



(12) 发明专利

(10) 授权公告号 CN 102714628 B

(45) 授权公告日 2015. 11. 25

(21) 申请号 201180005536. 7

代理人 宋鹤

(22) 申请日 2011. 01. 04

(51) Int. Cl.

(30) 优先权数据

H04L 12/70(2013. 01)

2010-000740 2010. 01. 05 JP

H04L 1/00(2006. 01)

(85) PCT国际申请进入国家阶段日

(56) 对比文件

2012. 07. 05

CN 101415023 A , 2009. 04. 22, 全文 .

(86) PCT国际申请的申请数据

EP 1653686 A1 , 2006. 05. 03, 全文 .

PCT/JP2011/050009 2011. 01. 04

US 2005111356 A1 , 2005. 05. 26, 说明书

(87) PCT国际申请的公布数据

0053 段、0056-0058 段、0062 段及图 1-4.

W02011/083780 JA 2011. 07. 14

US 2007211623 A1 , 2007. 09. 13, 全文 .

(73) 专利权人 日本电气株式会社

审查员 裴广坤

地址 日本东京都

(72) 发明人 大和纯一 浜崇之 稗田谕士

内田智士

(74) 专利代理机构 北京东方亿思知识产权代理
有限责任公司 11258

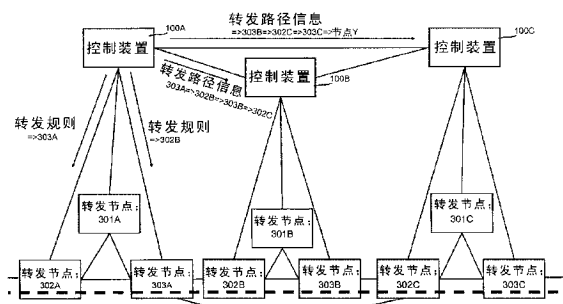
权利要求书3页 说明书17页 附图13页

(54) 发明名称

通信系统、控制装置、处理规则设置方法和分组传输方法

(57) 摘要

本发明提供了一种结构,其中即使当存在许多转发节点时也可以实现某一或更高层次的性能。一种通信系统包括多个控制服务器,并且还包
括多个转发节点,每个转发节点包括分组处理单元,用于基于预先建立的处理规则或者从与本地装置相关联的控制装置接受的处理规则来处理接收到的分组。多个控制装置中的第一控制装置响应于给定情形参考每个转发节点和相应控制装置之间的对应关系,并向另一控制装置发送至少包括与本地装置相关联的转发节点的一条转发路径信息,另一控制装置与接收到的分组的转发路径上、不与本地装置相关联的一个转发节点相关联。接收到转发路径信息的第二控制服务器基于转发路径信息创建并传输处理规则到位于转发路径上的与本地装置相关联的那些转发节点。



1. 一种通信系统,包括多个转发节点和多个控制装置,其中

所述多个转发节点中的每一个均包括分组处理单元,该分组处理单元基于预先设置的
处理规则或从与自身转发节点相关联的控制装置接收到的处理规则对接收到的分组进行
处理,以及

所述多个控制装置包括第一控制装置,所述第一控制装置包括:

对应关系管理单元,其存储所述多个转发节点中的每一个与所述多个控制装置中的每
一个之间的对应关系,其中所述对应关系识别将由所述第一控制装置控制的至少一个第一
转发节点和将由第二控制装置控制的至少一个第二转发节点;以及

路径传输单元,其响应于一个预先设定的触发,参考所述多个转发节点中的每一个与
所述多个控制装置中的每一个之间的对应关系,将转发路径信息中至少包含与第一控制装
置相关联的转发节点的转发路径信息传输到第二控制装置,第二控制装置与接收到的分组
的转发路径上的转发节点中和所述第一控制装置不相关联的一个转发节点相关联。

2. 根据权利要求 1 所述的通信系统,其中多个控制装置中的每一个还包括:

数据库,其存储一个描述多个转发节点之间的连接关系的网络拓扑结构,且所述多个
转发节点至少包括与自身控制装置关联的转发节点;

处理规则生成单元,其基于从路径传输单元接收到的转发路径信息,参考网络拓扑结
构,为与自身控制装置相关联的转发节点中位于接收到的分组的转发路径上的转发节点生
成处理规则;以及

传输单元,其将生成的处理规则传输到与自身控制装置相关联的转发节点中位于接收
到的分组的转发路径上的转发节点。

3. 根据权利要求 1 所述的通信系统,其中包括所述路径传输单元的所述第一控制装置
还包括:

数据库,其存储网络拓扑结构,该网络拓扑结构描述了至少包括与自身第一控制装置
相关联的转发节点的多个转发节点之间的连接关系;以及

路径计算单元,当接收到与所述自身第一控制装置连接的转发节点发来的传输有关新
分组的处理规则的请求时,其参考网络拓扑结构并计算转发路径。

4. 根据权利要求 3 所述的通信系统,其中

以包括所述路径计算单元的所述第一控制装置为起点,与所述转发路径上的转发节点
相关联的控制装置与处理规则的传输处理独立地转发新分组,并且

与位于转发路径终点的转发节点相关联的控制装置将新分组传输到位于所述终点的
转发节点,并指示所述位于终点的转发节点从指定端口输出新分组。

5. 根据权利要求 1 至 4 中任意一项所述的通信系统,

所述转发路径信息包括传输给和自身第一控制装置不相关联的转发节点的处理规则,
并且

接收转发路径信息的第二控制装置将包括在转发路径信息内的处理规则传输至位于
接收到的分组的转发路径上的转发节点中、与自身第二控制装置相关联的转发节点。

6. 根据权利要求 1 至 4 中任意一项所述的通信系统,其中所述转发路径是针对任意选
择的起点节点和终点节点的组合预先计算的转发路径。

7. 根据权利要求 3 或 4 所述的通信系统,其中所述转发路径被缓存一预定的时间段,且

当相同起点节点和终点节点的组合被缓存时,省略转发路径的计算。

8. 根据权利要求 2 至 4 中任意一项所述的通信系统,其中:

所述多个控制装置的每一个均包括:所述路径传输单元和路径计算单元,当接收到与自身控制装置连接的转发节点发来的传输有关新分组的处理规则的请求时,该路径计算单元参考所述网络拓扑结构并计算转发路径;

所述第一控制装置的路径传输单元向第二控制装置传输新分组和起点而非转发路径信息,第二控制装置与一转发节点相关联,该转发节点从连接至自身第一控制装置的位于转发路径上的转发节点接收分组,该起点用于计算与第二控制装置相关联的转发节点的转发路径;以及

接收到起点和新分组的所述第二控制装置基于所述起点和新分组重新计算转发路径,并生成和传输处理规则。

9. 根据权利要求 1 至 4 中任意一项所述的通信系统,其中:

所述第一控制装置包括数据同步单元,其获取与自身第一控制装置连接的多个转发节点的信息,并基于获取到的数据至少更新多个转发节点中的每一个的连接关系,以及多个转发节点中的每一个与多个控制装置中的每一个之间的连接关系中的一个。

10. 一种通信系统,包括多个转发节点和多个控制装置,其中:

多个转发节点中的每一个均包括分组处理单元,分组处理单元基于预先设置的处理规则或从与自身转发节点相关联的控制装置接收的处理规则对接收到的分组进行处理;

多个控制装置中的每一个均包括对应关系管理单元,对应关系管理单元存储多个转发节点中的每一个与多个控制装置中的每一个之间的对应关系,其中所述对应关系识别将由第一控制装置控制的至少一个第一转发节点和将由第二控制装置控制的至少一个第二转发节点;

当从某个转发节点转发接收到的分组时,与转发路径上的任一或多个转发节点相关联的控制装置参考多个转发节点中的每一个与多个控制装置中的每一个之间的对应关系,来依次传输接收的分组;并且

与位于转发路径终点的转发节点相关联的控制装置将新接收到的分组传输到位于所述终点的转发节点,并指示该位于所述终点的转发节点从指定端口输出所述新接收到的分组。

11. 一种第一控制装置,其连接至多个转发节点,

所述多个转发节点中的每一个均包括分组处理单元,其基于预先设置的处理规则或从与转发节点相关联的控制装置接收的处理规则对接收到的分组进行处理,其中

所述第一控制装置包括:

存储包括多个转发节点的转发节点组中的每一个与多个控制装置中的每一个之间的对应关系,其中所述对应关系识别将由自身第一控制装置控制的至少一个第一转发节点和将由第二控制装置控制的至少一个第二转发节点;以及

路径传输单元,响应于预先设定的触发,其参考所述多个转发节点中的每一个与多个控制装置中的每一个之间的对应关系,将转发路径信息中至少包含与第一控制装置相关联的转发节点的转发路径信息传输到第二控制装置,第二控制装置与位于接收到的分组的转发路径上的转发节点中和所述第一控制装置不相关联的一个转发节点相关联。

12. 一种通信系统中的处理规则设置方法,所述通信系统包括多个控制装置和多个转发节点,每个转发节点均包括分组处理单元,该分组处理单元基于预先设置的处理规则或从与自身转发节点相关联的控制装置接收到的处理规则对分组进行处理,该处理规则设置方法包括:

通过多个控制装置中的第一控制装置,响应于预先设置的触发,参考多个转发节点中的每一个与多个控制装置中的每一个之间的对应关系,其中所述对应关系识别将由所述第一控制装置控制的至少一个第一转发节点和将由第二控制装置控制的至少一个第二转发节点,将转发路径信息中至少包含与自身第一控制装置相关联的转发节点的转发路径信息传输到第二控制装置,第二控制装置与接收到的分组的转发路径上的转发节点中和第一控制装置不相关联的一个转发节点相关联;以及

通过接收转发路径信息的第二控制装置,通过参考网络拓扑结构生成处理规则并将处理规则传输到与第二控制装置相关联的转发节点中位于接收到的分组转发路径上的转发节点,该网络拓扑结构描述了转发节点间的连接关系,所述转发节点至少包含与第二控制装置相关联的转发节点。

13. 一种通信方法,所述方法是在包括多个控制装置和多个转发节点的通信系统中传输分组的方法,所述多个转发节点中的每一个均包括分组处理单元,该分组处理单元基于预先设置的处理规则或从与自身转发节点相关联的控制装置接收到的处理规则对接收到的分组进行处理,该通信方法包括:

通过与转发路径上的转发节点相关联的控制装置,在从某个转发节点转发接收到的分组时,参考多个转发节点中的每一个与多个控制装置中的每一个之间的对应关系,来依次传输接收到的分组,其中所述对应关系识别将由第一控制装置控制的至少一个第一转发节点和将由第二控制装置控制的至少一个第二转发节点,以及

通过与转发路径终点的转发节点相关联的控制装置,将接收的分组传输到位于所述终点的转发节点,并指示所述位于所述终点的转发节点从一个指定端口处输出接收的分组。

通信系统、控制装置、处理规则设置方法和分组传输方法

技术领域

[0001] (相关申请的参考)

[0002] 本申请基于下述专利申请并要求其优先权并通过引用将该申请公开整体结合于此:2010年1月5日提交的日本专利申请 No. 2010-000740。

[0003] 本发明涉及一种通信系统、控制装置、处理规则设置方法、分组传输方法和程序,具体而言涉及通过设置在网络中的转发节点转发分组来实现通信的一种通信系统、控制装置、处理规则设置方法、分组传输方法和程序。

背景技术

[0004] 近年来,人们已提出了一项被称为“OpenFlow”(开放流)的技术(参见非专利文献1)。在 OpenFlow 技术中,通信被看成一个点到点的流,其中路径控制、故障恢复、负载均衡以及优化均在流单元中完成。充当转发节点的 OpenFlow 交换机配有一个安全通道,该安全通道用于与被视为控制装置的 OpenFlow 控制器进行通信。OpenFlow 交换机基于流表运行,可通过 OpenFlow 控制器的指示对该流表适当地添加或改写。在流表中,对于每一个流(参见图10)具有用于与分组头部进行校对的多组规则的定义、定义处理内容的动作(Actions)、以及流统计信息(Stats)。

[0005] 图23示出了非专利文献2中定义的动作名称和动作内容的一个示例。“OUTPUT”(输出)是一个向特定端口(接口)进行输出的动作。从 SET_VLAN_VID 到 SET_TP_DST 是校正分组头部字段的动作。

[0006] 例如,当 OpenFlow 交换机接收到分组时,就会在流表中搜索一个条目,该条目具有与接收到的分组头部信息相匹配的规则(流密钥)。作为搜索结果,在找到了与接收到的分组相匹配的条目的情况下,OpenFlow 交换机就对接收到的分组执行所关注的条目的动作字段中描述的处理内容。另一方面,作为搜索结果,在未找到与接收到的分组相匹配的条目的情况下,OpenFlow 交换机就通过安全通道将其接收到的分组转发到 OpenFlow 控制器,并基于接收到的分组的来源地和目的地请求确定分组路径,然后接收实现上述目的的流条目,并更新流表。通过这种方式,OpenFlow 交换机将流表中包含的表条目用作处理规则进行分组转发。

[0007] 【非专利文献1】

[0008] Nick McKeown 和 7 个其他人的“OpenFlow:Enabling Innovation in Campus Networks,” [online],[search conducted December 14,2009](2009年12月14日进行检索)

[0009] 因特网 URL:

[0010] <http://www.openflowswitch.org/documents/openflow-wp-latest.pdf>

[0011] 【非专利文献2】

[0012] “OpenFlow:Switch Specification” Version 0.9.0.(Wire Protocol 0x98)(版本0.9.0,有线协议0x98), [search conducted December 14,2009](2009年12月14

日进行检索)

[0013] 因特网 URL :

[0014] <http://www.openflowswitch.org/documents/openflow-spec-v0.9.0.pdf>

发明内容

[0015] 本发明要解决的问题

[0016] 上述非专利文献 1 和 2 的公开通过引用整体结合于此。本发明给出如下分析。然而,在多个 OpenFlow 交换机与一个网络连接的情况下,就会出现一个问题:由于在一个 OpenFlow 控制器内进行的路径计算以及在一个基于计算结果的流表中条目分配的处理量增加,导致从接收到 OpenFlow 交换机发来的创建路径的请求到给出响应的时间变长。

[0017] 对于这一点,非专利文献 1 对其图 2 的解释描述为“所有的流表都由同一个控制器管理”,其仅仅描述了在 OpenFlow 协议中,一个特定的 OpenFlow 交换机可以由两个或更多个控制器控制以便改进性能和稳健性。另外,在非专利文献 2 的 4.3 节“Connection Interruption”中,仅仅描述了在某个 OpenFlow 交换机错过一个控制器的情况下会尝试连接到备用控制器。

[0018] 针对上述情况提出了本发明,其目的之一就是要提供一个配置来确保在有多个转发节点存在的情况下(正如上述存在多个 OpenFlow 交换机的例子),也能保证其性能不低于预定的水平。

[0019] 解决问题的手段

[0020] 根据本发明的第一方面,提供了一种通信系统,其包括多个转发节点和多个控制装置,其中

[0021] 多个转发节点中的每一个均包括一个分组处理单元,该分组处理单元基于预先设置的处理规则或从与自身转发节点相关联的控制装置接收到的处理规则对接收到的分组进行处理,以及

[0022] 所述多个控制装置包括第一控制装置,该第一控制装置包括:

[0023] 对应关系管理单元,其存储所述多个转发节点中的每一个与所述多个控制装置中的每一个之间的对应关系;以及

[0024] 路径传输单元,响应于预先设定的触发,其参考所述多个转发节点中的每一个与多数多个控制装置中的每一个之间的对应关系,将转发路径信息中包含至少一个与自身第一控制装置相关联的转发节点的转发路径信息传输到第二控制装置,第二控制装置与接收到的分组的转发路径上的转发节点中和自身第一控制装置不相关联的一个转发节点相关联。

[0025] 根据本发明的第二方面,提供了与多个转发节点连接的第一控制装置,

[0026] 多个转发节点中的每一个均包括分组处理单元,该分组处理单元基于预先设置的处理规则或从与转发节点关联的控制装置接收到的处理规则对接收到的分组进行处理,其中

[0027] 第一控制装置包括:

[0028] 对应关系管理单元,其存储包括多个转发节点的转发节点组中的每一个与多个控制装置中的每一个之间的对应关系;以及

[0029] 路径传输单元, 响应于预先设定的触发, 其参考所述多个转发节点中的每一个与多数多个控制装置中的每一个之间的对应关系, 将转发路径信息中包含至少一个与自身第一控制装置相关联的转发节点的转发路径信息传输到第二控制装置, 第二控制装置与位于接收到的分组的转发路径上的转发节点中和第一控制装置不相关联的一个转发节点相关联。

[0030] 根据本发明的第三方面, 提供了第二控制装置, 包括:

[0031] 数据库, 该数据库存储着描述转发节点间连接关系的网络拓扑结构, 所述转发节点包括至少一个与自身第二控制装置相关联的转发节点;

[0032] 处理规则生成单元, 其基于从第一控制单元的路径传输单元接收到的转发路径信息, 参考所述网络拓扑结构, 并为与自身第二控制装置相关联的转发节点中位于接收到的分组的转发路径上的转发节点生成处理规则; 以及

[0033] 传输单元, 将生成的处理规则传输至与第二控制装置的多个转发节点中位于接收到的分组的转发路径上的转发节点, 其中

[0034] 基于从上述第一控制装置接收到的转发路径信息, 处理规则被生成并被传输至多个与所述自身第二控制装置连接的转发节点中位于转发路径上的转发节点。

[0035] 根据本发明的第四方面, 提供了一种通信系统中的处理规则设置方法, 该通信系统包括多个控制装置和多个转发节点, 每个转发节点包括一个分组处理单元, 该分组处理单元基于预先设置的处理规则或从与自身转发节点相关联的控制装置接收到的处理规则对分组进行处理, 其中:

[0036] 通过多个控制装置中的第一控制装置, 响应于预先设定的触发, 参考多个转发节点中的每一个与多个控制装置中的每一个之间的对应关系, 将转发路径信息中包含至少一个与自身第一控制装置相关联的转发节点的转发路径信息传输到第二控制装置, 第二控制装置与接收到的分组的转发路径上的转发节点中和第一控制装置不相关联的一个转发节点相关联; 以及

[0037] 通过接收转发路径信息的第二控制装置, 通过参考网络拓扑结构生成处理规则, 该网络拓扑结构描述了转发节点间的连接关系, 所述转发节点包含至少一个与第二控制装置相关联的转发节点; 将处理规则传输到与自身第二控制装置相关联的转发节点中、位于接收到的分组转发路径上的转发节点。此方法与特定的装置相关, 这些设备被称为上文所述的转发节点和控制装置。

[0038] 根据本发明的第五方面, 提供了一种在用于构成上述第一和第二控制装置的计算机上执行的程序。需要指出的是, 这个程序可以被记录到计算机可读的存储介质上。也就是说, 本发明可以体现为计算机程序产品。

[0039] 根据本发明的第六方面, 提供了一种包括多个转发节点和多个控制装置的通信系统, 其中

[0040] 多个转发节点中的每一个均包括分组处理单元, 该分组处理单元基于预先设置的处理规则或从与自身转发节点相关联的控制装置接收到的处理规则对接收到的分组进行处理,

[0041] 所述多个控制装置的每一个均包括对应关系管理单元, 其存储所述多个转发节点中的每一个与所述多个控制装置中的每一个之间的对应关系;

[0042] 当从某个转发节点转发接收到的分组时,与转发路径上的转发节点相关联的控制装置参考多个转发节点中的每一个与多个控制装置的每一个之间的对应关系来传输接收到的分组;以及

[0043] 与位于转发路径终点的转发节点相关联的控制装置将接收的分组传输到位于终点的转发节点,并指示所述位于终点的转发节点从一个指定端口输出接收的分组。

[0044] 根据本发明的第七方面,提供了一种通信方法,该通信方法是在一个包括多个控制装置和多个转发节点的通信系统中传输分组的方法,所述多个转发节点中的每一个均包括一个分组处理单元,该分组处理单元基于预先设置的处理规则或从与自身转发节点相关联的控制装置接收到的处理规则对接收到的分组进行处理,该通信方法包括:

[0045] 通过与转发路径上的转发节点相关联的控制装置,在从一个特定的转发节点转发接收到的分组时,参考多个转发节点中的每一个与多个控制装置中的每一个之间的对应关系,来依次传输接收到的分组,以及

[0046] 通过与转发路径终点的转发节点相关联的控制装置,将接收的分组传输到位于终点的转发节点,并指示所述位于终点的转发节点从一个指定端口输出接收的分组。此方法与特定的工具相关,这些工具被称为上文所述的转发节点和控制装置。

[0047] 根据本发明,可以在大量(多个)OpenFlow 交换机存在的情况下确保性能不低于预定水平。这主要依赖于本发明使用的一个配置,在该配置中使用了多个控制装置,处理工作被分配给了多个控制装置。

附图说明

[0048] 图 1 是概括描述本发明的图;

[0049] 图 2 是描述了图 1 所示控制装置(第一控制装置)的大概配置的图;

[0050] 图 3 是描述了图 1 所示控制装置(第二控制装置)的大概配置的图;

[0051] 图 4 是另一个概括描述本发明的图;

[0052] 图 5 是示出了本发明的第一示例性实施例配置的图;

[0053] 图 6 是示出了本发明的第一示例性实施例中的 OpenFlow 控制器配置的图;

[0054] 图 7 是表示图 5 中的节点与 OpenFlow 交换机的连接关系的表格;

[0055] 图 8 是表示图 5 中的 OpenFlow 交换机与 OpenFlow 控制器之间对应关系的表格;

[0056] 图 9 是示出了本发明第一示例性实施例的 OpenFlow 交换机配置的图;

[0057] 图 10 是表示本发明第一示例性实施例的 OpenFlow 交换机的流表中记录的处理规则(流条目)的图;

[0058] 图 11 是图 5 的 OpenFlow 交换机的流表中记录的处理规则(流条目)的一个具体示例;

[0059] 图 12 是说明了本发明第一示例性实施例的操作的顺序图;

[0060] 图 13 是在图 5 中记录根据图 12 的顺序进行的操作的图;

[0061] 图 14 是表示本发明第一示例性实施例的操作的第二顺序图;

[0062] 图 15 是图 5 中记录的根据图 14 的顺序进行的操作的图;

[0063] 图 16 是示出了本发明第二示例性实施例中的 OpenFlow 控制器结构的图;

[0064] 图 17 是示出了本发明第三示例性实施例中的 OpenFlow 控制器配置的图;

- [0065] 图 18 是示出了本发明第三示例性实施例的操作的顺序图；
- [0066] 图 19 是示出了本发明第四示例性实施例的 OpenFlow 控制器配置的图；
- [0067] 图 20 是说明了本发明第四示例性实施例的操作的顺序图；
- [0068] 图 21 是描述了本发明第五示例性实施例的图；
- [0069] 图 22 说明了本发明第五示例性实施例的操作的顺序图；以及
- [0070] 图 23 是示出了动作名称和动作内容的示例的图。

[0071] 优选实施方式

[0072] 首先,参考图 1 至图 4 对本发明进行概括性描述。如图 1 所示,根据本发明的一种通信系统包括:多个控制装置 100A 至 100C,所述多个控制装置通过参考描述转发节点的连接关系的网络拓扑结构,生成一个确定接收到的分组的转发目的地的处理规则并将其传输至连接到每个控制装置的转发节点中位于计算的转发路径上的转发节点;以及多个转发节点 301A 至 303A、301B 至 303B、和 301C 至 303C,其中每个转发节点均与任意一个控制装置 100A 至 100C 连接并包括分组处理单元,该分组处理单元基于从与自身转发节点相关联的控制装置接收的处理规则对接收到的分组进行处理。应当注意的是,附图的参考标记仅为辅助理解,而不是要将本发明局限于附图中描述的模式。

[0073] 如图 2 所示,多个控制装置 100A 至 100C 中至少一个装置(例如控制装置 100A)包括对应关系管理单元 16A 和路径传输单元 13,对应关系管理单元 16A 存储转发节点 301A 至 303A、301B 至 303B、以及 301C 至 303C 与控制装置 100A 至 100C 之间的对应关系,路径传输单元 13 响应于一个预定的触发(例如接收来自转发节点的分组,对于该分组而言,处理规则未知),参考多个转发节点中的每一个与多个控制装置中的每一个之间的对应关系,并将转发路径信息中包括与至少一个自身控制装置相关联的至少一个转发节点的转发路径信息传输至其他控制装置(例如控制装置 100B 或 100C),该其他控制装置与接收到的分组的转发路径上的转发节点中与自身控制装置不相关联的转发节点相关联。

[0074] 在从 100A 到 100C 的多个控制装置中,如图 3 所示,处于转发路径信息接收侧的控制装置(例如,控制装置 100B 或 100C)包括:数据库(拓扑数据库)15,其存储一个网络拓扑结构,该网络拓扑描述各个转发节点之间的连接关系,所述转发节点包括至少一个与自身控制装置相关联的转发节点;处理规则生成单元 18A,其基于接收到的转发路径信息,参考网络拓扑结构,为与控制装置相关联的转发节点中位于接收到的分组的转发路径上的转发节点生成处理规则;传输单元 17A,其将生成的处理规则传输到与自身控制装置相关联的多个转发节点中位于接收到的分组的转发路径上的转发节点。

[0075] 这里,例如,控制装置 100A 响应于与转发节点 302A 连接的特定客户节点的请求,计算从转发节点 302A、303A、302B、303B、302C 和 303C 转发的分组的转发路径,如图 1 中的虚线所示。在这种情况下,控制装置 100A 为与自身控制装置 100A 连接的转发节点 302A 和 303A 生成并传输实现所述转发路径的处理规则。另外,控制装置 100A 参考多个转发节点中的每一个与多个控制装置中的每一个之间的对应关系,并将用于生成处理规则的转发路径传输至控制装置 100B 和 100C,该控制装置 100B 和 100C 与转发节点 301B-303B 以及 301C-303C 相关联,上述转发节点不与自身控制装置 100A 连接。接收到转发路径信息的控制装置 100B 和 100C 基于该转发路径信息生成处理规则并将其传输到与自身控制装置(100B、100C)连接的转发节点中位于转发路径上的转发节点 302B、303B、302C 和 303C。

[0076] 综上所述,控制装置 100B 和 100C 使用从控制装置 100A 接收的转发路径,且通过生成和传输处理规则,为转发路径上的所有转发节点设置一个处理规则。

[0077] 同样地,如图 4 所示,假设响应另一个客户节点的请求,控制装置 100A 计算由如图 4 中虚线所示的转发节点 302A、303A、302C 和 303C 转发的分组的转发路径。在这种情况下,控制装置 100A 生成实现所述转发路径的处理规则并将其传输至与控制装置连接的转发节点 302A 和 303A。此外,控制装置 100A 参考上述各个转发节点与各个控制装置之间的对应关系,并将转发路径信息传输到控制装置 100C 上,控制装置 100C 和与自身控制装置 100A 不相关联的转发节点 301C 和 303C 相关联。接收到上述转发路径信息的控制装置 100C 基于所述转发路径信息生成处理规则并将其传输到与自身控制装置 100C 连接的转发节点中位于转发路径上的转发节点 302C 和 303C。

[0078] 如上所述,在已经生成一个未通过控制装置 100B 的转发路径的情况下,省略了向控制装置 100B 传输转发路径信息。在这种情况下,所述控制装置 100C 使用从控制装置 100A 接收到的转发路径生成并传输处理规则,从而可为所述转发路径上的所有节点设置处理规则。

[0079] 本发明可以采用下述模式:

[0080] (模式 1)

[0081] 根据上述第一方面的通信系统。

[0082] (模式 2)

[0083] 在上述通信系统中,多个控制装置中的每一个还可包括:

[0084] 数据库,用于存储网络拓扑结构,该网络拓扑结构描述了包括至少一个与自身控制装置相关联的转发节点的多个转发节点之间的连接关系;

[0085] 处理规则生成单元,其基于从路径传输单元接收到的转发路径信息,参考网络拓扑结构,为与自身控制装置相关联的转发节点中位于接收到的分组的转发路径上的转发节点生成处理规则;以及

[0086] 传输单元,其将生成的处理规则传输到与自身控制装置相关联的转发节点中位于接收到的分组的转发路径上的转发节点。

[0087] (模式 3)

[0088] 在上述通信系统中,包括路径传输单元的第一控制装置还可包括:

[0089] 存储网络拓扑结构的数据库,该网络拓扑结构描述了包括至少一个与自身第一控制装置相关联的转发节点的多个转发节点之间的连接关系;以及

[0090] 路径计算单元,当接收到与第一控制装置连接的转发节点发来的传输有关新分组的处理规则的请求时,其参考网络拓扑结构并计算转发路径。

[0091] (模式 4)

[0092] 在该通信系统中,以包括路径计算单元的第一控制装置为起点,与位于转发路径上的转发节点相关联的控制装置可以转发新分组,而不依赖处理规则的转发处理,并且

[0093] 与位于转发路径终点的转发节点相关联的控制装置可以将新分组传输到位于终点的转发节点上,并指示位于终点的转发节点从指定端口输出所述新分组。

[0094] (模式 5)

[0095] 在上述通信系统中,所述转发路径信息可以包括传输至与自身第一控制装置不相

关联的转发节点的处理规则,并且

[0096] 接收到转发路径信息的第二控制装置可以将包括在转发路径信息中的处理规则传输至位于接收到的分组的转发路径上的转发节点中、与自身第二控制装置相关联的转发节点。

[0097] (模式 6)

[0098] 在上述通信系统中,转发路径可以是针对任意选择的起点节点和终点节点的组合预先计算出的转发路径。

[0099] (模式 7)

[0100] 在上述通信系统中,上述转发路径可以被缓存一预定时间段,且当相同起点节点和终点节点的组合被缓存时,可以省略转发路径的计算。

[0101] (模式 8)

[0102] 在上述通信系统中,所述多个控制装置中的每一个可以包括:所述路径传输单元和路径计算单元,当接收到与自身控制装置连接的转发节点发来的传输有关新分组的处理规则的请求时,所述路径计算单元参考所述网络拓扑结构并计算转发路径,

[0103] 第一控制装置的路径传输单元可以向第二控制装置传输新分组和起点而非转发路径信息,第二控制装置与一转发节点相关联,该转发节点从与自身第一控制装置连接的位于转发路径上的转发节点接收分组,该起点用于计算与第二控制装置相关联的转发节点的转发路径;以及

[0104] 接收到所述起点和新分组的第二控制装置可以根据所述起点和新分组重新计算转发路径,并生成和传输处理规则。

[0105] (模式 9)

[0106] 在通信系统中,控制装置可以包括数据同步单元,用于获取与自身控制装置连接的多个转发节点的信息,并基于所获得的数据更新每个转发节点中的每一个的至少一个连接关系以及多个转发节点中的每一个与多个控制装置的每一个之间的对应关系。

[0107] (模式 10)

[0108] 根据上述第二方面的第一控制装置。

[0109] (模式 11)

[0110] 根据所述第三方面的第二控制装置。

[0111] (模式 12)

[0112] 根据所述第四方面的处理规则设置方法。

[0113] (模式 13)

[0114] 根据所述第五方面的程序。

[0115] (模式 14)

[0116] 根据所述第六方面的通信系统。

[0117] (模式 15)

[0118] 根据所述第七方面的通信方法。

[0119] 应该注意的是,上述第二到第七方面中提到的第一和第二控制装置、处理规则设置方法、程序、通信系统和通信方法对应的元件和步骤与方式 1 中的通信系统类似,可以扩展到模式 2 到模式 9 的内容。

[0120] < 第一示例性实施例 >

[0121] 接下来,将参照附图对本发明的第一个实施例进行详细描述。图 5 是示出了本发明第一示例性实施例的配置的图。参见图 5,示出了四个 OpenFlow 交换机(以下简称“OFS”)301A、302A、301B、302B 和两个 OpenFlow 控制器(以下简称为“OFC”)100A 和 100B。应当注意的是,图 5 中的配置是说明性的,其只是为了简单地描述本发明。对 OFS 和 OFC 的数量没有什么限制,且没有必要使得与每个对应 OFC 连接的 OFS 数量相同。另外,节点 A 和节点 D 是用于传输和接收分组的装置,例如用户终端、各种类型的服务器等。

[0122] OFS 301A、302A、301B 和 302B 对应于上述转发节点,包含由 OFC100A 和 100B 在各自的流表中设置的流条目(规则+动作),并执行具有与接收到的分组匹配的规则的流条目中的动作。

[0123] 在上述 OFS 中,OFS 301A 和 302A 与 OFC 100A 连接,并根据 OFC 100A 的控制运行。同样地,在这些 OFS 中,OFS 301B 和 302B 与 OFC 100B 连接,并根据 OFC 100B 发来的控制而操作。具体地说,当未能在流表中找到与接收到的分组匹配的规则时,OFS 301A 和 302A 就请求 OFC 100A 为所述接收到的分组生成一个转发路径,并传输(设置)一个流表(规则+动作)以实现所述转发路径。同样地,当未能在流表中找到与接收到的分组匹配的规则时,OFS 301B 和 302B 就请求 OFC 100B 为所述接收到的分组生成转发路径,并传输(设置)流表(规则+动作)以实现所述转发路径。这些类型的 OFS 配置可以通过向相应供应商的交换机添加上述功能来实现,如非专利文献 1 和 2 中所述。

[0124] OFC 100A 和 100B 对应于上述控制装置,并通过将流条目(规则+动作)传输(设置)到分别与其连接的 OFS 301A、302A、301B 和 302B 而控制这些 OFS。此外,本实施例中的 OFC 100A 和 100B 相连,在下文中将描述,由一个 OFC 生成的转发路径信息可以传输至另一个 OFC。

[0125] 图 6 是示出了第一示例性实施例中 OFC 的配置的图。参见图 6,OFC 100 包括分组转发单元 11、路径计算单元 12、路径传输单元 13、OFC 间通信单元 14、拓扑数据库(DB)15、OFS-OFC 对应关系管理单元 16、OFS 通信单元 17、流生成单元 18 和路径接受单元 19。

[0126] 当 OFS 中不存在相关的流条目,且接收到处理规则(流条目)传输(设置)请求时,分组转发单元 11 就通过 OFC 间通信单元 14 对另外一个 OFC 执行传输分组的操作,该分组附有用于传输(设置)处理规则(流条目)的请求。

[0127] 基于处理规则(流条目)传输(设置)请求的来源的交换机 ID 和 OFS 通信单元 17 输出的分组,路径计算单元 12 参考描述了网络拓扑数据库 15 中记录的转发节点之间的连接关系的网络拓扑结构,并计算出连续流中的转发路径,该连续流包括从 OFS 转发来的分组。此路径计算可以利用例如 Dijkstra 最短路径算法。此外,在路径计算时,也可以考虑网络中流量状况。

[0128] 路径传输单元 13 将接收到的分组以及在路径计算单元 12 中计算出的转发路径输出到 OFS-OFC 对应关系管理单元 16。另外,路径传输单元 13 参考 OFS-OFC 对应关系管理单元 16 中记录的 OFS 和与该 OFS 连接的 OFC 的对应关系,通过 OFC 间通信单元 14 执行将转发路径信息传输至与路径计算单元 12 计算出的转发路径上的特定 OFS 连接的 OFC(与该转发路径上的特定 OFS 相关联的 OFC)的操作,所述路径转发信息包含接收到的分组和路径计算单元 12 计算出的转发路径。应当注意的是,对于这里用到的转发路径信息,只要另一个

OFC 的流生成单元 18 可以生成处理规则（流条目），任何格式都是可以的，并且可以是这样的内容，其中另一个 OFC 制作一个将在相应的 OFS 中进行设置的处理规则（流条目）的列表。

[0129] OFC 间通信单元 14 与另一个 OFC 的 OFC 间通信单元 14 连接，并发出以及接收转发路径信息和分组，所述分组附有用于传输（设置）所关注的流条目的请求。

[0130] 拓扑数据库 15 记录描述包括 OFS 在内的各节点（各种类型的服务器和外部网络）之间的连接关系的网络拓扑结构。图 7 通过表格形式描述了图 5 中包括 OFC 在内的各个节点之间的连接关系。在图 7 中，“OFS ;302A”等实际上包含着 OFS 302A 的 IP 地址和 MAC 地址。应当指出的是，在此示例性实施例中，OFC 100A 和 100B 各自都具有一个拓扑数据库 15，但是 OFC 100A 和 100B 共享拓扑数据库 15 的配置也是可以的。

[0131] OFS-OFC 对应关系管理单元 16 记录 OFS 和与 OFS 连接的 OFC 之间的对应关系，并将路径传输单元 13 输出的转发路径和接收到的分组以及与 OFC 连接的 OFS 的信息输出至流生成单元 18。图 8 用表格的形式示出了图 5 中的多个 OFS 和多个 OFC 之间的连接关系。应当注意的是，在本示例性实施例中，OFS-OFC 对应关系管理单元 16 独立于拓扑数据库 15，但是把拓扑数据库 15 和 OFS-OFC 对应关系管理单元 16 合并起来也是可以的。例如，除了图 7 中示出的 OFS 连接，也可以提供描述连接的 OFC 的字段。

[0132] 由于没有与接收到的分组相对应的流条目，OFS 通信单元 17 将流条目传输（设置）请求连同从 OFS 传输的接收到的分组传输到路径计算单元 12。OFS 通信单元 17 将流生成单元 18 中生成的处理规则（流条目）传输至转发路径上与所述 OFC 连接的多个 OFS 中的一个特定的 OFS 中，并指示处理规则（流条目）的设置。

[0133] 流生成单元 18 基于从 OFS-OFC 对应关系管理单元 16 或路径接受单元 19 输出的接收到的分组和转发路径（信息），参考与 OFC 连接的 OFS 的信息以及拓扑数据库 15 的网络拓扑结构，并生成处理规则，所述处理规则将在与多个 OFC 连接的 OFS 中位于转发路径上的特定的 OFS（与 OFC 相关联的 OFS）中进行设置。

[0134] 路径接受单元 19 将包括从另一个 OFC 接收的分组的转发路径信息输出至流生成单元 18，并执行生成处理规则（流条目）的操作。

[0135] 应该注意的是，上述分组转发单元 11、路径计算单元 12、路径传输单元 13、OFC 间通信单元 14、拓扑数据库 15、OFS-OFC 对应关系管理单元 16、OFS 通信单元 17、流生成单元 18 和路径接受单元 19 可以通过存储装置或者设置在构成相应的 OFC 的计算机中的装置以及将存储装置用作硬件的计算机程序来实现。

[0136] 图 9 是示出了第一示例性实施例中 OFS 的配置。参见图 9，OFS 300 包括 OFC 通信单元 31、流设置单元 32、分组处理单元 33 和流表 34。另外，在 OFS 300 中，与 OFS 300 连接（与 OFS 300 相关联）的 OFC 是预先设置的。

[0137] OFC 通信单元 31 从 OFC 的 OFS 通信单元 17 接收处理规则（流条目），除了向流设置单元 32 输出外，在没有与从分组处理单元 33 接收到的分组相对应的流条目的情况下，请求 OFC 生成一个处理规则（流条目），用于计算所述接收到的分组的转发路径以及该转发路径的实现。

[0138] 流设置单元 32 执行一个操作以记录或更新流表 34 中通过 OFC 通信单元 31 从 OFC 传输的处理规则（流条目）的操作。

[0139] 分组处理单元 33 参考流表 34, 搜索一个包含与接收到的分组的头部信息相匹配的处理规则 (流条目), 并执行与接收到的分组相匹配的处理规则 (流条目) 的动作字段中描述的处理内容。

[0140] 流表 34 是一个表格, 包含通过 OFC 通信单元 31 从 OFC 传输的处理规则 (流条目)。图 10 是示出了流表 34 的字段配置的图。例如, 如果一个流具有如下特性: 在一个源 IP 地址字段 (IP SA) 中描述了节点 A 的 IP 地址, 在目标 IP 地址字段 (IP DA) 中描述了节点 D 的 IP 地址, 通过设置从传输接口 (端口) 到临近的 OFS 中沿着转发路径的 OFS 进行转发为动作 (图 23 中的 “OUTPUT”), 就实现了转发路径为从节点 A 到节点 D 的流的转发。否则, 通过设置图 23 中所示的不同类型的动作, 或将某个动作字段设置为空, 可以改变一个特定流的 MAC 地址或 IP 地址或丢弃某个特定流。

[0141] 图 11 示出了在 OFS 301A、302A、301B 和 302B 中设定处理规则 (流条目) 的示例。这里, 设定通过 OFS 301A、302A、301B 和 302B 在节点 A 和节点 D 之间转发分组的转发路径。例如, 在 OFS 301A 中被设定从节点 A 到节点 D, 以节点 D 作为正向的目的地 (接口 2 (连接到 OFS 302A; 参见图 7)) 转发分组的动作, 以及从节点 D 和节点 A 沿着 (接口 1 (连接到接口 A; 参考图 7)) 以节点 A 为反向的目的地转发分组的动作。同样地, 在 OFS 302A、301B 和 302B 中, 通过设置正向和反向的处理规则 (流条目), 在上述转发路径上进行节点 A 和节点 D 之间的分组转发。

[0142] (设置处理规则 (流条目) 的方法)

[0143] 接下来, 将参照附图详细描述此示例性实施例的操作。图 12 是表示本示例性实施例的操作的顺序图。在图 12 的示例中, 从与 OFS 301A 连接的节点 A 传输以节点 D 为目的地的分组, 节点 D 与 OFS 301A 的流表 34 中存储的所有处理规则 (流条目) 都不匹配。

[0144] 首先, 一旦接收到新分组 (步骤 S001), OFS 301A 就向与 OFS 301A 连接的 OFC 100A 发出请求, 请求传输 (设置) 包括所述新分组的处理规则 (流条目) (步骤 S002)。应当注意的是在步骤 S001 中, 在接收到的分组与 OFS 301A 的流表 34 中存储的某个处理规则 (流条目) 相匹配的情况下, 将不执行后续的处理, 而是根据节点处理规则执行分组转发处理, 正如图 12 最下面部分所示。

[0145] 被请求传输 (设定) 处理规则 (流条目) 的 OFC 100A 参考拓扑数据库 15 的网络拓扑结构, 并基于路径计算单元 12 中的请求源 OFS 301A 和新分组的目的地节点 D, 计算包括 OFS 301A 转发的分组的顺序流的转发路径 (步骤 S003)。这里, 假定计算除了转发路径, 根据该路径对来自节点 A 且以节点 D 为目的地的分组按照 OFS 301A、302A、301B 和 302B 的顺序进行转发。

[0146] 接着, OFC 100A 参考存储在 OFS-OFC 对应关系管理单元 16 中的 OFS 与 OFC 之间的对应关系, 并搜索和在计算出的转发路径上的 OFS 中未与 OFC 100A 连接的 OFS 相关联的 OFC (步骤 S004)。这里, 检索出的是与 OFS 301B 和 302B (它们未与图 5 中的 OFC 100A 连接) 连接的 OFC 100B。

[0147] 然后, OFC 100A 将转发路径信息 (包括在步骤 S003 计算出的转发路径和接收分组) 发送到上述检索出的 OFC 100B。

[0148] 此后, OFC 100A 参考拓扑数据库 15 的网络拓扑结构以及与自身 OFC 100A 连接的 OFS 的信息, 生成处理规则 (流条目), 所述规则将在与 OFC 100A 连接的 OFS 中位于转发路

径上的 OFS 中设置,并针对 OFS 301A 和 302A 执行传输(设置)(步骤 S006-1)。

[0149] 同样地,OFC 100B 也参考拓扑数据库 15 的网络拓扑结构以及与自身 OFC 100B 连接的 OFS 的信息,生成处理规则(流条目),所述规则将在与 OFC 100A 连接的 OFS 中位于转发路径上的 OFS 中设置,并针对 OFS 301B 和 302B 执行传输(设置)(步骤 S006-2)。

[0150] 此后,在 OFS 301A、302A、301B 和 302B 的每一个中,都执行处理以将接收到的处理规则记录或更新到流表 34 中(步骤 S007)。

[0151] 综上所述,在 OFS 301A、302A、301B 和 302B 的每一个中,都设置(参见图 11)与计算出的转发路径相对应的处理规则。同时,节点 A 和节点 B 之间的分组转发成为可能。

[0152] 应当注意的是,在上述步骤 S006-1 和 S006-2 中,OFC 在与自身 OFC 连接的 OFS 中位于转发路径上的 OFS 内同时传输(设置)处理规则(流条目)。但是,在这些 OFS 中,也可以设置为只在部分 OFS(例如,在转发路径上游侧的一个或几个 OFS)中传输(设置)处理规则。在这种情况下,每一个 OFC 都在从各自的 OFS 接收到传输(设置)处理规则的请求后传输(设置)处理规则(流条目)。

[0153] 图 13 是在图 5 的配置图中记录根据图 12 的顺序进行的操作的图。在接收到传输(设置)一个特定分组的处理规则的请求后,OFC 100A 执行转发路径的计算,并将计算结果设置为关于 OFS 301A 和 302A 的处理规则(流条目)。此外,通过 OFC 100A 将转发路径信息传输到 OFC 100B,对 OFS 301B 和 302B 的处理规则(流条目)进行设置。从图 13 可以明确看出,在本发明中没有特别限制 OFS/OFC 的数量,在转发路径上出现其他 OFS 的情况下,可以将转发路径信息传输到与该 OFS 相连的相关 OFC 上,并设置处理规则(流条目)。

[0154] 此外,图 13 中的 OFC 100A 和 100B 都不需要设置路径计算单元 12 和路径接受单元 19;例如,可以采用如下配置:在 OFC 100A 接收到一个新分组的情况下,路径计算单元 12 设置在 OFC 100A 中而路径接受单元 19 设置在 OFC 100B 中。

[0155] (新分组的旁路传输)

[0156] 接下来,将描述有关与上述处理规则(流条目)独立地并行执行新分组的旁路转发方法。图 14 是表示本示例性实施例中新分组的旁路转发的顺序图。由于图 14 的步骤 S001 到步骤 S003 的流程与图 12 的步骤 S001 到步骤 S003 的流程是一样的,下面的描述将针对转发路径计算之后的顺序流。

[0157] 参见图 14,已经计算出转发路径的 OFC 100A 查找与所述计算出的转发路径上的目标节点连接的 OFS(位于转发路径终点的 OFS),然后参考存储在 OFS-OFC 对应关系管理单元 16 中的 OFS-OFC 对应关系,并检索与涉及的 OFS 相关联的 OFC(步骤 S104)。这里,计算出一个转发路径,根据该路径,按照 OFS 301A、302A、301B 和 302B 的顺序将分组从节点 A 转发到目的地节点 D。这时,位于转发路径终点的 OFS 是 OFS 302B,并检索出与 OFS 302B 连接的 OFC 100B。

[0158] 然后,OFC 100A 将步骤 S001 中接收到的新分组传输至被检索出的 OFC 100B,并请求从从属 OFS 302B 的一指定接口进行传输(步骤 S105)。

[0159] OFC 100B 请求从属 OFS 302B 通过一个指定接口传输所述新分组(步骤 S106)。

[0160] OFS 302B 根据 OFC 100B 的指示从指定接口传输所述新分组(步骤 S107)。很多方法都可以被视为 OFS 302B 通过指定接口传输所述新分组的设置;从指定接口传输具有特定头部信息的分组的处理规则(流条目)可以存储在每一个 OFS 中,并且 OFC 100B 可以

执行将头部信息写入新分组的处理。

[0161] 综上所述,分组的旁路转发是可能实现的,其中,以转发路径的计算为触发,通过 OFC 转发分组(新分组),而不经转发路径上的 OFS,并通过终点将其输出。应该注意的是,图 15 所示的上述程序顺序可以独立于处理规则(流条目)的传输(设置)而单独执行。这样取得的效果是,可以将新分组传输到目标节点,而无需等待所述单元设置处理规则(流条目)。然而,此操作可与图 13 所示的传输(设置)处理规则(流条目)并行进行。例如,可以在图 12 的步骤 S004 至 S006-1 的处理之前或之后插入图 14 的步骤 S104 和 S105 的处理,以在 OFC 100A 中执行。

[0162] 图 15 清楚地显示,在本发明中没有限制 OFS/OFC 的数量,且有可能不经过 OFS 沿着转发路径传输分组。

[0163] 以上通过一个转发(即正常)路径的示例描述了本发明第一示例性实施例,该转发路径上的分组按照 OFS 301A、301B、302A 和 302B 的顺序(称作“正常方向”)从节点 A 转发到目标节点 D。然而,转发路径并不仅仅限于上述描述。例如,在计算的转发路径要通过 OFS 302A 和 302B 将分组从 OFS 301A 转发到与 OFS 302B 相连的节点的情况下,操作大致与上述描述相同。另外,例如,如果计算的转发路径要通过 OFS 301B、302A 和 301A 将分组从 OFS 302B 转发到节点 A,仅需交换上述 OFC 100B 与 OFC 100A,操作大致也与上述描述相同。

[0164] <第二示例性实施例>

[0165] 接下来,将参照附图详细描述本发明的第二示例性实施例,其中,对上文描述的一个 OFC 添加了修改。在此示例性实施例中,由于只有一个 OFC 的配置与第一示例性实施例不同,下面的描述将集中在其不同点上。

[0166] 图 16 是示出了本发明第二示例性实施例的 OFC 101 的配置的图。与图 6 所示的第一示例性实施例中的 OFC 100 的一个不同点就是添加了数据同步单元 20,用于获取与 OFC 100 连接的节点的信息。

[0167] 具体地说,数据同步单元 20 收集采用 LLDP(链路层发现协议)或各种类型的供应商协议的网上的各种节点中的装置信息及装置配置,获取节点之间的物理连接,并执行将上述结果映射到拓扑数据库 15 或 OFS-OFC 对应关系管理单元 16 中的操作。另外,也可以采用不在所有的 OFC 中安装数据同步单元 20 的配置,使某个特定 OFC 中的数据同步单元 20 通过 OFC 间通信单元 14 更新另一个 OFC 的拓扑数据库 15 和 OFS-OFC 对应关系管理单元 16。

[0168] 如上所述,本示例性实施例中使用带有数据同步单元 20 的 OFC,根据本实施例,在对网络配置添加了修改,或者相应节点发生故障的情况系,可以生成恰当的转发路径,也可以基于转发路径执行处理规则(流条目)分配。

[0169] <第三示例性实施例>

[0170] 接下来,将参照附图详细描述本发明第三示例性实施例,其中,对上文描述的 OFC 实施例添加了修改。在本示例性实施例中,对第一示例性实施例中的 OFC 添加了转发路径缓存功能,下面的描述将集中在其不同点上。

[0171] 图 17 是示出了本发明第三示例性实施例 OFC 102 的配置。其与图 6 所示的第一示例性实施例的 OFC 100 的一个不同点就是为路径计算单元 12 添加了路径记录单元 22。

[0172] 路径记录单元 22 用以在预定时间到达前记录路径计算单元 12 计算出的转发路径。

[0173] (设置缓存处理规则(流条目)的方法)

[0174] 接着,将参照附图详细描述本示例性实施例的一个操作。图 18 是表示本发明操作的顺序图。在图 18 的示例中,分组(新分组)以与 OFS301A 的流表 34 中存储的任何处理规则(流条目)都不匹配的节点 D 为目的节点,分组从与 OFS 301A 相连的节点 A 进行传输。

[0175] 下述操作与在第一示例性实施例中的操作相同:OFS 301A 接收新分组(步骤 S001)并请求与自身 OFS 301A 连接的 OFC 100A 传输(设置)一个包括该新分组的处理规则(流条目)(步骤 S002)。本示例性实施例中还应当注意的是,在步骤 S001 中,在接收到的分组与 OFS 301A 的流表 34 中存储的一个处理规则(流条目)匹配的情况下,就不执行之后的处理,而是执行基于节点处理规则的分组转发处理,如图 18 最下面的部分所示。

[0176] 在路径计算单元 12 计算转发路径之前,被请求传输(设置)处理规则(流条目)的 OFC 100A 参考路径记录单元 22,并搜索具有与新分组(例如:一个目的地为节点 D 且来自 OFS 301A 的分组)相同特性的转发路径信息(步骤 S201)。

[0177] 作为搜索结果,在找到了具有与新分组(例如:一个目的地为节点 D 且来自 OFS 301A 的分组)相同特性的转发路径信息的情况下,OFC 100A 就省略转发路径计算并执行步骤 S004 的操作以及之后的操作。

[0178] 另一方面,作为搜索结果,在没有找到具有与新分组(例如:一个目的地为节点 D 且来自 OFS 301A 的分组)相同特性的转发路径信息的情况下,OFC 100A 基于新分组的目的地节点 D 和所述请求的来源 OFS 301A,参考拓扑数据库 15 的网络拓扑结构,并计算包括从 OFS 301A 转发的分组的顺序流的转发路径(步骤 S202)。之后,OFC 100A 将计算出的转发路径和所述新分组的特性一起存储在路径记录单元 22 中(步骤 S203)。

[0179] 后续的操作与上述第一示例性实施例是相同的。根据本示例性实施例,关于生成了转发路径的分组,只要其内容存储在路径记录单元 22 中,由于可以省略转发路径生成,因此只要传输(设置)处理规则(流条目),就可以降低 OFC 负载并改进响应性。

[0180] 此外,在本示例性实施例中,也可以并行地执行图 14 所示的新分组旁路转发过程。在这种情况下,在图 18 的步骤 S004 至 S006-1 的处理之前或之后可以插入图 14 的步骤 S104 和 S105 的处理,以执行 OFC 100A 中的操作。

[0181] <第四示例性实施例>

[0182] 然后,将参照附图详细描述本发明第四示例性实施例。在本示例性实施例中,对上述 OFC 添加了修改。在本示例性实施例中,为第一示例性实施例添加了转发路径缓存功能和推测路径计算功能,下面的描述将集中在其不同点上。

[0183] 图 19 是示出了本发明第四示例性实施例中的 OFC 103 的配置。其与图 6 所示的第一示例性实施例的 OFC 100 的不同点是,为路径计算单元 12 添加了路径记录单元 22 和推测路径计算单元 23,推测路径计算单元用于生成转发路径并将其存储在路径记录单元 22 中。

[0184] 路径记录单元 22 用以存储路径计算单元 12 计算出的转发路径,直到预定时间到达。

[0185] 当 OFC 第一次启动或其负载低的时候,当网络拓扑结构中发生变化的时候,或在

一个预定的时间内,例如在固定的时间间隔等等,推测路径计算单元 23 参考拓扑数据库 15,计算从一个特定 OFS 到一个特定节点的转发路径,并记录在路径记录单元 22 中。这里“推测”的意思是不等待对处理规则(流条目)的传输(设置)请求,而是先生成一个处理规则(流条目)。

[0186] 应当注意的是,作为推测路径计算单元 23 针对其计算转发路径的节点和该 OFS 的组合,期望作出未存储在路径存储单元 22 中的选择。通过这种设置,就可以在生成处理规则(流条目)传输(设置)请求时缩减缓存缺失。

[0187] (使用推测传输路径设置处理规则(流条目)的方法)

[0188] 接下来,将参照附图详细描述本示例性实施例的操作。图 20 是表示本示例性实施例的操作的顺序图。在图 20 的一个示例中,从与 OFS 301A 连接的节点 A 传输以节点 D 为目的地的分组,节点 D 与 OFS 301A 的流表 34 中存储的所有处理规则(流条目)都不匹配。

[0189] 首先,当 OFC 第一次启动或负载低的时候,当网络拓扑结构发生变化的时候,或在预定的时间内,例如固定的时间间隔等等,OFC 100A 和 100B 的推测路径计算单元 23 执行推测转发路径计算(步骤 S300-1 和 S300-2)。

[0190] 下述操作与在第一示例性实施例中的操作相同:OFS 301A 接收新分组(步骤 S001)并请求与自身 OFS 301A 连接的 OFC 100A 传输(设置)一个包括该新分组的处理规则(流条目)(步骤 S002)。本示例性实施例中还应当注意的是,在步骤 S001 中,在接收到的分组与 OFS 301A 的流表 34 中存储的一个处理规则(流条目)匹配的情况下,就不执行之后的处理,而是执行基于节点处理规则的分组转发处理,如图 20 最下面的部分所示。

[0191] 在路径计算单元 12 计算转发路径之前,被请求传输(设置)处理规则(流条目)的 OFC 100A 参考路径记录单元 22,并搜索具有与新分组(例如:一个目的地为节点 D 且来自 OFS 301A 的分组)相同特性的转发路径信息(步骤 S301)。

[0192] 作为搜索结果,在找到了具有与新分组(例如:一个目的地为节点 D 且来自 OFS 301A 的分组)相同特性的转发路径信息的情况下,OFC 100A 就省略转发路径计算并执行步骤 S004 的操作以及之后的操作。

[0193] 另一方面,作为搜索结果,在没有找到具有与新分组(例如:一个终点为节点 D 且来自 OFS 301A 的分组)相同特性的转发路径信息的情况下,OFC 100A 基于新分组的目的地节点 D 和所述请求的来源 OFS 301A,参考拓扑数据库 15 的网络拓扑结构,并计算包括从 OFS 301A 转发的分组的顺序流中的转发路径(步骤 S302)。此后,OFC 100A 将计算出的转发路径和所述新分组的特性一起存储在路径记录单元 22 中(步骤 S303)。

[0194] 后续的操作与上述第一示例性实施例是相同的。根据本示例性实施例,关于生成了转发路径的分组以及统计路径计算单元计算的转发路径,只要其内容存储在路径记录单元 22 中,由于可以省略转发路径生成,因此就可以降低 OFC 负载并改进对处理规则(流条目)的传输。

[0195] 此外,在本示例性实施例中,也可以并行地执行图 14 所示的新分组旁路转发程序。在这种情况下,在图 18 的步骤 S004 至 S006-1 的处理之前或之后可以插入图 14 中的步骤 S104 和 S105 的操作,以执行 OFC 100A 中的操作。

[0196] <第五示例性实施例>

[0197] 接下来,将参照附图详细描述本发明的第五示例实施例。在此实施例中,上文中描

述的 OFC 相互之间不交换转发路径信息。由于此实施例可以通过一个与第一实施例中相同的 OFC 配置实现,以下描述将以其不同点为中心。

[0198] 图 21 是表示本发明第五示例性实施例的大致操作的图。关于第一示例性实施例描述中使用的图 13,有一点是:OFC 100A 传输的是新分组和作为起点(OFC 100B 针对该起点计算转发路径)的 OFS(301B),而不是转发路径信息。OFC 100B 生成从指定为起点的 OFS 301B 到目的地节点(节点 D)的转发路径,并将相应的处理规则(流条目)传输(设置)至与 OFC 100B 连接的 OFS。

[0199] (设置分布式处理规则(流条目)的方法)

[0200] 接下来,将参照图详细描述本示例性实施例的操作。图 22 是表示本实施例的操作的顺序图。

[0201] 下述操作与在第一示例性实施例中的操作相同:OFS 301A 接收新分组(步骤 S001)并请求与自身 OFS 301A 连接的 OFC 100A 传输(设置)一个包括该新分组的处理规则(流条目)(步骤 S002)。本示例性实施例中还应当注意的是,在步骤 S001 中,在接收到的分组与 OFS 301A 的流表 34 中存储的一个处理规则(流条目)匹配的情况下,就不执行之后的处理,而是执行基于节点处理规则的分组转发处理,如图 22 最下面的部分所示。

[0202] 被请求传输(设置)处理规则(流条目)的 OFC 100A 参考拓扑数据库 15 的网络拓扑结构,并基于路径计算单元 12 中的请求源 OFS 301A 和新分组的目的地节点 D,计算包括从 OFS 301A 转发的分组的顺序流的转发路径(步骤 S403)。这里,OFC 100A 并不计算从节点 A 到节点 D 的所有转发路径,而是计算一个部分转发路径,该部分转发路径显示在转发过程中将通过哪个与自身 OFC(100A)相连的 OFS 以及通过哪个接口进行输出。

[0203] 接下来,OFC 100A 参考存储在 OFS-OFC 对应关系管理单元 16 中的 OFS 和 OFC 之间的对应关系,并搜索与 OFS(一个片段的起始点 OFS,该片段的转发路径还未被计算出)连接的 OFC(步骤 S404),所述 OFS 与计算出的转发路径终点的 OFS 的输出接口相连。这里,检索与图 21 的 OFS 301B 连接的 OFC 100B。

[0204] OFC 100A 将步骤 S001 中接收的新分组和 OFS(一个片段的起点 OFS,该片段的转发路径尚未被计算出)传输至 OFC 100B(步骤 S405),所述 OFS 与计算出的转发路径终点的 OFS 的输出接口连接。

[0205] 作为转发路径尚未被计算出的片段的起点 OFS 的 OFS 和接收步骤 S001 中接收到的新分组的 OFC 100B 参考拓扑数据库 15 的网络拓扑图结构,基于路径计算单元 12 中被指定为起点的 OFS 301B 和作为新分组的目的地节点 D,根据与 OFC 100B 连接的 OFS 组计算转发路径(步骤 S406)。这里,OFC 100B 计算从 OFS 301B 至 OFS 302B 进行转发的转发路径,以及从 OFS 302B 的接口 #2 至节点 D 进行转发的转发路径。

[0206] 此后,OFC 100A 参考拓扑数据库 15 的网络拓扑结构以及与 OFC 100A 连接的 OFS 的信息,生成处理规则(流条目)(所述规则将在与 OFC 100A 连接的 OFS 中位于转发路径上的 OFS 中进行设置),并针对 OFS 301A 和 302A 执行传输(设置)(步骤 S407-1)。

[0207] 同样地,OFC 100B 也参考拓扑数据库 15 的网络拓扑结构以及与 OFC 100B 连接的 OFS 的信息,生成处理规则(流条目)(所述处理规则将在与自身控制装置 OFC 100B 连接的 OFS 中位于转发路径上的 OFS 中进行设置),并针对 OFS 301B 和 302B 执行传输(设置)(步骤 S407-2)。

[0208] 此后,在 OFS 301、302A、301B 和 302B 中的每一个中,均执行处理以在流表 34(步骤 S007) 中登记或更新接收到的处理规则(流条目)。

[0209] 从上文看出,在 OFS 301、302A、301B 和 302B 中的每一个中,都设置了与计算出的转发路径相对应的处理规则(参见图 11),使得节点 A 和节点 B 之间的分组转发成为可能。

[0210] 如上所述,根据本示例性实施例,通过将计算分配给几个 OFC 来执行转发路径计算是可能的。此外,由于采用了上述配置,因此可以分布式设置拓扑数据库 15。

[0211] 此外,本示例性实施例可以被配置为与第二至第四示例性实施例的组合。例如,如果与第三示例性实施例组合,OFC 100A 使用缓存的转发路径,由 OFC 100B 执行路径计算的修改实施例是可行的。同样,如果与第四示例性实施例组合,OFC 100A 使用预先计算出的推测转发路径,由 OFC 100B 执行路径计算的修改实施例是可行的。

[0212] 以上描述了本发明优选示例性实施例及其具体操作,但是本发明并不局限于上述示例性实施例,可以在不脱离本发明基础性技术概念的范围内进行进一步修改、替换和调整。

[0213] 例如,在上述各示例性实施例中,利用非专利文献 1 和 2 中的技术描述了使用 OFC 和 OFS 的引用示例,但是利用各种类型进行同等操作的节点和控制装置实现也是可以的。

[0214] 例如,从上述第一到第四示例性实施例中,在描述中每一个 OFC 都分别设有路径计算单元 12。然而,从上述描述中可以理解,在接收转发路径信息一侧的 OFC 中没有进行转发路径计算处理。因此,可以省去没有定期发生转发路径计算处理的 OFC 中的路径计算单元 12。基于类似的原因,也可以省去不接收转发路径信息的 OFC 中的路径接受单元 19。

[0215] 标号说明

[0216] 11 分组转发单元

[0217] 12 路径计算单元

[0218] 13 路径传输单元

[0219] 14 OFC 间通信单元

[0220] 15 拓扑数据库

[0221] 16 OFS-OFC 对应关系管理单元

[0222] 16A 对应关系管理单元

[0223] 17 OFS 通信单元

[0224] 17A 传输单元

[0225] 18 流生成单元

[0226] 18A 处理规则生成单元

[0227] 19 路径接受单元

[0228] 20 数据同步单元

[0229] 22 路径记录单元

[0230] 23 推测路径计算单元

[0231] 31 OFC 通信单元

[0232] 32 流设置单元

[0233] 33 分组处理单元

[0234] 34 流表

[0235] 100 至 103、100A、100B、100C 控制装置 (OFC)

[0236] 300、301A 至 303A、301B 至 303B、301C 至 303C 转发节点 (OFS)

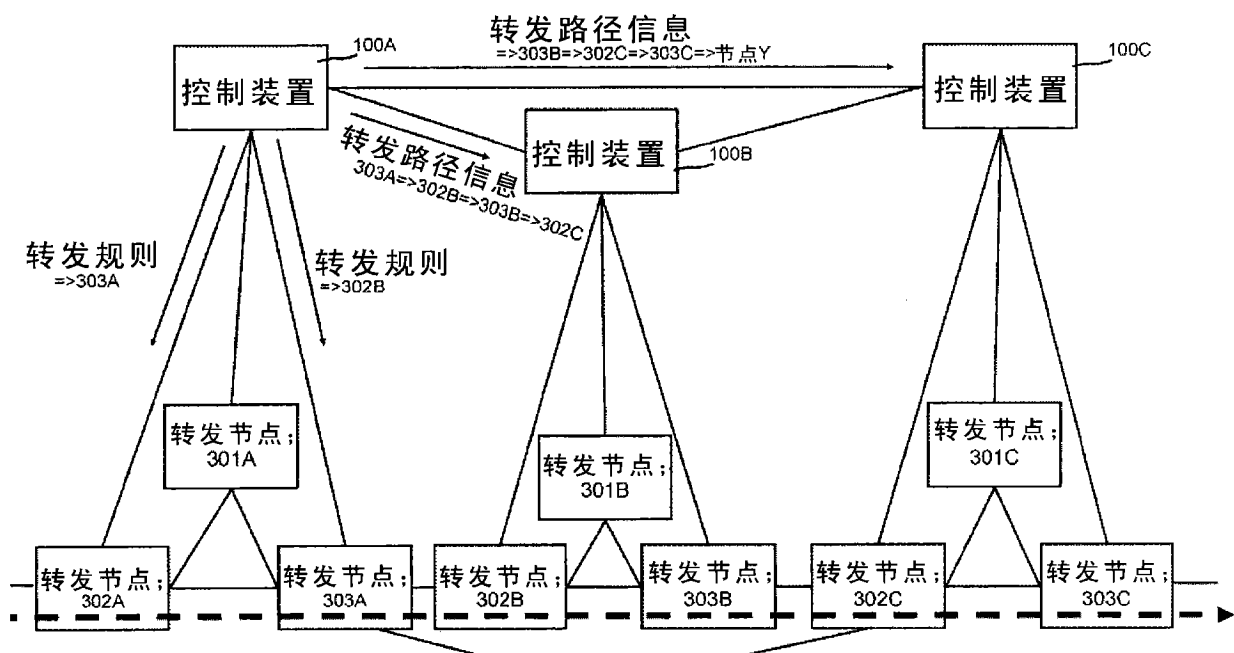


图 1

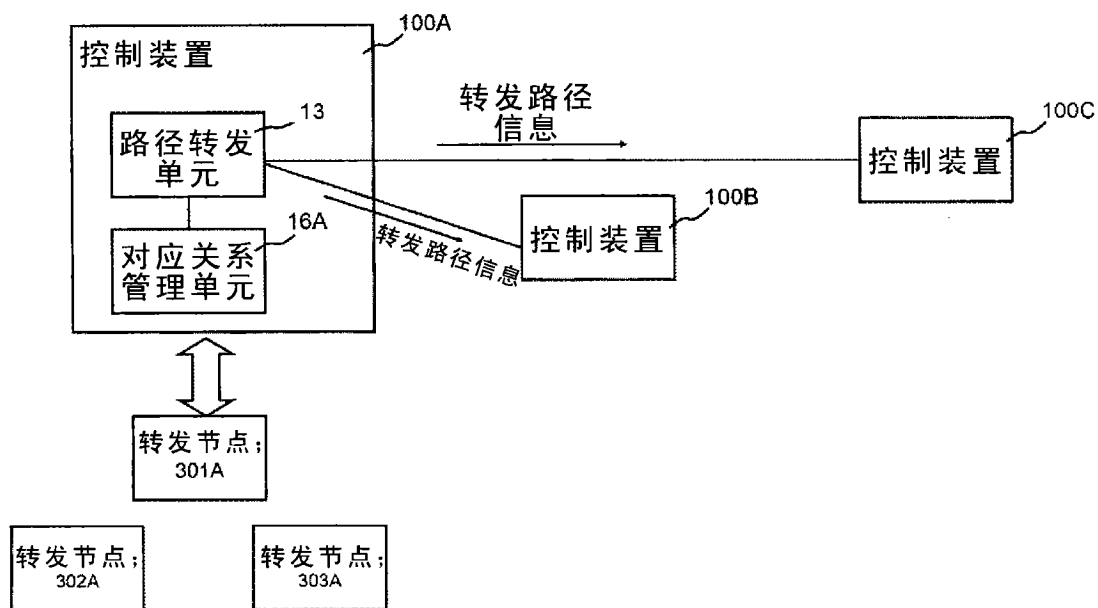


图 2

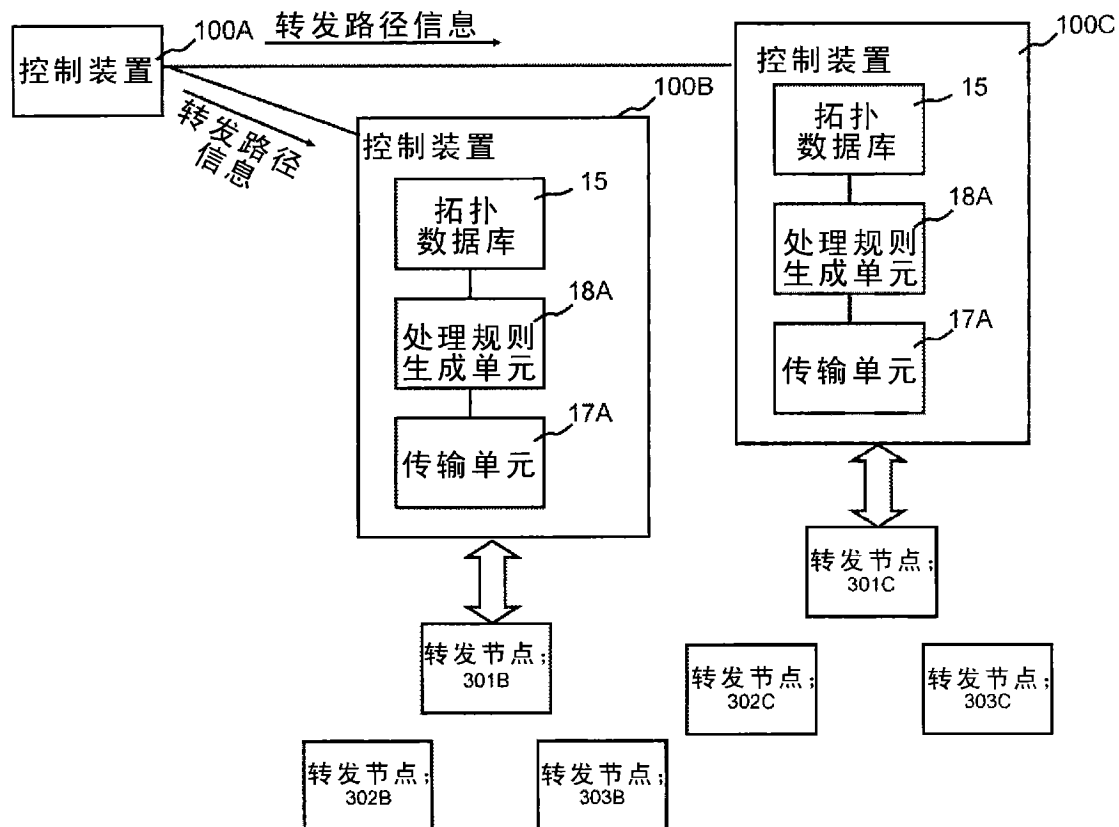


图 3

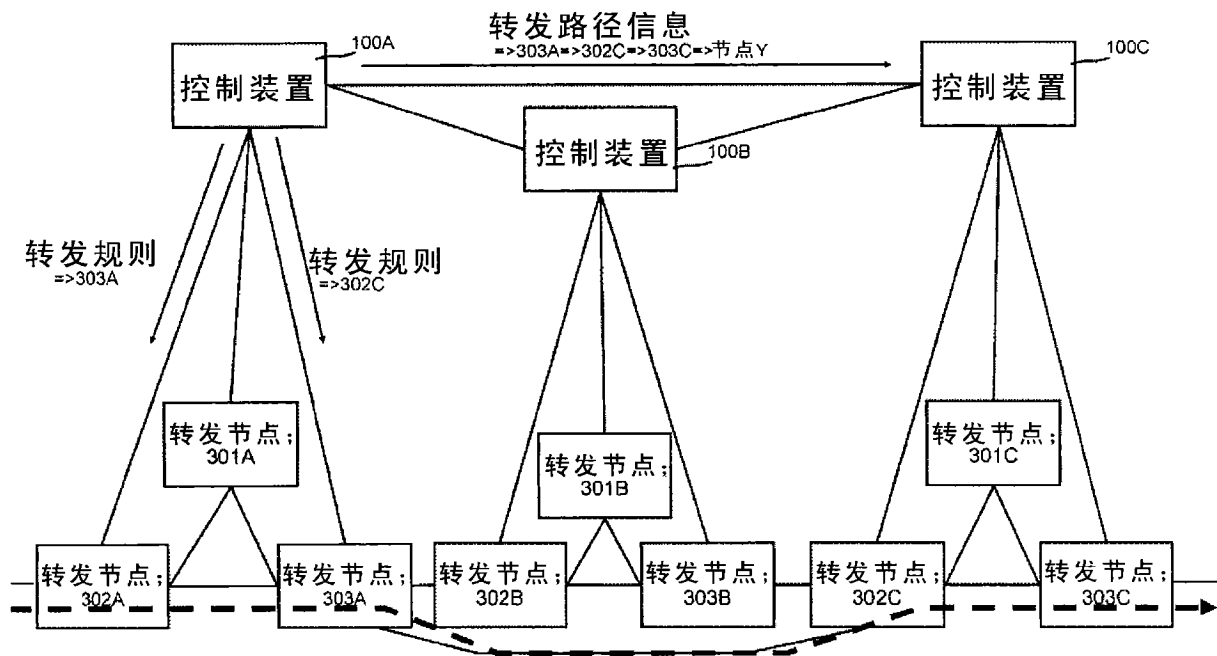


图 4

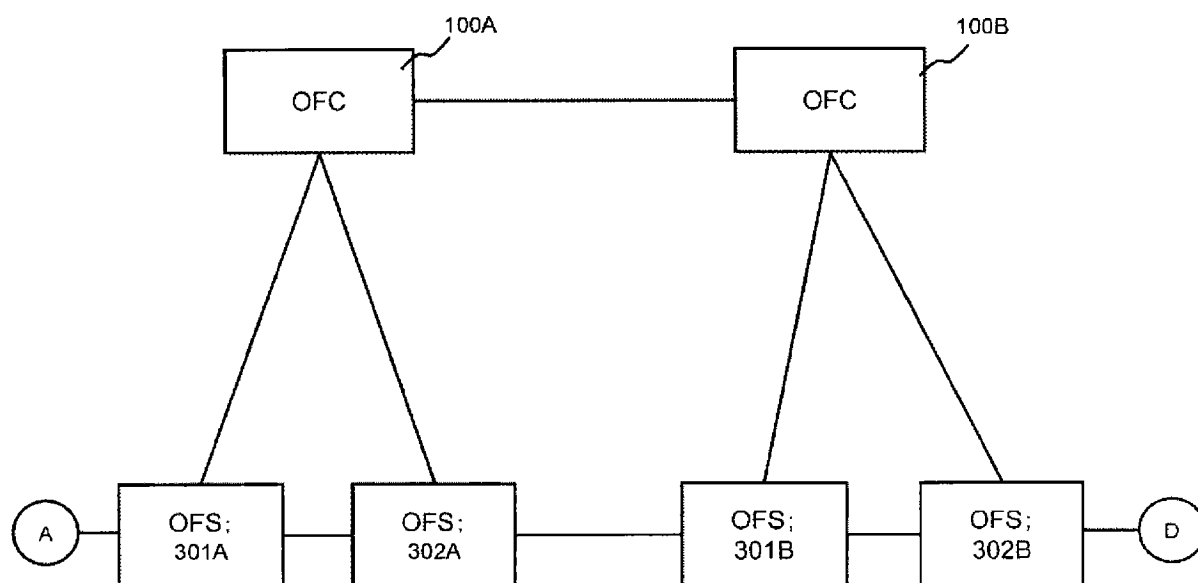


图 5

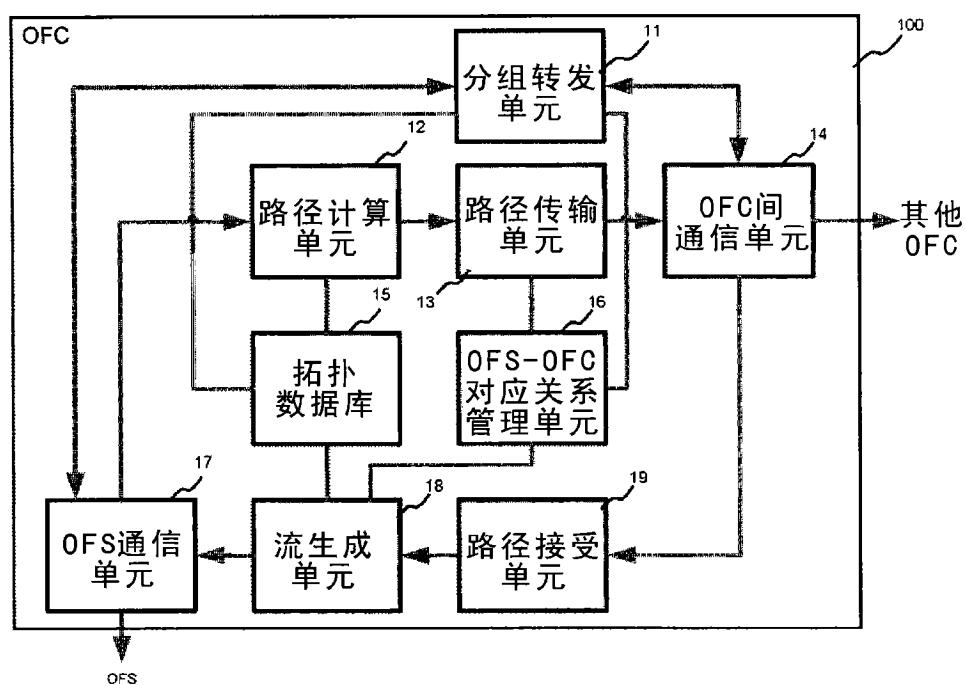


图 6

	接口1	接口2
OFS301A	节点A	OFS302A
OFS302A	OFS301A	OFS301B
OFS301B	OFS302A	OFS302B
OFS302A	OFS301B	节点D

图 7

	所连接的交换机	
OFC;100A	OFS;301A	OFS;302A
OFC;100B	OFS;301B	OFS;302B

图 8

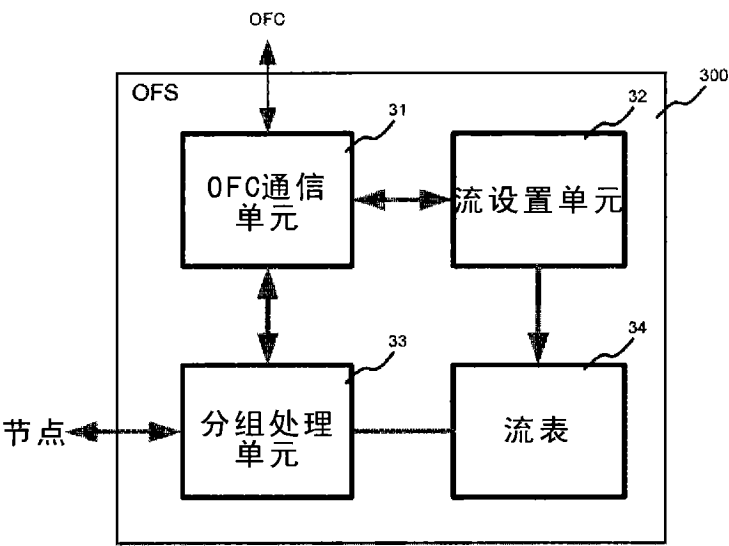


图 9

规则												动作 (Actions)	Stats
Wildcards	In Port	Ether SA	Ether DA	Ether type	VLAN ID	IP SA	IP DA	IP proto	Src port	Dst port			

图 10

-
- OFS;301A
 - 规则
 - IP SA: 节点A的IP
 - IP DA: 节点D的IP
 - 保留wildcard (通配符)
 - 动作
 - 转发至接口2
 - OFS;302A
 - 规则
 - IP SA: 节点A的IP
 - IP DA: 节点D的IP
 - 保留wildcard (通配符)
 - 动作
 - 转发至接口2
 - OFS;301B
 - 规则
 - IP SA: 节点A的IP
 - IP DA: 节点D的IP
 - 保留wildcard (通配符)
 - 动作
 - 转发至接口2
 - OFS;302B
 - 规则
 - IP SA: 节点A的IP
 - IP DA: 节点D的IP
 - 保留wildcard (通配符)
 - 动作
 - 转发至接口2
- OFS;301A
 - 规则
 - IP SA: 节点D的IP
 - IP DA: 节点A的IP
 - 保留wildcard (通配符)
 - 动作
 - 转发至接口1
 - OFS;302A
 - 规则
 - IP SA: 节点D的IP
 - IP DA: 节点A的IP
 - 保留wildcard (通配符)
 - 动作
 - 转发至接口1
 - OFS;301B
 - 规则
 - IP SA: 节点D的IP
 - IP DA: 节点A的IP
 - 保留wildcard (通配符)
 - 动作
 - 转发至接口1
 - OFS;302B
 - 规则
 - IP SA: 节点D的IP
 - IP DA: 节点A的IP
 - 保留wildcard (通配符)
 - 动作
 - 转发至接口1

图 11

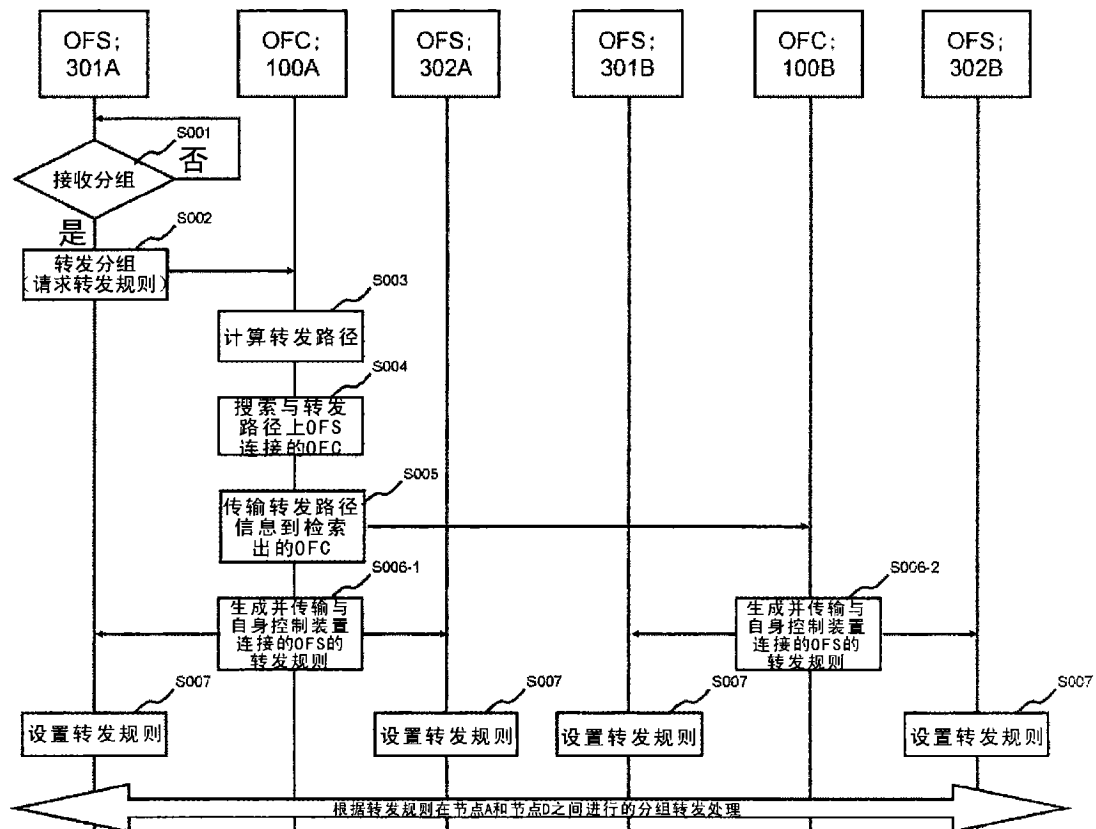


图 12

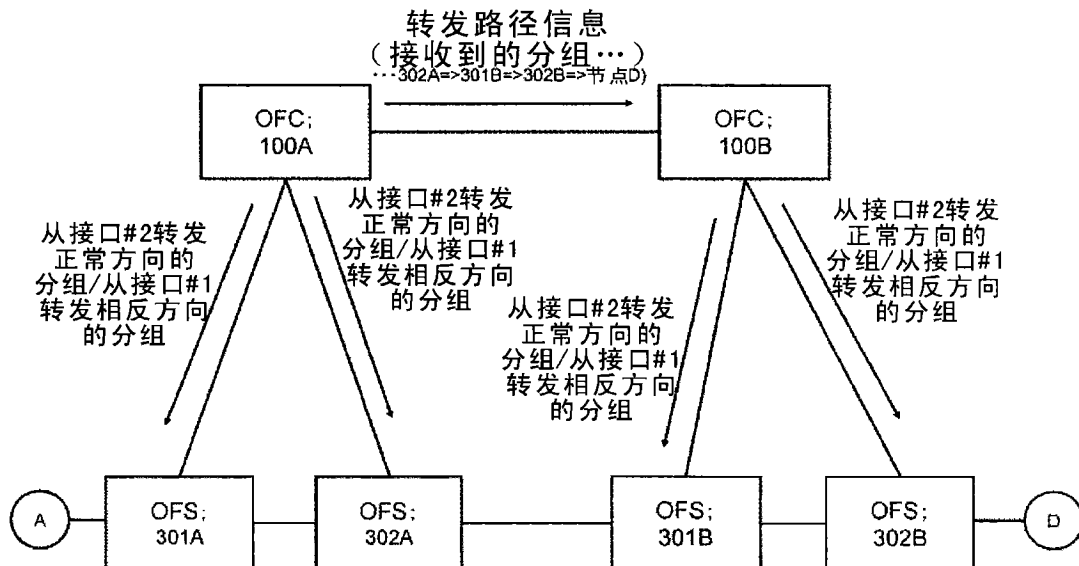


图 13

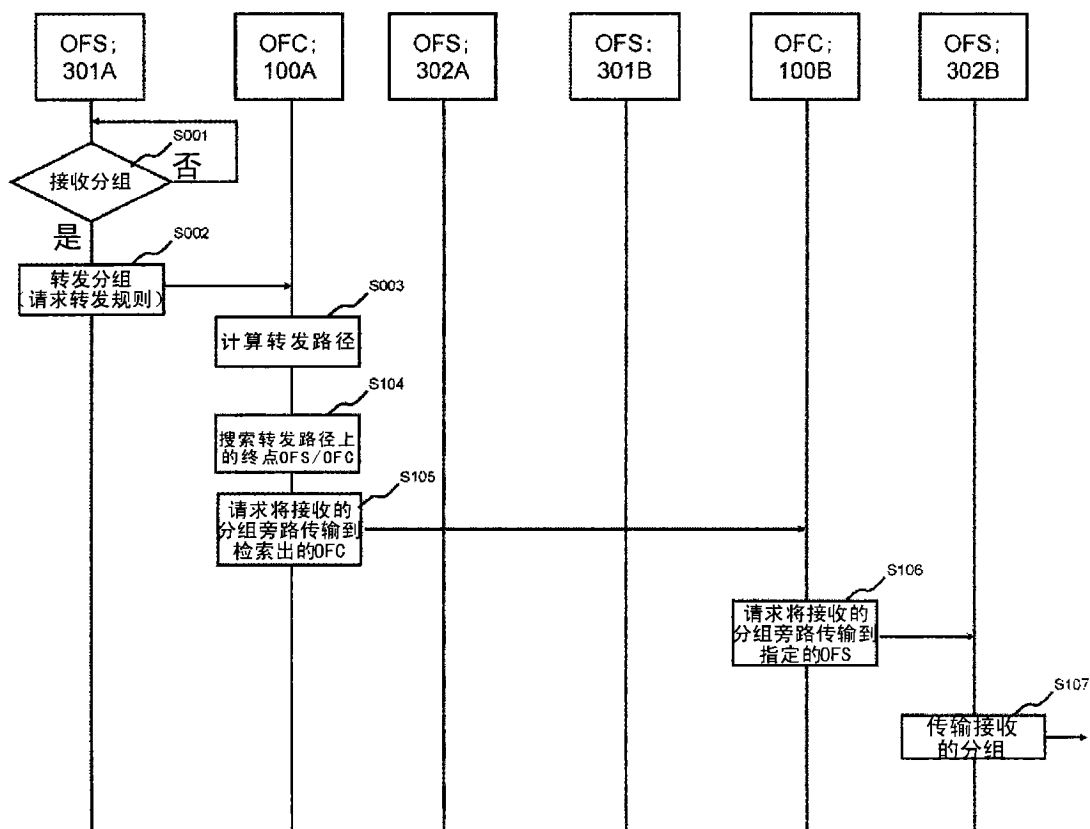


图 14

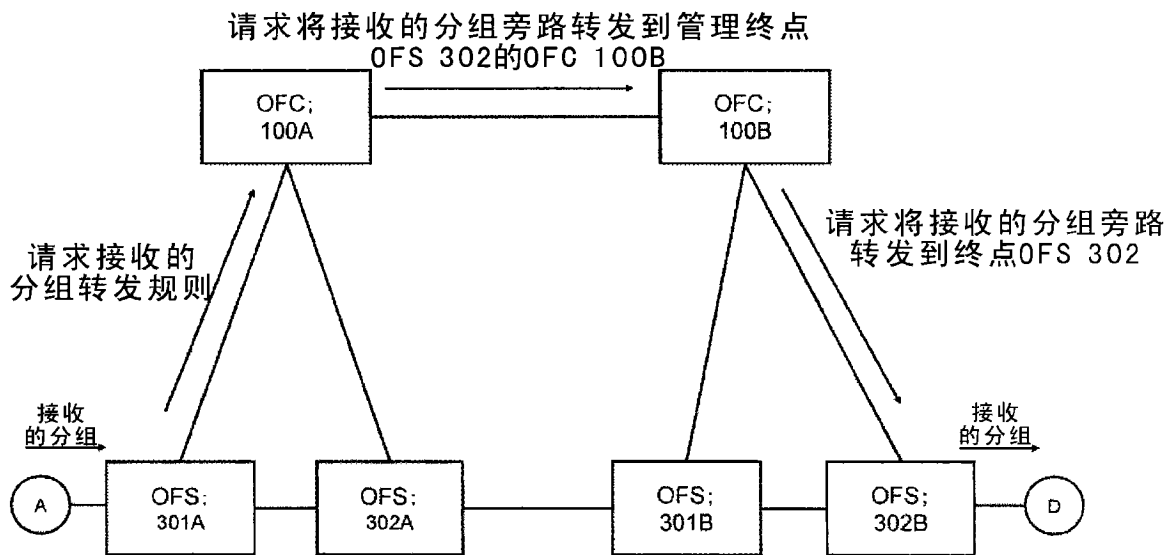


图 15

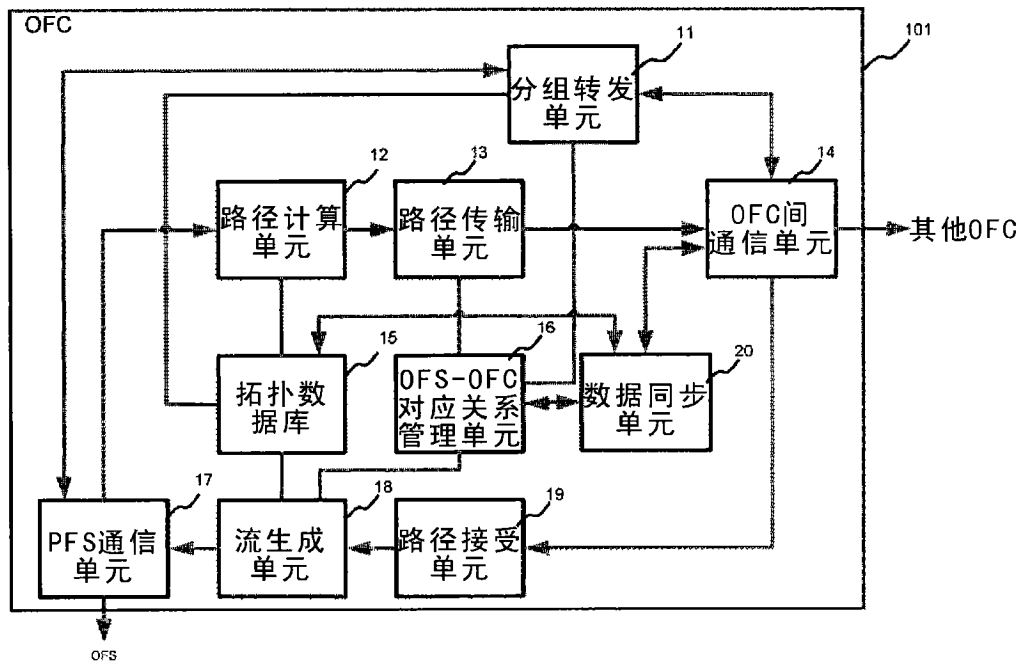


图 16

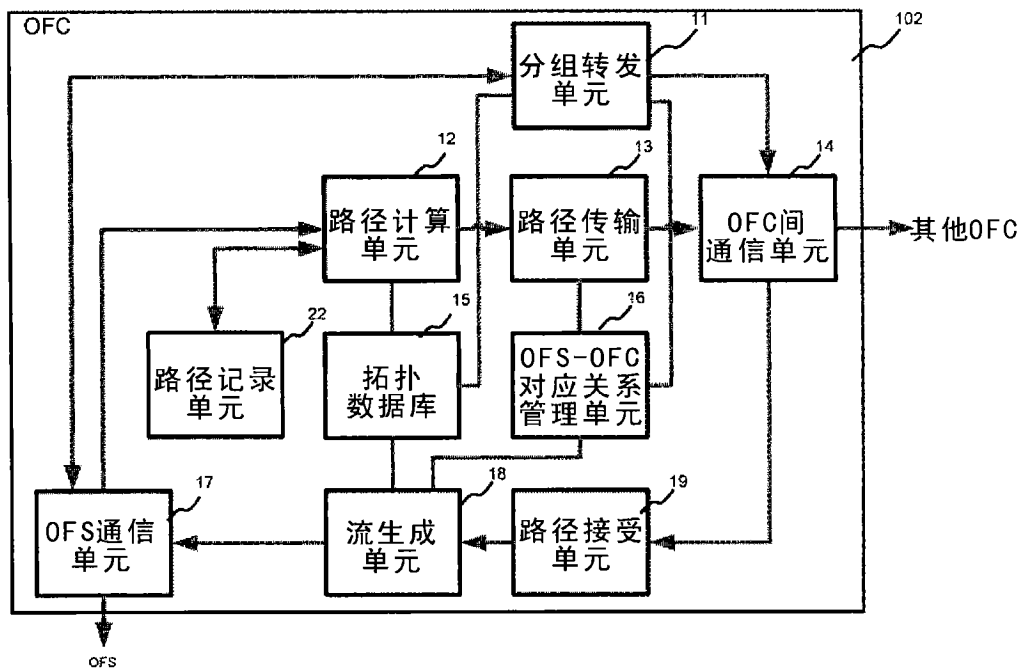


图 17

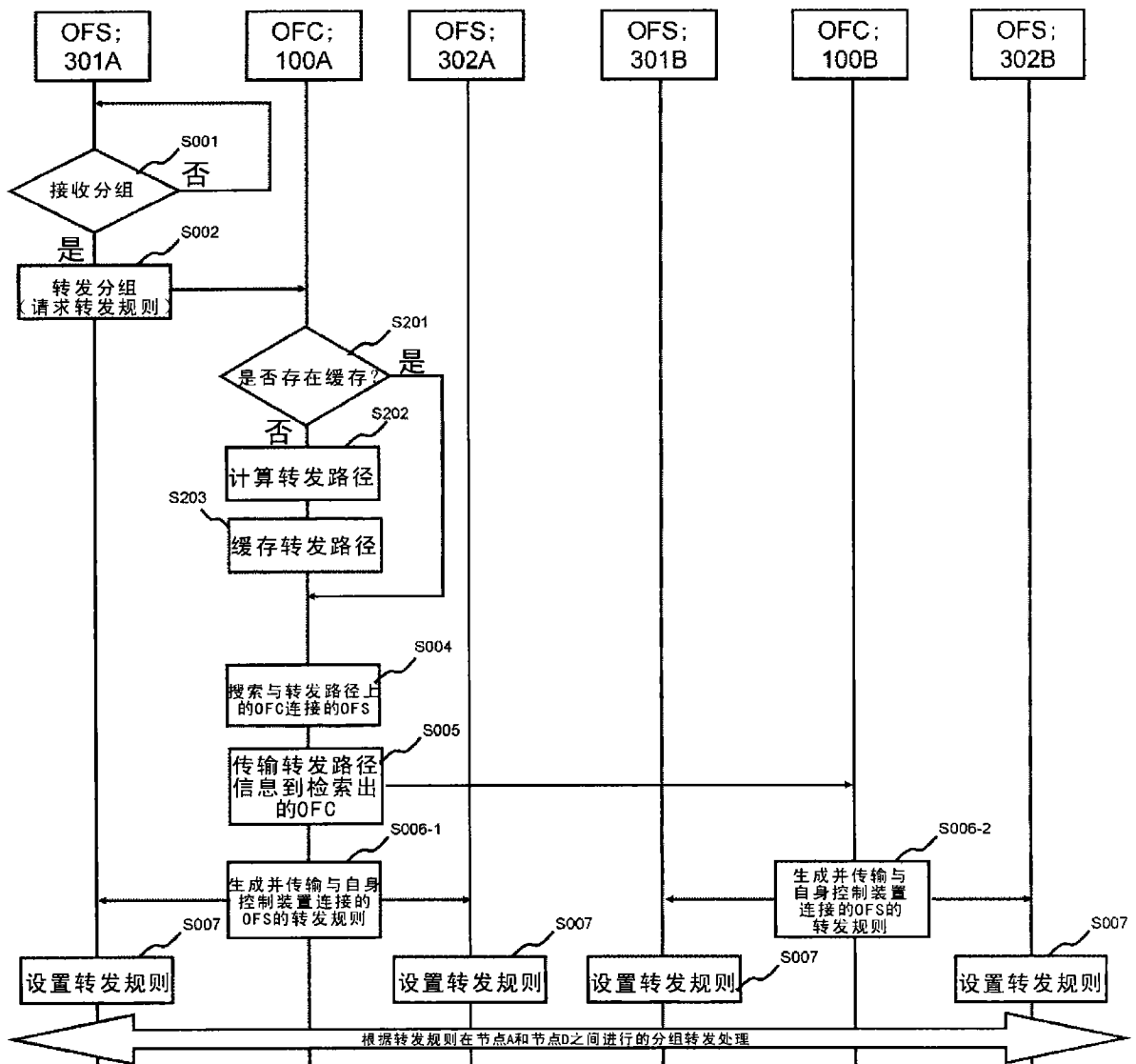


图 18

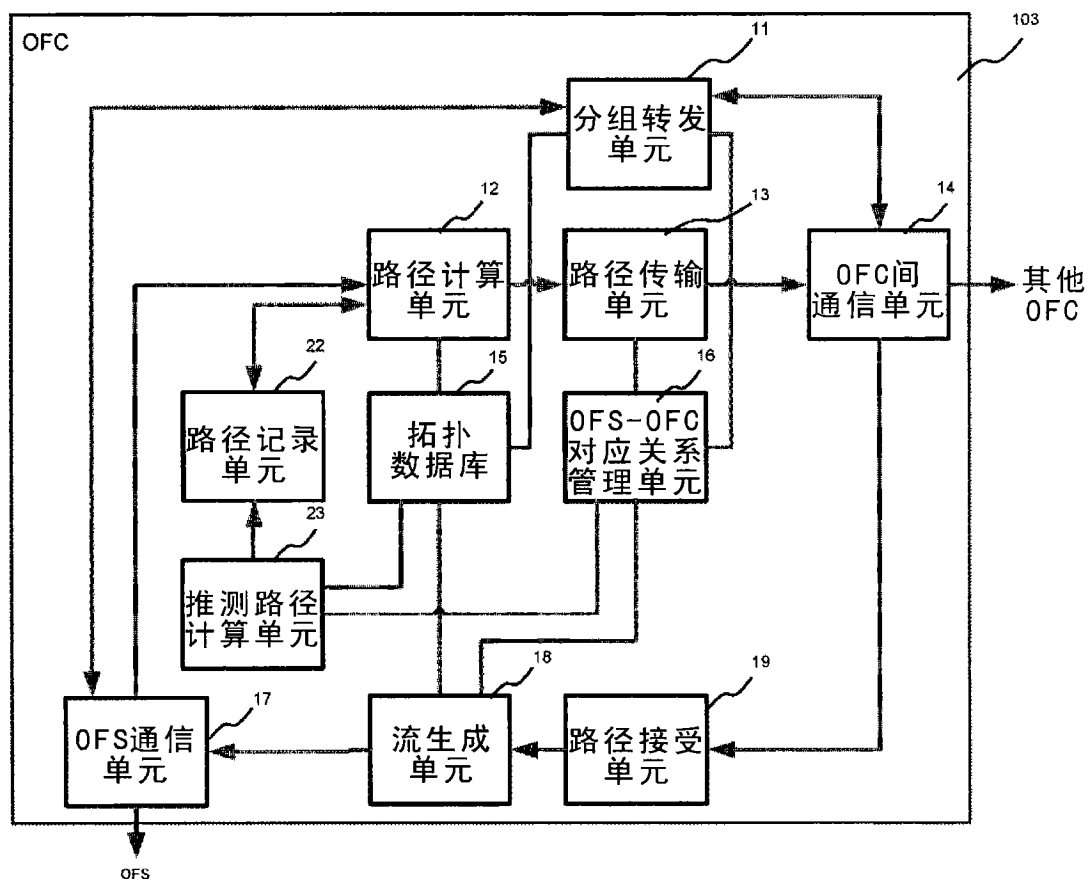


图 19

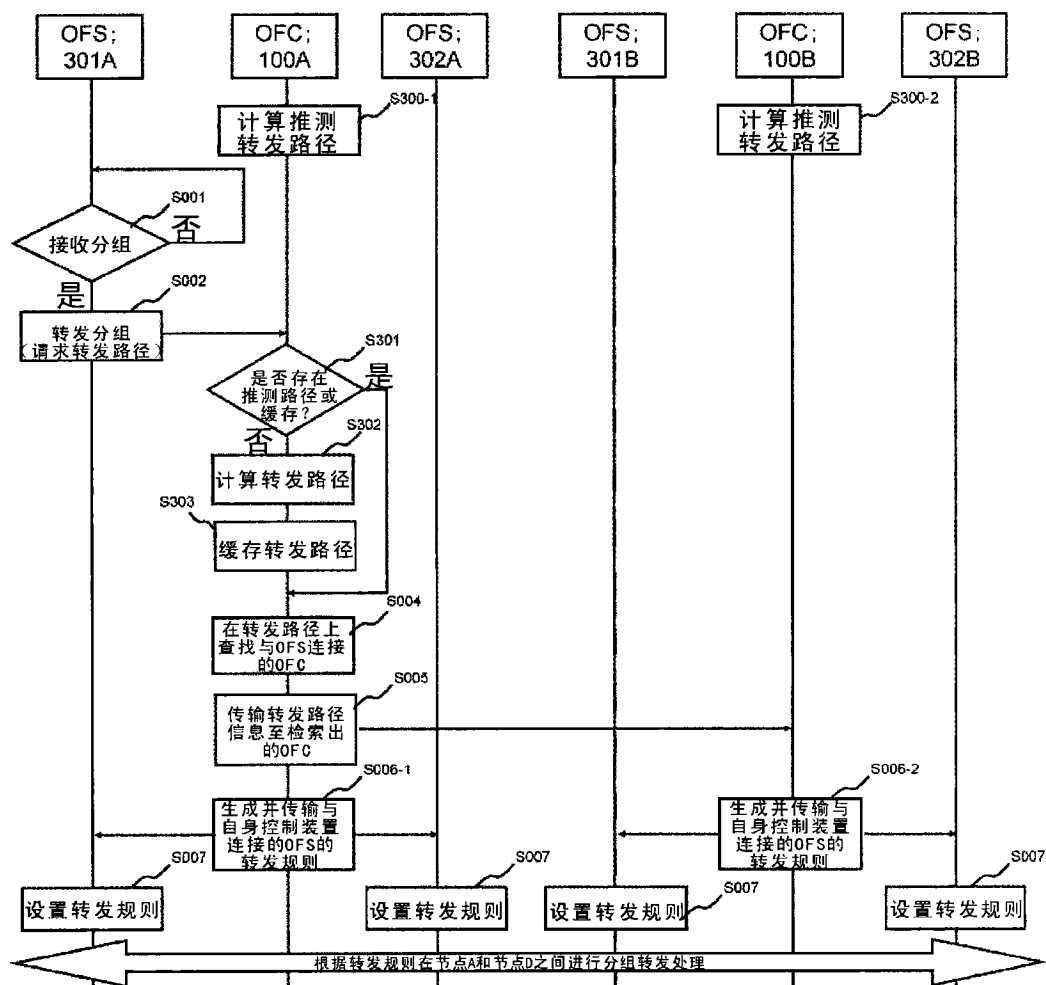


图 20

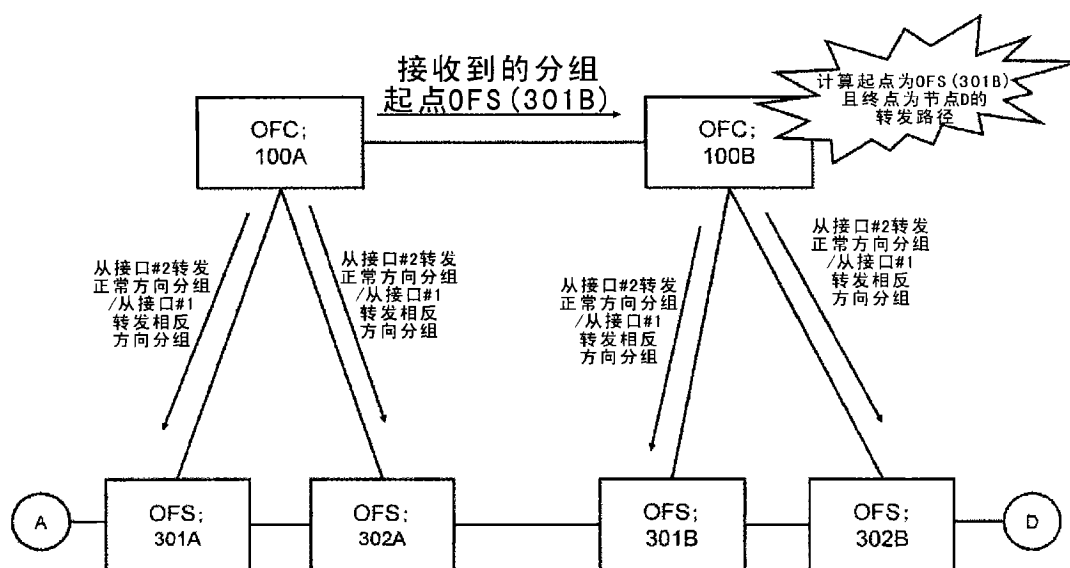


图 21

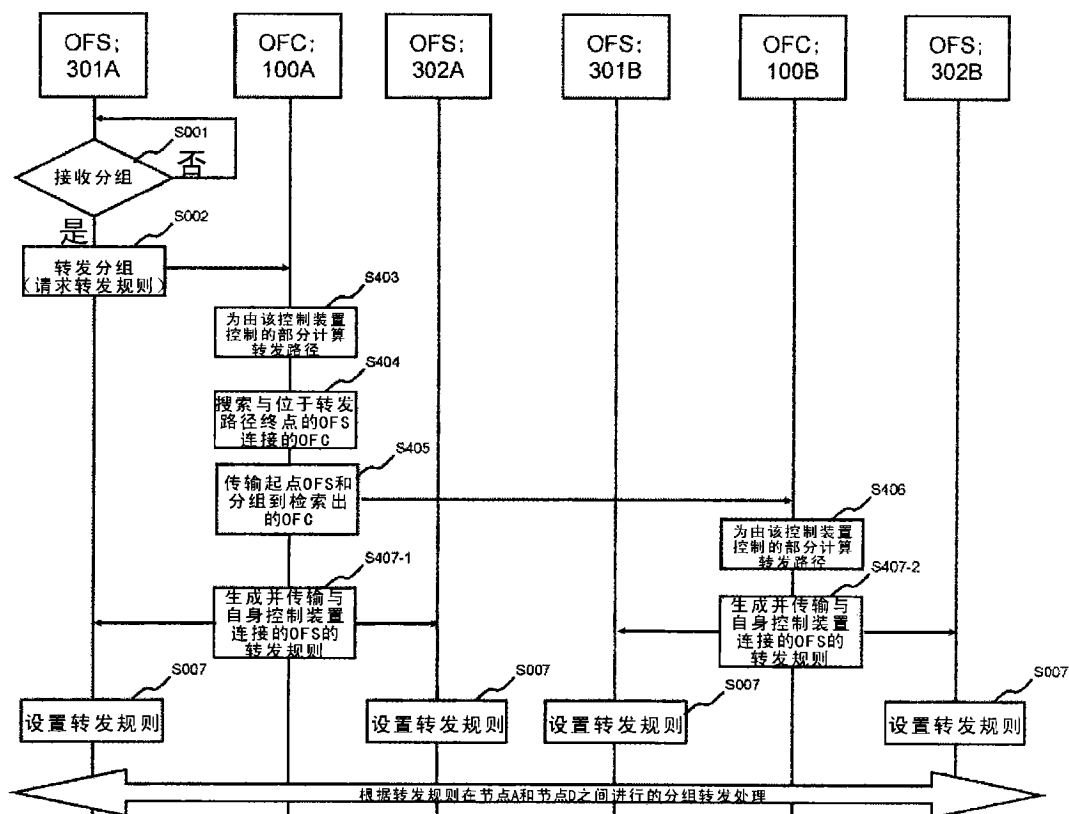


图 22

动作名称	动作内容
OUTPUT	向指定端口输出
SET_VLAN_VID	通过指定的VLAN ID添加/更新VLAN标签
SET_VLAN_PCP	通过指定的VLAN优先级添加/更新VLAN标签
STRIP_VLAN	发布IEEE802.1q Vlan标签
SET_DL_SRC	更新源MAC地址
SET_DL_DST	更新目的地MAC地址
SET_NW_SRC	更新源IP地址
SET_NW_DST	更新目的地IP地址
SET_TP_SRC	更新TCP/UDP源端口
SET_TP_DST	更新TCP/UDP目的地端口
VENDOR	供应商定义的动作

图 23