基于OSI/RM的多层交换技术与应用

集美大学计算机系　郭大伟

▲ 基于OSI参考模型（OSI/RM）的多层交换技术是近年来发展最为迅速的网络技术。多层交换技术是交换和路由技术智能化的组合，它为局域网互连或广域网互连提供了一个完整、集成的解决方案。本文较深入地分析了多层交换技术概念以及所采用的技术，为用户在构造自己的Intranet时提供了可参考的方案。

1. 基于OSI/RM的分层交换

当今在网络界出现了众多的新技术、新名词和新标准，令人眼花缭乱，甚至无所适从。但是我们如果揭开各种新技术的面纱，花些时间认真研析一下的话，就会发现所有这些技术都只是在网络交换的方式上进行的创新和变革，利用交换技术的高速性能来传输数据，其核心依旧是基本的网络交换技术。正如80年代Sun公司提出具有前瞻性的口号"Network is computer：网络就是计算机"一样，当网络技术发展到今天时，我们可以有信心地宣称："Network is switching：网络就是交换"。

OSI制定的网络的七层结构模型对从事网络技术的人员来说并不陌生。现代网络已变得越来越复杂，为了便于分析和规划，将一个网络的体系结构从垂直方向分解为若干独立的网络层，单独地设计和运行每一层网络要比将整个网络作为单个实体来设计和运行要简单方便得多。基于这种认识，网络交换技术也是分层次实现的。我们这里所讨论的"多层交换"，实际就是指在OSI/RM的数据链路层（第2层）、网络层（第3层）和传输层（第4层）上实现的交换技术。

网络技术在其发展进程中曾经历过两次重大的变革：第一次是从主机/终端方式转为共享式（Shared Network）网络，第二次就是从共享式网络转变为交换式网络（Switching Network）。多层交换是相对于传统交换的概念而提出的。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，一般限定于局域网的应用，所以有人把第二层交换又称为LAN交换。然而，LAN交换技术并没为大规模的LAN建设提供一个完整、普遍的解决方案。这主要是由于传统的LAN交换技术不是完全可以扩充的。进入90年代以来，互联网络成为应用焦点，利用路由器通过WAN连接不同类型的局域网络已成为网络应用的主流。在大部分实际运行的网络中，LAN交换机必须与路由器相结合。但是路由器的端口价格及延迟是较严重的问题。今天，即使是最高档的主干路由器也难以应付由于VLAN、Intranet以及其他基于IP的应用所带来的迅猛增长的数据流量。怎样减少网络堵塞、优化网络结构、提高网络性能且扩大网络吞吐量，是网络管理员必须考虑的问题。

多层交换技术的出现有助于以下问题的解决：

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。

对于大型企业网和校园网，多层交换技术用来解决有多个子网而不同子网之间需要互通的场合。如不同的IP子网、IPX子网和虚拟子网的互连。

满足基于策略的服务，根据不同的数据流或网络流量分配不同的带宽或优先级。例如，高层交换机可依据基于TCP和UDP协议端口号判定Web传输流的类型，并可以将此数据包分类映射到服务质量保证中。可以为诸如话音和视频之类的时间敏感应用保留带宽，可以在更高的水平上应用安全策略。

多层交换技术是交换和路由技术智能化的组合，它为局域网互连或广域网互连提供了一个完整、集成的解决方案，这就是越来越多的用户在构造自己的Intranet时采用多层交换技术的原因。

2．基于数据链路层的第二层交换

在网络应用初期，大部分用户使用局域网，网络之间连接的带宽矛盾并不突出。作为一种简便易行的网络，共享式网络大行其道。这个时期，传统的连接设备是我们熟知的共享式集线器。随着对网络带宽的进一步的需求，工作组交换机替代了共享式集线器，随之也出现了第二层交换机，即我们所说的LAN交换机。

第二层交换(Layer 2 Switching)技术在数据链路层上进行操作，是面向局域网（LAN）的交换技术。LAN交换是解决网络堵塞、扩展网络带宽的主要选择之一。交换机作为一个真正的多端口桥接器，对共享式局域网提供了有效的网段划分解决方案，它可以使每个用户尽可能地分享到最大带宽。

LAN交换机中具有一定数量的物理端口来接连LAN网段，通常为8～128个。这些端口通过提取每个发送到交换机的数据包的源MAC(Media Access Control)地址，经交换引擎，然后在MAC地址表中得到端口目的地址，然后直接转送到目的端口。其过程如图1所示。

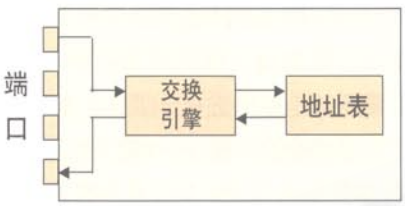

图 1　端口交换方式

第二层交换机产品大多数是基于端口的交换，其交换机的接口模块都是通过高速背板／总线(Core Bus)交换数据的，速率可以达到10Gb/s以上。背板／总线是指交换机中每个模块都通过连线直接连至其他模块，形成了全网状背板(图2)。由于每个模块都有自己的一组连接线，因而不必设置中央交换阵列。背板总容量等于连接线的总线(N×(N-1))乘以一条点对点链路的传输速度。

图 2　全网状背板式交换机结构

第二层网络交换技术虽然极大地扩展了网络，其不足之处也是明显的。它使网络恢复到了网桥的平铺拓扑结构，容易形成广播风暴。例如，当交换机收到一个不认识的IP包，也就是说目的MAC地址不能在地址表中找到时，交换机会把IP封包“扩散”出去，即把该包传到所有其他端口去，哪怕有些端口上连的是IPX工作站。这样一来，非TCP/IP节点的带宽便会受到影响。另外，第二层交换还有对异种网络之间互联的限制、安全性等问题，这促使我们在更高的网络层次上采用新的交换技术或设备。

3．基于网络层的第三层交换

网络层是OSI参考模型的第三层，它作为通信子网的最高层，负责将数据从物理连接的一端传送到另一端，包括寻址、路由选择、连接的建立、保持和终止等。基于OSI网络层的第三层交换(Layer 3 Switching)，是指在交换机内部完成不同子网间和虚拟网间的互连，从而改变了传统网络解决方案中由交换机外接路由器来完成局域网中不同IP子网、IPX子网和虚拟子网的互连。这样可大大减少原来采用路由器连接不同子网所带来的延迟、瓶颈等弊病。

从原理上讲，第三层交换技术是将第二层交换机和第三层路由器的优势结合成一个灵活的解决方案，可在各个层次提供线速性能(线速就是使交换速度达到传输线路上的数据传输速度)，消除交换瓶颈。实现线速交换的核心是ASIC（专用集成电路芯片）技术，用硬件实现协议解析和包转发。尽管目前似乎对第三层交换这一术语的解释不尽相同，但实际的做法就是在原有的第二层交换机内加入最新的ASIC的路由模块，即把与路由器有关的第三层路由硬件模块插接在交换机的高速背板/总线上，使路由从软件之中移出并移入ASIC之中，其成本远低于传统路由器的成本，优势显而易见。一个真正的第三层交换机并不是简单地将传统路由器加入到第二层交换机中的产品，它要使得路由模块可以与需要路由的其他模块间高速交换数据，使路由模块能真正达到线速的路由能力。这种改变不仅仅是硬件上的改变，还表现在软件的服务质量方面以及能提供极强的对网络流量的控制能力。

就广域网应用而言，IP实际上已经成为了网际互连事实上的标准协议。目前在Internet上，几乎所有新的应用

程序、以及现有的许多应用程序，都是在IP网络上编写并运行。IP在任何平台上都得到支持，已被证明是非常稳定、具有很强可扩展性的协议。第三层交换机在基于IP协议的应用中有着广泛的发展前景。Ipsilon网络公司最先提出利用ATM硬件来转发IP包，并称之为IP交换技术（IP Switching）。IP交换算法是通过ATM交换机把IP包分类，以加快路由的速度，甚至实现无阻塞路由。具体作法是在具有IP交换能力的交换机上只需读取同类型包的第一个包头，通过第一个包中所包含的信息可以掌握后续的包是和第一个包属同一类型的，这样就不用对后续包计算路由，使得它们可以快速通过路由器或者交换机，进而大大提高路由的效率。

值得注意的是，基于IP的第三层交换的实现技术目前还没有公认的标准，只有一些企业标准。主要有以下几种基于IP的交换技术及其应用领域：Ipsilon公司的IP Switching主要是利用ATM作为IP交换的硬件平台，并将其定位于局域网和广域网。Cisco公司的Tag Switching，它是一种典型的标记交换方法，Cisco公司将其定位于广域网，尤其是Internet和大型Intranet。Cascade公司的IP Navigator，它也是一种面向广域网的IP交换技术。3Com公司的Fast IP，定位于Intranet。ATM论坛的MPOA标准，定位于Intranet。

作为目前网络交换技术的热点，已出现较为成熟的各种基于第三层交换的交换机产品。例如Cisco的CoreBuilder 3500第三层高功能交换机可用作主干局域网络由设备来提供第三层的转发功能，从而取代局域网中的传统路由器。同样地，它也可以作为接入千兆以太网或ATM网的边缘设备。CoreBuilder 3500交换机可以用于以太网、快速以太网、千兆以太网、FDDI、ATM网，支持IP路由、IP组播路由、IPX和Apple Talk协议路由。同时，由于使用了先进的基于策略的服务机制，该交换机可以支持实时的多媒体网络通信。

4．基于传输层的第四层交换

OSI模型的第四层是传输层，它的作用是利用下面三层所提供的服务向高层提供可靠的端到端的透明数据传输，主要任务是提供进程间通信机制和保证数据传输的可靠性。传输层也是TCP（传输控制协议）和UDP（用户数据报协议）所在的协议层。

第四层交换（Layer 4 Switching）的实质是：在决定传输时不仅仅依据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。TCP协议是目前在Internet上广泛使用的一种基于连接的对话协议，例如FTP、Telnet等；而UDP协议则是一种在目标计算机描述信息如何到达应用软件的协议，UDP用于无连接通信，例如SNMP或SMTP。在协议层的应用中，网络可以通过监听协议所使用的端口来确定所接收到的IP包的类型，而端口号和设备IP地址的组合通常称作"套接字（socket）"。既然第四层交换使用了与特定应用有关的信息，利用这个信息可以完成大量的服务。例如通过查询其所接受的每个包内诸如TCP端口号之类的应用级信息，第四层交换机可以做出比第二层和第三层交换机更为明智的发送决策。

根据第四层交换的定义，我们认为第四层交换机应具有以下特性：

- 包过滤／安全控制

许多路由器被用于建立基于包过滤式的防火墙，第四层交换也具有这种能力，它对包的过滤能力是在ASIC中实现的。第四层交换机允许用户以LAN速度对通信量和防火墙功能进行优先考虑，可消除与防火墙认证有关的延迟，第四层交换可以在所有端口以线速操作，即使在千兆以太网上连接上也是如此，而其他交换机则会出现此类延迟。如果用户计划在自己的网络上支持没有延迟的防火墙安全性，那么采用第四层交换机是值得考虑的。

- 服务质量（QoS）

如果没有第四层交换，网络的服务质量／服务级别必然受制于第二层和第三层提供的信息。当因缺乏第四层信息而受到妨碍时，紧急应用的优先权就无从谈起。借助于第四层交换，TCP和UDP协议端口号告诉交换器生成传输流的应用程序的类型，然后，交换器则可以将此数据包分类映射到服务质量保证中。对关键应用流量可以设定与基于HTTP的Internet流量不同的发送规则，以区分优先级，于是紧急的应用可以获得网络的高级别服务。

- 网管统计

提供附加的硬件手段来以每端口为基础收集应用层流量统计。管理员使用第四层交换支持的统计特性，能够获得丰富网管信息。管理员不仅可以跟踪服务器和客户之间的数据，也可以很好地跟踪哪一个应用服务在工作、服务器上的活动和被打开的对话数等。

然而，严格地讲，术语"第四层交换"并不是很准确。ISO的第四层传输协议（包括UDP、TCP以及XNS），旨在确保可靠的数据传输。交换过程意味着源地址和目标地址之间的连接，然而第四层上并没有相应的寻址结构。事实上，若把所谓的"第四层交换"叫作"第二（或第三）层应用交换"，并将其功能视为一种路由过滤过程或许更为准确。

目前市场上已推出基于第四层交换的第1代产品，如Berkeley网络公司的exponeNT e4和Alteon网络公司的ACEswitch 180。实验测试表明，这两种性能良好、使用灵活的交换产品均

可很好地满足用户对更大网络吞吐能力的需求。

5．多层交换技术互相包容

我们在划分多层交换技术时，并不能简单地将某个第N层交换机认定为只能在第N层进行交换。实际上，交换机具有"向下包容"和"向上包容"的能力。一般来讲，位于高层（如第四层）的交换机都具有在低层次上（第三层或第二层）上进行交换的能力。而不具备一般性，某些交换机也具备有限的在高层交换的能力。

例如，第三层交换机不仅要承担路由功能，还要在数据链路层完成数据包的存储-转发任务，并且要达到与第二层交换机同样的包处理速度。在对一些第三层交换机性能所进行的测试中，记录到的最低延时与传统的主干路由器相比几乎低了一个数量级，证实了这种能力。

也可以在第二层采用"直通（Cut through）"技术的交换机上实现第三层交换。这种技术只对数据流的第一个包（或前几个包）作路由处理，而对其余的包进行交换处理，以提高交换性能。有些第三层交换机附加了一些增值软件，也具有辨别第四层协议端口的能力，但从严格的意义上讲，它还是在第三层进行交换操作，只不过是对第三层交换更加敏感而已。

基于策略的网络是刚刚崭露头角的新技术。在第二层上，我们还可以实现基于策略的服务。大多数的交换机和路由器都是把不同的数据流放入不同的队列来实现优先级策略。在完成转发之前，它们对数据包进行优先级标识，以通知其他设备如何处理该数据包。在具体实现有多种机制，其中最简单的就是IEEE 802.1p协议。由于工作在第二层，其实现较为容易，也不必关心网络上传输的是什么协议。

多协议标记交换(MPLS)也是近来研究的热点，Internet工程任务组(IETF)正在制定MPLS标准。MPLS把第2层性能与3层连接和网络业务组合在一起，以减少复杂性和其他IP/ ATM连接选择方案的成本。它还能在多种传输形式下提高网络服务质量QoS。

6．采用专用芯片构造多层交换机

一台交换机实际上就是一台计算机，因此也有自己的处理器。与衡量PC的标准是CPU一样，评价交换机性能的关键因素是其所采用的芯片。芯片作为交换机的引擎，其交换能力直接决定着交换机的处理能力。所以，芯片或芯片组的设计及它们如何与该交换机的其他部分集成，都是十分重要的。目前大多数交换机都是基于专用ASIC的，例如CoreBuilder 3500第三层高功能交换机，依赖于成熟的第三代ASIC技术，提供了线速的第二层和第三层通信能力，大幅度地提高了高端交换设备和路由设备的性能和性能／价格比。也有些交换机使用RISC芯片，但由于通用CPU芯片不是专为交换机设计的，所以工作效率比较低，如果多个端口同时工作，则会引起丢包、堵塞等状况发生。

现在最主要的交换机芯片生产商是TI，它的交换机套片称为Thunder Switch，有多种型号。芯片内部集成了交换引擎、网管监测引擎和多个MAC控制器。除TI外，还有一些公司提供从属性套片，如LevelOne公司提供以太网端口收发器芯片。

位于美国硅谷的NPL公司(Neo Paradigm Labs)推出的NP31X系列以太网交换机控制芯片，就是一种面向不同应用层次的低成本、高性能的产品。系统级IC NP315是一种5端口的快速以太网交换及可级联双速交换式集线器解决方案。其集成片上的MAC(介质访问控制器)支持5个10/100Mbps以太网端口。

Prizma是IBM专为交换技术定造的处理芯片。简而言之，Prizma芯片就是做在一片芯片上的交换机，实现了"交换在芯片"。与其他实现交换功能的技术相比，Prizma芯片的最大的特点是能够在单个芯片上完成全部的交换功能，而一般的方法是采用通用的CPU二次编码或ASIC芯片组合实现。Prizma芯片作为高速、大容量、支持多种协议的网络核心交换芯片，在一块芯片上，能完成16 × 16路、每路400Mbps ~ 1.78Gbps的信元交换能力。

由上所述，因为局域网交换机的端口和内部协议都已标准化，所以从原理上说，我们没有什么新东西需要自己设计。我们所要做的工作是如何高性能、低成本地实现它，而且在这方面也不用从零做起，因为有越来越多的厂商开始提供集成的交换机套片，以目前的商用套片实现桌面交换机是毫无问题的。一个芯片就是一个单片交换机，多个芯片组合成一个完整的交换机，不需要太多的外围电路。以NP315为例，使用NP315并配以适当的外围芯片可以构成面向不同层次应用的交换机产品。单独一块NP315芯片加外围电路可以构成5端口的以太网交换机。两块NP315级联可以构成8端口的以太网交换机。事实上，目前的交换机研制完全类似微机设计，利用功能完备的芯片组，做电路板级工作。

参考文献

[1] 鄂大伟，交换式网络的技术性能分析，集美大学学报，1998第3期
[2] 张一凡，多层交换，计算机世界，1998第一期
[3] 网络交换技术与产品，微电脑世界，1998第50期
[4] http://www.cisco.com
[5] 贾巍，第四层交换概述，网络世界，1999第12期