高负荷下分支型轮询排队网络的极限性态

刘再明, 储育青*, 吴锦标

中南大学数学与统计学院, 长沙 410083

E-mail: math_lzm@csu.edu.cn, chuyuqing@csu.edu.cn, wujinbiao@csu.edu.cn

收稿日期: 2015-01-25; 接受日期: 2015-03-25; * 通信作者

国家自然科学基金 (批准号: 11271373 和 11201489) 资助项目

摘要 本文研究服务策略具有一般分支性的轮询排队网络. 所有队列共享一个服务器, 且服务器依次服务每个队列. 队列 i 中的顾客服务完后, 以概率 $p_{i,j}$ 转移到队列 j , 或以概率 $p_{i,0}$ 离开系统. 利用带移民的多型分支过程的极限定理, 本文给出高负荷下服务器进入队列服务时队长联合分布的极限性态; 结合平均准则方法, 还得到平均等待时间的极限定理.

关键词 轮询排队网络 多型分支过程 极限性态 后代集方法

MSC (2010) 主题分类 60K25, 60K20

1 引言

经典轮询排队系统由多个队列和一个服务器组成, 服务器按顺序依次服务每个队列. 近年来, 轮询排队系统在计算机通信、交通运输和制造系统等领域中的广泛应用, 使得轮询排队日益受到学者们的关注. 迄今为止, 已经有很多关于轮询排队系统的论文和专著. 基于我们以前对轮询排队系统的研究 (参见文献 [1]), 本文将研究一类服务策略具有分支性的轮询排队网络.

本文研究的轮询排队网络中, 所有队列共享一个服务器, 服务器按顺序依次服务每个队列. 队列 i 中的顾客服务完后以概率 $p_{i,j}$ 转移到队列 j , 或者以概率 $p_{i,0}$ 离开系统. 这类轮询排队网络在计算机网络和生产制造等系统中都有广泛的应用, 例如, 在 FWLAN (ferry based wireless local area network) 系统中, 同一个区域内的一些节点与中心点间通过一个信息轮渡接收和发送信息; 又如对汽车保险杠上色的 BumperCo 系统过程和小区内邮递员收发邮件的过程等, 参见文献 [2, 3].

由于每类顾客的总到达过程不是 Poisson 过程, 所以, 由 Little 公式求等待时间的传统方法已不适用于此系统. 2013 年, 利用多分支性和后代集方法 (descendant set approach), Boon 等人^[4] 首次给出了内部顾客和外部顾客等待时间的 LST (Laplace-Stieltjes transform), 而在此之前, 学者们只给出了队长和平均等待时间等性能指标, 如文献 [5–7]. 但是, 文献 [4] 中给出的等待时间的计算相当复杂, 因此, 我们希望通过文献 [8] 中的拟合方法, 利用低负荷和高负荷下等待时间的极限分布函数, 给出等待时间在任意负荷下的近似分布函数, 而本文则重点研究系统在高负荷下的一些极限性态.

van der Mei^[9] 研究了服务策略具有一般分支性的经典轮询排队系统在高负荷下的极限性态; Boon 等人^[5] 也曾给出门限服务策略下队长的高负荷极限性态. 本文利用带移民的多型分支过程在 $\rho \uparrow 1$ 时的极限定理, 考虑分支型轮询排队网络的高负荷极限性态, 文献 [5, 9] 中的结果是我们的特殊情形.

英文引用格式: Liu Z M, Chu Y Q, Wu J B. The asymptotic behavior of a branching-type polling network in heavy traffic (in Chinese). Sci Sin Math, 2015, 45: 515–526, doi: 10.1360/N012015-00044

2 模型描述

考虑一个由 N 个队列 $Q_1, \dots, Q_N$ 和一个服务器组成的排队网络, 服务器依次服务每个队列. 外部顾客输入 Q_i 是参数为 λ_i 的 Poisson 过程. Q_i 中顾客的服务时间 B_i 服从一般分布. 记

$$\Lambda = \sum_{i=1}^N \lambda_i, \quad b_i = \mathbb{E}B_i, \quad b_i^{(2)} = \mathbb{E}B_i^2.$$

服务器从队列 Q_i 转移到队列 Q_{i+1} 的转移时间 R_i 服从一般分布. 记

$$r_i = \mathbb{E}R_i, \quad r_i^{(2)} = \mathbb{E}R_i^2, \quad r = \sum_{i=1}^N r_i.$$

假设顾客的到达过程、每个顾客的服务时间以及服务器的转移时间相互独立. 队列 Q_i 中的顾客服务完后以概率 $p_{i,j}$ 转移到队列 Q_j 的尾部, 或以概率 $p_{i,0}$ 离开系统, 其中

$$\sum_{i=1}^N p_{i,0} > 0, \quad \sum_{j=0}^N p_{i,j} = 1,$$

且顾客转移时间忽略不计. 假设每个队列的服务策略具有分支性, 这里, 分支性定义如下:

定义 1 ^[10] 当服务器从其他队列转移到 Q_i 时, 发现 Q_i 中有 k_i 个顾客, 服务器在 Q_i 中服务直到离开时, 如果这 k_i 个顾客均将独立同分布地被 N 类顾客有效取代, 这 N 类顾客数联合分布的概率母函数为 $h_i(\mathbf{z}) = h_i(z_1, \dots, z_N)$, 则称 Q_i 中的服务策略满足分支性.

由文献 [4] 知, 队列 Q_i 的“有效到达率” γ_i 是下述方程组的唯一非负解:

$$\gamma_i = \lambda_i + \sum_{j=1}^N \gamma_j p_{j,i}, \quad i = 1, \dots, N.$$

定义队列 Q_i 的负荷率为 $\rho_i = \gamma_i b_i$, 系统总的负荷率为

$$\rho = \sum_{i=1}^N \rho_i.$$

假设系统稳定, 即 $\rho < 1$ (参见文献 [11]). 记 $\bar{B}_i$ 表示队列 Q_i 中的顾客在系统逗留期间接受的总服务时间, 并引入随机变量 X_i ($i = 1, \dots, N$):

$$X_i = \begin{cases} j, & \text{若 } Q_i \text{ 中的顾客服务完后转移到 } Q_j, \quad j = 1, \dots, N, \\ 0, & \text{若 } Q_i \text{ 中的顾客服务完后离开系统,} \end{cases}$$

则有 $\mathbb{P}\{X_i = j\} = p_{i,j}$. 由全概率公式有

$$\begin{aligned} \mathbb{E}\bar{B}_i &= \mathbb{E}[\bar{B}_i | X_i = 0] \mathbb{P}\{X_i = 0\} + \sum_{j=1}^N \mathbb{E}[\bar{B}_i | X_i = j] \mathbb{P}\{X_i = j\} \\ &= p_{i,0} \mathbb{E}B_i + \sum_{j=1}^N p_{i,j} \mathbb{E}[B_i + \bar{B}_j] = \mathbb{E}B_i + \sum_{j=1}^N p_{i,j} \mathbb{E}\bar{B}_j, \end{aligned} \quad (2.1)$$

$$\begin{aligned}
\mathbb{E}\bar{B}_i^2 &= \mathbb{E}[\bar{B}_i^2 | X_i = 0]\mathbb{P}\{X_i = 0\} + \sum_{j=1}^N \mathbb{E}[\bar{B}_i^2 | X_i = j]\mathbb{P}\{X_i = j\} \\
&= p_{i,0}\mathbb{E}B_i^2 + \sum_{j=1}^N p_{i,j}(\mathbb{E}B_i^2 + 2\mathbb{E}[B_i\bar{B}_j] + \mathbb{E}\bar{B}_j^2) \\
&= \mathbb{E}B_i^2 + 2\sum_{j=1}^N p_{i,j}\mathbb{E}B_i\mathbb{E}\bar{B}_j + \sum_{j=1}^N p_{i,j}\mathbb{E}\bar{B}_j^2,
\end{aligned} \tag{2.2}$$

(2.2) 中最后一个等号利用了顾客服务时间的相互独立性. 记

$$\bar{b}_i = \mathbb{E}\bar{B}_i, \quad \bar{b}_i^{(2)} = \mathbb{E}\bar{B}_i^2, \quad \bar{b} = \sum_{i=1}^N \frac{\lambda_i}{\Lambda} \bar{b}_i, \quad \bar{b}^{(2)} = \sum_{i=1}^N \frac{\lambda_i}{\Lambda} \bar{b}_i^{(2)}.$$

由 (2.1) 和 (2.2) 得 $\rho = \sum_{i=1}^N \lambda_i \bar{b}_i$, 且 $\bar{b}_i$ 和 $\bar{b}_i^{(2)}$ 满足下面的方程:

$$\begin{aligned}
\bar{b}_i &= b_i + \sum_{j=1}^N p_{i,j} \bar{b}_j, \\
\bar{b}_i^{(2)} &= b_i^{(2)} + 2 \sum_{j=1}^N p_{i,j} b_i \bar{b}_j + \sum_{j=1}^N p_{i,j} \bar{b}_j^{(2)}.
\end{aligned}$$

为方便起见, 我们规定: (1) $X^*(\cdot)$ 表示随机变量 X 的 LST; (2) 对于任意函数 $f(x)$, 记 $\hat{f} = f(1)$.

3 带移民的多型分支过程

我们考虑一个带有 N 类粒子的多型分支过程 $\mathbf{Z} = \{\mathbf{Z}_n, n = 0, 1, \dots\}$, 其中 $\mathbf{Z}_n = (Z_n^{(1)}, \dots, Z_n^{(N)})$, 这里 $Z_n^{(i)}$ 表示第 n 代 i 型粒子的个数. i 型粒子产生下一代粒子数的联合分布的概率母函数记为

$$f^{(i)}(\mathbf{z}) = \sum_{j_1, \dots, j_N \geq 0} p^{(i)}(j_1, \dots, j_N) z_1^{j_1} \cdots z_N^{j_N}, \quad |z_k| \leq 1, \quad i, k = 1, \dots, N,$$

其中 $\mathbf{z} = (z_1, \dots, z_N)$, $p^{(i)}(j_1, \dots, j_N)$ 表示一个 i 型粒子产生 j_k 个 k ($k = 1, 2, \dots, N$) 型粒子的概率. 移民分布的概率母函数记为

$$g(\mathbf{z}) = \sum_{j_1, \dots, j_N \geq 0} q(j_1, \dots, j_N) z_1^{j_1} \cdots z_N^{j_N}, \quad |z_k| \leq 1, \quad k = 1, \dots, N,$$

其中 $q(j_1, \dots, j_N)$ 表示移民由 j_k 个 k ($k = 1, 2, \dots, N$) 型粒子组成的概率. 为方便起见, 我们引入以下记号:

$$\begin{aligned}
\mathbf{g} &= (g_1, \dots, g_N)^T, \quad \text{其中 } g_i = \left. \frac{\partial g(\mathbf{z})}{\partial z_i} \right|_{\mathbf{z}=\mathbf{1}}, \quad \mathbf{1} = (1, \dots, 1), \\
\mathbf{M} &= (m_{i,j}), \quad \text{其中 } m_{i,j} = \left. \frac{\partial f^{(i)}(\mathbf{z})}{\partial z_j} \right|_{\mathbf{z}=\mathbf{1}}, \quad i, j = 1, \dots, N, \\
\mathbf{K}^{(i)} &= (K_{j,k}^{(i)}), \quad \text{其中 } K_{j,k}^{(i)} = \left. \frac{\partial^2 f^{(i)}(\mathbf{z})}{\partial z_j \partial z_k} \right|_{\mathbf{z}=\mathbf{1}}, \quad i, j, k = 1, \dots, N.
\end{aligned}$$

记 $\mathbf{M}$ 的最大非负实值特征值为 θ (简称最大特征值), $\mathbf{v} = (v_1, \dots, v_N)^T$ 和 $\mathbf{w} = (w_1, \dots, w_N)^T$ 分别是 θ 对应的标准化左特征向量和右特征向量. 定义函数 $\pi_n(x)$ 如下:

$$\pi_0(x) = 0, \quad \pi_n(x) = \sum_{i=1}^n x^{i-2}, \quad x > 0, \quad n = 1, 2, \dots$$

由文献 [11], 若均值矩阵 $\mathbf{M}$ 是一个素阵 (即存在整数 $k > 0$ 使得 $\mathbf{M}^k > \mathbf{0}$), 最大特征值 $\theta < 1$, 且 Markov 链 $\mathbf{Z}$ 是遍历的, 则带移民的分支过程 $\mathbf{Z}$ 存在稳态分布的充要条件为

$$\sum_{\substack{j_1, \dots, j_N \geq 0 \\ j_1 + \dots + j_N > 0}} q(j_1, \dots, j_N) \log(j_1 + \dots + j_N) < \infty.$$

下面给出当 $\theta \uparrow 1$ 时, 分支过程 $\mathbf{Z}$ 的极限定理.

定理 1 假设 $\mathbf{f}(\mathbf{z}) = (f^{(1)}(\mathbf{z}), \dots, f^{(N)}(\mathbf{z}))^T$ 在点 $\mathbf{z} = \mathbf{1}$ 处二阶导数存在, 且 $0 < g_i < \infty$ ($i = 1, \dots, N$), 则

$$\lim_{n \rightarrow \infty} \frac{1}{\pi_n(\theta)} \begin{pmatrix} Z_n^{(1)} \\ \vdots \\ Z_n^{(N)} \end{pmatrix} \rightarrow_d A \begin{pmatrix} \hat{v}_1 \\ \vdots \\ \hat{v}_N \end{pmatrix} \Gamma(\alpha, 1), \quad \theta \uparrow 1,$$

其中

$$\alpha = \frac{1}{A} \hat{\mathbf{g}}^T \hat{\mathbf{w}} = \frac{1}{A} \sum_{i=1}^N \hat{g}_i \hat{w}_i, \quad A = \sum_{i=1}^N \hat{v}_i (\hat{\mathbf{w}}^T \hat{\mathbf{K}}^{(i)} \hat{\mathbf{w}}). \quad (3.1)$$

定理 1 的证明可参见文献 [11]. 接下来, 我们将利用这一极限定理给出分支型轮询排队网络在高负荷下队长的极限性态.

4 高负荷下队长的极限性态

由分支性的定义, 服务器在队列 Q_i 服务期间, 这 k_i 个顾客均会引出一个独立同分布的“子忙期”, 记为 T_i . 令 M_i 为在 T_i 期间服务完离开 Q_i 的顾客数, L_i 为产生子忙期 T_i 的原始顾客在子忙期 T_i 结束时其后代中的 i 类顾客数, 则有

$$\mathbb{E}L_i = 1 + \lambda_i \mathbb{E}T_i - \mathbb{E}M_i. \quad (4.1)$$

为进一步刻画这一过程, 我们引入下面的联合 PGF-LST 函数:

$$\psi_i(u, z_1, z_2) = \mathbb{E}[e^{-uT_i} z_1^{L_i} z_2^{M_i}], \quad \operatorname{Re}(u) > 0, \quad |z_1| \leq 1, \quad |z_2| \leq 1, \quad i = 1, \dots, N.$$

记 B_i^E 为 Q_i 中的顾客在转移到其他队列或离开系统之前接受的总服务时间, Y_i 为 Q_i 中顾客在转移到其他队列或离开系统之前总的服务次数, 显然, $\mathbb{P}\{Y_i = n\} = p_{i,i}^{n-1}(1 - p_{i,i})$, $n = 1, 2, \dots$, 且有

$$b_i^E = \mathbb{E}B_i^E = \mathbb{E}\left[\sum_{j=1}^{Y_i} B_{i,j}\right] = \mathbb{E}Y_i \mathbb{E}B_i = \frac{b_i}{1 - p_{i,i}},$$

其中 $B_{i,j}$ ($j = 1, 2, \dots$) 是一列与 B_i 独立同分布的随机变量. 假设不考虑转移时间时, 系统具有工作量保守性, 即只要系统有顾客, 服务器就一定忙, 则有 $\mathbb{E}T_i = b_i^E \mathbb{E}M_i$. 结合 (4.1), 得到

$$\mathbb{E}T_i = b_i^E \mathbb{E}M_i = (1 - \mathbb{E}L_i) \frac{b_i^E}{1 - \lambda_i b_i^E}. \quad (4.2)$$

令 $X_{n,j}^{(i)}$ 表示服务器第 n 次进入 Q_i 时 Q_j 中的顾客数, $X_j^{(i)}$ 表示稳态下服务器进入 Q_i 时 Q_j 中的顾客数, 则过程 $\mathbf{X} = \{(X_{n,1}^{(1)}, \dots, X_{n,N}^{(1)}), n = 0, 1, \dots\}$ 是一个带移民的多型分支过程. i 型粒子产生下一代粒子数的联合分布的概率母函数 $f^{(i)}(\mathbf{z})$ 为

$$f^{(i)}(\mathbf{z}) = h_i(z_1, z_2, \dots, z_i, f^{(i+1)}(\mathbf{z}), \dots, f^{(N)}(\mathbf{z})), \quad |z_k| \leq 1, \quad i = 1, \dots, N,$$

其中

$$h_i(\mathbf{z}) = \psi_i\left(\sum_{j \neq i} \lambda_j(1 - z_j), z_i, \frac{p_{i,0}}{1 - p_{i,i}} + \sum_{j \neq i} \frac{p_{i,j}}{1 - p_{i,i}} z_j\right), \quad i = 1, \dots, N.$$

移民分布的概率母函数 $g(\mathbf{z})$ 为

$$g(\mathbf{z}) = \prod_{i=1}^N R_i^* \left(\sum_{k=1}^i \lambda_k(1 - z_k) + \sum_{k=i+1}^N \lambda_k(1 - f^{(k)}(\mathbf{z})) \right), \quad |z_k| \leq 1.$$

定义队列 Q_i 的空竭率 f_i 为

$$f_i = 1 - \frac{\partial h_i(\mathbf{z})}{\partial z_i} \Big|_{\mathbf{z}=\mathbf{1}} = 1 - \mathbb{E}L_i, \quad i = 1, \dots, N.$$

令 $\varphi_i = \frac{b_i^E}{1 - \lambda_i b_i^E}$, 由 (4.2) 得

$$\mathbb{E}T_i = b_i^E \mathbb{E}M_i = (1 - \mathbb{E}L_i)\varphi_i = f_i \varphi_i.$$

注 1 令 $E = \{i : Q_i \text{ 服从空竭服务策略}\}$, $G = \{i : Q_i \text{ 服从门限服务策略}\}$, 我们有

$$\psi_i(u, z_1, z_2) = B_i^*(u + \lambda_i(1 - z_1))[p_{i,i}z_1 + (1 - p_{i,i})z_2], \quad i \in G,$$

$$\psi_i(u, z_1, z_2) = z_2 B_i^{E*}(u + \lambda_i(1 - \psi_i(u, 1, z_2))), \quad i \in E.$$

若 $i \in G$, 则 $f_i = 1 - p_{i,i} - \rho_i$; 若 $i \in E$, 则 $f_i = 1$.

引理 1 对于分支型轮询排队网络, 粒子生成的均值矩阵 $\mathbf{M}$ 有如下分解:

$$\mathbf{M} = \mathbf{M}_1 \cdots \mathbf{M}_N, \quad (4.3)$$

其中 $\mathbf{M}_k = (m_{i,j}^{(k)})$ 且

$$m_{i,j}^{(k)} = \begin{cases} I_{\{i=j\}}, & i \neq k, \\ 1 - f_k, & i = k = j, \\ f_k \varphi_k \left(\lambda_j + \frac{p_{i,j}}{b_i} \right), & i = k \neq j, \end{cases}$$

这里 I_F 表示事件 F 上的示性函数. 如果我们定义

$$u_i = \frac{(1 - f_i)\gamma_i(1 - p_{i,i} - \rho_i)}{f_i} + \sum_{k=i+1}^N \rho_k \left(\lambda_i + \frac{p_{k,i}}{b_k} \right), \quad (4.4)$$

则 $\hat{\mathbf{M}}$ 的标准化左特征向量和右特征向量分别为

$$\hat{\mathbf{w}} = |\bar{\mathbf{b}}|^{-1} \begin{pmatrix} \bar{b}_1 \\ \vdots \\ \bar{b}_N \end{pmatrix}, \quad \hat{\mathbf{v}} = \frac{|\bar{\mathbf{b}}|}{\eta} \begin{pmatrix} \hat{u}_1 \\ \vdots \\ \hat{u}_N \end{pmatrix}, \quad (4.5)$$

其中

$$|\bar{\mathbf{b}}| = \sum_{k=1}^N \bar{b}_k, \quad \eta = |\bar{\mathbf{b}}| \hat{\mathbf{u}}^T \hat{\mathbf{w}}.$$

证明 由均值矩阵 $\hat{\mathbf{M}}$ 的定义可知,

$$m_{i,j} = h_{i,j} I_{\{j \leq i\}} + \sum_{k=i+1}^N h_{i,k} m_{k,j}, \quad i, j = 1, \dots, N, \quad (4.6)$$

$$h_{i,j} = \left. \frac{\partial h_i(\mathbf{z})}{\partial z_j} \right|_{\mathbf{z}=\mathbf{1}} = \begin{cases} 1 - f_i, & i = j, \\ f_k \varphi_k \left(\lambda_j + \frac{p_{i,j}}{b_i} \right), & i \neq j. \end{cases} \quad i, j = 1, \dots, N. \quad (4.7)$$

由 (4.6) 和 (4.7) 可以得到 (4.3).

将 $\hat{\mathbf{w}}$ 代入 (4.3) 中计算可得 $\hat{\mathbf{M}}_i \hat{\mathbf{w}} = \hat{\mathbf{w}}$, 因而有 $\hat{\mathbf{M}} \hat{\mathbf{w}} = \hat{\mathbf{w}}$, 又 $|\hat{\mathbf{w}}| = 1$, 故 $\hat{\mathbf{w}}$ 是 $\hat{\mathbf{M}}$ 的标准化右特征向量. 下面证明 $\hat{\mathbf{v}}$ 是 $\hat{\mathbf{M}}$ 的标准化左特征向量即可. 首先定义向量 $\hat{\mathbf{u}}^{(k)} = (\hat{u}_1^{(k)}, \dots, \hat{u}_N^{(k)})^T$, 其中

$$\hat{u}_i^{(k)} = \begin{cases} \frac{(1 - f_i) \gamma_i (1 - p_{i,i} - \rho_i)}{f_i} + \sum_{j=i+1}^{k-1} \rho_j \left(\lambda_i + \frac{p_{j,i}}{b_j} \right), & i \neq k-1, \\ \frac{(1 - f_i) \gamma_i (1 - p_{i,i} - \rho_i)}{f_i}, & i = k-1. \end{cases}$$

当 $k = 1, \dots, N$ 时, 可以证明 $\hat{\mathbf{u}}^{(k+1)} = \hat{\mathbf{M}}_k^T \hat{\mathbf{u}}^{(k)}$ 且 $\hat{\mathbf{u}}^{(1)} = \hat{\mathbf{u}}^{(N+1)}$. 由此, 我们得

$$\hat{\mathbf{u}}^{(1)} = \hat{\mathbf{u}}^{(N+1)} = \hat{\mathbf{M}}_N^T \hat{\mathbf{u}}^{(N)} = \hat{\mathbf{M}}_N^T \dots \hat{\mathbf{M}}_1^T \hat{\mathbf{u}}^{(1)} = \hat{\mathbf{M}}^T \hat{\mathbf{u}}^{(1)},$$

即 $\hat{\mathbf{u}}$ 是 $\hat{\mathbf{M}}$ 的左特征向量, 将其标准化后即得到 $\hat{\mathbf{v}}$ 是 $\hat{\mathbf{M}}$ 的标准化左特征向量. 证毕. $\square$

注 2 对于工作量保守的排队系统, $\hat{\mathbf{M}}$ 的右特征向量 $\hat{\mathbf{w}}$ 体现了工作量保守的思想.

将引理 1 中的 $\hat{\mathbf{w}}$ 和 $\hat{\mathbf{v}}$ 代入 (3.1) 计算得下面的引理.

引理 2 对于一般分支型轮询排队网络, 有

$$\hat{\mathbf{g}}^T \hat{\mathbf{w}} = |\bar{\mathbf{b}}|^{-1} r, \quad A = |\bar{\mathbf{b}}|^{-1} \eta^{-1} \frac{\bar{b}^{(2)}}{2b}. \quad (4.8)$$

引理 3 对于一般分支型轮询排队网络, 最大特征值 $\theta = \theta(\rho)$ 有下面的性质:

- (1) $\theta < 1$ 当且仅当 $\rho < 1$, $\theta = 1$ 当且仅当 $\rho = 1$, $\theta > 1$ 当且仅当 $\rho > 1$;
- (2) $\theta(\rho)$ 是 ρ 的连续函数;
- (3) $\lim_{\rho \uparrow 1} \theta(\rho) = \theta(1) = 1$;
- (4) $\theta(\rho)$ 在 $\rho = 1$ 处的导数为

$$\theta'(1) = \lim_{\rho \uparrow 1} \frac{1 - \theta(\rho)}{1 - \rho} = \frac{1}{\eta}.$$

证明 显然性质 (2) 成立, 性质 (3) 可由性质 (1) 得到, 故这里只给出性质 (1) 和 (4) 的证明. 要证明性质 (1), 只需证明以下三条:

- (a) $\rho < 1 \Rightarrow \mathbf{H} \hat{\mathbf{w}} < \hat{\mathbf{w}} \Rightarrow \mathbf{M} \hat{\mathbf{w}} < \hat{\mathbf{w}} \Rightarrow \theta < 1$;
- (b) $\rho = 1 \Rightarrow \mathbf{H} \hat{\mathbf{w}} = \hat{\mathbf{w}} \Rightarrow \mathbf{M} \hat{\mathbf{w}} = \hat{\mathbf{w}} \Rightarrow \theta = 1$;
- (c) $\rho > 1 \Rightarrow \mathbf{H} \hat{\mathbf{w}} > \hat{\mathbf{w}} \Rightarrow \mathbf{M} \hat{\mathbf{w}} > \hat{\mathbf{w}} \Rightarrow \theta > 1$,

其中 $\mathbf{H} = (h_{i,j})_{N \times N}$. 将 (4.6) 和 (4.7) 代入计算即可得到 (a) 中的第一和二推导符, 而第三个推导符可参见文献 [12]. (b) 和 (c) 的证明同 (a).

接下来, 我们用新后代集方法 (new descendent set approach) (参见文献 [13]) 证明性质 (4). 利用时间逆向递推定义服务器进入 Q_i 的时刻 $P_{i,c}$ ($i = 1, \dots, N, c = 0, 1, 2, \dots$): 取服务器任意一次进入队列 Q_1 服务的时刻为基准点, 记为 $P_{1,0}$; 服务器在基准点后首次进入队列 Q_i 的时刻记为 $P_{i,0}$ ($i = 2, \dots, N$); $P_{i,c}$ ($c = 1, 2, \dots$) 是服务器在时刻 $P_{i,c-1}$ 之前最后一次进入队列 Q_i 的时刻. 新后代集方法主要用来研究服务器进入队列 Q_k 服务时的队长 $X_k^{(k)}$ ($k = 1, \dots, N$). 令 $A_{(i,c),k}$ 为时刻 $P_{i,c}$ 时的一个 i 类顾客在时刻 $P_{k,0}$ 时其后代中 k 类顾客的个数, 记 $\alpha_{(i,c),k} = \mathbb{E}A_{(i,c),k}$. 当 $c \geq 1, i = 1, \dots, N$ 或 $c = 0, i = 1, \dots, k-1$ 时, 通过对紧邻的母代进行迭代, 可以得到下面的迭代方程:

$$\begin{aligned} \alpha_{(i,c),k} = & \sum_{j=1}^{k-1} \left(\lambda_k \mathbb{E}T_j + \frac{p_{j,k}}{1-p_{j,j}} \mathbb{E}M_j \right) \alpha_{(i,c),j} \\ & + \sum_{j=k+1}^N \left(\lambda_k \mathbb{E}T_j + \frac{p_{j,k}}{1-p_{j,j}} \mathbb{E}M_j \right) \alpha_{(i,c-1),j} + \mathbb{E}L_k \alpha_{(i,c-1),k}, \end{aligned} \quad (4.9)$$

其边界条件为

$$\begin{aligned} \alpha_{(k,0),k} &= 1; \\ \alpha_{(i,0),k} &= 0, \quad i = k+1, \dots, N; \\ \alpha_{(i,-1),k} &= 0, \quad i = 1, \dots, k-1. \end{aligned}$$

当 $\rho < 1$ 时, 记

$$\Delta = \sum_{i=1}^N \lambda_i \sum_{c=1}^{\infty} \alpha_{(i,c),1}, \quad (4.10)$$

$$\Delta_k = \sum_{i=1}^N \lambda_i \sum_{c=0}^{\infty} \alpha_{(i,c),k}, \quad k = 1, \dots, N, \quad (4.11)$$

则有 $\Delta = \Delta_1 - \lambda_1$. 将 (4.9) 代入 (4.11) 得

$$\begin{aligned} \Delta_k &= \sum_{i=1}^N \lambda_i \sum_{c=1}^{\infty} \alpha_{(i,c),k} + \sum_{i=1}^{k-1} \lambda_i \alpha_{(i,0),k} + \lambda_k \\ &= \sum_{j \neq k} \Delta_j \left(\lambda_k \mathbb{E}T_j + \frac{p_{j,k}}{1-p_{j,j}} \mathbb{E}M_j \right) + \lambda_k + \Delta_k \mathbb{E}L_k \\ &= \sum_{j \neq k} \left(\lambda_k + \frac{p_{j,k}}{b_j} \right) \Delta_j \mathbb{E}T_j + \lambda_k + \Delta_k \mathbb{E}L_k, \quad k = 1, \dots, N. \end{aligned}$$

进一步化简得

$$\frac{1}{\lambda_k} \left[\frac{1}{b_k} \Delta_k \mathbb{E}T_k - \sum_{j=1}^N \frac{p_{j,k}}{b_j} \Delta_j \mathbb{E}T_j \right] = 1 + \sum_{j=1}^N \mathbb{E}\Delta_j T_j, \quad k = 1, \dots, N. \quad (4.12)$$

解方程 (4.12) 得

$$\Delta_k \mathbb{E}T_k = \frac{\rho_k}{1-\rho}, \quad k = 1, \dots, N.$$

从而有

$$\Delta_k = \frac{\gamma_k(1 - p_{k,k} - \rho_k)}{f_k(1 - \rho)} = \hat{u}_k.$$

故

$$\Delta = \hat{u}_1 - \lambda_1. \quad (4.13)$$

类似新后代集方法, 在后代集方法中, 通过对紧邻的子代迭代得

$$\begin{aligned} \alpha_{(i,c),1} &= \sum_{j=1}^{i-1} \left(\lambda_j \mathbb{E} T_i + \frac{p_{i,j}}{1 - p_{i,i}} \mathbb{E} M_i \right) \alpha_{(j,c-1),1} + \sum_{j=k+1}^N \left(\lambda_j \mathbb{E} T_i + \frac{p_{i,j}}{1 - p_{i,i}} \mathbb{E} M_i \right) \alpha_{(j,c),1} \\ &\quad + \mathbb{E} L_i \alpha_{(i,c-1),1}. \end{aligned} \quad (4.14)$$

由 (4.14) 得 $\alpha_{(i,c),1} = \mathbf{e}_i^T \mathbf{M}^c \mathbf{e}_1$, 其中 $\mathbf{e}_i$ 是第 i 个分量为 1 的单位向量. 由文献 [13, 引理 4.1] 知,

$$\mathbf{M}^c = \theta^c \mathbf{v} \mathbf{w}^T + \mathbf{S}^{(c)}, \quad c = 1, 2, \dots,$$

其中 $S^{(c)} = (s_{i,j}^{(c)})$, 且存在 K ($0 < K < \infty$) 和 θ_* ($0 < \theta_* < \theta$) 使得 $|s_{i,j}^{(c)}| < K\theta_*^c$. 因此, 当 $\rho < 1$ 时, $\alpha_{(i,c),1}$ 可以分解为

$$\alpha_{(i,c),1} = \theta^c w_i v_1 + s_{i,1}^{(c)}, \quad c = 1, 2, \dots \quad (4.15)$$

将 (4.15) 代入 (4.10) 得

$$\Delta = \frac{\theta}{1 - \theta} v_1 \sum_{i=1}^N \lambda_i w_i + \sum_{i=1}^N \lambda_i \sum_{c=1}^N s_{i,1}^{(c)}. \quad (4.16)$$

在方程 (4.13) 和 (4.16) 两端分别乘以因子 $1 - \rho$, 然后令 $\rho \uparrow 1$, 化简得

$$\lim_{\rho \uparrow 1} \frac{1 - \theta(\rho)}{1 - \rho} = \frac{1}{\eta}.$$

证毕. □

定理 2 对于分支型轮询排队网络, 服务器进入队列 Q_1 时的联合队长有下面的极限性态:

$$(1 - \rho) \begin{pmatrix} X_1^{(1)} \\ \vdots \\ X_N^{(1)} \end{pmatrix} \rightarrow_d \frac{\bar{b}^{(2)}}{2\bar{b}} \frac{1}{\eta} \begin{pmatrix} \hat{u}_1 \\ \vdots \\ \hat{u}_N \end{pmatrix} \Gamma(\alpha, 1), \quad \rho \uparrow 1, \quad (4.17)$$

其中 $\alpha = 2r \frac{\bar{b}}{\bar{b}^{(2)}}$, η 和 $\hat{u}_i$ ($i = 1, \dots, N$) 见引理 1 中的定义.

证明 计算得

$$\begin{aligned} \lim_{\rho \uparrow 1} (1 - \rho) \begin{pmatrix} X_1^{(1)} \\ \vdots \\ X_N^{(1)} \end{pmatrix} &= \lim_{\rho \uparrow 1} \lim_{n \rightarrow \infty} (1 - \rho) \begin{pmatrix} X_{n,1}^{(1)} \\ \vdots \\ X_{n,N}^{(1)} \end{pmatrix} \\ &= \lim_{\rho \uparrow 1} \lim_{n \rightarrow \infty} (1 - \rho) \pi_n(\theta(\rho)) \times \lim_{\rho \uparrow 1} \lim_{n \rightarrow \infty} \frac{1}{\pi_n(\theta(\rho))} \begin{pmatrix} X_{n,1}^{(1)} \\ \vdots \\ X_{n,N}^{(1)} \end{pmatrix} \end{aligned}$$

$$= \eta A \begin{pmatrix} \hat{v}_1 \\ \vdots \\ \hat{v}_N \end{pmatrix} \Gamma(\alpha, 1) = \frac{\bar{b}^{(2)}}{2\bar{b}} \frac{1}{\eta} \begin{pmatrix} \hat{u}_1 \\ \vdots \\ \hat{u}_N \end{pmatrix} \Gamma(\alpha, 1).$$

证毕. $\square$

注 3 $\mathbb{E}X_i^{(k)}$ ($i, k = 1, \dots, N$) 也可由缓冲区容量直接计算得到, 参见文献 [7]. 当 $\rho < 1$, $i, k = 1, \dots, N$ 时,

$$\mathbb{E}X_i^{(i)} = \lambda_i r + \lambda_i \sum_{j \neq i} \mathbb{E}X_j^{(j)} \mathbb{E}T_j + \sum_{j \neq i} \mathbb{E}X_j^{(j)} \mathbb{E}M_j \frac{p_{j,i}}{1 - p_{j,j}} + \mathbb{E}X_i^{(i)} \mathbb{E}L_i, \quad (4.18)$$

$$\mathbb{E}X_i^{(k)} = \lambda_i \sum_{j=i}^{k-1} r_j + \lambda_i \sum_{j=i+1}^{k-1} \mathbb{E}X_j^{(j)} \mathbb{E}T_j + \sum_{j=i+1}^{k-1} \mathbb{E}X_j^{(j)} \mathbb{E}M_j \frac{p_{j,i}}{1 - p_{j,j}} + \mathbb{E}X_i^{(i)} \mathbb{E}L_i. \quad (4.19)$$

化简 (4.18) 得

$$\frac{1}{\lambda_i} \left[\frac{1}{b_k} \mathbb{E}X_i^{(i)} \mathbb{E}T_i - \sum_{j=1}^N \frac{p_{j,i}}{b_j} \mathbb{E}X_j^{(j)} \mathbb{E}T_j \right] = r + \sum_{j=1}^N \mathbb{E}X_j^{(j)} \mathbb{E}T_j. \quad (4.20)$$

方程 (4.20) 与 (4.12) 类似, 直接得方程 (4.20) 的解为

$$\mathbb{E}X_i^{(i)} = \frac{\gamma_i r (1 - p_{i,i} - \rho_i)}{f_i (1 - \rho)}, \quad i = 1, \dots, N. \quad (4.21)$$

故

$$\lim_{\rho \uparrow 1} (1 - \rho) \mathbb{E}X_i^{(i)} = \frac{\gamma_i r (1 - p_{i,i} - \rho_i)}{f_i}, \quad i = 1, \dots, N. \quad (4.22)$$

将 (4.21) 代入 (4.19) 得

$$\lim_{\rho \uparrow 1} (1 - \rho) \mathbb{E}X_i^{(k)} = \frac{\gamma_i r (1 - p_{i,i} - \rho_i)}{f_i} + \sum_{j=i+1}^{k-1} \rho_j r \left[\lambda_i + \frac{p_{j,i}}{b_j} \right] = \hat{u}_i^{(k)} r, \quad i, k = 1, \dots, N. \quad (4.23)$$

注 4 由定理 2 的证明知,

$$\lim_{\rho \uparrow 1} (1 - \rho) \mathbb{E}X_i^{(1)} = \eta \cdot A \cdot \hat{v}_i \cdot \alpha = \hat{u}_i r, \quad (4.24)$$

其中第二个等号是将 (4.8) 代入上式化简得到的. 显然, (4.23) 和 (4.24) 提供了求解 $\hat{\mathbf{u}}$ 的另一种方法.

推论 1 对于分支型轮询排队网络, 服务器进入队列 Q_k ($k = 1, \dots, N$) 时的联合队长有下面的极限性态:

$$(1 - \rho) \begin{pmatrix} X_1^{(k)} \\ \vdots \\ X_N^{(k)} \end{pmatrix} \xrightarrow{d} \frac{\bar{b}^{(2)}}{2\bar{b}} \frac{1}{\eta} \begin{pmatrix} \hat{u}_1^{(k)} \\ \vdots \\ \hat{u}_N^{(k)} \end{pmatrix} \Gamma(\alpha, 1), \quad \rho \uparrow 1, \quad (4.25)$$

其中 $\alpha = 2r \frac{\bar{b}}{\bar{b}^{(2)}}$, η 和 $\hat{u}_i^{(k)}$ ($i = 1, \dots, N$) 见引理 1 中定义.

证明 由 (4.23)、(4.24) 和定理 2 可以直接得到上面的结论. $\square$

注 5 当 $f_i = 1 - p_{i,i} - \rho_i$ 时, 即得到文献 [5] 中门限服务策略下联合队长在高负荷下的极限性态. 当 $p_{i,0} = 1$, $i = 1, \dots, N$ 时, 本文的排队系统退化为一般的分支型轮询排队系统, 得到的结论与文献 [9] 中的结论一致.

5 高负荷下平均等待时间的极限性态

本节将给出高负荷下任意时刻各队列工作量的极限性态, 并由 Little 公式给出平均等待时间的极限. 记 $V(t)$ 为 t 时刻系统的总工作量, $V_i(t)$ 为 t 时刻队列 Q_i 的工作量, W_i 为进入队列 Q_i 中顾客在 Q_i 中的等待时间, $i = 1, \dots, N$.

定义服务器相邻两次进入队列 Q_i 服务的时间间隔为一个周期, 记为 C_i . 由平衡方程, 显然有

$$\mathbb{E}C_i = \frac{r}{1-\rho},$$

即周期的长度与起始点没有关系. 现在考虑一列有相同服务率、转移时间及顾客转移概率的分支型轮询排队网络 $\{S^{(n)}, n = 1, 2, \dots\}$, 记 $\rho^{(n)}$ 为排队网络 $S^{(n)}$ 的总负荷率, $V^{(n)}(t)$ 为其在 t 时刻的总工作量. 当 $n \rightarrow \infty$ 时, 有 $\rho^{(n)} \uparrow 1$, 且

$$\lim_{n \rightarrow \infty} \sqrt{n}(\rho^{(n)} - 1) = c, \quad -\infty < c < \infty.$$

记

$$U^{(n)}(t) = \frac{1}{\sqrt{n}}V^{(n)}(nt), \quad n \geq 1, \quad 0 \leq t \leq 1.$$

由文献 [14, 15] 中的高负荷平均准则 (heavy traffic averaging principle), 我们得到

$$U^{(n)}(t) \rightarrow_d U(t), \quad n \rightarrow \infty, \quad (5.1)$$

其中 $U(t)$ 是一个 Bessel 型扩散过程, 漂移项

$$c(t) = c + \frac{r\eta}{t},$$

方差

$$\sigma^2 = \frac{\bar{b}^{(2)}}{\bar{b}}.$$

令 $n = \frac{1}{(1-\rho)^2}$, 则有 $c = -1$. 高负荷平均准则显示, 每个队列的工作量的变化速度比总工作量的要快很多, 且在一个周期中, 可以认为总工作量保持不变, 而每个队列的工作量则看成是一个流体模型.

记 $v_i^{(j)}$ 为服务器进入 Q_j 服务时队列 Q_i 的工作量. 由 (4.18)、(4.19) 和 (4.23) 可得

$$v_i^{(j)} = \hat{u}_i^{(j)} \bar{b}_i, \quad i, j = 1, \dots, N. \quad (5.2)$$

在流体模型中, 当服务器在 Q_j ($j \neq i$) 服务时, Q_i 的工作量流入率为 $\hat{\lambda}_i \bar{b}_i + \frac{p_{j,i} \bar{b}_i}{b_j}$, 流出率为 0; 当服务器在 Q_i 服务时, Q_i 的工作量流入率为 $\hat{\lambda}_i \bar{b}_i$, 流出率为 $\frac{(1-p_{i,i}) \bar{b}_i}{b_i}$. 任取时刻 t , 服务器以 $\hat{\rho}_j$ 的概率在 Q_j 服务, 且 Q_i 的工作量等于服务器刚进入 Q_j 服务时 Q_i 的工作量 $v_i^{(j)}$ 加上工作量的净流入量. 记任意时刻队列 Q_i 的工作量为 η_i , 由全概率公式可得

$$\begin{aligned} \eta_i &= \hat{\rho}_i \left[v_i^{(i)} + \int_0^{\hat{\rho}_i} \frac{1}{\hat{\rho}_i} \left(\hat{\lambda}_i \bar{b}_i - \frac{(1-p_{i,i}) \bar{b}_i}{b_i} \right) t dt \right] + \sum_{j \neq i} \hat{\rho}_j \left[v_i^{(j)} + \int_0^{\hat{\rho}_j} \frac{1}{\hat{\rho}_j} \left(\hat{\lambda}_i \bar{b}_i + \frac{p_{j,i} \bar{b}_i}{b_j} \right) t dt \right] \\ &= \sum_{j=1}^N \hat{\rho}_j v_i^{(j)} + \frac{1}{2} \hat{\rho}_i^2 \bar{b}_i \left(\hat{\lambda}_i - \frac{1-p_{i,i}}{b_i} \right) + \frac{1}{2} \sum_{j \neq i} \hat{\rho}_j^2 \bar{b}_i \left(\hat{\lambda}_i + \frac{p_{j,i}}{b_j} \right). \end{aligned} \quad (5.3)$$

由全概率公式可知, 任意时刻总工作量的净流入率为 0, 因此有

$$\sum_{i=1}^N \eta_i = \sum_{i=1}^N v_i^{(1)} = \sum_{i=1}^N \hat{u}_i^{(1)} \bar{b}_i = |\bar{\mathbf{b}}| \hat{\mathbf{u}}^T \hat{\mathbf{w}} = \eta.$$

故任意时刻队列 Q_i 的工作量

$$V_i(t) = \frac{\eta_i}{\eta} V(t).$$

由 (5.1) 和 Little 公式, 得下述定理:

定理 3 对于分支型轮询排队网络, 在高负荷下, 队列 Q_i 中顾客的平均等待时间有以下极限行为:

$$\lim_{\rho \uparrow 1} (1 - \rho) \mathbb{E} W_i = \left(r + \frac{\bar{b}^{(2)}}{2\eta \bar{b}} \right) \frac{\eta_i}{\bar{b}_i \hat{\gamma}_i}, \quad i = 1, \dots, N. \quad (5.4)$$

6 结论

本文对分支型轮询排队网络在高负荷下的极限行为进行了分析, 我们的结论是文献 [5, 9] 中的一般化. 此外, 本文的结论也为内部顾客和外部顾客等待时间的极限性态研究提供了部分理论基础. 虽然该方法仅限于求解 Poisson 到达的轮询排队系统, 但依然可以基于这些结论对具有一般更新到达的轮询排队系统的极限行为进行探究. 由于轮询排队系统中大部分服务策略都具有分支性, 因此, 本文的方法还可以进一步推广到更广泛的模型.

参考文献

- 1 储育青, 刘再明, 吴锦标. 带有阈值和优先权的三队列轮询排队系统. 中国科技论文, 2014, 9: 434–440
- 2 Kavitha V, Altman E. Queuing in space: Design of message ferry routes in static ad hoc networks. In: Teletraffic Congress, 2009. ITC 21 2009. 21st International. Paris: IEEE, 2009, 1–8
- 3 Grasman S E, Olsen T L, Birge J R. Setting basestock levels in multi-product systems with setups and random yield. IIE Trans, 2008, 40: 1158–1170
- 4 Boon M A, van der Mei R D, Winands E M. Waiting times in queueing networks with a single shared server. Queueing Syst, 2013, 74: 403–429
- 5 Boon M A, van der Mei R D, Winands E M. Queueing networks with a single shared server: Light and heavy traffic. ACM SIGMETRICS Performance Evaluation Review, 2011, 39: 44–46
- 6 Sidi M, Levy H. customer routing on polling systems. In: Proceedings of the 14th IFIP WG 7.3 International Symposium on Computer Performance Modelling, Measurement and Evaluation. Amsterdam: North-Holland Publishing Co, 1990, 319–331
- 7 Sidi M, Levy H, Fuhrmann S W. A queueing network with a single cyclically roving server. Queueing Syst, 1992, 11: 121–144
- 8 Boon M A, Winands E M, Adan I J, et al. Closed-form waiting time approximations for polling systems. Perform Evaluation, 2011, 68: 290–306
- 9 van der Mei R D. Towards a unifying theory on branching-type polling systems in heavy traffic. Queueing Syst, 2007, 57: 29–46
- 10 Resing J A C. Polling systems and multitype branching processes. Queueing Syst, 1993, 13: 409–426
- 11 Quine M. The multitype Galton-Watson process with ρ near 1. Adv Appl Probab, 1972, 14: 429–452
- 12 Seneta E. Non-Negative Matrices and Markov Chains. New York: Springer, 1981
- 13 van der Mei R D, Levy H. Expected delay analysis of polling systems in heavy traffic. Adv Appl Probab, 1998, 30: 586–602
- 14 Coffman Jr E, Puhalskii A, Reiman M. Polling systems with zero switchover times: A heavy-traffic averaging principle. Ann Appl Probab, 1995, 5: 681–719

- 15 Coffman Jr E, Puhalskii A, Reiman M. Polling systems in heavy traffic: A Bessel process limit. *Math Oper Res*, 1998, 23: 257–304

The asymptotic behavior of a branching-type polling network in heavy traffic

LIU ZaiMing, CHU YuQing & WU JinBiao

Abstract We consider a polling network with a single server that serves the queues in a cyclic order. Upon the completion of his service at queue i , a customer is routed to queue j with probability $p_{i,j}$, or leaves the system with probability $p_{i,0}$. The service policy is branching-type. Using the limit theorem of a multi-type branching process with immigration, we derive the asymptotic behavior of joint queue lengths at polling epochs in the heavy traffic. With the help of heavy traffic averaging principle, the limit behavior of the mean waiting times is also presented.

Keywords polling network, multi-type branching process, asymptotic behavior, descendant set approach

MSC(2010) 60K25, 60K20

doi: 10.1360/N012015-00044