RSTP 概要

关于此文有任何疑问或者建议请到www.51cto.com

或者游来游去岛留言,也可以直接给我发邮件.

crazycat84@sina.com

二〇〇六年五月

目录

1．协议简介(Overview)

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。

由于生成树协议本身比较小，所以并不像路由协议那样广为人知。但是它却掌管着端口的转发大权—“小树枝抖一抖，上层协议就得另谋生路”。真实情况也确实如此，特别是在和别的协议一起运行的时候，生成树就有可能断了其他协议的报文通路，造成种种奇怪的现象。

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。本章标题中的“生成树协议”是一个广义的概念，并不是特指 IEEE 802.1D 中定义的 STP 协议，而是包括 STP 以及各种在 STP 基础上经过改进了的生成树协议。

在生成树协议发展过程中，老的缺陷不断被克服，新的特性不断被开发出来。按照大功能点的改进情况，我们可以粗略地把生成树协议的发展过程划分成三代，下面一一道来。

1.1 开天辟地的第一代生成树协议：STP/RSTP

在网络发展初期，透明网桥是一个不得不提的重要角色。它比只会放大和广播信号的集线器聪明得多。它会悄悄把发向它的数据帧的源 MAC 地址和端口号记录下来，下次碰到这个目的 MAC 地址的报文就只从记录中的端口号发送出去，除非目的 MAC 地址没有记录在案或者目的 MAC 地址本身就是多播地址才会向所有端口发送。通过透明网桥，不同的局域网之间可以实现互通，网络可操作的范围得以扩大，而且由于透明网桥具备 MAC 地址学习功能而不会像 Hub 那样造成网络报文冲撞泛滥。

但是，金无足赤，透明网桥也有它的缺陷，它的缺陷就在于它的透明传输。透明网桥并不能像路由器那样知道报文可以经过多少次转发，一旦网络存在环路就会造成报文在环路内不断循环和增生，甚至造成恐怖的“广播风暴”。之所以用“恐怖”二字是因为在这种情况下，网络将变得不可用，而且在大型网络中故障不好定位，所以广播风暴是二层网络中灾难性的故障。

在这种大环境下，扮演着救世主角色的 STP（Spanning Tree Protocol）协议来到人间，其中以 IEEE 的 802.1D 版本最为流行。

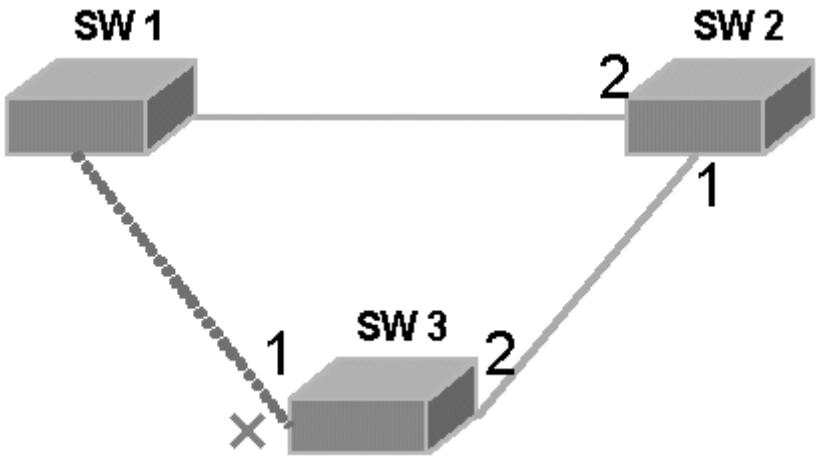


图 1-1　生成树工作过程示意图

STP 协议的基本思想十分简单。大家知道，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 协议中定义了根桥(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自然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 SPA（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流，这些信息交流单元就称为配置消息 BPDU（Bridge Protocol Data Unit)。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 协议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。

要了解生成树协议的工作过程也不难，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（Bridge ID)，桥 ID 最小的网桥将成为网络中的根桥。在图 1-1 所示的网络中，各网桥都以默认配置启动，在网桥优先级都一样（默认优先级是 32768）的情况下，MAC 地址最小的网桥成为根桥，例如图 1-1 中的 SW1，它的所有端口的角色都成为指定端口，进入转发状态。

接下来，其他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。假设图 1-1 中 SW2 和 SW2、SW3 之间的链路是千兆 GE 链路，SW1 和 SW3 之间的链路是百兆 FE 链路，SW3 从端口 1 到根桥的路径开销的默认值是 19，而从端口 2 经过 SW2 到根桥的路径开销是 4+4=8，所以端口 2 成为根端口，进入转发状态。同理，SW2 的端口 2 成为根端口，端口 1 成为指定端口，进入转发状态。

根桥和根端口都确定之后一棵树就生成了，如图中实线所示。下面的任务是裁剪冗余的

环路。这个工作是通过阻塞非根桥上相应端口(非根端口,非指定端口)来实现的，例如 SW3 的端口 1 的角色成为禁用端口，进入阻塞状态（图中用“×”表示)。

生成树经过一段时间（默认值是 30 秒左右）稳定之后，所有端口要么进入转发状态，要么进入阻塞状态。STP BPDU 仍然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。

当然生成树协议还有很多内容，在这里不可能一一介绍。之所以花这么多笔墨介绍生成树的基本原理是因为它太“基本”了，其他各种改进型的生成树协议都是以此为基础的，基本思想和概念都大同小异。

STP 协议给透明网桥带来了新生。但是，随着应用的深入和网络技术的发展，它的缺点在应用中也被暴露了出来。STP 协议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，生成树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！

为了解决 STP 协议的这个缺陷，在世纪之初 IEEE 推出了 802.1w 标准，作为对 802.1D 标准的补充。在 IEEE 802.1w 标准里定义了快速生成树协议 RSTP（Rapid Spanning Tree Protocol)。RSTP 协议在 STP 协议基础上做了三点重要改进，使得收敛速度快得多（最快 1 秒以内)。

第一点改进：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/指定端口失效的情况下，替换端口/备份端口就会无时延地进入转发状态。图 1-2 中所有网桥都运行 RSTP 协议，SW1 是根桥，假设 SW2 的端口 1 是根端口，端口 2 将能够识别这种拓扑结构，成为根端口的替换端口，进入阻塞状态。当端口 1 所在链路失效的情况下，端口 2 就能够立即进入转发状态，无需等待两倍 Forward Delay 时间。

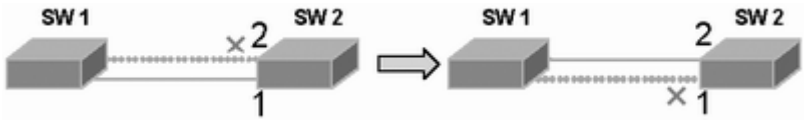


图 1-2　RSTP 冗余链路快速切换示意图

第二点改进：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。

第三点改进：直接与终端相连而不是把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。

可见，RSTP 协议相对于 STP 协议的确改进了很多。为了支持这些改进，BPDU 的格式做了一些修改，但 RSTP 协议仍然向下兼容 STP 协议，可以混合组网。虽然如此，RSTP 和 STP 一样同属于单生成树 SST（Single Spanning Tree），有它自身的诸多缺陷，主要表现在三个方面。

第一点缺陷：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，拓扑改变的影响面也较大。

第二点缺陷：近些年 IEEE 802.1Q 大行其道，逐渐成为交换机的标准协议。在网络结构对称的情况下，单生成树也没什么大碍。但是，在网络结构不对称的时候，单生成树就会影响网络的连通性。

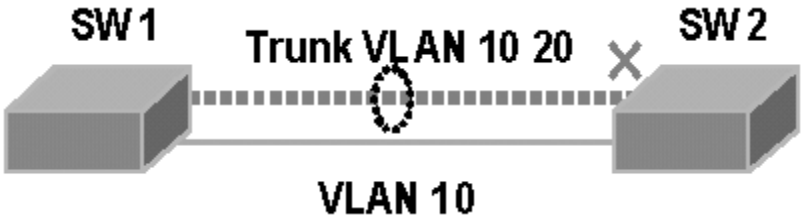


图 1-3　非对称网络示意图

图 1-3 中假设 SW1 是根桥，实线链路是 VLAN 10，虚线链路是 802.1Q 的 Trunk 链路，Trunk 了 VLAN 10 和 VLAN 20。当 SW2 的 Trunk 端口被阻塞的时候，显然 SW1 和 SW2 之间 VLAN 20 的通路就被切断了。

第三点缺陷：当链路被阻塞后将不承载任何流量，造成了带宽的极大浪费，这在环行城域网的情况下比较明显。

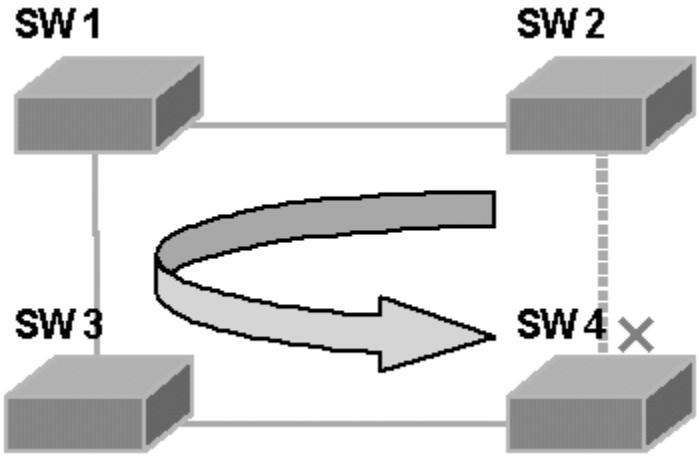


图 1-4　SST 带宽利用率低下示意图

图 1-4 中假设 SW1 是根桥，SW4 的一个端口被阻塞。在这种情况下，SW2 和 SW4 之间铺设的光纤将不承载任何流量，所有 SW2 和 SW4 之间的业务流量都将经过 SW1 和 SW3 转发，增加了其他几条链路的负担。

这些缺陷都是单生成树 SST 无法克服的，于是支持 VLAN 的多生成树协议出现了。

1.2　聪明伶俐的第二代生成树协议：PVST/PVST+

每个 VLAN 都生成一棵树是一种比较直接，而且最简单的解决方法。它能够保证每一个 VLAN 都不存在环路。但是由于种种原因，以这种方式工作的生成树协议并没有形成标准，而是各个厂商各有一套，尤其是以 Cisco 的 VLAN 生成树 PVST（Per VLAN Spanning Tree）为代表。

为了携带更多的信息，PVST BPDU 的格式和 STP/RSTP BPDU 格式已经不一样，发送的目的地址也改成了 Cisco 保留地址 01-00-0C-CC-CC-CD，而且在 VLAN Trunk 的情况下 PVST BPDU 被打上了 802.1Q VLAN 标签。所以，PVST 协议并不兼容 STP/RSTP 协议。

Cisco 很快又推出了经过改进的 PVST＋协议，并成为了交换机产品的默认生成树协议。经过改进的 PVST＋协议在 VLAN 1 上运行的是普通 STP 协议，在其他 VLAN 上运行 PVST 协议。PVST＋协议可以与 STP/RSTP 互通，在 VLAN 1 上生成树状态按照 STP 协议计算。在其他 VLAN 上，普通交换机只会把 PVST BPDU 当作多播报文按照 VLAN 号进行转发。但这并不影响环路的消除，只是有可能 VLAN 1 和其他 VLAN 的根桥状态可能不一致。

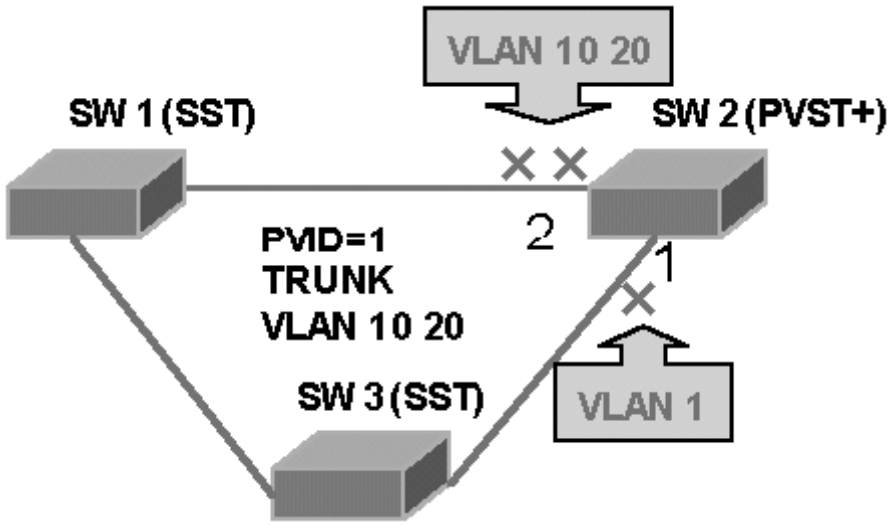


图 1-5　PVST+与 SST 对接示意图

图 1-5 中所有链路默认 VLAN 是 VLAN 1，并且都 Trunk 了 VLAN 10 和 VLAN 20。SW1 和 SW3 运行单生成树 SST 协议，而 SW2 运行 PVST＋协议。在 VLAN 1 上，可能 SW1 是根桥，SW2 的端口 1 被阻塞。在 VLAN 10 和 VLAN 20 上，SW2 只能看到自己的 PVST BPDU，所以在这两个 VLAN 上它认为自己是根桥。VLAN 10 和 VLAN 20 的 PVST BPDU 会被 SW1 和 SW3 转发，所以 SW2 检测到这种环路后，会在端口 2 上阻塞 VLAN 10 和 VLAN 20。这就是 PVST＋协议提供的 STP/RSTP 兼容性。可以看出，网络中的二层环路能够被识别并消除，强求根桥的一致性是没有任何意义的。

由于每个 VLAN 都有一棵独立的生成树，单生成树的种种缺陷都被克服了。同时，PVST 带来了新的好处，那就是二层负载均衡。

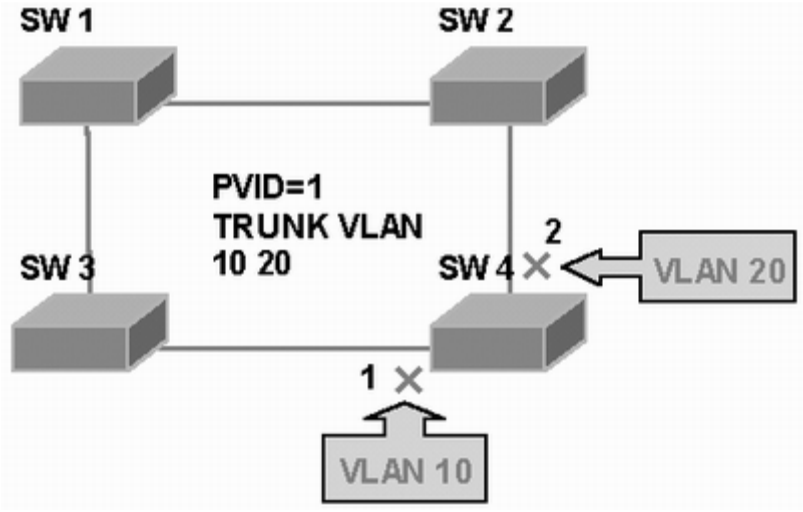


图 1-6　PVST＋负载均衡示意图

图 1-6 中四台设备都运行 PVST＋协议，并且都 Trunk 了 VLAN 10 和 VLAN 20。假设 SW1 是所有 VLAN 的根桥，通过配置可以使得 SW4 端口 1 上的 VLAN 10 和端口 2 上的 VLAN 20 阻塞，SW4 的端口 1 所在链路仍然可以承载 VLAN 20 的流量，端口 2 所在链路也可以承载 VLAN 10 的流量，同时具备链路备份的功能。这在以往的单生成树情况下是无法实现的。

聪明伶俐的 PVST/PVST＋协议实现了 VLAN 认知能力和负载均衡能力，但是新技术也带来了新问题，PVST/PVST＋协议也有它们的“难言之隐”。

第一点缺陷：由于每个 VLAN 都需要生成一棵树，PVST BPDU 的通信量将正比于 Trunk 的 VLAN 个数。

第二点缺陷：在 VLAN 个数比较多的时候，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 Trunk 了很多 VLAN 的接口状态变化的时候，所有生成树的状态都要重新计算，CPU 将不堪重负。所以，Cisco 交换机限制了 VLAN 的使用个数，同时不建议在一个端口上 Trunk 很多 VLAN。

第三点缺陷：由于协议的私有性，PVST/PVST＋不能像 STP/RSTP 一样得到广泛的支持，不同厂家的设备并不能在这种模式下直接互通，只能通过一些变通的方式实现，例如 Foundry 的 IronSpan。IronSpan 默认情况下运行的是 STP 协议，当某个端口收到 PVST BPDU 时，该端口的生成树模式会自动切换成 PVST/PVST＋兼容模式。

一般情况下，网络的拓扑结构不会频繁变化，所以 PVST/PVST＋的这些缺点并不会很致命。但是，端口 Trunk 大量 VLAN 这种需求还是存在的。于是，Cisco 对 PVST/PVST＋又做了新的改进，推出了多实例化的 MISTP 协议。

1.3 多实例化的第三代生成树协议：MISTP/MSTP

多实例生成树协议 MISTP（Multi-Instance Spanning Tree Protocol）定义了“实例”（Instance）的概念。简单的说，STP/RSTP 是基于端口的，PVST/PVST＋是基于 VLAN 的，而 MISTP 就是基于实例的。所谓实例就是多个 VLAN 的一个集合，通过多个 VLAN 捆绑到一个实例中去的方法可以节省通信开销和资源占用率。

在使用的时候可以把多个相同拓扑结构的 VLAN 映射到一个实例里，这些 VLAN 在端口上转发状态将取决于对应实例在 MISTP 里的状态。值得注意的是网络里的所有交换机的 VLAN 和实例映射关系必须都一致，否则会影响网络连通性。为了检测这种错误，MISTP BPDU 里

除了携带实例号以外，还要携带实例对应的 VLAN 关系等信息。MISTP 协议不处理 STP/RSTP/PVST BPDU，所以不能兼容 STP/RSTP 协议，甚至不能向下兼容 PVST/PVST＋协议，在一起组网的时候会出现环路。为了让网络能够平滑地从 PVST＋模式迁移到 MISTP 模式，Cisco 在交换机产品里又做了一个可以处理 PVST BPDU 的混合模式 MISTP-PVST＋。网络升级的时候需要先把设备都设置成 MISTP-PVST＋模式，然后再全部设置成 MISTP 模式。

MISTP 带来的好处是显而易见的。它既有 PVST 的 VLAN 认知能力和负载均衡能力，又拥有可以和 SST 媲美的低 CPU 占用率。不过，极差的向下兼容性和协议的私有性阻挡了 MISTP 的大范围应用。

多生成树协议 MSTP（Multiple Spanning Tree Protocol）是 IEEE 802.1s 中定义的一种新型多实例化生成树协议。这个协议目前仍然在不断优化过程中，现在只有草案（Draft）版本可以获得。不过 Cisco 已经在 CatOS 7.1 版本里增加了 MSTP 的支持，华为公司的三层交换机产品 Quidway 系列交换机也即将推出支持 MSTP 协议的新版本。

MSTP 协议精妙的地方在于把支持 MSTP 的交换机和不支持 MSTP 交换机划分成不同的区域，分别称作 MST 域和 SST 域。在 MST 域内部运行多实例化的生成树，在 MST 域的边缘运行 RSTP 兼容的内部生成树 IST（Internal Spanning Tree）。

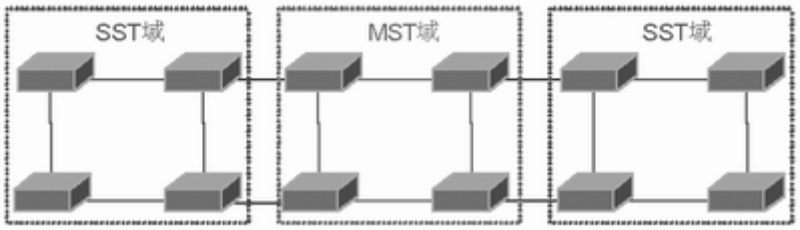


图 1-7　MSTP 工作原理示意图

图 1-7 中间的 MST 域内的交换机间使用 MSTP BPDU 交换拓扑信息， SST 域内的交换机使用 STP/RSTP/PVST＋ BPDU 交换拓扑信息。在 MST 域与 SST 域之间的边缘上，SST 设备会认为对接的设备也是一台 RSTP 设备。而 MST 设备在边缘端口上的状态将取决于内部生成树的状态，也就是说端口上所有 VLAN 的生成树状态将保持一致。

MSTP 设备内部需要维护的生成树包括若干个内部生成树 IST，个数和连接了多少个 SST 域有关。另外，还有若干个多生成树实例 MSTI（Multiple Spanning Tree Instance）确定的 MSTP 生成树，个数由配置了多少个实例决定。

MSTP 具有 VLAN 认知能力，可以实现负载均衡，可以实现类似 RSTP 的端口状态快速切换，可以捆绑多个 VLAN 到一个实例中以降低资源占用率。最难能可贵的是 MSTP 可以很好

地向下兼容 STP/RSTP 协议。而且，MSTP 是 IEEE 标准协议，推广的阻力相对小得多。

可见，各项全能的 MSTP 协议能够成为当今生成树发展的一致方向是当之无愧的。

1.4 生成树协议的未来之路

任何技术的发展都不会因为某项“理想”技术的出现而停滞，生成树协议的发展历程本身就说明了这一点。随着应用的深入，各种新的二层隧道技术不断涌现，例如 Cisco 的 802.1Q Tunneling，华为 Quidway S8016 的 QinQ，以及基于 MPLS 的二层 VPN 技术等。在这种新形势下，用户和服务提供商对生成树协议又会有新的需求。生成树协议该往何处走？这个问题虽然现在还没有一个统一的答案，但是各厂商已经开始了这方面的积极探索。也许不久的将来，支持二层隧道技术的生成树协议将成为交换机的标准协议。

- 本章出自《网管员世界》2003 年第 7 期
- 以下重点介绍 RSTP,以 802.1D 2004, 802.1w 2001 为参考标准
- 在本文中,网桥(Bridge)与交换机(Switch)是等同的

2. 基本概念(Definition)

2.1 端口角色(Port Role)

2.1.1 根端口(Root Port)

收到最好(依次比较 BPDU 之间或者 BPDU 与端口存储的 Root Id, Root Path Cost, Bridge ID, Port ID,数字越小,就是更好.处理更好的,丢弃更差的)BPDU 的端口成为根端口.它提供到 ROOT 最近的距离,也就是 Root Path Cost 最小.

2.1.2 指定端口(Designated Port)

在每个网段,发送最好 BPDU 的端口成为指定端口,对应的网桥成为指定网桥.每个网段通过指定端口到达 ROOT 的距离最近.

2.1.3 备份端口(Backup Port)

如果一个端口收到同一个网桥的更好 BPDU,那么这个端口成为备份端口.

2.1.4 替换端口(Alternate Port)

如果一个端口收到另外一个网桥的更好的 BPDU,但不是最好的,那么这个端口成为替换端口.

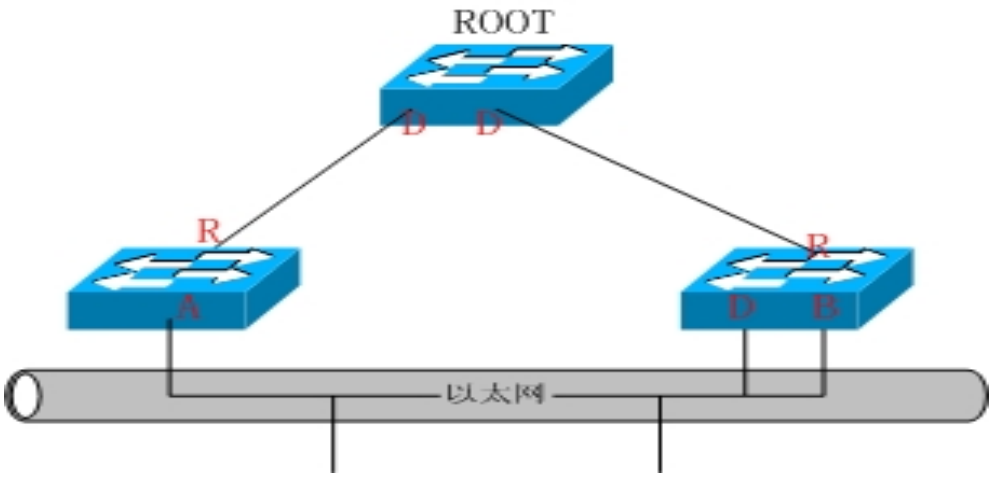


图 2-1 端口角色示意图

2.2 端口状态(Port State)

Relationship between Port State values in STP and RSTP

STP Port State	Administrative Bridge Port State	MAC Operational	RSTP Port State	Active Topology (Port Role)
DISABLED	Disabled	FALSE	Discarding	Excluded (Disabled)
DISABLED	Enabled	FALSE	Discarding	Excluded (Disabled)
BLOCKING	Enabled	TRUE	Discarding	Excluded (Alternate, Backup)
LISTENING	Enabled	TRUE	Discarding	Included (Root, Designated)
LEARNING	Enabled	TRUE	Learning	Included (Root, Designated)
FORWARDING	Enabled	TRUE	Forwarding	Included (Root, Designated)

图 2-2 RSTP 与 STP 端口状态比较示意图

2.2.1 丢弃(Discarding)

丢弃接收帧,不转发帧,不进行地址学习.能够处理和发送 BPDU.

2.2.2 学习(Learning)

丢弃接收帧,不转发帧,进行地址学习.能够处理和发送 BPDU.

2.2.3 转发(Forwarding)

转发帧.发送帧.进行地址学习, 能够处理和发送 BPDU.
拓扑稳定时,根端口,指定端口处于 Forwarding 状态,替换端口,备份端口,Disable 端口处于 Discarding 状态.

2.3 端口参数(Per-Port variables)

2.3.1 端口 ID(Port Identifier)

在一个交换机中,端口 ID 唯一标识一个端口.端口 ID 由端口优先级(Port Priority)和端口编号组成.端口优先级可以手动配置.端口编号不能手动配置,在交换机内唯一.

2.3.2 边缘端口(Edge Port)

与终端直接相连的端口就是边缘端口.边缘端口不会形成环路.边缘端口 UP/DOWN 不会引发 TCN.边缘端口也不会 FLUSH MAC 表.如果边缘端口收到一个 BPDU,那么它就不再是边缘端口,而是一个通常的生成树端口.边缘端口无法自动识别,需要手动配置(802.1w).

2.3.3 链路类型(Link Type)

端口只与一个网桥相连的链路是点对点(Point-to-Point)链路.端口与多个网桥相连的链路是共享(share medium)链路.链路类型对替换端口转换为根端口没有影响.指定端口快速迁移到转发状态只发生在点对点链路,除非是边缘端口.链路类型根据双工来判断.全双工认为

是点对点链路,半双工认为是共享链路.链路类型可以手动配置.

2.3.4 端口链路代价(Port Path Cost)

用于计算最短路径,取决于链路带宽,可以手动配置.

2.4 网桥参数(Per-Bridge variables)

2.4.1 网桥 ID(Bridge Identifier)

每个网桥都有个唯一的网桥 ID.网桥 ID 由网桥 ID 优先级(Bridge Identifier Priority)和网桥地址(Bridge Address)组成.网桥 ID 优先级可以手动配置,在比较网桥 ID 时,先比较网桥 ID 优先级,数字越小,优先级越高(better),相同时再比较网桥地址.网桥地址全球唯一,确保网桥 ID 全球唯一.

2.4.2 最大消息生存时间(maximum age)

消息生存时间指配置消息(Configuration Message)产生后经历的时间,配置消息由 ROOT 产生,ROOT 发送的 BPDU 中,消息生存时间总是 0.最大消息生存时间也是由 ROOT 设置,全网网桥共用.为了避免旧消息在冗余链路上无休止的计算,也为了避免新消息的传输,每个 Configuratoin Message 都包含了 message age 和 maximum age, message age 每经过一个交换机加 1,当 message age > maximum age 时,消息被丢弃.所以网络上的交换机数量受到限制.

2.4.3 转发延迟(Forward Delay)

Discarding-->Learning--->Forwarding 的状态转换 Timeout 时间,由 ROOT 设置,全网统一.这个时间也用于 MAC 表项的快速老化时间(Short Ageing Timer).

2.4.4 保活时间(Hello Time)

BPDU 发送的时间间隔,由 ROOT 设置,全网统一.

2.5 状态机参数(State machine parameters)

2.5.1 老化时间(Ageing Time)

当 mac 地址项由 Learning Process 创建或者刷新后,在 Ageine Time 内,没有同样的源 MAC 进入这个端口,那么这个 mac 地址项被删除.默认是 300 秒(normally Ageing Time),如果端口工作在 stpVersion, Ageing Time 被拓扑变化状态机(Topology Change State Machine)设置为 15(Forward Delay)秒(Rapid Ageing Time).

2.5.2 迁移时间(Migrate Time)

用于设置 mdelayWhile 和 edgeDelayWhile(802.1D 2004). mdelayWhile 是 RSTP<--->STP 之间的转换最小间隔时间,如果在 edgeDelayWhile 时间内,端口没有收到 BPDU,那么这个端口就会自动成为 edge port.

2.5.3 拓扑变化通知时间(TC While)

在 TC While 时间内,端口发送 TCN Messages.在点对点链路上 TC While=2*2(Hello Time),在共享链路和 STP 端口上,TC While=20(Max Age)+15(Forward Delay). 在 RSTP 中,只有 LINK UP 才会引发拓扑变化.

2.5.4 传输间隔(Transmit Hold Count)

Transmit Hold Count 用来限制 BPDU 的发送速率.每发一个 BPDU,txCount+1, T 每秒 txCount-1. 当 txCount=TransmitHold 时,延时发送 BPDU.

2.5.5 协议版本(Force Protocol Version)

Force Protocol Version<2, stpVersion=TRUE. 网桥工作在生成树模式.

Force Protocol Version>1, rstpVersion=TRUE. 网桥工作在快速生成树模式.

2.6 参数取值(Parameter Value)

2.6.1 性能参数(Performace Parameter)

Parameter	Recommended or Default value	Permitted Range	Compatibility Range
Migrate Time (17.13.9)	3.0	—[a]	—[a]
Bridge Hello Time (17.13.6)	2.0	—[a]	1.0–2.0
Bridge Max Age (17.13.8)	20.0	6.0–40.0	6.0–40.0
Bridge Forward Delay (17.13.5)	15.0	4.0–30.0	4.0–30.0
Transmit Hold Count (17.13.12)	6	1–10	1–10

All times are in seconds. —[1] Not applicable, value is fixed.

图 2-3 性能参数取值范围

Parameter	Recommended or default value	Range
Bridge Priority	32 768	0–61 440 in steps of 4096
Port Priority	128	0–240 in steps of 16

- 2 * (Bridge_Forward_Delay - 1.0 seconds) >= Bridge_Max_Age
- Bridge_Max_Age >= 2 * (Bridge_Hello_Time + 1.0 seconds)

2.6.2 端口链路代价(Port Path Cost)

Link Speed	Recommended value	Recommended range	Range
<=100 Kb/s	200 000 000*	20 000 000–200 000 000	1–200 000 000
1 Mb/s	20 000 000[a]	2 000 000–200 000 000	1–200 000 000
10 Mb/s	2 000 000[a]	200 000–20 000 000	1–200 000 000
100 Mb/s	200 000[a]	20 000–2 000 000	1–200 000 000
1 Gb/s	20 000	2 000–200 000	1–200 000 000
10 Gb/s	2 000	200–20 000	1–200 000 000
100 Gb/s	200	20–2 000	1–200 000 000
1 Tb/s	20	2–200	1–200 000 000
10 Tb/s	2	1–20	1–200 000 000

图 2-4 端口链路代价取值范围

2.7 消息封装(Encoding of BPDUs)

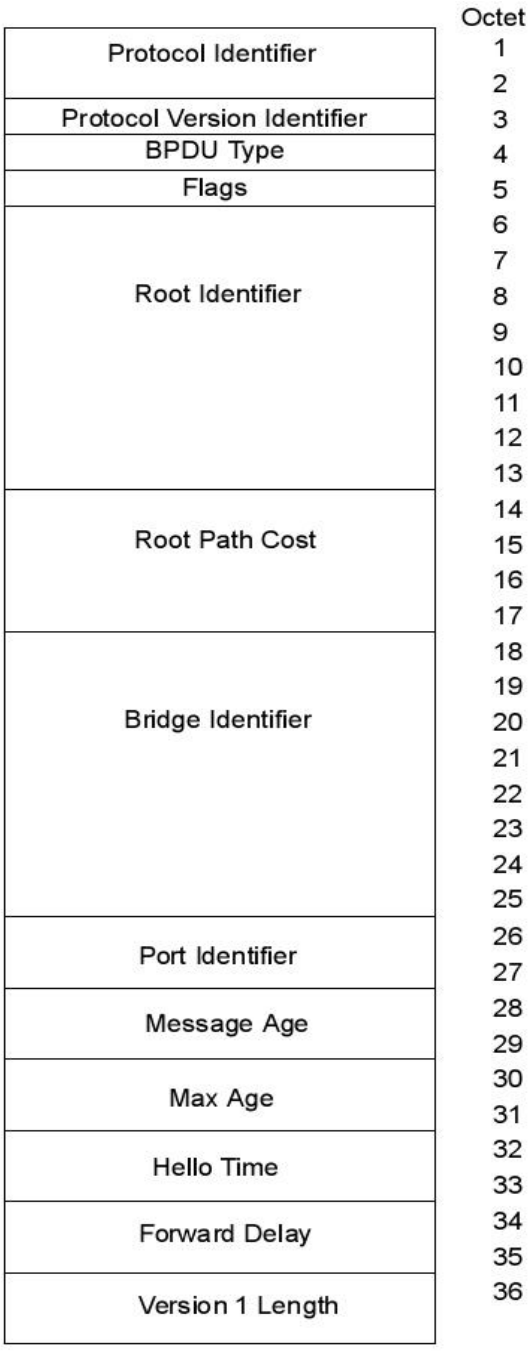


图 2-5 协议封装格式

所有 BPDU 都包含整数个字节(Octet).BPDU 中的字节序号根据它们进入 Data Link Service Date Unit(DLSDU)的顺序编号.在一个字节里,高位(Higher Bit)更重要,在连续字节中,低序号字节(Lower Octet)更重要.

2.7.1 Encoding of Protocol Identifiers

STP,RSTP 都是 0000 0000 0000 0000.

2.7.2 Encoding of Protocol Version Identifiers

0000 0000 代表 STP, 0000 0010 代表 RSTP.数字越大,版本越新.

2.7.3 Encoding of BPDU Types

0000 0000 代表 Configuration BPDU, 0000 0010 代表 RST BPDU, 1000 0000 代表 TCN BPDU.

2.7.4 Encoding of Flags

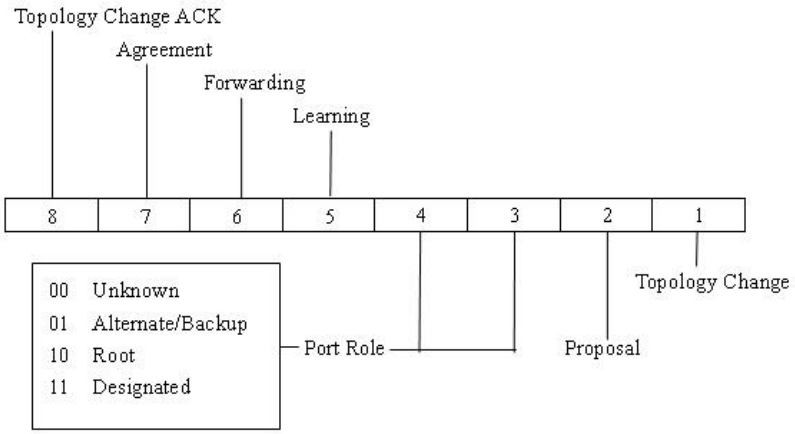


图 2-6 RST Flag 示意图

拓扑发生变化时,Topology Change 置 1.

指定端口要快速迁移到 Forwarding 状态时,Proposal 置 1.

Port Role 封装发送此 BPDU 的 Port Role,具体值见上图.

端口可以进行地址学习时,Learning 置 1.

端口进入 Forwarding 状态时,Forwarding 置 1.

根端口收到 Proposal 时,Agreement 置 1.

指定端口收到 TCN BPDU(STP)时,Topology Change ACK 置 1.

Proposal, Agreement 设置参见 3.

Topology Change, Topology Change ACK 设置参见 3.

2.7.5 Encoding of Bridge Identifiers

Root Identifier 里封装 Root Bridge 的 Bridge Identifier.

Bridge Identifier 里封装 Designated Bridge Identifier.

Bridge Identifier 由 2 个字节的优先级和 6 个字节的 MAC 地址组成.

2.7.6 Encoding of Root Path Cost

封装网桥到根网桥的总距离.

2.7.7 Encoding of Port Identifiers

Port Identifier 里封装 Designated Port Identifier.

Port Identifier 由优先级和端口编号组成,共 2 个字节.

2.7.8 Encoding of Timer Values

Max Age, Hello Time, Forward Delay 由 ROOT 设置,所有 BPDU 都采用同样的值.

Message age 每经过一个网桥+1.

2.7.9 Encoding of Length Values

0000 0000 代表非 Version 1 Protocol. STP BPDU 没有此项.

2.7.10 Validation of Received BPDUs

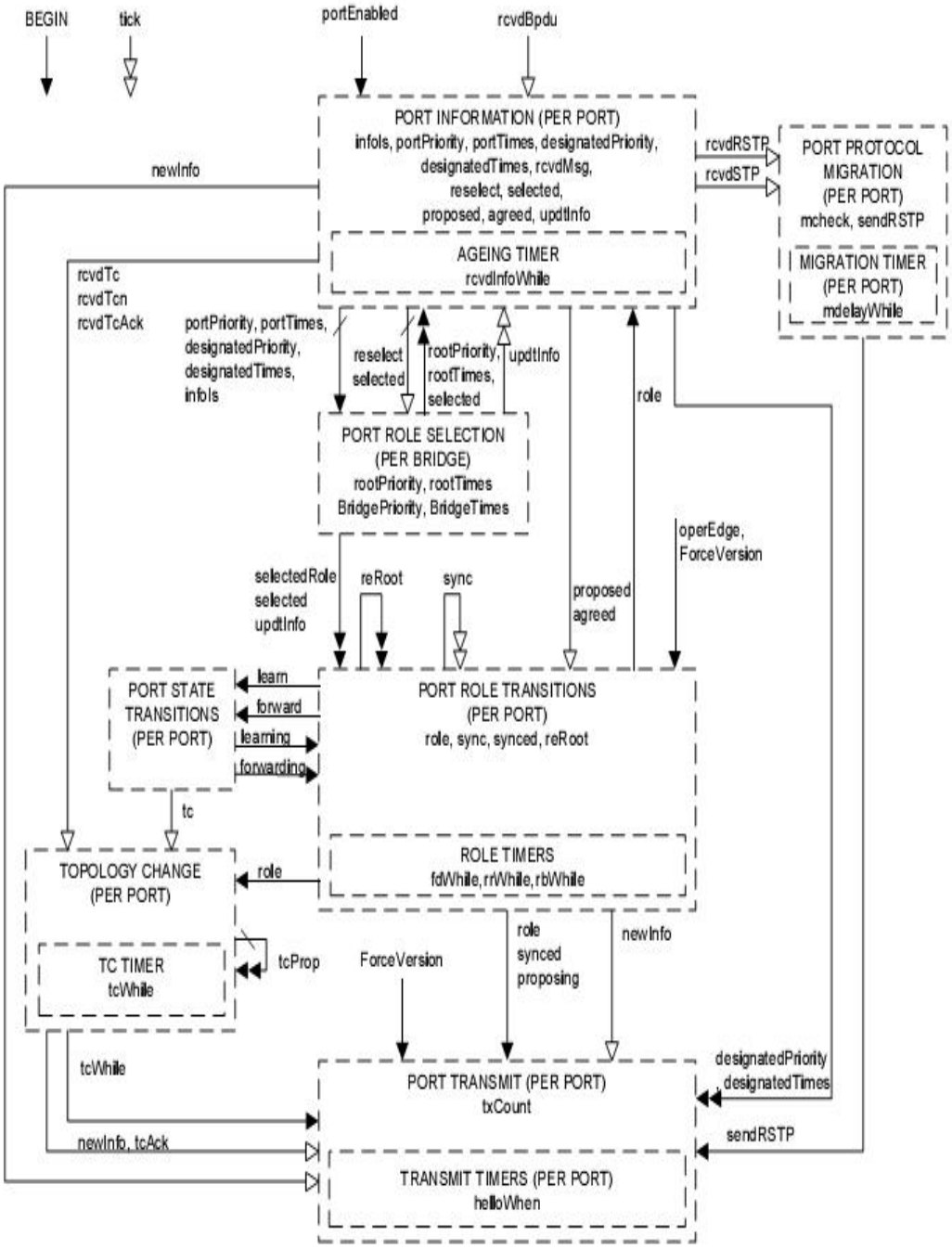
配置(Configuration)BPDU 至少有 35 个字节,Message Age<Max Age, 网桥 ID 和端口 ID 不能与接收到 BPDU 中的同时一样.

RSTP BPDU 至少有 36 个字节.

3. 状态机(State Machine)

状态机是深层次的知识,一般情况不必深究.这里只作简单描述,如果需要进一步学习,请参见 802.1w 2001 或者 802.1D 2004.只须关注3.11 Handshake.

3.1 overview and interrelationships



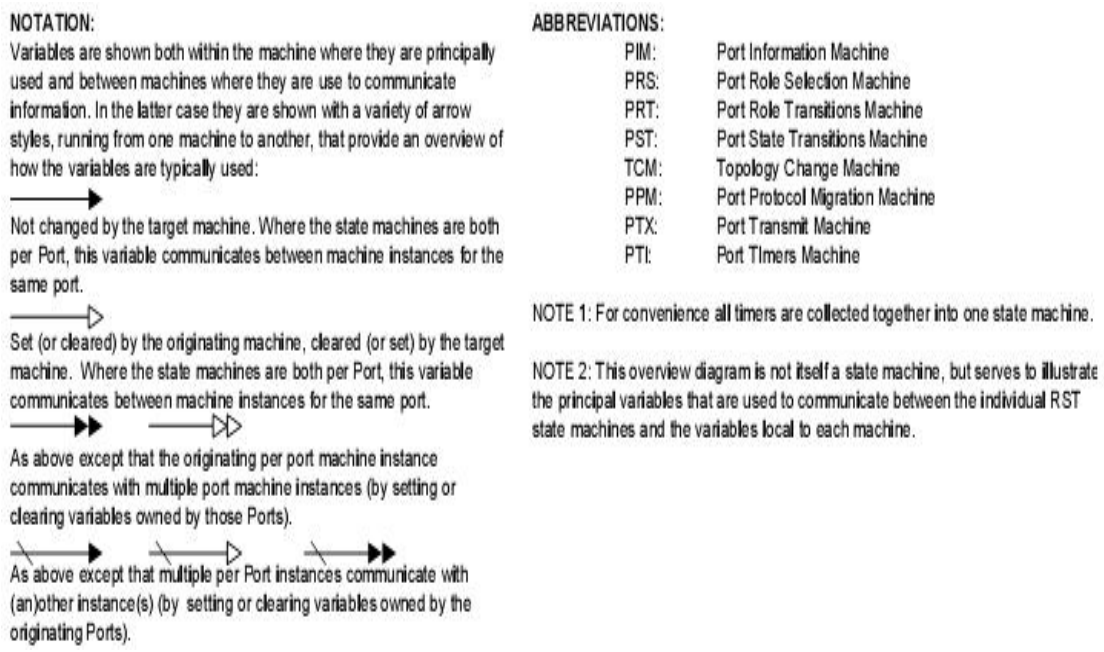


图 3-1 RSTP state machines - overview and interrelationships

3.2 Notational Conventions

状态图(State Diagrams)将一些互斥的状态连接,描述某个功能的操作方式.在任一时刻,只有一个状态是活动的(Active).

在状态图里,每个状态用一个矩形框表示.矩形框用水平线分为两部分.上面是状态ID(State Identifier),用大写表示.下面是进入状态后要执行的程序.

状态之间的转换用带箭头的线表示.箭头表示转换的目的.线旁边的字代表发生转换的前提条件.如果一个状态可能由其它任一个状态转换而来,那么这条线用一个没有源的箭头线表示(Open Arrow).

进入一个状态后,马上执行相应的程序.每个程序都是原子的(atomic),不可分的,只有当前面的程序执行完后,才会执行下一个程序.程序不会在状态之外执行.当所有程序执行完毕后,开始等待退出此状态的条件.UCT 代表无条件转换(Unconditional Transition).

当同一级别的退出条件同时发生多个时,状态转换是随机的.

状态框里的变量值保持不变,除非其它状态框进行了设置.

如果文字说明和状态图相矛盾,以状态图为准.

3.3 Port Timers State Machine

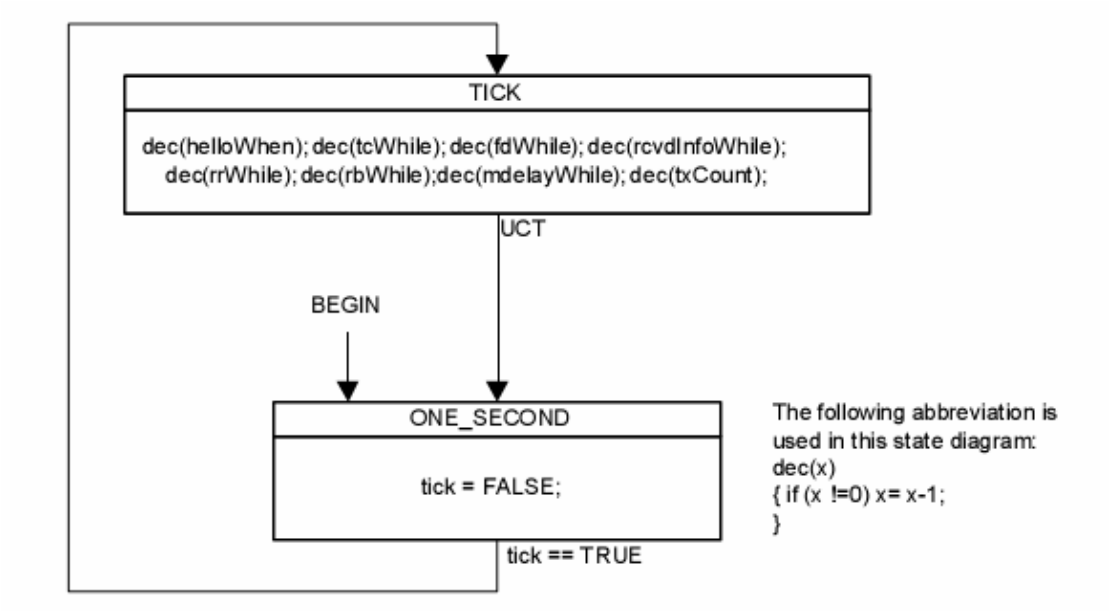


图 3-2 Port Timers State Machine

每一秒,helloWhen, tcWhile, fdWhile, rcvdInfoWhile, rrWhile, rbWhile, mdelayWhile,txCount 的值减 1,如果这些变量的值不等于 0.

helloWhen 用来确保每个 Hello Time 时间内都有 BPDU 发送.

rcvdInfoWhile:接收到 BPDU 信息的剩余生存时间.如果 message Age < maximum age,rcvdInforWhile 的值为下面 2 个中较小的一个:

- maximum age - message age
- 3 * 2(hello time)

rrWhile=recent root while.根端口或者最近是根端口的 rrWhile=15.当端口变成 Discarding 时,rrWhile=0.

rbWhile=recent backup while. 备份端口或者最近是备份端口的 rbWhile=4.

mdelayWhile 是 RSTP<--->STP 之间的转换最小间隔时间.

txCount 用来控制 BPDU 的发送速率.

3.4 Port Information State Machine

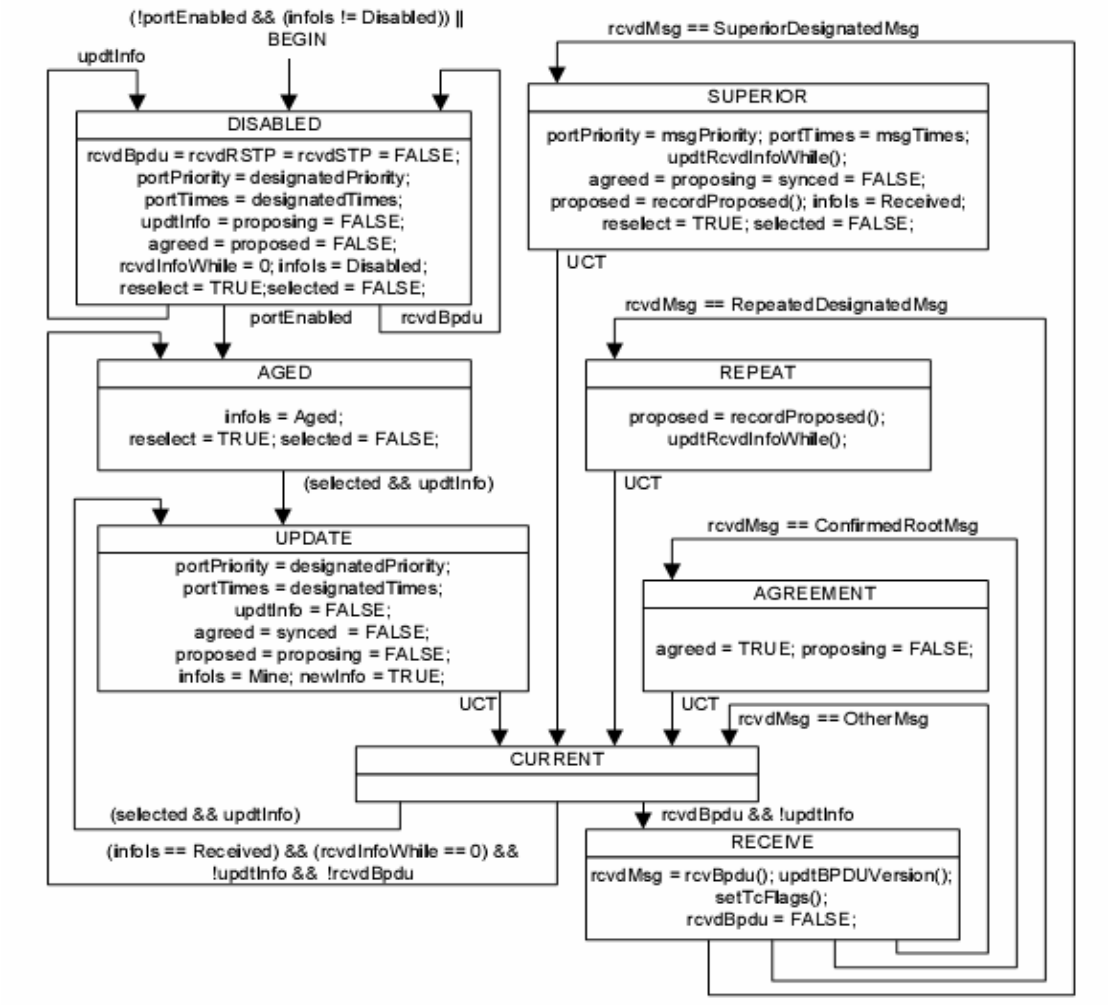


图 3-3 Port Information State Machine

在 RECEIVE 状态,先对 rcvdMsg 变量进行设置,然后 updtBPDUVersion()结合 Port Protocol Migration State 进行版本设置.setTcFlags()记录 BPDU 中的拓扑变化信息.

如果接收到的 BPDU 来自一个新的指定网桥,或者当前指定网桥,但是包含 change message prioriy,进入 SUPERIOR 状态,如果 BPDU 来自当前指定网桥,信息与以前不变,进入 REPEAT 状态,如果 BPDU 包含端口可以进入 Forwarding 的 agreement 信息,进入 AGREEMETN 状态.

else,进入 CURRENT 状态.

在 SUPERIOR 状态,先把 msgPriority 保存在 portPriority,时间信息保存在 portTimers,更新 rcvdInfoWhile,reselct=TRUE 使交换机重新计算端口角色.执行完所有程序后,进入 CURRENT 状态.

3.5 Port Role Selection State Machine

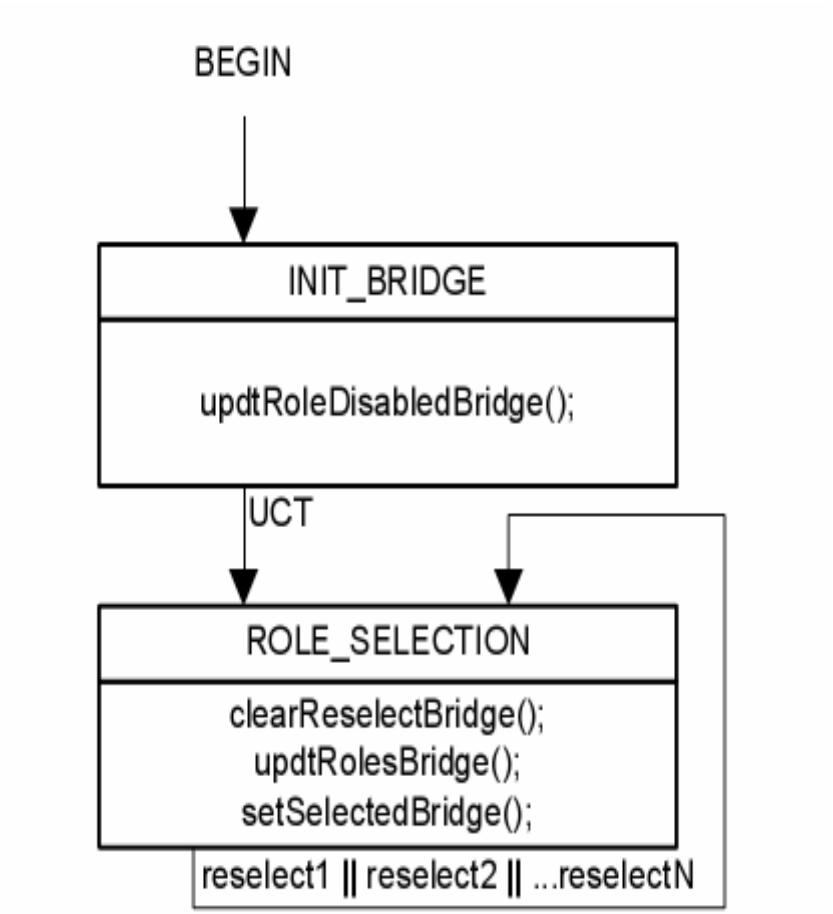


图 3-4 Port Role Selction State Machine

初始化时,进入 INIT_BRIDGE 状态,所有端口的端口角色设置为 Disable Port.

在 ROLE-SELECTION 状态, clearReselectBridge()将所有端口的 reselct 变量设置为 FALSE, 是为了在计算过程中,如果接收到新的消息,可以立刻进行新的计算.

不论什么时候,交换机的任何一个端口的 reselct 变量为 TRUE 时,都会重新进入 ROLE-SELECTION 状态,updtRolesBridge()进行端口角色的计算,计算完成后,设置 selected 变量为 TRUE.

3.6 Port States Transitions State Machine

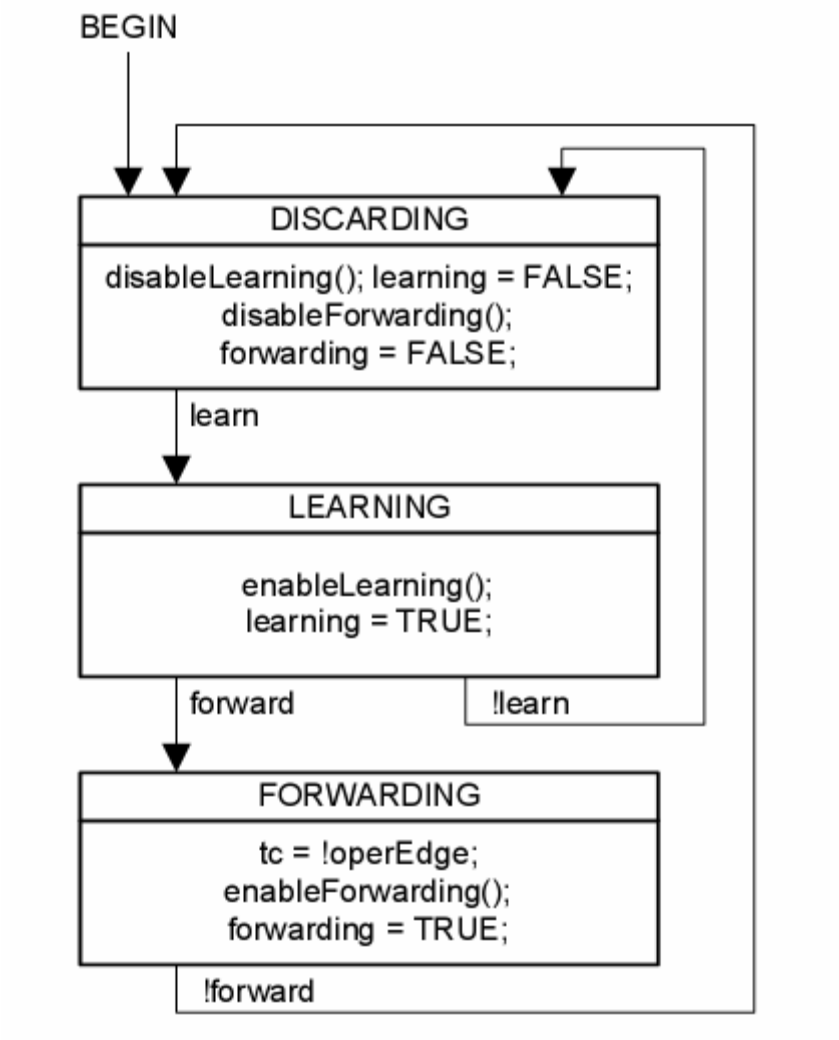


图 3-5 Port States Transition State Machine

DISCARDING, LEARNING, FORWARDING 状态之间的变化通过 learn,forward 变量来进行.这两个变量来自 Port Role Transition State Machine. Port Role Transition State Machine 又要利用 Port States Transtions State Machine 设置的 learning, forwarding 变量.

状态机首先进入 DISCARDING 状态,learning=FALSE, forwarding=FALSE.如果 learn=TRUE,进入 LEARNING 状态.在 LEARNING 状态,learning=TRUE.如果 learn=FLASE,回到 DISCARDING 状态,如果 forward=TRUE,进入 FORWARDING 状态.如果不是边缘端口,设置 TC=1,forwarding=TRUE.在 FORWARDING 状态,如果 forward=FALSE.回到 DISCARDING 状态.

3.7 Port Role Transitions State Machine

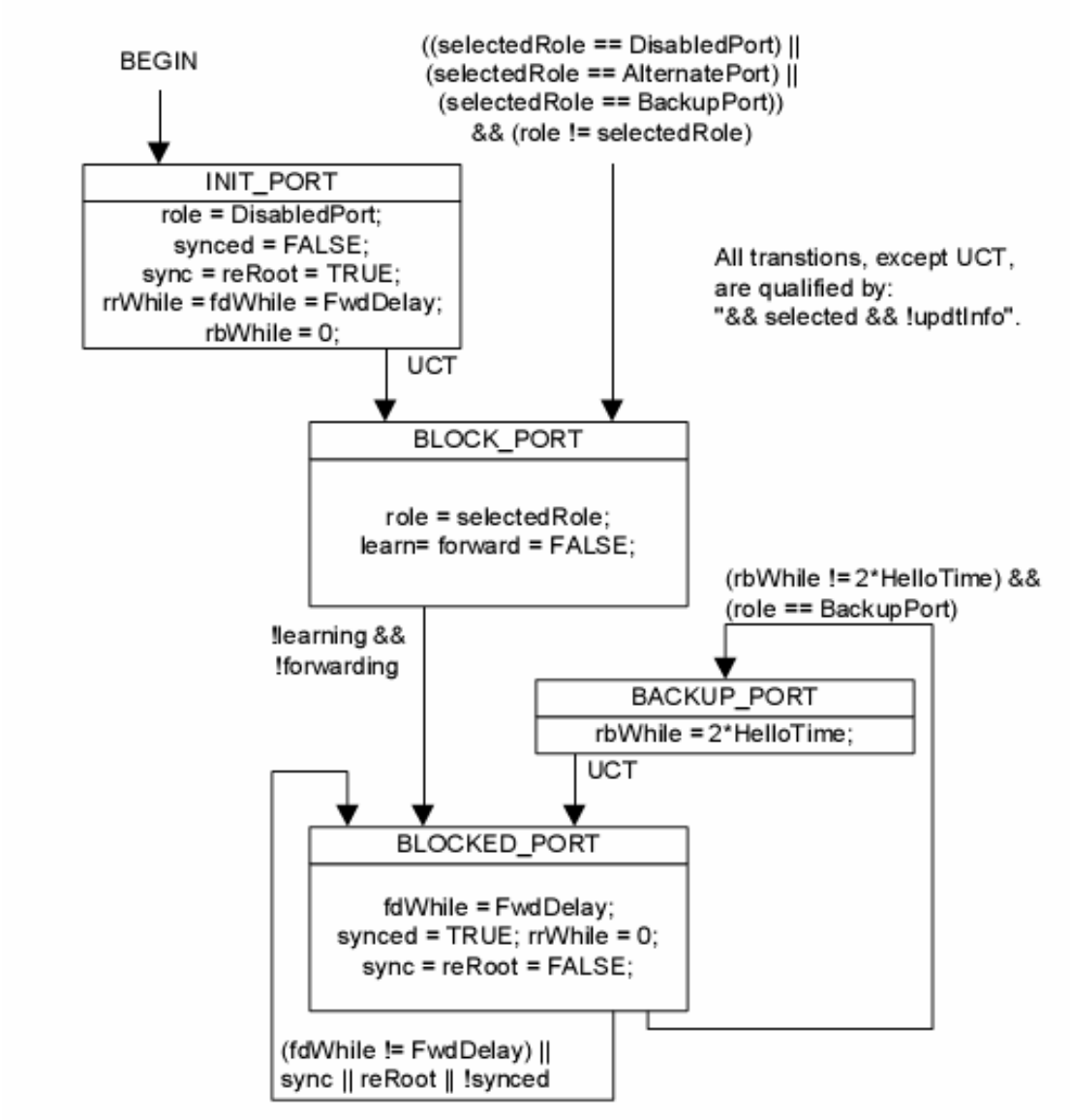


图 3-6 Port Role Transition: Disabled, Alternate, and Backup Role

状态机首先进入 INIT_PORT 状态,设置端口角色为 DisablePort, 设置完变量后进入 BLOCK_PORT 状态.如果 learning=forwarding=FALSE, 进入 BLOCKED_PORT 状态.

在 BLOCKED_PORT 状态.rrWhile=0, fdWhile=FwdDelay. 当 fdWhile!=FwdDelay 时,重新进入 BLOCKED_PORT 状态,确保 fdWhile=FwdDelay. 当(rbWhile!=2 * HelloTime) && (role== BackupPort)时,进入 BACKUP_PORT.

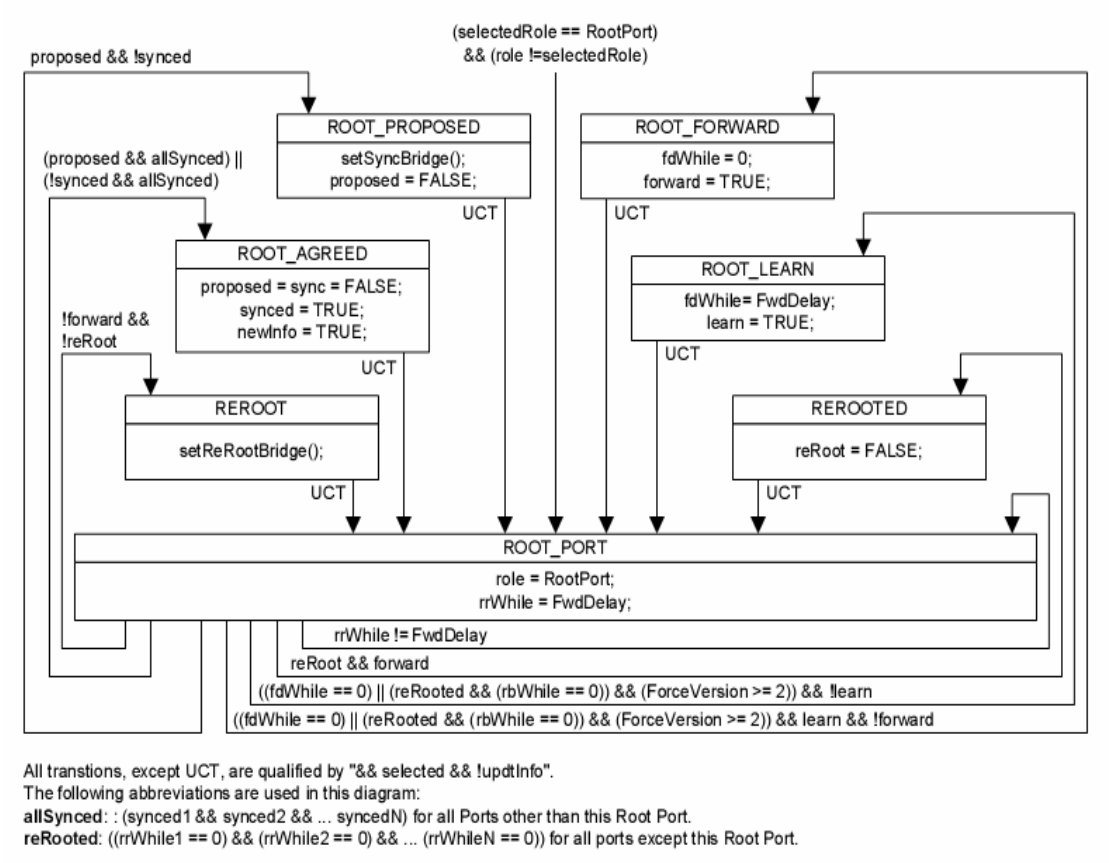


图 3-7 Port Role Transition: Root Port Role

当 Port Role Section State Machine 设置 selcetedRole=RootPort 时,ROOT_PORT 状态可从图 3-6 和图 3-8 的任何一个状态进入.如果端口以前的角色是 DisablePort, BackupPort, AlternatePort, fdWhile=FwdDelay; 如果端口以前是 DesignatedPort, fwdWhile 的取值范围是 0 到 FwdDelay.当 rrWhile!=FwdDelay 时,重新进入 ROOT_PORT 状态.

当 learn=FALSE,且满足下面其中一个条件:

a) The fdWhile timer has expired; or

b) The rbWhile timer for this Port is zero, rrWhile is zero for all Ports except this Root Port, and the protocol version selected by ForceVersion is version 2 or greater.

进入 ROOT_LEARN 状态.

当从指定网桥收到 Proposal=1 且!synced 时,进入 ROOT_PROPOSED 状态.

当(proposed && allSynced)||(!synced && allSynced),进入 ROOT_AGREED 状态.

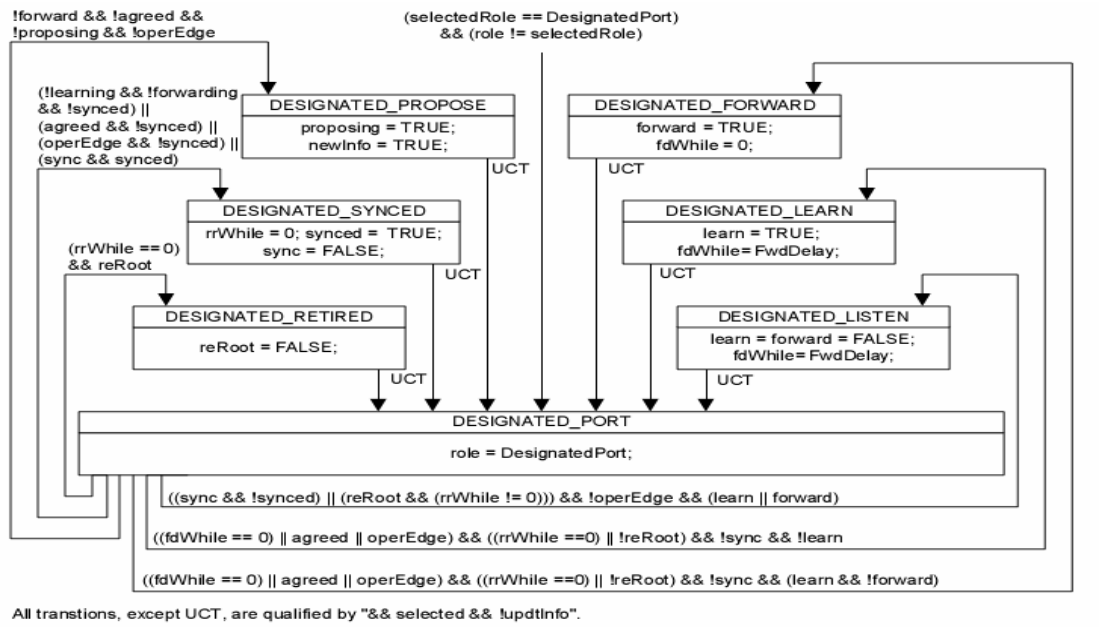


图 3-8 Port Role Transition: Designated Port Role

当 Port Role Selection State Machine 设置 selectedRole=DesignatedPort 后,DESIGNATED_PORT 状态可从图 3-6, 图 3-7 中的任一状态进入.也可以由图 3-8 的其它状态进入. 如果端口以前的角色是 DisablePort, BackupPort,AlternatePort, fdWhile=FwdDelay; 如果端口以前是 RootPort,fwdWhile 的取值范围是 0 到 FwdDelay.

Entry to the DESIGNATED_LISTEN state from the DESIGNATED_PORT state occurs if either learn or forward is TRUE, and the Port is not an edge Port, and either:

a) The rrWhile timer is running, and the current Root Port has requested that recent Root Ports revert to the Discarding Port State; or

b) The Port is not in agreement with current Spanning Tree information, and the current Root Port has instructed Designated Ports that are not in agreement with current Spanning Tree information to revert to the Discarding Port State.

The learn and forward variables are set FALSE to indicate to the Port State Transition state machine that the Port State should be set to Discarding, and the fdWhile timer is set equal to FwdDelay. Entry to the DESIGNATED_LEARN state from the DESIGNATED_PORT state occurs if learn and sync are both FALSE, and either rrWhile is not running or there is no outstanding request from the Root Port to retire recent Root Ports, and either:

c) The forwarding delay has expired; or

d) The Port is an edge Port; or

e) An agreement has been received in a BPDU from the Root Port of the Bridge attached to the LAN (agreed == TRUE).

Entry to the DESIGNATED_FORWARD state from the DESIGNATED_PORT state occurs if forward and sync are both FALSE, and learn is TRUE, and either rrWhile is not running or there is no outstanding request from the Root Port to retire recent Root Ports, and either:

f) The forwarding delay has expired; or

g) The Port is an edge Port; or

h) An agreement has been received in a BPDU from the Root Port of the Bridge attached to the LAN (agreed == TRUE).

Entry to the DESIGNATED_PROPOSE state from the DESIGNATED_PORT state occurs if the Port is not an edge Port, and forward is FALSE, and the Port is not in agreement with current Spanning Tree information, and the Port has not already sent a Proposal flag to the Bridge on the LAN to which it is connected.

Entry to the DESIGNATED_SYNCED state from the DESIGNATED_PORT state occurs if any of the following are true:

i) The Port is neither Learning nor Forwarding, but the Port is not indicating that it is in agreement with current Spanning Tree information (the synced variable is FALSE).

j) A response has been received from the Bridge on the LAN to which the Port is connected, indicating that the Port may proceed to the Forwarding Port State, but the Port is not indicating that it is in agreement with current Spanning Tree information.

k) The Port is an edge Port, but the Port is not indicating that it is in agreement with current Spanning Tree information.

l) The Root Port has requested this Port to establish agreement with current Spanning Tree information, and the Port is already in agreement with current Spanning Tree information.

3.8 Topology Change State Machine

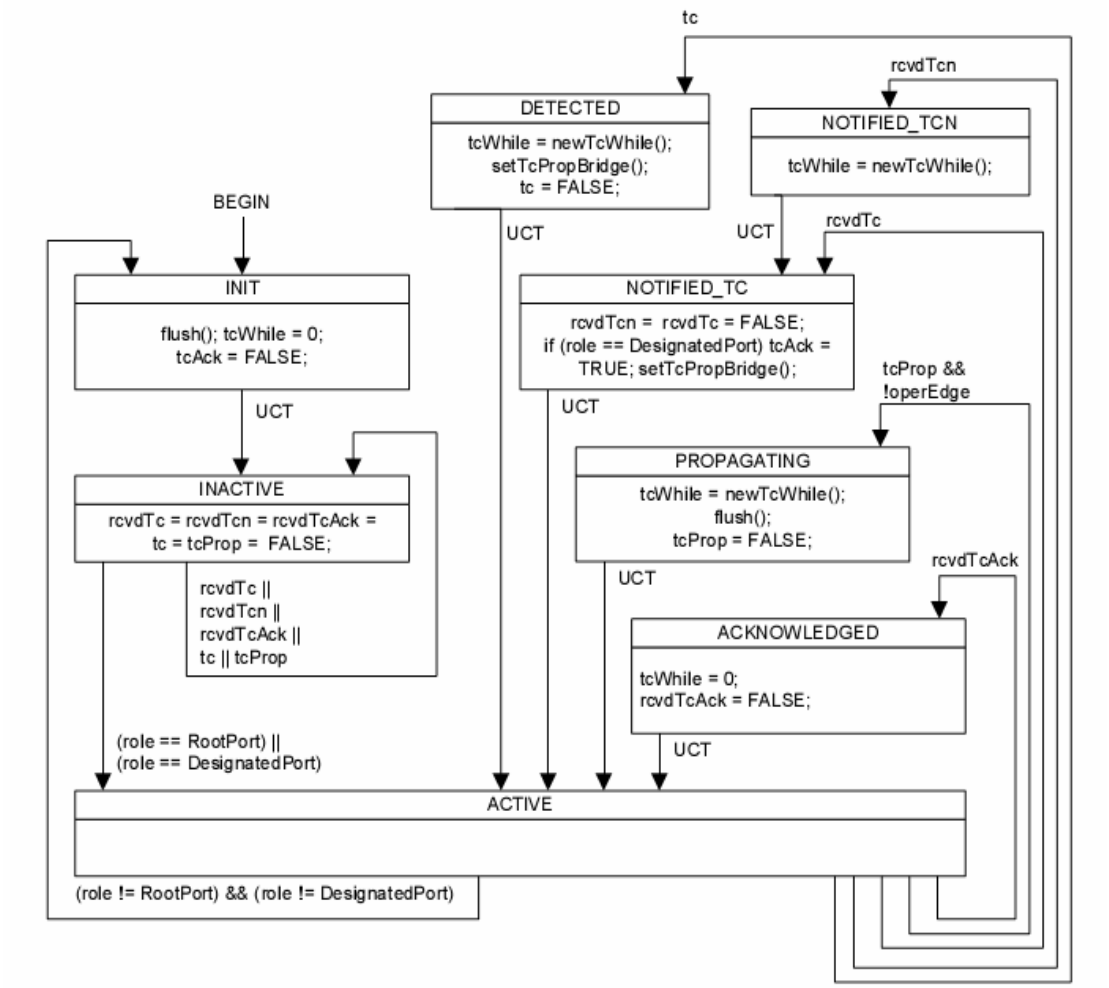


图 3-9 Topology Change State Machine

当根端口收到 TC=1 时,设置 TcWhile,向其它非边缘端口发送 TC=1.

当指定端口收到 TC=1 时,设置 TcACK(txRstp()又会设置为 0),设置 tcWhile,向其它非边缘端口发送 TC=1.

当根端口收到 TcACK 时,设置 tcWhile=0.

当指定端口收到 TCN=1 时,设置 tcACK, 并设置这个端口的 tcWhile,开始向下流发送 TC=1.

然后进入 NOTIFIED_TC,PROPAGATING 状态,设置其它非边缘端口的 tcWHile,并发送 TC=1.

所有非边缘端口收到 TC=1 后,FLUSH MAC TABLE.

当端口角色既不是根端口,也不是指定端口的时候,端口会变成 Discarding 状态,并且 FLUSH MAC TABLE.

3.9 Port Protocol Migration State Machine

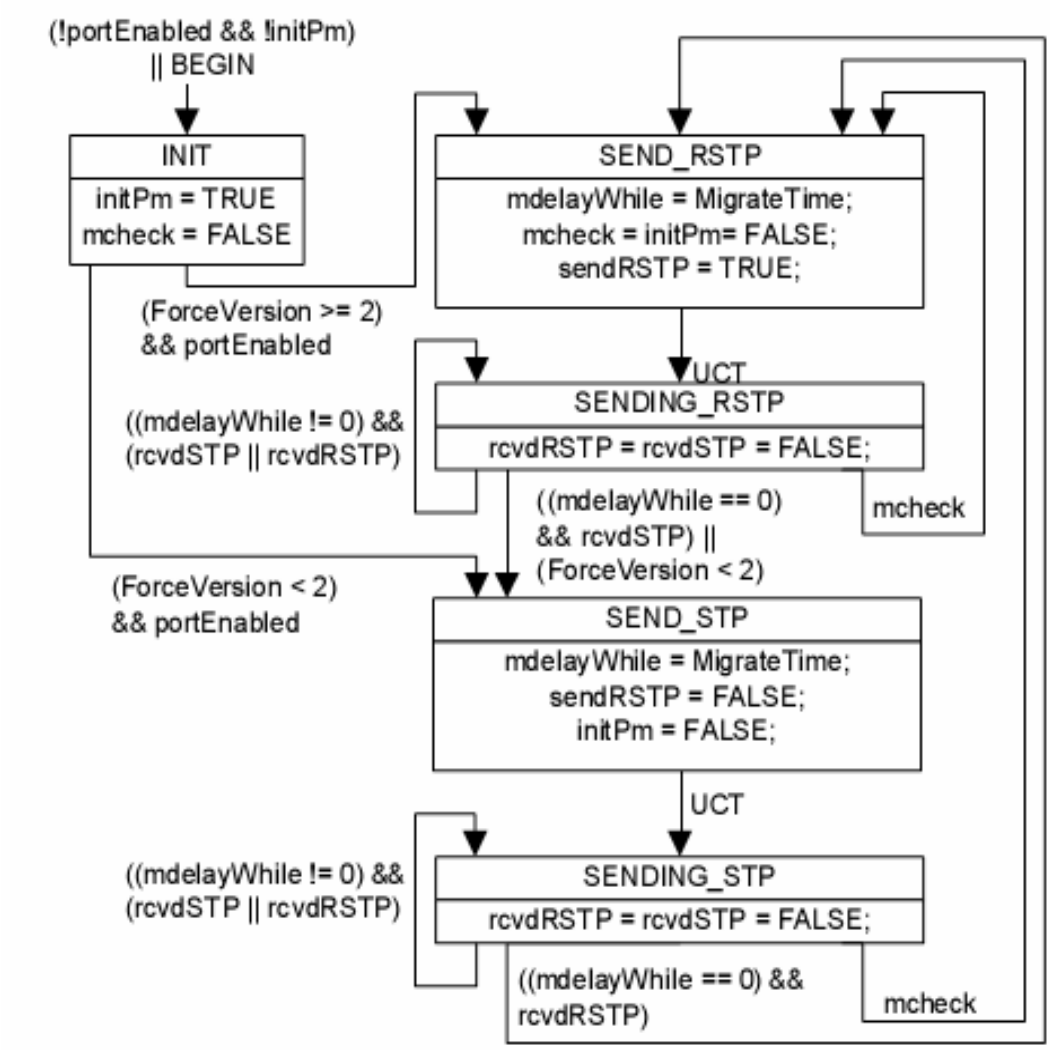


图 3-10 Port Protocol Migration State Machine

当(ForceVersion >= 2) && portEnabled 时,端口发送 RSTP BPDU.

当(ForceVersion < 2) && portEnabled 时,端口发送 STP BPDU.

端口发送 RSTP BPDU, 如果(mdelayWhile==0 && rcvdSTP), 则端口变成发送 STP BPDU.

端口发送 STP BPDU, 如果(mdelayWhile==0 && rcvdRSTP && ForceVersion>= 2), 则端口变成发送 RSTP BPDU.

当 mdelayWhile!=0 时,状态机忽略所有 BPDU.

当 ForceVersion >=2, mcheck=TRUE 时,端口发送 RSTP BPDU.

3.10 Port Transmit State Machine

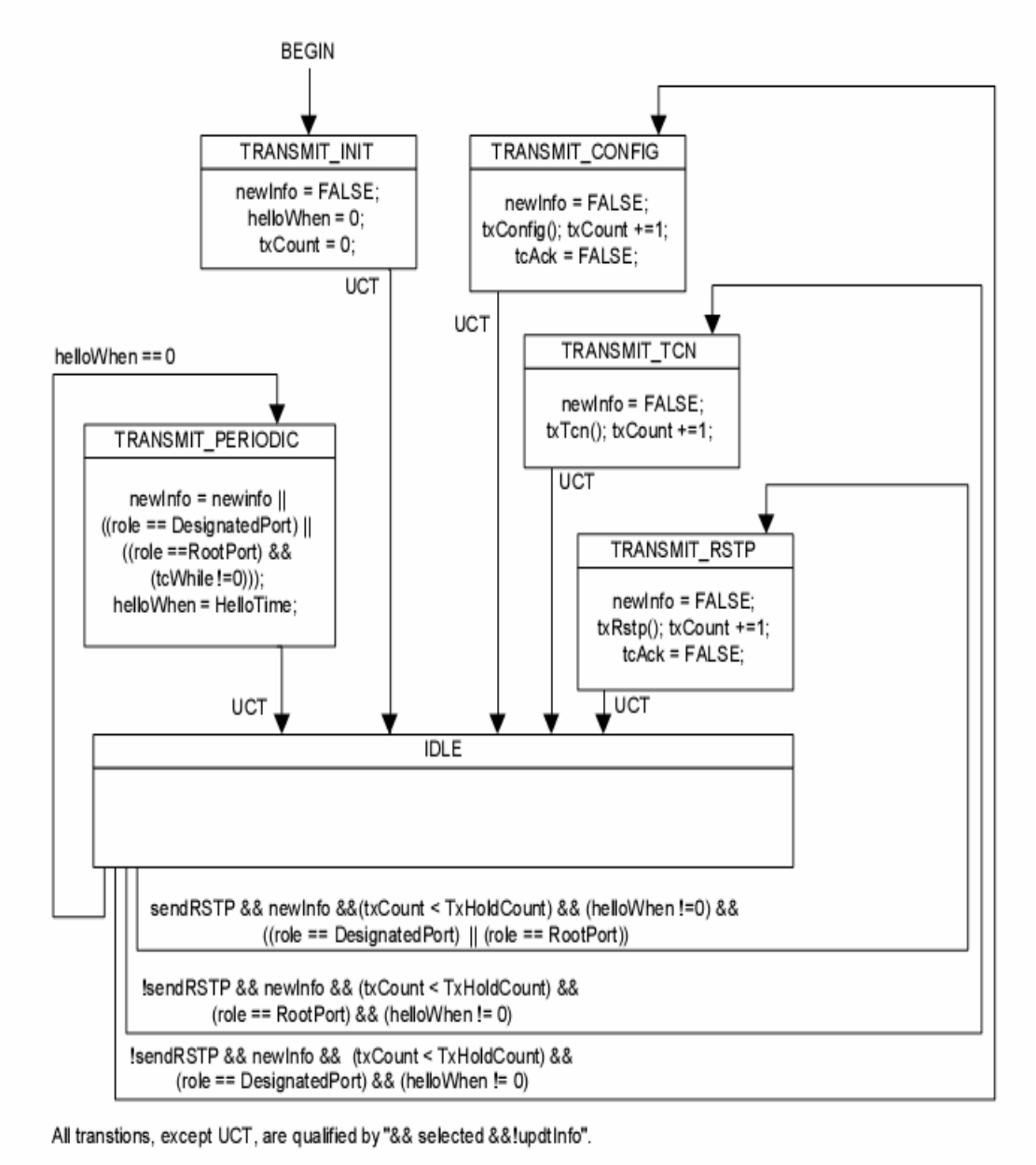


图 3-11 Port Transmit State Machine

正常情况下,每个 Hello Time 时间内发送一个 BPDU.

每发送一个 BPDU, txCount+1, 每过 1 秒, Port Timers State Machine 使 txCount-1.

当 txCount=TxHoldCount 时,不会再发送 BPDU.

3.11 Handshake

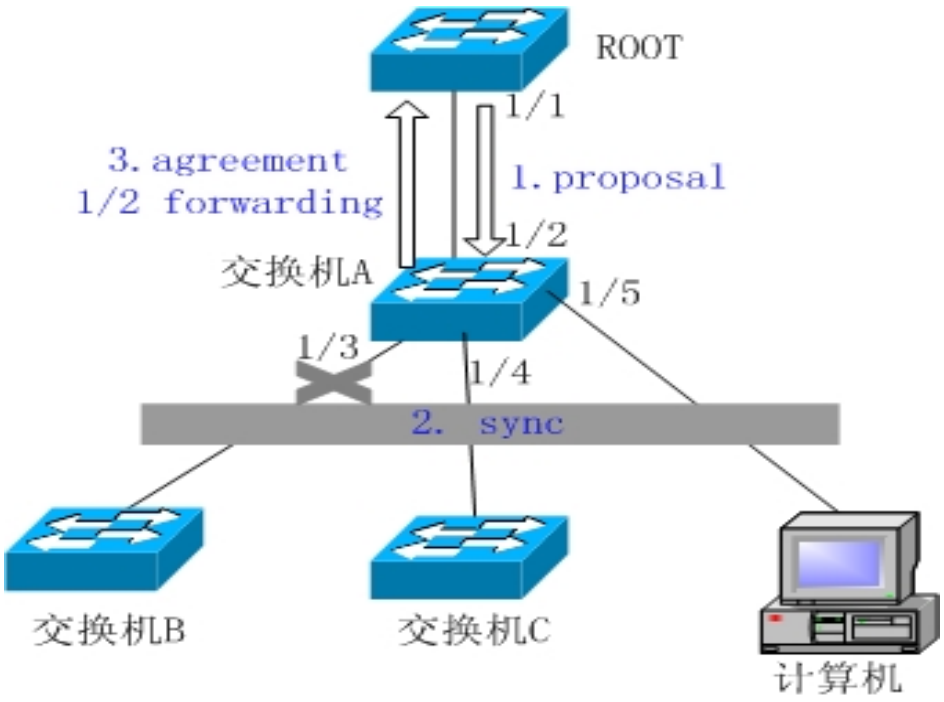


图 3-12 Handshake

如图 3-12 所示,最上面的交换机是 ROOT BRIDGE, 1/1, 1/4, 1/5 是指定端口,1/2 是根端口,1/3 是替换端口,1/5 是边缘端口.

当新增 1/1---1/2 这条链路时,1/1, 1/2 都处于 Discarding 状态.

第一步,指定端口发送 Proposal=1,即 1/1 发送 Proposal=1.(1/1, 1/2 都认为自己是指定端口,都发 Proposal=1,谁先到达对方是随机的.1/1 收到 1/2 的 BPDU 后,忽略.1/2 收到 1/1 的 BPDU 后,成为根端口,进入第二步).

第二步,交换机 A 开始 sync,即迫使其它端口 sync=1.当端口是边缘端口,或者端口处于 Discarding 状态时,sync=1.如果有端口处于 Forwarding 状态,状态机迫使它进入 Discarding 状态.在图 3-12 中,1/3 处于 Discarding 状态,1/5 是边缘端口,Handshake 对它们没有任何影响.1/4 处于 Forwarding 状态,状态机使 1/4 进入 Discarding 状态,sync=1.

第三步,其它所有端口 sync=1 后,1/2 进入 Forwarding 状态,并发送 Agreement=1.

第四步,指定端口 1/1 收到 Agreement=1 后,进入 Forwarding 状态.此时,交换机 A 的 1/4 处于 Discarding 状态,它的操作与原来 ROOT 的 1/1 一样,开始发送 Proposal=1,交换机 C 则与原来的交换 A 一样,先 sync 其它端口,然后发送 Agreement.如此循环.

4. 典型案例

4.1 链路备份

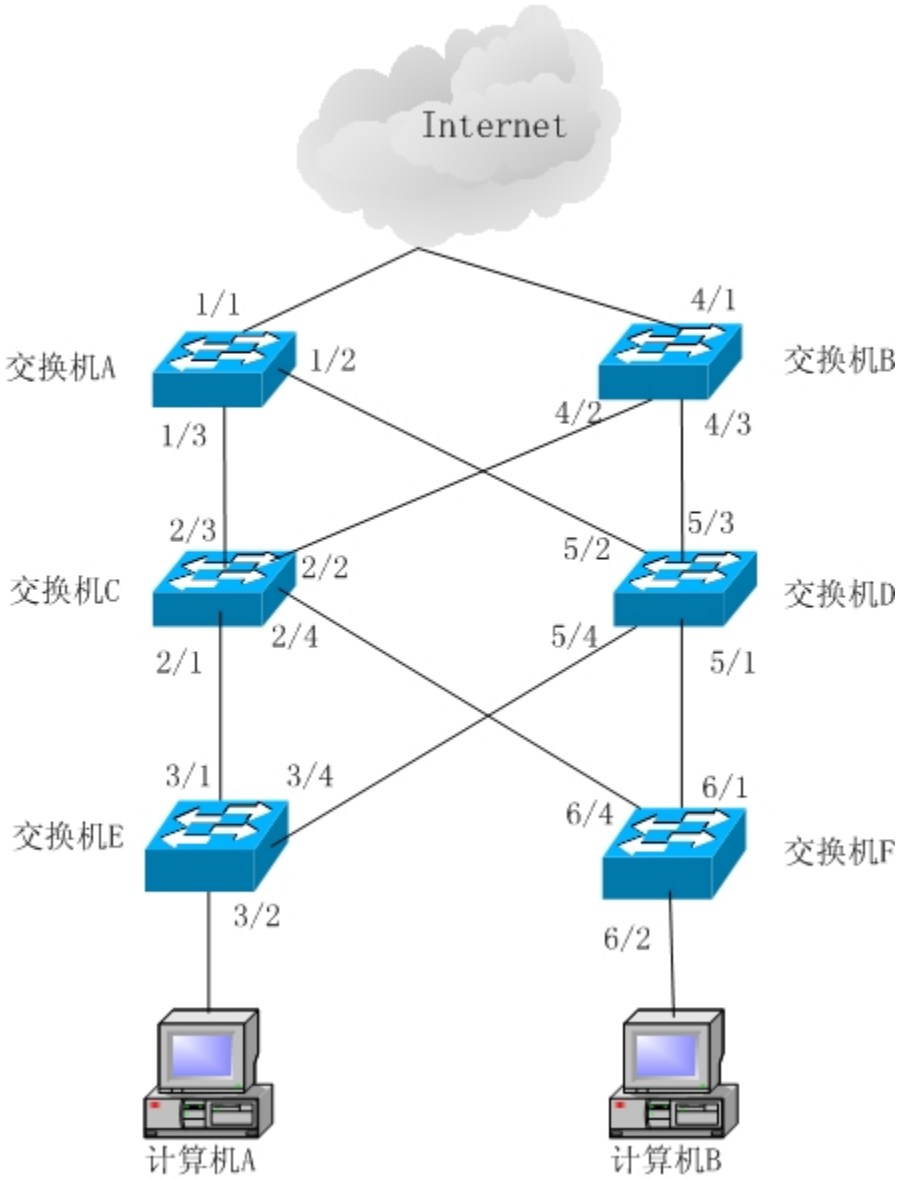


图 4-1 链路备份组网示意图

如图 4-1,交换机 A 和交换机 B 作为核心层,连接 Internet.交换机 C 和交换机 D 作为汇聚层,交换机 E 和交换机 F 作为接入层,连接用户终端设备.全部交换机运行 RSTP(交换机 E 和交换机 F 不支持生成树协议对组网功能没有影响). 在实际组网中,汇聚层和核心层常用千兆,万兆链路.为简单起见,图 4-1 所有链路都是百兆以太链路.

组网配置有三个注意事项:

a) RSTP 组网最重要的配置就是网桥优先级的配置.因为根网桥在拓扑中的特殊地位,有大量数据都需要经过根网桥.需要把处理能力强的高端交换机的优先级设置为全网最低,以使之成为根网桥.

b) 还有一个要特别注意的就是边缘端口的配置.对于与终端相连的交换机端口,需要设置为边缘端口,不仅可以加快拓扑收敛,还可以大量减少拓扑变化次数.一般不要手动配置端口为非边缘端口,没有任何益处,只会带来不良后果(延迟拓扑收敛,增加拓扑变化次数).

c) 在和 STP 组网的时候,还要特别注意链路代价(Port Cost),由于对于相同的链路,STP 的 cost 比 RSTP 的 cost 小,所以会优先选择 STP, 甚至会弃用 RSTP 的千兆链路而选择 STP 的百兆链路. 而 STP 的收敛速度慢,而且生产时间早,性能比 RSTP 差.建议修改 STP 的 port cost,百兆链路修改为 200 100,千兆链路修改为 20 100.这样在同样链路情况下,会优先选择 RSTP.

在图 4-1 中,对交换机 A 和交换机 B 的网桥优先级进行设置,A:0, B:4096.确保 ROOT 是交换机 A 和交换机 B 中的一个.在其它参数都是默认值.

在所有交换机互发 BPDU 形成稳定拓扑后,根端口,指定端口都工作在 Forwarding 状态,替换端口,备份端口都工作在 Discarding 状态.整个网络拓扑是一棵树,没有环路.

计算机 A 和计算机 B 各有 4 条链路可到达 Internet. 这 4 条链路互相备份,任一时刻只有一条链路在转发数据.计算机 A 和计算机 B 的环境一样,以计算机 A 为例.

计算机 A 通过 3/2---3/1---2/1---2/3---1/3---1/1 访问 Internet.

如果 3/1---2/1---2/3---1/3 发生故障,

计算机 A 自动通过 3/2---3/4---5/4---5/2---1/2---1/1 访问 Internet.

如果 1/3---1/1 或者交换机 A 发生故障,

计算机 A 自动通过 3/2---3/1---2/1---2/2---4/2---4/1 访问 Internet.

如果 1/3---1/1 或者交换机 A 发生故障且 3/1---2/1---2/2---4/2 发生故障,

计算机 A 自动通过 3/2---3/4---5/4---5/3---4/3---4/1 访问 Internet.

通过手动配置网桥优先级和端口代价,可以选择上面任意一条链路工作.

4.2 版本兼容

RSTP 能够识别 STP BPDU, STP 不能识别 RSTP BPDU.

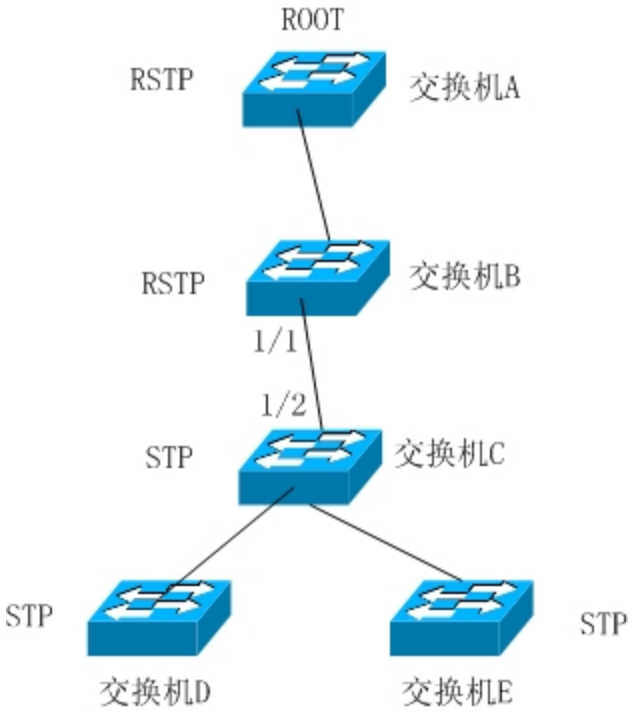


图 4-2 RSTP 与 STP 组网示意图

如图所示,交换机 A 和交换机 B 运行 RSTP,交换机 C, 交换机 D 和交换机 E 运行 STP.交换机 B 的 1/1 端口收到 1/2 的 STP BPDU 后,自动发送 STP BPDU,交换机 B 的其它端口还是发送 RSTP BPDU.所有交换机都认为交换机 A 是根网桥.当交换机 D 或者交换机 E 监测到拓扑变化时,向交换机 C 发送 TCN, 交换机 C 收到 TCN 后,回发 ACK,并向交换机 B 发送 TCN. 交换机 B 收到 TCN 后,向交换机 A 发送 TC,并回发 TC,交换机 C 收到 TC 后,向所有端口发送 TC.

5. 注意事项

5.1 Protocol Design Requirements

a> 配置一些端口转发数据,另外一些端口丢弃数据,建立一棵全连通树.

b> 阻止环路,自动备份.

c> 快速收敛.

d> 拓扑可预计,可重现,可管理.

e> 对终端透明.

f> RSTP 协议包占很小带宽,而且与其它流量独立.

g> 端口需要的内存独立于网络或者网桥的数量.

h> 网桥加入网络前不需要配置.

i> 由点对点链路组成的网络,建立稳定拓扑的时间与协议时间参数无关.

5.2 Symmetric Connectivity

在拓扑图中,终端 A 到终端 B 的链路与终端 B 到终端 A 的链路是相同的.

5.3 Temporary Loops

RST 算法和协议不能防止由不支持 MAC Internal Sublayer Service 的设备(如中继器,HUB)形成的环路.

5.4 Root

根网桥没有 Root Port, Root Path Cost=0.

5.5 Root Bridge Selction

拥有 best Bridege Identifier 的网桥成为 Root Bridge.

5.6 Root Port Selection

拥有最小 Root Path Cost 的端口被选为 Root Port, 如果有多个 Port 有同样的 Root Path Cost, 则拥有 best Port Identifier 的端口成为 Root Port.除了 Root Bridge, 其它网桥有且只有一个 Root Port.

5.7 Designated Bridge Selction

在每个 LAN, 拥有 best Root Path Cost 的网桥成为 Designated Bridge, 如果有多个网桥有相同的 best Root Path Cost, 则拥有 best Bridge Identifier 的网桥成为 Desiganted Bridge, Designated Bridge 与 LAN 相连的 Port 成为 Designated Port, 如果有多个 Port 与 LAN 相连,则拥有 best Port Identifier 的端口成为 Designated Port.在收到其它 BPDU 前,网桥默认是 Root Bridge, 所有端口都是 Designated Port.每个 LAN 有且只有一个 Designated Brdige 和 Designated Port.

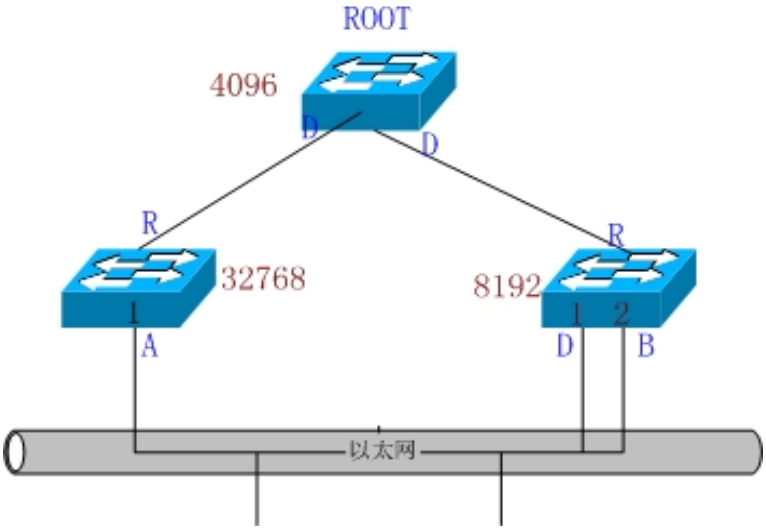


图 5-1 端口角色选举示意图

5.8 STP compatibility

RSTP 可以自动兼容 STP, 不需要任何配置.STP 不识别 RST BPDU. 与 STP 相连的端口发送 Configuration BPDU, Topology Change Notification, 其它端口不受影响.

5.9 Updating learned station information

a> 为了避免频繁的地址学习和广播风暴,MAC 地址在 MAC 表中需要保留较长一段时间.

b> 终端移动后,需要重新学习地址.拓扑发生变化后,即使终端没有移动,从网桥端口的角度看,终端位置发生了变化.原来可以到达路径变成不可达,到终端走新的路径.

c> 网桥端口收到 TCN 后,FLUSH 其它非边缘端口的 MAC 项,并且发送 TCN.

5.10 Protocol Version

如果交换机支持协议版本 A,当它接收到协议版本 B 时,

a) 如果 B >= A, 交换机把 BPDU 当作版本 A, 具体如下:

1) 版本 A 中已经定义的所有参数按照版本 A 处理.

2) 版本 A 中没有定义的参数全部忽略.

3) 超过版本 A 规定的字节数后的内容全部忽略.

b) 如果 B < A, 交换机把 BPDU 当作版本 B, 具体如下:

1) 版本 B 中已经定义的所有参数按照版本 B 处理.

2) 版本 B 中没有定义的参数全部忽略.

3) 超过版本 B 规定的字节数后的内容全部忽略.

当交换机只支持 STP 时,会忽略收到的所有 RSTP BPDU.

当交换机支持 RSTP 时,收到 STP BPDU 后,可以识别所有参数.

当交换机支持 STP,RSTP, 且端口工作在 STP 时,收到 RSTP BPDU 后,可以识别所有参数.

5.11 Port Role---Unknown

一般说明,交换机不会发出 Port Role 为 Unknown 的 BPDU,但是交换机可以接收.交换机接收到 Port Role 为 Unkown 的 BPDU 时,会把它当作配置(Configration)BPDU.

5.12 Configuration Message and TCN Message

Configuration Message: Configuration BPDU, RST BPDU.

TCN Message: TCN BPDU, RST BPDU with TC=1.

6．疑难解答

6.1 优先级设置

在 802.1D 1998 中,网桥优先级是 16 位,端口优先级是 8 位.而在 802.1w 2001 中,网桥优先级必须是 4096 的倍数,端口优先级必须是 16 的倍数.原因如下:

a) 在 802.1Q Multiple Spanning Trees Protocol(MSTP)中,用 12 位的系统 ID(System ID)作为一个 VLAN 的虚拟网桥 ID. 考虑到 16 个优先级已经够用,于是把优先级最重要的 4 位保留,其余 12 位用来存放系统 ID. 为了与 RSTP 等兼容,考虑到管理方面的因素,仍然认为网桥优先级是 16 位,只是低 12 位全是 0.

b) 端口 ID 一共 16 位,端口编号只有 8 位,一个交换机只能有 255 个端口,对于端口数量高速增加的交换机来说是不够用的.考虑到 16 个端口优先级已经够用,于是把优先级中最重要的 4 位保留,其它 4 位用来存放端口编号. 为了与 RSTP 等兼容,考虑到管理方面的因素,仍然认为端口优先级是 8 位,只是低 4 位全是 0.

6.2 端口角色选择

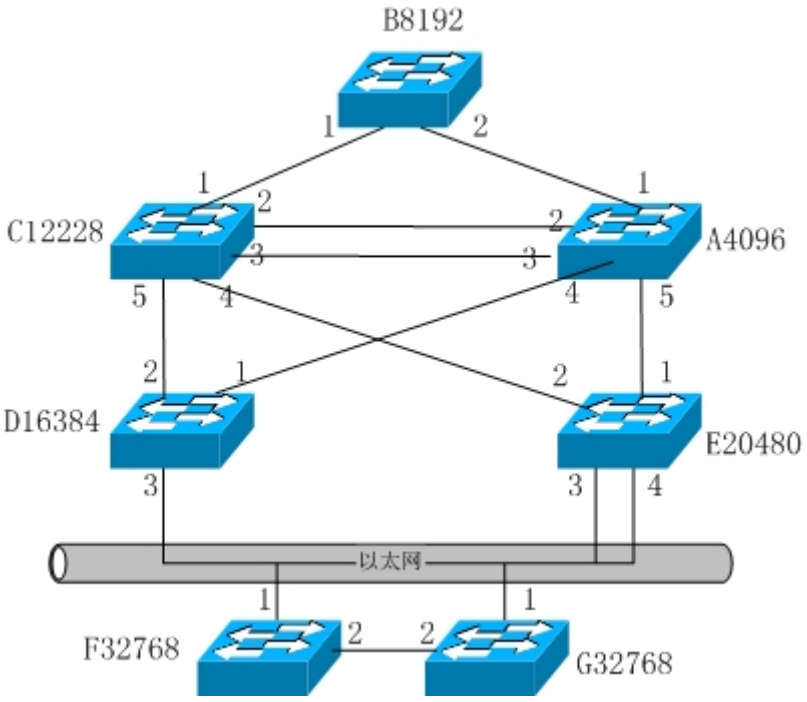


图 6-1 端口角色选择初始状态示意图

A4096 代表交换机 A 的优先级是 4096,其它参数都是默认值.所有链路都是百兆以太链路.

根据Root Bridge Selction, Root Port Selection, Designated Bridge Selction, 备份端口(Backup Port), 替换端口(Alternate Port)介绍的理论,各端口角色如图 6-2.

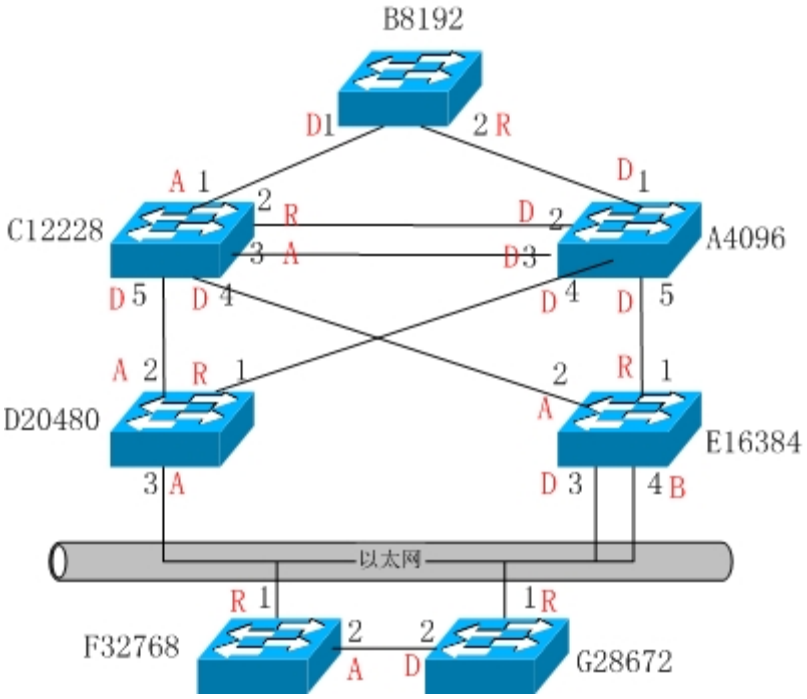


图 6-2 各端口角色示意图

最终形成的拓扑图见图 6-3.

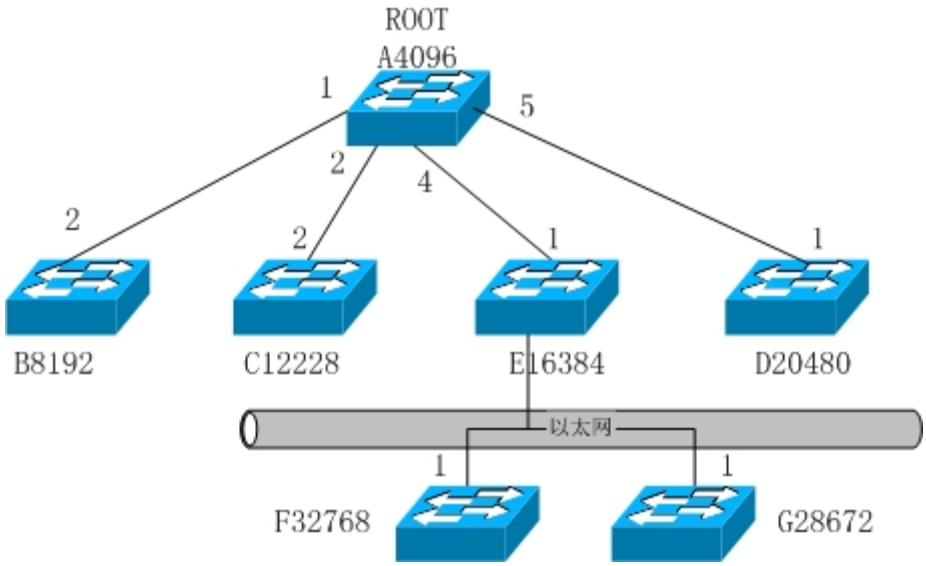


图 6-3 最终拓扑示意图