

# 一类具有随机离去策略的排队系统及其在 Internet 拥塞控制中的应用<sup>\*</sup>)

汪浩 李晓明 严伟

(北京大学信息科学技术学院网络实验室 北京 100871)

**摘要** 当顾客到达一个排队系统时,可以根据当前队列的长度,按照一定的概率选择加入排队系统等待队列,或选择离开排队系统且不再返回。这样的排队系统称为“具有随机离去策略的排队系统”。我们利用输入流稀疏化(Thinning of recurrent flows)和输入流阻行(Balking)两种方法分别扩充了 GI/M/1/N 和  $GI^X/M/1/N$  两个排队系统,建立了 3 个具有随机离去策略的排队系统,给出了这类排队系统的顾客丢失率、系统利用率、队列长度的均值/方差、平均等待时间等性能指标。最后,讨论了这类排队系统在 Internet 拥塞控制中的应用。

**关键词** 排队系统,输入流稀疏化,输入流阻行,GI/M/1/N, $GI^X/M/1/N$ ,Internet,拥塞控制

## Some Queueing System with Randomly Going Away Policy and its Application to the Congestion Control of the Internet

WANG Hao LI Xiao-Ming YAN Wei

(Network Computing Lab, School of Electronics Engineering and Computer Science, Peking University, Beijing 100871)

**Abstract** When a customer arrives at a queueing system, he can decide whether to join it or go away and never return according to the length of the queue. We call a queueing system with such mechanism “a queueing system with randomly going away mechanism”. In this paper, we extend two queueing systems GI/M/1/N and  $GI^X/M/1/N$  by means of thinning and balking techniques respectively and establish three queueing systems with randomly going away mechanism. In addition, some performance metrics are given, such as customer loss rate, system utilization, the mean and variance of queue length, the mean waiting time. The application to the Internet congestion control is discussed concerning these queueing systems.

**Keywords** Queueing system, Thinning of recurrent flows, Balking, GI/M/1/N,  $GI^X/M/1/N$ , Computer network, Congestion control

### 1 引言

如果在 GI/M/1/N 或  $GI^X/M/1/N$  排队系统前加一个控制机制,该机制根据当前队列的长度,对每一个到达的顾客,按一定的概率决定是否允许他加入等待队列;也可以等价地认为每一个到达的顾客要根据当前队列的长度,按照一定的概率决定是否离开排队系统且不再返回,如图 1 所示。其中的“菱形”表示顾客在那里要做一个决定,决定是否加入或离开排队系统等待队列。我们称具有这种机制的排队系统为“具有随机离去策略的排队系统”。

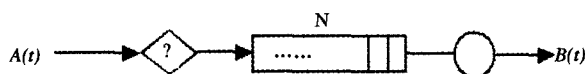


图 1 具有随机离去策略的排队系统

这样一种考虑有什么意义?

在日常生活中,这样的考虑是有意义的。例如人们去银行办事,供等待的座位是有限的,但许多顾客通常是看到当时

人比较多(而不是座位都占满了),就不参加排队了。

在 Internet 拥塞控制中,也有类似机制。例如,主动队列管理(Active Queue Management)机制中的 RED 算法在 IP 分组(也称“IP 包”)到达队列前端时,即便缓存队列未填满,也会根据缓存队列的长度(以分组为单位)随机地丢弃到达的 IP 分组(IP 包)。

#### 相关研究

许多作者讨论过具有随机离去策略的排队系统。Haight 讨论了顾客到达间隔服从指数分布、服务时间也服从指数分布且具有随机离去策略的排队系统<sup>[2,3]</sup>;Ancker 讨论了顾客到达间隔服从指数分布、服务时间也服从指数分布且具有随机离去和随机违约<sup>1)</sup>策略的排队系统<sup>[4,5]</sup>;Rao 讨论了顾客到达间隔服从指数分布、服务时间服从任意分布且具有随机离去和随机违约策略的排队系统<sup>[6]</sup>;Drekic 讨论了顾客到达间隔服从指数分布、具有抢先优先级和随机离去策略的排队系统<sup>[7]</sup>。以上作者所涉及的具有随机离去策略的排队系统,顾客的到达间隔均假定服从指数分布。但由于互联网中数据包(通常称作“分组”)到达网络交换设备(如路由器或交换机的

<sup>\*</sup>) 本项工作受国家自然科学基金项目(90412010)和 973 项目(2004CB318204)资助。汪浩 博士生,主要研究方向:QoS、队列管理、拥塞控制;李晓明 博士,教授,博士生导师,主要研究方向:网络与分布式系统、并行计算;严伟 副教授,主要研究方向:无线 Ad hoc 网、无线 mesh 网。

<sup>1)</sup> 即顾客进入排队系统等待队列后,仍然可以随机离开的策略。

到达间隔不是指数分布,而是 Pareto 或 Weibull 分布<sup>[8,15,16]</sup>,并且随着网络流量的增大,到达网络交换设备的分组具有成批到达的特征<sup>[9]</sup>,因此为了用排队论的方法分析网络交换设备的性能,我们特别地关注顾客的到达间隔服从一般分布、顾客单个或成批到达、服务时间服从指数分布、1 个服务台、等待队列容量有限且具有随机离去策略的排队系统。Finch 讨论了顾客的到达间隔服从一般分布、服务时间服从指数分布、1 个服务台、等待队列长度有限且具有随机离去策略的排队系统 GI/M/1/N<sup>[10]</sup>,但只讨论了顾客单个到达时的情形。

为了讨论诸如 GI/M/m 或 M/GI/m(等待队列容量有限或无穷)的排队系统, Cox 提出了“辅助变量法”的方法<sup>[11]</sup>。利用辅助变量法, Hokstad 讨论了顾客的到达间隔服从一般分布、服务时间服从指数分布、多个( $m$  个)服务台且等待队列容量有限的排队系统 GI/M/m/N<sup>[12]</sup>。利用辅助变量法和 Markov 嵌入链, Vijaya 等人讨论了顾客到达间隔服从一般分布、顾客成批到达、服务时间服从指数分布、多个服务台( $c$  个)、等待队列容量有限的排队系统 GIX/M/c/N<sup>[13]</sup>。但 Hokstad 和 Vijaya 等人讨论的排队系统均不具有随机离去策略。到目前为止,我们还未发现有文献讨论过顾客成批到达、到达间隔服从一般分布、服务时间服从指数分布、1 个服务台、等待队列容量有限且具有随机离去策略的排队系统。

我们在 Hokstad、Vijaya 和 Finch 工作的基础上,分别利用输入流稀疏化(Thinning of recurrent flows)<sup>[14]</sup>和输入流阻行(Balking)<sup>[2~5]</sup>两种方法扩充了 GI/M/1/N 和 GI<sup>X</sup>/M/1/N 两个排队系统,建立了 3 个具有随机离去策略的排队系统(见表 1),讨论了该排队系统的顾客丢失率、系统利用率、队列长度的均值/方差、平均等待时间等性能指标<sup>[17~19]</sup>。本文的主要目的是综述这 3 个排队系统及其主要结论和相互关系,相关定理的证明见我们已发表的论文<sup>[17~19]</sup>。

本文后续安排如下:第 2 节综述了上述 3 个排队系统的主要结论和相互关系,以及排队系统性能指标的计算公式;第 3 节指出了这类排队系统在 Internet 拥塞控制中的应用。

表 1 本文建立的 3 个具有随机离去策略的排队系统

	输入流稀疏化	输入流阻行
顾客单个到达	基于输入流稀疏化的具有随机离去策略的 GI/M/1/N 排队系统 <sup>[17]</sup> (作者建立)	基于输入流阻行的具有随机离去策略的 GI/M/1/N 排队系统 <sup>[10]</sup> (Finch 建立)
顾客成批到达	基于输入流稀疏化的具有随机离去策略的 GI <sup>X</sup> /M/1/N 成批到达排队系统 <sup>[18]</sup> (作者建立)	基于输入流阻行的具有随机离去策略的 GI <sup>X</sup> /M/1/N 成批到达排队系统 <sup>[19]</sup> (作者建立)

## 2 一类具有随机离去策略的排队系统

从本节开始,我们约定如下记号:

$N$ :表示排队系统能容纳的顾客的个数,其中一个顾客正在接受服务,其余的顾客( $\leq N-1$ )在等待队列等待。

df:表示概率分布函数

pdf:表示概率密度函数

r. v.:一个随机变量

r. v. s:多个随机变量

i. i. d:独立同分布

我们先给出与随机离去策略相关的顾客离去概率序列和顾客阻行概率序列的概念及其作用。

**定义** 称满足条件  $0=d_0 \leq d_1 \leq \dots \leq d_{N-1} < d_N=1$  的序列  $\{d_i\}$  ( $i=0,1,\dots,N$ ) 为顾客离去概率序列;而称序列  $\{r_i: r_i=1-d_i, i=0,1,\dots,N\}$  为顾客阻行概率序列。

这两个序列的作用如下:假设排队系统中当前有  $i$  个顾客( $i=0,1,\dots,N$ ),当一个(或一批)顾客到达时,该个顾客(或该批顾客)以概率  $d_i$  离开排队系统,并且不再回来,或者说以概率  $r_i$  加入排队系统等待队列。

我们称顾客的以上策略是“随机离去策略”,对应的排队系统为“具有随机离去策略的排队系统”。

为了全面刻画一个排队系统,需要知道排队系统的如下两个概率分布。

$q_k$ :在稳态情况下,一个顾客到达前一瞬间,系统中有  $k$  个顾客的概率分布,其中  $0 \leq k \leq N$ 。

$p_k$ :在稳态情况下,在任意时刻,系统中有  $k$  个顾客的概率分布,其中  $0 \leq k \leq N$ 。

为了求解具有随机离去策略的排队系统的队列长度的概率分布,有两种方法可供选择,它们分别是白箱法和黑箱法。“白箱法”对应“输入流稀疏化”的方法;“黑箱法”对应“输入流阻行”的方法。以下分别叙述这两种方法。

**白箱法:**对具有随机离去策略的排队系统,当一个(一批)顾客到达排队系统,并看到排队系统中有  $i$  ( $i=0,1,\dots,N$ ) 个顾客时,该个(该批)顾客将以概率  $d_i$  离开排队系统。因此,真正进入排队系统等待队列的顾客的到达间隔已不是  $A(t)$ ,而是一个与  $d_i$  (或  $r_i$ ) 相关的分布函数,我们将其记为  $A_i(t)$ ,见图 2。此时,进入排队系统等待队列的顾客流比到达排队系统的顾客流更稀疏,因此称为输入流的稀疏化。一旦有了  $A_i(t)$ ,我们就能设法求解  $q_k$  和  $p_k$ ,参见 2.1 和 2.2 小节。

假设(一个或一批)顾客到达排队系统的时刻依次为  $\tau_0, \tau_1, \tau_2, \dots, \tau_n, \dots$ ,其中  $\tau_0 < \tau_1 < \tau_2 < \dots < \tau_n < \dots$ ,且  $\tau_0=0$ ;到达间隔: $e_i = \tau_i - \tau_{i-1}$  ( $i=1, 2, \dots, n, \dots$ ) 是独立同分布随机变量,其 df 为  $A(t)$ 。当一个(一批)顾客到达排队系统,并看到排队系统中有  $i$  ( $i=0,1,\dots,N$ ) 个顾客时,该个(该批)顾客将以概率  $d_i$  离开排队系统。此时,顾客进入等待队列的时间间隔变为

$$e_1 + e_2 + \dots + e_\gamma$$

其中, $\gamma$  是与  $d_i$  (或  $r_i$ ) 相关的取值于集合  $\{1, 2, \dots\}$  的离散型随机变量。如果  $\gamma$  服从参数  $r_i$  的几何分布,即  $\Pr(\gamma=l) = r_i(1-r_i)^{l-1}$ ,  $l=1, 2, \dots$ ,此时称顾客按几何分布随机离去。

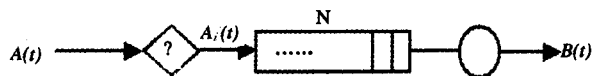


图 2 输入流稀疏化的基本思想

1)当  $i=0,1,2,\dots,N-1$  时,  $A_i(t)$  与  $d_i$  (或  $r_i$ )、 $A(t)$  之间具有如下关系:

$$A_i(t) = \Pr(e_1 + e_2 + \dots + e_\gamma \leq t) \stackrel{\infty}{=} \Pr(\gamma = l) A^l(t) \quad (1)$$

如果  $\gamma$  服从参数  $r_i$  的几何分布,则(参见文[14], pp68~69):

$$A_i(t) = \sum_{l=1}^{\infty} r_i(1-r_i)^{l-1} A^l(t) = \sum_{l=1}^{\infty} (1-d_i) d_i^{l-1} A^l(t) \quad (2)$$

其中  $A^l(t)$  表示  $A(t)$  的  $l$  次卷积运算,且  $r_i$  (或  $d_i$ ) 和  $A(t)$  为已知量。式(2)的 Laplace-Stieltjes 变换如下:

$$a_i(s) = \frac{r_i a(s)}{1 - (1-r_i) a(s)} = \frac{(1-d_i) a(s)}{1 - d_i a(s)} \quad (3)$$

其中  $a_i(s)$  和  $a(s)$  分别是  $A_i(t)$  和  $A(t)$  的 Laplace-Stieltjes 变换。

2)当  $i=N$  时,  $A_N(t)$  可用下式估计(参见文[14]68~69页):

$$A_N(t) = 1 - e^{-\frac{t}{a}} \quad (4)$$

其中  $a$  是  $A(t)$  的数学期望。

**黑箱法:**如果顾客按几何分布随机离去,我们可以利用辅助变量法和 Markov 嵌入链法直接求解  $q_k$  和  $p_k$ ,而不必知道  $A_i(t)$ 。这就是“输入流阻行”的方法,也是文[2~7]所用的方法,参见 2.3 小节。也就是说,“输入流阻行”方法是“输入流稀疏化”方法当“顾客按几何分布随机离去”时的特例。

### 2.1 基于输入流稀疏化的具有随机离去策略的排队系统 GI/M/1/N<sup>[17]</sup>

一个基于输入流稀疏化的具有随机离去策略的排队系统 GI/M/1/N 可表述为:

1)顾客到达排队系统的时刻依次为  $\tau_0, \tau_1, \tau_2, \dots, \tau_n, \dots$ , 其中  $\tau_0 < \tau_1 < \tau_2 < \dots < \tau_n < \dots$ , 且  $\tau_0 = 0$ ; 到达间隔:  $e_i = \tau_i - \tau_{i-1} (i=1, 2, \dots, n, \dots)$  是独立同分布随机变量,其 df 为  $A(t)$ , pdf 为  $a(t)$ , 数学期望为  $a$ 。

2)一个服务台,服务时间服从参数为  $\mu$  的负指数分布,其 df 记为  $B(t) = 1 - \exp(-\mu t)$ 。服务规则采用先到先服务 FCFS。

3)如果排队系统中当前有  $i (i=0, 1, \dots, N)$  个顾客,当一个新的顾客到达时,该顾客以概率  $r_i$  进入排队系统等待队列,或者说以概率  $d_i$  离开排队系统,其中  $r_i + d_i = 1$ 。此时,顾客到达排队系统等待队列的间隔时间的 df 为  $A_i(t) (i=0, 1, \dots, N)$ , pdf 为  $a_i(t) (i=0, 1, \dots, N)$ , 均值为  $a_i (i=0, 1, \dots, N)$ 。  $A_i(t)$  与  $A(t), r_i$  (或  $d_i$ ) 的关系见公式(1)~(3)。

4)排队系统最多能容纳  $N$  个顾客,其中一个顾客正在接受服务,其余的顾客( $\leq N-1$ )在等待队列等待服务。

5)排队系统的服务时间、到达间隔时间相互独立;排队系统的输入强度为  $\rho = \frac{1}{a\mu}$ 。

在上述条件下,我们利用辅助变量法和 Markov 嵌入链证明了以下定理。

**定理 1** 如果已知  $q_k (k=0, 1, \dots, N)$ , 则计算  $p_k$  的公式为

$$p_{k+1} = \frac{1}{\mu a_k} q_k, \quad k=0, 1, \dots, N-1 \quad (5)$$

$$p_0 = 1 - \sum_{k=1}^N p_k \quad (6)$$

其中  $a_k$  为分布函数  $A_k(t)$  的均值,  $1/\mu$  为服务时间的均值。

**注记** 如果顾客按几何分布随机离去,由于  $a_k, a$  与  $r_k$  之间具有如下关系( $k=0, 1, 2, \dots, N$ ):  $a_k = \frac{a}{r_k}$  (参见文[14]69

页); 因此,式(5)可改写为:  $p_{k+1} = \frac{r_k}{\mu a} q_k = \frac{1-d_k}{\mu a} q_k$ ; 此公式与文[10]的公式(12)一致(注:文[10]中的  $\lambda$  相当于本文中的  $\frac{1}{a}, b_k$  相当于本文中的  $r_k$ )。可见“输入流阻行”方法是“输入流稀疏化”方法当“顾客按几何分布随机离去”时的特例。

**定理 2**  $q_k (k=0, 1, \dots, N)$  可以通过求解以下线性方程组获得:

$$q_j = \sum_{i=0}^N q_i h_{i,j}, \quad j=0, 1, \dots, N \quad (7)$$

$$\sum_{j=0}^N q_j = 1 \quad (8)$$

其中,计算该线性方程组系数  $h_{i,j}$  的公式如下:

$$h_{i,j} = \int_0^{\infty} \frac{e^{-\mu t} (\mu t)^{i+1-j}}{(i+1-j)!} dA_i(t); \quad 0 \leq i \leq N-1, 1 \leq j \leq i+1 \quad (9)$$

$$h_{i,0} = 1 - \sum_{j=1}^{i+1} h_{i,j}; \quad 0 \leq i \leq N-1 \quad (10)$$

$$h_{N,j} = \int_0^{\infty} \frac{e^{-\mu t} (\mu t)^{N-j}}{(N-j)!} dA_N(t); \quad 1 \leq j \leq N \quad (11)$$

$$h_{N,0} = 1 - \sum_{j=1}^N h_{N,j}$$

### 2.2 基于输入流稀疏化的具有随机离去策略的排队系统 GI<sup>X</sup>/M/1/N<sup>[18]</sup>

一个基于输入流稀疏化的具有随机离去策略的排队系统 GI<sup>X</sup>/M/1/N 可表述为:

1)顾客成批到达排队系统且每批中顾客的个数是随机的。用随机变量  $X$  表示每批中顾客的个数,且  $X$  具有如下分布:  $\Pr(X=j) = g_j$ , 其中  $j$  是正整数;其数学期望记为  $E(X) = \bar{g}$ 。

2)各批次顾客依次在时刻  $\tau_0, \tau_1, \tau_2, \dots, \tau_n, \dots$ , 到达排队系统,其中  $\tau_0 < \tau_1 < \tau_2 < \dots < \tau_n < \dots$ , 且  $\tau_0 = 0$ 。到达间隔:  $e_i = \tau_i - \tau_{i-1} (i=1, 2, \dots, n, \dots)$  是独立同分布随机变量,其 df 为  $A(t)$ , pdf 为  $a(t)$ , 数学期望为  $a$ 。

3)一个服务台,服务时间服从参数为  $\mu$  的负指数分布,其 df 记为  $B(t) = 1 - \exp(-\mu t)$ 。服务规则采用先到先服务 FCFS。

4)排队系统最多能容纳  $N$  个顾客,其中一个顾客正在接受服务,其余的顾客( $\leq N-1$ )在等待队列等待服务。

5)顾客到达的间隔时间,服务台的服务时间和每批中顾客的个数相互独立。系统的输入强度为  $\rho = \frac{\bar{g}}{a\mu}$ 。

6)假设等待队列中当前有  $i (i=0, 1, \dots, N)$  个顾客,当一批新的顾客到达时,该批次顾客以概率  $r_i$  进入排队系统等待队列,或者说该批次顾客以概率  $d_i$  永远离开排队系统,其中  $r_i + d_i = 1$ 。此时,顾客成批到达等待队列的间隔时间的 df 为  $A_i(t) (i=0, 1, \dots, N)$ , pdf 为  $a_i(t) (i=0, 1, \dots, N)$ , 均值为  $a_i (i=0, 1, \dots, N), t \geq 0$ 。由于等待队列容量有限,当一批顾客进入等待队列时,如果等待队列的剩余位置不够容纳该批次的所有顾客,允许该批次的顾客进入等待队列的剩余位置,多余的顾客则永远离开。

在上述条件下,我们利用辅助变量法和 Markov 嵌入链证明了以下定理。

**定理 3** 如果已知  $q_i (i=0, 1, \dots, N)$ , 则计算  $p_k$  的公式为

$$p_{k+1} = \frac{1}{\mu} \sum_{i=0}^k \left( \frac{q_i}{a_i} \sum_{j=k-i+1}^{\infty} g_j \right), \quad k=0, 1, \dots, N-1 \quad (12)$$

$$p_0 = 1 - \sum_{k=1}^N p_k \quad (13)$$

其中  $a_i$  为分布函数  $A_i(t) (i=0, 1, \dots, N-1)$  的均值,  $1/\mu$  为服务时间的均值。

**注记**

1)如果顾客按几何分布随机离去,由于  $a_i, a$  与  $r_i$  之间具有如下关系( $i=0, 1, 2, \dots, N-1$ ):  $a_i = \frac{a}{r_i}$  (参见文[14]69页)。因此,式(12)可改写为

$$p_{k+1} = \frac{1}{\mu a} \sum_{i=0}^k (r_i q_i \sum_{j=k-i+1}^{\infty} g_j) = \frac{1}{\mu a} \sum_{i=0}^k ((1-d_i) q_i \sum_{j=k-i+1}^{\infty} g_j)$$

这与定理 5 的结论一致,表明“输入流阻行”方法是“输入流稀疏化”方法当“顾客按几何分布随机离去”时的特例。

2)当排队系统不具备随机离去策略时,即  $r_1 = r_2 = \dots = r_{N-1} = 1$  时,式(12)可化简为

$$p_{k+1} = \frac{1}{\mu a} \sum_{i=0}^k (q_i \sum_{j=k-i+1}^{\infty} g_j)$$

这就是当  $c=1$  时,文[13]的式(15)。

3)当顾客单个到达时,即  $\Pr(X=1)=1$  且  $\Pr(X>1)=0$  时,式(12)可化简为

$$p_{k+1} = \frac{q_k}{\mu a_k}$$

这就是定理 1 的结论,说明顾客单个到达是成批到达时的特例。

**定理 4**  $q_k (k=0, 1, \dots, N)$  可以通过求解以下线性方程组获得:

$$q_j = \sum_{i=0}^N q_i h_{i,j}, j=0, 1, \dots, N \quad (14)$$

$$\sum_{j=0}^N q_j = 1 \quad (15)$$

计算该线性方程组系数  $h_{i,j}$  的公式如下:

$$h_{i,j} = \begin{cases} \sum_{k=j-i}^{N-i} \beta_{i+k-j}^{(i)} g_k + \beta_{N-j}^{(i)} \sum_{k=N-i+1}^N g_k, & j>i \geq 0, j \geq 1 \\ \sum_{k=1}^{N-i} \beta_{i+k-j}^{(i)} g_k + \beta_{N-j}^{(i)} \sum_{k=N-i+1}^N g_k, & 1 \leq j \leq i \end{cases} \quad (16)$$

$$h_{i,0} = 1 - \sum_{j=1}^N h_{i,j} \quad (17)$$

其中

$$\beta_r^{(i)} = \begin{cases} \int_0^{\infty} \frac{e^{-\mu t} (\mu t)^r}{r!} dA_i(t), & r \geq 0 \\ 0, & r < 0 \end{cases} \quad (18)$$

### 2.3 基于输入流阻行的具有随机离去策略的排队系统 $GI^X/M/1/N$ <sup>[19]</sup>

一个基于输入流阻行的具有随机离去策略的排队系统  $GI^X/M/1/N$  可表述为:

1)顾客成批到达排队系统且每批中顾客的个数是随机的。用随机变量  $X$  表示每批中顾客的个数,且  $X$  具有如下分布:  $\Pr(X=j)=g_j$ , 其中  $j$  是正整数,其数学期望记为  $E(X)=\bar{g}$ 。

2)各批次顾客依次在时刻  $\tau_0, \tau_1, \tau_2, \dots, \tau_n, \dots$ , 到达排队系统,其中  $\tau_0 < \tau_1 < \tau_2 < \dots < \tau_n < \dots$ , 且  $\tau_0=0$ 。到达间隔:  $e_i = \tau_i - \tau_{i-1} (i=1, 2, \dots, n, \dots)$  是独立同分布随机变量,其 df 为  $A(t)$ , pdf 为  $a(t)$ , 数学期望为  $a$ 。

3)一个服务台,服务时间服从参数为  $\mu$  的负指数分布,其 df 记为  $B(t) = 1 - \exp(-\mu t)$ 。服务规则采用先到先服务 FCFS。

4)排队系统最多能容纳  $N$  个顾客,其中一个顾客正在接受服务,其余的顾客( $\leq N-1$ )在等待队列等待服务。

5)顾客到达的间隔时间,服务台的服务时间和每批次中顾客的个数相互独立。排队系统的输入强度为  $\rho = \frac{\bar{g}}{a\mu}$ 。

6)假设等待队列中当前有  $i (i=0, 1, \dots, N)$  个顾客,当一批新的顾客到达时,该批次顾客以概率  $r_i$  进入排队系统等待队列,或者说该批顾客以概率  $d_i$  永远离开排队系统,其中  $r_i + d_i = 1$ 。由于等待队列容量有限,当一批顾客进入等待队列时,如果等待队列的剩余空间不够容纳该批次的所有顾客,允许该批次的顾客进入等待队列的剩余空间,多余的顾客则永远离开。

在上述条件下,我们利用辅助变量法和 Markov 嵌入链证明了以下定理。

**定理 5** 如果已知  $q_i (i=0, 1, \dots, N)$ , 则计算  $p_k$  的公式为

$$p_{k+1} = \frac{1}{\mu a} \sum_{i=0}^k [r_i q_i \sum_{j=k+1-i}^{\infty} g_j] = \frac{1}{\mu a} \sum_{i=0}^k [(1-d_i) q_i \sum_{j=k+1-i}^{\infty} g_j], \quad 0 \leq k \leq N-1 \quad (19)$$

$$p_0 = 1 - \sum_{k=1}^N p_k \quad (20)$$

**注记**

1)当排队系统不具备随机离去策略时,即  $r_1 = r_2 = \dots = r_{N-1} = 1$  时,式(19)可化简为

$$p_{k+1} = \frac{1}{\mu a} \sum_{i=0}^k (q_i \sum_{j=k-i+1}^{\infty} g_j)$$

这就是当  $c=1$  时,文[13]的式(15)。

2)当顾客单个到达时,即  $\Pr(X=1)=1$  且  $\Pr(X>1)=0$  时,式(19)可化简为

$$p_{k+1} = \frac{r_k q_k}{\mu a}$$

这就是文[10]的式(12),说明顾客单个到达是成批到达时的特例。

**定理 6**  $q_k (k=0, 1, \dots, N)$  可以通过求解以下线性方程组获得:

$$q_j = \sum_{i=0}^N q_i h_{i,j}, j=0, 1, \dots, N \quad (21)$$

$$\sum_{j=0}^N q_j = 1 \quad (22)$$

计算该线性方程组系数  $h_{i,j}$  的公式如下:

$$h_{i,j} = \begin{cases} \sum_{k=j-i}^{N-i} \beta_{i+k-j} g_k + \beta_{N-j} \sum_{k=N-i+1}^N g_k, & j>i \geq 0, j \geq 1 \\ r_i [\sum_{k=1}^{N-i} \beta_{i+k-j} g_k + \beta_{N-j} \sum_{k=N-i+1}^N g_k] + d_i \beta_{i-j}, & 1 \leq j \leq i \end{cases} \quad (23)$$

$$h_{i,0} = 1 - \sum_{j=1}^N h_{i,j} \quad (24)$$

$$\beta_r = \begin{cases} \int_0^{\infty} \frac{e^{-\mu t} (\mu t)^r}{r!} dA(t), & r \geq 0 \\ 0, & r < 0 \end{cases} \quad (25)$$

本文中所讨论的 3 个排队系统以及文[10]讨论的排队系统之间的关系可用图 3 说明。方框表示相应的排队系统,方框之间的连线表示相应排队系统之间的联系。

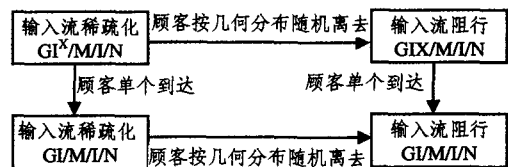


图 3 本文所讨论的排队系统之间的关系

### 2.4 评价排队系统性能的若干指标

一旦我们获得了排队系统队列长度的概率分布  $q_k (k=0, 1, \dots, N)$  和  $p_k (k=0, 1, \dots, N)$ ; 我们就能据此计算诸如顾客丢失率、系统利用率、队列长度的均值和方差、平均等待时间等排队系统的性能指标。具体的计算公式见我们已发表的文[17~19]。

## 3 应用讨论

Internet 是一种无连接的、尽力服务的分组交换网络,这种网络模型在提高网络资源利用率的同时,也不可避免会造成网络拥塞。当大量的分组(也称“IP 包”)被注入到网络并超过网络的处理能力时,网络性能会急剧下降,这种现象称为拥塞。拥塞会使路由器缓存队列溢出、IP 包大量丢弃、数据吞吐率下降,从而导致网络性能急剧下降。拥塞控制是指网络中的节点采取适当的措施尽量避免拥塞的发生或当拥塞发生时做出响应的过程。

(下转第 64 页)

- Flexible Web Services. IEEE MultiMedia, Jan 2001, 8(1): 58~65
- 18 Koppol P, Tai K. An Incremental Approach to Structural Testing of Concurrent Software. In: Proc. Int'l Symp Software Testing and Analysis, Jan 1996. 14~23
  - 19 Lee S, Offutt J. Generating Test Cases for XML-based Web Component Interactions Using Mutation Analysis. In: Proc. 12th IEEE Int'l Symp. Software Reliability Eng, Nov 2001. 200~209
  - 20 Lee T. World Wide Web Consortium. <http://www.w3.org/>, 2004
  - 21 Ricca F, Tonella P. Analysis and Testing of Web Applications. In: Proc. Int'l Conf Software Eng, May 2001. 25~34
  - 22 Macromedia Inc. Developing ColdFusion MX Applications. 2003. <http://www.macromedia.com/>
  - 23 Manley S, Seltzer M. Web Facts and Fantasy. In: Proc. 1997 Usenix Symp. Internet Technologies and Systems, 1997
  - 24 Nikora A, Munson J. Software Evolution and the Fault Process. In: Proc. 23rd Ann Software Eng Workshop, 1998
  - 25 Business Internet Group of San Francisco. The BIG-SF Report on Government Web Application Integrity. [http://www.tealeaf.com/downloads/news/analyst\\_report/BIG-SF\\_Report\\_Gov\\_2003-05.pdf](http://www.tealeaf.com/downloads/news/analyst_report/BIG-SF_Report_Gov_2003-05.pdf), 2004
  - 26 Business Internet Group of San Francisco. The Black Friday Report on Web Application Integrity. [http://www.tealeaf.com/downloads/news/analyst\\_report/BIG-SF\\_BlackFriday\\_Report.pdf](http://www.tealeaf.com/downloads/news/analyst_report/BIG-SF_BlackFriday_Report.pdf), 2004
  - 27 Parasoft. WebKing. <http://www.parasoft.com/jsp/products>, 2004
  - 28 Pressman R. Software Engineering: A Practitioner's Approach, Fifthed; McGraw-Hill, 2001
  - 29 Rational-Corporation. Rational Testing Robot. <http://www.rational.com/products/robot/>, 2004
  - 30 Testing a Website: Best practices. <http://www.reveregroup.com>, 2004
  - 31 Elbaum S, Karre S, Rothermel G. Improving Web Application Testing with User Session Data. In: Proc. Int'l Conf Software Eng, May 2003. 49~59
  - 32 Tai K C, Carver R H. A Specification-based Methodology for Testing Concurrent Programs. In: Proc. Fifth European Software Eng Conf, Sept 1995. 154~172
  - 33 Taylor R, Levine D, Kelly C. Structural Testing of Concurrent Programs. IEEE Trans Software Eng, 1992, 18(3): 206~215
  - 34 Tilley S, Shihong H. Evaluating the Reverse Engineering Capabilities of Web Tools for Understanding Site Content and Structure: A Case Study. In: Proc. Int'l Conf Software Eng, May 2001. 514~523
  - 35 Tzay J, Huang J, Wang F, et al. Constructing an Object-oriented Architecture for Web Application Testing. J Information Science and Eng, Jan 2002. 59~84

(上接第 40 页)

早期的 Internet 不具备拥塞控制机制, 从而产生“拥塞崩溃”。1989 年 10 月, Internet 发生了第一次拥塞崩溃, 导致从 LBL(Lawrence Berkeley Lab.) 到 UC Berkeley 的数据吞吐率从 32kbps 下降到 40bps。1988 年, Jacobson 基于网络包的守恒原理, 即“流入网络的包=流出网络的包”, 提出了端到端的 TCP 拥塞控制算法。然而仅有端到端的 TCP 拥塞控制算法, 还不能很好地解决 Internet 的拥塞问题。鉴于网络中间节点(路由器)具有及时检测拥塞是否发生的能力, 1993 年 Floyd 提出了一个用于网络中间节点的拥塞控制方法: RED 算法(Random Early Detection)。1998 年, IETF 在 RFC 2309 中建议在路由器中使用主动队列管理(Active Queue Management, 以下简称 AQM)算法, 并推荐 RED 算法为 AQM 的唯一候选算法。AQM 通过与 TCP 端到端拥塞控制算法相配合, 随机丢弃或标记欲进入路由器缓存队列的 IP 包, 从而实现拥塞控制。

Internet 业务流量由响应业务流量(即 TCP 长流 long lived TCP flows)和非响应业务流量(即 TCP 短流 short lived TCP flows 和 UDP 流)构成。由于非响应业务流量约占 Internet 业务流量的 70%~80%, 且不能响应 RED 算法的丢包信号, 因此本文提出的排队系统, 可用于评价 RED 算法在非响应业务流量下性能, 即分析 RED 算法的丢包率、链路利用率、队列长度的均值和方差等评价指标。详细的讨论请见我们已录用的论文(文[20])。

### 参 考 文 献

- 1 Kleinrock L. Queueing Systems. Volume I: Theory. New York: Wiley Interscience, 1975
- 2 Haight F A. Queueing with Balking. Biometrika, 1957, 44(3/4): 360~369
- 3 Haight F A. Queueing with Balking II. Biometrika, 1960, 47(3/4): 285~296
- 4 Ancker C J, Gafarian A V. Some Queueing Problems with Balking and Reneging, I. Operations Research, 1963, 11(1): 88~100
- 5 Ancker C J, Gafarian A V. Some Queueing Problems with Balking and Reneging, II. Operations Research, 1963, 11(6): 928~937
- 6 Rao S S. Queueing Models with Balking, Reneging, and Interruptions. Operations Research, 1965, 13(4): 596~608
- 7 Dreik A, Woolford D G. A Preemptive Priority Queue with Balking. European Journal of Operational Research, 2005, 164(2): 387~401
- 8 Younsuk K, Kiseon K. Loss Probability Behavior of Pareto/M/1/K Queue. IEEE Communication Letters, 2003, 7(1): 39~41
- 9 Cao J, Cleveland W S, Lin D, et al. The Effect of Statistical Multiplexing on the Long-Rang Dependence of Internet Packet Traffic; [Technical Report]. Bell Labs, 2001
- 10 Finch P D. Balking in the queueing system GI/M/1. Acta Math Acad Sci Hungar. 1959, 10: 241~247
- 11 Cox D R. The analysis of non-Markovian stochastic processes by the inclusion of supplementary variables. Proc. Cambridge Philos Soc., 1955, (51): 433~441
- 12 Hokstad P. The G/M/m Queue with Finite Waiting Room. Journal of Applied Probability, 1975, 12(4): 779~792
- 13 Vijaya P L, Gupta U C. Analysis of finite-buffer multi-server queues with group arrivals:  $G^X/M/c/N$ . Queueing System, 2000, 36(1/3): 125~140
- 14 Vladimir V K. Mathematical Methods in Queueing Theory. Dordrecht: Kluwer Academic, 1994
- 15 Cao J, Cleveland W S, Lin Dong, et al. On the nonstationarity of Internet traffic. In: Proc. SIGMETRICS'2001, ACM, 2001, 29(1): 102~112
- 16 Tudjarov A, Temkov D, Janevski T, et al. Empirical modeling of Internet traffic at middle-level burstiness. In: Proc. 12th IEEE Mediterranean, IEEE, 2004. 535~538
- 17 汪浩, 李晓明. 一个具有随机丢弃分组机制的 GI/M/1/N 排队系统. 通信学报, 2006, 27(1): 14~20
- 18 汪浩. 一个具有随机丢弃分组机制且分组成批到达的  $G^X/M/1/N$  排队系统. 数学的实践与认识, 2005, 35(9): 113~120
- 19 汪浩, 李晓明, 严伟. 一个具有阻行机制的成批到达排队系统  $G^X/M/1/N$ . 北京大学学报, 2006, 42(3): 406~411
- 20 汪浩, 严伟. 基于自相似聚合业务流量的 AQM 算法性能评价. 软件学报(已录用)