H3C SR6600/SR6600-X 路由器

OpenFlow 配置指导（V7）

杭州华三通信技术有限公司
http://www.h3c.com.cn

资料版本：6W101-20170512
产品版本：SR6600_SR6600X-CMW710-R7607

前 言

本配置指导主要介绍 OpenFlow 协议工作原理及相关配置。OpenFlow 允许控制器直接访问和操作网络设备的转发平面，将控制平面和数据平面分离。交换机依据控制器下发的流表（Flow Table）对报文进行匹配和转发，在同一个流表中按照流表项的优先级大小进行匹配。一个 OpenFlow 交换机可以包含一个或者多个流表。

前言部分包含如下内容：

读者对象

本手册主要适用于如下工程师：

- 网络规划人员
- 现场技术支持与维护人员
- 负责网络配置和维护的网络管理员

本书约定

1. 命令行格式约定

格　式	意　义
粗体	命令行关键字（命令中保持不变、必须照输的部分）采用**加粗**字体表示。
斜体	命令行参数（命令中必须由实际值进行替代的部分）采用*斜体*表示。
[ ]	表示用“[ ]”括起来的部分在命令配置时是可选的。
{ x | y | ... }	表示从多个选项中仅选取一个。
[x | y | ...]	表示从多个选项中选取一个或者不选。
{ x | y | ... }*	表示从多个选项中至少选取一个。
[x | y | ...]*	表示从多个选项中选取一个、多个或者不选。
&<1-n>	表示符号&前面的参数可以重复输入1～n次。
#	由“#”号开始的行表示为注释行。

2. 图形界面格式约定

格　式	意　义
< >	带尖括号“< >”表示按钮名，如“单击<确定>按钮”。
[ ]	带方括号“[ ]”表示窗口名、菜单名和数据表，如“弹出[新建用户]窗口”。
/	多级菜单用“/”隔开。如[文件/新建/文件夹]多级菜单表示[文件]菜单下的[新建]子菜单下的[文件夹]菜单项。

3. 各类标志

本书还采用各种醒目标志来表示在操作过程中应该特别注意的地方，这些标志的意义如下：

 警告	该标志后的注释需给予格外关注，不当的操作可能会对人身造成伤害。
 注意	提醒操作中应注意的事项，不当的操作可能会导致数据丢失或者设备损坏。
 提示	为确保设备配置成功或者正常工作而需要特别关注的操作或信息。
 说明	对操作内容的描述进行必要的补充和说明。
 窍门	配置、操作、或使用设备的技巧、小窍门。

4. 图标约定

本书使用的图标及其含义如下：

	该图标及其相关描述文字代表一般网络设备，如路由器、交换机、防火墙等。
	该图标及其相关描述文字代表一般意义下的路由器，以及其他运行了路由协议的设备。
	该图标及其相关描述文字代表二、三层以太网交换机，以及运行了二层协议的设备。
	该图标及其相关描述文字代表无线控制器、无线控制器业务板和有线无线一体化交换机的无线控制引擎设备。
	该图标及其相关描述文字代表无线接入点设备。
	该图标及其相关描述文字代表无线Mesh设备。
	该图标代表发散的无线射频信号。
	该图标代表点到点的无线射频信号。
	该图标及其相关描述文字代表防火墙、UTM、多业务安全网关、负载均衡等安全设备。

	该图标及其相关描述文字代表防火墙插卡、负载均衡插卡、NetStream插卡、SSL VPN插卡、IPS插卡、ACG插卡等安全插卡。

5. 端口编号示例约定

本手册中出现的端口编号仅作示例，并不代表设备上实际具有此编号的端口，实际使用中请以设备上存在的端口编号为准。

资料获取方式

您可以通过H3C网站（www.h3c.com.cn）获取最新的产品资料：

H3C 网站与产品资料相关的主要栏目介绍如下：

- [服务支持/文档中心]：可以获取硬件安装类、软件升级类、配置类或维护类等产品资料。
- [产品技术]：可以获取产品介绍和技术介绍的文档，包括产品相关介绍、技术介绍、技术白皮书等。
- [解决方案]：可以获取解决方案类资料。
- [服务支持/软件下载]：可以获取与软件版本配套的资料。

技术支持

用户支持邮箱：service@h3c.com

技术支持热线电话：400-810-0504（手机、固话均可拨打）

网址：http://www.h3c.com.cn

资料意见反馈

如果您在使用过程中发现产品资料的任何问题，可以通过以下方式反馈：

E-mail：info@h3c.com

感谢您的反馈，让我们做得更好！

目 录

1 OpenFlow

说明

本文中的交换机均指代支持 OpenFlow 功能的路由器。

1.1 OpenFlow简介

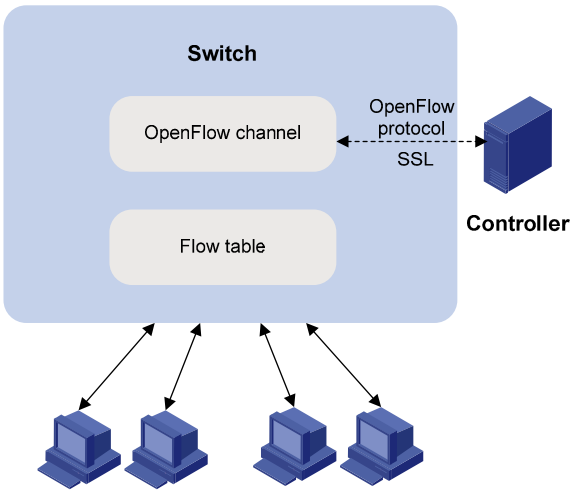
1.1.1 OpenFlow概述

OpenFlow 是 SDN（Software Defined Network，软件定义网络）架构中定义的一个控制器与转发层之间的通信接口标准。OpenFlow 允许控制器直接访问和操作网络设备的转发平面，这些网络设备可能是物理上的，也可能是虚拟的。

OpenFlow 的思想是分离控制平面和数据平面，二者之间使用标准的协议通信；数据平面采用基于流的方式进行转发。

OpenFlow网络由OpenFlow设备（Switch）和控制器（Controller）通过安全通道（OpenFlow channel）组成，如 图 1-1 所示。Switch与Controller通过TLS或者TCP建立安全通道，进行OpenFlow消息交互，实现表项下发、查询以及状态上报等功能。下文如果没有特殊说明，交换机指的就是OpenFlow设备。

图1-1 OpenFlow 网络组成



1.1.2 OpenFlow Switch类型

OpenFlow Switch 有下面两种：

- OpenFlow-Only Switch：仅支持 OpenFlow 转发。
- OpenFlow-Hybrid Switch：既支持 OpenFlow 转发，也支持正常转发。

1.1.3 OpenFlow接口

OpenFlow 接口有下面三类：

- 物理接口：比如以太网接口。可以作为入接口和出接口。
- 逻辑接口：比如聚合接口、Tunnel 接口等。可以作为入接口和出接口。
- 保留接口：由转发动作定义的接口，实现OpenFlow转发功能。除Any接口外，其他接口都可以作为出接口，仅Controller和Local可以作为入接口。具体类型请参见 表 1-1。

表1-1 保留接口类型

类型	说明
ALL	报文从所有接口发送
Controller	报文上送控制器
Table	报文重新进入流表进行匹配
In Port	报文从入接口转发
Any	接口通配描述，不能作为入接口以及出接口
Local	报文上送本地CPU
Normal	报文正常转发
Flood	报文广播发送

1.1.4 OpenFlow实例

OpenFlow 支持多实例。每个 OpenFlow 实例可以单独连接控制器，相当于一台独立的交换机，根据控制器下发的流表项指导流量转发。下文如果没有特殊说明，交换机指的就是一个 OpenFlow 实例。

1. 全局实力

全局实例即对设备上所有流量进行 OpenFlow 处理。

2. 实例激活

需要激活实例后，OpenFlow 才能将设备的支持能力、当前的接口信息等设备信息上报给控制器，控制器才能够下发流表项指导转发。

3. 实例所属接口

OpenFlow 协议规定需要将接口信息上报给控制器，这些接口包括物理接口、逻辑接口以及保留接口中的 Local。

1.1.5 OpenFlow流表

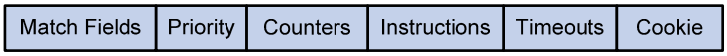
1. 流表项组成

OpenFlow 通过流表（Flow Table）来匹配和处理报文，在同一个流表中按流表项的优先级进行先后匹配。一台交换机上可以包含一个或者多个流表。

流表分为两种类型：

- MAC-IP流表：通过MAC地址表和FIB表实现。只能匹配目的MAC地址、VLAN以及目的IP地址，动作也仅支持修改目的MAC地址、源MAC地址、VLAN、TUNNEL ID以及指定出接口。具体请参见 1.10 附录 B MAC-IP流表。
- Extensibility 流表：扩展流表，使用 TCAM（Ternary Content Addressable Memory，三态内容寻址存储器）或者软件实现。

图1-2 流表项结构



流表项如 图 1-2 所示：

- Match Fields：匹配规则。可以匹配入接口、报文头等字段。
- Priority：优先级。定义流表项之间的匹配顺序，优先级高的先匹配。
- Counters：统计计数。统计有多少个报文和字节匹配到该流表项。
- Instructions：动作指令集。定义匹配到该流表项的报文需要进行的处理。流表项动作指令集是对动作进行操作，流表项的动作有两种执行类型：
  - 动作集（Action Set）：一系列动作的组合，不会立刻修改报文内容，直到报文不再需要进入下一级流表，动作集里每种动作仅能存在一个，并且按照 表 1-2 从上到下的顺序执行。
  - 动作序列（Action List）：需要立即执行的一系列动作，其动作内容与 Action Set 相同，但是会立即修改报文的内容，其效果是累加的，并且执行顺序是按照下发的顺序执行的。

表1-2 动作指令集定义

Instruction	处理
Meter	对匹配到流表项的报文进行限速
Apply-Actions	立即执行动作序列中的动作
Clear-Actions	清除动作集中的所有动作
Write-Actions	更改动作集中的所有动作
Write-Metadata	更改流表间数据，在支持多级流表时使用
Goto-Table	进入下一级流表

具体动作类型如 表 1-3 所示。

表1-3 动作类型（1.3.1 版本）

动作名称	可选/必选	描述
Output	必选	Output动作转发报文到特定的OpenFlow端口，比如物理端口，逻辑端口以及OpenFlow保留端口
Drop	必选	并没有直接的动作来代表Drop，当动作集中不含有Output指令时，报文会被丢弃。通常来说空指令集，空动作集或者执行清空动作集后，报文会被丢弃
Group	必选	将报文转交给指定的Group处理，该动作的确切含义由Group的类型定义

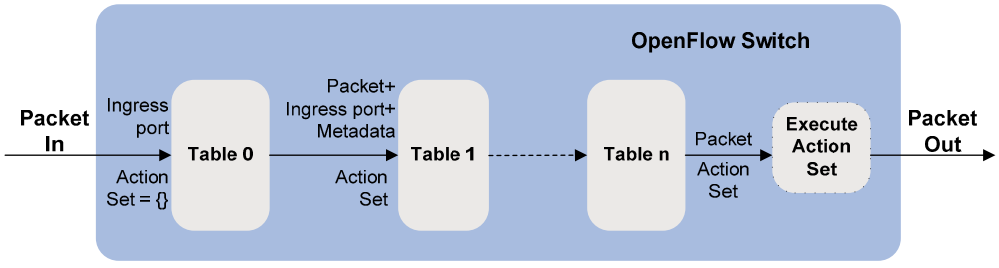
动作名称	可选/必选	描述
Set-Queue	可选	Set-Queue动作为报文指定队列ID。当报文被转发到特定端口时，队列ID通常被用于基本QoS
Push-Tag/Pop-Tag	可选	Push-Tag和Pop-Tag动作适用于VLAN头、MPLS头
Set-Field	可选	Set-Field动作可以识别报文字段的类型，并且可以修改该字段的值。Set-Field动作通常只适用于最外层的字段（比如当内外层均有VLAN tag时，该动作只修改最外层的VLAN Tag）
Change-TTL	可选	Change-TTL动作可以改变报文中IPv4的TTL，IPv6的Hop Limit或者MPLS的TTL。同样，Change-TTL也只适用于最外层的字段。该动作可以设置TTL（TTL必须已经存在）、减少TTL、TTL值拷贝（inwards/outwards）

- Timeouts：超时时间。包括了 idle time 和 hard time。
  - idle time：在 idle time 时间内，如果没有报文匹配到该流表项，则此流表项被删除。
  - hard time：在 hard time 时间超时后，无论是否有报文匹配到该流表项，此流表项都会被删除
- Cookie：控制器下发的流表项的标识。

2. 流表处理流程

如 图 1-3 所示，当报文进入交换机后，必须从流表ID最小的流表开始依次匹配；流表可以按次序从小到大越级跳转，但不能从某一流表向前跳转至流表ID更小的流表。一旦在某个流表匹配到后，会更新此报文的动作集（允许被下一级流表覆盖），到了最后一个流表后，所有的动作会被执行，此时报文的内容会被修改，指定出接口进行转发。如果在某个流表处理时，处理指令包含Action List，则报文的一份拷贝立即执行Action List而不需要在最后一个流表处理结束后再执行。

图1-3 OpenFlow 转发示意图



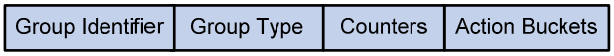
3. Table Miss表项

每个流表都包含一个 Table Miss 流表项，该表项用于定义在流表中没有匹配的报文的处理方式，该表项的匹配域为通配，即匹配任何报文，优先级为 0，动作指令与正常表项相同。

1.1.6 Group Table

Group Table 由 Group 表项组成，Group 表项被流表项所引用，提供额外的报文转发功能。

图1-4 Group 表项结构



- Group Identifier：Group ID，用于识别 Group，32bits。
- Group Type：Group 类型。
  - All：执行所有动作桶，用于组播或者广播。
  - Select：自动选择一个动作桶执行。
  - Indirect：始终执行固定的动作桶。
  - Fast failover：始终执行第一个活跃的动作桶。
- Counters：当报文被 Group 处理时，更新计数器。
- Action Buckets：一个由动作桶组成的有序列表。每个动作桶由许多动作组成。

1.1.7 Meter Table

Meter Table 由 Meter 表项组成，Meter 表项被流表项所引用，为所有引用 Meter 表项的流表项提供报文限速的功能。

图1-5 Meter 表项结构

- Meter Identifier：Meter ID，用于识别 meter，32bits。
- Meter Bands：一个 Meter 表项可以包含一个或者多个 Meter Bands，每个 Meter Band 定义了速率以及动作。当报文的速率超过了某些 Meter Band，根据这些 Meter Band 中速率最大的那个定义的动作进行处理。
- Counters：当报文被 Meter 处理时，更新计数器。

图1-6 Meter Bands 结构

- Band Type：Band 类型，定义报文如何处理。为可选，可使用丢弃（drop），即报文高于该速率会被丢弃；以及重新标记 DSCP（dscp remark）。
- Rate：Meter 用于选择 Band 的最低速率，即报文速率高于该速率并最接近该速率，该 Band 将被应用。
- Counters：当 Band 处理报文时，更新计数器。
- Type Specific arguments：某些 Band 含有的特定参数。

1.1.8 OpenFlow channel

交换机与控制器通过 TLS 或者 TCP 建立 Channel，进行 OpenFlow 消息交互，实现表项下发、查询以及状态上报等功能。

OpenFlow 协议中定义了三种消息类型 Controller to Switch 消息、异步消息和同步消息，每种报文类型都有很多子类型。

1. Controller to Switch消息

Controller to Switch 消息是指由控制器产生并发送到交换机，用来查询交换机的消息，可以不需要交换机响应。这些消息主要由控制器用来对交换机进行状态查询和修改配置等操作。

表1-4 Controller to Switch 消息（1.3.1 版本）

子类型	描述
Features	用于控制器发送请求来了解交换机的能力，交换机必须回应该报文
Configuration	用于控制器设置，查询交换机的配置，交换机只有在控制器查询时回应
Modify-State	用于管理交换机的状态，如流表项和端口状态。该命令主要用于增加、删除、修改、交换机内的流表表项，组表表项以及交换机端口的属性
Multipart	用于控制器收集交换机各方面的信息，例如当前配置，统计信息等
Packet-Out	用于通过交换机特定端口发送报文，这些报文可以通过Packet-In消息触发，也可以通过控制器直接发送。通常Packet-Out消息包含整个之前接收到的Packet-In消息所携带的报文或者buffer ID（用于指示存储在交换机内的特定报文）。这个消息需要包含一个动作列表，当交换机收到该动作列表后会对Packet-Out消息所携带的报文执行该动作列表。如果动作列表为空，Packet-Out消息所携带的报文将被交换机丢弃
Barrier	用于确认之前下发动作是否成功。控制器发送Barrier请求消息，当交换机确认之前下发的流表等操作都已经成功时会回复Barrier应答消息
Role-Request	用于设定或查询OpenFlow channel的角色。通常用于交换机和多个控制器相连的情况
Asynchronous-Configuration	控制器使用该报文设定异步消息过滤器来接收其只希望接收到的异步消息报文，或者向交换机查询该过滤器。通常用于交换机和多个控制器相连的情况

2. 异步（Asynchronous）消息

异步（Asynchronous）消息是由交换机发送给控制器，用来通知交换机上发生的某些异步事件的消息。例如，当某一条规则因为超时而被删除时，交换机将自动发送一条 Flow-Removed 消息通知控制器，以方便控制器作出相应的操作，如重新设置相关规则等。

表1-5 异步消息（1.3.1 版本）

子类型	描述
Packet-In	转移报文的控制权到控制器。对于所有通过匹配流表项或者Table Miss后转发到保留端口Controller端口的报文均要通过Packet-in消息送到控制器。也有部分其他流程，如TTL检查等，也需要通过该消息和控制器交互。Packet-In既可以携带报文，也可以通过在交换机内部设置报文的Buffer来仅携带报文头以及其Buffer ID传输给控制器。控制器在接收到Packet-In消息后会对其接收到的报文或者报文头和Buffer ID进行处理，并发回Packet-out消息通知交换机如何处理该报文
Flow-Removed	通知控制器将某个流表项从流表的移除。通常该消息在控制器发送删除流表项的消息或者流表项的两个定时器其中之一超时时产生

子类型	描述
Port-Status	通知控制器端口状态或设置的改变
Error	通知控制器交换机出现的问题或错误

3. 对称（Symmetric）消息

对称（Symmetric）消息，就是双向对称的消息，主要用来建立连接和检测对方是否在线等。

表1-6 对称消息（1.3.1 版本）

子类型	描述
Hello	当连接启动时交换机和控制器会发送Hello交互
Echo	用于验证控制器与交换机之间连接的存活，控制器和交换机都会发送Echo request/reply消息，而且对于接受到的Echo request消息必须能返回Echo reply消息。也可用于测量控制器与交换机之间链路的延迟和带宽
Experimenter	为将来新加入的特性预留的消息

1.1.9 协议规范

- OpenFlow Switch Specification Version 1.3.3

1.1.10 OpenFlow配置限制和指导

FIP-600 业务板和 SAP-4EXP 业务板不支持 OpenFlow 功能。

1.2 OpenFlow配置任务简介

表1-7 OpenFlow 配置任务简介

配置任务			说明	详细配置
配置OpenFlow实例	创建OpenFlow实例		必选	1.3.1
	配置OpenFlow实例的基本能力	配置全局实例	必选	1.3.2 1.
		配置流表ID	可选	1.3.2 2.
		配置控制器模式	可选	1.3.2 3.
		配置Extensibility表的流表项的最大个数	可选	1.3.2 4.
		配置Datapath ID	可选	1.3.2 5.
		配置SSL服务器	可选	1.3.2 6.
		配置缺省的table miss动作	可选	1.3.2 7.
		配置禁止上送Controller的端口类型	可选	1.3.2 8.
	激活OpenFlow实例		必选	1.3.3
配置连接控制器	配置主连接		必选	1.4.1

配置任务		说明	详细配置
	配置辅助连接	可选	1.4.2
	配置连接中断模式	可选	1.4.3
配置OpenFlow定时器		可选	1.5
配置支持动态MAC地址		可选	1.6

1.3 配置OpenFlow实例

1.3.1 创建OpenFlow实例

表1-8 创建 OpenFlow 实例

操作	命令	说明
进入系统视图	**system-view**	-
创建OpenFlow实例，并进入OpenFlow实例视图/创建	**openflow instance** *instance-id*	缺省情况下，不存在OpenFlow实例
（可选）配置OpenFlow实例描述	**description** *text*	缺省情况下，未配置OpenFlow实例的描述信息

1.3.2 配置OpenFlow实例的基本能力

下面配置用于定义 OpenFlow 实例的基本能力，交换机在与控制器建立连接后会上报这些基本能力，控制器根据这些能力下发表项。

1. 配置OpenFlow全局实例

表1-9 配置 OpenFlow 全局实例

操作	命令	说明
进入系统视图	**system-view**	-
进入OpenFlow实例视图	**openflow instance** *instance-id*	-
配置OpenFlow全局实例	**classification global**	缺省情况下，未配置OpenFlow全局实例

2. 配置流表ID

说明

- MAC-IP 表仅支持一个；Extensibility 表最多可以支持 254 个。
- 输入的 Extensibility 流表 ID 要大于 MAC-IP 流表 ID。

OpenFlow 实例中可以配置多个流表。

表1-10 配置流表 ID

操作	命令	说明
进入系统视图	**system-view**	-
进入OpenFlow实例视图	**openflow instance** *instance-id*	-
配置流表ID	**flow-table** { **extensibility** *table-id* | **mac-ip** *table-id* }&<1-n>	缺省情况下，实例包含了一个Extensibility流表，流表ID为0

3. 配置控制器模式

支持两种模式与控制器建立连接。

- Single 模式：同一时刻，仅与一个控制器建立连接，配置的多个控制器之间互为备份。当且仅当当前的连接断开后，交换机会连接下一个控制器，直到连接成功。
- Multiple 模式：同一时刻，允许与多个控制器建立连接。交换机会连接配置的所有控制器，在与某个控制器连接失败或者断开连接后，在重连时间间隔后重新与之进行连接，直到连接成功。

表1-11 配置控制器模式

操作	命令	说明
进入系统视图	**system-view**	-
进入OpenFlow实例视图	**openflow instance** *instance-id*	-
配置实例内的多个控制器的连接模式	**controller mode** { **multiple** | **single** }	缺省情况下，连接模式为Multiple

4. 配置Extensibility表的流表项的最大个数

在 OpenFlow 实例中允许定义 Extensibility 流表支持的表项最大值，当控制器下发的流表表项个数超过最大值的时候，向控制器返回失败。

本配置是针对单个 Extensibility 流表的，当存在多个 Extensibility 流表时，每个流表都单独受此配置限制。

表1-12 配置 Extensibility 表的流表项的最大个数

操作	命令	说明
进入系统视图	**system-view**	-
进入OpenFlow实例视图	**openflow instance** *instance-id*	-
配置Extensibility表的流表项的最大个数	**flow-entry max-limit** *limit-value*	缺省情况下，Extensibility表的流表项的最大个数为65535

5. 配置Datapath ID

Datapath ID 用来在唯一标识交换机（OpenFlow 实例），不同交换机（OpenFlow 实例）的 Datapath ID 不能相同，配置时请注意。

表1-13 配置 Datapath ID

操作	命令	说明
进入系统视图	**system-view**	-
进入OpenFlow实例视图	**openflow instance** *instance-id*	-
配置Datapath ID	**datapath-id** *id*	缺省情况下，OpenFlow实例的Datapath ID由实例ID与设备桥MAC组成，其中前16个比特为实例ID，后48个比特为设备桥MAC

6. 配置SSL服务器

说明

关于 SSL 的介绍和基本功能配置，请参见“安全配置指导”中的“SSL”。

OpenFlow 实例启动 SSL 服务器之后，SSL 客户端可以连接 SSL 服务器，并使用这个连接的通道完成 OpenFlow 协议的通信。

没有启动 SSL 服务器时，设备作为 TCP/SSL 客户端主动连接控制器（SSL 服务器，需要相应配置）；启动 SSL 服务器之后，设备作为 SSL 服务器端被动等待控制器（SSL 客户端）连接。

表1-14 配置 SSL 服务器

操作	命令	说明
进入系统视图	**system-view**	-
进入OpenFlow实例视图	**openflow instance** *instance-id*	-
配置SSL服务器	**listening port** *port-number* **ssl** *ssl-policy-name*	缺省情况下，未配置SSL服务器 不能通过重复执行本命令修改OpenFlow实例启动的SSL服务器。如需修改，请先通过**undo listening port**命令删除OpenFlow实例启动的SSL服务器，再执行**listening port**命令重新启动SSL服务器

7. 配置缺省的table miss动作

如果配置了本命令，则实例下所有流表的缺省 table miss 动作为走正常二三层转发。如果没有配置本命令，则实例下所有流表的缺省 table miss 动作为丢弃。

表1-15 配置缺省的 table miss 动作

操作	命令	说明
进入系统视图	**system-view**	-

操作	命令	说明
进入OpenFlow实例视图	**openflow instance** *instance-id*	-
配置缺省的table miss动作	**default table-miss permit**	缺省情况下，缺省的table miss动作为丢弃

8. 配置禁止上送Controller的端口类型

配置该功能后，交换机不再向控制器上送对应 VLAN 或 VSI 端口的信息。

表1-16 配置禁止端口上送

操作	命令	说明
进入系统视图	**system-view**	-
进入OpenFlow实例视图	**openflow instance** *instance-id*	-
配置禁止上送Controller的端口类型	**forbidden port** { **vlan-interface** | **vsi-interface** } *	缺省情况下，所有接口都上送Controller

1.3.3 激活OpenFlow实例

此功能用于激活实例。如果实例已经与控制器建立了连接，此时修改了实例基本能力配置并重新激活，交换机会断开与所有控制器的连接，清除已经下发的流表，根据当前的配置更新 OpenFlow 实例的能力集，重新与控制器建立连接。

表1-17 激活 OpenFlow 实例

操作	命令	说明
进入系统视图	**system-view**	-
进入OpenFlow实例视图	**openflow instance** *instance-id*	-
激活OpenFlow实例	**active instance**	缺省情况下，未激活OpenFlow实例

1.4 配置连接Controller

一个OpenFlow交换机可以与多个Controller建立连接，初始连接时，多个Controller的角色相同，权限相同，Controller可以通过OpenFlow消息设置本Controller的角色，各种角色的权限如 表 1-18 所示。

表1-18 Controller 角色

角色	权限
Master	处于该角色的Controller拥有全部权限，可以下发流表项，查询统计信息，接收设备上报的状态信息，在多个Controller中仅能有一个Controller是Master角色
Equal	处于该角色的Controller同样拥有全部权限，相对于Master角色，唯一不同的是可以有多个Controller处于Equal角色

角色	权限
Slave	处于该角色的Controller仅拥有部分权限，Controller to switch消息中不能下发流表项，Group表项以及Meter表项，不允许修改接口配置和设备配置，不允许执行Packet Out操作。异步消息中，缺省情况下设备不会上送Flow Remove消息和Packet In消息，仅能上送接口状态变化消息，但是异步消息的上送能力可以通过Controller的设置异步消息进行修改

1.4.1 配置主连接

交换机可以连接多个控制器，但仅允许与每个控制器建立一个主连接，一般用于控制消息的处理（下发流表项、获取数据、信息上报等），需要使用 TCP/SSL 保持可靠的连接。

如果交换机与控制器之间存在多条路由可达，当交换机进行主备倒换时或者重启后，希望通过原来的路由重新建立连接，而不是新选择的路由，此时可以通过配置交换机与控制器连接的源 IP 地址来实现。

表1-19 配置主连接

操作	命令	说明
进入系统视图	**system-view**	-
进入OpenFlow实例视图	**openflow instance** *instance-id*	-
配置主连接	**controller** *controller-id* **address** { **ip** *ipv4-address* | **ipv6** *ipv6-address* } [**port** *port-number*] [**local address** { **ip** *local-ipv4-address* | **ipv6** *local-ipv6-address* } [**port** *local-port-number*]] [**ssl** *ssl-policy-name*] [**vrf** *vrf-name*]	缺省情况下，不存在主连接 建议控制器的IP地址使用单播地址，否则交换机和控制器之间可能无法建立连接 建议源IP地址使用单播地址，且该IP地址是OpenFlow实例下一个端口的IP地址，否则交换机和控制器之间可能无法建立连接

1.4.2 配置辅助连接

OpenFlow 通道可以由一个主连接和多个辅助连接组成。辅助连接用于提高控制器和 OpenFlow 交换机的通信能力。辅助连接的目的地址和接口号可以和主连接不一致。

提示

- 辅助连接命令行和主连接命令行不做额外的检查处理。如果配置冲突，辅助连接将无法建立。
- 地址或接口号未指定时，和主连接一致。

表1-20 配置辅助连接

操作	命令	说明
进入系统视图	**system-view**	-
进入OpenFlow实例视图	**openflow instance** *instance-id*	-

操作	命令	说明
配置实例辅助连接	**controller** *id* **auxiliary** *auxiliary-id* **transport** { **tcp** | **udp** | **ssl** *ssl-policy-name* } [**address** { **ip** *ipv4-address* | **ipv6** *ipv6-address* }] [**port** *port-number*]	缺省情况下，不存在辅助连接

1.4.3 配置连接中断模式

一旦交换机与所有控制器断开连接，则交换机必须进入连接中断模式，模式分为两种：

- Secure 模式：连接断开后，交换机根据流表项转发。不主动删除控制器下发的表项，而是等待表项超时后进行删除，一旦连接建立成功，未超时的表项依然存在。
- Standalone 模式：连接断开后，交换机正常转发。

如果交换机与控制器重新连接成功，则继续作为 OpenFlow 设备根据流表项进行转发。

表1-21 配置连接中断模式

操作	命令	说明
进入系统视图	**system-view**	-
进入OpenFlow实例视图	**openflow instance** *instance-id*	-
配置连接中断模式	**fail-open mode** { **secure** | **standalone** }	缺省情况下，OpenFlow实例建立时，缺省为Secure模式，且为该实例下发Table Miss表项（动作为drop）

1.5 配置OpenFlow定时器

控制器和交换机都会发送 Echo request/reply 报文，用于验证连接是否正常。

连接检测定时器用来定义发送 Echo request 报文时间间隔，当超过三次 Echo request 报文发送并且没有收到 Echo reply 报文，则交换机与控制器的连接断开。

重连定时器用来定义 OpenFlow 实例与控制器断开连接后下次开始重新连接的时间。

表1-22 配置 OpenFlow 定时器

操作	命令	说明
进入系统视图	**system-view**	-
进入OpenFlow实例视图	**openflow instance** *instance-id*	-
配置连接检测定时器	**controller echo-request interval** *interval*	缺省情况下，发送Echo request报文的时间间隔为5秒
配置重连定时器	**controller connect interval** *interval*	缺省情况下，OpenFlow实例与控制器重连尝试的时间间隔为60秒

1.6 配置支持动态MAC地址

此功能仅在支持 MAC-IP 流表情况下，决定是否支持控制器在查询或者删除流表项时包含动态 MAC 地址。

表1-23 配置支持动态 MAC 地址

操作	命令	说明
进入系统视图	**system-view**	-
进入OpenFlow实例视图	**openflow instance** *instance-id*	-
配置支持动态MAC地址	**mac-ip dynamic-mac aware**	缺省情况下，不支持动态MAC地址，即忽略控制器下发的此类消息

1.7 OpenFlow显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 OpenFlow 的运行情况。

表1-24 OpenFlow 显示和维护

操作	命令
显示OpenFlow实例的详细信息	**display openflow instance** [*instance-id*]
显示OpenFlow实例的流表信息	**display openflow instance** *instance-id* **flow-table** [*table-id*]
显示OpenFlow实例的控制器信息	**display openflow instance** { *instance-id* { **controller** [*controller-id*] | **listened** } }
显示OpenFlow实例的Group信息	**display openflow instance** *instance-id* **group** [*group-id*]
显示OpenFlow实例的Meter信息	**display openflow instance** *instance-id* **meter** [*meter-id*]
显示OpenFlow实例的概要信息	**display openflow summary**
显示OpenFlow实例的辅助连接信息和收发的报文统计信息等	**display openflow instance** *instance-id* **auxiliary** [*controller-id* [**auxiliary** *auxiliary-id*]]
清除控制器发送和接收报文的统计计数	**reset openflow instance** { *instance-id* { **controller** [*controller-id*] | **listened** } } **statistics**

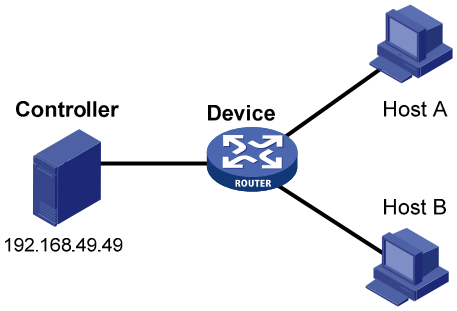
1.8 OpenFlow典型配置举例

1. 组网需求

- 创建全局实例类型的 OpenFlow 实例 1，并激活实例。
- 配置 OpenFlow 实例 1 连接的控制器，用来控制 Switch 上的流量转发。

2. 组网图

图1-7 OpenFlow 配置组网图



3. 配置步骤

(1) 创建实例 1 并配置其为全局实例

```
[Switch] openflow instance 1
[Switch-of-inst-1] classification global
```

(2) 配置控制器 1 的 IP 地址为 192.168.49.49，并激活实例

```
[Switch-of-inst-1] controller 1 address ip 192.168.49.49
[Switch-of-inst-1] active instance
```

4. 验证配置

显示实例详细信息。

```
[Switch-of-inst-1] display openflow instance 1
Instance 1 information:

Configuration information:
 Description   : --
 Active status : Active
 Inactive configuration:
  None
 Active configuration:
  Classification: Global(Standard)
  In-band management VLAN, total VLANs(0)
   Empty VLAN
  Connect mode: Multiple
  Mac-address learning: Enabled
  Flow table:
   Table ID(type): 0(Extensibility), count: 0
  Flow-entry max-limit: 65535
  Datapath ID: 0x0064001122000101
  Default table-miss: Drop
  Forbidden port: None
  Qinq Network: Disabled
Port information:
 GigabitEthernet2/1/3
 Route-Aggregation1
```

```
Active channel information:
 Controller 1 IP address: 192.168.49.49 port: 6633
```

1.9 附录 A 应用限制

1.9.1 Flow Entry的限制

1. 匹配的限制

(1) 协议报文的匹配

对于协议报文，一旦相关的协议配置使能后，协议报文不会进入 OpenFlow 转发处理，仍然由相关协议进行处理。

(2) MetaData 的匹配

MetaData 用于流表间的匹配信息传递，在非第一级流表支持下发 MetaData 的匹配，如果 Controller 在第一级流表下发了 MetaData 的匹配项，Switch 返回不支持。

2. Instruction的限制

(1) Clear actions 的限制

- 单级流表的情况下，支持 Clear actions。
- 多级流表的情况下，仅第一级流表支持 Clear actions 与其它 instruction 的动作配合，后续流表仅支持单独下发 Clear Actions。

(2) Apply actions 的限制

不支持Action List中包含多个Output的情况，仅支持一个Output时，请参见 1.9.2 。

(3) Write MetaData/MetaMask

在且仅在非最后一级流表的情况下，Switch 支持 Write MetaData/MetaMask 的操作，否则 Switch 返回不支持。

(4) Go To Table

在且仅在非最后一级流表的情况下，Switch 支持 Go To Table 的操作，否则 Switch 返回不支持。

1.9.2 Action List和Action Set整合的限制

OpenFlow Switch 设备整合 Action Set 和 Action List 为 Action Set，其整合原则如下。

1. 非Output Action

Action List 和 Action Set 中的 Action（除 Output 和 Group 外）如果不存在冲突，则全部保留为 Action Set；如果存在冲突，则以 Action Set 的动作替换 Action List 中的动作（其原因是 Action List 要执行在 Action Set 之前）。

2. Output Action

- 当 Action List 和 Action Set 中都存在一个 Output 的 Action 时，Action List 中的 Output 发送的报文不会对报文进行任何修改，其执行顺序最优，Action Set 中的 Ouput 会执行 Action List 和 Action Set 中的所有修改。
- 当 Action List 和 Action Set 中仅存在一个 Output 的 Action 时，该 Output 为报文出接口，执行顺序按照 Action Set 的顺序。

- 当 Action List 中存在一个 Output 的 Action，Action Set 中存在一个 Group 的 Action（Output 的 Action 存在与否都可以）时，Action List 中的 Output 发送的报文不会对报文进行任何修改，Group 在 Action Set 中。
- 其它情况，不支持。

1.9.3 Packet Out的处理限制

1. 入接口限制

在 Packet out 消息中 Output 为 Normal、Local、In port 或 To Controller 时，入接口只能是设备上的物理接口或者逻辑接口，不能是 OpenFlow 保留口。

2. Buffer ID和报文内容同时存在的处理

在 Packet Out 消息中如果同时存在 Buffer ID 和报文，OpenFlow Switch 只会获取 Buffer ID 对应的缓存报文进行处理，忽略消息中携带的报文。

1.9.4 Packet in的处理限制

报文缓存限制如下：

- 对于上送原因是 No Match 的报文支持缓存，缓存大小是 1K 个报文。
- 对于其它上送原因的报文不支持缓存，整个报文都会被上送，并且 Cookie 是全 F。

1.9.5 Flow Mod的限制

1. Table Miss表项的添加、修改和删除

- Switch 在激活后缺省会生成 Table Miss 表项，其动作是 Drop，此表项不能被 Controller 通过 Modify 的动作修改，不能被 Controller 通过 Mulipart 消息查询到，仅能由 Controller 通过 Add 进行添加 Table Miss 的动作进行修改。
- Table Miss 表项仅能通过严格匹配进行修改和删除，在非严格匹配的情况下，即使匹配项是通配也不能够操作 Table Miss 表项。
- Table Miss 表项被删除后，会生成缺省的 Table Miss 表项，其动作是 Drop。

2. 普通表项的添加、修改和删除

- 在非严格匹配的情况下，不支持通过 match 域为通配修改所有普通流表项。

1.10 附录 B MAC-IP流表

OpenFlow Switch 支持两种类型的 Flow Table，MAC-IP 类型和 Extensibility 类型。允许通过命令行指定 Table ID，Flow Table 会根据 Table ID 进行排序。Flow Table 需要重新激活后才能生效。

MAC-IP 流表是使用 MAC 地址表项和路由表项实现 Flow Table；Extensibility 表使用 TCAM 或者软件实现 Flow Table。

1.10.1 MAC-IP流表支持能力

必选的能力是 Controller 下发时必须携带的匹配或者动作项，可选能力是下发时可携带可不携带，如果不携带的话由 Switch 添加缺省的匹配或者动作项。

三层表项使用路由表实现，其支持能力如 表 1-25 所示。

表1-25 MAC-IP 流表三层表项支持能力

支持项	能力
必选匹配项	• VLAN • 单播目的 IP 地址 • 单播目的 MAC 地址（必须是匹配 VLAN 对应的 VLAN 接口的 MAC 地址）
可选匹配项	无
必选动作项	• 指定出接口 • 修改 VLAN • 修改目的 MAC 地址
可选动作项	• 修改源 MAC 地址(源 MAC 地址会修改为目的出接口所在 VLAN 对应的 VLAN 接口的 MAC 地址） • TTL 减 1 • Go to table(在多级流表存在的情况下，即使Controller不下发，Swtich上缺省下发该动作） • Write Meta(在多级流表存在的情况下，即使 Controller 不下发，Switch 上缺省下发目的 IP 地址匹配的 MetaData）

VXLAN三层表项使用路由表实现，其支持能力如 表 1-26 所示。

表1-26 MAC-IP 流表 VXLAN 三层表项支持能力

支持项	能力
必选匹配项	• 以太网类型 eth_type • 单播目的 IP 地址
可选匹配项	无
必选动作项	• 指定 TUNNEL 出接口 • 指定 TUNNEL ID • 修改目的 MAC 地址
可选动作项	• Go to table(在多级流表存在的情况下，即使Controller不下发，Swtich上缺省下发该动作） • Write Meta(在多级流表存在的情况下，即使 Controller 不下发，Switch 上缺省下发目的 IP 地址匹配的 MetaData）

1.10.2 MAC-IP流表的限制

MAC-IP 流表的 Flow Entry 有一定的限制，Controller 需要遵循这些限制下发表项，否则可能会造成转发错误。

三层表项的限制如 表 1-27 所示。

表1-27 MAC-IP 流表三层表项限制

表项类型	限制
匹配项限制	• 匹配的 VLAN 所对应的 VLAN 接口 UP • 目的 MAC 地址是匹配 VLAN 对应的 VLAN 接口的 MAC 地址 • 目的 IP 地址不是本机 IP 地址
动作项限制	• 指定出接口属于目的 VLAN • 目的 MAC 地址不是本机 MAC 地址 • 如果修改源 MAC 地址，源 MAC 地址必须是目的出接口所在 VLAN 对应的 VLAN 接口的 MAC 地址

说明

三层表项能够下发的前提是匹配 VLAN 所对应的 VLAN 接口存在并且处于 UP 状态，且 VLAN 接口会作为 OpenFlow 接口上报（包括了 VLAN 接口的链路状态和 MAC 地址），在 VLAN 接口删除时同时也会上报给 Controller，需要由 Controller 删除对应的三层流表项，因此需要 Controller 保证三层表项的正确性，Switch 端不对三层表项的匹配项限制进行检查。

VXLAN三层表项的限制如 表 1-28 所示。

表1-28 MAC-IP 流表 VXLAN 三层表项限制

表项类型	限制
匹配项限制	• 目的 IP 地址不是本机 IP 地址
动作项限制	• 出接口必须为存在的 TUNNEL 接口 • 指定的 TUNNEL ID（VNI）对应的 VXLAN 和 VSI 必须存在 • 目的 MAC 不是本机的 MAC 地址

1.10.3 MAC-IP流表的Table Miss

MAC-IP 流表的 Table Miss 支持下列 Output Action：

- Go To Table：进入下一级流表；
- Drop：丢弃报文；
- Controller：报文上送 Controller；
- Normal：报文正常转发。

1.10.4 MAC-IP Table与Extensibility Table的配合

1. MetaData/Mask

MAC-IP Table 和 Extensibility Table 通过 MetaData/Mask 可以实现多级流表。

MAC-IP Table 支持 Write MetaData/Mask，Extensibility 支持 Match MetaData/Mask。

MetaData Mask每个Bit表示不同的含义，MetaData中对应的Bit位置位表示匹配，未置位表示通配，具体参见 表 1-29。

表1-29 MetaData Mask 含义

MetaData Mask Bit	含义	MetaData
Bit 0	目的MAC	1，置位，表示匹配到目的MAC
		0，未置位，表示未匹配到目的MAC
Bit 1	源MAC	1，置位，表示匹配到源MAC
		0，未置位，表示未匹配到源MAC
Bit 2	目的IP	1，置位，表示匹配到目的IP
		0，未置位，表示未匹配到目的IP
其他	保留	保留