



SPOLM 2007

ISSN 2175-6295

Rio de Janeiro- Brasil, 08 e 09 novembro de 2007.

## PROCESSOS MARKOVIANOS INTERMEDIÁRIOS: O MODELO $M/M/c(n)/\infty/FIFO$

**Glauco Barbosa da Silva**

Programa de Engenharia de Produção - COPPE - UFRJ  
Cidade Universitária – Centro de Tecnologia – Bloco F  
Rio de Janeiro – RJ - Brasil  
[glauco\\_bar@hotmail.com](mailto:glauco_bar@hotmail.com)

**Manoel Batista dos Santos Junior**

Programa de Engenharia de Produção - COPPE - UFRJ  
Cidade Universitária – Centro de Tecnologia – Bloco F  
Rio de Janeiro – RJ - Brasil  
[mbsj03@yahoo.com.br](mailto:mbsj03@yahoo.com.br)

**Maria Cristina Fogliatti de Sinay**

Mestrado em Engenharia de Transportes - IME  
Rio de Janeiro – RJ - Brasil  
[cristinasinay@ime.eb.br](mailto:cristinasinay@ime.eb.br)

### Resumo

Neste artigo apresentam-se as vantagens do modelo de Filas  $M/M/c(n)/\infty/FIFO$ , isto é, de um sistema com filas onde o número de servidores é variável ao longo do período de oferta do serviço. Este tipo de modelo permite onerar menos os usuários com menores tempos no sistema ao mesmo tempo em que permite a utilização mais racional dos servidores do sistema.  
**Palavras-chave:** Teoria das Filas, Processos Estocásticos, Apoio à Decisão.

### Abstract

In this article it is present advantages of Queueing Model  $M/M/c(n)/\infty/FIFO$ , i.e., queueing system with variable number of servers throughout period of offered service. This type of model allows to reduce burdensome customers with smaller time in the system as well allows a more reasonable use of servers in the system.

**Keywords:** Queueing Theory, Stochastic Models, Decision Support Methodologies.

### 1. Introdução

A formação de filas de espera é um problema real e comum que ocorre sempre que a demanda atual por um serviço excede a capacidade de atendimento. Espera-se em uma fila para comprar ingressos no cinema, efetuar um depósito bancário, postar uma carta no correio, servir-se de um prato em um restaurante “*self-service*”, entre outros.

O emprego da Teoria das Filas permite a modelagem analítica de processos que resultam em espera (filas) e posterior obtenção de resultados ou medidas de desempenho que traduzam a operacionalidade ou produtividade dos processos.

Estes modelos são, então, importantes ferramentas nos processos decisórios, pois possibilitam o dimensionamento da infra-estrutura, dos recursos humanos e financeiros dos mais variados sistemas, atendendo enfoques conflitantes, como são os dos usuários dos sistemas, que preferenciam filas curtas e tempos de permanência reduzidos, e os dos gerentes, que procuram maiores volumes de serviços prestados.

A prestação de serviço, seja por pessoas e/ou por máquinas, requer investimentos de capital e custos operacional e de manutenção. Assim, oferecer serviços além dos necessários provoca custos que devem ser evitados e abaixo dos necessários, a perda de clientes, que leva a falência do sistema.

Um sistema corretamente dimensionado permite manter o equilíbrio entre o capital disponibilizado no sistema e os retornos sociais e financeiros do mesmo, ou seja, um balanceamento econômico entre o custo do serviço e o custo associado à espera pelo serviço.

Este balanceamento pode ser obtido pela aplicação da Teoria das Filas e do tratamento estocástico de informações referentes a tempos entre chegadas sucessivas de clientes ao sistema e tempos de serviços despendidos para estes clientes. Amostras de dados permitem caracterizar os processos “chegada” e “atendimento”.

Observações físicas do sistema sob análise complementam as informações necessárias para aplicações da Teoria das Filas, quais sejam, número de servidores em paralelo (atendimento simultâneo), existência ou não de limitações físicas para a formação da fila e disciplina adotada.

Para a utilização desta Teoria, algumas simplificações da realidade, bem como algumas suposições básicas devem ser estabelecidas. Dentre estas, destacam-se o fato do sistema estar vazio no instante do início de sua operação, a necessidade dos clientes chegarem separadamente mesmo que apenas por pequenos intervalos de tempo, chegadas e saídas serem estatisticamente independentes e taxa de chegadas ao sistema constante. Estas hipóteses são essenciais para que a teoria seja aplicável.

Tem sido demonstrado, até mesmo com o uso de Simulação, que se qualquer uma dessas simplificações (exceto a última) não forem estritamente cumpridas, os resultados finais não se distanciam significativamente da realidade. Entretanto, a taxa de chegada flutuando ao longo do tempo leva a medidas de desempenho sub/superestimadas. Surge, assim, a necessidade de modelos para avaliar este caso.

O objetivo deste trabalho é apresentar as medidas de desempenho associadas ao modelo  $M/M/c(n)/\infty/FIFO$  e comparar com as correspondentes do modelo tradicional  $M/M/c/\infty/FIFO$ .

## 2. O Modelo M/M/c(n)/∞/FIFO

Segundo KLEINROCK[9], considerando-se os conceitos inerentes aos Processos Markovianos (mais especificamente os Processo de Nascimento-Morte, cujas transições de um estado qualquer somente são possíveis para estados vizinhos, ou seja, de um estado  $i$  para estados  $i + 1$  ou  $i - 1$ ), o modelo M/M/c/∞/FIFO representa uma realidade onde usuários ingressam no sistema, com tempos entre chegadas sucessivas seguindo uma distribuição exponencial de parâmetro( $\lambda$ ), buscando por um determinado serviço.

Como não há limitação física do sistema, ou seja, não há limitação do tamanho da fila ou da quantidade de usuários no sistema, não existem rejeições ou desistências. Esses usuários, ao ingressarem no sistema, entram numa fila simples(única) e são atendidos, na ordem de chegada, por  $c$  servidores em paralelo à medida que estes ficam disponíveis para atendimento. Assume-se a hipótese de que os tempos de atendimento dos  $c$  servidores são idênticos e seguem uma distribuição exponencial de parâmetro ( $\mu$ ).

Esquemáticamente, o modelo por ser visualizado como no gráfico da Figura 1 a seguir.

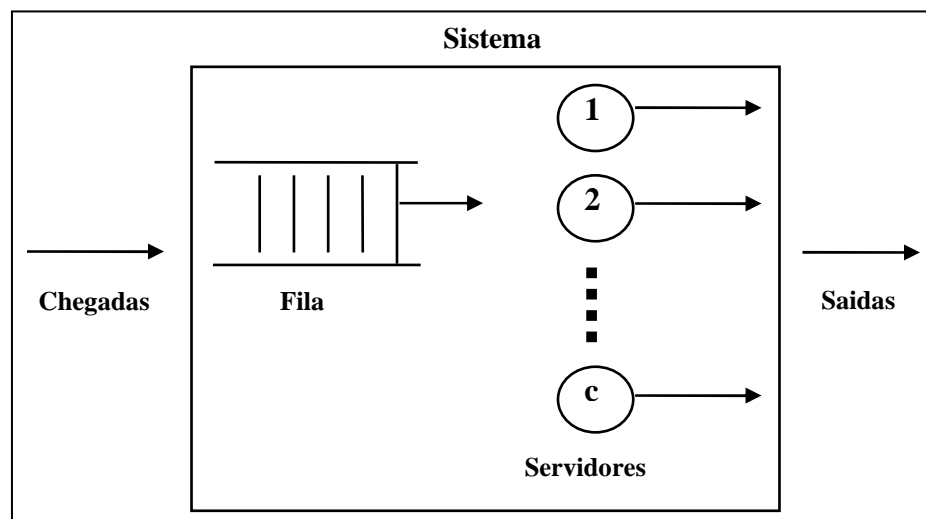


Figura 1 – Modelagem M/M/c/∞/FIFO

Formalmente, segundo a notação de Kendall, para o modelo M/M/c/∞/FIFO, tem-se:

- M – os tempos entre chegadas sucessivas seguem distribuições exponenciais de parâmetro  $\lambda$ ;
- M – os tempos de atendimentos de cada servidor seguem distribuições exponenciais de parâmetro  $\mu$ ;
- c – o número de postos de atendimento (servidores) em paralelo;
- ∞ – não são impostas limitações físicas ao modelo;
- FIFO – o atendimento dos usuários respeita a ordem das chegadas;

Para um sistema com diferentes situações de congestionamento ao longo do período observado, onde a fluutuabilidade da demanda não pode ser controlada, uma forma de controlar o desempenho do sistema é variar o número de servidores de modo a atender a demanda satisfatoriamente e reduzir custos operacionais.

Este modelo busca equilibrar a fluutuabilidade da demanda de clientes com a variação do número de servidores em operação, isto é, o número de servidores passa a ser flexível, não ultrapassando o número limite de servidores disponíveis, além de atender alguma restrição estabelecida pela gerência do sistema, que pode ser uma função do número de usuários

existentes no sistema (ou fila) ou do tempo de permanência de um usuário qualquer no sistema (ou fila).

Neste trabalho apresenta-se o caso particular limitado a dois servidores ( $1 \leq c \leq 2$ ), que atende a restrição de um determinado número de usuários no sistema ( $k$  usuários). Trata-se de um sistema onde o número de servidores em paralelo depende do número de elementos no sistema, quando a fila atinge determinado comprimento, um novo servidor é posto em operação, sem que o número de postos de atendimento ultrapasse o número máximo de servidores disponíveis. Quando o número de usuários decresce ( $i < k$ ) e o serviço é completado, o servidor livre é retirado de operação.

O estado estacionário sob análise, supondo que este existe, é representado pelo diagrama de fluxo da Figura 2, a seguir:

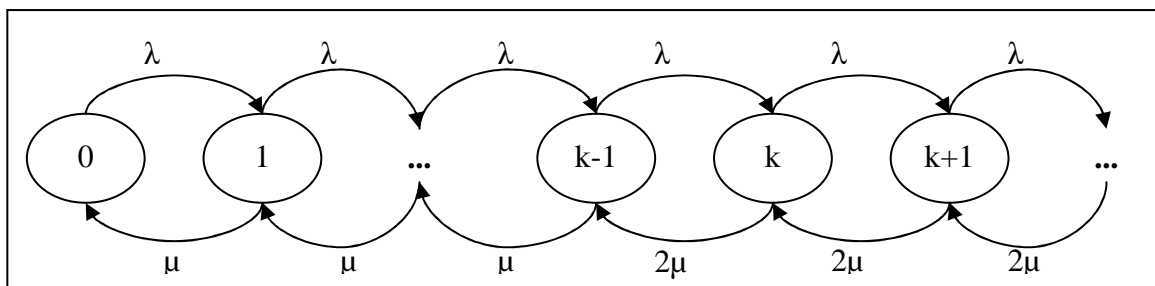


Figura 2 – Diagrama de Fluxos para o modelo M/M/c(n)/∞/FIFO

Considerando-se o Princípio de Conservação de Energia (fluxo que entra em um nó é igual ao fluxo que sai desse mesmo nó) e sendo  $P_i$  a probabilidade de um sistema estar no estado  $i$ , do diagrama da figura 2, tem-se:

Estado	Fluxo de entrada	Fluxo de saída
0	$\mu P_1$	$\lambda P_0$
1	$\mu P_2 + \lambda P_0$	$\mu P_1 + \lambda P_1$
2	$\mu P_3 + \lambda P_1$	$\mu P_2 + \lambda P_2$
:	:	:
:	:	:
k-1	$2\mu P_k + \lambda P_{k-2}$	$\mu P_{k-1} + \lambda P_{k-1}$
k	$2\mu P_{k+1} + \lambda P_{k-1}$	$2\mu P_k + \lambda P_k$
k+1	$2\mu P_{k+2} + \lambda P_k$	$2\mu P_{k+1} + \lambda P_{k+1}$

de onde, MODER[10] e FOGLIATTI e MATTOS [5] mostram que as correspondentes probabilidades dos estados são dadas por:

$$P_i = \left(\frac{\lambda}{\mu}\right)^i P_0, \quad \text{se } 1 \leq i < k$$

$$P_i = \frac{1}{2^{i-k+1}} \left(\frac{\lambda}{\mu}\right)^i P_0, \quad \forall i \geq k \quad (1)$$

Substituindo as equações (1) em:

$$\sum_{i=0}^{\infty} P_i = \sum_{i=0}^{k-1} P_i + \sum_{i=k}^{\infty} P_i$$

a condição necessária para que a soma infinita acima seja convergente e o sistema entre no estado estacionário, é :

$$\frac{\rho}{2} < 1 \quad \text{onde } \rho = \lambda/\mu;$$

Satisfeita a condição acima, FOGLIATTI e MATTOS [5] mostram que:

se  $\rho \neq 1$

$$P_0 = \frac{(1-\rho)(2-\rho)}{2-\rho-\rho^k} \quad (2)$$

e se  $\rho = 1$

$$P_0 = \frac{1}{k+1} \quad (3)$$

Substituindo (2) e (3) em (1), tem-se:

para  $\rho \neq 1$

$$P_i = \begin{cases} \rho^i \frac{(1-\rho)(2-\rho)}{2-\rho-\rho^k}, & 0 \leq i < k; \\ \rho^i \frac{1}{2^{i-k+1}} \frac{(1-\rho)(2-\rho)}{2-\rho-\rho^k}, & i \geq k; \end{cases} \quad (4)$$

e para  $\rho = 1$

$$P_i = \begin{cases} \frac{1}{k+1}, & 0 \leq i < k; \\ \frac{1}{2^{i-k+1}(k+1)}, & i \geq k; \end{cases} \quad (5)$$

A partir das probabilidades apresentadas, podem ser obtidos os parâmetros de desempenho do sistema, isto é,  $L, L_q, W, W_q$ , entre outros.

### 3. Comparação do Modelo Tradicional, $c$ constante, com o Modelo com $c=c(n)$

Por meio de um exemplo, será mostrado que o uso do modelo com  $c=c(n)$  trás algumas vantagens em relação ao uso do modelo tradicional, isto é,  $c=$ constante.

Seja um sistema que se comporta como um M/M/1/∞/FIFO, com  $\lambda = 19$  clientes por hora e  $\mu = 20$  clientes por hora.

Denotando as medidas de desempenho operacional para o modelo como: Probabilidade do Sistema estar ocioso ( $P_0$ ), Número Médio de Usuários no Sistema ( $L$ ), Número Médio de Usuários na Fila-Comprimento da Fila ( $L_q$ ), Tempo Médio de Espera no Sistema ( $W$ ), Tempo Médio de Espera na Fila ( $W_q$ ), para este modelo, vide FOGLIATTI e MATTOS [5], tem-se:

Taxa de Utilização do Sistema	$\rho$	0,95
Probabilidade do Sistema estar Ocioso	$P_0$	0,05
Número Médio de elementos na Fila (Comprimento da Fila)	$L_q$	18
Número Médio de elementos no Sistema	$L$	19
Tempo Médio de Espera na Fila(Min)	$W_q$	57
Tempo Médio de Espera no Sistema(Min)	$W$	60

Tabela 1 – Medidas de Desempenho para o Modelo M/M/1/∞/FIFO

Analisando os resultados obtidos pode-se observar que a taxa de ocupação do sistema é alta (95%), apenas 5% do tempo o sistema encontra-se ocioso. Em média existem 18 clientes na fila e 19 no sistema, um número bastante elevado, sendo que esses clientes esperam em média quase uma hora na fila / sistema, sendo o serviço realizado em apenas 3 minutos. Os resultados mostram que o sistema encontra-se próximo de um colapso.

Com base na avaliação apresentada a gerência do sistema deveria contratar um segundo servidor com características semelhantes ao primeiro, a ser colocado em paralelo.

Considerando-se ainda que a taxa de chegada de clientes utilizada nos cálculos representa uma taxa média das chegadas e que o sistema possui demanda flutuante com diferentes situações de congestionamento, nos momentos de pico de demanda o sistema não comporta a demanda.

Preservando-se as condições do cenário, após a inclusão do segundo servidor, são apresentadas as medidas de desempenho na Tabela 2 a seguir:

Taxa de Utilização do Sistema	$\rho$	0,46
Probabilidade do Sistema estar Ocioso	$P_0$	0,36
Número Médio de elementos na Fila (Comprimento da Fila)	$L_q$	0,28
Número Médio de elementos no Sistema	$L$	1,23
Tempo Médio de Espera na Fila(Min)	$W_q$	0,87
Tempo Médio de Espera no Sistema(Min)	$W$	3,87
Probabilidade de um cliente chegar e ser logo atendido	$P(i < c)$	0,69
Probabilidade de Ócio de um servidor qualquer	$P_s$	0,53

Tabela 2 – Medidas de Desempenho para o Modelo M/M/2/∞/FIFO

Avaliando-se os resultados obtidos e apresentados na Tabela 2, com a contratação do segundo servidor, a taxa de ocupação do sistema passa a ser muito baixa (46%), praticamente não existem clientes na fila.

Comparando-se os resultados obtidos nos dois casos, tem-se situações conflitantes de colapso e de excesso de serviço.

Como alternativa para equilibrar a situação encontrada propõe-se a variação do número de servidores em operação, isto é, o número de servidores passa a ser flexível, não ultrapassando o número limite de servidores disponíveis, estabelecendo-se como restrição, que quando o número de usuários no sistema for maior ou igual a um dado  $k$  será utilizado o segundo servidor no atendimento. Tem-se, então, caracterizado o sistema com o modelo  $M/M/c(k)/\infty/FIFO$ , para o que as medidas de desempenho são as apresentadas na Tabela 3 a seguir:

	<b>k= 5</b>	<b>k=8</b>
<b>P<sub>0</sub></b>	0,19	0,14
<b>L<sub>q</sub></b>	1,5	2,8
<b>L</b>	2,5	3,7
<b>W<sub>q</sub>(min)</b>	4,8	8,7
<b>W(min)</b>	7,8	11,7
<b>P(i&lt;c)</b>	0,37	0,27
<b>P<sub>s</sub></b>	0,28	0,20

Tabela 3 – Medidas de Desempenho para o Modelo  $M/M/c(k)/\infty/FIFO$

Avaliando-se os resultados obtidos e apresentados na tabela 3, com a utilização do segundo servidor “flexível”, consegue-se uma significativa redução da ociosidade do sistema(19%) sem acarretar em um aumento significativo do número médio de clientes na fila/sistema ou no tempo médio de espera na fila/sistema. O modelo permite ao gerente estabelecer de maneira conveniente o ponto de entrada/saída do servidor( $k$ ).

Na Figura 3 a seguir são apresentadas as variações das medidas de desempenho x variação do  $k$ .

Para  $\rho=0.95$ , tem-se:

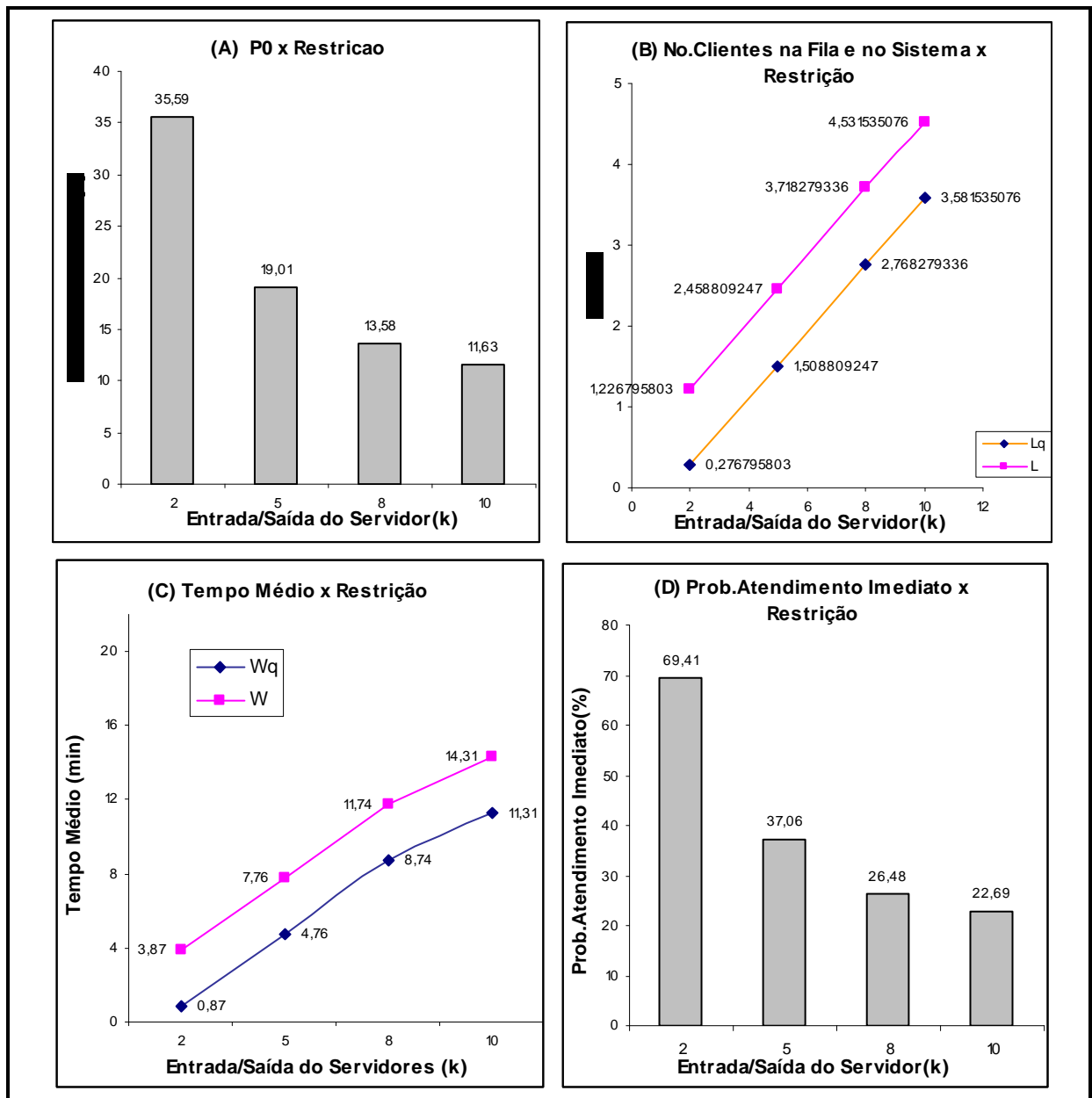


Figura 3 – Medidas de Desempenho x Ponto de Entrada/Saída do Servidor

Para os diferentes pontos de entrada/saída do segundo servidor, as medidas de desempenho podem ser administradas pela gerência de modo a não haver perda de clientes.

No caso particular de se trabalhar com uma taxa  $\lambda$  variável, mantendo-se  $\mu$  constante, têm-se os valores da Tabela 3 e as representações da Figura 4 para as medidas de desempenho  $P_0, L_q, L, W_q, W$  respectivamente e para o modelo  $M/M/c(k)/\infty/FIFO$  com  $k=5, 8$  e  $10$  e para  $c=1$  e  $c=2$ , respectivamente.



<b>P<sub>0</sub>(%)</b>						
<b>ρ</b>	<i>0,5</i>	<i>0,75</i>	<i>1</i>	<i>1,25</i>	<i>1,5</i>	<i>1,75</i>
<b>k=5</b>	51,0%	30,0%	16,7%	8,2%	3,5%	1,2%
<b>k=8</b>	50,1%	27,2%	11,1%	3,6%	1,0%	0,2%
<b>k=10</b>	50,0%	26,2%	9,1%	2,2%	0,4%	0,0%
<b>c =1</b>	50,0%	25,0%	0,0%	0,0%	0,0%	0,0%
<b>c =2</b>	60,0%	45,0%	33,0%	23,0%	14,0%	6,0%
<b>L</b>						
<b>ρ</b>	<i>0,5</i>	<i>0,75</i>	<i>1</i>	<i>1,25</i>	<i>1,5</i>	<i>1,75</i>
<b>k=5</b>	0,89	1,69	2,67	3,84	5,56	9,85
<b>k=8</b>	0,98	2,25	4,11	6,06	8,22	12,67
<b>k=10</b>	0,99	2,50	5,09	7,69	10,11	14,63
<b>c =1</b>	1,00	3,00	∞	∞	∞	∞
<b>c =2</b>	0,53	0,87	1,33	2,05	3,43	7,45
<b>L<sub>q</sub></b>						
<b>ρ</b>	<i>0,5</i>	<i>0,75</i>	<i>1</i>	<i>1,25</i>	<i>1,5</i>	<i>1,75</i>
<b>k=5</b>	0,39	0,94	1,67	2,59	4,06	8,10
<b>k=8</b>	0,48	1,50	3,11	4,81	6,72	10,92
<b>k=10</b>	0,49	1,75	4,09	6,44	8,61	12,88
<b>c =1</b>	0,5	2,25	∞	∞	∞	∞
<b>c =2</b>	0,03	0,12	0,33	0,80	1,93	5,70
<b>W<sub>q</sub> (min)</b>						
<b>ρ</b>	<i>0,5</i>	<i>0,75</i>	<i>1</i>	<i>1,25</i>	<i>1,5</i>	<i>1,75</i>
<b>k=5</b>	2,32	3,75	5,00	6,21	8,13	13,89
<b>k=8</b>	2,87	6,01	9,33	11,54	13,43	18,72
<b>k=10</b>	2,96	7,00	12,27	15,45	17,22	22,08
<b>c =1</b>	3,00	9,00	∞	∞	∞	∞
<b>c =2</b>	0,20	0,49	1,00	1,92	3,86	9,77
<b>W(min)</b>						
<b>ρ</b>	<i>0,5</i>	<i>0,75</i>	<i>1</i>	<i>1,25</i>	<i>1,5</i>	<i>1,75</i>
<b>k=5</b>	5,32	6,75	8,00	9,21	11,13	16,89
<b>k=8</b>	5,87	9,01	12,33	14,54	16,43	21,72
<b>k=10</b>	5,96	10,00	15,27	18,45	20,22	25,08
<b>c =1</b>	6,00	12,00	∞	∞	∞	∞
<b>c =2</b>	3,20	3,49	4,00	4,92	6,86	12,77

Tabela 3 – Medidas de Desempenho x Variação Taxa de Chegadas

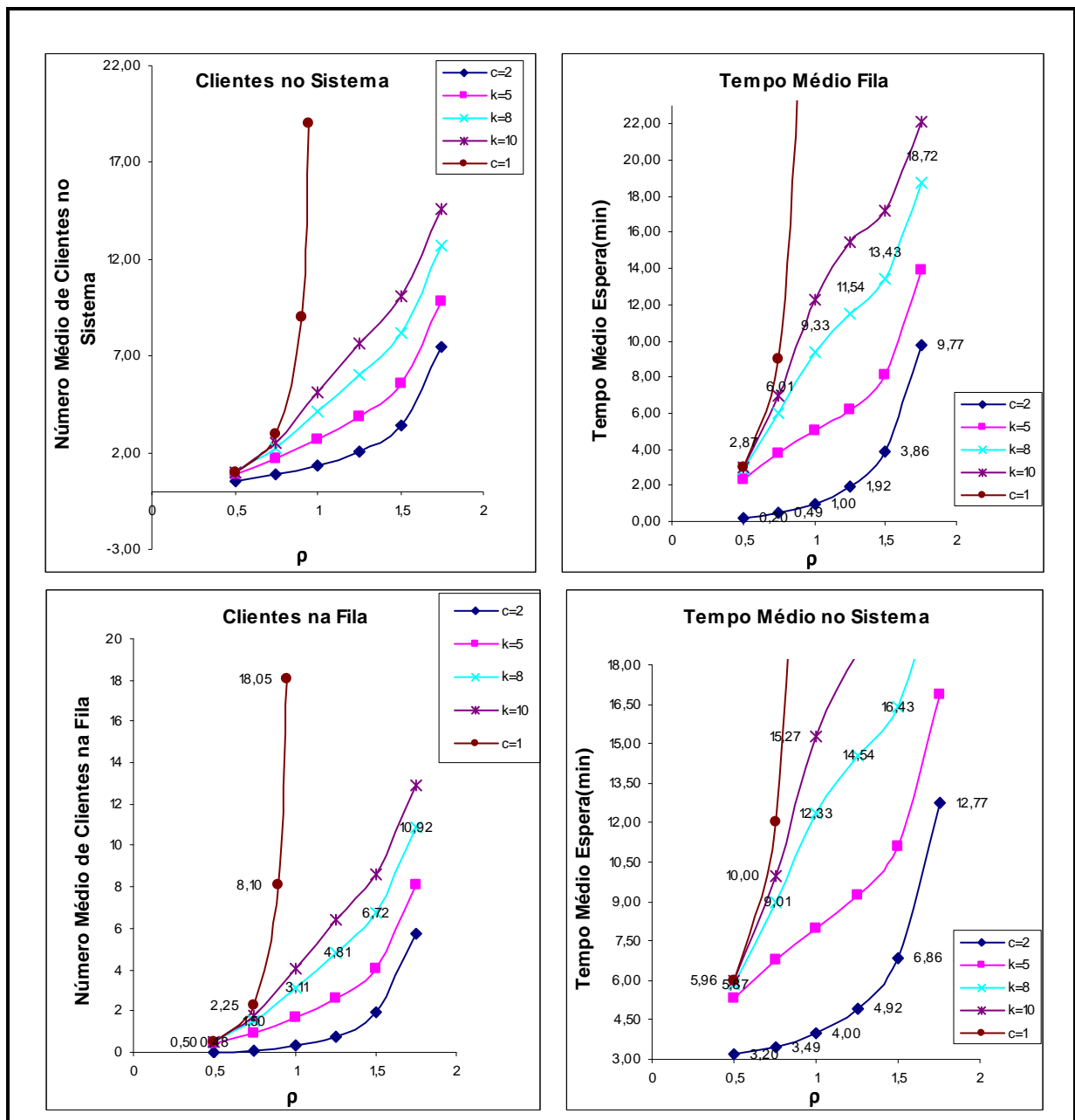


Figura 4 – Medidas de Desempenho x Taxa de Chegadas

Analisando-se estes dados, pode-se observar que:

Se a demanda por serviço é baixa, o ponto de entrada/saída do servidor(k) não influencia significativamente nos resultados obtidos, ou seja, o segundo servidor pode permanecer em outra atividade sem comprometer o funcionamento do sistema.

Entretanto, à medida que a demanda cresce, esse ponto de entrada/saída precisa ser reavaliado de modo a não representar perdas, pois os resultados distanciam-se mais significativamente do modelo tradicional.

#### 4. Considerações Finais

Apresentou-se o comportamento do modelo  $M/M/c(n)/\infty/FIFO$  como alternativa ao modelo tradicional  $M/M/c/\infty/FIFO$ , mostrando que o primeiro modelo pode proporcionar um atendimento eficiente ao usuário e flexibilidade ao gerente, permitindo este último re-alocar os servidores que não estejam em operação em outras necessidades da gerência.

Considerando-se a fluuabilidade da demanda em determinados períodos, constata-se que para demandas baixas a utilização de apenas um número fixo de servidores atende as necessidades dos clientes, porém não são contemplados os períodos de pico.

Os resultados para os modelos tradicionais e proposto foram obtidos com a utilização de planilhas eletrônicas do Microsoft Excel 2003.

O modelo proposto demonstrou-se uma relevante ferramenta de apoio a decisão, propiciando diferentes alternativas para gerenciamento do sistema com a mesma quantidade de recursos e qualidade de serviço.

#### 5. Referências

- [1] ALVES, R.A.F.S; DELGADO, C.J.M. Processos Estocásticos. Faculdade de Economia, FEP, 1997. Disponível em <http://nuieee.fe.up.pt/~ei00046/1/susy/Texto4.pdf> Último acesso em 25 de abril de 2007.
- [2] BOLCH, G.; GREINER, S.; DE MEER, H.; TRIVEDI, K.S. Queueing networks and Markov chains: modeling and performance evaluation with computer science applications. John Wiley & Sons, Inc., New York, 1998.
- [3] CLARK, A.B.; DISNEY, R.L. Probabilidade e Processos Estocásticos. Rio de Janeiro. Livros Técnicos e Científicos Editora S.A., 1979.
- [4] De GROOT, M.H.; SCHERVISH, M. J. Probability and Statistics, 3ª edição. New York: Addison-Wesley, 2002.
- [5] FOGLIATTI, M.C.; MATTOS, N.M.C. Teoria das Filas. Rio de Janeiro. Editora Interciência, 2007.
- [6] GROSS, D.; HARRIS, C.M. Fundamentals of queueing theory, John Wiley & Sons Inc., Chichester, 1985.
- [7] HEYMAN, D.P.; SOBEL, M.J. Stochastic Models – Handbooks in Operations Research and Management Science Vol.2. North-Holland. Amsterdam, 1990.
- [8] HILLIER, F.S.; LIEBERMAN, G.J. Introdução a Pesquisa Operacional, Editora Campus/Edusp, 1988.
- [9] KLEINROCK, L. Queueing Systems, Vol. I: Theory. John Wiley & Sons, Inc. New York, 1975.
- [10] MODER, J.J.; PHILLIPS Jr. C.R. Queueing with Fixed and Variable Channels. Operations Research, Vol. 10, No. 2. (Mar. - Apr., 1962), pp. 218-231.
- [11] NAIN, P. Notas de aula da Disciplina Performance Evaluation of Computer Systems. University of Massachusetts, Amherst, MA, 1994.
- [12] RESING, J.; ADAN, I. Queueing Theory - Department of Mathematics and Computing Science, Eindhoven University of Technology, 2001.
- [13] ROSS, SHELDON M. Stochastic Processes, 2ª edição. New York: Wiley & Sons, 1996.
- [14] ROSS, SHELDON M. Introduction to Probability Models, 5ª ed. Academic Press, 1993.