第 25 卷 第 3 期 电 子 与 信 息 学 报 Vol.25 No.3
2003 年 3 月 JOURNAL OF ELECTRONICS AND INFORMATION TECHNOLOGY Mar. 2003

Internet 路由器中拥塞控制机制研究的现状与展望 [1]

魏蛟龙 张 驰

(华中科技大学电子与信息工程系 武汉 430074)

摘 要 目前，Internet 路由器中拥塞控制机制研究的热点在于：在不破坏网络可扩展性的前提下，如何在路由器上提供拥塞控制机制。该文从信息与激励的角度出发，以部分流状态、核心路由器状态无关、拥塞计费等为 3 个基本思路，对该领域中已提出的主要研究方案进行了分类阐述和比较分析，总结了其最新研究进展，为下一步的研究提出了新的课题和设想。

关键词 网络拥塞控制，部分流状态，核心 (路由器) 状态无关，拥塞计费

中图号 TN919.3

1 引 言

当前 Internet 上的拥塞控制主要是基于用户端的 TCP 而实现的，网络上的路由器仅是简单地采用先进先出 (FIFO) 的调度算法和当缓冲区满时丢弃后到分组的丢尾 (tail drop) 策略。当网络发生拥塞时，终端主机检测到分组被路由器丢弃，将按 TCP 的 AIMD(Additive Increase Multiplicative Decrease) 算法降低分组发送速率。由于路由器不需要保持任何关于个别通信流的状态信息，因而是状态无关的 (Stateless)[1]。

但是，随着 Internet 从一个教育科研型网络转变成一个商业网络，这种完全基于终端主机 (host-based) 的拥塞控制机制正面临严峻的挑战：一是随着网络用户的增加和所承载业务的多元化，越来越多的业务流是非拥塞响应的，如 UDP(User Datagram Protocol)；二是这种机制不能支持一些对网络服务质量 (Quality of Service, QoS) 有特殊需要的新业务，而这正是 Internet 下一步发展的关键所在 [2]。

网络本身要参与到拥塞控制中去，已成为一个不可回避的研究方向，出现了越来越多的基于路由器 (router-based) 的拥塞控制机制，使网络具有在不同的业务流之间分配稀缺资源的能力。由于路由器对分组的控制主要体现在缓冲区管理和队列调度上，因而这些方法又可相应地分为两类，分别以 FRED(Flow Random Early Detection) 和 WFQ(Weighted Fair Queue) 为代表。但这些方法并没有在 Internet 中得到广泛的实施，关键原因有：(1) 它们都要求网络是状态相关的，即要求路由器保持每个流的状态信息。因而这些方法都破坏了 Internet 赖以成功的基本特性：可扩展性、鲁棒性。(2) 它们都使得路由器变得十分地复杂。(3) 它们要求整个网络的软件和硬件有较大的改变，这种改变所需的投资对于广域网是不可想象的。

因而需要在 Internet 路由器上实现一种新的拥塞控制机制，在不破坏网络可扩展性的前提下，提供流保护 (flow protection)、带宽公平分配，甚至支持区分业务 [3]。本文将以 3 种方法为主线，总结最近一年来在此领域的最新进展。在介绍每种方法时以其中最具潜力实现方案为主，并比较其它方案。最后，对这 3 种方法从信息与激励的角度进行总结。

2 部分流状态法

2.1 部分流状态法的基本思路与实现方案

部分流状态法 (partial per-flow state) 的基本思路是：路由器中缓存区的分组提供了关于每

[1] 2002-01-28 收到，2002-07-08 改回
湖北省自然科学基金资助课题 (No. 99J041 和 No. 2001ABB104)

个流发送速率的准确信息，问题在于我们应当如何处理这些信息，尽量减少路由器需要保存的流的状态。体现这一思路的实现方案有：

(1) 对 WFQ 和 FRED 的改进：在 M 个活动的长期流中，只有 N 个流在任意时间间隔 T 内有分组在路由器中排队，且 N 小于 M。对于任何给定时间间隔 T，调度器只需在 N 个队列间均匀分配链路带宽，就可在长时间内，成功地调度 M 个流。根据这一思想，文献 [4] 对 WFQ 进行改进，用能以线速产生和删除队列的高速队列管理器实现了路由器，只需要 64k 个队列，就能在几十万个 OC-12c 速率的流之间保持公平的链路共享。基于同样的理由，文献 [5] 针对 FRED 提出，仅保留那些目前有分组在 RED 管理的队列中的流的状态。

(2) RED-PD(RED with Preferential Dropping) 模型：在路由器中只保留非 TCP 友好流的状态信息 [6]。

(3) CHOKe(CHOose and Keep for responsive flows, CHOose and Kill for unresponsive flows) 模型：文献 [7] 则提出不用保留任何流的状态信息，当一个分组到达路由器时，从缓冲区中随机抽取一个分组与之配对。若这 2 个分组属于同一流，则 2 个分组均被丢弃。若 2 个分组不属于同一流，则所到分组按一定概率 p 丢弃，p 由当前队列长度决定 (类似 RED)。显然，非 TCP 友好流的分组被丢弃的概率高。仿真试验表明：当流的数目与缓冲区大小相比较大时，CHOKe 模型的性能不佳，而且对占据带宽较大的 CBR 流控制能力有限。

显然，在上述 3 种方案中，RED-PD 模型在保存的流状态与流保护能力之间取得了较好的平衡，因而在部分流状态法中具体讨论这一模型。

2.2 RED-PD 模型

由于目前 TCP 仍是 Internet 上的主导性传输协议 [8]，因而 Internet 上流保护以及拥塞控制的关键，就在于抑制那些在带宽竞争上比 TCP 流占优势的非 TCP 友好流 (TCP unfriendly flow)。据此，Floyd 和 Fall 首先提出用带选择性丢弃功能的 RED(RED-PD) 来提供更可靠的拥塞控制。其方案是：路由器鉴别出那些表现不良的流 (ill-behaved flow)，然后通过有选择地丢弃这些流的分组，直到这些流获得的实际带宽小于相同情况下 TCP 流所能获得的带宽。这种惩罚措施将给终端用户提供足够的激励去使用 TCP 或与 TCP 等效的拥塞控制机制。

非 TCP 友好流是相对于 TCP 流而言的，因而描述 TCP 流是设计鉴别算法的前提。

2.2.1 TCP 流稳态条件下发送速率的稳态数学模型　当 TCP 控制的数据流传输处于稳态 (steady-state) 时，假设 TCP 发送端的拥塞窗口 (congestion windows, cwnd) 最大值为 W，则 cwnd 在 W 和 $W/2$ 之间呈周期性的锯齿状变化 (如图 1 所示)。当一个分组被网络丢弃，发送端发现后 cwnd 减半为 $W/2$，并在随后的每一个往返时间 (Round Trip Time, RTT) 中 cwnd 加 1，直至 cwnd 恢复为 W，又发现一个分组丢失。设 RTT 为一常数 r，分组大小为 M，则在 cwnd 变化的一个周期内，共发送了 $\sum_{i=0}^{W/2}\left(\frac{W}{2}+i\right)=\frac{3}{8}W^2$ 个分组，从而分组丢失率 (packet drop rate) 为

$$p = 1/[(3/8)W^2] = 8/(3W^2) \qquad (1)$$

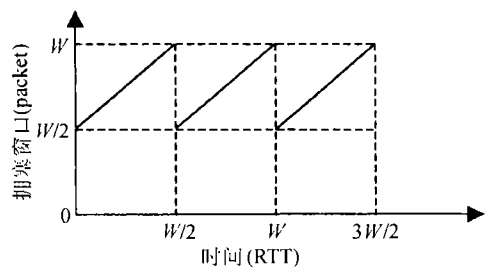

图 1　稳态下 TCP 发送端的拥塞窗口变化

又 cwnd 变化一个周期的时间为 $W/2$，因而平均发送速率 (带宽) 为

$$B = [(3/8)W^2M]/(Wr/2) = 3WM/(4r) \qquad (2)$$

又由 (1) 式得 $W=\sqrt{8/(3p)}$，代入 (2) 式得

$$B = (\sqrt{3}/\sqrt{2})[M/(r\sqrt{p})] = 1.22M/(r\sqrt{p}) \qquad (3)$$

上式给出了在网络稳态环境下 TCP 流占据的带宽 B 与分组丢失率 p 间的数量关系. 对于一个 TCP 流, 平均发生一次分组丢失的时间为

$$\mathrm{CL}(r,p) = M/(Bp) = r/\sqrt{1.5p} \tag{4}$$

为了更精确, 规定在 $k\times$ CL(r,p) 秒内, 被丢弃的分组数目多于 k 个的流为非 TCP 流, 或称为表现不良的流 (在相同情况下发送速率大于 TCP 流).

2.2.2 RED-PD 模型中路由器的功能 RED-PD 模型中路由器的功能框图如图 2 所示. 主要功能模块如下:

(1) RED : 由于 RED 随机丢弃分组, 而且防止了缓存器的溢出, 因而被 RED 丢弃的分组可以视为对到来的整个通信流的抽样. RED-PD 将 RED 丢弃分组作为鉴别非 TCP 友好流的依据还有一个好处: 由于被丢弃的分组不处于高速向前转发路径上, 因而在鉴别过程中并不需要高速存贮器保存丢弃记录. 另外, 根据到达分组和被丢弃分组数目可估算出分组丢弃率 p, 每到一个新分组更新一次 p, 采用加权指数平滑算法.

(2) 鉴别引擎: 保留 $k\times$ CL(r,p) 秒内被 RED 丢弃的分组, 若其中属于同一流的分组个数超过 k, 便对该流实施监控. (4) 式是从确定性模型中导出的, 而分组丢弃往往是随机的, 因而 k 越大, 鉴别越准确, 但判别时间越长.

(3) 监控器: 如果到来的分组属于鉴别引擎判定的非 TCP 友好流, 则将该分组转入前置过滤器; 否则, 直接转入 RED .

(4) 前置过滤器: 前置过滤器 (PF) 对鉴别引擎判别为非 TCP 友好流进行处罚. 它执行两个功能: 一是确定对非 TCP 友好流的丢弃概率, 使之占用带宽小于相同情况下 TCP 流带宽; 二是确定对流解除监控.

在 RED-PD 路由器中需保存的状态信息有:

(1) 鉴别引擎保存了 $k\times$ CL(r,p) 秒的分组丢弃信息; (2) PF 保存的非 TCP 友好流的状态信息. 由于 RED 以及对非 TCP 友好流的处罚可以降低 p, 要保存的状态信息很少.

2.2.3 RED-PD 模型存在的问题 (1) RED-PD 模型的有效性主要取决于鉴别算法. 文献 [9] 认为 (3) 式对 TCP 流的描述过于理想化, 在考虑了超时重传等因素后, 提出了一个更接近实际情况的 TCP 吞吐量公式.

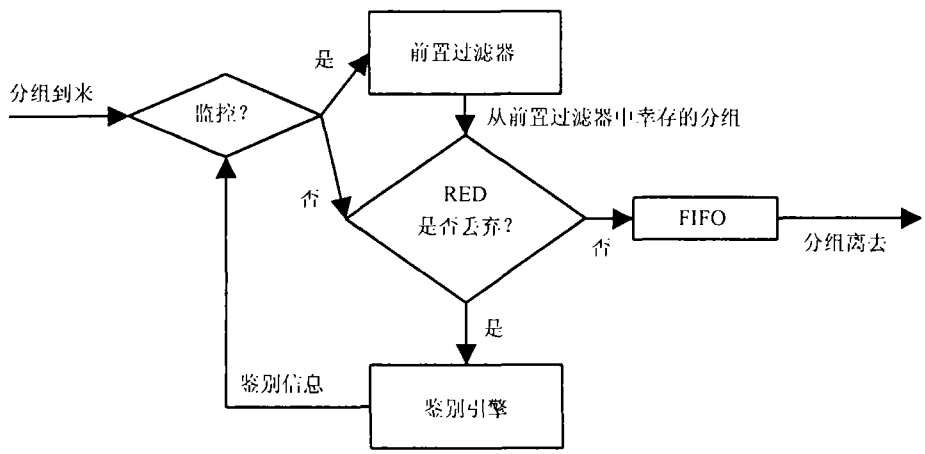

图 2 RED-PD 模型中路由器的功能框图

(2) 无论是最初的还是改进的非友好 TCP 流的鉴别公式中都涉及到往返时间 r . RTT 的准确估计是一项十分复杂的工作, 一般以单向传播时延的 2 倍近似之, 这使得鉴别算法过松, 部份非 TCP 流可能判别不出来. 文献 [10] 对 RTT 估计算法最新的研究进展作了一个综述.

(3) RED-PD 模型要求终端用户必须使用与 TCP 等效的流量控制机制. 考虑到 TCP 仍是传输协议的主流, 因而问题的关键在于能否给其它新兴业务提供 TCP 友好的流量控制机制. 文献 [11] 总结了为实时业务等新兴网络应用提供 TCP 友好的流量控制机制的研究进展.

3 状态无关法

3.1 状态无关法的基本思路与实现方案

Stoica 首先提出 "核心 (路由器) 状态无关" (Stateless CORE, SCORE) , 其基本思路是: 为了避免让核心路由器 (Core Router, CR) 保持每个流的状态信息, 可以在边缘路由器 (Edge Router, ER) 上把每个流的状态信息插入其分组的包头, 让分组携带这些信息通过网络的核心部分, 供 CR 进行基于流的操作而又不用保存这些流的信息 [12] .

文献 [13] 提出另一种实现机制: 彩虹公平队列 (Rainbow Fair Queueing, RFQ) . 在 ER 对每个流通过彩色标记 (color marked) 进行分层描述. 进入不同层的分组被标上不同的数字 (称彩色标签) . 一个流的分组所分布的层越多, 其发送速率越大. 在 CR , 当发生拥塞时首先丢弃高层分组, 这样发送速率大的流受到抑制. RFQ 很容易支持区分业务, 一个流可以通过与 ER 协商, 将其分组的层级标低, 从而分组被丢弃的概率减少, 业务流实际获得的带宽增加.

一般的, RFQ 可视为 SCORE 模型的一个特例. 因而下面仅讨论 SCORE 模型.

3.2 SCORE 模型

3.2.1 SCORE 模型中边缘路由器的功能　在 SCORE 模型中, ER 是状态相关的: 它要保留经过的每个通信流的状态信息, 由于一般 ER 比 CR 转发分组的速度要慢得多, 所要处理的通信流数目也少得多, 因而这并不影响 SCORE 网络的可扩展性. 当有新分组到达 ER 时, ER 首先将其分类, 归入其所属的流; 然后更新所保存的这个流的传输速率的估计值 $\hat{r}_i$(流 i 的状态信息) ; 并将 $\hat{r}_i$ 插入分组的包头内. ER 使用指数加权平均算法来更新 $\hat{r}_i$. 之所以由 ER 而不是用户端来估计流的传输速率并作标记, 是为了防止用户端为自身利益而有意错标.

3.2.2 SCORE 模型中核心路由器的功能　以 SCORE 模型实现公平队列 (Fair Queuing, FQ) 为例, 说明 CR 的功能. 首先考虑一个实施公平队列调度机制的路由器, 其输出链路速率为 C . 在某一时刻有 n 个流经过该路由器, 每个流的速率为 $\hat{r}_i, i \in [1,n]$. 定义公平速率 α , 满足 (5) 式

$$\sum_{i=1}^{n} \min(r_i, \alpha) = C \tag{5}$$

它表示在最大最小公平准则下, 每个流所能获得的最大带宽. 流 i 在 FQ 调度机制下, 获得的实际带宽为 $\min(r_i, \alpha)$.

SCORE 模型在 CR 上要达到同样的效果, 其策略是: 当 CR 收到一个分组, 依据该分组头部携带的 $\hat{r}_i$ 和 CR 保持的信息 α 计算

$$p = \min(1, \alpha/\hat{r}_i) \tag{6}$$

CR 以概率 p 转发该分组, 以概率 $1-p$ 丢弃之. 若 $p < 1$, 则 CR 还要将该分组包头中的 $\hat{r}_i$ 改为 α . 因而流 i 获得的带宽为 $\hat{r}_i \times p = \min(\hat{r}_i, \alpha)$, 与 FQ 等效.

现在的问题是如何在 CR 上保持 α . α 并不能由其定义 (6) 式直接得到，因为这需要每个流的状态信息 $\hat{r}_i$. 因而只能从经过 CR 的总流量的信息中估计 α . 每当一个新分组到来时，CR 就更新一次到来的数据总流量 $\hat{A}$ 和被转发 (未被丢弃) 的数据总流量 $\hat{F}$ ，采用的同样是指数加权平均算法，若 $\hat{A} \le C$ ，表明 CR 上无拥塞，由 α 的含义可估计： $\hat{\alpha} = \max_{1 \le i \le m}(r_i(t))$. 若 $\hat{A} \ge C$ ，表明 CR 上有拥塞，这表明原有的 α 估计值 $\hat{\alpha}_{\text{old}}$ 过大，规定新的估计值 $\hat{\alpha}_{\text{new}}$ 为

$$\hat{\alpha}_{\text{new}} = \hat{\alpha}_{\text{old}} C / \hat{F} \tag{7}$$

容易证明这是用线性函数近似 $F(\cdot)$ 后，方程 $F(\hat{\alpha}) = C$ 的解.

3.2.3 用 SCORE 模型支持多业务 文献 [14] 对 SCORE 模型进行简单地扩展，使其支持保证最小发送速率 (Minimum Guaranteed Rate, MGR) 的新业务. 假设第 i 个流的 MGR 为 $\text{MGR}(i)$ ，则第 i 个流应分得的带宽为

$$B(i) = \text{MGR}(i) + \left(C - \sum_{j=1}^{n} \text{MGR}(j) \right) \Big/ n \tag{8}$$

为实现这一理想分配方案，首先将 SCORE 模型中的 RED 扩展为具有 in/out 的 RED(RED with In/Out, RIO) . ER 根据流 i 的发送速率是否符合业务描述 $\text{MGR}(i)$ 对分组标记 in 或 out . 当一个 in 分组到达路由器时，只要缓冲区没有溢出，路由器都会接受之. 当一个 out 分组到达路由器，以概率 p 丢弃之. 令 m 为队列长度目标值与当前队列长度之比，则 p 是 m 的函数.

3.2.4 SCORE 模型存在的问题 SCORE 模型涉及到对 IP 包头的改造. 另外它需要路由器间的协同工作，而不像 RED-PD 中各个路由器可独自工作. 这使其逐步部署的能力受到约束. 由于 Internet 是分散管理的，因而任何一种方案都只可能是逐步部署的，但用户对方案性能评价又是全局性的. 而 RED-PD 即使是在 Internet 的一部份上部署，也能明显改善网络性能.

4 拥塞计费

4.1 拥塞计费法的基本思路与实现方案

目前 Internet 中的计费方式被称为平坦计费 (flat-rate pricing) ，与用户对网络资源的使用无关. 这种计费方式只能为 ISP 回收成本，而无法利用价格杠杆调节用户对网络资源的使用. 而且，随着 Internet 的商业化，这种与使用脱节的计费方式只会鼓励更多的用户滥用网络资源. 因而，越来越多的研究关注按照资源的实际社会价值即“影子价格” (shadow price) 计费. 这时网络并不直接控制每个用户的传输行为，而是间接地在网络边缘通过计费来给用户提供正确的货币激励，使之合理地使用稀缺的带宽 [15] .

文献 [16] 总结了早期的拥塞计费方案. 但是基于使用的计费 (usage-based pricing) 往往需要路由器保存每个流的状态信息，会破坏网络的可扩展性. 为此， Odlyzko 提出 PMP(Paris Metro Pricing) 方案，将网络分为几个逻辑通道，对不同的通道收取不同的费用. 但每个通道的计费方式仍为平坦计费. PMP 方案可在网络中维持有不同拥塞水平的逻辑通道，但仍无法控制由信源的突发性造成的拥塞 [17] .

Kelly 提出的基于 ECN(Explicit Congestion Notification) 的拥塞计费 (ECN-based congestion pricing) 方案，对现有网络的 CR 改造最小：它只要求 CR 采用 ECN 来对分组进行标记 (设置分组的 ECN 位) ，不需要路由器保持任何关于流的状态信息 [18,19] . 因而下面仅讨论基于 ECN 的拥塞计费.

4.2 基于 ECN 的拥塞计费

4.2.1 实施方案　Kelly 的方案中，在 ER 上，除支持 ECN 外，还要执行计费功能：对每个被标记的分组收取一个小额的固定费用 Pr．由于 ECN 的标记算法依据路由器的平均队列长度来确定标记概率，因而被标记的分组总数 (相应于总的收费) 反映了网络的拥塞程度．又由于采用的是随机标记，因而每个流被标记的分组数 (相应于对每个流的收费) 与该流占据的瓶颈带宽成正比．对带宽资源评价高的用户，就愿意在单位时间收到更多被标记的分组 (相应地支付更高的费用)，从而获得更大的带宽．

4.2.2 用户端的优化行为　现在讨论用户端如何根据实际需要发送分组．一般地，考察一个分组大小相同的离散时间系统，设用户 i 在一个时间片里的支付意愿为 w_i，在第 n 个时间片里用户 i 发送的分组数目为 $x_i(n)$，收到的被标记的分组数目为 $f_i(n)$，则用户在下一时间片发送的分组数目为

$$x_i(n+1) = x_i(n) + k_i(w_i - \Pr f_i(n)) \tag{9}$$

其中 k_i 为一常数，上式的含义是，当支付意愿大于实际收费时，增大发送速率；反之则减少发送速率．

下面讨论 (9) 式给出的用户端行为是否合理：在该式中取 $n = 0, 1, \cdots, N$，可得

$$[x_i(N) - x_i(0)]/k_i = Nw_i - \Pr(f_i(1) + \cdots + f_i(N)) \tag{10}$$

上式右边表示在前 N 个时间片中，用户的支付意愿与实际支付费用之差．在等式两边都除以 N 并令 $N \to \infty$，则右边趋于零．这意味着从长远来看，实际支付费用将收敛于用户支付意愿．显然 k_i 越大，收敛速度越快，但给流量和实际收费带来的波动也大．用户应根据实际情况来设置 k_i，比如对数据量小的业务设置较大的 k_i 值．文献 [20] 进一步给出用户如何根据自己的效用函数来调节 w_i．

4.2.3 基于 ECN 的拥塞计费方案中存在的问题　基于 ECN 的计费机制中，单位带宽价格是一个随时间变化的随机变量．对于用户而言，某些业务需要持续一段时间，这给用户决策带来困难：用户不但要看当前价格，还要估计其后一段时间内价格的变化，而且还要承担预期与实际不符而带来的超出预算的风险．文献 [21] 对此给予了详细讨论，并通过引入衍生市场机制得到初步解决．

5 总结及展望

5.1 各种方案性能的综合比较

我们首先对 3 种方法的代表性解决方案的性能作一个综合比较 (表 1)．考察的内容涉及到 Internet 路由器中拥塞控制机制的主要设计目标：

(1) 拥塞控制的效果，包括路由器对各个流保护的能力，支持多业务的能力．

(2) 对网络可扩展性的影响，主要指其正常运作给 CR 带来的额外开销．

(3) 对网络改造的成本 (包括软、硬件) 及分步部署能力．

表 1　路由器上主要拥塞控制机制的性能比较

模式	考察项目					
	流保护能力	支持多业务的能力	可扩展性	对硬件改造	对软件改造	分步布署能力
	性能评价					
RED-PD	一般	弱	弱	较大	少	强
SCORE	强	一般	一般	较大	较大	一般
ECN-based pricing	一般	强	强	少	较大	弱

5.2 研究展望

通过上述比较，我们认为这一领域今后的研究应重点注意：

(1) 路由器对带宽分配的控制应当是通过缓冲区管理机制来实现。路由器对带宽分配的控制主要有两种手段：一是队列调度机制，直接控制与输出链路的接口；二是缓存区管理机制，通过控制队列中分组的数目间接影响对输出链路的占用，这种方法对带宽的分配是概率意义上的。由于队列调度机制一般需要更多的流状态信息，且涉及到的分类也是路由器中复杂度最大的操作，可扩展性差，成本高。因而上述 3 种方案都是使用缓冲区管理机制中的主动式队列管理 (AQM)，即 RED 的各种改进型再配合上传统的 FIFO 调度机制。

(2) 路由器在拥塞控制中的作用是为用户端的行为提供正确的激励。为此路由器需要用户使用资源的信息 (流状态信息)，但为了保证可扩展性，又要尽可能减少这些信息。表 2 比较了 3 种方案中信息处理和激励机制的异同，显然，ECN-based pricing 是比较成功的。因为它正确地为网络与用户间划分了职责：网络只是为用户正确使用网络提供信息和激励 (就是 ECN bit 及由此带来的计费)，让用户自己进行决策，因为只有用户自己才知道对业务的评价，ECN-based pricing 在不需要任何流状态信息的条件下能够提供基于流的服务，是 3 种方案中支持多业务能力最强的。RED-PD 和 SCORE 似乎不需要计费，但它们支持多业务时，在 ER 上往往要为不同的流设置不同的优先级或允许占用更大带宽的标识。而这些标识在 ER 上的分配最终要涉及到计费问题。因而支持多业务必然意味着对不同流的区分计费。而在拥塞控制设计阶段即考虑到计费 (或使用价格作为一种控制手段) 往往比在拥塞控制机制完成后再设计配套的计费机制更能提高效能，简化设计方案。

表 2 3 种方案的信息机制和激励机制比较

模式	比较项目	
	流状态信息	激励机制
RED-PD	来自网络自身	网络对非 TCP 友好流处罚
SCORE	分组携带	网络直接控制带宽分配
ECN-based pricing	无	货币激励

参 考 文 献

[1] S. Floyd, A report on recent developments in TCP congestion control, IEEE Communications Magazine, 2001, 39(4), 84–90.

[2] 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.

[3] R. Kapoor, C. Casetti, M. Gerla, Core-stateless fair bandwidth allocation for TCP flows, IEEE International Conference on Communications, Helsinki, Finland, 2001, 1, 146–150.

[4] V. P. Kumar, T. V. Lakshman, Beyond best effort: router architectures for the differentiated services of tomorrow's Internet, IEEE Communications Magazine, 1998, 36(5), 152–164.

[5] D. Lin, R. Morris, Dynamics of random early detection, Proc. from ACM SIGCOMM 97, Cannes, France, October 1997, 127–137.

[6] R. Mahajan, S. Floyd, Controlling high-bandwidth flows at the congested router, ICSI Tech Report TR-01-001, April 2001.

[7] P. Rong, B. Prabhakar, CHOKe, A stateless active queue management scheme for approximating fair bandwidth allocation, INFOCOM 2000, Tel Aviv, Israel, 2000, Vol. 2, 942–951.

[8] P. Gevros, J. Crowcroft, P. Kirstein, S. Bhatti, Congestion control mechanisms and the best effort service model, IEEE Network, 2001, 15(3), 235–277.

[9] J. Padhye, V. Firoiu, D. Towsley, J. Kurose, Modeling TCP throughput: a simple model and its empirical validation, ACM SIGCOMM'98, Vancouver, USA, 1998, 303–314.

[10] R. Roy, R. C. Mudumbai, S. S. Panwar, Analysis of TCP congestion control using a fluid model, IEEE International Conference on Communications, Helsinki, Finland, 2001, Vol. 8, 151–160.

[11] J. Widmer, R. Denda, M. Mauve, A survey on TCP-friendly congestion control, IEEE Network, 2001, 15(3), 28–37 .

[12] I. Stoica, H. Zhang, Providing guaranteed services without per flow management, proc. of SIGCOMM'99, Boston, MA, 1999, 81–94.

[13] Cao Zhiruo, Wang Zheng, E. Zegura, Rainbow fair queueing: fair bandwidth sharing without per-flow state, IEEE Proc. of INFOCOM 2000, Tel Aviv, Israel, 2000, vol.2, 922–931.

[14] M. Nabeshima, T. Shimizu, I. Yamasaki, Fair queuing with in/out bit in core stateless networks, IEEE Proc. of IWQoS, 2000, Pittsburgh, USA, 2000, 3–9

[15] T. Henderson, J. Crowcroft, S. Bhatti, Congestion pricing, paying your way in communication networks, IEEE Internet Computing, 2001, 5(5), 85–89.

[16] R. Cocchi, S. Shenker, D. Estrin, L. Zhang, Pricing in computer networks: Motivation, formulation and example, IEEE/ACM Trans. on Networking, 1993, 1(6), 614–627.

[17] A. M. Odlyzko, Paris Metro Pricing: The minimalist differentiated services solution, IEEE Proc. of IWQoS '99, London, UK, 1999, 159–161.

[18] F. Kelly, A. Maulloo, Rate control in communication networks: shadow prices, proportional fairness and stability, Journal of the Operational Research Society, 1998, 49(3), 237–252.

[19] F. Kelly, Mathematical modelling of the Internet, In Mathematics Unlimited–2001 and Beyond (Editors B. Engquist, W. Schmid), Berlin, Springer-Verlag, 2001, 685–702.

[20] R. J. Gibbens, F. Kelly, Resource pricing and the evolution of congestion control, Automatica, 1999, 35(10), 1969–1985.

[21] Zhang Chi, Wei Jiaolong, An economic model for QoS guarantee in the Internet, Proc. of SPIE, Wuhan, 2001, Vol. 4556, 116–121.

RESEARCH ON ROUTER MECHANISMS FOR CONGESTION CONTROL IN THE INTERNET

Wei Jiaolong Zhang Chi

(*Dept. of Electron. and Info., Huazhong Univ. of Sci. and Tech., Wuhan* 430074, *China*)

Abstract This paper summaries the recent research advance in the field providing congestion control mechanisms as powerful and flexible as the ones implemented by a stateful network. In a partial stateful or stateless network, three kinds of mechanisms are introduced for Internet routers: partial per-flow state, stateless core, congestion pricing and their representative implementation schemes are discussed in detail. The paper summaries this research field from the information and incentive point of view, discusses the internal relationships between these mechanisms and analyzes their performance characteristics.

Key words Network congestion control, Partial per-flow state, Stateless core, Congestion pricing

魏蛟龙: 男, 1965 年生, 在职博士生, 副教授, 现主要从事广域网拥塞控制、IP 网服务质量保障、人工智能与专家系统、故障诊断等方面的研究.

张 驰: 男, 1977 年生, 硕士生, 现主要研究领域为网络拥塞和流量控制.