

PROCESSOS MARKOVIANOS INTERMEDIÁRIOS:
UM MODELO DE FILAS COM NÚMERO VARIÁVEL DE SERVIDORES

Glauco Barbosa da Silva

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS
PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE
FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS
NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS
EM ENGENHARIA DE PRODUÇÃO.

Aprovada por:

Prof. Samuel Jurkiewicz, D.Sc.

Prof. Maria Cristina Fogliatti de Sinay, Ph.D.

Prof. Cláudia Marcela Justel, D.Sc.

Prof. Virgílio José Martins Ferreira Filho, D.Sc.

RIO DE JANEIRO, RJ - BRASIL

MAIO DE 2008

DA SILVA, GLAUCO BARBOSA

Processos Markovianos Intermediários:
Um Modelo de Filas com Número Variável
de Servidores [Rio de Janeiro] 2008

XI, 75 p. 29,7 cm (COPPE/UFRJ, M.Sc.,
Engenharia de Produção, 2008)

Dissertação - Universidade Federal do Rio
de Janeiro, COPPE

1. Teoria das Filas
2. Processos Markovianos Intermediários
 - I. COPPE/UFRJ II. Título (série)

Agradecimentos

A Deus.

À professora Cristina, por quem descobri a essência da palavra orientação, pela confiança, segurança, apoio, incentivo, e, principalmente, pelo exemplo de educadora.

À minha namorada e companheira, Maria Fernanda, pela serenidade e tolerância nas ausências.

Aos amigos *Altas Horas*: Gazu, Frangão, Manel e Juliana pela paciência, companheirismo, confiança e apoio nos momentos difíceis.

Ao professor Samuel Jurkiewicz pelo apoio em um momento decisivo desta dissertação.

Ao professor José Paulo Dieguez, *The Series Killer*, pela valorosa contribuição no *extermínio* de uma série complexa.

Aos professores da banca examidora: Cláudia Marcela Justel e Virgílio José Martins Ferreira Filho por aceitarem o convite de participar da banca examinadora.

Finalmente, agradeço à Marinha do Brasil pela oportunidade de realizar este curso.

Com todo amor, para:
Denira e Antônio, meus pais;

Resumo da Dissertação apresentada a COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

PROCESSOS MARKOVIANOS INTERMEDIÁRIOS:
UM MODELO DE FILAS COM NÚMERO VARIÁVEL DE SERVIDORES

Glauco Barbosa da Silva

Maio/2008

Orientadores: Samuel Jurkiewicz, D.Sc.

Maria Cristina Fogliatti de Sinay, Ph.D

Programa: Engenharia de Produção

Fila é uma característica comum a variados campos, está presente em muitas situações do dia-a-dia até ambientes mais técnicos como redes de computadores e sistemas de telecomunicações. Os primeiros trabalhos desenvolvidos em Teoria de Filas foram motivados por problemas de Telefonia. Em estágios mais recentes, a Teoria de Filas tem sido aplicada extensivamente aos problemas reais de produção, transporte público, entre outros. A Teoria de Filas cresceu buscando ferramentas teóricas (modelos matemáticos) para os fenômenos de filas. Em consequência, a despeito da aplicabilidade potencial desta Teoria, o distanciamento entre os desenvolvimentos teóricos e as aplicações reais também cresceram. Neste trabalho é apresentado um modelo de filas com número variável de servidores, que visa contribuir no estreitamento entre a teoria e a prática. Os resultados obtidos a partir do modelo desenvolvido permitem avaliar o potencial do modelo como ferramenta de apoio à decisão.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

INTERMEDIATE MARKOV PROCESS:
A VARIABLE-CHANNEL QUEUING MODEL

Glauco Barbosa da Silva

May/2008

Advisors: Samuel Jurkiewicz, D.Sc.

Maria Cristina Fogliatti de Sinay, Ph.D

Department: Industrial Engineering

Queueing is a common feature in a wide range of fields a spectrum from many daily-life situations to more technical environments such as computer networks and telecommunications systems. The early works were motivated by problems concerning telephone traffic. Nowadays, queueing has been extensively applied to real problems arising in manufacturing, public transportation, etc. The Queueing Theory has grown looking for theoretic and algorithmic tools, and mathematical models of queueing phenomena. As a result of this, despite of the potencial applicability of queueing, the gap between theoretical developments and real applications also has grown. In this work is present a model of queueing with variable number of servers, that aims to contribute to make narrow the gap between theory and practice in queueing. The measure from the development model allow to evaluate the potential of the model as tool of support to the decision.

Sumário

Lista de Figuras	ix
Lista de Tabelas	xi
1 Introdução	1
2 Características de um Modelo de Filas	7
2.1 Estrutura Básica	7
2.2 Notação e Terminologia	9
2.3 Medidas de Desempenho	10
3 O Modelo $M/M/c/\infty/FIFO$	12
3.1 Caracterização do Sistema no Regime Estacionário	14
3.2 Medidas de Desempenho para o Modelo $M/M/c/\infty/FIFO$	16
3.2.1 Número Médio de Elementos na Fila(L_q)	16
3.2.2 Número Médio de Elementos no Sistema(L)	17
3.2.3 Tempo Médio de Espera na Fila(W_q)	17
3.2.4 Tempo Médio de Permanência no Sistema(W)	17
3.2.5 Probabilidade de Ócio Parcial do Sistema	17
3.2.6 Probabilidade de Ócio de um Servidor Qualquer(P_{oc})	18
3.2.7 Considerações Finais	19

4	O Modelo $M/M/c(k) \leq 2/\infty/FIFO$	20
4.1	$M/M/c(k) \leq 2/\infty/FIFO$	22
4.2	Medidas de Desempenho	29
5	O Modelo $M/M/c(k) \leq c/\infty/FIFO$	32
5.1	Medidas de Desempenho	42
5.2	Considerações Finais	46
6	Estudo de Caso: Sistema de Pagamento de uma Loja de Departamentos	47
6.1	Caracterização do Estudo	48
6.2	Medidas de Desempenho	49
6.2.1	Resultados obtidos usando o Modelo $M/M/c/\infty/FIFO$	49
6.2.2	Resultados para o Modelo $M/M/c(k) \leq c/\infty/FIFO$	51
7	Conclusão	58
	Referências Bibliográficas	60
	Apêndices	64
A		64

Lista de Figuras

2.1	Processo de Fila Básico	7
3.1	Modelagem $M/M/c/\infty/FIFO$	13
3.2	Diagrama de Fluxo para o modelo $M/M/c/\infty/FIFO$	14
4.1	Diagrama de Fluxos para o modelo $M/M/c(k) \leq 2/\infty/FIFO$	22
4.2	Estado Inicial	23
4.3	Estado i	23
4.4	Estado k	24
4.5	Estado $k+1$	24
4.6	Estado $k+2$	25
4.7	Estado $k+3$	26
4.8	Estado i com $i \geq k + 2$	26
5.1	Diagrama de Fluxos para o modelo $M/M/c(k)/\infty/FIFO$	32
5.2	Estado Inicial	33
5.3	Estado i	33
5.4	Estado k	34
5.5	Estado $k+1$	34
5.6	Estado $k+2$	35

5.7	Estado $k+3$	36
5.8	Estado i ($k + 1 < i < k + c$)	36
5.9	Estado $k+c-1$	37
5.10	Estado $k+c$	38
5.11	Estado $k + c + 1$	39
5.12	Estado i ($i > k + c$)	39
6.1	Medidas de Desempenho x Servidores Fixos	49
6.2	Medidas de Desempenho x <i>Shift Point</i>	51
6.3	Número Médio de Clientes no Sistema.	53
6.4	Número Médio de Clientes na Fila.	55
6.5	Tempo Médio no Sistema x ρ	56
6.6	Tempo Médio na Fila.	57
A.1	Diagrama de Fluxo de uma Cadeia de Markov para 2 estados	67
A.2	Diagrama de Fluxo de um Processo de Nascimento e Morte	69
A.3	Diagrama de Fluxo de um Processo de Poisson com taxa λ	72

Lista de Tabelas

5.1	Resumo das Probabilidades de Estado	41
6.1	Medidas de Desempenho	49
6.2	Medidas de Desempenho - aumento 10% na taxa de chegadas.	50
6.3	Medidas de Desempenho - redução 10% na taxa de chegadas.	50
6.4	Medidas de Desempenho x <i>Shift Point</i>	51
6.5	Medidas de Desempenho x Servidores Fixos e <i>Shift Points</i>	52
6.6	Medidas de Desempenho x <i>Shift Point</i> - aumento 10% na taxa de chegadas.	52
6.7	Medidas de Desempenho x <i>Shift Point</i> - redução 10% na taxa de chegadas.	52
6.8	Número Médio de Clientes no Sistema	53
6.9	Número Médio de Clientes na Fila.	55
6.10	Tempo Médio no Sistema.	56
6.11	Tempo Médio na Fila.	57
A.1	Eventos possíveis no instante $t + \Delta t (i > 0)$	70
A.2	Eventos possíveis no instante $t + \Delta t (i = 0)$	71
A.3	Eventos possíveis no instante $t + \Delta t$	73

Capítulo 1

Introdução

A Pesquisa Operacional é uma ciência aplicada voltada para a resolução de problemas reais, utilizando diversas técnicas quantitativas. Visa introduzir elementos de objetividade e racionalidade nos processos decisórios, sem descuidar-se dos elementos subjetivos e de enquadramento organizacional que caracterizam os problemas, utilizando-se de modelos científicos, tipicamente matemáticos, que representem a essência do problema real.

A Pesquisa Operacional surgiu durante a Segunda Guerra Mundial com problemas urgentes (de natureza logística, tática e de estratégia militar) de grande dimensão e complexidade. Por causa do esforço de guerra foram então criados grupos multidisciplinares de cientistas (matemáticos, físicos, engenheiros e cientistas sociais) que aplicaram métodos científicos aos problemas que lhes foram sendo colocados. Desenvolveram assim a idéia de criar modelos matemáticos, apoiados em dados e fatos, que lhes permitissem perceber os problemas em estudo e simular e avaliar o resultado hipotético de estratégias ou decisões alternativas.

O sucesso e credibilidade alcançados durante a guerra fizeram com que, terminado o conflito, esses grupos de cientistas e a nova metodologia de abordagem dos problemas se transferissem para as empresas emergentes com a explosão econômica que se seguiu.

Surgiram em seguida grandes desenvolvimentos técnicos e metodológicos que apoiados na crescente capacidade e disseminação dos meios computacionais, permitiram trabalhar enormes volumes de dados e, por meio de adequados modelos de base quan-

titativa, simular e avaliar linhas de ação alternativas e encontrar as soluções que melhor serviam aos objetivos dos indivíduos ou organizações.

A Pesquisa Operacional apresenta contribuições que se estendem por praticamente todos os domínios da atividade humana, mas com ligações particularmente fortes com a Engenharia de Produção em vista do seu potencial de abordagem dos problemas envolvendo os meios de produção: seres humanos, matérias-primas, equipamentos, recursos financeiros e meio ambiente. As aplicações da Pesquisa Operacional vão desde o balanceamento de linhas de produção em fábricas, ou fluxo ótimo de pacientes em hospitais, até ao estudo de estruturas sociais e de processos psicológicos.

A Teoria de Filas constitui um importante ramo da ciência Pesquisa Operacional. Esta Teoria consiste na modelagem analítica de processos que resultam em espera (filas), buscando obter resultados, medidas de desempenho (comprimento médio da fila, tempo médio de espera na fila, utilização média do sistema, entre outros) que traduzam a operacionalidade ou produtividade dos processos, atuando assim como importante ferramenta nos processos decisórios.

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

No início do Século *XX*, o matemático dinamarquês A.K.Erlang, considerado o pai da Teoria de Filas, desenvolveu os primeiros estudos teóricos para o problema do congestionamento de linhas telefônicas. Erlang aplicou conceitos de probabilidade ao problema e em 1909 apresentou o primeiro trabalho publicado neste assunto *The Theory of Probabilities and Telephone Conversations*, provando que ligações telefônicas distribuídas aleatoriamente seguiam a lei de distribuição de Poisson. Em 1917, publicou o *Solution of some Problems in the Theory of Probabilities of Significance in Automatic Telephone Exchanges*. Dos estudos de Erlang, muitos modelos quantitativos foram desenvolvidos para ajudar à tomada de decisão envolvendo filas de espera, com ênfase a partir da Segunda Guerra Mundial, quando a Teoria de Filas foi aplicada a outros problemas de filas. Dentre eles, podem-se citar: - *The Delay to Pedestrians Crossing a Road*(TANNER,1951), para o cálculo do tempo médio de espera de pedestres para atravessar uma rua sem sinal; - *Seaport Operation as a Stochastic*

Process(EVERETT,1953), para o problema de escoamento de fluxo de barcos em terminais portuários; - *Priority Assignment in Waiting-Line Problems*(COBHAM,1954), para o problema de reparo de maquinárias; - *Queues, Inventories and Maintenance*(MORSE,1962) e - *Queues and Inventories*(PRABHU,1965), para o problema de formação de estoques; - *On the Algebra of Queues*(KINGMAN,1966), para problemas de fluxo de tráfego; - *Queues Served in Cyclic Order*(COOPER & MURRAY,1969), para sistemas com filas múltiplas e servidor único; - *Probabilistic Models of Computer Systems*(GELENBE & MUNTZ,1976) e - *Decomposability, Queueing and Computer System Applications*(COURTOIS,1977), para problemas concernentes à Ciência da Computação; - *Open Queueing Networks: Optimization and Performance Evaluation Models for Discrete Manufacturing Systems*(BITRAN & MORABITO,1996), para sistemas de manufaturas; e - *Queues with a variable number of servers*(LI & YANG,2000), para problemas de utilização de servidores em redes digitais de comunicação. Mesmo com a evolução ao longo dos anos, a complexidade matemática envolvida nessa teoria tem deixado muitos problemas sem solução.

Os resultados obtidos pela modelagem de filas, podem auxiliar no dimensionamento da infra-estrutura física, dos recursos humanos e financeiros, possibilitando, assim, um melhor desempenho global dos mais variados sistemas. A avaliação desses resultados podem ter dois enfoques, na maioria das vezes bastante conflitantes:

- Usuário (cliente): deseja filas curtas e tempos de permanência(ou espera) reduzidos, menor tempo na fila.
- Gerente: deseja maiores volumes de serviços prestados.

A prestação de serviço seja ela feita por pessoas e/ou máquinas requer investimentos de capital e custos operacional e de manutenção. Assim, oferecer serviços além do necessário provoca custos que devem ser evitados. Por outro lado, não oferecer capacidade de serviço suficiente implica em formação de filas de espera excessivamente longas, que também pode ser cara por motivos diversos (custo social, perda de clientes).

À medida que o nível de serviço aumenta, o custo de espera diminui (redução da fila) e o custo de serviço aumenta (mais unidades de atendimento). A função de custo total é uma composição do custo de serviço com o custo de espera. Um sistema corretamente dimensionado deve permitir manter o equilíbrio entre o capital disponibilizado no sistema e os retornos sociais e financeiros do mesmo, ou seja, o balanceamento econômico entre o custo do serviço e o custo associado à espera pelo serviço

O balanceamento econômico pode ser obtido pela aplicação da Teoria de Filas e o 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 de *chegadas* e *atendimento*.

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

Para o desenvolvimento 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 separados apenas por pequenos intervalos de tempo, chegadas e saídas serem estatisticamente independentes e taxa média de chegadas ao sistema constante. Estas hipóteses são essenciais para que a Teoria de Filas seja aplicável.

Objetivo

Na maioria dos modelos de filas estudados é assumida a existência de um número fixo de servidores. Contudo, na prática, em muitas situações o número de servidores varia em função de algum critério estabelecido (ROMANI, 1957).

Segundo Moder & Phillips (1962), um exemplo de sistema com número de servidores variáveis é um serviço de cópias, que têm dois servidores, mas usualmente mantém um servidor no balcão de atendimento. Quando as chegadas acontecem e a fila cresce até

determinado tamanho, o segundo servidor deixa o serviço interno e ajuda no balcão.

Em alguns sistemas de filas é requerido que um certo nível de desempenho, como a média do tempo de espera, seja garantido para os clientes. No modelo clássico $M/M/c/\infty/FIFO$, a garantia de alguns requisitos acarreta na utilização ineficiente dos servidores. Uma solução para o conflito entre requisitos de desempenho do sistema e alta utilização dos servidores é ajustar o número de servidores dinamicamente com o tempo. Quando a mudança freqüente no número de servidores no sistema é economicamente viável, a taxa de utilização pode ser aumentada pelo ajuste do número de servidores de acordo com o grau de congestionamento(LI & YANG,2000).

O objetivo deste trabalho é apresentar o desenvolvimento da modelagem analítica, baseado em Teoria de Filas, de sistemas que apresentam flutuabilidade na taxa de chegadas e onde o número de servidores em paralelo passa a depender do número de elementos no sistema, isto é, quando a fila atingir 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 a fila decresce, os servidores são retirados de operação na mesma ordem que ficaram ociosos.

Justificativa

Alguns sistemas são caracterizados por apresentarem diferentes situações de congestionamento ao longo dos períodos avaliados, como bancos, supermercados, postos de pedágios, emergências médicas, entre outras. Dimensionar o sistema para os momentos de pico, pode representar um custo muito alto. Ignorar os momentos de maior movimento, pode gerar filas caóticas com possíveis perdas de usuários, o que também pode representar perdas.

É possível mostrar que se qualquer uma das hipóteses essenciais para aplicação de Teoria de Filas mencionadas anteriormente(exceto a taxa de chegadas ao sistema constante) não forem estritamente cumpridas, os resultados finais não se distanciam significativamente da realidade. Entretanto, a taxa de chegada flutuando significativamente ao longo do tempo leva a medidas de desempenho sub/superestimadas. Surge, assim, a necessidade de modelos como aqui proposto para avaliar este caso.

Organização do Trabalho

Após este capítulo introdutório, o trabalho é estruturado nos seguintes capítulos:

Capítulo 2 onde são apresentados os conceitos básicos dos modelos de filas;

Capítulo 3 consta da descrição do modelo tradicional de filas: $M/M/c/\infty/FIFO$;

Capítulo 4 onde se descreve um caso particular do modelo proposto, $M/M/c(k) \leq 2/\infty/FIFO$, de forma a possibilitar de maneira gradativa o entendimento do genérico;

Capítulo 5 onde se apresenta o modelo genérico, $M/M/c(k) \leq c/\infty/FIFO$;

Capítulo 6 que consta de um estudo de caso, onde as diferenças entre algumas medidas de desempenho dos modelos apresentados são verificadas e

Capítulo 7 que contém as considerações finais e propostas para trabalhos futuros.

O trabalho inclui ainda o **Apêndice A** onde são apresentados conceitos básicos necessários a dissertação para fonte de consulta quando necessário.

Capítulo 2

Características de um Modelo de Filas

2.1 Estrutura Básica

A estrutura básica dos modelos de filas compreende um processo onde os clientes¹ são oriundos de uma fonte (população) e buscam por determinado serviço. Sendo a demanda por serviço maior que a capacidade de atendimento, o cliente deve aguardar, saindo do sistema após ter o atendimento concluído.

Fisicamente, o sistema de filas é composto por: fonte de origem de clientes, espaço para a espera pelo serviço e postos de atendimento.

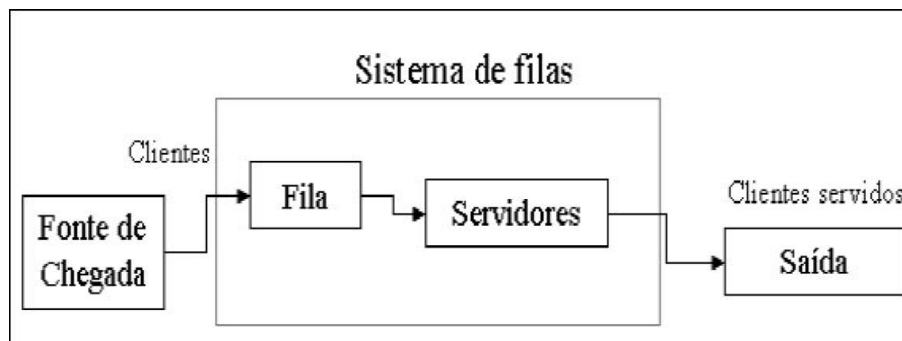


Figura 2.1: Processo de Fila Básico

¹O termo cliente é usado de maneira geral e não implica necessariamente num cliente humano.

O Processo de Chegada representa o comportamento do fluxo de chegada dos clientes ao sistema. Pode ser determinístico(ex.linha de montagem de automóveis) ou aleatório (processo estocástico com parâmetro λ =taxa de chegadas, representando o número médio de usuários que chegam ao sistema por unidade de tempo - ex.automóveis chegando em uma praça de pedágio).

Os usuários podem ser:

1. Originários de uma população finita(ex.computadores de uma *Lan House* que necessitam reparo) ou infinita(ex.elerodomésticos que chegam a uma oficina para reparo) ; e
2. Chegar individualmente(ex.chegada de navios em um terminal) ou em grupos(ex.pessoas em uma excursão chegando de ônibus a pontos turísticos).

O Processo de Atendimento representa o comportamento do fluxo de usuários atendidos, podendo ser individual(ex.médico atendendo pacientes) ou em grupo(ex.carros sendo transportados em um *Ferry Boat*).

Os Postos de Atendimento (servidores) são os locais(físicos ou não) onde os usuários são atendidos. Pode haver um número finito ou infinito de postos(diz-se que há um número infinito de postos de atendimento quando o serviço é do tipo *Self-Service*).

A capacidade do sistema é o número máximo de usuários que o sistema comporta, limitação física, que considera tanto os clientes na fila como aqueles em atendimento. Quando o sistema não possui um setor limitado é comum utilizar-se um valor infinito para representar esta capacidade física(ex.um porto onde os navios chegam para descarregamento, aguardando, se necessário, no mar).

A disciplina de atendimento refere-se ao critério com que os usuários de uma fila são selecionados para serem atendidos quando um posto de atendimento fica livre. Podem-se citar como as disciplinas mais utilizadas:

- FIFO(First In - First Out): os usuários são atendidos na ordem da chegada(ex.venda de ingressos numa bilheteria);
- LIFO(Last In - First Out): o último usuário a chegar é o primeiro a ser atendido(utilização de estoques verticais);

- PRI(Priority Service): os usuários são atendidos conforme a prioridade estabelecida pela gerência do sistema(agendamento de cirurgias hospitalares);
- SIRO(Service in Random Order): os usuários são atendidos em uma ordem aleatória(contemplação de consórcios);

2.2 Notação e Terminologia

Kendall(1953) propôs a notação $A/B/C/D/E$ para descrever um modelo de filas, onde:

A - distribuição dos tempos entre chegadas sucessivas;

B - distribuição dos tempos de atendimento;

C - número de postos de atendimento, quantidade de servidores, em paralelo;

D - capacidade física do sistema;

E - disciplina de atendimento empregada.

Quando somente os três primeiros símbolos forem utilizados, assume-se que o sistema possui capacidade ilimitada e disciplina FIFO.

As distribuições dos intervalos entre chegadas e dos tempos de atendimento possuem como formas mais comuns:

D - distribuição Determinística ou Degenerada;

M - distribuição Exponencial (*Memoryless* ou Markoviana);

E_k - distribuição Erlang do tipo k;

G - distribuição Geral (não especificada), onde média e desvio-padrão são conhecidos.

2.3 Medidas de Desempenho

Segundo Fogliatti & Mattos(2007), a Teoria de Filas permite avaliar a eficiência de um sistema por meio da análise de suas características utilizando medidas de desempenho.

Na maioria das vezes, essas características mudam ao longo do tempo, sendo então representadas por variáveis aleatórias cujos valores esperados podem ser utilizados como medidas de desempenho no regime estacionário.

Quando as características do sistema mudam significativamente, o sistema é dito estar em regime transiente. Caso contrário, isto é, a partir do instante de tempo em que essas características se estabilizam, diz-se que o sistema alcança o regime estacionário.

No regime estacionário, as principais medidas de desempenho são representadas por: Número médio de elementos no sistema(L), Número médio de elementos na fila(L_q), Tempo médio de permanência de um cliente qualquer no sistema(W), Tempo médio de espera um cliente qualquer na fila(W_q). Estas medidas de desempenho são dadas por:

$$L = E[i] = \sum_{i=0}^{\infty} iP_i \quad (2.1)$$

$$L_q = E[i_q] = \sum_{i=c}^{\infty} (i - c).P_i \quad (2.2)$$

Onde:

i = número de clientes no sistema;

i_q = número de clientes na fila;

P_i = probabilidade de i clientes no sistema.

No início da década de 60, Little(1961) relacionou o tamanho médio de um sistema em regime estacionário com o tempo médio de espera dos clientes:

$$L = \lambda W \quad (2.3)$$

A partir da relação (2.3), obtém-se as seguintes:

$$L_q = \lambda W_q \quad (2.4)$$

$$W = W_q + \frac{1}{\mu} \quad (2.5)$$

de onde obtém-se também:

$$L_q = L - \frac{\lambda}{\mu} \quad (2.6)$$

Destas relações pode ser concluído que basta calcular L ou L_q para obter-se W e W_q . Como L e L_q dependem das probabilidades de estados do sistema(P_i), pode ser observado que as medidas de desempenho dos modelos com filas dependem de P_i .

As medidas de desempenho podem, ainda, incluir a porcentagem de tempo que um servidor particular está ocioso ou o sistema está desprovido de usuários. Sendo $\lambda =$ taxa de chegadas e $\mu =$ taxa de atendimento de cada um dos c servidores disponíveis, tem-se que a taxa de ocupação do sistema(ρ) é dada por :

$$\rho = \frac{\lambda}{c\mu} \quad (2.7)$$

Se $\rho > 1$, o número médio de chegadas no sistema excede a taxa de atendimento do sistema($\rho > c\mu$), ou seja, a medida que o tempo passa, a fila se torna maior, sem qualquer perspectiva de equilíbrio. Assim, a condição para uma situação estacionária é que ρ deve ser estritamente menor que 1.

Se $\rho = 1$, a menos que seja determinístico e bem escalonado, o sistema não atinge o regime estacionário.

Capítulo 3

O Modelo $M/M/c/\infty/FIFO$

Este capítulo tem como propósito apresentar o modelo de filas $M/M/c/\infty/FIFO$, descrevendo e caracterizando-o, considerando os conceitos inerentes aos Processos de Nascimento-Morte¹, cujas transições de um estado qualquer somente são possíveis para estados vizinhos(ou seja, de um estado i para estados $i + 1$ ou $i - 1$).

Após a caracterização do modelo, são obtidas as medidas de desempenho: número médio de clientes na fila, número médio de clientes no sistema, tempo médio de espera na fila, tempo médio de ocupação do sistema e o tempo médio de ocupação do servidor, entre outras.

O modelo $M/M/c/\infty/FIFO$ representa uma realidade onde clientes ingressam no sistema individualmente, com tempo entre chegadas sucessivas seguindo uma distribuição exponencial de parâmetro(λ), buscando por um serviço. Esses clientes, ao ingressarem no sistema individualmente, entram numa fila única para serem atendidos por c servidores em paralelo à medida que estes ficam disponíveis para atendimento.

Assume-se a hipótese que os tempos de atendimento de cada servidor seguem uma distribuição exponencial de parâmetro(μ). Ainda referente ao cenário modelado, pode-se afirmar que não há limitação física do sistema, ou seja, independente do tamanho da fila ou da quantidade de clientes, não existem rejeições ou desistências e os clientes são atendidos na ordem de chegada.

¹Conceitos principais descritos no apêndice A.

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

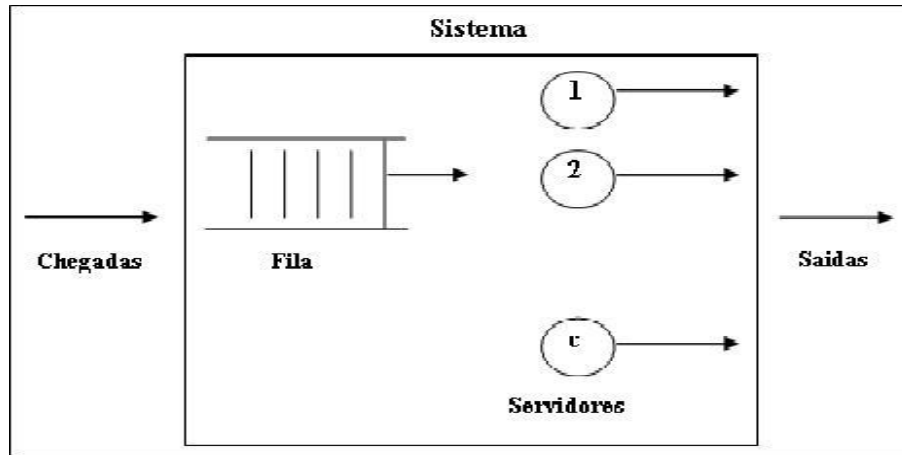


Figura 3.1: Modelagem M/M/c/∞/FIFO

Formalmente, segundo a notação proposta por Kendall(1953), para o modelo M/M/c/∞/FIFO, tem-se:

- M: tempos entre chegadas sucessivas seguem distribuição exponencial;
- M: tempos de atendimentos por servidor seguem distribuição exponencial;
- c: o número de postos de atendimento (servidores) em paralelo
- ∞: não são impostas limitações físicas ao modelo;
- FIFO: o atendimento dos clientes respeita a ordem das chegadas.

Os processos de chegada e de atendimento do sistema caracterizam um processo de nascimento e morte, ou seja, somente um único evento pode ocorrer em curtos períodos de tempo.

3.1 Caracterização do Sistema no Regime Estacionário

Seja um sistema em que os clientes chegam com taxa λ constante e são atendidos por um dos c servidores disponíveis com taxa individual μ . O sistema sob análise, supondo que este alcança o regime estacionário, pode ser representado graficamente pelo diagrama de fluxo da Figura (3.2).

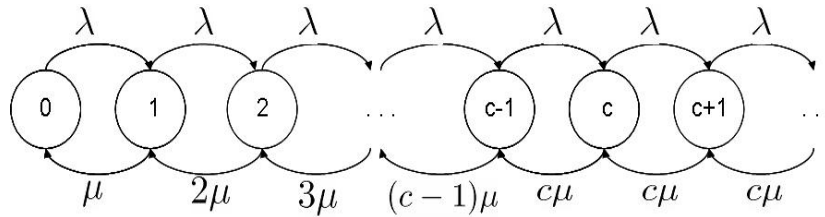


Figura 3.2: Diagrama de Fluxo para o modelo M/M/c/∞/FIFO

Denotando-se $r = \frac{\lambda}{\mu}$, tem-se:

$$\rho = \frac{r}{c} . \tag{3.1}$$

Aplicando o Princípio de Conservação de Energia(fluxo que entra em um nó é igual ao fluxo que sai desse mesmo nó) podem ser calculados as probabilidades P_i do sistema da Figura(3.2). Tem-se, assim:

Estado	Fluxo de entrada	Fluxo de saída
0	μP_1	λP_0
1	$2\mu P_2 + \lambda P_0$	$\mu P_1 + \lambda P_1$
2	$3\mu P_3 + \lambda P_1$	$2\mu P_2 + \lambda P_2$
⋮	⋮	⋮
c	$c\mu P_{c+1} + \lambda P_{c-1}$	$c\mu P_c + \lambda P_c$
⋮	⋮	⋮
i	$c\mu P_{i+1} + \lambda P_{i-1}$	$c\mu P_i + \lambda P_i$

Por recorrência, tem-se:

$$P_i = \begin{cases} P_0 \frac{1}{i!} \left(\frac{\lambda}{\mu} \right)^i, & \text{se } 1 \leq i \leq c; \\ P_0 \frac{c^c}{c!} \left(\frac{\lambda}{c\mu} \right)^i, & \text{se } i \geq c. \end{cases} \quad (3.2)$$

Substituindo $r = \frac{\lambda}{\mu}$ em (3.2), tem-se:

$$P_i = \begin{cases} P_0 \frac{r^i}{i!}, & \text{se } 1 \leq i \leq c \\ P_0 \left(\frac{c^c}{c!} \right) \left(\frac{r}{c} \right)^i, & \text{se } i \geq c; \end{cases} \quad (3.3)$$

Para calcular P_0 , lembra-se que $\sum_{i=0}^{\infty} P_i = 1$ e que

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

de onde:

$$\sum_{i=0}^{\infty} P_i = P_0 \left[\sum_{i=0}^{c-1} \frac{r^i}{i!} + \sum_{i=c}^{\infty} \frac{r^i}{c^{i-c} c!} \right] = P_0 \left[\sum_{i=0}^{c-1} \frac{r^i}{i!} + \frac{r^c}{c!} \sum_{i=c}^{\infty} \frac{r^{i-c}}{c^{i-c}} \right]. \quad (3.4)$$

Fazendo $i - c = k$ e substituindo em (3.4), tem-se:

$$1 = P_0 \left(\sum_{i=0}^{c-1} \frac{r^i}{i!} + \frac{r^c}{c!} \sum_{k=0}^{\infty} \frac{r^k}{c^k} \right) = P_0 \left(\sum_{i=0}^{c-1} \frac{r^i}{i!} + \frac{r^c}{c!} \sum_{k=0}^{\infty} \rho^k \right). \quad (3.5)$$

Para que a soma infinita da equação(3.5) seja convergente, ρ deve ser menor que um ($\rho < 1$), condição do sistema alcançar o regime estacionário, portanto:

$$1 = P_0 \left(\sum_{i=0}^{c-1} \frac{r^i}{i!} + \frac{r^c}{c!} \frac{1}{(1 - \rho)} \right) \quad (3.6)$$

de onde:

$$P_0 = \left(\sum_{i=0}^{c-1} \frac{r^i}{i!} + \frac{cr^c}{c!(c-r)} \right)^{-1}. \quad (3.7)$$

3.2 Medidas de Desempenho para o Modelo $M/M/c/\infty/FIFO$

A seguir são apresentadas as principais medidas de desempenho para o modelo $M/M/c/\infty/FIFO$: número médio de usuários na fila, número médio de usuários no sistema, tempo médio de espera na fila, tempo médio de permanência de um usuário no sistema, assim como a probabilidade de ócio parcial do sistema e a probabilidade de ócio de um servidor qualquer.

3.2.1 Número Médio de Elementos na Fila(L_q)

Seja i_q a variável aleatória discreta que representa o número de usuários na fila, o número médio de elementos na fila(L_q) é seu valor esperado.

Somente acontecerá a formação de fila, a partir do momento em que o número de usuários for maior ou igual ao número de servidores, ou seja, $i \geq c$. Tem-se, neste caso:

$$L_q = E[i_q] = \sum_{i=c}^{\infty} (i-c)P_i \quad (3.8)$$

Substituindo (3.3) em (3.8), tem-se:

$$\begin{aligned} L_q &= P_0 \frac{r^{c+1}}{c} \frac{1}{c!} \sum_{i=c}^{\infty} \frac{(i-c)r^{i-c-1}}{c^{i-c-1}} \\ &= P_0 \frac{r^{c+1}}{c} \frac{1}{c!} \frac{d}{d\left(\frac{r}{c}\right)} \sum_{i=0}^{\infty} \left(\frac{r}{c}\right)^i \\ &= P_0 \frac{r^{c+1}}{c} \frac{1}{c!} \frac{d\left(\frac{1}{1-\frac{r}{c}}\right)}{d\left(\frac{r}{c}\right)} \end{aligned}$$

$$= P_0 \frac{r^{c+1}}{c} \frac{1}{c!} \frac{1}{\left(1 - \frac{r}{c}\right)^2} .$$

Assim,

$$L_q = P_0 \frac{r^{c+1}c}{c!(c-r)^2} \quad (3.9)$$

3.2.2 Número Médio de Elementos no Sistema(L)

Seja \underline{i} a variável aleatória discreta que representa o número de clientes no sistema. O número médio de elementos no sistema(L) é seu valor esperado.

Substituindo (3.9) em (2.6) , tem-se:

$$L = r + P_0 \frac{r^{c+1}c}{c!(c-r)^2} . \quad (3.10)$$

3.2.3 Tempo Médio de Espera na Fila(W_q)

Substituindo (3.9) em (2.5) e desenvolvendo, tem-se:

$$W_q = P_0 \frac{r^c \mu}{(c-1)!(c\mu - \lambda)^2} . \quad (3.11)$$

3.2.4 Tempo Médio de Permanência no Sistema(W)

A partir de 2.5, tem-se

$$W = P_0 \frac{r^c \mu}{(c-1)!(c\mu - \lambda)^2} + \frac{1}{\mu} . \quad (3.12)$$

3.2.5 Probabilidade de Ócio Parcial do Sistema

A probabilidade do sistema encontrar-se em ócio parcial, ou seja, de um cliente qualquer chegar e ser atendido sem ter que aguardar equivale a ter no sistema um número de clientes estritamente menor que o número de servidores disponíveis, ou

seja,

$$P(i < c) = \sum_{i=0}^{c-1} P_i = P_0 \sum_{i=0}^{c-1} \frac{r^i}{i!} . \quad (3.13)$$

3.2.6 Probabilidade de Ócio de um Servidor Qualquer (P_{oc})

Para um servidor qualquer encontrar-se ocioso devem existir no sistema um número de clientes ($i < c$). Satisfeita esta condição, a probabilidade de um servidor qualquer encontrar-se ocioso, P_{oc} , é a probabilidade condicional do servidor estar ocioso dado que o sistema encontra-se em ócio parcial.

$$P_{oc} = \sum_{i=0}^{c-1} P(\text{de servidor ocioso entre } c \text{ servidores disponíveis}) P_i \quad (3.14)$$

onde $P(\text{de servidor ocioso entre } c \text{ servidores disponíveis}) = \frac{c-i}{c}$.

Assim, tem-se:

$$\begin{aligned} i = 0 & \quad P_{oc} = \frac{c-0}{c} P_0 \\ i = 1 & \quad P_{oc} = P_0 + \frac{c-1}{c} P_1 \\ i = 2 & \quad P_{oc} = P_0 + \frac{c-1}{c} P_1 + \frac{c-2}{c} P_2 \\ & \quad \vdots \\ i = c-1 & \quad P_{oc} = P_0 + \frac{c-1}{c} P_1 + \frac{c-2}{c} P_2 + \dots + \frac{c-i}{c} P_i . \end{aligned}$$

Para $i < c$, tem-se:

$$P_{oc} = \sum_{i=0}^{c-1} \frac{(c-i)}{c} P_i = P_0 \frac{1}{c} \sum_{i=0}^{c-1} \frac{(c-i)r^i}{i!} \quad (3.15)$$

3.2.7 Considerações Finais

Caracterizado o modelo $M/M/c/\infty/FIFO$ e apresentadas as correspondentes medidas de desempenho, deve ser lembrado que para a obtenção destas, algumas simplificações da realidade, bem como algumas suposições básicas, foram estabelecidas. Dentre estas, destaca-se o fato da taxa de chegada ser constante.

No regime estacionário, se a taxa de chegada ao sistema apresentar um comportamento oscilante, ou seja, para um sistema com diferentes situações de congestionamento ao longo do período observado, as medidas de desempenho obtidas podem ser sub/superestimadas. Assim, faz-se necessário adotar algum procedimento que permita equilibrar as flutuabilidades, o que será apresentado no capítulo seguinte onde se apresenta a variação do número de servidores em paralelo.

Capítulo 4

O Modelo $M/M/c(k) \leq 2/\infty/FIFO$

Na maioria dos modelos de filas, o número de servidores é assumido fixo. Entretanto, na prática são encontradas muitas situações em que o número de servidores varia em função dos clientes em espera.

Este capítulo tem como propósito apresentar um modelo de filas com número variável de servidores, representando um sistema onde a flutuabilidade de demanda não pode ser controlada, isto é, a taxa de chegadas sofre variações significativas ao longo do tempo que podem comprometer a análise do sistema usando modelos como o do capítulo anterior. Uma forma de analisar mais realisticamente o desempenho de sistemas deste novo tipo é variando o número de servidores até um dado limite(c), de modo a atender a demanda satisfatoriamente, reduzindo custos operacionais.

Vários pesquisadores têm tratado modelos deste tipo:

Romani(1957) apresenta um modelo que não permite a fila aumentar além de um tamanho M especificado. Sempre que a fila atingir M clientes e houver uma nova chegada, o número de servidores é aumentado de um, não havendo limite para a quantidade de servidores que possam ser adicionados. Quando não há clientes na fila e os serviços são completados, os servidores são retirados de operação na mesma ordem que ficaram ociosos, até que reste um servidor, que permanece em operação mesmo que o sistema se encontre vazio.

Davis et al.(1959) apresentam um modelo de filas com número variável de servidores, limitado a dois, que atende a restrição de um determinado número N de

clientes na fila para a entrada(ou retirada) do segundo servidor em operação. O *Shift Point*¹ é o mesmo para entrada ou retirada do servidor. Quando ambos servidores estão em operação a fila é ilimitada.

Moder & Phillips(1962) apresentam um modelo de filas com número variável de servidores limitado e com *Shift Points* diferentes para entrada e retirada de servidores, ou seja, os servidores são adicionados de modo a atender a restrição de um número N de clientes na fila até o limite disponível. Toda vez que a fila atinge o valor especificado um novo servidor é posto em operação, quando todos os servidores disponíveis estão em operação a fila passa a ser ilimitada. Quando a fila diminui até um valor $\underline{\gamma}$ especificado, os servidores são retirados de operação à medida que ficam ociosos. O *Shift Point* para entrada e retirada de servidores são diferentes.

Al-Seedy(2004,2006) apresentam modelos de filas com servidores variáveis que introduzem os conceitos de desistência e rejeição.

Silva, Santos & Fogliatti(2007) apresentam um modelo limitado até dois servidores, mesmo *Shift Point* para entrada e retirada do novo servidor, atendendo a restrição do número de clientes no sistema.

Em todos os modelos apresentados é assumido que o tempo de adição ou retirada de um novo servidor é imediato.

Neste trabalho, o número de servidores é uma variável aleatória e uma função do comprimento da fila. Segundo Fogliatti & Mattos(2007), a modelagem analítica desse tipo de sistema não é trivial. Portanto, de forma a possibilitar um melhor entendimento, apresenta-se de maneira gradativa a caracterização do modelo dividida em duas etapas: o caso particular, limitado até dois servidores, apresentado neste capítulo e o modelo completo apresentado no capítulo seguinte.

Como no modelo tradicional (M/M/c/ ∞ /FIFO), é mantida a representação de uma realidade onde os clientes ingressam no sistema com tempo entre chegadas sucessivas seguindo uma distribuição exponencial de parâmetro(λ), buscando por um serviço.

Não há limitação física, ou seja, não há limitação para o tamanho da fila ou da quantidade de clientes no sistema e não existem rejeições ou desistências. Neste modelo, os clientes entram numa fila simples e são atendidos, na ordem de

¹Estado no qual o número de servidores se altera.

chegada, por até 2 servidores em paralelo. Assume-se a hipótese de que os tempos de atendimento dos 2 servidores seguem distribuição exponencial de parâmetro (μ).

4.1 $M/M/c(k) \leq 2/\infty/\text{FIFO}$

O caso particular representa um sistema limitado a dois servidores ($1 \leq c(k) \leq 2$), que atende a restrição de um determinado número de clientes na fila (k clientes), que equivale a $k + 1$ clientes no sistema. Assim, toda vez que o sistema atinge o estado $k + 1$ e ocorre uma nova chegada, o segundo servidor é posto em operação e a fila passa a ser ilimitada. Quando o número de clientes no sistema decresce ($i \leq k + 1$) e o serviço é completado, o primeiro servidor a ficar livre é retirado de operação. O *Shift Point* é o mesmo para entrada ou saída do segundo servidor.

O estado estacionário sob análise, supondo que este existe, é representado pelo diagrama de fluxo da Figura 4.1, a seguir:

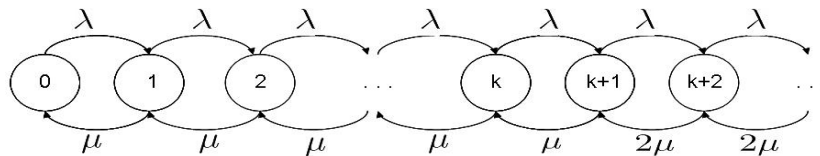


Figura 4.1: Diagrama de Fluxos para o modelo $M/M/c(k) \leq 2/\infty/\text{FIFO}$

Observa-se deste diagrama que existem estados que têm comportamento semelhante, quais sejam: o grupo de estados i ($0 < i < k + 1$ e $i \geq k + 2$), enquanto os estados 0 e $k + 1$ têm comportamento diferenciado. Portanto, estes estados devem ser analisados separadamente.

Pelo Princípio de Conservação de Energia aplicado ao diagrama da figura 4.1, tem-se:

Estado	Fluxo de entrada	Fluxo de saída
0	μP_1	λP_0
$0 < i < k + 1$	$\mu P_{i+1} + \lambda P_{i-1}$	$(\mu + \lambda) P_i$
$i = k + 1$	$2\mu P_{k+2} + \lambda P_k$	$(\mu + \lambda) P_{k+1}$
$i \geq k + 2$	$2\mu P_{i+1} + \lambda P_{i-1}$	$(2\mu + \lambda) P_i$

Tem-se então,

para o estado(0),

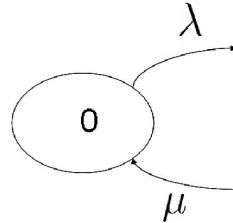


Figura 4.2: Estado Inicial

$$P_1 = P_0 \left(\frac{\lambda}{\mu} \right). \quad (4.1)$$

Para o estado(i), com $0 < i < k$

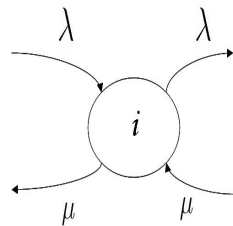


Figura 4.3: Estado i

$$\mu P_{i+1} + \lambda P_{i-1} = P_i(\lambda + \mu)$$

Então:

$$\mu P_{i+1} = P_0 \left(\frac{\lambda}{\mu} \right)^{i-1} \left(\frac{\lambda}{\mu}(\lambda + \mu) - \lambda \right)$$

$$= P_0 \left(\frac{\lambda}{\mu} \right)^{i-1} \left(\frac{\lambda^2}{\mu} + \lambda - \lambda \right)$$

de onde:

$$P_{i+1} = P_0 \left(\frac{\lambda}{\mu} \right)^{i+1} \quad (4.2)$$

Para o estado (k) , tem-se

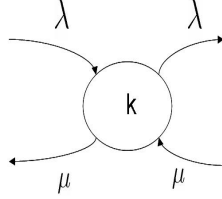


Figura 4.4: Estado k

$$\mu P_{k+1} + \lambda P_{k-1} = P_k(\lambda + \mu)$$

de onde

$$\begin{aligned} P_{k+1} &= P_0 \left(\frac{\lambda}{\mu} \right)^{k-1} \left(\frac{\lambda}{\mu} \right)^2 \\ &= P_0 \left(\frac{\lambda}{\mu} \right)^{k+1} \end{aligned} \quad (4.3)$$

Para o estado $(k + 1)$, tem-se

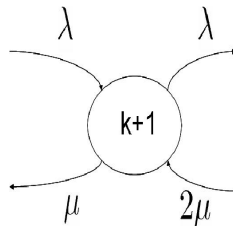


Figura 4.5: Estado $k+1$

$$2\mu P_{k+2} + \lambda P_k = P_{k+1}(\lambda + \mu)$$

então:

$$2\mu P_{k+2} = P_{k+1}(\lambda + \mu) - \lambda P_k$$

$$= P_k \left(\frac{\lambda}{\mu} (\lambda + \mu) - \lambda \right)$$

de onde:

$$\begin{aligned} P_{k+2} &= P_0 \left(\frac{\lambda}{\mu} \right)^k \left(\frac{\lambda^2}{\mu} \right) \left(\frac{1}{2\mu} \right) \\ &= P_0 \left(\frac{\lambda}{\mu} \right)^{k+2} \left(\frac{1}{2} \right) \end{aligned} \quad (4.4)$$

Para o estado ($i \geq k + 2$), tem-se

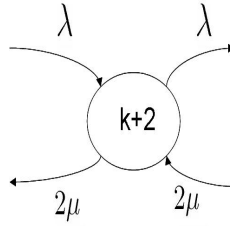


Figura 4.6: Estado k+2

$$2\mu P_{k+3} + \lambda P_{k+1} = P_{k+2}(\lambda + 2\mu)$$

então:

$$\begin{aligned} 2\mu P_{k+3} &= P_{k+2}(\lambda + 2\mu) - \lambda P_{k+1} \\ &= P_{k+1} \left(\frac{\lambda}{2\mu} (\lambda + 2\mu) - \lambda \right) \\ &= P_0 \left(\frac{\lambda}{\mu} \right)^{k+1} \left(\frac{\lambda^2}{2\mu} + \lambda - \lambda \right) \end{aligned}$$

de onde:

$$\begin{aligned} P_{k+3} &= P_0 \left(\frac{\lambda}{\mu} \right)^{k+1} \left(\frac{\lambda^2}{2\mu} \right) \left(\frac{1}{2\mu} \right) \\ &= P_0 \left(\frac{\lambda}{\mu} \right)^{k+3} \left(\frac{1}{2} \right)^2 \end{aligned} \quad (4.5)$$

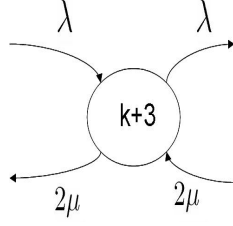


Figura 4.7: Estado $k+3$

$$2\mu P_{k+4} + \lambda P_{k+2} = P_{k+3}(\lambda + 2\mu)$$

então:

$$\begin{aligned} 2\mu P_{k+4} &= P_{k+3}(\lambda + 2\mu) - \lambda P_{k+2} \\ &= P_{k+2} \left(\frac{1}{2} \right) \left(\frac{\lambda}{2\mu}(\lambda + 2\mu) - \lambda \right) \\ &= P_0 \left(\frac{\lambda}{\mu} \right)^{k+2} \left(\frac{1}{2} \right) \left(\frac{\lambda^2}{2\mu} + \lambda - \lambda \right) \end{aligned}$$

de onde:

$$\begin{aligned} P_{k+4} &= P_0 \left(\frac{\lambda}{\mu} \right)^{k+2} \left(\frac{1}{2} \right) \left(\frac{\lambda^2}{2\mu} \right) \left(\frac{1}{2\mu} \right) \\ &= P_0 \left(\frac{\lambda}{\mu} \right)^{k+4} \left(\frac{1}{2} \right)^3 \end{aligned} \tag{4.6}$$

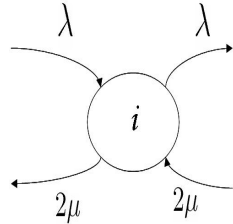


Figura 4.8: Estado i com $i \geq k + 2$

$$2\mu P_{i+1} + \lambda P_{i-1} = P_i(\lambda + 2\mu)$$

de onde:

$$\begin{aligned} P_{i+1} &= P_0 \left(\frac{\lambda}{\mu} \right)^{i+1} \left(\frac{1}{2} \right)^{i+1-(k+1)} \\ P_i &= P_0 \left(\frac{\lambda}{\mu} \right)^i \left(\frac{1}{2} \right)^{i-(k+1)}. \end{aligned} \tag{4.7}$$

Tem-se, então:

$$P_i = \begin{cases} P_0 \rho^i, & \text{se } 1 \leq i < k+2; \\ P_0 \rho^i \left(\frac{1}{2}\right)^{i-(k+1)}, & \forall i \geq k+2 \text{ com } \rho = \frac{\lambda}{\mu}. \end{cases} \quad (4.8)$$

Substituindo as equações (4.8) em:

$$\sum_{i=0}^{\infty} P_i = \sum_{i=0}^{k+1} P_i + \sum_{i=k+2}^{\infty} P_i$$

tem-se

$$\begin{aligned} & P_0 \sum_{i=0}^{k+1} \rho^i + P_0 \sum_{i=k+2}^{\infty} \rho^i \left(\frac{1}{2}\right)^{i-(k+1)} = \\ & = P_0 \left[\sum_{i=0}^{k+1} \rho^i + 2^{k+1} \left(\sum_{i=0}^{\infty} \left(\frac{\rho}{2}\right)^i - \sum_{i=0}^{k+1} \left(\frac{\rho}{2}\right)^i \right) \right]. \end{aligned}$$

A condição necessária para que a soma infinita acima seja convergente e o sistema alcançar o regime estacionário é $\frac{\rho}{2} < 1$.

Satisfeita esta condição, obtém-se:

para $\rho \neq 1$

$$\begin{aligned} \sum_{i=0}^{\infty} P_i &= P_0 \left[\frac{1 - \rho^{k+2}}{1 - \rho} + 2^{k+1} \left(\frac{1}{1 - \frac{\rho}{2}} - \frac{1 - \left(\frac{\rho}{2}\right)^{k+2}}{1 - \frac{\rho}{2}} \right) \right] = \\ &= P_0 \left[\frac{1 - \rho^{k+2}}{1 - \rho} + \frac{\rho^{k+2}}{2 - \rho} \right] = P_0 \frac{2 - \rho - \rho^{k+2}}{(1 - \rho)(2 - \rho)}. \end{aligned}$$

Como $\sum_{i=0}^{\infty} P_i = 1$,

obtém-se

$$P_0 = \frac{(1 - \rho)(2 - \rho)}{2 - \rho - \rho^{k+2}}. \quad (4.9)$$

Para $\rho = 1$

$$\begin{aligned}
\sum_{i=0}^{\infty} P_i &= P_0 \left[\sum_{i=0}^{k+1} 1^i + 2^{k+1} \left(\sum_{i=0}^{\infty} \left(\frac{1}{2}\right)^i - \sum_{i=0}^{k+1} \left(\frac{1}{2}\right)^i \right) \right] \quad (4.10) \\
&= P_0 \left\{ k + 2 + 2^{k+1} \left[2 - 2 + 2 \left(\frac{1}{2}\right)^{k+2} \right] \right\} = P_0 \left\{ k + 2 + 2^{k+1} \left(\frac{1}{2}\right)^{k+1} \right\} = \\
&= P_0(k + 3) .
\end{aligned}$$

Como $\sum_{i=0}^{\infty} P_i = 1$, tem-se

$$P_0 = \frac{1}{k + 3} . \quad (4.11)$$

Substituindo (4.9) e (4.11) em (4.8), tem-se então:

para $\rho \neq 1$

$$P_i = \begin{cases} \rho^i \left(\frac{(1-\rho)(2-\rho)}{2-\rho-\rho^{k+2}} \right), & \text{se } 0 \leq i < k+2; \\ \rho^i \left(\frac{1}{2}\right)^{i-(k+1)} \left(\frac{(1-\rho)(2-\rho)}{2-\rho-\rho^{k+2}} \right), & \forall i \geq k+2 \end{cases} \quad (4.12)$$

e para $\rho = 1$

$$P_i = \begin{cases} \frac{1}{k+3}, & \text{se } 0 \leq i < k+2; \\ \left(\frac{1}{2}\right)^{i-(k+1)} \left(\frac{1}{k+3} \right), & \forall i \geq k+2 . \end{cases} \quad (4.13)$$

A partir destas probabilidades podem ser obtidos as medidas de desempenho de interesse do sistema.

4.2 Medidas de Desempenho

Número Médio de Elementos no Sistema(L)

$$L = \sum_{i=0}^{\infty} iP_i = \sum_{i=0}^{k+1} iP_i + \sum_{i=k+2}^{\infty} iP_i \quad (4.14)$$

Satisfeita a condição de convergência($\frac{\rho}{2} < 1$), tem-se:

$$L = \sum_{i=0}^{k+1} i\rho^i P_0 + \sum_{i=k+2}^{\infty} i\rho^i \left(\frac{1}{2}\right)^{i-(k+1)} P_0 .$$

Para $\rho \neq 1$

$$\begin{aligned} L &= P_0 \left[\sum_{i=0}^{k+1} i\rho^i + 2^{k+1} \left(\sum_{i=k+2}^{\infty} i \left(\frac{\rho}{2}\right)^i \right) \right] = \\ &= P_0 \left[\sum_{i=0}^{k+1} i\rho^i + 2^{k+1} \left(\sum_{i=0}^{\infty} i \left(\frac{\rho}{2}\right)^i - \sum_{i=0}^{k+1} i \left(\frac{\rho}{2}\right)^i \right) \right] = \\ &= P_0 \left[\rho \sum_{i=1}^{k+1} i\rho^{i-1} + 2^{k+1} \left(\frac{\rho}{2} \sum_{i=1}^{\infty} i \left(\frac{\rho}{2}\right)^{i-1} - \frac{\rho}{2} \sum_{i=1}^{k+1} i \left(\frac{\rho}{2}\right)^{i-1} \right) \right] = \\ &= P_0 \left[\rho \frac{d}{d\rho} \sum_{i=0}^{k+1} \rho^i + \frac{\rho}{2} 2^{k+1} \left(\frac{d}{d\frac{\rho}{2}} \sum_{i=0}^{\infty} \left(\frac{\rho}{2}\right)^i - \frac{d}{d\frac{\rho}{2}} \sum_{i=0}^{k+1} \left(\frac{\rho}{2}\right)^i \right) \right] = \\ &= \rho P_0 \left[\frac{1 + \rho^{k+1}(k\rho + \rho - k - 2)}{(1 - \rho)^2} + \left(\frac{2^{k+2}}{(2 - \rho)^2} \right) \left(\left(\frac{\rho}{2}(k - 1) + k + 2\right) \left(\frac{\rho}{2}\right)^{k+1} \right) \right] = \\ &= \rho P_0 \left[\frac{1 + \rho^{k+1}(k\rho + \rho - k - 2)}{(1 - \rho)^2} + \frac{\rho^{k+1}}{(2 - \rho)^2} (k\rho - \rho + 2k + 4) \right] . \\ L &= P_0 \left[\frac{\rho (1 + \rho^{k+1}(k\rho + \rho - k - 2))}{(1 - \rho)^2} + \frac{\rho^{k+2} (4 + 2k - k\rho - \rho)}{(2 - \rho)^2} \right] . \quad (4.15) \end{aligned}$$

E para $\rho = 1$, tem-se:

$$\begin{aligned} L &= P_0 \left(\sum_{i=0}^{k+1} i \rho^i + \sum_{i=k+2}^{\infty} i \rho^i \left(\frac{1}{2} \right)^{i-(k+1)} \right) = \\ &= P_0 \left[\sum_{i=0}^{k+1} i + 2^{k+1} \left(\sum_{i=k+2}^{\infty} i \left(\frac{1}{2} \right)^i \right) \right]. \end{aligned}$$

$$\sum_{i=0}^{k+1} i = \frac{(k+1)(k+2)}{2}$$

$$\sum_{i=k+2}^{\infty} i \left(\frac{1}{2} \right)^i = \frac{k+3}{2^{k+1}} \text{ (Soma dos termos de uma Série Aritmético-Geométrica}^2\text{)}.$$

Assim, tem-se:

$$L = \frac{1}{k+3} \left[\frac{(k+1)(k+2)}{2} + (k+3) \right]$$

ou ainda,

$$L = P_0 \frac{(k+1)(k+2)}{2} + 1 \tag{4.16}$$

A partir das fórmulas de Little e suas decorrentes, (2.3),(2.4),(2.5) e (2.6), obtém-se: L_q , W_q e W .

Probabilidade de Ócio de um Servidor Qualquer(P_{oc})

Para o intervalo de tempo em que o sistema trabalha com um servidor ativo: ambos servidores estarão ociosos para atendimento quando o sistema estiver vazio e um dos servidores estará ocioso para atendimento quando houver no sistema entre 1 e $k+1$ clientes. Já para o intervalo de tempo durante o qual os dois servidores estão em operação, qualquer que seja o estado $i \geq k+2$, não existirá ócio.

²O desenvolvimento desta série encontra-se no apêndice A.

Assim, sendo P_{oc} a probabilidade de ócio de um servidor qualquer, tem-se:

$$\begin{aligned} P_{oc} &= P_0 + \frac{1}{2} \sum_{i=1}^{k+1} P_i \\ &= P_0 + \frac{1}{2} \sum_{i=1}^{k+1} \rho^i P_0 = P_0 \left(1 + \frac{1}{2} \sum_{i=1}^{k+1} \rho^i \right) \end{aligned}$$

Para $\rho \neq 1$

$$\begin{aligned} P_{oc} &= P_0 \left[1 + \frac{1}{2} \left(\frac{1 - \rho^{k+2}}{1 - \rho} - 1 \right) \right] = P_0 \left[1 + \frac{1}{2} \left(\frac{1 - \rho^{k+2}}{1 - \rho} \right) - \frac{1}{2} \right] = \\ &= P_0 \left(\frac{2 - \rho - \rho^{k+2}}{2(1 - \rho)} \right) . \end{aligned}$$

Como $P_0 = \frac{(1 - \rho)(2 - \rho)}{2 - \rho - \rho^{k+2}}$,

$$\begin{aligned} P_{oc} &= \frac{1}{2} \left(\frac{(1 - \rho)(2 - \rho)}{2 - \rho - \rho^{k+2}} \right) \left(\frac{2 - \rho - \rho^{k+2}}{1 - \rho} \right) = \\ &= 1 - \frac{\rho}{2} \end{aligned} \tag{4.17}$$

E para $\rho = 1$

$$P_{oc} = P_0 \left[1 + \frac{1}{2}(k + 1) \right]$$

de onde

$$P_{oc} = \frac{1}{k + 3} \left(\frac{2 + k + 1}{2} \right) = \frac{1}{2} . \tag{4.18}$$

Capítulo 5

O Modelo $M/M/c(k) \leq c/\infty/FIFO$

Neste capítulo apresentam-se as características e as medidas de desempenho do modelo $M/M/c(k) \leq c/\infty/FIFO$. Partindo-se de um servidor fixo em operação, agregam-se novos servidores até o limite c , atendendo a restrição de até k clientes na fila, à medida que existam na fila $k+1, k+2, k+3, \dots, k+c-1, k+c$ clientes. Quando a quantidade de clientes diminui, os servidores são retirados à medida que ficam ociosos até que reste um servidor em operação, obedecendo os mesmos *Shift Points*, ou seja, $k+c, k+c-1, \dots, k+3, k+2, k+1$. Verifica-se então, que ao agregar-se um servidor se o total de servidores em operação não forem capazes de manterem até k clientes na fila, um novo servidor é agregado até o limite de c servidores disponíveis para o processo. Considera-se ainda que os tempos de atendimento dos c servidores disponíveis estão exponencialmente distribuídos com taxa μ .

O estado estacionário sob análise, supondo que este existe, é representado pelo Diagrama de Fluxo da Figura 5.1, a seguir:

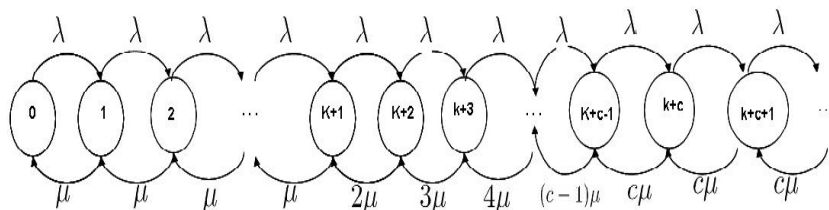


Figura 5.1: Diagrama de Fluxos para o modelo $M/M/c(k)/\infty/FIFO$

Pelo diagrama de fluxo pode-se observar que o estado(0) tem comportamento único, os estados $0 < i < k + 1$ tem comportamento idêntico, o estado $(k + 1)$ tem comportamento particular, os estados $k + 1 < i < k + c$ comportam-se de maneira semelhante e os estados $i \geq k + c$ apresentam comportamento idênticos. Assim, estes estados serão tratados separadamente.

Analisando o Diagrama de Fluxo e utilizando o Princípio de Conservação de Energia:

para o estado (0),

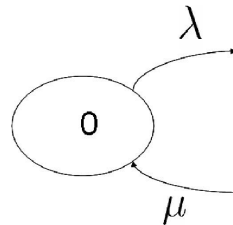


Figura 5.2: Estado Inicial

obté-m-se:

$$P_1 = P_0 \left(\frac{\lambda}{\mu} \right)$$

para o estado(i) $0 < i < k + 1$

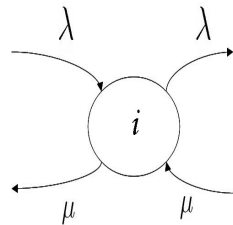


Figura 5.3: Estado i

$$P_i = P_0 \left(\frac{\lambda}{\mu} \right)^i . \tag{5.1}$$

Analisando o estado(k), tem-se:

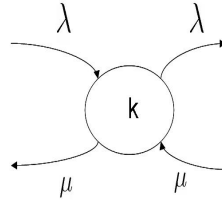


Figura 5.4: Estado k

de onde

$$\mu P_{k+1} + \lambda P_{k-1} = (\lambda + \mu)P_k \ .$$

Assim,

$$\begin{aligned} \mu P_{k+1} &= (\lambda + \mu)P_k - \lambda P_{k-1} \\ &= P_{k-1} \left(\frac{\lambda}{\mu}(\lambda + \mu) - \lambda \right) \\ &= P_0 \left(\frac{\lambda}{\mu} \right)^{k-1} \left(\frac{\lambda^2}{\mu} + \lambda - \lambda \right) \ . \end{aligned}$$

$$P_{k+1} = P_0 \left(\frac{\lambda}{\mu} \right)^{k+1} \ . \tag{5.2}$$

Analisando o estado($k + 1$), tem-se:

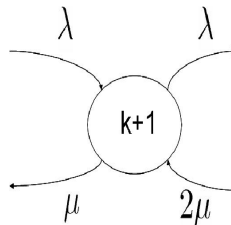


Figura 5.5: Estado $k+1$

de onde

$$(\lambda + \mu)P_{k+1} = \lambda P_k + 2\mu P_{k+2} \ .$$

Então

$$2\mu P_{k+2} = (\lambda + \mu)P_{k+1} - \lambda P_k .$$

Substituindo P_{k+1} e P_k segundo (5.1), tem-se:

$$\begin{aligned} 2\mu P_{k+2} &= (\lambda + \mu)P_0 \left(\frac{\lambda}{\mu}\right)^{k+1} - \lambda P_0 \left(\frac{\lambda}{\mu}\right)^k = \\ &= P_0 \left(\frac{\lambda}{\mu}\right)^k \left(\frac{\lambda^2}{\mu} + \lambda - \lambda\right) . : \end{aligned}$$

$$P_{k+2} = \left(\frac{1}{2}\right) P_0 \left(\frac{\lambda}{\mu}\right)^{k+2} \quad (5.3)$$

Análise do estado($k + 2$),

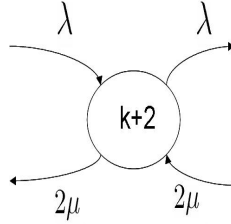


Figura 5.6: Estado $k+2$

tem-se:

$$3\mu P_{k+3} + \lambda P_{k+1} = (\lambda + 2\mu)P_{k+2}$$

de onde

$$3\mu P_{k+3} = P_0 \left(\frac{\lambda}{\mu}\right)^{k+1} \left[\frac{\lambda}{2\mu}(\lambda + 2\mu) - \lambda \right]$$

Assim,

$$P_{k+3} = \left(\frac{1}{2}\right) \left(\frac{1}{3}\right) P_0 \left(\frac{\lambda}{\mu}\right)^{k+3} \quad (5.4)$$

Análise do estado($k + 3$)

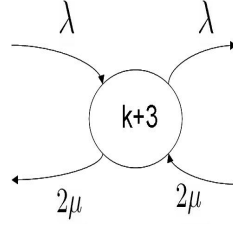


Figura 5.7: Estado $k+3$

tem-se:

$$4\mu P_{k+4} + \lambda P_{k+2} = (\lambda + 3\mu)P_{k+3}$$

de onde,

$$4\mu P_{k+4} = P_0 \left(\frac{\lambda}{\mu}\right)^{k+2} \left(\frac{1}{2}\right) \left[\frac{\lambda}{3\mu}(\lambda + 3\mu) - \lambda\right]$$

Assim,

$$P_{k+4} = \left(\frac{1}{2}\right) \left(\frac{1}{3}\right) \left(\frac{1}{4}\right) P_0 \left(\frac{\lambda}{\mu}\right)^{k+4} \quad (5.5)$$

Análise do estado(i), $k + 1 < i < k + c$

Denotando-se por s o número de servidores em operação, para $k + 1 < i < k + c$,

$2 \leq s < c$, tem-se:

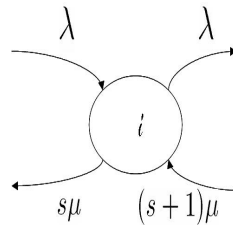


Figura 5.8: Estado i ($k + 1 < i < k + c$)

$$\lambda P_{i-1} + (s + 1)\mu P_{i+1} = (\lambda + s\mu)P_i$$

de onde,

$$P_{i+1} = \frac{1}{(s+1)!} P_0 \left(\frac{\lambda}{\mu} \right)^{i+1} .$$

Assim,

$$P_i = \left(\frac{1}{s!} \right) P_0 \left(\frac{\lambda}{\mu} \right)^i , \quad \forall \quad k+2 \leq i < k+c .$$

Pelo diagrama de fluxo pode-se observar que:

Estado(i)	Servidores em Operação(s)
$k+2$	2
$k+3$	3
$k+4$	4
\vdots	\vdots
$k+c$	c

Assim, $s = i - k$, então:

$$P_i = \frac{1}{(i-k)!} P_0 \left(\frac{\lambda}{\mu} \right)^i , \quad \forall \quad k+2 \leq i < k+c . \quad (5.6)$$

Análise do estado($k+c-1$)

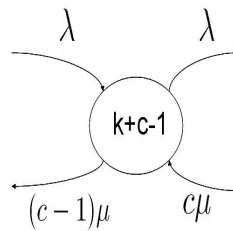


Figura 5.9: Estado $k+c-1$

tem-se:

$$c\mu P_{k+c} + \lambda P_{k+c-2} = (\lambda + (c-1)\mu) P_{k+c-1} .$$

Assim,

$$c\mu P_{k+c} = (\lambda + (c-1)\mu) P_{k+c-1} - \lambda P_{k+c-2}$$

$$= P_0 \left(\frac{\lambda}{\mu} \right)^{k+c-2} \left(\frac{1}{(c-2)!} \right) \left[\frac{\lambda^2}{(c-1)\mu} + \lambda - \lambda \right] .$$

$$P_{k+c} = \left(\frac{1}{c!} \right) P_0 \left(\frac{\lambda}{\mu} \right)^{k+c} . \quad (5.7)$$

Análise do estado $(k+c)$

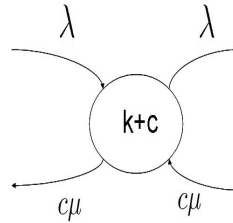


Figura 5.10: Estado $k+c$

tem-se:

$$c\mu P_{k+c+1} + \lambda P_{k+c-1} = (\lambda + c\mu) P_{k+c}$$

de onde,

$$\begin{aligned} c\mu P_{k+c+1} &= (\lambda + c\mu) P_{k+c} - \lambda P_{k+c-1} \\ &= P_0 \left(\frac{\lambda}{\mu} \right)^{k+c-1} \left(\frac{1}{(c-1)!} \right) \left(\frac{\lambda^2}{c\mu} \right) \end{aligned}$$

Assim,

$$P_{k+c+1} = \left(\frac{1}{c!} \right) \left(\frac{1}{c} \right) P_0 \left(\frac{\lambda}{\mu} \right)^{k+c+1} . \quad (5.8)$$

Análise do estado $(k + c + 1)$

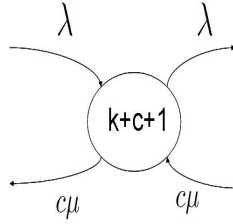


Figura 5.11: Estado $k + c + 1$

tem-se:

$$c\mu P_{k+c+2} + \lambda P_{k+c} = (\lambda + c\mu)P_{k+c+1}$$

de onde,

$$\begin{aligned} c\mu P_{k+c+2} &= (\lambda + c\mu)P_{k+c+1} - \lambda P_{k+c} \\ &= P_0 \left(\frac{\lambda}{\mu}\right)^{k+c} \left(\frac{1}{c!}\right) \left(\frac{\lambda^2}{c\mu}\right) \end{aligned}$$

Assim,

$$P_{k+c+2} = \left(\frac{1}{c!}\right) \left(\frac{1}{c}\right)^2 P_0 \left(\frac{\lambda}{\mu}\right)^{k+c+2} \quad (5.9)$$

Análise do estado (i) , $i > k + c$

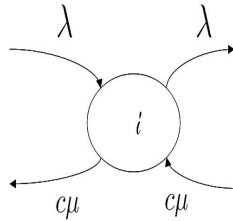


Figura 5.12: Estado i ($i > k + c$)

tem-se:

$$c\mu P_{i+1} + \lambda P_{i-1} = (\lambda + c\mu)P_i$$

$$P_i = \left(\frac{1}{c!}\right) \left(\frac{1}{c}\right)^{i-(k+c)} P_0 \left(\frac{\lambda}{\mu}\right)^i \quad (5.10)$$

Como $\sum_{i=0}^{\infty} P_i = 1$, tem-se:

$$1 = \sum_{i=0}^{k+1} P_i + \sum_{i=k+2}^{k+c-1} P_i + \sum_{i=k+c}^{\infty} P_i . \quad (5.11)$$

Chamando $\rho = \frac{\lambda}{\mu}$, tem-se:

$$\sum_{i=0}^{k+1} P_i = P_0 \sum_{i=0}^{k+1} \rho^i = \begin{cases} P_0 \frac{1 - \rho^{k+2}}{1 - \rho}, & \text{se } \rho \neq 1 \\ P_0(k+2), & \text{se } \rho = 1 \end{cases} .$$

$$\sum_{i=k+2}^{k+c-1} P_i = P_0 \sum_{i=k+2}^{k+c-1} \frac{\rho^i}{(i-k)!}$$

e

$$\sum_{i=k+c}^{\infty} P_i = P_0 \sum_{i=k+c}^{\infty} \frac{\rho^i}{c!} \left(\frac{1}{c}\right)^{i-(k+c)} = \left(\frac{c^{k+c}}{c!}\right) P_0 \sum_{i=k+c}^{\infty} \left(\frac{\rho}{c}\right)^i \quad (5.12)$$

Se $\frac{\rho}{c} < 1$, tem-se:

$$\sum_{i=k+c}^{\infty} \left(\frac{\rho}{c}\right)^i = \frac{c}{c-\rho} \left(\frac{\rho}{c}\right)^{k+c} \quad (5.13)$$

Substituindo (5.13) em (5.12),

$$\sum_{i=k+c}^{\infty} P_i = P_0 \frac{\rho^{k+c}}{(c-1)!(c-\rho)} . \quad (5.14)$$

Assim, para $\rho \neq 1$, tem-se:

$$1 = P_0 \left(\frac{1 - \rho^{k+2}}{1 - \rho} + \sum_{i=k+2}^{k+c-1} \frac{\rho^i}{(i-k)!} + \frac{\rho^{k+c}}{(c-1)!(c-\rho)} \right)$$

de onde,

$$P_0 = \left(\frac{1 - \rho^{k+2}}{1 - \rho} + \sum_{i=k+2}^{k+c-1} \frac{\rho^i}{(i-k)!} + \frac{\rho^{k+c}}{(c-1)!(c-\rho)} \right)^{-1} \quad (5.15)$$

E para $\rho = 1$, tem-se:

$$P_0 = \left(k + 2 + \sum_{i=k+2}^{k+c-1} \frac{i}{(i-k)!} + \frac{1}{(c-1)(c-1)!} \right)^{-1} \quad (5.16)$$

Na tabela(5.1) a seguir é apresentado um resumo dos resultados obtidos quanto as probabilidades dos estados para este modelo.

Tabela 5.1: Resumo das Probabilidades de Estado

Estado(i)	Probabilidade(P_i)
$0 \leq i < k + 2$	$P_i = P_0 \left(\frac{\lambda}{\mu} \right)^i$
$k + 2 \leq i < k + c$	$P_i = P_0 \frac{\rho^i}{(i-k)!}$
$i \geq k + c$	$P_i = P_0 \left(\frac{\rho^i}{c!} \right) \left(\frac{1}{c} \right)^{i-(k+c)}$
$\rho \neq 1$	$P_0 = \left(\frac{1 - \rho^{k+2}}{1 - \rho} + \sum_{i=k+2}^{k+c-1} \frac{\rho^i}{(i-k)!} + \frac{\rho^{k+c}}{(c-1)!(c-\rho)} \right)^{-1}$
$\rho = 1$	$P_0 = \left(k + 2 + \sum_{i=k+2}^{k+c-1} \frac{i}{(i-k)!} + \frac{1}{(c-1)(c-1)!} \right)^{-1}$

Obtidas as Probabilidades de Estado para o caso geral, a seguir é apresentada a obtenção do número médio de clientes no sistema(L).

5.1 Medidas de Desempenho

Número Médio de Elementos no Sistema(L)

$$L = \sum_{i=0}^{\infty} iP_i = \sum_{i=0}^{k+1} iP_i + \sum_{i=k+2}^{k+c-1} iP_i + \sum_{i=k+c}^{\infty} iP_i \quad (5.17)$$

Para $\rho \neq 1$ e usando o mesmo raciocínio que para o caso particular ($c(k) \leq 2$), tem-se:

$$\sum_{i=0}^{k+1} iP_i = P_0 \sum_{i=0}^{k+1} i\rho^i = P_0 \left(\frac{\rho(1 + \rho^{k+1}(k\rho + \rho - k - 2))}{(1 - \rho)^2} \right) \quad (5.18)$$

$$\sum_{i=k+2}^{k+c-1} iP_i = P_0 \sum_{i=k+2}^{k+c-1} \frac{i\rho^i}{(i-k)!} \quad (5.19)$$

$$\sum_{i=k+c}^{\infty} iP_i = P_0 \left(\frac{c^{k+c}}{c!} \right) \sum_{i=k+c}^{\infty} i \left(\frac{\rho}{c} \right)^i \quad (5.20)$$

$$= P_0 \left(\frac{c^{k+c}}{c!} \right) \left(\sum_{i=0}^{\infty} i \left(\frac{\rho}{c} \right)^i - \sum_{i=0}^{k+c-1} i \left(\frac{\rho}{c} \right)^i \right) =$$

$$= P_0 \left(\frac{\rho^{k+c}}{(c-1)!} \right) \left[\frac{\rho}{(c-\rho)^2} + \frac{k+c}{c-\rho} \right] \quad (5.21)$$

Assim,

$$L = P_0 \left[\frac{\rho(1 + \rho^{k+1}(k\rho + \rho - k - 2))}{(1 - \rho)^2} + \sum_{i=k+2}^{k+c-1} \frac{i\rho^i}{(i-k)!} + \left(\frac{\rho^{k+c}}{(c-1)!} \right) \left(\frac{\rho}{(c-\rho)^2} + \frac{k+c}{c-\rho} \right) \right] \quad (5.22)$$

E para $\rho = 1$, tem-se:

$$\sum_{i=0}^{k+1} iP_i = P_0 \left(\frac{(k+2)(k+1)}{2} \right)$$

$$\sum_{i=k+2}^{k+c-1} iP_i = P_0 \sum_{i=k+2}^{k+c-1} \frac{i}{(i-k)!}$$

$$\sum_{i=k+c}^{\infty} iP_i = P_0 \left(\frac{1}{(c-1)!} \right) \left(\frac{1}{(c-1)^2} + \frac{k+c}{c-1} \right)$$

Assim,

$$L = P_0 \left[\frac{(k+2)(k+1)}{2} + \sum_{i=k+2}^{k+c-1} \frac{i}{(i-k)!} + \left(\frac{1}{(c-1)!} \right) \left(\frac{1}{(c-1)^2} + \frac{k+c}{c-1} \right) \right] \quad (5.23)$$

A partir das fórmulas de Little (2.3),(2.4),(2.5) e (2.6), as demais medidas de desempenho podem ser obtidas.

Probabilidade de Ócio de um Servidor Qualquer(P_{oc})

Seja um sistema que opera com um número variável de servidores em relação ao número médio de clientes no sistema, conforme a seguir:

Clientes no Sistema(i)	Servidores em Operação(s)
$i \leq k + 1$	1
$k + 2$	2
$k + 3$	3
\vdots	\vdots
$k + c - 1$	$c - 1$
$i \geq k + c$	c

onde c é o número máximo de servidores disponíveis para atendimento e k é o número máximo de clientes na fila, para que um servidor qualquer esteja ocioso deve existir no sistema um número de clientes $i < k + c$. Satisfeita esta condição, sendo P_{oc} a probabilidade de ócio de um servidor qualquer, tem-se:

$$P_{oc} = \sum_{i=0}^{k+c-1} P(\text{servidor estar ocioso})P_i$$

Quando o sistema está vazio($i = 0$), todos os servidores estão ociosos, então

$$P(\text{servidor estar ocioso}) = \frac{c - 0}{c} = 1, \quad P_i = P_0 \quad .$$

Assim,

$$P_{oc} = P_0$$

Quando há um cliente no sistema($i = 1$), apenas um servidor dos c disponíveis está em operação. Então,

$$P(\text{servidor estar ocioso}) = \frac{c - 1}{c}, \quad P_i = P_1 \quad .$$

Assim,

$$P_{oc} = \frac{c - 1}{c}P_1$$

Quando s servidores estão em operação e o sistema está em estado entre 1 e $k + 1$, um servidor está em operação ($s = 1$) e a probabilidade de ócio de servidor é:

$$P_{oc} = \left(\frac{c-1}{c}\right) \sum_{i=1}^{k+1} P_i .$$

Quando $s = 2$, $i = k + 2$, tem-se:

$$P_{oc} = \left(\frac{c-2}{c}\right) P_{k+2}$$

Quando $s = 3$, $i = k + 3$, tem-se:

$$P_{oc} = \left(\frac{c-3}{c}\right) P_{k+3}$$

Assim, para o intervalo $0 \leq i < k + c$, tem-se:

$$\begin{aligned} P_{oc} &= P_0 + \left(\frac{c-1}{c}\right) \sum_{i=1}^{k+1} P_i + \left(\frac{c-2}{c}\right) P_{k+2} + \left(\frac{c-3}{c}\right) P_{k+3} + \cdots + \left(\frac{1}{c}\right) P_{k+c-1} = \\ &= P_0 + \frac{1}{c} \left[(c-1) \sum_{i=1}^{k+1} P_i + (c-2)P_{k+2} + (c-3)P_{k+3} + \cdots + (c-c+1)P_{k+c-1} \right] = \\ &= P_0 + \frac{1}{c} \left[(c-1) \sum_{i=1}^{k+1} P_i + \sum_{s=2}^{c-1} (c-s)P_{k+s} \right] = \\ &= P_0 + \frac{1}{c} \left[(c-1) \sum_{i=1}^{k+1} P_i + \sum_{s=2}^{c-1} (c-s)P_0 \frac{\rho^{k+s}}{s!} \right] \end{aligned}$$

Para $\rho \neq 1$, tem-se:

$$\begin{aligned} P_{oc} &= P_0 + \frac{1}{c} \left[(c-1)P_0 \frac{\rho - \rho^{k+2}}{1 - \rho} + \sum_{s=2}^{c-1} (c-s)P_0 \frac{\rho^{k+s}}{s!} \right] = \\ &= P_0 \left[1 + \frac{1}{c} \left((c-1) \left(\frac{\rho - \rho^{k+2}}{1 - \rho} \right) + \rho^k \sum_{s=2}^{c-1} (c-s) \frac{\rho^s}{s!} \right) \right] . \end{aligned} \quad (5.24)$$

E para $\rho = 1$, tem-se:

$$\sum_{i=1}^{k+1} P_i = P_0(k+1)$$

de onde:

$$P_o = P_0 \left[1 + \frac{1}{c} \left((c-1)(k+1) + \rho^k \sum_{s=2}^{c-1} (c-s) \frac{\rho^s}{s!} \right) \right] . \quad (5.25)$$

5.2 Considerações Finais

Após a caracterização do modelo proposto $M/M/c(k) \leq c/\infty/FIFO$ e a obtenção das medidas de desempenho, o capítulo seguinte apresenta, por meio de um estudo de caso aplicado ao sistema de pagamento de uma loja de departamentos, uma comparação dos resultados obtidos segundo os modelos tradicional e o proposto nesta dissertação.

Capítulo 6

Estudo de Caso:

Sistema de Pagamento de uma Loja de Departamentos

Nos capítulos anteriores foram apresentados dois modelos teóricos de filas, quais sejam, o modelo tradicional com número fixo de servidores e o modelo proposto com número de servidores em função do número de clientes no sistema, cujas medidas de desempenho permitem avaliar a funcionalidade do sistema.

Neste capítulo é apresentado um estudo de caso aplicado a um Sistema de Pagamento de Mercadorias de uma Loja de Departamentos, cujos dados foram extraídos do trabalho¹ apresentado em Fogliatti & Mattos(2007) e que serão utilizados nos modelos apresentados nos capítulos 3 e 5 com a finalidade de comparar e comentar as medidas de desempenho obtidas.

¹Desenvolvido por Sergio Pacífico Soncim e Rafael de Bruns

6.1 Caracterização do Estudo

Numa Loja de Departamentos observou-se que o sistema de fila única para os caixas disponibilizados no atendimento ao público para pagamento comporta-se de maneira ineficiente. Devido a flutuabilidade de demanda apresentada ao longo do dia, ora tem-se ociosidade dos servidores, ora insatisfação dos clientes com a demora no atendimento. As ocorrências descritas acarretam prejuízo à empresa, seja pela ociosidade dos servidores que são pagos mesmo estando livres, seja pela perda dos clientes.

Segundo Fogliatti & Mattos(2007), os dados necessários para a aplicação dos modelos foram coletados em um período de baixa flutuabilidade de demanda, onde 11 servidores atendiam em paralelo. Sem diferenças significativas entres eles no que se refere ao número médio de clientes atendidos por minuto. Assim, considerou-se que os tempos de atendimentos eram exponencialmente distribuídos.

Realizados os teste estatísticos e validados os dados, as características do sistema sob análise são as seguintes:

- 11 servidores em paralelo, com tempos de atendimentos exponencialmente distribuídos, $\mu = 0,42$ clientes por minuto;
- os clientes ao chegarem ao setor de pagamento ingressam em uma fila simples;
- o atendimento dos clientes é feito na ordem de chegada pelos servidores, à medida que estes ficam livres para atendimento;
- não foram observadas desistências e
- as chegadas são regidas por um processo de Poisson com $\lambda = 3,014$ clientes por minutos.

A condição necessária para que o sistema entre em regime estacionário, $\frac{\lambda}{c\mu} < 1$, indica que o número mínimo de servidores para o cenário em análise é $c \geq 8$.

6.2 Medidas de Desempenho

6.2.1 Resultados obtidos usando o Modelo $M/M/c/\infty/FIFO$

Caracterizado o sistema do tipo $M/M/c/\infty/FIFO$, com $\lambda = 3,014$ clientes por minuto e $\mu = 0,42$ clientes por minuto, a Taxa de Ocupação(ρ) e as medidas de desempenho: Número Médio de Clientes no Sistema(L), Número Médio de Clientes na Fila(L_q), Tempo Médio de Permanência no Sistema e na Fila(W, W_q) para $c = 8, 9, 10$ e 11 são apresentadas na tabela(6.1) e as relações entre as medidas de desempenho e o número de servidores ativos são apresentadas na figura(6.1).

Tabela 6.1: Medidas de Desempenho

c	ρ	L	L_q	W (min)	W_q (min)
8	90%	13,21	6	4,39	2
9	80%	8,85	1,68	2,94	0,56
10	72%	7,81	0,63	2,59	0,21
11	65%	7,43	0,26	2,46	0,09

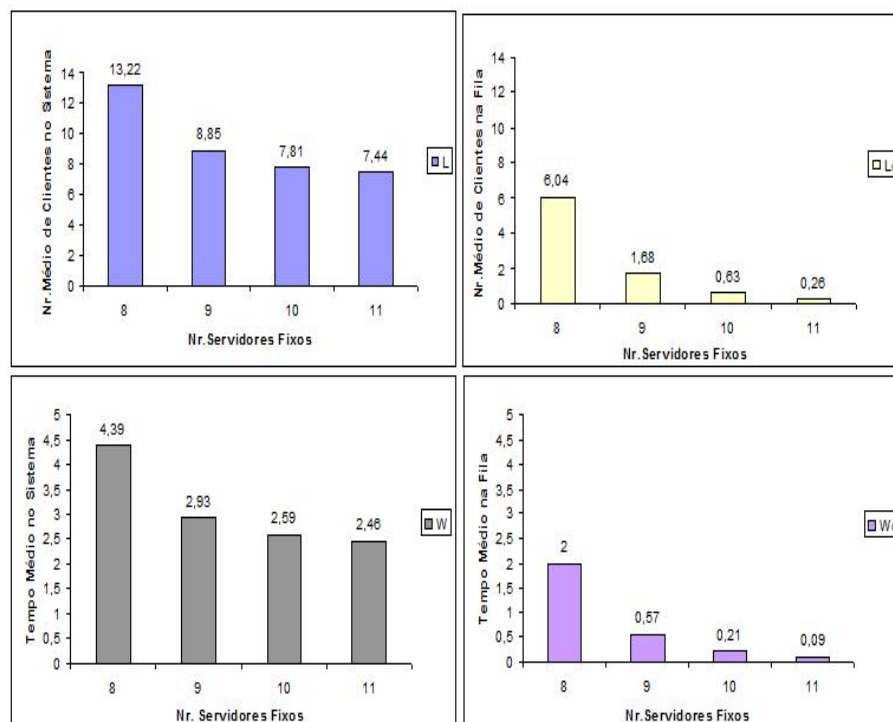


Figura 6.1: Medidas de Desempenho x Servidores Fixos

Análise dos Resultados

Comparando-se os resultados obtidos para os diferentes números de servidores, apesar da aparente condição satisfatória de funcionamento para $c = 8$, os números médios de clientes no sistema e na fila são significativamente maiores do que para o sistema trabalhando com 11 servidores.

Considerando um aumento da demanda de 10%, obtiveram-se os valores apresentados na tabela 6.2 a seguir, onde pode-se constatar que a taxa de ocupação foi elevada para 99% e os números médios de clientes passaram para 79 e 71 no sistema e na fila, respectivamente, tendo cada cliente que aguardar em média 21,5 minutos na fila para um atendimento de cerca de 3 minutos. Este aumento da demanda tornou o sistema caótico.

Tabela 6.2: Medidas de Desempenho - aumento 10% na taxa de chegadas.

c	ρ	L	L_q	$W(\text{min})$	$W_q(\text{min})$
8	99%	79	71	24	21,5

Em sentido contrário, diminuindo-se a demanda em 10%, para o sistema configurado com 11 servidores fixos, obtiveram-se os resultados apresentados na tabela(6.3), onde houve uma redução da taxa de ocupação para 59%, sendo que o número médio de clientes no sistema e o tempo médio nele não sofreram variações significativas.

Tabela 6.3: Medidas de Desempenho - redução 10% na taxa de chegadas.

c	ρ	L	L_q	$W(\text{min})$	$W_q(\text{min})$
11	59%	6,57	0,11	2,42	0,04

Observa-se então que o sistema é fortemente sensível a flutuabilidade da demanda. Uma maneira de compensar-se essa flutuabilidade é a variação do número de servidores disponibilizados para o atendimento.

Em Fogliatti & Mattos(2007) foram apresentados resultados para o problema em função dos valores da taxa média de chegadas. Entretanto, na prática, este tipo de controle torna-se inviável.

6.2.2 Resultados para o Modelo $M/M/c(k) \leq c/\infty/FIFO$

Como alternativa a situação encontrada, preservando-se as taxas de chegadas ($\lambda = 3,014$ clientes por minuto) e de atendimento ($\mu = 0,42$ clientes por minuto), aplicou-se o modelo proposto nesta dissertação, configurado com até 11 servidores e *Shift Point-k*. Para diferentes valores de k , foram obtidos os resultados apresentados na tabela(6.4), sendo que na figura(6.2) foram apresentadas as variações das medidas de desempenho em função da mudança dos *Shift Points*.

Tabela 6.4: Medidas de Desempenho x *Shift Point*.

k	5	6	7	8	9	10
L	12,44	13,44	14,44	15,44	16,44	17,44
L_q	5,26	6,26	7,26	8,26	9,26	10,26
$W(\text{min})$	4,13	4,46	4,79	5,12	5,45	5,78
$W_q(\text{min})$	1,75	2,08	2,41	2,74	3,07	3,40

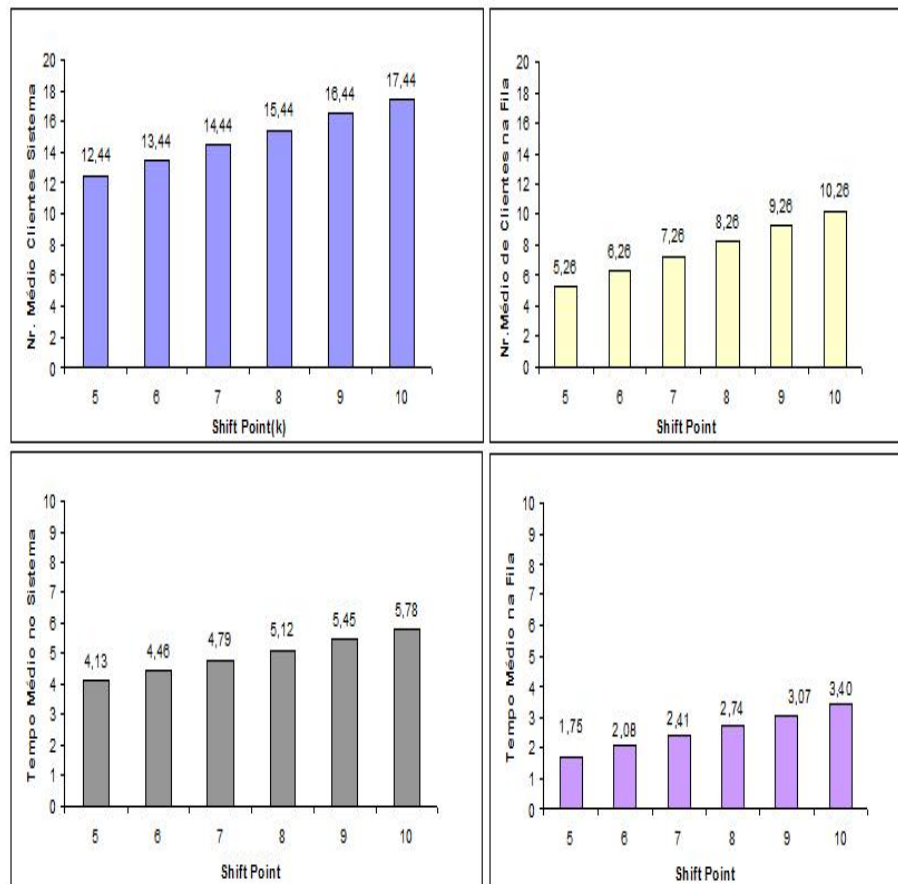


Figura 6.2: Medidas de Desempenho x *Shift Point*

Tabela 6.5: Medidas de Desempenho x Servidores Fixos e *Shift Points*.

	c				k					
	8	9	10	11	5	6	7	8	9	10
L	13,21	8,85	7,81	7,43	12,44	13,44	14,44	15,44	16,44	17,44
L_q	6	1,68	0,63	0,26	5,26	6,26	7,26	8,26	9,26	10,26
$W(\text{min})$	4,39	2,94	2,59	2,46	4,13	4,46	4,79	5,12	5,45	5,78
$W_q(\text{min})$	2	0,56	0,21	0,09	1,75	2,08	2,41	2,74	3,07	3,4

Observa-se na tabela(6.5) que os valores obtidos para $k = 6$ aproximam-se bastante do modelo anterior com 8 servidores fixos.

Considerando aumento/redução da demanda de 10%, como no caso anterior, obtiveram-se os valores apresentados nas tabelas (6.6) e (6.7) a seguir, onde pode-se constatar que o sistema apresentou menor sensibilidade às alterações da taxa de chegadas.

Tabela 6.6: Medidas de Desempenho x *Shift Point* - aumento 10% na taxa de chegadas.

k	5	6	7	8	9	10
L	13,48	14,48	15,48	16,48	17,48	18,48
L_q	5,58	6,58	7,58	8,58	9,58	10,58
$W(\text{min})$	4,06	4,37	4,67	4,97	5,27	5,57
$W_q(\text{min})$	1,68	1,99	2,29	2,74	2,89	3,19

Tabela 6.7: Medidas de Desempenho x *Shift Point* - redução 10% na taxa de chegadas.

k	5	6	7	8	9	10
L	11,57	12,57	13,57	14,57	15,57	16,57
L_q	5,11	6,11	7,11	8,11	9,11	10,11
$W(\text{min})$	4,26	4,63	5,00	5,37	5,74	6,11
$W_q(\text{min})$	1,88	2,25	2,62	2,99	3,36	3,73

Considerando-se que o desafio da administração do sistema é equilibrar o sistema às diferentes situações de congestionamento, foram obtidas as medidas de desempenho para diferentes valores da taxa média de chegadas, de modo a possibilitar uma análise de sensibilidade das configurações estabelecidas. Os resultados obtidos são apresentados nas tabelas (6.8), (6.9), (6.10) e (6.11) e nos gráficos (6.3), (6.4), (6.5) e (6.6), a seguir.

Tabela 6.8: Número Médio de Clientes no Sistema

	λ						
	1,51	2,26	3,014	3,35	3,77	4,07	4,52
c=8	3,62	5,85	13,21	236,11	∞	∞	∞
c=9	3,60	5,56	8,85	12,33	306,93	∞	∞
c=10	3,59	5,45	7,81	9,54	14,72	37,28	∞
c=11	3,59	5,41	7,43	8,59	10,84	14,09	52,74
k=5	8,54	10,40	12,44	13,59	15,84	19,09	57,74
k=10	13,54	15,40	17,44	18,59	20,84	24,09	62,74
k=15	18,54	20,40	22,44	23,59	25,84	29,09	67,74
k=20	23,54	25,40	27,44	28,59	30,84	34,09	72,74
k=25	28,54	30,40	32,44	33,59	35,84	39,09	77,74

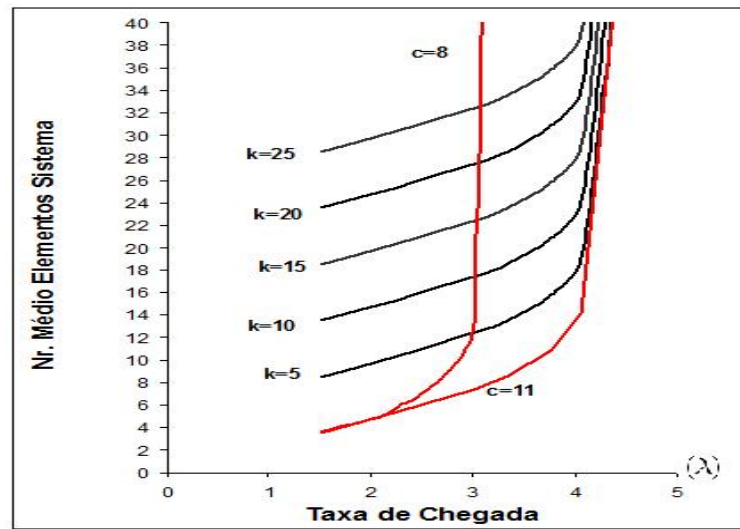


Figura 6.3: Número Médio de Clientes no Sistema.

Comparando-se os resultados do número médio de clientes no sistema para as diferentes configurações estabelecidas, constata-se que:

1. Para valores menores da taxa de chegadas o modelo tradicional operando com 8 ou 11 servidores fixos não apresenta diferenças significativas em relação ao número médio de clientes no sistema e o modelo proposto com até 11 servidores consegue manter um maior número médio de clientes no sistema, de acordo com o interesse da administração e

2. À medida que são atribuídos valores maiores para a taxa de chegadas, o modelo tradicional deixa de operar satisfatoriamente, enquanto o modelo proposto mantém relativa estabilidade até atingir comportamento idêntico ao modelo fixo com 11 servidores.

Tabela 6.9: Número Médio de Clientes na Fila.

	λ						
	1,51	2,26	3,014	3,35	3,77	4,07	4,52
$c = 8$	0,03	0,47	6,04	228,14	∞	∞	∞
$c = 9$	0,01	0,17	1,68	4,95	297,96	∞	∞
$c = 10$	0,00	0,06	0,63	1,57	5,75	27,59	∞
$c = 11$	0,00	0,02	0,26	0,63	1,87	4,40	41,98
$k = 5$	4,95	5,02	5,26	5,63	6,87	9,40	46,98
$k = 10$	9,95	10,02	10,26	10,63	11,87	14,40	51,98
$k = 15$	14,95	15,02	15,26	15,63	16,87	19,40	56,98
$k = 20$	19,95	20,02	20,26	20,63	21,87	24,40	61,98
$k = 25$	24,95	25,02	25,26	25,63	26,87	29,40	66,98

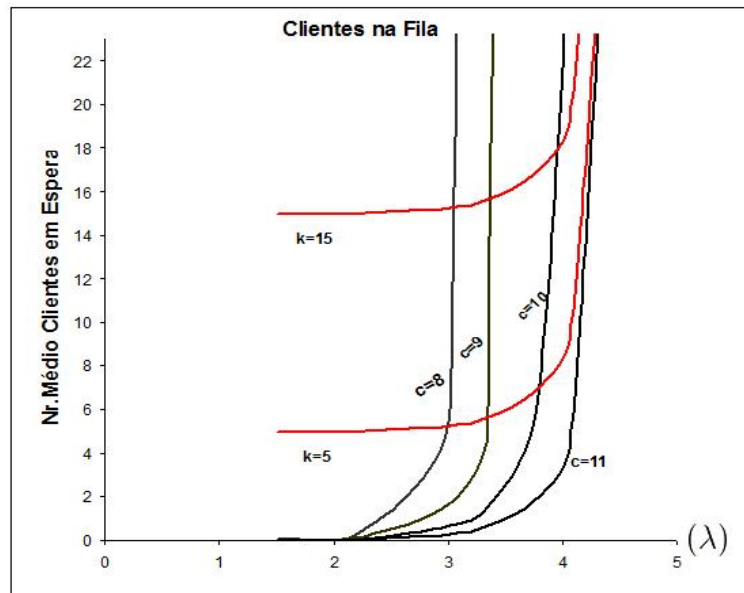


Figura 6.4: Número Médio de Clientes na Fila.

Neste caso, pode-se observar que para valores reduzidos da taxa de chegadas o modelo proposto mantém k clientes na fila como estabelecido pela administração, enquanto que para o modelo tradicional praticamente não existem clientes na fila.

Tabela 6.10: Tempo Médio no Sistema.

	λ						
	1,51	2,26	3,014	3,35	3,77	4,07	4,52
$c = 8$	2,40	2,59	4,39	70,57	∞	∞	∞
$c = 9$	2,39	2,46	2,94	3,86	81,47	∞	∞
$c = 10$	2,38	2,41	2,59	2,85	3,91	9,16	∞
$c = 11$	2,38	2,39	2,47	2,57	2,88	3,46	11,67
$k = 5$	5,66	4,60	4,13	4,06	4,21	4,69	11,67
$k = 10$	8,98	6,81	5,78	5,56	5,53	5,92	12,58
$k = 15$	12,30	9,02	7,44	7,05	6,86	7,15	13,68
$k = 20$	15,62	11,24	9,10	8,55	8,19	8,38	14,79
$k = 25$	18,94	13,45	10,76	10,04	9,51	9,61	15,90

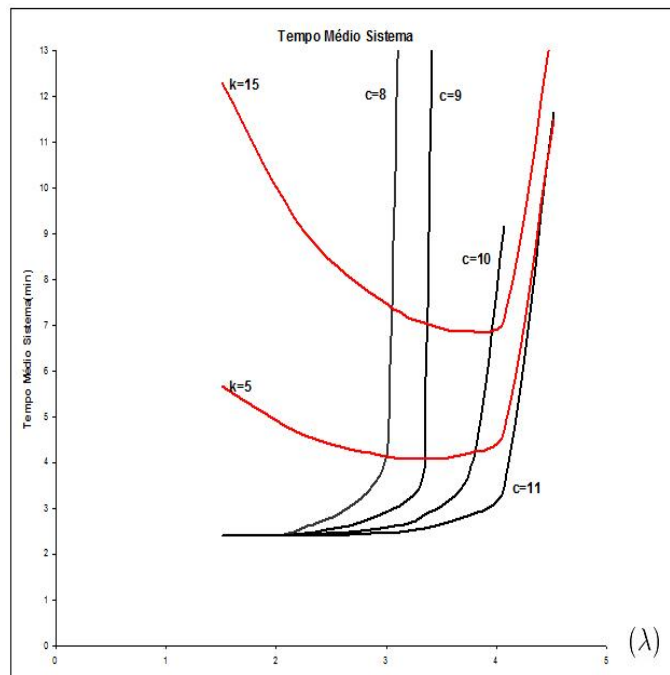


Figura 6.5: Tempo Médio no Sistema \times ρ .

Os tempos médios de espera no sistema apresentam inicialmente valores mais altos para o modelo proposto nesta dissertação e vão diminuindo à medida que novos servidores são agregados.

Para demandas maiores da taxa de chegadas, o modelo tende a comportar-se de modo semelhante ao modelo tradicional.

Tabela 6.11: Tempo Médio na Fila.

	λ						
	1,51	2,26	3,014	3,35	3,77	4,07	4,52
$c = 8$	0,02	0,21	2,00	68,19	∞	∞	∞
$c = 9$	0,01	0,08	0,56	1,48	79,09	∞	∞
$c = 10$	0,00	0,03	0,21	0,47	1,53	6,78	∞
$c = 11$	0,00	0,01	0,09	0,19	0,50	1,08	9,29
$k = 5$	3,28	2,22	1,75	1,68	1,82	2,31	10,39
$k = 10$	6,60	4,43	3,40	3,18	3,15	3,54	11,50
$k = 15$	9,92	6,64	5,06	4,67	4,48	4,77	12,60
$k = 20$	13,24	8,85	6,72	6,17	5,81	6,00	13,71
$k = 25$	16,55	11,07	8,38	7,66	7,13	7,23	14,81

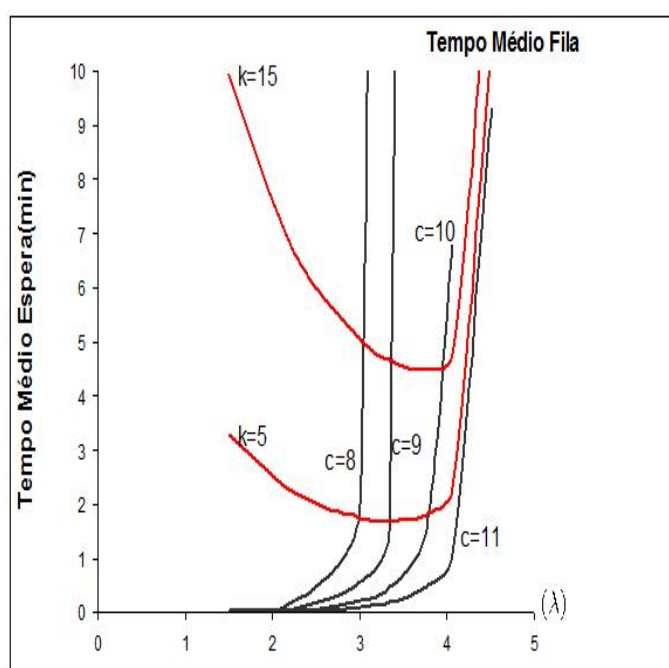


Figura 6.6: Tempo Médio na Fila.

de modo análogo, os tempos médios de espera na fila apresentam inicialmente valores mais altos para o modelo proposto nesta dissertação e vão diminuindo à medida que novos servidores são agregados.

Para demandas maiores da taxa de chegadas, o modelo tende a comportar-se de modo semelhante ao modelo tradicional.

Os resultados obtidos permitem afirmar que o modelo proposto apresenta menor sensibilidade a flutuabilidade de demanda e maior nível de estabilidade quando comparado ao modelo tradicional .

Capítulo 7

Conclusão

Neste trabalho apresentou-se o desenvolvimento de um modelo analítico, baseado em Teoria de Filas, para sistemas que apresentam diferentes situações de congestionamento e onde o número de servidores em paralelo depende do número de elementos no sistema ou na fila, isto é, quando a fila atingir 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 a fila decresce, os servidores são retirados de operação na mesma ordem que ficarem ociosos.

Para avaliação do modelo, foi realizada uma aplicação no estudo de caso de uma Loja de Departamentos, que possibilitou mostrar que o modelo proposto pode proporcionar um atendimento eficiente ao cliente dando flexibilidade a administração do sistema, permitindo a realocação de servidores que não estejam envolvidos com atendimento a outras atividades de interesse da administração.

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 envolvidos e qualidade de serviço.

As medidas de desempenho dos modelos apresentados foram obtidas por meio da utilização das fórmulas desenvolvidas e apresentadas ao longo do trabalho. Para a realização dos cálculos foram utilizadas planilhas eletrônicas confeccionadas no software Microsoft Excel 2003.

Durante o desenvolvimento do trabalho observou-se a falta de bibliografia disponível sobre Teoria de Filas em língua portuguesa, destacando-se os trabalhos de Novaes(1975) e Fogliatti & Mattos(2007).

Uma modelagem analítica para sistemas com *Shift Points* de entrada de servidores maiores que um foi iniciada ao longo deste trabalho, isto é, a inserção de novos servidores em pontos distintos(k, k_0, k_1, \dots) com ($k \neq k_1 \neq k_2$). Contudo, face o escopo inicial estabelecido e disponibilidade de tempo para o desenvolvimento, optou-se por não prosseguir em tal desenvolvimento.

Como proposta para desenvolvimento de trabalhos futuros, sugere-se a modelagem de sistemas com mecanismos de troca de servidores(*Shift Points*) distintos para entrada e retirada de novos servidores.

Referências Bibliográficas

ALVES, R.A.F.S; DELGADO, C.J.M. *Processos Estocásticos*. Faculdade de Economia, FEP, 1997. Disponível em [http : //www.de.ufpe.br/ ~ rjdsc/teaching/2005.2/et301/handouts/](http://www.de.ufpe.br/~rjdsc/teaching/2005.2/et301/handouts/), acesso em 25 de abril de 2007.

AL-SEEDY,R.O. Queuing with Fixed and Variable Channels Considering Balking and Reneging Concepts. *Applied Mathematics and Computation*, Vol.156, pp.755-761, 2004.

AL-SEEDY,R.O. A Transient Solution of Queues with Variable Channel Considering Balking Concept When $S = 2$ and $\sigma = 1$. *Applied Mathematics and Computation*, Vol.174, pp.337-344, 2006.

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.

BITRAN,G.; MORABITO,R. Open Queueing Networks: Optimization and Performance Evaluation Models for Discrete Manufacturing Systems, *Production and Operations Management*, pp.163-193, 1996.

CARNEIRO, J.P.; MOREIRA, C.G. Sequências Aritmético-Geométricas, *Eureka*, Nr.14, pp.32-34, 2002. Disponível em <http://www.obm.org.br/eureka.htm>, acesso em 25 de fevereiro de 2008.

CLARK, A.B.;DISNEY,R.L. *Probabilidade e Processos Estocásticos*. Rio de Janeiro.Livros Técnicos e Científicos Editora S.A., 1979.

COBHAM, A. "Priority Assignment in Waiting-Line Problems", *Journal of the Operations Research Society of America*,1954.

COOPER, R.B.; MURRAY G. "Queues Served in Cyclic Order", *The Bell System Technical Journal*, pp.675-689,1969.

COURTOIS, P.J. *Decomposability, Queueing and Computer System Applications*, New-York, Academic Press,1977.

DAVIS,E.W.; PHILLIPS,C.R.;TALBIRO,J.B. *An application of queuing theory in a hospital emergency Clinic*. School of industrial engineering, Georgia Institute of Technology, 1959.

De GROOT, M.H.; SCHERVISH, M. J. *Probability and Statistics* 3a edição. New York: Addison-Wesley, 2002.

EVERETT, J.L. - "Seaport Operation as a Stochastic Process", *The Journal of the Operations Research Society of America*,Nr.1.

FOGLIATTI, M.C.; MATTOS, N.M.C. *Teoria das Filas*. Rio de Janeiro. Editora Interciência, 2007.

GELENBE, E.; MUNTZ, R.R. Probabilistic Models of Computer Systems. *Acta Informatica*, Vol.7, Nr.1, pp.35-60, 1976.

GROSS, D.; HARRIS,C.M., *Fundamentals of queueing theory*. Chichester, John Wiley Sons Inc., 1985.

HERSH , M. Allocation Process With Variable Channel Queues. *Management Science*,Vol.20, 1973.

HEYMAN , D.P.; SOBEL, M.J. *Stochastic Models - Handbooks in Operations Research and Management Science* Vol.2. North-Holland. Amsterdam, 1990.

HILLIER, F.S.;LIEBERMAN,G.J. *Introdução a Pesquisa Operacional*, Editora Campus/Edusp,1988.

JENNINGS, O.B.; MANDELBAUM, A.; MASSEY, W.A.; WHITT,W. Server Staffing to Meet Time-Varying Demand. *Management Science*,Vol.42, 1996.

KENDALL, D.G. Some Problems in the Theory of Queues. *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 13, Nr.2, pp. 151-185,1951.

KENDALL, D.G. Stochastic Process Occurring in The Theory of Queues and their Analysis by Method of Imbedded Markov Chain. *The Annals of Mathematical Statistics*, Vol.24, Nr.3,pp.338-354,1953.

KINGMAN, J.F.C. *On the Algebra of Queues*. London, Methuen,1966.

KLEINROCK, L. *Queueing Systems, Vol. I: Theory*. New York, John Wiley Sons, Inc. 1975.

LI,H.; YANG,T. "Queues with a variable number of servers". *European Journal of Operational Research*, Vol.124, 2000, pp. 615-628.

LITTLE, J.D.C. "A Proof for the Queuing Formula: $L = \lambda W$ ". *Operations Research*, Vol.9, 1961, pp. 383-387.

MODER, J.J.; PHILLIPS Jr.C.R. Queuing with Fixed and Variable Channels. *Operations Research*, Vol. 10, 1962, pp. 218-231.

MORSE, P.M.. *Queues, Inventories and Maintenance*. New York, John Wiley Sons Inc., 1962.

NAIN, P. *Notas de aula da Disciplina Performance Evaluation of Computer Systems*. University of Massachusetts, Amherst, MA, 1994.

NOVAES, A.G.N; *Pesquisa Operacional e Transportes: Modelos Probabilísticos*. São Paulo. Editora da Universidade de São Paulo, 1975.

OVUWORIE, G.C. Multi-Channel Queues: A Survey and Bibliography, *International Statistical Review / Revue Internationale de Statistique*, Vol.48, Nr.1, pp. 49-71,1980.

PRABHU, N.U. *Queues and Inventories*. New York, John Wiley Sons Inc., 1965.

PAPOULIS, A. *Probability, Random Variables, and Stochastic Processes*, 2nd ed. New York: McGraw-Hill, 1984.

PERUCHA, V.T.; JUÁREZ, I.U.; MORENO, J.S.M. *Problemas Resueltos de Cálculo en Una Variable*. Thomson Learning Ibero, 2005.

RESING, J.; ADAN,I. *Queueing Theory* - Department of Mathematics and Computing Science, Eindhoven University of Technology, 2001.

ROMANI, J. Un Modelo de la Teoría de Colas con Número Variable de Canales, *Trabajos Estadística 8*, 175-189,1957.

ROSS, SHELDON M. *Introduction to Probability Models*, 5th ed. Academic Press, 1993.

ROSS, SHELDON M. *Stochastic Processes*, 2nd ed. New York: Wiley Sons,1996.

SILVA,G.B.;SANTOS,M.B.J.;SINAY,M.C.F., Processos Markovianos Intermediários: O Modelo $M/M/c(n)/\infty/FIFO$. *Simpósio de Pesquisa Operacional e Logística da Marinha*, 2007.

TANNER, J.C. "The Delay to Pedestrians Crossing a Road", *Biometrika* v.38, partes 3 e 4, pp. 383-392, 1951.

Apêndice A

Este Apêndice contém conceitos básicos referentes a Processos Estocásticos e a definição de Série Aritmético-Geométrica útil para determinar algumas medidas de desempenho(vide Capítulo 4).

PROCESSOS ESTOCÁSTICOS

Segundo Hillier & Lieberman(1988), um processo estocástico é a abstração matemática de um processo empírico cujo desenvolvimento é governado por leis probabilísticas. Genericamente, é um fenômeno que varia em algum grau, de maneira aleatória, isto é, não é possível saber que evento irá acontecer no futuro, mesmo sob condições presumidamente idênticas, à medida que o tempo passa, resultando em diferentes seqüências observadas. Por exemplo, no lançamento de uma moeda não há como saber qual lado ficará à vista quando a moeda chegar ao solo, mesmo após repetidos lançamentos. O nível de estoque de um artigo no fim de um dia, o número de máquinas avariadas em determinado instante, a cotação de uma ação ao fim do dia, a variação de tráfego em um certo cruzamento com a formação e dissipação de congestionamento, o número de usuários em uma fila de banco em determinado instante, o índice pluviométrico em cada dia do mês, o número de dias que choveram em cada mês do ano, entre outros, são exemplos de processos estocásticos.

Um processo estocástico pode ser definido como um conjunto de variáveis aleatórias, $\mathbf{E} = \{\mathbf{X}(\mathbf{t}), \mathbf{t} \in \mathbf{T}\}$, que descreve o comportamento de um processo em análise durante um período de tempo \mathbf{t} .

- \mathbf{T} → conjunto de valores assumidos pela variável \mathbf{t} .
- $\mathbf{X}(\mathbf{t})$ → estado do processo no momento \mathbf{t} .
- \mathbf{E} → conjunto de valores possíveis de $\mathbf{X}(\mathbf{t})$.

Os conjuntos \mathbf{T} e \mathbf{E} podem assumir valores discretos ou contínuos, produzindo diferentes tipos de processos estocásticos.

Se \mathbf{E} é um conjunto:

- Discreto(\mathbf{D}), Enumerável ou finito - o processo estocástico é denominado *Cadeia Estocástica*.
- Contínuo(\mathbf{C}) - o processo estocástico é denominado *Processo Contínuo*.

Se \mathbf{T} é um conjunto:

- Discreto(\mathbf{D}), Enumerável ou finito - o processo estocástico é denominado *Processo de Parâmetro Discreto*.
- Contínuo(\mathbf{C}) - o processo estocástico é denominado *Processo de Parâmetro Contínuo*.

Um processo estocástico é dito estacionário se o seu comportamento dinâmico é invariante no tempo, ou seja, se $X(t_1 + s) - X(t_1)$ tem a mesma distribuição de $X(t_2 + s) - X(t_2)$, para todo valor de $t \in T$.

Dentre os vários processos estocásticos possíveis de acontecerem, a seguir serão apresentados os Processos Markovianos, os Processos de Nascimento-Morte e o Processo de Poisson que são mencionados ao longo desta dissertação.

PROCESSOS MARKOVIANOS

Segundo Fogliatti & Mattos(2007), os Processos Markovianos são um tipo especial de processo estocástico, cuja seqüência de tempos ($t_0 < t_1 < \dots < t_n < t$) tem a distribuição de probabilidade condicional de $X(t)$ para dados valores de $X(t_0), X(t_1), \dots, X(t_n)$ dependente unicamente de $X(t_n)$, ou seja:

$$P[X(t) = x \mid X(t_n) = x_n, X(t_{n-1}) = x_{n-1}, \dots, X(t_0) = x_0] = P[X(t) = x \mid X(t_n) = x_n] \quad (\text{A.1})$$

A principal característica de um Processo Markoviano é a ausência de memória(*Memoryless*). Isto é, se o valor presente é conhecido, o conhecimento futuro não é alterado por informação adicional sobre o passado. O comportamento futuro é condicionado apenas pelo estado presente. Tem-se como consequência, que num processo Markoviano todos os tempos entre acontecimentos sucessivos têm de ser exponencialmente distribuídos.

Se a variável aleatória $X(t)$ está definida em um espaço de estados discretos, o Processo Markoviano é denominado Cadeia de Markov, que podem ser de parâmetro discreto ou contínuo.

Uma Cadeia de Markov de parâmetro discreto é definida por:

- um espaço de estados (E) finito ou enumerável; e
- probabilidades de transição entre os estados em um período.

Denomina-se probabilidade de transição($P_{ij}^{(n)}$), a probabilidade de transição que um processo no estágio i estará no estágio j após n passos. Define-se a probabilidade de transição em n passos como:

$$P_{ij}^{(n)} = P\{X(t+n) = j \mid X(t) = i\} = P\{X(n) = j \mid X(0) = i\}, \quad \forall t = 0, 1, 2, \dots \text{ e } i, j \in X \quad (\text{A.2})$$

As cadeias de Markov podem ser representadas:

1. Graficamente por *Diagramas de Fluxo*, como na figura(A.1), onde os *nós* representam os estados e os *arcos* representam as transições entre esses estados no tempo; e

2. Por matriz de transição. $P = \begin{bmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{bmatrix}$

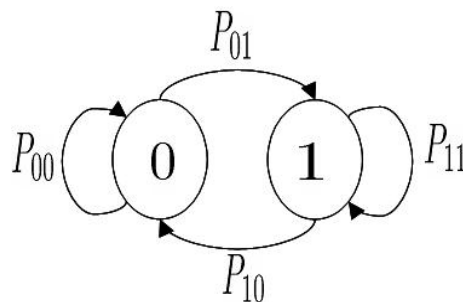


Figura A.1: Diagrama de Fluxo de uma Cadeia de Markov para 2 estados

Na matriz P , cada linha i representa o estado atual e cada coluna j representa o estado futuro (a ordem dos estados atuais deve ser igual à dos estados futuros, respectivamente, nas linhas e nas colunas de P). Deste modo, o elemento P_{ij} da matriz representa a probabilidade de ocorrência da transição do estado i para o estado j em um período.

Na matriz de transição P , seus elementos são probabilidades, assim, $P_{ij} \geq 0$ e $\sum P_{ij} = 1$.

Classificação dos estados

Um caminho do estado i para o estado j é uma seqüência de transições com probabilidades de ocorrência positivas que ligam os dois estados. Não existe qualquer limite para o número de transições nem há a obrigação de a seqüência ser a mais curta entre os dois estados.

O estado j é dito atingível a partir do estado i , se e somente se, houver pelo menos um caminho de i para j .

Dois estados i e j são *comunicantes*, se e somente se, j for atingível a partir de i e i for atingível a partir de j .

Uma Cadeia de Markov é dita *irredutível*, se e somente se, qualquer estado j for atingível a partir de qualquer estado i , ou seja, se todos os estados forem comunicantes.

Um estado j é dito *absorvente* se após o processo entrar em j não sair de j , isto é, se $P_{jj} = 1$.

Probabilidades Estacionárias

Para algumas Cadeias de Markov, a probabilidade do processo estar no estado j é *constante e independente* do estado inicial i . Quer dizer que, espera-se que, ao longo de um grande período de tempo, a influência do estado inicial diminua de tal forma que as probabilidades limites não dependem desse estado.

A esta probabilidade, constante e independente, chama-se *probabilidade estacionária do estado j* e representada por π_j .

$$P_{ij}^{(n+1)} \approx P_{ij}^{(n)} \approx \pi_j, \quad n \rightarrow \infty \quad (\text{A.3})$$

Para que isto aconteça, a cadeia de Markov deve ser irredutível e ergódica (recorrente e aperiódica), isto é, qualquer estado j é atingível a partir de qualquer estado i , partindo-se de qualquer estado j sempre é possível regressar a j .

PROCESSO DE NASCIMENTO-MORTE

Segundo Fogliatti & Mattos(2007), os Processos de Nascimento-Morte são processos estocásticos, cadeias de Markov de parâmetro contínuo, cujas transições de um estado qualquer somente são possíveis para estados vizinhos, ou seja, de um estado i para estados $i + 1$ ou $i - 1$.

A maioria dos modelos de filas elementares supõe que as chegadas e as saídas de

clientes ocorrem segundo um Processo de Nascimento-Morte, ou seja, a chegada de um cliente representa o nascimento para o sistema e a saída do cliente representa a morte.

Estes modelos têm a representação gráfica da figura(A.2) a seguir.

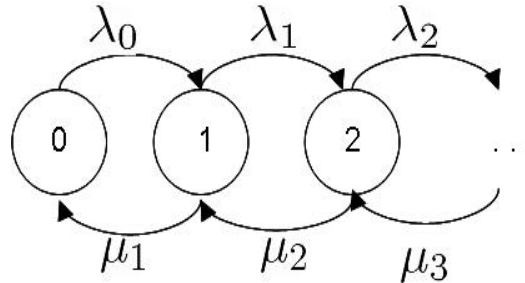


Figura A.2: Diagrama de Fluxo de um Processo de Nascimento e Morte

Hipóteses básicas consideradas no estudo dos Processos de Nascimento-Morte:

1. No instante inicial(t_0), o sistema está vazio;
2. Nascimentos e Mortes são eventos estatisticamente independentes;
3. Para um sistema no estado i , no intervalo de tempo $(t, t + \Delta t)$, Δt tão pequeno quanto se queira, a probabilidade de ocorrer:
 - a. $i + 1$, um nascimento, é $\lambda_i \Delta t + o(\Delta t)$;
 - b. $i - 1$, uma morte, é $\mu_i \Delta t + o(\Delta t)$;
 - c. $i + 1$ e $i - 1$, mais de um evento, é $o(\Delta t)$;

Definindo-se:

$P_i^n(\Delta t)$ = Probabilidade de ocorrência de i nascimentos em Δt ; e

$P_j^m(\Delta t)$ = Probabilidade de ocorrência de j mortes em t ,

tem-se,

$$P_1^n(\Delta t) = \lambda_i \Delta t + o(\Delta t), \quad \forall i = 0, 1, 2, 3, \dots;$$

$$P_1^m(\Delta t) = \mu_i \Delta t + o(\Delta t), \quad \forall i = 0, 1, 2, 3, \dots;$$

$$P_1^n(\Delta t)P_1^m(\Delta t) = o(\Delta t), \quad \forall i, j = 0, 1, 2, 3, \dots;$$

Sabendo-se que a probabilidade de não ocorrência de A é um menos a probabilidade de ocorrer A, obtém-se a probabilidade de não ocorrer nascimento ou morte em intervalos pequenos de tempo, Δt :

$$P_0^n(\Delta t) = \text{Probabilidade de não ocorrência de nascimentos em } \Delta t \text{ é } 1 - P_1^n(\Delta t),$$

$$P_0^n(\Delta t) = 1 - P_1^n(\Delta t) = 1 - \lambda_i \Delta t - O(\Delta t), \quad \forall i = 0, 1, 2, 3, \dots; \text{ e}$$

$$P_0^m(\Delta t) = \text{Probabilidade de não ocorrência de } i \text{ mortes em } \Delta t \text{ é } 1 - P_1^m(\Delta t),$$

$$P_0^m(\Delta t) = 1 - P_1^m(\Delta t) = 1 - i \Delta t - o(\Delta t), \quad \forall i = 0, 1, 2, 3, \dots;$$

Das considerações anteriores, pode-se determinar, para todo i , as probabilidades P_i dos estados do processo.

Dividindo-se um intervalo de observação em dois sub-intervalos disjuntos $(0, t]$ e $(t, t + \Delta t)$. Para o estágio inicial vazio e independência entre nascimentos e mortes, pode-se calcular a probabilidade de um sistema estar no estado i no instante $t + \Delta t$:

Para $i > 0$, no instante $t + \Delta t$ e com $\Delta t \rightarrow 0$:

Tabela A.1: Eventos possíveis no instante $t + \Delta t (i > 0)$

t	$t + \Delta t$	Nascimento	Morte
i	i	0	0
$i - 1$	i	1	0
$i + 1$	i	0	1

$$P_i(t + \Delta t) = P_i(t)P_0^n(\Delta t)P_0^m(\Delta t) + P_{i-1}(t)P_1^n(\Delta t)P_0^m(\Delta t) + P_{i+1}(t)P_0^n(\Delta t)P_1^m(\Delta t) + o(\Delta t) \quad (\text{A.4})$$

Para $i = 0$, no instante $t + \Delta t$ e com $\Delta t \rightarrow 0$:

Tabela A.2: Eventos possíveis no instante $t + \Delta t (i = 0)$

t	$t + \Delta t$	Nascimento	Morte
0	0	0	–
1	0	0	1

$$P_0(t + \Delta t) = P_1(t)P_0^n(\Delta t)P_1^m(\Delta t) + P_0(t)P_0^n(\Delta t) + o(\Delta t) \quad (\text{A.5})$$

Como os processos de nascimento-morte são cadeias de Markov de parâmetro contínuo, existe um momento t^* a partir do qual o processo entra em regime estacionário, com características estáveis. Desenvolvendo-se as equações A.4 e A.5, detalhadas em Fogliatti & Mattos(2007), chega-se em:

$$0 = -\lambda_i P_i - \mu_i P_i + \lambda_{i-1} P_{i-1} + \mu_{i+1} P_{i+1}, \quad \forall i \geq 1 \quad (\text{A.6})$$

$$0 = -\lambda_0 P_0 + \mu_1 P_1 \quad (\text{A.7})$$

As equações (A.6) e (A.7) são denominadas equações de balanço e podem ser obtidas também pelo do diagrama de fluxo (Figura 5) aplicando-se o princípio da conservação de energia, em cada nó (estado), o fluxo que entra em um nó é igual ao fluxo que sai.

Das equações de balanço, obtém-se:

$$P_i = P_0 \prod \left(\frac{\lambda_{n-1}}{\mu_n} \right), \quad i \geq 1 \text{ e } n = 1, 2, 3, \dots, i \quad (\text{A.8})$$

Como $\sum P_i = 1$, tem-se:

$$P_0 = \frac{1}{1 + \prod \left(\frac{\lambda_{n-1}}{\mu_n} \right)}, \quad i \geq 1 \text{ e } n = 1, 2, 3, \dots, i \quad (\text{A.9})$$

PROCESSO DE POISSON

Um processo de nascimento-morte é dito ser um processo de nascimento puro se a taxa de atendimento ou serviço (μ_i) é nula para todo i .

Segundo Ross(1996), o Processo de Poisson é o exemplo mais simples de um processo de nascimento puro, pois possui taxa de chegada constante $\lambda_i = \lambda$.

Sendo este processo uma cadeia de Markov de parâmetro contínuo, onde a única transição possível a partir de um estado i qualquer é para o estado $i + 1$, a modelagem do processo pode ser feito como para um processo de nascimento-morte com $\lambda_i = \lambda$ e $\mu_i = 0, \forall i$. O diagrama de fluxo correspondente é apresentado na figura(??)a seguir.

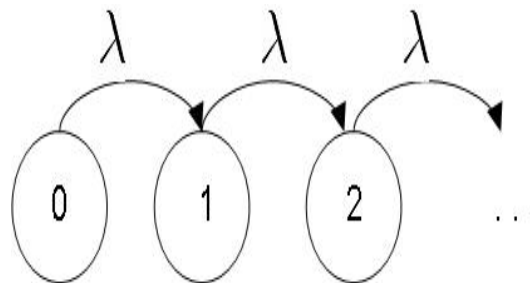


Figura A.3: Diagrama de Fluxo de um Processo de Poisson com taxa λ

As hipóteses associadas aos Processos de Poisson são:

1. No instante inicial(t_0), o sistema está vazio;
2. Nascimentos são eventos estatisticamente independentes;
3. Para um sistema no estado i , no intervalo de tempo $(t, t + \Delta t)$, Δt tão pequeno quanto se queira, a probabilidade de ocorrer:
 - a. $i + 1$, um nascimento, é $\lambda_i \Delta t + o(\Delta t)$,
 - b. mais de um nascimento, é $o(\Delta t)$.

Para $i > 0$, no instante $t + \Delta t$ e com $\Delta t \rightarrow 0$:

$$P_1^n(t + \Delta t) = \lambda \Delta t + O(\Delta t) \rightarrow \text{Probabilidade de ocorrência de 1 nascimento;}$$

$$P_0^n(t + \Delta t) = 1 - \lambda \Delta t - O(\Delta t) \rightarrow \text{Probabilidade de não ocorrência de nascimento;}$$

Tabela A.3: Eventos possíveis no instante $t + \Delta t$

t	$t + \Delta t$	Nascimento
i	i	0
$i - 1$	i	1

$P_i^n(t + \Delta t) = O(\Delta t) \rightarrow$ Probabilidade de ocorrência de i nascimentos, $i > 1$.

Das considerações acima, pode-se calcular a probabilidade de um sistema estar no estado i no instante $t + \Delta t$:

$$P_i(t + \Delta t) = P_i(t)P_0^n(\Delta t) + P_{i-1}(t)P_1^n(\Delta t) + O(\Delta t), \quad i \geq 1 \quad (\text{A.10})$$

$$P_0(t + \Delta t) = P_0(t)P_0^n(\Delta t) + O(\Delta t) \quad (\text{A.11})$$

Para $\Delta t \rightarrow 0$, fazendo-se as devidas substituições e desenvolvendo as equações A.10 e A.11, detalhadas em Fogliatti & Mattos(2007), chega-se a:

$$P_i(t) = \frac{(\lambda t)^i e^{-\lambda t}}{i!}, \quad \forall i \geq 0 \quad (\text{A.12})$$

SÉRIE ARITMÉTICO-GEOMÉTRICA

É uma seqüência(a_n) cujo termo geral satisfaz a fórmula:

$$a_n = (a + nr)q^n, \quad \forall n \in \mathbb{N} \quad (\text{A.13})$$

No caso de $a = 0$ e $r = 1$, tem-se

$$S = \sum_{i=k+2}^{\infty} \frac{i}{x^i}$$

que pode ser representado por:

$$S = \frac{k+2}{x^{k+2}} + \frac{k+3}{x^{k+3}} + \frac{k+4}{x^{k+4}} + \frac{k+5}{x^{k+5}} + \dots = S_1 + S_2$$

$$S_1 = \frac{k}{x^{k+2}} + \frac{k}{x^{k+3}} + \frac{k}{x^{k+4}} + \frac{k}{x^{k+5}} + \dots = \frac{k}{x^{k+2}} \left(\frac{1}{1 - \frac{1}{x}} \right) .$$

$$S_1 = \frac{k}{(x-1)x^{k+1}}, \text{ Progressão Geométrica.}$$

$$\begin{aligned} \text{E } S_2 &= \frac{2}{x^{k+2}} + \left(\frac{2}{x^{k+2}} + \frac{1}{x^{k+2}} \right) \frac{1}{x} + \left(\frac{2}{x^{k+2}} + \frac{2}{x^{k+2}} \right) \frac{1}{x^2} + \\ &\left(\frac{2}{x^{k+2}} + \frac{3}{x^{k+2}} \right) \frac{1}{x^3} + \dots \end{aligned}$$

$$\begin{aligned} S_2 &= a + (a+r)q + (a+2r)q^2 + (a+2r)q^2 + (a+3r)q^3 + \dots = \\ &= a + aq + aq^2 + aq^3 + \dots + rq + 2rq^2 + 3rq^3 + \dots, \text{ Progressão Aritmético-Geométrica.} \end{aligned}$$

Multiplicando S_2 por q , tem-se:

$$S_2q = aq + aq^2 + aq^3 + aq^4 + \dots + rq^2 + 2rq^3 + 3rq^4 + \dots$$

Subtraindo S_2q de S_2 , tem-se:

$$S_2(1-q) = a + (rq + rq^2 + rq^3 + rq^4 + \dots) \quad (\text{A.14})$$

A parcela $S^* = (rq + rq^2 + rq^3 + rq^4 + \dots) = a + S^*$

S^* é uma Progressão Geométrica infinita de termo inicial rq e razão q , portanto

$$S^* = rq \left(\frac{1}{1-q} \right) \quad (\text{A.15})$$

Substituindo (A.15) em (A.15), tem-se:

$$S_2(1 - q) = a + rq \left(\frac{1}{1 - q} \right) = \frac{a(1 - q) + rq}{1 - q} .$$

$$S_2 = \frac{a(1 - q) + rq}{(1 - q)^2} \quad (\text{A.16})$$

Substituindo a , r , e q por seus valores, tem-se:

$$S_2 = \frac{\frac{2}{x^{k+2}} \left(1 - \frac{1}{x} \right) + \frac{1}{x^{k+2}} \frac{1}{x}}{\left(1 - \frac{1}{x} \right)^2} =$$

$$= \frac{\frac{2x - 2 + 1}{x^{k+3}}}{\frac{(x - 1)^2}{x^2}} = \frac{2x - 1}{x^{k+1}(x - 1)^2}$$

Como

$$S = \sum_{i=k+2}^{\infty} \frac{i}{x^i} = S_1 + S_2 =$$

$$= \frac{k}{(x - 1)x^{k+1}} + \frac{2x - 1}{x^{k+1}(x - 1)^2} .$$

de onde,

$$S = \frac{(k + 2)x - (k + 1)}{x^{k+1}(x - 1)^2}$$

Índice Remissivo

- Shift Point*, 21, 32
- Al-Seedy, 21
- Análise dos resultados, 50, 51, 53
- Davis et al, 20
- Diagrama de fluxo, 14, 22, 32
- Disciplina de atendimento, 8
- Erlang, 2
- Estudo de caso, 47
- Fogliatti & Mattos, 10, 21
- Kendall, 9, 13
- Li & Yang, 5
- Little, 10
- Modelo de filas
- $M/M/c(k) \leq 2/\infty/FIFO$, 20, 22
 - estrutura básica, 7
 - $M/M/c/\infty/FIFO$, 12
 - notação, 9
- Moder & Phillips, 4, 21
- Pesquisa Operacional, 1, 2
- Princípio de Conservação de Energia, 14,
23, 33
- Processo de Atendimento, 8
- Processo de Chegada, 8
- Processo de Nascimento e Morte, 13
- Regime estacionário, 10, 14, 48
- Regime transiente, 10
- Romani, 4, 20
- Silva,Santos & Fogliatti, 21
- Teoria de Filas, 2, 4