



冗余链路管理

本章重点围绕冗余链路管理展开相关论述。从阐述冗余链路产生的原因及所带来的问题入手,阐明生成树协议的作用场景,深入讲解生成树协议的基本原理,并通过实训剖析生成树的作用机制。然后,论述链路聚合技术的概念与作用,列举链路聚合的配置命令,并通过实训对链路聚合配置进行说明。通过对生成树协议和链路聚合技术的全面阐述及实训操作指导,为读者提供关于冗余链路管理的详细知识和实践方法。

5.1 生成树协议

5.1.1 冗余链路产生原因与问题

冗余链路是在网络或系统中存在的多余的、备份的链路连接。在计算机网络中,冗余链路的主要目的是提高网络的可靠性和可用性。当主链路出现故障时,冗余链路能够立即接替工作,确保数据传输不会中断,从而减少因链路故障导致的业务损失。

冗余链路可以应用在多种场景。例如,在一个企业的网络架构中,如果只有一条链路连接总部和分支机构,一旦这条链路出现问题,如线缆损坏、设备故障等,两地之间的通信就会中断。但如果存在冗余链路,即使主链路出现故障,数据仍可以通过备份的链路进行传输,保障业务的连续性。又如在数据中心与服务器之间的连接通常会设置冗余链路,以应对可能的硬件故障或网络拥塞情况。

下面通过一个示例来直观阐明冗余链路的应用。假设有一个网络,其网络拓扑如图 5-1 所示,主机 A 通过“交换机 1—交换机 2”的单链路与服务器 C 连接。如果这条链路上的任何一点出现故障(线缆故障或端口故障),主机 A 与服务器 C 的通信就会中断,即网络中的单点故障会致使网络无法访问。

冗余链路可以有效解决单点故障问题,如图 5-2 所示。另外架设一台交换机 3,“交换机 1—交换机 3—交换机 2”形成一条冗余链路,当“交换机 1—交换机 2”这条链路上出现故障时,主机 A 到服务器 C 的数据可经由“交换机 1—交换机 3—交换机 2”到达服务器 C。

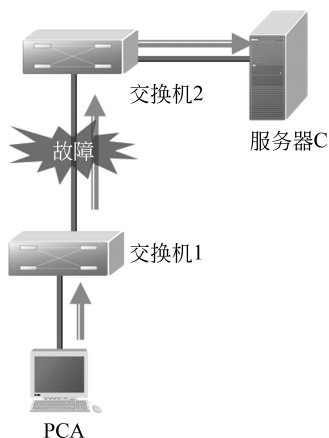


图 5-1 单链路网络拓扑

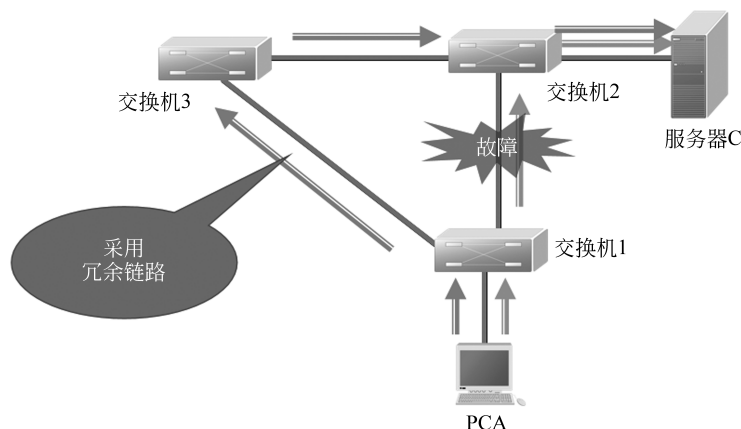


图 5-2 冗余链路网络拓扑

冗余链路在一定程度上增加了网络建设和维护的成本,但与因链路故障可能带来的巨大损失相比,其带来的稳定性和可靠性价值往往更为重要。

冗余链路完美地解决了单点故障问题,但是又引出了新问题。冗余链路会造成网络环路,如图 5-2 所示,“交换机 1—交换机 3—交换机 2—交换机 1”链路形成了一个环路,而环路的存在会引起一系列问题。

1. 广播风暴

通常交换机对网络中的广播帧不会进行任何数据过滤,交换机总是直接将这些信息广播到所有端口,如果网络中存在环路,这些广播帧将在网络中不停地转发,直到导致交换机出现超负荷运转,如 CPU 过度使用,内存耗尽,最终耗尽所有带宽资源、阻塞全网通信,这一过程被称为广播风暴,如图 5-3 所示。

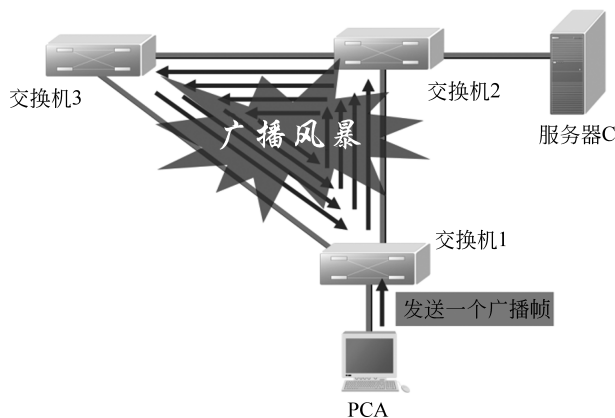


图 5-3 环路中的广播风暴

2. 重复非广播帧传输

广播帧在环路中发送会产生问题,单播帧同样也会产生问题。例如,如图 5-4 所示,主机 A 发送单播帧给服务器 C,如果交换机 1 上还没有服务器 C 的 MAC 地址表映射,就会同时向交换机 2、交换机 3 转发该帧,继而再同时交付给服务器 C,服务器同时会收到这个数据帧的多个副本,此时会导致上层协议在处理这些数据帧时无从选择,产生迷惑:究竟该处理哪个帧?严重时还可能导致网络连接的中断。

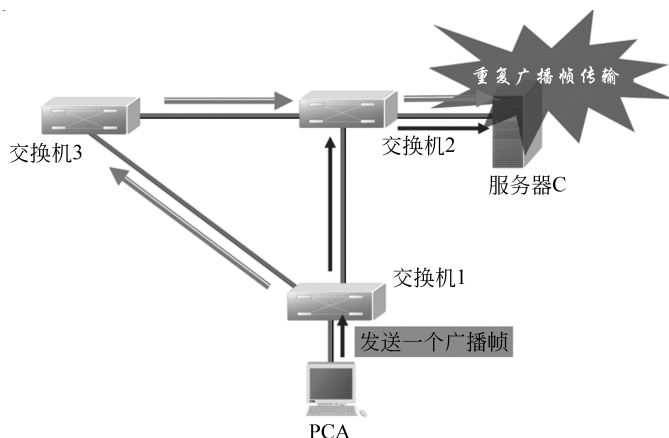


图 5-4 重复非广播帧传输

3. MAC 地址表不稳定性

当主机 A 经交换机 1、交换机 2 发送数据帧给服务器 C 时,交换机 2 会在 MAC 地址表中记录主机 A 的 MAC 地址与端口 1 的映射关系。而当数据经交换机 1、交换机 3、交换机 2 发送至服务器 C 时,在交换机 2 的 MAC 地址表中,主机 A 的 MAC 地址的对应端口又会变更为端口 0。也就是说,多个副本来自不同链路会导致交换机 MAC 地址表的多次刷新,如图 5-5 所示。这种持续的更新、刷新过程会严重耗费内存资源,影响该交换机的交换能力,同时降低整个网络的运行效率。严重时,将消耗掉整个网络资源,并最终造成网络瘫痪。

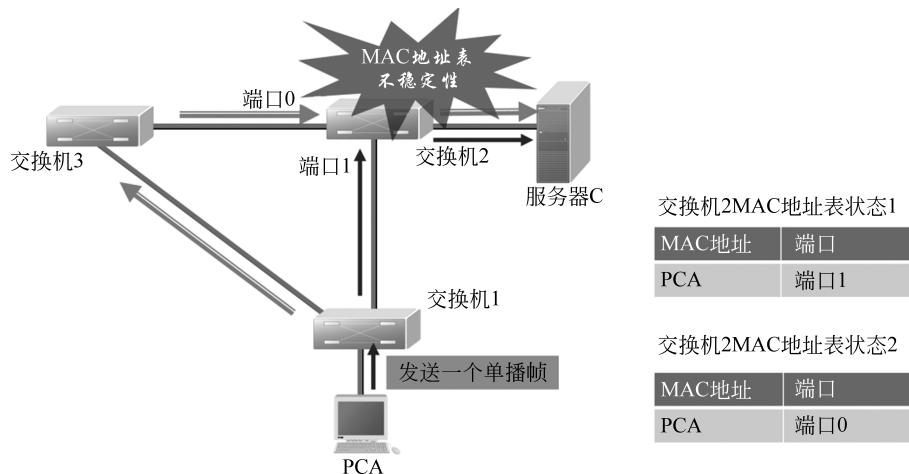


图 5-5 MAC 地址表不稳定性

通过上述例子可以看出,尽管备份链路带来了许多好处,但环路的出现同时也引发了诸多问题。

5.1.2 生成树协议基本原理

1. 网桥协议数据单元

生成树协议(Spanning Tree Protocol,STP)是 IEEE 802.1d 中定义的数据链路层协议,

主要用于解决在交换网络中冗余链路所产生的网络环路问题,生成树协议的存在,既解决了核心网络需要冗余链路的网络健壮性要求,又解决了因为冗余链路形成的物理环路导致的一系列问题。

在交换网络中通过生成树算法生成一个没有环路的网络,当主要链路出现故障时,能够自动切换到备份链路,保证网络的正常通信。要实现上述过程,交换机之间必须进行一些信息的交流,这个信息的最小数据单元称为网桥协议数据单元(Bridge Protocol Data Unit, BPDU)。

BPDU 属于数据链路层报文,其目的 MAC 地址为组播地址 01-80-C2-00-00-00。所有支持生成树协议的交换机,均会接收并处理接收到的 BPDU 报文。BPDU 在网络拓扑构建和维护中扮演着关键角色,主要用于根桥选举、端口角色确定,以及在拓扑结构发生变化时进行通知。在网络初始化阶段,各交换机尚未选举出根桥之前,都会积极发送 BPDU 报文参与根桥的竞选。而一旦完成根桥选举,此后便仅由根桥负责发送 BPDU 报文,以维持网络拓扑的稳定状态和相关信息的更新。在 STP 的运行机制下,根桥每隔 2s 便会发送一次 BPDU 报文。若非根桥连续三次未收到 BPDU 报文,则表明网络中可能发生了故障。

BPDU 由 12 个字段构成,如图 5-6 所示。

Protocol ID	协议ID: 为0
Version	协议版本号: 0-STP 1-RSTP 2-MSTP
Message Type	BPDU类型
Flags	标记域
Root ID	根桥BID: 优先级+MAC地址
Cost of Path	根路径开销
Bridge ID	发送者的BID
Port ID	发送端口的PID
Message Age	消息的老化时间
Maximum Time	BPDU最大寿命
Hello Time	发送间隔
Forward Delay	listening到learning的转换时间

图 5-6 BPDU 字段

(1) Protocol ID。协议 ID,一直为 0。

(2) Version。协议版本号,生成树协议有多个不同版本,值为 0 代表标准的生成树协议 STP,1 代表快速生成树协议 RSTP,2 代表多生成树协议 MSTP。

(3) Message Type。BPDU 类型。值为 00(十六进制)表示配置 BPDU,主要用于选举根桥及端口角色、维护端口状态、确认接收到的 TCN BPDU 报文、选举根桥及端口角色。值为 02 表示配置 TCN,即网络拓扑变化通知。正常情况下,交换机只会从它的根端口上接收配置 BPDU 包,但是绝不会主动发送配置 BPDU 包给根桥。当一台交换机检测到拓扑变化后,它就可以发送 TCN 给根桥,注意 TCN 是通过根端口向根桥方向发出的。

(4) Flags 标记域。第一个 bit 表示拓扑改变响应,最后一个 bit 表示拓扑改变。

(5) Root ID。根交换机桥 ID,由 2 字节优先级和 6 字节 MAC 组成。优先级数值范围为 0~61 440,默认值是 32 768,取值只能是 0 或 4096 的倍数,共 16 个,值越小优先级越高。

(6) Cost of Path。本交换机的根路径花费,路径开销是从本交换机到根桥的方向叠加的。

(7) Bridge ID。发送者的桥 ID,桥 ID 与根桥 ID 一样,由优先级与交换机 MAC 地址组成。

(8) Port ID。端口 ID,端口信息由 1 字节端口优先级和 1 字节端口 ID 组成。端口优先级为 0~240,默认 128,取值为 0 或 16 的倍数,共 16 个。

(9) Message Age。消息老化时间,表示 BPDU 在网络中传播所经过的时间。当 BPDU 在交换机之间传递时,Message Age 的值会不断增加。当一个交换机收到一个 BPDU 时,它会检查 Message Age 的值。如果这个值超过了网络中设定的最大老化时间,交换机可能会认为该 BPDU 已经过时或不可靠。

(10) Maximum Time。最大寿命,用于记录保留对方有效 BPDU 消息的最长时间,当一段时间未收到任何 BPDU,交换机认为该端口连接的链路发生故障,默认 20s。

(11) Hello Time。用于根桥定期发送 BPDU 的时间间隔,默认 2s。

(12) Forward Delay。转发延迟,通常是指网桥端口由 listening 到 learning 的转换时间,默认 15s。

2. 生成树协议工作过程

生成树协议的工作基于 BPDU 的交互,过程分为以下六步。

- (1) 选举根交换机。
- (2) 所有非根交换机选择一条到达根交换机的最短路径。
- (3) 所有非根交换机产生一个根端口。
- (4) 每个 LAN 确定指定端口。
- (5) 将所有根端口和指定端口设为转发状态。
- (6) 非根端口和指定端口设为阻塞状态。

通过这样的过程,网络逻辑结构从环变为树。如果网络拓扑发生改变,将重新计算生成树拓扑。每一步的实现过程如下所述。

1) 选举根交换机

选举的依据是交换机优先级和交换机 MAC 地址组合成的桥 ID,桥 ID 最小的交换机将成为网络中的根交换机。如图 5-7 所示的网络中,各交换机都以默认配置启动,默认优先级为 32768,在交换机优先级都一样的情况下,MAC 地址最小的交换机成为根交换机。交换机 A 的 MAC 地址为 0007.ec37.1d4b,小于交换机 B 的 00d0.bae2.35ed 和交换机 C 的 0060.3e01.4567,从而可以判定交换机 A 为根交换机。

2) 判断最短路径

其他交换机将各自选择一条“最粗壮”的树枝作为到根交换机的路径,也就是最短路径,如图 5-8 所示。判断最短路径的依据有三个,首先判断交换机到根桥的开销大小,开销最小的便成为最短路径,IEEE 802.1d 中规定了不同带宽的相应开销值,如带宽 10Mbps,则开销为 100;带宽 100Mbps,则开销为 19;带宽 1Gbps,则开销为 4;带宽 10Gbps,则开销为 2。假设三个交换机间的带宽皆为 1Gbps,那么,交换机 B 经 1 端口到 A 的开销为 4,经 3 端口、交换机 C 到交换机 A 的开销为 8,所以前者为最短路径。若开销相同,则选择对端网桥桥 ID 最小的路径。若桥 ID 还一样,再选择对端端口 ID 最小的路径。

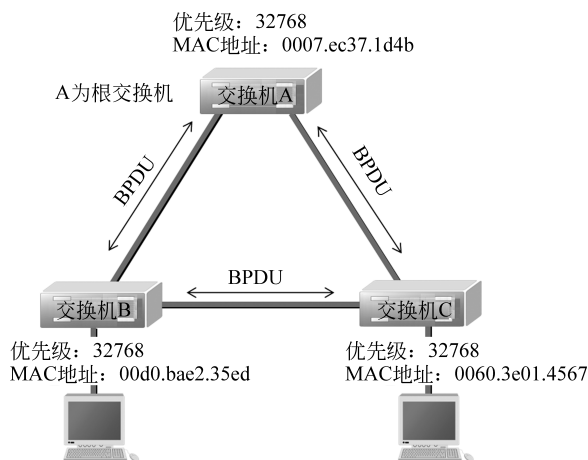


图 5-7 选举根交换机

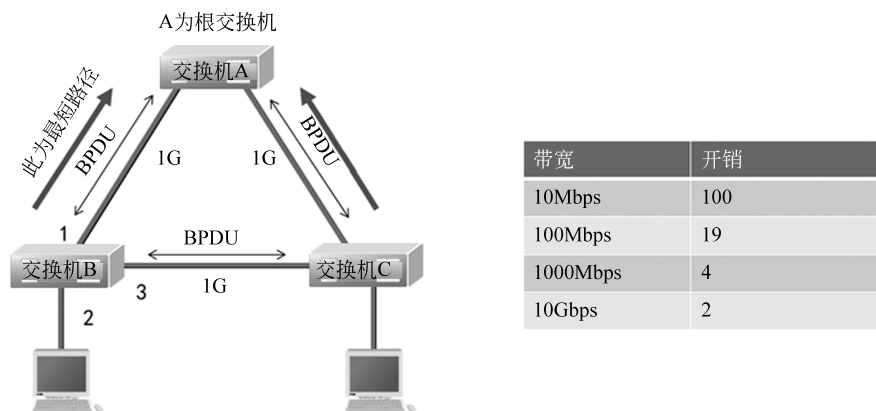


图 5-8 判断最短路径

3) 产生根端口

所有的非根交换机都会产生一个根端口，根端口需要在非根交换机的所有端口之中进行选择。实际上，当非根交换机计算出到根交换机的最短路径后，根端口自然就处于这条最短路径之上，从而实现根端口的定位。如图 5-9 所示，交换机 B 与交换机 C 连接交换机 A 的端口皆为根端口。

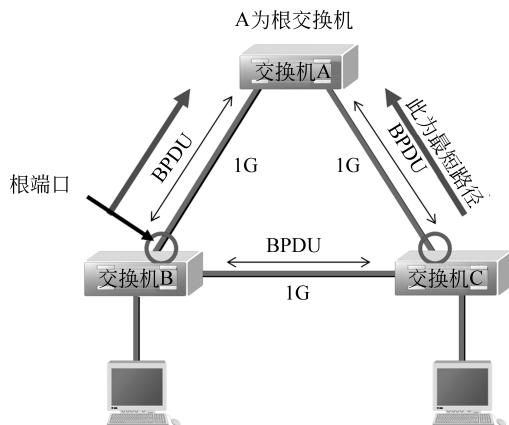


图 5-9 产生根端口

4) 每个 LAN 确定指定端口

除了每台交换机有一个根端口,每个网段还有一个用来到达根交换机的端口,该端口成为指定端口。路径成本是确定指定端口的重要考量因素,路径成本通常与链路的带宽、延迟等特性相关。例如,高带宽、低延迟的链路可能具有较低的路径成本,使其更有可能成为指定端口。在根交换机上,由于其每个活动端口到自身(即根交换机)的路径成本均为 0,所以根交换机上的每个活动端口都被认定为指定端口。当交换机直接连接网络终端时,这个连接端口也会被设置为指定端口。对于其他链路,指定端口的判断依据按优先级从高到低排序,依次为路径成本最低、端口所属交换机的网桥 ID 较小、直连端口 ID 较小。如图 5-10 所示,根交换机 A 的两个端口和交换机 B、交换机 C 的下联端口都是指定端口。

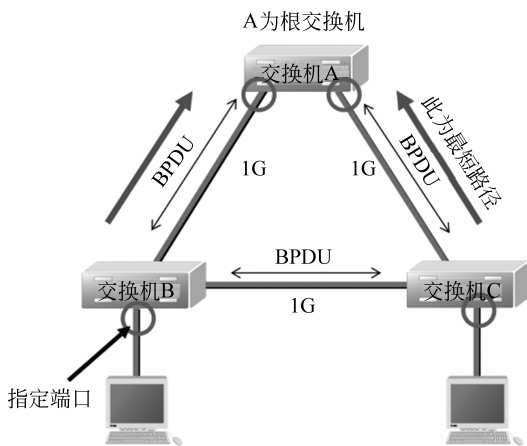


图 5-10 每个 LAN 确定指定端口

5) 将所有根端口和指定端口设为转发状态

在生成树协议中,根端口是连接非根交换机到根交换机的最短路径上的端口,而指定端口是每个网段中用于到达根交换机的端口。如图 5-11 所示,交换机所有根端口和指定端口设置为转发状态,这样做的目的是允许这些端口正常转发数据流量,以实现网络的连通性。

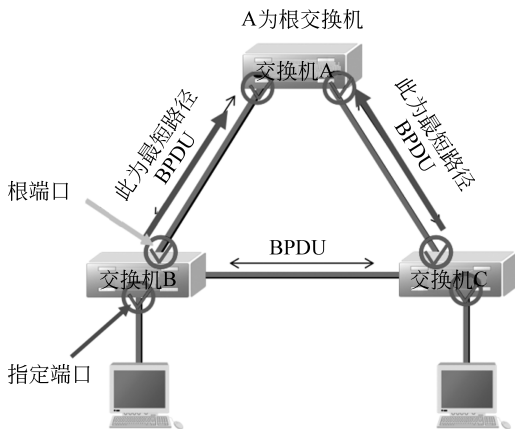


图 5-11 将所有根端口和指定端口设为转发状态

6) 非根端口和指定端口设为阻塞状态

如图 5-12 所示,交换机 B 和交换机 C 间的连接端口被设为阻塞状态,用来有效地打破

环路,保证网络的稳定性和可靠性。同时,当网络拓扑发生变化时,生成树协议会重新计算根端口和指定端口,并相应地调整端口的状态,以确保网络的最优性能。

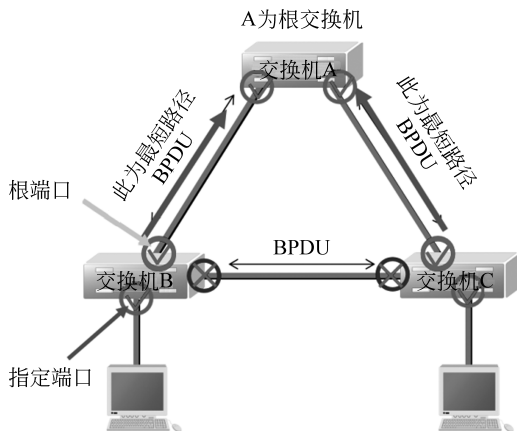


图 5-12 非根端口和指定端口设为阻塞状态

5.1.3 生成树协议配置命令

1. 开启生成树协议

```
Switch(config)#spanning-tree
```

该命令在交换机全局配置模式执行,用于开启生成树协议,进入生成树协议的配置子模式。为了防止环路,交换机一般默认开启生成树协议,如果没有开启,则采用 spanning-tree 命令开启生成树协议。

2. 关闭 spanning-tree 协议

```
Switch(config)#no spanning-tree
```

该命令在交换机全局配置模式执行,用于关闭交换机上的生成树协议。关闭生成树协议可能会导致网络中存在二层环路的风险。在执行此命令之前,务必确保充分了解网络拓扑结构,并且确认关闭生成树协议不会对网络的正常运行造成负面影响。对于一些关键业务或对网络稳定性要求较高的环境,应谨慎操作,避免因关闭生成树协议而导致不可预测的网络故障。

3. 配置生成树协议模式

```
Switch(config)#spanning-tree mode {stp|rstp|pvst}
```

该命令在全局配置模式下执行,用于设置生成树模式。5.1.2 节重点阐述了由 IEEE 802.1d 标准所定义的最早的生成树协议 STP。生成树协议如同其他协议一样,伴随着网络的持续发展而不断演进升级,RSTP 便是从 STP 发展而来并加以改进的快速生成树协议。而且,不同厂家所支持的生成树协议存在差异。例如,思科的 PVST 技术属于一种基于 VLAN 的生成树技术,不再像 STP 那样将整个交换网络视作一个生成树实例,而是把每个

VLAN 都当作一个生成树实例。

(1) spanning-tree mode stp 表示将交换机的生成树协议模式设置为传统的生成树协议(STP)。

(2) spanning-tree mode rstp 表示将模式设置为快速生成树协议(RSTP)。RSTP 相较于 STP,在收敛速度上有了显著的提升,能够更快地适应网络拓扑的变化。

(3) spanning-tree mode pvst 是将模式设置为每个 VLAN 生成树协议(PVST)。PVST 允许为每个 VLAN 独立地运行生成树协议,提供了更精细的网络控制和优化。如果一个企业网络中有多个 VLAN,并且对不同 VLAN 的流量和可靠性有不同的要求,那么使用 PVST 可以针对每个 VLAN 进行特定的配置和优化。

4. 配置交换机优先级

```
Switch(config)#spanning-tree(vlan 1)priority <0-61440>
```

优先级的高低是决定根桥的重要因素,优先级必须为 4096 的倍数。最小为 4096,最大为 61 440,默认优先级是 32 768。如要人为调整根桥,就需要通过该命令调整优先级,数字越小,优先级越高。思科设备以 VLAN 为单位建立生成树,所以命令中要加关键词 vlan VLAN 号。

在设置端口优先级时,需要根据网络拓扑和业务需求进行合理规划。例如,在一个有冗余链路的网络中,如果希望某条链路上的端口更有机会成为根端口,继而引导流量的走向,可以将该端口的优先级设置得较低。如果多个端口的优先级相同,还会基于其他因素(如端口 ID、路径开销等)来选择根端口。

5. 配置端口优先级

```
Switch(config-if)#spanning-tree(vlan 1)port-priority <0-240>
```

此命令在交换机的接口配置模式执行,用于针对 VLAN 1 来设置端口优先级。端口优先级用于在生成树协议计算中决定端口的角色。取值范围是 0~240,端口优先级是 16 的倍数,默认优先级 128,同样值越小优先级越高。例如,可以设置为 0、16、32 等,其中 0 优先级最高。

当网络中有多个端口处于相同的条件时,端口优先级将起到决定性作用。优先级高的端口更有可能成为指定端口或根端口。例如,在一个网络中,有两个交换机通过两个端口相连,一个端口的默认优先级为 128,另一个端口通过配置将优先级设置为 32,在生成树计算时,优先级为 32 的端口就更有可能被选为指定端口,从而影响数据的转发路径。

6. 显示交换机生成树的状态

```
Switch#show spanning-tree
```

该命令在特权模式下执行,用于在交换机上显示生成树协议的相关信息。通过执行该命令,可以获取以下重要信息。

(1) 生成树的类型(如 PVST、RSTP 等)。

(2) 每个 VLAN 中根桥的标识符(包括网桥优先级、MAC 地址等)。

- (3) 端口的状态(如转发、阻塞等)以及相关的端口开销。
- (4) 生成树的拓扑结构信息,包括每个端口连接的设备等。

7. 显示交换机接口的状态

```
Switch#show spanning-tree interface <interface>
```

此命令用于在交换机上显示特定接口的生成树协议相关详细信息。通过这条命令可以获取关于该接口在生成树协议中的以下具体内容。

- (1) 接口的角色(如根端口、指定端口或阻塞端口)。
- (2) 接口的状态(是否处于转发或阻塞状态)。
- (3) 接口的开销值。
- (4) 该接口接收到和发送的 BPDU 帧的统计信息。

当怀疑某个特定接口存在与生成树相关的问题时,使用此命令能够快速聚焦该接口的详细状态,以确定是否存在异常。同时,了解接口的开销值有助于评估该链路在生成树拓扑中的重要性以及对流量转发的影响。例如,如果发现接口本应是转发状态却处于阻塞状态,可能意味着生成树的计算出现偏差,或者该接口连接的链路存在问题。再如,通过查看接口的开销值,如果发现其过高,可能需要考虑调整网络拓扑或优化链路配置,以实现更高效地流量转发。

5.1.4 实训 生成树配置

实训任务

建立如图 5-13 所示的网络拓扑,观察生成树状态;改变影响生成树的参数,观察其变化。

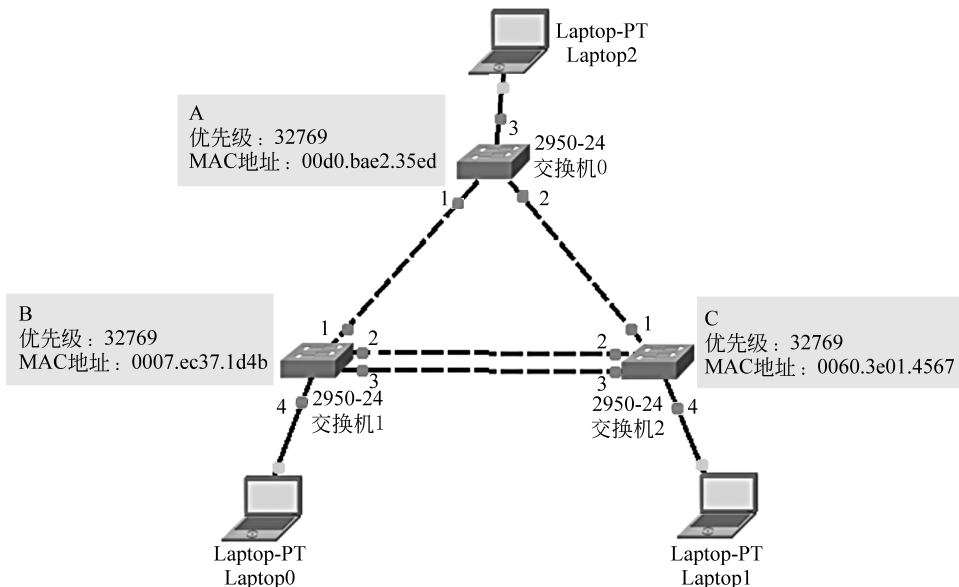


图 5-13 实训网络拓扑