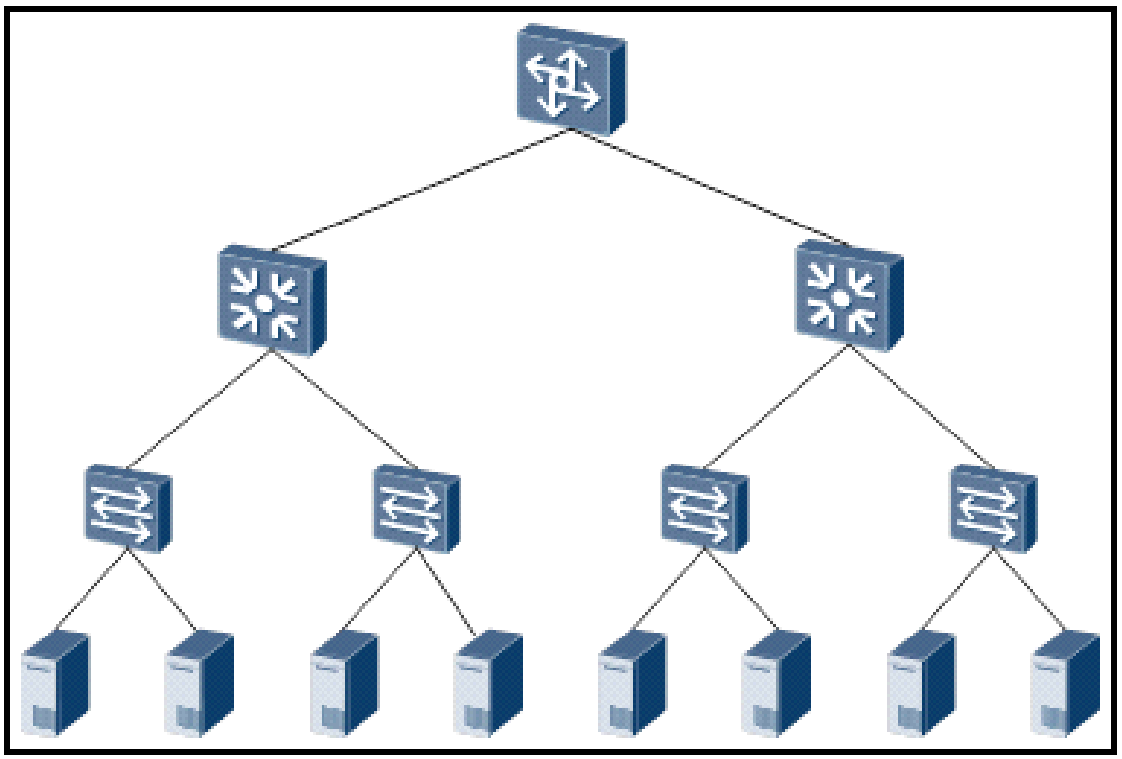
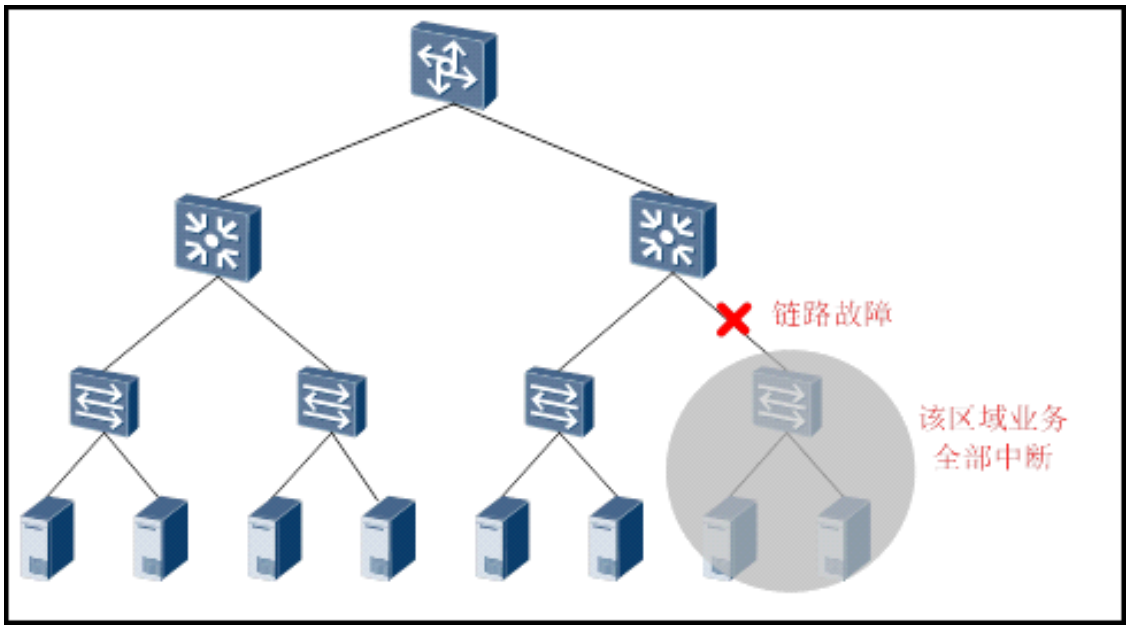
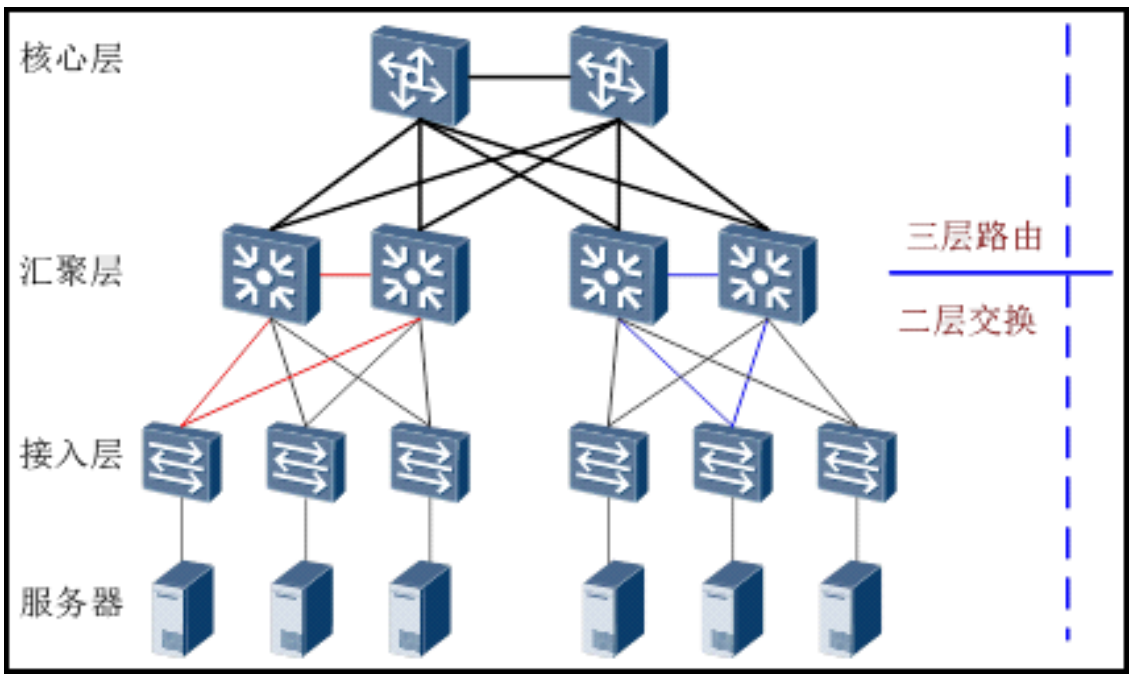
STP

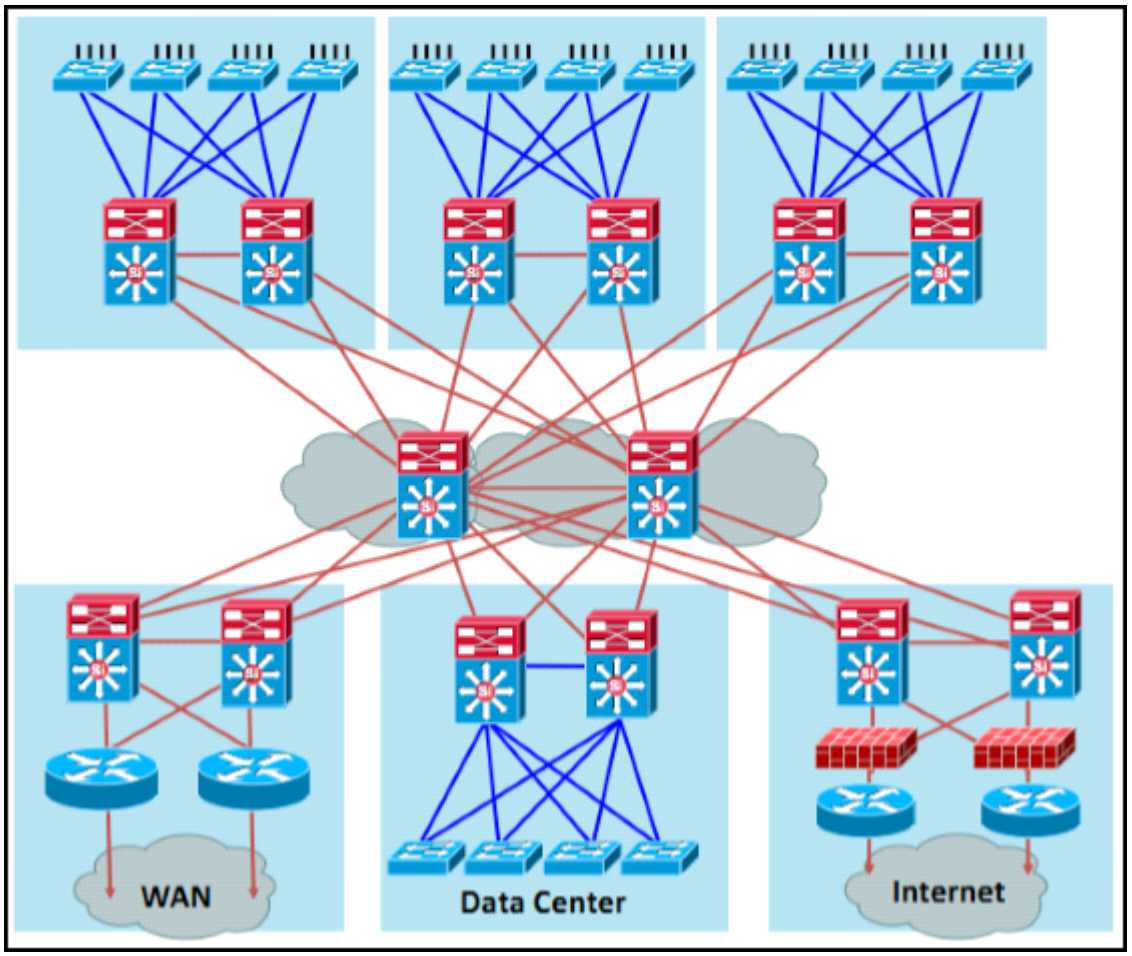
前言

为了提高网络可靠性，交换网络中通常会使用冗余链路。然而，冗余链路会给交换网络带来环路风险，并导致广播风暴以及MAC地址表不稳定等问题，进而会影响到用户的通信质量。生成树协议STP（Spanning Tree Protocol）可以在提高可靠性的同时又能避免环路带来的各种问题。

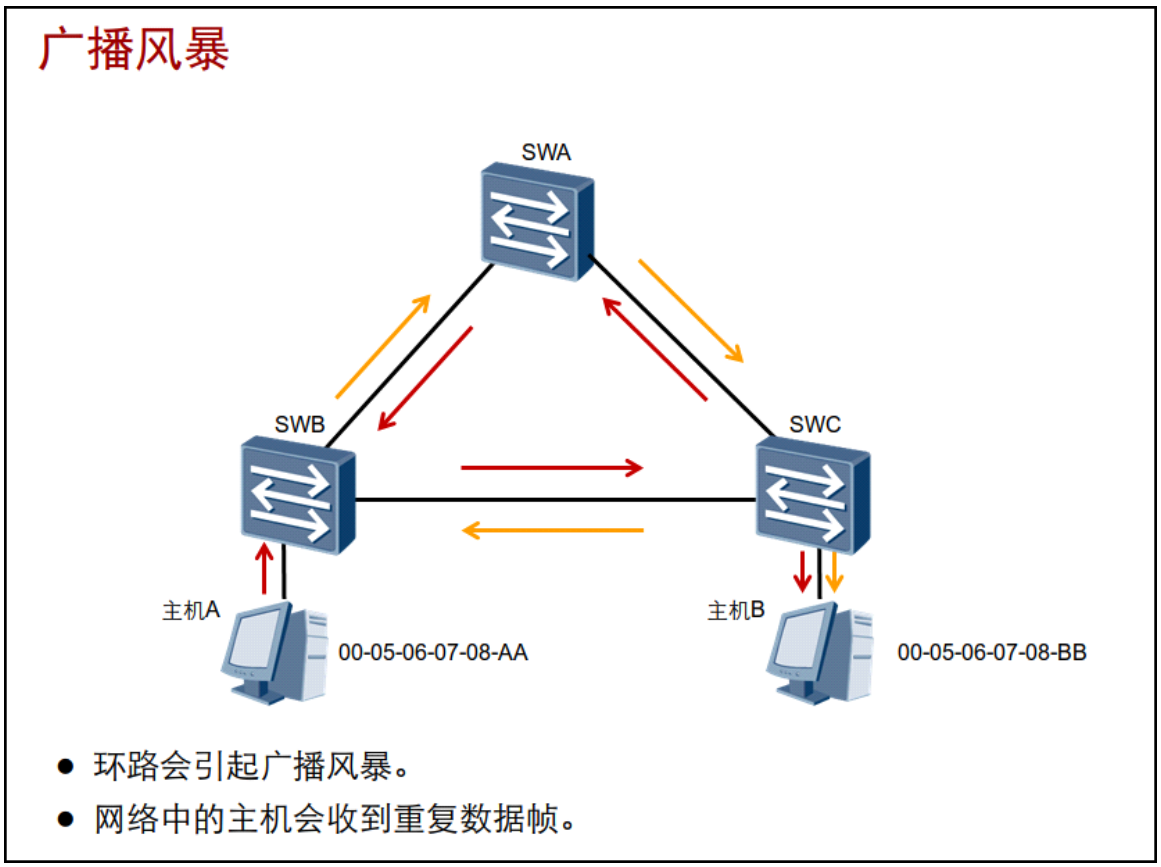
二层网络设计需求和问题：

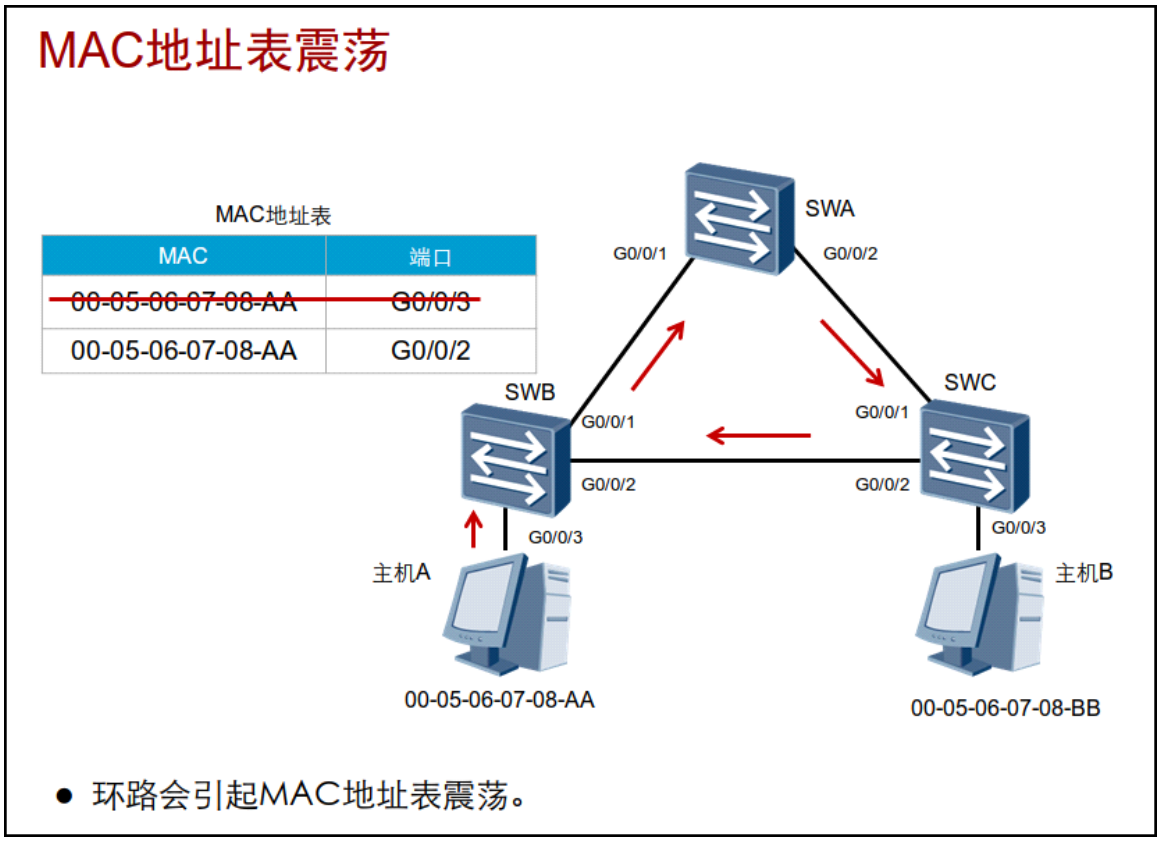
- 为了提高可靠性，交换机之间会通过多条链路相连，从而避免单点故障。
- 但同时会带来一些灾难性的环路问题。
- 

- 
链路故障
该区域业务
全部中断
- 
核心层
汇聚层
接入层
服务器
三层路由
二层交换

- 
WAN
Data Center
Internet

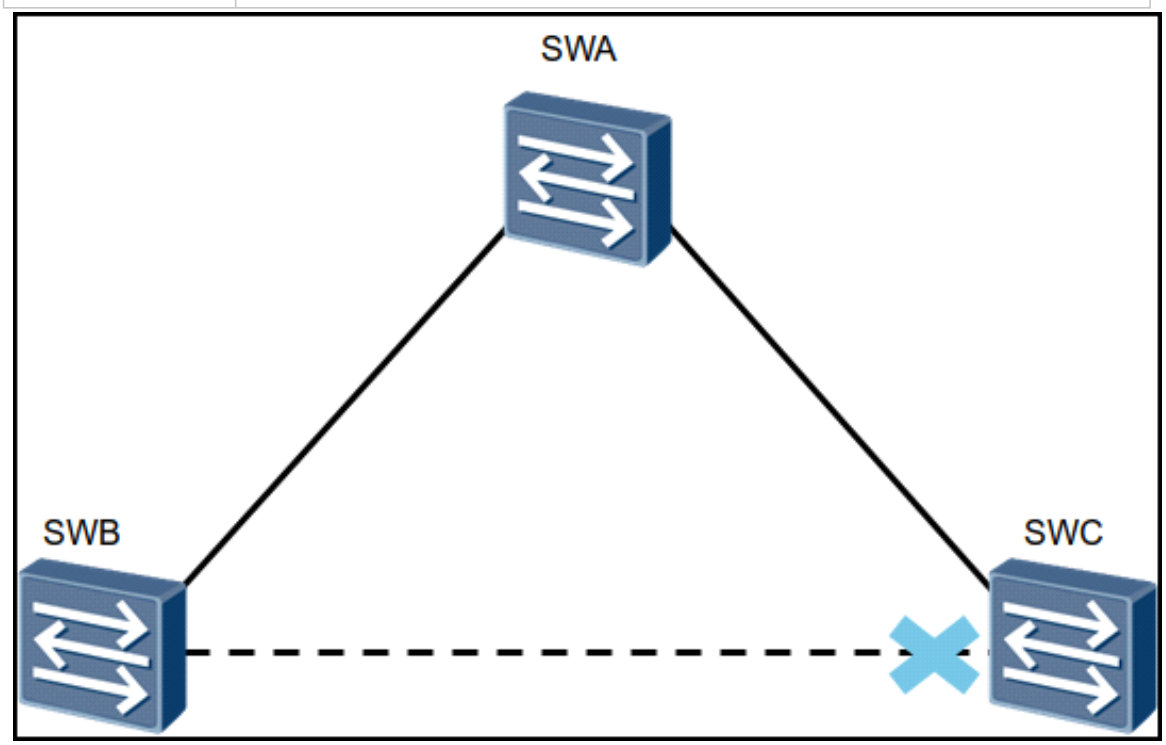
环路问题：

- 
广播风暴
SWA
SWB
SWC
主机A
00-05-06-07-08-AA
主机B
00-05-06-07-08-BB
环路会引起广播风暴。
网络中的主机会收到重复数据帧。

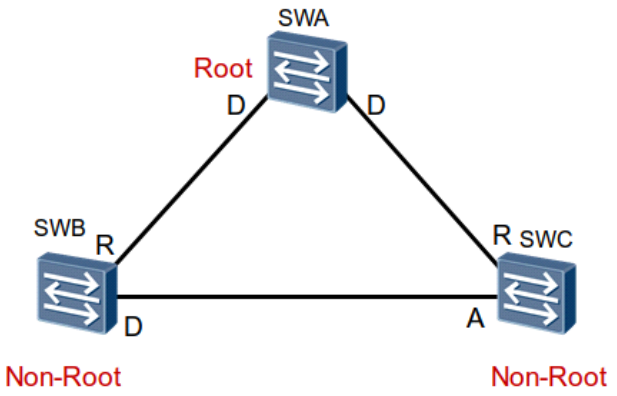
- 

STP：Spanning Tree Protocol，生成树协议，提供两大功能：

消除环路	通过阻断冗余链路来消除网络中可能存在的环路。
链路备份	当活动路径发生故障时，激活备份链路，及时恢复网络连通性。

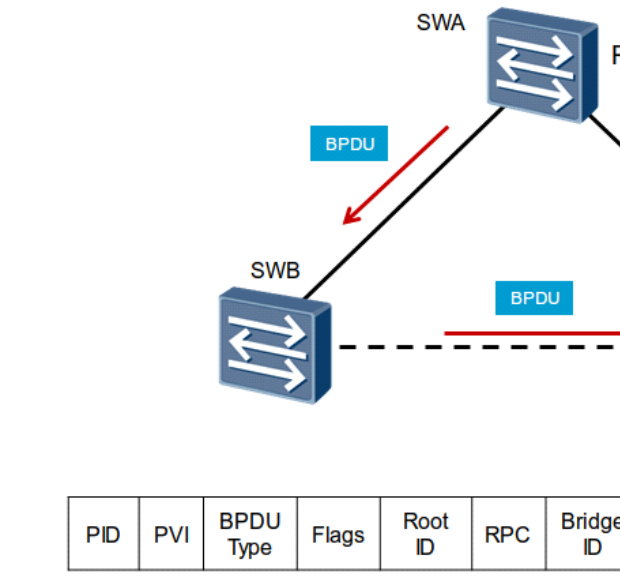
- 

STP操作：通过构造一棵树来消除交换网络中的环路。

- 

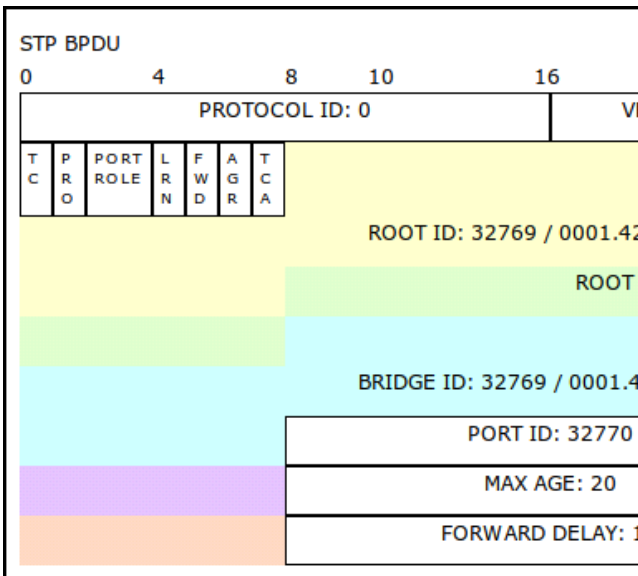
1. 选举一个根桥。
2. 每个非根交换机选举一个根端口。
3. 每个网段选举一个指定端口。
4. 阻塞非根、非指定端口。

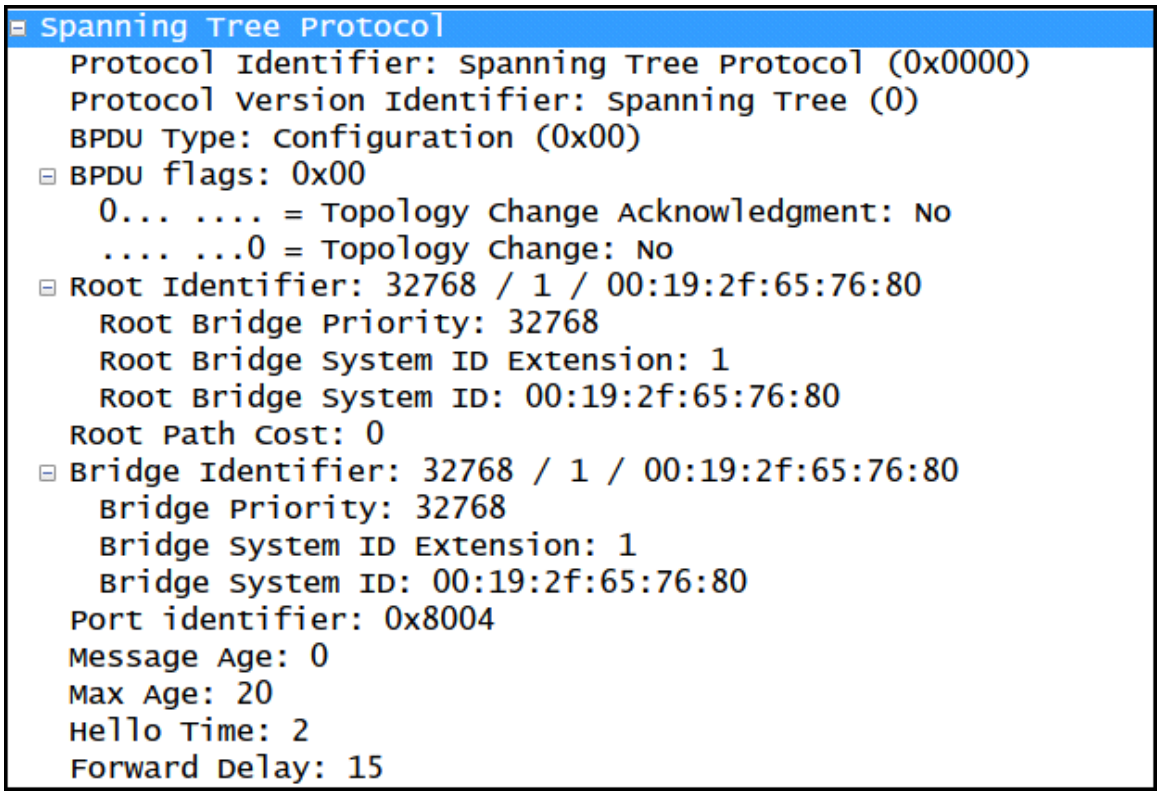
BPDU：Bridge Protocol Data Unit - 桥协议数据单元，STP工作协议

- 使用组播 - 01-80-C2-00-00-00
- 

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

- BPDU包含桥ID、路径开销、端口ID、计时器等参数。

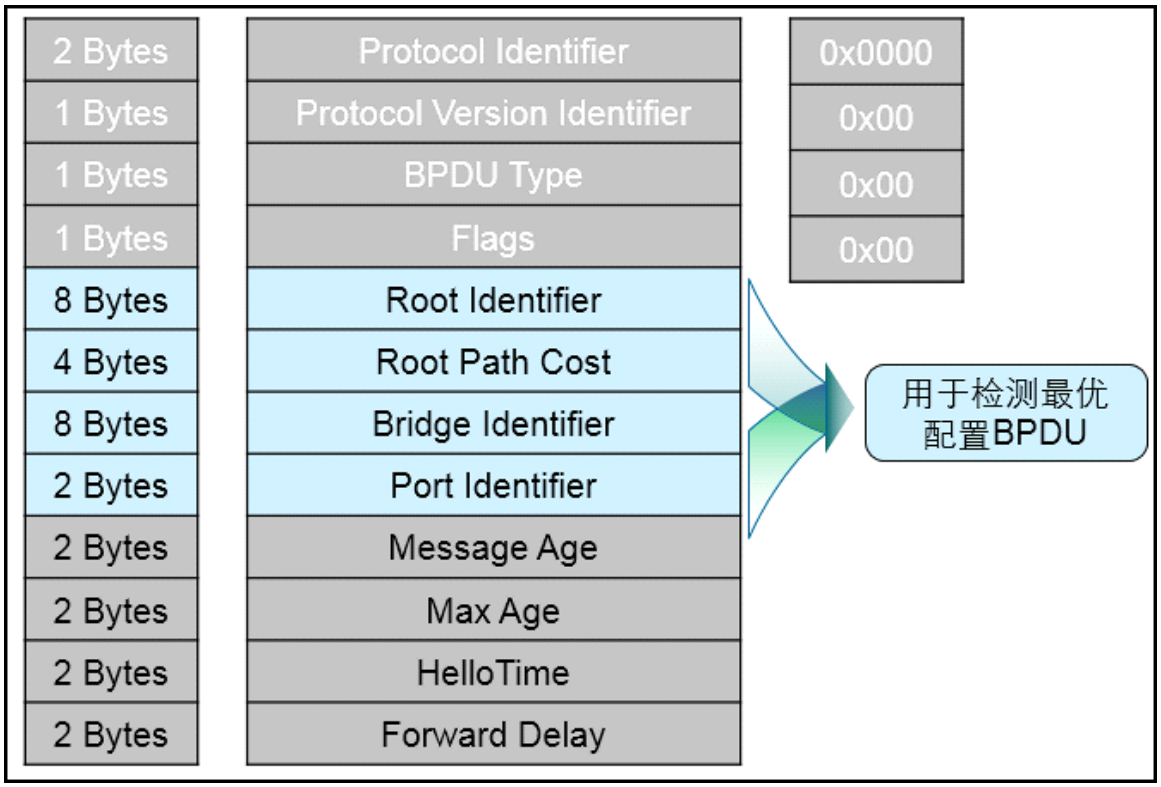
- 

- 

BPDU类型：

- 配置BPDU：
  - 选举根交换机以及确定每个交换机端口的角色和状态。
  - 在初始化过程中，每个桥都主动发送配置BPDU。
  - 在网络拓扑稳定以后，只有根桥主动发送配置BPDU，其他交换机在收到上游传来的配置BPDU后，才会发送自己的配置BPDU。
  - 发送周期为Hello Time。
  - 老化时间为Max Age。
- 拓扑变更通告BPDU-TCN BPDU
  - 下游交换机感知到拓扑发生变化时向上游发送的拓扑变化通知。

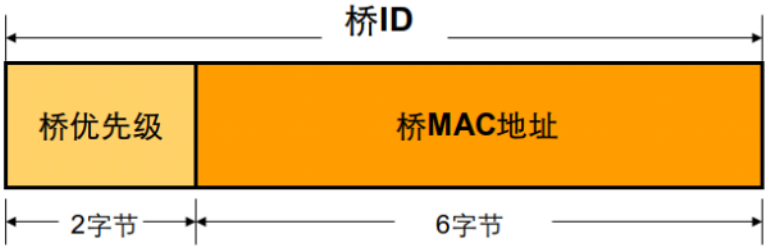
BPDU字段详解：



参 数	描 述
Root Identifier	发送此配置BPDU的交换机所认为的根交换机的交换机标识
Root Path Cost	从发送此配置BPDU的交换机到达根交换机的最短路径总开销，含交换机根端口的开销，不含发送此配置BPDU的端口的开销
Bridge Identifier	发送此配置BPDU的交换机的交换机标识
Port Identifier	发送此配置BPDU的交换机端口的端口标识

桥ID

- BID - Bridge ID
  - 用于在STP中唯一标识一个交换机，由两部分组成：
    - 桥优先级：高16位
    - 桥MAC地址：低48位

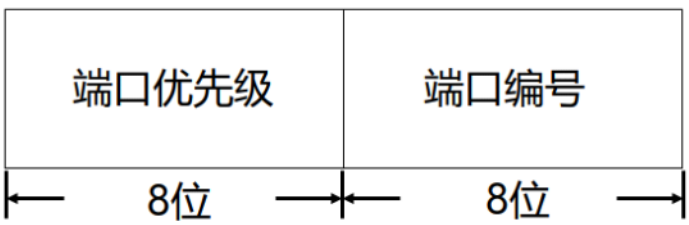


优先级取值范围：0 ~ 65535

缺省值：32768

端口ID

- PID – Port ID
  - 用于在STP中唯一标识一个交换机上的端口，由两部分组成：
    - 端口优先级：高8位
    - 端口编号：低8位



优先级取值范围：0 ~ 255

缺省值：128

路径开销

- Path Cost
  - 路径开销用于衡量桥与桥之间路径的优劣，越低越好
  - STP中每条链路都具有开销值

链路速度	开销（修订的 IEEE 规范）	开销（之前的 IEEE 规范）

路径开销

Path Cost

- 路径开销用于衡量桥与桥之间路径的优劣，越低越好
- STP中每条链路都具有开销值

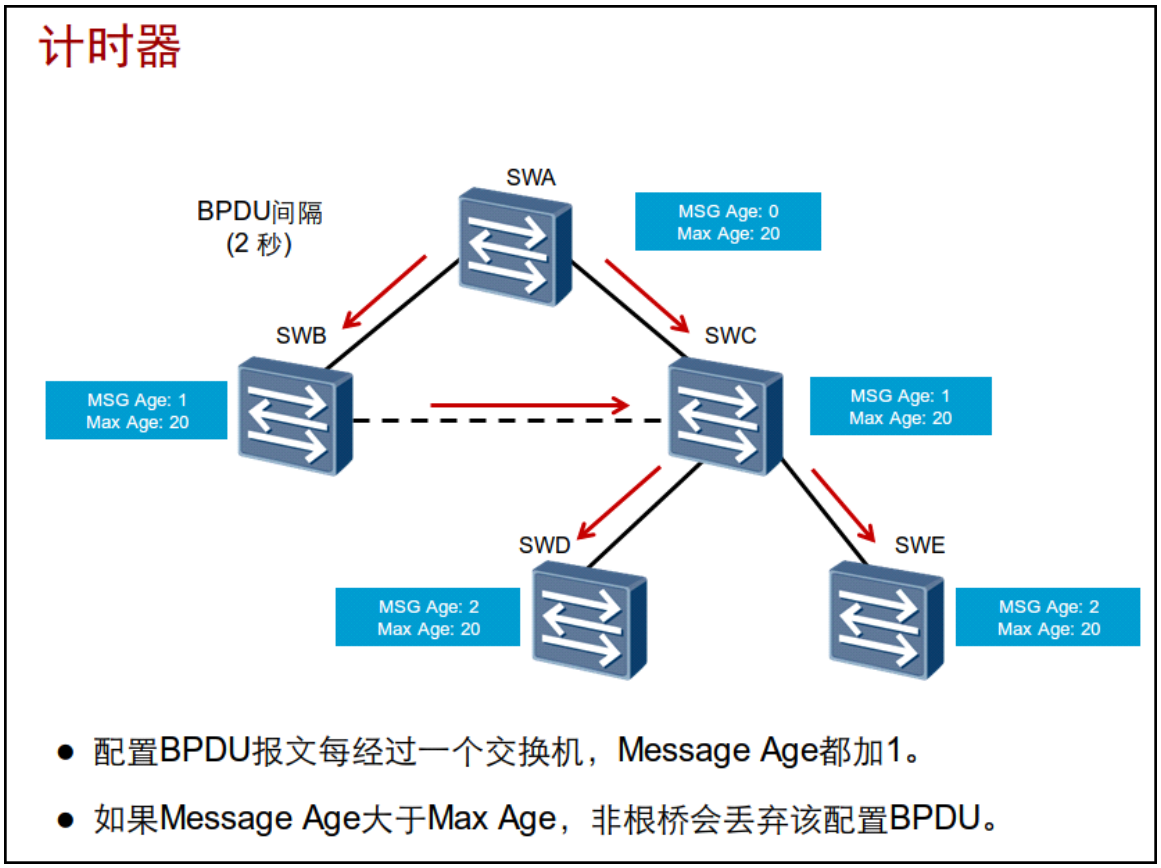
链路速度	开销（修订的 IEEE 规范）	开销（之前的 IEEE 规范）
10 Gb/s	2	1
1 Gb/s	4	1
100 Mb/s	19	10
10 Mb/s	100	100

根路径开销

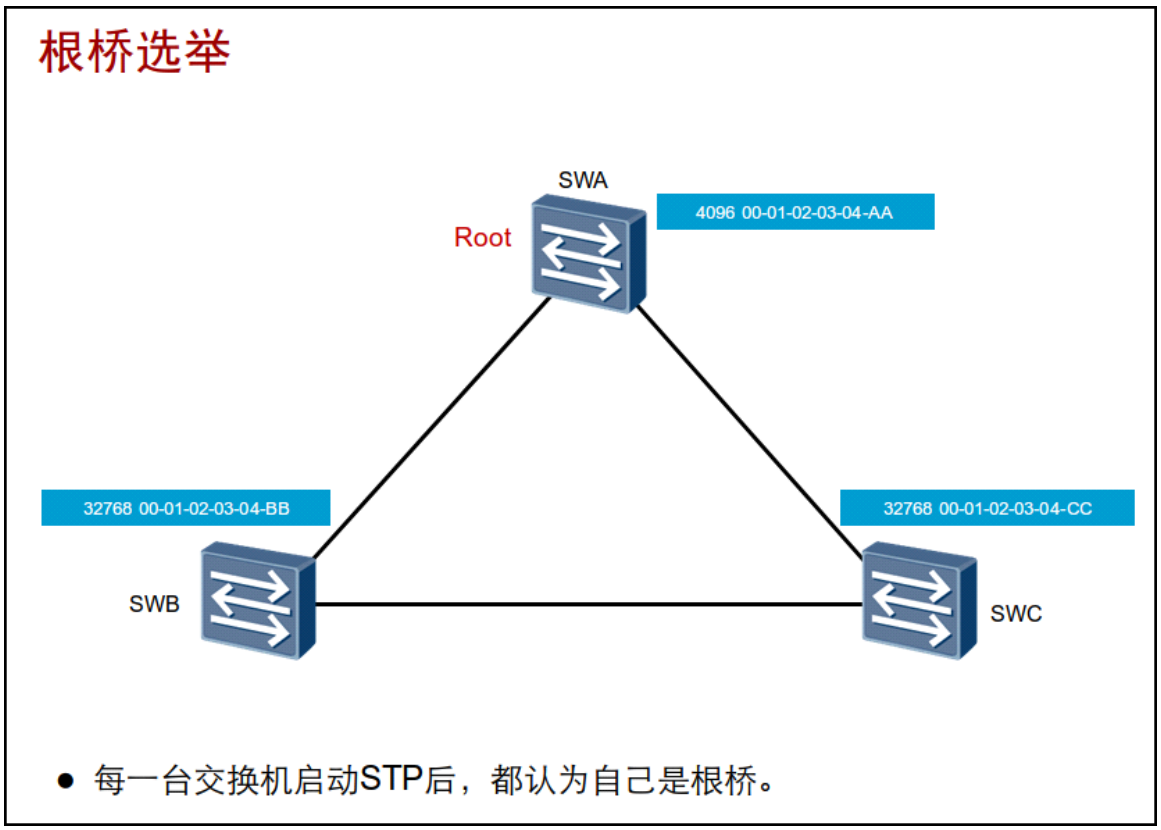
Root Path Cost

- 确定到达根桥的最短路径，并生成无环树状网络。
- 到根桥的路径上所有路径开销之和

SW3的Port 1根路径开销＝19＋100＝119

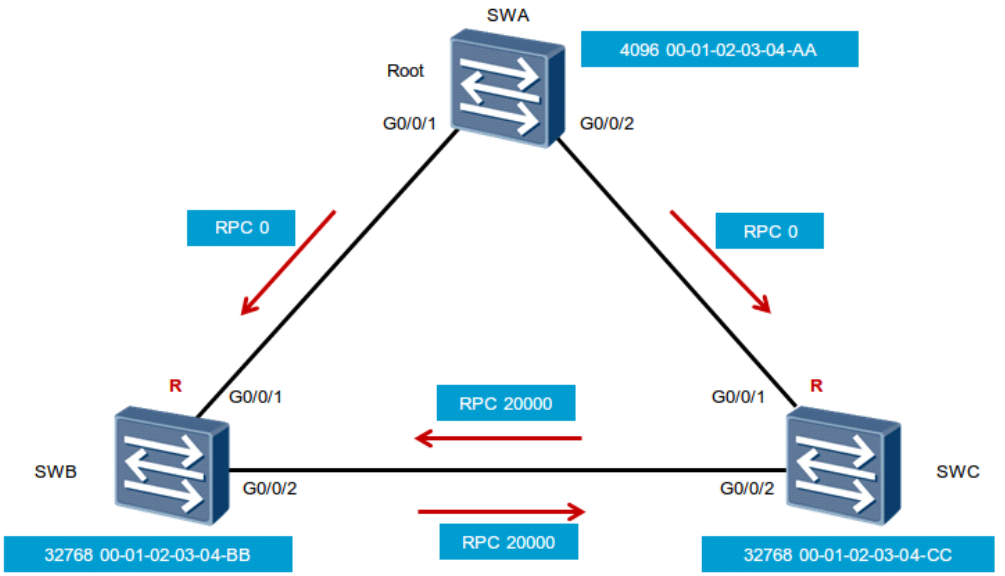
- 

STP选举过程：

- 

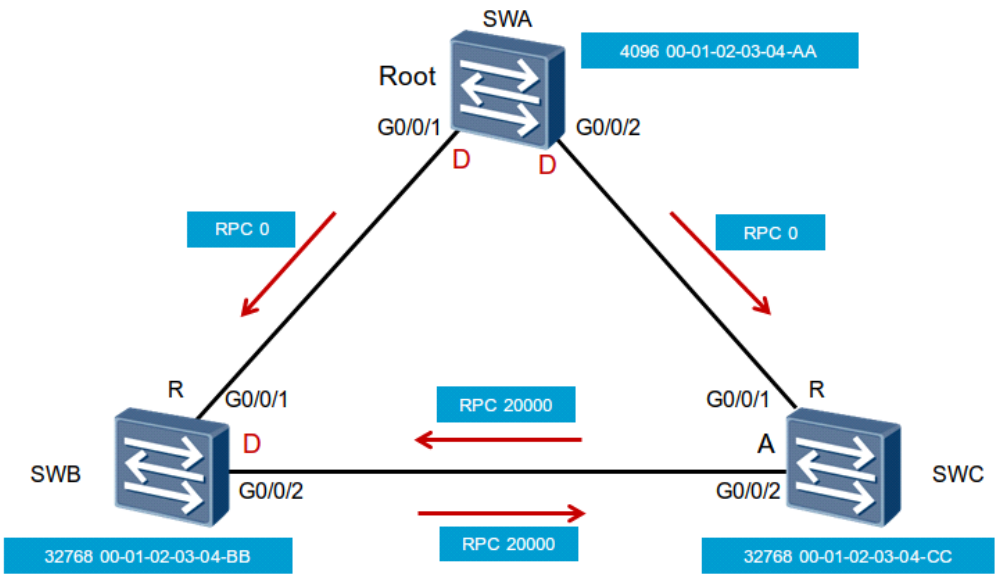
- BID最小的成为根桥（先比较优先级，再比较MAC）

- 根端口选举



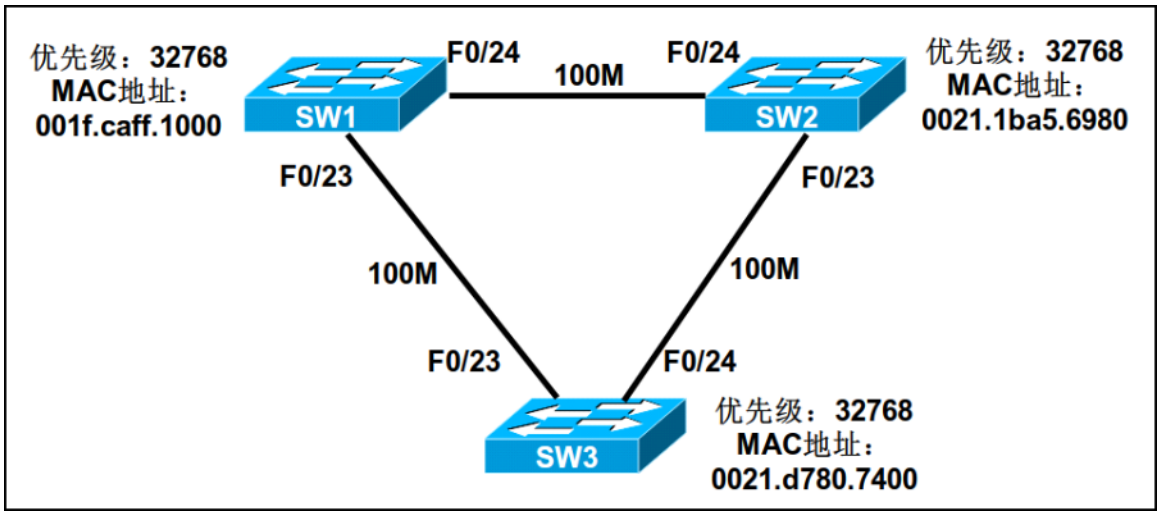
  - 非根交换机在选举根端口时分别依据该端口的根路径开销、对端BID、对端PID和本端PID。

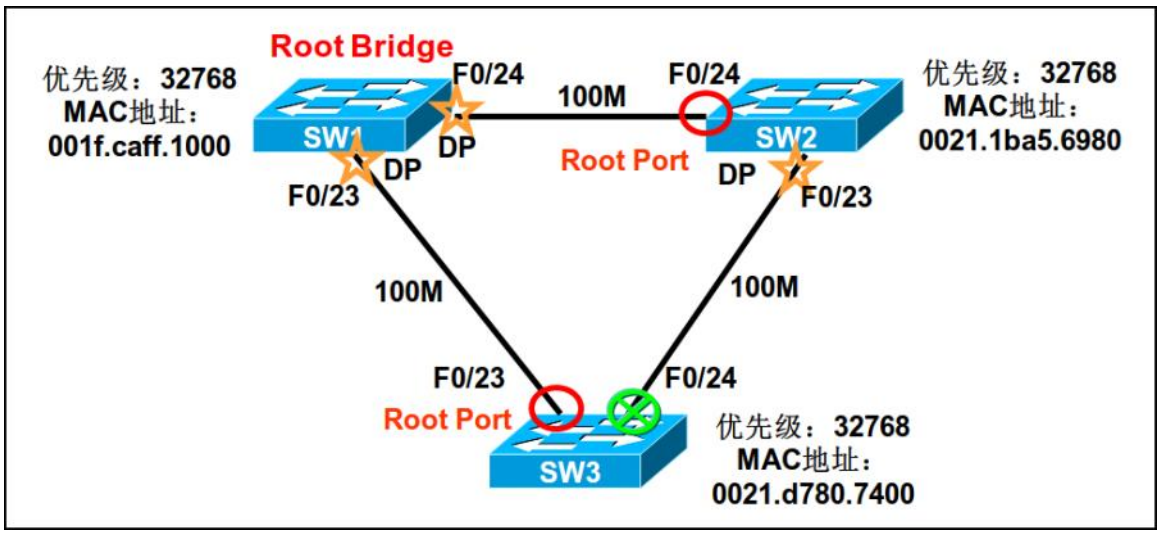
- 指定端口选举

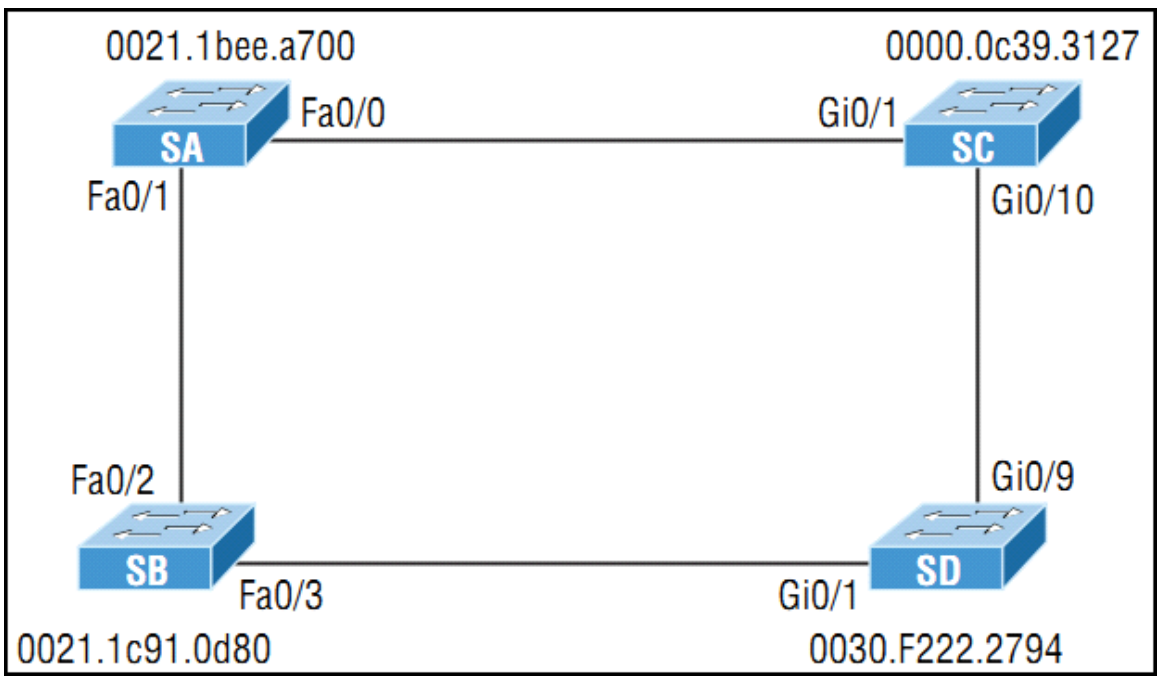


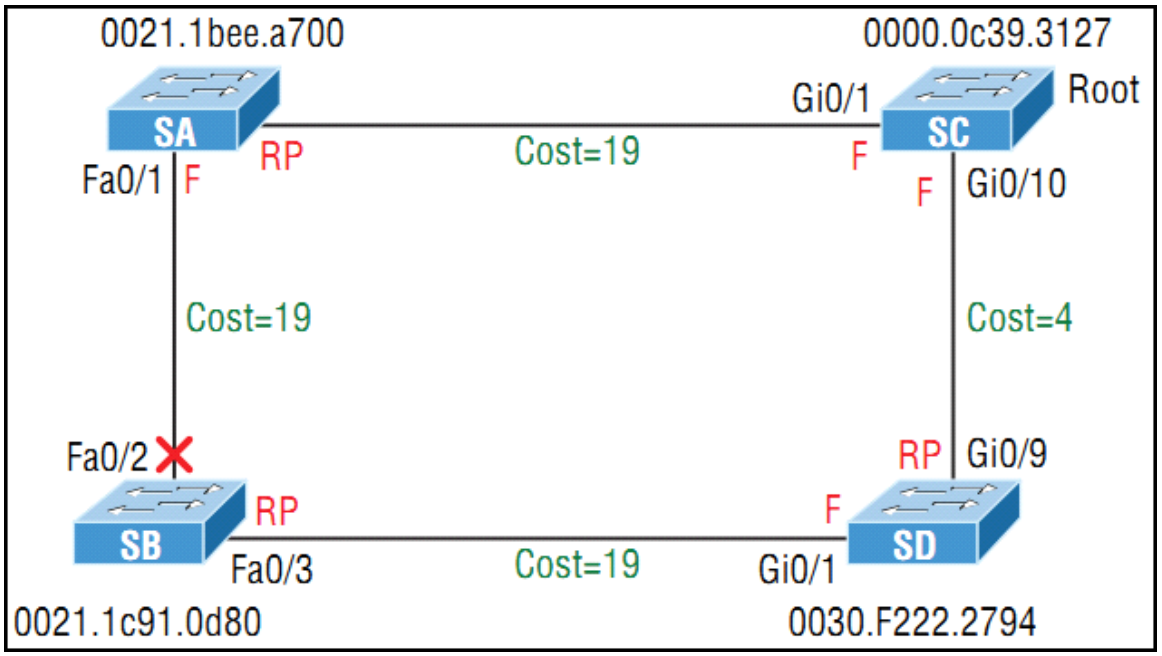
  - 非根交换机在选举指定端口时分别依据根路径开销、BID、PID。
  - 未被选举为根端口或指定端口的端口为预备端口，将会被阻塞。

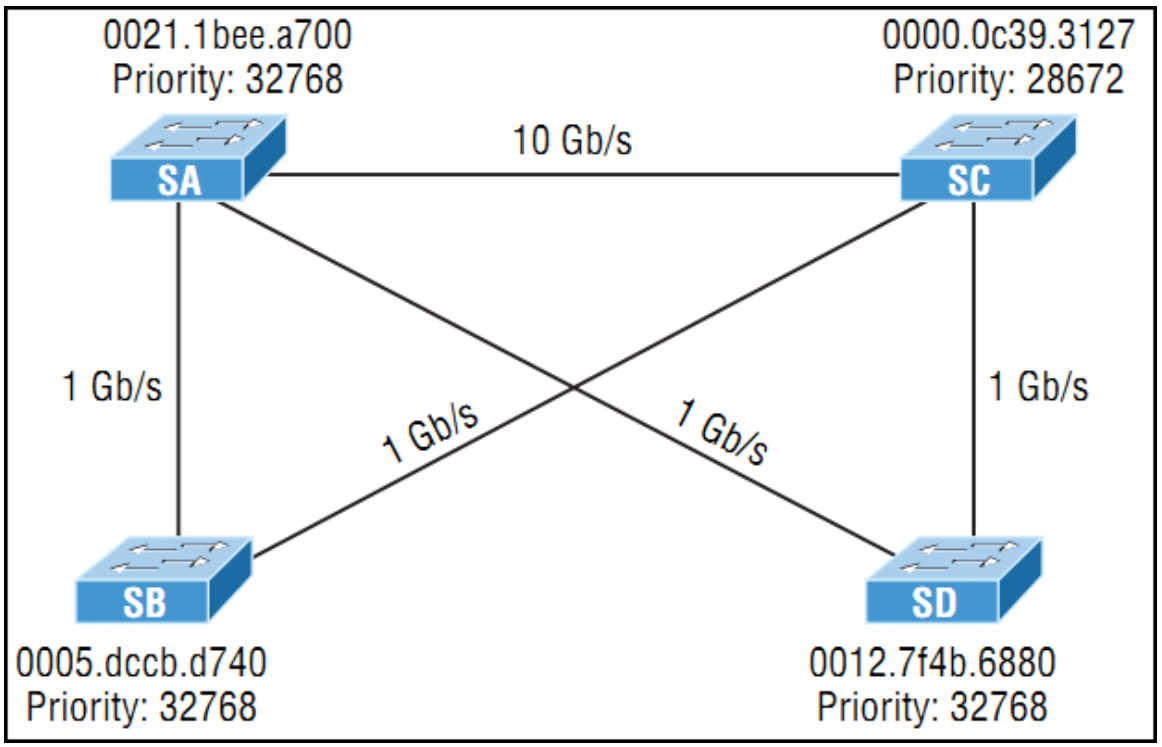
STP选举案例：

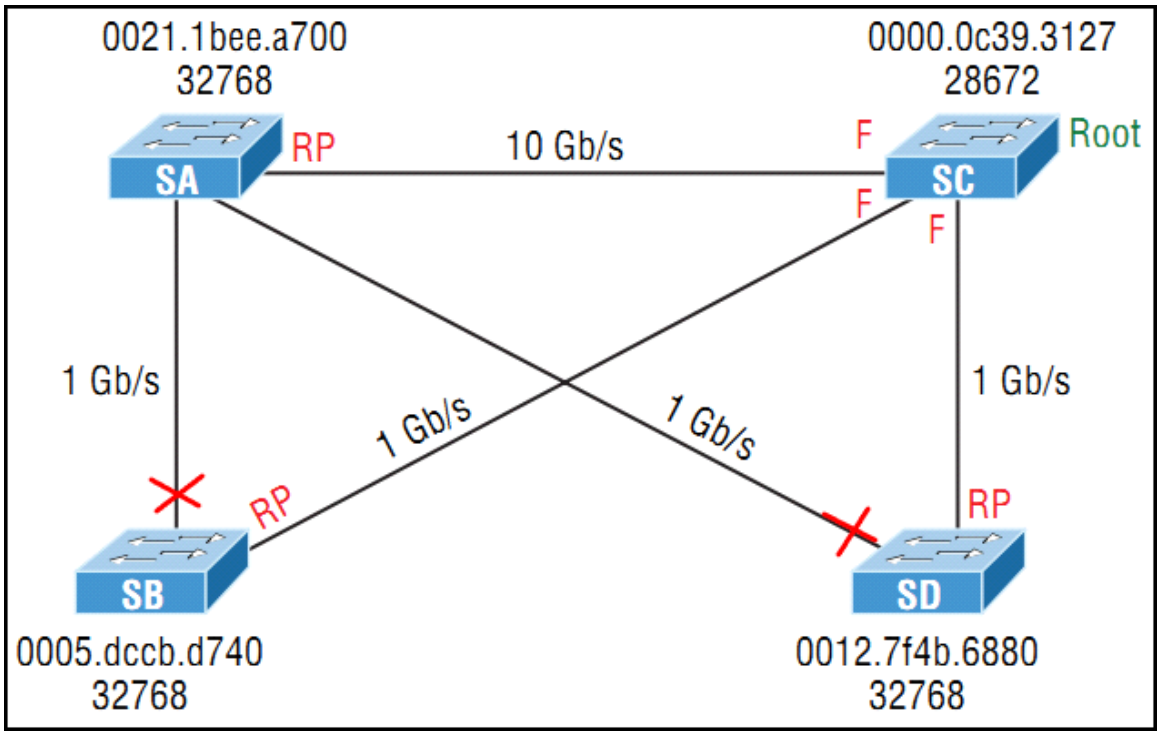

优先级：32768
MAC地址：
001f.caff.1000
SW1
F0/24
100M
F0/24
SW2
优先级：32768
MAC地址：
0021.1ba5.6980
F0/23
F0/23
100M
100M
F0/23
F0/24
SW3
优先级：32768
MAC地址：
0021.d780.7400

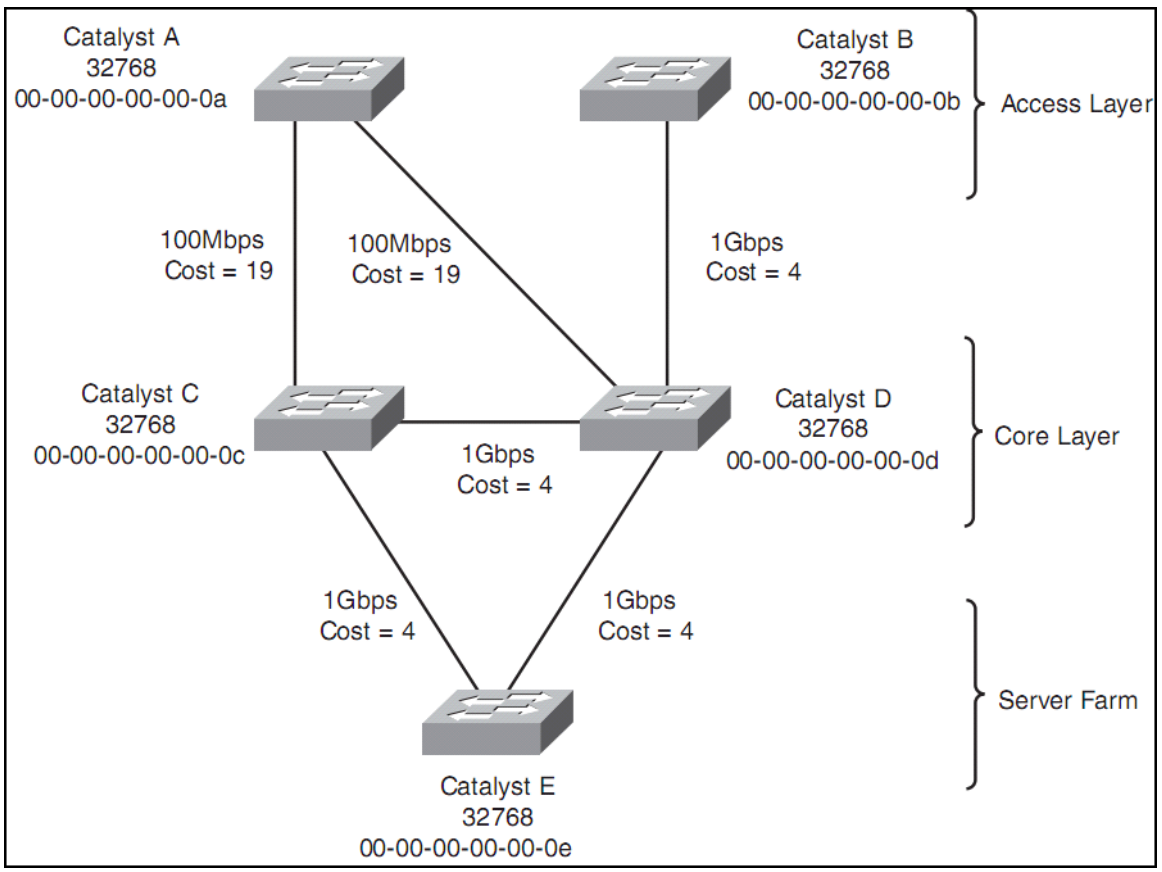

Root Bridge
优先级：32768
MAC地址：
001f.caff.1000
F0/24
100M
F0/24
SW1
SW2
优先级：32768
MAC地址：
0021.1ba5.6980
DP
DP
Root Port
DP
F0/23
F0/23
100M
100M
F0/23
F0/24
Root Port
SW3
优先级：32768
MAC地址：
0021.d780.7400

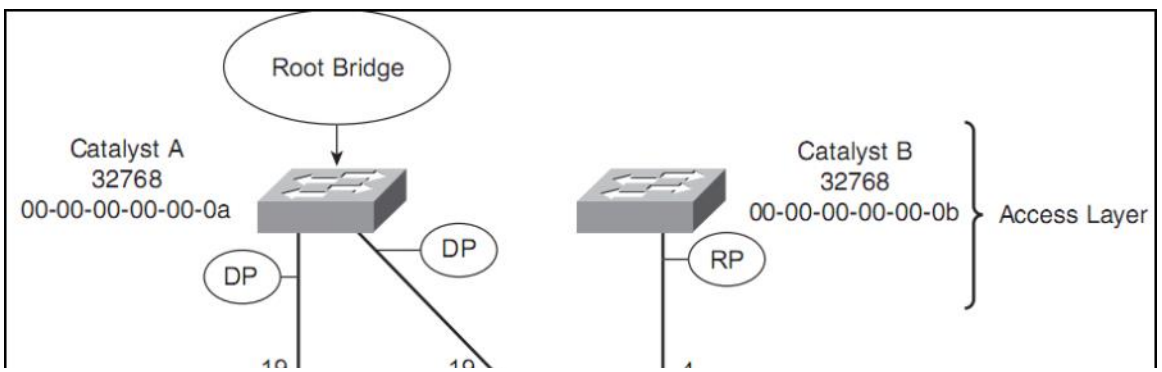

0021.1bee.a700
0000.0c39.3127
SA
Fa0/0
Gi0/1
SC
Fa0/1
Gi0/10
Fa0/2
Gi0/9
SB
SD
Fa0/3
Gi0/1
0021.1c91.0d80
0030.F222.2794

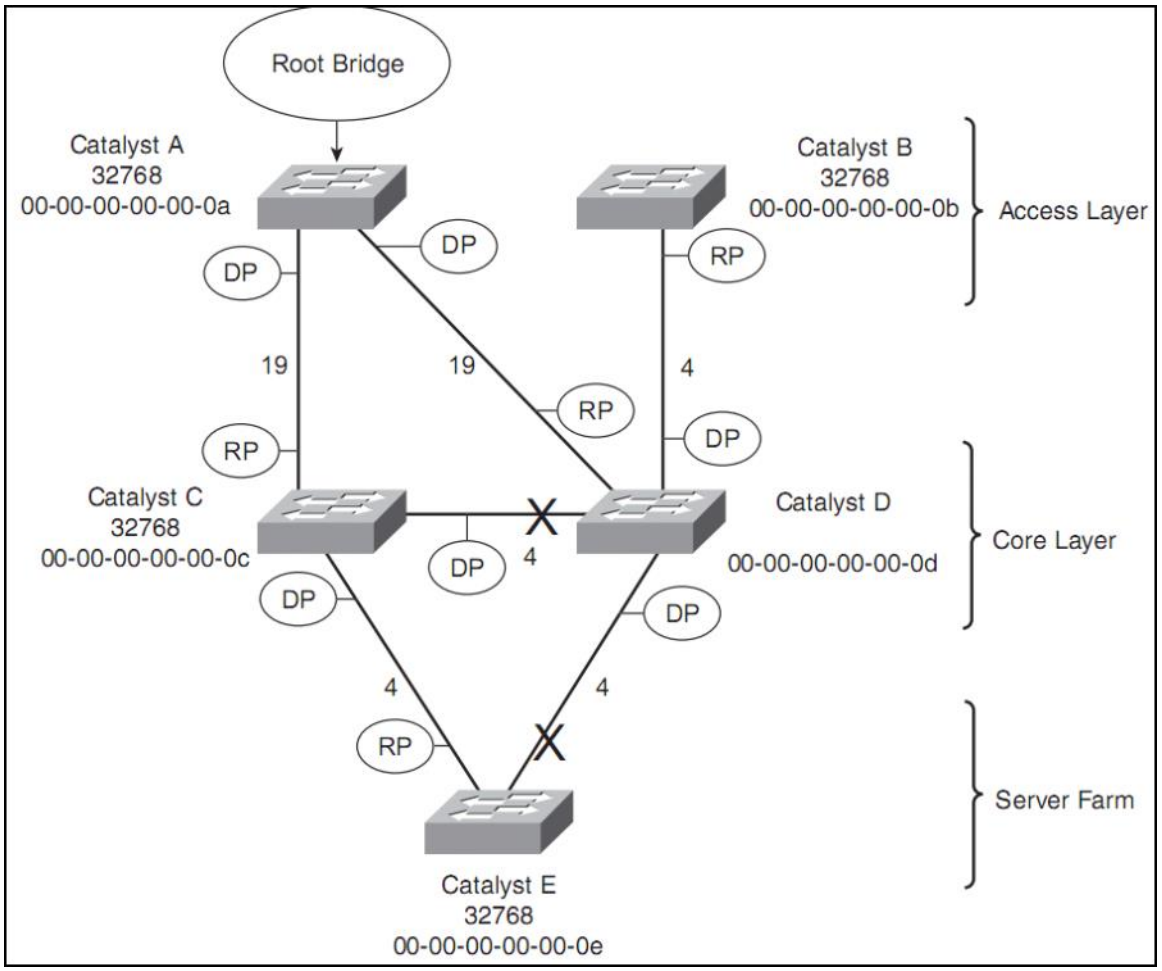
- 
0021.1bee.a700
0000.0c39.3127
SA
SC
Root
Gi0/1
RP
Cost=19
F
Fa0/1
F
F
Gi0/10
Cost=19
Cost=4
Fa0/2
RP
Gi0/9
SB
RP
F
SD
Fa0/3
Cost=19
Gi0/1
0021.1c91.0d80
0030.F222.2794

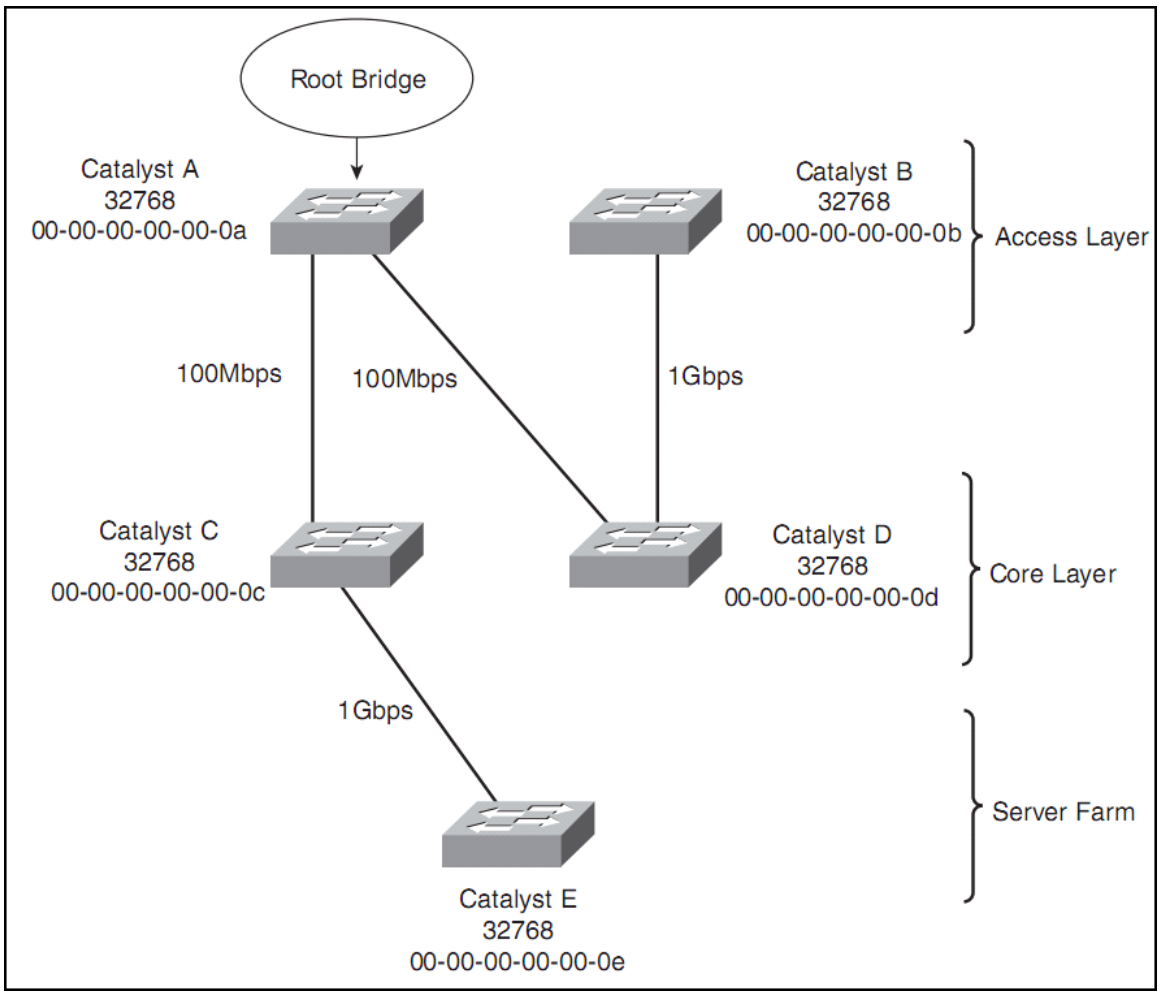
- 
0021.1bee.a700
Priority: 32768
0000.0c39.3127
Priority: 28672
SA
10 Gb/s
SC
1 Gb/s
1 Gb/s
1 Gb/s
1 Gb/s
SB
SD
0005.dccb.d740
Priority: 32768
0012.7f4b.6880
Priority: 32768

- 
0021.1bee.a700
32768
0000.0c39.3127
28672
SA
RP
10 Gb/s
F
Root
SC
F
F
1 Gb/s
1 Gb/s
1 Gb/s
1 Gb/s
RP
RP
SB
SD
0005.dccb.d740
32768
0012.7f4b.6880
32768

- 
Catalyst A
32768
00-00-00-00-00-0a
Catalyst B
32768
00-00-00-00-00-0b
Access Layer
100Mbps
Cost = 19
100Mbps
Cost = 19
1Gbps
Cost = 4
Catalyst C
32768
00-00-00-00-00-0c
Catalyst D
32768
00-00-00-00-00-0d
Core Layer
1Gbps
Cost = 4
1Gbps
Cost = 4
1Gbps
Cost = 4
Server Farm
Catalyst E
32768
00-00-00-00-00-0e


Root Bridge
Catalyst A
32768
00-00-00-00-00-0a
Catalyst B
32768
00-00-00-00-00-0b
Access Layer
DP
DP
RP

- 
Root Bridge
Catalyst A
32768
00-00-00-00-00-0a
Catalyst B
32768
00-00-00-00-00-0b
Access Layer
DP
DP
RP
19
19
4
RP
RP
DP
Catalyst C
32768
00-00-00-00-00-0c
Catalyst D
00-00-00-00-00-0d
Core Layer
DP
4
DP
DP
4
4
RP
Server Farm
Catalyst E
32768
00-00-00-00-00-0e

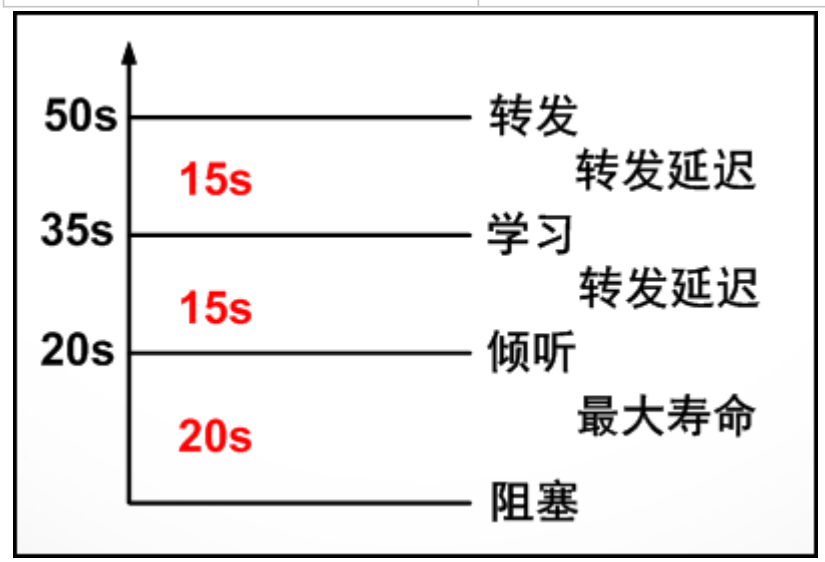
- 
Root Bridge
Catalyst A
32768
00-00-00-00-00-0a
Catalyst B
32768
00-00-00-00-00-0b
Access Layer
100Mbps
100Mbps
1Gbps
Catalyst C
32768
00-00-00-00-00-0c
Catalyst D
32768
00-00-00-00-00-0d
Core Layer
1Gbps
Server Farm
Catalyst E
32768
00-00-00-00-00-0e

STP端口状态：

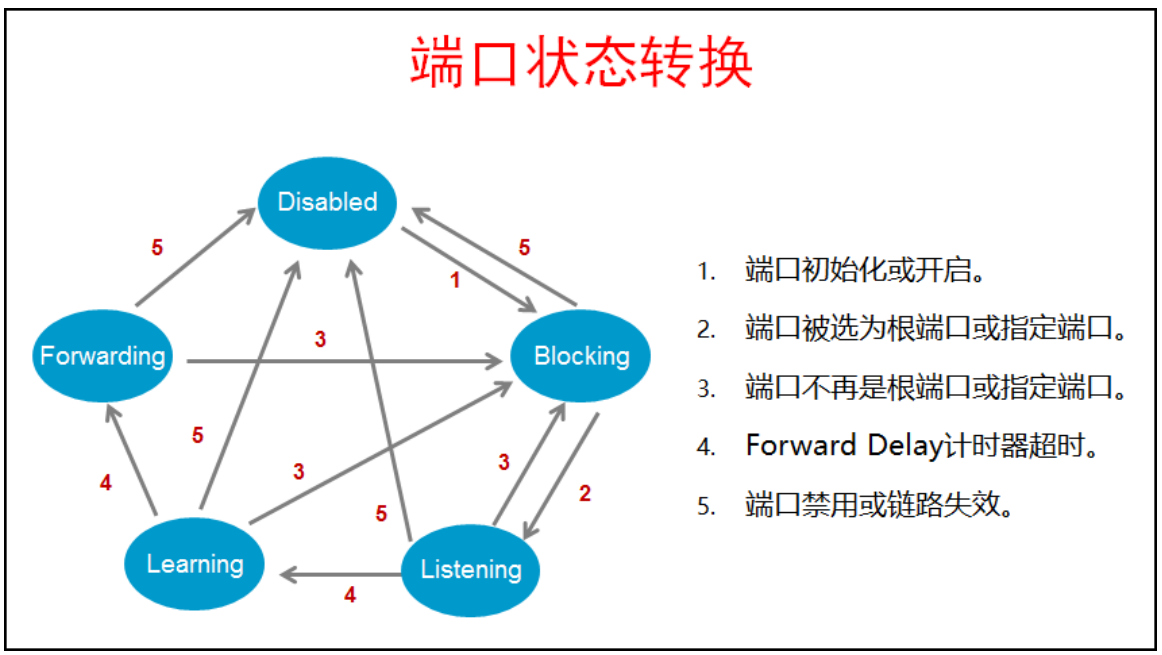
STP端口状态	接收BPDU	转发BPDU	学习MAC	转发数据	过渡状态	稳定状态
Disabled	×	×	×	×	×	√
Blocking	√	×	×	×	×	√
Listening	√	√	×	×	√	×
Learning	√	√	√	×	√	×
Forwarding	√	√	√	√	×	√

STP的计时器：

Hello	2s，根桥发送BPDU的间隔
Forward Delay	15s，监听和学习的持续时间
Max Age	20s，保持阻塞的最大时间（没有收到BPDU）

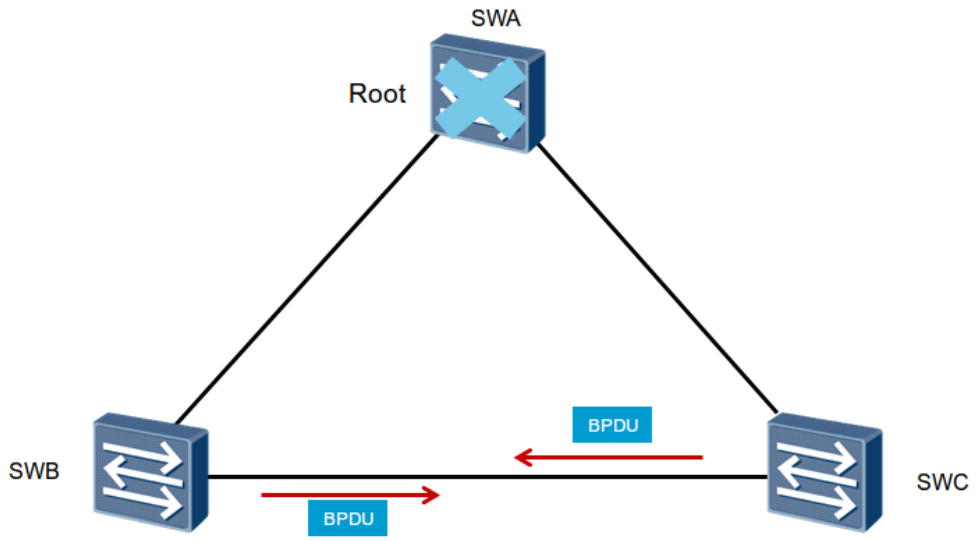


STP端口转换：



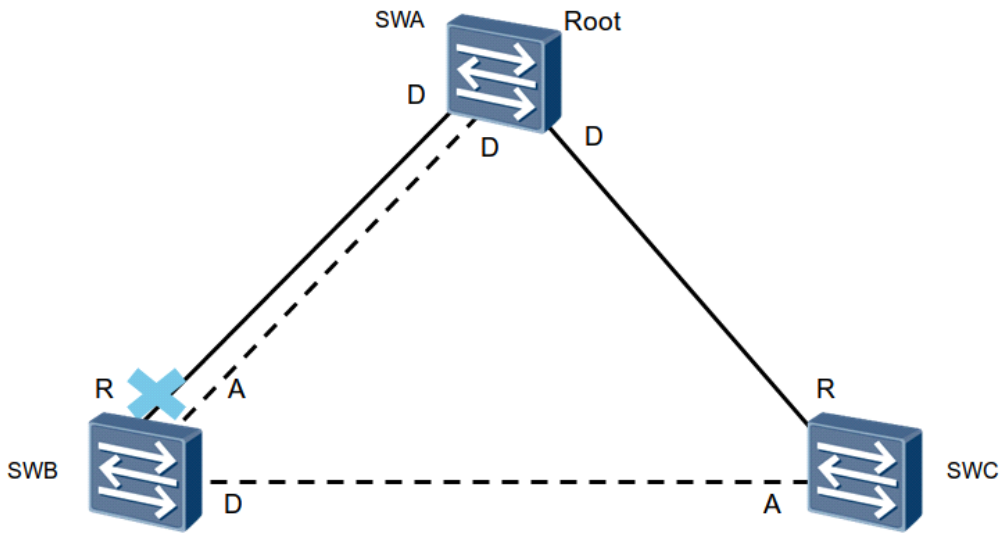
STP拓扑变化：

- 根桥故障



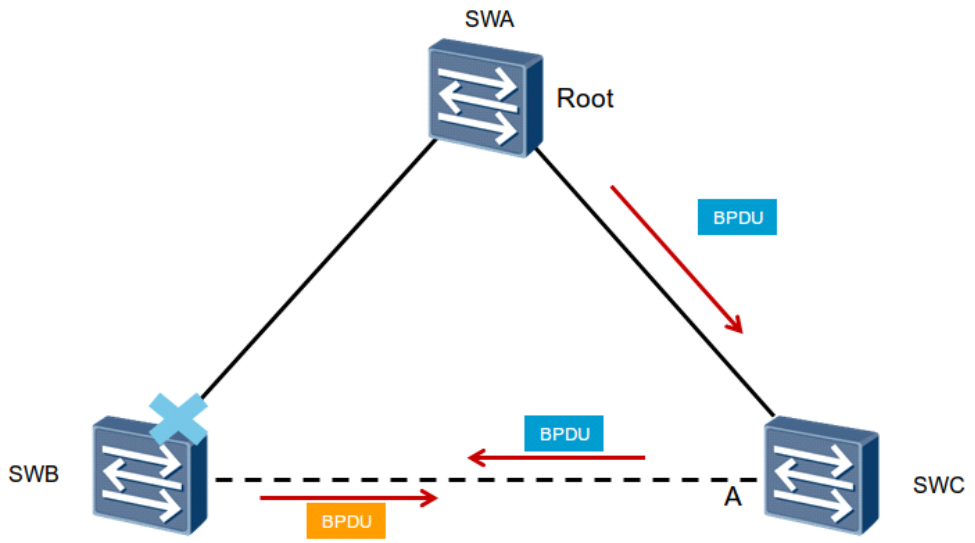
- 非根桥会在BPDU老化之后开始根桥的重新选举。

- 直连链路故障



- SWB检测到直连链路物理故障后，会将预备端口转换为根端口。
- SWB的预备端口会在30 秒后恢复到转发状态。

非直连链路故障

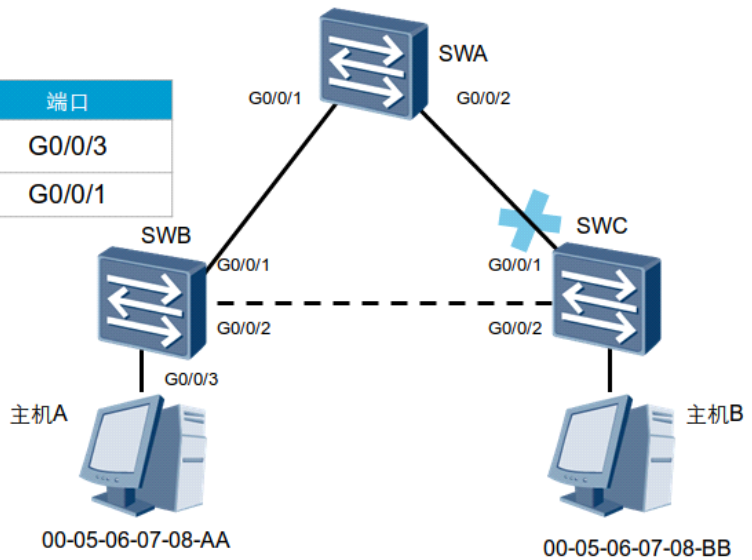


- 非直连链路故障后，SWC的预备端口恢复到转发状态大约需要50秒。

拓扑改变导致MAC地址表错误

MAC地址表

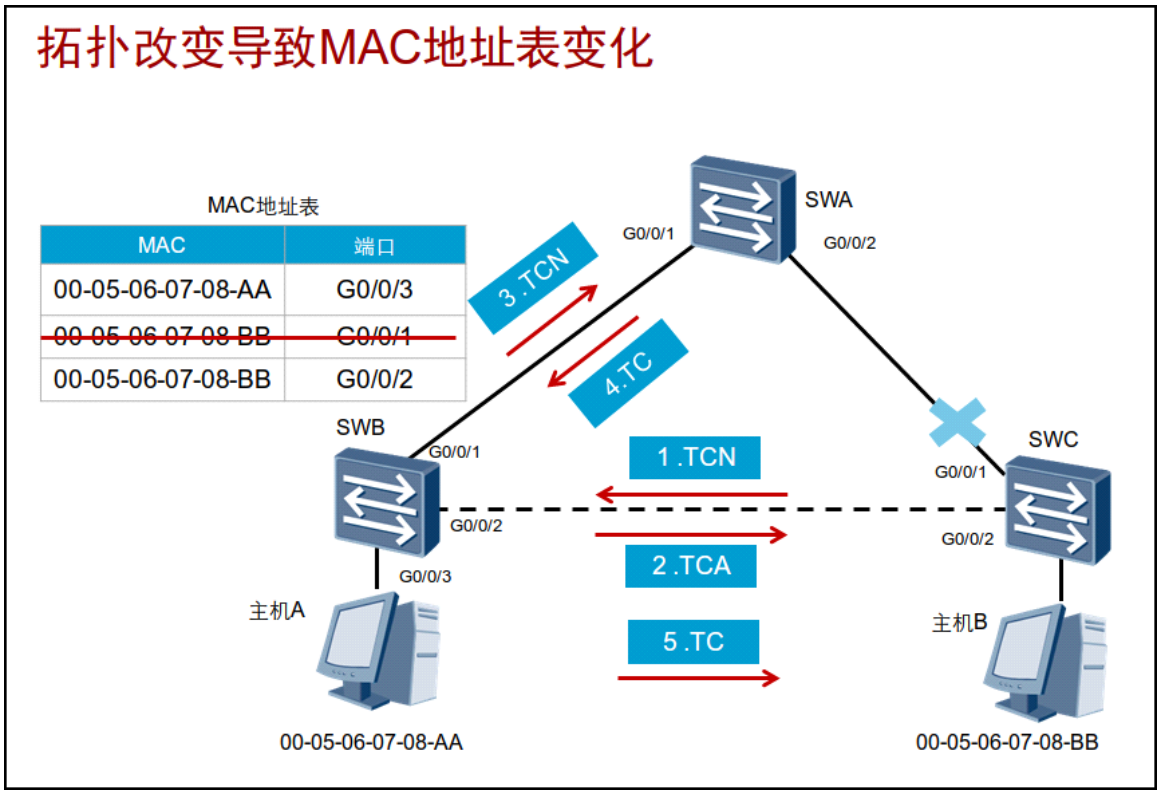
MAC	端口
00-05-06-07-08-AA	G0/0/3
00-05-06-07-08-BB	G0/0/1



- MAC地址表项的默认老化时间是300秒。在这段时间内，SWB无法将数据从G0/0/2端口转发给主机B。

拓扑改变导致MAC地址表变化

MAC地址表

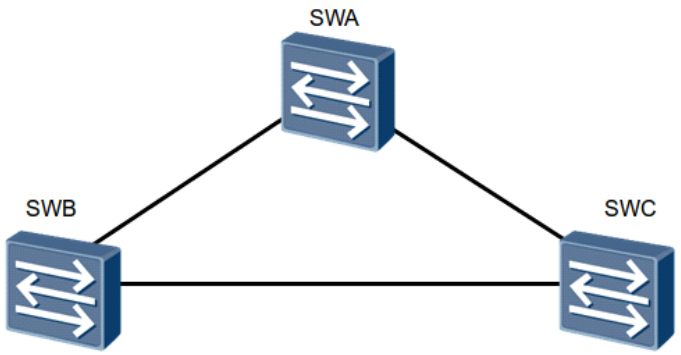
- 

- 检测到拓扑改变的交换机通过根端口向根桥发送TCN，上游交换机收到TCN后回应TCA，让后下游交换机停止发送TCN，再通过根端口发送TCN直到根桥收到，根桥通过指定端口发送TC通知所有下游交换机把MAC地址表记录老化时间从300秒变为15秒

STP配置：

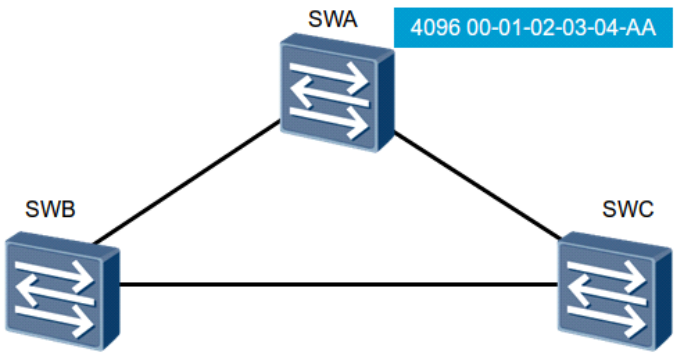
stp mode { mstp | stp | rstp }	配置STP模式，缺省为MSTP
stp priority 4096	配置优先级值，0~61440，步长为4096
stp root primary/secondary	自动修改优先级，指定主/备根桥
stp pathcost-standard { dot1d-1998 | dot1t | legacy }	配置路径开销值的标准
开销标准：	• legacy：cost=1～200000，华为的私有 • 802.1d标准：cost=1～65535 • 802.1t标准：cost=1～200000000，默认
stp cost	修改STP开销
display stp 【brief】	显示STP配置信息和参数

- ## STP模式




```
[SWA]stp mode ?
  mstp  Multiple Spanning Tree Protocol (MSTP) mode
  rstp  Rapid Spanning Tree Protocol (RSTP) mode
  stp   Spanning Tree Protocol (STP) mode
[SWA]stp mode stp
```

- ## 配置交换机优先级

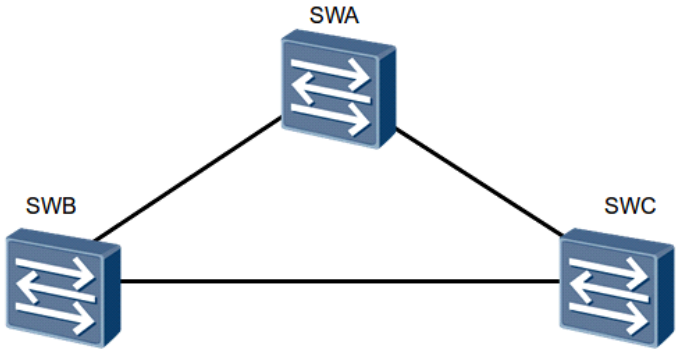



```
[SWA]stp priority 4096
Apr 15 2013 16:15:33-08:00 SWA DS/4/DATASYNC_CFGCHANGE:OID
1.3.6.1.4.1.2011.5.25.191.3.1 configurations have been
changed. The current change number is 4, the change loop
count is 0, and the maximum number of records is 4095.
```

  - 通过修改交换机的优先级，可以配置交换机为根交换机。

- 建议指定企业内配置高、性能好的交换机为根桥。

配置路径开销




```
[SWC]stp pathcost-standard ?
  dot1d-1998   IEEE 802.1D-1998
  dot1t        IEEE 802.1T
  legacy       Legacy
[SWC]interface GigabitEthernet 0/0/1
[SWC-GigabitEthernet0/0/1]stp cost 2000
```

配置验证

```
[SWA]display stp
-------[CIST Global Info][Mode STP]-------
CIST Bridge          :4096 .00-01-02-03-04-BB
Bridge Times         :Hello 2s MaxAge 20s FwDly 15s MaxHop 20
CIST Root/ERPC       :4096 .00-01-02-03-04-BB / 0
CIST RegRoot/IRPC    :4096 .00-01-02-03-04-BB / 0
CIST RootPortId      :0.0
BPDU-Protection      :Disabled
TC or TCN received   :37
TC count per hello   :0
STP Converge Mode    :Normal
Share region-configuration :Enabled
Time since last TC   :0 days 0h:1m:29s
……
```

```
[SWA]display stp
……
----[Port1(GigabitEthernet0/0/1)][FORWARDING]----
 Port Protocol             :Enabled
 Port Role                 :Designated Port
 Port Priority             :128
 Port Cost(Dot1T )         :Config=2000 / Active=2000
```

```
[SWA]display stp
……
----[Port1(GigabitEthernet0/0/1)][FORWARDING]----
 Port Protocol            :Enabled
 Port Role                :Designated Port
 Port Priority            :128
 Port Cost(Dot1T )        :Config=2000 / Active=2000
 Designated Bridge/Port   :4096.00-01-02-03-04-BB / 128.1
 Port Edged               :Config=default / Active=disabled
 Point-to-point           :Config=auto / Active=true
 Transit Limit            :147 packets/hello-time
 Protection Type          :None
……
```