

Y 950381

学科门类 工 学
分类号 TP393

单位代码 10293
密 级

南京邮电大学

工程硕士研究生学位论文

论文题目：南京 IP 城域网网络优化及其
应用研究

研 究 生 李 磊
指 导 教 师 张登银 副研究员 叶祥云 高工
学 科 专 业 计算机技术
研 究 方 向 IP 城域网及相关技术
申 请 学 位 级 别 硕 士

二〇〇六年四月

南京邮电大学
工程硕士学位论文摘要

学科、专业：工 学 计算机技术
研究方向： IP 城域网及相关技术

作 者： 2004 级研究生 李 磊 指导教师 张登银 叶祥云

题 目： 南京 IP 城域网网络优化及其应用研究

英文题目： The optimization and application research of Nanjing
Metropolitan Area Network

主题词： 城域网 核心骨干网 业务接入控制层
IP 网络优化 TCP/IP QoS 组播 IPV6
Keywords: MAN Backbone Network BRAS
IP Network Optimization TCP/IP QoS
IP Multicast IPV6

摘 要

IP 城域网 (MAN, Metropolitan Area Network) 的概念是由计算机网深化而来, 指介于广域网和局域网之间, 在城市及郊区范围内实现信息传输与交换, 为全市各类用户提供宽带 (通常是指 2Mbit/s 以上) 接入的数字通信网络。对一个城市而言, IP 城域网的建设是其信息化基础设施的重要组成部分; 从技术和运营模式来看, IP 城域网是计算机网络和传统电信网络的融合; 从技术发展的趋势来看, IP 城域网是传统电信体系发展的必然趋势。

当前的网络是以千兆以太网为基础构筑的, 考虑 Internet 上的数据流量带有明显的逐层向上汇聚特征, 采用分层汇接式的组织结构, 网络采拓扑结构以星型为主, 进一步分为核心层、汇聚层、小区层和楼宇层四个层次。

IP 城域网的建设使城市的通信业务 (包括传统电信所能提供的电话、有线电视提供的电视节目传送、以及通过计算机网络所实现的上网业务、租用业务、虚拟专用网业务<VPN>、窄带拨号接入、视频业务、语音业务等等) 有机地整合在同一个 IP 宽带网络平台上。上述目标并非可以一蹴而就, IP 城域网的建设必须以现实技术为基础, 以客户的需求为导向, 采取循序渐进的策略逐步实施, 并且要不断的优化 IP 城域网以满足新业务对网络的要求。

主 题 词: 城域网、核心骨干网、 业务接入控制层、IP 网络优化、
TCP/IP 、QOS 、组播、IPV6

Abstract

The conception of IP Metropolitan Area Network(MAN) is the deepening of Computer Network. It refers to the digital communication network between Broad Area Network and Local Area Network, which accomplishes the information transferring and exchanging as well as providing Broad Band (Commonly above 2Bit/s) switch on service. To a metropolis, the construction of IP MAN is an important ingredient of its normalization infrastructure construction; from the standpoint of technology and operation mode, IP MAN is the amalgamation of Computer Network and traditional communication network; from the standpoint of technology development trend, IP MAN is the inevitable trend of traditional telecommunication system development.

The present network is constructed based on the Thousand Mbit ethernet. Concerning the data outflow of Internet, this network has obvious character of assembling upward layer of layer and adopts the organizational structure of delamination tandem. The structure of Network is further divides to four layers: core layer, assembling layer, building area layer and building layer.

The construction of IP MAN makes the communication businesses of a metropolis (including the telephone provided by traditional telecommunication; the TV program transferring provided by cable TV; the Internet business, renting business, VPN, narrowband dialing business, video frequency business, audio business accomplished through computer network etc.) integrated on the same IP broadband network platform. The target mentioned above cannot be accomplished at one move; the construction of IP MAN must be based on present technology and guided by the client's request, adopting the step-by-step strategy and continually optimize the IP MAN to satisfy the network requests of new businesses.

Keywords: MAN、Backbone Network、BRAS、IP Network Optimization、
TCP/IP、QOS、IP Multicast、IPV6

南京邮电大学学位论文独创性声明

本人声明所呈交的学位论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知,除了文中特别加以标注和致谢的地方外,论文中不包含其他人已经发表或撰写过的研究成果,也不包含为获得南京邮电大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

研究生签名: 杨 日期: 2006.4.6

南京邮电大学学位论文使用授权声明

南京邮电大学、中国科学技术信息研究所、国家图书馆有权保留本人所送交学位论文的复印件和电子文档,可以采用影印、缩印或其他复制手段保存论文。本人电子文档的内容和纸质论文的内容相一致。除在保密期内的保密论文外,允许论文被查阅和借阅,可以公布(包括刊登)论文的全部或部分内容。论文的公布(包括刊登)授权南京邮电大学研究生部办理。

研究生签名: 杨 导师签名: 张学银 日期: 2006.4.6

第一章 概 述

随着国民经济的持续高速发展以及计算机的日益普及特别是数据通信的爆炸性增长, 社会各行各业对通信的需求越来越高。同时, 通信技术的发展又为用户提供了丰富的业务, 其中数据业务的发展对整个社会的影响尤为突出。电子商务、视频点播、远程医疗, 远程教育等新业务的诞生促进了宽带网络的发展。因此, 新一代的宽带网络不仅具有多样化的接入方式, 而且是快速转发数据的核心, 并能够提供丰富的增值业务。

南京电信宽带 IP 城域网从 2000 年开始建设, 几年来宽带用户发展迅速, 截至 2004 年 10 月 20 号, 宽带网 ADSL+LAN 已有用户 23 万左右, 其中 ADSL 用户大约 18.5 万, LAN 用户近 4.5 万。预计到 2005 年年底宽带用户总数将发展到 40 万, 其中 ADSL 用户数和 LAN 用户数预计为 4:1 的比例, 即 ADSL 用户 32 万, LAN 用户为 8 万, 同时到了 2005 年年底将有 60%ADSL 高端用户的下行带宽从 1M 提升至 2M。

为了保证在宽带用户数迅速扩大以及增值业务、VPN 等多种业务迅速发展的情况下, 南京 IP 城域网有足够的承载能力和转发性能为用户提供服务, 我们在对现有 IP 城域网的运行情况进行了全面的分析得到的客观数据的基础上提出了南京 IP 城域网优化方案并加以了实施。本文将研究南京 IP 城域网的优化方案。首先, 先介绍一下 IP 城域网的相关概念及业务特点。

1.1 城域网的基本概念

1、城域网的定义

城域网(Metropolitan Area Network, MAN)是在一个城市范围内所建立的计算机通信网, 简称 MAN。这是 90 年代初, 在局域网(LAN)的发展基础上提出的, 在技术上与 LAN 有许多相似之处, 而与广域网(WAN)区别较大。

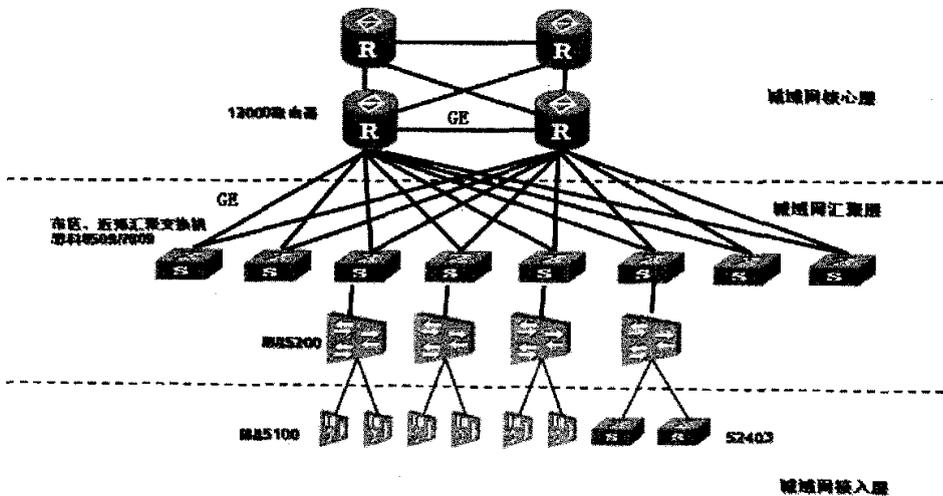
MAN 的传输媒介主要采用光缆, 传输速率在 1000 兆比特/秒以上。所有联网设备均通过专用连接装置与媒介相联, 只是媒质访问控制在实现方法上与

LAN 不同。

MAN 的一个重要用途是用作骨干网，通过它将位于同一城市内不同地点的主机、数据库，以及 LAN 等互相联接起来，这与 WAN 的作用有相似之处，但两者在实现方法与性能上有很大差别。MAN 不仅用于计算机通信，同时可用于传输语音、图像等信息，成为一种综合利用的通信网，但属于计算机通信网的范畴，不同于综合业务通信网(ISDN)。

2. 城域网的基本结构

城域网一般分为骨干层、汇接层和接入层(如图 1-1)。



如图 1-1 城域网基本结构

骨干层的主要功能是给业务汇接点提供高容量的业务承载与交换通道，实现各叠加网的互联互通；汇接层主要是给业务接入点提供业务的汇聚、管理和分发处理；接入层则是利用光纤、同轴电缆、无线接入技术等传输介质，实现与用户连接，并进行业务和带宽的分配。

3. 宽带 IP 城域网的特点

IP 城域网是可以用来承载各种不同的业务，比如数据业务、语音业务、视

频业务等，但是除了用途更广泛、带宽更高以外，宽带 IP 城域网还有许多特色。

(1) 宽带(Bandwidth)

宽带 IP 城域网拥有比过去传统网络更高的带宽，比如过去网络接入速率上大多使用 64K/128K/256Kbps，而现在宽带接入则多使用 1M/10M/100M/1000Mbps；过去在骨干网上大多使用百兆或数个百兆网，而现在则多使用千兆或数千兆网，甚至 10 千兆网。带宽增加会激发新型业务的出现，而新型业务的推动也会带动带宽的需求，带宽与业务因此会有良性的互动与增长。因此，城域网的建设必须兼顾现有的带宽处理能力(Performance)与未来的扩充能力 (Scalability)，才能在不断成长的需求中体现业务平滑升级与设备投资保障的好处。此外在城域网中提供无阻碍(non-blocking)的传输及处理能力也是服务提供商的竞争优势之一。

(2) 带宽管理(Bandwidth Management)

带宽的扩充并不是惟一的途径，有效地使用或管理带宽也是一个重要的课题，比如 OSPF 或 BGP4 协议中的负载分流、MPLS 协议中的流量工程 (TE)、接入网中的指定速率(Rate Limiting)。根据用户的不同等级分别给予不同的带宽，根据使用者的付费分配带宽，都是带宽管理的例子。

(3)服务质量 (Quality of Service)

服务质量(QoS) 是指 IP 网络的一种能力，是为特定的业务提供其所需要的服务。在有限的 IP 城域网资源下，按使用的原则付费，也可以对用户划分不同的服务等级，享有不同的资源；即使在同一用户中，也可以依据业务不同而有不同的服务等级，如多媒体业务(Voice、Video)优于数据(Data)业务，E-mail 服务高于 Web Browser 服务等。

(4)冗余备份(Redundancy)

IP 城域网逐渐成为各种不同业务的传输平台，人们对它的可靠性也就要求更高，因此城域网必须有电信级的冗余配置，包括设备、部件（电源、风扇、模块、软件）的冗余；线路的冗余，如 PoS 的自动保护转换(APS)、百兆/千兆以太网的捆绑技术(Tracking)；第二层的扩展树 (STP) 或快速扩展树(Rapid STP)、第二层的热备份协议(VRRP、VRRP-E)、或路由备份(OSPF 及 BGP4)。在城域网的各个地方配置相应的冗余备份可大大提高网络的可靠度。

(5)安全与控制(Security & Control)

整个城域网是一个大型的开放平台,网络设备必须提供多种安全功能以保护自己及用户不受黑客攻击,多种等级的登录(Login)、Secure Shell(SSH)、Secure Copy(SCP)及 HTTPS 是设备自我保护诸多功能中较为人所熟悉的,而防止阻绝服务(DoS)攻击、防止 Ping to Death 攻击、防止地址欺骗(Smurf)攻击则是较常见的用户保护功能。除此之外,如果网络设备能够提供线速访问控制功能,也可以对网络的安全或控制起到相当大的作用,同时也可以在网络环境中制定不同的策略(Policy),进行所谓策略路由的功能。

(6)IP 统计与计费功能(IP Statistics & IP Billing)

IP 的统计及计费功能可以分为不同的层次,最简单的可以利用 SNMP 中的 MIB 来实现,其次可以利用基于取样技术的扩展 RMON(Extended RMON:XRMON)来实现,或者更进一步用 Netflow 的技术基于数据流(traffic flow)的方式来采集统计资料与计费,IP 城域网中有了统计与计费功能后,不但可以对用户进行细微的计费,同时也可以收集网络的流量资料,作为网络优化与升级的参考。

(7)多种业务的传输平台(Multi-Services)

IP 城域网的建设可以提供传统的 DDN 数据服务(低速专线、帧中继),窄带接入(RAS)、宽带接入(Broad band RAS)、住宅小区、局域网到局域网,企业虚拟专线服务(VPN)、IP 电话、视讯组播(Multicast IP)、视频点播(Video on Demand)等各种不同的业务。

4、宽带 IP 城域网提供的业务

(1) 宽带上网

随着各种宽带技术的成熟和普及,可以向用户提供 xDSL、LAN、Wireless 等多种方式的接入选择,高带宽的接入使用户享受高质量的网络访问,促进数据业务的不断进步。用户快速便捷地通过网络获得更多更丰富的信息,WWW 不再是"World Wide Waiting"。

(2) 企业 VOIP (FOIP) 业务提供

(3) VOD、MP3 音乐、交互游戏、IPTV

网络带宽和服务质量的提高,使 VOD、MP3 音乐、交互游戏等对带宽、服务质量、延时要求较高的业务得以大量增长,这也是宽带网所能带给人们的最接近生活的享受之一。

(4) 远程教育、远程医疗

多媒体远程教育系统以 Video/Audio 工业的最复杂的压缩和传送技术为基础,采用标准 RTP 协议并且把高品质的视频/音频和 HTML 页面紧密结合,实现视频、音频、图像和文字教学材料在网上的实时同步传输。远程医疗是使用远程通信技术和计算机多媒体技术提供医学信息和服务。

(5) 会议电视、可视电话

会议电视就是通过传输信道在多个地点进行实时性视频交互通讯的一种联系手段。宽带城域网的建设为电视会议提供了可靠的网络平台,基于 IP、基于 ATM 的电视会议在企业中的应用越来越多。可视电话使人们的远距沟通更加生动和亲切。

(6) VPN

VPN 的需求一直是各类企业,尤其是跨地区的企业所急需的,城域网的宽带化为大量开展 VPN 业务提供了网络保证。

(7) 端口出租、主机托管、虚拟 ISP

对于网络运营商来说,对各类 ISP 和 ICP 提供多样化的服务和灵活的网络资源组合,是取得效益增长的一个有效途径。

1.2 本文的内容及安排

从目前网络技术的发展趋势看,通信网络向 IP 的融合是不可逆转的大趋势,IP 将成为各种通信应用统一的网络平台。英国老牌运营商 BT 正在打造其新一代统一网络平台(21 世纪网络),中国电信也在建设其新一代的 IP 骨干网 CN2,意图作为下一代多业务承载平台,统一承载 NGN、3G、IPTV、企业互联及互联网等各种业务。城域网作为 IP 网络最贴近用户和业务的部分,其多业务承载能力,对于通信网络的 IP 融合至关重要。以多业务承载能力的提升为目标的城域网改

造优化，将是国内外电信运营商关注的重点。

相对于国家骨干网，城域网规模庞大、技术复杂，其优化改造工作涉及大量的投资，本论文将对南京 IP 城域网优化改造中的一些技术热点问题进行探讨和分析。

本论文具体内容安排如下：

第一章为绪论，阐述了城域网的概念，并对 IP 城域网的网络结构、网络特点及所承载的业务做了简单介绍。

第二章介绍了南京 IP 城域网的现状，分析了 IP 核心骨干层、业务汇聚层及边缘接入层的问题，并通过网络流量分析模型对现网的 MSR、LSR 及增值业务网络设备进行了流量分析，得出了南京宽带城域网优化改造的必要性。

第三章介绍了南京 IP 城域网网络优化的原则、思路，提出了南京 IP 网优化改造的网络模型。随后，对南京 IP 城域网核心骨干层、业务接入控制层、用户边缘接入层分别提出了优化改造方案，并对其中的主要设备进行了介绍。

第四章介绍了当前 Qos 和组播的技术，并根据南京 IP 城域网的特点制定了 Qos 及组播部署方案。通过 Qos 及组播部署，我们要把南京城域网建设为一个能提供全面业务支撑能力、能满足不同客户要求、具备优良扩展性和稳定性的综合业务支撑平台。

第五章介绍了南京 IP 城域网 Qos 和组播的部署结果及还存在的问题，最后介绍了南京 Qos 和组播的相关组网方式及应用。

最后总结了南京电信 IP 城域网优化改造的系统化的解决思路，并展望了下一代 IP 网络协议—IPV6 技术的特点及优点，采用 IPV6 技术来组建 IP 电信网势在必行。

新庄、云南路、长乐路、大行宫的 MSR 作为城域网的总出口，每节点通过 1 条 2.5G POS 与宽带网出口路由器 R4 或 R5 相连。由 R4/R5 通过 10G POS 连到省网路由器 A1/A2(CISCO12816), 实现与省内其它地市及出省用户的互通。此外, 2 台 Foudry8000 与 R4/R5 通过双 GE 互连, 提供 163 窄带拨号用户的接入。

2、会聚层节点 LSR

共设置 LSR 27 个, 均采用 CISCO Catalyst 6509。分别为:

东区局——大行宫、金鹰、光华门、后宰门、苜蓿园

西区局——龙江、南湖、汉中门

南区局——长乐路、红花村、能仁里

北区局——新庄、中央门、吉祥庵、尧化门

中区局——鼓楼、云南路、马鞍山、兴中门

江北区局——泰山新村、大厂、浦口

五县局——江宁、溧水、高淳、六合、江浦

27 台 LSR 设备, 形成区域汇聚层, 分别负担全市东、西、南、中、北、江北及五县的接入。LSR 通过 GE 接入相应的 MSR 设备(双归属), 其中一对 GE 接入云南路或新庄的出网千兆核心节点上, 另一对 GE 接入长乐路或大行宫的千兆核心节点上。

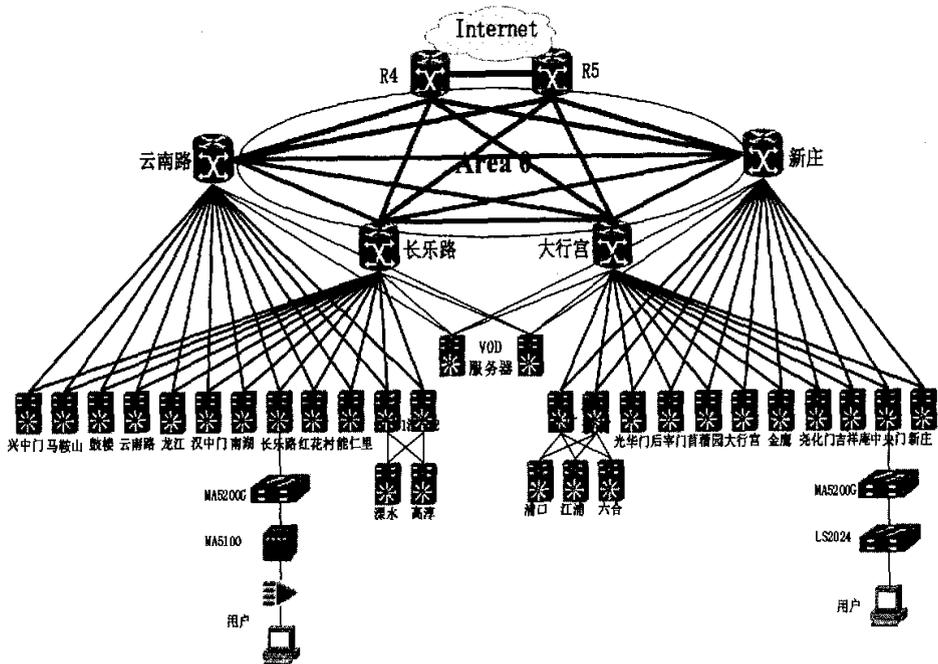
由于传输资源原因, 高淳 LSR 挂在溧水 LSR 下; 浦口、江浦、六合 LSR 挂在大厂 LSR 下。

3、接入层节点 ZAN

前期的宽带建设工程在所有的模块局、分纤点及小区中心安装 ZAN 设备, 使用的华为公司的 MA5200, 总共安装 ZAN 交换机 200 多个, 形成市区和郊县的全面覆盖。各模块 ZAN 节点通过一条光路千兆上连到 LSR 上。

4、边缘接入层

ZAN (MA5200) 下连的是多个楼道交换机和 ADSL 设备, 它们都是百兆上连到接入层节点 ZAN 上的。



图(2-2)南京电信 IP 城域网网络拓扑现状

南京市 IP 城域网建成以后，覆盖了整个南京市的东、西、南、中、北、江北区局及五县，已经成为南京市的高速 IP 通讯基础平台，满足了大量用户的宽带上网、VPN、局域网互连等业务需求。

但是现有的城域网自 2001 年宽带工程竣工以来一直没有改变，而南京宽带用户的数量和需求都在不断变化，整个 IP 网络的结构及性能都已暴露出了一些需要改进的地方。

2.2 南京 IP 城域网结构分析

1、核心扩展性问题

核心 7609 空槽位不够，不利于 BIP 网络扩展。现有 4 台 7609 均只剩余一个空槽位，可扩展性有限。

2、MPLS VPN 网络结构问题

核心 7609 不支持 MPLS VPN，导致现有 VPN 网络没有核心层，无法扩展。

现有的 MPLS VPN 网络是在长乐路，鼓楼，大行宫和新庄的 6509 上增加一块 OSM-4GE 模块，做为 PE 接入 VPN 用户。由于 OSM 共只有 4 个端口，加上没有核心 P 设备对 PE 进行汇聚，VPN 网络的规模和可扩展性都比较差。对 VPN 用户来说，从接入点到 PE 要经过 2~3 个 6509 的 VLAN 透传转接，不仅浪费局间光路中继，而且在网络繁忙的情况下服务质量很难保证；对运营商来说，为了接入 N 个用户，就要在核心网中透传 N 个 2 层 VLAN，用户自家的 2 层网络可能会和运营商的 2 层网络的生成树连在一起，容易引起网络振荡，而关闭生成树协议，又不能自动检测 2 层网络可能产生的环路，容易引发广播风暴，不仅影响 VPN 用户，也影响普通的宽带用户。

3、策略路由（PBR）的问题

目前网上保留地址还有几百个用户，导致 PIX 和 7609 的策略路由不能去掉，大量的 ACE（Access List Entry）占用了 7609 的 TCAM 资源。7609 使用 TCAM 硬件处理访问列表，处理性能和普通包转发性能相同，即在硬件处理的情况下 7609 对访问列表的处理可以达到线速 30Mpps。而在用户流量比较大的云南路节点，PBR 使用的 ACE 对 TCAM 的占用率已经超过了 90%，一旦 ACE 超出 TCAM 的容量，访问列表的处理将从硬件处理转为 SUP2 引擎上的 CPU 纯软件处理，会造成节点崩溃，网络阻断。

4、6509 上联带宽的问题

6509 上联到 7609 的总带宽为 4G，分别连到 2 台 7609。但是为了保证使用保留地址的用户可以进过 PIX 上网，必须保证一个用户流出网和入网在同一台 PIX 上完成（在同一台 7609 上），这样只能把一台 7609 做为主用，另外一台做为备用，带宽闲置 50%。

5、MA5200 性能不足

城域网在用的分布式 BRAS 和部分集中式 BRAS 不能满足新业务（如 NGN、IPTV、VPN）承载的需求。现有 MA5200 容量偏小，由于 MA5200/MA5200F 支持的并发 PPPOE 数 6000-2000，以致 BRAS 网元设备数量多；老 MA5200 的上行处理能力为 350Mpps，不能满足宽带用户的发展。

6、边缘接入层的性能不足

现网的边缘接入层都是采用百兆上连 MA5200 的，边缘接入层的设备性能不足，不能承载 IPTV 业务，不支持 QOS 和组播。

2.3 网络业务需求分析

目前南京宽带 IP 骨干网支持多种宽带业务，其中已经开放的业务包括：

- 1) ADSL, LAN 方式接入的宽带上网业务
- 2) 以 3 层 VPN 为主的 VPN 业务
- 3) VOD、IPTV、网络游戏等增值业务

其中宽带上网业务用户最多，增长最快，增值业务流量不太确定，VPN 正在发展中。在前几次讨论的基础上，我们将根据不同的业务分类，对宽带骨干网（汇聚层+核心层）的带宽需求进行分类分析。

2.4 网络数据分析模型

由于目前 IP 网络的分析模型还处于起步阶段，一般以工程经验模型为网络分析的依据。具体的说，就是：

1、带宽扩容可以链路峰值利用率做为参考，峰值利用率大于 70%~80%就必须扩容带宽；

2、转发性能分析可以峰值转发性能和转发总性能的百分比做为参考，若峰值转发性能大于总性能的 70%~80%，必须对系统进行升级。

3、设备 CPU 占用率：因为 CISCO 6509 采用 CEF (CISCO Express Forwarding) 技术和 dCEF (Distributed CEF) 分布式 CEF 技术，包转发由接口上的专用 ASIC 完成，SUP2 引擎和 DFC 上的 CPU 不参与具体的数据包转发，而只从事配置更新，计算路由，SNMP 管理，IP Accounting 等管理工作，所以单看 6509 的 CPU 占用率和用户流量没有什么实际对应关系，一般而言，6509 只运行静态路由，考虑到运行了 SNMP，NDE 数据采集等功能，CPU 占用率应该在 15%以下，而 7609 运行了动态路由协议，CPU 占用率应该在 30%以下。

4、设备内存占用率：SUP2 和 DFC 上的内存用于放置路由表、转发信息表和部分端口缓冲。

5、TCAM 占用率：TCAM 是用于处理访问列表的硬件，只要 TCAM 的占

用率不超过 100%，7609 访问列表的处理就是硬件方式，线速处理。

6、网络平均时延，一般来说从宽带网网管 ping 到宽带网中任意一台设备，平均时延应该小于 10ms，如果平均时延超出 10ms 太多（如大于 150ms）就因该仔细检查该设备的工作情况，是负载过大还是受到恶意攻击。

2.5 现有网络核心设备的标称性能

目前南京 BIP 城域网中的 CISCO 6509/7609 都配置了 SUP2 引擎和 SFM 交换模块。系统背板交换容量均为 256Gbps，系统转发能力根据 DFC 分布式转发卡的配置不同而稍有差别，SUP2 引擎有 30Mpps 的转发性能，每块 DFC 可以给系统增加 15Mpps 的转发性能（以包大小 64 字节为测试标准）。

此外，OSM 模块不占用系统的转发性能，每一块 OSM 模块具有 6Mpps 的转发性能。概括如下：

7609：SUP2 引擎+3 块 DFC+1 块 OSM，总的转发性能为 $30+3\times 15+6=81\text{Mpps}$

普通的 **6509：**SUP2 引擎+2 块 DFC，总的转发性能为 $30+2\times 15=60\text{Mpps}$

做 PE 的 **6509：**SUP2 引擎+2 块 DFC+1 块 OSM，总的转发性能为 $30+2\times 15+6=66\text{Mpps}$ 。

2.6 现有网络设备流量数据采集分析

2.6.1 流量采集说明及分析模型

我们从 2004 年 2 月 1 号开始，对南京宽带城域网中的 4 台 MSR，27 台 LSR，101 台 MA5200（ZAN）的用户流量数据进行了采集、分析，在此基础上对现有宽带网能承载多少用户，以及为了满足到 2005 年年底用户的需求，网络需要做什么调整进行了分析讨论。

截至 2004 年 1 月底，南京宽带网已有 ADSL+LAN 用户 11 万左右，其中 ADSL 用户数大约占 80%。预计到 2004 年年底，用户总数将增加 12 万，达到 23 万左右，2005 年年底预计达到 40 万左右。如果假设 ADSL 用户数占总用户的 80%，同时可能有 60%的 ADSL 用户下行带宽从 1M 提升到 2M，那么用户流量计算公式为：

2004 年用户流量 = ADSL 流量 + 非 ADSL 流量

$$= \text{目前流量} \times \text{ADSL 流量百分比} \times (23/11) \times 0.6 \times 2 + \text{目前流量} \times \text{ADSL 流量百分比} \times (23/11) \times 0.4 \times 1 + \text{目前流量} \times \text{LAN 流量百分比} \times (23/11)$$

$$2005 \text{ 年用户流量} = \text{ADSL 流量} + \text{非 ADSL 流量}$$

$$= \text{目前流量} \times \text{ADSL 流量百分比} \times (40/11) \times 0.6 \times 2 + \text{目前流量} \times \text{ADSL 流量百分比} \times (40/11) \times 0.4 \times 1 + \text{目前流量} \times \text{LAN 流量百分比} \times (40/11)$$

$$\text{带宽} = \text{用户流量} / \text{带宽占用率}$$

$$= \text{用户流量} / 0.7 \quad (\text{假设带宽占用率到了 } 70\% \text{ 就应该考虑扩容})$$

因为各个分局业务发展不平均, 有的地方 LAN 用户多, 有的地方 ADSL 用户多, 有的地方个人用户多, 有的地方网吧用户多, 所以每个 LSR 上 ADSL 流量的比例都不相同, 而且很难预测发展趋势, 但是我们可以目前的数据为基准, 根据用户流量模型, 分析各个 LSR、MSR 到了 2004 年和 2005 年的用户流量和带宽需求, 从而判断网络是否需要扩容。

2.6.2 LSR 流量分析结果

目前各个 LSR 上联带宽为 4G, 分别接到 2 个 MSR 上, 由于有 PIX 的限制, 实际只用了 2G, 待 PiX 退网完成以后, 可以充分使用 4G 的上联带宽。

根据最新的数据分析, 27 台 LSR 在 2004 年和 2005 年的上联带宽需求为:

LSR	预测 2004 用户总流量(考虑 ADSL 提速)带宽需求(Gbps)(按峰值流量占带宽 70%计算)	预测 2005 用户总流量(考虑 ADSL 提速)带宽需求(Gbps)(按峰值流量占带宽 70%计算)
鼓楼 LSR	2	3
云南路 LSR	4	6
马鞍山 LSR	2	2
兴中门 LSR	2	3
汉中门 LSR	2	3
龙江 LSR	2	2
南湖 LSR	2	2

长乐路 LSR	2	3
能仁里 LSR	2	2
红花村 LSR	2	2
江宁 LSR	2	2
溧水 LSR	2	2
高淳 LSR	2	2
新庄 LSR	2	2
中央门 LSR	2	2
尧化门 LSR	2	2
吉祥庵 LSR	2	2
大厂 LSR	2	2
江浦 LSR	2	2
浦口 LSR	2	2
六合 LSR	2	2
泰山新村 LSR	2	2
大行宫 LSR	2	3
金鹰 LSR	2	2
光华门 LSR	2	3
后宰门 LSR	2	3
苜蓿园 LSR	2	2

可以看出,目前 4G 上联可以满足各个 LSR 到 2004 年底的用户需要,到了 2005 年只有云南路 LSR 可能需要增加 2 个 GE 上联。

2.6.3 MSR 流量分析结果

1、带宽需求分析

对于核心层 MSR 来说,从 LSR 接入的用户流量不一定全部是访问 Internet 的,也有一部分分流在 MSR 之间互联的链路上,其中很大一部分是访问南京城域网内部的服务。对于访问城域网内部服务的流量,我们考虑到除了电信局提供的 VOD 等,还有用户托管主机以及各种私人服务器,所以不好预测 2004 年和 2005 年访问城域网内部服务的流量模型,只能按照访问 Internet 的用户流量模型

进行估计。MSR 上访问 Internet 的用户流量模型和 LSR 的相同。

(1)MSR 上访问 Internet 的用户流量对带宽的需求如下:

核心层设备	应有物理带宽 Gbps(2004年)按照带宽占用率 70%计算	应有物理带宽 Gbps (2005年)按照带宽占用率 70%计算
云南路 7609	4	7
长乐路 7609	6	9
新庄 7609	3	4
大行宫 7609	5	8

可见,为了满足用户流量需求,最终需要采用 10G 端口,全网出口共 40G。

同时,我们从流量模型公式可以反过来推出现有 10G 的网络出口可以容纳用户数(假设 ADSL 不提速)约为 21 万户。

(2)MSR 之间互联流量对带宽的需求:

预计 2004 年城域网核心互联应有带宽 (Gbps):

核心层设备	云南路 7609 (Mbps)	长乐路 7609 (Mbps)	新庄 7609 (Mbps)	大行宫 7609 (Mbps)
云南路 7609	N/A	2	2	2
长乐路 7609	2	N/A	2	2
新庄 7609	2	2	N/A	2
大行宫 7609	2	2	2	N/A

预计 2005 年城域网核心互联应有带宽 (Gbps):

核心层设备	云南路 7609 (Mbps)	长乐路 7609 (Mbps)	新庄 7609 (Mbps)	大行宫 7609 (Mbps)
云南路 7609	N/A	3	2	4
长乐路 7609	3	N/A	2	2
新庄 7609	2	2	N/A	3
大行宫 7609	4	2	3	N/A

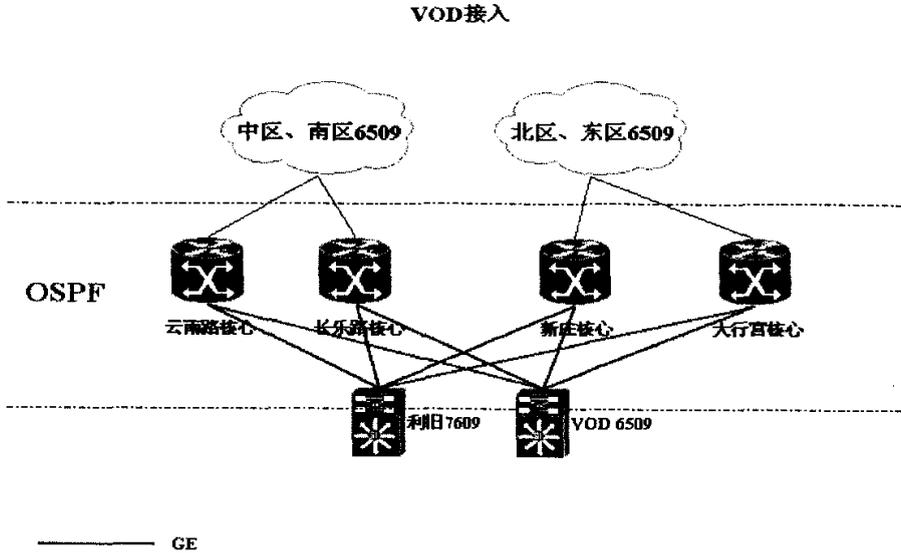
可见,核心设备互联可以保持现有 2GE 不变。每台核心设备上有 6 个 GE 口用户核心互联。

2、MSR 转发性能分析

到了 2005 年底各个 MSR 的模块转发性能都将超过 70%,必须对现有 MSR 进行系统升级,提高转发性能。

2.7 增值业务带宽需求分析

为了减少增值业务流量对 MSR 之间互联带宽需求的影响, 以及提高 VOD 接入系统的性能, 我们利旧一台 7609 给增值业务部。



图(2-3) 增值业务网络图

考虑增值业务流量, 4 台核心设备每台需要再增加 2 个 GE 端口。

2.8 总结

为了接应企业战略转型要求, 积极稳妥推进网络平滑演进, 将南京 IP 城域网打造成“用户可识别、业务可区分、质量可控制、网络可管理”的电信级网络, 我们必须对南京城域网进行优化改造, 使之成为一张结构清晰、功能齐全、控制集中, 能够满足语音、视频、数据业务承载要求, 具备差异化服务能力的综合业务承载网络。

第三章 南京 IP 城域网网络优化

3.1 南京电信 IP 城域网优化的目标

IP 城域网优化建设的目的是立足现有业务，使 IP 城域网高度可控、易管理和可持续发展，简而言之，城域网优化需要解决的几个突出问题就是：用户的管理（包括用户定位、控制）、业务的管理（包括 QoS、业务策略）、网络融合（多网融合和多业务融合）。基于以上分析，目标网络应该具备清晰的网络层次，便于规划、管理和维护；具备业务的智能，可以对业务进行识别，并根据业务类型实施多样化的业务策略和服务质量；具备多业务能力，可以支持多个业务网络，满足未来网络和业务发展的需求。

1、网络层次清晰化：二三层网络分离，构建物理和逻辑层次清晰的三层路由网络（IP 城域网）和二层宽带接入网。

2、网络结构扁平化：大容量、少节点、广覆盖，减少物理和逻辑级联级数。

3、管理控制集中化：用宽带接入服务器（BRAS）、业务路由器（SR）构建清晰的 IP 城域网边缘—业务接入控制层（简称接入层），实现集中的业务提供和控制；同时规范设备的网管接口要求，加强集中网管系统的建设，提高网络的可管理性，实现运营高级网络管理。

4、网络质量差异化：网络部署 Diffserv 机制，为不同用户和不同业务提供不同 QoS 等级的服务。

5、商用规模的组播能力：城域网承载 IPTV 业务，须重点解决对组播的支持、接入网的带宽和 Qos 保证。

6、接入汇聚层采用大容量的 L3 设备：高密度 GE/FE 端口汇聚，IPTV 终端管理、QinQ。

7、满足未来 2--3 年宽带业务发展需求。

3.2 南京电信 IP 城域网优化思路及优化模型

1、建立明确的 IP 路由网层面：由核心层、业务接入控制层组成，BRAS 采用直挂方式接入核心层设备。

2、新增大容量、转发性能好、端口密度高、单端口平均价格低的 BRAS 设备。

3、根据业务发展需要，逐步将现网 DSLAM 割接到新 BRAS 设备，原有 MA5200/5200F 将轻载 LAN 用户。

4、逐步启用 QoS、组播等技术，支持 IPTV 等业务。

5、逐步部署及新建 SR 设备，实现 PE 及以太专线网关等功能。

6、根据核心层设备流量递增情况和工程建设周期，逐步提升链路带宽。

7、采用大容量高性能设备作为汇聚交换机，支持 SVLAN，部署支持 802.1P（至少支持两个队列）的二层交换机，需具备组播复制能力。

8、部分分布式小 BRAS 关闭 BRAS 功能，作为汇聚层二层交换机使用。

9、对商务写字楼宇的电话、宽带等业务的满足，逐步引入宽窄带综合接入设备。

10、加强 MSTP 网络的建设和利用，作为大客户的接入平面以及重要的企业互连客户互通平面。

11、增强 VPN 网络接入能力。

根据南京电信 IP 城域网优化思路的要求，我们制定了相应的网络优化模型如图(3-1)

目标网络结构模型框图

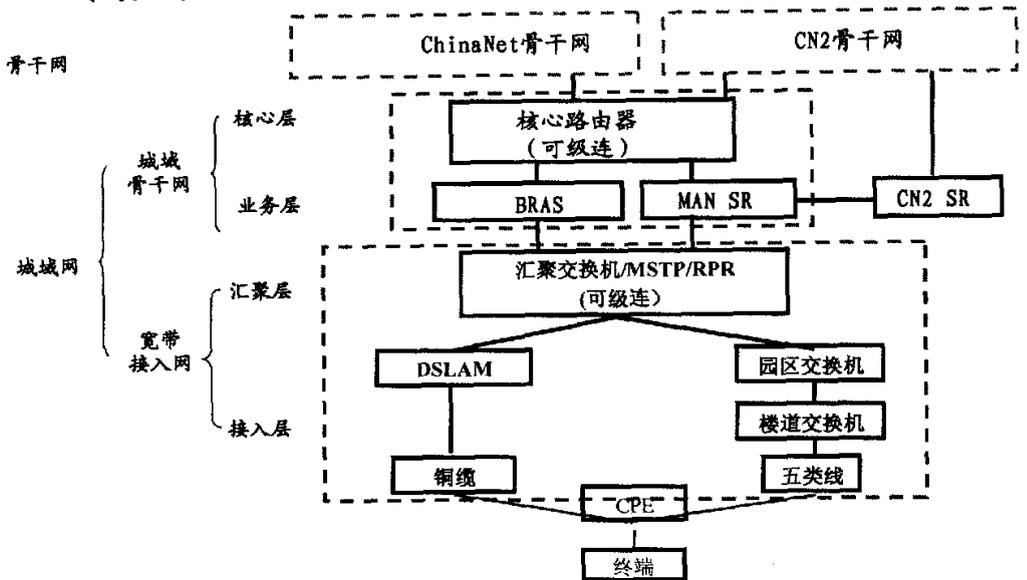


图 (3-1) 网络优化模型

3.3 南京 IP 城域网核心骨干层优化改造

3.3.1 核心骨干层优化改造方案

在几乎所有的大型 ISP 的网络中基本上都使用了大型路由器作为组网的核心设备，总的来看，路由器经历了组建大型 IP 网络的考验，无论从支持的协议类型、端口的缓冲能力、设备的稳定性、对大型 IP 网络的适应性(大量的、无序的、突发数据流的交换)、所支持端口的多样性(OC-48 POS/OC-192POS、DPT、ATM、GE/10GE)等方面来看，都要优于交换机的组网方式。而且，采用路由器作为组网的骨干设备，将会为网络的多用途打下基础(尤其是 MPLS VPN)。

南京核心骨干层优化改造采用 4 台 CISCO12416 替换掉原来的 4 台核心交换机：CISCO 7609。网络仍然以 R4、R5 做为出口核心路由器，4 个核心 GSR 各有 2 个 10G POS 端口和 R4、R5 相连。由于传输资源有限，现在泰山新村、大厂、江宁到 GSR 采用 ZX GE 裸光纤连接的方式。溧水 LSR 上联到长乐路 MSR，高淳 LSR 挂在溧水 LSR 之下。

南京 IP 城域网骨干层优化改造后的拓扑图如图 (3-2)：

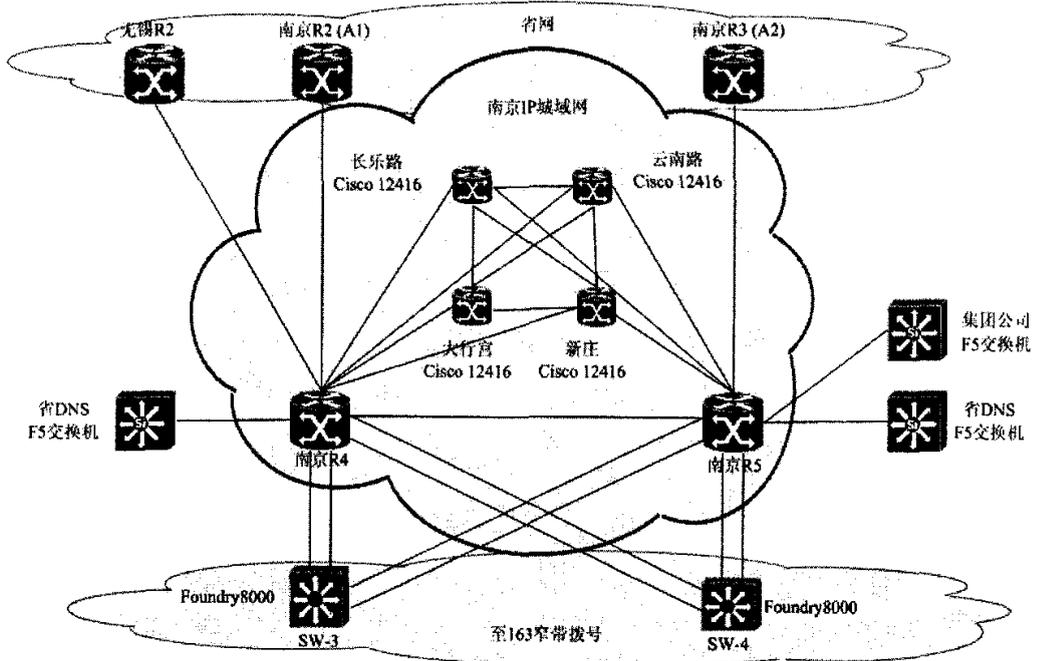


图 (3-2) 南京 IP 城域网骨干层优化改造后的拓扑图

3.3.2 吉比特处理路由器（GSR）的特点：

1、可用槽位多，除了主控板 PRP 占用 2 个槽位以外，剩下 14 个槽位都可以用于接入用户，加上现在 GSR 12416 系列已经支持大密度的 GE 模块（每个模块 10 个 GE 端口），可持续扩展性更强；

2、GSR 的所有端口都支持 MPLS，支持 3 层 MPLS VPN 和 2 层 MPLS VPN，业务功能丰富；

3、转发性能强。GSR 是建立在高速分布式路由体系之上的，该体系包含先进的交换核心，能以千兆位的速率传送第三层路由。PRP 运算出的转发表被发布到每个板卡中，每个板卡在自己的转发表的基础上独立地寻找接收到的数据包的目的地址。各个模块在持续峰值流量下仍然能以线速进行转发。

考虑到设备性能，业务能力和网络可持续发展等多种因素，本次工程改造如下：现有核心层设备不再采用 7609 路由交换机，而采用 CISCO 的核心路由器 GSR12416，以提供丰富的业务功能和可持续发展性。

3.3.3 CISCO12416 (GSR) 的性能分析

本次配置了 4 台核心路由器：CISCO12416。均配置了直流系统，并配置了 320Gbps 的交换矩阵。产品号为：GSR16/320-DC。

GSR 各单元介绍如下：

1、机架子单元

CISCO12416 包含了 3 个子框（上子框、下子框和交换矩阵框）和两个风扇模块。系统的电源架是在机架上方附属的分离单元。

上子框有 8 个用户可配置的槽道，支持线路卡和 RP 的安装。最左边的槽道是装告警卡的专用槽道。最右边的槽道留给 RP 安装（槽道 7）。剩余的槽道（槽道 0 至 6）可装路由器支持的任意线路卡类型。

下子框也有 8 个用户可配置的槽道，支持更多的线路卡并有一个可选的/冗余的 RP。下子框是和上子框是反转的关系，卡以相同的方式装，但左右顺序是反转的。下子框的最右边的槽道用来安装第二块告警卡。由于路由器配置了冗余的 RP，它必须安装在下子框的最左边的槽位：槽道 8。

2、机架背板

机架后部的无源系统背板提供了三个子框的电气连接。机架背板分配直流电源给所有的卡和风扇模块，并提供了卡间的物理通讯通路，不仅是网络数据而且还有内部系统维护总线间的系统通信通路。

3、制冷

机架上方和底部的风扇模块给三个子框提供冷风。在电源单元中，每个电源模块包含了一个风扇从前面抽入冷风并将热风强排到后方。

4、电源

DC 输入电源子系统有四个 DC 电源输入模块。在全冗余电源配置下，模块 A1 和 B1 提供了系统负载区 1 的电源冗余：上风扇模块和上插板子框；模块 A2 和 B2 提供了系统负载区 2 的电源冗余：交换矩阵框、下风扇模块和下插板子框。

5、交换矩阵

路由器的核心是一个交换矩阵，提供线路卡和 RP 间的同步吉比特速度连接。CISCO12416 有 10G bps 交换矩阵（320Gbps 交换系统带宽）——包含 GSR16/320-CSC 和 GSR16/320-SFC。每个 SFC 或 CSC 卡提供 10Gbps 全双工连

接到每个线路卡。因此，有 16 个槽道板卡的 CISCO12416，每槽有 2*10Gbps 的容量(全双工)，系统交换带宽是 16*20Gbps=320Gbps。可支持线路卡接口的速率至 OC-192c/STM-48c。

交换矩阵子框包含 2 块时钟和调度卡 (CSC) 和 3 块交换矩阵卡 (SFC)。交换矩阵子框中的 1 块 CSC 和 3 块 SFC 处于激活状态，第 2 块 CSC 提供冗余。CSC 安装在槽道 0 和 1 (标为 CSC0、CSC1)；SFC 卡安装在槽 2、3、4 (标为 SFC0、SFC1、SFC2)。

钟和调度卡 (CSC) 提供下列功能：

- 1) 交换矩阵——运载线路卡之间的业务或 RP 和线路卡之间的业务。
- 2) 调度——处理从线路卡来的请求以接入到交换矩阵并决定何时允许线路卡接入到交换矩阵。
- 3) 系统时钟——同步通过交换矩阵的线路卡之间的转发数据或线路卡和 RP 之间的转发数据。
- 4) 交换矩阵卡仅包含交换矩阵电路，携带了线路卡或 RP 和线路卡之间的业务。SFC 接受 CSC 来的调度信息和系统时钟。

6、路由处理器

每个 CISCO12416 路由器有一个主系统处理器 (或称为路由处理器)。路由处理器 (RP) 处理网络路由协议并发布更新到线路卡上的 CEF 表。RP 也执行一般的维护功能，如：诊断，Console 支持，线路卡监控。路由器的 RP 有两种类型：Gigabit Route Processor (GRP)和 Performance Route Processor (PRP)。本次优化方案中 CISCO12416 均配置 PRP 卡。

PRP 产品号是 PRP-1=，包含了一个 PRP (512MB SDRAM 以及一个 64MB ATA Flash disk)。

PRP 的主要功能如下：

- 1) 加电时向所有的安装线路卡下载 Cisco IOS 软件；
- 2) 为路由器的配置提供一个 console 端口；
- 3) 为其它外设提供一个 aux 端口；
- 4) 提供 2 个 IEEE 802.3 10/100M 以太网口以实现 Telnet 功能；
- 5) 运行路由协议；
- 6) 建立和分布路由表到线路卡；

7) 提供一般系统维护功能;

8) 提供线路卡通过交换矩阵或 MBus 之间的通讯。

9) MBus 连接允许 PRP 下载系统 bootstrap image, 收集或调用诊断信息, 执行一般/内部维护功能。交换矩阵连接是线路卡和 RP 之间路由表发布以及包转发的主要的数据通路。

7、维护总线 (MBus)

MBus 管理着系统全部的维护功能。MBus 包含了两个冗余的总线, 连接了所有的线路卡、RP、交换矩阵卡、告警卡、电源单元和风扇单元。

MBus 模块执行如下功能: 上电和断电控制; 设备发现功能; 代码下载; 环境监测和告警; 诊断功能。

3.4 南京 IP 城域网业务接入控制层优化改造

业务接入控制层是宽带 IP 城域网业务接入的控制点, 是城域骨干网和宽带接入网的衔接点, 完成业务区分、用户管理、计费信息采集等业务接入控制功能。业务接入控制层由宽带接入服务器 (BRAS) 和业务路由器 (SR) 两种业务接入控制点组成, 是 IP 城域网实现多业务承载, 向用户可识别、业务可区分、质量可控制、网络可管理的电信级网络演进的关键环节, 在建设时需要重点考虑设备在业务实现、识别、控制、保障等方面的综合能力。

3.4.1 BRAS 的定位及性能要求

1、BRAS 的定位

BRAS 的功能分为两部分: 业务功能模块和接入功能模块。业务功能模块包括互联网网关业务功能、组播网关功能和 MPLS PE 业务功能; 接入功能模块包括拨号接入功能、专线接入功能。BRAS 主要实现中小客户拨号接入互联网业务、专线接入互联网业务、拨号接入 MPLS VPN 业务、专线接入 MPLS VPN 业务和组播类业务。在 BRAS 功能和性能满足 SR 要求的前提下, 对于大客户专线、重要业务接入需求较少的节点, BRAS 可兼作 SR。

2、BRAS 性能要求

1、新建 BRAS 应重点关注下列性能和功能要求:

2、满足 MPLS PE 的业务功能、具备较强的 MPLS VPN 处理能力, 支持至

少 500 个 VRF、VPN 路由表容量至少 5 万条；

3、具备 DHCP Server、Relay 和 DHCP 认证模式下的安全控制功能；

4、实现 DiffServ QoS 机制，包括速率限制、分类、标记和流量整形，同时提供对 DSCP、802.1P、MPLS EXP 位进行标记的手段，并支持三者之间的相互映射，具备层次化 QOS 能力；

5、具备基于 PPP、Vlan 和端口的组播包复制功能，应支持组播频道的快速加入和离开、10K 组播路由条目、200 个组播组，组播用户数量应大于最大用户数量的 1/3。

6、动态带宽调整，支持 COPS 协议。

3、BRAS 对 MPLS VPN 和 IPTV 业务的支持

随着 L3 MPLS VPN 业务的发展，通过 ADSL 接入 MPLS VPN 的业务需求迅速增长。业务接入模式有两种：对于 ADSL 拨号用户，BRAS 终结 PPPoE Session 后，将用户 PVC/Vlan 绑定到相应的虚拟路由表（VRF）里；对于 ADSL 专线用户，BRAS 直接将用户的 PVC/Vlan 绑定到相应的 VRF 里。在建设应根据集中式 BRAS 的 MPLS VPN 业务接入能力，综合规划 L3 MPLS VPN 的覆盖范围，优先覆盖有业务需求的端局。

BRAS 作为 IPTV 组播业务 PIM 路由协议的边界，IGMP 请求的终结点，通过静态组加入配置应可将所有组播流量推送到 BRAS。在业务开展初期少量用户阶段，采用 PPPoE 双 Session 接入方式，由 BRAS 完成基于 PPPoE Session 的组播复制。在密集用户阶段，组播报文由 BRAS 向用户接入点逐级复制，由用户接入点实现面向用户的跨 Vlan 组播复制。

3.4.2 SR 的定位及性能要求

1、业务路由器（SR）的定位

SR 的功能分为两部分：业务功能模块和接入功能模块。业务功能模块包括互联网网关业务功能、组播网关功能和 MPLS PE 业务功能；接入功能模块包括专线接入功能。SR 主要为大客户提供专线接入互联网业务、专线接入 MPLS VPN 业务和组播类业务。

2、SR 性能要求

新建 SR 应重点关注下列性能和功能要求：

- 1、考虑未来业务量的发展，应采用 10G 平台核心路由器；
- 2、具备 NSF（不间断转发）、GR（平稳重启）功能，提高网络可靠性；
- 3、具备支持 L3 MPLS VPN 能力，支持至少 500 个 VRF、VPN 路由表容量至少 20 万条；
- 4、具备 BFD 链路级故障快速检测功能和路由快速收敛、MPLS FRR 网络级故障快速恢复功能；
- 5、实现 DiffServ QoS 机制，包括速率限制、分类、标记和流量整形，同时提供对 DSCP、802.1P、MPLS EXP 位进行标记的手段，并支持三者之间的相互映射；具备层次化 QoS 能力，满足大客户差异化服务的需求；
- 6、具备 MPLS 多路径 LSP 的负荷分担，暂不支持的 SR 应能通过软件升级满足需求。

3、SR 对 MPLS VPN 业务的支持

L2 MPLS VPN 技术用 IP/MPLS 网络承载二层数据链路包（如 ATM 信元、FR 帧、以太网帧），可以实现 IP 网和基础数据网的融合。新建 SR 应具备点到点模式 VLL 的业务开放能力、多点模式(VPLS)的业务开放或后续支持的能力。在 VPLS 标准化完成后，具备快速开放 VPLS 业务的能力。在设备选型时应重点关注不同厂商 SR 的业务互通性和 MAC 地址表容量。

3.4.3 南京 IP 城域网业务接入控制层优化改造

我们决定在南京 IP 城域网中引进大容量、高性能的 BRAS（以下简称大 BRAS），大 BRAS 共计 42 台，设备为：华为 MA5200G 和 Juniper ERX1440，分别建在鼓楼、汉中门、大行宫、光华门、长乐路等 42 各局点，从而做到南京全区的宽带业务覆盖。每台大 BRAS 分别千兆上连两台 GSR，千兆下连接入层交换机。同时，基于投资保护的原则，原有的网络不做改动。

接入控制层网络拓扑示意图如图（3-3）：

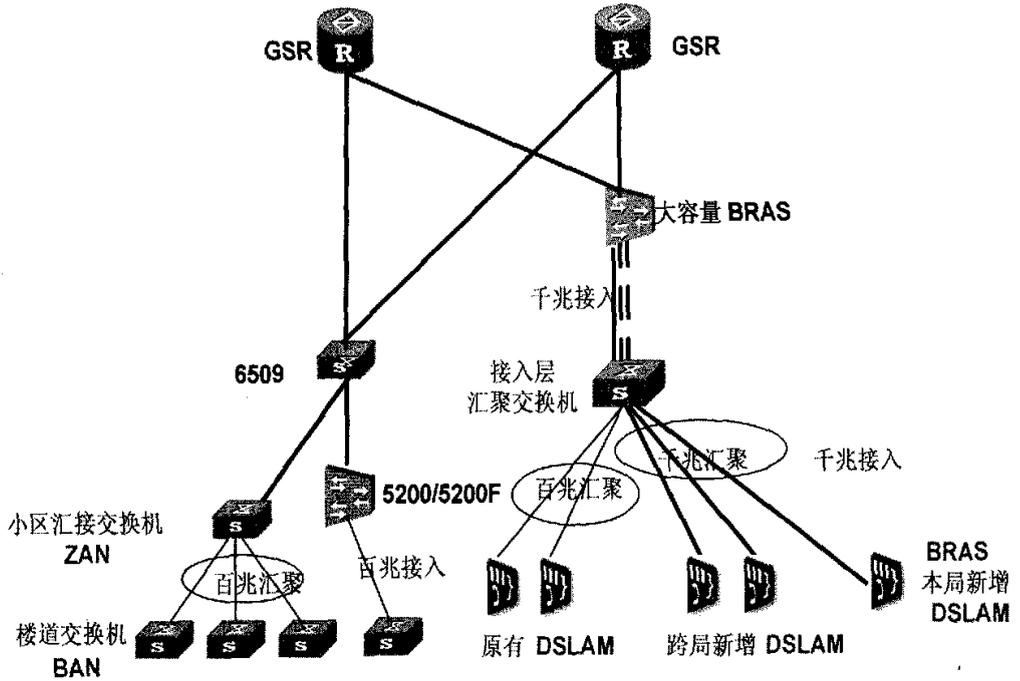


图 (3-3) 接入控制层网络拓扑示意图

1、MA5200G 的体系架构及性能

MA5200G 的无源背板交换容量为 256Gbps；无阻塞交换网： $8 \times 8/4 \times 4$ 交叉矩阵；接口和通道容量 8Gbps；交换容量 64Gbps。它的全分布业务板为高性能 CPU+NP 和业务处理+ASIC 硬件转发，且每单板转发性能达 6Mpps。5200G 的体系架构如图 (3-4) 所示：

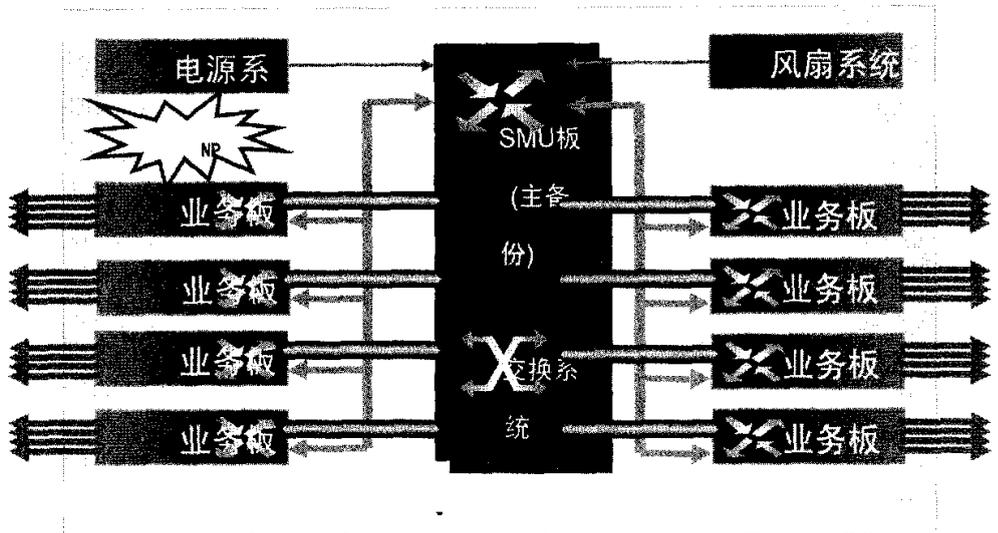


图 (3-4) 5200G 的体系架构

(1)、MA5200G 强大的用户接入能力:

- 并发用户数: 每单板 8K, 整机 48K;
- VLAN: 每端口 4K VLAN, 单板 64K, 整机 512K ;
- QinQ 每端口 64K VLAN ;
- PVC: 每端口 64K PVC, 单板 128K, 整机 512K;
- ISP 数 (用户和业务群分类): 1K ;
- IP 地址: 4K 地址池, 整机 96K 地址;
- VPDN 业务: 支持 LAC (接口板直接提供), 整机 16K 隧道、48K 会话;
- 支持 LNS (专门隧道板), 每块 8K 隧道、8K 会话;
- 设备容量 (单板容量): 槽位数/业务槽位: 10/8;
- 100Base-Tx/Fx 口 16; GE 口 4;
- 带宽 CAR 控制: 范围 8Kbps~1Gbps, 粒度 8Kbps, 误差<3% 。

(2)、MA5200G 具有边缘路由器功能:

- 转发能力: 端口线速转发, 整机 48Mpps;
- 路由协议: 静态、RIP、OSPF、IS-IS、BGP, 512K 路由表;
- 组播协议: IGMP、PIM-SM/DM、MSDP、MBGP, 组播路由表 8K, 每端口可复制 8K 组播流;
- 支持 MPLS L2 和 L3 VPN: 支持 LDP 协议, 10K LSP 隧道;
- 全分布的 MPLS VPN: 接口板直接提供, 1K 个 VRF, 每个 VRF 64K 路由;
- 将 PPP 拨号用户映射进 MPLS VPN; GRE 隧道: 每板 1K GRE 隧道;
- ACL 规则: 每板 64K, 整机 512K 条, 并支持 EAACL 和 UCL;
- 支持 IPTN 组网和 SSG: 基于 COPS+ 协议, 接受承载控制层的业务转发 QoS 策略;
- 对业务流识别分类, 按策略优先级控制, 多层标签转发;
- 业务电信级 QoS 保证: 5 级 QoS 调度、64K 用户优先级队列。

(3)、安全保证功能:

- 用户终端安全保证: 终端之间由 VLAN/PVC 二层隔离;

- 防止帐号被盗用；
- 防止用户 IP 地址被仿冒；
- 防止用户非法访问网络服务；
- 防止用户终端被非法访问；
- 终端之间的互访可以控制。
- 网络资源保护：防用户私接，防资费流失；
- 防用户恶意申请 IP 地址；
- 防止恶意 PPP 拨号呼叫；
- 防止不合理侵占带宽；
- 对访问网络设备的流量/速度控制；
- 对重要业务服务器屏蔽/保护；
- 防止 ARP 广播风暴的“冲击波”病毒；
- 网络自免疫：对终端健康状况检测；
- 不健康终端被隔离；为不健康终端提供杀毒、打补丁等服务；
- 攻击追查：端口镜像；流量采样；用户日志。

(4)、业务控制功能：

- 业务分类；识别来自不同终端的业务；
- 识别来自同一终端的不同业务；
- 对接入的各类业务进行分类；
- 业务区分服务：对不同业务流实施不同的带宽、时延、抖动、丢包率的服务；
- 按业务实施拥塞管理、拥塞避免、Diff_Serv 标识；
- 对业务流按照 TR059 的 QoS 模型进行多级队列调度，保证业务在接入网中的 QoS；
- 业务分流：按策略分流到相应的业务网络；
- 按策略分流到相应的 MPLS VPN 业务平面；
- 多 ISP 业务批发：将属于不同 ISP 的业务通过隧道技术批发给相应的 ISP；对不同 ISP 的业务，由隧道实施区分 QoS 服务。

(5)、MA5200G 产品组播能力：

- 协议支持能力：支持 IGMPv1/v2、PIM-DM、PIM-SM、MBGP、

MSDP、IGMP Proxy 等各类组播协议；

- 对 IGMP 协议，用户可以通过 IGMP over VLAN 协议、也可以通过 IGMP 的 PPPoE 封装来加入/离开某个组播组，也就是说 MA5200G 支持双协议栈的 STB 的频道点播与切换。
- 组播路由表：支持 8K 个组播组，也就是最大可以支持 8K 个频道。
- 性能：转发性能：组播业务在每个端口的转发性能和单播报文的转发性能相同，GE 端口可以达到线速转发。组播转发时延低，平均时延在 20us-70us。同时点播用户数：每秒可支持 500 个用户的频道点播和频道切换；每个端口/每块业务板可以支持 8K 用户同时点播 BTV 业务频道，在 GE 端口上达到线速转发。
- 复制能力：面向端口复制：组播数据报文在每个端口上可以以组播 MVLAN 来复制，由 DSLAM 或者以太网交换机面向用户复制。面向用户复制：可以支持面向用户的组播复制，在 5200G 上通过 PPPoX Session/VLAN/PVC 来复制组播，可以直接面向用户复制 BTV 业务，每端口可以同时面向 8K 用户复制，在 GE 端口线速复制。复制的可控性：可以控制每个用户接收组播的权限。
- 组播业务的安全保证：可以对组播源进行控制，控制某个频道只能来自某个指定的源地址，对来自其它源地址的组播报文一律过滤。对组播接收端口，可以通过 RPF 反向路径检查，过滤来自非法路径的组播报文。防止用户的点播攻击，对每块接口板的 IPTV 点播报文中送 CPU 的速度进行控制。对来自某个 VLAN/PVC 的所有上送 CPU 的控制报文（包括 PPP 拨号握手、ARP 报文、IGMP 频道切换申请报文）的速度进行控制。上送的流量速度可以 1K-1Gbps 之间任意设置。

2、爱立信 ERX-1410/1440 的体系架构及性能

ERX-1410/1440 具有较大的交换能力及宽带用户接入能力，具有较好的 QoS 处理能力，端口种类丰富，支持 L2TP（其中 L2TP LNS 需要另加模块）、Q in Q 和 MPLS VPN、VR 功能。

ERX 1440 采用 40G 交换矩阵结构，14 槽：2 个槽插入 SRP，其余 12 槽插入业务卡。ERX 1410 采用 10G 交换矩阵结构，14 槽：2 个槽插入 SRP，其余

12 槽插入业务卡。

ERX-1410/1440 系统结构如图 (3-5) 所示:

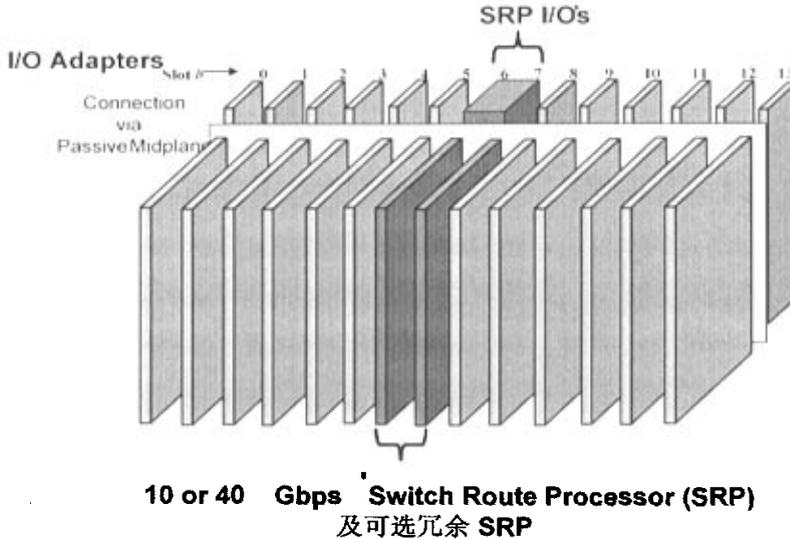


图 (3-5) ERX-1410/1440 系统结构

ERX-1410/1440 使用模块化的运营商级设计, 拥有无源中平面、有源前端插入线路模块和高可靠性后端插入输入/输出(I/O)模块。系统体系结构包括 3 个主要组件: 以 10 或 40 Gbps 运行的共享交换结构; 分布到每个线路模块的硬件辅助转发引擎; 用于路由表维护和系统配置的高性能路由处理器。

系统使用高度分布的多处理器体系结构, 可向系统中的每个端口分布处理能力, 加速决策并扩大网络增长。因此, ERX-1410/1440 可支持下一代路由特性; 这些特性可提供差分服务、线速数据包过滤、缓冲器管理和调度功能。

线路模块负责数据包处理和转发。交换结构执行高速内部数据包交换。路由处理器收集路由表信息, 并向线路模块发送路由表和更新。

ERX-1410/1440 的创新设计将路由处理器从转发路径中移出并在每个端口上维护整个路由表, 为 IP 流量提供一条快速路径。交换路由处理器(SRP)模块上的路由处理器使用通过交换结构的专用组播信道来更新转发引擎上的路由表并只在发生变化时发布更新。对于每个数据包, 系统使用自己强大的硬件以线速执行全表检索。这种设计使系统可以避免缓存检索的不确定性能和数据包延迟相关的不确定性。

ERX 1440 性能指标:

- ◆ 整机交换容量：40Gbps；整机最大并发 PPP session 数：48K。
- ◆ 设备容量(单板数量)：槽位数/业务槽位：14/12；100Base-Fx 口 8;GE 口 2。
- ◆ VPN 功能：支持 L2TP(LAC/LNS);32K 用户 (L2TP)、Ipsec、二层/三层 MPLS VPN、VR。
- ◆ 可靠性：支持电源冗余、主控及交换模块冗余。

3.5 南京 IP 城域网边缘接入层的优化改造

采用支持 ADSL2+技术的 DSLAM 设备，并把用户边缘接入层设备的上连带宽由百兆升级为千兆。必须支持每用户多个 PVC 及其到不同 VLAN 的一一绑定、802.1P 和跨 VLAN 组播复制功能。交换机必须实现基于 VLAN 的 QoS 控制功能。

3.6 增值业务网络优化

增值业务网络结构优化如图 (3-6)：

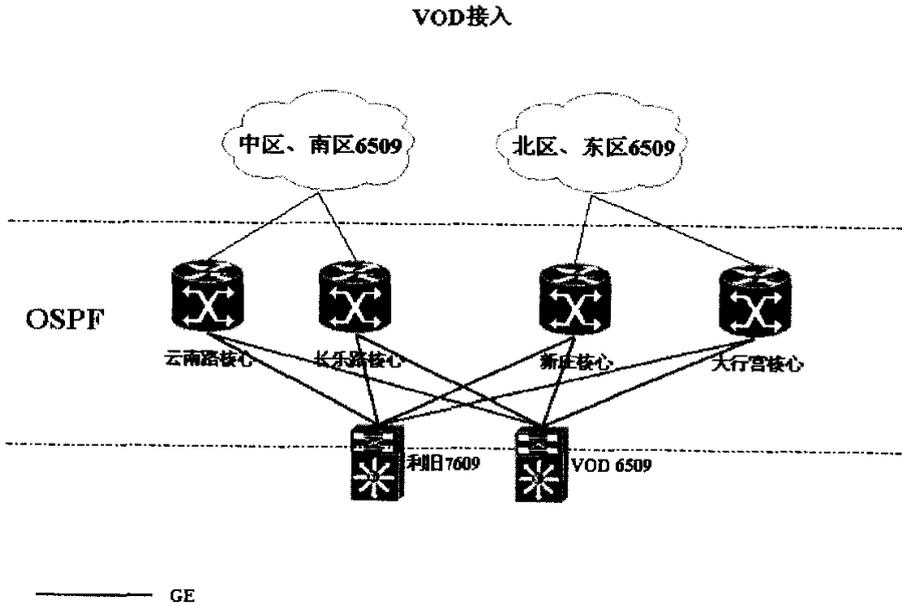


图 (3-6) 增值业务网络结构优化

增值业务部所有的 VOD6509 原来有 3 条 GE 接省网、4 条 GE 接苜蓿园 LSR。本工程改造为：以 GE 分别直接连接 4 台 CISCO 12416。待系统运行正常后，拆除原有 VOD6509 与苜蓿园 LSR 的连接。利旧 7609 提供冗余和负荷分担。

3.7 总结

为了接应中国电信战略转型要求，积极稳妥推进网络平滑演进，将城域网打造成“用户可识别、业务可区分、质量可控制、网络可管理”的电信级网络，南京电信对 IP 城域网进行了全面的优化改造。改造后的南京 IP 城域网成为了一张结构清晰、功能齐全、控制集中，能够满足语音、视频、数据业务承载要求，具备差异化服务能力的综合业务承载网络。

第四章 南京城域网 QoS 和组播部署策略

为了将南京 IP 城域网建设为一个能提供全面业务支撑能力、能满足不同客户要求、具备优良扩展性和稳定性、全程业务与设备可管理可分析的综合业务支撑平台。我们根据集团公司和江苏省公司的要求对城域网现有设备进行了组播及 QoS 支持能力评估与测试。在前期评估测试的基础上, 针对南京 IP 城域网的重点问题, 结合我们对集团公司“城域网优化”的理解, 我们提出在南京城域网上进行组播及 QoS 部署。

4.1 QoS 部署策略

IP 城域网是 CN2 和 ChinaNet 在本地的延伸, 将建成具有 QoS 保证的、可运营、可维护、可管理的多业务承载网络, 满足 CN2 和 ChinaNet 的业务在本地延伸要求。

南京城域网将引入以 DiffServ 为主的 QoS 技术, 以 COPS 协议为主的业务管理技术, 加强对城域网设备的技术评估和选型, 使南京城域网最终成为具有支持多业务能力的承载平台。

4.1.1 QoS 概述

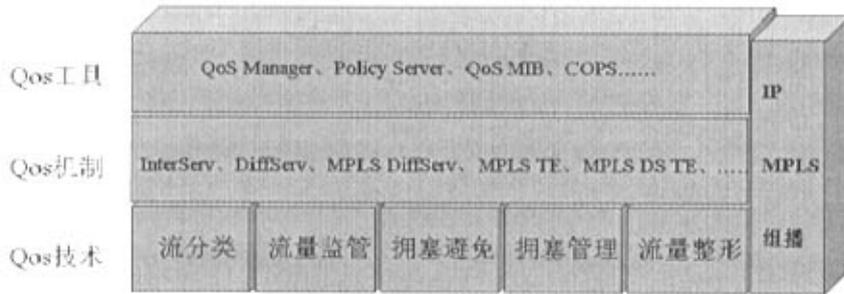
IP QoS (Quality of Service) 是指 IP 网络的一种能力, 是为特定的业务提供其所需要的服务。衡量 IP QoS 的技术指标包括:

- 1、带宽/吞吐量——指网络的两个节点之间特定应用业务流的平均速率;
- 2、时延 - 指数据包在网络的两个节点之间传送的平均往返时间;
- 3、抖动 - 指时延的变化;
- 4、丢包率 - 指在网络传输过程中丢失报文的百分比, 用来衡量网络正确转发用户数据的能力;
- 5、可用性 - 指网络可以为用户提供服务的时间的百分比。

不同的宽带用户及业务对 IP QoS 技术指标的要求是不同的, 通过有效地实施各项 IP QoS 技术, 使得我们能够有效地控制网络资源及其使用, 并有可能在 IP 网络上开展语音、视频及数据等多种业务, 能够在现有网络上细分客户、针

对不同的客户需求提供特色的差别业务、以便能迅速获得利益回报、从而进一步扩大市场占有率、提高市场竞争力。

IP 网络的 QoS 基本要素包括 QoS 基本技术、QoS 机制以及 QoS 工具，如图（4-1）所示。



图（4-1）IP 网络的 QoS 基本要素

QoS 技术是前提条件，主要是网络设备关于 QoS 的实现技术；QoS 机制是端到端实现的方法，主要是在设备实现 QoS 基本技术基础上，从整个网络的角度考虑端到端的 QoS 实现方法；QoS 工具，主要用于简化 QoS 部署难度而开发工具，可以大大简化人为的工作。

4.1.2 QoS 部署总体规划

在 IP 网络的环境中，我们在实施 QoS 方案时，应该首先考虑以下几个问题：

1. 应该为我的宽带客户提供什么样的业务水平
2. 如何提供端到端的 IP QoS、如何经济高效地管理 IP QoS
3. 如何充分利用现有的 IP 基础设施
4. 如何将 IP QoS 与虚拟专用网业务结合

针对第一个问题，中国电信集团公司已经做了总体的规划，可以作为南京电信在部署时的依据。对于后面的几个问题，在本次核心和汇聚层的部署中只是进行一个初步的实施，还需要我们在实际的实施过程中进行不断的反馈和修正。因此，QoS 的部署是一个长期调优的过程，这是由其自身的特性所决定的。

表 1 和表 2 是集团公司对 QoS 部署的总体规划，也是我们在本次南京 IP 城域网内部署 QoS 的基本依据。

表 1: 关键业务性能要求

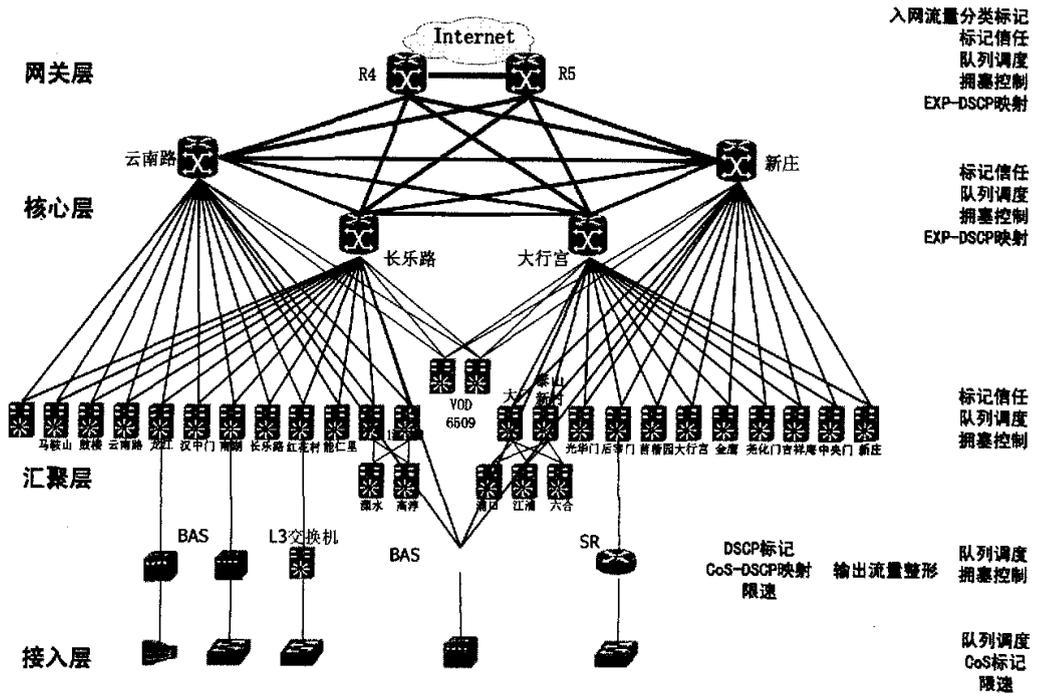
应用类型	典型业务	端到端单向平均时 延上限	端到端单向平均 抖动上限	端到端单向丢包 率上限	端到端单向包 误差率上限
即时语音	软交换语音	100ms	20ms	0.1%	0.01%
即时视频(视频 电话、视频会 议)	视频电话、视频 会议	100ms	20ms	0.1%	0.01%
流媒体视频	IPTV	1000ms	U	0.1%	0.01%
即时交互数据	游戏、信令	100ms	U	0.1%	0.01%
一般数据		U	U	U	U

表 2: 关键业务分类表

服务 等级	业务性质	业务	DSCP	IP Pre/EXP	与骨干网 衔接	队列类型
一级	关键业务	NGN 语音、管理控制信 息	111000	111	对应 CN2 等级 1 (EXP110)	Low latency Queue
二级	CN2 金接入	新视通、全球眼、IPTV 等视讯业务、跨越城域网 和 CN2 的高等级 VPN	110000	110	对应 CN2 等级 2 (EXP100)	Round-robin Queue
三级	同城互连 金业务	高 QoS 的同城互连	101000	101	城域网专用	
四级	CN2 银接入	Vnet 高等级业务、跨越城 域网和 CN2 的中等级 VPN	100000	100	对应 CN2 等级 3(EXP010)	Round-robin Queue
五级	CN2 铜接入	Vnet 低等级业务、跨越城 域网和 CN2 的低等级 VPN	011000	011	对应 CN2 等级 4(EXP001)	
六级	CN2 普通接入	城域网个人 VIP 业务	010000	010	对应 CN2 等级 5 (EXP000)	
七级	同城互连 银业务	低QoS的同城互连	001000	001	城域网专用	
八级	普通业务	普通互联网业务	000000	000	ChinaNet 尽力转发	BE Queue

4.1.3 QoS 具体部署方案

QoS 的部署主要包括数据包的分类和标记、流量的拥塞管理、流量的拥塞避免、流量监管和流量整形这几个方面。本期的部署方案也是从这几个方面入手进行。网络中各层次 QoS 部署策略如图 4-2 所示：



如图 4-2 网络中各层次 QoS 部署策略

1、数据包的分类和标记

数据包分类是将数据报文划分为多个优先级或多个服务类，如使用 IP 报文头的 TOS 字段（Type of service）的前三位（即优先权子字段），可以将报文最多分成八类。在报文分类后，就可以将其它的 QoS 特性应用到不同的分类，如拥塞管理、带宽分配等。

报文分类的依据是五元组（源地址、源端口号、协议号码、目的地址、目的端口号）

一般在网络的边界，对报文进行分类，同时设置报文的 IP 头的 TOS 字段的

前三位作为报文的 IP 优先级，这样，在网络的内部就可以简单的使用 IP 优先级作为分类的标准。而队列技术如 WFQ 就可以使用这个优先级来对报文进行不同的处理。下游（downstream）网络可以选择接收上游（upstream）网络的分类结果，也可以按照自己的分类标准重新进行分类。

所有业务的数据分类标记均在接入层或业务控制层完成，尽量使用接入层设备进行数据标记，如果接入层设备不能支持标记功能则在业务控制层完成。

IPTV 业务数据流是单向数据流，流向是从服务器至用户终端，因此只在服务器接入点（VOD6509）根据源 IP 地址进行标记。

其它业务需要在服务器接入点和用户接入点同时标记（没有服务器的 P2P 型业务除外）。

QoS 标记在城域网 QoS 域内传递时，相关设备采取信任的原则。

在接入层部署时主要就采用在设备对业务类型进行分类标记的方法，具体的部署为：

- 1、尽可能靠近源实现标记，后续设备对其标记予以信任
- 2、接入层网络：所有城域网内产生的流量尽量在接入层网络使用“物理端口/802.1p/PVC”等特征实现分类标记（包括 IDC 的接入层网络）
- 3、业务接入控制点：如不能在接入层网络实现分类标记，则在业务接入控制点上使用五元组实现分类标记。
- 4、核心汇聚层不需要标记。
- 5、出口路由器 R4/R5：对 ChinaNet 进入城域网的流量，其 DSCP 标记不予信任，需使用五元组进行重新分类标记；但对 MPLS 帧的 EXP 标记信任并作映射。

2、流量的拥塞管理

（1）排队和策略技术的选择

当业务始终能够以线速通过网络时，任何 QoS 保证机制都是多余的，因为这时所有的业务都可以得到最佳的保证。但是在真实网络环境中，优良的交换性能并不能保证业务以线速通过，复杂的网络业务量模型、动态的流量分布都会造成网络频繁地瞬间拥塞。拥塞一般分为 3 个等级：轻度拥塞、严重拥塞和完全拥塞。在网络利用率达到一定程度时会产生轻度拥塞，这种状态是网络的最佳状态，

因为这时网络在保持服务质量的同时, 承载的业务量也达到最大。所谓严重拥塞就是网络业务量已经达到或超出网络所能承受的范围; 完全拥塞是指业务量完全阻塞了网络, 网络已经无法正常工作。

对于拥塞的管理主要有以下几种方法:

首先是排队。出现拥塞意味着有数据需要缓存, 排队就是一种缓存数据的方法。排队是根据一定的策略将业务分别缓存, 这是一切 QoS 或 CoS 处理的基础和前提。

然后是根据队列优先级决定发送数据的策略, 如 WRR (Weighted Round Robin) 或 MDRR (Modified Deficit Round Robin)。通过为高优先级业务提供更多的转发机会可以保证轻度拥塞时业务的服务质量。

在出现严重拥塞时, 网络必须丢弃一定的数据以缓解拥塞, 避免出现进一步拥塞。丢弃数据也可以根据优先级, 对于丢失敏感型业务可以进行一定程度的保护, 这种实现机制就是 WRED。

对于南京电信的网络情况, 根据目前的监控情况, DSLAM-BRAS-6509 (或 GSR) 之间的峰值带宽利用率 < 50%, 4 个 GSR 之间的互联链路的峰值带宽利用率 < 10%, 4 个 GSR-R4、R5 之间的链路的峰值带宽利用率 > 50%, R4、R5-省网路由器的链路的峰值带宽利用率 > 50%。从以上数据可以得出结论: DSLAM-BRAS-核心汇聚层的链路属于轻载状态, 核心汇聚层-R4、R5 的链路属于轻度拥塞状态, 以上结论不能排除有瞬间严重拥塞的可能, 因为目前的监控手段和采样周期不能保证实时的监控到网络的真实流量情况。

本期的 QoS 进行部署, 也就是为了保证在链路可能出现拥塞时能够确保高等级业务的服务质量, 其核心技术就是围绕着排队和队列处理技术来进行。

拥塞管理技术对比:

	优 点	缺 点
FIFO	1、缺省队列机制, 队列长度配置简单, 易于使用。 2、处理简单, 处理延迟小。	1、所有报文同等对待, 报文到来的次序决定了报文可占用的带宽、报文的延迟、报文的丢失。 2、对不配合的数据源 (如

		<p>UDP报文发送)无约束力,不配合的数据源会造成配合的数据源(如TCP报文发送)带宽受损失。</p> <p>3、对时间敏感的实时应用(如VoIP)的延迟得不到保证。</p>
PQ	<p>可对不同业务数据提供绝对的优先,对时间敏感的实时应用(如VoIP)的延迟可以得到保证。对优先业务的报文的带宽占用可以绝对优先。</p>	<p>1、需配置,处理速度慢。</p> <p>2、如果不对高优先级的报文的带宽加限制,会造成低优先级的报文得不到带宽。</p>
CQ	<p>1、可对不同业务的报文按带宽比例分配带宽。</p> <p>2、当没有某些类别的报文时,能自动增加现存类别的报文可占的带宽。</p>	<p>1、需配置,处理速度慢。</p> <p>2、不适于解决对时延敏感的实时业务。</p>
WFQ	<p>1、配置简单,报文分类自动完成。</p> <p>2、可以保护配合(交互)的数据源(如TCP报文发送)的带宽。</p> <p>3、可以使延迟的抖动减小。</p> <p>4、可以减小数据量小的交互式应用的延迟。</p> <p>5、可以为不同优先级的流分配不同的带宽。</p> <p>6、当流的数目减少时,能自动增加现存流可占的带宽。</p>	<p>1、处理速度比FIFO要慢。但比PQ、CQ要快。</p> <p>2、不适于解决对时延敏感的实时业务。</p>
CBWFQ	<p>1、对报文进行分类,为每类报文提供确保带宽。</p>	<p>系统开销比较大。</p>

	<p>2、可为部分报文提供快速转发服务 (Expedited Forwarding, EF), 使延迟降低为最小。</p> <p>3、各类数据的流确保带宽的总和小于接口的带宽时, 能自动增加各类流的带宽, 从而充分利用线路的带宽。</p> <p>4、为缺省类的报文提供WFQ服务。</p> <p>5、可以为非优先类的报文提供WRED的丢弃策略。</p>	
--	--	--

针对南京城域网的具体情况, 不同设备对以上各种排队技术的支持程度不同, 根据我们的综合比较, 结合集团公司的总体规划, 我们的部署方案是排队采用 PQ+WFQ 的方式, 具体的队列是 1P3Q。转发是采用 MDRR/WRR 的策略, MDRR 是高端路由器所独有的调度机制, 能够避免 WRR 会出现的被大包传输造成的关键数据的延迟, 目前可以在 GSR 上部署, 其它设备 (6509 和 BRAS) 采用 WRR 机制。

对于接入层网络, 目前基本上支持两个队列, 因此就部署两个队列的方式, 将优先等级高的业务放入高等级的队列中传输。

具体的部署为:

LLQ 策略

1. 网关层和核心层 Cisco12416: 由 MDRR 实现 1 个 LLQ, 使用 Alternate 模式, 带宽权重设为 1, 保证 10%的带宽
2. 汇聚层 C6509 和业务接入点各种 BAS/SR/L3 交换机: 由 PQ 实现
3. 接入层网络: LS2024 可使用 PQ 实现, 其它设备由 WRR 实现, 分配给该 WRR 较大带宽 (如 50%)
4. 为保证 LLQ 不过分占用带宽, 应在流量源头进行限速

WRR 策略

1. 网关层和核心层 Cisco12416: 由 MDRR 实现, 三个 WRR 队列的带宽权重分别设为 9, 6, 12, 相应的带宽比例分配分别为 30%, 20% 和 40%
2. 汇聚层 C6509 和业务接入点各种 BAS/SR/L3 交换机: 由 WRR 实现, 带宽比例分配分别为 30%, 60% (目前 6509 的 GE 口仅支持 1P2Q)

3. 接入层网络：大部分设备只有两个队列，其中一个用于 LLQ，剩下的这一个用于其余流量，占用剩下的带宽

(2) 队列分配的部署

关键业务（NGN 语音、管理控制信息）的 DSCP 值设为 56，这类业务属于语音类业务，对时延和抖动比较要求较高但流量较小，使用 LLQ 队列。

新视通、全球眼、IPTV 等视讯业务、跨越城域网和 CN2 的高等级 VPN 业务的 DSCP 值设为 48，高 QOS 的同城互连业务的 DSCP 值设为 40，这两类业务使用同一个 MDRR 队列。根据带宽估算的结果，提供 30%的保证带宽。

Vnet 业务、跨越城域网和 CN2 的中低等级 VPN 业务、城域网个人 VIP 业务、低 QOS 同城互连业务（DSCP：8，16，24，32）使用同一个 MDRR 队列。根据带宽估算的结果，提供 20%的保证带宽。

普通 BE 业务 DSCP 值设为 0，使用同一个队列。提供 40%的保证带宽

(3) GSR12416 部署策略：

由于 GSR 支持 MDRR，所以它的 LLQ 不同于其它设备的 LLQ，能够支持 PQ 功能，并支持两种模式的 PQ（Alternate/Strict）。

在 Strict 模式下，LLQ 中只要有数据需要服务，总会得到服务，只有 LLQ 处于空闲状态时，其它队列才能得到服务。这种模式能够最大限度减小数据的时延和抖动，但是可能会影响其它队列数据的服务。

在 Alternate 模式下，LLQ 与其它普通队列根据权重进行轮询服务，但是 LLQ 具有优先服务功能：如果所有队列都有数据等待服务，轮询的顺序为 LLQ-Q0-LLQ-Q1-LLQ-Q2-LLQ-Q3.....。这种模式虽然在拥塞状态下相对于 Strict 模式来说时延和抖动稍大，但是不会完全占用其它队列的带宽，且 LLQ 的时延和抖动相对于其它队列来说仍然是最小的。

综合上述两种模式的特点，使用 Alternate 模式。

MDRR 队列中唯一可修改的参数是每个队列的带宽权重（weight），在一次轮询中每个队列得到服务的数据量为 $MTU * (weight-1) * 512byte$ ，也就是说 weight 值会影响队列轮询的时间，从而会影响 LLQ 数据转发的时延和抖动，因此将保证带宽最小的队列的 weight 设为最小值 1。

在 Alternate 模式下，LLQ 得到的保证带宽比例为 $n * WL / (n * WL + Sum(W0, Wn))$ 。WL 为 LLQ 的带宽权重；n 为其它队列数量；W0, W1...Wn

为其它队列的权重。根据上述公式在 WL 为 1 的情况下,其它三个队列的带宽权重应该设为 9, 6, 12, 从而能够使得带宽保证符合上述值。

(4) Catalyst6509 部署策略:

南京 BIP 网络中 Catalyst 6509 主要使用的板卡是 WS-X6516-GBIC, 该板卡只能支持 1P2Q2T, 不能完全满足队列需求, 因此将普通 BE 业务与 Vnet 业务, 跨越城域网和 CN2 的中低等级 VPN 业务, 城域网个人 VIP 业务, 低 QOS 同城互连业务 (DSCP: 8, 16, 24, 32) 使用同一个队列 (Queue1), 并保证 60% 的带宽, Queue2 保证 30% 的带宽, Queue 为 PQ 队列。

由于 Catalyst 6509 的 PQ 是绝对优先队列, 对带宽的占用没有限制, 因此对其进行限速, 通过 CAR 保证 PQ 占用 10% 的带宽。(目前 6509 由于使用是 Sup2, 只能支持 INPUT 方向的限速功能)

两个 WRR 队列通过设置带宽权重 (weight) 来分配带宽, weight1 和 weight2 分别设为 2 和 1 (PQ 队列不参与计算, 必须通过限速来限制其对带宽的占用)。

(5) BRAS/SR 部署策略:

南京电信目前使用的 BRAS 主要有 ERX1440、MA5200G。使用的 SR 主要有 Quidway3026、Catalyst3550, 还有相当数量的 2948。但是由于 2948 的型号太老, cisco 已经停产, 且 2948 本身不支持 COS, 因此在此次测试中未将 2948 纳入测试范围, 对 SR 的测试仅针对 Quidway3026 和 Catalyst3550。

ERX 可为每个用户提供独立的 8 个硬件队列, 包括一个 Strict Priority Queue。在队列调度算法上提供 WRR 算法。

MA5200G 为每个端口定义了六个优先级队列, 分别是 BE、AF1 ~ AF4、EF, 根据上行流量分类出的服务等级, 入相应队列, 保证输出带宽 (流量整形)。

Quidway3026 和 Catalyst3550 每个端口均支持 4 个队列, 其中一个可以设为 Strict Priority Queue, 在队列调度算法上提供 WRR 算法。

因此, BRAS/SR 的部署策略完全可以满足省公司的规划, 具体如下:

ERX1410/1440/Quidway3026/Catalyst3550 将关键业务 (NGN 语音、管理控制信息) 分配给 Strict Priority Queue, 并通过 CAR 保证 PQ 只占用 10% 的带宽。其它三类业务分配到不同的三个队列, 采用 WRR 的调度算法, 带宽权值设为 3: 2: 4。

MA5200G 将关键业务定义为 EF 业务, 带宽权值设为 1, 普通互联网业务定

义为 BE 业务，带宽权值设为 4，其它两类业务分别定义为 AF1~AF2 业务，带宽权值设为 3 和 2。

(6) 接入层设备部署策略:

由于接入层网络大部分设备只有两个队列，不能完全满足总体规划的要求，因此在部署中将其中一个用于设备为 PQ，将前三级业务分配到 PQ 中，剩下的业务使用另一个队列。

(7) 带宽分配的选择依据

南京电信目前和将要开展的业务主要有 IPTV、新视通、全球眼、同城专线互联、VPN 和公众上网等。

1、IPTV 带宽需求可以按照下列公式来计算

TV 节目*频道带宽 + 最大并发 VoD*节目带宽

根据 IPTV 业务发展的计划，最多开展 20 套组播节目，每套节目 2M 带宽，约需要 40M 带宽。

VoD 点播方面计划发展 5000 个用户，按照 1:3 的收敛比计算，同时在线约 1700 个用户，如果每个用户都点播 VoD，总共需要占用 3.4G 的带宽，目前增值业务部 VOD6509 同时连接到 BIP 网络的四台 GSR 上（云南路、长乐路、新庄、大行宫），每台 GSR 有两条 GE 线路，共 8G 带宽，考虑到正常情况下用户分布应该是相对平均的，因此每条 GE 最多将会有 425M 的 VoD 点播流量。

四台 GSR 总共与 23 台 6509 相连，每台 6509 有四条 GE 线路，因此按用户平均分布来计算，每条 GE 最多将会有 36M 的 VoD 点播流量。

综上所述，由于 IPTV 业务流量是单向的下行流量，GSR 需要保证下行 GE 接口提供至少 76M 的带宽。

2、新视通业务流量主要存在于 MCU 与客户终端之间，终端速率有 384K，512K，768K 三种，协议速率*1.2 即为每个终端与 MCU 之间通信所需要带宽，最大约为 1M，目前南京城域网的新视通用户基本以 LAN 的形式接入，接入带宽基本上为 10M。。

3、全球眼业务保证带宽为 512K，目前用户很少，因此需要的带宽很小。

4、高 QOS 同城互连业务目前主要是 28 家大用户，由于不是集中式业务且数量较少，因此正常情况下所需要的带宽比例很小。

其余带宽可供其它业务使用。

3、 流量的拥塞避免

当拥塞发生时，按照传统的处理方法，所有到来的报文都被丢弃。对于 TCP 报文，如果大量的报文被丢弃，将造成 TCP 超时，从而引发 TCP 的慢启动和拥塞避免机制，使 TCP 减少报文的发送。当队列同时丢弃多个 TCP 连接的报文时，将造成多个 TCP 连接同时进入慢启动和拥塞避免，称之为：TCP 全局同步。这样多个 TCP 连接发向队列的报文将同时减少，使得发向队列的报文的量不及线路发送的速度，减少了线路带宽的利用。并且，发向队列的报文的流量总是忽大忽小，使线路的上的流量总在极少和饱满之间波动。

为了避免这种情况的发生，队列可以采用加权随机早期检测 WRED (Weighted Random Early Detection) 的报文丢弃策略 (WRED 与 RED 的区别在于前者引入 IP 优先权, DSCP 值, 和 MPLS EXP 来区别丢弃策略)。采用 WRED 时，用户可以设定队列的阈值 (threshold)。当队列的长度小于低阈值时，不丢弃报文；当队列的长度在低阈值和高阈值之间时，WRED 开始随机丢弃报文 (队列的长度越长，丢弃的概率越高)；当队列的长度大于高阈值时，丢弃所有的报文。

WRED 可以感知 QoS 的带内信令，包括 IP 优先级、DSCP，和 MPLS EXP。可以为不同 IP 优先级，DSCP，或 MPLS EXP 的报文设定不同的队列长度滤波系数、队列阈值、丢弃概率，从而对不同优先级的报文提供不同的丢弃特性。

加权随机早期检测 (WRED) 参数选择

一般情况下，WRED 的参数都有缺省值，而且缺省值均是经过测试的性能最佳的值，因此一般不需要重设置，可以根据实际情况做一些修改，下面是一些设备的 WRED 参数配置：

1、GSR12416:

WRED 可配置参数共有以下几个：

(1)、exponential-weighting-constant

该参数的作用是计算队列的平均长度，用来与 minimum-threshold、maximum-threshold 比较，确定是否需要丢包。具体公式为：

$$\text{average} = (\text{old_average} * (1-1/2^n)) + (\text{current_queue_size} * 1/2^n)$$

这里 n 表示 exponential-weighting-constant 值，从上述公式中可以看出： n 取较大的值时，average 越接近 old_average，受 current_queue_size 的影响较小，其变化比较平滑； n 较小时，average 受 current_queue_size 的影响较大，其变化比较灵敏。

因此针对变化较大的流量，exponential-weighting-constant 应取较小的值，而针对变化较小只是偶然突发的流量应取较大的值，以避免突发流量引起不必要的丢包。

Cisco 建议 exponential-weighting-constant 按以下方式取值：

$$\text{exponential-weighting-constant} = 10/B$$

$$B = 1000\text{Mbps}/(8\text{bits/byte})/1500(\text{bytes/packet}) = 83335$$

其中 1000Mbps 表示 GE 接口带宽，1500 表示 GE 接口 MTU，B 表示 GE 接口每秒能转发的数据包最大数量（MTU 不是 1500 的接口也按照 1500 来计算）。

根据以上公式 GE 接口的 exponential-weighting-constant 约为 $1/10^{13}$ ，配置为 13；2.5G POS 接口的配置 exponential-weighting-constant 约为 $1/10^{14}$ ，配置为 14

(2)、minimum-threshold

该参数用来定义 WRED 的最小门限值，建议最小值是 0.03B。根据 B 的计算公式，GE 接口的最小门限值建议最小使用 2500，2.5G POS 建议最小使用 6000。建议在配置该参数时根据 DSCP 级别配置不同的值，级别最低配置上述最小值，级别越高 minimum-threshold 值越大。

(3)、maximum-threshold

该参数用来定义 WRED 的最小门限值，建议最小值是 0.1B。根据 B 的计算公式，GE 接口的最小门限值建议最小使用 8334，2.5G POS 建议最小使用 20000。在配置该参数时根据 DSCP 级别配置不同的值，级别最低配置上述最小值，级别越高 maximum-threshold 值越大。

(4)、mark-probability

该参数用来定义 Queue 长度在最小门限和最大门限之间时的丢包概率，建议

根据 DSCP 值不同, 建议根据 DSCP 级别配置不同的值, 级别最低配置上述最小值, 级别越高 maximum-threshold d 值越大, 8 个等级的数据分别取 1-8。

2、Catalyst6509:

南京 BIP 网络中 Catalyst 6509 主要使用的板卡是 WS-X6516-GBIC, 该板卡只能支持 1P2Q2T, 其中 PQ 队列没有门限, 每个 WRR 队列都各有两组门限值, 每组分别有最小门限 (min-thresholds) 和最大门限 (max-thresholds) 两个值, 同一个队列中的数据可根据不同的优先级选取相应的组从而获得 min-thresholds 和 max-thresholds。

每组门限中只有最大门限 (max-thresholds) 可以配置, 最小门限 (min-thresholds) 根据最大门限值自动配置。

Catalyst6509 的一个 WRR 队列中缺省的第一组门限值为 40 (min-thresholds) 和 70 (max-thresholds), 第二组门限值为 70 (min-thresholds) 和 100 (max-thresholds)。同一个 WRR 队列中的数据可以根据不同的优先级选用不同组的门限值。一般来说缺省值均是经过测试的性能最优的值, 因此不建议对其进行修改, 只需要根据不同的优先级进行选用。

4、流量监管与流量整形

流量监管 (traffic policing) 的典型作用是限制进入某一网络的某一连接的流量与突发。在报文满足一定的条件时, 如某个连接的报文流量过大, 流量监管就可以对该报文采取不同的处理动作, 例如丢弃报文, 或重新设置报文的优先级等。通常的用法是使用 CAR 来限制某类报文的流量, 例如限制 HTTP 报文不能占用超过 50% 的网络带宽。

流量整形 (traffic shaping) 的典型作用是限制流出某一网络的某一连接的流量与突发, 使这类报文以比较均匀的速度向外发送。流量整形通常使用缓冲区和令牌桶来完成, 当报文的发送速度过快时, 首先在缓冲区进行缓存, 在令牌桶的控制下, 再均匀地发送这些被缓冲的报文。

根据南京城域网的情况, 在业务控制点设备 (BRAS/SR) 上部署流量整形和流量监管 (如果接入层设备能够实现限速功能, 则业务控制点设备不需要部署流

量监管)。根据规划中各类业务的带宽分配比例及接口物理带宽大小进行部署。

4.2 组播部署

随着南京电信宽带业务的发展,流媒体信息服务的需求也越来越多,如实时的新闻、娱乐转播、流媒体形式的视频节目等。而已往的 TCP/IP 技术一般都是针对点到点的传递或者广播的方式传递,对于这样的业务需求,传统的技术需要将这些大量的流业务复制多份,分别向不同的用户发送,这样对整个网络带宽的消耗以及对网络设备的压力是可想而知的,无法应付大规模的多媒体业务。部署方法

4.2.1 组播路由协议介绍

目前主要的组播路由协议为 PIM。PIM 协议的发展目的是在 Internet 上提供足够规模的域间组播路由。PIM 协议主要有两种工作模式:

1、密集模式 (PIM-DM)

此模式与 DVMRP 相似,都属于密集模式协议,采用了 DVMRP 一样的方式,建立 PIM-DM 环境中建造起来的基于资源的组播树。PIM-DM 独立于网络选择的 IP 路由协议,"协议无关组播"由此而得名。

协议工作原理如下:路由器收到发自资源的一个组播包,对资源 IP 地址进行检查,进而得知在"反转路径转发"接口是否已收到该组播包。然后,路由器将该组播包流传到除已收到该包的接口外的各接口。

PIM-DM 适用于以下集中情况:

- 1、发送者和接收者非常接近,且只有一小部分发送者和大量的接收者。
- 2、组播的流量大。
- 3、组播是持续的。

2、分散模式 (PIM-SM)

与 PIM-DM 相似,两者都是建立在组播路由协议基础上用于决定 RPF 接口的协议。PIM-SM 协议假定在网络中接收者的人数很稀疏或者组播组被广域网分割开。

PIM-SM 适用于以下集中情况:

- 1、在一个组里接收者很少。
- 2、发送者和接收者被广域网分开。
- 3、组播流量断断续续。

由这两种模式的特点，我们在本期初次部署时，采用 PIM-SM 的方式，随着业务的发展，可改为 PIM-DM 模式。

在本期部署中，采用了大行宫和新庄 GSR 作为组播的 RP，配置 anycast RP，RP 及组播源准入控制。其余设备上配置 PIM-SM 路由协议。

4.2.2 组播具体部署

核心层 GSR 之间及核心层与汇聚层之间的互联电路从 GE PortChannel 改造为 GE 互联。

主备 RP：核心层网络中选取云南路 GSR 和大行宫 GSR 担任主备 RP

核心层设备、汇聚层设备及业务接入控制点设备均开启组播，配置 PIM-SM 业务接入控制点除配置 PIM-SM 外，还可选地配置静态组播组以加快 IPTV 节目切换速度。

业务接入控制点配置组播过滤。

PPPOE/PPPOA 接入时，由 BAS 实现组播流量复制，二层网络设备（L2 交换机和 DSLAM）无需组播配置；路由方式接入时，二层网络设备配置 igmp-snooping。

各层次组播实施策略如图（4-3）所示：

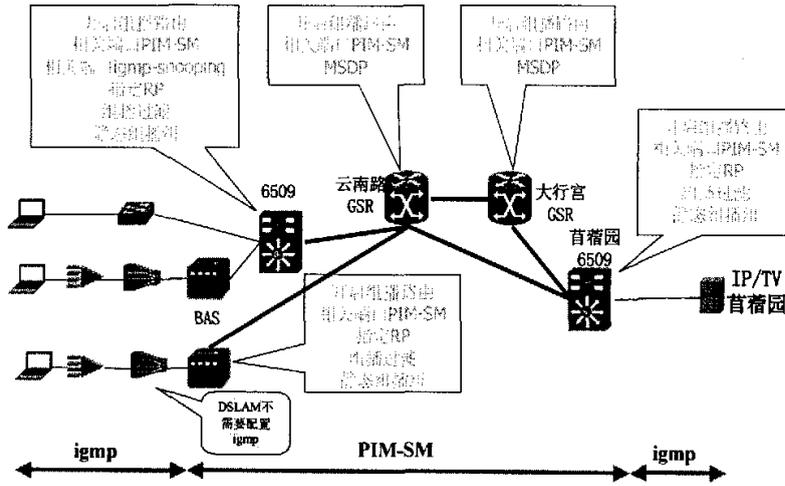


图 (4-3) 各层次组播实施策略

4.3 总结

为了将南京电信城域网建设为一个能提供全面业务支撑能力、能满足不同客户要求、具备优良扩展性和稳定性、全程业务与设备可管理可分析的综合业务支撑平台。我们根据集团公司和江苏省电信公司的要求对城域网现有设备进行了组播及 Qos 支持能力评估与测试。

在前期评估测试的基础上，针对南京电信城域网的重点问题，结合我们对“IP 城域网优化”的理解和多年来的 IP 网络建设经验，我们提出在南京电信城域网上进行组播及 Qos 部署的建议，希望能对南京电信承载网的组播和 Qos 部署提供帮助。

第五章 QoS 和组播实施结果及应用分析

5.1 QoS 部署结论及应用分析

1、改造的范围和改造方法：

(1)、接入层设备

- 1、MA5100 上行板卡需更新为 LAND 板，并升级软件，需补充数量和所需费用。
- 2、2403F 需要逐步退网，但是目前南京城域网上有超过 2000 台，要短时间内全线退网，难以实现。
- 3、MA5200 不支持 QoS，且全网目前有 186 台 MA5200，我们建议将其上的 ADSL 用户逐步迁移至 BRAS 上，用 5200 上的链路轻载来保证视频类业务的带宽，不建议在其上部署 QoS。

(2)、SR和BRAS

ERX1440 和 华为 5200G 不需要额外的硬件和软件升级

(3)、核心汇聚层

需要将目前南京城域网 GSR 和 6509 下联广泛使用 GE PortChannel 改造为多 GE 下联。目前不涉及对设备的硬件和软件升级改造。

2、部署结论：

- 1、大部分设备具有的宣称的 QoS 功能，且性能达到要求。
- 2、开启设备的诸项 QoS 功能未造成设备资源大量占用，设备间 QoS 功能配合未出现异常。
- 3、配置 QoS 手段可保证高等级数据流的带宽、延迟和丢包率满足要求；反之则不能。
- 4、PQ 占有绝对优先发送权，可支持要求保证带宽、低延迟、低丢包率的业务，但滥用将可能造成其它流量的严重阻塞，有必要限速；WRR 可为不同等级的流量分配争用时的带宽占用比例；各队列闲置带宽可被其它队列使用。
- 5、部署所使用的 QoS 手段可有效地支持 IPTV/新视通在内的对 QoS 要求很高的

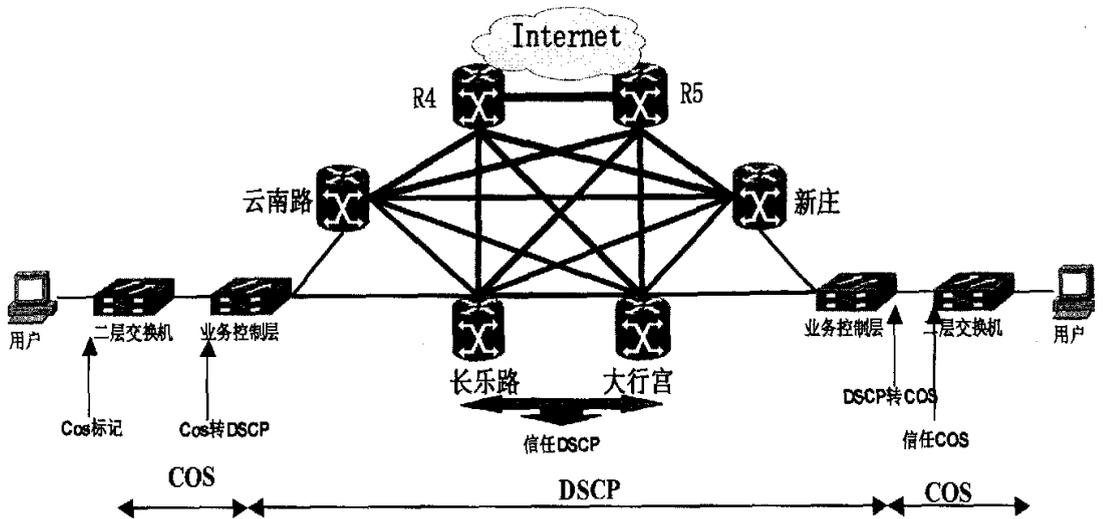
业务。

3、QoS业务典型组网方式

南京电信城域网内目前主要有以下 QoS 业务组网方式：

(1)、LAN-LAN 用户；

当用户数据进通过二层交换机进入网络时，二层交换机根据相应网络服务级别利用 COS 将数据包标记成相应服务级别，交换机根据 COS 级别进行队列调度与拥塞管理，在业务控制层（BAS 或 C6509）网络设备将 COS 转换成 DSCP 服务级别并进行队列调度与拥塞管理，核心层网络设备根据 DSCP 值进行队列调度与拥塞管理并将数据包传输至对端业务控制层，业务控制层将 DSCP 值转换成 COS 并传输给二层交换机，二层交换同根据 COS 进行队列调度与拥塞管理，并将数据传送给用户。网络拓扑如图（5-1）：



图（5-1）LAN-LAN QoS业务组网

(2)、ADSL-LAN 用户；

网络拓扑如图（5-2）：

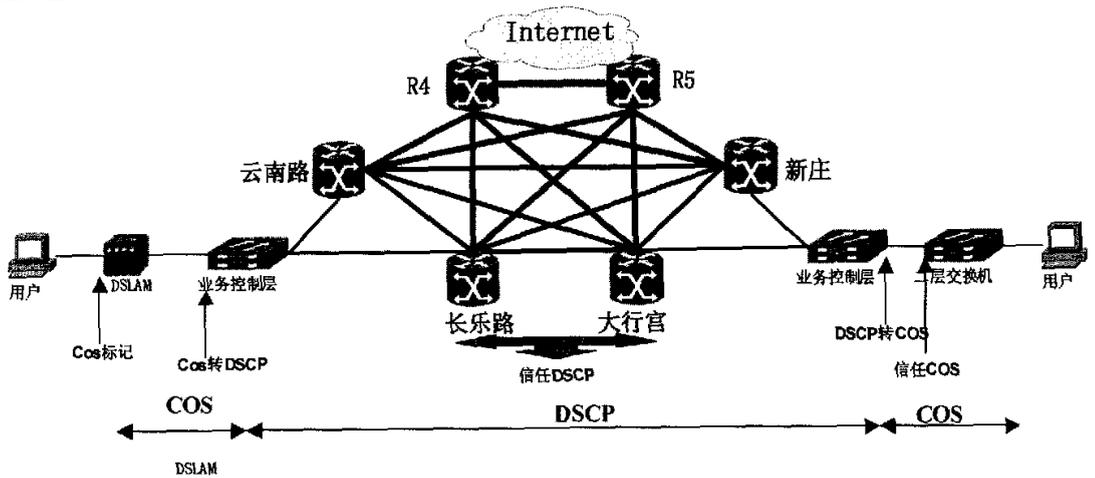


图 (5-2) ADSL-LAN QoS业务组网

在该种组网方式下，当用户数据进通过 DSLAM 进入网络时，DSLAM 根据相应网络服务级别利用 COS 将数据包标记成相应服务级别，DSLAM 根据 COS 级别进行队列调度与拥塞管理，在业务控制层（BAS）网络设备将 COS 转换成 DSCP 服务级别并进行队列调度与拥塞管理，核心层网络设备根据 DSCP 值进行队列调度与拥塞管理并将数据包传输至对端业务控制层，业务控制层将 DSCP 值转换成 COS 并传输给二层交换机，二层交换同根据 COS 进行队列调度与拥塞管理，并将数据传送给用户。

(3)、ADSL-ADSL 用户；

网络拓扑如图 (5-3)：

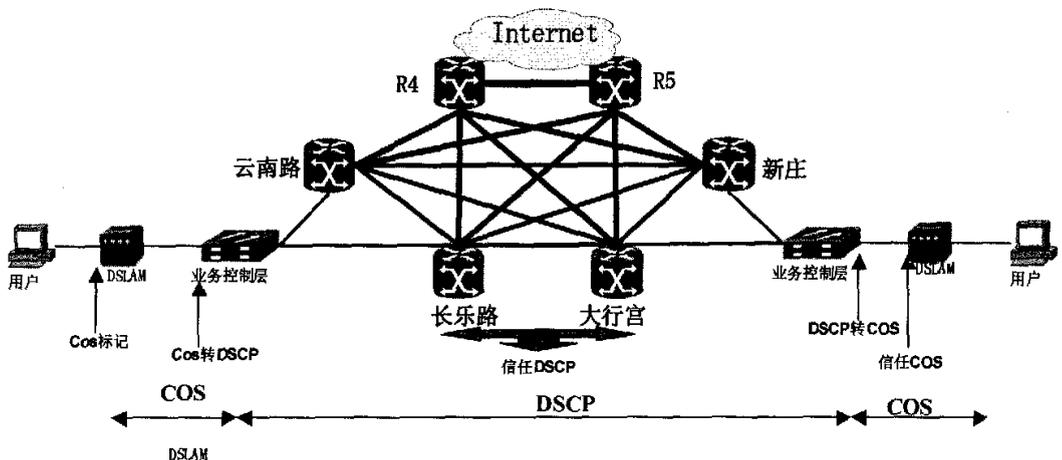


图 (5-3) ADSL-ADSL QoS 业务组网

在该种组网方式下，当用户数据进通过 DSLAM 进入网络时，DSLAM 根据相应网络服务级别利用 COS 将数据包标记成相应服务级别，DSLAM 根据 COS 级别进行队列调度与拥塞管理，在业务控制层（BAS）网络设备将 COS 转换成 DSCP 服务级别并进行队列调度与拥塞管理，核心层网络设备根据 DSCP 值进行队列调度与拥塞管理并将数据包传输至对端业务控制层，业务控制层将 DSCP 值转换成 COS 并传输给 DSLAM，DSLAM（DSLAM 下行接口没有队列调度与拥塞管理功能）将数据传送给用户。

以上仅是同南京目前 QoS 应用组网方式，随着业务的不断发展与 CN2 网络的建设，将会有新的 QoS 应用组网方式。

5.2 组播部署结论及应用分析

1、核心汇聚层

GSR 支持 8 条队列，从电总要求看，目前只需启用 4 条队列，其一为 LLQ 所有 GSR 均配置 MDRR、WRED 和 EXP-DSCP 映射。

(1) 网关层 R4/R5

不需要开启组播

对入网流量进行重新分类标记，对出网流量标记信任

(2) 云南路和大行宫 GSR

开启组播且担任主备 RP，配置 PIM-SM，另外配置组播源准入控制以维护组播源的合法性

标记信任

(3) 新庄和长乐路 GSR

开启组播，配置 PIM-SM

标记信任

2、业务接入控制点

(1) 各设备均开启组播并配置 PIM-SM。

(2) 三层交换机上加上 igmp-snooping 配置。

- (3) 各设备用户接入端口均配置静态组播组和组播过滤。
- (4) 各设备配置分类标记策略, 根据五元组完成 DSCP 标记, 接入层设备携入 QoS 标记的予以信任并作 QoS-DSCP 映射。
- (5) 各设备上行端口均配置 IP3Q 和 WRED (MA5200F 和 QW3026 不支持 WRED), 对用户上行流量实施 QoS 策略。
- (6) 由于限速任务主要在接入层网络完成, 目前未在 BAS 上配置限速策略。
- (7) BAS 下行端口配置流量整形。

3、边缘接入层网络

1. L2交换机配置igmp-snooping
2. L2交换机配置根据端口标记QoS策略
3. DSLAM配置根据PVC标记QoS策略
4. L2交换机和DSLAM上联端口均配置队列调度策略
5. L2交换机和DSLAM上行端口配置限速策略

4、组播业务组网方式

南京电信城域网内目前主要有以下组播业务组网方式:

- 1、ADSL终端用户;
- 2、LAN终端用户;

在组播网络组网方式中, 核心层与业务控制层启用 PIM-SM 组播协议, 在二层交换机及三层交换机中启用 IGMP 协议, DSLAM 不建议启用 IGMP。并在整个网络中配置长乐路与大行宫 GSR 为主备 RP。组播业务服务器通过 C6509 接入南京城域网中。网络拓扑如图 (5-4):

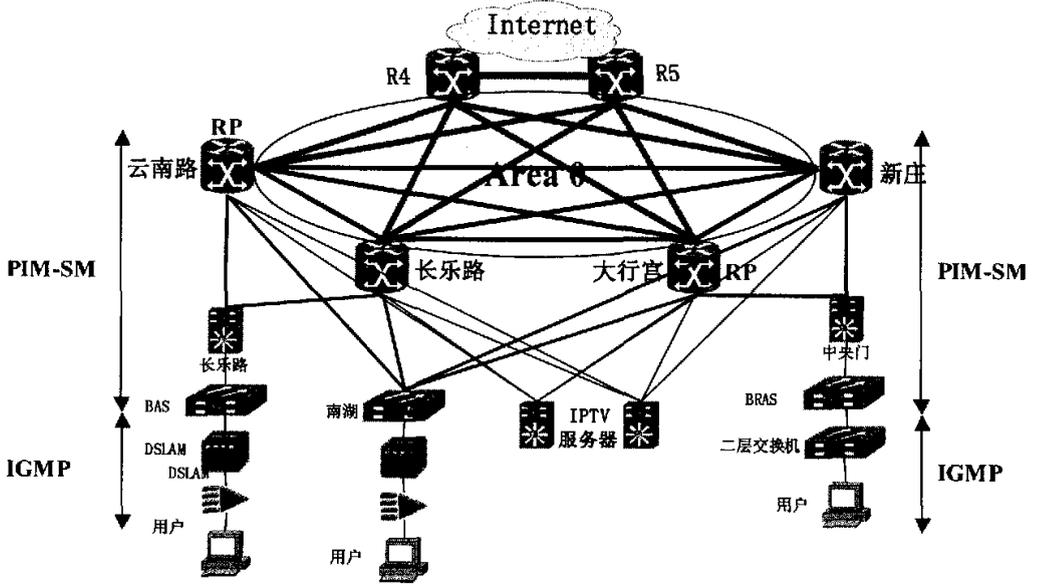


图 (5-4) 组播业务组网

5.3 典型应用

1、IPTV 业务

根据 IPTV 业务规划,其节目源将从首蓿园 6509 接入,上连到城域网的 GSR。

目前城域网的用户有两种接入方式:LAN 接入与 ADSL 接入。如图 (5-5):

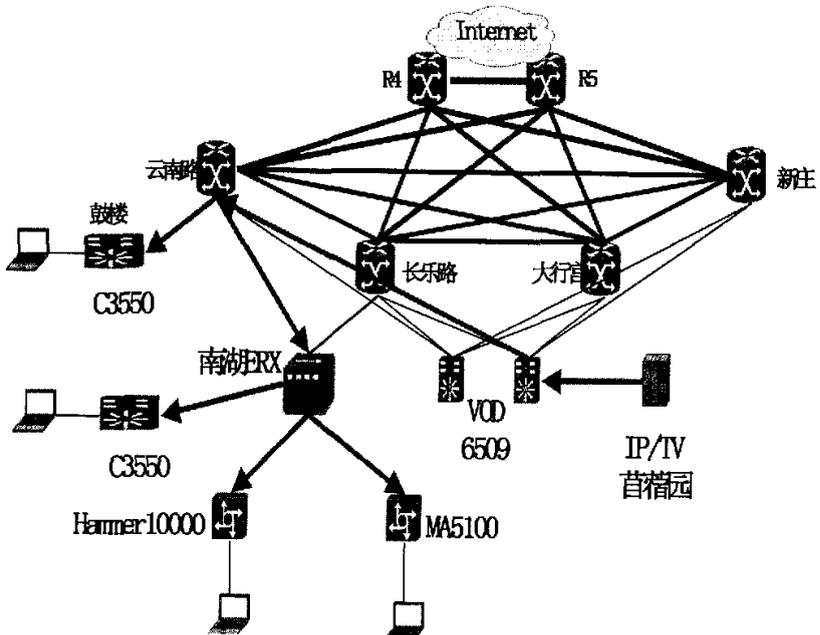


图 (5-5) IPTV业务

为了保证 IPTV 业务能够正常服务，并提供很好的服务质量，需要对组播功能及 QoS 功能作如下部署：

(1) QoS部署：

由于 IPTV 业务是一种中心型业务，其数据主要是从节目源到用户终端的单向数据流。因此在 QoS 部署过程重点是保证下行方向数据的服务质量。具体部署如下：

1. 在苜蓿园6509上完成IPTV数据的DSCP标记，根据规划设为48。由于IPTV服务器的源IP地址是固定的，因此根据源IP地址进行标记。
2. GSR对数据的DSCP值不作改写，下行使用MDRR (PQ) 进行队列调度。
3. 6509对数据的DSCP值不作改写，下行使用WRR (PQ) 进行队列调度。
4. BAS对数据的DSCP值不作改写，并完成DSCP-COS的转换。下行接口使用WRR (PQ) 进行队列调度。(ERX1440/1410不具备标记信任功能，必须进行重标记)
5. 对于LAN用户，如果L2交换机能够支持QoS，在下行端口上使用WRR (PQ) 进行队列调度。由于LAN用户接入层的二层交换机上联基本为华为的5200，华为虽然支持COS-DSCP的转换，但是QoS的部署必须针对用户的帐号来做，而无法象ERX那样在端口上做。因此如果途径5200的LAN用户，则不易部署QoS。
6. 对于ADSL用户，由于DSLAM下行接口不能支持QoS，因此不作部署。
7. 在各设备的下行接口部署WRED，但是由于IPTV数据是UDP数据，因此WRED拥塞避免功能对IPTV数据没有效果，考虑到在可能会有其它TCP业务数据在整体QoS部署中与IPTV同级，因此还是有必要部署WRED。

(2) 组播部署：

组播是 IPTV 业务必须部署的功能，具体部署方式如下：

1. 用云南路GSR和大行宫GSR作为RP，启用组播路由和组播过滤，配置MSDP协议使用两台设备作为AnycastRP，提供RP备份功能，在下行到6509/BAS的端口及互连到其它GSR的端口上配置PIM-SM。

2. 在首蓆园6509上启用组播路由和组播过滤, 指定RP, 在上行至GSR的端口上配置PIM-SM, 在下连服务器的端口上配置igmp, 并启用速率限制, 保证组播数据流在一定范围之内。
3. 在汇聚层6509上启用组播路由和组播过滤, 指定RP, 在上行至GSR和下行到BAS的端口上配置PIM-SM
4. 在BRAS上启用组播路由和组播过滤, 指定RP, 在上行至GSR的端口上配置PIM-SM, 在下行到DSLAM和L2交换机的接口上启用igmp。随着用户量的增长, 可以考虑部署静态组播。5200对组播的支持能力弱, 不考虑部署。
5. 在L2交换机上启用igmp-snooping (目前ADSL用户的VLAN是独立的, 因此DSLAM上不必部署igmp-snooping功能。

(3) 部署结论:

1. 用户通过ADSL+PPPOE、LAN PPPOE方式可以接入IPTV业务, 但是LAN 路由接入方式由于是接入到华为5200上, 不支持组播, 因此无法支持。
2. 接入IPTV业务时, 用户等待时间为组播建立时间和节目流缓冲时间之和
3. 使用静态组播组方式对上述时间没有明显改善, 说明组播建立时间几乎不为用户察觉
4. 目前大多链路处于轻载, 无压力/无QoS时, IPTV节目收视正常, 播放器默认开始缓冲时间约6秒, 收视中节目无中断/跳跃现象
5. 无QoS/有压力时, 受压力流量影响, IPTV客户端出现缓冲时间延长、节目中断缓冲及节目跳跃等影响IPTV收视的现象
6. 有QoS/无压力时, 由于链路轻载, QoS机制不发挥作用, IPTV节目收视正常
7. 有QoS/有压力时, QoS手段作用明显, IPTV节目收视正常
8. 需和IPTV自身的业务特性相结合

2、新视通业务

新视通业务是一种中心型视频业务, 用户使用新视通时, 由MCU控制数据的转发, 所有终端都与MCU交换数据, 目前MCU位于云南路, 通过一台NE40路由器接入到云南路6509。用户终端有两种接入方式: LAN接入与ADSL接入, 目前大部分用户是LAN用户。如图(5-6):

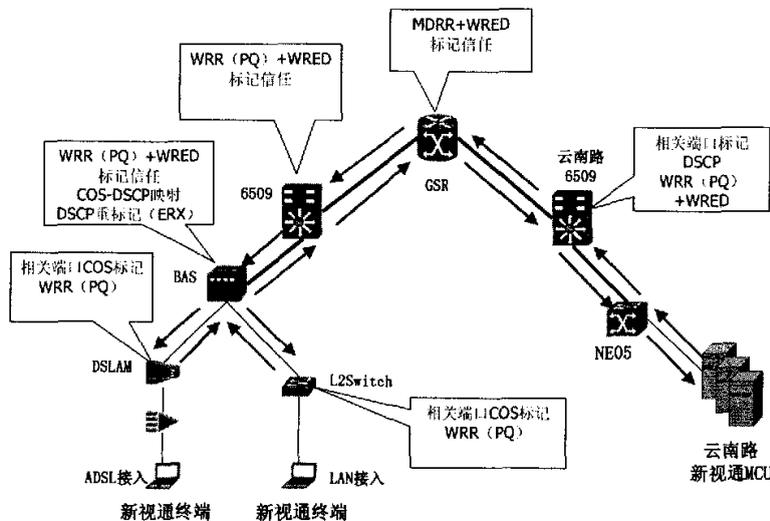


图 (5-6) 新视通业务

新视通业务数据是双向传输，为保证新视通业务的服务质量，必须对上行和下行流量均作 QoS 部署，具体如下：

- 1、在云南路 6509 上标记新视通业务的下行数据（MCU—客户端），上连 GSR 的端口信任入流量数据的 DSCP 值，同时使用 WRR (PQ) 对出流量进行队列调度。
- 2、GSR 对数据的 DSCP 值不作改写，连接 6509 的端口使用 MDRR (PQ) 进行队列调度。
- 3、汇聚层 6509 对数据的 DSCP 值不作改写，上连到 GSR 的端口和下连到 BAS 的端口使用 WRR(PQ)进行队列调度。BAS 对数据的 DSCP 值不作改写(ERX1440/1410 不具备标记信任功能，必须进行重标记)，并完成 DSCP-COS 的转换。上连到 6509 的端口和下连到 DSLAM 的端口使用 WRR (PQ) 进行队列调度。以上 BAS 的测试是基于 ERX，由于 LAN 用户接入层的二层交换机上联基本为华为的 5200，华为虽然支持 COS-DSCP 的转换，但是 QoS 的部署必须针对用户的帐号来做，而无法象 ERX 那样在端口上做。因此如果途径 5200 的 LAN 用户，则不易部署 QoS。
- 4、L2 交换机与 DSLAM 完成终端用户的 COS 标记，上行接口使用 WRR (PQ) 进行队列调度。如果不具备 COS 标记功能，则标记工作在 BAS 上完成。

5、在各设备的相关接口上部署 WRED。但是由于新视通数据是 UDP 数据，因此 WRED 拥塞避免功能对 IPTV 数据没有效果，考虑到在可能会有其它 TCP 业务数据在整体 QoS 部署中与 IPTV 同级，因此还是有必要部署 WRED。

部署结论：

- 1、无 QoS，无压力：由于链路轻载，业务正常，丢包率为 0，无马赛克现象
- 2、无 QoS，有压力：压力流量作用下，业务质量明显下降，丢包率为 1-4%，马赛克现象明显
- 3、有 QoS，无压力：由于链路轻载，业务正常，QoS 机制不发挥作用，丢包率为 0，无马赛克现象
- 4、有 QoS，有压力：QoS 作用明显，业务恢复正常，丢包率为 0，无马赛克现象
- 5、需和新视通自身的业务特性相结合

5.4 南京电信城域网主要网络设备 QoS 和组播功能测试总结

1、设备型号与版本

本次 QoS 及组播部署涉及到的设备的硬件和软件版本如表(5-1)：

表(5-1)：

	设备型号	引擎硬件版本	软件版本	板卡型号
骨干设备	Cisco 12416	C12KPRP-P-M	12.0(26)S4	EPA-GE/FE-BBRD
				EPA-3GE-SX/LH-LC
业务接入	Catalyst 6509	SUP2+PFC2	12.1(13)E9	WS-XG516-GBIC
	ERX1410	ERX-10G2GECC-SRP	5.1.2	8口 100M/1口 GE 的处理线卡 1+1GE 的接口卡
控制点	MA5200F		2.10 RELEASE 7127	
	MA5200G		3.30 RELEASE 2209	GE 板和 FE 板

L2 交换机	Catalyst 3550		12. 2. 25. SEA	
	3206		VRP3. 10	
	2403H		VRP3. 10	
	LS 2024		BOOT0. 99(49)	
DSLAM	MA5100	V100R003	B10D003	LAND 卡
	ZTE8220	V1. 1. 2CORES	v1. 1. 2k	
	ASAM7300	UD	4. 3. 30	
	Hammer10000	Version 2. 30	V2. 0	

2、QoS 和组播支持能力

根据本次设备功能测试的结果, 下表中的设备对 QoS 及组播各项功能的支持能力如表 (5-2):

表 (5-2):

功能		核心层	汇聚层	业务接入控制点				
		Cisco12416	C6509	ERX1410	MA5200F	MA5200G	C3550	QW3026
组播功能	igmp	√	√	√	√	√	√	√
	anycast RP	√	√	NA	NA	NA	√	NA
	PIM-SM	√	√	√	√	√	√	√
	静态组播路由	√	√	√	√	√	√	√
	组播过滤	√	√	√	NA	√	√	NA
	igmp snooping	NA	√	NA	√	√	√	√
QoS 功能	限速	√	√	√	√	√	√	√
	分类标记	√	√	√	√	√	√	√
	标记信任	√	√	√	√	√	√	√
	COS 映射	√	√	√	√	√	√	√
	EXP 映射	√	NA	√	NA	NA	NA	NA

队列调度	√	√	√	√	√	√	√
拥塞控制	√	√	√	NA	√	√	√
流量整形	√	√	√	NA	√	NA	NA

功能		二层接入交换机		DSLAM			
		QW2403H	LS2024	MA5100	ZTE8220	ASAM7300	H10000
组播 功能	igmp	√	√	√	√	√	√
	anycast RP	NA	NA	NA	NA	NA	NA
	PIM-SM	NA	NA	NA	NA	NA	NA
	静态组播路由	NA	NA	NA	NA	NA	NA
	组播过滤	NA	NA	NA	NA	NA	NA
	igmp snooping	√	√	√	√	√	√
QOS 功能	限速	√	√	√	√	√	√
	分类标记	√	√	√	√	√	√
	标记信任	NA	NA	NA	NA	NA	NA
	COS 映射	NA	NA	NA	NA	NA	NA
	EXP 映射	NA	NA	NA	NA	NA	NA
	队列调度	√	√	√	√	√	√
	拥塞控制	NA	NA	NA	NA	NA	NA
流量整形	NA	NA	NA	NA	NA	NA	

3、支持的业务分类标记方式为：

设备		分类依据	测试结果	备注
骨干网络	Cisco12416	五元组	√	
	C6509	五元组	√	
业务接入控制点	ERX1410	五元组	√	
		802.1p	√	
	MA5200F	用户子端口	√	
	MA5200G	五元组	√	
		802.1p	√	
	C3550	五元组	√	
		802.1p	√	
QW3026	五元组	√		
L2交换机	QW2403H	物理端口	√	
	LS2024	物理端口	√	
DSLAM	MA5100	用户PVC	√	
	ZTE8220	用户PVC	X	配置不成功
	ASAM7300	用户PVC	√	
	H10000	用户PVC	X	配置不成功

3、 队列调度和拥塞控制能力如表 (5-3) 所示:

表 (5-3):

设备		队列调度测试			拥塞控制测试		
		队列数量	测试结果	备注	拥塞控制	测试结果	备注
骨干网络	Cisco12416	8, 支持MDRR	√		WRED	√	
	C6509	3, 1PQ+2WRR	√		WRED	√	
业务接入控制点	ERX1410	8, PQ+WRR	√		WRED	√	
	MA5200F	4, PQ+WRR	√		NA	NA	
	MA5200G	4, PQ+WRR	√		WRED	√	
	C3550	4, PQ+WRR	√		WRED	√	仅GE口
	QW3026	4, PQ+WRR	√		NA	NA	
L2交换机	QW2403H	2, WRR	√		NA	NA	
	LS2024	2, PQ+WRR	√		NA	NA	
DSLAM	MA5100	2, WRR	√	仅LAND板	NA	NA	
	ZTE8220	WRR	X	配置不成功	NA	NA	
	ASAM7300	WRR	√	仅上行板卡	NA	NA	
	H10000	WRR	X	配置不成功	NA	NA	

4、设备资源占用情况表为：

设备		CPU占用率		内存占用率		测试结论	备注
		无QoS	有QoS	无QoS	有QoS		
骨干网络	Cisco12416	4%	4%	77M	77M	√	
	C6509	20%	20%	40M	40M	√	
业务接入控制点	ERX1410	27%	27%	540M	540M	√	
	MA5200F	9.75%	10.40%	10M	10M	√	
	MA5200G	11%	11%	98M	98M	√	
	C3550	2%	2%	10M	10M	√	
	QW3026	12%	14%	40%	40%	√	
L2交换机	QW2403F	67%	69%	59%	60%	√	
	LS2024	54%	55%	40%	43%	√	
DSLAM	MA5100	35%	35%	50%	51%	√	
	ZTE8220	41%	43%	47%	47%	√	
	ASAM7300	53%	53%	45%	47%	√	
	H10000	36%	36%	41%	41%	√	

上述所有设备的 QoS 和组播功能除了有某些不支持之外，还有一些不能完全支持，存在一定的限制。具体如下：

1. Cisco 12416 和 6509 在 GE PortChannel 上不支持组播和 QoS
2. ERX1410 不信任 CoS 标记，但可根据 CoS 作 DSCP 标记
3. MA5200F 只能根据用户（子端口）进行标记
4. Catalyst 3550 只在 GE 端口上支持 WRED
5. L2 交换机华为 2403H 和联创 2024 只支持根据物理端口标记
6. MA5100 配置 QoS 时必须配置 LAND 板卡，且必须删除原有的所有配置
7. ZTE8220 和 Hammer10000 宣称支持 CoS 和队列调度，但测试中原厂商工程师未配置成功
8. ASAM7300 只能根据 PVC 的特征（cbr/abr/ubr/vbr 等）作 CoS 标记

设备配置规范

总结与展望

总结

IP网络的全业务化对城域网提出了很多新的要求，南京电信在这方面做了大量的研究工作，提出了城域网优化改造的系统化的解决思路，归纳为下面几个方面：

1、将交换型城域网改造为路由型城域网

路由型城域网的要求是“网络层次清晰化”，二层和三层网络分离，构建物理和逻辑层次清晰的三层路由网络和二层宽带接入网，在城域网接入层以上为三层路由网络，三层路由网络的边缘是BRAS及SR（业务路由器），它们作为城域网二层到三层转换的网关设备，构成了城域网的业务网关层。因此，路由型城域网的网络层次可以分为四层：核心层、汇聚层、业务网关层和接入层。

2、网络结构的扁平化

扁平化网络的特点是大容量、少节点、广覆盖，减少物理和逻辑级联级数，提高网络的可靠性、减少报文转发的时延、提高实时多媒体业务的QoS。网络扁平化的设计并非简单地减少网络层次，还需要考虑光纤资源、业务流向、用户分布等多方面因素。

3、业务隔离和QoS差异化

从安全性、地址规划和QoS等角度考虑，全业务承载的城域网需要对不同的业务实施隔离，业务隔离的模式有两种：一种是为每种业务建立一张物理专网，这种模式可以较好地提供业务的QoS保障，但是成本较高；另一种是采用MPLS VPN技术，在物理网上为每个业务隔离出一张虚拟的VPN承载网，成本较低，且可以利用资源预留VPN技术为每种业务分配带宽，是一种切实可行的方案。

4、业务控制的集中化

对于各种多媒体业务,其业务权限管理、业务逻辑控制以及业务分流和VPN映射都需要由网络设备来承担;业务控制集中化的思路就是把业务控制点定位在业务网关层,由BRAS和SR来承担。

展望

现有的 IP 城域网是在 IPv4 协议的基础上运行。IPv6 是下一版本的互联网协议,它的提出最初是因为随着互联网的迅速发展,IPv4 定义的有限地址空间将被耗尽,地址空间的不足必将影响互联网的进一步发展。为了扩大地址空间,拟通过 IPv6 重新定义地址空间。IPv4 采用 32 位地址长度,只有大约 43 亿个地址,估计在 2005~2010 年间将被分配完毕,而 IPv6 采用 128 位地址长度,几乎可以不受限制地提供地址。按保守方法估算 IPv6 实际可分配的地址,整个地球每平方米面积上可分配 1000 多个地址。在 IPv6 的设计过程中除了一劳永逸地解决地址短缺问题以外,还考虑了在 IPv4 中解决不好的其它问题。IPv6 的主要优势体现在以下几方面:扩大地址空间、提高网络的整体吞吐量、改善服务质量(QoS)、安全性有更好的保证、支持即插即用和移动性、更好实现多播功能。

显然,IPv6 的优势能够对上述挑战直接或间接地作出贡献。其中最突出的是 IPv6 大大地扩大了地址空间,恢复了原来因地址受限而失去的端到端连接功能,为互联网的普及与深化发展提供了基本条件。当然,IPv6 并非十全十美、一劳永逸,不可能解决所有问题。IPv6 只能在发展中不断完善,也不可能在一夜之间发生,过渡需要时间和成本,但从长远看,IPv6 有利于互联网的持续和长久发展。

1、IPv6 与 IPv4 相比有什么特点和优点?

1) 更大的地址空间。IPv4 中规定 IP 地址长度为 32, 即有 $2^{32}-1$ 个地址;而 IPv6 中 IP 地址的长度为 128, 即有 $2^{128}-1$ 个地址。

2) 更小的路由表。IPv6 的地址分配一开始就遵循聚类(Aggregation)的原则,这使得路由器能在路由表中用一条记录(Entry)表示一片子网,大大减小了路由器中路由表的长度,提高了路由器转发数据包的速度。

3) 增强的组播(Multicast)支持以及对流的支持(Flow-control)。这使得网络上的多媒体应用有了长足发展的机会,为服务质量(QoS)控制提供了良好的网络平台。

4) 加入了对自动配置(Auto-configuration)的支持.这是对 DHCP 协议的改进和扩展,使得网络(尤其是局域网)的管理更加方便和快捷.

5) 更高的安全性.在使用 IPv6 网络中用户可以对网络层的数据进行加密并对 IP 报文进行校验,这极大的增强了网络安全.

2、 IPv6 的地址资源

IPv6 拥有十分丰富的地址资源.由于汲取了 IPv4 地址资源不足的教训,IPv6 的设计者一下子将地址长度扩大了 4 倍,即从 IPv4 的 32bit 地址,扩展到了 IPv6 的 128bit 地址.如果这些地址平均分配,地球上的每一平方米即可获得几百万个 IP 地址.从人们的实际需求来看,很难将这样大的地址空间消费掉. IPv6 具有足够宽广的地址空间,在 Internet 的爆炸性发展, IPv4 地址资源紧缺的今天,无疑是极为重要的.

不过,从网络设计的角度来看,地址空间并非越大越好,因为地址空间大是要付出代价的,即在信息的传输和地址的处理等方面都将付出代价.正确的做法是,先对需求作出客观的估计,同时确定地址的分配原则,然后再来确定真正的地址空间大小.与 IPv4 地址的设计存在同样问题的是, IPv6 地址的设计也没有严格遵照需求来规划,而是通过“拍脑袋”定下来的.想当年,如果 IPv4 不是定为 32bit,而是与 MAC 地址一样定为 48bit, IPv4 将不会像现在这样成为“鸡肋”.而 IPv6 的设计者同样是以‘想当然’来确定地址长度的,一下子将地址长度扩大了 4 倍,而在地址分配方面却只给出了一个极为粗略的原则. IPv6 的地址肯定够用,但地址设计的合理性则是有疑问的.

目前, IPv6 标准已经颁布, 128bit 地址也已确定. IPv6 启用后,再也不会为地址而发愁了.但是, IPv6 地址的规划仍然是大问题,需要下大力量去做.合理的地址规划将大大提高地址的寻址效率,大大减小路由表中的路由条目,从而极大地提高 IP 网络的性能.

3、 移动 IP

移动 IP 是一个十分热门的话题,它涉及 IP 网上对个人的移动性、终端的移动性和系统的移动性的支持.与目前无线移动通信不同的是,移动 IP 主要涉及 IP 协议对 IP 网中移动性的支持问题.由于目前广泛使用的 IP 网是 IPv4 网,因而移动 IP 必然是先从 IPv4 入手.

对于 IP 网的移动性来说,它涉及的业务有二类:

- 非实时业务；
- 实时业务。

这两类业务对协议的要求是不同的。对于非实时业务，传输时延不是关键参数，非实时业务在传输过程中采用迂回路由是可以接受的。传输非实时业务的要求是，在通信过程中，不论个人、终端或系统移动到何处，业务的通信连接（TCP 连接）必须保持不断，对于非实时业务移动性的支持，IPv4 协议已经可以做到。对实时业务而言，传输时延是一个极其重要的参数，它不能容忍采用三角迂回路由进行通信。在这方面，IPv4 很难处理，IPv6 的解决办法则有可取之处。因此，在移动 IP 方面，应将 IPv6 作为首选目标。

4、IP 网的 QoS 问题

IP 网的 QoS 问题是一个极大的问题，从协议来说，IPv4 考虑到了 QoS 问题，它的 TOS 字段就是用于区分服务类型，并以此来提供不同的服务。遗憾的是，IP 网的设计者将 IP 网定义为是一个“尽力而为”地提供传输服务的网，因而 IP 网不能提供对不同类型的业务提供分类服务的手段。在实际网络中，网络设备甚至不对 TOS 作任何处理。当然，由于 TOS 字段是在 IP 报头之中，对 TOS 的处理亦是一个不小的开销。

关于 IPv6 对 QoS 的考虑，主要是设定了通信流类型（8bit）和数据流标号（20bit）。当然，这 28 bit 只是用来指示特定的数据流，真正 QoS 的实现还需要网络设备采用特定技术来实现。从本质上看，IPv6 的 28bit 与 IPv4 的 6bit 的 TOS 的用途是类似的。对 IP 网来说，特别是在传输成本大大下降的今天，是否有必要引入数据流的标记，是一个十分值得商榷的问题，IPv6 在 QoS 方面的改善尚有疑问。

5、IP 网中的安全问题

安全问题是 IP 网中永远存在的问题。当然，它与 IP 网的网络协议有关，更重要的是与 IP 网的管理机制有关。从本质上讲，基于分组交换技术的网络的安全强度是相仿的。但大家一致认为，X.25 网的安全性大大优于 IP 网，其真正的原因在于网络的管理机制而不是协议机制。举例来说，在 X.25 网中，网络设备要对源地址进行严格检查，凡是与源地址不符的分组将会被抛弃，源地址与接入端口是严格绑定的。这样一来，X.25 网的安全性确实会好得多。IP 网没有这些管理机制（源地址检查、源地址与接入端口绑定等），可能是造成 IP 网安全问

题的原因。当然 X.25 网与 IP 网不能简单类比，IP 网中的情况要复杂得多。

IPv6 对安全性有周密的考虑。Ipsec 最初是专门为 IPv6 设计的，有 Ipsec 的配合，IPv6 的安全性的确大为改善，它已经有能力提供用户与用户之间端到端的安全性。从网络提供者来说，做到这一点已经足够了。但是，由于种种原因，IPv6 迟迟不能推广使用。尽管 Ipsec 是伴随 IPv6 诞生的，但从目前来看，已无特色可言。当然，从使用方面看，由于在 IPv6 中有专用的安全报头，还是比 IPv4 方便。

IP 网的高速发展已成为不可阻挡的趋势。从发展的眼光看，IPv6 势在必行。采用 IPv6 技术来组建 IP 电信网，从而推动 IPv6 的发展，为国内 IP 网络打下可持续高速发展的基础。

致 谢

在即将结束课题之际，谨向我的导师张登银副研究员表示衷心的感谢！张登银老师严谨的治学态度、求实创新的指导思想、渊博的学识和扎实的理论功底使我受益匪浅，并将对我今后的工作和学习产生重要影响，感谢他给予我的悉心教导和无私帮助！

同时要感谢叶祥云主任，在参与南京 IP 城域网优化的工作过程中，自始至终得到了叶主任的悉心指导，在此表示深深的谢意！

感谢郜建军、印海峰及项目组的其他成员在南京 IP 城域网优化项目的各个阶段与我进行的有益探讨。

最后，感谢在百忙中抽出时间来审阅论文，参加答辩并给予我批评和指导的各位专家和教授。

李 磊

二零零六年三月

参考文献

- [1]、方晖. 计算机网络技术基础. 北京: 中国电力出版社, 2005. 2.
- [2]、吴自容. IP 路由技术基础. 北京: 清华大学出版社, 1999. 4.
- [3]、陈启美. IP 宽带通信网络技术. 北京: 北京邮电大学出版社, 2006. 1.
- [4]、法罗赞恩, 费根著, 谢希仁, Chung Fegan 译. TCP/IP 协议族(第 2 版).
北京:清华大学出版社, 2003. 11
- [5]、(美) Douglas E. Comer David L. Stevens. TCP/IP 网络互连(第 3 卷英文版). 北京:人民邮电出版社, 2004. 1.
- [6]、(美) 赖利 (Riley C.). CISCO 网络核心技术精解. 北京: 中国水利水电出版社, 2005. 5.
- [7]、(美) Cisco Systems 公司. 网络互连技术手册(第四版). 北京: 人民邮电出版社, 2004. 10.
- [8]、翟禹, 唐宝民, 彭木根. 宽带通信网与组网技术. 北京: 人民邮电出版社, 2004. 7.
- [9]、强磊. 基于软交换的下一代网络组网技术. 北京:人民邮电出版社, 2005. 5.
- [10]、侯自强. 宽带 IP 技术进展. 北京:人民邮电出版社, 2005. 9.

南京IP城域网网络优化及其应用研究

作者: [李磊](#)
学位授予单位: [南京邮电大学](#)

本文链接: http://d.g.wanfangdata.com.cn/Thesis_Y950381.aspx

授权使用: 上海海事大学(wf1shyxy), 授权号: ea82e5c2-7d98-4419-bc00-9e01013a2baf

下载时间: 2010年9月30日