TP 547- Princípios de Simulação de Sistemas de Comunicação

Prof. Samuel Baraldi Mafra

Inatel
Instituto Nacional de Telecomunicações

Créditos: - Prof. Edson J. C. Gimenez - Prof. Dr. Antônio Marcos Alberti - Prof. Dr. José Marcos Câmara Brito

Introdução à Teoria das Filas:

Teoria de Filas

A teoria de filas é uma das mais interessantes aplicações da teoria da probabilidade, sendo de grande importância para a análise e dimensionamento de sistemas de comunicações e também em sistemas ligados à ciência da computação.

O estudo da teoria de filas é o estudo da espera.

Agner Krarup Erlang
The Theory of Probabilities and
Telephone Conversations

Os sistemas de filas aparece em diversas situações do nosso cotidiano, tais como:

- Fila de pessoas em supermercados e bancos.
- Fila de pessoas para embarcar em um avião.
- Fila de carros em um semáforo.
- Fila de carros aguardando por conserto em uma oficina.
- Fila de containers a serem descarregados em um porto.

Sistemas de fila também são formados em sistemas e redes de comunicações:

- Fila de pacotes aguardando por transmissão.
- Fila de pacotes aguardando por roteamento/comutação.
- Fila de pacotes recebidos na placa de rede de um terminal.
- Fila de chamadas telefônicas aguardando por linha em um PABX.
- Fila de amostras de voz recebidas em um telefone IP.
- Fila de símbolos a serem codificados em um transmissor de TV Digital.

Métricas de Desempenho

O desempenho das redes de comunicações é tipicamente medido através das seguintes métricas de desempenho:

- Ocupação
- Atraso
- Perda
- Bloqueio
- Utilização
- Vazão
- Eficiência

Métricas de Desempenho

- Ocupação
  - É uma contagem de elementos existentes em um sistema.
  - Exemplo: quantidade de pacotes armazenados em um buffer.
  - Para cada fila lógica é mantido um contador que indica a ocupação atual.

Métricas de Desempenho

- Atraso
    - É o tempo parcial ou total gasto para realizar as funções implementadas em cada camada da rede.
    - Praticamente todas as atividades realizadas para enviar ou processar informações implicam na ocorrência de atrasos.
    - Dentre os principais atrasos existentes estão:
        - Atraso de Estabelecimento de Conexões
        - Atraso de Empacotamento
        - Atraso de Transmissão/Propagação
        - Atraso de Armazenamento
        - Atraso de Roteamento/Comutação
        - Atraso de Segmentação/Remontagem
        - Atraso de Retransmissão
        - Atraso de Processamento

Métricas de Desempenho

- Variação de Atraso
  - É a variação do atraso sofrido na rede.
  - É uma métrica importante para aplicações em tempo real interativas.

Métricas de Desempenho

- Perda
  - Indica se a rede está perdendo pacotes.
  - As principais causas de perda de pacotes são:
    - Falta de recursos de armazenamento.
    - Erros ou falhas.
    - Congestionamento.
    - Atraso elevado.
    - Colisão no acesso ao meio.
- Bloqueio
  - Em algumas tecnologias, se a rede não possui mais recursos para atender uma nova conexão, esta poderá ser bloqueada.
  - Estes recursos podem ser tanto em termos de espaço para armazenamento dos pacotes quanto em termos de largura de banda.
- Utilização: Indica a fração dos recursos de transmissão/armazenamento que estão sendo utilizados.

Métricas de Desempenho

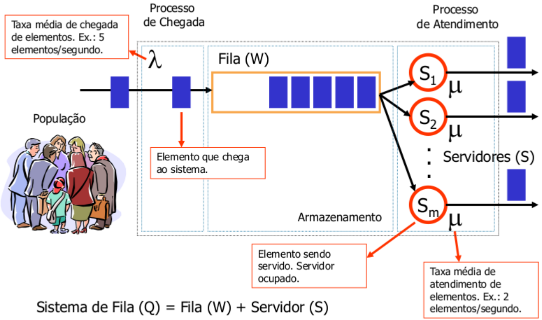
- Vazão
  - Indica a taxa efetiva de transmissão.
  - Algumas tecnologias que utilizam o compartilhamento do meio físico podem possuir uma vazão consideravelmente abaixo da capacidade de transmissão disponível.
- Eficiência
  - Indica a razão entre a quantidade de informação útil transmitida e a quantidade total de informações utilizadas para efetivar tal transmissão (carga útil + overhead).
  - Tecnologias que possuem recursos dedicados possuem uma eficiência menor do que tecnologias que compartilham recursos.

Sistema de filas

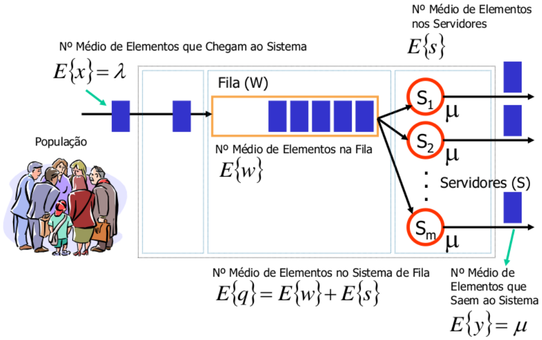
O que é um Sistema de Filas?
Um sistema de filas (Q - Queuing System) é um sistema composto por:

- Uma ou mais filas (W - Waiting Line) onde são armazenados os elementos que aguardam por atendimento.
- Um ou mais servidores (S - Servers) que atendem os elementos.
- Um processo de chegada, que define como os elementos chegam ao sistema.
- Um processo de atendimento, que define como os elementos são atendidos pelo sistema.
- O tamanho da população que gera os elementos.

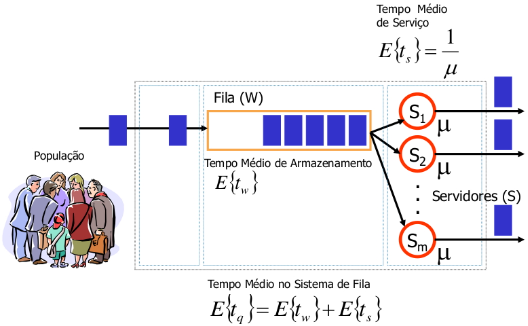
Sistema de Filas com 1 Fila e vários Servidores:



Métricas de Desempenho: Ocupação



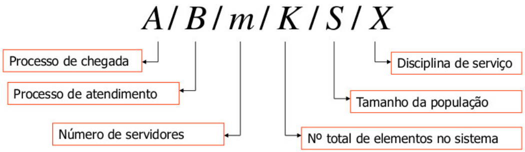
Métricas de Desempenho: Atraso



Sistema de filas

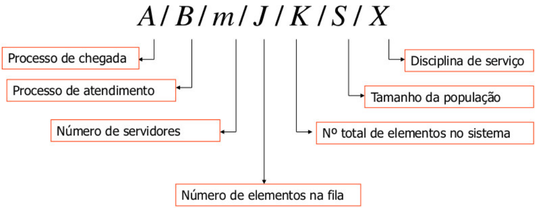
Notação de Kendall e Notação Expandida

A notação de Kendall (David Kendall) foi desenvolvida em 1951 para descrever o comportamento de um sistema de fila em uma única frase:



Sistema de filas

Notação de Kendall Expandida:



Notação de Kendall e Notação Expandida

- É comum vermos sistemas definidos com a notação simplificada: A/B/m.
- Neste caso assume-se que não há limite para o tamanho da fila (fila infinita), a população é infinita, e a disciplina de tratamento é FIFO: $A/B/m/\infty/\infty/FIFO$

Notação de Kendall e Notação Expandida

- Processo de Chegada (A)
- Descreve o processo que modela as chegadas de elementos ao sistema.
- As seguintes opções são utilizadas:
  - M: Markoviano (Distribuição Exponencial)
  - D: Determinístico (Constante)
  - E_k: Erlang (Erlang-k)
  - H_k: Hiperexponencial
  - G: Genérico (Intervalo de tempo entre chegadas é tratado de forma genérica, independente da distribuição).

Notação de Kendall e Notação Expandida

- Processo de Atendimento (B)
- Descreve o processo que modela o atendimento de elementos no sistema.
- As seguintes opções são utilizadas:
  - M: Markoviano (Distribuição Exponencial)
  - D: Determinístico (Constante)
  - E_k: Erlang (Erlang-k)
  - H_k: Hiperexponencial
  - G: Genérico (Processo de atendimento é tratado de forma genérica, independente da distribuição).

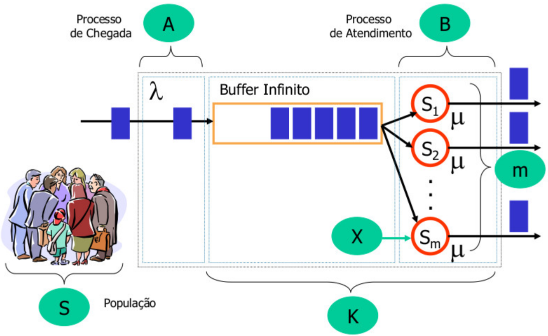
Tamanho da População (S)

- Descreve o tamanho da população que gera elementos para o sistema. Tipicamente é considerada como infinito.

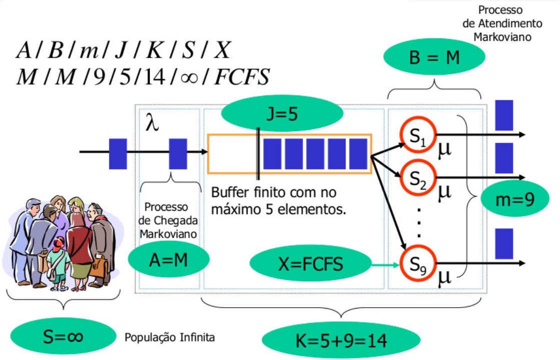
Disciplina de Serviço (X)

- Os elementos que aguardam por serviço na fila podem ser selecionados de acordo com uma regra chamada disciplina de serviço. Dentre as principais disciplinas estão:
    - FCFS - First Come First Served (**FIFO) Primeiro elemento que chega é o primeiro a ser atendido.
    - LCFS - Last Come First Served Último elemento que chega é o primeiro a ser atendido.
    - SIRO - Service In a Random Order Elementos são atendidos em ordem aleatória.

Notação de Kendall e Notação Expandida

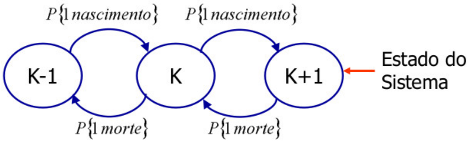


Notação de Kendall e Notação Expandida



Processo de Nascimento e Morte e Diagrama de Estado

- É uma classe especial de processos estocásticos em que são permitidas somente transições aos estados vizinhos.



- As probabilidades de transição são determinadas em função do estado atual e das médias das distribuições dos processos de chegada (λ) e de atendimento (μ).

Processo de Nascimento e Morte e Diagrama de Estado

- Assim, para um sistema em equilíbrio tem-se:

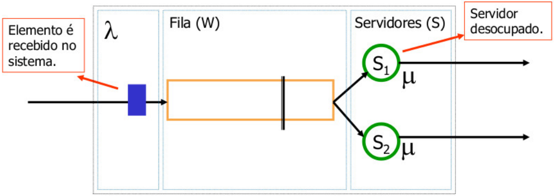
$$P(1\ nascimento) = \lambda_k$$

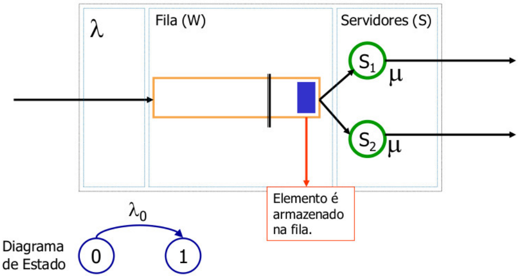
$$P(1\ morte) = \mu_k$$

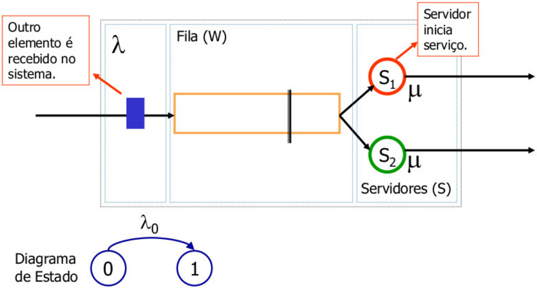
sendo

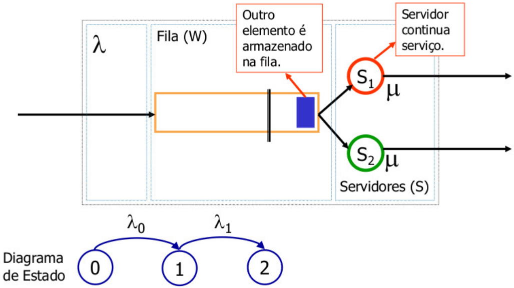
- λ_k - Média de chegada de elementos no estado K.
- μ_k- Média de saída de elementos no estado K.

$M/M/2/2/4/\infty/FCFS$

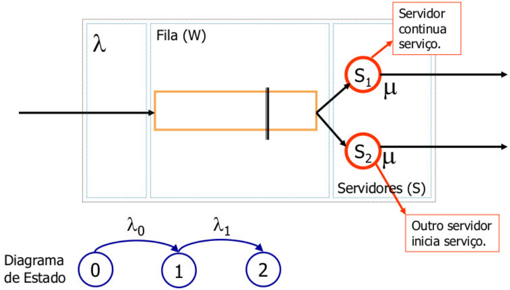


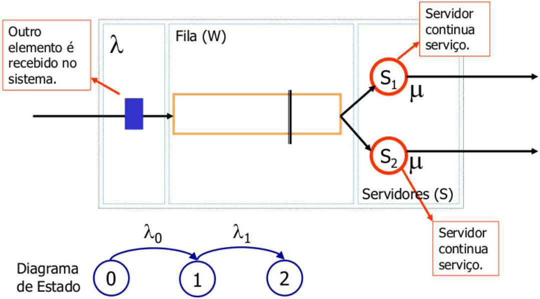

λ
Fila (W)
Servidores (S)
S_1
μ
S_2
μ
Elemento é armazenado na fila.
λ_0
Diagrama de Estado
0
1

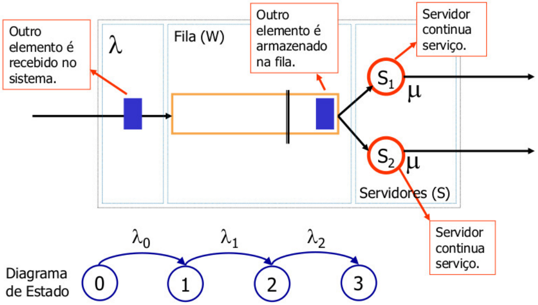

Outro
elemento é
recebido no
sistema.
λ
Fila (W)
Servidor
inicia
serviço.
S_1
μ
S_2
μ
Servidores (S)
λ_0
Diagrama
de Estado
0
1

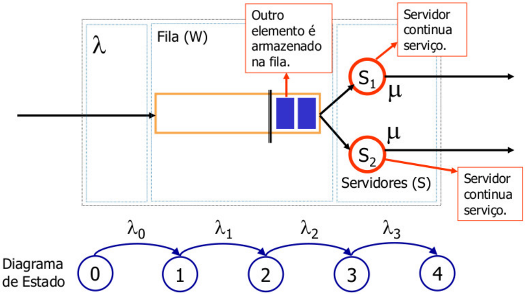


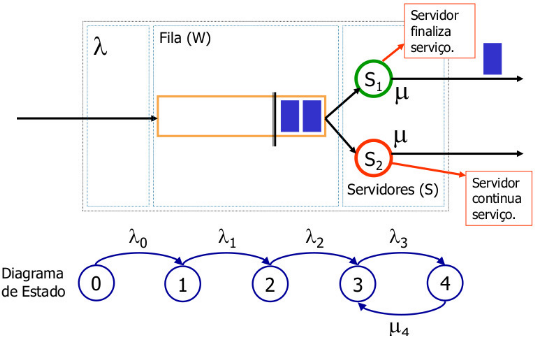
λ
Fila (W)
Outro elemento é armazenado na fila.
Servidor continua serviço.
S_1
μ
S_2
μ
Servidores (S)
λ_0
λ_1
Diagrama de Estado
0
1
2

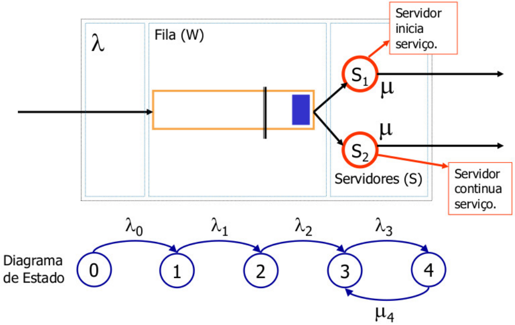

λ
Fila (W)
Servidor continua serviço.
S_1
μ
S_2
μ
Servidores (S)
Outro servidor inicia serviço.
λ_0
λ_1
Diagrama de Estado
0
1
2

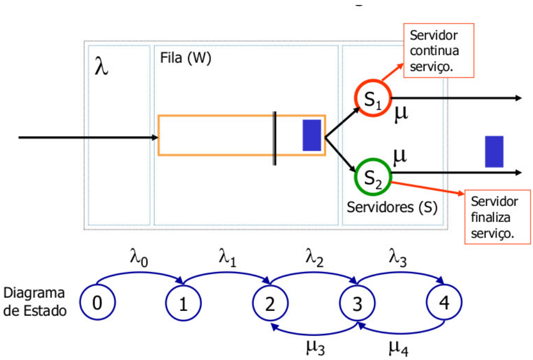

Outro elemento é recebido no sistema.
λ
Fila (W)
Servidor continua serviço.
S1
μ
S2
μ
Servidores (S)
Servidor continua serviço.
λ0
λ1
Diagrama de Estado
0
1
2


Outro elemento é recebido no sistema.
λ
Fila (W)
Outro elemento é armazenado na fila.
Servidor continua serviço.
S1
μ
S2
μ
Servidores (S)
Servidor continua serviço.
λ0
λ1
λ2
Diagrama de Estado
0
1
2
3

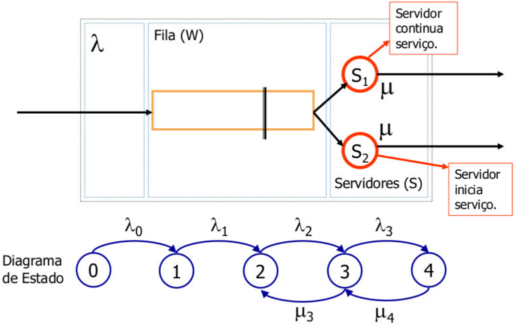

λ
Fila (W)
Outro elemento é armazenado na fila.
Servidor continua serviço.
S_1
μ
μ
S_2
Servidores (S)
Servidor continua serviço.
λ_0
λ_1
λ_2
λ_3
Diagrama de Estado
0
1
2
3
4

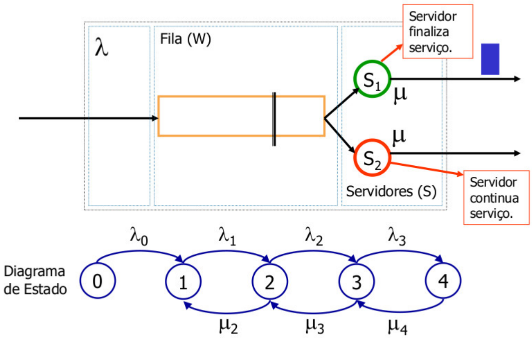

λ
Fila (W)
Servidor
finaliza
serviço.
S_1
μ
μ
S_2
Servidores (S)
Servidor
continua
serviço.
λ_0
λ_1
λ_2
λ_3
Diagrama
de Estado
0
1
2
3
4
μ_4

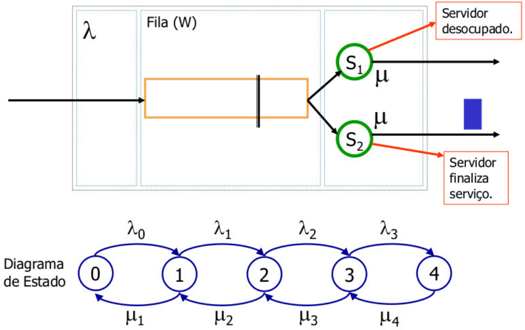

λ
Fila (W)
Servidor
inicia
serviço.
S_1
μ
μ
S_2
Servidores (S)
Servidor
continua
serviço.
λ_0
λ_1
λ_2
λ_3
Diagrama
de Estado
0
1
2
3
4
μ_4



Servidor continua serviço.
λ
Fila (W)
S_1
μ
μ
S_2
Servidores (S)
Servidor finaliza serviço.
λ_0
λ_1
λ_2
λ_3
Diagrama de Estado
0
1
2
3
4
μ_3
μ_4

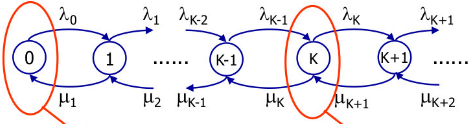

λ
Fila (W)
Servidor continua serviço.
S_1
μ
μ
S_2
Servidores (S)
Servidor inicia serviço.
λ_0
λ_1
λ_2
λ_3
Diagrama de Estado
0
1
2
3
4
μ_3
μ_4


λ
Fila (W)
Servidor finaliza serviço.
S_1
μ
μ
S_2
Servidores (S)
Servidor continua serviço.
λ_0
λ_1
λ_2
λ_3
Diagrama de Estado
0
1
2
3
4
μ_2
μ_3
μ_4


Servidor desocupado.
λ
Fila (W)
S_1
μ
μ
S_2
Servidor finaliza serviço.
λ_0
λ_1
λ_2
λ_3
Diagrama de Estado
0
1
2
3
4
μ_1
μ_2
μ_3
μ_4

Equações de Equilíbrio

- Em equilíbrio, a soma dos fluxos que saem de um determinado estado (λ_k), deve ser igual a soma dos fluxos que chegam a este mesmo estado (μ_{k+1}).
- Ou seja: $\sum$ Fluxo de Entrada $=\sum$ Fluxo de Saída



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

$$(\lambda_K + \mu_K)P_K = \mu_{K+1}.P_{K+1} + \lambda_{K-1}.P_{K-1}$$

- A soma das probabilidades deve ser igual a um:

$$\sum_{k=0}^{\infty} P_k = 1$$

;

- Considerando-se esta equação, o sistema de equações pode ser resolvido como:

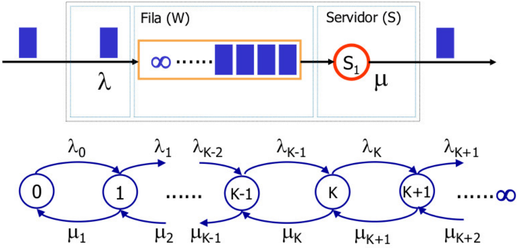
$$P_1 = \frac{\lambda_0}{\mu_1}.P_0 \qquad P_2 = \frac{\lambda_1}{\mu_1}.\frac{\lambda_0}{\mu_2}P_0 \qquad \ldots\ldots \qquad P_K = P_0.\prod_{i=0}^{K-1}\frac{\lambda_i}{\mu_{i+1}}$$

$$\sum_{K=1}^{\infty} P_0.\prod_{i=0}^{K-1}\frac{\lambda_i}{\mu_{i+1}} + P_0 = 1 \qquad P_0 = \frac{1}{\left(\sum_{K=1}^{\infty}\prod_{i=0}^{K-1}\frac{\lambda_i}{\mu_{i+1}}\right)+1}$$

Filas com Servidor Único

Sistema de Fila com Servidor Único e Buffer Infinito

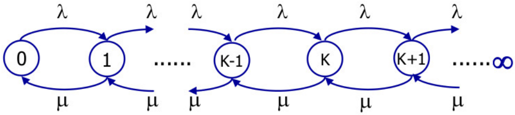
- Este sistema é conhecido como M/M/1, ou na notação expandida $M/M/1/\infty/\infty/\infty/FCFS$.



No sistema M/M/1, todas as transições de nascimento tem valor igual a λ, e como existe somente um servidor, todas as transições de morte são iguais a μ. Ou seja:

$$\lambda_K = \lambda, \text{ para } K=0,1,\ldots,\infty$$

$$\mu_K = \mu, \text{ para } K=1,\ldots,\infty$$



As equações de equilíbrio, neste caso, são:

$$\begin{cases} (\lambda+\mu)P_K = \lambda.P_{K-1} + \mu.P_{K+1} & \mathrm{K} \geq 1 \\ \lambda P_0 = \mu.P_1 & \mathrm{K} = 0 \\ \sum_{K=0}^{\infty} P_K = 1 & \end{cases}$$

Resolvendo, tem-se:

$$P_1 = \frac{\lambda}{\mu}.P_0 \qquad P_2 = \frac{\lambda}{\mu}P_1 = \left(\frac{\lambda}{\mu}\right)^2 P_0 \qquad P_3 = \frac{\lambda}{\mu}P_2 = \left(\frac{\lambda}{\mu}\right)^3 P_0$$

$$\boxed{P_K = \left(\frac{\lambda}{\mu}\right)^K P_0}$$

Podemos encontrar P_0 fazendo-se:

$$P_0 + \sum_{K=1}^{\infty} P_K = 1 \qquad P_0 + \sum_{K=1}^{\infty} \left(\frac{\lambda}{\mu}\right)^K .P_0 = 1$$

$$P_0 \left(1 + \sum_{K=1}^{\infty} \left(\frac{\lambda}{\mu}\right)^K \right) = 1 \qquad P_0 = \frac{1}{\left(1 + \sum_{K=1}^{\infty} \left(\frac{\lambda}{\mu}\right)^K \right)}$$

Sabendo-se que:

$$\sum_{K=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^K = \frac{1}{1 - \frac{\lambda}{\mu}}$$

Então:

$$P_0 = 1 - \frac{\lambda}{\mu}$$

Fator de utilização

- Relação entre o tempo em que a facilidade está ocupada e o tempo total disponível.
- Relação entre a carga real do sistema e a máxima carga que o sistema pode manusear

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

Portanto:

$$P_0 = 1 - \rho$$

$$P_k = \rho^k(1 - \rho)$$

Exemplo

Um comutador de pacotes recebe em média 200 pacotes/segundo, cada um com um comprimento médio de 128 bytes. O comutador possui uma única linha de saída com capacidade de 256 kbps. Qual a utilização da linha de saída?

$$\lambda = 200 \text{ pc/s}$$

Exemplo

Um comutador de pacotes recebe em média 200 pacotes/segundo, cada um com um comprimento médio de 128 bytes. O comutador possui uma única linha de saída com capacidade de 256 kbps. Qual a utilização da linha de saída?

$$\lambda = 200 \text{ pc/s}$$

$$L = 128 * 8 = 1024 \text{ bits}$$

Exemplo

Um comutador de pacotes recebe em média 200 pacotes/segundo, cada um com um comprimento médio de 128 bytes. O comutador possui uma única linha de saída com capacidade de 256 kbps. Qual a utilização da linha de saída?

$$\lambda = 200 \text{ pc/s}$$

$$L = 128 * 8 = 1024 \text{ bits}$$

$$C = 256000 \text{ bps}$$

$$\mu = \frac{C}{L} = \frac{256000}{1024} = 250 \text{ pc/s}$$

Exemplo

Um comutador de pacotes recebe em média 200 pacotes/segundo, cada um com um comprimento médio de 128 bytes. O comutador possui uma única linha de saída com capacidade de 256 kbps. Qual a utilização da linha de saída?

$$\lambda = 200 \text{ pc/s}$$

$$L = 128 * 8 = 1024 \text{ bits}$$

$$C = 256000 \text{ bps}$$

$$\mu = \frac{C}{L} = \frac{256000}{1024} = 250 \text{ pc/s}$$

$$\rho = \frac{\lambda}{\mu} = \frac{200}{250} = 0.8$$

Número médio de Clientes no Sistema

$$E\{q\} = \sum_{K=0}^{\infty} K.P_K$$

$$E\{q\} = \sum_{K=0}^{\infty} K.\rho^K (1-\rho)$$

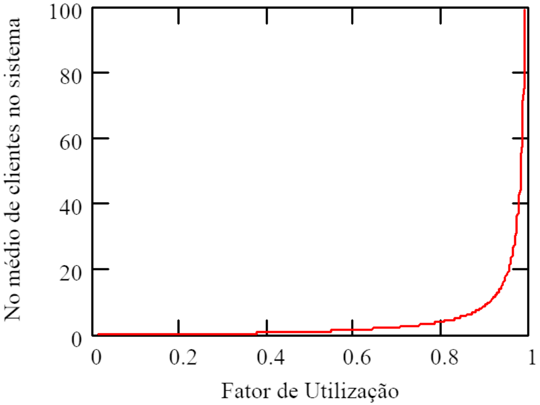
Tem-se que :

$$\sum_{K=0}^{\infty} K\rho^K = \frac{\rho}{(1-\rho)^2}$$

$$E[q] = \frac{\rho}{1-\rho}$$

Fila M/M/1 Numero Médio de Clientes no Sistema

$$E[q] = \frac{\rho}{1-\rho}$$



Teorema de Little:

- Diz que o número médio de elementos no sistema é igual a taxa média efetiva de chegadas no sistema multiplicada pelo tempo médio de permanência no sistema.

$$E[q] = \lambda E[t_q]$$

$$E[t_q] = \frac{E[q]}{\lambda}$$

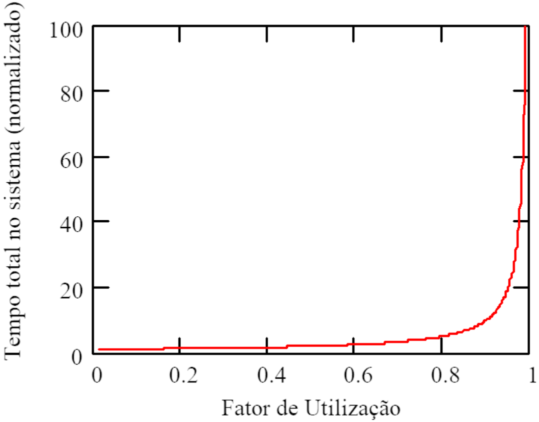
- Também é válido para as demais médias de elementos no sistema:

$$E[t_s] = \frac{E[s]}{\lambda}$$

$$E[t_w] = \frac{E[w]}{\lambda}$$

Tempo Médio de Permanência no Sistema

$$E[T_q] = \frac{E[q]}{\lambda} = \frac{1}{\mu - \lambda}$$



Exemplo

O nó de uma rede de comutação de pacotes recebe em média 480 pacotes por minuto (segundo uma distribuição de Poisson) para uma das suas linhas de saída. A linha tem uma taxa de transmissão de 64 kbps. A distribuição do tamanho da mensagem é aproximadamente exponencial com um tamanho médio de 4.000 bits. Considerando o buffer do comutador infinito, calcule:

- a) O tempo de serviço.
- b) A utilização da facilidade.
- c) O número médio de pacotes no sistema
- d) O tempo médio que um pacote gasta no comutador.
- e) O tempo médio que um pacote gasta na fila.
- f) O número médio de pacotes na fila

Exemplo

O nó de uma rede de comutação de pacotes recebe em média 480 pacotes por minuto (segundo uma distribuição de Poisson) para uma das suas linhas de saída. A linha tem uma taxa de transmissão de 64 kbps. A distribuição do tamanho da mensagem é aproximadamente exponencial com um tamanho médio de 4.000 bits. Considerando o buffer do comutador infinito, calcule:

- a) O tempo de serviço.

$$E[t_s] = \frac{L}{C} = \frac{4000}{64000} = 0.0625\text{s}$$

- b) A utilização da facilidade.

$$\rho = \frac{\lambda}{\mu} = \lambda E[t_s] = \frac{480 * 0.0625}{60} = 0.5$$

Exemplo

O nó de uma rede de comutação de pacotes recebe em média 480 pacotes por minuto (segundo uma distribuição de Poisson) para uma das suas linhas de saída. A linha tem uma taxa de transmissão de 64 kbps. A distribuição do tamanho da mensagem é aproximadamente exponencial com um tamanho médio de 4.000 bits. Considerando o buffer do comutador infinito, calcule:

- c) O número médio de pacotes no sistema

$$E[q] = \frac{\rho}{1-\rho} = \frac{0.5}{1-0.5} = 1 \text{ pacote}$$

- d) O tempo médio que um pacote gasta no comutador.

$$E[t_q] = \frac{1}{\mu - \lambda} = \frac{1}{16-8} = 1/8 = 0.125\text{s}$$

Exemplo

O nó de uma rede de comutação de pacotes recebe em média 480 pacotes por minuto (segundo uma distribuição de Poisson) para uma das suas linhas de saída. A linha tem uma taxa de transmissão de 64 kbps. A distribuição do tamanho da mensagem é aproximadamente exponencial com um tamanho médio de 4.000 bits. Considerando o buffer do comutador infinito, calcule:

- e) O tempo médio que um pacote gasta na fila.

$$E[t_w] = E[t_q] - E[t_s] = 0.125 - 0.0625 = 0.0625\text{s}$$

- f) O número médio de pacotes na fila

$$E[w] = \lambda * E[t_w] = 8 * 0.0625 = 0.5 \text{ pacote}$$

código: queuemm1.py e queuemm1partidas.py