

Análise de uma Rede de Sistemas de Fila com Dois Servidores em Série

** Projeto Final.*

Adrian Alejandro Chavez Alanes, Everton Vilhena Cardoso

Prof. Samuel Baraldi Mafra
Instituto Nacional de Telecomunicações – INATEL

30 de junho de 2025

A modelagem de sistemas de filas é essencial para compreender e otimizar o desempenho de serviços sujeitos à variabilidade de chegada e atendimento. Este trabalho analisa uma rede com dois servidores em série: o primeiro com buffer finito ($M/M/1/K$) e o segundo com buffer infinito ($M/M/1$). Essa configuração é comum em redes de comunicação, centros de dados e linhas de produção. São avaliadas métricas como taxa de bloqueio e tempo médio de espera, considerando três cenários operacionais: sistema subdimensionado, equilibrado e com gargalo. A abordagem combina análise analítica e simulação para avaliar o impacto do dimensionamento.

- **Contexto:** Rede com dois servidores em série ($M/M/1/K \rightarrow M/M/1$), aplicável a redes de telecomunicações, linhas de produção e sistemas computacionais (Kleinrock, 1975).
- **Objetivo:** Analisar métricas de desempenho (bloqueio, vazão efetiva, ocupação média) em três cenários:
 - Subdimensionado.
 - Equilibrado.
 - Gargalo.
- **Resultados-Chave:**
 - Alto bloqueio no Servidor 1.
 - Estabilidade.
 - Crescimento ilimitado da Fila 2.
- **Contribuição:** Validação analítica-simulativa para otimização de recursos, alinhada a aplicações em QoS (Gupta & Zhang, 2019) e AQM (Smith & Li, 2020).

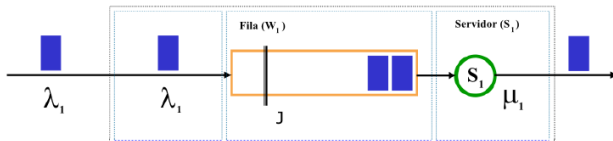
Fonte: Modelo baseado em Kleinrock (1975) e Lakken (2013).

Descrição Geral do Sistema

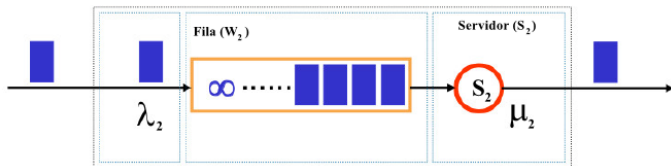
- Rede com dois servidores em série.
- Servidor 1: fila com buffer finito de tamanho J , taxa de serviço μ_1 .
- Servidor 2: fila com buffer infinito, taxa de serviço μ_2 .
- Chegadas externas seguem processo de Poisson com taxa γ_1 .
- Quando fila 1 cheia, novas chegadas são bloqueadas.
- Saída do servidor 1 alimenta a fila 2.



- Primeiro estágio: sistema **M/M/1/K**
 - Chegadas: Poisson com taxa λ_1
 - Serviço: exponencial com taxa μ_1
 - Capacidade total: $K = J + 1$



- Segundo estágio: sistema **M/M/1**
 - Chegadas: saída do servidor 1 com taxa λ_2
 - Serviço: exponencial com taxa μ_2
 - Buffer infinito



Variáveis e Eventos Principais

- Variáveis:
 - λ_1 : taxa de chegadas externas
 - K : capacidade total da fila 1
 - μ_1, μ_2 : taxas de serviço dos servidores 1 e 2
 - $n_1(t), n_2(t)$: número de clientes nas filas 1 e 2 no tempo t
- Eventos:
 1. Chegada externa: cliente entra se $n_1 < K$, se não é bloqueado
 2. Término no servidor 1: cliente vai para fila 2
 3. Término no servidor 2: cliente sai do sistema

Representação do Estado do Sistema

- Estado representado pelo par (i, j) :
 - i : clientes na fila 1 (incluindo em atendimento), $0 \leq i \leq K$
 - j : clientes na fila 2 (incluindo em atendimento), $j \geq 0$
- Evolução do sistema:
 - Chegada externa: $(i, j) \rightarrow (i + 1, j)$ se $i < K$, se não bloqueado
 - Término servidor 1: $(i, j) \rightarrow (i - 1, j + 1)$ se $i > 0$
 - Término servidor 2: $(i, j) \rightarrow (i, j - 1)$ se $j > 0$

Análise Analítica – M/M/1/K

- Fator de utilização do servidor 1:

$$\rho_1 = \frac{\lambda_1}{\mu_1}$$

- Probabilidade de ocupação de cada estado n :

$$P(n) = \frac{(1-\rho_1)\rho_1^n}{1-\rho_1^{K+1}}$$

- Probabilidade de bloqueio (fila cheia):

$$P_{bloq1} = P_1(K) = \frac{(1-\rho_1)\rho_1^K}{1-\rho_1^{K+1}}$$

- Número médio de clientes no sistema:

$$E[q_1] = \frac{\rho_1}{1-\rho_1} - \frac{(K+1)\rho_1^{K+1}}{1-\rho_1^{K+1}}$$

- Tempo médio de permanência no sistema 1:

$$E[tq_1] = \frac{E[q_1]}{(1-P_{bloq1})\lambda_1}$$

- Taxa efetiva de chegada no segundo estágio:

$$\lambda_2 = \lambda_1 (1 - P_{bloq1})$$

- Fator de utilização do servidor 2:

$$\rho_2 = \frac{\lambda_2}{\mu_2}$$

- Número médio de clientes na fila 2:

$$E[q_2] = \frac{\rho_2}{1 - \rho_2}$$

- Tempo médio de permanência no sistema 2:

$$E[tq_2] = \frac{\rho_2}{\mu_2 - \lambda_2}$$

- Foram simulados 3 cenários principais:
 - **Cenário 1 (subdimensionado)**: alta taxa de bloqueio e filas instáveis
 - **Cenário 2 (equilibrado)**: baixa perda, filas estáveis
 - **Cenário 3 (gargalo no servidor 2)**: crescimento infinito da fila 2
- Parâmetros chave:

Cenário	λ_1	μ_1	μ_2	K
1	4.5	4.0	4.5	3
2	1.5	2.0	2.0	5
3	1.5	2.0	1.0	5

Simulação - Cenário 1: Sistema Subdimensionado

- Cenário com alta demanda e servidor lento, gerando alta probabilidade de bloqueio e instabilidade nas filas.
- A ocupação das filas apresenta alta volatilidade e perdas significativas na fila 1.

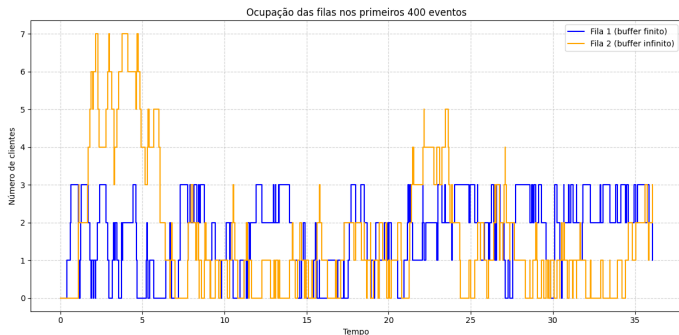


Figura: Ocupação das filas no Cenário 1, evidenciando alta taxa de bloqueio e instabilidade.

Simulação - Cenário 2: Sistema Equilibrado

- Sistema bem dimensionado, baixa probabilidade de bloqueio e filas estáveis.
- A ocupação das filas é estável, com pequenas flutuações naturais e fluxo contínuo.

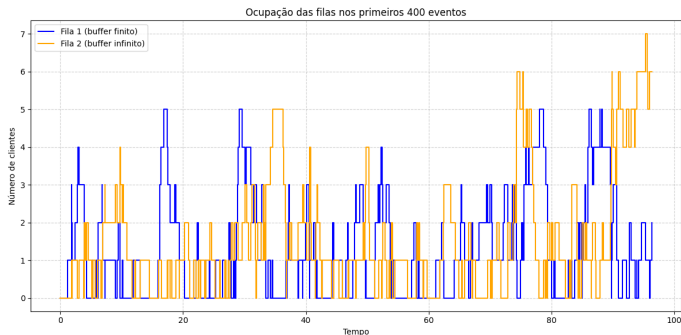


Figura: Ocupação das filas no Cenário 2, mostrando estabilidade e baixa perda.

Simulação - Cenário 3: Gargalo no Servidor 2

- Gargalo no servidor 2 provoca crescimento ilimitado da fila 2 e instabilidade do sistema.
- A fila 2 cresce continuamente, resultando em tempos médios de espera muito elevados.

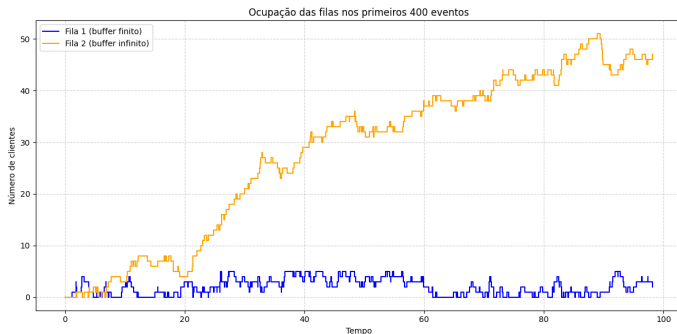


Figura: Ocupação das filas no Cenário 3, com crescimento ilimitado da fila 2 devido ao gargalo.

Resultados Analíticos e Simulados dos Cenários

Principais resultados analíticos e simulados para os três cenários avaliados.

Cenário	P_{bloq}	λ_2	ρ_2	$E[q_1]$	$E[tq_1]$	$E[q_2]$	$E[tq_2]$	$E[tq]_{total}$
<i>Resultados Analíticos</i>								
1	0.2957	3.169	0.704	1.6466	0.5195	2.3818	0.5293	1.0488
2	0.0722	1.3917	0.696	1.7009	1.2221	2.2873	1.1438	2.3659
3	0.0722	1.3917	1.3917	1.7009	1.2221	–	–	–
<i>Resultados Simulados</i>								
1	0.2960	3.169	0.704	1.6474	0.5199	2.0371	0.6428	1.1627
2	0.0740	1.3909	0.695	1.7110	1.2302	2.1057	1.5139	2.7441
3	0.0728	1.3943	1.3987	1.7119	1.2278	19896.9	14332.1	14333.4

Conclusões

Principais conclusões

- Demonstra a importância do dimensionamento para evitar bloqueios e gargalos.
- Os cenários evidenciam impacto da capacidade e taxas de serviço na estabilidade.
- E os resultados reforçam a necessidade de balanceamento para garantir qualidade de serviço.

Relevância

Então, qual é a grande contribuição do nosso trabalho? Nós demonstramos um modelo validado tanto por análise matemática quanto por simulação, que serve como uma ferramenta para a otimização de recursos.

Isso é especialmente relevante para tecnologias atuais, como *QoS* (Qualidade de Serviço), para garantir uma boa experiência em redes, e *AQM* (Gerenciamento Ativo de Filas), que são técnicas para evitar congestionamentos. Nosso trabalho se alinha com pesquisas recentes nessas áreas, como as de Gupta & Zhang e Smith & Li.

Referências



Kleinrock, L.

Queueing Systems, Volume I: Theory.

Wiley-Interscience, 1975.

ISBN: 0471491101.

Disponível em: [https:](https://www.wiley.com/en-us/Queueing+Systems%2C+Volume+I-p-9780471491101)

[//www.wiley.com/en-us/Queueing+Systems%2C+Volume+I-p-9780471491101.](https://www.wiley.com/en-us/Queueing+Systems%2C+Volume+I-p-9780471491101)



Kleinrock, L.

Queueing Systems, Volume II: Computer Applications.

Wiley-Interscience, 1976.

ISBN: 047149111X.

Disponível em: [https://www.wiley.com/en-us/Queueing+Systems%2C+Volume+2%3A+Computer+Applications-p-9780471491118.](https://www.wiley.com/en-us/Queueing+Systems%2C+Volume+2%3A+Computer+Applications-p-9780471491118)



Lakatos, L.

Queueing Theory: A Linear Algebraic Approach.

Wiley, 2013.

Disponível em:

[https://onlinelibrary.wiley.com/doi/book/10.1002/9781118625651.](https://onlinelibrary.wiley.com/doi/book/10.1002/9781118625651)