

# 区分服务网络中直接拥塞反馈机制与动态流量调节

邬海涛<sup>1</sup>,程时端<sup>1</sup>,隆克平<sup>1</sup>,马 健<sup>2</sup>

(1. 北京邮电大学交换技术与通信网国家重点实验室,北京 100876;2. 诺基亚中国研发中心,和平里东街11号,北京 100013)

**摘 要:** IP服务质量是目前国际上的一个研究热点,IETF为此而定义了区分服务体系结构.然而,该体系中存在不公平问题.本文提出了一种直接拥塞反馈机制(DCCS),配合网络边缘的动态流量调节(DTC)算法,能够为该体系提供很好的公平性.该机制不仅能保证响应流与非响应流之间的公平性,而且能解决响应流(如TCP)自身由于分组长度,微流数目及环回时延(RTT)等因素而产生的差异.

**关键词:** 服务质量;区分服务;拥塞控制;公平性

**中图分类号:** TN915.04 **文献标识码:** A **文章编号:** 0372-2112(2003)04-0481-06

## Direct Congestion Control Scheme and Dynamic Traffic Conditioning in Differentiated Services IP Networks

WU Hai-tao<sup>1</sup>,CHENG Shi-duan<sup>1</sup>,LONG Ke-ping<sup>1</sup>,MA Jian<sup>2</sup>

(1. National Laboratory, Beijing Univ. of Posts & Telecommunications, Beijing 100876, China;2. Nokia China R&D Center, Beijing, 100013, China)

**Abstract:** IP QoS is a research focus. Recently IETF propose the Differentiated Services (DS) architecture to implement QoS in Internet. However, there exists unfairness in DS network. This paper proposes to use a Direct Congestion Control Scheme (DCCS) and Dynamic Traffic Conditioning (DTC) to achieve the fairness. It maintains the fairness between responsive and unresponsive aggregates as well as better allocations for responsive flows (TCP) with different packet length, micro-flow number, Round Trip Time, etc.

**Key words:** quality of service (QoS); differentiated service; congestion control; fairness

### 1 引言

近年来 Internet 飞速发展,新的应用层出不穷.而传统的 IP 网络只提供尽力而为 (Best Effort) 的业务,伴随着传输和处理能力的增强,以及多媒体业务的引入,人们越来越期望网络能提供某种程度的服务质量 (QoS) 保证.

IETF 为了实现 IP QoS 而定义的区分服务体系模型<sup>[1]</sup>. 这种体系可以为不同的业务需求设置不同的区分服务编码点 (DSCP), 核心路由器采用不同逐跳属性 (Per Hop Behavior) 转发方式处理数据包. 区分服务体系的 RFC 建议了相关的业务量调节以及资源管理等策略, 用于系统地解决不同业务在网络中的需求差别. 在区分服务网络中, 与 QoS 保证相关的一系列复杂机制被推到网络的边缘; 边缘路由器需要完成流量调节 (Traffic Conditioning)<sup>[1]</sup> 的功能, 其中包括计量, 整形和标记等机制, 而核心路由器则只需要完成相应的分组调度和丢弃策略. 目前 IETF 定义的 PHB 有加速转发 (EF), 确保转发 (AF)<sup>[2]</sup> 以及 BE. IETF 的 RFC<sup>[2]</sup> 中定义了 4 个 AF 类, 每个 AF 类包括三个丢弃优先级. 当网络发生拥塞时, 这些丢弃优先级标识了数据包在同一 AF 类中的相对重要性. 近期的研究表明, AF 不能很好的保证用户间的公平性<sup>[3~9]</sup>.

Floyd 等人提出了 TCP 友好的速率控制<sup>[10]</sup> (TFRC) 的概念, 目的是为了使用非 TCP 的应用能够以某个相对平滑的速率, 与使用 TCP 作为端到端拥塞控制的应用, 以友好的方式分享网络中的带宽. 然而, 对拥塞不作响应的实时流或恶意流的加入, 可能会导致响应流和非响应流间的不公平. 此外, 由于不同的分组长度, 不同的微流数目以及环回时延等因素的影响, TCP 自身也存在着不公平现象.

本文认为: 区分服务的网络必须能够保证用户间的公平性, 才能促进用户使用端到端的拥塞控制或 TCP 友好的速率控制, 网络资源利用率才能提高. 根据前人成果<sup>[11~12]</sup>, 以及作者提出的区分服务网络中 TCP 友好公平性的概念<sup>[13]</sup>, 本文提出了一种直接拥塞反馈机制<sup>[14]</sup>, 结合网络边缘的动态流量调节, 目的是为了在提高公平性的同时优化网络资源的利用率. 本文的创新之处在于, 反馈机制是将网络的状态反馈给边缘, 由边缘负责利用该信息完成动态流量调节; 这是符合区分服务的根本思想的, 核心网络没有微流状态记录, 是简化核心, 复杂边缘的思想. 大量的仿真结果证实了这种机制及其实现算法的可行性和有效性, 该机制是在区分服务网中保证 TCP 友好公平性的有效方式.

本文的其他部分安排如下: 第二节简要描述区分服务网

收稿日期: 2001-10-15; 修回日期: 2002-01-15

基金项目: 高等学校博士学科点专项科研基金 (No. 20010013003); 国家自然科学基金 (No. 69972008) 和 Nokia-BUPT 合作项目资助

络中 AF 的性能问题, 第三节为直接拥塞反馈机制; 第四节为动态流量调节的算法. 第五节为仿真结果及分析; 最后为结论.

## 2 区分服务网络中的 AF 的性能和公平性

对具有目标速率, 又称确保速率的 TCP 的性能分析是从区分服务的思想被提出时就开始的. 通过使用 RIO<sup>[3]</sup> (RED with In and Out) 算法, Clark<sup>[3]</sup> 给出了一个 AF 的网络结构. RIO 中对标为 In (符合流量描述) 和标为 Out (不符合流量描述) 的分组制定了两套不同的 RED 参数, 通过丢包概率的不同来保证一定的目标速率. 而 Ibanez<sup>[4]</sup> 则用令牌桶标记算法进一步仿真分析了 AF 和 BE 业务混合时的情况, 其结论是 AF 无法很好的保证带宽分配. Goyal<sup>[5]</sup> 的仿真结论是三色标记有助于区分拥塞敏感和拥塞不敏感的业务流, 但对 TCP 的性能改进不大, 并且瓶颈处目标速率的总和不能超过瓶颈带宽.

区分服务网络中 AF 类的公平性问题可以分为两类: 一是响应流和非响应流之间的公平性问题, 这是由于二者对拥塞信号的不同处理方式所致; 二是响应流之间的公平性问题, 如 TCP 由于不同的分组长度, 微流数目而导致的对资源占用能力的不同.

Sedigh<sup>[6]</sup> 将 TCP 视为响应流, 而 UDP 视为非响应流, 并对 TCP 和 UDP 的 In/Out 的分组标记 (Marking) 方式提出了 6 种映射方案, 但结果都不理想. Sedigh<sup>[7]</sup> 对影响 TCP 获得目标速率的因素归纳为: 1) 环回时延 (Round Trip Time); 2) 汇聚 (复合流) 中微流的数目; 3) 目标速率的大小; 4) 分组长度; 5) 非响应流的影响. 但并没有给出能够实现区分服务中确保带宽的具体方案. 文 [13] 对三色标记器 srTCM<sup>[8]</sup> 及 trTCM<sup>[9]</sup> 下的公平性进行了分析, 提出了响应流和非响应流的符合流量描述的部分应该同样保护, 而超出流量描述的部分应该以 TCP 友好的方式分享网络的剩余带宽的观点. 文 [6] 中提到将 TCP 和 UDP 映射为不同的 PHB 类的方法来保证其公平性, 但实际上这很难做到, 本文认为主要受以下几点原因的影响: 1) 不能简单的通过协议类型来区分一个流是响应的还是非响应的, UDP 对拥塞的响应取决于应用, 而恶意用户很可能改变 TCP 流的响应性, 因此, 本文认为判断一个响应流的标准应该是它对拥塞所做出的反应; 2) 由于网络中流量的动态变化 (IP 是无连接方式), 核心路由器缺少状态记录, 类间的调度很难得到保证; 3) 响应流之间由于不同分组长度, 微流数目等因素造成的不公平仍然没有解决.

综合以上分析, 区分服务网络中没有一种好的机制来为这两种公平性提供保证, 而资源分配的不公平显然不利于端到端流控与速率控制的提倡, 在区分服务这样缺少状态记录及资源预留的网络中, 很容易造成拥塞坍塌. 因此, 本文提出了直接拥塞控制机制和动态流量调节机制来保证公平性, 同时保证网络资源的高效利用.

## 3 直接拥塞反馈机制 (DCCS)

由于区分服务中没有 Inte-sev 中的资源预留, 边缘又无法获知网络中的流量信息, 因此边缘路由器在执行标记算法时存在着一定的盲目性. 文 [14] 中提出了一种直接拥塞反馈

机制让网络能保证不同业务流间的公平性. 该机制的根本思想是由网络中央的路由器将控制消息动态的反馈给网络的边缘设备, 网络边缘的路由器根据这些信息来执行动态的流量控制 (DTC). 该机制的原理如图 1 所示.

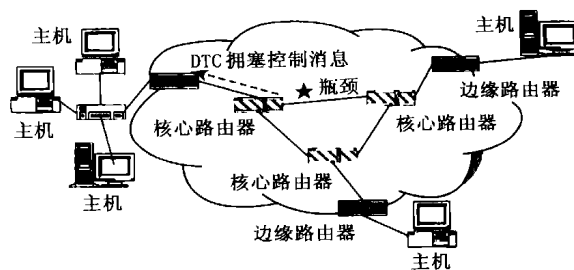


图 1 直接拥塞反馈机制 (DCCS) 原理图

在文 [14] 中已描述了该机制的框架, 这里只作简要说明. 核心路由器按一定的规则发送拥塞控制消息, 边缘路由器负责截获该消息并调整其流量调节 (TC) 的参数. 该消息可以用多种方式来标识, 如 ICMP 报文, IP 选项等, 本文选用的方法是用一个新的 DSCP 来标识. 当核心路由器要对某个数据分组发送控制消息时, 将该数据分组的源 IP 和目的 IP 分别作为控制消息的目的 IP 和源 IP, 用新的 DSCP 标识该控制消息, 控制信息的内容则写入该消息的净荷中, 包括当前的负荷, 各 PHB 的丢失率等信息. 按选路规则, 该消息将通过原数据分组进入该 DS 域的边缘路由器; 此时由边缘路由器负责截获该消息并调整 TC 参数.

在区分服务网络中的核心路由器必须能按业务类型对分组实施相应的转发机制, 对 AF 业务而言, 由于 AF 中包含 4 个类, 每个类包含 3 个优先级, 不同类是不相关的, 因此可采用调度策略, 而为了保证类内部的分组不失序, 须采用多门限 RED (mRED<sup>[4]</sup>) 或 RIO<sup>[3]</sup>. 本文中使用了三个丢弃优先级, 为 mRED. 为尽量减少核心路由器的复杂度, 本文的策略是核心路由器不保存任何有关连接的信息. 核心路由器只需将近期的丢包率反馈给网络边缘, 若丢包率很低, 则不进行反馈.

反馈的方法可以采用一般的时间区间的方法, 也可以是隔一定分组数目的方法. 本文中反馈的粒度是  $M:1$ , 及每转发  $M$  个数据分组, 就检查一次丢包率, 若有丢包则反馈, 然后将计数器清零. 这种反馈方式的选择是为了减小路由器的开销, 反馈时分组的选择是随机的. 从下一节的 DTC 可以看到, 反馈分组只是将网络的情况反馈给边缘路由器, 并没有决定边缘如何控制 TC, 由边缘的 DTC 判断 TCP 在该情况下的反应; 因此, 响应流并不会因为收到反馈分组而接受不公正的控制. 反馈的粒度是反向带宽开销和控制速度的折衷, 动态的调整反馈的时间区间或反馈粒度的方法有待进一步的研究.

这种方式给反向流量带来的影响是很小的. 根据对 Internet 流量的分析, 数据分组的长度集中分布在 552 字节和 576 字节处; 而反馈的分组长度很小, 假定为 50 字节, 若反馈粒度为 100, 则反向最大将增加不到 0.09% 的流量, 从该机制带来的性能提高与公平性来看, 这是值得的. 核心路由器负责简单操作, 而把有关微流的处理推到网络边缘是为了符合区分服

务体系简化核心,复杂边缘的根本思想。由于区分服务是正在研究中的体系,该体系中的核心和边缘路由器都需要加入新的功能来支持区分服务。DCCS 中所有的操作都只与区分服务将要引入变化的部分相关,如边缘及核心的操作,新 DSCP 等,因此,该机制是可行的。

#### 4 动态流量调节

区分服务网络的边缘路由器需执行流量调节(TC)的功能,包括标记、整形等策略。对 AF 类一般采用标记算法,主要是为了在保证目标速率的同时让超出流量描述的部分能竞争网络中的剩余资源以最大化利用率。本文认为,为了保证网络中资源分配的公平性,网络应该执行动态的 TC 功能。并且为了保证响应流和非响应流能公平的竞争网络的资源,TC 应该采用类似于 TFRC<sup>[10]</sup>的策略,但区分服务网络中 TCP 的建模不同于传统 BE 的情况。由于本文的思想是要将简化的建模结果应用于 TC 的动态调整,因此建模时有一些简化。

##### 4.1 区分服务中的 TCP 友好流量方程

由于 TCP 用慢启动和拥塞避免的方法来探测网络中的资源,当 TCP 检测到拥塞时,将其拥塞窗口减半。本文采用的是基于 TCP 方程的策略<sup>[10]</sup>。

根据 J. Padhye<sup>[10]</sup>等人的结果,TCP 在 BE 环境中稳态下的吞吐量可以用下面的方程来表示:

$$r = \frac{s}{T \sqrt{2p/3} + T_0(3 \sqrt{2p/8}) p(1+32p^2)} \quad (1)$$

其中: $r$ 给出了 TCP 在尽力而为(BE)方式下的发送速; $s$ 为分组长度; $T$ 为环回时延(RTT); $p$ 为稳态丢包率; $T_0$ 为 TCP 重传超时值。但由于区分服务中有目标速率的概念,上式不能直接套用,文[11]中给出了 TCP 在时间滑动窗口三色标记器(TSWTCM<sup>[3]</sup>)下 TCP 的建模结果,而文[12]对令牌桶两色标记的 TCP 进行了建模。由于本文中的 DTC 是基于双速率三色标记器(trTCM<sup>[9]</sup>)的,因此需对 trTCM 三色令牌标记器的 TCP 结果建模。按照 IETF 对 AF 类 PHB 的定义,一个 AF PHB 组中包含三个丢弃优先级,可以用绿、黄、红三种颜色来标记,拥塞时首先丢弃红包,然后是黄包和绿包。

由于 TCP 本身也存在很多变种,如 Reno, SACK 等,而這些变种获得的吞吐量是有差别的。本文中的 TCP 建模的目的是为了能让不同参数的流以 TCP 友好的方式分享网络的带宽,而不是对 TCP 动态变化的精确建模。因此,本文假定 TCP 处于理想的更新过程。

首先不考虑超时。在单速率的情况,只有一个令牌桶,深度设为  $B$ ,TCP 的目标速率为  $A$ ,设  $W_a = A \times T$ 。假定 TCP 处于稳定的更新过程,窗口增长到  $2W$  时出现分组丢失,减半为  $W$ 。当网络欠载时,TCP 获得的吞吐量大于其目标速率,根据  $W_a$  与  $W$  的关系可分为两种情况: $W > W_a$  和  $W < W_a$ 。区分两种情况,可以得到<sup>[12,14]</sup>

$$W = \begin{cases} (W_a + \sqrt{A^2 + 6s^2/p_2})/3, \\ A \leq (\sqrt{2(B \times s + s^2/p_2)} + 2 \sqrt{2B \times s})/T \\ (W_a + \sqrt{2B \times s + 2s^2/p_2})/2, \\ A > (\sqrt{2(B \times s + s^2/p_2)} + 2 \sqrt{2B \times s})/T \end{cases} \quad (2)$$

$$r = \begin{cases} (A + \sqrt{A^2 + 6s^2/p_2})/2, \\ A \leq (\sqrt{2(B \times s + s^2/p_2)} + 2 \sqrt{2B \times s})/T \\ (3A/4 + (3 \sqrt{B \times s + s^2/p_2})/(2\sqrt{2}T), \\ A > (\sqrt{2(B \times s + s^2/p_2)} + 2 \sqrt{2B \times s})/T \end{cases} \quad (3)$$

其中  $p_1$  和  $p_2$  分别为绿包和黄包的丢包率。而超载时的情况较简单,为<sup>[12]</sup>

$$W = \min(2W_a/3, (W_a + \sqrt{2B \times s})/2, s \sqrt{2/3p_1}) \quad (4)$$

$$r = \min(A, 3(A + \sqrt{2B \times s}/T)/4, s \times \sqrt{3/2p_1}/T) \quad (5)$$

引入超时的因素,设分组丢失后 TCP 进入超时的概率为  $Q$ ,超时的平均时间为<sup>[10]</sup>

$$Z^{T_0}(p) = T_0 \frac{1+p+2p^2+4p^3+8p^4+16p^5+32p^6}{1-p} \quad (6)$$

其中  $p$  在网络欠载时取  $p_2$ ,超载时取  $p_1$ 。根据文[11],超时概率的计算采用近似方法,设

$$q(W, p) = \frac{1 - (1-p)^W - Wp(1-p)^{W-1} - \frac{W(W-1)}{2} p^2(1-p)^{W-2}}{1 - (1-p)^W} \quad (7)$$

则超时概率为

$$Q = \begin{cases} q(2W - W_a)/s, p_2 & \text{网络欠载} \\ q(2W/s, p_1) & \text{网络超载} \end{cases} \quad (8)$$

引入超时因素后吞吐量又可表示为

$$r = 3W^2/2(TW + QZ^{T_0}) \quad (9)$$

根据式(2)、(4)、(6)、(8)、(9),区分服务中令牌桶标记算法下 TCP 的吞吐量可以表示<sup>[14]</sup>为

$$r = f(A, s, T, B, p_1, p_2) \quad (10)$$

双速率三色标记器(trTCM)中有两个令牌桶,分别称为桶 C 和桶 P,其桶深分别为 CBS 和 PBS,目标速率分别为 CIR 和 PIR。假定红包的丢失率为  $p_3$ ,则

$$r = \begin{cases} f(\text{PIR}, s, T, \text{PBS}, p_2, p_3) & p_1 = p_2 = 0 \\ \min[f(\text{CIR}, s, T, \text{CBS}, p_1, p_2), \text{PIR}] & \text{else} \end{cases} \quad (11)$$

用式(11)得到的速率为区分服务网络中 trTCM 标记下的 TCP 友好速率。其中  $p_1$ ,  $p_2$  和  $p_3$  是从网络核心路由器反馈的拥塞控制信息中得到的,CIR 为用户的目标速率,而桶深 CBS 和 PBS 则需根据用户流量突发度及其对网络的影响适当选取。根据网络中分组长度的分布, $s$  取其平均值, $T$  则取网络中环回时延的平均值,从后面的仿真结果将看到,这种设置不受某个连接参数的影响,主要是为了消除不同 TCP 流由于不同参数而导致的竞争能力的差异。作为 TCP 友好速率的  $r$  将用于 PIR 的设置,以达到不影响 TCP 而控制非响应流使其对 TCP 友好的目的。

##### 4.2 边缘路由器的动态流量调节

由于区分服务的思想是让超出业务量描述的流量竞争网络的剩余资源,以最大化网络的利用率;另外,黄包和红包的丢失对 TCP 性能的影响与绿包是相同的。因此,对黄包和红包的处理更应该谨慎。通过 TC 中的计量功能可计算流的当前速率,将式(11)计算的结果与之相比,可以判定流的响应性。本文采取平滑的方法对流量进行调整,算法如下:

采用一次滤波来计算某个汇聚的平均速率,设当前测得速率为  $S_{curr}$ ,则平滑后速率为

$$S_{rate} = \alpha S_{curr} + (1 - \alpha) S_{rate} \quad (12)$$

这样平滑的原因是为了尽量消除 TCP 中突发的影响,缺省的取值为 0.125.这样,DTC 的功能可以用下面的伪码表示(这里采用了调整 trTCM<sup>[9]</sup>中峰值信息速率 PIR 的方法):

1) 若收到拥塞控制消息,则按式(11)计算友好速率  $S_{fri}$ ;

2) 若  $(S_{fri} > S_{rate})$  则  $PIR = \text{Max}(S_{fri}, PIR)$

否则  $PIR = \text{Min}(S_{fri}, PIR)$

3)  $PIR = \text{Max}(PIR, CIR)$

实际仿真中计算 TCP 友好速率  $S_{fri}$  时没有考虑超时的影响,主要原因如下:1) TCP 友好速率的计算是为了设置令牌桶参数,因此需要的是 TCP 流量可能的最大值,而不是精确的建模结果;2) 由于目标速率的引入,区分服务中不同于 BE 的情况, TCP 能维持一定的速率,超时概率较低;3) 由于 TCP 自身流量的突发性,短期速率即使平滑后仍有较大波动,而不考虑超时可减弱 TCP 突发性造成的影响;4) 引入超时后公式过于复杂,不利于网络边缘设备的实时计算.

## 5 仿真结果及分析

### 5.1 响应流和非响应流间的公平性

本文采用仿真方法来验证 DCCS 与 DTC 的性能,仿真工具为目前 TCP 性能分析中较知名的 NS-2.1b6<sup>[15]</sup>,本文按 IETF 的与区分服务相关的 RFC 对 NS 仿真软件增加了所有有关区分服务的实现.需要再次强调的一点是不能简单的通过协议类型来区分一个流是响应的还是非响应的,如执行 TFRC<sup>[10]</sup> 算法的 UDP 流仍然是对 TCP 友好的.本文中只是采用 UDP 来产生固定比特率(CBR)的业务量,但这决不是把 UDP 和非响应流在实际中等同起来.DTC 算法的执行是与协议类型无关的,也未利用任何与协议有关的信息.

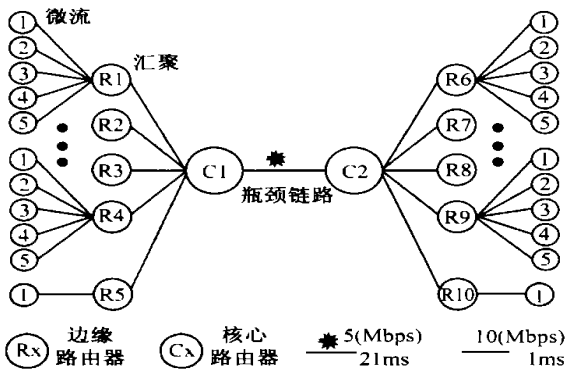


图2 仿真网络拓扑

本文仿真中的拓扑如图2所示.核心路由器  $C1$  与  $C2$  之间的链路是唯一的瓶颈,带宽为 5Mbps,传播时延为 21ms;其余链路的带宽均为 10Mbps,时延 1ms,因此环回时延为 50ms.图中有 4 个响应流汇聚,本文用 TCP 产生响应流,每个 TCP 汇聚的目标带宽为  $X$ Mbps,应用为 FTP 并且有足够需求.非响应流为 CBR 方式,用 UDP 产生,仿真中有 1 个 CBR 流,其目标速率与 TCP 的保持相同.

通过改变目标速率的方法可改变瓶颈处的负荷,这里负荷的定义为通过瓶颈链路流量的目标速率的总和与瓶颈链路速率的比值.本文中一个汇聚的目标速率为  $X$ Mbps,有 5 个汇聚(4 个 TCP,1 个 UDP),因此负荷为  $X \times 100\%$ .本文仿真中所有令牌桶深度均设为  $100\text{ms} \times X\text{Mbps}$ .作者对桶深采取了多种设置,得到的结果类似(桶深过小时例外).仿真中所有 IP 分组的长度均设为 500bytes,路由器中队列长度为 100 个分组,三色 RED 的参数设置见表 1.

表 1 三色 RED 门限参数

	Minth	Maxth	Maxp
绿	30	60	0.02
黄	15	30	0.1
红	5	15	0.1

我们对 TCP 的版本为 Reno 和 SACK,以及 TCP 汇聚中含有不同数目流的情况都进行了仿真,DCCS 算法取得的性能都明显优于其他没有反馈的算法.本文采用 RIO,srTCM 和 trTCM(未使用反馈和 DTC)的结果来与 DCCS 的结果进行比较.限于篇幅,我们只给出每个 TCP 汇聚含有 2 个 SACK 连接的结果(见表 2,UDP 发送速率为 2Mbps);由于丢包对 Reno 性能的影响远大于 SACK,因此,Reno 的仿真结果中 DCCS 对公平性的提高更为明显.

表 2 中用有效吞吐量来比较各种算法的性能,有效吞吐量(goodput)是指能正确到达目的端且不包括重传的流量.由于 TCP 与 UDP 的目标速率保持相等,因此,公平份额为 1Mbps.从表 2 的结果可以看出,DCCS 能极大的提高公平性,TCP 的性能可以得到很大程度上的保证.在网络欠载(负荷 < 100%)的情况下,公平性提高是极为明显的,TCP 离公平份额的偏差很小;而 UDP 由于发送速率是公平份额的 2 倍,因此丢包率接近 50%.这样,可以有效的保证公平性并能防止恶意流影响其他用户的性能.DCCS 下有效吞吐量总和也得到提高,因此,带宽利用率也被提高.另外,我们对 UDP 发送速率为 1Mbps 的情况也进行了仿真,结果是公平性很好且 UDP 丢包率极低.这样,能促使用户以公平的方式最大化利用网络的资源.

公平性得到保证的原因是由于对 trTCM 中峰值信息速率 PIR 的动态调整.trTCM 用双令牌桶来标记分组的颜色,两个令牌桶的速率分别为 CIR(Committed Information Rate)和 PIR(Peak Information Rate),PIR 必须大于等于 CIR.当发送速率小于等于 CIR 时,标记颜色为绿色;超出 PIR 时,标记颜色为红;二者之间为黄;桶深可控制标记的突发性.CIR 可设为目标速率,这样符合流量的部分被标为高优先级的绿色;但 RFC 中对 PIR 的设置没有建议,而 PIR 的正确设置正是公平性保证的关键.若 PIR 值比理想值过大,则本应标为红色的分组被标为黄色,过多的黄色分组使得黄包的丢失率上升,对 TCP 而言,会使得吞吐量降低,而非响应流会占用 TCP 损失的流量,使得公平性劣化.若 PIR 值比理想值过小,由于超出 PIR 的分组被标为红,而过小的 PIR 会使得应标为黄色的分组被标为红色,使红包的丢失率很大,TCP 的流量会受 PIR 的控制,使得公平性劣化.因此,PIR 过大不能很好的保护响应流,PIR 过

小会限制响应流的吞吐量,理想的情况是 PIR 随着网络中流量的变化而变化.但由于路由,流量等因素的综和影响,在目前区分服务网络中这是无法做到的,这也是 DCCS 要解决的根本问题.根据目前网络的状态计算 TCP 友好速率,结合连接的当前流量动态调整 PIR,是 DTC 保证公平性的根本原因,对于不同参数的 TCP 间的公平性可得到保证的原因分析类似.

表 2 TCP 汇聚中包含 2 个 SACK 微流的结果

		算法	TCP 平均 吞吐量	UDP 平均 吞吐量	总吞 吐量
目标 速率 x 及 负荷	0.4 Mbps (40%)	RIO	761.6k	1883.5k	4929.8k
		srTCM	740.1k	1883.1k	4843.5k
		trTCM	757.3k	1861.7k	4890.8k
		DCCS	957.2k	1163.3k	4992.3k
	0.7 Mbps (70%)	RIO	802.8k	1785.3k	4996.3k
		srTCM	810.8k	1734.0k	4977.1k
		trTCM	859.9k	1547.8k	4987.3k
		DCCS	954.4k	1147.2k	4964.7k
	1.0 Mbps (100%)	RIO	939.2k	1236.3k	4993.3k
		srTCM	987.2k	1040.3k	4989.2k
		trTCM	975.0k	1089.6k	4989.4k
		DCCS	979.7k	1077.6k	4996.3k
1.3 Mbps (130%)	RIO	925.1k	1299.3k	4999.8k	
	srTCM	926.8k	1277.9k	4985.2k	
	trTCM	921.6k	1299.7k	4986.1k	
	DCCS	924.0k	1279.9k	4976.0k	

表 3 不同分组长度下的仿真结果

		Reno		SACK		
		算法	平均 吞吐量	方差	平均 吞吐量	方差
目标 速率 x 及 负荷	0.4 Mbps (40%)	RIO	777.0k	173.1k	845.3k	164.6k
		srTCM	769.3k	173.1k	885.0k	224.0k
		trTCM	785.0k	152.9k	882.5k	168.3k
		DCCS	852.0k	29.9k	933.4k	86.1k
	0.7 Mbps (70%)	RIO	874.6k	98.3k	924.9k	70.3k
		srTCM	875.0k	113.8k	915.9k	88.6k
		trTCM	885.9k	81.5k	931.2k	39.6k
		DCCS	911.2k	62.7k	974.7k	36.7k
	1.0 Mbps (100%)	RIO	938.9k	56.3k	993.2k	42.9k
		srTCM	933.3k	57.0k	986.9k	20.1k
		trTCM	953.7k	49.1k	992.8k	27.9k
		DCCS	944.4k	58.7k	986.0k	25.8k
	1.3 Mbps (130%)	RIO	978.1k	127.2k	989.2k	227.0k
		srTCM	973.9k	90.5k	987.9k	206.0k
		trTCM	977.8k	94.4k	985.6k	214.0k
		DCCS	964.7k	79.3k	989.3k	202.4k

5.2 不同分组长度 TCP 的公平性

对文[7]中提到的影响 TCP 间公平性能的因素,我们都进行了仿真.这里仿真拓扑与图 2 类似,但不使用 UDP,瓶颈带宽改为 10Mbps,有 10 个 TCP 汇聚,各个 TCP 的分组长度从 200 字节到 1100 字节.通过比较有效吞吐量的方差的方法,可以看到 DCCS 对公平性的改善(表 3).不论是使用 Reno 还是 SACK,DCCS 在提高有效吞吐量的均值的同时,降低了方差,

且负荷越低,效果越好.因此,公平性和资源利用率都得到了很大的提高.

5.3 不同微流数目 TCP 的公平性

微流数目对公平性的影响是很大的.如果用户使用的连接越多,得到的带宽就越多,就会促使用户大量的增大连接数目;在区分服务的网络中,因为已经有目标速率的概念,过多的连接只是浪费资源,而且用户间的公平性也得不到保证.仿真中有 10 个 TCP 汇聚,每个汇聚包含的连接数目从 1 到 10 不等.

从表 4 的结果可以看出,DCCS 对资源利用率及公平性的改善都是很大的,且网络负荷越低,效果越好.由于区分服务网络为保证用户的 QoS,网络必须保持一定的剩余资源,而 DCCS 对于此时的公平性和利用率都有大的改进.对不同 RTT 等情况的仿真结果都较类似,限于篇幅,不一一列举.

从以上结果可知,由于 DCCS 将网络中心的信息反馈到边缘,边缘路由器是根据从核心路由器收到的拥塞控制消息来对流入的汇聚执行流量控制(TC)功能,因此能够根据某个流对拥塞的反应来进行相应的处理.这种机制能克服区分服务网络中中心和边缘缺少交互,流量调节不考虑网络中瓶颈的缺点,作者进行了不同拓扑,不同参数的大量仿真,结果表明该机制对公平性和网络资源利用率有很大的改善.该机制能对非响应流进行很好的控制,要强调的一点是这时的 DTC

表 4 不同微流数目下的仿真结果

		算法	平均 吞吐量	方差	平均 吞吐量	方差
目标 速率 x 及 负荷	0.4 Mbps (40%)	RIO	892.4k	276.8k	926.4k	237.4k
		srTCM	881.0k	281.4k	931.6k	262.2k
		trTCM	890.2k	282.1k	922.0k	233.7k
		DCCS	911.2k	142.0k	933.4k	86.9k
	0.7 Mbps (70%)	RIO	943.6k	202.2k	958.5k	106.7k
		srTCM	933.2k	166.5k	949.5k	111.6k
		trTCM	943.6k	202.2k	958.5k	106.7k
		DCCS	946.0k	118.7k	971.0k	50.6k
	1.0 Mbps (100%)	RIO	976.0k	135.9k	990.9k	69.6k
		srTCM	956.1k	132.7k	986.3k	24.8k
		trTCM	976.0k	135.9k	990.9k	69.6k
		DCCS	957.9k	97.4k	988.2k	26.0k
1.3 Mbps (130%)	RIO	984.9k	251.7k	992.1k	292.5k	
	srTCM	986.5k	263.3k	993.1k	307.1k	
	trTCM	984.9k	251.7k	992.1k	292.5k	
	DCCS	983.2k	251.6k	993.4k	277.1k	

是根据反馈的信息进行的,而不是根据不同的协议类型来进行的,因此有很强的适应性.并且,反馈消息是在区分服务的路由器间传递和处理的,所有区分服务网络以外的主机,路由器中协议都不受影响.这样,公平性将由网络来保证.

6 结束语

本文的贡献主要如下:1)基于区分服务简化核心,复杂边缘的思想提出了 DCCS 直接拥塞反馈机制,将网络中的拥塞信息及时的反馈到网络的边缘;2)对区分服务体系中三色令

牌桶 trTCM 下 TCP 性能的建模,将简化的建模结果用于网络边缘的流量调节功能;3)提出了一种 DTC 动态流量调节算法与 DCCS 相配合,这样做的目的是使得边缘的动态控制对 TCP 流没有影响,但对非响应流的控制能使其对 TCP 友好;4)用大量的仿真验证了 DCCS 与 DTC 可以对区分服务体系的响应流与非响应流之间的公平性,不同分组长度,微流数目,RTT 等条件下的响应流(TCP)之间的公平性都有很好的保证,并且,提高了网络的资源利用率。

#### 参考文献:

- [ 1 ] S Blake ,D Black ,M Carlson. An architecture for differentiated services [S]. IETF RFC 2475 ,Dec. 1998.
- [ 2 ] J Heinanen ,F Baker. Assured Forwarding PHB Group [S]. IETF RFC 2597 ,June 1999.
- [ 3 ] D Clark ,W Fang. Explicit allocation of best effort packet delivery service [J]. IEEE Trans. on Networking ,1998 ,6(4) :362 - 373.
- [ 4 ] J Ibanez ,K Nichols. Preliminary Simulation Evaluation of an Assured Service [S]. Internet Draft ,Aug. 1998.
- [ 5 ] M Goyal ,A Durrezi ,et al. Performance Analysis of Assured Forwarding [S] Internet Draft ,Feb 2000.
- [ 6 ] N Seddigh ,B Nandy. Study of TCP and UDP interaction for the AF PHB ,Internet Draft [S]. Aug. 1999.
- [ 7 ] N Seddigh ,B Nandy. Bandwidth assurance issues for TCP flows in a differentiated services network [A]. IEEE Globecom [C]. Riode Janeiro ,Brazil ,1999.
- [ 8 ] J Heinanen. A Single Rate Three Color Marker [S]. IETF RFC 2697 ,September 1999.
- [ 9 ] J Heinanen ,R Guerin. A Two Rate Three Color Marker [S]. IETF RFC 2698 ,September 1999.
- [ 10 ] S Floyd ,M Handley ,J Padhye. Equation-based congestion control for unicast applications [A]. Proceedings of SIGCOMM [C]. 2000.
- [ 11 ] I Yeom A Narasimaha. Modeling TCP behavior in a differentiated services network [R]. TX77843 - 3128 ,1999.
- [ 12 ] S Sahu ,P Nain ,et al. On achievable service differentiation with token bucket marking for TCP [A]. ACM SIGMETRICS 2000 [C]. Aug. 2000.
- [ 13 ] Haitao Wu ,Keping Long ,Shiduan Cheng ,et al. Fairness of responsive and unresponsive aggregate for assured forwarding PHB in differentiated services IP networks [J]. Journal of Software ,2001 ,12(11) :1601 - 1607.
- [ 14 ] Haitao Wu ,Keping Long ,Shiduan Cheng ,Jian Ma. Direct congestion control scheme (DCCS) for differentiated services IP networks [A]. IEEE GLOBECOM 2001 [C]. Nov. 2001.
- [ 15 ] NS(Network Simulator) ,http://www.isi.edu/nsnam/ns

#### 作者简介:



**郭海涛** 男,1976年9月生于江西省南昌市,北京邮电大学交换技术与通信网国家重点实验室博士研究生,1998年毕业于北京邮电大学电信工程工程系,目前研究方向为宽带网络服务质量,TCP/IP改进,区分服务,流控和拥塞控制,及无线分组网络性能。xiaodan@bupt.edu.cn



**程时端** 女,1940年6月生于上海,北京邮电大学交换技术与通信网国家重点实验室教授、博导,1963年7月毕业于北京邮电大学有线系,曾两次到比利时贝尔公司作高级访问学者,是我国最早从事 ISDN 和 ATM 技术研究的科研人员之一,1992 - 1999 年任 863 专家和国家实验室主任,作为负责人主持过 10 余项重大科研课题,已发表学术论文百余篇,专著 2 部,译著 2 部。