应用数学和力学,第 28 卷 第 10 期
2007 年 10 月 15 日出版

Applied Mathematics and Mechanics
Vol. 28, No. 10, Oct. 15, 2007

文章编号: 1000-0887(2007) 10-1185-12

抢占型优先服务机制下多类排队网络的扩散逼近*

戴万阳

(南京大学 数学系, 南京 210093)

(郭懋正推荐)

摘要: 证明一个满负荷交通极限定理以证实在抢占型优先服务机制下多类排队网络的扩散逼近,进而为该系统提供有效的随机动力学模型. 所研究的排队网络典型地出现在现代通讯系统中高速集成服务分组数据网络,其中包含分组数据包的若干交通类型,每个类型涉及若干工作处理类(步骤),并且属于同一交通类型的工作在可能接受服务的每一个网站被赋予相同的优先权等级,更进一步地,在整个网络中,属于不同交通类型的分组数据包之间无交互路由.

关 键 词: 排队网络; 抢占型优先权; 满负荷交通; 半鞅反射布朗运动; 流体模型; 扩散逼近; Liapunov 函数

中图分类号: O211; O226 文献标识码: A

引 言

受到现代通讯系统中高速集成服务分组数据网络的启迪,我们研究一类多类排队网络并建立相应的逼近随机动力学模型以用于系统的性能评估与预测. 该排队网络的一个重要特性便是每个网站能够处理多类别的工作并具有复杂的反馈结构,而满负荷(输入负荷靠近服务容量)的网络尤其令人感兴趣,其中网络的拥挤与堵塞现象是亟待解决的问题. 然而,这些网络的精确分析常常是难以得到的,因而,寻求其可处理的逼近便是一种很自然的方法. 与此相连,在满负荷交通条件下,被称为半鞅反射布朗运动(SRBM)的一类扩散过程已被证明可以用来逼近许多单类排队网络以及某些多类排队网络规范化的队长或工作量过程(比如: W. Dai[1-2], J. G. Dai 和 W. Dai[3], Reiman[4], Bramson[5-6] 和 Williams[7-8]. Bramson 和 J. G. Dai[9], Chen 和 Zhang[10-11]). 逼近的布朗模型(SRBM)仅需要网络中相继到达时间,服务时间与路由向量的前二阶矩的信息,更进一步地,对于这些布朗模型,包括平稳分布在内的许多量可以被精确或者数值计算出(比如: Harrison 和 Williams[12], Dai 和 Harrison[13], W. Dai[2] 及其合作者[14]),因而,人们可以通过布朗模型得到原网络的诸如平均队长与平均延时等性能估计. 不幸的是,现在已知: 在满负荷交通条件下,并非所有具有反馈的多类网络能够被 SRBM 逼近(比如: Dai 和 Wang[15]). 事实上,在排队网络的现代研究中,极具挑战性的问题之一便是识别能有如此逼近

* 收稿日期: 2005-10-03; 修订日期: 2007-07-11
基金项目: 国家自然科学基金资助项目(10371053)
作者简介: 戴万阳(1963—),男,江苏盐城人,教授,博士(E-mail: nan5lu8@ netra. nju. edu. cn).

的诸多网络并证明一个满负荷交通极限定理以证实这样的逼近. 因此, 本文识别出一类具有一般性相继到达时间与服务时间分布并在抢占型优先服务机制下的多类排队网络可以有这样的逼近, 同时我们也通过证明一个满负荷交通极限定理来证实这样的逼近.

传统的证明上述逼近的方法依赖于相应于排队网络的 Skorohod 问题的解是存在的且是唯一的(比如: Reiman[4]). 然而, 在具有有限缓存的单类网络与具有反馈的多类网络中, 唯一性并非总是成立的(比如: W. Dai[1-2] J. G. Dai 和 W. Dai[3], Dai 和 Wang[15]). 对于这些情形, 传统的方法不能轻易地被推广. 为了克服这些困难, 一些作者(W. Dai[1-2], J. W. Dai 和 W. Dai[3]) 首次建立起了一般性的弱收敛方法来证明具有有限缓存的单类网络的扩散逼近问题, 之后, Williams[7-8, 16] 和 Bramson[5-6] 则采用了与之类似的路径证明了某些类型具有反馈的多类网络的逼近问题. 沿着上述路径, 针对具有反馈的多类网络, 证明一个满负荷交通极限定理的关健在于证明一个状态空间倒塌性质与某反射矩阵的完全 S 性质. Bramson 和 J. W. Dai[9] 进一步归纳了文献[5]、文献[7]、文献[16] 中的结果并证明了以下结论: 只要能证明与排队网络相应的流体模型的一致收敛性便能证明网络状态空间倒塌性. 利用这种方法, 他们在文献[9] 中针对单个网站的抢占型优先服务多类排队系统证明了一个满负荷交通极限定理. 但是, 对于前面提及的较为一般的网络模型, 其一致收敛性与完全 S 的证明却不是一件简单的事情. 因而, 它们的证明构成了本文主要定理论证的主体部分.

本文的其余部分按下列顺序编排: 在第 1 节, 我们描述网络模型; 在第 2 节, 我们陈述我们的主要定理; 在第 3 节, 我们证明我们的主要定理.

1 排队网络模型

我们所考虑的排队网络由 J 个单服务员网站组成的, 且每各网站有一个无限容量的等待缓存. 在此网络中, 有 h 个交通类型且每个类型由分布在不同网站的 J 个工作类(简称类) 组成, 因而, 该网络中有 $K(=Jh)$ 个类. 网站由 $j(=1, \cdots, J)$ 标记, 而类则由 $k(=1, \cdots, K)$ 标识. 在属于某个交通类型的一个工作(比如一个分组数据包) 从网络外部到达时, 它可以仅受到 J 个类当中的部分服务而且可以访问某一网站多次, 然后离开网络. 在该工作逗留于网络的任何给定时间点上, 它一定属于某个类• 当该工作在网络中各网站之间移动时, 它改变着自身的类且成为改变类后的一个新的工作, 而这种改变发生在完成某类一次服务的时刻. 属于某类的所有工作均在特定的一个网站接受服务, 且多个类可以在同一网站接受服务. 我们假设每个工作最终将离开该网络且属于不同交通类型的工作之间无交互路由.

有关控制工作在每个网站接受服务次序的服务机制, 我们认定属于同一交通类型的工作在所有它们可能接受服务的网站中具有相同的优先权等级(例如, 在图 1 中, 我们给了 1 个例子, 这里交通类型 1 可能需要接受类 1 和类 2 的服务; 交通类型 2 可能需要接受类 3 和类 4 的服务; 在它们相应的网站中, 属于交通类型 1 的工作具有较低的优先权.). 在服务员从一个工作转换到另一个工作时, 新接受服务的工作将是该网站中优先权最高且非空类中最前面(即等待时间最长) 的那一个工作. 更进一步地, 我们假设服务机制

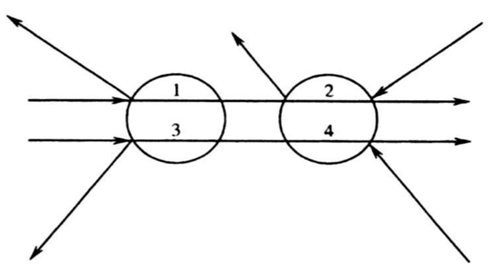

图 1 1 个由两网站两交通类型和 4 个工作类组成的网络

是抢占型恢复的，即当一个比正在接受服务的工作优先权等级高的工作到达该网站时，那个正在接受服务的工作将被终止服务直到所有比它优先权等级高的工作完成服务，它才恢复未完成的服务．最后，我们假设我们的策略是非空闲的，即只要有工作在某个网站里，则该网站的服务员便不能空闲着．

我们用 $\mathscr{C}(j)$ 表示属于网站 j 的所有工作类(简称类)的集合，用 $s(k)$ 表示类 k 所属的网站．当 j 和 k 一齐出现时，则蕴涵着取 $j = s(k)$．对于网络中的每一个类 k，存在着下列的一些量与之相对应：两个独立同分布(i. i. d.)的随机变量序列 $u_k = \{u_k(i), i \geqslant 1\}$ 和 $v_k = \{v_k(i), i \geqslant 1\}$，一个 K 维 i. i. d. 随机向量序列 $\phi^k = \{\phi^k(i), i \geqslant 1\}$，两个实数 $\alpha_k \geqslant 0$ 和 $m_k > 0$．我们假设下列 $3K$ 序列

$$u_1, \cdots, u_K; v_1, \cdots, v_K; \phi^1, \cdots, \phi^K \tag{1}$$

之间是相互独立的．令 $a_k = \mathrm{var}(u_k(1))$，$b_k = \mathrm{var}(v_k(1))$，并假设 $a_k < \infty$ 和 $b_k < \infty$，u_k 和 v_k 是被单位化了的，即 $E[u_k(1)] = 1$ 且 $E[v_k(1)] = 1$．对于每一个 i，$u_k(i)/\alpha_k$ 表示第 $(i-1)$ 个和第 i 个由网络外部到达类 k 的工作的相继到达时间，$m_k v_k(i)$ 表示第 i 个类 k 工作的服务时间，$\phi^k(i)$ 表示第 i 个类 k 工作的路由向量．这样，对于每一个 k，m_k 是类 k 工作的平均服务时间，α_k 是网络外部到达类 k 的速率，a_k 和 b_k 分别是相继到达与服务时间的平方变差系数(即方差除以期望的平方．对某些类 k，我们允许 $\alpha_k = 0$ 并令 $E = \{k: \alpha_k \neq 0\}$．我们假设路由向量 $\phi^k(i)$ 在集 $\{\boldsymbol{e}_0, \boldsymbol{e}_1, \cdots, \boldsymbol{e}_K\}$ 中取值，这里 $\boldsymbol{e}_0$ 是分量均为 0 的 K 维向量，而 $\boldsymbol{e}_l (l = 1, \cdots, K)$ 是第 l 个分量为 1 其余分量为 0 的 K 维向量．如果 $\phi^k(i) = \boldsymbol{e}_l$，则第 i 个离开类 k 的工作变为一个类 l 的工作．若用 $P_{kl} = \boldsymbol{P}\{\phi^k(i) = \boldsymbol{e}_l\}$ 表示一个离开类 k 的工作变为(属于同一交通类型)类 l 工作的概率，则称 $K \times K$ 阶矩阵 $\boldsymbol{P} = (P_{kl})$ 为网络的路由矩阵．由于我们假设了每个工作最终离开网络，因而我们的网络是一种开型网络，即下面的矩阵是有限的

$$\boldsymbol{Q} = \boldsymbol{I} + \boldsymbol{P}' + (\boldsymbol{P}')^2 + \cdots \tag{2}$$

其等价于 $(\boldsymbol{I} - \boldsymbol{P}')$ 是可逆阵且 $\boldsymbol{Q} = (\boldsymbol{I} - \boldsymbol{P}')^{-1}$，这里 $\boldsymbol{I}$ 表示单位阵而 $\boldsymbol{P}'$ 则表示 $\boldsymbol{P}$ 的转置矩阵.

为了研究开型多类排队网络，人们通常引入下列交通方程的解 λ_l，$l = 1, \cdots, K$，

$$\lambda_l = \alpha_l + \sum_{k=1}^{K} \lambda_k P_{kl} \tag{3}$$

或写成等价的向量形式：$\lambda = \alpha + \boldsymbol{P}'\lambda$．由于对应于 $\boldsymbol{P}$ 的网络是开型的，因而，交通方程(3)的唯一解由 $\lambda = \boldsymbol{Q}\alpha$ 给出．λ_k 被称为是类 k 名义上的总的到达率，它不仅取决于网络外部的到达也取决于网络内部的到达．对于每一个类 k，如果存在该类的一个长期平均输入速率等于该类的长期平均输出速率，则该速率便是 λ_k．利用向量 $\boldsymbol{m} = (m_1, \cdots, m_K)'$ 和 λ，我们定义 ρ_j 为第 j 个服务员的交通强度，即

$$\rho_j = \sum_{k \in \mathscr{C}(j)} \lambda_k m_k \tag{4}$$

或写成向量形式 $\rho = \boldsymbol{C}\boldsymbol{M}\lambda$，这里 $\boldsymbol{M} = \mathrm{diag}(\boldsymbol{m})$ (它表示一个 $K \times K$ 对角矩阵，其主对角上的元素由 $\boldsymbol{m}$ 的分量给出，且其它元素为 0) 且 $\boldsymbol{C}$ 的元素由下式给出

$$C_{jk} = \begin{cases} 1, & \text{若 } k \in \mathscr{C}(j)\text{，即类 } k \text{ 在网站 } j \text{ 接受服务} \\ 0, & \text{其它．} \end{cases} \tag{5}$$

当 $\rho_j \leqslant 1$ 时，它也被称为服务员 j 忙于工作的那部分时间的名义比例．在该文中，我们将研究

相应于每个网站的 ρ_j 靠近 1 的网络, 即所谓的满负荷网络.

2 一个满负荷交通极限定理

在叙述我们的满负荷交通极限定理之前, 我们还需要一些术语和定义.

2.1 量化与满负荷交通条件

用 α^r 和 $\boldsymbol{m}^r$ 分别表示由 r 标记的一列与网络相应的外部到达率与平均服务时间的向量, 这里 $r \in \{1, 2, \ldots\}$. 令 $\lambda^r = \boldsymbol{Q}\alpha^r$, $\rho^r = \boldsymbol{C}M^r\lambda^r$ 且 $M^r = \mathrm{diag}(\boldsymbol{m}^r)$. 假设集合 $E = \{k: \alpha_k^r \neq 0\}$ 和路由矩阵 $\boldsymbol{P}$ 不依赖于 r, 且在 $r \to \infty$ 时, α^r 和 $\boldsymbol{m}$ 满足下列满负荷交通条件

$$\alpha_k^r \to \alpha_k > 0 \text{ 当 } k \in E, \quad m_k \to m_k > 0 \text{ 当 } k = 1, \ldots, K, \tag{6}$$

且 $\rho^r \to \boldsymbol{e}$ 并具有下列收敛速率

$$r(\rho^r - \boldsymbol{e}) \to \gamma, \tag{7}$$

这里 $\boldsymbol{e}$ 是每个分量都为 1 的 J 维向量, 而 γ 是某一 J 维向量. 由(6) 式和(7) 式, 我们有

$$\rho = \boldsymbol{CM}\lambda = \boldsymbol{e}, \tag{8}$$

即在极限情况下, 每个网站被临界负荷了. 对于类 k, 相继到达时间与服务时间分别由 $\{u_k(i)/\alpha_k^r, i = 1, 2, \ldots\}$ 和 $\{m_k^r v_k(i), i = 1, 2, \ldots\}$ 给出, 因而, 相继到达时间与服务时间的平方变差系数 a_k 和 b_k 不依赖于 r.

对于第 r 个网络, 下列过程 $\boldsymbol{Z}^r$, $\boldsymbol{W}^r$ 和 $\boldsymbol{Y}^r$ 将被用来评估它的性能. $\boldsymbol{Z}^r = \{\boldsymbol{Z}^r(t), t \geqslant 1\}$ 是 K 维的, 且 $Z_k^r(t)$ 表示 t 时刻处的类 k 工作的数目, 被称为队长过程. 另外两个过程 $\boldsymbol{W}^r = \{\boldsymbol{W}(t), t \geqslant 1\}$ 和 $\boldsymbol{Y}^r = \{\boldsymbol{Y}^r(t), \geqslant 1\}$ 都是 J 维的. 对于每一个网站 j, $W_j^r(t)$ 表示服务员 j 完成在 t 时刻出现在网站 j 中所有工作所需的时间, 称为瞬时工作量; $Y_j^r(t)$ 表示区间 $[0, t]$ 中, 服务员 j 总的闲期, 称为累积闲期过程. 队长与工作量过程用来评估网络中的拥挤与时滞状况, 而闲期过程用来评估网络中的资源利用状况. 在 $\rho^r \to \boldsymbol{e}\,(r \to \infty)$ 时, 队长过程, 工作量过程与闲期过程将变大. 因而, 鉴于泛函中心极限定理, 我们定义下列量化了的队长过程 $Z^r(t) = (Z_1^r(t), \ldots, Z_K^r(t)))'$, 量化了的工作量过程 $\boldsymbol{W}(t) = (W_1^r(t), \ldots, W_K^r(t)))'$ 与量化了的闲期过程 $\boldsymbol{Y}^r(t) = (Y_1^r(t), \ldots, Y_K^r(t)))'$ 如下:

$$Z_k^r = \frac{1}{r}Z_k^r(r^2 t), \quad W_k^r(t) = \frac{1}{r}W_k^r(r^2 t), \quad Y_k^r(t) = \frac{1}{r}Y_k^r(r^2 t)).$$

2.2 一些定义

在本节中, $\mathscr{B}$ 表示由 $\mathscr{R}_+^J$ 的波雷尔子集生成的 σ 代数且 $\mathscr{R}_+ = [0, \infty)$, θ 是 $\mathscr{R}^J$ 中的一个向量而 $\mathscr{R}^J$ 则表示 J 维欧氏空间, Γ 是一个 $J\times J$ 对称正定矩阵, $\boldsymbol{R}$ 是一个 $J\times J$ 矩阵, ν 是 $(\mathscr{R}_+^J, \mathscr{B})$ 上的一个概率测度. 由文献 Williams[16], 我们有下列定义.

定义 1(SRBM) 具有参数 $(\mathscr{R}_+^J, \theta, \Gamma, \boldsymbol{R}, \nu)$ 的 SRBM 是一个定义在某流域概率空间 $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}, \mathscr{P})$ 上的 $\{\mathscr{F}_t\}$– 适应的 J 维随机过程 $\boldsymbol{W}$ 满足: $\mathscr{P}$ a. s. .

1. $\boldsymbol{W}$ 具有连续轨道且 $\boldsymbol{W}(t) \in \mathscr{R}_+^J\ (t \geqslant 1)$;
2. 对于合适的 J 维过程 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$, 有 $\boldsymbol{W} = \boldsymbol{X} + \boldsymbol{RY}$;
3. $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 具有下列性质: 在 $\mathscr{P}$ 下

(a) $\boldsymbol{X}$ 是一个具有飘移向量 θ, 协方差矩阵 Γ 的布朗运动, 且 $\boldsymbol{X}(0)$ 具有初始分布 ν;

(b) $\{\boldsymbol{X}(t) - \boldsymbol{X}(0) - \theta t, \mathscr{F}_t, t \geqslant 0\}$ 是一个鞅;

4. $\boldsymbol{Y}$ 是一个 $\{\mathscr{F}_t\}$– 适应的 J 维过程满足: 对每一 $j = 1, \ldots, J$, $\mathscr{P}$ a. s.

(a) $Y_j(0)=0$;

(b) Y_j 是连续非降的;

(c) Y_j 仅在 $W_j(t)=0$ 的时刻 t 能递增.

定义 2(完全S) 一个 $J\times J$ 矩阵被称为是完全S 的充分必要条件是: 对于 $\boldsymbol{R}$ 的每一主子矩阵 $\boldsymbol{R}$, 存在 $\boldsymbol{x}>0$ 使的 $\boldsymbol{Rx}>0$, 这里的向量不等式是以分量比较形式出现的.

2.3 主要定理

为了陈述我们的主要定理, 我们需要一些一般性的假设. 由于 α 和 $\boldsymbol{m}$ 是(6) 式中的极限且 $\lambda=\boldsymbol{Q}\alpha$. 我们假设(6) 式成立, 且对所有 k, $\lambda_k>0$. 用 Λ、Σ 和 Π 表示主对角元素分别由 λ_k、b_k 和 $\alpha_k^3 a_k(k=1,\cdots,K)$ 组成的对角矩阵, 并令 Γ^k 是下列 $K\times K$ 矩阵

$$\Gamma_{ll'}^k=\begin{cases}P_{kl}(1-P_{kl}), & \text{当 } l=l',\\ -P_{kl}P_{kl'}, & \text{当 } l\neq l',\end{cases}$$

这里, $l,l'=1,\cdots,K$. 我们可以验证: Γ^k 是路由向量 $\phi^k(1)$ 的协方差矩阵. 因此, 它是对称且是非负定的. 令

$$\boldsymbol{H}=\boldsymbol{C}\left(\Lambda\Sigma+\boldsymbol{MQ}\left(\Pi+\sum_{k=1}^K\lambda_k\Gamma^k\right)\boldsymbol{Q}'\boldsymbol{M}\right)\boldsymbol{C}'. \tag{9}$$

由(9) 式可知: 既然 $\sum_{k=1}^K\lambda_k\Gamma^k$ 与两个对角阵是对称且非负定的, $\boldsymbol{H}$ 也是对称且非负定的. 我们将假设 $\boldsymbol{H}$ 是正定的以保证网络是具有随机性的. 用 Δ 表示一个 $K\times J$ 非负矩阵, 其元素由下式给出:

$$\Delta_{kj}=\begin{cases}1/m_k, & \text{若 } k \text{ 是网站} j \text{ 中优先权最低的类},\\ 0, & \text{其它}.\end{cases} \tag{10}$$

进一步地, 对于一个向量 $\boldsymbol{a}=(a_1,\cdots,a_d)'$, 定义 $\|\boldsymbol{a}\|=\max_{i=1}^d|a_i|$ 且用 ⇒ 表示 Skorohod 拓扑空间中的弱收敛(比如, 参见文献 Eithier 和 Kurtz[17]).

定理 1 假设(6) 式和(7) 式成立, 初始数据满足: 当 $r\to\infty$ 时, 有

$$\boldsymbol{W}^r(0)\Rightarrow\boldsymbol{W}^*(0), \tag{11}$$

这里, $\boldsymbol{W}^*(0)$ 为某一非负随机向量, 而且当 $r\to\infty$ 时, 有下列依概率收敛,

$$\|\boldsymbol{Z}^r(0)-\Delta\boldsymbol{W}^r(0)\|\to 0, \tag{12}$$

令 $\boldsymbol{X}^r=\boldsymbol{W}^r-\boldsymbol{RY}$, 则当 $r\to\infty$ 时, 存在过程 $\boldsymbol{W}^*$、$\boldsymbol{X}^*$、$\boldsymbol{Y}^*$ 和 $\boldsymbol{Z}^*$ 使得

$$(\boldsymbol{W}^r,\boldsymbol{X}^r,\boldsymbol{Y}^r,\boldsymbol{Z}^r)\Rightarrow(\boldsymbol{W}^*,\boldsymbol{X}^*,\boldsymbol{Y}^*,\boldsymbol{Z}^*), \tag{13}$$

这里 $\boldsymbol{W}^*=\boldsymbol{X}^*+\boldsymbol{RY}^*$ 是一个$(\mathscr{R}_+^J,\theta,\Gamma,\boldsymbol{R},\nu)$-SRBM, 且上述的极限满足下列状态空间倒塌性质

$$\boldsymbol{Z}^*=\Delta\boldsymbol{W}^*, \tag{14}$$

$\boldsymbol{R}$ 是由下式给出的完全S 矩阵

$$\boldsymbol{R}=(\boldsymbol{I}+\boldsymbol{CMQP}'\Delta)^{-1}, \tag{15}$$

参数 θ 和 Γ 为

$$\theta=\boldsymbol{R}\gamma,\quad\Gamma=\boldsymbol{RHR}'. \tag{16}$$

3 主要定理的证明

引理 1 在定理 1 的条件下, 矩阵 $\boldsymbol{I}+\boldsymbol{CMQP}'\Delta$ 是可逆的, 且由(15) 式给出的逆矩阵 $\boldsymbol{R}$ 是完全 S 的.

证明 不失一般性,我们用连续整数去标识在网站 1 到 J 中具有相同优先权等级的工作类,比如,网站 1 到 J 中的最低等级的工作类可以被标记为 $\mathscr{T}_1=\{1,\ \cdots,\ J\}$,倒数第二等级的可以被标记为 $\mathscr{T}_2=\{J+1,\ \cdots,\ 2J\}$,$\cdots$,最高等级的可以被标记为 $\mathscr{T}_h=\{K-J+1,\ \cdots,\ K\}$,这里 $h=K/J$ 是交通类型的数目,且在工作类 $\mathscr{T}_1,\mathscr{T}_2,\ \cdots,\ \mathscr{T}_h$ 之间无交互路由. 因而,基于上述标识方法,可知路由矩阵 $\boldsymbol{P}$ 是由位于主对角线上的 h 个 $J\times J$ 子矩阵 $P_{\mathscr{T}_i\mathscr{T}_i}(i\in\{1,\ \cdots,\ h\})$ 组成的分块对角矩阵,这里 $P_{\mathscr{T}_i\mathscr{T}_i}$ 是相应于在各自网站中具有优先权等级 i 的那些工作类工作的路由概率,即

$$\boldsymbol{P}=\operatorname{diag}(\boldsymbol{P}_{\mathscr{T}_1\mathscr{T}_1},\ \cdots,\ \boldsymbol{P}_{\mathscr{T}_h\mathscr{T}_h}). \tag{17}$$

因而矩阵 $\boldsymbol{Q}$ 也是一个分块对角矩阵并能被表示为:

$$\boldsymbol{Q}=(\boldsymbol{I}-\boldsymbol{P}')^{-1}=\operatorname{diag}((\boldsymbol{I}-\boldsymbol{P}'_{\mathscr{T}_1\mathscr{T}_1})^{-1},\ \cdots,\ (\boldsymbol{I}-\boldsymbol{P}'_{\mathscr{T}_h\mathscr{T}_h})^{-1}). \tag{18}$$

矩阵 $\boldsymbol{CM}$ 能够被分块为

$$\boldsymbol{CM}=[\operatorname{diag}(m_1,\ \cdots,\ m_J),\ \cdots,\ \operatorname{diag}(m_{K-J+1},\ \cdots,\ m_K)]. \tag{19}$$

矩阵 Δ 是由 h 个 $J\times J$ 阶矩阵组成的分块矩阵,具有下列形式:

$$\Delta=[\operatorname{diag}(m_1,\ \cdots,\ m_J)^{-1},\boldsymbol{0},\ \cdots,\boldsymbol{0}]', \tag{20}$$

这里 $\boldsymbol{0}$ 是 $J\times J$ 阶 $\boldsymbol{0}$ 矩阵. 由(18)式、(19)式和(20)式,有

$$\begin{aligned}\boldsymbol{I}+\boldsymbol{G}&=\boldsymbol{I}+\boldsymbol{CM}(\boldsymbol{I}-\boldsymbol{P}')^{-1}\Delta-\boldsymbol{CM}\Delta=\\&\operatorname{diag}(m_1,\ \cdots,\ m_J)(\boldsymbol{I}-\boldsymbol{P}'_{\mathscr{T}_1\mathscr{T}_1})^{-1}[\operatorname{diag}(m_1,\ \cdots,\ m_J)]^{-1}.\end{aligned} \tag{21}$$

由(21)式可以看出 $\boldsymbol{I}+\boldsymbol{G}$ 是可逆的且其逆 $\boldsymbol{R}$ 由下式给出:

$$\boldsymbol{R}=\operatorname{diag}(m_1,\ \cdots,\ m_J)(\boldsymbol{I}-\boldsymbol{P}'_{\mathscr{T}_1\mathscr{T}_1})[\operatorname{diag}(m_1,\ \cdots,\ m_J)]^{-1}. \tag{22}$$

既然网络是开型的,我们知道矩阵 $(\boldsymbol{I}-\boldsymbol{P}'_{\mathscr{T}_1\mathscr{T}_1})$ 是完全 S 的(比如参考 Bernard 和 Kharroubi[18], Harrison 和 Reiman[19]). 因而,容易检验 $\boldsymbol{R}$ 是完全 S 的. □

我们主要定理证明的其余部分在相当程度上依赖于下面的流体模型(正象在文献 Bramson 和 J. G. Dai[9] 中解释的,该模型是排队网络过程的类推),即对所有的 $t\geq 0$,

$$A(t)=\alpha t+\boldsymbol{P}'\boldsymbol{D}(t), \tag{23}$$

$$\boldsymbol{Z}(t)=\boldsymbol{Z}(0)+\boldsymbol{A}(t)-\boldsymbol{D}(t), \tag{24}$$

$$\boldsymbol{W}(t)=\boldsymbol{CM}(\boldsymbol{A}(t)+\boldsymbol{Z}(0))-\boldsymbol{CT}(t), \tag{25}$$

$$\boldsymbol{CT}(t)+\boldsymbol{Y}(t)=\boldsymbol{e}^J t, \tag{26}$$

$$Y_j(t)\text{ 仅在 }W_j(t)=0\text{ 的时刻 }t\text{ 能够递增}(j=1,\ \cdots,\ J), \tag{27}$$

这里假设 $\alpha=(\alpha_1,\ \cdots,\ \alpha_K)'$ 具有非负分量且 $\boldsymbol{m}=(m_1,\ \cdots m_K)'$ 具有正分量,而 $\boldsymbol{M}=\operatorname{diag}(\boldsymbol{m})$ 且 $\boldsymbol{P}$ 是一个流的转换矩阵. 更进一步地,有

$$\boldsymbol{T}(t)=\boldsymbol{MD}(t). \tag{28}$$

我们假设所有的流体过程是连续非负的且 $\boldsymbol{A}(\cdot)$、$\boldsymbol{D}(\cdot)$、$\boldsymbol{T}(\cdot)$ 和 $\boldsymbol{Y}(\cdot)$ 是非降的,并从(23)式至(26)式,可得

$$\boldsymbol{A}(0)=\boldsymbol{D}(0)=\boldsymbol{T}(0)=\boldsymbol{Y}(0)=0, \tag{29}$$

且从(24)式,(25)式和(28)式可得对所有的 $t\geq 0$, 有

$$\boldsymbol{W}(t)=\boldsymbol{CMZ}(t). \tag{30}$$

利用(23)式至(28)式,可证明如此的每一个流体过程是 Lipschitz 连续的,即如果 f 是任一这样的过程,则对某一 $N>0$(仅依赖于$(\alpha,\boldsymbol{m},\boldsymbol{P})$)与对所有的 t_1、$t_2\geq 0$, 有下式成立:

$$|f(t_2)-f(t_1)|\leq N|t_2-t_1|.$$

特别地,它们是绝对连续的,因而它们是关于 $[0, \infty)$ 上的 Lebesgue 测度几乎处处可微的. 如果流体模型(23)式至(30)式的一个解 $(\boldsymbol{A}, \boldsymbol{D}, \boldsymbol{T}, \boldsymbol{W}, \boldsymbol{Z}, \boldsymbol{Y})$ 在时刻一个 $t > 0$ 是可微的,则称该点 t 为此解的一个规则点. 今后,无论何时在我们利用一个流体过程在一点 t 的导数时,我们蕴含地假设了该时刻 t 是一个规则点. 符号 $\dot{f}(t)$ 将被用去表示函数 f 在时刻 t 的导数. 因而,由我们的服务机制可得,对所有规则点 t,

$$\dot{T}^{+}_{k}(t) = 1 \text{ 当 } Z^{+}_{k}(t) > 0, \qquad k = 1, \cdots, K, \tag{31}$$

$Z^{+}_{k}(\cdot)$ 是对应于出现在优先权等级至少为 k 的工作类中的工作的总数,而 $T^{+}_{k}(\cdot)$ 是对应于服务员 $s(k)$ 花在服务优先权等级至少为 k 的工作类的累积时间.

定义 3 令 Δ 是一个 $K \times J$ 的非负矩阵. 一个流体模型被称为是具有升高矩阵 Δ 一致收敛的,如果存在这样的一个函数 $g: \mathscr{R}_{+} \to \mathscr{R}_{+}$,当 $t \to \infty$ 时,$g(t) \to 0$,使得对每一个满足 $\|\boldsymbol{Z}(0)\| = 1$ 的流体模型的解 $(\boldsymbol{A}(\cdot), \boldsymbol{D}(\cdot), \boldsymbol{T}(\cdot), \boldsymbol{W}(\cdot), \boldsymbol{Y}(\cdot), \boldsymbol{Z}(\cdot))$ 及所有 $t \geqslant 0$,有

$$\|\boldsymbol{Z}(t) - \boldsymbol{Z}(\infty)\| \leqslant g(t), \tag{32}$$

这里 $\boldsymbol{Z}(\infty) \in \mathscr{R}^{K}_{+}$ 并满足

$$\boldsymbol{Z}(\infty) = \Delta w, \qquad \text{对某一 } \boldsymbol{w} \in \mathscr{R}^{K}_{+}. \tag{33}$$

命题 2 假定矩阵 Δ 由(10)式给出且条件(6)式、(7)式、(11)式和(12)式都成立,则流体模型(23)式至(31)式是具有升高矩阵 Δ 一致收敛的.

证明 利用引理 1 证明中的符号,我们知道 $\mathscr{K}_1$ 是各网站中优先权等级最底的所有工作类组成的指标集,$\mathscr{K}_i$($i \in \{2, \cdots, h\}$)是各网站中优先权等级为 i 的所有工作类组成的指标集. 我们将证明在初始数据 $\|\boldsymbol{Z}(0)\| = 1$ 的条件下,在有限时间内,流体水平 $Z_k(t)$($k \in \mathscr{K}_i$, $i \in \{2, \cdots, h\}$)达到 0,并且在那有限时间以后,$Z_k(t)$ 保持常数.

第 1 步 我们证明在有限时间内,流体水平 $Z_k(t)$($k \in \mathscr{K}_i$, $i \in \{2, \cdots, h\}$)达到 0. 对于一个 K 维向量 $\boldsymbol{x}$,我们用 $(\boldsymbol{x}_{\mathscr{K}_1}, \cdots, \boldsymbol{x}_{\mathscr{K}_h})'$ 来表示相应的分割,而用 $M_{\mathscr{K}_i}$($i \in \{2, \cdots, h\}$)表示主对角线由 m_k($k \in \mathscr{K}_i$)组成的 $J \times J$ 矩阵,并令

$$\boldsymbol{Q}_{\mathscr{K}_i} = (\boldsymbol{I} - \boldsymbol{P}'_{\mathscr{K}_i\mathscr{K}_i})^{-1}.$$

这样,我们可以构造一个由所有优先权等级为 i 的工作类的流体水平组成的 Liapunov 函数(总的工作量)如下:

$$f_i(t) = \boldsymbol{e}' M_{\mathscr{K}_i} \boldsymbol{Q}_{\mathscr{K}_i} \boldsymbol{Z}_{\mathscr{K}_i}(t), \tag{34}$$

这里 $\boldsymbol{e}$ 表示所有分量均为 1 的 J 维向量. 注意到引理 1 证明中的(17)式和(18)式,我们可以构造一个由较高优先权等级工作类的流体水平组成的 Liapunov 函数,

$$f(t) = \sum_{i=2}^{h} f_i(t). \tag{35}$$

现在,我们想证明存在一个合适的 $\delta \geqslant 0$,使得对所有 $t \geqslant \delta$,有 $f(t) = 0$ 成立. 为了达此目的,我们对 $i \in \{2, \cdots, h\}$ 采用归纳法.

在进行归纳假设分析之前,我们需导出一些与集 $\mathscr{K}_i$($i \in \{1, 2, \cdots, h\}$)有关的方程. 由(23)式和(24)式及路由矩阵 $\boldsymbol{P}$ 的结构,我们有

$$\boldsymbol{Z}_{\mathscr{K}_i}(t) = \boldsymbol{Z}_{\mathscr{K}_i}(0) + \alpha_{\mathscr{K}_i} t - (\boldsymbol{I} - \boldsymbol{P}'_{\mathscr{K}_i\mathscr{K}_i}) \boldsymbol{D}_{\mathscr{K}_i}(t). \tag{36}$$

之后,我们要考虑 $k \in \mathscr{K}_i$ 及在一个固定的 $t \geqslant 0$ 时,$Z_k(t) = 0$ 和 $Z_k(t) \neq 0$ 的情况,因而,我们引入下面的符号

$$\mathscr{K}^{o} = \{k: Z_k(t) = 0, k \in \mathscr{K}_i\}, \quad \mathscr{K}^{n} = \{k: Z_k(t) \neq 0, k \in \mathscr{K}_i\}. \tag{37}$$

注意到我们的网络是开型的, 因而我们可以求解(36) 式并得到

$$D_{\mathscr{T}_i^o}(t)=Q_{\mathscr{T}_i^o}(Z_{\mathscr{T}_i^o}(0)+\alpha_{\mathscr{T}_i^o}t+P'_{\mathscr{T}_i^n\mathscr{T}_i^o}D_{\mathscr{T}_i^n}(t)-Z_{\mathscr{T}_i^o}(t)),\tag{38}$$

这里, 由(37) 式, 有 $Q_{\mathscr{T}_i^o}=(I-P'_{\mathscr{T}_i^o\mathscr{T}_i^o})^{-1}$ 和 $Z_{\mathscr{T}_i^o}(t)=0$. 再由交通方程(3) 式与路由矩阵 P 的结构, 我们有

$$\lambda_{\mathscr{T}_i}=\alpha_{\mathscr{T}_i}+P'_{\mathscr{T}_i\mathscr{T}_i}\lambda_{\mathscr{T}_i}.\tag{39}$$

求解(39) 式, 我们得到

$$\lambda_{\mathscr{T}_i^o}=Q_{\mathscr{T}_i^o}(\alpha_{\mathscr{T}_i^o}+P'_{\mathscr{T}_i^n\mathscr{T}_i^o}\lambda_{\mathscr{T}_i^n}).\tag{40}$$

下面我们进行归纳证明. 在归纳证明的第1步, 我们考虑 $i=h$ 的情况, 并证明在有限时间内, 流体水平 $Z_k(t)\ (k\in\mathscr{T}_h)$ 达到0, 这里 $\mathscr{T}_h$ 是各网站中优先权等级最高的工作类相应的指标集. 具体的, 我们证明一个替代的结论, 即证明存在合适的 $\delta_h\geqslant 0$, 使得对于所有 $t\geqslant\delta_h$, 有 $f_h(t)=0$, 因而, 对于所有的 $t\geqslant\delta_h$ 和 $k\in\mathscr{T}_h$, 有 $Z_k(t)=0$. 事实上, 在一个时刻 t, 当 $f_h(t)>0$ 时, 则至少存在一个 $k\in\mathscr{T}_h$, 使得 $Z_k(t)>0$, 故 $\mathscr{T}_h^n$ 是非空的. 这样, 由(34) 式, (36) 式和(39) 式, 可得对每一个规则点 t, 我们有

$$\begin{aligned}
\dot f_h(t)={}&e'M_{\mathscr{T}_h}(\lambda_{\mathscr{T}_h}-\dot D_{\mathscr{T}_h}(t))=\\
&e'_{\mathscr{T}_h^o}M_{\mathscr{T}_h^o}(\lambda_{\mathscr{T}_h^o}-\dot D_{\mathscr{T}_h^o}(t))+e'_{\mathscr{T}_h^n}M_{\mathscr{T}_h^n}(\lambda_{\mathscr{T}_h^n}-\dot D_{\mathscr{T}_h^n}(t))=\\
&e'_{\mathscr{T}_h^o}M_{\mathscr{T}_h^o}Q_{\mathscr{T}_h^o}P'_{\mathscr{T}_h^n\mathscr{T}_h^o}(\lambda_{\mathscr{T}_h^n}-\dot D_{\mathscr{T}_h^n}(t))+e'_{\mathscr{T}_h^n}M_{\mathscr{T}_h^n}(\lambda_{\mathscr{T}_h^n}-\dot D_{\mathscr{T}_h^n}(t))=\\
&(e'_{\mathscr{T}_h^o}M_{\mathscr{T}_h^o}Q_{\mathscr{T}_h^o}P'_{\mathscr{T}_h^n\mathscr{T}_h^o}M^{-1}_{\mathscr{T}_h^n}+e'_{\mathscr{T}_h^n})(M_{\mathscr{T}_h^n}\lambda_{\mathscr{T}_h^n}-M_{\mathscr{T}_h^n}\dot D_{\mathscr{T}_h^n}(t))=\\
&\sum_{k\in\mathscr{T}_h^n}(1+(e'_{\mathscr{T}_h^o}M_{\mathscr{T}_h^o}Q_{\mathscr{T}_h^o}P'_{\mathscr{T}_h^n\mathscr{T}_h^o}M^{-1}_{\mathscr{T}_h^n})_k)(\lambda_k m_k-m_k\dot D_k(t))=\\
&-\sum_{k\in\mathscr{T}_h^n}(1+(e'_{\mathscr{T}_h^o}M_{\mathscr{T}_h^o}Q_{\mathscr{T}_h^o}P'_{\mathscr{T}_h^n\mathscr{T}_h^o}M^{-1}_{\mathscr{T}_h^n})_k)\left(\sum_{l\in\mathscr{C}(s(k))\setminus\{k\}}\lambda_l m_l\right),
\end{aligned}\tag{41}$$

这里 $e_{\mathscr{T}_h^o}$ 和 $e_{\mathscr{T}_h^n}$ 是各分量为1, 其维数分别相应于集 $\mathscr{T}_h^o$ 和 $\mathscr{T}_h^n$ 中指标的数目; 在上述的第3个等式中, 我们用了(38) 式、(40) 式和 $Z_{\mathscr{T}_i^o}(t)=0$ 的事实; 在第5个等式中, $(\cdot)_k$ 表示相应向量的第 k 个分量; 在最后一个等式中, 我们用了满负荷交通条件(8) 和流体模型性质(31) 式及(28) 式推断出 $\dot D_l(t)=0\ (l\in\mathscr{C}(s(k))\setminus\{k\})$ 且 $m_k\dot D_k(t)=1$, 这里 k 是网站 $s(k)$ 中优先权等级最高的那个工作类的指标, $\mathscr{C}(s(k))\setminus\{k\}$ 是网站 $s(k)$ 中除了工作类 k 以外的所有工作类的指标集. 由于 $e'_{\mathscr{T}_h^o}M_{\mathscr{T}_h^o}Q_{\mathscr{T}_h^o}P'_{\mathscr{T}_h^n\mathscr{T}_h^o}M^{-1}_{\mathscr{T}_h^n}\geqslant 0$, 因此

$$\delta'_h\equiv\min_{\mathscr{T}_h^n\in\mathscr{G}_h}\left\{\sum_{k\in\mathscr{T}_h^n}(1+(e'_{\mathscr{T}_h^o}M_{\mathscr{T}_h^o}Q_{\mathscr{T}_h^o}P'_{\mathscr{T}_h^n\mathscr{T}_h^o}M^{-1}_{\mathscr{T}_h^n})_k)\left(\sum_{l\in\mathscr{C}(s(k))\setminus\{k\}}\lambda_l m_l\right)\right\}>0,\tag{42}$$

这里 $\mathscr{G}_h$ 是由 $\mathscr{T}_h$ 中工作类指标所有可能的非空集合组成的集类. 这样, 由(41) 式和(42) 式, 可得下述结论: 只要 $f_h(t)>0$ 就有下式成立 $\dot f_h(t)\leqslant-\delta'_h$. 下面定义

$$\delta_h\equiv f_h(0)/\delta'_h,\tag{43}$$

则既然 $f_h(\cdot)$ 是绝对连续的, 因而对所有 $t\geqslant\delta_h$, 有 $f_h(t)=0$ 成立, 并且对所有 $t\geqslant\delta_h$ 和 $k\in\mathscr{T}_h$, 有 $Z_k(t)=0$ 成立.

在归纳证明的第2步中, 我们假设对一个固定的 $i\in\{2,\ldots,h-1\}$, 命题对 $i+1\leqslant j\leqslant h$ 是成立的, 即存在一个非负常数 δ_{i+1} 使得对 $t\geqslant\delta_{i+1}$ 和 $k\in\bigcup_{j=i+1}^h\mathscr{T}_j$ (集 $\mathscr{T}_j$ 的并), 有 $Z_k(t)=0$ 成立. 这样, 我们可以推断在所有 $t\geqslant\delta_{i+1}$ 的规则点上, 有 $\dot Z_k(t)=0\ (k\in\bigcup_{j=i+1}^h\mathscr{T}_j)$ 成立.

因而，由(36)式和(39)式，我们得到对 $i+1\leqslant j\leqslant h$，有下式成立

$$\dot{D}_{\mathcal{T}_j}(t)=(I-P'_{\mathcal{T}_j\mathcal{T}_j})^{-1}\alpha_{\mathcal{T}_j}=\lambda_{\mathcal{T}_j}. \tag{44}$$

在归纳证明的第 3 步中，我们证明对于第 2 步中固定的 i，命题对 $i\leqslant j\leqslant h$ 成立，即存在一个非负常数 δ_i，使得对于所有 $t\geqslant\delta_i$ 和 $k\in\bigcup_{j=i}^{h}\mathcal{T}_j$，有 $Z_k(t)=0$ 成立．与前面的解释类似，我们证明存在一个合适的常数 $\delta_i\geqslant 0$ 使得对所有 $t\geqslant\delta_i$ 和 $i\leqslant j\leqslant h$，有 $f_j(t)=0$ 成立．为达此目标，我们考虑 $t\geqslant\delta_{i+1}$．由第 2 步中的归纳假设，我们知道对于 $i+1\leqslant j\leqslant h$，有 $f_j(t)=0$ 成立．这样，我们仅需讨论当 $t\geqslant\delta_{i+1}$ 时 $f_i(t)$ 的情况．由于在一个时刻 t，至少存在一个 $k\in\mathcal{T}_i$ 使得 $Z_k(t)>0$，因而当 $f_i(t)>0$ 时，$\mathcal{T}_i^n$ 非空．这样，由(34)式，(36)式和(39)式可得，对每一个规则点 t，有下式成立：

$$\begin{aligned}
\dot{f}_i(t)=&\ e'M_{\mathcal{T}_i}(\lambda_{\mathcal{T}_i}-\dot{D}_{\mathcal{T}_i}(t))=\\
&e'_{\mathcal{T}_i^o}M_{\mathcal{T}_i^o}(\lambda_{\mathcal{T}_i^o}-\dot{D}_{\mathcal{T}_i^o}(t))+e'_{\mathcal{T}_i^n}M_{\mathcal{T}_i^n}(\lambda_{\mathcal{T}_i^n}-\dot{D}_{\mathcal{T}_i^n}(t))=\\
&e'_{\mathcal{T}_i^o}M_{\mathcal{T}_i^o}Q_{\mathcal{T}_i^o}P'_{\mathcal{T}_i^n\mathcal{T}_i^o}(\lambda_{\mathcal{T}_i^n}-\dot{D}_{\mathcal{T}_i^n}(t))+e'_{\mathcal{T}_i^n}M_{\mathcal{T}_i^n}(\lambda_{\mathcal{T}_i^n}-\dot{D}_{\mathcal{T}_i^n}(t))=\\
&(e'_{\mathcal{T}_i^o}M_{\mathcal{T}_i^o}Q_{\mathcal{T}_i^o}P'_{\mathcal{T}_i^n\mathcal{T}_i^o}M^{-1}_{\mathcal{T}_i^n}+e'_{\mathcal{T}_i^n})(M_{\mathcal{T}_i^n}\lambda_{\mathcal{T}_i^n}-M_{\mathcal{T}_i^n}\dot{D}_{\mathcal{T}_i^n}(t))=\\
&\sum_{k\in\mathcal{T}_i^n}(1+(e'_{\mathcal{T}_i^o}M_{\mathcal{T}_i^o}Q_{\mathcal{T}_i^o}P'_{\mathcal{T}_i^n\mathcal{T}_i^o}M^{-1}_{\mathcal{T}_i^n})_k)(\lambda_k m_k-m_k\dot{D}_k(t))=\\
&-\sum_{k\in\mathcal{T}_i^n}(1+(e'_{\mathcal{T}_i^o}M_{\mathcal{T}_i^o}Q_{\mathcal{T}_i^o}P'_{\mathcal{T}_i^n\mathcal{T}_i^o}M^{-1}_{\mathcal{T}_i^n})_k)\left(\sum_{l\in\mathcal{C}(s(k))\setminus\mathcal{H}_i(s(k))}\lambda_l m_l\right),
\end{aligned} \tag{45}$$

这里我们需要给出一些有关上述最后一个等式的解释．$\mathcal{H}_i(s(k))$ 是网站 $s(k)$ 中优先权等级至少为 i 的工作类的指标集，$\mathcal{C}(s(k))\setminus\mathcal{H}_i(s(k))$ 表示网站 $s(k)$ 中优先权等级低于 i 的那些工作类的指标集．由于在网站 $s(k)$ 中，工作类 k 具有优先权等级 i 且是非空的，这样由(28)式和 SBP 流体模型性质(31)式，我们知道对 $l\in\mathcal{C}(s(k))\setminus\mathcal{H}_i(s(k))$，有 $\dot{D}_l(t)=0$ 成立且

$$\sum_{l\in\mathcal{H}_i(s(k))}m_l\dot{D}_l(t)=1. \tag{46}$$

从而，由满负荷交通条件(8)，第 2 步中的规纳假设(44)式，及(46)式可得，对 $k\in\mathcal{T}_i^n$，有下式成立：

$$\lambda_k m_k-m_k\dot{D}_k(t)=\sum_{l\in\mathcal{H}_i(s(k))}(\lambda_l m_l-m_l\dot{D}_l(t))=-\sum_{l\in\mathcal{C}(s(k))\setminus\mathcal{H}_i(s(k))}\lambda_l m_l.$$

因而我们为(45)式提供了一个证明．下面因为 $e'_{\mathcal{T}_i^o}M_{\mathcal{T}_i^o}Q_{\mathcal{T}_i^o}P'_{\mathcal{T}_i^n\mathcal{T}_i^o}M^{-1}_{\mathcal{T}_i^n}\geqslant 0$，从而有

$$\delta'_i\equiv\min_{\mathcal{T}_i^n\in\mathcal{G}_i}\left\{\sum_{k\in\mathcal{T}_i^n}(1+(e'_{\mathcal{T}_i^o}M_{\mathcal{T}_i^o}Q_{\mathcal{T}_i^o}P'_{\mathcal{T}_i^n\mathcal{T}_i^o}M^{-1}_{\mathcal{T}_i^n})_k)\left(\sum_{l\in\mathcal{C}(s(k))\setminus\mathcal{H}_i(s(k))}\lambda_l m_l\right)\right\}>0, \tag{47}$$

这里 $\mathcal{G}_i$ 是由 $\mathcal{T}_i$ 中工作类指标所有可能的非空集合组成的集类．那么，由(45)式和(47)式，可得对 $t\in[\delta_{i+1},\infty)$，有下式成立：$\dot{f}_i(t)\leqslant-\delta'_i$ 若 $f_i(t)>0$．定义

$$\delta_i\equiv\delta_{i+1}+f_i(\delta_{i+1})/\delta'_i, \tag{48}$$

则既然 $f_i(\cdot)$ 是绝对连续的，那么对 $t\geqslant\delta_i$，有 $f_i(t)=0$ 成立，因而对所有 $t\geqslant\delta_i$，有 $Z_k(t)=0(k\in\bigcup_{j=i}^{h}\mathcal{T}_j)$ 成立．

在定理该步证明的最后，我们取由上面归纳证明部分得到的 δ_2，即令 $\delta=\delta_2$，则对于 $t\geqslant\delta$，有 $f(t)=0$ 成立，因此对所有的 $t\geqslant\delta$ 和 $k\in\bigcup_{j=2}^{h}\mathcal{T}_j$，有 $Z_k(t)=0$ 成立，这里 $f(t)$ 由(35)式给出．

第 2 步　我们证明在由定理证明的第一步中得到的时刻 δ 之后，流体水平 $Z_k(t)(k\in\mathcal{T}_1)$

保持常数. 为了证明此结论, 我们在时刻 δ 处重新开启流体模型的解如下:

$$
\begin{aligned}
&(\boldsymbol{A}^{\delta}(t), \boldsymbol{D}^{\delta}(t), \boldsymbol{T}^{\delta}(t), \boldsymbol{W}^{\delta}(t), \boldsymbol{Y}^{\delta}(t), \boldsymbol{Z}^{\delta}(t)) = \\
&\quad (\boldsymbol{A}(t+\delta) - \boldsymbol{A}(\delta), \boldsymbol{D}(t+\delta) - \boldsymbol{D}(\delta), \boldsymbol{T}(t+\delta) - \\
&\quad \boldsymbol{T}(\delta), \boldsymbol{W}(t+\delta), \boldsymbol{Y}(t+\delta) - \boldsymbol{Y}(\delta), \boldsymbol{Z}(t+\delta)),
\end{aligned} \tag{49}
$$

这里 $t \geqslant 0$. 可以看出(49) 式的左边仍旧是一个初值为 $\boldsymbol{Z}^{\delta}(0) = \boldsymbol{Z}(\delta)$ 的流体模型的解. 由(36) 式, (40) 式和 $\boldsymbol{Z}^{\delta}_{\mathscr{T}_i}(t) = 0\ (i \in \{2, \ldots, h\})$, 可得

$$
\boldsymbol{D}^{\delta}_{\mathscr{T}_i}(t) = (\boldsymbol{I} - \boldsymbol{P}'_{\mathscr{T}_i\mathscr{T}_i})^{-1}\boldsymbol{Z}^{\delta}_{\mathscr{T}_i}(0) + \lambda_{\mathscr{T}_i} t. \tag{50}
$$

因而, 我们有

$$
\begin{aligned}
\boldsymbol{Z}^{\delta}_{\mathscr{T}_1}(t) &= \boldsymbol{Z}^{\delta}_{\mathscr{T}_1}(0) + \alpha_{\mathscr{T}_1} t - (\boldsymbol{I} - \boldsymbol{P}'_{\mathscr{T}_1\mathscr{T}_1})\boldsymbol{M}^{-1}_{\mathscr{T}_1}\left(\boldsymbol{e}_{\mathscr{T}_1} t - \sum_{i=2}^{h}\boldsymbol{T}^{\delta}_{\mathscr{T}_i}(t) - \boldsymbol{Y}^{\delta}(t)\right) = \\
&\quad \boldsymbol{X}^{\delta}_{\mathscr{T}_1}(t) + \left(\alpha_{\mathscr{T}_1} - (\boldsymbol{I} - \boldsymbol{P}'_{\mathscr{T}_1\mathscr{T}_1})\boldsymbol{M}^{-1}_{\mathscr{T}_1}\left(\boldsymbol{e}_{\mathscr{T}_1} - \sum_{i=2}^{h}\boldsymbol{M}_{\mathscr{T}_i}\lambda_{\mathscr{T}_i}\right)\right) t + \\
&\quad (\boldsymbol{I} - \boldsymbol{P}'_{\mathscr{T}_1\mathscr{T}_1})\boldsymbol{M}^{-1}_{\mathscr{T}_1}\boldsymbol{Y}^{\delta}(t) = \boldsymbol{X}^{\delta}_{\mathscr{T}_1}(t) + (\boldsymbol{I} - \boldsymbol{P}'_{\mathscr{T}_1\mathscr{T}_1})\boldsymbol{M}^{-1}_{\mathscr{T}_1}\boldsymbol{Y}^{\delta}(t),
\end{aligned} \tag{51}
$$

这里

$$
\boldsymbol{X}^{\delta}_{\mathscr{T}_1}(t) \equiv \boldsymbol{Z}^{\delta}_{\mathscr{T}_1}(0) + (\boldsymbol{I} - \boldsymbol{P}'_{\mathscr{T}_1\mathscr{T}_1})\boldsymbol{M}^{-1}_{\mathscr{T}_1}\sum_{i=2}^{h}\boldsymbol{M}_{\mathscr{T}_i}(\boldsymbol{I} - \boldsymbol{P}'_{\mathscr{T}_i\mathscr{T}_i})^{-1}\boldsymbol{Z}^{\delta}_{\mathscr{T}_i}(0).
$$

在(51) 式的第 1 个等式中, 我们用了(36) 式、(28) 式和(26) 式; 在第 2 个等式中, 我们用了(28) 式和(50) 式; 第 3 个等式是由满负荷交通条件(8) 与交通方程(39) 得到. 进一步地, 我们有下列关系成立:

$$
\int_0^{\infty} \boldsymbol{Z}_k^{\delta}(t)\,\mathrm{d}Y_k^{\delta}(t) = 0, \qquad k \in \mathscr{K}, \tag{52}
$$

$$
Y_k^{\delta}(\cdot)\ \text{为非减序列},\ Y_k^{\delta}(0) = 0, \qquad k \in \mathscr{K}. \tag{53}
$$

因而(51) 式到(53) 式构成一个所谓的连续确定性的规则(Skorohod) 问题. 既然我们的网络是开型的, 则矩阵 $(\boldsymbol{I} - \boldsymbol{P}'_{\mathscr{T}_1\mathscr{T}_1})$ 是完全S 的(比如: 参见文献 Bernard 和 Kharroubi[18], Harrison 和 Reiman[19]). 这样可知(51) 式中的反射矩阵 $(\boldsymbol{I} - \boldsymbol{P}'_{\mathscr{T}_1\mathscr{T}_1})\boldsymbol{M}^{-1}_{\mathscr{T}_1}$ 是完全S 的. 因而, 由文献 Bernard 和 Kharroubi[18] 中的一个振荡不等式及(51) 式可得: 对任何 $0 \leqslant t_1 < t_2 < \infty$, 有

$$
\begin{aligned}
&\mathrm{Osc}(\boldsymbol{Z}^{\delta}_{\mathscr{T}_1}(\cdot), [t_1, t_2]) \equiv \sup\left\{\|\boldsymbol{Z}^{\delta}_{\mathscr{T}_1}(t) - \boldsymbol{Z}^{\delta}_{\mathscr{T}_1}(s)\| : t_1 \leqslant s < t \leqslant t_2\right\} \leqslant \\
&\quad \kappa\,\mathrm{Osc}(\boldsymbol{X}^{\delta}_{\mathscr{T}_1}(\cdot), [t_1, t_2]) = 0
\end{aligned} \tag{54}
$$

成立, 这里 κ 是一个正常数, 它仅依赖于(51) 式中的反射矩阵$(\boldsymbol{I} - \boldsymbol{P}'_{\mathscr{T}_1\mathscr{T}_1})\boldsymbol{M}^{-1}_{\mathscr{T}_1}$. 从而有下式成立: $\boldsymbol{Z}_{\mathscr{T}_1}(\delta + t) = \boldsymbol{Z}_{\mathscr{T}_1}(\delta)$ 对所有 $t \geqslant 0$. 因此, 我们完成了定理的第 2 步证明.

最后, 令 $\boldsymbol{Z}_{\mathscr{T}_1}(\infty) = \boldsymbol{Z}_{\mathscr{T}_1}(\delta)$ 和 $\boldsymbol{Z}_{\mathscr{T}_i}(\infty) = 0\ (i \in \{2, \ldots, h\})$, 则有 $\boldsymbol{Z}(t) = \boldsymbol{Z}(\infty)\ (t \geqslant \delta)$. 若我们能够证明$\sup_{\|\boldsymbol{Z}(0)\|=1}\delta < \infty$, 则我们可完成命题2 的证明. 事实上, 该论断也可以由归纳法证明. 首先, 对于 $j = h$, 由(43) 式和(34) 式可得: $\sup_{\|\boldsymbol{Z}(0)\|=1}\delta_h < \infty$. 其次, 假设对一个固定的 $i \in \{2, \ldots, h-1\}$, 有 $\sup_{\|\boldsymbol{Z}(0)\|=1}\delta_{i+1} < \infty$成立. 再者, 我们考虑 $j = i$ 的情况: 既然 $\boldsymbol{Z}(\cdot)$ 是Lipschitz 连续的, 则有下式成立

$$
\|\boldsymbol{Z}(\delta_{i+1})\| \leqslant \|\boldsymbol{Z}(\delta_{i+1}) - \boldsymbol{Z}(0)\| + \|\boldsymbol{Z}(0)\| = N\delta_{i+1} + 1, \tag{55}
$$

这里 N 是一个仅依赖于$(\alpha, \boldsymbol{M}, \boldsymbol{P})$ 的常数. 因而, 由(48) 式、(34) 式和归纳假设, 可知$\sup_{\|\boldsymbol{Z}(0)\|=1}\delta_i < \infty$成立. □

定理 1 的证明 基于引理 1 和命题 2, 再由文献 Bramson 和 J. G. Dai[9] 中的有关结论可知我们的主要定理是成立的. □

[参 考 文 献]

[1] Dai W. A heavy traffic limit theorem for queueing networks with finite capacity[A]. Presentation With Preprint at INFORMS Applied Probability Conference[C]. Atlanta, USA, 1995.

[2] Dai W. Brownian approximations for queueing networks with finite buffers: modeling, heavy traffic analysis and numerical implementations[D]. Ph D Thesis. School of Mathematics, Georgia Institute of Technology, 1996. Aslo published in UMI Dissertation Services, A Bell & Howell Company, 300 N. Zeeb Road, Ann Arbor, Michican 48106, USA, 1997.

[3] Dai J G, Dai W. A heavy traffic limit theorem for a class of open queueing networks with finite buffers[J]. Queueing Systems, 1999, **32**(1/3), 5-40.

[4] Reiman M I. Open queueing networks in heavy traffic [J]. Mathematics of Operations Research, 1984, **9**(3): 441-458.

[5] Bramson M. State space collapse with application to heavy traffic limits for multiclass queueing networks[J]. Queueing Systems, 1998, **30**(1/2): 89-148.

[6] Bramson M. State space collapse for queueing networks[A]. Proceedings of the International Congress of Mathematicians [C]. Bielefeld, Germany: Documenta mathematica, Vol Ⅲ. 1998, , 213-222.

[7] Williams R J. Diffusion approximations for open multiclass queueing networks: sufficient conditions involving state space collapse[J]. Queueing Systems: Theory and Applications, 1998, **30**(1/2): 27-88.

[8] Williams R J. Reflecting diffusions and queueing networks[A]. Proceedings of the International Congress of Mathematicians [M]. Bielefeld, Germany: Documenta mathematica, Vol Ⅲ. 1998, 321-330.

[9] Bramson M, Dai J. G. Heavy traffic limits for some queueing networks[J]. Annals of Applied Probability, 2001, **11**(1): 49-90.

[10] Chen H, Zhang H. A sufficient condition and a necessary condition for the diffusion approximations of multiclass queueing networks under priority service displines[J]. Queueing Systems, 2000, **34**(1/4): 237-268.

[11] Chen H, Zhang H. Diffusion approximations for some multiclass queueing networks with FIFO service disciplines[J]. Mathematics of Operations Research, 2000, **25**(4): 679-707.

[12] Harrison J M, Williams R J. Multidimensional reflected Brownian motions having exponential stationary distributions[J]. Annals of Probability, 1987, **15**(1): 115-137.

[13] Dai J G, Harrison J M. Reflected Brownian motion in an orthant: numerical methods for steady-state analysis[J]. Annals of Applied Probability, 1992, **2**(1): 65-86.

[14] Shen X, Chen H, Dai J G, et al. The finite element method for computing the stationary distribution of an SRBM in a hypercube with applications to finite buffer queueing networks[J]. Queueing Systems, 2002, **42**(1): 33-62.

[15] Dai J G, Wang Y. Nonexistence of Brownian models of certain multicalss queueing networks[J]. Queueing Systems, 1993, **13**(1/3): 41-46.

[16] Williams R J. An invariance principle for semimartingale reflecting Brownian motions in an orthant [J]. Queueing Systems, 1998, **30**(1/2): 5-25.

[17] Ethier S N, Kurtz T G. Markov Processes: Charaterization and Convergence[M]. New York: Wiley, 1986.

[18] Bernard A, Kharroubi A El. Regulation deterministes et stochastiques dans le premier "orthant" de R^n [J]. Stochastics Stochastics Rep, 1991, **34**(3/4): 149-167.

[19] Harrison J M, Reiman M I. Reflected Brownian motion on an orthant[J]. Annals of Probability, 1981, **9**(2): 302-308.

Diffusion Approximations for Multiclass Queueing Networks Under Preemptive Priority Service Discipline

DAI Wan-yang

(Department of Mathematics, Nanjing University, Nanjing 210093, P.R. China)

Abstract: A heavy traffic limit theorem is proved to justify diffusion approximations for multiclass queueing networks under preemptive priority service discipline and provide effective stochastic dynamical models for the systems. Such queueing networks typically appear in high-speed integrated services packet networks in telecommunication system. In the network, there are a number of packet traffic types. Each type needs a number of job classes (stages) of processing and each type of jobs is assigned the same priority rank at every station where it possibly receives service. Moreover, there is no inter-routing among different traffic types throughout the entire network.

Key words: multiclass queueing network; preemptive priority; heavy traffic; semimartingale reflecting Brownian motion; fluid model; diffusion approximation; Lyapunov function