

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

APLICAÇÃO DE MODELOS DE REDES DE FILAS ABERTAS NO PROJETO
E PLANEJAMENTO DE SISTEMAS DISCRETOS DE MANUFATURA

CLAUDIO ROGERIO NEGRI DA SILVA

TESE DE DOUTORADO

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

**APLICAÇÃO DE MODELOS DE REDES DE FILAS ABERTAS NO PROJETO
E PLANEJAMENTO DE SISTEMAS DISCRETOS DE MANUFATURA**

Claudio Rogerio Negri da Silva

**Tese de Doutorado apresentada ao
Programa de Pós-Graduação em
Engenharia de Produção da
Universidade Federal de São
Carlos, como parte dos requisitos
para obtenção do título de Doutor
em Engenharia de Produção.**

Orientador: Prof. Dr. Reinaldo Morabito Neto

Agência Financiadora: FAPESP

SÃO CARLOS

2005

**Ficha catalográfica elaborada pelo DePT da
Biblioteca Comunitária/UFSCar**

S586am

Silva, Claudio Rogerio Negri da.

Aplicação de modelos de redes de filas abertas no projeto e planejamento de sistemas discretos de manufatura / Claudio Rogerio Negri da Silva. -- São Carlos : UFSCar, 2005.

243 p.

Tese (Doutorado) -- Universidade Federal de São Carlos, 2005.

1. Pesquisa operacional. 2. Fábrica focalizada. 3. Rede de filas aberta. 4. Modelos de otimização. 5. Projeto de sistema de manufatura. I. Título.

CDD: 658.4034 (20^a)



UNIVERSIDADE FEDERAL DE SÃO CARLOS
Centro de Ciências Exatas e de Tecnologia
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO
Rod. Washington Luís, Km. 235 - Caixa Postal 676
CEP. 13565-905 - São Carlos - SP - Brasil
Fones: (016) 3351-8238 – (ramal 232)
Fax: (016) 3351-8238 (r. 232)
E-mail: ppgep@dep.ufscar.br

FOLHA DE APROVAÇÃO

Aluno: CLAUDIO ROGERIO NEGRI DA SILVA

TESE DE DOUTORADO DEFENDIDA E APROVADA EM 29/04/2005
PELA COMISSÃO JULGADORA:

PROF. DR. REINALDO MORÁBITO NETO
(Orientador - PPGE/UFSCar)

PROF. DR. SOLON VENÂNCIO DE CARVALHO
(INPE/LAC)

PROF. DR. HUGO TSEGUNOBU YOSHIDA YOSHIKAZI
(POLI/USP)

PROF. DR. MARCOS NASCIMENTO MAGALHÃES
(IME/USP)

PROF. DR. EDUARDO VILA GONÇALVES FILHO
(EESC/USP)

PROF. DR. HENRIQUE PACCA LOUREIRO LUNA
(UFAL)

Presidente da Coordenação de Pós-Graduação
Prof. Dr. Dário Henrique Alliprandini

Time of Your Life

*Another turning point a fork stuck in the road
Time grabs you by the wrist directs you where to go
So make the best of this test and don't ask why
It's not a question but a lesson learned in time*

*It's something unpredictable
But in the end is right
I hope you had the time of your life*

*So take the photographs and still frames in your mind
Hang it on a shelf in good health and good time
Tatoos of memories and dead skin on trial
For what it's worth it was worth all the while*

*It's something unpredictable
But in the end is right
I hope you had the time of your life*

*It's something unpredictable
But in the end is right
I hope you had the time of your life*

*It's something unpredictable
But in the end is right
I hope you had the time of your life*

DEDICATÓRIA

A meus pais, irmãos e agregados à família (Eliel, “Elielzinho”, Ana Beatriz, Edléssia e Alice) pelos gestos carinhosos. Em especial, gostaria de agradecer ao meu pai (“Seu Pereira”), mãe (“D. Cidinha”) e irmãos (Vitor, Gislaine e Wellington) pela força e esperança em mim depositada, pelo apoio incondicional e irrestrito, pelas palavras certas e pelos valores ensinados. Três palavras resumem todos meus sentimentos de admiração, respeito e gratidão a vocês: EU OS AMO. Vocês foram o meu porto seguro nos momentos mais difíceis, e a vocês dedico o mérito deste trabalho.

A Deus por sua infindável fonte de bondade e sabedoria.

AGRADECIMENTOS

Ao amigo e orientador Reinaldo Morabito Neto que, com paciência e humildade, soube me direcionar com maestria em momentos cruciais da pesquisa, mostrando que além de excelente professor e pesquisador também é uma pessoa extremamente humana, digna de minha admiração e respeito.

Ao Prof Manoel Fernando Martins (DEP-UFSCar) pela indicação da empresa Imart Marrara, utilizada como estudo de caso nesta tese. Particularmente, gostaria de agradecer imensamente ao eng. Sr. Iveraldo (responsável pelo PCP) e ao diretor industrial Sr. Nelson Marrara pelo acesso permitido às informações e tempo dispensado, cruciais para o desenvolvimento desta pesquisa.

Aos membros da banca de qualificação (Prof Solon Venâncio de Carvalho, Prof Marcos Nascimento Magalhães, Prof Hugo T Y Yoshizaki e Prof Eduardo Vila Gonçalves Filho) pelas contribuições e sugestões ao desenvolvimento da pesquisa.

Aos meus amigos Júnior (“Jack”), Alex (“Profeta do Caos”) e Gustavo (“Japa”) pelas orelhas e ombros amigos nos momentos difíceis, imprescindíveis para a continuidade deste trabalho. Muito obrigado pela paciência e sinceridade.

À Leonara (“Léo”) por compartilhar anos agradáveis, pela atenção, sinceridade e dedicação em boa parte desta caminhada, dignas de minha gratidão e amizade eternas.

À Larissa (“Lá”) pelo carinho, apoio, compreensão e paciência na reta final deste trabalho.

Aos amigos do “Trio Canalha” (Rui Tadashi, Ronaldo Mardegan e Marcelo Rui), à Regiane Máximo e à Simone pelo relacionamento de trabalho sadio que, de forma indireta, influenciou positivamente no desenvolvimento deste trabalho. Aos amigos de viagem (turma da “latrina”) Belluzzo, Norberto, Flávio e Ana Claudia. Obrigado pelos papos prazerosos, momentos de riso e festas *bus*.

Aos amigos de laboratório Ana Paula Iannoni, Lilian, Beto, Renata Takeda, Eli e Gisele pelo auxílio e momentos de distração no decorrer da pesquisa.

À FAPESP pela seriedade e apoio financeiro, primordiais ao desenvolvimento desta pesquisa.

SUMÁRIO

1 INTRODUÇÃO	1
1.1 Objetivos e Justificativas.....	3
1.2 Estrutura da Tese.....	4
2 MODELOS DE AVALIAÇÃO DE DESEMPENHO	7
2.1 Sistemas de Manufatura Discretos.....	7
2.1.1 <i>Job-shop</i>	7
2.1.2 <i>Flow-shop</i>	8
2.1.3 Fábrica Focalizada e Manufatura Celular.....	8
2.1.4 <i>Job-shop versus</i> Manufatura Celular.....	10
2.2 Redes de Filas.....	14
2.3 Redes de Filas Markovianas (Método Exato de Decomposição).....	19
2.3.1 Redes de Jackson.....	20
2.3.2 Redes de Kelly e BCMP.....	25
2.4 Método Aproximado de Decomposição.....	27
2.5 Redes Generalizadas (Método Aproximado de Decomposição).....	30
2.5.1 Modelos de Redes com Filas $GI/G/1$, Classe Única e Rota Probabilística.....	31
2.5.1.1 Passo 1: Análise das interações entre estações da <i>OQN</i>	32
2.5.1.2 Passo 2: Decomposição da rede de filas em subsistemas com uma única estação e avaliação das medidas de desempenho para cada estação.....	38
2.5.1.3 Passo 3: Recomposição dos resultados obtidos pelas estações decompostas com avaliação das medidas de desempenho da rede.....	39
2.5.2 Modelos de Redes com Filas $GI/G/m$, Classe Única e Rota Probabilística.....	39
2.5.2.1 Passo 1: Análise das interações entre estações da <i>OQN</i>	40
2.5.2.2 Passo 2: Decomposição da rede de filas em subsistemas com uma única estação e avaliação das medidas de desempenho para cada estação.....	41
2.5.2.3 Passo 3: Recomposição dos resultados obtidos pelas estações	

decompostas com avaliação das medidas de desempenho da rede.....	43
2.5.3 Modelos de Redes com Filas $GI/G/m$, Múltiplas Classes e Rota Determinística.....	43
2.5.3.1 Passo 1: Análise das interações entre estações da OQN	46
2.5.3.2 Passo 2: Decomposição da rede de filas em subsistemas com uma única estação e avaliação das medidas de desempenho para cada estação.....	47
2.5.3.3 Passo 3: Recomposição dos resultados obtidos pelas estações decompostas com avaliação das medidas de desempenho da rede.....	47
2.5.4 Interferência entre Classes e Aproximação de Tráfego Leve.....	48
2.6 Modelos de Simulação.....	56
3 MODELOS DE OTIMIZAÇÃO EM REDES DE FILAS: ALOCAÇÃO E REALOCAÇÃO DE SERVIDORES.....	60
3.1 Problemas Tático-Estratégicos Relacionados ao Projeto/Planejamento de Sistemas Discretos de Manufatura.....	60
3.1.1 Desempenho Desejado do Sistema ($SP1$).....	62
3.1.2 Desempenho Ótimo do Sistema ($SP2$).....	63
3.1.3 Partição da Instalação ($SP3$).....	63
3.2 Formulação dos Problemas $SP1$ e $SP2$	64
3.3 Modelos $\cdot / J / \cdot / \cdot$	66
3.3.1 Modelos $\cdot / J / \cdot / R$	66
3.3.2 Modelos $\cdot / J / M / N$	68
3.3.2.1 Modelo $SP1.1/J/M/N$	69
3.3.2.2 Modelo $SP2.1/J/M/N$	71
3.4 Modelos $\cdot / G / \cdot / \cdot$	72
3.4.1 Modelos $\cdot / G / \cdot / R$	72
3.4.1.1 Modelo $SP1.1/G/S/R$	73
3.4.1.2 Modelo $SP2.1/G/S/R$	77
3.4.1.3 Modelo $\cdot / G / \cdot / R$ com alternativas discretas.....	79
3.4.2 Modelos $\cdot / G / M / N$	84
3.4.2.1 Modelo $SP1.1/G/M/N$	85

3.4.2.2 Modelo <i>SP2.1/G/M/N</i>	86
4. RESULTADOS COM ESTUDO DE CASO	89
4.1 Empresa Estudada.....	89
4.2 Representação do Sistema de Manufatura nos Modelos de Redes de Filas e de Simulação.....	90
4.2.1 O Sistema de Manufatura como um Conjunto de Classes de Produtos, Estações de Trabalho e Roteiros de Fabricação.....	90
4.2.2 Caracterização do Processo de Chegadas Externas.....	94
4.2.3 Caracterização do Processo de Serviço.....	98
4.3 Custos e Alternativas Discretas para Alocação de Capacidade.....	101
4.4 Avaliação do Desempenho da Configuração Inicial da Rede.....	104
4.4.1 Avaliação da Sensibilidade das Aproximações a Alterações nas Taxas de Chegadas.....	113
4.4.2 Sensibilidade das Medidas de Desempenho da Rede a Variações nas Disciplinas de Atendimento nas Estações.....	115
4.5 Alocação de Capacidade Ótima na Rede.....	117
4.5.1 Sensibilidade do Modelo a Variações nos Parâmetros de Variabilidade..	125
4.5.2 Sensibilidade do Modelo a Variações nas Taxas de Chegadas Externas..	128
4.5.3 Sensibilidade do Modelo a Variações no <i>Mix</i> de Produção.....	130
4.6 Comentários Gerais Acerca dos Resultados para o Estudo de Caso.....	134
5 MODELOS DE OTIMIZAÇÃO EM REDES DE FILAS: PARTIÇÃO DA INSTALAÇÃO	139
5.1 O Sistema de Produção Modelado.....	140
5.2 Medidas de Complexidade.....	140
5.2.1 Complexidade do Ponto de Vista de Gestão do Produto: Previsibilidade.	141
5.2.2 Complexidades do Ponto de Vista da Estação.....	142
5.3 Definição do Problema de Redução de Complexidade (PRC).....	147
5.3.1 Problema de Redução de Complexidade PRC-1.....	148

5.3.2 Problema de Redução de Complexidade PRC-2.....	148
5.3.3 Problema de Redução de Complexidade PRC-3.....	149
5.4 Solução do Problema de Redução de Complexidade (PRC).....	149
6. RESULTADOS COM PROBLEMA DE PARTIÇÃO DA INSTALAÇÃO	153
6.1 Problema de Redução de Complexidade 1 (PRC-1).....	157
6.1.1 Rede Formada por Um Estágio Único $GI/G/m$	158
6.1.1.1 Experimentos com Rede Ilustrativa (Rede 1).....	158
6.1.1.2 Experimentos com Rede Mais Realista (Rede 3).....	163
6.1.2 Rede Formada por Vários Estágios ($GI/G/m$).....	172
6.1.2.1 Experimentos com Rede Ilustrativa (Rede 2).....	172
6.1.2.2 Experimentos com Rede mais Realista (Rede 4).....	176
6.2 Problema de Redução de Complexidade 2 (PRC-2).....	186
6.2.1 Rede Formada por Um Estágio Único $GI/G/m$ (Rede 3).....	186
6.2.2 Rede Formada por Vários Estágios $GI/G/m$ (Rede 4).....	195
6.3 Problema de Redução de Complexidade 3 (PRC-3).....	207
6.3.1 Rede Formada por Um Único Estágio $GI/G/m$ (Rede 3).....	207
6.3.2 Rede Formada por Vários Estágios $GI/G/m$ (Rede 4).....	209
7 CONCLUSÕES E PERSPECTIVAS FUTURAS	213
7.1 Conclusões.....	213
7.2 Perspectivas de Pesquisa Futura.....	218
REFERÊNCIAS	221
APÊNDICE A	225
APÊNDICE B	230
APÊNDICE C	236

LISTA DE TABELAS

TABELA 4.1 – Tamanho da amostra, tamanho médio e desvio padrão do lote (entre colchetes), $E(a'_k)$, λ'_k , ca'_k , n_k e estações visitadas para operação l da rota da classe k , n_{kl}	97
TABELA 4.2 – Tempos médios de serviço de um lote na operação l do roteiro da classe k , $E(s_{kl})$	99
TABELA 4.3 – Custo acumulado nas estações de um lote de produto qualquer para a rede estudada.....	103
TABELA 4.4 – Alternativas de adição/remoção de capacidade nas estações j : taxa de produção (μ_{jt}) e custo (f_{jt}) na estação j e alternativa t	103
TABELA 4.5 – Comparação entre os <i>leadtimes</i> médios amostrais e obtidos por simulação considerando capacidade efetiva igual a 100% e 95% da nominal para classes com tamanho de amostra maior ou igual a 10.....	105
TABELA 4.6 – Dados para a rede com classe única (classe agregada).....	106
TABELA 4.7 – Resultados para a classe agregada nas estações j utilizando as aproximações do sistema 2.....	107
TABELA 4.8 – Número médio de lotes, $E(L_k)$, obtidos por simulação e pelas aproximações.....	108
TABELA 4.9 – <i>Leadtime</i> médio obtido por simulação e aproximações e desvios relativos dos <i>leadtimes</i> obtidos pelas aproximações com relação ao obtido por simulação.....	109
TABELA 4.10 – Desvio-padrão do <i>leadtime</i> obtido para cada classe utilizando as aproximações.....	110
TABELA 4.11 – Número médio de lotes para cada classe de produto, ao supor que a rede é determinística ($E(L_k)^D$), Jackson ($E(L_k)^J$), ou Genérica ($E(L_k)^G$), sistema 2) com os respectivos desvios percentuais relativos à simulação ($E(L_k)^0$).....	111
TABELA 4.12 – <i>Leadtime</i> médio para cada classe de produto, ao supor que a rede é determinística ($E(T_k)^D$), Jackson ($E(T_k)^J$), ou Genérica ($E(T_k)^G$), sistema 2) com os respectivos desvios percentuais relativos à simulação ($E(T_k)^0$).....	111
TABELA 4.13 – Número médio de lotes, $E(L_k)$, obtidos da simulação e das	

aproximações, ao reduzir em 10% as taxas médias de chegadas de produtos.....	114
TABELA 4.14 – <i>Leadtime</i> médio obtido por simulação e aproximações, ao reduzir em 10% as taxas médias de chegadas dos produtos.....	114
TABELA 4.15 – Número médio de lotes, $E(L_k)$, e <i>leadtime</i> médio, $E(T_k)$, para cada classe na rede com disciplinas de atendimento <i>FCFS</i> e <i>SPT</i>	116
TABELA 4.16 – <i>Work-in-process (WIP)</i> em cada estação e total para a configuração inicial da rede.....	117
TABELA 4.17 – Alternativa t escolhida pela solução do modelo $SPI.1/G/S/R^{discreto}$ para limitante superior do <i>WIP</i> igual a \$27000 e investimento de -\$68500.....	120
TABELA 4.18 – Alternativa t escolhida pela solução do modelo $SPI.1/G/S/R^{discreto}$ para limitante superior do <i>WIP</i> igual a \$26000 e investimento de -\$6500.....	120
TABELA 4.19 – Limitantes superiores para o <i>leadtime</i> de cada classe de produtos, $E(T_k)$, nas configurações com necessidade de $F=\$180500$ e $F=\$35500$ de investimento em capacidade.....	122
TABELA 4.20 – Tempos médios de espera nas filas das estações para a configuração inicial ($F=\$0,0$) e com investimento de $F=\$35500$	124
TABELA 4.21 – Alternativas ótimas do modelo $SPI.2/G/S/R^{discreto}$ para $F=\$35500$ e $F=\$180500$	124
TABELA 4.22 – Alternativas ótimas para $F=\$180500$ obtidas com os modelos $SPI.1/G/S/R^{discreto}$ e $SPI.2/G/S/R^{discreto}$	124
TABELA 4.23 – Alternativas ótimas do modelo $SPI.1/G/S/R^{discreto}$ para limitante superior do <i>WIP</i> igual a \$22000 para os cenários 1.1 e 1.2.....	126
TABELA 4.24 – <i>Leadtime</i> ótimo do modelo $SPI.1/G/S/R^{discreto}$ para cada classe de produto para limitante superior do <i>WIP</i> igual a \$22000 para os cenários 1.0, 1.1 e 1.2.....	127
TABELA 4.25 – Alternativas ótimas do modelo $SPI.1/G/S/R^{discreto}$ para limitante superior do <i>WIP</i> igual a \$25400 para os cenários 2.0 e 2.1.....	129
TABELA 4.26 – Alternativas ótimas para os cenários 3.0, 3.1 e 3.2 com limitante superior para o <i>WIP</i> de \$22000 no modelo $SPI.1/G/S/R^{discreto}$	132

TABELA 6.1 – Dados para as duas classes de produtos na rede formada por um único estágio $GI/G/m$ (Rede 1).....	154
TABELA 6.2 – Dados para as duas classes de produtos na rede formada por dois estágios $GI/G/m$ (Rede 2).....	154
TABELA 6.3 – Dados para as 10 classes de produtos para a rede com um único estágio $GI/G/m$ (Rede 3).....	155
TABELA 6.4 – Parâmetros da rede referentes ao roteiro (n_{kl}) , ao número de operações (n_k) , ao peso da variância do <i>leadtime</i> (w_k) , ao coeficiente quadrático de variação de intervalo de tempo entre chegadas (ca'_k) e à taxa média de chegada (λ'_k) para as 10 classes de produtos (Rede 4).....	157
TABELA 6.5 – Parâmetros referentes ao tempo médio de serviço da classe k e operação $l, E(s_{kl})$, para as 10 classes de produtos (Rede 4).....	157
TABELA 6.6 – Número de máquinas (MQ) e variância da rede $V(T)$ obtidos para as diferentes partições em cada iteração $(iter)$ para $T^* = 1,0$ na Rede 1.....	158
TABELA 6.7 – Medidas obtidas para as partições [1],[2] e [1,2], com número de máquinas na Rede 1 igual a 2.....	159
TABELA 6.8 – Variância do <i>leadtime</i> da rede, $V(T)$, inventário em processo, $E(L_k)$, e <i>leadtime</i> , $E(T_k)$, para as classes 1 e 2 e para a rede, $E(T)$, para duas partições da Rede 1 com $ca'_2 = 0,4$ e número de máquinas igual a 2.....	160
TABELA 6.9 – Variância do <i>leadtime</i> , $V(T)$, ao mudar a variabilidade do processo de chegadas da Rede 1, mantendo o número de máquinas da rede igual a 2.....	161
TABELA 6.10 – Variância do <i>leadtime</i> , $V(T)$, ao mudar o nível de utilização das estações da Rede 1, mantendo o número de máquinas da rede igual a 2.....	162
TABELA 6.11 – Número total de máquinas (MQ) e variância do <i>leadtime</i> da rede $V(T)$ obtida para as diferentes partições em cada iteração $(iter)$ do algoritmo 6 para limitante superior da variância do <i>leadtime</i> $T^* = 0,2$ (Rede 3).....	164
TABELA 6.12 – Número de máquinas (MQ) e variância do <i>leadtime</i> da rede $V(T)$ obtidos para partições fixas (PF) e $T^* = 0,2$ (Rede 3).....	165
TABELA 6.13 – <i>Leadtime</i> , $E(T_k)$, e número médio de itens, $E(L_k)$, obtidos para as classes e para a rede ($E(T)$ e $E(L)$) nas partições [1,2],[3-6],[7-9],[10] e [1-	

10], com $T^*=0,2$ e número de máquinas na Rede 3 igual a 16.....	166
TABELA 6.14 – Número total de máquinas (MQ) e variância do <i>leadtime</i> da rede $V(T)$ obtida para as diferentes partições em cada iteração (<i>iter</i>) do algoritmo 6 para $T^*=0,2$ e processo de chegadas das classes 7 – 10 determinístico (Rede 3 modificada).....	167
TABELA 6.15 – Número de máquinas (MQ) e variância do <i>leadtime</i> da rede $V(T)$ obtidos para partições fixas (PF), com $T^*=0,2$ e processo de chegadas das classes 7 – 10 determinístico (Rede 3 modificada).....	168
TABELA 6.16 – Número total de máquinas (MQ) e variância do <i>leadtime</i> da rede $V(T)$ obtida para as diferentes partições em cada iteração (<i>iter</i>) do algoritmo 6 para $T^*=0,2$ e processo de chegadas das classes 1 e 2 determinístico (Rede 3 modificada).....	169
TABELA 6.17 – Número de máquinas (MQ) e variância do <i>leadtime</i> da rede $V(T)$ obtidos para partições fixas (PF) com $T^*=0,2$ e processo de chegadas das classes 1 e 2 determinístico (Rede 3 modificada).....	170
TABELA 6.18 – <i>Leadtime</i> , $E(T_k)$, e número médio de itens, $E(L_k)$, obtidos para as classes e para a rede ($E(T)$ e $E(L)$) nas partições [1,2],[3-6],[7-10] e [1,2],[3-10], com $T^*=0,2$ e número de máquinas na Rede 3 modificada (classes 1 e 2 com processo de chegadas determinístico) igual a 15.....	171
TABELA 6.19 – <i>Leadtime</i> , $E(T_k)$, e número médio de itens, $E(L_k)$, obtidos para as classes e para a rede ($E(T)$ e $E(L)$) nas partições [1,2],[3-6],[7-10] e [1-10], com $T^*=0,2$ na Rede 3 modificada (classes 1 e 2 com processo de chegadas admitido determinístico).....	172
TABELA 6.20 – Número total de máquinas (MQ) e variância do <i>leadtime</i> da rede $V(T)$ obtidos para as diferentes partições em cada iteração (<i>iter</i>) do algoritmo 6 para $T^*=0,2$ (Rede 2).....	173
TABELA 6.21 – Coeficiente de variação do processo de chegadas e nível de utilização das estações das linhas [1] e [2] (Rede 2).....	173
TABELA 6.22 – Número de máquinas (MQ) e variância da rede $V(T)$ obtidos para partições fixas (PF) e $T^*=0,2$ (Rede 2).....	174
TABELA 6.23 – <i>Leadtime</i> , $E(T_k)$, e número médio de itens, $E(L_k)$, obtidos para	

as classes e para a rede ($E(T)$ e $E(L)$) nas partições [1],[2] e [1,2], com $T^*=0,2$ na Rede 2.....	174
TABELA 6.24 – Número total de máquinas (MQ) e variância do <i>leadtime</i> da rede $V(T)$ obtidos para as diferentes partições em cada iteração (<i>iter</i>) do algoritmo 6 para $T^*=0,2$ e $ca'_2=0,250$ (Rede 2 modificada).....	175
TABELA 6.25 – Número de máquinas (MQ) e variância do <i>leadtime</i> da rede $V(T)$ obtidos para partições fixas (<i>PF</i>) com $T^*=0,2$ e $ca'_2=0,250$ (Rede 2 modificada).....	175
TABELA 6.26 – Coeficiente de variação do processo de chegadas e nível de utilização das estações das linhas [1] e [2] com $ca'_2=0,250$ (Rede 2 modificada).	176
TABELA 6.27 – Número total de máquinas (MQ) e variância do <i>leadtime</i> da rede $V(T)$ obtidos para as diferentes partições em cada iteração (<i>iter</i>) do algoritmo 6 para $T^*=8,0$ (Rede 4).....	177
TABELA 6.28 – <i>Leadtime</i> médio, $E(T_k)$, número médio de itens para cada classe, $E(L_k)$, e total na Rede 4, $E(T)$ e $E(L)$, na partição [1,2,3,4,5,6,7,8],[9,10]..	178
TABELA 6.29 – Número de máquinas (MQ) e variância do <i>leadtime</i> da rede $V(T)$ obtidos para partições fixas (<i>PF</i>) com $T^*=8,0$ (Rede 4).....	178
TABELA 6.30 – <i>Leadtime</i> médio, $E(T_k)$, número médio de itens para cada classe, $E(L_k)$, e para a Rede 4, $E(L)$ e $E(T)$, na partição [1,2,3,4,5,6,7,8,9,10].....	179
TABELA 6.31 – Número de máquinas (MQ) e variância do <i>leadtime</i> ($V(T)$) da Rede 4 com $T^*=8,0$ para a partição obtida pelo algoritmo 6 e para a partição com uma linha única.....	180
TABELA 6.32 – Nível de utilização (ρ_j) e coeficiente quadrático dos intervalos de tempo entre chegadas (ca_j) das 13 estações das linhas [1,2,3,4,5,6,7,8] e [9,10] para o caso com $ca'_9 = 0$ e $ca'_{10} = 0$ (Rede 4 modificada com $T^*=8,0$).....	181
TABELA 6.33 – Nível de utilização (ρ_j) e coeficiente quadrático de variação do intervalo de tempo entre chegadas (ca_j) nas 13 estações das linhas [1,2,3,4,5,6,7,8] e [9,10] para o caso com $2^*E(s_{k9})$ e $2^*E(s_{k10})$ (Rede 4 modificada com $T^*=8,0$).....	182
TABELA 6.34 – Número total de máquinas (MQ) e variância da rede $V(T)$	

obtidos para as diferentes partições em cada iteração (<i>iter</i>) do algoritmo 6 para $T^*=8,0$ na Rede 4, utilizando a aproximação de Bitran e Tirupati (1988) na Rede 4.....	183
TABELA 6.35 – <i>Leadtime</i> médio, $E(T_k)$, número médio de itens para cada classe, $E(L_k)$, e para a rede, $E(L)$ e $E(T)$, na partição [1,2,3,4,5,6,7,8,9,10] ao utilizar a aproximação de Bitran e Tirupati (1988) na Rede 4.....	184
TABELA 6.36 – Número de máquinas (MQ) e variância do <i>leadtime</i> ($V(T)$) ao mudar alguns parâmetros do caso base (Rede 4 com $T^*=8,0$), utilizando a aproximação de Bitran e Tirupati (1988) para a partição obtida pelo algoritmo 6 e para a partição com uma linha única.....	185
TABELA 6.37 – Número total de máquinas (MQ) obtido para as diferentes partições em cada iteração (<i>iter</i>) do algoritmo 6 para $G_{jp}=1,0$ (Rede 3).....	187
TABELA 6.38 – Número total de máquinas (MQ) obtido em diferentes partições (PF) para $G_{jp}=1,0$ (Rede 3).....	188
TABELA 6.39 – Tempo médio de espera em fila, dado que espera, $E(W_{q_{jp}} W_{q_{jp}}>0)$, e número médio de itens no sistema para as linhas ($E(L_p)$) da partição [1,2],[3,4,5,6],[7,8,9,10] obtidos por simulação para a Rede 3 com $G_{jp}=1,0$	190
TABELA 6.40 – <i>Leadtime</i> médio, $E(T_k)$, e número médio de itens, $E(L_k)$, para as 10 classes de produtos e para a rede ($E(L)$ e $E(T)$) na partição [1,2],[3,4,5,6],[7,8,9,10] da Rede 3 obtidos por simulação e pelas aproximações com $G_{jp}=1,0$	190
TABELA 6.41 – Número total de máquinas (MQ) obtido para as diferentes partições em cada iteração (<i>iter</i>) do algoritmo 6 para $G_{jp}=0,6$ (Rede 3).....	191
TABELA 6.42 – Número total de máquinas (MQ) obtido em diferentes partições (PF) para $G_{jp}=0,6$ (Rede 3).....	192
TABELA 6.43 – Tempo médio de espera em fila, dado que espera, $E(W_{q_{jp}} W_{q_{jp}}>0)$, e número médio de itens no sistema para as linhas ($E(L_p)$) da partição [1,2],[3,4,5,6],[7,8,9,10] obtidos por simulação para a Rede 3 com $G_{jp}=0,6$	193
TABELA 6.44 – <i>Leadtime</i> médio, $E(T_k)$, e número médio de itens, $E(L_k)$, para	

as 10 classes de produtos e para a rede ($E(L)$ e $E(T)$) obtidos por simulação e pelas aproximações para a Rede 3 com $G_{jp}=0,6$	193
TABELA 6.45 – Número de máquinas (MQ) ao mudar alguns parâmetros do caso base (Rede 3) para a partição obtida pelo algoritmo 6 com $G_{jp}=1,0$ e para a partição com uma linha única.....	194
TABELA 6.46 – Tempos médios de processamento para cada classe k para os casos base, 1, 2 e 3 (Rede 3).....	195
TABELA 6.47 – Tempo médio de espera em fila, dado que espera, $E(Wq_{jp} Wq_{jp}>0)$, e número médio de itens nas linhas ($E(L_p)$), obtidos por simulação, para os casos base, 1, 2 e 3 quando $G_{jp}=1,0$ na Rede 3.....	195
TABELA 6.48 – Número total de máquinas (MQ) obtido para as diferentes partições em cada iteração ($iter$) do algoritmo 6 para a Rede 4 com $G_{jp}=0,9$	196
TABELA 6.49 – <i>Leadtime</i> médio, $E(T_k)$, e número médio de itens, $E(L_k)$, obtidos por simulação e pelas aproximações para as 10 classes de produtos e para a rede ($E(T)$ e $E(L)$) na partição [1-8],[9,10] da Rede 4 com $G_{jp}=0,90$	197
TABELA 6.50 – Número de máquinas na rede (MQ) obtido para partições fixas (PF) com $G_{jp}=0,9$ (Rede 4).....	198
TABELA 6.51 – <i>Leadtime</i> médio, $E(T_k)$, e número médio de itens, $E(L_k)$, obtidos por simulação e pelas aproximações para as 10 classes de produtos e para a rede ($E(T)$ e $E(L)$) na partição [1-10] da Rede 4 com $G_{jp}=0,90$	198
TABELA 6.52 – Tempo médio de espera em fila, dado que espera, $E(Wq_{jp} Wq_{jp}>0)$, e número médio de itens no sistema para as linhas, $E(L_p)$, obtidos por simulação para a partição [1,2,3,4,5,6,7,8],[9,10] da Rede 4 com $G_{jp}=0,9$	199
TABELA 6.53 – Número total de máquinas (MQ) obtido para as diferentes partições em cada iteração ($iter$) do algoritmo 6 para a Rede 4 com $G_{jp}=4,00$	200
TABELA 6.54 – Número de máquinas na rede (MQ) obtido para partições fixas (PF) com $G_{jp}=4,00$ (Rede 4).....	201
TABELA 6.55 – <i>Leadtime</i> médio, $E(T_k)$, e número médio de itens, $E(L_k)$, obtidos por simulação e pelas aproximações para as 10 classes de produtos e para a rede ($E(T)$ e $E(L)$) na partição [1-8],[9],[10] da Rede 4 com $G_{jp}=4,00$	201

TABELA 6.56 – Tempo médio de espera em fila, dado que espera, $E(Wq_{jp} Wq_{jp}>0)$, e número médio de itens no sistema para as linhas ($E(L_p)$) da partição [1,2,3,4,5,6,7,8],[9],[10] da Rede 4 obtidos por simulação com $G_{jp}=4,00$	202
TABELA 6.57 – Número de máquinas (MQ) ao mudar alguns parâmetros do caso base (Rede 4) para a partição com menor número de máquinas resultante do algoritmo 6 com $G_{jp}=0,9$ e para a partição com uma linha única.....	203
TABELA 6.58 – Tempos médios de serviço para as operações l das 10 classes de produtos (k) do caso 1 com $G_{jp}=0,9$ (Rede 4).....	203
TABELA 6.59 – Número total de máquinas (MQ) obtido para as diferentes partições em cada iteração ($iter$) do algoritmo 6 para a Rede 4 com $G_{jp}=0,90$, empregando as aproximações de Bitran e Tirupati (1988).....	204
TABELA 6.60 – <i>Leadtime</i> médio, $E(T_k)$, e número médio de itens, $E(L_k)$, obtidos por simulação e pelas aproximações de Bitran e Tirupati (1988) para as 10 classes de produtos e para a rede ($E(T)$ e $E(L)$) na partição [1-10] da Rede 4 com $G_{jp}=0,90$	206
TABELA 6.61 – Tempo médio de espera em fila, dado que espera, $E(Wq_{jp} Wq_{jp}>0)$, e número médio de itens no sistema para as linhas ($E(L_p)$) da partição [1-10] da Rede 4 obtidos por simulação com $G_{jp}=0,9$, utilizando as aproximações de Bitran e Tirupati (1988).....	206
TABELA 6.62 – Número total de máquinas (MQ) obtido para as diferentes partições da Rede 3 em cada iteração ($iter$) do algoritmo 6 para $h_{jp}=10$ (Rede 3).	208
TABELA 6.63 – Número total de máquinas (MQ) obtido em diferentes partições (PF) para $h_{jp}=10$ (Rede 3).....	208
TABELA 6.64 – Coeficientes quadráticos de variação do intervalo de tempo entre chegadas (ca_{jp}) e de serviço (cs_{jp}) nas estações da partição [1],[2],[3],[4,5,6],[7,8,9,10] (Rede 3).....	209
TABELA 6.65 – Número total de máquinas (MQ) obtido para as diferentes partições da Rede 4 em cada iteração ($iter$) do algoritmo 6 para $h_{jp}=0,9$	210
TABELA 6.66 – Número total de máquinas (MQ) obtido em diferentes partições (PF) para $h_{jp}=0,9$ (Rede 4).....	211

TABELA 6.67 – Coeficientes quadráticos de variação do intervalo de tempo entre chegadas (ca_{jp}) e de serviço (cs_{jp}) nas estações da partição [1-8],[9],[10] (Rede 4).....

LISTA DE FIGURAS

FIGURA 2.1 – Representação esquemática de sistemas de manufatura – a) Sistema <i>job-shop</i> ; b) Sistema <i>flow-shop</i> e; c) manufatura celular.....	10
FIGURA 2.2 – Sistema de fila com único estágio.....	15
FIGURA 2.3 – Diagrama de taxas de transição de estados.....	23
FIGURA 2.4 – (a) Processo de superposição de chegadas; (b) Processo de partidas e; (c) Processo de separação das partidas.....	29
FIGURA 2.5 – Ilustração do processo de chegadas (partidas) da classe agregada entre duas chegadas (partidas) sucessivas da classe de interesse na estação n_{kl}	50
FIGURA 4.1 – Sistema <i>job-shop</i> em estudo como uma rede de filas (0 é a estação fictícia representando a chegada externa): (a) representa o roteiros das classes $4, 5, 9, 10, 11, 13$ e 14 ; (b) representa os roteiros das classes $1, 2, 3$ e 7 ; (c) representa os roteiros das classes $6, 8$ e 12	93
FIGURA 4.2 – Roteiro de fabricação da classe de produtos $6: 3, ST, 14, ST, 14, 15$	100
FIGURA 4.3 – Curva de <i>trade-off</i> entre investimento em máquina e <i>WIP</i>	118
FIGURA 4.4 – <i>Trade-off</i> entre investimento em máquina e <i>leadtime</i> médio da rede.....	118
FIGURA 4.5 – Relação entre <i>WIP</i> e número médio de lotes na rede.....	119
FIGURA 4.6 – Relação entre <i>WIP</i> e <i>leadtime</i> médio na rede.....	119
FIGURA 4.7 – <i>Leadtime</i> das classes de produtos para três configurações de rede diferentes, obtidos ao resolver o modelo $SPI.2/G/S/R^{discreto}$	123
FIGURA 4.8 – <i>Trade-off</i> entre investimento em recursos e <i>WIP</i> desejado para os cenários 1.0 (quando os coeficientes são os mesmos da configuração inicial) e 1.2 (quando o coeficiente quadrático de variação do processo de chegada para todas as classes é reduzido à metade).....	127
FIGURA 4.9 – <i>Trade-off</i> entre investimento em recursos e <i>WIP</i> desejado para os cenários 2.0 (quando a taxa é mantida de acordo com a configuração inicial) e 2.1 (quando a taxa de chegada total à rede é aumentada em 10%).....	130
FIGURA 4.10 – <i>Trade-off</i> entre investimento em recursos e <i>WIP</i> desejado para os cenários 3.0 (quando o <i>mix</i> de produção é o da configuração inicial da rede),	

3.1 (quando a classe de produtos 13 é extinta, com acréscimo da taxa de chegada da classe 2) e 3.2 (quando a classe de produtos 13 é extinta, com acréscimo da taxa de chegada da classe 3).....	133
FIGURA 6.1 – Rede formada por um estágio único $GI/G/m$, processando 2 classes de produtos (Rede 1).....	154
FIGURA 6.2 – Rede formada por dois estágios $GI/G/m$, processando 2 classes de produtos (Rede 2).....	154
FIGURA 6.3 – Rede formada por um estágio único $GI/G/m$, processando 10 classes de produtos (Rede 3).....	155
FIGURA 6.4 – Rede formada por vários estágios $GI/G/m$, processando as classes 1,2,...,7 (Figura 6.4a) e 8,9,10 (Figura 6.4b) de produtos (Rede 4).....	156

LISTA DE PRINCIPAIS SIGLAS, SÍMBOLOS E ABREVIATURAS

$1\{.\}$	Função indicadora que resulta 1 se a expressão $\{.\}$ é verdadeira e 0 caso contrário.
a_0	Intervalo de tempo entre chegadas externas (da estação 0) na rede.
a_{0j}	Intervalo de tempo entre chegadas externas na estação j .
a_{ji}	Intervalo de tempo entre chegadas na estação j da estação i .
a'_k	Intervalo de tempo entre chegadas externas da classe de produtos k .
a_{kl}	Intervalo de tempo entre chegadas da classe de interesse k para a operação l na estação n_{kl} .
a'_{kl}	Intervalo de tempo entre chegadas da agregação de todas as classes que chegam entre duas chegadas sucessivas da classe de interesse k para a operação l na estação n_{kl} .
a_j^{iter}	Intervalo de tempo entre chegadas na estação j na $iter$ -ésima iteração dos algoritmos 3, 4 e 5.
C_{kl}	Custo acumulado (operacional + matéria-prima) de um lote médio da classe k com a realização da l -ésima operação.
cx	Coefficiente quadrático de variação da variável aleatória x , definido por $cx = V(x)/E(x)^2$.
$Cov(x,y)$	Covariância entre as variáveis aleatórias x e y .
CQN	Rede de filas fechada (<i>closed queueing network</i>)
D	Indica que o processa de chegadas ou de serviço em um sistema de fila é determinístico.
d_j	Intervalo de tempo entre partidas da estação j .
d_{ji}	Intervalo de tempo entre partidas da estação j para a estação i .
d_{kl}	Intervalo de tempo entre partidas da classe de interesse k após a operação l na estação n_{kl} .
$DP(x)$	Desvio-padrão da variável aleatória x .
$E(x)$	Valor esperado da variável aleatória x .
EDD	Disciplina de atendimento da fila data devida mais próxima (<i>earliest due date</i>)

e_i	Vetor (e_1, e_2, \dots, e_n) com $e_i=1$ e $e_j=0, j=1, \dots, n, j \neq i$
e_j	Intervalo de tempo entre chegadas externas das classes na estação j , considerando o número de visitas de cada classe na estação j .
ET	Limitante superior para o <i>leadtime</i> médio da rede
ET_k	Limitante superior para o <i>leadtime</i> médio da classe k
F	Investimento (ou custo) de capacidade da rede.
$F(x)$	Função distribuição acumulada empírica da variável aleatória x .
$FCFS$	Disciplina de atendimento de fila primeiro a chegar, primeiro a ser servido (<i>first come, first served</i>).
F_T	Limitante superior para o custo de capacidade da rede, F .
G	Processo de chegada ou serviço genérico num sistema de fila.
GI	Processo de chegada genérico independente (processo de renovação) num sistema de fila.
$GAMS$	<i>General Algebraic Modeling System</i>
$F_j(\mu_j, m_j)$	Custo de alocação da capacidade (μ_j, m_j) na estação j .
G'_j	Igual a $1/2G_j$.
G_j	Limitante superior para o tempo médio de espera de um item em fila, dado que ele espera $E(W_{qj} W_{qj}>0)$ na estação j .
G_{jp}	Limitante superior para o tempo médio de espera de um item em fila, dado que ele espera $E(W_{qjp} W_{qjp}>0)$ na estação j da linha p .
γ_j	Constante indicando de forma aproximada o nível de serviço da estação j
γ_i	Representa a criação (se $\gamma_i > 1$) ou a combinação (se $\gamma_i < 1$) de itens.
h_{jp}	Limitante superior para o tempo de espera na fila da estação j da linha p .
i	Índice utilizado para indicar estações ($i=1, \dots, n$).
iid	Independente e identicamente distribuído.
j	Índice utilizado para indicar estações ($j=1, \dots, n$).
k	Índice utilizado para indicar as classes de produtos ($k=1, \dots, r$)
l	Índice utilizado para indicar operações ($l=1, \dots, n_k$).
L	Número de itens (<i>jobs</i>) na rede.
\mathbf{L}	Vetor (L_1, L_2, \dots, L_n) representando o estado da rede.
L_j	Número de <i>jobs</i> na estação j .

L_{jt}	Número de <i>jobs</i> na estação <i>j</i> e alternativa <i>t</i> .
L_k	Número de <i>jobs</i> da classe <i>k</i> na rede.
Lq_j	Número de <i>jobs</i> em espera (em fila) na estação <i>j</i> .
L_T	Limitante superior para o número de itens (<i>jobs</i>) na rede, <i>L</i> .
λ'_k	Taxa média de chegada externa da classe <i>k</i> na rede.
λ_0	Taxa média de chegada externa (<i>throughput</i>) da rede.
λ_{0j}	Taxa média de chegada externa na estação <i>j</i> .
λ_{ij}	Taxa média de chegada na estação <i>j</i> da estação <i>i</i> .
λ_j	Taxa média de chegada na estação <i>j</i> .
λ_{jp}	Taxa média de chegada na estação <i>j</i> da linha <i>p</i> .
m_j^0	Limitante inferior para o número de máquinas idênticas e paralelas na estação <i>j</i> .
m	Número de máquinas idênticas e paralelas em um sistema de filas monoestágio.
M	Na notação de sistemas de filas indica que o processo de chegada ou serviço é Markoviano. Na notação dos modelos de otimização pertencentes às classes <i>SP1</i> , <i>SP2</i> ou <i>SP3</i> indica que as estações possuem múltiplas máquinas.
m_j	Número de máquinas paralelas e idênticas na estação <i>j</i> .
m_{jp}	Número de máquinas na estação <i>j</i> da linha (planta) <i>p</i> .
m_{jt}	Número de máquinas da estação <i>j</i> na alternativa <i>t</i> .
μ_j^0	Limitante inferior da taxa média de processamento da estação <i>j</i> .
μ_j^{iter}	Taxa média de processamento na estação <i>j</i> na <i>iter</i> -ésima iteração dos algoritmos 3, 4 e 5.
μ	Vetor $(\mu_1, \mu_2, \dots, \mu_n)$ da taxa média de processamento em cada estação da rede.
μ_j	Taxa média de processamento da estação <i>j</i> .
μ_{jp}	Taxa média de serviço na estação <i>j</i> da linha <i>p</i> .
μ_{jt}	Taxa média de processamento da estação <i>j</i> na alternativa <i>t</i> .
μ_{kl}	Taxa média de processamento da classe <i>k</i> na operação <i>l</i> .

N	Número de estações internas da rede.
N_k	Número de operações da classe k .
N_{kl}	Estação visitada pela classe k para operação l .
v_j	Custo médio de um lote médio de uma classe qualquer na estação j .
OF	Ordem de fabricação.
OQN	Rede de filas aberta (<i>open queueing network</i>).
p	Índice indicando a linha (planta), $p=1, \dots, r$.
$P(A)$	Probabilidade de que o evento A ocorra.
PCP	Planejamento e controle da produção.
PI_j	Índice de prioridade na estação j .
$PRC-1$	Problema de redução de complexidade 1.
$PRC-2$	Problema de redução de complexidade 2.
$PRC-3$	Problema de redução de complexidade 3.
$\pi(L)$	Vetor $(\pi_1(L_1), \pi_2(L_2), \dots, \pi_n(L_n))$ das distribuições de equilíbrio das estações da rede.
π_j	Distribuição de equilíbrio da estação j .
Q	Matriz de transição sub-estocástica $\{q_{ij}, i, j=1, \dots, n\}$.
Q_{0i}	Probabilidade de um <i>job</i> entrar na rede pela estação i .
Q_{i0}	Probabilidade de um <i>job</i> sair da rede pela estação i .
Q_{ij}	Probabilidade de um <i>job</i> , após ser atendido na estação i , seguir para a estação j .
Q_{kl}	Proporção de <i>jobs</i> da classe k (de interesse) na operação l , dentre os <i>jobs</i> na estação n_{kl} .
r	Número de classes na rede.
R	Matriz de transição subestocástica $\{r_{ij}, i, j=0, \dots, n\}$.
R	Indica que as variável de decisão nos modelos de otimização $SP1$, $SP2$ e $SP3$ é a taxa de produção da estação.
R_{ij}	Probabilidade de um <i>job</i> após ser atendido na estação i seguir para a estação j .
ρ_j	Utilização média (intensidade de tráfego) da estação j .
ρ_{jp}	Utilização média da estação j da linha (planta) p .

ρ_{jt}	Utilização média da estação j na alternativa t .
S	Indica que as estações possuem um único servidor nos modelos de otimização $SP1$, $SP2$ e $SP3$.
s_j	Tempo de processamento na estação j .
s_{kl}	Tempo de processamento da l -ésima operação da classe k .
s_{klt}	Tempo de processamento da classe k na operação l da alternativa t .
SPT	Disciplina de atendimento do item em fila com menor tempo de processamento (<i>shortest processing time</i>).
ST	Serviço terceirizado.
T	<i>Leadtime</i> de produção da rede.
t	Índice utilizado para indicar a alternativa discreta de adição/remoção de capacidade.
T^*	Limitante superior para a variância ponderada da rede.
T_k	<i>Leadtime</i> de produção da classe k .
T_{kl}	Tempo de processamento de cada item (e não de um lote) da classe k na operação l .
u_{jt}	Alternativa discreta de adição/remoção de capacidade t na estação j .
$V(x)$	Variância da variável aleatória x .
WIP	Valor esperado do estoque em processo (\$) (<i>work-in-process</i>).
W_j	Peso atribuído ao método assintótico e ao método dos intervalos estacionários para estimativa de ca na estação j .
W_j	Tempo de espera (fila + serviço) na estação j .
W_k	Peso atribuído à classe k para estimativa da variância ponderada da rede.
Wq_j	Tempo de espera em fila na estação j .
Wq_{jp}	Tempo de espera em fila na estação j da linha p .
Wq_{jt}	Tempo de espera em fila na estação j com a alternativa discreta de capacidade t .
x_{kp}	Variável de decisão binária do modelo $SP3$, assumindo valor igual a 1 se a classe k é atribuída à linha p ; e valor igual 0, em caso contrário.
y_k	Tamanho de um lote de produção.
z_{kl}	Número de chegadas/partidas de produtos de todas as outras classes (classe

agregada) entre duas chegadas/partidas sucessivas da classe de interesse k na operação l executada na estação n_{kl} .

RESUMO

APLICAÇÃO DE MODELOS DE REDES DE FILAS ABERTAS PARA O PROJETO E PLANEJAMENTO DE SISTEMAS DISCRETOS DE MANUFATURA.

A gestão de sistemas de manufatura tem se tornado mais complexa na medida em que novos produtos estão surgindo, a demanda de produtos é incerta, os ciclos de vida são mais curtos e uma grande variedade de produtos competem pelos mesmos recursos. Esta tese trata o projeto e planejamento de sistemas discretos de manufatura, baseados em modelos de redes de filas abertas, para auxiliar no processo de tomada de decisões. Porque sistemas de manufatura podem ser representados por redes de filas genéricas, e não existem métodos exatos de solução, aqui é empregado o método aproximado de decomposição para avaliar o desempenho dos sistemas em diferentes configurações. Mostra-se nesta tese que estas aproximações são adequadas e efetivas para estimar os estoques em processo (*WIP*) e o *leadtime* de produção de uma rede de manufatura real de uma indústria metal-mecânica. Mostra-se ainda que os modelos de alocação de capacidade discreta, baseados nestas aproximações, também são efetivos para avaliar e otimizar o desempenho da rede do estudo de caso em diferentes configurações. Curvas de *trade-off* entre investimento em capacidade e *WIP* são geradas e são úteis não somente para auxiliar um gerente a estimar quanto alocar de capacidade, mas também para decidir onde alocá-la na rede de filas. As curvas também auxiliam a tomada de decisões em termos de capacidade, se a variabilidade das chegadas externas, o *mix* de produtos e/ou a taxa de produção da rede mudam. Além de adicionar capacidade, a partição da instalação é outra alternativa para reduzir a complexidade do sistema. Esta tese também aborda o problema de projeto de fábrica focalizada, envolvendo a partição da instalação em subplantas e a alocação de capacidade em cada estação das subplantas. Novamente, as aproximações por decomposição foram utilizadas para avaliar e otimizar o desempenho do sistema. Apesar de sua importância, este problema tem sido muito pouco reportado em literatura. Nesta tese, o objetivo é reduzir a complexidade do sistema do ponto de vista da gestão do produto, ou do ponto de vista da gestão da estação. Do ponto de vista da gestão do produto, é apresentado um modelo cuja restrição de complexidade é um limitante superior para a variância do *leadtime* dos produtos na rede. Do ponto de vista da gestão da estação, a restrição de complexidade do modelo fixa o tempo médio de espera de um produto na estação, quando há espera. Mostra-se, por meio destes modelos para alguns casos que, a partição da instalação em subplantas reduz a complexidade do sistema, sem necessidade de investimentos adicionais em capacidade. Além disso, algumas vezes, é possível manter (ou até melhorar) o desempenho da rede, particionando-a em subplantas que necessitam de menos capacidade do que a configuração original com uma planta única.

Palavras-chave: Fábrica Focalizada. Projeto de Sistemas de Manufatura.
Modelos de Avaliação de Desempenho. Redes de Filas Abertas.
Modelos de Otimização. Curvas de *trade-off*.

ABSTRACT

APPLICATION OF OPEN QUEUEING NETWORK MODELS FOR THE DESIGN AND PLANNING OF DISCRETE MANUFACTURING SYSTEMS.

The management of manufacturing systems have become more complex, once that new products are arising, product demands are uncertain, life cycles get shorter, and a wide variety of products compete for common resources. This thesis deals with the design and planning of discrete manufacturing systems, based on open queueing network models to support the decision making of capacity allocation. As manufacturing systems may be represented by generalized queueing networks, and there are no exact solution methods, here is employed the decomposition approximate method to evaluate the performance of systems under different configurations. It is shown in the thesis how these approximations are suitable and effective to estimate the work-in-process (WIP) and the production leadtime of an actual metallurgical industry queueing network. It is also shown that discrete capacity allocation models, based on the approximations, are effective to evaluate and optimize the performance of the case study under different configurations. Trade-off curves between capacity investment and WIP are generated and are useful not only to support a manager to estimate how much capacity he/she should allocate, but also to decide where it should be allocated in the queueing network. These curves also support the decision making in terms of capacity, if the variability of the external arrivals, the product mix and/or the throughput for the network change. Besides adding capacity, partitioning the facility is another alternative to reduce the system complexity. This thesis also approaches the focused factory design problem, involving the partition of the facility into smaller shops and the capacity allocation in each shop. Again, the decomposition approximations were employed to evaluate the system performance. Despite its importance, this problem has rarely been reported in the literature. In this thesis, the goal is to reduce the system complexity either from the product management point of view or from the workstation management point of view. From the product management point of view, a model whose complexity constraint is na upper limit on the production leadtime variance of the products passing through the network was studied. From the workstation management point of view, the complexity constraint of the model keeps constant the expected waiting time of a product at a workstation, once it waits for being served. It is shown through these models for some instances that the partition of the facility into smaller shops decreases the complexity system without necessary capacity additional investments. Futhermore, sometimes it is possible to keep the network performance (or even to improve it), partitioning the network into smaller shops which need less capacity than the original configuration with a single shop.

Key-words: Focused Factory. Manufacturing System Design.
Performance Evaluation Models. Open Queueing Network.
Optimization Models. Trade-off Curves.

1. INTRODUÇÃO

A gestão de sistemas de manufatura tem se tornado cada vez mais complexa na medida em que uma grande variedade de produtos competem pelos mesmos recursos, demandas de produtos são incertas e o ciclo de vida dos produtos torna-se cada vez mais curto. Redes de filas podem ser utilizadas para representar sistemas de manufatura e auxiliar na tomada de decisões. São formadas de vários sistemas de filas conectados entre si, com usuários deslocando-se entre eles para receber serviço. Os nós da rede representam as estações (*shops*); os produtos (*jobs*), os usuários requerendo serviço das estações e; os arcos conectando os nós, as rotas dos produtos.

As redes de filas são classificadas como redes de Jackson, aquelas cujos processos de chegadas externas são Poisson e os de serviço em cada estação são exponenciais, ou redes Generalizadas, aquelas cujos processos de chegadas externas não necessariamente são Poisson e/ou os de serviço não necessariamente são exponenciais, ou seja, são descritos por distribuições genéricas. As redes podem ser abertas (número de *jobs* circulando na rede pode variar a cada instante) ou fechadas (o número de *jobs* na rede é constante), sendo as abertas o foco desta tese.

Métodos exatos de avaliação de desempenho existem somente para redes de Jackson, mas a representação de sistemas de manufatura como redes de Jackson, em geral, superestima a variabilidade do processo de serviço (Bitran e Tirupati, 1988). Portanto, métodos aproximados são utilizados para estimar o desempenho de redes mais gerais, particularmente os sistemas discretos de manufatura tipo *job-shop*. As aproximações são modelos descritivos, ou seja, para uma dada configuração descrevem o desempenho da rede.

Nesta tese são abordadas aproximações geradas pelo método de decomposição paramétrico, cujo objetivo é estender para redes mais gerais o resultado obtido para as redes de Jackson: cada nó da rede pode ser tratado como se fosse estocasticamente independente e, desta maneira, a análise é simplificada. O método usa dois parâmetros para caracterizar cada processo de chegada e de serviço, um para descrever a taxa média e o outro para descrever a variabilidade. Medidas de congestão para a rede toda são obtidas admitindo que os nós sejam estocasticamente independentes. Vários trabalhos em literatura abordam as aproximações e utilizam o

método de decomposição paramétrico para avaliar o desempenho de redes de filas abertas. Dentre os quais pode-se citar Albin (1982, 1984), Bitran e Dasu (1992), Bitran e Morabito (1995, 1996), Bitran e Sarkar (1993, 1994a, 1994b), Bitran e Tirupati (1988, 1989a, 1989b), Boxma (1990), Morabito (1998), Segal e Whitt (1989), Souza et al (2001), Van Vliet e Rinnooy Kan (1991) e Whitt (1983a, 1983b).

Aqui, os modelos descritivos de redes de filas abertas são aplicados no projeto de sistemas de manufatura discretos. Uma questão central é a seleção entre as várias configurações para a rede ou, mais especificamente, como os recursos devem ser adequadamente distribuídos para alocar capacidade nas várias estações. Para isso, são utilizados modelos prescritivos (de otimização), ou seja, prescrevem a melhor configuração do sistema, otimizando certo critério e satisfazendo determinadas restrições. Tais modelos são resolvidos por abordagem que combinam modelos estocásticos de redes de filas com procedimentos de otimização. A alocação insuficiente de capacidade pode causar altos níveis de inventário em processo (*work-in-process* ou *WIP*) e longos *leadtimes* de produção (tempo total que um produto leva para ser produzido, ou seja, a soma dos tempos de espera em fila mais os tempos de processamento). Por outro lado, o excesso de capacidade pode resultar em desperdício de recursos onerosos, devido aos baixos níveis de utilização.

Bitran e Dasu (1992) classificam os problemas de projeto de sistemas de manufatura em três classes:

- i) desempenho desejado do sistema (*SP1, Strategic Problem 1*)
- ii) desempenho ótimo do sistema (*SP2, Strategic Problem 2*)
- iii) partição da instalação (*SP3, Strategic Problem 3*)

Na classe *SP1*, o objetivo é minimizar o investimento no sistema sujeito às restrições de desempenho desejado para o sistema. Medidas de desempenho típicas são *WIP*, *leadtime* médio dos produtos e utilização média (intensidade de tráfego) das estações de trabalho. Na classe *SP2* deseja-se otimizar o desempenho do sistema sujeito a limitações orçamentárias. Tanto o problema *SP1* quanto o *SP2* envolvem um conflito (*trade-off*) entre capital de investimento e capital de trabalho se o *WIP* for a medida de desempenho adotada. A classe *SP3* busca particionar a planta em subplantas menores (plantas dentro da planta) para reduzir a complexidade do sistema e melhorar o desempenho global. Esta medida pode requerer adição de capacidade e há, portanto, um

trade-off entre o custo de adição de capacidade e a redução de complexidade gerencial do sistema. Estes problemas são objetos de estudo nesta tese.

1.1 Objetivos e Justificativas

O foco desta tese está no estudo e aplicação de modelos descritivos (aproximações) e prescritivos (de otimização) de redes de filas abertas, para projeto e planejamento de sistemas de manufatura discretos. Deste modo, objetiva-se:

- a) Fazer uma revisão abrangente da literatura acerca de modelos e métodos de solução para avaliação de desempenho e otimização de redes de filas abertas, voltados para representação de sistemas de manufatura discretos tipo *job-shop*.
- b) Verificar a aplicabilidade e adequação dos modelos de avaliação de desempenho (baseados em redes de filas generalizadas, especificamente os aproximados por decomposição) para estimar o *leadtime* (ou o *WIP*) e o nível de utilização das estações de um sistema *job-shop* na prática. Para isso, realizou-se um estudo de caso em uma empresa do setor metal-mecânico. Os modelos baseados no método aproximado por decomposição paramétrico são validados por amostras, modelos de simulação e análise do tomador de decisões da empresa.
- c) Verificar a aplicabilidade e adequação dos modelos de otimização *SP1* e/ou *SP2* no estudo de caso. Curvas de fronteira ótima (curvas de *trade-off* ou de Pareto), obtidas através de avaliação paramétrica das medidas de desempenho dos problemas *SP1* e/ou *SP2*, são geradas para o estudo de caso com vistas a auxiliar no processo de tomada de decisões em problemas de projeto e planejamento. Especificamente, aquelas que dizem respeito à alocação de capacidade e que a considera no formato de alternativas discretas. Ainda, observar a sensibilidade das soluções apresentadas a variações nos parâmetros de entrada da rede, tais como a variabilidade do processo de chegadas dos produtos, a taxa de produção da rede (*throughput*) e o *mix* de produção.
- d) Estudar e desenvolver modelos e métodos de solução para problemas da classe *SP3*, que envolve o projeto de fábrica focalizada. O conceito de fábrica focalizada é considerado como uma alternativa possível ao sistema *job-shop* com máquinas intercambiáveis, visando diminuir a complexidade deste sistema. Particularmente, o

projeto de fábrica focalizada é abordado neste trabalho do ponto de vista de partição da instalação de um *job-shop* em plantas menores e alocação de capacidade (máquinas) a cada planta. O foco é os modelos de otimização para analisar o *trade-off* entre investir em capacidade e reduzir a complexidade do sistema (seja do ponto de vista da gestão do produto ou da gestão da estação), ao particionar a planta em linhas de produção ou subplantas focalizadas. Além disso, analisar sob quais condições é vantajoso particionar uma instalação em subplantas focalizadas, sem necessariamente envolver aumento de capacidade na instalação, podendo, em certos casos, até reduzi-la.

Os objetivos desta tese vêm de encontro às perspectivas para pesquisa futura propostas por Morabito (1998). Poucos trabalhos analisaram e reportaram a aplicação prática de modelos de redes de filas em sistemas *job-shops* reais, tanto para avaliar o desempenho quanto para descrever configurações para a rede que otimizam um determinada medida. Do ponto de vista quantitativo, o problema de partição da instalação também é muito pouco abordado em literatura, apesar de sua importância, e daí a contribuição deste trabalho. Os modelos aqui tratados visam auxiliar um gerente de produção em algumas de suas principais atividades, tais como avaliar o desempenho do sistema ao prever com maior precisão o *leadtime* de produção (ou o *WIP*). Também permitem prescrever novas configurações para o sistema por meio de curvas de *trade-off* que auxiliem o gerente nas decisões relacionadas ao projeto (ou reprojeto) e planejamento do sistema, tais como adicionar capacidade e/ou particionar a instalação de modo a atingir uma medida de desempenho desejada ao mínimo custo.

1.2 Estrutura da Tese

A tese está estruturada em 7 capítulos.

O capítulo 2 apresenta uma revisão bibliográfica acerca de modelos de avaliação de desempenho para redes de filas abertas. A seção 2.1 descreve tipos de sistemas de manufatura discretos que podem ser representados por redes de filas. A seção 2.2 define a terminologia e os tipos de redes de filas. A seção 2.3 trata as redes de filas Markovianas, enquanto a seção 2.4 aborda o método aproximado de decomposição utilizado para resolver os modelos para as redes de filas generalizadas apresentadas na

seção 2.5. Na seção 2.6, são definidos os principais conceitos relacionados à técnica de simulação, utilizada neste trabalho para validar os resultados obtidos pelas aproximações por decomposição.

O capítulo 3 reporta uma revisão de literatura acerca de modelos referentes a alocação e realocação de servidores. A seção 3.1 trata os tipos de problemas tático-estratégicos relacionados ao projeto e planejamento de sistemas de manufatura (classes *SP1*, *SP2* e *SP3*). A seção 3.2 formula os problemas *SP1* e *SP2*, e as seções 3.3 e 3.4 apresentam os modelos das classes de problemas *SP1* e *SP2*, para redes de Jackson e Generalizadas, utilizando capacidade discreta e contínua.

O capítulo 4 é importante para os objetivos deste trabalho porque apresenta os resultados computacionais obtidos ao aplicar as aproximações e os modelos de alocação de capacidade, considerando alternativas discretas, em um estudo de caso real. Inicialmente, há uma breve descrição da empresa estudada (seção 4.1), seguida por uma descrição de como o sistema de manufatura *job-shop* da empresa foi representado por uma rede de filas aberta e um modelo de simulação. Nesta seção, são relatadas todas as hipóteses estabelecidas para caracterizar o roteiro de fabricação, os processos de chegadas e de serviço para cada classe de produto. A seção 4.3 apresenta os custos e as alternativas discretas de capacidade considerados como parâmetros para os modelos de otimização. Na seção 4.4 mostra-se o desempenho da configuração inicial do sistema *job-shop* estudado na empresa. Seis sistemas lineares diferentes de aproximações são testados, e os resultados de medidas de desempenho obtidos são comparados com resultados obtidos por amostras e um modelo de simulação, para verificar qual dos sistemas é o mais apropriado. O sistema mais adequado é utilizado no procedimento de otimização para alocação de capacidade na rede, conforme a seção 4.5. Nesta seção, verifica-se a sensibilidade do modelo de alocação de capacidade discreta a mudanças nos parâmetros de variabilidade, nas taxas de chegadas externas das classes de produtos e no *mix* de produção.

O capítulo 5 apresenta os modelos de otimização referentes ao problema de partição da instalação (*SP3*). A seção 5.1 descreve as hipóteses e considerações relacionadas ao sistema de produção modelado. Não foi possível utilizar o sistema de produção da empresa estudada e, portanto, utilizou-se redes geradas neste trabalho e redes encontradas na literatura. A seção 5.2 apresenta as medidas de complexidade do

ponto de vista da gestão do produto e da gestão da estação para estas redes. A seção 5.3 define três modelos para o projeto de fábrica focalizada, em que cada modelo possui uma medida de complexidade diferente como restrição. Por fim, a seção 5.4 apresenta uma heurística de economias utilizada para resolver os modelos de projeto de fábrica focalizada.

O capítulo 6 mostra os resultados computacionais com o problema de partição da instalação aplicado nas redes acima, considerando a medida de complexidade do ponto de vista de gestão do produto (PRC-1, na seção 6.1) e do ponto de vista da gestão da estação (PRC-2 e PRC-3, nas seções 6.2 e 6.3, respectivamente). Neste capítulo, mostra-se algumas condições de uma rede generalizada que favorecem a partição da instalação em subplantas, sem a necessidade de adição de capacidade, mantendo em nível desejado as medidas de complexidade do sistema.

Por fim, o capítulo 7 discute as conclusões e perspectivas de pesquisas futuras.

2. MODELOS DE AVALIAÇÃO DE DESEMPENHO

Neste capítulo apresenta-se uma revisão dos modelos de avaliação de desempenho para redes de filas abertas representando sistemas de manufatura discretos. No final do capítulo são discutidos resumidamente os modelos de simulação utilizados para validação dos modelos revisados. Antes da apresentação e discussão dos modelos, faz-se necessário apresentar conceitos e definições importantes sobre sistemas de manufatura discretos e como eles podem ser representados por redes de filas.

2.1 Sistemas de Manufatura Discretos

Sistemas de manufatura consistem basicamente de máquinas e estações de trabalho, onde operações são realizadas sobre partes, itens, submontagens, e montagens, para criar produtos que serão distribuídos para clientes. Os componentes de movimentação de materiais e de estocagem permitem que os itens se movam de estação para estação, que partes apropriadas estejam disponíveis para montagem e que o trabalho seja mantido até poder entrar nas estações para processamento.

Os sistemas discretos de manufatura são aqueles onde cada item processado é distinto. Por simplicidade, um item, parte, submontagem ou montagem processada por uma máquina ou estação é chamado simplesmente de usuário ou item/produto/tarefa, *job* (Buzacott e Shanthikumar, 1993).

Sistemas de manufatura discretos podem ser classificados em função do volume e variedade das tarefas. Tanto a variedade dos tipos de itens/produtos manufaturados pelo sistema (ou seja, escopo) quanto o volume de cada tipo de item/produto produzido (ou seja, escala) interferem diretamente no projeto e operação do sistema. Convém, para os propósitos do trabalho, definir algumas formas tradicionais de se organizar um sistema de manufatura discreto.

2.1.1 *Job-shop*

É uma forma organizacional do sistema de manufatura com capacidade de escopo (ou seja, capacidade de processamento de vários tipos de *jobs*), mas com limitação em capacidade de escala devido aos problemas associados ao controle e

movimentação dos *jobs* e ao tempo de preparação (*setup time*) das máquinas. Este sistema também é referenciado como *layout* por processo (ou *layout* funcional) dado que máquinas desempenhando funções semelhantes são agrupadas em um mesmo departamento. As máquinas podem ser intercambiáveis, ou seja, uma mesma máquina pode realizar operações em itens/produtos pertencentes a classes diferentes, o que exige um certo tempo de preparação das máquinas. Daí, o sistema ser caracterizado por longos *leadtimes* e altos níveis de inventário em processo (*WIP* ou *work-in-process*). Vantagens e desvantagens do emprego deste tipo de sistema de manufatura podem ser vistas em Buzacott e Shanthikumar (1993) e Krajewski e Ritzman (1996). A figura 2.1a ilustra este tipo de arranjo.

2.1.2 Flow-shop

Diferentemente do *job-shop*, o *flow-shop* é uma forma de organização do sistema de produção com capacidade de escala (capaz de processar altos volumes), mas com capacidade limitada de escopo. Tem-se, para este caso, o *layout* por produto em que as estações de trabalho ou departamentos são arranjados de uma maneira linear (vide figura 2.1b). Todos os produtos/itens visitam máquinas e estações na mesma seqüência. São caracterizados por baixos *leadtimes* e baixos níveis de *WIP*. Outras vantagens e desvantagens podem ser vistas em Buzacott e Shanthikumar (1993) e Krajewski e Ritzman (1996).

2.1.3 Fábrica Focalizada e Manufatura celular

Segundo Skinner (1974), a abordagem mais prática de fábrica focalizada é o de divisão, organizacional e física, da instalação existente em plantas dentro de uma planta (*PWP*, do inglês, *plant within a plant*). Cada *PWP* tem suas próprias instalações podendo se concentrar em suas tarefas de produção particulares, usando abordagens de gestão de recursos humanos, controle de produção e estrutura organizacional próprias. Equipamentos e manutenção de materiais são especializados quando necessários. As vantagens são que cada *PWP* ganha experiência facilmente ao focalizar e concentrar todo elemento de seu trabalho naqueles objetivos essenciais e limitados que constituem sua tarefa de manufatura. A abordagem de *PWP* torna mais fácil o realinhamento, com o

decorrer do tempo, de operações essenciais e de elementos do sistema com mudanças propostas pela estratégia corporativa e de mercado.

Conforme Bitran e Dasu (1992), para particionar uma planta em várias fábricas focalizadas, deve-se determinar os produtos que serão produzidos em cada planta e alocar capacidade a cada uma delas. Idealmente, cada planta não deveria partilhar equipamentos e cada produto deveria ser atribuído a somente uma planta. Entretanto, se as demandas não são suficientemente altas ou se o custo das máquinas é muito alto, então vários produtos/itens terão de dividir o mesmo equipamento, tal como ocorre em fábricas de semicondutores nos processos de metalização e fotolitografia. Mesmo assim, há várias vantagens em particionar plantas grandes em sub-unidades menores e mais focalizadas. Dentre as quais, pode-se citar fluxos de produtos simplificados, aperfeiçoamentos do controle, respostas mais rápidas à demanda e maior previsibilidade dos *leadtimes* de produção.

Souza et al. (2001) definem fábrica focalizada como uma maneira de lidar com complexidade em uma operação de manufatura. Processos ou produtos partilhando características similares são agrupados e produzidos em plantas isoladas ou em plantas dentro de uma planta. O agrupamento pode ocorrer de acordo com tecnologia de processo, demanda de mercado, volume de produto, nível de qualidade e equipamento de manufatura.

Abrangida pelo conceito de fábrica focalizada, uma maneira utilizada para arranjar o sistema de manufatura é o emprego de células responsáveis pela produção de uma ou mais classes de produtos, conhecida por manufatura celular.

Este sistema pode ser usado para converter um sistema de disposição por processo em disposição por produto através da adoção da tecnologia de grupo. Primeiramente, realiza-se a classificação e codificação dos itens/produtos para que sejam formadas as classes de itens/produtos, baseadas nas suas semelhanças de projeto e/ou manufatura. Em seguida, máquinas que realizam operações diferentes são identificadas e dedicadas à produção de uma classe ou a um conjunto de classes de itens/produtos. Estas máquinas dedicadas são agrupadas para formar a célula de manufatura e o arranjo para este caso é denominado *layout* celular. Em verdade, esta é uma disposição por produto com arranjo físico das máquinas em uma célula (vide figura 2.1c). O uso de células para processar um conjunto específico de itens/produtos

simplifica a programação e o controle e reduz substancialmente o tempo de preparação da máquina, a movimentação de materiais, o *leadtime* de produção e o nível de *WIP*.

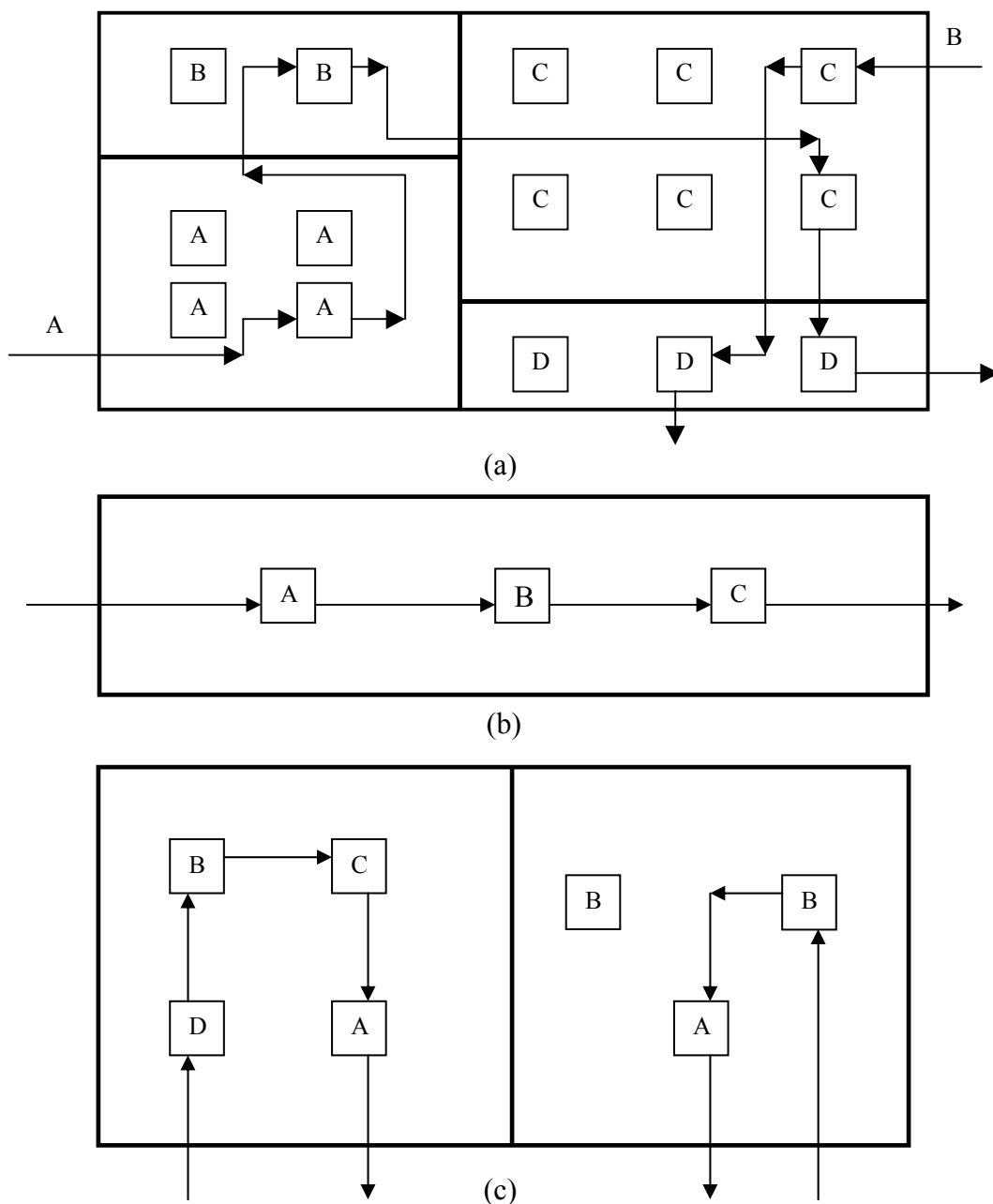


FIGURA 2.1 – Representação esquemática de sistemas de manufatura – a) Sistema *job-shop* ; b) Sistema *flow-shop* e; c) manufatura celular.

2.1.4 *Job-shop* versus Manufatura celular

Diversos pesquisadores tem se preocupado com o tipo de sistema de manufatura mais apropriado, quando opera-se com baixa escala e alto escopo. Vários

trabalhos na literatura comparam o desempenho de um sistema tipo *job-shop* com células de manufatura. Para conhecer melhor o impacto nas medidas de desempenho de se adotar um ou outro tipo de sistema de manufatura, alguns desses trabalhos de pesquisa foram estudados.

Flyn e Jacobs (1986, 1987) realizaram simulações combinando duas características marcantes e distintas nos sistemas tipo *job-shop* e manufatura celular: arranjo físico das máquinas (disposição celular ou por processo) e máquinas do mesmo tipo dedicadas ou intercambiáveis. Os resultados alcançados foram inesperados para os autores. Embora as estações de trabalho (*shops*) que utilizaram disposição celular tenham apresentado tempo médio de preparação de máquina e distância média viajada entre pares de máquinas sucessivas menores (comparados a *job-shops* tradicionais), surgiram problemas em relação a outros aspectos. A dedicação de máquinas, decorrente da aplicação da tecnologia de grupo, levou à formação de longas filas, maior estoque em processo (*WIP*) e tempos de espera mais elevados. A flexibilidade de rota (característica existente nos sistemas que adotam disposição por processo) perdida pela dedicação de máquinas foi atribuída como causa destes efeitos negativos. Estações com máquinas intercambiáveis tiveram desempenho global superior, enquanto células de manufatura limitaram a flexibilidade das estações e resultaram em desempenhos piores. Flynn e Jacobs (1987) reportam também que o efeito de dedicação de máquinas pode ser mais importante que o efeito de como as máquinas são fisicamente agrupadas.

Morris e Tersine (1990), por meio de experimentos com dados hipotéticos em um modelo de simulação, compararam disposição por processo com disposição celular. Quatro experimentos de simulação foram conduzidos para analisar a influência das seguintes variáveis operacionais na escolha do tipo de disposição: a) razão (tempo de preparação/tempo de processamento); b) tempo de transferência de material entre centros de trabalho; c) estabilidade da demanda e; d) fluxo de trabalho dentro das células. Os experimentos foram planejados para testar a robustez da disposição escolhida a variáveis operacionais baseada nas medidas de desempenho: estoque de trabalho em processo e taxa média de produção do sistema (*throughput*). Os autores concluíram em seus experimentos que nenhuma das mudanças individuais em níveis de variáveis operacionais produziram uma clara vantagem para disposições celulares.

Suresh (1991) desenvolveu aproximações analíticas baseadas em modelos de filas $M/M/m$ (a notação utilizada para sistemas de filas é definida adiante neste capítulo) para várias características de desempenho de um centro de trabalho tipo *job-shop*. Essas expressões foram usadas para analisar os efeitos de partição e introdução de elementos de tecnologia de grupo. O autor utiliza o termo partição para referir-se à implementação formal de tecnologia de grupo, envolvendo a abordagem de dedicação de máquina, sem movimentação do equipamento em um primeiro momento e, em seguida, a abordagem celular, envolvendo a dedicação da máquina assim como a movimentação do equipamento para criação das células. Foi mostrado por meio de simulação e pelas aproximações analíticas que a partição da instalação resulta em vantagens significativas para os arranjos celulares em redução de tempo de preparação, mas também efeitos adversos na *remainder cell* (como, na prática, raramente é possível dividir todos itens/produtos e máquinas em células independentes, tal que cada célula seja responsável pela produção total de uma classe ou de um conjunto de itens/produtos, torna-se necessária a criação de um *job-shop* - a *remainder cell* - que se refere ao grupo residual de itens/produtos e máquinas).

Suresh (1991) concluiu que efeitos globais para situações híbridas (configuração entre a disposição por processo e a disposição celular pura, como por exemplo partilhar equipamentos entre células) não são favoráveis quando comparadas à disposição por processo. A pesquisa também conclui que células maiores, com múltiplos servidores dentro delas, são superiores em desempenho a células menores, excessivamente particionadas.

Suresh (1992) utilizou modelos analíticos de filas Markovianas para investigar o efeito de particionar centros de trabalho tipo *job-shop* para implementar células de manufatura e comparou os dados com experimentos de simulação. Foi mostrado que a partição de uma estação de trabalho conduz a efeitos adversos nas características do fluxo. Para um nível específico de *WIP*, a utilização da máquina pode ser menor em sistemas particionados. O autor ainda fornece implicações práticas dos resultados alcançados: há um nível mínimo de redução de tempo de preparação requerido para considerar a partição; políticas operacionais como a permissão de movimentos intercelulares foram avaliadas como sendo muito úteis, quando baixos graus de redução de tempo de preparação podem ser atingidos; células maiores

possuindo certa flexibilidade de rota possuem melhor desempenho que células em sistemas excessivamente particionados, pois assim como a permissão de movimentos intercelulares, elas são capazes de absorver desbalanceamentos de cargas melhor que sistemas excessivamente particionados.

Suresh e Meredith (1994) observaram que a redução dos tempos de preparação e de processamento resultaram em impactos maiores no *WIP* e na utilização da máquina do que o tamanho de lote e as variabilidades em tempos de processamento e em intervalos de tempo entre chegadas de tarefas/itens/produtos. Os resultados obtidos por meio da simulação foram consistentes com os previstos pelo modelo analítico.

Jensen et al. (1996) investigaram o conflito (*trade-off*) no desempenho entre a flexibilidade de rota de máquinas não dedicadas em um *job-shop* funcional, e a eficiência de tempo de preparação em máquinas dedicadas em sistemas que têm máquinas com disposição celular. Concluíram que em geral a disposição celular pura pode não operar tão bem, mesmo se uma grande percentagem de redução de preparação puder ser obtida pela dedicação. Mostraram que uma disposição híbrida (uma configuração intermediária à disposição por processo e celular pura) pode operar bem se a dedicação das máquinas resultar em redução significativa de tempos de preparação, tal como concluiu Suresh (1992). Ainda afirmaram que a abordagem de disposição celular pura seria benéfica somente se a capacidade existente para qualquer tipo de máquina pudesse ser facilmente dividida entre as células, visto que os efeitos negativos de utilização desbalanceada de máquinas não são superadas pelos ganhos de redução de tempo de preparação devido à dedicação. Reportaram ainda que a utilização desbalanceada de máquinas também pode resultar de variabilidade na demanda e a alta variabilidade na mistura (*mix*) de produtos é mais danosa ao desempenho de disposições celulares que variabilidade no volume do produto.

Jensen et al. (1996) propõem como alternativa a adoção de disposições celulares com flexibilidade de rotas, permitindo balancear a utilização das máquinas e superar os efeitos negativos da variabilidade da mistura (*mix*) de produtos. Para altos níveis de redução de tempo de preparação, um sistema com disposição celular e flexibilidade de rotas mostrou melhor desempenho do que um sistema eficiente com disposição funcional. A habilidade das células com flexibilidade de rota em trocar eficiência de tempo de preparação por uma quantidade moderada de flexibilidade de

processo foi vantajosa em várias circunstâncias. Por exemplo, quando utilizações de máquinas não podiam ser balanceadas entre as células, ao invés de forçar uma máquina em um célula operar com utilização muito alta e sua duplicata em outra célula operar com utilização muito baixa, as capacidades destas máquinas e seus requisitos de processo foram reunidos. Os autores alertam que dedicação de máquinas sem altos ganhos de especialização pode não ser valiosa.

Dos trabalhos revisados, é perceptível que, a depender da medida de desempenho de interesse, uma forma de arranjo (*job-shop* ou manufatura celular) possui vantagens e desvantagens sobre a outra. Nesta tese, um dos objetivos é considerar o conceito de fábrica focalizada como uma alternativa possível ao sistema *job-shop* com máquinas intercambiáveis, visando diminuir a complexidade deste sistema. Entretanto, o foco está em decisões de médio a longo prazo (problemas tático-estratégicos), como por exemplo no projeto e planejamento de sistemas *job-shop* e no projeto de fábricas focalizadas. Aqui, o projeto de fábrica focalizada é abordado do ponto de vista da partição da instalação de um *job-shop* em sub-plantas, e da alocação de capacidade em cada sub-planta. Por ser este um tópico importante para a presente pesquisa, é dedicado um capítulo (capítulo 5) sobre o problema estratégico de partição da instalação.

2.2 Redes de Filas

Redes de filas são formadas de vários sistemas de filas conectados entre si, com usuários deslocando-se entre eles para receber serviço. Geralmente, as filas são referenciadas como nós, centros de serviço ou estações, e o caminho interno dos usuários na rede como rota ou roteiro.

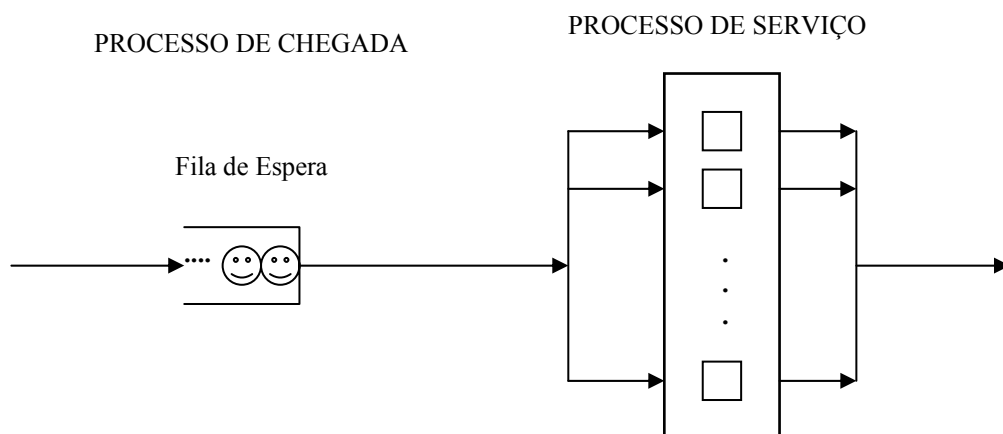
Redes de filas podem ser usadas para representar sistemas de manufatura onde os nós representam as estações (*shops*); os itens/produtos (*jobs*), os usuários requerendo serviço das estações e; os arcos conectando os nós, as rotas dos produtos. Uma rede de filas é caracterizada pelo número de estações (nós), roteiro dos itens/produtos (seqüência de operações) e pelo tipo de rede (aberta, fechada ou mista).

O número de nós na rede corresponde ao número de estações. O roteiro, determinístico ou probabilístico, pode ser seqüencial, seqüencial com realimentação, de montagem, cíclico e acíclico. Rotas probabilísticas podem ser empregadas para

representar, por exemplo, a probabilidade de quebra de máquina ou a necessidade de retrabalho em um item/produto. Uma estação pode ser visitada mais de uma vez por um mesmo item/produto para realizar operações diferentes.

A rede é aberta ou fechada dependendo de poder enviar e receber ou não usuários de fora da rede. Isto é, em uma rede fechada (*CQN* ou *closed queueing network*), o número total de usuários não se altera, havendo apenas uma permutação nas suas posições. Entretanto, pode-se definir uma estação artificial externa com carga e descarga instantânea, tal que para cada chegada de um item/produto nesta estação haja uma partida de outro desta estação, de modo a manter o número de itens/produtos na *CQN* constante. Deste modo, é possível representar um sistema de manufatura flexível (*FMS* ou *flexible manufacturing systems*) por meio de uma *CQN*. Já em uma rede aberta (*OQN* ou *open queueing network*), o número total de itens/produtos circulando é uma variável aleatória e pode variar pela chegada ou saída externa de itens/produtos. Se a rede tiver múltiplas classes de produtos, pode-se redefinir sub-redes abertas para algumas classes e sub-redes fechadas para outras. Neste caso a rede é denominada mista. Neste trabalho, sistemas de manufatura são modelados por meio de redes de filas abertas (*OQN*).

Para especificar um nó na rede, devem ser identificados os processos que descrevem a chegada e o serviço, assim como a estrutura e disciplina de serviço do servidor. A figura 2.2, ilustra um sistema de filas mono estágio.



Fonte: Adaptada de Morabito (1998).

FIGURA 2.2 – Sistema de fila com único estágio.

O processo de chegada é descrito pelo intervalo dos tempos entre chegadas de usuários no sistema e pode ser determinístico ou probabilístico. Se for probabilístico, os intervalos de tempos entre chegadas podem ser dependentes de outros intervalos de tempo entre chegadas e/ou do processo de serviço, ou consistir de intervalos de tempo entre chegadas que formam um processo de renovação (ou seja, independentes e identicamente distribuídos, *iid*), hipótese freqüentemente admitida em teoria das filas. Quando há vários tipos de itens/produtos, é possível agregá-los em uma única classe (supostos idênticos estatisticamente) ou em várias classes diferentes. A chegada na estação pode ser individual ou em lotes (*batch*). Para determinadas situações é conveniente estabelecer um lote como sendo um item/produto, como foi estabelecido no estudo de caso (vide capítulo 4).

A segunda grandeza estatística que deve ser descrita é a quantidade de tempo demandada por uma chegada no servidor, ou o tempo gasto pelo usuário no servidor, usualmente referida como tempo de serviço e pode ser determinístico ou probabilístico. Se for probabilístico, pode depender de outros tempos de processamento e/ou do processo de chegadas, ou então ser independente e identicamente distribuído (*iid*). Itens/produtos podem ser processados individualmente ou em lotes, que também podem ser considerados para determinadas situações como um único item/produto. Cada máquina pode representar um conjunto de recursos como máquinas, operadores, ferramentas, etc.

Com relação à estrutura e disciplina de serviço do servidor, é preciso especificar:

- Capacidade da fila;
- Número de servidores;
- Número de classes de usuários/itens/produtos (*jobs*);
- Disciplina de atendimento: por exemplo, *FCFS* (*first-come-first-serve*) e *LCFS* ou *LIFO* (*last-come-first-serve* e *last-in-first-out*); e em caso de existência de prioridades entre classes de produtos, é preciso especificar se é preemptiva (interrupção do serviço de um item/produto com a chegada de outro na fila com maior prioridade) ou não-preemptiva (o item/produto com maior prioridade

chegando na fila só entra em serviço após o término do serviço do item/produto com menor prioridade).

Um sistema de filas, neste trabalho, é identificado (codificado) na forma $A/B/m/K/M$, onde:

- A e B são letras simbólicas representando as distribuições dos tempos entre chegadas e de serviço, respectivamente. As distribuições A e B podem ser de Poisson (M), determinística (D), Erlang de ordem k (E_k), Hiperexponencial (H_k) ou qualquer distribuição (G ou GI , de genérica e genérica independente, respectivamente);
- m , K e M são constantes inteiras representando o número de servidores, a capacidade de usuários no sistema (incluindo em serviço) e o tamanho da população, respectivamente. É interessante frisar que quando algumas destas constantes não são especificadas, elas são consideradas como infinitas. Por exemplo, a notação $M/D/4$ representa um sistema de filas com chegadas de acordo com distribuição de Poisson, distribuição de tempo de serviço determinística, com quatro servidores e capacidade de usuários no sistema e população infinitas.

Tipicamente, algumas medidas de desempenho de interesse em um sistema de filas são:

- tempo que o usuário permanece no sistema: é a soma do tempo de espera em fila com o gasto em serviço. Em sistemas de manufatura esta grandeza é conhecida como *leadtime* de produção.
- número de usuários/itens/produtos no sistema: é a soma dos usuários/itens/produtos em fila e em serviço. Para sistemas de manufatura, esta grandeza também pode ser representada por um valor monetário correspondente ao capital de trabalho associado ao inventário em processo (*WIP*), ou simplesmente pelo número de usuários/itens/produtos em processo (*jobs* em processo).
- período de tempo em que o servidor está ocupado ou vazio. Esta grandeza é útil em sistemas de manufatura para avaliar o que é denominado carga de trabalho (*workload*), ou intensidade de tráfego (*traffic intensity*), ou nível de utilização da capacidade.

Bitran e Tirupati (1988) apresentaram uma classificação para os modelos de avaliação de desempenho segundo o método de abordagem:

- métodos exatos;
- métodos aproximados;
- simulação e técnicas relacionadas.

Métodos exatos existem para sistemas Markovianos (Redes de Jackson, de Kelly e *BCMP*). Estes métodos são tratados adiante com maior ênfase nas Redes de Jackson (seção 2.3).

A falta de sucesso em obter soluções exatas para redes mais gerais tem motivado pesquisadores a desenvolver aproximações para avaliar medidas de desempenho. Os métodos de aproximação podem ser classificados em quatro categorias (Bitran e Dasu, 1992; Bitran e Morabito, 1996):

- métodos de decomposição paramétricos;
- aproximação por difusão;
- análise operacional;
- análise do valor médio.

Para os objetivos deste trabalho foram estudadas as aproximações geradas pelo método de decomposição, por serem bastante flexíveis e de grande potencial para aplicações práticas. Métodos aproximados de decomposição são tentativas de generalizar a noção de independência e resultados na forma de produto dos sistemas Markovianos para sistemas mais gerais. Estes sistemas são denominados Redes Generalizadas ou Genéricas. O método e aproximações provenientes são descritos nas seções 2.4 e 2.5.

A simulação discreta baseada em eventos é muito utilizada na ausência de resultados analíticos exatos. Simulação de Monte-Carlo tem sido muito empregada para avaliar grandes sistema de redes de filas. É uma abordagem que permite o uso de hipóteses mais detalhadas e que estejam mais próximas da realidade operacional. Entretanto, o custo está relacionado aos requisitos computacionais, exigindo tempos longos de processamento, o que implica em um pequeno número de alternativas sendo testadas (Bitran e Tirupati, 1988). Além disso, há dificuldades para integrar modelos de simulação dentro de procedimentos de otimização, para prescrever a melhor configuração para o sistema. Neste trabalho, a técnica de simulação é empregada como

uma forma de validar resultados gerados pelas aproximações analíticas e, portanto, somente os tópicos necessários e indispensáveis à pesquisa são reportados.

Para deixar mais clara a terminologia adotada neste trabalho, convém definir os termos aqui empregados.

- Operação: uma operação é uma tarefa elementar que requer recursos tal como máquinas, ferramentas e trabalho.
- Item: é uma parte física distinta produzida pela instalação. Associado a cada item está um conjunto de operações e uma relação de precedência, que pode restringir a seqüência em que as operações podem ser executadas. Por simplicidade, um item, parte, submontagem ou montagem processada por uma máquina é chamado simplesmente por usuário ou item/produto/tarefa (*job*).
- Produto: um *commodity* distinto produzido pela instalação. Geralmente um produto é produzido através da montagem de vários itens.
- Classe de produtos: conjunto de produtos que tem requisitos de manufatura similares. Para os propósitos deste trabalho, todos os produtos pertencentes à mesma classe seguem rotas similares pela instalação.
- Centro de máquinas ou estação de trabalho: um conjunto de máquinas capaz de executar operações similares, embora não necessariamente idênticas. Um exemplo de uma estação seria um conjunto de máquinas perfuradeiras com capacidades variadas.

2.3 Redes de Filas Markovianas (Método Exato de Decomposição)

Primeiramente, por se tratar de uma seção carregada de símbolos e índices, convém definir a notação aqui empregada:

n	número de estações internas na rede
$E(a_0)$	valor esperado do intervalo de tempo entre chegadas externas, a_0 .
λ_0	taxa média de chegada externa (nó 0) na rede ($\lambda_0 = 1/E(a_0)$).
λ_{ij}	taxa média de chegada na estação j a partir da estação i .
λ'_k	taxa média de chegada externa de itens/produtos da classe k , ($\lambda'_k = 1/E(a'_k)$).

r_{ij}	probabilidade de um item/produto, após ser atendido na estação i , seguir para a estação j .
m_j	número de máquinas na estação j .
$E(s_j)$	valor esperado do tempo de serviço de um item/produto na estação j , s_j .
μ_j	taxa média de serviço para cada máquina na estação j ($\mu_j = 1/E(s_j)$).
L_j	número de itens/produtos em cada estação j . L é o número total de <i>jobs</i> no sistema ou seja, $L = \sum_{j=1}^n L_j$
ρ_j	é a intensidade de tráfego (ou nível de utilização média) na estação j ($\rho_j = \lambda_j / m_j \mu_j$)

2.3.1 Redes de Jackson

Uma das grandes contribuições em redes de filas foi dada por Jackson (1957, 1963), que fornece resultados para distribuição de equilíbrio do número de usuários para uma variedade de sistemas, referidos por isso como redes de Jackson. São aplicáveis quando trata-se de estações com processos de chegada Poisson e serviço exponencial, roteiro markoviano e apenas uma classe de itens/produtos. As principais características destes sistemas são (Magalhães, 1996; Kleinrock, 1975; Bitran e Tirupati, 1988):

- processo de chegada é Poisson com parâmetro $\lambda(L)$. Ao permitir que o processo de chegadas possa depender do número total de itens/produtos no sistema, L , redes fechadas (*CQN*) podem ser consideradas casos particulares de *OQN* pois podem ser criadas a partir delas.
- para cada nó j pertencendo à rede, o atendimento é em ordem de chegada com distribuição exponencial de parâmetro $\mu_j(L_j)$, $\mu_j(0) = 0$.
- após o usuário completar seu serviço na estação i , ele move-se para mais serviço na estação j com probabilidade $r_{i,j}$, onde i, j são nós quaisquer da rede, permitindo inclusive que haja retornos à mesma estação. Por outro lado, se o usuário já completou sua demanda de serviço na rede, ele deixa o sistema após ser atendido na estação i , com probabilidade $r_{i,0} = 1 - \sum_{j=1}^n r_{i,j}$, onde 0 é o nó artificial representando o exterior da rede. A figura 4.1 (seção 4.2.1) ilustra a rede utilizada para representar o estudo de caso

com o nó artificial. Todas as transferências são supostas instantâneas e, assim, o tempo gasto no sistema é proveniente de esperas e atendimentos nas estações.

d) todos os nós da rede podem receber usuários externos de forma direta ou indireta. Isto é, qualquer que seja j ($j=1, \dots, n$), ou $\lambda_{0j} > 0$ ou existe um i tal que $\lambda_{0i} > 0$ e i se comunica com j (ou seja, é possível, em algum número de etapas, partir de i e alcançar j).

e) o usuário servido em j tem probabilidade positiva de sair da rede em uma ou mais etapas. Ou seja, qualquer que seja j ($j=1, \dots, n$), ou $r_{j,0} > 0$ ou existe um i tal que $r_{i,0} > 0$ e i se comunica com j .

O fluxo de itens/produtos entre as estações é descrito pela cadeia de Markov com matriz de transição $\mathbf{R} = \{r_{ij}, 0 \leq r_{ij} \leq 1, i = 0, \dots, n, j = 0, \dots, n\}$, onde $\sum_{j=0}^n r_{ij} = 1$ para $i = 0, \dots, n$ e $r_{00} = 0$ por definição. Seja $\mathbf{Q} = \{q_{ij}, i = 1, \dots, n, j = 1, \dots, n\}$ e $q_{i0} = 1 - \sum_{j=1}^n q_{ij}$. Em palavras, \mathbf{Q} é a matriz \mathbf{R} sem a linha 0, que descreve a probabilidade de um item/produto entrar no sistema por uma estação j , e sem a coluna 0, que descreve a probabilidade de um item/produto deixar o sistema a partir de uma estação j . Ou seja, \mathbf{Q} é a matriz subestocástica representando a rota probabilística de itens/produtos entre as estações internas da rede.

As características (d) e (e), acima relatadas, garantem que todos os centros de serviço podem receber usuários e que todos que entram têm probabilidade positiva de sair em algum momento. Com estas condições, a cadeia de Markov associada com a matriz \mathbf{Q} será recorrente não-nula e as equações de tráfego abaixo terão solução única:

$$\lambda_j = \lambda_{0j} + \sum_{i=1}^n \gamma_i q_{ij} \lambda_i \quad j = 1, 2, \dots, n \quad (1)$$

Pode-se interpretar λ_j como sendo a taxa total de entrada ao nó j , resultando da composição de chegadas externas (primeiro termo do lado direito de (1)) com as saídas de estações internas da rede que se movimentam para a estação j (segundo termo do lado direito de (1)). Em (1), γ_i representa a criação (se $\gamma_i > 1$) ou combinação (se $\gamma_i < 1$) de itens. Se itens não são criados ou combinados, então $\gamma_i = 1$.

Um teorema importante provado por Jackson (1957, 1963) afirma que a distribuição estacionária conjunta do número de usuários presentes no sistema pode ser

calculada pelo produto das distribuições estacionárias marginais (referentes a cada estação). Esta solução é denominada solução na forma produto.

Teorema de Jackson: Seja o vetor $\mathbf{L} = (L_1, L_2, \dots, L_n)$, o número de usuários em fila e em serviço na estação j , definindo o estado do sistema. Jackson (1957, 1963) mostrou que se $\rho_j = \lambda_j / m_j \mu_j < 1$ para $j = 1, 2, \dots, n$, então a probabilidade do sistema estar em equilíbrio no estado \mathbf{L} , $\pi(\mathbf{L})$, é dada pela seguinte forma de produto:

$$\pi(\mathbf{L}) = \prod_{j=1}^n \pi_j(L_j) \quad (2)$$

com:

$$\pi_j = f_j(L_j) \pi_j(0) \quad \text{para } L_j = 0, 1, \dots$$

$$f_j(L_j) = f_j(L_j - 1) \frac{\lambda_j}{\mu_j r_j(L_j)} \quad \text{para } L_j = 1, 2, \dots$$

$$f_j(0) = 1$$

$$\pi_j(0) = \frac{1}{\sum_{L_j=0}^{\infty} f_j(L_j)}$$

em que

$\pi(\mathbf{L})$ é a distribuição conjunta do número de *jobs* em cada estação j ;

$\mu_j r_j(L_j)$ é a taxa de serviço na estação j como função do número de usuários;

$\pi_j(L_j)$ é a probabilidade da estação j ter L_j usuários.

Particularmente, quando $r_j(L_j) = \min\{L_j, m_j\}$ e o sistema de filas possui capacidade e população infinitas, então $\pi_j(L_j)$ e $\pi_j(0)$ têm a seguinte forma:

$$\pi_j(L_j) = \begin{cases} \frac{(\lambda_j / \mu_j)^{L_j}}{L_j!} \pi_j(0), & \text{se } L_j \leq m_j \\ \frac{(\lambda_j / \mu_j)^{L_j}}{m_j^{L_j - m_j} m_j!} \pi_j(0), & \text{se } L_j > m_j \end{cases} \quad (3)$$

com

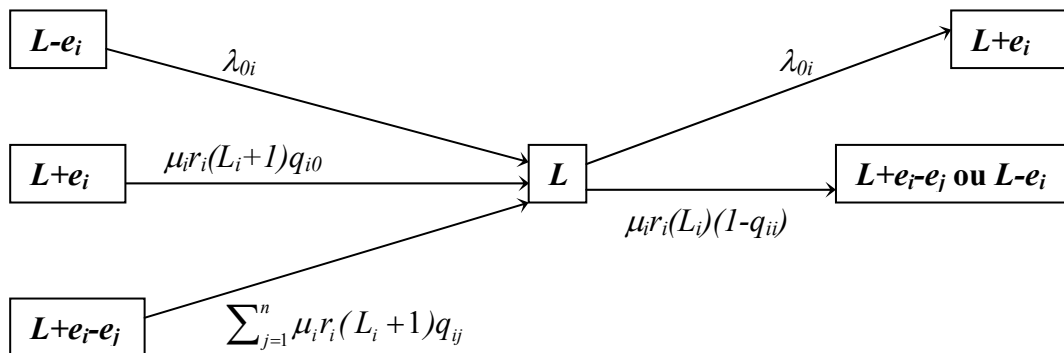
$$\pi_j(0) = \left\{ \sum_{t=0}^{m_j-1} \frac{(\lambda_j / \mu_j)^t}{t!} + \frac{(\lambda_j / \mu_j)^{m_j}}{(1 - \rho_j) m_j!} \right\}^{-1}$$

A prova deste teorema consiste em mostrar que a solução na forma produto, expressão (2), satisfaz as equações de balanço do sistema:

$$\begin{aligned} & \sum_{i=1}^n \lambda_{0i} \pi(\mathbf{L} - \mathbf{e}_i) + \sum_{i=1}^n \mu_i r_i(L_i + 1) q_{i0} \pi(\mathbf{L} + \mathbf{e}_i) + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \mu_i r_i(L_i + 1) q_{ij} \pi(\mathbf{L} + \mathbf{e}_i - \mathbf{e}_j) = \\ & = \sum_{i=1}^n [\lambda_{0i} + \mu_i r_i(L_i)(1 - q_{ii})] \pi(\mathbf{L}) \quad \mathbf{L} \in N^n \dots \end{aligned} \quad (4)$$

com \mathbf{e}_i denotando um vetor n -dimensional com 1 na posição i e 0 nas demais posições.

A equação de balanço (4) pode ser interpretada da seguinte forma: o primeiro termo do lado esquerdo corresponde à taxa de transição do estado $\mathbf{L} - \mathbf{e}_i$ para o estado \mathbf{L} com a chegada externa de um usuário; o segundo termo corresponde à taxa de transição do estado $\mathbf{L} + \mathbf{e}_i$ para o estado \mathbf{L} com o término de um serviço sobre um usuário que sai do sistema pela estação i ; o terceiro termo corresponde à taxa de transição do estado $\mathbf{L} + \mathbf{e}_i - \mathbf{e}_j$ para o estado \mathbf{L} pela transferência de um usuário que completa seu serviço na estação i e segue para a estação j , não saindo do sistema. Já o lado direito da equação (4) corresponde à taxa de transição do estado \mathbf{L} para outros estados pela chegada externa de um usuário ou o término de serviço de um usuário na estação i que não retorna imediatamente à estação i . A figura 2.3 ilustra a transição de estados.



Fonte: Morabito (1998).

FIGURA 2.3 – Diagrama de taxas de transição de estados.

“O resultado na forma produto da solução do teorema de Jackson é extraordinário, pois indica que o sistema se comporta como se cada estação j fosse uma fila com chegada e serviços exponenciais com taxas λ_j e $\mu_j r_j(L_j)$, respectivamente, independente das demais estações de serviço” (Magalhães, 1996). Assim no caso de m_j servidores idênticos, o sistema se comportaria como se fosse uma coleção independente

de filas $M/M/m_j$ e as distribuições marginais teriam a forma da equação (3). “Contudo é importante frisar que devido ao roteiro interno, as chegadas a cada centro não formam, em geral, um processo de Poisson” (Magalhães, 1996). Basta, por exemplo, que haja realimentação em uma estação. Portanto, a solução na forma produto deve ser aplicada com cuidado, pois sugere independência entre o número de usuários em cada estação de trabalho para um determinado tempo t . Entretanto, os números de usuários em instantes de tempos diferentes não são independentes apesar de serem identicamente distribuídos no equilíbrio.

Assim, partindo-se dos parâmetros iniciais $\{m_j, \lambda_{0j}, \mu_j\}$ e da matriz \mathbf{Q} , após resolver o sistema de equações (1) para determinar cada λ_j , pode-se decompor a rede em n estações individuais $M/M/m_j$, cada uma descrita por $\{m_j, \lambda_j, \mu_j\}$. Para o cálculo das medidas de desempenho, cada estação é considerada separadamente das demais. Por exemplo, o número médio de usuários, $E(L_j)$, na estação j é dado por (Tijms, 1986):

$$E(L_j) = \sum_{L_j}^{\infty} L_j \pi_j(L_j) = \frac{\rho_j (\lambda_j / \mu_j)^{m_j} \pi_j(0)}{(1 - \rho_j)^2 m_j!} + \frac{\lambda_j}{\mu_j} \quad (5)$$

com $\pi_j(0)$ dado em (3).

Outras medidas de desempenho tais como número médio de usuários em fila, $E(Lq_j)$, o tempo médio de espera em fila, $E(Wq_j)$, e o tempo médio de permanência em fila e em serviço $E(W_j)$ na estação j para quaisquer sistemas de filas (não necessariamente $M/M/m_j$) são calculadas a partir da lei de Little (6.a) e das relações subsequentes (6.b e 6.c) (Kleinrock, 1975):

$$E(L_j) = \lambda_j E(W_j) \quad (6.a)$$

$$E(L_j) = E(Lq_j) + \lambda_j / \mu_j \quad (6.b)$$

$$E(W_j) = E(Wq_j) + E(s_j) \quad (6.c)$$

O número médio de *jobs* na rede pode então ser calculado por:

$$E(\mathbf{L}) = \sum_{j=1}^n E(L_j) \quad (7)$$

Dado que o número de usuários na rede é estatisticamente independente (resultado extraído de (2)), a variância pode ser calculada da seguinte forma:

$$V(\mathbf{L}) = \sum_{j=1}^n V(L_j) \quad (8)$$

O *leadtime* médio (tempo médio total, incluindo tempo de espera e de processamento, gasto pela rede para processar um usuário ou um lote qualquer), $E(T)$, pode ser calculado por:

$$E(T) = \sum_{j=1}^n E(V_j)E(W_j) = \sum_{j=1}^n E(V_j)[E(Wq_j) + E(s_j)] \quad (9)$$

com $E(Wq_j)$ representando o atraso médio em uma fila $M/M/m_j$ da estação j , obtido da manipulação de (5), (6a) e (6c), sendo dado por:

$$E(Wq_j) = \frac{(\lambda_j / \mu_j)^{m_j} \pi_j(0)}{\mu_j m_j (1 - \rho_j)^2 m_j!} \quad (10)$$

com $\pi_j(0)$ definido em (3) e $E(V_j)$ representando o número médio de visitas de um usuário arbitrário na estação j , dado por:

$$E(V_j) = \frac{\lambda_j}{\lambda_0} \quad (11)$$

onde $\lambda_0 = \sum_{j=1}^n \lambda_{0j} = \sum_{j=1}^n \lambda_{j0}$, no equilíbrio, é a taxa média de produção da rede (*throughput*).

2.3.2 Redes de Kelly e BCMP

Extensões da solução na forma produto das redes de Jackson foram formuladas por Kelly (1975, 1979) e Bastkett, Chandy, Muntz e Palacios (1975), e ficaram conhecidas como redes de Kelly e *BCMP*, respectivamente. Apesar de alguns autores se referenciam às redes de Jackson, Kelly e *BCMP* genericamente por Redes de Jackson (veja por exemplo Bitran e Morabito, 1996; Bitran e Dasu, 1992; Whitt, 1983), aqui far-se-á a distinção. Quando desnecessária e em momento oportuno, estas redes serão referenciadas simplesmente por Redes Markovianas.

Para os propósitos deste trabalho, são reportadas as hipóteses e extensões destas redes com relação as de Jackson, sem muita atenção aos aspectos quantitativos dos modelos uma vez que, como as Redes de Jackson, são pouco aplicáveis a sistemas de manufatura, dado que a variabilidade do tempo de serviço é superestimada pelas distribuições exponenciais (Bitran e Tirupati, 1988). Entretanto, o método de

decomposição, muito adotado em literatura (veja por exemplo Whitt, 1983a; Whitt, 1983b, Shanthikumar e Buzacott, 1981; Bitran e Tirupati, 1988; Bitran e Morabito, 1996; Bitran e Dasu, 1992) para tratar redes que representam sistemas de manufatura, é uma extensão da solução na forma de produto obtida para estas redes.

Redes de Kelly e *BCMP* são convenientes quando as necessidades de processamento em cada estação são significativamente diferentes entre as classes (o que inviabiliza a agregação de classes para aplicação de modelo para rede de Jackson) e a disciplina de atendimento é diferente da *FCFS*, disciplina das redes de Jackson.

A Rede de Kelly estendeu a solução na forma produto para redes com múltiplas classes de produtos, roteiros determinísticos para algumas ou todas as classes de *jobs*, e disciplina de atendimento com prioridades. As principais características dessas redes são (Bitran e Tirupati, 1988; Magalhães, 1996):

- a) usuários da classe k chegam a rede de acordo com processo de Poisson com taxa λ'_k ;
- b) tempo de serviço em cada estação é homogêneo e tem distribuição exponencial;
- c) é permitido roteiro determinístico para algumas ou todas as classes;
- d) é possível definir uma variedade de disciplinas de filas. Por exemplo, seja $\beta_j(l, L_j)$ a proporção de esforço dirigida ao usuário na posição l da fila na estação j com L_j usuários. A taxa de serviço para este usuário é, portanto, $\mu_j \beta_j(l, L_j)$. Ao deixar a estação, todos os usuários da posição $l+1$ a L_j movem-se uma posição à frente. É importante deixar claro que $\sum_{l=1}^{L_j} \beta_j(l, L_j) = 1$. No caso de $\beta_j(1, L_j) = 1$ significa que o usuário na posição 1 é servido primeiro exclusivamente;
- e) $\delta(l, L_j)$ define a probabilidade de que uma nova chegada na estação j com L_j usuários ocupe a posição l na fila. Particularmente, quando $\delta(L_j + 1, L_j) = 1$, a nova chegada na estação j ocupa a última posição na fila. Novamente, $\sum_{l=1}^{L_j+1} \delta(l, L_j) = 1$;
- f) estado do sistema é descrito pela especificação do número de usuários em cada estação e a classe do usuário em cada posição da fila.

Pode-se mostrar que, mesmo para essa classe mais geral de rede, a distribuição de equilíbrio é da forma produto. O leitor mais interessado pode recorrer aos trabalhos de Kelly (1975, 1979), Bitran e Dasu (1992), Bitran e Tirupati (1988) e Magalhães (1996) para mais detalhes.

Baskett et al. (1975) estenderam os modelos anteriores buscando maior flexibilidade no atendimento, além de permitir que a rede seja mista, fechada para algumas e aberta para outras classes de usuários.

Bitran e Dasu (1992) resumem as principais características em redes que permitem uma análise exata (Jackson, Kelly e *BCMP*):

- a) distribuição de tempo de serviço exponencial;
- b) processo de chegada à rede é um processo de Poisson;
- c) prioridade é independente da classe do usuário;
- d) requisição de serviço homogêneo em todas estações (a distribuição do tempo de serviço é independente da classe do usuário).

Resultados na forma produto são interessantes e úteis, porém de difícil implementação devido ao tamanho do espaço de estados. Além disso, as distribuições exponenciais superestimam a variabilidade dos tempos de serviço encontrados em muitas operações de manufatura (Bitran e Tirupati, 1988). Daí, o motivo pelo desenvolvimento de aproximações. Para os objetivos deste trabalho foram estudadas as aproximações geradas pelo método de decomposição, por serem bastante flexíveis e de grande potencial para aplicações práticas.

2.4 Método Aproximado de Decomposição

Os métodos de decomposição são em parte motivados pelas propriedades de redes de Jackson e podem ser vistos como tentativas para estender a forma produto da solução para redes mais gerais (Redes Generalizadas ou Genéricas). Resumidamente, o método aproxima os processos de chegada por processos de renovação, decompõe a rede em estações e as considera como se fossem estocasticamente independentes.

Whitt (1983a) afirma que o método aproximado de decomposição vai além dos métodos exatos por tratar redes não-Markovianas, ou seja, redes cujos processos de chegadas externas não são processos de Poisson e/ou as distribuições dos

tempos de serviço não são exponenciais. A abordagem geral do método é a representação de todos os processos de chegada e distribuições de tempos de serviço por poucos parâmetros, daí o nome método de decomposição paramétrico. As medidas de congestionamento de cada estação são então descritas por fórmulas aproximadas que dependem somente destes parâmetros.

O método envolve três passos essenciais (Shanthikumar e Buzacott, 1981):

Passo 1- Análise das interações entre estações da *OQN*.

Passo 2- Decomposição da rede de filas em subsistemas com uma única estação, e avaliação das medidas de desempenho para cada estação.

Passo 3- Recomposição dos resultados obtidos pelas estações decompostas com avaliação das medidas de desempenho da rede.

Bitran e Tirupati (1988) e Whitt (1983a) ressaltam a importância do Passo 1 e relatam os três processos básicos para analisar a interação entre estações em uma rede:

- a) Superposição das chegadas: representa a chegada agregada em uma estação, combinação dos fluxos individuais de outras estações.
- b) Partidas: é o resultado da combinação do processo de chegada agregada com o processo de serviço.
- c) Separação ou decomposição das partidas: este processo decompõe o processo de partida agregada nos fluxos individuais de partida dessa estação para outras estações (inclusive a externa).

Os processos (a) e (c) descrevem a rota dos *jobs* na rede. A equação (1) é resultado do processo de superposição nas chegadas em uma estação *j*. A figura 2.4 ilustra cada um dos processos acima.

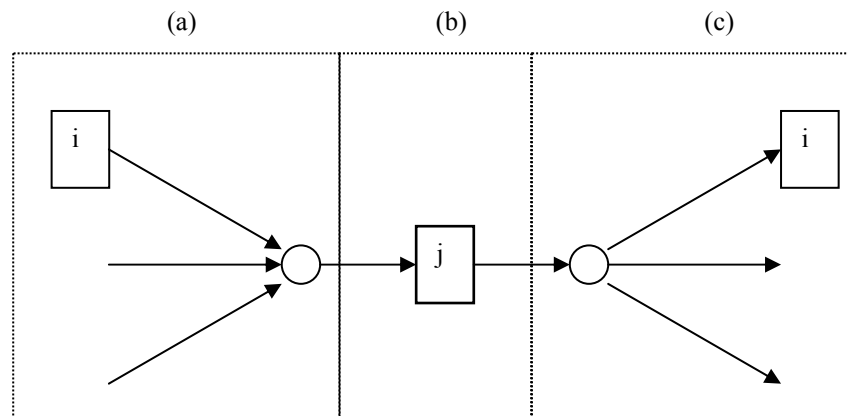


FIGURA 2.4 – (a) Processo de superposição de chegadas; (b) Processo de partidas e; (c) Processo de separação das partidas

Os parâmetros que descrevem os processos de chegada e de serviço em cada estação são parâmetros relacionados ao fluxo interno, determinados através de cálculos elementares nos três processos básicos existentes em cada estação da rede, descritos pelo Passo 1 do método. O método usa dois parâmetros para caracterizar cada processo de chegada e de serviço. Um para descrever a taxa e outro para descrever a variabilidade. Para o processo de serviço, os dois parâmetros são a média e o coeficiente quadrático de variação dos tempos de serviço, ou seja, funções dos dois primeiros momentos da distribuição de probabilidade. Para o processo de chegada, os parâmetros estão associados com aproximações a processos de renovação. Os dois primeiros parâmetros são funções dos dois primeiros momentos da distribuição do intervalo de renovação (intervalo entre ocorrências sucessivas) na aproximação do processo de renovação. Estes dois parâmetros são a taxa de chegada, que é o inverso do intervalo de renovação médio, e o coeficiente quadrático de variação dos intervalos de tempo entre chegadas. Em outras palavras, os parâmetros para descrever os processos de chegada e de serviço consistem na média e no coeficiente quadrático de variação dos intervalos de tempo entre chegadas e dos tempos de serviço, respectivamente. Bitran e Tirupati (1988) e Whitt (1983a) afirmam que estes dois parâmetros são suficientes para descrever redes de filas genéricas com precisão usualmente aceitável para as decisões tomadas aqui.

Uma vez que os parâmetros dos fluxos internos são determinados (Passo 1 do método), os nós são analisados separadamente (Passo 2). Além disso, quando as medidas de congestão são calculadas para a rede como um todo, os nós são tratados como se fossem aproximadamente independentes estocasticamente. Esta independência pode ser interpretada como uma generalização da solução na forma produto, válida para redes Markovianas conforme provado por Jackson (1957) e estendido por Kelly (1975, 1979) e Baskett et al. (1975). Enquanto o método de decomposição pode ser visto como uma extensão da solução na forma produto, esforço é feito para capturar a dependência entre os nós da rede. A idéia é representar aproximadamente esta dependência por meio dos parâmetros de fluxo interno, uma vez que estes são calculados no Passo 1 do método, que analisa a interação entre as estações da rede.

A seguir são apresentadas as aproximações geradas pelo método de decomposição abordados neste trabalho. A revisão baseia-se em Shanthikumar e Buzacott (1981, 1993), Whitt (1983a, 1983b), Bitran e Tirupati (1988), Segal e Whitt (1989), Albin (1982, 1984), Kuehn (1979), Bitran e Morabito (1996), Morabito (1998), Bitran e Dasu (1992) e Tijms (1986).

2.5 Redes Generalizadas (Método Aproximado de Decomposição)

Conforme abordado anteriormente, o método de decomposição aproxima os processos de chegada a um processo de renovação. Ao assumir que todo processo de chegada (não necessariamente Poissoniano) é aproximado por processos de renovação, é natural pensar, por exemplo, no modelo de uma *OQN* com filas *GI/G/m* (também referenciada como Rede Generalizada ou Genérica *GI/G/m*) como uma generalização a *OQN's* de Jackson com filas *M/M/m*. Na rede generalizada, cada nó é aproximado por um sistema de filas *GI/G/m* tendo um processo de chegadas de renovação independente dos tempos de serviço, que são independentes e identicamente distribuídos (*iid*) com uma distribuição genérica. Com esta aproximação, o método assume que toda rede pode ser representada por uma rede genérica, em que as redes Markovianas são um caso particular. Obviamente, se esta rede genérica apresenta uma única classe de usuários, se todos os processos de chegada externa são Poisson e se todos os tempos de serviço são exponencialmente distribuídos, então os resultados devem coincidir com as de uma rede

de Jackson exata com filas $M/M/m$, já que as distribuições exponenciais e os processos de Poisson são casos particulares de processos de renovação (vide Çinlar, 1975).

Com o intuito de generalizar as redes Markovianas para redes mais genéricas, aquelas em que os processos de chegadas externas não são necessariamente Poisson e a distribuição dos tempos de serviço não são necessariamente exponenciais, muitos pesquisadores (Shanthikumar e Buzacott (1981); Whitt, (1983a, 1983b); Bitran e Tirupati (1988); Segal e Whitt (1989); Albin (1982, 1984); Kuehn (1979); Bitran e Morabito (1996) e; Morabito (1998), dentre outros) aplicaram os três passos do método de decomposição e desenvolveram/aplicaram os modelos para redes com processos de chegada e de serviço mais genéricos. A seção 2.5.1 apresenta os modelos para redes com filas $GI/G/1$, classe única de produtos (usuários) e rota (roteiro de fabricação) probabilística. A seção 2.5.2 apresenta os modelos para redes com filas $GI/G/m$, com classe única de produtos e rota probabilística. A seção 2.5.3 apresenta os modelos para redes com filas $GI/G/m$, com múltiplas classes de produtos e rota determinística, incluindo as aproximações de tráfego leve.

2.5.1 Modelos de Redes com Filas $GI/G/1$, Classe Única e Rota Probabilística

Esta seção admite que todos os produtos pertencem a uma única classe de produtos (ou podem ser convenientemente agregados) e movem-se pelas estações da rede segundo um roteiro probabilístico. Cada estação possui um único servidor com processo de chegada e de serviço genéricos.

Como se tratam de aproximações geradas pelo método de decomposição, os processos de chegada externa e de serviço são aproximados por processos de renovação, ou seja, os intervalos de tempo entre chegadas externas (a_{0j}) e os tempos de serviço (s_j) em cada estação j são supostos *iid*.

Considere a seguinte notação para os dados de entrada (Bitran e Morabito, 1996):

n número de estações internas na rede.

Para cada estação $j=1, \dots, n$:

λ_{0j} taxa média de chegada externa na estação j ($\lambda_{0j}=1/E(a_{0j})$);

ca_{0j} coeficiente quadrático de variação do intervalo de tempo entre chegadas externas na estação j , ($ca_{0j}=V(a_{0j})/(E(a_{0j}))^2$), onde $V(x)$ e $E(x)$ representam, respectivamente, a variância e o valor esperado da variável aleatória x ;

μ_j taxa média de serviço na estação j ($\mu_j=1/E(s_j)$);

cs_j coeficiente quadrático de variação do tempo de processamento na estação j ($cs_j=V(s_j)/(E(s_j))^2$).

Para cada par (i,j) , $i=1,\dots,n$; $j=1,\dots,n$:

q_{ij} probabilidade de um usuário após terminar serviço em i , seguir para estação j ;

Então os dados de entrada se resumem a $4n \{\lambda_{0j}, ca_{0j}, \mu_j, cs_j\} + n^2 \{\mathbf{Q}\}$, onde $\mathbf{Q} = \{q_{ij}, i=1, \dots, n, j=1, \dots, n\}$ é a matriz subestocástica discutida anteriormente, com $q_{i0} = 1 - \sum_{j=1}^n q_{ij}$ e $q_{0j} = 1 - \sum_{i=1}^n q_{ij}$. Assume-se que a rede não possui estações com realimentação imediata. Caso a rede possua, Whitt (1983a) fornece um método de ajuste nos parâmetros iniciais para removê-los. O procedimento de remoção aperfeiçoa a qualidade das aproximações.

Na prática, os parâmetros iniciais de entrada (λ_{0j} e μ_j) podem ser estimados da seguinte maneira:

$$\lambda_{0j} = q_{0j} \lambda_0 \quad (12)$$

com

$$\lambda_0 = \lim_{t \rightarrow \infty} \frac{N(t)}{t} \quad (13)$$

sendo $N(t)$ o número de usuários que chegam na rede da estação externa 0 , durante o intervalo de tempo $(0,t]$ e $\mu_j=1/E(s_j)$ com (Bitran e Morabito, 1996):

$$E(s_j) = \lim_{u \rightarrow \infty} \frac{\sum_{v=1}^u s_{j,v}}{u} \quad (14)$$

onde $s_{j,v}$ é o tempo de processamento do v -ésimo usuário servido na estação interna j (sem incluir o tempo em fila).

2.5.1.1 Passo 1: Análise das interações entre estações da OQN

Esta etapa consiste em descrever cada estação j pelos parâmetros $\{\lambda_j, ca_j, \mu_j, cs_j\}$ e pela matriz $\{Q\}$ que descreve a rota, partindo dos parâmetros iniciais $\{\lambda_{0j}, ca_{0j}, \mu_j, cs_j\}$ e $\{Q\}$. Em outras palavras, determinar pelas operações de superposição das chegadas, do processo de partidas e da decomposição das partidas, a taxa média de chegada λ_j , definida como $\lambda_j=1/E(a_j)$ e o coeficiente quadrático de variação ca_j , sendo a_j o intervalo de tempo entre chegadas na estação j . Os dois parâmetros são a solução de dois sistemas lineares: as equações de taxa de tráfego (1) e as equações de variabilidade de tráfego, definidas a seguir. Primeiramente, obtém-se as taxas médias (exatas) por meio da solução das equações de taxa de tráfego e, em seguida, estas taxas são utilizadas para obter os parâmetros de variabilidade aproximados por meio de um segundo sistema linear, as equações de variabilidade de tráfego.

As equações de variabilidade de tráfego fornecem aproximações para o *scv* do intervalo de tempo entre chegadas ca_j em cada estação j . Estas aproximações combinam dois métodos básicos: o método assintótico e o método dos intervalos estacionários (conforme Whitt, 1983a; Bitran e Morabito, 1996).

O método assintótico, também chamado de macro, toma uma visão macroscópica do processo de chegada e tenta descrever seu comportamento num intervalo de tempo longo, na tentativa de levar em conta a dependência entre os intervalos de tempo entre chegadas. Suponha que chegadas venham ocorrendo em uma estação j desde $t=-\infty$. Seja $S_{j,p}$ o tempo decorrido desde a chegada ocorrida no instante $t=0$ até a ocorrência da p -ésima chegada na estação j . Ou seja, $S_{j,p}=a_{j,1} + a_{j,2} + \dots + a_{j,p}$, em que $a_{j,q}$ é o intervalo de tempo, suposto identicamente distribuído, entre a $(q-1)$ -ésima e a q -ésima chegadas na estação j (Bitran e Dasu, 1992; Bitran e Morabito, 1996). Assim, a média e variância dos intervalos de tempo entre chegadas na estação j ficam, respectivamente:

$$E(a_j) = \lim_{p \rightarrow \infty} \frac{E(S_{j,p})}{p} = \lim_{p \rightarrow \infty} \frac{\sum_{q=1}^p E(a_{j,q})}{p} = \frac{pE(a_{j,1})}{p} = E(a_{j,1})$$

$$V(a_j) = \lim_{p \rightarrow \infty} \frac{V(S_{j,p})}{p} = \lim_{p \rightarrow \infty} \frac{pV(a_{j,1}) + \sum_{q=1}^p \sum_{\substack{r=1 \\ r \neq q}}^p Cov(a_{j,q}, a_{j,r})}{p}$$

É perceptível pela expressão da variância que o método assintótico leva em conta a dependência dos intervalos de tempo entre chegadas, pois considera todos os termos de covariância.

Já o método dos intervalos estacionários ignora qualquer dependência no processo de chegada e toma uma visão microscópica, daí o porquê de também ser chamado de micro, buscando descrever o comportamento do processo num intervalo de tempo curto. As expressões, respectivamente, para a média e variância dos intervalos de tempo na estação j usando este método seguem:

$$E(a_j) = E(a_{j,1})$$

$$V(a_j) = V(S_{j,1}) = V(a_{j,1})$$

Ambos os métodos resultam na mesma média, mas as variâncias podem diferir substancialmente. Para o caso particular de o processo de chegadas ser de renovação, então ambos os métodos coincidem e resultam na mesma variância

$$V(a_j) = V(a_{j,1}), \text{ já que } \sum_{q=1}^p \text{Cov}(a_{j,q}, a_{j,r}) = 0 \text{ para variáveis iid.}$$

Whitt (1982) mostrou que o método assintótico é assintoticamente correto quando $\rho_j \rightarrow 1$ (intensidade de tráfego pesado), enquanto o método dos intervalos estacionários é assintoticamente correto quando o número de processos de chegadas (não necessariamente de Poisson) na estação j tende a infinito, quando a superposição destes processos tende ao processo de Poisson.

a) Superposição das Chegadas

Sevcik et al. (1977), baseando-se no método assintótico, mostraram que ca_j é uma combinação convexa de ca_{ij} (definido como o coeficiente quadrático de variação do intervalo de tempo entre chegadas na estação j a partir da estação i ; vide figura 2.4a):

$$ca_j = \sum_{i=0}^n \frac{\lambda_{ij}}{\lambda_j} ca_{ij} \quad (15)$$

onde

$$\lambda_{ij} = q_{ij} \lambda_i \quad (16)$$

e λ_j é obtido através de (1). Observa-se de (15) que se todos processos de chegada são Poisson ($ca_{ij} = 1$, para $i=0, \dots, n$), então $ca_j = 1$ como era de se esperar, uma vez que a superposição de processos de Poisson independentes resulta em um processo de Poisson (Çinlar, 1975).

Entretanto, nem o método assintótico nem o método dos intervalos estacionários funcionam muito bem sozinhos para aproximar o ca_j para um número variado de casos. Albin (1982) mostrou que uma melhoria considerável podia ser obtida usando uma combinação convexa entre o valor obtido pelo método assintótico e o valor obtido pelo método dos intervalos estacionários, e propôs o método híbrido:

$$ca_j = w_j ca_j^{AS} + (1 - w_j) ca_j^{IE} \quad (17)$$

em que ca_j^{AS} é o *scv* aproximado pelo método assintótico, conforme (15), enquanto o ca_j^{IE} é o *scv* aproximado pelo método dos intervalos estacionários. Whitt (1983a) aproximou o método dos intervalos estacionários a um processo de Poisson ($ca_j^{IE} = 1$), simplificando o refinamento proposto por Albin (1982). Esta simplificação apresentou erro médio absoluto de 4%, enquanto o refinamento de Albin (1982) apresentava erro médio absoluto de 3%. Substituindo (15) em (17), obtém-se a simplificação adotada por Whitt (1983a):

$$ca_j = w_j \sum_{i=0}^n \frac{\lambda_{ij}}{\lambda_j} ca_{ij} + (1 - w_j) \quad (18)$$

$$\text{com } w_j = \frac{1}{1 + 4(1 - \rho_j)^2 (w'_j - 1)} \text{ e } w'_j = \frac{1}{\sum_{i=0}^n \left(\frac{\lambda_{ij}}{\lambda_j} \right)^2}$$

A expressão para w_j foi obtida por Albin (1982) após extensiva simulação.

b) Processo de Partida

Esta etapa determina a taxa média de partida agregada e o parâmetro de variabilidade da partida agregada da estação j , cd_j , combinando os efeitos da chegada superposta (taxa média e ca_j) com o processo de serviço (taxa média de serviço e cs_j), conforme pode ser visualizado pela figura 2.4b.

Uma hipótese para o estudo da rede é a de que esteja em estado estacionário. Deste modo, a taxa de partida agregada pode ser determinada e é igual à taxa média de chegada superposta (λ_j), uma vez que supõe-se que toda estação j da rede esteja operando em condições de equilíbrio ($\rho_j < 1$). Já cd_j pode ser aproximado pelo método dos intervalos estacionários (Buzacott e Shanthikumar, 1993):

$$cd_j = ca_j + 2\rho_j^2 cs_j - 2\rho_j(1 - \rho_j)\mu_j E(Wq_j) \quad (19)$$

sendo $E(Wq_j)$, o tempo médio de espera na estação j . Desde que $E(Wq_j)$ aparece em (19), o congestionamento da estação j afeta a variabilidade do processo de partida. A aproximação para cd_j pelo método dos intervalos estacionários é obtida inserindo uma aproximação para $E(Wq_j)$ em uma fila $GI/G/1$. Esta medida de congestionamento pode ser estimada pela fórmula de Kraemer e Lagenbach-Belz (Tijms, 1986):

$$E(L_j) = \frac{\rho_j^2 (ca_j + cs_j)g(\rho_j, ca_j, cs_j)}{2(1 - \rho_j)} + \frac{\lambda_j}{\mu_j} \quad (20)$$

com

$$g(\rho_j, ca_j, cs_j) = \begin{cases} \exp\left[\frac{-2(1 - \rho_j)(1 - ca_j)^2}{3\rho_j(ca_j + cs_j)}\right], & \text{se } ca_j < 1 \\ 1, & \text{se } ca_j \geq 1 \end{cases}$$

que, após a aplicação da lei de Little e relações subsequentes (6a-6c), resulta em:

$$E(Wq_j) = \frac{\rho_j(ca_j + cs_j)g(\rho_j, ca_j, cs_j)}{2\mu_j(1 - \rho_j)} \quad (21)$$

Fazendo a simplificação $g(\rho_j, ca_j, cs_j)=1$ em (21) e substituindo em (19), obtém-se a aproximação obtida pelo método dos intervalos estacionários (Whitt, 1983a):

$$cd_j = \rho_j^2 cs_j + (1 - \rho_j^2)ca_j \quad (22)$$

Conforme (22), cd_j é uma combinação convexa de cs_j e ca_j e resulta exatamente $cd_j=1$, quando $cs_j=ca_j=1$ (ou seja, quando os intervalos de tempo entre chegadas e de serviço são exponenciais, os intervalos de tempo entre partidas também é exponencial; veja este resultado em Kleinrock, 1975). Além disso, quando $\rho_j \rightarrow 1$, então $cd_j \rightarrow cs_j$. Ou seja, longas filas resultantes de altos níveis de utilização da estação j tendem a diminuir o efeito do scv do intervalo de tempo entre chegadas. Por outro lado,

quando $\rho_j \rightarrow 0$, então $cd_j \rightarrow ca_j$. Ou seja, em níveis de utilização baixos da estação j em que as filas de espera não são esperadas, o scv dos tempos de serviço não surte efeito sobre o cd_j e este tende a ca_j .

O método assintótico resulta em uma aproximação bem mais elementar que a aproximação gerada pelo método dos intervalos estacionários. O número de partidas em um longo intervalo de tempo é apenas o número de chegadas menos o número em fila, mas o número em fila oscila ao redor de sua distribuição no estado estacionário ($\rho_j < 1$), enquanto que o número de chegadas tende a infinito. Daí, o porquê de Whitt (1983a) sugerir que o processo de partida é apenas o processo de chegada, segundo o método assintótico:

$$cd_j = ca_j \quad (23)$$

Esta aproximação também é exata quando os processos de chegada e de serviço são exponenciais, e torna-se mais precisa na estação j quando as utilizações nas subseqüentes estações aumentam.

c) Separação das Partidas

Esta etapa decompõe as taxas médias (λ_j) e os coeficientes quadráticos de variação das partidas agregadas cd_j em λ_{ji} (taxa média de partida da estação j para estação i) e cd_{ji} (coeficiente quadrático de variação de partida da estação j para a estação i). Cada taxa média de partida λ_{ji} é calculada conforme (16).

Com respeito ao parâmetro de variabilidade, nenhuma aproximação seria necessária para separação das partidas (vide figura 2.4c) porque um processo de renovação decomposto por probabilidades independentes (roteiro Markoviano) é ainda um processo de renovação. Entretanto, é indispensável utilizar aproximações nesta etapa porque o processo real sendo decomposto (separado) não é geralmente um processo de renovação, e/ou a separação não é freqüentemente um roteiro Markoviano.

Desde que um processo de renovação separado de acordo com uma rota Markoviana é um processo de renovação, o método assintótico e o método dos intervalos estacionários coincidem, conforme visto anteriormente. Se uma partida agregada com parâmetro cd_j é separada em n partidas, sendo selecionadas

independentemente de acordo com probabilidades q_{ji} , então o i -ésimo processo de partida tem scv igual a (Whitt, 1983a):

$$cd_{ji} = q_{ji}cd_j + (1 - q_{ji}) \quad (24)$$

A expressão (24) é exata quando o processo de partida agregada é Poisson (processo de renovação), isto é $cd_j=1$, resultando em $cd_{ji}=1$. Além disso, quando $q_{ji} \rightarrow 1$ (o que implica em $\lambda_{ji} \rightarrow \lambda_j$ de acordo com (16)), então $cd_{ji} \rightarrow cd_j$. Ainda, quando $q_{ji} \rightarrow 0$ (o que implica em λ_{ji} ser uma proporção muito pequena da partida agregada λ_j), então $cd_{ji} \rightarrow 1$, ou seja, o processo de partida entre as estações i e j é um processo de Poisson.

d) Equações de Variabilidade de Tráfego

O sistema formado, por exemplo, pelas expressões:

$$ca_j = \sum_{i=0}^n \frac{\lambda_{ij}}{\lambda_j} ca_{ij} \quad j=1, \dots, n \quad (15)$$

$$cd_j = \rho_j^2 cs_j + (1 - \rho_j^2) ca_j \quad j=1, \dots, n; \quad (22)$$

$$cd_{ji} = q_{ji}cd_j + (1 - q_{ji}) \quad j=1, \dots, n; \quad i=1, \dots, n \quad (24)$$

é conhecido como equações de variabilidade de tráfego. Este sistema é linear nas variáveis ca_j , cd_j e ca_{ij} (ou cd_{ij} , dado que cd_{ij} , o coeficiente quadrático de variação do intervalo de tempo entre partidas da estação i para a estação j é o mesmo que o coeficiente quadrático de variação do intervalo de tempos entre chegadas na estação j a partir da estação i , ca_{ij}) e fornece boas aproximações para ca_j , $j=1, \dots, n$. Desta maneira, pode-se descrever cada estação j pelos parâmetros $\{\lambda_j, ca_j, \mu_j, cs_j\}$, resolvendo os sistemas lineares de equações de taxa de tráfego (1) e de equações de variabilidade de tráfego, por exemplo, (15), (22) e (24). Pode-se então prosseguir para o Passo 2 do método de decomposição.

2.5.1.2 Passo 2: Decomposição da rede de filas em subsistemas com uma única estação e avaliação das medidas de desempenho para cada estação.

Tendo calculado os parâmetros de taxa e de variabilidade associados a cada processo de chegada interna, pode-se calcular as medidas de congestionamento

para cada nó. Em outras palavras, com os parâmetros $\{\lambda_j, ca_j, \mu_j, cs_j\}$ para $j=1, \dots, n$, calcula-se as medidas de desempenho separadamente para cada nó da rede, como se fossem estocasticamente independentes. Note que a forma com que os nós interagem já foi considerada nos cálculos de λ_j e ca_j no Passo 1.

Whitt (1983a) afirma que procedimentos de estimativas de medidas de desempenho mais detalhados não apresentam grandes ganhos, quando comparados aos que caracterizam cada nó j da rede pelos parâmetros $\{\lambda_j, ca_j, \mu_j, cs_j\}$, ou similarmente, por funções do primeiro e segundo momentos das distribuições dos intervalos de tempo entre chegadas e dos tempos de serviço.

A partir de fórmulas aproximadas da teoria de filas (vide por exemplo Kleinrock, 1975; Tijms, 1986; Whitt, 1983a), pode-se estimar as medidas de desempenho para uma fila $GI/G/1$ com disciplina $FCFS$ tal como em (20) e (21).

Shanthikumar e Buzacott (1981) apresentam aproximações para filas $GI/G/1$ com disciplina de serviço SPT (*shortest processing time*). Experimentos conduzidos em *job-shops* simétricos e *flow-shops* quando as distribuições de tempo de serviço são exponenciais, Erlang-2 ou Erlang-5, mostraram que as aproximações tiveram desempenho satisfatório. Entretanto, quando os tempos de serviço são determinísticos, o intervalo com 95% de confiança obtido pela simulação não continha os resultados obtidos pelas aproximações.

2.5.1.3 Passo 3: Recomposição dos resultados obtidos pelas estações decompostas com avaliação das medidas de desempenho da rede.

Neste passo, avalia-se as medidas de desempenho para toda rede como, por exemplo, a taxa média de produção da rede no equilíbrio (*throughput* ou taxa total de chegada externa à rede) λ_0 :

$$\lambda_0 = \sum_{j=1}^n \lambda_{0j} \quad (25)$$

A média e a variância do número de usuários na rede podem ser estimados, respectivamente, por (7) e (8). É importante salientar que (8) é válido para modelos de redes Markovianas como uma consequência da solução na forma produto, mas é uma aproximação para redes mais gerais. O *leadtime* médio para um

item/produto/usuário qualquer e o tempo médio de espera em fila na estação j podem ser estimados de acordo com (9) e (10), respectivamente.

2.5.2 Modelos de Redes com Filas $GI/G/m$, Classe Única e Rota Probabilística

Diferentemente da seção anterior, esta seção modela cada nó da rede como um sistema de filas $GI/G/m$, com várias máquinas paralelas e idênticas. Entretanto, ainda admite que todos os itens/produtos pertencem a uma única classe de produtos (ou podem ser convenientemente agregados) e movem-se através das estações da rede segundo um roteiro probabilístico.

2.5.2.1 Passo 1: Análise das interações entre estações da OQN .

Considere a mesma notação definida para a rede com filas $GI/G/1$ para os dados de entrada. Cada estação é inicialmente caracterizada pela matriz $\{Q\}$ que descreve a rota, e pelos cinco parâmetros iniciais $\{m_j, \lambda_{0j}, ca_{0j}, \mu_j, cs_j\}$, sendo $m_j, m_j \geq 1$, o número de máquinas idênticas e paralelas na estação j . Novamente, esta etapa consiste em descrever cada estação j pelos parâmetros $\{m_j, \lambda_j, ca_j, \mu_j, cs_j\}$. Calcula-se λ_j por (1), e ca_j por meio de aproximações para os processos de superposição das chegadas, partidas e de decomposição das partidas.

Estendendo (22) para o processo de partidas em filas $GI/G/m$ (Whitt, 1983a):

$$cd_j = 1 + (1 - \rho_j^2)(ca_j - 1) + \frac{\rho_j^2(cs_j - 1)}{m_j^{1/2}} \quad (26)$$

Note que se $m_j=1$, então (26) se reduz a (22). Além disso, sistemas $M/M/m$ ($ca_j=cs_j=1$) e $M/G/\infty$ ($ca_j=1, m_j \rightarrow \infty$) resultam em $cd_j=1$, como era de se esperar. Neste último, como o número de servidores tende a infinito, praticamente não há formação de fila de espera e o processo de partida não sofre efeito do processo de serviço, coincidindo com o processo de chegada que, para este caso, é um processo de Poisson.

De (22) é claro que a variabilidade do processo de partida é uma média ponderada apropriada das variabilidades dos processos de chegada e de serviço. Portanto, quando o processo de serviço é determinístico ($cs_j=0$), o processo de partida é

menos variável que o processo de chegada. Entretanto, as reduções reais de variabilidade em uma rede, causada por processos de serviço determinísticos, não são em geral tão grandes quanto previsto por (26). Whitt (1983a), com vistas a reduzir esta distorção, propôs uma modificação em (26):

$$cd_j = 1 + (1 - \rho_j^2)(ca_j - 1) + \frac{\rho_j^2(\max\{cs_j, 0,2\} - 1)}{m_j^{1/2}} \quad (27)$$

Finalmente, combinando (27), (15) e (24) obtém-se o sistema linear (equações de variabilidade de tráfego) em ca_j para redes de filas abertas $GI/G/m$:

$$ca_j = \alpha_j + \sum_{i=1}^n \beta_{ij} ca_i \quad j=1, \dots, n \quad (28)$$

em que:

$$\alpha_j = 1 + w_j \left\{ p_{0j} ca_{0j} - 1 + \sum_{i=1}^n p_{ij} [(1 - q_{ij}) + q_{ij} \rho_i^2 x_i] \right\} \quad e$$

$$\beta_{ij} = w_j p_{ij} q_{ij} (1 - \rho_i^2)$$

com w_j definido em (18) e

$$p_{ij} = \frac{\lambda_{ij}}{\lambda_j}, \quad q_{ij} = \frac{\lambda_{ij}}{\lambda_i}, \quad x_i = 1 + \frac{\max\{cs_i, 0,2\} - 1}{m_j^{1/2}}$$

2.5.2.2 Passo 2: Decomposição da rede de filas em subsistemas com uma única estação e avaliação das medidas de desempenho para cada estação.

Similarmente à seção anterior, tendo calculado os parâmetros de taxa e de variabilidade para os processos de chegada em cada estação, o Passo 2 consiste em calcular as medidas de desempenho para cada nó da rede, supostamente independentes estocasticamente. Para isso, recorre-se às aproximações derivadas para filas $GI/G/m$. Por exemplo, o número médio de usuários na estação j pode ser aproximado por:

$$E(L_j) = \frac{\lambda_j(ca_j + cs_j)}{2} E(Lq_j)_{M/M/m} + \frac{\lambda_j}{\mu_j} \quad (29)$$

onde $E(Lq_j)_{M/M/m}$ denota o número médio de usuários em uma fila $M/M/m_j$, definido como (Tijms, 1986):

$$E(Lq_j)_{M/M/m} = \frac{\rho_j (\lambda_j / \mu_j)^{m_j} \pi_j(0)}{(1 - \rho_j)^2 m_j!} \quad \text{com } \pi_j(0) \text{ dado em (3).}$$

O tempo médio de espera em fila na estação j pode ser aproximado por (utilizando as relações 6a-6c):

$$E(Wq_j) = \frac{(ca_j + cs_j)}{2} E(Wq_j)_{M/M/m} \quad (30)$$

com $E(Wq_j)_{M/M/m}$ representando o tempo médio de espera em uma fila $M/M/m_j$, definido de acordo com (10). Já o tempo médio de permanência na estação j (tempo de espera em fila mais o tempo em serviço) é dado por (utilizando 6c):

$$E(W_j) = \frac{(ca_j + cs_j)}{2} E(Wq_j)_{M/M/m} + \frac{1}{m_j \mu_j} \quad (31)$$

Outras aproximações para $E(L_j)$ e $E(Wq_j)$ podem ser obtidas em Buzacott e Shanthikumar (1993):

$$E(L_j) = \frac{E(Lq_j)_{M/M/m}}{E(Lq_j)_{M/M/1}} E(Lq_j)_{GI/G/1} + \frac{\lambda_j}{\mu_j}$$

$$E(Wq_j) = \frac{E(Wq_j)_{M/M/m}}{E(Wq_j)_{M/M/1}} E(Wq_j)_{GI/G/1}$$

em que $E(Lq_j)_{GI/G/1}$ e $E(Wq_j)_{GI/G/1}$ correspondem, respectivamente, a (20) sem a carga ofertada, segundo termo do lado direito, e (21). As variâncias de L_j e Wq_j são dadas, respectivamente, por (Whitt (1983a, 1993)):

$$V(L_j) = E(L_j^2) - E(L_j)^2 = E(L_j)^2 c_{L_j} \quad (32)$$

e

$$V(Wq_j) = E(Wq_j)^2 c_{Wq_j} \quad (33)$$

com:

$$c_{L_j} = \frac{\frac{\lambda_j}{\mu_j} (1 + P(L_j \geq m_j)) + [P(L_j \geq m_j) \rho_j + P(L_j \geq m_j) (1 - P(L_j \geq m_j))]}{\left[\frac{\lambda_j}{\mu_j} + P(L_j \geq m_j) \frac{\rho_j}{(1 - \rho_j)} \right]^2} \frac{\rho_j^2}{(1 - \rho_j)^2}$$

e

$$c_{Wq_j} = \frac{2 - P(Wq_j > 0)}{P(Wq_j > 0)}$$

sendo

$$P(L_j \geq m_j) = P(Wq_j > 0) = \frac{\left(\frac{\lambda_j}{\mu_j}\right)^{m_j} \pi_j(0)}{(1 - \rho_j)m_j!} e$$

$\pi_j(0)$ definido em (3).

2.5.2.3 Passo 3: Recomposição dos resultados obtidos pelas estações decompostas com avaliação das medidas de desempenho da rede.

Esta etapa consiste em avaliar a medida de desempenho para toda a rede. Por exemplo, o número médio de usuários/itens/produtos no sistema pode ser calculado de acordo com (7), e o *leadtime* médio para um item/produto arbitrário pode ser calculado através de (9).

2.5.3 Modelos de Redes com Filas *GI/G/m*, Múltiplas Classes e Rota Determinística

Esta seção apresenta os modelos estudados em literatura para redes de filas cujos nós podem ser modelados como filas *GI/G/m*. Entretanto, agora assume-se que a rede processa itens/produtos pertencentes (ou convenientemente agregados) a múltiplas classes que perfazem roteiro determinístico. Em geral, estas redes de filas são mais adequadas a sistemas reais de manufatura tipo *job-shop*, pois processam uma variedade muito alta de produtos com roteiros de fabricação bem definidos (a seqüência de nós visitados são definidos *a priori*).

Como anteriormente, o processo de chegada externa é caracterizado pelos parâmetros de taxa média de chegada e de variabilidade dos intervalos de tempo entre chegadas, ou seja, em função do primeiro e segundo momentos da distribuição de chegadas. Cada classe de produto pode ter sua própria distribuição de serviço em cada nó de sua rota. A distribuição de serviço pode ser diferente, não só para classes diferentes, mas também para visitas diferentes ao mesmo nó da mesma classe. Estas distribuições de tempo de serviço também são caracterizadas pelos parâmetros de taxa média e de variabilidade dos tempos de serviço.

É importante definir a notação utilizada nesta seção para os dados de entrada. Sejam:

n número de estações internas na rede

r	número de classes na rede
m_j	número de máquinas idênticas e paralelas na estação j
n_k	número de operações no roteiro da classe k
λ'_k	taxa média de chegada externa da classe k ($\lambda'_k = 1/E(a'_k)$)
ca'_k	coeficiente quadrático de variação do intervalo de tempo entre chegadas externas da classe k ($ca'_k = V(a'_k)/(E(a'_k))^2$).
n_{kl}	estação visitada para a operação l do roteiro da classe k ;
$E(s_{kl})$	média do tempo de serviço da operação l do roteiro da classe k , s_{kl} .
μ_{kl}	taxa média de serviço da operação l do roteiro da classe k ($\mu_{kl} = 1/E(s_{kl})$)
cs_{kl}	coeficiente quadrático de variação do tempo de serviço da operação l do roteiro da classe k ($cs_{kl} = V(s_{kl})/(E(s_{kl}))^2$)

É importante notar que, neste caso, o roteiro para cada classe k é determinado por n_k e n_{kl} . Por exemplo, a sequência (1,3,2,1,4) na qual $n_k=5$ define um roteiro em que $n_{k1}=1$, $n_{k2}=3$, $n_{k3}=2$, $n_{k4}=1$ e $n_{k5}=4$ é a sequência de estações visitadas pelos itens/produtos da classe k .

A seção anterior, que trata o modelo de rede de filas $GI/G/m$ com única classe de produto e roteiro probabilístico, parte dos parâmetros iniciais $\{m_j, \lambda_{0j}, ca_{0j}, \mu_j, cs_j\}$. Como para este caso os dados de entrada estão em função das classes dos produtos e do roteiro, é preciso utilizar um procedimento que permita que, partindo dos dados de entrada, se obtenha os parâmetros iniciais $\{m_j, \lambda_{0j}, ca_{0j}, \mu_j, cs_j\}$. Os dados de entrada, em função das classes, são convertidos nos parâmetros iniciais, em função das estações j ($j=1, \dots, n$), de acordo com o procedimento de agregação das múltiplas classes em uma única classe agregada, descrito por Whitt (1983a). Feito isso, pode-se empregar o método de decomposição tal como empregado nas seções anteriores, e aplicar o Passo 1 do método para calcular o valor exato de λ_j e o aproximado para ca_j . Estes valores são utilizados no Passo 2 para avaliar a medida de desempenho de cada estação da rede, supostas estocasticamente independentes. Posteriormente, no Passo 3, as medidas de desempenho da rede para esta classe agregada são avaliadas (se necessárias) e, em

seguida, retorna-se à rede original para estimar as medidas de desempenho para cada classe na rede.

Whitt (1983a) apresentou um procedimento para agregar todas as classes em uma única e utilizar o modelo de classe única. Desta maneira, a *OQN* original, com múltiplas classes, é reduzida a uma *OQN* com uma única classe agregada.

Quando se trata da agregação de itens/produtos em uma única classe com roteiro determinístico, a taxa média de chegadas externas de todas as classes, cuja primeira operação ocorre na estação j , é dada por:

$$\lambda_{0j} = \sum_{k=1}^r \lambda'_k 1\{n_{k1} = j\} \quad (34)$$

sendo $1\{\cdot\}$ uma função indicadora resultando em 1 se a expressão $\{\cdot\}$ é verdadeira e 0, caso contrário.

As taxas λ_j e λ_{ij} (definidas respectivamente por taxa média de chegada agregada na estação j e taxa média de chegada na estação j a partir da estação i , com $i, j = 0, \dots, n$ e $i \neq j$) são computadas por:

$$\lambda_j = \sum_{k=1}^r \sum_{l=1}^{n_k} \lambda'_k 1\{n_{kl} = j\} \quad (35)$$

$$\lambda_{ij} = \sum_{k=1}^r \sum_{l=1}^{n_k-1} \lambda'_k 1\{n_{kl} = i, n_{k,l+1} = j\} \quad (36)$$

Deste modo, a matriz \mathbf{Q} (conforme definida anteriormente), cujos elementos q_{ij} ($i=1, \dots, n; j=1, \dots, n$) representam a proporção de produtos da classe agregada que chegam na estação i e partem para a estação j , pode ser obtida pela expressão:

$$q_{ij} = \frac{\lambda_{ij}}{\lambda_i} \quad (37)$$

O tempo de serviço na estação j , s_j , é igual a s_{kl} (o tempo de serviço da l -ésima operação do item/produto da classe k) se o item/produto pertence à classe k e está realizando a operação l , cuja probabilidade é $\frac{\lambda'_k 1\{n_{kl} = j\}}{\lambda_j}$, ou seja,

$$P[s_j = s_{kl}] = \frac{\lambda'_k 1\{n_{kl} = j\}}{\lambda_j} . \text{ Portanto, o tempo médio de serviço da classe agregada na}$$

estação j , $E(s_j)$, é obtido através da média ponderada:

$$E(s_j) = \frac{\sum_{k=1}^r \sum_{l=1}^{n_k} E(s_{kl}) \lambda'_k 1\{n_{kl} = j\}}{\lambda_j} \quad (38)$$

com λ_j definida em (35).

O coeficiente quadrático de variação dos tempos de serviço agregado na estação j , cs_j , é estimado usando (38) e a propriedade de que o segundo momento da soma de distribuições independentes é igual à soma dos respectivos segundos momentos:

$$cs_j = \frac{V(s_j)}{(E(s_j))^2} = \frac{E(s_j^2)}{(E(s_j))^2} - 1 = \frac{\sum_{k=1}^r \sum_{l=1}^{n_k} E(s_{kl}^2) \lambda'_k 1\{n_{kl} = j\}}{\lambda_j (E(s_j))^2} - 1$$

Aplicando (35), a definição de variância $V(s_{kl}) = E(s_{kl}^2) - (E(s_{kl}))^2$ e a definição de coeficiente quadrático de variação para s_{kl} ($cs_{kl} = V(s_{kl}) / (E(s_{kl}))^2$), chega-se a:

$$cs_j = \frac{\sum_{k=1}^r \sum_{l=1}^{n_k} (E(s_{kl}))^2 (cs_{kl} + 1) \lambda'_k 1\{n_{kl} = j\}}{\sum_{k=1}^r \sum_{l=1}^{n_k} \lambda'_k 1\{n_{kl} = j\} \cdot (E(s_j))^2} - 1 \quad (39)$$

Para obtenção do coeficiente quadrático de variação do intervalo de tempo entre chegadas externas da classe agregada na estação j , ca_{0j} , pode-se utilizar o método híbrido, refinado por Whitt (1983b) e reportado por Bitran e Morabito (1996), para superpor todos parâmetros de variabilidade dos intervalos de tempo entre chegadas externas à estação j :

$$ca_{0j} = w_j \sum_{k=1}^r \frac{\lambda'_k 1\{n_{k1} = j\} ca'_k}{\sum_{l=1}^{n_k} \lambda'_l 1\{n_{l1} = j\}} + 1 - w_j \quad (40)$$

$$\text{onde: } w_j = \frac{1}{1 + 4(1 - \rho_j)^2 (w'_j - 1)} \quad \text{e} \quad w'_j = \frac{1}{\sum_{k=1}^r \left(\frac{\lambda'_k 1\{n_{k1} = j\}}{\sum_{l=1}^{n_k} \lambda'_l 1\{n_{l1} = j\}} \right)^2}$$

As expressões (34) a (40) produzem os parâmetros iniciais $\{m_j, \lambda_{0j}, ca_{0j}, \mu_j, cs_j\}$ da classe agregada em cada estação j e a matriz \mathbf{Q} para a OQN de classe única, com filas $GI/G/m$. Em seguida, aplica-se o método de decomposição.

2.5.3.1 Passo 1: Análise das interações entre estações da OQN

Semelhantemente aos modelos de redes de filas $GI/G/m$ com classe única, tratados na seção anterior 2.5.2, esta etapa consiste em descrever cada nó da rede pelos parâmetros $\{m_j, \lambda_j, ca_j, \mu_j, cs_j\}$ a partir dos parâmetros iniciais $\{m_j, \lambda_{0j}, ca_{0j}, \mu_j, cs_j\}$. O cálculo de λ_j é dado por (1) ou (35) e ca_j estimado, por exemplo, de acordo com o sistema linear (28).

2.5.3.2 Passo 2: Decomposição da rede de filas em subsistemas com uma única estação e avaliação das medidas de desempenho para cada estação

Uma vez que os parâmetros de taxa e de variabilidade associados aos processos de chegada em cada estação j ($j=1, \dots, n$) foram calculados no Passo 1, esta etapa avalia as medidas de desempenho para cada estação j da rede, utilizando fórmulas de medidas de desempenho derivadas da teoria de filas $GI/G/m$. É possível, por exemplo, estimar o número médio de itens/produtos na estação j por meio de (29), o tempo médio de espera em fila na estação j por meio de (30), e o tempo médio de permanência na estação j com (31).

2.5.3.3 Passo 3: Recomposição dos resultados obtidos pelas estações decompostas com avaliação das medidas de desempenho da rede

Esta etapa avalia as medidas de desempenho para a classe agregada na rede. Por exemplo, o número médio de usuários/itens/produtos na rede é dado de acordo com (7) e o *leadtime* médio para um usuário arbitrário pode ser calculado com (9).

Ao retornar à rede original, pode-se obter medidas para cada classe k individualmente a partir da classe agregada. Por exemplo, o *leadtime* médio $E(T_k)$, tempo de espera em fila mais o tempo em serviço, para um produto arbitrário da classe k na rede, é dado por:

$$E(T_k) = \sum_{l=1}^{n_k} E(Wq_{n_{kl}}) + \sum_{l=1}^{n_k} E(s_{kl}) \quad (41)$$

com $E(Wq_{n_{kl}})$, o tempo de espera em fila na estação n_{kl} é aproximado por (30) no qual $n_{kl}=j$ e $E(s_{kl})$ é um dado de entrada. Observe que o primeiro termo de (41) corresponde ao tempo total de espera em fila consumido por um item/produto da classe k em todas as estações pertencentes ao seu roteiro. O segundo termo corresponde ao tempo total de

processamento gasto por um item/produto da classe k em serviço para realizar todas as l operações do seu roteiro.

Similarmente, pode-se aproximar a variância do *leadtime*, $V(T_k)$, admitindo que os processos de chegada e de serviço são *iid* e que as estações são estocasticamente independentes:

$$V(T_k) = \sum_{l=1}^{n_k} V(Wq_{n_{kl}}) + \sum_{l=1}^{n_k} V(s_{kl}) = \sum_{l=1}^{n_k} V(Wq_{n_{kl}}) + \sum_{l=1}^{n_k} (E(s_{kl}))^2 cs_{kl} \quad (42)$$

com $V(Wq_{n_{kl}})$ estimado de acordo com (33) em que $n_{kl}=j$.

2.5.4 Interferência entre Classes e Aproximação de Tráfego Leve

Baseada nos artigos de Bitran e Tirupati (1988), Bitran e Dasu (1992), Whitt (1988, 1994), Bitran e Morabito (1996) e Morabito (1998), esta seção aborda o problema de interferência entre classes para gerar aproximações mais efetivas para o processo de separação de partidas quando trata-se de *OQN's* que processam múltiplas classes de produtos e roteiros determinísticos, cenário típico para sistemas de manufatura *job shop*.

Bitran e Tirupati (1988) observaram que a aproximação para o processo de separação das partidas (24) é pouco efetiva quando se trata de *OQN's* com roteiros determinísticos e múltiplas classes de produtos. Nestas situações, (24) ignora a interação entre as classes. Para capturar este efeito, o *scv* das partidas de uma estação processando múltiplas classes é caracterizado para cada classe. Neste processo, a distorção do *scv* de um dado produto devido a presença de outros produtos é identificada. Esta distorção é referenciada como interferência entre classes.

Para cada classe de produto em uma estação, o método apresentado por Bitran e Tirupati (1988) consiste na redução em duas classes: a classe de interesse e o agregado de todas as outras classes. Deseja-se caracterizar o *scv* das partidas para cada classe. O resultado é análogo e generaliza o processo de separação das partidas dado por (24). Assim, as interações entre as estações na rede (Passo 1 do método de decomposição) podem ser, por exemplo, descritas pelas aproximações (18) para superposição das chegadas, pela aproximação (22) para o processo de partidas e pela aproximação desenvolvida a seguir para a separação das partidas.

As principais hipóteses admitidas para a aproximação que é apresentada adiante são (Bitran e Tirupati, 1988):

- estação com único servidor com disciplina de fila *FCFS* (juntas, garantem que a seqüência de chegada de itens/produtos, por classe de produto, é idêntica a de partidas);
- intervalos de tempo entre partidas da estação são *iid*;
- as chegadas de produtos são independentes e os tempos entre chegadas para cada produto são *iid*.

Para cada classe k e operação l em uma estação, sejam (k, l) a classe de interesse e (k', l') a agregação de todas as outras classes, $(k', l') \neq (k, l)$, que chegam (ou saem) entre duas chegadas (partidas) sucessivas da classe de interesse (k, l) numa certa estação n_{kl} de uma *OQN* com múltiplas classes e roteiros determinísticos. Uma vez que só há roteiros determinísticos na rede, pode-se facilmente obter $\lambda_{n_{kl}}$ (a taxa de chegada total na estação n_{kl}) por:

$$\lambda_{n_{kl}} = \sum_{k=1}^r \sum_{l=1}^{n_k} \lambda'_k 1\{n_{kl} = n_{kl}\} \quad (43)$$

A proporção da classe de interesse (k, l) na estação n_{kl} é dada por:

$$q_{kl} = \lambda'_k / \lambda_{n_{kl}} \quad (44)$$

em que λ'_k é a taxa média de chegada externa da classe k ($\lambda'_k = 1/E(a'_k)$)

Sejam d_{kl} o intervalo de tempo entre partidas da classe de interesse (k, l) na estação n_{kl} , $d_{n_{kl}}$ o intervalo de tempo entre partidas de todas as classes da estação n_{kl} e z_{kl} o número de itens/produtos da classe agregada, (k', l') , que chega (ou que sai) na estação n_{kl} durante um intervalo de tempo entre chegadas (entre partidas) da classe de interesse (k, l) . A figura 2.5 ilustra as chegadas (partidas) de itens/produtos da classe agregada entre duas chegadas (partidas) sucessivas da classe de interesse. Definindo $z'_{kl} = z_{kl} + 1$, a variável aleatória d_{kl} é a soma de z'_{kl} variáveis aleatórias, ou seja $d_{kl} = \sum_{i=1}^{z'_{kl}} d_{n_{kl}}$, (vide figura 2.5). A variável aleatória z'_{kl} corresponde ao número de intervalos de tempo entre chegadas (ou de partidas) entre duas chegadas (partidas) sucessivas da classe de interesse (k, l) .

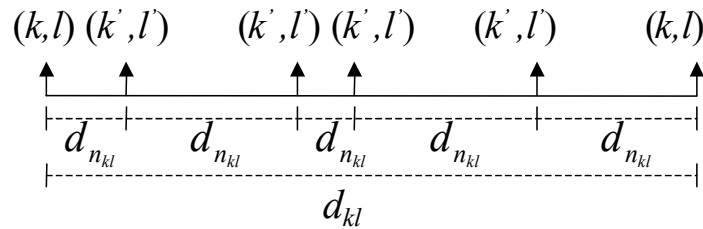


FIGURA 2.5 – Ilustração do processo de chegadas (partidas) da classe agregada entre duas chegadas (partidas) sucessivas da classe de interesse na estação n_{kl} .

A média e a variância da soma de z'_{kl} variáveis aleatórias são (vide Kleinrock, 1975), respectivamente:

$$E(d_{kl}) = E(z'_{kl})E(d_{n_{kl}}) \quad (45)$$

$$V(d_{kl}) = E(z'_{kl})V(d_{n_{kl}}) + V(z'_{kl})(E(d_{n_{kl}}))^2 \quad (46)$$

com $E(d_{kl}) = 1/\lambda'_k$ e $E(d_{n_{kl}}) = 1/\lambda_{n_{kl}}$. Daí, segue de (45) que:

$$E(z'_{kl}) = \lambda_{n_{kl}} / \lambda'_k \quad (47)$$

Dividindo ambos os lados de (46) por $(E(d_{n_{kl}}))^2$ e $(E(z'_{kl}))^2$ e utilizando (45), (47) e (44), obtém-se (Bitran e Tirupati, 1988):

$$cd_{kl} = q_{kl}cd_{n_{kl}} + cz'_{kl} \quad (48)$$

O coeficiente quadrático de variação das partidas de cada classe de produto é caracterizado por (48). É importante notar que (48) expressa o cd_{kl} da classe de produto k que realiza a l -ésima operação (aqui chamada de classe de interesse (k,l)) como a soma de dois termos. O primeiro reflete o efeito do processo de fila pois considera o coeficiente quadrático de variação do processo de partida da estação n_{kl} . O segundo termo não depende do processo de serviço e captura o efeito da distribuição das chegadas da classe agregada entre duas chegadas da classe de interesse (k,l) , o que contrasta com (24) que não levava em consideração este efeito. Dada a dificuldade de se avaliar cz'_{kl} , algumas aproximações baseadas em hipóteses de distribuição de chegadas da classe agregada são propostas.

Motivado pela noção de que a superposição de um grande número de processos de renovação independentes pode ser aproximado por um processo de

Poisson, supõe-se que z_{kl} (o número de chegadas/partidas de produtos da classe agregada entre duas chegadas/partidas sucessivas da classe de interesse na estação n_{kl}) segue a distribuição de Poisson com taxa $(1 - q_{kl})\lambda_{n_{kl}}$. Sob esta hipótese, pode ser mostrado que (Morabito, 1998; Bitran e Tirupati, 1988):

$$cz'_{kl} = (1 - q_{kl})[q_{kl} + (1 - q_{kl})ca_{kl}]$$

que, substituindo em (48), resulta em:

$$cd_{kl} = q_{kl}cd_{n_{kl}} + (1 - q_{kl})q_{kl} + (1 - q_{kl})^2 ca_{kl} \quad (49)$$

em que q_{kl} é dado por (44) e ca_{kl} é o scv do intervalo de tempo entre chegadas da classe de interesse (k, l) na estação n_{kl} . Observe que $ca_{kl} = cd_{k, l-1}$. A expressão (18) pode ser reescrita como (Bitran e Morabito, 1996):

$$ca_j = w_j \sum_{k=1}^r \sum_{l=1}^{n_k} \frac{\lambda'_k}{\lambda_j} ca_{kl} 1\{n_{kl} = j\} + (1 - w_j) \quad (50)$$

$$\text{com } w_j = \frac{1}{1 + 4(1 - \rho_j)^2 (w'_j - 1)} \text{ e } w'_j = \frac{1}{\sum_{k=1}^r \sum_{l=1}^{n_k} \left(\frac{\lambda'_k}{\lambda_j}\right)^2 1\{n_{kl} = j\}}$$

com $j = n_{kl}$ e λ_j obtido de acordo com (43). Combinando (50), (27) e (49) tem-se um sistema linear alternativo em função de ca_j , cd_j e cd_{kl} (ou $ca_{k, l+1}$) para determinar ca_j (Passo 1). Caracterizando cada estação j da rede com os parâmetros $\{m_j, \lambda_j, ca_j, \mu_j, cs_j\}$, procede-se aos Passos 2 e 3, conforme anteriormente.

Bitran e Tirupati (1988), além de aproximar z_{kl} por um processo de Poisson, também assumiram distribuições Erlang para o intervalo de tempo entre chegadas das classes de interesse e da classe agregada. Utilizando dados reais de uma fábrica de semicondutores, concluíram que essas aproximações tem desvios de 5-10% dos valores obtidos em experimentos de simulação. As aproximações que não consideravam o efeito da interferência entre classes resultaram em desvios acima de 20% na maioria dos casos.

De acordo com Morabito (1998), a aproximação (49) pode ser vista como uma generalização de (24). Basta considerar uma situação particular na qual itens/produtos da classe de interesse k entram na rede pela estação j , esperam em fila junto com itens/produtos de outras classes e, após receberem serviço, somente os

itens/produtos da classe k prosseguem para uma certa estação i . Assim, a taxa média de partida do arco (j,i) é $\lambda_{ij} = \lambda'_k$, e a proporção de itens/produtos que fluem da estação j para a estação i é $q_{ji} = \lambda'_k / \lambda_j$. Definindo d_{ji} como o intervalo de tempo entre partidas da estação j para a estação i ; z_{ji} como o número de itens/produtos da classe agregada que chega (ou que sai) na estação j durante um intervalo de tempo entre chegadas (entre partidas) da classe de interesse e ; z'_{ji} como o número de intervalos de tempo entre chegadas (ou de partidas) entre duas chegadas (partidas) sucessivas da classe de interesse, pode-se reescrever (48) da seguinte maneira (Bitran e Morabito, 1996):

$$cd_{ji} = q_{ji}cd_j + cz'_{ji} \quad (51)$$

Admitindo que z_{ji} tenha distribuição de Poisson com taxa $(1-q_{ji})\lambda_j$, (49) pode ser reescrita como:

$$cd_{ji} = q_{ji}cd_j + (1-q_{ji})q_{ji} + (1-q_{ji})^2 ca'_k \quad (52)$$

Observe que se o processo de chegadas da classe k for Poisson ($ca'_k = 1$), então (52) é reduzida a (24). Daí (52) ser considerada como uma generalização de (24). Entretanto, (24) difere de (52): quando q_{ji} tende a zero, (24) sugere que $cd_{ji} \rightarrow 0$ enquanto (52) indica que $cd_{ji} \rightarrow ca'_k$. Este resultado é exato no limite (Bitran e Tirupati, 1988) e possibilitou que Whitt (1988) enunciasse os princípios de aproximação por tráfego leve. Aqui, tráfego leve não significa que a intensidade de tráfego global da estação é baixa, mas que a contribuição da classe para a intensidade de tráfego global da estação é negligenciável, ou seja, $q_{ji} \rightarrow 0$.

“Princípio de Aproximação de Tráfego Leve para Processos de Partida de Classe Única de Filas com Múltiplas Classes. Se a taxa de chegada de certa classe em uma visita a uma determinada estação é uma pequena proporção da taxa total de chegada àquela estação, então o processo de partida para aquela classe correspondente àquela visita àquela estação tende a ser aproximadamente o mesmo que o processo de chegada para aquela classe para aquela visita àquela estação”.

“Princípio de Aproximação de Tráfego Leve para Redes de Filas com Múltiplas Classes e Roteiros Determinísticos. Considere uma classe que segue um roteiro determinístico na rede. Se a contribuição para a taxa de chegada para esta classe em cada visita a cada estação da rede é uma pequena proporção da taxa de chegada em cada estação, então o processo de partida desta classe correspondente à visita em cada

estação, e assim para toda rede, é aproximadamente o mesmo que o processo de chegada externa desta classe à rede”.

Os dois princípios de aproximação por tráfego leve acima foram enunciados sem cuidadosa definição das conclusões, ou seja, sem explicar o que significa, por exemplo, o processo de partida ser aproximadamente o mesmo que o processo de chegada. Whitt (1988) reporta que esta conclusão implica em afirmar que a distribuição de todo o processo de partida é aproximadamente a mesma que a distribuição de todo o processo de chegada. Como consequência, isto significa que a distribuição estacionária dos intervalos de tempos entre partidas é aproximadamente a mesma que a distribuição estacionária dos intervalos de tempo entre chegadas.

Além disso, os dois princípios acima foram enunciados sem condições especificando o número de servidores, a capacidade da fila de espera, as disciplinas de serviço, as distribuições dos tempos de serviço ou mesmo o processo de chegada. “Algumas condições são necessárias, mas de fato as condições são muito gerais de modo que as afirmações presentes nos princípios acima capturam o espírito da verdade sem estas condições extras”, conclui Whitt (1988). Entretanto, há uma condição importante digna de menção. A taxa de chegada ($\lambda_{ij} = \lambda'_k$) e a intensidade de tráfego para a classe de interesse ($q_{ij}=q_{kl}$) devem ser pequenas, de modo que $cd_{kl} \rightarrow ca'_k$ em (52) ou $cd_{kl} \rightarrow ca_{kl}$ em (49).

Segal e Whitt (1989), baseados nas aproximações acima, propuseram uma expressão alternativa para o processo de separação das partidas para *OQN*'s de múltiplas classes e roteiros determinísticos. Seja ce_j a média dos parâmetros de variabilidade do intervalo de tempo entre chegadas externas na estação j , ponderada pelo número de visitas de cada classe na estação j :

$$ce_j = \frac{\sum_{k=1}^r \sum_{l=1}^{n_k} \lambda'_k 1\{n_{kl} = j\} ca'_k}{\sum_{k=1}^r \sum_{l=1}^{n_k} \lambda'_k 1\{n_{kl} = j\}} \quad (53)$$

Utilizando (53), o coeficiente quadrático de variação das partidas entre as estações j e i é redefinido por:

$$cd_{ji} = q_{ji} cd_j + (1 - q_{ji}) q_{ji} ca_j + (1 - q_{ji})^2 ce_j \quad (54)$$

Observe que para $ce_j=ca_j=1$, (54) se reduz a (24). Segal e Whitt (1989) sugeriram a substituição de (24) por (54) se todas as classes tiverem roteiros puramente determinísticos. Sugeriram ainda utilizar uma combinação convexa de (24) e (54) para capturar o efeito de roteiros probabilísticos no processo de separação.

Com mais esta alternativa para o processo de separação, ca_j pode ser determinado no Passo 1 do método de decomposição, resolvendo por exemplo o sistema linear composto pelas expressões (18), (27) e (54).

Whitt (1994) fornece uma aproximação para o coeficiente quadrático de variação da separação das partidas alternativa ao procedimento numérico apresentado por Bitran e Tirupati (1988), quando aproximaram por distribuições Erlang os intervalos de tempo entre chegadas da classe de interesse e da classe agregada (agregação de todas as outras classes). Whitt (1994) propôs para cz'_{kl} em (48) a aproximação:

$$cz'_{kl} = (1 - q_{kl})(q_{kl}ca_{k'l'} + (1 - q_{kl})ca_{kl}) \quad (55)$$

em que $ca_{k'l'}$ é o *scv* aproximado do intervalo de tempo entre chegadas para a classe agregada entre duas chegadas sucessivas da classe de interesse (k,l) na estação n_{kl} , e ca_{kl} é o coeficiente quadrático de variação aproximado do intervalo de tempo entre chegadas da classe de interesse (k,l). Deste modo, (48) fica:

$$cd_{kl} = q_{kl}cd_{n_{kl}} + (1 - q_{kl})(q_{kl}ca_{k'l'} + (1 - q_{kl})ca_{kl}) \quad (56)$$

Observe que se $ca_{k'l'}=1$, então (56) se reduz a (49). A expressão (56) também é uma alternativa mais simples que o procedimento numérico apresentado por Bitran e Tirupati (1988) quando $ca_{k'l'} \leq 1$ e $ca_{kl} \leq 1$ (caso dos processos de chegada Erlang). Além disso, também se aplica aos casos em que $ca_{k'l'} > 1$, $ca_{kl} > 1$, não tratados por Bitran e Tirupati (1988). Whitt (1994) apresentou resultados computacionais sugerindo que (56) é mais efetiva que (49) no processo de separação.

O método de decomposição utilizado para gerar aproximações que permitem avaliar o desempenho de redes de filas é bastante flexível, permitindo que seja empregado para analisar o desempenho de redes que representam sistemas de manufatura reais. Este aspecto inspirou o desenvolvimento de vários pacotes de *software* que utilizam este método (como podem ser vistos em Bitran e Morabito, 1996), tal como o *Queueing Network Analyzer (QNA)*, *ManuPlan*, *MPX*, e *Queueing Network Analysis Package (QNAP)*. Por este motivo é o método adotado neste trabalho.

Com vistas a facilitar a compreensão do leitor, a seguir resumimos as principais aproximações aqui reportadas que descrevem a variabilidade dos processos de superposição de chegada, partida e de separação das partidas e as principais hipóteses são relatadas.

i) superposição de chegadas

A aproximação (15) é baseada no método assintótico e, portanto, considera a dependência dos intervalos de tempo entre chegadas da classe agregada na estação j (Sevcik et al., 1977) para a estimativa de ca_j .

A aproximação (18) consiste na combinação convexa dos valores obtidos pelo método assintótico (considera a dependência dos intervalos de tempo entre chegadas da classe agregada na estação j) e o método dos intervalos estacionários (ignora qualquer dependência no processo de chegada admitindo que o processo de chegadas da classe agregada na estação é exponencialmente distribuído), conforme Albin (1982) e Whitt (1983a).

A aproximação (50) é uma forma de reescrever (18) para redes processando múltiplas classes (Bitran e Morabito, 1996).

ii) processo de partidas

Whitt (1983a) propôs a aproximação para o processo de partidas (27). Nota-se que o processo de partida é uma média ponderada apropriada dos processos de chegada e de serviço.

iii) separação das partidas

A aproximação (24) admite que a partida agregada é separada independentemente de acordo com probabilidade q_{ji} (Whitt, 1983a). Entretanto, a aproximação (24) não é efetiva para redes que processam múltiplas classes com roteiros determinísticos em que a contribuição de cada classe na ocupação total da estação é pequena (ou seja, $q_{kl} \rightarrow 0$). Para estas condições (tráfego leve), a aproximação (49) é recomendada por Bitran e Tirupati (1988).

A aproximação (49) supõe ainda que o processo de chegadas (partidas) da classe agregada pode ser aproximado a um processo de Poisson. Segal e Whitt (1989) propuseram a expressão alternativa (54) para o processo de separação de partidas baseados nas aproximações de tráfego leve (49).

A seção seguinte aborda alguns conceitos e definições básicas sobre modelos de simulação.

2.6 Modelos de Simulação

Simular é imitar as operações de vários tipos de instalações ou processos (chamados de sistemas) do mundo real com o auxílio de um computador. As relações lógicas e matemáticas que descrevem as hipóteses estabelecidas na tentativa de descrever o mundo real constituem um modelo que é uma forma de tentar conhecer o comportamento do sistema. Se as relações que compõem o modelo são simples, métodos matemáticos analíticos são utilizados para obter informações exatas do comportamento do sistema. Se as relações são complexas, como acontece com modelos mais próximos da realidade, a técnica de simulação é utilizada para avaliar um modelo numericamente e os dados são coletados para estimar o comportamento do sistema (Law e Kelton, 1991).

Neste trabalho é utilizada a simulação discreta baseada em eventos, e os modelos por ela tratados são dinâmicos (representa a evolução do sistema no tempo), estocásticos (possuem alguns componentes de entrada aleatórios) e discretos (o sistema muda em um número discreto de pontos no tempo, ou seja, muda quando ocorre um evento).

Law e Kelton (1991) destacam os passos principais para conduzir um estudo de simulação:

- i) Formular o problema e planejar o estudo: definição clara dos objetivos globais e específicos a serem alcançados
- ii) Coletar dados e definir o modelo: coleta de informações e dados do sistema de interesse devem ser usados para especificar procedimentos operacionais e distribuições de probabilidade para as variáveis aleatórias usadas no modelo. O modelo deve capturar a essência do sistema para os propósitos do estudo e, em princípio, deve inicialmente ser moderadamente detalhado, podendo ser mais sofisticado posteriormente, se necessário.
- iii) Validar o modelo: nesta etapa é imperativo a quem modela envolver pessoas familiarizadas com o sistema e interagir com quem toma as decisões. Isto aumenta a

validade e credibilidade do modelo. Além disso, a adequação das distribuições de probabilidade para geração de variáveis de entrada aleatórias devem ser testadas por meio de testes de aderência (*goodness of fit*).

- iv) Construir e verificar um programa computacional: nesta etapa quem modela deve escolher se usa linguagens de programação de propósito geral (*Pascal*, *C* ou *Fortran*), ou linguagens especialmente desenvolvidas para simulação tais como *GPSS*, *SIMAN* ou *SIMSCRIPT*. Neste trabalho é utilizado o *software* de simulação *GPSS/H*[®] (para mais detalhes sobre o *software*, vide Schriber, 1991).
- v) Realizar “corridas” piloto: “corridas” piloto do modelo verificado podem ser usadas para testar a sensibilidade da saída do modelo a pequenas mudanças em um parâmetro de entrada.
- vi) Validação final do modelo: se pequenas variações dos parâmetros de entrada acarretam consideráveis mudanças na saída do modelo, então melhores estimativas dos parâmetros de entrada devem ser obtidas.
- vii) Planejar experimentos: nesta etapa deve-se decidir quais sistemas simular quando há mais de uma alternativa, o que em geral acontece na prática. Para cada alternativa, quem modela deve definir as condições iniciais, o comprimento do período para que o estado transiente seja desconsiderado (*warm-up period*), o comprimento das corridas de simulação e o número de simulações independentes (réplicas).
- viii) Realizar corridas: esta etapa visa fornecer dados de desempenho dos projetos de sistemas de interesse.
- ix) Analisar dados de saída: utilizar técnicas estatísticas para analisar os dados de saída das “corridas” da etapa 8. Geralmente, o objetivo é construir um intervalo de confiança para uma determinada medida de desempenho, ou decidir qual sistema simulado é melhor com relação a uma medida de desempenho.
- x) Documentar, apresentar e implementar resultados.

Obviamente, estas etapas não necessariamente devem seguir esta seqüência e podem existir outras etapas ou inexistir algumas das acima citadas, dependendo do estudo de simulação. Por se tratar de um tema extenso, somente os tópicos imprescindíveis ao entendimento do trabalho são aqui abordados. Leitores mais

interessados em simulação podem consultar Law e Kelton (1991), Schriber (1991) e Banks et al (1996).

Para conduzir um estudo de simulação usando variáveis aleatórias de entrada, tais como intervalos de tempo entre chegadas e tempos de serviço, é necessário especificar as respectivas distribuições de probabilidade. Porque, dado que as variáveis aleatórias de entrada do modelo de simulação seguem distribuições particulares, a simulação avança no tempo gerando valores aleatórios destas distribuições. Quando é possível coletar dados das variáveis aleatórias de interesse, estes dados podem ser utilizados em uma das seguintes abordagens, em ordem crescente de preferência, para especificar uma distribuição (Law e Kelton, 1991):

- a) Os próprios valores dos dados são utilizados diretamente na simulação. Se, por exemplo, os dados representam tempos de serviço, então um dos valores é utilizado quando um tempo de serviço é necessário na simulação.
- b) Os valores dos dados são utilizados para definir uma função de distribuição empírica. Se os dados representam tempos de serviço, então um valor é amostrado desta distribuição quando o tempo de serviço é necessário na simulação.
- c) Técnicas padrões de inferência estatística são usadas para ajustar uma forma de distribuição “teórica”, por exemplo a distribuição de Poisson, e executar o teste de aderência (*goodness of fit*). Se uma distribuição teórica particular com certos valores de seus parâmetros é um bom modelo para os dados de tempo de serviço, um ponto amostral desta distribuição deve ser sorteado quando um tempo de serviço for necessário na simulação.

Dois inconvenientes da abordagem (a) são que a simulação só pode reproduzir o que aconteceu historicamente e que raramente há dados suficientes para realizar todas as corridas de simulação desejadas. A abordagem (b) evita esta falha desde que, pelo menos para dados contínuos, qualquer valor entre os pontos mínimo e máximo observados possa ser gerado. Deste modo, a abordagem (b) é preferível à abordagem (a). A abordagem (a) é recomendada para validação do modelo ao comparar a saída do modelo para um sistema existente com a correspondente saída para um sistema real. Se uma distribuição teórica que se ajuste razoavelmente bem aos dados observados puder ser determinada, abordagem (c), então ela é preferível à abordagem (b), (Law e Kelton, 1991).

Os testes de aderência utilizados neste trabalho para avaliar formalmente, com níveis de significância pré-determinados, se as observações são uma amostra independente de uma função de distribuição particular são: os testes de Kolmogorov-Smirnov e Anderson-Darling. Ambos os testes estão presentes no *software BestFit[®] 2.0*, utilizado neste trabalho. O Teste de Qui-quadrado não foi empregado devido à dificuldade de usá-lo no caso contínuo, uma vez que é necessário definir o número e o tamanho dos intervalos (classes).

Na possibilidade de não se encontrar uma distribuição “teórica” que se ajuste adequadamente aos dados, Law e Kelton (1991) sugerem a seguinte distribuição empírica para variáveis aleatórias contínuas, quando se tem os valores reais das observações individuais X_1, X_2, \dots, X_n . Se o dados originais estiverem disponíveis, pode-se definir uma função distribuição contínua e linear por partes F , classificando os X_i 's ($i=1, \dots, n$) em ordem crescente. Seja $X_{(i)}$ o i -ésimo menor dos X_j 's, tal que $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$. Então F é dada por:

$$F(x) = \begin{cases} 0 & \text{se } x < X_{(1)} \\ \frac{i-1}{n-1} + \frac{x - X_{(i)}}{(n-1)(X_{(i+1)} - X_{(i)})} & \text{se } X_{(i)} < x \leq X_{(i+1)} \text{ para } i=1,2,\dots,n-1 \\ 1 & \text{se } X_{(n)} < x \end{cases} \quad (57)$$

É importante notar que $F(x)$ cresce mais rapidamente sobre aqueles intervalos de x em que os X_i 's são mais densamente distribuídos, como é desejado. Ainda, para cada i , $F(X_{(i)}) = (i-1)/(n-1)$ é, para n grande, aproximadamente igual a proporção de X_j 's que são menores que $X_{(i)}$, característica também desejada para o comportamento da distribuição contínua. Entretanto, uma desvantagem clara ao especificar esta distribuição empírica em particular é que os valores aleatórios gerados a partir dela durante a simulação nunca podem ser menores que $X_{(1)}$ ou maiores do que $X_{(n)}$. Além disso, a média de $F(x)$ pode não ser igual à média amostral $\bar{X}(n)$ dos X_i 's.

O capítulo 4 apresenta e analisa os resultados obtidos com a aplicações de modelos de avaliação de desempenho (aproximações e simulação) para o estudo de caso deste trabalho.

3 MODELOS DE OTIMIZAÇÃO EM REDES DE FILAS: ALOCAÇÃO E REALOCAÇÃO DE SERVIDORES

Este trabalho está preocupado em analisar questões estratégicas de médio a longo prazo, relacionadas ao projeto/reprojeto de sistemas de manufatura *job-shop* representados por redes de filas abertas. Este capítulo estuda os modelos de otimização que tratam de problemas de projeto e planejamento. A revisão bibliográfica baseia-se em Bitran e Dasu (1992), Boxma et al. (1990), Bitran e Tirupati (1989a), Bitran e Tirupati (1989b), Van Vliet e Rinnooy Kan, (1991), Morabito (1998), Bitran e Morabito (1996, 1999) e Bitran e Sarkar (1994b). Porém, antes de revisar os modelos propriamente ditos, são introduzidos os problemas de projeto e planejamento abordados neste trabalho.

3.1 Problemas Tático-Estratégicos Relacionados ao Projeto/Planejamento de Sistemas Discretos de Manufatura

Um dos objetivos do planejamento tático-estratégico é identificar objetivos de médio a longo prazo da companhia. Para alcançar estes objetivos, o plano tático-estratégico determina recursos que devem ser utilizados e a política que governa o uso destes recursos. Um componente importante deste plano é o projeto do sistema produtivo. O projeto do sistema envolve (Bitran e Dasu, 1992):

- escolha de tecnologia;
- seleção de equipamento;
- alocação de produtos a diferentes plantas;
- escolha da localização das plantas;
- projeto do sistema de distribuição.

Neste trabalho, o problema de projeto está restrito aos três primeiros fatores. O projeto de uma planta é a principal restrição sob a qual o gerente operacional toma decisões. Por exemplo, se várias estações de trabalho operam próximas à capacidade e há uma considerável diversidade nos requisitos de processamento dos diferentes produtos, então filas nas estações são provavelmente longas; além disso, a variância do tempo que um produto gasta no sistema é provavelmente alta. As tarefas de

um gerente operacional incluem atribuir e prever prazos de produção para itens/produtos, controlar o fluxo de itens/produtos na instalação tal que os prazos sejam respeitados, controlar os níveis de estoque em processo (*WIP*), responder a perdas não antecipadas de capacidade devido a fatores como quebra de máquina ou incertezas de rendimento, e empenhar-se para melhorar processos e produtos. Dado que os problemas de programação, seqüenciamento e rota podem se tornar difíceis de serem tratados, o sistema deve ser projetado tal que regras de controle simples sejam adequadas para obter um bom desempenho. Daí a importância do projeto de sistemas.

O projeto de sistemas envolve conflitos (*trade-offs*) entre (Bitran e Dasu, 1992):

- custo fixo das instalações e equipamento;
- custo variável operacional da instalação;
- capacidade de processamento das máquinas;
- *throughput* (capacidade média de processamento da planta);
- *leadtime* (tempo médio que um produto gasta para ser produzido);
- complexidade de gerir a instalação.

Para os propósitos deste trabalho, considerar-se-á somente aquelas variáveis e objetivos que se relacionam mais diretamente à formação de filas nas estações. Há várias variáveis que determinam as características de desempenho de uma instalação, incluindo:

- tecnologia – tipo de equipamento, flexibilidade das máquinas, controle do processo, confiabilidade, etc.
- características do produto – nível de padronização entre os produtos, rigor de especificações, flexibilidade de processamento (rota e requisitos de máquinas), variabilidade de demanda, etc.
- capacidade do sistema – número de máquinas de diferentes tipos, níveis de força de trabalho e;
- regras de controle empregadas para regular o fluxo de itens/produtos nas estações.

Bitran e Dasu (1992) agrupam os problemas de projeto aqui tratados em 3 classes:

- desempenho desejado do sistema (*SP1- Strategical Problem 1*)
- desempenho ótimo do sistema (*SP2- Strategical Problem 2*)
- partição da instalação (*SP3- Strategical Problem 3*)

3.1.1 Desempenho Desejado do Sistema (*SP1*)

Nesta classe de problemas, certas características de desempenho do sistema são impostas externamente como metas, e a tarefa do projetista é determinar o tipo e capacidade das estações que resultam em uma instalação a custo global mínimo, atendendo estas metas. Os problemas pertencentes a esta classe podem ser formulados como problemas de otimização, nos quais o objetivo é minimizar o custo da planta e a restrição é o desempenho desejado do sistema. Pode-se pensar no exemplo de uma meta em que o *leadtime* dos produtos não exceda quatro semanas, e o problema então seria projetar a planta tal que esta restrição fosse alcançada ao menor custo possível. Abaixo segue um exemplo de um problema pertencendo a esta classe (Morabito 1998):

(*SP1.1*) *WIP* desejado:

- *Objetivo*: minimizar o custo de aquisição de equipamentos
- *Variáveis de decisão*: capacidade de cada estação, tecnologia
- *Restrições*: limitante superior para o *WIP*.

Estes problemas tratam a relação entre o capital de trabalho e o custo de equipamento (capital de investimento). Desde que o projeto de um sistema é baseado em múltiplos critérios, é útil desenvolver curvas que refletem o *trade-off* entre custos com *WIP* e o custo de aquisição de máquinas. Morabito (1998) estuda a importância destas curvas para a tomada de decisões em projeto de sistemas de manufatura *job-shop*.

O conceito de *trade-off* é aqui entendido como a troca de um benefício por um outro visto como mais vantajoso. Curvas de *trade-off* são curvas de fronteira ótima que indicam vantagens (e desvantagens) da troca de um ponto por outro desta fronteira. No caso de uma curva de *trade-off* entre os custos de *WIP* e capacidade nas estações, para cada custo de *WIP*, a curva indica o ponto com mínimo custo de recursos nas estações. Isto pode ser feito resolvendo o problema *SP1* parametricamente, variando o limitante superior para o *WIP*.

3.1.2 Desempenho Ótimo do Sistema (*SP2*)

Na classe *SP2*, há um limite de orçamento que restringe a quantidade de capital disponível para aquisição de máquinas. Esta restrição pode ser importante, quando uma instalação existente está sendo projetada ou ampliada. Sob tais circunstâncias, os modelos devem redistribuir o capital de investimento disponível (capacidade instalada), de modo a fornecer o melhor desempenho do sistema. Abaixo, segue um exemplo de problema pertencente a esta classe (Morabito, 1998).

(*SP2.1*) *WIP ótimo*:

- *Objetivo*: minimizar o *WIP*
- *Variáveis de decisão*: capacidade de cada estação, tecnologia
- *Restrições*: limitante superior para o custo de aquisição de equipamentos.

Neste trabalho, os problemas das classes *SP1* e *SP2* são abordados do ponto de vista de alocação ou realocação ótima de servidores às estações. Particularmente, dados o *mix* de produção, *throughput* e tecnologia, o objetivo é alocar (*SP1*) ou realocar (*SP2*) otimamente servidores nas estações da rede. Boxma et al. (1990) chamam os problemas das classes *SP1* e *SP2*, respectivamente, por problemas de alocação e realocação de servidores. Bitran e Tirupati (1989a) chamam o problema *SP2* de problema de balanceamento da capacidade.

3.1.3 Partição da Instalação (*SP3*)

A gestão de um sistema de manufatura torna-se complexa à medida que o número de produtos produzidos na instalação aumenta. Portanto, pode ser vantajoso particionar uma instalação grande em menores, com subunidades (sub-plantas) mais focalizadas. Entretanto, particionar uma instalação pode resultar em duplicação de equipamentos. Conseqüentemente, há um *trade-off* entre o aumento dos custos com aquisição de equipamento e a redução da complexidade gerencial. A principal decisão considerada nesta classe de problemas é a partição da instalação em sub-plantas menores. Admite-se que a instalação já existe e, portanto, trata-se de um problema de reprojeto ou replanejamento.

Para particionar uma planta em várias fábricas focalizadas, deve-se determinar as classes de produtos que serão produzidos em cada fábrica, e alocar apropriadamente equipamentos em cada uma delas. Idealmente, as fábricas não

deveriam dividir equipamentos e cada produto deveria ser atribuído (produzido) em uma única planta. Entretanto, se a demanda não é suficientemente alta ou se o custo dos equipamentos é muito alto, então pode ser mais vantajoso ter várias classes de itens/produtos partilhando o mesmo equipamento. O *trade-off* para esta classe de problemas está entre custo extra de capital para aquisição de equipamentos e a redução da complexidade gerencial. Um exemplo deste tipo de problema é (Morabito, 1998):

(SP3.1) *Número de produtos e WIP desejados em cada unidade de produção:*

- *Objetivo:* minimizar o custo de aquisição de equipamentos
- *Variáveis de decisão:* número de unidades de produção, *mix* de produtos em cada unidade, e capacidade de cada estação
- *Restrições:* limitante superior para o número de produtos e o *WIP* em cada unidade.

Para desenvolver modelos de otimização que auxiliem gerentes a particionar a instalação em plantas dentro da planta, Bitran e Sarkar (1993c) sugerem desenvolver boas medidas quantitativas de complexidade. Por se tratar de um tópico importante ao presente trabalho, é dedicado um capítulo a parte para o estudo deste assunto (capítulo 5).

3.2 Formulação dos Problemas *SP1* e *SP2*

Várias medidas de desempenho podem ser utilizadas tais como *WIP*, *leadtime* ou *throughput* (taxa média de produção do sistema). Nas seções seguintes, utilizou-se o *WIP* como medida de desempenho para formular problemas das classes *SP1.1* e *SP2.1*. Desde que *WIP* e *leadtime* são linearmente relacionados pela lei de Little (6a), os algoritmos aqui apresentados também se aplicam para o *leadtime*.

Sejam:

μ_j : a taxa de serviço média de cada máquina na estação j ;

m_j : número de máquinas idênticas na estação j ;

$F_j(\mu_j, m_j)$: custo de alocação da capacidade (μ_j, m_j) na estação j ;

F_T : orçamento disponível para aquisição de capacidade na rede;

$L_j(\mu_1, m_1; \mu_2, m_2; \dots; \mu_n, m_n)$: número médio de itens/produtos na estação j como função da capacidade da rede. No texto que segue, em certos casos, o número médio de itens/produtos na estação j é denotado simplesmente por L_j ;

v_j : valor monetário médio de um item/produto na estação j (independente de sua classe);

L_T : limitante superior do *WIP* na rede.

O *WIP* é um valor médio monetário do número de itens/produtos na rede, aqui definido como $\sum_{j=1}^n v_j L_j(\mu_1, m_1; \dots; \mu_n, m_n)$. Cada valor monetário v_j associado a um item/produto na estação j pode ser estimado usando experiência prática, ou pela média ponderada proporcional à taxa média de chegada e ao tempo médio de espera de cada classe. Sem perda de generalidade, admite-se que o valor de um produto na estação j , em fila ou em serviço, é independente do tipo de produto e igual a v_j . Obviamente, se $v_j = 1$ para todo j , então o *WIP* corresponde ao número de itens/produtos na rede.

O problema de *WIP* desejado *SP1.1* é o problema de determinar a capacidade (μ_j, m_j) para cada estação j , de maneira a minimizar o custo total e satisfazer a restrição do *WIP* desejado para a rede. O problema *SP1.1* é formulado como (Morabito, 1998):

$$\begin{aligned} & \min \sum_{j=1}^n F_j(\mu_j, m_j) \\ \text{sujeito a} & \quad \sum_{j=1}^n v_j L_j(\mu_1, m_1; \dots; \mu_n, m_n) \leq L_T \\ \text{com} & \quad (\mu_j, m_j) \in P_j, \quad j = 1, \dots, n \end{aligned}$$

sendo P_j o domínio das variáveis. Similarmente, o problema de desempenho ótimo *SP2.1* é o problema de determinar a capacidade (μ_j, m_j) para cada estação j , de maneira a minimizar o *WIP* total e satisfazer a restrição orçamentária para a rede. *SP2.1* é formulado como (Morabito, 1998):

$$\begin{aligned} & \min \sum_{j=1}^n v_j L_j(\mu_1, m_1; \dots; \mu_n, m_n) \\ \text{sujeito a} & \quad \sum_{j=1}^n F_j(\mu_j, m_j) = F_T \\ \text{com} & \quad (\mu_j, m_j) \in P_j, \quad j = 1, \dots, n \end{aligned}$$

Para ambos os problemas *SP1.1* e *SP2.1* as variáveis de decisão são a capacidade de processamento de cada estação j , determinadas pelo número de servidores (m_j) e pela taxa de processamento de cada servidor (μ_j) . A taxa de

processamento deveria ser um vetor com m_j elementos, porém supõe-se que todos os servidores sejam idênticos e, portanto, μ_j é um escalar.

Adotando a notação sugerida por Bitran e Dasu (1992), neste trabalho cada classe de problema é denotada por $\alpha/\beta/\chi/\delta$, em que $\alpha \in \{SP1.1, SP2.1, SP3.1\}$, $\beta \in \{J, G\}$, $\chi \in \{S, M\}$ e $\delta \in \{R, N\}$. O símbolo α indica o tipo de problema, β indica o tipo de rede aberta ao qual o problema é aplicado (J para redes de Jackson e G para redes genéricas), χ indica se as estações possuem múltiplas máquinas (M) ou uma única (S), e δ indica se as variáveis de decisão correspondem à taxa média de serviço (R) ou ao número de máquinas em cada estação (N).

3.3 Modelos $\cdot / J / \cdot / \cdot$

Uma consequência do resultado na forma produto do teorema de Jackson (discutido no capítulo 2) é a análise das estações de uma rede de Jackson (a partir deste momento, o termo redes de Jackson referir-se-á às redes de Jackson, às de Kelly e às *BCMP*, salvo quando se fizer necessária uma definição mais apropriada) como se fossem estocasticamente independentes. Assim, L_j em *SP1.1* e *SP2.1* torna-se uma função apenas de μ_j e m_j somente, ao invés de $\mu_1, m_1; \mu_2, m_2; \dots; \mu_n, m_n$.

3.3.1 Modelos $\cdot / J / \cdot / R$

Kleinrock (1964, 1976) considerou o problema de determinar a taxa de processamento μ_j em cada nó de uma rede de Jackson aberta de classe única, de modo a minimizar o número esperado de *jobs* em cada estação. Ele admitiu todas as estações da rede com servidor único (*M/M/1*) e que o custo de uma máquina é proporcional à taxa de processamento: $F_j(\mu_j) = f_j \mu_j$, onde f_j é o custo unitário de aquisição de capacidade na estação j . Aplicando a Lei de Little ($L_j = \lambda_j W_{qj}$) em (10) e adicionando a utilização média (λ_j / μ_j), obtém-se o número esperado de *jobs* em um sistema *M/M/1* definido como $L_j = \frac{\lambda_j}{(\mu_j - \lambda_j)}$, com λ_j computado em (1). O modelo *SP2.1/J/S/R* é

formulado como:

$$\begin{aligned} & \min \sum_{j=1}^n L_j(\mu_j) \\ \text{sujeito a} & \quad \sum_{j=1}^n f_j \mu_j = F_T \\ \text{com} & \quad \mu_j \geq 0, \quad j = 1, \dots, n \end{aligned}$$

Kleinrock (1976) mostrou que este modelo tem solução ótima μ_j^* , $j=1, \dots, n$ definida em uma forma fechada por

$$\mu_j^* = \lambda_j + \frac{\sqrt{f_j \lambda_j}}{\sum_{i=1}^n \sqrt{f_i \lambda_i}} \frac{(F_T - \sum_{i=1}^n f_i \lambda_i)}{f_j} \quad (58)$$

Note que se o custo unitário de capacidade é o mesmo para todas as estações, (58) primeiro aloca capacidade suficiente à estação j para satisfazer a taxa de chegada esperada, e depois aloca capacidade à estação j na proporção da raiz quadrada de sua taxa de chegada esperada. Observe que a alocação ótima não resulta em um sistema onde todos os servidores possuem intensidade de tráfego iguais. Bitran e Dasu (1992) observaram cinco condições no modelo acima:

- i) o número médio no sistema é uma função convexa da taxa de processamento ($L_j(\mu_j)$ é uma função convexa de μ_j);
- ii) adição de capacidade em uma estação não tem efeito no tamanho médio da fila de outras estações ($L_j(\mu_j)$ é independente de μ_i , $i \neq j$, $i=1, \dots, n$);
- iii) taxas de serviço μ_j (variáveis de decisão) são contínuas;
- iv) os custos de capacidade são funções convexas das taxas de serviço ($F_j(\mu_j)$ é uma função convexa de μ_j);
- v) $L_j(\mu_j)$ pode ser expressa na forma fechada.

As condições de (i) – (iv) reduzem o problema *SP2.1* a um problema convexo, que pode ser resolvido otimamente por meio de métodos de ótimo local. A condição (v) fornece uma forma fechada à solução do problema. Estas condições são amplamente utilizadas nas classes de problemas a seguir.

O problema *SP2.1/J/M/R* pode ser formulado exatamente como o problema *SP2.1/J/S/R*, mas com $L_j(\mu_j)$ definido para uma fila *M/M/m_j*. As condições (i) – (iv) também são satisfeitas.

3.3.2 Modelos $J/M/N$

O conteúdo apresentado nesta seção foi baseado em Boxma et al. (1990), Bitran e Morabito (1996) e Morabito (1998). Admite-se que o roteiro dos produtos, a localização das máquinas e a tecnologia estão especificados, e a questão relacionada aos problemas *SP1.1* e *SP2.1* que os modelos buscam responder refere-se a capacidade: quantas máquinas atribuir a cada estação? Aqui, os problemas são formulados com variáveis de decisão inteiras (o número de máquinas idênticas atribuídas a cada estação j). O processo de produção é modelado por redes de filas abertas, com diferentes classes de produtos que seguem roteiros determinísticos, e estações com múltiplos servidores em paralelo, cujos processos de chegadas externas e de serviço são considerados exponencialmente distribuídos.

Agregando todas as classes em uma única, tem-se cada estação j descrita pelos parâmetros $\{\lambda_j, \mu_j, m_j\}$. Kelly (1979) mostrou que a distribuição de equilíbrio do número de itens/produtos/usuários na rede pode ser expressa na forma produto, e que cada estação j pode ser analisada como se fosse uma fila $M/M/m_j$ estocasticamente independente. Aplicando a lei de Little em (10) e adicionando a utilização (λ_j/μ_j) , obtém-se o número médio de usuários L_j como uma função de m_j, λ_j, μ_j dado por (veja a expressão (5)):

$$L_j(m_j, \lambda_j, \mu_j) = \frac{\lambda_j \left(\frac{\lambda_j}{\mu_j}\right)^{m_j} \pi(0)}{\left(1 - \frac{\lambda_j}{\mu_j m_j}\right)^2 m_j!} + \frac{\lambda_j}{\mu_j} \quad \text{com}$$

$$\pi(0) = \left\{ \sum_{t=0}^{m_j-1} \frac{\left(\frac{\lambda_j}{\mu_j}\right)^t}{t!} + \frac{\left(\frac{\lambda_j}{\mu_j}\right)^{m_j}}{\left(1 - \frac{\lambda_j}{\mu_j m_j}\right) m_j!} \right\}^{-1}$$

Como as taxas de chegada e de serviço (λ_j e μ_j , respectivamente) são dadas, $L_j(m_j, \lambda_j, \mu_j)$ pode ser considerada como uma função somente de m_j , isto é, $L_j(m_j)$. Note que $L_j(m_j)$ é uma função convexa e decrescente na variável de decisão m_j e, portanto, as condições (i) e (ii) são satisfeitas (Boxma et al., 1990). O *WIP* foi adotado como medida de desempenho da rede mas, desde que o *WIP* e o *leadtime*

podem ser relacionados pela lei de Little, a análise também pode ser estendida ao *leadtime*. O *WIP* da rede (o valor total dos produtos que estão na rede) é dado por:

$$L(m_1, \dots, m_n) = \sum_{j=1}^n v_j L_j(m_j) \quad (59)$$

Além disso, supõe-se que a alocação de m_j máquinas na estação j gera custos de investimentos $F_j(m_j)$, uma função convexa e não-decrescente em m_j . Para prevenir a instabilidade do sistema, é requerida que a intensidade de tráfego em cada estação j ($\lambda_j/\mu_j m_j$) seja menor que 1. É fácil verificar que esta condição é satisfeita se m_j é maior ou igual ao limitante inferior m_j^0 :

$$m_j \geq m_j^0 = \left\lfloor \frac{\lambda_j}{\mu_j} \right\rfloor + 1 \quad (60)$$

onde $\lfloor x \rfloor$ representa o maior inteiro menor que (x) .

3.3.2.1 Modelo *SPI.1/J/M/N*

Também chamado de problema de alocação de servidores, este problema atribui servidores a estações, de maneira que o *WIP* seja menor ou igual a um limite especificado L_T , com $L_T < L(\mathbf{m}^0)$, e \mathbf{m} é o vetor das variáveis de decisão (m_1, m_2, \dots, m_n). A configuração que se busca para a rede é a de mínimo custo, ou seja, o custo com alocação de servidores deve ser mínimo. A formulação matemática para este problema pode ser:

$$\begin{aligned} \min F(\mathbf{m}) &= \sum_{j=1}^n F_j(m_j) \\ \text{sujeito a } L(\mathbf{m}) &\leq L_T && \text{SPI.1/J/M/N} \\ \text{com } m_j &\geq m_j^0, \quad m_j \text{ inteira} && j = 1, \dots, n \end{aligned}$$

Este é um problema convexo com variáveis inteiras e, apesar das condições (i), (ii) e (iv) serem satisfeitas, o uso de análise marginal não conduz necessariamente à otimalidade (condição (iii) não é satisfeita).

Seja $PI_j(m_j)$ um índice de prioridade, resultante da análise marginal de F_j e L_j , definido como o quociente de aumento no custo e a redução do *WIP* na estação j por máquina adicionada:

$$PI_j(m_j) = \frac{\Delta F_j(m_j + 1)}{-v_j \Delta L_j(m_j + 1)} \quad (61)$$

com $\Delta F_j(m_j + 1) = F_j(m_j + 1) - F_j(m_j) \geq 0$ (não-decrescente)

$$\Delta L_j(m_j + 1) = L_j(m_j + 1) - L_j(m_j) < 0 \text{ (decrescente)}$$

Boxma et al. (1990) apresentaram um algoritmo simples baseado na heurística gulosa para resolver este problema (Algoritmo 1). O algoritmo começa com a menor alocação possível dada por (60) para cada estação. Em geral, esta alocação é infactível. Em cada iteração, adiciona uma máquina na estação cujo índice de prioridade (61) é o menor possível. Ou seja, onde a melhoria local é máxima. O algoritmo pára tão logo a adição de uma máquina torna a alocação factível.

Algoritmo 1:

Passo 1: Comece com \mathbf{m}^0 , onde cada m_j^0 é dado por (60).

Passo 2: Faça $iter=1$

Passo 3: Faça $\mathbf{m}^{iter} = \mathbf{m}^{iter-1} + \mathbf{e}_{j^*}$ onde \mathbf{e}_{j^*} é o vetor de dimensão n com 1 na posição j^* e 0 nas demais posições. O índice j^* corresponde à estação com o menor índice de prioridade PI_{j^*} , dado por:

$$PI_{j^*} = \min \{PI_j(m_j), j=1, \dots, n\} \quad (62)$$

Passo 4: Se $L(\mathbf{m}^{iter}) \leq L_T$, pare (solução factível); senão $iter=iter+1$, vá para o passo 3.

Observe que a estação j^* em (62) resulta no menor aumento de $F(\mathbf{m})$ por unidade de redução de $L(\mathbf{m})$, indicado por PI_{j^*} .

As alocações geradas pelo Algoritmo 1 são, em geral, infactíveis (Teorema 1 citado por Boxma et al., 1990). Isto não necessariamente implica que a alocação com o qual o algoritmo termina seja também uma alocação ótima. Sejam $m^1, m^2, \dots, m^{p-1}, m^p$ as alocações geradas pelo Algoritmo 1 e m^* a alocação ótima para o problema $SPI.I/J/M/N$. Boxma et al. (1990) mostraram (por meio do Teorema 2 citado no referido trabalho) que $F(m^{p-1})$ e $F(m^p)$ são limitantes para o valor da solução ótima. Ou seja, $F(m^{p-1}) < F(m^*) \leq F(m^p)$. Alguns de seus resultados computacionais de redes de manufatura reais resultaram em erros relativos da ordem de 5% entre $F(m^{p-1})$ e $F(m^p)$, sugerindo que o Algoritmo 1 gera uma alocação de servidores suficientemente próxima da alocação ótima.

3.3.2.2 Modelo SP2.1/J/M/N

Também chamado de problema de realocação de servidores, este problema trata da alocação de servidores de modo a minimizar o *WIP* da rede. O número de servidores homogêneos que podem ser alocados é fixo e igual a M , com $M > \sum_{j=1}^n m_j^0$. Esta situação pode ocorrer, por exemplo, no projeto de um sistema *FMS* (*Flexible Manufacturing System*), em que um número fixo de máquinas (servidores) homogêneas é disponível. Cada máquina pode ser adequada para qualquer tipo de operação atribuindo diferentes ferramentas à máquina e o problema é distribuir as máquinas disponíveis sobre a rede de modo a otimizar a medida de desempenho da rede. A formulação matemática para este problema pode ser:

$$\begin{aligned} & \min L(\mathbf{m}) \\ \text{sujeito a} & \quad \sum_{j=1}^n m_j = M && (SP2.1/J/M/N) \\ \text{com} & \quad m_j \geq m_j^0, \quad m_j \text{ inteira} \quad j = 1, \dots, n \end{aligned}$$

Novamente, tem-se um programa convexo com variáveis inteiras (condições (i), (ii) e (iv) satisfeitas, mas a (iii) não) e o uso de algoritmos baseados em análise marginal podem não produzir uma solução ótima de *SP2.1/J/M/N*. Seja $PI_j(m_j)$ um índice de prioridade definido como o decréscimo de *WIP* por máquina na estação j , dado por:

$$PI_j(m_j) = -v_j \Delta L_j(m_j + 1) \quad (63)$$

$$\text{com } \Delta L_j(m_j + 1) = L_j(m_j + 1) - L_j(m_j) < 0$$

Boxma et al. (1990) apresentaram um algoritmo simples (Algoritmo 2), similar ao Algoritmo 1, também baseado na estratégia gulosa para resolver o problema *SP2.1/J/M/N*. O algoritmo começa com a menor alocação possível (60) para cada estação j . Em cada iteração, adiciona-se uma máquina (servidor) à estação com o maior índice de prioridade (63), ou seja, aquela estação que gera a maior redução de $L(\mathbf{m})$ por máquina. O algoritmo pára quando todas as máquinas ($M - \sum_{j=1}^n m_j^0$) forem alocadas.

Algoritmo 2:

Passo 1: Comece com \mathbf{m}^0 , onde m_j^0 para cada estação j da rede é dado por (60).

Passo 2: Faça $iter=1$

Passo 3: Faça $\mathbf{m}^{iter} = \mathbf{m}^{iter-1} + \mathbf{e}_{j^*}$ onde \mathbf{e}_{j^*} é o vetor de dimensão n com 1 na posição j^* e 0 nas demais posições. O índice j^* corresponde à estação com o maior índice de prioridade PI_{j^*} dado por:

$$PI_{j^*} = \max\{PI_j(m_j), j=1, \dots, n\} \quad (64)$$

Passo 4: Se $iter = M - \sum_{j=1}^n m_j^0$, então pare; senão $iter = iter + 1$ e vá para o passo 3.

A estação j^* escolhida em (64) gera a maior redução de $L(\mathbf{m})$ por máquina indicada pelo índice PI_{j^*} . Boxma et al. (1990) provaram (vide teorema 3 do referido trabalho) que a alocação \mathbf{m}^* (em que o Algoritmo 2 termina) é uma alocação ótima para $SP2.1/J/M/N$, apesar da violação da condição (iii).

3.4 Modelos $\cdot/G/\cdot/\cdot$

Esta seção trata os modelos $SP1.1/G/S/R$ e $SP2.1/G/S/R$ (Bitran e Tirupati, 1989a), $SP1.1/G/M/R$ com variáveis discretas (Bitran e Tirupati, 1989b) e $SP1.1/G/M/N$ e $SP2.1/G/M/N$ (Van Vliet e Rinooy Kan, 1991).

3.4.1 Modelos $\cdot/G/\cdot/R$

Inicialmente são apresentados dois algoritmos introduzidos por Bitran e Tirupati (1989a) para resolver o modelo $SP1.1/G/S/R$ (chamado de problema de medida de desempenho desejado, ou do inglês *targeting problem*, cujo objetivo é alcançar um nível de desempenho desejado) e o modelo $SP2.1/G/S/R$ (chamado de problema de balanceamento, ou do inglês *balancing problem*, referindo à redistribuição de capacidade na instalação de modo a otimizar uma determinada medida de desempenho). Estes algoritmos podem ser facilmente estendidos para tratar modelos $SP1.1/G/M/R$ e $SP2.1/G/M/R$.

3.4.1.1 Modelo $SP1.1/G/S/R$

Considera-se redes de filas abertas, em que cada nó da rede pode ser representado por um sistema de filas ($GI/G/I$), processando múltiplas classes de produtos com roteiros determinísticos. Na ausência de métodos exatos para avaliar o desempenho de redes genéricas, o método aproximado de decomposição é utilizado

para estimar os parâmetros de variabilidade de cada estação. Mesmo com o uso de aproximações, os problemas de otimização encontrados são programas não-lineares e não-convexos. A importante contribuição de Bitran e Tirupati (1989a), conforme é visto mais adiante, é fornecer condições realísticas para que os problemas *SPI.1/G/S/R* e *SP2.1/G/S/R* possam ser aproximados por problemas convexos de fácil solução.

Para calcular a média e o parâmetro de variabilidade (*scv*) do intervalo de tempo entre chegadas em cada estação $j, j=1, \dots, n$ (Passo 1 do método de decomposição) pode-se, por exemplo, resolver o sistema linear formado pelas aproximações (15), (27), (49), e (1) ou (35), representado simplesmente (quando $m_j=1, j=1, \dots, n$) por:

$$\Phi(\lambda, ca, \mu, cs) = \theta \quad (65)$$

em que os vetores λ, ca, μ e cs denotam $\{\lambda, ca, \mu, cs\}$ para toda estação $j, j=1, \dots, n$. O número esperado de itens/produtos L_j para cada estação j modelada como uma fila *GI/G/1*, é aproximado por (20), por meio de uma função de λ_j, ca_j, μ_j e cs_j :

$$L_j(\lambda_j, ca_j, \mu_j, cs_j) = \frac{\left(\frac{\lambda_j}{\mu_j}\right)^2 (ca_j + cs_j) g(\lambda_j, ca_j, \mu_j, cs_j)}{2\left(1 - \frac{\lambda_j}{\mu_j}\right)} + \frac{\lambda_j}{\mu_j} \quad (20)$$

com $g(\lambda_j, ca_j, \mu_j, cs_j)$ definido conforme (20).

Assim, uma formulação para o problema *SPI.1*/G/S/R* é:

$$\min F(\mu) = \sum_{j=1}^n F_j(\mu_j)$$

sujeito a (65), (20),

$$\sum_{j=1}^n v_j L_j(\lambda_j, ca_j, \mu_j, cs_j) \leq L_T \quad (66) \quad (SPI.1*/G/S/R)$$

$$e \quad \mu_j \geq \mu_j^0, \quad j = 1, \dots, n \quad (67)$$

onde μ_j^0 é um limitante inferior para a capacidade na estação j , de modo a garantir a estabilidade do sistema ($\rho_j < 1$), ou seja, $\mu_j^0 > \lambda_j$ e L_T o nível desejado de *WIP* da rede.

No problema *SPI.1*/G/S/R*, a restrição (66) especifica que o *WIP* não deve exceder o nível desejado L_T . As restrições (67) restringem mudanças em capacidade somente por adições (mas esta restrição não é essencial para os resultados

alcançados por Bitran e Tirupati (1989a) e pode ser modificada de modo a permitir mudanças em capacidade também por remoções).

As restrições (66) e (67) são lineares e a complexidade do problema $SPI.1*/G/S/R$ é devida à função-objetivo, (65) e (20). Desde que L_j é uma função de λ_j , ca_j , μ_j e cs_j , e λ , ca , μ e cs são relacionados em (65), obtém-se L_j como uma função de λ , ca , μ e cs . Bitran e Tirupati (1989a) consideraram cada capacidade $\mu_j, j=1, \dots, n$, como uma variável de decisão contínua (condição (iii) satisfeita), supondo que adições de capacidade nas estações podem ser em pequenos incrementos quando comparados à capacidade total. Para um dado λ , (65) e (20) sugerem que mudanças na capacidade μ resultam em mudanças em ca e cs . Portanto, L_j é uma função de $\mu_1, \mu_2, \dots, \mu_n$. Entretanto, esta relação funcional é complexa desde que o sistema de equações em (65) é não-linear e de difícil análise com respeito a μ . Daí, as hipóteses simplificadoras propostas por Bitran e Tirupati (1989a):

Hipótese 1 (H1) – supõe-se que $F_j(\mu_j)$, o custo de alocar a capacidade μ_j à estação j , é uma função convexa, não-decrescente e diferenciável de μ_j . “O que é razoável considerando que opções mais baratas são usadas primeiro e capacidade possa ser adicionada em pequenos incrementos a um certo custo” (Bitran e Tirupati, 1989a);

Hipótese 2 (H2) – os coeficientes quadráticos de variação (ca e cs) são independentes de mudanças de capacidade. Admite-se que enquanto μ é modificado, o quadrado da média e variância do tempo de serviço variam na mesma proporção e, portanto, cs permanece praticamente constante. Quando este não for o caso, os resultados propostos ainda se aplicam, contanto que a variação no cs seja tal que a aproximação para estimar L_j é convexa em μ_j . Além disso, a sensibilidade de ca a mudanças em μ parece ser pequena, resultado suportado pelo Princípio de Aproximação de Tráfego Leve (vide capítulo2). Este princípio sugere que, sob circunstâncias de tráfego leve (número de classes de produtos alta e proporção da intensidade de tráfego devida a cada classe pequena), a distribuição dos intervalos de tempo entre partidas para cada classe de produto tende a ser igual a distribuição dos intervalos de tempo entre chegadas, e torna-se independente do processo de serviço e de fila. Em outras palavras, ca torna-se independente de μ .

A consequência dessas hipóteses é a possibilidade de se resolver o sistema (65) inicialmente para uma dada capacidade inicial e então tratar ca e cs como parâmetros conhecidos em (20). Deste modo, a restrição (65) torna-se redundante no problema $SPI.1*/G/S/R$, podendo ser excluída do problema. Além disso, Bitran e Tirupati (1989a) também mostraram que, sobre as hipóteses H1 e H2, $L_j(\lambda_j, ca_j, \mu_j, cs_j)$ em (20) é uma função convexa em μ_j , agora denotada simplesmente por $L_j(\mu_j)$ (condição (i) satisfeita). Expressando o WIP da rede como

$$L(\boldsymbol{\mu}) = L(\mu_1, \dots, \mu_n) = \sum_{j=1}^n v_j L_j(\mu_j) \quad (68)$$

e simplificando o problema $SPI.1*/G/S/R$, obtém-se o problema de programação convexa $SPI.1/G/S/R$ (as hipóteses H1 e H2 garantem que o problema abaixo é convexo):

$$\begin{aligned} \min F(\boldsymbol{\mu}) &= \sum_{j=1}^n F_j(\mu_j) \\ \text{sujeito a } L(\boldsymbol{\mu}) &\leq L_T && SPI.1/G/S/R \\ \text{com } \mu_j &\geq \mu_j^0, \quad j = 1, \dots, n \end{aligned}$$

onde $L(\boldsymbol{\mu})$ é uma função convexa, pois é a combinação positiva de funções convexas $L_j(\mu_j)$, em que cada $L_j(\mu_j)$ é dada por (20) assumindo λ_j , ca_j e cs_j como parâmetros conhecidos.

Bitran e Tirupati (1989a) apresentaram uma heurística (Algoritmo A1, no Apêndice A) para resolver o problema $SPI.1/G/S/R$ e para gerar curvas de *trade-off* entre $F(\boldsymbol{\mu})$ e $L(\boldsymbol{\mu})$.

O algoritmo A1 é semelhante ao Algoritmo 1 apresentado neste capítulo. Experimentos computacionais indicaram que a hipótese de considerar ca e cs independentes de mudanças em $\boldsymbol{\mu}$ é razoável (é importante observar que no Algoritmo A1 ca e cs são mantidos constantes, hipótese H2). Em um experimento de Bitran e Tirupati (1989a), o WIP foi reduzido de 69972 para 30000 com uma mudança máxima em ca de 3%. Esta mudança foi obtida atualizando ca e resolvendo o sistema linear (65) para a configuração final da rede. Além disso, os experimentos mostraram que o erro relativo entre a solução encontrada pela aplicação da heurística e a solução ótima foi de 0,6%, o que é aceitável para muitas situações reais

Bitran e Sarkar (1994b) propuseram um algoritmo iterativo exato para resolver o problema $SPI.1/G/S/R$, relaxando a hipótese de que ca é insensível a mudanças de capacidade (problema $SPI.1^*/G/S/R$). Além disso, estabelece condições suficientes para que o algoritmo convirja para redes de interesse prático.

Quando a hipótese H2 é relaxada, o problema $SPI.1^*/G/S/R$ não necessariamente é convexo, dadas as relações funcionais complexas entre μ e ca em (65). Mas uma vez que ca (ca_j para $j=1, \dots, n$) é determinado, o problema torna-se convexo porque $L_j(\mu_j)$ é convexo em μ_j , conforme discutido anteriormente. O algoritmo iterativo proposto por Bitran e Sarkar (1994b) usa este fato para redes de filas com classe única (ou múltiplas classes convenientemente agregadas), à medida que em cada iteração considera ca independente a mudanças na capacidade das estações (veja passo 1 do algoritmo 3).

Algoritmo 3 (Bitran e Sarkar, 1994b)

Passo 1: Escolha ca^0 , os valores iniciais de ca_j^0 para $j=1, \dots, n$. Os valores iniciais de ca_j^0 ($j=1, \dots, n$) para Redes Generalizadas processando múltiplas classes de produtos podem ser obtidos pela aplicação do método de decomposição, conforme aborda o capítulo 2. Faça $iter=1$.

Passo 2: Resolva o problema $SPI.1/G/S/R$, convexo em μ_j :

$$\begin{aligned} \min F(\mu) &= \sum_{j=1}^n F_j(\mu_j) \\ \text{sujeito a} \quad L(\mu) &\leq L_T \\ \text{com} \quad \mu_j &\geq \mu_j^0, \quad j = 1, \dots, n \end{aligned}$$

utilizando valores atualizados de ca (ca_j^{iter-1} , $j=1, \dots, n$). Denote a solução ótima por $\mu_j^{iter}(ca_j^{iter-1})$.

Passo 3: Recalcule ca (ca_j^{iter} , para $j=1, \dots, n$) utilizando as novas capacidades $\mu_j^{iter}(ca_j^{iter-1})$ no sistema (65). Pare se ca_j^{iter-1} e ca_j^{iter} forem muito próximos. Senão, faça $iter=iter+1$ e volte ao passo 2.

3.4.1.2 Modelo $SP2.1/G/S/R$

Também chamado de problema de balanceamento ou de realocação de servidores, esta seção analisa o problema de redistribuição de capacidade existente nas estações para minimizar *WIP*. Esta redistribuição é significativa em redes com capacidade homogênea, ou seja, recursos que podem ser compartilhados por diferentes estações. Um exemplo de capacidade transferível é a de um funcionário treinado, que pode executar tarefas em diferentes estações. Para toda esta seção, a capacidade será admitida homogênea, apesar de se permitir que os resultados sejam estendidos para casos mais gerais em que as capacidades possam ser transferidas somente para um subconjunto restrito de estações. O problema *SP2.1*/G/S/R* pode ser formulado da seguinte maneira:

$$\begin{aligned}
 & \min L(\lambda, ca, \mu, cs) \\
 \text{sujeito a} & \quad (65), (20), & \text{SP2.1*/G/S/R} \\
 & \sum_{j=1}^n \mu_j = \sum_{j=1}^n \mu_j^1 \\
 \text{e} & \quad \mu_j \geq \lambda_j, \quad j = 1, \dots, n
 \end{aligned}$$

em que μ_j^1 é a capacidade inicial existente na estação j tal que $\mu_j^1 \geq \mu_j^0$, para que o sistema esteja inicialmente estável.

Similarmente ao problema *SP1.1*/G/S/R*, considerando a hipótese H2 e o fato de que, sob esta hipótese, $L_j(\mu_j)$ em (20) é convexa em μ_j , o problema *SP2.1*/G/S/R* se reduz ao problema convexo *SP2.1/G/S/R*:

$$\begin{aligned}
 & \min L(\mu) = \sum_{j=1}^n v_j L_j(\mu_j) \\
 \text{sujeito a} & \quad \sum_{j=1}^n \mu_j = \sum_{j=1}^n \mu_j^1 & \text{(SP2.1/G/S/R)} \\
 \text{com} & \quad \mu_j \geq \lambda_j, \quad j = 1, \dots, n
 \end{aligned}$$

com $L_j(\mu_j)$ dada por (20) e supondo λ_j , ca_j e cs_j como parâmetros conhecidos (calculados inicialmente pelo sistema linear (65)).

Bitran e Tirupati (1989a) apresentaram uma heurística (Algoritmo A2, no Apêndice A) também baseada na estratégia gulosa para resolver o problema *SP2.1/G/S/R* (condições (i) – (iv) são satisfeitas). Realizaram experimentos computacionais utilizando o mesmo conjunto de dados utilizado no Algoritmo A1 (vide Apêndice A). Assumindo incrementos de capacidade fixados em 0,02, foi possível reduzir o *WIP* de 69972 para 40256. A redução dramática foi parcialmente

atribuída à hipótese de que a capacidade homogênea é completamente transferível. O Algoritmo A2 e seus resultados também podem ser modificados para casos em que transferência de capacidade estão restritas a algum subconjunto de estações. Observaram também que a mudança máxima em ca foi de 3%. Novamente, ca e cs são assumidos invariantes no Algoritmo A2 e a mudança em ca foi calculada através de (65) com a configuração final da rede, alcançada após o término do Algoritmo A2. O Algoritmo A2 apresentou erro relativo inferior a 2% entre a solução ótima e a solução encontrada pela heurística após p -iterações. Estes valores, associados ao baixo requisito computacional, indicam que o algoritmo é eficiente para resolver problemas reais.

Novamente, pode-se relaxar a hipótese de que ca_j é constante a mudanças na capacidade para $j=1, \dots, n$, e empregar o algoritmo iterativo (Algoritmo 4) proposto por Bitran e Sarkar (1994b).

Algoritmo 4

Passo 1: Escolha ca^0 , os valores iniciais de ca_j^0 para $j=1, \dots, n$. Os valores iniciais de ca_j^0 ($j=1, \dots, n$) para Redes Generalizadas processando múltiplas classes de produtos podem ser obtidos pela aplicação do método de decomposição, por meio de (65) por exemplo. Faça $iter=1$.

Passo 2: Resolva o problema $SP2.I/G/S/R$, convexo em μ_j :

$$\begin{aligned} \min L(\boldsymbol{\mu}) &= \sum_{j=1}^n v_j L_j(\mu_j) \\ \text{sujeito a} \quad & \sum_{j=1}^n \mu_j = \sum_{j=1}^n \mu_j^1 \\ \text{com} \quad & \mu_j \geq \lambda_j, \quad j = 1, \dots, n \end{aligned}$$

utilizando valores atualizados de ca (ca_j^{iter-1} , $j=1, \dots, n$). Denote a solução ótima por $\mu_j^{iter}(ca_j^{iter-1})$.

Passo 3: Recalcule ca (ca_j^{iter} , para $j=1, \dots, n$) utilizando as novas capacidades $\mu_j^{iter}(ca_j^{iter-1})$ no sistema (65). Pare se ca_j^{iter-1} e ca_j^{iter} forem muito próximos. Senão, faça $iter=iter+1$ e volte ao passo 2.

As hipóteses de convexidade para $L_j(\mu_j)$ e de convergência para o Algoritmo 4 são as mesmas já discutidas no Algoritmo 3.

Os Algoritmos 2, A1 e 4 (problema de medida de desempenho ótima ou problema de realocação de servidores) auxiliam no balanceamento do sistema de manufatura, enquanto os algoritmos 1, A2 e 3 (problema de medida de desempenho desejada ou problema de alocação de servidores) acrescentam recursos eficientemente ao sistema. Estes problemas são complementares e curvas de *trade-off* podem ser geradas entre capital de trabalho (*WIP*) e capital de investimento ($F(\mu)$): “primeiro aplica-se o Algoritmo A2 à configuração do sistema original e, utiliza-se a solução obtida, μ^p , como capacidade inicial para o Algoritmo A1 (isto é, $\mu^0 \leftarrow \mu^p$ onde μ^0 é a capacidade inicial no passo 1 do Algoritmo A1) para promover uma redução maior no *WIP* da rede” (Morabito, 1998). Bitran e Tirupati (1989a) e Bitran e Morabito (1999) realizaram diversos experimentos computacionais e geraram e analisaram as curvas *trade-off* obtidas por meio destes procedimentos.

3.4.1.3 Modelo G/R com alternativas discretas

Esta seção trata o problema de adicionar capacidade com o objetivo de alcançar um nível de *WIP* desejado, L_T , a custo mínimo. Supõe-se que adição de capacidade em cada estação possa ser modelado por um conjunto de alternativas discretas. Tratar capacidade como variável discreta pode fornecer considerável flexibilidade na modelagem de sistemas de manufatura. Adições de máquinas e/ou mão-de-obra (funcionários) e introdução de novo turno de operação são exemplos de adições de capacidades que podem ser modeladas como variáveis discretas.

Novamente, sistemas de manufatura que processam uma variedade de produtos com roteiros determinísticos são modelados como redes de filas abertas com distribuições gerais para os processos de chegada e de serviço. Admite-se que os intervalos de tempo entre chegadas dentro de cada classe e os tempos de serviço para todas as classes têm distribuições independentes e idênticas.

A falta de soluções exatas para redes cujos nós são modelados como sistemas de filas genéricas ($GI/G/m$) motiva a utilização de aproximações para avaliar o desempenho da rede. Especificamente, são utilizadas aproximações baseadas no método de decomposição.

O sistema formado, por exemplo, pelas aproximações (15), (27), (49), e (1) ou (35) do Passo 1 do método de decomposição é denotado simplesmente por:

$$\Phi(\mathbf{m}, \boldsymbol{\lambda}, \mathbf{ca}, \boldsymbol{\mu}, \mathbf{cs}) = \mathbf{0} \quad (69)$$

em que os vetores \mathbf{m} , $\boldsymbol{\lambda}$, \mathbf{ca} , $\boldsymbol{\mu}$ e \mathbf{cs} correspondem, respectivamente, aos parâmetros $\{m_j, \lambda_j, ca_j, \mu_j, cs_j\}$ para toda estação $j, j=1, \dots, n$. O número esperado de itens/produtos na estação j (para outras aproximações de L_j , veja Whitt, 1993) pode ser dado por (29):

$$L_j(m_j, \lambda_j, ca_j, \mu_j, cs_j) = \frac{\lambda_j(ca_j + cs_j)}{2} Lq_j(M / M / m_j) + \frac{\lambda_j}{\mu_j} \quad (29)$$

com Lq_j , o número esperado de *jobs* em uma fila $M/M/m_j$, definido em (29). Similarmente a (20), $L_j(m_j, \lambda_j, ca_j, \mu_j, cs_j)$ em (29) é uma função de $m_j, \lambda_j, ca_j, \mu_j$ e cs_j satisfazendo (69).

Na seqüência, supõe-se que o roteiro dos produtos, localização das máquinas e a tecnologia estão especificados, e o problema de projeto resume-se a adicionar capacidade em estações a mínimo custo de modo a minimizar o *WIP*. Ao invés de escolher \mathbf{m} ou $\boldsymbol{\mu}$ como variáveis de decisão do modelo, Bitran e Tirupati (1989b) consideraram um número finito de alternativas para mudanças de capacidade em cada estação.

Seja n_j o número total de alternativas na estação j . Para cada alternativa $t, t=1, \dots, n_j$, o seguinte dado de entrada é fornecido:

m_{jt} = número de máquinas idênticas na estação j na alternativa t ;

μ_{jt} = taxa média de serviço de cada máquina na estação j na alternativa t ;

f_{jt} = custo da estação j na alternativa t .

Defina u_{jt} a variável de decisão binária 0 – 1 (condição (iii) violada) tal que

$$u_{jt} = \begin{cases} 1, & \text{se alternativa } t \text{ é escolhida para estação } j \\ 0, & \text{caso contrário} \end{cases}$$

em que $\sum_{t=1}^{n_j} u_{jt} = 1, j=1, \dots, n$. Observe que, deste modo, a variável binária u_{jt} “carrega” as informações de taxa de serviço (μ_{jt}) e de número de máquinas (m_{jt}) para um conjunto discreto de alternativas $t (t=1, \dots, n_j)$ para cada estação j . Em cada estação j , a escolha de capacidade é representada pelo vetor $(u_{j1}, u_{j2}, \dots, u_{jn_j})$, no qual todos

elementos são nulos exceto um. Desta maneira, tem-se $m_j = \sum_{t=1}^{n_j} m_{jt} u_{jt}$ e $\mu_j = \sum_{t=1}^{n_j} \mu_{jt} u_{jt}$.

Uma formulação para o problema de adicionar capacidade de modo a minimizar o WIP a custo mínimo (*SPI.1*/G/M/R*) é:

$$\begin{aligned} \min F(\mathbf{u}) &= \sum_{j=1}^n \sum_{t=1}^{n_j} f_{jt} u_{jt} \\ \text{sujeito a } (29), (69), \\ &\sum_{j=1}^n v_j L_j(m_j, \lambda_j, ca_j, \mu_j, cs_j) \leq L_T \quad (70) \\ &\mu_j - \sum_{t=1}^{n_j} \mu_{jt} u_{jt} = 0, \quad j = 1, \dots, n \quad (71) \quad (\text{SPI.1*/G/M/R}) \\ &m_j - \sum_{t=1}^{n_j} m_{jt} u_{jt} = 0, \quad j = 1, \dots, n \quad (72) \\ &\sum_{t=1}^{n_j} u_{jt} = 1, \quad j = 1, \dots, n \quad (73) \\ &u_{jt} \in \{0,1\}, \quad t = 1, \dots, n_j; \quad j = 1, \dots, n \quad (74) \end{aligned}$$

A função-objetivo representa o custo total com adição de capacidade; (29) e (69) representam as aproximações para estimar o número médio de *jobs* em cada estação, enquanto (70) representa *WIP* da rede. Convém observar que a expressão (70) também pode ser substituída por uma restrição que representa o *leadtime* médio da rede $E(T)$, ou o *leadtime* médio de cada classe $E(T_k)$, conforme discutido no capítulo 4. Já (71) e (72) determinam a capacidade em cada estação. As expressões (73) e (74) juntas restringem a escolha de exatamente uma alternativa t para cada estação j . *SPI.1*/G/M/R* é um problema misto e não-linear, difícil de resolver. As restrições (29) e (73) são não-lineares e contêm um número de funções implícitas que tornam o problema particularmente difícil. Segue daí, a necessidade de hipóteses simplificadoras.

Pelos mesmos motivos já explicitados na seção anterior, assume-se que ca e cs são invariantes a mudanças de capacidade (hipótese H2). Uma consequência imediata desta hipótese é que o sistema de equações (69) pode ser resolvido inicialmente e os vetores ca e cs podem ser tratados como parâmetros para o problema *SPI.1*/G/M/R*. Além disso, o número médio de itens/produtos em cada estação é função de sua capacidade m_j e μ_j . Assim, enquanto se escolhe uma alternativa t na estação j (ou seja, $u_{jt}=1$ e $u_{jl}=0, l \neq t$), pode-se referir a $L_j(m_j, \lambda_j, ca_j, \mu_j, cs_j)$ em (29) simplesmente como L_{jt} onde $L_{jt}=L_j(m_{jt}, \lambda_j, ca_j, \mu_{jt}, cs_j)$. Desde que a hipótese H2 se

aplique a $SPI.1^*/G/M/R$, então este problema é equivalente ao problema $SPI.1/G/M/R$, definido a seguir:

$$\begin{aligned} \min F(\mathbf{u}) &= \sum_{j=1}^n \sum_{t=1}^{n_j} f_{jt} u_{jt} \\ \text{sujeito a } \sum_{t=1}^{n_j} u_{jt} &= 1, \quad j = 1, \dots, n \\ \sum_{j=1}^n \sum_{t=1}^{n_j} (v_j L_{jt}) u_{jt} &\leq L_T \\ u_{jt} &\in \{0,1\}, \quad t = 1, \dots, n_j; \quad j = 1, \dots, n \end{aligned} \quad (SPI.1/G/M/R)$$

Na formulação de $SPI.1/G/M/R$, assume-se, sem perda de generalidade, que se $L_{jt} > L_{jl}$, então $f_{jt} < f_{jl}$, $t \neq l$, para $t, l = 1, \dots, n$.

Mesmo com a hipótese H2, $SPI.1/G/M/R$ é um problema inteiro difícil de ser resolvido. Relaxando as condições de integralidade de $SPI.1/G/M/R$ tem-se o problema $SPI.1^{relaxado}/G/M/R$:

$$\begin{aligned} \min F(\mathbf{u}) &= \sum_{j=1}^n \sum_{t=1}^{n_j} f_{jt} u_{jt} \\ \text{sujeito a } \sum_{t=1}^{n_j} u_{jt} &= 1 \\ \sum_{j=1}^n \sum_{t=1}^{n_j} (v_j L_{jt}) u_{jt} &\leq L_T \\ u_{jt} &\geq 0, \quad t = 1, \dots, n_j; \quad j = 1, \dots, n \end{aligned} \quad (SPI.1^{relaxado}/G/M/R)$$

Bitran e Tirupati (1989b) apresentaram uma heurística (Algoritmo A3, Apêndice A) baseada no problema $SPI.1^{relaxado}/G/M/R$ para obter uma solução aproximada para o problema $SPI.1/G/M/R$. Novamente, no modelo $SPI.1^{relaxado}/G/M/R$, a expressão relacionando o WIP da rede L_T pode ser substituída por uma expressão relacionando o $leadtime$ médio da rede $E(T)$, ou o $leadtime$ médio de cada classe k da rede $E(T_k)$, conforme discutido no capítulo 4.

Experimentos computacionais com um exemplo de uma rede real composta de 13 estações e 10 classes de produtos com roteiros determinísticos indicaram que o Algoritmo A3 é uma boa aproximação para $SPI.1/G/M/R$ quando o número de classes é relativamente grande. Além disso, a maior mudança em ca foi de 4,6% (é importante frisar que os valores de ca e cs são mantidos constantes no Algoritmo A3 e que esta mudança foi calculada com os resultados obtidos para a configuração final da rede, obtida após o término do algoritmo).

Morabito (1998) adaptou o algoritmo 3 para $SPI.1/G/S/R$ com alternativas discretas (Algoritmo 5).

Algoritmo 5:

Passo 1: Escolha \mathbf{ca}^0 , os valores iniciais de ca_j^0 para $j=1, \dots, n$. Os valores iniciais de ca_j^0 ($j=1, \dots, n$) para Redes Generalizadas processando múltiplas classes de produtos podem ser obtidos pela aplicação do método de decomposição, ou seja, resolvendo (69) por exemplo. Faça $iter=1$.

Passo 2: Resolva o problema $SP1.1/G/S/R^{discreto}$:

$$\begin{aligned} \min F(\mathbf{u}) &= \sum_{j=1}^n \sum_{t=1}^{n_j} f_{jt} u_{jt} \\ \text{sujeito a } L(\mathbf{u}) &\leq L_T \\ \sum_{t=1}^{n_j} u_{jt} &= 1 \quad j = 1, \dots, n \\ u_{jt} &\in \{0,1\}, \quad t = 1, \dots, n_j; \quad j = 1, \dots, n \end{aligned}$$

utilizando valores atualizados de \mathbf{ca} (ca_j^{iter-1} , $j=1, \dots, n$). Denote a solução ótima de $SP1.1/G/S/R^{discreto}$ por $u_{jt}^{iter}(ca_j^{iter-1})$. Observe que se $u_{jt}=1$, então a capacidade μ_{jt} é alocada na estação j . Seja $\mu_j^{iter}(ca_j^{iter-1})$ a capacidade alocada na estação j .

Passo 3: Recalcule \mathbf{ca} (ca_j^{iter} , para $j=1, \dots, n$) utilizando as novas capacidades $\mu_j^{iter}(ca_j^{iter-1})$ no sistema (69). Pare se ca_j^{iter-1} e ca_j^{iter} forem suficientemente próximos. Senão, faça $iter=iter+1$ e volte ao passo 2.

De maneira similar, o Algoritmo 4 também pode ser adaptado para $SP2.1/G/S/R$ com alternativas discretas.

O modelo $SP1.1/G/S/R^{discreto}$ também pode ser adaptado de forma que o *leadtime* seja a medida de desempenho a ser alcançada, ao invés do *WIP*, e que cada classe de produto para uma rede processando múltiplas classes não exceda o limitante superior imposto ao modelo. O modelo adaptado (aqui denotado por $SP1.2/G/S/R^{discreto}$) também foi utilizado no estudo de caso (vide seção 4.5), pois mostrou-se bastante conveniente para apoiar decisões de alocação de capacidade, uma vez que para manter-se competitivo no mercado, a empresa deve saber onde e quanto alocar de capacidade de modo a alcançar o *leadtime* requerido pelo mercado (clientes).

3.4.2 Modelos $\cdot /G/M/N$

Esta seção aborda os modelos *SP1.I/G/M/N* (chamados de problema de alocação de servidores) e *SP2.I/G/M/N* (chamados de problema de realocação de servidores). Nestes problemas, a variável de decisão inteira corresponde ao número de máquinas idênticas em cada estação, m_j . Na seqüência, novamente, admite-se que rotas de produtos, intensidade de tráfego, localização de máquinas e tecnologia (tempos de processamento) já estão especificados. A questão de projeto que os modelos buscam responder refere-se à capacidade: quantas máquinas atribuir a cada estação?

O sistema de manufatura é modelado como uma rede de filas aberta e genérica, com diferentes classes de produtos que perfazem roteiros de fabricação determinísticos. As estações consistem de várias máquinas idênticas paralelas.

Os algoritmos apresentados por Van Vliet e Rinnooy Kan (1991) são uma combinação de otimização combinatória e teoria de redes de filas. Os autores estenderam os algoritmos apresentados por Boxma et al. (1990) para redes mais gerais. A falta de soluções exatas para redes cujos nós são modelados como sistemas de filas genéricas (*GI/G/m*) motiva a utilização de aproximações para avaliar o desempenho da rede, especificamente as baseadas no método de decomposição.

Para o Passo 1 do método de decomposição, Van Vliet e Rinnooy Kan (1991) resolveram o sistema formado por (26), (35), (49) e a expressão (75), adaptada de (15):

$$ca_j = \sum_{k=1}^r \sum_{l=1}^{n_k} \frac{\lambda'_k}{\lambda_j} ca_{kl} 1\{n_{kl} = j\} \quad (75)$$

O sistema foi representado simplesmente por (69). A solução do sistema (69) aproxima ca_j e calcula exatamente λ_j para cada estação j da rede.

No passo 2, o número esperado de *jobs* na estação j foi aproximado por (29). A expressão de $L_j(m_j, \lambda_j, ca_j, \mu_j, cs_j)$ em (29) é função de m_j , λ_j , ca_j , μ_j e cs_j satisfazendo (69). Baseados nos resultados de Bitran e Tirupati (1989a), Van Vliet e Rinnooy Kan (1991) adotaram a hipótese H2 (ca e cs independentes de mudanças de capacidade). Isto significa que o sistema (69) não tem de ser recalculado quando mudanças nas variáveis de decisão m_j ocorrem. Assim, uma vez que o sistema (69) é calculado, ca pode ser considerado como parâmetro para os problemas abordados a

seguir e $L_j(m_j, \lambda_j, ca_j, \mu_j, cs_j)$ pode ser considerada uma função convexa de m_j , $L_j(m_j)$ (para mais detalhes vide Van Vliet e Rinnooy Kan, 1991).

Novamente, o WIP da rede, $L(\mathbf{m})$, é dado por (59), com $L_j(m_j)$ dado por (29).

3.4.2.1 Modelo $SPI.1/G/M/N$

Também chamado de problema de alocação de servidores (máquinas), este modelo aloca servidores de tal forma que o WIP seja menor ou igual a um limite especificado L_T . A configuração que se busca é a de mínimo custo. A formulação matemática para o problema $SPI.1/G/M/N$, tal como para o problema $SPI.1/J/M/N$, é:

$$\begin{aligned} \min F(\mathbf{m}) &= \sum_{j=1}^n F_j(m_j) \\ \text{sujeito a } L(\mathbf{m}) &\leq L_T && SPI.1/G/M/N \\ \text{com } m_j &\geq m_j^0, \quad m_j \text{ inteira} && j = 1, \dots, n \end{aligned}$$

Novamente, $F(\mathbf{m})$ é uma função convexa e não-decrescente em \mathbf{m} , enquanto $L(\mathbf{m})$ é uma função convexa e não-crescente nesta variável.

Este é um problema convexo com variáveis inteiras e, apesar das condições (i), (ii) e (iv) serem satisfeitas, o uso de análise marginal não conduz necessariamente à otimalidade (condição (iii) não é satisfeita). Van Vliet e Rinnooy Kan (1991) utilizaram o Algoritmo 1 (apresentado anteriormente para o problema $SPI.1/J/M/N$) para resolver $SPI.1/G/M/N$. O algoritmo inicia com alocação infactível (com a menor alocação possível \mathbf{m}^0 , onde cada m_j^0 é dado por (60)) e, em cada iteração, adiciona uma máquina na estação cujo quociente entre o aumento da função-objetivo e a redução de WIP é o menor possível. O algoritmo pára tão logo a adição de uma máquina torna a alocação factível. Para estabilidade do algoritmo deve-se garantir que $\rho_j < 1$ para toda estação $j, j=1, \dots, n$.

Desde que $F(\mathbf{m})$ é uma função convexa e não decrescente em \mathbf{m} e $L(\mathbf{m})$ é uma função convexa e não crescente em \mathbf{m} , então $F(m^{p-1})$ e $F(m^p)$ são limitantes para o valor da solução ótima. Ou seja, $F(m^{p-1}) < F(m^*) \leq F(m^p)$, com m^p e m^{p-1} sendo as alocações geradas pelo Algoritmo 1, e m^* sendo a alocação ótima para o problema

SP1.1/G/M/N (vide Teoremas 1 e 2 em Van Vliet e Rinnooy Kan (1991) e Boxma et al. (1990), respectivamente).

Van Vliet e Rinnooy Kan (1991) aplicaram o Algoritmo 1 a dois exemplos de redes de manufatura. Obtiveram erros médios relativos para as soluções dos dois exemplos de 7,24% e 4,71%. O erro máximo em *ca* foi inferior a 6%. Este erro foi obtido recalculando *ca* para a configuração final, resolvendo o sistema linear (69). Além disso, derivaram curvas de *trade-off* entre *WIP* e custos de investimento em capacidade, executando o Algoritmo 1 para vários valores de L_T .

3.4.2.2 Modelo *SP2.1/G/M/N*

Também chamado de problema de realocação de servidores (máquinas), este modelo realoca os servidores a estações de modo a minimizar o *WIP*. O número de servidores que pode ser alocado é fixo e igual a M ($M > \sum_{j=1}^n m_j^0$). A formulação matemática para o problema *SP2.1/G/M/N* segue (semelhante à formulação para o problema *SP2.1/J/M/N*):

$$\begin{aligned} & \min L(\mathbf{m}) \\ \text{sujeito a} \quad & \sum_{j=1}^n m_j = M && \text{SP2.1/G/M/N} \\ \text{com} \quad & m_j \geq m_j^0, \quad m_j \text{ inteira} \quad j = 1, \dots, n \end{aligned}$$

Van Vliet e Rinnooy Kan (1991) utilizaram o Algoritmo 2 para resolver o problema *SP2.1/G/M/N*. O algoritmo começa com a menor alocação possível dada por (60). Em cada iteração, uma máquina é adicionada à estação onde a maior redução de *WIP* é alcançada. O procedimento é repetido até que todos os servidores ($M - \sum_{j=1}^n m_j^0$) tenham sido alocados. Para estabilidade do algoritmo deve-se garantir que $\rho_j < 1$ para toda estação $j, j=1, \dots, n$.

Desde que $L(\mathbf{m})$ é uma função convexa e não crescente em m , o algoritmo termina com a solução ótima \mathbf{m}^* (vide Teoremas 2 e 3 em Van Vliet e Rinnooy Kan (1991) e Boxma et al. (1990), respectivamente).

Van Vliet e Rinnooy Kan (1991) também geraram curvas de *trade-off* entre *WIP* e número de servidores, executando o Algoritmo 2 para vários valores de M .

Experimentos computacionais para os dois exemplos de rede também confirmaram a hipótese H2 de que ca é pouco sensível a mudanças na capacidade.

Para ambos os problemas $SP1.1/G/M/N$ e $SP2.1/G/M/N$, Van Vliet e Rinnoy Kan (1991) também utilizaram uma extensão de (29), desenvolvida por Whitt (1993) para o número esperado de *jobs* na estação j :

$$L_j(m_j, \lambda_j, ca_j, \mu_j, cs_j) = \lambda_j \Theta(\rho_j, ca_j, cs_j, m_j) \frac{(ca_j + cs_j)}{2} Lq_j(M/M/m_j) + \frac{\lambda_j}{\mu_j} \quad (76)$$

com $Lq_j(M/M/m_j)$ definido em (29) e

$$\Theta(\rho_j, ca_j, cs_j, m_j) = \begin{cases} \frac{4(ca_j - cs_j)}{4ca_j - 3cs_j} \varpi_1(m_j, \rho_j) + \frac{cs_j}{4ca_j - 3cs_j} \Psi\left(\frac{ca_j + cs_j}{2}, m_j, \rho_j\right), & \text{se } ca_j \geq cs_j \\ \frac{cs_j - ca_j}{2(ca_j + cs_j)} \varpi_1(m_j, \rho_j) + \frac{cs_j + 3ca_j}{2(ca_j + cs_j)} \Psi\left(\frac{ca_j + cs_j}{2}, m_j, \rho_j\right), & \text{se } ca_j < cs_j \end{cases}$$

com

$$\Psi\left(\frac{ca_j + cs_j}{2}, m_j, \rho_j\right) = \begin{cases} 1 & \text{se } \frac{(ca_j + cs_j)}{2} \geq 1 \\ \varpi_4(m_j, \rho_j)^{2-(ca_j+cs_j)} & \text{se } \frac{(ca_j + cs_j)}{2} < 1 \end{cases}$$

e

$$\varpi_4(m_j, \rho_j) = \min\left\{1, \frac{\varpi_1(m_j, \rho_j) + \varpi_3(m_j, \rho_j)}{2}\right\}$$

$$\varpi_3(m_j, \rho_j) = \varpi_2(m_j, \rho_j) \exp\left(\frac{-2(1-\rho_j)}{3\rho_j}\right)$$

$$\varpi_2(m_j, \rho_j) = 1 - 4\eta(m_j, \rho_j)$$

$$\varpi_1(m_j, \rho_j) = 1 + \eta(m_j, \rho_j)$$

$$\eta(m_j, \rho_j) = \min\left\{0.24, (1-\rho_j)(m_j-1) \frac{(4+5m_j)^{1/2} - 2}{16m_j\rho_j}\right\}$$

O multiplicador $\Theta(\rho_j, ca_j, cs_j, m_j)$ em (76) é usado como filtro para situações em que a aproximação (29) não é efetiva. Entretanto para os dois exemplos computacionais, os resultados obtidos pelas aproximações (29) e (76) foram muito próximos.

Estes são os modelos de otimização relacionados aos problemas $SP1$ e $SP2$, revisados neste capítulo por serem importantes para o estudo dos próximos

capítulos. O capítulo 4 apresenta e analisa os resultados computacionais das aplicações destes modelos e métodos no estudo de caso. Modelos e métodos de otimização envolvendo o problema de partição da instalação (classe *SP3*) são reportados no capítulo 5.

4 RESULTADOS COM ESTUDO DE CASO

Este capítulo analisa a aplicação dos modelos e métodos de solução abordados nos capítulos 2 e 3 em um estudo de caso. Particularmente, os resultados mostram a aplicabilidade e a adequação dos modelos de avaliação de desempenho e de alocação ótima de capacidade discreta, discutidos nos capítulos anteriores, em uma empresa com sistema de manufatura *job-shop* e política de fabricação sob pedido (*make-to-order*). Por este trabalho tratar de problemas tático-estratégicos, cujas decisões possuem um horizonte de tempo de médio a longo prazo, a primeira dificuldade foi selecionar uma empresa que se dispusesse a colaborar com o experimento e fornecer informações. A seguir, um breve histórico da empresa que se dispôs a colaborar com a pesquisa.

4.1 Empresa Estudada

De origem familiar, a empresa *IMART-MARRARA Tornearia de Peças Ltda.*, sediada na cidade de São Carlos (SP), especializada em usinagem de peças metálicas para aplicação industrial, emprega cerca de 65 funcionários, operando 17,6 horas por dia divididos em dois turnos. Fabrica peças metálicas usinadas para aplicação industrial nos ramos eletrodoméstico (denominado no meio industrial por linha branca) e automotivo. Possui como principais clientes empresas como Multibrás S/A Eletrodomésticos, Electrolux do Brasil S/A, CCE Eletrodomésticos (ambas da linha branca) e Ideal Standard Wabco Indústria e Comércio Ltda (setor automotivo). O faturamento está entre R\$ 300 a R\$ 450 mil por mês (dados de junho/2002).

O sistema de manufatura possui algumas máquinas desempenhando funções diferentes arranjadas para compor duas células de manufatura, e as demais máquinas arranjadas fisicamente por função (disposição funcional) para compor o sistema *job-shop*. As células de manufatura visam atender demandas mensais altas de poucas classes de produtos e as máquinas correspondentes são dedicadas a produzir somente aquelas poucas classes. Cada célula produz entre 50-65 mil peças/mês.

O sistema tipo *job-shop* foi adotado pela empresa para satisfazer demandas mensais menores (pequenos lotes) para uma grande variedade de produtos.

Para alguns produtos, lotes de 50-60 peças são solicitados somente uma vez ao ano, enquanto que para outros são pedidos lotes com 400-1600 peças/mês.

Atualmente, o sistema *job-shop* satisfaz quase que exclusivamente a empresa Ideal Standard Wabco Indústria e Comércio Ltda. Os pedidos são feitos pela Wabco à Imart-Marrara através de um sistema de intercâmbio de dados eletrônicos (*EDI*) que são atualizados semanalmente. Buscando permitir planejamento e programação da produção por parte da Imart-Marrara, a Wabco disponibiliza os tamanhos dos lotes (que podem variar semanalmente) e o mês de entrega em um horizonte de tempo de aproximadamente um ano. Entretanto, as datas específicas (dia e mês) e o tamanho exato dos lotes são estabelecidos em um horizonte de tempo de dois meses. Ou seja, os dados disponibilizados via *EDI* pela Wabco para produtos que devem ser produzidos e entregues em um horizonte de dois meses possuem datas específicas e tamanho de lote fixos (geralmente não mudam semanalmente).

4.2 Representação do Sistema de Manufatura nos Modelos de Redes de Filas e de Simulação

Uma vez que a empresa foi selecionada, as principais dificuldades encontradas em aplicar os modelos foram: coletar e tratar estatisticamente os dados disponíveis na empresa de modo a utilizá-los nos modelos e validar os resultados. A primeira atividade para modelar o estudo de caso foi representar o sistema de manufatura real em estações visitadas por classes, formadas por produtos que perfazem roteiros de fabricação similares. Em seguida, cada classe foi caracterizada por seu processo de chegadas externas e processo de serviço nas estações do roteiro de fabricação. Essas atividades envolveram este autor, o diretor industrial e o engenheiro responsável pelo Planejamento e Controle da Produção (PCP) da empresa estudada, e são detalhadas nas seções a seguir.

4.2.1 O Sistema de Manufatura como um Conjunto de Classes de Produtos, Estações de Trabalho e Roteiros de Fabricação

O sistema de manufatura *job-shop* foi modelado como uma rede com filas *GI/G/1* de múltiplas classes e roteiro de fabricação determinístico. Estações com m_j

máquinas operando em paralelo, por simplicidade, foram modeladas como uma estação com uma “super máquina” com capacidade equivalente. A razão desta simplificação é que em certas estações as máquinas não são idênticas, ou são idênticas mas não necessariamente trabalham em paralelo (por exemplo, duas máquinas e apenas um operador para operá-las ao mesmo tempo). Deste modo, todas estações da rede foram modeladas como possuindo uma única máquina ($m_j=1, j=1, \dots, n$). Convém mencionar que o *leadtime* de produção, assim como o inventário em processo na rede, são superestimados ao se admitir que toda estação da rede é modelada como uma fila $GI/G/1$, ao invés de uma fila $GI/G/m$. Ora, uma estação j com m_j servidores é mais flexível (por exemplo, em termos de manutenção) e, portanto, possui em média menor fila de espera. Entretanto, esta hipótese foi adotada com vistas a simplificar o processo de coleta de dados e parece razoável considerando-se que as decisões de capacidade a serem suportadas pelos modelos são de médio a longo prazo (tático-estratégicas).

Juntamente com o diretor industrial e o engenheiro responsável pelo PCP, o sistema *job-shop* foi dividido em 15 estações de trabalho, a saber:

- Estação 1: Máquina Serra Fita;
- Estação 2: Torno Automático;
- Estação 3: Torno CNC (controlado numericamente);
- Estação 4: Torno Mecânico;
- Estação 5: Torno Revólver;
- Estação 6: Máquinas Furadeiras;
- Estação 7: Estação de Fresas;
- Estação 8: Forno de Indução;
- Estação 9: Centro de Usinagem;
- Estação 10: Retíficas;
- Estação 11: Rosqueadeiras de Bancada;
- Estação 12: Tamboreador Vibratório;
- Estação 13: Balancim;
- Estação 14: Estação para Inspeção Intermediária ou Final e;
- Estação 15: Estação responsável pela embalagem.

Assim, para este sistema, a rede possui 15 estações de trabalho ($j=1, \dots, 15$), cada uma com uma única super-máquina e descrita por uma fila $GI/G/1$.

Operações realizadas por terceiros (tratamento superficial das peças, por exemplo) não foram consideradas como estações de trabalho da rede, embora tenha-se considerado 3 dias de operação efetiva para efeitos de estimativa de *leadtime* para aquelas classes de produtos que necessitam de serviço terceirizado. De acordo com o gerente industrial, com pequenas variações, esse é o tempo médio gasto para que um lote deixe a empresa, realize o serviço e retorne à empresa. Por isso, o tempo de serviço terceirizado foi considerado igual a 3 dias. Além disso, há várias pequenas empresas disponíveis para realizar esse serviço, e quando uma empresa está congestionada, o lote é enviado para outra.

No total, o *job-shop* produz 20 tipos de produtos diferentes. Entretanto, no período em que a amostra foi coletada (mais de 1 ano), alguns produtos sequer foram fabricados, enquanto outros foram pedidos somente uma ou duas vezes. Assim, 5 produtos (aqueles que consumiram menos de 1,5% em hora-máquina do total consumido no período e que tiveram poucos pedidos) foram desconsiderados na análise. Os demais que possuíam roteiros de fabricação similares foram agregados em uma mesma classe. Para este caso, a rede processa 14 classes de produtos ($k=1, \dots, 14$), listadas a seguir com os respectivos produtos:

Classe 1: núcleo da bobina;

Classe 2: haste de inox I e haste de inox II (são produtos com roteiros de fabricação similares, mas com especificações técnicas diferentes);

Classe 3: garfo;

Classe 4: corpo da válvula;

Classe 5: pino;

Classe 6: pistão I;

Classe 7: eixo;

Classe 8: batente;

Classe 9: adaptador I;

Classe 10: adaptador II;

Classe 11: pistão II;

Classe 12: bucha de inox;

Classe 13: luva e;

Classe 14: parafuso de regulagem.

Conforme pode-se notar, todas as classes, com exceção da número 2, são formadas por um único produto. Cada classe perfaz um roteiro de fabricação diferente. A figura 4.1 ilustra o roteiro das classes no sistema *job-shop* estudado.

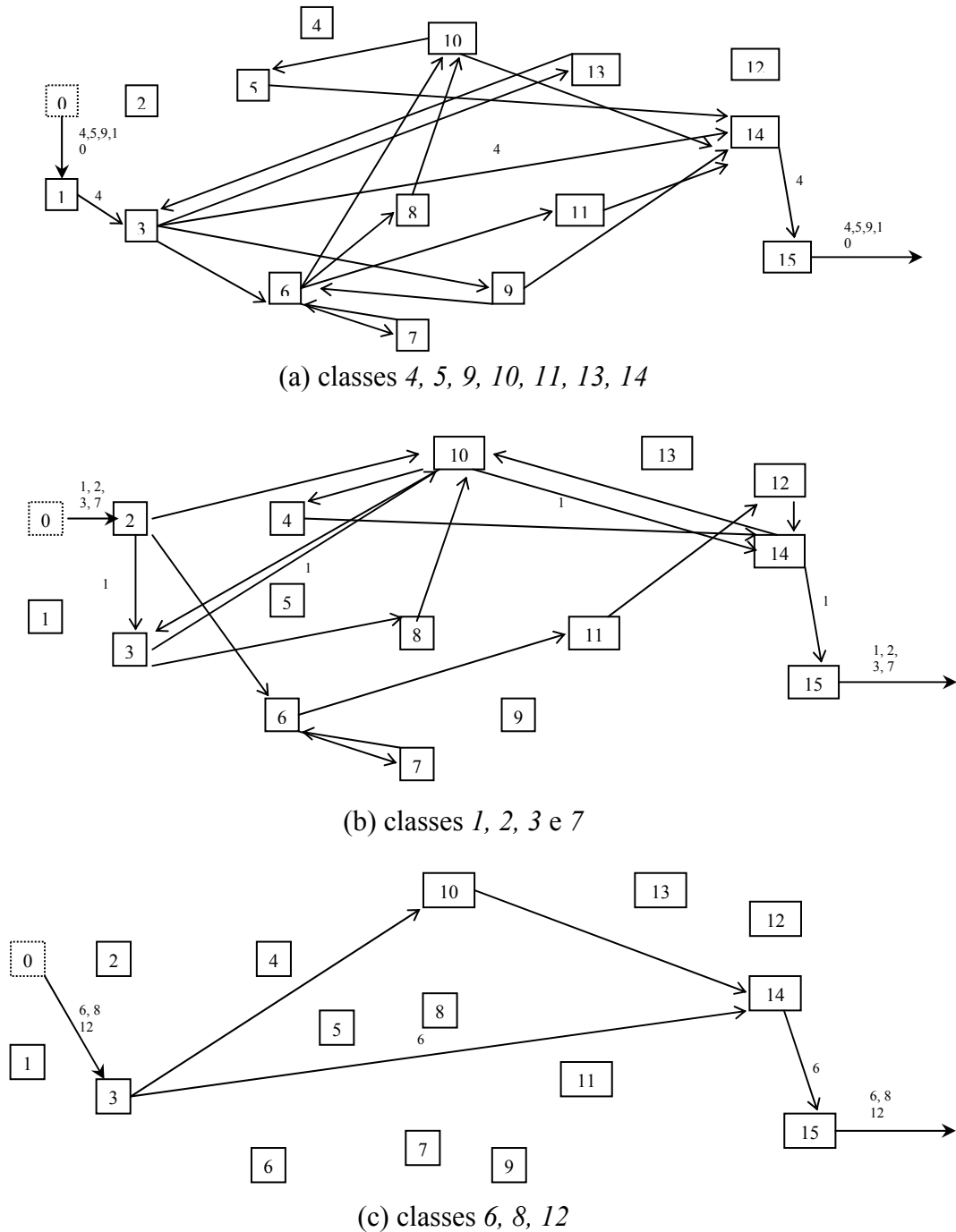


FIGURA 4.1 – Sistema *job-shop* em estudo como uma rede de filas (0 é a estação fictícia representando a chegada externa): (a) representa o roteiros das classes 4, 5, 9, 10, 11, 13 e 14; (b) representa os roteiros das classes 1, 2, 3 e 7 ; (c) representa os roteiros das classes 6, 8 e 12.

Na figura 4.1a, por exemplo, os arcos ilustram o roteiro de fabricação das classes 4, 5, 9, 10, 11, 13 e 14. Para que as figuras ficassem legíveis, identificou-se somente os arcos referentes ao roteiro da classe 4, da classe 1 e da classe 6 nas figuras 4.1a, 4.1b e 4.1c, respectivamente. Em ambas as figuras, 0 é a estação fictícia representando as chegadas externas à rede. O roteiro de cada classe também é dado na tabela 4.1, a seguir.

Para caracterizar os processos de chegada e de serviço de cada classe de produto para ambos os modelos (aproximações de rede de filas e de simulação) foi utilizada uma amostra compreendendo 14 meses (de setembro/2001 a outubro/2002). Os dados de entrada necessários foram coletados através de relatórios emitidos de uma base eletrônica disponível na empresa. A inviabilidade da coleta direta pelo autor no chão de fábrica deu-se pelo fato de boa parte das classes de produtos serem fabricadas quatro ou cinco vezes em um ano (o que as deixaria de fora em um processo de coleta de dias, semanas ou poucos meses).

4.2.2 Caracterização do Processo de Chegadas Externas

Com relação ao modelo de simulação foram considerados, para cada classe de produto k ($k=1, \dots, 14$), dados de intervalos de tempo (em horas) entre chegadas externas, nos 14 meses analisados. Testes de aderência (*goodness of fit*) foram aplicados na amostra com o objetivo de avaliar se as observações, para cada classe, são uma amostra independente de uma distribuição particular “teórica”. Os testes de aderência de Kolmogorov-Smirnov (K-S) e Anderson-Darling (A-D) foram conduzidos no *software BestFit*[®] em níveis de significância de até 15% (para mais detalhes dos testes, vide Law e Kelton, 1991). O *BestFit*[®] avalia 26 famílias de distribuições “teóricas”, tais como a família Exponencial, a Erlang, a Weibull, a Uniforme, a Normal, a Lognormal e a Qui-quadrado, dentre outras.

Como a maioria das amostras foi rejeitada nos testes de K-S e A-D para distribuições teóricas, a distribuição empírica (57), discutida na seção 2.6, foi adotada para representar as distribuições do processo de chegadas. Estas distribuições foram utilizadas pelo simulador *GPSS/H*[®] para gerar as chegadas de cada classe no modelo de simulação (vide Apêndice B). Uma causa possível para que a maioria das observações

dos intervalos de tempo entre chegadas externas de cada classe tenham sido rejeitadas nos testes de hipóteses como amostras independentes de distribuições conhecidas (“teóricas”), pode ser o fato de que das 14 classes de produtos, 13 possuem amostras com menos de 30 pontos amostrais, e 8 possuem amostras com menos de 10 pontos amostrais. Como ilustração, o intervalo de tempo entre chegadas para a classe de produtos 10, que possui apenas 4 pontos amostrais, possui a distribuição empírica:

$$F(x) = \begin{cases} 0, & \text{se } x < 457,6 \\ \frac{x-457,6}{3(809,6-457,6)}, & \text{se } 457,6 \leq x \leq 809,6 \\ \frac{1}{3} + \frac{x-809,6}{3(950,4-809,6)}, & \text{se } 809,6 \leq x \leq 950,4 \\ \frac{2}{3} + \frac{x-950,4}{3(1689,6-950,4)}, & \text{se } 950,4 \leq x \leq 1689,6 \\ 1, & \text{se } x > 1689,6 \end{cases}$$

Com relação ao modelo de rede de filas, para cada classe k ($k=1, \dots, 14$), a taxa média de chegada externa λ'_k e o coeficiente quadrático de variação do intervalo de tempo entre chegadas externas ca'_k (definidos como $\lambda'_k = 1/E(a'_k)$ e $ca'_k = V(a'_k)/(E(a'_k))^2$, em que $V(a'_k)$ e $E(a'_k)$ são a variância e a média do intervalo de tempo entre chegadas externas da classe k , respectivamente) foram estimados a partir da amostra.

Por falta de informações mais precisas, para ambos os modelos (rede de filas e de simulação), o instante de chegada de um lote de fabricação de uma determinada classe de produto na primeira estação da sua rota foi considerado como o instante (dia) em que, segundo relatórios, foi aberta uma ordem de fabricação (*OF*) para aquela classe de produto. Obviamente, à abertura de uma *OF* antecedem regras de programação que são aplicadas pelo PCP da empresa, tais como a *Earliest Due Date* (*EDD*). Desta forma, o processo de chegada é o resultado desta programação e não das chegadas externas de pedidos propriamente dita. Mesmo assim, a hipótese de considerar a chegada de um produto como o instante em que é aberta uma *OF* parece ser bastante razoável para caracterizar o processo de chegada no sistema *job-shop*, ao considerar que todas as classes possuem um atraso (*time lag*) aproximadamente constante entre o instante de abertura da *OF* e o início da fabricação. Deste modo, considerou-se que as

chegadas externas de classes de produtos ocorrem nos dias úteis em que a manufatura opera. Portanto, uma classe de produto com ordem de fabricação aberta em uma segunda-feira, que possui outra ordem aberta na segunda-feira da semana seguinte, possui intervalo de tempo entre chegadas externas (em horas) igual a $17,6 * 5 = 88$ horas. Isto ocorre se o *job-shop* opera de segunda a sexta-feira normalmente (por exemplo, sem feriados), com 17,6 horas de operação efetiva por dia, que corresponde à quantidade de horas nominalmente trabalhadas pela empresa.

Entretanto, ao considerar cada dia como sendo constituído de 17,6 horas, supõe-se que todas estações estejam disponíveis e operem efetivamente 17,6 horas por dia. Para os modelos, quando uma estação está desocupada e ocorre a chegada de um lote de produtos nela, o serviço inicia-se instantaneamente. Esta situação nem sempre ocorre na prática, pois pode acontecer da estação estar desocupada e o lote ter de aguardar por falta de operador, quebra de máquina ou falta de energia. Em outras palavras, ao considerar que a estação opera efetivamente 17,6 horas por dia, superestima-se a capacidade da estação, pois são desconsiderados os fatores que ocasionam perda de capacidade. Por exemplo, a falta de um operador (que pode faltar ou estar operando em outra estação), quebra/manutenção de uma máquina, etc.

Uma maneira de considerar a perda de capacidade é a redução das horas efetivamente trabalhadas. Em outras palavras, a redução do número de horas que compõem o dia efetivamente trabalhado. De acordo com o gerente industrial da empresa, cerca de 5% das horas disponíveis por dia não são efetivamente trabalhadas, ou seja, o dia efetivamente trabalhado possui $17,6 * (1 - 0,05) = 16,7$ horas. Uma maneira mais precisa de considerar tais perdas de capacidade é a criação de classes de produtos fictícias (Morabito, 1998) com roteiros probabilísticos que podem visitar todas as estações de trabalho que compõem o *job-shop*. Infelizmente, a empresa estudada não dispunha de dados dos instantes e durações de quebras de máquinas, faltas de operador, energia, etc., em cada estação, durante o período de coleta de dados, para que esta alternativa pudesse ser considerada nas aproximações analíticas e na simulação. Portanto, neste trabalho, a perda de capacidade foi considerada simplesmente pela redução das horas efetivamente trabalhadas para 16,7 horas/dia.

Ainda, para ambos os modelos, a chegada de cada lote de tamanho variável (vide média e desvio-padrão do tamanho dos lotes, entre colchetes, na terceira

coluna da tabela 4.1) foi considerada como se fosse de um único produto (*batch arrival*). Assim, uma *OF* aberta representa a chegada de um produto/item. Nota-se pelas classes 8 e 2 que o tamanho dos lotes é bastante variável. O roteiro de fabricação foi descrito pelas estações visitadas para cada operação de cada classe. Os dados de entrada caracterizando os processos de chegadas externas e o roteiro de fabricação para cada classe são mostrados na tabela 4.1.

TABELA 4.1 – Tamanho da amostra, tamanho médio e desvio padrão do lote (entre colchetes), $E(a'_k)$, λ'_k , ca'_k , n_k e estações visitadas para operação l da rota da classe k , n_{kl} .

Classe	Tamanho da Amostra	Tamanho do lote (itens)	$E(a'_k)$ (h)	λ'_k (lote/h)	ca'_k	n_k	n_{kl}
1	13	493,6 [182,1]	280,03	0,003571	0,3818	5	2, 3, 10, 14, 15
2	36	896,0 [603,3]	131,29	0,007617	0,3377	8	2, 10, 3, 8, 10, 4, 14, 15
3	18	1583,7 [702,9]	257,00	0,003891	0,2569	9	2, 6, 7, 6, 11, 12, 14, 14, 15
4	9	621,0 [213,9]	473,26	0,002113	0,5092	4	1, 3, 14, 15
5	8	468,2 [236,1]	436,30	0,002292	0,2853	8	1, 3, 9, 6, 10, 5, 14, 15
6	5	1083,2 [104,1]	517,60	0,001932	0,4048	4	3, 14, 14, 15
7	8	900,1 [332,6]	538,50	0,001857	0,2939	7	2, 10, 14, 14, 10, 14, 15
8	5	4801,5 [3395,9]	671,14	0,001490	0,2934	5	3, 14, 14, 14, 15
9	13	640,6 [229,1]	357,14	0,002800	0,4131	8	1, 3, 9, 6, 8, 10, 14, 15
10	4	876,6 [238,6]	926,78	0,001079	0,2817	5	1, 3, 14, 14, 15
11	15	495,7 [327,0]	289,44	0,003455	0,4351	5	1, 3, 9, 14, 15
12	13	1069,2 [590,2]	349,41	0,002862	0,3607	4	3, 10, 14, 15
13	5	877,2 [499,4]	697,84	0,001433	0,3037	9	1, 3, 6, 7, 6, 11, 14, 14, 15
14	5	1108,5 [380,0]	738,01	0,001355	0,3528	7	1, 3, 13, 3, 9, 14, 15

Note que os ca'_k são bem diferentes de 1, ou seja, os processos de chegadas externas são bem diferentes do processo de Poisson, caracterizando uma rede de filas genérica. Observe, por exemplo, que o roteiro da classe 6 é definido por 4 operações ($n_6=4$) em que $n_{61}=3$, $n_{62}=14$, $n_{63}=14$ e $n_{64}=15$ é a seqüência de estações visitadas, com duas operações consecutivas diferentes na estação 14.

4.2.3 Caracterização do Processo de Serviço

Como para o processo de chegadas, para caracterizar o processo de serviço para ambos os modelos, foram coletadas amostras no mesmo período de 14 meses, via relatórios da base eletrônica, dos tempos de serviço para cada classe k em cada operação l , s_{kl} , do seu roteiro.

As amostras coletadas de tempos de serviço para cada classe k e uma determinada operação l foram alimentadas no *software BestFit*[®] para que os testes de aderência fossem realizados. Como para os dados de chegada externa, a maioria das amostras foi rejeitada com níveis de significância de até 15% nos testes de K-S e A-D para 26 famílias de distribuições teóricas. As observações foram então utilizadas para definir uma distribuição empírica que descrevesse o processo de serviço da operação l no roteiro da classe k no modelo de simulação, conforme ilustra o Apêndice B.

A partir dessas amostras, a taxa média de serviço e o coeficiente quadrático de variação do tempo de serviço da operação l do roteiro da classe k , respectivamente $\mu_{kl} = 1/E(s_{kl})$ e $cs_{kl} = V(s_{kl})/(E(s_{kl}))^2$, foram determinados para os lotes processados, para servirem como dados de entrada dos modelos de redes de filas. Conforme mencionado anteriormente, admitiu-se que cada *OF* aberta com lote de tamanho variável corresponde à chegada de um único produto e, portanto, o tempo de serviço em cada estação corresponde ao tempo de processamento do lote todo. A tabela 4.2 mostra os tempos médios de serviço para cada classe e operação da rede.

TABELA 4.2 – Tempos médios de serviço de um lote na operação l do roteiro da classe k , $E(s_{kl})$.

Classe	Operação (h/lote)								
	1	2	3	4	5	6	7	8	9
1	5,761	27,831	4,773	5,507	2,346	-	-	-	-
2	19,197	8,664	27,989	12,488	31,368	14,848	6,971	12,639	-
3	26,935	33,828	35,662	47,808	9,139	4,744	4,519	3,247	2,557
4	0,718	17,596	1,770	2,197	-	-	-	-	-
5	1,080	22,038	18,353	14,022	7,604	6,580	8,919	2,700	-
6	42,332	3,968	3,028	2,558	-	-	-	-	-
7	14,606	18,966	3,293	8,052	32,346	4,937	6,638	-	-
8	38,146	3,891	2,033	6,490	3,103	-	-	-	-
9	23,900	30,973	28,480	25,128	6,399	9,095	6,281	3,620	-
10	54,625	41,868	4,530	7,158	4,320	-	-	-	-
11	1,261	11,395	21,002	5,216	1,820	-	-	-	-
12	24,265	8,565	3,731	2,676	-	-	-	-	-
13	3,207	19,688	10,119	17,693	10,978	4,761	2,623	3,399	1,600
14	0,566	24,156	13,566	19,469	48,550	6,274	6,881	-	-

É importante mencionar que ao considerar o tempo de processamento do lote todo, supõe-se que o processo de serviço ocorre em lote (*batch service*), ou seja, o lote é processado completamente na estação em que está sendo servido, sem interrupções que permitam a estação servir outro lote de classe de produto diferente. Esta política é, sempre que possível, mantida pela programação da produção para tirar vantagem do tempo de preparação da máquina/estação, conforme declarações do gerente industrial e do engenheiro responsável pelo PCP.

A base de dados da empresa não continha os tempos de processamento dos lotes da classe k na operação l , s_{kl} , mas apenas os tempos de processamento de cada item da classe k em cada operação l . Como as variâncias dos tempos de processamento de cada item são muito pequenos, utilizou-se a média destes tempos unitários (t_{kl}) para estimar os tempos de processamento dos lotes da seguinte maneira: o tempo de processamento s_{kl} de cada lote da classe k em cada operação l foi estimado multiplicando-se o tempo médio unitário, t_{kl} , pelo tamanho do lote (vide a média e o desvio-padrão do tamanho dos lotes na tabela 4.1). Ou seja, admitiu-se que o tempo de processamento do lote, s_{kl} , é proporcional ao tamanho do lote, y_k . Desta forma, a variabilidade dos tempos de serviço, cs_{kl} , deve-se somente à variação dos tamanhos dos lotes (cy_k), conforme é mostrado a seguir:

$$cs_{kl} = \frac{V(s_{kl})}{E(s_{kl})^2} = \frac{V(t_{kl}y_k)}{E(t_{kl}y_k)^2} = \frac{t_{kl}^2 V(y_k)}{t_{kl}^2 E(y_k)^2} = cy_k$$

Para os produtos estudados, os coeficientes quadráticos de variação para o tempo de serviço de cada classe ($cs_{kl} = cy_k$) foram de 0,1361, 0,4534, 0,1970, 0,1186, 0,2544, 0,0092, 0,1366, 0,5002, 0,1250, 0,0741, 0,7535, 0,3047, 0,3242 e 0,1175, respectivamente, para as classes de 1 a 14. Ou seja, variaram de 0,0092 a 0,5002 e, portanto, bem diferentes de 1, ou seja, não são exponencialmente distribuídos e deve-se utilizar um modelo de rede de filas genérica.

Algumas classes de produtos tais como a 3, 6, 7, 8, 10 e 13 realizam operações tais como tratamento superficial, que são terceirizadas. A figura 4.2 ilustra em detalhes o roteiro de fabricação da classe 6, juntamente com o serviço terceirizado (ST).

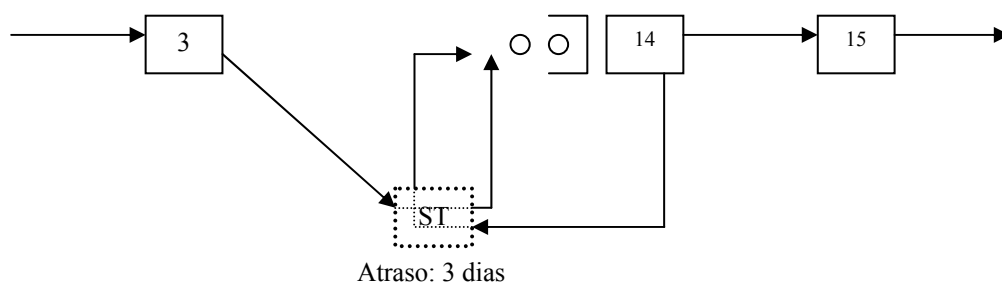


FIGURA 4.2 – Roteiro de fabricação da classe de produtos 6: 3, ST, 14, ST, 14, 15.

Conforme pode-se perceber da figura 4.2, a classe de produto 6 realiza sua primeira operação na estação 3, sendo em seguida encaminhada para realização de serviço terceirizado (ST), onde é feito o tratamento superficial. Após 3 dias (ou seja, $3 \cdot 16,7 = 50,1$ horas), o produto é devolvido para a fábrica e encaminhado para o fim da fila da estação 14, para ser inspecionado. Após inspeção na estação 14, o produto é enviado para outro serviço terceirizado (ST), que pode inclusive empregar outro ou o mesmo prestador de serviços, retornando à fábrica após 3 dias (ou 50,1 horas) no fim da fila da estação 14, para inspeção final. A partir daí, segue para embalagem na estação 15, para então aguardar a expedição.

De acordo com o diretor industrial, todo prestador de serviço terceirizado possui prazo de 3 dias para pegar o produto, realizar a operação e devolvê-lo à fábrica. Ainda segundo o diretor industrial, raramente este prazo não é cumprido. Por este motivo, foi considerado – somente para efeito de estimativa de *leadtime* para aquelas

classes que realizam operações terceirizadas – um atraso (*time lag*) determinístico de 50,1 horas para cada visita a terceiros que a classe de produtos realiza.

É conveniente esclarecer que para o modelo, o roteiro de fabricação considerado para a classe 6 foi a seqüência de estações: 3, 14, 14, 15 (vide tabela 4.1). Ou seja, não se considera no roteiro as visitas no serviço terceirizado, e a seqüência é tratada como retroalimentação imediata na estação 14 (*immediate feedback*, conforme Whitt, 1983a), visto que, após realização do serviço na estação 3, o produto segue para a estação 14, retorna novamente à fila da estação 14 e, posteriormente, segue para a estação 15. No entanto, o que acontece na prática é retroalimentação após atraso de 3 dias (ou 50,1 horas), que só é considerado na estimativa do *leadtime* das classes de produtos. Entretanto, tratar no roteiro como retroalimentação imediata a retroalimentação com atraso de 3 dias parece ser uma simplificação razoável, dado que a análise do sistema é feita em equilíbrio (*steady-state*) e espera-se que o estado da fila em um certo instante seja, em média, semelhante ao seu estado após 3 dias.

Convém ressaltar que não se deve confundir a retroalimentação de um roteiro determinístico com a de um roteiro probabilístico, que traz imprecisões na rede e deve ser eliminada (Whitt, 1983a; Morabito, 1998).

Essas foram as hipóteses admitidas e os dados coletados para implementar os modelos de redes de filas e de simulação e avaliar o desempenho do sistema de manufatura.

4.3 Custos e Alternativas Discretas para Alocação de Capacidade

Dentre os problemas de alocação/realocação de capacidade estudados, o modelo referente à alocação de capacidade discreta para uma rede genérica com uma única máquina (“super-máquina”) em cada estação ($SP1.1/G/S/R^{discreto}$ ou $SP1.2/G/S/R^{discreto}$) mostrou-se o mais adequado para aplicação. Primeiro, as máquinas e mão-de-obra pertencentes ao *job-shop* estudado são distintas e especializadas quando se muda de uma estação para outra, o que inviabiliza o intercâmbio de capacidade entre estações em um problema de realocação de capacidade. Segundo, adições de máquinas e/ou mão-de-obra (funcionários) são exemplos de adições de capacidades que podem ser modeladas como variáveis discretas.

Assim, o primeiro passo para implementação do modelo $SP1.1/G/S/R^{discreto}$ no algoritmo 5, discutido na seção 3.4.1.3, foi a estimativa do custo de um lote de produto qualquer na estação j , v_j (conforme (66)). Para cada classe de produto k e operação l foram levantados os custos com matéria-prima e operação. Por exemplo, sabe-se que cada item da classe 14 , gasta \$0,12 com matéria-prima (dados de 2002). Como cada lote da classe 14 possui, em média, 1108,5 itens (vide tabela 4.1), então o custo com matéria-prima para produzir um lote de tamanho médio foi considerado $1108,5 * 0,12 = \$133,73/\text{lote}$. Já o custo relacionado à operação l ($l=1$) para a classe 14 foi estimado da seguinte maneira: o custo operacional por hora (\$/hora) na estação em que ocorre a primeira operação da classe 14 ($n_{14,1}=1$, vide tabela 4.1) foi estimado e multiplicado pelo inverso da taxa de produção de um lote (hora/lote) da classe 14 na estação. Deste modo, o valor obtido para o custo operacional foi de \$13,13. Logo, o custo acumulado da classe 14 com a operação 1 , $C_{14,1}$, foi de \$146,86/lote (\$13,13 + \$133,73). Da mesma forma, o custo operacional para a classe 14 na estação que realiza a operação 2 ($n_{14,2}=3$, vide tabela 4.1) foi estimado em \$972,04/lote, resultando em um custo acumulado da classe 14 com a 2ª operação do roteiro de fabricação, $C_{14,2}$, de \$1118,90/lote (\$146,86 + \$972,04). Este procedimento foi aplicado para cada classe k e operação l do roteiro de fabricação. Note que C_{kl} representa o custo acumulado operacional e com matéria-prima de um lote médio da classe k com a realização da l -ésima operação.

Entretanto, para estimativa do capital de trabalho associado ao inventário em processo, de acordo com a expressão (66), deve-se estimar o custo médio de um lote de produto de uma classe qualquer na estação j , v_j . Para cada estação j ($j=1, \dots, n$), foi computado o custo médio acumulado de todas as classes de produtos que utilizam a estação j , ponderado pela proporção de chegada dos produtos à estação, ou seja:

$$v_j = \frac{\sum_{k=1}^r \sum_{l=1}^{n_k} C_{kl} \lambda'_k 1\{n_{kl} = j\}}{\sum_{k=1}^r \sum_{l=1}^{n_k} \lambda'_k}, \quad j=1, \dots, n$$

com

$1\{n_{kl}=j\}$ representando uma função indicadora com valor igual a 1 se a operação l da classe k é realizada na estação j ; e 0, em caso contrário;

λ'_k a taxa média de chegadas externas de lotes da classe k .

A tabela 4.3 mostra os custos acumulados, v_j , para um lote de uma classe de produto qualquer na estação j da rede (dados de 2002).

TABELA 4.3 – Custo acumulado nas estações de um lote de produto qualquer para a rede estudada.

	Custo (\$/lote)							
Estação	1	2	3	4	5	6	7	8
Custo	799,0	2955,4	3369,9	10698,5	2825,0	5008,5	4361,2	6847,1
Estação	9	10	11	12	13	14	15	
Custo	2991,9	5888,5	7678,1	9664,2	1496,3	5526,9	5850,1	

Em conjunto com o gerente industrial e o engenheiro encarregado do PCP, foram consideradas as alternativas para adição ou remoção de capacidade nas estações da tabela 4.4 (dados de 2002).

TABELA 4.4 – Alternativas de adição/remoção de capacidade nas estações j : taxa de produção (μ_{jt}) e custo (f_{jt}) na estação j e alternativa t .

Estação	Alternativa 1		Alternativa 2		Alternativa 3	
	μ_{j1} (lote/h)	f_{j1} (\$1000)	μ_{j2} (lote/h)	f_{j2} (\$1000)	μ_{j3} (lote/h)	f_{j3} (\$1000)
1	0,104	0,0	0,208	9,0	0,104	0,0
2	0,057	0,0	0,066	22,0	0,047	-12,0
3	0,038	0,0	0,042	190,0	0,038	0,0
4	0,067	0,0	0,084	12,0	0,052	-7,0
5	0,152	0,0	0,228	6,0	0,076	-3,0
6	0,035	0,0	0,042	0,5	0,028	-0,3
7	0,032	0,0	0,037	40,0	0,027	-20,0
8	0,092	0,0	0,138	120,0	0,046	-60,0
9	0,038	0,0	0,057	130,0	0,019	-60,0
10	0,063	0,0	0,076	225,0	0,050	-110,0
11	0,126	0,0	0,252	2,0	0,126	0,0
12	0,211	0,0	0,422	2,0	0,211	0,0
13	0,074	0,0	0,148	0,5	0,074	0,0
14	0,199	0,0	0,199	0,0	0,199	0,0
15	0,201	0,0	0,201	0,0	0,201	0,0

A alternativa 1 corresponde à configuração inicial e, por isso, o custo com adição de capacidade é nulo para todas estações. A alternativa 2 considera a possibilidade de se adicionar uma máquina, exceto na estações 14 e 15 porque correspondem às atividades de inspeção e embalagem, cujas tarefas são executadas por funcionários e não são de maior interesse da empresa estudada. A alternativa 3

considera a hipótese de remover uma máquina naquelas estações em que a taxa de utilização continua sendo menor que 1 ($\rho_j < 1$), que é a condição de equilíbrio da rede. Os custos negativos significam a remoção. Os dados de custos de máquinas foram pesquisados junto aos fornecedores dos equipamentos. Para o caso de remoção, foram considerados os custos de equipamentos obtidos dos fornecedores, levando-se em conta a depreciação do equipamento usado, uma vez que ao considerar a possibilidade de a empresa remover a máquina, ela possa vendê-la a preço de mercado.

Esses foram os dados coletados para implementação dos modelos de otimização.

4.4 Avaliação do Desempenho da Configuração Inicial da Rede

Uma vez que o sistema de manufatura foi representado em um conjunto de estações, classes de produtos, roteiros de fabricação, e os processos de chegadas externas e de serviço foram caracterizados por meio da coleta e tratamento dos dados necessários, o próximo passo foi a implementação computacional dos modelos de avaliação de desempenho e as respectivas soluções.

Objetivou-se, inicialmente, determinar as aproximações que melhor descrevem o desempenho do sistema de manufatura, dentre as tratadas no capítulo 2 e resumidas no fim da seção 2.5.4. Assim, foram examinados os seguintes sistemas lineares de equações de variabilidade de tráfego (aproximações):

Sistema linear 1 – aproximações (15), (27) e (24);

Sistema linear 2 – aproximações (15), (27) e (49);

Sistema linear 3 – aproximações (15), (27) e (54);

Sistema linear 4 – aproximações (18), (27) e (24);

Sistema linear 5 – aproximações (50), (27) e (49);

Sistema linear 6 – aproximações (18), (27) e (54).

O modelo de simulação foi utilizado com vistas a validar os sistemas de aproximações. A tabela 4.5 compara os *leadtimes* estimados pelo modelo de simulação com *leadtimes* médios obtidos a partir de amostras, cujos tamanhos eram maior ou igual a 10. Note na tabela 4.5 que os *leadtimes* médios obtidos a partir das amostras mudam com mudança na capacidade efetiva. Isto ocorre porque as informações de *leadtimes* de

produção estavam disponíveis em dias e tiveram de ser convertidos para horas, e ao mudar de capacidade efetiva de 100% para 95% da nominal, um dia, que correspondia a 17,6 horas, equivale a 16,7 horas, respectivamente. Conforme pode-se perceber, ao considerar a capacidade efetiva igual a 95% da nominal, o desvio médio absoluto entre os *leadtimes* obtidos por simulação e amostra real é reduzido de 12,7% para 7,7%, o que é aceitável para os níveis de decisões que se pretende apoiar. Ainda, esse resultado confirma a opinião do gerente industrial de que cerca de 5% da capacidade diária em horas é perdida com manutenção ou equipamento ocioso por falta de funcionário.

TABELA 4.5 – Comparação entre os *leadtimes* médios amostrais e obtidos por simulação considerando capacidade efetiva igual a 100% e 95% da nominal para classes com tamanho de amostra maior ou igual a 10.

Classe	Tamanho Amostra	Capacidade efetiva = 100% nominal			Capacidade efetiva = 95% nominal		
		Simulação (horas)	Amostra (horas)	Desvio (%)	Simulação (horas)	Amostra (horas)	Desvio (%)
1	10	109,0	133,8	-18,5	128,6	126,9	-1,3
2	78	192,7	189,1	1,9	212,8	179,4	18,6
3	21	241,0	274,9	-12,3	239,9	271,0	-11,5
9	17	200,0	204,0	-2,0	219,3	193,5	13,3
11	15	103,0	130,2	-20,9	122,4	123,6	-0,9
12	13	98,1	123,2	-20,4	117,2	116,9	0,3
Média dos Desvios Absolutos (%)			12,7%	7,7%			

Os dados das tabelas 4.1, 4.2 e cs_{kl} ou cy_k ($k=1,\dots,r$; $l=1,\dots,n_k$) foram utilizados como entradas para utilizar aproximações para uma rede de filas $GI/G/1$, com múltiplas classes e roteiro determinístico. A tabela 4.6 mostra os dados obtidos após o procedimento de agregação dado pelas expressões (35), (36), (38), (39) e (40).

TABELA 4.6 – Dados para a rede com classe única (classe agregada).

Estação	λ_{0j} (lote/h)	$E(a_{0j})$ (h)	ca_{0j}	μ_j (lote/h)	$E(s_j)$ (h)	cs_j
1	0,015	66,67	0,962	0,104	9,61	2,948
2	0,017	58,82	0,874	0,057	17,64	0,527
3	0,006	166,67	0,431	0,038	25,99	0,366
4	-	-	-	0,067	14,85	0,453
5	-	-	-	0,152	6,58	0,254
6	-	-	-	0,035	28,61	0,471
7	-	-	-	0,032	30,83	0,288
8	-	-	-	0,092	10,85	0,513
9	-	-	-	0,038	26,27	0,359
10	-	-	-	0,063	15,91	1,009
11	-	-	-	0,126	7,96	0,280
12	-	-	-	0,211	4,74	0,197
13	-	-	-	0,074	13,57	0,117
14	-	-	-	0,199	5,03	0,444
15	-	-	-	0,201	4,97	1,319

Da segunda à quarta colunas da tabela 4.6 nota-se que nenhum produto realiza a primeira operação nas estações 4 a 15. Pode-se notar na última coluna também que algumas estações apresentaram variabilidade maior do que 1 para os tempos de serviço, ao se agregar todas as classes em uma única, como foi o caso das estações 1, 10 e 15. Os dados da tabela 4.6 foram utilizados como entradas para o primeiro passo do método de decomposição. A taxa é calculada exatamente por meio da solução do sistema linear dado por (1). Com vistas a avaliar as diversas aproximações, os sistemas de equações lineares 1 a 6, citados anteriormente, foram resolvidos com o auxílio do *software GAMS/CPLEX*[®] para calcular aproximadamente o parâmetro de variabilidade. Os sistemas lineares 1, 3, 4 e 6 são compostos de 255 variáveis e 255 equações lineares. Já os sistemas 2 e 5 são compostos de 156 variáveis e 156 equações. Para todos os sistemas, o *GAMS/CPLEX*[®] levou menos de um segundo para executar o programa em um microcomputador Pentium III, com processador de 800 MHz e 128MB RAM. A rapidez com que os sistemas são resolvidos reforça a utilidade e potencial de aplicação das aproximações.

A tabela 4.7 reporta ρ_j , λ_j , $E(a_j)$, ca_j , μ_j , $E(s_j)$ e cs_j obtidos pelas aproximações do sistema 2, cujos resultados (em termos do número de lotes e *leadtime*) estiveram mais próximos dos resultados obtidos por meio de simulação, conforme é

visto nas tabelas 4.8 e 4.9. Na coluna 2 da tabela 4.7 fica evidente que a estação 3 é a mais utilizada dentre as estações ($\rho_3=0,867$). As colunas 5 e 8 evidenciam que a hipótese admitida para tratar o sistema de manufatura como uma rede de Jackson superestima a variabilidade, visto que os processos de chegada e de serviço nas estações possuem, em sua maioria, coeficientes quadráticos de variação menores que 1 (quando obtidos pelas aproximações). As soluções dos sistemas lineares foram utilizadas em (20) e (21) - Passo 2 – permitindo a estimativa do número de lotes nas estações da rede (tabela 4.8) e do *leadtime* (tabela 4.9) para cada classe de produto por (41).

TABELA 4.7 – Resultados para a classe agregada nas estações j utilizando as aproximações do sistema 2.

Estação	ρ_j	λ_j (lote/h)	$E(a_j)$ (h)	ca_j	μ_j (lote/h)	$E(s_j)$ (h)	cs_j
1	0,140	0,015	68,837	0,386	0,104	9,608	2,949
2	0,299	0,017	59,046	0,324	0,057	17,638	0,527
3	0,867	0,033	29,981	0,488	0,037	25,992	0,365
4	0,113	0,008	131,285	0,760	0,067	14,848	0,453
5	0,015	0,002	436,300	0,639	0,152	6,580	0,254
6	0,450	0,016	63,532	0,535	0,035	28,613	0,471
7	0,164	0,005	187,829	0,524	0,032	30,825	0,288
8	0,113	0,010	95,997	0,651	0,092	10,851	0,513
9	0,260	0,010	100,990	0,515	0,038	26,273	0,430
10	0,485	0,030	32,816	0,568	0,063	15,905	1,009
11	0,042	0,004	257,003	0,650	0,126	7,961	0,197
12	0,018	0,005	187,829	0,719	0,211	4,744	0,231
13	0,018	0,001	738,007	0,438	0,074	13,566	0,117
14	0,265	0,053	18,948	0,593	0,199	5,030	0,466
15	0,188	0,038	26,492	0,640	0,201	4,972	1,322

Os resultados de simulação foram obtidos a partir de 10 réplicas aceitas pelo teste de uniformidade de qui-quadrado realizado pelo *GPSS/H*[®]. Foram gerados 200000 *jobs* para cada réplica, dos quais os 100000 primeiros foram descartados com vistas a evitar os efeitos do estado transiente, e os 100000 restantes foram utilizados para gerar as estatísticas, admitidas em estado estacionário. O modelo de simulação consumiu poucos segundos para ser executado no mesmo microcomputador. O número médio de lotes para cada classe na rede e de todas as classes e os desvios com relação aos resultados obtidos por simulação constam na tabela 4.8. Nesta tabela, as colunas

$E(L_k)^0, \dots, E(L_k)^6$ representam o número médio de itens de cada classe k obtidos pelos modelos de simulação e pelas aproximações de 1 a 6, respectivamente.

TABELA 4.8 – Número médio de lotes, $E(L_k)$, obtidos por simulação e pelas aproximações.

Classe	Número médio de lotes na rede ($E(L_k)$, em lotes)						
	$E(L_k)^0$	$E(L_k)^1$	$E(L_k)^2$	$E(L_k)^3$	$E(L_k)^4$	$E(L_k)^5$	$E(L_k)^6$
1	0,477	0,623	0,465	0,502	0,631	0,577	0,554
2	1,709	2,123	1,754	1,817	2,146	2,031	1,975
3	0,948	1,023	0,946	0,937	1,032	1,029	1,018
4	0,217	0,283	0,202	0,222	0,287	0,256	0,243
5	0,405	0,518	0,407	0,419	0,525	0,498	0,471
6	0,314	0,409	0,337	0,352	0,413	0,384	0,373
7	0,304	0,331	0,306	0,306	0,333	0,332	0,329
8	0,344	0,395	0,338	0,349	0,397	0,376	0,367
9	0,671	0,781	0,646	0,660	0,791	0,758	0,725
10	0,241	0,297	0,255	0,265	0,299	0,283	0,277
11	0,457	0,534	0,402	0,432	0,540	0,502	0,467
12	0,341	0,465	0,349	0,368	0,471	0,428	0,409
13	0,356	0,391	0,314	0,320	0,396	0,374	0,362
14	0,363	0,460	0,359	0,383	0,464	0,430	0,408
$E(L)$	7,147	8,633	7,082	7,333	8,723	8,257	7,980
Desvio	0%	20,8%	-0,9%	2,6%	22,1%	15,5%	11,7%

Nas 10 réplicas do modelo de simulação foram obtidos, em média, 7,147 lotes na rede (tabela 4.8), com desvio padrão amostral de 0,066 e intervalo de 95% de confiança [7,193, 7,099], construído a partir da distribuição *t-Student*, conforme recomenda Schriber (1991). Percebe-se da tabela 4.8 que os valores obtidos pelas aproximações não foram muito discrepantes do obtido em simulação. O maior desvio foi de 22,1%, enquanto os resultados gerados pelos sistemas 2, 3 e 6 desviaram, respectivamente, em -0,9%, 2,6% e 11,7% do obtido em simulação. Morabito (1998) realizou experimentos computacionais com uma rede genérica com 10 classes de produtos e 13 estações de trabalho *GI/G/I* representando a manufatura de uma fábrica de semicondutores. Ao utilizar os sistemas de aproximações 1 a 6 na rede estudada por Morabito (1998), os três menores desvios relativos à simulação para o número médio de *jobs* na rede foram 2,58%, 8,36% e -10,94% para os sistemas 6, 2 e 3, respectivamente. Ou seja, as mesmas aproximações que estiveram mais próximas dos valores obtidos por simulação para a rede tratada neste trabalho foram também as que melhor descreveram o comportamento da rede abordada em Morabito (1998).

Na tabela 4.9, as colunas $E(T_k)^0, \dots, E(T_k)^6$ representam o *leadtime* médio de itens de cada classe k obtidos pelos modelos de simulação e pelas aproximações de 1 a 6, respectivamente. O número entre parênteses representa o desvio do *leadtime* obtido pelas aproximações dos sistemas 1 a 6, com relação ao *leadtime* resultante do modelo de simulação. Os sistemas 2 e 3 foram os que apresentaram menores desvios com relação à simulação (desvios relativos médios absolutos de 4,14% e 6,72% para os sistemas 2 e 3, respectivamente). Note que, assim como para a estimativa do número de lotes na rede, o sistema de aproximações 2 foi o que apresentou desvio relativo do *leadtime* médio da rede ($E(T)$, com $E(T) = \sum_{k=1}^r \lambda'_k E(T_k) / \sum_{k=1}^r \lambda'_k$) mais próximo do obtido por simulação, ou seja, 3,1%. A tabela 4.9 mostra o desvio-padrão dos *leadtimes* dos lotes para cada classe (estimado como a raiz quadrada de (42)), empregando as aproximações de 1 a 6 ($DP(T_k)^1, \dots, DP(T_k)^6$ representam os desvios-padrões obtidos pelas aproximações de 1 a 6, respectivamente). A ordem de grandeza significativa do desvio padrão justifica-se pelo fato dos tamanhos de lotes de cada classe ser bastante variável (vide tabela 4.1).

TABELA 4.9 – *Leadtime* médio obtido por simulação e aproximações e desvios relativos dos *leadtimes* obtidos pelas aproximações com relação ao obtido por simulação.

Classe	<i>Leadtime</i> médio ($E(T_k)$, em horas/ote) e desvios relativos (%) estimados com relação a simulação						
	$E(T_k)^0$	$E(T_k)^1$	$E(T_k)^2$	$E(T_k)^3$	$E(T_k)^4$	$E(T_k)^5$	$E(T_k)^6$
1	128,6 (0%)	174,5 (35,7%)	130,2 (1,2%)	140,6 (9,3%)	176,6 (37,3%)	161,7 (25,7%)	155,2 (20,7%)
2	212,8 (0%)	278,8 (31,0%)	230,3 (8,4%)	238,6 (12,1%)	281,8 (32,4%)	266,6 (25,3%)	259,3 (21,9%)
3	239,9 (0%)	263,0 (9,6%)	243,0 (1,3%)	240,9 (0,4%)	265,2 (10,5%)	264,5 (10,3%)	261,7 (9,1%)
4	99,6 (0%)	133,9 (34,4%)	95,6 (-4,0%)	105,1 (5,5%)	135,6 (36,1%)	120,9 (21,4%)	115,1 (15,6%)
5	179,0 (0%)	225,8 (26,1%)	177,8 (-0,7%)	183,0 (2,2%)	228,9 (27,9%)	217,2 (21,3%)	205,6 (14,9%)
6	163,8 (0%)	211,9 (29,4%)	174,4 (-6,4%)	182,1 (11,2%)	213,6 (30,4%)	198,9 (21,4%)	193,0 (17,8%)
7	163,3 (0%)	178,2 (9,12%)	165,0 (1,0%)	164,5 (0,7%)	179,2 (9,7%)	178,7 (9,4%)	177,4 (8,6%)

8	215,1 (0%)	265,1 (23,2%)	227,0 (5,5%)	234,3 (8,9%)	266,8 (24,0%)	252,1 (17,2%)	246,1 (14,4%)
9	219,3 (0%)	279,0 (27,2%)	230,8 (5,2%)	235,6 (7,4%)	282,4 (28,8%)	270,8 (23,5%)	259,0 (18,1%)
10	215,0 (0%)	275,6 (28,2%)	236,7 (10,1%)	245,7 (14,3%)	277,2 (28,9%)	262,5 (22,1%)	256,7 (19,4%)
11	122,4 (0%)	154,6 (26,3%)	116,5 (-4,8%)	125,1 (2,2%)	156,3 (27,7%)	145,3 (18,7%)	135,3 (10,5%)
12	117,2 (0%)	162,3 (38,5%)	121,8 (3,9%)	128,5 (9,6%)	164,4 (40,3%)	149,5 (27,6%)	143,1 (22,1%)
13	231,3 (0%)	272,5 (17,8%)	219,0 (-5,3%)	223,3 (-3,5%)	276,3 (19,5%)	261,0 (12,8%)	252,4 (9,1%)
14	264,7 (0%)	339,4 (28,2%)	265,3 (0,2%)	282,8 (6,8%)	342,7 (29,5%)	317,0 (19,8%)	301,2 (13,8%)
$E(T)$	182,0 (0%)	228,7 (25,7%)	187,6 (3,1%)	194,3 (6,8%)	231,1 (27,0%)	218,8 (20,2%)	211,4 (16,2%)

TABELA 4.10 – Desvio-padrão do *leadtime* obtido para cada classe utilizando as aproximações.

Classe	Desvio-Padrão do <i>leadtime</i> , $DP(T_k)$, na rede, em horas					
	$DP(T_k)^1$	$DP(T_k)^2$	$DP(T_k)^3$	$DP(T_k)^4$	$DP(T_k)^5$	$DP(T_k)^6$
1	125,1	83,2	92,3	127,0	110,8	104,2
2	132,2	92,0	100,7	134,3	118,9	112,5
3	56,4	42,5	40,9	58,4	57,6	55,3
4	122,0	80,7	91,2	123,8	107,2	100,6
5	129,4	86,7	95,3	131,6	116,7	109,0
6	121,5	80,3	90,4	123,3	106,5	100,0
7	41,8	31,7	31,0	42,9	42,3	41,0
8	124,5	84,9	94,4	126,3	110,0	103,7
9	129,8	87,3	95,8	132,1	117,2	109,5
10	123,3	82,6	92,9	125,2	108,7	102,3
11	123,1	82,3	92,5	124,9	109,3	101,9
12	124,8	83,7	92,9	126,8	110,5	103,9
13	130,6	86,4	94,9	133,2	117,5	110,4
14	173,2	115,5	129,8	175,8	153,0	143,1

A tabela 4.11 compara estimativas do número de lotes para cada classe na rede, ao supô-la determinística ($ca'_k = cs_{kl} = 0$, $k=1, \dots, r$; $l=1, \dots, n_k$), ao supor que se comporte como uma rede de Jackson ($ca'_k = cs_{kl} = 1$, $k=1, \dots, r$; $l=1, \dots, n_k$) e ao considerá-la genérica, conforme sistema de aproximação 2 da tabela 4.8.

TABELA 4.11 – Número médio de lotes para cada classe de produto, ao supor que a rede é determinística ($E(L_k)^D$), Jackson ($E(L_k)^7$), ou Genérica ($E(L_k)^2$, sistema 2) com os respectivos desvios percentuais relativos à simulação ($E(L_k)^0$).

Classe	Simulação $E(L_k)^0$	Rede Determinística $E(L_k)^D$	Rede de Jackson $E(L_k)^7$	Rede Genérica $E(L_k)^2$
1	0,477	0,248	0,960	0,465
2	1,709	1,232	2,914	1,754
3	0,948	0,879	1,162	0,946
4	0,217	0,088	0,464	0,202
5	0,405	0,248	0,781	0,407
6	0,314	0,234	0,571	0,337
7	0,304	0,275	0,372	0,306
8	0,344	0,258	0,521	0,338
9	0,671	0,451	1,105	0,646
10	0,241	0,196	0,391	0,255
11	0,457	0,208	0,859	0,402
12	0,341	0,178	0,724	0,349
13	0,356	0,215	0,557	0,314
14	0,363	0,213	0,698	0,359
Total $E(L)$	7,147	4,922	12,080	7,082
Desvio	0%	-31,1%	69,0%	-0,9%

Ao comparar os resultados obtidos pelos modelos analíticos com os gerados pela simulação, fica claro pela tabela 4.11 que ao supor a rede determinística (sem incertezas nos processos de chegada e de serviço), o modelo subestima o número de lotes, como era esperado. Por outro lado, ao considerar a rede como uma de Jackson, o número de lotes é superestimado, o que também era esperado. Já o uso do sistema de aproximações 2 é o que fornece número de lotes mais próximos dos obtidos por simulação. Similarmente à tabela 4.11, a tabela 4.12 compara os resultados obtidos para o *leadtime* médio e os desvios relativos com relação à simulação.

TABELA 4.12 – *Leadtime* médio para cada classe de produto, ao supor que a rede é determinística ($E(T_k)^D$), Jackson ($E(T_k)^7$), ou Genérica ($E(T_k)^2$, sistema 2) com os respectivos desvios percentuais relativos à simulação ($E(T_k)^0$).

Classe	Simulação $E(T_k)^0$	Rede Determinística $E(T_k)^D$	Rede de Jackson $E(T_k)^7$	Rede Genérica $E(T_k)^2$
1	128,6 (0%)	69,3 (-46,1%)	268,9 (109%)	130,2 (1,2%)

2	212,8 (0%)	161,7 (-24,0%)	382,5 (79,7%)	230,3 (8,4%)
3	239,9 (0%)	225,8 (-5,9%)	298,7 (24,5%)	243,0 (1,3%)
4	99,6 (0%)	41,6 (-58,2%)	219,6 (120%)	95,6 (-4,0%)
5	179,0 (0%)	108,3 (-39,5%)	340,7 (90,3%)	177,8 (-0,7%)
6	163,8 (0%)	121,0 (-26,1%)	295,8 (80,6%)	174,4 (-6,4%)
7	163,3 (0%)	148,1 (-9,3%)	200,3 (22,7%)	165,0 (1,0%)
8	215,1 (0%)	173,0 (-19,6%)	349,8 (62,6%)	227,0 (5,5%)
9	219,3 (0%)	160,9 (-26,6%)	394,7 (80,0%)	230,8 (5,2%)
10	215,0 (0%)	182,0 (-15,3%)	362,0 (68,4%)	236,7 (10,1%)
11	122,4 (0%)	60,1 (-50,9%)	248,6 (103%)	116,5 (-4,8%)
12	117,2 (0%)	62,4 (-46,8%)	253,1 (116%)	121,8 (3,9%)
13	231,3 (0%)	150,4 (-35,0%)	388,5 (68,0%)	219,0 (-5,3%)
14	264,7 (0%)	157,3 (-40,6%)	515,3 (94,7%)	265,3 (0,2%)
<i>E(T)</i>	182,0 (0%)	130,4 (-28,4%)	320,0 (78,5%)	187,6 (3,1%)

Como esperado, novamente, ao supor a rede determinística, os *leadtimes* médios para as classes foram subestimados (*leadtime* médio da rede com desvio de -28,4% do obtido por simulação), enquanto que ao supô-la como uma rede de Jackson, os *leadtimes* médios foram superestimados (*leadtime* médio da rede com desvio de 78,5% do obtido por simulação). Por outro lado, o uso do sistema de aproximações 2 é o que fornece *leadtimes* das classes mais próximos dos obtidos por simulação (*leadtime* médio da rede com desvio de 3,1% do obtido por simulação). Os resultados das tabelas 4.11 e 4.12 mostram que admitir que uma rede se comporte como determinística ou como uma rede de Jackson pode acarretar imprecisões significativas nas medidas de desempenho de redes genéricas. Daí, a importância das aproximações para estimar o desempenho dessas redes.

Convém ressaltar que ao apresentar os resultados das tabelas 4.5, 4.8 e 4.9 ao engenheiro responsável pelo PCP da empresa, ele mencionou que os resultados apresentados pelos modelos de simulação para o *leadtime* estão próximos do que ocorre na empresa. Ele acrescentou ainda que diferenças de 10% na estimativa do *leadtime* pelas aproximações com relação à simulação (como ocorre com o sistema 2 para a classe 10, tabela 4.9) são admissíveis do ponto de vista prático, dado que 10% de desvio representa cerca de um dia, o que seria admissível em caso de atraso na entrega para a empresa ao qual fornecem. Além disso, o engenheiro não pôde afirmar a respeito do *WIP* estimado já que o controle desta variável pela empresa é visual, sem registros que permitem comparação. Por estes motivos, o sistema linear de aproximações 2 é utilizado a seguir para avaliar o desempenho da rede em diferentes configurações, ao empregar os modelos de otimização para alocação ótima de capacidade.

4.4.1 Avaliação da Sensibilidade das Aproximações a Alterações nas Taxas de Chegadas

Com vistas a analisar a sensibilidade do modelo a alterações nos parâmetros de entrada, as taxas de chegadas externas (lotes/hora) para as classes de 1 a 14 foram reduzidas para 0,00325, 0,00693, 0,00354, 0,00192, 0,00208, 0,00176, 0,00169, 0,00136, 0,00254, 0,00098, 0,00314, 0,00260, 0,00130 e 0,00123, respectivamente, no modelo de redes de filas (compare com as taxas de chegadas da tabela 4.1). Ou seja, uma redução de 10%, com relação ao caso base. Por isso, no modelo de simulação, o intervalo de tempo entre chegadas externas para todas classes de produtos foi aumentado em 10% para que correspondessem às taxas de chegadas externas citadas acima. Os resultados são mostrados nas tabelas 4.13 e 4.14. Nas colunas destas tabelas, os índices de 0 a 6 representam as medidas de desempenho obtidas por simulação e pelos sistemas de aproximações 1 a 6, respectivamente.

Conforme pode-se perceber das tabelas 4.13 e 4.14, as aproximações 2, 3 e 6 foram as que apresentaram menores desvios de número de lotes e de *leadtime* na rede com relação à simulação. Ao reduzir as taxas de chegadas externas das classes em 10%, os desvios obtidos pelos sistemas lineares de aproximações 2, 3 e 6 para o número

médio de lotes na rede foram de -4,5%, -1,7% e 8,1%, respectivamente, isto é, parecem ser pouco sensíveis e razoáveis para apoiar decisões tático-estratégicas.

TABELA 4.13 – Número médio de lotes, $E(L_k)$, obtidos da simulação e das aproximações, ao reduzir em 10% as taxas médias de chegadas de produtos.

Classe	Número médio de lotes na rede ($E(L_k)$, em lotes)						
	$E(L_k)^0$	$E(L_k)^1$	$E(L_k)^2$	$E(L_k)^3$	$E(L_k)^4$	$E(L_k)^5$	$E(L_k)^6$
1	0,342	0,411	0,314	0,339	0,417	0,400	0,383
2	1,348	1,582	1,348	1,390	1,600	1,562	1,523
3	0,859	0,909	0,844	0,839	0,915	0,913	0,906
4	0,145	0,170	0,123	0,136	0,173	0,163	0,155
5	0,312	0,367	0,296	0,303	0,372	0,367	0,348
6	0,246	0,293	0,251	0,260	0,296	0,287	0,278
7	0,272	0,291	0,271	0,271	0,292	0,292	0,290
8	0,278	0,298	0,265	0,271	0,300	0,293	0,287
9	0,536	0,583	0,496	0,505	0,591	0,584	0,560
10	0,198	0,226	0,201	0,207	0,227	0,222	0,218
11	0,324	0,344	0,265	0,285	0,349	0,342	0,317
12	0,239	0,299	0,230	0,242	0,304	0,291	0,277
13	0,285	0,289	0,239	0,243	0,294	0,286	0,278
14	0,266	0,307	0,249	0,265	0,311	0,302	0,287
$E(L)$	5,650	6,369	5,392	5,556	6,441	6,304	6,107
Desvio	0%	12,7%	-4,5%	-1,7%	14,0%	11,6%	8,1%

TABELA 4.14 – *Leadtime* médio obtido por simulação e aproximações, ao reduzir em 10% as taxas médias de chegadas dos produtos.

Classe	<i>Leadtime</i> médio ($E(T_k)$, em horas/lote) e desvios relativos (%) estimados com relação a simulação						
	$E(T_k)^0$	$E(T_k)^1$	$E(T_k)^2$	$E(T_k)^3$	$E(T_k)^4$	$E(T_k)^5$	$E(T_k)^6$
1	101,6 (0%)	126,5 (24,5%)	96,6 (-4,5%)	104,5 (2,9%)	128,4 (26,4%)	123,2 (21,3%)	118,1 (16,2%)
2	184,5 (0%)	228,4 (23,8%)	194,6 (5,5%)	200,7 (8,8%)	231,0 (25,2%)	225,5 (22,2%)	219,9 (19,2%)
3	237,9 (0%)	257,0 (8,0%)	238,6 (0,3%)	237,1 (-0,3%)	258,7 (8,7%)	258,1 (8,5%)	256,1 (7,7%)
4	73,6 (0%)	88,6 (20,4%)	64,0 (-13,0%)	70,9 (-3,7%)	90,2 (22,6%)	85,1 (15,6%)	80,5 (9,4%)
5	151,2 (0%)	176,0 (16,4%)	142,1 (-6,0%)	145,6 (-3,7%)	178,7 (18,2%)	175,9 (16,3%)	166,8 (10,3%)
6	140,8 (0%)	166,7 (18,4%)	143,0 (1,6%)	148,2 (5,3%)	168,3 (19,5%)	163,2 (15,9%)	158,6 (12,6%)
7	160,1 (0%)	172,4 (7,7%)	160,5 (0,2%)	160,3 (0,1%)	173,2 (8,2%)	172,8 (7,9%)	171,8 (7,3%)

8	191,9 (0%)	219,7 (14,5%)	195,4 (1,8%)	200,3 (4,4%)	221,3 (15,3%)	216,2 (12,7%)	211,6 (10,3%)
9	192,0 (0%)	229,2 (19,4%)	195,1 (1,6%)	198,3 (3,3%)	232,1 (20,9%)	229,3 (19,4%)	220,1 (14,6%)
10	194,3 (0%)	230,0 (18,4%)	205,0 (5,5%)	211,5 (8,9%)	231,6 (19,2%)	226,6 (16,6%)	221,9 (14,2%)
11	95,3 (0%)	109,4 (14,8%)	84,4 (-11,4%)	90,8 (-4,7%)	111,1 (16,6%)	108,8 (14,2%)	100,8 (5,8%)
12	90,4 (0%)	115,0 (27,2%)	88,6 (-2,0%)	93,0 (2,9%)	116,9 (29,3%)	111,7 (23,6%)	106,6 (17,9%)
13	202,7 (0%)	222,1 (9,6%)	183,5 (-9,5%)	186,2 (-8,1%)	225,3 (11,1%)	219,6 (8,3%)	213,1 (5,1%)
14	214,1 (0%)	249,5 (16,5%)	202,2 (-5,6%)	214,9 (0,4%)	252,7 (18,0%)	245,3 (14,6%)	232,8 (8,7%)
$E(T)$	158,1 (0%)	185,6 (17,4%)	157,1 (-0,6%)	161,9 (2,4%)	187,7 (18,7%)	183,7 (16,2%)	177,9 (12,6%)

Quando a produção total da rede é reduzida em 10%, mantendo a capacidade constante, o nível médio de utilização das estações diminui. O sistema linear 2 (formado pelas aproximações (15), (27) e (49)) foi o que melhor descreveu o desempenho do sistema em termos de *leadtime* e número médio de lotes na rede sob níveis de utilização mais altos, ou seja, sem a redução da taxa de produção em 10%, conforme pode ser visto nas tabelas 4.8 e 4.9. Altos níveis de utilização é uma característica esperada para um sistema com configuração prescrita por um modelo de otimização, cujo objetivo é minimizar o investimento em capacidade, e satisfazer certas restrições de desempenho. Ainda, o sistema 2 apresentou pouca sensibilidade a variações nas taxas de chegada (de -0,9% para -4,5% no número médio de lotes na rede, $E(L)$, e de 3,1% para -0,9% para o *leadtime* $E(T)$), ao reduzir a taxa de produção da rede em 10%) e desvios relativamente baixos para as decisões tático-estratégicas aqui apoiadas. As aproximações do sistema 2 consideram a dependência dos intervalos de tempos entre chegadas nas estações, hipótese suposta pela aproximação (15). Além disso, o efeito da interferência entre classes para o caso de uma rede processando múltiplas classes com roteiros determinísticos – cuja contribuição de cada classe no nível de utilização total da estação é pequena – é uma hipótese admitida por (49) e que satisfaz as condições da rede estudada. Por esses motivos, o sistema 2 é utilizado para avaliar o desempenho nos modelos de otimização utilizados no estudo de caso.

4.4.2 Sensibilidade das Medidas de Desempenho da Rede a Variações nas Disciplinas de Atendimento nas Estações.

Os resultados apresentados nas seções anteriores, tanto para os modelos de simulação quanto para os modelos de redes de filas genéricos, admitiram disciplina de atendimento em cada uma das estações como sendo *FCFS* (*first-come, first-served*). Uma questão interessante a ser respondida é: para a rede estudada, como se comportam as medidas de desempenho, *WIP* e *leadtime*, se a disciplina de atendimento para a fila de cada estação mudar, por exemplo, para *SPT* (*shortest processing time*)?

Com o objetivo de responder a esta questão, dois modelos de simulação foram construídos no *GPSS/H*[®]: um com disciplina *FCFS* e outro com disciplina *SPT*. É importante ressaltar que na disciplina *FCFS* outros fatores, como prazos de entregas (*EDD, earliest due date*) e prioridades de produção, tais como cliente preferencial, nível de serviço ao usuário e importância do produto no faturamento, já estão sendo levados em consideração uma vez que a chegada de produtos na rede corresponde ao instante de tempo em que a ordem de fabricação (OF) é enviada à produção e o produto começa a ser produzido. A tabela 4.15 reporta o *leadtime* e o *WIP* obtidos pelos dois modelos de simulação. Novamente, os resultados de simulação foram obtidos a partir de 10 réplicas aceitas pelo teste de uniformidade de qui-quadrado realizado pelo *GPSS/H*[®]. Foram gerados 200000 *jobs* para cada réplica, dos quais os 100000 primeiros foram descartados com vistas a evitar os efeitos do estado transiente e os 100000 restantes foram utilizados para gerar as estatísticas, consideradas de estado estacionário. O modelo de simulação consumiu cerca de 80 segundos para ser executado em um microcomputador Pentium III, 800 MHz e 128 MB RAM.

TABELA 4.15 – Número médio de lotes, $E(L_k)$, e *leadtime* médio, $E(T_k)$, para cada classe na rede com disciplinas de atendimento *FCFS* e *SPT*.

Classe de Produto	$E(L_k)$ – lote		Classe de Produto	$E(T_k)$ – hora	
	<i>FCFS</i>	<i>SPT</i>		<i>FCFS</i>	<i>SPT</i>
Classe 1	0,477	0,316	Classe 1	128,6	84,8
Classe 2	1,709	1,531	Classe 2	212,8	191,2
Classe 3	0,948	0,957	Classe 3	239,9	242,1
Classe 4	0,217	0,096	Classe 4	99,6	42,9
Classe 5	0,405	0,278	Classe 5	179,0	123,0
Classe 6	0,314	0,653	Classe 6	163,8	340,0
Classe 7	0,304	0,308	Classe 7	163,3	166,2

Classe 8	0,344	0,412	Classe 8	215,1	257,3
Classe 9	0,671	0,708	Classe 9	219,3	230,8
Classe 10	0,241	0,340	Classe 10	215,0	304,3
Classe 11	0,458	0,238	Classe 11	122,4	63,6
Classe 12	0,341	0,204	Classe 12	117,2	70,2
Classe 13	0,356	0,266	Classe 13	231,3	172,9
Classe 14	0,363	0,227	Classe 14	264,7	165,6
$E(L)$ (lote)	7,147	6,533	$E(T)$ (hora)	182,0	166,6
Desvio %	$100(FCFS-SPT)/(FCFS)$	8,59	Desvio %	$100(FCFS-SPT)/(FCFS)$	8,46

Conforme pode-se perceber da tabela 4.15, em termos de *leadtime*, a regra *SPT* favorece algumas classes de produtos (por exemplo, as classes 1, 2, 4, 5, 11, 12, 13 e 14), mas prejudica outras (por exemplo, as classes 3, 6, 7, 8, 9 e 10). Por este motivo, fica difícil fazer comentários mais genéricos sobre qual disciplina seria melhor para as classes de produtos produzidas na rede, mas de uma forma geral a disciplina *FCFS* apresenta um aumento relativo de 8,46% no *leadtime* das classes com relação à disciplina *SPT*. Ao analisar o número de lotes na rede, a disciplina *FCFS* resultou em 7,147 lotes, enquanto a *SPT* resultou num número menor 6,533 lotes, como era esperado. Em termos financeiros, o *WIP* total da rede utilizando a regra *FCFS* corresponde a \$27033,04, enquanto que para a regra *SPT* corresponde a \$24789,52. A redução relativa, correspondendo a 8,59% é razoavelmente significativa e estimula pesquisas futuras em modelos analíticos considerando disciplina *SPT* em redes de filas.

4.5 Alocação de Capacidade Ótima na Rede

Inicialmente, implementou-se em *GAMS/CPLEX*[®] o sistema linear 2 e estimou-se o *WIP* ($v_j E(L_j)$) em cada uma das estações j para a configuração inicial da rede. Observe na tabela 4.16 que o *WIP* da configuração inicial da rede é de \$26943,6. Nesta configuração, o *leadtime* médio da rede corresponde a 187,6 horas.

TABELA 4.16 – *Work-in-process (WIP)* em cada estação e total para a configuração inicial da rede.

Estação	<i>WIP</i> (\$/lote) em cada uma das estações e na rede toda							
	1	2	3	4	5	6	7	8
<i>WIP</i> (\$)	130,5	951,7	10799,4	1282,8	42,6	3035,4	737,9	807,1

Estação	9	10	11	12	13	14	15	Total
WIP (\$)	849,9	4799,6	326,3	178,5	27,5	1671,4	1303,0	26943,7

Em seguida, o algoritmo 5 (vide seção 3.4.1.3) foi implementado no software *GAMS/CPLEX*[®] (vide Apêndice C) e resolvido para limitantes superiores de *WIP* iguais a \$19500, \$20000, \$21000, \$22000, \$23000, \$24000, \$25000, \$26000, \$27000 e \$28000. Os resultados, obtidos em poucos minutos em microcomputador, foram utilizados para gerar a curva de *trade-off* entre investimento em capacidade e inventário em processo (figura 4.3) na rede. O *leadtime* médio da rede, correspondente ao inventário em processo obtido é mostrado na figura 4.3, foi estimado e consta na figura 4.4. As figuras 4.5 e 4.6 mostram, respectivamente, o número médio de lotes ($E(L)$) e o *leadtime* médio da rede ($E(T)$) para os correspondentes limitantes superiores de *WIP* citados anteriormente.

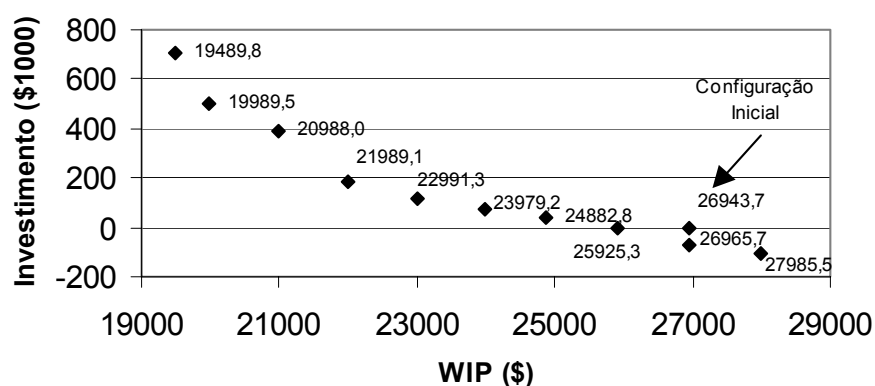


FIGURA 4.3 – Curva de *trade-off* entre investimento em máquina e *WIP*.

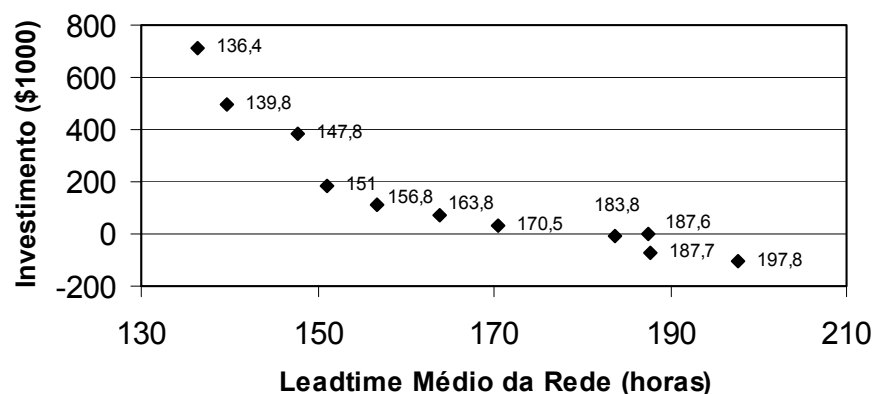


FIGURA 4.4 – *Trade-off* entre investimento em máquina e *leadtime* médio da rede.

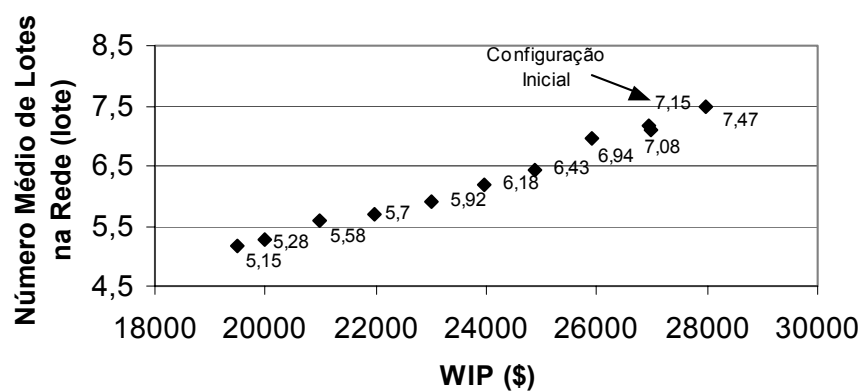


FIGURA 4.5 – Relação entre *WIP* e número médio de lotes na rede.

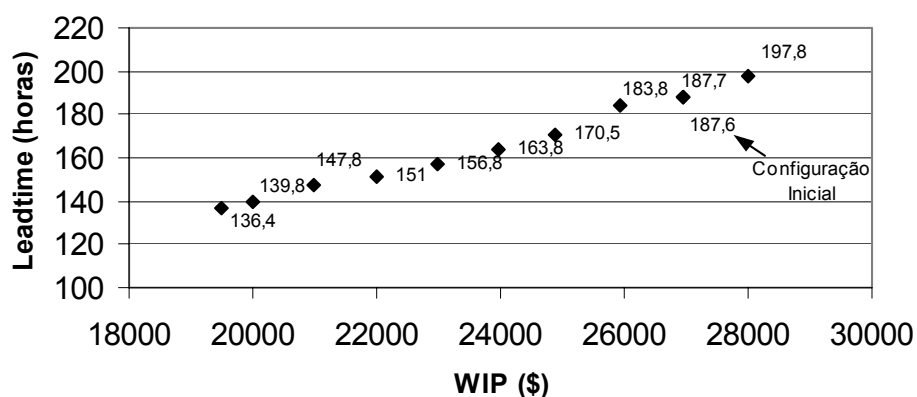


FIGURA 4.6 – Relação entre *WIP* e *leadtime* médio na rede.

Percebe-se nas figuras 4.3 e 4.4 que, à medida em que se aumenta o investimento em máquina, o *WIP* e o *leadtime* médio da rede são reduzidos, como era esperado. Além disso, por se tratar de um problema de otimização discreta, o *WIP* da solução ótima pode ser menor que o *WIP* máximo desejado (vide figura 4.3), imposto no modelo $SPI.1/G/S/R^{discreto}$ (seção 3.4.1.3). Se o problema de adição de capacidade fosse contínuo, o *WIP* da solução ótima coincidiria com o limitante superior. Note nas figuras 4.3 e 4.4 que a curva de *trade-off* pode ser não convexa.

A figura 4.3 é bastante útil na tomada de decisões estratégicas, pois para cada custo de *WIP*, a curva indica o ponto com mínimo custo de recursos. Nota-se na figura 4.3 que a configuração inicial da rede (26943,6; 0) é um ponto “fora” da curva de

Pareto e, portanto, pode ser melhorada. Por exemplo, o gestor pode migrar o estado do sistema de manufatura para o ponto (26965,7; -68,5). Ao migrar da configuração inicial para esta configuração, o *WIP* e o *leadtime* médio da rede são mantidos aproximadamente constantes (*WIP* de 26943,6 para 26965,7 e *leadtime* de 187,6 para 187,7 horas, vide figuras 4.3, 4.4 e 4.6), mas o investimento necessário em capacidade é menor. Além disso, houve pequena variação no número médio de lotes na rede (de 7,15 para 7,08, conforme figura 4.5). A tabela 4.17 mostra as alternativas ótimas para a configuração ótima com *WIP* máximo igual a \$27000.

TABELA 4.17 – Alternativa *t* escolhida pela solução do modelo $SP1.1/G/S/R^{discreto}$ para limitante superior do *WIP* igual a \$27000 e investimento de -\$68500.

Estação	1	2	3	4	5	6	7	8
Alternativa	1	1	1	2	3	2	3	3
Estação	9	10	11	12	13	14	15	
Alternativa	1	1	2	3	3	2	1	

Observa-se na tabela 4.17 que foi adicionada capacidade nas estações 4, 6, 11 e 14, mas as estações 5, 7, 8, 12 e 13 necessitaram de menos, compensando a aquisição e resultando em uma economia de \$68500 em máquinas.

Por outro lado, caso o gestor tenha interesse em reduzir o *WIP* da configuração inicial de \$26943,6 para cerca de \$26000, ele deve migrar o sistema de manufatura da configuração inicial (26943,6; 0) para o ponto (25925,3; -6,5), vide figura 4.3. Neste caso, deve-se adicionar máquinas nas estações 4, 6, 11, 12 e 14 e disponibilizar capacidade das estações 3, 7 e 13 para venda, (vide tabela 4.18), resultando em economia com capital de investimento de -\$6500. A redução do *leadtime* médio da rede foi de 3,8 horas (de 187,6 horas para 183,8 horas, conforme as figuras 4.4 e 4.6). Além disso, há redução no número médio de lotes na rede de 7,15 para 6,94 (figura 4.5).

TABELA 4.18 – Alternativa *t* escolhida pela solução do modelo $SP1.1/G/S/R^{discreto}$ para limitante superior do *WIP* igual a \$26000 e investimento de -\$6500.

Estação	1	2	3	4	5	6	7	8
Alternativa	1	1	1	2	3	2	3	1
Estação	9	10	11	12	13	14	15	
Alternativa	1	1	2	2	3	2	1	

Observe que se o gestor quiser reduzir o *WIP* da configuração inicial para menos de \$22000, ele deve investir \$180500 em capacidade (vide figura 4.3 e tabela 4.22 adiante). Ou seja, para reduzir o *WIP* em \$4954,6 (de \$26943,7 para \$21989,1) há a necessidade de investimento da ordem de grandeza bem maior, que provavelmente não seja justificado. Entretanto, a redução do *leadtime* médio da rede foi de 36,6 horas (de 187,6 para 151,0 horas, vide figura 4.6), que corresponde a 2,19 dias de operação da empresa (lembre que a empresa opera 16,7 horas por dia). Esta redução pode ser significativa do ponto de vista de atendimento dos pedidos do cliente e, portanto, a análise do ponto de vista do *leadtime* pode ser mais importante para o gestor desta empresa do que do ponto de vista do *WIP*. Em outras palavras, pode ser mais interessante para o gestor um modelo em que seja estabelecido o limitante superior para o *leadtime*, ao invés do limitante superior para o *WIP*. Por este motivo, doravante o *leadtime* médio da rede ou dos produtos também será considerado nas observações e análise.

Neste trabalho, foi feita uma adaptação do modelo $SPI.1/G/S/R^{discreto}$ para permitir que o gestor obtenha o mínimo investimento em capacidade, sujeito a restrições de limitante superior para o *leadtime* das classes de produtos. Assim, definiu-se o modelo $SPI.2/G/S/R^{discreto}$:

$$\begin{aligned} \min F(\mathbf{u}) &= \sum_{j=1}^n \sum_{t=1}^{n_j} f_{jt} u_{jt} \\ \text{sujeito a } \sum_{t=1}^{n_j} u_{jt} &= 1, \quad j = 1, \dots, n \\ \sum_{l=1}^{n_k} \sum_{j=1}^n \sum_{t=1}^{n_j} \{ (EWq_{jt} + E(s_{klt})u_{jt}) 1[n_{kl} = j] \} &\leq ET_k, \quad k = 1, \dots, r \quad (77) \\ u_{jt} &\in \{0, 1\}, \quad t = 1, \dots, n_j; \quad j = 1, \dots, n \end{aligned}$$

O lado esquerdo de (77) é a soma dos tempos de serviço e de fila para todas as operações e alternativas das estações j que pertencem ao roteiro de fabricação de uma determinada classe de produto k . Em (77), o termo $1[n_{kl}=j]$ é uma função indicadora que assume valor igual a 1, se o produto k realiza a operação l na estação j ; e valor igual a 0, em caso contrário. O lado direito de (77) representa o limitante superior para o *leadtime* da classe k , ET_k . Este modelo é útil para situações em que o gestor deve, por exemplo, reduzir o *leadtime* de alguns produtos para se manter competitivo, e necessita saber quanto e em qual(is) estação(ões) investir.

Convém observar que a expressão (77) permite que o gestor estabeleça o *leadtime* médio para cada classe de produto. Entretanto, se as k restrições (77) forem manipuladas de modo a resultar na expressão $E(L)/\lambda \leq ET$, ou especificamente

$$\sum_{k=1}^r \lambda'_k \left(\sum_{l=1}^{n_k} \sum_{j=1}^n \sum_{t=1}^{n_j} \{ (EWq_{jt} + E(s_{klt})u_{jt}) [n_{kl} = j] \} \right) / \sum_{k=1}^r \lambda'_k \leq \sum_{k=1}^r \lambda'_k E(T_k) / \sum_{k=1}^r \lambda'_k,$$

o modelo permite que o gestor especifique o limitante superior para o *leadtime* médio da rede ET , e desta forma o modelo resultante corresponde ao modelo com a restrição agregada para o WIP , discutido anteriormente.

A figura 4.7 mostra o *leadtime* das classes de produtos para a configuração inicial da rede e para duas configurações diferentes ($WIP=\$21679,0$, $F=\$180500$ e $WIP=\$24596,0$, $F=\$35500$, com F representando o custo com adição/remoção de recursos), obtidas ao estabelecer o limitante superior para o *leadtime* de cada classe na rede, $E(T_k)$, no modelo $SP1.2/G/S/R^{discreto}$. Os limitantes superiores estabelecidos para cada uma das classes nas configurações com $F=\$180500$ e $F=\$35500$ são dados na tabela 4.19.

TABELA 4.19 – Limitantes superiores para o *leadtime* de cada classe de produtos, $E(T_k)$, nas configurações com necessidade de $F=\$180500$ e $F=\$35500$ de investimento em capacidade.

Classe	Limitantes superiores para o <i>leadtime</i> das classes, $E(T_k)$, em horas	
	$F=\$35500$	$F=\$180500$
1	96,0	96,0
2	214,0	195,3
3	236,0	224,0
4	60,0	59,0
5	179,0	141,0
6	137,0	138,0
7	170,0	171,0
8	190,0	191,0
9	243,0	173,0
10	199,0	171,0
11	123,0	80,0
12	87,0	87,0
13	175,0	169,0
14	261,0	192,0

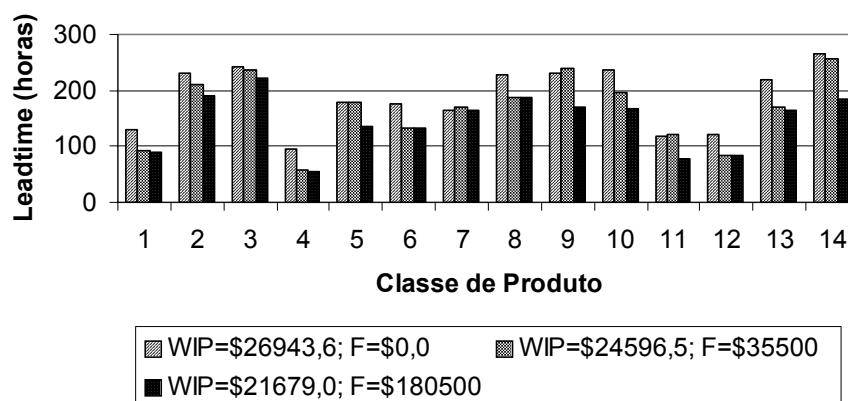


FIGURA 4.7 – *Leadtime* das classes de produtos para três configurações de rede diferentes, obtidos ao resolver o modelo $SP1.2/G/S/R^{discreto}$.

Note na figura 4.7 que, geralmente, quanto menor o investimento em recursos, maior é o *leadtime* das classes de produtos, como era de se esperar. Entretanto, as classes 7, 9 e 11 apresentam *leadtime* na configuração inicial menor que o *leadtime* quando se investe $F=\$35500$. Isto ocorre porque as estações visitadas pelas respectivas classes apresentam maior nível de ocupação e, portanto, maior é o tempo de espera nas respectivas filas das estações, apesar do investimento na rede. Conforme a tabela 4.1, a classe de produtos 7 realiza as sete operações do roteiro de fabricação nas estações 2, 10, 14, 14, 10, 14 e 15, respectivamente. De acordo com a tabela 4.20, a estação 2 apresenta maior tempo de espera, apesar do investimento na rede de $\$35500$. Já os produtos 9 e 11 também passam pelas estações 8 e 9 que, apesar do investimento na rede, apresentam maior tempo de espera em fila (vide tabela 4.20). Isto acontece porque o modelo empregado ao avaliar a melhor das três alternativas para as estações 2, 8 e 9, escolheu a alternativa 3 para ambas as estações (vide soluções ótimas para $F=\$35500$ e $F=\$180500$ na tabela 4.21), com vistas a minimizar o investimento total na rede. Conforme a tabela 4.4, esta alternativa corresponde à remoção de capacidade.

TABELA 4.20 – Tempos médios de espera nas filas das estações para a configuração inicial ($F=\$0,0$) e com investimento de $F=\$35500$.

	Tempo de Espera nas Estações (hora)							
Estação	1	2	3	4	5	6	7	8
$F=\$0,0$	1,634	1,376	70,626	0,895	0,000	9,900	0,956	0,466
$F=\$35500$	1,634	2,646	33,315	0,894	0,008	5,666	1,744	2,931
Estação	9	10	11	12	13	14	15	
$F=\$0,0$	2,718	10,849	0,023	0,002	0,000	0,721	0,932	
$F=\$35500$	23,147	10,860	0,023	0,002	0,000	0,721	0,932	

TABELA 4.21 – Alternativas ótimas do modelo $SPI.2/G/S/R^{discreto}$ para $F=\$35500$ e $F=\$180500$.

	Alternativa Ótima							
Estação	1	2	3	4	5	6	7	8
$F=\$35500$	1	3	2	1	3	2	3	3
$F=\$180500$	2	1	2	1	3	2	3	1
Estação	9	10	11	12	13	14	15	
$F=\$35500$	3	1	3	1	1	1	1	
$F=\$180500$	1	1	2	2	1	1	1	

A tabela 4.22 compara as alternativas ótimas para as estações, obtidas ao estabelecer o limitante superior para o WIP no modelo $SPI.1/G/S/R^{discreto}$ igual a $\$22000$ ($WIP=\$21989,1$ e $F=\$180500$), e ao estabelecer o limitante superior para o $leadtime$ das classes de produtos, $E(T_k)$, no modelo $SPI.2/G/S/R^{discreto}$ ($WIP=\$21679,0$ e $F=\$180500$), conforme a tabela 4.19. O objetivo é comparar as soluções ótimas obtidas pelos dois modelos ($SPI.1/G/S/R^{discreto}$ e $SPI.2/G/S/R^{discreto}$) que resultam no mesmo investimento ($F=\$180500$).

TABELA 4.22 – Alternativas ótimas para $F=\$180500$ obtidas com os modelos $SPI.1/G/S/R^{discreto}$ e $SPI.2/G/S/R^{discreto}$.

Estações	Alternativas	
	$SPI.1/G/S/R^{discreto}$	$SPI.2/G/S/R^{discreto}$
1	2	2
2	3	1
3	2	2
4	2	1
5	3	3
6	2	2
7	3	3
8	1	1

9	1	1
10	1	1
11	2	2
12	2	2
13	3	1
14	2	1
15	1	1

Compare os resultados da tabela 4.22. Com exceção das estações 2, 4, 13 e 14 todas as outras tiveram a mesma alternativa ótima e resultaram no investimento mínimo de \$180500. Ora, o limitante para o *leadtime* de cada classe imposto no modelo $SP1.2/G/S/R^{discreto}$ correspondia ao *leadtime* ótimo obtido ao impor o *WIP* máximo para a rede no modelo $SP1.1/G/S/R^{discreto}$, com um pequeno incremento (visto que trata-se de um problema inteiro) para que as soluções de ambos os modelos fossem as mesmas. Entretanto, enquanto em um modelo ($SP1.1/G/S/R^{discreto}$) é imposta uma restrição referente ao limitante superior do *WIP* da rede toda, no outro ($SP1.2/G/S/R^{discreto}$) são impostas 14 restrições referentes ao limitante superior do *leadtime* de cada classe de produto. Logo, o último modelo é mais restrito, o que gerou as diferentes alternativas para as estações 2, 4, 13 e 14. Convém mencionar que os resultados de ambos os modelos foram obtidos em poucos segundos utilizando-se um computador pessoal com processador Pentium III de 800MHz e 128MB RAM no *GAMS/CPLEX*[®]. Este fato reforça a aplicabilidade da metodologia aqui empregada.

4.5.1 Sensibilidade do Modelo a Variações nos Parâmetros de Variabilidade

Uma maneira de reduzir o investimento em capacidade é a redução da variabilidade do intervalo de tempo entre chegadas externas das classes de produtos, ca'_k . Neste trabalho, a chegada externa foi considerada como o instante em que a equipe responsável pelo PCP da empresa abre uma ordem de fabricação e o produto começa a ser produzido na primeira estação de seu roteiro de fabricação. Por isso, reduzir a variabilidade do processo de chegadas é uma estratégia possível de ser implementada, uma vez que depende de ajustes na política de programação da produção e, obviamente, de negociação de prazos e entregas com os clientes.

Dois cenários são propostos e comparados com o cenário original:

- Cenário 1.0 – as taxas de chegada, o *mix* de produção e o coeficiente quadrático de variação do intervalo de tempo entre chegadas externas são mantidos constantes, conforme a tabela 4.1 (cenário original).
- Cenário 1.1 – a taxa e o *mix* de produção são mantidos constantes, mas o coeficiente quadrático de variação do intervalo de tempo entre chegadas externas para a classe de produto 2 (ca_2') é reduzido de 0,3377, vide tabela 4.1, para 0 (determinístico).
- Cenário 1.2 – a taxa e o *mix* de produção são mantidos constantes, mas o coeficiente quadrático de variação do intervalo de tempo entre chegadas externas para todas as classes é reduzido à metade.

O investimento mínimo necessário para que o *WIP* não ultrapasse, por exemplo, o limitante superior de \$22000 foi obtido pelo modelo $SPI.1/G/S/R^{discreto}$ em \$180500 para o cenário 1.0, em \$159500 para o cenário 1.1 e em \$97500 para o cenário 1.2. Ou seja, houve economia de investimento em capacidade de \$21000 para o cenário 1.1 e \$83000 para o cenário 1.2, comparados ao cenário 1.0, que não considerou redução na variabilidade no processo de chegada. Além disso, o *WIP* ótimo foi de \$21989,1 para o cenário 1.0 (com número médio de lotes e *leadtime* médio na rede de 5,70 lotes e 151,0 horas, respectivamente, conforme figuras 4.5 e 4.6), de \$21954,2 para o cenário 1.1 (com número médio de lotes e *leadtime* médio na rede de 5,73 lotes e 151,7 horas, respectivamente) e de \$21930,2 para o cenário 1.2 (com número médio de lotes e *leadtime* médio na rede de 5,58 lotes e 147,9 horas, respectivamente). As alternativas ótimas para os cenários 1.1 e 1.2 podem ser vistas na tabela 4.23.

TABELA 4.23 – Alternativas ótimas do modelo $SPI.1/G/S/R^{discreto}$ para limitante superior do *WIP* igual a \$22000 para os cenários 1.1 e 1.2.

	Alternativas Ótimas para Estações nos Cenários 1.1 e 1.2							
Estação	1	2	3	4	5	6	7	8
Cenário 1.1	1	3	2	1	3	2	3	1
Cenário 1.2	1	3	2	1	3	2	3	3
Estação	9	10	11	12	13	14	15	
Cenário 1.1	1	1	2	2	3	2	1	
Cenário 1.2	1	1	2	3	3	2	1	

A tabela 4.24 reporta os *leadtimes* médios ótimos obtidos para as classes de produtos e para a rede ($E(T)$) nos cenários 1.0, 1.1 e 1.2.

TABELA 4.24 – *Leadtime* ótimo do modelo $SPI.1/G/S/R^{discreto}$ para cada classe de produto para limitante superior do *WIP* igual a \$22000 para os cenários 1.0, 1.1 e 1.2.

<i>Leadtime</i> Médio (hora/lote) para Classes de Produtos e para a Rede nos Cenários 1.0, 1.1 e 1.2								
Classe	1	2	3	4	5	6	7	8
Cenário 1.0	90,5	190,6	221,9	55,2	137,0	133,3	165,1	186,3
Cenário 1.1	92,4	194,4	226,7	58,5	139,4	134,7	166,5	187,7
Cenário 1.2	85,1	201,0	225,7	51,2	129,3	127,9	164,7	180,7
Classe	9	10	11	12	13	14	<i>E(T)</i> da Rede	
Cenário 1.0	169,3	166,8	76,6	82,5	165,7	185,7	151,0	
Cenário 1.1	183,0	197,0	79,9	83,3	169,9	190,2	151,7	
Cenário 1.2	181,4	189,6	71,5	76,1	159,5	175,3	147,9	

Nota-se da tabela 4.24 que, apesar do cenário 1.2 exigir menor investimento em capacidade, o *leadtime* médio da rede também é, em geral, menor que para o cenário 1.0 devido à redução à metade dos parâmetros de variabilidade.

A figura 4.8 ilustra o *trade-off* entre investimento em recursos e *WIP* para os cenários 1.0 e 1.2 a limitantes superiores de *WIP* iguais a \$20000, \$20500, \$21000, \$21500, \$22000, \$22500, \$23000, \$23500, \$24000, \$24500, \$25000, \$25500 e \$26000.

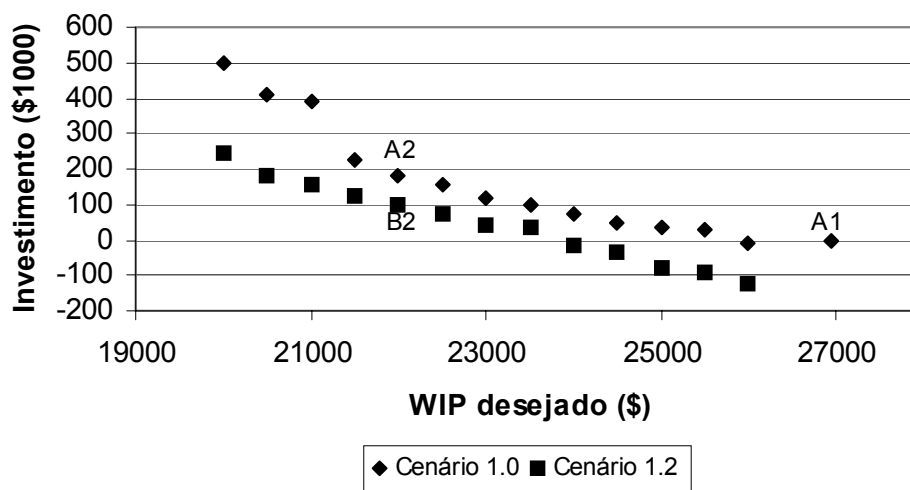


FIGURA 4.8 – *Trade-off* entre investimento em recursos e *WIP* desejado para os cenários 1.0 (quando os coeficientes são os mesmos da configuração inicial) e 1.2 (quando o coeficiente quadrático de variação do processo de chegada para todas as classes é reduzido à metade).

Percebe-se da figura 4.8 que para níveis desejados de *WIP* bem inferiores

a \$22000, a diferença de investimento necessário entre os cenários 1.0 e 1.2 é bem significativa. Por exemplo, para *WIP* desejado de \$21000, o investimento em recursos é de $F=\$387500$ para o cenário 1.0 (o *leadtime* médio da rede é 147,8 horas e o número médio de itens é 5,58 lotes) e de $F=\$158000$ para o cenário 1.2 (o *leadtime* médio da rede é 144,1 horas e o número médio de itens é 5,44 lotes). Este resultado sugere que para a rede estudada e respectivas alternativas discretas de capacidade, o investimento em capacidade é mais sensível à variabilidade dos processos de chegada quando a rede opera em tráfego mais intenso, que é o caso quando um limitante superior para *WIP* inferior a \$21000 é imposto.

O ponto A1 na figura 4.8 corresponde à configuração inicial da rede, ou seja, $F=\$0$ e $WIP=\$26943,6$. Se o gestor quiser reduzir o *WIP* para um nível inferior a \$22000, ele pode mover o estado do sistema para o ponto A2, investindo $F=\$180500$ (o número médio de lotes e o *leadtime* médio da rede para esta configuração são 5,70 lotes e 151,0 horas, respectivamente), ou mover o sistema para o ponto B2, reduzindo a variabilidade do processo de chegada das classes de produtos à metade e investindo $F=\$97500$ em capacidade (o número médio de lotes e o *leadtime* médio da rede para esta configuração são 5,58 lotes e 147,9 horas, respectivamente). Esta última solução geraria uma economia de \$83000 em aquisição de capacidade. Uma análise útil para o gestor seria avaliar se o investimento necessário para reduzir a variabilidade do processo de chegadas é compensado pela redução com investimento em capacidade.

Os resultados aqui alcançados enfatizam a sensibilidade dos modelos de investimento em capacidade estudados a parâmetros de variabilidade, particularmente o coeficiente quadrático de variação dos intervalos de tempo entre chegadas externas das classes de produtos.

4.5.2 Sensibilidade do Modelo a Variações nas Taxas de Chegadas Externas

Para avaliar a sensibilidade do modelo a variações nas taxas de chegadas, os seguintes cenários foram propostos:

- Cenário 2.0 – as taxas de chegada, o *mix* de produção e o coeficiente quadrático de variação do intervalo de tempo entre chegadas externas são mantidos constantes (os parâmetros de entrada são mantidos de acordo com a configuração inicial, vide

tabela 4.1). O limitante superior para o *WIP* foi imposto em \$25400, cujo número médio de lotes e *leadtime* médio da rede correspondem a 6,48 lotes e 171,6 horas.

- Cenário 2.1 – o *mix* de produção, o coeficiente quadrático de variação do intervalo de tempo entre chegadas externas e o limitante superior para o *WIP* são os mesmos do cenário 2.0, mas as taxas de chegadas externas das classes foram aumentadas em 10%. Ou seja, a taxa de produção da rede foi aumentada em 10%.

As alternativas ótimas para ambos os cenários podem ser vistas na tabela 4.25.

TABELA 4.25 – Alternativas ótimas do modelo $SPI.1/G/S/R^{discreto}$ para limitante superior do *WIP* igual a \$25400 para os cenários 2.0 e 2.1.

	Alternativas Ótimas para Estações nos Cenários 2.0 e 2.1							
Estação	1	2	3	4	5	6	7	8
Cenário 2.0	1	3	2	3	3	2	3	3
Cenário 2.1	2	2	2	2	3	2	2	2
Estação	9	10	11	12	13	14	15	
Cenário 2.0	3	1	3	3	3	2	1	
Cenário 2.1	2	2	2	2	3	2	1	

Nota-se na tabela 4.25 que, para a maior parte das estações, a alternativa ótima para o cenário 2.1 foi a 2, que representa a adição de capacidade (vide tabela 4.4). Convém mencionar que ao impor um *WIP* igual a \$25350 no modelo $SPI.1/G/S/R^{discreto}$ para o cenário 2.1, o problema torna-se infactível, ou seja, o modelo não consegue reduzir o *WIP* da rede para valor inferior ou igual a \$25350, para as alternativas de capacidade dadas da tabela 4.4. O investimento mínimo necessário foi de \$28500 para o cenário 2.0 e de \$749500 para o cenário 2.1 (ou seja, ao aumentar as taxas de chegadas de todos os produtos em 10%, o investimento aumenta em \$721000 para manter o *WIP* da rede menor que \$25400). Este resultado mostra o quão sensível é o modelo a variações nas taxas de chegadas externas. Os inventários em processo para os cenários 2.0 e 2.1 foram de \$25284,1 e \$25383,9, respectivamente.

A figura 4.9 ilustra o *trade-off* entre investimento em recursos e *WIP* desejados variando de \$25500 a \$30000 em intervalos de \$500 para os cenários 2.0 e 2.1.

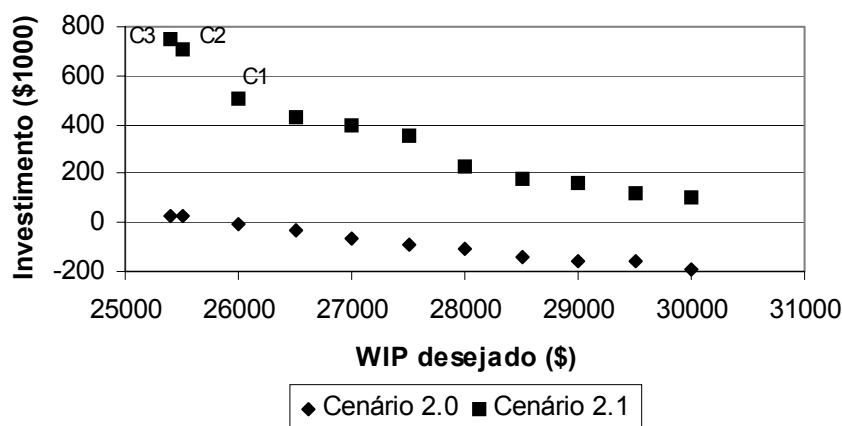


FIGURA 4.9 – *Trade-off* entre investimento em recursos e *WIP* desejado para os cenários 2.0 (quando a taxa é mantida de acordo com a configuração inicial) e 2.1 (quando a taxa de chegada total à rede é aumentada em 10%).

Nota-se pela figura 4.9 que, para a rede estudada e respectivas alternativas de investimento de capacidade consideradas, a taxa de investimento ($\Delta F/\Delta WIP$) é maior quanto maior a intensidade de tráfego. Por isso, a curva para o cenário 2.0 é mais achatada que a curva para o cenário 2.1, cujas taxas e intensidades de tráfego são maiores para uma dada capacidade. No cenário 2.1 para manter o *WIP* em torno de \$27000 (*WIP* da configuração inicial) é necessário cerca de \$400000 em investimento em capacidade (vide figura 4.9), cujo *leadtime* médio da rede correspondente é 175,9 horas. O ponto C1 corresponde a $F=\$503000$ e *WIP* desejado igual a \$26000 (número médio de lotes e *leadtime* médio da rede iguais a 6,93 lotes e 166,8 horas, respectivamente). Ao mudar o estado do sistema para o ponto C2 ($F=\$709500$ e *WIP* desejado igual a \$25500, cujo número médio de lotes e *leadtime* médio da rede são 6,82 lotes e 164,3 horas, respectivamente), o investimento aumenta em \$206500 para uma redução em *WIP* desejado de \$500. Em C2, a rede opera em condições de tráfego pesado, pois para limitantes superiores, inferiores a \$25400 (que corresponde ao ponto C3, $F=\$749500$, $E(L)= 6,80$ lotes e $E(T)= 163,6$ horas), a rede torna-se instável (pelo menos uma estação com $\rho_j \geq 1$) e o algoritmo não converge.

4.5.3 Sensibilidade do Modelo a Variações no *Mix* de Produção

Para a empresa estudada, as classes de produtos 2 e 3 são as mais

interessantes considerando a regularidade com que são requisitadas e a margem de lucro, segundo o engenheiro responsável pelo PCP. Ainda de acordo com o engenheiro, a classe 13 é pouco interessante para o PCP da empresa porque possui poucos pedidos (foram abertas 4 ordens de fabricação no período de 14 meses de amostra, conforme tabela 4.1) cujos lotes eram de tamanhos pequenos e bastante variáveis (vide tabela 4.1). Isto prejudica e dificulta a programação da produção, pois algumas vezes a formação do lote mínimo (número mínimo de itens compondo um lote que viabiliza a abertura de uma ordem de fabricação) é desrespeitada para atender às necessidades do cliente. Por esses motivos, os cenários a seguir são propostos para analisar a sensibilidade do modelo a mudanças no *mix* de produção.

- Cenário 3.0 – a taxa de chegada, o *mix* de produção e o coeficiente quadrático de variação do intervalo de tempo entre chegadas das classes são mantidos de acordo com a configuração inicial (vide tabela 4.1), e o *WIP* é imposto em \$22000.
- Cenário 3.1 – a taxa de chegada da rede, o coeficiente quadrático de variação do intervalo de tempo entre chegadas das classes e o limitante para o *WIP* são mantidos constantes, mas a classe de produtos 13 é extinta e a taxa de chegada da classe de produtos 2 é acrescida em 0,001433 lotes/hora (correspondente à taxa de chegadas externas da classe 13, para manter a taxa de chegadas na rede constante).
- Cenário 3.2 – a taxa de chegada da rede e o coeficiente quadrático de variação do intervalo de tempo entre chegadas das classes e o limitante para o *WIP* são mantidos constantes, mas a classe de produtos 13 é extinta e a taxa de chegada da classe de produtos 3 é acrescida em 0,001433 lotes/hora (correspondente à taxa de chegadas externas da classe 13, para manter a taxa de chegadas na rede constante).

As alternativas ótimas para ambos os cenários constam na tabela 4.26.

TABELA 4.26 – Alternativas ótimas para os cenários 3.0, 3.1 e 3.2 com limitante superior para o WIP de \$22000 no modelo $SP1.1/G/S/R^{discreto}$.

Alternativas Ótimas para Estações nos Cenários 3.0, 3.1 e 3.2								
Estação	1	2	3	4	5	6	7	8
Cenário 3.0	2	3	2	2	3	2	3	1
Cenário 3.1	3	1	2	2	3	2	3	1
Cenário 3.2	3	3	2	2	3	2	3	1
Estação	9	10	11	12	13	14	15	
Cenário 3.0	1	1	2	2	3	2	1	
Cenário 3.1	1	2	2	3	3	2	2	
Cenário 3.2	1	1	2	2	2	2	2	

Os custos com aquisição de capacidade e o WIP associados às soluções ótimas da tabela 4.26 para os cenários 3.0, 3.1 e 3.2 foram de $F=\$180500$ e $WIP=\$21989,1$ (com número médio de lotes, $E(L)$, e *leadtime* médio da rede, $E(T)$, iguais a 5,70 lotes e 151,0 horas); $F=\$406500$ e $WIP=21952,2$ (com $E(L) = 5,54$ lotes e $E(T) = 146,8$ horas); e $F=\$172000$ e $WIP=21993,0$ (com $E(L)= 5,65$ lotes e $E(T)= 149,6$ horas), respectivamente.

Observa-se a partir deste resultado que o investimento em capacidade para manter o WIP menor ou igual a \$22000 é muito maior para o cenário 3.1 do que para o cenário 3.2. Ou seja, ao alterar o *mix* de produção da rede, aumentando a participação da classe 2, o nível de WIP da rede aumenta mais que ao aumentar a participação da classe 3. Isto ocorre porque, de acordo com a tabela 4.1, a estação 3, que possui o maior nível de ocupação de todas estações, pertence ao roteiro de fabricação da classe 2. Assim, aumentando a taxa de chegada desta classe de produto em 0,001433 lotes/hora ao extinguir a classe 13 (cenário 3.1), o tempo médio de espera em fila na estação 3 obtido foi de 36,96 horas, enquanto que para o cenário 3.2 o tempo de espera na fila da estação 3 é de 29,58 horas. Ora, quanto maior o tempo de espera em fila, maior o inventário em processo e, portanto, maior deve ser o investimento em capacidade nas estações da rede para manter o WIP abaixo de um nível pré-determinado.

Comparando os resultados dos cenários 3.0 e 3.2, nota-se que ao aumentar a participação da classe 3 no *mix* de produção às custas da extinção da classe 13, o investimento mínimo em capacidade teve uma redução de \$8500 (as reduções em

$E(L)$ e $E(T)$ foram 0,05 lotes e 1,4 horas, respectivamente). Por outro lado, comparado ao cenário 3.0, o cenário 3.1 teve um aumento em investimento em capacidade de \$226000, embora as reduções em $E(L)$ e $E(T)$ tenham sido pouco significativas (reduções de 0,16 lotes e 4,2 horas, respectivamente). Este resultado mostra que o gestor deve realizar análise de viabilidade econômica e verificar se o investimento em capacidade é compensado pelos ganhos com o aumento da participação da classe 2 no *mix* de produção, uma vez que o investimento mostrou ser bastante sensível a aumentos no *mix* de produção da classe de produtos 2 para nível máximo de *WIP* pré-determinado em \$22000.

A figura 4.10 ilustra o *trade-off* entre investimento em capacidade e diferentes níveis de *WIP* desejado para os cenários 3.0, 3.1 e 3.2.

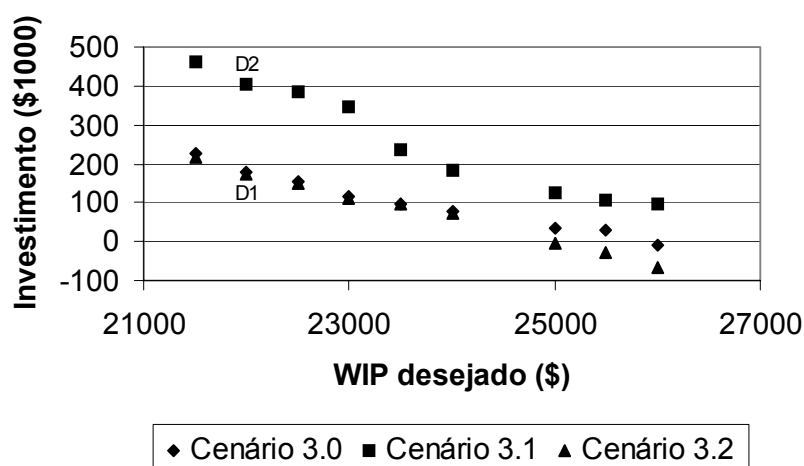


FIGURA 4.10 – *Trade-off* entre investimento em recursos e *WIP* desejado para os cenários 3.0 (quando o *mix* de produção é o da configuração inicial da rede), 3.1 (quando a classe de produtos 13 é extinta, com acréscimo da taxa de chegada da classe 2) e 3.2 (quando a classe de produtos 13 é extinta, com acréscimo da taxa de chegada da classe 3).

Percebe-se da figura 4.10 que a extinção da classe 13, com o aumento da participação da classe 3 no *mix* de produção (cenário 3.2), exige praticamente o mesmo investimento mínimo em capacidade que o *mix* de produção da configuração inicial (cenário 3.0) para diferentes limitantes superiores de *WIP*. Diferentemente, ao aumentar a participação no *mix* de produção da classe de produtos 2, com a extinção da classe 13 (cenário 3.1), o investimento mínimo necessário para manter os mesmos níveis de *WIP*

desejados é bem maior que o investimento mínimo com o *mix* de produção da configuração inicial (cenário 3.0). Ou seja, do ponto de vista de investimento e inventário em processo, para o gestor é indiferente manter o mesmo *mix* da configuração da rede inicial ou extinguir a classe de produto 13 e aumentar a participação da classe de produto 3. Entretanto, extinguir a classe 13 e aumentar a participação no *mix* de produção da classe de produtos 2 pode representar altos investimentos. Por exemplo, se o gestor decide manter o *WIP* em nível inferior a \$22000, deve investir no mínimo \$180500 (vide ponto D1, figura 4.10), mantendo o mesmo *mix* da configuração inicial – cenário 3.0. Se optar por extinguir a classe de produto 13 e aumentar a participação da classe 2 – cenário 3.1 – deve investir no mínimo \$406500 (vide ponto D2), o que representa um aumento de \$226000 em investimento em capacidade. Esta última alternativa deve ser avaliada cuidadosamente para verificar se o aumento no investimento é compensado por outros fatores, tais como atendimento a outros mercados com ampliação das vendas e aumento de faturamento. Ainda, este resultado de certa forma surpreendeu o engenheiro responsável pelo PCP que, inicialmente, esperava que a extinção da classe 13 iria liberar capacidade para produzir mais produtos da classe 2, uma vez que é a classe requisitada mais frequentemente e com boa margem de lucro, o que poderia aumentar o lucro da empresa. Estes resultados reforçam a importância do modelo no processo de tomada de decisões.

4.6 Comentários Gerais Acerca dos Resultados para o Estudo de Caso

Com vistas a verificar a aplicabilidade e adequação das aproximações para avaliar o desempenho do sistema, e estudar a classe de problemas *SPI* ou *SP2* em um caso real, foi escolhida uma empresa fabricante de peças metálicas para uma indústria automobilística e de eletrodomésticos para servir de estudo de caso. O sistema de manufatura *job-shop*, responsável por produzir peças para uma indústria automobilística, foi representado por uma rede de filas aberta, em que os nós correspondem às estações, e os arcos conectando os nós aos roteiros de fabricação das classes, conhecidos *a priori* (roteiros determinísticos). O sistema foi dividido em 15 estações de trabalho (modeladas como filas *GI/G/1*), visitadas por 14 classes de

produtos. Para avaliar o desempenho da rede, foram coletados dados para caracterizar os processos de chegada e de serviço nas estações por um período de 14 meses. Estes dados foram utilizados como entrada para o método aproximado de decomposição paramétrico (método analítico) e para a simulação discreta (método experimental), utilizada para validar os resultados obtidos pelas aproximações analíticas. Seis sistemas de aproximações diferentes que estimam o coeficiente quadrático de variação do intervalo de tempo entre chegadas nas estações da rede foram testados neste trabalho, com o objetivo de identificar as aproximações que melhor descrevem a rede. O sistema linear formado pelas aproximações (15), (27) e (49) apresentou resultados mais próximos de estoque em processo na rede (desvio relativo de $-0,9\%$) e de *leadtime* médio da rede (desvio relativo de $3,1\%$) aos obtidos por simulação, conforme as tabelas 4.8 e 4.9. Além disso, segundo opinião do engenheiro responsável pelo PCP da empresa, os resultados fornecidos pelas aproximações e pela simulação para o *leadtime* dos produtos são próximos dos observados no dia-a-dia da empresa, mostrando que o método aproximado de decomposição é bastante útil e adequado para descrever o desempenho de redes reais. Por este motivo, esse sistema linear foi utilizado nos modelos de otimização para avaliar, a cada iteração dos algoritmos de solução, o desempenho do sistema de manufatura do estudo de caso em diferentes configurações.

Dentre os problemas de alocação/relocação de capacidade estudados (*SP1* e *SP2*), os modelos referentes à alocação de capacidade discreta para uma rede ($SP1.1/G/S/R^{discreto}$ e $SP1.2/G/S/R^{discreto}$) mostraram-se mais adequados para aplicação ao estudo de caso. Para estes modelos, juntamente com o engenheiro responsável pelo PCP e o diretor industrial da empresa, foram propostas três alternativas de capacidade para cada estação (capacidade atual, adição e remoção de capacidade), tendo em vista manter em níveis desejados o *WIP* e o *leadtime* para cada classe de produto. Curvas de *trade-off* entre *WIP* e capital de investimento foram construídas (figura 4.3), variando parametricamente o limitante superior para o *WIP* no modelo $SP1.1/G/S/R^{discreto}$. Estas curvas demonstraram ser bastante úteis para auxiliar um gerente a tomar decisões. Por exemplo, se o gerente quiser reduzir o inventário da rede, que inicialmente era de aproximadamente \$27 mil para cerca de \$22 mil, ele sabe que deve investir cerca de \$180 mil em capacidade, conforme a tabela 4.22 e figura 4.3. O modelo $SP1.2/G/S/R^{discreto}$ mostrou-se útil pois auxilia um gerente a tomar decisões de alocação

de capacidade quando a medida de desempenho desejada é o *leadtime* para as classes de produtos. Com este modelo, conforme é observado na figura 4.4, o gerente pode saber quanto e onde investir o mínimo em capacidade na rede para manter o *leadtime* das classes de produtos em níveis inferiores aos estabelecidos que, por sua vez, podem refletir condições de competição de mercado, conforme visto nas tabelas 4.19 e 4.21. Este modelo mostrou-se mais interessante para o gestor da rede estudada do que o modelo $SP1.1/G/S/R^{discreto}$, uma vez que os investimentos em aquisição de capacidade são altos comparados à redução do *WIP*, mas podem ser representativos na redução do *leadtime* médio da rede. Por exemplo, o investimento de cerca de \$180 mil em capacidade reduz o *WIP* aproximadamente em \$5000 (de aproximadamente \$27 mil para cerca de \$22 mil), o que pode ser pouco significativo. Todavia este investimento reduz o *leadtime* médio da rede em mais de 2 dias de operação da empresa (redução de 187,6 para 151,0 horas, vide figura 4.6), o que é significativo para atendimento ao mercado. Os resultados de ambos os modelos foram obtidos em poucos segundos utilizando um computador pessoal (processador Pentium III de 800 MHz) o que reforça a aplicabilidade da metodologia aqui empregada: aplicação de modelos de otimização analíticos, utilizando métodos aproximados de decomposição para avaliar o desempenho da rede em cada iteração dos algoritmos de solução dos modelos.

As soluções do modelo $SP1.1/G/S/R^{discreto}$ também foram analisadas quando os coeficientes quadráticos de variação do intervalo de tempo entre chegadas externas das classes de produtos (ca'_k), as taxas de chegadas externas (λ'_k) da rede e o *mix* de produção são alterados. Com relação à variabilidade do intervalo de tempo entre chegadas, os resultados (vide figura 4.8) mostram, por exemplo, que se o gestor quiser reduzir o *WIP* inicial da rede de cerca de \$27 mil para cerca de \$22 mil, ele pode ou investir cerca de \$180 mil em capacidade (o *leadtime* médio da rede corresponde a 151,0 horas), mantendo a variabilidade dos processos de chegadas constantes, ou reduzir a variabilidade dos processos de chegadas externas à metade (o *leadtime* correspondente é 147,9 horas) e investir cerca de \$97 mil em máquinas (economia de \$83 mil e redução de *leadtime* médio da rede de 3,1 horas). Esta solução é viável de ser implementada, uma vez que a chegada foi considerada como o instante em que o lote inicia sua fabricação na primeira estação de trabalho, e bastaria alterar regras de

programação da produção. Este resultado enfatiza a sensibilidade do modelo de investimento a parâmetros de variabilidade, e cabe ao gestor avaliar se o investimento na redução da variabilidade do processo de chegadas é compensado pela economia em máquinas.

Quando as taxas de produção da rede (*throughput*) são aumentadas em, por exemplo, 10%, os resultados mostram que se o gestor quiser manter o *WIP* da rede em torno de \$27 mil (com *leadtime* médio da rede correspondente a 175,9 horas), ou seja, a configuração inicial, ele deve investir cerca de \$400 mil (figura 4.9), dadas as alternativas de adição/remoção de capacidade da tabela 4.4. Portanto, pode-se concluir que o modelo de alocação de capacidade discreta é bastante sensível a alterações na taxa de produção da rede estudada, considerando as alternativas estabelecidas.

Com relação ao *mix* de produção da rede, os resultados mostram que se, por exemplo, o gestor quiser manter o *WIP* em torno de \$22 mil, mantendo o *mix* inicial da rede, ele deve investir cerca de \$180 mil em máquinas (vide figura 4.10), com *leadtime* médio correspondente à rede de 151 horas. Entretanto, se quiser manter o mesmo nível de *WIP*, mas excluir a classe 13 e aumentar a produção da classe 2 de modo a manter a taxa de produção da rede constante, o investimento necessário é de aproximadamente \$400 mil (146,8 horas de *leadtime* médio da rede correspondente). Já se a classe 13 for extinta e a taxa da classe 3 aumentada (ao invés da classe 2), o investimento necessário seria de cerca de \$170 mil (149,6 horas de *leadtime* médio da rede correspondente), ou seja, menor até do que manter a fabricação da classe de produtos 13. Por este exemplo, pôde-se notar que a classe 3, ao contrário do que pensava o engenheiro responsável pelo PCP da empresa, exige menos investimento em capacidade, e aumentar a sua participação no *mix* de produção pode ser uma alternativa melhor do ponto de vista de redução do *WIP* que aumentar a participação da classe 2, até então tratada pela empresa como classe mais privilegiada que a classe 3. A empresa estava considerando, segundo depoimentos do engenheiro, reduzir a produção da classe 13, para liberar capacidade e aumentar a produção da classe 2. Novamente, este resultado enfatiza a sensibilidade do modelo de investimento em capacidade com o *mix* de produção para a rede estudada.

Este capítulo mostra a adequação e aplicabilidade dos modelos e métodos de solução discutidos nos capítulos 2 e 3, e a importância das soluções geradas para a

tomada de decisões em um estudo de caso. Especificamente, mostra-se a efetividade das aproximações analíticas em avaliar o desempenho de sistemas discretos de manufatura, e o quanto elas podem ser importantes para avaliar e otimizar o desempenho do sistema em diferentes configurações, quando inseridas em procedimentos de otimização que envolvem a alocação de capacidade, sujeita a níveis desejados de certas medidas de desempenho.

5 MODELOS DE OTIMIZAÇÃO EM REDES DE FILAS: PARTIÇÃO DA INSTALAÇÃO

Sistemas de manufatura têm se tornado mais complexos à medida em que uma maior variedade de produtos competem pelos mesmos recursos, demandas são incertas e o ciclo de vida dos produtos fica mais curto. Alternativas para reduzir essa complexidade têm sido sugeridas: partição de plantas de produção em subplantas, duplicação de recursos e reprojetado de produtos e processos de manufatura. A complexidade de um *job-shop* pode ser medida de muitas maneiras, tais como similaridade entre roteiros de produtos, tempos de operações em estações, etc. Apesar de úteis, estas medidas não são suficientes para fornecer uma compreensão adequada na gestão ou projeto de *job-shops*.

Bitran e Sarkar (1993c) afirmam que a previsibilidade de um *job-shop* é uma poderosa medida de complexidade. Primeiro, porque reflete o impacto de medidas de complexidade específicas (por exemplo, quanto mais similar os tempos de processamento, maior a previsibilidade). Segundo, porque é uma medida útil para um gerente de produção, permitindo-lhe prever razoavelmente bem o *leadtime* de produção. Quando muitos tipos diferentes de produtos são produzidos em um sistema, a previsibilidade tende a ser reduzida, mesmo que os roteiros sejam determinísticos, devido a efeitos de interferência nas estações. Assim, a previsibilidade está ligada à variação dos intervalos de partida do produto do sistema. Bitran e Sarkar (1993c) propõem medidas de complexidade que procuram captar esta noção.

Aqui, diferentemente dos modelos tratados anteriormente (Boxma et al., 1990; Bitran e Morabito, 1996; Bitran e Tirupati, 1989a e 1989b; Van Vliet e Rinooy Kan, 1991; Morabito, 1998), que abordaram a alocação e realocação de servidores a uma dada instalação, o projeto de fábrica focalizada integra a partição da instalação em subplantas (plantas menores dentro de uma planta) e a alocação de servidores em cada subplanta. A seguir são apresentadas medidas de complexidade e o problema de redução de complexidade para sistemas de produção propostos por Bitran e Sarkar (1993c).

5.1 O Sistema de Produção Modelado

O sistema de produção é representado por uma rede de filas aberta. Cada nó na rede representa uma estação que consiste de uma ou mais máquinas idênticas. Cada estação é modelada como uma fila $GI/G/m$. Desde que resultados exatos para avaliação de desempenho não existem para modelos de redes de filas abertas, aproximações são obtidas utilizando o método de decomposição abordado no capítulo 2.

Admite-se que o roteiro de fabricação de cada produto seja determinístico, e que haja r classes de produtos. Seja n_{kl} a estação visitada para a operação l do roteiro da classe k e s_{kl} o tempo de serviço (admitido determinístico por simplicidade) da operação l do roteiro da classe k . A metodologia também se aplica se s_{kl} é uma variável aleatória. Os tempos de serviço agregados nas estações, s_j , são caracterizados pela média e pelo coeficiente quadrático de variação ($E(s_j)$ e cs_j , respectivamente).

Supõe-se que máquinas idênticas possam ser empregadas em cada estação para processar produtos. Sob a abordagem de decomposição paramétrica, a variabilidade de serviço na estação j dependerá dos tempos de serviço dos produtos passando pela estação. Ainda, o coeficiente quadrático de variação do intervalo de tempo entre partidas e o tempo de espera no nó depende de ambas as variabilidades de serviço e de chegada. As variabilidades dos tempos de serviço e do intervalo de tempo entre chegadas podem ser controladas por agregação apropriada dos produtos em grupos, e atribuição de capacidade a cada estação dentro de um grupo. Cada grupo de produtos assim determinado representa uma linha de produção ou uma fábrica focalizada. Os produtos são agregados por Bitran e Sarkar (1993c) com vistas a minimizar a complexidade global do sistema.

5.2 Medidas de Complexidade

Bitran e Sarkar (1993c) consideram dois tipos de medidas de complexidade: uma do ponto de vista de gestão do produto e outra do ponto de vista de gestão da estação.

Do ponto de vista de gestão do produto, um gerente deveria ser capaz de prever o *leadtime* do produto o mais precisamente possível. Em outras palavras, a variância do tempo total gasto por um produto no sistema deveria ser a menor possível. A variância pode ser reduzida por meio de adição de máquinas na estação (e aí há o *trade-off* entre previsibilidade e investimento requerido), ou por meio de partição apropriada dos produtos.

Do ponto de vista de gestão da estação, quanto maior a diversidade de produtos ou incertezas na chegada em uma estação, maior deveria ser a flexibilidade para operá-la, em termos de manutenção e programação. Maior flexibilidade pode ser alcançada adquirindo mais máquinas. Bitran e Sarkar (1993c) mediram a diversidade de produtos ou pela variância dos tempos de serviço dos produtos, ou pelo coeficiente quadrático de variação do tempo de serviço na estação. A incerteza da chegada é medida pelo coeficiente quadrático dos intervalos de tempo entre chegadas.

A seguir são abordadas medidas de complexidade que podem compor o conjunto de restrições de problemas de otimização relacionados à partição da instalação.

5.2.1 Complexidade do Ponto de Vista de Gestão do Produto: Previsibilidade

O objetivo das restrições de previsibilidade é reduzir a variância do tempo que cada item/produto permanece no sistema. Se $V(T_k)$ é a variância do *leadtime* do produto k ($k=1, \dots, r$) na rede, então requer-se que:

$$\sum_{k=1}^r w_k V(T_k) \leq T^* \quad (78)$$

onde w_k representa um peso associado ao produto k , e T^* é um limitante superior para a variância ponderada do *leadtime*. O peso w_k pode depender do valor ou importância do produto. A metodologia é diretamente aplicável em situações com múltiplas restrições de complexidade como (78), ou com limitantes individuais para cada produto k . Note que quanto maior o lado esquerdo de (78), menor o potencial de previsibilidade do sistema.

Na abordagem de decomposição paramétrica, cada estação é tratada como se fosse aproximadamente independente das demais. Como resultado, $V(T_k)$ é aproximada simplesmente como a soma das variâncias de tempos de espera em fila

(uma vez que os tempos de serviço para os produtos em cada estação são considerados determinísticos e, portanto, a variância do tempo de serviço para cada classe k , $V(s_k)$, é nula, ou seja: $V(s_k) = \sum_{l=1}^{n_k} \sum_{j=1}^n V(s_{n_{kl}}) I\{n_{kl} = j\} = 0$). Logo,

$$V(T_k) = \sum_{l=1}^{n_k} \sum_{j=1}^n V(Wq_{n_{kl}}) I\{n_{kl} = j\} \quad (79)$$

em que:

$V(Wq_{n_{kl}})$ é a variância do tempo de espera na estação n_{kl} ;

n_k é o número de operações no roteiro da classe k ;

n_{kl} estação visitada para operação l do roteiro da classe k ;

$I\{\cdot\}$: função indicadora igual a 1 se $\{\cdot\}$ é verdadeiro e 0, caso contrário.

Utilizando aproximações de tráfego pesado (em que as utilizações médias de capacidade ρ_j são altas, isto é, próximas de 1) para $V(Wq_j)$, tem-se (Bitran e Sarkar, 1993c):

$$V(Wq_j) = \frac{(ca_j + cs_j)}{4} E(Wq_j)_{M/M/m_j}^2 \quad (80)$$

com $E(Wq_j)_{M/M/m_j}$, dado por (10), representando o tempo médio de espera para um sistema $M/M/m_j$ em equilíbrio.

5.2.2 Complexidades do Ponto de Vista da Estação

Para gerir uma estação eficientemente, espera-se que o número de máquinas alocadas às estações sejam proporcionais às incertezas de chegada e de tempo de serviço dos produtos que passam pela estação, porque quanto maior o número de máquinas disponíveis para lidar com as incertezas em uma única estação, maior é a flexibilidade da estação. Bitran e Sarkar (1993c) sugerem que a mesma flexibilidade pode ser alcançada com menos máquinas, quando classes de produtos passando pela estação são particionadas adequadamente.

Whitt (1992) propôs a seguinte equação como uma aproximação (e um limitante inferior aproximado) para a maneira com que ρ_j deveria aumentar com m_j na estação j (esta equação considera que a taxa de serviço de cada servidor é igual a 1):

$$(1 - \rho_j) \sqrt{m_j} = \gamma_j \quad (81)$$

em que γ_j é uma constante, indicando de forma aproximada o nível de serviço da estação.

A expressão (81) relaciona ρ_j com m_j tal que uma medida de congestionamento se mantenha aproximadamente fixa, no caso a probabilidade do tempo de espera em fila ser positivo, $P(Wq_j > 0)$. Whitt (1992) mostra que se ρ_j e m_j são mudados em (81) de forma que γ_j se mantenha fixo, então $P(Wq_j > 0)$ deve permanecer aproximadamente constante.

Whitt (1992) também propõe aproximações de difusão de tráfego pesado para caracterizar aproximadamente o tempo médio de espera de um usuário, dado que ele espera antes de iniciar o serviço, isto é, $E(Wq_j | Wq_j > 0)$, para uma fila $GI/G/m_j$ com taxa $\mu_j=1$. Ou seja, a taxa de serviço da estação é dada por m_j . A análise de tráfego pesado sugere a seguinte aproximação para $E(Wq_j | Wq_j > 0)$, quando (81) é satisfeita (Whitt, 1992):

$$P(Wq_j > x | Wq_j > 0) \approx e^{-\beta_j \sqrt{m_j}} \quad (82) \quad e$$

$$E(Wq_j | Wq_j > 0) \approx 1/\beta_j \sqrt{m_j} \quad (83)$$

com

$$\beta_j = \frac{2\gamma_j}{ca_j + cs_j} \quad (84)$$

Substituindo (84) em (83), tem-se:

$$E(Wq_j | Wq_j > 0) \approx \frac{ca_j + cs_j}{2\gamma_j \sqrt{m_j}} \Rightarrow \gamma_j \approx \frac{ca_j + cs_j}{2E(Wq_j | Wq_j > 0) \sqrt{m_j}} \quad (85)$$

A equação (85) implica que, para um dado nível de serviço γ_j , a medida de congestionamento $E(Wq_j | Wq_j > 0)$ é mantida aproximadamente constante.

Considerando igualdade em (85), tem-se de (81) e (85):

$$\gamma_j = (1 - \rho_j) \sqrt{m_j} = \frac{ca_j + cs_j}{2E(Wq_j | Wq_j > 0) \sqrt{m_j}} \Rightarrow \frac{(1 - \rho_j) m_j}{ca_j + cs_j} = \frac{1}{2E(Wq_j | Wq_j > 0)} \quad (86)$$

Segue de (86) que ao se adicionar máquinas e aumentar a taxa média de chegada na estação j , a utilização média, $\rho_j = \lambda_j / m_j \mu_j$, cresce para manter $E(Wq_j | Wq_j > 0)$ constante, supondo ca_j e cs_j independentes a mudanças de

capacidade. A seguinte restrição de complexidade, baseada em (86), pode ser imposta para cada estação da rede:

$$\frac{(1 - \rho_j)m_j}{ca_j + cs_j} \geq G_j \quad (87)$$

Esta restrição implica que G_j é um limitante inferior para o lado direito de (86), ou seja, é um limitante superior para o tempo médio de espera, dado que espera. Desta maneira, G_j pode ser interpretado como a mínima flexibilidade exigida na estação j . Pode-se ainda abstrair de (87) que pode-se satisfazer a mínima flexibilidade desejada nas estações sem alterar o número total de máquinas na rede. Ao particionar apropriadamente a planta em subplantas com um *mix* de produtos mais homogêneo, pode-se obter ca_{jp} e cs_{jp} para cada estação j de cada subplanta p , tal que o lado esquerdo de (87) cresça para todo j e p .

Manipulando (86) e (87), $1/(2E(Wq_j | Wq_j > 0)) \geq G_j$, tem-se que $1/(2G_j) \geq E(Wq_j | Wq_j > 0)$. Definindo, $G'_j = 1/(2G_j)$, obtém-se $G'_j \geq E(Wq_j | Wq_j > 0)$ e, portanto, G'_j é um limitante superior para a medida de congestionamento $E(Wq_j | Wq_j > 0)$:

$$E(Wq_j | Wq_j > 0) \leq \frac{1}{2G_j} = G'_j. \quad (88)$$

Quanto maior G_j (menor G'_j), maior a flexibilidade mínima da estação j , e menor o tempo médio de espera de um usuário, dado que ele espera (de acordo com (88)). E, portanto, menor o inventário em processo na rede, uma vez que o tempo médio de espera tende a ser menor (lembre-se da lei de Little que o inventário em processo e o tempo de espera são proporcionais).

O número de máquinas em cada estação j , necessário para manter a mínima flexibilidade na estação, pode ser estimado substituindo $\rho_j = \lambda_j / \mu_j m_j$ em (87):

$$m_j \geq \frac{\lambda_j}{\mu_j} + G_j(ca_j + cs_j) \quad (89)$$

Convém ressaltar que as expressões de (81) a (89), abordadas em Morabito (1998) e Whitt (1992), admitem que a fila $GI/G/m_j$ possui cada servidor com

taxa de processamento $\mu_j=1$, ou seja, a taxa de processamento da estação é dada por m_j . Para situações em que as taxas de processamento dos servidores são diferentes de 1, as hipóteses estabelecidas e os resultados permanecem válidos, mas as expressões (81) – (89) resultam, respectivamente em (compare com as expressões anteriores):

$$(1 - \rho_j)\sqrt{\mu_j m_j} = \gamma_j \quad (90)$$

$$P(Wq_j > x | Wq_j > 0) \approx e^{-\beta_j \sqrt{\mu_j m_j}} \quad (91) \text{ e}$$

$$E(Wq_j | Wq_j > 0) \approx 1/\beta_j \sqrt{\mu_j m_j} \quad (92)$$

com

$$\beta_j = \frac{2\gamma_j}{ca_j + cs_j} \quad (93)$$

$$E(Wq_j | Wq_j > 0) \approx \frac{ca_j + cs_j}{2\gamma_j \sqrt{\mu_j m_j}} \Rightarrow \gamma_j \approx \frac{ca_j + cs_j}{2E(Wq_j | Wq_j > 0)\sqrt{\mu_j m_j}} \quad (94)$$

$$\gamma_j = (1 - \rho_j)\sqrt{\mu_j m_j} = \frac{ca_j + cs_j}{2E(Wq_j | Wq_j > 0)\sqrt{\mu_j m_j}} \Rightarrow$$

$$\frac{(1 - \rho_j)\mu_j m_j}{ca_j + cs_j} = \frac{1}{2E(Wq_j | Wq_j > 0)} \quad (95)$$

$$\frac{(1 - \rho_j)\mu_j m_j}{ca_j + cs_j} \geq G_j \quad (96)$$

$$E(Wq_j | Wq_j > 0) \leq \frac{1}{2G_j} \quad (97)$$

$$m_j \geq \frac{\lambda_j}{\mu_j} + \frac{G_j}{\mu_j}(ca_j + cs_j) \quad (98)$$

Note que (82), (83), (85), (86), (87) e (89) são casos particulares de (91), (92), (94), (95), (96) e (98) em que $\mu_j=1$.

Bitran e Sarkar (1993c) observaram que a inequação (89) (similarmente para (98)), estabelece uma relação entre os parâmetros de tempo de serviço e o número de máquinas requerida na estação. Uma característica interessante de (89), ou (98), é que quando ca_j é grande com relação a cs_j , a partição de produtos pode aumentar o número de máquinas significativamente. Por exemplo, considere um conjunto de produtos com tempos de serviço determinísticos e idênticos ($cs_j=0$). Se particionarmos

os produtos em n grupos, então $\lambda_{jp}/\mu_{jp} + (G_{jp}/\mu_{jp})ca_{jp}$ máquinas são requeridas para o grupo p , enquanto que $(\sum_{p=1}^n \lambda_{jp}/\mu_j) + (G_j/\mu_j)ca_j$ máquinas são necessárias quando os produtos formam um grupo único. Portanto, dependendo dos valores de G_j , ca_j e μ_j , o número de máquinas para o conjunto de produtos pode ser significativamente maior quando as linhas são particionadas.

Bitran e Sarkar (1993c) ainda propõem $\lambda_j V(s_j)$ como uma terceira medida de complexidade em uma estação ($V(s_j)$ é a variância do tempo de serviço na estação j), para situações em que ca_j é pequeno comparado a cs_j . O objetivo é encontrar uma expressão simples para m_j como uma função dos parâmetros de chegada e de serviço, tal que o tempo de espera em fila, $E(Wq_j)$, na estação j permaneça aproximadamente constante. Para isso, a aproximação (21), supondo $g(\rho_j, ca_j, cs_j)=1$, foi utilizada:

$$E(Wq_j) = \frac{\frac{\lambda_j}{m_j} \frac{1}{\mu_j^2} (ca_j + cs_j)}{2 \left[1 - \frac{\lambda_j}{\mu_j m_j} \right]}$$

Desde que o ca_j ($j=1, \dots, n$) seja suficientemente pequeno comparado a cs_j , ele pode ser ignorado, o que resulta em:

$$E(Wq_j) = \frac{\lambda_j V(s_j)}{2 \left[m_j - \frac{\lambda_j}{\mu_j} \right]} \Rightarrow m_j = \frac{\lambda_j}{\mu_j} + \frac{\lambda_j V(s_j)}{2E(Wq_j)} \quad (99)$$

Definindo, $h_j = 1/(2E(Wq_j))$, obtém-se:

$$m_j = \frac{\lambda_j}{\mu_j} + h_j \lambda_j V(s_j) \quad (100)$$

Na expressão seguinte, a interpretação natural é que, para a complexidade, $\lambda_j V(s_j)$, permanecer baixa em um ambiente de produção de alto volume, os tempos de serviço deveriam ser similares (portanto, $V(s_j)$ baixa), enquanto que em um ambiente onde os tempos de serviço são muito diferentes, o volume de produção deveria ser relativamente pequeno (λ_j baixa). A seguinte restrição de complexidade pode ser imposta a cada estação j :

$$m_j \geq \frac{\lambda_j}{\mu_j} + h_j \lambda_j V(s_j) \quad (101)$$

Um modelo de projeto de fábrica focalizada utilizando essas restrições de complexidade é apresentado em seguida.

5.3 Definição do Problema de Redução de Complexidade (PRC)

Bitran e Sarkar (1993c) formularam um problema para o projeto de fábrica focalizada, cujo objetivo é encontrar a partição da instalação que reduz o número de máquinas em todas linhas (subplantas) e atender uma determinada medida de complexidade.

Sejam:

$$x_{kp} = \begin{cases} 1, & \text{se produto } k \text{ é atribuído à linha } p; \\ 0, & \text{caso contrário} \end{cases};$$

m_{jp} : número de máquinas na estação j da linha p .

O objetivo do PRC é atribuir cada produto k a uma linha tal que a medida de complexidade desejada seja alcançada e o número de máquinas requeridas seja mínimo. Ou seja:

$$\min \sum_{j=1}^n \sum_{p=1}^r c_{jp} m_{jp} \quad (102a)$$

$$\text{sa: restrição de complexidade} \quad (102a)$$

$$\sum_{p=1}^r x_{kp} = 1, \quad k = 1, 2, \dots, r \quad (102b)$$

$$\lambda_{jp} / \mu_{jp} = \sum_{k=1}^r \sum_{l=1}^{n_k} \lambda'_k E(s_{kj}) 1\{n_{kl} = j\} x_{kp} < m_{jp}, \quad j = 1, \dots, n; \quad p = 1, \dots, r \quad (102c)$$

$$x_{kp} = 0, 1; \quad m_{jp} = \text{inteiro}, \quad k, p = 1, \dots, r; \quad j = 1, \dots, n \quad (102d)$$

sendo $E(s_{kj})$ o tempo médio de serviço do produto k na estação j . Note que λ_{jp} é a taxa média de chegada total na estação j da linha p , ou seja, $\lambda_{jp} = \sum_{k=1}^r \sum_{l=1}^{n_k} \lambda'_k 1\{n_{kl} = j\} x_{kp}$, $\rho_{jp} = \lambda_{jp} / m_{jp} \mu_{jp}$ é o nível médio de utilização da estação j da planta p e c_{jp} o custo unitário da máquina na estação j da linha p . Por simplicidade de análise, neste trabalho considera-se $c_{jp}=1$ ($j=1, \dots, n; p=1, \dots, r$), ou seja, o objetivo é simplesmente minimizar o número total de máquinas.

Observa-se do modelo que o objetivo do PRC é minimizar o total de máquinas em todas estações e linhas. As restrições (102b) e (102d) garantem juntas que

todo produto deve ser atribuído a uma e somente uma linha de produção. A restrição (102c) garante que o sistema esteja em equilíbrio, ou seja, $\rho_{jp} = \frac{\lambda_{jp}}{m_{jp}\mu_{jp}} < 1$ para toda estação j ($j=1, \dots, n$) e linha p ($p=1, \dots, r$). Numa implementação computacional, esta restrição em geral deve satisfazer uma certa tolerância ($\rho_{jp} \leq 1 - \varepsilon$, com $\varepsilon > 0$ representando a tolerância). É importante observar também que o modelo assume que podem existir tantas linhas p quantas forem as classes de produtos k , uma vez que p varia de 1 a r , o número de classes de produtos.

A depender da medida de complexidade que se pretende utilizar, o PRC tem uma formulação diferente, conforme discutido a seguir.

5.3.1 Problema de Redução de Complexidade PRC-1

Quando a restrição de complexidade (102a) é dada por (78):

$$\sum_{k=1}^r w_k V(T_k) \leq T^*$$

com

$$V(T_k) = \sum_{l=1}^{n_k} \sum_{j=1}^n \sum_{p=1}^r V(Wq_{jp}) \mathbb{1}\{n_{kl} = j\} x_{kp} \quad (103)$$

$$V(Wq_{jp}) = \frac{(ca_{jp} + cs_{jp})^2}{4} E(Wq_{jp})_{M/M/m_j}^2 \quad (104)$$

em que $V(Wq_{jp})$ é a variância do tempo em espera em fila na estação j da linha p ; e

$$E(Wq_{jp})_{M/M/m_j} = \frac{(\lambda_{jp} / \mu_{jp})^{m_{jp}} \pi_{jp}(0)}{\mu_{jp} m_{jp} (1 - \rho_{jp})^2 m_{jp}!} \quad (105)$$

Neste caso, o PRC é chamado de PRC-1.

5.3.2 Problema de Redução de Complexidade PRC-2

Quando a restrição de complexidade é dada por (98) e adaptada para uma rede com p plantas, a restrição (102a) fica:

$$\sum_{k=1}^r \sum_{l=1}^{n_k} \lambda'_k E(s_{kj}) \mathbb{1}\{n_{kl} = j\} x_{kp} + \frac{G_{jp}}{\mu_{jp}} (ca_{jp} + cs_{jp}) \leq m_{jp}, \quad p=1, \dots, r; j=1, \dots, n \quad (106)$$

em que os ca_{jp} 's são os coeficientes quadráticos de variação dos intervalos de tempo entre chegadas na estação j da linha p , G_{jp} é uma constante indicando a mínima flexibilidade da estação j e linha p e

$$cs_{jp} = V(s_{jp}) / E(s_{jp})^2$$

com

$$V(s_{jp}) = \frac{\sum_{k=1}^r \sum_{l=1}^{n_k} \lambda'_k x_{kp} 1\{n_{kl} = j\} (E(s_{kj}) - E(s_{jp}))^2}{\sum_{k=1}^r \sum_{l=1}^{n_k} \lambda'_k x_{kp} 1\{n_{kl} = j\}} \quad (107)$$

e

$$E(s_{jp}) = \frac{\sum_{k=1}^r \sum_{l=1}^{n_k} \lambda'_k x_{kp} 1\{n_{kl} = j\} E(s_{kj})}{\sum_{k=1}^r \sum_{l=1}^{n_k} \lambda'_k x_{kp} 1\{n_{kl} = j\}} \quad (108)$$

onde $V(s_{jp})$ e $E(s_{jp})$ são a variância e a média dos tempos de serviço na estação j e linha p . Neste caso, o PRC é chamado de PRC-2.

5.3.3 Problema de Redução de Complexidade PRC-3

Finalmente, o PRC-3 é formulado utilizando a definição de medida de complexidade expressa na relação (101). Assim, a restrição (102a) no PRC-3 fica:

$$\sum_{k=1}^r \sum_{l=1}^{n_k} \lambda'_k E(s_{kj}) 1\{n_{kl} = j\} x_{kp} + h_{jp} \lambda_{jp} V(s_{jp}) \leq m_{jp}, \quad p = 1, \dots, r; j = 1, \dots, n \quad (109)$$

5.4 Solução do Problema de Redução de Complexidade (PRC)

Os problemas PRC-1 – PRC-3 acima definidos são em geral problemas de programação inteira não-convexos e difíceis para resolver otimamente. Bitran e Sarkar (1993c) propuseram um método de solução que emprega um algoritmo (Algoritmo 6 adiante) para alocar produtos às linhas de produção (particionar a instalação), requerendo o menor número de máquinas (admitindo $c_{jp}=1$ na função-objetivo) e satisfazer as restrições de complexidade. Este algoritmo utiliza idéias similares às do algoritmo de economias de Clarke-Wright (*apud* Larson e Odoni, 1981) para o problema de roteirização de veículos. Basicamente, a heurística do tipo gulosa é iniciada com uma linha para cada classe de produto em que se calcula o número mínimo de máquinas necessárias para atender às restrições e, em seguida, agrega-se o par de

linhas (e, conseqüentemente, os produtos) que proporciona maior redução (economia) no número de máquinas. A partição da instalação é atualizada e, novamente, agrega-se o par de linhas que proporciona maior redução no número de máquinas. Este processo continua enquanto houver economia no número de máquinas com a agregação de um par de linhas qualquer.

Algoritmo 6:

Passo 0: Faça $iter=0$. Aloque cada produto a uma linha separada. Aloque máquinas às estações de modo a satisfazer as restrições com o menor número de máquinas possível (os respectivos métodos utilizados para alocar máquinas para os problemas PRC-1, PRC-2 e PRC-3 são definidos a seguir). Calcule o número de máquinas requeridas.

Passo 1: Faça $iter=iter+1$. Escolha duas linhas (*clusters* ou conjuntos) quaisquer e agregue-as. Aloque máquinas às estações de modo a satisfazer as restrições com o menor número de máquinas possível (os respectivos métodos utilizados para alocar máquinas para os problemas PRC-1, PRC-2 e PRC-3 são definidos a seguir). Seja $\Delta m(p,p1)$ a mudança no número de máquinas quando as linhas p e $p1$ são agregadas e, conseqüentemente, os respectivos produtos pertencentes às linhas. Calcule $\Delta m(p,p1)$ para cada par de linhas da rede.

Passo 2: Se $\Delta m(p,p1) < 0$, então faça $(P,P1) = \arg \min_{p,p1} \Delta m(p,p1)$, agregue as linhas P e $P1$, atualize a lista de linhas (*clusters*), e volte para o passo 1. Senão, pare. A agregação dos produtos nas linhas da iteração anterior (partição da instalação) é a que resulta no menor número de máquinas.

Note na heurística que ao agregar um par de linhas qualquer, é necessário calcular o número mínimo de máquinas requeridas para calcular a mudança no número de máquinas com a agregação.

No PRC-1, para uma dada partição (x_{kp} conhecidos para $k=1,\dots,r$; $p=1,\dots,r$), ca_{jp} e cs_{jp} podem ser determinados e (104) é convexo em m_{jp} . Entretanto, PRC-1 permanece um modelo não-linear e inteiro em m_{jp} , o que inviabiliza sua resolução direta por meio de um *solver* linear inteiro (por exemplo, o *GAMS/CPLEX*[®]) para determinar o número mínimo de máquinas requerido para a partição dada. Por isso, no presente trabalho propõe-se utilizar o algoritmo de Boxma et al (1990), vide

Algoritmo 1 na seção 3.3.2.1, com uma pequena variação no índice de prioridade. Em outras palavras, conhecidos x_{kp} , $k=1, \dots, r$; $p=1, \dots, r$, em PRC-1, o algoritmo de Boxma et al (1990) é aplicado para determinar o número mínimo de máquinas nas estações das linhas. O algoritmo começa com o menor número de máquinas possível para cada estação, satisfazendo (102c). Em cada iteração, adiciona uma máquina na estação com o menor índice de prioridade. Ou seja, aquela estação que conduz à máxima redução na medida de variância ponderada. O algoritmo pára tão logo a restrição (102a) ou, particularmente, a expressão (78) seja satisfeita. Para o PRC-1, o índice de prioridade, determinando a estação j de uma linha p na qual uma máquina será adicionada, é dado por:

$$PI_{jp}(m_j) = \frac{\Delta m_{jp}}{-\Delta \sum_{k=1}^r w_k V(T_k)} \quad j=1, \dots, n \quad (110)$$

com $\Delta m_{jp} = (m_{jp} + 1) - (m_{jp})$, representando a adição de máquinas e

$$\Delta \sum_{k=1}^r w_k V(T_k) = \sum_{k=1}^r w_k V(T_k)(m_{jp} + 1) - \sum_{k=1}^r w_k V(T_k)(m_{jp}),$$

representando a mudança de variância total na rede com a adição de uma máquina na estação j da linha p .

Já para os problemas PRC-2 e PRC-3, para uma dada partição da instalação (x_{kp} , $k=1, \dots, r$; $p=1, \dots, r$ conhecidos), os modelos são lineares inteiros em m_{jp} , o que permite a utilização direta de um *solver* linear inteiro do tipo *GAMS/CPLEX*[®] para alocar o número mínimo de máquinas necessárias para satisfazer as respectivas restrições. Ou seja, aqui para uma dada partição, a capacidade mínima requerida que satisfaz as restrições de PRC-2 e PRC-3 é determinada utilizando o *solver* do *GAMS/CPLEX*[®], ao invés da heurística gulosa proposta por Boxma et al (1990).

O método aproximado de decomposição, discutido no capítulo 2, foi empregado para avaliar o desempenho de instalações, em uma dada configuração, representadas por uma rede com várias estações. A menos que seja feita alguma menção em contrário, foi utilizado o sistema de aproximações composto pelas aproximações:

$$ca_{jp} = \sum_{i=0}^n \frac{\lambda_{ijp}}{\lambda_{jp}} ca_{ijp} \quad j=1, \dots, n; \quad p=1, \dots, r \quad (111)$$

$$cd_{jp} = \frac{\rho_{jp}^2}{\sqrt{m_{jp}}} cs_{jp} + [1 - \rho_{jp}^2] ca_{jp} + \rho_{jp}^2 \left[1 - \frac{1}{\sqrt{m_{jp}}} \right] \quad j=1, \dots, n; \quad p=1, \dots, r \quad (112)$$

$$cd_{jip} = q_{jip} cd_{jp} + (1 - q_{jip}) q_{jip} ca_{jp} + (1 - q_{jip})^2 ce_{jp} \quad i, j=1, \dots, n; \quad p=1, \dots, r \quad (113)$$

Este sistema foi empregado para que os resultados computacionais obtidos neste trabalho pudessem ser comparados com os obtidos por Bitran e Sarkar (1993c), que empregaram as mesmas aproximações. Note que (111) e (113) são adaptações das aproximações (15) e (54), à medida que especifica a estação j ou i da linha p . A expressão (54) foi proposta por Segal e Whitt (1989). Já a aproximação (112) foi desenvolvida por Whitt (1983a) *apud* Bitran e Sarkar (1993c).

Os resultados computacionais obtidos ao aplicar estes métodos de solução para particionar a instalação (modelada por uma rede composta por vários estágios $GI/G/m$ e por um estágio único) são apresentados e analisados no capítulo 6.

6 RESULTADOS COM PROBLEMA DE PARTIÇÃO DA INSTALAÇÃO

Neste capítulo são apresentados os resultados referentes aos problemas de partição da instalação (classe *SP3*) que, diferentemente dos tratados nos capítulos 4 e 5, envolvem a partição da planta em subplantas, com alocação dos produtos e alocação de capacidade nas subplantas. Analisa-se aqui a adequação e aplicabilidade dos modelos e métodos de solução para o problema de partição da instalação. Os resultados reportados neste capítulo mostram que particionar uma planta em subplantas, pode ser mais vantajoso do que não particionar, e que a partição não necessariamente requer capacidade adicional. Ao contrário, pode requerer até menos capacidade do que a opção de não particionar a instalação.

Os modelos PRC-1, PRC-2 e PRC-3 (tratados no capítulo 5) não puderam ser diretamente aplicados à rede do estudo de caso por esta se tratar de uma rede de filas *GI/G/1*. A rede que representa o estudo de caso foi modelada por um conjunto de estações *GI/G/1*, com vistas a simplificar o processo de coleta e análise de dados, o que impossibilitou sua aplicação. Por exemplo, convém lembrar que algumas estações desta rede continham máquinas diferentes, operadas por um ou mais operadores, o que tornaria imprecisa a análise de sistemas de filas *GI/G/m*, supondo máquinas paralelas idênticas em cada estação. Os modelos PRC-1 – PRC-3 foram então inicialmente aplicados a duas redes ilustrativas, uma composta de uma única estação *GI/G/m* e a outra de duas estações *GI/G/m*. Em seguida, aplicou-se os modelos a redes mais realistas disponíveis em literatura, ambas utilizadas em Bitran e Sarkar (1993c).

As redes ilustrativas são exemplos simples e foram criadas pelo autor com o intuito de entender e ilustrar sob quais condições é mais vantajoso particionar ou não particionar a instalação. As figuras 6.1 e 6.2 apresentam as redes ilustrativas, denominadas doravante por Rede 1 e Rede 2, respectivamente.

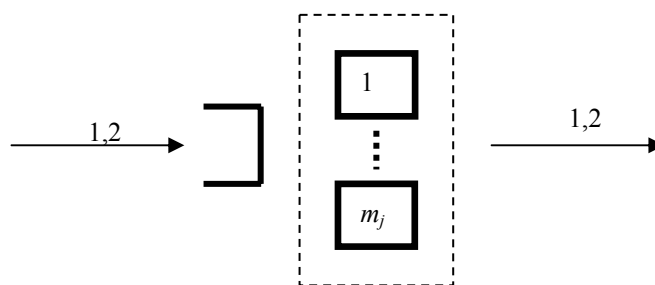


FIGURA 6.1 – Rede formada por um estágio único $GI/G/m$, processando 2 classes de produtos (Rede 1).

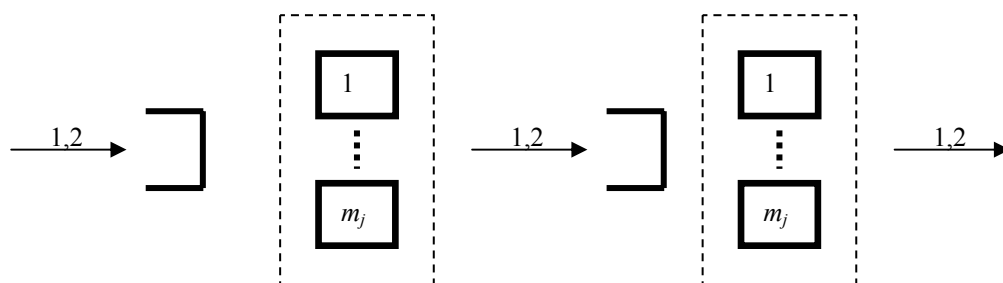


FIGURA 6.2 – Rede formada por dois estágios $GI/G/m$, processando 2 classes de produtos (Rede 2).

As tabelas 6.1 e 6.2 mostram os dados referentes às redes 1 e 2, respectivamente.

TABELA 6.1 – Dados para as duas classes de produtos na rede formada por um único estágio $GI/G/m$ (Rede 1).

Classe	λ'_k	$E(s_{k1})$	ca'_k	cs_{k1}	w_k
1	4	0,20	2,0	0	1
2	1	0,90	0,1	0	1

TABELA 6.2 – Dados para as duas classes de produtos na rede formada por dois estágios $GI/G/m$ (Rede 2).

Classe	λ'_k	$E(s_{k1})$	$E(s_{k2})$	ca'_k	cs_{k1}	cs_{k2}	w_k
1	4	0,15	0,15	0,500	0,000	0,000	1
2	1	0,85	0,85	0,100	0,000	0,000	1

A figura 6.3 ilustra a terceira rede em que os modelos PRC-1, PRC-2 e

PRC-3 foram aplicados. A rede é formada por um único estágio $GI/G/m$, visitada por 10 classes de produtos e doravante chamada de Rede 3.

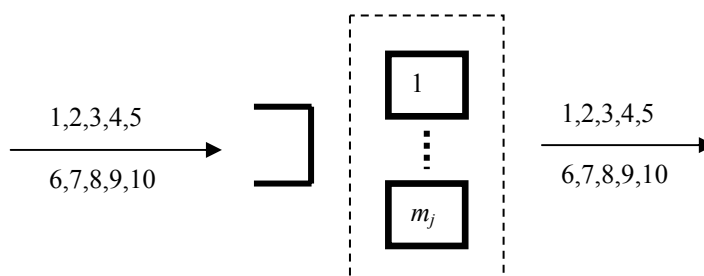


FIGURA 6.3 – Rede formada por um estágio único $GI/G/m$, processando 10 classes de produtos (Rede 3).

Os parâmetros de entrada que caracterizam os produtos que visitam a Rede 3 são dados na tabela 6.3. Note que o tempo de serviço foi considerado determinístico ($cs_{kl}=0$).

TABELA 6.3 – Dados para as 10 classes de produtos para a rede com um único estágio $GI/G/m$ (Rede 3).

Classe	λ'_k	$E(s_{kl})$	ca'_k	cs_{kl}	w_k
1	0,075	30,50	0,133	0	0,5
2	0,100	27,84	0,133	0	0,5
3	0,125	10,27	0,150	0	0,5
4	0,150	7,72	0,150	0	0,5
5	0,175	7,68	0,150	0	0,5
6	0,179	7,65	0,150	0	0,5
7	0,192	1,12	0,150	0	0,5
8	0,200	1,10	0,130	0	0,5
9	0,225	1,08	0,100	0	0,5
10	0,250	1,07	0,100	0	0,5

Os dados para a Rede 3 foram adaptados de Bitran e Sarkar (1993c) para permitir que o PRC-1 pudesse ser resolvido utilizando o *software GAMS*. Para estimar a variância do *leadtime* foi necessário estimar a variância do tempo de espera em fila, dado por (80), que é função do tempo de espera em uma fila $M/M/m$, $E(Wq_j)_{M/M/m}$, que, conforme (10), depende do cálculo de $m_j!$. O *software GAMS 2.0* não trata números com valores absolutos maiores que $1,000.10^{20}$ (Brooke et al, 1998, pg 59), e todos os valores derivados deles são inutilizados no modelo. Por isso, calcular o tempo de espera para uma estação com mais de 21 máquinas não é tratável, dado que $22! \approx 1,124.10^{21}$. Por

este motivo, na rede original proposta por Bitran e Sarkar (1993c), sem perda de generalidade, os tempos de serviço e taxa de chegada foram arbitrariamente divididos por 2 e 4, respectivamente, para que a rede pudesse ser tratada computacionalmente pelo *GAMS*.

A figura 6.4 ilustra a rede composta por um conjunto de estações *GI/G/m* que doravante é chamada de Rede 4. A Rede 4, proposta por Bitran e Sarkar (1993c), foi utilizada neste trabalho para os problemas PRC-1, PRC-2 e PRC-3. As características da rede referentes ao processo de chegadas e roteiro dos produtos são dados na tabela 6.4.

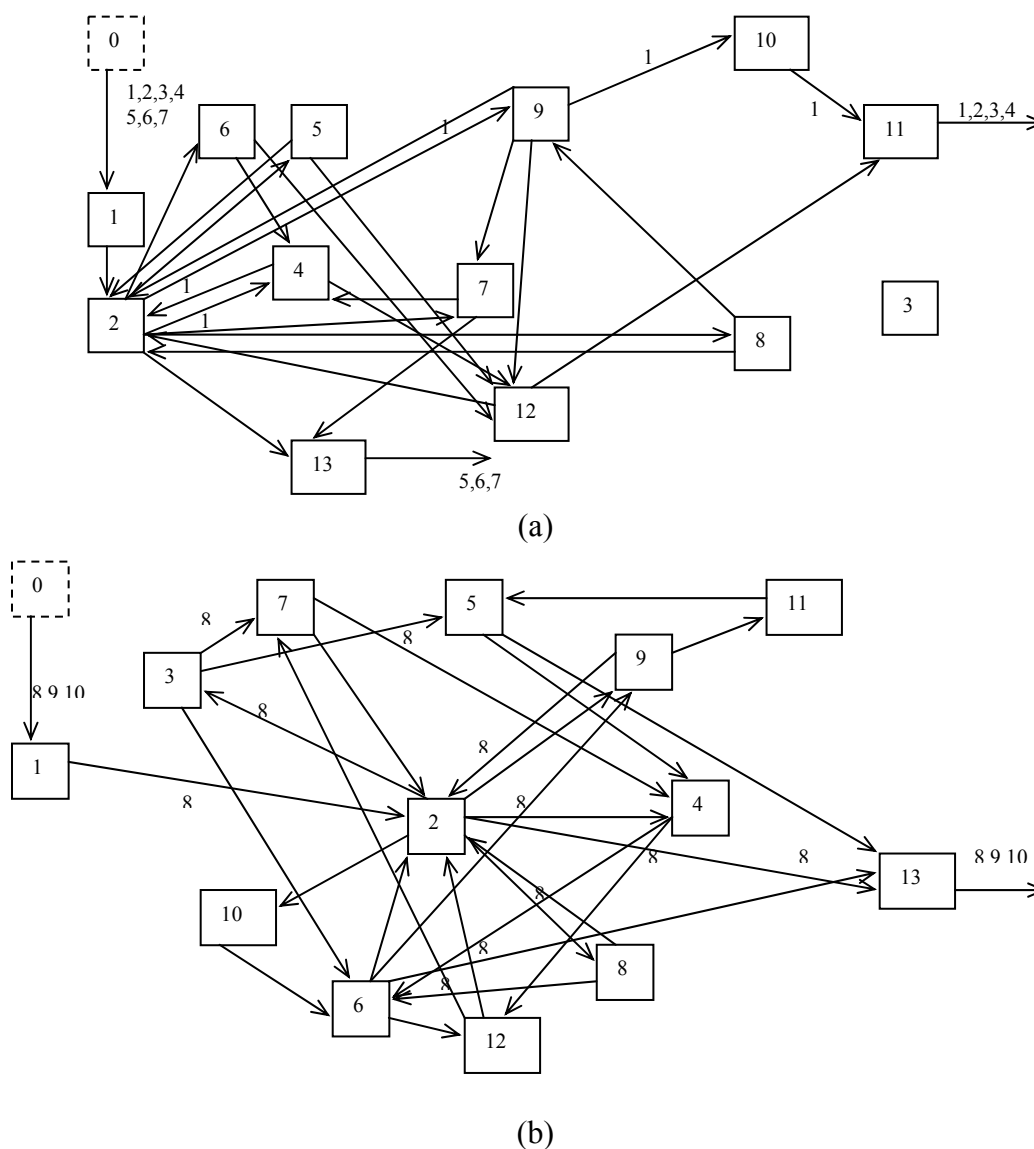


FIGURA 6.4 – Rede formada por vários estágios *GI/G/m*, processando as classes 1,2,...,7 (Figura 6.4a) e 8,9,10 (Figura 6.4b) de produtos (Rede 4).

TABELA 6.4 – Parâmetros da rede referentes ao roteiro (n_{kl}), ao número de operações (n_k), ao peso da variância do *leadtime* (w_k), ao coeficiente quadrático de variação de intervalo de tempo entre chegadas (ca'_k) e à taxa média de chegada (λ'_k) para as 10 classes de produtos (Rede 4).

Classe	λ'_k (item/h)	ca'_k	w_k	n_k	n_{kl}
1	0,10	0,333	0,5	7	1, 2, 4, 2, 9, 10, 11
2	0,10	0,500	0,5	8	1, 2, 5, 2, 8, 9, 10, 11
3	0,10	0,333	0,5	8	1, 2, 6, 4, 2, 9, 12, 11
4	0,10	0,333	0,5	8	1, 2, 7, 4, 2, 9, 10, 11
5	0,10	0,250	0,5	8	1, 2, 4, 12, 2, 9, 2, 13
6	0,10	0,500	0,5	8	1, 2, 5, 12, 2, 9, 7, 13
7	0,10	0,250	0,5	8	1, 2, 6, 12, 2, 8, 2, 13
8	0,10	0,333	0,5	12	1, 2, 3, 7, 4, 12, 2, 8, 6, 9, 2, 13
9	0,10	0,050	0,1	13	1, 2, 3, 5, 4, 6, 12, 2, 8, 2, 10, 6, 13
10	0,10	0,100	0,1	13	1, 2, 3, 6, 2, 4, 12, 7, 2, 9, 11, 5, 13

A tabela 6.5 caracteriza o processo de serviço da rede. O tempo de serviço para todas as operações das classes foi considerado determinístico, ou seja, $cs_{kl}=0$, $k=1,\dots,10$; $l=1,\dots,n_k$, conforme em Bitran e Sarkar (1993c).

TABELA 6.5 – Parâmetros referentes ao tempo médio de serviço da classe k e operação l , $E(s_{kl})$, para as 10 classes de produtos (Rede 4).

Classe	Operação (hora/item)												
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0,2	0,1	0,2	0,2	0,9	3,4	3,8						
2	0,6	0,2	3,0	0,6	1,8	0,9	0,8	2,4					
3	0,8	0,4	1,8	3,5	0,6	0,4	1,0	0,6					
4	1,8	0,4	1,8	0,2	0,2	0,4	0,8	0,4					
5	0,5	0,3	1,5	0,6	0,2	0,8	0,5	0,6					
6	0,6	0,7	0,7	2,6	0,7	0,8	0,4	3,0					
7	0,7	0,2	1,0	1,2	0,3	1,8	0,3	0,7					
8	0,3	0,2	0,6	1,8	0,2	0,7	0,7	0,6	2,2	0,4	0,8	1,3	
9	1,0	0,8	8,0	15,0	1,2	2,8	2,8	1,2	14,0	1,6	9,6	2,8	3,6
10	1,3	0,8	21,6	8,8	1,2	12,0	4,0	14,0	0,8	9,2	1,6	15,0	6,8

6.1 Problema de Redução de Complexidade 1 (PRC-1)

Nesta seção são apresentados e discutidos os resultados para o PRC-1 aplicados em uma rede formada por um único e por vários estágios $GI/G/m$.

6.1.1 Rede Formada por Um Estágio Único $GI/G/m$

O PRC-1 (modelo (102a)– (102d) com restrição (78)) foi solucionado empregando os algoritmos 6 e 7 (seção 5.4), que geram as partições, e com o algoritmo proposto por Boxma et al (1990) para alocar capacidade às linhas de produção. Nesta seção são apresentados e discutidos os resultados para uma rede formada por um estágio único.

6.1.1.1 Experimentos com Rede Ilustrativa (Rede 1)

Com o objetivo de ilustrar situações em que a partição da instalação em linhas de produção é vantajosa em relação a não particionar, inicialmente aplicou-se o PRC-1 à Rede 1 com $T^*=1,0$. Note na tabela 6.1 que a classe 2 visitando a Rede 1 é bem comportada, ou seja, a variabilidade do processo de chegadas é pequena comparada à classe 1.

A tabela 6.6 mostra o número mínimo de máquinas requeridas e a correspondente variância ao aplicar o algoritmo 6 na Rede 1. O algoritmo pára com a partição [1],[2], com 2 máquinas. A primeira coluna da tabela indica a iteração do algoritmo, as demais colunas $L1$, $L2$ indicam as classes de produtos contidas em cada uma das linhas e o número de máquinas (entre parênteses) obtido para cada linha, a coluna MQ indica o número total de máquinas obtido pelo algoritmo e a coluna $V(T)$ apresenta a correspondente variância ponderada do *leadtime* da rede, T .

TABELA 6.6 – Número de máquinas (MQ) e variância da rede $V(T)$ obtidos para as diferentes partições em cada iteração (*iter*) para $T^*= 1,0$ na Rede 1.

<i>Iter</i>	Classes de produtos alocados às linhas de produção ($L1$, $L2$) em cada partição e número de máquinas em cada linha – entre parênteses		MQ	$V(T)$
	$L1$	$L2$		
0	1 (1)	2 (1)	2	0,804

Nota-se na partição [1],[2] da tabela 6.6 que são necessárias no mínimo 2 máquinas (uma em cada linha) para que a variância ponderada do *leadtime* da rede não seja superior a 1,0 ($V(T)=0,804$). Se os dois produtos forem agregados em uma linha única, a partição [1,2] requer, no mínimo, 3 máquinas para que a variância do *leadtime* seja inferior a 1,0 (neste caso, $V(T)=0,018$). Quando o número de máquinas é igual a 2,

obtém-se $V(T)=2,07$, requerendo, portanto, uma terceira máquina para que $V(T) \leq T^*=1,0$. Logo, este exemplo ilustra uma situação em que particionar a instalação é mais vantajoso que considerar uma linha única para o problema PRC-1 em uma rede formada por uma única estação.

A tabela 6.7 mostra algumas medidas obtidas para as partições [1],[2] e [1,2] da Rede 1 quando o número de máquinas na rede é 2.

TABELA 6.7 – Medidas obtidas para as partições [1],[2] e [1,2], com número de máquinas na Rede 1 igual a 2.

Medidas	Partição [1],[2]	Partição [1,2]
$V(T)$	0,804	2,070
$E(L_1)$ – itens	4,000	4,869
$E(L_2)$ – itens	1,305	1,917
$E(T_1)$ – horas	1,000	1,217
$E(T_2)$ – horas	1,305	1,917
$E(T)$ – horas	1,061	1,357

Nota-se na tabela 6.6 que tanto a variância do *leadtime* quanto o número de itens na rede e o *leadtime* médio aumentaram quando a Rede 1 não foi particionada, mantendo o número de máquinas igual a 2. Ou seja, para uma mesma capacidade (2 máquinas), particionar a linha é melhor do que não particionar do ponto de vista da variância do *leadtime* (0,804 ao invés de 2,070), do ponto de vista da média do *leadtime* (1,000 e 1,305 horas, ao invés de 1,217 e 1,917 horas, respectivamente, para as classes 1 e 2, e 1,061 e 1,357 horas para a rede considerando as configurações [1],[2] e [1,2], respectivamente), e do ponto de vista do número de itens na rede (5,305 itens, ao invés de 6,786 itens, respectivamente). Convém salientar que este não é um resultado intuitivamente esperado, pelo menos para o autor.

Com o objetivo de avaliar a sensibilidade da variância e da média do *leadtime*, assim como do inventário em processo médio com ca'_k , variou-se ca'_2 de 0,1 para 0,4 na Rede 1, mantendo o número de máquinas na rede constante e igual a 2. A rede foi analisada em duas configurações: uma e duas linhas (respectivamente, partições [1,2] e [1],[2]). A tabela 6.8 mostra os resultados.

TABELA 6.8 – Variância do *leadtime* da rede, $V(T)$, inventário em processo, $E(L_k)$, e *leadtime*, $E(T_k)$, para as classes 1 e 2 e para a rede, $E(T)$, para duas partições da Rede 1 com $ca'_2 = 0,4$ e número de máquinas igual a 2.

Medidas	Partição [1],[2]	Partição [1,2]
$V(T)$	3,264	2,180
$E(L_1)$ – itens	4,000	4,975
$E(L_2)$ – itens	2,520	1,944
$E(T_1)$ – horas	1,000	1,244
$E(T_2)$ – horas	2,520	1,944
$E(T)$ – horas	1,304	1,384

Nota-se na tabela 6.8 que se as duas linhas forem agregadas em uma única, a variância do *leadtime* é reduzida de 3,264 para 2,180 horas, mas o número de itens na rede aumenta de 6,520 para 6,919 itens (diferentemente do caso apresentado na tabela 6.7), assim como o *leadtime* médio da rede (de 1,304 para 1,384 horas). Em outras palavras, se a rede original (com $ca'_2=0,1$) sofrer alterações no processo de chegadas de modo a aumentar o ca'_2 (por exemplo, de 0,1 para 0,4), e se o gestor não puder investir em aquisição de máquinas, uma alternativa seria simplesmente agregar as duas linhas em uma única, se o objetivo for reduzir a variância do *leadtime*. Caso contrário, a partição da rede em duas linhas aumentaria a variância do *leadtime*, mas reduziria o número médio de itens na rede e o *leadtime* médio da rede.

Outros experimentos foram realizados na Rede 1 com o objetivo de avaliar o efeito da variabilidade do processo de chegadas externas na variância do *leadtime* e, conseqüentemente, na configuração da rede. Experimentos com as partições [1],[2] e [1,2] foram conduzidos, mantendo o número de máquinas constante (2 máquinas), conforme mostra a tabela 6.9.

TABELA 6.9 – Variância do *leadtime*, $V(T)$, ao mudar a variabilidade do processo de chegadas da Rede 1, mantendo o número de máquinas da rede igual a 2.

Variabilidade do processo de chegadas externas	$V(T)$	
	Partição [1],[2]	Partição [1,2]
Base: $ca'_1=2,0$ e $ca'_2=0,1$	0,804	2,070
a) $ca'_1=0,1$ e $ca'_2=0,1$	0,166	0,240
b) $ca'_1=0,3$ e $ca'_2=0,3$	1,491	0,370
c) $ca'_1=0,5$ e $ca'_2=0,5$	4,141	0,544
d) $ca'_1=1,0$ e $ca'_2=1,0$	16,563	1,103
e) $ca'_1=1,5$ e $ca'_2=1,5$	37,266	1,860
f) $ca'_1=0,1$ e $ca'_2=2,0$	65,612	0,526

Observa-se na tabela 6.9 que, quando a variabilidade do processo de chegadas externas para as duas classes é pequena (tendendo a um processo de chegadas determinístico) e bastante similares, convém particionar a instalação (caso (a)) em termos da variância do *leadtime* da rede. Já quando a variabilidade do processo de chegadas assume valores maiores (por exemplo, casos (b), (c), (d) e (e)) e similares, a partição da instalação é menos conveniente que a agregação das linhas. O caso (f) mostra que o fato dos coeficientes serem bastante distintos, isoladamente, não definem a partição [1],[2] como sendo mais conveniente em termos de variância do *leadtime* que a partição [1,2], conforme foi observado no caso base. Por isso, outros experimentos com a Rede 1 foram conduzidos de modo a avaliar a influência do nível de ocupação das estações (balanceamento das estações) ao particionar ou não a instalação. O nível de utilização das estações para as partições [1],[2] e [1,2] em todos os casos da tabela 6.9 são $\rho_1=0,8$, $\rho_2=0,9$ e $\rho=0,85$, respectivamente. Os experimentos ilustrados na tabela 6.10 foram realizados, mantendo o número de máquinas constante.

TABELA 6.10 – Variância do *leadtime*, $V(T)$, ao mudar o nível de utilização das estações da Rede 1, mantendo o número de máquinas da rede igual a 2.

Partição [1],[2]		Partição [1,2]	
Utilização das estações	$V(T)$	Utilização das estações	$V(T)$
Base: $\rho_1=0,8$ e $\rho_2=0,9$	0,804	Base: $\rho=0,85$	2,070
a) $\rho_1=0,4$ e $\rho_2=0,4$	0,005	a) $\rho=0,40$	0,002
b) $\rho_1=0,6$ e $\rho_2=0,6$	0,053	b) $\rho=0,60$	0,043
c) $\rho_1=0,7$ e $\rho_2=0,7$	0,173	c) $\rho=0,70$	0,172
d) $\rho_1=0,8$ e $\rho_2=0,8$	0,666	d) $\rho=0,80$	0,771
e) $\rho_1=0,4$ e $\rho_2=0,9$	0,168	e) $\rho=0,65$	0,178
f) $\rho_1=0,8$ e $\rho_2=0,4$	0,640	f) $\rho=0,60$	0,027
g) $\rho_1=0,4$ e $\rho_2=0,8$	0,030	g) $\rho=0,60$	0,081

Observando os casos (a), (b) e (c), pode-se notar que convém agregar os produtos, em termos de variância de *leadtime*, quando as linhas estão balanceadas. Entretanto, quando a utilização para as estações da partição [1],[2] atinge níveis mais altos como 80% (vide caso (d)), particionar a instalação torna-se mais vantajoso. O caso (f) ilustra que, quando as estações da partição [1],[2] estão desbalanceadas e a estação com a classe menos comportada (aquela com maior variabilidade no processo de chegadas, vide tabela 6.1) possui nível de utilização alto, então é conveniente não particionar a instalação. Por outro lado, se as estações da partição [1],[2] estão desbalanceadas, mas o nível de ocupação da estação menos comportada está baixo (vide casos (e) e (g)), então convém particionar. Note no caso (f) que ao agregar as duas linhas em uma única (partição [1,2]), a variância do *leadtime* é reduzida porque a partição [1,2] apresenta maior flexibilidade operacional que a partição [1],[2]. Ora, a linha [2] que processa o produto bem comportado (produto 2) também apresenta baixo nível de utilização ($\rho_2=0,4$). A linha [1] processa o produto cuja variabilidade do processo de chegadas é alta (produto 1) e também apresenta nível de utilização ($\rho_1=0,8$) mais alto que a linha [2]. Ao agregar as linhas [1] e [2], a linha resultante ([1,2]) apresenta nível de utilização intermediário ($\rho=0,60$), porém com maior flexibilidade operacional, uma vez que agora são 2 máquinas. Esta mesma evidência pode ser observada no caso (f) da tabela 6.9 ao agregar a linha 1, que processa o produto mais comportado e com menor nível de utilização, à linha 2, que processa o produto menos comportado e com maior nível de utilização.

De uma forma geral, observa-se na Rede 1 que é conveniente particionar as classes (partição [1],[2]) quando pelo menos uma das classes tem baixo ca'_k (isto é, ela é bem comportada em termos de variabilidade), e tem alto nível de utilização (veja casos base, (d), (e) e (g) na tabela 6.10). Nestes casos, a agregação das classes (partição [1,2]) não contribui para o aumento da flexibilidade operacional da linha agregada e, conseqüentemente, para a redução da variância do *leadtime*.

6.1.1.2 Experimentos com Rede Mais Realista (Rede 3)

O passo seguinte foi verificar as observações obtidas na Rede 1 em um exemplo de rede mais realista, com um estágio único. Por isso, experimentos foram conduzidos na Rede 3 (vide figura 6.3 e tabela 6.3). Ao fixar o limitante superior para a variância ponderada do *leadtime* da rede em $T^*=0,2$ no modelo PRC-1 e utilizar o algoritmo 6, o número de máquinas para as linhas e a variância da rede para as diferentes configurações geradas em cada iteração do algoritmo são mostradas na tabela 6.11.

Nota-se na tabela 6.11 que, no passo 0 do algoritmo 6 ($iter=0$), cada classe de produto é atribuída a uma linha de produção (ou seja, 10 partições), e que o número de máquinas requeridas para que a variância da rede seja inferior ao limitante superior de 0,2 é 24. A variância de *leadtime* para esta configuração foi 0,107. Note em $iter=1$ que a agregação das classes de produtos 1 e 2 foi a que resultou na maior redução no número de máquinas (economia de 24 para 22 máquinas), satisfazendo o limitante superior para a variância ($V(T) = 0,110 < T^*$) e as demais restrições do modelo PRC-1. Note também que, em cada iteração do algoritmo, duas linhas são agregadas em uma, reduzindo em um o número de linhas (tabela 6.11). O algoritmo 6, conforme proposto por Bitran e Sarkar (1993c), interrompe o processo iterativo de busca por novas partições quando elas não geram redução no número de máquinas. Por isso, a partição [1,2], [3,4,5,6], [7,8,9], [10], requerendo o menor número de máquinas, 16, e variância do *leadtime* igual a 0,179, foi a última encontrada pelo algoritmo na iteração 6.

TABELA 6.11 – Número total de máquinas (MQ) e variância do *leadtime* da rede $V(T)$ obtida para as diferentes partições em cada iteração (*iter*) do algoritmo 6 para limitante superior da variância do *leadtime* $T^*=0,2$ (Rede 3).

<i>Iter</i>	Classes de produtos alocados às linhas de produção ($L1, \dots, L10$) em cada iteração e número de máquinas em cada linha – entre parênteses										MQ	$V(T)$
	$L1$	$L2$	$L3$	$L4$	$L5$	$L6$	$L7$	$L8$	$L9$	$L10$		
0	1 (4)	2 (5)	3 (3)	4 (2)	5 (3)	6 (3)	7 (1)	8 (1)	9 (1)	10 (1)	24	0,107
1	1-2 (7)		3 (3)	4 (2)	5 (3)	6 (3)	7 (1)	8 (1)	9 (1)	10 (1)	22	0,110
2	1-2 (7)		3 (3)	4 (2)	5-6 (4)		7 (1)	8 (1)	9 (1)	10 (1)	20	0,173
3	1-2 (7)		3-4 (4)		5-6 (4)		7 (1)	8 (1)	9 (1)	10 (1)	19	0,200
4	1-2 (7)		3-4 5-6 (7)				7 (1)	8 (1)	9 (1)	10 (1)	18	0,174
5	1-2 (7)		3-4 5-6 (7)				7-8 (1)		9 (1)	10 (1)	17	0,158
6	1-2 (7)		3-4 5-6 (7)				7-8 9 (1)			10 (1)	16	0,179

Se, ao invés de parar o algoritmo neste ótimo local (16 máquinas) conforme em Bitran e Sarkar (1993c), o processo de agregação de produtos nas linhas continuar a partir da iteração 6 da tabela 6.11, obtém-se as partições com 16 máquinas da tabela 6.12. Note na tabela que cada partição corresponde à agregação mais econômica de duas linhas da partição anterior, até atingir a partição única. Note também que, ao continuar o processo de agregação de produtos nas linhas, o número de máquinas requeridas se manteve em 16, e a variância ponderada do *leadtime* da rede diminuiu para as respectivas partições. Estes resultados mostram que, para este exemplo de rede, não particionar (ou seja, apenas uma partição) é melhor do que particionar (ou seja, para as mesmas 16 máquinas, obtém-se menor variância do *leadtime*: 0,111). Também mostram que, se ao invés de interromper o algoritmo em um mínimo local, continuá-lo enquanto a solução não piorar (isto é, sem aumentar o número de máquinas), pode resultar em soluções melhores, que eventualmente caminham até a

decisão de não particionar (isto não foi notado em Bitran e Sarkar, 1993c).

TABELA 6.12 – Número de máquinas (MQ) e variância do *leadtime* da rede $V(T)$ obtidos para partições fixas (PF) e $T^*=0,2$ (Rede 3).

PF	Classes de produtos alocados às linhas de produção em cada partição e número de máquinas nas linhas – entre parênteses										MQ	$V(T)$
	$L1$	$L2$	$L3$	$L4$	$L5$	$L6$	$L7$	$L8$	$L9$	$L10$		
1	1-2 (7)		3-4 5-6 (7)				7-8 9-10 (2)				16	0,155
2	1-2 3-4 5-6 (14)						7-8 9-10 (2)				16	0,125
3	1-2 3-4 5-6 7-8 9-10 (16)										16	0,111

A tabela 6.13 mostra o *leadtime* e o número médio de itens para cada classe obtidos para as partições [1,2],[3-6],[7-9],[10] (obtida pelo algoritmo 6) e [1-10] (obtida por agregação de todos os produtos em uma única linha, vide tabela 6.12). Note que a partição com uma linha única apesar de apresentar menor variância de *leadtime* (vide tabelas 6.11 e 6.12), apresenta maior número de itens na rede (11,425 para a partição com um única linha, enquanto a partição formada por 4 linhas apresenta 11,413 itens), e *leadtime* médio da rede aproximadamente igual à rede particionada em 4 linhas (6,830 e 6,831 horas para as partições [1,2],[3-6],[7-9],[10] e [1-10], respectivamente).

TABELA 6.13 – *Leadtime*, $E(T_k)$, e número médio de itens, $E(L_k)$, obtidos para as classes e para a rede ($E(T)$ e $E(L)$) nas partições [1,2],[3-6],[7-9],[10] e [1-10], com $T^*=0,2$ e número de máquinas na Rede 3 igual a 16.

Classe	Medidas de Desempenho para as Partições			
	$E(T_k)$		$E(L_k)$	
	[1,2],[3-6],[7-9],[10]	[1-10]	[1,2],[3-6],[7-9],[10]	[1-10]
1	30,846	30,650	2,313	2,299
2	28,181	27,895	2,818	2,798
3	10,406	10,424	1,301	1,303
4	7,856	7,874	1,178	1,181
5	7,816	7,834	1,368	1,371
6	7,782	7,800	1,395	1,398
7	1,268	1,272	0,243	0,244
8	1,247	1,251	0,249	0,250
9	1,228	1,232	0,276	0,277
10	1,087	1,217	0,272	0,304
Rede	6,830 horas	6,831 horas	11,413 itens	11,425 itens

Se, por outro lado, a variabilidade do processo de chegadas de algumas classes for reduzida, por exemplo, se o processo de chegadas das classes 7, 8, 9 e 10 na Rede 3 forem considerados determinísticos ($ca'_k=0$, $k=7, \dots, 10$), e o algoritmo 6 for aplicado, a variância do *leadtime* obtida, assim como o número de máquinas para as diferentes configurações, são dadas na tabela 6.14.

TABELA 6.14 – Número total de máquinas (MQ) e variância do *leadtime* da rede $V(T)$ obtida para as diferentes partições em cada iteração (*iter*) do algoritmo 6 para $T^*=0,2$ e processo de chegadas das classes 7 – 10 determinístico (Rede 3 modificada).

<i>Iter</i>	Classes de produtos alocados às linhas de produção ($L1, \dots, L10$) em cada iteração e número de máquinas em cada linha – entre parênteses										MQ	$V(T)$
	$L1$	$L2$	$L3$	$L4$	$L5$	$L6$	$L7$	$L8$	$L9$	$L10$		
0	1 (4)	2 (5)	3 (3)	4 (2)	5 (3)	6 (3)	7 (1)	8 (1)	9 (1)	10 (1)	24	0,106
1	1-2 (7)		3 (3)	4 (2)	5 (3)	6 (3)	7 (1)	8 (1)	9 (1)	10 (1)	22	0,106
2	1-2 (7)		3 (3)	4 (2)	5-6 (4)		7 (1)	8 (1)	9 (1)	10 (1)	20	0,169
3	1-2 (7)		3-4 (4)		5-6 (4)		7 (1)	8 (1)	9 (1)	10 (1)	19	0,197
4	1-2 (7)		3-4 5-6 (7)				7 (1)	8 (1)	9 (1)	10 (1)	18	0,171
5	1-2 (7)		3-4 5-6 (7)				7-8 (1)		9 (1)	10 (1)	17	0,155
6	1-2 (7)		3-4 5-6 (7)				7-8 9 (1)			10 (1)	16	0,155
7	1-2 (7)		3-4 5-6 (7)				7-8 9-10 (1)				15	0,155

Note que, a menos de pequenas diferenças na variância do *leadtime*, o número de máquinas e as partições até a 6ª iteração da tabela 6.14 são iguais as encontradas na tabela 6.11.

Entretanto, na 7ª iteração o algoritmo 6 obtém uma solução melhor (com apenas 15 máquinas), particionando a rede em três linhas. Se, ao invés de parar o algoritmo neste ótimo local (15 máquinas), o processo de agregação de produtos nas linhas continuar a partir da iteração 7 da tabela 6.14, obtém-se as partições apresentadas na tabela 6.15. A última delas ($PF=2$) corresponde a opção de não particionar. Note que a partição [1,2,3,4,5,6],[7,8,9,10] da tabela 6.15 requer o mesmo número de máquinas e envolve menor variância do *leadtime* ($V(T)=0,125$) do que a partição do algoritmo 6 (vide tabela 6.14), o que mostra mais uma vez as vantagens de não parar o algoritmo em um ótimo local. Os resultados da tabela 6.15 mostram que, para este exemplo de rede,

particionar ainda é melhor que não particionar, diferente do exemplo anterior. A partição [1,2,3,4,5,6],[7,8,9,10], apesar de apresentar uma variância do *leadtime* um pouco maior ($V(T)=0,125$) que a opção de não particionar, [1-10], com $V(T)=0,104$, requer 15 máquinas, enquanto que a opção de não particionar requer 16 máquinas.

TABELA 6.15 – Número de máquinas (MQ) e variância do *leadtime* da rede $V(T)$ obtidos para partições fixas (PF), com $T^*=0,2$ e processo de chegadas das classes 7 – 10 determinístico (Rede 3 modificada).

PF	Classes de produtos alocados às linhas de produção ($L1, \dots, L10$) em cada partição e número de máquinas em cada linha – entre parênteses										MQ	$V(T)$	
	$L1$	$L2$	$L3$	$L4$	$L5$	$L6$	$L7$	$L8$	$L9$	$L10$			
1	1-2 3-4 5-6 (14)						7-8 9-10 (1)					15	0,125
2	1-2 3-4 5-6 7-8 9-10 (16)											16	0,104

É importante observar que a agregação das duas linhas [1,2,3,4,5,6] e [7,8,9,10] numa única linha não foi vantajosa em termos de número de máquinas (vide tabela 6.15), uma vez que o níveis de ocupação das estações das linhas foram, respectivamente, 0,731 e 0,946, e os coeficientes de variação do processo de chegadas foram 0,144 e 0,000, respectivamente. Ora, esta situação é similar ao caso (a) da Rede 1 em que particionar os produtos em duas linhas foi mais conveniente do que a agregá-los em uma linha única (vide tabela 6.9), quando os níveis de utilização das linhas são altos e os coeficientes de variação são baixos, tendendo a processos de chegadas determinísticos.

Outros experimentos foram realizados com a Rede 3. Se, ao invés de considerar os processos de chegadas das classes 7-10 como determinísticos, os processos de chegadas das classes 1 e 2 forem admitidos como determinísticos no algoritmo 6, o número de máquinas e a variância do *leadtime* para as diferentes configurações obtidas em cada iteração são dados na tabela 6.16.

TABELA 6.16 – Número total de máquinas (MQ) e variância do *leadtime* da rede $V(T)$ obtida para as diferentes partições em cada iteração (*iter*) do algoritmo 6 para $T^*=0,2$ e processo de chegadas das classes 1 e 2 determinístico (Rede 3 modificada).

<i>Iter</i>	Classes de produtos alocados às linhas de produção ($L1, \dots, L10$) em cada iteração e número de máquinas em cada linha – entre parênteses										MQ	$V(T)$
	$L1$	$L2$	$L3$	$L4$	$L5$	$L6$	$L7$	$L8$	$L9$	$L10$		
0	1 (3)	2 (3)	3 (3)	4 (2)	5 (2)	6 (3)	7 (1)	8 (1)	9 (1)	10 (1)	20	0,161
1	1 (3)	2 (3)	3-6 (4)	4 (2)	5 (2)		7 (1)	8 (1)	9 (1)	10 (1)	18	0,164
2	1 (3)	2 (3)	3-6 (4)	4 (2)	5 (2)		7-8 (1)		9 (1)	10 (1)	17	0,200

O algoritmo 6 pára ao agregar 2 linhas e requerer 17 máquinas para atingir variância do *leadtime* da rede igual a 0,200 (iteração 2 da tabela 6.16). Se, novamente ao invés de parar o algoritmo neste ótimo local (17 máquinas), o processo de agregação de produtos nas linhas continuar a partir da iteração 2 da tabela 6.16, obtém-se, entre outras, as partições da tabela 6.17, que também inclui a opção de não particionar ($PF=7$), com os respectivos números de máquinas e variâncias do *leadtime*.

Note que a partição $PF=4$ (tabela 6.17) requer apenas 15 máquinas (2 a menos do que a obtida pelo algoritmo 6 na tabela 6.16) e apresenta a menor variância do *leadtime* ($V(T)=0,035$) das partições mostradas na tabela 6.17. Ao agregar as linhas [1,2] e [3,4,5,6,7,8,9,10] em uma linha única, o número de máquinas aumenta de 15 ($PF=6$) para 16 ($PF=7$). Ora, a linha [1,2] apresenta coeficiente de variação do processo de chegadas e nível de utilização da estação iguais a 0,000 e 0,845, respectivamente, enquanto a linha [3,4,5,6,7,8,9,10] apresenta 0,130 e 0,678, respectivamente. Novamente, esta situação é similar ao caso da Rede 1: a agregação da linha [1,2], processando produtos bem comportados (processo de chegada determinístico) e com alto nível de utilização, com a outra linha [3,4,5,6,7,8,9,10], menos comportada e com nível de utilização mais baixo, resulta em aumento na variância do *leadtime* da rede e aumento do número de máquinas (compare $PF=6$ e 7 na tabela 6.17).

TABELA 6.17 – Número de máquinas (MQ) e variância do *leadtime* da rede $V(T)$ obtidos para partições fixas (PF) com $T^*=0,2$ e processo de chegadas das classes 1 e 2 determinístico (Rede 3 modificada).

PF	Classes de produtos alocados às linhas de produção ($L1, \dots, L10$) em cada partição e número de máquinas em cada linha – entre parênteses										MQ	$V(T)$
	$L1$	$L2$	$L3$	$L4$	$L5$	$L6$	$L7$	$L8$	$L9$	$L10$		
1	1 (3)	2 (3)	3-4-6 (5)		5 (3)		7-8 (1)		9 (1)	10 (1)	17	0,125
2	1 (3)	2 (3)	3-4 5-6 (7)				7-8 (1)		9 (1)	10 (1)	16	0,039
3	1 (3)	2 (3)	3-4 5-6 (7)				7-8 9 (1)			10 (1)	15	0,067
4	1 (3)	2 (3)	3-4 5-6 (7)				7-8 9-10 (2)				15	0,035
5	1 (3)	2 (3)	3-4 5-6 7-8 9-10 (9)								15	0,072
6	1-2 (6)		3-4 5-6 7-8 9-10 (9)								15	0,072
7	1-2 3-4 5-6 7-8 9-10 (16)										16	0,110

Convém ressaltar que a Rede 3 modificada reforça o resultado anterior obtido com a Rede 1: particionar uma rede pode ser melhor, do ponto de vista do número de máquinas e da variância do *leadtime* do que não particionar (compare os resultados para $PF=4$ e $PF=7$ na tabela 6.17).

A tabela 6.18 mostra resultados para as partições [1,2],[3-6],[7-10] e [1,2],[3-10] para a Rede 3 modificada (processos de chegadas das classes 1 e 2 admitidos determinísticos). Nota-se que ao aumentar a variância do *leadtime* de 0,035

(partição [1,2],[3-6],[7-10]) para 0,072 (partição [1,2],[3-10]), mantendo a capacidade (15 máquinas), o número médio de itens na rede, assim como o *leadtime* médio da rede sofrem pequeno aumento.

TABELA 6.18 – *Leadtime*, $E(T_k)$, e número médio de itens, $E(L_k)$, obtidos para as classes e para a rede ($E(T)$ e $E(L)$) nas partições [1,2],[3-6],[7-10] e [1,2],[3-10], com $T^*=0,2$ e número de máquinas na Rede 3 modificada (classes 1 e 2 com processo de chegadas determinístico) igual a 15.

Classe	Medidas de Desempenho para as Partições			
	$E(T_k)$		$E(L_k)$	
	[1,2],[3-6],[7-10]	[1,2],[3-10]	[1,2],[3-6],[7-10]	[1,2],[3-10]
1	30,846	30,846	2,313	2,313
2	28,181	28,181	2,818	2,818
3	10,406	10,404	1,301	1,300
4	7,856	7,854	1,178	1,178
5	7,816	7,813	1,368	1,367
6	7,782	7,780	1,395	1,395
7	1,132	1,252	0,217	0,240
8	1,111	1,230	0,222	0,246
9	1,092	1,212	0,246	0,273
10	1,077	1,197	0,269	0,299
Rede	6,778 horas	6,839 horas	11,327 itens	11,429 itens

Analogamente à tabela 6.18, a tabela 6.19 mostra os resultados para duas partições da Rede 3 modificada: a partição [1,2],[3-6],[7-10] com 15 máquinas, e a partição [1-10] com 16 máquinas. Nota-se que a partição formada por 3 linhas apresenta menor variância do *leadtime*, menor número de itens na rede e menor *leadtime* médio da rede do que a opção de não particionar os produtos (linha única) que apresenta uma máquina a mais que a opção de particionar a instalação em 3 linhas. Este resultado não era intuitivamente esperado pelo autor.

TABELA 6.19 – *Leadtime*, $E(T_k)$, e número médio de itens, $E(L_k)$, obtidos para as classes e para a rede ($E(T)$ e $E(L)$) nas partições [1,2],[3-6],[7-10] e [1-10], com $T^*=0,2$ na Rede 3 modificada (classes 1 e 2 com processo de chegadas admitido determinístico).

Classe	Medidas de Desempenho para as Partições			
	$E(T_k)$		$E(L_k)$	
	[1,2],[3-6],[7-10]	[1-10]	[1,2],[3-6],[7-10]	[1-10]
1	30,846	30,647	2,313	2,299
2	28,181	27,982	2,818	2,798
3	10,406	10,422	1,301	1,303
4	7,856	7,872	1,178	1,181
5	7,816	7,831	1,368	1,370
6	7,782	7,798	1,395	1,398
7	1,132	1,270	0,217	0,243
8	1,111	1,248	0,222	0,250
9	1,092	1,230	0,246	0,277
10	1,077	1,215	0,269	0,304
Rede	6,778 horas	6,835 horas	11,327 itens	11,423 itens

6.1.2 Rede Formada por Vários Estágios ($GI/G/m$)

Nesta seção são apresentados e discutidos os resultados referentes à aplicação do algoritmo 6 ao PRC-1 em redes envolvendo vários estágios $GI/G/m$. Inicialmente o algoritmo foi aplicado em uma rede ilustrativa mais simples (Rede 2), e então em uma rede mais realista (Rede 4), para confirmação dos resultados obtidos com a rede mais simples.

6.1.2.1 Experimentos com Rede Ilustrativa (Rede 2)

O algoritmo 6 foi aplicado à Rede 2 com limitante superior para a variância do *leadtime* $T^*=0,2$. Conforme a tabela 6.2, a classe de produtos 2 é mais comportada ($ca'_2=0,100$) que a classe 1 ($ca'_1=0,500$), e também possui tempos médios

de serviço maiores.

A tabela 6.20 ilustra as configurações obtidas ao aplicar o algoritmo 6 na Rede 2 com os respectivos número de máquinas e variância do *leadtime* da rede.

TABELA 6.20 – Número total de máquinas (MQ) e variância do *leadtime* da rede $V(T)$ obtidos para as diferentes partições em cada iteração (*iter*) do algoritmo 6 para $T^*=0,2$ (Rede 2).

<i>Iter</i>	Classes de produtos alocados às linhas de produção ($L1$, $L2$) em cada partição e número de máquinas em cada linha – entre parênteses		MQ	$V(T)$
	$L1$	$L2$		
0	1 (2)	2 (2)	4	0,067

O algoritmo pára com a partição [1],[2], cuja variância do *leadtime* igual a 0,067 (menor que $T^*=0,2$) é alcançada com 4 máquinas. A tabela 6.21 mostra os coeficientes de variação dos processos de chegadas e o nível de utilização das estações das duas linhas, obtidos com o método de decomposição (capítulo 2).

TABELA 6.21 – Coeficiente de variação do processo de chegadas e nível de utilização das estações das linhas [1] e [2] (Rede 2).

Estação j	Partição [1]		Partição [2]	
	ca_j	ρ_j	ca_j	ρ_j
1	0,500	0,60	0,100	0,85
2	0,320	0,60	0,028	0,85

Note na tabela 6.21 que a partição [2] apresenta estações com alto nível de utilização e coeficientes de variação do intervalo de tempo entre chegadas baixos, se comparados à partição [1]. Espera-se, portanto, que o processo de agregação das linhas [1] e [2] em uma única, [1,2], seja desvantajoso, conforme foi observado com as Redes 1 e 3.

A tabela 6.22 confirma este fato. Ao agregar as duas linhas em uma única, o número de máquinas necessárias para que $V(T) \leq 0,2$ aumentou de 4 para 5. Ora, conforme observado para as Redes 1 e 3, a agregação de duas linhas é desvantajosa quando pelo menos uma delas possui baixa variabilidade no processo de chegadas (tendendo a um processo de chegadas determinístico) e suas estações operam com alto nível de utilização. Ou seja, a partição da linha em duas linhas na Rede 2 reduz o

número de máquinas (de 5 para 4) necessárias para manter a variância do *leadtime* abaixo de $T^*=0,2$. Além disso, particionar a rede reduz a variância do *leadtime* (de 0,097 para 0,067, conforme as tabelas 6.20 e 6.22), o número médio de itens na rede e o *leadtime* médio da rede (vide tabela 6.23).

TABELA 6.22 – Número de máquinas (MQ) e variância da rede $V(T)$ obtidos para partições fixas (PF) e $T^*=0,2$ (Rede 2).

PF	Classes de produtos alocados às linhas de produção ($L1$, $L2$) em cada partição e número de máquinas em cada linha – entre parênteses		MQ	$V(T)$
	$L1$	$L2$		
0	1-2 (5)		5	0,097

TABELA 6.23 – *Leadtime*, $E(T_k)$, e número médio de itens, $E(L_k)$, obtidos para as classes e para a rede ($E(T)$ e $E(L)$) nas partições [1],[2] e [1,2], com $T^*=0,2$ na Rede 2.

Classe	Medidas de Desempenho para as Partições			
	$E(T_k)$		$E(L_k)$	
	[1],[2]	[1,2]	[1],[2]	[1,2]
1	0,392	0,551	1,569	2,203
2	2,008	1,951	2,008	1,951
Rede	0,715 horas	0,831 horas	3,577 itens	4,154 itens

Aumentando o coeficiente de variação do processo de chegadas externas de $ca'_2=0,100$ para $ca'_2=0,250$ na Rede 2 e aplicando o algoritmo 6, nota-se pelas tabelas 6.24 e 6.25 que a partição da instalação torna-se menos vantajosa (comparado a $ca'_2=0,100$), pois necessita de 5 máquinas, assim como a opção de não particionar. A tabela 6.25 mostra a variância do *leadtime* e o número de máquinas da opção de não particionar ($PF=1$).

TABELA 6.24 – Número total de máquinas (MQ) e variância do *leadtime* da rede $V(T)$ obtidos para as diferentes partições em cada iteração (*iter*) do algoritmo 6 para $T^*=0,2$ e $ca'_2=0,250$ (Rede 2 modificada).

<i>Iter</i>	Classes de produtos alocados às linhas de produção ($L1, L2$) em cada partição e número de máquinas em cada linha – entre parênteses		MQ	$V(T)$
	$L1$	$L2$		
0	1 (2)	2 (3)	5	0,033

TABELA 6.25 – Número de máquinas (MQ) e variância do *leadtime* da rede $V(T)$ obtidos para partições fixas (PF) com $T^*=0,2$ e $ca'_2=0,250$ (Rede 2 modificada).

PF	Classes de produtos alocados às linhas de produção ($L1, L2$) em cada partição e número de máquinas em cada linha – entre parênteses		MQ	$V(T)$
	$L1$	$L2$		
1	1-2 (5)		5	0,101

Analisando as tabelas 6.24 e 6.25 conclui-se que a partição da instalação em duas linhas ainda é mais conveniente por apresentar menor variância do *leadtime*, embora a capacidade necessária em ambas as partições seja a mesma (5 máquinas). Entretanto, ao aumentar ca'_2 de 0,100 para 0,250, o número de máquinas aumentou de 4 para 5 na partição [1],[2], composta de duas linhas.

A tabela 6.26 mostra o coeficiente de variação do processo de chegadas e nível de utilização das estações das linhas [1] e [2], admitindo $ca'_2=0,250$. Note que a linha [2] já não possui nível de utilização tão alto e coeficientes de variação do processo de chegadas tendendo a determinístico (compare as tabelas 6.21 e 6.26). Por este motivo, a desagregação das linhas [1] e [2] da Rede 3 com $ca'_2=0,250$ é menos vantajosa que no caso de $ca'_2=0,100$.

TABELA 6.26 – Coeficiente de variação do processo de chegadas e nível de utilização das estações das linhas [1] e [2] com $ca'_2=0,250$ (Rede 2 modificada).

Estação j	Partição [1]		Partição [2]	
	ca_j	ρ_j	ca_j	ρ_j
1	0,500	0,600	0,250	0,425
2	0,320	0,600	0,258	0,85

Para estimar o coeficiente de variação do intervalo de tempo entre chegadas nas estações j das linhas p pelo método de decomposição, foi empregado o sistema de aproximações (111)-(113) que, conforme mencionado na seção 5.4, também foi utilizado por Bitran e Sarkar (1993c).

Outros experimentos foram conduzidos na Rede 2 empregando outros sistemas de aproximações, com o objetivo de avaliar a sensibilidade dos resultados alcançados com o sistema utilizado. Para isso, a aproximação que descreve o processo de desagregação de partidas (113) foi substituída pela (49), vide seção 2.5.4. A aproximação (49) foi desenvolvida por Bitran e Tirupati (1988) e foi a mais adequada para descrever o estudo de caso no capítulo 4, sendo por este motivo a aproximação escolhida aqui. Os resultados para a Rede 2 e para a Rede 2 modificada ($ca'_2=0,250$ ao invés de $ca'_2=0,100$) utilizando a aproximação (49) foram idênticos aos obtidos utilizando a aproximação (113) para ambas as redes. Para este exemplo de rede ilustrativa, pode-se concluir que o modelo PRC-1 é pouco sensível ao sistema de aproximações utilizado.

6.1.2.2 Experimentos com rede mais realista (Rede 4)

O algoritmo 6 foi aplicado à Rede 4 com T^* fixado em 8,0, conforme em Bitran e Sarkar (1993c). As partições com menor número de máquinas em cada iteração do algoritmo 6, com as respectivas variâncias ponderadas do *leadtime*, são mostradas na tabela 6.27.

TABELA 6.27 – Número total de máquinas (MQ) e variância do *leadtime* da rede $V(T)$ obtidos para as diferentes partições em cada iteração (*iter*) do algoritmo 6 para $T^*=8,0$ (Rede 4).

<i>Iter</i>	Classes de produtos alocados às linhas de produção ($L1, \dots, L10$) em cada iteração e número de máquinas em cada linha – entre parênteses										MQ	$V(T)$
	$L1$	$L2$	$L3$	$L4$	$L5$	$L6$	$L7$	$L8$	$L9$	$L10$		
0	1 (6)	2 (7)	3 (7)	4 (7)	5 (6)	6 (7)	7 (6)	8 (10)	9 (13)	10 (17)	86	7,698
1	1 (6)	2 (7)	3 (7)	4 (7)	5 (10)	6 (7)	7-8 (6)		9 (13)	10 (17)	80	4,740
2	1 (6)	2 (7)	3 (7)	4 (7)	5-7-8 (10)	6 (7)			9 (13)	10 (17)	74	4,752
3	1-4 (7)	2 (7)	3 (7)		5-7-8 (10)	6 (7)			9 (13)	10 (18)	69	4,798
4	1-4 (7)	2 (7)	3-5 7-8 (11)			6 (7)			9 (13)	10 (18)	63	5,582
5	1-3 4-5 7-8 (12)	2 (7)				6 (7)			9 (13)	10 (18)	57	6,759
6	1-2 3-4 5-7 8 (14)					6 (7)			9 (13)	10 (18)	52	6,917
7	1-2 3-4 5-6 7-8 (15)								9 (13)	10 (18)	46	6,516
8	1-2 3-4 5-6 7-8 (15)								9-10 (26)		41	7,908

Nota-se na tabela 6.27 que o algoritmo 6 inicia (*iter*=0) com uma classe de produto em cada linha, requerendo 86 máquinas, no mínimo, para que $V(T) \leq 8,0$ e termina (*iter*=8) na partição [1,2,3,4,5,6,7,8],[9,10], quando o número de máquinas requerido é 41 e $V(T)=7,908$. A tabela 6.28 mostra o *leadtime* médio e o número médio de itens de cada classe na rede para a partição [1,2,3,4,5,6,7,8],[9,10].

TABELA 6.28 – *Leadtime* médio, $E(T_k)$, número médio de itens para cada classe, $E(L_k)$, e total na Rede 4, $E(T)$ e $E(L)$, na partição [1,2,3,4,5,6,7,8],[9,10].

Classe	$E(T_k)$, horas	$E(L_k)$, itens
1	8,959	0,896
2	11,662	1,166
3	10,484	1,048
4	10,555	1,056
5	9,474	0,947
6	11,300	1,130
7	11,172	1,117
8	15,358	1,536
9	92,161	9,216
10	87,491	8,749
Rede	26,862 horas	26,862 itens

A tabela 6.29 mostra o número mínimo de máquinas e a variância ponderada para a rede formada por uma linha única ($PF=1$), e a tabela 6.30 mostra o *leadtime* médio e o número médio de itens de cada classe e para a rede nesta partição.

TABELA 6.29 – Número de máquinas (MQ) e variância do *leadtime* da rede $V(T)$ obtidos para partições fixas (PF) com $T^*=8,0$ (Rede 4).

PF	Classes de produtos alocados às linhas de produção ($L1, \dots, L10$) em cada partição e número de máquinas em cada linha – entre parênteses										MQ	$V(T)$
	$L1$	$L2$	$L3$	$L4$	$L5$	$L6$	$L7$	$L8$	$L9$	$L10$		
1	1-2 3-4 5-6 7-8 9-10 (41)										41	6,489

Comparando a partição referente à iteração 8 ($iter=8$) da tabela 6.27, com os resultados da tabela 6.29, pode-se constatar que a agregação das duas linhas em uma única, além de requerer as mesmas 41 máquinas, apresentou redução da variância ponderada do *leadtime* (de 7,908 para 6,489), assim como redução no número médio de itens (de 26,862 para 25,886, vide tabelas 6.28 e 6.30) e do *leadtime* médio da rede (de 26,862 para 25,886 horas, vide tabelas 6.28 e 6.30). Portanto, para a Rede 4 a partição da instalação não é vantajosa.

TABELA 6.30 – *Leadtime* médio, $E(T_k)$, número médio de itens para cada classe, $E(L_k)$, e para a Rede 4, $E(L)$ e $E(T)$, na partição [1,2,3,4,5,6,7,8,9,10].

Classe	$E(T_k)$, horas	$E(L_k)$, itens
1	10,405	1,041
2	12,742	1,274
3	11,481	1,148
4	8,819	0,882
5	7,050	0,705
6	13,133	1,313
7	8,555	0,855
8	14,569	1,457
9	69,451	6,945
10	102,658	10,266
Rede	25,886 horas	25,886 itens

Convém ressaltar que os resultados encontrados neste trabalho diferem dos resultados apresentados por Bitran e Sarkar (1993c) para a mesma rede (Rede 4) e valor de T^* ($T^*=8,0$). Bitran e Sarkar (1993c) encontraram a seguinte partição [1,2,3,4,5,6,7,8],[9,10], requerendo 73 máquinas, como sendo a que satisfaz o limitante superior para a variância ponderada com o menor número de máquinas. A razão desta diferença aparentemente foi o emprego incorreto da aproximação (113) por parte de Bitran e Sarkar (1993c). Apesar de mencionarem que a aproximação utilizada para descrever o processo de partidas corresponde a de Segal e Whitt (1989), constatou-se que a expressão citada no artigo não coincide com a do artigo de Segal e Whitt (1989), nem a qualquer outra das revisadas no capítulo 2. Isto, além de invalidar o processo de aferimento dos resultados aqui apresentados com os de Bitran e Sarkar (1993c), também sugeriu que o processo de partição da instalação para esta rede pode ser sensível ao sistema de aproximações utilizado.

Outros experimentos foram realizados com a Rede 4. A tabela 6.31 reporta o número de máquinas (MQ) e a variância ponderada do *leadtime* ($V(T)$) para a partição gerada pelo algoritmo 6, e para a opção de não particionar a Rede 4. As classes 9 e 10 foram escolhidas para alteração nos parâmetros de entrada porque, conforme a tabela 6.27, o algoritmo pára com uma linha dedicada para as classes 1 a 8 e outra dedicada às classes 9 e 10, e que ao agregar estas duas linhas o número de máquinas não é reduzido. Isso sugere que as classes de produtos 9 e 10 possuem características em

comum e que a agregação da linha a qual pertencem com a linha [1,2,3,4,5,6,7,8] é pouco vantajosa. Note que as classes 9 e 10 possuem tempos médios de serviço maiores que as demais classes, mas coeficientes de variação do processo de chegadas externas menores (vide tabelas 6.4 e 6.5). Por isso, com o objetivo de tornar as classes 9 e 10 mais distintas das demais classes, o processo de chegadas externas das classes 9 e 10 foram consideradas determinísticas, e os tempos médios de serviço em cada operação foram duplicados.

TABELA 6.31 – Número de máquinas (MQ) e variância do *leadtime* ($V(T)$) da Rede 4 com $T^*=8,0$ para a partição obtida pelo algoritmo 6 e para a partição com uma linha única.

Mudanças nos Parâmetros de Entrada das Tabelas 6.4 e 6.5	Partição obtida pelo algoritmo 6			Partição [1-10]	
	MQ	$V(T)$	Partição	MQ	$V(T)$
Caso base	41	7,908	[1-8], [9,10]	41	6,489
$ca'_9 = 0$ e $ca'_{10} = 0$	40	7,125	[1-8] [9,10]	41	6,292
$2^*E(s_{k9})$ e $2^*E(s_{k10})$	62	7,050	[1-8] [9,10]	66	6,824

Percebe-se que ao tornar o processo de chegadas das classes 9 e 10 determinística, a partição com menor capacidade (40 máquinas) é composta de duas linhas [1,2,3,4,5,6,7,8],[9,10], ao invés de uma única (41 máquinas), apesar desta diferença não ser muito significativa (menor que 2,5%). Note também que, assim como nas redes anteriores, particionar as linhas na Rede 4 modificada pode reduzir o número de máquinas necessárias para manter a variância do *leadtime* abaixo de $T^*=8,0$, em relação a não particionar a rede.

A tabela 6.32 reforça os resultados obtidos para as Redes 1 e 2: a agregação de uma linha que possui estações com alto nível de utilização e variabilidade do processo de chegadas baixo (linha [9,10]) com outra que possui variabilidade do processo de chegadas maior (linha [1,2,3,4,5,6,7,8]) não é vantajosa. A tabela 6.32 mostra o nível de utilização (ρ_j) e os coeficientes de variação do intervalo de tempo entre chegadas das estações (ca_j) das linhas [1,2,3,4,5,6,7,8] e [9,10].

TABELA 6.32 – Nível de utilização (ρ_j) e coeficiente quadrático dos intervalos de tempo entre chegadas (ca_j) das 13 estações das linhas [1,2,3,4,5,6,7,8] e [9,10] para o caso com $ca'_9 = 0$ e $ca'_{10} = 0$ (Rede 4 modificada com $T^*=8,0$).

Estação j	Partição [1,2,3,4,5,6,7,8]		Partição [9,10]	
	ca_j	ρ_j	ca_j	ρ_j
1	0,354	0,55	0,000	0,23
2	0,370	0,38	0,101	0,64
3	0,346	0,06	0,053	0,74
4	0,338	0,28	0,108	0,66
5	0,349	0,37	0,098	0,75
6	0,348	0,50	0,139	0,72
7	0,349	0,40	0,103	0,70
8	0,353	0,42	0,029	0,70
9	0,350	0,46	0,029	0,92
10	0,335	0,50	0,029	0,96
11	0,375	0,36	0,005	0,16
12	0,354	0,61	0,185	0,68
13	0,355	0,56	0,166	0,52

Note que a linha [9,10] possui estações com maior nível de utilização (exceção para as estações 1, 11 e 13) e menor coeficiente quadrático de variação do intervalo de tempo entre chegadas do que a linha [1,2,3,4,5,6,7,8]. Do caso em que $ca'_9 = 0$ e $ca'_{10} = 0$ para o caso base, houve aumento nos coeficientes de variação do intervalo de tempo entre chegadas (para a classe 9 de 0 para 0,050 e para a classe 10 de 0 para 0,100). A partição [1-10], que no caso $ca'_9 = ca'_{10} = 0$ era menos vantajosa, no caso base requer menor número de máquinas e apresenta menor variância do *leadtime* (41 máquinas e $V(T)=6,489$, vide tabela 6.31) do que a partição [1-8],[9,10] (41 máquinas e $V(T)=7,908$, vide tabela 6.31). Isto também foi observado na Rede 2 quando o coeficiente de variação do intervalo de tempo entre chegadas para a classe 2 aumentou de 0,100 para 0,250 e a partição tornou-se menos vantajosa.

Ao duplicar os tempos de processamento das classes de produtos 9 e 10, a partição [1,2,3,4,5,6,7,8],[9,10] obtida pelo algoritmo 6 necessita de 62 máquinas, enquanto que a partição [1,2,3,4,5,6,7,8,9,10] necessita de 66 máquinas (vide tabela 6.31). Portanto, para este exemplo, acentuar a diferença entre as classes 9 e 10 e as demais, duplicando os tempos médios de serviço, favoreceu mais o particionamento da

instalação do que tornar o processo de chegadas das classes 9 e 10 determinístico. A tabela 6.33 mostra os níveis de utilização, assim como os coeficientes quadráticos de variação do intervalo de tempo entre chegadas para as estações das linhas [1,2,3,4,5,6,7,8], [9,10] e [1-10].

TABELA 6.33 – Nível de utilização (ρ_j) e coeficiente quadrático de variação do intervalo de tempo entre chegadas (ca_j) nas 13 estações das linhas [1,2,3,4,5,6,7,8] e [9,10] para o caso com $2^*E(s_{k9})$ e $2^*E(s_{k10})$ (Rede 4 modificada com $T^*=8,0$).

Estação j	Partição [1,2,3,4,5,6,7,8]		Partição [9,10]		Partição [1-10]	
	ca_j	ρ_j	ca_j	ρ_j	ca_j	ρ_j
1	0,354	0,55	0,075	0,46	0,298	0,51
2	0,370	0,38	0,217	0,64	0,461	0,68
3	0,346	0,06	0,172	0,74	0,349	0,66
4	0,338	0,28	0,236	0,66	0,388	0,53
5	0,349	0,37	0,256	0,86	0,349	0,71
6	0,348	0,50	0,291	0,72	0,362	0,68
7	0,349	0,40	0,243	0,70	0,349	0,53
8	0,353	0,42	0,127	0,70	0,372	0,54
9	0,350	0,46	0,127	0,61	0,397	0,46
10	0,335	0,50	0,127	0,64	0,476	0,48
11	0,375	0,36	0,238	0,32	0,511	0,52
12	0,354	0,61	0,292	0,68	0,445	0,66
13	0,355	0,56	0,301	0,69	0,355	0,53

Note na tabela 6.33 que ao agregar as partições [1,2,3,4,5,6,7,8],[9,10] em uma única [1-10], o coeficiente quadrático do intervalo de tempo entre chegadas nas estações (ca_{jp}) aumenta, uma vez que a variação dos tempos de serviço nas estações (cs_{jp}) também aumenta com a agregação das classes 1-8, cujos tempos médios de serviço são bem menores que os produtos 9 e 10. Ora, o aumento dos coeficientes de variação dos tempos de serviço e do intervalo de tempo entre chegadas aumenta a variação do tempo de espera em fila (vide expressões (79) e (80), seção 5.2.1) que, por sua vez, aumenta a variância do *leadtime*, implicando em um número maior de máquinas para que a variância da rede seja menor que T^* . Por isso, particionar a instalação em duas linhas, [1,2,3,4,5,6,7,8],[9,10] requer um número menor de máquinas. Pode-se concluir, por este exemplo, que a agregação de produtos, com tempos de serviços bem diferentes, em uma mesma linha aumenta ca_{jp} e cs_{jp} nas estações que, por sua vez, aumenta a variância do *leadtime* e requer mais capacidade

para mantê-la a níveis inferiores a T^* . Para estes casos, a partição da instalação é vantajosa se comparada a não particionar.

Com o objetivo de avaliar a sensibilidade do modelo a mudanças no sistema de aproximações (utilizado para avaliar o desempenho da rede em diferentes configurações), também foi empregada na Rede 4 a aproximação (49), adaptada para o caso de múltiplas linhas p , em substituição à expressão (113), que descreve o processo de desagregação das partidas. A tabela 6.34 mostra os resultados obtidos ao aplicar o algoritmo 6 com a aproximação de Bitran e Tirupati (1988) à Rede 4 com $T^*=8,0$.

TABELA 6.34 – Número total de máquinas (MQ) e variância da rede $V(T)$ obtidos para as diferentes partições em cada iteração ($iter$) do algoritmo 6 para $T^*=8,0$ na Rede 4, utilizando a aproximação de Bitran e Tirupati (1988) na Rede 4.

<i>Iter</i>	Classes de produtos alocados às linhas de produção ($L1, \dots, L10$) em cada iteração e número de máquinas em cada linha – entre parênteses										MQ	$V(T)$
	$L1$	$L2$	$L3$	$L4$	$L5$	$L6$	$L7$	$L8$	$L9$	$L10$		
0	1 (6)	2 (7)	3 (7)	4 (7)	5 (6)	6 (7)	7 (6)	8 (10)	9 (13)	10 (18)	87	7,438
1	1 (6)	2 (7)	3 (7)	4 (7)	5-8 (10)	6 (7)	7 (6)		9 (13)	10 (18)	81	6,636
2	1 (6)	2 (7)	3 (7)	4 (7)	5-7 8 (10)	6 (7)			9 (13)	10 (18)	75	6,684
3	1-4 (7)	2 (7)	3 (7)		5-7 8 (10)	6 (7)			9 (14)	10 (18)	70	6,827
4	1-4 5-7 8 (12)	2 (7)	3 (7)			6 (7)			9 (14)	10 (18)	65	6,808
5	1-2 4-5 7-8 (14)		3 (7)			6 (7)			9 (14)	10 (18)	60	6,798
6	1-2 3-4 5-7 8 (16)					6 (7)			9 (14)	10 (18)	55	6,925
7	1-2 3-4 5-6								9 (14)	10 (18)	51	7,534

	7-8 (19)											
8	1-2 3-4 5-6 7-8 9 (28)									10 (18)	46	7,514
9	1-2 3-4 5-6 7-8 9-10 (42)										42	6,624

A tabela 6.34 mostra que a partição obtida pelo algoritmo 6, utilizando a aproximação de Bitran e Tirupati (1988) é a mesma obtida pelo algoritmo 6 utilizando a aproximação de Segal e Whitt (1989), conforme a tabela 6.27. Entretanto, ao utilizar a aproximação de Bitran e Tirupati (1988), foram necessárias 42 máquinas ao invés de 41. Ou seja, uma diferença menor que 2,5%, o que é aceitável para as decisões envolvidas.

A tabela 6.35 mostra o *leadtime* médio e o número médio de itens para cada classe e para a rede, para a partição [1,2,3,4,5,6,7,8,9,10].

TABELA 6.35 – *Leadtime* médio, $E(T_k)$, número médio de itens para cada classe, $E(L_k)$, e para a rede, $E(L)$ e $E(T)$, na partição [1,2,3,4,5,6,7,8,9,10] ao utilizar a aproximação de Bitran e Tirupati (1988) na Rede 4.

Classe	$E(T_k)$, horas	$E(L_k)$, itens
1	10,723	1,072
2	13,343	1,334
3	12,062	1,206
4	8,203	0,820
5	7,547	0,755
6	12,879	1,288
7	9,227	0,923
8	14,351	1,435
9	70,647	7,065
10	102,712	10,271
Rede	26,170 horas	26,170 itens

Comparando os resultados da tabela 6.30 com os da tabela 6.35, nota-se que a aproximação de Bitran e Tirupati (1988) superestima em cerca de 1% o número

médio de itens e o *leadtime* médio da rede, com relação à aproximação de Segal e Whitt (1989). Porém, convém lembrar que a partição obtida com a aproximação de Bitran e Tirupati (1988) necessita de 42 máquinas, enquanto que a obtida utilizando a aproximação de Segal e Whitt (1989) necessita de 41 máquinas.

A tabela 6.36 apresenta os resultados da partição obtida pelo algoritmo 6 e da partição com linha única, quando os coeficientes de variação do intervalo de tempo entre chegadas e tempos médios de serviços das classes 9 e 10 são alterados, utilizando a aproximação de Bitran e Tirupati (1988). A motivação para a realização destes experimentos é a comparação dos resultados utilizando a aproximação desenvolvida por Segal e Whitt (1989), com os obtidos pela aproximação de Bitran e Tirupati (1988), para a Rede 4.

TABELA 6.36 – Número de máquinas (MQ) e variância do *leadtime* ($V(T)$) ao mudar alguns parâmetros do caso base (Rede 4 com $T^*=8,0$), utilizando a aproximação de Bitran e Tirupati (1988) para a partição obtida pelo algoritmo 6 e para a partição com uma linha única.

Mudanças nos Parâmetros de Entrada da Rede 4	Partição obtida pelo algoritmo 6			Partição [1-10]	
	MQ	$V(T)$	Partição	MQ	$V(T)$
Caso base (aproximação (49), Bitran e Tirupati, 1988)	42	6,624	[1-10]	42	6,624
Aproximação (49) e $ca'_9 = 0$ e $ca'_{10} = 0$	42	6,531	[1-10]	42	6,531
Aproximação (49) e $2 * E(s_{k9})$ e $2 * E(s_{k10})$	67	7,140	[1-10]	67	7,140

Nota-se na tabela 6.36 que o resultado obtido com o algoritmo 6 coincide em todos os casos com a partição de uma linha única [1-10]. Diferentemente dos resultados apresentados na tabela 6.31, ao utilizar a aproximação proposta por Bitran e Tirupati (1988), em nenhuma situação a partição da instalação foi mais econômica do que não particionar, do ponto de vista do número de máquinas, conforme mostra a tabela 6.36. Este resultado indica uma certa sensibilidade no modelo PRC-1 proposto por Bitran e Sarkar (1993c): o modelo não é robusto com relação ao sistema de aproximações utilizado. Ora, se a aproximação utilizada é a proposta por Segal e Whitt

(1989), é conveniente particionar a rede em duas linhas ([1,2,3,4,5,6,7,8],[9,10]) para os casos $ca'_9 = 0$, $ca'_{10} = 0$ e $2 * E(s_{k9})$, $2 * E(s_{k10})$, conforme a tabela 6.31. Entretanto, se a aproximação utilizada for a proposta por Bitran e Tirupati (1988), ao particionar a instalação nas duas linhas [1,2,3,4,5,6,7,8],[9,10], o caso em que $ca'_9 = 0$ e $ca'_{10} = 0$ resulta em 48 máquinas (enquanto o caso de uma linha única resulta em 42, conforme a tabela 6.36), uma diferença maior que 10%. Já o caso em que $2 * E(s_{k9})$ e $2 * E(s_{k10})$ resulta em 69 máquinas (enquanto o caso de uma linha única resulta em 67, conforme a tabela 6.36). Diante desse resultado, a questão é: em qual resultado um tomador de decisão deveria confiar, uma vez que eles são divergentes? Em outras palavras, a instalação deve ser particionada ou não? Este resultado estimula pesquisas futuras, além do escopo do presente trabalho, envolvendo o projeto de experimentos com diferentes configurações de rede e aproximações de decomposição para avaliar e otimizar o seu desempenho. O objetivo seria identificar para redes com determinadas características (processando grande ou pequena quantidade de produtos, tráfego pesado ou leve, produtos muito ou pouco similares em termos de roteiro e de tempos de processamento, produtos com alta ou pouca variabilidade no processo de chegadas externas à rede, etc.), a aproximação que melhor avalia o desempenho da rede. A validação dos resultados obtidos pelas aproximações poderia ocorrer por meio de experimentos de simulação.

6.2 Problema de Redução de Complexidade 2 (PRC-2)

Nesta seção são apresentados e discutidos os resultados para o modelo PRC-2 aplicados em redes formadas por um único (Rede 3) e por vários (Rede 4) estágios $GI/G/m$.

6.2.1 Rede formada por um estágio único $GI/G/m$ (Rede 3)

O PRC-2 foi solucionado empregando o algoritmo 6, que gera as partições, em conjunto com o *solver* GAMS/CPLEX[®] para alocar capacidade às linhas de produção nas diferentes configurações da rede. Ou seja, uma vez definida a partição (x_{kp} conhecidos para $k=1, \dots, 10$ e $p=1, \dots, 10$) pelo algoritmo 6, o PRC-2 torna-se um problema de programação linear inteira nas variáveis m_{jp} , permitindo a utilização do

solver GAMS/CPLEX[®] para sua solução.

Ao fixar $G_{jp}=1,0$ na expressão (106) para $j=1$ e $p=1,\dots,10$, o limitante superior para o tempo médio de espera em fila, dado que espera ($E(W_{q_{jp}}|W_{q_{jp}}>0)$), fica fixado em $1/(2*1)=0,5$ horas para as estações j das respectivas linhas p , conforme expressão (97). O número de máquinas para as linhas da Rede 3 nas diferentes configurações geradas em cada iteração do algoritmo são mostradas na tabela 6.37.

TABELA 6.37 – Número total de máquinas (MQ) obtido para as diferentes partições em cada iteração (*iter*) do algoritmo 6 para $G_{jp}=1,0$ (Rede 3).

<i>Iter</i>	Classes de produtos alocados às linhas de produção ($L1, \dots, L10$) em cada iteração e número de máquinas em cada linha – entre parênteses										MQ
	$L1$	$L2$	$L3$	$L4$	$L5$	$L6$	$L7$	$L8$	$L9$	$L10$	
0	1 (7)	2 (7)	3 (3)	4 (3)	5 (3)	6 (3)	7 (1)	8 (1)	9 (1)	10 (1)	30
1	1-2 (9)		3 (3)	4 (3)	5 (3)	6 (3)	7 (1)	8 (1)	9 (1)	10 (1)	25
2	1-2 (9)		3-4 (4)		5 (3)	6 (3)	7 (1)	8 (1)	9 (1)	10 (1)	23
3	1-2 (9)		3-4 (4)		5-6 (4)		7 (1)	8 (1)	9 (1)	10 (1)	21
4	1-2 (9)		3-4 5-6 (7)				7 (1)	8 (1)	9 (1)	10 (1)	20
5	1-2 (9)		3-4 5-6 (7)				7-8 (1)		9 (1)	10 (1)	19
6	1-2 (9)		3-4 5-6 (7)				7-8 9 (1)			10 (1)	18

Na tabela 6.37, a partição com menor número de máquinas, que satisfaz $E(W_{q_{jp}}|W_{q_{jp}}>0) \leq 0,5$ e as demais restrições do PRC-2, é [1,2],[3,4,5,6],[7,8,9],[10]. A tabela 6.38 mostra o número mínimo de máquinas requeridas usando outras partições além das obtidas pelo algoritmo 6, inclusive a opção de não particionar.

TABELA 6.38 – Número total de máquinas (MQ) obtido em diferentes partições (PF) para $G_{jp}=1,0$ (Rede 3).

PF	Classes de produtos alocados às linhas de produção ($L1, \dots, L10$) em cada partição e número de máquinas em cada linha – entre parênteses										MQ
	$L1$	$L2$	$L3$	$L4$	$L5$	$L6$	$L7$	$L8$	$L9$	$L10$	
1	1-2 (9)		3-4 5-6 (7)				7-8 9-10 (2)				18
2	1-2 (9)		3-4 5-6 7-8 9-10 (10)								19
3	1-2 3-4 5-6 7-8 9-10 (23)										23

Para o PRC-2, diferentemente do resultado encontrado em PRC-1, para satisfazer a restrição de complexidade que garante a mínima flexibilidade nas estações, é mais vantajoso particionar a Rede 3 em 3 linhas ([1,2],[3,4,5,6],[7,8,9,10]). Note que o algoritmo 6 interrompeu a busca por uma partição melhor que [1,2],[3,4,5,6],[7,8,9],[10] (tabela 6.37), porque não encontra economia de máquinas a partir desta partição. Ou seja, não encontra a partição apresentada na tabela 6.38, que também necessita de 18 máquinas. Convém lembrar que a Rede 3 foi adaptada de Bitran e Sarkar (1993c). A diferença aqui é que os tempos médios de serviço e as taxas médias de chegadas de todas as classes de produtos foram, arbitrariamente, divididas por 2 e 4, respectivamente, para serem tratáveis computacionalmente no *GAMS* (conforme discussão no início deste capítulo). A partição com o menor número de máquinas obtida aqui é igual à partição reportada por Bitran e Sarkar (1993c), ou seja, [1,2],[3,4,5,6],[7,8,9,10]. Enquanto neste trabalho foram necessárias 18 máquinas, Bitran e Sarkar (1993c) necessitaram de 101 máquinas. Convém mencionar que a partição [1,2],[3,4,5,6],[7,8,9,10] agrega produtos com tempos médios de serviço similares em ordem de grandeza em uma mesma linha.

Mesmo do ponto de vista do *leadtime* médio ($E(T)$), do número médio de itens ($E(L)$) e da variância do *leadtime* na rede ($V(T)$), com $w_k=0,5$, $k=1,\dots,10$) a opção de particionar a instalação ainda é mais vantajosa que a opção de não particioná-la. Ora, para a configuração [1,2],[2-6],[7-10] ($PF=1$, tabela 6.38), $E(L)=11,285$, $E(T)=6,754$ e $V(T)=0,039$. Já para a configuração [1-10] ($PF=3$, tabela 6.38), $E(L)=11,177$, $E(T)=6,690$ e $V(T)=2,11 \cdot 10^{-6}$. Em palavras, o número médio de itens e o *leadtime* médio da rede sofrem aumentos pouco significativos ao optar por particionar a instalação (aumento de 0,108 itens e 0,064 horas). A variância ponderada do *leadtime* da rede, apesar do aumento ao optar em particionar a instalação, permanece em ordem de grandeza pouco significativa e, portanto, aceitável. Entretanto, a opção de particionar a instalação requer 18 máquinas (vide figura 6.38, $PF=1$), enquanto a opção de não particioná-la requer 23 máquinas (vide figura 6.38, $PF=3$). Note também que ao limitar o $E(W_{q_{jp}}|W_{q_{jp}}>0)$, a variância do *leadtime* da rede obtida para ambas as configurações é pequena (compare $iter=4$, tabela 6.11), resultado intuitivamente esperado uma vez que o tempo médio de espera de um item, dado que ele espera, é limitado.

A partição [1,2],[3,4,5,6],[7,8,9,10] foi simulada no *GPSS/H*[®] com o objetivo de verificar se o tempo médio de espera em fila, dado que espera, era de fato inferior a 0,5 horas, conforme hipótese do modelo PRC-2. Os resultados de simulação foram obtidos a partir de 10 réplicas aceitas pelo teste de uniformidade de qui-quadrado realizado pelo *GPSS/H*[®]. Foram gerados 200000 *jobs* para cada réplica, dos quais os 100000 primeiros foram descartados com vistas a evitar os efeitos do estado transiente, e os 100000 restantes foram utilizados para gerar as estatísticas, admitidas em estado estacionário. Como a tabela 6.4 fornece funções somente do primeiro e segundo momentos da distribuição do intervalo de tempo entre chegadas para cada classe (respectivamente, λ'_k e ca'_k , $k=1,\dots,10$), por simplicidade, foram assumidas distribuições uniformes para o processo de geração das chegadas das classes de produtos no modelo de simulação, com as correspondentes funções do primeiro e segundo momentos. A tabela 6.39 mostra os resultados.

TABELA 6.39 – Tempo médio de espera em fila, dado que espera, $E(Wq_{jp}|Wq_{jp}>0)$, e número médio de itens no sistema para as linhas ($E(L_p)$) da partição [1,2],[3,4,5,6],[7,8,9,10] obtidos por simulação para a Rede 3 com $G_{jp}=1,0$.

Medidas de desempenho	Medidas para as linhas		
	[1,2]	[3,4,5,6]	[7,8,9,10]
$E(Wq_{jp} Wq_{jp}>0)$ – hora	0,0000	0,3997	0,3495
$E(L_p)$ – item	5,07	5,16	0,99

Pode-se observar que os tempos de espera em cada linha foram inferiores a 0,5 horas, o limitante superior imposto ao especificar $G_{jp}=1,0$. Isso evidencia a eficácia da restrição de complexidade em manter a medida de desempenho $E(Wq_{jp}|Wq_{jp}>0)$ no modelo abaixo do limitante superior G_{jp} .

A tabela 6.40 compara o número médio de itens e o *leadtime* médio para cada classe na partição [1,2],[3,4,5,6],[7,8,9,10], obtidos por simulação e pelas aproximações utilizadas para avaliar o desempenho da rede sob diferentes configurações.

TABELA 6.40 – *Leadtime* médio, $E(T_k)$, e número médio de itens, $E(L_k)$, para as 10 classes de produtos e para a rede ($E(L)$ e $E(T)$) na partição [1,2],[3,4,5,6],[7,8,9,10] da Rede 3 obtidos por simulação e pelas aproximações com $G_{jp}=1,0$.

Classe	$E(T_k)$ – hora		$E(L_k)$ – item	
	Simulação	Aproximações	Simulação	Aproximações
1	30,487	30,543	2,288	2,291
2	27,823	27,878	2,784	2,788
3	10,275	10,406	1,285	1,301
4	7,739	7,856	1,161	1,178
5	7,688	7,816	1,346	1,368
6	7,654	7,782	1,372	1,395
7	1,184	1,139	0,227	0,218
8	1,127	1,118	0,225	0,224
9	1,144	1,099	0,257	0,247
10	1,139	1,084	0,285	0,271
Rede	6,717 horas (0%)	6,750 horas (0,49%)	11,230 itens (0%)	11,281 itens (0,45%)

Pode-se observar na tabela 6.40 que os resultados obtidos pelas aproximações são bastante próximos dos obtidos pela simulação (desvios relativos médios de 0,49% e de 0,45%, para o *leadtime* médio e número médio de itens na rede,

respectivamente).

Outros experimentos foram realizados com a Rede 3, alterando-se o valor de alguns parâmetros. A tabela 6.41 mostra o número de máquinas para a Rede 3, alterando-se G_{jp} de 1,0 para 0,6. Isto implica que o limitante superior para $E(W_{q_{jp}}|W_{q_{jp}}>0)$ é $1/(2*0,6)=0,833$ horas (ao invés de 0,5 horas).

TABELA 6.41 – Número total de máquinas (MQ) obtido para as diferentes partições em cada iteração ($iter$) do algoritmo 6 para $G_{jp}=0,6$ (Rede 3).

<i>Iter</i>	Classes de produtos alocados às linhas de produção ($L1, \dots, L10$) em cada iteração e número de máquinas em cada linha – entre parênteses										MQ
	$L1$	$L2$	$L3$	$L4$	$L5$	$L6$	$L7$	$L8$	$L9$	$L10$	
0	1 (5)	2 (6)	3 (3)	4 (2)	5 (3)	6 (3)	7 (1)	8 (1)	9 (1)	10 (1)	26
1	1-2 (8)		3 (3)	4 (2)	5 (3)	6 (3)	7 (1)	8 (1)	9 (1)	10 (1)	23
2	1-2 (8)		3-5 (4)	4 (2)		6 (3)	7 (1)	8 (1)	9 (1)	10 (1)	21
3	1-2 (8)		3-5-6 (5)	4 (2)			7 (1)	8 (1)	9 (1)	10 (1)	19
4	1-2 (8)		3-4 5-6 (6)				7 (1)	8 (1)	9 (1)	10 (1)	18
5	1-2 (8)		3-4 5-6 (6)				7-8 (1)		9 (1)	10 (1)	17
6	1-2 (8)		3-4 5-6 (6)				7-8 9 (1)			10 (1)	16

Observe que a partição em $iter=6$ é a mesma obtida para a tabela 6.37, mas o número de máquinas necessárias para $G_{jp}=0,6$ é menor, conforme esperado, uma vez que é permitido que o tempo de espera de um item, dado que ele espera, seja maior. A tabela 6.42 mostra que o algoritmo cessou a busca por uma partição melhor ($[1,2],[3,4,5,6],[7,8,9,10]$) porque o número de máquinas se manteve constante e não houve economia. A tabela inclui também a opção de não particionar ($PF=3$).

TABELA 6.42 – Número total de máquinas (MQ) obtido em diferentes partições (PF) para $G_{jp}=0,6$ (Rede 3).

PF	Classes de produtos alocados às linhas de produção ($L1, \dots, L10$) em cada partição e número de máquinas em cada linha – entre parênteses										MQ
	$L1$	$L2$	$L3$	$L4$	$L5$	$L6$	$L7$	$L8$	$L9$	$L10$	
1	1-2 (8)		3-4 5-6 (6)				7-8 9-10 (2)				16
2	1-2 (8)		3-4 5-6 7-8 9-10 (9)								17
3	1-2 3-4 5-6 7-8 9-10 (18)										18

Novamente, se as partições $PF=1$ e $PF=3$ (tabela 6.42) forem analisadas do ponto de vista do número médio de itens ($E(L)$), do *leadtime* médio ($E(T)$) e da variância do *leadtime* da rede ($V(T)$), observa-se que, devido a diferença de 2 máquinas, ainda é mais vantajoso particionar a instalação do que não particioná-la. Na configuração [1,2],[3-6],[7-10], $E(L)=11,541$ itens, $E(T)=6,908$ horas e $V(T)=0,561$ horas², enquanto na configuração [1-10], $E(L)=11,234$ itens, $E(T)=6,724$ e $V(T)=0,006$ horas². Ou seja, o aumento de 0,307 itens, 0,184 horas e 0,555 na variância da rede ao particionar a instalação são pouco representativos se comparados à necessidade de mais 2 máquinas ao optar por não particioná-la. Aqui, conforme observado anteriormente com as partições $PF=1$ e $PF=3$ na tabela 6.38, ao limitar $E(Wq_{jp}|Wq_{jp}>0)$, a variância do *leadtime* da rede obtida para ambas as configurações é indiretamente limitada. Conforme esperado, as variâncias obtidas para as configurações da tabela 6.38 (0,039 e $2,11 \cdot 10^{-6}$, para $PF=1$ e $PF=3$, respectivamente) são menores que as obtidas para a configuração da tabela 6.42. Ora, na tabela 6.42 G_{jp} é igual a 0,6 e, portanto, $E(Wq_{jp}|Wq_{jp}>0) \leq 0,833$ (ao invés de 0,5 horas quando $G_{jp}=1,0$ como é o caso da tabela 6.38), ocasionando maiores variâncias no *leadtime* uma vez que o limitante superior para o tempo médio de espera de um item, dado que ele espera, é maior.

Pela tabela 6.43 pode-se perceber que ao simular a rede sob a configuração [1,2],[3,4,5,6],[7,8,9,10] e $G_{jp}=0,6$, a medida $E(Wq_{jp}|Wq_{jp}>0)$ foi inferior ao limitante superior 0,83 horas em todas as estações da rede, o que valida a hipótese do modelo PRC-2.

TABELA 6.43 – Tempo médio de espera em fila, dado que espera, $E(Wq_{jp}|Wq_{jp}>0)$, e número médio de itens no sistema para as linhas ($E(L_p)$) da partição [1,2],[3,4,5,6],[7,8,9,10] obtidos por simulação para a Rede 3 com $G_{jp}=0,6$.

Medidas de desempenho	Medidas para as linhas		
	[1,2]	[3,4,5,6]	[7,8,9,10]
$E(Wq_{jp} Wq_{jp}>0)$ – hora	0,0000	0,6924	0,3495
$E(L_p)$ – item	5,07	5,27	0,99

Pode-se notar pela tabela 6.44 que as aproximações utilizadas para esta rede são efetivas em descrever o desempenho do sistema, uma vez que tanto o *leadtime* médio, quanto o número médio de itens na rede, obtidos pelas aproximações são bastante próximos dos obtidos por simulação (desvios relativos médios de 1,80% e 1,77%, respectivamente).

TABELA 6.44 – *Leadtime* médio, $E(T_k)$, e número médio de itens, $E(L_k)$, para as 10 classes de produtos e para a rede ($E(L)$ e $E(T)$) obtidos por simulação e pelas aproximações para a Rede 3 com $G_{jp}=0,6$.

Classe	$E(T_k)$ – hora		$E(L_k)$ – item	
	Simulação	Aproximações	Simulação	Aproximações
1	30,487	30,619	2,286	2,296
2	27,823	27,954	2,784	2,795
3	10,445	10,786	1,306	1,348
4	7,930	8,236	1,190	1,235
5	7,847	8,196	1,373	1,434
6	7,797	8,162	1,398	1,463
7	1,185	1,139	0,227	0,218
8	1,127	1,118	0,225	0,224
9	1,144	1,099	0,257	0,247
10	1,139	1,084	0,285	0,271
Rede	6,779 horas (0%)	6,901 horas (1,80%)	11,331 itens (0%)	11,531 itens (1,77%)

Outros experimentos foram realizados com a Rede 3. Com o objetivo de avaliar a sensibilidade da partição [1,2],[3,4,5,6],[7,8,9,10] com 18 máquinas e $G_{jp}=1,0$,

variou-se o tempo médio de algumas classes de produtos, conforme mostra a tabela 6.45.

TABELA 6.45 – Número de máquinas (MQ) ao mudar alguns parâmetros do caso base (Rede 3) para a partição obtida pelo algoritmo 6 com $G_{jp}=1,0$ e para a partição com uma linha única.

Mudanças nos Parâmetros de Entrada da Tabela 6.3	Partição com menor número de máquinas		Partição [1-10]
	MQ	Partição	MQ
Caso base	18	[1,2],[3,4,5,6],[7,8,9,10]	23
$E(s_{11})/3$ e $E(s_{21})/3$ - caso 1	11	[1,2,3,4,5,6],[7,8,9,10]	12
$E(s_{31})/7$; $E(s_{41})/7$; $E(s_{51})/7$; $E(s_{61})/7$ – caso 2	11	[1,2],[3-10]	26
$E(s_{11})/3$; $E(s_{21})/3$ $10 * E(s_{71})$; $10 * E(s_{81})$; $10 * E(s_{91})$; $10 * E(s_{10,1})$ – caso 3	18	[1-10]	18

Os tempos médios de processamento dos casos base, 1, 2 e 3 da tabela 6.45 constam na tabela 6.46. Note que, comparado ao caso base, ao dividir o tempo médio de processamento das classes 1 e 2 por 3 (caso 1), os tempos médios de processamento das classes 1, 2, 3, 4, 5 e 6 tornam-se mais semelhantes (vide tabela 6.46). Note também que ao dividir $E(s_{31})$, $E(s_{41})$, $E(s_{51})$ e $E(s_{61})$ por 7 (caso 2), os tempos médios de processamento das classes 3 a 10 tornam-se mais semelhantes (tabela 6.46). O mesmo ocorre para todas as classes ao manipular os tempos de processamento no caso 3. Diante dessa observação e dos resultados mostrados na tabela 6.45, pode-se concluir que a configuração com número mínimo de máquinas, que satisfaz todas restrições do PRC-2, tende a agregar classes de produtos com tempos de processamento semelhantes nas mesmas linhas. Resultado semelhante foi notado para o PRC-1, ou seja, é vantagem do ponto de vista do número de máquinas agregar produtos com tempos de processamento semelhantes.

TABELA 6.46 – Tempos médios de processamento para cada classe k para os casos base, 1, 2 e 3 (Rede 3).

Caso	Tempos médios de processamento para as classes de produtos para cada caso									
	1	2	3	4	5	6	7	8	9	10
Base	30,50	27,84	10,27	7,72	7,68	7,65	1,12	1,10	1,08	1,07
1	10,17	9,28	10,27	7,72	7,68	7,65	1,12	1,10	1,08	1,07
2	30,50	27,84	1,47	1,10	1,10	1,09	1,12	1,10	1,08	1,07
3	10,17	9,28	10,27	7,72	7,68	7,65	11,23	11,01	10,83	10,68

A tabela 6.47 mostra a medida $E(Wq_{jp}|Wq_{jp}>0)$ e $E(L_p)$ para as estações das linhas referentes aos casos base, 1, 2 e 3 da tabela 6.45.

TABELA 6.47 – Tempo médio de espera em fila, dado que espera, $E(Wq_{jp}|Wq_{jp}>0)$, e número médio de itens nas linhas ($E(L_p)$), obtidos por simulação, para os casos base, 1, 2 e 3 quando $G_{jp}=1,0$ na Rede 3.

Caso	Medidas de desempenho	Medidas de desempenho para as linhas dos respectivos casos					
		[1,2]	[3-6]	[7-10]	[1-6]	[3-10]	[1-10]
Base	$E(Wq_{jp} Wq_{jp}>0)$	0,0000	0,3997	0,3495	-----	-----	-----
	$E(L_p)$	5,07	5,16	0,99	-----	-----	-----
1	$E(Wq_{jp} Wq_{jp}>0)$	-----	-----	0,3495	0,3935	-----	-----
	$E(L_p)$	-----	-----	0,99	6,85	-----	-----
2	$E(Wq_{jp} Wq_{jp}>0)$	0,000	-----	-----	-----	0,6496	-----
	$E(L_p)$	5,07	-----	-----	-----	2,27	-----
3	$E(Wq_{jp} Wq_{jp}>0)$	-----	-----	-----	-----	-----	0,4137
	$E(L_p)$	-----	-----	-----	-----	-----	16,41

Note que, com exceção da linha [3-10] do caso 2, todas as outras linhas apresentaram $E(Wq_{jp}|Wq_{jp}>0) \leq 0,5$ horas, como resultado do modelo de simulação desenvolvido no GPSS/H[®]. É bom lembrar que o limitante superior para o tempo médio de espera de um item, dado que ele espera, foi derivado de uma aproximação (vide as expressões 83 e 92). Por isso, era de se esperar que $E(Wq_{jp}|Wq_{jp}>0)$, obtido de um modelo de simulação, pudesse eventualmente superar um pouco o limitante superior estabelecido. Como pôde-se perceber, houveram poucas ocorrências deste tipo.

6.2.2 Rede formada por vários estágios GI/G/m (Rede 4)

Aplicou-se o algoritmo 6 à Rede 4 com $G_{jp}=0,9$, o que implica em impor

um limitante superior para $E(W_{q_{jp}}|W_{q_{jp}}>0)$ igual a $1/(2*0,9)= 0,56$ horas. As partições com menor número de máquinas em cada iteração do algoritmo 6 são mostradas na tabela 6.48.

TABELA 6.48 – Número total de máquinas (MQ) obtido para as diferentes partições em cada iteração ($iter$) do algoritmo 6 para a Rede 4 com $G_{jp}=0,9$.

<i>Iter</i>	Classes de produtos alocados às linhas de produção ($L1, \dots, L10$) em cada iteração e número de máquinas em cada linha – entre parênteses										MQ
	$L1$	$L2$	$L3$	$L4$	$L5$	$L6$	$L7$	$L8$	$L9$	$L10$	
0	1 (8)	2 (9)	3 (8)	4 (7)	5 (6)	6 (9)	7 (6)	8 (10)	9 (17)	10 (31)	111
1	1-2 (11)		3 (8)	4 (7)	5 (6)	6 (9)	7 (6)	8 (10)	9 (17)	10 (31)	105
2	1-2 (11)		3 (8)	4 (7)	5-8 (10)	6 (9)	7 (6)		9 (17)	10 (31)	99
3	1-2 (11)		3-5-8 (12)	4 (7)		6 (9)	7 (6)		9 (17)	10 (31)	93
4	1-2 (11)		3-4 5-8 (14)			6 (9)	7 (6)		9 (17)	10 (31)	88
5	1-2 3-4 5-8 (20)					6 (9)	7 (6)		9 (17)	10 (31)	83
6	1-2 3-4 5-6 8 (24)						7 (6)		9 (17)	10 (31)	78
7	1-2 3-4 5-6 7-8 (26)								9 (17)	10 (31)	74
8	1-2 3-4 5-6 7-8 (26)								9-10 (44)		70

Note que a partição [1-8],[9,10] obtida pelo algoritmo 6 requer 70 máquinas. Bitran e Sarkar (1993c), aplicando o algoritmo 6, encontraram a mesma partição como solução para a Rede 4, mas o número de máquinas requerido foi 73

máquinas. Provavelmente, esta diferença é devida à aproximação descrevendo o processo de separação das partidas que é diferente da aproximação (113), conforme já discutido anteriormente. Esta observação sugere que o modelo é sensível ao sistema de aproximações utilizado.

A tabela 6.49 disponibiliza $E(T_k)$ e $E(L_k)$ obtidos por simulação e pelas aproximações. Novamente, as aproximações utilizadas para este caso (sugeridas por Bitran e Sarkar, 1993c) foram efetivas em descrever medidas de desempenho funções do primeiro momento (média), visto que os valores são bastante próximos (desvios relativos médios de $-0,56\%$ para o *leadtime* e de $-0,62\%$ para o número de itens da rede).

TABELA 6.49 – *Leadtime* médio, $E(T_k)$, e número médio de itens, $E(L_k)$, obtidos por simulação e pelas aproximações para as 10 classes de produtos e para a rede ($E(T)$ e $E(L)$) na partição [1-8],[9,10] da Rede 4 com $G_{jp}=0,90$.

Classe	$E(T_k)$ – hora		$E(L_k)$ – item	
	Simulação	Aproximações	Simulação	Aproximações
1	9,095	9,066	0,909	0,907
2	10,606	10,599	1,059	1,060
3	9,494	9,386	0,950	0,939
4	6,350	6,283	0,635	0,628
5	5,397	5,318	0,540	0,532
6	9,830	9,828	0,981	0,983
7	6,424	6,421	0,642	0,642
8	10,220	10,181	1,022	1,018
9	65,737	65,254	6,577	6,525
10	98,346	97,864	9,848	9,786
Rede	23,150 horas (0%)	23,020 horas (-0,56%)	23,163 itens (0%)	23,020 itens (-0,62%)

A tabela 6.50 apresenta os resultados obtidos sem particionar a Rede 4. Observe que ao agregar a linha [1,2,3,4,5,6,7,8] à linha [9,10], ocorreu um aumento de 18 máquinas (compare as tabelas 6.48 e 6.50), ou seja, total de 88 máquinas. Convém mencionar que o resultado encontrado por Bitran e Sarkar (1993c), com a opção de não particionar a Rede 4, foi de 90 máquinas.

TABELA 6.50 – Número de máquinas na rede (MQ) obtido para partições fixas (PF) com $G_{jp}=0,9$ (Rede 4).

PF	Classes de produtos alocados às linhas de produção ($L1, \dots, L10$) em cada partição e número de máquinas em cada linha – entre parênteses										MQ
	$L1$	$L2$	$L3$	$L4$	$L5$	$L6$	$L7$	$L8$	$L9$	$L10$	
1	1-2 3-4 5-6 7-8 9-10 (88)										88

A tabela 6.51 disponibiliza $E(T_k)$, $E(T)$, $E(L_k)$ e $E(L)$ obtidos por simulação e pelas aproximações para a opção de não particionar a Rede 4. Novamente, as aproximações utilizadas para este caso (sugeridas por Bitran e Sarkar, 1993c) foram efetivas em descrever medidas de desempenho, visto que os valores são bastante próximos (desvios relativos menores que 1,27% para o *leadtime* e 0,02% para o número de itens da rede).

TABELA 6.51 – *Leadtime* médio, $E(T_k)$, e número médio de itens, $E(L_k)$, obtidos por simulação e pelas aproximações para as 10 classes de produtos e para a rede ($E(T)$ e $E(L)$) na partição [1-10] da Rede 4 com $G_{jp}=0,90$.

Classe	$E(T_k)$ – hora		$E(L_k)$ – item	
	Simulação	Aproximações	Simulação	Aproximações
1	9,336	9,277	0,936	0,928
2	10,848	10,778	1,086	1,078
3	9,656	9,614	0,967	0,961
4	9,505	6,477	0,651	0,648
5	5,743	5,680	0,574	0,568
6	10,012	9,975	1,001	0,998
7	6,964	6,905	0,694	0,691
8	10,499	10,505	1,050	1,051
9	64,965	65,142	6,496	6,514
10	97,637	97,839	9,770	9,784
Rede	23,517 horas (0%)	23,219 horas (1,27%)	23,225 itens (0%)	23,221 itens (0,02%)

Comparando o número médio de itens ($E(L)$) e o *leadtime* médio da rede ($E(T)$) obtidos para as configurações [1-8],[9,10], vide tabela 6.49, e [1-10], vide tabela 6.51, observa-se que o *leadtime* médio e o número médio de itens na rede diminuem

quando a instalação é particionada. O mesmo ocorre com a variância do *leadtime* da rede: ao particionar a instalação a variância diminui de 0,433 (configuração [1-10]) para 0,110 (configuração [1-8],[9,10]). Ainda, há redução de 18 máquinas, quando a instalação é particionada, pois a configuração [1-10] necessita de 88 máquinas, ao invés de 70 máquinas (configuração [1-8],[9,10]). Em outras palavras, a partição da instalação é vantajosa comparada a opção de não particionar, pois requer menos máquinas e reduz o *leadtime* médio, o número médio de itens e a variância do *leadtime* da rede. Além disso, conforme foi observado na rede 3, a variância ponderada do *leadtime* da rede foi indiretamente limitada, ao limitar superiormente $E(Wq_{jp}|Wq_{jp}>0)$, uma vez que as variâncias de *leadtime* obtidas para ambas as configurações são modestas (pequenas).

A tabela 6.52 mostra as medidas $E(Wq_{jp}|Wq_{jp}>0)$ e $E(L_p)$ para a partição [1,2,3,4,5,6,7,8],[9,10].

TABELA 6.52 – Tempo médio de espera em fila, dado que espera, $E(Wq_{jp}|Wq_{jp}>0)$, e número médio de itens no sistema para as linhas, $E(L_p)$, obtidos por simulação para a partição [1,2,3,4,5,6,7,8],[9,10] da Rede 4 com $G_{jp}=0,9$.

Medidas de desempenho das estações e das linhas	Medidas de Desempenho para as linhas	
	[1,2,3,4,5,6,7,8]	[9,10]
$E(Wq_{1p} Wq_{1p}>0)$ – hora	0,2754	0,5860
$E(Wq_{2p} Wq_{2p}>0)$ – hora	0,1824	0,8381
$E(Wq_{3p} Wq_{3p}>0)$ – hora	0,0000	0,0000
$E(Wq_{4p} Wq_{4p}>0)$ – hora	0,4133	0,0000
$E(Wq_{5p} Wq_{5p}>0)$ – hora	0,5415	0,2614
$E(Wq_{6p} Wq_{6p}>0)$ – hora	0,5717	0,3129
$E(Wq_{7p} Wq_{7p}>0)$ – hora	0,4358	0,0000
$E(Wq_{8p} Wq_{8p}>0)$ – hora	0,4307	0,0000
$E(Wq_{9p} Wq_{9p}>0)$ – hora	0,5362	0,8519
$E(Wq_{10p} Wq_{10p}>0)$ – hora	0,4794	0,6995
$E(Wq_{11p} Wq_{11p}>0)$ – hora	0,4589	0,4145
$E(Wq_{12p} Wq_{12p}>0)$ – hora	0,4453	0,6018
$E(Wq_{13p} Wq_{13p}>0)$ – hora	0,4982	0,2714
$E(L_p)$ – item	6,74	16,42

Note que das 13 estações da linha [1,2,3,4,5,6,7,8], nenhuma delas ultrapassa o limitante superior de 0,56 horas de tempo de espera, dado que espera. Entretanto, 5 estações da linha [9,10] apresentaram $E(Wq_{jp}|Wq_{jp}>0)>0,56$ horas Os piores casos foram das estações 2 e 9, com 0,84 horas e 0,85 horas, respectivamente). Convém lembrar que a medida $E(Wq_{jp}|Wq_{jp}>0)$ foi derivada de uma aproximação (vide

expressões 83 e 92) e que, portanto, era esperado que alguns resultados obtidos por um modelo de simulação pudessem ultrapassar 0,56 horas.

Com vistas a analisar a sensibilidade do modelo ao parâmetro G_{jp} , a tabela 6.53 reporta o número de máquinas nas diferentes partições obtidas em cada iteração do algoritmo 6 ao fixar o limitante superior para $E(W_{q_{jp}}|W_{q_{jp}}>0)$ em 0,125 horas ($G_{jp}=4,00$).

TABELA 6.53 – Número total de máquinas (MQ) obtido para as diferentes partições em cada iteração ($iter$) do algoritmo 6 para a Rede 4 com $G_{jp}=4,00$.

<i>Iter</i>	Classes de produtos alocados às linhas de produção ($L1, \dots, L10$) em cada iteração e número de máquinas em cada linha – entre parênteses										<i>MQ</i>
	<i>L1</i>	<i>L2</i>	<i>L3</i>	<i>L4</i>	<i>L5</i>	<i>L6</i>	<i>L7</i>	<i>L8</i>	<i>L9</i>	<i>L10</i>	
0	1 (16)	2 (25)	3 (15)	4 (12)	5 (7)	6 (22)	7 (9)	8 (17)	9 (31)	10 (68)	222
1	1-2 (32)		3 (15)	4 (12)	5 (7)	6 (22)	7 (9)	8 (17)	9 (31)	10 (68)	213
2	1-2 4 (35)		3 (15)		5 (7)	6 (22)	7 (9)	8 (17)	9 (31)	10 (68)	204
3	1-2 4-6 (50)		3 (15)		5 (7)		7 (9)	8 (17)	9 (31)	10 (68)	197
4	1-2 4-6 8 (57)		3 (15)		5 (7)		7 (9)		9 (31)	10 (68)	187
5	1-2 4-6 7-8 (54)		3 (15)		5 (7)				9 (31)	10 (68)	175
6	1-2 3-4 6-7 8 (62)				5 (7)				9 (31)	10 (68)	168
7	1-2 3-4 5-6 7-8 (61)								9 (31)	10 (68)	160

Comparando os resultados da tabela 6.53 com a 6.48, constata-se que foram necessárias muito mais máquinas (aumento de 70 para 160) ao reduzir o limitante

superior para tempo médio de espera de um item, dado que ele espera, de 0,56 horas para 0,125 horas.

A tabela 6.54 mostra que a partição [1,2,3,4,5,6,7,8],[9],[10] obtida pelo algoritmo 6 (tabela 6.53) necessita de um número menor de máquinas e que, portanto, para a Rede 4 com $G_{jp}=4,00$, é mais vantajoso particioná-la do que agregar todos os produtos em uma linha única.

TABELA 6.54 – Número de máquinas na rede (MQ) obtido para partições fixas (PF) com $G_{jp}=4,00$ (Rede 4).

PF	Classes de produtos alocados às linhas de produção ($L1, \dots, L10$) em cada partição e número de máquinas em cada linha – entre parênteses										MQ
	$L1$	$L2$	$L3$	$L4$	$L5$	$L6$	$L7$	$L8$	$L9$	$L10$	
0	1-2 3-4 5-6 7-8 (61)								9-10 (115)		176
1	1-2 3-4 5-6 7-8 9-10 (289)										289

A tabela 6.55 mostra que tanto o *leadtime* como o número médio de itens de cada classe na rede, estimados pelas aproximações empregadas, estão bem próximos dos resultados obtidos por simulação (desvios relativos de $-0,03\%$ para o *leadtime* e $-0,06\%$ para o número médio de itens na rede).

TABELA 6.55 – *Leadtime* médio, $E(T_k)$, e número médio de itens, $E(L_k)$, obtidos por simulação e pelas aproximações para as 10 classes de produtos e para a rede ($E(T)$ e $E(L)$) na partição [1-8],[9],[10] da Rede 4 com $G_{jp}=4,00$.

Classe	$E(T_k)$ – hora		$E(L_k)$ – item	
	Simulação	Aproximações	Simulação	Aproximações
1	8,877	8,855	0,884	0,885
2	10,378	10,355	1,033	1,035
3	9,179	9,155	0,917	0,915
4	6,080	6,055	0,609	0,605
5	5,102	5,077	0,509	0,508
6	9,565	9,555	0,955	0,955
7	6,295	6,268	0,630	0,627

8	9,879	9,884	0,988	0,988
9	64,377	64,468	6,449	6,447
10	97,241	97,229	9,728	9,723
Rede	22,697 horas (0%)	22,690 horas (-0,03%)	22,702 itens (0%)	22,688 itens (-0,06%)

Note, na tabela 6.56, que somente 4 estações da linha [1,2,3,4,5,6,7,8], 1 estação da linha [9] e 1 estação da linha [10] apresentam $E(Wq_{jp}|Wq_{jp}>0)$ ligeiramente maiores que 0,125 horas, o limitante superior imposto ao fixar $G_{jp}=4,00$ (os piores casos foram as estações 6 e 9 com 0,264 e 0,235 horas, respectivamente). Conforme mencionado anteriormente, porque $E(Wq_{jp}|Wq_{jp}>0)$ foi derivado de uma aproximação (vide expressões 83 e 92), era esperado que alguns resultados obtidos por um modelo de simulação pudessem ultrapassar o limitante superior.

TABELA 6.56 – Tempo médio de espera em fila, dado que espera, $E(Wq_{jp}|Wq_{jp}>0)$, e número médio de itens no sistema para as linhas ($E(L_p)$) da partição [1,2,3,4,5,6,7,8],[9],[10] da Rede 4 obtidos por simulação com $G_{jp}=4,00$.

Medidas de desempenho das estações e das linhas	Medidas de Desempenho para as linhas		
	[1,2,3,4,5,6,7,8]	[9]	[10]
$E(Wq_{1p} Wq_{1p}>0)$ – hora	0,166	0,000	0,000
$E(Wq_{2p} Wq_{2p}>0)$ – hora	0,188	0,374	0,464
$E(Wq_{3p} Wq_{3p}>0)$ – hora	0,000	0,000	0,000
$E(Wq_{4p} Wq_{4p}>0)$ – hora	0,000	0,000	0,000
$E(Wq_{5p} Wq_{5p}>0)$ – hora	0,000	0,000	0,000
$E(Wq_{6p} Wq_{6p}>0)$ – hora	0,264	0,000	0,000
$E(Wq_{7p} Wq_{7p}>0)$ – hora	0,000	-----	0,000
$E(Wq_{8p} Wq_{8p}>0)$ – hora	0,021	0,000	-----
$E(Wq_{9p} Wq_{9p}>0)$ – hora	0,235	-----	0,000
$E(Wq_{10p} Wq_{10p}>0)$ – hora	0,000	0,000	-----
$E(Wq_{11p} Wq_{11p}>0)$ – hora	0,000	-----	0,000
$E(Wq_{12p} Wq_{12p}>0)$ – hora	0,080	0,000	0,000
$E(Wq_{13p} Wq_{13p}>0)$ – hora	0,000	0,000	0,000
$E(L_p)$ – item	6,53	6,45	9,73

Outros experimentos foram realizados com a Rede 4 e $G_{jp}=0,9$ para avaliar a sensibilidade do modelo quando ocorrem mudanças nos tempos de processamento e nos coeficientes de variabilidade. A tabela 6.57 mostra a partição com o menor número de máquinas resultante do algoritmo 6 e o número de máquinas para a partição de uma única linha.

TABELA 6.57 – Número de máquinas (MQ) ao mudar alguns parâmetros do caso base (Rede 4) para a partição com menor número de máquinas resultante do algoritmo 6 com $G_{jp}=0,9$ e para a partição com uma linha única.

Mudanças nos Parâmetros de Entrada da Tabela 6.4 e 6.5	Partição com menor número de máquinas		Partição [1-10]
	MQ	Partição	MQ
Caso base	70	[1-8],[9,10]	88
$E(s_{9l})/3$ e $E(s_{10l})/3$ - caso 1	36	[1-10]	36
$ca'_9 = ca'_{10} = 0,5$ - caso 2	91	[1-10]	91
$ca'_9 = ca'_{10} = 0$ - caso 3	66	[1-8],[9,10]	88

Observe nas tabelas 6.4 e 6.5 que as classes de produtos 9 e 10 se destacam das demais por apresentarem menor coeficiente de variação do intervalo de tempo entre chegadas e tempos médios de processamento maiores. Por isso, no caso 1, ao se dividir por 3 os tempos médios de processamento, os tempos de serviço das classes 9 e 10 tornam-se similares aos das demais classes (vide tabela 6.58). Já os casos 2 e 3 foram criados para avaliar o efeito do coeficiente de variação em particionar ou não a instalação.

TABELA 6.58 – Tempos médios de serviço para as operações l das 10 classes de produtos (k) do caso 1 com $G_{jp}=0,9$ (Rede 4).

Classe	Operação (hora/item)												
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0,2	0,1	0,2	0,2	0,9	3,4	3,8						
2	0,6	0,2	3,0	0,6	1,8	0,9	0,8	2,4					
3	0,8	0,4	1,8	3,5	0,6	0,4	1,0	0,6					
4	1,8	0,4	1,8	0,2	0,2	0,4	0,8	0,4					
5	0,5	0,3	1,5	0,6	0,2	0,8	0,5	0,6					
6	0,6	0,7	0,7	2,6	0,7	0,8	0,4	3,0					
7	0,7	0,2	1,0	1,2	0,3	1,8	0,3	0,7					
8	0,3	0,2	0,6	1,8	0,2	0,7	0,7	0,6	2,2	0,4	0,8	1,3	
9	0,33	0,27	1,60	5,00	0,40	0,93	0,93	0,40	4,67	0,53	3,20	0,93	1,20
10	0,43	0,27	7,20	2,93	0,40	4,00	1,33	4,67	0,27	3,07	0,53	5,00	2,27

Analisando os resultados da tabela 6.57, pode-se concluir que, quanto mais semelhantes as classes de produtos em termos de tempo médio de processamento e de variabilidade do processo de chegadas, é mais vantajoso agregá-las em uma mesma

linha (vide resultados dos casos 1 e 2 da tabela 6.57). Os casos base e 3 evidenciam que se algumas classes de produtos apresentarem tempos médios de processamento e variabilidade razoavelmente distintas, convém particioná-las em linhas diferentes. Ora, as classes 9 e 10 apresentam (nos casos base e 3) menor variabilidade no processo de chegada e tempos médios de processamento maiores que as demais classes.

Para verificar se o modelo PRC-2 é robusto ao sistema de aproximações utilizado para estimar ca_{jp} nas estações j de cada linha p , empregou-se a aproximação proposta por Bitran e Tirupati (1988), em substituição à de Segal e Whitt (1989), para descrever o processo de desagregação de partidas no método aproximado de decomposição. A tabela 6.59 mostra o número mínimo de máquinas para as configurações obtidas em cada iteração do algoritmo 6 para a Rede 4 com $G_{jp}=0,90$.

TABELA 6.59 – Número total de máquinas (MQ) obtido para as diferentes partições em cada iteração ($iter$) do algoritmo 6 para a Rede 4 com $G_{jp}=0,90$, empregando as aproximações de Bitran e Tirupati (1988).

<i>Iter</i>	Classes de produtos alocados às linhas de produção ($L1, \dots, L10$) em cada iteração e número de máquinas em cada linha – entre parênteses										MQ
	$L1$	$L2$	$L3$	$L4$	$L5$	$L6$	$L7$	$L8$	$L9$	$L10$	
0	1 (10)	2 (11)	3 (10)	4 (8)	5 (6)	6 (10)	7 (7)	8 (12)	9 (37)	10 (54)	165
1	1 (10)	2 (11)	3-8 (16)	4 (8)	5 (6)	6 (10)	7 (7)		9 (37)	10 (54)	159
2	1 (10)	2 (11)	3-4-8 (17)		5 (6)	6 (10)	7 (7)		9 (37)	10 (54)	152
3	1 (10)	2-3 4-8 (21)			5 (6)	6 (10)	7 (7)		9 (37)	10 (54)	145
4	1 (10)	2-3 4-6 8 (23)			5 (6)		7 (7)		9 (37)	10 (54)	137
5	1-2 3-4 6-8 (27)				5 (6)		7 (7)		9 (37)	10 (54)	131
6	1-2 3-4 6-7 8 (27)				5 (6)				9 (37)	10 (54)	124
7	1-2								9	10	119

	3-4 5-6 7-8 (28)								(37)	(54)	
8	1-2 3-4 5-6 7-8 9 (58)									10 (54)	112
9	1-2 3-4 5-6 7-8 9-10 (96)										96

Comparando os resultados da tabela 6.59 com os das tabelas 6.48 e 6.50, percebe-se que não somente o número de máquinas na partição com menor número de máquinas mudou substancialmente (de 70 para 96), como também o fato de particionar a instalação já não ser mais vantajoso.

Note na tabela 6.60 que, apesar da aproximação proposta por Bitran e Tirupati (1988) apresentar partição e número de máquinas diferentes dos apresentados pela aproximação de Segal e Whitt (1989) (tabelas 6.48 e 6.50), ela é bastante eficiente em descrever o desempenho da rede na melhor configuração obtida pelo algoritmo 6. Os resultados obtidos pelas aproximações são bastante próximos dos obtidos por simulação (desvios relativos de 0,56% para o *leadtime* e 0,52% para o número de itens na rede, conforme tabela 6.60). Convém frisar que as aproximações de Segal e Whitt (1989) também foram bastante eficientes em descrever o desempenho da Rede 4 com $G_{jp}=0,90$ (vide tabela 6.51). Observe também que nenhuma estação da partição [1-10] com 96 máquinas ultrapassou o limitante superior para o tempo médio de espera de um item, dado que ele espera, de 0,56 horas, ao impor $G_{jp}=0,9$ (vide tabela 6.60).

TABELA 6.60 – *Leadtime* médio, $E(T_k)$, e número médio de itens, $E(L_k)$, obtidos por simulação e pelas aproximações de Bitran e Tirupati (1988) para as 10 classes de produtos e para a rede ($E(T)$ e $E(L)$) na partição [1-10] da Rede 4 com $G_{jp}=0,90$.

Classe	$E(T_k)$ – hora		$E(L_k)$ – item	
	Simulação	Aproximações	Simulação	Aproximações
1	9,331	9,373	0,935	0,937
2	10,840	10,873	1,085	1,087
3	9,639	9,697	0,966	0,970
4	6,489	6,573	0,649	0,657
5	5,742	5,849	0,574	0,585
6	10,012	10,088	1,001	1,009
7	6,960	7,056	0,694	0,706
8	10,497	10,657	1,050	1,066
9	64,962	65,274	6,496	6,527
10	97,629	97,971	9,769	9,797
Rede	23,210 horas (0%)	23,341 horas (0,56%)	23,219 itens (0%)	23,341 itens (0,52%)

A tabela 6.61 mostra o tempo médio de espera em fila, dado que espera, e o número médio de itens no sistema para as linhas obtidos por simulação para a Rede 4.

TABELA 6.61 – Tempo médio de espera em fila, dado que espera, $E(Wq_{jp}|Wq_{jp}>0)$, e número médio de itens no sistema para as linhas ($E(L_p)$) da partição [1-10] da Rede 4 obtidos por simulação com $G_{jp}=0,9$, utilizando as aproximações de Bitran e Tirupati (1988).

Medidas de desempenho das estações e das linhas	Linha [1-10]
$E(Wq_{1p} Wq_{1p}>0)$ – hora	0,362
$E(Wq_{2p} Wq_{2p}>0)$ – hora	0,430
$E(Wq_{3p} Wq_{3p}>0)$ – hora	0,000
$E(Wq_{4p} Wq_{4p}>0)$ – hora	0,000
$E(Wq_{5p} Wq_{5p}>0)$ – hora	0,000
$E(Wq_{6p} Wq_{6p}>0)$ – hora	0,341
$E(Wq_{7p} Wq_{7p}>0)$ – hora	0,000
$E(Wq_{8p} Wq_{8p}>0)$ – hora	0,000
$E(Wq_{9p} Wq_{9p}>0)$ – hora	0,000
$E(Wq_{10p} Wq_{10p}>0)$ – hora	0,000
$E(Wq_{11p} Wq_{11p}>0)$ – hora	0,302
$E(Wq_{12p} Wq_{12p}>0)$ – hora	0,322
$E(Wq_{13p} Wq_{13p}>0)$ – hora	0,204
$E(L_p)$ – item	23,22

O fato de substituir o sistema de aproximações (de Segal e Whitt (1989))

para Bitran e Tirupati (1988)) mudar sensivelmente o número de máquinas (de 70 para 96 máquinas) e a partição (de [1-8],[9,10] para [1-10]), e ainda assim ambas as aproximações gerarem resultados próximos aos obtidos por simulação (conforme mostram as tabelas 6.49, 6.51 e 6.52 ao empregar a aproximação de Segal e Whitt (1989) e as tabelas 6.60 e 6.61, empregando a aproximação de Bitran e Tirupati (1988)), indicam que elas são eficientes em descrever o desempenho da rede, mas que devem ser utilizadas com cautela em um procedimento de otimização. Muito provavelmente a restrição de complexidade (expressão (106)) é razoavelmente sensível ao coeficiente de variabilidade do intervalo de tempo entre chegadas nas estações (ca_{jp}) que são estimados pelas aproximações no procedimento iterativo do algoritmo 6. Este resultado evidencia a sensibilidade do modelo PRC-2 com respeito à escolha das aproximações, o que também foi observado no modelo do PRC-1.

6.3 Problema de Redução de Complexidade 3 (PRC-3)

Nesta seção são apresentados alguns resultados referentes à aplicação do modelo PRC-3 às Redes 3 e 4. Aqui, o parâmetro de entrada da restrição de complexidade (h_{jp}) estabelece um limitante superior para o tempo de espera na estação j da linha p ($E(Wq_{jp})$). Infelizmente, a hipótese principal estabelecida no desenvolvimento da restrição de complexidade que caracteriza o PRC-3 não é totalmente satisfeita para estas redes, o que prejudicou a análise da aplicabilidade do modelo, conforme discutido adiante. Os experimentos com as Redes 3 e 4 foram aqui mantidos a título de comparação com os resultados obtidos em Bitran e Sarkar (1993c).

6.3.1 Rede formada por um único estágio $GI/G/m$ (Rede 3)

A tabela 6.62 mostra o número mínimo de máquinas para as diferentes configurações obtidas pela aplicação do algoritmo 6 para a Rede 3 com $h_{jp}=10$ no modelo PRC-3. A tabela 6.63 mostra o número mínimo de máquinas necessárias ao continuar o processo de agregação das linhas a partir da quinta iteração ($iter=5$) da tabela 6.62.

Ao estimar em cada iteração ca_{jp} e cs_{jp} por meio das aproximações (111)-(113), para as partições que resultaram em maior economia de número de máquinas, notou-se que os valores de ca_{jp} , em geral, não são desprezíveis quando comparados aos valores de cs_{jp} . Ao contrário, frequentemente cs_{jp} pode ser considerado desprezível com relação a ca_{jp} . Portanto, a principal hipótese estabelecida no desenvolvimento da expressão (109) não é satisfeita para a Rede 3. Para fins ilustrativos, a tabela 6.64 mostra ca_{jp} e cs_{jp} para a partição [1],[2],[3],[4,5,6],[7,8,9,10] da tabela 6.62.

TABELA 6.64 – Coeficientes quadráticos de variação do intervalo de tempo entre chegadas (ca_{jp}) e de serviço (cs_{jp}) nas estações da partição [1],[2],[3],[4,5,6],[7,8,9,10] (Rede 3).

Coeficiente	Coeficientes quadráticos de variação nas linhas				
	[1]	[2]	[3]	[4,5,6]	[7,8,9,10]
ca_{1p}	0,133	0,133	0,133	0,150	0,107
cs_{1p}	0,000	0,000	0,000	$1,502 \cdot 10^{-5}$	$3,530 \cdot 10^{-4}$

É importante observar que as ordens de grandeza de ca_{jp} e cs_{jp} para as demais partições da tabela 6.62 são bastante similares. Logo, a Rede 3 não é adequada para o estudo e análise do modelo PRC-3, e fica como sugestão para pesquisa futura a aplicação e análise do modelo PRC-3 em outras redes satisfazendo a hipótese acima.

6.3.2 Rede Formada por Vários Estágios $GI/G/m$ (Rede 4)

Ao aplicar o algoritmo 6 à Rede 4 com $h_{jp}=0,9$, o número mínimo de máquinas nas diferentes configurações obtidas em cada iteração é dado na tabela 6.65. Novamente, o algoritmo cessou a busca por uma partição com o menor número de máquinas (vide tabela 6.66). Bitran e Sarkar (1993c) obtiveram o mesmo resultado para a Rede 4 com $h_{jp}=0,9$: a partição [1-8],[9],[10] requerendo a menor capacidade (43 máquinas), reforçando que a implementação computacional neste trabalho está de acordo com a dos autores acima. Entretanto, conforme mostra a tabela 6.67, novamente a hipótese principal no desenvolvimento da restrição de complexidade (109) não é totalmente satisfeita, embora Bitran e Sarkar (1993c) não tenham discutido esta questão.

TABELA 6.65 – Número total de máquinas (MQ) obtido para as diferentes partições da Rede 4 em cada iteração ($iter$) do algoritmo 6 para $h_{jp}=0,9$.

$Iter$	Classes de produtos alocados às linhas de produção ($L1, \dots, L10$) em cada iteração e número de máquinas em cada linha – entre parênteses										MQ
	$L1$	$L2$	$L3$	$L4$	$L5$	$L6$	$L7$	$L8$	$L9$	$L10$	
0	1 (6)	2 (7)	3 (7)	4 (7)	5 (6)	6 (7)	7 (6)	8 (10)	9 (12)	10 (16)	84
1	1-4 (7)	2 (7)	3 (7)		5 (6)	6 (7)	7 (6)	8 (10)	9 (12)	10 (16)	78
2	1-4 (7)	2 (7)	3-8 (11)		5 (6)	6 (7)	7 (6)		9 (12)	10 (16)	72
3	1-4 (7)	2 (7)	3-6-8 (12)		5 (6)		7 (6)		9 (12)	10 (16)	66
4	1-4 (7)	2-3 6-8 (13)			5 (6)		7 (6)		9 (12)	10 (16)	60
5	1-4 (7)	2-3 6-7 8 (13)					7 (6)		9 (12)	10 (16)	54
6	1-2 3-4 6-7 8 (15)						7 (6)		9 (12)	10 (16)	49
7	1-2 3-4 5-6 7-8 (15)								9 (12)	10 (16)	43

Observe que a partição da tabela 6.67 é a que requer o menor número de máquinas na tabela 6.65 e foi escolhida em caráter ilustrativo. Os valores de ca_{jp} e cs_{jp} obtidos para as demais partições das tabelas 6.66 e 6.67 possuem ordens de grandeza similares. Na tabela 6.67, opostamente à hipótese estabelecida para o desenvolvimento da restrição de complexidade que caracteriza o modelo PRC-3, cs_{jp} são em geral desprezíveis com relação à ca_{jp} . Portanto, a Rede 4 (assim como a Rede 3) não é adequada para o modelo PRC-3, e o estudo da aplicabilidade do modelo fica como proposta de pesquisa futura.

TABELA 6.66 – Número total de máquinas (MQ) obtido em diferentes partições (PF) para $h_{jp}=0,9$ (Rede 4).

PF	Classes de produtos alocados às linhas de produção ($L1, \dots, L10$) em cada partição e número de máquinas em cada linha – entre parênteses										MQ
	$L1$	$L2$	$L3$	$L4$	$L5$	$L6$	$L7$	$L8$	$L9$	$L10$	
1	1-2 3-4 5-6 7-8 (15)								9-10 (39)		54
2	1-2 3-4 5-6 7-8 9-10 (116)										116

TABELA 6.67 – Coeficientes quadráticos de variação do intervalo de tempo entre chegadas (ca_{jp}) e de serviço (cs_{jp}) nas estações da partição [1-8],[9],[10] (Rede 4).

Estação j	Partição [1-8]		Partição [9]		Partição [10]	
	ca_{j1}	cs_{j1}	ca_{j2}	cs_{j2}	ca_{j3}	cs_{j3}
1	0,354	0,447	0,050	0,000	0,100	0,000
2	0,397	0,289	0,097	0,074	0,126	0,041
3	0,344	0,000	0,075	0,000	0,112	0,000
4	0,336	1,331	0,177	0,000	0,112	0,000
5	0,346	0,386	0,027	0,000	0,017	0,000
6	0,345	0,090	0,090	0,000	0,273	0,000
7	0,348	0,245	-----	-----	0,149	0,000
8	0,348	0,163	0,075	0,000	-----	-----
9	0,343	0,118	-----	-----	0,112	0,000
10	0,331	0,541	0,075	0,000	-----	-----
11	0,375	0,599	-----	-----	0,017	0,000
12	0,391	0,350	0,066	0,000	0,177	0,000
13	0,351	0,472	0,066	0,000	0,172	0,000

Diante dos resultados expostos nas seções deste capítulo, pode-se

concluir que, para certos casos, é possível melhorar o desempenho de uma rede de manufatura apenas particionando-a em subplantas, ou seja, sem necessidades de investimentos adicionais em capacidade. Ou alternativamente, em certos casos é possível manter (ou até melhorar) o desempenho da rede de manufatura, particionando-a em subplantas (sublinhas) que necessitam de menos capacidade do que a configuração com uma única linha (vide por exemplo, os casos da tabela 6.31; os casos base, 1 e 2 da tabela 6.45; e os casos base e 3 da tabela 6.57). Esta é uma observação importante desta tese.

7. CONCLUSÕES E PERSPECTIVAS FUTURAS

7.1 Conclusões

Esta tese estudou a aplicação de aproximações para avaliar e otimizar o desempenho de sistemas de manufatura discretos representados por redes de filas abertas. As aproximações foram utilizadas como alternativa à falta de abordagens exatas para tratar redes de filas genéricas (aquelas cujas distribuições de tempos de serviços e/ou de intervalo de tempo entre chegadas nas estações não são necessariamente exponenciais). Para alcançar este objetivo, inicialmente foi realizada uma pesquisa bibliográfica abrangente acerca da aplicação de aproximações de redes de filas abertas em modelos de otimização para projeto e planejamento de sistemas de manufatura discretos tipo *job-shop*.

Aqui foram estudadas três classes de problemas de otimização, envolvendo decisões de médio a longo prazo (decisões tático-estratégicas). Na classe *SP1* (problema estratégico 1, também chamado de problema de alocação de capacidade ou desempenho desejado), o objetivo é minimizar o investimento em capacidade sujeito a restrições de desempenho desejado para o sistema, tais como limitantes superiores para o *WIP* (inventário em processo) ou para o *leadtime* das classes de produtos que visitam a rede. Na classe *SP2* (problema estratégico 2, também chamado de realocação de capacidade), o objetivo é otimizar o desempenho do sistema (*WIP* ou *leadtime*, por exemplo), sujeito a restrições de limitação (orçamentária) de capacidade. As classes *SP1* e *SP2* envolvem o *trade-off* entre capital de investimento e capital de trabalho, se a medida de desempenho adotada for *WIP*. Já a classe *SP3* envolve a partição da instalação em subplantas (plantas dentro da planta), e alocação de capacidade nas subplantas para reduzir a complexidade do sistema (seja em termos da gestão dos produtos ou da gestão da estação) e melhorar o desempenho global da instalação.

Com vistas a verificar a aplicabilidade e adequação das aproximações para avaliar o desempenho do sistema, e estudar a classe de problemas *SP1* ou *SP2* em um caso real, foi escolhida uma empresa fabricante de peças metálicas, representada por uma rede de filas aberta, para servir de estudo de caso. Para avaliar o desempenho da rede, foram coletados dados para caracterizar os processos de chegada e de serviço nas

estações por um período de 14 meses. Estes dados foram utilizados como entrada para o método aproximado de decomposição paramétrico (método analítico) e para a simulação discreta (método experimental), utilizada para validar os resultados obtidos pelas aproximações analíticas. As aproximações (15), (27) e (49) (vide capítulo 2) apresentou resultados mais próximos de estoque em processo na rede (desvio relativo de $-0,9\%$) e de *leadtime* médio da rede (desvio relativo de $3,1\%$) aos obtidos por simulação, conforme as tabelas 4.8 e 4.9, mostrando que o método aproximado de decomposição é bastante útil e adequado para descrever o desempenho de redes reais.

Dentre os problemas de alocação/realocação de capacidade estudados (*SP1* e *SP2*), os modelos referentes à alocação de capacidade discreta para uma rede ($SP1.1/G/S/R^{discreto}$ e $SP1.2/G/S/R^{discreto}$) mostraram-se mais adequados para aplicação ao estudo de caso. Curvas de *trade-off* entre *WIP* e capital de investimento foram construídas (figura 4.3), variando parametricamente o limitante superior para o *WIP* no modelo $SP1.1/G/S/R^{discreto}$. O modelo $SP1.2/G/S/R^{discreto}$ auxilia um gerente a tomar decisões de alocação de capacidade quando a medida de desempenho desejada é o *leadtime* para as classes de produtos. Estes modelos demonstraram ser bastante úteis para um gerente tomar decisões quando ele deve reduzir o inventário em processo (modelo $SP1.1/G/S/R^{discreto}$) ou manter o *leadtime* das classes de produtos em níveis inferiores aos estabelecidos pelo mercado, por exemplo (modelo $SP1.2/G/S/R^{discreto}$). Os resultados de ambos os modelos foram obtidos em poucos segundos utilizando um computador pessoal (processador Pentium III de 800 MHz) o que reforça a aplicabilidade da metodologia aqui empregada: aplicação de modelos de otimização analíticos para prescrever configurações da rede, utilizando métodos aproximados de decomposição para avaliar o desempenho da rede em cada configuração.

As soluções do modelo $SP1.1/G/S/R^{discreto}$ também foram analisadas quando os coeficientes quadráticos de variação do intervalo de tempo entre chegadas externas das classes de produtos (ca'_k), as taxas de chegadas externas (λ'_k) da rede e o *mix* de produção são alterados. Em ambas as análises, o modelo demonstrou sensibilidade ao variar os parâmetros.

Convém destacar, neste ponto, algumas das principais contribuições desta tese. Mostra-se aqui a aplicabilidade e adequação das aproximações para avaliar o

desempenho do sistema em um estudo de caso real. Mostra-se também a utilidade da inserção das aproximações em procedimentos de otimização na tomada de decisões, especificamente nos problemas de alocação de capacidade (*SP1*), variando ou não alguns parâmetros da rede, para otimizar o desempenho do sistema em um caso real.

Com relação à classe *SP3*, foram estudados três tipos de problemas de redução de complexidade: PRC-1, PRC-2 e PRC-3. O modelo PRC-1 trata a redução de complexidade do ponto de vista de gestão do produto. A premissa básica é a de que um gerente deve ser capaz de prever de forma precisa o *leadtime* de um produto. Isto ocorre com a redução da variância do *leadtime* deste produto, que pode ser alcançada com aquisição de capacidade ou por meio da partição apropriada da instalação. Os modelos PRC-2 e PRC-3, por outro lado, tratam da redução de complexidade do ponto de vista de gestão da estação. De acordo com estes modelos, quanto maior a diversidade de produtos ou incertezas na chegada em uma estação, maior deve ser a flexibilidade da estação, que pode ser alcançada com aquisição de capacidade ou partição apropriada da instalação. Estes modelos não puderam ser aplicados à rede do estudo de caso, modelada neste trabalho como uma rede de filas *GI/G/1*. Por isso, foram utilizadas redes formadas por um único e por vários estágios *GI/G/m*, algumas concebidas para este trabalho e outras encontradas em literatura (Bitran e Sarkar, 1993c). O modelo PRC-3 não foi devidamente analisado neste trabalho, como os modelos PRC-1 e PRC-2, porque as redes usadas nos experimentos não satisfazem uma hipótese do modelo. A análise dos modelos PRC-1 e PRC-2 ocorreu no sentido de identificar sob quais condições é mais vantajoso particionar ou não a instalação, sem necessidade de aumentar a capacidade necessária, ou até reduzir esta capacidade, mantendo ou aumentando o desempenho da instalação.

No modelo PRC-1, é imposto o limitante superior para a variância ponderada do *leadtime* da rede. Em algumas redes com um e vários estágios *GI/G/m*, verificou-se que a agregação de uma linha que possui estações com alto nível de utilização e variabilidade do processo de chegadas baixo (processo quase determinístico), com outra linha que possui maior variabilidade do processo de chegadas não é vantajosa (vide tabelas 6.9, 6.20, 6.21 e 6.22). Ainda, ao realizar experimentos com uma das redes formada por vários estágios (Rede 4, proposta por Bitran e Sarkar, 1993c), observou-se que a agregação de produtos com tempos de

serviços bastante diferentes em uma mesma linha aumenta os valores de ca_{jp} e cs_{jp} nas estações que, por sua vez, aumentam a variância do *leadtime* e requerem mais capacidade para manter a rede a níveis inferiores ao limitante superior fixado no modelo PRC-1. Portanto, para estes casos, particionar a instalação parece ser mais vantajoso do que não particionar.

Entretanto, experimentos realizados com o modelo PRC-1 na Rede 4 mostraram que, se a aproximação utilizada para descrever o processo de desagregação das partidas for a de Segal e Whitt (1989), é vantajoso particionar a instalação (vide tabela 6.31) em termos de capacidade requerida. Por outro lado, se a aproximação utilizada para descrever o processo de desagregação das partidas for a de Bitran e Tirupati (1988), então a partição não é vantajosa (vide tabela 6.36) em termos de capacidade necessária. Convém mencionar que este experimento também foi realizado para a Rede 2 (rede, com 2 estágios e duas classes de produtos, criada pelo autor), mas os resultados obtidos empregando as aproximações de Segal e Whitt (1989) e Bitran e Tirupati (1988) foram iguais (vide fim da seção 6.1.2.1). Logo, o modelo PRC-1 demonstrou sensibilidade ao sistema de aproximações utilizado na Rede 4, e sugere-se pesquisa futura para estudar melhor a adequação e aplicação deste modelo em situações gerais.

No modelo PRC-2, o tempo médio de espera, dado que espera, de um item nas estações da rede ($E(Wq_{jp}|Wq_{jp}>0)$) é fixado na restrição de complexidade, ao impor o parâmetro G_{jp} na expressão (106). Observou-se, em algumas redes formadas por um único estágio e vários estágios $GI/G/m$ aqui estudadas, que é mais vantajoso agregar em uma mesma linha produtos semelhantes em termos de tempo médio de processamento e/ou de variabilidade do processo de chegadas. Portanto, a partição em linhas diferentes é viabilizada para classes de produtos apresentando tempos médios de processamento e/ou variabilidade razoavelmente distintos. A restrição de complexidade foi verificada por modelos de simulação para as configurações obtidas pelo modelo PRC-2. Constatou-se que, para a grande maioria das estações da rede, $E(Wq_{jp}|Wq_{jp}>0)$ foi inferior ao limitante superior fixado no modelo, demonstrando a eficácia da restrição. Além disso, notou-se que ao limitar superiormente o $E(Wq_{jp}|Wq_{jp}>0)$, a variância ponderada do *leadtime* da rede também é limitada.

Entretanto, assim como no modelo PRC-1, experimentos foram

conduzidos no PRC-2 para avaliar a sensibilidade do modelo ao sistema de aproximações utilizado. Novamente, na Rede 4, ao mudar-se a expressão que descreve o processo de desagregação das partidas de Segal e Whitt (1989) para a de Bitran e Tirupati (1988), a partição que requeria menor número de máquinas passou de duas linhas ([1-8],[9,10]) para uma única linha ([1-10]), respectivamente. Este resultado mostra a sensibilidade do modelo PRC-2 ao sistema de aproximações utilizado na Rede 4, e sugere-se pesquisa futura para estudar melhor a adequação e aplicação deste modelo em situações gerais.

Diante dos resultados obtidos para problema da classe *SP3*, cabe destacar algumas contribuições desta tese. Observa-se que, em certos casos, é possível melhorar o desempenho de uma rede de manufatura (planta) apenas particionando-a em subplantas, ou seja, sem a necessidade de investimentos adicionais em capacidade. Em alguns casos, é possível manter (ou até melhorar) o desempenho da rede, particionando os produtos em linhas e necessitando de menos capacidade do que a configuração com uma linha única (vide por exemplo os casos da tabela 6.31; os casos base, 1 e 2 da tabela 6.45; e os casos base e 3 da tabela 6.57). Ainda, realizou-se uma pesquisa exploratória inicial acerca das características da rede que sugerem quando é vantajoso particioná-la ou não. Mostra-se que, para minimizar o número de máquinas e manter o desempenho da rede, pode ser vantajoso agregar produtos com tempos de serviço médio e/ou variabilidade similares, e que não é vantajoso agregar uma linha com baixa variabilidade no processo de chegadas externas (tendendo a processo determinístico) e alta utilização nas estações, com outra linha com variabilidade no processo de chegadas moderada. Outra contribuição desta tese é a verificação da sensibilidade dos modelos PRC-1 e PRC-2 da classe *SP3* em certas redes, quando muda-se o sistema de aproximações utilizado. Convém ressaltar que estas questões do problema de partição da instalação não foram levantadas por Bitran e Sarkar (1993c).

De uma forma geral, a utilização de aproximações por decomposição para avaliar o desempenho de sistemas de manufatura discretos foi bastante útil. Primeiro, consome poucos recursos computacionais, conforme pôde-se notar com a rapidez com que os modelos foram resolvidos. Segundo, os resultados obtidos pelas aproximações descreveram razoavelmente bem o desempenho das redes testadas neste trabalho (inclusive a rede representando um sistema de manufatura real), quando os

resultados foram comparados aos obtidos por amostras e modelos de simulação. Terceiro, permitiu o emprego de modelos de otimização que auxiliam a tomada de decisão no projeto e planejamento de capacidade de sistemas produtivos. Entretanto, durante o desenvolvimento deste trabalho, algumas questões foram levantadas, sugerindo tópicos para pesquisa futura, conforme destacado a seguir.

6.2 Perspectivas de Pesquisa Futura

Uma perspectiva de pesquisa futura interessante é estudar melhor em projeto de experimentos a relação entre as características de redes de filas abertas e a aplicabilidade das diversas aproximações encontradas em literatura para avaliação de desempenho, por meio do método de decomposição paramétrico. O objetivo seria identificar a melhor adequação de um determinado sistema de aproximação para uma certa configuração de rede. Ou seja, dada uma configuração da rede (por exemplo, processando grande ou pequena quantidade de produtos, tráfego pesado ou leve, produtos muito ou pouco similares em termos de roteiro e de tempos de processamento, *mix* de produtos, produtos com alta ou pouca variabilidade no processo de chegadas externas à rede, etc.), qual sistema de aproximações seria o mais adequado e robusto para descrevê-la, dentre todos os sistemas disponíveis na literatura?

Outro tópico importante seria identificar a relação entre os sistemas de aproximações e as restrições de complexidade tratadas nos modelos PRC-1 e PRC-2. Ora, acredita-se que o estudo deste tema possa explicar porque a mudança do sistema de aproximações nos modelos PRC-1 e PRC-2, em certas redes, pode produzir resultados significativamente diferentes, apesar de os dois sistemas de aproximações descreverem razoavelmente bem o desempenho da rede (quando comparados a simulação). Este tópico também poderia envolver um planejamento estatístico de experimentos.

Outra perspectiva é a busca em literatura de aproximações e métodos de solução para avaliar o desempenho de sistemas de manufatura que consideram outras disciplinas de atendimento nas filas. Aqui, foram empregadas aproximações que consideram a disciplina *FCFS* (*first-come, first-served*), que descrevem suficientemente bem o desempenho de um sistema de manufatura real. Entretanto, é muito comum em sistemas de manufatura se empregar outras disciplinas, como *EDD* (*Earliest Due Date*)

ou *SPT* (*Shortest Processing Time*), no processamento dos itens na estação. No presente trabalho, para a rede do estudo de caso, foi estimado por simulação o *WIP* e o *leadtime* das classes de produtos utilizando as disciplinas *FCFS* e *SPT*. Em média, ao empregar a disciplina *SPT*, o *WIP* foi reduzido em 8,6% (desvio relativo) e o *leadtime* em 8,5% (desvio relativo) com relação a disciplina *FCFS*, conforme mostra a tabela 4.15. Para esta rede, essas diferenças são relativamente pequenas para o nível de decisão em consideração (tático-estratégicas), e a disciplina parece interferir pouco na estimativa das medidas. Entretanto, surgem outras perguntas: este resultado poderia ser muito diferente para outras redes? Como ficariam os modelos de alocação e realocação de capacidade para redes em que a disciplina da fila influencie significativamente a medição de desempenho? Daí, a importância de testar aproximações que consideram outras disciplinas de fila. Há poucos trabalhos em literatura reportando o assunto, mas sugere-se como referência de partida o trabalho de Shanthikumar e Buzzacot (1981), que apresenta uma aproximação para o tempo médio de espera em uma fila *GI/G/1* com disciplina *SPT*.

Outra perspectiva de pesquisa interessante seria a busca por uma rede real em que as hipóteses admitidas nos modelos PRC-1, PRC-2 e PRC-3 fossem satisfeitas. Uma questão a ser respondida seria se as restrições de complexidade realmente são efetivas em limitar a variabilidade do *leadtime* da rede (PRC-1), o tempo médio de espera dos itens em filas (PRC-3), ou o tempo médio de espera em fila, dado que esperam para serem servidos (PRC-2). A validação da restrição pode ocorrer por meio de simulação. Ainda, poderia ser estabelecida uma comparação efetiva entre os benefícios atingidos ao aplicar os modelos PRC-1, PRC-2 e/ou PRC-3 com os benefícios propostos pela filosofia de produção enxuta que, de certa forma, preconiza a partição da instalação para formação de células. O objetivo seria observar a influência da redução de complexidade da instalação (ao particioná-la ou não, com a aplicação dos modelos PRC-1, PRC-2 e PRC-3) nas 7 formas de desperdícios da produção enxuta: superprodução, defeitos, inventários desnecessários, processamento inapropriado, movimentação desnecessária de bens ou informação, esperas e movimentação excessiva de pessoas (Rother e Shook, 1998).

Por último, pesquisar outras formas de resolver o modelo (102a) – (102d). Por exemplo, melhorando a heurística construtiva descrita no algoritmo 6, ou

utilizando outros métodos, tais como heurísticas de busca local e metaheurísticas do tipo algoritmo genético e busca tabu. Além disso, poderia ser considerada a adaptação e aplicação dos modelos e métodos de solução aqui empregados em outras áreas além de sistemas de manufatura como redes de serviço (por exemplo, centrais de chamadas), redes computacionais e redes de comunicação.

REFERÊNCIAS

- Albin, S. L. (1982). "Poisson Approximations for Superposition Arrival Processes in Queues". *Management Science*, 28(2), pg 126-137.
- Albin, S. L. (1984). "Approximating a Point Process by a Renewal Process, II: Superposition Arrival Process of Queues". *Operations Research*, 32(5), pg 1133-1162.
- Banks, J.; Carson, J. S.; Nelson, B. L. (1996). *Discrete-Event System Simulation*. Prentice Hall, 2ª edição, New Jersey, 548pg.
- Baskett, F.; Chandy, K. M.; Muntz, R. R.; Palacios, F. G. (1975). "Open, closed, and mixed networks of queues with different classes of customers." *Journal of the Association for Computing Machinery* 22(2), 248-260pg.
- Bitran G. R. e Dasu, S. (1992). "A review of open queueing network models of manufacturing systems." *Queueing systems* 12, pg. 95-134.
- Bitran, G. R. e Morabito, R. (1995). "Um exame dos modelos de redes de filas abertas aplicados a sistemas de manufatura discretos – Parte I". *Gestão & Produção*, 2(2), pg 192-219.
- Bitran, G. R. e Morabito, R. (1996). "Open Queueing networks: Optimization and performance evaluation models for discrete manufacturing systems". *Production and Operations Management* 5(2), pg.163-193.
- Bitran, G. R. e Morabito, R. (1999). "An overview of trade-off curves in manufacturing systems design". *Production and Operations Management* 8(1), pg.56-75.
- Bitran, G. R. e Sarkar, D. (1993). "Focused factory design: complexity, capacity and inventory trade-offs". *Technical Memorandum*. AT&T Bell Lab., Holmdel, NJ.
- Bitran, G. R. e Sarkar, D. (1994a). "Throughput Analysis in Manufacturing Networks.". *European Journal of Operational Research*, n° 74, pg 501-510.
- Bitran, G. R. e Sarkar, D. (1994b). "Targeting Problems in Manufacturing Queueing Networks – An Iterative Scheme and Convergence". *European Journal of Operational Research*, n° 76, pg 501-510.
- Bitran, G. R. e Tirupati, D. (1988). "Multiproduct queueing networks with deterministic routing: Decomposition approach and the notion of interference". *Management Science*. 34(1), pg. 75-100.
- Bitran, G. R. e Tirupati, D. (1989a). "Tradeoff Curves, Targeting and Balancing in Manufacturing Queueing Networks". *Operations Research*, 37(4), pg. 547-564.
- Bitran, G. R. e Tirupati, D. (1989b). "Capacity Planning in Manufacturing Networks with Discrete Options". *Annals of Operations Research*, 17, pg 119-136.

- Boxma, O. J.; Rinnooy Kan, A.; Van Vliet, M. (1990). "Machine Allocation Problems in Manufacturing Networks". *European Journal of Operational Research*, 45, pg 47-54.
- Brooke, A.; Kendrick, D.; Meeraus, A.; Raman, R.; Rosenthal, R. E. (1998). *GAMS: A User's Guide*. Gams Development Corporation, Washington. 262 pg
- Buzacott, J. A. ; Shanthikumar, J. G. (1993). *Stochastic models of manufacturing systems*, Prentice-Hall, Englewood Cliffs, New Jersey.
- Çınlar, E. (1975). *Introduction to Stochastic Processes*. Prentice Hall, Inc., New Jersey. 402pg.
- Flyn, B. B. ; Jacobs, F. R. (1986). "A simulation comparison of group technology with traditional job shop manufacturing". *International Journal of Production Research*, 34(5) , pg. 1171-1192.
- Flyn, B. B.; Jacobs, F. R. (1987). "An experimental comparison of cellular (group technology) layout with process layout". *Decision Sciences*, 18(4), pg. 562-581.
- Jackson, J. R. (1957). "Networks of waiting lines." *Operations Research* 5(4), 518-521pg.
- Jackson, J. R. (1963). "Job shop-like queueing systems". *Management Science* 10(1), 131-142pg.
- Jensen, J. B. et al. (1996). "Machine dedication and process flexibility in a group technology environment". *Journal of Operations Management*, vol 14, pg 19-39.
- Kelly, F. P. (1975). "Networks of queues with customers of different types." *Journal of Applied Probability* 12, 542-554pg.
- Kelly, F. P. (1975). *Reversibility and Stochastic Processes*. John Wiley & Sons, New York.
- Kleinrock, L. (1964). *Communication Nets: Stochastic Message Flow and Delay*, Dover Publishing, New York.
- Kleinrock, L. (1975). *Queueing Systems, Volume 1*. John Wiley & Sons, New York. 417pg.
- Kleinrock, L. (1976). *Queueing Systems, Volume 2*. John Wiley & Sons, New York. 417pg.
- Krajewski, L. J. ; Ritzman, L. P. (1996) . *Operations management: strategy and analysis*. 4^a ed. Addison-Wesley Publishing Company, Massachusetts.
- Kuehn, P. J. (1979). "Approximate Analysis of General Networks by Decomposition". *IEEE Transactions Commun*, 27(1), pg 113-126.
- Larson, R. C.; Odoni, A. R. (1981). *Urban Operations Research*. 1^a ed. Prentice-Hall Inc., New Jersey, 573pg.

- Law, A. M. ; Kelton, W. D. (1991). *Simulation Modeling and Analysis*. 2^a ed. McGraw-Hill. New York, 759pg.
- Magalhães, M. N. (1996). *Introdução à Rede de Filas*. 12^o SINAPE (Simpósio Nacional de Probabilidade e Estatística). Associação Brasileira de Estatística, Caxambu, SP, 116pg.
- Morabito, R. (1998). *Análise de curvas de trade-off baseada em teoria de redes de filas para o projeto e planejamento de sistemas discretos de manufatura*. São Carlos, tese livre-docência, EESC/USP, 136pg.
- Morris, J. S.; Tersine, R. J. (1990). "A simulation analysis of factors influencing the attractiveness of group technology cellular layouts". *Management Science* 36(12), pg. 1567-1578.
- Rother, M. ; Shook, John (1998). *Aprendendo a Enxergar: mapeando o fluxo de valor para agregar valor e eliminar o desperdício*. Lean Institute Brasil, São Paulo, 100pg.
- Schriber, T. J. (1991). *An Introduction to Simulation Using GPSS/H*. John Wiley & Sons, New York, 437pg.
- Segal, M.; Whitt, W. (1989). "A Queueing Network Analyzer for Manufacturing", in *Teletraffic Science for New-Effective System, Networks and Services*, ITC-12, M. Bonatti (ed.), Elsevier, North-Holland, Amsterdam, pg 1146-1152.
- Sevcik, K. C.; Levy, A. I.; Tripathi, S. K.; Zahoran, J. L. (1977). "Improving Approximations of Aggregated Queueing Network Systems", in *Computer Performance*, K. Chandy e M. Reiser (eds). North-Holland, pg 1-22.
- Shanthikumar, J. G.; Buzacott, J. A. (1981). "Open queueing network models of dynamic job shops". *International Journal of Production Research* 19(3), 255-266pg.
- Skinner, W. (1974). "The focused factory". *Harvard Business Review*. pg 113-121.
- Souza, G. S; Wagner, H. M.; Whybark, D. C. (2001). "Evaluating focused factory benefits with queueing theory". *European Journal of Operational Research*, vol 128, pg 597-610.
- Suresh, N. C. (1991). "Partitioning work center for group technology: insights from an analytical model". *Decision Sciences*, 22(4), pg 772-791.
- Suresh, N. C. (1992). "Partitioning work center for group technology: analytical extension and shop-level simulation investigation". *Decision Sciences*, 23(2), pg 267-290.
- Suresh, N. C. e Meredith, J. R. (1994). "Coping with the loss of pooling synergy in cellular manufacturing systems". *Management Science*, 40(4), pg 466-483.

- Taha, H. A. (1997). *Operations Research: An Introduction*. 6^a ed. Prentice Hall, New Jersey.
- Tijms, H. C. (1986). *Stochastic modeling and analysis: A computational approach*. John Wiley & Sons, New York.
- Van Vliet, M. e Rinnooy Kan, A. (1991). "Machine Allocation Algorithms for Job-Shop Manufacturing". *Journal of Intelligent Manufacturing*, vol 2, pg 83-94
- Whitt, W. (1982). "Approximating a Point Process by a Renewal Process, I: Two Basic Methods." *Operations Research*, 30(1), pg 125-147.
- Whitt, W. (1983a). "The queueing network analyzer". *The Bell System Technical Journal*, 62(9), 2779-2815 pg.
- Whitt, W. (1983b). "The queueing network analyzer". *The Bell System Technical Journal*, 63(9), 1911-1979pg.
- Whitt, W. (1988). "A Light-Traffic Approximation for Single-Class Departure Processes from Multi-Class Queues". *Management Science*, 34(11), pg 1333-1346.
- Whitt, W. (1992). "Understanding the Efficiency of Multi-Server Service Systems". *Management Science*, 38(5), pg 708-723.
- Whitt, W. (1993). "Approximations for the GI/G/m Queue". *Production and Operations Management*. 2(2), pg 114-161.
- Whitt, W. (1994). "Towards Better Multi-Class Parametric-Decomposition Approximations for Open Queueing Networks". *Annals of Operations Research*, 48, pg 221-248.
- Wolff, R. W. (1989). *Stochastic Modeling and the Theory of Queues*. Prentice Hall, Inc., New Jersey. 556pg.

APÊNDICE A – ALGORITMOS DE OTIMIZAÇÃO

Este apêndice apresenta os algoritmos desenvolvidos por Bitran e Tirupati (1989a) para resolver os problemas $SP1.I/G/S/R$ e $SP2.I/G/S/R$ e o algoritmo proposto por Bitran e Tirupati (1989b) baseado no problema $SP1.I^{relaxado}/G/M/R$ para obter uma solução aproximada para o problema $SP1.I/G/M/R$, quando as alternativas de adição de capacidade são discretas.

Algoritmo A1

Bitran e Tirupati (1989a) apresentaram uma heurística, baseada na estratégia gulosa para resolver o problema $SP1.I/G/S/R$ e para gerar curvas de *trade-off* entre $F(\boldsymbol{\mu})$ e $L(\boldsymbol{\mu})$. Seja $PI_j(\mu_j)$ um índice de prioridade definido como o quociente entre o aumento do custo marginal e a redução do *WIP* marginal na estação j , dado por:

$$PI_j(\mu_j) = \frac{\partial F_j(\mu_j) / \partial \mu_j}{-v_j \partial L_j(\mu_j) / \partial \mu_j} \quad (\text{A.1})$$

Seja Δ um incremento previamente especificado de capacidade em cada iteração. Inicia-se com uma capacidade inicial satisfazendo (60) – vide capítulo 3 – para todas as estações. Em cada iteração, aumenta-se em Δ (valor fixado *a priori*) a capacidade da estação com o menor índice de prioridade (A.1). Este procedimento é repetido até que o *WIP* desejado (L_T) é atingido.

Algoritmo A1

Passo 1: Calcule o número médio de *jobs* em cada estação da rede ($L_j(\mu_j)$, $j=1, \dots, n$), resolvendo o sistema (65) e (20) do capítulo 3 e o *WIP* da rede, $L(\boldsymbol{\mu})$, conforme (68), capítulo 3, para a configuração inicial da rede. (Esta solução é infactível, $L(\boldsymbol{\mu}^0) > L_T$, e o custo correspondente, $F(\boldsymbol{\mu}^0)$, é menor do que o custo mínimo de $SP1.I/G/S/R$. Calcule o índice de prioridade para cada estação conforme (A.1)).

Passo 2: Se $L(\boldsymbol{\mu}) \leq L_T$, então pare. Senão vá para o passo 3.

Passo 3: Faça:

a) escolha a estação j^* , correspondente à estação com menor índice de prioridade (A.1), para adicionar o incremento de capacidade:

$$PI_{j^*} = \min \{PI_j(\mu_j), j=1, \dots, n\} \quad (A.2)$$

b) $\mu_{j^*} \leftarrow \mu_{j^*} + \Delta$, atualize L_{j^*} , PI_{j^*} , ρ_j e calcule $L(\mu)$. Vá para o passo 2.

Bitran e Tirupati (1989a) mostraram que se as hipóteses H1 e H2 (vide seção 3.4.1.1) são satisfeitas, então a solução apresentada pelo Algoritmo A1 resolve o problema *SP1.1/G/S/R* otimamente (condições (i)–(iv) são satisfeitas) no limite quando $\Delta \rightarrow 0$ e $PI_j(\mu_j)$ obtido na última iteração corresponde ao multiplicador dual associado com a restrição de *WIP* na estação j .

Bitran e Tirupati (1989a) também apresentaram um limite para o erro para o valor da solução aproximada produzida pelo Algoritmo A1. Suponha que o algoritmo encontra uma solução factível após p iterações e denote esta solução por μ^p e a solução ótima de *SP1.1/G/S/R* por μ^* . Então,

$$0 \leq F(\mu^p) - F(\mu^*) \leq \frac{(L_T - L(\mu^p))}{PI_{j^*}^p} + \delta \quad (A.3)$$

onde $\delta = \Delta \sum_{i=1}^p (1 - PI_{j^*}^i / PI_{j^*}^p)$ e $PI_{j^*}^i$ é o quociente obtido por (A.2) na iteração i , $i=1, \dots, p$. Experimentos computacionais com $\Delta=0,1$ para dados de uma rede real com 13 estações e 10 classes de produtos resultaram em erro relativo de 0,6% entre $F(\mu^p)$ e $F(\mu^*)$, o que é aceitável para muitas situações reais. Estes experimentos também indicaram que a hipótese de considerar *ca* e *cs* independentes de mudanças em μ é razoável.

Algoritmo A2

Bitran e Tirupati (1989a) apresentaram uma heurística também baseada na estratégia gulosa para resolver o problema *SP2.1/G/S/R* (condições (i) – (iv) são satisfeitas).

Algoritmo A2

Passo 1: Seja J_0 o conjunto de todas estações e $J_1=J_2=\phi$. J_1 é o conjunto de estações para as quais capacidade é adicionada e J_2 o conjunto das quais capacidade é removida. Calcule as medidas de desempenho da rede $L_j(\mu_j)$, utilizando (20) do

capítulo 3, e $L(\boldsymbol{\mu})$ para a configuração inicial da rede que deve ser factível ($\mu_j = \mu_j^1$). Calcule também, para todas estações da rede, o índice de prioridade $PI_j(\mu_j)$ definido como a redução marginal de WIP na estação j dado por:

$$PI_j(\mu_j) = -v_j \frac{\partial L_j(\mu_j)}{\partial \mu_j} \quad (\text{A.4})$$

Calcule $\varepsilon_j, j=1, \dots, n$; tal que $PI_j(\mu_j)(\lambda_j + \varepsilon_j) = \max\{PI_j(\mu_j), j \in J_0\}$

Passo 2: Se $J_0 = \emptyset$ ou $[J_0] = 1$ (conjunto unitário) ou $PI_j(\mu_j)$ é igual para todas estações em J_0 então pare. Senão vá para o passo 3.

Passo 3:

a) Encontre a estação j_1 que resulte no menor PI_{j_1} (aquela estação cuja capacidade será removida) dado por:

$$PI_{j_1} = \min \{PI_j(\mu_j), j \in J_0\} \quad (\text{A.5})$$

e a estação j_2 que resulta no maior PI_{j_2} (aquela estação na qual ocorrerá adição de capacidade) dado por:

$$PI_{j_2} = \max \{PI_j(\mu_j), j \in J_0\} \quad (\text{A.6})$$

Se $j_1 \notin J_1$ e $j_2 \notin J_2$ então vá para o passo 4.

b) Se $j_1 \in J_1$, então faça $J_0 \leftarrow J_0 - \{j_1\}$. Se $j_2 \in J_2$, então faça $J_0 \leftarrow J_0 - \{j_2\}$.

Vá para o passo 2.

Passo 4: Faça: $\Delta_l = \min\{\Delta, \mu_{j_1} - \lambda_{j_1} - \varepsilon_{j_1}\}$ e faça

$$\mu_{j_1} \leftarrow \mu_{j_1} - \Delta_l, \mu_{j_2} \leftarrow \mu_{j_2} + \Delta_l, J_1 \leftarrow J_1 \cup \{j_2\} \text{ e } J_2 \leftarrow J_2 \cup \{j_1\}.$$

Calcule os novos valores de L_{j_1} , L_{j_2} , PI_{j_1} , PI_{j_2} e $L(\boldsymbol{\mu})$. Vá para o passo 2.

O cálculo de ε_j no passo 1 do algoritmo em conjunção com o cálculo de Δ_l no passo 4 garante que, sem perda de generalidade, $\mu_j \geq \lambda_j + \varepsilon_j$ para todo j por todos passos do algoritmo. O passo 3b tenta evitar ciclos devido ao uso de passos (Δ_l) de tamanho finito. Na solução ótima, todos os $PI_j(\mu_j)$'s são iguais e interpretados como os multiplicador dual associado à restrição de capacidade.

Bitran e Tirupati (1989a, Proposição 4 do referido trabalho) também apresentaram um limitante para o erro no valor da solução aproximada produzida pelo

Algoritmo A2. Seja μ^p a solução heurística encontrada na iteração p e seja μ^* a solução ótima de $SP2.1/G/S/R$. Então,

$$0 \leq F(\mu^p) - F(\mu^*) \leq n\Delta PI_{j_2}^p \quad (\text{A.7})$$

em que $PI_{j_2}^p$ é o índice de prioridade obtido de (A.6) na última iteração p . Observe em (A.7) que no limite quando $\Delta_l \rightarrow 0$ a solução aproximada é exata.

Na formulação do problema $SP2.1/G/S/R$ (vide seção 3.4.1.2), a restrição

$$\sum_{j=1}^n \mu_j = \sum_{j=1}^n \mu_j^1 \quad (\text{A.8})$$

implica que a taxa total de serviço na rede é fixa. Em muitas situações reais, o problema de balanceamento consiste em manter a disponibilidade total fixa (adotando como medida hora-homem ou hora-máquina) ao invés da taxa de serviço. Para este caso, (A.8) pode ser substituída por:

$$\sum_{j=1}^n b_j \mu_j = \sum_{j=1}^n b_j \mu_j^1 \quad (\text{A.9})$$

em que b_j é uma constante positiva, representando, por exemplo, a quantidade média de horas-homem requerida por item/produto na estação j . Os resultados dessa seção, com poucas modificações, podem ser estendidos a esta situação também. O Algoritmo A2 é modificado da seguinte maneira: o índice de prioridade é redefinido como:

$$PI_j(\mu_j) = -\frac{v_j}{b_j} \frac{\partial L_j(\mu_j)}{\partial \mu_j}$$

e o incremento de capacidade Δ (e Δ_l), transferido entre estações no passo 4 do algoritmo, deve estar em unidades de disponibilidade, $b_j \mu_j$, ao invés de taxas de serviço. O limitante para o erro (Proposição 4) permanece o mesmo.

Algoritmo A3

Bitran e Tirupati (1989b) apresentaram uma heurística baseada no problema $SPI.1^{relaxado}/G/M/R$ para obter uma solução aproximada para o problema $SPI.1/G/M/R$. Mostraram que: a) a solução ótima tem nenhuma ou no máximo duas variáveis $u_{j,k}$ com valores fracionários (Proposição 3.1 do referido trabalho) e; b) se a

solução ótima tem duas variáveis $u_{j,k}$ com valores fracionários, então elas correspondem à mesma estação (Corolário). O Algoritmo A3 produz a solução aproximada \mathbf{u}^1 ($u_{jk}^1, j=1, \dots, n; k=1, \dots, n_j$). Nota: a notação aqui utilizada é a mesma utilizada na seção 3.4.1.

Algoritmo A3

Passo 1: Seja \mathbf{u}^0 a solução ótima ao problema linear $SPI.1^{\text{relaxado}}/G/M/R$, satisfazendo a Proposição 3.1, enunciada em Bitran e Tirupati (1989b). Se \mathbf{u}^0 é uma solução factível para $SPI.1/G/M/R$, então $\mathbf{u}^1 = \mathbf{u}^0$ é uma solução ótima para $SPI.1/G/M/R$. Senão, vá para o passo 2.

Passo 2: Seja i a estação cujas variáveis são valores fracionários para algumas alternativas k_1 e k_2 ($0 < u_{ik_1}^0 < 1, 0 < u_{ik_2}^0 < 1$). Uma solução factível para $SPI.1/G/M/R$ é dada por:

Para as estações $j \neq i, j=1, \dots, n; k=1, \dots, n_j$

$$u_{jk}^1 = u_{jk}^0$$

Para a estação $j=i$

$$u_{ik}^1 = \begin{cases} 1, & \text{se } k = l \\ 0, & \text{caso contrário} \end{cases}$$

onde l é tal que: $L_{il} = \max \{L_{ik}/L_{ik} \leq L_{ik_1} u_{ik_1}^0 + L_{ik_2} u_{ik_2}^0, k=1, \dots, n_i\}$.

Os autores também apresentaram um limitante para o erro para o valor da solução aproximada gerada pelo Algoritmo A3. Sem perda de generalidade, admita que $L_{ik_1} > L_{ik_2}$ e denote por \mathbf{u}^* a solução ótima de $SPI.1/G/M/R$. Então,

$$0 \leq F(\mathbf{u}^1) - F(\mathbf{u}^*) \leq f_{ik_2} - f_{ik_1} \leq \max\{f_{jk}, j=1, \dots, n, k=1, \dots, n_j\}$$

APÊNDICE B – MODELO DE SIMULAÇÃO EM GPSS/H®

Este anexo apresenta o modelo de simulação desenvolvido no *software GPSS/H®*. Como a finalidade é ilustrar como o modelo foi construído, no programa abaixo constam somente as distribuições empíricas (de chegada e de serviço) e o roteiro de fabricação da classe de produtos 1. As demais classes de produtos foram omitidas, por economia de espaço, mas a implementação é análoga à classe 1.

O modelo de simulação considerando um dia de operação efetiva igual a 16,7 horas, é semelhante ao modelo abaixo, embora as distribuições empíricas de chegada de cada classe de produto mudem. Foram realizadas 10 réplicas, sendo que em cada uma as 100 mil primeiras chegadas, das 200 mil geradas, foram desprezadas para desconsiderar o estado transiente (*warm-up period*).

Modelo de Simulação implementado para dia de operação efetiva igual a 17,6 horas.

```

SIMULATE
REALLOCATE COM,1000000
*****
INTEGER  &I,&J
REAL &A(16),&B(16),&C(16),&D(16),&F(14),&H(14),&L(14),&N(14),&M,&T
CHAR*8 &E(16),&G(14)
*****
*   DISTRIBUICOES EMPIRICAS DE CHEGADA E SERVICO   *
*****
CHEG1  FUNCTION RN7,C13 distribuicao chegada classe 1
0,88/0.083,105.6/0.167,123.2/0.250,140.8/0.333,193.6
0.417,211.2/0.500,281.6/0.583,352/0.667,369.6/0.750,387.2
0.833,387.2/0.917,457.6/1,739.2
*
SERV11 FUNCTION RN7,C13 distribuicao classe 1 operacao 1
0,2.58/0.077,3.50/0.154,3.57/0.231,3.59/0.308,4.27
0.385,4.46/0.462,5.27/0.538,6.02/0.615,7/0.692,7.24
0.769,7.7/0.923,8.34/1,8.75
*
SERV12 FUNCTION RN7,C13 distribuicao classe 1 operacao 2
0,12.46/0.077,16.91/0.154,17.25/0.231,17.36/0.308,20.63
0.385,21.54/0.462,25.48/0.538,29.09/0.615,33.83/0.692,34.95
0.769,37.21/0.923,40.31/1,42.28
*
SERV13 FUNCTION RN7,C13 distribuicao classe 1 operacao 3
0,2.14/0.077,2.90/0.154,2.96/0.231,2.98/0.308,3.54
0.385,3.69/0.462,4.37/0.538,4.99/0.615,5.8/0.692,5.99
0.769,6.38/0.923,6.91/1,7.25
*
SERV14 FUNCTION RN7,C13 distribuicao classe 1 operacao 4

```

0,2.47/0.077,3.35/0.154,3.41/0.231,3.44/0.308,4.08
 0.385,4.26/0.462,5.04/0.538,5.76/0.615,6.69/0.692,6.92
 0.769,7.36/0.923,7.98/1,8.37

*

SERV15 FUNCTION RN7,C13 distribuicao classe 1 operacao 5

0,1.05/0.077,1.43/0.154,1.45/0.231,1.46/0.308,1.74
 0.385,1.82/0.462,2.15/0.538,2.45/0.615,2.85/0.692,2.95
 0.769,3.14/0.923,3.40/1,3.56

...

As distribuições de chegada e de serviço das classes 2-14 para todas as operações / foram omitidas.

...

* NUMERO DE SERVIDORES EM PARALELO NAS 16 ESTACOES

STATION1 STORAGE 1
 STATION2 STORAGE 1
 STATION3 STORAGE 1
 STATION4 STORAGE 1
 STATION5 STORAGE 1
 STATION6 STORAGE 1
 STATION7 STORAGE 1
 STATION8 STORAGE 1
 STATION9 STORAGE 1
 STATIO10 STORAGE 1
 STATIO11 STORAGE 1
 STATIO12 STORAGE 1
 STATIO13 STORAGE 1
 STATIO14 STORAGE 1
 STATIO15 STORAGE 1
 STATIO16 STORAGE 1

* Roteiro da Classe 1 *

GENERATE FN(CHEG1) geracao de chegadas
 QUEUE 1 entrada do job no sistema
 QUEUE STATION2
 ENTER STATION2
 DEPART STATION2
 ADVANCE FN(SERV11)
 LEAVE STATION2

*

QUEUE STATION3
 ENTER STATION3
 DEPART STATION3
 ADVANCE FN(SERV12)
 LEAVE STATION3

*

QUEUE STATIO10
 ENTER STATIO10
 DEPART STATIO10
 ADVANCE FN(SERV13)
 LEAVE STATIO10

*

QUEUE STATIO15
 ENTER STATIO15
 DEPART STATIO15
 ADVANCE FN(SERV14)
 LEAVE STATIO15

```

*
  QUEUE  STATIO16
  ENTER  STATIO16
  DEPART STATIO16
  ADVANCE FN(SERV15)
  LEAVE  STATIO16
  DEPART 1          saida do job do sistema
*
  TERMINATE 1

*...
* Roteiros de Fabricação das classes de produtos 2-14 foram omitidos
*....
*****
*   Comandos Run-Control                                     *
*****
  DO    &I=1,10,1   controle para 10 replicas
        START 100000
        PUTPIC LINES=3,FILE=SYSPRINT,(&I)
0=====
O Relatorio acima eh para o periodo "transiente" da replica *.
=====
  RESET          reset para a i-esima replica
  START 100000   re-start para a i-esima replica
  PUTPIC LINES=3,FILE=SYSPRINT,(&I)
0=====
O Relatorio acima eh para o periodo "estacionario" da replica *.
=====
* Atribuicao do numero medio de jobs em servico nas estacoes ao vetor &A
  LET &A(1)=SA(STATION1)
  LET &A(2)=SA(STATION2)
  LET &A(3)=SA(STATION3)
  LET &A(4)=SA(STATION4)
  LET &A(5)=SA(STATION5)
  LET &A(6)=SA(STATION6)
  LET &A(7)=SA(STATION7)
  LET &A(8)=SA(STATION8)
  LET &A(9)=SA(STATION9)
  LET &A(10)=SA(STATIO10)
  LET &A(11)=SA(STATIO11)
  LET &A(12)=SA(STATIO12)
  LET &A(13)=SA(STATIO13)
  LET &A(14)=SA(STATIO14)
  LET &A(15)=SA(STATIO15)
  LET &A(16)=SA(STATIO16)
* Atribuição do numero de jobs nas filas das estacoes ao vetor &B
  LET &B(1)=QA(STATION1)
  LET &B(2)=QA(STATION2)
  LET &B(3)=QA(STATION3)
  LET &B(4)=QA(STATION4)
  LET &B(5)=QA(STATION5)
  LET &B(6)=QA(STATION6)
  LET &B(7)=QA(STATION7)
  LET &B(8)=QA(STATION8)
  LET &B(9)=QA(STATION9)
  LET &B(10)=QA(STATIO10)

```

```

LET &B(11)=QA(STATIO11)
LET &B(12)=QA(STATIO12)
LET &B(13)=QA(STATIO13)
LET &B(14)=QA(STATIO14)
LET &B(15)=QA(STATIO15)
LET &B(16)=QA(STATIO16)

```

* Atribuição dos caracteres correspondentes aas estacoes ao vetor &E

```

LET &E(1)='EST 1'
LET &E(2)='EST 2'
LET &E(3)='EST 3'
LET &E(4)='EST 4'
LET &E(5)='EST 5'
LET &E(6)='EST 6'
LET &E(7)='EST 7'
LET &E(8)='EST 8'
LET &E(9)='EST 9'
LET &E(10)='EST 10'
LET &E(11)='EST 11'
LET &E(12)='EST 12'
LET &E(13)='EST 13'
LET &E(14)='EST 14'
LET &E(15)='EST 15'
LET &E(16)='EST 16'

```

* Atribuição do leadtime de cada classe de job ao vetor &F

```

LET &F(1)=QT(1)
LET &F(2)=QT(2)
LET &F(3)=QT(3)
LET &F(4)=QT(4)
LET &F(5)=QT(5)
LET &F(6)=QT(6)
LET &F(7)=QT(7)
LET &F(8)=QT(8)
LET &F(9)=QT(9)
LET &F(10)=QT(10)
LET &F(11)=QT(11)
LET &F(12)=QT(12)
LET &F(13)=QT(13)
LET &F(14)=QT(14)

```

* Atribuição do numero medio de jobs de cada classe no sistema ao vetor &L

```

LET &L(1)=QA(1)
LET &L(2)=QA(2)
LET &L(3)=QA(3)
LET &L(4)=QA(4)
LET &L(5)=QA(5)
LET &L(6)=QA(6)
LET &L(7)=QA(7)
LET &L(8)=QA(8)
LET &L(9)=QA(9)
LET &L(10)=QA(10)
LET &L(11)=QA(11)
LET &L(12)=QA(12)
LET &L(13)=QA(13)
LET &L(14)=QA(14)

```

* Atribuição dos caracteres correspondentes a cada classe ao vetor &G

```

LET &G(1)='CLAS 1'
LET &G(2)='CLAS 2'
LET &G(3)='CLAS 3'

```

```

LET &G(4)='CLAS 4'
LET &G(5)='CLAS 5'
LET &G(6)='CLAS 6'
LET &G(7)='CLAS 7'
LET &G(8)='CLAS 8'
LET &G(9)='CLAS 9'
LET &G(10)='CLAS 10'
LET &G(11)='CLAS 11'
LET &G(12)='CLAS 12'
LET &G(13)='CLAS 13'
LET &G(14)='CLAS 14'
DO &J=1,16,1
LET &C(&J)=&A(&J)+&B(&J) soma os usuarios em operacao+fila espera
ENDDO
PUTPIC LINES=5,FILE=RESP,(&I)
0=====
DADOS DA REPLICA *.
NUMERO MEDIO EM NUMERO MEDIO NA NUMERO MEDIO NO SISTEMA
ESTACAO OPERACAO (S) FILA DE ESPERA (F) (F+S)
-----
DO &J=1,16,1
PUTPIC FILE=RESP,(&E(&J),&A(&J),&B(&J),&C(&J))
***** ** **** ** **** ** ****
LET &D(&J)=&D(&J)+&C(&J) soma (S+F) p/ calculo de media nas 10 replicas
LET &M=&M+&C(&J) soma o numero de jobs em todas estacoes para cada replica
ENDDO
PUTPIC LINES=2,FILE=RESP,(&I,&M)
-----
O NUMERO TOTAL DE LOTES NO SISTEMA NA REPLICA ** EH DE ** ****
PUTPIC LINES=4,FILE=RESP,(&I)
0=====
DADOS DA REPLICA *.
CLASSE LEADTIME MEDIO (HORAS)
-----
DO &J=1,14,1
PUTPIC FILE=RESP,(&G(&J),&F(&J))
***** **** **
LET &H(&J)=&H(&J)+&F(&J) soma do leadtime p/ calculo de media nas 10 replicas
ENDDO
PUTPIC LINES=4,FILE=RESP,(&I)
0=====
DADOS DA REPLICA *.
CLASSE NUMERO MEDIO DE LOTES POR CLASSE NO SISTEMA (LOTES)
-----
DO &J=1,14,1
PUTPIC FILE=RESP,(&G(&J),&L(&J))
***** *** ****
LET &N(&J)=&N(&J)+&L(&J) soma o numero de lotes p/ calculo da media nas 10 replicas
ENDDO
CLEAR
LET &M=0
ENDDO procede para a proxima iteracao
PUTPIC FILE=RESP,LINES=3
ESTACAO NUMERO MEDIO DE USUARIOS
DAS 10 REPLICAS NO SISTEMA
-----
DO &J=1,16,1

```



```

      LET &D(&J)=&D(&J)/10  valor medio do numero de lotes no sistema nas 10 replicas
      PUTPIC FILE=RESP,(&E(&J),&D(&J))
*****
      LET &T=&T+&D(&J)    total na rede apos 10 replicas
      ENDDO
      PUTPIC FILE=RESP,(&T)
      O NUMERO MEDIO DE USUARIOS NA REDE APOS 10 REPLICAS EH **.****.
      PUTPIC FILE=RESP,LINES=3
0=====
CLASSE    LEADTIME MEDIO (HORAS) DAS 10 REPLICAS
-----
      DO  &J=1,14,1
          LET &H(&J)=&H(&J)/10  valor medio do leadtime nas 10 replicas
          PUTPIC FILE=RESP,(&G(&J),&H(&J))
*****
          ENDDO
          PUTPIC FILE=RESP,LINES=3
0=====
CLASSE    NUMERO MEDIO (LOTES) DAS 10 REPLICAS PARA CADA CLASSE
-----
      DO  &J=1,14,1
          LET &N(&J)=&N(&J)/10  numero medio de lotes obtido nas 10 replicas
          PUTPIC FILE=RESP,(&G(&J),&N(&J))
*****
          ENDDO
      END          fim

```

APÊNDICE C – MODELO EM *GAMS/CPLEX*[®]

Este apêndice apresenta o modelo desenvolvido em *GAMS/CPLEX*[®] para solucionar o modelo *SPI.1/G/S/R^{discreto}*, cujos resultados e análise constam no capítulo 4.

```

$offtext;
$eolcom #
* optcr eh o criterio de parada do MIP - gap de otimalidade;
* iterlim eh o limite de iteracoes no solver
* reslim eh o tempo limite para o solver (em segundos).
options OPTCR = 0.0, ITERLIM = 100000, reslim=10000;
options limrow=0, limcol=0, solprint=off;
sets
  j estacoes /1*15/
  k classes /1*14/
  l no maximo de operacoes considerando todas as classes /1*9/
  p no de alternativas /1*3/
  iter iteracoes /1*30/;
alias(j,i); alias(k,k1);
parameters
Lk(k) taxa media de chegada externa da classe k
  /1 0.003571
  2 0.007617
  3 0.003891
  4 0.002113
  5 0.002292
  6 0.001932
  7 0.001857
  8 0.001490
  9 0.002800
 10 0.001079
 11 0.003455
 12 0.002862
 13 0.001433
 14 0.001355/
CAk(k) scv do intervalo de tempo entre chegadas externas da classe k
  /1 0.3818
  2 0.3377
  3 0.2569
  4 0.5092
  5 0.2853
  6 0.4048
  7 0.2939
  8 0.2934
  9 0.4131
 10 0.2817
 11 0.4351
 12 0.3607
 13 0.3037
 14 0.3528/
Nk(k) no de operacoes no roteiro da classe k

```

/1 5
 2 8
 3 9
 4 4
 5 8
 6 4
 7 7
 8 5
 9 8
 10 5
 11 5
 12 4
 13 9
 14 7/

*\$ontext;

STk(k) tempo medio de servico terceirizado

/1 0
 2 0
 3 50.1
 4 0
 5 0
 6 50.1
 7 50.1
 8 100.2
 9 0
 10 50.1
 11 0
 12 0
 13 50.1
 14 0/

table CustAcu(k,l) custo acumulado ateh a l-esima operacao da classe k

	1	2	3	4	5	6	7	8	9
1	570.03	2809.86	2970.36	3146.21	3197.85	0	0	0	0
2	5060.36	5351.72	6739.40	7206.94	10371.68	10698.49	10921.08	11199.27	0
3	1622.66	3857.01	5149.06	9358.59	9559.83	9664.24	9808.53	9808.53	9968.49
4	185.57	893.6	950.13	998.49	0	0	0	0	0
5	240.53	1127.35	1792.27	2409.66	2665.39	2825.03	3109.80	3169.23	0
6	4029.07	4155.76	4252.44	4308.74	0	0	0	0	0
7	1701.29	2976.97	3082.10	3339.21	6602.62	6760.25	6906.35	0	0
8	1502.80	1627.05	1691.97	1899.21	1967.50	0	0	0	0
9	1945.81	4342.51	5075.59	5628.66	5868.24	6174.11	6374.64	6454.31	0
10	3616.59	7768.23	7912.89	8141.44	8236.52	0	0	0	0
11	160.07	618.60	1379.49	1546.02	1586.09	0	0	0	0
12	2178.18	2754.27	2873.40	2932.29	0	0	0	0	0
13	391.88	1367.98	1813.60	2221.95	2463.68	2568.52	2652.28	2760.80	2796.02
14	146.86	1118.90	1496.31	3063.18	4826.99	5027.30	5178.76	0	0;

table Nkl(k,l) estacao visitada para l-esima operacao da classe k

	1	2	3	4	5	6	7	8	9
1	2	3	10	14	15	0	0	0	0
2	2	10	3	8	10	4	14	15	0
3	2	6	7	6	11	12	14	14	15
4	1	3	14	15	0	0	0	0	0
5	1	3	9	6	10	5	14	15	0
6	3	14	14	15	0	0	0	0	0
7	2	10	14	14	10	14	15	0	0
8	3	14	14	14	15	0	0	0	0
9	1	3	9	6	8	10	14	15	0

10 1 3 14 14 15 0 0 0 0
 11 1 3 9 14 15 0 0 0 0
 12 3 10 14 15 0 0 0 0 0
 13 1 3 6 7 6 11 14 14 15
 14 1 3 13 3 9 14 15 0 0;

table Skl(k,l) tempo de servico medio da l-esima operacao da classe k

	1	2	3	4	5	6	7	8	9
1	5.761	27.831	4.773	5.507	2.346	0	0	0	0
2	19.197	8.664	27.989	12.488	31.368	14.848	6.971	12.639	0
3	26.935	33.828	35.662	47.808	9.139	4.744	4.519	3.247	2.557
4	0.718	17.596	1.770	2.197	0	0	0	0	0
5	1.080	22.038	18.353	14.022	7.604	6.580	8.919	2.700	0
6	42.332	3.968	3.028	2.558	0	0	0	0	0
7	14.606	18.966	3.293	8.052	32.346	4.937	6.638	0	0
8	38.146	3.891	2.033	6.490	3.103	0	0	0	0
9	23.900	30.973	28.480	25.128	6.399	9.095	6.281	3.620	0
10	54.625	41.868	4.530	7.158	4.320	0	0	0	0
11	1.261	11.395	21.002	5.216	1.820	0	0	0	0
12	24.265	8.565	3.731	2.676	0	0	0	0	0
13	3.207	19.688	10.119	17.693	10.978	4.761	2.623	3.399	1.600
14	0.566	24.156	13.566	19.469	48.550	6.274	6.881	0	0;

table CSkl(k,l) scv do tempo de servico da l-esima operacao da classe k

	1	2	3	4	5	6	7	8	9
1	0.1361	0.1361	0.1361	0.1361	0.1361	0	0	0	0
2	0.4534	0.4534	0.4534	0.4534	0.4534	0.4534	0.4534	0.4534	0
3	0.1970	0.1970	0.1970	0.1970	0.1970	0.1970	0.1970	0.1970	0.1970
4	0.1186	0.1186	0.1186	0.1186	0	0	0	0	0
5	0.2544	0.2544	0.2544	0.2544	0.2544	0.2544	0.2544	0.2544	0
6	0.0092	0.0092	0.0092	0.0092	0	0	0	0	0
7	0.1366	0.1366	0.1366	0.1366	0.1366	0.1366	0.1366	0	0
8	0.5002	0.5002	0.5002	0.5002	0.5002	0	0	0	0
9	0.1280	0.1280	0.1280	0.1280	0.1280	0.1280	0.1280	0.1280	0
10	0.0741	0.0741	0.0741	0.0741	0.0741	0	0	0	0
11	0.4351	0.4351	0.4351	0.4351	0.4351	0	0	0	0
12	0.3047	0.3047	0.3047	0.3047	0	0	0	0	0
13	0.3242	0.3242	0.3242	0.3242	0.3242	0.3242	0.3242	0.3242	0.3242
14	0.1175	0.1175	0.1175	0.1175	0.1175	0.1175	0.1175	0	0;

Table Mjp(j,p) no de maquinas na estacao j na alternativa p

	1	2	3
1	1	1	1
2	1	1	1
3	1	1	1
4	1	1	1
5	1	1	1
6	1	1	1
7	1	1	1
8	1	1	1
9	1	1	1
10	1	1	1
11	1	1	1
12	1	1	1
13	1	1	1
14	1	1	1
15	1	1	1;

table MIjp(j,p) a taxa de servico na estacao j da alternativa p

	1	2	3
1	0	0.20800	0.10400

2 0 0.06650 0.04750
 3 0 0.04222 0.03800
 4 0 0.08375 0.05250
 5 0 0.22800 0.07600
 6 0 0.04200 0.02800
 7 0 0.03730 0.02667
 8 0 0.13800 0.04600
 9 0 0.05700 0.01900
 10 0 0.07560 0.05040
 11 0 0.25200 0.12600
 12 0 0.42200 0.21100
 13 0 0.14800 0.07400
 14 0 0.19900 0.19900
 15 0 0.20100 0.20100;

* A alternativa p=0 envolve investir nada em maquinas; p=1 envolve adicionar uma maquina;

* p=2 envolve remover uma maquina, quando a estacao tiver mais de uma maquina.

Table Fjp(j,p) custo de capacidade(em R\$1000) para estacao j na alternativa p

	1	2	3
1	0	9.0	0
2	0	22.0	-12.0
3	0	190.0	0.0
4	0	12.0	-7.0
5	0	6.0	-3.0
6	0	0.5	-0.3
7	0	40.0	-20.0
8	0	120.0	-60.0
9	0	130.0	-60.0
10	0	225.0	-110.0
11	0	2.0	0
12	0	2.0	0
13	0	0.5	0
14	0.0	0.0	0.0
15	0.0	0.0	0.0;

*estacoes 14 e 15 envolvem homens e nao maquinas. Por isso tomei 0 como referencia.

* A alternativa p=0 envolve investir nada em maquinas; p=1 envolve adicionar uma maquina;

* p=2 envolve remover uma maquina, quando a estacao tiver mais de uma maquina.

* MODELO DE AVALIACAO DE DESEMPENHO: CALCULO DE CAj

parameters

Qkl(k,l) proporcao de chegadas da classe (k l) na estacao Nkl

MAXAUXj(j) variavel usada na equacao partida(j) para substituir max(CSj(j) 0.2)

Lj(j) chegada total na estacao j

Sj(j) tempo medio de servico na estacao j

CSj(j) scv do tempo de servico na estacao j

Mj(j) no de maquinas na estacao j

MIj(j) taxa media de servico da estacao j

ROj(j) nivel de ocupacao da estacao j;

$L_j(j) = \sum(k, \sum(l, \sum(N_k(k) \geq \text{ord}(l)), L_k(k) \$(N_{kl}(k,l) = \text{ord}(j))))$;

$Q_{kl}(k,l) \$(N_k(k) \geq \text{ord}(l)) = L_k(k) / \sum(j \$(N_{kl}(k,l) = \text{ord}(j)), L_j(j))$;

$S_j(j) = (\sum(k, \sum(l, \sum(N_k(k) \geq \text{ord}(l)), Skl(k,l) * L_k(k) \$(N_{kl}(k,l) = \text{ord}(j)))) /$

$L_j(j)$;

$CS_j(j) = (\sum(k, \sum(l, \sum(N_k(k) \geq \text{ord}(l)), ((CSkl(k,l) + 1) * L_k(k) * (Skl(k,l) ** 2)) \$(N_{kl}(k,l) = \text{ord}(j)))) / (L_j(j) * (S_j(j) ** 2)) - 1$;

MAXAUXj(j)\$(CSj(j)<0.2)=0.2;

MAXAUXj(j)\$(CSj(j)>=0.2)=CSj(j);

variables

Z funcao objetivo ficticia

```

CAj(j) scv do intervalo de tempo entre chegadas na estacao j
CDj(j) scv do intervalo de tempo entre partidas da estacao j
CDkl(k,l) scv do int de tempo entre partidas da estacao j para a i (=CAji);
positive variables CAj, CDj, CDkl;
equations
  objfun1 funcao objetivo ficticia
  superpos(j) equacao de superposicao das chegadas na estacao j
  partida(j) equacao de partida agregada da estacao j
  desagreg(k,l) equacao de desagregacao da partida da estacao j;
  objfun1.. Z =e= sum(j, Mj(j));
superpos(j)..
* metodo estacionario (dependencia dos intervalos de tempo entre chegadas)
  CAj(j)=e=sum(k,sum(l$(Nk(k)>=ord(l)),(Lk(k)/Lj(j))*(CAk(k)$ord(l)=1)+
  CDkl(k,l-1)$ord(l)>1)$Nkl(k,l)=ord(j)));
partida(j)..
*Whitt (1983a)
  CDj(j)=e=1+(1-ROj(j)**2)*(CAj(j)-1)+ROj(j)**2*(MAXAUXj(j)-1)/sqrt(Mj(j));
desagreg(k,l)$Nk(k)>=ord(l)..
*Aproximacao Trafego leve: Bitran e Tirupati (1988)
  CDkl(k,l)=e=Qkl(k,l)*(sum(j$(Nkl(k,l)=ord(j)),CDj(j)))+(1-Qkl(k,l))+
  (1-Qkl(k,l))**2*(CDkl(k,l-1)$ord(l)>1)+CAk(k)$ord(l)=1);
model passo1 / objfun1, superpos, partida, desagreg/;
*****
* MODELO DE OTIMIZACAO: Yjp
scalars
  WIPmax limitante superior para o WIP na rede /22000/;
parameters
  Vj(j) valor monetario associado a um lote de uma classe qualquer na estacao j
  ELjp(j,p) numero medio de lotes na estacao j com alternativa p - nao considera terceirizados;
  Vj(j)=sum(k,sum(l$(Nk(k)>=ord(l)),CustoAcu(k,l)*Lk(k)$Nkl(k,l)=ord(j)))/
  sum(k, sum(l$(Nk(k)>=ord(l)),Lk(k)$Nkl(k,l)=ord(j)));
variables
  CUSTOT custo otimo total
  Yjp variavel binaria: =1 se alternativa p eh escolhida =0 caso contrario;
binary variables Yjp;
*Yjp eh inteira e binaria
* Yjp.LO(j,p)=0;
* Yjp.UP(j,p)=1;
equations
  objfun2 funcao objetivo
  WIPrest restricao para o limitante superior para o WIP
  alt(j) uma e somente uma alternativa eh escolhida para cada estacao j;
objfun2..
  CUSTOT =e= sum(j,sum(p, Fjp(j,p)*Yjp(j,p)));
WIPrest..
  sum(j,sum(p, Vj(j)*ELjp(j,p)*Yjp(j,p)))=l= WIPmax;
alt(j)..
  sum(p, Yjp(j,p))=e=1;
model convexo /objfun2, WIPrest, alt/;
*$ontext;
*****
* PROGRAMA PRINCIPAL: SOLUCAO ITERATIVA (CAj VARIA) DO MODELO DE
OTIMIZACAO
scalars
  TOL nivel de tolerancia do erro em CAj de uma iteracao para outra /0.001/;
parameters
  CA1j(j) scv dos intervalos de tempo entre chegadas OTIMO na estacao j

```

```

ROjp(j,p) intensidade de trafego na estacao j para alternativa p
GAUXjp(j,p) funcao g da formula de Kraemer para estacao j e alternativa p
WIPjp(j,p) WIP otimo na estacao j e alternativa p;
scalars
WIPtot WIP total otimo
WIPtin WIP total inicial
custotin custo total inicial de capacidade
prec precisao do scv da solucao atual;
*Taxa de servico, no maquinas e intensidade de trafego da configuracao inicial
MIjp(j,"1")=1/Sj(j);
ROjp(j,"1")=Lj(j)/(MIjp(j,"1")*Mjp(j,"1"));
MIj(j)=MIjp(j,"1");
ROj(j)=ROjp(j,"1");
Mj(j)=Mjp(j,"1");
*CAj inicial
solve passo1 minimizing Z using LP;
display CAj.l, CDj.l, CDkl.l;
*
prec=100; # precisao inicial para forcar iteracoes pois TOL = 0.001
CA1j(j)=CAj.l(j);
*****
*inicio das iteracoes
loop(iter$(prec >= TOL),
ROjp(j,p)=Lj(j)/(MIjp(j,p)*Mjp(j,p));
GAUXjp(j,p)$(CA1j(j)>=1)=1;
GAUXjp(j,p)$(CA1j(j)<1)=exp((-2*(1-ROjp(j,p))*(1-CA1j(j))**2)/
(3*ROjp(j,p)*(CA1j(j)+CSj(j))));
ELjp(j,p)=(ROjp(j,p)**2*(CA1j(j)+CSj(j))*GAUXjp(j,p)/
(2*(1-ROjp(j,p))+ROjp(j,p));#Kraemer para GI/G/m
display GAUXjp, MIjp, ROjp, ELjp;
*Custo de capacidade (custotin) e de WIP (WIPtin) iniciais
WIPjp(j,"1")$(ord(iter)=1)= Vj(j)*ELjp(j,"1");
WIPtin$(ord(iter)=1)= sum(j,WIPjp(j,"1"));
custotin$(ord(iter)=1)= sum(j,Fjp(j,"1"));
display$(ord(iter)=1) WIPtin, custotin;
*Escolha das alternativas mais viaveis (Yjp)
solve convexo minimizing CUSTOT using MIP;
WIPjp(j,p)=Vj(j)*ELjp(j,p);
WIPtot=sum(j,sum(p$(Yjp.l(j,p)=1),WIPjp(j,p)));
display WIPjp, Yjp.l, CUSTOT.l, WIPtot;
*Atualizacao MIj, Mj e ROj
MIj(j)=sum(p$(Yjp.l(j,p)=1), MIjp(j,p));
Mj(j)=sum(p$(Yjp.l(j,p)=1), Mjp(j,p));
ROj(j)=Lj(j)/(MIj(j)*Mj(j));
*Atualizacao CAj
solve passo1 minimizing Z using LP;
display CAj.l, CDj.l, CDkl.l, MIj, ROj;
prec=sum(j,abs(CAj.l(j)-CA1j(j)))/card(j);
display prec;
display$(prec<=TOL) Yjp.l, CUSTOT.l, WIPtot, WIPtin, custotin, ROj, ELjp, WIPjp;
CA1j(j)=CAj.l(j); #fim das iteracoes
abort$(prec >= TOL) "precisao violada"
*$ontext;
*****
*AVALIACAO DE DESEMPENHO DA CLASSE AGREGADA NAS ESTACOES PARA SOLUCAO
OTIMA

```

*NUMERO MEDIO DE JOBS EM CADA ESTACAO J: MODELO DE KRAEMER E LAGENBACH-BELZ

*VARIANCIA DO NUMERO DE JOBS EM CADA ESTACAO J: $(V(L_j) = (E(L_j))^2) * CL_j$

*

parameters

GAUXj(j) funcao auxiliar para calculo de EWqj

EWqj(j) o tempo medio de espera em fila na estacao j otimo

EWqjp(j,p);

GAUXjp(j,p) $(CA1j(j) \geq 1) = 1;$

GAUXjp(j,p) $(CA1j(j) < 1) = \exp((-2 * (1 - ROjp(j,p)) * (1 - CA1j(j)) ** 2) / (3 * ROjp(j,p) * (CA1j(j) + CSj(j))));$

GAUXj(j) = sum(p\$(Yjp.l(j,p) = 1), GAUXjp(j,p));

EWqjp(j,p) = $(ROjp(j,p) * (CA1j(j) + CSj(j)) * GAUXjp(j,p)) / (2 * MIjp(j,p) * (1 - ROjp(j,p)));$

EWqj(j) = $(ROj(j) * (CA1j(j) + CSj(j)) * GAUXj(j)) / (2 * MIj(j) * (1 - ROj(j)));$ #Kraemer

display EWqjp, EWqj;

* PASSO 4: AVALIACAO DE DESEMPENHO PARA CADA CLASSE PARA A SOLUCAO OTIMA

* LEADTIME MEDIO NA REDE, NUMERO MEDIO DE JOBS DA CLASSE K NA REDE

* VARIANCIA DO LEADTIME DA CLASSE K NA REDE

scalar

ELrede numero de lotes na rede obtido pelas estacoes (desconsidera os terceirizados)

ETrede leadtime medio da rede considerando os tempos de servico terceirizados

Lote numero total de lotes na rede obtido pelas classes (considera servico terceirizado);

parameter

ETk(k) leadtime medio da classe k na solucao otima (pode considerar servico terceirizado)

ETkp(k,p) leadtime medio da classe k na alternativa p

Lotek(k) numero de lotes na rede obtido a partir da classe k - considera terceirizados;

table Sklp(k,l,p) tempo de servico medio da l-esima operacao da classe k na alternativa p

	1.1	1.2	1.3	2.1	2.2	2.3	3.1	3.2	3.3
1	5.761	4.938	6.913	27.831	25.048	27.831	4.773	3.978	5.966
2	19.197	16.455	23.036	8.664	7.220	10.830	27.989	25.190	27.989
3	26.935	23.087	32.322	33.828	28.190	42.285	35.662	30.567	42.794
4	0.718	0.359	0.718	17.596	15.836	17.596	1.770	1.770	1.770
5	1.080	0.540	1.080	22.038	19.834	22.038	18.353	12.235	36.706
6	42.332	38.099	42.332	3.968	3.968	3.968	3.028	3.028	3.028
7	14.606	12.519	17.527	18.966	15.805	23.708	3.293	3.293	3.293
8	38.146	34.331	38.146	3.891	3.891	3.891	2.033	2.033	2.033
9	23.900	11.950	23.900	30.973	27.876	30.973	28.480	19.987	56.960
10	54.625	27.313	54.625	41.868	37.681	41.868	4.530	4.530	4.530
11	1.261	0.631	1.261	11.395	10.256	11.395	21.002	14.001	42.004
12	24.265	21.839	24.265	8.565	7.138	10.706	3.731	3.731	3.731
13	3.207	1.604	3.207	19.688	16.407	19.688	10.119	8.433	12.649
14	0.566	0.283	0.566	24.156	21.740	24.156	13.566	6.783	13.566

+	4.1	4.2	4.3	5.1	5.2	5.3	6.1	6.2	6.3
1	5.507	5.507	5.507	2.346	2.346	2.346	0	0	0
2	12.488	8.325	24.976	31.368	26.140	39.210	14.848	11.878	19.797
3	47.808	39.840	59.760	9.139	4.570	9.139	4.744	2.372	4.744
4	2.197	2.197	2.197	0	0	0	0	0	0
5	14.022	11.685	17.528	7.604	6.337	9.505	6.580	4.387	13.160
6	2.558	2.558	2.558	0	0	0	0	0	0
7	8.052	8.052	8.052	32.346	26.955	40.433	4.937	4.937	4.937
8	6.490	6.490	6.490	3.103	3.103	3.103	0	0	0
9	25.128	20.940	31.410	6.399	4.266	12.798	9.095	7.579	11.369
10	7.158	7.158	7.158	4.320	4.320	4.320	0	0	0
11	5.216	5.216	5.216	1.820	1.820	1.820	0	0	0
12	2.676	2.676	2.676	0	0	0	0	0	0

13	17.693	15.165	21.232	10.978	9.148	13.723	4.761	2.381	4.761
14	19.469	17.522	19.469	48.550	32.367	97.100	6.274	6.274	6.274

	7.1	7.2	7.3	8.1	8.2	8.3	9.1	9.2	9.3
1	0	0	0	0	0	0	0	0	0
2	6.971	6.971	6.971	12.639	12.639	12.639	0	0	0
3	4.519	4.519	4.519	3.247	3.247	3.247	2.557	2.557	2.557
4	0	0	0	0	0	0	0	0	0
5	8.919	8.919	8.919	2.700	2.700	2.700	0	0	0
6	0	0	0	0	0	0	0	0	0
7	6.638	6.638	6.638	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0
9	6.281	6.281	6.281	3.620	3.620	3.620	0	0	0
10	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0
13	2.623	2.623	2.623	3.399	3.399	3.399	1.600	1.600	1.600
14	6.881	6.881	6.881	0	0	0	0	0	0;

$ETk(k) = \sum(l, \$Nk(k) \geq ord(l), \sum(j, \sum(p, ((EWqjp(j,p) + Sklp(k,l,p)) * Yjp.l(j,p)) \$Nkl(k,l) = ord(j)))) + S$
 $Tk(k)$; # no leadtime foi considerado o tempo em servico terceirizado

$ETkp(k,p) = \sum(l, \$Nk(k) \geq ord(l), \sum(j, ((EWqjp(j,p) + Sklp(k,l,p)) \$Nkl(k,l) = ord(j)))) + STk(k)$; # no
leadtime foi considerado o tempo em servico terceirizado

$Lote = \sum(k, Lk(k) * ETk(k))$; # aqui o no. de lotes na rede considera o no. de lotes em servico terceirizado

$Lotek(k) = Lk(k) * ETk(k)$;

$ELrede = \sum(j, \sum(p, ELjp(j,p) * Yjp.l(j,p)))$;

$ETrede = Lote / \sum(k, Lk(k))$;

display ETk , $ETkp$, $Lote$, $Lotek$, $ELrede$, $ETrede$, $WIPmax$, $WIPtot$, $CUSTOT.l$;

