

Análise de filas $M | M | n$ com servidores lentos e usuários não informados

Fabricio B. Cabral

Resumo— Neste trabalho, analisamos sistemas que podem ser modelados por filas $M | M | n$ com servidores heterogêneos e usuários não informados. As equações de equilíbrio são resolvidas. Mostramos que existe um valor limiar de taxa de chegada, abaixo do qual, o sistema sem o servidor lento é melhor, e, acima do qual, o sistema com o servidor lento é melhor.

Palavras-Chave— sistemas de filas, sistemas de filas com vários servidores, servidores heterogêneos, processos Markovianos, política de limiar, servidor lento.

Abstract— Systems that can be modelled by $M | M | n$ queues with heterogeneous servers and non informed customers are studied. The balance equations are solved. A threshold result for a system with an arbitrary number n of servers is presented, i.e. it is shown that there is a value of arrival rate below which the slow server should not be used and above which it should be used.

Keywords— queues; multiserver queues; heterogeneous servers; Markovian processes; threshold policy; slow server.

I. INTRODUÇÃO

Quando operamos um sistema, é natural nos questionarmos se devemos ou não usar um servidor lento. Consideremos um sistema com: n servidores operando em paralelo; tempos de serviço com distribuição exponencial; exatamente um linha de chegada de usuários; processo de chegada Poisson com taxa de chegada λ ; usuários na fila de espera atendidos de acordo com a política *primeiro a chegar, primeiro a ser servido*. Este sistema foi estudado por Rubinovitch [1] para o caso de $n = 2$ servidores. As noções de usuários informados e não informados foram então definidas. Em ambos os casos, se apenas um servidor estiver ocioso, este atenderá o próximo usuário. Se mais do que um usuário estiver ocioso, o mais rápido atenderá o próximo usuário no caso de usuários informados. No caso de usuários não informados, um servidor será escolhido aleatoriamente através de um ensaio de Bernoulli, e, conseqüentemente, cada servidor terá a mesma probabilidade de ser escolhido.

Para um sistema com $n = 2$ servidores e usuários não informados, com taxas de serviço médias (μ_1, μ_2) , os resultados de Rubinovitch estabelecem que existe um número ρ^c tal que, se a intensidade de tráfego estiver abaixo deste número, o servidor

lento não deverá ser utilizado. Este número é dado por uma função $\rho^c = f(\mu_1, \mu_2)$.

Para um sistema com $n = 2$ servidores e usuários informados, por outro lado, os resultados de Rubinovitch estabelecem um valor de limiar apenas para os casos onde tenhamos $\mu_2/\mu_1 > 1/2$. Para usuários informados, o único resultado conhecido é o de Rubinovitch. Para usuários informados que também conheçam o tamanho da fila de espera, existe um resultado para sistemas com $n = 2$ servidores, o qual mostra que uma política de limiar minimiza o tempo médio de serviço [4] ([5],[6]). Mais especificamente, dadas a taxa de chegada λ e as taxas de serviço médias (μ_1, μ_2) , existe um número $L = l(\lambda, \mu_1, \mu_2)$ tal que um usuário usará o servidor lento se, e somente se, o comprimento da fila de espera exceder L .

Mostramos que um resultado de limiar pode ser estabelecido, também, para o caso de sistemas com n servidores e usuários não informados. Note que se tivéssemos apenas dois servidores, o sistema poderia ser modelado por uma fila $M | M | 2$ [3]. Assim, para estudarmos o fenômeno de limiar para sistemas com n servidores e usuários não informados, modelamos o sistema por uma fila $M | M | n$ com servidores heterogêneos e usuários não informados. Derivamos as distribuições de probabilidade relevantes utilizando as equações de equilíbrio. O inverso da probabilidade de que o sistema esteja vazio é dado por um polinômio em λ . Relações entre os coeficientes deste polinômio para o sistema com $n - 1$ servidores e os coeficientes correspondentes para o sistema com n servidores são derivadas. Estas relações são úteis na determinação das derivadas da razão entre estes polinômios. O resultado desejado é, assim, derivado como um caso particular de um resultado mais geral.

II. FILAS $M | M | n$ COM SERVIDORES HETEROGÊNEOS E USUÁRIOS NÃO INFORMADOS

O sistema para o caso de usuários não informados e sob hipóteses Markovianas pode ser modelado por um processo de Markov. Os estados deste processo são associados às $(n + 1)$ -uplas $(x_1, x_2, \dots, x_n, x_f)$, onde x_i , para $i = 1$ a n , é uma variável Booleana, a qual é igual a um se, e somente se, o servidor i está ocupado e x_f denota o comprimento da fila do sistema. Indexaremos os estados do sistema por

$$i = x_1 2^{n-1} + x_2 2^{n-2} + \dots + x_n 2^0 + x_f,$$

onde

$$x_f > 0 \Rightarrow x_1 = x_2 = \dots = x_n = 1.$$

TABELA I
 TABELA DE NOTAÇÕES

N_k^n	conjunto de estados com k usuários para o sistema com n servidores
S_i^n	conjunto de servidores ocupados quando o sistema com n servidores está no estado i
$g^n(i, j)$	estados vizinhos: dois estados i e j que diferem por exatamente uma componente da $(n+1)$ -upla para o sistema com n servidores para dois estados i e j que diferem em exatamente uma componente da $(n+1)$ -upla (estados vizinhos), $g^n(i, j)$ denota o índice daquela componente, para o sistema com n servidores
D^n	conjunto de pares ordenados de estados vizinhos para o sistema com n servidores
D_i^n	conjunto de estados vizinhos do estado i para o sistema com n servidores
μ_i	taxa de serviço média do servidor i
λ	$\mu_1 \geq \mu_2 \geq \dots \geq \mu_n$ taxa de chegada
N^n	número médio de usuários no sistema com n servidores
T^n	tempo médio de permanência de um usuário no sistema com n servidores
μ^n	$\sum_{i=1}^n \mu_i$
p_i^n	probabilidade em regime estacionário p_i^n que o sistema com n servidores esteja no estado i
P_k^n	probabilidade que o número de usuários no sistema com n servidores seja k
S^n	$\{\mu_1, \mu_2, \dots, \mu_n\}$
S^{-n}	$\{\mu_1^{-1}, \mu_2^{-1}, \dots, \mu_n^{-1}\}$
$P(n, k)$	denota a soma de todos os produtos distintos de k elementos do conjunto S^{-n}
$Q(n, k)$	denota a soma de todos os produtos distintos de k elementos do conjunto S^n
$Q(S, k)$	se o primeiro argumento n é um número denota a soma de todos os produtos distintos de k elementos do conjunto S se o primeiro argumento S é um conjunto

A. Notação

As notações definidas na tabela I são introduzidas de modo a possibilitar o estudo do comportamento em regime estacionário do processo Markoviano.

B. Equações de equilíbrio

As probabilidades em regime estacionário p_i^n , probabilidade de que o sistema esteja no estado i , é dada pela solução do sistema linear abaixo:

$$\left(\lambda + \sum_{k \in S_i^n} \mu_k\right) p_i^n - \sum_{j \in N_{|S_i^n|-1}^n \cap D_i^n} \frac{\lambda}{n - (|S_i^n| - 1)} p_j^n - \sum_{j \in N_{|S_i^n|+1}^n \cap D_i^n} \mu_{g^n(i,j)} p_j^n = 0,$$

$$0 \leq i \leq 2^n - 1;$$

$$\lambda p_{i-1}^n = \mu^n p_i^n, \quad i \geq 2^n;$$

$$\sum_{i=0}^{\infty} p_i^n = 1.$$

A solução deste sistema é dada pelo próximo teorema.

Teorema 2.1: A distribuição de probabilidade de uma fila $M | M | n$ com servidores heterogêneos e usuários não informados é dada por

$$p_i^n = \frac{(n - |S_i^n|)!}{n!} \frac{\lambda^{|S_i^n|}}{\prod_{j \in S_i^n} \mu_j} p_0^n, \quad 0 \leq i \leq 2^n - 1.$$

A probabilidade de que o número de usuários no sistema seja k , $P_k^n = \sum_{i \in N_k^n} p_i^n$, é dada por $P_k^n = \lambda^k K_k^n p_0^n$, onde

$$K_k^n = \frac{(n-k)! Q(n, n-k)}{n! Q(n, n)} = \frac{(n-k)! P(n, k)}{n!},$$

$$1 \leq k \leq n-1;$$

$$K_k^n = \frac{K_{n-1}^n}{\mu^n}, \quad k \geq n;$$

$$P_0^n = p_0^n = \left(\sum_{k=0}^{\infty} \lambda^k K_k^n \right)^{-1}.$$

Prova. Substituindo os valores de p_i^n e de p_j^n nas equações de equilíbrio e notando que

$$\begin{aligned} & \sum_{j \in N_{|S_i^n|-1}^n \cap D_i^n} \frac{\lambda}{n - (|S_i^n| - 1)} p_j^n = \\ & \frac{\lambda}{n - (|S_i^n| - 1)} \sum_{j \in N_{|S_i^n|-1}^n \cap D_i^n} \frac{\lambda^{|S_i^n|-1} (n - (|S_i^n| - 1))!}{n! \prod_{l \in S_j^n} \mu_l} p_0^n \\ & = \frac{\lambda^{|S_i^n|}}{n!} \sum_{j \in N_{|S_i^n|-1}^n \cap D_i^n} \frac{(n - |S_i^n|)!}{\prod_{l \in S_j^n} \mu_l} p_0^n \\ & = p_i^n \sum_{k \in S_i^n} \mu_k \end{aligned}$$

e

$$\begin{aligned} & \sum_{j \in N_{|S_i^n|+1}^n \cap D_i^n} \mu_{g^n(i,j)} p_j^n = \\ & \lambda^{|S_i^n|+1} \frac{(n - |S_i^n|)(n - (|S_i^n| + 1))!}{n! \prod_{l \in S_i^n} \mu_l} p_0^n \\ & = \lambda^{|S_i^n|+1} \frac{(n - |S_i^n|)!}{n! \prod_{l \in S_i^n} \mu_l} p_0^n \\ & = \lambda p_i^n, \end{aligned}$$

a tese segue. \square

III. PROPRIEDADES DOS COEFICIENTES E DAS DERIVADAS

Desejamos comparar o comportamento do sistema com todos os servidores com o comportamento do sistema sem o servidor mais lento. Primeiramente, derivamos as propriedades dos coeficientes e, em seguida, derivamos as propriedades das derivadas.

A. Propriedades dos coeficientes

Nesta seção, restringir-nos-emos ao caso de servidores heterogêneos, i.e. $\mu_1 > \mu_n$.

Lema 3.1: As seguintes desigualdades são válidas

$$K_i^n K_{i+1}^{n-1} - K_{i+1}^n K_i^{n-1} < 0, \quad 0 \leq i \leq n-2.$$

Prova. Notemos que

$$\begin{aligned} & n!(n-1)!(K_i^n K_{i+1}^{n-1} - K_{i+1}^n K_i^{n-1}) = \\ & (n-i)!P(n,i)(n-i-2)!P(n-1,i+1) \\ & - (n-i-1)!P(n,i+1)(n-i-1)!P(n-1,i) \\ & = (n-i)!(\mu_n^{-1}P(n-1,i-1) \\ & + P(n-1,i))(n-i-2)!P(n-1,i+1) \\ & - (n-i-1)!(\mu_n^{-1}P(n-1,i) \\ & + P(n-1,i+1))(n-i-1)!P(n-1,i) \\ & = (n-i-2)!^2(n-i-1)(P(n-1,i)P(n-1,i+1) \\ & + \mu_n^{-1}((n-i)P(n-1,i-1)P(n-1,i+1) \\ & - (n-i-1)P(n-1,i)^2)). \end{aligned}$$

Esta expressão é igual à soma de monômios de grau $2i+1$. Cada um destes monômios é um produto de $i+k+1$ variáveis distintas, onde k satisfaz

$$\begin{aligned} k & \geq 0 \\ 2k+1+(i-k) & \leq n-1, \end{aligned}$$

ou, mais simplesmente, $0 \leq k \leq n-2-i$.

Os monômios que tem μ_n^{-1} como fator possuem coeficientes dados por

$$\begin{aligned} & (n-i)(C_{k+1}^{2k}) - (n-i-1)C_k^{2k} \\ & = (n-i)(C_{k+1}^{2k} - C_k^{2k}) + C_k^{2k} \\ & = (n-i)C_k^{2k} \left(\frac{k}{k+1} - 1 \right) + C_k^{2k} \\ & = -C_k^{2k} \frac{n-i-(k+1)}{k+1} \\ & = -C_k^{2k} \frac{n-1-(i+k)}{k+1}. \end{aligned}$$

Os outros monômios têm coeficientes dados por

$$C_k^{2k+1} = C_k^{2k} \frac{2k+1}{k+1}.$$

Associando estes monômios, temos que

$$n!(n-1)!(K_i^n K_{i+1}^{n-1} - K_{i+1}^n K_i^{n-1})$$

é dado por uma soma de monômios da seguinte forma $C_k^{2k} \frac{1}{k+1} \mu_{j_1}^{-2} \dots \mu_{j_{i-k}}^{-2} \mu_{j_{i-k+1}}^{-1} \dots \mu_{j_{i+k}}^{-1} (-n-1-(i+k))\mu_n^{-1} + \sum_{l=i+k+1}^{n-1} \mu_{j_l}^{-1}$, os quais são menores ou iguais a zero. Como, pelo menos um destes termos é negativo quando o sistema é heterogêneo, i.e. quando $\mu_1 > \mu_n$, a tese segue. \square

B. Propriedades das derivadas

Nesta subseção não nos restringiremos aos sistemas heterogêneos. Esta restrição é explicitada no caso do único lema que utiliza resultados da subseção anterior.

Lema 3.2: A seguinte desigualdade é válida:

$$\frac{d^i \frac{1}{p_0^n}}{d\lambda^i} \Big|_{\lambda=0^+} = i! K_i^n > 0. \quad \square$$

Lema 3.3: A expressão seguinte é válida

$$d \frac{\frac{d^i \frac{1}{p_0^n}}{d\lambda^i}}{\frac{p_0^{n-1}}{d\lambda^i}} = \left(\frac{d^{i+1} \frac{1}{p_0^n}}{d\lambda^{i+1}} - \frac{d^i \frac{1}{p_0^n}}{d\lambda^i} \right) \frac{d^{i+1} \frac{1}{p_0^{n-1}}}{d\lambda^{i+1}}$$

Prova. Notemos que

$$\begin{aligned} d \frac{\frac{d^i \frac{1}{p_0^n}}{d\lambda^i}}{\frac{p_0^{n-1}}{d\lambda^i}} & = \frac{d^{i+1} \frac{1}{p_0^n}}{d\lambda^{i+1}} \frac{d^i \frac{1}{p_0^{n-1}}}{d\lambda^i} - \frac{d^i \frac{1}{p_0^n}}{d\lambda^i} \frac{d^{i+1} \frac{1}{p_0^{n-1}}}{d\lambda^{i+1}} \\ & = \frac{d^{i+1} \frac{1}{p_0^n}}{d\lambda^{i+1}} - \frac{d^i \frac{1}{p_0^n}}{d\lambda^i} \frac{d^{i+1} \frac{1}{p_0^{n-1}}}{d\lambda^{i+1}} \\ & = \left(\frac{d^{i+1} \frac{1}{p_0^n}}{d\lambda^{i+1}} - \frac{d^i \frac{1}{p_0^n}}{d\lambda^i} \right) \frac{d^{i+1} \frac{1}{p_0^{n-1}}}{d\lambda^{i+1}}. \end{aligned} \quad \square$$

Lema 3.4: Para $0 < \bar{\lambda} < \mu^{n-1}$ tal que

$$\frac{d^i \frac{1}{p_0^n}}{d\lambda^i} \Big|_{\lambda=\bar{\lambda}} = 0,$$

temos que

$$d^2 \frac{\frac{d^i \frac{1}{p_0^n}}{d\lambda^i}}{\frac{p_0^{n-1}}{d\lambda^i}} \Big|_{\lambda=\bar{\lambda}} = \frac{d^{i+1} \frac{1}{p_0^n}}{d\lambda^{i+1}} \frac{d^{i+1} \frac{1}{p_0^{n-1}}}{d\lambda^{i+1}} \Big|_{\lambda=\bar{\lambda}}.$$

Prova. Este resultado segue diretamente do lema 3.3. \square

Lema 3.5: Supondo que $\mu_1 > \mu_n$, temos que

$$\frac{d^i \frac{1}{p_0^n}}{d\lambda^i} \Big|_{\lambda=0^+} > 0, \quad 0 \leq i \leq n-2.$$

Prova. Este resultado segue diretamente dos lemas 3.3 e 3.1. \square

Lema 3.6: A expressão

$$\frac{d^i \frac{1}{p_0^n}}{d\lambda^i}, \quad i \geq n-1, \quad 0 < \lambda < \mu^{n-1}$$

é uma função decrescente de λ , a qual tende a 0 quando λ tende a μ^{n-1} . Como uma consequência direta deste fato, temos que

$$\frac{d^i \frac{1}{p_0^n}}{d\lambda^i} < 0, \quad i \geq n-1, \quad 0 < \lambda < \mu^{n-1}.$$

Prova. Notemos que, para $i \geq n-1$,

$$\frac{d^i \frac{1}{p_0^n}}{d\lambda^i} = \frac{K_{n-1}^n (1 - \frac{\lambda}{\mu^n})^{-1-i} (\frac{1}{\mu^n})^i}{K_{n-1}^{n-1} (1 - \frac{\lambda}{\mu^{n-1}})^{-1-i} (\frac{1}{\mu^{n-1}})^i}.$$

e

$$\frac{1 - \frac{\lambda}{\mu^{n-1}}}{1 - \frac{\lambda}{\mu^n}} = \frac{\mu^n}{\mu^{n-1}} \frac{\lambda - \mu^{n-1}}{\lambda - \mu^n} = \frac{\mu^n}{\mu^{n-1}} \left(1 + \frac{\mu_n}{\lambda - \mu^n}\right)$$

é uma função decrescente de λ a qual tende a 0 quando λ tende a μ^{n-1} . \square

IV. RESULTADOS PRINCIPAIS

Teorema 4.1: Para um sistema heterogêneo com n servidores ($\mu_1 > \mu_n$), para cada $0 \leq i \leq n-2$, existe exatamente um $0 < \bar{\lambda}_i < \mu^{n-1}$ tal que

$$\frac{d^i \frac{1}{p_0^n}}{d^i \frac{1}{p_0^{n-1}}} \Big|_{\lambda=\bar{\lambda}_i} = 0.$$

Além disso, $\bar{\lambda}_i$ é um ponto de máximo local, e

$$\bar{\lambda}_i \geq \bar{\lambda}_{i+1}, \quad 0 \leq i \leq n-3.$$

Prova. A existência segue diretamente dos lemas 3.5, 3.6, e 3.3.

A unicidade pode ser provada por indução, tendo como base $i = n-2$. Notemos que os lemas 3.5, 3.6 e 3.3 implicam que imediatamente após o primeiro ponto crítico de

$$\frac{d^i \frac{1}{p_0^n}}{d^i \frac{1}{p_0^{n-1}}},$$

sua derivada tem que ser negativa. Deste resultado e do lema 3.5, temos que $\bar{\lambda}_i$ é um ponto de máximo local, e, conseqüentemente, que

$$\frac{d^2 \frac{1}{p_0^n}}{d^2 \frac{1}{p_0^{n-1}}} \Big|_{\lambda=\bar{\lambda}_i} = \frac{d^{i+1} \frac{1}{p_0^n}}{d^{i+1} \frac{1}{p_0^{n-1}}} \frac{d^{i+1} \frac{1}{p_0^{n-1}}}{d^{i+1} \frac{1}{p_0^n}} \Big|_{\lambda=\bar{\lambda}_i} \leq 0.$$

Deste resultado e do lema 3.4, temos que se

$$\frac{d^{i+1} \frac{1}{p_0^n}}{d^{i+1} \frac{1}{p_0^{n-1}}}$$

tem um ponto crítico $\bar{\lambda}_{i+1}$ então $\bar{\lambda}_{i+1} \leq \bar{\lambda}_i$. Supondo, por contradição, que

$$\frac{d^i \frac{1}{p_0^n}}{d^i \frac{1}{p_0^{n-1}}}$$

tenha um outro ponto crítico, as asserções anteriores, referentes às derivadas primeira e segunda, implicam que este seria um outro ponto de máximo local, o que seria uma contradição. O passo da indução é concluído e a tese segue. \square

Teorema 4.2: Para um sistema heterogêneo com n servidores ($\mu_1 > \mu_n$), existe $\lambda^c \in (0, \mu_1 + \mu_2 + \dots + \mu_{n-1})$ tal que

$$T^n \leq T^{n-1} \Leftrightarrow \lambda \geq \lambda^c.$$

Prova. Lembremos que o número médio de usuários no sistema é dado por

$$\begin{aligned} N^n &= \sum_{i=1}^{\infty} iP_i^n = \sum_{k=1}^{\infty} k\lambda^k K_k^n p_0^n \\ &= \lambda p_0^n \frac{d \frac{1}{p_0^n}}{d\lambda}. \end{aligned}$$

Além disso, pelo teorema de Little, temos que T^n , tempo médio de permanência de um usuário no sistema com n servidores, é dado por

$$T^n = p_0^n \frac{d \frac{1}{p_0^n}}{d\lambda}.$$

Assim,

$$\begin{aligned} T^n - T^{n-1} &= p_0^n \frac{d \frac{1}{p_0^n}}{d\lambda} - p_0^{n-1} \frac{d \frac{1}{p_0^{n-1}}}{d\lambda} \\ &= \left(\frac{d \frac{1}{p_0^n}}{d\lambda} - \frac{1}{p_0^{n-1}} \right) \frac{d \frac{1}{p_0^{n-1}}}{d\lambda} \\ &= \frac{\frac{1}{p_0^n}}{\frac{1}{p_0^{n-1}}} \frac{d \frac{1}{p_0^{n-1}}}{d\lambda} \\ &= \frac{1}{p_0} \frac{d \frac{1}{p_0^{n-1}}}{d\lambda} \end{aligned}$$

onde a última igualdade segue do lema 3.3. A tese segue do teorema 4.1 para $i = 0$. \square

V. CONCLUSÃO

O problema do servidor lento com usuários não informados foi resolvido para um número arbitrário de servidores. Este resultado foi derivado como um simples corolário de um teorema referente a razões entre derivadas da probabilidade de que o sistema com n servidores esteja vazio e as derivadas correspondentes para o sistema com $n-1$ servidores.

REFERÊNCIAS

- [1] Rubinovitch, M. *The Slow Server Problem*. J. Appl. Prob. 22,205-213. (1985)
- [2] Rubinovitch, M. *The Slow Server Problem: A Queue With Stalling*. J. Appl. Prob. 22,879-892. (1985)
- [3] Kleinrock, L. *Queueing Systems*. New York, John Wiley, 1975-76, 2v.
- [4] Lin, Woei and Kumar, P. R. *Optimal Control of a Queueing System with Two Heterogeneous Servers*. IEEE Transactions on Automatic Control, AC- 29(8); 696-703. (1983)
- [5] Walrand, J. *A note on "Optimal control of a queueing system with two heterogeneous servers"*. Syst. Control Letters 4, 131-134. (1984)
- [6] Koole, G. *A simple proof of the optimality of a threshold policy in a two-server queueing system* Syst. Control Letters 26, 301-303. (1995)