区分服务结构及其 TCP 性能分析

隆克平[1,2],白 刚[2],程时端[2],陈俊亮[2],张润彤[3]
(1. 澳大利亚墨尔本大学,ARC 超亮带信息网研究中心;2. 北京邮电大学程控交换与通信网国家重点实验室,北京 100876;3. Nokia 北京研究开发中心,北京 100013)

摘 要: 本文首先对区分服务(DiffServ)结构中的边缘路由器和核心路由器机制进行了系统的分类研究,分析并比较了各种机制.接着,对区分服务结构中 TCP 的性能问题进行了研究,总结了国内外对这一问题的仿真、解析模型分析和实验研究的成果.找出了关键的问题所在,并根据我们的研究成果提出了一些区分服务结构本身的改进建议.

关键词: 服务质量; 区分服务; 标记器; 主动队列管理; TCP 性能
中图分类号: TN915.04 **文献标识码**: A **文章编号**: 0372-2112 (2001) 11-1540-06

Implementing Mechanisms and TCP Performance in DiffServ Capable IP Networks

LONG Ke-ping[1,2],BAI Gang[2],CHENG Shi-duan[2],CHEN Jun-liang[2],ZHANG Run-tong[3]
(1. *ARC Special Research Centre for Ultra Broadband Information Networks (CUBIN), Department of Electrical and Electronic Engineering, The University of Melbourne*; 2. *Beijing University of Posts & Telecommunications (BUPT), Beijing* 100876, *China*; 3. *Nokia China R&D Center, No.* 11, *He Ping Li Dong Jie, Beijing* 100013, *China*)

Abstract: Firstly, the Diffserv QoS architecture and various edge router marking mechanisms and core router drop mechanisms are discussed in detail. Secondly, This paper examines the several analytic model studies and many simulations on TCP(Transport control protocol) performance in DiffServ Capable IP(Internet protocol) networks. Through analysis, we find that TCP flow using current DiffServ AF PHB can't always achieve assured rate whether by srTCM/trTCM or by TSWTCM. The factors that influence TCP performance are: RTT, target rate, TCP/UDP coexistence and the number of flows in aggregate. What wrong with DiffServ and TCP? We conclude that current DiffServ architecture needs to be enhanced and some per-flow state must be maintained to provide predictive fair service for TCP flows.

Key words: QoS; diffServ; marker; active queue management; TCP performance

1 引言

今天的 Internet 是基于传统的 Transport Control Protocol/Internet Protocol(TCP/IP)技术,仅能提供尽力而为(Best-effort)传送的业务,没有明确的时间和可靠性传送保障.对于传统的数据业务是充分的,但随着音频/视频会议、视频点播(VoD)、远程教育等实时多媒体业务、WWW 业务、电子商务在 Internet 网上传送,这些不同的应用将有不同的 QoS 需求,如:不同的带宽、延迟和抖动需求.因此,提供不同业务 QoS 支持将是 Internet 网络能否成功商用化的一个关键性和挑战性课题[1~6],并已成为国际上的一个研究热点.

IETF 已经标准化了支持 IP QoS 的几个新的协议和技术,包括:综合服务/资源预留(IntServ/RSVP)[3,4]对每一个流提供定量的保证,而区分服务(DiffServ)结构[5~7]通过采用分组头 DSCP 标记的每一跳转发行为(Per-hop Behavior,PHB)提供相对的 QoS 保证.多协议标签交换(MPLS)[8,9]将三层路由的灵活性与二层交换的 QoS 保证及快速紧密结合起来,并提供显式路由和业务量会聚.业务量工程能更优化的分配网络中的业务量,以达到负载均衡和提高网络吞吐量.服务质量路由(QoS Routing)[10,11]可以找到满足 QoS 参数(如:带宽、延迟、抖动和丢失率)要求的路由.这些结构既可相互独立,又可协同工作.

其中,区分服务结构由于其好的扩展性和实现简单已成为首选方式和研究的热点[5~7,15,16,22,23,29~36],最大的研究计划是由美国的 170 所大学、40 家公司和 30 个其他组织参加的 Internet2 Project(已组建了 Qbone 实验床对上述问题进行研究[12,13]).但是,由于区分服务是一个新提出的 IP QoS 结构,目前这些研究尚处于不成熟阶段,值得研究的课题很多.急待解决的问题包括:如何确保 TCP 和 UDP 流的公平性、如何实现不同 Diffserv Domain 之间的有效互通及 SLA、区分服务结构的计费和安全问题、如何有效的结合 Interserv 和区分服务结构以实现端到端 QoS 保障.

收稿日期:2000-12-25;修回日期:2001-05-08
基金项目:国家自然科学基金项目(No. 69972008)和 Nokia. BUPT 合作项目联合资助

本文对区分服务结构及其实现机制进行了系统的分类比较研究,对区分服务结构下的 *TCP* 性能进行了分析,并根据我们的研究成果得出了重要结论和提出了一些新的建议.

2　区分服务结构中的路由器机制研究

2.1　区分服务边缘路由器和和核心路由器机制分类研究

如图 1[31]所示,区分服务结构有三个主要部件:(1) QoS 策略/资源管理器,(2)边缘器件功能块,(3)核心路由器功能块.其中,策略/资源管理器允许网络运营商确定 QoS 策略——即:哪些业务接受网络的哪一类服务.这些高级的策略被翻译成器件级策略并被下载到路由器.最初,策略/资源管理器只包括一些简单的资源管理,理想情况下策略/管理器应支持动态 QoS 请求和业务分配,然而动态网络资源分配领域仍有许多问题急待研究和解决.边缘器件可以是典型的路由器也可以是网关,还可以是信任的终端主机,它根据网络运营商确定的策略执行复杂的基于 Per-flow 的分类,还执行计量、业务整形或管制、标记等功能——又称为业务量调节.业务整形是限制用户业务量服从业务描述的一种方法,管制是整形的另一种选择,它追踪用户业务量并和业务描述比较,超出业务描述的分组可能被丢弃或被标记,标记是根据计量或整形的结果在 DS 域编码以标识会聚流在核心路由器接受的 PHB (Per-Hop Behavior)转发.提供给特定分组流的服务是通过边缘路由器(入口或出口)的业务量调节和核心路由器的一些 PHBs 来共同实现的.不同的 PHB 和不同的业务量调节机制,区分服务将提供不同业务以区分的服务.由于将复杂的分类和业务量调节推到边缘而维持核心路由器的简单性,使得 DiffServ 比 IntServ 更具扩展性.

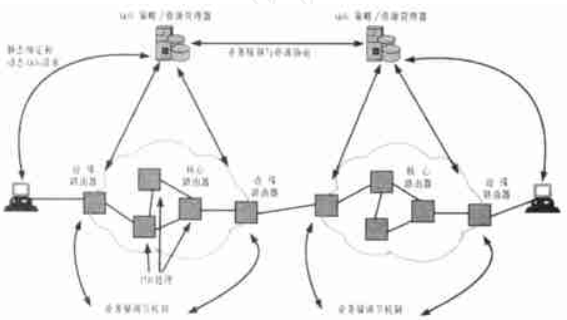


图 1　区分服务结构的网络模型

尽管有许多不同的区分服务实现方法和结构[5～7,15,16,22,23,29～36],就边缘器件和核心器件的功能实现的本质上讲,大体可以分类为:边缘路由器机制和核心路由器机制.核心路由器机制根据缓存管理和调度策略又分为优先权调度(PS)和阈值丢弃(TD).图 2[29]所示为两个优先级的阈值丢弃(TD)机制,其中,B 为整个缓存大小、B_l^{TD} 表示低优先级的缓存阈值、$B(t)$ 为时刻 t 的缓存占有量.所有接受的分组在单个缓存队列中排队并根据 FIFO 调度策略被服务,本文后面讨论的 RED/RIO 及其改进机制均属于 TD.只要 $B(t) < B$,则较高优先级分组被接受;而只有 $B(t) < B_l^{TD}$ 时,低优先级分组才被接受,此时允许两个优先级分组共享缓存.图 3[29]所示优先权调度(PS)将缓存分为多个队列,此处为两个:一个为高优先级队列,另一个是低优先级队列,根据队列被服务的策略不同而有 PQ、WFQ 等.

边缘路由器根据是否转发超出业务量描述的分组进入网络又细分为:边缘丢弃(ED)和边缘标记(EM);ED 只转发 IN 包进入网络而丢弃所有 OUT 包,EM 是将 IN 和 OUT 包都转发到网络内部,只不过打上不同标记.一个业务实际上可以定义为边缘标记机制和内部路由器机制的一种组合,即:ED 或 EM 和 TD 或 PS 的组合,分别如图 4 和图 5 所示[29,40].

这就提出一个问题,给定应用的 QoS 要求,哪种边缘标记机制和核心路由器机制能使区分服务结构更好地满足应用的要求?Sahu[29]经过两种边缘机制和核心路由器机制建立的服务的 QoS(延迟和损失)性能的定量分析表明:更准确的答案将依赖于业务模型和应用要求.也就是说不同的业务量模型和应用的要求需要不同的边缘和核心机制,很难找到一种万能的组合.为此,本文对适合不同应用和业务量模型的边缘机制进行比较研究,并讨论它们适合的应用,以为区分服务结构的研究和开发提供指导.

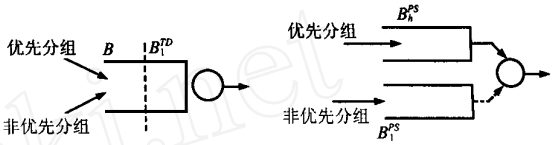


图 2　域值丢弃(TD)

图 3　优先权调度(PS)

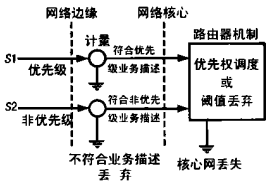


图 4　边缘丢弃机制(ED)

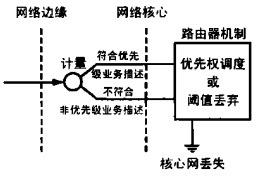


图 5　边缘标记机制(EM)

2.2　边缘标记机制研究

这里只讨论边缘标记机制,区分服务结构中更准确的定义是边缘业务量分类和调节机制(Traffic Conditioning),但大多数文献都简称为边缘标记.目前出现的标记机制大体可以分为如下三类:

(1)基于令牌桶的标记器,包括:RFC2597 规定的单速率三色标记器(srTCM)[16]、RFC2598 定义的双速率三色标记器(trTCM)[17]、公平标记器(FM)[18]和多媒体颜色标记器(mmCM)等.以工作在颜色盲模式下的 srTCM 为例,其工作原理是:srTCM 计量业务流量并根据三个业务量参数(CIR、CBS 和 EBS)用两个令牌桶标记分组为绿(DP0)、黄(DP1)和红(DP2),核心路由器对不同颜色的包提供不同等级的服务,如提供不同的丢弃优先级.它必须设置其计量器工作模式和三个业务量参数:承诺的信息速率(CIR)、承诺的突发度(CBS)和过度突发度(EBS).其中,CIR 为令牌桶的令牌产生速率,等于目标速率或协商的确保速率.两个令牌桶的产生速率均为 CIR,故称为单速率.CBS 和 EBS 分别为两个令牌桶的大小.$T_c(t)$ 和 $T_e(t)$ 分别表示两个桶在 t 时刻的令牌数(以 bytes 计量),当一个长为 B btyes 的分组到达时:

® 如果 $T_c(t) - B \geq 0$,则分组标记为绿且 T_c 减少 B;否则:

® 如果 $T_e(t) - B \geq 0$,则分组标记为黄且 T_e 减少 B;否则:

® 分组是标记为红,且 T_c 和 T_e 都不减少.

该类标记器的优点是:不需要记录每一个流的状态和测量平均速率.但缺点是实现相对复杂,而且很难确保 TCP 性能,这在本文后面将详细讨论.

(2) 基于速率的标记器,包括 Wu[20]提出的自适应分组标记器(APM)和 TSWTCM[19].其中,TSMTCM 是目前区分服务结构中推荐使用的一种标记机制,其工作原理是根据比例调节反馈控制的简单的控制论原理,在网络拥塞时以可控的方式分配带宽,很适合 AF PHB 转发的业务量,如图 6 所示[19].它由速率估计器和标记器两个独立的部件组成,其得名源于速率估计器是对应一个记录历史状态并随时间滑动的时间窗口.速率估计器提供业务量到达速率的估计值,该速率是在一个时间窗口内到达的业务量平均值,故可以平滑业务量突发.一种适合 TCP 的速率估计算法如文献[20]所述,标记器是根据速率估计器的平均速率估计值以及两个特定的速率参数(承诺目标速率 CTR 和峰值目标速率 PTR)而给分组打上不同颜色的标记.

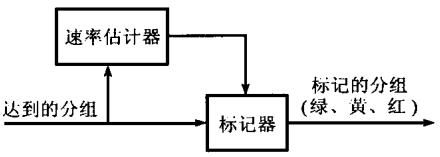


图 6 TSWTCM 原理图

TSWTCM 的优点是实现简单,且以一定的概率给分组打标记,这对 TCP 流非常有益,它减少了同一个窗口中丢弃多个包的可能性.此外,它可以在一个时间窗口平滑突发.但缺点是速率估计器不能准确的计量业务量速率.

(3) 基于策略的标记器,如 TCP-friendly 标记器[21]和我们的 TFTCM[22,23],是根据满足用户应用和 ISPs 要求的适当策略对基于令牌的标记器作适当的改进.尽管前者更强调保护小窗口流,这两种 TCP-friendly 标记策略的主要目的是减少突发丢包和 TCP 同步的概率,并将会聚的突发丢包转换成间隔的每一连接的非突发丢包,从而降低同一窗口的多个包丢弃($\geq$3)导致的超时次数,改善 TCP 流的吞吐量和流之间的公平性,增加网络的利用率.其工作原理为:根据可用令牌数和流的发送速率等参数,计算出优化的 IN 包标记间隔,然后根据这个 IN 标记间隔优化值间叉标记 IN 和 OUT 包.

显然,基于策略的标记器比单纯的令牌标记更能改善性能,而且可以根据应用和 ISP 的要求灵活的选择标记策略.但缺点是适当的策略确定较难.

综上所述,这三种标记机制各有利弊和各自的适用范围.事实上,区分服务中应用(TCP)的服务实现依赖于确保速率、标记机制和 TCP 动态性等多种因素.本文后面将详细研究这一问题.

2.3 核心路由器阈值丢弃机制研究

当业务到达速率超过了队列调度的服务速率时,路由器或交换机的队列长度增加,开始出现拥塞.减少队长就需要一些方法触发拥塞避免机制,通常,路由器的队列管理器必须接受两种类型的拥塞:瞬时拥塞和长期拥塞.当路由器采用基于 ToS/DS 的分类,几十、几百甚至几千个应用流都映射到同一个队列,必须应用反向的反馈控制系统给传输协议提供反馈信号以便降低发送速率,从而减少队长.反馈结构必须大致知道在给定时间内哪一个流实际上造成长期拥塞,反馈也有两种方式:分组标记(ECN)和分组丢弃方式.分组丢弃是一种好的方式,因为在 IP 网络中 TCP 用丢包来触发拥塞避免机制,本文主要讨论分组丢弃的方式.有一些结构通过引入统计反馈信号解决这一问题,其反馈信号强度是由如队列平均占有率和上游路由器在分组中的标记共同决定的.于是提出了一种被称为随机早期探测(RED)的反馈机制.

RED[23]作为一种主动排队管理技术(AQM)[25],它不同于传统的尾部丢弃(Tail Drop),而是探测初期的拥塞并通过丢包给源端主机提供反馈.RED 的动机是保持小的排队队长、减小突发和解决全局同步问题.自 1993 年 Floyd 提出 RED[24]以来,它一直是一个研究热点.先后出现了几种改进:ARED[26],FRED[27],以及近来讨论最多的 MRED,MRED 是一个用于描述不同丢弃优先级(不同颜色)的分组的丢弃概率需要独立计算的通用术语.图 7 示出了 RED 及其可能的变化形式分类.其中,WRED[28]和 RIO[7]是目前区分服务结构中核心路由器中广泛采用的实现 AF PHB 的两种丢包机制[14,15,29~31].下面结合 TCP 分别对 RED、WRED 和 RIO 给予讨论.

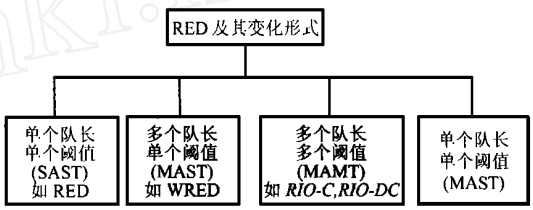


图 7 RED 及其变化形式分类

2.3.1 随机早期探测

Internet 研究任务署(IRTF)花了很多年研究产生,旨在控制由不同流共享的队列的占有率的主动的反向反馈的问题.该组得出的结论是:最好的解决方案(考虑已经存在的路由器实现)应是统计的随机的发布反馈信号,其强度是平均队列长度的递增函数(RFC2309[25]).IRTF 用于解释这个行为的一个特例被称为随机早期探测(RED)[24],有时被别称为随机早期丢弃,因为拥塞探测的结果常常导致丢弃.RED 将队列的平均对长作为决定拥塞避免机制是否应该被触发的随机函数的参数(为后面的讨论方便,假定触发器是"分组丢弃").当平均队列占有率增加,分组丢弃的概率增大,它属于单个平均队长单个阈值的 RED 类.图 8 示出了一个丢包概率函数的例子:

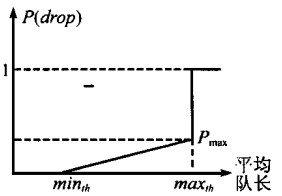


图 8 RED 丢弃概率随队长变化

® 当队列长度小于低的门限值 min_{th},分组丢弃概率为 0.

® 队列长度大于 min_{th},分组丢弃概率线性增加直到达到队列长度高门限 max_{th}时的丢弃概率 max_p.

® 队列长度大于 max_{th},分组全部被丢弃.

这三个阶段有时分别称作正常、拥塞避免和拥塞控制级. 最坏情况的最长队列大小被限制为 max_{th}(分组丢弃概率跳到 1). RED 在队列满之前提前开始触发拥塞标识. 路由器可对不同的队列支持不同 min_{th}、max_{th}和 max_p 值 ——平衡队列可用的空间、要求的队列数和使用每一个队列的业务类的延迟/抖动范围. 此外, W_q 应该对每一队列应适当不同.

平均队列长度在每一次分组到达时重新计算,且基于低通滤波器或瞬时队列长度的指数权值滑动平均(EWMA)公式如下:

$$Q_{avg} = (1 - W_q) \times Q_{avg} + Q_{inst} \times W_q$$

其中: Q_{avg}是平均队列长度; Q_{inst}是瞬时队列长度; W_q 是滑动平均函数的权值. W_q 影响平均队列长度参数追踪瞬时队列长度的紧密程度 ——较高的值是更积极,而较低的值是更保守. 选择该值的目的是:允许 RED 能忽略短期的瞬时拥塞而不至于引入分组丢失,但在每一个分组的延迟被影响或多个 TCP 流的拥塞避免同步之前,对可维持级的队列长度作出反应.

RED 统计丢弃策略具有如下优点:

® 它通过丢包对 TCP 产生一个反向的反馈机制,其反馈强度是路由器内部的拥塞程度的函数.

® 消耗一个队列容量的比例大的流需要更强烈的反馈.

® 在独立的会话共享同一个队列的拥塞避免时的同步必须最小化.

随机早期丢弃(在队列实际耗尽它允许的空间之前)增加了在队列长度变得太高之前平滑瞬时拥塞的可能性. 随机早期丢弃减少了同时使多个流受分组丢弃影响的可能性.

两个关键的假设成为基于分组丢弃的主动排队管理的基础:

® 许多或大多数造成瞬时拥塞的流是 TCP 流,因此能对早期分组丢弃的反向反馈作出反应.

® 丢弃的分组实际上是属于造成拥塞的 TCP 流.

缺乏每一流分类和排队意味着这些假设并不总是有效,尽管它们通常是合理的. 如果在拥塞期间到达的大多数分组实际上属于非响应流(恶意流)或 UDP 流,可以推断更合理的拥塞期间分组丢弃算法是可能惩罚导致拥塞的流. 造成拥塞的流的瞬时特性允许 RED 和它的改进机制注意相关的流,甚至在缺乏清晰的流的分组上下文时. 为此,出现两种 RED 改进:ARED[26]和 FRED[27].

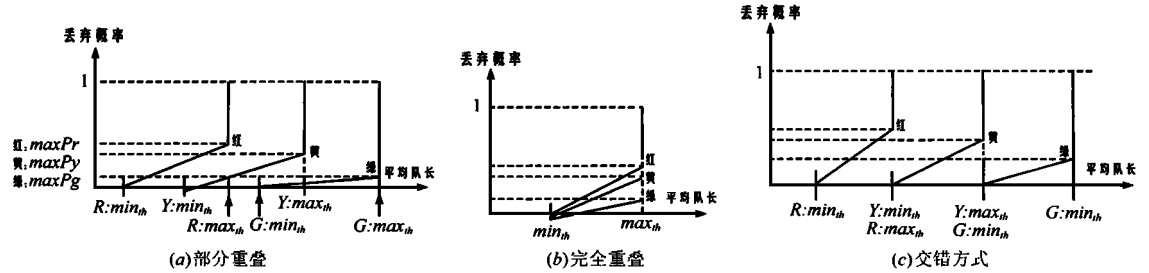
ARED 基于 Q_{avg}最近的滑动而动态调节 max_p,从而可以动态追踪队列中变化的业务量负载,负载变化是由于任何时间通过队列的 TCP 流的增加和减少,该算法可工作在不需要任何清晰地知道流数量情况下. FRED 算法通过基于短期的每一流状态调节流应该在队列中的分组数量,解决了当队列是由对早期拥塞通知具有不同反应的流(响应/非响应)共享时 RED 的不公平问题.

2.3.2　权值随机早期探测(WRED)

WRED[28]是目前区分服务(DiffServ)核心路由器广泛用于实现 AF PHBs 的队列管理机制[14,15,29~31]. WRED 根据所有颜色的分组到达数计算一个平均对长,并根据这些颜色(绿、黄、红)的整个分组数的变化更新这个单一的队列长度. 但多个 RED 阈值参数和丢弃概率参数被采用以便为不同颜色的包,不同颜色的参数设置又有三种方式,如图 9 所示,(a)部分重叠,(b)完全重叠,(c)交错方式. 文献[30]比较了这三种参数设置对 AF PHB 的实现性能的影响,结论是部分重叠方式更好.

2.3.3　具有 IN/OUT的随机早期丢弃(RIO)

传统的 RIO[7](具有 IN/OUT 的 RED)采用和 RED 同样的机制,但采用两套参数,一套用于标记为 IN 的分组,另一套用于标记为 OUT 的分组. IN 和 OUT 包分别对应于 DP0(绿)和 DP1(黄). RIO 的目的是在拥塞期间对 OUT 分组区分处理. 它是通过在同一个队列并行执行两个 EWMA 队列长度算法来实现该功能 ——$Qavg_{IN}$对 IN 分组和 $Qavg_{OUT}$对 OUT 分组. 我们称这种 RIO 为 RIO-C,即不同颜色的分组的平均队长计算是不同的,但是存在关联关系,如绿包的平均队长只用绿色包的数目计算,而黄包的平均队长的计算则依赖于绿包和黄包的数量,依次类推. 每一个颜色的包的丢弃依赖于为该颜色计算的平均队长及其参数设置,参数设置可以类似于图 8 所示. 结果是不仅 OUT 分组的丢弃概率曲线增长更快,而且 OUT 分组的滑动平均决定反应进入队列的 IN 和 OUT 分组的曲线. 进入队列的 OUT 分组的数量不影响 IN 分组的丢弃概率. 这一因素在一定程度上阻止了 OUT 分组的持续突发触发 IN 分组的流的不必要的拥塞避免.



(a)部分重叠　(b)完全重叠　(c)交错方式

图 9　WRED 不同的参数设置分组标记修改丢弃功能

另一种 RIO 可以被称为 RIO-DC,即:每一颜色的包的队列长度是完全独立的计算,如绿包的队长计算只依赖于绿包数,黄包的队长只依赖于黄包数,依次类推. 不同颜色的包有一套自己的参数,其余同 RIO-C 一样. 大量的研究表明:RIO 比 WRED 更能达到 AF PHB 的性能要求.

3 区分服务结构下的 TCP 性能问题及其改进

区分服务结构下的 TCP 性能是研究的热点,包括仿真和解析模型的研究,也提出了许多边缘标记机制和核心路由器的丢弃机制改进方案. 本文总结了这些研究成果并找出存在的关键问题,并提出了新的建议.

Clark 和 Fang[7]报告了不同的区分丢弃结构的仿真研究结果,他们提出了 RIO、时间滑动窗口(TSW)速率估计器和相应的标记结构,结论是:源目标速率可以在单个容量分配的网络中获得保证,但依赖于统计复用. 该文的结果表明:TCP 流的 IN 部分被保护不受 RTT 和非响应的 UDP 流的影响. Feng 等人[32]考察了类似于 RIO 的自适应标记结构,研究结果表明:服从业务描述的 TCP 流部分(IN)的吞吐量独立于 RTT,然而,超出部分(OUT)和尽力而为的业务流的吞吐量受 RTT 影响. 最后,该文指出:在资源充分情况下,预定的目标速率被忽略,而所有 RIO TCP 流等量共享确保的带宽.

Ibanez 和 Nichols[33]通过仿真研究表明:RTT,目标速率和 TCP/UDP 相互作用是在 RIO 类似的 AF 结构中影响一个流获取确保速率的几个关键因素,并得出结论:"确保服务不能给 TCP 流提供定量的服务",即:AF 不能清晰的定义和一致的速率保证. Seddigh 等人[30,34,35]用实验证实了是上述因素在网络资源充分时是影响剩余带宽不公平分配的关键因素,此外,会聚中的微流数和分组大小在网络资源充分情况下是决定获取带宽的主要因素. Pieda 等人[36]采用六种丢弃优先级分配方式和两种 RED 参数设置,研究了在区分服务结构中共享同一个 AF PHB 类的 TCP 和 UDP 流的公平性问题,结论是:无论是在资源充分和不充分情况下,目前的三个丢弃优先级均无法完全获得二者的公平性.

最近的研究[18,22,36～38]报告了一些新的方法以减轻上述因素的影响. 文献[37]提出了一个改进的 TSW 标记器和两个改进的 RIO 排队管理算法,仿真结果表明:改进的算法在不同目标速率、不同 RTT 和与 UDP 共存时改进了 TCP 吞吐量和公平性. 然而,提出的算法由于网络核心需要大量的状态而扩展性太差. Yeom[38]提出了改善在会聚中单个不同 RTT 流的公平性的算法.

与此同时,大量的解析模型[29,39～41]的通过不同业务量达到模型的详细分析也表明:目前的区分服务结构 AF 不能提供确保的业务区分. Sahu[29,40]试图通过解析模型考察:(1)是否可能通过适当的标记机制在 TCP 流之间提供业务区分?(2)在什么情况下,标记机制影响 TCP 流接受的服务?(3)如何选择适当的标记参数以提供给定的服务级? Sahu 在过分预定和未过分预定情况下,导出了边缘令牌标记和核心 RED 丢弃机制下的 TCP 流获得速率的简单而准确的解析模型(未考虑 TO),结果表明:(1)获得的速率不与确保的目标速率成正比,(2)不可能总是能获得确保的速率,(3)存在一个获得速率的范围值,在这个范围外,令牌漏桶参数变化无影响. 这说明通过唯一的设置令牌桶参数难以调节 TCP 流获得的业务级. 并通过在边缘标记和核心丢弃情况下的 TCP 吞吐量的随机模型分析表明:要获得好的业务特性和充分利用边缘标记的好处,TCP 窗口必须考虑边缘标记机制. Yeom 在另一些文章中则分析了成比例标记的 TSWTCM 对于不同 RTT 的 TCP 流是很不公平的.

本文作者在文献[23]中从控制论的角度建立了区分服务 AF 类的会聚控制 PID 模型,通过该模型的分析得出了重要结论:区分服务结构中 AF PHB 如果不同特性的 TCP 流以及 TCP/UDP 共享 AF PHB 的同一个类,将无法真正保证彼此的性能和完全的公平性. 而且,在发送速率和 RTT 相差太大的多个流共享同一类时,很可能导致系统的不稳定.

为此,作者根据这些研究和我们的研究成果[22,23,44～46]认为:

® 目前的区分服务结构的 AF PHB 不可能完全确保 TCP 和 UDP 流之间的公平性,也不可能保证会聚流中不同 RTT 的 TCP 流之间的公平性,更不可能同时解决 RTT、目标速率、会聚中大数量流、UDP/TCP 共存等因素对 TCP 流性能的影响.

® 要真正消除上述因素的影响,必须是对 TCP 窗口机制进行改进,使得窗口机制把标记结构考虑进去. 同时,必须研究 TCP-friendly 的速率控制机制[22].

® 目前的两种标记机制均不能有效保证 AF 的性能,应该在区分服务结构中增加边缘的智能,标记机制必须是 TCP RTT-Aware 的,并在标记时考虑 RTT 的因素. 同时增加边缘标记和 TCP 源端之间的反馈或通信功能,或者将标记工程集成在源端.

® 核心路由器需适当增加一定功能,处理部分的每一流状态. 即新的区分服务结构可以在核心维持少量的每一流状态,需要研究介于综合服务(IntServ)和区分服务结构之间的新的结构.

综上所述,尽管区分服务结构由于其实现简单和具有好的扩展性正被研究结构和各路由器厂商广泛关注和研究开发. 但目前的区分服务结构还需要改进,还存在许多研究课题:扩展性和 TCP 流性能保证等综合特性很好的区分服务结构及其有效的实现机制、业务级协商(SLA)、QoS 策略管理和 QoS 计费问题. 此外,有效的端到端的 QoS 策略和实现机制仍需要继续探索.

4 结束语

本文对区分服务结构的边缘路由器和核心路由器实现机制进行了系统的分类比较研究,并对区分服务结构下的 TCP 性能的解析和模拟研究进行深入分析,并根据我们的研究成果得出了一些重要的结论和提出了一些新的建议. 本文对这一领域的研究具有重要的指导意义.

作者简介:

隆克平 男. 1968 年 5 月出生于四川省通江县. 1998 年获电子科技大学博士学位,1998 年 9 至 2000 年 8 月,北京邮电大学博士后,现为副教授、硕士生导师. 主持和承担过国家级、省部级及国际合用项目 8 项. 发表学术论文近 50 篇. 主要研究方向:SDH/ATM 网络生存性、TCP/IP 协议改进机制及性能分析、增强 Internet 实时多媒体业务 QoS 保障的策略及其实现机制、IP/ATM 综合技术、移动 IP 技术及

应用、路由器的队列调度和缓存管理策略及算法等.

白　刚　男. 1973 年出生于陕西省. 现为北京邮电大学通信网国家重点实验室博士研究生,研究方向:VoIP 技术、ATM 技术、Internet 业务流量的模型及其对拥塞控制的影响.

参考文献:

[1] D Ferrari, L Delgrossi. Charging for QoS [DB/OL]. May 1998. http://www-ece.rice.edu/conf/iwqos98.

[2] P Ferguson, G Huston. Quality of Service [M]. John Wiley & Sons, 1998.

[3] Braden R, Clark D, et al. Integrated Services in the Internet Architecture: an Overview. RFC1633 [S]. IETF, Jun. 1994.

[4] R Braden, L Zhang, S Berson, S Herzog, S Jamin. Resource ReSerVation Protocol (RSVP) ——Version 1 Functional Specification. RFC2205 [S]. IETF, Sept. 1997.

[5] Steven Blake, David Black, et al. An Architecture for Differentiated Services. RFC2475 [S]. IETF, Oct. 1998.

[6] Y Bernet, S Blake, et al. A Framework for Differentiated Services, Internet draft <draft-ietf-diffserv-framework-02.txt> [S]. IETF, Feb. 1999.

[7] D Clark, W Fang. Explicit allocation of best-effort packet delivery service [J]. IEEE/ACM Transactions on Networking, Aug. 1998, 1(4): 397 - 413.

[8] E Rosen, A Viswanathan, R Callon. Multiprotocol Label Switching Architecture. Internet draft <draft-ietf-mpls-arch-07.txt> [S]. IETF, Mar. 2000.

[9] D Awduche, J Malcolm, et al. Requirements for Traffic Engineering over MPLS. RFC2702 [S]. IETF, Sept. 1999.

[10] E Crawley, R Nair, et al. A Framework for QoS-based Routing in the Internet. RFC 2386 [S]. IETF, Aug. 1998.

[11] G Apostolopoulos, R Guerin, et al. Quality of Service Based Routing: A Performance Perspective [DB/OL]. ACM SIGCOMM98, Sept. 1998. http://www.acm.org/sigcomm/sigcomm98/tp/paper02.pdf.

[12] J Sikora, B Teitelbaum. Differentiated services for internet2 [DB/OL]. Internet2 QoS Working Group Draft, http://www.internet2.edu/qos/wg/.

[13] B Teitelbaum. QBone Architecture (v1.0) [DB/OL]. Internet2 QoS Working Group Draft, http://www.internet2.edu/qos/wg/papers/qbArch/1.0/draft-i2-qbone-arch-1.0.html.

[14] Heinanen J, Baker F, Weiss W, Wroclawski J. Assured Forwarding PHB Group. RFC 2597 [S]. IETF, June 1999.

[15] Jacobson V, Nichols K, Poduri K. An Expedited Forwarding PHB. RFC2598 [S]. IETF, June 1999.

[16] J Heinanen, R Guerin. A Single Rate Three Color Marker. RFC2697 [S]. IETF, Sept. 1999.

[17] J Heinanen, R Guerin. A Two Rate Three Color Marker. RFC2698 [S]. IETF, Sept. 1999.

[18] Hyogon Kim. A Fair Marker. Internet draft <draft-kimr-fairmarker-diffserv-00.txt> [S]. IETF, April 1999.

[19] W Fang, N Seddigh, B Nandy. A Time Sliding Window Three Color Marker (TSWTCM). RFC2859 [S]. IETF, June 2000.

[20] W Feng, D Kandlur, et al. Adaptive packet marking for providing differentiated services in the internet [DB/OL]. http://citeseer.nj.nec.com/feng98adaptive.html.

[21] A Feroz, A Rao, et al. A TCP-friendly traffic marker for IP differentiated services [A]. Proc. IWQoS'2000 [C], Pittsburgh, PA, June 2000: 35 - 48.

[22] Qian Wang, Keping Long, et al. A TCP-friendly three color marker [A]. Submitted to ICC2001 [C], August, 2000.

[23] **隆克平. IP 网络的 QoS 机制及其网络生存性策略研究** [R]. **博士后研究报告. 北京邮电大学**. 2000 **年** 8 **月**.

[24] S Floyd, V Jacobson. Random early detection gateways for congestion avoidance [J]. IEEE/ACM Trans. on Networking, 1993, 1: 397 - 413.

[25] B Braden, D Clark, et al. Recommendations on Queue Management and Congestion Avoidance in the Internet. RFC2309 [S]. IETF, April 1998.

[26] W Feng, et al. A self-configuring RED gateway [A]. Proceedings of IEEE INFOCOM'99 [C], San Franscisco, CA, April 1999: 1320 - 1328.

[27] Dong Lin, Robert Morris. Dynamics of random early detection [A]. Proceedings of ACM SIGCOMM'97 [C], Cannes, France, Oct. 1997: 127 - 137.

[28] Distributed random early detection, reports of cisco [DB/OL]. http://www.cisco.com.

[29] S Sahu, D Towsley, J Kurose. A quantitative study of differentiated services for the internet [A]. Proceedings of IEEE Global Internet'99 [C], Rio de Janeiro, Brazil, Dec. 1999: 1808 - 1817.

[30] N Seddigh, B Nandy, et al. An experimental study of assured services in a diffserv IP QoS network [A]. Proceedings of SPIE'98 [C], Boston, Nov. 1998: 217 - 230.

[31] A Basu, Z Wang. A comparative study of schemes for differentiated services [R]. Bell labs technical report. August 1998.

[32] W Feng, D KAndlur, D Saha, K Shin. Adaptive packet marking for maintaining end-to-end throughput in a differentiated-services internet [J]. IEEE/ACM Transactions on Networking, Oct. 1999, 7(5): 685 - 697.

[33] Ibanez J, Nichols K. Preliminary Simulation Evaluation of an Assured Service. Internet Draft <draft-ibanez-diffserv-assured-eval-00.txt> [S]. IETF, Aug. 1998.

[34] Seddigh N, Nandy B, Pieda P. Bandwidth assurance issues for TCP flows in a differentiated services network [A]. Proceedings of Globecom'99 [C], Rio De Janeiro, December 1999: 148.

[35] N Seddigh, B Nandy, et al. An experimental study of buffer management scheme for diffserv assured forwarding PHB [A]. IC3C'2000 [C], Las Vegas, Oct. 2000.

[36] P Pieda, N Seddigh, B Nandy. The dynamics of TCP and UDP interaction in IP-QoS differentiated services networks [A]. The 3rd Canadian Conference on Broadband Research [C], November 1999. http://kabru.eecs.umich.edu/qos_network/diffserv/DiffServ_papers/papers/CCBR-Pie_06_fin.pdf.

(下转第 1548 页)

FCFS 策略,FCFS 的峰值吞吐量为 0.59.

表 4 正常吞吐量和业务量的对照关系

策略 3 (交换尺寸: $N=8$; 突发业务长度 = 30)						
业务量 ρ	0.1	0.3	0.5	0.7	0.9	0.999
正常吞吐量	0.957	0.877	0.8066	0.7418	0.6834	0.6566

4 结束语

本文用排队理论对输入队列 ATM 交换机的几种调度策略作了定性分析进行研究,并给出相应的一些改进的解析式.通过理论分析和仿真模拟表明这些解析式和实际吻合,为选择不同的调度策略提供了参考依据,也对改善吞吐量、保证 QoS 和公平服务等具有现实指导意义.研究输入队列调度策略不仅是设计高速 ATM 交换机的需要,而且对未来高速路由器的设计也有借鉴作用.

参考文献:

[1] 黄立群,黄载禄. FQLP:ATM 网中一种新的实时业务调度算法[J]. 电子学报. 2000,28(4):20 - 23.

[2] 戴礼森,洪佩琳. 高速信元交换调度算法研究[J]. 电子学报. 2000,28(5):96 - 98.

[3] Karol M J,et al. Input versus output queueing on a space division packet switch [J]. IEEE Trans. on Commmun,1988,35(12):1347 - 1356.

[4] HLUCHJ M G, KAROL M J. Queueing in high-performance packet-switching [J]. IEEE J. Sel. Areas Commun,1998,6(9):1587 - 1597.

[5] LEE T T. A modular architecture for very large packet switches [J]. IEEE Trans. on Commun,1990,38(7):1097 - 1106.

作者简介:

刘宴兵 男. 1971 年生于四川省遂宁市. 讲师,取得 Cisco 网络技术教员资格,北京邮电大学硕士研究生. 主要研究方向为宽带网络性能分析,发表论文近 20 篇.

李秉智 男. 1946 年生于河北省唐山市. 教授,博士生导师,重庆邮电学院计算机系主任. 主要研究方向为宽带网络技术. 曾获国家科技成果进步二等奖等多项奖励.

幸云辉 女. 1938 年 5 月出生于湖南省长沙市. 教授,主要从事计算机网络及应用研究.

(上接第 1545 页)

[37] Lin W, Zheng R, Hou J. How to make assured services more assured [A]. Proceedings of ICNP [C], Toronto, Canada, Oct. 1999: 182 - 191.

[38] I Yeom, A Reddy. Imapct of marking strategy on aggregated flows in a differentiated services network [A]. IWQoS'99 [C], May 1999.

[39] M May, J Bolot, A Jean-Marie, C Diot. Simple performance models of differentiated services schemes for the internet [A]. Proc. INFOCOM'99, New York City, NY, March 1999: 1385 - 1394.

[40] S Sahu, P Nain, D Towsley, C Diot, V Firoiu. On achievable service differentiation with token bucket marking for TCP [R]. UMASS CMPSCI technical report, pp. 99 - 72.

[41] Yeom I, Reddy A. Modeling TCP behavior in a differentiated services network [J]. To appear in IEEE/ACM Transactions on Networking, Feb. 2001.

[42] R Gibbens, et al. An approach to service level agreements for IP networks with differentiated services [DB/OL], http://www.statslab.cam.ac.uk/~richard/research/papers/sla/, January 2000.

[43] Haitao Wu, Keping Long, Shiduan Cheng, Jian Ma. A direct congestion control scheme for non-responsive flow control in diff serv IP networks. Internet Draft <draft-wuht-diffserv-dccs-00.txt> [S]. IETF, Aug. 2000.

[44] Haitao Wu, Keping Long, Shiduan Cheng, Jian Ma. Direct congestion control scheme for non-responsive flow control in diff serv IP networks [P]. Invention Reports Applying for Patent, March, 2000.

[45] Haitao Wu, Keping Long, Shiduan Cheng, Jian Ma. A self-configuring TCP-friendly marker based on random token bucket [P]. Invention Reports Applying for Patent, March, 2000.

[46] Haitao Wu, Keping Long, Shiduan Cheng, Jian Ma. A general three color marker [P]. Invention Reports Applying for Patent, July, 2000.