

# Dissertação de Mestrado

## **Modelagem e Otimização de Redes de Filas Finitas**

por

**Helgem de Souza Ribeiro Martins**

Orientador:

**Frederico R. B. Cruz**

Setembro de 2016

Helgem de Souza Ribeiro Martins

# Modelagem e Otimização de Redes de Filas Finitas

Dissertação de Mestrado apresentada ao Departamento de Estatística do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Mestre em Estatística.

Orientador: Frederico R. B. Cruz

Universidade Federal de Minas Gerais  
Belo Horizonte, setembro de 2016



UNIVERSIDADE FEDERAL DE MINAS GERAIS

PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

UFMG

## ATA DA DEFESA DA DISSERTAÇÃO DO ALUNO HELGEM DE SOUZA RIBEIRO MARTINS

Realizou-se, no dia 16 de setembro de 2016, às 14:00 horas, 2077 - ICEX, da Universidade Federal de Minas Gerais, a 219ª defesa de dissertação, intitulada "Modelagem e Otimização de Redes de Filas Finitas", apresentada por HELGEM DE SOUZA RIBEIRO MARTINS, número de registro 2015656167, graduado no curso de ESTATÍSTICA, como requisito parcial para a obtenção do grau de Mestre em ESTATÍSTICA, à seguinte Comissão Examinadora: Prof. Frederico Rodrigues Borges da Cruz - Orientador (DEST/UFMG), Prof. Anderson Ribeiro Duarte (DEST-UFOP), Prof. André Luiz Fernandes Cançado (DEST/UnB).

A Comissão considerou a dissertação:

Aprovada

Reprovada

Finalizados os trabalhos, lavrei a presente ata que, lida e aprovada, vai assinada por mim e pelos membros da Comissão.  
Belo Horizonte, 16 de setembro de 2016.

Prof. Frederico Rodrigues Borges da Cruz (Doutor)

Prof. Anderson Ribeiro Duarte (Doutor)

Prof. André Luiz Fernandes Cançado (Doutor)



UNIVERSIDADE FEDERAL DE MINAS GERAIS

PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

UFMG

## FOLHA DE APROVAÇÃO

**"Modelagem e Otimização de Redes de Filas Finitas"**

**HELGEM DE SOUZA RIBEIRO MARTINS**

Dissertação submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ESTATÍSTICA, como requisito para obtenção do grau de Mestre em ESTATÍSTICA, área de concentração ESTATÍSTICA E PROBABILIDADE.

Aprovada em 16 de setembro de 2016, pela banca constituída pelos membros:

Prof. Frederico Rodrigues Borges da Cruz - Orientador  
DEST/UFMG

Prof. Anderson Ribeiro Duarte  
DEST-UFOP

Prof. André Luiz Fernandes Cançado  
DEST/UnB

Belo Horizonte, 16 de setembro de 2016.

# Dedicatória

Dedico este trabalho a todos os meus amigos e familiares.

# Agradecimentos

Em primeiro lugar, à minha família, por todo apoio recebido durante toda minha formação, em especial à minha namorada *Rosielle*, uma grande companheira, que suportou com paciência meu mau humor e meu *stress* durante todo o mestrado.

Ao prof. Frederico, por todo apoio e conhecimento comigo compartilhado e por compreender as minhas limitações de tempo, sem jamais colocá-las como dificuldade à minha orientação.

Aos meus companheiros de trabalho, pelo suporte oferecido todas as vezes que tive de ausentar-me do meu posto de trabalho, além do apoio e motivação que sempre me dedicaram.

Aos membros das bancas de qualificação e de defesa da dissertação, prof. Anderson Duarte, prof. André Cançado e prof. Fernando Luiz, pela disponibilidade e pelas sugestões apresentadas ao nosso trabalho.

Um agradecimento especial ao prof. Anderson, por ser uma referência para mim, além de um grande amigo, a quem sempre pude recorrer nos momentos de dificuldade.

Ao Gabriel, pela amizade e companhia nas viagens de Ouro Preto para BH.

Ao ASSUFOP, por lutar por benefícios para os TA's da UFOP, aos quais recorri neste momento: 30h para servidores estudantes e auxílio à capacitação.

Aos colegas de mestrado, professores e servidores do PPGEST/UFMG.

# Resumo

O problema de alocação conjunta de servidores e áreas de espera (BCAP) é um problema de programação não-linear inteira que visa obter uma configuração ótima em redes de filas que garanta um limiar mínimo de desempenho pré-estabelecido. Este trabalho apresenta uma revisão detalhada da bibliografia atual que trata dos problemas de alocação de recursos em redes de filas, sobretudo do BCAP, e uma metodologia bastante eficaz empregada na solução deste problema que consiste na combinação de métodos aproximados com o algoritmo de Powell, um algoritmo de otimização livre de derivadas. Tal metodologia foi aplicada a redes nas topologias básicas série, divisão e fusão, que apresentaram resultados bastante animadores, no que diz respeito à robustez e homogeneidade das soluções.

**Palavras-chaves:** Alocação de servidores e áreas de espera, filas finitas, redes de filas, método da expansão generalizado.

# Abstract

The joint buffer and server optimization problem (BCAP) is a non-linear optimization problem with integer decision variables that optimizes the number of buffers and servers such that the resulting throughput is greater than a pre-defined threshold throughput. This work presents a detailed review of the current literature that deals with allocation problems, especially the BCAP, and a quite effective methodology in solving this problem, which consists in a combination of approximate methods and Powell algorithm, a derivative-free optimization algorithm. The methodology was applied to networks of queues in series, split, and merge basic topologies, producing very encouraging results that pointed at robust and homogeneous solutions.

**Keywords:** Buffer and server allocation, finite queues, queueing networks, generalized expansion method.



# Glossário

Algumas definições de símbolos e abreviações, usadas frequentemente ao longo deste texto, são apresentadas neste glossário. As definições estão listadas em ordem alfabética, com as letras gregas inseridas de acordo com seu nome em português.

$A/B/X/Y/Z$ : Notação usual para descrever um modelo de filas, em que  $A$  representa o padrão de chegada,  $B$ , o de serviço,  $X$ , o número de canais de serviço,  $Y$ , a capacidade total do sistema (incluindo os itens em serviço) e  $Z$ , a disciplina da fila;

$C_c$ : Custo de um servidor;

$C_B$ : Custo de uma área de espera (*buffer*);

$C_c/C_B$ : Relação entre o custo unitário dos servidores e o custo unitário das áreas de espera (em alguns pontos do texto é apresentada simplificada como  $c/B$ );

$K$ : Capacidade total (incluindo os itens em serviço) de um sistema de filas;

$\lambda_j$ : Taxa de chegada na fila  $j$ ;

$M$ : Processo markoviano (quando inserido na notação  $A/B/X/Y/Z$ , para descrever um modelo de filas);

$\mu_j$ : Taxa de serviço na fila  $j$ ;

$N$ : Número de nós na rede de filas;

$p_j$ : Probabilidade de roteamento para a fila  $j$ ;

$\rho$ : Intensidade de tráfego, definida como  $\rho \equiv \lambda/\mu$ ;

$\theta$ : Taxa de atendimento da fila;

# Índice

<b>Resumo</b>	<b>vii</b>
<b>Abstract</b>	<b>viii</b>
<b>Glossário</b>	<b>ix</b>
<b>Lista de Figuras</b>	<b>xiii</b>
<b>Lista de Tabelas</b>	<b>xiv</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Motivação . . . . .	1
1.2 Principais Contribuições . . . . .	3
1.3 Organização . . . . .	4
<b>2 O Problema de Alocação Conjunta de Servidores e Áreas de Espera (BCAP)</b>	<b>5</b>
2.1 Introdução . . . . .	5
2.2 Trabalhos Anteriores . . . . .	6
2.3 Formulação Matemática . . . . .	7
2.4 Relaxação Lagrangeana . . . . .	9
2.5 Observações Finais . . . . .	11

<b>3</b>	<b>Algoritmos para Análise de Desempenho e Otimização em Redes de Filas Finitas</b>	<b>12</b>
3.1	Introdução . . . . .	12
3.2	Análise de Desempenho em Filas Simples . . . . .	14
3.3	Método da Expansão Generalizado . . . . .	15
3.4	Algoritmo de Otimização . . . . .	23
3.5	Observações Finais . . . . .	24
<b>4</b>	<b>Resultados e Discussão</b>	<b>25</b>
4.1	Introdução . . . . .	25
4.2	Análise dos Padrões de Alocação Ótima . . . . .	28
4.3	Fatores de Influência nas Alocações Ótimas . . . . .	42
4.4	Observações Finais . . . . .	49
<b>5</b>	<b>Conclusões e Observações Finais</b>	<b>52</b>
5.1	Sumário . . . . .	52
5.2	Propostas de Continuidade . . . . .	53
	<b>Referências Bibliográficas</b>	<b>55</b>

# Lista de Figuras

1.1	Exemplo de rede de filas na produção de alimentos congelados . . .	2
2.1	Grafo representativo de uma rede de filas . . . . .	7
3.1	Método da expansão generalizado . . . . .	18
3.2	Algoritmo de Powell . . . . .	24
4.1	Rede de filas na topologia série . . . . .	26
4.2	Rede de filas na topologia divisão . . . . .	27
4.3	Rede de filas na topologia fusão . . . . .	28
4.4	Alocação de servidores e áreas de espera para topologia série . . . .	34
4.5	Alocação de servidores e áreas de espera para topologia divisão . . .	41
4.6	Alocação de servidores e áreas de espera para topologia fusão . . . .	48
4.7	Alocação total em função de $s^2$ e $C_c/C_B$ para topologia série . . . .	50

# Lista de Tabelas

3.1	Notação básica para redes de filas . . . . .	17
4.1	Resultados do BCAP para topologia série e $N = 2$ . . . . .	30
4.2	Resultados do BCAP para topologia série e $N = 3$ . . . . .	31
4.3	Resultados do BCAP para topologia série e $N = 4$ . . . . .	32
4.4	Resultados do BCAP para topologia série e $N = 5$ . . . . .	33
4.5	Resultados do BCAP para topologia divisão, $p_1 = 0,1$ e $p_2 = 0,9$ . .	36
4.6	Resultados do BCAP para topologia divisão, $p_1 = 0,2$ e $p_2 = 0,8$ . .	37
4.7	Resultados do BCAP para topologia divisão, $p_1 = 0,3$ e $p_2 = 0,7$ . .	38
4.8	Resultados do BCAP para topologia divisão, $p_1 = 0,4$ e $p_2 = 0,6$ . .	39
4.9	Resultados do BCAP para topologia divisão, $p_1 = 0,5$ e $p_2 = 0,5$ . .	40
4.10	Resultados do BCAP para topologia fusão, $\lambda_1 = 0,5$ e $\lambda_2 = 4,5$ . . .	43
4.11	Resultados do BCAP para topologia fusão, $\lambda_1 = 1,0$ e $\lambda_2 = 4,0$ . . .	44
4.12	Resultados do BCAP para topologia fusão, $\lambda_1 = 1,5$ e $\lambda_2 = 3,5$ . . .	45
4.13	Resultados do BCAP para topologia fusão, $\lambda_1 = 2,0$ e $\lambda_2 = 3,0$ . . .	46
4.14	Resultados do BCAP para topologia fusão, $\lambda_1 = 2,5$ e $\lambda_2 = 2,5$ . . .	47

# Capítulo 1

## Introdução

### 1.1 Motivação

Um dos grandes interesses ao modelar sistemas por filas é encontrar uma configuração ótima, de servidores e áreas de espera, que atenda aos requisitos de desempenho e minimize o custo envolvido (van Woensel et al., 2010). Dado que a grande maioria dos sistemas possui limitação de recursos, tanto de servidores quanto de áreas de espera, que geram impactos financeiros direto nos custos do processo, é importante seu uso racional.

Neste contexto podemos destacar o *problema de alocação conjunta de servidores e áreas de espera*, conhecido como BCAP (do inglês *Buffer and Server (c) Allocation Problem*), um desafiador problema na área de teoria das filas (van Woensel et al., 2010), cuja metodologia de solução pode ser aplicada em diversas situações reais modeladas na forma de filas ou redes de filas, tais como linhas de produção, serviços telefônicos, filas de estabelecimentos (bancos, supermercados), sistemas computadorizados, *webshops*, dentre outras aplicações análogas. Sua principal característica é a relação intrínseca existente entre o custo dos servidores e áreas de espera. Conhecer uma configuração ótima desta relação de custo permite o de-

envolvimento de redes de filas cujo serviço seja eficiente e o desempenho máximo possa ser obtido.

**Exemplo 1.1** *Uma empresa de gêneros alimentícios irá iniciar a produção de três variedades de alimentos congelados. Para tanto, ela receberá as matérias primas processadas e congeladas, que serão armazenadas em refrigeradores até o momento da produção. Após a montagem do produto, eles serão novamente congelados até a distribuição para os clientes. Para armazenar as matérias primas e os produtos finais, serão necessárias câmaras frigoríficas, cujo custo de fabricação e funcionamento aumentam conforme seu tamanho. A distribuição também é realizada por caminhões frigoríficos, que assim como as câmaras frigoríficas, são bastante caros. Qual seria a configuração de tamanho de câmara frigorífica e número de caminhões que otimizaria a produção e o custo envolvido?*

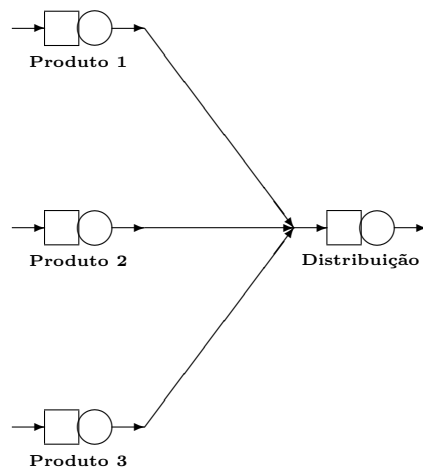


Figura 1.1: Exemplo de rede de filas na produção de alimentos congelados

No problema apresentado, uma proposta inadequada de dimensionamento das áreas de espera (câmaras frigoríficas) ou do número de servidores (linhas de montagem e caminhões frigoríficos) poderiam levar à perda significativa da eficiência e



rentabilidade do processo, dados os altos custos envolvidos. Em casos como estes, a utilização de métodos que otimizem simultaneamente o número de servidores e a capacidade das áreas de espera se faz necessária.

O *problema de alocação conjunta de servidores e áreas de espera* (BCAP) é um problema de difícil solução do ponto de vista computacional, por se tratar de um problema de programação não-linear, cuja solução analítica não apresenta forma fechada e conseqüentemente devem ser utilizados métodos aproximados em sua solução. A solução deste problema se torna ainda mais complexa quando o objeto de interesse deixa de ser um sistema de filas simples e se torna uma rede de filas complexa, como no caso do exemplo apresentado.

## 1.2 Principais Contribuições

Além da aplicabilidade já mencionada, são escassos na bibliografia especializada estudos específicos sobre o *problema de alocação conjunta de áreas de espera e servidores* (BCAP). Neste sentido, as principais contribuições deste estudo incluem:

- apresentação de uma revisão da bibliográfica recente na área de modelagem e otimização de redes de filas finitas;
- apresentação de métodos clássicos de estimação de medidas de desempenho em filas finitas configuradas em redes, em particular do consagrado *método da expansão generalizado* (ou GEM, do inglês, *Generalized Expansion Method*);
- apresentação de um modelo de programação matemática para o *problema de alocação conjunta de servidores e de áreas de espera* (ou BCAP, do inglês *Buffer and Server (c) Allocation Problem*), ainda pouco difundido na literatura da área;

- proposta do algoritmo de Powell, com múltiplos reinícios (para evitar mínimos locais), acoplado ao GEM, como ferramenta de resolução do BCAP;
- análise do BCAP via experimentos computacionais, para identificação de padrões nas soluções ótimas;
- identificação de fatores importantes nas alocações ótimas;
- discussão de tópicos para futuros trabalhos na área de alocação em filas finitas configuradas em redes.

### 1.3 Organização

No Capítulo 2 é apresentada uma revisão da literatura disponível sobre o *problema de alocação conjunta de servidores e áreas de espera* (BCAP), sendo este descrito matematicamente.

No Capítulo 3 é realizada uma revisão da literatura específica sobre redes de filas finitas e os métodos empregados na sua análise de desempenho, destacando o *método da expansão generalizado* (GEM), além da apresentação de uma proposta de otimização combinando o GEM e o algoritmo de Powell.

No Capítulo 4 são apresentados resultados para o BCAP, em redes de filas nas topologias básicas *séries*, *divisão* e  *fusão*. São também verificadas a existência de padrões de alocação para cada topologia e de fatores que impactam na alocação ótima de servidores e de áreas de espera.

Para concluir, no Capítulo 5 serão discutidos os resultados obtidos e apresentadas propostas para continuação dos estudos na área de otimização em redes de filas finitas.

## Capítulo 2

# O Problema de Alocação Conjunta de Servidores e Áreas de Espera (BCAP)

### 2.1 Introdução

Apesar de meramente ilustrativo, o Exemplo 1.1 apresenta uma cenário em que não basta efetuar a otimização isolada do número de servidores ou do tamanho das áreas de espera para que seja garantido um funcionamento ótimo da produção. Deve-se considerar a melhor configuração possível entre servidores e áreas de espera que garanta um bom funcionamento do sistema. De situações semelhantes, emerge a necessidade de solucionar o *problema de alocação conjunta de servidores e áreas de espera* (BCAP).

De maneira simplificada, o BCAP pode ser descrito como o problema de determinação de quantos servidores e qual capacidade de áreas de espera a ser disponibilizada garantirá que o desempenho do sistema, em termos da taxa de atendimento  $\theta$ , seja superior a um limiar mínimo de aceitação pré-estabelecido  $\theta_{\min}$ ,

considerando a relação de custos existente entre os servidores e as áreas de espera (representada pela razão entre os custos de servidores e áreas de espera  $C_c/C_B$ ). Tal alocação ótima será obtida através de uma configuração que minimize a probabilidade  $p_k$  de que clientes encontrem o sistema bloqueado, probabilidade esta que está diretamente vinculada à taxa de atendimento do sistema através da seguinte expressão:

$$\theta = \lambda(1 - p_k), \quad (2.1)$$

que em última análise indica que minimizar a probabilidade de bloqueio implica em maximizar a taxa de atendimento.

## 2.2 Trabalhos Anteriores

A otimização de sistemas de manufatura complexos via teoria das filas tem sido o foco de numerosos estudos nas últimas décadas, dada a grande aplicabilidade desta teoria na sua representação (Suri, 1985).

Dentre estes estudos, diversos tratam da alocação ótima de elementos que compõem tais sistemas, buscando o melhor desempenho possível. De modo geral, os principais estudos envolvem a alocação ótima de servidores (MacGregor Smith et al., 2010), áreas de espera (MacGregor Smith & Cruz, 2005; Cruz et al., 2008), roteamento ótimo (van Woensel & Cruz, 2014), taxas de serviço ou uma combinação destes elementos. Como exemplos de otimização simultânea destes elementos, podemos citar as análises conjuntas de taxas de serviço e áreas de espera (Cruz et al., 2012), roteamento e áreas de espera (Daskalaki & MacGregor Smith, 2004), número de servidores e áreas de espera (van Woensel et al., 2010), dentre outros.

Observa-se na literatura especializada que, dentre as possíveis otimizações conjuntas, a que envolve o número de servidores e tamanho das áreas de espera em

sistemas representados por filas ainda foi pouco explorada. van Woensel et al. (2010) apresentam o *problema de alocação conjunta de servidores e áreas de espera* (BCAP) em redes de filas do tipo  $M/G/c/K$ , que na notação de Kendall (1953) significa um sistema de filas com taxas de chegada markovianas,  $\Lambda$ ; tempos de serviço com distribuição geral  $G$  e taxa de atendimento  $\mu$  e coeficiente de variação  $s^2$ ,  $c$  servidores em paralelo e capacidade total de  $K$  usuários na fila, incluindo aqueles que estão sendo atendidos. Em Cruz & van Woensel (2014b) é apresentada uma revisão sobre modelagem e otimização em redes de filas finitas, incluindo discussão e resultados sobre o BCAP.

Em ambos trabalhos citados, a metodologia de solução do BCAP se apresenta na forma de problema de otimização cuja solução envolve técnicas de programação inteira não-linear.

## 2.3 Formulação Matemática

Defina uma rede de filas como um grafo  $G = (N, A, \mathbf{p})$ , em que  $N$  é o conjunto de todos os nós (filas) que compõe a rede,  $A$  é o conjunto de arcos que interconectam pares de nós e  $\mathbf{p}$  é o vetor das respectivas probabilidades de roteamento nos arcos.

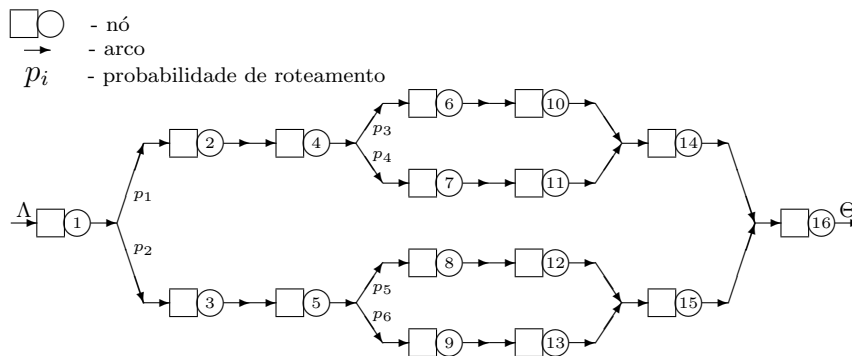


Figura 2.1: Grafo representativo de uma rede de filas

O BCAP busca otimizar o número de áreas de espera e servidores de forma que a taxa de atendimento resultante seja maior que um limiar predefinido de taxa de atendimento, e pode ser definido matematicamente pela seguinte formulação de programação não-linear, conforme apresentado por van Woensel et al. (2010).

(BCAP):

$$Z = \min \left[ \sum_{\forall i \in N} \omega_i c_i + \sum_{\forall i \in N} (1 - \omega_i) B_i \right], \quad (2.2)$$

sujeito a

$$\theta(\mathbf{c}, \mathbf{B}) \geq \theta_{\min}, \quad (2.3)$$

$$c_i \in \{1, 2, \dots\}, \forall i \in N, \quad (2.4)$$

$$B_i \in \{0, 1, \dots\}, \forall i \in N, \quad (2.5)$$

em que  $c_i$  é o número de servidores no nó  $i$ ,  $B_i$  representa o tamanho das áreas de espera puras, ou seja, excluindo aqueles que encontram-se em atendimento (a capacidade total do nó  $i$  é representada por  $K_i = c_i + B_i$ ),  $\theta(\mathbf{c}, \mathbf{B})$  é a taxa de saída resultante, que é dada em função dos servidores ( $\mathbf{c}$ ) e das áreas de espera ( $\mathbf{B}$ ),  $\theta_{\min}$  é o limiar mínimo aceitável para a taxa de atendimento resultante e  $\omega_i$  é uma variável que representa o custo relativo entre servidores e áreas de espera ( $0 \leq \omega_i \leq 1$ ).

O valor de  $\omega_i$  pode ser alterado para refletir o custo de servidores em comparação ao custo das áreas de espera. Quando o valor de  $\omega_i$  diminui, o custo dos servidores se torna relativamente menor se comparado ao custo das áreas de espera. Em contrapartida, se o valor de  $\omega_i$  sofre acréscimos, os servidores se tornam relativamente mais caros se comparados às áreas de espera. Oscilando os valores

possíveis de  $\omega_i$ , é possível verificar as variações existentes nas alocações de servidores e áreas de espera e efetuar uma busca por padrões existentes nas alocações de recursos resultantes. Outra possibilidade que emerge da relação de custos é a verificação de quais valores de  $\omega_i$  tornam significativas as alterações em alocações de servidores e áreas de espera. Se definirmos  $\omega_i = 0, \forall i \in N$ , o BCAP se reduz ao *problema de alocação de áreas de espera* (ou BAP, do inglês *Buffer Allocation Problem*). Por outro lado, se definirmos  $\omega_i = 1, \forall i \in N$ , obtém-se o *problema de alocação de servidores* (ou CAP, do inglês *Server (c) Allocation Problem*).

É importante observar a existência de soluções que não apresentem áreas de espera (conhecidos, em inglês, como sistemas *zero-buffer* ou *bufferless*), ou seja, em determinados nós pode ocorrer  $B_i = 0$ . Redes de filas sem áreas de espera são observadas em diversos sistemas de produção reais e em determinados ambientes eles são estritamente necessários, devido às características do processo em si ou simplesmente pela ausência de capacidade de armazenamento entre duas etapas consecutivas de operação em um processo produtivo. Mais detalhes sobre redes de filas sem áreas de espera podem ser obtidos no trabalho de Andriansyah et al. (2010).

Além das restrições apresentadas, uma restrição adicional é necessária para garantir a existência de uma solução ótima finita:

$$\rho_i \equiv \frac{\lambda_i}{c_i \mu_i} < 1, \quad (2.6)$$

em que  $\mu_i$  é a taxa média de serviço no servidor  $i$ .

## 2.4 Relaxação Lagrangeana

O BCAP, definido pelas expressões (2.2), (2.3), (2.4) e (2.5), apresenta restrições complexas, sobretudo a representada pela expressão (2.3).

A incorporação das restrições mais complexas na função objetivo via relaxação lagrangeana têm sido aplicada com sucesso em problemas de alocação, como o BAP (Cruz et al., 2008) e o CAP (MacGregor Smith et al., 2010). Assim, tal metodologia se torna uma alternativa viável na redução da complexidade do problema em estudo. Uma apresentação detalhada do método pode ser encontrada no trabalho de Lemaréchal (2007).

De forma simplificada, a relaxação lagrangeana é uma técnica que consiste em relaxar as restrições complexas e incorporá-las diretamente na função objetivo na forma de uma penalidade. No BCAP, a restrição complexa da Eq. (2.3) pode ser relaxada em termos da variável lagrangeana  $\alpha > 0$ , resultando no BCAP relaxado, ou RBCAP (do inglês *Relaxed Buffer and Server (c) Allocation Problem*), cuja formulação é apresentada a seguir.

(RBCAP):

$$Z_\alpha = \min \left[ \sum_{\forall i \in N} \omega_i c_i + \sum_{\forall i \in N} (1 - \omega_i) B_i + \alpha \left( \theta_{\min} - \theta(\mathbf{c}, \mathbf{B}) \right) \right], \quad (2.7)$$

sujeito às restrições (2.4) e (2.5).

É importante notar que na formulação RBCAP, o termo  $\alpha \left( \theta_{\min} - \theta(\mathbf{c}, \mathbf{B}) \right)$  é sempre não-positivo, para qualquer solução factível da formulação original do BCAP. Ou seja, se as restrições (2.3), (2.4) e (2.5) forem satisfeitas, elas devem garantir que  $\alpha \left( \theta_{\min} - \theta(\mathbf{c}, \mathbf{B}) \right) \leq 0$ , e assim, temos  $Z_\alpha \leq Z$ . Consequentemente,  $Z_\alpha$  será utilizado como uma cota inferior para o valor ótimo de  $Z$ . Espera-se que tal cota inferior esteja o mais próximo possível do verdadeiro valor de  $Z$ .

Vale salientar que  $\alpha$  fornece uma penalidade para os casos nos quais a restrição não for atendida. Com base em resultados presentes em Cruz et al. (2008), o valor da variável lagrangeana  $\alpha$  foi fixado em  $10^3$ . Definido o modelo de otimização para a solução do RBCAP, será estabelecido o limiar mínimo  $\theta_{\min}$  igual à taxa de chegada  $\lambda$  e utilizada uma aproximação em dois momentos aliada ao *método da*



*expansão generalizado* (GEM), a ser detalhado na Seção 3.3, pág. 15, para obtenção das probabilidades de bloqueio  $p_k$ , necessárias ao cálculo aproximado da taxa de saída  $\theta(\mathbf{c}, \mathbf{B})$ , para uma particular configuração de servidores e tamanho de áreas de espera.

## 2.5 Observações Finais

Neste capítulo foi apresentada uma breve revisão da literatura sobre problemas de otimização em teoria das filas, destacando os problemas de alocação de servidores (CAP), de alocação de áreas de espera (BAP) e de alocação conjunta de servidores e áreas de espera (BCAP).

O BCAP foi definido matematicamente e suas principais características foram exploradas. Para finalizar, a técnica da relaxação lagrangeana foi brevemente descrita e sua aplicação na solução do BCAP discutida, culminando na forma relaxada do problema de alocação conjunta de servidores e áreas de espera, denominada RB-CAP.

## Capítulo 3

# Algoritmos para Análise de Desempenho e Otimização em Redes de Filas Finitas

### 3.1 Introdução

Conforme descrito anteriormente, a solução do BCAP está diretamente vinculada ao desempenho da rede de filas para a qual deseja-se obter a solução ótima e a configuração do número de servidores e tamanho das áreas de espera em cada nó da referida rede. Para tal, serão necessários resultados sobre obtenção de medidas de desempenho em filas e redes de filas finitas.

Diversas são as ferramentas existentes para a análise de desempenho em redes de filas. Dentre elas, destacam-se os *métodos forma-produto*, *métodos numéricos*, *métodos aproximados* e *simulações*.

Os *métodos forma-produto* evitam efetuar a análise do sistema de uma só vez, decompondo o sistema em pares ou triplas de nós e os analisando individualmente (maiores detalhes em Perros, 1984). Cada nó decomposto é tratado como um

provedor independente de serviço para os quais os resultados de filas simples podem ser utilizados. Jackson (1957) demonstra que a distribuição conjunta da rede completa pode ser fatorada no produto das distribuições marginais de cada um dos nós sob determinadas condições de regularidade (por exemplo, tempos de chegada e atendimento exponencialmente distribuídos). Técnicas de decomposição geram resultados exatos para redes de filas com soluções na forma-produto. Para redes que não apresentam soluções na forma-produto, a técnica da decomposição fornece boas aproximações (Balsamo et al., 2001).

*Métodos numéricos* são bastante úteis, pois teoricamente podem ser utilizados na solução de qualquer problema de natureza markoviana. Sua desvantagem é o custo computacional envolvido, que cresce exponencialmente com o aumento da rede de filas, seu número de nós, número de clientes e áreas de espera. Assim, métodos numéricos são utilizados preferencialmente em redes de tamanho reduzido (Balsamo et al., 2001).

Entre os *métodos aproximados*, os métodos de decomposição são bastante populares. Estes métodos são considerados aproximados devido às sub-redes serem apenas partes da rede e como tais, não apresentam necessariamente o mesmo comportamento. Considerando o custo computacional na obtenção de soluções exatas, a utilização de tais métodos são absolutamente justificáveis, sendo seu principal desafio se aproximar ao máximo dos valores exatos. Sua precisão pode ser verificada utilizando soluções numéricas (em redes pequenas) ou através de simulações. A ideia central dos métodos de decomposição é tentar generalizar as ideias de independência e das soluções na forma-produto para redes Jackson para sistemas mais genéricos. Reiser & Kobayashi (1974) e Kuehn (1979) foram pioneiros nestes métodos.

Outra metodologia amplamente empregada na obtenção de medidas de desempenho para redes de filas são os *métodos de simulação*. Diversos são os estudos

bem sucedidos baseados em tais métodos, dos quais pode-se destacar os resultados apresentados em Cruz et al. (2005; 2010); Dorda & Teichmann (2013), dentre diversos outros trabalhos.

Uma ferramenta bastante empregada conjuntamente com as metodologias para análise de desempenho anteriormente descritas são os métodos de otimização numérica. A otimização é aplicada aos modelos de programação matemática para resolução de problemas nos quais deseja-se minimizar ou maximizar uma função objetivo através de uma escolha sistemática de valores para suas variáveis dentre os valores pertencentes ao seu conjunto de possíveis valores. Dentre os métodos de otimização, os não-lineares destacam-se na solução de problemas de otimização em redes de filas, devido à natureza da função objetivo, ou das restrições existentes nos modelos de otimização, geralmente não-lineares. Maiores detalhes sobre otimização podem ser encontrados no livro de Zhang et al. (2015).

## 3.2 Análise de Desempenho em Filas Simples

Uma das medidas mais habitualmente utilizadas na análise de desempenho de filas é a sua taxa de atendimento (ou *throughput*, em inglês), representada por  $\theta$ . Quando se trata de filas únicas finitas, tal taxa está diretamente relacionada com a taxa de entrada da fila e da sua probabilidade de bloqueio,  $p_k$ , sendo o bloqueio o momento em que o número de entidades presentes na fila preenche toda sua capacidade  $K$ . O desempenho da fila, em termos de taxa de atendimento, é dado pela expressão que se segue.

$$\theta = \lambda(1 - p_k). \quad (3.1)$$

A Eq. (3.1) indica que, conhecendo a taxa de entrada da fila, a obtenção da taxa de atendimento fica condicionada à determinação do valor da sua probabilidade

de bloqueio  $p_k$ . Assim, definindo-se um limiar mínimo para a taxa de saída,  $\theta_{\min}$ , é possível expressar a probabilidade de bloqueio  $p_k$  a partir da seguinte expressão:

$$p_k \leq 1 - \frac{\theta_{\min}}{\lambda}. \quad (3.2)$$

Como a taxa de atendimento  $\theta_{\min}$  é necessariamente menor ou igual à taxa de entrada na fila  $\lambda$  ( $\theta_{\min} \leq \lambda$ ), analisando a Eq. (3.2), pode-se concluir que quando  $\theta_{\min} \rightarrow 0$  então  $p_k \rightarrow 1$ . Em contrapartida, quando  $\theta_{\min} \rightarrow \lambda$ , temos  $p_k \rightarrow 0$ . Considerando tais fatos, para obter um desempenho ótimo da fila, deve-se buscar configurações de servidores e tamanhos de áreas de espera que garantam que a taxa mínima de atendimento  $\theta_{\min}$  seja próxima o suficiente da taxa de entrada  $\lambda$ , para que sejam minimizadas as chances de que clientes sofram bloqueio durante o processo de atendimento na fila.

Assim como no caso de filas únicas, uma *rede* de filas finitas também tem a taxa de saída  $\theta$  como um das mais populares medidas de desempenho. Será apresentado a seguir um método bastante eficaz e eficiente para determinação da taxa de saída em filas finitas configuradas em redes acíclicas.

### 3.3 Método da Expansão Generalizado

Conforme discutido previamente, o desempenho de uma rede de filas geralmente é avaliado considerando sua taxa de atendimento  $\theta$ . Para a obtenção de tal medida de desempenho, a rede de filas deve ser ‘resolvida’, ou seja, devem ser encontrados valores para os parâmetros de interesse que levem a rede de filas ao desempenho esperado. Uma ferramenta bastante utilizada na avaliação de tal desempenho e obtenção dos parâmetros de interesse é o *método da expansão generalizado*, que a partir deste ponto, por simplicidade, será referido como GEM (do inglês, *Generalized Expansion Method*).

A ideia central do GEM é transformar a rede de filas de forma que ela equivalha a uma rede do tipo Jackson, que poderá ser decomposta e ter seus nós solucionados de forma independente dos demais. Assim a solução seria similar a uma solução do tipo forma-produto. Enquanto um método para avaliação de desempenho de redes de filas altamente efetivo e robusto, o GEM foi tema de uma série de estudos que comprovaram tais características, incluindo os trabalhos de Kerbache & MacGregor Smith (1987; 1988; 2000), Jain & MacGregor Smith (1994) e Andriansyah et al. (2010), dentre outros.

O GEM utiliza o bloqueio após o serviço (ou BAS, do inglês *Blocking After Service*), ou seja, ele avalia a probabilidade de um cliente encontrar o próximo nó (fila) completo após ter sido atendido por um nó anterior. Este tipo de bloqueio é comum na maioria dos sistemas de produção e manufatura (Dallery & Gershwin, 1992), transportes (van Woensel & Vandaele, 2006; 2007) e outros sistemas similares.

Apresentado em Kerbache & MacGregor Smith (1987), o GEM se tornou um método bastante utilizado para análise de desempenho em redes de filas, devido à sua precisão e relativa simplicidade, além da restrição existente no número de soluções exatas para tal problema, de modo geral restritas apenas a redes de filas bastante simplificadas e também a complexidade dos métodos simulados anteriores, que exigiam grande esforço computacional.

De forma resumida, o GEM é basicamente uma combinação de dois métodos aproximados que envolve tentativas repetidas e decomposição nó a nó da rede de filas. Para efetuar a avaliação do desempenho da fila, o método primeiramente divide a rede em nós simples, com tempos de serviço e parâmetros de chegada revisados. Clientes que encontram um nó bloqueado são registrados em um nó de espera artificial e são reenviados a este nó até que sejam atendidos. A adição deste nó de espera transforma a rede de filas em uma rede de comportamento

equivalente a redes do tipo Jackson, onde cada nó pode ser solucionado de forma independente.

Tabela 3.1: Notação básica para redes de filas

Variável	Descrição
$\Lambda$	taxa de chegada externa Poisson na rede;
$\lambda_j$	taxa de chegada Poisson no nó $j$ ;
$\tilde{\lambda}_j$	taxa efetiva de chegada no nó $j$ ;
$\mu_j$	taxa média de atendimento no nó finito $j$ ;
$\tilde{\mu}_j$	taxa efetiva de serviço no nó $j$ devido ao bloqueio;
$p_{K_j}$	probabilidade de bloqueio no nó finito $j$ de capacidade $K$ ;
$p'_{K_j}$	probabilidade de novo bloqueio no método da expansão generalizado;
$h_j$	nó artificial de espera criado pelo GEM antes do nó $j$ ;
$c_j$	número de servidores no nó $j$ ;
$K_j$	capacidade total do nó $j$ , incluindo os itens em atendimento;
$B_j \equiv K_j - c_j$	capacidade do nó $j$ excluindo os itens em serviço;
$N$	conjunto de nós(filas) da rede de filas;
$V$	conjunto de arcos(pares de nós) da rede de filas;
$\rho_j \equiv \lambda_j / (c_j \mu_j)$	intensidade de tráfego no nó $j$ ;
$\theta_j$	taxa média de atendimento no nó $j$ ;
$s_j^2$	quadrado do coeficiente de variação da distribuição dos tempos de serviço no nó $j$ .

Existem três etapas principais na implementação do GEM, denominadas *reconfiguração da rede*, *estimação dos parâmetros* e *eliminação de retroalimentação*. Para uma melhor compreensão do método, a Tabela 3.1 apresenta a notação básica utilizada na descrição do método.

A seguir, serão apresentadas as três etapas que compõem o método da expansão generalizado.

### 3.3.1 Reconfiguração da Rede

Inicialmente, para cada nó finito que compõe a rede de filas, é criado um nó artificial cuja função será registrar os clientes bloqueados. A introdução de tais nós implica na criação de novas probabilidades de roteamento em toda a rede de filas. O resultado de tal reconfiguração pode ser verificado na Figura 3.1.

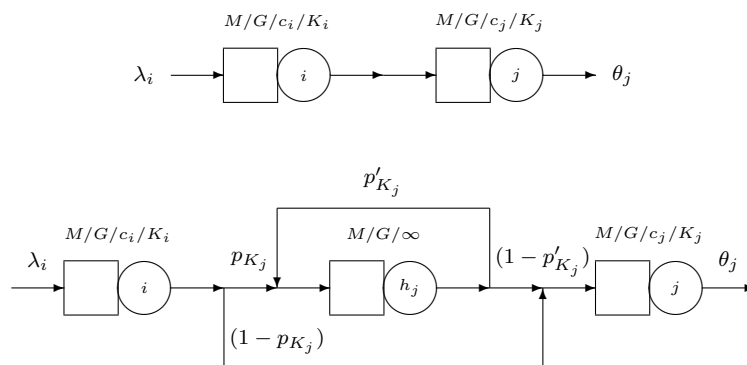


Figura 3.1: Método da expansão generalizado

Clientes atendidos no nó  $i$  podem encontrar o nó  $j$  em dois possíveis estados: *saturado*, quando o nó  $j$  encontra-se repleto e o novo cliente não poderá ser atendido, e *não-saturado*, quando o novo cliente encontra o sistema capacitado a lhe receber. O cliente encontrará o sistema não-saturado com probabilidade  $(1 - p_{k_j})$  e conseqüentemente acessará a fila e será atendido. Porém, se o cliente encontrar o nó saturado, com probabilidade  $p_{k_j}$  de encontrá-lo em tal estado, então ele será direcionado ao nó artificial de espera  $h_j$ , onde será atrasado. O atraso no nó artificial  $h_j$  é modelado utilizando uma fila do tipo  $M/G/\infty$ , representando um atraso no tempo sem filas. Após o atraso, o cliente tentará novamente acessar a fila finita, com probabilidade de reingresso  $(1 - p'_k)$ . Conseqüentemente, existe uma probabilidade  $p'_k$  de que o nó ainda esteja saturado após o tempo de atraso. Neste caso, o cliente será novamente direcionado ao nó artificial  $h_j$ . Este processo será repetido até que o cliente consiga reingressar na fila.



### 3.3.2 Estimação dos Parâmetros

No estágio de estimação dos parâmetros, os parâmetros de interesse  $p_k$ ,  $p'_k$  e  $\mu_h$  serão determinados. Por simplificação, nesta subseção o índice  $j$  será omitido.

#### Estimação de $p_k$

A determinação de  $p_k$  deve ser realizada utilizando fórmulas analíticas exatas sempre que possível (Kerbache & MacGregor Smith, 2000). Para os casos nos quais as fórmulas para obtenção de  $p_k$  não estão disponíveis, aproximações para o valor de  $p_k$  em filas do tipo  $M/G/c/K$  são apresentadas em MacGregor Smith (2003). Tais aproximações são baseadas em expressões fechadas derivadas de expressões para filas do tipo exponencial com capacidade finita ( $M/M/c/K$ ) usando a aproximação a dois momentos de Kimura (1996). Como exemplo, a fórmula para  $p_k$  em uma fila  $M/G/2/K$  é apresentada a seguir.

$$p_K = \frac{2\rho^2 \left( \frac{\sqrt{\frac{\rho}{e}} s^2 - \sqrt{\frac{\rho}{e} + K}}{2 + \sqrt{\frac{\rho}{e}} s^2 - \sqrt{\frac{\rho}{e}}} \right) (2\mu - \lambda)}{-2\rho^2 \left( \frac{\sqrt{\frac{\rho}{e}} s^2 - \sqrt{\frac{\rho}{e} + K}}{2 + \sqrt{\frac{\rho}{e}} s^2 - \sqrt{\frac{\rho}{e}}} \right) \lambda + 2\mu + \lambda}. \quad (3.3)$$

As equações para obtenção de  $p_k$ , em filas  $M/G/c/K$ , para valores de  $c = 3, 4, \dots$ , não serão apresentadas neste texto, porém estão disponíveis para consulta no trabalho de MacGregor Smith (2003).

#### Estimação de $p'_k$

Não há método exato a ser empregado no cálculo de  $p'_k$ . Para estimação do valor de tal parâmetro, é utilizada uma aproximação proposta por Labetoulle & Pujolle (1980), que é baseada em técnicas de difusão. A expressão é dada por

$$p'_K = \left[ \frac{\mu_j + \mu_h}{\mu_h} - \frac{\lambda [(r_2^K - r_1^K) - (r_2^{K-1} - r_1^{K-1})]}{\mu_h [(r_2^{K+1} - r_1^{K+1}) - (r_2^K - r_1^K)]} \right]^{-1}, \quad (3.4)$$

em que  $r_1$  e  $r_2$  são as raízes do polinômio

$$\lambda - (\lambda + \mu_h + \mu_j)x + \mu_h x^2 = 0, \quad (3.5)$$

no qual  $\lambda = \lambda_j - \lambda_h(1 - p'_k)$ ,  $\lambda_j$  e  $\lambda_h$  são as taxas de chegada ao nó finito  $j$  e ao nó artificial  $h_j$ , respectivamente. Além disto, pode-se argumentar que

$$\lambda_j = \tilde{\lambda}_i(1 - p_K) = \tilde{\lambda}_i - \lambda_h. \quad (3.6)$$

### Estimação de $\mu_h$

A distribuição do tempo de atraso no nó de espera artificial  $h$  nada mais é do que o tempo de serviço restante no nó finito  $j$ . Baseado na teoria da renovação, pode-se formular o tempo remanescente de serviço como a seguinte taxa  $\mu_h$ :

$$\mu_h = \frac{2\mu_j}{1 + \sigma_j^2 \mu_j^2}, \quad (3.7)$$

em que  $\sigma_j^2$  é a variância do tempo de serviço no nó finito  $j$ . Note que, se o tempo de serviço no nó  $j$  for exponencialmente distribuído com taxa  $\mu_j$ , devido à propriedade da falta de memória da distribuição exponencial, pode-se concluir que

$$\mu_h = \mu_j. \quad (3.8)$$

### 3.3.3 Eliminação da Retroalimentação

O laço de retroalimentação que ocorre no nó de espera tem como resultado a criação de uma forte dependência no processo de chegada. Para eliminar tal dependência, a taxa de atendimento no nó de espera  $h$  deverá sofrer um reajuste, a partir da seguinte relação:

$$\mu'_h = (1 - p'_K)\mu_h. \quad (3.9)$$

ou seja, a taxa de atendimento ajustada será proporcional à probabilidade de que o cliente consiga acessar a fila após ser registrado no nó de espera. Como consequência, a taxa de serviço no nó  $i$  que antecede o nó finito  $j$  também será afetada. É fácil ver que a taxa média de serviço no nó  $i$  é dada por  $\mu_i^{-1}$  quando o nó finito não encontra-se saturado e vale  $\mu_i^{-1} + \mu_h'^{-1}$  quando o nó finito está saturado. Assim, em média, a taxa média de serviço no nó  $i$  que precede o nó finito  $j$  é dada por:

$$\mu_i^{-1} = \mu_i^{-1} + p_K \mu_h'^{-1},$$

As etapas descritas acima deverão ser aplicadas a todos os nós presentes na rede de filas cujo desempenho será avaliado.

### 3.3.4 Sumário

Para sumarizar o GEM, a obtenção de todas as medidas de desempenho da rede de filas pode ser realizada solucionando o seguinte conjunto de equações simultaneamente para todos os nós da rede:

$$\lambda = \lambda_j - \lambda_h(1 - p'_K), \quad (3.10)$$

$$\lambda_j = \tilde{\lambda}_i(1 - p_K), \quad (3.11)$$

$$\lambda_j = \tilde{\lambda}_i - \lambda_h, \quad (3.12)$$

$$\lambda_j = \tilde{\lambda}_i - \lambda_h, \quad (3.13)$$

$$p'_K = \left[ \frac{\mu_j + \mu_h}{\mu_h} - \frac{\lambda[(r_2^K - r_1^K) - (r_2^{K-1} - r_1^{K-1})]}{\mu_h[(r_2^{K+1} - r_1^{K+1}) - (r_2^K - r_1^K)]} \right]^{-1}, \quad (3.14)$$

$$z = (\lambda + 2\mu_h)^2 - 4\lambda\mu_h, \quad (3.15)$$

$$r_1 = \frac{[(\lambda + 2\mu_h) - z^{\frac{1}{2}}]}{2\mu_h}, \quad (3.16)$$

$$r_2 = \frac{[(\lambda + 2\mu_h) + z^{\frac{1}{2}}]}{2\mu_h}, \quad (3.17)$$

$$p_K = f(M/G/c/K), \quad (3.18)$$

Expressões  $f(M/G/c/K)$ , ou seja, para obtenção de  $p_K$ , para filas  $M/G/c/K$ , com  $c = 3, 4, \dots$ , foram desenvolvidas por MacGregor Smith (2003) e podem ser utilizadas no conjunto de equações simultâneas, Eq. (3.10)–(3.18), por conveniente substituição na Eq. (3.18), para a configuração de filas de interesse. Observe que a Eq. (3.3), pág. 19, somente aplica-se às filas  $M/G/2/K$ .

Recapitulando o GEM, primeiramente a rede é expandida, seguido pela aproximação das probabilidades de roteamento, devido ao bloqueio e à taxa de atraso no nó de espera. Após estas etapas, o nó de espera é eliminado e, completando-se tais estágios, obtém-se a rede expandida, que será então utilizada na obtenção das medidas de desempenho para a rede original. É importante ressaltar que este processo não modifica fisicamente a rede de filas, sendo todo o processo de expansão realizado via *software*. A taxa de atendimento no nó  $i$  que é sucedido pelo nó  $j$ , denotada por  $\theta_i$ , é obtida a partir da seguinte expressão:

$$\theta_i = \lambda_i(1 - p_k). \quad (3.19)$$

Note que a probabilidade de bloqueio  $p_k$  no nó  $i$  é uma função complexa da intensidade de tráfego  $\rho$ , dentre outros parâmetros, os quais são dependentes da taxa de atendimento corrigida  $\tilde{\mu}_i$ , obtida na última etapa do GEM. Finalmente, a taxa de atendimento geral da rede de filas é obtida pela soma de todas as taxas de atendimento obtidas no(s) último(s) nó(s) da rede de filas.

### 3.4 Algoritmo de Otimização

Para otimizar o RBCAP, definido pela função objetivo, Eq. (2.7), sujeito às restrições, Eq. (2.4) e (2.5), será utilizado o método de Powell, um algoritmo clássico de otimização não-linear que não utiliza derivadas, em conjunto com o GEM, que calculará as medidas de desempenho de interesse. Neste caso, será a taxa de atendimento  $\theta$ .

O método de Powell, apresentado de forma detalhada por Himmelblau (1972), localiza o mínimo de uma função não linear  $f(x)$  através de sucessivas buscas unidimensionais a partir de um ponto de partida inicial  $x_{(0)}$  ao longo de um conjunto de direções conjugadas. Estas direções conjugadas são geradas dentro do próprio procedimento. O método de Powell é baseado na ideia de que se um mínimo de uma função não-linear  $f(x)$  é observado ao longo de  $p$  direções conjugadas em algum estágio da busca, um passo apropriado deverá ser dado em cada direção. O passo geral do início do procedimento até o  $p$ -ésimo sub-passo é conjugado em todas as  $p$  sub-direções. A Figura 3.2 descreve o algoritmo de otimização irrestrita de Powell, que foi utilizado nos experimentos que serão apresentados na sequência.

Alguns cuidados adicionais devem ser tomados quando se utiliza o método de Powell conjuntamente com o GEM na solução do problema de alocação de servidores e áreas de espera, pois o algoritmo de Powell é um processo de busca irrestrito. Assim, a menos que a busca seja controlada, alguns vetores de soluções que podem violar as restrições impostas pelo problema poderão ser considerados, quando na verdade tais soluções não são desejáveis. Além deste cuidado, para que sejam evitados pontos ótimos locais, o algoritmo foi reiniciado diversas vezes (20 vezes) com diferentes pontos de partida aleatoriamente selecionados.

Além das vantagens supracitadas, o algoritmo de Powell foi utilizado devido à sua disponibilidade em Himmelblau (1972) e compatibilidade com o código do GEM em FORTRAN fornecido pelos seus criadores (Kerbache & MacGregor

```

algoritmo
  leia  $G(N, A, \mathbf{p}), \boldsymbol{\lambda}, \boldsymbol{\mu}$  e  $\mathbf{x}^{(0)}$ 
  /* escolha direções de busca */
  escolha  $\mathbf{d}^{(1)}, \dots, \mathbf{d}^{(n)}$ 
   $\mathbf{x}^{(\text{opt})} \leftarrow \mathbf{x}^{(0)}$ 
repita
   $\mathbf{x}^{(1)} \leftarrow \mathbf{x}^{(\text{opt})}$ 
  para  $i = 1$  até  $n$  faça
    /* busca unidimensional */
    /* calcule  $f(\bullet)$  via GEM */
     $\mathbf{x}^{(i+1)} \leftarrow \underset{\gamma \in \mathcal{R}}{\text{argmin}} f(\mathbf{x}^{(i)} + \gamma \mathbf{d}^{(i)})$ 
  fim para
   $\mathbf{x}^{(n+2)} \leftarrow 2\mathbf{x}^{(n+1)} - \mathbf{x}^{(1)}$ 
  se  $f(\mathbf{x}^{(n+2)}) \geq f(\mathbf{x}^{(1)})$  então
     $\mathbf{x}^{(\text{opt})} \leftarrow \mathbf{x}^{(n+1)}$ 
  senão
     $\mathbf{x}^{(\text{opt})} \leftarrow \underset{\gamma \in \mathcal{R}}{\text{argmin}} f(\mathbf{x}^{(n+1)} + \gamma(\mathbf{x}^{(n+1)} - \mathbf{x}^{(1)}))$ 
    escolha nova direção de busca  $\mathbf{d}^{(1)}, \dots, \mathbf{d}^{(n)}$ 
  fim se
  até  $\|\mathbf{x}^{(\text{opt})} - \mathbf{x}^{(1)}\| < \epsilon$ 
  imprima  $\mathbf{x}^{(\text{opt})}$ 
fim algoritmo

```

Figura 3.2: Algoritmo de Powell

Smith, 1987). Uma discussão mais detalhada sobre a implementação do método de Powell em conjunto com o GEM na análise do RBCAP, bem como suas restrições e características de desempenho, foram descritas por van Woensel et al. (2010).

### 3.5 Observações Finais

Neste capítulo foi apresentado um método popular e eficiente para a estimação das medidas de desempenho em redes de filas finitas gerais multi-servidores, o GEM, com ênfase nas taxas de atendimento. Este método foi combinado com o algoritmo de otimização irrestrita de Powell, um clássico algoritmo de otimização não-linear. No capítulo seguinte serão apresentados alguns resultados experimentais baseados nesta metodologia.

# Capítulo 4

## Resultados e Discussão

### 4.1 Introdução

O algoritmo de otimização conjunta foi testado em redes nas topologias *série*, *divisão* e *fusão*, com o objetivo de identificar possíveis padrões de alocação existentes, através da variação de parâmetros de interesse presentes em cada uma das topologias. Todos os algoritmos descritos foram implementados em FORTRAN e estão disponíveis, a pedido, para fins de pesquisa. Apresenta-se a seguir uma breve descrição das topologias utilizadas, com as respectivas configurações utilizadas nas simulações. Na sequência serão apresentados os resultados obtidos.

#### 4.1.1 Topologia Série

A topologia de rede de filas em *série* implica que o cliente após sair de um nó, não possui possibilidade de seleção do próximo nó, ou seja, o serviço será recebido de forma serial. Supondo um sistema com três nós em série, o cliente ao sair do nó 1 automaticamente seria destinado ao nó 2 e conseqüentemente ao nó 3. A Figura 4.1 apresenta um exemplo de rede de filas em série.

Um exemplo desta topologia seria um sistema de atendimento de emergência

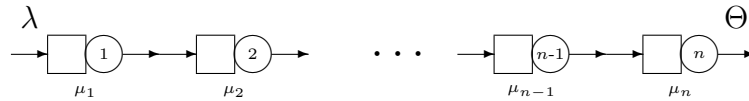


Figura 4.1: Rede de filas na topologia série

hospitalar, onde o paciente primeiramente se registra no guichê de atendimento, após efetuado o registro aguarda pela avaliação inicial geralmente realizada por uma enfermeira e após a triagem aguarda o atendimento médico, que poderá levar o paciente a uma série de novos procedimentos sequenciais, tais como exames laboratoriais, exames de imagem, etc.

Para as simulações na topologia em série, foram consideradas redes de filas com o número de nós  $N = 2, 3, 4, 5$ , taxa de entrada na rede  $\Lambda = 5$  e servidores com taxas de atendimento  $\mu_j = 10$ .

### 4.1.2 Topologia Divisão

A topologia *divisão*, diferentemente da topologia em *série*, apresenta ao término do atendimento em um nó a possibilidade de seleção do próximo nó. Supondo que após o término do atendimento, o cliente possua  $n$  nós subjacentes, o mesmo pode selecionar o próximo nó de destino com probabilidade  $p_i, i = 1, 2, \dots, n$ . A Figura 4.2 representa tal topologia.

O guichê de companhia aérea poderia ser representado por uma rede de filas do tipo divisão. Cada cliente acessa uma única fila onde irá adquirir seu bilhete de embarque. Após a aquisição do bilhete, cada cliente se destinará à fila para seu respectivo voo, pois não necessariamente todos os clientes da companhia aérea compram passagens para o mesmo voo. Neste caso, existe uma probabilidade  $p_i$  associada ao embarque de cada passageiro após o atendimento inicial, indicando uma divisão entre o grupo de clientes, anteriormente agrupados em uma única fila.



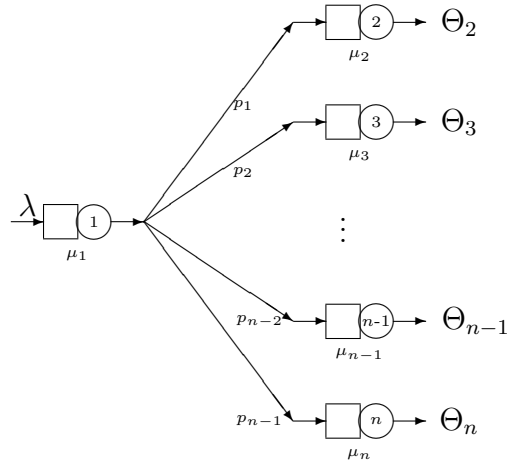


Figura 4.2: Rede de filas na topologia divisão

As redes de filas com a topologia divisão utilizadas nos experimentos computacionais contam com um número de nós fixo  $N = 3$ , onde um nó inicial  $h_1$  recebe as entidades a serem atendidas e após o atendimento elas são encaminhadas para os nós  $h_2$  e  $h_3$  com probabilidades de roteamento  $p_1$  e  $p_2$ , respectivamente. Foram configuradas redes com probabilidades de roteamento variando dentre as seguintes opções:  $p_1 = 0.1$  e  $p_2 = 0.9$ ,  $p_1 = 0.2$ ,  $p_2 = 0.8$ ,  $p_1 = 0.3$ ,  $p_2 = 0.7$ ,  $p_1 = 0.4$ ,  $p_2 = 0.6$  e a situação de equilíbrio  $p_1 = p_2 = 0.5$ . Assim como nas redes de filas em série, foram consideradas taxas de entrada na rede  $\Lambda = 5$  e servidores com taxas de atendimento  $\mu_j = 10$ .

### 4.1.3 Topologia Fusão

Uma rede de filas com topologia *fusão* se caracteriza pela união de clientes atendidos em servidores distintos em um único servidor subsequente, ou seja, supondo uma série de filas paralelas, todos os clientes destas ao término do atendimento serão direcionados para uma única fila, independentemente do atendimento anterior. Um esquema deste tipo de topologia é apresentado na Figura 4.3.

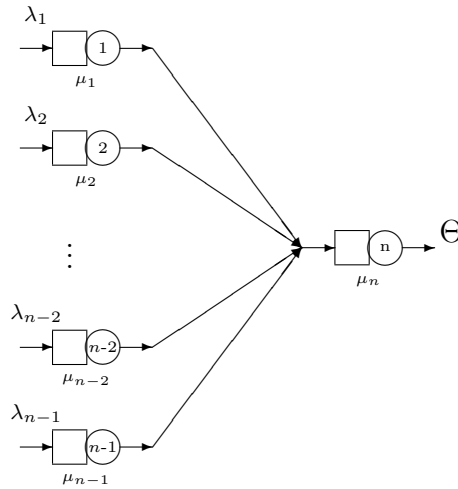


Figura 4.3: Rede de filas na topologia fusão

O Exemplo 1.1, apresentado na introdução, pág. 2, representa uma rede de filas do tipo fusão, em que gêneros alimentícios distintos são produzidos em paralelo e após a conclusão de sua fabricação, todos eles são destinados a um mesmo centro de distribuição.

Nas redes do tipo fusão, também foram considerados 3 nós, onde as entidades a serem atendidas ingressam na rede pelos nós  $h_1$  e  $h_2$  e após atendimento, são necessariamente destinadas ao nó  $h_3$ . Foi mantida a taxa total de entrada no sistema  $\Lambda = 5$ , porém, tal taxa foi dividida entre os nós iniciais nas seguintes proporções variáveis  $\lambda_1$  (taxa de entrada no nó  $h_1$ ) e  $\lambda_2$  (taxa de entrada no nó  $h_2$ ):  $\lambda_1 = 0.5$  e  $\lambda_2 = 4.5$ ,  $\lambda_1 = 1.0$  e  $\lambda_2 = 4.0$ ,  $\lambda_1 = 1.5$  e  $\lambda_2 = 3.5$ ,  $\lambda_1 = 2.0$  e  $\lambda_2 = 3.0$  e  $\lambda_1 = \lambda_2 = 2.5$ . A taxa de atendimento em cada servidor é  $\mu_j = 10$ .

## 4.2 Análise dos Padrões de Alocação Ótima

Para as três topologias básicas, foram considerados os seguintes valores para os quadrados dos coeficientes de variação dos tempos de serviço,  $s^2 = 0,5, 1,0$  e

1,5. Assim, estão representados sistemas *hipoexponenciais* (isto é, com  $s^2 = 0,5$ ), *markovianos* (com  $s^2 = 1,0$ ) e *hiperexponenciais* (ou seja, com  $s^2 = 1,5$ ).

### 4.2.1 Topologia Série

Nas Tabelas 4.1 a 4.4 encontram-se as soluções sub-ótimas de redes em topologia série e na Figura 4.4 tem-se uma representação gráfica dos padrões de alocação para as diferentes configurações de números de nós e quadrados dos coeficientes de variação dos tempos de serviço ( $s^2$ ), em que observamos os seguintes aspectos. As alocações, tanto de servidores quanto das áreas de espera (ou *buffers*), são homogêneas, conforme esperado. Não poderia ser diferente, pois as filas possuem servidores com a mesma taxa de serviço ( $\mu_j = 10$ ). É interessante observar que as configurações sub-ótimas não possuem área de espera ( $\sum_i B_i = 0$ ) até uma relação de custos  $C_c/C_B$  superior a 2 : 1 (ou seja,  $w_i/(1 - w_i) > 2$ ). Infelizmente, esse limiar é bastante imprevisível e dependente da configuração da rede e é, até o presente momento, determinável apenas pela aplicação do algoritmo. Em outras palavras, não há ainda uma regra que preveja essa mudança de estado (sem *buffers*  $\rightarrow$  com *buffers*).

No que diz respeito ao quadrado do coeficiente de variação do tempo de serviço ( $s^2$ ), um aumento no seu valor (isto é, na variabilidade) conduz a uma maior alocação (tanto de servidores, quanto de áreas de espera). Tal fato é confirmado pelo crescente valor do  $Z_\alpha$ , o que é também bastante intuitivo. É interessante o comportamento da rede de filas no sistema hipoexponencial, que apresenta queda no número de servidores e aumento brusco do número das áreas de espera na transição  $C_c/C_B = 4 : 1 \rightarrow C_c/C_B = 8 : 1$ , sendo que para os sistemas markovianos houve estabilidade tanto para servidores quanto para áreas de espera e para o caso em que  $s^2 = 1,5$ , observa-se manutenção do número de servidores e menor aumento no número de áreas de espera. Reforça-se aqui a convicção, já mencionada

Tabela 4.1: Resultados do BCAP para topologia série e  $N = 2$

$s^2$	$C_c/C_B$	$\mathbf{c}$	$\mathbf{K}$	$\sum_i c_i$	$\sum_i K_i$	$\sum_i B_i$	$\theta(\mathbf{c}, \mathbf{B})$	$Z_\alpha$
0.5	1:16	( 4 4 )	( 4 4 )	8	8	0	5.0000	0.47
	1:8	( 4 4 )	( 4 4 )	8	8	0	5.0000	0.89
	1:4	( 4 4 )	( 4 4 )	8	8	0	5.0000	1.60
	1:2	( 3 3 )	( 3 3 )	6	6	0	4.9996	2.41
	1:1	( 3 3 )	( 3 3 )	6	6	0	4.9996	3.41
	2:1	( 3 3 )	( 3 3 )	6	6	0	4.9996	4.41
	4:1	( 2 2 )	( 5 5 )	4	10	6	4.9996	4.81
	8:1	( 1 1 )	( 11 11 )	2	22	20	4.9998	4.24
1.0	1:16	( 4 4 )	( 4 4 )	8	8	0	5.0000	0.47
	1:8	( 4 4 )	( 4 4 )	8	8	0	5.0000	0.89
	1:4	( 4 4 )	( 4 4 )	8	8	0	5.0000	1.60
	1:2	( 3 3 )	( 3 3 )	6	6	0	4.9994	2.58
	1:1	( 3 3 )	( 3 3 )	6	6	0	4.9994	3.58
	2:1	( 3 3 )	( 3 3 )	6	6	0	4.9994	4.58
	4:1	( 2 2 )	( 6 6 )	4	12	8	4.9998	4.98
	8:1	( 2 2 )	( 6 6 )	4	12	8	4.9998	4.63
1.5	1:16	( 4 4 )	( 4 4 )	8	8	0	5.0000	0.48
	1:8	( 4 4 )	( 4 4 )	8	8	0	5.0000	0.89
	1:4	( 4 4 )	( 4 4 )	8	8	0	5.0000	1.61
	1:2	( 4 4 )	( 4 4 )	8	8	0	5.0000	2.67
	1:1	( 3 3 )	( 3 3 )	6	6	0	4.9992	3.80
	2:1	( 3 3 )	( 3 3 )	6	6	0	4.9992	4.80
	4:1	( 2 2 )	( 6 6 )	4	12	8	4.9997	5.13
	8:1	( 2 2 )	( 7 7 )	4	14	10	4.9999	4.76

Tabela 4.2: Resultados do BCAP para topologia série e  $N = 3$

$s^2$	$C_c/C_B$	$\mathbf{c}$	$\mathbf{K}$	$\sum_i c_i$	$\sum_i K_i$	$\sum_i B_i$	$\theta(\mathbf{c}, \mathbf{B})$	$Z_\alpha$
0.5	1:16	( 4 4 4 )	( 4 4 4 )	12	12	0	5.0000	0.71
	1:8	( 4 4 4 )	( 4 4 4 )	12	12	0	5.0000	1.34
	1:4	( 4 4 4 )	( 4 4 4 )	12	12	0	5.0000	2.40
	1:2	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9994	3.62
	1:1	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9994	5.12
	2:1	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9994	6.62
	4:1	( 2 2 2 )	( 5 5 5 )	6	15	9	4.9994	7.22
	8:1	( 1 1 1 )	( 11 11 11 )	3	33	30	4.9996	6.36
1.0	1:16	( 4 4 4 )	( 4 4 4 )	12	12	0	5.0000	0.71
	1:8	( 4 4 4 )	( 4 4 4 )	12	12	0	5.0000	1.34
	1:4	( 4 4 4 )	( 4 4 4 )	12	12	0	5.0000	2.41
	1:2	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9991	3.88
	1:1	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9991	5.38
	2:1	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9991	6.88
	4:1	( 2 2 2 )	( 6 6 6 )	6	18	12	4.9997	7.47
	8:1	( 2 2 2 )	( 6 6 6 )	6	18	12	4.9997	6.94
1.5	1:16	( 4 4 4 )	( 4 4 4 )	12	12	0	5.0000	0.71
	1:8	( 4 4 4 )	( 4 4 4 )	12	12	0	5.0000	1.34
	1:4	( 4 4 4 )	( 4 4 4 )	12	12	0	5.0000	2.41
	1:2	( 4 4 4 )	( 4 4 4 )	12	12	0	5.0000	4.01
	1:1	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9988	5.70
	2:1	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9988	7.20
	4:1	( 2 2 2 )	( 6 6 6 )	6	18	12	4.9995	7.69
	8:1	( 2 2 2 )	( 7 7 7 )	6	21	15	4.9999	7.14

Tabela 4.3: Resultados do BCAP para topologia série e  $N = 4$

$s^2$	$C_c/C_B$	$\mathbf{c}$	$\mathbf{K}$	$\sum_i c_i$	$\sum_i K_i$	$\sum_i B_i$	$\theta(\mathbf{c}, \mathbf{B})$	$Z_\alpha$
0.5	1:16	( 4 4 4 4 )	( 4 4 4 4 )	16	16	0	5.0000	0.95
	1:8	( 4 4 4 4 )	( 4 4 4 4 )	16	16	0	5.0000	1.78
	1:4	( 4 4 4 4 )	( 4 4 4 4 )	16	16	0	5.0000	3.21
	1:2	( 3 3 3 3 )	( 3 3 3 3 )	12	12	0	4.9992	4.82
	1:1	( 3 3 3 3 )	( 3 3 3 3 )	12	12	0	4.9992	6.82
	2:1	( 3 3 3 3 )	( 3 3 3 3 )	12	12	0	4.9992	8.82
	4:1	( 2 2 2 2 )	( 5 5 5 5 )	8	20	12	4.9992	9.63
	8:1	( 1 1 1 1 )	( 11 11 11 11 )	4	44	40	4.9995	8.47
1.0	1:16	( 4 4 4 4 )	( 4 4 4 4 )	16	16	0	5.0000	0.95
	1:8	( 4 4 4 4 )	( 4 4 4 4 )	16	16	0	5.0000	1.79
	1:4	( 4 4 4 4 )	( 4 4 4 4 )	16	16	0	5.0000	3.21
	1:2	( 3 3 3 3 )	( 3 3 3 3 )	12	12	0	4.9988	5.17
	1:1	( 3 3 3 3 )	( 3 3 3 3 )	12	12	0	4.9988	7.17
	2:1	( 3 3 3 3 )	( 3 3 3 3 )	12	12	0	4.9988	9.17
	4:1	( 2 2 2 2 )	( 6 6 6 6 )	8	24	16	4.9996	9.97
	8:1	( 2 2 2 2 )	( 6 6 6 6 )	8	24	16	4.9996	9.26
1.5	1:16	( 4 4 4 4 )	( 4 4 4 4 )	16	16	0	5.0000	0.95
	1:8	( 4 4 4 4 )	( 4 4 4 4 )	16	16	0	5.0000	1.79
	1:4	( 4 4 4 4 )	( 4 4 4 4 )	16	16	0	5.0000	3.21
	1:2	( 4 4 4 4 )	( 4 4 4 4 )	16	16	0	5.0000	5.34
	1:1	( 3 3 3 3 )	( 3 3 3 3 )	12	12	0	4.9984	7.60
	2:1	( 3 3 3 3 )	( 3 3 3 3 )	12	12	0	4.9984	9.60
	4:1	( 2 2 2 2 )	( 6 6 6 6 )	8	24	16	4.9993	10.26
	8:1	( 2 2 2 2 )	( 7 7 7 7 )	8	28	20	4.9998	9.51

Tabela 4.4: Resultados do BCAP para topologia série e  $N = 5$

$s^2$	$C_c/C_B$	$\mathbf{c}$	$\mathbf{K}$	$\sum_i c_i$	$\sum_i K_i$	$\sum_i B_i$	$\theta(\mathbf{c}, \mathbf{B})$	$Z_\alpha$
0.5	1:16	( 4 4 4 4 4 )	( 4 4 4 4 4 )	20	20	0	5.0000	1.18
	1:8	( 4 4 4 4 4 )	( 4 4 4 4 4 )	20	20	0	5.0000	2.23
	1:4	( 4 4 4 4 4 )	( 4 4 4 4 4 )	20	20	0	5.0000	4.01
	1:2	( 3 3 3 3 3 )	( 3 3 3 3 3 )	15	15	0	4.9990	6.02
	1:1	( 3 3 3 3 3 )	( 3 3 3 3 3 )	15	15	0	4.9990	8.52
	2:1	( 3 3 3 3 3 )	( 3 3 3 3 3 )	15	15	0	4.9990	11.02
	4:1	( 2 2 2 2 2 )	( 5 5 5 5 5 )	10	25	15	4.9990	12.04
	8:1	( 1 1 1 1 1 )	( 11 11 11 11 11 )	5	55	50	4.9994	10.59
1.0	1:16	( 4 4 4 4 4 )	( 4 4 4 4 4 )	20	20	0	5.0000	1.19
	1:8	( 4 4 4 4 4 )	( 4 4 4 4 4 )	20	20	0	5.0000	2.23
	1:4	( 4 4 4 4 4 )	( 4 4 4 4 4 )	20	20	0	5.0000	4.01
	1:2	( 3 3 3 3 3 )	( 3 3 3 3 3 )	15	15	0	4.9985	6.46
	1:1	( 3 3 3 3 3 )	( 3 3 3 3 3 )	15	15	0	4.9985	8.96
	2:1	( 3 3 3 3 3 )	( 3 3 3 3 3 )	15	15	0	4.9985	11.46
	4:1	( 2 2 2 2 2 )	( 6 6 6 6 6 )	10	30	20	4.9995	12.46
	8:1	( 2 2 2 2 2 )	( 6 6 6 6 6 )	10	30	20	4.9995	11.57
1.5	1:16	( 4 4 4 4 4 )	( 4 4 4 4 4 )	20	20	0	5.0000	1.19
	1:8	( 4 4 4 4 4 )	( 4 4 4 4 4 )	20	20	0	5.0000	2.24
	1:4	( 4 4 4 4 4 )	( 4 4 4 4 4 )	20	20	0	5.0000	4.01
	1:2	( 4 4 4 4 4 )	( 4 4 4 4 4 )	20	20	0	5.0000	6.68
	1:1	( 3 3 3 3 3 )	( 3 3 3 3 3 )	15	15	0	4.9980	9.50
	2:1	( 3 3 3 3 3 )	( 3 3 3 3 3 )	15	15	0	4.9980	12.00
	4:1	( 2 2 2 2 2 )	( 6 6 6 6 6 )	10	30	20	4.9992	12.82
	8:1	( 2 2 2 2 2 )	( 7 7 7 7 7 )	10	35	25	4.9998	11.89

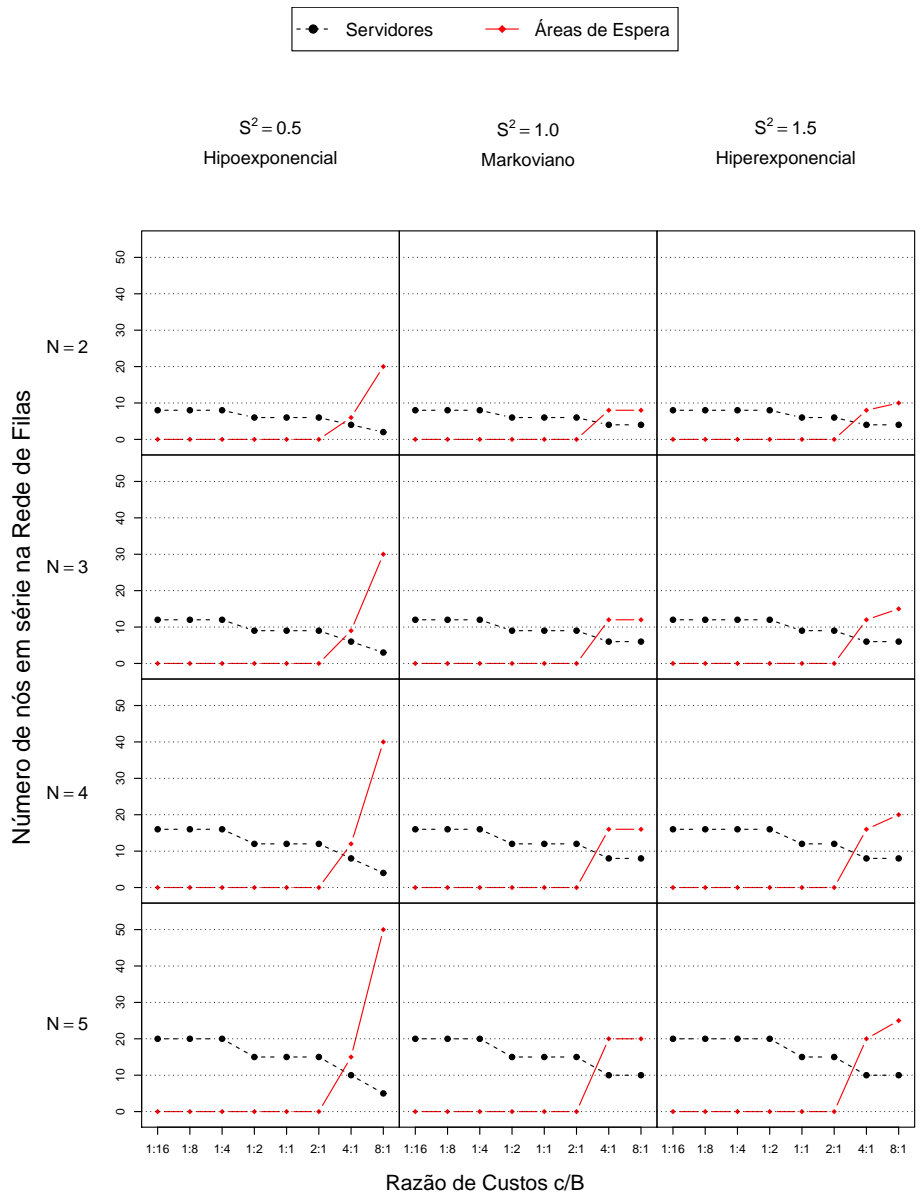


Figura 4.4: Alocação de servidores e áreas de espera para topologia série



na literatura da área (van Woensel et al., 2010; Cruz & van Woensel, 2014a;c), que os servidores são mais eficazes para lidar com a variabilidade que as áreas de espera, possivelmente pelo fato de que eles têm a função dupla de segurar o cliente e o atender. Finalmente, em relação à configuração em série, notamos que o aumento do número de filas na linha não altera o padrão homogêneo da alocação nem a transição sem/com *buffers*, o que demonstra a robustez da solução com a configuração do sistema de filas a ser otimizado.

### 4.2.2 Topologia Divisão

Os resultados da topologia divisão podem ser vistos nas Tabelas 4.5 a 4.9 e na Figura 4.5. Como observado na topologia série, as alocações são sem áreas de espera até um certa proporção de custos  $C_c/C_B$  (na maioria dos casos observa-se ausência de áreas de espera para proporções iguais ou inferiores a 2:1), independente do quadrado do coeficiente de variação do tempo de serviço ( $s^2$ ). Também interessante é que a proporção de clientes destinados para cada braço da divisão após o atendimento inicial não se verifica na alocação de recursos. Por exemplo, a razão das probabilidades de roteamento  $0,1 : 0,9$  não resulta em uma divisão de recursos nesta mesma proporção, mas na de  $2 : 3$ . Entretanto, à medida em que as probabilidades de roteamento tornam-se próximas (ou seja,  $p_1 \rightarrow p_2$ ), a alocação torna-se homogênea.

Também digno de nota é que a alocação na fila antes da divisão em geral *não* é a soma das alocações nos ramos após a divisão, tanto no que diz respeito a  $c$  quanto  $K$ . Neste aspecto, observa-se que quanto maior o custo dos servidores, maior será a diferença entre o número de áreas de espera do ramo inicial e a soma dos ramos que receberão as entidades atendidas pelo primeiro nó. Quando necessárias, para valores de  $s^2 = 0.5$ , observa-se uma alocação superior de áreas de espera para os ramos finais da rede, enquanto para valores de  $s^2 = 1.0$  e  $1.5$ , o maior contingente

Tabela 4.5: Resultados do BCAP para topologia divisão,  $p_1 = 0,1$  e  $p_2 = 0,9$

$s^2$	$C_c/C_B$	$\mathbf{c}$	$\mathbf{K}$	$\sum_i c_i$	$\sum_i K_i$	$\sum_i B_i$	$\theta(\mathbf{c}, \mathbf{B})$	$Z_\alpha$
0.5	1:16	( 5 2 3 )	( 5 2 3 )	10	10	0	4.9999	0.64
	1:8	( 5 2 3 )	( 5 2 3 )	10	10	0	4.9999	1.17
	1:4	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9997	2.05
	1:2	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9997	3.25
	1:1	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9997	4.75
	2:1	( 3 1 3 )	( 3 3 3 )	7	9	2	4.9997	5.64
	4:1	( 2 1 2 )	( 5 3 4 )	5	12	7	4.9995	5.87
	8:1	( 1 1 1 )	( 11 3 8 )	3	22	19	4.9997	5.10
1.0	1:16	( 5 2 3 )	( 5 2 3 )	10	10	0	4.9999	0.67
	1:8	( 5 2 3 )	( 5 2 3 )	10	10	0	4.9999	1.19
	1:4	( 5 2 3 )	( 5 2 3 )	10	10	0	4.9999	2.08
	1:2	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9996	3.36
	1:1	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9996	4.86
	2:1	( 3 1 3 )	( 3 3 3 )	7	9	2	4.9996	5.78
	4:1	( 3 1 2 )	( 4 3 5 )	6	12	6	4.9998	6.21
	8:1	( 1 1 1 )	( 16 3 10 )	3	29	26	4.9998	5.76
1.5	1:16	( 5 2 3 )	( 5 2 3 )	10	10	0	4.9999	0.69
	1:8	( 5 2 3 )	( 5 2 3 )	10	10	0	4.9999	1.22
	1:4	( 5 2 3 )	( 5 2 3 )	10	10	0	4.9999	2.11
	1:2	( 5 2 3 )	( 5 2 3 )	10	10	0	4.9999	3.44
	1:1	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9995	4.99
	2:1	( 3 1 3 )	( 3 3 3 )	7	9	2	4.9994	5.98
	4:1	( 3 1 2 )	( 4 3 5 )	6	12	6	4.9997	6.34
	8:1	( 1 1 1 )	( 16 4 11 )	3	31	28	4.9997	6.06

Tabela 4.6: Resultados do BCAP para topologia divisão,  $p_1 = 0,2$  e  $p_2 = 0,8$

$s^2$	$C_c/C_B$	$\mathbf{c}$	$\mathbf{K}$	$\sum_i c_i$	$\sum_i K_i$	$\sum_i B_i$	$\theta(\mathbf{c}, \mathbf{B})$	$Z_\alpha$
0.5	1:16	( 5 2 3 )	( 5 2 3 )	10	10	0	4.9999	0.64
	1:8	( 5 2 3 )	( 5 2 3 )	10	10	0	4.9999	1.17
	1:4	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9997	2.05
	1:2	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9997	3.25
	1:1	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9997	4.75
	2:1	( 3 1 3 )	( 3 3 3 )	7	9	2	4.9997	5.64
	4:1	( 2 1 2 )	( 5 3 4 )	5	12	7	4.9995	5.87
	8:1	( 1 1 1 )	( 11 3 8 )	3	22	19	4.9997	5.10
1.0	1:16	( 5 2 3 )	( 5 2 3 )	10	10	0	4.9999	0.67
	1:8	( 5 2 3 )	( 5 2 3 )	10	10	0	4.9999	1.19
	1:4	( 5 2 3 )	( 5 2 3 )	10	10	0	4.9999	2.08
	1:2	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9996	3.36
	1:1	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9996	4.86
	2:1	( 3 1 3 )	( 3 3 3 )	7	9	2	4.9996	5.78
	4:1	( 3 1 2 )	( 4 3 5 )	6	12	6	4.9998	6.21
	8:1	( 1 1 1 )	( 16 3 10 )	3	29	26	4.9998	5.76
1.5	1:16	( 5 2 3 )	( 5 2 3 )	10	10	0	4.9999	0.69
	1:8	( 5 2 3 )	( 5 2 3 )	10	10	0	4.9999	1.22
	1:4	( 5 2 3 )	( 5 2 3 )	10	10	0	4.9999	2.11
	1:2	( 5 2 3 )	( 5 2 3 )	10	10	0	4.9999	3.44
	1:1	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9995	4.99
	2:1	( 3 1 3 )	( 3 3 3 )	7	9	2	4.9994	5.98
	4:1	( 3 1 2 )	( 4 3 5 )	6	12	6	4.9997	6.34
	8:1	( 1 1 1 )	( 16 4 11 )	3	31	28	4.9997	6.06

Tabela 4.7: Resultados do BCAP para topologia divisão,  $p_1 = 0,3$  e  $p_2 = 0,7$

$s^2$	$C_c/C_B$	$\mathbf{c}$	$\mathbf{K}$	$\sum_i c_i$	$\sum_i K_i$	$\sum_i B_i$	$\theta(\mathbf{c}, \mathbf{B})$	$Z_\alpha$
0.5	1:16	( 5 2 3 )	( 5 2 3 )	10	10	0	4.9999	0.67
	1:8	( 5 2 3 )	( 5 2 3 )	10	10	0	4.9999	1.20
	1:4	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9998	2.02
	1:2	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9998	3.22
	1:1	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9998	4.72
	2:1	( 3 2 3 )	( 3 2 3 )	8	8	0	4.9997	5.62
	4:1	( 3 1 2 )	( 3 4 4 )	6	11	5	4.9997	6.14
	8:1	( 1 1 1 )	( 11 4 7 )	3	22	19	4.9997	5.08
1.0	1:16	( 5 2 3 )	( 5 2 3 )	10	10	0	4.9999	0.70
	1:8	( 5 2 3 )	( 5 2 3 )	10	10	0	4.9999	1.22
	1:4	( 5 2 3 )	( 5 2 3 )	10	10	0	4.9999	2.11
	1:2	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9997	3.32
	1:1	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9997	4.82
	2:1	( 3 2 3 )	( 3 2 3 )	8	8	0	4.9996	5.73
	4:1	( 3 1 2 )	( 4 4 4 )	6	12	6	4.9997	6.29
	8:1	( 1 1 1 )	( 16 4 9 )	3	29	26	4.9998	5.73
1.5	1:16	( 4 4 4 )	( 4 4 4 )	12	12	0	5.0000	0.71
	1:8	( 5 2 3 )	( 5 2 3 )	10	10	0	4.9999	1.25
	1:4	( 5 2 3 )	( 5 2 3 )	10	10	0	4.9999	2.14
	1:2	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9996	3.44
	1:1	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9996	4.94
	2:1	( 3 2 3 )	( 3 2 3 )	8	8	0	4.9995	5.87
	4:1	( 3 1 2 )	( 4 4 4 )	6	12	6	4.9995	6.48
	8:1	( 1 1 1 )	( 16 5 10 )	3	31	28	4.9998	6.00

Tabela 4.8: Resultados do BCAP para topologia divisão,  $p_1 = 0,4$  e  $p_2 = 0,6$

$s^2$	$C_c/C_B$	$\mathbf{c}$	$\mathbf{K}$	$\sum_i c_i$	$\sum_i K_i$	$\sum_i B_i$	$\theta(\mathbf{c}, \mathbf{B})$	$Z_\alpha$
0.5	1:16	( 4 4 4 )	( 4 4 4 )	12	12	0	5.0000	0.71
	1:8	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9998	1.21
	1:4	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9998	2.01
	1:2	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9998	3.21
	1:1	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9998	4.71
	2:1	( 3 2 2 )	( 3 2 3 )	7	8	1	4.9993	5.70
	4:1	( 3 1 1 )	( 3 4 6 )	5	13	8	4.9994	6.18
	8:1	( 1 1 1 )	( 11 5 6 )	3	22	19	4.9997	5.08
1.0	1:16	( 4 4 4 )	( 4 4 4 )	12	12	0	5.0000	0.71
	1:8	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9997	1.30
	1:4	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9997	2.10
	1:2	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9997	3.30
	1:1	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9997	4.80
	2:1	( 3 2 2 )	( 3 2 3 )	7	8	1	4.9990	5.96
	4:1	( 2 2 2 )	( 6 3 4 )	6	13	7	4.9998	6.37
	8:1	( 1 1 1 )	( 16 5 7 )	3	28	25	4.9997	5.70
1.5	1:16	( 4 4 4 )	( 4 4 4 )	12	12	0	5.0000	0.71
	1:8	( 4 4 4 )	( 4 4 4 )	12	12	0	5.0000	1.34
	1:4	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9996	2.21
	1:2	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9996	3.41
	1:1	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9996	4.91
	2:1	( 3 2 3 )	( 3 3 3 )	8	9	1	4.9995	6.12
	4:1	( 2 2 2 )	( 6 3 4 )	6	13	7	4.9997	6.49
	8:1	( 1 1 1 )	( 16 5 8 )	3	29	26	4.9995	6.02

Tabela 4.9: Resultados do BCAP para topologia divisão,  $p_1 = 0,5$  e  $p_2 = 0,5$

$s^2$	$C_c/C_B$	$\mathbf{c}$	$\mathbf{K}$	$\sum_i c_i$	$\sum_i K_i$	$\sum_i B_i$	$\theta(\mathbf{c}, \mathbf{B})$	$Z_\alpha$
0.5	1:16	( 4 4 4 )	( 4 4 4 )	12	12	0	5.0000	0.71
	1:8	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9998	1.21
	1:4	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9998	2.01
	1:2	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9998	3.21
	1:1	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9998	4.71
	2:1	( 3 2 2 )	( 3 3 3 )	7	9	2	4.9996	5.70
	4:1	( 3 1 1 )	( 3 5 5 )	5	13	8	4.9995	6.14
	8:1	( 1 1 1 )	( 11 6 6 )	3	23	20	4.9998	5.08
1.0	1:16	( 4 4 4 )	( 4 4 4 )	12	12	0	5.0000	0.71
	1:8	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9997	1.30
	1:4	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9997	2.10
	1:2	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9997	3.30
	1:1	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9997	4.80
	2:1	( 3 2 2 )	( 3 3 3 )	7	9	2	4.9995	5.86
	4:1	( 2 2 2 )	( 6 3 3 )	6	12	6	4.9997	6.33
	8:1	( 1 1 1 )	( 16 6 6 )	3	28	25	4.9998	5.69
1.5	1:16	( 5 3 3 )	( 5 3 3 )	11	11	0	5.0000	0.66
	1:8	( 5 3 3 )	( 5 3 3 )	11	11	0	5.0000	1.23
	1:4	( 5 3 3 )	( 5 3 3 )	11	11	0	5.0000	2.21
	1:2	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9996	3.41
	1:1	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9996	4.91
	2:1	( 3 2 2 )	( 3 3 3 )	7	9	2	4.9993	6.06
	4:1	( 2 2 2 )	( 6 3 3 )	6	12	6	4.9995	6.49
	8:1	( 1 1 1 )	( 16 7 7 )	3	30	27	4.9997	5.94

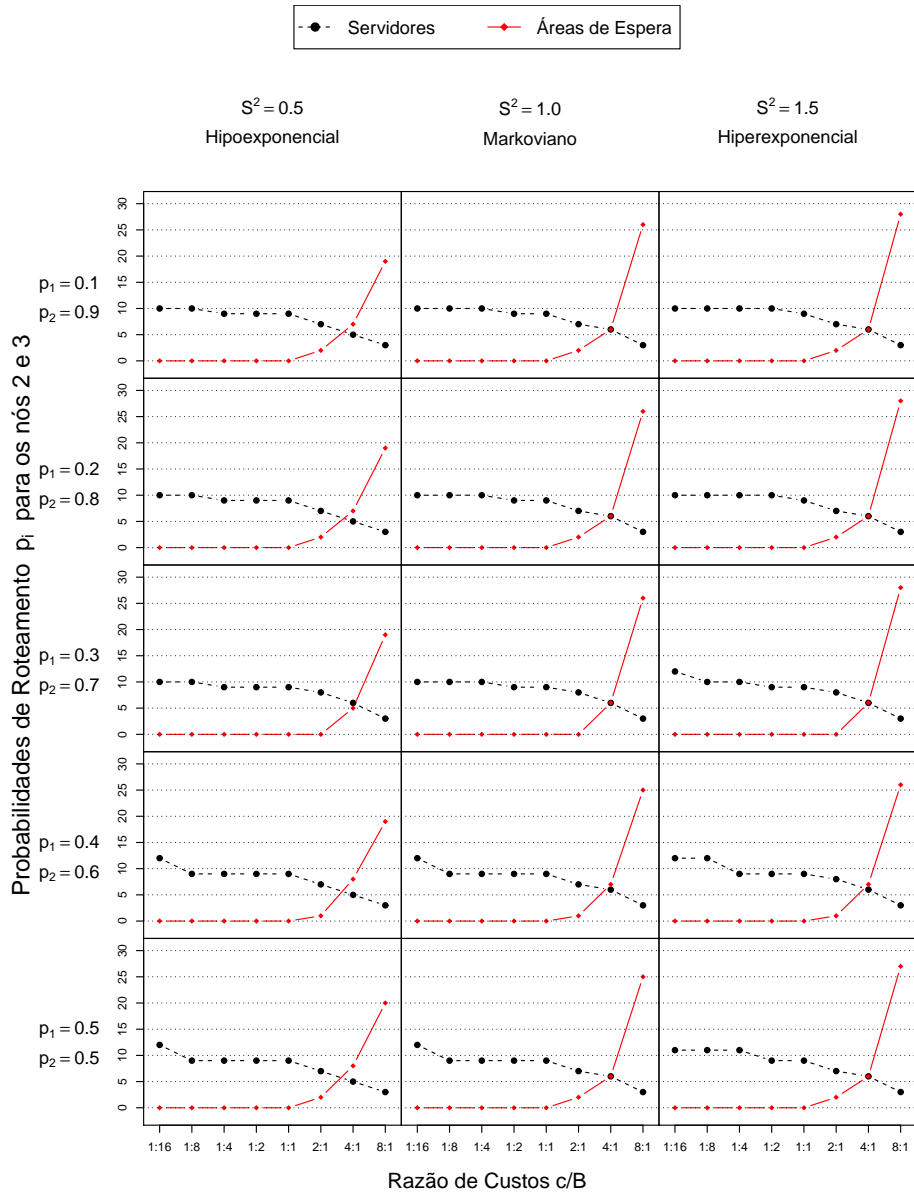


Figura 4.5: Alocação de servidores e áreas de espera para topologia divisão

de áreas de espera é destinado para o nó inicial da rede. Finalmente, notamos que, com o aumento do número de nós na rede (resultados não apresentados), a sequência de filas após a divisão comporta-se como uma topologia série (isto é, com alocação homogênea), como seria de se esperar.

### 4.2.3 Topologia Fusão

As Tabelas 4.10 a 4.14 e a Figura 4.6 apresentam os resultados de alocação obtidos para redes em topologia fusão. Notamos uma simetria entre a topologia fusão e a anterior (topologia divisão), conforme seria esperado. É interessante notar que, apenas no caso hiperexponencial ( $s^2 = 1,5$ ), com  $\lambda_1 = 0,5$  e  $\lambda_2 = 4,5$ , houve um acréscimo de um servidor no nó que recebe o maior número de clientes e no nó final. O acréscimo destes dois servidores impactou e reduziu as áreas de espera, em comparação com as demais configurações.

Também aqui, o nó após a divisão necessita de uma alocação que em geral não é a soma das alocações nos ramos antes da divisão. Verifica-se portanto um efeito do tipo economia de escala neste nó compartilhado. Assim como ocorrido nas demais topologias, as oscilações provocadas no sistema, neste caso o desequilíbrio entre as taxas de chegada  $\lambda_1$  e  $\lambda_2$  não afetaram o padrão homogêneo das alocações nem a transição da rede do estado sem áreas de espera para o estado com áreas de espera.

## 4.3 Fatores de Influência nas Alocações Ótimas

Na análise dos resultados até este ponto, padrões foram observados que indicaram a existência de homogeneidade na alocação ótima de servidores e áreas de espera, para valores de  $s^2 = 0,5$ ,  $1,0$  e  $1,5$ . A fim de verificar a possível influência de alguns fatores em tais padrões de alocação foi construído o diagrama de dispersão



Tabela 4.10: Resultados do BCAP para topologia fusão,  $\lambda_1 = 0,5$  e  $\lambda_2 = 4,5$

$s^2$	$C_c/C_B$	$\mathbf{c}$	$\mathbf{K}$	$\sum_i c_i$	$\sum_i K_i$	$\sum_i B_i$	$\theta(\mathbf{c}, \mathbf{B})$	$Z_\alpha$
0.5	1:16	( 4 4 4 )	( 4 4 4 )	12	12	0	5.0000	0.71
	1:8	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9997	1.31
	1:4	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9997	2.11
	1:2	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9997	3.31
	1:1	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9997	4.81
	2:1	( 1 3 3 )	( 2 3 3 )	7	8	1	4.9997	5.35
	4:1	( 1 2 2 )	( 2 5 5 )	5	12	7	4.9997	5.74
	8:1	( 1 1 1 )	( 2 10 11 )	3	23	20	4.9998	5.12
1.0	1:16	( 4 4 4 )	( 4 4 4 )	12	12	0	5.0000	0.71
	1:8	( 2 3 5 )	( 2 3 5 )	10	10	0	4.9999	1.26
	1:4	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9996	2.24
	1:2	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9996	3.44
	1:1	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9996	4.94
	2:1	( 1 3 3 )	( 2 3 3 )	7	8	1	4.9995	5.50
	4:1	( 1 2 2 )	( 2 5 6 )	5	13	8	4.9997	5.92
	8:1	( 1 1 1 )	( 2 11 13 )	3	26	23	4.9996	5.60
1.5	1:16	( 2 5 5 )	( 2 5 5 )	12	12	0	5.0000	0.71
	1:8	( 2 5 5 )	( 2 5 5 )	12	12	0	5.0000	1.33
	1:4	( 2 5 5 )	( 2 5 5 )	12	12	0	5.0000	2.40
	1:2	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9994	3.60
	1:1	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9994	5.10
	2:1	( 1 3 3 )	( 2 3 3 )	7	8	1	4.9993	5.68
	4:1	( 1 2 2 )	( 2 6 6 )	5	14	9	4.9997	6.11
	8:1	( 1 2 2 )	( 2 6 7 )	5	15	10	4.9998	5.75

Tabela 4.11: Resultados do BCAP para topologia fusão,  $\lambda_1 = 1,0$  e  $\lambda_2 = 4,0$

$s^2$	$C_c/C_B$	$\mathbf{c}$	$\mathbf{K}$	$\sum_i c_i$	$\sum_i K_i$	$\sum_i B_i$	$\theta(\mathbf{c}, \mathbf{B})$	$Z_\alpha$
0.5	1:16	( 4 4 4 )	( 4 4 4 )	12	12	0	5.0000	0.71
	1:8	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9997	1.25
	1:4	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9997	2.05
	1:2	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9997	3.25
	1:1	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9997	4.75
	2:1	( 1 3 3 )	( 3 3 3 )	7	9	2	4.9997	5.64
	4:1	( 1 2 2 )	( 3 4 5 )	5	12	7	4.9995	5.87
	8:1	( 1 1 1 )	( 3 8 11 )	3	22	19	4.9997	5.10
1.0	1:16	( 2 3 5 )	( 2 3 5 )	10	10	0	4.9999	0.67
	1:8	( 2 3 5 )	( 2 3 5 )	10	10	0	4.9999	1.19
	1:4	( 2 3 5 )	( 2 3 5 )	10	10	0	4.9999	2.08
	1:2	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9996	3.36
	1:1	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9996	4.86
	2:1	( 1 3 3 )	( 3 3 3 )	7	9	2	4.9996	5.78
	4:1	( 1 2 3 )	( 3 5 4 )	6	12	6	4.9998	6.21
	8:1	( 1 1 1 )	( 3 10 13 )	3	26	23	4.9997	5.57
1.5	1:16	( 2 3 5 )	( 2 3 5 )	10	10	0	4.9999	0.69
	1:8	( 2 3 5 )	( 2 3 5 )	10	10	0	4.9999	1.22
	1:4	( 2 3 5 )	( 2 3 5 )	10	10	0	4.9999	2.11
	1:2	( 2 3 5 )	( 2 3 5 )	10	10	0	4.9999	3.44
	1:1	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9995	4.99
	2:1	( 1 3 3 )	( 3 3 3 )	7	9	2	4.9994	5.98
	4:1	( 1 2 3 )	( 3 5 4 )	6	12	6	4.9997	6.34
	8:1	( 1 1 1 )	( 4 11 15 )	3	30	27	4.9996	6.02

Tabela 4.12: Resultados do BCAP para topologia fusão,  $\lambda_1 = 1,5$  e  $\lambda_2 = 3,5$

$s^2$	$C_c/C_B$	$\mathbf{c}$	$\mathbf{K}$	$\sum_i c_i$	$\sum_i K_i$	$\sum_i B_i$	$\theta(\mathbf{c}, \mathbf{B})$	$Z_\alpha$
0.5	1:16	( 4 4 4 )	( 4 4 4 )	12	12	0	5.0000	0.71
	1:8	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9998	1.22
	1:4	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9998	2.02
	1:2	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9998	3.22
	1:1	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9998	4.72
	2:1	( 2 3 3 )	( 2 3 3 )	8	8	0	4.9997	5.62
	4:1	( 1 2 3 )	( 4 4 3 )	6	11	5	4.9997	6.14
	8:1	( 1 1 1 )	( 4 7 11 )	3	22	19	4.9997	5.08
1.0	1:16	( 2 3 5 )	( 2 3 5 )	10	10	0	4.9999	0.70
	1:8	( 2 3 5 )	( 2 3 5 )	10	10	0	4.9999	1.22
	1:4	( 2 3 5 )	( 2 3 5 )	10	10	0	4.9999	2.11
	1:2	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9997	3.32
	1:1	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9997	4.82
	2:1	( 2 3 3 )	( 2 3 3 )	8	8	0	4.9996	5.73
	4:1	( 1 2 3 )	( 4 4 4 )	6	12	6	4.9997	6.29
	8:1	( 1 1 1 )	( 4 9 13 )	3	26	23	4.9997	5.53
1.5	1:16	( 4 4 4 )	( 4 4 4 )	12	12	0	5.0000	0.71
	1:8	( 2 3 5 )	( 2 3 5 )	10	10	0	4.9999	1.25
	1:4	( 2 3 5 )	( 2 3 5 )	10	10	0	4.9999	2.14
	1:2	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9996	3.44
	1:1	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9996	4.94
	2:1	( 2 3 3 )	( 2 3 3 )	8	8	0	4.9995	5.87
	4:1	( 1 2 3 )	( 4 4 4 )	6	12	6	4.9995	6.48
	8:1	( 1 1 1 )	( 5 10 15 )	3	30	27	4.9997	5.97

Tabela 4.13: Resultados do BCAP para topologia fusão,  $\lambda_1 = 2,0$  e  $\lambda_2 = 3,0$

$s^2$	$C_c/C_B$	$\mathbf{c}$	$\mathbf{K}$	$\sum_i c_i$	$\sum_i K_i$	$\sum_i B_i$	$\theta(\mathbf{c}, \mathbf{B})$	$Z_\alpha$
0.5	1:16	( 4 4 4 )	( 4 4 4 )	12	12	0	5.0000	0.71
	1:8	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9998	1.21
	1:4	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9998	2.01
	1:2	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9998	3.21
	1:1	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9998	4.71
	2:1	( 2 2 3 )	( 2 3 3 )	7	8	1	4.9993	5.70
	4:1	( 1 1 3 )	( 4 6 3 )	5	13	8	4.9994	6.18
	8:1	( 1 1 1 )	( 5 6 11 )	3	22	19	4.9997	5.08
1.0	1:16	( 4 4 4 )	( 4 4 4 )	12	12	0	5.0000	0.71
	1:8	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9997	1.30
	1:4	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9997	2.10
	1:2	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9997	3.30
	1:1	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9997	4.80
	2:1	( 2 2 3 )	( 2 3 3 )	7	8	1	4.9990	5.96
	4:1	( 2 2 2 )	( 3 4 6 )	6	13	7	4.9998	6.37
	8:1	( 1 1 1 )	( 5 7 13 )	3	25	22	4.9996	5.50
1.5	1:16	( 4 4 4 )	( 4 4 4 )	12	12	0	5.0000	0.71
	1:8	( 4 4 4 )	( 4 4 4 )	12	12	0	5.0000	1.34
	1:4	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9996	2.21
	1:2	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9996	3.41
	1:1	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9996	4.91
	2:1	( 2 3 3 )	( 3 3 3 )	8	9	1	4.9995	6.12
	4:1	( 2 2 2 )	( 3 4 6 )	6	13	7	4.9997	6.49
	8:1	( 1 1 1 )	( 6 8 15 )	3	29	26	4.9996	5.92

Tabela 4.14: Resultados do BCAP para topologia fusão,  $\lambda_1 = 2,5$  e  $\lambda_2 = 2,5$

$s^2$	$C_c/C_B$	$\mathbf{c}$	$\mathbf{K}$	$\sum_i c_i$	$\sum_i K_i$	$\sum_i B_i$	$\theta(\mathbf{c}, \mathbf{B})$	$Z_\alpha$
0.5	1:16	( 4 4 4 )	( 4 4 4 )	12	12	0	5.0000	0.71
	1:8	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9998	1.21
	1:4	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9998	2.01
	1:2	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9998	3.21
	1:1	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9998	4.71
	2:1	( 2 2 3 )	( 3 3 3 )	7	9	2	4.9996	5.70
	4:1	( 1 1 3 )	( 5 5 3 )	5	13	8	4.9995	6.14
	8:1	( 1 1 1 )	( 6 6 11 )	3	23	20	4.9998	5.08
1.0	1:16	( 4 4 4 )	( 4 4 4 )	12	12	0	5.0000	0.71
	1:8	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9997	1.30
	1:4	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9997	2.10
	1:2	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9997	3.30
	1:1	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9997	4.80
	2:1	( 2 2 3 )	( 3 3 3 )	7	9	2	4.9995	5.86
	4:1	( 2 2 2 )	( 3 3 6 )	6	12	6	4.9997	6.33
	8:1	( 1 1 1 )	( 6 6 13 )	3	25	22	4.9996	5.49
1.5	1:16	( 3 3 5 )	( 3 3 5 )	11	11	0	5.0000	0.66
	1:8	( 3 3 5 )	( 3 3 5 )	11	11	0	5.0000	1.23
	1:4	( 3 3 5 )	( 3 3 5 )	11	11	0	5.0000	2.21
	1:2	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9996	3.41
	1:1	( 3 3 3 )	( 3 3 3 )	9	9	0	4.9996	4.91
	2:1	( 2 2 3 )	( 3 3 3 )	7	9	2	4.9993	6.06
	4:1	( 2 2 2 )	( 3 3 6 )	6	12	6	4.9995	6.49
	8:1	( 1 1 1 )	( 7 7 15 )	3	29	26	4.9997	5.91

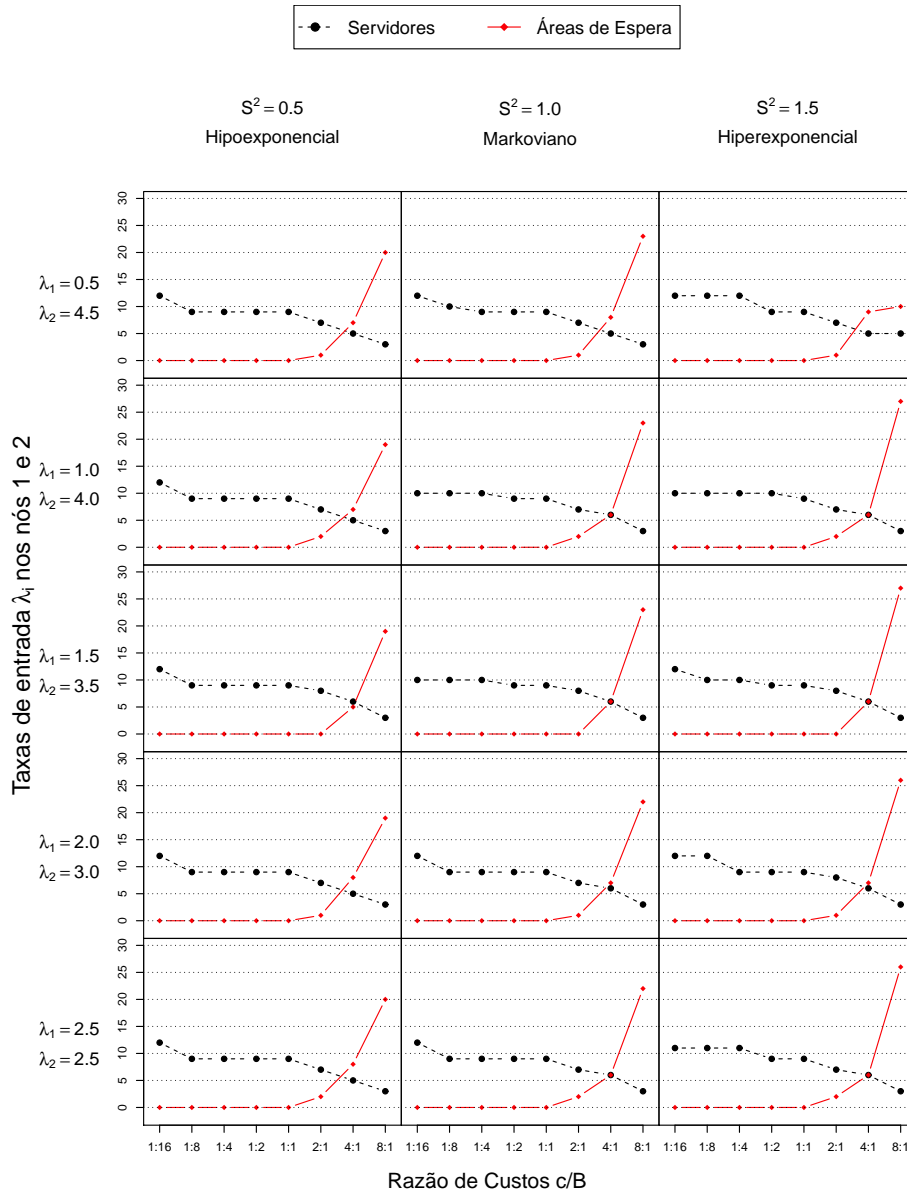


Figura 4.6: Alocação de servidores e áreas de espera para topologia fusão

apresentado na Figura 4.7. Na construção do diagrama, para cada topologia apresentada na seção anterior, as soluções foram estratificadas em  $s^2 = 0,5$  (sistemas hipoexponenciais) e  $s^2 \geq 1,0$  (sistemas markovianos e hiperexponenciais). Também foram estratificadas considerando as relações de custo  $C_c/C_B$  como inferiores a 4:1 ou iguais ou superiores a esta relação (isto é, 4:1 ou 8:1).

Nota-se que, quando se tem uma relação de custos entre servidores e áreas de espera  $C_c/C_B$  igual ou superior a 4:1, a alocação ótima total (considerando servidores e áreas de espera conjuntamente) tende a ser mais alta em relação à alocação ótima de servidores. Ademais, para relação de custos  $C_c/C_B$  iguais ou inferiores a 2:1, resultam-se soluções sem áreas de espera (ou, do inglês, *zero-buffer systems*), na forma de pares ordenados exatamente sobre a bissetriz do primeiro quadrante do plano cartesiano. O quadrado do coeficiente do tempo de serviço,  $s^2$ , parece não ter influência no padrão de alocação, uma vez que as soluções estratificadas por este fator não se concentram em nenhuma sub-região específica do primeiro quadrante. Resultados obtidos para outras topologias e diferentes taxas de chegada e probabilidades de bloqueio (não apresentados) conduzem a conclusões similares.

## 4.4 Observações Finais

Neste capítulo foram apresentados resultados experimentais do BCAP para diversas redes de filas configuradas nas topologias básicas *série*, *divisão* e *fusão*. As soluções para as configurações testadas foram bastante consistentes sob o aspecto da existência de um padrão na alocação de servidores e áreas de espera, independente das variações efetuadas nos parâmetros taxa de entrada ( $\Lambda$ ), quadrado do coeficiente de variação do tempo de serviço ( $s^2$ ), relação de custo servidor/áreas de espera ( $C_c/C_B$ ) e número de nós ( $N$ ). De modo geral, para relações de custo

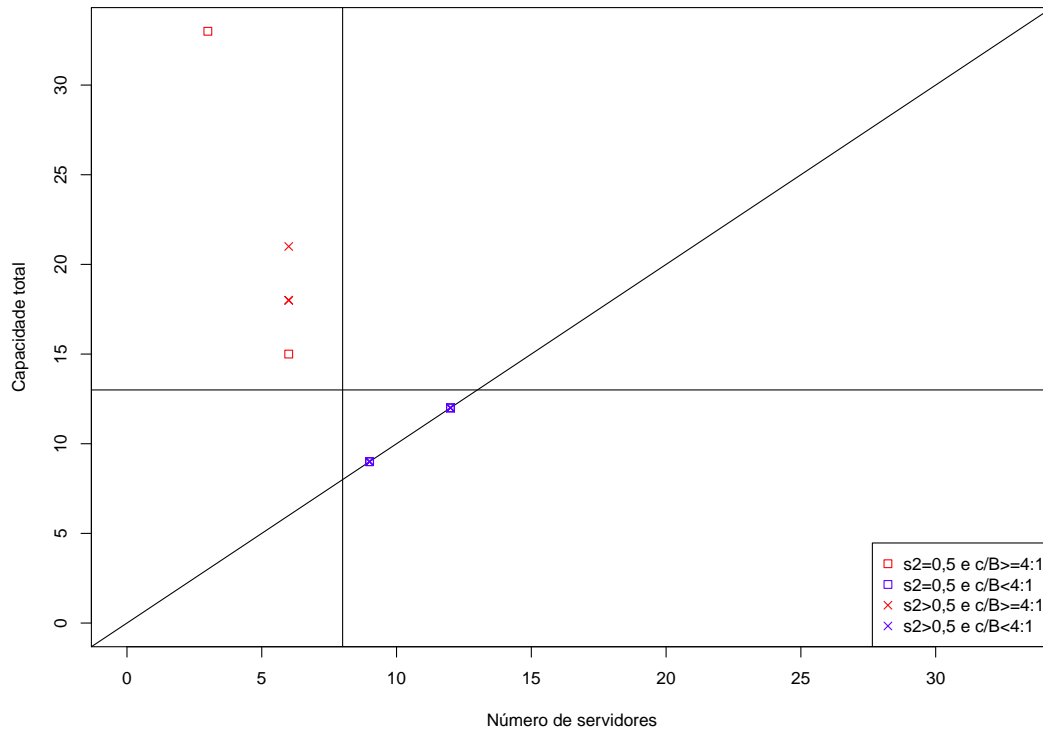


Figura 4.7: Alocação total em função de  $s^2$  e  $C_c/C_B$  para topologia série

$C_c/C_B$  iguais ou superiores a 4 : 1, observa-se crescimento na alocação de servidores em detrimento das áreas de espera. Relações de custo inferiores a 4 : 1, em sua grande maioria, resultaram em soluções sem áreas de espera. Constatou-se que para as soluções com áreas de espera, em redes na topologia divisão, o nó inicial necessita de mais recursos que a soma dos nós de destino, comportamento semelhante ao apresentado pelas redes com topologia fusão, nas quais o nó de destino recebe alocação superior à soma dos nós de partida.

Para completar, foi apresentada uma análise de possíveis fatores de influência nas alocações ótimas, que culminou na constatação de que, para os casos estudados, o padrão de alocações é afetado principalmente pela relação de custos  $C_c/C_B$ .



Por outro lado, a variabilidade dos tempos de serviço (medida pelo quadrado do coeficiente de variação do tempo de serviço,  $s^2$ ) não se configura como fator preponderante na alteração dos padrões de alocação.

# Capítulo 5

## Conclusões e Observações Finais

### 5.1 Sumário

Nesta dissertação, foi apresentada uma continuidade dos trabalhos apresentados em van Woensel et al. (2010) e Cruz & van Woensel (2014a;c), que tratam da análise do *problema de alocação conjunta de servidores e áreas de espera* (BCAP). Foi realizado um levantamento da bibliografia recente na área de alocação em redes das filas, com destaque para os textos que tratam do BCAP e da sua formulação em termos de um modelo de programação matemática, bem como estratégias para relaxação das complexas restrições impostas pelo problema.

Também foi apresentada uma revisão detalhada de uma metodologia clássica de análise de desempenho em filas simples e redes de filas, o *método da expansão generalizado* (GEM) e apresentada uma proposta de otimização em redes de filas combinando o GEM com o algoritmo de Powell, um algoritmo de otimização que dispensa o uso de derivadas e é apropriado para a resolução de problemas de programação não linear inteira como o BCAP.

A principal contribuição deste trabalho se refere à identificação de padrões de alocação de servidores em redes de filas configuradas sob as topologias em série,

divisão e fusão, pois os resultados gerados são bastante animadores, tanto em termos de robustez, quanto no que diz respeito aos padrões gerados. Observa-se que apesar das diferentes topologias, a alocação de servidores e áreas de espera apresenta resultados homogêneos para todos os casos. É indicada ausência de áreas de espera para situações nas quais os seus custos sejam superiores aos (ou não tão distantes dos) custos de servidores. Por outro lado, ocorre crescimento da alocação de tais áreas de espera quanto o custo do servidor é elevado. Tal constatação pode colaborar para a resolução do problema para redes de filas mais gerais, dado que se pode construir sistemas complexos baseados na junção de tais topologias básicas. Pode também auxiliar na compreensão de problemas reais modelados por redes de filas.

## 5.2 Propostas de Continuidade

Como trabalhos futuros, podem-se citar os seguintes:

- Análise dos padrões obtidos em redes de filas complexas combinando as três topologias básicas;
- Verificação da homogeneidade das soluções com oscilações nas taxas de serviço  $\mu_j$ ;
- Investigação das soluções do BCAP em redes de filas com tempos de chegada gerais e independentes, ou seja em redes do tipo  $G/G/c/K$ , na notação de Kendall (1953);
- Alocação ótima em redes de filas com ciclos, que podem modelar o retrabalho, dentre diversas outras possibilidades.

Investigações futuras também incluem a avaliação da qualidade da correção via *bootstrap* em outras medidas de desempenho da fila, tais como a probabilidade de

servidor ocioso,  $P(M = 0)$ , o tempo esperado no sistema,  $W$ , e o tempo médio na fila,  $W_q$ . Também outros tipos de filas devem ser investigados, tais como filas markovianas multi-servidoras infinitas,  $M/M/c$ , finitas,  $M/M/c/K$ , e assim por diante. Estes são apenas alguns tópicos para trabalhos futuros nesta instigante linha de pesquisa.

# Referências Bibliográficas

- Andriansyah, R., van Woensel, T., Cruz, F. R. B. & Duczmal, L. (2010). Performance optimization of open zero-buffer multi-server queueing networks, *Computers & Operations Research* **37**(8): 1472–1487.
- Balsamo, S., de Nitto Personé, V. & Onvural, R. (2001). *Analysis of Queueing Networks with Blocking*, Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Cruz, F. R. B., Duarte, A. R. & van Woensel, T. (2008). Buffer allocation in general single-server queueing networks, *Computers & operations research* **35**(11): 3581–3598.
- Cruz, F. R. B., Kendall, G., While, L., Duarte, A. R. & Brito, N. L. C. (2012). Throughput maximization of queueing networks with simultaneous minimization of service rates and buffers, *Mathematical Problems in Engineering* **2012**(Article ID 692593): 19 p.
- Cruz, F. R. B., MacGregor Smith, J. & Medeiros, R. O. (2005). An  $M/G/C/C$  state dependent network simulation model, *Computers & Operations Research* **32**(4): 919–941.
- Cruz, F. R. B., Oliveira, P. C. & Duczmal, L. (2010). State-dependent stochastic

- mobility model in mobile communication networks, *Simulation Modelling Practice and Theory* **18**(3): 348–365.
- Cruz, F. R. B. & van Woensel, T. (2014a). Buffers and servers allocation in general finite queueing networks, *Proceeding Series of the Brazilian Society of Computational and Applied Mathematics*, v. 2, p. 0101081–0101086.
- Cruz, F. R. B. & van Woensel, T. (2014b). Finite queueing modeling and optimization: A selected review, *Journal of Applied Mathematics* **2014**(Article ID 374962).
- Cruz, F. R. B. & van Woensel, T. (2014c). Joint buffer and server allocation in general finite queueing networks, *XLVI Simpósio Brasileiro de Pesquisa Operacional - XLVI SBPO [CD-ROM]*, Salvador, Brasil, p. 2213–2220.
- Dallery, Y. & Gershwin, S. B. (1992). Manufacturing flow line systems: A review of models and analytical results, *Queueing Systems* **12**(1-2): 3–94.
- Daskalaki, S. & MacGregor Smith, J. (2004). Combining routing and buffer allocation problems in series-parallel queueing networks, *Annals of Operations Research* **125**(1-4): 47–68.
- Dorda, M. & Teichmann, D. (2013). Modelling of freight trains classification using queueing system subject to breakdowns, *Mathematical Problems in Engineering* **2013**(Article ID 307652): 11 p.
- Himmelblau, D. M. (1972). *Applied Nonlinear Programming*, McGraw-Hill Book Company, New York.
- Jackson, J. R. (1957). Networks of waiting lines, *Operations Research* **5**(4): 518–521.

- Jain, S. & MacGregor Smith, J. (1994). Open finite queueing networks with  $M/M/C/K$  parallel servers, *Computers & Operations Research* **21**(3): 297–317.
- Kendall, D. G. (1953). Stochastic processes occurring in the theory of queues and their analysis by the method of embedded Markov chains, *Annals Mathematical Statistics* **24**: 338–354.
- Kerbache, L. & MacGregor Smith, J. (1987). The generalized expansion method for open finite queueing networks, *European Journal of Operational Research* **32**: 448–461.
- Kerbache, L. & MacGregor Smith, J. (1988). Asymptotic behavior of the expansion method for open finite queueing networks, *Computers & Operations Research* **15**(2): 157–169.
- Kerbache, L. & MacGregor Smith, J. (2000). Multi-objective routing within large scale facilities using open finite queueing networks, *European Journal of Operational Research* **121**(1): 105–123.
- Kimura, T. (1996). A transform-free approximation for the finite capacity  $M/G/s$  queue, *Operations Research* **44**(6): 984–988.
- Kuehn, P. (1979). Approximate analysis of general queueing networks by decomposition, *IEEE Transactions on Communications* **27**(1): 113–126.
- Labetoulle, J. & Pujolle, G. (1980). Isolation method in a network of queues, *IEEE Transactions on Software Engineering* **SE-6**(4): 373–381.
- Lemaréchal, C. (2007). The omnipresence of lagrange, *Annals of Operations Research* **153**(1): 9–27.
- MacGregor Smith, J. (2003).  $M/G/c/K$  blocking probability models and system performance, *Performance Evaluation* **52**(4): 237–267.

- MacGregor Smith, J. & Cruz, F. R. B. (2005). The buffer allocation problem for general finite buffer queueing networks, *IIE Transactions* **37**(4): 343–365.
- MacGregor Smith, J., Cruz, F. R. B. & van Woensel, T. (2010). Optimal server allocation in general, finite, multi-server queueing networks, *Applied Stochastic Models in Business & Industry* **26**(6): 705–736.
- Perros, H. G. (1984). Queueing networks with blocking: A bibliography, *ACM Sigmet.* **12**: 8–12.
- Reiser, M. & Kobayashi, H. (1974). Accuracy of the diffusion approximation for some queueing systems, *IBM Journal of Research and Development* **18**(2): 110–124.
- Suri, R. (1985). An overview of evaluative models for flexible manufacturing systems, *Annals of Operations Research* **3**: 13–21.
- van Woensel, T., Andriansyah, R., Cruz, F. R. B., MacGregor Smith, J. & Kerbache, L. (2010b). Buffer and server allocation in general multi-server queueing networks, *International Transactions in Operational Research* **17**(2): 257–286.
- van Woensel, T. & Cruz, F. R. B. (2014). Optimal routing in general finite multi-server queueing networks, *PLoS ONE* **9**(7): e102075.
- van Woensel, T. & Vandaele, N. (2006). Empirical validation of a queueing approach to uninterrupted traffic flows, *JOR* **4**(1): 59 – 72.
- van Woensel, T. & Vandaele, N. (2007). Modelling traffic flows with queueing models: A review, *Asia-Pacific Journal of Operational Research* **24**(4): 435–461.
- Zhang, G., Lu, J. & Gao, Y. (2015). *Multi-Level Decision Making*, Springer.