



(12) 发明专利

(10) 授权公告号 CN 103428031 B

(45) 授权公告日 2016. 04. 13

(21) 申请号 201310335883. 7

审查员 翁平

(22) 申请日 2013. 08. 05

(73) 专利权人 浙江大学

地址 310058 浙江省杭州市西湖区余杭塘路
866 号

(72) 发明人 吴春明 李徐艳 周伯阳 姜明

(74) 专利代理机构 杭州求是专利事务所有限公
司 33200

代理人 周烽

(51) Int. Cl.

H04L 12/24(2006. 01)

H04L 12/721(2013. 01)

(56) 对比文件

CN 1859156 A, 2006. 11. 08,

CN 1798051 A, 2006. 07. 05,

US 7471625 B2, 2008. 12. 30,

CN 101060497 A, 2007. 10. 24,

CN 101534239 A, 2009. 09. 16,

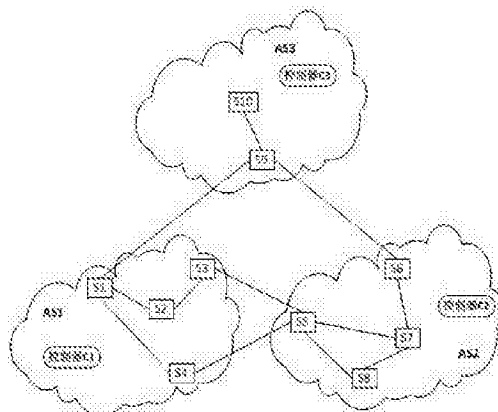
权利要求书2页 说明书5页 附图2页

(54) 发明名称

一种基于软件定义网络的域间链路快速故障
恢复方法

(57) 摘要

本发明公开了一种基于软件定义网络的域间
链路快速故障恢复方法,该方法在域间链路出现
故障的时候,根据存储在 OpenFlow 交换机本地
的邻接节点表选择绕过故障链路的路径而不是
由控制器根据最新的拓扑来选择一条新路径。
本发明与由控制器介入的链路故障恢复方法
相比,有恢复速度快的优点,能减少丢包率,
其实现方法简便,灵活性好,具有可扩展性。



1. 一种基于软件定义网络的域间链路快速故障恢复方法,其特征在于,包括以下步骤:

步骤 1:初始网络中的所有的 OpenFlow 交换机都和控制器相连,每个 OpenFlow 交换机有全网唯一的标识符,由同一个控制器管理的所有 OpenFlow 交换机组成一个域,域由控制器的 IP 地址标识;OpenFlow 交换机在与控制器建立连接时,控制器会把域间 OpenFlow 交换机选择策略加载到 OpenFlow 交换机;

步骤 2:各 OpenFlow 交换机向邻接 OpenFlow 交换机广播邻接消息,该邻接消息封装了消息类型、域 ID、物理节点 ID、端口 ID 和端口 IP;OpenFlow 交换机收到从邻接节点发来的邻接消息后将邻居 OpenFlow 交换机的信息加入到自己维护的邻接节点表中;邻接节点表由邻接节点域 ID、邻接节点 ID、邻接节点端口 ID、邻接节点端口 IP、与邻接节点相连的本 OpenFlow 交换机的端口 ID 和定时器项组成;

步骤 3:每个 OpenFlow 交换机把用 OpenFlow 消息封装的节点信息消息发送给控制器;该节点信息消息由消息类型、OpenFlow 交换机信息和邻接 OpenFlow 交换机信息组成;控制器收到各 OpenFlow 交换机发来的封装了节点信息消息的 OpenFlow 消息后,就获得了本域的物理拓扑信息以及与本域相连的其他域的边界 OpenFlow 交换机信息(如果一个域的 OpenFlow 交换机与其他域的 OpenFlow 交换机相连,则该 OpenFlow 交换机为边界 OpenFlow 交换机),控制器将本域的物理拓扑存储到本地,把域间的物理拓扑存储至网络视图,每个控制器根据网络视图获取全网域间拓扑;

步骤 4:维护域间链路,快速恢复故障链路;该步骤通过以下子步骤来实现:

(4.1)OpenFlow 交换机周期性地向邻接 OpenFlow 交换机广播邻接消息,OpenFlow 交换机收到邻接消息后,提取出物理节点 ID 字段,并与邻接节点表中的邻接节点 ID 相匹配,如果匹配成功,就重置命中表项的定时器;如果匹配不成功,表示有新的 OpenFlow 交换机与之相连,则更新邻接节点表,为刚连入的 OpenFlow 交换机增加一个表项,并由 OpenFlow 交换机发送一个封装了节点信息消息的 OpenFlow 消息给控制器,以告知新节点的加入以及网络拓扑的变化;

(4.2) 每个 OpenFlow 交换机定期检查邻接节点表中的定时器项,如果某个表项的定时器超过预先设定的最大值,表示此 OpenFlow 交换机在一定时间内没有收到表项中 OpenFlow 交换机发来的邻接消息,表项中的 OpenFlow 交换机为不可达;

(4.3) 如果 OpenFlow 交换机检测到某个邻接 OpenFlow 交换机不可达,说明两个 OpenFlow 交换机之间的链路发生了故障;OpenFlow 交换机从邻接节点表中找出邻接节点域 ID 为与不可达 OpenFlow 交换机域 ID 相同的候选表项,如果有多个候选表项,根据域间 OpenFlow 交换机选择策略从多个候选表项中选择一个最佳表项,转到步骤(4.4);若没有找到,转到步骤(4.5);

(4.4) OpenFlow 交换机根据最佳表项修改流表:先找出流表项中数据包发出端口为与不可达 OpenFlow 交换机相连的端口的流表项,再修改这些流表项把发出端口修改为最佳表项中的与邻接节点相连的本 OpenFlow 交换机的端口 ID,从而把发往不可达 OpenFlow 交换机的流量定向到与不可达 OpenFlow 交换机在同一个域中的边界 OpenFlow 交换机;

(4.5) OpenFlow 交换机采用洪泛法发送寻找路径消息以找到此 OpenFlow 交换机到不可达 OpenFlow 交换机所在域的路径;该寻找路径消息封装了消息类型、不可达 OpenFlow 交

交换机的域D_ID和源OpenFlow交换机的域S_ID和路径；寻找路径消息每到达一个OpenFlow交换机的时候需要把OpenFlow交换机的ID和收到寻找路径消息的端口ID写入路径，以方便寻找路径回复消息的发送，每从OpenFlow交换机的端口发送前，需要把发出端口ID写入到路径中，以方便随后流表更新消息的发送；OpenFlow交换机首先向除了与不可达OpenFlow交换机相连的端口外的其他端口发送寻找路径消息，当邻接OpenFlow交换机收到寻找路径消息后，查找邻接节点表以确定邻接节点表中是否存在邻接节点域ID为D_ID的OpenFlow交换机；如果找到，根据寻找路径消息中的路径发送寻找路径回复消息；如果没有找到，OpenFlow交换机把寻找路径消息从除了收到寻找路径消息端口之外的端口发送出去直到找到通往域D_ID的一条路径；当源OpenFlow交换机收到寻找路径回复消息后，提取出寻找路径回复消息中的路径，若OpenFlow交换机收到多条寻找路径回复消息，则从所有路径中选择出跳数最少的路径为最佳路径，如果有多条跳数相同且最少的路径，则随机从多条跳数最少的路径中选择一条为最佳路径；

(4.6)OpenFlow交换机找到发往不可达OpenFlow交换机的最佳路径后，往最佳路径发送流表更新消息，以往最佳路径上的OpenFlow交换机增加流表项使发往不可达OpenFlow交换机的数据包往最佳路径发送；该流表更新消息封装了消息类型、不可达OpenFlow交换机的域ID、路径和匹配域；路径为从寻找路径回复消息中提取出的最佳路径，匹配域为在源OpenFlow交换机流表中找出的数据包发出端口为与不可达OpenFlow交换机相连的端口的流表项中的匹配域，匹配域有一个或多个；OpenFlow交换机根据流表更新消息中的路径来发送流表更新消息，每到达一个OpenFlow交换机的时候，为OpenFlow交换机增加流表项，如果流表更新消息中有多个匹配域，则增加多个流表项，流表项的匹配域为流表更新消息中的匹配域，从流表更新消息的路径中提取下一跳的发出端口并写入流表项的行为部分，并把该流表项的优先级设为最高；这样就为发往不可达OpenFlow交换机所在的域的数据包建立了一条新的路径；

(4.7)OpenFlow交换机向控制器发送封装了节点信息消息的OpenFlow消息报告此OpenFlow交换机与不可达OpenFlow交换机之间的链路故障，控制器更新拓扑，从而实现域间链路的故障恢复。

一种基于软件定义网络的域间链路快速故障恢复方法

技术领域

[0001] 本发明涉及计算机网络技术领域,尤其涉及一种基于软件定义网络的域间链路快速故障恢复方法。

背景技术

[0002] 经过数十年的发展之后,互联网无论是在商业方面还是在通信领域都取得了十分巨大的成功,也为人类的生活与生产方式带来了自工业革命以来最有影响力的变革。然而规模巨大的网络同时也限制了它的发展,导致了网络体系结构的“僵化”。软件定义网络(SDN)通过把路由器中的控制面与数据面分离开来并把控制面放到中心控制器上实现网络的可编程性,控制器根据所承载网络业务的路由策略来决定如何转发数据包并把转发规则通过控制器与转发设备(OpenFlow 交换机)间的通信标准(如 OpenFlow 协议)写入转发设备的流表中。

[0003] 分布式 SDN 控制器改进了单一 SDN 控制器控制底层转发设备的可伸缩性,将整个网络分割成多个域,每个域的转发设备由各自的控制器来控制,域间物理链路状态存储到一致的网络视图(Network View,由各个控制器通过一致性分布式算法来维护,如分布式哈希表等)上。当域间的链路发生故障的时候,发往故障链路的数据包都将丢弃。链路的恢复需要控制器的参与,控制器收到拓扑变化的消息后,首先要把故障链路状态存储至网络视图中,然后重新为发往故障链路的数据包计算一条路径并写入相应的转发设备,以确保发往故障链路的数据包从新的路径转发而不被丢弃。这种域间故障链路恢复的方法需要控制器的参与,因此速度比较慢,会使更多的数据包被丢弃,而且控制器的参与会占用控制器与交换设备之间的带宽。

发明内容

[0004] 本发明的目的在于针对现有技术的不足,提供一种基于软件定义网络的域间链路快速故障恢复方法,该方法效率高,灵活性好,实现简便。

[0005] 本发明的目的是通过以下技术方案来实现的:一种基于软件定义网络的域间链路快速故障恢复方法,包括以下步骤:

[0006] 步骤 1:初始网络中的所有 OpenFlow 交换机(OFS)都和控制器相连,每个 OFS 有全网唯一的标识符,由同一个控制器管理的所有 OFS 组成一个域,域由控制器的 IP 地址标识。OFS 在与控制器建立连接时,控制器会把域间 OFS 选择策略加载到 OFS。

[0007] 步骤 2:各 OFS 向邻接 OFS 广播邻接消息(NDM),该邻接消息封装了消息类型、域 ID、物理节点 ID、端口 ID 和端口 IP。OFS 收到从邻接节点发来的 NDM 后将邻居 OFS 信息加入到自己维护的邻接节点表中。邻接节点表由邻接节点域 ID、邻接节点 ID、邻接节点端口 ID、邻接节点端口 IP、与邻接节点相连的本 OFS 的端口 ID 和定时器项组成。

[0008] 步骤 3:每个 OFS 把用 OpenFlow 消息封装的节点信息消息(NIM)发送给控制器。该节点信息消息封装了消息类型、OFS 信息和邻接 OFS 信息。控制器收到各 OFS 发来的封

装了 NIM 的 OpenFlow 消息后,就获得了本域的物理拓扑信息以及与本域相连的其他域的边界 OFS 信息(如果一个域的 OFS 与其他域的 OFS 相连,则该 OFS 为边界 OFS),控制器将本域的物理拓扑存储到本地,把域间的物理拓扑存储至网络视图,每个控制器根据网络视图获取全网域间拓扑。

[0009] 步骤 4:维护域间链路,快速恢复故障链路。该步骤通过以下子步骤来实现:

[0010] (4.1) OFS 周期性地向邻接 OFS 广播 NDM, OFS 收到 NDM 后,提取出物理节点 ID 字段,并与邻接节点表中的邻接节点 ID 项匹配,如果匹配成功,就重置命中表项的定时器;如果匹配不成功,表示有新的 OFS 与之相连,则更新邻接节点表,为刚连入的 OFS 增加一个表项,并由 OFS 发送一个封装了 NIM 的 OpenFlow 消息给控制器,以告知新节点的加入以及网络拓扑的变化。

[0011] (4.2) 每个 OFS 定期检查邻接节点表中的定时器项,如果某个表项的定时器超过预先设定的最大值,表示此 OFS 在一定时间内没有收到表项中 OFS 发来的 NDM,表项中的 OFS 为不可达。

[0012] (4.3) 如果 OFS 检测到某个邻接 OFS 不可达,说明两个 OFS 之间的链路发生了故障。OFS 从邻接节点表中找出邻接节点域 ID 为与不可达 OFS 域 ID 相同的候选表项,如果有多个候选表项,根据域间 OFS 选择策略从多个候选表项中选择一个最佳表项,转到步骤(4.4);若没有找到,转到步骤(4.5)。

[0013] (4.4) OFS 根据最佳表项修改流表:先找出流表项中数据包发出端口为与不可达 OFS 相连的端口的流表项,再修改这些流表项把发出端口修改为最佳表项中的与邻接节点相连的本 OpenFlow 交换机的端口 ID,从而把发往不可达 OFS 的流量定向到与不可达 OFS 在同一个域中的边界 OFS。

[0014] (4.5) OFS 采用洪泛法发送寻找路径消息(FPM)以找到此 OFS 到不可达 OFS 所在域的路径。该寻找路径消息封装了消息类型、不可达 OFS 的域 D_ID 和源 OFS 的域 S_ID 和路径。FPM 每到达一个 OFS 的时候需要把 OFS 的 ID 和收到 FPM 的端口 ID 写入路径,以方便 FPM 回复消息的发送,每从 OFS 的端口发送前,需要把发出端口 ID 写入到路径中,以方便随后流表更新消息的发送。OFS 首先向除了与不可达 OFS 相连的端口外的其他端口发送 FPM,当邻接 OFS 收到 FPM 后,查找邻接节点表以确定邻接节点表中是否存在邻接节点域 ID 为 D_ID 的 OFS,如果找到,根据 FPM 中的路径发送 FPM 回复消息;如果没有找到,OFS 把 FPM 从除了收到 FPM 端口之外的端口发送出去直到找到通往域 D_ID 的一条路径。当源 OFS 收到 FPM 回复消息后,提取出 FPM 回复消息中的路径,若 OFS 收到多条 FPM 消息,则从所有路径中选择出跳数最少的路径为最佳路径,如果有多条跳数相同且最少的路径,则随机从多条跳数最少的路径中选择一条为最佳路径。

[0015] (4.6) OpenFlow 交换机找到发往不可达 OpenFlow 交换机的最佳路径后,往最佳路径发送流表更新消息(FUM),以往最佳路径上的 OpenFlow 交换机增加流表项使发往不可达 OpenFlow 交换机的数据包往最佳路径发送;该流表更新消息封装了消息类型、不可达 OpenFlow 交换机的域 ID、路径和匹配域;路径为从寻找路径回复消息中提取出的最佳路径,匹配域为在 OpenFlow 交换机流表中找出的数据包发出端口为与不可达 OpenFlow 交换机相连的端口的流表项中的匹配域,匹配域有一个或多个;OpenFlow 交换机根据 FUM 中的路径来发送流表更新消息,每到达一个 OpenFlow 交换机的时候,为 OpenFlow 交换机增加流

表项,如果 FUM 中有多个匹配域,则增加多个流表项,流表项的匹配域为流表更新消息中的匹配域,从 FUM 的路径中提取下一跳的发出端口并写入流表项的行为部分,并把该流表项的优先级设为最高。这样就为发往不可达 OpenFlow 交换机所在的域的数据包建立了一条新的路径。

[0016] (4.7)OFS 向控制器发送封装了 NIM 的 OpenFlow 消息报告此 OFS 与不可达 OFS 之间的链路故障,控制器更新拓扑,从而实现域间链路的故障恢复。

[0017] 本发明的有益效果是:本发明提出了一种基于软件定义网络的域间链路快速故障恢复方法。由于新路由的计算不需要通过控制器,所以链路故障能快速恢复,从而减少数据包丢弃的数量。

附图说明

[0018] 图 1 是网络拓扑结构示意图。

[0019] 图 2 是基于软件定义网络的域间链路快速故障恢复方法的流程图。

具体实施方式

[0020] 下面根据附图详细描述本发明,本发明的目的和效果将变得更加明显。

[0021] 本发明基于软件定义网络的域间链路的快速故障恢复方法,包括以下步骤:

[0022] 步骤 1:初始网络中的所有 OFS 都和控制器相连,每个 OFS 有全网唯一的标识符,由同一个控制器管理的所有 OFS 组成一个域,域由控制器的 IP 地址标识。OFS 在与控制器建立连接时,控制器会把域间 OFS 选择策略加载到 OFS。在图 1 中,S1-S10 为 OpenFlow 交换机,S1、S2、S3、S4 组成一个域 AS1,由控制器 C1 控制;S5、S6、S7、S8 组成一个域 AS2,由控制器 C2 控制;S9、S10 组成一个域 AS3,由控制器 C3 控制。

[0023] 步骤 2:所有 OFS 向邻接 OFS 广播邻接消息(NDM)。该邻接消息封装了消息类型、域 ID、物理节点 ID、端口 ID 和端口 IP。OFS 收到邻接消息 NDM 后将该 OFS 信息加入到自己维护的邻接节点表(邻接节点域 ID、邻接节点 ID、邻接节点端口 ID、邻接节点端口 IP、与邻接节点相连的本 OFS 的端口 ID 和定时器)中。

[0024] 步骤 3:每个 OFS 把用 OpenFlow 消息封装的节点信息消息(NIM)发送给控制器。该节点信息消息封装了消息类型、OFS 信息、邻接 OFS 信息。控制器收到各 OFS 发来的封装了 NIM 的 OpenFlow 消息后,就获得了本域的物理拓扑信息以及与本域相连的其他域的边界 OFS 信息,控制器将本域的物理拓扑存储到本地,把域间的物理拓扑存储至网络视图,每个控制器根据网络视图获取全网域间拓扑。

[0025] 步骤 4:维护域间链路,快速恢复故障链路。该步骤通过以下子步骤来实现:

[0026] (4.1) OFS 每隔 30s 向邻接 OFS 广播 NDM, OFS 收到 NDM 后,提取出物理节点 ID 字段,并与邻接节点表中的邻接节点 ID 项匹配,如果匹配成功,就重置命中表项的定时器。如果匹配不成功,表示有新的 OFS 与之相连,则更新邻接节点表,为刚连入的 OFS 增加一个表项,并由 OFS 发送一个封装了 NIM 的 OpenFlow 消息给控制器,以告知新节点的加入以及网络拓扑的变化。

[0027] (4.2) 每个 OFS 定期检查邻接节点表中的定时器项,如果某个表项的定时器超过 120s,表示此 OFS 在一定时间内没有收到表项中 OFS 发来的 NDM,表项中的 OFS 为不可达。

[0028] (4.3) 如果 OFS 检测到某个邻接 OFS 不可达,说明两个 OFS 之间的链路发生了故障。OFS 从邻接节点表中找出邻接节点域 ID 为与不可达 OFS 域 ID 相同的候选表项,如果有多个候选表项,根据域间 OFS 选择策略从多个候选表项中选择一个最佳表项,转到步骤(4.4);若没有找到,转到步骤(4.5)。

[0029] (4.4) OFS 根据最佳表项修改流表:先找出流表项中数据包发出端口为与不可达 OFS 相连的端口的流表项,再修改这些流表项把发出端口修改为最佳表项中的与邻接节点相连的本 OFS 的端口 ID,从而把发往不可达 OFS 的流量定向到与不可达 OFS 在同一个域中的边界 OFS。在图 1 中,如果 S5 检测到 S3 不可达,但是 S5 在本地的邻接节点表中能找到与 S3 在同一个域 AS1 的 S4,于是修改流表项,把发往 S3 的流量定向到 S4。

[0030] (4.5) OFS 采用洪泛法发送寻找路径消息(FPM),以找到此 OFS 到不可达 OFS 所在域的路径。该寻找路径消息封装了消息类型、不可达 OFS 的域 D_ID、源 OFS 的域 S_ID 和路径。FPM 每到达一个 OFS 的时候需要把 OFS 的 ID 和收到 FPM 的端口 ID 写入路径,以方便 FPM 回复消息的发送,每从 OFS 的端口发送前,需要把发出端口 ID 写入到路径中,以方便随后流表更新消息的发送。OFS 首先向除了与不可达 OFS 相连的端口外的其他端口发送 FPM,当邻接 OFS 收到 FPM 后,查找邻接节点表确定邻接节点表中是否存在邻接节点域 ID 为 D_ID 的 OFS,如果找到,根据 FPM 中的路径发送 FPM 回复消息;如果没有找到,OFS 把 FPM 从除了收到 FPM 端口之外的端口发送出去直到找到通往域 D_ID 的一条路径。当源 OFS 收到 FPM 回复消息后,提取出 FPM 回复消息中的路径,若 OFS 收到多条 FPM 消息,则从所有路径中选择出跳数最少的路径为最佳路径,如果有多条跳数相同且最少的路径,则随机从多条跳数最少的路径中选择一条为最佳路径。在图 1 中,如果 S3 检测到 S5 不可达而且在 S3 的邻接节点表中找不到 S5 所在域 AS2 中的 OFS,于是 S3 向除了与 S5 相连的端口以外的其他端口发送 FPM,S2 收到 S3 发来的 FPM 后,由于在 S2 的邻接节点表中不存在域 AS2 的 OFS,S2 就把 S2 和收到消息的端口加入到 FPM 的路径部分中再次发送出去,就这样一直发送下去,到最后 S4 和 S9 收到了 FPM,在 S4 和 S9 的邻接节点表中都存在域 AS2 中的 OFS,于是 S4 和 S9 都会送 FPM 回复消息报告到达 S5 所在的域的路径已经找到。

[0031] (4.6) OFS 找到发往不可达 OFS 的最佳路径后,往最佳路径发送流表更新消息(FUM),以往最佳路径上的 OFS 增加流表项使发往不可达 OFS 的数据包往最佳路径发送;该流表更新消息封装了消息类型、不可达 OFS 的域 ID、路径和匹配域;路径为从寻找路径回复消息中提取出的最佳路径,匹配域为在 OpenFlow 交换机流表中找出的数据包发出端口为与不可达 OFS 相连的端口的流表项中的匹配域,匹配域有一个或多个;OFS 根据 FUM 中的路径来发送流表更新消息,每到达一个 OFS 的时候,为 OFS 增加流表项,如果有多个匹配域,则增加多个流表项,流表项的匹配域为流表更新消息中的匹配域,从 FUM 的路径中提取下一跳的发出端口并写入流表项的行为部分,并把该流表项的优先级设为最高。这样就为发往不可达 OpenFlow 交换机所在的域的数据包建立了一条新的路径。在图 1 中,S3 找到到达 S5 所在的域的最佳路径为 S3、S2、S1、S4、S5 后,S3 往最佳路径发送 FUM,在最佳路径上的 OpenFlow 交换机增加流表项,使发往 S5 的数据包往最佳路径发送。

[0032] (4.7) OFS 向控制器发送封装了 NIM 的 OpenFlow 消息报告此 OFS 与不可达 OFS 之间的链路故障,控制器更新拓扑。

[0033] 本发明提出的域间链路故障恢复方法在域间链路出现故障的时候, OpenFlow 交

交换机直接根据存储在本地邻接节点表中的信息为发往不可达 OpenFlow 交换机的流量选择一条路径,而不用控制器的参与,不占用 OpenFlow 交换机与控制器之间有限的带宽,恢复速度快,减少了网络的丢包率。

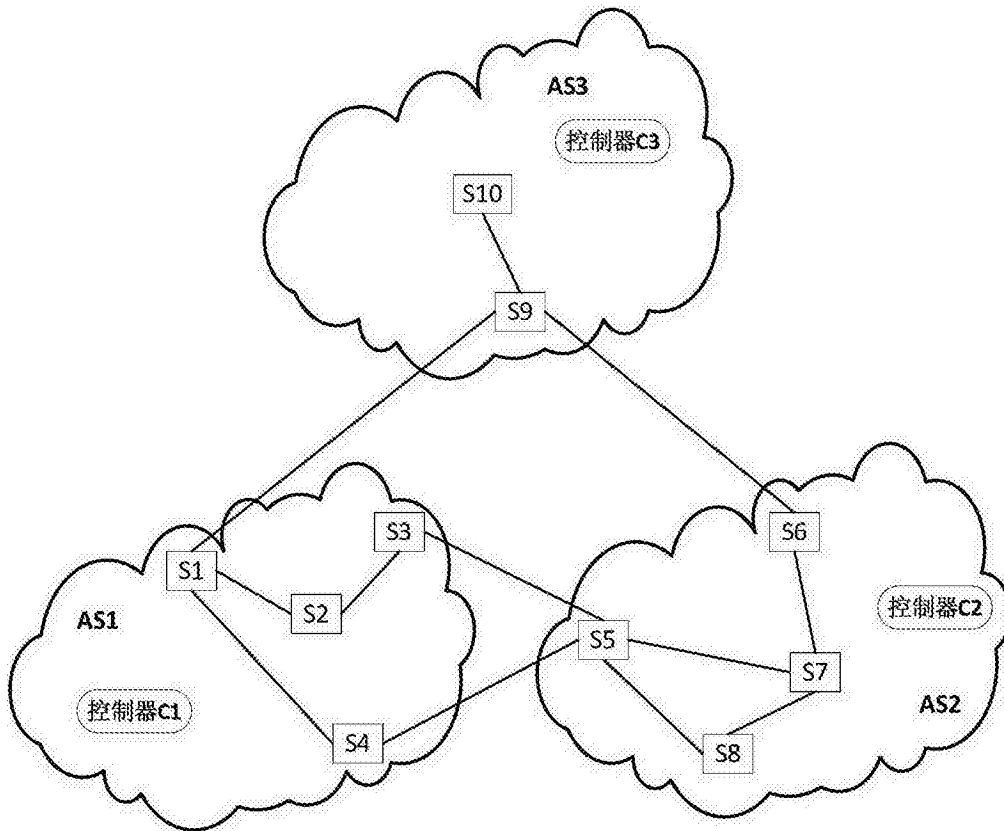


图 1

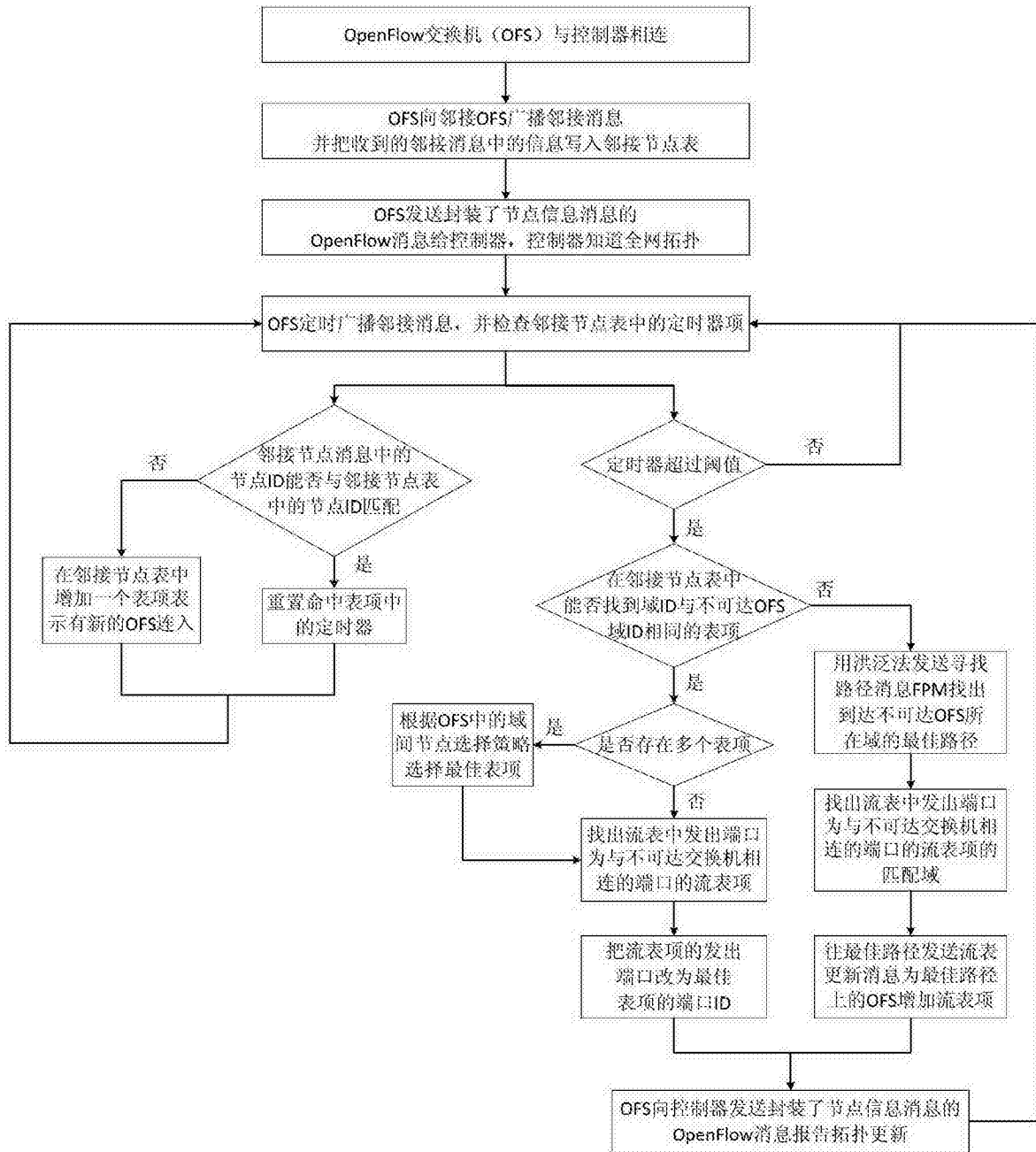


图 2