合到所述入口模块和出口模块,所述网络处理器配置成执行一组模块,包括控制器模块、拓扑学习模块、最短路径计算模块、邻居发现模块和链路状态管理模块,

所述拓扑学习模块,配置成确定所述分离式架构网络中的所述控制器的所述区域的拓扑,所述拓扑包括所述控制器的所述区域中的各边界交换机,其中各边界交换机具有至少一个外部端口,所述至少一个外部端口将所述控制器的所述区域链接到所述网络的另一个区域或者所述网络中的传统路由器,

所述控制器模块,配置成为所述分离式架构网络中的控制器的所述区域提供控制平面功能性,

所述最短路径计算模块,配置成识别将要与相邻控制器和传统路由器共享的所述控制器的所述区域中的各边界交换机之间的最短路径,以及以所述控制器作为树的根来计算所述网络的最短路径树,

所述邻居发现模块,配置成使用问候协议来识别所述网络中的各相邻控制器和传统路由器,以及

所述链路状态管理模块,配置成与所述网络中的各相邻控制器和传统路由器交换链路状态数据库,所述链路状态数据库包括所述控制器的所述区域中的各边界交换机对之间的各最短路由的成本;以及

路由选择表存储装置,在通信上耦合到所述网络处理器,所述路由选择表存储装置配置成存储用于所述控制器模块的路由选择表,包括所述控制器的所述区域以及所述控制器的所述区域的边界交换机之间的最短路径信息。

10.如权利要求9所述的网络元件,其中,所述邻居发现模块配置成在所述区域的各外部端口上发送问候分组,以及通过所述区域的至少一个外部端口从各相邻控制器接收问候分组。

11.如权利要求9所述的网络元件,其中,所述链路状态管理模块还配置成向各相邻控制器发送链路状态通告,所述链路状态通告包括到所述控制器的各外部链路的所述成本。

12.如权利要求9所述的网络元件,其中,所述链路状态管理模块还配置成向各相邻控制器发送链路状态通告,所述链路状态通告包括所述控制器的所述区域中的各边界交换机对的最短路径的所述成本。

13.如权利要求9所述的网络元件,其中,所述链路状态管理模块还配置成向各相邻控制器通告更新链路状态数据。

14.如权利要求9所述的网络元件,其中,所述链路状态管理模块还配置成向各相邻控制器发送具有更新链路状态数据的链路状态通告消息。

15.如权利要求9所述的网络元件,其中,所述控制器模块还配置成使用OpenFlow协议来更新所述控制器的所述区域中的各交换机的转发表。

16. 如权利要求9所述的网络元件,其中,所述最短路径计算模块还配置成在所述控制器初始化开放最短路径优先(OSPF)协议数据结构。

实现分离式架构网络中的OSPF

技术领域

[0001] 本发明的实施例涉及分离式架构网络中的路径查找或路由选择。具体来说,本发明的实施例涉及用于实现开放最短路径优先过程(其用于在包括分离式架构网络 and 传统路由器的网络中路由数据)的方法和系统。

背景技术

[0002] 分离式架构网络设计引入网络的控制与转发组件(其又称作控制平面和转发或数据平面)之间的分隔。分离式架构网络能够用于电信级网络的接入/聚合域、移动回程、云计算和多层(L3和L2和L1、光传输网络(OTN)、波分复用(WDM))支持中,其全部处于网络架构的构件块之中。

[0003] 与传统网络架构(其将转发(数据)和控制平面整合在同一个框中)不同,分离式架构分离这两个功能,并且在服务器(控制器)(其可能处于与转发元件(交换机)不同的物理位置中)上运行控制平台。分离式架构简化转发平台的功能和硬件,并且将网络的智能和管理集中到一组控制器(其监督交换机)中。传统网络架构中的转发和控制平台的紧密结合通常引起极复杂的控制平面和复杂网络管理。这使新的连网装置的创建成本高,并且造成对这些装置中潜在部署的新协议和技术的进入的极大障碍。尽管对线路速度、端口密度和性能的迅速改进,但是用于管理这些特征的网络控制平面机制以慢许多的步调发展。

[0004] 在分离式架构网络中,控制器从交换机收集信息,以及计算并且向交换机分配适当转发判定。控制器和交换机使用控制平面协议来传递和交换信息。这种协议的一个示例是OpenFlow,其提供使交换机与控制器进行通信的开放式标准方法。图1是交换机与控制器之间的OpenFlow接口的概述的简图。OpenFlow控制器使用安全信道与OpenFlow交换机进行通信,以配置转发表(流表)。

[0005] OpenFlow交换机中的转发表装载有由下列项所述组成的条目:定义分组报头中的字段的匹配的规则;在检测规则所定义的匹配时将要执行的动作;以及关于数据平面中的数据分组的处理的统计的集合。当入局数据分组匹配特定规则时,对该数据分组执行关联动作。规则包含来自协议栈中的若干报头的关键字段,例如以太网MAC地址、IP地址、IP协议、TCP/UDP端口号以及入局端口号。能够将具有相似特性的一组数据分组作为流来管理。能够使用数据分组中的任何数量的可用字段或者其组合来定义流。还有可能通过将通配符用于不希望的字段,来将规则限制成对可用字段的子集进行匹配。

[0006] 分离式架构的控制平面和数据平面的分离减轻修改网络控制逻辑的任务,并且提供程序接口,开发人员在其上能够构建大量新协议和管理应用。在这个模型中,数据和控制平面能够单独演进和缩放,同时降低数据平面元件的成本。

发明内容

[0007] 描述一种方法,其在用作分离式架构网络的多个区域之一的多个控制器之一的网络元件中实现。控制器为分离式架构网络的区域提供控制平面,其中控制器远离多个交换

机(其为分离式架构网络的区域提供数据平面)。通过将有限区域内链路成本数据提供给分离式架构网络的其它区域的其它控制器以及包括分离式架构网络的网络的传统路由器,控制器促进跨分离式架构网络的多个区域的优化路由选择。有限区域内链路成本数据提供控制器的区域的各可能最短路径遍历的成本,而无需提供所有内部链路成本数据。该方法包括以下步骤,包括学习包括控制器的区域中的各边界交换机的分离式架构网络中的控制器的区域的拓扑,其中控制器的区域中的各边界交换机具有至少一个外部端口,其将控制器的区域链接到分离式架构网络的另一个区域或者网络中的传统路由器之一。步骤包括计算控制器的区域中的各边界交换机对之间的最短路径。步骤包括在控制器的路由选择表中存储各边界交换机对之间的各最短路径的成本。步骤包括使用问候协议来识别分离式架构网络中的各相邻控制器或者网络中的相邻传统路由器,其中各相邻控制器控制分离式架构网络的另一个区域中的交换机(其是通过控制器的区域的至少一个外部端口可访问的)。步骤包括与各相邻控制器交换链路状态数据库,链路状态数据库包括各边界交换机对之间的各最短路径的成本。步骤包括以控制器作为树的根来计算网络的最短路径树,并且更新控制器的区域的交换机中的转发表,以实现按照最短路径树的转发。

[0008] 网络元件用作分离式架构网络的多个区域之一的多个控制器之一。控制器为分离式架构网络的区域提供控制平面,其中控制器远离多个交换机(其为分离式架构网络的区域提供数据平面)。通过将有限区域内链路成本数据提供给分离式架构网络的其它区域的其它控制器以及包括分离式架构网络的网络的传统路由器,控制器促进跨分离式架构网络的多个区域的优化路由选择。有限区域内链路成本数据提供控制器的区域的各可能最短路径遍历的成本,而无需提供所有内部链路成本数据。网络元件包括:入口模块,配置成通过网络接收数据;以及出口模块,配置成通过网络传送数据。网络元件还包括网络处理器(其耦合到入口模块和出口模块),网络处理器配置成执行一组模块,包括控制器模块、拓扑学习模块、最短路径计算模块、邻居发现模块和链路状态管理模块。拓扑学习模块配置成确定包括控制器的区域中的各边界交换机的分离式架构网络中的控制器的区域的拓扑,其中各边界交换机具有至少一个外部端口,其将控制器的区域链接到网络的另一个区域或者网络中的传统路由器。控制器模块配置成为分离式架构网络中的控制器的区域提供控制平面功能性。最短路径计算模块配置成识别将要与相邻控制器和传统路由器共享的控制器的区域中的各边界交换机之间的最短路径,以及以控制器作为树的根来计算网络的最短路径树。邻居发现模块配置成使用问候协议来识别分离式架构网络中的各相邻控制器,以及链路状态管理模块配置成与各相邻控制器交换链路状态数据库。链路状态数据库包括控制器的区域中的各边界交换机对之间的各最短路由的成本,以及路由选择表存储装置在通信上耦合到网络处理器。路由选择表存储装置配置成存储用于控制器模块的路由选择表,其中包括控制器的区域以及控制器的区域的边界交换机之间的最短路径信息。

附图说明

[0009] 通过附图、作为示例而不是限制来说明本发明,附图中,相似的参考标号表示相似的元件。应当注意,本公开中的“一”、“一个”实施例的不同说法不一定表示同一个实施例,并且这类说法表示至少一个。此外,在结合一个实施例来描述特定特征、结构或特性时,无论是否明确描述,均认为结合其它实施例来实现这种特征、结构或特性是在本领域的技术

人员的知识范围之内。

[0010] 图1是简单OpenFlow网络的示例架构的一个实施例的简图。

[0011] 图2是开放最短路径优先(OSPF)分组报头的简图。

[0012] 图3是OSPF问候(hello)分组的简图。

[0013] 图4是链路状态通告(LSA)报头格式的简图。

[0014] 图5是路由器-LSA消息的简图。

[0015] 图6是其中交换OSPF外部链路成本信息的一个示例网络的简图。

[0016] 图7是其中交换OSPF内部链路成本信息的一个示例网络的简图。

[0017] 图8是具有多个控制器的示例多区域OpenFlow网络的一个实施例的简图。

[0018] 图9是实现OSPF过程的网络元件的一个实施例的简图。

[0019] 图10是具有分离式架构区域的网络中的OSPF路由选择的过程的一个实施例的流程图。

[0020] 图11是OSPF链路状态通告报头格式的简图。

[0021] 图12是路由器LSA消息的简图。

[0022] 图13是分离式架构OSPF问候消息的简图。

[0023] 图14是从分离式架构控制器向传统路由器发送网络中的问候消息的过程的一个实施例的简图。

具体实施方式

[0024] 在以下描述中提出许多具体细节。但是要理解,即使没有这些具体细节也可实施本发明的实施例。在其它情况下,没有详细示出众所周知的电路、结构和技术,以免影响对本描述的了解。但是,本领域的技术人员将会理解,即使没有这类具体细节,也可实施本发明。通过所包含的描述,本领域的技术人员将能够实现适当的功能性而无需过分实验。

[0025] 将参照图7-9和图14的示范实施例来描述流程图的操作。但是应当理解,图10中的流程图的操作能够通过除了参照图7-9和图14所述之外的本发明的实施例来执行,以及参照图7-9和图14所述的实施例能够执行与参照图10的流程图所述操作不同的操作。

[0026] 附图所示的技术能够使用在一个或多个电子装置(例如端站、网络元件、服务器或者类似电子装置等)上存储和运行的代码及数据来实现。这类电子装置使用非暂时机器可读或计算机可读介质、例如非暂时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、闪速存储器装置和相变存储器)来存储和传递(内部和/或通过网络与其它电子装置)代码及数据。另外,这类电子装置通常包括一组一个或多个处理器,其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)和网络连接之类的一个或多个其它组件。该组处理器和其它组件的耦合通常通过一个或多个总线和桥接器(又称作总线控制器)进行。存储装置表示一个或多个非暂时机器可读或计算机可读存储介质和非暂时机器可读或计算机可读通信介质。因此,给定电子装置的存储装置通常存储供该电子装置的该组一个或多个处理器上执行的代码和/或数据。当然,本发明的一个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] 本文所使用的“网络元件”(例如路由器、交换机、桥接器或者类似连网装置)是包括硬件和软件的一个连网设备,其在通信上互连网络上的其它设备(例如其它网络元件、端

站或者类似连网装置)。一些网络元件是“多服务网络元件”,其提供对多个连网功能(例如路由选择、桥接、交换、第2层聚合、会话边界控制、多播和/或订户管理)的支持,和/或提供对多个应用服务(例如数据收集)的支持。

[0028] 分离式架构区域

[0029] 单个接入/聚合网络能够由多个不同的分离式架构区域(其与多个传统路由器协同工作)组成。如本文所使用的“分离式架构区域”是具有与域相似的独立路由选择的分离式架构网络的一部分。这能够进行,以简化对大地理区域的管理,以获得网络健壮性或者控制平面可缩放性。各分离式架构区域能够由独立控制器来管理。取决于特定应用,这些不同分离式架构区域的控制器需要共享和交换一些信息,以供分离式架构网络的正确管理。

[0030] 开放最短路径优先路由选择

[0031] 传统网络和分离式架构网络均必须计算交换机与网络所服务的其它装置之间的路由。开放最短路径优先(OSPF)是内部网关路由选择协议。OSPF(在RFC 2328中定义)是一种链路状态协议,其中路由器向路由选择域中的所有节点广播其邻居的链路状态信息。使用这个信息,每一个路由器构造域中的整个网络的拓扑图。各路由器保持反映整个网络拓扑的链路状态数据库。基于这个拓扑图和链路成本量度,路由器使用Dijkstra算法来确定到所有其它路由器的最短路径。这个信息又用于创建路由选择表,路由选择表用于因特网协议(IP)分组的转发。

[0032] OSPF实现将路由选择域分为多个“区域”,其由区域边界路由器(ABR)来分隔,以便于管理。各区域通过32位标识符(通常是区域中的主路由器的IP地址)来识别,其中标识符“0”保留用于OSPF网络的核心或主干。链路状态信息的广播限于该区域,并且不在该区域之外发送。在OSPF过程中,各路由器(交换机)为它所属的各区域运行OSPF协议过程的单独副本。如果路由器具有属于不同区域的多个接口,则它运行该过程的多个副本,各接口一个。在开机时,路由器的OSPF过程初始化所有路由选择协议数据结构,并且从下层得到与该路由器的活动接口有关的信息。然后,路由器使用OSPF的问候协议分组来检测其邻居。路由器向其邻居传送问候分组,并且从其邻居接收问候分组。在广播和点对点网络,问候分组在多播地址224.0.0.5上发出。在非广播网络,要求用户配置以发现邻居。在广播和非广播多接入网络(NBMA),问候协议用于推选网络的段的“指定路由器”和“备用指定路由器”。

[0033] 在检测邻居时,路由器将设法建立与其新检测邻居的“相邻性”。相邻性确定区域中的路由选择信息的分配。仅对相邻性传送和接收路由选择更新。在建立相邻性时,路由器将使其“链路状态数据库”与相邻性的另一端的对应路由器同步。对于广播和NBMA网络,指定路由器确定哪些路由器变成相邻的。

[0034] 路由器使用链路状态通告(LSA)来周期地通告其状态,又称作“链路状态”。当路由器的状态发生变化时,也通告链路状态。LSA包含通告路由器的相邻性。路由器在整个区域扩散(flood)其LSA。扩散算法确保信息的可靠性,以确保区域中的所有路由器具有相同的准确链路状态数据库。链路状态数据库由属于该区域的各路由器所始发的LSA的集合组成。

[0035] 各路由器使用这个数据库、以其本身作为根来计算最短路径树。最短路径树又用于创建路由器的路由选择表。OSPF消息以协议号89直接封装在IP分组中传送,而无需使用传输层协议、例如TCP或UDP。另外,OSPF使用其自己的检错和纠错。

[0036] OSPF定义五种不同分组类型。问候分组:OSPF的问候协议分组用于发现和保持相

邻关系。数据库描述分组：这些分组用于形成相邻性。链路状态请求分组：这些分组用于下载相邻路由器之间的链路状态数据库。链路状态更新分组：这些分组用于OSPF的可靠更新机制。单个链路状态更新分组可包含若干路由器的LSA。链路状态Ack分组：这些分组与链路状态更新分组结合用于OSPF的可靠更新机制。仅对相邻性发送OPF协议分组(除了问候分组之外)。因此，所有OSPF协议分组传播单因特网协议(IP)跳，其中一个路由器的IP地址作为源地址，而另一路由器的IP地址作为目的地地址。

[0037] RFC 2328中定义的OSPF规定5种不同类型的LSA。路由器-LSA：这些LSA由区域中的所有路由器来发送。每个LSA包含到区域的路由器接口的状态。路由器-LSA仅在单一区域来扩散。网络-LSA：这些LSA由广播或者NBMA网络的指定路由器来始发。这个LSA包含连接到网络的路由器的列表。与路由器-LSA相似，这个LSA也仅在单一区域来扩散。网络汇总LSA：这个LSA由区域边界路由器(ABR)始发，并且每个LSA描述到该区域外部、但仍然在自主系统(AS)内部的目的地网络的路由。边界汇总LSA：与网络汇总LSA相似，这个LSA由边界区域路由器(ABR)始发，并且它描述到AS边界路由器的路由。AS外部LSA：这个LSA由AS边界路由器始发，并且它们在整个AS扩散。这些LSA的每个描述到另一个AS中的目的地的路由。OSPF分组报头格式如图2所示。问候分组格式在图3中示出。LSA报头格式如图4所示。路由器LSA分组格式如图5所示。

[0038] 分离式架构网络中的OSPF

[0039] 在分离式架构网络中，在控制器之间交换路由选择消息，如同所有其它控制消息。实现分离式架构中的OSPF的一种简单方式是忽略内部成本，并且仅交换与外部成本—即连接不同网络区域的链路的成本—有关的信息。通过这样做，能够在控制器之间交换传统OSPF消息，并且各分离式架构区域能够被认为是单个节点。

[0040] 在图6所示的网络中，例如，控制器B宣布它具有成本为10个单位、到区域C的链路。但是，基于这个信息所找到的路由是未达最佳标准的。在这个实现中，因为内部成本没有提供给其它控制器，所以控制器将无法基于实际端对端成本进行判定。当一个区域中的内部成本从一个内部路径到另一个内部路径极大地改变时，这产生未达最佳标准的路由。

[0041] 在一个备选实施例中，能够通过将内部成本包含在控制器之间所交换的信息中，来修复未达最佳标准的路由的上述问题。各控制器能够计算内部链路成本，并且查找其边界交换机的任何两个之间的最短路径(即，最小费用路径)。例如，又参照图6，当控制器C宣布其自己的区域与控制器F的区域之间的链路成本时，它添加边界交换机S1与S2之间的最短路径成本，并且将其添加到发送给控制器B的成本信息。类似地，控制器C将交换机S2与S3之间的最短路径成本添加到区域C与F之间的成本，并且将这个信息发送给控制器D。

[0042] 这个备选实施例也造成问题。这个实施例要求对同一链路发送多个(以及可能不同的)链路成本消息。例如，在将这个信息发送给控制器B时，与将它发送给控制器D相比，由控制器C所宣布的到达区域F的成本是不同的，因为它们各基于不同的内部路径成本(例如，如果S1与S2之间的内部最短路径成本与S2与S3之间的成本是不同的，则控制器C将宣布其自己的区域与控制器F的区域之间的两个不同链路成本)。最后，这两个矛盾消息将由其它控制器接收，这些其它控制器将第二消息解释为对第一接收消息的更新，而不是将它们解释为两个独立消息。

[0043] 其它实施例例如通过修改OSPF协议以允许将多个成本归于同一链路来尝试修复

这个多成本问题,由分离式架构中的控制器所找到的最短路径可能不是最佳路径。例如,在图7所示的情形中,其中各链路上的数值表明任一个方向的链路的成本,会确定未达最佳标准的路由。在计算从区域A到区域F的最短路径中,将采取下列步骤:(1) C向B和D宣布:到达F的成本 $2(1+1)$;(2) B向A宣布:到达C的成本 $11(10+1)$;(3) D向B宣布:到达C的成本 $21(1+20)$ (注意,在区域D中的边界交换机S4与S5之间仅存在一个内部路径);以及(4) B向A宣布:到达C的成本 $2(1+1)$ 。

[0044] 当控制器A基于上述信息来计算其到区域F的最短路径时,它因其最小成本(为14)而选择路径A-B-C-F而不是其它路径。但是,基于图7所示的所有成本信息,A与F之间的最佳路由选择是A-B-D-E-D-C-F,其成本仅为10。控制器A无法选择这个路径,因为在控制器之间所交换的信息对于这种选择不是足够的。因此,对于最佳路由选择方案,在控制器之间所交换的OSPF消息需要传送与内部链路成本有关的更多信息,如下文进一步论述。

[0045] 具有抽象区域内路径成本的OSPF

[0046] 本发明的实施例提供用于避免现有技术的缺点的方法和系统。如上所述的分离式架构网络中的OSPF的现有技术和简单实现提供未达最佳标准的路由选择解决方案,其中不一定始终准确地确定最短路径,和/或提供过量信息以识别路由,其是低效或者不是缩放的。

[0047] 本发明的实施例克服了现有技术的这些缺点。本发明的实施例将每对边界交换机之间的区域内路径成本抽象为与成本值关联的直接链路。这个解决方案提供最佳路径,能够有效地执行,是可缩放的,并且与非分离式架构中的常规路由器后向兼容。

[0048] 在一个实施例中,在分离式架构中的控制器之间实现OSPF路由选择协议,以建立网络中的任何两个转发元件之间的最佳(最短)路径。本实施例向整个多区域分离式架构网络揭示区域内路径成本和区域间链路成本的必要信息。如上所述,实施例提供最优性、效率、可缩放性和后向兼容性。

[0049] 路由选择协议提供充分和准确信息,其允许各分离式架构控制器单独计算到达任何其它目的地的最佳路径。这是传统域内路由选择协议(OSPF、IS-IS)在传统网络中提供的最优性性质。但是,最佳路径的定义对于在分离式架构网络中的使用略微发生变化。在传统网络中,它是具有最小路由器间成本的路径。在分离式架构网络上下文中,最佳路径是具有最小路由器间成本加上路由器内成本(或者SA区域内成本)的路径。

[0050] 分离式架构OSPF(SA-OSPF)路由选择协议的实施例允许各路由器单独计算最短路径判定。OSPF快速收敛。换言之,与传统网络OSPF相比,分离式架构OSPF协议没有引入附加收敛开销。分离式架构OSPF协议缩放到具有数百交换机的大网络。能够通过所交换的消息数量的开销以及对控制器的存储要求来量化可缩放性。递增部署对于采用关于网络协议的任何新建议是关键的。分离式架构OSPF协议与常规路由器后向兼容。这个性质增加其在实际网络环境中的潜在使用。

[0051] 分离式架构OSPF的实施例将任何边界交换机对之间的区域内路径成本抽象为与成本值关联的直接链路。存在促进这个功能性的分离式架构OSPF的若干方面。区域内成本采用常规OSPF消息的形式来嵌入,以确保后向兼容性。将区域内成本独立于区域间成本传播给所有区域中的所有控制器。这确保最短路径计算的最优性。它确保各控制器具有整个网络(区域间和区域内)的完整图像。对于这种充分区域,计算最短路径能够易于使用

Dijkstra算法进行,如同传统OSPF中一样。

[0052] 分离式架构OSPF是可缩放的,因为将要向外部控制器传播哪一个内部信息经过仔细管理。一种自然方式是发送任何分离式架构交换机对之间的内部成本。但是,这会引起可缩放性问题。此外,这个信息的大多数对于对外部控制器进行的判定不是有用的。分离式架构OSPF仅传播任何边界分离式架构交换机对的总聚合成本。

[0053] 图8是分离式架构网络的一个示例实施例的简图。示例分离式架构网络分为独立分离式架构区域(SA)801A-C。各区域801A-C包含一组交换机。同一区域中的所有交换机由单个逻辑控制器803A-C来控制。在一个实施例中,为了冗余度目的,SA能够实现为主控制器和一组备用控制器。

[0054] 每个SA中的交换机能够是任何类型的路由器、交换机或者类似连网装置,其能够实现分离式架构网络的数据平面。交换机能够包括边界分离式架构交换机和内部分离式架构交换机。边界分离式架构交换机通过连接到不同SA区域中的另一个交换机的接口支持分离式架构功能。边界分离式架构交换机通常由单个SA区域的控制器来控制。在其它实施例中,边界分离式架构交换机能够处于多个SA中,并且具有由各相应SA控制器所控制的接口。内部分离式架构交换机支持分离式架构协议。它由其区域中的控制器来控制。其所有邻居处于同一SA区域中。

[0055] 交换机通过一组链路相互通信。这些链路能够是任何类型的通信介质,包括有线或无线通信介质以及它们的任何组合。链路能够分类为内部链路或者外部链路。内部链路是SA区域中的两个交换机之间的链路,这些交换机可能是边界交换机或者内部SA区域交换机,其属于同一SA区域。外部链路是属于不同SA区域的两个SA交换机之间的链路。在这种情况下,两种SA交换机均为边界SA交换机。

[0056] 链路状态通告(LSA)805是实现分离式架构OSPF的LSA的示例集。每个LSA是一对边界交换机以及遍历这两个边界交换机之间的SA区域的关联成本的抽象。这些LSA由各控制器生成,并且传送给相邻控制器。示例LSA集805是SA区域801B的控制器803B的LSA的集。

[0057] 图9是实现控制器的网络元件的一个实施例的简图。在一个实施例中,控制器901是路由器、交换机或者类似连网装置。控制器901能够包括入口模块903、出口模块905、网络处理器907和存储装置911。入口模块903在物理和链路级操控入局数据业务,并且将这个数据提供给网络处理器供进一步处理。类似地,出口模块905在物理和链路级操控出局数据业务,以便通过连接网络将其传送给其它装置。这两个模块共同起作用,以实现通过网络与其它装置的通信。

[0058] 网络处理器907是处理装置或者一组处理装置,其执行网络元件的功能,包括与网络的控制平面相关的每个功能,其管理网络的数据平面。网络处理器907能够执行一组模块,包括邻居发现模块913、OSPF模块915、拓扑学习模块917、链路状态管理模块919和控制器模块、例如OpenFlow控制器921。

[0059] 另外,网络处理器907能够访问存储装置911中存储的数据。存储装置911中存储的数据能够包括路由选择表923和链路状态数据库925。在其它实施例中,存储装置911能够包括任何数量的独立逻辑或者分布式存储装置以及跨这些装置的所存储数据的任何布置。由网络过程所执行的其它模块也能够从存储装置911来加载或者加载在存储装置911上。

[0060] 邻居发现模块913能够管理用于使用问候协议或者类似协议来与网络中的其它装

置进行通信的协议,以便得到与网络中的每个相邻控制器有关的信息,以实现由控制器所管理的SA区域的交换机之间的正确通信及其配置。任何问候协议或过程能够用于识别SA区域的相邻控制器和交换机。

[0061] 拓扑学习模块917利用由邻居发现模块913所采集的信息来确定控制器在其中进行操作的网络的拓扑。这个拓扑信息由OSPF模块915用于计算通过网络的最佳路由。拓扑信息也由链路状态管理模块919用于跟踪和确定网络中的链路成本。

[0062] OSPF模块917计算源或始发装置与网络中的目的地装置之间的最佳路由。OSPF模块917能够将路由选择信息存储在了一组路由选择表923中。OSPF模块917能够使用由拓扑学习模块917所生成的拓扑信息来计算网络元件之间的路由。在一些实施例中,OSPF模块917还能够以控制器处于根来计算网络的生成树,以建立到每个其它装置的路由。

[0063] 链路状态管理模块919在链路状态数据库925中管理网络或者网络的区域的链路状态信息。链路状态管理模块919还能够生成链路状态通告,以传播给相邻控制器,从而提供与控制器和其它控制器之间以及控制器的SA区域中的连接有关的信息。这个信息能够封装为各边界交换机对的链路状态通告集,其被发送给网络中的其它控制器。

[0064] 控制器921能够是管理SA区域的控制平面的任何类型的分离式架构控制器。控制器能够实现用于管理分离式架构网络的OpenFlow协议或者类似协议。控制器模块921能够与SA区域中的交换机进行通信,以配置数据平面上的分组的转发。控制器921还与其它控制器进行通信,以交换邻居信息、链路状态通告,并且向对等体提供类似信息。

[0065] 图10是实现和支持分离式架构网络中的OSPF的控制器的操作的一个实施例的流程图。为了清楚起见,该过程根据由控制器所执行来描述,但是,该过程能够由控制器的组件(例如OSPF模块)结合特定控制器模块(例如OpenFlow控制器)来执行。能够在激活或重置控制器的时间来初始化该过程。在其它实施例中,该过程是连续或周期的。控制器能够初始化一组数据结构以用于存储和操纵数据,例如OSPF路由选择协议相关数据结构(框1001)。

[0066] 控制器学习其所指配SA的拓扑,包括识别SA区域中作为边界交换机的那些交换机(框1003)。能够通过与其它控制器和传统路由器交换链路状态通告或者链路状态数据库以及通过类似机制来学习拓扑。不同SA区域中的控制器能够通过SA-OSPF协议进行通信。所交换的消息与传统OSPF消息相似,包括问候消息、数据库描述、链路状态请求、链路状态更新和链路状态确认。具体来说,图8中,示出链路状态更新消息805。单个链路状态更新分组可包含若干链路状态通告(LSA)。SA-OSPF的基本路由选择算法的独立副本在各区域中运行。具有到多个区域的接口的路由器运行算法的多个副本。将这个数据汇编到SA区域以及SA区域所在的更宽网络的拓扑图中。

[0067] 控制器(例如控制器的OSPF模块)计算控制器的SA区域中的每对边界交换机之间的最短路径(框1005)。这能够结合计算交换机之间的其它区域内路由来执行。但是,边界交换机对成本预计将与其它控制器和传统路由器共享,以便使那些控制器和传统路由器能够确定跨越控制器的SA区域的最佳路径。使用SA区域的学习拓扑来计算最短路径。在一个实施例中,Dijkstra的算法用于计算任何边界交换机对之间的最短路径。每对的最短路径的计算成本则随拓扑数据或者随OSPF路由选择数据来存储(框1007)。

[0068] 然后,控制器识别相邻控制器(框1009)。在一个实施例中,控制器能够使用SA-OSPF的问候协议来获取邻居。控制器将问候分组发送给其相邻区域中的另一个控制器或者

相邻传统路由器,并且又从其它控制器和传统路由器接收问候分组(框1111)。在广播和点对点网络,控制器通过将其问候分组发送给多播地址,来动态检测其相邻区域中的控制器或传统路由器。在非广播网络,能够提供一些配置信息,以发现相邻控制器。

[0069] 随后,控制器将尝试形成与其新获取相邻控制器的一些相邻控制器的相邻性。通过与相邻控制器交换LSA消息(框1115),在相邻控制器对之间同步链路状态数据库(框1113)。每个LSA消息包含其完整外部链路的成本。控制器不仅向特定邻居发送连接到这个邻居的链路的成本,而且还向所有邻居发送链路的成本。除了发出外部链路的成本之外,控制器还将发出内部链路的成本。LSA消息能够使用两个边界交换机的SA交换机ID连同两个相关接口ID一起来构建。控制器能够周期地通告其SA区域的状态,其又称作链路状态。当区域(控制器)的状态发生变化时,还通告链路状态。区域的相邻性在其LSA的内容中反映。相邻性与链路状态之间的这种关系允许协议及时地检测故障和性能变化。

[0070] LSA跨多个SA区域扩散。扩散过程是可靠的,从而确保整个网络中的所有控制器具有完全相同的链路状态数据库。这个数据库由各区域(其中包括区域间拓扑和区域内拓扑)所始发的LSA的集合组成。

[0071] 从这个链路状态数据库,各控制器以其本身作为根来计算整个网络的最短路径树(框1117)。这个最短路径树又产生OSPF协议的路由选择表。基于这个路由选择信息,控制器将其SA区域中的交换机编程为使用所计算最短路径、通过SA区域路由数据分组(框1119)。交换机的更新能够使用控制平面协议(例如OpenFlow)、通过安全信道或者使用类似过程来完成。

[0072] 当变化在网络中发生时,更新链路状态通告能够由控制器发送(框1121)。将LSA消息发送给相邻控制器,以通知它们关于控制器的SA区域中的变化,使得相邻控制器能够根据需要更新其路由选择。

[0073] 分离式架构OSPF与传统OSPF的不同之处在于,传统路由器看起来像到网络中的其它路由器的单个路由器。但是,SA区域将看起来像到网络中的其它区域/路由器的N个路由器,其中N是区域的边界分离式架构交换机的数量。差别在于,虽然传统OSPF在每个区域的单个路由器上运行单个过程并且看起来像单个路由器,但是SA-OSPF在SA区域的单个控制器上运行单个过程并且看起来像到其它控制器的N个路由器。

[0074] 在一个实施例中,对于由传统路由器中的OSPF过程所发送的每一个问候分组,单个SA-OSPF过程发送N个不同问候分组,以表示N个边界SA交换机的每个。这些问候分组在连接到其它路由器/SA区域的接口上发送。类似地,对于由传统路由器中的OSPF过程所发送的每一个路由器-LSA分组,单个SA-OSPF过程发送N个不同路由器-LSA分组,以表示N个边界SA交换机的每个。这些LSA分组在连接到其它路由器/SA区域的接口上发送。

[0075] 简言之,SA区域的控制器管理OSPF,好像它具有N个OSPF控制器过程,其中N是SA区域中的边界分离式架构交换机的数量。这确保与运行于连接传统路由器的OSPF的后向兼容性。因此,控制器发出N个问候分组、N个LSA消息等。对于传统网络中的相同数量的路由器,在分离式架构网络中,与传统OSPF相比,SA-OSPF使用更少数量的协议分组。节省来自无需在属于同一SA区域的边界SA交换机之间发送OSPF协议分组。

[0076] LSA消息

[0077] 图11是OSPF链路状态通告报头格式的简图。为了确保SA-OSPF是后向兼容的,利用

与标准OSPF消息格式相同的格式(即,字段的相同序列和字段的相同长度)。但是,向一些字段指配的值能够是不同的。

[0078] 图12是SA-OSPF LSA消息、具体来说路由器-LSA类型消息的简图。路由器-LSA由区域中的各交换机生成。它描述区域中的区域接口的状态。这个消息由所有控制器始发。这个LSA描述到多域SA网络的路由器接口的所收集状态。它在全部控制器扩散。

[0079] 在一个实施例中,路由器-LSA消息与传统路由器-LSA消息在消息的四个字段中不同(图12所示):自身交换机ID、自身接口ID、相邻接口ID和相邻交换机ID。交换机ID是唯一的32位标识符(其指配给分离式架构中的各交换机)。这个ID在多区域SA域中的所有交换机之间必须是唯一的。在一个实施例中,交换机的最高IP地址用作交换机ID。自身交换机ID是通告交换机的标识符,以及相邻交换机ID是相邻交换机的标识符。接口ID是32位标识符(其指配给交换机的各接口)。这个标识符仅需要在单个交换机中是唯一的。自身接口ID是通告交换机的接口ID,以及相邻接口ID是(相邻交换机的)相邻接口的接口ID。

[0080] 问候消息的操控

[0081] OSPF问候协议是一种机制,其允许路由器建立和保持与相邻路由器的相邻性。相邻路由器交换问候消息。在广播和点对点网络中,路由器能够通过向多播地址发送问候消息,来动态检测其相邻路由器。在接收问候分组时,路由器形成邻居之间的相邻性。在相邻路由器对之间同步链路状态数据库以及路由选择更新。在建立相邻之后,路由器仍然需要周期地交换问候消息,以指示它是有效的。总之,问候消息的目的是发现和保持相邻关系。

[0082] 图13是SA-OSPF中的问候消息的简图。在SA-OSPF协议中,问候消息能够未改变。但是,分离式架构控制器需要构建附加问候消息,以保持相邻路由器与它所连接的分离式架构交换机之间的相邻性。

[0083] 图14是分离式架构控制器与传统常规路由器之间所交换的问候消息的简图。在SA-OSPF中,控制器首先代表边界交换机来构建OSPF问候消息。它在交换机ID的字段中使用边界交换机的最高IP地址。分组首先被注入交换机S1,并且然后在出口链路上发送给相邻路由器。

[0084] 这样,描述了用于实现分离式架构网络中的OSPF的方法、系统和设备。要理解,预计以上描述是说明性而不是限制性的。通过阅读和了解以上描述,许多其它实施例将是本领域的技术人员显而易见的。因此,本发明的范围应当参照所附权利要求连同这类权利要求涵盖的完整等效范围共同确定。

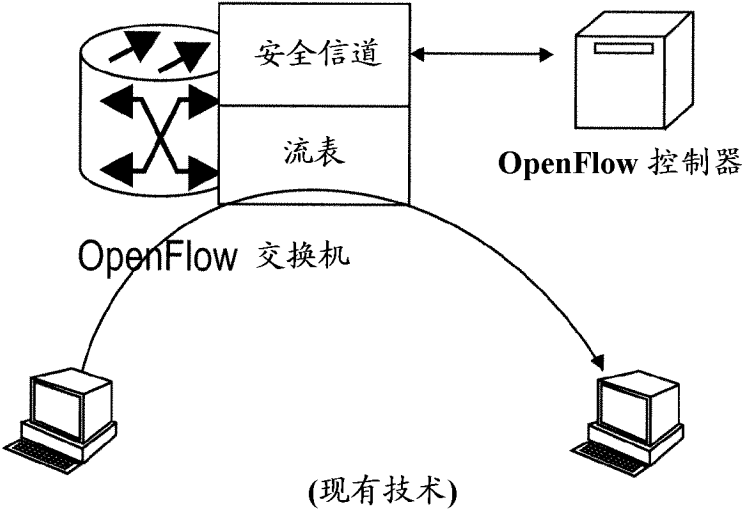


图 1



图 2

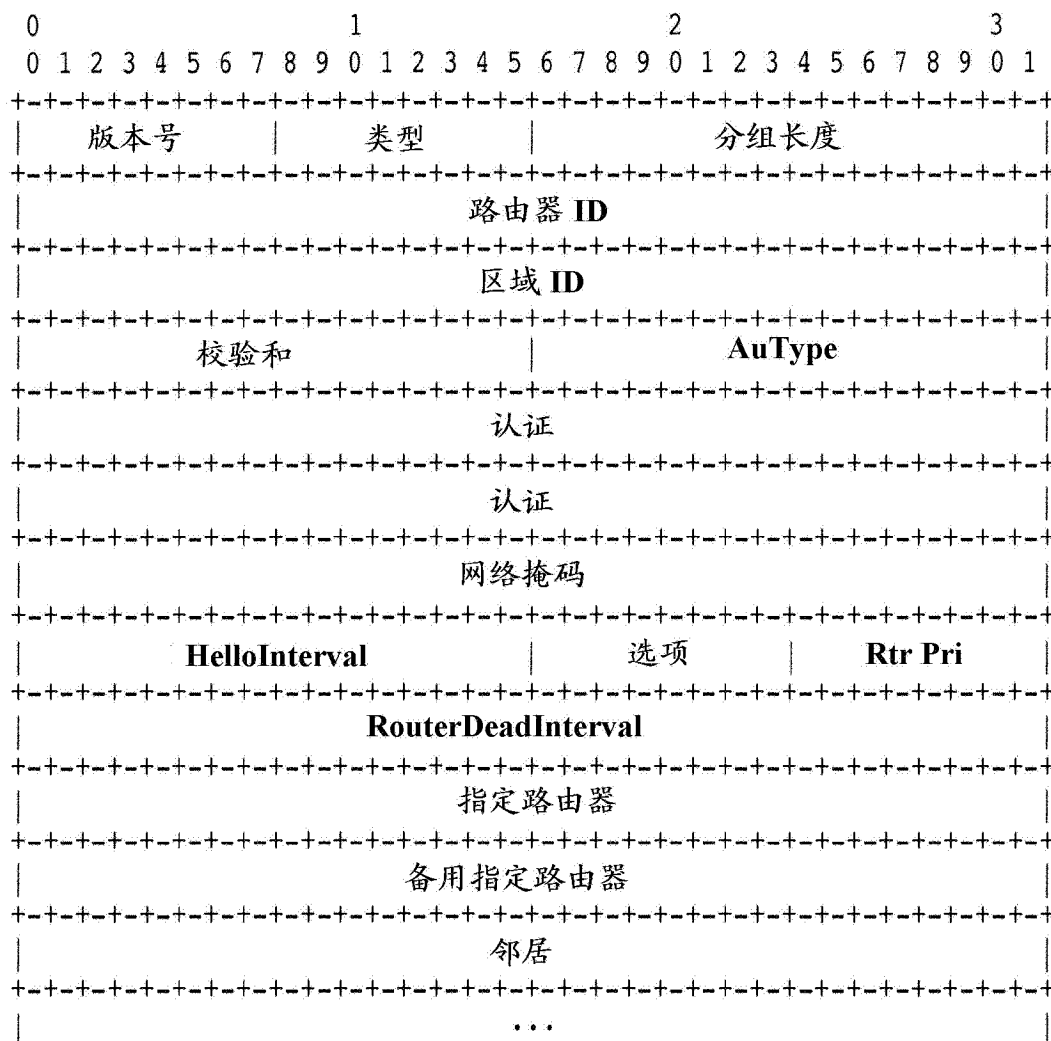


图 3

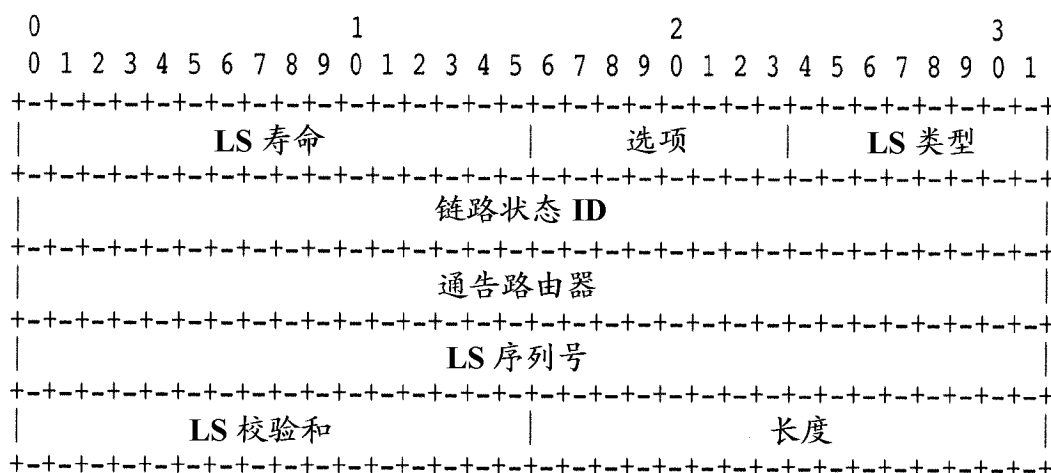


图 4



图 5

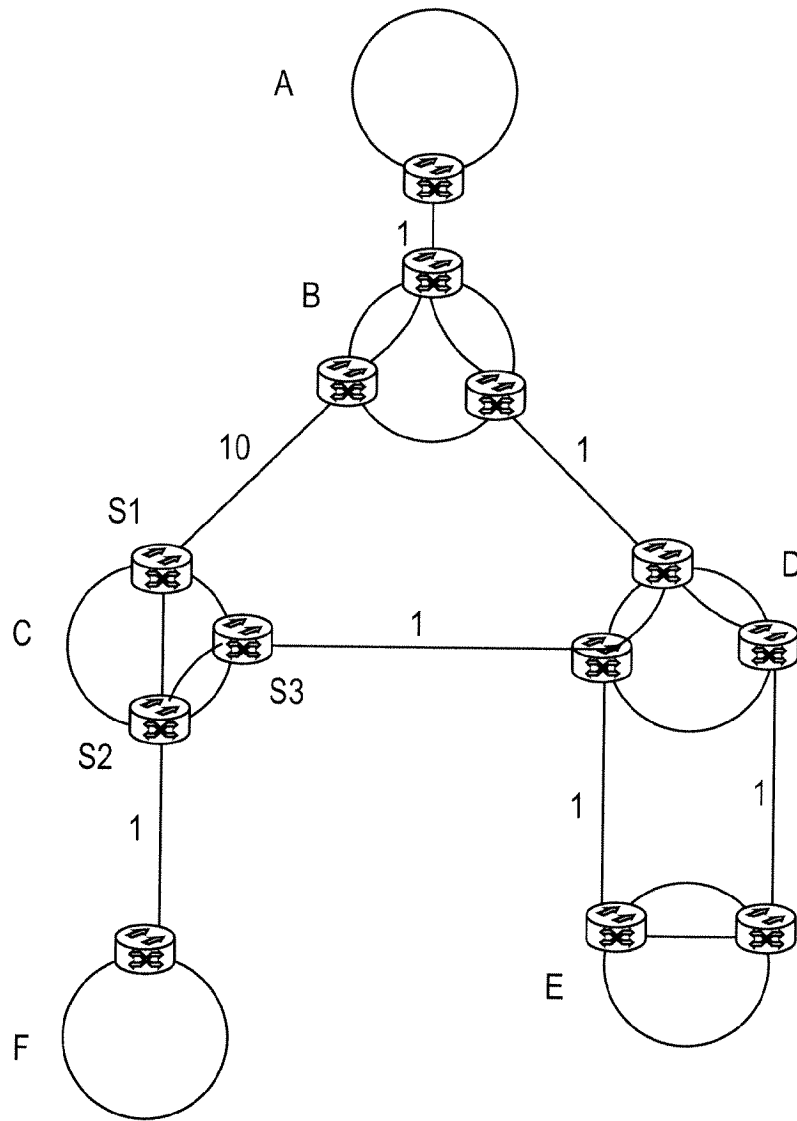


图 6

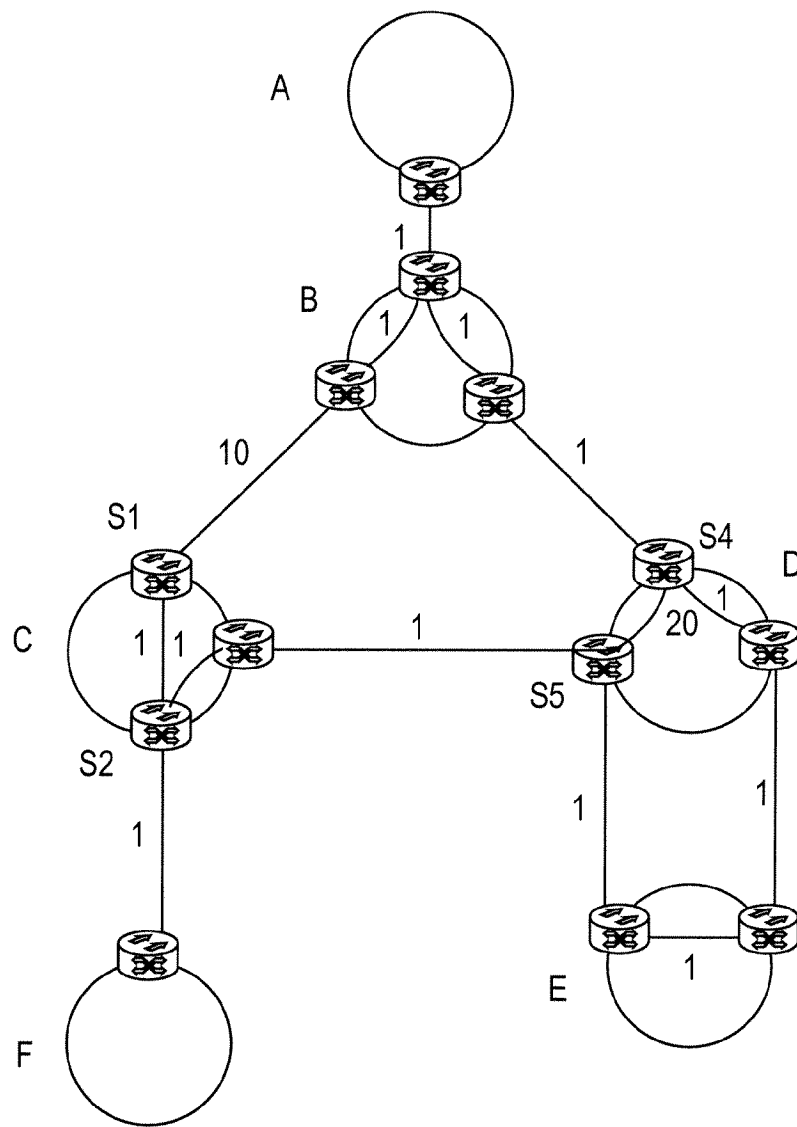


图 7

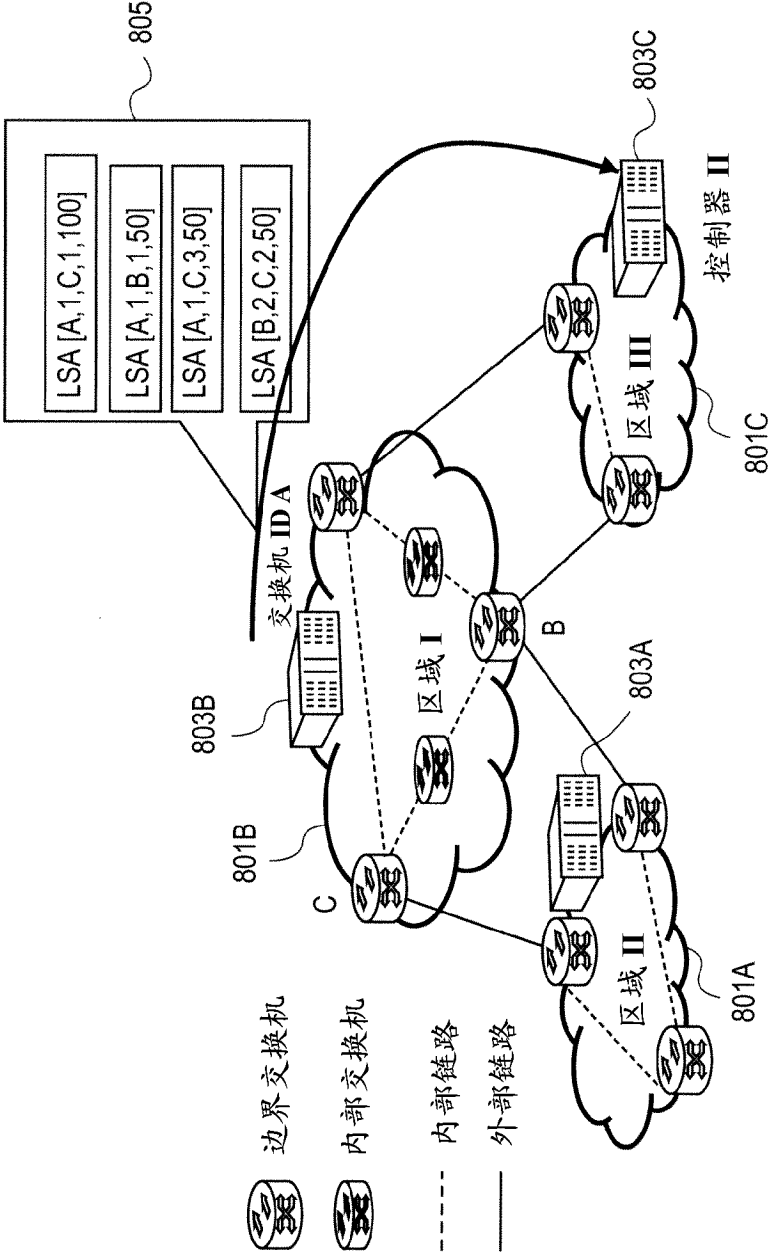


图 8

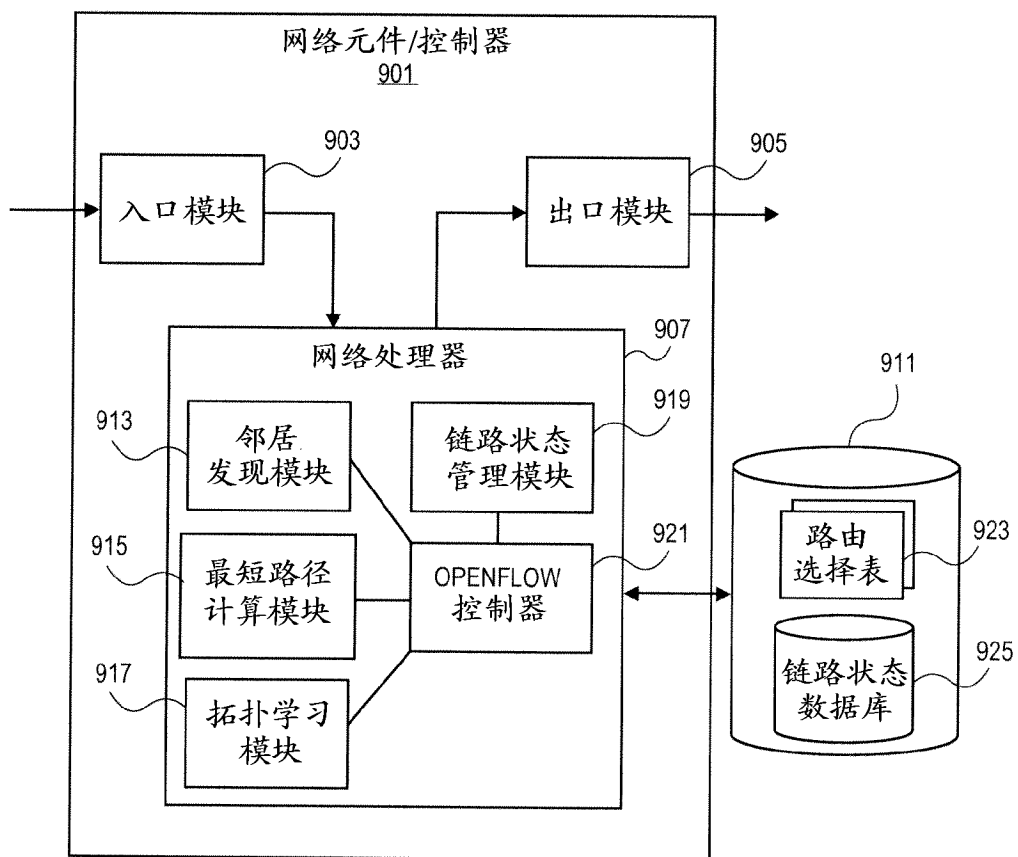


图 9

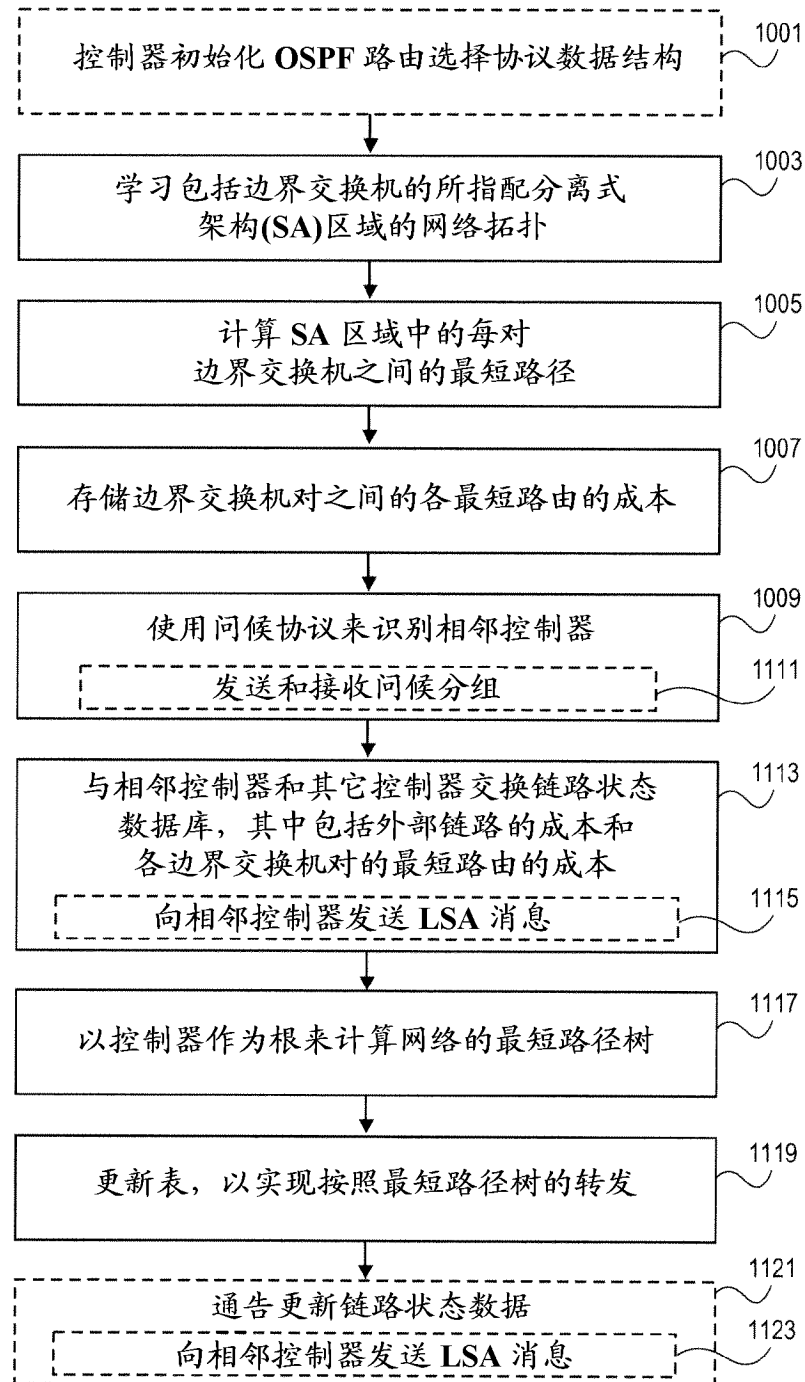


图 10

位/字节	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
0	LS 寿命																LS 类型															
32	链路状态 ID																															
64	通告路由器																															
96	LS 序列号																															
128	LS 校验和																长度															
160	根据“LS 类型”，内容细节将改变。																															
192																																
-																																
-																																
-																																
-	...																															

图 11

位/字节	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
0	LS 寿命																0	0	1	1												
32	链路状态 ID																															
64	自身交换机 ID																															
96	LS 序列号																															
128	LS 校验和																长度															
160	0	Nt	x	V	E	B	选项																									
192	类型						0						量度																			
224	自身接口 ID																															
256	相邻接口 ID																															
288	相邻交换机 ID																															
320	...																															
-	类型						0						量度																			
-	接口 ID																															
-	相邻接口 ID																															
-	相邻路由器 ID																															
-	...																															

图 12

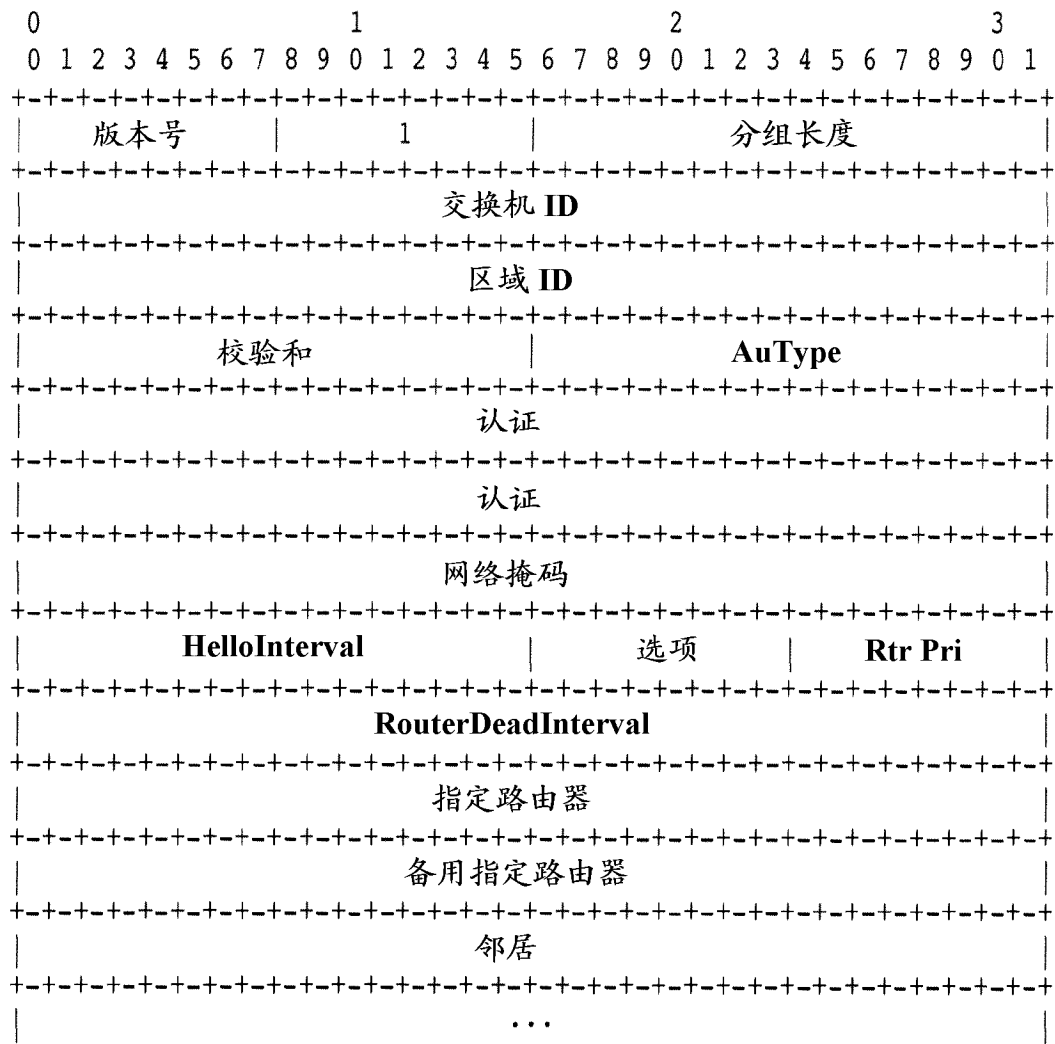


图 13

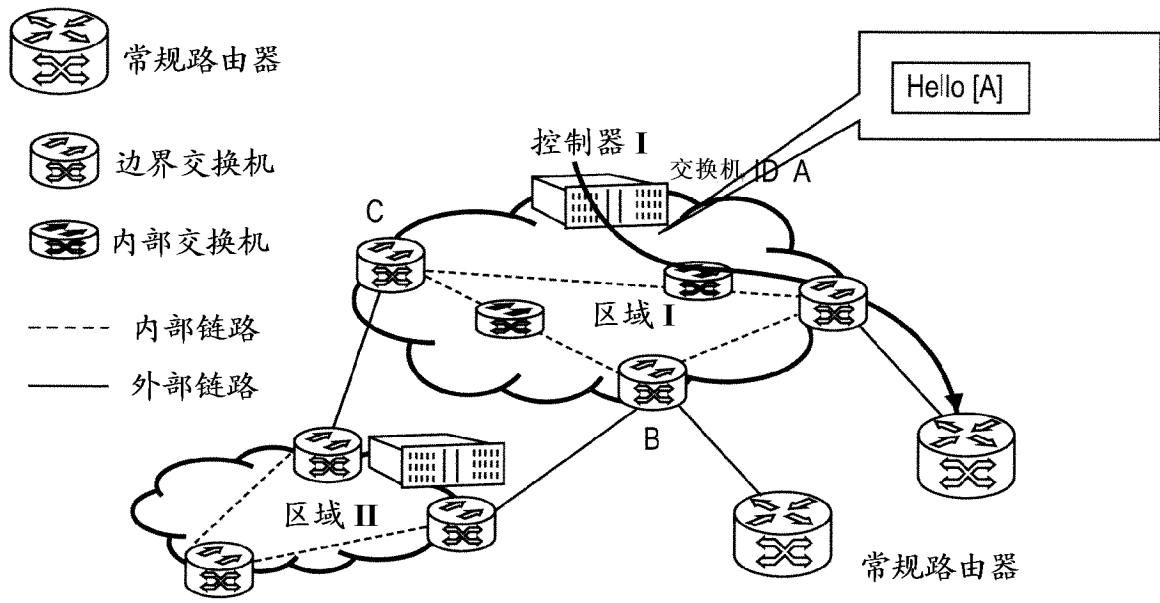


图 14