

一种基于速率和队列长度的主动队列管理机制

张鹤颖¹, 刘宝宏², 窦文华¹

(11 国防科学技术大学计算机学院, 湖南长沙 410073; 21 国防科学技术大学自动化所, 湖南长沙 410073)

摘要: 本文研究了拥塞产生的原因及其表现形式, 提出了一种基于报文到达速率和队列长度的随机早丢弃算法(RQ). 根据拥塞的严重程度和变化趋势将拥塞划分为六个级别, 每个级别采用合适的丢弃概率, 从而将队列控制在理想工作点附近. 利用经典控制理论, 分析了系统的稳定性, 给出了参数配置的原则. 最后, 用 NS 网络仿真器对算法性能进行了验证.

关键词: 拥塞控制; 主动队列管理; 报文到达速率; 队列长度

中图分类号: TP393 **文献标识码:** A **文章编号:** 0372-2112 (2003) 12-1743-04

A Rate and Queue Controlled Active Queue Management

ZHANG He2ying¹, LIU Bao2hong², DOU Wen2hua¹

(1. School of Computer Science, National University of Defense Technology, Changsha, Hunan 410073, China;

2. Institute of Automation, National University of Defense Technology, Changsha, Hunan 410073, China)

Abstract: Active Queue Management is an active research area. Its main objective is to provide low delay and low loss service in packet-switched network. RQ is a new random early detection algorithm. It can divide the seriousness of congestion into six levels according to the aggregate packet arrival rate and the average queue length in the router. Each level has appropriate packet drop probability. The responsiveness and stability properties of RQ are superior to RED in favor of promptly reacting of rate-based dropping than queue-based. TCP and RQ are modeled as a closed loop feedback control system and the stability is analyzed. The performance of RQ was compared with RED by simulations. The results show that RQ can keep low queuing delay and low packet loss in wide variable load conditions.

Key words: congestion control; active queue management; packet arrival rate; queue length

1 引言

主动队列管理(AQM)^[1]是基于路由器的拥塞控制领域的研究热点之一. AQM 主要是针对传统路由器中的/丢弃尾0队列提出来的, 目的是设计一种有效的拥塞早期检测机制, 将队列长度控制在较低的水平, 避免缓冲区溢出, 并减少报文丢失. 典型的 AQM 机制是 RED 算法^[2]. RED 根据平均队列长度调节报文丢弃概率, 在保持高吞吐率和低延迟的同时, 能够容纳一定的突发数据. RED 算法最大的缺陷是性能与参数设置密切相关, 静态参数设置只适用于特定网络条件, 而通用的参数配置很难用分析方法得到. 目前 RED 参数配置的研究主要集中在两个方向: 其一是利用随机过程, 自动控制理论等工具建立 TCP/RED 系统的模型^[3,4], 通过稳定性分析给出参数配置的原则. 这些参数配置能够保证系统的稳态性能, 但不能保证暂态性能; 其二是动态调节 RED 的关键参数^[5], 使丢弃概率与网络拥塞程度相吻合, 消除 RED 的参数敏感性. 这种方法主要通过动态调节最大丢弃概率 \max_p , 将队列长度控制在目标值附近. 实现的难点在于参数调节的频率难以把握, 需要在

稳定性和响应速度之间做出折衷.

本文研究了拥塞产生的原因, 提出了一种基于报文到达速率和平均队列长度检测拥塞的 RQ(Rate and Queue based random early detection)算法. 该算法能够较精确地跟踪网络状态变化, 使报文丢弃概率符合网络拥塞程度和拥塞变化的趋势, 从而改善了系统的响应速度, 并且能够保证低延迟和低丢失率.

2 拥塞产生的根本原因及直观表现

假设有 n 个流经过同一个路由器, 流 i 在 t 时刻到达路由器的速率是 $r_i(t)$, 路由器输出链路的带宽是 C , 报文到达速率是 $A(t) = \sum_{i=1}^n r_i(t)$. 如果 $A(t) \leq C$, 所有到达的报文无需等待, 立刻被发送出去. 反之, 如果 $A(t) > C$, 部分到达的报文就需要经过排队等待, 才能被发送出去. 如果这种状态维持下去, 排队等待的报文将逐渐增多. 存储容量有限的缓冲区最终将会溢出, 以后到达的报文将被丢弃.

上述分析表明: 报文到达速度大于链路输出带宽是拥塞

发生的根本原因, 队列长度的增长是拥塞累积到一定程度的结果. 所以, 报文到达速率和平均队列长度分别反映拥塞的瞬时状况和持续状况.

在实际中, 路由器的缓冲区容量总是有限的. 为了保持链路的高利用率, 需要将报文到达速率控制在链路容量附近, 即 $A(t) = C$, 这是路由器中队列的理想工作状态. 如果 $A(t) < C$, 网络不会发生拥塞, 但是链路利用率低; 反之, 如果 $A(t) > C$, 网络容易发生拥塞, 队列性能类似于传统的/ 弃尾0 队列, 无法体现 AQM 的优越性.

3 RQ 算法

基于上述分析, 我们提出了一种混合拥塞检测方法 RQ. RQ 算法基于报文到达速率和平均队列长度决定报文丢弃概率, 使丢弃概率符合当前的拥塞程度和未来拥塞变化的趋势. 假设平均队列长度为 avg , 队列最小和最大阈值分别为: min_{th} 和 max_{th} . 报文丢弃概率按下列公式计算:

(1) 当 $avg \leq min_{th}$ 时:

$$pd = \begin{cases} 0, & \text{if } A \leq C_t, \\ 1 - C_t / A, & \text{if } A > C_t. \end{cases}$$

(2) 当 $min_{th} < avg < max_{th}$ 时:

$$pd = \begin{cases} \frac{avg - min_{th}}{max_{th} - min_{th}} max_p, & \text{if } A \leq C_t, \\ \min \left\{ 1, 1 - \frac{C_t}{A} + \frac{avg - min_{th}}{max_{th} - min_{th}} max_p \right\}, & \text{if } A > C_t. \end{cases}$$

(3) 当 $avg \geq max_{th}$ 时:

$$pd = \begin{cases} 2max_p - 1 + \frac{avg}{max_{th}} (1 - max_p), & \text{if } A \leq C_t, \\ 1, & \text{if } A > C_t. \end{cases}$$

为了保护瞬时突发流, 报文到达速率按下式计算:

$$A_{new} = \left(1 - e^{-\frac{D^k}{K}} \right) \frac{1^k}{D^k} + e^{-\frac{D^k}{K}} A_{old}$$

其中 $D^k = t^k - t^{k-1}$, t^k 为第 k 个报文的到达时间, 1^k 为第 k 个报文的长度, K 为常数.

平均队列长度的计算公式为:

$$avg_{new} = (1 - w_q) @ avg_{old} + w_q @ q$$

目标链路带宽 $C_t = C @ C$, 目标利用率 $C[1]$.

从报文丢弃概率计算公式可以看出, RQ 中丢弃概率的计算方法更符合网络状态, RQ 具有良好的稳定性. 我们将通过 TCP/RQ 系统的稳定性分析给出证明.

4 系统模型及稳定性分析

TCP 与 AQM 所的闭环反馈系统可描述为^[6]:

$$\begin{cases} \dot{\hat{W}}(t) = \frac{1}{R(t)} - \frac{W(t)W(t-R(t))}{2R(t-R(t))} p(t-R(t)) \\ \dot{\hat{q}}(t) = \frac{W(t)}{R(t)} N(t) - C \end{cases} \quad (1)$$

其中, C 为链路容量, $W(t)$ 为 TCP 连接的窗口大小, $q(t)$ 为队列长度, $p(t)$ 为丢弃概率, $N(t)$ 为激活的连接数目, $R(t)$ 为传送延迟, 包括排队延迟和传输延迟 T_p .

若令报文到达速率为: $A(t) = \frac{W(t)}{R(t)} N(t)$, 则方程(1)可

改写为:

$$\begin{cases} \dot{\hat{A}}(t) = \frac{N(t)}{R^2(t)} - \frac{A(t)A(t-R(t))}{2N(t)} - p(t-R(t)) \\ \dot{\hat{q}}(t) = A(t) - C \end{cases} \quad (2)$$

将 (A, q) 作为系统的状态变量, p 作为系统的输入. 由 $\dot{\hat{A}} = 0$, $\dot{\hat{q}} = 0$ 可求得系统的工作点 (A_0, q_0, p_0) :

$$p_0 = \frac{2N^2}{A_0^2 R_0^2}, A_0 = C, \text{ 其中 } R_0 = \frac{q_0}{C} + T_p$$

对系统在工作点附近进行线性化, 忽略延迟, 并进行拉氏变换, 可得到:

$$\begin{cases} sDA(s) = -\frac{2N}{CR_0^2} DA(s) - \frac{2N}{CR_0^3} Dq(s) - \frac{C^2}{2N} e^{-sR_0} Dp(s) \\ sDq(s) = DA(s) \end{cases} \quad (3)$$

RED 的控制率为^[6]:

$$C_{red}(s) = \frac{L_{red}}{1 + s/K_1} \quad (4)$$

其中 $L_{red} = \frac{max_p}{max_{th} - min_{th}}$, $K_1 = -\frac{\ln(1 - w_q)}{D}$, D 为采样间隔.

基于速率的控制率为:

$$C_r(s) = \frac{1/C_t}{1 + s/K_2} \quad (5)$$

其中 $K_2 = 1/K$.

联立方程(3)、(4)、(5)可得系统框图, 如图 1 所示.

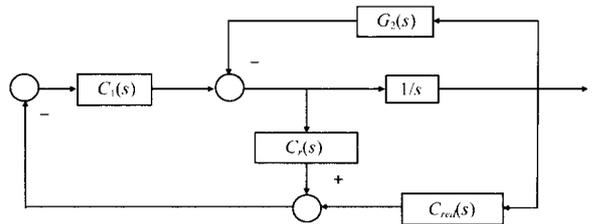


图 1 系统框图

图 1 中, $G_1(s) = \frac{C^2 e^{-sR_0}}{s + \frac{2N}{CR_0^2}}$, $G_2(s) = \frac{2N}{s + \frac{2N}{CR_0^2}}$.

经过等效变换, 可得到系统的开环传递函数:

$$G(s) = \frac{s^2 + K_1 \left[1 + \frac{C_t L_{red}}{K_2} \right] s + K_1 C_t L_{red}}{\left[s^2 + \frac{2N}{CR_0^2} s + \frac{2N}{CR_0^3} \right] (s^2 + (K_1 + K_2) s + K_1 K_2)} \frac{C^2 K_1}{2N} \frac{1}{C_t} e^{-sR_0}$$

由于 K_1 和 K_2 同量纲同数量级, 不妨令 $K_1 = K_2 = K_c$, 系统的传递函数可以简化为

$$G(s) = \frac{s / (C_t L_{red}) + 1}{((s/w_n)^2 + 2F(s/w_n) + 1)(s/K_c + 1)} \frac{C^3 R_0^3}{4N^2} L_{red} e^{-sR_0}$$

其中, $w_n = \sqrt{\frac{2N}{CR_0^3}}$, $F = \sqrt{\frac{N}{2CR_0}}$.

对于该系统的稳定性, 有如下定理:

定理 1 如果 L_{red} 和 K_c 满足

$$\frac{(CR_0)^3}{4N^2} \sqrt{L_{red}^2 + w_g^2 / C_t^2} \left[\sqrt{1 + w_g^2 / K_c^2} \right]$$

其中 $w_g = 0.01 \min \left\{ \sqrt{\frac{2N}{CR_0^3}}, \frac{1}{R_0} \right\}$, 则对所有 $N \setminus N_c$, $R[R_c$

闭环系统是稳定的.

证明 当 $w \ll w_n$ 时, 振荡环节可忽略. 考虑系统的频率特性:

$$G(j\omega) U \frac{j\omega + 1}{C_t L_{red}} \frac{C^3 R_0^3}{4N^2 L_{red}} P w I [0, w_g]$$

$$|G(j\omega)| = \sqrt{(1 + \omega^2 / C_t^2 L_{red}^2) / (1 + \omega^2 / K_c^2)} \frac{C^3 R_0^3}{4N^2 L_{red}}$$

根据定理中的条件, 对所有 $N \ll N_c, R \ll R_c$, 有 $|G(j\omega_g)| \ll 1$.

$$N G(j\omega_g) = \arctg \frac{w_g}{C_t L_{red}} - \arctg \frac{w_g}{K_c} - w_g R_0$$

$$\ll - \arctg \frac{w_g}{K_c} - w_g R_0$$

$$\ll - 90^\circ - 0.01 \frac{180}{P}$$

$$> - 180^\circ$$

由奈奎斯特稳定性判据可知, 闭环系统是稳定的. 证毕.

从定理 1 的证明过程可以看出, TCP/RQ 系统的相角裕量大于 TCP/RED 系统的相角裕量^[9], 表明 RQ 算法的稳定性优于 RED 算法. 其原因是 RQ 基于速率的丢弃在时间上比基于队列长度的丢弃早. 从系统框图容易看出, 基于队列长度的反馈要经过一个积分环节, 其相位比基于速率的反馈滞后 90°.

5 仿真结果分析

我们用 NS 仿真器对 RQ 算法的性能进行了验证. 仿真中采用的拓扑结构如图 2 所示. 端节点 s_i 和 d_i 之间建立连接, 路由器 r_1 和 r_2 之间构成瓶颈链路, 链路带宽为 10Mbps, 传输延迟为 10ms. 其他链路的带宽为 30Mbps. 发送端采用 TCP

Reno 协议, 报文平均长度为 1000B, TCP 连接的传输延迟分布在 40ms 到 220ms 之间. 路由器中的缓冲区最多可容纳 1000 个报文, r_1 采用标记策略进行拥塞反馈. 用 FTP 流模拟长连接.

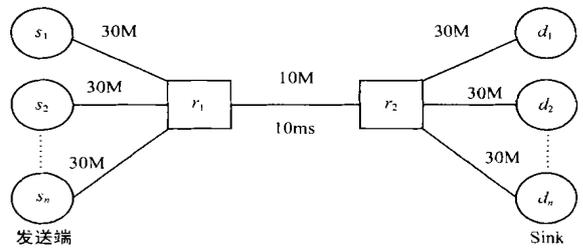


图 2 实验网络拓扑

根据定理 1, 取 $N_c = 20, R_c = 0.23\text{sec}, C = 1250\text{pkt/s}, D = 1/C, K_c = 215$. 可求得 $w_g = 0.016217\text{rad/sec}, w_q = 0.002, K = 0.4, L_{red} [6.62 @ 10^{-5}]$, 如果 $\max_p = 0.05$, 则 $\max_{th} - \min_{th} \ll 756$. 仿真中令 $\min_{th} = 90, \max_{th} = 850$.

实验 1 仿真开始时 FTP 流的数目为 60, 50 秒时 30 个 FTP 流退出. 图 3(a), 图 3(b) 分别表示瞬时队列和平均队列变化情况. 从图中可以看出, RQ 的稳态队列长度小于 RED. 当负载增加或减少时, RQ 收敛到稳定状态的时间比 RED 短, 而且没有瞬时过冲, 表明 RQ 的稳定性和响应速度优于 RED. 实验结果与模型分析结论一致.

实验 2 FTP 流的数目从 20 变化到 160 时, 每种情况下进入稳态后, 100 秒内丢失的报文总数如图 4 所示. 在低负载情况下, 由于 RED 和 RQ 的平均队列长度小于 \max_{th} , 所以报文丢弃率为 0. 当负载增加, 平均队列长度大于 \max_{th} 时, 报文丢弃率相应增加. RQ 的丢弃率小于 RED.

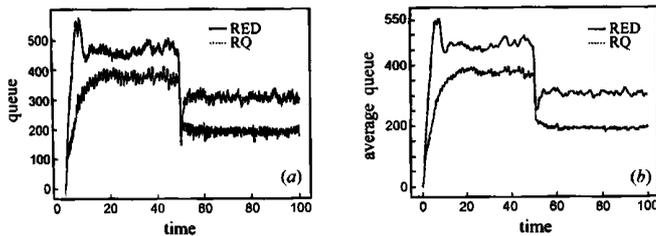


图 3 (a) 瞬时队列长度; (b) 平均队列长度

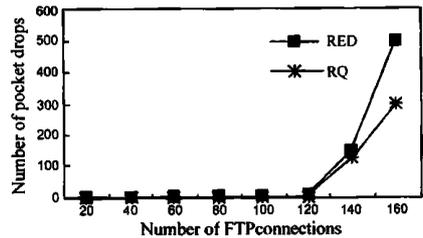


图 4 报文丢失率

实验 3 FTP 流的数目为 60, \min_{th} 从 20 变化到 100, 图 5(a) 比较了 RED 与 RQ 的平均队列长度. RED 的平均队列长度随着 \min_{th} 的增加而增长, RQ 的平均队列长度保持不变. 图

5(b) 表示 $\min_{th} = 90, \max_p$ 从 0.05 增长到 0.1 时平均队列长度的变化. RQ 的平均队列长度受 \max_p 的影响较小, RED 的队列长度变化剧烈, 表明 RQ 降低了对参数的敏感度.

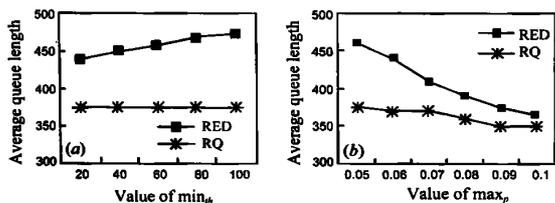


图 5 算法性能对主要参数的敏感性

6 总结

本文分析了拥塞产生的根本原因及直观表现, 提出了一种基于报文到达速度和队列长度的 AQM 算法, 建立了系统的模型, 并分析了系统的稳定性. 由于 RQ 算法能够较早检测拥塞, 并且采用符合网络状态的丢弃概率, 仿真结果表明该算法具有良好的稳态和暂态性能. 此外, RQ 算法没有增加额外的参数, 易于实现. 值得指出的是, 基于速率的丢弃可以作为一个独立的功能单元与其他基于队列长度的 AQM 机制相结合

以改善其响应速度.

参考文献:

- [1] B Braden, D Clark, J Crowcroft, B Davie, S Deering, D Estrin, S Floyd, V Jacobson, G Minshall, C Partridge, L Peterson, K Ramakrishnan, S Shenker, J Wroclawski, L Zhang. Recommendations on queue management and congestion avoidance in the internet[Z]. RFC2309, April 1998.
- [2] S Floyd, V Jacobson. Random early detection gateways for congestion avoidance[J]. IEEE/ACM Transactions on Networking, 1993, 1(4): 397- 413.
- [3] V Firoiu, M Borden. A study of active queue management for congestion control[A]. In Proceedings of IEEE INFOCOM 2001[C]. TelAviv Is2rael, 2000. 1435- 1444.
- [4] C Hollot, V Misra, D Towsley, W Gong. A control theoretic analysis of RED[A]. In Proceedings of IEEE INFOCOM 2001[C]. Anchorage, Alaska, USA, 2001. 1510- 1519.
- [5] W Feng, D Kandlur, D Saha, K Shin. A self configuring RED gateway [A]. in Proceedings of IEEE INFOCOM 1999[C]. New York, USA, 1999. 1320- 1328.
- [6] Vishal Misra, Weibo Gong, Don Towsley. Fluidbased analysis of a

network of AQM routers supporting TCP flows with an application to RED[A]. in Proceedings of ACM SIGCOMM 2000[C]. Stockholm, Sweden, 2000. 151- 160.

作者简介:



张鹤颖 女, 1976 年 10 月生于陕西户县, 1998 年毕业于国防科技大学机电工程与自动化学院, 同年直接攻读博士学位, 主要研究方向是高速网络互连, 网络拥塞控制.

刘宝宏 男, 1974 年 10 月生于天津, 1998 年毕业于国防科技大学机电工程与自动化学院, 同年直接攻读博士学位, 主要研究方向是控制系统的建模与仿真.

窦文华 男, 1946 年 11 月生于山西平定, 国防科技大学计算机学院教授, 博士生导师, 主要研究方向: 网络安全, 高速网络互连, CIMS, 并行与分布处理等.