

生成树协议技术

目 录

1 概述	1
1.1 产生背景	1
1.1.1 STP	1
1.1.2 RSTP	2
1.1.3 PVST	2
1.2 MSTP技术优点	3
2 MSTP技术实现	3
2.1 相关术语	3
2.2 MSTP算法实现	7
2.2.1 初始状态	7
2.2.2 端口角色的选择原则	7
2.2.3 优先级向量计算	7
2.2.4 角色选择过程	9
2.2.5 计算结果	11
3 Comware实现的技术特色	11
3.1 生成树协议的四种工作模式	11
3.2 Path Cost缺省值的计算	12
3.3 设置超时因子特性	12
3.4 指定根桥和备份根桥	12
3.5 BPDU保护	13
3.6 Root保护和角色限制	13
3.7 Loop保护和Dispute机制	13
3.8 TC保护	14
3.9 TC传播限制	14
3.10 配置摘要侦听	15
3.11 No Agreement Check特性实现	16
3.12 MSTP标准报文格式特性实现	17
3.13 VLAN Ignore特性	17
4 典型组网应用	17
5 参考文献	18

1 概述

1.1 产生背景

在二层交换网络中，一旦存在环路就会造成报文在环路内不断循环和增生，产生广播风暴，从而占用所有的有效带宽，使网络变得不可用。

在这种环境下生成树协议应运而生，生成树协议是一种二层管理协议，它通过有选择性地阻塞网络冗余链路来达到消除网络二层环路的目的，同时具备链路的备份功能。

生成树协议和其他协议一样，是随着网络的不断发展而不断更新换代的。最初被广泛应用的是 IEEE 802.1D-1998 STP（Spanning Tree Protocol，生成树协议），随后以它为基础产生了 IEEE 802.1w RSTP（Rapid Spanning Tree Protocol，快速生成树协议）、PVST（Per VLAN Spanning Tree，每 VLAN 生成树）协议和 IEEE 802.1s MSTP（Multiple Spanning Tree Protocol，多生成树协议）。

1.1.1 STP

STP 包含了两个含义，狭义的 STP 是指 IEEE 802.1D-1998 中定义的 STP 协议，广义的 STP 是指包括 IEEE 802.1D-1998 定义的 STP 协议以及各种在它的基础上经过改进的生成树协议。本文中的 STP 均指狭义的 STP。

STP 的基本思想十分简单。自然界中生长的树是不会出现环路的，如果网络也能够像一棵树一样生长就不会出现环路。于是，STP 中定义了根桥（Root Bridge）、根端口（Root Port）、指定端口（Designated Port）、路径开销（Path Cost）等概念，目的就在于通过构造一棵树的方法达到裁剪冗余环路的目的，同时实现链路备份和路径最优化。用于构造这棵树的算法称为生成树算法（Spanning Tree Algorithm）。

要实现这些功能，网桥之间必须要进行一些信息的交互，这些信息交互单元就称为 BPDU（Bridge Protocol Data Unit，桥协议数据单元），也称为配置消息。本文中将把生成树协议的协议报文均简称为 BPDU。STP BPDU 是一种二层报文，目的 MAC 是多播地址 01-80-C2-00-00-00，所有支持 STP 的网桥都会接收并处理收到的 BPDU 报文。该报文的数据区里携带了用于生成树计算的所有有用信息。

STP 的工作过程是：首先进行根桥的选举。选举的依据是网桥优先级和网桥 MAC 地址组合成的桥 ID，桥 ID 最小的网桥将成为网络中的根桥，它的所有端口都连接到下游桥，所以端口角色都成为指定端口。接下来，连接根桥的下游网桥将各自选择一条“最粗壮”的树枝作为到根桥的路径，相应端口的角色就成为根端口。循环这个过程到网络的边缘，指定端口和根端口确定之后一棵树就生成了。生成树经过一段时间（默认值是 30 秒左右）稳定之后，指定端口和根端口进入转发状态，其他端口进入阻塞状态。STP BPDU 会定时从各个网桥的指定端口发出，以维护链路的状态。如果网络拓扑发生变化，生成树就会重新计算，端口状态也会随之改变。这就是生成树的基本原理。

随着应用的深入和网络技术的发展，STP 的缺点在应用中也暴露了出来。STP 的缺陷主要表现在收敛速度上。

当拓扑发生变化，新的配置消息要经过一定的时延才能传播到整个网络，这个时延称为 Forward Delay，协议默认值是 15 秒。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，则可能存在临时环路。为了解决临时环路的问题，STP 使用了一种定时器策略，即在端口从阻塞状态到转发状态中间加上一个只学习 MAC 地址但不

参与转发的中间状态，两次状态切换的时间长度都是 **Forward Delay**，这样就可以保证在拓扑变化的时候不会产生临时环路。但是，这个看似良好的解决方案实际上带来的却是至少两倍 **Forward Delay** 的收敛时间，这在某些实时业务（如语音视频）中是不能接受的。

1.1.2 RSTP

为了解决 **STP** 的收敛速度缺陷，2001 年 **IEEE** 定义了基于 **IEEE 802.1w**（已经合入 **IEEE 802.1D-2004**）标准的快速生成树协议 **RSTP**。**RSTP** 在 **STP** 基础上做了三点重要改进，加快了收敛速度（最快可在 1 秒以内）：

- (1) 为根端口和指定端口设置了快速切换用的替换端口(**Alternate Port**)和备份端口(**Backup Port**)两种角色。当根端口失效的情况下，替换端口就会快速转换为新的根端口并无时延地进入转发状态；当指定端口失效的情况下，备份端口就会快速转换为新的指定端口并进入转发状态。
- (2) 在只连接了两个交换端口的点对点链路中，指定端口只需与下游网桥进行一次握手就可以无时延地进入转发状态。如果是连接了三个以上网桥的共享链路，下游网桥是不会响应上游指定端口发出的握手请求的，只能等待两倍 **Forward Delay** 时间进入转发状态。
- (3) 直接与终端相连而不与其他网桥相连的端口定义为边缘端口（**Edge Port**）。边缘端口可以直接进入转发状态，不需要任何延时。由于网桥无法知道端口是否是直接与终端相连，所以需要人工配置。

RSTP 相对于 **STP** 的确有很多改进，并且向下兼容 **STP**，可以混合组网。但是，**RSTP** 和 **STP** 一样同属于 **SST**（**Single Spanning Tree**，单生成树），有它自身的诸多缺陷，主要表现在三个方面：

- (1) 由于整个交换网络只有一棵生成树，在网络规模比较大的时候会导致较长的收敛时间。
- (2) 因为 **RSTP** 是单生成树协议，所有 **VLAN** 共享一棵生成树，为了保证 **VLAN** 内部可以正常通信，网络内每个 **VLAN** 都必须沿着生成树的路径方向连续分布，否则将会出现有的 **VLAN** 由于内部链路被阻塞而被分隔开，从而导致 **VLAN** 内部无法通信的问题。
- (3) 当某条链路被阻塞后将不承载任何流量，无法实现负载均衡，造成了带宽的极大浪费。

这些缺陷都是单生成树无法克服的，于是支持 **VLAN** 的 **PVST** 和 **MSTP** 出现了。

1.1.3 PVST

在每个 **VLAN** 中生成一棵树，是一种直接简单的解决 **RSTP** 缺陷的方法。基于此思想，**PVST** 协议应运而生，它能够保证每一个 **VLAN** 都不存在环路。

PVST 协议可以简单理解为在每个 **VLAN** 上运行一个 **RSTP**，不同 **VLAN** 之间的生成树完全独立。根据端口类型的不同，**PVST** 收发的 **BPD** 的格式也有所不同：

- 对于 **Access** 端口，**PVST** 协议收发标准的 **RSTP BPD**。
- 对于 **Trunk** 和 **Hybrid** 端口，在所有允许通过的 **VLAN** 内收发各自的 **PVST BPD**，同时在其中的 **VLAN 1** 内收发标准的 **RSTP BPD**。

PVST BPD 和 **RSTP BPD** 的内容相同，仅以太网的封装格式不同。

PVST 协议在 **VLAN** 间实现了数据的负载均衡，有效地提高了链路带宽的利用率，同时也带来了新问题，主要有以下缺点：

- (1) 由于每个 **VLAN** 都需要生成一棵树，**PVST BPD** 的通信量将正比于 **Trunk** 端口允许通过的 **VLAN** 数量。

- (2) 在 VLAN 数量比较多时，维护多棵生成树的计算量和资源占用量将急剧增长。特别是当允许通过很多 VLAN 的 Trunk 端口状态变化时，所有生成树的状态都要重新计算，网络设备的 CPU 将不堪重负。

上述缺点限制了运行 PVST 的 VLAN 端口数量，对 PVST 的部署产生较大影响，这些缺点在 MSTP 中得到了解决。

1.2 MSTP技术优点

MSTP 是 IEEE 802.1s（已经合入 IEEE 802.1Q-2011）中定义的一种新型生成树协议，相对于之前的生成树协议，优势非常明显。MSTP 的特点如下：

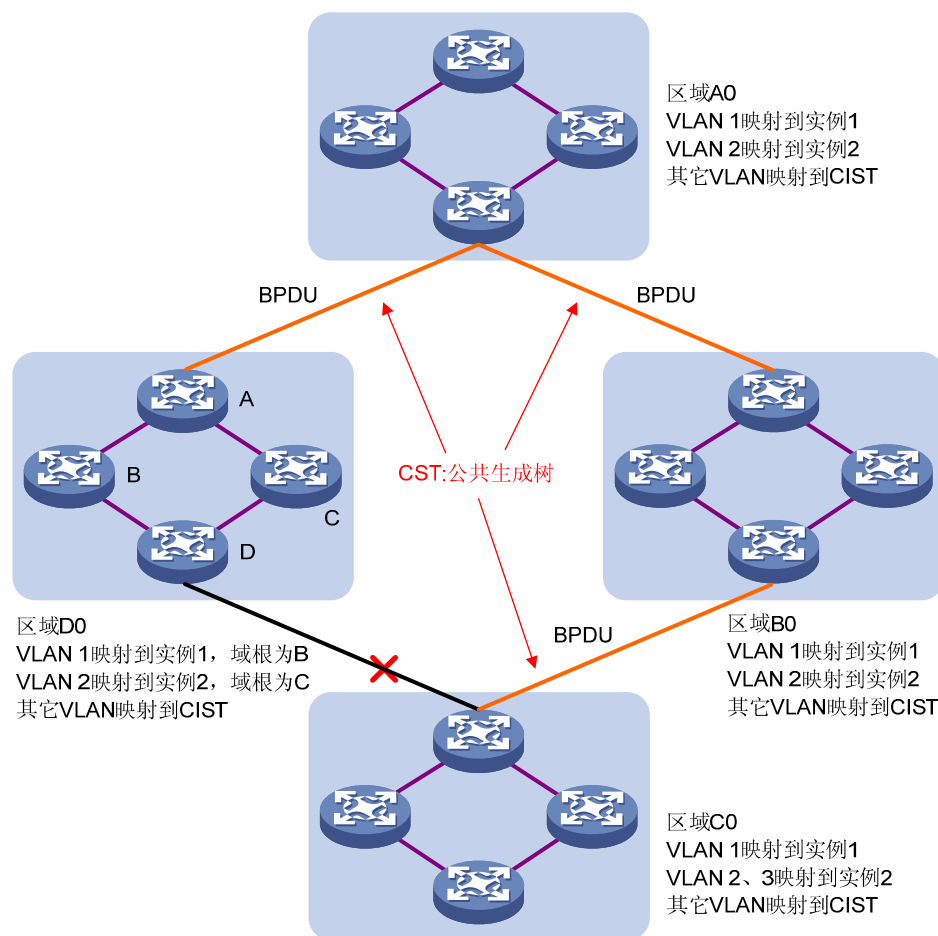
- MSTP 引入“域”的概念，把一个交换网络划分成多个域。每个域内形成多棵生成树，生成树之间彼此独立；在域间，MSTP 利用 CIST 保证全网络拓扑结构的无环路存在。
- MSTP 引入“实例（Instance）”的概念，将多个 VLAN 映射到一个实例中，以节省通信开销和资源占用率。MSTP 各个实例拓扑的计算是独立的（每个实例对应一棵单独的生成树），在这些实例上就可以实现 VLAN 数据的负载分担。
- PVST 最多支持 4096 个 VLAN，而 MSTP 最多可支持 65 个实例，因此其所需资源和计算量都大大降低。
- MSTP BPDU 中包含所有实例的信息，因此 MSTP 相比于 PVST，报文数量大大降低。
- MSTP 可以实现类似 RSTP 的端口状态快速迁移机制。
- MSTP 兼容 STP 和 RSTP。

2 MSTP技术实现

2.1 相关术语

在 [图 1](#) 中的每台设备都运行 MSTP。下面将结合 [图 1](#) 解释 MSTP 的一些基本概念。

图1 MSTP 基本概念示意图



(1) MST 域

MST域是由交换网络中的多台设备以及它们之间的网段所构成。这些设备具有下列特点：都启动了MSTP；具有相同的域名；具有相同的VLAN到生成树实例映射配置；具有相同的MSTP修订级别配置；这些设备之间在物理上有链路连通。例如，图1中的区域A0就是一个MST域。

(2) VLAN 映射表

VLAN映射表是MST域的一个属性，用来描述VLAN和生成树实例的映射关系。例如，图1中MST域A0的VLAN映射表就是：VLAN 1映射到生成树实例1，VLAN 2映射到生成树实例2，其余VLAN映射到CIST。

(3) IST

IST是域内实例0上的生成树。IST和CST共同构成整个交换网络的CIST。IST是CIST在MST域内的片段。图1中，CIST在每个MST域内都有一个片段，这个片段就是各个域内的IST。

(4) CST

CST是连接交换网络内所有MST域的单生成树。如果把每个MST域看作是一个“设备”，CST就是这些“设备”通过STP、RSTP计算生成的一棵生成树。图1中红色线条描绘的就是CST。

(5) CIST

CIST是连接一个交换网络内所有设备的单生成树，由IST和CST共同构成。[图 1](#)中，每个MST域内的IST加上MST域间的CST就构成整个网络的CIST。

(6) MSTI

一个MST域内可以通过MSTP生成多棵生成树，各棵生成树之间彼此独立。每棵生成树都称为一个MSTI。例如 [图 1](#)中，每个域内可以存在多棵生成树，每棵生成树和相应的VLAN对应。这些生成树就被称为MSTI。

(7) 域边界端口

域边界端口是指位于 MST 域的边缘，用于连接不同 MST 域、MST 域和运行 STP 的区域、MST 域和运行 RSTP 的区域的端口。

(8) 桥 ID

由桥的优先级和 MAC 地址组成。

(9) 总根

总根是指 CIST 实例中桥 ID 最优的桥。

(10) 外部根路径开销

外部根路径开销指的是端口到总根的最短路径开销。

(11) 域根

MST 域内的 IST 和每个 MSTI 的根桥都是一个域根。MST 域内各棵生成树的拓扑不同，域根也可能不同。

(12) 内部根路径开销

到域根的最短路径开销。

(13) 指定桥 ID

由指定桥的优先级和 MAC 地址组成。

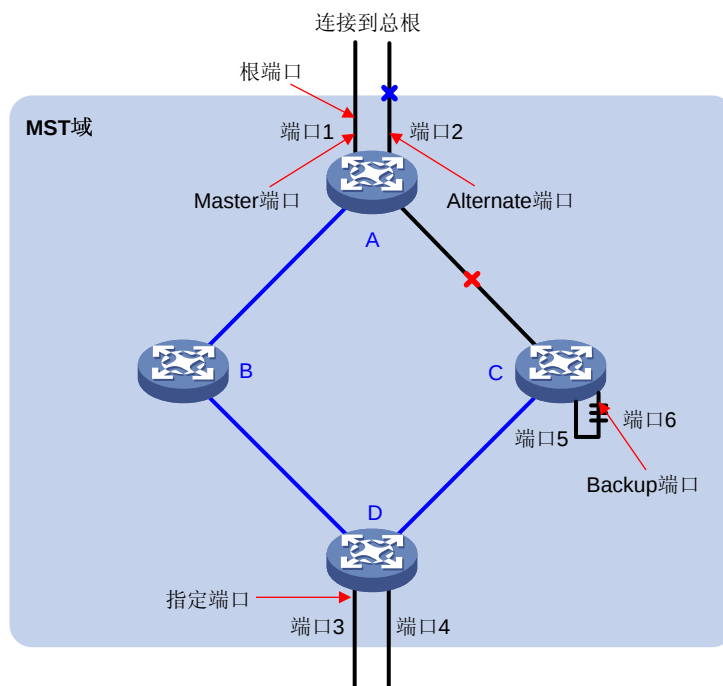
(14) 指定端口 ID

由指定端口的优先级和端口号组成。

(15) 端口角色

在MSTP的计算过程中，端口角色有根端口、指定端口、Master端口、Alternate端口和Backup端口。端口在不同的生成树实例中可以担任不同的角色。端口角色示意如 [图 2](#)所示。

图2 端口角色示意图



- 根端口：负责向根桥方向转发数据的端口。
- 指定端口：负责向下游网段或设备转发数据的端口。
- Master 端口：连接 MST 域到总根的端口，位于整个域到总根的最短路径上。Master 端口在 IST/CIST 上的角色是根端口，而在其它 MSTI 上的角色则是 Master 端口。
- Alternate 端口：根端口和 Master 端口的备份端口。当根端口或 Master 端口被阻塞后，Alternate 端口将成为新的根端口或 Master 端口。
- Backup 端口：当开启了 MSTP 的同一台设备的两个端口互相连接时就存在一个环路，此时设备会阻塞端口 ID 较小的端口，此阻塞端口称为 Backup 端口，而另外一个端口则处于转发状态，成为指定端口。Backup 端口是指定端口的备份端口，当指定端口被阻塞且无法发送协议报文后，Backup 端口的报文超时后就会快速转换为新的指定端口，并无时延的转发数据。

(16) 端口状态

MSTP 中，同一端口在不同的 MSTI 中的端口状态可以不同。

根据端口是否学习 MAC 地址和是否转发用户流量，可将端口状态划分为以下三种：

- Forwarding 状态：学习 MAC 地址，转发用户流量；
- Learning 状态：学习 MAC 地址，不转发用户流量；
- Discarding 状态：不学习 MAC 地址，不转发用户流量。

2.2 MSTP算法实现

2.2.1 初始状态

各台设备的各个端口在初始时会生成以自己为根桥的配置消息，总根和域根都是本桥 ID，外部根路径开销和内部根路径开销全为 0，指定桥 ID 为本桥 ID，指定端口为本端口，接收 BPDU 报文的端口为 0。

2.2.2 端口角色的选择原则

端口角色的选择原则中涉及到多种优先级向量，这些优先级向量的含义以及计算方法的介绍，请参考“[2.2.3 优先级向量计算](#)”。只要端口收到的消息优先级向量优于其端口优先级向量，就会引起所有优先级向量的重新计算，并且也会重新计算每个端口的角色。

端口角色的选择原则如 [表 1](#) 所示。

表1 端口角色的选择原则

端口角色	选择原则
根端口	端口的端口优先级向量优于其指定优先级向量，且设备的根优先级向量取自该端口的根路径优先级向量
指定端口	端口的指定优先级向量优于其端口优先级向量
Master端口	域边界根端口在MSTI实例上的角色就是Master端口
Alternate端口	端口的端口优先级向量优于其指定优先级向量，但设备的根优先级向量不是取自该端口的根路径优先级向量
Backup端口	端口的端口优先级向量优于其指定优先级向量，但端口优先级向量中的指定桥ID为本设备的桥ID

2.2.3 优先级向量计算

所有网桥的 MSTP 角色都是通过报文中携带的信息计算出来的，其中报文中携带的最重要的信息就是生成树的优先级向量。下面将分别介绍一下 CIST 优先级向量和 MSTI 优先级向量的计算方法。

1. CIST优先级向量计算

在 CIST 中优先级向量由总根、外部根路径开销、域根、内部根路径开销、指定桥 ID、指定端口 ID 和接收 BPDU 报文的端口 ID 组成。

为了方便后续描述，现做如下假设：

- 初始情况下，网桥 B 的端口 PB 对外发送报文中携带的信息如下：总根为 RB，外部根路径开销为 ERCB，域根为 RRB，内部根路径开销为 IRCB，指定桥 ID 为 B，指定端口 ID 为 PB，接收 BPDU 报文的端口 ID 为 PB；
- 网桥 B 的端口 PB 收到网桥 D 的端口 PD 发送过来的报文中携带的信息如下：总根为 RD，外部根路径开销为 ERCD，域根为 RRD，内部根路径开销为 IRCD，指定桥 ID 为 D，指定端口 ID 为 PD，接收 BPDU 报文的端口 ID 为 PB；
- 网桥 B 的端口 PB 收到的网桥 D 的端口 PD 发送过来的报文的优先级较高。

根据上述假设，下面将逐一介绍各优先级向量的计算方法。

(1) 消息优先级向量

消息优先级向量是 MSTP 报文中所携带的优先级向量。根据假设，网桥 B 的端口 PB 收到的消息优先级向量即为： $\{RD : ERCD : RRD : IRCD : D : PD : PB\}$ 。如果网桥 B 和网桥 D 不在同一个域，那么内部根路径开销对网桥 B 而言是毫无意义的，它会被赋值为 0。

(2) 端口优先级向量

在初始情况下，端口优先级向量的信息是以自己为根。端口 PB 的端口优先级向量为： $\{RB : ERCD : RRB : IRCD : B : PB : PB\}$ 。

端口优先级向量是随端口收到的消息优先级向量更新的：如果端口收到的消息优先级向量优于端口优先级向量，则将端口优先级向量更新为消息优先级向量；否则，端口优先级向量保持不变。由于端口 PB 收到的消息优先级向量优于端口优先级向量，所以端口优先级向量更新为： $\{RD : ERCD : RRD : IRCD : D : PD : PB\}$ 。

(3) 根路径优先级向量

根路径优先级向量由端口优先级向量计算所得：

- 如果端口的优先级向量来自不同域的网桥，根路径优先级向量的外部根路径开销为端口的路径开销和端口优先级向量的外部根路径开销之和，根路径优先级向量的域根为本桥的域根，内部根路径开销为 0。假设网桥 B 的端口 PB 的路径开销为 PCPB，则端口 PB 的根路径优先级向量为： $\{RD : ERCD + PCPB : B : 0 : D : PD : PB\}$ ；
- 如果端口优先级向量来自同一域的网桥，根路径优先级向量的内部路径开销为端口优先级向量的内部根路径开销和端口路径开销之和，计算后端口 PB 的根路径优先级向量为： $\{RD : ERCD : RRD : IRCD + PCPB : D : PD : PB\}$ 。

(4) 桥优先级向量

桥优先级向量中总根 ID、域根 ID 以及指定桥 ID 都是本桥 ID，外部根路径开销和内部根路径开销为 0，指定端口 ID 和接收端口 ID 也全为 0。网桥 B 的桥优先级向量为： $\{B : 0 : B : 0 : B : 0 : 0\}$ 。

(5) 根优先级向量

根优先级向量是桥优先级向量和所有指定桥 ID 和本桥 ID 值不相同的根路径优先级向量的最优值，如果本桥优先级向量比较优，那么本桥就为 CIST 总根。假设网桥 B 的桥优先级向量最优，则网桥 B 的根优先级向量为： $\{B : 0 : B : 0 : B : 0 : 0\}$ 。

(6) 指定优先级向量

端口的指定优先级向量由根优先级向量计算所得，将根优先级向量的指定桥 ID 替换为本桥 ID，指定端口 ID 替换为自己的端口 ID。网桥 B 的端口 PB 的指定优先级向量为： $\{B : 0 : B : 0 : B : PB : 0\}$ 。

2. MSTI 优先级向量计算

MSTI 的各优先级向量计算的规则和 CIST 优先级向量计算规则是基本一致的，存在两点区别：

- MSTI 优先级向量中没有总根和外部根路径开销，仅由域根、内部根路径开销、指定桥 ID、指定端口 ID 和接收 BPDU 报文的端口 ID 组成。
- MSTI 只处理来自同一域的消息优先级向量。

2.2.4 角色选择过程

下面结合 图 3 的组网对CIST实例的计算过程进行简要说明。假设，网桥的优先级为Switch A优于Switch B，Switch B优于Switch C，4、5、10 分别为链路的路径开销。Switch A和Switch B属于同一域，Switch C单独一个域。

图3 MSTP 算法计算过程组网图

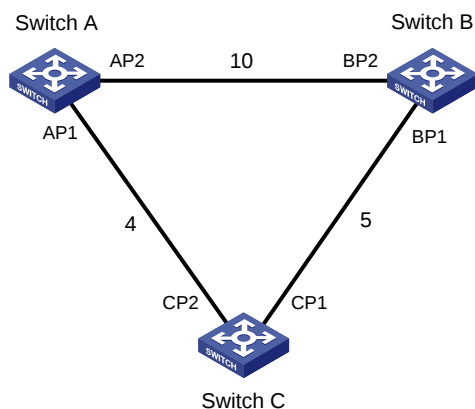


图 3 中各设备的初始情况下对外发送的报文中携带的消息优先级向量如 表 2 所示。

表2 各台设备的初始状态

设备	端口	报文中的消息优先级向量
Switch A	AP1	{A:0:A:0:A:AP1:0}
	AP2	{A:0:A:0:A:AP2:0}
Switch B	BP1	{B:0:B:0:B:BP1:0}
	BP2	{B:0:B:0:B:BP2:0}
Switch C	CP1	{C:0:C:0:C:CP2:0}
	CP2	{C:0:C:0:C:CP2:0}

设备各端口的端口优先级向量与消息优先级向量在初始情况下是保持一致的。

在初始情况下各设备的端口都会被计算为指定端口且对外发送以自己为根桥的消息优先级向量。

1. Switch A的角色选择过程

Switch A 的端口 AP1 和端口 AP2 会分别收到来自 Switch B 和 Switch C 的报文，Switch A 会将端口 AP1 以及 AP2 的端口优先级向量和收到的来自其它交换机的消息优先级向量进行比较，由于 AP1 和 AP2 的端口优先级向量优于报文中携带的消息优先级向量，端口 AP1 和 AP2 端口角色不变仍为指定端口，设备 Switch A 为总根且为 Switch A 和 Switch B 所在域的域根。此后端口定时对外传播以自己为根的消息。



说明

端口优先级向量和消息优先级向量的比较、处理过程为：

- 逐一比较端口优先级向量和消息优先级向量中的各元素，元素值较小的优先级向量较优，当各元素都相等时，端口优先级向量和消息优先级向量相等；
- 当消息优先级向量优于端口优先级向量或者消息优先级向量中的指定桥 ID 的桥 MAC 和指定端口 ID 分别和端口优先级向量中的指定桥 ID 的桥 MAC 和指定端口 ID 一致时，用消息优先级向量替换端口优先级向量。

2. Switch B的角色选择过程

Switch B 的端口 BP1 收到来自 Switch C 的端口 CP1 的报文后，将消息优先级向量和端口优先级向量比较，由于端口优先级向量优于消息优先级向量，端口角色不更新。

Switch B 的端口 BP2 收到来自 Switch A 的端口 AP2 的报文后，处理过程如下：

- (1) 将端口的消息优先级向量和端口优先级向量进行比较。由于端口的消息优先级向量优于端口优先级向量，将端口的端口优先级向量更新为消息优先级向量{A:0:A:0:A:AP2:BP2}；
- (2) 计算端口的根路径优先级向量。Switch A 和 Switch B 在同一域内，端口的根路径优先级向量为{A:0:A:10:A:AP2:BP2}；
- (3) 计算 Switch B 的根优先级向量。只有端口 BP2 的根路径优先级向量是来自其它设备，由于端口 BP2 的根路径优先级向量优于 Switch B 的桥优先级向量，Switch B 的根优先级向量为{A:0:A:10:A:AP2:BP2}；
- (4) 指定优先级向量计算。端口 BP1 的指定优先级向量为{A:0:A:10:B:BP1:BP2}，端口 BP2 的指定优先级向量为{A:0:A:10:B:BP2:BP2}。

端口角色的确定：将端口 BP1 和 BP2 的指定优先级向量和端口优先级向量进行比较，由于 BP1 的指定优先级向量优于端口优先级向量，则 BP1 角色为指定端口，定时对外发送以 Switch A 为总根和域根的指定优先级向量{A:0:A:10:B:BP1:BP2}；由于 BP2 的端口优先级向量优于指定优先级向量、且根优先级向量取自端口 BP2 的根路径优先级向量，则 BP2 角色为根端口。

3. Switch C的角色选择过程

Switch C 的端口 CP1 收到来自 Switch B 未更新前的消息优先级向量{B:0:B:0:B:BP1:CP1}，端口 CP2 收到来自 Switch A 的消息优先级向量{A:0:A:0:A:AP1:CP2}，经过分别比较，CP1 和 CP2 的消息优先级向量均优于端口优先级向量，因此分别更新 CP1 和 CP2 的端口优先级向量为{B:0:B:0:B:BP1:CP1}和{A:0:A:0:A:AP1:CP2}。由于 Switch C 与 Switch A 和 Switch B 不在同一域，端口 CP1 的根路径优先级向量为{B:5:C:0:B:BP1:CP1}，端口 CP2 的根路径优先级向量为{A:4:C:0:A:AP1:CP2}，CP2 的根路径优先级向量优于 CP1 的根路径优先级向量，则根优先级向量为{A:4:C:0:A:AP1:CP2}。端口 CP1 和 CP2 的指定优先级向量分别为{A:4:C:0:C:CP1:CP2}和{A:4:C:0:C:CP2:CP2}，端口 CP1 被计算为指定端口，CP2 被计算为根端口。

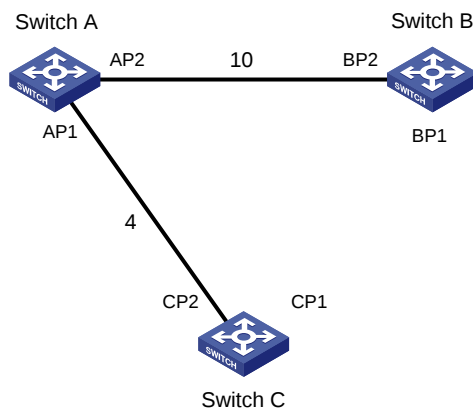
Switch C 的端口 CP1 收到来自 BP1 更新后的消息优先级向量{A:0:A:10:B:BP1:CP1}后，经过比较 CP1 的消息优先级向量优于端口优先级向量，更新端口优先级向量为{A:0:A:10:B:BP1:CP1}，端口 CP1 计算后的根路径优先级向量为{A:5:C:0:B:BP1:CP1}。由于端口 CP2 收到的消息优先级向量没有变化，根据前面的计算，端口 CP2 的根路径优先级向量保持为{A:4:C:0:A:AP1:CP2}，CP2 的根

路径优先级向量优于 CP1 的根路径优先级向量，则根优先级向量为{A:4:C:0:A:AP1:CP2}。端口 CP1 和 CP2 的指定优先级向量分别为{A:4:C:0:C:CP1:CP2}和{A:4:C:0:C:CP2:CP2}。CP1 的端口优先级向量优于其指定优先级向量、但根优先级向量不是取自端口 CP1 的根路径优先级向量，故 CP1 角色为 Alternate 端口。CP2 仍为根端口。

2.2.5 计算结果

设备和端口的角色确定之后，整个树形拓扑就建立完毕了。经过上述计算后的流量转发线路如 [图 4](#) 所示。

图4 计算后流量转发线路



3 Comware实现的技术特色

3.1 生成树协议的四种工作模式

生成树的工作模式有以下四种：

- STP 模式：设备的所有端口都将向外发送 STP BPDU。
- RSTP 模式：设备的所有端口都向外发送 RSTP BPDU。当端口收到对端设备发来的 STP BPDU 时，会自动迁移到 STP 模式；如果收到的是 MSTP BPDU，则不会进行迁移。
- MSTP 模式：设备的所有端口都向外发送 MSTP BPDU。当端口收到对端设备发来的 STP BPDU 时，会自动迁移到 STP 模式；如果收到的是 RSTP BPDU，则不会进行迁移。
- PVST 模式：设备的所有端口都向外发送 PVST BPDU，每个 VLAN 维护一棵生成树。

工作在 RSTP/MSTP 模式的设备可以自动迁移到 STP 模式下工作，但是工作在 STP 模式下的设备不能自动迁移到 RSTP/MSTP 模式，此时需要用户执行 mCheck 操作来迫使工作模式发生迁移。假设在一个交换网络中，运行 MSTP（或 RSTP）的设备的端口连接着运行 STP 的设备，该端口会自动迁移到 STP 模式下工作；但是此时如果运行 STP 的设备被拆离，该端口不能自动迁移到 MSTP（或 RSTP）模式下运行，仍然会工作在 STP 模式下。此时可以通过执行 mCheck 操作迫使其迁移到 MSTP（或 RSTP）模式下运行。

在 STP 模式和 RSTP 模式下可以配置多实例，MSTI 各端口状态和 CIST 保持一致。为了减小 CPU 的负担，建议在 STP 和 RSTP 模式下最好不要配置多实例。

3.2 Path Cost缺省值的计算

Comware STP 支持 3 种 Path Cost 缺省值的计算方法：IEEE 802.1D-1998 标准方法、IEEE 802.1t 标准方法和 Comware 的私有计算方法。

IEEE 802.1D-1998 和 IEEE 802.1t 标准的 Path Cost 缺省值的基本计算请参考协议文本，下面主要介绍对标准协议的一些扩充以及 Comware 的私有计算方法。

(1) 对 IEEE 802.1D-1998 标准方法的扩充

对于聚合链路，IEEE 802.1D-1998 并没有具体的规定，它没有区分聚合链路和单端口链路的优先级别的不同，因此对于 IEEE 802.1D-1998 中聚合链路 STP 的 Path Cost 值不用考虑聚合链路数。

(2) 对 IEEE 802.1t 标准方法的扩充

IEEE 802.1t Path Cost 计算标准中，端口 Path Cost 值计算公式为：20,000,000,000 / Link Speed in Kbps，聚合链路速率为聚合链路中所有选中端口速率相加。

(3) Comware 的私有计算方法

$$\text{Path Cost} = \begin{cases} 200,000 & (\text{Link Speed} = 0) \\ 2,200 - 20 * \text{Link Speed} & (0 < \text{Link Speed} \leq 100) \\ 220 - 0.2 * \text{Link Speed} & (100 < \text{Link Speed} \leq 1000) \\ 22 - 0.002 * \text{Link Speed} & (1000 < \text{Link Speed} \leq 10,000) \\ 1 & (\text{Link Speed} > 10,000) , \text{Link Speed in Kbps} \end{cases}$$

聚合链路速率为所有 unblock 端口速率相加。

3.3 设置超时因子特性

协议中规定，端口收到的 STP 报文超时时间 rcvdinfowhile 小于等于 3 倍 hello time，超过 rcvdinfowhile 后端口还未收到 STP 报文就会重新计算 STP 拓扑。在实际组网中，可能会出现由于设备负载较重端口不能及时收到 STP 报文导致 rcvdinfowhile 定时超时而引起网络拓扑震荡，为了解决该问题引入了超时因子。可以根据实际的组网情况设置超时因子，从而改变报文的超时时间，增强网络的稳定性。

超时时间 = 超时因子 × 3 × hello time。

3.4 指定根桥和备份根桥

STP 可以通过计算来确定生成树的根桥，用户也可以通过交换机提供的命令来指定当前交换机为根桥。

当根桥出现故障或被关机时，备份根桥可以取代根桥成为生成树的根桥；但是此时如果用户设置了新的根桥，则备份根桥将不会成为根桥。如果用户为生成树配置了多个备份根桥，当根桥失效时，STP 将选择 MAC 地址最小的那个备份根桥作为根桥。当根桥和备份根桥都失效时，STP 将通过协议计算来自动选举根桥。

在 MSTP 或 PVST 中，交换机上可以基于 MSTI 或 VLAN 运行多个生成树实例，用户可以将当前交换机指定为生成树实例的根桥或备份根桥。当前交换机在各棵生成树实例中的根类型互相独立，它

可以作为一棵生成树实例的根桥或备份根桥，同时也可以作为其他生成树实例的根桥或备份根桥。在同一棵生成树实例中，同一台交换机不能既作为根桥，又作为备份根桥。

3.5 BPDU保护

对于接入层设备，接入端口一般直接与用户终端（如 PC 机）或文件服务器相连，此时可以设置接入端口为边缘端口以实现这些端口的快速迁移。正常情况下，边缘端口不会收到生成树协议的配置消息（BPDU 报文），但是，如果有人伪造配置消息恶意攻击交换机，当边缘端口接收到配置消息时，系统会自动将这些端口设置为非边缘端口，重新进行生成树的计算，这将引起网络拓扑的震荡。BPDU 保护功能可以防止这种网络攻击。

交换机上启动了 BPDU 保护功能以后，如果边缘端口收到了配置消息，系统就将这些端口关闭，同时通知网管。被关闭端口在一定时间之后可以自动恢复。推荐用户在配置了边缘端口的交换机上配置 BPDU 保护功能。

3.6 Root保护和角色限制

由于维护人员的错误配置或网络中的恶意攻击，网络中的合法根桥有可能会收到优先级更高的配置消息，这样当前根桥会失去根桥的地位，引起网络拓扑结构的错误变动。假设原来的流量是经过高速链路转发的，这种不合法的变动，会导致原来通过高速链路的流量被牵引到低速链路上，导致网络拥塞。Root 保护功能和角色限制功能可以防止这种情况的发生。

对于设置了 Root 保护功能的端口，端口角色只能保持为指定端口。一旦这种端口上收到了优先级高的配置消息，这些端口的状态将被设置为 Discarding 状态，不再转发报文。当在足够长的时间内没有收到更优的配置消息时，端口会恢复原来的状态。

对于设置了角色限制功能的端口，端口角色被限制不能成为根端口。正常情况下这些端口被计算为指定端口，一旦端口上收到了优先级高的配置消息，这些端口的角色将被计算为 Alternate 端口、状态将被设置为阻塞状态，不再转发报文。当在足够长的时间内没有收到更优的配置消息时，端口会恢复原来的角色和状态。

角色限制功能在 IEEE 802.1Q 中定义，优先推荐使用。在 MSTP 中，此功能对所有实例都起作用；在 PVST 中，此功能对所有 VLAN 都起作用。

3.7 Loop保护和Dispute机制

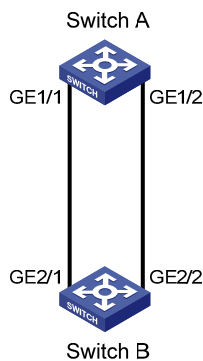
交换机各端口的端口状态依靠不断接收上游交换机发送的 BPDU 来维持。但是由于链路拥塞或者单向链路故障，根端口会收不到上游交换机的 BPDU。此时下游交换机会重新选择根端口，原来的根端口经过计算后会变为指定端口，而原来的阻塞端口重新计算后会变为根端口且迁移到转发状态，从而交换网络中会产生环路。环路保护功能和 Dispute 机制会抑制这种环路的产生。

在启动了环路保护功能后，根端口、Alternate 端口和 Backup 端口的角色一旦发生变化就会变为 Discarding 状态，并一直保持在 Discarding 状态，不转发报文，从而不会在网络中形成环路。

角色限制功能默认在所有端口生效，当端口收到这样的报文——报文中携带指定端口角色和 Learning/Forwarding 状态并且报文中的优先级向量低于接收端口的优先级向量，端口的 Dispute 机制生效，端口被设置为 Discarding 状态，阻止网络形成环路。

下面结合 [图 5](#) 举例说明 Loop 保护功能和 Dispute 机制。

图5 Loop 保护功能和 Dispute 机制



Switch A 和 Switch B 都为 Comware 交换机。假设，Switch A 为根交换机，Switch B 的端口 GigabitEthernet2/1 角色为根端口，GigabitEthernet2/2 角色为 Alternate 端口。如果端口 GigabitEthernet2/1 由于链路拥塞收不到 Switch A 的 BPDU 报文，经过一段时间后，Switch B 重新计算角色，端口 GigabitEthernet2/1 被计算为指定端口且状态为 Forwarding，GigabitEthernet2/2 被计算为根端口且状态为 Forwarding，此时 Switch A 与 Switch B 之间的两条链路均可以转发报文，只要链路中存在广播报文就会引起广播风暴。

对于 图 5 的组网，在 GigabitEthernet2/1 启动环路保护功能，当端口 GigabitEthernet2/1 在一定时间内收不到 BPDU 报文时，角色就会被计算为指定端口并被设置变为 Discarding 状态，从而阻止环路的形成，端口 GigabitEthernet2/2 会迁移到转发状态，进行报文转发。

对于 图 5 的组网，当端口 GigabitEthernet2/1 在一定时间内收不到 BPDU 报文时，角色就会被计算为指定端口并被设置变为 Forwarding 状态，端口 GigabitEthernet1/1 收到端口 GigabitEthernet2/1 发送的报文时，Dispute 机制会将端口 GigabitEthernet1/1 设置为 Discarding 状态，从而阻止环路的形成。端口 GigabitEthernet2/2 会迁移到转发状态，进行报文转发。

3.8 TC 保护

交换机在接收到 TC-BPDU 报文（网络拓扑发生变化的通知报文）后，会执行转发地址表项的删除操作。在有人伪造 TC-BPDU 报文恶意攻击交换机时，交换机短时间内会收到很多的 TC-BPDU 报文，频繁的删除操作给交换机带来很大负担，给网络的稳定带来隐患。

TC 保护功能使能后，设备在收到 TC-BPDU 报文后的 10 秒内，允许收到 TC-BPDU 报文后立即进行地址表项删除操作的次数可以由用户控制（假设次数限制为 X）。同时系统会监控在该时间段内收到的 TC-BPDU 报文数是否大于 X，如果大于 X，则设备在该时间超时后再进行一次地址表项删除操作。这样就可以避免频繁地删除转发地址表项。

3.9 TC 传播限制

用户接入网络的拓扑改变会引起核心网络的转发地址更新，当用户接入网络的拓扑因某种原因而不稳定时，就会对核心网络形成冲击。为了避免这种情况，可以在与用户接入网络相连的核心网络端口上使能 TC-BPDU 传播限制功能，此后当该端口收到 TC-BPDU 时，不会再向其它端口传播。

3.10 配置摘要侦听

根据 IEEE 802.1s 的规定，相连交换机若实现 MSTP 域内 MSTI 的互通，它们的域配置（域名、修订级别、VLAN 与实例的映射关系）必须完全一致。MSTP 在发送 BPDU 报文的时候，会把配置 ID（配置 ID 由域名、修订级别和配置摘要组成，其中配置摘要是由 VLAN 与实例的映射关系经过 HMAC-MD5 运算生成的 16 字节签名）放到报文中传输，相连的交换机就是根据这些信息来判断发送报文的交换机和自己是否处于同一个域内。

如果其它厂商的配置摘要计算方法和标准中列举的参考例子不一致，那么 Comware 交换机和其它厂商交换机即使域配置相同，各自计算出的配置摘要也会不相同，所以它们不会认为在一个域内，这样就只能实现 CIST 的互通，不能实现 MSTI 的互通。

Comware MSTP 提供如下方法可以和配置摘要计算方法和标准协议不一致的厂商交换机实现域内 MSTI 的互通。

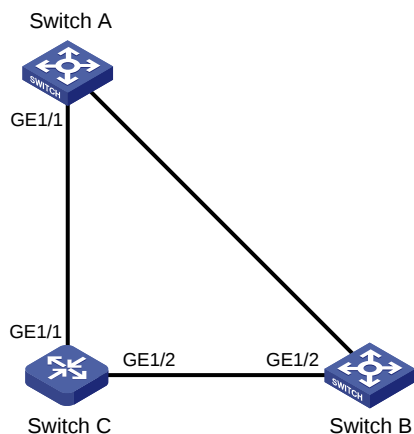
在保证相连 Comware 交换机域配置和其它厂商交换机域配置完全一致的前提下，可以通过命令在每一个和其它厂商交换机相连的端口上启动配置摘要侦听功能。对于启动了配置摘要侦听功能的端口，在接收到其它厂商交换机 MSTP 报文时，直接认为报文来自域内，同时记录下报文中的配置摘要；在发送 MSTP 报文时，将之前记录的配置摘要填充到发送的报文中，这就保证了其它厂商交换机接收到该报文时也认为它来自域内，这样 Comware MSTP 和其它厂商交换机 MSTP 就可以在 MSTI 域内互通了。

配置摘要侦听时，需要注意：

- 配置摘要侦听功能一定要在 Comware 交换机和其它厂商交换机的域配置完全相同的条件下启动，否则可能因为各交换机 VLAN 与实例映射关系不一致导致广播风暴。
- 域内和其它厂商交换机相连的每一个端口都必须启动配置摘要侦听功能；在域的边界端口上不能使能配置摘要侦听功能。
- 不要直接更改启动了配置摘要侦听功能的 Comware 交换机及其相连的其它厂商交换机的域配置。请在更改域配置之前将配置摘要侦听功能关闭，否则在更改域配置的过程中可能因为各交换机 VLAN 与实例映射关系不一致导致广播风暴。
- 如果域内都是 Comware 交换机，则不必启动配置摘要侦听功能。

下面结合 [图 6](#) 举例说明配置摘要侦听功能。

图6 配置摘要侦听功能



Switch A、Switch B 为 Comware 交换机，Switch C 为配置摘要计算方法非标准的其它厂商交换机，所有设备都启用 MSTP，并且域配置都相同。

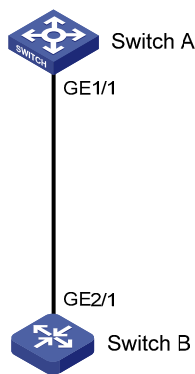
在组网中，为了实现 Switch A 与 Switch C 以及 Switch B 与 Switch C 之间的域内互通，必须在 Switch A 的端口 GigabitEthernet1/1 和 Switch B 的端口 GigabitEthernet1/2 上启动配置摘要侦听功能。Switch A 与 Switch B 由于都是 Comware 设备，不需要启动配置摘要侦听功能。

3.11 No Agreement Check特性实现

IEEE 802.1Q-2003 MSTP 标准协议中规定，指定端口快速迁移的条件就是收到下游根端口发送的携带 agreement 标志报文，而根端口发送携带 agreement 标志报文的前提又是收到上游指定端口发送携带 agreement 标志的报文。当交换机（如：RSTP 交换机）的指定端口发送的报文中不携带 agreement 标志时，该交换机作为上游桥设备与 Comware 设备 MSTP 互通时，因为 Comware 设备根端口无法收到上游指定端口发送的携带 agreement 标志的报文，所以不向上游桥指定端口回应 agreement 标志报文，从而导致该交换机的指定端口无法快速迁移。可以通过在 Comware 设备的端口上启动 No Agreement Check 特性来避免端口连接的上游桥设备为 RSTP 交换机或者与 MSTP 实现存在私有性差异的厂商设备时，上游桥设备不能快速迁移问题。

下面结合 [图 7](#) 举例说明 No Agreement Check 特性。

图7 No Agreement Check 特性



Switch A 为根交换机并为 RSTP 交换机，Switch B 的 GigabitEthernet2/1 为根端口。为了使 Switch A 的端口 GigabitEthernet1/1 能够快速迁移，应该在 Switch B 的 GigabitEthernet2/1 上启动 No Agreement Check。

IEEE 802.1Q-2005 MSTP 标准协议对指定端口快速迁移条件做了修订，根端口发送携带 agreement 标志报文不再依赖收到上游指定端口发送携带 agreement 标志的报文。因此当上游桥设备为 RSTP 交换机时，上游桥设备也能快速迁移。

3.12 MSTP标准报文格式特性实现

本特性主要实现 Comware 设备与支持标准 MSTP 报文格式设备之间的互通。有了此特性，当标准协议报文格式设备和私有协议报文格式设备混合组网时，就能够正确地进行网络拓扑计算。

Comware 设备在缺省配置下能自动识别接收的 MSTP 报文格式，其发送报文格式可以根据接收到的报文格式自动更改，按接收到的报文格式向外发送报文。也可按用户在端口上的实际配置收发用户指定格式的报文。Comware 设备如果工作在 RSTP 或 STP 模式下与其它设备组网，在公共生成树实例上也能正常互通。

3.13 VLAN Ignore特性

在一般情况下，某个 VLAN 都会映射到一个 MSTP 的实例中，此 VLAN 中的端口在该实例上的转发状态由 MSTP 计算得出。在网络拓扑比较复杂的情况下，某些 VLAN 的拓扑可能会被生成树阻塞，造成该 VLAN 的业务流量不通，为了解决该问题引入了 VLAN Ignore 特性。当在该 VLAN 上启动 VLAN Ignore 特性后，该 VLAN 中每个端口的实际转发状态不再遵从 MSTP 计算出的状态，而是一直保持 Forwarding 的状态。当该 VLAN 关闭 VLAN Ignore 特性后，该 VLAN 中每个端口的实际转发状态仍然会遵从 MSTP 计算出的状态。

4 典型组网应用

MSTP 可以使得同一组网中的不同 VLAN 的报文按照不同的生成树进行转发，从而实现不同 VLAN 数据的负载分担和冗余备份。

图8 MSTP 典型组网图

