一种用于QoS控制的报文分组调度与丢弃算法*

王晓春 张尧学

(清华大学计算机科学与技术系 北京 100084)

E-mail: wxc@sun475.cs.tsinghua.edu.cn

摘要 提出了一种用于Internet中服务质量控制的报文分组调度与丢弃算法.该方法根据用户要求的服务质量(quality of service,简称QoS)参数和多媒体应用的类型以及等待时间等因素,综合调度路由器中所到达的报文分组和分配缓冲,使其能够满足用户所要求的服务质量.计算机模拟表明,该算法的转发性能等指标要优于当前常用的调度算法——加权公平队列法(weighted fair queuing,简称WFQ),从而提高了网络传输多媒体信息的能力.

关键词 网络,QoS,多媒体,路由器.

中图法分类号 TP393

随着Internet的飞速发展,人们对于在Internet上传输分布式多媒体应用的需求越来越大.一般来说,用户对不同的分布式多媒体应用有着不同的服务质量要求,这就要求网络应能根据用户的要求分配和调度资源.然而,传统的Internet只提供尽力而为(best effort)服务的转发机制,无法为用户提供高质量的声音、图像等多媒体传输服务.为了解决在Internet等计算机网上高质量地传输多媒体信息的问题,美国于1996年底开始了以提高网络服务质量研究为核心的Internet II以及NGI(下一代Internet)等研究项目.IETF(Internet Engineering Task Force)也成立了专门的工作小组来研究多媒体服务质量的定义及相关标准[1~3].

网络服务质量(quality of service,简称QoS)是网络与用户之间以及网络上互相通信的用户之间关于信息传输与共享的质的约定[4],例如,传输延迟允许时间、最小传输画面失真度以及声像同步等.在Internet等计算机网络上为用户提供高质量的QoS必须解决以下问题:

(1) QoS的分类与定义.对QoS进行分类和定义的目的是使网络可以根据不同类型的QoS进行管理和分配资源.例如,给实时服务分配较大的带宽和较多的CPU处理时间等.另一方面,对QoS进行分类定义也方便用户根据不同的应用提出QoS需求.

(2) 准入控制与协商.即根据网络中资源的使用情况,允许用户进入网络进行多媒体信息传输并协商其QoS.

(3) 资源预约.为了给用户提供满意的QoS,必须对端系统、路由器以及传输带宽等相应的资源进行预约,以确保这些资源不被其他应用所抢用.

(4) 资源调度与管理.对资源进行预约之后,是否能得到这些资源,还依赖于相应的资源调度与管理系统.

迄今为止,Braden等人提出了网络资源预约协议RSVP(resource reservation protocol)[1],Shenker等人提出和定义了保证式服务[3]和负载控制式服务[2].但是,这两种服务都只考虑了平均延迟和平均丢失率,不利于多媒体信息传输时的抖动及断续等服务质量的反映,我们在文献[5]中提出了考虑抖动等服务质量的适时服务并

* 本文研究得到国家自然科学基金(No.69873024)、国家863高科技项目基金(No.863 306-ZD-05-01)和国家973高科技项目基金(No.G1998030406)资助.作者**王晓春**,1968年生,博士生,工程师,主要研究领域为计算机网络中服务质量的控制方法.**张尧学**,1956年生,博士后,教授,博士生导师,主要研究领域为计算机网络,网络路由器,网络协议工程,服务质量控制方法,网络操作系统.

本文通讯联系人:王晓春,北京 100084,清华大学计算机科学与技术系结构教研组

本文1999-03-23收到原稿,1999-05-21收到修改稿

定义了相关参数.本文以RSVP协议为基础,使用文献[5]中所定义的QoS参数,提出和设计了一种报文分组调度与丢弃算法.该算法用于路由器中,按照RSVP传递的相关消息中的参数,为用户调度和分配缓冲区与CPU资源,提供用户满意的适时服务.

1 QoS控制与服务类型

如何在Internet上提供集成服务的关键是QoS控制.为此,IETF把QoS控制问题划分为两大部分,即集成服务模型与QoS实现框架.IETF是这样定义QoS的:用带宽、分组延迟和分组丢失率等参数描述的关于分组传输的质量.传统的Internet只提供单一的服务质量,即尽力而为(best effort)服务.在该服务中,可利用的带宽以及相应的延迟特性取决于网络中的负载状况.

为了进一步地描述QoS的控制过程和服务类型,IETF把QoS定义为一个两维空间:

〈服务类型〉,〈参数类型〉

服务类型与参数类型两者都用整数表示.

〈服务类型〉的取值范围为[1,254],但目前IETF只提供两种服务:编号为2的保证式服务和编号为5的负载控制式服务.保证式服务保证所传输的信息流能在所要求的延迟时间内到达,主要用于对时间要求非常严格的多媒体应用.而负载控制式服务则并不保证所传输的信息流都能在所要求的延迟时间内到达,但其到达的百分比会相当高,它主要对应于Internet上广范围的多媒体应用.

正是为了支持广范围的应用,负载控制式服务对提供的服务质量所作的描述是不定量的,本质上不能为有明确QoS要求的多媒体应用提供QoS保证.并且,负载控制式服务与保证式服务都只考虑了应用的带宽和延迟要求,不利于多媒体信息传输时的抖动及断续等服务质量的反映,因此,它们都难以满足多媒体信息的QoS要求.为此,我们设计了一种编号为3的Internet集成服务类型:适时服务[5],它具有以下特点:

- 充分考虑应用的QoS要求,包括延迟、抖动、丢失率以及同步;
- 充分考虑应用对QoS要求的不同级别,便于网络系统采用最有效的资源分配和管理机制;
- 允许网络系统通过价格方法指导用户对网络资源的使用,提高网络的整体性能;
- 能够为应用提供明确的不同级别的QoS担保.

与〈服务类型〉相同,〈参数类型〉的取值范围也是[1,254].区间[1,12]是保留区间,专门用于指定那些供所有服务公用和共享的参数.例如,当前可利用的带宽等.该区间的参数值与服务类型值1一起组成公用共享参数.例如,〈1,5〉表示一个可供各种服务共享的QoS参数.区间[128,254]由服务规范的设计人员给定,它们不是共享的,只针对相应的服务类型.

2 适时服务QoS参数与RSVP消息传递

2.1 适时服务QoS参数

在文献[5]中,我们提出了适时服务的QoS描述方法,即把QoS定义为一个由应用(服务)类型、流量和性能参数、不同媒体流之间的同步度以及反映网络资源使用状况的价格参数等组成的4元组:

$$(S,F,T,C)$$

其中S是描述应用(服务)类型的集合,它包括CBR(constant bit rate)类应用(例如,高保真音频视频)和RT-VBR(variable bit rate:real time)类应用(例如,电视会议类实时音频视频).它还包括NRT VBR(variable bit rate:non-real time)类应用和UBR(unspecified bit rate)与ABR(available bit rate)类应用.

F是描述信息流的流量特性与性能特性的集合.$F=(d,j,l,p,x_0,x_1,i,b)$,其中d表示端到端的延迟,j表示抖动,l表示丢失率,p表示过度延迟率.参数d,j,l,p为信息流的性能特性参数,一般由信息接收端提出.

x_0表示传送分组的峰值速率,x_1表示传送分组的平均速率,i表示传送分组的平均周期,b表示传送分组时的最大突发长度.参数x_0,x_1,i,b表示信息流的流量特性参数,一般为信息发送端,即源端提出.

T和C分别表示信息流之间的同步程度和网络资源的价格参数.

支持适时服务的 QoS 控制参数主要有 4 类对象:SENDER-TSPEC,FLOWSPEC,ADSPEC 和 POLICY DATA[11]. 前 3 类对象的意义已由 IETF 所规定,只是实际的格式定义不同. 而第 4 类对象 POLICY_DATA 是用于传输用户对请求的服务可以接受的价格,以便网络根据当前的资源定价策略完成用户 QoS 请求的准入控制,并在应用运行时监视应用的信息流量及 QoS 要求的变化.

2.2 RSVP 消息传递[11]

上述 QoS 参数是利用 RSVP 协议进行传递的. RSVP 是于 1997 年 9 月通过的 Internet 国际标准,它由多媒体信息的接收端开始进行资源预约,可适用于组广播和点对点通信. 对适时服务进行 QoS 控制的 RSVP 消息传递的过程如下:

Step 1. 发送端主机根据用户输入的流量特性参数生成 SENDER-TSPEC 对象,并生成相应的 PATH 报文,发往下一个路由器;

Step 2. 路由器将 PATH 报文中的信息保存在相应的数据结构 PSB[11]中,生成带有本地服务支持信息的 ADSPEC 对象,并修改 PATH 报文,然后根据接收主机的地址寻径,直至将 PATH 报文送达接收端主机为止;

Step 3. 接收端主机根据用户输入的 QoS 参数以及 PATH 报文中所携带的 SENDER-TSPEC 对象,生成 FLOWSPEC 对象和 POLICY_DATA 对象,并生成相应的 RESV(reservation)报文,发往上一个路由器;

Step 4. 路由器根据所接收到的 RESV 报文进行 QoS 协商,并将信息保存在相应的数据结构 RSB[11]中;

Step 5. 若该路由器上的资源无法满足用户的 QoS 要求,它将生成预留错误报文 RESV_ERR,发回接收端主机,用户可以等待,或者更改 QoS 要求进行 QoS 重协商,或者放弃服务请求;若该路由器上的资源可以满足用户的 QoS 要求,它将为该用户应用建立资源预留软状态,然后向上一个路由器发送 RESV 报文;

Step 6. 重复上述过程,直至为应用建立一条发送端到接收端的适时通道为止.

3 报文分组调度与丢弃算法

本文提出的报文分组调度算法就是基于前面所定义的 QoS 参数,并且和报文分组丢弃策略一起使用. 这是因为,即使 RSVP 协议根据 QoS 要求对资源进行了预约,但仍不能排除某些媒体流会在某个时间段产生大量突发报文分组从而使得缓冲产生拥塞的可能. 在这种情况下,必须放弃缓冲中已存在的部分报文分组,从而将缓冲资源让给优先级更高的报文分组.

报文分组调度算法.

Step 1. 决定不同类型报文分组队列的权重(报文分组队列类型由第 2.1 节中 QoS 的参数 S 给出)

If 在路由器已建立了该媒体流的软状态

Then a_0 或 a_1 类型报文分组队列的权重$=(x_0/r)+c$

a_2,a_3 或 a_4 类型报文分组队列的权重$=x_1/r$

这里,x_0,x_1 分别为第 2.1 节给出的报文分组传输峰值速率与平均速率,而 r 则是路由器 CPU 的处理率,c 为调节常数,a_0,a_1,a_2,a_3 与 a_4 分别表示队列类型.

Step 2. 根据权重和等待时间决定每个队列的优先级 Pri:

$$Pri=\text{权重}+a_1\times\text{等待时间}.$$

这里,a_1 是一个大于 0 的常数,该公式意味着大权重和长等待时间的队列具有高优先级.

Step 3. 分配 CPU 处理时间 f 给具有最高优先级的报文分组队列. 这里,分给不同队列的 CPU 处理时间是不同的,其计算公式为

$$f=a_2b-a_3d+a_4j-a_5l-a_6p.$$

这里,a_2,a_3,a_4,a_5 和 a_6 是大于 0 的常数,b,d,j,l 和 p 分别是第 2.1 节所给出的 QoS 参数的最大报文分组突发长度、端到端传输延迟、抖动、丢失率和允许过度延迟的概率. 该公式意味着一个具有大的分组突发长度、大的抖动以及小的传输延迟、小的丢失率和小的过度延迟率要求的报文分组队列将获得大的 CPU 服务时间.

不同的 CPU 处理时间段防止某个队列占用 CPU 时间过长从而使其他队列无法即时获得

CPU.

Step 4. Goto Step 1.

报文分组丢弃算法.

If 一个到达报文分组属于具有最大丢失率的队列,且该队列已放弃的报文分组小于所规定的分组丢失率 l_i;

Then 放弃该分组;

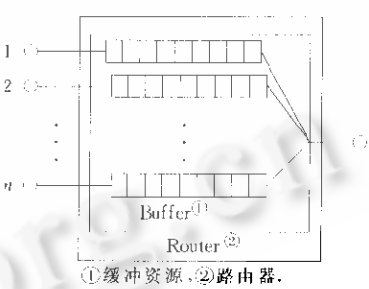
Else 从已存在的具有最大丢失率的队列中选择最后到达的分组丢弃.

上述调度与丢弃算法的主要思想是:在路由器中为每类流开设专门的缓冲队列,并根据各类流的服务质量要求为这些队列分配权值.调度时,综合考虑各队列的权值和等待时间,以决定此次调度要处理哪个队列.然后根据该队列中报文分组的性能参数分配CPU的处理时间.这可以与当前较常用的调度算法WFQ(weighted fair queuing)进行比较,WFQ也是在路由器中为每类流开设专门的缓冲队列,并根据各类流的服务质量要求为这些队列分配权值,但此权值是用来决定相应队列所应分配的CPU处理时间的,至于调度哪个队列,则采用公平轮转的办法.由此可见,我们的算法比WFQ考虑了更多的因素,可以根据不同类型的流及等待时间,灵活地调度较急迫的队列进行处理,并且可以灵活地分配对选中队列的处理时间,所以其性能也应更优.

4 调度性能模拟

4.1 模型的建立

为了评价我们所提出的报文分组调度与丢弃算法的性能,我们建立了以下模型来进行计算机模拟.假设在路由器中有固定数目的缓冲资源,该路由器有 n 个输入口,各类流从各个入口同时到达,争用同一条输出链路.如图1所示.



①缓冲资源,②路由器.

Fig. 1 Experiment model

图1 实验模型

在模拟过程中,我们假设有3类流同时到达:第1类流为CBR流,它要求没有传输延迟、不允许有报文分组丢失;第2类流为RT VBR流,它要求有很小的传输延迟和很低的丢失率;第3类流为ABR流,我们假设它的到达过程服从泊松分布.

评价的方法主要是将我们的算法与现有算法中较常用的一种WFQ进行比较,比较从以下4个方面进行:

- 总体性能,对我们的算法与WFQ的性能有一个量的概念;
- 在满足其他流性能要求的前提下,测量某一类流的平均延迟与平均丢失率;
- 在需要达到相同性能指标的前提下,两种算法对资源的需求情况;
- 两种算法在各种负载情况下的吞吐量.

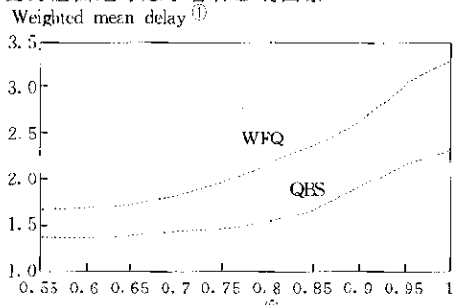
下面,我们分别给出模拟结果,为简单起见,在下面的叙述与图示中,我们以记号QBS(Qos based scheduling)来表示我们的算法.

4.2 总体性能

我们已经知道,不同的流应有不同的性能测量尺度.例如,语音(CBR流)要求极小的传输延迟和很低的丢失率,而电子邮件(UBR流)则要求传输时不能有任何差错或丢失,但对延迟要求不高.所以,我们对各种负载情况下的加权平均延迟和加权平均丢失率进行了测量,以评价QBS和WFQ的总体性能.模拟结果如图2所示.

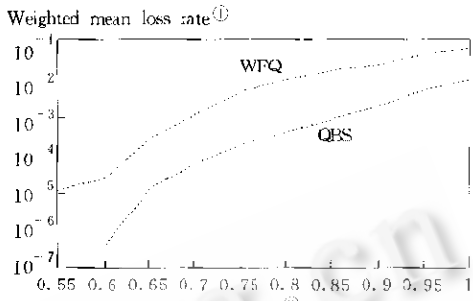
在图2(a)中,横坐标为归一化的负载量,纵坐标为两种算法分别处理上述3类流的加权平均延迟.从图中可以看出,无论在何种负载状况下,QBS的加权平均延迟都比WFQ的加权平均延迟要小.与此相似,在图2(b)中,横坐标为归一化的负载量,纵坐标为两种算法分别处理上述3类流的加权平均丢失率.从图中也可以看出,无论在何种负载状况下,QBS的加权平均丢失率都比WFQ的加权平均丢失率要小.所以,QBS的性能要优于WFQ的性能,这是因为QBS在进行调度时,充分利用了报文分组所提供的QoS参数和各队列的等待时间等信息,比

WFQ 更为全面地考虑了各种影响因素.



①加权平均延迟,②负载.

(a) Weighted mean delay
(a) 加权平均延迟



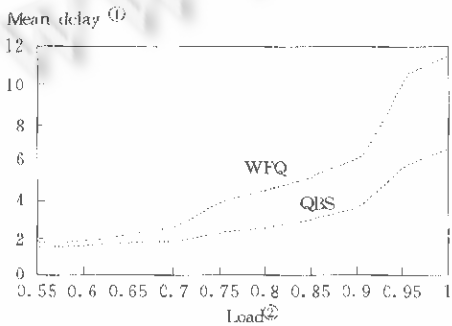
①加权平均丢失率,②负载.

(b) Weighted mean loss rate
(b) 加权平均丢失率

Fig. 2 Three traffics' performance
图2 3类服务流的总体性能

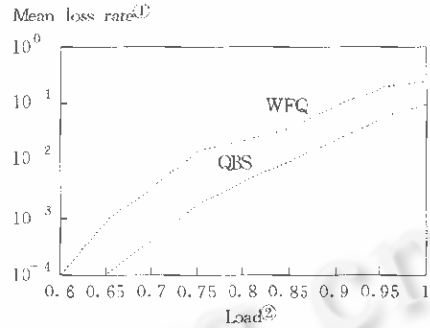
4.3 某一类流的性能

在这项模拟实验中,无论负载多重,我们都保证满足 CBR 流和 VBR_RT 流的性能指标要求,然后测量 ABR 流的平均延迟与平均丢失率,结果如图 3 所示.



①平均延迟,②负载.

(a) Mean delay
(a) 平均延迟



①平均丢失率,②负载.

(b) Mean loss rate
(b) 平均丢失率

Fig. 3 One certain traffics' performance
图3 某一类服务流的性能

在图 3(a)中,横坐标为归一化的负载量,纵坐标为两种算法分别处理 ABR 流的平均延迟.从图中可以看出,无论在何种负载状况下,QBS 的平均延迟都比 WFQ 的平均延迟要小.与此相似,从图 3(b)中,横坐标为归一化的负载量,纵坐标为两种算法分别处理 ABR 流的平均丢失率.从图中也可以看出,无论在何种负载状况下,QBS 的平均丢失率都比 WFQ 的平均丢失率要小.这是因为,在网络资源固定的情况下,QBS 能"适可而止"地为各类流提供服务:首先保证 CBR 流的要求,然后根据 VBR_RT 流的性能要求分配资源,其余的资源则都用于服务 ABR 流.但 WFQ 因不能充分利用报文分组中的信息,过多地为 VBR_RT 流分配了资源,从而导致 ABR 流的性能较差,这种情况在负载越重,即资源越紧张时越明显.

4.4 资源需求

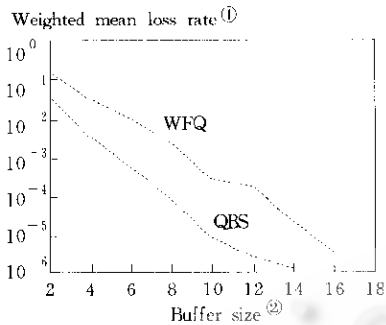
为了比较 QBS 与 WFQ 对资源的需求情况,我们固定负载量,然后测量在各种缓冲资源条件下的加权平均丢失率,结果如图 4 所示.

在图 4 中,横坐标为路由器中总共的缓冲个数,纵坐标为两种算法分别处理上述 3 类流的加权平均丢失率.从图中可以看出,在相同的缓冲资源条件下,QBS 的加权平均丢失率比 WFQ 的加权平均丢失率小,或者说,为了达到相同的丢失率指标,QBS 所需的缓冲资源比 WFQ 所需的缓冲资源要少.例如,在路由器中缓冲个数均为 10 时,QBS 算法实现的加权平均丢失率约为 10^{-5},而 WFQ 则要达到 10^{-3} 数量级;或者说,要想使该路由器具有

10^{-5}的加权平均丢失率,用QBS实现时需10个缓冲资源,而WFQ则要用15个缓冲资源.

4.5 吞吐量

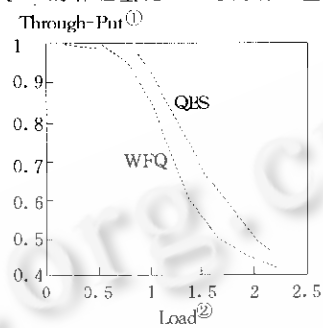
最后,我们对QBS与WFQ算法的吞吐量进行了比较,模拟结果如图5所示.图中,横坐标为归一化的负载量,纵坐标为两种算法分别处理上述三类流的吞吐量,结果表明,QBS的吞吐量比WFQ的吞吐量要大.



①加权平均丢失率,②缓冲大小.

Fig. 4 Resource requirement

图4 资源需求



①吞吐量,②负载.

Fig. 5 Through Put

图5 吞吐量

5 小 结

本文提出和定义了能够在网络上为用户提供适时服务的QoS参数,并根据该定义设计了一种报文分组调度与丢弃算法.该方法根据用户要求的服务质量(QoS)参数和多媒体应用的类型以及等待时间等因素,综合调度路由器中所到达的报文分组和分配缓冲,使其能够满足用户所要求的服务质量.并在计算机上建模,将我们的算法与目前较常用的WFQ算法进行了多方面的比较,结果表明,该算法的各项性能指标均优于WFQ.

将此算法应用于当前Internet或新一代高速Internet的路由器中,可更好地为用户提供满意的适时服务,提高网络传输多媒体信息的能力.

参考文献

1 Braden R, Zhang L, Berson S *et al*. Resource reservation protocol (RSVP). Version 1. Functional Specification, IETF RFC2205, Sept. 1997. http://www.retf.org

2 Wrodawski J. Specification of the controlled load network element service. IETF RFC2211, Sept. 1997. http://www.retf.org

3 Shenker S, Partridge C, Guerin K. Specification of guaranteed quality of service. IETF RFC2212, Sept. 1997. http://www.retf.org

4 Chen Hua, Zhang Yao-xue. Multimedia QoS classification and negotiation manager. Chinese Journal of Electronics, 1998, 7(1):44~48

5 Zhang Yao-xue, Gai Feng. QoS control and multicast Mbone. ACTA Electronica Sinica, 1995,23(10):32~36
(张尧学,盖峰.QoS控制与成组广播Mbone.电子学报,1995,23(10):32~36)

A Scheduling and Dropping Algorithm for Packets in QoS Controlling

WANG Xiao-chun ZHANG Yao-xue

(*Department of Computer Science and Technology Tsinghua University Beijing* 100084)

Abstract In this paper, the authors propose a scheduling and dropping algorithm for packets in QoS (quality of service) controlling for Internet. This method is based on the QoS parameters required by users, the types of multimedia and the waiting time, etc. It schedules the packets arriving in a router and allocates the buffers to meet the quality of services. The computer simulation shows that the forwarding performance of the proposed algorithm is superior to the WFQ's (weighted fair queuing), which is the common scheduling algorithm used in routers. So, this algorithm improves the transmission of multimedia information on Internet.

Key words Network, QoS, multimedia, router.