交换机基础

第1章 形形色色的交换机

1.1 概述

随着网络技术的发展，各种各样的通信设备应运而生，交换机就是其中一员。在一些技术类书籍或文章中，我们经常可以看到很多交换机的名词，它们中很多是由英语直接翻译过来，也有一些是厂商为了某种目的命名的。这些形形色色的交换机名很容易让人混淆，以下我们将介绍一下交换机的不同分类情况，并对其中一些常见的名词作一分析。

交换机包括电话交换机（PBX）、数据交换机（Switch），以下我们所提到的交换机都是指数据交换机，对传统的电话交换机就不作讨论了。

从广义上来看，交换机分为两种：广域网交换机和局域网交换机。广域网交换机主要应用于电信领域，提供通信用的基础平台。而局域网交换机则应用于局域网络，用于连接终端设备，如 PC 机及网络打印机等。以下内容都是基于局域网交换机来说的。

1.2 按规模应用分

按照最广泛的普通分类方法，即从规模应用上，局域网交换机可分为**企业级交换机、部门级交换机和工作组交换机**等。作为骨干交换机时，支持 500 个信息点以上大型企业应用的交换机为企业级交换机，支持 300 个信息点以下中型企业的交换机为部门级交换机，而支持 100 个信息点以内的交换机为工作组级交换机。但这也不是绝对的标准，正是由于没有统一的划分的尺度标准，又出现了**桌面型交换机（Desktop Switch）、校园网交换机（Campus Switch）**等概念。下面

对以上几个概念作一简介：

桌面型交换机，这是最常见的一种交换机，它区别于其他交换机的一个特点是支持的每端口 MAC 地址很少。广泛的使用于一般办公室、小型机房和业务受理较为集中的业务部门、多媒体制作中心、网站管理中心等部门。在传输速度上，现代桌面型交换机大都提供多个具有 10/100Mbps 自适应能力的端口。

工作组交换机，常用来作为扩充设备，在桌面型交换机不能满足需求时，大多直接考虑工作组型交换机。虽然工作组型交换机只有较少的端口数量，但却支持较多的 MAC 地址，并具有良好的扩充能力，端口的传输速度基本上为 100Mbps。

部门交换机，它通常不比工作组交换机更贵，而且与工作组交换机不同的是它们的端口数量和性能级别有所差异。一个部门交换机通常有 8～16 个端口，通常在所有端口上支持全双工操作。它们的性能要好于一个工作组交换机的性能，而且有一个等于或超过所有端口带宽的半双工汇集带宽。

校园网交换机，这种交换机应用相对较少，仅应用于大型网络，且一般作为网络的骨干交换机，并具有快速数据交换能力和全双工能力，可提供容错等智能特性，还支持扩充选项及第三层交换中的虚拟局域网(VLAN)等多种功能。

企业交换机，虽然非常类似于校园网交换机，但最大的不同是企业交换机还可以接入一个大底盘。

这些底盘产品通常支持许多不同类型的组件，比如快速以太网和以大网中继器、FDDI 集中器、令牌环 MAU 和路由器。企业交换机在建设企业级别的网络时非常有用，尤其是对需要支持一些网络技术和以前的系统。基于底盘设备通常有非常强大的管理特征，因此非常适合于企业网络的环境。不过，基于底盘设备的缺点是它们的成本都非常高。

1.3 按架构特点分

根据架构特点，人们还将局域网交换机分为**机架式、带扩展槽固定配置式、不带扩展槽固定配置式** 3 种产品。

机架式交换机

这是一种插槽式的交换机，这种交换机扩展性较好，可支持不同的网络类型，如以太网、快速以太网、千兆以太网、ATM、令牌环及 FDDI 等，但价格较贵，高端交换机有不少采用机架式结构。

带扩展槽固定配置式交换机

它是一种有固定端口数并带少量扩展槽的交换机，这种交换机在支持固定端口类型网络的基础上，还可以通过扩展其他网络类型模块来支持其他类型网络。这类交换机的价格居中。

不带扩展槽固定配置式交换机

这类交换机仅支持一种类型的网络（一般是以太网），可应用于小型企业或办公室环境下的局域网，价格最便宜，应用也最广泛。

1.4 按传输介质和传输速度分

从传输介质和传输速度上看，局域网交换机可以分为以太网交换机、快速以太网交换机、千兆以太网交换机、FDDI 交换机、ATM 交换机和令牌环交换机等多种，这些交换机分别适用于以太网、快速以太网、FDDI、ATM 和令牌环网等环境。

1.5 按层次结构分

从 ISO/OSI 的分层结构上说，交换机可分为二层交换机、三层交换机等。

二层交换机指的就是传统的工作在 OSI 参考模型的第二层 — 数据链路层上交换机，主要功能包括物理编址、错误校验、帧序列以及流控。一个纯第二层的解决方案，是最便宜的方案，但它在划分子网和广播限制等方面提供的控制最少。传统的路由器与外部的交换机一起使用也能解决这个问题，但现在路由器的处理速度已跟不上带宽要求。因此三层交换机、Web 交换机等应运而生。

三层交换机是一个具有三层交换功能的设备，即带有第三层路由功能的第二层交换机，但它是二者的有机结合，并不是简单地把路由器设备的硬件及软件叠加在局域网交换机上。

Web 交换机为数据中心设备（包括 Internet 服务器、防火墙、高速缓冲服务器和网关等）提供管理、路由和负载均衡传输。不同于传统网络设备的是，传统网络设备注重高速完成单个帧和数据包的交换，而 Web 交换侧重于跟踪和处理 Web 会话。除了由传统第二/三层交换机所提供的连接和封包路由外，Web 交换机还可提供传统局域网交换机和路由器所缺乏的完备策略，将局部和全球服务器负载均衡、存取控制、服务质量保证（QoS）以及带宽管理等管理能力结合起来。目前，Web 交换机已由纯粹的传输层（第四层）设备发展到具有基于内容（第七层）的交换的智能。利用内容或用户分类进行 Web 请求重定向是 Web 服务器的一项功能。不过，Internet 传输和商业的发展远远超过计算机处理能力的提高。把内容分类卸到 Web 交换机可平衡整个网站的基础设施。

随着技术的发展，肯定还会有更多的新名词涌现出来，但是只要掌握好原理，有清楚的概念，就不会被它们搞昏头脑。

1.6 新一代交换机前景展望

数字化、宽带化、传输光纤化、分组化是今后电信网络和信息网络的发展趋势，因此，很有必要构建一个与业务类型无关或弱相关的综合业务平台，使网络可以承载多业务。这样既能在设计、建设、运营和维护上取得较大的经济效益，又能够对业务的发展保持良好的适应性，从而最大限度地保护运营者与用户的投资利益。目前，随着硬件技术的发展，这种想法在技术上已经完全可以实现。

1.6.1 多层交换

目前，一二层交换技术已经不能够满足用户的需要，因此出现了第三层交换技术。

第三层交换将二层交换机和三层路由器两者的优势有机而智能化地结合成一个灵活的解决方案，可在各个层次提供线速性能。随着三层交换机在市场的不断推广和应用，三层交换技术及其产品在企业网/校园网建设、宽带 IP 网络建设（如城域网、智能社区接入）中得到了大量的应用，市场的需求和技术的发展双重拉动这种应用的纵深发展。三层交换的应用在从最初骨干层、中间的汇聚层一直渗透到边缘的接入层。

下一代网络将更加智能化，假如引入第四到第七层交换，那么网络就可识别网络上每一个数据包所属的应用和服务，然后应用这种信息把数据包传送到正确的路径。因此，在第三层交换技术走向成熟的基础上，第四到第七层交换技术也开始逐步被接受，并在一定的范围内获得了应用。

从技术角度来说，目前三层交换机虽然具备了企业级路由器的大多数功能，但路由器较三层交换机功能更为强大，如网络地址转换等，仍然无法由三层交换完全替代。但是随着技术的发展，三层交换机的功能肯定会越来越丰富。另外，从国内的情况来看，三层交换机虽然发展势头良好，但想取代企业级路由器还要经历一个漫长的过程。

1.6.2 光交换

业务的高速持续增长，需要更大的网络带宽来满足要求。DWDM 技术能有效解决带宽问题，但随之而来的是光纤中信道数量急剧增加，需要大容量的光交换机。密集波分复用（DWDM）能充分利用光纤的巨大带宽，使得“光纤耗尽”问题得到有效解决。近年来，DWDM 技术已经从长途干线系统渗透到城域网。由于复用的波长信道数急剧增加，光纤的传输容量可以以指数形式增长。

目前光交换机的市场规模并不大，这是因为光交换机技术上的相对不成熟，是限制光交换机市场迅速增大的主要因素。其主要体现在光交换矩阵技术的研究和开发上的不成熟。但是，随着技术的发展，光交换机的市场规模将日益壮大。可以预测，在未来的十年中，具有路由选择和出错恢复等功能的全光交换机将在交换机中逐渐占据主要地位。在以低于 1Gb／s 速率传输的网络（如局域网）中，OEO 还是有着明显的优势，这种可靠性高且成本较低的光电光(OEO)光交换机在低速网络中仍会得以广泛应用。

可以预见，光路交换技术构成的光交换机，将逐步在电信领域得到大规模推广使用。随着各类光开关技术及其他单元技术的日趋成熟，小规模的光交换机设备如 OXC 和 OADM 已经问世。现以 MEMS 为突破口的大规模光电集成和由此实现的大规模全光交换技术已初见端倪。全光交换将在未来的光通信领域占据重要地位和巨大的市场，并孕育着又一次网络技术的革命。

ATM 与 IP 结合

尽管 ATM 作为一种全能技术的神话已经破灭，但迄今为止，ATM 仍然是最适于多业务、多比特率应用环境的通信协议，因而作为多业务平台，汇接各种业务是其未来的主要角色。

而且，IP 技术也会面临大量的问题。比如，IP 的 QoS 问题。目前解决方案的思想很多都是借鉴于电信网络，但是实现起来难度很大。

所以，ATM 与 IP 结合是其另一重要的发展方向。IP 与 ATM 的结合是面向连接的 ATM 与无连接 IP 的统一，也是选路与交换的优化组合。

目前，还没有哪一种像 ATM 那样具有多业务高速率支持能力。因而对于电信网而言，在一段时间内，ATM 作为多业务平台是比较理想的。即使在可预见的未来，在网络边缘地带，ATM 作为业务汇集点仍然是不可缺少的。为此不少 ATM 厂家仍然在努力改进 ATM 交换机的性能和容量，使其在下一代电信网中占据一席之地。

ATM 与 IP 的结合有两种方式，第一是两者重迭，这样虽然可以保证 QoS，但是会增加协议的复杂度;第二是两者综合，这种方案能够最大限度地同时利用 ATM 与 IP 的优点。

ATM 技术能够提供二层的高速交换，ATM 标准已逐渐被完善，其产品也已基本成熟，广泛地推向市场。但是 ATM 和 IP 的结合仍然在不断演变中，工业标准的一致化也需要相当的时间。因此，ATM 与 IP 技术相结合的交换机会成为以后市场的主流。

1.6.3 MPLS 交换机

MPLS 交换机则是顺应 ATM 与 IP 结合这个潮流而产生的。多协议标记交换（Multiprotocol Label Switching，MPLS）是一种介于第二层和第三层之间的标记交换技术，是专门为 IP 设计的，可以将第二层的高速交换能力和第三层的灵活特性结合起来，使 IP 网具备高速交换、流量控制、QoS 等性能。它的产生伴随着网络的发展。

随着 MPLS 应用的不断升温，无论是产品还是网络，对 MPLS 的支持已不再是额外的要求，而应该是必备的功能。此外，MPLS 从骨干网走向边缘网也是一种越来越明显的趋势，这一进程将给边缘网带来更多的带宽、更高的智能和更多的服务。在接入网中，利用 MPLS 的技术承载以太网，会使网络更易升级和富有弹性。普通的以太网在每个骨干网中只能处理 4000 个 VLAN，MPLS 能使每个路由器支持最多 100 万个标记。因此，核心路由器厂商支持 MPLS 自然是毫无疑问的，如华为 Quidway NetEngine 系列产品等等;边缘路由器厂商也开始关注 MPLS。

从整个网络发展方向来看，在未来的核心网上，所有新的运营商在第一时间内建立的骨干 Internet 网都是光结点。MPLS 不再单一存在，它将与底层的光设备相辅相成。以前的 IP 是第一层、第二层、第三层在一起，现在，利用 MPLS 的基础，IP 与底层的光设备结合起来，让光去识别 IP 路由，即光是基于 IP 来驱动，将来的网络核心是波长路由，外面是一种大路由，这是以后大网核心的必然。对运营商来讲，今天的网络与以后的网络的关系是，所有今天的电信的其他网，如 DDN 专线网、ATM 的中继网等，都是将来整个大网络的接入结点。这个网不会摒弃以前所有的技术和产品，而是把它们结合进来，只是所有的应用都要以 IP 的形式来做，所有的东西都会以 IP 的形式终结在 Internet 和路由阶层中去。

目前，中国的骨干网带宽的利用率在 10%以下，因而，如何吸引更多的用户使用网络资源，是运营商、服务商关心的话题。路由器制造商都看到 MPLS 的最佳用武之地是，把承载多种不同类型服务的网络集成为一个单一的网络。网络运营商和服务商大多认为，用 MPLS 统一各种服务不失为一种长远的发展方向。

1.6.4 以太网交换机

另外，以太网交换机则是交换机市场的另外一个重要的角色。随着信息通信业的发展以及国民经济信息化的推进，以太网交换机市场呈稳步上升态势。以太网的特点是:性能价格比高、高度灵活、相对简单、易于实现。所以，以太网技术成为当今最重要的一种局域网建网技术。在全球数据网络中，以太网在数量上占有绝对的优势。所以，很明显，采用千兆以太网技术可以避免网络升级可能带来的兼容性等问题，并且可以节约由于网络升级而可能带来的高额费用。

因此，由于千兆/万兆以太网在技术上和实际应用中都有着非常大的优势，所以。以太网交换机的应用范围也必然越来越广泛。

同时，由于网络应用的普及，导致了用户的迅速增加，所以随着接入网技术的发展，使得以太网接入交换机也成为了有巨大发展潜力的市场，成为了另外一个新的市场增长点。

第2章 交换机工作原理

2.1 概述

1993 年，局域网交换设备出现，1994 年，国内掀起了交换网络技术的热潮。其实，交换技术是一个具有简化、低价、高性能和高端口密集特点的交换产品，体现了桥接技术的复杂交换技术在 OSI 参考模型的第二层操作。与桥接器一样，交换机按每一个包中的 MAC 地址相对简单地决策信息转发。而这种转发决策一般不考虑包中隐藏的更深的其他信息。与桥接器不同的是交换机转发延迟很小，操作接近单个局域网性能，远远超过了普通桥接互联网络之间的转发性能。

交换技术允许共享型和专用型的局域网段进行带宽调整，以减轻局域网之间信息流通出现的瓶颈问题。现在已有以太网、快速以太网、FDDI 和 ATM 技术的交换产品。

类似传统的桥接器，交换机提供了许多网络互联功能。交换机能经济地将网络分成小的冲突网域，为每个工作站提供更高的带宽。协议的透明性使得交换机在软件配置简单的情况下直接安装在多协议网络中；交换机使用现有的电缆、中继器、集线器和工作站的网卡，不必作高层的硬件升级；交换机对工作站是透明的，这样管理开销低廉，简化了网络节点的增加、移动和网络变化的操作。

利用专门设计的集成电路可使交换机以线路速率在所有的端口并行转发信息，提供了比传统桥接器高得多的操作性能。如理论上单个以太网端口对含有 64 个八进制数的数据包，可提供 14880bps 的传输速率。这意味着一台具有 12 个端口、支持 6 道并行数据流的“线路速率”以太网交换器必须提供 89280bps 的总体吞吐率（6 道信息流 X 14880bps／道信息流）。专用集成电路技术使得交换器在更多端口的情况下以上述性能运行，其端口造价低于传统型桥接器。

2.2 三种交换技术

2.2.1 端口交换

端口交换技术最早出现在插槽式的集线器中，这类集线器的背板通常划分有多条以太网段（每条网段为一个广播域），不用网桥或路由连接，网络之间是互不相通的。以大主模块插入后通常被分配到某个背板的网段上，端口交换用于将以太模块的端口在背板的多个网段之间进行分配、平衡。根据支持的程度，端口交换还可细分为：

- 模块交换：将整个模块进行网段迁移。
- 端口组交换：通常模块上的端口被划分为若干组，每组端口允许进行网段迁移。
- 端口级交换：支持每个端口在不同网段之间进行迁移。这种交换技术是基于 OSI 第一层上完成的，具有灵活性和负载平衡能力等优点。如果配置得当，那么还可以在一定程度进行客错，但没有改变共享传输介质的特点，自而未能称之为真正的交换。

2.2.2 帧交换

帧交换是目前应用最广的局域网交换技术，它通过对传统传输媒介进行微分段，提供并行传送的机制，以减小冲突域，获得高的带宽。一般来讲每个公司的产品的实现技术均会有差异，但对网络帧的处理方式一般有以下几种：

- 直通交换：提供线速处理能力，交换机只读出网络帧的前 14 个字节，便将网络帧传送到相应的端口上。
- 存储转发：通过对网络帧的读取进行验错和控制。

前一种方法的交换速度非常快，但缺乏对网络帧进行更高级的控制，缺乏智能性和安全性，同时也无法支持具有不同速率的端口的交换。因此，各厂商把后一种技术作为重点。

有的厂商甚至对网络帧进行分解，将帧分解成固定大小的信元，该信元处理极易用硬件实现，处理速度快，同时能够完成高级控制功能（如美国 MADGE 公司的 LET 集线器）如优先级控制。

2.2.3 信元交换

ATM 技术代表了网络和通讯技术发展的未来方向，也是解决目前网络通信中众多难题的一剂“良药”，ATM 采用固定长度 53 个字节的信元交换。由于长度固定，因而便于用硬件实现。ATM 采用专用的非差别连接，并行运行，可以通过一个交换机同时建立多个节点，但并不会影响每个节点之间的通信能力。ATM 还容许在源节点和目标、节点建立多个虚拟链接，以保障足够的带宽和容错能力。ATM 采用了统计时分电路进行复用，因而能大大提高通道的利用率。ATM 的带宽可以达到 25M、155M、622M 甚至数 Gb 的传输能力。

2.3 局域网交换机的种类和选择

2.3.1 按网络技术分

局域网交换机根据使用的网络技术可以分为：

- 以大网交换机；
- 令牌环交换机；
- FDDI 交换机；
- ATM 交换机；
- 快速以太网交换机等。

2.3.2 按应用领域分

如果按交换机应用领域来划分，可分为：

- 台式交换机；
- 工作组交换机；
- 主干交换机；
- 企业交换机；
- 分段交换机；
- 端口交换机；
- 网络交换机等。

2.3.3 交换机的选择

局域网交换机是组成网络系统的核心设备。对用户而言，局域网交换机最主要的指标是端口的配置、数据交换能力、包交换速度等因素。因此，在选择交换机时要注意以下事项：

(1) 交换端口的数量；
(2) 交换端口的类型；
(3) 系统的扩充能力；
(4) 主干线连接手段；
(5) 交换机总交换能力；
(6) 是否需要路由选择能力；
(7) 是否需要热切换能力；
(8) 是否需要容错能力；
(9) 能否与现有设备兼容，顺利衔接；
(10) 网络管理能力。

2.4 交换机应用中几个值得注意的问题

2.4.1 交换机网络中的瓶颈问题

交换机本身的处理速度可以达到很高，用户往往迷信厂商宣传的 Gbps 级的高速背板。其实这是一种误解，连接入网的工作站或服务器使用的网络是以大网，它遵循 CSMA／CD 介质访问规则。在当前的客户／服务器模式的网络中多台工作站会同时访问服务器，因此非常容易形成服务器瓶颈。有的厂商已经考虑到这一点，在交换机中设计了一个或多个高速端口（如 3COM 的 Linkswitch1000 可以配置一个或两个 100Mbps 端口），方便用户连接服务器或高速主干网。用户也可以通过设计多台服务器（进行业务划分）或追加多个网卡来消除瓶颈。交换机还可支持生成树算法，方便用户架构容错的冗余连接。

2.4.2 网络中的广播帧

目前广泛使用的网络操作系统有 Netware、Windows NT 等，而 Lan Server 的服务器是通过发送网络广播帧来向客户机提供服务的。这类局域网中广播包的存在会大大降低交换机的效率，这时可以利用交换机的虚拟网功能（并非每种交换机都支持虚拟网）将广播包限制在一定范围内。
每台文交换机的端口都支持一定数目的 MAC 地址，这样交换机能够“记忆”住该端口一组连接站点的情况，厂商提供的定位不同的交换机端口支持 MAC 数也不一样，用户使用时一定要注意交换机端口的连接端点数。如果超过厂商给定的 MAC 数，交换机接收到一个网络帧时，只有其目的站的 MAC 地址不存在于该交换机端口的 MAC 地址表中，那么该帧会以广播方式发向交换机的每个端口。

2.4.3 虚拟网的划分

虚拟网是交换机的重要功能，通常虚拟网的实现形式有三种：

2.4.3.1 静态端口分配

静态虚拟网的划分通常是网管人员使用网管软件或直接设置交换机的端口，使其直接从属某个虚拟网。这些端口一直保持这些从属性，除非网管人员重新设置。这种方法虽然比较麻烦，但比较安全，容易配置和维护。

2.4.3.2 动态虚拟网

支持动态虚拟网的端口，可以借助智能管理软件自动确定它们的从属。端口是通过借助网络包的 MAC 地址、逻辑地址或协议类型来确定虚拟网的从属。当一网络节点刚连接入网时，交换机端口还未分配，于是交换机通过读取网络节点的 MAC 地址动态地将该端口划入某个虚拟网。这样一旦网管人员配置好后，用户的

计算机可以灵活地改变交换机端口，而不会改变该用户的虚拟网的从属性，而且如果网络中出现未定义的 MAC 地址，则可以向网管人员报警。

2.4.3.3 多虚拟网端口配置

该配置支持一用户或一端口可以同时访问多个虚拟网。这样可以将一台网络服务器配置成多个业务部门（每种业务设置成一个虚拟网）都可同时访问，也可以同时访问多个虚拟网的资源，还可让多个虚拟网间的连接只需一个路由端口即可完成。但这样会带来安全上的隐患。虚拟网的业界规范正在制定当中，因而各个公司的产品还谈不上互操作性。Cisco 公司开发了 Inter－Switch Link（ISL）虚拟网络协议，该协议支持跨骨干网（ATM、FDDI、Fast Ethernet）的虚拟网。但该协议被指责为缺乏安全性上的考虑。传统的计算机网络中使用了大量的共享式 Hub，通过灵活接入计算机端口也可以获得好的效果。

2.4.3.4 高速局域网技术的应用

快速以太网技术虽然在某些方面与传统以大网保持了很好的兼容性，但 100BASE-TX、100BASAE-T4 及 100BASE-FX 对传输距离和级连都有了比较大的限制。通过 100Mbps 的交换机可以打破这些局限。同时也只有交换机端口才可以支持双工高速传输。

目前也出现了 CDDI／FDDI 的交换技术，另外该 CDDI／FDDI 的端口价格也呈下降趋势，同时在传输距离和安全性方面也有比较大的优势，因此它是大型网络骨干的一种比较好的选择。

3COM 的主要交换产品有 Linkswitch 系列和 LANplex 系列；BAY 的主要交换产品有 LattisSwitch2800，BAY stack workgroup、System3O00／5000（提供某些可选交换模块）；Cisco 的主要交换产品有 Catalyst 1000／2000／3000／5000 系列。

三家公司的产品形态看来都有相似之处，产品的价格也比较接近，除了设计中要考虑网络环境的具体需要（强调端口的搭配合理）外，还需从整体上考虑，例如网管、网络应用等。随着 ATM 技术的发展和成熟以及市场竞争的加剧，帧交换机的价格将会进一步下跌，它将成为工作组网的重要解决方案。

第3章 交换式以太网技术

3.1 交换式技术发展过程

3.1.1 引言

以太网交换机，英文为 SWITCH，也有人翻译为开关，交换器或称交换式集线器。我们首先回顾一下局域网的发展过程。

计算机技术与通信技术的结合促进了计算机局域网络的飞速发展，从六十年代末 ALOHA 的出现到九十年代中期 1000MBPS 交换式以太网的登台亮相，短短的三十年间经过了从单工到双工，从共享到交换，从低速到高速， 从简单到复杂，从昂贵到普及的飞跃。

八十年代中后期，由于通信量的急剧增加，促使技术的发展，使局域网的性能越来越高，最早的 1MBPS 的速率已广泛地被今天的 100BASE－T 和 100CG－ANYLAN 替代，但是，传统的媒体访问方法都局限于使大量的站点共享对一个公共传输媒体的访问， 既 CSMA/CD。

九十年代初，随着计算机性能的提高及通信量的聚增，传统局域网已经愈来愈超出了自身的负荷，交换式以太网技术应运而生，大大提高了局域网的性能。与现在基于网桥和路由器的共享媒体的局域网拓扑结构相比，网络交换机能显著的增加带宽。交换技术的加入，就可以建立地理位置相对分散的网络，使局域网交换机的每个端口可平行、安全、同时的互相传输信息，而且使局域网可以高度扩充。

3.1.2 从网桥、多端口网桥到交换机

局域网交换技术的发展要追溯到两端口网桥。桥是一种存储转发设备，用来连接相似的局域网。从互联网络的结构看，桥是属于 DCE 级的端到端的连接；从协议层次看，桥是在逻辑链路层对数据帧进行存储转发；与中继器在第一层、路由器在第三层的功能相似。两端口网桥几乎是和以太网同时发展的。

以太网交换技术（SWITCH）是在多端口网桥的基础上与九十年代初发展起来的，实现 OSI 模型的下两层协议，与网桥有着千丝万缕的关系，甚至被业界人士称为"许多联系在一起的网桥"，因此现在的交换式技术并不是什么新的标准，而是现有技术的新应用而已，是一种改进了的局域网桥，与传统的网桥相比，它能提供更多的端口（4～88）、更好的性能、更强的管理功能以及更便宜的价格。现在某些局域网交换机也实现了 OSI 参考模型的第三层协议，实现简单的路由选择功能，目前很热的第三层交换就是指此。以太网交换机又与电话交换机相似，除了提供存储转发（STORE ANG FORWORD）方式外还提供了其它的桥接技术，如：直通方式（CUT THROUGH）。

3.2 交换式以太网技术的优点

交换式以太网不需要改变网络其它硬件，包括电缆和用户的网卡，仅需要用交换式交换机改变共享式 HUB，节省用户网络升级的费用。

可在高速与低速网络间转换，实现不同网络的协同。目前大多数交换式以太网都具有 100MBPS 的端口，通过与之相对应的 100MBPS 的网卡接入到服务器上，暂时解决了 10MBPS 的瓶颈，成为网络局域网升级时首选的方案。

它同时提供多个通道，比传统的共享式集线器提供更多的带宽，传统的共享式 10MBPS/100MPS 以太网采用广播式通信方式，每次只能在一对用户间进行通信，如果发生碰撞还得重试，而交换式以太网允许不同用户间进行传送，比如，一个 16 端口的以太网交换机允许 16 个站点在 8 条链路间通信。

特别是在时间响应方面的优点，使的局域网交换机倍受青睐。它以比路由器低的成本却提供了比路由器宽的带宽、高的速度，除非有上广域网（WAN）的要求，否则，交换机有替代路由器的趋势。

3.3 第二层和第三层交换及其与路由器方案的竞争

如前所述，局域网交换机是工作在 OSI 第二层的，可以理解为一个多端口网桥，因此传统上称为第二层交换；目前，交换技术已经延伸到 OSI 第三层的部分功能，既所谓第三层交换，第三层交换可以不将广播封包扩散，直接利用动态建立的 MAC 地址来通信，似乎可以看懂第三层的信息，如 IP 地址、ARP 等，具有多路广播和虚拟网间基于 IP、IPX 等协议的路由功能，这方面功能的顺利实现得力于专用集成电路（ASIC）的加入，把传统的由软件处理的指令改为 ASIC 芯片的嵌入式指令，从而加速了对包的转发和过滤，使得高速下的线性路由和服务质量都有了可靠的保证。目前，如果没有上广域网的需要，在建网方案中一般不再应用价格昂贵、带宽有限的路由器。

3.4 虚拟局域网技术

交换技术的发展，允许区域分散的组织在逻辑上成为一个新的工作组，而且同一工作组的成员能够改变其物理地址而不必重新配置节点，这就是用到所谓的虚拟局域网技术（VLAN）。用交换机建立虚拟网就是使原来的一个大广播区（交换机的所有端口）逻辑的分为若干个"子广播区"，在子广播区里的广播封包只会在该广播区内传送，其它的广播区是收不到的。VLAN 通过交换技术将通信量进行有效分离，从而更好地利用带宽，并可从逻辑的角度出发将实际的 LAN 基础设施分割成多个子网，它允许各个局域网运行不同的应用协议和拓扑结构，对这部分详细内容感兴趣的读者可以参考 IEEE802.10 规定。

第4章 以太网交换机技术

4.1 交换机原理

以太网交换机的原理很简单，它检测从以太端口来的数据包的源和目的地的 MAC（介质访问层）地址，然后与系统内部的动态查找表进行比较，若数据包的 MAC 层地址不在查找表中，则将该地址加入查找表中，并将数据包发送给相应的目的端口。

交换机使用一种虚拟连接技术来连接通信的双方。所谓虚拟连接，就是指通信时通信双方建立一个逻辑上的专用连接，这个连接直到数据传送至目的节点后结束。虚拟连接是通过交换机的端口-地址表来实现的：交换机在工作过程中不

断地建立和维护它本身的一个地址表，这个地址表标明了节点的 MAC 地址和交换机端口的对应关系。当交换机收到一个数据包，它便会去查看自身的地址表以验明数据包中的目的 MAC 地址究竟对应于哪个端口。一旦验证完毕，就将发送节点与该端口建立一个专用连接，发送方的数据仅发送到目的 MAC 地址所对应的交换机端口。

交换机在同一时刻可进行多个端口对之间的数据传输。当节点 A 向节点 D 发送数据时，节点 B 可同时向节点 C 发送数据，而且这两个传输都享有网络的全部带宽，都有着自己的虚拟连接。这样，假使使用的是 10Mbps 的以太网交换机，那么该交换机的总流通量就等于 2×10Mbps=20Mbps，而在 10Mbps 的共享式以太网中，一个 Hub 的总流通量是不会超出 10Mbps。

4.2 交换机的内部结构

交换机卓越的性能表现来源于其内部独特的结构，目前交换机采用的内部结构主要有以下几种：

4.2.1 共享内存结构

依赖中心交换引擎来提供全端口的高性能连接。由核心引擎检查每个输入包以决定路由。这种方法需要很大的内存带宽，很高的管理费用。尤其是随着交换机端口的增加，由于需要内存容量更大，速度也更快，中央内存的价格变得很高。交换机内核成为性能实现的瓶颈。

4.2.2 交叉总线结构

交叉总线式结构在端口间建立直接的点对点连接。这种结构对于单点(unicast)传输来讲性能很好，但并不适合点对多点传输。由于在实际的网络应用环境中广播和多播传输很常见，这种标准的交叉总线方式会带来一些问题。例如：当端口 A 向端口 D 传输数据时，端口 B 和 C 就只能等待。而当端口 A 向所有的端口广播消息时，可能会引起目标端口的排队等候。这样会消耗掉大量的带宽，从而影响了交换机的性能。而且很容易理解的是要连接 n 个端口，就需要 n×(n+1)条交叉总线，因而实现成本会随端口数的增加急剧上升。

4.2.3 混合交叉总线结构

鉴于标准交叉总线存在的缺陷，一种混合交叉总线实现方式被提了出来。此种方式的设计思路是将一体的交叉总线矩阵划分成小的交叉矩阵，中间通过一条高性能的总线连接。优点是减少了交叉总线数，降低了成本，还减少了总线争用。但连接交叉矩阵的总线成为新的性能瓶颈。

4.2.4 环形总线结构

这种结构在1个环内最多支持4个交换引擎并且允许不同速度的交换矩阵互连，环与环间通过交换引擎连接。由于采用环形结构，所以很容易聚集带宽。当端口数增加的时候,带宽就相应增加了。与前述几种结构不同的是此种结构有独立的一条控制总线,用于搜集总线状态、处理路由、流量控制和清理数据总线。另外在环形总线上可以加入管理模块，提供完整的 SNMP 管理特性。根据需要，还可以选用第三层交换的模块使交换机具有第三层交换的功能。这种结构的最大优点是扩展能力强,实现成本低,而且有效地避免了系统扩展时造成的总线瓶颈。

4.3 交换机的交换方式

如果我们把集线器看成是一条内置的以太网总线的话,交换机可以被视为多条总线和交换矩阵互连。具体地说，交换机把每一个端口都挂在一条带宽很高的背板总线（Core Bus）上（至少比端口带宽高出一个数量级），Core Bus 与一个 Switch Engining 相连，由端口丢进来的封装数据包经 Core Bus 进入 Switch Engining，通过以下三种方式进行交换:

4.3.1 直通式（Cut Through）

直通方式的以太网络交换机可以理解为在各端口间是纵横交叉的线路矩阵电话交换机。它在输入端口检测到一个数据包时，检查该包的包头，获取包的目的地址，启动内部的动态查找表转换成相应的输出端口，在输入与输出交叉处接通，把数据包直通到相应的端口，实现交换功能。由于不需要存储，延迟非常小、交换非常快，这是它的优点。它的缺点是，因为数据包内容并没有被以太网交换机保存下来，所以无法检查所传送的数据包是否有误，不能提供错误检测能力。由于没有缓存，不能将具有不同速率的输入/输出端口直接接通，而且容易丢包。

4.3.2 存储转发（Store & Forward）

存储转发方式是计算机网络领域应用最为广泛的方式。它把输入端口的数据包先存储起来,然后进行 CRC 检查,在对错误包处理后才取出数据包的目的地址,通过查找表转换成输出端口送出包。正因如此，存储转发方式在数据处理时延时大，这是它的不足，但是它可以对进入交换机的数据包进行错误检测，有效地改善网络性能。尤其重要的是它可以支持不同速度的输入输出端口间的转换，保持高速端口与低速端口间的协同工作。随着交换技术的不断发展和成熟，存储转发交换机和直通式交换机之间的速度差距越来越小。此外，许多厂商已经推出了可以根据网络的运行情况，自动选择不同交换技术的混合型交换机。

4.3.3 碎片隔离（Fragment Free）

这是介于前两者之间的一种解决方案。它检查数据包的长度是否够 64 个字节，如果小于 64 字节，说明是假包，则丢弃该包；如果大于 64 字节，则发送该包。这种方式也不提供数据校验。它的数据处理速度比存储转发方式快，但比直通式慢。

从硬件上看，交换机比集线器多出 Core Bus 和 Switch Engining 两大部分，从而多出三种交换方式，这就是相同接口、相同带宽的交换机比集线器贵很多的原因。

4.3.4 直通式和存储转发的比较

直通方式的以太网络交换机可以理解为在各端口间是纵横交叉的线路矩阵电话交换机。它在输入端口检测到一个数据包时，检查该包的包头，获取包的目的地址，启动内部的动态查找表转换成相应的输出端口，在输入与输出交叉处接通，把数据包直通到相应的端口，实现交换功能。由于不需要存储，延迟（LATENCY）非常小、交换非常快，这是它的优点；它的缺点是：因为数据包的内容并没有被以太网交换机保存下来，所以无法检查所传送的数据包是否有误，不能提供错误检测能力，由于没有缓存，不能将具有不同速率的输入/输出端口直接接通，而且，当以太网络交换机的端口增加时，交换矩阵变的越来越复杂，实现起来相当困难。

存储转发方式是计算机网络领域应用最为广泛的方式，它把输入端口的数据包先存储起来，然后进行 CRC 检查，在对错误包处理后才取出数据包的目的地址，通过查找表转换成输出端口送出包。正因如此，存储转发方式在数据处理时延时大，这是它的不足，单是它可以对进入交换机的数据包进行错误检测，尤其重要的是它可以支持不同速度的输入输出端口间的转换，保持高速端口与低速端口间的协同工作。

4.4 更上层楼的交换机

局域网交换机是工作在 OSI 第二层的，可以理解为一个多端口网桥，因此传统上称为第二层交换。但是，尽管具有更大的带宽，目前的第二层交换机创建的巨大网络会很快被 ARP（地址转换协议）、服务存取点请求、RIP（路由信息协议）升级版之类的东西所淹没。在第三层交换出现前，网络管理人员对上述状态唯一的解决办法就是利用 VLAN 或路由器将网络进行分割，这种办法并不理想。

多层交换机的出现是网络应用发展的必然结果。类似点击式 Web 浏览这种瞬息万变和无法预测信息流的新型共享型应用不断涌现，已经彻底改变了传统的 80/20 法则(即 80%的信息流保持在一个局部子网内)。非本地子网间大量数据的传输使传统的路由器严重超载。我们知道，路由器是工作在 OSI 模型第三层的网络设备，任务是完成基于第三层地址的路由计算，提供多协议支持、WAN 连接以及完善的安全特性，但需要更多的反应时间和复杂的管理。第三层工作在包含客户机/服务器逻辑地址的通信协议一层，即 IP/IPX 地址层，也称网络层。第三层交

换机将传统的第二层交换机和第三层路由技术在单一产品中进行了功能合并。第三层交换可以不将广播封包扩散，直接利用动态建立的 MAC 地址来通信，似乎可以看懂第三层的信息，如 IP 地址、ARP 等，具有多路广播和虚拟网间基于 IP、IPX 等协议的路由功能。这主要得益于最新的高集成度 ASIC (ApplicationSpecific Integrated Circuit)专用电路，使第三层交换机能够实时进行线速交换和转发数据。其优势在于：克服普通路由器的性能瓶颈，解决了 Intranet 和 Internet 应用的 IP 阻塞呈爆炸性增长的难题。由于 ASIC 技术的发展允许以很经济的价格直接在高速硬件上实现复杂的路由功能，而不是像传统的路由器通过低速的软件实现，所以也避免了扩展性能时采用高性能路由器所付出的高昂价格。随着网络技术的不断发展，多层交换机有望在大规模网络中取代现有路由器的位置。

伴随着第三层交换机的出现，所谓的第四层、第七层交换机也粉墨登场。第四层(传输层)交换机通过阅读、处理第四层报头信息而提供应用级的控制，即支持安全过滤，提供对应用流施加特定的 QoS 策略以及应用层记账功能，从而优化数据传输和实现多台服务器间的负载均衡。第七层交换则更进一步，提供了基于内容的智能交换，能够根据实际的应用类型做出决策，进一步提高了应用性能。

以太网交换机的连接速度也突飞猛进。1999 年 6 月，IEEE 讨论通过了基于双绞线的 1000Base-T 标准 802.3ab，配合支持光纤的 802.3z 千兆标准，千兆以太网标准已经能支持目前所有主流的传输介质。目前市场上已经有成熟的千兆产品推出，甚至支持 10G 速率的 IEEE 802.3ae 标准也已出台。从传输介质上看，交换机之间越来越多地使用光纤来连接。与双绞线相比，虽然光纤的成本较高，架设稍复杂，但其连接距离远，工作稳定，可扩展性强，已经成为高速网络连接无可替代的主要介质。同时，随着网络用户的急速增长，交换机的端口密度也在不断膨胀，48 口、96 口，甚至几百个端口，这种高端口密度主要通过扩展模块来实现。另外，交换机也更注重安全、服务保证和管理功能，VLAN、QoS（服务质量）、SNMP（简单网络管理协议）、RMON（远程监控）、安全控制等功能的实现，极大地提高了网络的性能、安全性和稳定性。

第5章 以太网交换机常用的功能及协议

5.1 概要

本小节主要介绍了适合应用于宽带接入网的以太网交换机所常用的功能及协议。

以太网由美国 Xerox 公司和 Stanford 大学联合开发，1975 年推出。由 Xerox 公司和 Stanford 大学合作于 1980 年 9 月第一次公布了以太网的物理层和数据链路层的详细技术规范，成为世界上第一个局域网工业标准。IEEE 802.3 国际标准是在以太网标准的基础上制定的。以太网是迄今为止应用最为广泛的局域网形式，它经历了以太－快速以太－千兆以太的持续发展过程，应用领域也从 LAN 逐步向 MAN 扩展。IEEE802. 3ae 工作组已经开始了 10 吉比特以太网(10 Gigabit)标准的制定工作。相信在不久的将来，10 吉比特以太网将把以太网技术推向一

个新的高峰。

5.2 以太网交换机常用功能及协议

区别于在传统计算机局域网中的应用，基于以太网技术的宽带接入网对以太网交换机提出了严格的要求，即具有高度的信息安全性、电信级的网络可靠性、强大的网管功能，并且能保证用户的接入带宽，支持 VLAN、组播、STP（生成树协议）、QoS 保证、SNMP、RMON 等功能和协议。

5.2.1 VLAN

与第 3 层（路由）相比，通过第 2 层（交换）LAN 的结构能够明显改善整个网的性能。传统的第 2 层 LAN 交换提倡平整网络结构，充分利用线速交换功能传输数据，不是通过传统的路由器降低网络速度。

同时，在结构平整的大型网络中，经常会受到大量广播和偶然发生的广播风暴的困扰，使网络性能降低。过去，只能将网络分为更小的网段，各网段之间通过路由器连接，因为路由器通常不传输广播数据。但 VLAN 提供了另一种解决方案。

VLAN 已经出现了几年，但并不是所有的 VLAN 都是标准的。VLAN 标准 802.1Q 是 1998 年出台的，802.1Q 定义了帧标记的标准。标准制定者希望 802.1Q 能够消除 VLAN 中专有性，标准简化了构建 VLAN 的方案。

VLAN 就是不考虑用户的物理位置而根据功能、应用等因素将用户逻辑上划分为一个个功能相对独立的工作组，每个用户主机都连接在一个支持 VLAN 的交换机端口上并属于一个 VLAN。同一个 VLAN 中的成员都共享广播，而不同 VLAN 之间广播信息是相互隔离的。这样，将整个网络分割成多个不同的广播域。

虚拟局域网（VLAN）的出现打破了传统网络的许多固有观念，使网络结构变得灵活、方便、随心所欲。

VLAN 优势

1. 缩小广播域，控制广播风暴，提高带宽的使用率
2. 不同广播域的成员不能相互访问，提供网络安全性
3. 逻辑划分 VLAN 组，打破了地域的限制，当组成员物理位置变迁时减小搬迁的复杂性

在帧中，标记头位于目的 MAC 地址和源 MAC 地址之后，它是实现数据流过滤的基础。标记头由标记协议标识符（Tag Protocol Identifier，TPID)和标记控制信息（Tag Control Information，TCI)两部分组成，其中 TPID 表示本帧是个标记帧，以太网格式的 TPID 长两个字节，其值为以太网 V2 协议类型 802.1Q Tag Type，协议规定该域的值为 0x81 00。

TCI 由以下部分组成（如图 1 所示）：

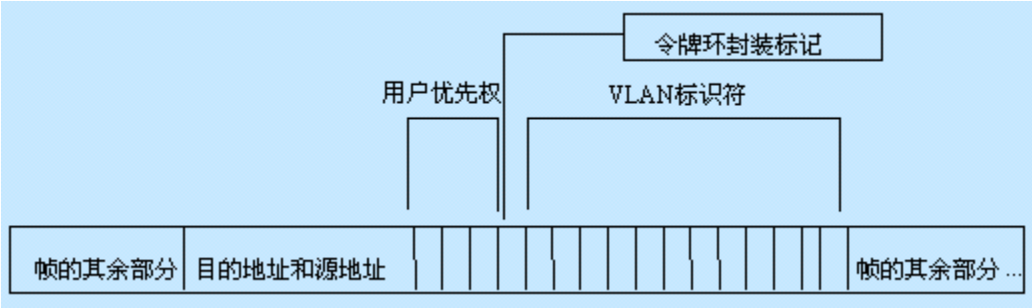

图 1 TCI 结构

用户优先权 ：user priority，用 3 位表示，取值范围从 0 至 7，表明帧的优先权。

一位令牌环封装标记 ：Token Ring encapsulation flag，用于指明该帧是否采用 IEEE 802.5 令牌的帧格式。

VLAN 标识符： VLAN identifier（VLAN ID），用 12 位表示，在帧与 VLAN 成员关系之间建立的关联。

网桥可以根据以上信息将帧仅转发到与特定 VLAN ID 相关的端口，能够依据优先权决定转发帧的顺序。更重要的是，交换机会保留该标记；即使桥接仍然是点到点的，该标记中的信息仍然能够帮助帧在非路由网络中"路由"。

通过 VLAN，你可以跨越多个 LAN，创建网络设备的逻辑组。这些逻辑组可能要跨越一个或多个第 2 层交换机，或者是建立在交换机到交换机基础之上的，逻辑组中可以传输广播数据。一个 VLAN 就定义一个广播域。

VLAN 通过交换技术将通信量进行有效分离，从而更好地利用带宽，并可从逻辑的角度出发将实际的 LAN 基础设施分割成多个子网，它允许各个局域网运行不同的应用协议和拓扑结构。

交换机可以支持 VLAN 的多种实现方式，如基于端口的 VLAN、基于 MAC 地址的 VLAN、基于网络层的 VLAN 等。基于端口的 VLAN 简单通用，可以任意指定多个端口为一个 VLAN，比较常用；基于 MAC 地址的 VLAN 适合于计算机经常移动位置的局域网，根据源 MAC 地址来设定和识别所属的 VLAN；基于网络层的 VLAN，可以根据主机的 IP 地址、子网掩码或协议类型来指定子网，这种定义方法的优点是当网络层的协议和 IP 地址改变时，交换机能自动识别和重新定义 VLAN。

5.2.2 组播技术

网络中的多媒体视频应可能在不同的网段/子网内，需要有多址广播路由协议才能使客户端工作站和服务器相连接，组播技术解决了网络带宽无谓浪费的问题。

基于 MAC 地址或第四类 IP 地址，将所需用户设置为固定的组，当这些组需要 VOD 视频点播等大数据量传输时，并不是单独为每个用户分别发送数据包，而是以单一数据包的形式，经层层复制，在带宽占用率相当低的情况下，最终用户却能以相当平滑的速度接收数据包，着就是组播的功能。ZONET 的 ZFS3024M 可分别基于 Layer2 的 MAC 地址或 Layer3 的 IP 地址进行接收广播组的设定，对多媒体业务的开展有良好的支持。

5.2.2.1 网络层组播协议实现

解决终端的动态登记用 Internet 组管理协议（IGMP，Internet Group Management Protocol）完成，该协议位于网络层，一个有组播功能的三层交换机定期向所有和它相连的子网系统发送 IGMP 查询报文来维护一个组播组的信息。终端通过发送 IGMP 应答报文（含组播组标识号）来确定参加某个组播组，也可以主动向交换机发送请求加入某个组播组的 IGMP 报文。交换机如果在一定时间内收不到从某个端口进来的某个组播组的应答报文，就从登记表中删除该项登记。

5.2.2.2 组播路由协议

IGMP 协议确定组播组和交换机之间的组成员关系，但是交换机之间还需要有组播路由算法使信息包能够遍历所有连接组播组成员的交换机，这需要组播路由协议来完成。

5.2.2.2.1 DVMRP 协议

DVMRP（Distance Vector Multicast Routing Protocol）名为距离向量组播选路协议。该协议分两步实现组播报文传送：第一步是反向路径转发，第二步是剪枝（Prune）。

5.2.2.2.2 PIM 协议

PIM（Protocol Independent Multicast）是 Internet 上唯一的商业使用的多址广播路由协议。它有两种模式：DM（Dense Mode 密集模式）和 SM（Sparse Mode 稀疏模式）。

Dense Mode 基于两种假设，一是组播成员分布比较密集，二是网络带宽足够。DVMRP、 MOSPF 和 PIM-DM 都是基于这种密集模式下的协议；Sparse Mode 假设组播成员分布比较稀疏，网络带宽也不充足的情况，PIM-SM 和 Core Based Trees (CBT)是基于这种稀疏模式下的组播路由协议。

PIM 变化时才广播，CPU 负荷小 ；DVMRP 周期性广播，CPU 负荷大。

5.2.2.3 第二层组播协议实现

5.2.2.3.1 IGMP Snooping

顾名思义，IGMP Snooping 就是当网络层发送 IGMP 报文时，第二层通过某种手段侦听 IGMP 请求或应答报文，自动记录下每个端口所属组播组和该组播组对应的 MAC 组播地址。当收到一个组播报文时，自动组播到那些表中有和该组播报文相同组播 MAC 地址的端口，实现组播功能，这种方法可以提高组播的效率。

IGMP snooping 采用软件的方法，中断交换机的 CPU，来告知交换机不必向

没有多址广播应用需求的工作站广播，从而降低交换机的负荷。

5.2.2.3.2 GMRP（GARP Multicast Registory Protocol）GARP 组播注册协议

所有 IGMP 响应报文和组播报文都具有相同的目的地址，必须通过分析 IP 头部才能区分，对所有组播报文进行分析以确定是否是 IGMP 响应报文，这不仅增加处理时间，而且增加了组播报文的传输延迟，所以 IGMP Snooping 对交换机造成处理上的负担；另外，IGMP 侦听只支持 TCP/IP 协议，不支持其它网络层协议，所以 GMRP 应运而生。

GMRP（通用组播注册协议）提供了向网络设备传播组播信息的机制,包括组成员信息和组服务要求信息，通过对组过滤数据库的创建、修改和删除，影响桥的组播过滤行为。GMRP 实际上是 GARP(Generic Attribute Registration Protocol 通用属性注册协议）的一个应用。GARP 是一种让终端和交换机通过交换域传播有用信息的二层传输机制。 GARP 为 GARP 应用提供了散发其属性登记或撤销的能力，使桥接局域网上的其它 GARP 应用能够获知该属性的改变。这些属性的类型、取值以及取值的语义则由具体的 GARP 应用来确定。

当一个终端想接收组播信息，用 GMRP 发一个请求加入报文到所连接的交换机，交换机收到该请求加入报文，在转发数据库中设置过滤表，并和相邻交换机交换组成员信息。当需要组播信息流时，交换机执行硬件查找，将报文转发到接收端口，由于交换机知道每一个 MAC 组播地址所对应的组播成员端口，这种转发不浪费带宽和处理时间。GMRP 用特定的 MAC 地址作为协议数据单元（PDUS）的 MAC 地址，不会和组播报文混淆，而且一个 PDUS 中可以包含多个组播组请求，减少了控制报文的通信流量。GMRP 是一个开放标准，它具有线速转发、节省带宽、良好的可扩充性为在交换域中提供了理想的途径。

5.2.3 容错技术

网络容错就是避免单点故障，在网络某处发生故障的情况下，系统能正常运行。

交换机应该在三个方面支持网络容错：链路冗余、生成树和主干（Trunking）通路。

5.2.3.1 链路冗余技术

链路冗余技术可以在交换机和网络之间定义两条链路：一条为主链路，另一条为备用链路。一般情况下，主链路工作，完成交换机和网络之间的信息交换，备用链路被阻塞不进行通信。一旦主链路发生故障,交换机能够自动切断主链路，而让备用链路进行交换机和网络之间的信息交换，主链路如果恢复，可以再重新切换到主链路上。

5.2.3.2 生成树（STP）技术

生成树（STP）技术允许交换机之间存在冗余链路，正常情况下，交换机之

间只允许一条链路工作，别的冗余链路被阻塞，这是通过生成树算法来实现的。一旦某个链路发生故障，交换机将自动启动生成树算法，将原来阻塞的冗余链路变为工作状态，保证交换机之间存在通信链路。一般是为每一个 VLAN 产生一个生成树，这样，每一个 VLAN 内都允许有冗余链路来保证系统的容错性。

生成树技术和冗余链路技术不能同时使用。冗余链路技术要求在需要容错的连接上定义冗余链路。这两条链路互为备份，只要有一条能正常通信，就能保证连接成功。而生成树技术可以在整个网络内设置冗余链路，只要网络连通性不被破坏，就能保证网络连接成功。

5.2.3.3 主干（Trunking）技术

第三种方法是主干（Trunking）技术，也叫链路聚合（Link Aggregation）技术。通过使用链路聚合，可以使一个或多个连接形成一个链路聚合组，对上层的 MAC Client 来说，链路的聚合在逻辑上等同于一条链路，只是链路通信容量增加了许多。

链路聚合（Trunking）功能是将交换机的多个低带宽交换端口捆绑成一条高带宽链路，通过几个端口进行链路负载平衡，避免链路出现拥塞现象，打比喻来说，链路聚合就如同超市设置多个收银台以防止收银台过少而出现消费者排队等候过长的现象。通过配置，可通过 2 个、3 个或 4 个端口进行捆绑，分别负责特定端口的数据转发，防止单条链路转发速率过低而出现丢包的现象。

主干技术用于完成两台交换机之间的连接。在没有故障的情况下，两台交换机之间的带宽可以随着主干中物理链路的增加而增加，信息流量均匀地分配给主干中的各条物理链路。当其中某条物理链路发生故障，自动失效该链路并停止传送信息，交换机不再把信息流分配给该失效链路所连端口。主干中一条或多条物理链路失效，不会影响两台交换机的连通性，只是链路带宽随着失效链路数的增加而下降，因此主干技术是提高网络带宽和容错的有效方法。

5.2.4 QoS 保证

QoS（Quality of Service）本来是 ATM 中的专用术语，在 IP 上原来是不谈 QoS 的，但利用 IP 传 VOD 等多媒体信息的应用需求越来越多，IP 作为一个打包的协议显得有点力不从心：延迟长且不为定值，丢包造成信号不连续且失真大。

交换式以太网不能保证数据传输延迟的确定性，在重负载情况下，由于阻塞导致报文延迟过大，甚至报文丢失。如果以太网不能对信息流分配不同的优先级，时间敏感的数据流和普通数据流享受同等级别服务，多媒体数据会常常因为等待普通数据转发而增加传输延迟。以太网 QoS 保证通过对数据流设置优先级来实现，实现优先级设置存在两种方法：一是基于 IEEE802.1p 标准的第二层链路帧优先级（3BITS）；二是基于 Ipv4 的 TOS，只适合 TCP/IP 协议（3BITS）。

数据帧可以根据协议类型、IP 源地址、目的地址及它们的组合、源或目的 TCP/UDP 端口地址及其它 IP 首部内容进行分类。

5.2.5 配置和管理及其安全性

网络管理者需要获得设备统计数据和对网络设备进行配置。设备统计数据完整反映出网络系统的运行状况，不能被随意访问。对网络设备的配置更是直接影响到网络结构和网络使用效益，除授权的网络管理员外，其它用户不能随便操作。

目前访问以太网交换机的途径主要有四种：一是通过设备的串行终端口；二是通过 Telnet；三是通过 SNMP 网管工作站；四是基于 WEB 的网络管理系统。

第一种方法中用户可以通过设备串行终端口外接一个终端，用终端命令来完成对网络设备的访问；

第二种方法中用户终端可用远程登录协议 Telnet，登录到网络设备上，通过用户终端命令完成对网络设备的访问。

对上述两种方法的安全问题，可通过对设备设置用户名和口令来解决，访问前必须输入正确的用户名和口令。

通过 SNMP 报文访问的方法如下：运行网管软件的网管工作站向设备发送信息报，信息报中包含网络设备中 SNMP 代理识别的公共字符串。如果代理能识别信息报中的公共字符串，则执行信息报指定的动作，否则，对信息报不予理睬。SNMPv2 采用了 MD5 加密算法，对 SNMP 信息报加密，防止公共字符串被泄密。

基于 Web 的网络管理系统通常采用 Web 嵌入的方式，将 Web 能力真正嵌入到以太网交换机中，每个设备都有它自己的 Web 地址，这样可以轻松地通过浏览器访问到该设备并管理之。在登录 Web 服务器时要提供口令，以保证安全性。

5.2.6 网络监测和管理

5.2.6.1 SNMP(Simple Network Management Protocol)

简单网络管理协议（SNMP）已成为管理分布的网络设备的主要手段，SNMP 用嵌入在网络设备中的代理软件收集网络流量信息和设备统计数据。每个代理不断地采集统计数据，并把它们记录在本地系统管理信息库（MIB）中，网络管理工作站通过轮询 MIB 得到这些信息。

MIB 只记录统计数据，它们不能提供网络流量的历史分析，为了得到一定时间内的网络流量视图和变化速度，网络管理工作站必须定时轮询 SNMP 代理。传统的 SNMP 的模式存在如下缺陷：

定时轮询浪费带宽，甚至有可能造成阻塞，随着网络规模增大问题更为突出。

SNMP 把所有负担都加在网络管理工作站上，CPU 承担大量网络段的流量计算、分析工作。

SNMP 代理只记录反映单个监测段的信息，从中很难发现整个网络系统的流量状态。

5.2.6.2 MON(Remote Monitoring)

参考标准：RFC1757 和 RFC1513

RMON 同 SNMP 一样，也是基于 C/S 结构，一个 RMON 探测器或代理作为服务

器，保存采集到的历史统计数据，避免了为得出网络流量历史视图，网络管理工作站的定时轮询。RMON 代理可以是一个单独的设备，也可以是嵌在网络设备中的智能部件。网络管理工作站作为客户，用代理得到的信息分析网络流量，查找存在问题。网络管理工作站和分布的代理之间通信用 SNMP 实现。

遍布在 LAN 网段之中的 RMON 代理，不会干扰网络，它能自动工作，无论何时出现意外的网络事件，它都能上报，RMON 代理的过滤功能使之能根据用户定义的参数来捕获特定类型的数据报文。当一个 RMON 代理发现一个网段处于一种不正常状态时，它会主动与网络管理工作站来联系，并将描述不正常状况的捕获信息转发。网络管理工作站对 RMON 数据从结构上进行分析，来诊断问题之所在。和传统的 SNMP 相比，RMON 有如下优点：大大减少了网络管理信息流量；提供了整个网络系统，包括网络设备、服务器、应用程序和用户总的网络流量关系；降低了网络管理成本。

RFC1757 为以太网定义了 RMON MIB，RMON MIB 分为 9 组，分别是：统计（statistics）、历史（history）、警报（alarm）、主机（host）、最高 N 台主机（hostTopN）、矩阵（matrix）、过滤（filter）、数据报捕获（packet capture）、事件（event）。

第6章 网络交换技术浅析

6.1 前言

当今世界已经步入信息时代，随着社会的迅速发展以及人们对网络应用需求的不断提高，对网络速度及带宽的要求不断上升。正是在这样的发展形势下，许多高速交换的新技术不断涌现。本文旨在对网络交换技术进行一些简单的分析，希望可以帮助一些有兴趣的朋友了解网络交换的基本原理。

6.2 集线器

集线器(HUB)是局域网 LAN 中重要的部件之一，它是网络连线的连接点。其基本的工作原理是使用广播技术，也就是 HUB 从任一个端口收到一个信息包后，它都将此信息包广播发送到其它的所有端口，而 HUB 并不记忆该信息包是由哪一个 MAC 地址挂在哪一个端口。接在 HUB 端口上的网卡 NIC 根据该信息包所要求执行的功能执行相应动作，这是由网络层之上控制的。上面所说的广播技术是指 HUB 将该信息包发以广播发送的形式发送到其它所有端口，并不是将该包改变为广播数据包。集线器的工作原理很类似于现实中投递员的工作，投递员只是根据信封上的地址传递信件，并不理会信的内容以及收信人是否回信，也不管是否收信人由于某种原因而没有回信，而导致发信人着急。唯一不同的就是投递员在找不到该地址时会将信退回，而 HUB 不管退信，仅仅负责转发而已。

6.3 路由器

路由器是在 OSI 七层网络模型中的第三层--网络层操作的。它的工作原理是，在网络中收到任何一个数据包(包括广播包在内)，都将该数据包第二层(数据链路层)的信息去掉(称为"拆包")，并查看第三层信息（IP 地址)。然后，再根据路由表来确定数据包的路由，然后检查安全访问表；如果能够通过，则进行第二层信息的封装(又称为"打包")，最后才将该数据包转发。此时，如果在路由表中不能查到对应 MAC 地址的网络地址，则路由器将向源地址的站点返回一个信息，然后将这个数据包丢弃。

如果从路由器的工作原理来看，路由器的作用与交换机、网桥非常相似。但是与工作在网络物理层，从物理上划分网段的交换机不同，路由器则是使用专门的软件协议从逻辑上对整个网络进行划分。例如，一台支持 IP 协议的路由器可以把网络划分成多个子网段，只有指向特殊 IP 地址的网络流量才被允许通过路由器。路由器对每一个接收到的数据包，都会重新计算其校验值，最后写入新的物理地址。因此，在网络中使用路由器来转发和过滤数据的速度往往要比只查看数据包物理地址的交换机慢一些。但是，对于那些网络结构较复杂的网络，采用路由器来连接网络却可以提高网络的整体效率。路由器的另外一个明显的优势就是可以自动过滤网络广播，但是从总体上说，在网络中添加路由器的安装过程要比即插即用的交换机复杂许多。

6.4 交换机

交换机能够检查每一个收到的数据包，并且对该数据包进行相应的动作处理。在交换机内保存着每一个网段上所有节点的物理地址，它只允许必要的网络流量通过交换机。例如，当交换机接收到一个数据包之后，它需要根据自身以保存的网络地址表来检验数据包内所包含的发送方地址和接收方地址。如果接收方地址位于发送方地址网段，那么该数据包将会被交换机丢弃，不会通过交换机传送到其它的网段；如果接收方地址与发送方地址是属于两个不同的网段内，那么该数据包就会被交换机转发到目标网段。这样，我们就可以通过交换机的过滤和转发功能，来避免网络广播风暴，减少误包和错包的出现。

在实际网络构件的过程中，是选择使用交换机还是选择其它的网络部件，主要还是要根据不同部件在网络中的不同作用来决定。在网络中交换机主要具有两方面的重要作用。第一，交换机可以将原有的网络划分成多个子网络，能够做到扩展网络有效传输距离，并支持更多的网络节点。第二，使用交换机来划分网络还可以有效隔离网络流量，减少网络中的冲突，缓解网络拥挤情况。但是，在使用交换机进行处理数据包的时候，不可避免的会带来处理延迟时间，所以如果在不必要的情况下盲目使用交换机就可能会在实际上降低整个网络的性能。

6.5 二层交换机介绍

该交换机根据每一个数据包中的目的 MAC 地址作简单的转发，转发决策并不需要要判断数据包深层的其他信息。与网桥不同的是，该交换机能以非常低的延迟转发数据包，能比桥接的网络提供更接近于单一局域网段的性能。它把网络分段成更小的冲突域，为每个终端站点提供更高的平均带宽。

二层交换机是协议透明的，当工作于多协议的网络环境时，不需要或只需要很少的软件配置。二层交换机可以使用现有的电缆系统、集线器、工作站网卡，不需要昂贵的硬件升级。

二层交换机要取的网络控制方法一般有两种：一种是“处处交换”的网络控制方法，采取这
种网络控制方法的二层交换机没有路由功能，而且所有的 LAN 信息流都经过交换机进行传输，所以，这种方法能最低的成本提供最高的性能。它对整个网络进行平整（即取消一切子网），使终端站点可以访问 LAN 的任何部分，不必经受路由器的延迟或控制，简化了某些方面的网络管理。但这种方法也有其局限性，由于其中没有任何减少广播流量的控制手段，所以广播信息流的泛滥会浪费网络的带宽。大型的平整化的网络还难以进行故障检修，也难以实现先进的安全功能。随着网络的扩大，由交换机连接的所有网段仍属于同一个广播域，广播数据包会在所有网段上传播，这在某些情况下会导致通信拥挤和出现安全漏洞，因而通常需要在某个层次上使用第三层控制。另一种是“尽可能交换、必要时路由”的网络控制方法。采取这种网络控制方法的二层交换机划分了逻辑子网，用虚拟局域网（VLAN）来产生逻辑子网，并将广播流量限制在子网内部，子网间的互通必须经过必要的路由处理，这样就要利用中币功能保证安全，减少广播信息流量和地址管理。

在 OSI 标准中将通讯分为 7 层(Layer)，包括 Physical，Logic，Network，Transport，Session，Present 及 Application 等，每层各司其职，最高层的应用程序(如 WWW / HTTP)逐层透过解析与封包而由第一层(如 Ethernet)之传输媒介载送信号传至接收端；值得注意的是 TCP/IP 的通信协议与各种应用程序通常省略了第五及第六二层，故只有五层。而 Layer 2 Switch 顾名思义，即是在局域网络通讯传输中仅以第二层(MAC 层)的信息来作为传输与资料交换之依据，通常此类交换器先以学习的方式(Learning) 在每一个 port 纪录该区段的 MAC Address 再根据 MAC 层封包中的目的地地址(Destination Address，DA)传送该封包至目的地的 port (或区段)，其它 port (或区段)将不会收到该封包，若目的地地址仍然在该(或区段)，则封包将不会被传送。Layer 2 的 Switch 由于只判断第二层的信息故其处理效能佳，且其有效隔绝区段间非往来封包(及独享频宽)，大大提升网络的传输效能，且因技术与 ASIC 芯片的功能日益强化，目前较高档的 Layer 2 Switch 每个 port 均可达到 Wiring Speed 的传输率(Ethernet 为 14880pps，Fast Ethernet 为 148800pps)。

6.6 第三层交换技术

6.6.1 概述

第三层交换技术（也被称做多层交换技术，或是 IP 交换技术）是相对于传统交换概念而提出的。众所周知，传统的交换技术是在 OSI 网络标准模型中的第二层--数据链路层进行操作的，而多层交换技术是在网络模型中的第三层实现了数据包的高速转发。简单地说，多层交换技术就是：第二层交换技术＋第三层转发技术，或者说是将传统路由器的数据包处理功能和交换机的速度优势结合在一起。

要了解第三层交换技术的原理并不困难，我们可以假设某主机 A 与 B 以前曾通过交换机进行通信，如果该交换机可以支持第三层交换的话，那么它便会将 A 和 B 的 IP 地址及它们的 MAC 地址记录下来，当其它主机 C 想要与 A 主机或 B 主机进行通信时，在交换机接收到 C 所发出的寻址封包后，会不假思索的送回给 C 一个回覆信息包，并告诉它主机 A 或主机 B 的 MAC 地址，那么以后主机 C 就会使用主机 A 或 B 的 MAC 地址"直接"通信。

因为通信双方并没有通过路由器进行"拆包"和"打包"的过程，所以那怕主机 A、B 或 C 分属于不同的子网，它们之间也可直接知道对方的 MAC 地址来进行通信，最重要的是，第三层交换机并没有像其它交换机一样把广播封包扩散，第三层交换机之所以叫三层交换机就是因为它可以看懂三层信息，比如 IP 地址、ARP 等。所以，三层交换机便能洞悉某一广播封包目的何在，在没有把它扩散出去的情形下，同时满足了发出该广播封包的人的需求，(不论它们在任何子网里)。因为第三层交换机没做任何"拆打"数据包的工作，所有经过它的数据包都不会被修改并以交换的速度传到目的地。所以，应用第三层交换技术即可实现网络路由的功能，又可以根据不同的网络状况做到最优的网络性能。我们可以相信，随着网络技术的不断发展，第三层交换机有望在大规模网络中取代现有路由器的位置。

6.6.2 三层交换的概念

什么是三层交换，简单地说，三层交换技术就是：二层交换技术＋三层转发技术。它解决了局域网中网段划分之后，网段中子网必须依赖路由器进行管理的局面，解决了传统路由器低速、复杂所造成的网络瓶颈问题。

三层交换（也称多层交换技术，或 IP 交换技术）是相对于传统交换概念而提出的。众所周知，传统的交换技术是在 OSI 网络标准模型中的第二层——数据链路层进行操作的，而三层交换技术是在网络模型中的第三层实现了数据包的高速转发。

三层交换技术的出现，解决了局域网中网段划分之后，网段中子网必须依赖路由器进行管理的局面，解决了传统路由器低速、复杂所造成的网络瓶颈问题。

Layer 3 Switch 又称为 IP Switch 或 Switch Router，意即其工作于第三层网络层的通信协议(如 IP)，并藉由解析第三层表头(Header)将封包传至目的地，有别于传统的路由器以软件的方式来执行路由运算与传送，Layer 3 Switch 是以硬件的方式(通常由专属 ASIC 构成)来加速路由运算与封包传送率并结合

Layer 2 的弹性设定，因此其效能通常可达每秒数百万封包(Million packet per second)的传送率，并具备数十个至上百个以上的高速以太网络(Fast Ethernet)连接端口，或数个至数十个超高速以太网络(Gigabit Ethernet)连接端口之容量。

传统路由器通常可处理 Multiprotocal 多重协议路由运算(如 IP，IPX AppleTalk，DEC Net...etc)但 Layer 3 Switch 通常只处理 IP 及 IPX，此乃为简化设计，降低路由运算与软件的复杂性以提升效能，并配合网络协议发展的单纯化(多重协议慢慢简化至 IP 一种协议)趋势所致。

由于 Layer 2 的 Switch 并无法有效的阻绝广播域(Broadcast Domain)如 ARP (Address Resolution Protocol)及 Win95/98 中大量使用的 NetBEUI 协议均大量使用广播封包，因此就算 Layer 2 Switch 以 VLAN (Virtual LAN)的方式(虚拟网络)将经常要通讯的群组构成一广播域(Broadcast Domain)来试图降低 broadcast 封包对网络层的影响，但仍无法完全避免广播风暴问题(同一个 VLAN 间仍会产生广播风暴)，再加上现今网络(尤其是 Campus 内部间流量及对外的 Internet/Intranet 流量)已不是 80/20 规则(80%流量在本地，20%是外地)，而是渐渐成为 20/80 规则，且加上 Client/Server 及 Distributor Server 之运用，因此单靠 Layer 2 Switch 或传统 Router 路由器便无法符合对效能(传统路由器变成瓶颈)及 Intranet 上对安全顾忌(Layer 2 Broadcast Domain，对因广播而使信息传送被盗取的安全疑虑)之要求，因此 Layer 3 Switch 便大量兴起，初期只运用 Core 端(骨干)，现在的趋势已渐渐走向桌面(Layer 3 down to desktop)。

如同传统路由器(Router)，Layer 3 Switch 的每一个连接埠(port)都是一个子网络(Subnet)，而一个子网络就单独是一个 Broadcast Domain 广播域，因此每一个 port 的广播封包并不会流窜到另一个 port，其仅负责传送要跨越子网络的封包(Routing Forward)，并以目的地的 IP 地址(目的地子网络的网络号码)来决定封包要转送至哪一个 port，并以 Routing Protocol(如 RIP 或 OSPF)来交换 Routing Table 并学习网络拓蹼，其通常存放于 Layer 3 Switch 的 Routing Forward Data-Base(FDB)，并以硬件及 Route Cache 的方式来加速 IP table lookup 并予以寻址与更新(目前大多以 ASICIC 来执行)，因此才得以提升运算效能达成 Wiring Speed Forward 之目的。

Layer 3 Switch 通常提供较大频宽的交换核心(Switch Fabric)以提供较大的容量(Port Capacity)与较高的交换效能，近来各厂家并不断附以 Layer 3 Switch 更强大的支持能力，如 Class of Service(服务等级优先权)，Quality of Service(服务品质保证)，Policy Management(策略分级品质与频宽管制与管理)，Multicast Routing(群组广播路由传送)等功能，以符合网络环境的快速变化与应用。

6.6.3 三层交换原理

三层交换不是一个新概念，但很多人只是把它当作路由器的替代品，其实，它将二层交换机和三层路由器两者的优势有机而智能化地结合起来，可在各个层次提供线速性能。

这种集成化的结构还引进了策略管理属性，不仅使二层与三层相互关联起来，而且还提供流量优先化处理、安全访问机制以及多种其它的灵活功能。

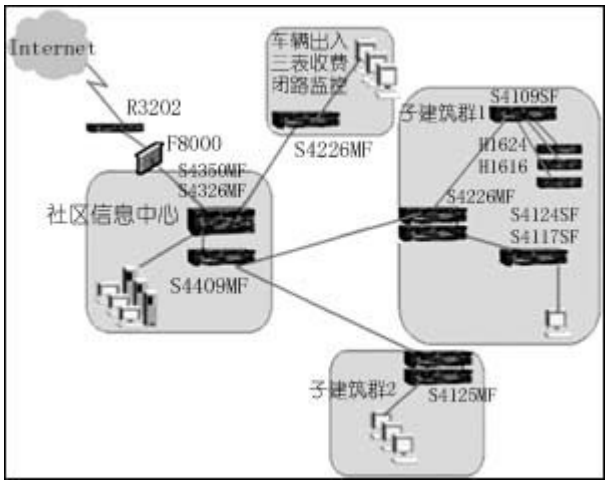

图 TCL 网络智能社区方案

TCL 网络能够组建具有社区特色的 IP 网络。这个网络既是一个相对独立的信息服务和信息管理网，同时又是一个与外部网络（如 Internet）连接的桥梁。作为一个独立的网络信息系统，具有内部高速交换的性能，并且利用其高速交换的能力，建立社区内部使用的、具有社区文化特色的多媒体信息系统；而作为一个与外部网络连接的桥梁，则具有透明的连通性及对社区内部网的安全防护功能。

其原理是：假设两个使用 IP 协议的站点 A、B 通过第三层交换机进行通信，发送站点 A 在开始发送时，把自己的 IP 地址与 B 站的 IP 地址比较，判断 B 站是否与自己在同一子网内。若目的站 B 与发送站 A 在同一子网内，则进行二层的转发。若两个站点不在同一子网内，如发送站 A 要与目的站 B 通信，发送站 A 要向“缺省网关”发出 ARP(地址解析)封包，而“缺省网关”的 IP 地址其实是三层交换机的三层交换模块。当发送站 A 对“缺省网关”的 IP 地址广播出一个 ARP 请求时，如果三层交换模块在以前的通信过程中已经知道 B 站的 MAC 地址，则向发送站 A 回复 B 的 MAC 地址。否则三层交换模块根据路由信息向 B 站广播一个 ARP 请求，B 站得到此 ARP 请求后向三层交换模块回复其 MAC 地址，三层交换模块保存此地址并回复给发送站 A,同时将 B 站的 MAC 地址发送到二层交换引擎的 MAC 地址表中。从这以后，当 A 向 B 发送的数据包便全部交给二层交换处理，信息得以高速交换。由于仅仅在路由过程中才需要三层处理，绝大部分数据都通过二层交换转发，因此三层交换机的速度很快，接近二层交换机的速度，同时比相同路由器的价格低很多。

6.6.4 三层交换机种类

三层交换机可以根据其处理数据的不同而分为纯硬件和纯软件两大类。

6.6.4.1 纯硬件的三层技术

纯硬件的三层技术相对来说技术复杂，成本高，但是速度快，性能好，带负载能力强。其原理是，采用 ASIC 芯片，采用硬件的方式进行路由表的查找和刷新。

6.6.4.1.1 纯硬件三层交换机原理

当数据由端口接口芯片接收进来以后，首先在二层交换芯片中查找相应的目的 MAC 地址，如果查到，就进行二层转发，否则将数据送至三层引擎。在三层引擎中，ASIC 芯片查找相应的路由表信息，与数据的目的 IP 地址相比对，然后发送 ARP 数据包到目的主机，得到该主机的 MAC 地址，将 MAC 地址发到二层芯片，由二层芯片转发该数据包。

6.6.4.2 纯软件的三层技术

基于软件的三层交换机技术较简单，但速度较慢，不适合作为主干。其原理是，采用 CPU 用软件的方式查找路由表。

6.6.4.2.1 软件三层交换机原理

当数据由端口接口芯片接收进来以后，首先在二层交换芯片中查找相应的目的 MAC 地址，如果查到，就进行二层转发否则将数据送至 CPU。CPU 查找相应的路由表信息，与数据的目的 IP 地址相比对，然后发送 ARP 数据包到目的主机得到该主机的 MAC 地址，将 MAC 地址发到二层芯片，由二层芯片转发该数据包。因为低价 CPU 处理速度较慢，因此这种三层交换机处理速度较慢。

6.6.5 市场产品选型

近年来宽带 IP 网络建设成为热点，下面以适合定位于接入层或中小规模汇聚层的第三层交换机产品为例，介绍一些三层交换机的具体技术。在市场上的主流接入第三层交换机，主要有 Cisco 的 Catalyst 2948G-L3、Extreme 的 Summit24 和 AlliedTelesyn 的 Rapier24 等，这几款三层交换机产品各具特色，涵盖了三层交换机大部分应用特性。当然在选择第三层交换机时，用户可根据自己的需要，判断并选择上述产品或其他厂家的产品，如北电网络的 Passport/Acceler 系列、原 Cabletron 的 SSR 系列（在 Cabletron 一分四后，大部分 SSR 三层交换机已并入 Riverstone 公司）、Avaya 的 Cajun M 系列、3Com 的 Superstack3 4005 系列等。此外，国产网络厂商神州数码网络、TCL 网络、上海广电应确信、紫光网联、首信等都已推出了三层交换机产品。下面就其中三款产品进行介绍，使您能够较全面地了解三层交换机，并针对自己的情况选择合适的机型。

Cisco Catalyst 2948G-L3 交换机结合业界标准 IOS 提供完整解决方案，在版本 12.0(10)以上全面支持 IOS 访问控制列表 ACL，配合核心 Catalyst 6000，可完成端到端全面宽带城域网的建设（Catalyst 6000 使用 MSFC 模块完成其多层交换服务，并已停止使用 RSM 路由交换模块，IOS 版本 6.1 以上全面支持 ACL）。

Extreme 公司三层交换产品解决方案，能够提供独特的以太网带宽分配能力，切割单位为 500kbps 或 200kbps，服务供应商可以根据带宽使用量收费，可实现音频和视频的固定延迟传输。

AlliedTelesyn 公司 Rapier24 三层交换机提供的 PPPoE 特性，丰富和完善了用户认证计费手段，可适合多种接入网络，应用灵活，易于实现业务选择，同

时又保护目前用户的已有投资，另可配合 NAT（网络地址转换）和 DHCP 的 Server 等功能，为许多服务供应商看好。

总之，三层交换机从概念的提出到今天的普及应用，虽然只历经了几年的时间，但其扩展的功能也不断结合实际应用得到丰富。随着 ASIC 硬件芯片技术的发展和实际应用的推广，三层交换的技术与产品也会得到进一步发展

第7章 什么是第四层交换？

在建设企业网时，一些产品商所创造的术语令人迷惑。正当网络管理员认为他们正逐步弄清楚不同的第三层交换技术的差别时，一种十分具意义的新概念提出了：第四层交换。许多第三层交换技术要求采用专有的或厂家专用的协议，而与此相反，第四层交换是“厂商中立”的，即使加入到现成的网络环境中亦可受益。

7.1 第二/三层交换

首先是第二层交换，这个概念数年前由 Kalpana(现在的 CISCO)等公司提出。第二层交换是多端口网桥技术的重新包装，其性能和可扩展能力显著提高。这些产品传输帧基基于第二层以及网、令牌环网、或 FDDI MAC(介质访问控制）地址。有两类通用的第二层交换机；工作组交换机和网段交换机。工作组交换机产生独占网段交换机。工作组交换机产生独占网络带宽，为每个端点设备（如工作站或服务器）提供专门的ＬＡＮ网段；实际上取代了共享介质集线器。网段交换机被优化设计在多用户共享介质ＬＡＮ网段间或在ＬＡＮ主干间桥接流量；一般每个端口必须支持大量的ＭＡＣ地址。

后来第三层交换及所有与之相关的术语（如多层交换、ＩＰ交换、路由交换机）提出来了。但第三层交换技术实质上是路由，譬如在ＩＰ子网间交换流量。第三层交换试图减轻传统路由器带来的性能瓶颈－－在企业网流量分布偏离 80/20 规则且大多数流量必须跨越子网边界时显得越来越重要。大多数第三层交换技术可以归结为“路由一次，交换多次”，或者是基于高性能硬件的线速路由器。

当两个设备在不同子网间通信时，“路由一次，交换多次”技术试图使路由次数降至最低。这种技术通过分离路由的两个功能组件 — 路由计算和帧发送 — 减轻了潜在的性能下降。交换机根据与一个数据“流”关联的第一个数据包计算并建立通信路径一次（“路由一次”），然后对此数据流剩余包交换至同一路径（“交换多次”）；这样消除了进一步的路由计算。“路由一次，交换多次”技术的实例包括 Cabletron 公司的 SecureFast 虚拟网。

与此对照，线速路由器在硬件上实现了传统的路由功能，消除了基于软件的路由器的性能瓶颈。通过建造专门的路由专用集成电路（ASICs），这些产品可以把路由性能提高一个量级－－线速路由器以每秒百万包的速率发送流量；而传统基于软件的路由器仅能以每秒数十万包速率发送数据包。线速路由器提供的吞吐速率足以以全介质速率驱动多条千兆以太网链路。

7.2 进入第四层交换

如果第二层交换是网桥的再现，第三层交换是路由，那么，什么是第四层交换？

OSI 模型的第四层是传输层。传输层负责端对端通信， 即在网络源和目标系统间协调通信。在ＩＰ协议中，这是传输协议（ＴＣＰ）和用户数据报协议（ＵＤＰ）所在的协议层。

在第四层中，ＴＣＰ和ＵＤＰ标题包含端口号（port number)，它们可以唯一区分每个数据包包含那些应用协议(例如 HTTP、SMTP、FTP 等等)。 端点系统利用这种信息来区分包中的数据，尤其是端口号使一个接收端计算机系统能够确定它所收到的ＩＰ包类型，并把它交给合适的高层软件。端口号和设备ＩＰ地址的组合通常称作“插口”（socket)。

1 和 225 之间的端口号被保留，它们称为" 熟知" 端口；也就是说，在所有主机 TCP/IP 协议实现中,这些端口号是相同的。表一提供了这些“熟知”端口的例子。除了“熟知”端口外，标准 UNIX 服务分配在 256 到 1024 端口范围结果， 定制的应用一般在 1024 以上分配端口号。 分配端口号的最近清单可以在 RFC 1700"AssignedNumbers"上找到。TCP/IP 端口号提供的附加信息可以为网络交换机所利用，这是第四层交换的基础。

第四层交换的一个简单定义是：它是一种功能，它决定传输不仅仅依据 MAC 地址(第二层网桥)或源/目标 IP 地址(第三层路由),而且依据 TCP/UDP(第四层)应用端口号。

为了在企业网中行之有效，第四层交换必须提供与第三层线速路由器可比拟的性能。也就是说，第四层交换必须在所有端口以全介质速度操作，即使在多个千兆以太网连接上亦如此。千兆以太网速度等于以每秒 1488000 个数据包的最大速度路由(假定最坏的情形,即所有包为以及网定义的最小尺寸,长 64 字节)。

这为网络提供了在决定路由时区分应用的功能。例如，对关键应用流量(即 SAP R/3、Peoplesoft、Baan 定做开发的客户/ 服务应用）可以设定与基于 HTTP 的 Internet 流量不同的发送规则,即是它们都需要穿越同一套交换机/路由器接口。第四层交换机－－依据第四层信息（加上第二/ 三层标题）发送数据包－－如果它们即使在多个千兆以太网连接情形下也要获得全介质速度， 必须通过硬件来实现。Cabletron 的 SmartSwitch Router 是支持基于硬件的第四层交换的线速路由器。

在决定路由时检查第四层信息并不是一个新的概念。依据第四层信息定义路由过滤器是传统基于软件路由器的一个标准功能。这些路由器倾向采用第四层信息仅作为安全性考虑。过滤器（即强制规则）可以允许或不允许穿越路由器接口的流量。这使路由器作为一个包过滤防火墙来运作，因为路由器依据源/ 目标地址和端口号（即应用协议）对流量进行审查。通过硬件专用集成电路来实现第四层交换，这些交换机能够以线速度实现安全过滤器。

但是，第四层交换不仅是为在网段间提供防火墙而建立肯定和否定列表。第四层交换可以根据专门的应用进行流量排队，这为基于规则的服务质量机制提供了一条更可操作的途径。除了在两个设备间允许或不允许应用层通信外，第四层交换学提供了一种区分不同应用类型的方法 ;这些信息在进行发送决定时可以考虑进去。

第四层交换提供了以应用层为基础配置网络的工具。这和依据第二/ 三层信息配置网络形成对照，这些信息受网络和跨越网络构架进行通信的设备的制约。依据应用和网络设备定义安全过滤器和服务级别使网络获得了非常好的控制－－这在网络管理员必须为不同类型的应用提供不同服务质量级别时尤为重要。

第四层交换提供附加的硬件手段以每端口为基础收集应用层流量统计、依据第四信息的流量收集（加上第三层ＩＰ标题）增强了网络管理员排除网络故障的能力，为他们提供了网络使用更详尽的记帐以及流量基本活性支持。 如SmartSwitchRouter 支持每个端口的 RMON 统计。

应注意的是，进行第四层交换的交换机需要有区分和存贮大量发送表项的能力。交换机在一个企业网的核心时尤其如此。许多第二/ 三层交换机倾向发送表的大小与网络设备的数量成正比。对第四层交换机，这个数量必须乘以网络中使用的不同应用协议和会话的数量。因而发送表的大小随端点设备和应用类型数量的增长而迅速增长。第四层交换机设计者在设计其产品时需要考虑表的这种增长。大的表容量对制造支持线速发送第四层流量的高性能交换机至关重要。

7.3 结论

第四层交换机提供了支持高性能、可调谐网络构架所必需的功能，这些产品提供了：

·基于第四层信息的线速性能，即使在多个千兆以太网连接时亦如此
·一个路由器所要求的全部功能（如路由协议支持、安全过滤器、组播支持）
·基于应用的不同服务级别
·大的路由表能够支持基于第四层应用信息的线速泫量发送
·以及网络流量的增强管理能力

在一个企业的网络构架中能够提供不同层次的服务对于网络管理员来说正变得越来越重要。第四层交换使一个企业能建立依据特殊应用的类型的流量控制。这种强大的功能有助在今天和明天的企业网中提供需要的服务水平。

第8章 快速 IP 交换技术

8.1 概要

快速 IP(Fast IP)交换技术利用 NHRP 协议使源站和目的站获得对方的 MAC 地址，从而建立起一条交换连接。在此交换路径上可以实现直接的数据传输，大大减轻了路由器的负担。

由于数据业务量的不断增长，网络管理员已将交换技术作为建设高速局域网的基本选择。交换机在不同的物理网段之间提供线速的转发功能并在终端系统（end system）之间建立单一的逻辑网络。需要控制网络业务量的地方，交换机还可以建立虚拟网，即将终端系统按照它们的 MAC 地址、物理端口或网络协议类型分组。广播消息只局限在虚拟网内，虚拟网间的信息则经过路由器转发。高速

网络的出现和新兴的 Internet 的应用提高了网络业务量并改变了传统的通信模式。客户/服务器系统的出现使通信模式具有更多的变化。当信息源不再集中时，数据业务流也不再是可预见的，对网络的数据转发压力也比以往大得多。80%的网络业务量不再是包含在一个子网或虚拟网内了，更多的情况下，数据包必须在子网之间进行传送。但传统的路由技术缺乏扩展至千兆比特速度的能力。

Fast IP 是 3Com 公司为各种网络主干技术（包括以太网、快速以太网、千兆以太网、FDDI、令牌环和 ATM 等）提供的 IP 交换的战略措施，它综合了路由技术的控制功能与交换技术的线速（wire-speed）转发功能。 Fast IP 通过提高网络边缘系统的性能来支持新边缘模型（New Edge Model）所要实现的策略。它实现了“ 先寻径，后交换 ”的模型。它使大部分业务流绕过路由器并且在一个交换路径上直接传输，从而消除了路由器的处理过程。Fast IP 既可用于每秒数十万包转发率的高速网络，也可用于下一代的每秒数千万包转发率的网络技术。

8.2 Fast IP 的技术基础

____Fast IP 基于几个新近出现的标准：下一步进解析协议（NHRP，Next Hop Resolution Protocol）和 IEEE 802.1p/Q 等。Fast IP 的基础是 NHRP。虽然 802.1p/Q 有助于交换环境中的高效传输，但是 Fast IP 并不要求一定要具备 802.1p/Q。

8.2.1 非广播多路访问（NBMA，Non-Broadcast Multiple Access）网络的 NHRP 协议

8.2.1.1 NHRP 的基本概念

____出于管理和策略方面的考试，一个物理的 NBMA 子网可以分为几个分离的“ 逻辑 NBMA 子网 ”，也就是逻辑独立 IP 子网（LIS，Logical Independent Subnet）。逻辑独立子网有以下几个特征：

①逻辑独立子网的所有成员有相同 IP 网络/子网号和地址掩码；

②逻辑独立子网的所有成员直接与同一 NBMA 子网相连；

③通过一个路由器对在逻辑独立子网以外的所有主机和路由器进行访问；

④同一个逻辑独立子网的所有成员直接相互访问（不经路由器）。

____NHRP 是一个逻辑独立 IP 子网的地址解析协议而不是路由协议，它使一个在 NBMA 子网中的源站点可以确定通往目的站点的“ NBMA 下一步进 ”的网络层地址和 NBMA 子网层地址（如 MAC 地址）。如果目的站点与 NBMA 子网直接连接，NBMA 下一步进就是目的站点本身，NHRP 向源站点提供目的站点的 NBMA 子网层地址；如果目的站点不与 NBMA 子网直接连接，NBMA 下一步进是离开 NBMA 子网且靠目的站点“ 最近 ”的出口路由器，NHRP 则提供可以连接到目的站点的出口路由器的 NBMA 子网层地址。

____为了防止概念混淆，这里定义以下几个用语：

- 网络层：无须依赖介质的层（TCP/IP 网络的情况下就是 IP 层）。

- 子网层：在网络层下层且依赖于介质的层，包含 NBMA 技术。
- 服务器：通常是指下一步进服务器（NHS,Next Hop Server)。下一步进服务器是在 NBMA 网络中实现下一步进解析协议的实体，它总是和路由实体紧密地结合在一起。
- 客户：除非特别声明，通常是指下一步进解析协议客户（NHC,Next Hop Resolution Protocol Client)。下一步进解析协议客户引发各种 NHRP 请示以获得 NHRP 的服务。
- 站点：一般指包含 NHRP 实体的主机或路由器。

____在 NBMA 子网中存在一个或多个实现 NHRP 协议的实体。能够应答 NHRP 解析请求的站点就是“ 下一步进服务器 ”。每个服务器为一组目的主机服务，下一步进服务器在它们的逻辑 NBMA 子网中共同解析 NBMA 下一步进。下一步进服务器中有一个高速缓冲存储器（cache)，它存储由网络层地址转化到子网层地址的解析信息。这个存储器中的信息可以由 NHRP 注册包获得，或者由 NHRP 解析请求/应答包等其它途径获得。当逻辑独立子网中的站点不提供下一步进服务器功能时，在这个 NBMA 子网中就必须设置一个或多个下一步进服务器为它提供权威地址解析信息。这时，我们称下一步进服务器为这个站点“ 提供服务 ”。如果下一步进服务器可以为下一步进解析协议客户提供地址解析信息，那么在做解析请求的下一步进解析协议客户和目的下一步进解析协议客户之间路径上的每个中继点都必须存有下一步进服务器。而且，一个为下一步进解析协议客户提供服务的下一步进服务器应该与这个客户间有直接的子网层的连接，以便将预定的 NHRP 信息直接发送给它。也就是说，在路由路径上的最后一个 NHRP 实体就是“ 提供服务的下一步进服务器（serving NHS）”，NHRP 解析请求并不被转发到目的下一步进解析协议客户，而是由提供服务的下一步进服务器处理。下一步进解析协议客户也有一个高速缓冲存储器，存储由网络地址转化到子网层地址的解析信息。除了 NHRP 解析请求和应答以外，还可能会有以下 NHRP 包：

- NHRP 注册请求：本站点向下一步进服务器发送报告站点的 NBMA 信息。
- NHRP 注册应答：下一步进服务器应答客户的 NHRP 注册请求。
- NHRP 清除请求：下一步进服务器发向一个站点，清除它以前存储的信息；也可能从下一步进解析协议客户发向下一步进服务器，取消下一步进解析协议客户在这个下一步进服务器中的注册。
- NHRP 清除应答：告诉 NHRP 清除请求的发送者本站点已删除了所有指定的存储信息。
- NHRP 错误指示：向 NHRP 包的发送者传送各种错误信息。下一步进服务器会丢弃相应的包。

____NHRP 注册请求、NHRP 清除应答和 NHRP 错误指示所采用的寻路方式分别与 NHRP 解析请求和 NHRP 解析应答的相同。“ 请求 ”和“ 指示 ”通过由源站到目的站的路由路径。另一方面，“ 应答 ”则是通过由目的站到源站的路由路径。但在 NHRP 注册应答和下一步进解析协议客户引发 NHRP 清除请求的情况下，数据包必须经一个直接的虚电路回送；如果虚电路不存在，就必须当即创建。

8.2.1.2 NHRP 协议的进程

____如图 1 所示，一个事件触发 S 站需要解析 D 站子网层地址（通常是 S 站发送出要去 D 站的数据包，当然也有其他可能)。S 站首先通过正常寻路过程决定到 D

站的下一步进（对一个主机，下一步进可能就是默认路由器；对于路由器，则是网络层地址的“ 下一步进 ”）；如果其存储器中已经存有目的站点的地址解析信息，那么就使用这个信息来转发包，否则 S 站应生成一个 NHRP 解析请求包，其中包含目的地址（D 站的网络层地址）、源地址（S 站的网络层地址）和 S 站的子网层地址信息。S 站可能会指示希望接收到一个权威性的 NHRP 解析应答，并将 NHRP 解析请求包发向目的站点。

____如果 NHRP 解析请求是由一个数据包引发，那么当 S 站等待 NHRP 解析应答时，它会选择以下方法中的一个来处理这个数据包：

- 释放数据包；
- 保留数据包直到 NHRP 解析应答到达并且另外一条最佳路径可用为止；
- 沿着到 D 的路由路径发送数据包。

____其中方法③被推荐为默认策略，因为它可以在解析子网层地址的同时让数据流向目的站点。

____NHRP 解析请求在到达产生响应的站点之前在 NBMA 子网内经过一个或多个中继点。每个站点，包括源站点，根据目的站网络层地址和网络层路由表转发 NHRP 解析请求到下游下一步进服务器。当下一步进服务器接收到 NHRP 解析请求时，它会检查自己是否为 D 站提供服务 如果这个下一步进服务器不为 D 站提供服务，它将这个 NHRP 解析请求转发给另一个下一步进服务器；如果这个下一步进服务器为 D 站提供服务，它便解析 D 站的子网层地址信息，代表 D 站产生一个肯定的 NHRP 解析应答（这种情况下的 NHRP 解析应答被标记为“ 权威性的 ”，应答包中包含有 D 站的地址解析信息），并利用 NHRP 包中的源站网络地址通过路由路径发送 NHRP 解析应答。源站收到 NHRP 解析应答后，就可以利用包中目的站点的子网层地址与目的站点建立起子网层的直接连接了。

____当下一步进服务器接收了一个 NHRP 解析应答后会将应答中包含的地址解析信息存储起来。对于后继的 NHRP 解析请求，如果被允许的话，这个下一步进服务器会以存储着的、“ 非权威性 ”的地址解析信息进行响应。非权威性 NHRP 解析应答可以和权威性 NHRP 解析应答区分开。如果建立在非权威性地址信息基础上的通信失败的话，源站点还可以选择发送一个权威性 NHRP 解析要求。下一步进服务器不能以存储信息应答权威性 NHRP 解析要求。

____如果在 NBMA 子网中没有下一步进服务器可以应答对 D 站的 NHRP 解析请求，则回送一个否定的 NHRP 解析应答（NAK）。通常是由于所有下一步进服务器都没有对 D 站的下一步进解析信息，或者是下一步进服务器不能转发 NHRP 解析请求（即丢失连接）的缘故。如果客户已经试图在路径上建立一个捷径（shortcut）并且失败，那么客户会将网络层路径作为默认路径。

____如果 D 站不在 NBMA 子网上，下一步进就应是出口路由器，到 D 站的数据包经此路由器转发。NHRP 请求和 NHRP 应答不能不穿越 NBMA 子网的边界。因此，进出 NBMA 子网的网络层数据流总要通过在它边界的一个网络层路由器。对于这一问题正在作进一步的研究。

8.2.1.3 NHRP 的优点

____NHRP 最重要的优点是，在包含多个逻辑独立子网的 NBMA 中，它消除额外的路由器中继点。NHRP 利用 NBMA 网络的底层交换结构在主机间交换包，而无须路由器的介入，使主机间的通信更便捷。图 2（a）的网络没有配置 NHRP，路由器

1 需要通过一个 NBMA 主干与路由器 3 进行通信。如果路由器 1 和路由器 3 位于同一的物理网络但在两个不同的逻辑独立子网中，就要求路由器 1 通过路由器 2 与路由器 3 进行通信，其中路由器 2 是两个逻辑独立子网中的成员。但在配置了 NHRP 以后（图 2（b）所示），路由器 1 可以获得路由器 3 的子网层地址。这样可使路由器 1 以捷径方式绕过路由器 2 与路由器 3 直接通信。

8.2.2 IEEE 802.1p/Q

____IEEE 802.1p 是局域网和城域网对媒体访问控制网桥的业务类型简化和动态组播过滤的补充标准，它提供了基本的帧格式和协议使用的语义。Fast IP 将使用 820.1p 通用属性注册协议(GARP,Generic Attribute Registration Protocol)为虚拟网提供成员注册机制，并使交换机能映射和交换虚拟网的拓扑结构信息。
____IEEE 802.1Q 是虚拟局域网标准(Standard for Virtual Bridged Local Area Networks)，提供在虚拟网间进行映射的结构和协议。它成为虚拟网的标识及虚拟网之间的通信的一个标准，定义了基本的帧格式和协议操作。
____总的说来，这些标准定义了虚拟局域网的结构和虚拟局域网提供的服务，提供了加快业务流以支持实时信息传输的一个方法，支持动态使用小组 MAC 地址的滤波服务（组播）并准备了在局域网环境下提供这些服务所需要的有关协议和算法。

8.3 Fast IP 的操作

____Fast IP 采用主机到主机的模式实现 NHRP，以减少（或消除）通信主机之间的路由中继点来提高网络性能。Fast IP 的操作过程由终端站点和支持 Fast IP 的交换机发起。如图 3，终端系统（主机 A）要将普通交换结构内的信息传送到位于另外一个子网或虚拟局域网的一个主机（主机 B）。主机 A 启动一个标准 IP 通信进程并传送数据包到它的默认网关路由器，同时主机 A 通过路由路径传送一个 NHRP 请求到主机 B。NHRP 请求是一个标准格式的包，含源 MAC 地址，源和目的 IP 地址，并指出帧类型是 NHRP 类型的域。包中的数据部分含有源站点的 MAC 地址和虚拟局域网 ID（如果配置有 802.1p/Q 的话），它们将被接收方用于直接向源站发 NHRP 响应。路由器仍将保持它作为控制点的功能，并根据所设置的策略对包进行过滤或转发。当默认网关接收到 NHRP 请求时，如果策略允许，它就将 NHRP 请求转发给主机 B；如果策略不允许，它便释放 NHRP 请求，且业务流由路由路径从主机 A 流向主机 B。当主机 B 接收到主机 A 的 NHRP 请求时，它利用 NHRP 请求的数据部分提供的主机 A 的 MAC 地址发送一个 NHRP 响应。NHRP 响应穿过底层交换结构（而不是由路由路径）被直接转发回主机 A，沿途所经过的交换机将根据源 MAC 地址或虚拟局域网 ID 转发数据包。这保证了只要网络的基础设施是交换式的，NHRP 响应就能到达源站点。源站点一旦收到 NHRP 响应，则意味着虚拟网之间有一条交换式的连接。源站点就可沿着这条连接用目的 MAC 地址将数据包直接发给目的端站点，因而有效的绕过路由器并实现线速交换。如果未收到响应，源站点仍将继续通过默认的路由器转发数据包。
____需要注意的是，如果局域网是已有的交换和路由技术的网络，只需根据主机

A 的 MAC 地址将 NHRP 响应从主机 B 转发到主机 A。但是在某些情况下交换机没有设置源站点的地址，这时会要求一些交换机将发往源站点主机 B 的 NHRP 响应复制到所有端口。有一类交换机，可以控制未知目的 MAC 溢出，它们只会将响应复制到下游链路端口（也就是从边缘或桌面连接至一个核心交换机的端口）。因为服务器一般位于核心且核心知道大部分的主干传输业务，所以核心交换机很有可能已经知道了主机 A 的 MAC 地址。这样，未知目的 MAC 溢出只会造成很小的影响。这一点对使用不支持 802.1p/Q 的交换机升级到 Fast IP 时是很重要的。

____Fast IP 也有一些拓扑结构设计的限制。因为 NHRP 响应是在第二层交换结构被转发，在源和目的站点之间必须存在一条端到端交换路径。这就是说，像图 4 这样的拓扑结构不支持 Fast IP。这个拓扑结构的问题在于源站点和目的站点之间没有端到端的交换路径。如果允许路由器 1 和路由器 2（路由器无交换功能）发出 NHRP 请求在它们之间建立 Fast IP 捷径路径，在源和目的站点之间就可以产生一个从源站点到路由器 1，从路由器 1 到路由器 2，从路由器 2 到目的站点的三步 Fast IP 路径。但是，以现有的设计水平，路由器不能发出 NHRP 请求，只能响应终端点的请求。

____Fast IP 为网络管理员提供平稳的升级途径。Fast IP 功能的主要部分在于新的终端站点的软件，该软件提供 802.1p 虚拟网注册及 NHRP 地址解析协议。在终端站点无法升级的环境下，支持 802.1p,802.1Q 和 NHRP 的 Fast IP 交换机可以建立 Fast IP 连接。

____Fast IP 能与不支持 802.1p，802.1Q 和 NHRP 的交换机操作。一个实际的升级策略是将核心交换机升级成支持 Fast IP（802.1p，802.1Q 及 NHRP）。不支持 802.1p，802.1Q 的交换机将透明地转发虚拟网注册的包，虚拟网的学习和拓扑映射只发生在核心交换机中。不能辨认目的 MAC 地址或 802.1p 虚拟局域网 ID 的边缘交换机会将包转发到所有端口上。在目前的许多情况下，边缘交换机可被设置成只将未知地址的包转发到下游链路端口。虚拟局域网或 MAC 地址的直接转发将只发生在 Fast IP 核心交换机上。

____路由器的硬件和软件都无须改变。在 Fast IP 结构中，路由器将保持其执行第一层的过滤/防火墙策略的角色。当收到初始的 NHRP 请求时，路由器将会实施配置好的过滤/防火墙策略。例如，路由器可以滤掉 NHRP 类型的包，以此迫使所有的数据包通过该路由器。其它安全措施可由第三方的安全服务器、交换机的过滤或主机本身的安全机制提供。

8.4 Fast IP 代理（proxy）

____运行 Fast IP 要求在源站点和目的站点都安装 FIP 软件。然而，在有些特殊情况下，网络管理员不能控制终端系统。这样，Fast IP 代理只要求终端系统安装 FIP 软件（如图 5 所示）。当服务器应答一个客户请求时，这个服务会发出一个 NHRP 请求以期建立一条服务器到客户方向的数据流捷径路径。但客户不能应答服务器的 NHRP 请求，因为客户没有安装 FIP 软件。但是，运行 Fast IP 代理的第三层交换机可以代替客户响应并向服务器提供客户的 MAC 地址和虚拟局域网 ID。对拓扑结构的唯一要求是第三层交换机只能为和它有直接交换连接的客户做代理。客户和服务器之间的捷径路径是不对称的，只在服务器到客户的方向存在，而不存在于客户到服务器的方向。虽然只在一个方向上受益，但这已经大

大提高了网络的性能，因为大多数的业务流是从服务器流向客户的。

8.5 Fast IP 的优点

____Fast IP 被设计成可在多种网络体系结构上工作，此外，它所使用的基础技术（802.1p，802.1Q 或 NHRP）均不限于 TCP/IP，因此很容易扩展到其它协议。Fast IP 是目前唯一可用在多种主干网络技术和多种协议上的 IP 交换的解决方案。

____与其它 IP 交换方案不同，Fast IP 尤其适用在交换式局域网结构上。利用 Fast IP 在传输数据时不需要路由器来建立和维护交换式的连接，这会大大降低费用和复杂性。而且，Fast IP 并不集中在 ATM 技术上，因为目前只有少部分的局域网使用 ATM，而更多的 ATM 环境是采用混合技术的方案。而且从长远的观点看，多种类型的局域网结构仍将继续存在。

____Fast IP 使交换环境实现路由功能，这样可以使网络的吞吐量和性能都提高 4～5 倍。

8.6 小节

____Fast IP 是为局域网结构所需要的高性能所开发的 IP 交换技术。由于 Fast IP 只在建立连接的时候使用一次路由器，而在传输数据时完全把路由器从数据通道中旁路出去，所以它能以千兆比特的速度转发数据，并在虚拟网之间提供线速的通信能力。整个解决方案结合了交换式网络、高速路由和直通的通信路径三种技术。

第9章 交换机常见问题解答：

9.1 交换机与集线器有何区别？

答：集线器上的所有端口争用一个共享信道的带宽，因此随着网络节点数量的增加，数据传输量的增大，每节点的可用带宽将随之减少。集线器采用广播的形式传输数据，即向所有端口传送数据。交换机上的所有端口均有独享的信道带宽，以保证每个端口上数据的快速有效传输。交换机为用户提供的是独占的、点对点的连接，数据包只被发送到目的端口，而不会向所有端口发送。

交换机和集线器都遵循 IEEE802.3 或 IEEE802.3u，其介质存取方式均为 CSMA/CD。它们之间的区别为：

集线器为共享方式，既同一网段的机器共享固有的带宽，传输通过碰撞检测

进行，同一网段计算机越多，传输碰撞也越多，传输速率会变慢.

交换机每个端口为固定带宽，有独特的传输方式，传输速率不受计算机增加影响，其独特的 NWAY、全双工功能增加了交换机的使用范围和传输速度，目前倍受用户的青睐。

Nway 表示此端口可以自动识别 10M、100M 联接，全双工，半双工。

9.2 如何理解 Port Trunking 和 Port Mirror？

答："Port Trunking"即多干路冗余连接，可以理解为：将交换机上的多个端口在物理上连接起来，在逻辑上捆绑（bundle）在一起，形成一个拥有较大带宽的端口，组成一个干路。可以均衡负载，并提供冗余连接。"Port Mirror"即端口镜像，端口镜像为网络传输提供了备份通道。此外，还可以用于进行数据流量监测。可以这样理解：在端口 A 和端口 B 之间建立镜像关系，这样，通过端口 A 传输的数据将同时通过端口 B 传输，即使端口 A 处因传输线路等问题造成数据错误，还有端口 B 处的数据是可用的。

9.3 什么是 VLAN？

答：VLAN 即虚拟网络，VLAN 的划分有三种方式：基于端口（Port）、基于 MAC 地址和基于 IP 地址。通过划分虚拟网，可以把广播限制在各个虚拟网的范围内，从而减少整个网络范围内广播包的传输，提高了网络的传输效率；同时各虚拟网之间不能直接进行通讯，而必须通过路由器转发，为高级的安全控制提供了可能，增强了网络的安全性。

9.4 什么是基于端口的 VLAN ？

答：基于端口的 VLAN 是最实用的 VLAN。它保持了最普通常用的 VLAN 成员定义方法，配置也相当直观简单。 缺点是不够灵活，当成员位置移动后，网管员不得不对其重新进行配置。在基于端口的 VLAN 中，每个交换端口可以属于一个或多个 VLAN 组，比较适用于连接服务器。A+Link AP-4316 交换机支持最多两个端口干路及 16 组 VLAN。

问：我使用 Web 配置完 AP-4316 的 VLAN 后，却再也不能访问交换机了，只能通过未配置 VLAN 的端口来访问配置。

答：当你启用（Enable）IEEE 802.1Q VLAN 后，再配置菜单中有一个称为"Management VLAN ID"项，此选项用于设置管理此交换机的 VLAN 组。也就是说，只有由"Management VLAN "请求的包才会被交换机接受。确省的 Management Vlan ID 设置是 VID=1，因此，如果你想通过你所在的 VLAN ID 来管理交换机，请记住修改"Management Vlan ID" 值，使其符合与你的 Vlan ID 相匹配。另外，你的网卡必须支持 802.1q，才能正确使用 80.21q VLAN。

问：配置完 AP-4316 的 VLAN 后，交换机的 IP 地址丢失了？

答：在"Port management"（端口管理）项中,，检查 CPU 的 PVID 设定是否与工作站的 PVID 设定相同。

问：不能与其他厂家或型号的交换机结合使用 Port Trunking 功能。

答：不同厂家使用的芯片组会由不同的限制，所提供的 Port Trunking 功能也会不同。建议使用两台相同的交换机来使用 Port Trunking 功能。

9.5 Spanning Tree 有何作用？

答："Spanning Tree"即生成树，它是用于在设备间形成冗余的数据传输通路，同时避免数据传输时环路的形成。

问：A+Link 系列交换机在全/半双工模式下是如何进行流量控制的？

答：在全双工模式下支持 IEEE802.3x 流量控制（Flow Control）功能，在半双工模式下支持背压（Back-Pressure）流量控制功能。

问：当启用生成树协议后，交换机挂起或丢失 IP。

答：这种现象是正常的，因为交换机需要一定时间来计算每个端口的状态。大约 30 秒后，交换机会自动恢复。