



(12)发明专利

(10)授权公告号 CN 103516632 B

(45)授权公告日 2017. 05. 10

(21)申请号 201310159634.7

(22)申请日 2013.04.22

(65)同一申请的已公布的文献号

申请公布号 CN 103516632 A

(43)申请公布日 2014.01.15

(30)优先权数据

13/538,328 2012.06.29 US

(73)专利权人 瞻博网络公司

地址 美国加利福尼亚州

(72)发明人 K·纳拉亚纳斯瓦米 J-M·弗雷朗

A·文卡特拉玛尼 S·加甘纳德汗

(74)专利代理机构 北京市金杜律师事务所

11256

代理人 王茂华

(51)Int.Cl.

H04L 12/931(2013.01)

H04L 12/70(2013.01)

(56)对比文件

US 2007180513 A1,2007.08.02,全文.

US 2007083625 A1,2007.04.12,全文.

审查员 胡智权

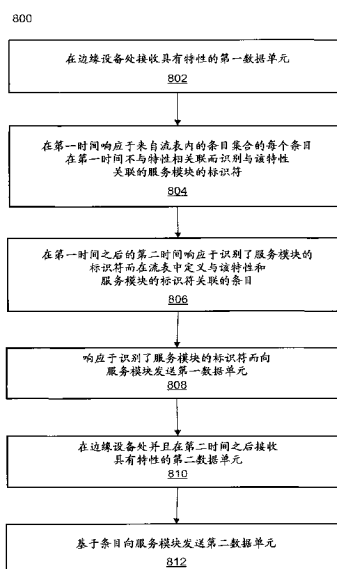
权利要求书3页 说明书18页 附图5页

(54)发明名称

用于在分布式交换机中提供服务的方法和装置

(57)摘要

本发明提供了一种用于在分布式交换机中提供服务的方法和装置。在一些实施例中,一种非瞬态处理器可读介质存储代表将由处理器执行的指令的代码。该代码使得处理器在边缘设备处接收具有特性的第一数据单元。该代码使得处理器在第一时间响应于来自流表内的条目集合的每个条目不与该特性关联而识别与该特性关联的服务模块的标识符。该代码使得处理器在流表中定义与该特性和服务模块的标识符关联的条目。该代码使得处理器向服务模块发送第一数据单元。该代码使得处理器在边缘设备处接收具有特性的第二数据单元并且基于所述条目向服务模块发送第二数据单元。



1. 一种用于提供服务的装置,包括:

存储器;以及

处理器,被配置为:

在边缘设备处接收具有一个特性的第一数据单元;

在第一时间响应于来自流表内的条目集合的每个条目在所述第一时间不与所述特性相关联而识别与所述特性相关联的服务模块的标识符;

在所述第一时间之后的第二时间,响应于识别了所述服务模块的所述标识符而在所述流表中定义与所述特性和所述服务模块的所述标识符相关联的条目;

响应于识别了所述服务模块的所述标识符而向所述服务模块发送所述第一数据单元;

在所述边缘设备处并且在所述第二时间之后接收具有所述特性的第二数据单元;以及基于所述条目向所述服务模块发送所述第二数据单元。

2. 根据权利要求1所述的装置,其中所述服务模块与安全服务或者负载平衡服务中的至少一个服务相关联。

3. 根据权利要求1所述的装置,其中所述特性包括源外围处理设备的标识符、目的地外围处理设备的标识符、所述第一数据单元的类型或者与所述第一数据单元相关联的协议中的至少一项。

4. 根据权利要求1所述的装置,其中所述处理器还被配置为:响应于在接收到所述第二数据单元之后的预定时间段内未接收到具有所述特性的第三数据单元而从所述流表中移除所述条目。

5. 根据权利要求1所述的装置,其中所述处理器被配置为通过进一步基于哈希函数来识别所述服务模块的所述标识符而进行识别,所述哈希函数使用源外围处理设备的标识符、目的地外围处理设备的标识符、所述第一数据单元的类型或者与所述第一数据单元相关联的协议中的至少一项作为输入。

6. 根据权利要求1所述的装置,其中所述处理器被配置为通过进一步基于哈希函数来识别所述服务模块的所述标识符而进行识别,并且所述处理器还被配置为基于所述条目并且在不使用所述哈希函数的情况下向所述服务模块发送所述第二数据单元。

7. 一种用于提供服务的装置,包括:

边缘设备,被配置为从第一外围处理设备接收第一数据单元,所述边缘设备被配置为基于与所述第一数据单元相关联的特性来识别与将对所述第一数据单元提供的服务相关联的第一服务模块的标识符,所述边缘设备被配置为向所述第一服务模块发送所述第一数据单元,

所述边缘设备被配置为从第二外围处理设备接收第二数据单元,所述边缘设备被配置为基于与所述第二数据单元相关联的特性来识别与将对所述第二数据单元提供的服务相关联的所述第一服务模块的所述标识符,所述边缘设备被配置为至少部分地基于所述边缘设备向所述第一服务模块发送所述第一数据单元来向与所述第一服务模块不同的第二服务模块发送所述第二数据单元。

8. 根据权利要求7所述的装置,其中所述边缘设备被配置为经由分布式交换结构向所述第一服务模块发送所述第一数据单元,以使得所述第一服务模块经由所述分布式交换结构向第三外围处理设备发送所述第一数据单元。

9. 根据权利要求7所述的装置,其中所述边缘设备被配置为在第一时间接收所述第二数据单元,所述边缘设备被配置为在所述第一时间之后的第二时间从所述第二外围处理设备接收第三数据单元,所述边缘设备被配置为至少部分地基于所述边缘设备向所述第二服务模块发送所述第二数据单元而向所述第二服务模块发送所述第三数据单元。

10. 根据权利要求7所述的装置,其中所述边缘设备被配置为在流表中存储所述第一服务模块的所述标识符和与所述第一数据单元相关联的所述特性之间的关联,所述边缘设备被配置为至少部分地基于所述关联向所述第二服务模块而不向所述第一服务模块发送所述第二数据单元。

11. 根据权利要求7所述的装置,其中所述边缘设备被配置为在流表中存储所述第一服务模块的所述标识符和与所述第一数据单元相关联的所述特性之间的关联,所述边缘设备被配置为至少部分地基于所述关联在包括所述第一服务模块和所述第二服务模块的多个服务模块之间对数据进行负载平衡。

12. 根据权利要求7所述的装置,其中所述边缘设备被配置为从所述第一外围处理设备并且在一个时间段期间接收包括所述第一数据单元的多个数据单元,所述边缘设备被配置为在所述时间段之后接收所述第二数据单元,所述边缘设备被配置为至少部分地基于所述边缘设备向所述第一服务模块发送来自所述多个数据单元的数据单元的速率而向所述第二服务模块而不向所述第一服务模块发送所述第二数据单元。

13. 根据权利要求7所述的装置,其中所述服务是安全服务或者负载平衡服务中的至少一个服务。

14. 根据权利要求7所述的装置,其中与所述第一数据单元相关联的所述特性包括所述第一外围处理设备的标识符、目的地外围处理设备的标识符、所述第一数据单元的类型或者与所述第一数据单元相关联的协议中的至少一项。

15. 根据权利要求7所述的装置,其中所述第二外围处理设备不同于所述第一外围处理设备。

16. 根据权利要求7所述的装置,其中所述第二外围处理设备是所述第一外围处理设备。

17. 一种用于提供服务的装置,包括:

边缘设备,被配置为接收第一数据单元,所述边缘设备被配置为基于与所述第一数据单元相关联的特性来识别与将对所述第一数据单元提供的服务相关联的服务模块的标识符,所述边缘设备被配置为向所述服务模块发送所述第一数据单元,以使得所述服务模块对所述第一数据单元执行所述服务,

所述边缘设备被配置为响应于对所述第一数据单元执行了所述服务并且从所述服务模块接收与对所述第一数据单元提供的所述服务相关联的指示,所述边缘设备被配置为接收具有所述特性的第二数据单元,所述边缘设备被配置为响应于接收到所述指示来对所述第二数据单元执行所述服务而不向所述服务模块发送所述第二数据单元。

18. 根据权利要求17所述的装置,其中所述边缘设备被配置为响应于接收到所述指示而在流表中存储所述特性和与所述服务相关联的所述指示之间的关联,所述边缘设备被配置为响应于所述关联而对所述第二数据单元执行所述服务。

19. 根据权利要求17所述的装置,其中所述边缘设备被配置为经由分布式交换结构向

所述服务模块发送所述第一数据单元,从而使得所述服务模块经由所述分布式交换结构向外围处理设备发送所述第一数据单元,所述边缘设备被配置为经由所述分布式交换结构向所述外围处理设备发送所述第二数据单元而不向所述服务模块发送所述第二数据单元。

20.根据权利要求17所述的装置,其中所述边缘设备被配置为响应于接收到所述指示而在流表中存储所述特性和与所述服务相关联的所述指示之间的关联,所述边缘设备被配置为响应于所述边缘设备未在预定时间段内接收到具有所述特性的数据而移除所述特性与所述指示之间的所述关联。

21.根据权利要求17所述的装置,其中所述服务是安全服务或者负载均衡服务中的至少一个服务。

22.根据权利要求17所述的装置,其中所述特性包括源外围处理设备的标识符、目的地外围处理设备的标识符、所述第一数据单元的类型或者与所述第一数据单元相关联的协议中的至少一项。

用于在分布式交换机中提供服务的方法和装置

[0001] 相关申请的交叉引用

[0002] 本申请与具有代理案号JUNI-189/00US108200-2425、提交于相同日期并且标题为“Methods and Apparatus for Providing Services in a Distributed Switch”的共同未决美国专利申请相关,通过引用将该申请整体结合于此。

技术领域

[0003] 这里描述的一些实施例总地涉及大型数据中心和企业网络,并且具体地涉及用于在例如单层数据中心内的分布式交换机中提供服务的方法和装置。

背景技术

[0004] 一些已知的数据中心以能够将典型三层架构(例如,接入层、聚合层和核心层)压缩成单层架构这样的方式布置网络设备。在这样的单层架构中,多个边缘设备,诸如架顶(TOR)网络设备通过多个链路互连。TOR网络设备连接到为数据中心的部件提供各种功能和服务的多个服务器设备。

[0005] 然而,存在对于能够在单层架构的数据中心中高效提供功能和服务的方法和装置的需要。

发明内容

[0006] 在一些实施例中,一种非瞬态处理器可读介质存储代表将由处理器执行的指令的代码。该代码使得处理器在边缘设备处接收具有一个特性的第一数据单元。该代码使得处理器在第一时间响应于来自流表内的条目集合中的每个条目不与该特性相关联来识别与该特性相关联的服务模块的标识符。该代码使得处理器定义流表中的与该特性和服务模块的标识符相关联的条目。该代码使得处理器向服务模块发送第一数据单元。该代码使得处理器在边缘设备处接收具有该特性的第二数据单元并且基于该条目向服务模块发送第二数据单元。

附图说明

[0007] 图1是根据一个实施例的被配置为向数据业务提供服务的交换结构系统的示意图示。

[0008] 图2是根据一个实施例的多级交换结构的示意图示。

[0009] 图3是根据一个实施例的边缘设备的框图。

[0010] 图4是根据一个实施例的管理模块的框图。

[0011] 图5是根据一个实施例的操作地耦合到服务模块集合的边缘设备的示意图示。

[0012] 图6是根据一个实施例的在边缘设备处实施的流表的示意图示。

[0013] 图7是根据另一实施例的在边缘设备处实施的流表的示意图示。

[0014] 图8是图示根据一个实施例的如下方法的流程图,该方法用于识别用于对跨分布

式交换结构传输的数据执行服务的服务模块。

具体实施方式

[0015] 在一些实施例中,一种非瞬态处理器可读介质存储代表将由处理器执行的指令的代码。在非瞬态处理器可读介质中存储的代码包括用于使得处理器在边缘设备处接收具有一个特性的第一数据单元(例如,数据单位、数据分组、数据分组的报头部分、数据分组的净荷部分等)的代码。这样的特性可以例如是源外围处理设备的标识符、目的地外围处理设备的标识符、第一数据单元的类型或者与第一数据单元关联的协议。

[0016] 非瞬态处理器可读介质包括用于使得处理器在第一时间响应于来自流表内的条目集合中的每个条目不与特性相关联来识别与该特性相关联的服务模块的标识符的代码。在一些实施例中,该代码使得处理器例如基于使用特性作为输入的哈希函数来识别服务模块的标识符。在一些实施例中,服务模块可以例如向数据单元提供安全服务或者负载平衡服务。

[0017] 非瞬态处理器可读介质包括用于使得处理器在第一时间之后的第二时间响应于识别了服务模块的标识符而在流表中定义与特性和服务模块的标识符相关联的条目的代码。非瞬态处理器可读介质也包括用于使得处理器响应于识别了服务模块的标识符而向服务模块发送第一数据单元的代码。

[0018] 非瞬态处理器可读介质还包括用于使得处理器在边缘设备处并且在第二时间之后接收具有该特性的第二数据单元并且基于条目向服务模块发送第二数据单元的代码。在一些实施例中,该代码使得处理器基于该条目并且在不使用例如哈希函数的情况下向服务模块发送第二数据单元。此外,在一些实施例中,非瞬态处理器可读介质包括用于使得处理器响应于在接收到第二数据单元之后的预定时间段内未接收到具有该特性的第三数据单元而从流表中去除该条目的代码。

[0019] 如这里所使用的术语“物理跳跃”可以包括在两个模块和/或设备之间的物理链路。例如,操作地将外围处理设备与边缘设备耦合的数据路径可以被称为物理跳跃。换言之,物理跳跃可以在物理上将外围处理设备与边缘设备链接。

[0020] 如这里所使用的术语“单个物理跳跃”可以包括在系统中的两个设备之间的直接物理连接。换言之,单个物理跳跃可以包括如下链路,经由该链路两个设备在没有任何中间模块的情况下被耦合。因而,例如如果外围处理设备经由单个物理跳跃耦合到边缘设备,则该外围处理设备可以向该边缘设备直接发送数据分组而无需通过任何居间模块来发送这些数据分组。

[0021] 如这里所使用的术语“单个逻辑跳跃”意指如下物理跳跃和/或物理跳跃组,该物理跳跃和/或物理跳跃组是与第一协议关联的网络拓扑内的单个跳跃。换言之,根据与第一协议关联的拓扑,没有居间节点存在于经由物理跳跃和/或物理跳跃组操作地耦合到第二模块和/或设备的第一模块和/或设备之间。经由单个逻辑跳跃连接到第二模块和/或设备的第一模块和/或设备无论在第一设备与第二设备之间的物理跳跃的数目如何,都可以使用与第一协议和第二模块和/或设备关联的目的地地址来向第二模块和/或设备发送数据分组。在一些实施例中,例如第二协议可以使用第一协议的目的地地址来通过单个逻辑跳跃向第二模块和/或设备路由来自第一模块和/或设备的数据分组和/或数据单位。换而

言之,在第一模块和/或设备经由第一协议的单个逻辑跳跃向第二模块和/或设备发送数据时,第一模块和/或设备将单个逻辑跳跃视为如同它向第二模块和/或设备直接发送数据一样。

[0022] 在一些实施例中,交换结构(例如,分布式交换结构)可以充当单个逻辑跳跃(例如,单个大规模联合L2/L3交换机)的一部分。交换结构的各部分可以在物理上跨例如通过多个物理跳跃互连的许多底盘和/或模块而分布。这样的交换结构被称为分布式交换结构。在一些实施例中,例如可以在第一底盘中包括分布式交换结构的一级并且可以在第二底盘中包括分布式交换结构的另一级。这两级可以在逻辑上充当单个联合交换机的一部分(例如,在根据第一协议的相同逻辑跳跃内)、但是包括在联合交换机内的相应成对级之间的分离的单个物理跳跃。换言之,物理跳跃可以操作地耦合分布式交换结构内的每一级,该每一级代表与用来在分布式交换结构以外路由数据的协议关联的单个逻辑跳跃。此外,与用来在单个逻辑跳跃以外路由数据的协议关联的分组分类和转发无需在单个逻辑跳跃内的每一级发生。在一些实施例中,例如与第一协议(例如,以太网)关联的分组分类和转发可以在模块和/或设备经由单个逻辑跳跃向另一模块和/或设备发送数据分组之前发生。

[0023] 如在本说明书中所使用的单数形式“一”和“该”包括复数引用,除非上下文另有明示。因此,例如术语“服务模块”旨在于意指单个服务模块或者服务模块的组合。

[0024] 图1是根据一个实施例的被配置为向数据业务提供服务的交换结构系统100的示意图示。交换结构系统100包括交换结构110、管理模块150和多个边缘设备(例如,边缘设备181-183)。交换结构系统100操作地将多个外围处理设备(例如,外围处理设备111-116)相互耦合。如图1中所示,每个外围处理设备111-116操作地耦合到交换结构系统100的边缘设备181-183。具体而言,外围处理设备111和112操作地耦合到边缘设备182;外围处理设备113和114操作地耦合到边缘设备181;外围处理设备115和116操作地耦合到边缘设备183。

[0025] 每个外围处理设备111-116可以是能够向和/或从操作地耦合到该外围处理设备111-116的边缘设备181-183发送数据和/或接收数据的任何设备。外围处理设备111-116可以例如是计算节点、服务节点、路由器和存储节点等。在一些实施例中,例如外围处理设备111-116包括服务器、存储设备、网关、工作站等。

[0026] 外围处理设备111-116可以使用任何适当连接,诸如例如光连接(例如,光线缆和光连接器)、电连接(例如,电线缆和电连接器)、无线连接(例如,无线收发器)等来操作地耦合到交换结构系统100的边缘设备181-183。这样,外围处理设备111-116可以被配置为经由包括边缘设备181-183和交换结构110的交换结构系统100向其它外围处理设备111-116发送数据(例如,数据分组、数据单位)。在一些实施例中,在外围处理设备111-116与边缘设备181-183之间的每个连接是直接链路(例如,有线链路、无线链路)。这样的链路可以视为单个物理跳跃链路。在其它实施例中,每个外围处理设备111-116可以经由中间模块(在图1中未示出)操作地耦合到边缘设备181-183。

[0027] 每个边缘设备181、182、183可以是被配置为将外围处理设备111-116操作地耦合到交换结构110的任何设备。在一些实施例中,边缘设备181-183可以例如是接入交换机、输入/输出模块、架顶(TOR)设备等。在结构上,边缘设备181-183可以充当源边缘设备和目的地边缘设备二者。因而,边缘设备181-183可以向和从交换结构110以及向和从连接的外围处理设备111-116发送数据(例如,数据分组或者数据单位的数据流)和接收数据。

[0028] 每个边缘设备181、182、183被配置为经由交换结构110与其它边缘设备181-183通信。具体而言,交换结构110被配置为以相对低的延时来提供在边缘设备181-183之间的任何设备到任何设备的连通。也就是说,交换结构110可以被配置为在边缘设备181-183之间传输(例如,传送)数据。在一些实施例中,交换结构110可以具有至少数以百计或者数以千计的端口(例如,外出端口和/或进入端口),边缘设备181-183可以通过这些端口发送和/或接收数据。在一些实施例中,每个边缘设备181、182、183可以被配置为在跨交换结构110的多个数据路径上与其它边缘设备181、182、183通信。换言之,在一些实施例中,多个数据路径在交换结构110内存在于边缘设备181、182或者183与另一边缘设备181、182或者183之间。在一些实施例中,交换结构系统100可以具有单层架构。在这样的单层架构中,每个边缘设备被配置为经由单个物理跳跃连接到每个其它边缘设备并且与每个其它边缘设备通信。换言之,每个边缘设备直接连接到交换结构系统中的每个其它边缘设备。

[0029] 图3是根据一个实施例的边缘设备300的框图。边缘设备300可以在结构上和功能上与关于图1示出和描述的边缘设备181-183相似。具体而言,边缘设备300可以将(在图3中未示出,但是与图1中的外围处理设备111-116相似的)一个或者多个外围处理设备操作地耦合到(在图3中未示出,但是与图1中的交换结构110相似的)交换结构。

[0030] 如图3中所示,边缘设备300包括处理器330;存储器350,该存储器包括过滤器模块355、选择模块353、服务执行模块356和流表354;线路卡(line card)310、320;以及端口331。处理器330操作地耦合到存储器350、线路卡310、线路卡320和端口331。线路卡310包括端口311和312。线路卡320包括端口321和322。在一些实施例中,线路卡310和/或320可以包括一个或者多个处理器和/或存储器。在一些实施例中,端口可以是任何如下实体,该实体可以活动地与耦合的设备通信(例如,向和/或从耦合的设备发送数据和/或接收数据)或者通过网络进行通信。这样的端口不必是硬件端口,而可以是虚拟端口或者软件定义的端口。

[0031] 在一些实施例中,端口311、312、321和/或322可以例如耦合到与边缘设备300耦合的多个外围处理设备并且与这些外围处理设备通信。在这样的实施例中,端口311、312、321和/或322可以实施一个或者多个物理层(例如,使用光纤信令的物理层、使用双绞线电信令的物理层)。另外,端口311、312、321和/或322可以允许边缘设备300经由一个或者多个通信协议(例如,光纤信道协议、以太网协议)与多个外围处理设备通信。因此,边缘设备300可以经由端口311、312、321和/或322使用同构或者异构物理层和/或通信协议来与多个外围处理设备通信。

[0032] 在一些实施例中,端口331可以连接到交换结构内的设备(例如,切换设备、路由设备),从而使得边缘设备300可以经由交换结构操作地耦合到其它边缘设备。端口331可以是一个或者多个网络接口(例如,40吉比特(Gb)以太网接口、100Gb以太网接口等)的一部分,边缘设备300可以通过该一个或者多个网络接口向和/或从交换结构发送数据和/或接收数据。可以例如经由操作地耦合到边缘设备300的光链路、电链路或者无线链路向和/或从交换结构发送和/或接收数据。在一些实施例中,边缘设备300可以基于一个或者多个通信协议(例如,光纤信道协议、以太网协议)向和/或从交换结构发送数据和/或接收数据。

[0033] 在一些实施例中,端口331可以实施与在端口311、312、321和322处实施的物理层和/或通信协议不同的物理层和/或通信协议。例如端口311、312、321和322可以使用基于数据分组的通信协议来与外围处理设备通信,并且端口331可以使用基于数据单位的通信协

议来与交换结构通信。在一些实施例中,边缘设备300可以是网络交换机、诸如分布式网络交换机的边缘设备。

[0034] 在一些实施例中,存储器350可以例如是随机存取存储器(RAM)(例如,动态RAM、静态RAM)、闪存、可移除存储器等。在一些实施例中,可以将流表354实施为例如关系数据库、表等。在一些实施例中,可以在存储器350内的位置处存储流表354。在一些实施例中,虽然在图3中未示出,但是边缘设备300可以包括一个或者多个流表。关于图6和7进一步示出和描述了流表(例如,流表354)的细节。

[0035] 在一些实施例中,过滤器模块355、服务选择模块353和服务执行模块356可以例如是(在硬件中存储和/或执行的)过程、应用、虚拟机和/或某一其它软件模块或者硬件模块。例如,可以在存储器350内存储并且在处理器330处执行实施过滤器模块355、选择模块353和服务执行模块356的指令。过滤器模块355可以被配置为识别期望的服务和/或用于执行该期望的服务的服务模块。选择模块353可以被配置为基于流表(例如,流表354)选择用于执行期望的服务的服务模块。服务执行模块356可以被配置为在边缘设备300处对数据单元执行期望的服务。下文关于图1描述与过滤器模块355、选择模块353和服务执行模块356关联的功能的细节。

[0036] 边缘设备300可以例如包括硬件模块与(在硬件中存储和/或执行的)软件模块的组合。在一些实施例中,例如边缘设备300可以包括现场可编程门阵列(FPGA)、专用集成电路(ASIC)、数字信号处理器(DSP)等。在一些实施例中,通过执行在边缘设备300中包括的模块(例如,过滤器模块355、选择模块353、服务执行模块356)并且使用流表354,边缘设备300可以被配置为通过交换结构路由数据业务,从而使得能够在向目的地(例如,目的地外围处理设备)发送数据之前对该数据执行一个或者多个期望的服务或者在边缘设备300处本地对该数据执行期望的服务。具体而言,边缘设备300可以被配置为识别将对在边缘设备300处接收的数据执行的一个或者多个服务、选择与识别的服务关联的一个或者多个服务模块并且向选择的服务模块发送数据,从而使得选择的服务模块对数据执行服务和/或在边缘设备300处对数据执行识别的服务。关于图1和图8进一步描述了这样的过程的细节。

[0037] 在一些实施例中,边缘设备可以操作地耦合到或者包括一个或者多个服务模块。图5是根据一个实施例的操作地耦合到服务模块561-567的机架的边缘设备550的示意图示。边缘设备550可以在结构上和功能上与关于图1示出和描述的边缘设备181-183相似。具体而言,边缘设备550可以是交换结构系统(与图1中的交换结构系统100相似)内的边缘设备。边缘设备550可以将外围处理设备(在图5中未示出,但是与图1中的外围处理设备111-116相似)集合操作地耦合到交换结构(在图5中未示出,但是与图1中的交换结构110相似)。在一些实施例中,例如边缘设备550可以是与服务模块561-567的机架互连的TOR网络设备、诸如网络交换机。

[0038] 服务模块561-567可以是被配置为提供一个或者多个功能和/或执行一个或者多个服务的任何模块、部件或者设备。例如,每个服务模块561-567可以被配置为提供与经由边缘设备550和交换结构向和/或从其它设备发送数据和/或接收数据关联的功能,其中边缘设备550可以是源边缘设备和/或目的地边缘设备。例如,服务模块561可以是被配置为托管网站并且经由边缘设备550和交换结构从用户操作的输入设备(例如,个人计算机、智能电话)接收数据的web服务器。再例如,服务模块562可以是被配置为执行应用并且经由边缘

设备550和交换结构向与用户关联的显示设备发送数据的应用服务器。

[0039] 来自服务模块561-567的机架的一个或者多个服务模块可以被配置为对跨交换结构传输或者将跨交换结构传输的数据执行服务。例如,服务模块563可以被配置为对向web服务器传输或者将向web服务器传输的数据执行安全服务(例如,防火墙)。再例如,服务模块564可以被配置为对从应用服务器生成的数据执行负载平衡服务(例如,通过多个数据路径分发数据业务的服务)。来自服务模块561-567的机架的一个或者多个服务模块可以是虚拟服务模块。这样的虚拟服务模块可以由管理模块(与图1中的管理模块150和图4中的管理模块400相似)管理(例如,实例化、配置、监视、维护、移动、去实例化等),该管理模块被配置为管理交换结构系统的边缘设备(包括边缘设备550)和虚拟服务模块。下文关于图4进一步描述了管理虚拟服务模块的细节。

[0040] 可以例如在来自操作地耦合到边缘设备550的外围处理设备集合的外围处理设备和/或操作地耦合到边缘设备550的任何其它类型的服务器设备处托管和执行一个或者多个服务模块561-567。在这样的实例中,一个或者多个服务模块561-567操作地耦合到如图5中所示的边缘设备550。托管服务模块的外围处理设备可以称为服务外围处理设备。类似地,虽然在图5中未示出,但是可以在边缘设备550内托管和执行一个或者多个服务模块561-567。在这样的实例中,可以在边缘设备550的存储器中存储并且在边缘设备550的处理器处执行一个或者多个服务模块561-567。托管服务模块的边缘设备可以称为服务边缘设备。

[0041] 返回图1,交换结构110可以是经由一个或者多个数据路径将边缘设备181-183操作地耦合到其它边缘设备181-183的任何适当交换结构。例如,交换结构110可以是具有多级交换模块(例如,集成以太网交换机)的Clos网络(例如,非阻塞Clos网络、严格意义上的非阻塞Clos网络、Benes网络)。在一些实施例中,例如交换结构110可以与(如下文关于图2示出和描述的)具有三级的交换结构200相似。在其它实施例中,图1中所示的交换结构110可以包括任何数目的级。在这样的实施例中,例如交换结构110可以包括五级、七级或者九级。

[0042] 在一些实施例中,交换结构110可以是(例如,可以充当)单个联合交换机(例如,单个大规模联合L2/L3交换机)。换言之,交换结构110可以被配置为作为单个逻辑实体(例如,单个逻辑网元)来操作。在这样的实施例中,交换结构110可以是在第一边缘设备181、182或者183与第二边缘设备181、182或者183之间的单个逻辑跳跃的一部分(例如,与在边缘设备181-183与交换结构110之间的数据路径一起)。交换结构110可以被配置为连接外围处理设备111-116(例如,操作地耦合外围处理设备111-116、促进在外围处理设备111-116之间的通信)。在一些实施例中,交换结构110可以被配置为经由如下接口设备(在图1中未示出)通信,这些接口设备被配置为在至少10Gb/s的速率下传输数据。在一些实施例中,交换结构110可以被配置为经由如下接口设备(例如,光纤信道接口设备)通信,这些接口设备被配置为在例如2Gb/s、4Gb/s、8Gb/s、10Gb/s、40Gb/s、100Gb/s和/或更快链路速度的速率下传输数据。

[0043] 虽然可以在逻辑上集中交换结构110,但是例如出于可靠性考虑,交换结构110的实现方式可以是高度分布式的。例如,交换结构110的各部分可以在物理上跨例如许多底盘(或者机架)而分布。在一些实施例中,例如可以在第一底盘中包括交换结构110的一级,并

且可以在第二底盘中包括交换结构110的另一级。这两级都可以在逻辑上充当单个联合式交换机(例如,在相同逻辑跳跃内)的一部分,但是在相应成对级之间具有分离的单个物理跳跃。

[0044] 图2是根据一个实施例的多级交换结构200的示意图示。交换结构200可以在结构上和功能上与关于图1示出和描述的交换结构110相似。交换结构200可以包括在单个逻辑跳跃内的多个物理跳跃。在一些实施例中,交换结构200可以是包括第一级240、第二级242和第三级244的多级非阻塞Clos网络。第一级240包括交换模块212;第二级242包括交换模块214;第三级244包括交换模块216。换言之,第一级240的交换模块212、第二级242的交换模块214和第三级244的交换模块216共同定义多级交换结构200。

[0045] 在一些实施例中,第一级240的每个交换模块212可以是电子部件和电路的组件。在一些实施例中,例如每个交换模块212是ASIC。在其它实施例中,在单个ASIC或者单个芯片封装上包含多个交换模块。在更多其它实施例中,每个交换模块是分立电部件的组件。

[0046] 在一些实施例中,第一级240的每个交换模块212是交换机(例如,分组交换机、帧交换机、集成以太网交换机和/或单元交换机)。交换机被配置为在数据(例如,数据分组、数据单位等)流过交换结构200时将它重定向。在一些实施例中,例如每个交换机包括操作地耦合到存储器缓冲器(在图2中未示出)上的写入接口的多个进入端口。类似地,外出端口集合操作地耦合到存储器缓冲器上的读取接口。在一些实施例中,存储器缓冲器可以是使用片上静态随机存取存储器(SRAM)来实施的共享存储器缓冲器,该存储器缓冲器用于向所有进入端口和所有外出端口提供充足带宽以在每个时间段(例如,一个或者多个时钟周期)写入一个传入数据单位(例如,数据分组的一部分)或者数据分组,并且在每个时间段读取一个传出数据单位或者数据分组。每个交换机与可以在后续每个时间段中重新配置的纵横(crossbar)交换机相似地进行操作。

[0047] 第一级240的每个交换模块212包括被配置为在数据(例如,数据单位、数据分组)进入交换结构200时接收该数据的进入端口260的集合。例如,每个进入端口260可以耦合到边缘设备(例如,关于图1示出和描述的边缘设备181-183)。在一些实施例中,交换模块212的多于一个进入端口260可以经由分离的物理连接(例如,多个电线电缆、多个光纤线缆等)耦合到公共边缘设备的不同端口。因而,边缘设备可以经由进入端口260向交换结构200发送数据。在图2的实施例中,第一级240的每个交换模块212包括相同数目的进入端口260。在其它实施例中,每个交换模块212可以具有不同数目的进入端口260。

[0048] 与第一级240相似,交换结构200的第二级242包括交换模块241。第二级242的交换模块214在结构上与第一级240的交换模块212相似。在一些实施例中,第二级242的每个交换模块214通过数据路径220操作地耦合到第一级240的每个交换模块212。在第一级240的给定交换模块212与第二级242的给定交换模块214之间的每个数据路径220被配置为促进从第一级240的交换模块212向第二级242的交换模块214的数据传送。

[0049] 可以用被配置为促进从第一级240的交换模块212向第二级242的交换模块214的数据传送的任何方式来构造在第一级240的交换模块212与第二级242的交换模块214之间的数据路径220。在一些实施例中,数据路径220可以例如是在交换模块之间的光连接器。在其它实施例中,数据路径220可以在中间平面(midplane)内。这样的中平面可以用来连接第二级242的每个交换模块214与第一级240的每个交换模块212。在更多其它实施例中,可以

在单个芯片封装内包含两个或者更多交换模块212和214,并且数据路径220可以是电迹线。

[0050] 在一些实施例中,交换结构200可以是非阻塞Clos网络。因此,交换结构200的第二级242的交换模块214的数目基于第一级240的每个交换模块212的进入端口260的数目而变化。在可再布置的非阻塞Clos网络(例如Benes网络)中,第二级242的交换模块214的数目大于或者等于第一级240的每个交换模块212的进入端口260的数目。在一些实施例中,例如第一级240的每个交换模块212具有五个进入端口。因此,第二级242具有至少五个交换模块214。第一级240的交换模块212中的每个交换模块由数据路径220操作地耦合到第二级242的所有交换模块214。换言之,第一级240的每个交换模块212可以向第二级242的任何交换模块214发送数据。

[0051] 交换结构200的第三级244包括交换模块216。第三级244的交换模块216在结构上与第一级240的交换模块212相似。第三级244的交换模块216的数目通常与第一级240的交换模块212的数目相等。第三级244的每个交换模块216包括被配置为允许数据退出交换结构200的外出端口262。例如,每个外出端口262可以耦合到边缘设备(例如,关于图1示出和描述的边缘设备181-183)。在一些实施例中,与第一级240的交换模块212相似,交换模块216的多于一个外出端口262可以经由分离的物理连接(例如,多个电线电缆、多个光纤线缆等)耦合到公共边缘设备的不同端口。因而,边缘设备可以经由外出端口262从交换结构200接收数据。在图2的实施例中,第三级244的每个交换模块216包括相同数目的外出端口262。另外,第三级244的每个交换模块216的外出端口262的数目通常与第一级240的每个交换模块212的进入端口260的数目相等。在其它实施例中,每个交换模块216可以具有不同数目的外出端口262,并且第三级244的交换模块216的外出端口262的数目可以不同于第一级240的交换模块212的进入端口260的数目。

[0052] 在一些实施例中,第三级244的每个交换模块216可以通过数据路径224连接到第二级242的每个交换模块214。在第二级242的交换模块214与第三级244的交换模块216之间的数据路径224被配置为促进从第二级242的交换模块214向第三级244的交换模块216的数据传送。

[0053] 与数据路径220相似,可以用被配置为促进从第二级242的交换模块214向第三级244的交换模块216的数据传送的任何方式来构造在第二级242的交换模块214与第三级244的交换模块216之间的数据路径224。在一些实施例中,例如数据路径224可以例如是在交换模块之间的光连接器。在其它实施例中,数据路径224可以在中间平面内。这样的中间平面可以用来连接第二级242的每个交换模块214与第三级244的每个交换模块216。在更多其它实施例中,可以在单个芯片封装内包含两个或者更多交换模块214和216,并且数据路径224可以是电迹线。

[0054] 在一些实施例中,可以使用哈希函数、查找表、路由表等通过交换结构200路由数据。例如,第一级交换模块212可以通过使用数据单位的报头值作为对哈希函数的输入来确定向哪个第二级交换模块214发送数据单位。哈希函数的结果可以是第二级交换模块214的标识符,并且第一级交换模块212可以相应地发送数据单位。类似地,第二级交换模块214和/或第三级交换模块216可以分别使用这样的哈希函数、查找表等来确定分别向哪个第三级交换模块216或者哪个(耦合到外出端口262的)边缘设备发送数据单位。

[0055] 返回图1,管理模块150可以例如是在交换结构系统100内的设备处托管的(在硬件

中存储和/或执行的)过程、应用、虚拟机和/或某一其它软件模块或者硬件模块。在一些实施例中,可以例如在边缘设备(例如,边缘设备181-183)、在交换结构110中的切换设备或者在交换结构系统100中的任何其它适当设备处托管管理模块150。托管管理模块150的设备操作地耦合到交换结构系统100的其余设备,包括边缘设备181-183。例如,可以在边缘设备(例如,边缘设备181-183)内的存储器处存储并且在该边缘设备的处理器处执行实施管理模块150的指令。

[0056] 管理模块150可以被配置为管理交换结构系统100的一个或者多个边缘设备(例如,边缘设备181-183)。例如,管理模块150可以管理和/或维护与边缘设备181-183关联的配置信息(例如,端口协议信息、网络分段分配信息、端口分配信息、外围处理设备信息等)、转发状态信息(例如,端口标识符、网络分段标识符、外围处理设备标识符等)和/或其它信息。管理模块150也可以例如监视与边缘设备181-183关联的外围处理设备(例如,外围处理设备111-116)的状态和/或状况和/或管理和维护与关联于边缘设备181-183的外围处理设备和/或端口关联的其它信息。具体而言,管理模块150可以被配置为管理和维护与对跨交换结构系统100传输或者将跨交换结构系统100传输的数据执行服务关联的信息。下文关于图4中的管理模块400的虚拟机(VM)调配模块453和供给模块454讨论关于在管理模块处管理和维护信息的更多细节。

[0057] 图4是根据一个实施例的管理模块400的框图。管理模块400可以在结构上和功能上与关于图1示出和描述的管理模块150相似。具体而言,可以在与图1中的交换结构系统100相似的交换结构系统的设备(例如,边缘设备、切换设备)处托管管理模块400。管理模块400可以操作地耦合到交换结构系统的其余设备(例如,边缘设备)。在一些实施例中,管理模块400可以被配置为管理交换结构系统中的边缘设备集合的操作。

[0058] 如图4中所示,管理模块400包括存储器450,该存储器包括VM调配模块453和供给模块454;处理器410;以及通信接口420。存储器450可以例如是RAM(例如,动态RAM、静态RAM)、闪存和/或可移除存储器。在存储器450中包括的模块(包括在图4中未示出的可能模块)中的每个模块可以例如是在管理模块400处执行的(在硬件中存储和/或执行的)数据库、过程、应用、虚拟机和/或某一其它软件模块或者硬件模块。例如,可以在存储器450中存储并且由处理器410执行实施这些模块的指令。处理器410可以是例如被配置为向存储器450中写入数据和从存储器450读取数据并且执行在存储器450内存储的指令的任何处理器。处理器410也可以被配置为例如控制VM调配模块453、供给模块454和通信接口420的操作。另外,通过执行在存储器450中存储的指令,管理模块400可以被配置为经由通信接口420与管理模块400所管理的设备(例如,边缘设备)通信(例如向和/或从这些设备发送信号和/或接收信号)。

[0059] 在处理器410的控制之下并且基于在存储器450中存储的指令,管理模块400也可以例如管理交换结构系统中的边缘设备集合的操作。具体而言,管理模块400的VM调配模块453、供给模块454和/或其它模块(在图4中未示出)可以被配置为共同管理与对跨交换结构系统传输或者将跨交换结构系统传输的数据执行服务关联的边缘设备的操作。

[0060] VM调配模块453可以被配置为管理(例如,实例化、配置、监视、维护、移动、去实例化等)在交换结构系统中的设备(例如,服务边缘设备、服务外围处理设备)处执行的虚拟服务模块。可以执行VM调配模块453所管理的虚拟服务模块以对跨交换结构系统传输或者将

跨交换结构系统传输的数据执行各种服务(例如,安全服务、负载平衡服务)。例如,VM调配模块453可以被配置为向边缘设备发送信号(例如,经由通信接口420),从而使得在该边缘设备处实例化用于执行防火墙服务的虚拟服务模块。再例如,VM调配模块453可以被配置为监视在服务外围处理设备处执行的虚拟服务模块的状况,从而使得VM调配模块453可以从该服务外围处理设备接收如下指示(例如,经由通信接口420),该指示指明该虚拟服务模块的状况的改变(例如,虚拟服务模块过加载)。

[0061] 供给模块454可以被配置为维护与可用于对跨交换结构系统传输或者将跨交换结构系统传输的数据执行服务的服务模块和/或对该数据活动地执行服务的服务模块关联的信息。例如,供给模块454可以被配置为维护如下数据结构(例如,列表、表),该数据结构存储可用服务的信息和执行服务的关联服务模块的位置。在一些实例中,供给模块454可以被配置为向每个边缘设备发送数据结构的副本或者数据结构的一部分,从而使得边缘设备可以基于数据结构选择适当服务模块。另外,在改变(例如,激活、去激活、过加载等)服务模块的状况之后,供给模块454可以被配置为接收该改变的指示、相应地更新该服务模块的信息、然后向每个边缘设备发送更新的信息(例如,数据结构或者数据结构的一部分的更新版本)。因此,可以用交换结构系统中的任何服务模块的任何改变来更新每个边缘设备。

[0062] 在一些实施例中,如下文进一步描述的那样,管理模块400管理的边缘设备可以被配置为响应于从执行服务的服务模块接收到指示而开始执行该服务。在这样的实施例中,边缘设备可以向管理模块400发送指示服务在边缘设备处可用的信号。响应于该信号,供给模块454可以被配置为在数据结构内存储与执行服务的边缘设备关联的信息。在一些实施例中,供给模块454可以被配置为向交换结构系统的其它设备(例如,边缘设备、外围处理设备)发送与边缘设备关联的更新的信息(例如,数据结构或者数据结构的一部分的更新版本)。

[0063] 返回图1,交换结构系统100可以被配置为经由耦合到源外围处理设备的边缘设备(例如,源边缘设备)、交换结构110、耦合到目的地外围处理设备的边缘设备(例如,目的地边缘设备)和/或对从源外围处理设备生成的数据执行服务的一个或者多个其它设备(例如,服务边缘设备、服务外围处理设备)向目的地外围处理设备转发从源外围处理设备生成的数据单元(例如,数据分组、数据单位、数据分组的一部分、数据单位的一部分、数据分组的报头部分、数据分组的净荷部分等)。因此,交换结构系统100可以被配置为在向目的地外围处理设备发送数据之前对该数据执行一个或者多个服务。

[0064] 例如,如图1中所示,交换结构系统100可以被配置为经由边缘设备181(源边缘设备和在一些场景中的服务边缘设备)、交换结构110、边缘设备182(目的地边缘设备)和/或边缘设备183和外围处理设备115(在一些场景中充当服务边缘设备或者服务外围处理设备)向外围处理设备111(目的地外围处理设备)转发从外围处理设备113(源外围处理设备)生成的数据。因此,可以在向外围处理设备111发送数据之前对该数据执行一个或者多个服务。在一些场景中,例如如图1中所示,从外围处理设备113经由数据路径172和数据路径174向外围处理设备111发送数据。

[0065] 具体而言,边缘设备181可以被配置为从外围处理设备113经由数据路径192例如接收数据单元(例如,数据分组、数据单位、数据分组的一部分、数据单位的一部分、数据分组的报头部分、数据分组的净荷部分等)的数据流。该数据流可以例如与在外围处理设备

113处执行的第一服务器(例如,web服务器)和在外围处理设备111处执行的第二服务器(例如,应用服务器)关联。作为结果,数据流可以在外围处理设备113的第一服务器处被生成并且去往在外围处理设备111处的第二服务器。

[0066] 数据流的每个数据单元包括与该数据流关联的至少一个特性。这样的特性可以例如是源外围处理设备113的标识符(例如,外围处理设备113的因特网协议(IP)地址或者媒体接入控制(MAC)地址)、目的地外围处理设备111的标识符(例如,外围处理设备111的IP地址或者MAC地址)、服务端口(例如,与第一服务器关联的外围处理设备113的端口)的标识符、目的地端口(例如,与第二服务器关联的外围处理设备111的端口)的标识符、与传输协议关联的信息、在数据流中包括的数据单元的类型(例如,数据分组、数据单位)、任何其它适当特性或者上述特性中的任何特性的组合。可以例如在每个数据单元的报头部分中包括该数据单元的特性的指示。

[0067] 响应于从数据流接收到数据单元,边缘设备181可以被配置为基于在数据单元中包括的特性和/或其它信息来识别将对数据单元执行的一个或者多个期望的服务和用于执行期望的服务的一个或者多个服务模块。在一些实施例中,边缘设备181可以基于流表(例如,图3中的边缘设备300的流表354)来识别期望的服务和用于执行期望的服务的服务模块。这样的流表存储和维护数据单元的特性与服务模块的关联和/或其它有关信息。

[0068] 图6是根据一个实施例的在源边缘设备处实施的流表600的示意图示。源边缘设备可以与图1中的边缘设备181-183和图3中的边缘设备300相似。与在图3中的边缘设备300的存储器350中实施的流表354相似,可以在源边缘设备的存储器中实施流表600。源边缘设备可以被包括在交换结构系统(例如,图1中的交换结构系统100)中,并因此操作地耦合到与交换结构系统关联的其它设备(例如,其它边缘设备、外围处理设备)。如图6中所示,流表600包括三列:服务模块标识符(ID)610、计数器620和特性标识符(ID)630。如这里进一步具体描述的那样,可以自动(例如,通过传输数据单元来触发)或者人工(例如,由交换结构系统的运营商或者管理员)来改变(例如,插入、修改、更新、去除等)在流表600中存储的值和/或关联。

[0069] 第一列,即服务模块ID610存储对数据单元执行一个或者多个服务的服务模块的标识符(例如,1、2、7、SM_N)。服务模块的每个标识符唯一地标识该服务模块。例如,服务模块标识符1标识在边缘设备处托管和执行的并且对数据单元执行防火墙服务的服务模块;服务模块标识符2标识在外围处理设备处托管和执行的并且对数据单元执行负载平衡服务的服务模块。

[0070] 第二列、即计数器620的每个条目存储如下值(例如,25、72、5、C_M),该值与在流表600中的具有该值的相同条目(或者行)中存储的服务模块标识符标识的服务模块的使用相关联。在一些实施例中,该值可以例如代表在特定时间段期间已经向用于服务的服务模块发送的数据单元的累计数目。特定时间段可以是可由例如交换结构系统的运营商或者管理员预定和/或配置的系统参数。例如,与服务模块标识符1关联的值25指示在最后10毫秒内已经从源边缘设备向服务模块标识符1标识的服务模块(例如,用于执行防火墙服务的服务模块)发送了25个数据单元。再例如,与服务模块标识符2关联的值72指示在最后10毫秒内已经从源边缘设备向服务模块标识符2标识的服务模块(例如,用于执行负载平衡服务的服务模块)发送了72个数据单元。

[0071] 在一些其它实施例中,该值可以例如是与从源边缘设备向服务模块标识符标识的服务模块传输一个或者多个数据流的数据单元关联的速率。例如,值5/毫秒指示源边缘设备已经在以每毫秒5个数据单元的速率向服务模块标识符7标识的服务模块发送数据单元。另外,在一些实施例中,可以定期地计算并且相应地更新速率。

[0072] 在一些实施例中,可以用任何适当方法来更新在计数器620中存储的值。例如,可以在源边缘设备在预定时间段中未向服务模块发送任何数据单元之后递增与该服务模块关联的累计数目。再例如,与服务模块关联的速率可以由交换结构系统的运营商或者管理员人工改变成任何任意值(例如,0)。在一些实施例中,可以响应于源边缘设备在预定时间段内未从数据流接收任何数据单元而从流表600移除流表600中的与该数据流关联的整个条目(包括服务模块标识符、计数器值和特性标识符)。

[0073] 第三列、即特性ID630的每个条目存储在数据流的每个数据单元中包括的特性的标识符(例如,10、37、5、21、17、54、CID_r)。每个特性标识符唯一地标识该特性。例如,特性标识符10标识用于数据流的源外围处理设备的标识符(例如,地址);特性标识符37标识用于数据流的目的地外围处理设备的标识符(例如,地址);特性标识符54标识用于传输数据流的数据单元的传输协议;特性标识符17标识数据流的数据单元的类型。

[0074] 根据流表600,向在与特性ID630的条目对应的、服务模块ID的条目中存储的服务模块标识符标识的服务模块发送如下数据单元,这些数据单元包括在特性ID630的该条目中存储的特性标识符标识的特性。也通过更新在与特性ID630的条目和服务模块ID610的条目关联的、计数器620的条目中存储的值来记录这样的数据单元的发送。例如,向(服务模块ID610)中的服务模块标识符1标识的服务模块发送来自如下数据流的数据单元,该数据流包括特性标识符10标识的特性(例如,公共源外围处理设备),并且在向该服务模块发送来自数据流的每个数据单元之后递增在计数器620的关联条目中存储的累计数目(在图6中示出为25)。再例如,向(服务模块ID610中的)服务模块标识符7标识的服务模块发送来自如下数据流的数据单元,该数据流包括特性标识符54标识的特性(例如,特定传输协议),并且(例如,在向该服务模块发送来自数据流的每个数据单元之后、在每个特定时间段之后等)相应地更新在计数器620的关联条目中存储的速率(例如,在图6中示出为5/毫秒)。

[0075] 虽然在流表600中示出为每个服务模块与一个计数器关联(例如,累计数目、速率),但是在一些其它实施例中,每个服务模块可以与多于一个计数器相关联。例如,服务模块可以与记录从源边缘设备向该服务模块发送的数据单元的累计数目的第一计数器和记录与从源边缘设备向该服务模块传输数据单元关联的速率的第二计数器关联。再例如,服务模块可以与多个计数器关联,每个计数器记录与从源边缘设备向该服务模块发送的不同数据流的数据单元关联的累计数目或者速率。虽然在图6中未示出,但是在一些实施例中,(服务模块ID610中的)服务模块标识符2标识的服务模块可以与三个计数器关联,每个计数器记录与三个数据流中的包括特性标识符5、21或者17标识的特性的数据流之一关联的累计数目或者速率。

[0076] 虽然在图6中未示出,但是在一些实施例中,流表可以包括与将对数据单元执行的服务关联的信息和/或与服务模块关联的其它信息。例如,流表可以包括服务ID列,该列存储将对跨交换结构系统传输或者将跨交换结构系统传输的数据单元执行的服务的标识符。这样的流表可以关联(通过特性标识符标识的)数据单元、将对这些数据单元执行的(通过

服务标识符标识的)服务、执行这些服务的(通过服务模块标识符标识的)服务模块和用于这些服务模块的计数器。

[0077] 返回图1,响应于从数据流接收到数据单元,边缘设备181可以检查流表(与上文关于图6描述的流表600相似)以确定该数据单元的信息(或者在一些实施例中等效地为该数据流的信息)是否存储于流表中。具体而言,边缘设备181可以被配置为比较数据单元中包括的特性与在流表中存储的特性。如果数据单元中包括的一个或者多个特性被识别为存储于流表中(例如,由在流表中存储的一个或者多个特性标识符标识),则先前已经在源边缘设备处接收并且由源边缘设备处理了该数据流的至少一个数据单元并且在流表中存储与对数据流的数据单元执行服务关联的信息。作为结果,可以识别一个或者多个服务模块以对数据单元执行期望的服务。在一些实施例中,例如边缘设备181的选择模块(与图3中的边缘设备300的选择模块353相似)可以被配置为基于流表中的包括数据单元的一个或者多个条目来识别用于对该数据单元执行期望的服务的服务模块。在这样的实施例中,边缘设备181无需识别将对数据单元执行的期望的服务和/或用于执行期望的服务的服务模块。

[0078] 例如,如图6中所示,响应于接收到如下数据单元,该数据单元包括特性标识符10标识的源外围处理设备的标识符,托管流表600的源边缘设备可以被配置为基于流表600来识别与特性标识符10关联的服务模块标识符。作为结果,可以将服务模块标识符1标识的服务模块作为用于对数据单元执行期望的服务的服务模块。因而,源边缘设备可以被配置为向服务模块标识符1标识的服务模块发送数据单元并且将在计数器620的关联条目中存储的累计数目(在图6中示出为25)递增1。

[0079] 再例如,如图6中所示,响应于基于特性标识符54标识的传输协议接收到数据单元,托管流表600的源边缘设备可以被配置为基于流表600来识别与特性标识符54关联的服务模块标识符7。作为结果,可以将服务模块标识符7标识的服务模块识别为用于对数据单元执行期望的服务的服务模块。因而,源边缘设备可以被配置为向服务模块标识符7标识的服务模块发送数据单元并且在适当时间更新在计数器620的关联条目中存储的速率(在图6中示出为5/毫秒)。

[0080] 否则,如果没有任何数据单元的特性被识别为存储于流表中(例如,未被流表中存储的任何特性标识符标识),则在流表中没有与对数据单元(或者在一些实施例中等效地为对来自数据流的数据单元)执行服务关联的信息被存储。在该情况下,边缘设备181可以被配置为使用另一方法、诸如哈希函数来识别期望的服务和用于执行期望的服务的服务模块。具体而言,过滤器模块(与图3中的边缘设备300的过滤器模块355相似)和选择模块(与图3中的边缘设备300的选择模块353相似)可以被配置为基于哈希函数来共同识别期望的服务和用于执行期望的服务的服务模块。例如,在数据单元中包括的一个或者多个特性(例如,源外围处理设备113的标识符、目的地外围处理设备111的标识符、数据单元的类型、用来传输数据单元的传输协议等)可以用作对哈希函数的输入,并且哈希函数的结果可以用来识别期望的服务和/或用于执行期望的服务的服务模块。在具有代理案号JUNI-189/00US108200-2425、提交于相同日期并且标题为“Methods and Apparatus for Providing Services in a Distributed Switch”的共同未决美国专利申请中描述了使用哈希函数来识别期望的服务和用于执行期望的服务的服务模块的细节。

[0081] 响应于在边缘设备181处识别了将对数据单元执行的期望的服务和用于执行期望的服务的服务模块,边缘设备181可以被配置为在流表中存储与识别的服务模块和/或识别的服务关联的信息。例如,边缘设备181可以被配置为在流表中定义如下条目,该条目包括标识用于对数据单元执行期望的服务的服务模块的服务模块标识符、标识在数据单元中包括的特性的特性标识符、与识别的服务模块关联的计数器值(例如,累计数目、速率)、标识期望的服务的服务标识符和/或服务模块关联的任何其它适当信息。具体而言,在流表的条目中存储的特性标识符与用来识别期望的服务和/或用于对数据单元执行期望的服务的服务模块的特性关联。例如,如果哈希函数用来识别期望的服务和/或用于针对数据单元执行期望的服务的服务模块,则在流表的条目中存储的特性标识符与用作对哈希函数的输入的特性关联。因此,可以基于流表的新定义的条目向与数据流的第一数据单元相同的服务模块发送数据单元的后续数据单元。

[0082] 在图1的示例中,边缘设备181从源外围处理设备113接收数据流的第一数据单元。响应于在流表内未识别在第一数据单元中包括的特性,边缘设备181被配置为通过使用哈希函数来识别在外围处理设备115处托管的服务模块,以对第一数据单元执行期望的服务。作为结果,边缘设备181被配置为经由数据路径172向外围处理设备115发送第一数据单元。服务模块在外围处理设备115处对第一数据单元执行期望的服务、然后经由数据路径174向目的地外围处理设备111发送第一数据单元。同时,边缘设备181被配置为在流表中定义条目以存储与对数据流的数据单元执行期望的服务关联的信息,该信息包括在外围处理设备115处托管的服务模块的标识符、在数据流的数据单元(包括第一数据单元)中包括的特性的标识符、计数器值等。另外,响应于向服务模块发送数据流的第一数据单元和每个后续数据单元,边缘设备181被配置为相应地更新与服务模块关联的计数器值(例如,递增累计数目、重新计算速率)。

[0083] 在一些实施例中,在发送第一数据单元之后的某一时间,边缘设备181被配置为从源外围处理设备接收数据流的第二数据单元。第二数据单元包括如上文描述的那样在流表的条目中存储的与第一数据单元相同的特性。在一些实例中,可以从与用于第一数据单元的源外围处理设备相同的源外围处理设备(即,外围处理设备113)接收第二数据单元。在这样的实例中,第二数据单元可以来自包括第一数据单元的相同数据流。在其它实例中,可以从与用于第一数据单元的源外围处理设备不同的源外围处理设备(即,外围处理设备114)接收第二数据单元。在这样的实例中,第二数据单元来自与包括第一数据单元的数据流不同的数据流。

[0084] 基于流表的条目,边缘设备181被配置为识别在外围处理设备115处托管的、可以对第二数据单元执行期望的服务的服务模块。在一些实施例中,边缘设备181可以被配置为基于在流表的条目中包括的信息来确定是否向识别的服务模块发送第二数据单元。具体而言,边缘设备181的选择模块(与图3中的边缘设备300的选择模块353相似)可以被配置为比较在条目中存储的计数器值与预定阈值。这样的预定阈值代表识别的服务模块可以提供用于对数据单元执行期望的服务的最大容量。在一些实施例中,可以基于与源边缘设备、服务模块、连接源设备与服务模块的链路和/或任何其它有关信息关联的信息来确定该预定阈值。预定阈值可以例如由交换结构系统的运营商或者管理员配置、定义或者设定。

[0085] 如果在条目中存储的计数器值少于预定阈值,则将在识别的服务模块处被服务的

数据单元尚未使识别的服务模块过载。例如,记录在特定时间段期间向服务模块发送的数据单元的累计数目的计数器值是比与该服务模块关联的阈值30更少的25。作为结果,边缘设备181可以被配置为经由数据路径172向外围处理设备115发送第二数据单元,该外围处理设备对第二数据单元执行期望的服务并且向与第二数据单元关联的目的地外围处理设备(例如,外围处理设备111)发送第二数据单元。

[0086] 否则,如果在条目中存储的计数器值等于或者大于预定阈值,则需要在识别的服务模块处被服务的数据单元已经使识别的服务模块过载。例如,记录与向服务模块传输数据单元关联的速率的计数器值是比与服务模块关联的阈值4/毫秒更大的5/毫秒。作为结果,边缘设备181的选择模块可以被配置为识别或者选择与用来对第一数据单元执行期望的服务的服务模块不同的第二服务模块,以对第二数据单元执行期望的服务。可以使用任何适当方法来识别或者选择第二服务模块。在一些实施例中,例如边缘设备181可以被配置为使用如上文描述的哈希函数来识别第二服务模块。在一些其它实施例中,例如边缘设备181可以被配置为从如下服务模块集合中选择第二服务模块,在边缘设备181中(例如,在边缘设备181内的流表或者另一存储器位置中)存储该服务模块集合的信息。因此,边缘设备181的选择模块可以被配置为基于流表中的条目在可以对数据单元执行期望的服务的服务模块集合(包括对第一数据单元执行服务的服务模块和对第二数据单元执行服务的第二服务模块)之中对数据单元(包括第一数据单元和第二数据单元)进行负载平衡。

[0087] 例如,边缘设备181可以被配置为选择在外围处理设备116处托管的服务模块以对第二数据单元执行期望的服务。因而,边缘设备181可以被配置为向外围处理设备116发送第二数据单元,该外围处理设备对第二数据单元执行期望的服务、然后向与第二数据单元关联的目的地外围处理设备(例如,外围处理设备111)发送第二数据单元。

[0088] 与上文描述的第一数据单元的场景相似,响应于向第二服务模块发送第二数据单元,边缘设备181被配置为在流表中定义与第二服务模块关联的条目(如果在流表中尚未存储该条目)或者相应地更新与第二服务模块关联的条目(如果在流表中已经存储了该条目)。与第二服务模块关联的条目包括与在第二数据单元中包括的特性关联的特性标识符。

[0089] 在一些实施例中,在发送第二数据单元之后的某一时间,边缘设备181被配置为从源外围处理设备(例如,外围处理设备113、外围处理设备114)接收第三数据单元。第三数据单元包括与第二数据单元相同的特性。与这里描述的方法相似,边缘设备181可以被配置为基于流表中的与第二服务模块关联的条目来识别用于第三数据单元的第二服务模块。具体而言,可以在该条目中识别在第三数据单元中包括的特性。作为结果,可以识别第二服务模块以对第三数据单元执行期望的服务。

[0090] 在一些实施例中,在边缘设备181向在外围处理设备115处的服务模块发送第一数据单元、从而使得在服务模块处执行期望的服务(例如,防火墙服务、负载平衡服务)之后,边缘设备181可以被配置为在边缘设备181处对适当数据单元执行期望的服务。具体而言,外围处理设备115可以被配置为向边缘设备181发送与期望的服务关联的指示。这样的指示可以例如包括与执行期望的服务关联的命令或者指令。响应于该指示,边缘设备181可以被配置为例如修改或配置边缘设备181的服务执行模块(例如,图3中的边缘设备300的服务执行模块356),从而使得服务执行模块可以在边缘设备181处对数据单元执行期望的服务。同时,边缘设备181可以被配置为在边缘设备181的流表中存储在期望的服务与在第一数据单

元中包括的特性之间的关联。在一些实施例中,这样的流表不同于如关于图6示出和描述的存储在服务模块、在数据单元中包括的特性和计数器值之间的关联的流表。

[0091] 图7是根据一个实施例的在源边缘设备处实施的流表700的示意图示,该流表用来存储在源边缘设备处执行的服务与在数据单元中包括的特性之间的关联。源边缘设备可以与图1中的边缘设备181-183和图3中的边缘设备300相似。与在图3中的边缘设备300的存储器350中实施的流表354相似,可以在源边缘设备的存储器中实施流表700。源边缘设备可以包含于交换结构系统(例如,图1中的交换结构系统100)中,并因此操作地耦合到与交换结构系统关联的其它设备(例如,其它边缘设备、外围处理设备)。如图7中所示,流表700包括两列:服务标识符(ID)710和特性标识符(ID)720。

[0092] 第一列、即服务ID710存储可以在源边缘设备处对数据单元执行的服务的标识符(例如100、120、250、S_N)。服务的每个标识符唯一地标识该服务。例如,服务标识符100标识在源边缘设备(例如,由源边缘设备的服务执行模块)对数据单元执行的防火墙服务。再例如,服务标识符120标识在源边缘设备(例如,由源边缘设备的服务执行模块)对数据单元执行的负载平衡服务。

[0093] 第二列、即特性ID720的每个标识符存储在数据单元中包括的特性的标识符(例如,10、17、54、CID_N)。每个特性标识符唯一地标识该特性。例如,特性标识符10标识用于数据单元的源外围处理设备(或者目的地外围处理设备)的标识符(例如地址);特性标识符20标识用于传输数据单元的传输协议;特性标识符30标识数据单元的类型。

[0094] 根据流表700,源边缘设备可以被配置为对如下数据单元执行在服务ID710的条目中存储的服务标识符标识的服务,该数据单元包括在与服务ID710的该条目对应(即在相同行中)的、特性ID720的条目中存储的特性标识符标识的特性。例如,源边缘设备可以被配置为对从如下源外围处理设备接收的数据单元执行服务标识符100标识的防火墙服务,该源外围处理设备是特性标识符100标识的特性。再例如,源边缘设备可以被配置为对基于如下传输协议接收的数据单元执行服务标识符120标识的负载平衡服务,该传输协议是特性标识符20标识的特性。

[0095] 在一些实施例中,可以自动(例如,通过接收到指示来触发)或者人工(例如,由交换结构系统的运营商或者管理员)来改变(例如,插入、修改、更新、移除等)在流表700中存储的值和/或关联。例如,响应于接收到与防火墙服务关联的指示,源边缘设备可以被配置为在流表700中定义如下条目,该条目关联防火墙服务的服务标识符与如下特性标识符,该特性标识符与在针对防火墙服务而指定的数据单元中包括的特性关联。再例如,响应于源边缘设备未接收到具有该特性的任何数据单元,源边缘设备可以被配置为从流表700中移除如下条目,该条目关联特性的特性标识符与包括该特性的数据单元所期望的负载平衡服务的服务标识符。

[0096] 返回图1,响应于接收到与在外围处理设备115处对数据单元执行的期望的服务关联的指示,边缘设备181可以被配置为在流表中定义条目以存储在期望的服务与在第一数据单元中包括的特性之间的关联。同时,配置或者修改边缘设备181的服务执行模块,从而使得服务执行模块可以在边缘设备181处对数据单元执行期望的服务。

[0097] 例如,边缘设备181被配置为接收具有与第一数据单元相同的特性的第四数据单元。与第一数据单元相同,第四数据单元从外围处理设备113被发送并且去往外围处理设备

111。基于流表中的新定义的条目,边缘设备181的服务执行模块被配置为对第四数据单元执行期望的服务而无需向外围处理设备115或者任何其它服务模块发送第四数据单元。边缘设备181然后被配置为如图1中所示经由数据路径176向外围处理设备111发送第四数据单元。

[0098] 图8是根据一个实施例的方法800的流程图,该方法用于标识用于对跨分布式交换结构传输的数据执行服务的服务模块。可以在结构上和功能上分别与关于图1和图3示出和描述的边缘设备181-183和边缘设备300相似的边缘设备处执行方法800。具体而言,可以在边缘设备的存储器(例如,图3中的边缘设备300的存储器350)中的非瞬态处理器可读介质中存储与执行方法800关联的指令。非瞬态处理器可读介质存储如下代码,该代码代表将由边缘设备的处理器(例如,图3中的边缘设备300的处理器330)执行的指令(包括与执行方法800关联的指令)。另外,与图1中的边缘设备181-183相似,执行方法800的边缘设备可以与分布式交换结构(例如,图1中的交换结构110)关联并且操作地耦合到一个或者多个外围处理设备(例如,图1中的外围处理设备111-116)。

[0099] 方法800包括在802接收具有一个特性的第一数据单元。在一些实施例中,这样的特性可以例如包括源外围处理设备的标识符、目的地外围处理设备的标识符、第一数据单元的类型、与第一数据单元关联的协议、任何其它适当特性或者上述特性中的任何特性的组合。

[0100] 在804,边缘设备可以被配置为在第一时间响应于来自流表内的条目集合中的每个条目在第一时间不与特性关联而识别与该特性关联的服务模块的标识符。也就是说,边缘设备可以被配置为搜索流表并且基于搜索结果来确定特性未包含于流表的任何条目中或者未与流表的任何条目关联。流表可以与关于图6示出和描述的流表600相似。因此,边缘设备不能基于流表来识别用于对第一数据单元执行期望的服务(例如,防火墙服务、负载平衡服务)的服务模块。作为结果,边缘设备被配置为使用某一其它方法来识别用于对第一数据单元执行期望的服务的服务模块。例如,边缘设备可以被配置为基于哈希函数来识别服务模块,该哈希函数使用在第一数据单元中包括的特性和/或其它信息作为对哈希函数的输入。

[0101] 在806,边缘设备可以被配置为在第一时间之后的第二时间响应于识别了服务模块的标识符而在流表中定义与该特性和服务模块的标识符关联的条目。具体而言,边缘设备可以被配置为在该条目内存储特性的特性标识符和服务模块的标识符和/或服务模块关联的其它信息或者数据(例如,计数器值)。

[0102] 在808,边缘设备可以被配置为响应于识别了服务模块的标识符而向服务模块发送第一数据单元。作为结果,服务模块可以对第一数据单元执行期望的服务、然后向第一数据单元的目的地外围处理设备发送第一数据单元。

[0103] 在810,边缘设备可以被配置为在第二时间之后接收具有该特性的第二数据单元。在一些实施例中,第二数据单元可以来自包括第一数据单元的数据流。在其它实施例中,第二数据单元可以来自与包括第一数据单元的数据流不同的数据流。

[0104] 在812,边缘设备可以被配置为基于条目向服务模块发送第二数据单元。具体而言,边缘设备可以被配置为基于服务模块的标识符与该条目中的特性的特性标识符关联来识别该服务模块。作为结果,与第一数据单元的场景相似,边缘设备可以被配置为向服务模

块发送第二数据单元,从而使得服务模块可以对第二数据单元执行期望的服务、然后向第二数据单元的目的地外围处理设备发送第二数据单元。

[0105] 虽然上文关于图1-图8将一些实施例示出和描述为能够为包括跨交换结构传输或者将跨交换结构传输的数据识别服务并且选择服务模块的边缘设备,但是应当理解其它实施例是可能的。在一些实施例中,例如外围处理设备可以能够执行识别服务和选择服务模块的至少一部分。例如,外围处理设备可以被配置为存储和维护将期望的服务和用于执行期望的服务的服务模块与数据单元的特性关联的流表。作为结果,这样的外围处理设备可以被配置为识别服务和/或为识别的服务识别服务模块。在一些实施例中,例如能够识别服务和服务模块的外围处理设备可以由具有多个交换节点而无任何边缘设备的多路径网络互连。

[0106] 这里描述的一些实施例涉及一种具有非瞬态计算机可读介质(也可以被称为非瞬态处理器可读介质)的计算机存储产品,该非瞬态计算机可读介质上具有用于执行各种计算机实施的操作的指令或者计算机代码。计算机可读介质(或者处理器可读介质)在它本身未包括瞬态传播信号(例如,在传输介质、诸如空间或者线缆上传播承载信息的电磁波)的意义上为非瞬态。介质和计算机代码(也可以称为代码)可以是被设计和构造用于一个或者多个特定目的的介质和计算机代码。非瞬态计算机可读介质的示例包括但不限于:磁存储介质、诸如硬盘、软盘和磁带;光学存储介质、诸如紧致盘/数字视频盘(CD-DVD)、紧致盘-只读存储器(CD-ROM)和全息设备;光磁存储介质、诸如光盘;载波信号处理模块;以及具体配置为存储和执行程序代码的硬件设备、诸如专用集成电路(ASIC)、可编程逻辑器件(PLD)、只读存储器(ROM)和随机存取存储器(RAM)设备。这里描述的其它实施例涉及一种可以例如包括这里讨论的指令和/或计算机代码的计算机程序产品。

[0107] 计算机代码的示例包括但不限于微代码或者微指令、诸如编译器产生的机器指令、用来产生web服务的代码和如下文件,这些文件包含计算机使用解译器来执行的更高级指令。例如,实施例可以使用Java、C++或者其它编程语言(例如,面向对象的编程语言)和开发工具来实施。计算机代码的附加示例包括但不限于控制信号、加密的代码和压缩的代码。这样的计算机代码也可以被称为计算机程序,并且一些实施例可以具有计算机程序的形式。

[0108] 在上文已经描述了各种实施例,但是应当理解仅通过示例而非限制来呈现了它们并且可以进行在形式和细节上的各种改变。可以在除了互斥组合之外的任何组合中组合这里描述的装置和/或方法的任何部分。这里描述的实施例可以包括描述的不同实施例的功能、部件和/或特征的各种组合/或子组合。

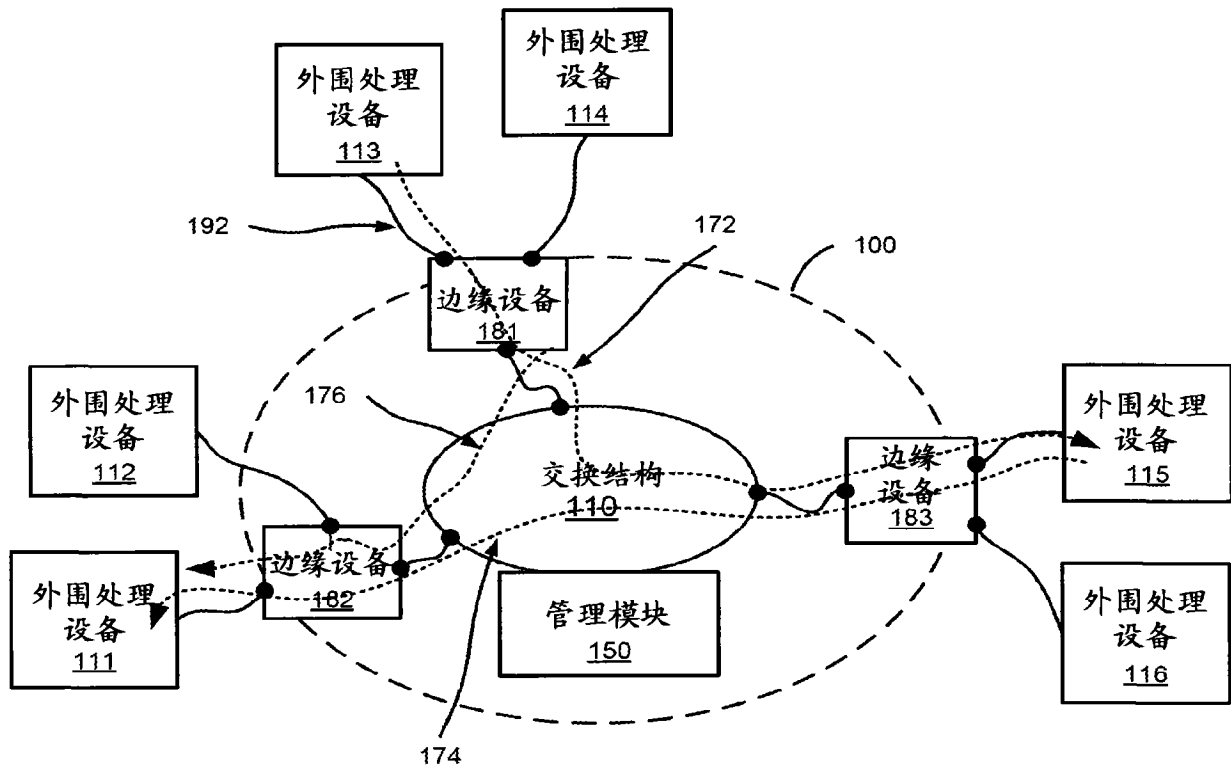


图1

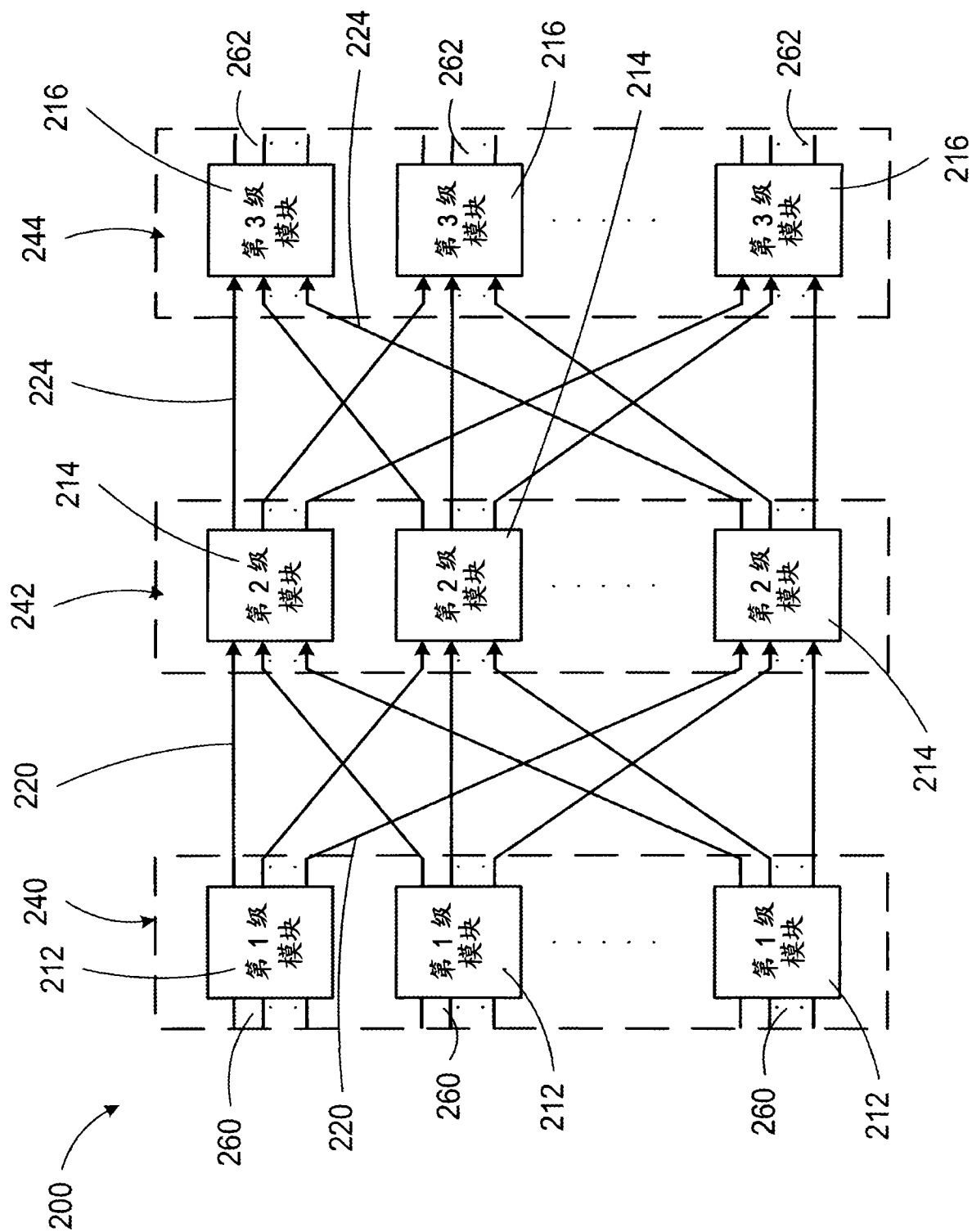


图2

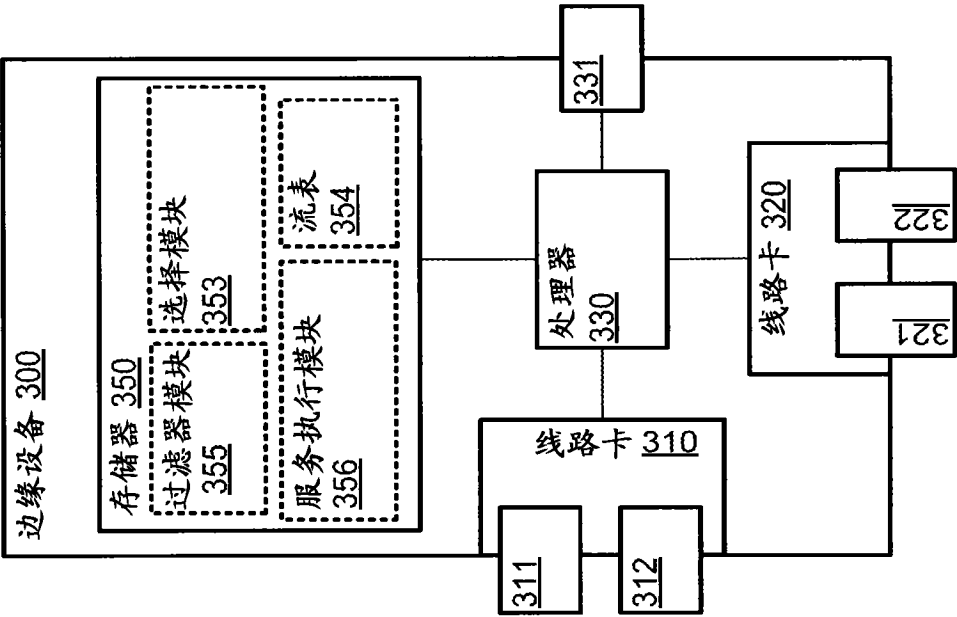


图3

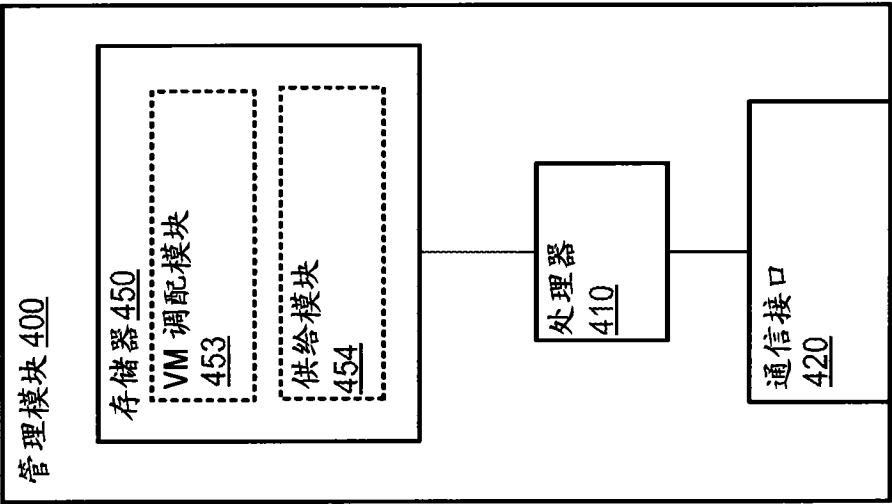


图4



图5

流表 600		
服务模块 ID 610	计数器 620	特性 ID 630
1	25	10
		37
2	72	5
		21
		17
7	5/ 毫秒	54
⋮		
SM _N	C _M	CID _T

图6

流表 700	
服务 ID 710	特性 ID 720
100	10
120	20
250	30
⋮	
S _N	CID _N

图7

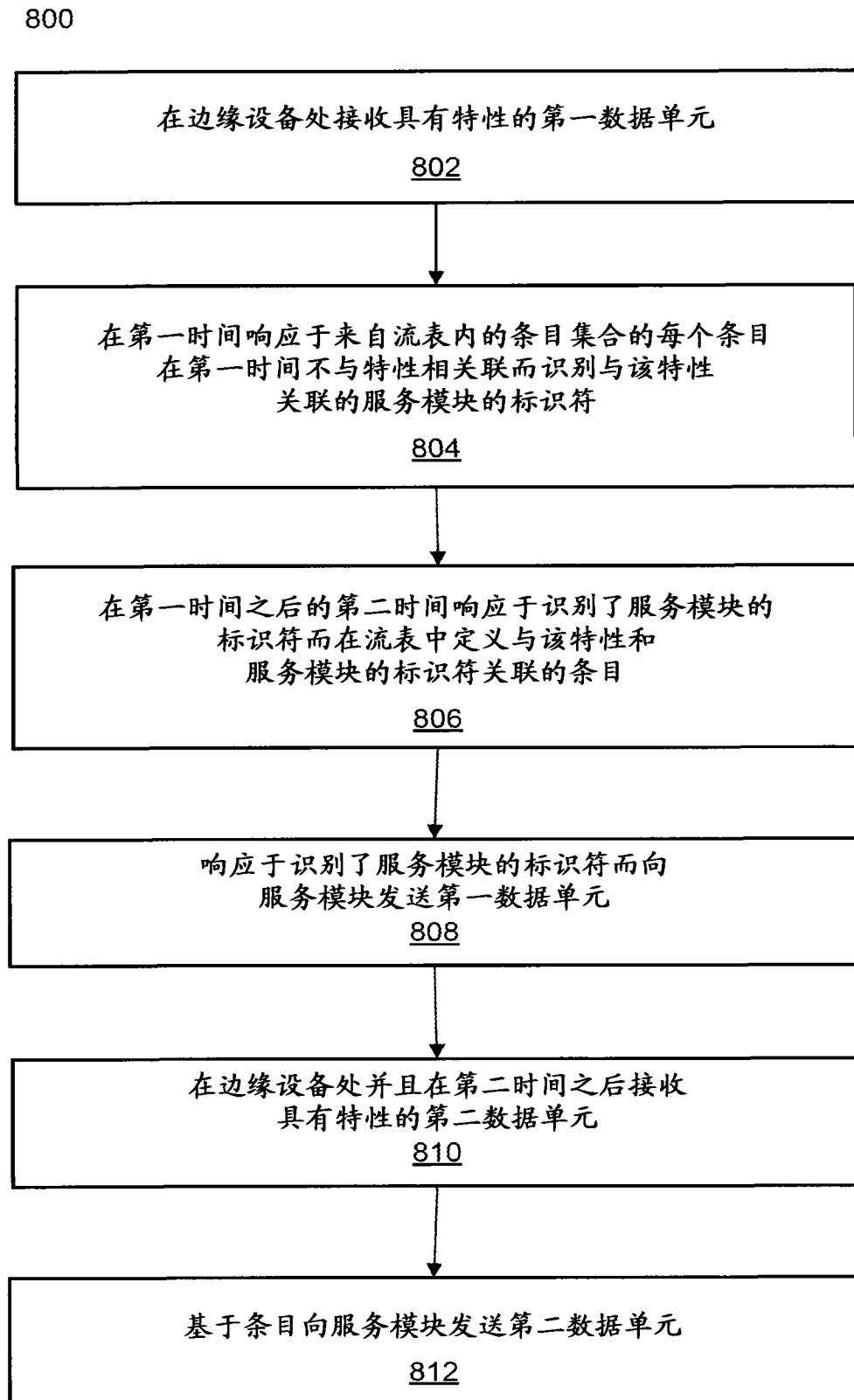


图8