点模式，Intserv 结构必须支持广大不同类型的“网络元素”。

在以上研究的基础上，提出了一个比较合理的基于 MPLS 网络的 Diffserv 网络区支持端到端 Intserv 的实现框架。提出使用 RSVP 协议在 Diffserv 网络区中实现显式接纳控制和有效、动态的资源的机制。在业务分类的基础上，我们给出了 IS 域与 DS 域之间的映射关系，并设计了一个合适的协作体系使三者结合起来提供可扩展的端到端 QoS 服务。这些映射允许被适当设计和配置过的 Diffserv 网络作为 Intserv 网络结构中的“网络元素”存在，并作为整个点到点 QoS 网络的组成部分。定义了 Diffserv 网络区内使用聚集传输控制的网络元素在支持 RSVP 信令时所需的功能。

第五章，对提出的方法应用网络仿真软件 ns 进行仿真。NS 是一个基于 IP 的仿真器，是由 UC Berkeley 在 1998 年开发出的对不同的真实网络进行仿真的一个平台。通过对 ns 功能的扩展设计实现在 ns 中 RSVP 信令及 RSVP 服务类型映射到 Diffserv 网络的实现方法，并嵌入 ns 中扩展 ns 功能。仿真结果验证了所提方法的可行性。

关键词：QoS；多协议标记交换(MPLS)；Diffserv；RSVP；End-to-End；IP；流量工程；

Abstract

This dissertation presents a framework for providing End-to-End *quality of service* (QoS) in the MPLS networks. Multiprotocol label switching (MPLS) is the convergence of connection-oriented forwarding techniques and the Internet routing protocol.

QoS is a set of technologies that enables network administrators to manage the effects of congestion on application traffic by using network resources optimally, rather than by continually adding capacity.

With the growth of the Internet and Intranets, QoS technology that has been developed over a span of several years is quickly becoming more relevant. The evolution of several major QoS mechanisms is described with a special focus on integrated services and differentiated services. Special attention is paid to the role of the IETF in developing QoS mechanisms.

First in Chapter 1, mainly presents the background of MPLS technology, and it's concept, development stage and characteristic.

Chapter 2, presents concepts and the architecture of MPLS , MPLS is a label-based message forwarding mechanism. By using labels, it can set up explicit routes within an MPLS domain. A packet's forwarding path is completely determined by its MPLS label. MPLS also also to route multiple network layer protocols within the same network and can be used as an efficient tunneling mechanism to implement traffic engineering.

Chapter 3, presents an overview for providing *quality of service* (QoS) in the Internet .QoS guarantee is a great challenge for IP network's development . Quality of Service has been one of the principal topics of research and development in packet networks for many years.For solving the problem in reason , connection-based MPLS technologies provide feasible solutions .This is the principles and characteristics of *integrated services*(Intserv) and *differentiated services* (Diffserv) models. The methods for MPLS implementing the two models are also discussed .The current actuality of MPLS and QoS is analyzed. Its future development is also forecasted.

If a packet crosses all MPLS domains, an end-to-end explicit path can be established for the packet. Label also serves as a faster and efficient method for packet classification and forwarding.

Chapter 4, we discuss the design of End-to-End QoS system through MPLS networks. We describe a QoS network that combines RSVP and differentiated

services in a manner that realizes the benefits of each . We propose a method for mapping QoS guarantees between Integrated Services and Differentiated services in the MPLS networks.

The Integrated Services (Intserv) architechure provides a means for the delivery of end-to-end Quality of Service to applications over heterogeneous networks. To support this end-to-end model, the Intserv architecture must be supports Diffserv may be viewed as a "network element" in the total end-to-end path.

Based on the existing classification, we gives the mapping rules and relations between IS and DS. These mappings allow appropriately engineered and configured differentiated service network clouds to play the role of "network elements" in the Integrated services framework, and thus to be used as components of an overall end-to-end Integrated services QoS solution.

Chapter 5, describes the purpose simulator which has been implemented by extending NS. NS is an IP based simulator, began as a variant of the real network simulator implemented by UC Berkeley in 1998. Now, NS version 2 is available as result of the VINT project.

Key Words: QoS, MPLS, Diffserv, RSVP, End-to-End, IP Switch

符号说明

AF: Assured Forwarding 确保转发
ATM: Asynchronous Transfer Mode 异步传输模式
BA: Behavior Aggregate 行为聚合
B-ISDN: Broadband Integrated Services Digital Network 宽带综合业务数字网
BGP: Border Gateway Protocol 边界网关协议
CAC: Connection Admission Control 连接接纳控制
CAR: Committed Access Rate 保证接入速率
CBQ: Class Based Queuing 基于分类队列
CBR: Constraint-Based Routing 约束路由
CBS: Committed Burst Size 提交组量大小
CIR: Committed Information Rate 保证信息速率
CoS: Classification on Flows 服务类型
CR-LDP: Constraint-Based LDP 基于约束路由是标签分发协议
DiffServ: Differentiated Services 区分服务
DSCP: Differentiated Services Code Point 区分服务代码点
EF: Expedited Forwarding 加速转发
EGP: Exterior Gateway Protocol 外部网关路由协议
ERP: Enterprise Resource Planning 企业资源计划
FEC: Forwarding Equivalent Class 转发等价类
FTN: FEC to NHLFE Map FEC到NHLFE映射
FTP: File Transfer Protocol 文件传输协议
GMPLS: Generalized MPLS 通用多协议标签交换
IETF: Internet Engineering Task Force Internet工程任务组
IGP: Interior Gateway Routing Protocol 内部网关路由协议
IntServ: Integrated Services 综合服务
ITU-T: International Union for Telecommunications, telecommunications 国际电信联盟
LDP: Label Distribute Protocol 标签分发协议

LER: Label Edge Router	标签边界路由器
LSR: Label Switch Router	标签交换路由器
MIB: Management Information Base	管理信息库
MPLS: Multi Protocol Label Switching	多协议标签交换
OSPF: Open Shortest Path First	开放式最短路径优先协议
OXC: Optical Cross-Connect	光交叉连接
PHB: Per Hop Behavior	每一跳行为
PSC: PHB Scheduling Class	PHB调度类型
QoS: Quality of Service	服务质量
RED: Random Early Detection	随机早期丢弃
RSVP: Resource Reservation Protocol	资源预留协议
RTP/RTCP: Real time Transport Protocol	实时传输协议
SDH: Synchronous Digital Hierarchy	同步数字序列
SLA: Service Level Agreement	服务水平约定
srTCM: Single Rate Three Color Marker	单速率三色标记
SVC: Switched Virtual Circuit	交换虚电路
TCP: Transfer Control Protocol	传输控制协议
TE: Traffic Engineering	流量工程
TLV: Type Length Value	类型长度值
ToS: Type of Service	服务类型
tr-TCM:Two Rate Three Color Marker	双速率三色标记
UDP: User Datagram Protocol	用户数据报协议
VoIP: Voice over IP	IP语音传输
VPN: Virtual Private Network	虚拟专用网
WFQ: Weighted Fair Queue	加权公平排队
WRED: Weighted Randomly Early Detected	加权随机早期丢弃

第一章　引言

随着信息技术的高速发展，Internet 已经成为一个巨大的公众数据网，快速增加的用户数使得 Internet 主干网的数据量扶摇直上。Internet 的发展对宽带化、多媒体化提出了越来越高的要求，IP 网络的建设迫切需要一种更为高效的技术。

Internet 的迅速发展使 IP 成为计算机网络的应用环境的“既成事实”的标准和开放系统平台。标签交换的想法来自于对 IP 网络中两种基本设备的考察，它们便是我们再熟悉不过的交换机与路由器。可以看到，若仅就交换速度，流量控制性能与性能价格比而言，交换机无疑要远远高于路由器，然而路由器却具有交换机所不能比拟的丰富而灵活的路由功能。这样，我们不禁要想，能不能让一部设备既有交换机的高速度与流量控制能力，又具备路由器灵活的功能呢？这实际上是标签交换技术产生的一个关键动机。

可以预见，IP 网络发展的转折点将是 IP 网络对于服务质量问题的解决以及对各种新兴的增值业务的支持。MPLS 一方面是目前唯一能够保证 IP 网络服务质量的网络技术，另一方面，使用 MPLS 将可以十分高效地实现各种增值业务，如 VPN 等。MPLS 技术成为下一代 IP 网络的基础技术。

标签交换（Label Switch）技术，是对现有的各种标签交换技术的统称，而这些技术，正是 MPLS 技术产生的基础。多协议标签交换，即 MPLS（Multiprotocol Label Switching），是 IP 通信领域中的一项技术，是对传统 IP 网络传输技术（如 IP over ATM，IP over SDH）的改进。它采用集成模型，将第三层 IP 技术与第二层的硬件交换技术结合在一起，并且使用一个定长的标签作为分组在 MPLS 网络传输时所需处理的唯一标志。这种技术兼具了 IP 的灵活性、可扩展性与 ATM 等硬件交换技术的高速性能、QoS 性能、流量控制性能。使用这一技术，将不仅能解决当前网络中存在的大量问题（如 N 平方问题、带宽瓶颈、QoS 保证、组播以及 VPN 支持等问题），而且能够实现许多崭新的功能（如流量工程、显示路由等），是一种理想的 IP 骨干网络技术。

当前，MPLS 技术是当前网络传输研究的热点，也是各大通信设备提供商力推的重点。国际上关于 MPLS 的研究十分活跃，IETF 这两年出台了一些

MPLS 的协议和协议草案，而 ITU-T 的各个工作组也在积极进行相关的研究以及标准的制定工作。

§1.1 MPLS 的发展过程及技术优势[1,2]

MPLS 是在一系列技术的发展和推动下不断发展起来的。其中主要有：

ATM 论坛推出的 IPOA、LANE 和 MPOA；

Toshiba 的信元交换路由器技术 CSR（Cell Switching Router）技术；

Ipsilon 公司的 IP Switching；

Cisco 公司的 Tag Switching；该方案与前两种技术有很大的差别。例如，它不是使用流驱动而是控制（拓扑）驱动方式来建立交换机中的转发表，而且该技术将不仅限于在 ATM 交换机之上使用。

IBM 公司的 ARIS 与 Nortel 公司的 VNS；在 Cisco 公司宣布其技术之后，很快 IBM 公司也推出了它们的标签交换技术 ARIS（Aggregate Route-based IP Switching），并且也形成了 RFC 文件。

国际标准组织认为 MPLS 作为最佳技术，主要有以下原因：

——多协议标签交换技术为解决 IP 网络上流量规划方面的局限性提供了一种新的可能性。尽管 MPLS 是一种非常简单的协议，但它为在 IP 网络上实现流量规划引入一套完整的控制手段。MPLS 通过显式的标签交换路径，能够有效的支持源端对连接的控制，通过与区分服务（Differential Services）及基于约束的寻找路由协议（Constraint-based routing）相结合，MPLS 能够在 IP 网络上支持 QoS；

——适应于较大规模的网络。众所周知，MPOA 非常适用于小规模的网络，然而在较大规模的网络中应用受到限制。而 MPLS 正是为满足大规模网络的各种要求（如灵活性、可扩充性与可管理性等要求）而设计的。

——适应于多种承载网络。大规模的网络可以使用包括 ATM 在内的多种承载技术。从一个较宽的范围来讲，应该选取一种对于 IPOA 是最优，而且对于其它的链路层技术也是最优的技术。而 MPLS 则可能正是能够覆盖这一范围的唯一技术。

——路由控制的灵活性。从选路的角度来讲，MPLS 技术允许我们可以选择使用固定寻路方式或者是动态寻路方式，具体使用哪种方式可以取决于网络操作者的选择。

MPLS 的协议和协议草案，而 ITU-T 的各个工作组也在积极进行相关的研究以及标准的制定工作。

§1.1 MPLS 的发展过程及技术优势[1,2]

MPLS 是在一系列技术的发展和推动下不断发展起来的。其中主要有：

ATM 论坛推出的 IPOA、LANE 和 MPOA；

Toshiba 的信元交换路由器技术 CSR（Cell Switching Router）技术；

Ipsilon 公司的 IP Switching；

Cisco 公司的 Tag Switching；该方案与前两种技术有很大的差别。例如，它不是使用流驱动而是控制（拓扑）驱动方式来建立交换机中的转发表，而且该技术将不仅限于在 ATM 交换机之上使用。

IBM 公司的 ARIS 与 Nortel 公司的 VNS；在 Cisco 公司宣布其技术之后，很快 IBM 公司也推出了它们的标签交换技术 ARIS（Aggregate Route-based IP Switching），并且也形成了 RFC 文件。

国际标准组织认为 MPLS 作为最佳技术，主要有以下原因：

——多协议标签交换技术为解决 IP 网络上流量规划方面的局限性提供了一种新的可能性。尽管 MPLS 是一种非常简单的协议，但它为在 IP 网络上实现流量规划引入一套完整的控制手段。MPLS 通过显式的标签交换路径，能够有效的支持源端对连接的控制，通过与区分服务（Differential Services）及基于约束的寻找路由协议（Constraint-based routing）相结合，MPLS 能够在 IP 网络上支持 QoS；

——适应于较大规模的网络。众所周知，MPOA 非常适用于小规模的网络，然而在较大规模的网络中应用受到限制。而 MPLS 正是为满足大规模网络的各种要求（如灵活性、可扩充性与可管理性等要求）而设计的。

——适应于多种承载网络。大规模的网络可以使用包括 ATM 在内的多种承载技术。从一个较宽的范围来讲，应该选取一种对于 IPOA 是最优，而且对于其它的链路层技术也是最优的技术。而 MPLS 则可能正是能够覆盖这一范围的唯一技术。

——路由控制的灵活性。从选路的角度来讲，MPLS 技术允许我们可以选择使用固定寻路方式或者是动态寻路方式，具体使用哪种方式可以取决于网络操作者的选择。

——IP 业务的业务量工程。目前，ATM 拥有最完整的业务量工程能力。然而，IPOA 的重迭模型无法高效地利用 ATM 的所有能力，而且在使用全连通的 PVC 方式时，其应用的可扩充性将受到“N 平方”问题的限制。MPLS 借用了一些 ATM 技术如 QoS、选路、资源管理等方面的特性，而且引入了显式路由（Explicit Routing）的概念，它有助于将业务量的要求映射到网络拓扑之上。这样，使用 MPLS 可以获得新的、更多的业务量管理性能。

——支持 VPN 业务。MPLS 的主要优势是能够以无连接方式或者显式路由方式提供面向连接的业务，这种特点使得 MPLS 尤其适用于动态隧道技术。而动态隧道技术是目前支持 VPN 业务的有效传送手段。但目前由于提供基于 MPLS 的 VPN 的方式不是唯一的，这使得将它同其它 IPOA 技术进行比较较为困难。

——QoS 方面。IP Diffserv 与 MPLS 具有明显的默契，因为它们在设计中都考虑了满足业务的需求。由于标签的扩展语义可以携带 Diffserv 信息，借助于标签与端到端的标签交换路径及一定的资源预留机制，可以保证 QoS 机制在特定 MPLS 域中的一致性。

§1.2 本论文工作

目前，IETF 在解决 IP　QoS 有两种基本模型，即综合服务模型(Intserv)及其相应的信令协议 RSVP 和区分服务模型(Diffserv)。但这两种 IP 网络的 QoS 控制都不能完全满足需要，它们各有自己是长处和局限。[7, 8]

Intserv 存在许多缺点，首先，可扩展性是 Intserv 的最严重的问题。由于使用了“软状态”的工作方式，同时 RSVP 进行资源预留需要对大量的状态信息进行刷新与储存，当网络规模扩大时，这一模型将无法实现。目前，对于 Intserv 模型，业界已经有了比较一致的意见。这一模型应当应用于网络规模较小，业务质量要求较高的边缘网络，如企业网、园区网等。对于骨干网络的 QoS 技术，则应当使用 Diffserv 模型

Diffserv 本身也还不完善。首先它并不提供全网端到端的服务质量保证。另外有关的许多技术细节 IETF 都还未给出具体明确的规定，例如业务类别的具体划分、每类业务性能的量化描述、IP 的业务类别等等。现在 IETF 的 MPLS 和 Diffserv 工作组都在研究 RSVP 与 Diffserv 框架的结合问题以进一步扩大 Diffserv 的与现有系统的可兼容性。

——IP 业务的业务量工程。目前，ATM 拥有最完整的业务量工程能力。然而，IPOA 的重迭模型无法高效地利用 ATM 的所有能力，而且在使用全连通的 PVC 方式时，其应用的可扩充性将受到“N 平方”问题的限制。MPLS 借用了一些 ATM 技术如 QoS、选路、资源管理等方面的特性，而且引入了显式路由（Explicit Routing）的概念，它有助于将业务量的要求映射到网络拓扑之上。这样，使用 MPLS 可以获得新的、更多的业务量管理性能。

——支持 VPN 业务。MPLS 的主要优势是能够以无连接方式或者显式路由方式提供面向连接的业务，这种特点使得 MPLS 尤其适用于动态隧道技术。而动态隧道技术是目前支持 VPN 业务的有效传送手段。但目前由于提供基于 MPLS 的 VPN 的方式不是唯一的，这使得将它同其它 IPOA 技术进行比较较为困难。

——QoS 方面。IP Diffserv 与 MPLS 具有明显的默契，因为它们在设计中都考虑了满足业务的需求。由于标签的扩展语义可以携带 Diffserv 信息，借助于标签与端到端的标签交换路径及一定的资源预留机制，可以保证 QoS 机制在特定 MPLS 域中的一致性。

§1.2 本论文工作

目前，IETF 在解决 IP　QoS 有两种基本模型，即综合服务模型(Intserv)及其相应的信令协议 RSVP 和区分服务模型(Diffserv)。但这两种 IP 网络的 QoS 控制都不能完全满足需要，它们各有自己是长处和局限。[7, 8]

Intserv 存在许多缺点，首先，可扩展性是 Intserv 的最严重的问题。由于使用了“软状态”的工作方式，同时 RSVP 进行资源预留需要对大量的状态信息进行刷新与储存，当网络规模扩大时，这一模型将无法实现。目前，对于 Intserv 模型，业界已经有了比较一致的意见。这一模型应当应用于网络规模较小，业务质量要求较高的边缘网络，如企业网、园区网等。对于骨干网络的 QoS 技术，则应当使用 Diffserv 模型

Diffserv 本身也还不完善。首先它并不提供全网端到端的服务质量保证。另外有关的许多技术细节 IETF 都还未给出具体明确的规定，例如业务类别的具体划分、每类业务性能的量化描述、IP 的业务类别等等。现在 IETF 的 MPLS 和 Diffserv 工作组都在研究 RSVP 与 Diffserv 框架的结合问题以进一步扩大 Diffserv 的与现有系统的可兼容性。

RSVP、Diffserv、MPLS 等协议都是在 QoS 管理的粒度和网络可扩展性这两个考虑因素之间寻求不同程度的折衷，RSVP 提供更细的 QoS 保障的粒度，而 Diffserv 和 MPLS 具有很好的可扩展性。因而可以将 RSVP 与 Diffserv 协议进行结合。RSVP 与 Diffserv 协议的有机结合和相互补充对 IP 网络 QoS 管理有着重要意义。

为此 IETF 的研究人员提出，可在 Intserv 体系结构上提供一种在异构网络元素之上提供端到端 QoS 的方法。一般来讲，网络元素可以是单独的节点或链路，更复杂的实体（如 ATM 云或 802.3 网络）也可以视为网络元素。在这种意义下，Diffserv 网络也可以视为更大的 Internet 网络中的一种网络元素。[28] 在该框架中，端到端的、定量的 QoS 是通过在含有一个或多个 Diffserv 区的端到端网络中应用 Intersev 模型来提供的。为了优化资源的分配和支持接纳控制，Diffserv 区可以（并不绝对要求）参加端到端的 RSVP 信令过程。从 Intserv 的角度看，网络中的 Diffserv 区被视为连接 Intserv 路由器和主机的虚电路。把 Diffserv 中的域当作 RSVP 中的预留节点，同 RSVP 原型的区别在于，域中的节点不必承担流的状态信息，而使 RSVP 只做建立连接和接纳控制方面的工作，减少了复杂性，同时又提高了灵活性。这一模型同时体现了两种思想的优点，是 Diffserv/IntServ 相结合的思路之一。但总的看来，目前已有的工作在研究这种方法具体实现的方面还不够深入。

因此为了在整个网络中实现端到端的 QoS。例如要解决一个 Diffserv 域的入口路由器(BR1)可以配置为从边缘网络 (EN1)只接受 10 M 的 DSCP 标记为 EF 的低延时、低丢包率服务的数据流量。当 EN1 中产生 2 0 个 1M 的 MPEG-1 视频流时，BR1 将丢弃一半的输入数据流量，很可能每一个视频流均有被丢弃的数据包。这是因为 Diffserv 是对数据包的聚集进行流量控制，而不是针对一个应用流。在这种情况下，即使 Diffserv 域有足够的资源可以为 10 个视频流服务，但由于没有显式的接入控制，这 20 个视频流没有一个可以得到满意的服务。如果加入 显式信令机制，可以拒绝其中 10 个视频流的接入请求，或通知其可选其它 DSCP 对应的服务。

如果应用资源预约机制 (RSVP)请求资源，网络可根据当前可用资源情况作出是否提供服务的决定并通知应用程序。这种动态资源分配方式对资源的利用率较高。而在 Diffserv 网络中，接入控制是以服务水平约定 (SLA)这种半静态的方式进行，只有对数据包的聚集进行流量控制的能力，而没有信

令的传递。为了实现资源预约机制与 Diffserv 相协作的体系从而在整个网络端到端的服务，以上两点的解决是必须的，这也是本文第四章所主要研究的问题。

基于以上的分析，我的硕士学位论文选题定为：基于 MPLS 的 QoS 应用研究。作为硕士学位论文，本文含有以下的创新性工作：

1. 提出了一个较合理的应用于 MPLS 网络的 Diffserv 域支持端到端 QoS 的实现框架。
2. 提出了 RSVP 协议在 Difffserv 域中实现显式接纳控制和有效、动态的资源调度机制。
3. 提出了 Intesev 服务类型到 Diffserv 网络提供的服务之间的映射关系。定义了 Diffserv 域内使用聚集传输控制的网络元素在支持 RSVP 信令时所需的功能。提出了在 MPLS 环境下，使用聚集 RSVP 将 Diffserv 区内的资源可用性信息传递到边界路由器的方法。
4. 对当前的主流开放式网络仿真平台 ns 进行了细致的探索，在 ns 平台上，设计、引入和实现了 RSVP 信令，以及把 RSVP 服务类型映射到 Diffserv 网络的功能模块，扩充了 ns 功能。仿真结果验证了所提方法的可行性和有效性。

第二章　MPLS 基本原理

§2.1　MPLS 网络结构

MPLS 工作组的基本目的是将标签交换的传递算法和网络层寻路结合起来，把基础技术标准化。标签交换技术可以提高网络层寻路的性能价格比，提高网络层的可扩展性，提供更多的灵活性，无须改变传递算法可以允许传送新的业务。

起初的 MPLS 集中处理 IPv4 和 IPv6，但核心技术可扩展到多种网络层协议（如 IPX, Appletalk, DECnet, CLNP）。通常，所有节点利用第三层寻路来决定路由。MPLS 可在网络层的任何介质上传递包。通过简化包的传递，可以降低高速传递的费用，提高传递的效率。

MPLS 是属于第三层交换技术，引入了基于标记的机制，它把选路和转发分开，由标签来规定一个分组通过网络的路径。MPLS 网络由核心部分的标签交换路由器（LSR）、边缘部分的标签边缘路由器（LER）组成。LSR 的作用可以看作是 ATM 交换机与传统路由器的结合，由控制单元和交换单元组成；LER 的作用是分析 IP 包头，用于决定相应的传送级别和标签交换路径

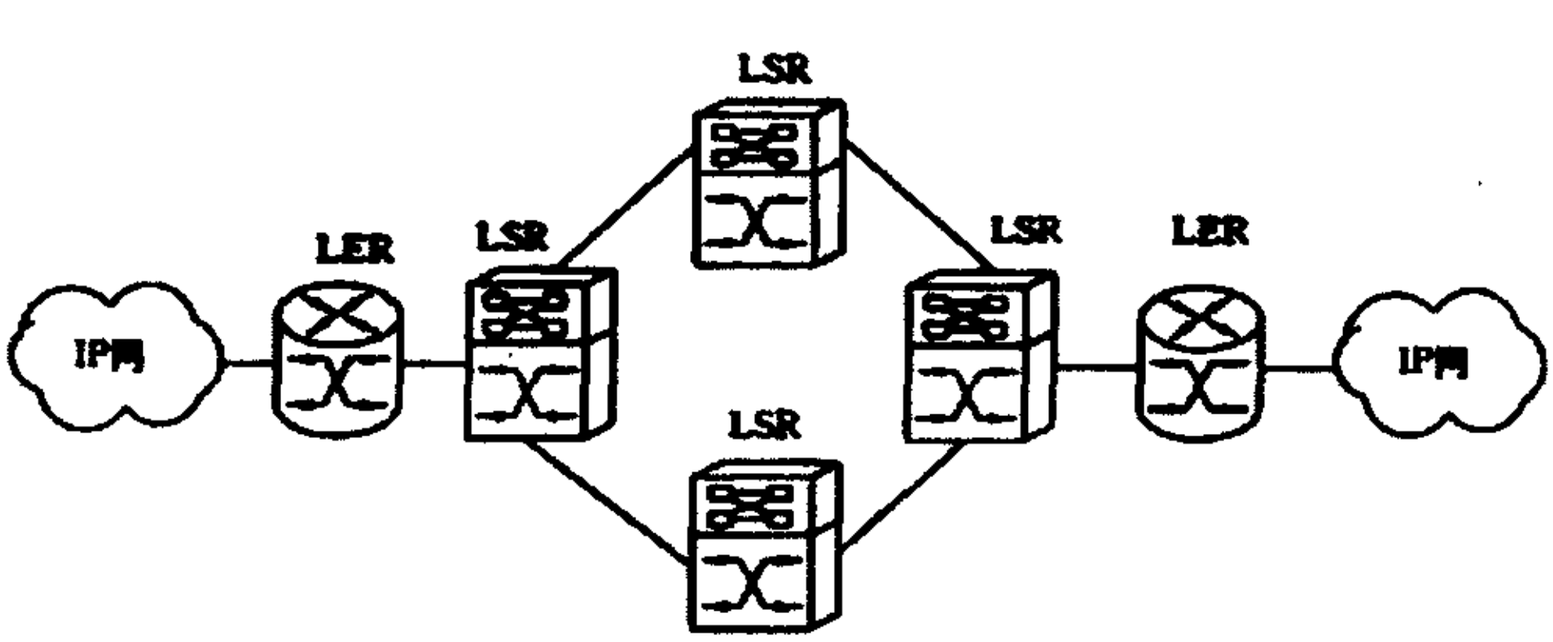

图 1　MPLS 网络示意

（LSP），MPLS 网络如图 1 所示。

LSR 就是实现了 MPLS 功能的 ATM 交换机；LER 可以是具有 MPLS 功能的 ATM 交换机，也可以是具有 MPLS 功能的路由器。标记交换的工作过程可概括为以下 3 个步骤：

第二章　MPLS 基本原理

§2.1　MPLS 网络结构

MPLS 工作组的基本目的是将标签交换的传递算法和网络层寻路结合起来，把基础技术标准化。标签交换技术可以提高网络层寻路的性能价格比，提高网络层的可扩展性，提供更多的灵活性，无须改变传递算法可以允许传送新的业务。

起初的 MPLS 集中处理 IPv4 和 IPv6，但核心技术可扩展到多种网络层协议（如 IPX, Appletalk, DECnet, CLNP）。通常，所有节点利用第三层寻路来决定路由。MPLS 可在网络层的任何介质上传递包。通过简化包的传递，可以降低高速传递的费用，提高传递的效率。

MPLS 是属于第三层交换技术，引入了基于标记的机制，它把选路和转发分开，由标签来规定一个分组通过网络的路径。MPLS 网络由核心部分的标签交换路由器（LSR）、边缘部分的标签边缘路由器（LER）组成。LSR 的作用可以看作是 ATM 交换机与传统路由器的结合，由控制单元和交换单元组成；LER 的作用是分析 IP 包头，用于决定相应的传送级别和标签交换路径

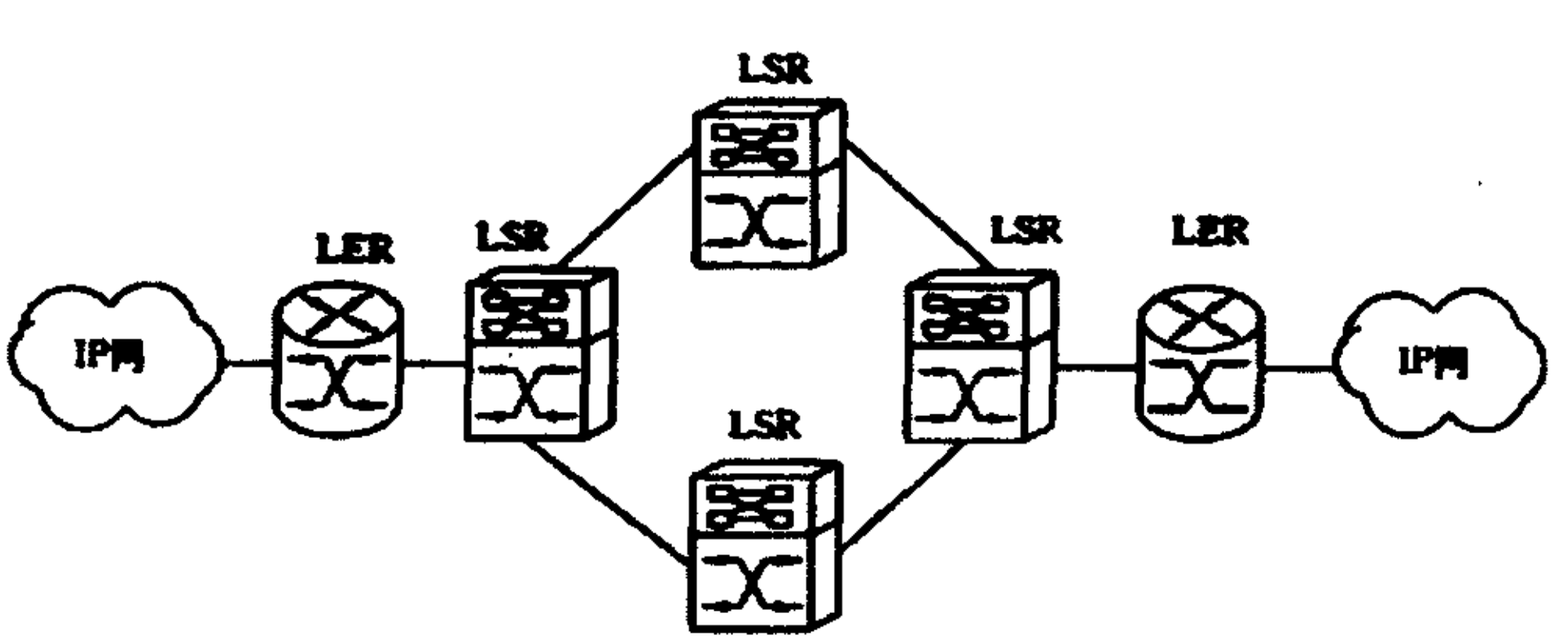

图 1　MPLS 网络示意

（LSP），MPLS 网络如图 1 所示。

LSR 就是实现了 MPLS 功能的 ATM 交换机；LER 可以是具有 MPLS 功能的 ATM 交换机，也可以是具有 MPLS 功能的路由器。标记交换的工作过程可概括为以下 3 个步骤：

1） 由 LDP（标记分布协议）和传统路由协议（OSPF 等）一起，在 LSR 中建立路由表和标记映射表。

2） LER 接收 IP 包，完成第三层功能，并给 IP 包加上标记；在 MPLS 出口的 LER 上，将分组中的标记去掉后继续进行转发。

3） LSR 对分组不再进行任何第三层处理，只是依据分组上的标记通过交换单元对其进行转发。

§2.2 标签分发协议（LDP）

1、标签的定义

标签是一个简短的，具有固定长度的，具有本地意义的标识符，它用与绑定转发等价类 FEC。将相同转发处理方式（目的地相同、使用的转发路径相同、具有相同的服务等级等）的分组归为一类，这就是转发等价类。分组到达 MPLS 网络入口时，被划分为不同的 FEC,根据分组所属 FEC，将适当的标签插入分组头中，然后在网络中按标签进行交换式转发。

如图所示，设 Ru 和 Rd 是标签交换路由器 LSR。标签值 L 表示从 Ru 至 Rd 的转发等价类 FEC F,当 Ru 有分组往 Rd 发送时，如果分组头的网络层头与 FEC F 相符，那么分组将被打上标签 L 后从 Ru 发往 Rd。在这里，标签 L 与转发等价类 F 进行绑定，L 是 Ru 的输出标签和 Rd 的输入标签。而标签只是在 Rd 和 Ru 之间有意义， Ru 称之为 Rd 的上游 LSR, Rd 为 Ru 的下游 LSR。

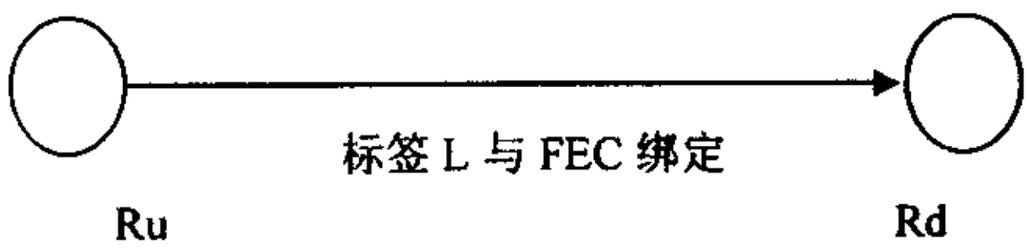

图 2.2 标签 L 与 FEC 绑定

2、标签分发

在 MPLS 网中，LSR 使用标签交换转发分组到另一个 LSR。在能使用两个标签之前，两个 LSR 间必须对标签的使用达成一致理解，这个过程称为标签分发。目前使用的标签分发方式是下游标签分发，它可分为两种

(1)“下游按需”标签分发

MPLS 结构允许 LSR 为某 FEC 从它的下游 LDP 对等体处显示地提出标签绑定申请。

1） 由 LDP（标记分布协议）和传统路由协议（OSPF 等）一起，在 LSR 中建立路由表和标记映射表。

2） LER 接收 IP 包，完成第三层功能，并给 IP 包加上标记；在 MPLS 出口的 LER 上，将分组中的标记去掉后继续进行转发。

3） LSR 对分组不再进行任何第三层处理，只是依据分组上的标记通过交换单元对其进行转发。

§2.2 标签分发协议（LDP）

1、标签的定义

标签是一个简短的，具有固定长度的，具有本地意义的标识符，它用与绑定转发等价类 FEC。将相同转发处理方式（目的地相同、使用的转发路径相同、具有相同的服务等级等）的分组归为一类，这就是转发等价类。分组到达 MPLS 网络入口时，被划分为不同的 FEC,根据分组所属 FEC，将适当的标签插入分组头中，然后在网络中按标签进行交换式转发。

如图所示，设 Ru 和 Rd 是标签交换路由器 LSR。标签值 L 表示从 Ru 至 Rd 的转发等价类 FEC F,当 Ru 有分组往 Rd 发送时，如果分组头的网络层头与 FEC F 相符，那么分组将被打上标签 L 后从 Ru 发往 Rd。在这里，标签 L 与转发等价类 F 进行绑定，L 是 Ru 的输出标签和 Rd 的输入标签。而标签只是在 Rd 和 Ru 之间有意义， Ru 称之为 Rd 的上游 LSR, Rd 为 Ru 的下游 LSR。

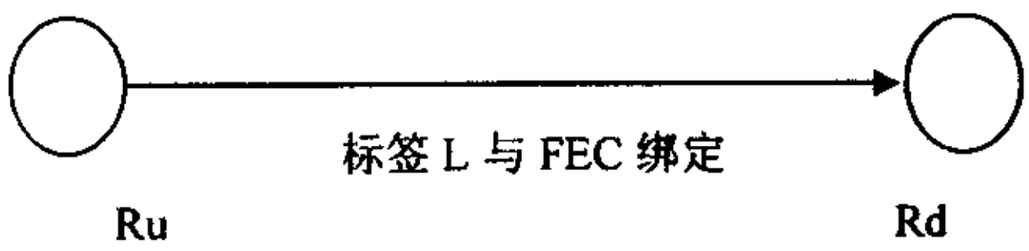

图 2.2 标签 L 与 FEC 绑定

2、标签分发

在 MPLS 网中，LSR 使用标签交换转发分组到另一个 LSR。在能使用两个标签之前，两个 LSR 间必须对标签的使用达成一致理解，这个过程称为标签分发。目前使用的标签分发方式是下游标签分发，它可分为两种

(1)“下游按需”标签分发

MPLS 结构允许 LSR 为某 FEC 从它的下游 LDP 对等体处显示地提出标签绑定申请。

(2)“下游自主”标签分发

MPLS 还允许下游的 LSR 主动告知上游的 LSR 有关的标签绑定信息，而不管上游的 LSR 是否向它提出标签请求。

这两种标签的分发方式可以共存于同一网络中，也可以使一部分网络采用按需分发，一部分采用自主分发，也可以在全网只使用一种分发方式。但对于某一对上、下游 LSR 则必须事先对采用何种标签分发方式达成共识。

3、信令方式的实现

目前 MPLS 实现信令的方式可分为两类，一类是 LDP / CR-LDP，另外一类是扩展的 RSVP，它们在协议特性上存在不同，有不同的消息集和信令处理规程。后面这两种信令方式进行详细的介绍

4、标签转换（Label Swap）

分组通过标签转换在 MPLS 域中进行转发。

标签转换要涉及到几个相关的概念

A. 下一跳标签选路入口（NHLFE）

NHLFE 中包含以下信息：

1）分组的下一跳

2）对标签堆栈应进行何种操作

3）发送分组时采用何种数据链路封装方式

4）发送分组时采用何种标签编码方式

如果收到的标签分组实际上并没包含有标签，那可能需要做 FEC 到 NHLFE 的映射，称之为 FTN(FEC-to-NHLFE MAP)。

B、输入标签映射（ILM:Incoming label Map）

标签转换：在转发有标签分组时，LSR 先检查标签堆栈的顶部标签，然后进行 ILM。接着根据 NHLFE 中的信息确定向何处转发分组，对标签堆栈进行相应的操作，然后将新形成的标签堆栈进行编码，完成转发。

在转发无标签分组时，LSR 先分析网络层包头确定分组属于哪个 FEC，然后进行 FTN。接着根据 NHLEF 中的信息确定向何处转发分组，对标签堆栈进行相应的操作，最后将新形成的标签堆栈进行编码，完成转发。

5、标签合并

来自不同输入端口但具有相同 FEC 映射的分组只使用一个标签，可减少标签的需求，提高网络扩展性。

6、标签交换路径 LSP

标签交换路径是指具有某特定 FEC 分组，在传输时经过的标签交换路由器集合构成的数据传输通路。由于 MPLS 支持层次化的网络拓扑结构，因此我们对某一分组传输路径进行描述时，还必须指明当前的标签交换路径位于第几层。

LSP 路由选择，MPLS 的路由选择是指一个 FEC 选择一条 LSP，MPLS 协议中支持：逐跳 LSP 路由和显示 LSP 路由

另外 MPLS 还支持多径路由。

7、LSP 建立的控制模式。

分组 FEC 建立一条 LSP 的方法上主要有两种模式：独立的 LSP 控制建立模式和有序的 LSP 控制建立模式。

在独立的 LSP 控制建立模式中，沿 LSP 上的每个 LSR 根据它所识别出的 FEC 来自己做出决定将某个标签绑定信息分发给它的 LDP 对等体。

在有序的 LSP 控制建立模式中，沿 LSP 的 LSR 除非它是某 FEC 的出口，否则它必须收到下游对等体分发给它的标签绑定后，才能对 FEC 做标签绑定。

8、环路控制

循环的预防、探测与避免是所有网络必须解决的问题之一。IETF MPLS 工作组提出了一种基于线程（thread）的 MPLS 环路控制方法。

当一个 LSR(标签交换路由器)发现其有一个新的特定的跳向 FEC（等效前传类）的下一跳时，它就创建一个线程并且将其扩展为下游。每一个这样的线程都被分配唯一的一种颜色来标识，这样就可保证网络上的任何两个线程都不会有相同的颜色。如果存在路由环，那么某一线程将会返回至它已经经过的 LSR 处。因为线程有特定的颜色，所以这一点很容易检测。

为了防止 LSP 路由环，通过使用线程："扩展"、"回绕"、"撤销"、"合并"、"滞留" 来定义线程的五个基本行为。

§2.3 扩展的 RSVP 和 CR-LDP 信令协议

2.3.1 扩展的 RSVP

RSVP 是为了改善 IP 网络的服务质量而设计的。RSVP 可以保证在路由器

6、标签交换路径 LSP

标签交换路径是指具有某特定 FEC 分组，在传输时经过的标签交换路由器集合构成的数据传输通路。由于 MPLS 支持层次化的网络拓扑结构，因此我们对某一分组传输路径进行描述时，还必须指明当前的标签交换路径位于第几层。

LSP 路由选择，MPLS 的路由选择是指一个 FEC 选择一条 LSP，MPLS 协议中支持：逐跳 LSP 路由和显示 LSP 路由

另外 MPLS 还支持多径路由。

7、LSP 建立的控制模式。

分组 FEC 建立一条 LSP 的方法上主要有两种模式：独立的 LSP 控制建立模式和有序的 LSP 控制建立模式。

在独立的 LSP 控制建立模式中，沿 LSP 上的每个 LSR 根据它所识别出的 FEC 来自己做出决定将某个标签绑定信息分发给它的 LDP 对等体。

在有序的 LSP 控制建立模式中，沿 LSP 的 LSR 除非它是某 FEC 的出口，否则它必须收到下游对等体分发给它的标签绑定后，才能对 FEC 做标签绑定。

8、环路控制

循环的预防、探测与避免是所有网络必须解决的问题之一。IETF MPLS 工作组提出了一种基于线程（thread）的 MPLS 环路控制方法。

当一个 LSR(标签交换路由器)发现其有一个新的特定的跳向 FEC（等效前传类）的下一跳时，它就创建一个线程并且将其扩展为下游。每一个这样的线程都被分配唯一的一种颜色来标识，这样就可保证网络上的任何两个线程都不会有相同的颜色。如果存在路由环，那么某一线程将会返回至它已经经过的 LSR 处。因为线程有特定的颜色，所以这一点很容易检测。

为了防止 LSP 路由环，通过使用线程："扩展"、"回绕"、"撤销"、"合并"、"滞留" 来定义线程的五个基本行为。

§2.3 扩展的 RSVP 和 CR-LDP 信令协议

2.3.1 扩展的 RSVP

RSVP 是为了改善 IP 网络的服务质量而设计的。RSVP 可以保证在路由器

6、标签交换路径 LSP

标签交换路径是指具有某特定 FEC 分组，在传输时经过的标签交换路由器集合构成的数据传输通路。由于 MPLS 支持层次化的网络拓扑结构，因此我们对某一分组传输路径进行描述时，还必须指明当前的标签交换路径位于第几层。

LSP 路由选择，MPLS 的路由选择是指一个 FEC 选择一条 LSP，MPLS 协议中支持：逐跳 LSP 路由和显示 LSP 路由

另外 MPLS 还支持多径路由。

7、LSP 建立的控制模式。

分组 FEC 建立一条 LSP 的方法上主要有两种模式：独立的 LSP 控制建立模式和有序的 LSP 控制建立模式。

在独立的 LSP 控制建立模式中，沿 LSP 上的每个 LSR 根据它所识别出的 FEC 来自己做出决定将某个标签绑定信息分发给它的 LDP 对等体。

在有序的 LSP 控制建立模式中，沿 LSP 的 LSR 除非它是某 FEC 的出口，否则它必须收到下游对等体分发给它的标签绑定后，才能对 FEC 做标签绑定。

8、环路控制

循环的预防、探测与避免是所有网络必须解决的问题之一。IETF MPLS 工作组提出了一种基于线程（thread）的 MPLS 环路控制方法。

当一个 LSR(标签交换路由器)发现其有一个新的特定的跳向 FEC（等效前传类）的下一跳时，它就创建一个线程并且将其扩展为下游。每一个这样的线程都被分配唯一的一种颜色来标识，这样就可保证网络上的任何两个线程都不会有相同的颜色。如果存在路由环，那么某一线程将会返回至它已经经过的 LSR 处。因为线程有特定的颜色，所以这一点很容易检测。

为了防止 LSP 路由环，通过使用线程："扩展"、"回绕"、"撤销"、"合并"、"滞留" 来定义线程的五个基本行为。

§2.3 扩展的 RSVP 和 CR-LDP 信令协议

2.3.1 扩展的 RSVP

RSVP 是为了改善 IP 网络的服务质量而设计的。RSVP 可以保证在路由器

无连接的情况下通过软状态方式实现 IP 网的资源预留。传统的 RSVP 存在的主要问题是其稳定性和可扩展性，这主要是由于 RSVP 是一个软状态协议，需要定时刷新而造成的，扩展 RSVP 针对这些问题进行一些修改和扩展。扩展 RSVP 的基本原理是对于有相同入口和出口 LSP 的流量中继(Traffic Trunk),MPLS 区域可为它们分配相同的标记，这样具有相同标记的业务流视为一组，而传统的 RSVP 要为每一个业务流都分别建立一套状态，这样可以大大节省路由器中的 RSVP 信息量。分组后的流量中继对于网络中间节点是透明传输的，即相同标记的流量中继统一使用由该标记决定的 LSP 与沿途所享受的网络资源与处理方式。扩展 RSVP 中使用的主要扩展技术有：消息合并技术、消息标识技术、摘要刷新技术和 HELLO 协议扩展技术等。

由于 RSVP 本身就是为了改善 IP 网络的服务质量而设计，扩展 RSVP 针对 RSVP 在 MPLS 中的应用扩展了某些功能，支持显示路由的建立与管理，对业务流分组而非每一个业务流都建立一套状态，节省了信息量，提高了网络的可扩展性，针对软状态特征所带来了问题引入许多扩展技术，降低时延和开销，减少了刷新信息的数量，扩展 RSVP 所建立的路由是基于各种约束条件的显示路由而非传统的面向目的地的路由。

扩展 RSVP 可以实现流量工程技术的全部要求，而且 RSVP 技术经过多年的完善与应用实践，是一种比较成熟的技术，但由于其存在的问题如软状态、资源消耗大、比较复杂、扩展性不好等使得扩展 RSVP 的应用受到了限制。

2.3.2 CR-LDP

信令协议从某种意义来说，CR-LDP 信令协议是基于现有标准的 LDP 信令加上显式路由，用于建立和维护基于约束的 LSP (Contrain-Based LDP)来保证 IP QoS 业务。CR-LDP 通过一套简单有效的硬状态控制与消息机制灵活地预留网络资源。CR-LDP 允许网管人员配置 QoS 的级别，并规范了与现有 ATM 交换机的业务等级和 QoS 之间的灵活映射。为了符合面向全网的流量工程要求，CR-LDP 采用约束路由机制，可提供严格、松散的显式路由；建立、保持优先级，路径挤占、重新优化路径等多种功能。由于信令基于可靠的 TCP 传输机制，因而 CR-LDP 可以确保快速响应节点故障，保证信息的可靠传输。CR-LDP 支持多种网络层协议，它利用 LSR 可传输任何特定类型的

无连接的情况下通过软状态方式实现 IP 网的资源预留。传统的 RSVP 存在的主要问题是其稳定性和可扩展性，这主要是由于 RSVP 是一个软状态协议，需要定时刷新而造成的，扩展 RSVP 针对这些问题进行一些修改和扩展。扩展 RSVP 的基本原理是对于有相同入口和出口 LSP 的流量中继(Traffic Trunk),MPLS 区域可为它们分配相同的标记，这样具有相同标记的业务流视为一组，而传统的 RSVP 要为每一个业务流都分别建立一套状态，这样可以大大节省路由器中的 RSVP 信息量。分组后的流量中继对于网络中间节点是透明传输的，即相同标记的流量中继统一使用由该标记决定的 LSP 与沿途所享受的网络资源与处理方式。扩展 RSVP 中使用的主要扩展技术有：消息合并技术、消息标识技术、摘要刷新技术和 HELLO 协议扩展技术等。

由于 RSVP 本身就是为了改善 IP 网络的服务质量而设计，扩展 RSVP 针对 RSVP 在 MPLS 中的应用扩展了某些功能，支持显示路由的建立与管理，对业务流分组而非每一个业务流都建立一套状态，节省了信息量，提高了网络的可扩展性，针对软状态特征所带来了问题引入许多扩展技术，降低时延和开销，减少了刷新信息的数量，扩展 RSVP 所建立的路由是基于各种约束条件的显示路由而非传统的面向目的地的路由。

扩展 RSVP 可以实现流量工程技术的全部要求，而且 RSVP 技术经过多年的完善与应用实践，是一种比较成熟的技术，但由于其存在的问题如软状态、资源消耗大、比较复杂、扩展性不好等使得扩展 RSVP 的应用受到了限制。

2.3.2 CR-LDP

信令协议从某种意义来说，CR-LDP 信令协议是基于现有标准的 LDP 信令加上显式路由，用于建立和维护基于约束的 LSP (Contrain-Based LDP)来保证 IP QoS 业务。CR-LDP 通过一套简单有效的硬状态控制与消息机制灵活地预留网络资源。CR-LDP 允许网管人员配置 QoS 的级别，并规范了与现有 ATM 交换机的业务等级和 QoS 之间的灵活映射。为了符合面向全网的流量工程要求，CR-LDP 采用约束路由机制，可提供严格、松散的显式路由；建立、保持优先级，路径挤占、重新优化路径等多种功能。由于信令基于可靠的 TCP 传输机制，因而 CR-LDP 可以确保快速响应节点故障，保证信息的可靠传输。CR-LDP 支持多种网络层协议，它利用 LSR 可传输任何特定类型的

业务流，而不要求 LSR 识别所传输的业务类型，从而有效地保证了传输报文的安全性。

CR-LDP 的工作原理基本与标准 LDP 一致，如对等体的发现，会话的建立与维护，标签分发.管理和故障处理等都沿用 LDP 的机制。因此 LDP/CR-LDP 实际上是一个统一的信令系统，它们一起为 MPLS 网络管理员提供了完整的标签分发和通路建立的操作模式。CR-LDP 只是在 LDP 的基础上增加了基于限制的选路，在 CR-LDP 中仍然沿用 LDP 中如下机制

*基本和/或扩展的发现机制；

*有序控制、下游按需标签分发模式中使用的标签请求消息；

*有序控制、下游按需标签分发模式中使用的标签映射消息；

*通知消息；

*标签撤销和释放消息；

*环路检测机制。

在 CR-LDP 中为了支持基于限制的选路，增加了许多其他的机制，如严格、松散的显式路由、业务特性描述、路由锁定、抢占、资源分类等。CR-LDP 正是通过这些机制使 MPLS 支持强大的流量工程能力。

1. 严格/松散显式路由

CR-LDP 主要通过显式路由 (ER/ Explicit Route)支持流量工程。显式路由是基于限制路由的子集，该限制即为显式路由，即 LSP 要经过的节点或节点组的列表，在形成时要求使之满足其它“限制”，进而使最终建立的显示路由 LSP(ER-LSP/Explicit Route LSP)可以满足流量工程的要求。

建立 ER-LSP 的原因可能是想要合理地使用网络资源，或给 LSP 分配一定的带宽以及其它业务等级特征，或者想要确保备用路由使用物理分离的路径。CR-LSP 建立之后，LSP 可能经过该列表中的所有的节点或其中的一个子集，沿途要执行的某些操作也可以包含在显式路由中。显式路由中不但可以包含特定节点，而且也可以包含节点组，这些节点组往往被称为抽象节点。抽象节点使系统在实现显式路由时有较大的灵活性。基于限制路由在基于限制路由 TLV(Type,L ength,Value)中的编码是一系列显式路由跳，在此 TLV 中各个显式路由跳出现的顺序就是各个节点或节点组在 ER-LSP 中出现的顺序。显式路由跳由显式路由跳 TLV 表征，如图 3 所示。

入口(b1, b2, b3)	c	d1, d2	出口 e

图 2.3 显式路由 TLV

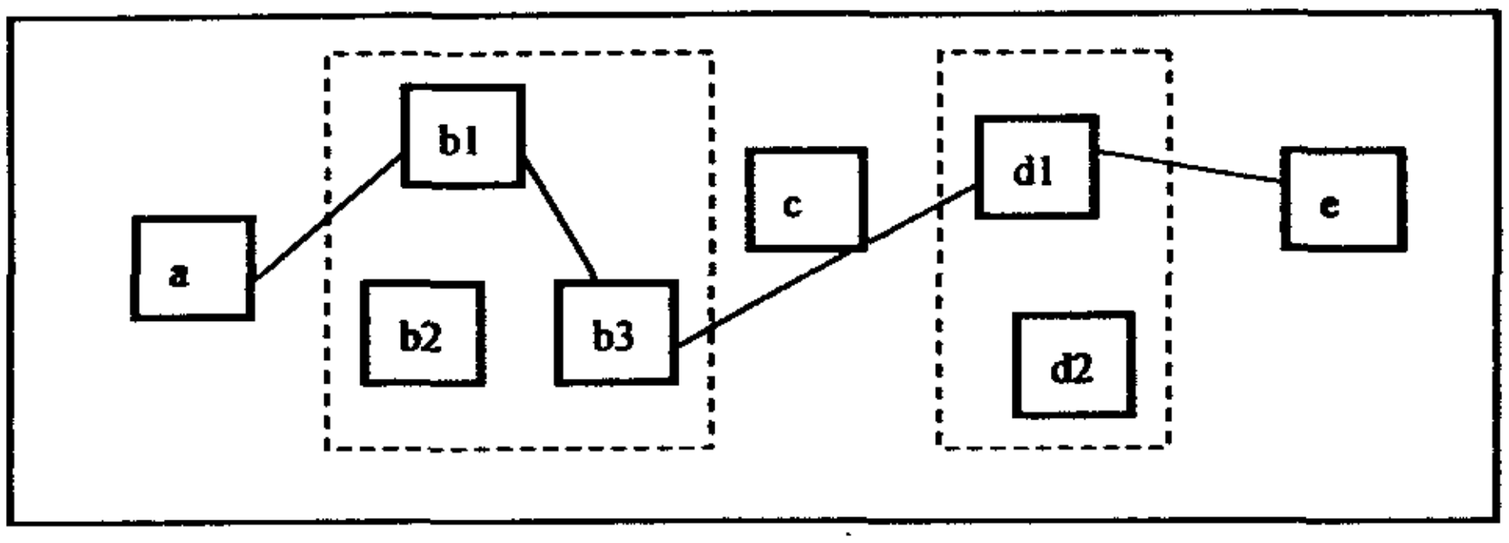

图 2.4 严格显式路由 LSP

图 2.4 所示为最终建立起来的严格路由的显式路由 LSP。CR-LDP 通过非严格路由又进一步赋予了 ER-LSP 更大的灵活性，在非严格显式路由所形成的 LSP 中的各个 LSR 不一定出现在显式路由 TLV 中，如图 2.5 所示，在第 4 跳中并没有选取抽象节点中的 d1、d2，而是选取了 f，f 并不包含在节点组 d 中。

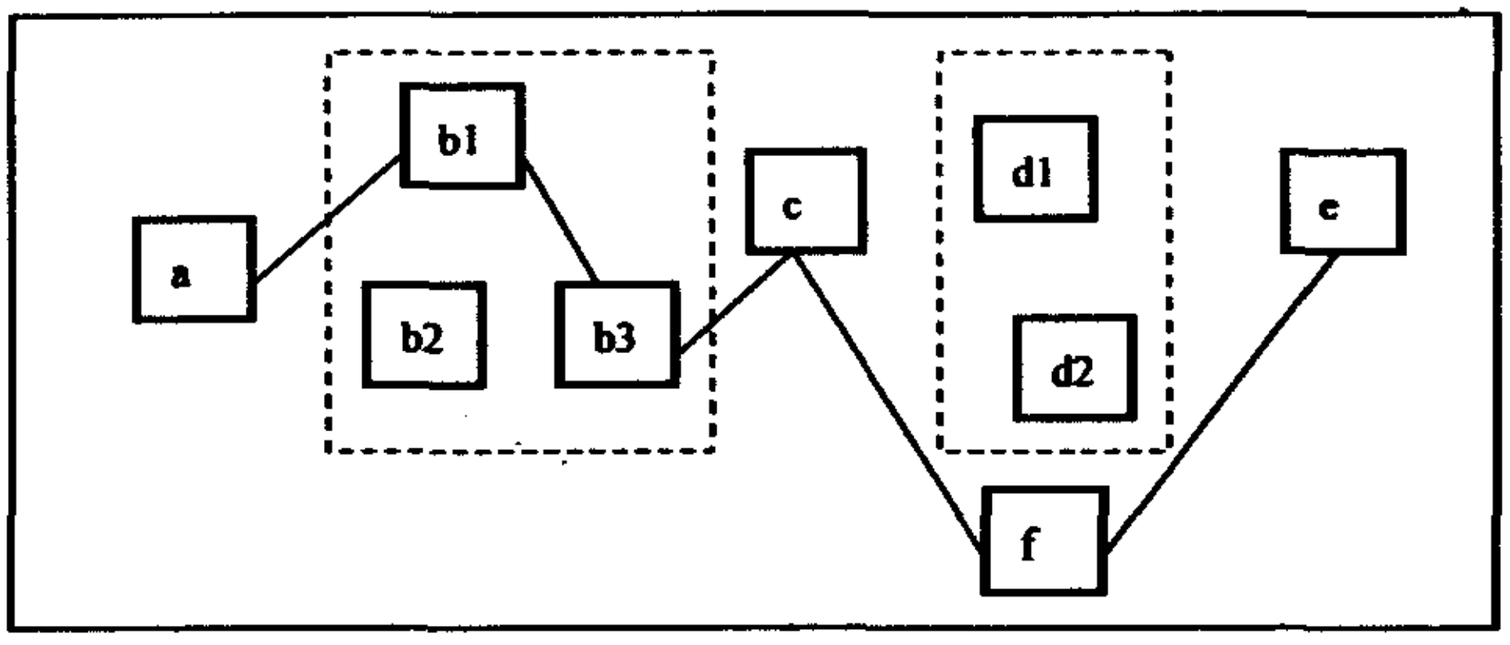

图 2.5 非严格显式路由 LSP

2. 流量特征描述

在 MPLS 域中，为了向 LSR 反映业务的“限制”，使 LSP 可以了解业务的特性，如峰值速率、协定速率和业务粒度等，从而使沿途的 LSR 可以为之预留资源，或者因资源不足及时地拒绝业务请求，在 CR-LDP 中增加了流量参数 TLV (Traffic Parameters TLV)对此进行描述。每条 CR-LDP 的流量特征通过流量参数 TLV 来描述。目前，流量参数 TLV 可以包含以下 7 种流

量参数：

频率(Frequency)：表明各种业务参数应能够保持的时长。

权重(Weight)：表明某一LSP在使用超出CDR以外带宽资源时的优先级。

PDR(峰值速率)：表示流量中继的最高速率。

PBS（峰值突发长度)：表示流量中继的最大突发分组长度。

CDR（协定速率)：表示LSP应当能够支持的速率。

CBS（协定突发长度)：表示LSP应当能够支持的最大分组长度。

EBS度（超标突发长度)：用于MPLS域边界上的流量调节，可以用于测定一个CR-LSP上发送的流量超过协定速率的程度。

3.路由锁定

路由锁定可用于非严格路由的LSP的段上，也就是那些用L比特置1的下一跳所指定的段或下一跳为抽象节点的段。当网络拓扑发生变化的时候，如果CR-LSP不希望改变自己使用的路径，那么可以锁定该路径，即使该LSP的非严格路由部分中的某些LSR有更好的下一跳时也不会影响该LSP所使用的路径。相应的在CR-LDP中增加了路由锁定TLV(Route Pinning TLV)完成对CR-LSP中非严格路由段的路径锁定。

4.抢占

CR-LDP在建立CR-LSP的过程中要告知路径上的每一跳特定业务所需的资源。如果没有足够的资源建立所需业务通道，那么就可为现存的路径重新选路，给新路径重新分配资源，这就是所说的“路径抢占”。在CR-LDP中，使用建立优先级为新路径分级；用保持优先级为现存路径分级。在抢占中，通过使用这两个优先级，确定新路径是否可以抢占现有路径。用信令告知一个较高的建立优先级，表示在资源不可用的情况下，路径抢占其它路径的可能性较大；同样用信令告知一个较高的保持优先级，表示一旦路径建立，它被抢占的机会就应该较小。在实际应用中要根据特定的网络策略确定应使用的抢占规则，即给路径分配建立优先级和保持优先级。在CR-LDP中使用抢占TLV(Pre-emption TLV)为相应的业务指定建立优先级和保持优先级。

CR-LSP的建立优先级不应该高于它的保持优先级。这是因为当CR-LSP的建立优先级高于保持优先级的时候，一个已经建立的LSP可能被某个等价LSP请求抢占。为了避免这种不合理的抢占给网络带来的不必要的负担，故建立优先级应小于保持优先级。这样当一个LSP建立后，等价LSP的建

立请求就不再会抢占已经建立的 LSP 了。当一个已经建立的 LSP 被抢占时，发起该抢占的 LSR 向上游发送一个撤销消息并向下游发送一个释放消息。当一个正在建立的 LSP(没有收到返回的标签映射)被抢占时，发起该抢占的 LSR 向上游发送一个通知消息并向下游发送一个中止消息。

5. 资源分类

网络管理者可以按照某种原则对网络资源进行分类，形象的可以理解为按照某种原则为网络中的各种资源进行染色，具有相同属性的资源将使用相同的颜色。对资源分类后就可按资源的不同类型进行策略调度。在 CR-LDP 中定义了资源分类 TLV(Resource Class(Color)TLV)，指定 CR-LSP 可以使用哪类资源。

6 .对 LDP 的消息的扩展

CR-LDP 在增加了上述多种 TLV 以支持流量工程的同时，对 LDP 使用的标签请求消息和标签映射也进一步作了扩展，使之可以有选择地携带各种新的 TLV,如：显式路由 TLV、流量参数 TLV、路由锁定 TLV、抢占 TLV、资源分类 TLV、以及 LSP ID TLV,进而使 CR-LSP 的建立过程尽可能满足流量工程的要求。其中 LSP ID TLV 是一个 CR-LSP 在 MPL S 域中的唯一标识。同样，CR-LDP 对标签映射消息也作了扩展，使之可以有选择的携带 LSP ID TLV 和流量参数 TLV,如图 2.6 所示

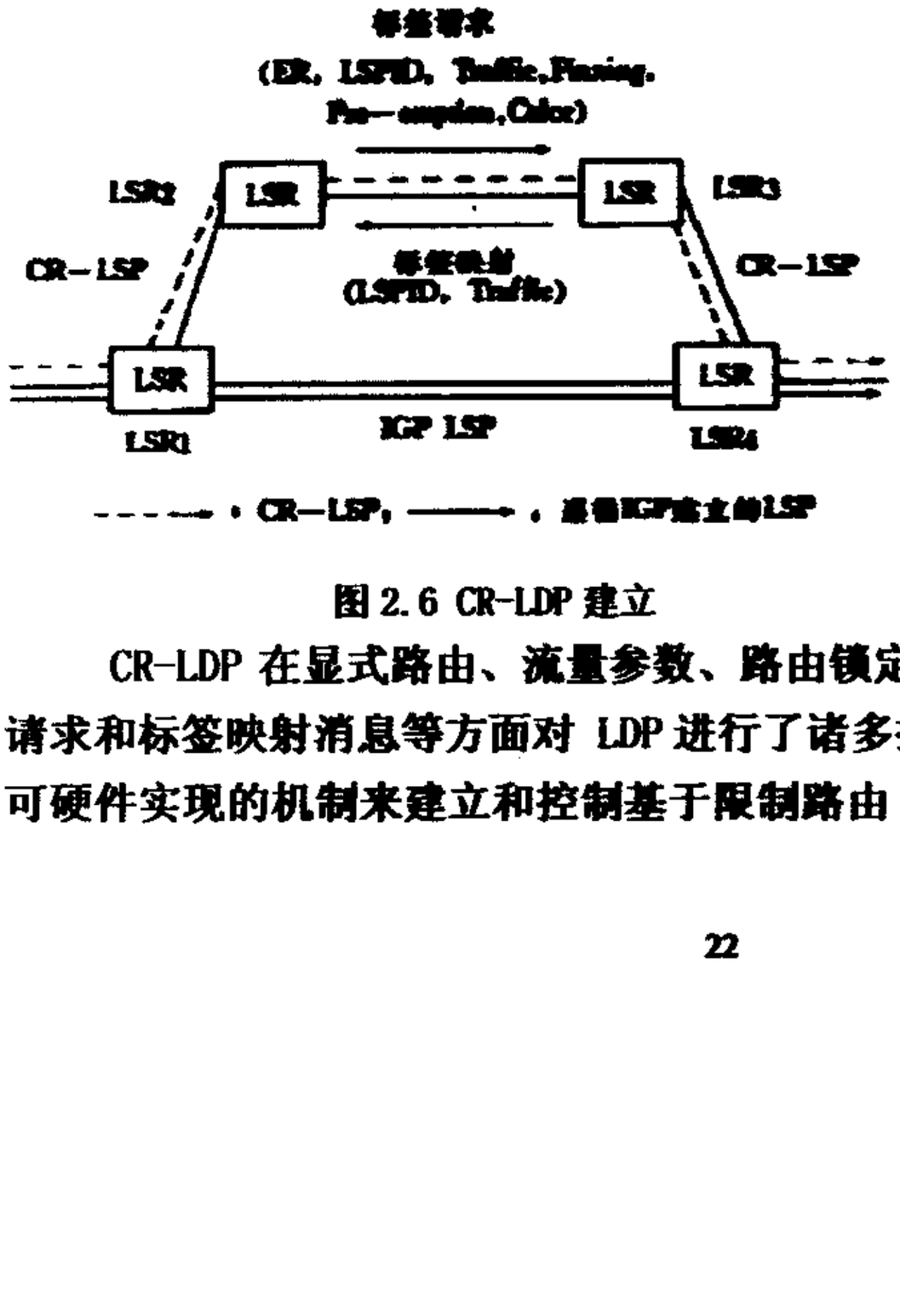

图 2.6 CR-LDP 建立

CR-LDP 在显式路由、流量参数、路由锁定、抢占、资源分类以及标签请求和标签映射消息等方面对 LDP 进行了诸多扩展，从而提供了一种简单、可硬件实现的机制来建立和控制基于限制路由 LSP,进而有效地支持了流量

工程。因此许多原电信厂商，如 LUCENT、NORTEL 都支持 CR-LDP。然而 IETF 的 MPLS 工作组同时还提出了另一种机制以支持流量工程，即扩展的资源预留协议 (RSVP-TE)。CR-LDP 与 RSVP-TE 相比在技术上有优势，但是 RSVP 作为一种出现较早的 IP 网络协议，许多大的路由器厂商，如 Cisco、JUNIPER 支持 RSVP。

从协议可靠性上来看，LDP / CR-LDP 是基于 TCP 的，当发生传输丢包时，利用 TCP 协议提供简单的错误指示，实现快速响应和恢复。而 RSVP 只是传送 IP 包。由于缺乏可靠的传输机制，RSVP 无法保证快速的失败通知。从网络可扩展性上看，LDP 较 RSVP 更有优势，一般电信级网络中，特别是 ATM 网络中，应采用 MPLS / LDP。ITU-T 倾向于在骨干网中采用 CR-LDP。从长远来看，CR-LDP 将因其技术优势而居于主导地位，有着广阔的应用前景。

第三章　服务质量 QoS

§3.1 介绍

随着高速网络和多媒体技术的飞速发展，电子商务、IP 语音、图像传送、视频传输等业务纷至沓来，这些新业务对网络的实时特性等有了更多的要求，它们的发展很大程度上取决于网络的有关服务指标，如延迟是否过大、画面是否抖动、声音和图像是否同步等，这也就是所谓的服务质量 QoS(Quality of Service)。按照国际电联 E.800 的建议，QoS 是业务性能的总体效果，它决定了用户对特定业务的满意程度，在这里 QoS 实质上是指 IP 包在一个或多个网络传输的过程中所表现的各种性能。

服务质量(Qos)是指在网络连接中由业务所决定的性能度量，如带宽、时延或连接中传送数据所允许的消息丢失率。Qos 需要通过新的连接发起时的接入控制机制获得保证。QoS 是指 IP 包在一个或多个网络中传输的过程中所表现的各种性能，它是对各种性能参数的具体描述。这些性能参数包括：业务可靠性、延迟、抖动、吞吐量和包丢失率。

IP QoS 的最终目的是要为各种业务（包括数据，图像，多媒体与语音业务等）提供可靠的端到端的服务质量保证。而实现这一目标的前提是要对各种 QoS 参数进行清楚的定义：

a. 业务可靠性：指用户与 Internet 业务的连接的可靠性。这包括，建立时间，保持时间等。

b. 时延：也称延迟，指在两个参考点间，某一 IP 包从发送到接收之间的时间间隔。

c. 时延抖动：是指沿同一路径传输的一个数据流中不同分组传输时延的变化。

d. 吞吐量：是指一个网络中 IP 包的传输速率，这一参数可以用平均速率或峰值速率来表征。

e. 包丢失率：是指某一业务在网络中传输时，可允许的最大丢包率。丢包主要是由网络拥塞引起的。

其中用户感知到的服务质量，传输的时延。时延主要从以下方面对不同服务产生不同的影响：端到端时延；时延变化(即抖动)。交互式的实时业务(如

第三章　服务质量 QoS

§3.1 介绍

随着高速网络和多媒体技术的飞速发展，电子商务、IP 语音、图像传送、视频传输等业务纷至沓来，这些新业务对网络的实时特性等有了更多的要求，它们的发展很大程度上取决于网络的有关服务指标，如延迟是否过大、画面是否抖动、声音和图像是否同步等，这也就是所谓的服务质量 QoS(Quality of Service)。按照国际电联 E.800 的建议，QoS 是业务性能的总体效果，它决定了用户对特定业务的满意程度，在这里 QoS 实质上是指 IP 包在一个或多个网络传输的过程中所表现的各种性能。

服务质量(Qos)是指在网络连接中由业务所决定的性能度量，如带宽、时延或连接中传送数据所允许的消息丢失率。Qos 需要通过新的连接发起时的接入控制机制获得保证。QoS 是指 IP 包在一个或多个网络中传输的过程中所表现的各种性能，它是对各种性能参数的具体描述。这些性能参数包括：业务可靠性、延迟、抖动、吞吐量和包丢失率。

IP QoS 的最终目的是要为各种业务（包括数据，图像，多媒体与语音业务等）提供可靠的端到端的服务质量保证。而实现这一目标的前提是要对各种 QoS 参数进行清楚的定义：

a. 业务可靠性：指用户与 Internet 业务的连接的可靠性。这包括，建立时间，保持时间等。

b. 时延：也称延迟，指在两个参考点间，某一 IP 包从发送到接收之间的时间间隔。

c. 时延抖动：是指沿同一路径传输的一个数据流中不同分组传输时延的变化。

d. 吞吐量：是指一个网络中 IP 包的传输速率，这一参数可以用平均速率或峰值速率来表征。

e. 包丢失率：是指某一业务在网络中传输时，可允许的最大丢包率。丢包主要是由网络拥塞引起的。

其中用户感知到的服务质量，传输的时延。时延主要从以下方面对不同服务产生不同的影响：端到端时延；时延变化(即抖动)。交互式的实时业务(如

语音通信等)对端到瑞时延和抖动很敏感。非交互式的实时业务(如单向广播等)对瑞到端时延不敏感，但对抖动敏感。非实时业务往往对时延不敏感，但由于这些应用可能采用时延指标来控制其流量速率(如 TCP)。或在应用得到响应前需要对数据进行缓存(如 FTP)，所以数值或变动大的时延网络传送数据的吞吐量。吞吐量决定业务可以在网络上传输的速率。吞吐量取决于以下因素，链路特性：带宽、误码率；节点特性：缓冲区容量、处理机能力。

QoS 主要通过以下方式实现：

连接接纳控制：在给定的全部资源中，网络需要控制接入用户的数量，使用户所需资源不超出网络的总资源，保证对已连接用户数据流提供优先服务。

资源预留：对给定数据流的业务特点和 QoS 要求，网络为其预留一定的网络资源（带宽、缓冲区等）。

资源分配：对已连接的数据流，怎么去保证它们公平的分配资源。

目前，网络中实现基于 IP 协议的 QoS 问题，IETF 提出了两种基本解决方案，即综合服务模型(Intserv)方案和区分服务模型(Diffserv)方案。

§3.2 综合服务模型

3.2.1 综合服务模型

为保证应用的 QoS, IETF 在 1994 年提出了综合服务模型。综合服务模型(Intserv)以标准的 RSVP 协议作为实现机制。通过 Intserv 将可以实现 IP 网中的 QoS 传输以及对于实时业务的支持，使得各种应用能够为其数据包选择服务等级。

综合服务是建立在流 (flow)的概念上。所谓流是指源于某一用户的特定行为的一串彼此相关的 IP 数据报 ,这些数据报具有相同的 QoS 要求 ,且可能有多个接收者。“流”的引入 ,使得一条流可以被理解为一条逻辑上的 IP 连接。

该模型的原理是对于每一个需要进行 QoS 处理的数据流，通过一定的信令机制，在其经由的每一个路由器上进行资源预留，以便实现端到端的 QoS 业务。它引入了资源预留协议 RSVP，从而保证了沿着该通道传输的数据流能够满足 QoS 要求。但为了支持这种能力,数据包所经过的每个网络元素

语音通信等)对端到瑞时延和抖动很敏感。非交互式的实时业务(如单向广播等)对瑞到端时延不敏感，但对抖动敏感。非实时业务往往对时延不敏感，但由于这些应用可能采用时延指标来控制其流量速率(如 TCP)。或在应用得到响应前需要对数据进行缓存(如 FTP)，所以数值或变动大的时延网络传送数据的吞吐量。吞吐量决定业务可以在网络上传输的速率。吞吐量取决于以下因素，链路特性：带宽、误码率；节点特性：缓冲区容量、处理机能力。

QoS 主要通过以下方式实现：

连接接纳控制：在给定的全部资源中，网络需要控制接入用户的数量，使用户所需资源不超出网络的总资源，保证对已连接用户数据流提供优先服务。

资源预留：对给定数据流的业务特点和 QoS 要求，网络为其预留一定的网络资源（带宽、缓冲区等）。

资源分配：对已连接的数据流，怎么去保证它们公平的分配资源。

目前，网络中实现基于 IP 协议的 QoS 问题，IETF 提出了两种基本解决方案，即综合服务模型(Intserv)方案和区分服务模型(Diffserv)方案。

§3.2 综合服务模型

3.2.1 综合服务模型

为保证应用的 QoS, IETF 在 1994 年提出了综合服务模型。综合服务模型(Intserv)以标准的 RSVP 协议作为实现机制。通过 Intserv 将可以实现 IP 网中的 QoS 传输以及对于实时业务的支持，使得各种应用能够为其数据包选择服务等级。

综合服务是建立在流 (flow)的概念上。所谓流是指源于某一用户的特定行为的一串彼此相关的 IP 数据报 ,这些数据报具有相同的 QoS 要求 ,且可能有多个接收者。“流”的引入 ,使得一条流可以被理解为一条逻辑上的 IP 连接。

该模型的原理是对于每一个需要进行 QoS 处理的数据流，通过一定的信令机制，在其经由的每一个路由器上进行资源预留，以便实现端到端的 QoS 业务。它引入了资源预留协议 RSVP，从而保证了沿着该通道传输的数据流能够满足 QoS 要求。但为了支持这种能力,数据包所经过的每个网络元素

语音通信等)对端到瑞时延和抖动很敏感。非交互式的实时业务(如单向广播等)对瑞到端时延不敏感，但对抖动敏感。非实时业务往往对时延不敏感，但由于这些应用可能采用时延指标来控制其流量速率(如 TCP)。或在应用得到响应前需要对数据进行缓存(如 FTP)，所以数值或变动大的时延网络传送数据的吞吐量。吞吐量决定业务可以在网络上传输的速率。吞吐量取决于以下因素，链路特性：带宽、误码率；节点特性：缓冲区容量、处理机能力。

QoS 主要通过以下方式实现：

连接接纳控制：在给定的全部资源中，网络需要控制接入用户的数量，使用户所需资源不超出网络的总资源，保证对已连接用户数据流提供优先服务。

资源预留：对给定数据流的业务特点和 QoS 要求，网络为其预留一定的网络资源（带宽、缓冲区等）。

资源分配：对已连接的数据流，怎么去保证它们公平的分配资源。

目前，网络中实现基于 IP 协议的 QoS 问题，IETF 提出了两种基本解决方案，即综合服务模型(Intserv)方案和区分服务模型(Diffserv)方案。

§3.2 综合服务模型

3.2.1 综合服务模型

为保证应用的 QoS, IETF 在 1994 年提出了综合服务模型。综合服务模型(Intserv)以标准的 RSVP 协议作为实现机制。通过 Intserv 将可以实现 IP 网中的 QoS 传输以及对于实时业务的支持，使得各种应用能够为其数据包选择服务等级。

综合服务是建立在流 (flow)的概念上。所谓流是指源于某一用户的特定行为的一串彼此相关的 IP 数据报 ,这些数据报具有相同的 QoS 要求 ,且可能有多个接收者。“流”的引入 ,使得一条流可以被理解为一条逻辑上的 IP 连接。

该模型的原理是对于每一个需要进行 QoS 处理的数据流，通过一定的信令机制，在其经由的每一个路由器上进行资源预留，以便实现端到端的 QoS 业务。它引入了资源预留协议 RSVP，从而保证了沿着该通道传输的数据流能够满足 QoS 要求。但为了支持这种能力,数据包所经过的每个网络元素

(子网和 IP 路由器)都必须能够支持 RSVP。

目前，Intserv 模型定义了三种业务类型：

a. 负载可控服务 (Controlled Load Service)

负载可控服务能保证在网络负载较重时提供与负载较轻时相同的 QoS。在轻载网络中这种业务类似于 Best-effort 业务。它与传统的因特网服务的主要区别在于它的性能不会随网络负载的加大而下降。它能够提供最小的传输时延，对于排队算法没有特别的要求。在控制负载业务网络中，应用可以假设网络传输的包差错率近似于下层传输媒质的基本包差错率；包平均传输延迟与网络绝对延迟（包括光传输延迟加路由器转发延迟）差别不大。

b. 保证服务 (Guaranteed Service)

保证服务要求提供一定的带宽和端到端延迟，且保证数据流中合法的数据包无排队丢失。该业务将提供时延、带宽与丢包率等参数的保证。该业务不能控制固定延迟（传输延迟等，它们取决于由连接建立机制所选的路由），它所能保证的是排队延迟的大小（排队延迟是令牌桶大小和数据速率的函数）。网络使用加权公平排队（WFQ）算法。

路由器将保证提供的服务抽象成分配一定的带宽 R 和缓冲区 B。服从漏桶 (b，r)的数据流有 b/R(R>=r)的延迟上限 (其中 b 为漏桶的容量，r 为令牌的生成速率)。考虑到路由器是按数据包而非纯比特流传输数据 ,是通过共享链路而非每个流有一条物理通路，因而会引入一些误差。 RFC2 2 1 2 定义两个误差参数 C 和 D。C 是与速率有关的误差, D 是与延时有关的误差。则延迟上限为 b/R + C/R + D，如果再考虑峰值速率和数据包的长度 ,那么端到端延迟上限为：

$$D=\begin{cases}\dfrac{(b-M)(p-R)}{R(p-r)}+\dfrac{M+C_{tot}}{R}+D_{tot} & p>R>=r \quad (1)\\ \dfrac{M+C_{tot}}{R}+D_{tot} & R>=p>=r \quad (2)\end{cases}$$

其中，Ctot 和 Dtot 表示该路径上各路由器的 C 和 D 的和。在 p>R>=r 的情况下，延迟由三部分组成：

1. 容量为 b 的漏桶以峰值速率 p 发送数据，以 R 速率输出而造成的延迟。
2. 每一跳的 C 造成的误差的累计。

3. 每一跳的 D 误差累计。当 R>=p>=r 时，由于输出链路的速率大于发送速率，因而无第一项误差。

c. 尽力而为型业务（Best-effort Service）：

实际就是传统的 Internet 所提供的业务，是一种尽力而为的工作方式，基本不提供任何 QoS 保证。

为了实现上述服务，IntServ 定义了四个功能部件：RSVP、访问控制(Adminsion Control)、分类器(Classifier)、队列调度器(Scheduler)。

3.2.2 RSVP

RSVP 是一种基于接收端，由接收端发起的资源预留协议。不同的接收端对 QoS 要求可能不同，由它向发送端指明所希望接收的数据流的 QoS 参数。在通信双方已经建立了的路径上，通过源端发出的 PATH 消息和接收端发出的 RESV 消息进行动态的 QoS 协商，达到资源预留的目的。

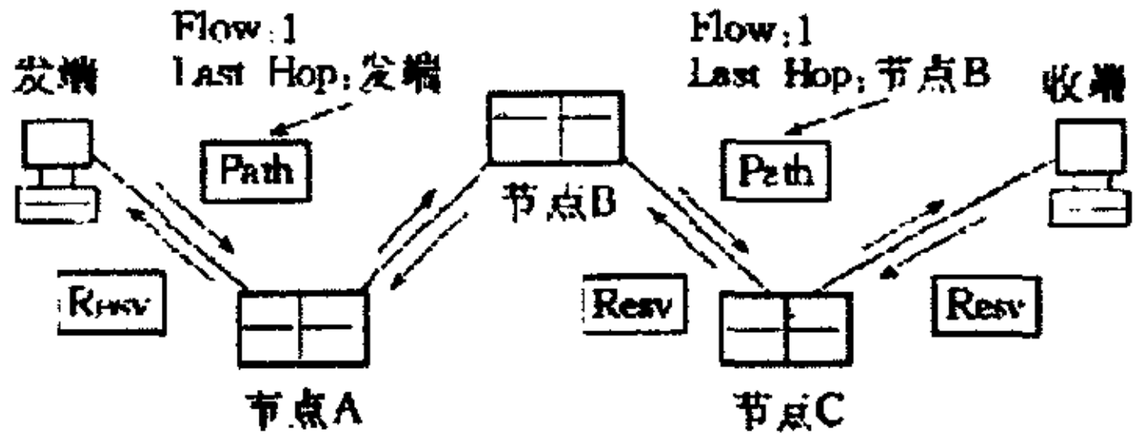

图 3.1 RSVP 协议的控制消息

图 3.1 说明了预留管道是如何工作的。沿着数据路径，每个 RSVP 发送主机通过路由协议所提供的单点传送或多址通信的路由下传 RSVP 的路径信息，这些路径信息存贮了传送中每个节点的路径状态。路径状态至少包括以前路由段节点的单点传送 IP 地址，这个 IP 地址过去常常按确定路线一段一段地反向传递预留请求(Resv)信息。每个接收主机向发送方上传 RSVP 的 Resv 信息，这些信息在 路径上的每个节点产生并保持了预留状态。Resv 信息自身最终也必须被传送到发送主机，因此主机可为第一个路由段设定适当的传输控制参数。与传送数据相同，路径信息也用相同的源和目的 单元地址来传送。另一方面，Resv 信息一段一段进行传送，每个 RSVP 节点向前面 RSVP 路由 段的单点传送地址转送一个 Resv 信息。

3. 每一跳的 D 误差累计。当 R>=p>=r 时，由于输出链路的速率大于发送速率，因而无第一项误差。

c. 尽力而为型业务（Best-effort Service）：

实际就是传统的 Internet 所提供的业务，是一种尽力而为的工作方式，基本不提供任何 QoS 保证。

为了实现上述服务，IntServ 定义了四个功能部件：RSVP、访问控制(Adminsion Control)、分类器(Classifier)、队列调度器(Scheduler)。

3.2.2 RSVP

RSVP 是一种基于接收端，由接收端发起的资源预留协议。不同的接收端对 QoS 要求可能不同，由它向发送端指明所希望接收的数据流的 QoS 参数。在通信双方已经建立了的路径上，通过源端发出的 PATH 消息和接收端发出的 RESV 消息进行动态的 QoS 协商，达到资源预留的目的。

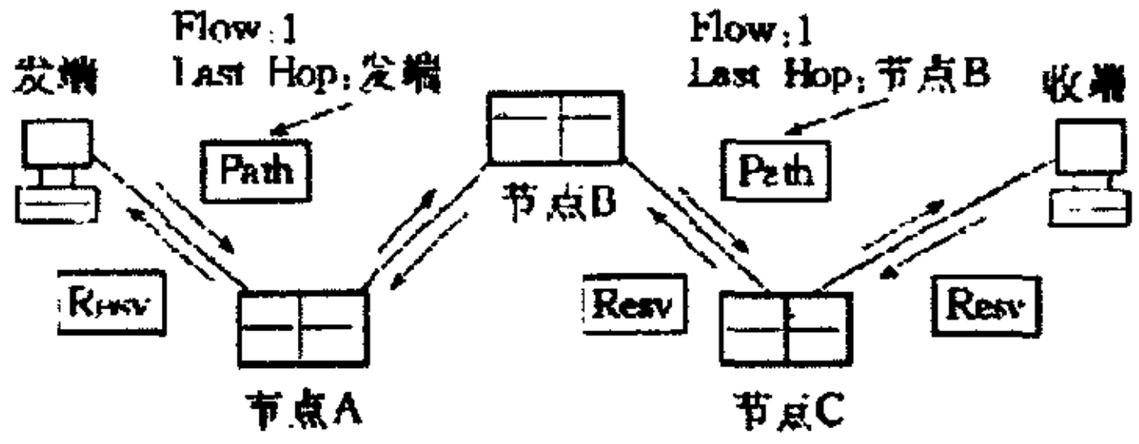

图 3.1 RSVP 协议的控制消息

图 3.1 说明了预留管道是如何工作的。沿着数据路径，每个 RSVP 发送主机通过路由协议所提供的单点传送或多址通信的路由下传 RSVP 的路径信息，这些路径信息存贮了传送中每个节点的路径状态。路径状态至少包括以前路由段节点的单点传送 IP 地址，这个 IP 地址过去常常按确定路线一段一段地反向传递预留请求(Resv)信息。每个接收主机向发送方上传 RSVP 的 Resv 信息，这些信息在 路径上的每个节点产生并保持了预留状态。Resv 信息自身最终也必须被传送到发送主机，因此主机可为第一个路由段设定适当的传输控制参数。与传送数据相同，路径信息也用相同的源和目的 单元地址来传送。另一方面，Resv 信息一段一段进行传送，每个 RSVP 节点向前面 RSVP 路由 段的单点传送地址转送一个 Resv 信息。

为了维持预留资源 ,RSVP 协议使路由器或交换节点维持在一个“软状态” ,这个状态周期性地由 PATH 和 RESV 消息来更新，也可以由拆卸消息来取消。如果在一段时间内没有收到更新报文，预留的资源也将被取消。

RSVP 协议的资源预留请求由流量说明(flowspec)和过滤器说明(filterspec)来定义。流量说明以定量的形式指定服务需要的 QoS，如最大延时、平均吞吐量、最大突发率等。过滤器说明定义了资源预留需要的分组数据的格式。流量说明和过滤器说明一起被称为流量描述器 (Folw Descriptor)。

RSVP 协议包含决策控制 (Policy control)、接纳控制 (Admission control)、分类控制器(Classifier)、分组调度器 (Scheduler)与 RSVP 处理模块等几个主要成分 (如图 3.2 所示)。决策控制用来判断用户是否拥有资源预留的许可权；接纳控制则用来判断可用资源是否满足应用的需求，主要用来减少网络负荷；分类控制器用来决定数据分组的通信服务等级，主要用来实现由 filterspec 指定的分组过滤方式；分组调度器则根据服务等级进行优先级排序，主要用来实现由 flowspec 指定的资源配置。当决策控制或接纳控制未能获得许可时，RSVP 处理模块将产生预留错误消息并传送给收发端点；否则将由 RSVP 处理模块设定分类与调度控制器所需的通信服务质量参数。

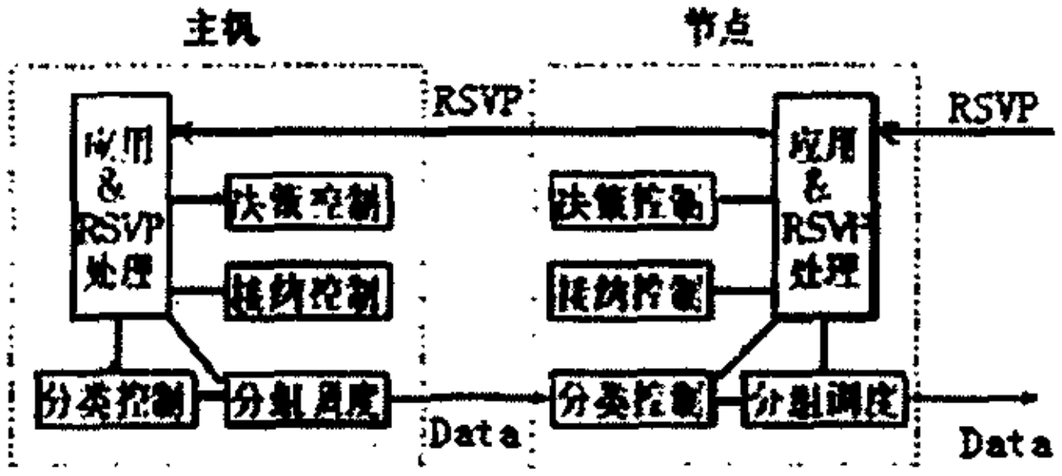

图 3.2　　RSVP 协议的相关组成

RSVP 只是一个信令协议，用来帮助建立端主机和路由器的资源保留状态。资源和服务管理算法主要依赖于所支持的服务级别。RSVP 将两个应用程序间的 IP 流当作网络层的连接来处理，它在 IP 层提供了与 ATM UNI 和信令在信元流层次上相似的功能。

目前实现综合服务的一般手段为采用 RSVP 进行资源预留。 RSVP 通过信令在应用程序和网络元素间进行 QoS 协商。RSVP 首先将发送端生成的业务特性 (SENDER-TSPEC)沿所选的路径朝接收端传输 ,并在沿途收集所经过的网络元素的信息。它包括最小可用带宽和最小路径延迟等。这些信

息保存在 ADSPEC 对象中。 SENDER-TSPEC 和 ADSPEC 对象被封装在 RSVP 的 PATH 消息中。

SENDER-TSPEC 包含的内容有 ：

- 漏桶的容量 b
- 令牌生成速率 r
- 数据流的峰值速率 p
- 最大数据包长度 M
- 最小控制单元 m

ADSPEC 中包含的信息有 ：

(1) 路径上的一般信息

- 沿途的网络元素是否都支持 RSVP
- 沿途最小的 MTU
- 最小的路径延迟
- 最小可用带宽

(2) 每种服务的特定信息

- 沿途的网络元素是否都支持这种服务
- 该服务可用的最小带宽
- 对保证服务 ,还必须包括 Ctot,Dtot 等

当 PATH 消息传到接收端后 ,接收端按照应用的延迟要求计算沿途允许的排队延迟 qdelay。

qdelay=应用允许的端到端延迟 － 最小路径延迟

然后选择满足要求所需要的带宽。一般的选择过程如下:

设 R=p，按公式 (2)计算沿途的延迟，如果结果大于 qdelay，说明当 R=p 时的延迟不能满足应用的要求，应增大 R，即 R 应大于 p。则令 D=qdelay 解方程(2)；否则说明 R 小于 p 就能满足要求，则解方程 (1)。求得能满足要求的 R。在某些情况下，即使 R 设置为其最小值 r 时，得到的端到端排队延迟仍小于 qdelay，则二者的差定义为松弛度。它可使路径上的某些网络元素预留小于 R 的带宽。有时为了给网络元素一定的灵活度，故意增大 R，使实际排队的延迟小于允许的排队延迟，而得到一定的松弛度。此时接收端可发出 RESV 消息申请资源预留。预留请求的内容包括：1、预留的带宽，2、松弛度。现在该 RESV 消息沿原路返回发送端，并完成在每个结点上的预

留。在中间结点上，当收到 RESV 时，它可以根据松弛度来预留小于或等于 R 的带宽。如果中间结点预留的带宽小于 R，则在该结点上必须满足：

$$Sout + b/Rout + Ctot/Rout \leqslant Sin + b/Rin + Ctot/Rin$$

3.2.3 Intserv 的优缺点

Intserv 的优点：

a、这种模型实现了绝对的服务质量保证。这种模型对于业务特征提供了充分的细节，使得 RSVP 服务器可以对各种业务类型的细节进行描述。由于在流所经由的所有路由器上都将运行 RSVP，网络将可以保证在任何一点都没有任何一个数据流能够过量地占用网络资源。

b、使用 RSVP 的软状态特性，可以支持网络状态的动态改变与组播业务中组员的动态加入与退出，同时，利用 PATH 与 RSVP 的刷新，还可以判断网络中相邻的产生与退出。

c、使用 RSVP 的资源预留模式，可以实现组播业务中网络资源的有效分配。

Intserv 存在的问题：

a、首先，可扩展性是 Intserv 的最严重的问题。由于使用了“软状态”的工作方式，同时 RSVP 进行资源预留需要对大量的状态信息进行刷新与储存，当网络规模扩大时，这一模型将无法实现。

b、使用这一模型进行端到端的资源预留要求从发送者到接收者之间的所有路由器都支持所实施的信令协议。

c、信令系统十分复杂，用户认证，优先权管理，计费等也需要一套复杂的上层协议。目前，对于 Intserv 模型，业界已经有了比较一致的意见。这一模型应当应用于网络规模较小，业务质量要求较高的边缘网络，如企业网、园区网等。对于骨干网络的 QoS 技术，则应当使用下面我们即将研究的 Diffserv 模型。

§3.3 区分服务模型

Intserv 存在的问题使其不适宜用于骨干网络。随着业务的不断扩展，网络资源日益紧张，越来越多的 ISP 希望能够在它们与客户之间建立以服务质

留。在中间结点上，当收到 RESV 时，它可以根据松弛度来预留小于或等于 R 的带宽。如果中间结点预留的带宽小于 R，则在该结点上必须满足：

$$Sout + b/Rout + Ctot/Rout \leqslant Sin + b/Rin + Ctot/Rin$$

3.2.3 Intserv 的优缺点

Intserv 的优点：

a、这种模型实现了绝对的服务质量保证。这种模型对于业务特征提供了充分的细节，使得 RSVP 服务器可以对各种业务类型的细节进行描述。由于在流所经由的所有路由器上都将运行 RSVP，网络将可以保证在任何一点都没有任何一个数据流能够过量地占用网络资源。

b、使用 RSVP 的软状态特性，可以支持网络状态的动态改变与组播业务中组员的动态加入与退出，同时，利用 PATH 与 RSVP 的刷新，还可以判断网络中相邻的产生与退出。

c、使用 RSVP 的资源预留模式，可以实现组播业务中网络资源的有效分配。

Intserv 存在的问题：

a、首先，可扩展性是 Intserv 的最严重的问题。由于使用了“软状态”的工作方式，同时 RSVP 进行资源预留需要对大量的状态信息进行刷新与储存，当网络规模扩大时，这一模型将无法实现。

b、使用这一模型进行端到端的资源预留要求从发送者到接收者之间的所有路由器都支持所实施的信令协议。

c、信令系统十分复杂，用户认证，优先权管理，计费等也需要一套复杂的上层协议。目前，对于 Intserv 模型，业界已经有了比较一致的意见。这一模型应当应用于网络规模较小，业务质量要求较高的边缘网络，如企业网、园区网等。对于骨干网络的 QoS 技术，则应当使用下面我们即将研究的 Diffserv 模型。

§3.3 区分服务模型

Intserv 存在的问题使其不适宜用于骨干网络。随着业务的不断扩展，网络资源日益紧张，越来越多的 ISP 希望能够在它们与客户之间建立以服务质

区分服务的模型可由图 3.3 表示：

区分服务在实现上由三个功能模块组成：每跳行为 PHB(Per Hop Behavior)、包的分类机制和流量控制功能 (测量、标记、整形、策略控制)。

区分服务实现可扩展性的重要策略是在网络中心节点只进行转发操作，将分类和大部分流控的复杂性操作转移到了网络边缘节点。同时，将同类的流聚集传输，避免了大量的流状态信息的保存，大大降低了网络实现的复杂性和网络负荷。

3.3.1 区分服务中的标记分类机制

区分服务中传输的是流聚集而不是单个的流，每一组流聚集都具有其相应的各自不同的流传输服务标准，在各个域内部根据不同的媒体传输要求提供不同的传输服务。这一过程是通过区分服务中 IP 包头的区分服务标记域(DS Field)来实现的，DS 的标记域在 IPv4 中定义为包头的 TOS 字节，在扩展的 IPv6 中定义为包头的流类型字节 (Traffic classoctet)的前六位，DS 标记域对应相应传输媒体的 PHB。

传输分类的过程是在边界节点上进行的，边界节点查询 DS 标记域并将其归入某一特定的流聚集中。DS 模型中边界调节分类的部分主要包括接入控制(Admissioncontrol)，判断是否有足够的资源来支持相应类型的控制；包分类器 (Packketclassifier)，确定源地址、目的地址、端口字段、判断包的类型；包调度器(Packetscheduler)，用来调度包的发送，在调度器中，负责主要的包流量的整形与调度，提供标记器 (Marker)、计量器(Meter)、丢包器(Dropper)三部分。计量器用来测量业务流的速率，标记器用来设置 DS 码点并对 IP 包头进行标记，整形器用于业务的发送以满足规定的业务发送级别的要求，决策操作通过丢弃数据包来强制执行规定的业务级别。如图 3.4 所示

计量器和丢包器已经有了许多比较成熟的调度算法，针对 IP 实现 QoS 而提出的几个较重要的算法及其思想将在下面进行介绍。

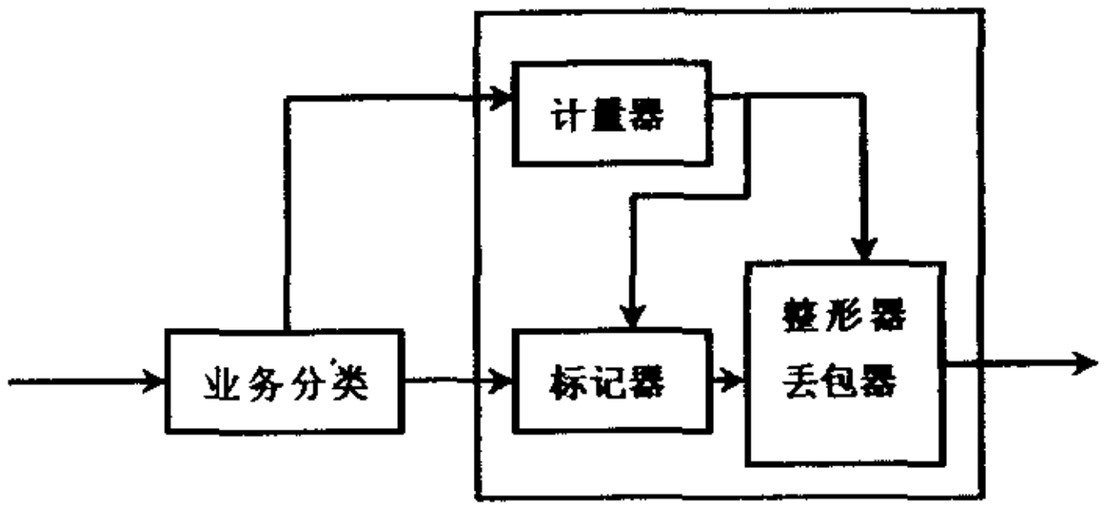

图 3.4 Diffserv 的业务分类器和业务调整器

3.3.2 区分服务当中的逐跳行为 PHB

在 Diffserv 域的路由器中，将对属于某一服务类别的业务流进行一致的处理。这种处理包括队列选择、排队、丢弃等。对属于同一服务类别的业务流进行的标准化处理的组合就构成了每一跳行为（PHB）组。PHB 中还包括了该 PHB 组与其他 PHB 组之间的互操作问题。

PHB 是一个 DS 节点调度转发处理包头标有 DS 标记的 IP 包流的外部行为描述。在 DS 域内，转发节点是按照 PHB 来进行的，在每一传输段逐段保证 PHB 行为是区分服务的最大特点，也是区分服务分段保证端到端 QoS 的基础。PHB 可以用一系列流的参数特性包括延迟、抖动、优先级等来描述。

PHB 是对路由器服务质量处理的总体描述，它并不对实现 PHB 的具体技术加以规定。这样，不同的厂商将可以使用自己的实现方式，只要结果能够满足标准 PHB 的要求就可以了。另外，通过对标准 PHB 的组合，各个厂商将可以实现自己所专有的业务。IETF 已经标准化了一部分 PHB ,包括尽力而为 PHB（BE-PHB），类别选择 PHB（CS-PHB），加速转发 PHB（EF-PHB）和可靠转发 PHB（AF-PHB）。具体如下：

1） 尽力而为 PHB（BE-PHB）

Internet 规定，当 DSCP 为全零（编码点为“000000”）时，对应的 PHB 就是尽力而为 PHB，也称为默认 PHB。当路由器收到 DSCP 为全零或者是无法识别的 DSCP 值时，都将使用尽力而为 PHB 对分组进行转发。但在后一种情况下，应当保持分组中的 DSCP 值不变。也就是说，尽力而为 PHB 是一种默认的服务质量。

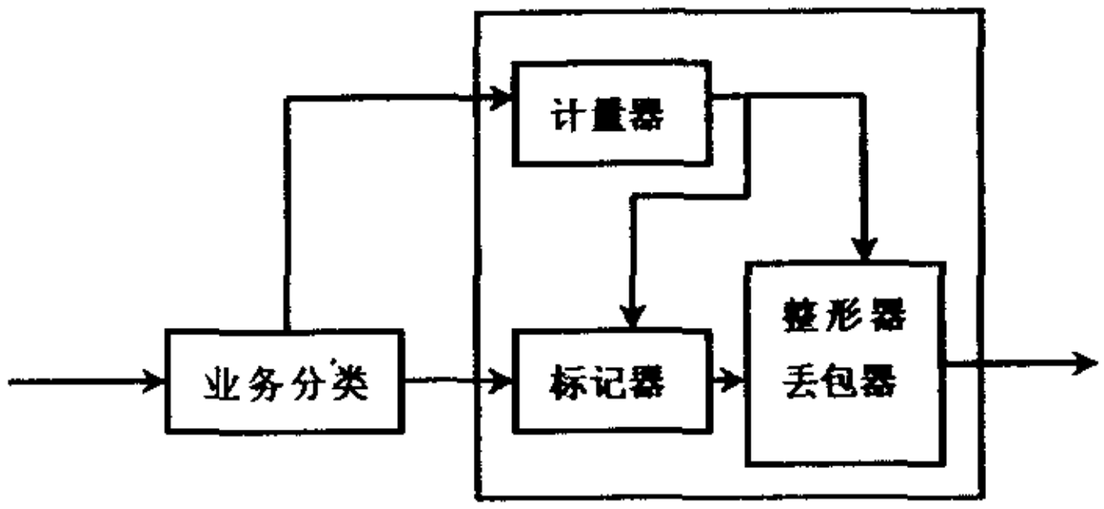

图 3.4 Diffserv 的业务分类器和业务调整器

3.3.2 区分服务当中的逐跳行为 PHB

在 Diffserv 域的路由器中，将对属于某一服务类别的业务流进行一致的处理。这种处理包括队列选择、排队、丢弃等。对属于同一服务类别的业务流进行的标准化处理的组合就构成了每一跳行为（PHB）组。PHB 中还包括了该 PHB 组与其他 PHB 组之间的互操作问题。

PHB 是一个 DS 节点调度转发处理包头标有 DS 标记的 IP 包流的外部行为描述。在 DS 域内，转发节点是按照 PHB 来进行的，在每一传输段逐段保证 PHB 行为是区分服务的最大特点，也是区分服务分段保证端到端 QoS 的基础。PHB 可以用一系列流的参数特性包括延迟、抖动、优先级等来描述。

PHB 是对路由器服务质量处理的总体描述，它并不对实现 PHB 的具体技术加以规定。这样，不同的厂商将可以使用自己的实现方式，只要结果能够满足标准 PHB 的要求就可以了。另外，通过对标准 PHB 的组合，各个厂商将可以实现自己所专有的业务。IETF 已经标准化了一部分 PHB ,包括尽力而为 PHB（BE-PHB），类别选择 PHB（CS-PHB），加速转发 PHB（EF-PHB）和可靠转发 PHB（AF-PHB）。具体如下：

1） 尽力而为 PHB（BE-PHB）

Internet 规定，当 DSCP 为全零（编码点为“000000”）时，对应的 PHB 就是尽力而为 PHB，也称为默认 PHB。当路由器收到 DSCP 为全零或者是无法识别的 DSCP 值时，都将使用尽力而为 PHB 对分组进行转发。但在后一种情况下，应当保持分组中的 DSCP 值不变。也就是说，尽力而为 PHB 是一种默认的服务质量。

2) 类别选择 PHB（CS-PHB）

为了与现在正在使用的 IP 优先级字段保持一定的后向兼容，在 Diffserv 中定义了类别选择 PHB。现有的 IP 优先级机制使用了 TOS 字段的前 3bit，从而可以提供 8 个 IP 优先级。可见，这种方式与 DSCP 的用法是十分相似的，其不同在于 Diffserv 使用了 TOS 字段中的前 6BIT，另外现有的路由器都能够理解 TOS 域的意义。所以，只要将 Diffserv 的一部分编码分配给传统 IP 优先级业务就可以很容易地实现上述的后向兼容。同时 Diffserv 的业务等级将可以与传统的 IP 优先级同时并存于网络之中。类别选择编码点的分配为“xxx000”，亦即“000000”到“111000”8 个编码点。可见类别选择编码点的位置与传统 IP 中的 IP 优先级字段的位置是完全一样的。

3) 加速转发 PHB（EF-PHB）

加速转发 PHB 所描述的是一组用于实现低丢包率、低延迟、低抖动、具有带宽保证的，以及在 DS 域中具有端到端服务质量的业务的服务策略。使用这一 PHB 组的业务流将获得 Diffserv 网络中最高的服务质量，具有最高的优先级，在转发过程中所使用的队列将是节点上最短的。当网络发生拥塞时，这类业务将获得最先处理，这样，便可以使得这类业务的时延最小，同时也改善了该业务的其他服务质量参数。这一 PHB 对应的 DSCP 编码为“1011110”。

4) 可靠转发 PHB（AF-PHB）

可靠转发 PHB 所要达到的目标实际上主要是要对相同业务中不同分组的丢失优先级进行一定的分级。在业务开始转发之前，发送方与网络节点之间将对业务流的速率作出一定的约定，这种约定称为业务流的轮廓(profile)。在 AF-PHB 中，网络节点将允许业务流的速率大于这一轮廓，但是，网络节点将对超出轮廓的业务流分组采用较大的丢弃优先级。根据这一思想 RFC2597 对可靠转发 PHB 作出了定义。RFC2597 规定，AF-PHB 组包括四个等级。网络中的节点将根据这些等级，为相应的业务流分配网络资源并进行相应的转发处理。另外对于每种不同类别的 AF，还分别规定了三种不同的丢包优先级，优先级越高，分组丢弃的机率就越大。也就是说，AF 目前一共有 12 种不同的编码点。

3.3.3 Diffserv 的优点和存在的问题

Diffserv 最主要的优势在于它弱化了对信令的依赖，中间节点只需依据一定的分组标志应用各种 PHB 就可以了。这样，将无须像在 Intserv 中那样在每个路由器上为每个业务流保存“软状态”，从而避免了大量的资源预留信息的传递，具有良好的可扩展性。

Diffserv 并不要求实现端到端的服务质量保证，而只要求域（domain）的范围服务质量的一致性，而在网络区域之间，对不同类别业务的服务质量的保证由一定的映射机制来保证。Diffserv 将服务质量的一致性范围缩小到了每个区域之中，从而降低了这种模型实现的复杂性。

Diffserv 的设计思想是希望使用一种与目前 IP 网络协议相结合的方式来实现对网络 QoS 的保证，因此其实现将比使用端到端控制的 Intserv 简单，网络额外负担也较小。

另外，在简化了网络实现的同时，Diffserv 使用的业务量组合模型也造成了服务质量某种意义上的不可预测性。这样，对于一个业务量组合来说，其中的业务量大小也是难以预测的，在这种情况下，要提供确定的服务质量可以说是不可能的。Diffserv 所提供的服务质量从本质上说只是一种相对的服务质量，也就是说，这种服务质量只是不同等级业务量之间服务质量的好坏关系。

目前，业界普遍认同了 Diffserv 的设计思想，可以预见的一点是，Diffserv 将成为未来广域网中居统治地位的 QoS 技术。在局域网中，各种应用可以依据各自的需求选择所要使用的 QoS 技术，这一技术既可以是 Intserv 或 Diffserv 也可以是现有的网络媒体所能够提供的 QoS 能力。而在广域网中，为了解决可扩展性问题，实现一定意义上的端到端的服务质量，则需要使用 Diffserv 技术。而 MPLS 将是实施 Diffserv 模型的重要载体。

§3.4 区分服务调度算法

基于IP的网络实现QoS的重要途径是确定良好的包调度与流控制算法。区分服务当中的流调度机制实现于边界节点的调度器和中间节点的丢包器，目前对提出的调度算法已经有了许多改进算法。

3.3.3 Diffserv 的优点和存在的问题

Diffserv 最主要的优势在于它弱化了对信令的依赖，中间节点只需依据一定的分组标志应用各种 PHB 就可以了。这样，将无须像在 Intserv 中那样在每个路由器上为每个业务流保存“软状态”，从而避免了大量的资源预留信息的传递，具有良好的可扩展性。

Diffserv 并不要求实现端到端的服务质量保证，而只要求域（domain）的范围服务质量的一致性，而在网络区域之间，对不同类别业务的服务质量的保证由一定的映射机制来保证。Diffserv 将服务质量的一致性范围缩小到了每个区域之中，从而降低了这种模型实现的复杂性。

Diffserv 的设计思想是希望使用一种与目前 IP 网络协议相结合的方式来实现对网络 QoS 的保证，因此其实现将比使用端到端控制的 Intserv 简单，网络额外负担也较小。

另外，在简化了网络实现的同时，Diffserv 使用的业务量组合模型也造成了服务质量某种意义上的不可预测性。这样，对于一个业务量组合来说，其中的业务量大小也是难以预测的，在这种情况下，要提供确定的服务质量可以说是不可能的。Diffserv 所提供的服务质量从本质上说只是一种相对的服务质量，也就是说，这种服务质量只是不同等级业务量之间服务质量的好坏关系。

目前，业界普遍认同了 Diffserv 的设计思想，可以预见的一点是，Diffserv 将成为未来广域网中居统治地位的 QoS 技术。在局域网中，各种应用可以依据各自的需求选择所要使用的 QoS 技术，这一技术既可以是 Intserv 或 Diffserv 也可以是现有的网络媒体所能够提供的 QoS 能力。而在广域网中，为了解决可扩展性问题，实现一定意义上的端到端的服务质量，则需要使用 Diffserv 技术。而 MPLS 将是实施 Diffserv 模型的重要载体。

§3.4 区分服务调度算法

基于IP的网络实现QoS的重要途径是确定良好的包调度与流控制算法。区分服务当中的流调度机制实现于边界节点的调度器和中间节点的丢包器，目前对提出的调度算法已经有了许多改进算法。

在 Internet 中一个包从源端到目的端可能要经过一个或多个网络域。在 Diffserv 框架中，相邻域之间对服务水平约定 SLA(service level agreement)进行协商,它定义了从一个域到另一个域每一个服务水平可以传送的业务量。更多的技术细节，例如协议速率、最大突发量等等由业务状态约定 TCA(traffic conditioning agreement)定义.流量调节 器(TC(traffic conditioner)) 安置在边界路由器中，用以保证全部的聚合流量不要超过 TCA 中的流量规定。

下面研究针对区分服务的特点提出的或改进的调度算法:

(1) 加权公平队列 (Weighted Fair Queuing)算法

WFQ 算法用来控制由路由器交换机向外发送数据包，每一个包是基于时间戳的，规定其到达、发送时间、以及它们的长度，WFQ 调度器的发送队列记录每一个包的到达时间,使得时间戳最小的包先发送。根据权值变量，允许某些特定的流能够得到比其它流更多的带宽。另外，WFQ 是工作存储(Work-conserving)的，就是说路由器将会不间断地发送当前的包而使链路不处于空闲状态。以上两点保证了 WFQ 调度器成为支持实时流量的理想算法。但研究同时表明，WFQ 调度算法也存在一些问题，如带宽利用不足、带宽 / 时延耦合、流量特性扭曲、流量粒度过细等。现在提出了一些新的算法 ,如 W^2 FQ 等。

(2) RED(Random Early Detection)算法

RED 算法是针对 Diffserv 模型展开的主要研究算法，也是使用最为广泛的一种，RED 是一种针对 FIFO 分组调度算法进行的分组缓冲管理和丢弃机制，其特点是通过一定丢弃规则丢弃分组，使得路由器的排队队列最短。图 3.5 是 RED 算法的检测队列示意图。其工作参数为两个阈值 min、max 和一个最大丢弃概率 Pmax。当缓冲排队长度小于 min 时，不对进入的分组进行丢弃；当缓冲排队长度大于 max， 丢弃所有进入分组；而当缓冲排队长度 S 在两阈值之间时，以概率((S- min)/ (max- min)) *Pmax 对进入的分组进行丢弃。现在已经出现了一些 RED 的改进型算法如 DWRED、GRED 等。

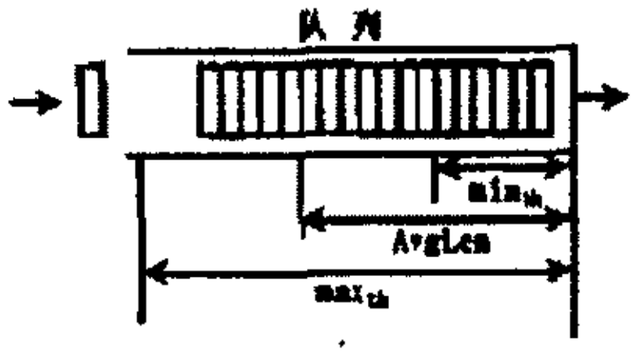

图 3.5 随机早期检测 RED

(3) Two drop 和 Three drop 过程算法

Two drop 过程算法是令牌桶过滤器 (Token Buckt Filter)的改进。

令牌桶过滤器算法是由 Clark, Shenker 和 Zhang 在其关于包交换网支持实时网络流的论文中提出的。图 3.5 给出了这一算法的示意图。其基本思想是，指定一个大小为 B 的令牌桶，当包到达时，从桶中减去一定数量的令牌。

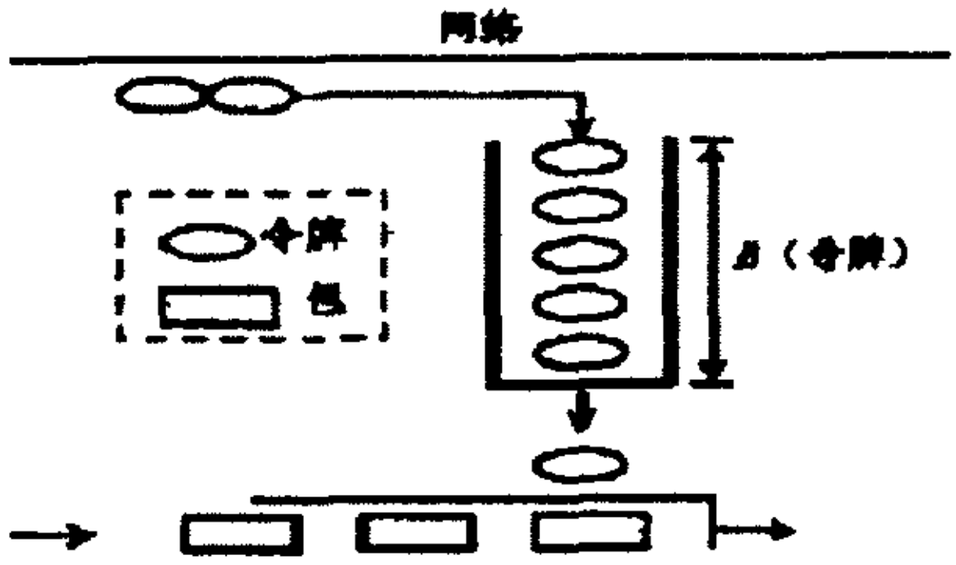

图 3.6 令牌桶过滤器

除非桶中有足够数量的令牌，否则就不能被发送。因此，控制发送端包的发送数量≤B，若流量源以小于等于 R 的速率发送包，称其为遵守令牌过滤器的参数。因此容易理解由令牌过滤器规划的网络，因为，遵守流量的数据包就应满足 R(t)≤B(对于任何时间增量为 t 而言)，实现上主要考虑怎样处理不遵守流量的数据流。令牌桶已被用在许多路由器的实现当中。

令牌桶改进算法源于 RIO 网络，在 RIO 网络中，网络边缘设备控制并标记进入的单个或聚集流包，若包到达时的发送速率在定义的范围以内，包被标志为 IN，否则被标记为 OUT。

漏桶标记是流量调节的其中一种，用它来实现随机早期丢弃（RED—in/out)。如图 3.7 所示。TCA 由上游域和下游域共同约定，由上游域进入的业务流量为 r bit/s,可以进入下游域的最大突发量为 b。在桶漏模式中，

即桶的大小为上游域可以流入下游域最大的突发量 b，桶漏速率为 $r$bit/s。当从上游域来的一个包到达，如果这个包被标记为“out”，流量控制器（TC）简单的以“out”转发。当这个包被标记为“in”，流量控制器（TC）检查漏桶看是否有足够的令牌。如果有，这个包按“in”转发，同时漏桶中减去包的大小对应的令牌数。否则的话，这个“in”标记的包降级为“out”进行转发。

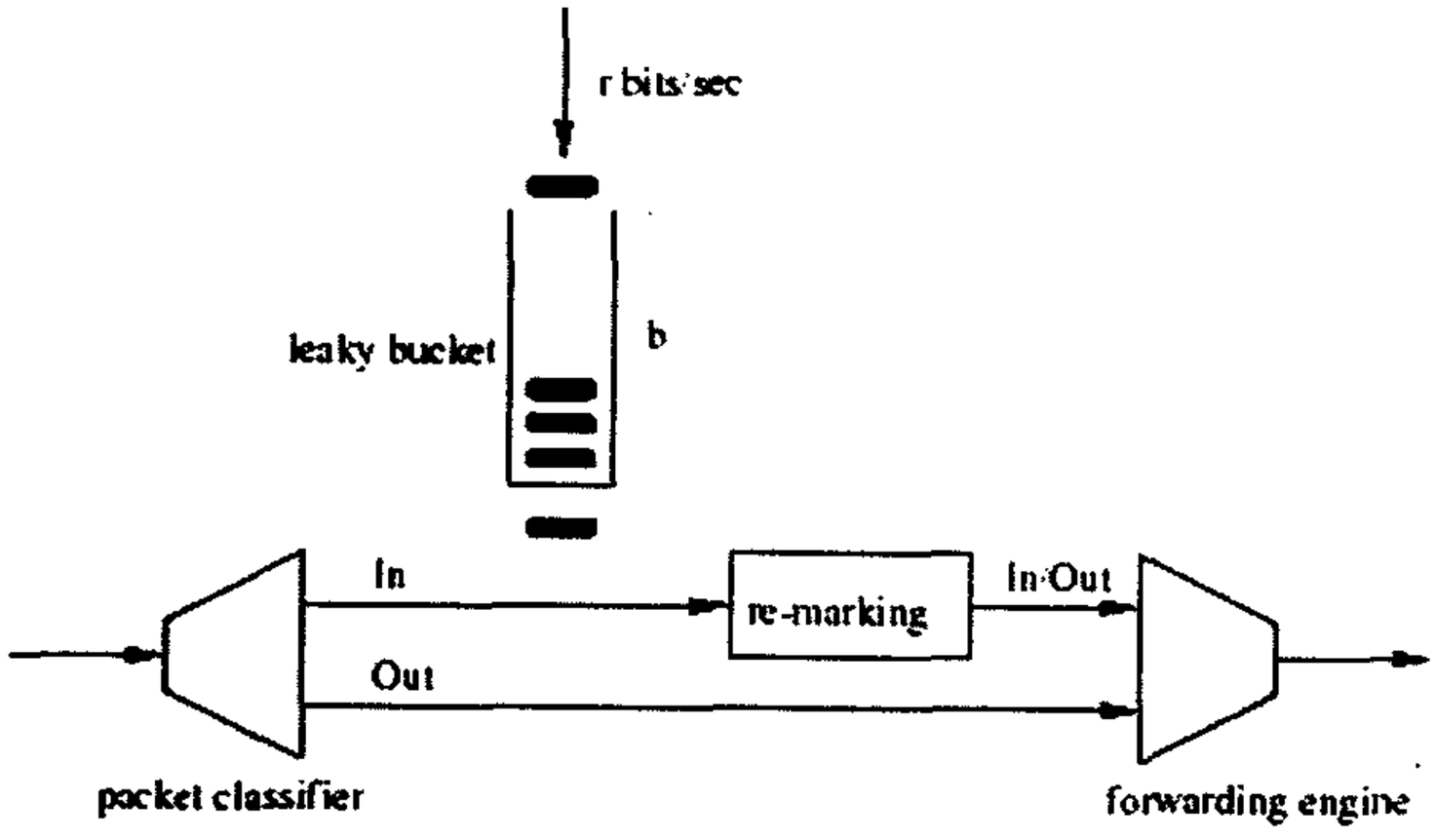

图 3.7 漏桶的标记模式

假定端与端之间域的数量为 n，一个包在通过每一个域降级的概论为 p，则这个包在端到端后降级的概率为 $1-(1-p)^n$ 。

Three drop 过程算法是在 Two drop 过程算法的基础上进行的扩展，将每一个包标记为红、绿、黄三种颜色。通常黄色代表的预留速率大于绿色代表的预留速率，绿色代表的预留速率大于红色代表的预留速率。两速三色(tr-TCM)标记方法定义了 AF 中的 3 种颜色的标记模式。这种调度算法我们在后面的仿真中用到。trTCM 测量信息流，并根据三种流量参数（提交信息速率，Committed Information Rate，CIR；提交组量大小 Committed Burst Size，CBS；超量组量大小 Excess Burst Size，EBS）对包进行标记，这三个参数我们分别称为绿，黄和红标记。如果包没有超过 CBS 就是绿的，如果超过 CBS 但未超过 EBS 就是黄的，如果超过 EBS 就是红的。

用更细的分级获得更好的执行控制效果。一些基于 * drop 的算法采用

Two drop 和 Three drop 的思想，通常被称为 Multi drop 过程， * drop 过程的思想被广泛地采纳于 Diffserv 的实现当中。

(4) 桶漏模式下的随机早期升降级算法(REDP)

下面使用三种不同的颜色作为标记模式，这样一个可能被标记为绿色、黄色或红色。假设一个用户发送一个期望速率为 r 的数据流，这个流通过一个漏桶标记器进行调度。这个流其中的一个包如果在要求区内被标记为绿色，不在要求区内被标记为红色。当经过域的边界时，如果聚合绿色包的速率超过协议速率，绿色的包将会被降级为黄色。如果聚合绿色包的速率低于协议速率，黄色的包将被升级为绿色。黄色包不会被降级为红色，红色也不会被升级为黄色。

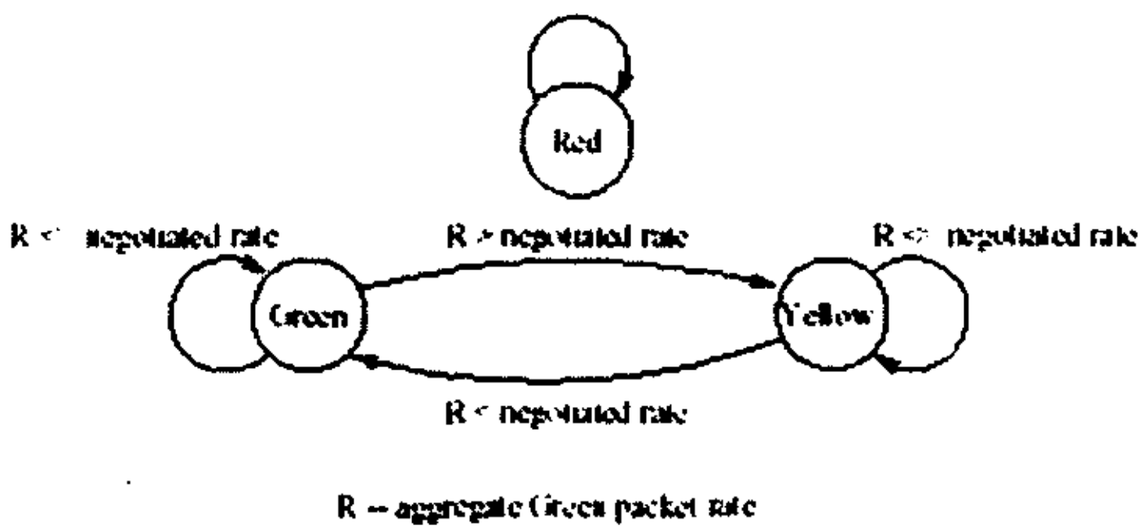

图 3.8　　三个颜色的升降级状态图

可能存在很多小的流量包从上游域通过中间的标记器传输到下游域，聚合 Green 包速率是这些小业务流的全部绿色包速率的总和。

标记模型如图 3.9 所示。T_L 和 T_H 作为两个界限将漏桶划分为三个区：降级区、平衡区和升级区。在标记过程中可能进行以下三种过程：

1)平衡：如果到达的 green 包速率等于令牌填充速率 *r*, 令牌的消耗速率就等于令牌填充速率，这样，桶内的令牌数量就保持，每一个包的转发都保持颜色不变。

2)降级：如果 green 包的速率超过 *r*，令牌消耗的速率大于令牌收回的速率，令牌的数量减少并且令牌数量跌入降级区。在降级区每一个到达的包以

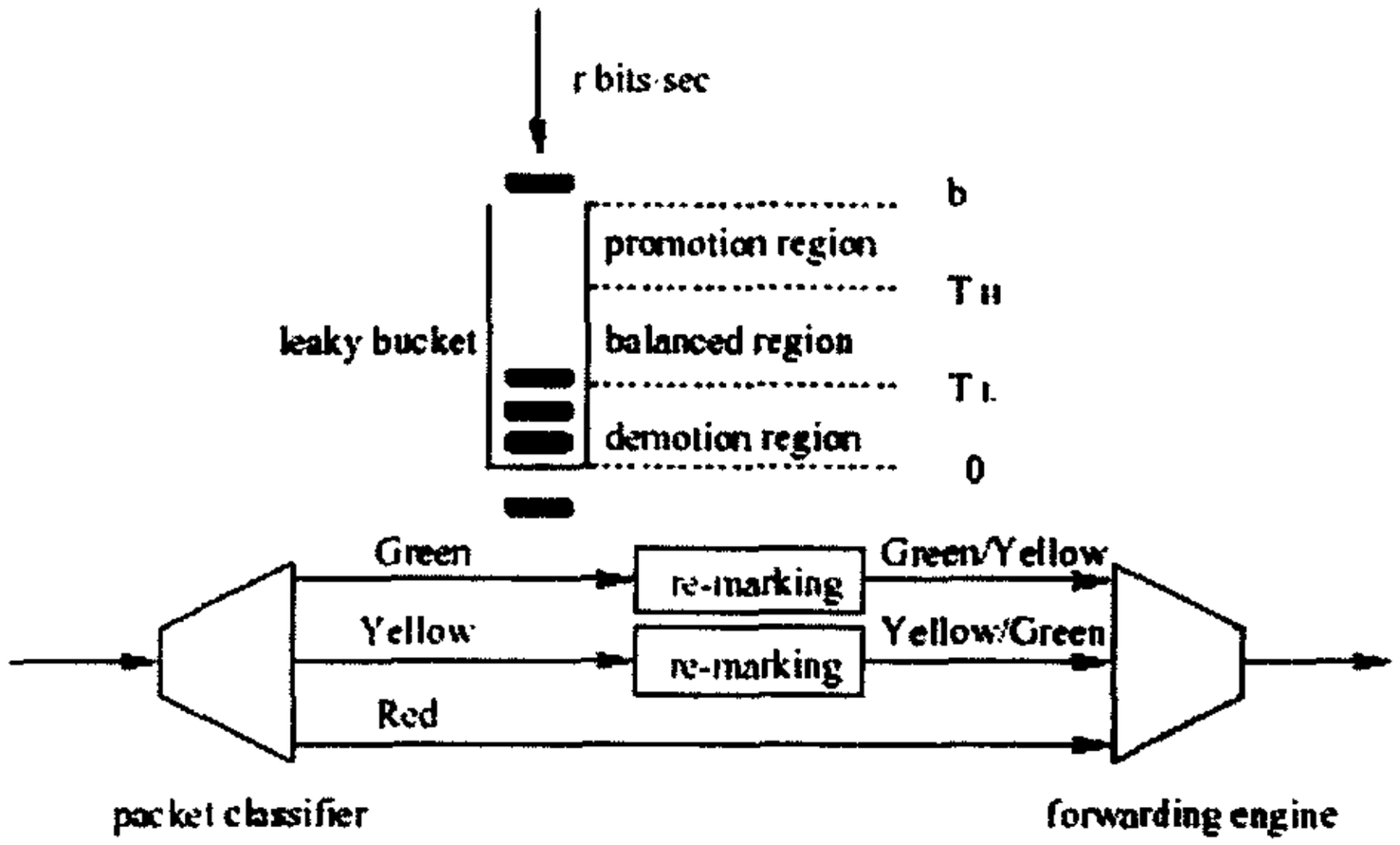

图 3.9　　REDP 标记

概论 P_{demo} 随意降级为 yellow，在这里概率 P_{demo} 是由令牌的数量（TK_{num}）决定的。

$$P_{demo}=(T_l-TK_{num})\cdot MAX_{demo}/\ T_l$$

MAX_{demo} 为最大的降级速率，当漏桶耗尽全部令牌，每一个 green 包都降级为 yellow。

3)升级：如果到达的 green 包速率小于 r,令牌回收速率大于令牌消耗速率。令牌数增加并且数量到达升级区。在升级区每一个 green 包转发为 green，消耗一定数量的令牌。每一个到达的 yellow 包以 p 的概论随意升级为 green。

$$P_{promo}=(\ TK_{num}-T_H)\cdot MAX_{promo}/(b-T_H)$$

以上讨论的是针对 IP　QoS 而设计的一些算法，流量整形是 Diffserv 当中重要的组成部分，因而良好的整形算法对于保证良好的服务质量至关重要。在实际应用中，往往是使用以上算法的基本思想，而采取多种方法相结合的方式来进行流量管理，特别是在 Diffserv 公平性方面的研究更是促进了以上各种算法的发展和丰富。

§3.5 MPLS 网络中 QoS 功能的实现

3.5.1 MPLS 网络中实现综合服务

当 MPLS 应用于边缘网络中时，就可能需要对 Intser 模型进行支持。MPLS 可以使用扩展的 RSVP 作为其控制协议。由于 Intserv 模型的信令协议就是 RSVP，已经有现成的标准可以使用。这样，使用扩展 RSVP 将可以很方便直接地实现各种 Intserv 所规定的业务。使用扩展 RSVP 信令的 MPLS 网络在与标准的 Intserv 网络互通时只需进行简单的适配就可以了。

当 Intserv 要通过使用 CR-LDP 信令的 MPLS 网络传输时，就需要实现 CR-LDP 协议与 Intserv 网络中 RSVP 协议的互通。RSVP 信令系统在进行资源预留的过程中 CR-LDP 的各种服务参数完全能够满足 Intserv 网络中的各种服务质量参数。

在 CR-LDP MPLS 网络的边缘，入口 LSR 收到服务类型为 Intserv 的业务请求时，将把这种请求转换为 MPLS 的连续请求，同时将 Intserv Tspec 中的 p、b、r 等参数映射为 CR-LDP 的 PDR、PBS、CBS 等参数，将业务的各种优先级映射为 CR-LDP 中的业务优先级参数。

在 MPLS 的连接建立过程中，标记请求信息中将包含上述参数。出口 LSR 收到标记请求消息之后，对上述参数转换回 Intserv 的 p、b、r 等参数，向下游发出 PATH 消息。出口 LSR 收到 RESV 消息后，将 Rspec 中的 R、S 等参数映射为 MPLS 的服务质量参数。在标记映射消息中，将包含上述服务质量参数。

当 MPLS 入口 LSR 收到标记映射消息时，该 LSR 将把收到的资源预留参数转换为 Intserv 的各种参数并向上游发出 RESV 消息。

在 Intserv 规范中，对于超出规定速率的流量使用 Best-effort 方式进行传输。在 MPLS 网络的边缘，对于超出规定速率的流量，将有边缘 LSR 中的业务量调整机制进行处理。

3.5.2 在 MPLS 网络中实现区分服务

区分服务在 MPLS 网络中实现，相对于传统 IP 网络有明显的优势。MPLS 是 IP 网络中引入面向连接的机制，采用建立标记交换通路来转发分组，能

§3.5 MPLS 网络中 QoS 功能的实现

3.5.1 MPLS 网络中实现综合服务

当 MPLS 应用于边缘网络中时，就可能需要对 Intser 模型进行支持。MPLS 可以使用扩展的 RSVP 作为其控制协议。由于 Intserv 模型的信令协议就是 RSVP，已经有现成的标准可以使用。这样，使用扩展 RSVP 将可以很方便直接地实现各种 Intserv 所规定的业务。使用扩展 RSVP 信令的 MPLS 网络在与标准的 Intserv 网络互通时只需进行简单的适配就可以了。

当 Intserv 要通过使用 CR-LDP 信令的 MPLS 网络传输时，就需要实现 CR-LDP 协议与 Intserv 网络中 RSVP 协议的互通。RSVP 信令系统在进行资源预留的过程中 CR-LDP 的各种服务参数完全能够满足 Intserv 网络中的各种服务质量参数。

在 CR-LDP MPLS 网络的边缘，入口 LSR 收到服务类型为 Intserv 的业务请求时，将把这种请求转换为 MPLS 的连续请求，同时将 Intserv Tspec 中的 p、b、r 等参数映射为 CR-LDP 的 PDR、PBS、CBS 等参数，将业务的各种优先级映射为 CR-LDP 中的业务优先级参数。

在 MPLS 的连接建立过程中，标记请求信息中将包含上述参数。出口 LSR 收到标记请求消息之后，对上述参数转换回 Intserv 的 p、b、r 等参数，向下游发出 PATH 消息。出口 LSR 收到 RESV 消息后，将 Rspec 中的 R、S 等参数映射为 MPLS 的服务质量参数。在标记映射消息中，将包含上述服务质量参数。

当 MPLS 入口 LSR 收到标记映射消息时，该 LSR 将把收到的资源预留参数转换为 Intserv 的各种参数并向上游发出 RESV 消息。

在 Intserv 规范中，对于超出规定速率的流量使用 Best-effort 方式进行传输。在 MPLS 网络的边缘，对于超出规定速率的流量，将有边缘 LSR 中的业务量调整机制进行处理。

3.5.2 在 MPLS 网络中实现区分服务

区分服务在 MPLS 网络中实现，相对于传统 IP 网络有明显的优势。MPLS 是 IP 网络中引入面向连接的机制，采用建立标记交换通路来转发分组，能

够明确指示从源端到目的端的路由。这使得在 MPLS 中的区分服务能够解决面向无连接的传统 IP 无法保证的 QoS 的问题。

不过，在 MPLS 网络中，LSP 建立之后，核心路由器只根据 Shim 头标中的标记来区分分组所属的 FEC 并进行转发服务，不对 IP 头进行解析。而标示区分服务级别的 DS 码点位于 IP 头标中，在 MPLS 网络中无法体现，因此必须将由 DS 字段所标示的 DSCP 正确映射到 MPLS 头标中才能使网络中的标记交换路由器正确识别 IP 分组的区分服务等级，提供相应的转发服务，实现对区分服务的支持。

要实现区分服务中的 DSCP 到 MPLS 头标的映射，可以利用头标中的 EXP 域。在建立一条 LSP 的同时，也在沿途 LSR 中建立 EXP 域到区分服务 PHB 的映射表。LSP 建立之后，标记边缘路由器 LER 对进入网络的分组再加上标记的同时，也能按照分组的 DSCP 与 EXP 的映射关系，在 EXP 域写入有关区分服务的信息。在网络内部，标记交换路由器 LSR 将综合分组的标记和 EXP 域的值提供相应的区分服务和转发处理。

Diffserv 定义了一组行为集合 BA(ehavior Aggregate)，该集合的相同约束条件形成一个有序集合(OA: Ordered Aggergate)。Diffserv 还定义了一个或多个 PHB 组，这些组形成 PHB 调度类型(PSC:PHB Scheduling Class)。

1、标记交换路由器 LSR 的标记转发模型

在支持区分服务的 MPLS 网络中，LSR 的分组标记转发模型如图 13 所示。LSR 的可根据收到的 IP 分组的标记并结合 shim 头标中的 EXP 域的值，对分组采取相应的区分服务支持和转发处理。

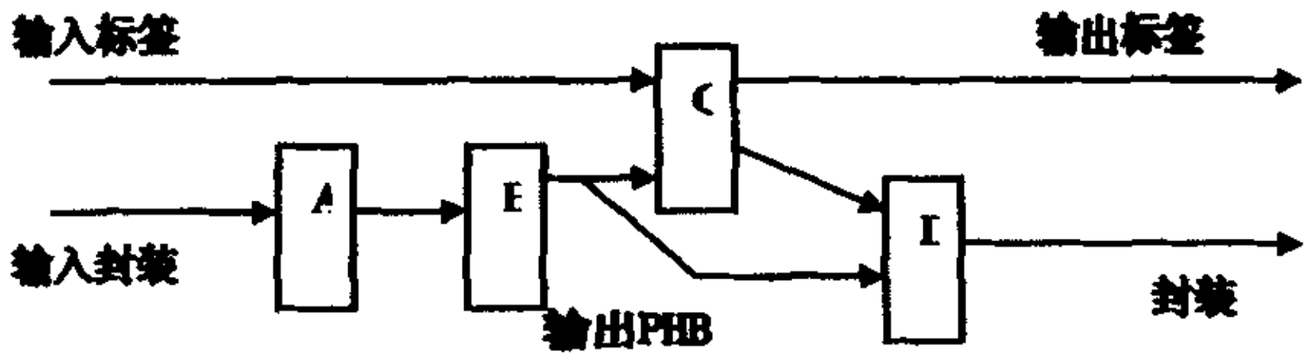

图 3.8　具有区分服务功能的 LSR 工作过程

如图 13 所示，在标记交换路由器中进行的分组标记转发过程分为四个步骤：

A) 输入 PHB 判决：接受到的分组标记和 EXP 域的值，根据输入标记映射表 ILM（Incoming Lable Map），确定对该分组采用区分服务处理 PHB。

B) 根据本地策略和流量情况确定输出 PHB：该阶段在 LSR 中是一个可选的功能，在此阶段主要完成由上阶段得到的 PHB 到输出 QoS 参数的映射；

C) 标签转换（Label Swaping）；

D) DS 信息的封装编码：本阶段主要完成将区分服务信息封装到发送分组的相关域，如 MPLS 中的 EXP 域、ATM 中的 CLP 域，帧中继的 DE 域等。

2、 两种类型的 LSP

图 13 中的转发模型利用 MPLS 头标中的 EXP 域标识分组的不同区分服务等级。但是，区分服务中的 DS 码点有 6 个 bit，最多可以区别 64 个不同服务等级。因此，IP 分组要在 MPLS 网络中实现区分服务需要分为两种情况：

1）E-LSP(EXP-Infered LSP)

当网络提供的服务等级少于 8 个，即不同的 BA 不服多于 8 个时，路由器建立称为 E-LSP 的通道。之所以称它为 E-LSP，是因为此类 LSP 中转发的分组属于哪个 PSC 完全由分组的 EXP 域值决定。在 E-LSP 中，分组的 DS 码点与 EXP 映射建立完全映射，仅由 EXP 域就可表达去分服务的信息，如丢弃优先级和排队调度处理等。沿途的 LSR 将根据 EXP←->PHB 的映射表，确定 IP 分组相应的区分服务等级。

对于 E-LSP，要在各个路由器预先配置一张适用于所有通过此路由器的 LSP 的 EXP<-->PHB 映射表。而在每个 LSP 建立时，应当建立独立的 EXP<-->PHB 映射表，并发送给此 LSP 沿途所有所有的 LSR。各路由器进行转发时根据此映射来确定经过这条 LSP 传输的 IP 分组的区分服务等级。如果 LSR 不能在路径建立时得到 EXP<-->PHB 的映射表，就根据预先配置的映射表确定 IP 分组的 PHB.

实际传送分组时，在网络入口处，LER 给分组分发标记的同时将根据输入标记映射表 ILM 确定头标中表达区分服务的 EXP 域值。在网络内部，LSR 收到携带标记的分组，将查找此分组 MPLS 头标的 EXP 域。根据 EXP 的值确定对此分组采用的 PHB，然后进行相应的转发服务，也就是说，分组的寻路

由头标中的标记域通过在沿途各个路由器上的交换来实现。而对分组采用的相应区分服务就由头标中的 EXP 域来决定。

2）L-LSP(Label-Only-Infered-PSC LSP)

由于 EXP 域只有 3 个 bit，所以当要传送分组所属的 BA 数目多于 8 个时，分组 DSP 所携带的区分服务信息就不能完全映射到 EXP 域，支持区分服务的 MPLS 网络将建立 L-LDP 通道。标记建立时，分组所属的 OA 将反映在标记中；标记建立后，LSR 仅仅根据标记的值来确定任何携带标记分组所属的 PSC(OA 与 PSC 一一对应)，并使以相应的调度处理，而由定义单个排序聚合体 OA 中包括的行为聚合体 BA 不会多于 8 个，这样 LSR 对携带标记分组采取的丢弃优先级就可以由头标中的 EXP 域决定。总而言之，在 L-LSP 中的 LSR 将根据分组 shim 头标中的标记值和 EXP 域的值综合决定对分组采用的区分服务。使较多等级的区分服务在 MPLS 网络中得以实现。这样的 LSP 之所以被称之为 L-LSP 是因为 IP 分组所属的 PSC 完全由标记来决定，而不需其它信息，比如 EXP 域的值等。

在 L-LSP 中传送分组时：

a)根据标记的内容确定该分组属于哪个 PSC，并建立此 PSC 中不同服务等级与 EXP 域值的映射表。

b)LSR 根据（PSC，EXP）←>PHB 的映射表，确定对此分组采用的 PHB，如调度处理和丢弃优先级等，达到区分服务的目的。

c)转发分组时，将根据以上映射表的逆映射，对转发分组的 EXP 域写入相应值，并由 PSC 对输出分组加上标记。表 1 是一个（PSC,EXP）←>PHB 映射表的例子。

服务	EXP 域	PSC	PHB
缺省	000	DF	DF
优先级转发	000	CSn	CSn
确定型转发	000	AFn	AFn1
	001	AFn	AFn2
	010	AFn	AFn3
加速转发	000	EF	EF

表 1 (PSC, EXP)<->PHB 映射表

3、对标记分发协议 LDP 进行扩展

为了在 MPLS 网络中实现区分服务，需要在建立通道分发标记时 LSR 之间能够交换有关区分服务的信息，因此需要对标记分发协议进行扩展。

LDP 协议用来在 MPLS 网络中分发标记建立 LSP，为了实现对区分服务的支持，主要新增了可选的 LDP TLV(Type-Length-Value)、Diffserv TLV 等项，包含以下内容：

LSP 类型标记，表示要建立的是 E-LSP 还是 L-LSP。

对 E-LSP 包括：在 LSP 中建立的 EXP←->PHB 映射表、在 LSP 的 ILM 包含的与区分服务有关的信息。对 L-LSP 包括要建立的 L-LSP 对应的 PSC 等。

在标记分发处理中也要增加对 Diffserv TLV 的标记分发信息，MPLS 网络中的接收节点将发送 Receive 信息作为应答。同时各个节点路由器将 ILM/FTN 中的区分服务内容（相应的映射表），并且记录对每一个转发标记应采取的转发处理。如果接收节点不能相应发送着的请求，将按照 Status TLV 中定义的出错代码返回错误信息。

MPLS 是在非连接的 IP 网中导入虚通道的协议，来提供 IP 的连接型服务。利用 Shim 头标中的标记域和 EXP 域，使 MPLS 网络能保证多个等级的区分服务质量。

第四章　MPLS 网络端到端 QoS 提供机制

§4.1 介绍

IP 网络的 QoS 研究导致了两种不同体系结构的出现：Intserv 体系结构及其相应的信令协议 RSVP 和 Diffserv 体系结构。根据上一章的研究可知，这两种 IP 网络的 QoS 控制都不能完全满足需要，它们各有自己是长处和局限。

Intserv 存在许多缺点，首先，可扩展性是 Intserv 的最严重的问题。由于使用了“软状态”的工作方式，同时 RSVP 进行资源预留需要对大量的状态信息进行刷新与储存，当网络规模扩大时，这一模型将无法实现。

Diffserv 本身也还不完善。首先它并不提供全网端到端的服务质量保证。另外有关的许多技术细节 IETF 都还未给出具体明确的规定，例如业务类别的具体划分、每类业务性能的量化描述、IP 的业务类别等等。现在 IETF 的 MPLS 和 Diffserv 工作组都在研究 RSVP 与 Diffserv 框架的结合问题以进一步扩大 Diffserv 的与现有系统的可兼容性。

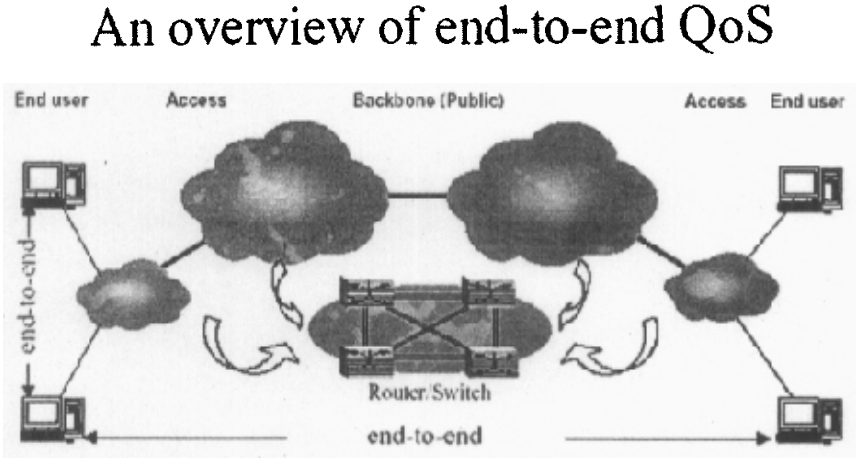

图 4.1 端到端 QoS

为了支持端到端的 QoS，考虑将 Diffserv 与 Intserv 相结合、互相补充、互相协同，实现端到端的 QoS 提供机制，最终达到既能提供类似状态相关网络的强有力的服务，又能实现与状态无关网络近似的可扩展性和鲁棒性。

第四章　MPLS 网络端到端 QoS 提供机制

§4.1 介绍

IP 网络的 QoS 研究导致了两种不同体系结构的出现：Intserv 体系结构及其相应的信令协议 RSVP 和 Diffserv 体系结构。根据上一章的研究可知，这两种 IP 网络的 QoS 控制都不能完全满足需要，它们各有自己是长处和局限。

Intserv 存在许多缺点，首先，可扩展性是 Intserv 的最严重的问题。由于使用了“软状态”的工作方式，同时 RSVP 进行资源预留需要对大量的状态信息进行刷新与储存，当网络规模扩大时，这一模型将无法实现。

Diffserv 本身也还不完善。首先它并不提供全网端到端的服务质量保证。另外有关的许多技术细节 IETF 都还未给出具体明确的规定，例如业务类别的具体划分、每类业务性能的量化描述、IP 的业务类别等等。现在 IETF 的 MPLS 和 Diffserv 工作组都在研究 RSVP 与 Diffserv 框架的结合问题以进一步扩大 Diffserv 的与现有系统的可兼容性。

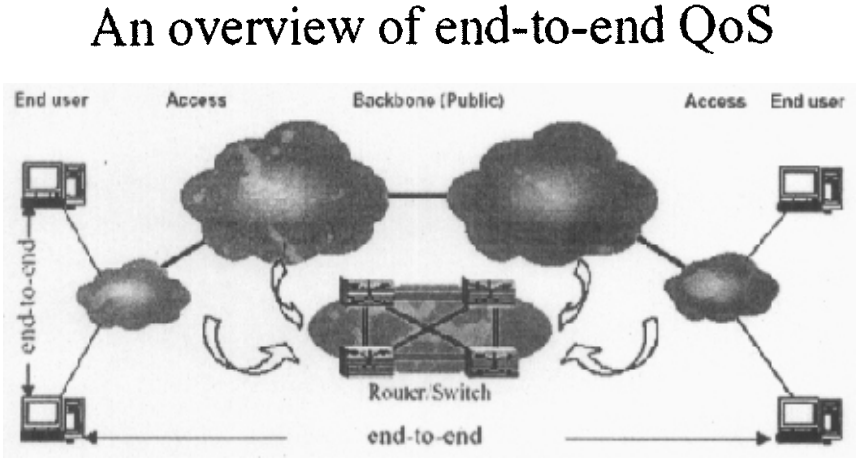

图 4.1 端到端 QoS

为了支持端到端的 QoS，考虑将 Diffserv 与 Intserv 相结合、互相补充、互相协同，实现端到端的 QoS 提供机制，最终达到既能提供类似状态相关网络的强有力的服务，又能实现与状态无关网络近似的可扩展性和鲁棒性。

RSVP、Diffserv、 MPLS 等协议都是在 QoS 管理的粒度和网络可扩展性这两个考虑因素之间寻求不同程度的折衷，RSVP 提供更细的 QoS 保障的粒度 ,而 Diffserv 和 MPLS 具有很好的可扩展性。因而可以将 RSVP 与 Diffserv 协议进行结合。RSVP 与 Diffserv 协议的有机结合和相互补充对 IP 网络 QoS 管理有着重要意义。

Intserv 体系结构提供了一种在异构网络元素之上提供端到端 QoS 的方法。一般来讲，网络元素可以是单独的节点或链路，更复杂的实体（如 ATM 云或 802.3 网络）也可以视为网络元素。在这种意义下，Diffserv 网络也可以视为更大的 Internet 网络中的一种网络元素。（RFC2998）

在该框架中，端到端的、定量的 QoS 是通过在含有一个或多个 Diffserv 区的端到端网络中应用 Intersev 模型来提供的。为了优化资源的分配和支持接纳控制，Diffserv 区可以（并不绝对要求）参加端到端的 RSVP 信令过程。从 Intserv 的角度看，网络中的 Diffserv 区被视为连接 Intserv 路由器和主机的虚电路。

下面研究 Intserv/ RSVP 与 Diffserv 的集成，达到 Intserv/ RSVP、Diffserv、MPL S 三者在设计思想上互相补充。设计一个合适的协作体系使三者结合起来提供可扩展的端到端 QoS 服务。把 Diffserv 中的域当作 RSVP 中的预留节点，同 RSVP 原型的区别在于，域中的节点不必承担流的状态信息，而使 RSVP 只做建立连接和接纳控制方面的工作，减少了复杂性 ,同时又提高了灵活性。

研究的目的为在整个网络中实现点到点的 QoS。解决例如一个 Diffserv 域的入口路由器(BR1)可以配置为从边缘网络(EN1)只接受 10M 的 DSCP 标记为 EF 的低延时、低丢包率服务的数据流量。当 EN1 中产生 20 个 1M 的 MPEG-1 视频流时，BR1 将丢弃一半的输入数据流量，很可能每一个视频流均有被丢弃的数据包。这是因为 Diffserv 是对数据包的聚集进行流量控制，而不是针对一个应用流。在这种情况下，即使 Diffserv 域有足够的资源可以为 10 个视频流服务，但由于没有显式的接入控制，这 20 个视频流没有一个可以得到满意的服务。如果加入显式信令机制，可以拒绝其中 10 个视频流的接入请求，或通知其可选其它 DSCP 对应的服务。

应用资源预约机制 (RSVP)请求资源，网络可根据当前可用资源情况作出是否提供服务的决定并通知应用程序。这种动态资源分配方式对资源的利

用率较高。而在 Diffserv 网络中，接入控制是以服务水平约定(SLA)这种半静态的方式进行，只有对数据包的聚集进行流量控制的能力，而没有信令的传递。因此，下面研究 RSVP 与 Diffserv 的结合，达到 RSVP、Diffserv、MPLS 三者在设计思想上互相补充。设计一个合适的协作体系使三者结合起来提供可扩展的端到端 QoS 服务。

为达到以上目的，以下问题需要解决

1) Intesev 服务类型到 Diffserv 网络提供的服务之间的映射；
2) 定义 Diffserv 网络区内使用聚集传输控制的网络元素在支持 RSVP 信令时所需的功能；
3) 定义在Diffserv网络区内有效、动态的资源提供机制(如聚集RSVP，隧道，MPLS 等，以及边界代理 BB 如何将 Diffserv 区内的资源可用性信息传递到边界路由器的定制协议。

§4.2 实现方法

Intserv提供了一种在异构网络元素之上提供端到端QoS方法。一般而言，网络元素可以是单独的节点(如路由器)或链路，更复杂的实体（ATM 云)也可从功能视为网络元素，在这种意义上来说，Diffserv 网络云也可视为巨大的 Intserv 网络中的一种网络元素。从 Intserv 的角度看，网络中的 Diffserv 区被视为连接 Intserv 路由器和主机的虚链路。另一方面、Diffserv 与其它服务质量保障技术(如 Intserv/RSVPMPLS、ATM 等)也有极好的兼容件。况且.两者在体系结构上存在相似和共同之处，它们都需要进行服务与流特性的描述机制，都是通过对流量进行控制实现不同等级的服务特性，都需要一种按分组头中一些域进行流分类的过程、都需要其它性能支持机制的配合、如 QoS 路由机制、基于测量的接纳控制机制等等。

4.2.1 实现框架

在该实现方案中，整个网络是 Intserv 节点(采用基于 MF 的分类和其于流的传输控制)和 Diffserv 区(采用聚集传输控制)的结合体。其参考网络框架的简化模型如图 4.2 所示。

用率较高。而在 Diffserv 网络中，接入控制是以服务水平约定(SLA)这种半静态的方式进行，只有对数据包的聚集进行流量控制的能力，而没有信令的传递。因此，下面研究 RSVP 与 Diffserv 的结合，达到 RSVP、Diffserv、MPLS 三者在设计思想上互相补充。设计一个合适的协作体系使三者结合起来提供可扩展的端到端 QoS 服务。

为达到以上目的，以下问题需要解决

1) Intesev 服务类型到 Diffserv 网络提供的服务之间的映射；
2) 定义 Diffserv 网络区内使用聚集传输控制的网络元素在支持 RSVP 信令时所需的功能；
3) 定义在Diffserv网络区内有效、动态的资源提供机制(如聚集RSVP，隧道，MPLS 等，以及边界代理 BB 如何将 Diffserv 区内的资源可用性信息传递到边界路由器的定制协议。

§4.2 实现方法

Intserv提供了一种在异构网络元素之上提供瑞到端QoS方法。一般而言，网络元素可以是单独的节点(如路由器)或链路，更复杂的实体（ATM 云)也可从功能视为网络元素，在这种意义上来说，Diffserv 网络云也可视为巨大的 Intserv 网络中的一种网络元素。从 Intserv 的角度看，网络中的 Diffserv 区被视为连接 Intserv 路由器和主机的虚链路。另一方面、Diffserv 与其它服务质量保障技术(如 Intserv/RSVPMPLS、ATM 等)也有极好的兼容件。况且.两者在体系结构上存在相似和共同之处，它们都需要进行服务与流特性的描述机制，都是通过对流量进行控制实现不同等级的服务特性，都需要一种按分组头中一些域进行流分类的过程、都需要其它性能支持机制的配合、如 QoS 路由机制、基于测量的接纳控制机制等等。

4.2.1 实现框架

在该实现方案中，整个网络是 Intserv 节点(采用基于 MF 的分类和其于流的传输控制)和 Diffserv 区(采用聚集传输控制)的结合体。其参考网络框架的简化模型如图 4.2 所示。

用率较高。而在 Diffserv 网络中，接入控制是以服务水平约定(SLA)这种半静态的方式进行，只有对数据包的聚集进行流量控制的能力，而没有信令的传递。因此，下面研究 RSVP 与 Diffserv 的结合，达到 RSVP、Diffserv、MPLS 三者在设计思想上互相补充。设计一个合适的协作体系使三者结合起来提供可扩展的端到端 QoS 服务。

为达到以上目的，以下问题需要解决

1) Intesev 服务类型到 Diffserv 网络提供的服务之间的映射；
2) 定义 Diffserv 网络区内使用聚集传输控制的网络元素在支持 RSVP 信令时所需的功能；
3) 定义在Diffserv网络区内有效、动态的资源提供机制(如聚集RSVP，隧道，MPLS 等，以及边界代理 BB 如何将 Diffserv 区内的资源可用性信息传递到边界路由器的定制协议。

§4.2 实现方法

Intserv提供了一种在异构网络元素之上提供瑞到端QoS方法。一般而言，网络元素可以是单独的节点(如路由器)或链路，更复杂的实体（ATM 云)也可从功能视为网络元素，在这种意义上来说，Diffserv 网络云也可视为巨大的 Intserv 网络中的一种网络元素。从 Intserv 的角度看，网络中的 Diffserv 区被视为连接 Intserv 路由器和主机的虚链路。另一方面、Diffserv 与其它服务质量保障技术(如 Intserv/RSVPMPLS、ATM 等)也有极好的兼容件。况且.两者在体系结构上存在相似和共同之处，它们都需要进行服务与流特性的描述机制，都是通过对流量进行控制实现不同等级的服务特性，都需要一种按分组头中一些域进行流分类的过程、都需要其它性能支持机制的配合、如 QoS 路由机制、基于测量的接纳控制机制等等。

4.2.1 实现框架

在该实现方案中，整个网络是 Intserv 节点(采用基于 MF 的分类和其于流的传输控制)和 Diffserv 区(采用聚集传输控制)的结合体。其参考网络框架的简化模型如图 4.2 所示。

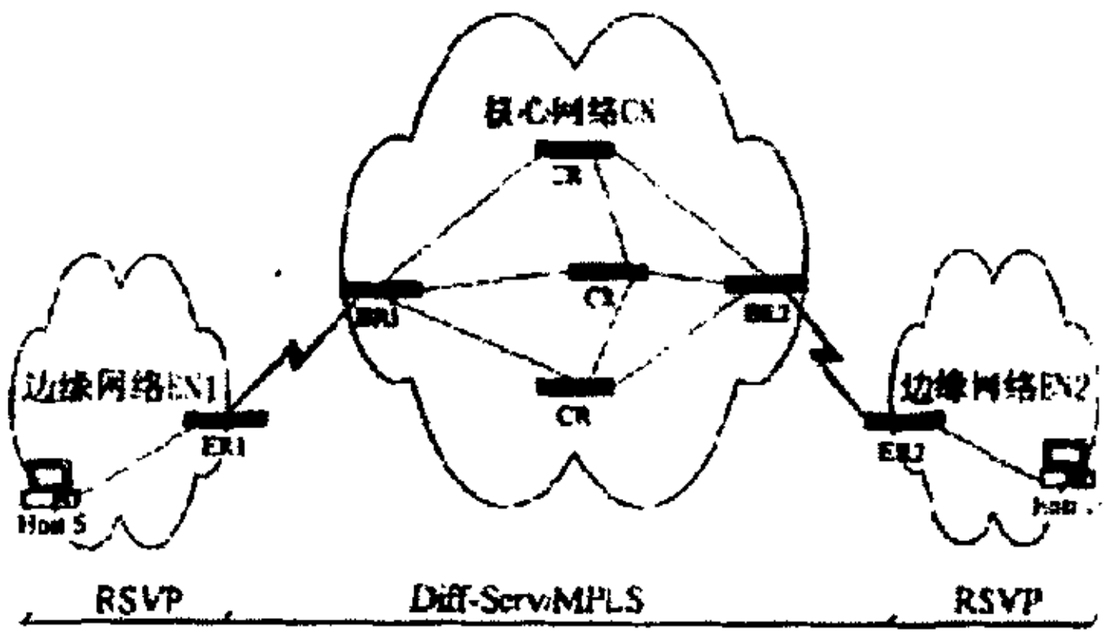

图 4.2 Diffserv 网络区支持端到端 Intserv 的实现框架

IntServ/ RSVP 与 DiffServ 的集成方式，这里只考虑一个 QoS 发送者 S 与一个 QoS 接收者 D 通过网络进行通信，邻近 DS 区域的边缘路由器 ER1、ER2 与 DS 区域内部的边界路由器 BR1、BR2 通过接口直接相连，S 和 D 主机都使用 RSVP 来传送主机应用的定量 QoS 请求。

主机操作系统的 QoS 进程生成 RSVP 信令，RSVP 消息在主机 S 和 D 之间端到端地传播以支持 DS 区外部的 RSVP 预留，端到端的 RSVP 信令至少应被透明地传送 DS 域。

EB1、ER2 和 BR1、BR2 的功能依赖于该框架的具体实现。若 DS 区域不识别 RSVP，则 ER1、ER2 作为 DS 区域的接纳控制代理处理来自 S 和 D 的 RSVP 信令消息，根据 DS 区域内部的资源信息和客户定义的策略实施动态接纳控制，但是资源管理却是静态方式，BR1、BR2 只作为纯粹的 Diffserv 路由器。而在 DS 区识别 RSVP 情况下，ER1、ER2 根据当地的资源情况和客户定义的策略实施接纳控制，而 BR1、BR2 参加 RSVP 信令过程并作为 DS 区域的接纳控制代理，同时完成动态接纳控制和动态资源管理。

为了支持上述端到端的 QoS 提供机制的集成框架。DS 网络区域必需满足以下需求：

(1) DS 区域的边界节点须能执行适当的管理、控制(如包括整形、重标记及策略等)；

(2) DS 区域的边界节点 BR 之间须能作为标准的 Intserv QoS 服务提供支持，而在 DS 区域内部使用际准的 PHB 来调用这些服务；

(3) DS 区域须基于资源可利用性，为非区分服务网络接纳控制机制；

(4) 此区域内节点至少应能透明传递 RSVP 消息，以便在 DS 区域出口可重新获取这些消息。

4.2.2 RSVP 协议在 Diffserv 区中实现显式接纳控制

1）显式接纳控制

在 Diffserv 网络中，接入控制是以服务水平约定(SLA)这种半静态的方式进行，只有对数据包的聚集进行流量控制的能力，而没有信令的传递。

例如DS区人口的某个网络元素对EF DSCP只允许接受50Kbps的传输。这时若有一聚集流要求带宽超过 50Kbps，则整个聚集流将被拒绝，即使其中共些微流只要求 10Kbps 也同样被拒绝。可见隐式接纳控制能够在某种程度上对网络起到保护作用，但同时效率低并会破坏瑞到端显式接纳控制的有效性。

为此借鉴 Intserv 网络中实现量化 QoS 应用显式接纳控制的方法，采用显式信令 RSVP 从网络请求资源，网络作出接受或拒绝的响应。这样便可保证对接纳的传输流实现资源预留(以损害未被接纳的流为代价)，因此为 DS 网络区指定一个支持 Intserv 的接纳控制代理可以优化资源利用，提高 DS 区对于定量 QoS 应用的服务质量。

在 DS 网络区采用 RSVP 接纳控制代理还可以实现基于策略的接纳控制。因为 RSVP 资源请求可以被识别 RSVP 的网络元素采取，并按照策略数据库的策略进行检查。由于资源请求标识了其所代表的用户和应用，所以在进行接纳控制时，网络元素可考虑基于每个用户或每个应用的策略。否则在没有 RSVP 信令的 DS 网络区，策略只能作用于发起传输的 DS 客户网络，而不是客户网络中的某个传输发起用户或应用。

2) 动态资源管理

通过 DS 区内部选定某些设备参加 RSVP 信令过程来实现，但值得提出的是，虽然这些路由器参加某种形式的 RSVP 信令。但它们仍使用 IP 分组头的 DSCP 值对聚集传输流进行识别、分类和调度，而不象 Intserv/RSVP 路由器使用基于流的 MF 分类准则。当一个新流要加入某行为聚集(BA: Behaviour Aggregate)时，就使用动态提供机制和显式信令进行接纳控制。此

(3) DS 区域须基于资源可利用性，为非区分服务网络接纳控制机制；

(4) 此区域内节点至少应能透明传递 RSVP 消息，以便在 DS 区域出口可重新获取这些消息。

4.2.2 RSVP 协议在 Diffserv 区中实现显式接纳控制

1）显式接纳控制

在 Diffserv 网络中，接入控制是以服务水平约定(SLA)这种半静态的方式进行，只有对数据包的聚集进行流量控制的能力，而没有信令的传递。

例如DS区人口的某个网络元素对EF DSCP只允许接受50Kbps的传输。这时若有一聚集流要求带宽超过 50Kbps，则整个聚集流将被拒绝，即使其中共些微流只要求 10Kbps 也同样被拒绝。可见隐式接纳控制能够在某种程度上对网络起到保护作用，但同时效率低并会破坏瑞到端显式接纳控制的有效性。

为此借鉴 Intserv 网络中实现量化 QoS 应用显式接纳控制的方法，采用显式信令 RSVP 从网络请求资源，网络作出接受或拒绝的响应。这样便可保证对接纳的传输流实现资源预留(以损害未被接纳的流为代价)，因此为 DS 网络区指定一个支持 Intserv 的接纳控制代理可以优化资源利用，提高 DS 区对于定量 QoS 应用的服务质量。

在 DS 网络区采用 RSVP 接纳控制代理还可以实现基于策略的接纳控制。因为 RSVP 资源请求可以被识别 RSVP 的网络元素采取，并按照策略数据库的策略进行检查。由于资源请求标识了其所代表的用户和应用，所以在进行接纳控制时，网络元素可考虑基于每个用户或每个应用的策略。否则在没有 RSVP 信令的 DS 网络区，策略只能作用于发起传输的 DS 客户网络，而不是客户网络中的某个传输发起用户或应用。

2) 动态资源管理

通过 DS 区内部选定某些设备参加 RSVP 信令过程来实现，但值得提出的是，虽然这些路由器参加某种形式的 RSVP 信令。但它们仍使用 IP 分组头的 DSCP 值对聚集传输流进行识别、分类和调度，而不象 Intserv/RSVP 路由器使用基于流的 MF 分类准则。当一个新流要加入某行为聚集(BA: Behaviour Aggregate)时，就使用动态提供机制和显式信令进行接纳控制。此

时，可以使用 RSVP 信令将流的描述和期望的 DSCP 传送 DS 区的路由器。可以说，DS 区路由器的控制平面是 RSVP 而数据平面仍是 Diffserv。这种方案既充分利用了 RSVP 信令的优越性．又保持了 Diffserv 的可扩展性。相比之下，若 DS 区内不含有能够识别 RSVP 的设备，则网络中的 DS 区以静态方式提供内部的资源管理。这时 DS 网络区的客户和网络所有者之间建立一种静态的契约——服务类型协定 SLA，保证在每个标准的 Diffserv 服务类型上向客户提供应有的传输能力。DS 区和其外部的网络元素之间没有信令，它们之间的 SLA 协商是唯一的关于资源可利用性信息的交换形式。在 DS 网络区的接纳控制代理配置 SLA 所表示的信息，这种方案灵活性差、不容易支持 SLA 的动态改变、因为 SLA 每次改变都需要重新配置接纳控制代理。另外，DS 网络区的资源也难以有效利用，因为接纳控制无法充分反映 DS 区中常受冲击的路径上的资源可利用性。

4) 此外，RSVP 在 DS 网络区的传输识别和分类中也有辅助作用。如 DSCP 标记可有两种实现机制：主机标记和路由器标记。

主机标记要求主机知道网络如何翻译 DSCP，这类信息可配置于每个主机，但会加重管理负担，而如果采用 RSVP 作为显式传令协议通过询问网络来获取是一种较好的方案，此时只要在 RSVP 的 Resv 消息中增加一 DCLASS 对象即可：其格式如下：

表 2

长度（>=8 字节）	C---Num(225)	1
保留	DSCP1	CE
保留	DSCP2	CE
…		

长度：是指整个 DCLASS 对象字节数，

c—Nums：DCLASS 对象的分类号为 255

一个 DCLASS 对象可根据需要包含多个 DSCP，个数为(长度—4) / 4。

第二种实现机制为路由器标记，这时要求在路由器配置 MF 分类准则，这可由主机操作系统通过请求动态完成，由手工配置或自动脚本完成。然而，静态配置困难很大，一种更好的选择是允许主机操作系统代表用户和应用通过信令 RSVP 将 MF 分类准则发送给路由器。所以 RSVP 为在 DS 网络区中实现网络资源优化可起到很大的作用，当然其本身固有的复杂性、不可扩展性都要求它在 DS 区中应用时需对其进行改进和简化。

4.2.3 集成服务到区分服务的映射

Intserv/ RSVP 与 DiffServ 的集成就是将 DS 域看成是 IntServ 端到端路径上的一个子网（network elements），集成的关键问题是如何为端系统产生的数据包打上合适的 DSCP 标记，即如何将 IntServ 服务的传输需求映射到DiffServ的PHB上，如何分别用Diffserv的Assured Forwarding(AF)和 Expedited Forwarding(EF) PHB 实施 Intserv Controled Load Service(CLS)和 Guaranteed Service(GS)的方式。这种映射不仅依赖于 DiffServ 网络的拓扑结构，而且必须考虑网络中通信流量特征和网络管理策略，因此相应参数（如 Guaranteed Service 中“C”和“D”参数）的计算是离线进行的。计算结果广播到 DiffServ 网络中所有的 BR(Border Router)和与之相连的 ER (Edge Router)中，然后使用 RSVP 与端系统交互进行接入控制并通知其应使用的 DSCP 标记。

IS 到 DS 的映射主要包括建立服务类之间的映射关系和标识（mark)数据包。映射是标识的基础，标识决定数据包对应的 PHB，不同 PHB 的 QoS 保证也是不同的。

我们把数据包的标识分为 IS 域预标识和 DS 域标识两步。IS 域预标识就是在 IS 域向数据包的 TOS(type of service)域写入反映数据包所属流的特性和 QoS 需求差异的预标识码。标识把数据包映射为有限个 PHB,它屏蔽了不同数据流的特性及 QoS 需求的差异。

在标识时标识控制算法依据预标识码控制标识过程，可以充分地考虑流特性的差异，数据包会更准确地映射成 PHB。利用数据包携带预标识码不增加 IS 域和 DS 域之间的通信量、实现代价小这些特点。

1、Internet 服务模型

1） 服务模型

由集成服务和区分服务构成的 Internet 网络服务模型如图 4.2 所示。ER1、ER2 是 IS 域支持 RSVP 的边缘路由器，BR1、BR2 是 DS 域支持区分服务的边界路由器,ER1 和 BR1 的关系可以是 ISP 与 ISP、企业网路由器和 ISP 之间的关系。

2） 集成服务的特征

Intserv 提供的服务分为 3 类 :保证服务 GS(guaranteed service)、

控制负载服务 CLS(control load service)和尽力转发服务 BE(best- effort service)。

不同的服务根据消息包中的服务编码进行辨识：根据数据流的 ID 值和 IP 地址等特征信息的关系辨认不同流的数据包。数据包的处理过程包括资源预约和数据包传递。资源预约使用 RSVP 实现。

FLOWSPEC 是包含在预留状态块 RSB(reservation state block)中的重要的数据对象之一。它的主要参数有数据流的峰值速率 P、令牌速率 r、令牌桶的大小 b 和最大数据包的体积 M,对于保证型服务 ,还有请求带宽 R 和延迟裕度 S。

3） 区分服务特征

典型的服务分类方法把区分服务分成 3 类 ,即奖赏服务 (premiun service,简称 PS)、确保服务 (asurred service,简称 AS)、尽力转发服务 BE。PS 由 EF(expedited forwarding)PHB 实现 ,AS 由 AF(assured forwarding)PHB 实现。

DS 结点在数据包的 IP 头的 TOS 域写入不同的编码 ,将使数据包对应不同的 PHB,数据包将得到不同的 QoS 保证 ,向 TOS 写编码的过程称为包标识。

EF PHB 组只有 1 个 EF PHB,DS 结点按 PS 的峰值速率要求给 EF PHB 分配资源。

AF PHB 组分为 4 个独立的类 ,每类又按丢弃优先级分为 3 级 ,共 1 2 个 PHB。每个 AF PHB 类独立于其他类来接受资源分配和调度 ,每个 AF PHB 要求最低的资源保证 ,同时可以使用其他类 AF PHB 或 EF PHB 剩余的资源。某个服务类的数据包可以使用一个或多个 AF 类数据包的处理过程有流量调节和 PHB 调度。包分类和包标识是流量调节功能的关键。

2、Intserv 到 Diffserv 的映射

在 Intserv 中，服务请求是以 RSVP Resv 消息中 Flowspec 对象来描述的，其中包括了 Intserv 服务类型和参数。Intserv 网络中每一网络元素都须将次服务请求映射成本地链路层介质相应地描述，DS 区域作为 Intserv 中的一个网络元素也应完成服务之间的映射，从而选择相应的 DSCP。

目前，Intserv 中支持三种业务流类型：保证服务(GS)、控制负荷(CLS)服务以及尽力而为(BE)服务，下面一一进行讨论：

1）GS 的映射　GS 能为符合该类型的分组提供专用的带宽和限定的时延，不会因队列溢出而丢失分组。它是惟一支持可计量 QoS 参数(时延)的业务流类型，它能够采用和收益于 GS 的应用包括高质量会议电话、实时财政事务处理等，它们对带宽和时延有极为严格的需求。GS 可采用 EF PHB 来实现，同时辅以整形和策略功能。若 DS 网络域内还存在其它 EF 业务，为严格确保然业务延时在最大限值之内。须采用缓冲及调度机制、分开 GS 用的 EF PHB 与其他业务用的 EF PHB。

2) CLS 的映射 CLS 业务是一种在轻负荷网络上类似于尽力而为性质的网络服务，它并不保证带宽和延时，但也不像 BE 服务那样性能能会在网络负载加重时严重下降。CLS 是为能容忍一定数量时延和丢失的应用而设计的、如中低质量的音频视频组播，根据数据流是否遵循其 Tspec 描述可分为两大类：一致流 CLS 和不一致流 CLS。网络元素在转发前者时应保证其时延。

CLS 业务在 DS 区域内可选择用以下两种方法：

a. 采用 AF PHB 来实现：

CLS 可根据 B/R 值(CLS 业务流可用带宽 B 和缓冲区大小 R 两个参数来表示）分成不同延迟优先级，对每一延迟优先级用聚合 A-Tspec 来描述其中已按纳业务流的特征。这个 A-Tspec 作为 DS 区域人口处对业务进行策略控制的依据，从而分离出哪些是一致流 CL 分组和哪些是不一致流 CL 分组，然后对它们进行 DSCP 标记。前者对应的 AF PHB 优先级最高，而后者对应最低。DS 区域内每个节点都将恰当配置每个 AF PHB：设置实际队列大小以限制队列时延；丢弃已超过延迟优先级时延限制的分组；为低优先级分组没置丢弃参数；设置该 AF PHB 的服务速率以便有足够的带宽来满足在仅有高优先级 CL 业务流通过时它们对时延及丢失宰的要求；最后实现接纳控制算法。可见每一延迟优先级 CLS 业务流均将映射成一单独的 AF PHB。这种方法最好。

b. 采用 EF PHB 来实现：

这只有在前一种方法不可用时才考虑的方案。因为 EF PHB 无优先级之分，所有 CLS 业务均将映射成问一 EF PHB，无法区分 CLS 业务中不同的延迟优先级，这就要求为其分配充裕的资源来满足其中要求最高的 CLS 流，显然没有最优地利用网络资。此外，采用 FF PHB 也无法很好处理不一致流

的 CLS。因为 EF PHB 实现用的是硬性指标，这样只得在入口处将它们丢弃或是重标记为 BE PHB 对应的 DSCP 值（但这会导致乱序）。

3) BE 的映射，采用 BE PHB 来实现

表 3 应用类、IS 服务类、DS 服务类、PHB 映射关系表

应用类	Intserv				Diffserv	
	服务类	CLASS	PRIORITY	LEVEL	服务类	PHB
高保真实时声频视频	GS	1	11	000-111	PS	EF PHB
实时声频视频			10	000-111		
非实时声频视频			01	000-111		
速率可控应用	CLS	0	11	000－111	AS	AF11－12
			10	000－111		AF21－22
			01	000－111	AS	AF31－32
			00	000－111		AF41－42
传统应用	BE	0	00	000	BE	

3、IS 域预标识

预标识在 IS 域完成，主要是确定和写入预标识码。预标识码的大小说明标识和转发的优先级的高低，DS 域的正式标识码将覆盖预标识码。

1) 预标识码的确定

GS 的 CLASS,PRIORITY 和 CLS 服务类的 CLASS 根据应用类型确定（见表 3)。GS 的 LEVEL 和 CLS 的 PRIORITY 、LEVEL 取决于同类应用中各数据流的相对特性。

用户特性取决于用户 QoS 需求的特殊性和付费等因素，作为预标识码因子。反应带宽需求的预标识码因子是令牌速率 (CLS 类流)或带宽(GS 类流)；突发性好 (K 值大)的流有快速增大数据流量的效果，所以突发特性也作为预标识码因子。GS 的 LEVEL 分为 U、R、Pr, CLS 中 PRIORITY 对应于 U 位，LEVEL 对应于 r 和 Pr,具体定义见表 4 和表 5。

表 4 GS 类服务的 LEVEL 定义

用户特性位	带宽需求位	突发特性位
U 意义	*R* 意义	*Pr* 意义
1 特殊用户	1 $R >= R_{mean}$	1 $K >= K_{mean}$
0 其他用户	0 $R < R_{mean}$	0 $K >= K_{mean}$

表 5　CLS 的 PRIORITY 和 LEVEL 的定义

用户特性位		令牌速率位		突发特性位	
U	意义	*R*	意义	*Pr*	意义
11.10	特殊用户	11	$R >= R_{mean}$	1	$K >= K_{mean}$
01.00	其他用户	10	$R < R_{mean}$	0	$K >= K_{mean}$

在 IPv6 中，利用“流标签”数据位较多的特点，对数据流参数完全可以进行量化编码。预标识码保存在预留状态块(RSB)中，预约消息 (Resv) 的刷新使预留状态块(RSB)中始终保存最新的码值。

2)　数据包的预标识

数据包的预标识可以在主机或边缘路由器(ER1)上进行，我们选择在端主机上完成预标识。因为端主机充分了解应用的特性，这样也可以减轻路由器的负担。端主机的 QoS 代理从预留状态块(RSB)中获取预标识码，把预标识码写入数据包的 IP 头的 TOS 中。

4、DS 域标识

1)　影响标识的相关因素

EF PHB 要求流入结点的信息流量小于流出结点的信息流量；AF PHB 只要求保证最小的带宽，但并不拒绝使用更多的带宽。这些都表明要有流量控制。当一个数据包被标识成某个 PHB 后，其资源分配和调度是 DS 域的职责，我们只研究它们被标识之前的流量控制。通过流量控制满足 EF PHB 和 AF PHB 的不同的 QoS 要求，实现带宽在多类数据流间的公平分配，提高链路利用率。

某一输出链路带宽为 B，某数据流的目标带宽 Ri 由公平分配获得的带宽 b 和处于高优先级获得的带宽 ri 构成。当有 N 个数据流经过该链路时，链路带宽的分配满足如下关系：

$$r_i + b = Ri$$

$$\sum_{i=1}^{N} r_i + Nb = B$$

当链路带宽充裕时，多个信源按需要分配带宽，r_i=0。若 b=0，则目标带宽均是通过特殊方法获得的(提高优先级、增大流量窗口)，低优先级的应用则可能被“饿”死。保持 ri 和 b 的合理比值可提高链路的流量和资源利用率。为体现带宽分配的公平性，网络管理者可确定每类数据流可使用的带宽的上、下限。

每次新接纳并准备标识的数据包应尽可能使流量快速增大，但不能在网中产生特别大的突发性数据流。这将反映标识算法对链路的负载变化的适应性和敏捷性。

2） 标识原理

流量窗口是确定时间内链路上信息的传输速率,每当一个数据包进入 DS 结点时，若检测到的数据包的传输速率低于其目标速率，则根据每类数据流的流量窗口目前的大小，有限制地增大它们的流量窗口。优先增大 PS 类的流量窗口，其次是 AS 类的，再次是 BE 类的。与此同时，标识器依据流量窗口目前的数值尽可能接收更多的数据包或数据位，按数据包的预标识码查表 3 得到标识码(标识码和 PHB 是一一对应的),对数据包进行标识。优先选择 PS 类的数据包进行标识，对于同类服务中的数据包，优先选择预标识码大的数据包进行标识。当数据流的目标速率得到满足时，逐步降低高优先级服务类的数据流窗口大小，使其他类的数据包能及时被标识和转发。

5、信令传递过程如下 ：

信令传递过程如下 ：

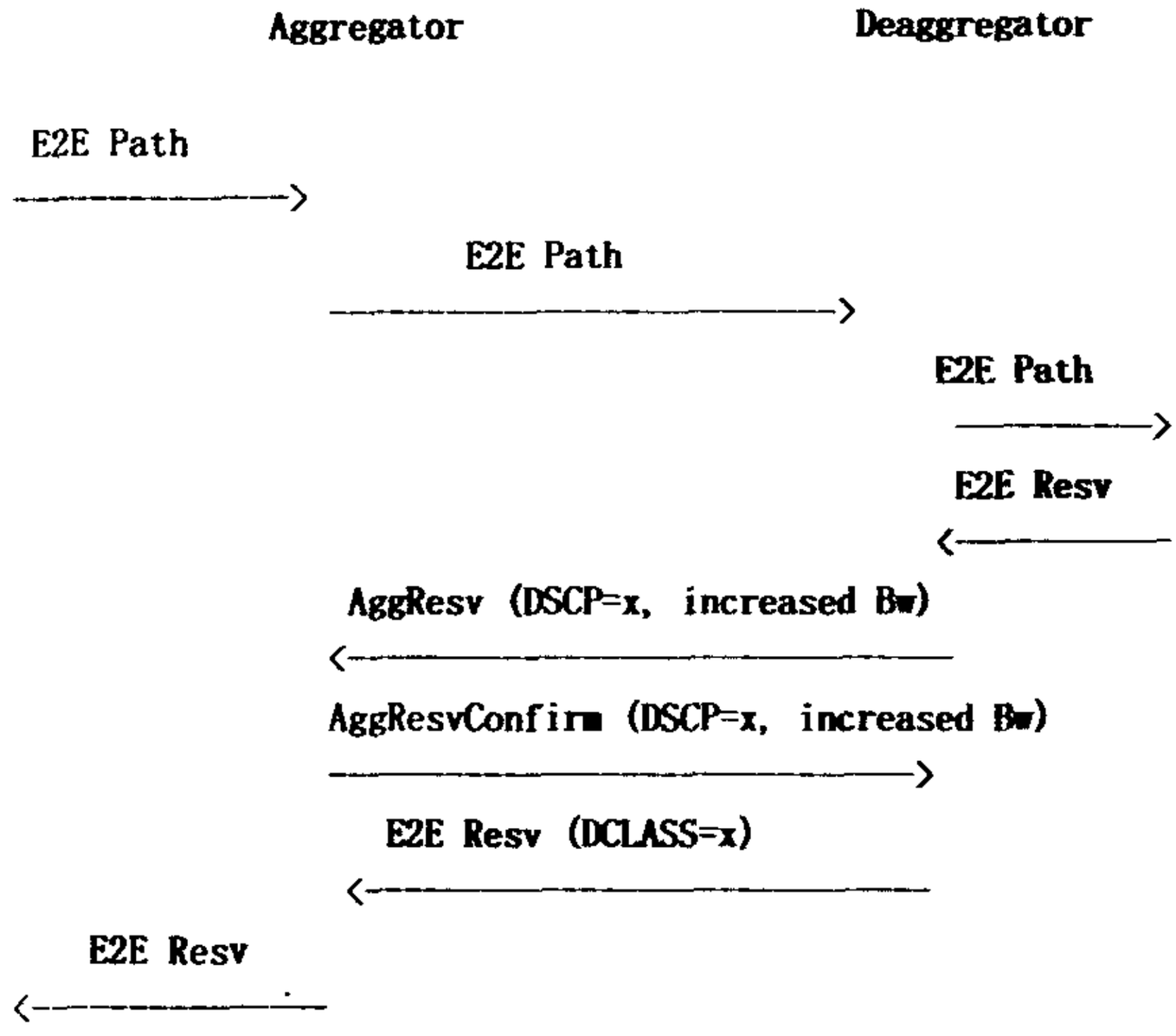

图 4.3 信令传递过程

1）端系统 S 构造一标准 RSVP PATH 消息，向 ER1 发送，在 ER1 建立关于此应用流的状态信息；

2）为保持可扩展性，Diffserv 域中的路由器不装 RSVP 模块 ，即只将 ER1 发来的 PATH 消息透明地传输到 ER2 ，而不记录应用流的状态信息；

3）ER2 处理 PATH 消息并建立此应用流的状态信息后，将 PATH 消息转发至目的端系统 D；

4）在 D 处构造一标准 RSVP RESV 消息，发起资源预约，并沿反向路径向 S 传播；

5)R2 接收到 RESV 消息后，根据本地下行链路资源情况进行接入控制，假设其接受预约，则将 RESV 消息通过 DiffServ 域透明地传输到 ER1；

6）ER1 根据 IntServ 服务传输需求到 DiffServ PHB 的映射关系决定使用何种水平的服务传输应用的数据包。在 ER1 和 BR1 之间存在一个服务水平约定(SLA)，规定了各种水平服务的允许传输流量；ER1 可以根据此 SLA

进行接入控制，在前述例子中就可以拒绝 10 个视频流使用 ER 服务的接入请求。对允许接入的应用流，在 RESV 消息中插入一个 DCLASS 对象，指明对应于应使用的 PHB 的 DSCP 值。

7）预约成功的应用将数据包 IP 分组头中的 DSCP 值设为 RESV 消息中指定的值，开始传输。

DiffServ 体系允许数据包在 Host S，ER1 域 BR1 处标记 DSCP 值。使用 DCLASS 对象使数据源 S 能对 DSCP 进行适当设置，从 DiffServ 服务的角度来看，这比在 ER1 或 BR1 处设置有两个好处：

1） 与 ER1 或 BR1 相比 ，S 有足够信息能够对不同应用类型产生的数据包的传输特性和优先级做出适当的判断；

2） 由于在单个节点的分类规则相对简单，在 S 处进行分类标记要比在 ER1 或 BR1 处聚集后再分类要简单有效得多。

在这个集成方案中，并不要求 DiffServ 域中的路由器参与 RSVP 信令交互，而只在边缘网络 EN1、EN2 中实现 RSVP 显式接入控制，这符合 IP 网络将复杂性限制在边缘，而使核心部分尽可能简单的一贯原则。如果 DiffServ 域也有部分路由器参与 RSVP 的信令交互，则资源的管理将从基于 SLA 的半静态方式转变为动态方式，利用率将大为提高。

第五章　网络仿真技术

网络仿真是一种利用数学建模和统计分析的方法模拟网络行为，从而获取特定的网络特性参数的技术。数学建模包括网络建模（网络设备、通信链路等）和流量建模两个部分。网络仿真获取的网络特性参数包括网络全局性能统计量、网络节点的性能统计量、网络链路的流量和延迟等，由此既可以获取某些业务层的统计数据，也可以得到协议内部的某些特殊的参数的统计结果。

网络仿真技术有两个显着的特点：

1)、首先，网络仿真能够为网络的规划设计提供可靠的定量依据。网络仿真技术能够迅速地建立起现有网络的模型，并能够方便地修改模型并进行仿真，这使得网络仿真非常适用于预测网络的性能。

2)、其次，网络仿真能够验证实际方案或比较多个不同的设计方案。在网络规划设计过程中经常出现多个不同的设计方案，它们往往是各有优缺点，仅凭主观判断，很难作出正确的选择。

目前世界上的网络仿真软件可以分为高端和低端两类产品。高端产品一般具有复杂的建模机制、比较完备的模型库、完善的外部接口、强大的功能并能够得到比较可靠的仿真结果，价位一般在数万美元/每个用户左右，其主流产品基本上都来自美国公司，例如 MIL3 公司的 OPNET、CACI 公司的 COMNET、UC Berkeley ns 等。低端产品一般只有简单的建模机制、较小的模型库、简单的外部接口，功能单一且仿真结果的可靠性较差，比较知名的产品也大都产自美国，例如 SES 公司的 Strategizer。

由于不同产品的定位不同、采用的仿真技术也有很大差异，因此呈现出不同的特点，也有其各自不同的适用领域。例如，COMNET 采用数学分析模拟方法，仿真效率很高，但是无法得到有关网络和协议细节的结果。因此，COMNET 适用于网络高层性能的仿真。MIL3 公司的 OPNET 综合采用基于包的建模方法和数学分析的建模方法，既可以得到非常细节的模拟结果，也可以获得比较快的仿真计算速度。UC Berkeley ns 是美国加州的 Lawrence Berkeley 国家实验室于 1989 年开始开发的软件，简称 ns。ns 从 S. Keshav's REAL 仿真器发展而来。目前 ns 在 Virtual InterNetwork Testbed（VINT）项

目的支持下由南加州大学、施乐公司、加州大学与 Lawrence Berkeley 国家实验室协作发展。目前最高版本为 ns2。可以应用于多种工作平台，Windows NT，UNIX，Sun，Linux 等，则特别适用于网络层，传输层及以上层的模拟仿真。我们应用其做有关 MPLS 标签分配协议的仿真。

§5.1 IP 业务的特点

IP 业务在未来的主导地位已毋庸置疑，因此充分理解 Internet 业务的特点对未来网络的设计是至关重要的。主要有互联网分析协会 CAIDA（Cooprative Association for Internet Data Analysis）和 IETF 的互联网协议性能工作组 IPPM WG（Internet Protocol Performance and Practices Measurement）在这些方面所做的工作。

1. Internet 业务的突发性（碎片性）

大量的研究表明 Internet 业务在本质上呈现突发性和自相似性。自相似性就是在一个给定的物理链路上，在不考虑通信量大小的情况下，Internet 网络上的业务呈现出基本相同的特性。

图 5.1 给出了 Internet 业务的碎片性和呈泊松分布的话音业务的特性比较。

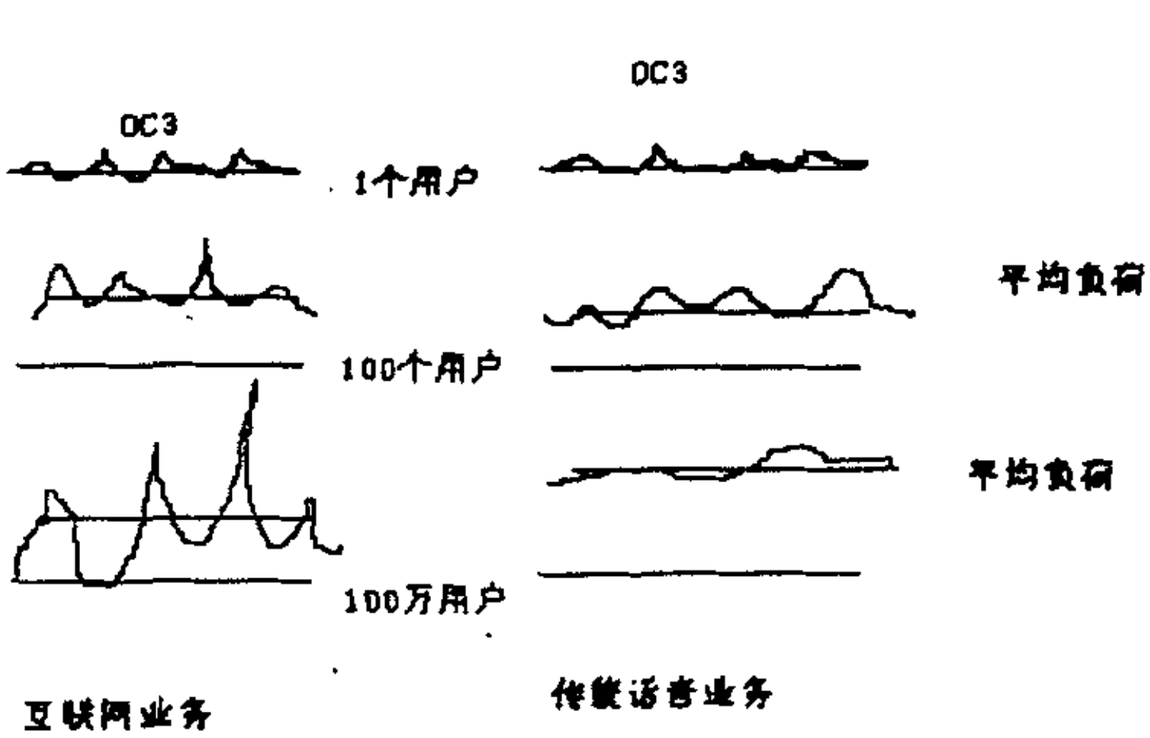

图 5.1 Internet 网络的业务特性与话音网络的业务特性对比

2. 发送和接收数据的非对称性

Internet 数据的另一特性就是在许多 Internet 链路上的发射和接收通道上带宽占用量存在非对称性。这种数据流的非对称性是由于大型服务器要

目的支持下由南加州大学、施乐公司、加州大学与 Lawrence Berkeley 国家实验室协作发展。目前最高版本为 ns2。可以应用于多种工作平台，Windows NT，UNIX，Sun，Linux 等，则特别适用于网络层，传输层及以上层的模拟仿真。我们应用其做有关 MPLS 标签分配协议的仿真。

§5.1 IP 业务的特点

IP 业务在未来的主导地位已毋庸置疑，因此充分理解 Internet 业务的特点对未来网络的设计是至关重要的。主要有互联网分析协会 CAIDA（Cooprative Association for Internet Data Analysis）和 IETF 的互联网协议性能工作组 IPPM WG（Internet Protocol Performance and Practices Measurement）在这些方面所做的工作。

1. Internet 业务的突发性（碎片性）

大量的研究表明 Internet 业务在本质上呈现突发性和自相似性。自相似性就是在一个给定的物理链路上，在不考虑通信量大小的情况下，Internet 网络上的业务呈现出基本相同的特性。

图 5.1 给出了 Internet 业务的碎片性和呈泊松分布的话音业务的特性比较。

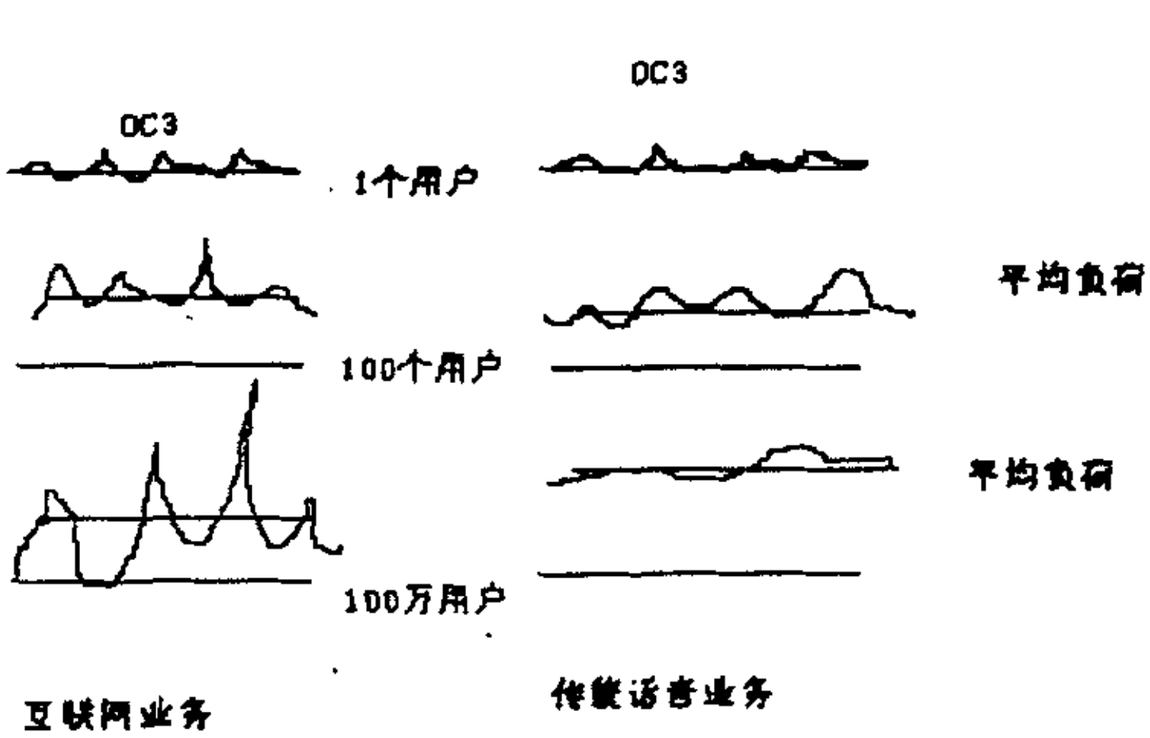

图 5.1 Internet 网络的业务特性与话音网络的业务特性对比

2. 发送和接收数据的非对称性

Internet 数据的另一特性就是在许多 Internet 链路上的发射和接收通道上带宽占用量存在非对称性。这种数据流的非对称性是由于大型服务器要

发送超量的数据给用户，与此同时需要下载大量 Web 数据的众多用户却只需发出很小的请求信息。

3. 网络中的服务瓶颈

越来越多的证据显示对 Internet 流量的限值因素不是网络本身而是服务器。Bellcore 的 Christian Huitema 最近对网络分析后给出的结论是超过 50%的 Web 拥塞原因于服务器有关。

§5.2 ns2 网络仿真软件

Network Simulator 是一个事件驱动的网络仿真器，NS 仿真软件是一种可扩展、容易配置的、可编程的事件驱动仿真引擎，支持多个流行的 TCP 和路由调度算法，其源代码全部公开，提供开发的用户接口。

图 5.2 NS 仿真器一般结构

ns 所用仿真语言是 Tool Command Language（tcl）语言的一个扩展，tcl 语言是一种简单的脚本语言，它的解释器与 c++语言相连结，tcl 具有强大功能的 X 工具包（tk），该工具包可以让用户开发具有图形用户界面的脚本，仿真通过 tcl 语言进行定义。

ns 由编译和解释两个层次组成，编译层次包括 C++类库，解释层次包括对应的 Otcl 类，用户以 Otcl 解释器作为前台使用 ns，ns 主要使用了六个类：Tcl 类、TclObject 类、TclClass 类、TclCommand 类、EmbeddedTcl 类和 InstVar 类，ns 内大部分类是 TclObject 的子类，用户在解释器环境创建新仿真对象，然后镜象到对应的编译层次对象。

利用 ns 命令编写脚本来定义网络拓扑结构、配置网络信息流量的产生和接收以及收集统计信息。软件配有仿真过程动态观察器，可以在仿真运行中动态察看仿真的运行过程，跟踪观察数据。软件还有图形显示器，显示从仿真中得到的结果数据，直观而形象。

目前 ns 可以应用于多种工作平台，Windows NT，UNIX，Sun，Linux 等。我们使用基于 Wndows 平台，需要 tcl8.3.2 、tk8.3.2 、tclcl-1.0b11 、otcl-1.0a7 和 Cygwin 支持。

发送超量的数据给用户，与此同时需要下载大量 Web 数据的众多用户却只需发出很小的请求信息。

3. 网络中的服务瓶颈

越来越多的证据显示对 Internet 流量的限值因素不是网络本身而是服务器。Bellcore 的 Christian Huitema 最近对网络分析后给出的结论是超过 50%的 Web 拥塞原因于服务器有关。

§5.2 ns2 网络仿真软件

Network Simulator 是一个事件驱动的网络仿真器，NS 仿真软件是一种可扩展、容易配置的、可编程的事件驱动仿真引擎，支持多个流行的 TCP 和路由调度算法，其源代码全部公开，提供开发的用户接口。

图 5.2 NS 仿真器一般结构

ns 所用仿真语言是 Tool Command Language（tcl）语言的一个扩展，tcl 语言是一种简单的脚本语言，它的解释器与 c++语言相连结，tcl 具有强大功能的 X 工具包（tk），该工具包可以让用户开发具有图形用户界面的脚本，仿真通过 tcl 语言进行定义。

ns 由编译和解释两个层次组成，编译层次包括 C++类库，解释层次包括对应的 Otcl 类，用户以 Otcl 解释器作为前台使用 ns，ns 主要使用了六个类：Tcl 类、TclObject 类、TclClass 类、TclCommand 类、EmbeddedTcl 类和 InstVar 类，ns 内大部分类是 TclObject 的子类，用户在解释器环境创建新仿真对象，然后镜象到对应的编译层次对象。

利用 ns 命令编写脚本来定义网络拓扑结构、配置网络信息流量的产生和接收以及收集统计信息。软件配有仿真过程动态观察器，可以在仿真运行中动态察看仿真的运行过程，跟踪观察数据。软件还有图形显示器，显示从仿真中得到的结果数据，直观而形象。

目前 ns 可以应用于多种工作平台，Windows NT，UNIX，Sun，Linux 等。我们使用基于 Wndows 平台，需要 tcl8.3.2 、tk8.3.2 、tclcl-1.0b11 、otcl-1.0a7 和 Cygwin 支持。

§5.3 网络仿真结果

我们应用仿真的机器为 PIII500/64M，应用操作系统为 windows2000。所用仿真软件版本 ns-2.1b8。

在 ns-2.1b8 中没有 RSVP 的功能，为使 ns 实现 RSVP 功能需要自己扩展 ns 功能，设计实现 RSVP 信令实现方法和把 RSVP 服务类型映射到 Diffserv 网络上。然后把扩展的 RSVP 信令功能嵌入 ns 中。

新功能嵌入过程

1)、在子目录 ns-2.1b8/tcl/lib 中加入

ns-rsvp.tcl

2)、在文件 ns-2.1b8/tcl/lib/ns-lib.tcl 中加入语句

```
source    ns-rsvp.tcl                                    <=加入
```

3)、在文件 ns-2.1b8/tcl/ns-packet.tcl

```
foreach prot {
rsvp                                                     <=加入
```

4)、在文件 ns-2.1b8/packet.h 中加入

```
enum packet_t {
// Pushback Messages
      PT_PUSHBACK,
      PT_RSVP,                                           <=加入
      PT_RSVP_PATH,                                      <=加入
      PT_RSVP_RESV,                                      <=加入
      PT_PATH_TEAR,                                      <=加入
      PT_RESV_TEAR,                                      <=加入
      PT_RESV_ERR,                                       <=加入
      PT_RESV_CONF,                                      <=加入
class p_info {
public:
      p_info() {
//pushback
name_[PT_RSVP] = "RSVP";                                 <=加入
name_[PT_RSVP_PATH] = "Path";                            <=加入
name_[PT_RSVP_RESV] = "Resv";                            <=加入
name_[PT_PATH_TEAR] = "PathTear";                        <=加入
```

```
name_[PT_RESV_TEAR] = "ResvTear";          <=加入
name_[PT_RESV_ERR] = "ResvErr";            <=加入
name_[PT_RESV_CONF] = "ResvConf";          <=加入
```

5)、在子目录 ns-2.1b8 中加入

rsvp.h

rsvp-link.h

rsvp-messages.h

rsvp-objects.h

wfq.h

rsvp.cc

rsvp-link.cc

rsvp-messages.cc

rsvp-objects.cc

wfq.cc

6)、在文件 ns-2.1b8/Makefile.vc 中加入

/rsvp.o / rsvp-link.o / rsvp-messages.o / rsvp-objects.o /wfq.o

和 tcl/ ns-rsvp.tcl

7)、nmake Makefile.vc

用 ns 命令编写仿真程序，并使用跟踪(trace)功能，收集仿真输出数据计录单个包在链路的到达、离开、丢弃，使用 trace-all 将每个包的情况记录在 out.tr 文件中。使用监视(monitor)功能，跟踪一个队列或其他对象中包的到达、发出、丢弃的统计数据及其平均值，将数值记录在 outfm.tr 中。

实现在 MPLS 和 Diffserv 环境下仿真，对业务各发送和丢弃的包的数量进行统计。在 RSVP 与 Diffserv 结合的新环境下，重新对相同网络拓扑结构进行仿真，对得到的 out.tr 和 outfm.tr 中的数据统计各发送和接收丢包率。对这两种情况的丢包率进行比较。

仿真中 r0、r1、r2 业务模型采用 CBR 业务模型，传输协议使用 RTP Agent，调度算法采用 RED 调度算法，策略模式采用二速三色的标示(trTCM Policer)。

拓扑结构如下图所示：

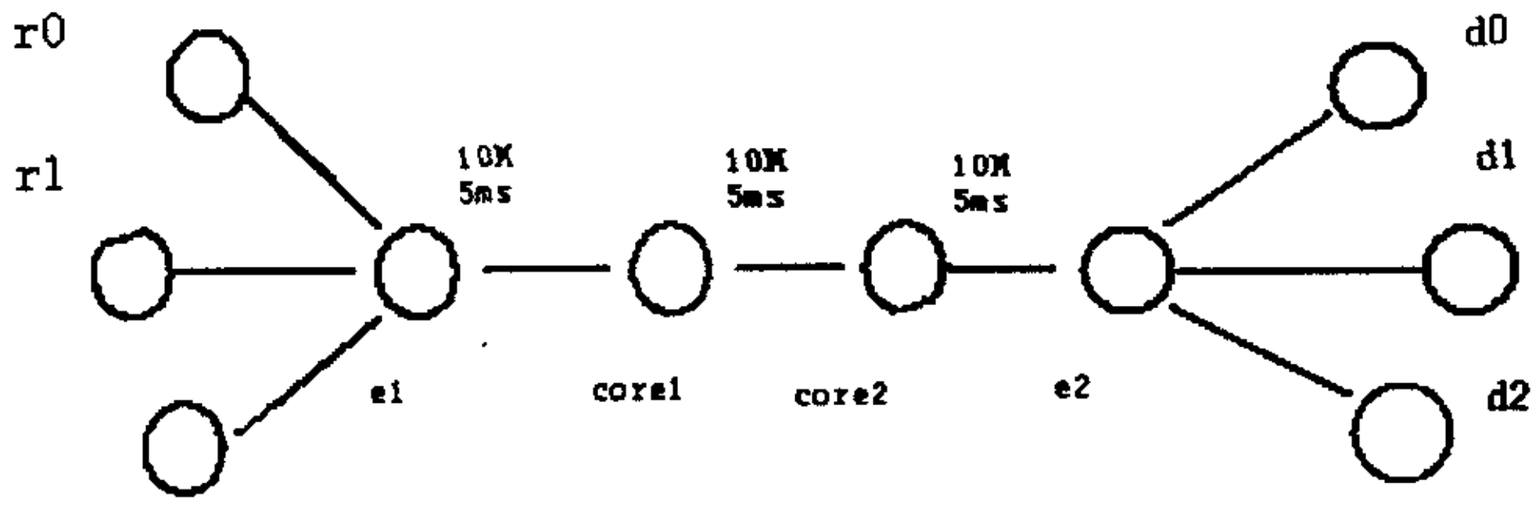

图 5.3　仿真拓扑图

仿真演示如下图所示，

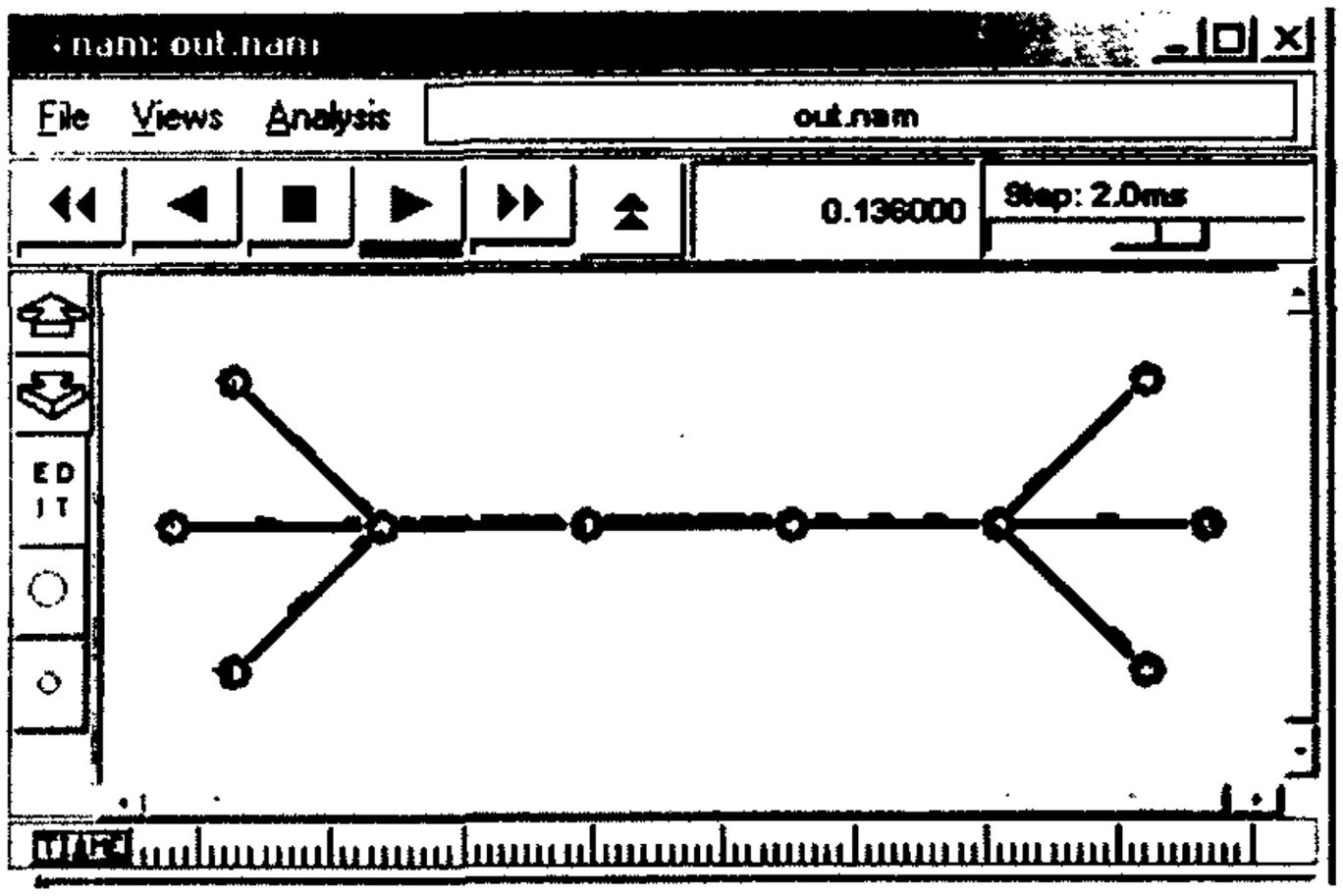

图 5.4　仿真演示图

仿真中 r0、r1、r2 的 CBR 业务模型中包大小选为 1000 bytes，当发送节点的发送速率 3000000 时，对仿真的拓扑结构的边界路由器节点进行跟踪和监视，对三种业务在传输中的包丢失统计如下：

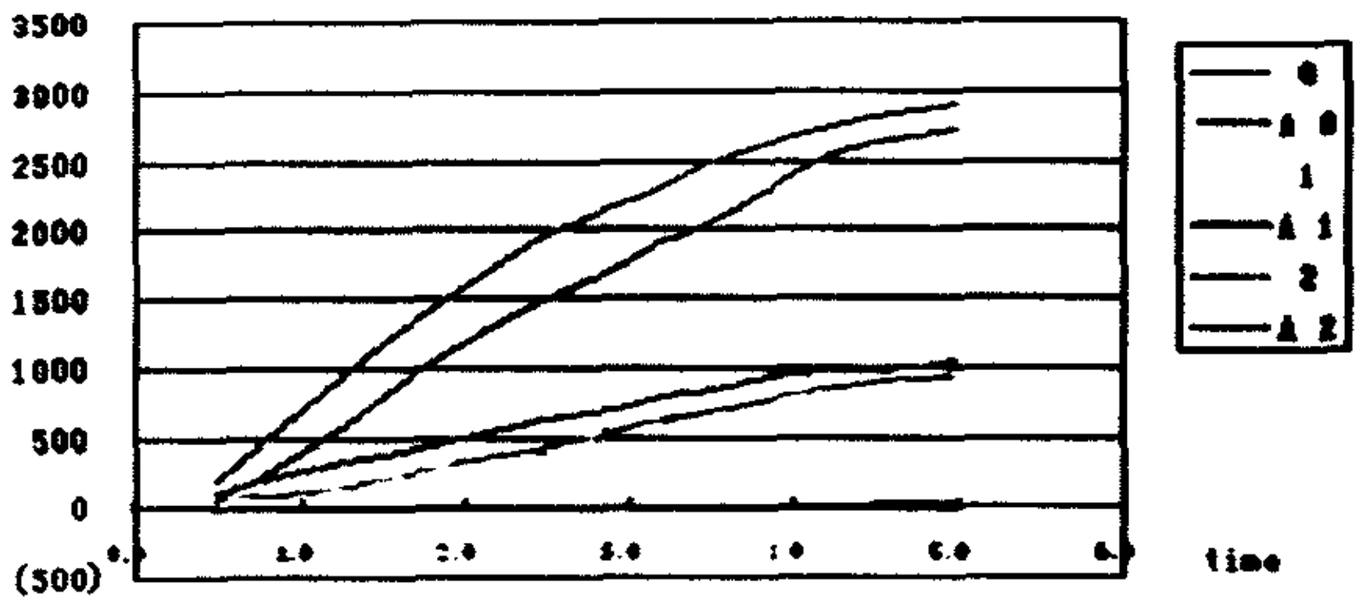

图 5.5　各业务在仿真时间内总丢包数

图中 0、1、2 为原环境下的丢包数，A0、A1、A2 为新环境下的丢包数。另对丢包率统计如下：

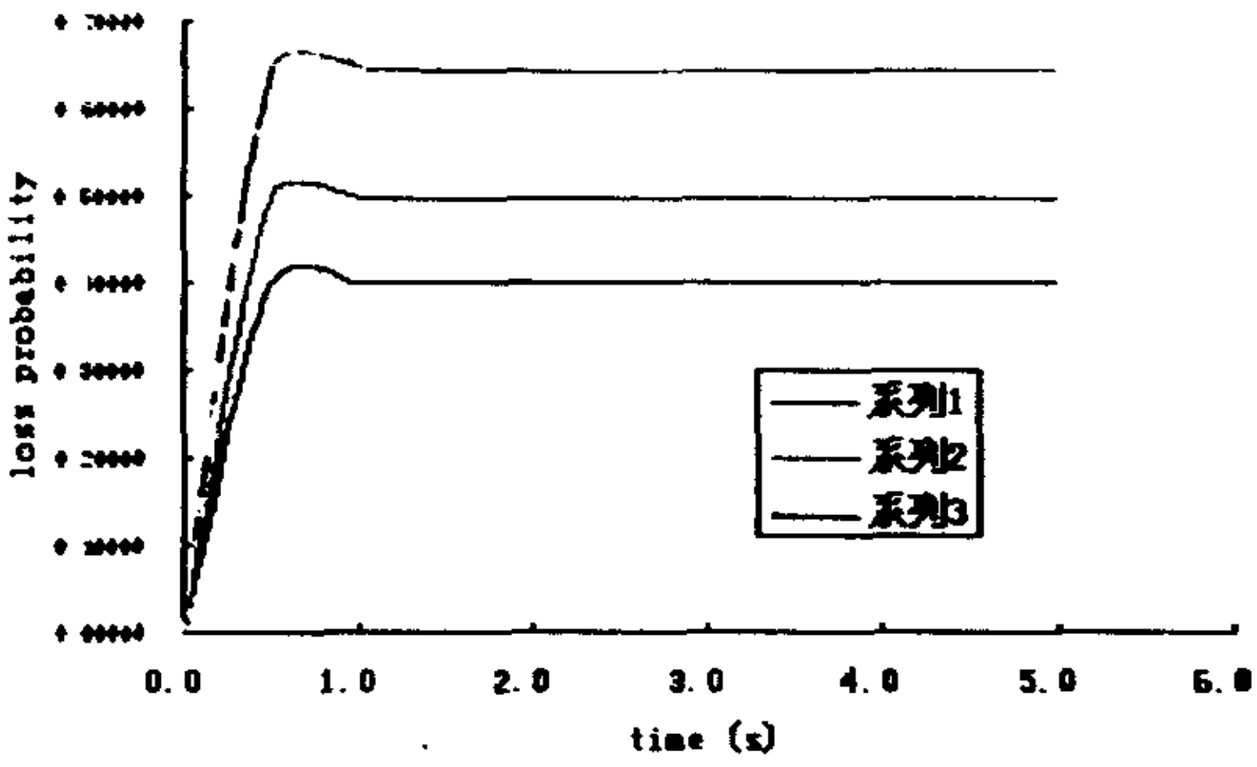

图 5.6　原环境下丢包率统计

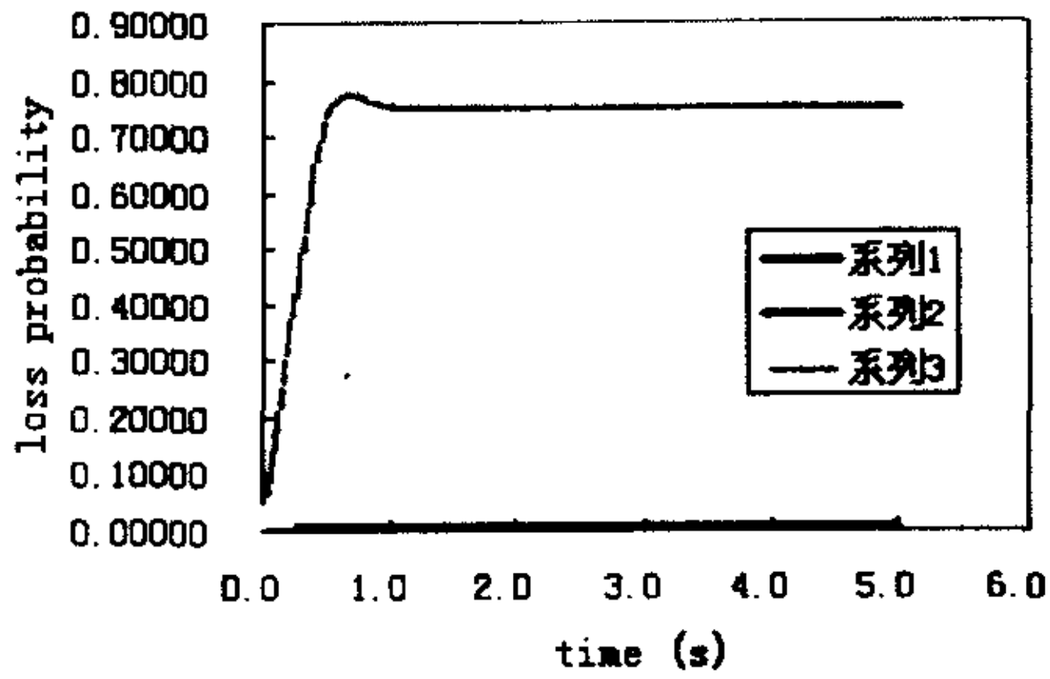

图 5.7 新环境下丢包率统计

当发送节点的发送速率 2000000 时，两种环境下的丢包率统计如下：

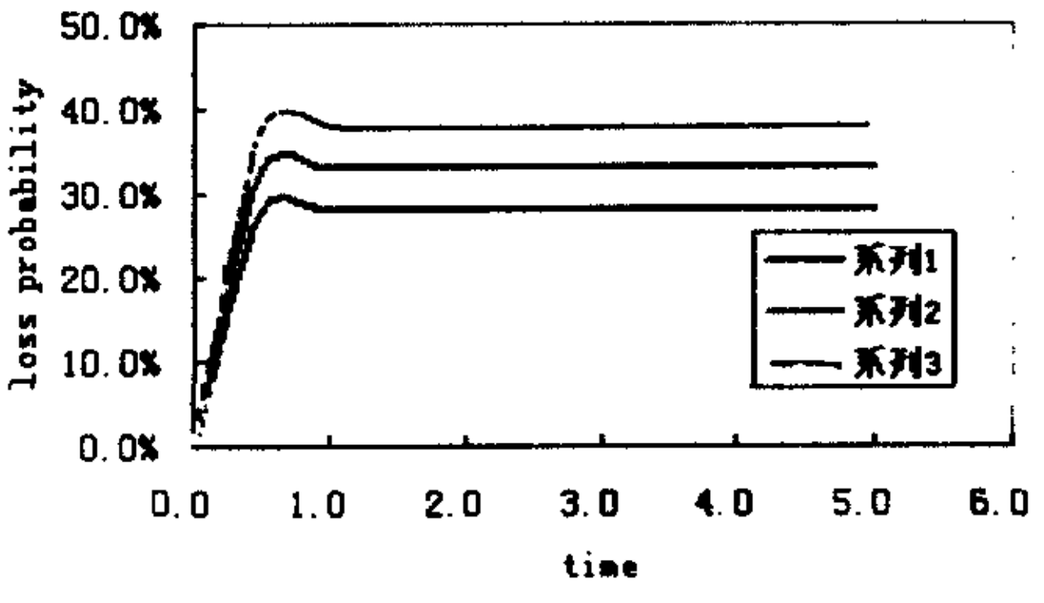

图 5.8 原环境下丢包率统计

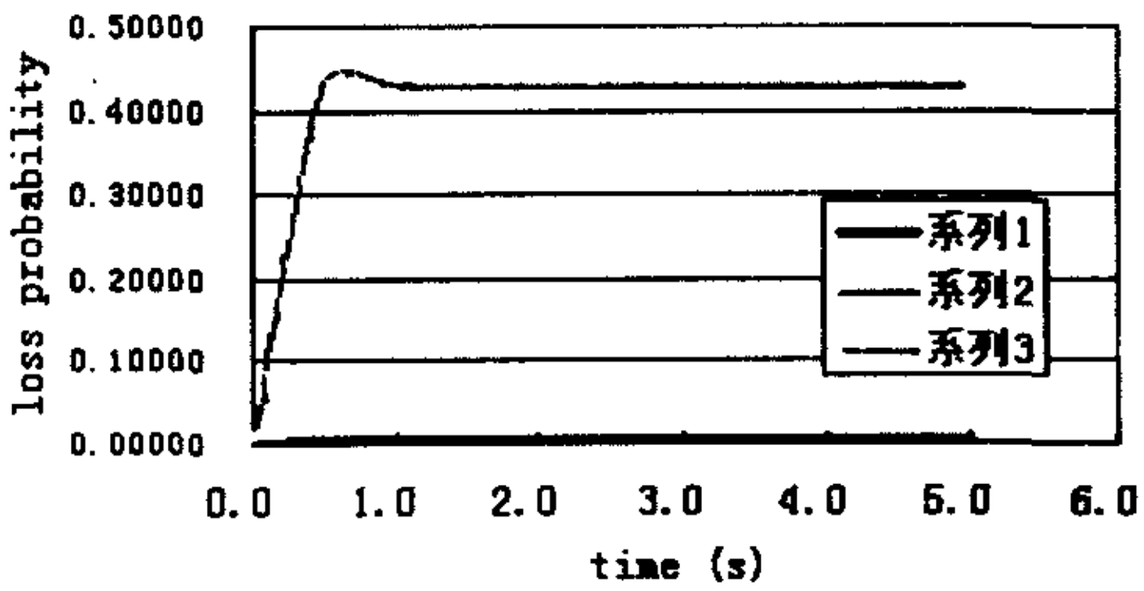

图 5.9 新环境下丢包率统计

对图 5.5 分析及图 5.6 于图 5.7 的比较、图 5.8 于图 5.9 的比较可知，在原环境中各业务的丢包率比较平均，环境使用 RSVP+Diffsrv 后，业务 3

的丢包率很大，业务 1、2 的丢包率很少，这就表明在使用新方法后重要业务可以优先的有效传送。这样就可以实现牺牲一部分数据流业务来换取另一部分数据留业务的有效传输，从而达到预定要求。

第六章 总结与展望

MPLS 技术是当前通信界研究的热点问题，是对传统 IP 网络传输技术（如 IP over ATM，IP over SDH）的改进。它采用集成模型，将第三层 IP 技术与第二层的硬件交换技术结合在一起，并且使用一个定长的标签作为分组在 MPLS 网络传输时所需处理的唯一标志。这种技术兼具了 IP 的灵活性、可扩展性与 ATM 等硬件交换技术的高速性能、QoS 性能、流量控制性能。使用这一技术，将不仅能解决当前网络中存在的大量问题（如 N 平方问题、带宽瓶颈、QoS 保证、组播以及 VPN 支持等问题），而且能够实现许多崭新的功能（如流量工程、显示路由等），是一种理想的 IP 骨干网络技术。

Internet 的发展对宽带化、多媒体化提出了越来越高的要求，IP 网络的建设迫切需要一种更为高效的技术。可以预见，IP 网络发展的转折点将是 IP 网络对于服务质量问题的解决以及对各种新兴的增值业务的支持。MPLS 一方面是目前唯一能够保证 IP 网络服务质量的网络技术，另一方面，使用 MPLS 将可以十分高效地实现各种增值业务，如 VPN 等。MPLS 技术将成为下一代 IP 网络的基础技术。

本论文对 MPLS 协议及 QoS 的两种实现模型 Intserv 协议、Diffserv 协议进行了深入的研究，对在 MPLS 网络中实现 Intserv 和 Difserv 的方式进行了深入的分析。在以上研究和分析的基础上，给出了在 MPLS 网络中实现端到端的 QoS 实现框架。提出了 RSVP 协议在 Difffserv 域中实现显式接纳控制和有效、动态的资源调度机制。提出了 Intesev 服务类型到 Diffserv 网络提供的服务之间的映射关系。定义了 Diffserv 域内使用聚集传输控制的网络元素在支持 RSVP 信令时所需的功能。提出了在 MPLS 环境下，使用聚集 RSVP 将 Diffserv 区内的资源可用性信息传递到边界路由器的方法。

并对当前的主流开放式网络仿真平台 ns 进行了细致的探索，在 ns 平台上，设计、引入和实现了 RSVP 信令，以及把 RSVP 服务类型映射到 Diffserv 网络的功能模块，扩充了 ns 功能。仿真结果验证了所提方法的可行性和有效性。

对于未来的研究工作，我们认为应从以下几个方面继续进行深入的探讨：

(1) MPLS 技术应用于 VPN 问题。

实现 VPN 的关键技术有四项，分别是隧道技术、解加密技术、密匙管理技术、使用者与设备认证技术。

MPLS 通过标记的使用来实现报文的正确转发，可以确保具有与外部 IP 报文相同 IP 地址的 VPN 内部 IP 报文不会被发到公网之上。MPLS 用一套特定的标记将任何一组抽象的网络层实体联系起来，再借助于这些标记对数据流进行转发。这样，便可以用一套标记以一种安全而且可预测的方式将 VPN 内的成员及其网络地址前缀联系起来。MPLS VPN 可以直接利用 MPLS 的流量工程和 QoS 能力。这样，对于具有不同的 QoS 要求的业务可以使用不同的技术组合来提供实现。

MPLS VPN 技术目前还只是一个笼统的框架，对于 MPLS 的这一应用还有待于进一步的研究。

(2) MPLS 向光网络扩展问题。为了适应对智能光网络进行动态控制和传送信令的要求，而产生了通用 MPLS（GMPLS）技术。它对传统的 MPLS 进行了扩展、更新。使用 GMPLS 可以为用户动态地提供网络资源，以及实现网络的保护和恢复功能。与传统的 MPLS 相比，GMPLS 在以下几个方面得到了改进。

向光网络进行了扩展的 GMPLS 不同于传统的 MPLS，主要在于它支持多种类型的交换单元，即 GMPLS 除了支持分组交换，还支持(1)时分多路复用 TDM（包括 SONET\SDH,ADMs）、波长复用（optical lambdas）和光纤交换（incoming port or fiber to outgoing port or fiber ）等

GMPLS 对传统 MPLS 协议的扩展和更新，为了支持这种新型的光交叉连接，GMPLS 不仅拓展了传统的 MPLS 的信令和路由协议，而且还增加了新的功能。这些变化影响了标签请求、标签分配、带宽分配的方式，以及 LSR 的双向特性和当网络发生故障时的通信机制等。

基于 GMPLS 的波长标签网络解决方案，最大的好处就是它在充分发挥已有光联网技术的基础上，也具有适应未来光联网网络技术发展的潜能，这种多方位的适应性包括了电路交换、分组交换、以及各种混合交换。

(3) DS 区域内的组播问题。传统的 IP 组播模型主要包括两部分：主机组模型和组播路由协议。在主机组模型中、 一组主机仅由一个组播组地址决定. 通过组地址来进行服务的订阅而后调度转发。组播路由算法用来维持

状态的数量和维护组播树。

在 Intserv/RSVP 体系结构中。由于 RSVP 可以为组播流提供资源预留的支持(单向预留、由接收者产生预留请求、预留是软状态的)，因而组播流的传输不会遇到问题。但在 DS 区域中、Diffserv 的简单性使其可能遇到异质组播的问题，既同一组内不同接收节点可能希望获得不问的 QoS。而恰好结合点在 DS 区域内部，它没有识别 RSVP 能力．也没有 MF 分类及存储基于流状态的功能。这样在复制分组流到不同分支过程中，就无法对它们按各接收收 QoS 要求重新标记不同的 DSCP 值。而是将它们简单地标记成未分支前的 DSCP 值(对应接收者要求最高的 QoS)，显然会浪费网络资源．同时要求低的用户也必须付高额费用。

为解决这一问题，有人提出让 DS 区域内节点增加功能模块使其和边界节点一样能 MF 分类重标记、存储流状态，这样其实就根本改变了 Diffserv 的最大的优点一可扩展性，不可取。目前如何解决 DS 区域异质组播问题仍是一个开放的课题．

参考文献

[1] 石晶林，丁炜 MPLS 宽带网络互连技术，人民邮电出版社，2001

[2] 吴江 赵慧玲 下一代的 IP 骨干网络技术 ，2001

[3] 徐荣 龚倩 高速宽带光互联网技术 ，2002

[4] " Multiprotocol Label Switching Architecture", IETF rfc3031

[5] "A Framework for Multiprotocol Label Switching" IETF Internet Draft draft-ietf-mpls-framework-05.txt

[6] "LDP Specification" IETF rfc3036

[7] "MPLS Support of Differentiated Services" IETF Internet Draft draft-ietf-mpls-diff-ext-09.txt

[8] "Generalized MPLS Signaling - CR-LDP Extensions" IETF Internet Draft draft-ietf-mpls-generalized-cr-ldp-04.txt

[9] "Generalized MPLS - Signaling Functional Description" IETF Internet Draft draft-ietf-mpls-generalized-signaling-06.txt

[10] D. Awduche, J. Malcolm, J. Agogbua, M. O'Dell and J. McManus, "Requirements for Traffic Engineering over MPLS", RFC 2702, Sept. 1999.

[11] "Extensions to RSVP for LSP Tunnels" IETF Internet Draft draft-ietf-mpls-rsvp-lsp-tunnel-07.txt

[12] "Multiprotocol Label Switching (MPLS) Traffic Engineering Management Information Base" IETF Internet Draft *draft-ietf-mpls-te-mib-07.txt*

[13] A.Banerjee et al.,"Generalized Multiprotocol Lable Switching : An Overview of Routing and Management Enhancement,"*IEEE Commun.Mag.,Jan.2001*

[14] Panos Trimintzios et al.. " A Management and Control Architecture for Providing IP Differentiated Services in MPLS-Based Networks" *IEEE Commun. Mag.,May.2001*

[15] D.Awduch and Y.Rakhter,"Multiprotocal Lambda Switching : Combining MPLS Traffic Engineering Control with Optical Cross-connects," *IEEE Commun.Mag.,Mar.2001*

[16] 万红生 易准 丁铁骑 "多协议标记交换的原理和应用" 2001.3

[17] 徐荣 龚倩 通用多协议标签交换技术， 通讯世界 ，2001（6）

[18] "An Architecture for Differentiated Services" IETF RFC2475

[19] "Resource Reservation Protocol(RSVP)—Version1 functional specification" IETF RFC2205

[20] "The Use of RSVP with IETF Integrated Services" IETF RFC2210

[21] "Specification of the Controlled-Load Network Element Services" IETF RFC211

[22] " Format of the RSVP DCLASS Objecg "IETF RFC2996

[23] "Aggregation of RSVP for IPv4 and IPv6 Reservations" IETF RFC3175

[24] "A Framework for Integrated Services Operation over Diffserv Networks " IETF RFC2998

[25] "Assured Forwarding PHB Group" IETF RFC2597

[26] "An Expedited Forwarding PHB" IETF RFC2598

[27] K. Nichols, S. Blake, F. Baker and D. Black, "Definition of the Differentiated services Field (DS Field) in the IPv4 and IPv6 Headers", RFC 2474, Dec. 1998.

[28] "Integrated Service Mapping for Differentiated Services Networks" IETF draft-ietf-issll-ds-map-01.txt

[29] Osama Aboul-Magd et al;"QoS and Service Interworking Using Constraint-Route Lable Distribution Protocol (CR-LDP)" *IEEE Commun.Mag.,May.2001*

[30] *Katsuyoshi et al* ; "Performance Evaluation of the Architecture for End-to-End Quality-of-Service Provisioning" *IEEE Commun.Mag.,April.2000*

[31] *Fugui Wang et al* ; "A Random Early Demotion and promotion Marker for Assured Services" *IEEE Joural of selected areas in communications.VOL.18.No12,Dec.2000*

[32] *Panos Trimintzios et al;* "A Management and Control Architecture for Providing IP Differentiated Services in MPLS-Based Networks" *IEEE Commun.Mag.,May.2001*

[33] *Atsushi Iwata et al* : "A Hierarchial Multilayer QoS Routing System with Dynamic SLA Management" *IEEE Joural of selected areas in communications.VOL.18.No12,Dec.2000*

[34] Yoram Bernet "The Complementary Roles of RSVP and Differentiated

Services in the Full-Service QoS Network" *IEEE Commun.Mag.,Feb.2000*

[35] X.Xiao and L.M.Ni, " Internet QoS: The big picture" *IEEE NetworkMag.,Mar/Apr.2001*

[36] "A Two-bit Differentiatd Services Architecture for the Internrt " IETF RFC2636

[37] Brian Williams, Ericsson Australia " Quality of Service Differentiated Services and Multiprotocol Label Switching "

[38] Nortel white paper "IP QoS---A Bold New Network : An IP Quality of Service backgrounder for service providers "

[39] Raquel Hill and HT Kung "A Diff-Serv Enhanced Admission Control Scheme" Proceedings of *IEEE Globecom 2001*,November 2001

[40] P. Trimintzios, I. Andrikopoulos, G. Pavlou, P. Flegkas, University of Surrey, " A Management and Control Architecture for Providing IP Differentiated Services in MPLS-based Networks"

[41] Atsushi Iwata and Norihito Fujita "A Hierarchical Multilayer QoS Routing System with Dynamic SLA Management"IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATION, VOL. 18, NO. 12, DECEMBER 2000 2603

[42] *Sang-Sik Yoon, *Deokjai Choi, **Gueesang Lee "Traffic Engineering System(TES) Design for Qbone over MPLS Network"

[43] Guy Almens, Philip F. Chimento, et al. "Qbone Architecture", Internet2 QoS Working Group draft, Aug, 1999.

[44] P.F. Chimento,et al. "BANDWIDTH BROKER", Interne2 QoS Working draft, May 3, 2000

[45] Ben Teitelbaum and Ted Hanss, "QoS Requirement for Internet2", Internet2 QoS Working Group, April 22, 1998

[46] http://www.internet2.edu/qos/wg/

[47] Kevin Fall, Kannan Varadhan "The *ns* Manual" The VINT Project

[48] http://www.topology.org/soft/sim.html

致　谢

论文的完成首先感谢我的导师赵珑高级工程师。在硕士研究生期间，赵老师在学习和工作中给予孜孜不倦的教诲和耐心的培养，使我在学术上、思想上有了长足的进步，顺利地完成了三年的研究生学习。赵老师渊博的学识、严谨的治学态度使我受益匪浅，老师学术上高瞻远瞩，为我今后的发展指明了道路。山东大学三年的研究生经历使我受益终生。

同时，研究生期间还得到张有志教授精心的指导以及杜岩、张晓敏老师的关怀与指导，和夏斌同学三年来同我在学术上的共同讨论，谨在此向所有帮助过我的老师和同学表示诚挚的感谢。

感谢母校对我的教育和培养，我将永远记住在山东大学度过的这这段美好时光。

最后，感谢我的家人对我学业上莫大的支持和鼓励。