

NGI/NGN 体系结构及其服务质量保障机制研究^{*}

窦 军 曾华荣 汪海鹰

(西南交通大学信息科学与技术学院 成都 610031)

摘 要 本文在深入分析国内外开展的下一代 Internet(NGI)或下一代网络(NGN)相关研究工作的基础上,重点研究了 NGI/NGN 体系结构及其服务质量保障机制。重点阐述了笔者所在的四川省网络通信技术重点实验室提出的一种潜在的 NGI 体系结构框架——SUPA(单物理层用户数据传输与交换平台体系结构)。研究结果表明,SUPA 使得容易实现从 Internet 到 NGI 的平滑过渡,并能够提供更好的服务质量保障。

关键词 NGI,单物理层用户数据传输与交换平台体系结构,面向以太网的物理帧时槽交换,服务质量

Research on NGI/NGN Architectures and their QoS Provisioning Mechanisms

DOU Jun ZENG Hua-Shen WANG Hai-Ying

(School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031)

Abstract This paper focuses on architectures of the Next Generation Internet/Network and their QoS provisioning mechanisms based on a comprehensive review of existing activities of relevant topics. A framework on NGI architecture called SUPA under development at Sichuan Network Communication Technology Laboratory is presented to demonstrate possible transition from current Internet to NGI with better QoS provisioning.

Keywords NGI, SUPA (Single physical layer User-data transfer & switching Platform Architecture), EPFTS (Ethernet-oriented Physical Frame Timeslot Switching), QoS

1 引言

30 多年来,计算机网络技术,特别是 Internet 技术取得了长足的进步并日益深入到人类社会的各个领域。另一方面,由于多媒体网络应用需求急剧增加,随着单信道传输能力多个数量级的提高,对实现综合网络业务传输的呼声越来越高。因此,Internet 面临着高速交换和服务质量保障问题的挑战。

尽管 20 世纪 90 年代中期,网络界已经开始提出下一代网络(NGN)或下一代 Internet(NGI)的问题,但在相当长的一段时间内,人们的视线都集中在以 IPv6 为基础的技术上,很少有人挑战现有的网络体系结构或意识到迎接所面临的挑战的关键在于重新审视现有网络体系结构本身。直到 2005 年,美国成立了全球网络创新环境(GENI)机构并构建相关的试验环境,相关文档中公开承认 IPv6 未能解决 Internet 面临的问题,必须重新审视 Internet 的体系结构^[1],有关网络体系结构的研究问题才引起学术界和业界的广泛重视。

本文从对 Internet 高速交换和服务质量保障的角度出发,在分析国内外相关研究工作的基础上,以笔者的研究工作为基础进一步探讨一种潜在的 NGI 体系结构及其服务质量保障机制。

2 国内外 NGI/NGN 研究工作一瞥

2.1 IETF 和部分标准化机构的工作

作为 20 世纪 70 年代末期产物,OSI/RM 和 Internet 自然地打上了当时“线路传输数率低(Mbps 数量级以下)、通信误码率高($10^{-6} \sim 10^{-5}$)和以文本数据为主体的网络应用背

景”这一烙印。其具体体现就是其典型的通信子网的用户数据传输与交换平台(U-platform,简称“用户数据平台”或“用户平台”)采用三层结构。当通信技术的发展使单波长的数据传输能力提高到 40/80/160 Gbps、单根光纤的传输能力提高到 Tbps 数量级之时,这一多层结构的用户数据平台显现出两大弊端:一方面,路由器两次对多层协议数据单元的处理降低了效率,增大了交换时延及其不确定性;另一方面,服务质量(QoS)参数本质上是物理层传输性能的度量值,而由于在多层平台上进行交换,缺少对物理传输性能的直接控制能力,因而很难满足不同数据流对 QoS 的需求。事实上,IETF 建议的以资源预留协议(RSVP)^[2]为基础的集成服务(Integrated Service)和区分服务(Differentiated Service)之所以未能取得预期的效果,其根本原因就在于受到多层用户数据传输平台层间依赖性的制约。

近年来,计算机网络界把 NGI 和 NGN 的期望寄托于多协议标签交换(MPLS—Multi-Protocol Label Switching)^[3]技术。MPLS 借用带外信令控制的思想,将 Internet 的 U-Platform 提供的服务从三层的无连接的 IP 服务简化为增强型两层的面向连接的服务。显然,MPLS 的 U-Platform 对用户数据的转发效率比 IP 路由高,其面向连接的服务也更适应多媒体应用面向数据流的特点。但是,由于 MPLS 本身并未提供任何服务质量的保障机制,而是将标签与标签子层下的所谓的前传等效类(FEC—Forward Equivalence Class)挂钩,期待现有的以数据链路层为基础的交换网络能够支持期望的 FEC。换言之,在现有具有良好的服务质量保障机制的数据链路层交换网(如 ATM 或帧中继)之上,MPLS 继承了其服

^{*} 本项目工作受国家自然科学基金支持(项目批准号:60372065)。窦 军 副教授,在职博士研究生,主要研究方向:网络体系结构;曾华荣教授,主要研究方向:网络体系结构、网络与通信技术;汪海鹰 博士研究生,主要研究方向:网络与通信技术。

务质量保障能力,也即是以额外的标签子层为代价去提供本来已可保障的服务质量;反之,如果它所依托的网络不提供服务质量保障机制,MPLS 丝毫不能改进其服务质量。文[4]详细分析了 MPLS 和 RSVP 这类采用“层上加层”打补丁的方式来改善服务质量的技术路线的致命弱点,提出应当充分利用带外信令思想进一步将 U-Platform 简化为单层结构,直接在物理层解决交换效率和服务质量保障问题。

采用这类“层上层”思想解决 NGI 问题的的工作还包括 ITU 的 T-MPLS^[5]和“城域以太网论坛”(MEF)的 MAC 帧封装 MAC 帧(MAC in MAC)技术^[6]和服务骨干传送(PBT - Provider Backbone Transport)技术^[7]。

2.2 国内外的“无层”(Layer-less)网络结构

21 世纪初,国内外的部分研究机构从网络传统的分层结构存在的弊端出发,开始探讨“无层”的网络体系结构。

MIT 计算机科学实验室和美国信息科学研究所的 NewArch^[8-10]项目是国外这类研究工作的典型代表。该项目受美国 DARPA 和美国国家自然科学基金的资助,提出以“角色”(Role)表征协议功能和组织通信结构。由于“角色”间不一定具有传统的上下层间关系(从这种意义上讲所谓的“无层结构”并不完全反对具有上下的层间关系),因此具有更大的灵活性,相应地可避免传统固定的网络分层结构中有时不可避免的“跨层操作”,如在无线网络中网络路径的选择常常利用物理层或数据链路层获得信道质量信息来改进路径的选择。

事实上,该项目提出的 FARA (Forwarding directive, Association, and Rendezvous Architecture)模型,就是由底层(Substrate)和上层(即 FARA 模型描述的内容)来界定的(两层)体系结构。FARA 用三个基本元素:前传命令(FD - Forwarding Directive)、关联 (Association)和会晤点(Rendezvous)来描述底层(即传统通信子网和操作系统)之上的协议行为。从长远来看,这一思路有较大的探讨和参考价值,但由于其网络上的模型与现在的分层结构模型差异较大,如何保护在 Internet 上已有的巨额投资以及如何实现从现有 Internet 向未来“基于 FARA 模型”的下一代网络平滑过渡问题必然将会面临较大的困难。

在国内,电子科技大学提出的“服务元体系结构”(SUNA - Service Unit based Network Architecture)^[11, 12]与 NewArch 研究工作的出发点有些类似。NewArch 中的基本功能单元称为“角色”,而 SUNA 中称为“服务元”(Service Unit),在实现上与软件技术中的面向对象技术结合直观而紧密。

与 NewArch 的研究报告类似,在已发表的与 SUNA 有关论文尚未涉及与物理通信子网间的关系和如何全面保障服务质量的问题的讨论。

2.3 GENI 的相关工作

2005 年 8 月 22 日,美国科学基金会(NSF)提出一项计划——GENI,其最初为“全球网络调研环境”(Global Environment for Network Investigations)的缩写,2006 年 3 月同一缩写被解释为“全球网络创新环境”(Global Environment for Network Innovations),从“调研”到“创新”,强化了该项工作的创新性。该行动计划由 NSF 下属的美国计算机与信息科学及工程(CISE, Computer and Information Science and Engineering)理事会来筹划,预计投入 3 亿美元,用 5 到 7 年的时间完成。NSF 将通过拨款或签订合作协议的方式来资

助 GENI 的研究。GENI 的参与者将包括美国的大学与学院、民营企业、国际合作机构,鼓励建立各种研究合作伙伴关系,尤其是产、学、研的合作,因此是一个开放研究与试验环境,目前尚处于草案规划阶段。

GENI 的行动计划包括研究性项目(Research Program)和网络创新的全局性试验设施(Global Experimental Facility)。前者与 NGI 体系结构的研究直接相关,相关文件中提到的创新性研究领域、理念和原则涉及广泛的领域,包括:

- 通过调查分析网络复杂性、可扩展性需求和经济激励因素,研究新的体系结构需求。

- 在交换技术研究上应当允许超越现有的数据报、分组和电路交换机制;设计新的命名、寻址和身份识别体系结构以及新的网络管理机制;支持新兴技术(如新的无线技术和光技术)和新计算方式(可通过广泛的设备实现)。

- 新体系结构设计必须注重安全性和可用性,必须在私密性与可溯源性之间进行平衡。

- 建立高级服务抽象框架,基于诸如信息目标、位置的服务和身份识别框架等;构建新的服务与应用,制定分布式应用的原则与模式,使大规模分布式应用变得安全、稳固和易于管理;必须充分考虑地区的差异与价值差异。

截至 2007 年 3 月,GENI 已发布文件 42 个^[13]。尽管相关文件对未来的网络体系结构相关的研究内容、设计原则和理念进行了广泛的探讨,但目前进入实质性阶段的工作仍限于创新试验(设施)环境的设计和建设,尚未发布任何与 NGI 新体系结构直接相关的具体建议方案。

3 单物理层用户数据传输与交换平台体系结构(SUPA)框架

30 多年来,通信技术的发展把传输误码率降低了 6 个数量级,因此 OSI/RM 中需要独立的数据链路层的理念需要重新审视。如果将分组交换网所需的分组(或帧)与现今已经广泛存在的“物理帧”融合,不但能够免去各层数据格式间的转换,提高处理效率,还可能将物理帧与对信道的时分复用机制结合在一起,在物理层实现服务质量能够得到保障的物理层的高速、高效交换。这就是提出 SUPA 的基本思路。图 1 为 SUPA 协议层次结构框架^[14]。

图 1 中的 SUPA 端系统代表支持 SUPA 的计算机或 SUPA 域与传统 Internet 域间的互连设备——半网关(Half-gateway)。该半网关在 Internet 侧,支持传统的 Internet 协议集;在 SUPANET 侧,支持 SUPA 协议。二者之间的中继功能实现两类网络的适配。这类网关能够为传统 Internet 设备借用 SUPANET 实现互联提供一种高效的“摆渡”服务(或称“装载服务”或“隧道服务”)。在这种情况下,由于用户数据的传输是在高速、高效的物理层 EPFTS (Ethernet-oriented Physical Frame Timeslot Switching,面向以太网的物理帧时槽交换)平台(图 1 中的 EPF(面向以太网的物理帧)子层)上进行的,对传统的 Internet 设备来讲,SUPANET 相当于一个高效传输“管道”,因此其互联效果就像 Internet 设备直接互联一样。SUPA 中间系统是 SUPA 域中的节点设备,SUPA 中间系统间的互联将通过“网络-网络接口”(NNI)来完成;而与 SUPA 端系统互联时,则通过“用户-网络接口”(UNI)来完成。

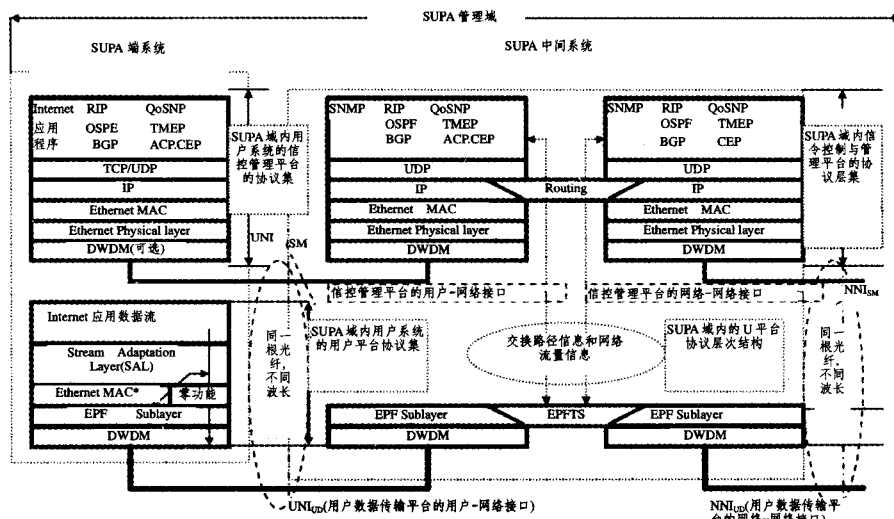


图 1 SUPA 模式下 SUPA 管理域内的接口及协议层次结构

图 1 的上半部分表示用户系统和 SUPA 节点的信控与管理平台(S&M-Platform)的协议层次结构以及用户与网络间的 S&M 接口(UNI_{SM}和 NNI_{SM}),下半部分为 U-Platform 的协议层次结构和接口(UNI_{UD}和 NNI_{UD})。在 SUPA 域内,用户数据传输平台被简化为单物理层,由 EPF 子层和 DWDM 子层组成。SUPA 节点利用 EPFTS 实现在多个端口间数据的高速转接^[15]。SUPANET 设备也可以工作于缺省 Internet 模式,此时接口将按带内信令方式工作。换言之,UNI 或 NNI 的信控管理接口与用户数据传输用接口概念上是同一个接口。此时,图 1 中上半部的 SUPA 专用协议(QoS, TMEP 等)将处于不活动状态,所提供的服务也就退化为传统的 Internet 服务。为了使 SUPA 节点能自动识别用户希望传统 Internet 服务还是 SUPANET 服务,选用 SUPANET 服务时,IP 报头中的特殊字段将使用特殊编码(IPv4 的 TOS 和 IPv6 中的 Traffic Class)。

为了解决在 SUPA 平台间信息的交换,我们在平台间定义了“节点内部访问接口”(IAI - Intra-node Access Interface)^[16]。完全实现 SUPA 平台整合前,利用 IAI 能够方便地利用现有的路由器设备与 EPF 交换平台紧耦合实现 SUPA 节点,从而易于实现从 Internet 向 SUPANET 的平滑过渡。

4 SUPANET 的服务质量保障体系

SUPANET 的服务质量保障由 S&M-Platform 和 U-Platform 的一系列机制协同完成^[17]。

4.1 S&M-Platform 的服务质量保障机制

• 服务质量协商协议(QoSNP—QoS Negotiation Protocol)^[18]

网络用户与网络节点间就服务质量达成“业务级共识”(SLA—Service-Level Agreement)的基本手段。

• 基于 QoS 的波长路由协议(QLR—QoS-based Lambda Routing Protocol)

解决 DWDM 环境中的多波长中服务质量信息交换与路径算法,为 QoSNP 提供路径选择依据和实现波长负载均衡^[19]。

• 波长内配额约束(Quota Constraint)机制与动态配额调整机制

仿真实验表明,不同优先级用户数据流的最大资源配

额对服务质量参数影响极大,根据对网络性能监控动态调整配额,以达到性能优化。

• 呼叫准入控制(CAC—Call Admission Control)是根据网络拥塞状况预防网络拥塞的机制之一。

4.2 U-Platform 的服务质量保障机制

• 新型交换结构(NISO-MEFS, N-Input-Single- Output)

传统的以交叉开关(Crossbar)为基础的交换结构采用面向交换结构输入端口排队方式,其仲裁调度算法较复杂,为解决队头堵塞问题需要与虚拟输出队列(VOQ)相结合方式,在多优先级队列情况下更是如此。采用 NISO-MEFS,因事先将指向同一输出端口的数据引导到相应的 NISO-MEFS 结构排队,因此直接采用面向输出排队(OOQ—Output-Oriented Queuing)方式,仲裁调度大大简化。

• 超前交换技术(SFS—Step-Forward Switching)

传统的面向链接服务的交换采用类似无连接方式查表选择输出端口的的方式。但是,在建立连接过程中下游节点输出端口实际已知,在传统交换中为充分利用。在 EPFTS 技术研究中,我们充分利用了这一特点,在 EPF 帧中携带了相关信息,不经过查表即可提前请求交换结构仲裁,在一定程度上减少等待时延。其重要意义还在于,有可能不经过传统的交换结构,在光域内直接转发。相关研究工作目前仍在进行中。

• 对用户数据的基于服务质量协商的入网控制(QUAC—QoS-based User-data Admission Control)

基于连接建立阶段 QoSNP 协商的参数对用户数据入网控制是履行网络承诺和保障服务质量的重要手段。QUAC 基于服务质量协商结果和网络拥塞状况,对违约的过量数据采用了不同的限流策略,包括挪用临近采样周期的准入额度的方式,如“借还”和“存留备用”算法。

5 面向以太网的物理帧时槽交换(EPFTS)

由于以太网的 MAC 是迄今为止 95% 以上的用户机和用户网络(CPN)都使用的数据链路层技术,因此本研究选择以“以太网 MAC 帧”作为 U-Platform 的物理帧(EPF, Ethernet-oriented Physical Frame)的运载对象具有较大的普适性。由于现代交换结构都毫无例外地采用定长信元交换,为了避免

(下转第 45 页)

提高效率的重要原因。这种分布式 Crossbar、分布式交换的设计理念是核心网络设备设计的发展方向,保证了现在的网络核心能支撑未来海量的数据交换和灵活的多业务支持的需求。

以太网交换机实际是一个为转发数据包优化的计算机。只要是计算机,就有被攻击的可能,比如非法获取交换机的控制权,导致网络瘫痪,另一方面也会受到攻击。传统交换机主要用于数据包的快速转发,强调转发性能。随着局域网的广泛互连,加上 TCP/IP 协议本身的开放性,网络安全成为一个突出问题,网络中的敏感数据、机密信息被泄露、重要数据设备被攻击,而交换机作为网络环境中重要的转发设备,其原来的安全特性已经无法满足现在的安全需求,因此传统的交换

机无论是结构体系还是软件系统都需要增加安全性。

参考文献

- 1 楚政. 以太网交换技术及发展[J]. 电信科学, 2003(5)
- 2 安学军, 张佩珩, 高文学, 等. 基于 UX8 交换芯片的机群互连网络设计[J]. 微电子学与计算机, 2003(7)
- 3 张磊. 基于消息队列的自治异构信息查询系统的研究与实现[D]. 国防科学技术大学, 2003
- 4 毕汝超. 以太网交换机 EAPS 系统的设计[D]. 西北工业大学, 2007
- 5 郑燕峰. 基于输入排队的可扩展交换结构调度算法的研究[D]. 中国科学院研究生院(计算技术研究所), 2006
- 6 荆元利. 基于片上网络的系统芯片研究[D]. 西北工业大学, 2005
- 7 林闯, 单志广, 盛立杰, 等. Internet 区分服务及其几个热点问题的研究[J]. 计算机学报, 2000(4)

(上接第 33 页)

“变长帧”在交换前、后的分割与重组, EPFTS 直接采用以定长帧 EPF 作为基本交换单元。EPFTS 的“交换时槽”(Time-slot)为交换结构转发一个 EPF 所需时间, 对输出波长的复用则以该波长上传输单个 EPF 所需的时间为“复用时槽”对该波长的传输能力进行复用, 收发双方以 EPF 为基础进行异步传输, 以避免频繁的时钟同步^[14, 20]。图 2 为 EPF 帧格式示意图:

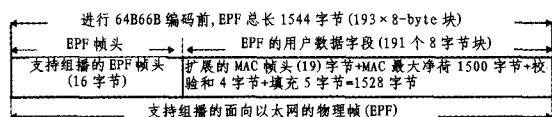


图 2 支持组播的面向以太网的物理帧格式

如图 2 所示, EPF 的载荷选择为以太网 MAC 帧的最大长度, 即 1523 字节(不包括前导 7 字节)。为了适应采用 64B66B 编码以 8 字节为基础的数据块编码的需要, 增加了 5 字节的填充字段, 使 EPF 的用户数据字段为 1528 字节(191 个 8 字节块)。这样可以避免像 MEF 建议的 MAC in MAC 技术那样需要对被承载的 MAC 帧进行分割。EPF 的 16 字节的帧头中定义了超前交换字段、多端口组播字段以及其他控制字段。

SUPA 的 U-Platform 为用户数据提供“交换虚线路”(VSL - Virtual Switched Line)或“永久虚线路”(VPL - Virtual Permanent Line)服务^[17]。对实时性要求高的单个数据流, EPFTS 可对单个数据流建立 VSL, 提供可保障服务质量的 VSL“集成服务”;对服务质量要求相对较低的数据流, 可为服务质量需求类似的多个数据流建立一个 VSL, 为用户提供类似于“区分服务”的统计复用服务, 以减少交换节点的 VSL 数量;对传统 Internet 能满足的数据传输业务(如文本业务), 可在多对边界节点对间建立若干 VLS 隧道(Tunnel), 供多个数据流共享, 提供“尽其所能”的服务。VSL 是在 S&M-Platform 上通过服务质量协商协议(QoSNP)和基于 QoS 的波长路径选择算法来建立的。

物理层的 VSL 配合组播功能, 很容易被用来构建虚拟专用网(VPN), 有助于提高 SUPANET 用户平台的安全性。

结束语 有关 NGI 和 NGN 的研究是一个长期的探索过程。尽管有关 SUPA 和 EPFTS 的研究工作仍处于前期阶段, 但我们的仿真实验结果初步表明这一研究路线具有较强的可行性和进一步研究的价值。由于充分考虑了保护在 Internet 上的投资问题, 较容易实现从 Internet 向 SUPANET 的平滑过渡。

参考文献

- 1 GENI Design Documents (GDD), 2005 -2007. <http://www.geni.net/documents.php>
- 2 RFC 2205 - Resource Reservation Protocol (RSVP) - Version 1 Functional Specification, IETF, Sep. 1997
- 3 RFC 3031 - Multiprotocol Label Switching Architecture, IETF, Jan. 2001
- 4 ZENG Huaxin, et al. Replace MPLS with EPFTS to Build a SUPANET. In: 2005 IEEE International Workshop on High Performance Switching and Routing (HPSR'05), Hong Kong, May 2005
- 5 ITU-T SG15, Recommendation G. 8110/Y. 1370: MPLS layer network architecture, Transport MPLS, February 2006
- 6 Chadha H. Capitalizing on the Evolution of Metro Nets: Understanding Multidimensional Ethernet, February 2006. <http://www.convergedigest.com/bp-me/bpl.asp?ID=388&ctgy=>
- 7 Bottorff P. PBT/DL. 0 - General Discussion of Provider Backbone Transport in 802.1ah Networks, May 2006. <http://www.ieee802.org/1/files/public/docs2006/ah-bottorff-pbt-v1-0506.pdf>
- 8 Clark D, Wroclawski J, Sollins K, et al. Tussle in cyberspace-defining tomorrow's Internet. In: Proc. of SIGCOMM, 2002
- 9 Clark D, Sollins K, Wroclawski J, et al. Addressing reality: An architectural response to real-world demands on the evolving Internet. ACM SIGCOMM Computer Communication Review, 2003, 33(4): 247~257
- 10 Clark D. FARA: Reorganizing the addressing architecture. In: Proc. of the ACM SIGCOMM, 2003. 313~321
- 11 ZENG J Z, XU J, WU Y, et al. Service Unit based Network Architecture. In: Fan Pingzhi, Shen Hong, eds. Proc. of PDCAT2003. IEEE Press, 2003. 12~16
- 12 曾家智, 徐洁, 吴跃, 等. 服务元网络体系结构和微通信元系统构架. 电子学报, 2004, 32(5): 745~749
- 13 GDD-05-01-GDD-05-06, GDD-06-07-GDD-06-041. available at <http://www.geni.net/>
- 14 Zeng Huaxin, Dou Jun, Xu Dengyuan. Single physical layer U-plane Architecture (SUPA) for next Generation Internet. In: Comprehensive Report on VoIP and enhanced IP Communications Services. IEC Publications, 2004. 197~227
- 15 Zeng Huaxin, Xu Dengyuan, Dou Jun. Promotion of Physical Frame Timeslot Switching (PFTS) over DWDM. In: Annual Review of Communications by IEC Publications, 2004, 57: 809~826
- 16 王慧. SUPANET 节点内部访问接口(IAI)研究:[学位论文]. 西南交通大学, 2006
- 17 DOU Jun, ZENG Huaxin, WANG Haiying. Single User-Plane Architecture and its QoS Provisioning Mechanisms in Signaling and Management (S&M) Planes. In: Proceedings of 5th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT2004), Singapore, December 2004. 429~440
- 18 赵君, 高雨. SUPANET 中 QoS 协商流程的研究. 计算机科学, 2004, 31(增刊)
- 19 王金兰. SUPANET 中基于服务质量的波长路径选择技术研究:[学位论文]. 西南交通大学, 2006. 5
- 20 XU Dengyuan, ZENG Huaxin, Li Ji, et al. Physical Frame Timeslot Switching (PFTS) in the Single User-Plane Architecture Network (SUPANET). In: Proceedings of 5th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT2004), Singapore, December 2004. 383~395