



# 数据中心 SDN 控制器

## 用户手册

**文档编号：** DCFabric- Controller-UserGuide-V1.3

**文档名称：** 数据中心 SDN 控制器用户手册

**版本信息：** V1.1

**建立日期：** 2014 年 12 月 31 日

**最后日期：** 2015 年 05 月 26 日

## 文档修订记录

版本编号 或者更改 记录编号	修订 原因	简要说明（变更 内容和变更范 围）	日期	变更人	批准日期	批准人
V0.1	创建		2014-11-18	邓超		
V1.0	补充		2014-11-23	邓超		
V1.1	修改		2014-12-31	邓超		
V1.2	修改		2015-05-24	赵良智		
V1.3	修改		2015-05-25	杨磊		

修订原因：创建；增加；修改；删除

## 文档审批信息

序号	审批人	角色	审批日期	签字	备注

# 目 录

<b>1 引言</b>	<b>5</b>
1.1 编写目的	5
1.2 读者对象	5
1.3 文档约定	5
1.4 术语定义	5
<b>2 WEB APP 使用</b>	<b>6</b>
2.1 控制器集群管理	6
2.1.1 添加控制器	6
2.1.2 删除控制器	7
2.2 交换机信息	7
2.2.1 交换机信息查询	7
2.2.2 交换机网口信息查询	8
2.3 网络拓扑	10
2.3.1 网络拓扑查询	10
2.3.2 sFabric 初始化	11
2.4 流表管理	12
2.4.1 动态流表查询	12
2.4.2 静态流表查询	14
2.4.3 静态流表下发	16
2.5 三层网关管理	25
2.5.1 查询三层网关	25
2.5.2 添加三层网关	26
2.6 METER 表配置	31
2.6.1 添加 meter 表	31
2.7 GROUP 表配置	32
<b>3 SFABRIC 功能</b>	<b>36</b>
3.1 sFABRIC 功能介绍	36
3.2 FABRIC 初始化	36
3.3 FABRIC 路径更新	37
3.3.1 添加交换机	37
3.3.2 删除交换机	37
3.3.3 连接数据线	37
3.3.4 移除数据线	37
3.4 FABRIC 路径查询	38
3.5 组 FABRIC 设置	40
<b>4 附录</b>	<b>41</b>
4.1 流表匹配项前置约束	41



# 1 引言

## 1.1 编写目的

本文档用于指导用户更加正确、高效的使用数据中心 SDN 控制器。文档是基于 WEB APP 层面的用户使用控制器指导。

## 1.2 读者对象

本文档预期读者包括：

- 数据中心 SDN 控制器开发项目管理和开发相关人员
- 系统需求组、系统架构组、系统设计组
- 业务人员、测试人员、用户

## 1.3 文档约定

本文档所有描述内容仅代表本数据中心 SDN 控制器支持的功能，并不表示其所管理的交换机也能够支持。用户需要结合交换机的功能、特性清单使用。

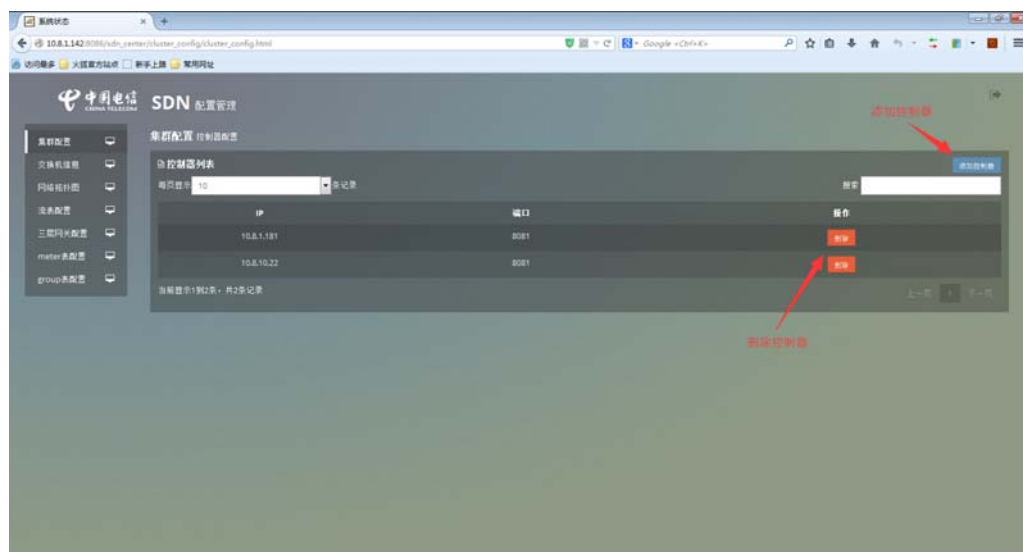
## 1.4 术语定义

术语和缩写	全拼	解释
静态流表	Static Flow Entry	用户手动制定下发的流表项。
动态流表	Dynamic Flow Entry	运行过程中控制器根据自身处理逻辑，自动产生并下发的流表项。

## 2 WEB APP 使用

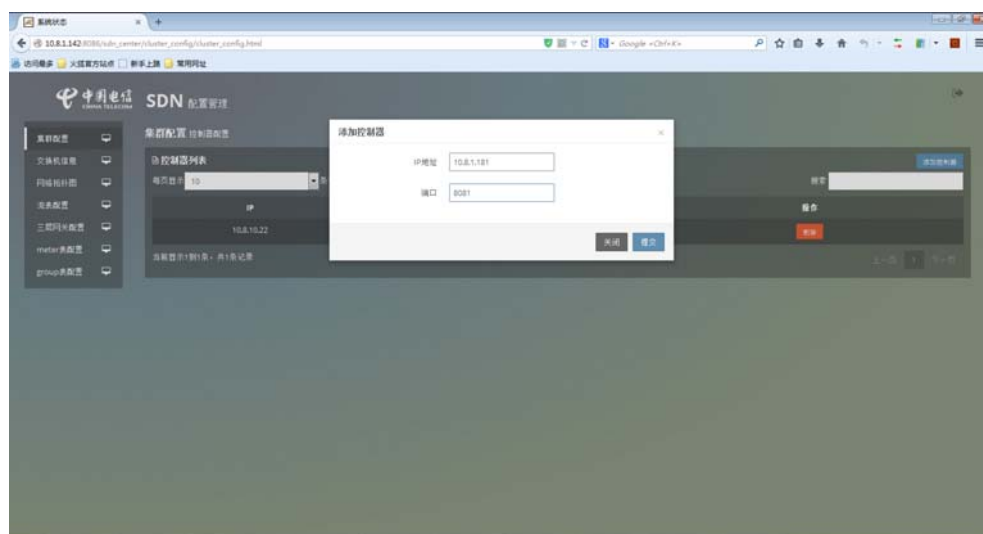
### 2.1 控制器集群管理

控制器管理包括控制器集群列表查询、添加和删除。其中列表查询动作在页面加载时自动完成。见下图：

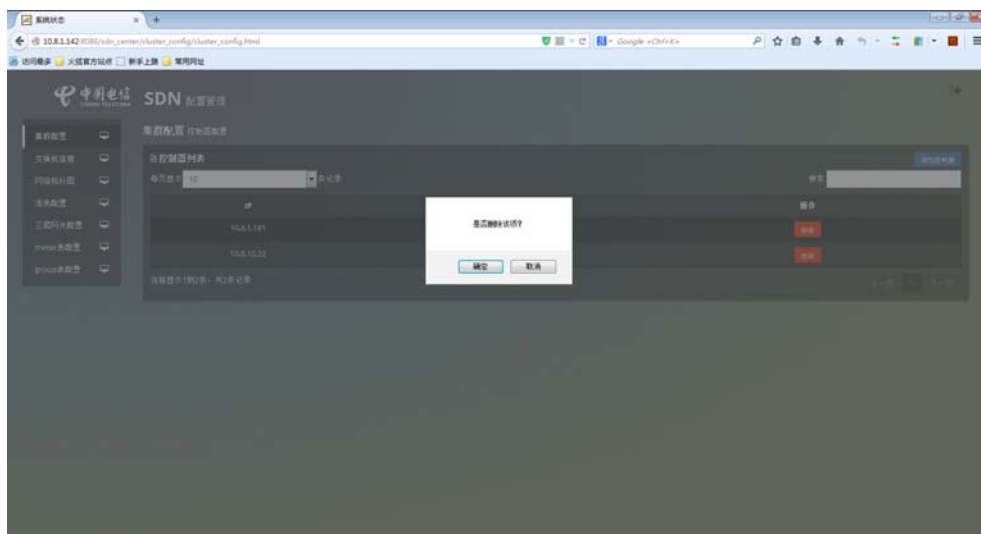


注：控制器管理中的“添加控制器”和“删除控制器”并不会启动或停止控制器服务。控制器服务的启停需要管理员后台完成。

#### 2.1.1 添加控制器



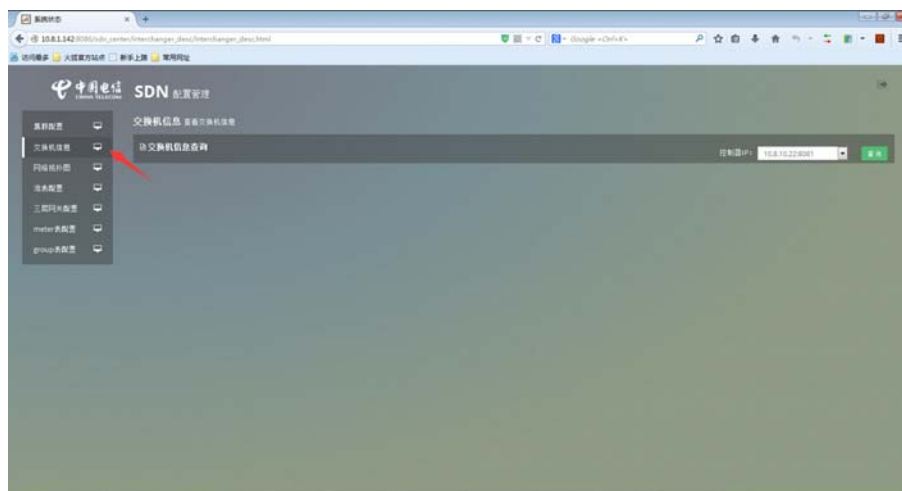
## 2.1.2 删除控制器



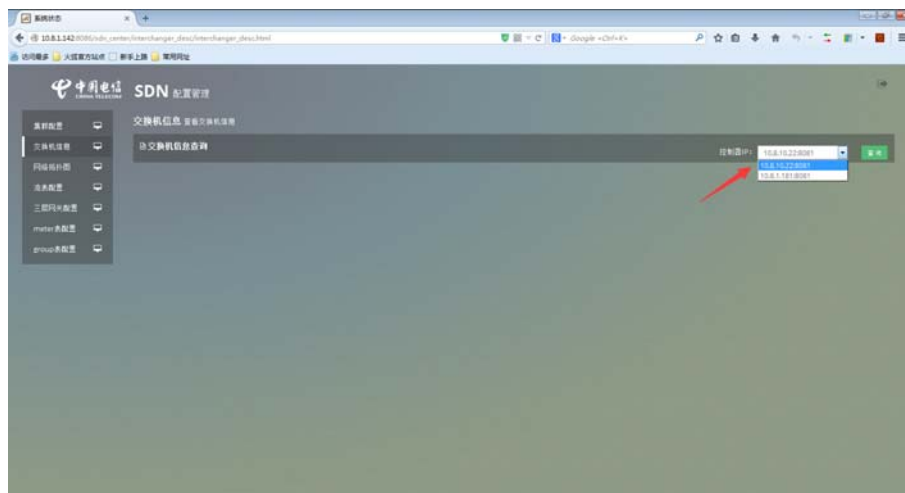
## 2.2 交换机信息

### 2.2.1 交换机信息查询

步骤一： 选择“交换机信息”菜单

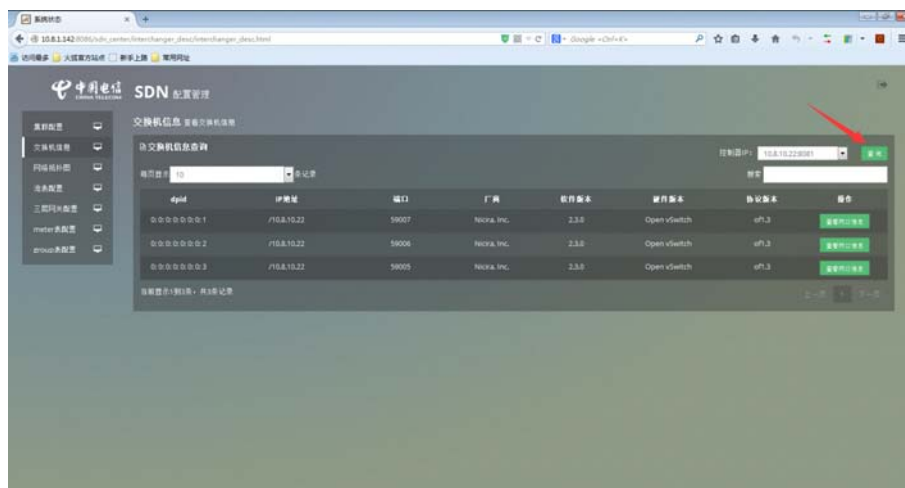


## 步骤二： 选择控制器



## 步骤三： 查询交换机信息

点击“查询按钮”：

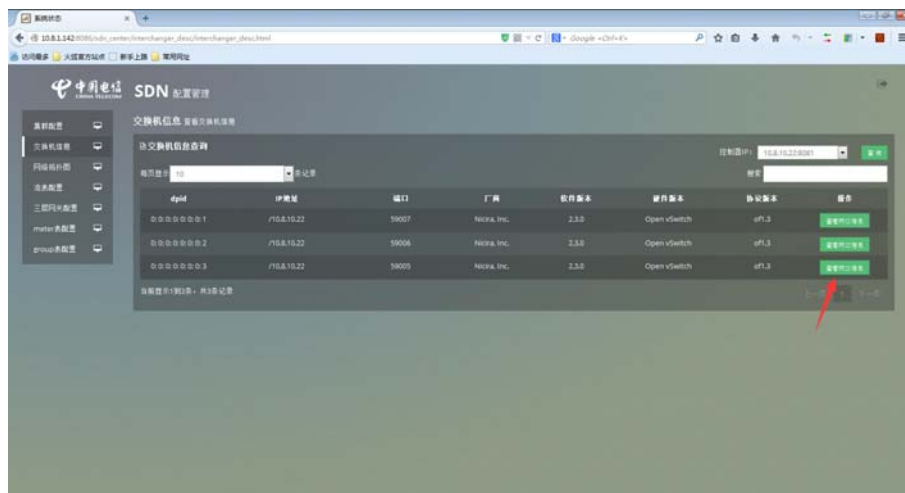


### 2.2.2 交换机网口信息查询

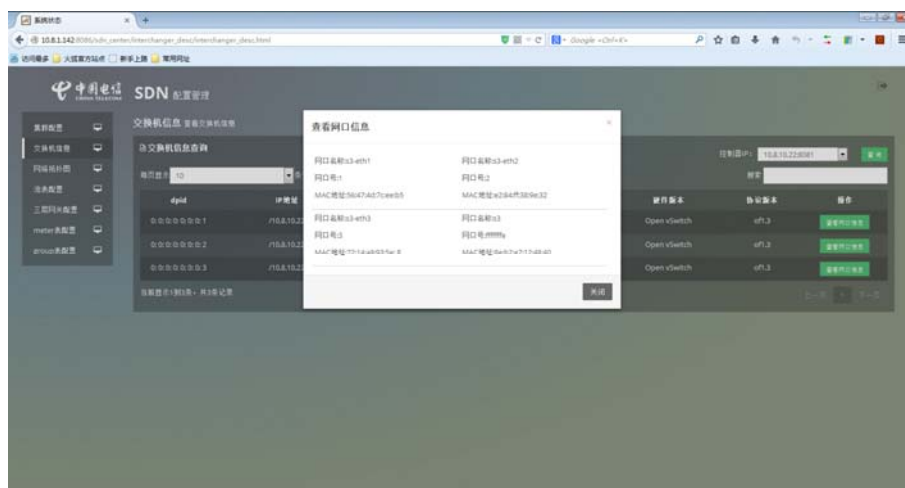
在“2.2.1 交换机信息查询”基础上，选择交换机查询其网口信息。



## 步骤一： 选择交换机



## 步骤二： 网口信息



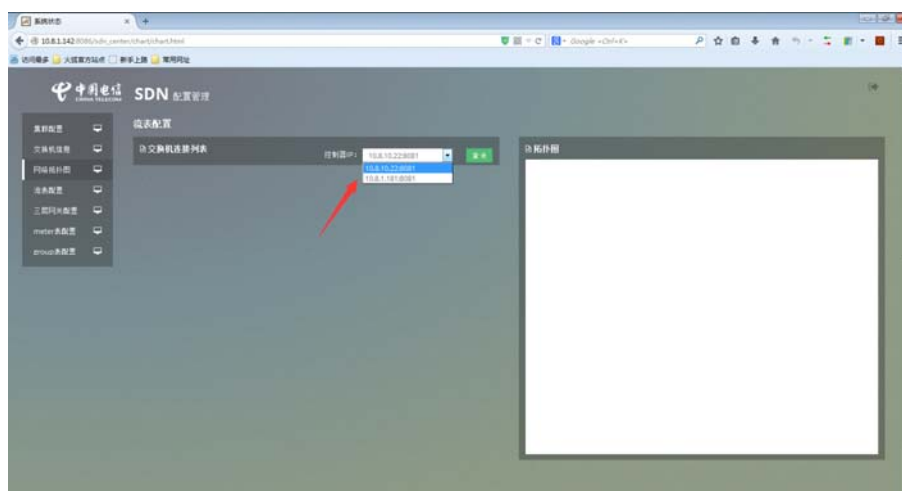
## 2.3 网络拓扑

### 2.3.1 网络拓扑查询

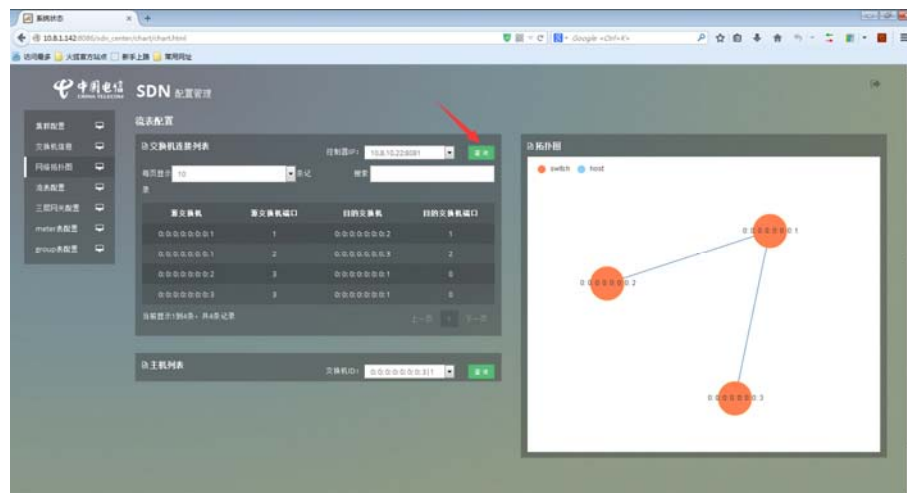
步骤一： 选择菜单栏“网络拓扑图”



步骤二： 选择控制器

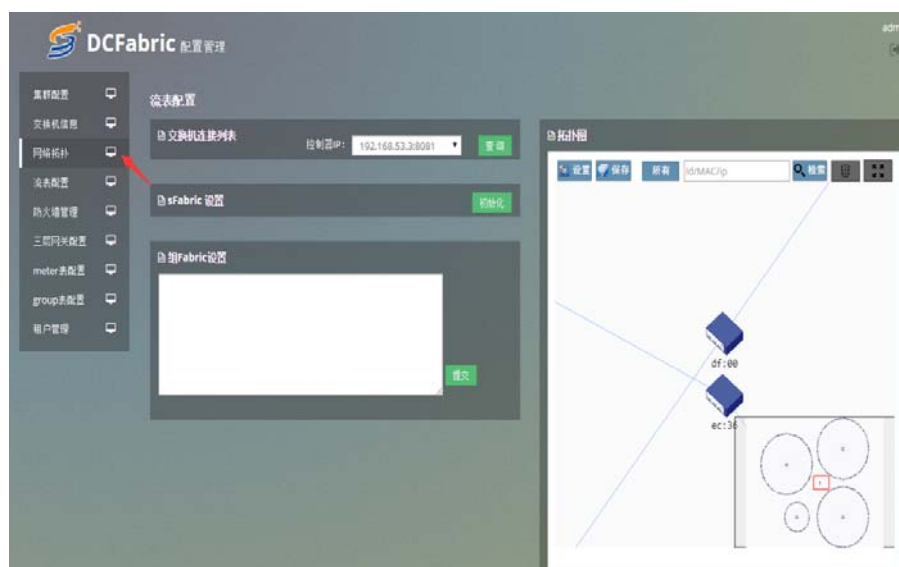


### 步骤三： 查询网络拓扑图

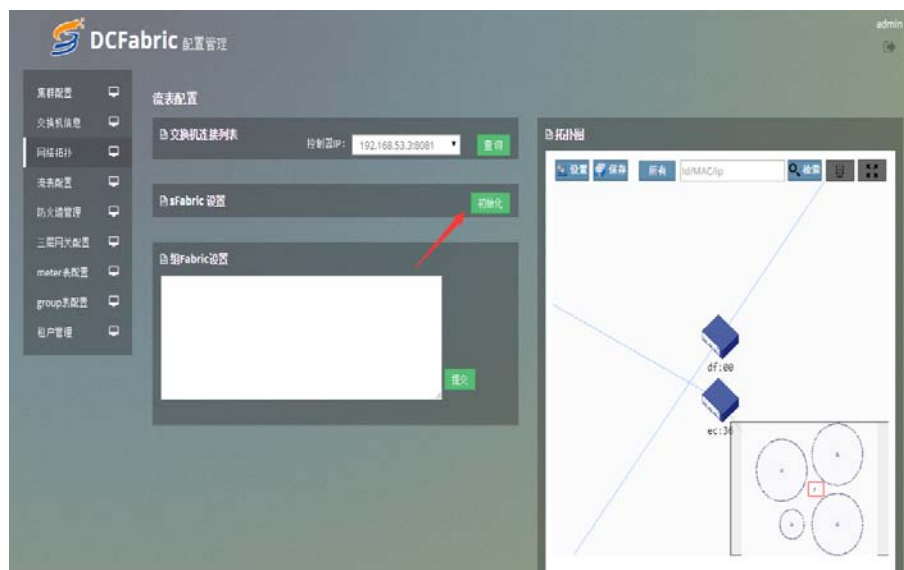


## 2.3.2 sFabric 初始化

### 步骤一： 选择菜单栏“网络拓扑图”



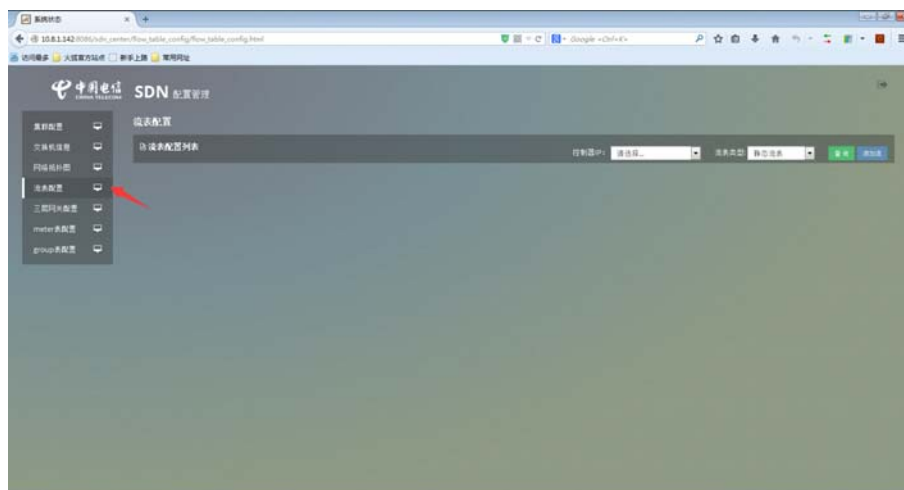
## 步骤二： 点击初始化



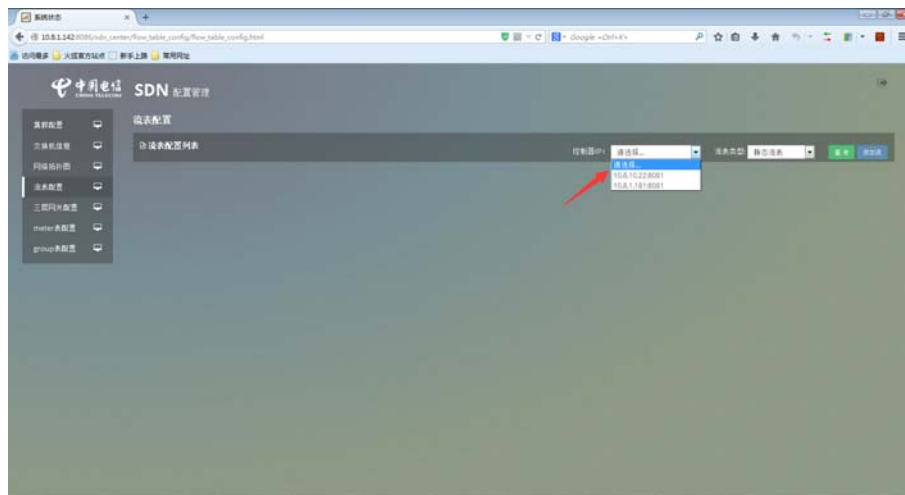
## 2.4 流表管理

### 2.4.1 动态流表查询

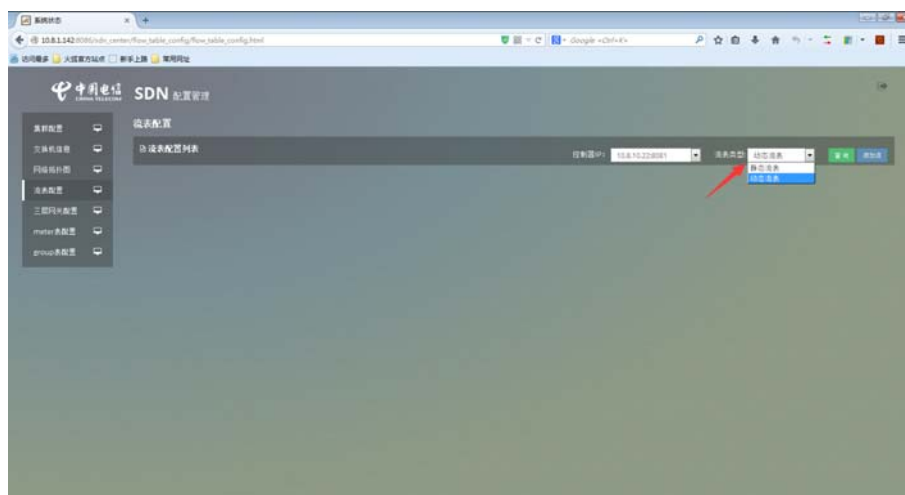
#### 步骤一： 选择菜单栏“流表配置”



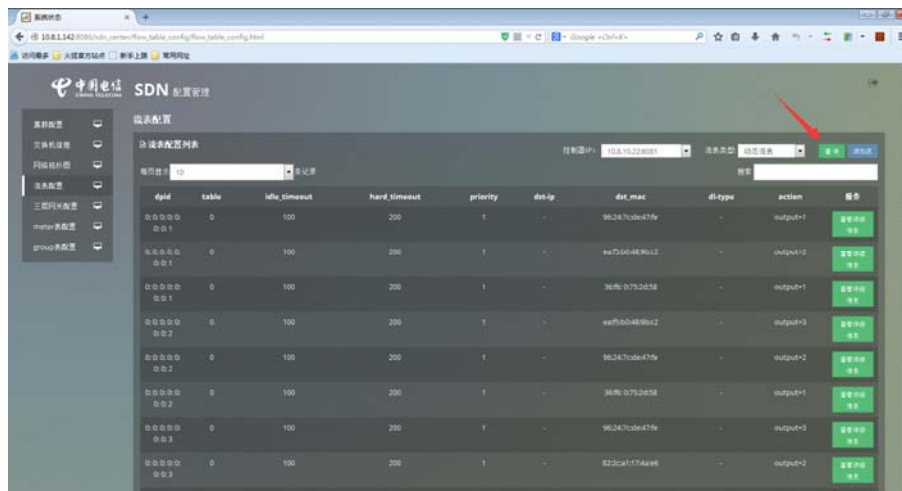
## 步骤二： 选择交换机



## 步骤三： 选择流表类型为“动态流表”

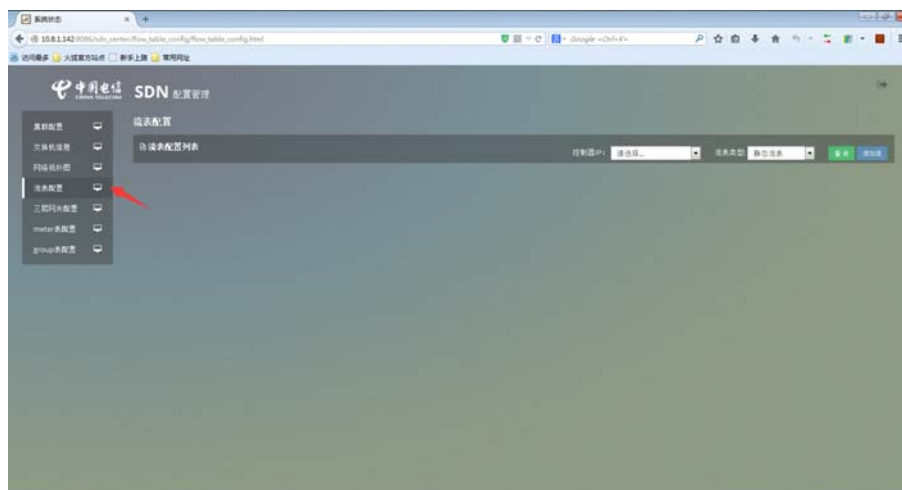


步骤四： 点击“查询”按钮

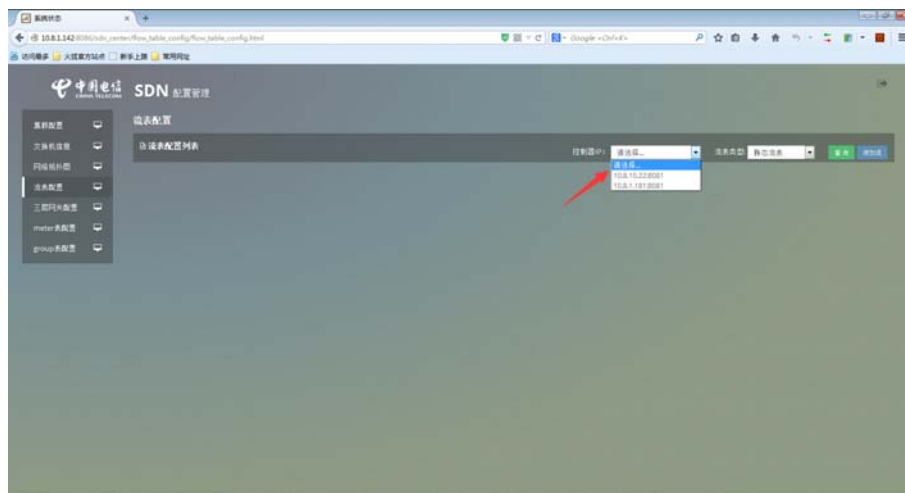


## 2.4.2 静态流表查询

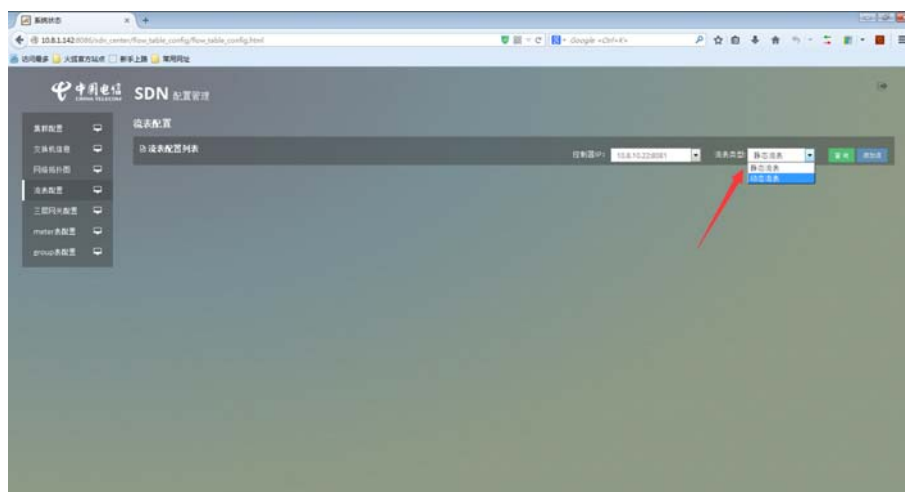
步骤一： 选择菜单栏“流表配置”



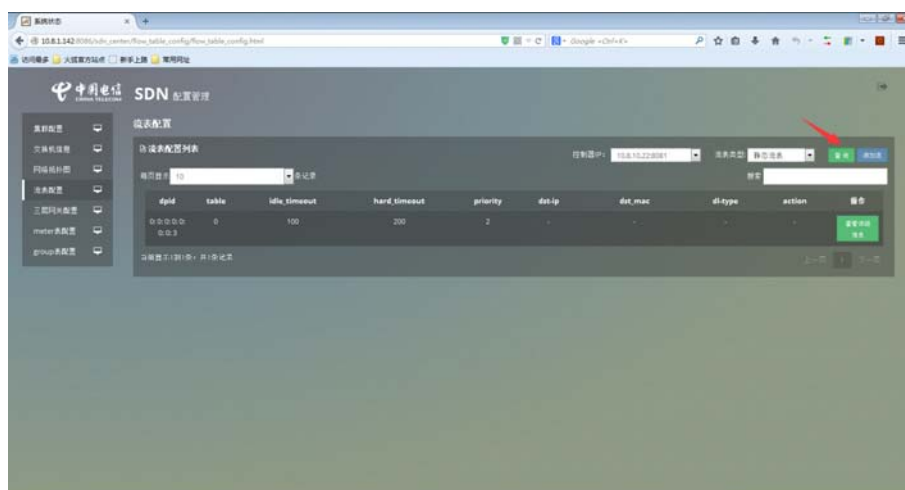
## 步骤二： 选择交换机



## 步骤三： 选择流表类型为“静态流表”

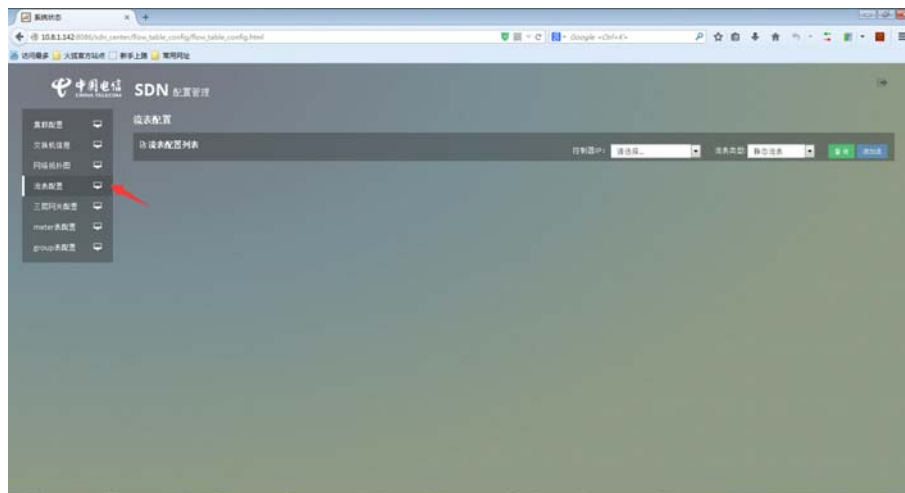


## 步骤四： 点击“查询”按钮

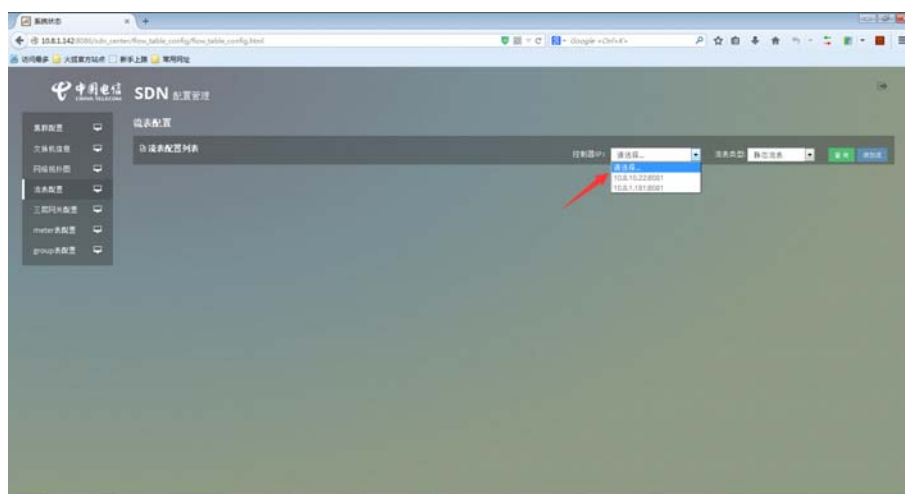


## 2.4.3 静态流表下发

步骤一： 选择菜单栏“流表配置”

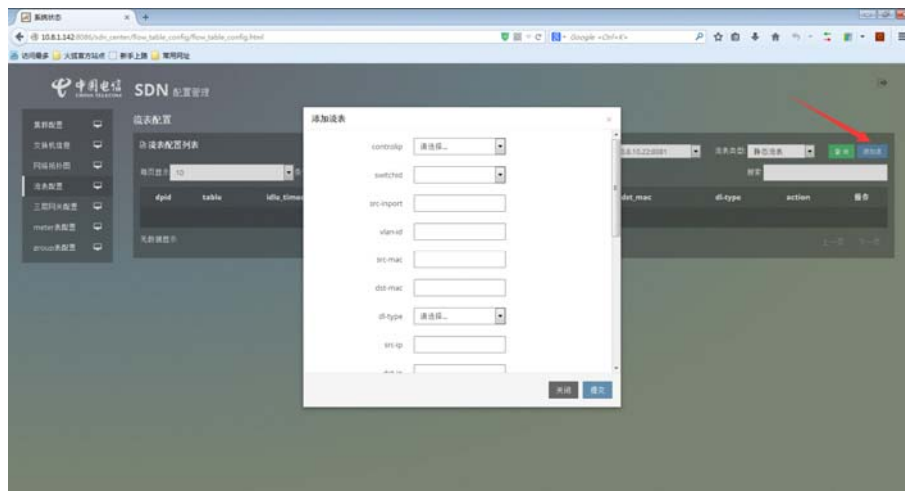


步骤二： 选择交换机





### 步骤三： 点击“添加流”按钮



### 步骤四： 填写流表内容并提交

本处仅描述流表下发操作步骤和各个填充项的含义，更多流表下发约束限制请参考“3sFabric 功能”

## 2.5 sFabric 功能介绍

## 2.6 Fabric 初始化

### 步骤一： 选择菜单栏“网络拓扑图”



## 步骤二： 点击初始化



## 2.7 Fabric 路径更新

### 2.7.1 添加交换机

观察控制器的日志界面，当出现下图的打印信息时，表示该交换机已经增加到拓扑中，并且 fabric 路径已经自动更新，该变化是即时发生的，响应时间<4 秒。

```
[INFO] New OpenFlow13 switch [00:00:00:00:00:01:95] connected: ip[192.168.52.134:53307]
[INFO] call_delete_switch_fun, sw num:1
[INFO] call_add_switch_fun, sw num:1
[INFO] call_add_switch_port_fun
[INFO] update_fabric_swap_flows begin!
[INFO] topo_change_event_thread start
[INFO] topo_change_event_thread start
[INFO] topo_change_event_thread start
[INFO] topo_change_event_thread start
[INFO] topo_change_event_thread start
```

### 2.7.2 删除交换机

观察控制器的日志界面，当出现下图的打印信息时，表示该交换机已经从拓扑中删除，并且 fabric 路径已经自动更新，该变化是即时发生的，响应时间<0.5 秒。

```
[INFO] topo_change_event_thread start
[WARNING] switch [00:00:00:00:00:01:95] disconnected
[INFO] topo_change_event_thread start
[INFO] call_delete_switch_port_fun
[INFO] update_fabric_swap_flows begin!
```

### 2.7.3 连接数据线

观察控制器的日志界面，当出现下图的打印信息时，表示一个新的边已经增加到拓扑中，并且 fabric 路径已经自动更新，该变化是即时发生的，响应时间<4 秒。

```
[INFO] topo_change_event_thread start  
[INFO] Port state change: eth1[new state: 0]  
[INFO] Port state change: eth-0-9[new state: 0]  
[INFO] topo_change_event_thread start  
[INFO] call_add_switch_port_fun  
[INFO] update_fabric_swap_flows begin!  
[INFO] topo_change_event_thread start
```

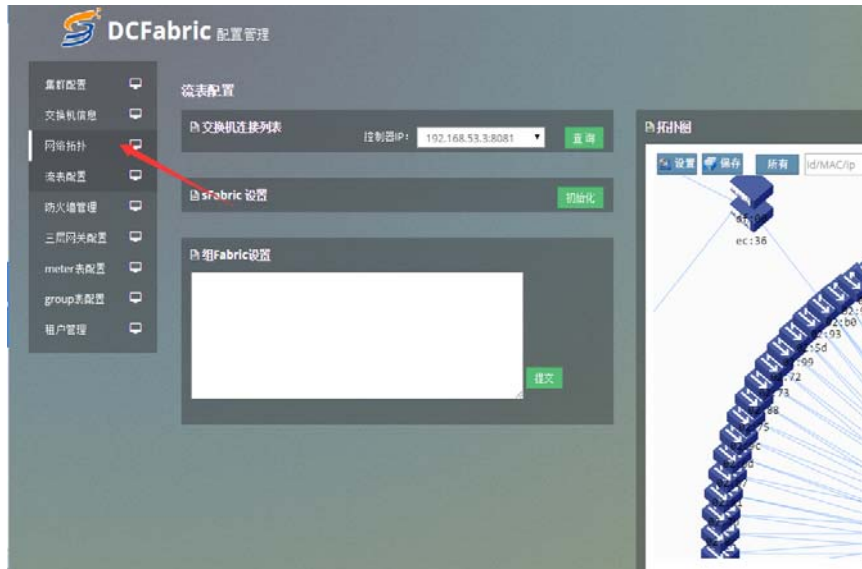
### 2.7.4 移除数据线

观察控制器的日志界面，当出现下图的打印信息时，表示一个边已经从拓扑中删除，并且 fabric 路径已经自动更新，该变化是即时发生的，响应时间<0.5 秒。

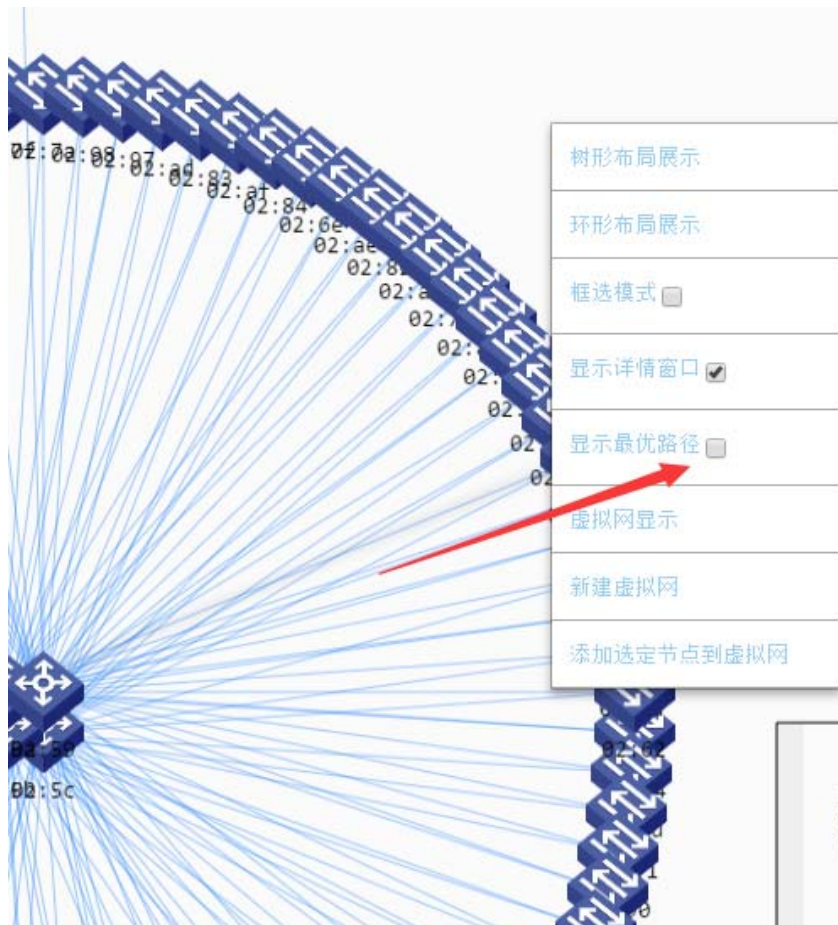
```
[INFO] topo_change_event_thread start  
[INFO] Port state change: eth1[new state: 1]  
[INFO] topo_change_event_thread start  
[INFO] Port state change: eth-0-9[new state: 1]  
[INFO] call_delete_switch_port_fun  
[INFO] update_fabric_swap_flows begin!  
[INFO] topo_change_event_thread start  
[INFO] Port state change: eth1[new state: 0]
```

## 2.8 Fabric 路径查询

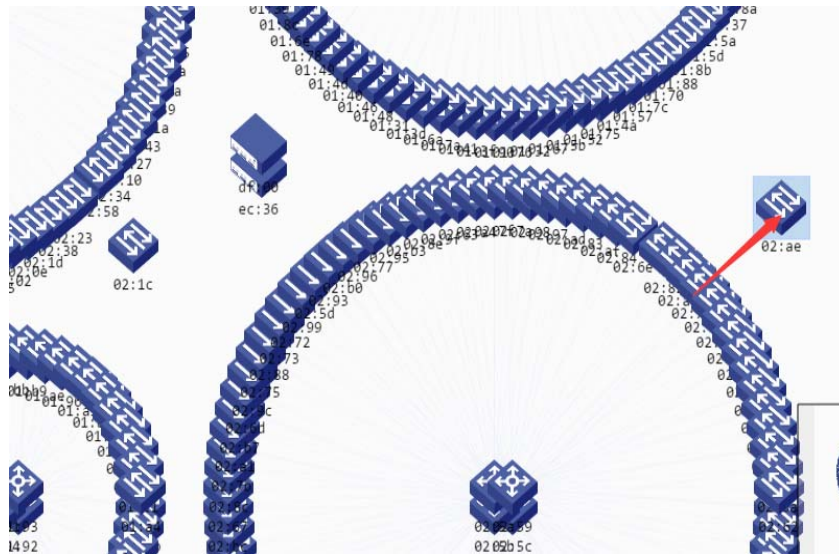
步骤一： 在菜单中选择“网络拓扑”，查看拓扑图



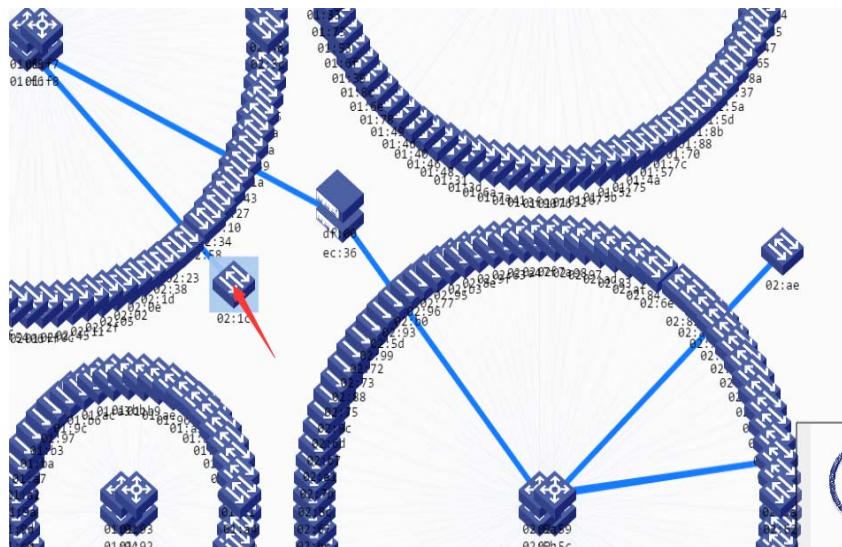
步骤二： 右击鼠标勾选上显示最优路径



### 步骤三： 点选起始位置交换机



### 步骤四： 点击终点位置交换机，即能显示该 2 点之间的 fabric 路径



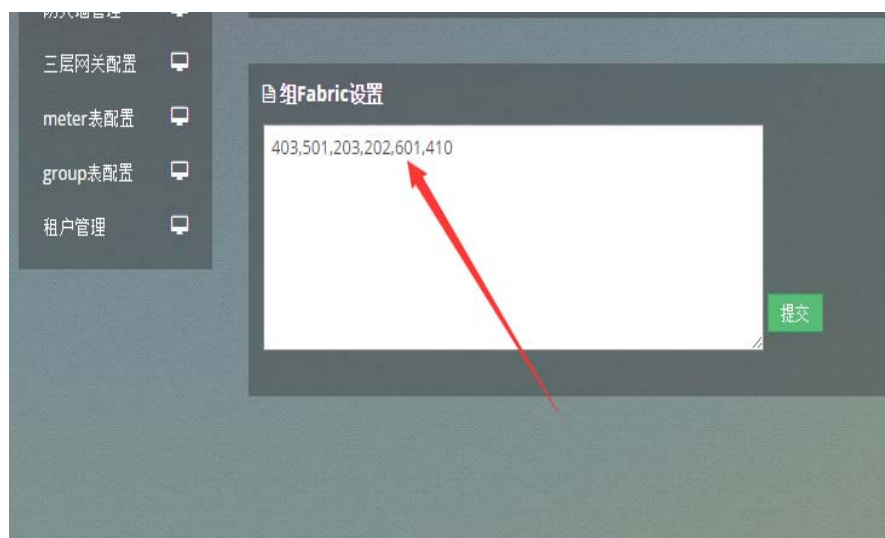


## 2.9 组 fabric 设置

步骤一： 在菜单中选择“网络拓扑”



步骤二： 在组 fabric 设置的多行文本框内输入想要设置的部分交换机的 dpid，以逗号隔开；



步骤三： 点击提交，该组交换机即可创建 fabric 路径。



附录 - 4.1 流表匹配项前置约束”中的约束限说明。

1) controllerip

从下拉表中选择控制器。

2) switchid

从下拉表中选择交换机。

3) src-input

用于匹配输入报文的交换机入口。网口号无限制，用户可以配置当前不存在的网口，当该网口出现时匹配规则便会生效。

4) vlan-id

用于匹配输入报文的 VLAN 标签。取值范围[0, 4095]。

5) src-mac

用于匹配输入报文的源 MAC。

6) dst-mac

用于匹配输入报文的目的 MAC。

7) dl-type

用于匹配输入报文的以太网类型：

- ARP
- IPv4

- IPv6

- MPLS

8) nw-proto

用于匹配输入报文的传输协议：

- TCP

- UDP

- ICMP

9) tp-src

用于匹配输入报文的 TCP/UDP 源端口。

10) tp-dst

用于匹配输入报文的 TCP/UDP 目的端口。

11) vlan-priority

匹配输入报文的 vlan 优先级。

12) nw-tos

用于匹配输入报文的 IP 包的 TOS 自段。

13) priority

流表优先级。

14) action

指定 action 类型为 output。

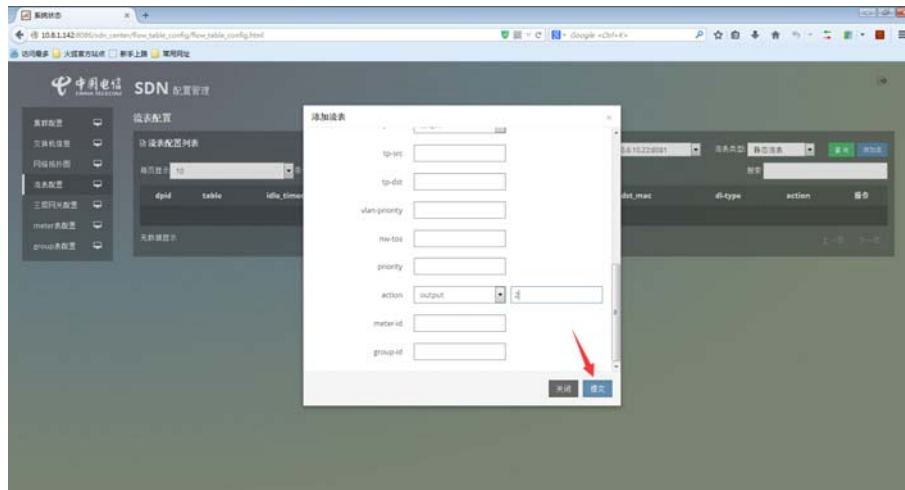
15) meter-id

匹配报文应用 meter 表。

16) group-id

匹配报文应用 group 表。



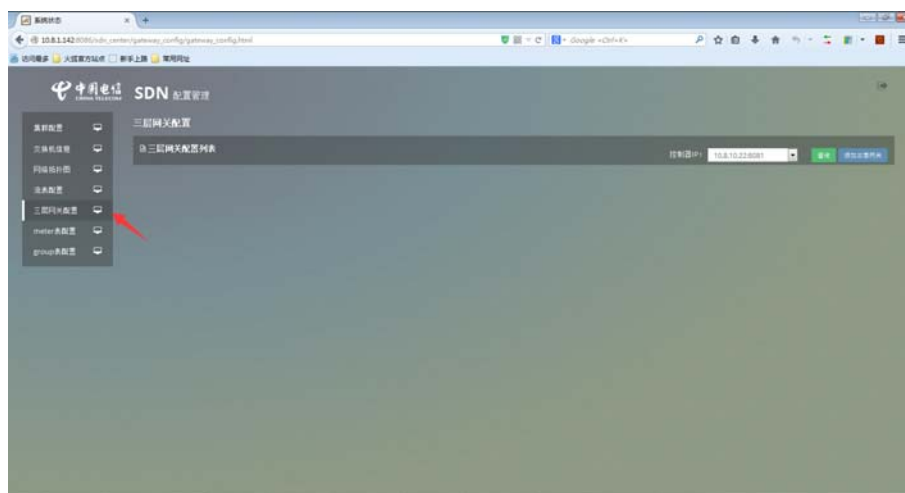


## 2.10 三层网关管理

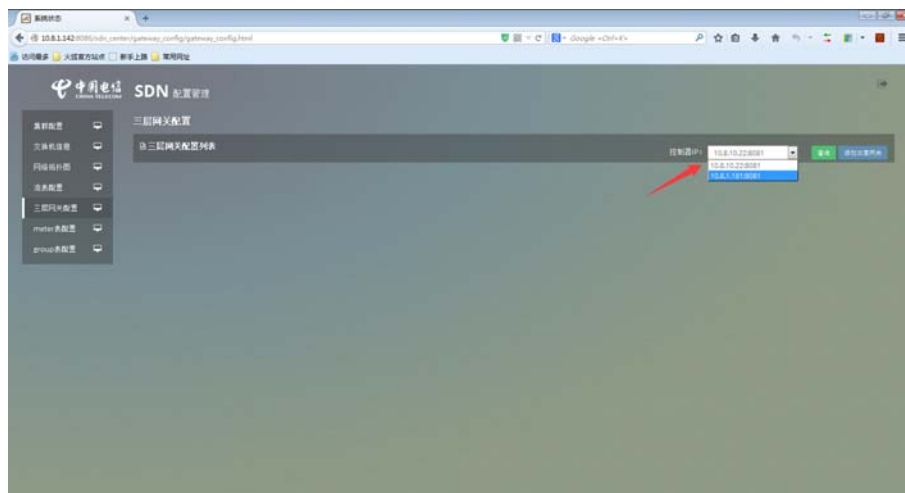
三层网关用于 L3 通信。当存在跨网段通信时，可以通过在控制器添加虚拟网关的方式达到不同网段互联互通的效果。

### 2.10.1 查询三层网关

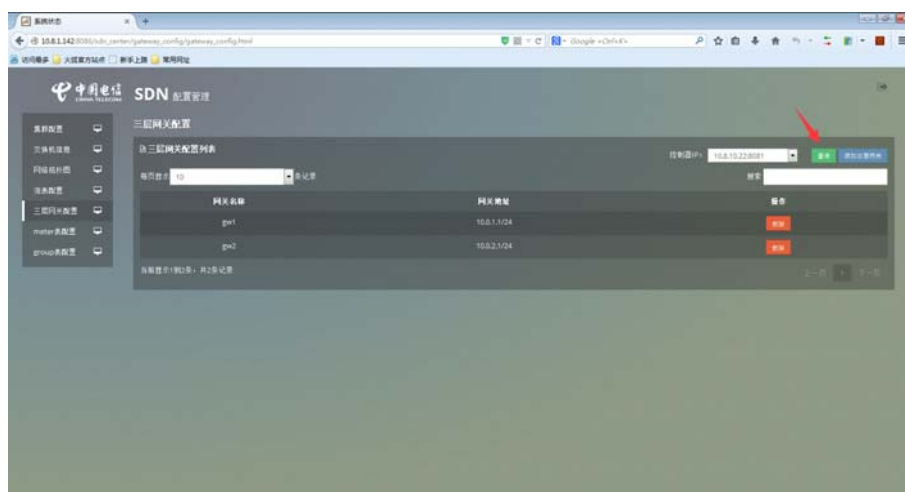
步骤一： 选择“菜单”中的“三层网关配置”



## 步骤二： 选择控制器

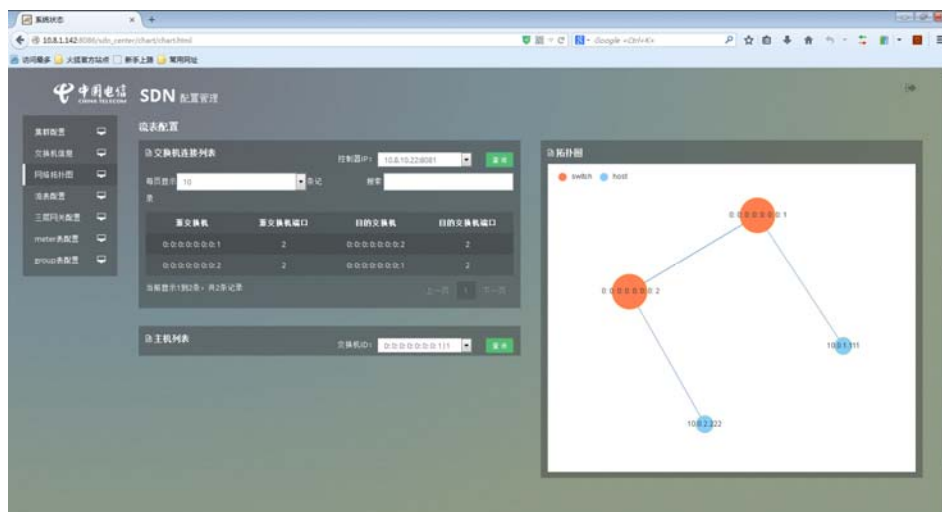


## 步骤三： 点击“查询”按钮



## 2.10.2 添加三层网关

例如根据“[2.3.1 网络拓扑查询](#)”查询到我当前的网络拓扑如下图：



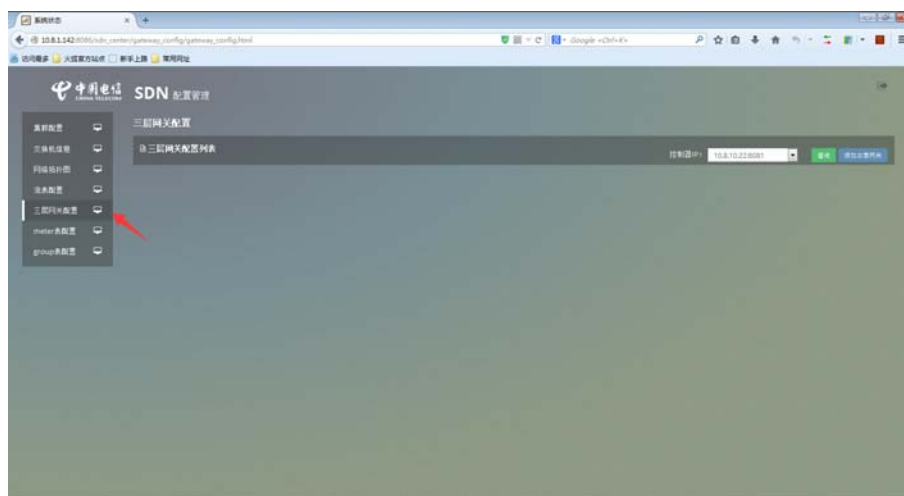
mininet 拓扑构造脚本:

```
#--l3.py--
from mininet.topo import Topo
class MyTopo( Topo ):
    "Simple topology example."
    def __init__( self ):
        "Create custom topo."
        # Initialize topology
        Topo.__init__( self )
        # Add hosts and switches
        H1 = self.addHost( 'h1' )
        H2 = self.addHost( 'h2' )
        S1 = self.addSwitch( 's1' )
        S2 = self.addSwitch( 's2' )
        # Add links
        self.addLink( H1, S1 )
        self.addLink( H2, S2 )
        self.addLink( S1, S2 )

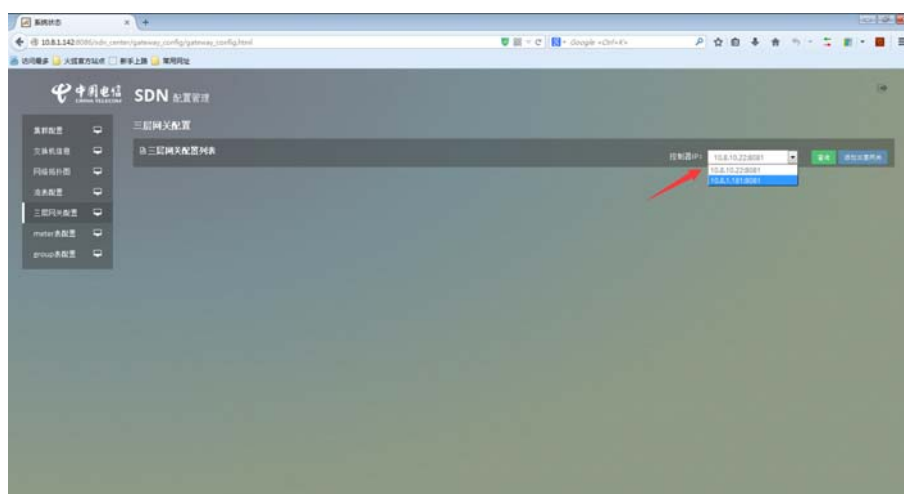
topos = { 'mytopo': ( lambda: MyTopo() ) }
```

由于交换机 s1 (0: 0: 0: 0: 0: 0: 0: 1) 上的 host1 (10.0.1.111) 和交换机 s2 (0: 0: 0: 0: 0: 0: 0: 2) 上的 host2 (10.0.2.222) 不在同一个网段, 无法通过二层完成通信。我们以此为例来添加虚拟三层通信网关。

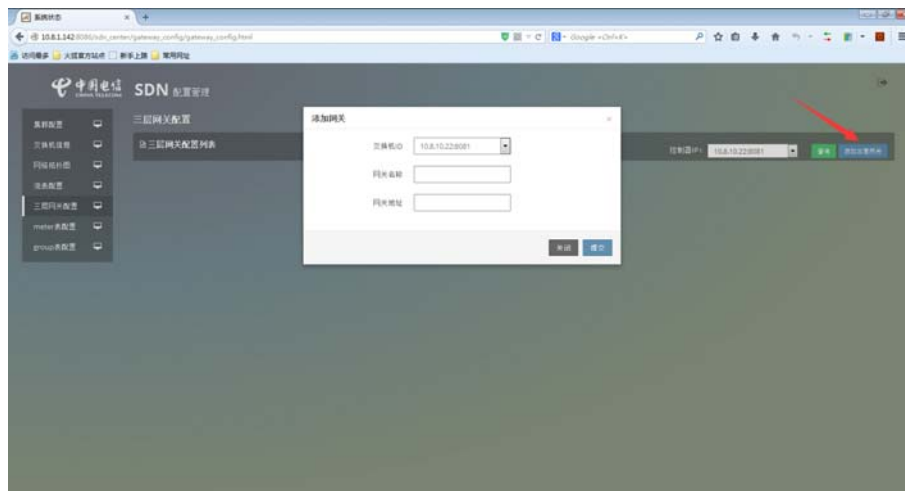
## 步骤一： 选择“菜单”中的“三层网关配置”



## 步骤二： 选择控制器



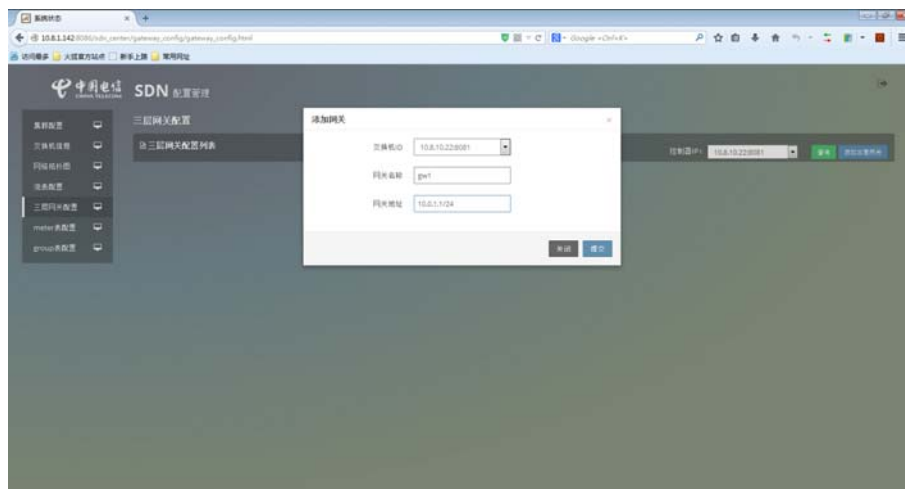
### 步骤三： 点击“添加三层网关”按钮

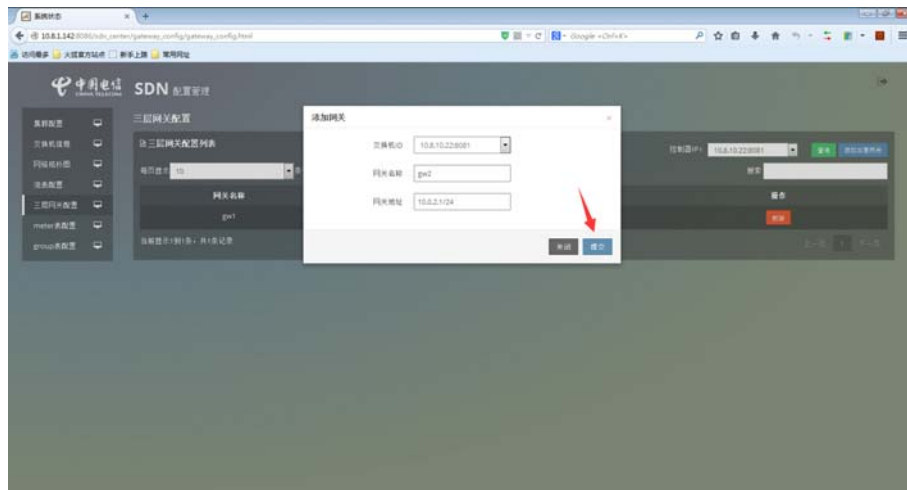


### 步骤四： 输入三层网关名称和网关地址后提交

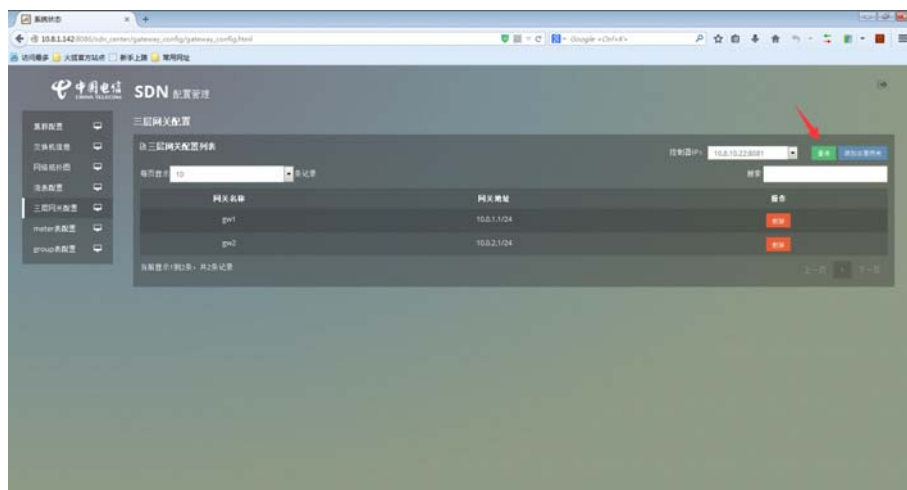
由于上面我们情况是两个不同网段通信，所以需要添加网关网段的虚拟网关。

网关地址一定要加上掩码，否则将导致出错。





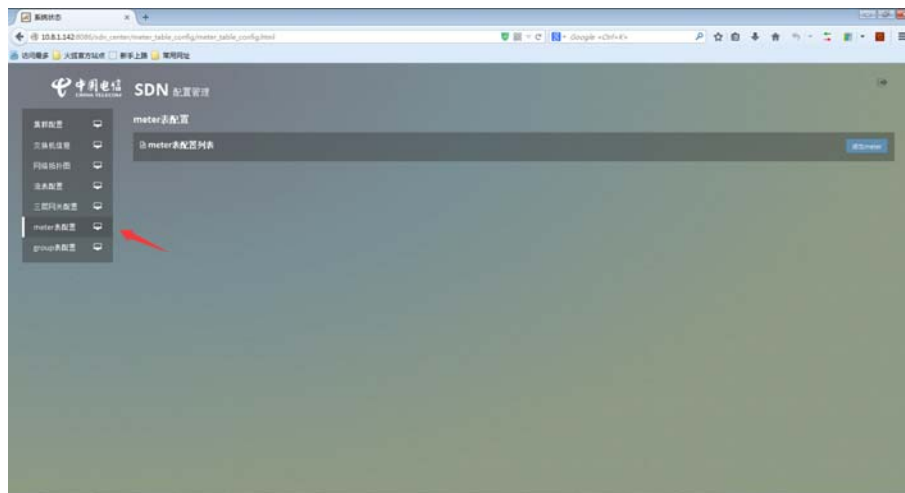
### 步骤五： 添加完成



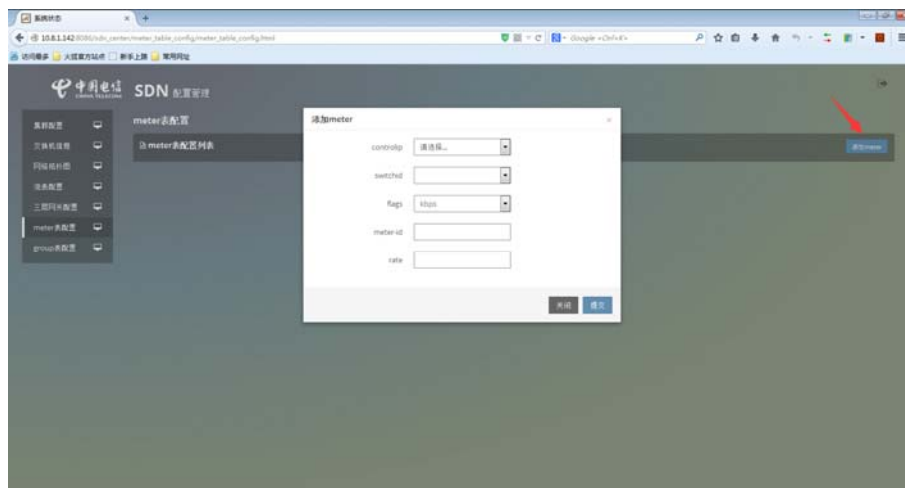
## 2.11 meter 表配置

### 2.11.1 添加 meter 表

步骤一： 在菜单中选择“meter 表配置”



步骤二： 点击“添加 meter”按钮



步骤三： 填充并提交 meter 表内容

- 1) 选择指定控制器
- 2) 选择控制器下的交换机
- 3) 选择限速方式

- kbps：限流量限速；

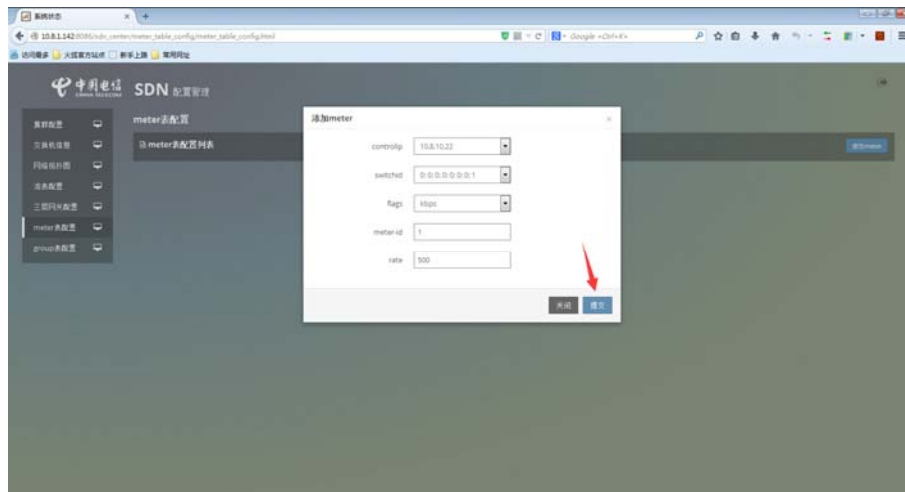
- pktps: 还是收发包数限速;

4) 输入 meter 表 id

唯一, 章节“[2.4.3 静态流表下发](#)”中流表下发时 meter-id 填充该值。

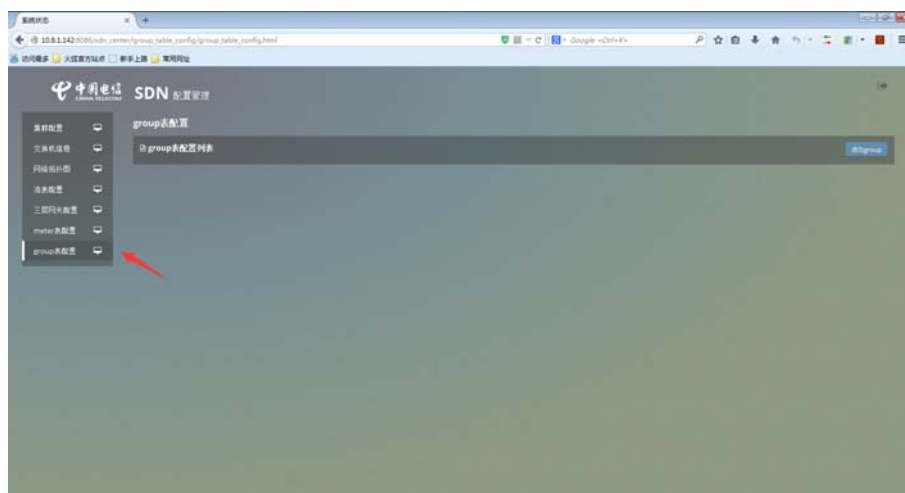
5) 输入速率值

6) 提交



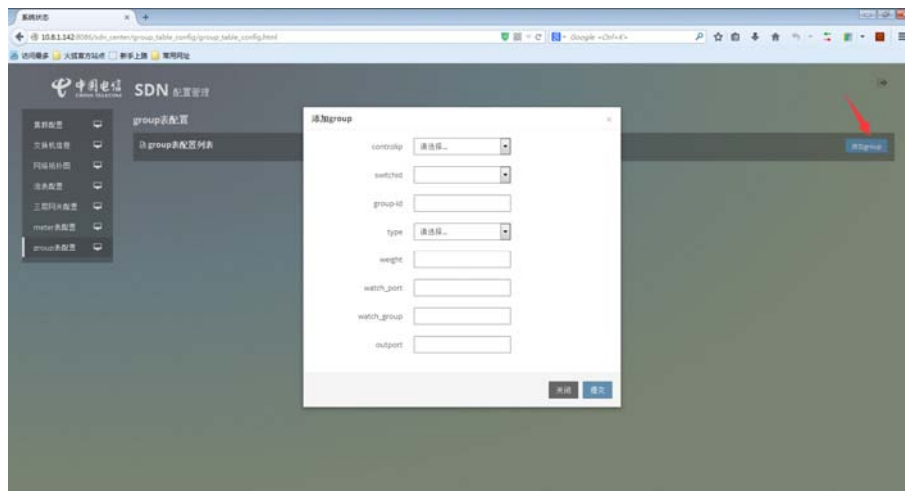
## 2.12 group 表配置

步骤一： 在菜单中选择“group 表配置”





## 步骤二： 点击“添加 group”按钮



## 步骤三： 填充并提交 group 表内容

- 1) 选择指定控制器
- 2) 选择控制器下的交换机
- 3) 输入 group 表 id

唯一，章节“[2.4.3 静态流表下发](#)”中流表下发时 group-id 填充该值。

- 4) 选择 group 表类型

某个交换机只支持那些标记为“Required”的组类型，至于“Optional”的组类型可以根据需要可选。

- Required: all: 执行组中的所有存储段 (Buckets)。这个组用于多播或广播的转发。数据包为每个存储段都有效地复制一份，然后被每个存储段处理。如果某个存储段中明确地指导数据包发往入端口，那么这个复制的包被丢弃。如果控制器希望数据包转发到入端口，那么这个组需要包含一个额外的存储段，这个存储段含有到 OFPP\_IN\_PORT 保留端口的输出行动。
- Optional: select: 执行组中的一个存储段。报文基于一个计算交换机选择算法 (如一些用户配置数组的哈希算法或简单的循环算法) 被组中的一个存储段处理。所有的配置和状态的选择算法都在 OpenFlow 外部运行。选择算法实现可以使用等负

荷分配，页可以有选择地根据存储段的权重进行。当存储段中指定的端口出现故障时，交换机可以选择剩余的存储段（生存端口的转发行为），而不是丢弃数据包。此行为可能会减少报文在链路或交换机的中断。

- Required: indirect: 执行此组中定义的一个存储段。这个组只支持单一的存储段。允许多个流表项或者组指向一个共同的组编号，这样可以使转发更快，更高效的聚集（例如：下一跳 IP 转发）。对于所有组的这个存储段，组类型应该是相同的。
- Optional: fast failover: 执行第一个有效的存储段。该 group 中每个的存储段均关联一个特定的端口或者 group，并按照该 group 定义的顺序进行排列，且每个的存储段均可以被控制是否生效。该 group 会选取第一个有效的存储段执行，如果没有有效的的存储段，报文会被丢弃。

5) 输入权重值

仅在 group 存储段类型为“select”时有效。

6) 输入“watch port”

指定端口，该端口状态影响当前存储段是否生效。仅在存储段类型为“fast failover”时有效。

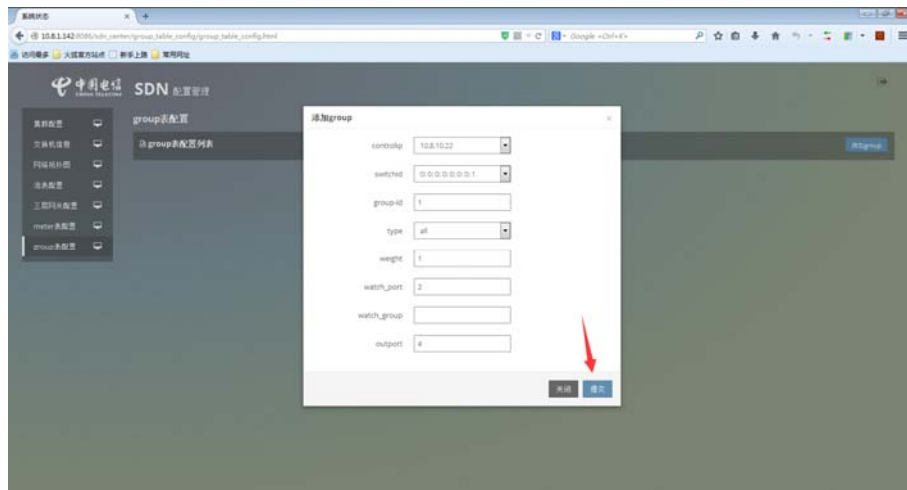
7) 输入“watch group”

指定 group，该 group 状态影响当前存储段是否生效。仅在存储段类型为“fast failover”时有效。

8) 输入“outport”

指定报文出口。

9) 提交



## 3 sFabric 功能

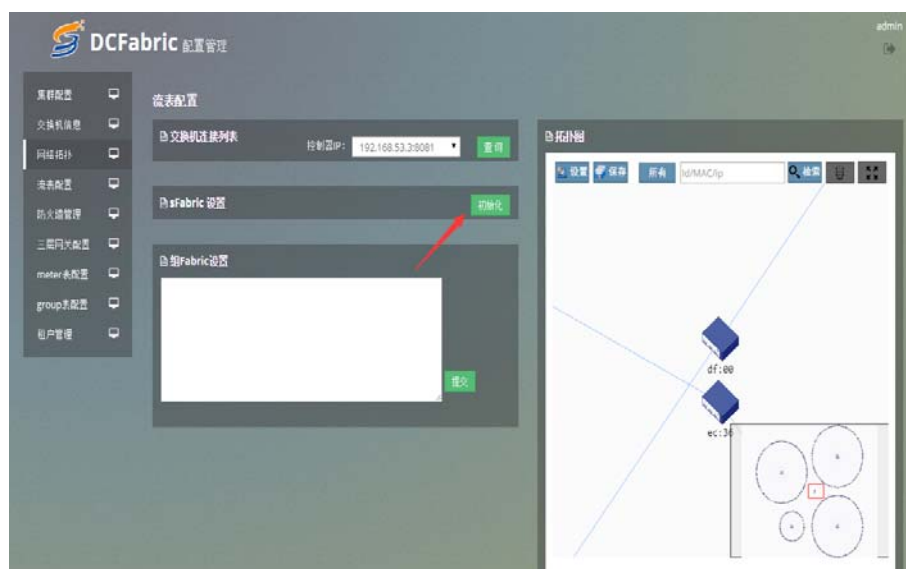
### 3.1 sFabric 功能介绍

### 3.2 Fabric 初始化

步骤四： 选择菜单栏“网络拓扑图”



步骤五： 点击初始化



## 3.3 Fabric 路径更新

### 3.3.1 添加交换机

观察控制器的日志界面，当出现下图的打印信息时，表示该交换机已经增加到拓扑中，并且 fabric 路径已经自动更新，该变化是即时发生的，响应时间<4 秒。

```
[INFO] New OpenFlow13 switch [00:00:00:00:00:00:01:95] connected: ip[192.168.52.134:53307]
[INFO] call_delete_switch_fun, sw num:1
[INFO] call_add_switch_fun, sw num:1
[INFO] call_add_switch_port_fun
[INFO] update_fabric_swap_flows begin!
[INFO] topo_change_event_thread start
[INFO] topo_change_event_thread start
[INFO] topo_change_event_thread start
[INFO] topo_change_event_thread start
[INFO] topo_change_event_thread start
```

### 3.3.2 删除交换机

观察控制器的日志界面，当出现下图的打印信息时，表示该交换机已经从拓扑中删除，并且 fabric 路径已经自动更新，该变化是即时发生的，响应时间<0.5 秒。

```
[INFO] topo_change_event_thread start
[WARNING] switch [00:00:00:00:00:00:01:95] disconnected
[INFO] topo_change_event_thread start
[INFO] call_delete_switch_port_fun
[INFO] update_fabric_swap_flows begin!
```

### 3.3.3 连接数据线

观察控制器的日志界面，当出现下图的打印信息时，表示一个新的边已经增加到拓扑中，并且 fabric 路径已经自动更新，该变化是即时发生的，响应时间<4 秒。

```
[INFO] topo_change_event_thread start
[INFO] Port state change: eth1[new state: 0]
[INFO] Port state change: eth-0-9[new state: 0]
[INFO] topo_change_event_thread start
[INFO] call_add_switch_port_fun
[INFO] update_fabric_swap_flows begin!
[INFO] topo_change_event_thread start
```

### 3.3.4 移除数据线

观察控制器的日志界面，当出现下图的打印信息时，表示一个边已

经从拓扑中删除，并且 fabric 路径已经自动更新，该变化是即时发生的，响应时间<0.5 秒。

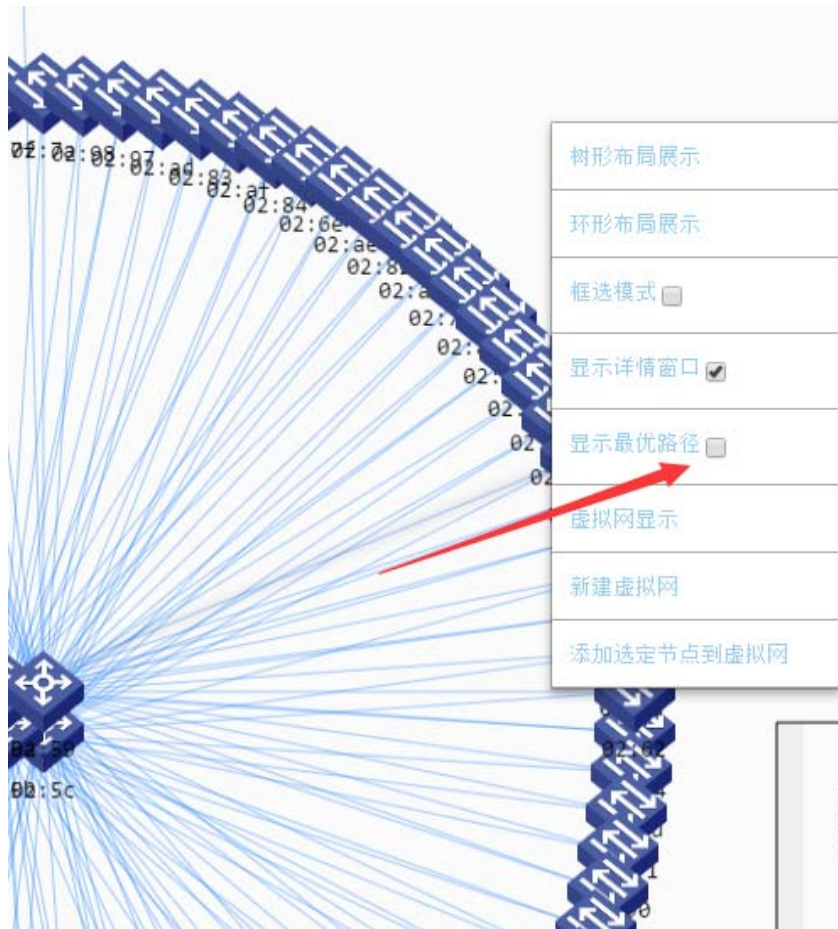
```
[INFO] topo_change_event_thread start
[INFO] Port state change: eth1[new state: 1]
[INFO] topo_change_event_thread start
[INFO] Port state change: eth-0-9[new state: 1]
[INFO] call_delete_switch_port_fun
[INFO] update_fabric_swap_flows begin!
[INFO] topo_change_event_thread start
[INFO] Port state change: eth1[new state: 0]
```

### 3.4 Fabric 路径查询

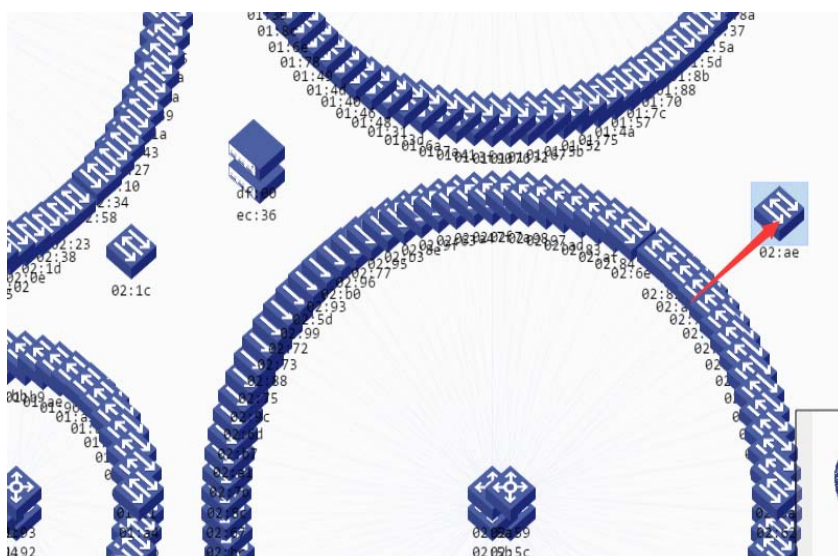
步骤六： 在菜单中选择“网络拓扑”，查看拓扑图



步骤七： 右击鼠标勾选上显示最优路径

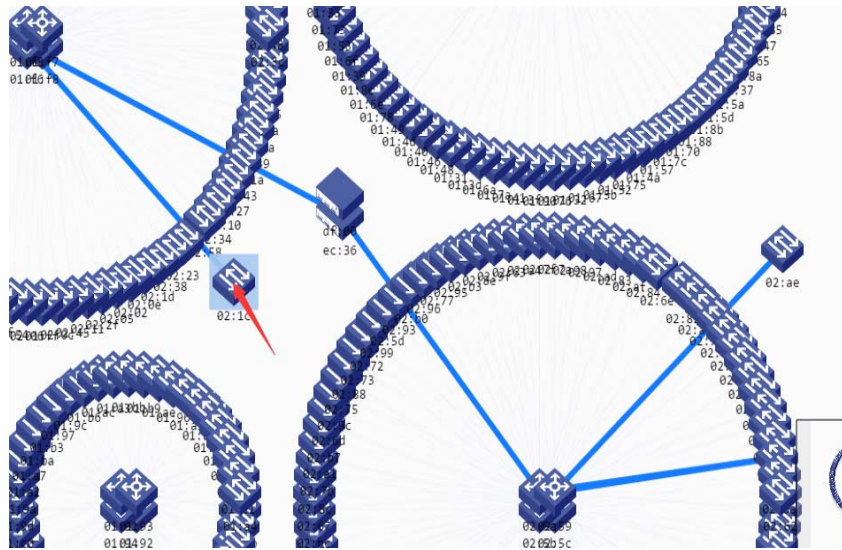


步骤八： 点选起始位置交换机





步骤九： 点击终点位置交换机，即能显示该 2 点之间的 fabric 路径



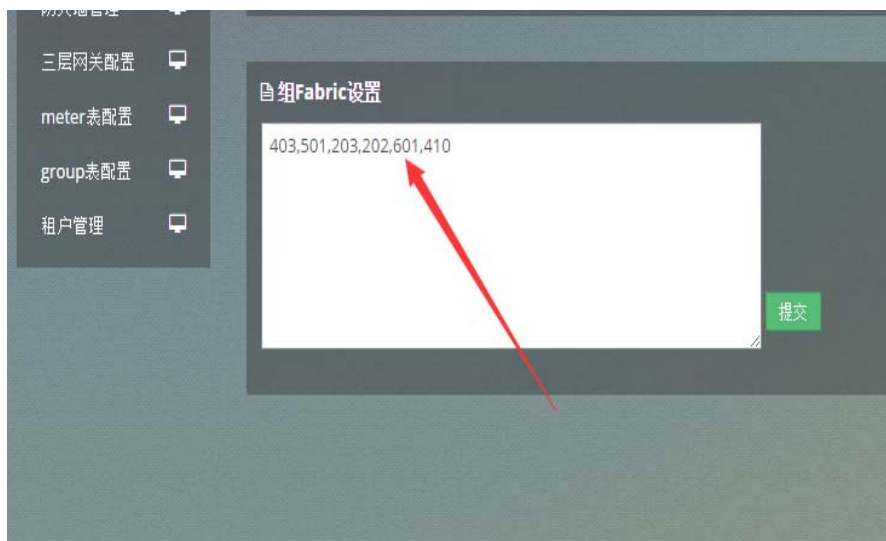
### 3.5 组 fabric 设置

步骤十： 在菜单中选择“网络拓扑”





步骤十一：在组 fabric 设置的多行文本框内输入想要设置的部分交换机的 dpid，以逗号隔开；



步骤十二：点击提交，该组交换机即可创建 fabric 路径。



## 4 附录

### 4.1 流表匹配项前置约束

匹配项	支持掩码	前置条件	备注
-----	------	------	----

IN_PORT	否	无	报文的入端口，可以是交换机的普通端口或逻辑端口
IN_PHY_PORT	否	IN_PORT	报文的物理入端口
METADATA	是	无	流表的元数据，用于不同流表间传递信息
ETH_DST	是	无	以太网目的 MAC 地址
ETH_SRC	是	无	以太网源 MAC 地址
ETH_TYPE	否	无	以太网类型
VLAN_VID	是	无	802.1Q 中定义的 VLAN ID
VLAN_PCP	否	VLAN_VID 不为空	802.1Q 中定义的 VLAN 优先级
IP_DSCP	否	ETH_TYPE=0x0800 ETH_TYPE=0x86dd	Diff Serv Code Point
IP_ECN	否	ETH_TYPE=0x0800 ETH_TYPE=0x86dd	IP 头中的 ECN 字段
IP_PROTO	否	ETH_TYPE=0x0800 ETH_TYPE=0x86dd	IPv4 及 IPv6 的协议号
IPv4_SRC	是	ETH_TYPE=0x0800	IPv4 源 IP 地址
IPv4_DST	是	ETH_TYPE=0x0800	IPv4 目的 IP 地址
TCP_SRC	否	IP_PROTO=6	TCP 源端口
TCP_DST	否	IP_PROTO=6	TCP 目的端口
UDP_SRC	否	IP_PROTO=17	UDP 源端口
UDP_DST	否	IP_PROTO=17	UDP 目的端口
SCTP_SRC	否	IP_PROTO=132	SCTP 源端口
SCTP_DST	否	IP_PROTO=132	SCTP 目的端口
ICMPV4_TYPE	否	IP_PROTO=1	ICMP 类型
ICMPV4_CODE	否	IP_PROTO=1	ICMP code
ARP_OP	否	ETH_TYPE=0x0806	ARP opcode
ARP_SPA	是	ETH_TYPE=0x0806	ARP 报文中的 IPv4 源地址
ARP_TPA	是	ETH_TYPE=0x0806	ARP 报文中的 IPv4 目标地址
ARP_SHA	是	ETH_TYPE=0x0806	ARP 报文中的以太网源 MAC 地址
ARP_THA	是	ETH_TYPE=0x0806	ARP 报文中的以太网目标 MAC 地址
IPv6_SRC	是	ETH_TYPE=0x86dd	IPv6 源 IP 地址
IPv6_DST	是	ETH_TYPE=0x86dd	IPv6 目的 IP 地址
IPv6_FLABEL	是	ETH_TYPE=0x86dd	IPv6 流标签
ICMPV6_TYPE	否	IP_PROTO=58	ICMPv6 类型

ICMPV6_CODE	否	IP_PROTO=58	ICMPv6 code
IPV6_ND_TARGET	否	ICMPV6 TYPE=135 ICMPV6 TYPE=136	IPv6 邻居发现报文中的目标地址
IPV6_ND_SLL	否	ICMPV6 TYPE=135	IPv6 邻居发现报文中的源链路层地址
IPV6_ND_TLL	否	ICMPV6 TYPE=136	IPv6 邻居发现报文中的目标链路层地址
MPLS_LABEL	否	ETH TYPE=0x8847 ETH TYPE=0x8848	首个 MPLS 头中的标签
MPLS_TC	否	ETH TYPE=0x8847 ETH TYPE=0x8848	首个 MPLS 头中的 TC
MPLS_BOS	否	ETH TYPE=0x8847 ETH TYPE=0x8848	首个 MPLS 头中的 BoS
PBB_ISID	是	ETH TYPE=0x88e7	首个 PBB 服务实例标签中的 I-SID
TUNNEL_ID	是	无	逻辑口分配的 metadata
IPV6_EXTHDR	是	ETH TYPE=0x86dd	IPv6 扩展头的虚拟字段