

区分服务网络中的 TCP 友好公平体系¹

邬海涛 隆克平 程时端 张润彤*

(北京邮电大学交换技术与通信网国家实验室 北京 100876)

*(诺基亚中国研发中心 和平里东街 11 号 北京 100013)

摘 要 区分服务 (Differentiated Services) 是 IETF 为实现 IP 服务质量 (QoS) 而定义的一个体系结构。研究表明, 在该体系中存在不公平问题。该文将 TCP 友好 (TCP Friendly) 的概念引入到 DS 网络中, 并定义了 DS 网络中的 TCP 友好的公平性。仿真验证了目前 IETF 定义的流量调节 (Traffic Conditioning) 以及丢包策略等机制不能很好地实现 TCP 友好公平性, 因而提出了直接拥塞控制机制来实现这一公平性。

关键词 服务质量, 区分服务, TCP 友好, 公平性

中图分类号 TN919.2

1 引言

在传统的 IP 网络中, 网络只提供尽力而为 (Best Effort, BE) 的业务。伴随着传输和处理能力的增强, 以及多媒体业务流的引入, 人们越来越期望网络能提供某种程度的服务质量 (QoS)。区分服务 (DS)^[1-4] 正是 IETF 为了实现 IP 中的 QoS 而提出的一个体系结构。

区分服务的根本思想是将与 QoS 保证的一系列复杂机制推到网络的边缘路由器 (Edge Router) 来完成, 而核心路由器 (Core Router) 只负责按分组的区分服务编码点 (DSCP)^[1] 将其转发, 从而实现某种逐跳属性 (Per Hop Behavior, PHB)。边缘路由器需要完成流量调节 (TC)^[2] 的功能, 而核心路由器则需要完成分组调度和丢弃功能。目前 IETF 定义的 PHB 有加速转发 (Expedited Forwarding, EF)^[3], 确保转发 (Assured Forwarding, AF)^[4] 及 BE。EF 提供的是一种类似于虚拟专线的业务, 而 AF 提供的则是一种确保带宽的业务。目前的研究集中于 AF 中的 TCP 的性能及公平性问题^[5-8]。

TCP 友好 (TCP Friendly)^[9-12] 的概念由 Floyd 等人提出, 目的是为了推进基于 BE 服务方式的 Internet 中的端到端拥塞控制的使用。本文认为端到端拥塞控制不仅对 BE 的服务方式有重要的意义, 对 DS 中的网络资源利用率以及公平性问题, 也存在重要的意义。但是, 由于对网络所承载的业务假设不同, TCP 友好的概念不能简单地照搬到 DS 的网络中; 同时, DS 网络中存在不公平的问题, 因此, 本文提出了区分服务中的 TCP 友好的公平性, 目的是为了在提高公平性的同时最大化网络的利用率。

本文安排如下: 第 2 节简要描述 DS 网络中的不公平的问题, 第 3 节介绍 TCP 友好的思想; 第 4 节定义了 DS 网络中的 TCP 友好公平性, 结合仿真进行分析; 第 5 节提出直接拥塞控制机制来实现 TCP 友好的公平性; 最后第 6 节总结全文。

2 区分服务网络中的性能和公平性问题

对区分服务中的性能分析是从区分服务的概念提出时就开始的。在 RED (Random Early Detection)^[13] 的基础上, 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

¹ 2001-03-19 收到, 2001-10-10 定稿

高等学校博士学科点专项资助课题 (No.20010013003), 国家自然科学基金项目 (批准号: 69972008) 和诺基亚-北邮合作项目资助

了两套参数,通过不同的丢包率来保证目标速率(Target Rate)可得到 RIO(RED with In and Out)算法。Clark^[5]用 RIO 给出了一个 AF 的网络结构,而 Ibanez^[6]则用令牌桶作标记算法进一步分析 AF 和 BE 业务混合时的情况,结论是 AF 无法保证带宽分配。

区分服务网络中的不公平问题可以分为两类:一是响应流和非响应流间的公平性问题,一是响应流之间的公平性问题。Seddigh^[7]将 TCP 视为响应流,而 UDP(User Datagram Protocol)视为非响应流,并对 TCP 和 UDP 的 In/Out 的分组标记(Marking)的方式提出了 6 种映射方案,但结果都不理想。Seddigh 在文献[8]中对影响 TCP 流量获得确保带宽的因素归纳为如下 5 点:(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;(4) 分组长度;(5) 非响应流的影响,但该文并没有给出能够实现区分服务中确保带宽的具体方案。

文献[7]中提到将 TCP 和 UDP 映射为不同的 PHB 类的方法可能可以保证其公平性,但实际上这是很难做到的,本文认为主要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,判断一个响应流的标准应该是它对拥塞所做出的反应;(2) 网络中流量的动态变化,核心路由由器缺少状态记录,类间的调度很难得到保证。

3 TCP 友好的概念

TCP 用慢启动和拥塞避免的方法来探测网络中的资源,而随着实时应用的增加,对拥塞不作响应的用户有可能在 TCP 进行退避时强占网络中的资源。Floyd 在文献[12]中将这不基于 TCP 的应用带来的威胁归纳为两点:(1) 这些应用会导致 Internet 的拥塞坍塌;(2) 与 TCP 竞争资源时极大的不公平现象。Floyd 详细地探讨了这种缺乏端到端拥塞控制的应用所带来的问题,即使是在使用 WFQ(Weighted Fair Queueing)等调度机制的情况下,也可能导致资源的极大浪费和拥塞恶化。然而, TCP 的这种基于窗口的传输方式是显然不适用于某些实时流的应用的。因此, TCP 友好的拥塞控制方式^[11,12]被提出。而 TCP 友好的定义却不是很严格,只是要求应用使用某种适应性的拥塞控制机制,从较长时间尺度来看获得的带宽与 TCP 在同等条件下大致相同。这种要求是很模糊的概念,因为即便 TCP 本身也存在很多变种^[14],而这些变种在丢包率较大时获得的吞吐量是有很大差别的。

本文中的 TCP 友好的概念与 Floyd 提倡的 BE 服务中的 TCP 友好的概念既有相同点,又有不同点。在要有对某些拥塞有响应,与 TCP 吞吐量可比上的要求是相同的;但由于区分服务中有目标速率的概念,因此二者有所区别。

4 区分服务网络中的 TCP 友好公平性及仿真

4.1 区分服务网络中的 TCP 友好公平性的定义

区分服务的根本目的是为了对用户实现某种程度的 QoS 保证。AF 类 PHB 的业务被理解为能提供某个目标速率,而对网络剩余资源的分配却缺乏一致的理解。本文为引出 TCP 友好的公平性,提出以下根据:

(1) 在 AF 类 PHB 中有目标速率的概念,本文认为对业务流符合业务量描述的部分,不论其是否对拥塞做出响应,都应该得到保护;(2) 类似于区分服务这样的体系缺乏对端到端的资源预留,由于 IP 网络中流量的动态变化,某个瓶颈的拥塞是不可避免的情况;因此, TCP 友好的端到端拥塞控制的思想对其资源利用率乃至性能,都有较大意义,仍然应该提倡;(3) 网络必须有能力保证使用 TCP 友好的拥塞控制的业务流与使用非响应流的业务流间的公平性;否则,如果使用非响应流就意味着好的性能,那 TCP 友好的概念就无法被提倡。

基于以上原因, 根据区分服务中 AF 类目标速率的概念, 本文提出区分服务中 TCP 友好的公平性定义如下:

(1) 在网络欠载时, 流量符合业务量描述的部分都应该得到保护, 而不管它是否响应; (2) 在网络欠载时, 超出业务量描述的流量部分应该按 TCP 友好的方式竞争网络的剩余资源; 网络必须能保证这种 TCP 友好的公平性; (3) 在网络超载时, 应该尽量保证所有符合业务量描述的流获取资源; 此时 QoS 已无法完全保证, 在丢包率很大时响应流和非响应流公平竞争是不可能的, 因此, 网络必须能保证此时所有的业务流不超过其目标速率。

4.2 区分服务网络中的仿真分析

本文采用仿真方法来验证区分服务中的机制能否保证本文提出的 TCP 友好公平性。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(TCP Friendly Rate Control)^[10] 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。

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。本文采用的仿真工具为 NS-2.1b6, 与文献 [5,7] 中采用的时间滑动窗口标记器 (TSW Marker) 不同, 本文中的 TC 算法都是基于令牌桶的, 包括令牌桶标记器, 单速率三色标记器 (srTCM)^[15] 及双速率三色标记器 (trTCM)^[16], 本文按 IETF 的与区分服务相关的标准对仿真软件增加了所有有关实现。

本文仿真中的拓扑如图 1 所示。拓扑和参数配置与文献 [7] 类似, 是为了便于与其比较结果。核心路由器 C1 与路由器 E3 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。图中有 4 对信源和信宿, 分别用 S1~S4 和 D1~D4 来表示。本文用 TCP 产生 4 个响应流, 每个信源宿对间均有一个 TCP 会聚 (Aggregate), 每个 TCP 会聚的目标带宽为 XMbps, 应用为 FTP 并且有足够需求, 通过改变 TCP 的目标带宽的方法可改变瓶颈处的负荷 (目标速率总和 / 瓶颈速率)。非响应流为 CBR 方式, 用 UDP 产生, 仿真中有两个 UDP 流, 分别是 S1 到 D1 和 S3 到 D3, UDP 流的目标速率为 0.5Mbps。仿真中所有 IP 分组的长度均设为 1kbyte, 绿黄红三色 RED 的参数设置 (即 Minth/Maxth/Maxp) 分别为 30/60/0.02, 15/30/0.1, 5/15/0.1。

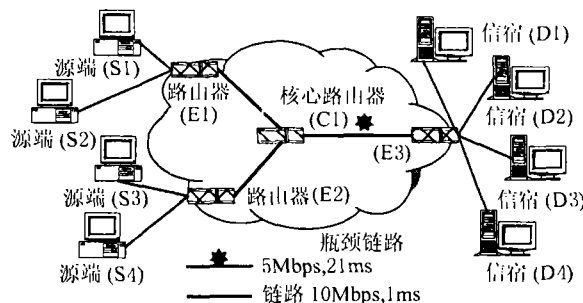
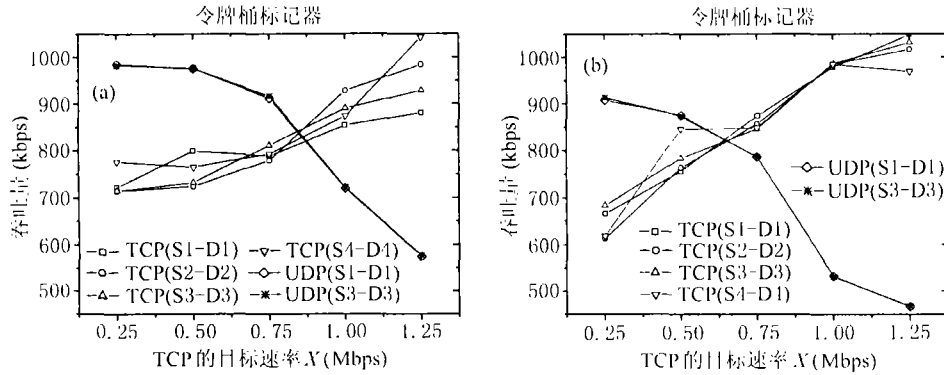


图 1 仿真网络拓扑

仿真结果按标记算法的不同可分为三类:

(1) 令牌桶标记器 (Token bucket based marker)

这也是 Ibanez 在文献 [6] 中采用的算法, 与 ATM 中通用速率控制算法是等效的。标记的颜色为 In 和 Out, 分别对应于 AF 类中的绿色和黄色。仿真结果如图 2 所示。



(a) TCP 汇聚中微流数目为 1

(b) TCP 汇聚中微流数目为 6

图 2 令牌桶标记器的仿真结果

由于只有两种颜色，这种情况和使用 RIO 时是相同的，这与文献 [7] 中的第一种映射方式也是相同的。图 2 中每个 TCP 汇聚分别包含 1 个和 6 个 SACK 微流，与 Reno 的结果类似。图中有两个点需特别注意，一是 X 为 0.5Mbps 时，此时所有会聚目标速率相同，应获得相同带宽，即 0.83Mbps；二是 X 为 1Mbps 时，目标速率的总和恰好等于瓶颈带宽，负荷为 100%，获得带宽应等于目标速率，即 TCP 应为 1Mbps，而 UDP 应为 0.5Mbps。按照本文对 TCP 友好的公平性定义，负荷小于 100% 时理想结果应为近似线性，并且 UDP 的流量不能超过这一值；而负荷大于 100% 后，应保证获得的吞吐量不超过其目标速率。与文献 [7] 不同，本文作者还对每个汇聚只包含一个微流的情况进行了仿真(图 3(a))，此时公平性极差，对拥塞不作相应的 UDP 获得了超过 TCP 友好的方式很多的吞吐量。图中 UDP 在负荷很低时出现抢占剩余带宽的趋势减小的原因是仿真中设定 UDP 发送速率为 1Mbps；若增大发送速率，则 UDP 占用的剩余带宽将增加，因此令牌桶的两色标记不能很好的保护 TCP 响应流。

本文对文献 [7] 中采用的 6 种标记映射方案都进行了相应的仿真，得到的结果与文献 [7] 类似。由于本文与文献 [7] 采用的标记器不同，后者使用的是时间滑动窗口标记器，因此可得出如下结论：对 TCP 和 UDP 的颜色映射方式决定了最后获得的公平性，而与具体的标记算法无关，这 6 种映射方式均不能保证 TCP 友好公平性。

(2) 单速率三色标记器 (srTCM)

srTCM 的仿真结果见图 3。图 3(a) 和图 3(b) 分别对应于微流数不同的设置。从图 3 的仿真结果可以看出：与用令牌桶标记器的结果不同，srTCM 对网络瓶颈负荷接近或超过 100% 时的公平性较好，即使是 TCP 中微流数目较少的情况。但是，当网络负荷较低时，结果和使用令牌桶算法接近，公平性得不到保障。

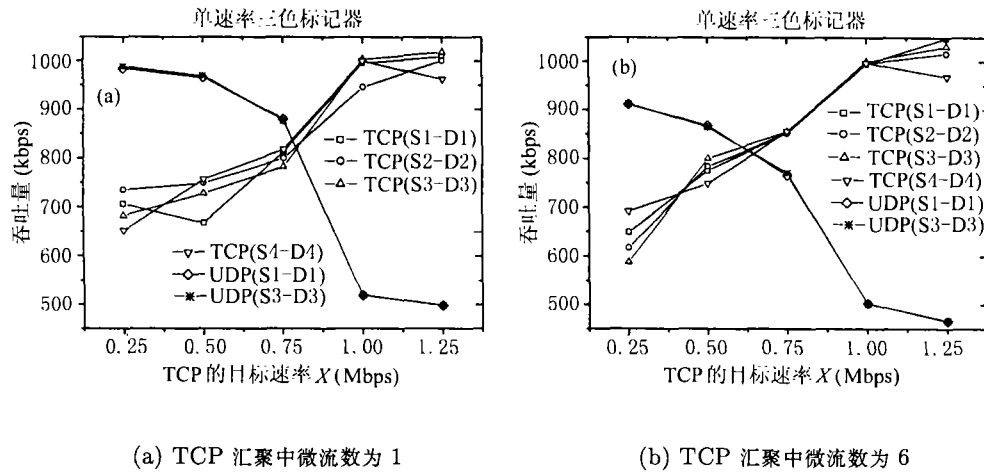


图3 srTCM 的仿真结果

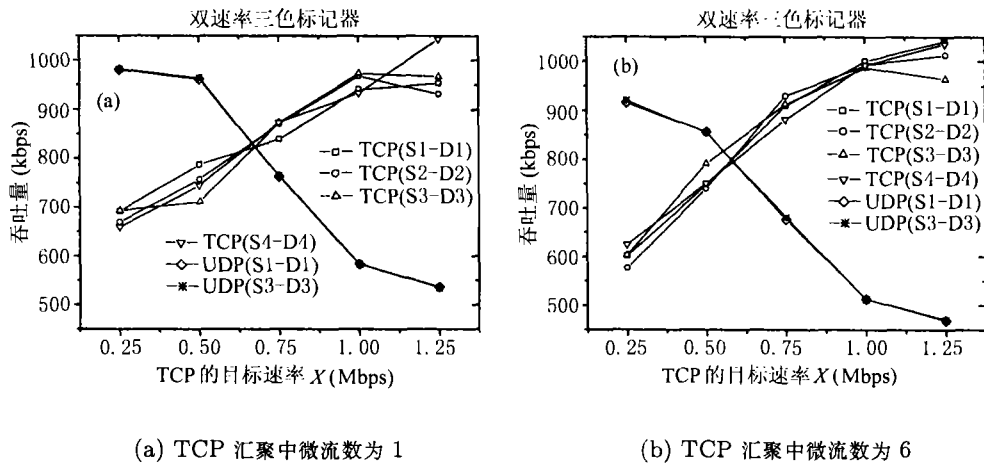


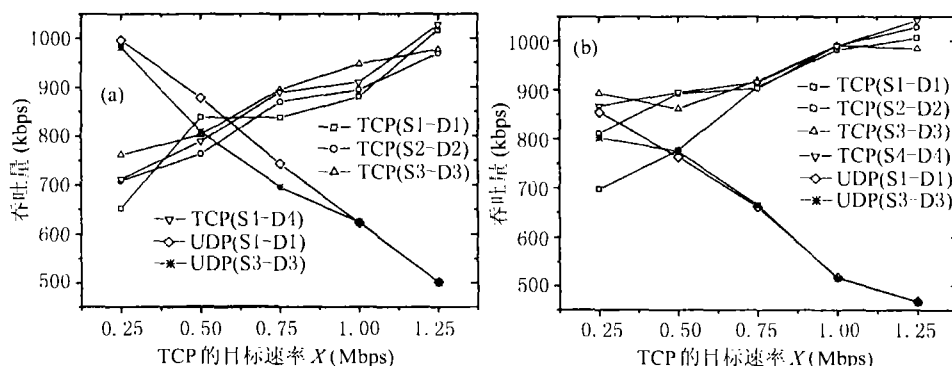
图4 trTCM 的仿真结果

(3) 双速率三色标记器 (trTCM)

当 TCP 一个汇聚中有 6 个微流时, 对 X 取值为 0.75~1.25Mbps 时 (对应瓶颈负荷为 80%~120%), 公平性较好, 而对超出该范围的情况, 公平性劣化; 当 TCP 一个汇聚中有 1 个微流时, 则范围更小, 只能在瓶颈负荷为 100%~120% 时公平性较好, 超出这一范围的公平性很差。

我们对 srTCM 和 trTCM 在不同参数, TCP 不同版本的情况下进行了大量仿真, 结果表明: 虽然使用三色标记器后得到的公平性要好于只有两色的情况, 但是当网络负荷稍低的情况下, 公平性都不理想, 尤其是 TCP 汇聚中微流数目较少时, 公平性迅速劣化。

出现上述情况的原因可以归纳为以下几点: (1) 区分服务网络中不区分微流, 由于某些流对资源的抢占会导致资源的不公平分配; (2) 在区分服务网络的边缘缺乏动态的流量控制功能; (3) 网络的边缘和中心没有交互。综合以上分析, 我们可以得出结论: 依靠区分服务现有的机制和算法, 是不可能获得很好的公平性的, 而本文定义的 TCP 友好的公平性网络更无法保证。因此, 本文提出一种适用于区分服务的机制来保证 TCP 友好的公平性。



(a) TCP 汇聚中微流数目为 1

(b) TCP 汇聚中微流数目为 6

图 5 直接拥塞控制机制仿真结果

5 直接拥塞控制机制 (DCCS)

本文作者提出了一种直接拥塞控制机制^[17,18]使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信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)。

该机制涉及的算法主要有反馈策略的算法和边缘路由器执行 DTC 的算法。由于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 TC 功能,因此能够根据某个流对拥塞的响应来进行相应的处理。这种机制能克服以上提到的区分服务网络中的缺点,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。从图 5 可以看出,对 TCP 中微流数目较多和较少的情况,该机制都能对 UDP 进行很好的控制,要强调的一点是这时的 TC 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,与简单的按协议类型将 TCP 和 UDP 进行不同的映射方式完全不同,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,路由器中协议都不用修改。这样, TCP 友好的公平性将由网络来保证,对端到端拥塞控制的提倡有很大意义。

6 结束语

目前 Internet 中的 QoS 问题是研究的热点。对于区分服务这样网络中没有微流状态记录及资源预留的体系,端到端的拥塞控制及 TCP 友好的观点具有重大意义。本文的贡献主要如下: (1) 结合区分服务的目标速率及 TCP 友好的概念,提出了区分服务体系下的 TCP 友好的公平性; (2) 用大量的仿真验证了区分服务不能支持这一公平性; (3) 提出了直接拥塞控制机制,可以对该体系的公平性有很好的支持,并用仿真验证。

参 考 文 献

- [1] K. Nichols, S. Blake, Definition of the differentiated services field (DS Field) in the IPv4 and IPv6 headers, RFC 2474, Dec. 1998.
- [2] S. Blake, D. Black, An architecture for differentiated services, IETF RFC 2475, Dec. 1998.
- [3] V. Jacobson, K. Nichols, K. Poduri, Expedited forwarding PHB, IETF RFC 2598, June 1999.
- [4] J. Heinanen, F. Baker, Assured forwarding PHB group, IETF RFC 2597, June 1999.

- [5] D. Clark D, W. Fang, Explicit allocation of best effort packet delivery service, *IEEE/ACM Trans. on Networking*, 1998, 6(4), 362-373.
- [6] J. Ibanez, K. Nichols, Preliminary simulation evaluation of an assured service, *Internet Draft*, Aug 1998.
- [7] N. Seddigh, B. Nandy, Study of TCP and UDP interaction for the AF PHB, *Internet Draft*, Aug 1999.
- [8] N. Seddigh, B. Nandy, P. Pieda, Bandwidth assurance issues for TCP flows in a differentiated services network, *IEEE Globecom*, Rio de Janeiro, 1999, 1792-1798.
- [9] J. Padhye, V. Firoiu, D. Towsley, Modeling TCP throughput: A simple model and its empirical validation, *Proc. of SIGCOMM*, Vancouver, 1998, 282-293.
- [10] S. Folyd, M. Handley, J. Padhye, J. Widmer, Equation-Based Congestion Control for Unicast Applications, *Proc. of SIGCOMM*, Stockholm, 2000, 38-51.
- [11] The TCP-Friendly Web Page. URL <http://www.psc.edu/networking/tcp-friendly.html>.
- [12] S. Floyd, K. Fall, Promoting the use of end-to-end congestion control in the Internet, *IEEE/ACM Trans. on Networking*, 1999, 7(4), 458-472.
- [13] S. Floyd, V. Jacobson, Random early detection gateways for congestion avoidance, *IEEE/ACM Trans. on Networking*, 1993, 1(4), 397-413.
- [14] K. Fall, S. Floyd, Simulation-based comparisons of Tahoe, Reno, and SACK TCP. *Computer Communication Review*, 1996, 26(3), 5-21.
- [15] J. Heinanen, A single rate three color marker, *IETF RFC 2697*, Sept 1999.
- [16] J. Heinanen, R. Guerin, A two rate three color marker, *IETF RFC 2698*, Sept 1999.
- [17] 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*, Aug 2000.
- [18] Haitao Wu, Keping Long, Shiduan Cheng, Jian Ma, Direct congestion control scheme(DCCS) for differentiated services IP networks, *IEEE GLOBECOM*, San Antonio, 2001, vol.4, 2290-2294.

A TCP FRIENDLY FAIRNESS ARCHITECTURE FOR DS IP NETWORKS

Wu Haitao Long Keping Cheng Shiduan Zhang Runtong*

(Nat. Lab of Switching Tech. and Telecom. Networks, BUPT, Beijing 100876, China)

**(Nokia China R&D Center, Nokia House 1, No.11, Hepingli Dong Jie, Beijing 100013, China)*

Abstract Differentiated Services(DS) architecture has been proposed by IETF to implement QoS in IP networks. However, research shows that there exists unfairness in DS network. This paper introduces the TCP friendly concept into DS network and defines TCP friendly fairness in DS architecture. Elaborate simulations are used to verify that current mechanisms in DS cannot guarantee the TCP friendly fairness; therefore, direct congestion control scheme is proposed to achieve the TCP friendly fairness.

Key words QoS, Differentiated Service(DS), Transport Control Protocol(TCP) friendly

邬海涛: 男, 1976 年生, 博士生, 研究方向为宽带网理论.

隆克平: 男, 1968 年生, 副教授, 博士后, 研究方向为网络生存性及服务质量.

程时端: 女, 1940 年生, 教授, 博士生导师, 研究方向为宽带通信网技术.

张润彤: 男, 1963 年生, 博士后, 研究方向为 AI 及网络服务质量.