



(12) 发明专利申请

(10) 申请公布号 CN 102349268 A

(43) 申请公布日 2012. 02. 08

(21) 申请号 201080011439. 4

(51) Int. Cl.

(22) 申请日 2010. 02. 23

H04L 12/56(2006. 01)

(30) 优先权数据

2009-055739 2009. 03. 09 JP

(85) PCT申请进入国家阶段日

2011. 09. 09

(86) PCT申请的申请数据

PCT/JP2010/052665 2010. 02. 23

(87) PCT申请的公布数据

W02010/103909 JA 2010. 09. 16

(71) 申请人 日本电气株式会社

地址 日本东京都

(72) 发明人 市野清久

(74) 专利代理机构 北京市金杜律师事务所

11256

代理人 吴立明 罗世娜

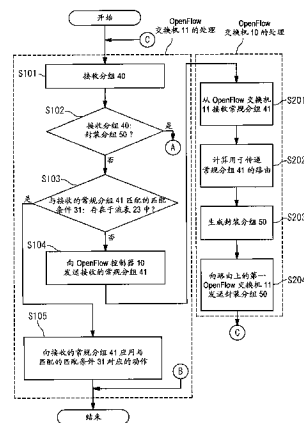
权利要求书 4 页 说明书 10 页 附图 15 页

(54) 发明名称

OpenFlow 通信系统和 OpenFlow 通信方法

(57) 摘要

OpenFlow 交换机根据流条目来控制分组的发送和接收。每个流条目包含表明分组的通信流的匹配条件并且也包括表明对分组的处理的动作。OpenFlow 控制器生成在位于通信流的路由上的具体 OpenFlow 交换机的流表中存储的登记流条目。通过将登记流条目与常规分组相关来生成封装分组。具体 OpenFlow 交换机响应于接收封装分组而从封装分组提取登记流条目，以产生新的流条目并且执行新的流条目中示出的动作。



1. 一种通信系统,包括:

至少一个交换机,其与网络连接以执行分组的传送;以及
控制器,被配置成控制所述交换机,

其中所述分组包括在所述网络中的终端之间发送和接收的常规分组以及从所述控制器供应的封装分组,

其中所述交换机包括:

流表,其可以保持多个流条目;以及

本地管理部,被配置成基于所述多个流条目中对应的一个流条目来控制每个所述分组的传送,

其中所述多个流条目中的每个流条目包括用于标识所述分组的通信流的匹配条件和示出所述分组的处理的动作,

其中所述控制器:

生成用于在存在于所述通信流的路由上的至少一个通道交换机的所述流表中存储的至少一个登记流条目,

将所述登记流条目和所述常规分组相关以生成所述封装分组,以及

向所述通道交换机发送所述封装分组,以及

其中所述通道交换机:

响应于接收所述封装分组而从所述封装分组提取所述登记流条目,以在所述通道交换机的所述流表中存储为新的流条目,以及

对所述封装分组执行所述新的流条目示出的动作。

2. 根据权利要求1所述的通信系统,其中所述交换机响应于接收所述常规分组而查阅所述流表,并且在包含与所述接收的常规分组的所述通信流匹配的所述匹配条件的所述流条目未登记于所述流表上时向所述控制器供应所述常规分组。

3. 根据权利要求2所述的通信系统,其中所述控制器:

响应于接收所述常规分组,指定传送所述常规分组时通过的在所述网络上的路由,

在基于所述通道交换机在所述路由上的位置而生成所述登记流条目之后,生成包含所述登记流条目的所述封装分组,以及

向所述常规分组已经被供应到的第一通道交换机供应所述封装分组。

4. 根据权利要求1至3中的任一权利要求所述的通信系统,其中所述通道交换机:

在多个所述登记流条目包含于所述封装分组中时,向所述路由上的下一个通道交换机发送所述封装分组,其中,已经从所述封装分组提取一个登记流条目,以及

在所述登记流条目包含于所述封装分组中时,向作为目的地的所述终端发送包含于所述封装分组中的所述常规分组。

5. 根据权利要求1至4中的任一权利要求所述的通信系统,其中所述通道交换机响应于接收所述常规分组来而查阅所述流表,并且在包含与所述接收的常规分组的所述通信流匹配的所述匹配条件的所述流条目已经登记于所述流表上时,基于所述登记流条目的所述动作对所述常规分组执行处理。

6. 一种通信方法,包括:

(a) 分组传送处理步骤:执行对在网络中的终端之间发送和接收的常规分组以及从控

制器供应的封装分组的传送处理；以及

(b) 封装分组供应步骤：供应所述封装分组，

其中所述分组传送处理步骤包括：

(a-1) 读取保持于交换机的流表中的流条目的步骤；

(a-2) 指定用于标识在所述读取的流条目中示出的所述常规分组或者所述封装分组的通信流的匹配条件和示出对所述常规分组或者所述封装分组的处理的动作的步骤；以及

(a-3) 基于所述流条目来控制所述常规分组或者所述封装分组的传送的步骤，

其中所述封装分组供应步骤包括：

(b-1) 生成一个或者多个登记流条目的步骤，所述一个或者多个登记流条目存储在所述通信流的所述路由上的一个或者多个通道交换机的流表中；

(b-2) 将一个或者多个所述登记流条目和所述常规分组相关以生成所述封装分组的步骤，以及

(b-3) 向所述通道交换机之一发送所述封装分组的步骤，

其中所述 (a-1) 步骤包括：

在所述通道交换机接收所述封装分组时从所述封装分组提取一个登记流条目的步骤；

以及

在所述通道交换机的所述流表中存储所提取的登记流条目作为新的流条目的步骤，

其中所述 (a-2) 步骤包括：

对所述封装分组执行所述新的流条目中示出的动作的步骤。

7. 根据权利要求 6 所述的通信方法，其中所述 (a-1) 步骤包括：

响应于接收所述常规分组而查阅所述流表的步骤；

在包含与所述接收的常规分组的所述通信流匹配的所述匹配条件的所述流条目未登记于所述流表上时向所述控制器供应所述常规分组的步骤。

8. 根据权利要求 7 所述的通信方法，其中所述 (b-1) 步骤包括：

由所述控制器响应于接收所述常规分组来而指定用于传送所述常规分组的在所述网络上的路由的步骤；并且以及

在基于所述通道交换机在所述指定的路由上的位置而生成一个或者多个所述登记流条目之后生成包含一个或者多个所述登记流条目的所述封装分组的步骤，

其中所述 (b-2) 步骤包括：

通过将所述常规分组已经被供应到的所述交换机设置为所述第一通道交换机来向所述第一通道交换机供应所述封装分组的步骤。

9. 根据权利要求 6 至 8 中的任一权利要求所述的通信方法，其中所述分组传送处理步骤还包括：

(a-4) 在等于或者多于两个的所述登记流条目包含于所述封装分组中时，在提取一个所述登记流条目之后向所述路由上的下一个通道交换机发送所述封装分组的步骤；以及

(a-5) 在一个所述登记流条目包含于所述封装分组中时向目的地的终端发送包含于所述封装分组中的所述常规分组的步骤。

10. 根据权利要求 6 至 9 中的任一权利要求所述的通信方法，其中所述分组传送处理步骤还包括：

(a-6) : 响应于接收所述常规分组而查阅所述流表并且在包含与所述接收的常规分组的所述通信流匹配的所述匹配条件的所述流条目已经登记于所述流表上时基于所述登记流条目的所述动作对所述常规分组执行处理的步骤。

11. 一种交换机程序, 其示出用于使计算机作为执行分组传送处理的交换机来工作的过程, 其中所述分组包括从所述控制器供应的封装分组或者在网络上的终端之间发送和接收的分组, 所述交换机程序包括:

(a) 接收所述封装分组的步骤, 其中所述常规分组与在所述常规分组的所述通信流的所述路由上存在的一个或者多个通道交换机的流表中存储的所述登记流条目相关;

(b) 在所述通道交换机接收所述封装分组时从所述封装分组提取一个所述登记流条目的步骤;

(c) 在所述通道交换机的所述流表中存储所提取的所述登记流条目作为新的流条目的步骤;

(d) 指定示出将对所述常规分组或者所述封装分组执行的处理并且在所述新的流条目中示出的动作的步骤;

(e) 基于所述指定的动作来控制所述常规分组或者所述封装分组的传送的步骤, 其中所述 (e) 步骤包括:

对所述封装分组执行所述新的流条目中示出的所述动作的步骤。

12. 根据权利要求 11 所述的交换机程序, 还包括:

(f) 响应于接收所述常规分组来而查阅所述流表的步骤; 并且以及

(g) 在包含与所述接收的常规分组的所述通信流匹配的所述匹配条件的所述流条目未登记于所述流表上时向所述控制器供应所述常规分组的步骤。

13. 根据权利要求 12 所述的交换机程序, 还包括:

(h) 在两个或者更多所述登记流条目包含于所述封装分组中时, 在提取一个所述登记流条目之后向所述路由上的下一个通道交换机发送所述封装分组的步骤; 以及

(i) 在一个所述登记流条目包含于所述封装分组中时, 向作为目的地的终端发送包含于所述封装分组中的所述常规分组的步骤。

14. 一种控制程序, 其示出用于使计算机作为执行分组传送处理的控制器来工作的过程, 其中所述分组包括从所述控制器供应的封装分组或者在网络上的终端之间发送和接收的常规分组, 所述控制程序包括:

(a) 生成用于在存在于所述常规分组的通信流的路由上的至少一个通道交换机的流表中存储的至少一个登记流条目的步骤;

(b) 通过将所述常规分组与所述登记流条目相关来生成所述封装分组的步骤;

(c) 向所述通道交换机之一发送所述封装分组的步骤,

其中所述 (a) 步骤包括:

响应于接收所述常规分组来而指定传送所述常规分组时通过的所述网络的路由的步骤; 以及

在基于所述通道交换机在所述指定的路由上的位置而生成所述登记流条目之后形成包含所述登记流条目的所述封装分组的步骤, 以及

其中所述 (b) 步骤包括:

向所述第一通道交换机供应所述封装分组的步骤,所述第一通道交换机作为所述常规分组被供应到的所述交换机。

OpenFlow 通信系统和 OpenFlow 通信方法

技术领域

[0001] 本发明涉及一种 OpenFlow(开放流)通信系统和 OpenFlow 通信方法。

背景技术

[0002] 已知关于计算机网络的各种技术(例如参阅专利文献 1 至 4)。在专利文献 1 中,描述了一种如下 IP 流表的技术,在该表中登记 IP/MAC 源地址和目的地地址以及输出物理端口号。在专利文献 2 中,描述了一种通信单元的技术,其中基于存储的路由信息确定通向另一通信单元的路由并且还根据更新信息来更新路由信息。在专利文献 3 中,描述了一种分组中继单元的技术,其中接收路由信息的改变并且登记、删除和改变路由表的对应信息。在专利文献 4 中,描述了一种传输路由信息添加功能的技术,其中向发往中继单元的分组添加设置的路由信息。还描述了一种中继单元的如下技术,其中从接收的分组提取路由信息并且当分组未以中继单元为目的地向另一中继单元传送该分组。

[0003] 计算机网络如以太网(注册商标)是交换机(或者路由器)独立操作的分布类型。出于这一原因,难以正确和快速地掌握在网络中发生的现象,并且指定故障出现部分并且从故障恢复需要长久时间。而且,由于每个交换机有必要具有足以独立操作的能力,所以交换机的功能变得复杂。

[0004] 为了解决这样的问题,已经提出一种称为 OpenFlow 的新网络架构(例如参阅非专利文献 1)。OpenFlow 通过如在电话网络中那样对网络执行集中控制来实现网络的高级别可视化。在 OpenFlow 中也有可能相对减少交换机将实现的功能。因此,交换机变得更廉价,从而可以降低整个网络的成本。

[0005] 图 1 是示出了基于 OpenFlow 的网络系统(下文称为“OpenFlow 通信系统”)的配置的框图。OpenFlow 通信系统具有作为主要部件的 OpenFlow 控制器 10、OpenFlow 交换机 11 和将它们链接的链路 13。图 1 示出了具有多个 OpenFlow 交换机 11 的 OpenFlow 通信系统。在下文描述中,为了相互区分多个 OpenFlow 交换机 11 而使用分支号,如第一 OpenFlow 交换机 11-1。

[0006] OpenFlow 控制器 10 承担多个角色。OpenFlow 控制器 10 执行网络拓扑的识别、OpenFlow 交换机 11 的控制、OpenFlow 交换机 11 和链路 13 中的故障监视、用于分组 40 的通信路由的确定等等。

[0007] OpenFlow 交换机 11 中继来自邻居终端 12 和另一 OpenFlow 交换机 11(比如现有以太网(注册商标)交换机和 IP 路由器)的分组 40。应当注意,在下文描述中,为了相互区分多个终端 12 而使用分支号,如第一终端 12-1。

[0008] 图 2 是示出了 OpenFlow 交换机 11 的配置的框图。OpenFlow 交换机 11 具有输入端口 20、输出端口 21、本地管理部 22、流表 23 和分组交换机 24 作为主要部件。

[0009] 输入端口 20 从另一 OpenFlow 交换机 11 或者终端 12 接收分组 40。输出端口 21 向另一 OpenFlow 交换机 11 或者终端 12 发送分组 40。

[0010] 本地管理部 22 与 OpenFlow 控制器 10 通信并且根据来自 OpenFlow 控制器 10 的

指令更新流表 23。本地管理部 22 也响应于来自 OpenFlow 控制器 10 的指令向分组交换机 24 供应分组 40。另外,本地管理部 22 根据需要向 OpenFlow 控制器 10 发送通过输入端口 20 从外部单元接收的分组 40。

[0011] 分组交换机 24 针对通过查阅流表 23 获得的输出端口 21 或者由本地管理部 22 指示的输出端口 21 传送分组 40。流表 23 存储用来处置向 OpenFlow 交换机 11 供应的分组 40 的数据。

[0012] 图 3 是示出了流表 23 的配置的框图。流表 23 保持流条目 30 的集合。每个流条目 30 具有匹配条件 31 和动作 32 这两个字段。

[0013] 每当 OpenFlow 交换机 11 从外部单元接收分组 40, OpenFlow 交换机 11 就查阅流表 23 以比较分组 40 与匹配条件 31。例如,当满足匹配条件 31 之一时,向分组 40 应用与匹配条件 31 对应的动作 32。当分组 40 未满足所有匹配条件 31 时,OpenFlow 交换机 11 向 OpenFlow 控制器 10 发送分组 40。每个匹配条件 31 包括网络层 (IP) 的协议号、源 / 目的地地址、传送层 (TCP 或者 UDP) 中的源 / 目的地端口号。数据链路层 (以太网 (注册商标)) 中的源 / 目的地 MAC 地址、类型值、VLAN-ID 条件方程等。

[0014] 诸如“针对具体输出端口 21 输出分组 40”、“丢弃分组 40”等等之类的操作由动作 32 来限定。

[0015] 图 4 是示出了 OpenFlow 控制器 10 和 OpenFlow 交换机 11 的操作的流程图。图 4 示出了在图 1 中所示 OpenFlow 通信系统中从作为发送源的第一终端 12-1 向作为目的地的第三终端 12-3 的通信流。还假设所有 OpenFlow 交换机 11 的流表 23 在初始状态下为空 (Empty)。

[0016] 第一终端 12-1 发送属于流的第一分组 40。在步骤 S1 处,第一 OpenFlow 交换机 11-1 在输入端口 20 处接收分组 40。然后在步骤 S2 处,第一 OpenFlow 交换机 11-1 检查与分组 40 匹配的匹配条件 31 是否存在于流表 23 中。由于流表 23 在这一点为空,所以流表 23 的搜索失败 (箭头指向“否”)。在步骤 S3 处,第一 OpenFlow 交换机 11-1 向 OpenFlow 控制器 10 发送分组 40。

[0017] 在步骤 S11 处, OpenFlow 控制器 10 从第一 OpenFlow 交换机 11-1 接收分组 40。在步骤 S12 处, OpenFlow 控制器 10 从分组 40 提取作为发送源的终端 12 (第一终端 12-1) 的地址和作为目的地的终端 12 (第三终端 12-3) 的地址等等,并且计算用于待传送的分组 40 的路由。OpenFlow 控制器 10 可以选择适当路由,因为它掌握网络的拓扑。参照图 1,将分组 40 的通信路由确定为从第一 OpenFlow 交换机 11-1 到第二 OpenFlow 交换机 12-1 到第三 OpenFlow 交换机 11-3 的路由。

[0018] 在步骤 S13 处, OpenFlow 控制器 10 在计算路由之后向路由上的所有 OpenFlow 交换机 11 发出指令以更新流表 23。

[0019] 在步骤 S4 处,第一 OpenFlow 交换机 11-1 到第三 OpenFlow 交换机 11-3 中的每个交换机响应于来自 OpenFlow 控制器 10 的指令向流表 23 添加新的流条目 30。

[0020] 图 5 是示出了向其添加新的流条目 30 的流表 23 的配置的框图。在第一 OpenFlow 交换机 11-1 到第三 OpenFlow 交换机 11-3 中的每个交换机中,在图 5 的 (a) 中示出了流表 23 在添加之前的状态,并且图 5 的 (b) 示出了流表 23 在添加之后的状态。

[0021] 再次参照图 4,在步骤 S14 处, OpenFlow 控制器 10 向第一 OpenFlow 交换机 11-1 发

回分组 40。这时,OpenFlow 控制器 10 指示第一 OpenFlow 交换机 11-1 通过与第二 OpenFlow 交换机 11-2 连接的输出端口 21 发送分组 40。其原因在于第二 OpenFlow 交换机 11-2 位于路由上的第二位置上。

[0022] 在步骤 S6 处,第一 OpenFlow 交换机 11-1 响应于指令针对第二 OpenFlow 交换机 11-2 发送从 OpenFlow 控制器 10 返回的分组 40。

[0023] 接着,控制流程转向第二 OpenFlow 交换机 11-2 的处理。第二 OpenFlow 交换机 11-2 在步骤 S1 处从第一 OpenFlow 交换机 11-1 接收分组、在步骤 S2 处检查与分组 40 匹配的匹配条件 31 是否存在于流表 23 中。这时已经将第二 OpenFlow 交换机 11-2 的流表 23 设置成图 5 的 (b) 中所示状态。因此,继续流表 23 的搜索(箭头在步骤 S2 处指向“是”)。向分组 40 应用与匹配条件 31 对应的动作 32(步骤 S5)。

[0024] 参照图 5,由于动作 32 在这一例子中是“针对第三 OpenFlow 交换机 11-3 的输出”,所以通过与第三 OpenFlow 交换机 11-3 连接的输出端口 21 向第三 OpenFlow 交换机 11-3 发送分组 40。由于第三 OpenFlow 交换机 11-3 的操作与第二 OpenFlow 交换机 11-2 的操作相同,所以省略描述。

[0025] 如上文提到的那样,该流的第一分组 40 被中继并且最终向作为目的地的第三终端 12-3 发送。在从第一 OpenFlow 交换机 11-1 依次向第二 OpenFlow 交换机 11-2、向第三 OpenFlow 交换机 11-3 传递之时向目的地传送属于相同流的后续分组 40 而不经 OpenFlow 控制器 10。

[0026] 具体而言,这时已经在这些 OpenFlow 交换机 11 的流表 23 上登记了与分组 40 匹配的匹配条件 31。因此,控制流程继续到图 4 的步骤 S5。然后,向分组 40 应用与匹配条件 31 对应的动作 32。因此实现上述流。

[0027] 引用列表:

[0028] [专利文献]

[0029] [专利文献 1] :JP 2000-295274A

[0030] [专利文献 2] :JP 2005-191922A

[0031] [专利文献 3] :JP 2005-354579A

[0032] [专利文献 4] :JP-A-Heisei 11-341060

[0033] [非专利文献]

[0034] [非专利文献 1] :OpenFlow Switch Consortium(<http://www.openflowswitch.org/>)

发明内容

[0035] 在由数百个或者更多交换机组成的的大规模网络中,通过多个交换机传送分组 40。在这样的情况下,每当生成新流,OpenFlow 控制器 10 指示路由上的多个 OpenFlow 交换机 11 中的每个交换机更新流表 23。

[0036] 例如在平均 10 个交换机居间于路由上并且每秒生成 10,000 个新流的情形下,流表 23 的更新频率变成每秒 100,000 次。由于单个 OpenFlow 控制器 10 执行处理并不实际,所以采用一种布置多个 OpenFlow 控制器 10 用于负荷分布的技术。然而,这种方案与 OpenFlow 控制器 10 的数目增加相伴地引入设施成本和管理成本的增加,以及 OpenFlow 控制器 10 的

复杂化,因为用于在多个 OpenFlow 控制器 10 之间建立同步的机制变得单独必需。

[0037] 鉴于上述境况而实现本发明,并且目的在于通过减少从 OpenFlow 控制器 10 到 OpenFlow 交换机 11 的用于更新流表 23 的指令数目来降低 OpenFlow 控制器 10 的处理负荷而不改变网络的规模和配置。

[0038] 为了获得上述主旨,提供一种通信系统,该通信系统至少包括与网络连接以执行分组传送的交换机和配置成控制交换机的控制器。希望分组包括在网络中的终端之间发送和接收的常规分组和从控制器供应的封装分组。希望交换机包括:流表,其可以保持多个流条目;以及本地管理部,其被配置成基于多个流条目中的对应流条目来控制每个分组的传送。希望多个流条目中的每个流条目包括用于标识分组的通信流的匹配条件和示出分组的处理的动作。希望控制器生成用于在存在于通信流的路由上的至少一个通道交换机 (passage switch) 的流表中存储的至少一个登记流条目,并且将登记流条目与常规分组相关以生成封装分组。通道交换机响应于接收封装分组从封装分组提取登记流条目以在通道交换机的流表中存储为新的流条目并且对封装分组执行新的流条目示出的动作。

[0039] 简述公开的本发明中的一项典型发明所获得的效果,可以构建如下 OpenFlow 通信系统,在该系统中可以减少 OpenFlow 控制器 10 的处理负荷。

[0040] 根据本发明,OpenFlow 控制器 10 向路由上的第一 OpenFlow 交换机 11 发送分组 40 的路由信息,并且在沿着路由将路由信息与分组 40 一起中继之时向路由上的 OpenFlow 交换机 11 的流表 23 登记路由信息。按照这一配置和操作,可以减少从 OpenFlow 控制器 10 向 OpenFlow 交换机 11 发出对流表 23 的更新指令的次数。

附图说明

[0041] 结合示例实施例的描述和附图可以使本发明的目的、效果和特征更清楚。

[0042] 图 1 是示出了基于 OpenFlow 的网络的配置的框图;

[0043] 图 2 是示出了 OpenFlow 交换机 11 的配置的框图;

[0044] 图 3 是示出了流表 23 的配置的框图;

[0045] 图 4 是示出了 OpenFlow 控制器 10 和 OpenFlow 交换机 11 的操作的流程图;

[0046] 图 5 是示出了流表 23 在新的流条目 30 被添加时的配置的框图;

[0047] 图 6 是示出了根据本发明第一示例实施例的 OpenFlow 通信系统的配置的框图;

[0048] 图 7 是示出了第一示例实施例中的 OpenFlow 控制器 10 和 OpenFlow 交换机 11 的配置和连接关系的框图;

[0049] 图 8 是示出了第一示例实施例中的 OpenFlow 通信系统的操作的流程图;

[0050] 图 9 是示出了第一示例实施例中的 OpenFlow 通信系统的操作的流程图;

[0051] 图 10 是示出了 OpenFlow 控制器 10 生成的封装分组 50 的配置的框图;

[0052] 图 11 是示出了第一封装分组 50-1 和第二封装分组 50-2 的配置的图;

[0053] 图 12 是示出了将对第二封装分组 50-2 执行的处理的框图;

[0054] 图 13 是示出了将对第三封装分组 50-3 执行的处理的框图;

[0055] 图 14 是示出了第二示例实施例中的 OpenFlow 交换机 11 的流表 23 中保持的流条目 30 的配置的框图;以及

[0056] 图 15 是示出了第二示例实施例中的 OpenFlow 通信系统的操作的流程图。

具体实施方式

[0057] 下文将参照附图详细描述本发明的示例实施例。应当注意：向相同部件分配相同标号并且省略重复描述。

[0058] [第一示例实施例]

[0059] 图 6 是示出了根据第一示例实施例的 OpenFlow 通信系统中的 OpenFlow 控制器 10 和 OpenFlow 交换机 11 的配置的框图。应当注意，在本示例实施例中，OpenFlow 通信系统的总体配置类似于图 1 中所示上述配置。在当前 OpenFlow 规范中还假设使用以太网 (Ethernet, 注册商标) 作为用于数据链路层的协议并且使用 TCP/IP 或者 UDP/IP 作为网络层和传送层上的协议。然而，在本示例实施例中，对所用网络协议无约束。也为了标识设置于 OpenFlow 通信系统中的多个 OpenFlow 交换机 11 和终端 12，在下文描述中通过添加分支号来标识部件（比如第一 OpenFlow 交换机 11-1）。

[0060] 参照图 6，本示例实施例的 OpenFlow 控制器 10 具有封装分组生成部 14。后文将描述封装分组生成部 14 的操作。OpenFlow 交换机 11 还具有输入端口 20、输出端口 21、本地管理部 22、流表 23 和分组交换机 24。

[0061] 输入端口 20 从 OpenFlow 控制器 10 或者另一 OpenFlow 交换机 11 接收分组 40。分组 40 是常规分组 41 或者封装分组 50 中的任一者。输入端口 20 也从终端 12 接收常规分组 41。输出端口 21 向 OpenFlow 控制器 10 或者另一 OpenFlow 交换机 11 发送常规分组 41 或者封装分组 50。输出端口 21 也向终端 12 发送常规分组 41。

[0062] 本地管理部 22 与 OpenFlow 控制器 10 通信并且响应于来自 OpenFlow 控制器 10 的指令更新流表 23。本地管理部 22 还向分组交换机 24 供应从 OpenFlow 控制器 10 指示的常规分组 41 或者封装分组 50。另外，本地管理部 22 按照需要向 OpenFlow 控制器 10 发送常规分组 41 或者封装分组 50，所述分组通过输入端口 20 从外部单元供应。

[0063] 分组交换机 24 向通过查阅流表 23 而获得的输出端口 21 传送常规分组 41 或者封装分组 50。分组交换机 24 也向从本地管理部 22 指示的输出端口 21 传送常规分组 41 或者封装分组 50。

[0064] 图 7 是示出了本示例实施例中的 OpenFlow 控制器 10 和 OpenFlow 交换机 11 的配置和连接关系的框图。参照图 7，OpenFlow 控制器 10 具有在控制器这一侧上的处理器 60、在控制器这一侧上的存储器 62 和在控制器这一侧上的网络接口 63。它们通过在控制器这一侧上的总线 61 来连接。

[0065] 在控制器这一侧上的处理器 60 (CPU：中央处理单元) 执行对 OpenFlow 控制器 10 的各种部分的控制以及数据的计算和处理。处理器 60 具有用于执行存储于存储器 62 中的程序的功能。处理器 60 从输入单元 (未示出) 和存储单元如 HDD 接收数据并且在计算和处理之后向输出单元 (未示出) 和存储单元输出。网络接口 63 是与链接 OpenFlow 控制器 10 和 OpenFlow 交换机 11 的链路 13 连接的接口。

[0066] 存储器 62 是存储器 60 直接存取的半导体存储器单元。存储器 62 存储 OpenFlow 控制器 10 中处理的数据和程序。存储器 62 具有 OpenFlow 交换机控制程序 70 和封装分组生成程序 73。OpenFlow 交换机控制程序 70 示出了 OpenFlow 交换机 11 的控制过程。封装分组生成程序 73 示出了用于生成封装分组的过程。

[0067] OpenFlow 交换机 11 具有第一网络接口 64、第二网络接口 65、在交换机这一侧上的处理器 66 和在交换机这一侧上的存储器 68，并且它们通过在交换机这一侧上的总线 67 来连接。

[0068] 第一网络接口 64 和第二网络接口 65 是用于连接 OpenFlow 交换机 11 与链路 13 的接口。应当注意：第一网络接口 64 和第二网络接口 65 可以集成为单个单元。处理器 66 执行对 OpenFlow 交换机 11 的各种部分的控制，以及在 OpenFlow 交换机 11 中处理的数据的计算和处理。处理器 66 也具有用于执行存储于存储器 68 中的程序的功能，并且从输入单元（未示出）和存储器如 HDD 接收数据并且在计算和处理之后向输出单元（未示出）和存储器输出。

[0069] 存储器 68 是处理器 66 直接存取的半导体存储器单元。处理器 66 存储 OpenFlow 交换机 11 中处理的数据和程序。存储器 68 具有本地管理程序 71、分组交换机程序 72 和流表 23。

[0070] 本地管理程序 71 示出了用于 OpenFlow 交换机 11 的本地管理部 22 的过程。分组交换机程序 72 示出了用于 OpenFlow 交换机 11 的分组交换机 24 的过程。流表 23 保持如下信息，该信息用于确定对向 OpenFlow 交换机 11 输入的分组 40 的处理。应当注意：可以用专用硬件实现在本示例实施例中的 OpenFlow 控制器 10 和 OpenFlow 交换机 11 中执行的每个计算机程序。

[0071] 图 8 和图 9 是用于示出本示例实施例的 OpenFlow 通信系统的操作的流程图。下述操作在从 OpenFlow 通信系统的终端 12 输出以另一节点为目的地的分组（常规分组 41）时开始。在步骤 S101 处，OpenFlow 交换机 11 通过输入端口接收常规分组 41 或者封装分组 50。

[0072] 在步骤 S102 处，OpenFlow 交换机 11 确定接收的分组 40 是常规分组 41 还是封装分组 50。应当注意：后文将描述封装分组 50 的配置。作为确定结果，当接收分组 40 为封装分组 50 时，控制流程继续到图 9 的步骤 S301，而当接收分组 40 为常规分组 41 时，控制流程继续到步骤 S103。这里将描述接收分组 40 为常规分组 41 的情况。

[0073] 在步骤 S103 处，OpenFlow 交换机 11 检查与常规分组 41 匹配的匹配条件 31 是否在流表 23 中。当与常规分组 41 匹配的匹配条件 31 在流表 23 中时，控制流程继续到步骤 S105，而当常规分组 41 匹配的匹配条件 31 不在流表 23 中时，控制流程继续到步骤 S104。

[0074] 在步骤 S105 处，由于接收分组 40 是常规分组 41 并且与常规分组 41 匹配的匹配条件 31 在流表 23 中，所以 OpenFlow 交换机 11 向常规分组 41 应用与匹配条件 31 对应的动作 32。

[0075] 在步骤 S104 处，由于接收分组 40 不是常规分组 41 并且与常规分组 41 匹配的任何匹配条件 31 都不在流表 23 中，所以 OpenFlow 交换机 11（第一 OpenFlow 交换机 11-1）向 OpenFlow 控制器 10 发送常规分组 41。此后，控制流程转向 OpenFlow 控制器 10。

[0076] 在步骤 S201 处，OpenFlow 控制器 10 接收从 OpenFlow 交换机 11 供应的常规分组 41。在步骤 S202 处，OpenFlow 控制器 10 从常规分组 41 提取作为发送源的终端 12（第一终端 12-1）和作为目的地的终端 12（第三终端 12-3）的地址等等，并且计算用于传送常规分组 41 的路由。在步骤 S203 处，OpenFlow 控制器 10 在路由计算之后生成封装分组 50。

[0077] 图 10 是示出了 OpenFlow 控制器 10 生成的封装分组 50 的配置的框图。封装分组

50 具有头部 51、一个或者多个登记流条目 52 和常规分组 41。

[0078] 头部 51 包括在 OpenFlow 交换机 11 区分常规分组 41 和封装分组 50 时使用的标识符。作为一种区分方法,有一种为封装分组 50 分配特殊目的地地址或者特殊源地址的方法和一种基于类型值区分以太网(注册商标)和 VLAN-ID 或者 MPLS 标签的方法等等。然而,方法不限于此。

[0079] 向路由上的第 X 个 OpenFlow 交换机 11 的流表 23 添加第 X 个登记流条目 52-X($1 \leq X \leq N$)。登记流条目 52 保持于流表 23 中作为流条目 30。应当注意:第 X 个登记流条目 52-X($1 \leq X \leq N$)的 N 是包含于封装分组 50 中的登记流条目 52 的数目。因此,第 X 个登记流条目 52-X($1 \leq X \leq N$)的内容与将向常规 OpenFlow 中的路由上的第 X 个 OpenFlow 交换机 11 添加的流条目 30 的内容重合。

[0080] 再次参照图 8,在步骤 S204 处,OpenFlow 控制器 10 向路由上的第一 OpenFlow 交换机 11(第一 OpenFlow 交换机 11-1)发送生成的封装分组 50。此后,控制流程转向 OpenFlow 交换机 11。

[0081] 在步骤 S101 处,OpenFlow 交换机 11 从 OpenFlow 控制器 10 接收封装分组 50。在步骤 S102 处,OpenFlow 交换机 11 确定通过输入端口 20 接收的分组 40 是常规分组 41 还是封装分组 50。这时,由于接收分组 40 为封装分组 50,所以控制流程继续到图 9 的步骤 S301。

[0082] 在步骤 S301 处,OpenFlow 交换机 11 从封装分组 50 查阅与本身对应的登记流条目 52 以向它自己的流表 23 添加登记流条目 52 的内容。换言之,OpenFlow 交换机 11 保持通过查阅而获得的登记流条目 52 的信息作为 OpenFlow 交换机 11 的流表 23 的流条目 30。

[0083] 在步骤 S302 处,OpenFlow 交换机 11 从封装分组 50 的与本身对应的登记流条目 52 取出动作 32。在下文描述中,将取出的动作 32 称为已存动作(这是暂时的,以便于理解本示例实施例的操作)。

[0084] 在步骤 S303 处,OpenFlow 交换机 11 确定包含于封装分组 50 中的登记流条目 52 的数目是否为一。作为确定结果,当包含多个登记流条目 52 时,控制流程继续到步骤 S304,而当登记流条目 52 为一时,控制流程继续到步骤 S305。

[0085] 在步骤 S304 处,OpenFlow 交换机 11 通过从封装分组 50 去除向流表 23 添加的登记流条目 52 来根据接收的封装分组 50 生成新封装分组 50。换言之,在每个 OpenFlow 交换机 11 中,通过去除与 OpenFlow 交换机 11 对应的登记流条目 52 来根据接收的封装分组 50 生成新封装分组 50。

[0086] 在步骤 S306 处,OpenFlow 交换机 11 向新封装分组 50 应用已存动作。

[0087] 在步骤 S305 处,OpenFlow 交换机 11 从接收的封装分组 50 取出常规分组 41 并且向常规分组 41 应用已存动作。

[0088] 在本示例实施例未应用于其上的 OpenFlow 通信系统中的 OpenFlow 控制器必须指示路由上的所有 OpenFlow 交换机 11 各个地对流表 23 进行更新。如上文所述,在根据本示例实施例的 OpenFlow 通信系统中的 OpenFlow 控制器 10 仅向路由上的第一 OpenFlow 交换机 11 发送路由信息。因而,OpenFlow 控制器 10 无需指示多个 OpenFlow 交换机 11 各个地对流表 23 进行更新。因此,在本示例实施例的 OpenFlow 通信系统中,有可能减少 OpenFlow 控制器 10 的处理负荷。

[0089] 下文将通过使用具体封装分组 50 的配置作为例子来描述本示例实施例的操作。在下文描述中,将描述在图 1 的上述 OpenFlow 通信系统中使用第一终端 12-1 作为发送源而第三终端 12-3 作为目的地的情况下的分组发送的通信流。也为了便于理解本示例实施例的操作,假设 OpenFlow 通信系统中的所有 OpenFlow 交换机 11 的流表 23 处于初始状态、即空状态。

[0090] 当终端 12(第一终端 12-1)发送属于流的第一常规分组 41 时,OpenFlow 交换机 11(第一 OpenFlow 交换机 11-1)通过输入端口 20 接收常规分组 41。OpenFlow 交换机 11(第一 OpenFlow 交换机 11-1)确定通过输入端口 20 接收的分组 40 是常规分组 41 还是封装分组 50。在初始状态,分组 40 是从第一终端 12-1 接收的常规分组 41 而不是封装分组 50。因此这一确定结果为否(假)。也由于流表 23 在初始状态为空,所以流表 23 的搜索失败。

[0091] 接收分组 40 为常规分组 41,并且与常规分组 41 匹配的匹配条件 31 不在流表 23 中。因此,OpenFlow 交换机 11(第一 OpenFlow 交换机 11-1)向 OpenFlow 控制器 10 发送常规分组 41。此后,控制流程转向 OpenFlow 控制器 10。

[0092] OpenFlow 控制器 10 接收从第一 OpenFlow 交换机 11-1 供应的常规分组 41。OpenFlow 控制器 10 从常规分组 41 提取作为发送源的终端 12(第一终端 12-1)和作为目的地的终端 12(第三终端 12-3)的地址等等并且计算用来传送常规分组 41 的通信路由。在本示例实施例中,选择常规分组 41 从第一 OpenFlow 交换机 11-1 到第二 OpenFlow 交换机 11-2 到第三 OpenFlow 交换机 11-3 的通信路由。这里,OpenFlow 控制器 10 在通过执行上述步骤 S203 的操作来计算通信路由之后生成封装分组 50。

[0093] 图 11 示出了 OpenFlow 控制器 10 生成的封装分组 50(第一封装分组 50-1)的配置和从第一 OpenFlow 交换机 11-1 输出的封装分组 50(第二封装分组 50-2)的配置。OpenFlow 控制器 10 生成的封装分组 50 的内容类似于图 11 的第一封装分组 50-1。如图 11 中所示,OpenFlow 控制器 10 向通信路由上的第一 OpenFlow 交换机 11(第一 OpenFlow 交换机 11-1)发送生成的封装分组 50(第一封装分组 50-1)。此后,控制流程转向 OpenFlow 交换机 11。

[0094] OpenFlow 交换机 11(第一 OpenFlow 交换机 11-1)从 OpenFlow 控制器 10 接收封装分组 50(第一封装分组 50-1)。OpenFlow 交换机 11(第一 OpenFlow 交换机 11-1)确定通过输入端口 20 接收的分组 40 是常规分组 41 还是封装分组 50。这时,由于接收分组 40 为封装分组 50,所以第一 OpenFlow 交换机 11-1 从封装分组 50 取出第一登记流条目 52(第一登记流条目 52-1)并且向它自己的流表 23 添加并登记取出的流条目。换言之,第一 OpenFlow 交换机 11-1 保持第一登记流条目 52-1 作为第一 OpenFlow 交换机 11-1 的流表 23 的流条目 23。

[0095] 这时已经在第一 OpenFlow 交换机 11-1 中登记的已存动作的内容与第一登记流条目 52-1 的动作 32 相同,并且“向第二 OpenFlow 交换机 11-2 输出”。也由于包含于封装分组 50 中的登记流条目 52 的数目为多个,所以 OpenFlow 交换机 11 通过从封装分组 50 删除已经向流表 23 添加的登记流条目 52、基于接收的封装分组 50 来生成新封装分组 50。OpenFlow 交换机 11 向新封装分组 50 应用已存动作。如上文所述,由于这时的已存动作是“针对第二 OpenFlow 交换机 11-2 输出”,所以通过与第二 OpenFlow 交换机 11-2 连接的输出端口 21 向第二 OpenFlow 交换机 11-2 发送新封装分组 50(第二封装分组 50-2)。

[0096] 图 12 是示出了第二封装分组 50-2 的处理的框图。另一 OpenFlow 交换机 11(第

二 OpenFlow 交换机 11-2) 从在前一级的 OpenFlow 交换机 11 (第一 OpenFlow 交换机 11-1) 接收分组 40 (第二封装分组 50-2), 并且在接收之后确定分组 40 是否为封装分组 50, 并且执行上述步骤 S301 至 S304。

[0097] 第二 OpenFlow 交换机 11-2 生成第三封装分组 50-3。由于这时的已存动作是“针对第三 OpenFlow 交换机 11-3 输出”, 所以向第三 OpenFlow 交换机 11-3 发送新封装分组 50 (第三封装分组 50-3)。

[0098] 图 13 是示出了第三封装分组 50-3 的处理的框图。第三 OpenFlow 交换机 11-3 从第二 OpenFlow 交换机 11-2 接收封装分组 50 (第三封装分组 50-3)。这时的动作是“针对第三终端 12-3 输出”。在第三 OpenFlow 交换机 11-3 接收的封装分组 50 (第三封装分组 50-3) 中包含的登记流条目 52 的数目为 1。因此, 从接收的封装分组 50 取出常规分组 41, 并且向常规分组 41 应用已存动作。如上文所述, 由于这时的动作是“针对第三终端 12-3 输出”, 所以向第三终端 12-3 发送常规分组 41。

[0099] 如上文提到的那样, 中继并且最终向作为目的地的第三终端 12-3 发送流中的第一常规分组 41。最终, 第一 OpenFlow 交换机 11-1 到第三 OpenFlow 交换机 11-3 的流表 23 的内容如图 5 的 (b) 中所示。

[0100] 属于相同流的后续常规分组 41 从第一 OpenFlow 交换机 11-1 依次传向第二 OpenFlow 交换机 11-2 又传向第三 OpenFlow 交换机 11-3 而不经 OpenFlow 控制器 10, 并且送往目的地。以这一方式操作的原因在于: 与常规分组 41 匹配的匹配条件 31 这时登记于这些 OpenFlow 交换机 11 的流表 23 上。步骤 S103 的确定结果在图 8 的流程图为真, 并且控制流程分支到步骤 S105。向常规分组 41 应用与匹配条件 31 对应的动作 32。

[0101] 应当注意, 在图 8 的流程图中, 步骤 S102 和步骤 S103 的执行顺序可改变。与上述操作描述相反, 可以比步骤 S102 更早地执行步骤 S103。

[0102] [第二示例实施例]

[0103] 下文将参照附图描述本发明的第二示例实施例。根据第二示例实施例的 OpenFlow 通信系统通过查阅 OpenFlow 交换机 11 的流表 23 来执行对接收分组 40 是常规分组 41 还是封装分组 50 的确定。

[0104] 图 14 是示出了第二示例实施例中的 OpenFlow 交换机 11 的流表 23 保持的流条目 30 的配置的框图。在第二示例实施例中, 图 14 的流条目 30 预先登记于所有 OpenFlow 交换机 11 的流表 23 上。

[0105] 图 15 是示出了第二示例系统中的 OpenFlow 通信系统的操作的流程图。第二示例实施例的操作不同于第一示例实施例的操作。基于流表 23 的流条目 30 执行相互区分常规分组 41 和封装分组 50 的步骤。在下文描述中, 当向 OpenFlow 交换机 11 输入从 OpenFlow 控制器 10 供应的封装分组 50 时, 将使用图 15 的流程图来描述第二示例实施例中的 OpenFlow 交换机 11 的操作。

[0106] 在步骤 S101 处, OpenFlow 交换机 11 接收通过输入端口 20 供应的分组 40 (在这一情况下为封装分组 50)。在步骤 S103 处, OpenFlow 交换机 11 检查与常规分组 41 匹配的匹配条件 31 是否在流表 23 中。如上文提到的那样, 图 14 的流表 30 登记于所有 OpenFlow 交换机 11 的流表 23 上。接收分组 40 也是封装分组 50。因此, 流表 23 的搜索成功, 并且控制流程继续到步骤 S107。

[0107] 在步骤 S107 处, 读出与匹配条件 31 对应的动作 32。动作 32 变成“作为封装分组 50 处理”。因此, 步骤 S107 的确定结果变成“是”(真), 并且将分组 40 作为封装分组 50 来处理。后续处理与第一实施例的后续处理相同。

[0108] 上文已经具体描述本发明的示例实施例。本发明不限于上述示例实施例, 并且有可能在不脱离本发明范围的范围内实现修改。因此可以容易实现上述示例实施例的各种修改。因此, 本发明决不限于上述示例实施例并且参照附图和说明书按最宽的范围来解释。应当注意, 本专利申请要求基于 2009 年 3 月 9 日提交的第 2009-055739 号日本专利申请的巴黎公约优先权, 并且通过引用将其公开内容结合于此。

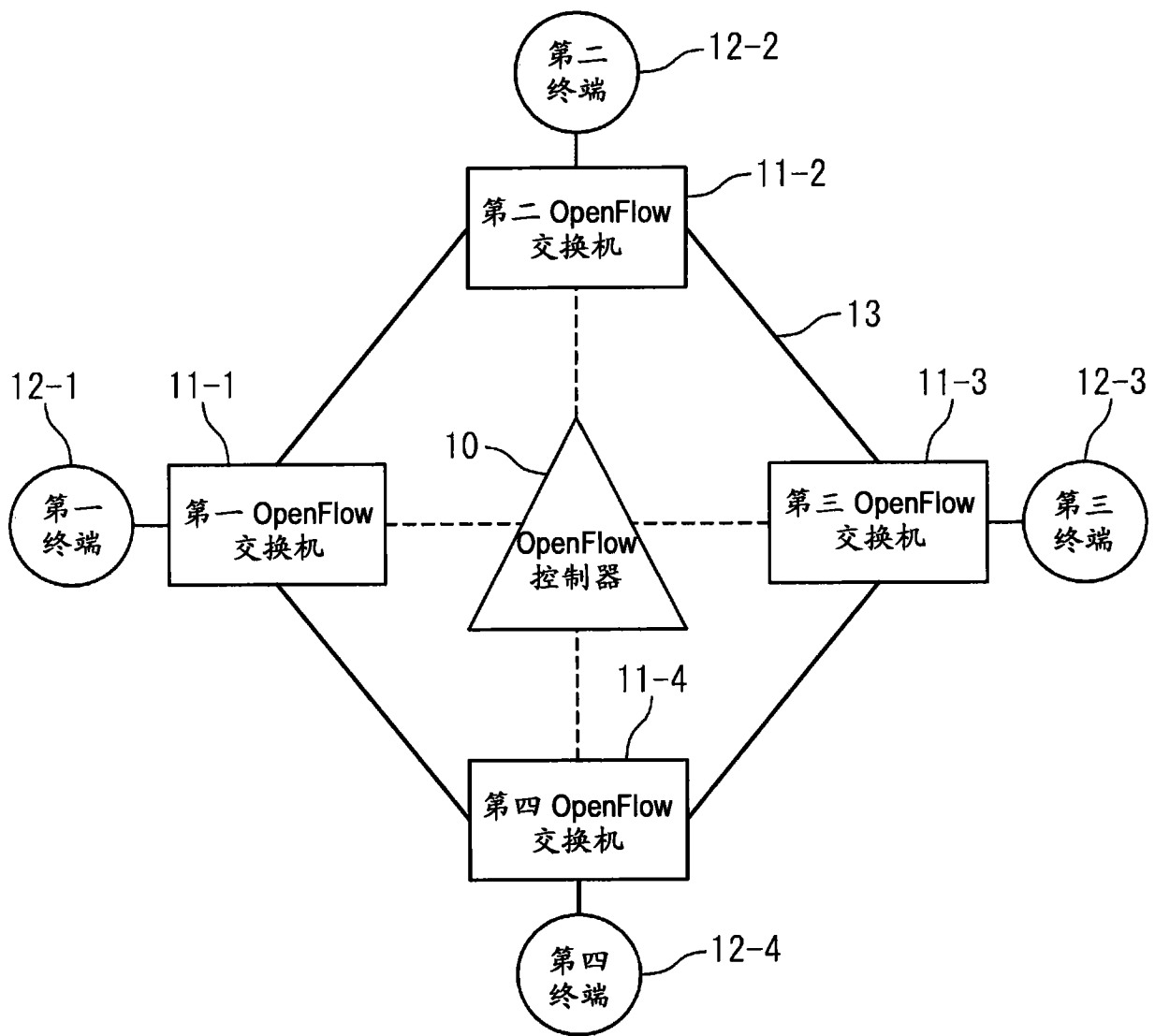


图 1

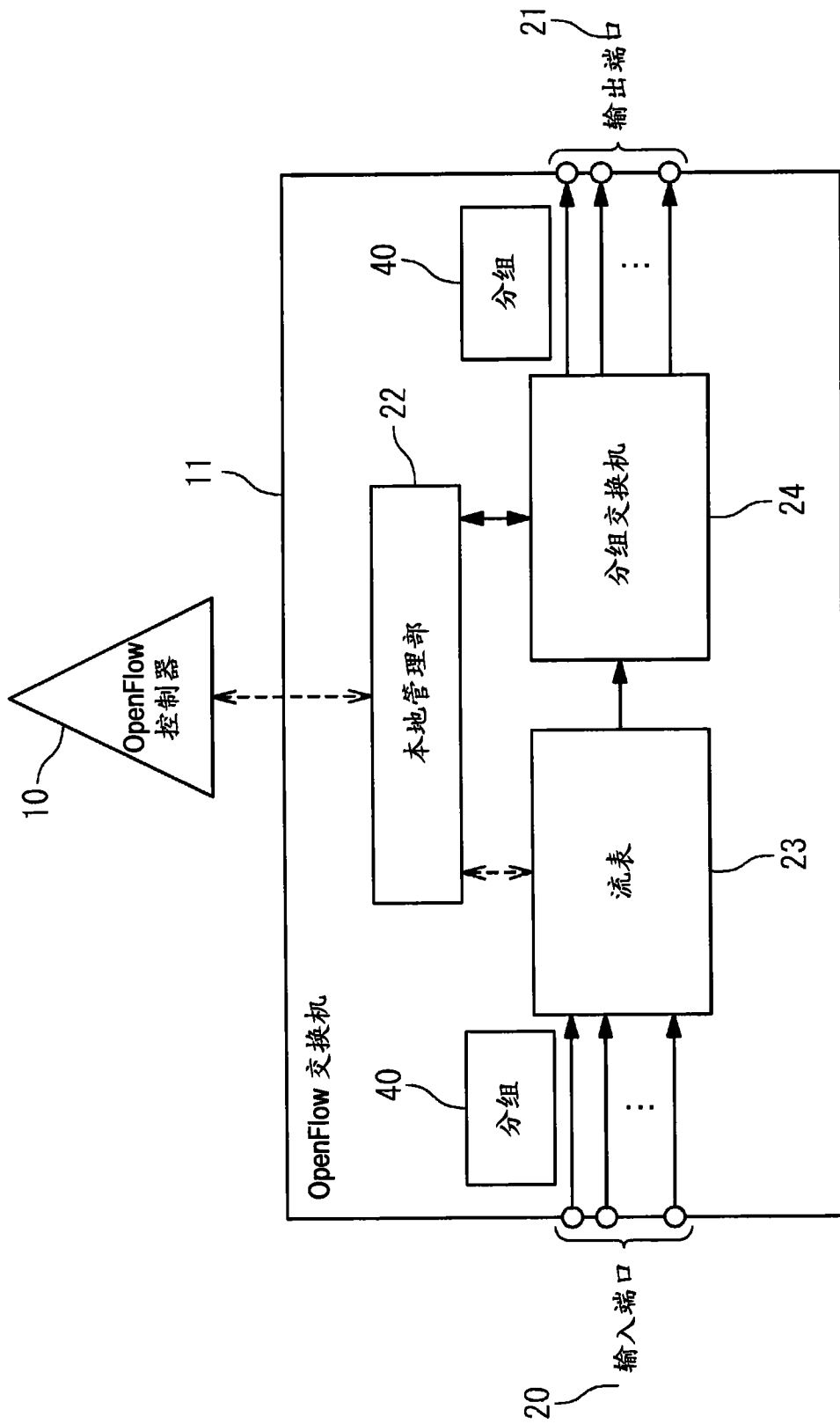


图 2

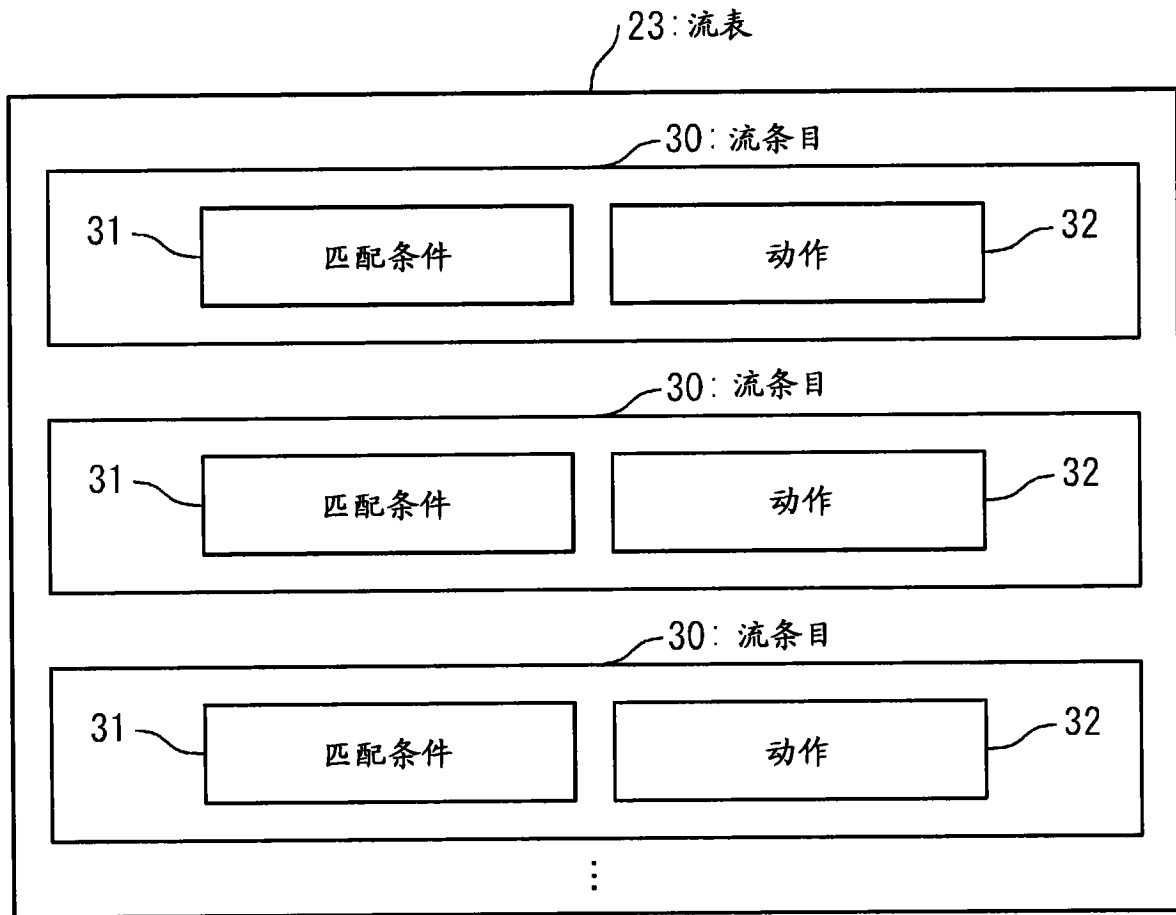


图 3

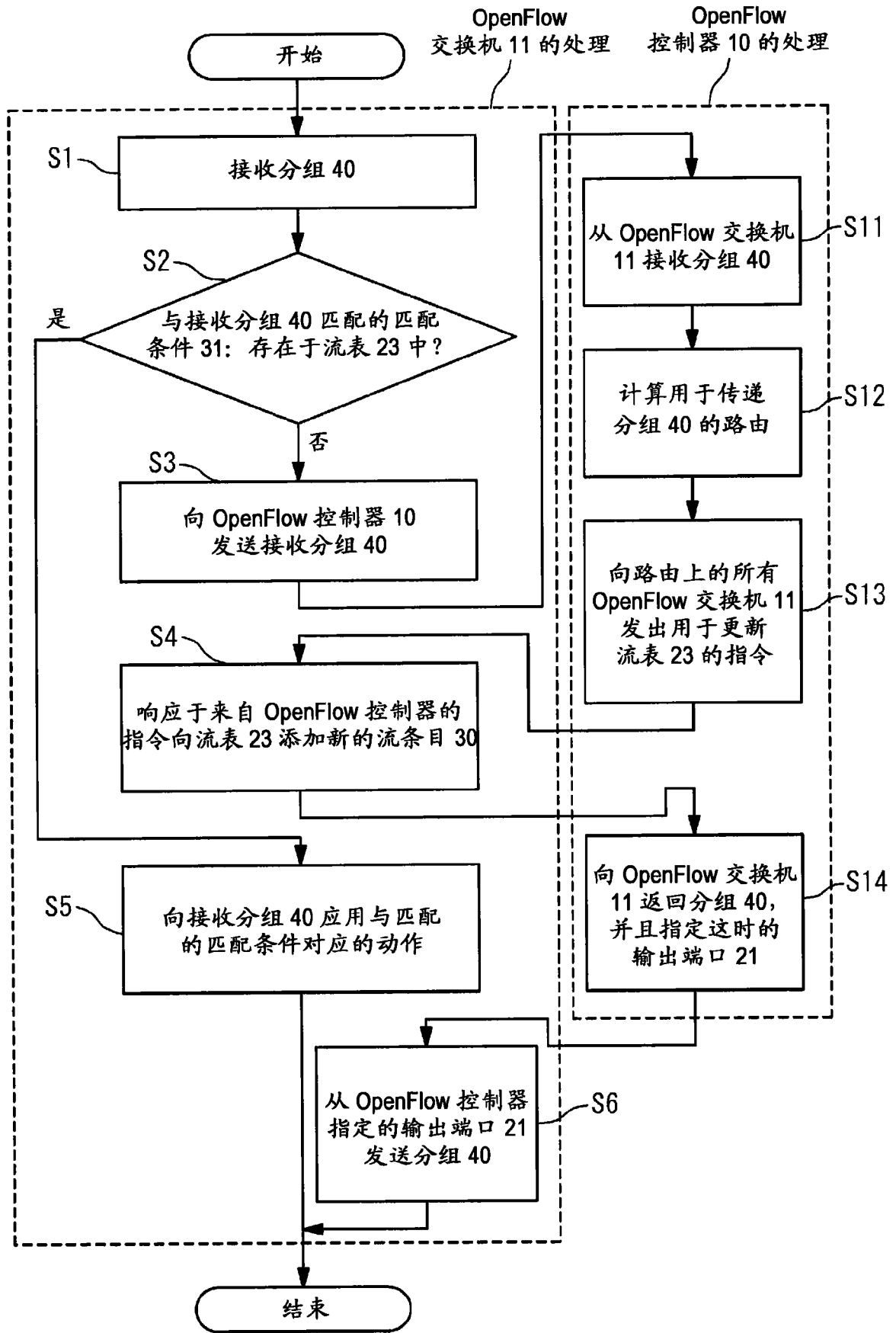


图 4

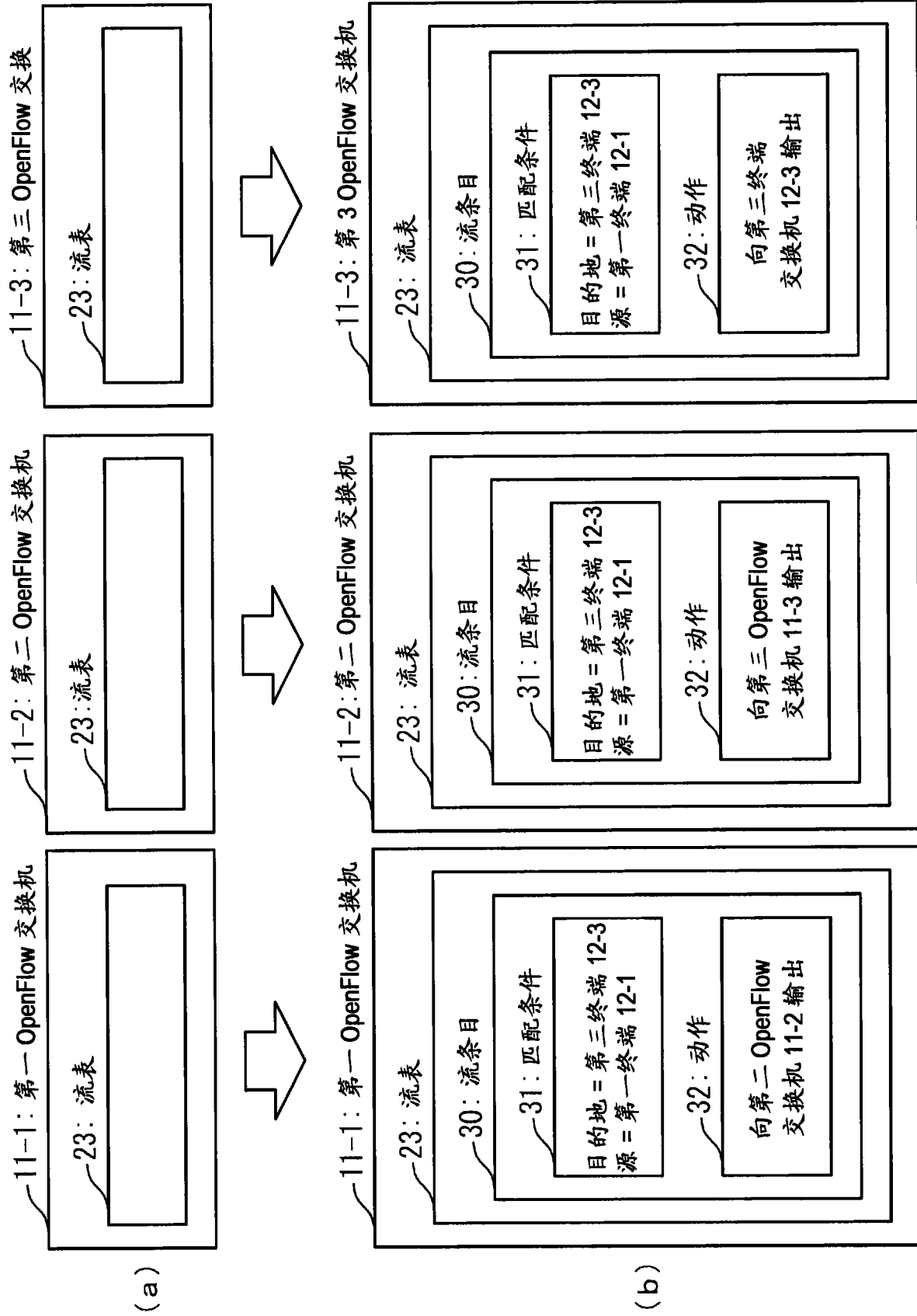


图 5

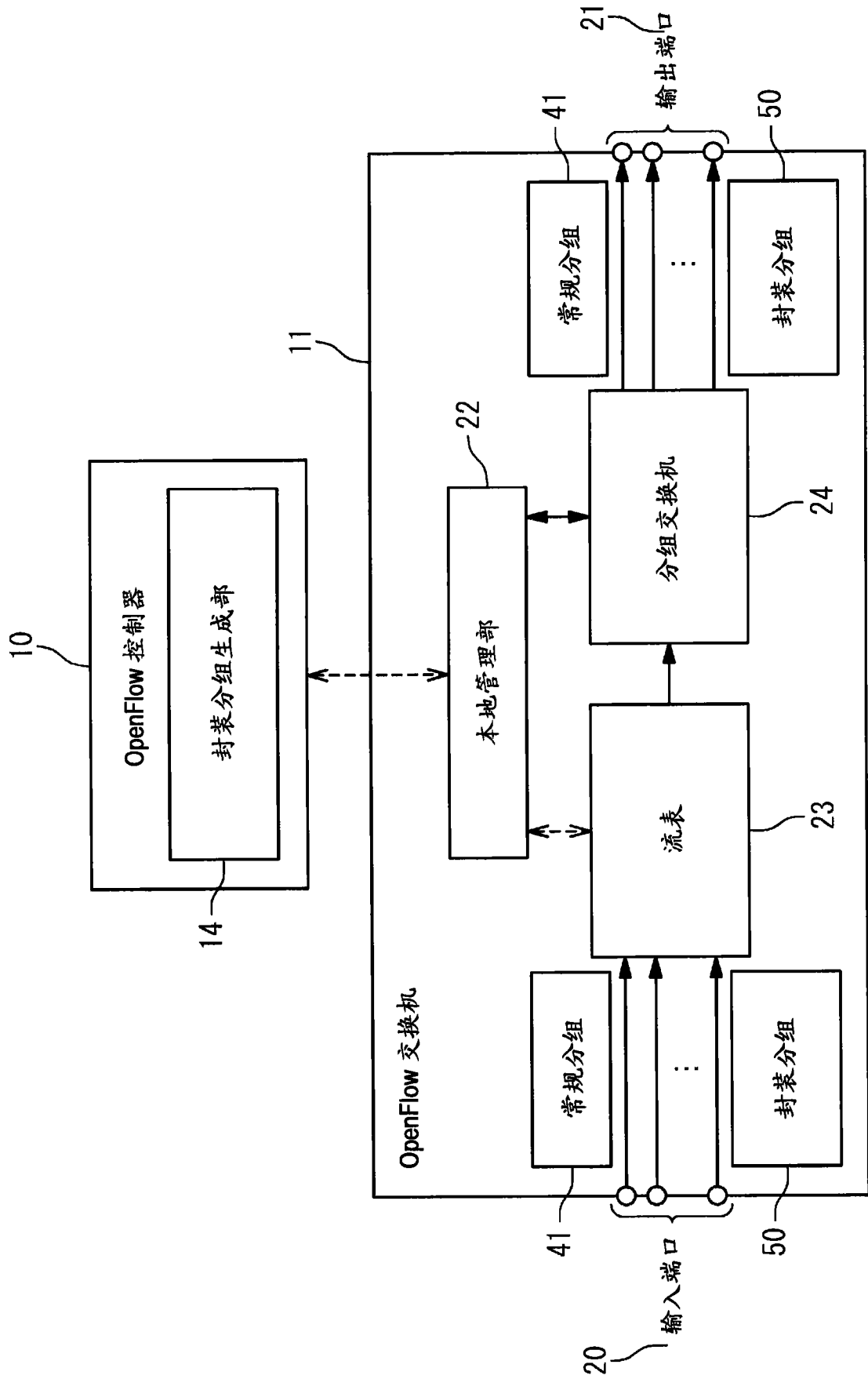


图 6

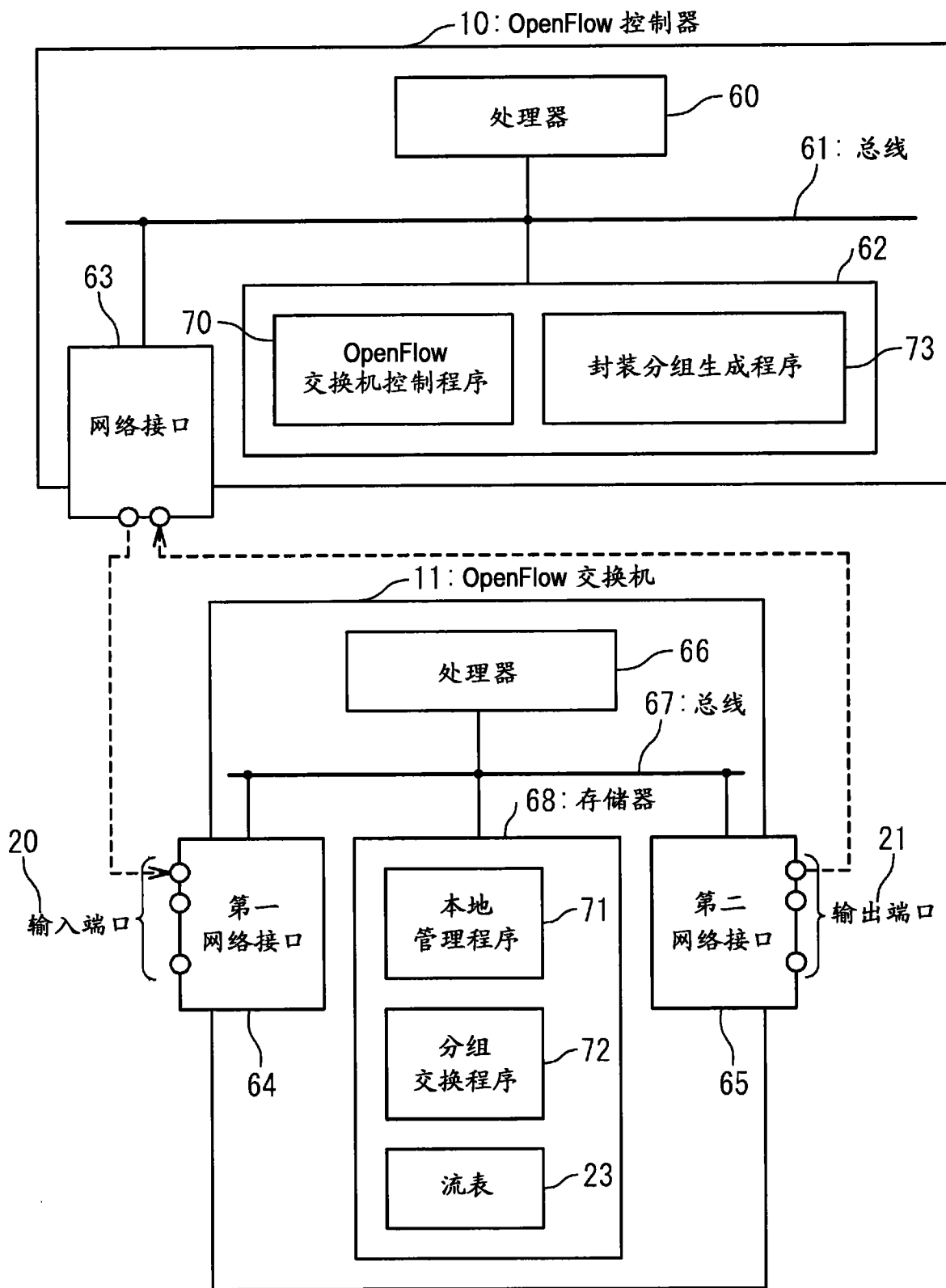


图 7

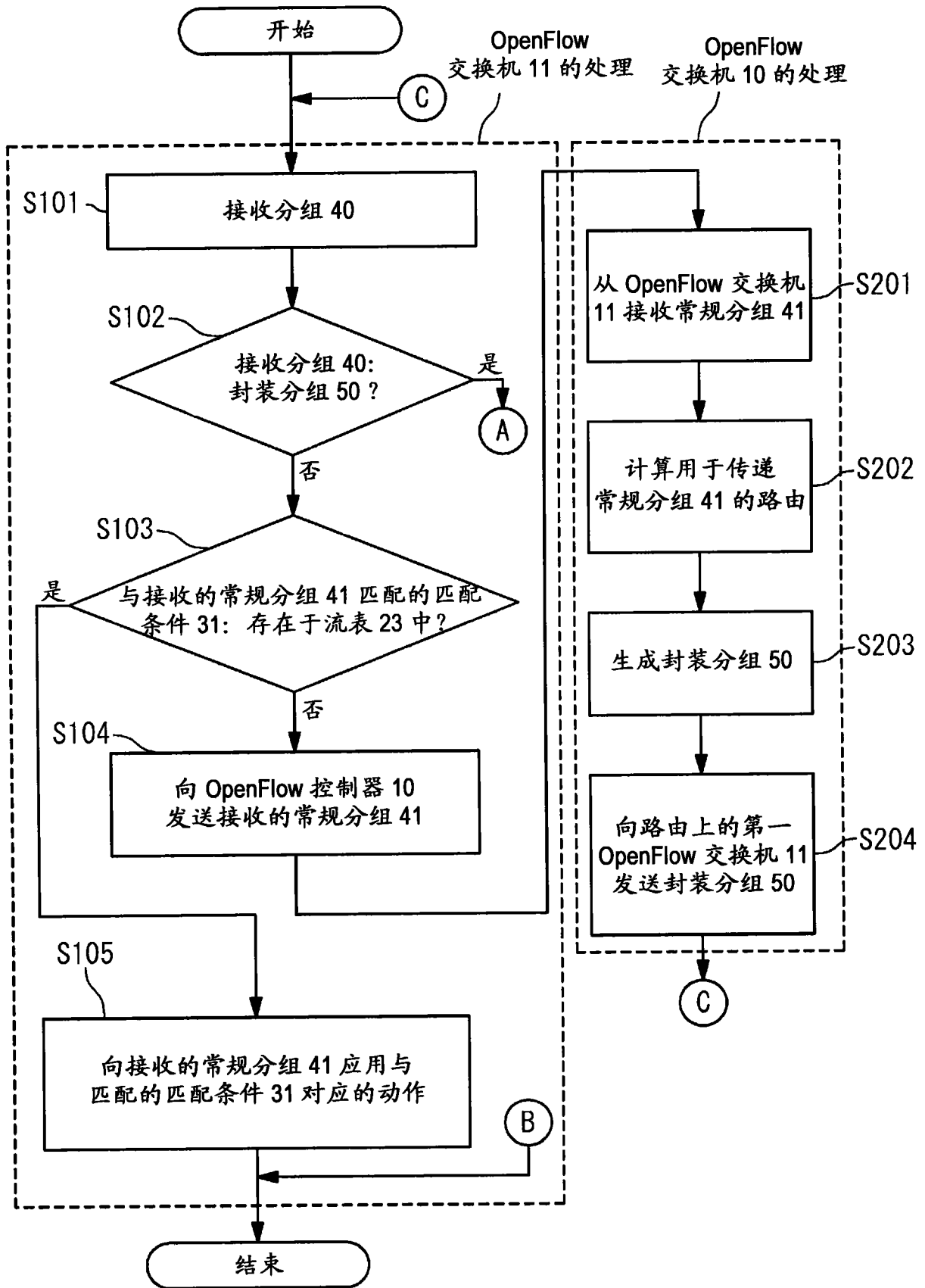


图 8

OpenFlow 交换机 11 的处理

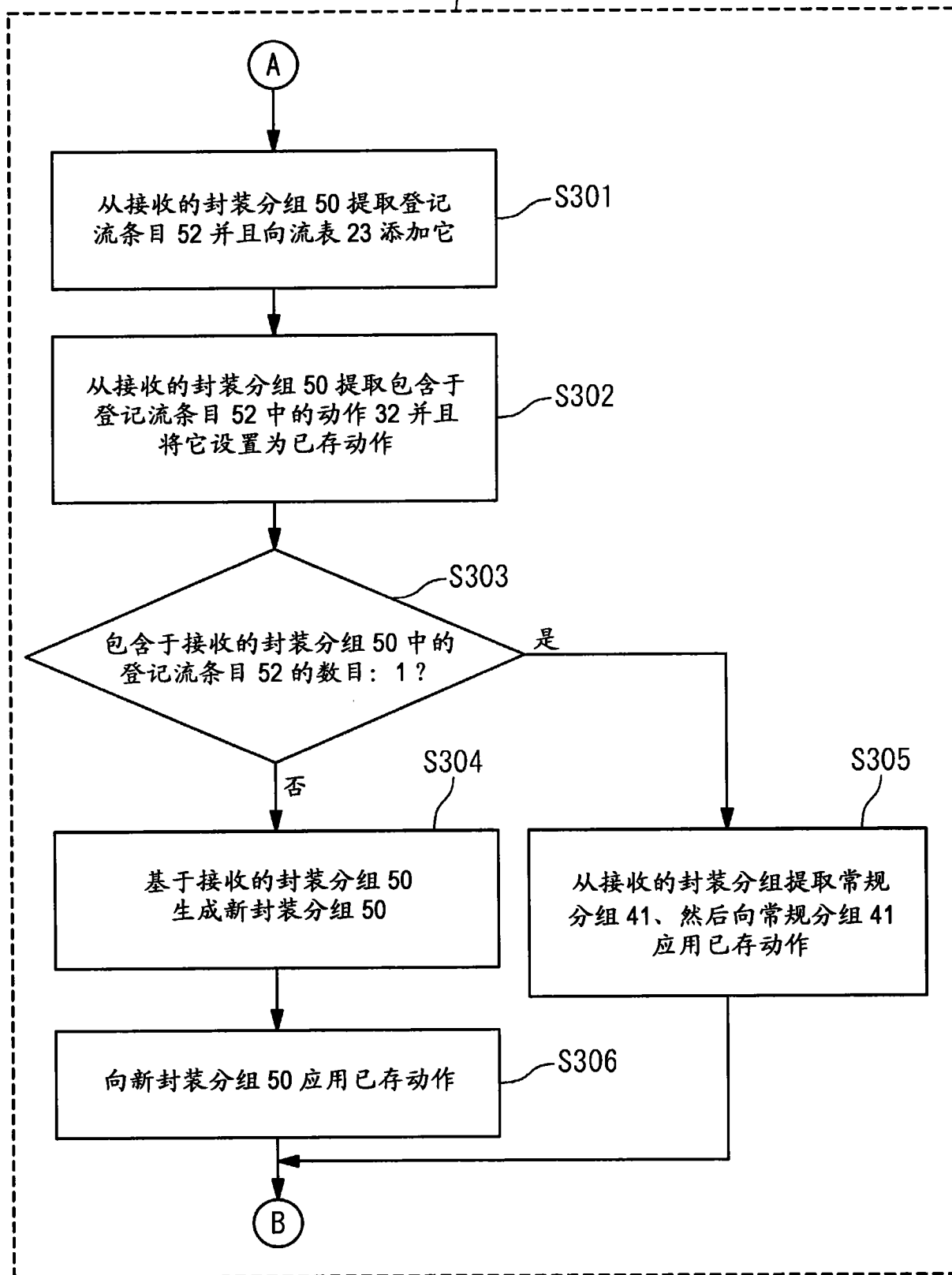


图 9

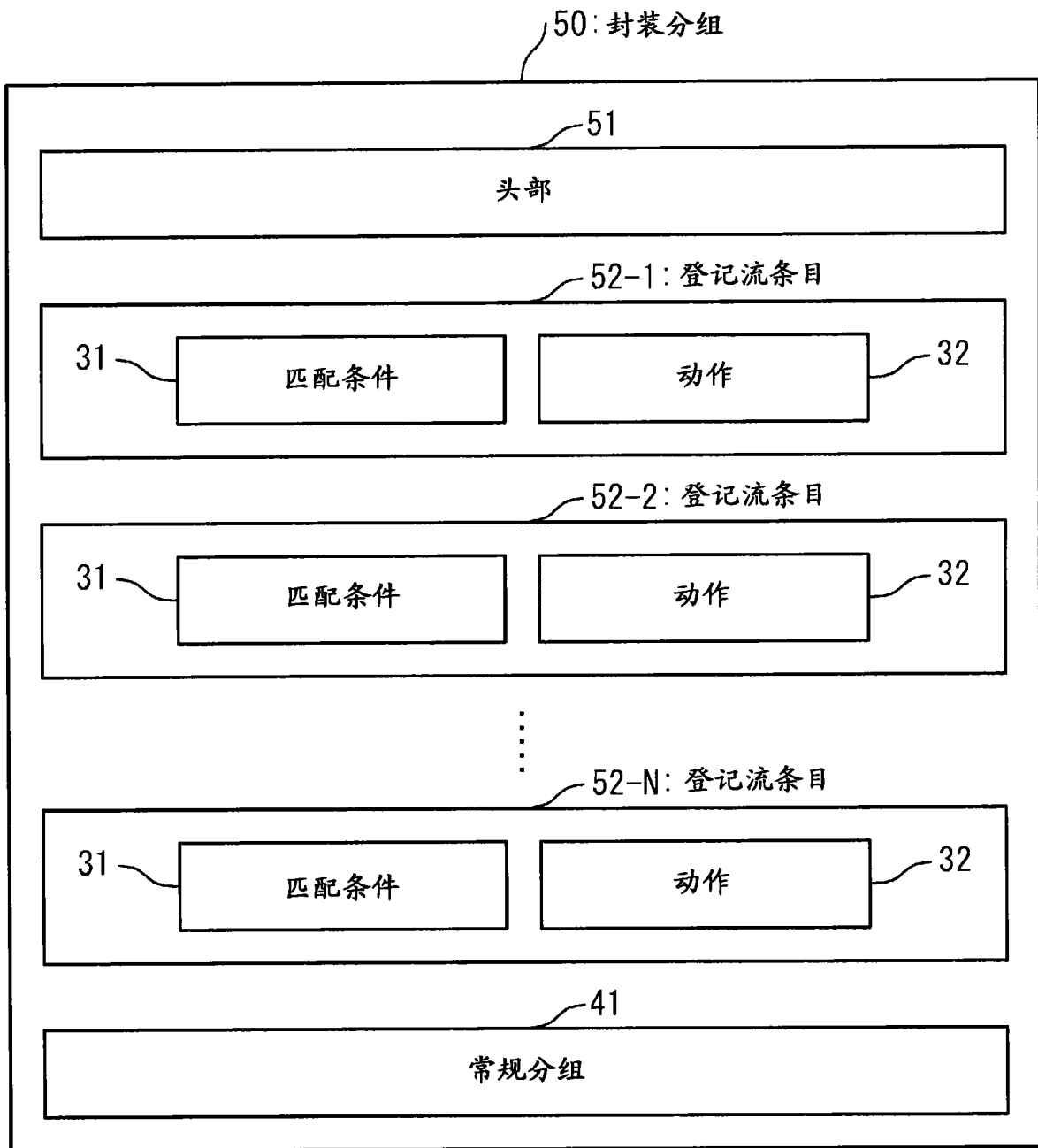


图 10

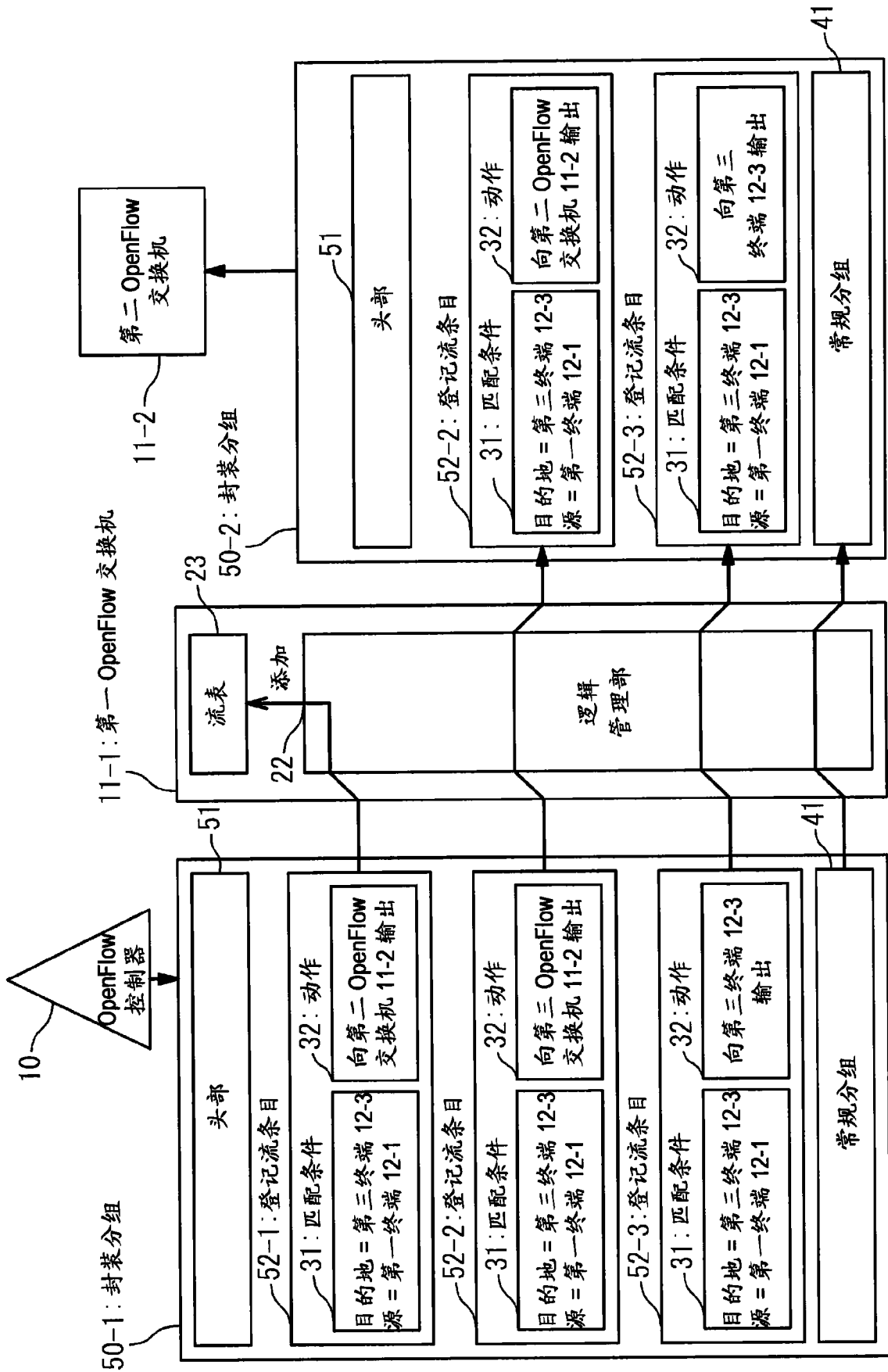


图 11

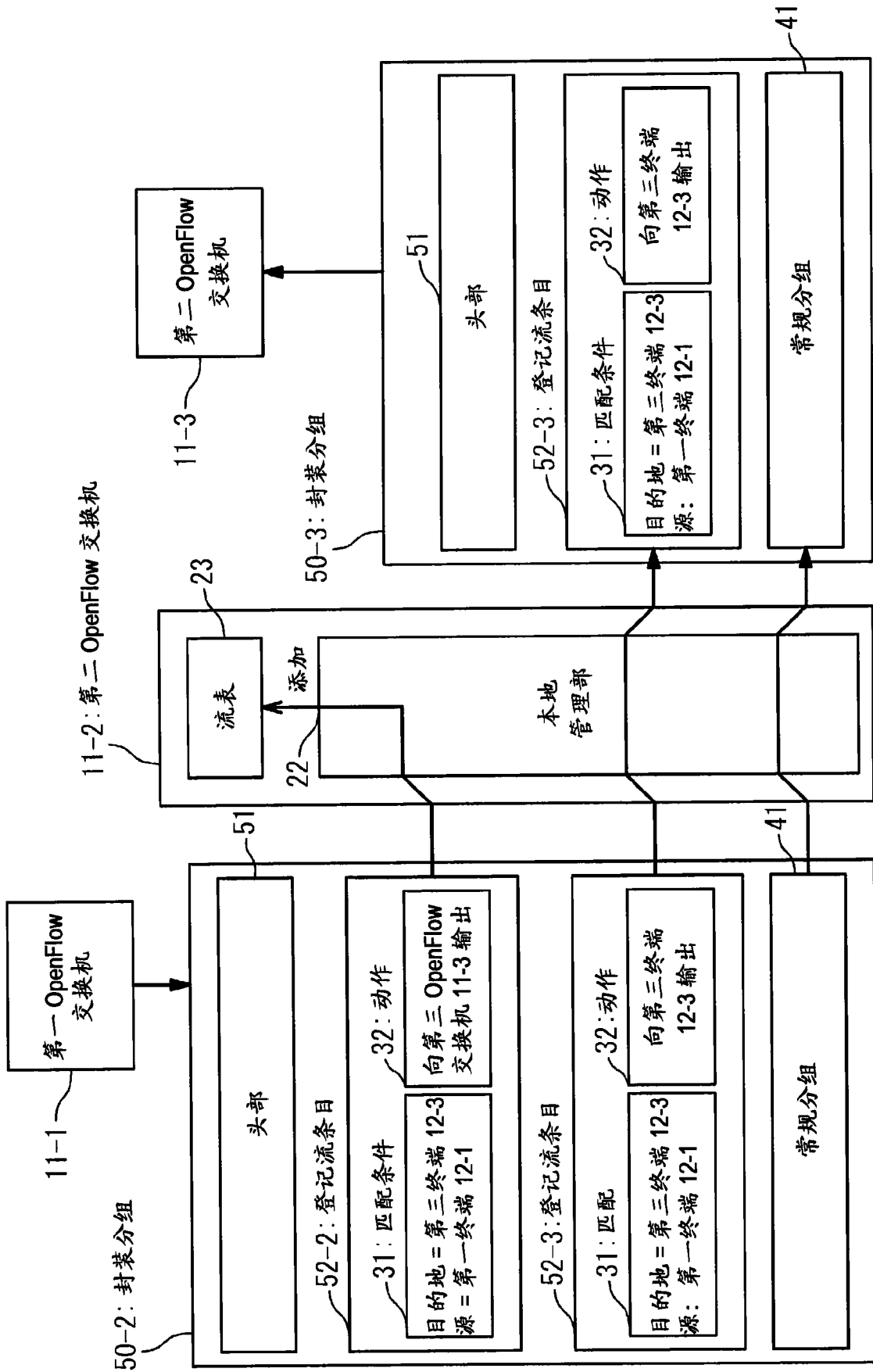


图 12

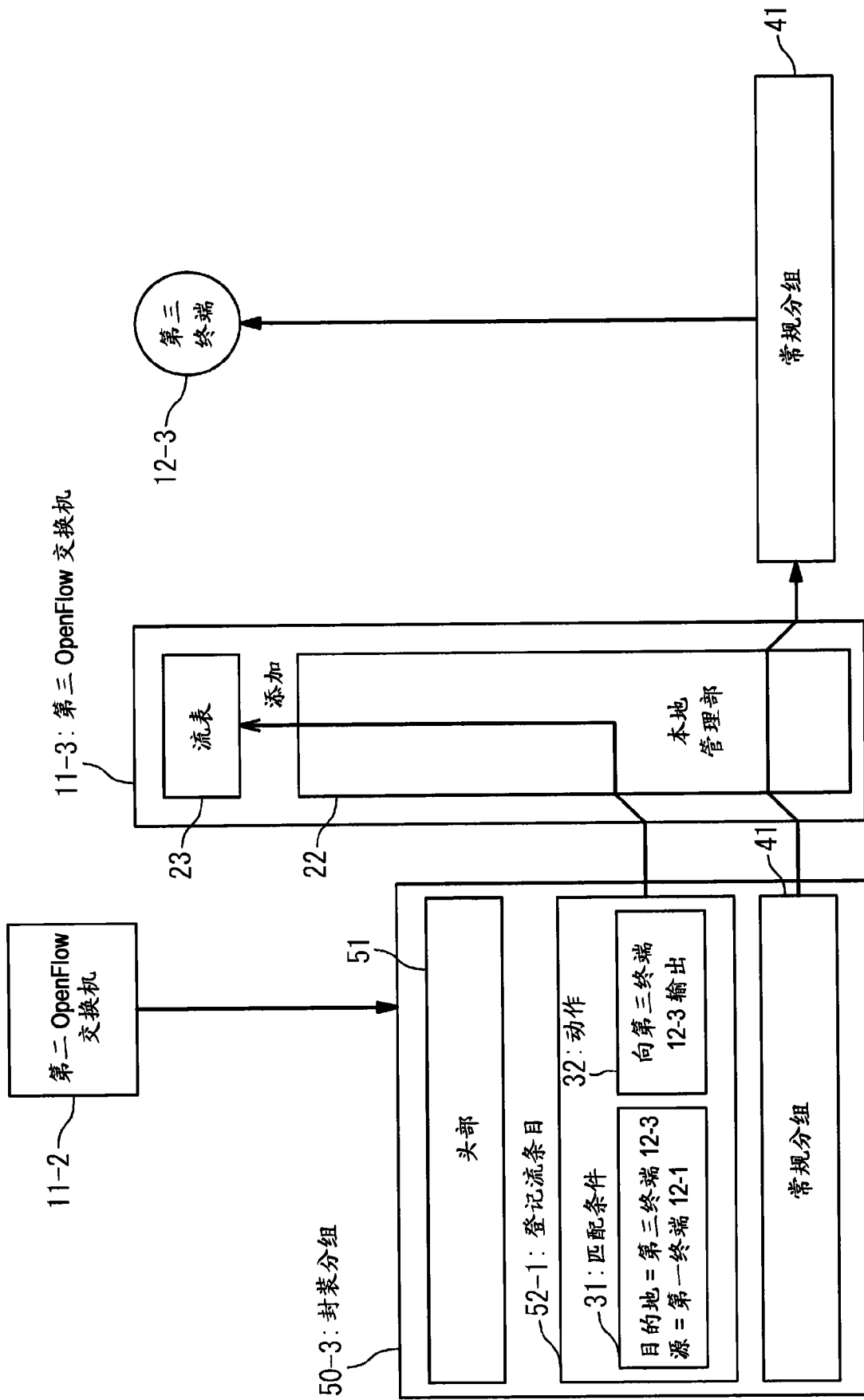


图 13

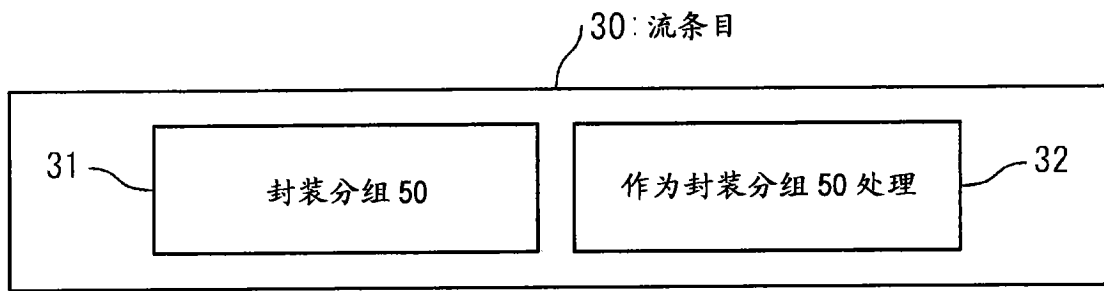


图 14

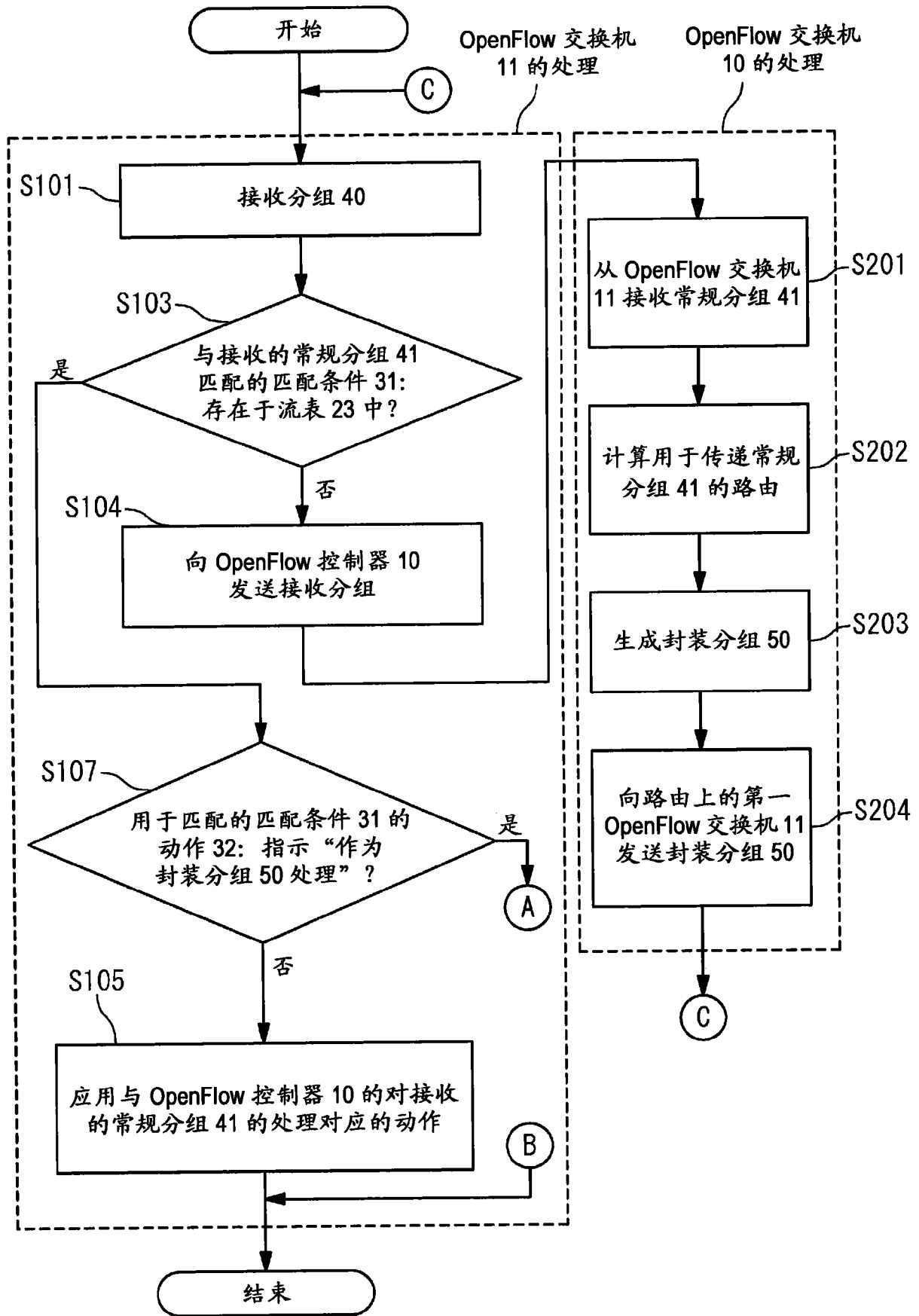


图 15