Security Level:

生成树协议原理

ISSUE 1.0

www.huawei.com

前 言

- 生成树协议解决了交换网络可能会出现的网络风暴问题

学习指南

- 开篇通过讲述交换网络可能面临的问题引出生成树协议
- 重点理解生成树协议的工作机制和不同版本的生成树协议所解决的问题

参考资料

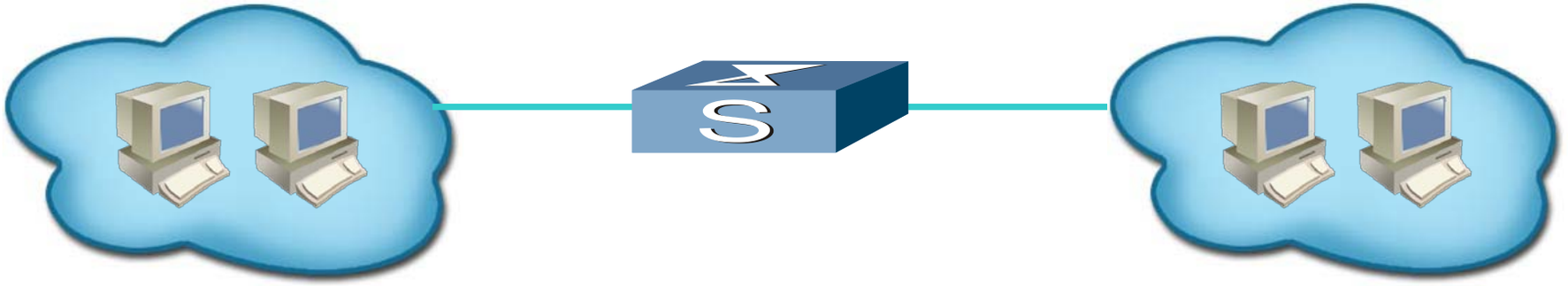
- IEEE 802.1d
- IEEE 802.1w

目 标

- 学习完此课程，您将会：
  - ⇨ 了解STP协议产生的背景
  - ⇨ 掌握STP工作原理
  - ⇨ 掌握RSTP工作原理

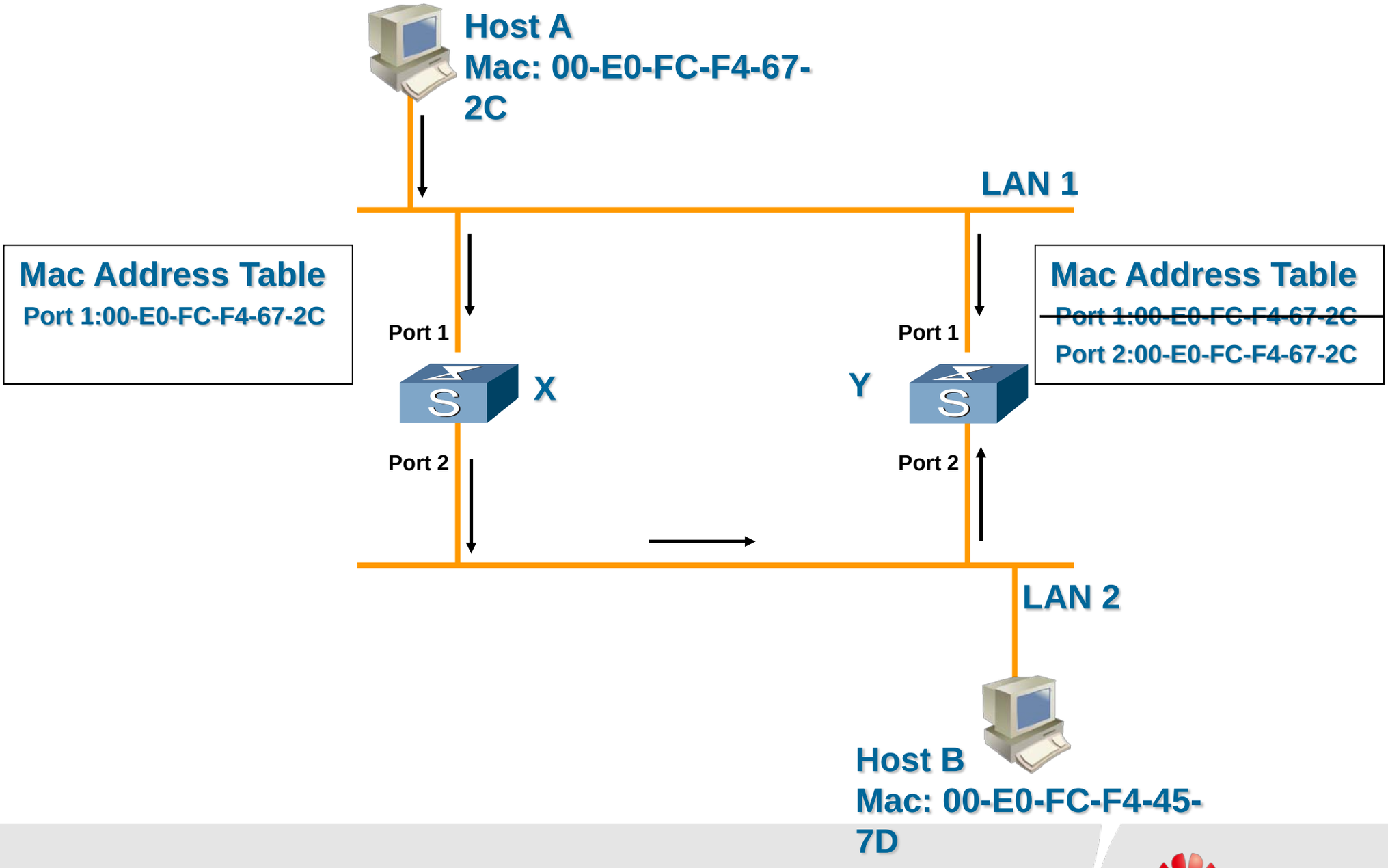
内容介绍

透明网桥的应用

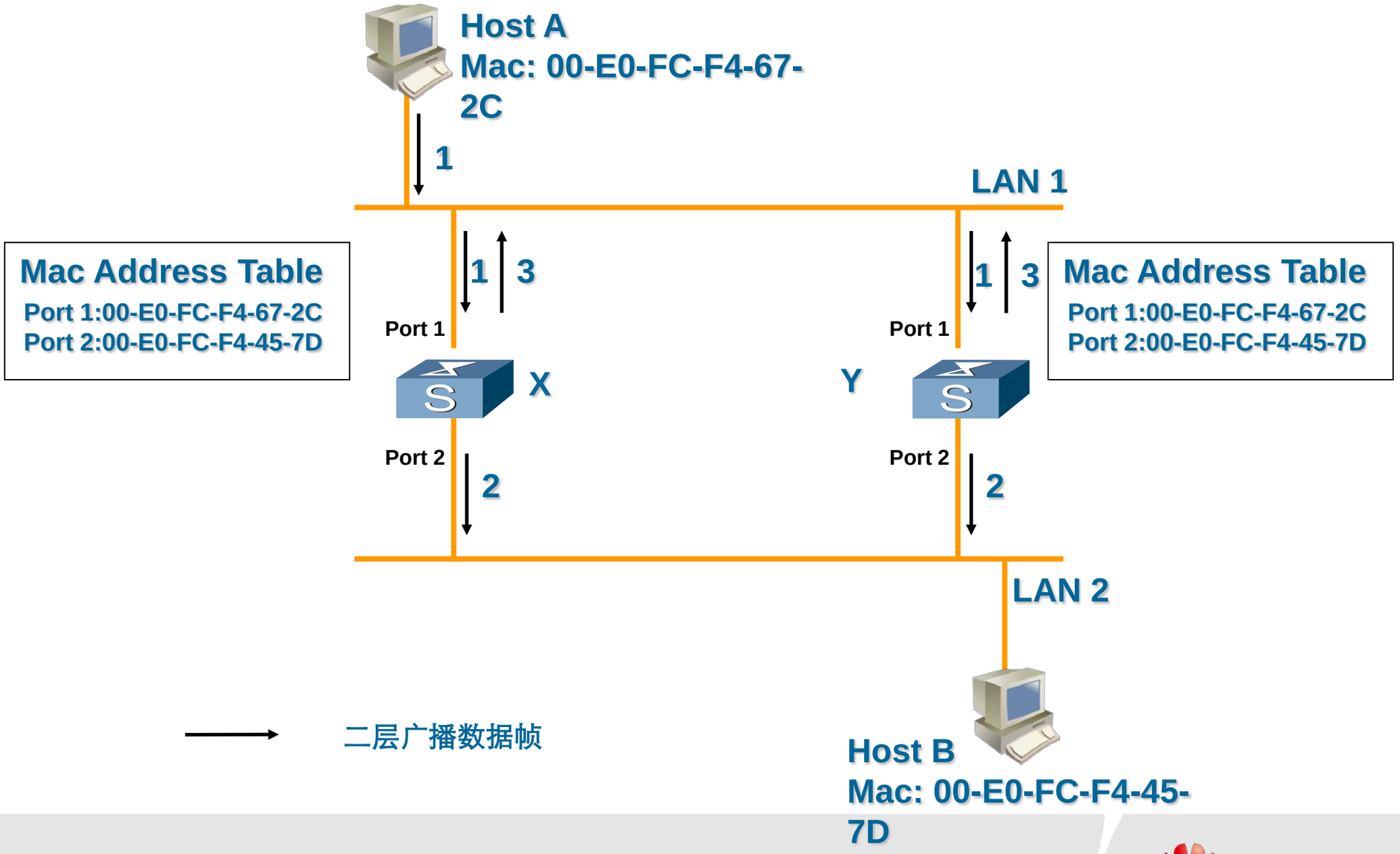


- 拓展LAN的能力。
- 自主动态学习站点的地址信息。
- 问题：一般的透明网桥不会对转发的报文做任何记号，这样，如果网络中存在回路，则有可能报文在回路中不断循环转发，造成网络拥塞。

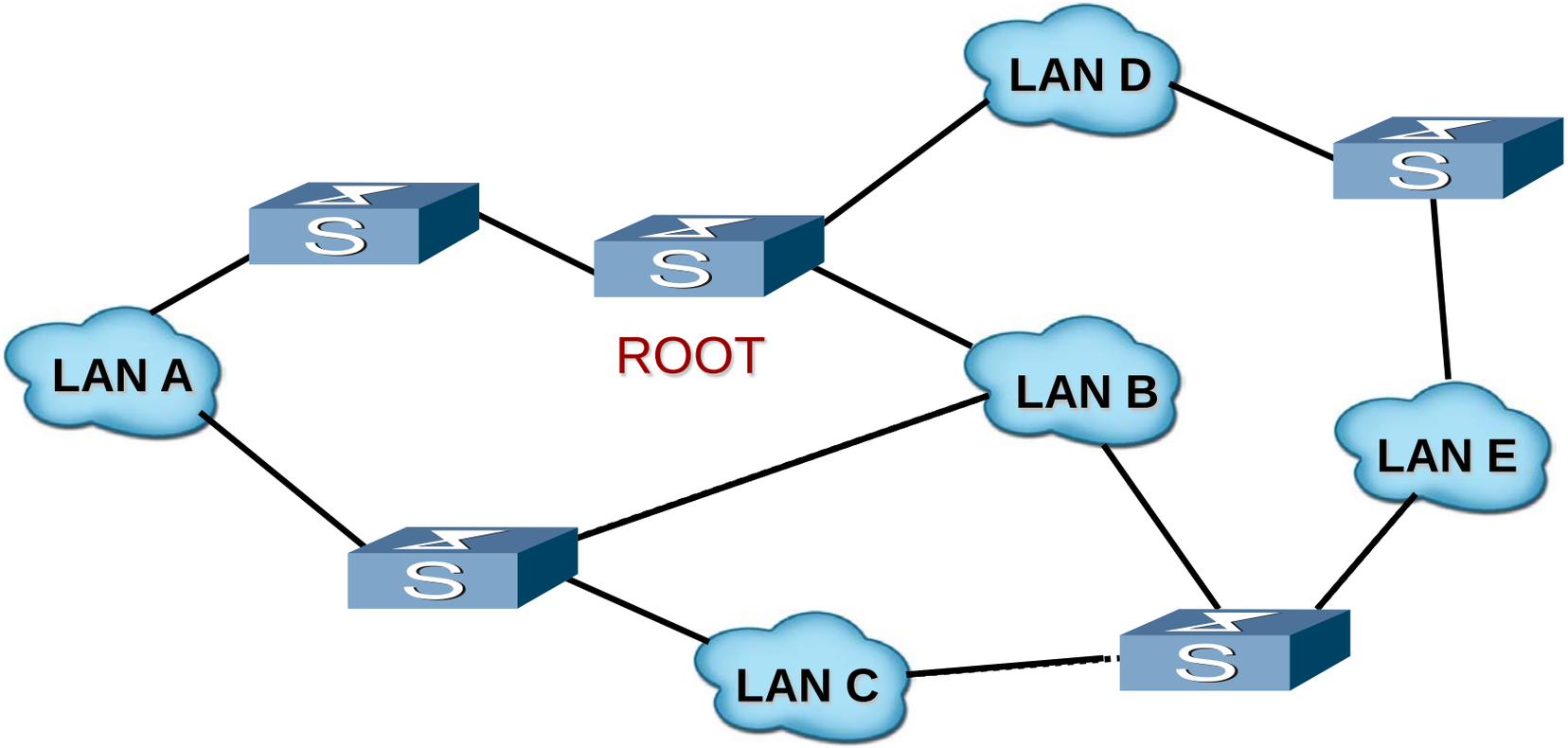
冗余链路产生的问题—Mac地址表不稳定



冗余链路产生的问题—广播风暴



为什么引入生成树协议



- 通过阻断冗余链路来消除桥接网络中可能存在的路径回环
- 当前活动路径发生故障时激活冗余备份链路恢复网络连通性

内容介绍

第1章 STP的产生原因

第2章 STP的基本原理

第3章 RSTP的基本原理

生成树协议的基本原理

- 基本思想：在网桥之间传递特殊的消息（配置消息），包含足够的信息做以下工作：
  - ⇨ 从网络中的所有网桥中，选出一个作为根网桥（Root）
  - ⇨ 计算本网桥到根网桥的最短路径
  - ⇨ 对每个LAN，选出离根桥最近的那个网桥作为指定网桥，负责所在LAN上的数据转发
  - ⇨ 网桥选择一个根端口，该端口给出的路径是此网桥到根桥的最佳路径
  - ⇨ 选择除根端口之外的包含于生成树上的端口（指定端口）

配置消息的内容

- 配置消息也被称作桥协议数据单元（BPDU）
- 主要内容包括
  - ⇨ 根网桥的Identifier（RootID）
  - ⇨ 从指定网桥到根网桥的最小路径开销（RootPathCost）
  - ⇨ 指定网桥的Identifier
  - ⇨ 指定网桥的指定端口的Identifier
  - ⇨ 即（RootID，RootPathCost，DesignatedBridgeID，DesignatedPortID）

配置消息格式

- DMA：目的MAC地址
  - ⇨ 配置消息的目的地址是一个固定的桥
  - ⇨ 的组播地址（0x0180c2000000）
- SMA：源MAC地址
  - ⇨ 即发送该配置消息的桥MAC地址
- L/T：帧长
- LLC Header：配置消息固定的链路头
- Payload：BPDU数据

值　域	占用字
协议ID	2
协议版本	1
BPDU类型	1
标志位	1
根桥ID	8
根路径开销	4
指定桥ID	8
指定端口ID	2
Message Age	2
Max Age	2
Hello Time	2
Forward Delay	2

配置消息格式

- 协议ID(2 字节)
  - ⇨ 当前保留没有被利用
- 协议版本(1 字节)
  - ⇨ 如果两大小不一的协议版本数字比较,则数字越大的将被认为最新定义的协议版本
- BPDU类型(1 字节)
  - ⇨ 类型域仅仅服务于区分BPDU的类型;在不同类型BPDU之间没有任何关系
- 标志位(1 字节)
  - ⇨ 被用来表示拓扑的变化,当拓扑发生变化时被置1,反之则置0
- 根桥ID(8 字节)
  - ⇨ 表示当前网络里的根桥,包括:
    - 网桥优先级 (2 字节)
    - 网桥的Mac地址 (6 字节)

配置端口开销

- 根路径开销(4 字节)
  - ⇨ 网桥到达根网桥的路径开销,数值大小可以由网桥自动配置或手动配置

参数	链路带宽	推荐值	推荐范围	范围
路径开销	4Mb/s	250	100-1000	1-65535
路径开销	10Mb/s	100	50-600	1-65535
路径开销	16Mb/s	62	40-400	1-65535
路径开销	100Mb/s	19	10-60	1-65535
路径开销	1Gb/s	4	3-10	1-65535
路径开销	10Gb/s	2	1-5	1-65535

配置消息格式

- 指定网桥ID(8 字节)
  - ⇨ 指发送BPDU的网桥,包括:
    - 网桥优先级 (2 字节)
    - 网桥的Mac地址 (6 字节)
- 指定端口ID(2 字节)
  - ⇨ 指发送BPDU的网桥端口,包括:
    - 端口优先级
    - 端口号

配置消息格式

- Message Age(2 字节)
  - ⇨ BPDU的有效存活时间
- Maximum Age(2 字节)
  - ⇨ BPDU的最大有效存活时间,默认为20秒
- Hello Time(2 字节)
  - ⇨ 周期发送BPDU的时间间隔,默认为2秒
- Forward Delay(2 字节)
  - ⇨ 端口转入发送状态的时延,默认为15秒

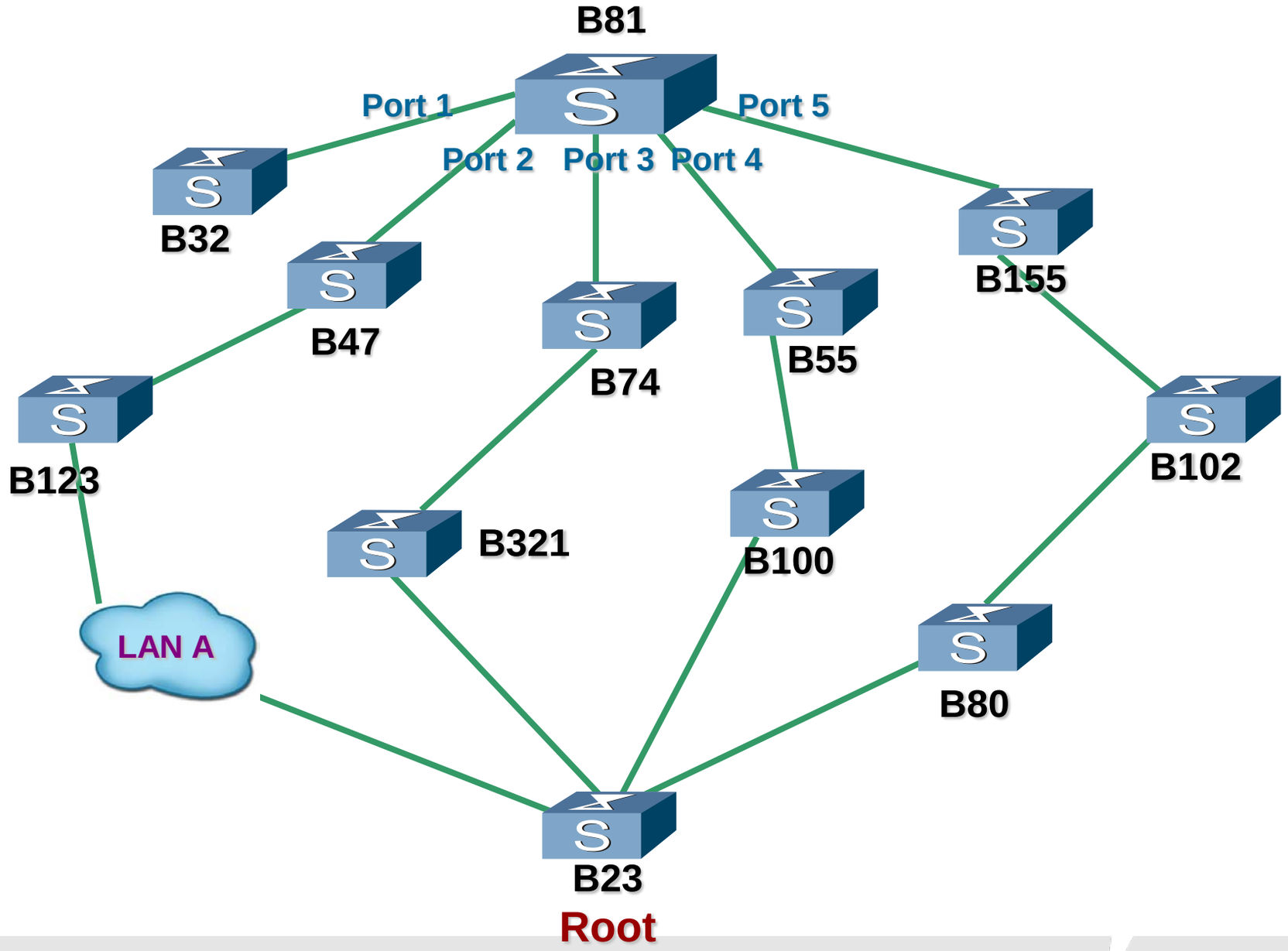
配置消息的处理

- 将各个端口收到的配置消息和自己的配置消息做比较，得出优先级最高的配置消息更新本身的配置消息，主要工作有：
  - ⇨ 选择根网桥RootID：最优配置消息的RootID
  - ⇨ 计算到根桥的最短路径开销RootPathCost：如果自己是根桥，则最短路径开销为0，否则为它所收到的最优配置消息的RootPathCost与收到该配置消息的端口开销之和
  - ⇨ 选择根端口RootPort：如果自己是根桥，则根端口为0，否则根端口为收到最优配置消息的那个端口
  - ⇨ 选择指定端口：包括在生成树上处于转发状态的其他端口
- 从指定端口发送新的配置消息

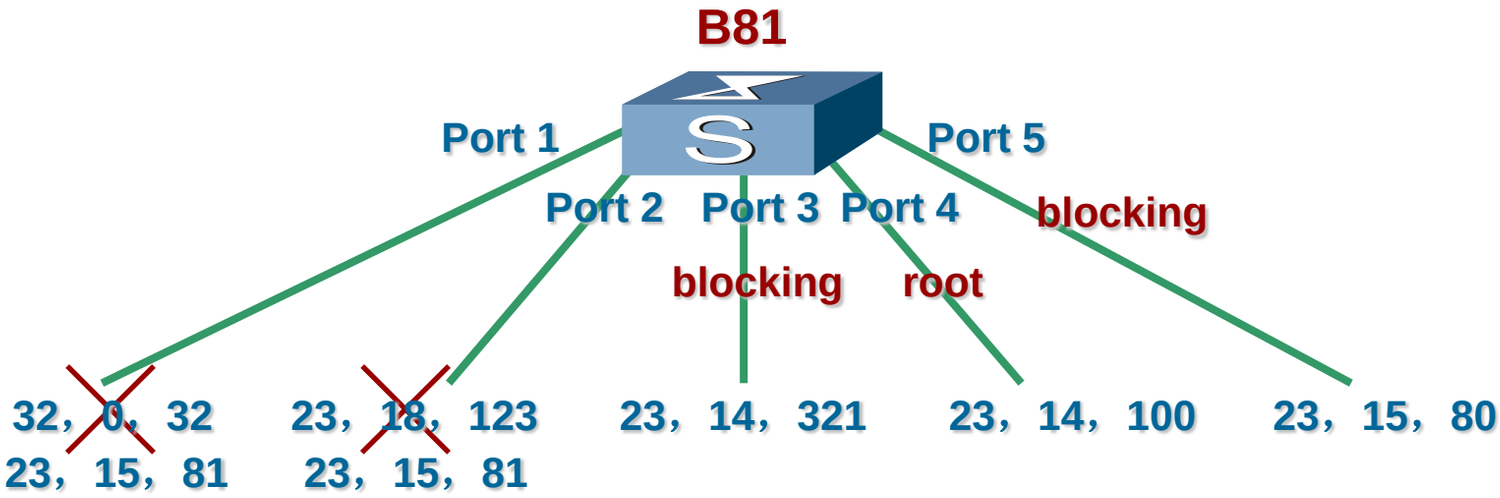
如何确定最优的配置消息

- 配置消息的优先级比较原则，假定有两条配置消息C1和C2，则：
  - ⇨ 如果C1的RootID小于C2的RootID，则C1优于C2
  - ⇨ 如果C1和C2的RootID相同，但C1的RootPathCost小于C2，则C1优于C2
  - ⇨ 如果C1和C2的RootID和RootPathCost相同，但C1的TransmitID小于C2，则C1优于C2
  - ⇨ 如果C1和C2的RootID、RootPathCost和TransimitId相同，但C1的PortID小于C2，则C1优于C2

一个接受并处理配置消息的例子



一个接受并处理配置消息的例子

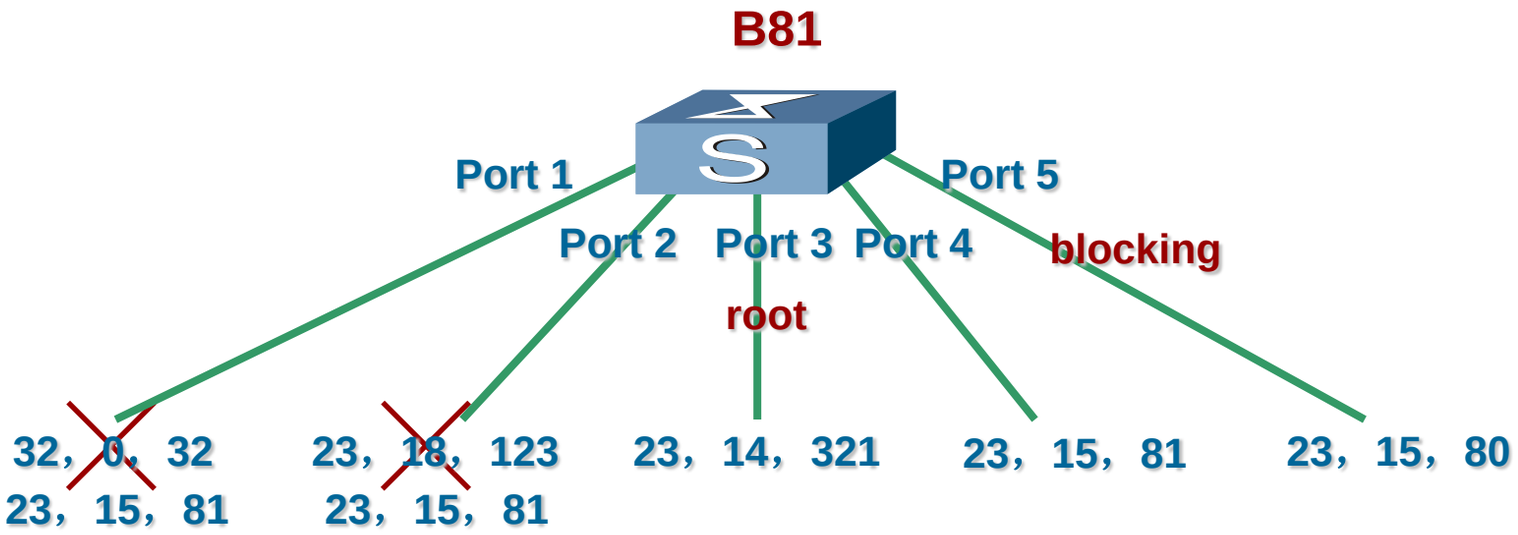


- 根据收到配置消息的优先级，选择Port4为根端口，选择Port1和Port2为指定端口，同时阻塞端口Port3和Port5。
- 从Port1和Port2发送新的配置消息：（23，15，81），其中，
  - ⇨ RootId = 23
  - ⇨ RootPathCost = 14+1 = 15
  - ⇨ RootPort = Port4

链路故障怎么办

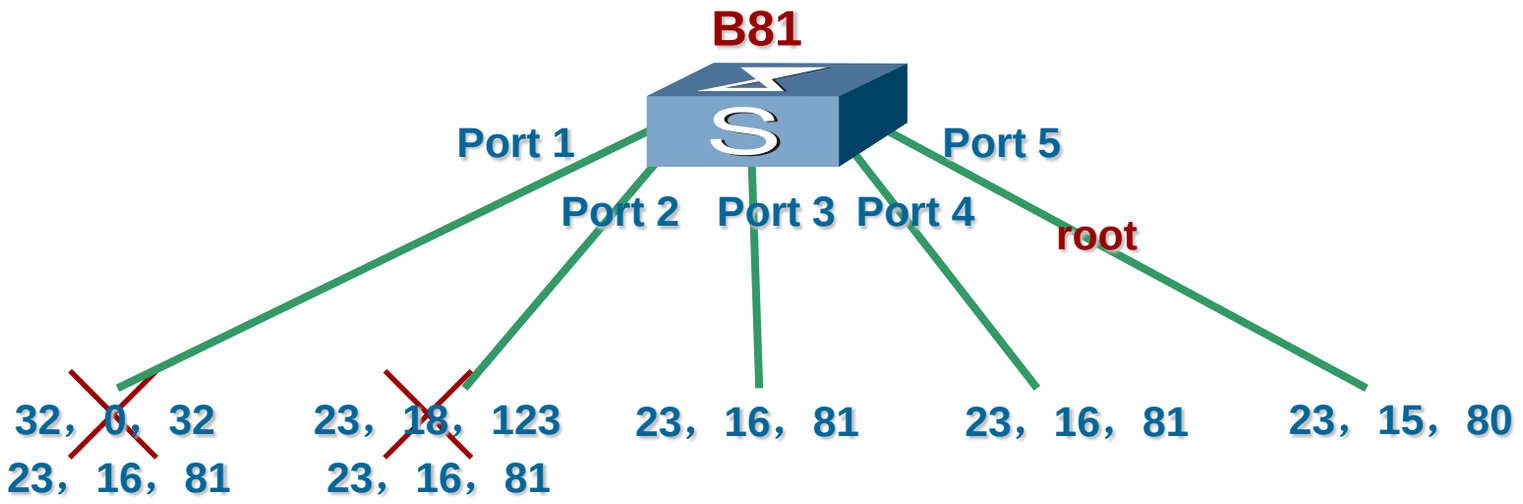
- Hello Time
  - ⇨ 网桥从指定端口以Hello Time为周期定时发送配置消息。
- Message Age和Max Age
  - ⇨ 端口保存的配置消息有一个生存期Message Age字段，并按时间递增。每当收到一个生存期更小的配置消息，则更新自己的配置消息。当一段时间未收到任何配置消息，生存期达到Max Age时，网桥则认为该端口连接的链路发生故障，进行故障的处理。

链路故障处理一



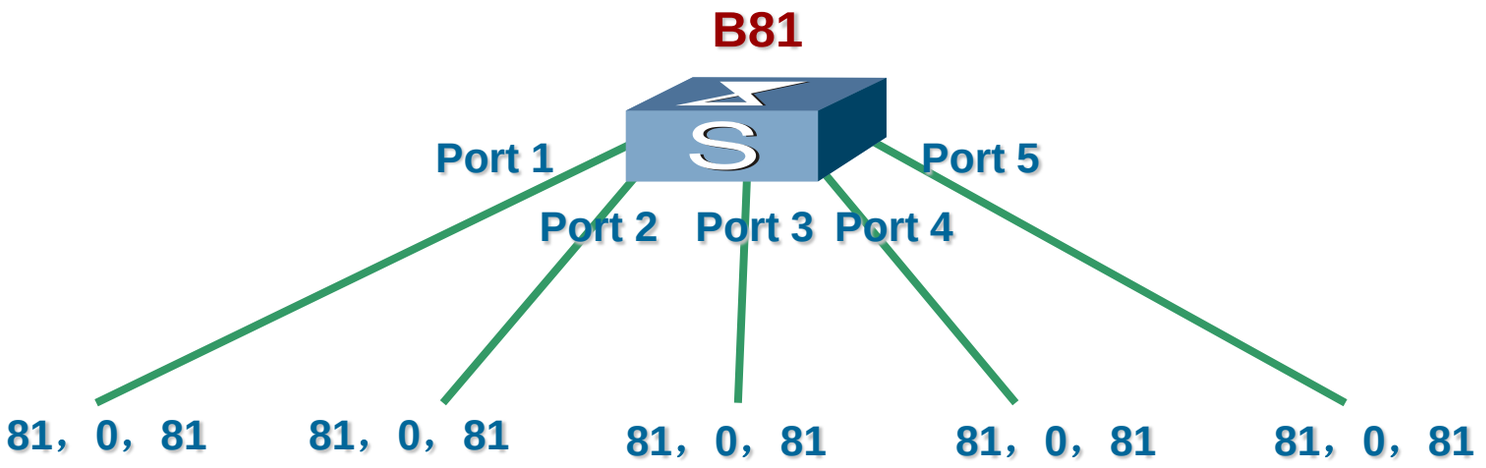
- Port4的配置消息生存期超时了，则抛弃该配置消息，重新进行生成树计算，选择Port3为新的根端口，而网桥81的配置消息没有变化

链路故障处理二



- Port3的配置消息生存期也超时了，则抛弃该配置消息，重新进行生成树计算，选择Port5为新的根端口，网桥81的配置消息变为（23，16，81）

链路故障处理三



- Port5的配置消息生存期也超时了，则抛弃该配置消息，以自己为根桥发送配置消息（81，0，81），直到从任一个端口收到优先级更高的配置消息

临时回路的问题

- 当拓扑结构发生变化，新的配置消息要经过一定的时延才能传播到整个网络，在所有网桥收到这个变化的消息之前：
  - ⇨ 若旧拓扑结构中处于转发的端口还没有发现自己应该在新的拓扑中停止转发，则可能存在临时的回环；
  - ⇨ 若旧的拓扑结构中阻塞的端口还没有发现自己应该在新的拓扑结构中开始转发，则可能造成网络暂时失去连通性。

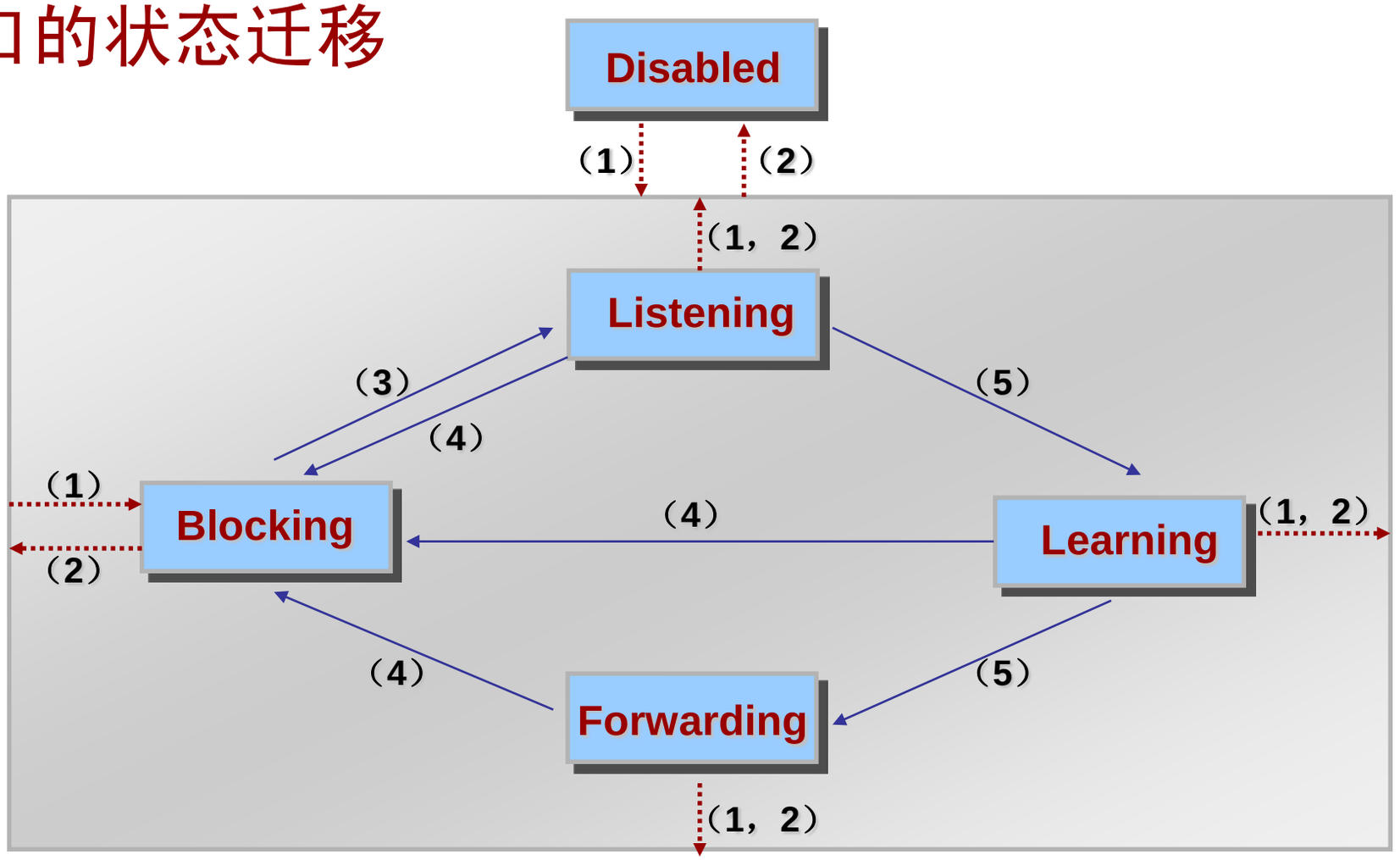
如何避免临时回路

- 端口由阻塞状态进入转发状态时，要经过一定时间的延时，这个时间起码是配置消息传播到整个网络所需最大时间的两倍。
- Forward Delay：配置消息传播到整个网络的最大时延
  - ⇨ 设计中间状态：处于中间状态的端口只是学习站点的地址信息，但不转发数据；
  - ⇨ 端口从阻塞状态经过Forward Delay的延时后进入中间状态；
  - ⇨ 再经过Forward Delay的延时后才能进入转发状态。

端口的几种状态

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BPDUs，不进行地址学习
Listening	不接收或转发数据，接收并发送BPDUs，不进行地址学习
Learning	不接收或转发数据，接收并发送BPDUs，开始地址学习
Forwarding	接收并转发数据，接收并发送BPDUs，进行地址学习

端口的状态迁移



1）端口enabled

2）端口disabled

3）端口被选为根端口或指定端口

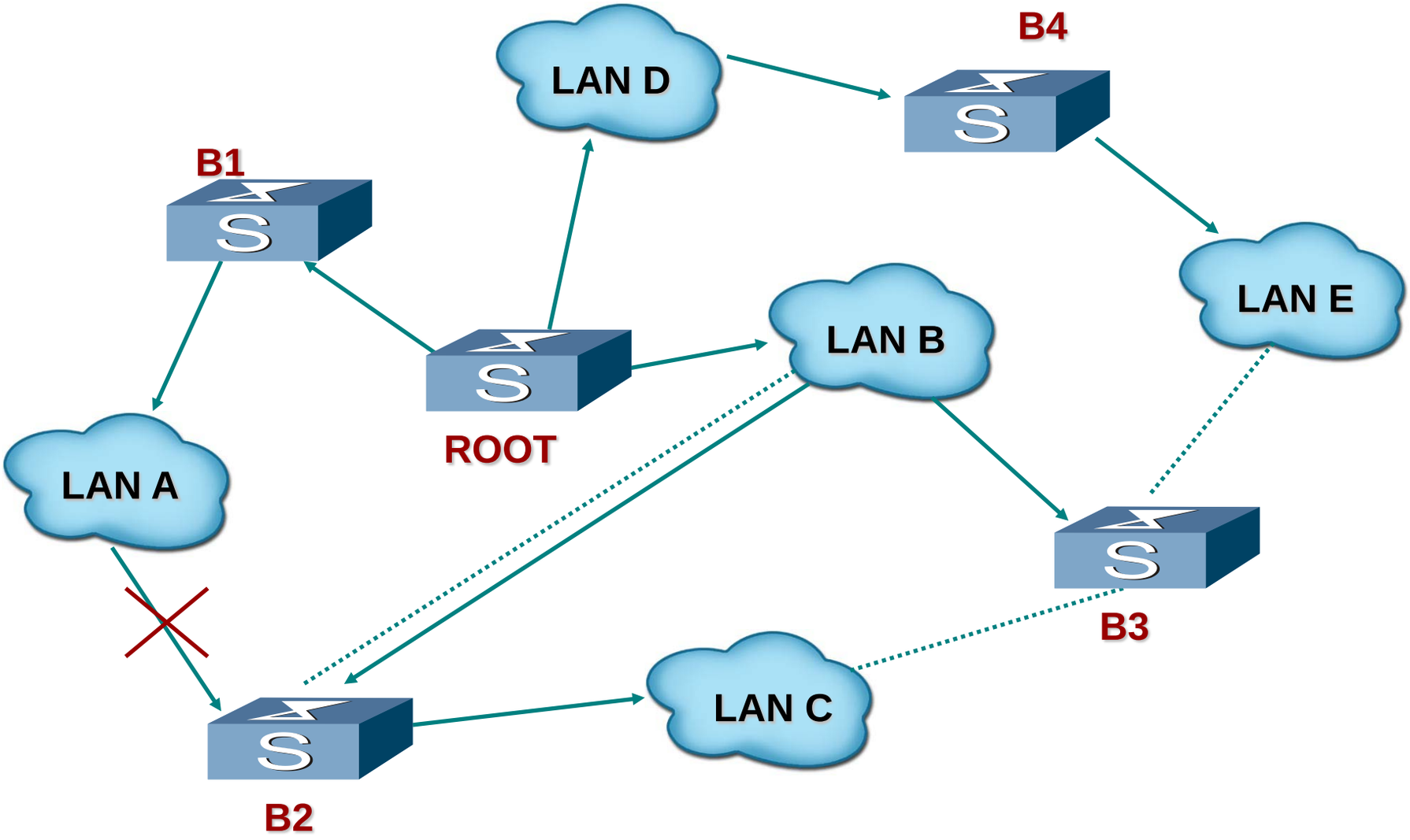
4）端口被选为备用端口（阻塞）

5）Forward Delay延时

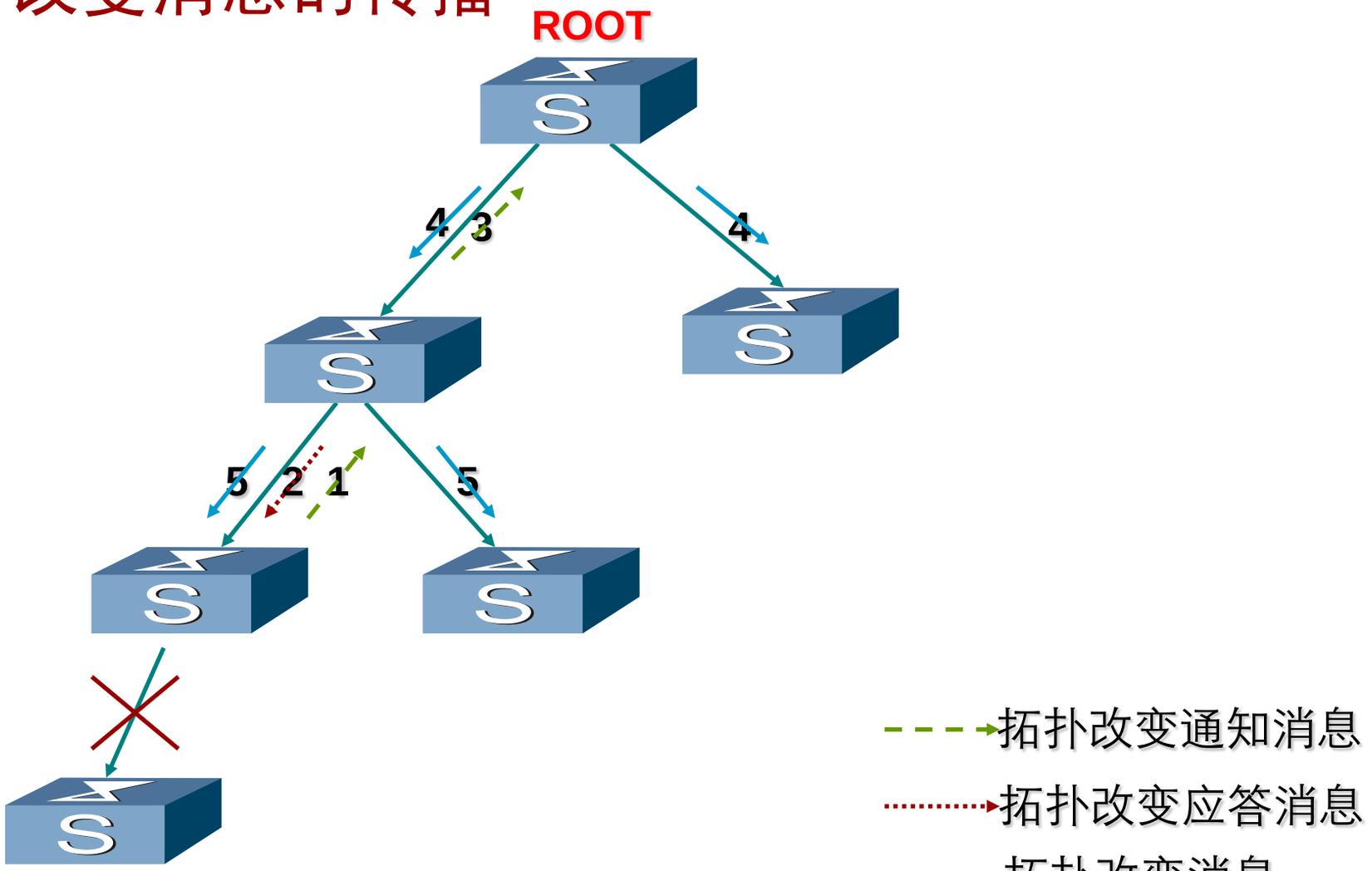
MAC地址信息的生存期

- 拓扑结构改变会使站点在生成树中的相对位置发生移动，那么网桥原来学习到的MAC地址信息就可能变得不正确，所以学习的MAC地址信息也要有生存期，如果该时间内没有证明地址的正确，则抛弃这条地址信息。
- 在生成树协议中有两个生存期：
  - ⇨ 拓扑稳定的时候用较长的生存期。
  - ⇨ 拓扑改变的时候用较短的生存期。
- 网络拓扑发生改变的时候，并 不是所有的网桥都 能够发现这一变化， 所以需要把拓扑改变的信息通知到整个网络。

站点的相对位置发生变化



拓扑改变消息的传播



内容介绍

第1章 STP的产生原因

第2章 STP的基本原理

第3章 RSTP的基本原理

生成树协议的不足

- 端口从阻塞状态进入转发状态必须经历两倍的Forward Delay时间，所以网络拓扑结构改变之后需要至少两倍的Forward Delay时间，才能恢复连通性。
- 如果网络中的拓朴结构变化频繁，网络会频繁的失去连通性，这样用户就会无法忍受。

快速生成树协议

- 快速生成树协议是从生成树协议发展而来，实现的基本思想一致；
- 快速生成树具备生成树的所有功能；
- 快速生成树改进目的就是当网络拓扑结构发生变化时，尽可能快的恢复网络的连通性。

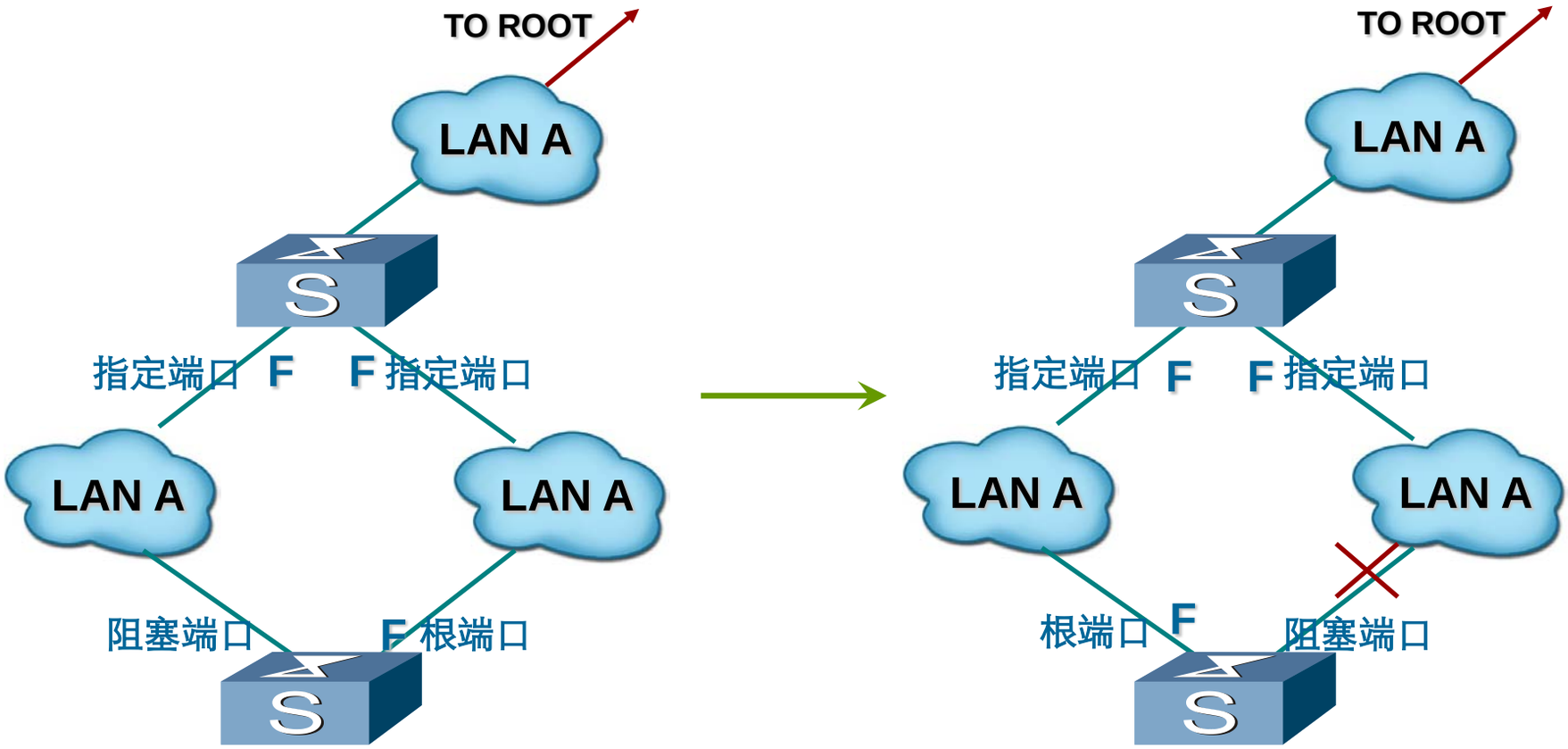
STP与RSTP端口的比较

STP端口类型	RSTP端口类型
Designated Port	Designated Port
Root Port	Root Port
Disabled Port	Disabled Port
	Alternate Port
	Backup Port

STP与RSTP状态机的比较

STP端口状态	RSTP端口状态
Disabled	Discarding
Blocking	Discarding
Listening	Discarding
Learning	Learning
Forwarding	Forwarding

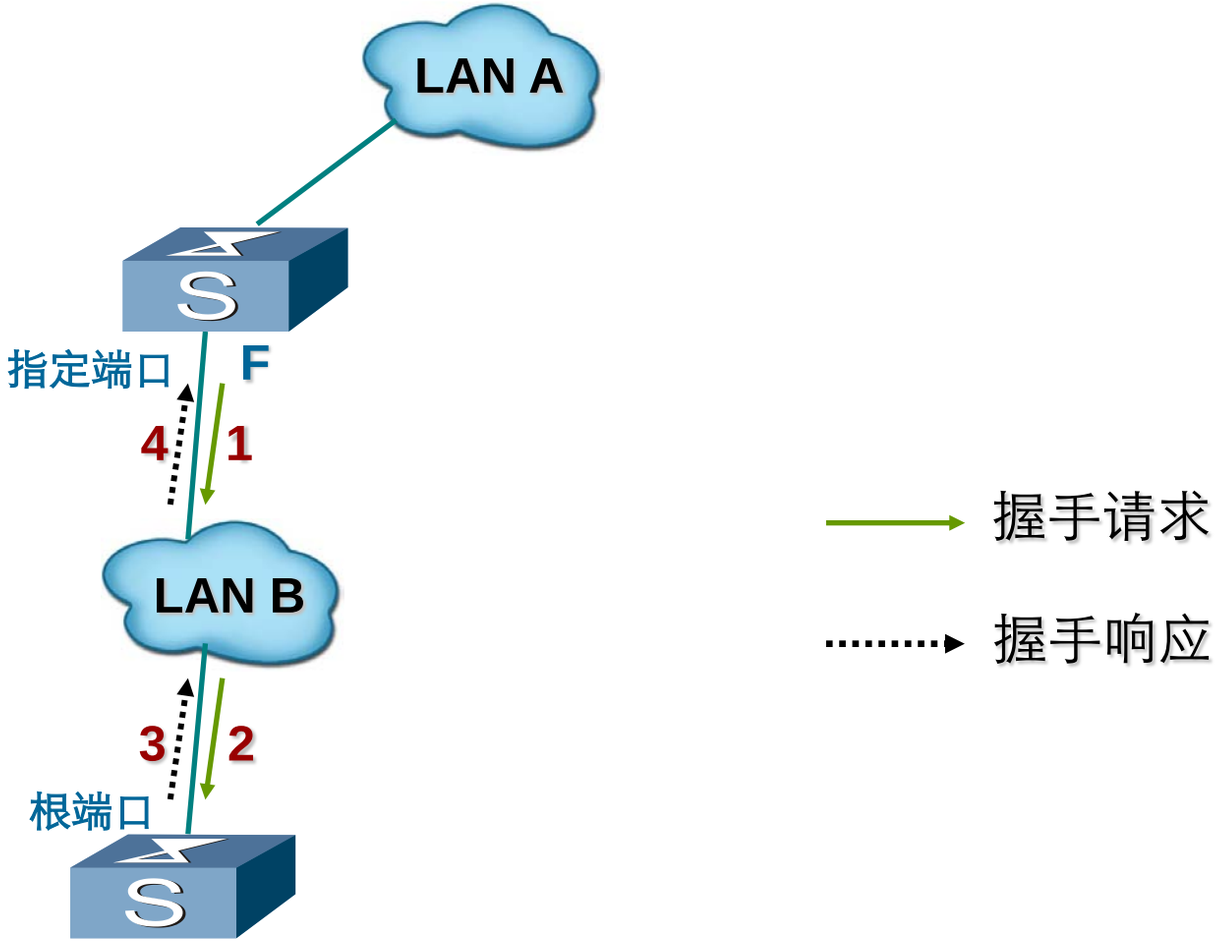
快速生成树的改进一



- 在新拓扑结构中的根端口可以立刻进入转发状态，如果旧的根端口已经进入阻塞状态，而且新根端口连接的对端交换机的指定端口处于Forwarding状态。

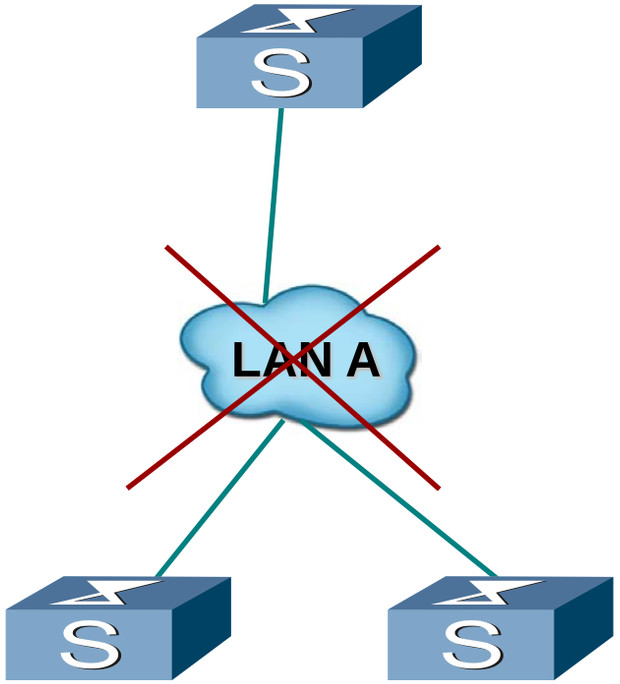
快速生成树的改进二

- 指定端口可以通过与相连的网桥进行一次握手，快速进入转发状态。

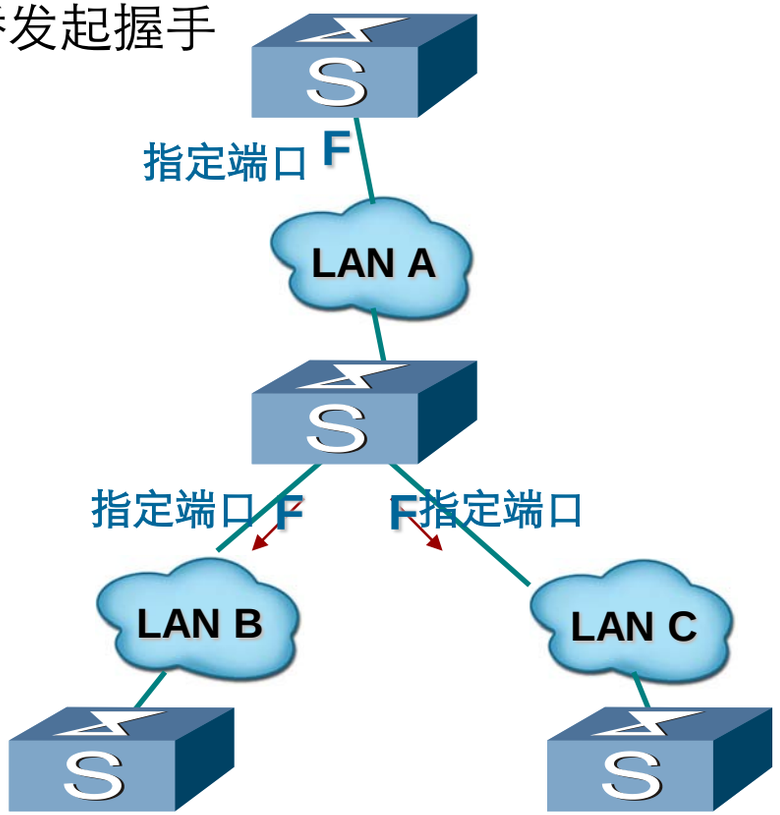


注意！

- 两点注意:
  - ⇨ 握手必须在点对点链路的条件下进行
  - ⇨ 一次握手之后，响应握手的网桥的非边缘指定端口将变为blocking状态，则需要继续向自己的邻接网桥发起握手



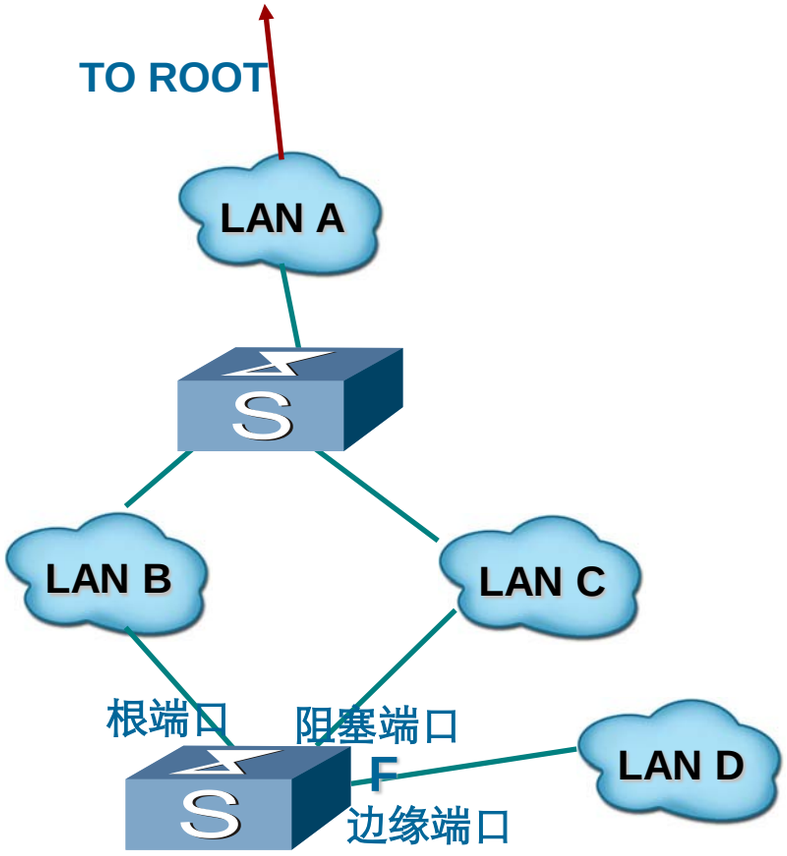
非点到点链路



握手的扩散

快速生成树的改进三

- 网络边缘的端口，即直接与终端相连，而不是和其他网桥相连的端口可以直接进入转发状态，不需要任何延时。



快速生成树的性能

- 第一种改进的效果：发现拓扑改变到恢复连通性的时间可达数毫秒，并且无需传递配置消息。
- 第二种改进的效果：网络连通性可以在交换两个配置消息的时间内恢复，即握手的延时；最坏的情况下，握手从网络的一边开始，扩散到网络的另一边缘的网桥，网络连通性才能恢复。比如当网络直径为7的时候，要经过6次握手。
- 第三种改进的效果：边缘端口的状态变化不影响网络连通性，也不会造成回路，所以进入转发状态无需延时。

生成树和快速生成树有何区别

- 协议版本不同；
- 端口状态转换方式不同；
- 配置消息报文格式不同；
- 拓扑改变消息的传播方式不同；

注意：快速生成树也是在整个交换网络应用单生成树实例，不能解决由于网络规模增大带来的性能降低问题。建议网络直径最好不要超过7。

问题

- 生成树协议解决了交换网络面临的什么问题?
- 生成树协议是如何工作的?
- RSTP相对STP做了那些改进?
- RSTP和STP有什么共同的问题需要解决?

总结

- 生成树协议解决了交换网络面临的网络风暴的问题
- RSTP再STP的基础上，加快了端口状态的迁移，提高了生成树的性能
- RSTP和STP维护的都是单生成树实例，它们如果与VLAN一起运行的话，会出现一些问题

谢谢

www.huawei.com