锐捷认证-RCNA

尚文知礼，科技竞才，技术改变命运　SKILL CHANGES YOUR DESTINY

RCNA-RCNP-RCIE

尚文网络

尚文网络
www.upwen.com

09

生成树技术原理与应用

- 生成树协议综述
- STP技术原理和配置
- STP技术高级特性

前言

- 本课程主要介绍了生成树协议的产生背景、工作原理、扩展技术特性，以及在实际网络中如何来使用生成树协议。

目标

- **通过本课程的学习，您将能够：**
  - 描述生成树协议的工作原理
  - 掌握生成树协议的特性
  - 掌握生成树协议的高级特性

尚文网络
www.upwen.com

09

生成树技术原理与应用

- 生成树协议综述
- STP技术原理和配置
- STP技术高级特性

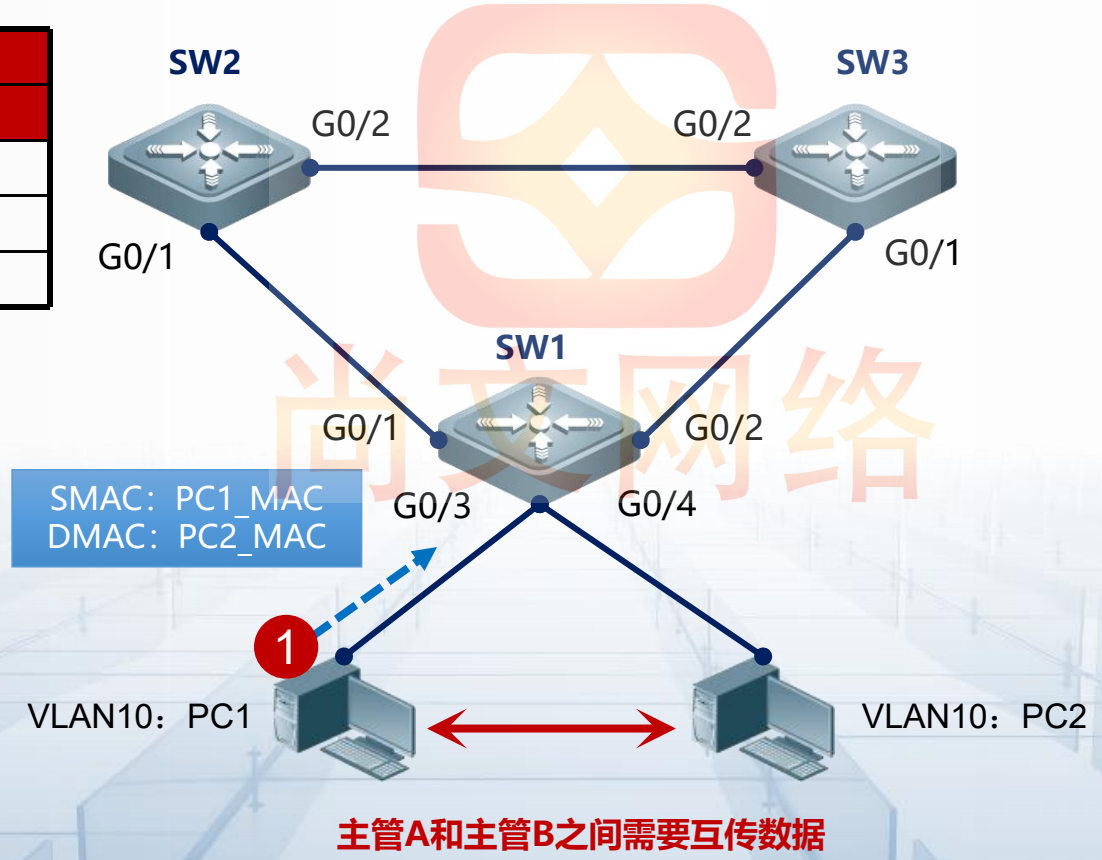
环路的现象与危害

- 环路的现象
  - 交换机端口指示灯以相同频率快速闪烁
  - 交换机MAC地址表震荡
  - 交换机因为资源耗尽，登陆操作异常
- 环路的危害
  - 链路堵塞：广播报文在二层网络中不断泛洪，所有链路都被大量的广播报文充斥
  - 主机系统响应迟缓：主机网卡接收到大量广播报文,操作系统调用大量CPU进程资源来识别这些广播报文
  - 二层交换机管理缓慢：大量广播报文需要CPU处理，浪费CPU大量资源，对正常的请求无法响应
  - 冲击网关设备的CPU：对网关IP地址的ARP请求报文，经过环路的复制转发，不断地发送到网关设备，网关设备的CPU压力不断增大，甚至崩溃

二层冗余网络面临的问题

- 在一个VLAN内，广播包向接收端口之外的所有端口洪泛
- 交换机基于工作原理进行：学习记录、查表转发

SW2　MAC地址表		
VLAN	MAC地址	端口



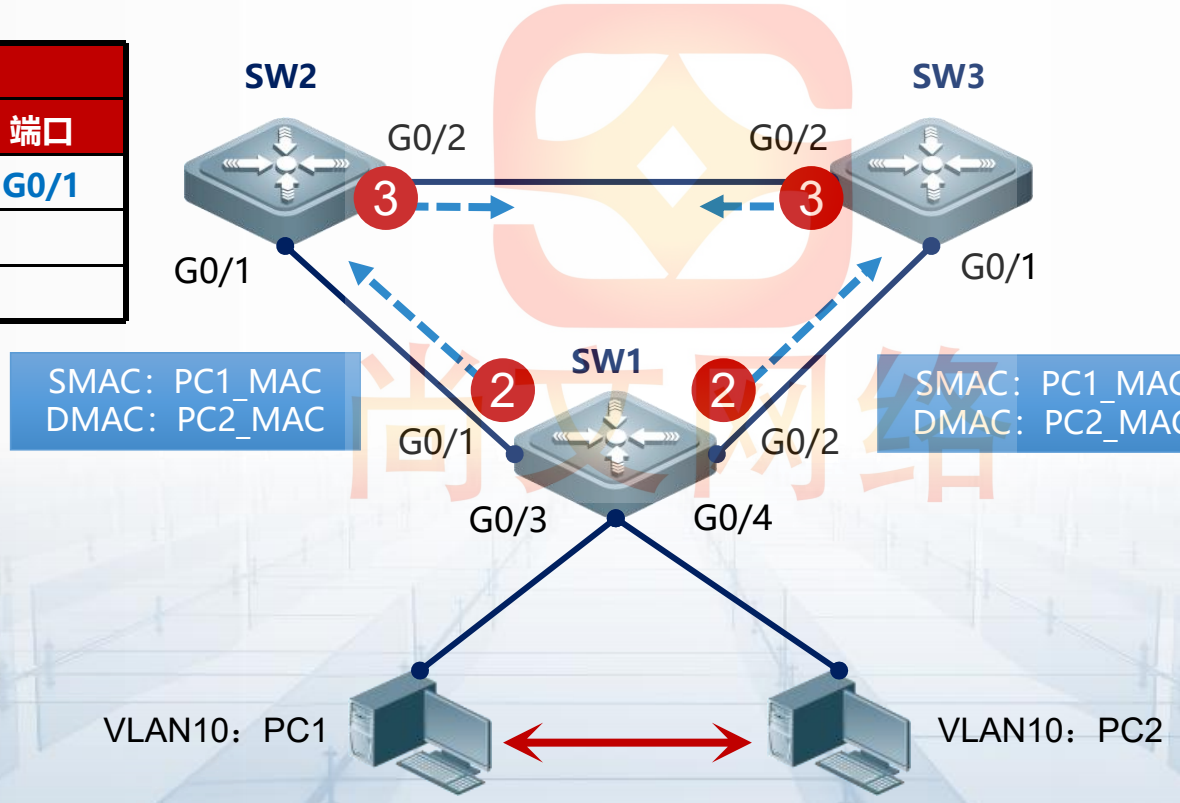
SW3　MAC地址表		
VLAN	MAC地址	端口

SW1　MAC地址表		
VLAN	MAC地址	端口
10	PC1_MAC	G0/3

二层冗余网络面临的问题

- 1、SW1学习PC1_MAC，查MAC地址表，无匹配目的MAC的表项，进行泛洪转发
- 2、SW2与SW3接收到数据帧，进行学习源MAC，并进行泛洪转发

SW2　MAC地址表		
VLAN	MAC地址	端口
10	PC1_MAC	G0/1



主管A和主管B之间需要互传数据

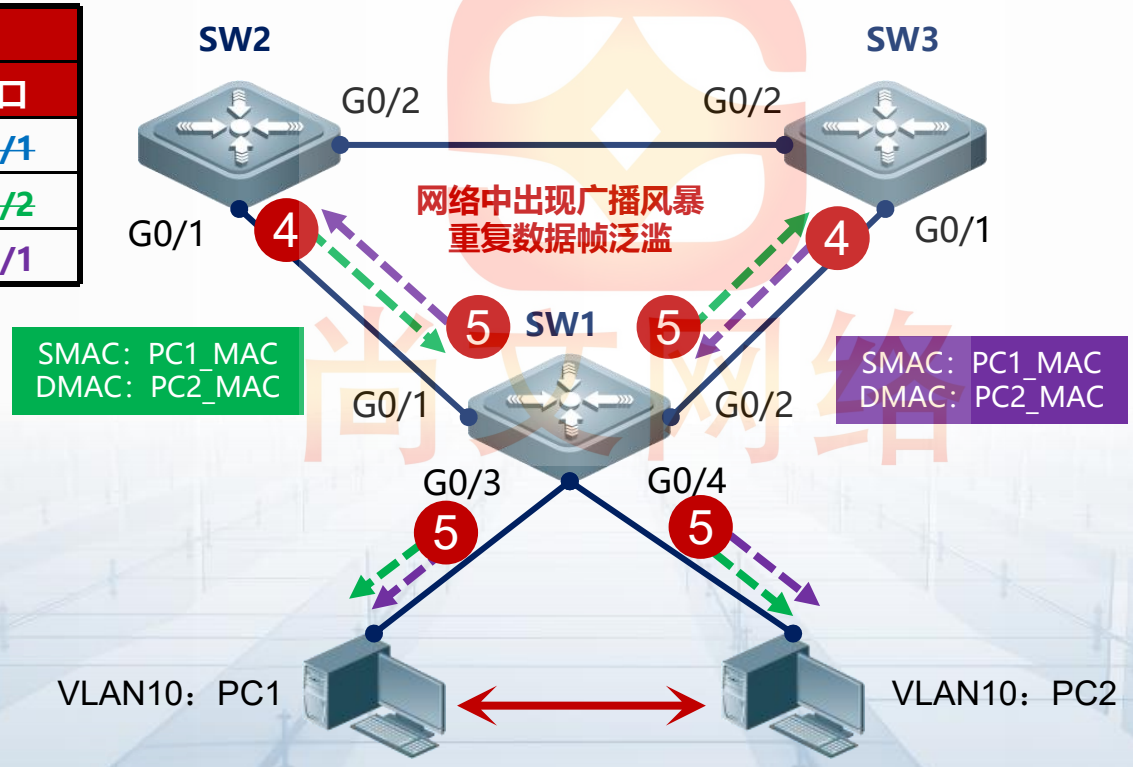
SW3　MAC地址表		
VLAN	MAC地址	端口
10	PC1_MAC	G0/1

SW1　MAC地址表		
VLAN	MAC地址	端口
10	PC1_MAC	G0/3

二层冗余网络面临的问题

- 3、SW2与SW3互相收到对方的相同帧，进行学习源MAC，将会把PC1_MAC重新关联接口
- 4、SW1根据接收到G0/1与G0/2的相同帧，按先后顺序进行学习与泛洪（MAC地址表不稳定）
- 5、SW2与SW3又会收到同样的帧，同理循环……

SW2　MAC地址表		
VLAN	MAC地址	端口
~~10~~	~~PC1_MAC~~	~~G0/1~~
~~10~~	~~PC1_MAC~~	~~G0/2~~
10	PC1_MAC	G0/1

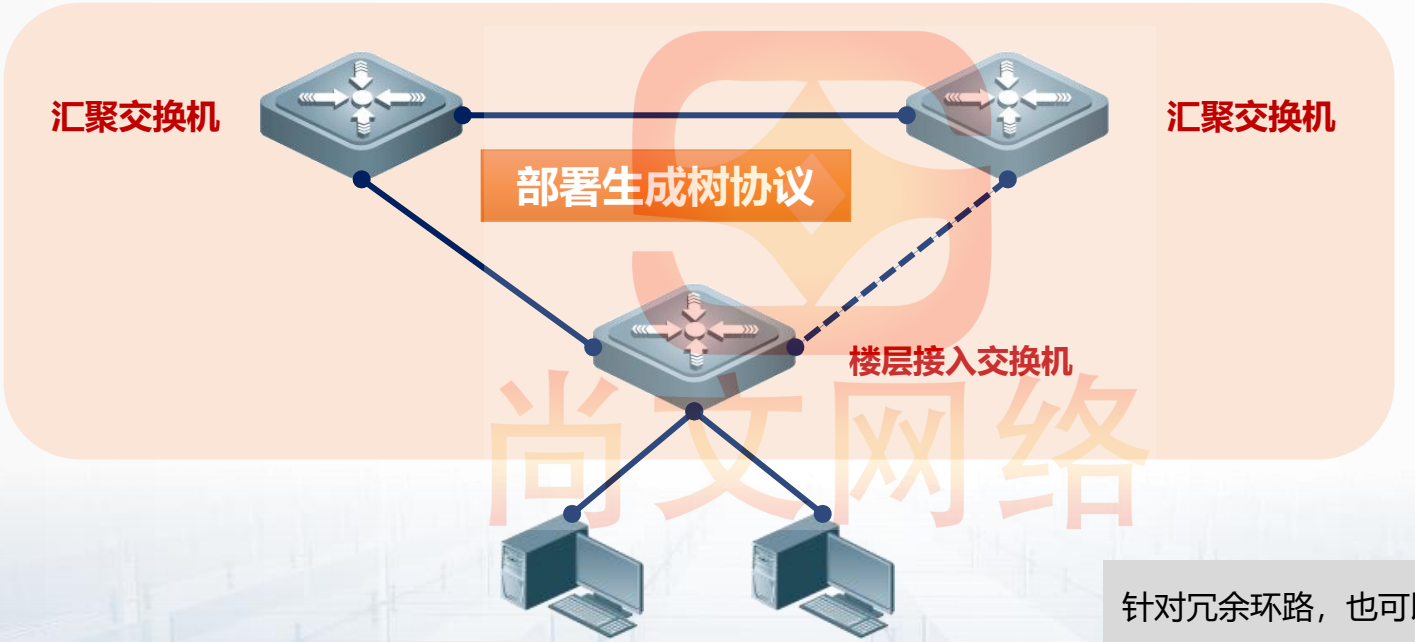


SW3　MAC地址表		
VLAN	MAC地址	端口
~~10~~	~~PC1_MAC~~	~~G0/1~~
~~10~~	~~PC1_MAC~~	~~G0/2~~
10	PC1_MAC	G0/1

SW1　MAC地址表		
VLAN	MAC地址	端口
~~10~~	~~PC1_MAC~~	~~G0/3~~
~~10~~	~~PC1_MAC~~	~~G0/1~~
10	PC1_MAC	G0/2

二层环路解决方案

- 在交换机上部署生成树协议，逻辑阻塞环路接口
- 当发生物理故障时，冗余链路可以恢复正常转发



针对冗余环路，也可以使用链路捆绑、VSU等技术消环
但针对故障环路，STP依然是有效解决方案

STP概述

- STP是怎样的一个协议呢?
  - 通过阻断冗余链路，将一个有环路的交换网络修剪成一个无环路的树型拓扑结构
  - 在某条活动(active)的链路断开时，通过激活被阻断的冗余链路重新修剪拓扑结构以恢复网络的连通。

生成树综述

- 生成树协议的分类
  - 生成树协议的分类，按照产生的时间先后顺序为STP、RSTP、MSTP
- 生成树协议所遵循的IEEE标准
  - 三种生成树所遵循的IEEE标准分别为：STP-IEEE 802.3D，RSTP-IEEE 802.3W，MSTP-IEEE 802.3S

09

生成树技术原理与应用

- 生成树协议综述
- **STP技术原理和配置**
- STP技术高级特性

STP工作原理

- STP技术的实现基于SPA（最短路径树算法 shortest path algorithm）算法，通过SPA在构建无环树形结构的时候经历四个步骤：
  - 选举根桥（Root），根桥是整个无环树形结构的根节点；
  - 选举根端口（RP），根端口是距离根桥最近的端口；
  - 选举指定端口（DP），指定端口是设备发出BPDU的接口，每条链路上必定有且只有一个指定端口；
  - 剩下的端口全部Block掉
- 实现SPA算法所需要的原始数据，则是通过交换机之间交互BPDU（bridge protocol data unit）报文来完成的。

BPDU报文介绍

- BPDU，Bridge Protocol Data Unit网桥协议数据单元。左图为BPDU报文结构简介，右图为真实报文结构展示

	Octet
Protocol Identifier	1-2
Protocol Version Identifier	3
BPDU Type	4
Flags	5
Root Identifier	6-13
Root Path Cost	14-17
Bridge Identifier	18-25
Port Identifier	26-27
Message Age	28-29
Max Age	30-31
Hello Time	32-33
Forward Delay	34-35

```
> Frame 2: 52 bytes on wire (416 bits), 52 bytes captured (416 bits)
> IEEE 802.3 Ethernet
> Logical-Link Control
v Spanning Tree Protocol
    Protocol Identifier: Spanning Tree Protocol (0x0000)
    Protocol Version Identifier: Spanning Tree (0)
    BPDU Type: Configuration (0x00)
  v BPDU flags: 0x00
      0... .... = Topology Change Acknowledgment: No
      .... ...0 = Topology Change: No
  v Root Identifier: 32768 / 0 / 00:d0:f8:8b:5d:d8
      Root Bridge Priority: 32768
      Root Bridge System ID Extension: 0
      Root Bridge System ID: 00:d0:f8:8b:5d:d8
    Root Path Cost: 0
  v Bridge Identifier: 32768 / 0 / 00:d0:f8:8b:5d:d8
      Bridge Priority: 32768
      Bridge System ID Extension: 0
      Bridge System ID: 00:d0:f8:8b:5d:d8
    Port identifier: 0x8026
    Message Age: 0
    Max Age: 20
    Hello Time: 2
    Forward Delay: 15
```

BPDU报文介绍

- BPDU，Bridge Protocol Data Unit网桥协议数据单元，报文中的各个字段长度以及内容如下：

Protocol Identifier	Version	Message Type	Flag	Root ID	Root Path Cost
Bridge ID	Port ID	Message Age	Max Age	Hello Time	Forward Delay

  - Protocol Identifier：2字节，总是为0；
  - Version：1字节，0为STP、2为RSTP、3为MSTP；
  - Message Type：1字节，0x00为C-BPDU，负责建立和维护STP拓扑，0x80为TCN-BPDU,传达拓扑变更；
  - Flags：1字节，最低位=TC（Topology Change，拓扑变化）标志，最高位=TCA（Topology Change Acknowledgement，拓扑变化确认）标志；
  - Root ID：8字节，指示当前根桥的RID（即“根ID”），由2字节的桥优先级和6字节MAC地址构成；
  - Root Path Cost：4字节，指示发送该BPDU报文的端口累计到根桥的开销（1G接口cost值为4，10G接口cost值为2）；

BPDU报文介绍

- BPDU，Bridge Protocol Data Unit网桥协议数据单元，报文中的各个字段长度以及内容如下：

Protocol Identifier	Version	Message Type	Flag	Root ID	Root Path Cost
Bridge ID	Port ID	Message Age	Max Age	Hello Time	Forward Delay

  - Bridge ID：8字节，指示该BPDU报文发送者的BID，是由2字节的桥优先级和6字节MAC地址构成；
  - Port ID：2字节，第1个字节为该端口优先级，默认128，第2个字节为发送的端口号。事实上端口优先级的后4bit和端口号的8bit，共同构成了端口号，由系统分配且不可修改；
  - Message Age：2字节，指示该BPDU报文的生存时间，即端口保存BPDU的最长时间；
  - Max Age：2字节，指示BPDU消息的最大生存时间，也即老化时间，默认20秒；
  - Hello Time：2字节，指示发送两个相邻BPDU的时间间隔，设备通过不断发送BPDU维持自己的地位，Hello time 是发送的间隔时间，默认2秒；
  - Forward Delay：2字节，指示控制listening和learning状态的持续时间，表示在拓扑结构改变后，交换机在发送数据包前维持在监听和学习状态的时间，默认15秒。

STP的收敛过程-选举根桥

选举根桥 | 选举RP | 选举DP | 其他端口Block

- 在一个交换广播域中，选举一个交换机为根桥，选举过程：
  - 比较自己收到BPDU和自己发出的BPDU的Bridge ID，Bridge ID更小的成为根桥：
    - Bridge ID的比较，先比较优先级，优先级更小的成为根桥；
    - 若优先级一致，则比较交换机的MAC地址（show sysmac显示的地址），MAC地址更小的成为根桥
- 交换机启动生成树协议，并配置根桥优先级命令：

```
Ruijie(config)#spanning-tree mode stp
Ruijie(config)#spanning-tree
Ruijie(config)#spanning-tree priority ?
  <0-61440>  Bridge priority in increments of 4096 (default value: 32768)
```

- 为什么根桥优先级只能是4096的整数倍?
  - Bridge ID：8字节，指示该BPDU报文发送者的BID，是由2字节的桥优先级和6字节MAC地址构成，而在2字节（16bit）的优先级字段中，后12bit用来标识VLAN ID，只有前4bit是通过配置修改的。2^{12}=4096。

STP的收敛过程-选举RP

选举根桥 → 选举RP → 选举DP → 其他端口Block

- 在广播域内所有的非根桥交换机选择根端口（RP，ROOT Port），RP是非根桥交换机上到根桥距离最近的端口，每个非根桥有且只有一个RP。非根桥从多个接口接收到BPDU，哪个接口是RP？选举过程：
  - 比较接收到的BPDU中的Cost值，最小Cost值的BPDU的接收端口成为RP；
  - 若Cost值一致，则比较多个BPDU中的Bridge ID，最小Bridge ID值的BPDU的接收端口成为RP。Bridge ID的比较，先比较优先级，再比较MAC，越小越优先；
  - 若Bridge ID一致，则比较多个BPDU中的Port ID，最小Port ID值的BPDU的接收端口成为RP。Port ID的比较，先比较端口优先级，后比较端口号，越小越优先。
- 修改端口cost值和端口优先级的命令（为什么端口优先级只能是16的整数倍？）：

```
Ruijie(config-if-GigabitEthernet 0/36)#spanning-tree cost ?
  <1-200000000>  Port path cost
Ruijie(config-if-GigabitEthernet 0/36)#spanning-tree port-priority ?
 <0-240>  Port priority in increments of 16 (default value: 128)
```

STP的收敛过程-选举DP

选举根桥 → 选举RP → 选举DP → **其他端口Block**

- 在所有的链路上选举指定端口（DP，Design Port），通过比较同一段链路上两个端口发送的BPDU中的字段来实现，选举过程：
  - 比较同一段链路上两个端口发送的BPDU中的Cost值，Cost值较小的BPDU的发送端口成为DP；
  - 若Cost值一致，则比较同一段链路上两个端口发送的BPDU中的的Bridge ID，Bridge ID值较小的BPDU发送端口成为DP；
  - 若Bridge ID一致，则比较同一段链路上两个端口发送的BPDU中的Port ID，Port ID值较小的BPDU发送端口成为DP。

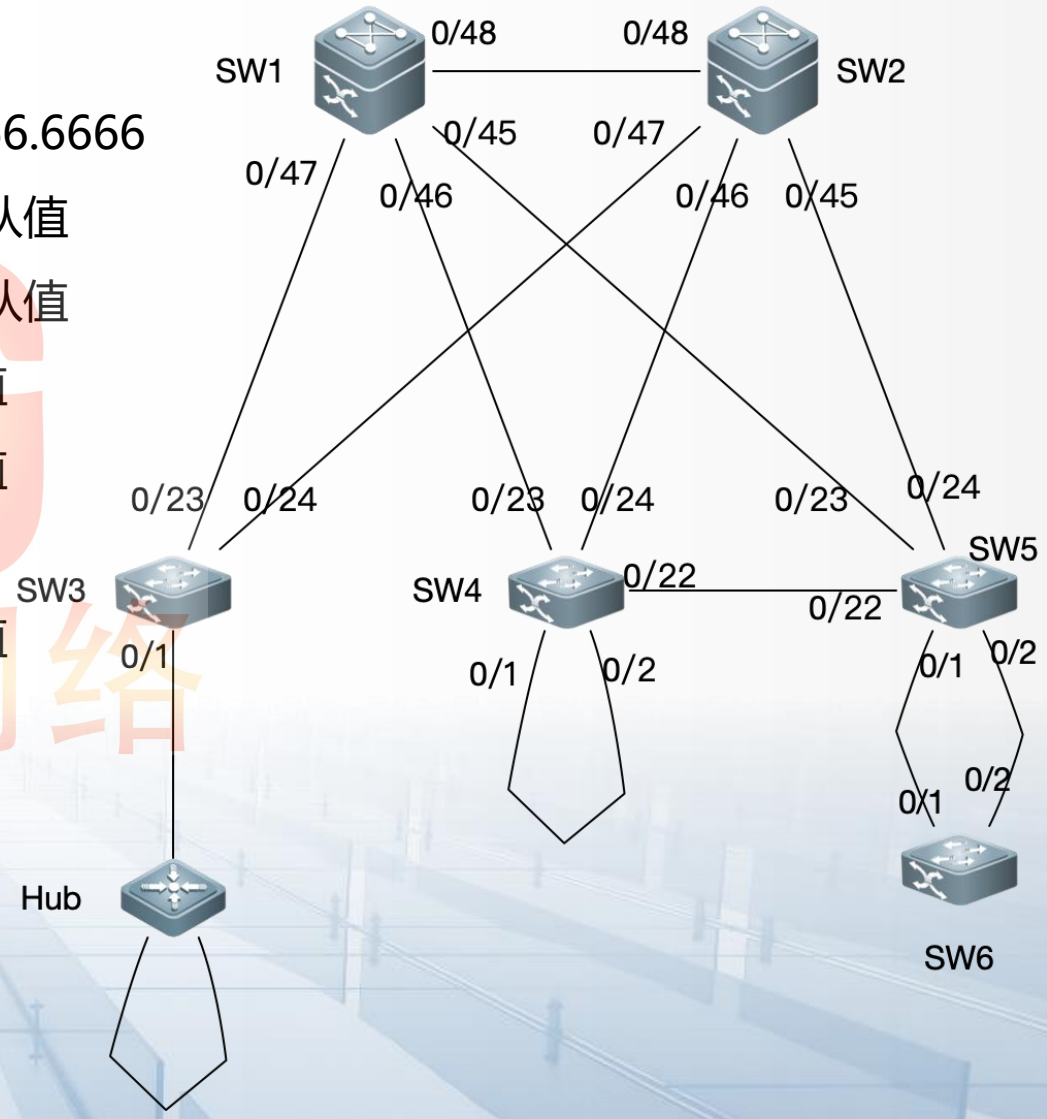
STP的收敛过程-剩余端口Block

选举根桥 → 选举RP → 选举DP → 其他端口Block

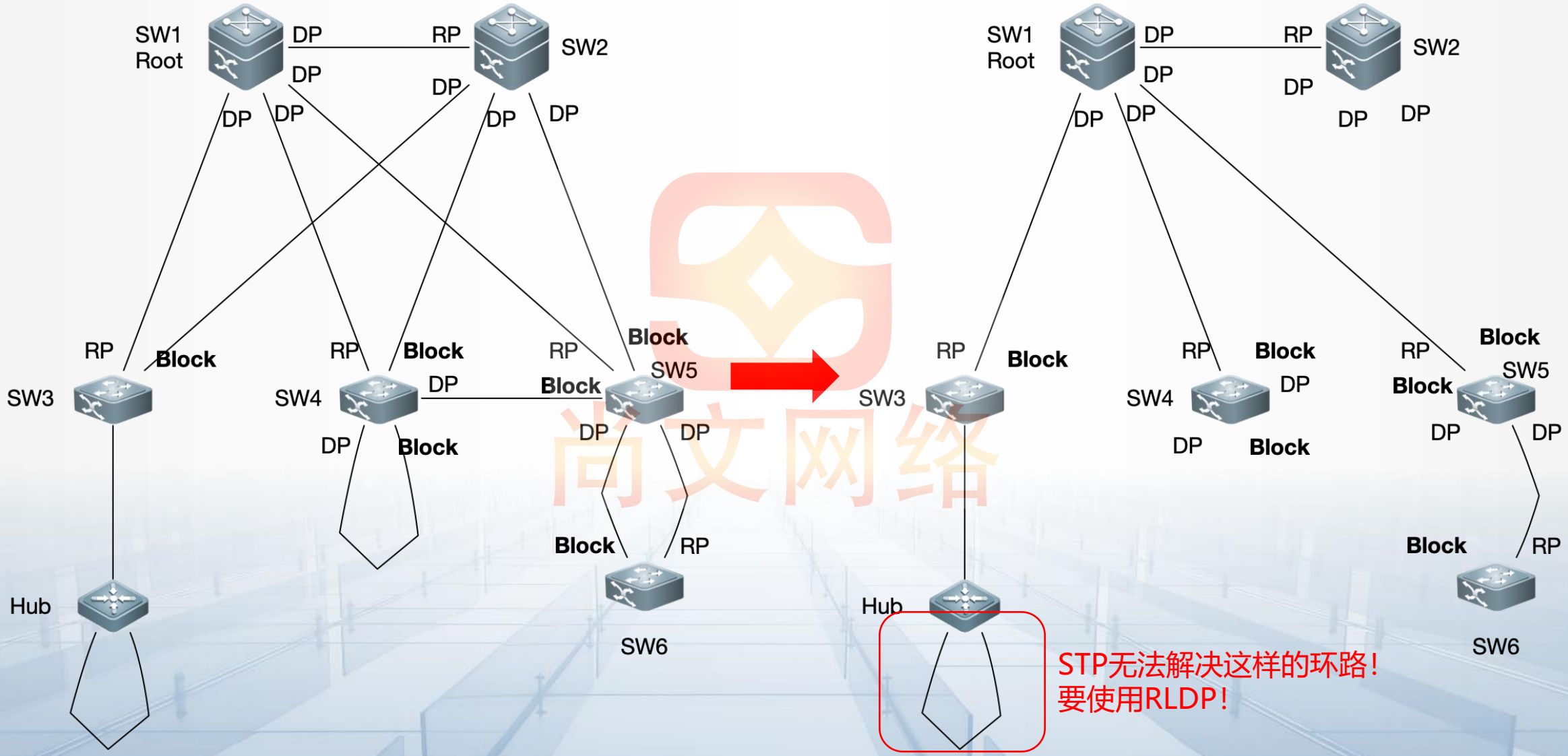
- RP和DP之外的端口全部Block

STP实际案例

- 拓扑图说明：
  - 6台交换机的MAC地址分别为1111.1111.1111-6666.6666.6666
  - SW1，桥优先级4096，所有接口优先级和Cost值均为默认值
  - SW2，桥优先级8192，所有接口优先级和Cost值均为默认值
  - SW3，桥优先级默认值，所有接口优先级和Cost为默认值
  - SW4，桥优先级默认值，所有接口优先级和Cost为默认值
  - SW5，桥优先级默认值，0/2口优先级为16，其余默认
  - SW6，桥优先级默认值，所有接口优先级和Cost为默认值
- 请根据以上信息，分析该拓扑图中的根桥、所有根端口、指定端口和Block端口。
- 请根据结果，分析该网络中还存在什么问题？



STP实际案例



STP的端口状态

- STP生成树收敛完成的过程中，接口会出现的五种状态，分别是：Disable、Blocking、Listening、Learning和Forwarding：
  - Disable状态：此时交换机端口没有激活，丢弃所有收到的帧，不学习mac地址，不接收BPDU报文；
  - Blocking状态：监听BPDU，但是不转发BPDU，丢弃所有收到的数据帧，不学习mac地址也不产生任何MAC地址表项；
  - Listening状态：持续15s，接受并发送BPDU，不转发用户数据，不产生该端口的MAC地址表项，在该状态下完成STP的收敛，此状态下交换机能决定根桥，并可以选择根端口、指定端口和非指定端口；
  - Learning状态：持续15s，接受并发送BPDU，不转发用户数据，完成部分端口的MAC地址表项，目的是为了减少当用户开始转发数据时，带来的大规模广播包泛洪；
  - Forwarding状态：接受并发送BPDU，转发用户数据。

端口状态

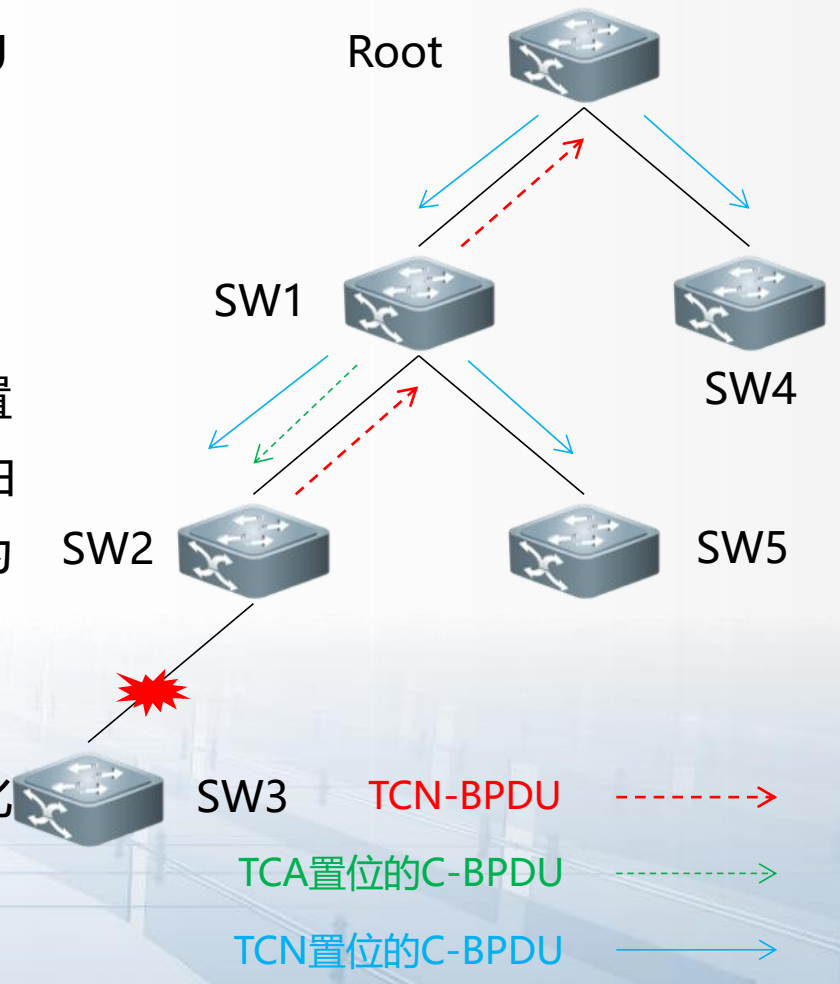
端口状态	端口能力
Disabled	不收发任何报文
Blocking	不接收或转发数据，接收但不发送BPDU，不进行地址学习
Listening	不接收或转发数据，接收并发送BPDU，不进行地址学习
Learning	不接收或转发数据，接收并发送BPDU，开始地址学习
Forwarding	接收并转发数据，接收并发送BPDU，进行地址学习

TCN-BPDU详解

- IEEE 802.1D协议规定，TCN-BPDU（下文简称TC报文）的产生条件有两个：
  - 网桥上有端口转变为Forwarding状态，且该网桥至少包含一个指定端口；
  - 网桥上有端口从Forwarding状态或Learning状态转变为Blocking状态；
- 若上述两个条件之一满足，就说明网络拓扑发生了变化，网桥就需要使用TC报文将拓扑发生变化的情况通知根桥。
- 在日常维护中，TC报文的产生通常有以下几种情况：
  - 设备或链路出现故障，引发STP重新计算，产生TC报文；
  - STP配置参数更改，引发STP重新计算，产生TC报文；
  - 连接终端的端口使能了STP，但没有配置为边缘端口，当终端发生重启等情况导致该端口链路状态变化时，该端口产生TC报文；
  - 来自用户设备的攻击TC报文也可能传入其所接入的二层网络。

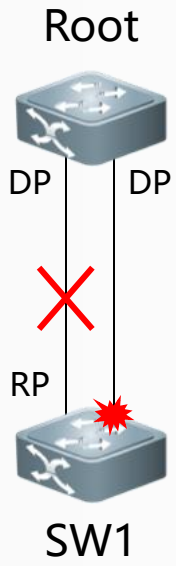
TCN-BPDU交互实例

- 网桥感知到拓扑变化，产生TCN-BPDU，从根端口发出，通知根桥；
- 如果上游网桥不是根桥，则上游网桥会将下一个要发送的配置BPDU中的TCA位置为1，作为对TCN的确认，发送给下游网桥；
- 上游网桥从根端口发送TCN-BPDU，通知根桥；
- 重复第2、3步，直到根桥收到TCN-BPDU；
- 根桥收到TCN-BPDU后，会将下一个要发送的配置BPDU的TCA位置为1作为对TCN的确认。同时根桥会将自己的MAC地址表老化时间由300秒修改为Forward Delay即15秒，同时根桥还会发出TCN位置为1的配置BPDU，用来通知网络中所有网桥网络拓扑发生了变化。
- 根桥在之后的Max Age+Forward Delay时间内，将发送TCN置为1的配置BPDU，当网桥收到该配置BPDU后，会将自己MAC地址老化时间由300s缩短为Forward Delay即15秒。



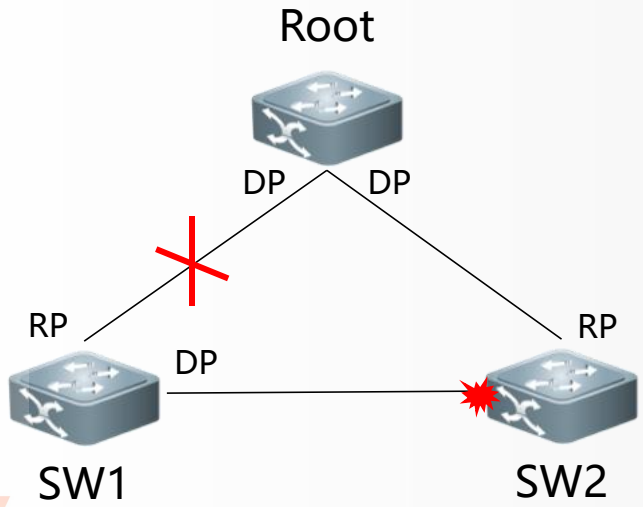
STP重新收敛过程分析

直连故障



- 被阻塞的端口会从Block状态，依次切换到Listening以及Learning状态，最终进入forwarding状态
- 直连链路故障，重新收敛需要Forwarding Delay*2=30s

非直连故障



- SW1上联线路异常，向SW2发送Root ID为自己的桥ID的BPDU，SW2的阻塞端口收到后，发现不比自己端口缓存的BPDU更优，因此忽略；
- Max Age后该端口依次切换到Listening以及Learning状态，最终进入forwarding状态
- 非直连故障，需要Max Age+Forwarding Delay*2=50s

09

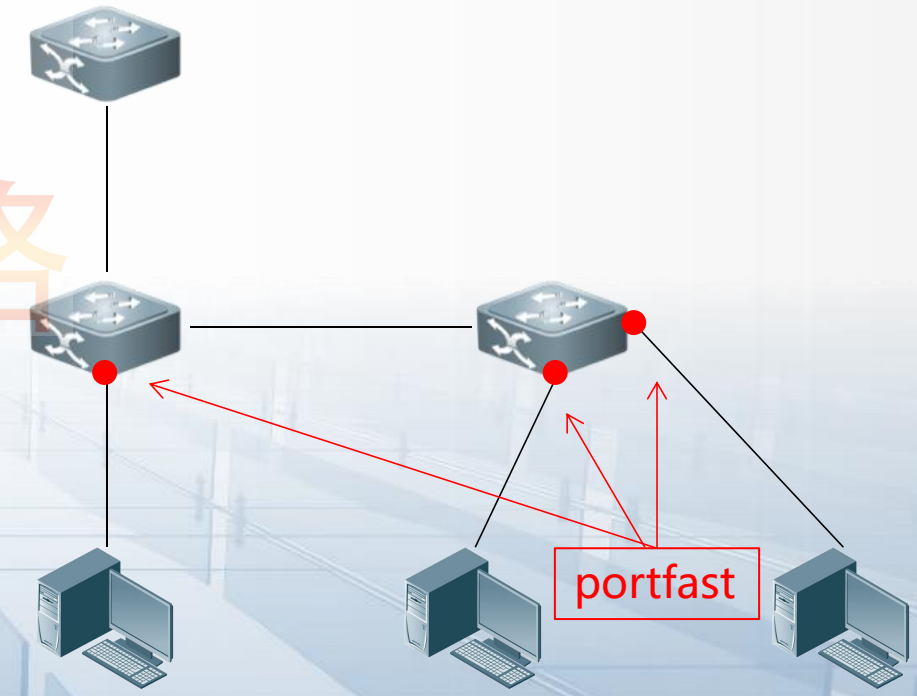
生成树技术原理与应用

- 生成树协议综述
- STP技术原理和配置
- **STP技术高级特性**

STP高级特性-Port Fast

- 交换机的某端口如果接入了用户终端，那么常规配置下，该端口需要经过2个Forwarding Delay才能够进入转发状态（Listening和Learning，共计30秒），很明显这是不合理的；
- 可以将接入用户终端的接口配置为Port Fast，这样该接口可以跳过30秒的等待时间，直接进入转发状态；
- 如果设置了Port Fast的端口还收到了BPDU，则该端口会经过2个Forwarding Delay后进入转发状态；
- 下图表示了一个设备的哪些端口可以配置为 Port Fast：
- 配置命令：

```
Ruijie(config)#interface gigabitEthernet 0/1
Ruijie(config-if-GigabitEthernet 0/1)#spanning-tree
portfast
```



STP高级特性-BPDU Guard

- BPDU Guard的意义就是一个不该接收BPDU的端口（比如portfast端口）一旦收到BPDU报文，那么该功能将会立即关闭该端口，并将端口状态置为error-disabled状态，该模式下，接口可以发出BPDU；
- BPDU Guard的两种配置模式
  - 全局模式启用：全局模式启用该功能之后，会在所有配置了portfast的接口生效。如果某个接口打开了 Port Fast，而该接口收到了 BPDU，那么该端口就会进入 Error-disabled 状态，表示网络中可能被非法用户增加了一台网络设备，使网络拓朴发生改变。配置命令：

```
Ruijie(config)#spanning-tree portfast bpduguard default
```

  - 接口模式启用：打开单个接口的 BPDU Guard(与该端口是否配置 portfast 无关)。在这个情况下如果该端口收到了 BPDU，就会进入 Error-disabled 状态

```
Ruijie(config)#interface gigabitEthernet 0/1
Ruijie(config-if-GigabitEthernet 0/1)#spanning-tree bpduguard enable
```

- Error-disabled如何恢复
  - 接口模式下，shutdown然后no shutdown
  - 全局模式下，errdisable recovery interval 300s

STP高级特性-BPDU Filter

- BPDU Filter可以过滤掉接口上收到或发出的BPDU；
- BPDU Guard的两种配置模式
  - 全局模式启用：打开全局的 BPDU Filter功能，在这种状态下，开启Port Fast的接口将既不收 BPDU，也不发出BPDU，连到portfast端口的主机就收不到BPDU。而如果开启portfast的端口收到了BPDU，那么该端口的portfast属性将失效，同时BPDU Filter也自动失效，配置命令：

```
Ruijie(config)#spanning-tree portfast bpdufilter default
```

  - 接口模式启用：在接口模式下，开启单个接口的BPDU Filter(与该端口 是否打开 Port Fast 无关)。在这个情况下该接口既不收 BPDU，也不发 BPDU，相当于关闭了该接口的STP功能，接口直接进入转发状态。

```
Ruijie(config)#interface gigabitEthernet 0/1
Ruijie(config-if-GigabitEthernet 0/1)#spanning-tree bpdufilter enable
```

STP高级特性-Tc-protection

- TC-BPDU 报文是指携带 TC 标志的 BPDU 报文，交换机收到这类报文表示网络拓扑发生了变化，会进行 MAC 地址表的删除操作。对三层交换机，还会引发快转模块的重新打通操作，并改变 ARP 表项的端口状态。
- 为避免交换机受到伪造 TC-BPDU 报文的恶意攻击时频繁进行以上操作，设备CPU负荷过重，影响网络稳定，可以使用 TC-protection 功能进行保护：
  - 在打开相应功能时，收到TC-BPDU 报文后的一定时间内(一般为 4 秒)，只进行一次删除操作，同时监控该时间段内是否收到 TC-BPDU报文。
  - 如果在该时间段内收到了TC-BPDU报文，则设备在该时间超时后再进行一次删除操作。这样可以避免频繁的删除 MAC 地址表项和 ARP表项，保护设备CPU资源。
- 配置命令：

```
Ruijie(config)#spanning-tree tc-protection
```

思考

1. 下列关于STP接口状态的描述，错误的是哪一项？（　　）

A. 被阻塞的接口不会侦听，也不发送BPDU

B. 处于Lerarning状态的接口会学习MAC地址，但是不会转发数据

C. 处于Listening状态的接口会持续侦听BPDU

D. 被阻塞的接口如果一定时间内收不到BPDU，则会自动切换到Listening状态

本章总结

- 生成树协议综述
- STP技术原理与配置
- STP技术高级特性

尚文网络RCNA

谢谢