维普资讯 http://www.cqvip.com

各种服务质量体系的机理及性能评价*)

Mechanism and Performance Evaluation of Various QoS Architectures

王雅丽 林 生

(华南师范大学计算机科学系 广州510631)

Abstract As the Internet is developing increasingly and used widely, more and more attentions are paid to the "Quality of Service" (QoS) of networks, which now becomes a research hotspot. Based on the summary of various QoS architectures currently existed, descriptions of their implementing mechanism are presented in detail. Furthermore, careful evaluations and comparisons on their performance are offered. Finally, from the aspect of their integrated utilities, a concept of "QoS Chain" is proposed and described.

Keywords QoS, DiffServ, QoS Cham, Internet, Internetworking

进入90年代后,互联网以超摩尔规律的速度增长,Internet 上的应用也逐步由单一的数据传送向数据、语音,图像等多媒体信息的综合传输演化。由于 Internet 采用 IP 协议只能提供简单的点到点"尽力而为"(best-effort)服务,无法提供服务质量(QoS)保证,为适应新的业务,提供对多种 QoS 业务的支持,必须改进当前的 Internet 网络,可以说,Internet 的迅猛发展使得如何在 Internet 中支持具有不同 QoS 要求的应用成为网络通信领域的研究热点。Internet 服务质量体系就是要提供质量稳定的、可预测的数据传输服务以满足客户应用请求。该体系使各网络要素(如发送数据的应用进程、主机或路由器)具有保障某项业务以一定的服务质量进行端到端传输的能力;使自顶向下的所有 OSI 层,从发送端到接收端的所有网络要素相互协作,共同实现某项服务质量请求。

Internet 服务质量体系包括服务质量机制 QoS 和服务类别机制 CoS 两大类。CoS 机制又包括数据链路层标记(如 IEEE802.1p)和网络层的业务优先级 ToS (Type of Service)两种类型,近年来最新提出的"区分服务"DiffServ 则是服务质量控制精度介于 QoS 和 CoS 之间的一种服务质量机制。所有这些服务质量机制并不创造带宽,而只是根据所传输的业务和网络环境有效地管理带宽,以各种策略最大限度地使用带宽,以提供优于目前尽力而为的服务质量保证。

1 各种服务质量机制及其实现机理

1.1 QoS 及 IntServ/RSVP 机制

QoS 机制为应用业务提供端到端的服务质量保证,控制精度高,可满足 Internet 多媒体应用等实时、大数据量的传输需求,是理想状态下最佳的服务质量保障机制。目前,对 QoS over IP 的研究借鉴了 QoS over ATM 的一系列解决方案,定义了 IntServ 服务质量保证模型,与资源预留协议 RSVP[1]共同实现 IP 网络的 QoS。

IntServ[2]提供基于流的、端到端的质量保证型服务。它在流传输之前在源节点和目的节点之间建立某种连接,以获得有效的资源预留。它依靠 RSVP 信令协议逐跳程(hop-by-hop)地建立或拆除每个单一流的资源预留"软状态";依靠接纳模块来判断是否接纳新业务,以减少网络负荷;依靠分类和分组控制模块将 IP 分组分类成传输流,并根据传输流的服务等级实施 QoS 路由,传输调度等控制。通过在端系统和网络中各节点中各个控制模块的协作,IntServ 能实现可与 ATM 网络相媲美的服务质量。

1.2 CoS(Class of Serrice)机制

QoS 机制提供了高精度的质量保证,而 CoS 机制则是将多个服务质量要求相同或相近的业务流合并成一种业务等级,得到较少的业务类别,从而降低了控制的复杂性。网络中的节点只需根据业务流的等级分配相应的网络资源,并做出相应的排队策略,路由转发等处理。

1.2.1 IEEE802.1p 的实现机理

IEEE802.1p 是一种子网带宽管理 SBM(Subnet Bandwidth Management)机制,它是在第2层上实现的服务类别机制。

*)本研究获得"综合业务网理论及关键技术"国家重点实验室(西安电子科技大学)开放研究基金资助。

维普资讯 http://www.cqvip.com

802.1p 应用于符合 IEEE802标准的交换共享型网络，在OSI 模型中的数据链路层上实现业务分类和优先级机制，其具体实现由 IEEE802.1p、802.1Q 和 802.1D 定义。IEEE802.1p 是基于优先级的服务质量保证机制，它使用了802.1Q 帧头中的三个比特来表示八个优先级别(表1)。

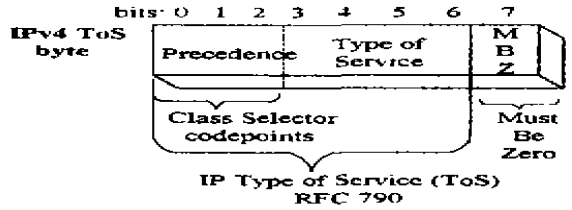

图1 ToS 域的位图

表1 CoS 业务类别与比特组合间的对应

三比特组合	优先级	对应的业务
000	0	尽力而为型(缺省)
001	1	准尽力而为型
010	2-3	保留
011	4	延迟敏感型，无时间限制
100	5	延迟敏感型，100ms 限制
101	6	延迟敏感型，10ms 限制
111	7	网络控制型

1.2.2 ToS 的实现机理 如图1所示，ToS 字段的0-2位为 IP 包优先级(Precedence)子域。三个比特位表示业务流的八种优先级，如优先级最高的比特组合为"111"，对应于网络控制业务流；优先级最低的比特组合为"000"，对应于普通业务流。ToS 字段的3-6位为服务类型子域，用大写的 TOS 表示。在业务流传输时，TOS 子域为路由决策服务，其各个比特位对应于不同的路由选择方式，其语义定义如表2所列。

表2 ToS 值与路由选择方式间的对应

D	T	R	C	路由选择方式
1	0	0	0	最小时延路由
0	1	0	0	最大吞吐量路由
0	0	1	0	最可靠路由
0	0	0	1	最小开销路由
0	0	0	0	普通路由(缺省值)

D:Delay，ToS 域的第4位 T：Throughout，ToS 域的第5位
R:Reliability，ToS 域的第6位 C:Cost，ToS 域的第7位

Precedence 子域并不影响路由的选择，它只标明业务的优先级别，并在子网内有效。当分组在同一通道排队时，具有高优先级的分组将先被转发，其转发的路径则是由 ToS 子域决定的。ToS 子域由源站设置，为网络如何在吞吐量、延迟、可靠性和开销中获得平衡提供解决方案。主机或网络节点根据 D、T、R、C 四位的组合选择相应的路由与转发方式，比如，如果请求了最小时延(D=1)的分组，就不应该让它经过包括卫星链路的子网。

此外，IETF 在 IPv6的制定中定义了8比特位的业务类别域和20比特位的流标记域[4]来实现基于优先级的业务分类和路由选择。其中，业务类别域不影响路由选择，源站利用它标识每一个分组与优先级有关的特性 从施行服务质量控制的对象出发，IPv6提出了流的概念，所谓"流"，是指从一个特定的源站向一个特定的目的站发送的一系列分组，这些分组共享如路径、资源分配、丢包率、开销等属性。在 IPv6中，一个流由一个源地址和一个非零的流标记的组合唯一标识。流标记由源站或路由器分配给分组，它与保存在路由器中的流的服务质量需求信息相对应，这样，路由器只要查看流标记所对应的需求，就可以决定给分组选路和处理该分组了。

1.3 DiffServ 机制

DiffServ[5]机制是一种基于业务分类及相关质量保证策略的机制。它不使用单独的资源预留信令机制和相应的显式资源分配过程，也不为每一种业务都提供相对应的 QoS 保证策略，而是根据分组中的"区分服务码点"(DSCP)域选择相应的"逐跳转发行为"(PHB)进行处理。它简化了网络内部节点的服务对象，采用聚集传输控制，即具有某些相同之处的单一流合并成聚集流进行传送。同时，DiffServ 简化节点控制，只有网络边界节点才需对进入网络的单一流进行分类、整形、聚合等操作，并将结果标记在 DSCP 域中，网络内部节点只需按 DSCP 域值执行 PHB 处理即可。

区分服务利用 IP 包头中的区分服务标记域(DS field)在 DS 区域的边界节点和内部节点间传递控制信息。DS 标记域占用了 IPv4包头的 ToS 字节或 IPv6包头的业务类别字节的前六位[6]，如图2所示。图中，DSCP 用来标识聚集流，每个分组根据包头中的 DSCP 值获取相应的网络资源和控制策略，并以此选择相应的 PHB。DSCP 中的前三位(0-2)用来标记业务类别，其作用与 ToS 中的 Precedence 子域相似，但 DiffServ 重新组织了优先级与 PHB 之间的映射。如优先级0对应于缺省型 BE(Best Effort)PHB；优先级1-4对应的均为确保型(Assured Forwarding)PHB；优先级5对应于加速型 EF(Expedited Forwarding)PHB。DSCP 中的第4和第5位标记丢包率，第6位总是为0。网络节点转发业务流时，先根据前3位判断业务类别，对具有相同类别的流再按丢包率进行区分。由此可见，DSCP 域的定义并不影响路由的选择，它所提供的服务仅限于流的分类与聚集、队列调度和缓冲管理。

维普资讯 http://www.cqvip.com

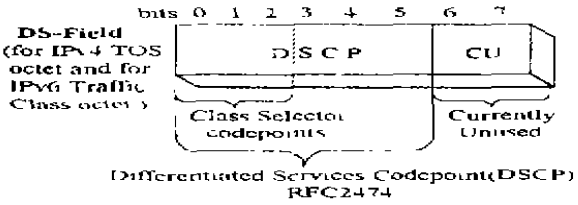

图2 DS field 的位图

2 各种机制的性能分析与比较

从上述对各种服务质量机制的描述，可以看出，服务质量体系就是要为不同的网络数据传送业务提供相应的服务质量保证。从实现的角度，可将现有的服务质量机制分为两个类型：

·资源预留型（QoS、IntServ/RSVP）：根据业务的服务质量要求，通过复杂的信令机制，在流传输前预约带宽等网络资源；

·服务类与优先级型（CoS、DiffServ）：将网络业务分类并标以一定的优先级，或选择相应的服务类型，根据带宽管理策略在流传送过程中分配网络资源，各网络要素根据业务类型按优先级分别对待。

从施行服务质量保证对象的角度，又可按是"单一流"（Per Flow）还是聚集流（Per Aggregate）来分类

·单一流（QoS、CoS）：每个流都是收发双方之间独立的、具有唯一目的地的业务流，由地址五元组（传输协议号、源地址、源端口号、目的地址、目的端口号）唯一标识；

·聚集流（DiffServ）：聚集流是由两个或以上的单一流组成。这些单一流必须具有某些相同的特征，如共同的目的地址，相同的优先级或标记等。

IntServ 依靠 RSVP 信令机制实现资源预留。IntServ/RSVP 需要基于单一流的、复杂的资源预留、QoS 路由和调度机制。毫无疑问，复杂的控制与高昂的开销带来的是高精度的服务质量保证，但由于这种保证具有某种面向连接的特性，在无连接的 IP 网络上实现 IntServ/RSVP 是不现实的。IntServ/RSVP 存在有两大缺陷，限制其在复杂的、大规模的网络中的应用。一是缺乏可扩展性，因为基于流的控制复杂性会随着流的增多而增加。为了保证端到端的服务质量，端系统和节点设备都要维护每个单一流的状态信息及其连接，如果业务过于繁忙，将会导致网络核心设备的崩溃。二是健壮性较差，在分布式网络环境中很难维持动态的传输流状态的一致性。尤其是如 Internet 这样大规模的网络，其链路状态是不确定的，难以有效地进行资源预留。如果网络中某个节点不支持 RSVP 控制协议，但其预约的 QoS 将得不到保证。所以，IntServ/RSVP 一般应用在组织性网络（如：接入网、企业网、校园网等）中，以获得局部的高服务质量保证。

IEEE802.1p 定义的八个优先级别中与延迟限制有关的有三个，而对要求低抖动、低丢失率的业务流并没有相应的优先级相对应。可见，802.1p 主要解决时延敏感（time-critical）的业务流传输，对于其它类型的业务流的支持并不完善。由于802.1p 是定义在第2层上的服务质量保证机制，所以它只应用于802协议族的局域网中。802.1p 只使用了帧头中的三个比特位，使其 QoS 控制过于粗糙，与之相类似，虽然 ToS 机制具有八种业务优先级别，十六种路由选择方案，可是对于多种多样的网络业务和多变的网络环境，依然显得简单和粗糙，所以它只是 IP 协议中提供的能力有限的服务质量保证机制。由此可见，CoS 机制是基于优先级的，它使用标头中较少位来标识业务优先级，所以控制机制简单、不需要额外的资源消耗、具有良好的可扩展性；但简单的控制机制，使得它提供的服务质量保障较差，通常与其它服务质量机制协作来完成服务质量粒度较低的业务流传输。

表3 各服务质量机制间的比较

		策略	流类型	作用范围	具体实现	可扩展性
QoS IntServ		基于流	单一流	端到端	RSVP 等信令机制进行资源预留，在主机和转接节点都需要控制模块，实现复杂	差，需改造现有 IP 网络及相关软/硬件设备
CoS	SBM（802.1p）	基于优先级	单一流	子网内部	利用帧头中的三个比特来表示八个优先级别，由第二层交换机实现，需要与上层协议优先级的映射	升级费用昂贵，与以前的网络不兼容，仅对第二层有效
	ToS	基于优先级	单一流	子网内部	由 IP 包头的 ToS 域提供八种业务优先级别，十六种路由选择方案	为某些网络控制协议、路由协议提供支持，一般不单独使用
DiffServ		基于策略和规则	流聚集	DS 区	IP 包头的6位 DSCP 域提供64种类型服务与 PHB 相对应	向后兼容 IPv4，无需大的软硬件改动，可通过协议升级实现

与 ToS 的7种业务优先级和16种路由选择方式相比，DiffServ 的64种 DSCP 值的分配和使用规范及其

维普资讯 http://www.cqvip.com

对应的PHB更能满足网络业务的多样性，更适应网络的管理和扩展需求。DiffServ服务质量控制的对象是粒度比单一流"粗"的聚集流，这种"粗"粒度的服务，使DiffServ具有良好的可扩展性，同时，也使得DiffServ将实现的复杂度大部分移到了子网的边缘，而在子网内部的节点只需进行简单的调度转发即可，通过这些既完善又简单的网络管理策略，DiffServ为业务流提供服务质量近似于QoS的高质量保证，表3比较了各服务质量机制之间的差别

从对各种服务质量机制的比较可以看出，QoS实现困难且成本昂贵，CoS只能提供针对某些方面的粗略的服务质量保证，二者的可扩展性都较差，而DiffServ由于向后兼容IPv4，提供的服务保证接近QoS但控制简单成本低廉，彻底贯彻QoS机制设计的基本原则"子网边缘复杂，内部简单"，因而具有进一步研究实践及应用推广的价值。

3 各种机制的联系与应用

值得注意的是，多种服务质量机制不是各自孤立地在互联网上实现服务质量保证，而是共同作用，为用户提供透明的自顶向下的端到端的服务质量保障体系，我们可将这种体系称为"服务质量链"。

所谓"自顶向下"，主要是指主机各个OSI层都要支持服务质量机制，而且各层机制的优先级别之间要能实行有效的映射。从上层传下来的具有高优先级的数据传输请求在下层也能得到优先对待，对于主机所在的子网，也要使优先级高的业务流在子网的第2层传输时获得优先对待，目前，在第2层实现QoS机制成熟的技术有ATM，在IEEE802网上有802.1p。如何实现802.1p中定义的优先级别与上层服务质量机制之间的对应，仍在研究与实践当中。

至于服务质量链中的"端到端"，则侧重于如何实现不同QoS机制之间的无缝衔接，因为互联网是由各种各样的子网在TCP/IP协议的基础上互联而成的，各子网使用的服务质量机制互有不同；作为网络节点的路由器、交换机的型号功能不同，支持的服务质量机制也有多有少，所以各服务质量机制如何识别其它QoS机制标记的业务流及其策略，并将其转换成自己的格式和策略，关系着是否能从发送端到接收端实现一致的服务质量。

图3描述了整个服务质量链的结构，并以IntServ/RSVP与DiffServ为例，说明不同的服务质量机制如何协作以实现端到端的服务质量，IntServ/RSVP为传输业务流预约资源，而DiffServ只是对业务流分类并标记优先级，IntServ的复杂性及对转接节点的要求远远高于DiffServ。当业务流从支持IntServ/RSVP的区域传输到不支持它的DS区时，位于DS区边界的入口节点(ingress points)能将RSVP资源预留配置映射到由DSCP标识的服务类别，当业务流到达DS区的出口节点(egress points)时，该节点将其转换成原来的配置，再进行转发。为了更好地实现服务质量链，可在各子网内加入带宽代理BB(Bandwidth Broker)。BB的应用是与DiffServ紧密联系的，是目前QoS研究中的新概念。

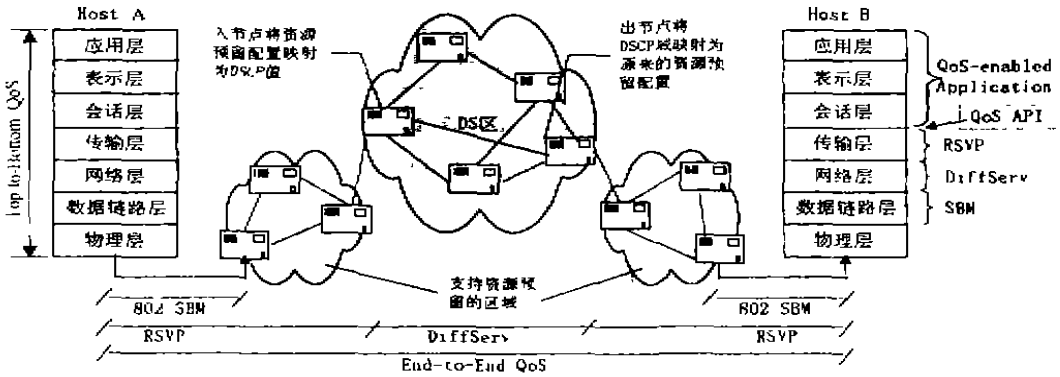

图3 涵盖"end-to-end"与"top-to-bottom"的QoS链

结束语 直到现在，TCP/IP网络提供的都是平均分配网络资源的"尽力而为"服务，把服务质量支持加入Internet体系将使其可靠传输的能力大大提高，从IntServ到DiffServ的发展，到追求多种服务质量机制的融合互补的服务质量链的提出，表明对Internet服务质量体系的研究不可能也不需要设计出一种能满足所有网络设施、所有业务需求的服务质量机制，而是要对众多的机制进行选择和折衷，并加入新的思路与构想，从而建立起一套自顶向下、端到端的服务质量保证体系。未来的Internet应是在流分类机制基础上，提供端到端自适应服务质量保证的智能型网络。

(下转第59页)

维普资讯 http://www.cqvip.com

为了更好地提高专用附网存储服务器的性能，在读写一个文件时，尽量少发读写命令，使每发一次读或写命令时，尽量读或写更多的数据块（因为每发一次读写命令，执行读或写命令时都需要一定的时间），在此基础上实现DMA传输机制，我们在Linux环境下开发了一个碎片整理程序，定期对服务器挂接的磁盘存储系统进行整理，以保证一个文件的数据块能尽量存放在相邻的扇区。

5 试验评估

5.1 服务器连续读速度测试

对我们的系统先进行试验评估代替数据分析，这里NR和ND分别代表RAID控制器数和总的硬盘数，文件读请求块尺寸用BS表示。在表1中，试验是对我们所设计的服务器和普通文件服务器进行的，CPU都是Pentium 500，内存都为128MB，试验时让它们分别挂接1-3个RAIDO，每个RAIDO所含有18硬盘，文件的分块尺寸分别从32k到64k，请求文件的大小分别为1MB到50MB，然后分别测试它们的读速度。

表1 在不同的磁盘阵列数和文件分块下的连续读速度测试

RAIDO		专用网络存储服务器 连续读速度(M字节/秒)			传统文件服务器 连续读速度(M字节/秒)		
文件分块	请求文件的块大小BS	RAID控制器数NR			RAID控制器数NR		
		1	2	3	1	2	3
32KB	1MB	31.1	45.2	58.2	20.2	39.9	50.8
	50kB	33.8	47.4	68.5	23.4	40.8	55.0
64kB	1MB	59.5	67.5	82.9	43.5	59.3	72.6
	50MB	61.8	77.9	92.5	44.1	67.2	80.2

从表中我们可以看到，在同样CPU和内存的情况下，专用的网络存储服务器比传统文件服务器读速度有大幅度提高，最大提高幅度63%，最小幅度为13%。

5.2 专用附网存储服务器连续写速度测试

在连续写速度测试时我们用有18个磁盘的RAIDO，文件分块大小为32kB-64kB，RAIDO数从1～3，测试结果如表2所示，从表中我们可以看出专用的附网存储服务器对连续写也有大幅度提高。

表2 文件块大小为64kB和RAIDO数为3连续写速度的测试

RAIDO	专用网络存储服务器 连续写速度(M字节/秒)			传统文件服务器 连续写速度(M字节/秒)		
请求块大小BS	每个RAIDO所含磁盘数DN			每个RAIDO所含磁盘数DN		
	9	12	18	9	12	18
1MB	40.1	48.2	54.2	28.6	38.2	46.1
50MB	59.5	55.3	65.3	48.1	45.3	54.9
100MB	64.3	60.1	76.4	52.8	48.9	63.2

从上表可以看出，在同样的条件下，专用的网络存储服务器连续写速度比普通文件服务器速度有大幅度的提高，最大提高幅度为40.2%，最小幅度为17.6%。

小结 网络存储是近年来国际上比较热门的一个研究方向，而附网存储技术又是网络存储中较新的课题，本文所设计的专用附网存储服务器系统，通过试验验证，读写性能良好，下一步我们将进一步完善其功能，并将其成熟的产品推向市场。

参考文献

1 ANSI SCSI-3 Fast-20 Parallel Interface X3T10/1047D Working Group, Revision

2 Benner A F. Fibre Channel Gigabit Communications and I/O for Computer Networks. McGraw Hill New York, 1996

3 McDoetal D. A Quantitative Analysis of Cache Policies for Scalable Network File Systems 15th SOSP Dec. 1995

4 Hu Yming, Yang Q DCD-disk caching disk: A New approach for boosting I/O performace. In Proc of th 23rd Int Symposium on Computer Architecture (Philadephia Pennsylvania), May 1996. 169～178

5 Tsujoka T, Onoda T An ultra high-speed file server with 105 Mbytes/s read performance based on a personal computer. IEICE Trans. Commun, 1998, E81-B(12): 2503～2508

（上接第75页）

参考文献

1 Braden R, Zhang L, Berson S, et al Resource ReSerVation Protocol (RSVP)-Version 1 Function Specification. RFC 2205, Sep. 1997

2 Wroclawski The Use of RSVP with IETF Integrated Services RFC2210, Sep. 1997

3 Almquist P. Type of Service in the Internet Protocol Suite RFC1349, July 1992

4 Deering S, Hinden R. Internet Protocol, Version 6 (IPv6) Specification. RFC2460, Dec 1998

5 Blake S, Black D, Carlson M, et al. An Architecture for Differentiated Services. RFC2475, Dec. 1998

6 Nichlis K, Blake S. Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Header. RFC 2474, Dec. 1998

7 邓志成，王茜，等. 区分服务体系的服务质量机制浅析. 计算机科学，2000，27(6)

8 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. 计算机学报，2000，23(4)