

# 了解多生成树协议 (802.1s)

## 目录

[简介](#)

[使用 MST 的位置](#)

[PVST+ 案例](#)

[标准 802.1q 案例](#)

[MST 案例](#)

[MST 区域](#)

[MST 配置和 MST 区域](#)

[区域边界](#)

[MST 实例](#)

[IST 实例](#)

[MSTIs](#)

[常见配置错误](#)

[所有端口 \( 中继或接入 \) 上的 IST 实例都是活动的](#)

[两个 VLAN 映射到同一实例将阻塞相同端口](#)

[MST 区域和外界之间的交互作用](#)

[推荐的配置](#)

[备选配置 \( 不推荐 \)](#)

[无效配置](#)

[迁移策略](#)

[结论](#)

[相关信息](#)

## 简介

多生成树 (MST) 是从 Cisco 专用多实例生成树协议 (MISTP) 实施得到的 IEEE 标准。本文档假定读者熟悉快速 STP (RSTP) (802.1w)，因为 MST 主要依赖于这一其他 IEEE 标准。下表显示了各种 Catalyst 交换机中对 MST 的支持：

Catalyst 平台	支持 RSTP 的 MST
Catalyst 2900 XL 和 3500 XL	离线
Catalyst 2950 和 3550	Cisco IOS 12.1(9)EA1
Catalyst 2955	所有 Cisco IOS 版本
Catalyst 2948G-L3 和 4908G-L3	离线
Catalyst 4000、2948G 和 2980G (Catalyst OS (CatOS))	7.1

Catalyst 4000 和 4500 (Cisco IOS)	12.1(12c)EW
Catalyst 5000 和 5500	离线
Catalyst 6000 和 6500 (CatOS)	7.1
Catalyst 6000 和 6500 (Cisco IOS)	12.1(11b)EX、12.1(13)E 和 12.2(14)SX
Catalyst 8500	离线

有关 RSTP (802.1w) 的详细信息，请参阅以下文档：

- [了解快速生成树协议 \(802.1w\)](#)

## 使用 MST 的位置

下图显示了一种常用设计，其特征是接入交换机 A 带 1000 个 VLAN，这些 VLAN 冗余连接到两台分布式交换机 D1 和 D2。在此设置中，用户连接到交换机 A，而网络管理员通常根据偶数或奇数 VLAN 或认为适当的任何其他方案，寻求实现接入交换机上行链路的负载均衡。

以下部分是对此设置使用各种类型的 STP 的案例示例：

### PVST+ 案例

在 Cisco 每 VLAN 生成树 (PVST+) 环境中，可以调整生成树参数，以便在每个上行链路中继上转发半数 VLAN。为轻松实现此目的，请选择网桥 D1 作为 VLAN 501 至 1000 的根网桥，选择网桥 D2 作为 VLAN 1 至 500 的根网桥。以下描述对于此配置是正确的：

- 在此案例中，可以获得最佳负载均衡结果。
- 为每个 VLAN 维护一个生成树实例，这意味着只为两个不同的最终逻辑拓扑维护 1000 个实例。这严重浪费了网络中所有交换机的 CPU 周期（除了每个实例发送各自的网桥协议数据单元 (BPDU) 所用的带宽以外）。

### 标准 802.1q 案例

原始 IEEE 802.1q 标准不仅仅定义了中继，还定义了许多其他内容。此标准定义了通用生成树 (CST)，后者仅假定整个桥接网络运行一个生成树实例，而不考虑 VLAN 的数量。如果 CST 适用于此图的拓扑，则结果将与下图类似：

在运行 CST 的网络中，以下描述是正确的：

- 无法实现负载均衡；需要针对所有 VLAN 阻塞一条上行链路。
- 可节省 CPU；只需计算一个实例。

**注意：** Cisco 实施增强了 802.1q 以便支持一个 PVST。此功能完全按照本示例中 PVST 的方式运行。Cisco 每 VLAN BPDU 通过纯 802.1q 网桥建立隧道。

### MST 案例

MST (IEEE 802.1s) 组合了 PVST+ 和 802.1q 的最佳优势。其观点在于，可以将若干个 VLAN 映射

到数量减少的生成树实例，因为大多数网络不需要太多逻辑拓扑。在第一个图描述的拓扑中，只有两种不同的最终逻辑拓扑，因此实际上只需要两个生成树实例。不需要运行 1000 个实例。如此图所示，如果将 1000 个 VLAN 中的一半映射到不同的生成树实例，则以下描述是正确的：

- 由于半数 VLAN 采用一个单独的实例，因此仍然可以实现所需的负载均衡方案。
- 由于只计算两个实例，因此可以节省 CPU。

从技术观点来看，MST 是最佳解决方案。从最终用户的角度来看，与迁移到 MST 相关联的主要缺点是：

- 该协议比常见的生成树更复杂，因此需要对员工进行额外培训。
- 与传统网桥的交互可能是一个难题。有关详细信息，请参阅本文档的 [MST 区域和外界之间的交互作用](#) 部分。

## [MST 区域](#)

如前所述，MST 引入的主要增强功能是可以将若干个 VLAN 映射到单个生成树实例。这样便会引发一个问题，即如何确定哪个 VLAN 将与哪个实例相关联。更准确地讲，如何标记 BPDU，以便接收设备识别每个设备适用的实例和 VLAN。

该问题在 802.1q 标准案例中是不相关的，在这一标准中，所有实例都映射到唯一的实例。在 PVST+ 实施中，该关联如下：

- 不同的 VLAN 传输各自实例的 BPDU ( 每个 VLAN 一个 BPDU ) 。

为解决该问题，Cisco MSTP 为每个实例发送了一个 BPDU ( 包括 BPDU 负责的 VLAN 列表 ) 。如果由于失误，两台交换机的配置不正确，并且同一实例相关的 VLAN 范围不相同，则协议很难从这种情形中正常恢复。

IEEE 802.1s 委员会采用了更容易且更简单的方法，该方法引入了 MST 区域。请设想某个区域相当于边界网关协议 (BGP) 自治系统，即接受常规管理的一组交换机。

## [MST 配置和 MST 区域](#)

在网络中运行 MST 的每台交换机具有由以下三个属性组成的单个 MST 配置：

1. 一个字母数字配置名称 ( 32 个字节 )
2. 配置修订版本号 ( 两个字节 )
3. 一个由 4096 个元素组成的表，用于将机箱上支持的每个潜在的 VLAN ( 4096 个 ) 关联到给定实例

若要成为通用 MST 区域的组成部分，交换机组必须共享相同的配置属性。网络管理员需要在整个区域中正确传播配置。当前，此步骤只能通过命令行界面 (CLI) 或简单网络管理协议 (SNMP) 来完成。由于 IEEE 规范没有明确指出如何完成此步骤，因此可以考虑采用其他方法。

**注意：** 如果由于任何原因两台交换机的一个或多个配置属性不同，则这两台交换机属于不同的区域。有关详细信息，请参阅本文档的 [区域边界](#) 部分。

## [区域边界](#)

为确保实现 VLAN 到实例的一致映射，协议必须能够准确识别区域的边界。为此，在 BPDU 中包括

了区域的特征。准确的 VLAN 到实例映射不会在 BPDU 中传播，因为交换机只需知道它们是否与邻居位于同一区域中。因此，只会发送 VLAN 到实例的映射表的摘要，以及修订版号和名称。交换机收到 BPDU 后，它将提取该摘要（通过数学函数从 VLAN 到实例映射表派生的数值），并将该摘要与自己的计算摘要进行比较。如果两个摘要不相同，则接收到 BPDU 的端口位于某个区域的边界。

通常而言，如果端口网段上的指定网桥位于不同区域，或者如果端口接收传统 802.1d BPDU，则该端口位于某个区域的边界。在下图中，B1 上的端口位于区域 A 的边界，而 B2 和 B3 上的端口位于区域 B 的内部：

## MST 实例

根据 IEEE 802.1s 规范，MST 网桥必须至少能够处理下面两个实例：

- 一个内部生成树 (IST)
- 一个或多个多个生成树实例(MSTIs)

由于 802.1s 实际上处于准标准阶段，因此术语将继续不断发展。这些名称有可能会在 802.1s 的最终版本中进行更改。Cisco 实施支持 16 个实例：一个 IST (实例 0) 和 15 个 MSTI。

## IST 实例

为清楚地了解 IST 实例的角色，请记住 MST 源自 IEEE。因此，MST 必须能够与基于 802.1q 的网络交互，这是因为 802.1q 是另一个 IEEE 标准。对于 802.1q，桥接网络仅实施单个生成树 (CST)。IST 实例只是一个在 MST 区域内扩展 CST 的 RSTP 实例。

IST 实例接收 BPDU 并将其发送到 CST。IST 可将整个 MST 区域表示为通向外界的 CST 虚拟网桥。

以下两个图的功能相当。注意各个阻塞端口的位置。在典型的桥接网络中，您期望看到交换机 M 和 B 之间的阻塞端口。除了在 D 上阻塞之外，您还期望在 MST 区域中的某个位置由阻塞端口中断第二个环路。但是由于 IST，整个区域显示为一个运行单个生成树 (CST) 的虚拟网桥。这样，就可以了解虚拟网桥阻塞 B 上的替代端口。而且，该虚拟网桥位于 C 到 D 网段，将会导致交换机 D 阻塞其端口。

使区域显示为一个虚拟 CST 网桥的具体机制不在本文档讨论范围内，IEEE 802.1s 规范中对该机制进行了详尽描述。然而，如果牢记 MST 区域的此虚拟网桥属性，将更容易理解与外界的交互作用。

## MSTIs

MSTI 是只位于区域内的简单 RSTP 实例。默认情况下，这些实例会自动运行 RSTP，而无需执行任何额外配置工作。与 IST 不同，MSTI 从不与区域外部进行交互。请记住，MST 只在区域外运行一个生成树，因此除 IST 实例以外，区域内部的常规实例都没有外部对等实例。另外，MSTI 不会在区域外部发送 BPDU，只有 IST 才会在区域外部发送 BPDU。

MSTI 不发送独立的单个 BPDU。在 MST 区域内部，网桥会交换可视为 IST 的正常 RSTP BPDU 的 MST BPDU，同时包含每个 MSTI 的其他信息。下图显示了 MST 区域内部的交换机 A 和 B 之间的 BPDU 交换。每台交换机只发送一个 BPDU，但包括位于端口上的每个 MSTI 的一个 Mrecord。

**注意：**在此图中，请注意 MST BPDU 传输的第一个信息字段包含有关 IST 的数据。这意味着 IST (实例 0) 总是位于 MST 区域内的各个位置。然而，由于网络管理员不必将 VLAN 映射到实例 0 上，因此不用担心此问题。

与常规的收敛生成树拓扑不同的是，链路的两端可以同时发送和接收 BPDU。如下图所示，这是因为每个网桥可以指定给一个或多个实例，并需要传输 BPDU。一旦在端口上指定单个 MST 实例，将立即发送包含所有实例 (IST + MSTI) 信息的 BPDU。下图演示了在 MST 区域内部和外部发送的 MST BPDU。

MRecord 包含足够的信息（主要是根网桥和发送者网桥优先级参数），可供对应的实例计算其最终拓扑。MRecord 不需要与计时器相关的任何参数（如 hello 时间、转发延迟和最大时限），这些参数通常位于常规的 IEEE 802.1d 或 802.1q CST BPDU 中。MST 区域中使用这些参数的唯一实例是 IST；hello 时间可确定 BPDU 的发送频率，转发延迟参数主要用于无法进行快速转换的情况（记住快速转换不会发生在共享链路上）。由于 MSTI 依靠 IST 传输其信息，因此 MSTI 不需要那些计时器。

## 常见配置错误

实例和 VLAN 之间保持独立是一个新概念，这意味着您必须仔细规划配置。[所有端口（中继或接入）上的 IST 实例都是活动的](#)部分说明了一些常见的缺陷以及如何避免它们。

## 所有端口（中继或接入）上的 IST 实例都是活动的

下图显示了使用接入端口（分别位于不同的 VLAN）连接的交换机 A 和交换机 B。VLAN 10 和 VLAN 20 映射到不同的实例。VLAN 10 映射到实例 0，而 VLAN 20 映射到实例 1。

此配置将导致 pcA 无法向 pcB 发送帧。**show** 命令表明交换机 B 要阻塞 VLAN 10 中指向交换机 A 的链路，如下图所示：

在此类没有明显循环的简单拓扑中，这怎么可能？

此问题通过以下事实进行解释：无论内部实例的数量是多少，仅使用 1 个 BPDU (IST BPDU) 传送 MST 信息。单个实例不发送单个 BPDU。当交换机 A 和交换机 B 交换 VLAN 20 的 STP 信息时，这两台交换机将为实例 1（因为这是映射 VLAN 20 的位置）发送包含 MRecord 的 IST BPDU。不过，由于它是 IST BPDU，因此该 BPDU 还包含实例 0 的信息。这意味着，无论 MST 区域内的所有端口是否传输映射到 IST 实例的 VLAN，IST 实例在这些端口上都是活动的。

下图显示了 IST 实例的逻辑拓扑：

交换机 B 从交换机 A 接收实例 0 的两个 BPDU（每个端口一个 BPDU）。很明显，交换机 B 必须阻塞其中一个端口，以避免循环。

首选解决方案是对 VLAN 10 使用一个实例，对 VLAN 20 使用另一个实例，从而避免将 VLAN 映射到 IST 实例。

备用解决方案是在所有链路上传输那些映射到 IST 的 VLAN（允许在两个端口上传输 VLAN 10，如[此图](#)所示）。

## 两个 VLAN 映射到同一实例将阻塞相同端口

请记住，VLAN 不再意味着生成树实例。拓扑取决于实例，而不管映射到该实例的 VLAN 如何。下图显示了一个问题，该问题是在[所有端口（中继或接入）上的 IST 实例都是活动的](#)部分中讨论的问题的另一种形式：

假定 VLAN 10 和 20 都映射到同一实例（实例 1）。网络管理员希望以手动方式在一条上行链路上修剪 VLAN 10，在另一条上行链路上修剪 VLAN 20，以便限制上行链路中继上从交换机 A 到分布式交换机 D1 和 D2 的流量（尝试实现上图所述的拓扑）。在完成此操作后，网络管理员会立即注意到 VLAN 20 中的用户已经与网络断开连接。

这是一个典型的配置错误问题。VLAN 10 和 20 都映射到实例 1，这意味着两个 VLAN 只有一个逻辑拓扑。无法实现负载共享，如此处所示：

由于手动修剪，因此在阻塞端口上只允许使用 VLAN 20，这就是连接断开的原因。为了实现负载均衡，网络管理员必须将 VLAN 10 和 20 映射到两个不同的实例。

要避开此问题，采用的简单规则是绝不从中继中手动修剪 VLAN。如果您决定从中继中删除某些 VLAN，请同时删除映射到给定实例的所有 VLAN。请勿从中继中删除单个 VLAN，也不要删除映射到同一实例的所有 VLAN。

## MST 区域和外界之间的交互作用

迁移到 MST 网络后，管理员可能必须处理 MST 和传统协议之间的互操作性问题。MST 可与标准的 802.1q CST 网络无缝交互操作；然而，由于网络的单个生成树限制，因此只有少数网络基于 802.1q 标准。Cisco 在宣布支持 802.1q 的同时发布了 PVST+。Cisco 还提供了 MST 和 PVST+ 之间的一种高效而简单的兼容性机制。此机制将在本文档后面进行介绍。

MST 区域的第一个属性是在边界端口不发送任何 MSTI BPDU，而只发送 IST BPDU。内部实例 (MSTI) 在边界端口总是自动采用 IST 拓扑，如下图所示：

在上图中，假定 VLAN 10 至 50 映射到绿色实例，该实例只是一个内部实例 (MSTI)。红色链路表示 IST，因此也表示 CST。拓扑中的各个位置都允许使用 VLAN 10 至 50。不会在 MST 区域外部发送绿色实例的 BPDU。这并不意味着 VLAN 10 至 50 中存在环路。MSTI 在边界端口采用 IST，并且交换机 B 上的边界端口还会阻塞绿色实例的流量。

运行 MST 的交换机能够自动在边界检测到 PVST+ 邻居。这些交换机能够检测到在中继端口的不同 VLAN 上接收到实例的多个 BPDU。

下图显示了一个互操作性问题。MST 区域仅与区域外的一个生成树 (CST) 进行交互。然而，PVST+ 网桥对每个 VLAN 运行一个生成树算法 (STA)，因此每两秒钟在每个 VLAN 上发送一个 BPDU。边界 MST 网桥不应接收太多 BPDU。MST 网桥应接收一个 BPDU 或发送一个 BPDU，具体取决于该网桥是否为 CST 的根网桥。

Cisco 建立了一种机制来处理上图中显示的问题。通过隧道传输 PVST+ 网桥在 MST 区域中发送的其他 BPDU 时，可能会产生这种可能性。然而实践证明，在 MSTP 中首次实施时，此解决方案过于复杂并且可能存在危险。因此，创建了一种更简单的方法。MST 区域复制所有 VLAN 上的 IST BPDU 来模拟 PVST+ 邻居。此解决方案包含在本文档中讨论的一些限制条件。

## 推荐的配置

由于 MST 区域现在复制边界的每个 VLAN 上的 IST BPDU，因此每个 PVST+ 实例都能从 IST 根获得 BPDU（这意味着根位于 MST 区域内部）。建议 IST 根的优先级高于网络中的其他任何网桥，以便 IST 根成为所有不同 PVST+ 实例的根，如下图所示：

在此图中，交换机 C 是冗余连接到 MST 区域的 PVST+。IST 根是交换机 C 上存在的所有 PVST+ 实例的根，因此，交换机 C 会阻塞其中一条上行链路以防止循环。在此特定情况下，PVST+ 和

MST 区域之间的交互是最佳的，这是因为：

- 可以调整交换机 C 的上行链路端口成本，以实现跨上行链路端口的不同 VLAN 的负载均衡（因为交换机 C 对每个 VLAN 运行一个生成树，能够选择阻塞每个 VLAN 上的哪个上行链路端口）。
- 可以在交换机 C 上使用 UplinkFast，以便在上行链路出现故障的情况下实现快速收敛。

## 备选配置（不推荐）

另一种可能性是让 IST 区域成为实例（绝不能是 PVST+）的根。这意味着所有 PVST+ 实例都拥有比 IST 实例更好的根，如下图所示：

此案例对应 PVST+ 核心和 MST 接入或分布层，这是一种很少见的情况。如果在区域外部建立根网桥，则与之前推荐的配置相比，存在以下缺点：

- MST 区域只运行一个与外界交互的生成树实例。这基本上意味着边界端口只能阻塞或转发所有 VLAN。换句话说，区域中通向交换机 C 的两条上行链路之间无法实现负载均衡。将针对所有 VLAN 阻塞实例的交换机 B 的上行链路，同时针对所有 VLAN 转发交换机 A。
- 此配置仍允许在区域内进行快速收敛。如果交换机 A 的上行链路发生故障，则需要快速切换到其他交换机的上行链路。尽管 IST 在区域内部运行以便使整个 MST 区域类似于 CST 网桥的方式不在本文详细讨论范围内，但您可以想象跨区域的切换绝不会像单个网桥上的切换那样高效。

## 无效配置

尽管 MST+ 仿真机制可在 MST 和 PVST+ 之间提供轻松的无缝互操作性，但这种机制意味着除前面提及的两种配置外的其他所有配置都是无效的。必须遵从以下基本规则，才能获得成功的 MST 和 PVST+ 交互：

1. 如果 MST 网桥是根网桥，该网桥必须是所有 VLAN 的根网桥。
2. 如果 PVST+ 网桥是根网桥，该网桥必须是所有 VLAN（包括总是在 VLAN 1 上运行的 CST，无论 CST 运行 PVST+ 时是否为本地 VLAN）的根网桥。
3. 如果 MST 网桥是 CST 的根网桥，而 PVST+ 网桥是一个或多个其他 VLAN 的根网桥，则模拟将失败并生成错误消息。失败的模拟会将边界端口置于根不一致模式。

在上图中，MST 区域中的网桥 A 是全部三个 PVST+ 实例（红色 VLAN 的实例除外）的根网桥。网桥 C 是红色 VLAN 的根网桥。假定在红色 VLAN 上创建的环路（网桥 C 是根网桥）被网桥 B 阻塞。这意味着，为红色 VLAN 之外的所有 VLAN 指定了网桥 B。MST 区域无法执行该操作。由于 MST 区域只对外界运行一个生成树，因此只能针对所有 VLAN 阻塞或转发边界端口。这样，当网桥 B 在其边界端口上检测到更好的 BPDU 时，它将调用 BPDU 防护以阻塞此端口。此端口将置于根不一致模式。这一完全相同的机制还会导致网桥 A 阻塞其边界端口。连接将会中断；然而，即使存在此类配置错误，也会保留无循环拓扑。

**注意：**当边界端口产生根不一致错误时，应立即调查 PVST+ 网桥是否已尝试成为某些 VLAN 的根网桥。

## 迁移策略

迁移到 802.1s/w 的第一步是正确识别点对点端口和边缘端口。确保需要进行快速转换的所有交换机对交换机链路都是全双工的。边缘端口可通过 Portfast 功能进行定义。请仔细决定交换网络中需要

的实例数目，并牢记实例将转换为逻辑拓扑。决定哪些 VLAN 要映射到这些实例上，并为每个实例仔细选择根和备份根。选择将对网络中的所有交换机通用的配置名称和修订版号。Cisco 建议您将尽可能多的交换机置于单个区域中；将网络划分为单独的区域是不利的。避免将任何 VLAN 映射到实例 0 上。应首先迁移核心。将 STP 类型更改为 MST，并逐步开始处理接入交换机。MST 可以在每个端口上与运行 PVST+ 的传统网桥交互，因此如果清楚了解交互作用，混合两种类型的网桥就不是什么问题。始终尝试将 CST 和 IST 的根保持在区域内部。如果通过中继与 PVST+ 网桥进行交互，请确保 MST 网桥是该中继上允许的所有 VLAN 的根网桥。

有关示例配置，请参阅：

- [将生成树从 PVST+ 迁移到 MST 的配置示例](#)
- [从 PVST+ 到快速 PVST 的生成树迁移配置示例](#)

## 结论

交换网络必须满足严格的稳健性、弹性和高可用性要求。随着 IP 语音 (VoIP) 和 IP 视频等技术的不断发展，链路或组件故障的快速收敛不再是理想特征。快速收敛是当务之急。然而直到最近，冗余交换网络还必须依赖相对缓慢的 802.1d STP 才能实现这些目标。这通常成为网络管理员的最富有挑战性的任务。使协议脱离时间的唯一方法是调整协议计时器，但这样往往会影响网络健康。Cisco 发布了许多 802.1d STP 新增功能（例如 UplinkFast、BackboneFast 和 PortFast），这些功能为加快生成树收敛铺平了道理。Cisco 还回答了与开发 MSTP 有关的、基于大型第二层 (L2) 的网络的可扩展性问题。IEEE 最近决定将这些概念的大部分整合到以下两个标准中：802.1w (RSTP) 和 802.1s (MST)。借助这些新协议的实施，在扩展到数以千计的 VLAN 的同时，有望实现几百毫秒的收敛时间。Cisco 一直保持着业界领先地位，并提供了这两个协议以及专有新增功能，以便促进传统网桥的迁移以及与传统网桥的互操作性。

## 相关信息

- [了解快速生成树协议 \(802.1w\)](#)
- [LAN 交换技术支持](#)
- [技术支持 - Cisco Systems](#)