



Modelagem de Filas a partir de Diagramas de Fluxos

Maria Cristina Fogliatti de Sinay

Índice

1	INTRODUÇÃO	1
1.1	Histórico	2
1.2	Potencialidade da Teoria de Filas	3
2	SISTEMAS COM FILA	6
2.1	Estrutura Básica	6
2.2	Notação	8
2.3	Medidas de Desempenho do Sistema	9
3	O PROCESSO DE NASCIMENTO E MORTE	10
4	MODELOS DE FILAS	15
4.1	O Modelo $M/M/1/\infty/FIFO$	15
4.2	Fórmula de Little	17
4.3	O Modelo $M/M/1/K/FIFO$	18
4.4	O Modelo $M/M/c/\infty/FIFO$	21
4.5	O Modelo $M/M/c/K/FIFO$	23
4.6	O Modelo $M/M/\infty/\infty/FIFO$ (“Self - Service”)	27
	BIBLIOGRAFIA	28

Capítulo 1

INTRODUÇÃO

As filas de espera por serviços fazem parte do dia a dia dos indivíduos na sociedade moderna e como não podem ser evitadas, tendem a ser toleradas, apesar dos atrasos e inconveniências que causam. Entretanto, os processos geradores de filas podem ser estudados e dimensionados de forma a aliviar os prejuízos em tempo e produtividade assim como as perdas financeiras que elas acarretam.

A **Teoria de Filas** consiste na modelagem analítica de processos ou sistemas que resultam em espera e tem como objetivo determinar e avaliar quantidades, denominadas **medidas de desempenho**, que expressam a produtividade/operacionalidade dos mesmos. Entre essas medidas podem-se citar: número de elementos na fila, tempo de espera pelo atendimento e tempo ocioso dos prestadores do serviço. O estudo dessas quantidades é importante na tomada de decisão quanto à modificação ou manutenção da operação do sistema no seu estado atual e facilita o dimensionamento racional da infraestrutura dos recursos humanos e financeiros, dos equipamentos e instalações, visando um melhor desempenho global. Dessa forma, os conceitos e a teoria básica de Filas são fundamentais para a gerência e administração de sistemas produtivos.

Em processos com determinadas características, após seu funcionamento durante um certo período de tempo, as medidas de desempenho mencionadas tendem a se estabilizar. Neste caso, o intervalo de tempo de funcionamento do sistema, $[t_0, t)$, pode ser dividido em dois sub-intervalos: $[t_0, t^*)$ e $[t^*, t)$ onde t_0 é o instante de entrada em operação e t^* é o instante a partir do qual as medidas de desempenho se mantêm estáveis. Diz-se que em $[t_0, t^*)$ o sistema se encontra no **regime transiente** enquanto que em $[t^*, t)$, o sistema se encontra no **regime estacionário**.

No regime transiente, a variabilidade das medidas de desempenho dificulta as representações analíticas das mesmas, sendo necessário para tal,

conhecimentos matemáticos avançados. No regime estacionário, a estabilização dessas medidas permite o uso dos seus respectivos valores médios para a avaliação do sistema.

Outra característica da maioria dos sistemas reais com fila é a utilização de um critério particular para o atendimento que determina que o primeiro cliente a entrar em atendimento quando um servidor fica disponível, é o primeiro cliente da fila (o atendimento é feito na ordem de chegadas). Estudaremos aqui processos que entram no regime estacionário e que adotam o critério de atendimento mencionado.

1.1 Histórico

Nesta seção é apresentado um breve resumo da evolução da Teoria de Filas e suas aplicações, desde o primeiro estudo realizado pelo matemático A.K. Erlang no ano 1909 para o problema de congestionamento de linhas telefônicas na Dinamarca. A.K. Erlang é considerado por alguns autores como o “pai” da Teoria de Filas, devido ao fato do seu trabalho ter-se antecipado por varias décadas aos conceitos modernos desta teoria. Já no ano de 1917 publicou o livro *“Solutions of Some Problems in the Theory of Probabilities of Significance in Automatic Telephone Exchanges”*, onde sua experiência ficou documentada.

Desde então, as áreas de economia, de administração e de processamento de fluxos usufruíram dessa técnica destacando-se entre outros, problemas de congestionamento de tráfego, de escoamento de fluxo de carga em terminais, de carregamento/descarregamento de veículos, de escoamento de fluxo de processamento de informações, de formação de estoques, de comunicação de computadores, etc.

Dentre os trabalhos desenvolvidos em diferentes áreas, podem-se citar as modelagens apresentadas por: J.C. Tanner (1951) para o cálculo do tempo médio de espera de pedestres para atravessar uma rua sem sinal, J.L. Everett (1953) para o problema de escoamento de fluxo de barcos em terminais portuários, A. Cobham (1954) para o problema de reparo de maquinárias, P.M. Morse (1962) e N.U. Prabhu (1965) para o problema de formação de estoques. A partir de 1960 a Teoria de Filas foi também utilizada para modelar problemas concernentes à Ciência da Computação (C.V. Ramamoorthy (1965), P.J. Courtois (1977)), destacando-se a partir de 1980, as aplicações a redes de filas (E. Gelembe e G. Pujolle (1987), J. Walrand (1988)), em comunicação de computadores (J. Daigle, (1991)).

Das publicações em português sobre Teoria de Filas podem-se mencionar o livro de A.G.N. Novaes (1975) que apresenta aplicações direcionadas ao

planejamento de transportes e a publicação de M. Magalhães (1996) que trata de modelos de redes de filas frequentemente aplicados em ciência da computação.

1.2 Potencialidade da Teoria de Filas

Após a identificação do modelo que se adequa ao sistema sob análise, são determinadas as medidas de desempenho do mesmo. Essas medidas devem ser analisadas levando-se em consideração dois pontos de vista muitas vezes conflitantes: o do usuário e o do gerente.

Do ponto de vista do usuário, as medidas essenciais indicativas do desempenho do sistema são o tamanho médio da fila e os tempos médios de espera na fila e de permanência no sistema. Do ponto de vista do gerente, as medidas de interesse são as que traduzem o volume de serviço realizado, dentre elas: tempo médio de serviço e tempo médio ocioso dos servidores.

O serviço oferecido pode depender de atividade humana, de equipamentos ou de ambos. Em qualquer dos casos existe capital investido e custos operacionais e de manutenção. O correto dimensionamento do serviço no que se refere ao número de postos de atendimento e à velocidade de processamento é essencial para se manter um equilíbrio entre o capital disponibilizado no sistema e os retornos financeiros/sociais do mesmo.

O dimensionamento do sistema considerando apenas um dos pontos de vista mencionados, implica em insatisfação/perdas para o outro setor envolvido. Para ilustrar esse fato, na Figura 1.1 são representados na ordenada o número de postos de atendimento em paralelo e na abcissa os custos, C_G e C_U dispendidos pela gerência (construção, manutenção, operação do sistema) e pelos usuários (espera pelo serviço ou atendimento) respectivamente. Essas funções crescem/decrescem respectivamente com o aumento do número de postos de atendimento.

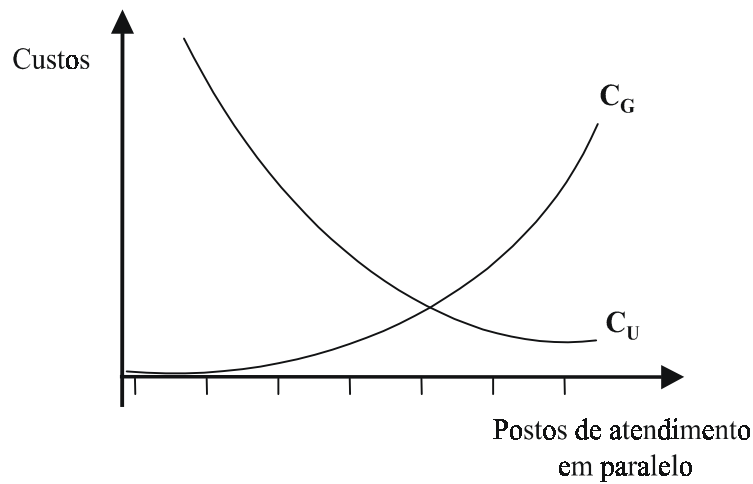


Figura 1.1 : Custos dos usuários e da gerência em função do número de postos de atendimento

Um número de postos de atendimento em excesso produz custos gerenciais elevados, que se repassados aos usuários podem inviabilizar o serviço, mesmo com a vantagem de se aguardar pouco tempo pelo mesmo. Poucos postos de atendimento, entretanto, provocam a insatisfação do usuário, que se tiver outra opção, geralmente abandona o sistema por não suportar um tempo de espera excessivo.

Por isso, se faz necessária a construção de uma função de custo total, C_T , que de alguma forma englobe os dois pontos de vista. Por exemplo:

$$C_T = aC_U + bC_G$$

ilustrada na Figura 1.2 a seguir.

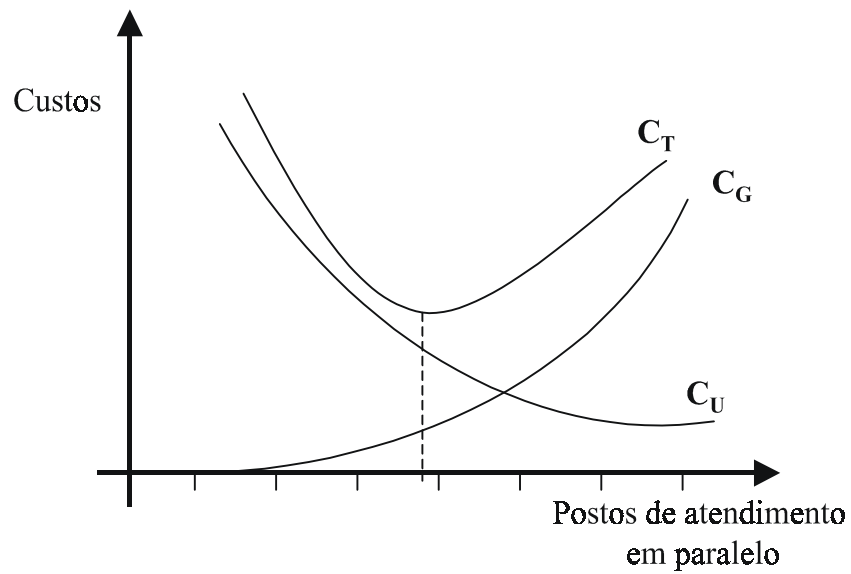


Figura 1.2 : Função de custo total

De posse da função C_T de custo total, o sistema deve ser dimensionado segundo o valor mínimo dessa função, atendendo assim usuários e gerência globalmente.

A Teoria das Filas pode ser usada para avaliar o desempenho operacional de sistemas em funcionamento ou para dimensionar eficientemente novos sistemas.

Capítulo 2

SISTEMAS COM FILA

Um **Sistema com Fila** é qualquer processo onde usuários oriundos de uma determinada população chegam para receber um serviço pelo qual esperam, se for necessário, saindo do sistema assim que o serviço é completado. Essa espera acontece quando a demanda é maior do que a capacidade de atendimento oferecido, em termos de fluxo.

2.1 Estrutura Básica

Um Sistema com Fila é composto fisicamente por usuários, por canais ou postos de serviço/atendimento e por um espaço designado para a espera.

Os usuários chegam segundo um determinado comportamento que caracteriza o processo de chegadas, para serem atendidos em canais ou postos de serviço (que funcionam em paralelo) segundo um padrão de atendimento. Enquanto os postos estão ocupados, os usuários aguardam na fila em um espaço designado para tal. Assim que um canal de serviço fica livre, um dos usuários da fila é chamado para atendimento segundo um critério estabelecido pela gerência. Uma vez completado o serviço, o usuário é liberado do sistema. Essa estrutura básica está representada na Figura 2.1 a seguir e os elementos necessários para a caracterização de um Sistema com Fila são apresentados nas Seções 2.2.1 a 2.2.5.

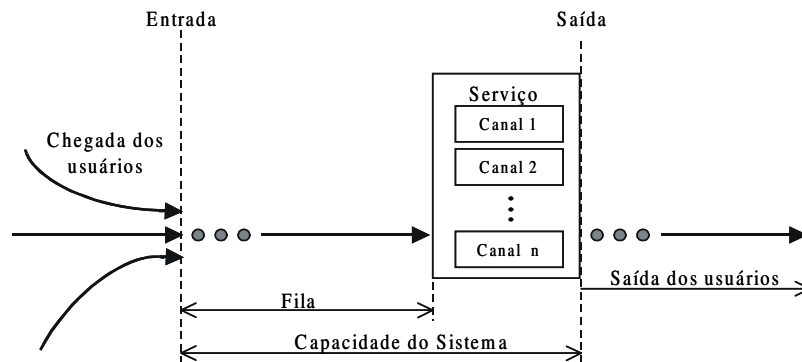


Figura 2.1: Representação esquemática de um Sistema com Fila

2.1.1 Processo de Chegadas dos Usuários

O processo de chegadas dos usuários é especificado pelo comportamento do fluxo de chegadas dos mesmos ao sistema. Se são conhecidos o número de chegadas e os instantes de tempo em que elas acontecem, esse processo é denominado determinístico; caso contrário, tem-se um comportamento aleatório constituindo um processo estocástico caracterizado por uma distribuição de probabilidade. Essa distribuição é especificada por um parâmetro denominado taxa que representa o número médio de usuários que chegam ao sistema por unidade de tempo.

A população de onde o usuário se origina pode ser finita ou infinita. Como exemplo do primeiro caso, tem-se a população constituída por poços de petróleo a serem explorados no instante em que suas vazões o indicarem e para o segundo caso, uma população constituída por carros chegando a uma oficina mecânica para reparo.

Os usuários podem chegar ao sistema individualmente ou em grupos cujos tamanhos podem ser constantes ou aleatórios. Como exemplo para o primeiro caso, pode-se citar a chegada de navios a um porto para carregamento/descarregamento e para o segundo caso, pessoas em excursão chegando de ônibus a pontos turísticos.

2.1.2 Processo de Atendimento

O processo de atendimento é especificado pelo comportamento do fluxo de usuários atendidos e a sua caracterização é análoga à do processo de chegadas

2.1.3 Canais ou Postos de Serviço/Atendimento

Os canais ou postos de serviço são os locais (físicos ou não) onde são atendidos os usuários. O número de postos de um sistema pode ser finito ou infinito. Como exemplo do primeiro caso podem-se citar os guichês de um posto de pedágio e como exemplo do segundo caso qualquer atendimento do tipo “self-service” onde cliente e servidor são a mesma pessoa e onde o serviço está sempre disponível.

2.1.4 Capacidade do Sistema

A capacidade do sistema é o número máximo de usuários que o mesmo comporta (incluindo fila e atendimento) que pode ser finita ou infinita. Como exemplo do primeiro caso, pode-se citar um posto de vistoria de carros que admite um número máximo de carros aguardando pelo serviço e do segundo caso, um porto onde navios chegam para descarregamento aguardando, se necessário, no mar. No caso de capacidade finita, quando esta é atingida, os usuários que chegam até o instante da próxima liberação, são rejeitados.

2.1.5 Disciplina de Atendimento

A disciplina de atendimento é o critério estabelecido pela gerência do sistema, segundo o qual os usuários que se encontram na fila são atendidos quando um posto fica disponível. Dentre as disciplinas mais utilizadas podem-se citar:

FIFO (“first in - first out”): os usuários são atendidos na ordem das chegadas. Essa disciplina de atendimento é a mais comumente adotada. Como exemplos tem-se a venda de ingressos num cinema, atendimento bancário, carregamento e descarregamento de navios num porto.

LIFO (“last in - first out”): o primeiro usuário a ser atendido é o que chegou por último. Como exemplos tem-se a utilização de estoques verticais ou horizontais e o carregamento de pilhas de containeres em navios.

PRI (“priority service”): o atendimento aos usuários segue uma ou mais prioridades pré-estabelecidas pela gerência do sistema. Como exemplos podem ser mencionados a internação hospitalar ou cirurgias, a exploração de poços petrolíferos e tarefas a serem processadas por um computador.

SIRO (“service in random order”): o atendimento aos usuários segue uma ordem aleatória. Como exemplos desta disciplina tem-se a contemplação de consórcios e seleção de ganhadores em concursos populares.

2.2 Notação

Para descrever um Sistema com Fila é utilizada neste livro a notação proposta por Kendall (1953), que é da forma $A/B/C/D/E$ onde A e B (literais)

denotam respectivamente as distribuições dos tempos entre chegadas sucessivas e de atendimento; C e D (numéricos), denotam o número de postos de atendimento em paralelo e a capacidade física do sistema, respectivamente e E uma das siglas que representa a disciplina de atendimento empregada conforme as definições apresentadas na Seção 2.2.5.

Entre as escolhas mais comuns de A e B podem-se citar:

D : distribuição determinística ou degenerada;

e para comportamento aleatório,

M : distribuição exponencial (“Memoryless” ou Markoviana);

E_k : distribuição Erlang do tipo k ;

G : distribuição geral (não especificada).

2.3 Medidas de Desempenho do Sistema

A utilização da Teoria de Filas permite avaliar a eficiência de um sistema por meio da análise de suas características utilizando medidas de operacionalidade/desempenho. Essas características na maioria das vezes, mudam ao longo do tempo, devendo então, ser representadas por variáveis aleatórias. Assim, os valores esperados dessas variáveis podem ser utilizados como medidas de desempenho do sistema. Dentre essas podem-se citar:

- Número médio de usuários na fila (L_q) e no sistema (L).
- Tempo médio de espera de um usuário qualquer na fila (W_q).
- Tempo médio de permanência de um usuário qualquer no sistema (W).

Outras medidas de desempenho que caracterizam o comportamento do sistema são:

- Probabilidade de se ter no máximo um número n_0 pré-fixado de usuários no sistema ($P(N \leq n_0)$).
- Probabilidade de um usuário qualquer ter que aguardar mais do que um determinado tempo t_0 na fila ($P(T_q > t_0)$).
- Probabilidade de se ter algum servidor ocioso ($P(N < c)$).

Capítulo 3

O PROCESSO DE NASCIMENTO E MORTE

Um processo é denominado de **nascimento e morte** (Birth-Death Process) se as únicas mudanças permitidas a partir de um determinado estado n do processo são para seus vizinhos imediatos, ou seja, para $n + 1$ ou $n - 1$. A mudança para o estado $n + 1$, $\forall n$ representa um nascimento e para o estado $n - 1$, $n > 0$, uma morte; essas transições se processam com taxas λ_n e μ_n respectivamente, que dependem somente do estado n do sistema (independem do tempo).

Para processos de nascimento e morte as seguintes hipóteses são assumidas:

- 1) no instante inicial $t_0 = 0$, o sistema está vazio, isto é, $N(0) = 0$;
- 2) nascimentos e mortes são eventos estatisticamente independentes;
- 3) dado que o sistema está no estado n , no intervalo de tempo $(t, t + \Delta t)$, Δt tão pequeno quanto se queira, a probabilidade de ocorrer:
 - a) um nascimento é igual a $\lambda_n \Delta t + o(\Delta t)$;
 - b) uma morte é igual a $\mu_n \Delta t + o(\Delta t)$;
 - c) mais de um evento (nascimento(s) e/ou morte(s)) é desprezível, igual a $o(\Delta t)$, onde,

$$\lim_{\Delta t \rightarrow 0} \frac{o(\Delta t)}{\Delta t} = 0.$$

Usando a notação:

$$P_i^n(\Delta t) = P\{\text{ocorrência de } i \text{ nascimentos em } \Delta t\}$$

e

$$P_j^m(\Delta t) = P\{\text{ocorrência de } j \text{ mortes em } \Delta t\},$$

a terceira hipótese pode ser-reescrita da forma a seguir:

$$P_1^n(\Delta t) = \lambda_n \Delta t + o(\Delta t) \quad \forall n = 0, 1, 2, \dots, \quad (3.1)$$

$$P_1^m(\Delta t) = \mu_n \Delta t + o(\Delta t) \quad \forall n = 1, 2, \dots, \quad (3.2)$$

e

$$P_i^n(\Delta t)P_j^m(\Delta t) = o(\Delta t), \quad \forall i, j \mid (i + j) > 1. \quad (3.3)$$

De (3.1), (3.2) e (3.3) obtém-se as probabilidades de não haver nascimentos (3.4) nem mortes (3.5) em intervalos pequenos de tempo:

$$P_0^n(\Delta t) = 1 - \lambda_n \Delta t - o(\Delta t) \quad \forall n = 0, 1, 2, \dots, \quad (3.4)$$

$$P_0^m(\Delta t) = 1 - \mu_n \Delta t - o(\Delta t) \quad \forall n = 1, 2, \dots, \quad (3.5)$$

Com essas hipóteses determinam-se, para todo n , as probabilidades P_n dos estados do processo, como mostrado a seguir.

Dividindo-se o intervalo de observação $(0, t + \Delta t)$ em dois subintervalos disjuntos $(0, t]$ e $(t, t + \Delta t)$, verifica-se que o sistema está no estado $n > 0$ no instante $(t + \Delta t)$, para Δt pequeno, se ocorre um dos seguintes eventos mutuamente excludentes:

- i) o sistema está no estado n no tempo t e no intervalo de tempo $(t, t + \Delta t)$ não há nenhum nascimento e nenhuma morte;
- ii) o sistema está no estado $n - 1$ no tempo t e no intervalo de tempo $(t, t + \Delta t)$ há um nascimento e nenhuma morte;
- iii) o sistema está no estado $n + 1$ no tempo t e no intervalo de tempo $(t, t + \Delta t)$ há uma morte e nenhum nascimento.

O sistema está no estado $n = 0$ no instante $(t + \Delta t)$, para Δt pequeno, se ocorre um dos seguintes eventos mutuamente excludentes:

- i) o sistema está no estado 0 no tempo t e não há nenhum nascimento no intervalo de tempo $(t, t + \Delta t)$;
- ii) o sistema está no estado 1 no tempo t e há uma morte e nenhum nascimento no intervalo de tempo $(t, t + \Delta t)$.

Utilizando as hipóteses 1, 2 apresentadas anteriormente e o Teorema da Probabilidade Total, pode-se calcular a probabilidade do sistema estar no estado n no instante $(t + \Delta t)$:

$$P_n(t + \Delta t) = P_n(t)P_0^n(\Delta t)P_0^m(\Delta t) + P_{n-1}(t)P_1^n(\Delta t)P_0^m(\Delta t) + P_{n+1}(t)P_0^n(\Delta t)P_1^m(\Delta t) + o(\Delta t), \quad \forall n \geq 1, \quad (3.6)$$

$$P_0(t + \Delta t) = P_1(t)P_0^n(\Delta t)P_1^m(\Delta t) + P_0(t)P_0^n(\Delta t) + o(\Delta t). \quad (3.7)$$

Substituindo (3.1), (3.2), (3.3), (3.4) e (3.5) em (3.6) e (3.7), tem-se:

$$\begin{aligned}
 P_n(t + \Delta t) &= P_n(t) [1 - \lambda_n \Delta t - o(\Delta t)] [1 - \mu_n \Delta t - o(\Delta t)] \\
 &\quad + P_{n-1}(t) [\lambda_{n-1} \Delta t + o(\Delta t)] [1 - \mu_{n-1} \Delta t - o(\Delta t)] \\
 &\quad + P_{n+1}(t) [1 - \lambda_{n+1} \Delta t - o(\Delta t)] [\mu_{n+1} \Delta t + o(\Delta t)] + o(\Delta t) \\
 &= P_n(t) - \lambda_n \Delta t P_n(t) - \mu_n \Delta t P_n(t) + P_{n-1}(t) \lambda_{n-1} \Delta t \\
 &\quad + P_{n+1}(t) \mu_{n+1} \Delta t + o(\Delta t), \quad \forall n \geq 1,
 \end{aligned} \tag{3.8}$$

$$\begin{aligned}
 P_0(t + \Delta t) &= P_0(t) [1 - \lambda_0 \Delta t - o(\Delta t)] \\
 &\quad + P_1(t) [1 - \lambda_1 \Delta t - o(\Delta t)] [\mu_1 \Delta t + o(\Delta t)] + o(\Delta t) \\
 &= P_0(t) - \lambda_0 \Delta t P_0(t) + \mu_1 \Delta t P_1(t) + o(\Delta t),
 \end{aligned} \tag{3.9}$$

lembrando que $(\Delta t)^2 \simeq o(\Delta t)$ e $o(\Delta t) \cdot \Delta t \simeq o(\Delta t)$.

De (3.8) e (3.9), obtém-se:

$$\begin{aligned}
 \frac{P_n(t + \Delta t) - P_n(t)}{\Delta t} &= -\lambda_n P_n(t) - \mu_n P_n(t) + \lambda_{n-1} P_{n-1}(t) \\
 &\quad + \mu_{n+1} P_{n+1}(t) + \frac{o(\Delta t)}{\Delta t}, \quad \forall n \geq 1, \\
 \frac{P_0(t + \Delta t) - P_0(t)}{\Delta t} &= -\lambda_0 P_0(t) + \mu_1 P_1(t) + \frac{o(\Delta t)}{\Delta t}.
 \end{aligned}$$

Tomando-se os limites quando $\Delta t \rightarrow 0$, tem-se:

$$\begin{aligned}
 \frac{dP_n(t)}{dt} &= -\lambda_n P_n(t) - \mu_n P_n(t) + \lambda_{n-1} P_{n-1}(t) \\
 &\quad + \mu_{n+1} P_{n+1}(t), \quad \forall n \geq 1,
 \end{aligned} \tag{3.10}$$

$$\frac{dP_0(t)}{dt} = -\lambda_0 P_0(t) + \mu_1 P_1(t), \tag{3.11}$$

que formam um sistema infinito de equações diferenciais que representa as probabilidades dos estados do sistema.

Como o processo de nascimento e morte é uma cadeia de Markov irreduzível (fato que pode ser facilmente observado na Figura 3.1), existe um tempo t^* a partir do qual ele entra no regime estacionário mantendo suas características estáveis. Neste caso,

$$\frac{dP_n(t)}{dt} = 0, \quad \forall n, \quad \forall t > t^*.$$

Dessa forma, o sistema de equações diferenciais (3.10) e (3.11) se converte no sistema de equações algébricas:

$$0 = -\lambda_n P_n - \mu_n P_n + \lambda_{n-1} P_{n-1} + \mu_{n+1} P_{n+1}, \quad \forall n \geq 1, \tag{3.12}$$

$$0 = -\lambda_0 P_0 + \mu_1 P_1. \tag{3.13}$$

Rearranjando (3.12) tem-se:

$$\lambda_n P_n - \mu_{n+1} P_{n+1} = \lambda_{n-1} P_{n-1} - \mu_n P_n, \quad \forall n \geq 1,$$

e usando-se recorrência,

$$\begin{aligned} \lambda_n P_n - \mu_{n+1} P_{n+1} &= \lambda_{n-1} P_{n-1} - \mu_n P_n \\ &= \lambda_{n-2} P_{n-2} - \mu_{n-1} P_{n-1} = \dots \\ &= \lambda_0 P_0 - \mu_1 P_1. \end{aligned} \quad (3.14)$$

De (3.13) e (3.14) tem-se

$$\lambda_{n-1} P_{n-1} - \mu_n P_n = 0 \quad \forall n \geq 1.$$

Então,

$$\begin{aligned} P_n &= \frac{\lambda_{n-1}}{\mu_n} P_{n-1} = \frac{\lambda_{n-1} \lambda_{n-2}}{\mu_n \mu_{n-1}} P_{n-2} = \dots = \frac{\lambda_{n-1} \lambda_{n-2} \lambda_{n-3} \dots \lambda_0}{\mu_n \mu_{n-1} \mu_{n-2} \dots \mu_1} P_0 \\ &= P_0 \prod_{i=1}^n \frac{\lambda_{i-1}}{\mu_i} \quad n \geq 1. \end{aligned} \quad (3.15)$$

Como $\sum_{n \geq 0} P_n = 1$, obtém-se:

$$P_0 = \frac{1}{1 + \sum_{n \geq 1} \prod_{i=1}^n \frac{\lambda_{i-1}}{\mu_i}}. \quad (3.16)$$

A partir de (3.15) e (3.16), tem-se a distribuição limite dos estados do sistema (P_0, P_1, P_2, \dots) que é também a distribuição do estado de regime estacionário do processo, totalmente determinada pelas taxas de nascimento e de morte.

As equações (3.12) e (3.13) são denominadas **equações de balanço** ou **de equilíbrio** e podem também ser obtidas diretamente do diagrama apresentado na Figura 3.1, utilizando o princípio de conservação de energia, isto é, para cada estado, “o fluxo que entra é igual ao fluxo que sai”.

Um processo de nascimento e morte pode ser esquematizado pelo diagrama apresentado na Figura 3.1, onde os nós representam os estados n (número de elementos no sistema) e os arcos, as transições entre os estados (taxas de nascimento e de morte λ_n e μ_n respectivamente).

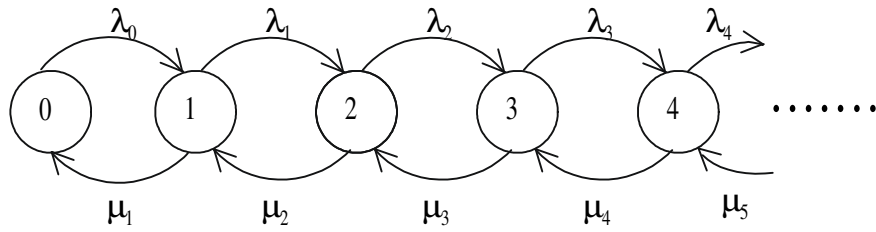


Figura 3.1: Diagrama de fluxo de um processo de nascimento e morte.

Dessa forma, para qualquer estado $n \geq 1$, tem-se:

$$\lambda_{n-1}P_{n-1} + \mu_{n+1}P_{n+1} = \lambda_n P_n + \mu_n P_n,$$

e para o estado $n = 0$,

$$\lambda_0 P_0 = \mu_1 P_1.$$

As probabilidades $P_n, n \geq 0$ podem ser utilizadas para analisar outras características da cadeia.

Capítulo 4

MODELOS DE FILAS

Neste capítulo são apresentados modelos de filas representando situações que se comportam como processos markovianos de nascimento e morte.

4.1 O Modelo $M/M/1/\infty/FIFO$

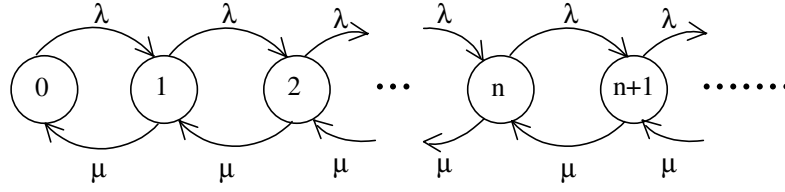
No modelo $M/M/1/\infty/FIFO$, os tempos entre chegadas sucessivas e os tempos de atendimento seguem distribuições exponenciais. As chegadas e os atendimentos caracterizam um processo de nascimento e morte, lembrando que somente um único evento pode acontecer em períodos pequenos de tempo. As taxas de chegada (ingresso) ao sistema e de atendimento são constantes e dadas respectivamente por:

$$\lambda_n = \lambda, \quad \forall n \geq 0$$

e

$$\mu_n = \mu \quad \forall n \geq 1,$$

existe um único posto de atendimento, não há limitação para o espaço reservado para a fila de espera e a ordem de acesso de usuários ao serviço segue a ordem das chegadas dos mesmos ao sistema. Este modelo é representado esquematicamente pelo diagrama de fluxo da Figura 4.1.

Figura 4.1: Diagrama de fluxo para o modelo $M/M/1/\infty/FIFO$

4.1.1 Equações de Balanço

$$\lambda P_0 = \mu P_1, \quad (4.1)$$

$$\lambda P_{n-1} + \mu P_{n+1} = \lambda P_n + \mu P_n, \quad \forall n \geq 1, \quad (4.2)$$

Então,

$$P_n = \frac{\lambda^n}{\mu^n} P_0, \quad \forall n \geq 1, \quad (4.3)$$

e

$$P_0 = \left[\sum_{n=0}^{\infty} \left(\frac{\lambda}{\mu} \right)^n \right]^{-1},$$

onde a soma geométrica só converge se $\frac{\lambda}{\mu} < 1$. Neste caso tem-se

$$P_0 = 1 - \frac{\lambda}{\mu}. \quad (4.4)$$

Quando $\frac{\lambda}{\mu} \geq 1$, não pode ser obtido um valor finito para P_0 nem para as probabilidades P_n , $\forall n > 0$, e portanto o processo não entra no regime estacionário.

O parâmetro ρ definido como:

$$\rho = \frac{\lambda}{\mu}$$

é denominado **taxa de ocupação/utilização do sistema**, que substituído em (4.4) e (4.3) leva a:

$$P_n = \rho^n (1 - \rho) \quad \forall n \geq 0. \quad (4.5)$$

4.1.2 Medidas de Desempenho

A seguir são calculadas algumas medidas de desempenho para o modelo $M/M/1/\infty/FIFO$.

Número médio de usuários no Sistema: L

Seja N a variável aleatória discreta que representa o número de usuários no sistema no regime estacionário, com distribuição de probabilidade $\{P_n\}$, $n \geq 0$ e valor esperado L . Tem-se então:

$$L = E[N] = \sum_{n=0}^{\infty} nP_n = (1-\rho) \sum_{n=0}^{\infty} n\rho^n = (1-\rho)\rho \sum_{n=1}^{\infty} n\rho^{n-1} = (1-\rho)\rho \sum_{n=1}^{\infty} \frac{d\rho^n}{d\rho}.$$

Supondo $\rho < 1$, $\sum_{n=0}^{\infty} \rho^n$ converge e então,

$$\sum_{n=1}^{\infty} \frac{d\rho^n}{d\rho} = \frac{d}{d\rho} \left(\sum_{n=0}^{\infty} \rho^n \right).$$

Dessa forma,

$$\begin{aligned} L &= (1-\rho)\rho \frac{d}{d\rho} \left(\sum_{n=0}^{\infty} \rho^n \right) = (1-\rho)\rho \frac{d}{d\rho} \left(\frac{1}{1-\rho} \right) \\ &= (1-\rho)\rho \frac{1}{(1-\rho)^2} = \frac{\rho}{(1-\rho)}. \end{aligned} \quad (4.6)$$

Número médio de usuários na Fila: L_q

$$\begin{aligned} L_q &= E[N_q] = \sum_{n=1}^{\infty} (n-1)P_n = \sum_{n=1}^{\infty} nP_n - \sum_{n=1}^{\infty} P_n = L - 1 + P_0 \\ &= \frac{\rho}{(1-\rho)} - 1 + 1 - \rho = \frac{\rho - \rho + \rho^2}{(1-\rho)} = \frac{\rho^2}{(1-\rho)}. \end{aligned} \quad (4.7)$$

4.2 Fórmula de Little

Um importante resultado geral que independe de propriedades específicas das distribuições dos tempos entre chegadas e de atendimento estabelece que: “O número médio de usuários num sistema é igual ao produto da taxa de ingresso no sistema pelo tempo médio de permanência de um usuário no mesmo”. Este resultado, conhecido como **fórmula de Little** (Little, 1961), é representado analiticamente por:

$$L = E[\Lambda]W, \quad (4.8)$$

onde $E[\Lambda]$ é a taxa média de ingressos no sistema.

Outras relações entre as medidas de desempenho do sistema válidas em geral são:

$$W = W_q + E[S], \quad (4.9)$$

onde S é o tempo que um usuário qualquer permanece em atendimento.

$$W_q = \frac{L_q}{E[\Lambda]} \quad (4.10)$$

e

$$L = L_q + E[\Lambda]E[S]. \quad (4.11)$$

As provas são intuitivas e a utilidade deste conjunto de relações reside no fato de que o conhecimento de uma das medidas de desempenho, implica no conhecimento das outras.

Para o modelo $M/M/1$, tem-se $E[\Lambda] = \lambda$ e $E[S] = \frac{1}{\mu}$, portanto,

$$L = \frac{\lambda}{\mu - \lambda}, \quad (4.12)$$

então,

$$W = \frac{\lambda}{\lambda(\mu - \lambda)} + \frac{1}{\mu} = \frac{1}{\mu - \lambda}, \quad (4.13)$$

$$W_q = \frac{\rho^2}{(1 - \rho)\lambda} = \frac{\lambda}{\mu(\mu - \lambda)} \quad (4.14)$$

e

$$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)}. \quad (4.15)$$

4.3 O Modelo $M/M/1/K/FIFO$

No modelo $M/M/1/K/FIFO$, os tempos entre chegadas sucessivas e os tempos de atendimento seguem distribuições exponenciais de parâmetros λ e μ respectivamente. Como no caso anterior, as chegadas e os atendimentos caracterizam um processo de nascimento e morte. Entretanto, a taxa de ingresso ao sistema, λ'_n difere da taxa de chegada para $n \geq K$ tendo em vista a existência de limitação na capacidade do sistema (igual a K). Neste caso, as taxas de ingresso e de atendimento são dadas respectivamente por:

$$\lambda'_n = \begin{cases} \lambda, & \forall \quad 0 \leq n < K, \\ 0, & n \geq K \end{cases}$$

e

$$\mu_n = \mu \quad \forall n \geq 1.$$

Existe um único posto de atendimento que atende os usuários na ordem das chegadas. Este modelo é representado esquematicamente pelo diagrama de fluxo da Figura 4.2.

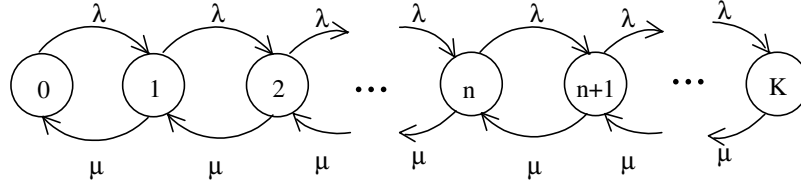


Figura 4.2: Diagrama de fluxo do modelo $M/M/1/K/FIFO$

Observa-se neste modelo, que o espaço de estados é um conjunto finito com $K + 1$ elementos.

4.3.1 Equações de Balanço

$$\lambda P_0 = \mu P_1, \quad (4.16)$$

$$\lambda P_{n-1} + \mu P_{n+1} = \lambda P_n + \mu P_n, \quad 1 \leq n < K, \quad (4.17)$$

$$\lambda P_{K-1} = \mu P_K. \quad (4.18)$$

Então,

$$P_n = \frac{\lambda^n}{\mu^n} P_0, \quad 1 \leq n < K \quad (4.19)$$

e

$$P_0 = \frac{1}{\sum_{n=0}^K \rho^n}.$$

A soma finita do denominador sempre converge, porém para valores distintos, dependendo de ρ . Dessa forma,

$$P_0 = \begin{cases} \frac{1}{K+1}, & \text{se } \rho = 1, \\ \frac{1-\rho}{1-\rho^{K+1}}, & \text{se } \rho \neq 1, \end{cases} \quad (4.20)$$

de onde, $\forall n \leq K$

$$P_n = \begin{cases} \frac{1}{K+1}, & \text{se } \rho = 1, \\ \frac{(1-\rho)\rho^n}{1-\rho^{K+1}}, & \text{se } \rho \neq 1. \end{cases} \quad (4.21)$$

4.3.2 Medidas de Desempenho

Número médio de usuários no Sistema: L

$$L = E[N] = \sum_{n=0}^K nP_n.$$

Se $\rho = 1$,

$$L = \sum_{n=0}^K n \frac{1}{(K+1)} = \frac{1}{(K+1)} \frac{K(K+1)}{2} = \frac{K}{2}. \quad (4.22)$$

Se $\rho \neq 1$

$$L = \frac{1-\rho}{(1-\rho^{K+1})} \sum_{n=0}^K n\rho^n = \rho \frac{[1 + K\rho^{K+1} - \rho^K(K+1)]}{(1-\rho)(1-\rho^{K+1})}.$$

Número médio de usuários na Fila: L_q

$$L_q = E[N_q] = \sum_{n=1}^K (n-1)P_n = \sum_{n=0}^K nP_n - \sum_{n=1}^K P_n = L - 1 + P_0. \quad (4.23)$$

Tempo médio de permanência no Sistema: W

Para se usar a fórmula de Little (4.8) deve ser feita uma modificação pois por existir limitação do espaço reservado para a fila, rejeições acontecem com taxa λP_K cada vez que o sistema atinge o estado K .

A taxa de efetivo ingresso λ' , é dada por:

$$\lambda' = \lambda - \lambda P_K = \lambda(1 - P_K),$$

que substituída em (4.8) leva a

$$W = \frac{L}{\lambda(1 - P_K)}. \quad (4.24)$$

Tempo médio de espera na Fila: W_q

De (4.9):

$$W_q = W - \frac{1}{\mu}$$

e usando (4.10) tem-se a fórmula equivalente:

$$W_q = \frac{L_q}{\lambda(1 - P_K)}. \quad (4.25)$$

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

No modelo $M/M/c/\infty/FIFO$, os tempos entre chegadas sucessivas seguem uma distribuição exponencial de parâmetro λ e há c servidores, cada um dos quais com tempos de atendimento que seguem distribuições exponenciais, de parâmetro μ . Como nos casos anteriores, as chegadas e os atendimentos caracterizam um processo de nascimento e morte, cujas taxas são dadas por:

$$\lambda_n = \lambda, \quad \forall n \geq 0$$

e

$$(4.26)$$

$$\mu_n = \begin{cases} n\mu, & \text{se } 0 \leq n < c, \\ c\mu, & \text{se } n \geq c. \end{cases}$$

Não há limitação física para o espaço reservado para a fila e a ordem de acesso de usuários ao serviço segue a ordem das chegadas dos mesmos ao sistema. Este modelo é representado esquematicamente pelo diagrama de fluxo da Figura 4.3.

Figura 4.3: Diagrama de fluxo do Modelo $M/M/c/\infty/FIFO$

4.4.1 Equações de Balanço

$$\lambda P_0 = \mu P_1, \quad (4.27)$$

$$\lambda P_{n-1} + (n+1)\mu P_{n+1} = \lambda P_n + n\mu P_n, \quad 1 \leq n < c, \quad (4.28)$$

$$\lambda P_{n-1} + c\mu P_{n+1} = \lambda P_n + c\mu P_n, \quad n \geq c. \quad (4.29)$$

Denotando $r = \frac{\lambda}{\mu}$, a taxa de utilização do sistema é dada por:

$$\rho = \frac{r}{c} = \frac{\lambda}{c\mu}$$

e então,

$$P_n = \begin{cases} P_0 \frac{r^n}{n!}, & 0 \leq n < c, \\ P_0 \frac{r^n}{c^{n-c}c!}, & n \geq c. \end{cases} \quad (4.30)$$

Então,

$$P_0 = \left(\sum_{n=0}^{c-1} \frac{r^n}{n!} + \frac{cr^c}{c!(c-r)} \right)^{-1}. \quad (4.31)$$

4.4.2 Medidas de Desempenho

Número médio de elementos na Fila: L_q

$$\begin{aligned} L_q &= E[N_q] = \sum_{n=c}^{\infty} (n-c)P_n \\ &= \frac{P_0 cr^{c+1}}{c!(c-r)^2}. \end{aligned} \quad (4.32)$$

Usando a fórmula de Little (4.8) e as relações (4.9) e (4.10) obtém-se as demais medidas de desempenho.

Número médio de elementos no Sistema: L

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

Tempo médio de espera na Fila: W_q

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

Tempo médio de permanência no Sistema: W

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

4.5 O Modelo $M/M/c/K/FIFO$

No modelo $M/M/c/K/FIFO$, os tempos entre chegadas sucessivas seguem uma distribuição exponencial de parâmetro λ e há c servidores, cada um dos quais com tempos de atendimento que seguem distribuições exponenciais, de parâmetro μ . Como nos casos anteriores, trata-se um processo de nascimento e morte. Entretanto, a taxa de ingresso ao sistema, λ'_n , difere da taxa de chegada λ para $n \geq K$ tendo em vista a existência de limitação na capacidade do sistema (igual a K). As taxas de ingresso e de atendimento são dadas respectivamente por:

$$\lambda'_n = \begin{cases} \lambda, & \text{se } 0 \leq n < K, \\ 0, & \text{se } n \geq K. \end{cases} \quad (4.33)$$

$$\mu_n = \begin{cases} n\mu, & \text{se } 0 \leq n < c, \\ c\mu, & \text{se } c \leq n \leq K. \end{cases} \quad (4.34)$$

A ordem de acesso de usuários ao serviço é a ordem das chegadas dos mesmos ao sistema. Este modelo é representado esquematicamente pelo diagrama de fluxo da Figura 4.4.

Figura 4.4: Diagrama de fluxo do Modelo $M/M/c/K/FIFO$

4.5.1 Equações de Balanço

$$\lambda P_0 = \mu P_1, \quad (4.35)$$

$$\lambda P_{n-1} + (n+1)\mu P_{n+1} = \lambda P_n + n\mu P_n, \quad 1 \leq n < c, \quad (4.36)$$

$$\lambda P_{n-1} + c\mu P_{n+1} = \lambda P_n + c\mu P_n, \quad c \leq n < K, \quad (4.37)$$

$$\lambda P_{K-1} = c\mu P_K. \quad (4.38)$$

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

$$P_n = \begin{cases} \frac{r^n}{n!} P_0, & \text{se } 0 \leq n \leq c-1, \\ \frac{r^n}{c! c^{n-c}} P_0, & \text{se } c \leq n \leq K \end{cases} \quad (4.39)$$

e

$$P_0 = \begin{cases} \left[\sum_{n=0}^{c-1} \frac{r^n}{n!} + \frac{r^c (K-c+1)}{c!} \right]^{-1}, & \text{se } \frac{r}{c} = 1, \\ \left[\sum_{n=0}^{c-1} \frac{r^n}{n!} + \frac{r^c \left(1 - \left[\frac{r}{c}\right]^{K-c+1}\right)}{c! \left(1 - \frac{r}{c}\right)} \right]^{-1}, & \text{se } \frac{r}{c} \neq 1. \end{cases} \quad (4.40)$$

4.5.2 Medidas de Desempenho

Número médio de usuários na Fila: L_q

$$L_q = E[N_q] = \sum_{n=c}^K (n-c) P_n.$$

Se $\frac{r}{c} = 1$, então,

$$L_q = \frac{P_0 r^c}{c!} \sum_{n=c}^K (n-c) = \frac{P_0 r^c}{c!} \sum_{i=0}^{K-c} i = \frac{P_0 r^c}{c!} \frac{(K-c+1)(K-c)}{2}. \quad (4.41)$$

Se $\frac{r}{c} \neq 1$ então

$$\begin{aligned} L_q &= \frac{P_0 r^{c+1}}{c! c} \sum_{n=c}^K \frac{d}{d\left(\frac{r}{c}\right)} \left(\frac{r}{c}\right)^{n-c} \\ &= \frac{P_0 r^{c+1}}{c! c} \frac{\left(\left[\left(\frac{r}{c}\right) - 1\right] (K-c+1) \left(\frac{r}{c}\right)^{K-c} + 1 - \left(\frac{r}{c}\right)^{K-c+1}\right)}{\left(1 - \left(\frac{r}{c}\right)\right)^2}. \end{aligned} \quad (4.42)$$

Número médio de usuários no Sistema: L

Como:

$$\begin{aligned}
 L_q &= \sum_{n=c}^K (n-c)P_n = \sum_{n=c}^K nP_n - c \sum_{n=c}^K P_n \\
 &= L - \sum_{n=0}^{c-1} nP_n - c + c \sum_{n=0}^{c-1} P_n = L - \sum_{n=0}^{c-1} (n-c)P_n - c,
 \end{aligned}$$

tem-se

$$L = L_q + c + \sum_{n=0}^{c-1} (n-c)P_n. \quad (4.43)$$

Tempo médio de espera na Fila: W_q

Utilizando as fórmulas (4.8) e (4.10) e lembrando que existe limitação no espaço físico reservado para a fila, tem-se

$$W_q = \frac{L_q}{\lambda'} \quad (4.44)$$

com $\lambda' = \lambda(1 - P_K)$.

Tempo médio de permanência no Sistema: W

$$W = \frac{L}{\lambda'}, \quad (4.45)$$

com $\lambda' = \lambda(1 - P_K)$.

4.5.3 O Caso Particular $M/M/c/c/FIFO$

Neste caso, não é permitida a formação de fila e as taxas de ingresso e de atendimento do sistema são dadas respectivamente por:

$$\lambda'_n = \begin{cases} \lambda, & 0 \leq n < c, \\ 0, & n \geq c, \end{cases}$$

e

$$\mu_n = n\mu, \quad 0 < n \leq c.$$

Este modelo é representado esquematicamente pelo diagrama de fluxo da Figura 4.5.

Figura 4.5: Diagrama de fluxo do modelo $M/M/c/c/FIFO$ **Equações de Balanço:**

$$\lambda P_0 = \mu P_1, \quad (4.46)$$

$$\lambda P_{n-1} + (n+1)\mu P_{n+1} = \lambda P_n + n\mu P_n, \quad 1 \leq n < c, \quad (4.47)$$

$$\lambda P_{c-1} = c\mu P_c. \quad (4.48)$$

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

$$P_n = \frac{r^n}{n!} P_0, \quad 0 < n \leq c,$$

de onde,

$$P_0 = \left[\sum_{n=0}^c \frac{r^n}{n!} \right]^{-1}$$

e

$$P_n = \frac{\frac{r^n}{n!}}{\sum_{i=0}^c \frac{r^i}{i!}}, \quad 0 \leq n \leq c. \quad (4.49)$$

A probabilidade do sistema se encontrar lotado,

$$P_c = \frac{\frac{r^c}{c!}}{\sum_{i=0}^c \frac{r^i}{i!}} \quad (4.50)$$

é conhecida como a **fórmula de perda de Erlang** (Erlang, 1917) e corresponde à percentagem de usuários rejeitados pela limitação física do sistema. Esta fórmula foi desenvolvida por Erlang quando estudava o problema de congestionamento de redes telefônicas.

4.6 O Modelo $M/M/\infty/\infty/FIFO$ (“Self - Service”)

Este modelo se caracteriza pelo fato de ter tantos servidores quanto usuários no sistema. Os tempos entre chegadas sucessivas e os tempos de atendimento de cada servidor seguem distribuições exponenciais de parâmetros λ e μ , respectivamente.

Para este modelo, tem-se:

$$\lambda_n = \lambda, \forall n \geq 0 \quad (4.51)$$

e

$$\mu_n = n\mu, \forall n \geq 1. \quad (4.52)$$

Este modelo é representado esquematicamente pelo diagrama de fluxo da Figura 4.6.

Figura 4.6: Diagrama de fluxo do modelo $M/M/\infty/\infty/FIFO$

4.6.1 Equações de Balanço

$$\begin{aligned} \lambda P_0 &= \mu P_1, \\ \lambda P_{n-1} + (n+1)\mu P_{n+1} &= \lambda P_n + n\mu P_n, \quad \forall n \geq 1. \end{aligned}$$

Então,

$$P_n = P_0 \prod_{k=1}^n \frac{\lambda_k}{k\mu} = P_0 \frac{\lambda^n}{n!\mu^n}, \quad \forall n \geq 1.$$

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

$$P_0 = e^{-\rho}$$

e

$$P_n = \frac{\rho^n e^{-\rho}}{n!}. \quad (4.53)$$

De (4.53) se deduz que o número de elementos no sistema é uma variável aleatória discreta com distribuição de Poisson de parâmetro $\rho = \frac{\lambda}{\mu}$.

4.6.2 Medidas de Desempenho

Para este modelo, tem-se:

$$L_q = 0$$

e

$$W_q = 0,$$

de onde

$$L = \frac{\lambda}{\mu}$$

e

$$W = \frac{1}{\mu}.$$

Bibliografia

- [1] Cobham, A., *Priority Assignment in Waiting-Line Problems*, Journal of the Operations Research Society of America (1954).
- [2] Courtois, P.J., *Decomposability: Queuing and Computer System Applications*, Ac. Press, New York (1977).
- [3] Daigle, A., *Queueing Theory for Computer Communications*, Addison-Wesley, New York (1991).
- [4] Erlang, A.K., *Probability and Telephone Calls*, Nyt Tidsskr., Mat.Sc.B, n°20, 33-39 (1909).
- [5] Erlang, A. K., *Solutions of Some Problems in the Theory of Probabilities of Significance in Automatic Telephone Exchanges*, Electroteknikeren n° 13 (1917).
- [6] Everett, J.L., *Seaport Operation as a Stochastic Process*, The Journal of the Operations Research Society of America, n°1 (1953).
- [7] Gelembel, E., Pujolle, G., *Introduction to Queueing Networks*, Wiley, Paris (1987).
- [8] Gross, D., Harris, C.M., *Fundamentals of Queueing Theory*, Third Edition, John Wiley & Sons Inc., NY (1998).
- [9] Hillier, F. S., Lieberman, G. J., *Introduction to Operations Research*. Holden-Day Inc. 12th Ed.(1974).
- [10] Kendall, D.G., *Stochastic Processes Occurring in the Theory of Queues and their Analysis by the Method of Imbedded Markov Chains*, Ann. Math. Statist. n°24, 338-354 (1953).
- [11] Kleinrock, L., *Queueing Systems Vol. 1: Theory*, Wiley Interscience, New York (1975).

- [12] Little, J. D. C., *A Proof for the Queueing Formula $L = \lambda W$* , Operations Research, 9, n°3 (1961).
- [13] Magalhães, M.N., *Introdução à Rede de Filas*, ABE - Associação Brasileira de Estatística (1996).
- [14] Morse, P.M., *Queues, Inventories and Maintenance*, capítulos 7 e 8, Wiley, New York (1962).
- [15] Novaes, A.G.N., *Pesquisa Operacional e Transportes: Modelos Probabilísticos*, Mc. Graw Hill do Brasil LTD (1975).
- [16] Page, E., *Queueing Theory in OR*, Butterworth & Co. Ltd. Londres (1972).
- [17] Prabhu, N.U., *Queues and Inventories*, Wiley, New York (1965).
- [18] Ramamoorthy, C.V., *Discrete Markov Analysis of Computer Programs*, Proc. ACM National Conference, 386-392 (1965).
- [19] Tanner, J.C., *The Delay to Pedestrians Crossing a Road*, Biometrika n°38 (parte 3 e 4), 383-392 (1951).
- [20] Walrand, J., *Introduction to Queueing Networks*, Prentice Halll, Englewood Cliffs, New Jersey (1988).