

生成树协议与交换网络环路研究

郭彦伟, 郑建德*

(厦门大学计算机科学系, 福建 厦门 361005)

摘要: 交换网络环路是为交换网络提供冗余链路时形成的, 环路的存在导致了交换网络中广播风暴的产生, 致使交换网络的 MAC 地址系统失效, 这是冗余链路所不期望产生的. 现代化的网络要求交换网络中必须具备冗余链路, 以避免因单点故障引起的网络中断所带来的损失. 本文讨论生成树协议在具有冗余链路的交换网络中的工作实现, 及随着网络应用要求而改进的生成树协议技术, 即快速生成树协议和多生成树协议的工作实现, 从而达到在交换网络中既保障了网络冗余链路, 又避免了环路危害的目的.

关键词: 冗余链路; STP; RSTP; MSTP

中图分类号: TP 393.1

文献标识码: A

文章编号: 0438-0479(2006)S-0301-04

目前, 各公司和组织越来越依赖于计算机网络进行工作, 如果网络出现了故障, 就会带来巨大的商业损失. 在局域网中, 为了提供可靠的网络连接, 需要网络具有冗余链路. 所谓冗余链路, 就是准备两条以上的通路, 如果哪一条不通了, 就从另外的路走. 网络冗余链路的目的是消除由于单点故障而引起的网络中断, 网络具有冗余路径和设备本来是一件很好的事情, 但它引起的问题比它要解决的问题还要多.

1 冗余链路的危害

交换机作为二层的交换设备都具有一个相当重要的功能, 那就是能够记住在每一个端口上所收到的每个数据帧的源设备的硬件地址, 也就是源设备的 MAC 地址, 而且还会将该数据帧的 MAC 地址和端口号等信息写到 MAC 地址表中. 当交换机某个端口收到数据帧的时候, 交换机就查看其目的硬件地址, 并在 MAC 地址表中查找其外出的端口, 进行数据的转发, 这就是交换机进行数据交换的原理.

交换机在启动之初, MAC 地址表数据库并不存在, 它是交换机在交换数据过程中不断学习生成的, 而且是动态变化的. 在一个存在冗余链路的交换网络中 (图 1), 从工作站发出的一个数据帧就会沿两条链路通过图示的端口 E1 和端口 E3 分别进入交换机 A 和交换机 B. 由于此时交换机 A 和 B 的 MAC 地址表尚未形成, 该数据帧就会以广播的形式被原封不动地发送到交换机的所有其它端口, 数据帧就会沿交换机 A 的

E2 端口被发送到交换机 B 的 E4 端口, 同样, 也会沿交换机 B 的 E4 端口被发送到交换机 A 的 E2 端口, 然后被两交换机再次转发, 这样就造成同一数据帧在两点之间的环路内不停地被交换机转发的状况, 形成了所谓的广播风暴. 另一方面, 当工作站发送数据帧到网络时, 交换机要将收到的数据帧的源 MAC 地址写进 MAC 地址表, A 和 B 交换机都会将该数据帧的源 MAC 地址和对应的端口号写进各自的 MAC 地址表, 并转发该数据帧到对应的交换机, 此时每台交换机会再次将该帧的源 MAC 和相应端口号再次写入自己的 MAC 地址表, 周而复始. 交换机完全被设备的源地址搞胡涂了, 它不断地用源 MAC 地址更新 MAC 地址表, 根本没有时间转发数据帧了, 这种现象称为 MAC 地址系统失效. 由于广播风暴和 MAC 地址系统失效的原因, 使得冗余链路的存在造成整个网络处于阻塞状态, 导致网络瘫痪.

2 生成树协议的工作方式

为了解决冗余链路引起的问题, 生成树协议 STP

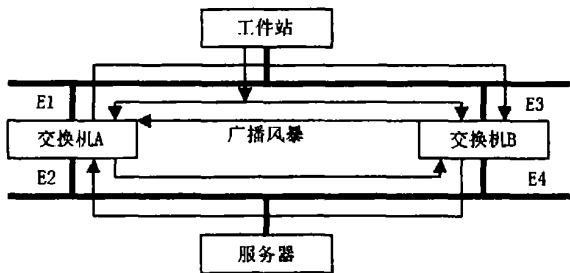


图 1 冗余链路交换网络

Fig 1 The network of redundancy link

收稿日期: 2006-02-20

基金项目: 厦门大学 985 二期信息创新平台项目资助

作者简介: 郭彦伟 (1971-), 女, 硕士研究生.

* 通讯作者: zhengjd@xmu.edu.cn

(Spanning Tree Protocol)应运而生. 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. 所以, 生成树协议的根本目的是将一个存在物理环路的交换网络变成一个没有环路的逻辑树型网络, 达到逻辑上裁剪冗余环路, 同时物理上实现链路备份和路径最优化. 生成树协议通过在交换机上运行 STA (Spanning Tree Algorithm) 算法来实现对链路的管理, 当发现网络中有环路时, 主动地在逻辑上阻断一个或多个冗余端口, 使得接入网络的计算机在与其它计算机通讯时, 只有一条链路生效, STP 持续地探测网络, 而当这个链路出现故障无法使用或者网络拓扑发生改变时, 生成树协议会重新计算网络链路, 重新配制端口, 这样既保障了网络的正常运行, 又保障了冗余能力.

对于一个存在物理环路的交换网络而言, 若想消除环路, 形成一个树型的逻辑网络, 实现的规则是这样的: 选举冗余链路中的一台交换机作为根(图 2), 在一个网络中只有一台交换机可以作为根交换机, 选举的依据是交换机的优先级和 MAC 地址组合成的 BID (Bridge ID), 它是 BPDU 的一部分, Bridge ID 是 8 个字节长, 包括 2 个字节的优先级和 6 个字节的设备 MAC 地址. 在一个冗余链路交换网络中桥 ID 最小的交换机将成为网络中的根交换机, 在交换机优先级都一样 (默认优先级是 32 768) 的情况下, MAC 地址最小的交换机成为根交换机. 选定了根交换机, 接下来要做的是交换机端口的设定. STP 在每个非根交换机上建立一个根端口, 选择根端口的依据是从非根交换机到根交换机的路径成本, 生成树的路径成本是路径上的所有链路的带宽的总路径成本的累积值. 一个非根交换机中到根交换机路径成本最低的端口将成为根端口, 根端口一般处于转发状态. 在建根端口的同时, 基于到根网桥的根路径成本的累计值的指定端口的选择也在进行, 指定端口是通过逐个考查与端口相连的网段来确定的, 选择指定端口的依据首先是路径成本, 路径成本低的端口将成为指定端口, 根交换机的所有端口都是指定端口 (除非根交换机自身存在环路), 因为根交换机的路径成本值为 0 在路径成本值相同的情况下, 再比较交换机的 BID 值, BID 值小的将成为指定端口, 否则将成为非指定端口. 指定端口通常处于转发状态, 当端口处于转发状态时, 它可以发送和接收流量. 非指定端口一般处于阻塞状态, 一个端口处于阻塞状态时, 不转发流量, 但仍能接收到流量. 交换机的端口信息记录在 BPDU (Bridge Protocol Data Unit) 里, 交换机之间是通过交换 BPDU, 获取各个交换机的参数信息, 通过这些信息来选择代价最小的成为通信的路

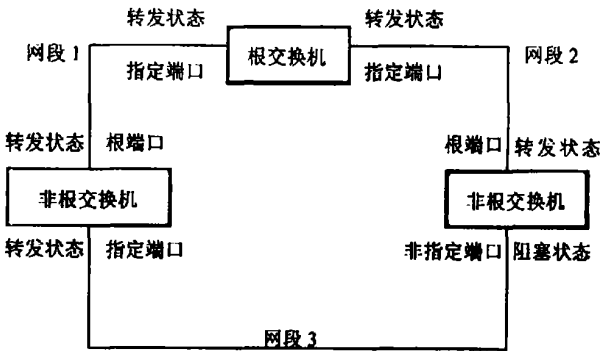


图 2 网络端口状态
Fig 2 The state of the port

径, 而另外的路径成为备份的路径. BPDU 每 2 s 发送一次, 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 协议的交换机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.

在一个启用了 STP 的交换网络中, 所有交换机在启动之后, 端口都要先经历阻塞状态, 再进入侦听和学习这两种过渡状态, 最终会稳定在转发或阻塞状态. 如果网络拓扑结构发生了变化, 处于转发状态的端口会暂时进入侦听状态, 处于阻塞状态的端口在指定时间没有收到新的 BPDU 时, 端口也会从阻塞状态转到侦听状态. 在侦听状态下, 交换机要完成选举根网桥, 选举非根网桥上的根端口以及各网段上的指定端口的工作, 这需要 15 s 的转发延迟. 另外, 为了防止旧的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应停止转发, 而继续转发造成的临时环路问题, 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学习状态. 在学习状态下, 端口只学习 MAC 地址但不参与数据帧的转发. 这虽然能够保证在网络拓扑发生变化时不会产生临时环路, 但是这个看似良好的解决方案实际上又增加了网络的收敛时间.

3 生成树协议的弊端和增强技术

生成树协议算法是广泛运用于二层以太网的收敛和自愈, 但由于它是在局域网的初期开发的技术, 因而对于现在的网络而言, 存在着一些弊端. 主要表现在以下几个方面:

(1) 当网络拓扑发生变化时, 端口由阻塞状态到转发状态, 要经过侦听状态和学习状态, 这两个状态需要约 30~50 s 的时间延迟, 这个时延对于早期的网络是可以接受的, 毕竟以前人们对网络依赖性不强, 但现在情况不同了, 50 s 的网络障碍足以带来巨大的损失,

尤其在一些关键的网络,如主机核心机房的连接,用户期望的值往往要短的多,这是生树协议的一个主要不足点.

(2) 由于在生成树的理论没有域的概念,网络拓扑的改变容易引起全局波动.网络中用户如果增加或减少设备,或者设备配置的改变往往会引起全局不必要的波动,用户改变其设备参数甚至会引起根桥的改变,致使通信网络中断.这造成用户在大规模的数据网络中不敢轻易使用生成树的算法.

(3) 缺乏对现有多 VLAN 环境的支持, 802.1d 没有阐明在一个存在多个 VLAN 情况下如何处理生成树协议算法,造成一个以太网交换机只支持单个 STP 运算,从而使双光纤链路的资源只能利用到一半.

针对生成树所带来的这些弊端,网络设备制造商开发了许多增强的技术,以使以太网用于现代高速的宽带网络.这些技术主要有:修改生成树的算法,提高其收敛时间;提出生成树划分域的概念,将核心网络与用户接入网络的 STP 域分开,保护核心网络的稳定性;针对以太网交换机开发的多 STP 协议,就是在每个 VLAN 上运行独立的 STP 算法,互不干扰.对应于这些技术,IEEE 在 802.1w 中定义了快速生成树协议 (RSTP),在 802.1s 中定义了多生成树协议 (MSTP).

3.1 快速生成树协议 (RSTP)

快速生成树协议是在生成树协议的基础上发展而来的,快速生成树协议不仅能够完成生成树协议的所有功能,优异之处在于当物理拓扑或者它的配置参数发生变化时,可以显著缩短网络的重新收敛时间,降低端口从阻塞到转发的延时.

与生成树协议相比,快速生成树协议额外为端口定义了替换和备份 2 种类型,将端口的状态定义为丢弃、学习和转发 3 种.与生成树协议不同的是,快速生成树协议在端口状态迁移方式、配置消息的格式、拓扑改变时消息的传播方式方面做了一些改进. 1) 为根端口和指定端口设置了快速切换用的替换端口 (为当前根端口所提供的根桥提供了替代路径)和备份端口 (指定端口向生成树提供的路径备份)两种角色,当根端口/指定端口失效的情况下,替换端口/备份端口通过它们之间的显示握手通信无时延地转入转发状态,并成为新的根端口和指定端口. 2) 为了防止环路, RSTP 利用网桥之间的“桥-桥握手机制”来确保通过网络分配的端口任务能够保持一致.由于这种握手机制不依赖于定时器,因此可以迅速地把信息传送到网络各处,随着拓扑结构的改变而在很短的时间内恢复连接. 3) 边缘端口 (直接与终端相连而不是把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,当边缘端口的链接稳定下来后,它将不会再引起拓扑的变化.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 3 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,还要等待两倍延迟时间进入转发状态.可见, RSTP 协议相对于 STP 协议的确改进了很多,加快网络的收敛,使得收敛速度最快达到 1 s 以内.但是, RSTP 协议同 STP 协议一样,仍然是单生成树协议,在网络规模比较大的时候会导致较长的收敛时间,在网络结构不对称的时候,单生成树还会影响网络的连通性.

3.2 多生成树协议 (MSTP)

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,现在只有草案版本可以获得,不过目前一些设备制造商已在产品中加进了对它的支持.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.在使用时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 MSTP 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, MSTP 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. 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 域内部运行多生成树协议,在 MST 域的边缘运行 RSTP 兼容的内部生成树协议.所以可以说, STP/RSTP 是基于端口的,而 MSTP 就是基于实例的.

图 3 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 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,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确定的 MSTP 生成树,个数由配置了多少个实例决定. MSTP 相对于之前的种种生成树协议而言,优势非常明显. 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

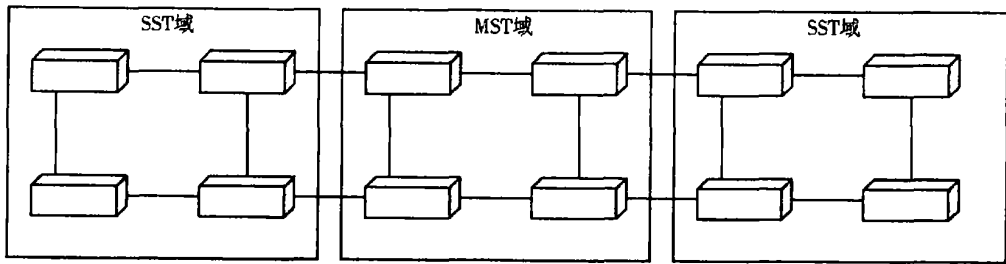


图 3 MSTP体系结构图
Fig 3 The architecture ofMSTP system

MSTP可以很好地向下兼容 STP /RSTP协议.

4 小结

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. 随着网络应用的不断深入,各种新的二层隧道技术不断涌现,如基于 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.

参考文献:

[1] 程庆梅, 刘天华. 计算机网络实训教程 [M]. 北京: 高等教育出版社, 2005
[2] Cisco Systems公司. 思科网络技术学院教程 [M]. 北京: 人民邮电出版社, 2004
[3] Tanenbaum. Andrew S. 计算机网络 [M]. 北京: 清华大学出版社, 2004

Research on Spanning Tree Protocol and Circle of Switch Network

GUO Yan-wei ZHENG Jian-de*

(Dept. of Computer Science, Xiamen University, Xiamen 361005, China)

Abstract The redundancy link of the switch network causes circle of network. And the circle causes not only the storm of the network broadcasting but also the failure of MAC address system. However, the advanced network needs the redundancy link, because it can avoid the break of network on single link. This paper is about how to avoid the damage of redundancy link through spanning tree protocol, and it discusses the principle of rapid spanning tree protocol and multiple spanning tree protocol. In the switch network of redundancy link, these protocols solve the damage of the circle and provide the redundancy link.

Key words redundancy link; STP; RSTP; MSTP