

# **Ciência da Computação**

## **Disciplina de Modelagem e Simulação**

### **Teoria das Filas**

#### **Aula 07**

Professor: André Flores dos Santos



# Sistemas de Filas Estão em Todo Lugar!

- Restaurante Universitário: Estudantes chegam, fazem fila, são atendidos no balcão
- Caixa do Supermercado: Clientes com carrinhos, múltiplos atendentes, filas paralelas
- Call Center: Ligações chegam, ficam em espera, são atendidas por operadores
- Laboratório de Informática: Alunos chegam, aguardam computador livre, fazem atividades
- Pronto Socorro: Pacientes chegam, triagem, atendimento por prioridade
- Impressora Compartilhada: Documentos chegam, fila de impressão, uma impressora

# Anatomia Completa de um Sistema de Filas

- ENTRADA: Clientes chegam com taxa  $\lambda$  (lambda)
- FILA DE ESPERA: Capacidade limitada ou ilimitada
- SERVIDORES: 'c' atendentes, cada um com taxa ' $\mu$  (mi)'
- DISCIPLINA: Regra de atendimento (FIFO, prioridade, etc.)
- SAÍDA: Clientes atendidos deixam o sistema

Fórmula Principal:  $\rho = \lambda / (c \times \mu)$

Legenda:  $\rho$  = ocupação;  $\lambda$  = chegadas/tempo;  
 $\mu$  = atendimentos/tempo por servidor; c = nº de servidores

# Medindo a Taxa de Chegada ( $\lambda$ )

- Como Medir: Conte quantas pessoas chegam em um período fixo
- Exemplo Prático: Em 10 minutos chegaram 25 pessoas
- Cálculo:  $\lambda = 25 \text{ pessoas} \div 10 \text{ minutos} = 2,5 \text{ pessoas/minuto}$
- Variações: Taxa pode mudar ao longo do dia (pico vs. vale)
- Dica Importante: Meça em diferentes horários para ter média realista

Fórmula:  $\lambda = \text{Número de Chegadas} \div \text{Tempo de Observação}$

Legenda:  $\lambda$  (lambda) = taxa de chegada em pessoas por unidade de tempo

## Medindo a Taxa de Atendimento ( $\mu$ )

- Como Medir: Cronometre vários atendimentos individuais
- Exemplo: 5 atendimentos duraram: 2min, 3min, 2,5min, 4min, 1,5min
- Tempo Médio:  $(2+3+2,5+4+1,5) \div 5 = 2,6$  minutos por atendimento
- Taxa  $\mu$ :  $\mu = 1 \div 2,6 = 0,38$  atendimentos/minuto por servidor
- Interpretação: Cada atendente consegue atender 0,38 pessoas por minuto

Fórmula:  $\mu = 1 \div \text{Tempo Médio de Atendimento}$

Legenda:  $\mu$  (mi) = taxa de atendimento por servidor em atendimentos por unidade de tempo

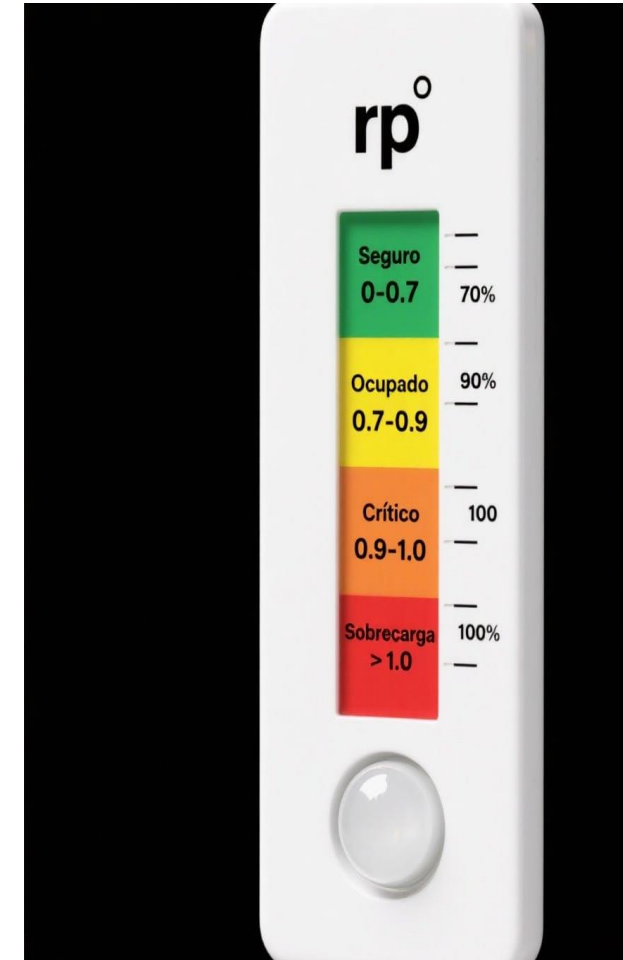
# Taxa de Ocupação ( $\rho$ ) - O Termômetro do Sistema

- $\rho < 0,7$  (Verde): Sistema tranquilo, servidores têm tempo livre
- $0,7 \leq \rho < 0,9$  (Amarelo): Sistema ocupado, servidores quase sempre trabalhando
- $\rho \geq 0,9$  (Vermelho): Sistema no limite, qualquer variação causa problemas
- $\rho \geq 1,0$  (Crítico): IMPOSSÍVEL manter! Chegam mais pessoas do que conseguimos atender

Por que  $\rho = 1,0$  é o limite?

- Se  $\rho = 1,0$ : chegadas = capacidade de atendimento (equilíbrio perfeito)
- Se  $\rho > 1,0$ : chegam mais pessoas do que conseguimos atender
- Resultado: A fila cresce sem parar, nunca conseguimos "zerar"

Fórmula:  $\rho = \lambda \div (c \times \mu)$



# Lei de Little - A Fórmula Mágica

- Conceito: Relaciona quantidade de pessoas no sistema com tempo de permanência
- Analogia: Como água em uma caixa - entrada, estoque, saída
- Aplicação Prática: Se 10 pessoas estão no sistema e cada uma fica 5 minutos...
- Resultado: Taxa de saída =  $10 \div 5 = 2$  pessoas/minuto

## Fórmulas Principais:

- $L = \lambda \times W$  (pessoas no sistema total)
- $Lq = \lambda \times Wq$  (pessoas apenas na fila)

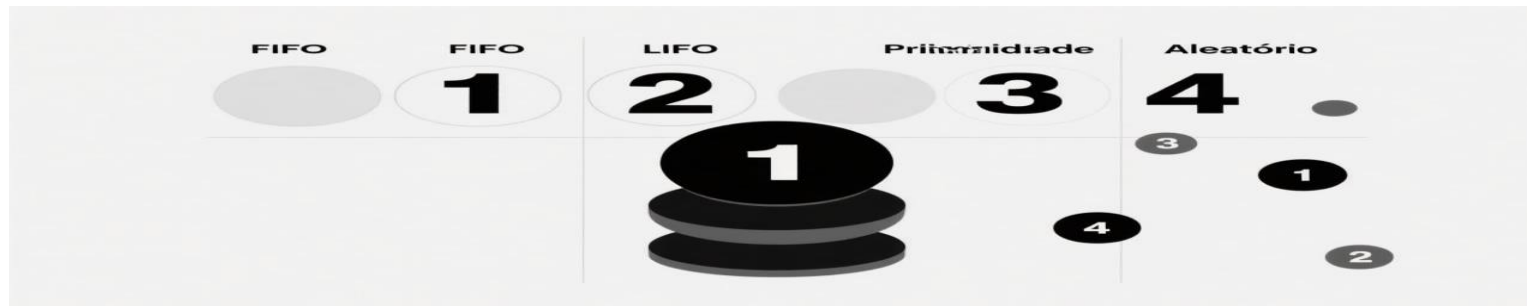
- **L** = pessoas **no sistema** (fila + atendimento) em média.
- **Lq** = pessoas **na fila** (apenas esperando) em média.
- **W** = tempo **no sistema** (espera + atendimento) em média.
- **Wq** = tempo **na fila** (só espera) em média.

Legenda: L = pessoas no sistema;  $\lambda$  = taxa de chegada ; W = tempo total no sistema;  
Lq = pessoas na fila; Wq = tempo apenas na fila;

# Disciplinas de Atendimento Comparadas

- FIFO (First In, First Out): Primeiro a chegar, primeiro a ser atendido
  - Exemplo: Fila do banco, caixa do supermercado
- LIFO (Last In, First Out): Último a chegar, primeiro a ser atendido
  - Exemplo: Pilha de documentos na mesa
- Prioridade: Atendimento por ordem de importância
  - Exemplo: Pronto socorro (casos graves primeiro)
- Aleatório: Atendimento em ordem aleatória
  - Exemplo: Sorteio de números

Mais Comum: FIFO é o padrão na maioria dos serviços





# Exercício Prático - Calculando e Entendendo

Legenda:  $\rho$  = ocupação;  $\lambda$  = chegadas/tempo;  
 $\mu$  = atendimentos/tempo por servidor;  
 $c$  = nº de servidores

Situação: Lanchonete da faculdade no horário do almoço

- Chegam 30 estudantes em 10 minutos
- Cada atendimento demora em média 2 minutos,  $\mu = 1/2$  minutos.
- Há 2 atendentes trabalhando,  $c=2$ .

Passo 1:  $\lambda = 30 \div 10 = 3$  pessoas/minuto chegando (Taxa de Chegada)

Passo 2:  $\mu = 1 \div 2 = 0,5$  atendimentos/minuto por atendente (Taxa de atendimento)

Passo 3: Capacidade total = 2 atendentes  $\times$  0,5 = 1 pessoa/minuto

Passo 4:  $\rho = \lambda \div (c \times \mu) \ggg \rho = 3 \div 1 = 3,0$  (Taxa de ocupação)

Por que está sobrecarregado?

- Chegam: 3 pessoas por minuto
- Conseguimos atender: apenas 1 pessoa por minuto
- Diferença: 2 pessoas por minuto se acumulam na fila!
- Em 10 minutos: 20 pessoas extras na fila!

# Entendendo o Problema - Visualização Prática

Cenário do Exercício Anterior:

Minuto 1: Chegam 3, atendemos 1 → Sobram 2 na fila

Minuto 2: Chegam +3, atendemos 1 → Sobram 4 na fila

Minuto 3: Chegam +3, atendemos 1 → Sobram 6 na fila

Minuto 4: Chegam +3, atendemos 1 → Sobram 8 na fila

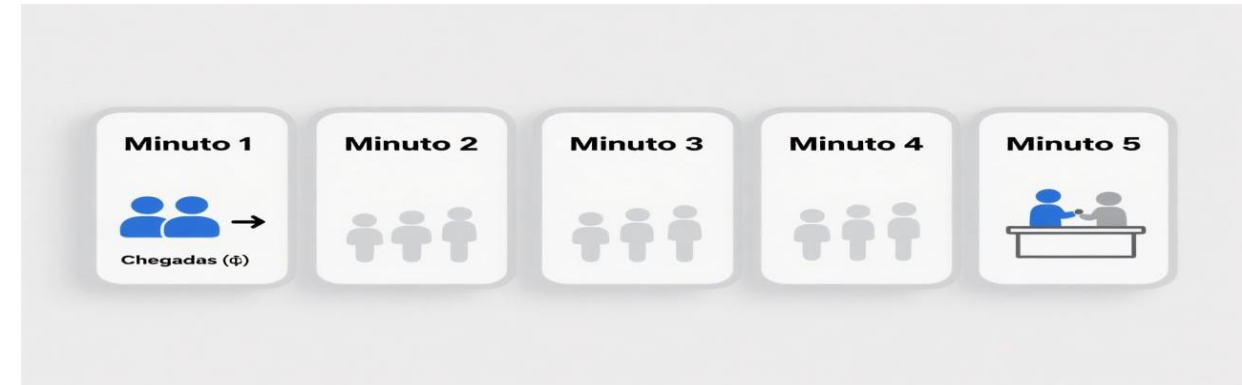
...

Minuto 10: Fila com 20 pessoas esperando!

Por isso  $\rho > 1,0$  é impossível de manter!

- A fila cresce 2 pessoas por minuto sem parar
- Nunca conseguimos "recuperar" o atraso
- Sistema entra em colapso

Solução: Precisamos que capacidade  $>$  chegadas



Legenda:  $\rho$  = ocupação;  $\lambda$  = chegadas/tempo;  
 $\mu$  = atendimentos/tempo por servidor;  $c$  = nº de servidores

# Qual Seria o Ideal? Calculando a Solução

Cenário Atual (Problemático):

- $\lambda = 3$  pessoas/min, capacidade = 1 pessoa/min
- $\rho = 3,0 \rightarrow$  Sistema colapsa

Cenário Ideal 1 - Adicionar 1 Atendente:

- $\lambda = 3$  pessoas/min,  $c = 3$  atendentes
- Capacidade =  $3 \times 0,5 = 1,5$  pessoas/min
- $\rho = 3 \div 1,5 = 2,0 \rightarrow$  Ainda sobrecarregado!

Cenário Ideal 2 - Adicionar 2 Atendentes:

- $\lambda = 3$  pessoas/min,  $c = 4$  atendentes
- Capacidade =  $4 \times 0,5 = 2,0$  pessoas/min
- $\rho = 3 \div 2,0 = 1,5 \rightarrow$  Ainda problemático!

Cenário Recomendado - 5 Atendentes Extras:

- $\lambda = 3$  pessoas/min,  $c = 7$  atendentes
- Capacidade =  $7 \times 0,5 = 3,5$  pessoas/min
- $\rho = 3 \div 3,5 = 0,86 \rightarrow$  Zona segura!

Legenda:  $\rho$  = ocupação;  $\lambda$  = chegadas/tempo;

$\mu$  = atendimentos/tempo por servidor;  $c$  = nº de servidores

Capacidade =  $c * \mu$

# Por Que $\rho = 0,8$ é Considerado Bom?

Margem de Segurança:

- $\rho = 0,8$ : Sistema usa 80% da capacidade, sobram 20%
- Por que não 100%?: Variações naturais no atendimento

Variações Reais:

- Alguns atendimentos demoram mais (pedido complicado)
- Alguns clientes fazem perguntas extras
- Atendente pode precisar de pausa
- Equipamento pode dar problema momentâneo

Exemplo Prático:

- $\rho = 0,8$ : Se capacidade cai 20%, ainda funciona
- $\rho = 0,95$ : Se capacidade cai 10%, sistema trava
- $\rho = 1,0$ : Qualquer problema causa colapso

Regra Prática: Mantenha  $\rho$  entre 0,7 e 0,85 para operação estável

# Interpretando os Resultados - Guia de Decisão

$\rho < 0,5$ : Sistema subutilizado

- Problema: Muitos atendentes ociosos, custo alto
- Ação: Considere reduzir pessoal ou expandir serviço

$0,5 \leq \rho < 0,8$ : Sistema eficiente

- Situação: Boa utilização, espera baixa
- Ação: Mantenha configuração atual

$0,8 \leq \rho < 1,0$ : Sistema no limite

- Problema: Filas longas em picos, clientes insatisfeitos
- Ação: Adicione 1 atendente ou melhore processo

$\rho \geq 1,0$ : Sistema em colapso

- Problema: Fila cresce infinitamente
- Ação: URGENTE! Adicione vários atendentes

Legenda:  $\rho$  = ocupação;  $\lambda$  = chegadas/tempo;  
 $\mu$  = atendimentos/tempo por servidor;  
 $c$  = nº de servidores

Lembre-se:  $\rho$  mede o quão "apertado" está o sistema!

# Exemplo Real - Banco no Horário de Almoço

Situação Observada:

- Período: 12h às 13h (1 hora = 60 minutos)
- Chegadas: 180 clientes em 60 minutos  $180/60 = \lambda = 3$  pessoas / minuto
- Atendentes: 3 caixas funcionando
- Tempo médio por atendimento: 2,5 minutos

Calculando:

- $\lambda = 180 \div 60 = 3$  clientes/minuto
- $\mu = 1 \div 2,5 = 0,4$  atendimentos/minuto por caixa
- Capacidade total =  $3 \times 0,4 = 1,2$  clientes/minuto
- $\rho = 3 \div 1,2 = 2,5$

Legenda:  $\rho$  = ocupação;  $\lambda$  = chegadas/tempo;  
 $\mu$  = atendimentos/tempo por servidor;  
 $c$  = nº de servidores

Resultado: Sistema sobrecarregado!

- Chegam 3 clientes/min, atendem apenas 1,2/min
- Diferença: 1,8 clientes/min se acumulam
- Em 1 hora: 108 clientes extras na fila!

# Solução para o Banco - Quantos Caixas Precisamos?

Para  $\rho = 0,8$  (ideal):  $c = \text{atendentes}$

- $\rho = \lambda \div (c \times \mu)$
- $0,8 = 3 \div (c \times 0,4)$
- $c \times 0,4 = 3 \div 0,8 = 3,75$
- $c = 3,75 \div 0,4 = 9,375$

Legenda:  $\rho$  = ocupação;  $\lambda$  = chegadas/tempo;  
 $\mu$  = atendimentos/tempo por servidor;  $c$  = nº de servidores

Arredondando: Precisamos de 10 caixas!

Verificando:

- Capacidade =  $10 \times 0,4 = 4$  clientes/minuto
- $\rho = 3 \div 4 = 0,75 \rightarrow$  Zona verde!

Comparação:

- Situação atual: 3 caixas,  $\rho = 2,5$  (colapso)
- Situação ideal: 10 caixas,  $\rho = 0,75$  (eficiente)
- Diferença: +7 caixas no horário de pico

# Comparando Cenários - 1 vs 2 Atendentes

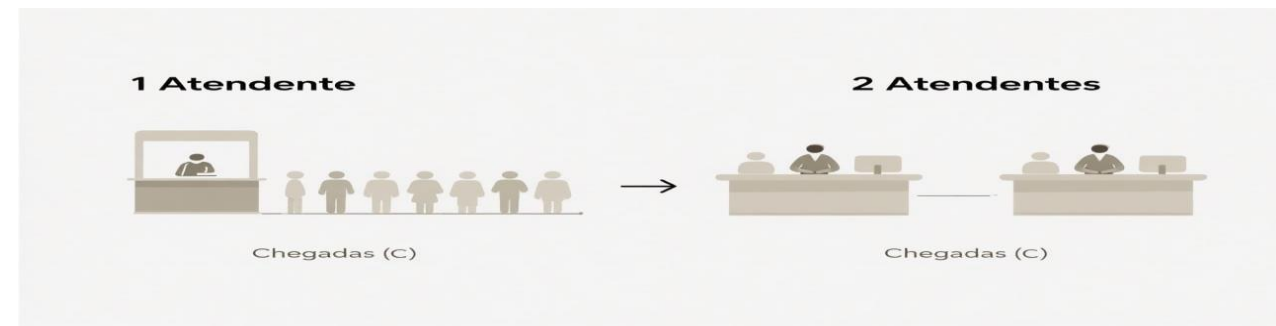
Legenda:  $\rho$  = ocupação;  $\lambda$  = chegadas/tempo;  
 $\mu$  = atendimentos/tempo por servidor;  $c$  = nº de servidores

## Cenário A - 1 Atendente:

- $\lambda = 2$  clientes/min
- $\mu = 3$  atend./min
- $c = 1$
- $\rho = 2/(1 \times 3) = 0,67$
- Resultado: Sistema estável, espera moderada, custo baixo

## Cenário B - 2 Atendentes:

- $\lambda = 2$  clientes/min
- $\mu = 3$  atend./min cada
- $c = 2$
- $\rho = 2/(2 \times 3) = 0,33$
- Resultado: Sistema subutilizado, espera muito baixa, custo alto



Conclusão: Para  $\lambda = 2$ , um atendente é suficiente!



# Variação da Demanda ao Longo do Dia

## Horários de Pico:

- 8h-9h: Abertura de contas ( $\lambda \approx 4$  pessoas/min)
- 12h-13h: Horário de almoço ( $\lambda \approx 5$  pessoas/min)
- 17h-18h: Saída do trabalho ( $\lambda \approx 3$  pessoas/min)

Legenda:  $\rho$  = ocupação;  $\lambda$  = chegadas/tempo;  
 $\mu$  = atendimentos/tempo por servidor;  
 $c$  = nº de servidores

## Horários Calmos:

- 10h-11h: Movimento baixo ( $\lambda \approx 1$  pessoa/min)
- 14h-16h: Tarde tranquila ( $\lambda \approx 1,5$  pessoas/min)

## Estratégias:

- Caixas extras nos picos
- Atendimento expresso para operações simples
- Agendamento online
- Self-service para consultas

Objetivo: Manter  $\rho < 0,9$  sempre

## Exercício de aula - Para entregar - Peso 2,0

Descrição no PDF em anexo a esta aula.

# Resumo dos Conceitos Principais

## Símbolos Importantes:

- $\lambda$  (lambda): Taxa de chegada (pessoas/tempo)
- $\mu$  (mi): Taxa de atendimento por servidor (atend./tempo)
- $c$ : Número de servidores
- $\rho$  (rô): Taxa de ocupação =  $\lambda/(c \times \mu)$

## Regras de Ouro:

- $\rho < 0,8$ : Sistema eficiente
- $\rho \geq 1,0$ : Sistema em colapso
- Capacidade total =  $c \times \mu$
- Para estabilidade: Capacidade > Chegadas

## Lei de Little:

- $L = \lambda \times W$  (pessoas no sistema)
- $Lq = \lambda \times Wq$  (pessoas na fila)

Legenda:  $L$  = pessoas no sistema;  $Lq$  = pessoas na fila;  
 $W$  = tempo total no sistema;  $Wq$  = tempo apenas na fila;  
 $\lambda$  = taxa de chegada

# Referências e material de apoio

- PIDD, Michael. Modelagem empresarial: ferramentas para tomada de decisão. Porto Alegre: Artes Médicas: Bookman, 1998. 314 p.
- PRADO, Darci; X PRADO, Darci Santos do. Teoria das filas e da simulação. Belo Horizonte, MG: Instituto de Desenvolvimento Gerencial - INDG, 1999. 122 p. (Pesquisa Operacional; 2).
- Vicente Falconi Campos. Usando o Arena em simulação, 2000. (Biblioteca Digital)
- KELTON, W. David; LAW, Averill M. Simulation modeling and analysis. 4. ed. Boston: Mc Graw Hill, 2007. 768 p.
- BANKS, Catherine M., 1960-; SOKOLOWSKI, John A., 1953-. Principles of modeling and simulation: a multidisciplinary approach . New Jersey: Wiley, 2010. xiii, 259 p. : il. ISBN 978-0-470-28943-3
- CHWIF, Leonardo; MEDINA, Afonso C. Modelagem e simulação de eventos discretos: teoria & aplicações. 2. ed. São Paulo, SP: Os Autores, c2007. 254 p.
- ZEIGLER, Bernard P.; PRAEHOFER, Herbert; KIM, Tag Gon. Theory of modeling and simulation: integrating discrete event and continuous complex dynamic systems. 2nd ed. San Diego, Califórnia: Academic Press, 2010. xxi, 510 p. ISBN 9780127784557.
- BARBETTA, Pedro Alberto; REIS, Marcelo Menezes; BORNIA, Antonio Cezar. Estatística para cursos de engenharia e informática. São Paulo, SP: Atlas, 2004. 410 p.

Thank you for your attention!!

---



Email: [andre.flores@ufn.edu.br](mailto:andre.flores@ufn.edu.br)