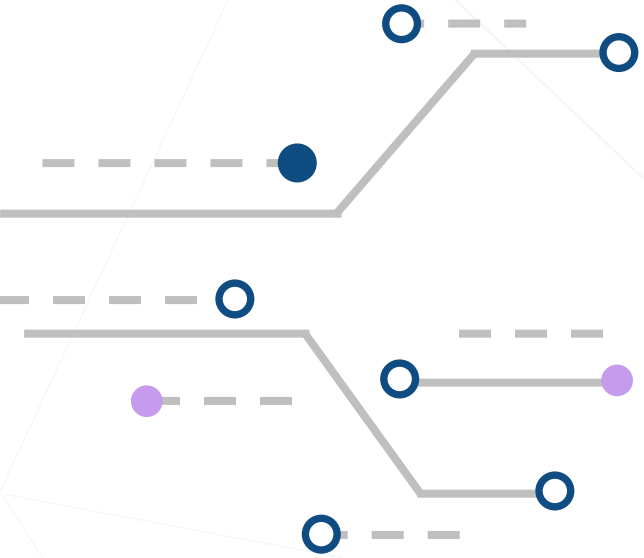


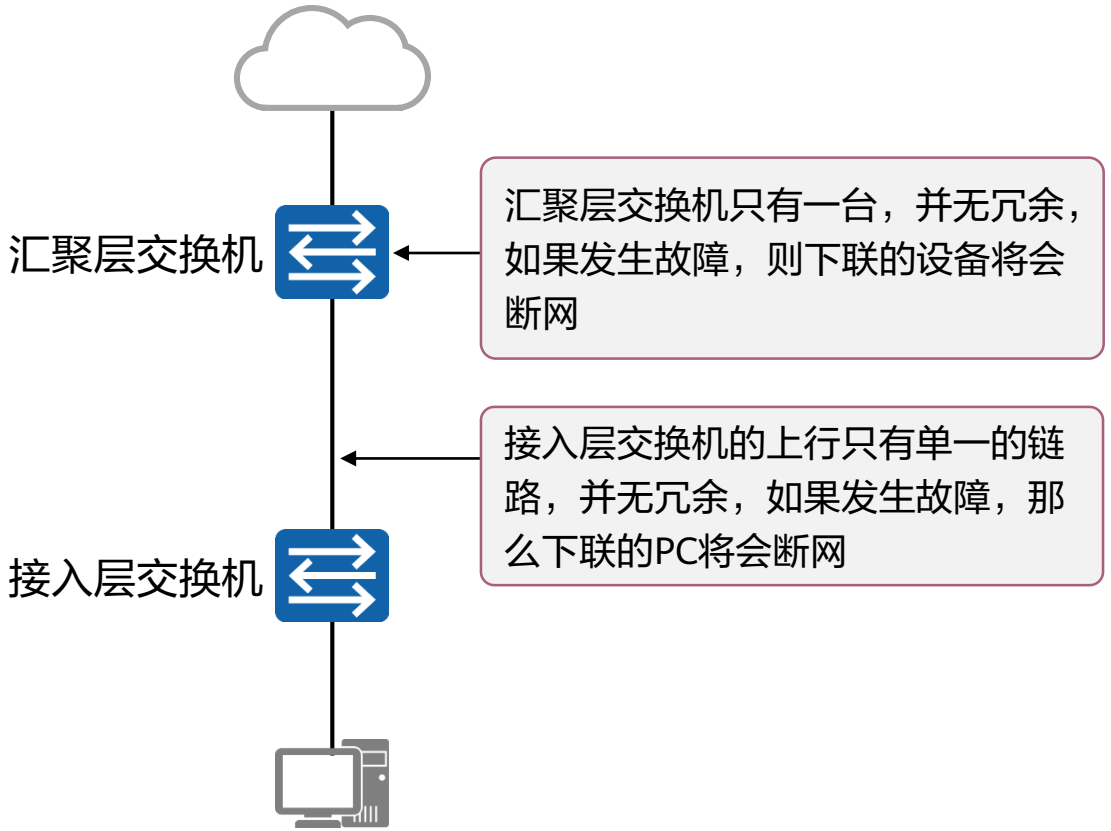
生 成 树

主讲人：鲍婷婷

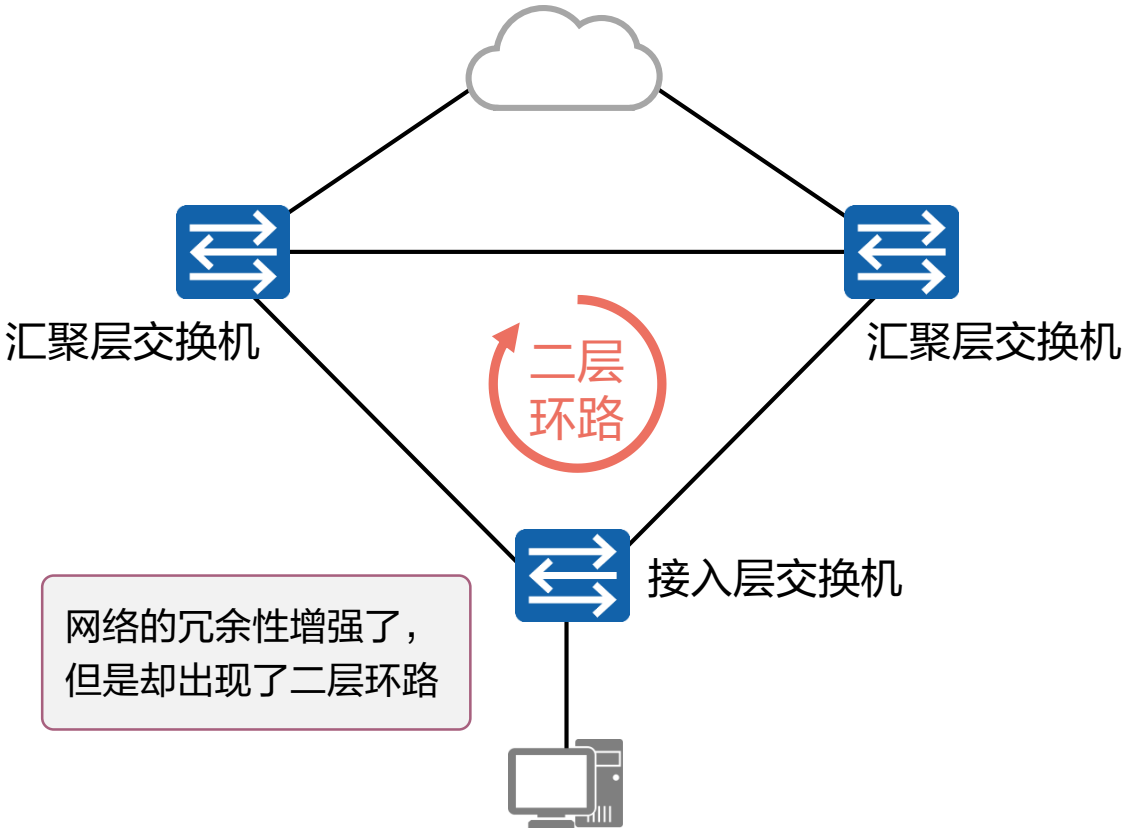
目录

技术背景：二层交换机网络的冗余性与环路

一个缺乏冗余性设计的网络

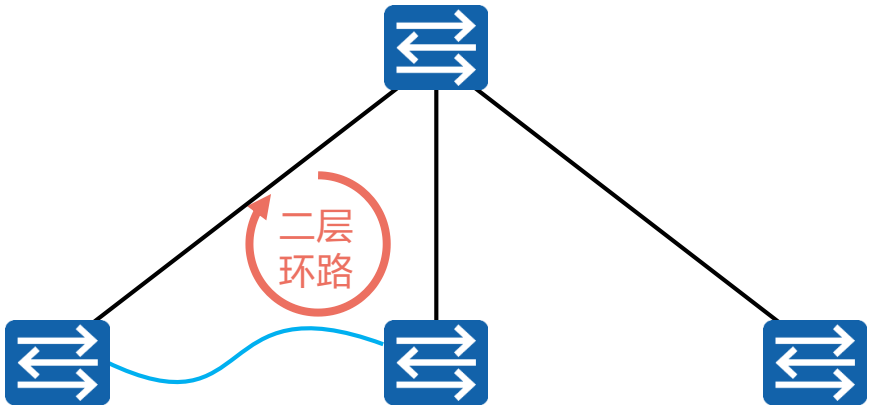


引入冗余性的同时也引入了二层环路



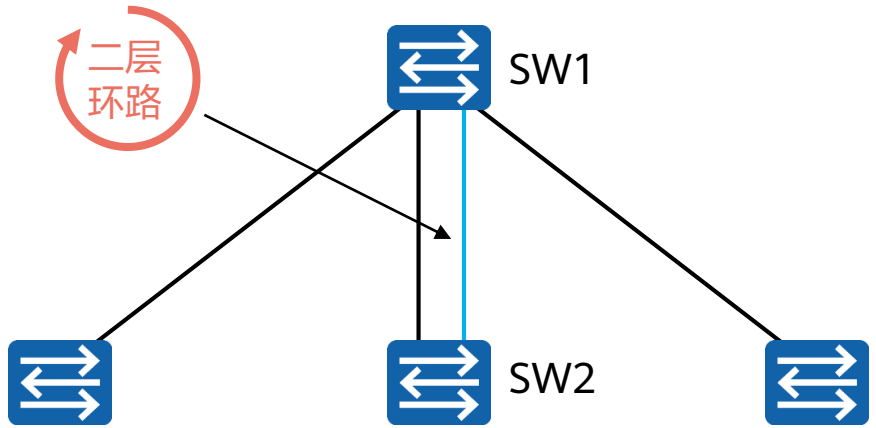
技术背景：人为错误导致的二层环路

人为错误导致的二层环路 案例1



设备之间的互联线缆连接错误

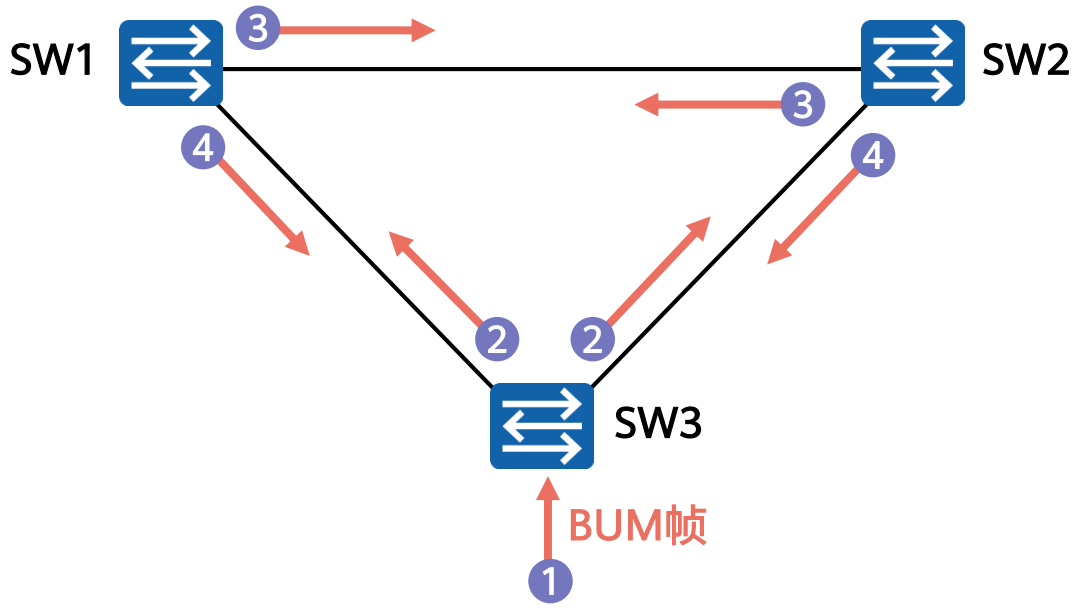
人为错误导致的二层环路 案例2



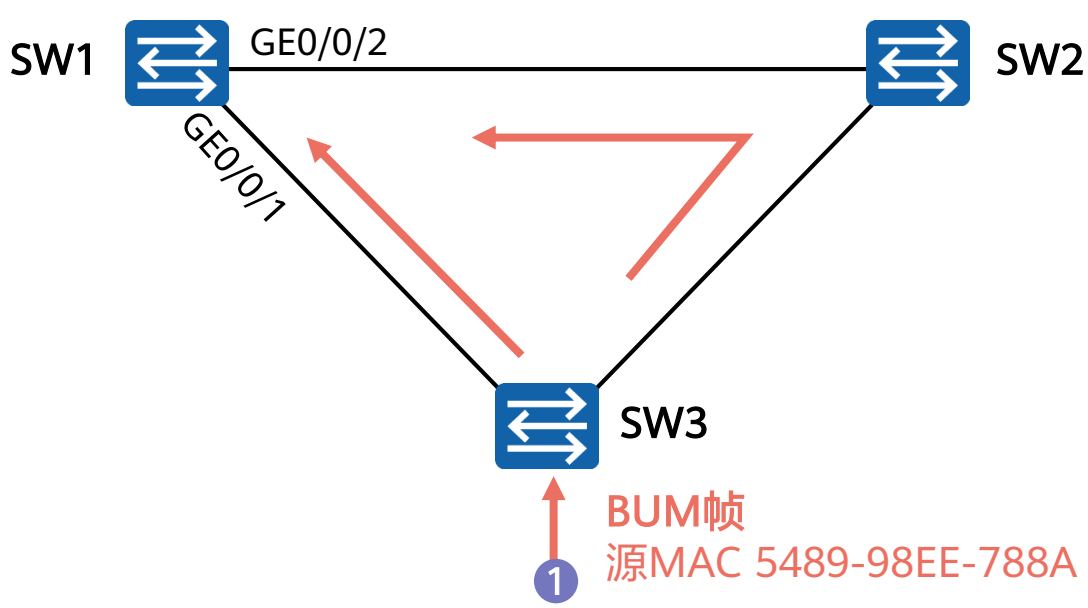
SW1与SW2之间的链路未绑定到一个逻辑链路（聚合链路）上

二层环路带来的问题

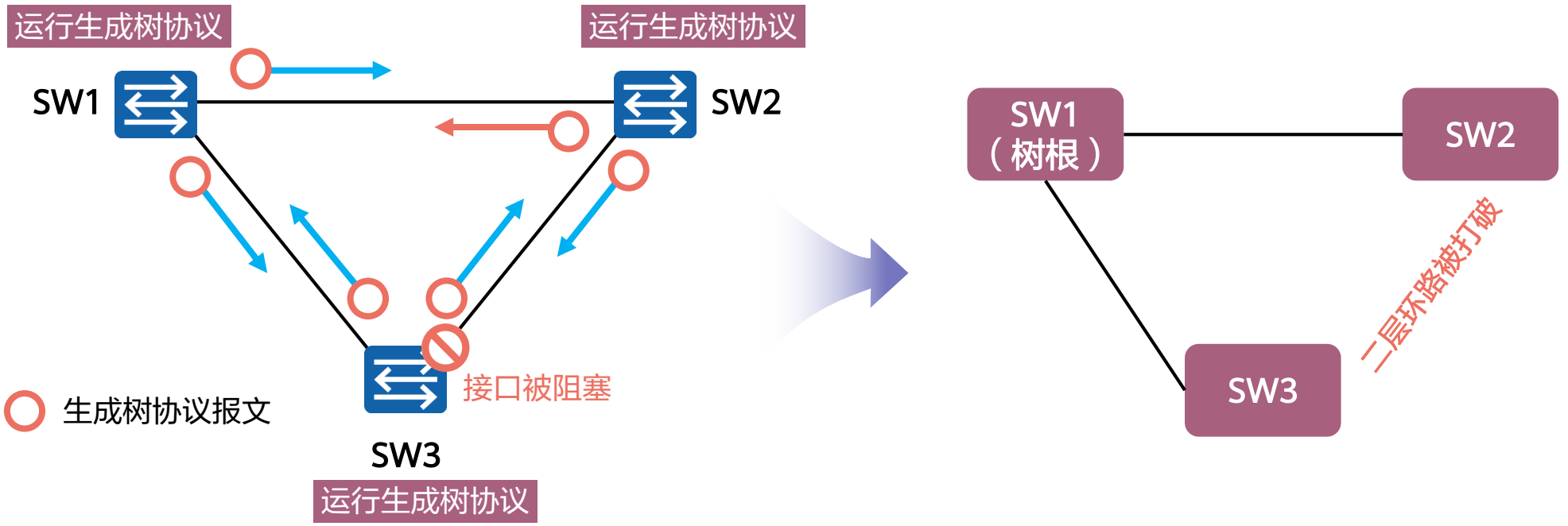
典型问题1：广播风暴



典型问题2：MAC地址漂移

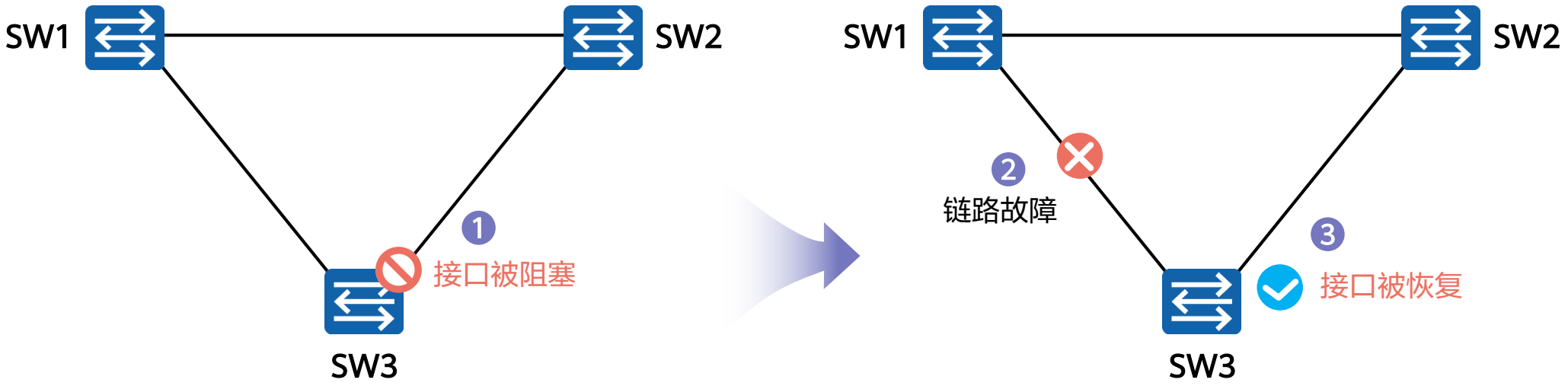


初识生成树协议



在网络中部署生成树后，交换机之间会进行生成树协议报文的交互并进行无环拓扑计算，最终将网络中的某个（或某些）接口进行阻塞（Block），从而打破环路。

生成树能够动态响应网络拓扑变化调整阻塞接口

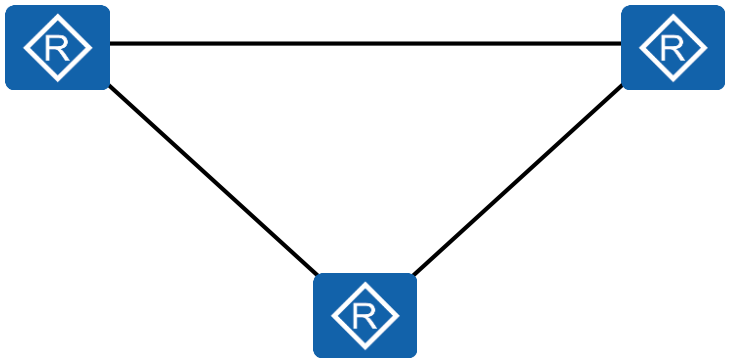


交换机上运行的生成树协议会持续监控网络的拓扑结构，当网络拓扑结构发生变化时，生成树能感知到这些变化，并且自动做出调整。

因此，生成树既能解决二层环路问题，也能为网络的冗余性提供一种方案。

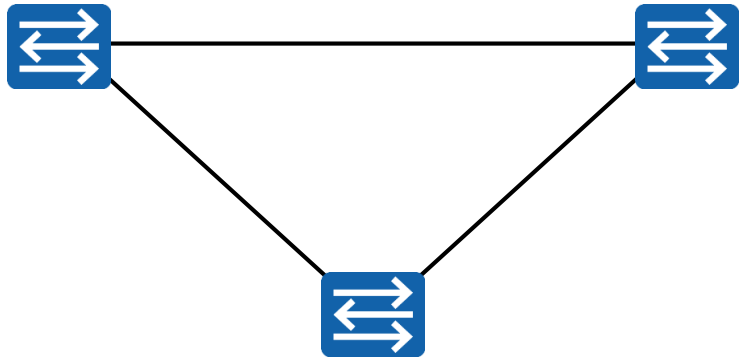
问答：二层及三层环路

三层环路（Layer 3 Loop）



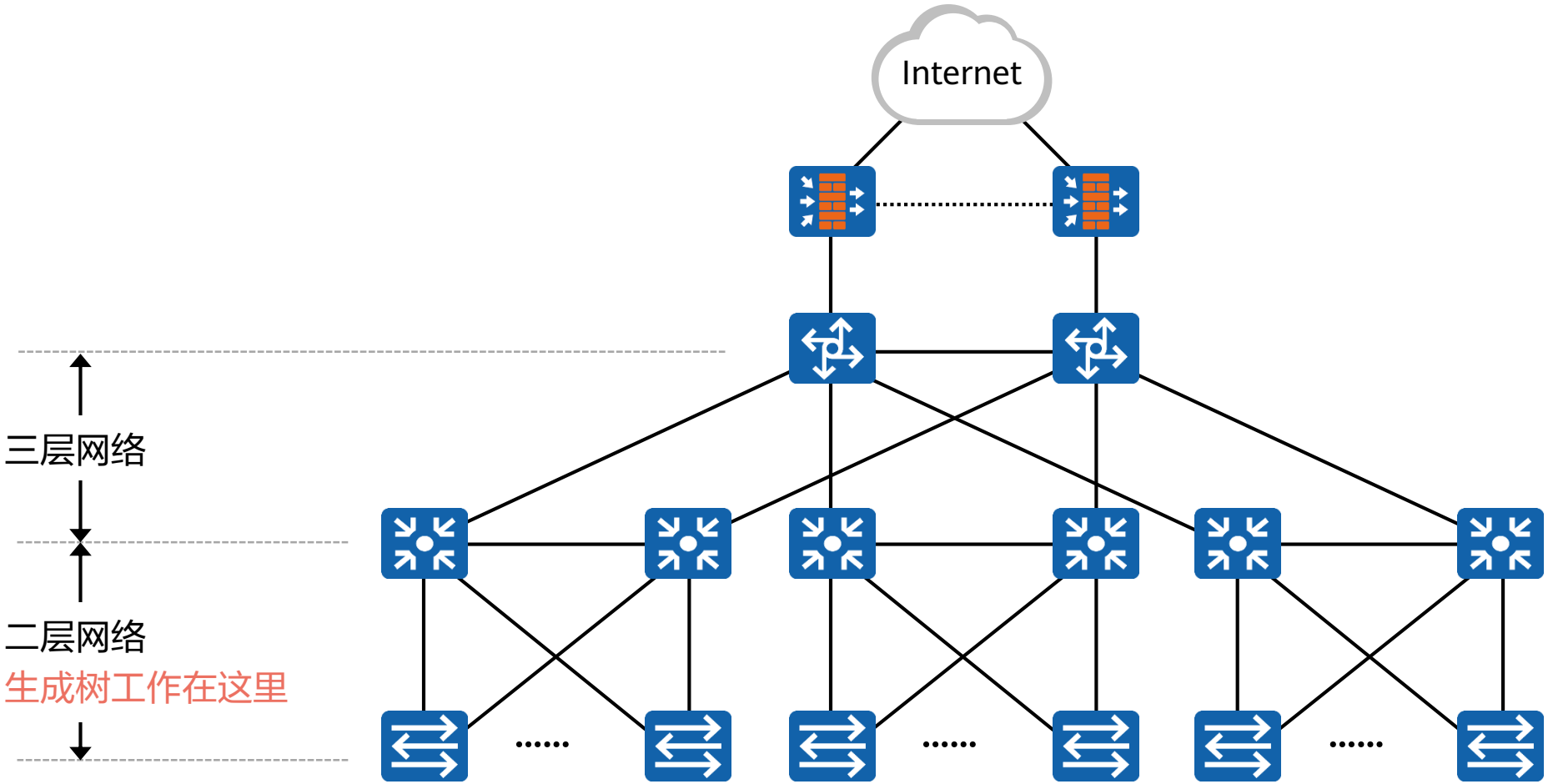
- 常见根因：路由环路
- 动态路由协议有一定的防环能力；
- IP报文头部中的TTL字段可用于防止报文被无止尽地转发。

二层环路（Layer 2 Loop）



- 常见根因：二层冗余环境
- 需借助特定的协议或机制实现二层防环；
- 二层帧头中并没有任何信息可用于防止数据帧被无止尽地转发。

生成树协议在园区网络中的应用位置

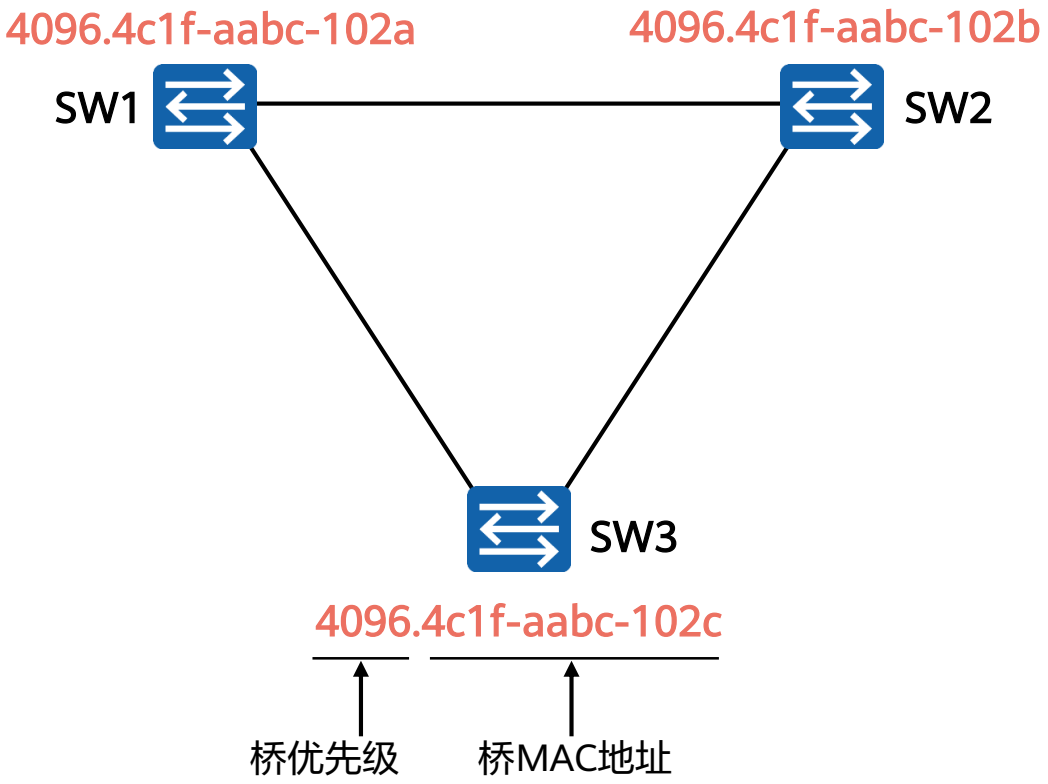


STP概述

- STP是一个用于局域网中消除环路的协议。
- 功能一：防止环路。
- 功能二：提供冗余备份链路。

目录

STP的基本概念：桥ID

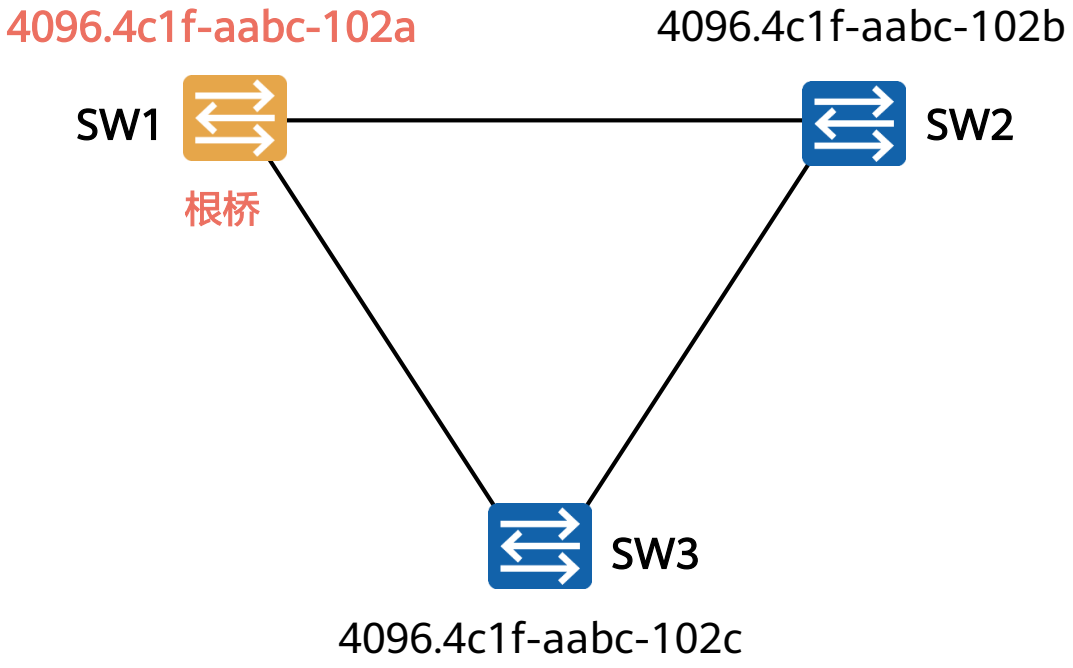


桥ID（Bridge ID，BID）

- IEEE 802.1D标准中规定BID由16位的桥优先级（Bridge Priority）与桥MAC地址构成。
- 每一台运行STP的交换机都拥有一个唯一的BID。
- BID桥优先级占据高16bit，其余的低48bit是桥MAC地址。
- 在STP网络中，BID最小的设备会被选举为根桥。

备注：此处网桥（Bridge），或者桥也就是交换机。

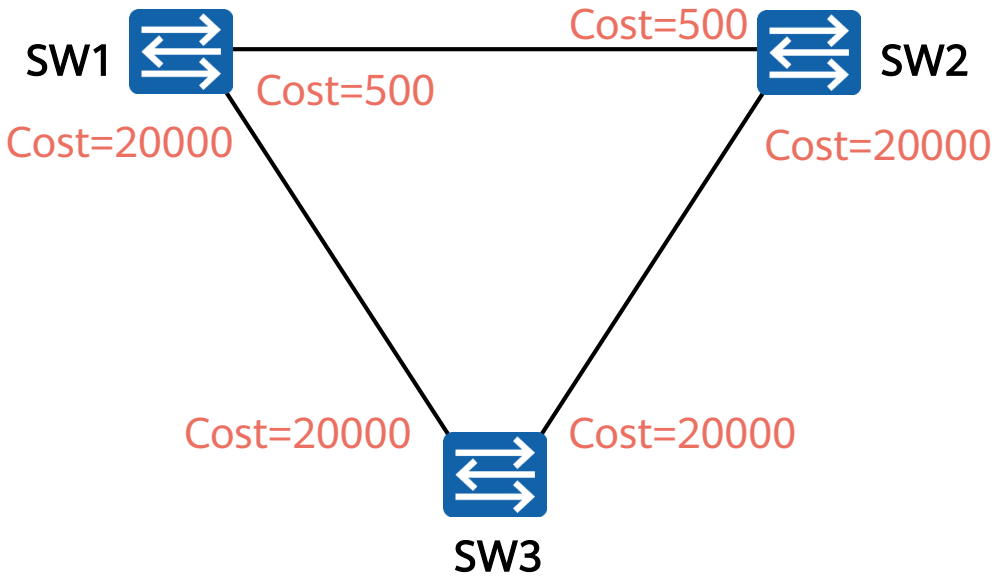
STP的基本概念：根桥



根桥（Root Bridge）

- STP的主要作用之一是在整个交换网络中计算出一棵无环的“树”（STP树）。
- 根桥是一个STP交换网络中的“树根”。
- STP开始工作后，会在交换网络中选举一个根桥，作为无环拓扑的“树根”。
- 在STP网络中，桥ID最小的设备会被选举为根桥。
  1. 首先比较桥优先级，优先级的值越小，则越优先；
  2. 如果优先级相等，那么再比较MAC地址，拥有最小MAC地址的交换机会成为根桥。

STP的基本概念：Cost



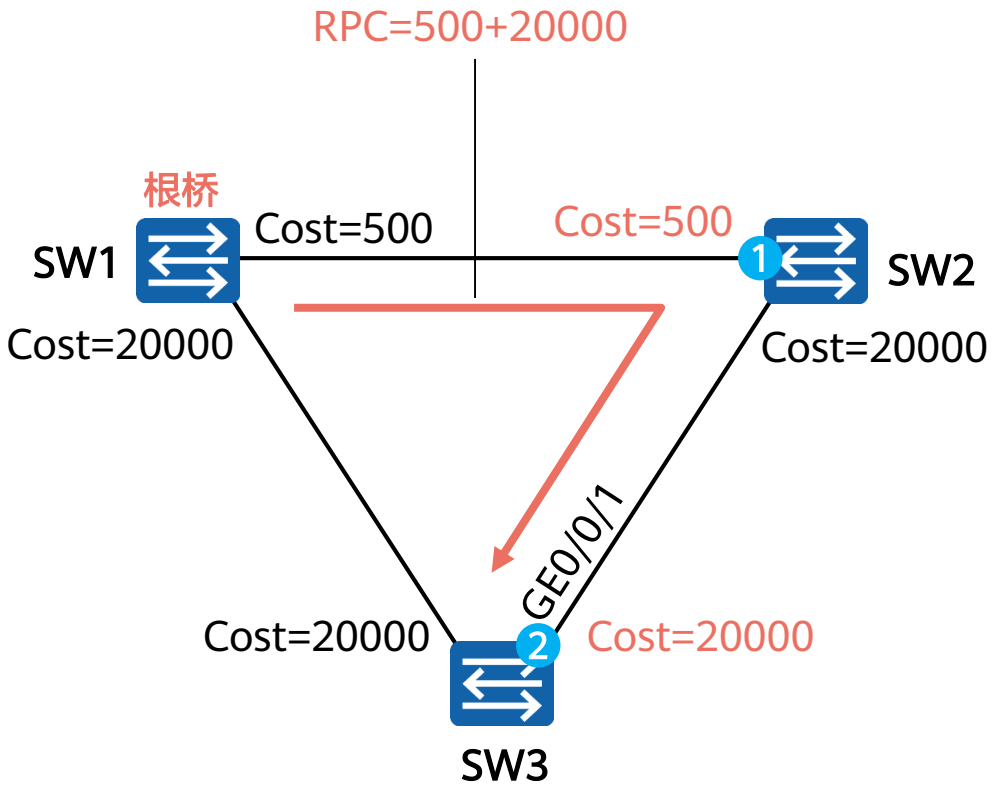
开销（Cost）

- 接口的Cost主要用于计算根路径开销，也就是到达根的开销。
- 接口的缺省Cost除了与其速率、工作模式有关，还与交换机使用的STP Cost计算方法有关。
- 接口带宽越大，则Cost值越小。
- 用户也可以根据需要通过命令调整接口的Cost。

STP的基本概念：Cost计算方法

接口速率	接口模式	STP开销（推荐值）		
		IEEE 802.1d-1998标准	IEEE 802.1t标准	华为计算方法
100Mbps	Half-Duplex	19	200,000	200
	Full-Duplex	18	199,999	199
	Aggregated Link 2 Ports	15	100,000	180
1000Mbps	Full-Duplex	4	20,000	20
	Aggregated Link 2 Ports	3	10,000	18
10Gbps	Full-Duplex	2	2000	2
	Aggregated Link 2 Ports	1	1000	1
40Gbps	Full-Duplex	1	500	1
	Aggregated Link 2 Ports	1	250	1
100Gbps	Full-Duplex	1	200	1
	Aggregated Link 2 Ports	1	100	1
.....				

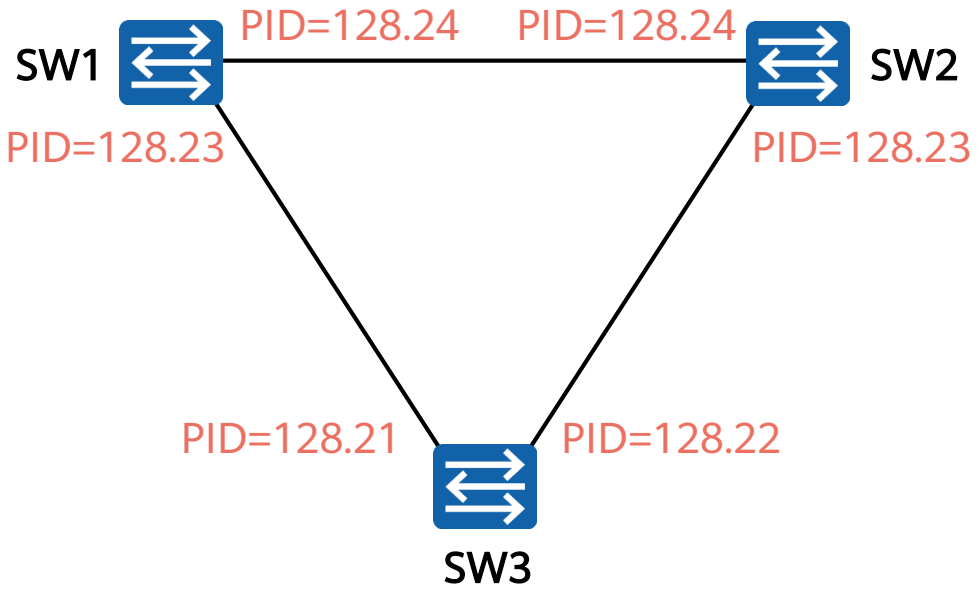
STP的基本概念：RPC



根路径开销（Root Path Cost）

- 一台设备从某个接口到达根桥的RPC等于从根桥到该设备沿途所有入方向接口的Cost累加。

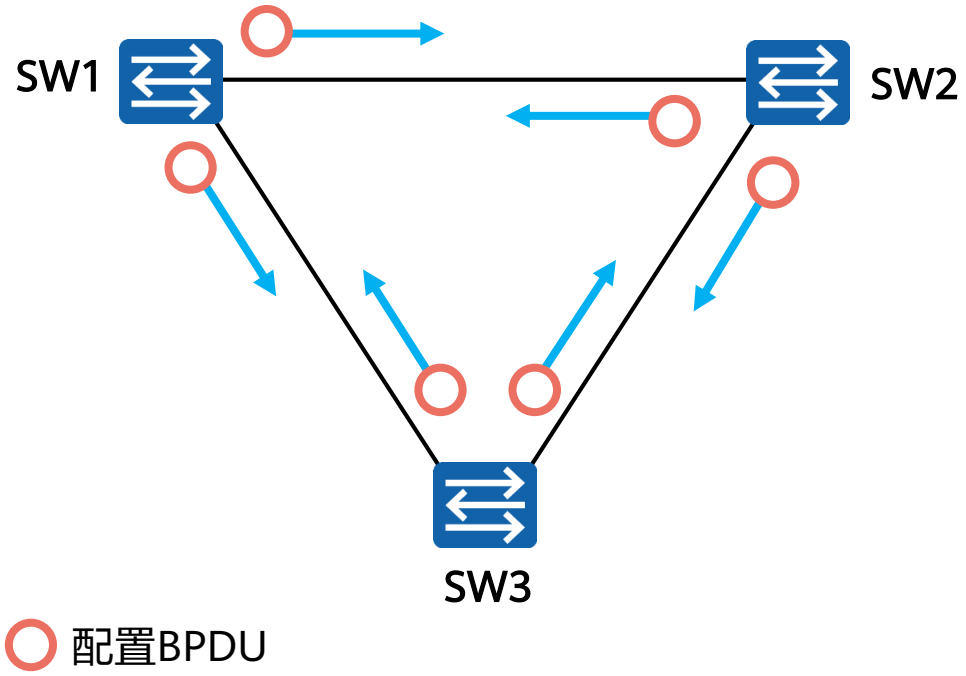
STP的基本概念：Port ID



接口ID（Port ID，PID）

- 接口ID由两部分构成的，高4 bit是接口优先级，低12 bit是接口编号。
- 激活STP的接口会维护一个缺省的接口优先级，在华为交换机上，该值为128。

STP的基本概念：BPDU



BPDU（Bridge Protocol Data Unit，网桥协议数据单元）

- BPDU是STP的协议报文。
- BPDU分为两种类型：
  - 配置BPDU（Configuration BPDU）
  - TCN BPDU（Topology Change Notification BPDU）
- 配置BPDU是STP进行拓扑计算的关键；
- TCN BPDU只在网络拓扑发生变更时才会被触发。

配置BPDU的报文格式

PID	PVI	BPDU Type	Flags	Root ID	RPC	Bridge ID	Port ID	Message Age	Max Age	Hello Time	Forward Delay

字节	字段	描述
2	PID	协议ID ，对于STP而言，该字段的值总为0
1	PVI	协议版本ID，对于STP而言，该字段的值总为0
1	BPDU Type	指示本BPDU的类型，若值为0x00，则表示本报文为配置BPDU；若值为0x80，则为TCN BPDU
1	Flags	标志，STP只使用了该字段的最高及最低两个比特位，最低位是TC（Topology Change，拓扑变更）标志，最高位是TCA（Topology Change Acknowledgment，拓扑变更确认）标志
8	Root ID	根网桥的桥ID
4	RPC	根路径开销，到达根桥的STP Cost
8	Bridge ID	BPDU发送桥的ID
2	Port ID	BPDU发送网桥的接口ID（优先级+接口号）
2	Message Age	消息寿命，从根网桥发出BPDU之后的秒数，每经过一个网桥都加1，所以它本质上是到达根桥的跳数
2	Max Age	最大寿命，当一段时间未收到任何BPDU，生存期到达最大寿命时，网桥认为该接口连接的链路发生故障。默认20s
2	Hello Time	根网桥连续发送的BPDU之间的时间间隔，默认2s
2	Forward Delay	转发延迟，在侦听和学习状态所停留的时间间隔，默认15s

配置BPDU的比较原则

字段
协议ID
协议版本ID
类型
标志
根桥ID
根路径开销
网桥ID
接口ID
消息寿命
最大寿命
Hello时间
转发延迟

STP操作：

1. 选举一个根桥。
2. 每个非根交换机选举一个根端口。
3. 每个网段选举一个指定端口。
4. 阻塞非根、非指定端口。

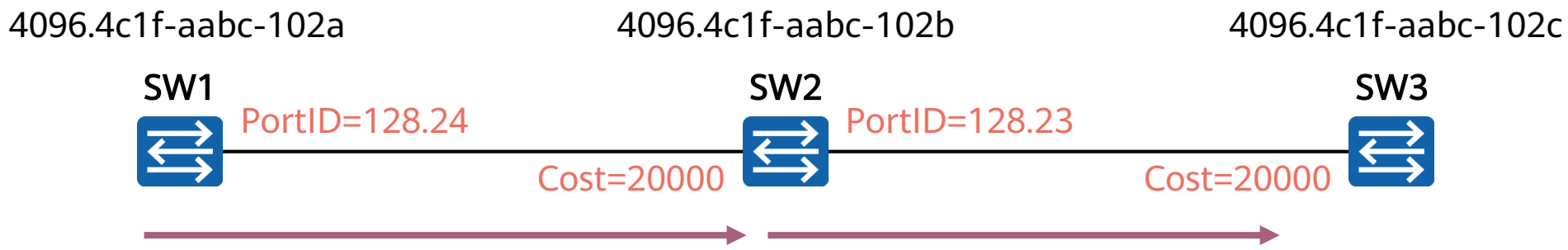
STP中定义了三种端口角色：指定端口，根端口和预备端口。

STP按照如下顺序选择最优的配置BPDU：

1. 最小的根桥ID
2. 最小的RPC
3. 最小的网桥ID
4. 最小的接口ID

在这四条原则中（每条原则都对应配置BPDU中的相应字段），第一条原则主要用于在网络中选举根桥，后面的原则主要用于选举根接口及指定接口。

配置BPDU的转发过程

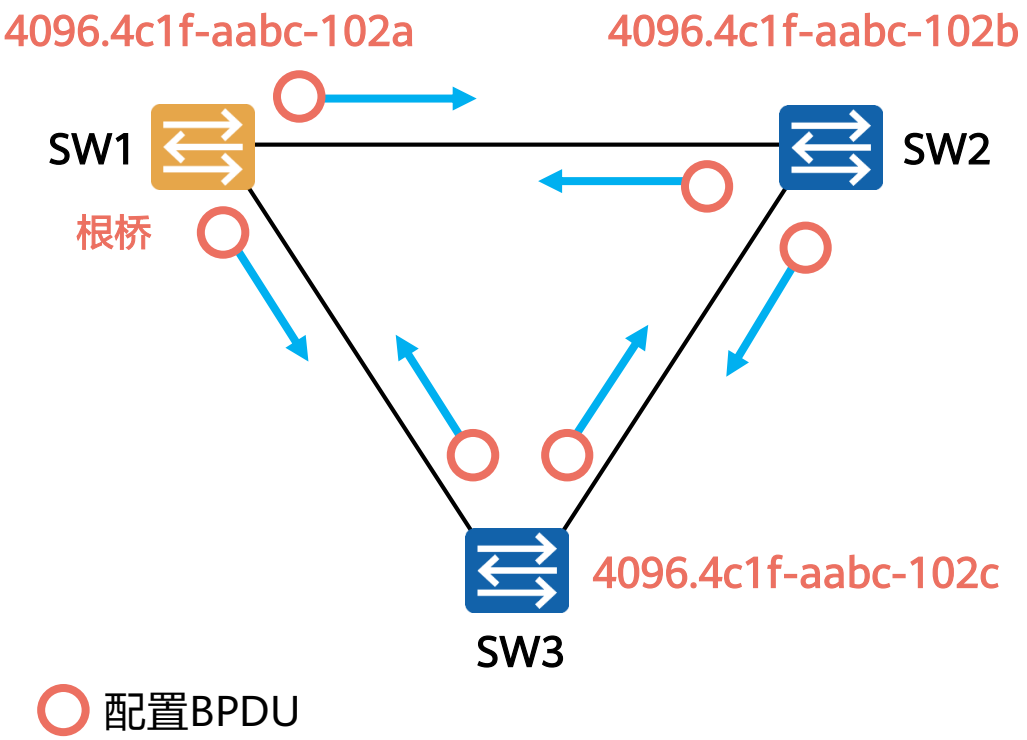


配置BPDU
……
根桥ID=4096.4c1f-aabc-102a
路径开销=0
桥ID=4096.4c1f-aabc-102a
接口ID=128.24
……

配置BPDU
……
根桥ID=4096.4c1f-aabc-102a
路径开销=0+20000
桥ID=4096.4c1f-aabc-102b
接口ID=128.23
……

目录

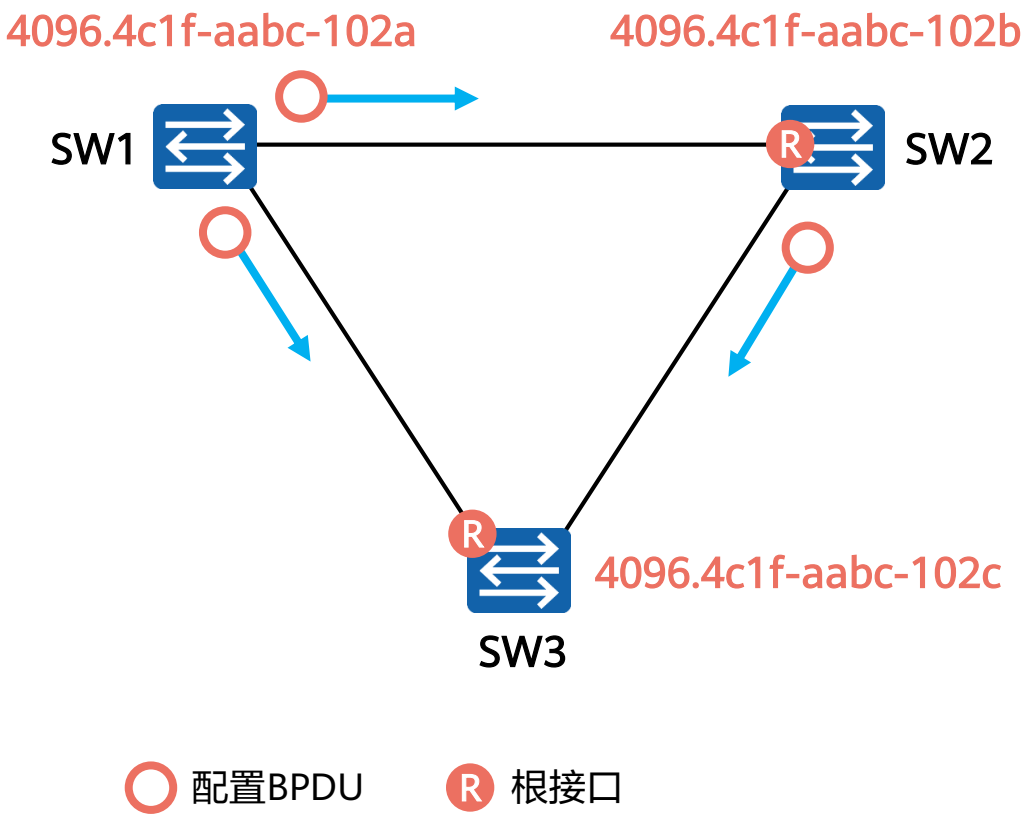
STP的计算过程 (1)



在交换网络中选举一个根桥

- STP在交换网络中开始工作后，每个交换机都会向网络中发送配置BPDU。配置BPDU中包含交换机自己的桥ID。
- 网络中拥有最小桥ID的交换机成为根桥。
- 在一个连续的STP交换网络中只会存在一个根桥。
- 根桥的角色是可抢占的。
- 为了确保交换网络的稳定，建议提前规划STP组网，并将规划为根桥的交换机的桥优先级设置为最小值0。

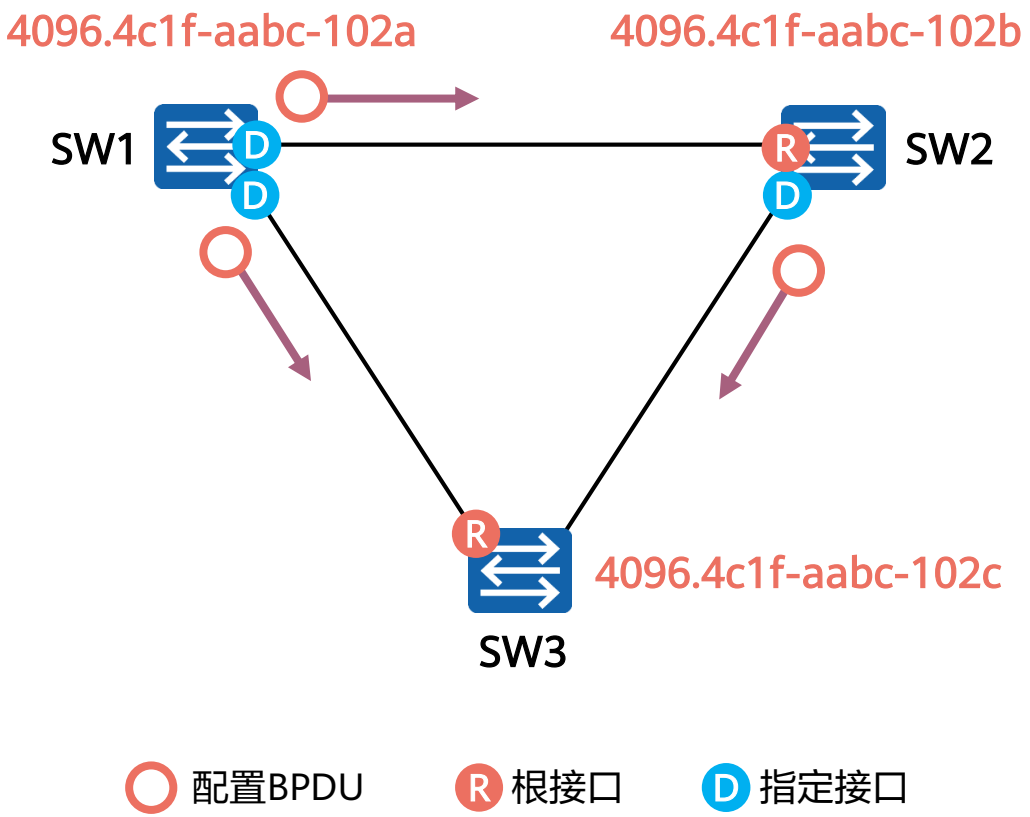
STP的计算过程 (2)



在每台非根桥上选举一个根接口

- 每一台非根桥交换机都会在自己的接口中选举出一个接口。
- 非根桥交换机上有且只会有一个根接口。
- 当非根桥交换机有多个接口接入网络中时，根接口是其收到最优配置BPDU的接口。
- 可以形象地理解为，根接口是每台非根桥上“朝向”根桥的接口。

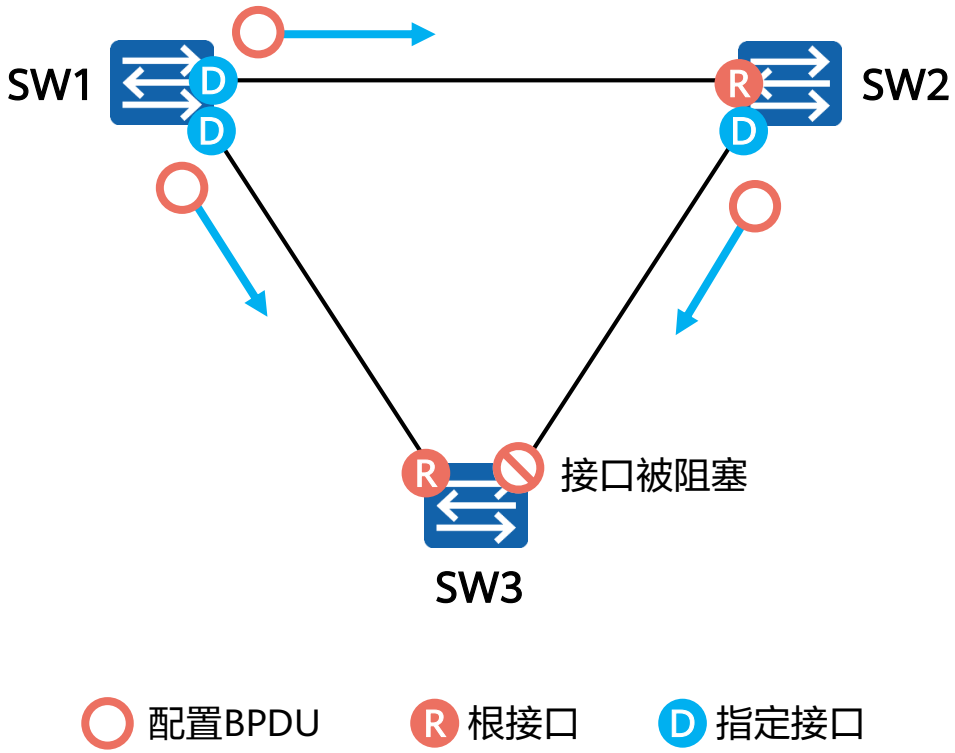
STP的计算过程 (3)



在每条链路上选举一个指定接口

- 根接口选举出来后，非根桥会使用其在该接口上收到的最优BPDU进行计算，然后将计算得到的配置BPDU与除了根接口之外的其他所有接口所收到的配置BPDU进行比较：
  - 如果前者更优，则该接口为指定接口；
  - 如果后者更优，则该接口为非指定接口。
- 一般情况下，根桥的所有接口都是指定接口。

STP的计算过程 (4)



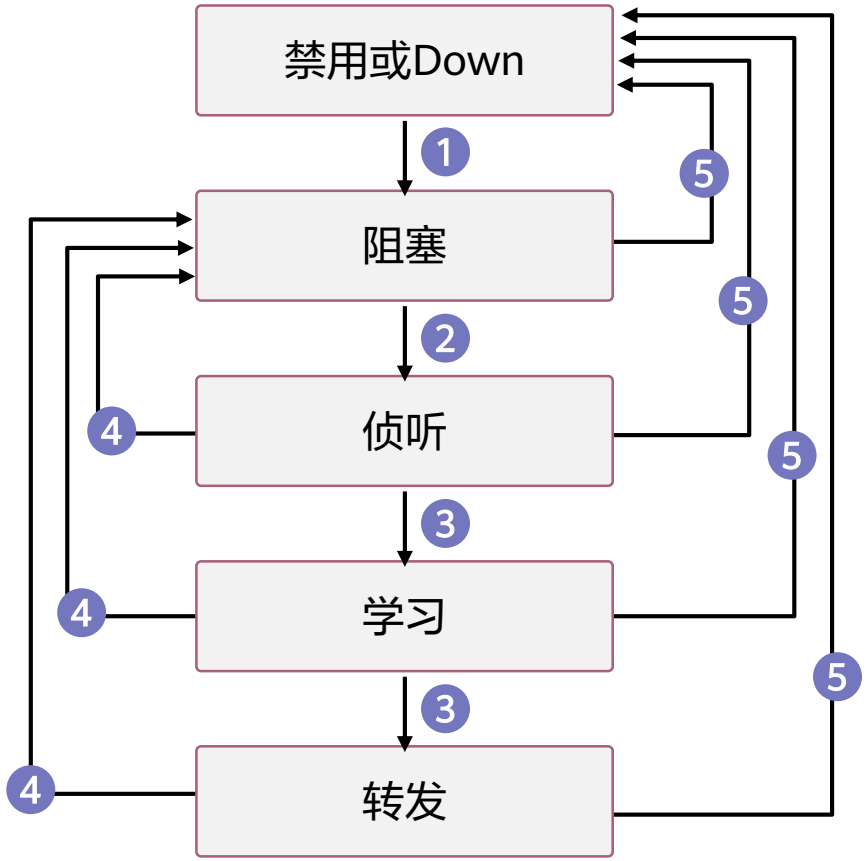
非指定接口被阻塞

- 一台交换机上，既不是根接口，又不是指定接口的接口被称为非指定接口。
- STP操作的最后一步是阻塞网络中的非指定接口。这一步完成后，网络中的二层环路就此消除。

STP的接口状态

状态名称	状态描述
禁用（Disable）	该接口不能收发BPDU，也不能收发业务数据帧，例如接口为down
阻塞（Blocking）	该接口被STP阻塞。处于阻塞状态的接口不能发送BPDU，但是会持续侦听BPDU，而且不能收发业务数据帧，也不会进行MAC地址学习
侦听（Listening）	当接口处于该状态时，表明STP初步认定该接口为根接口或指定接口，但接口依然处于STP计算的过程中，此时接口可以收发BPDU，但是不能收发业务数据帧，也不会进行MAC地址学习
学习（Learning）	当接口处于该状态时，会侦听业务数据帧（但是不能转发业务数据帧），并且在收到业务数据帧后进行MAC地址学习
转发（Forwarding）	处于该状态的接口可以正常地收发业务数据帧，也会进行BPDU处理。接口的角色需是根接口或指定接口才能进入转发状态

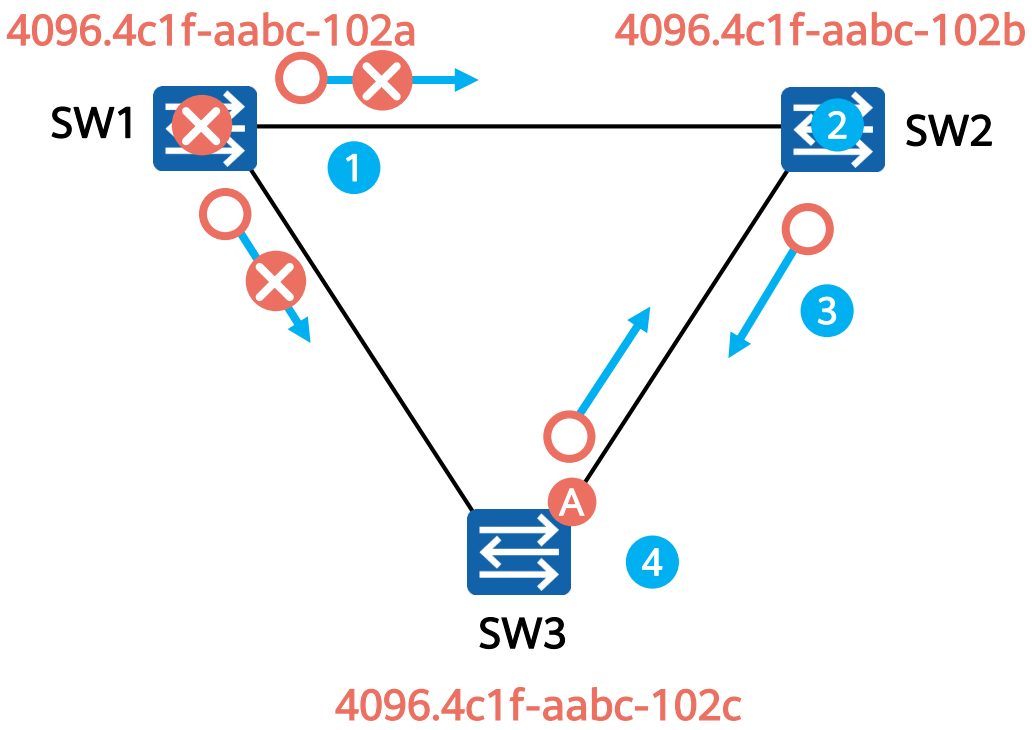
STP的接口状态迁移



1. 接口初始化或激活，自动进入阻塞状态
2. 接口被选举为根接口或指定接口，自动进入侦听状态
3. 转发延迟计时器超时且接口依然为根接口或指定接口
4. 接口不再是根接口或指定接口或指定状态
5. 接口被禁用或者链路失效

目录

拓扑变化 – 根桥故障

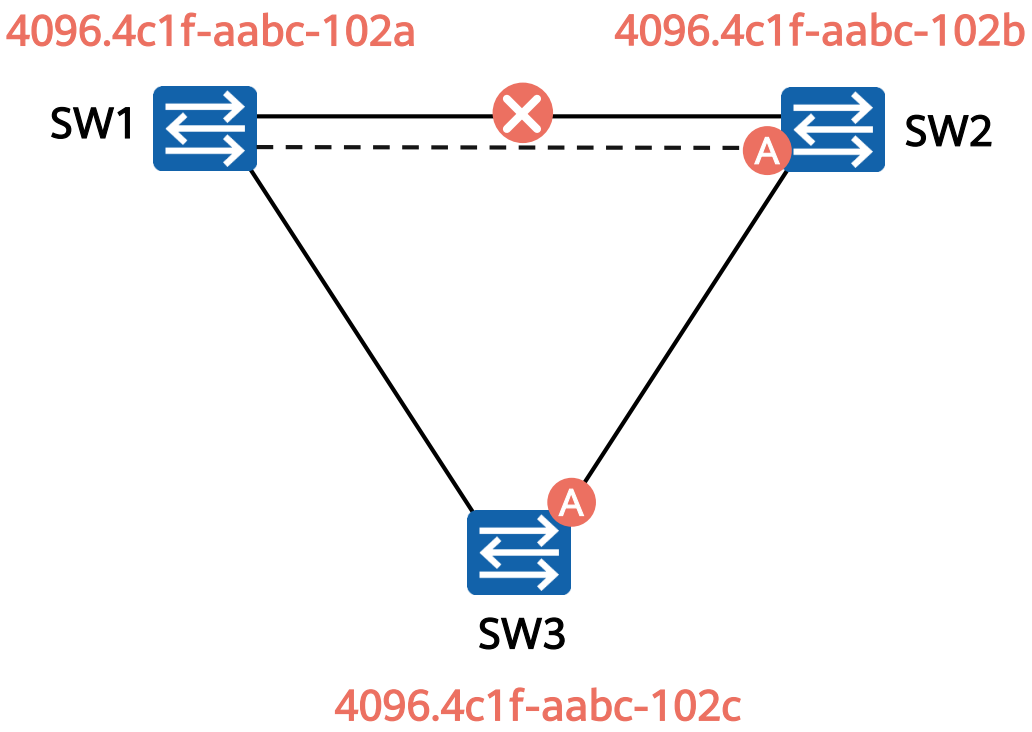


根桥故障恢复过程

1. SW1根桥发生故障，停止发送BPDU报文。
2. SW2等待Max Age计时器（20 s）超时，从而导致已经收到的BPDU报文失效，又接收不到根桥发送的新的BPDU报文，从而得知上游出现故障。
3. 非根桥会互相发送配置BPDU，重新选举新的根桥。
4. 经过重新选举后，SW3的A端口经过两个Forward Delay（15 s）时间恢复转发状态。

- 非根桥会在BPDU老化之后开始根桥的重新选举。
- 根桥故障会导致50 s左右的恢复时间。

拓扑变化 – 直连链路故障



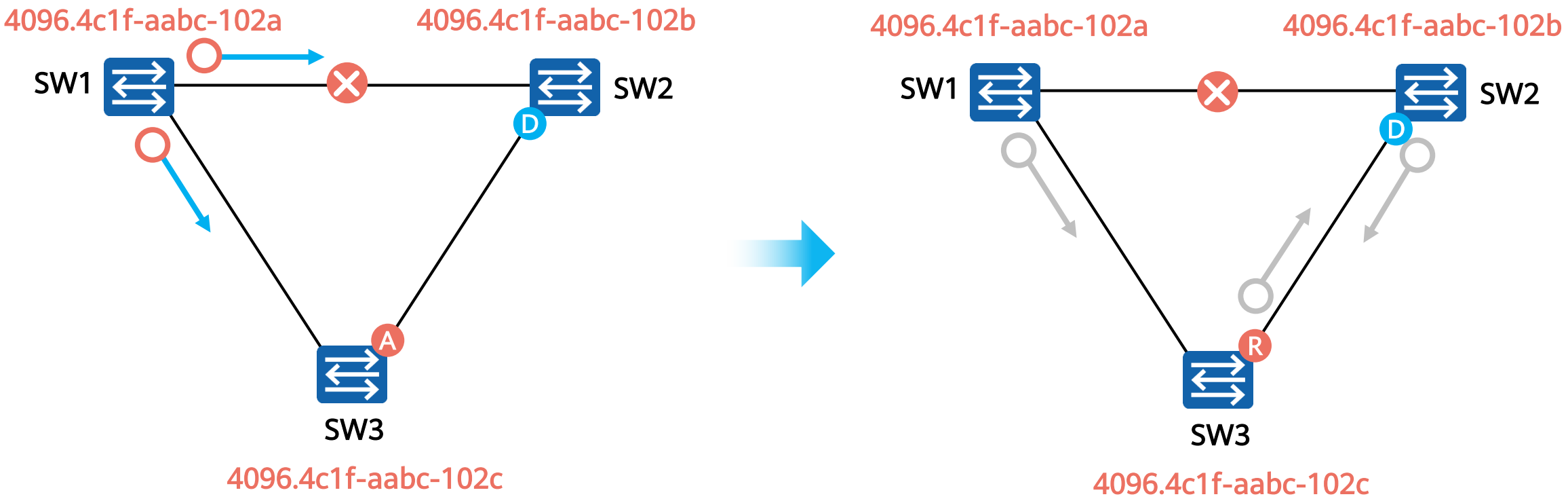
直连链路故障恢复过程

当交换机SW2网络稳定时检测到根端口的链路发生故障，则其备用端口会经过两倍的Forward Delay（15s）时间进入用户流量转发状态。

- SW2检测到直连链路物理故障后，会将预备端口转换为根端口。
- 直连链路故障，备用端口会经过30s后恢复转发状态。

拓扑变化 – 非直连链路故障

- 非直连链路故障后，SW3的备用端口恢复到转发状态，非直连故障会导致50s左右的恢复时间。

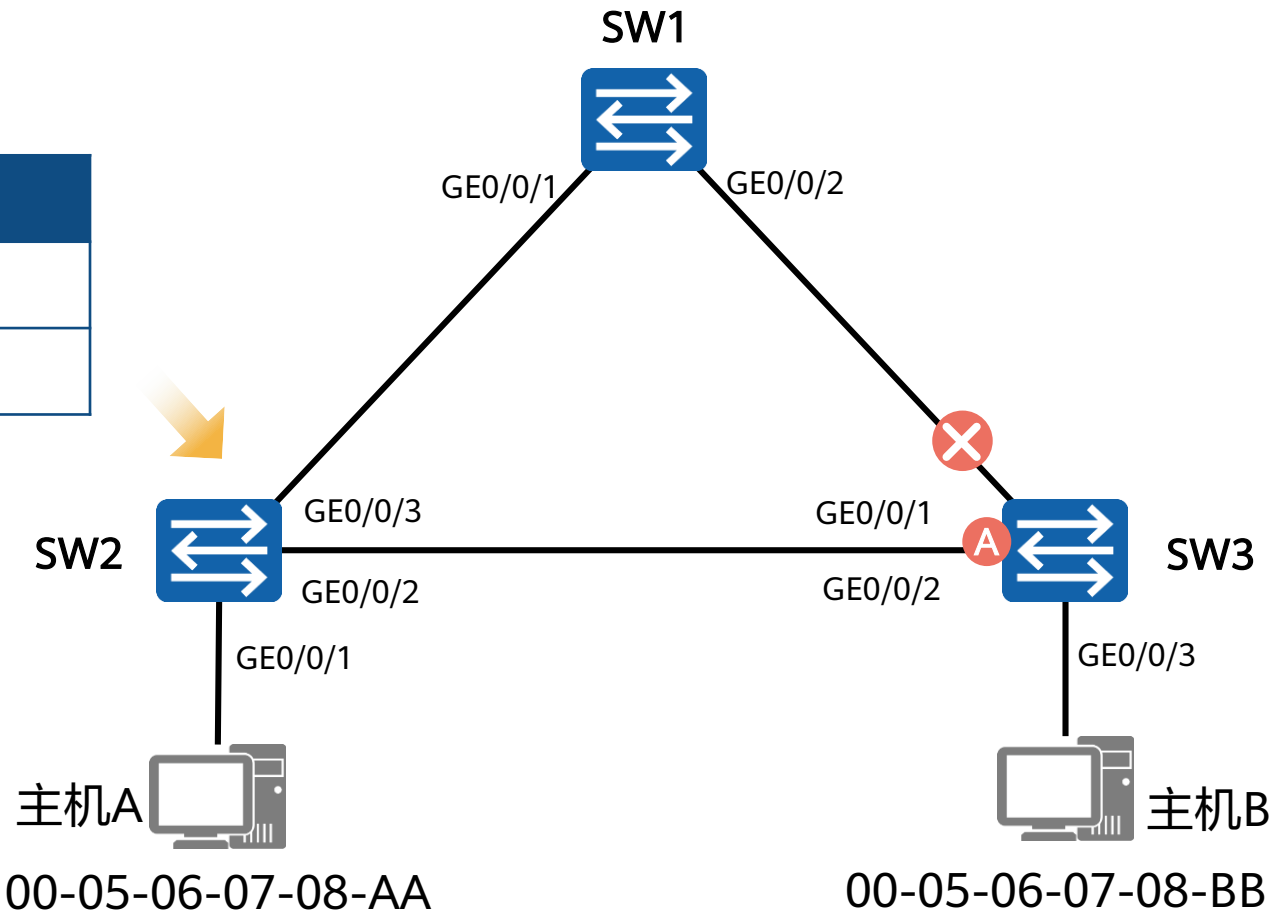


拓扑改变导致MAC地址表错误

MAC地址表

MAC	端口
00-05-06-07-08-AA	GE0/0/1
00-05-06-07-08-BB	GE0/0/3

如图，SW3的根端口发生故障，导致生成树拓扑重新收敛，在生成树拓扑完成收敛之后，从主机A到主机B的帧仍然不能到达目的地。这是因为交换机依赖MAC地址表转发数据帧，缺省情况下，MAC地址表项的老化时间是300秒。那么该怎么快速恢复转发？

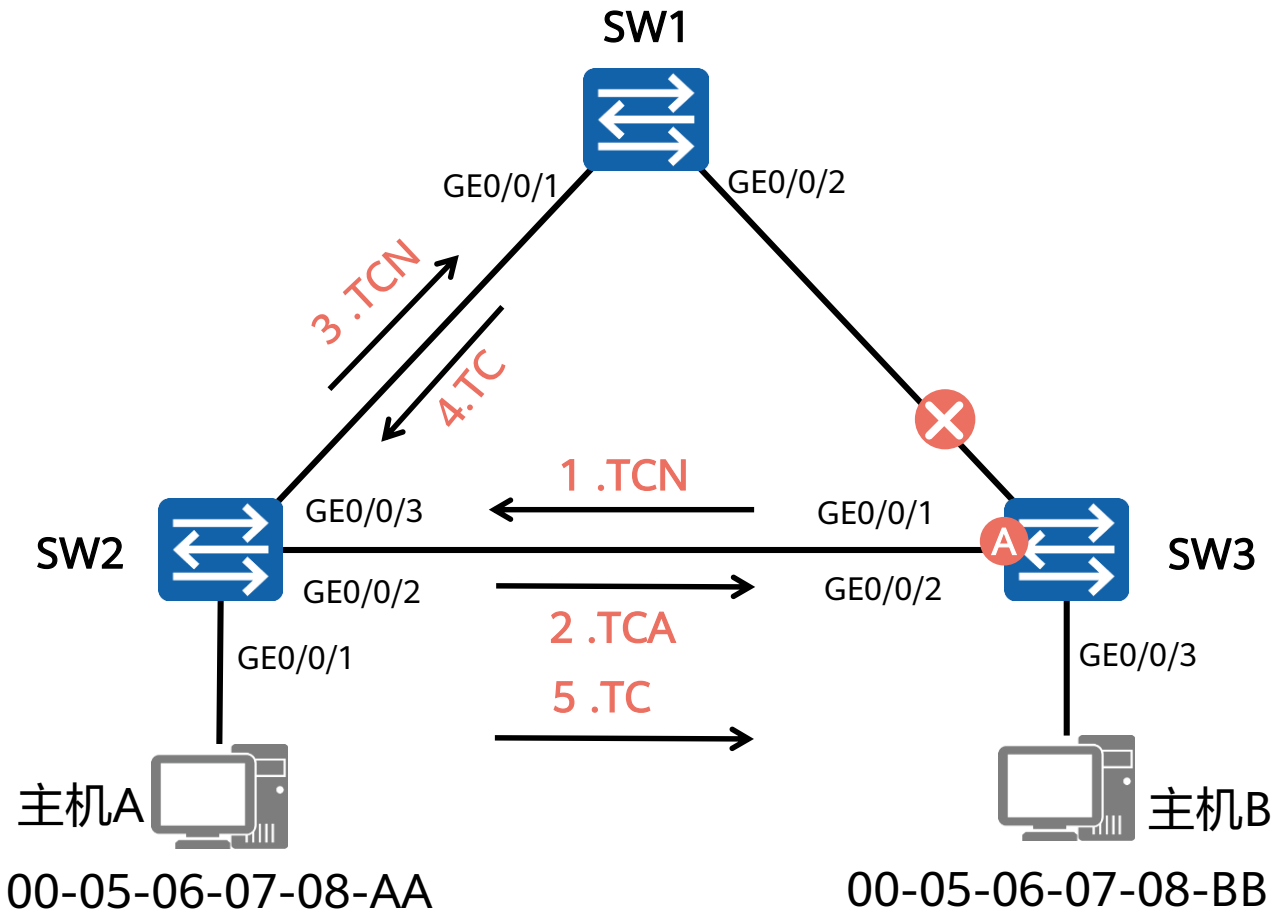


拓扑改变导致MAC地址表错误

MAC地址表

MAC	端口
00-05-06-07-08-AA	GE0/0/3
00-05-06-07-08-BB	GE0/0/1
00-05-06-07-08-BB	GE0/0/2

- TCN BPDU在网络拓扑变化的时候产生。
- 报文格式：协议标识、版本号和类型。
- 拓扑变化：会使用到配置BPDU中Flags的TCA和TC位。



目录

STP的基础配置命令 (1)

1. 配置生成树工作模式

```
[Huawei] stp mode { stp | rstp | mstp }
```

交换机支持STP、RSTP和MSTP（Multiple Spanning Tree Protocol）三种生成树工作模式，默认情况工作在MSTP模式。

2. （可选）配置根桥

```
[Huawei] stp root primary
```

配置当前设备为根桥。缺省情况下，交换机不作为任何生成树的根桥。配置后该设备优先级数值自动为0，并且不能更改设备优先级。

3. （可选）备份根桥

```
[Huawei] stp root secondary
```

配置当前交换机为备份根桥。缺省情况下，交换机不作为任何生成树的备份根桥。配置后该设备优先级数值为4096，并且不能更改设备优先级。

STP的基础配置命令 (2)

1. （可选）配置交换机的STP优先级

```
[Huawei] stp priority priority
```

缺省情况下，交换机的优先级取值是32768。

2. （可选）配置接口路径开销

```
[Huawei] stp pathcost-standard { dot1d-1998 | dot1t | legacy }
```

配置接口路径开销计算方法。缺省情况下，路径开销值的计算方法为IEEE 802.1t（dot1t）标准方法。
同一网络内所有交换机的接口路径开销应使用相同的计算方法。

```
[Huawei-GigabitEthernet0/0/1] stp cost cost
```

设置当前接口的路径开销值。

STP的基础配置命令 (3)

1. （可选）配置接口优先级

```
[Huawei-intf] stp priority priority
```

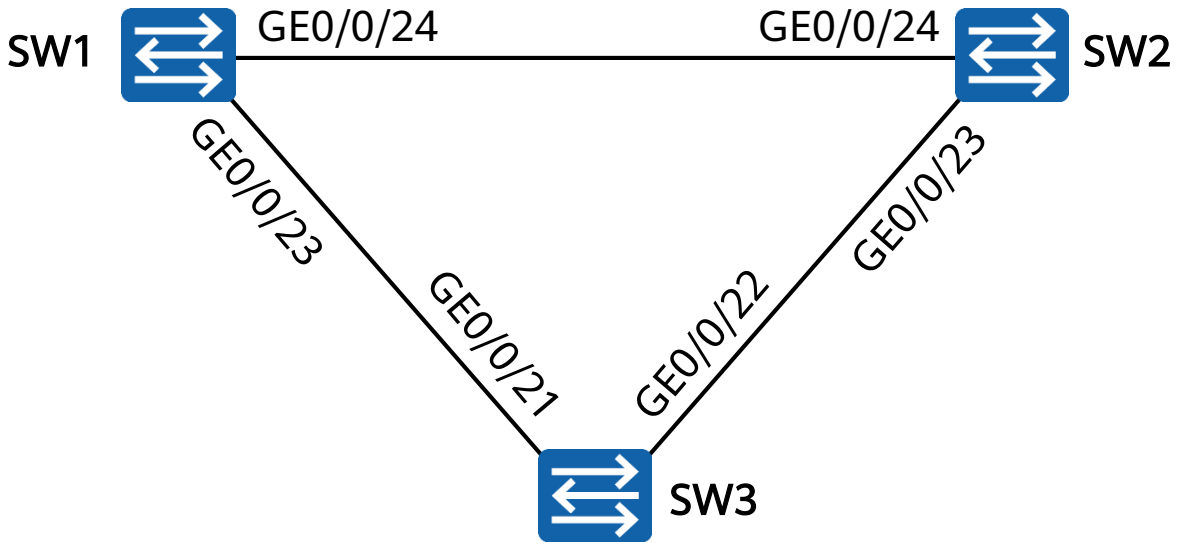
配置接口的优先级。缺省情况下，交换机接口的优先级取值是128。

2. 启用STP/RSTP/MSTP

```
[Huawei] stp enable
```

使能交换机的STP/RSTP/MSTP功能。缺省情况下，设备的STP/RSTP/MSTP功能处于启用状态。

案例1：STP的基础配置



- 在上述三台交换机上部署STP，以便消除网络中的二层环路。
- 通过配置，将SW1指定为根桥，并使SW3的GE0/0/22接口被STP阻塞。

SW1的配置如下：

```
[SW1] stp mode stp
[SW1] stp enable
[SW1] stp priority 0
```

SW2的配置如下：

```
[SW2] stp mode stp
[SW2] stp enable
[SW2] stp priority 4096
```

SW3的配置如下：

```
[SW3] stp mode stp
[SW3] stp enable
```

案例1：STP的基础配置

在SW3上查看STP接口状态摘要：

```
<SW3> display stp brief
 MSTID     Port                        Role      STP State     Protection
   0       GigabitEthernet0/0/21       ROOT      FORWARDING    NONE
   0       GigabitEthernet0/0/22       ALTE      DISCARDING    NONE
```

目录

STP的不足之处

- STP协议虽然能够解决环路问题，但是由于网络拓扑收敛慢，影响了用户通信质量。
- STP没有细致区分接口状态和接口角色，不利于初学者学习及部署。
- 网络协议的优劣往往取决于协议是否对各种情况加以细致区分。
  - 从用户角度来讲，Listening、Learning和Blocking状态并没有区别，都同样不转发用户流量。
  - 从使用和配置角度来讲，接口之间最本质的区别并不在于接口状态，而是在于接口扮演的角色。
  - 根接口和指定接口可以都处于Listening状态，也可能都处于Forwarding状态。
- STP算法是被动的算法，依赖定时器等待的方式判断拓扑变化，收敛速度慢。
- STP算法要求在稳定的拓扑中，根桥主动发出配置BPDU报文，而其他设备进行处理，传遍整个STP网络。这也是导致拓扑收敛慢的主要原因之一。

RSTP概述

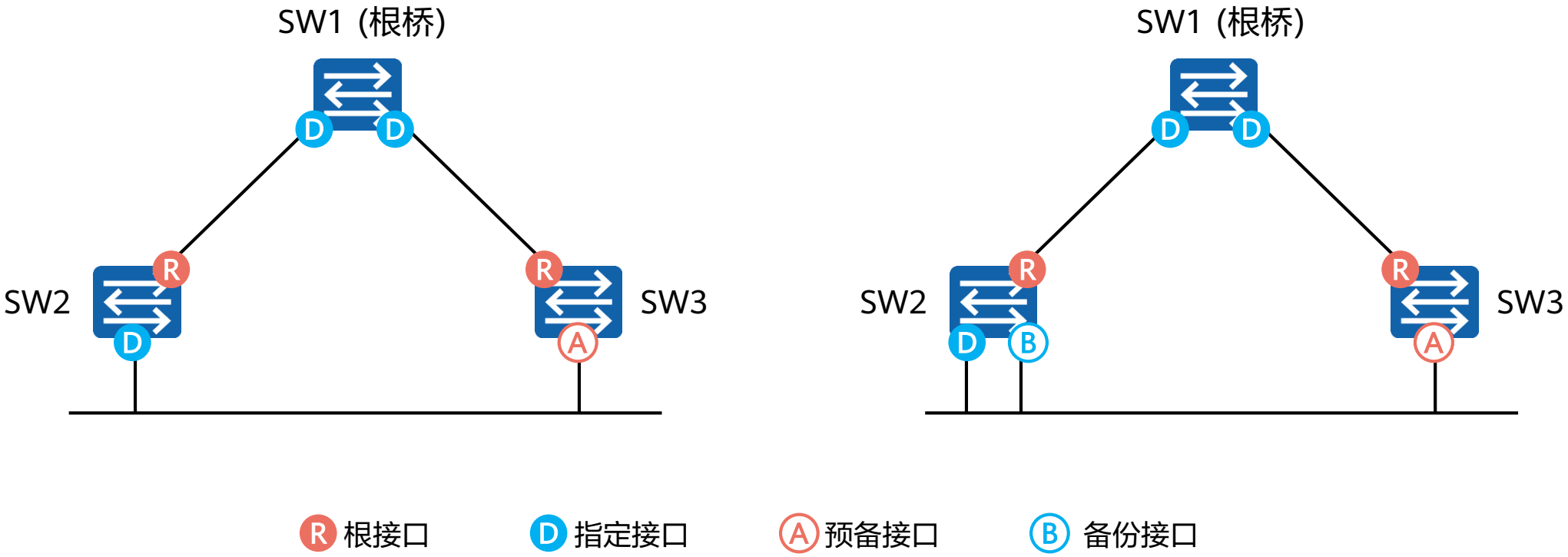
- IEEE 802.1w中定义的RSTP可以视为STP的改进版本，RSTP在许多方面对STP进行了优化，它的收敛速度更快，而且能够兼容STP。
- RSTP引入了新的接口角色---备份端口，边缘端口。
- RSTP的状态规范---3种状态。

RSTP对STP的其他改进

- 配置BPDU的处理发生变化：
  - 拓扑稳定后，配置BPDU报文的发送方式进行了优化
  - 使用更短的BPDU超时计时
  - 对处理次等BPDU的方式进行了优化
- 配置BPDU格式的改变，充分利用了STP协议报文中的Flag字段，明确了接口角色

端口角色不同

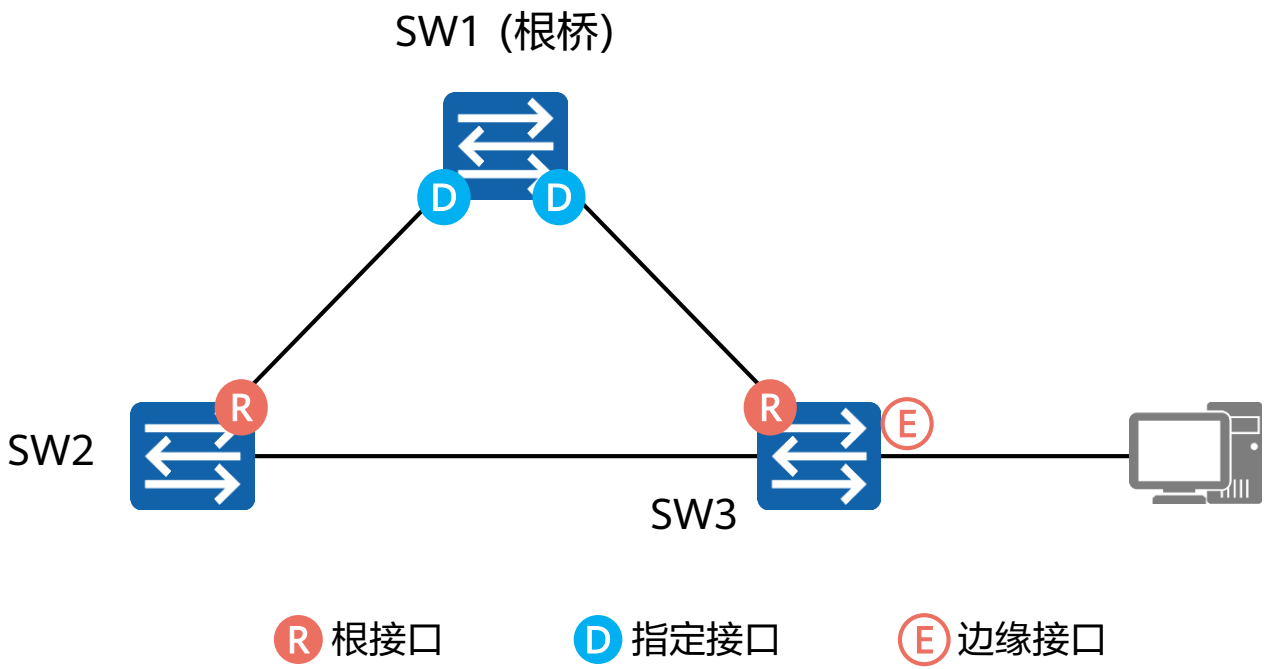
- 通过接口角色的增补，简化了生成树协议的理解及部署



RSTP的接口角色共有4种：根接口、指定接口、预备接口和备份接口

边缘端口

- 如果指定端口位于整个域的边缘，不再与任何交换设备连接，这种端口叫做边缘端口。



边缘端口一般与用户终端设备直接连接，可以由Disabled状态直接转到Forwarding状态。

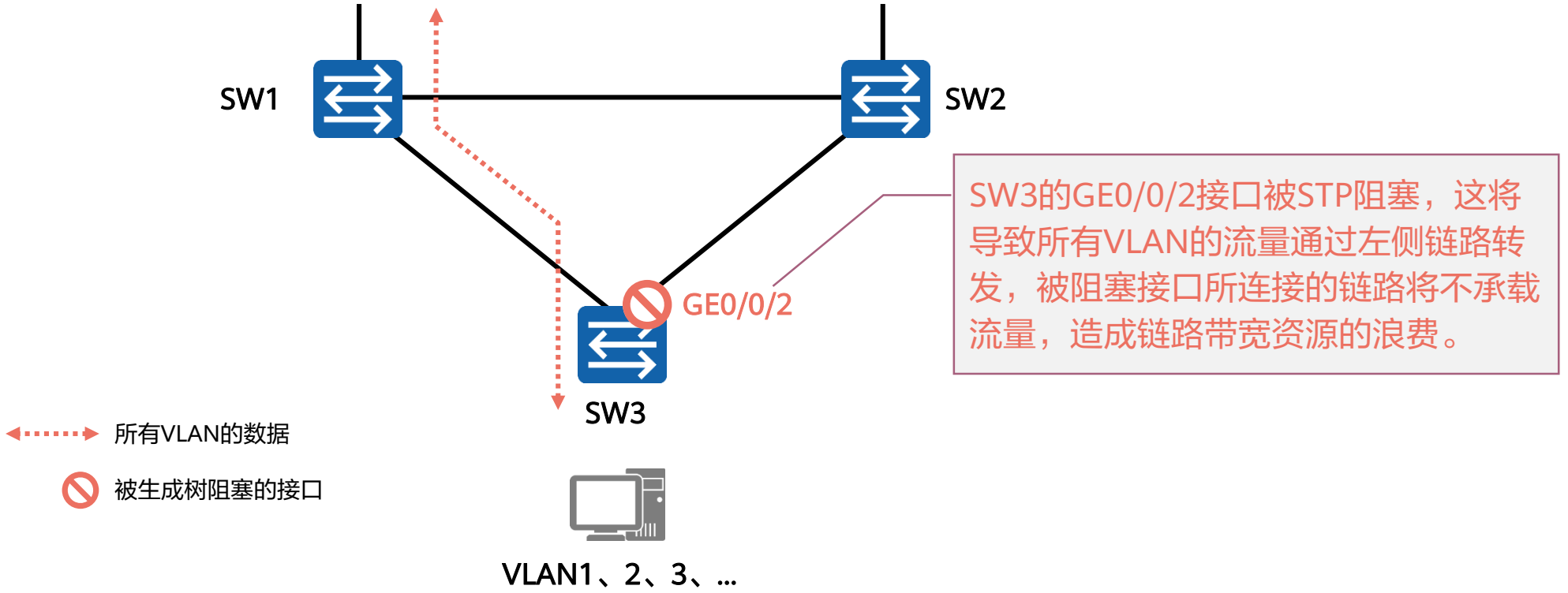
端口状态不同

- RSTP的状态规范把原来的5种状态缩减为3种。

STP接口状态	RSTP接口状态	接口在拓扑中的角色
Forwarding	Forwarding	包括根接口、指定接口
Learning	Learning	包括根接口、指定接口
Listening	Discarding	包括根接口、指定接口
Blocking	Discarding	包括Alternate接口、Backup接口
Disabled	Discarding	包括Disable接口

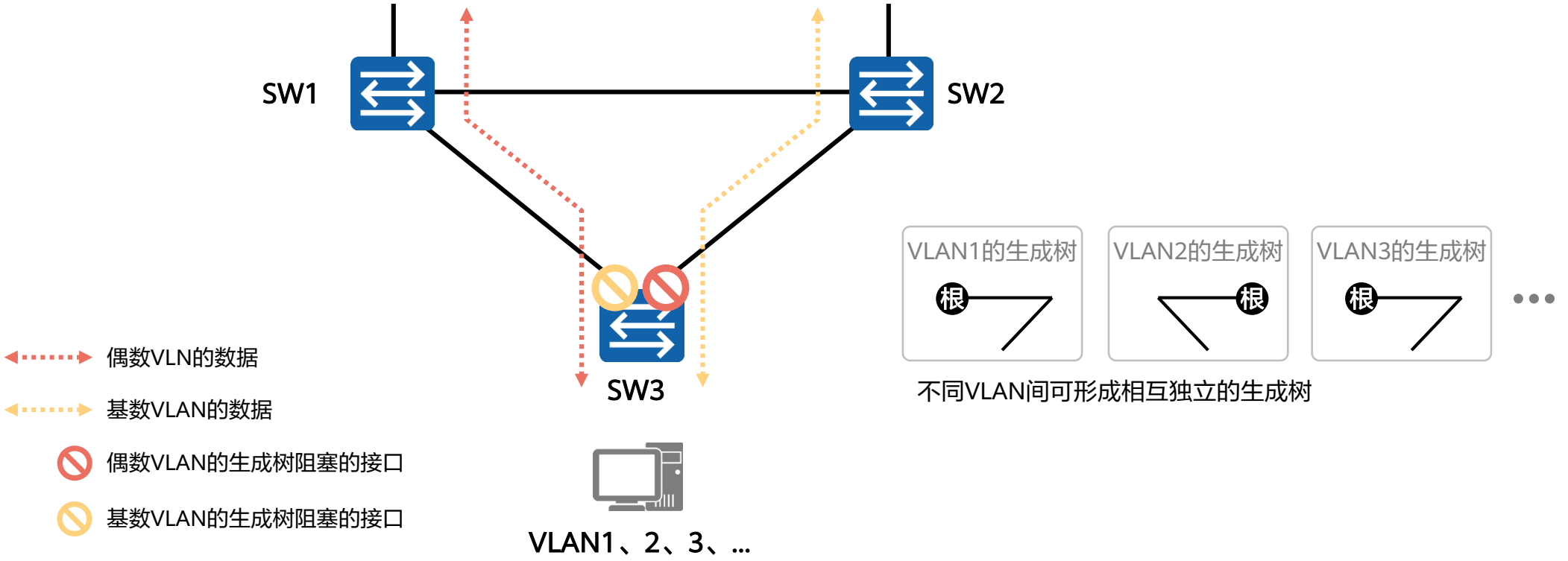
目录

STP/RSTP的缺陷：所有的VLAN共享一棵生成树

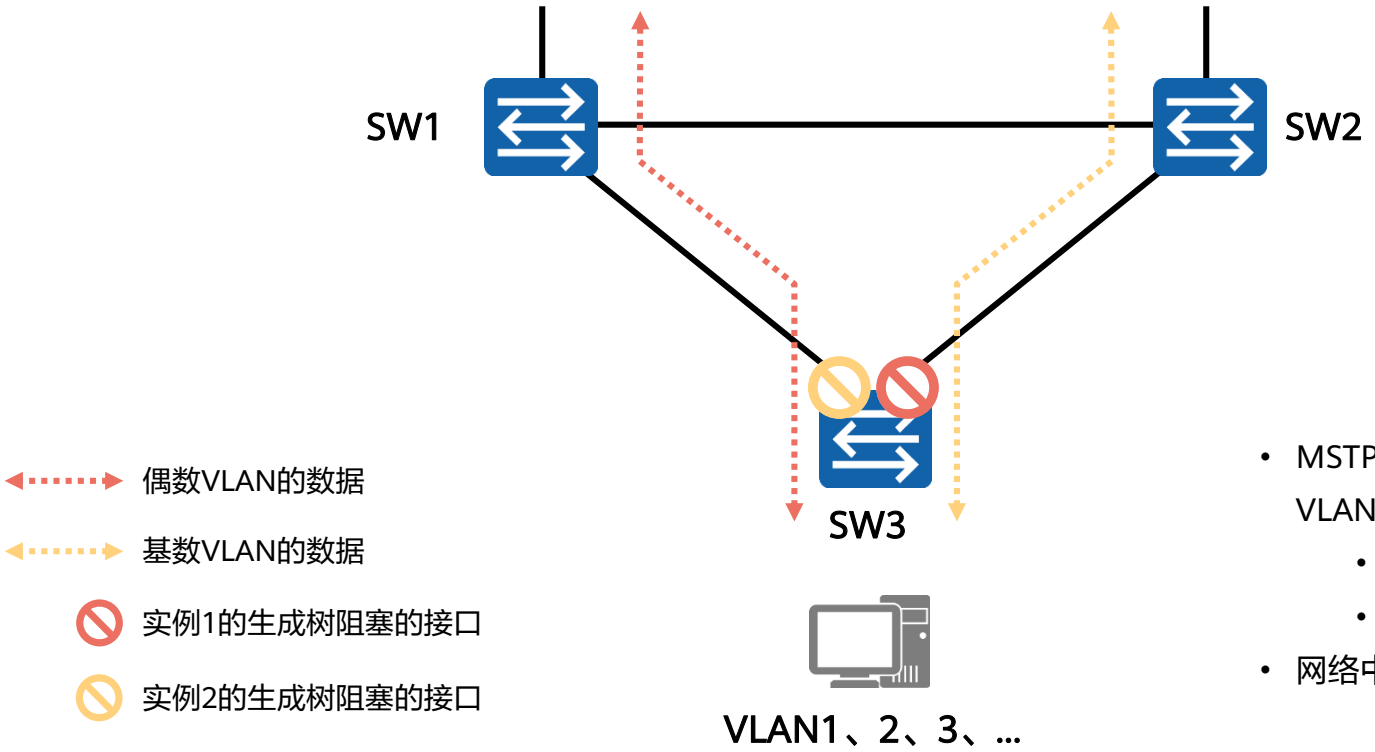


SW3的GE0/0/2接口被STP阻塞，这将导致所有VLAN的流量通过左侧链路转发，被阻塞接口所连接的链路将不承载流量，造成链路带宽资源的浪费。

VBST：基于VLAN的生成树



MSTP：多生成树



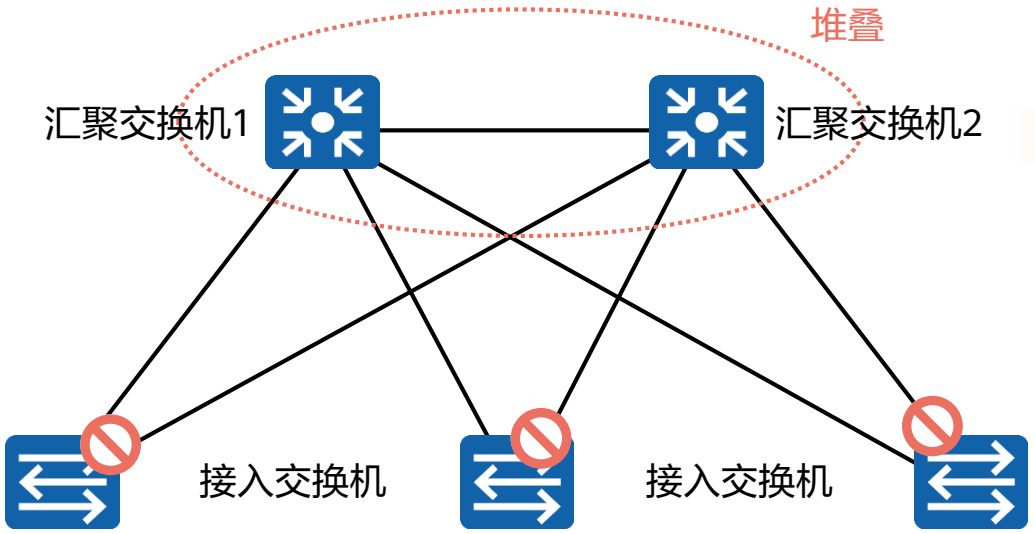
- MSTP将VLAN映射到一个生成树的实例，若干个VLAN可共用一棵生成树。例如：
  - 将偶数VLAN映射到实例1
  - 将基数VLAN映射到实例2
- 网络中将只维护2棵生成树。

MSTP概述

- MSTP把一个交换网络划分成多个域，每个域内形成多棵生成树，生成树之间彼此独立。
- 每棵生成树叫做一个多生成树实例MSTI（Multiple Spanning Tree Instance）。
- 所谓生成树实例就是多个VLAN的集合所对应的生成树。
- 通过将多个VLAN捆绑到一个实例，可以节省通信开销和资源占用率。
- MSTP各个实例拓扑的计算相互独立，在这些实例上可以实现负载均衡。
- 可以把多个相同拓扑结构的VLAN映射到一个实例里，这些VLAN在接口上的转发状态取决于接口在对应实例的状态。

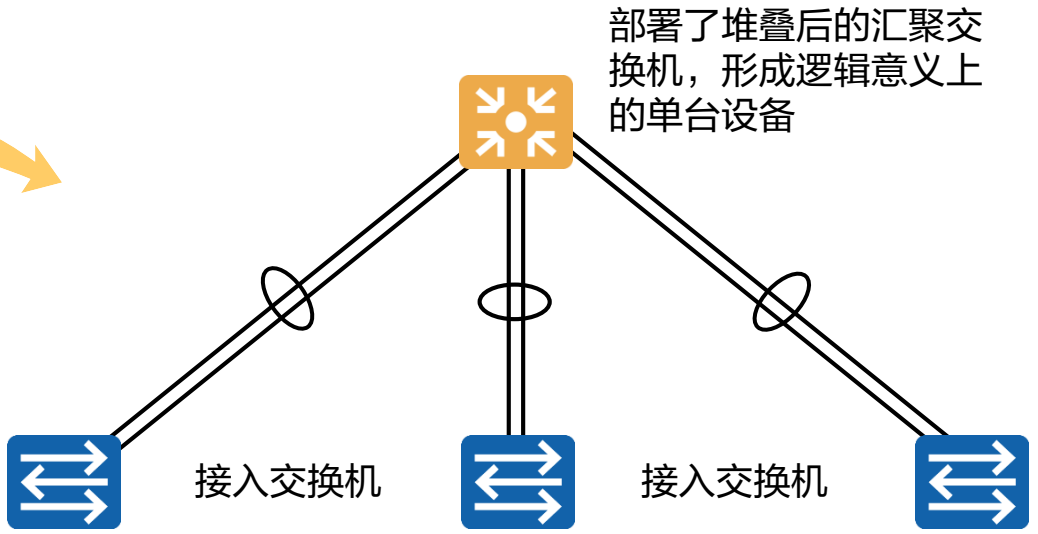
堆叠与园区网络树形结构组网形态

传统STP组网



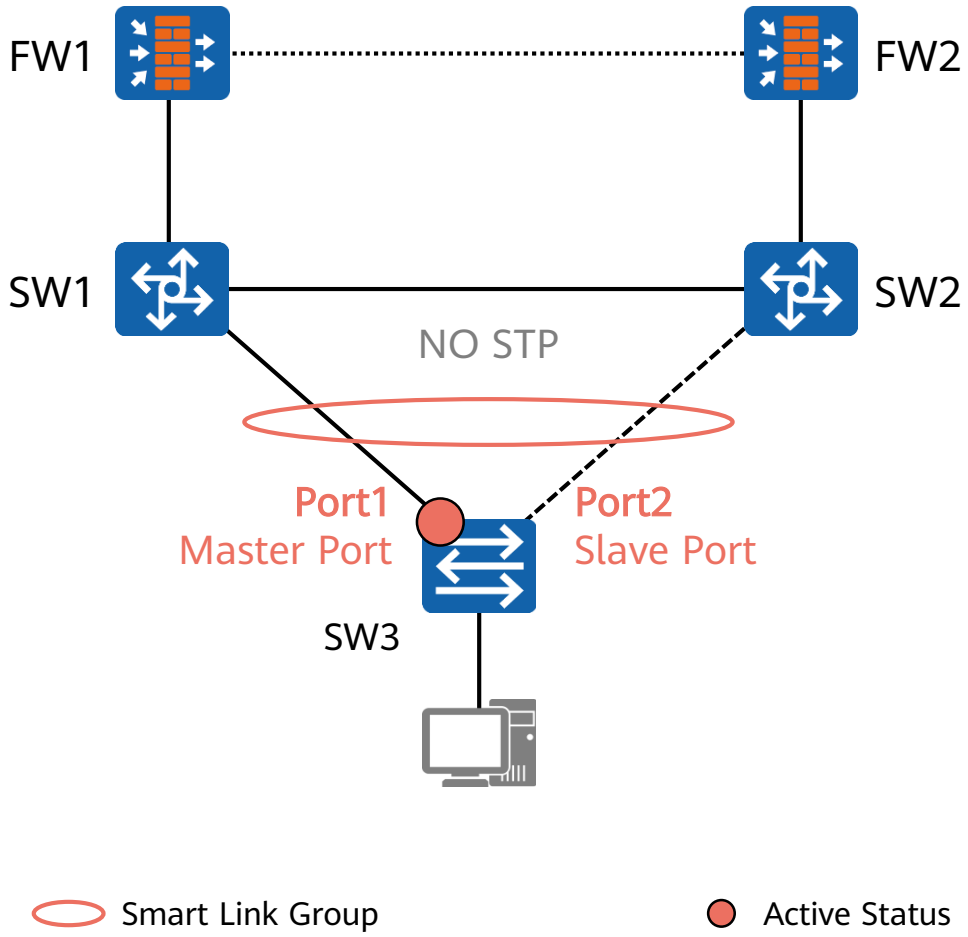
STP将阻塞网络中的接口，造成链路带宽无法充分利用。

交换机堆叠组网



将汇聚交换机部署堆叠，形成逻辑意义上的单台设备，部署链路聚合，可将拓扑进一步简化为“树形结构”，消除二层环路，同时充分提高链路带宽利用率

Smart Link



- Smart Link是一种为双上行组网量身定做的解决方案：
  - 在双向行的设备上部署，当网络正常时，两条上行链路中，一条处于活跃状态，而另一条则处于备份状态（不承载业务流量）。如此一来二层环路就此打破。
  - 当主用链路发生故障后，流量会在毫秒级的时间内迅速切换到备用链路上，保证了数据的正常转发。
  - Smart Link配置简单，便于用户操作。
  - 无需协议报文交互，收敛速度及可靠性大大提升。

本章总结

- 生成树是一个用于局域网中消除环路的协议。
- 消除环路
- 链路备份