

ABORDAGEM MULTIOBJETIVO PARA OTIMIZAÇÃO DE REDES DE FILAS FINITAS

Multi-objective Approach for Finite Queuing Network Optimization

**Frederico Rodrigues Borges da Cruz¹, Fernando Luiz Pereira de Oliveira²,
José Hissa Ferreira³, Oriane Magela Neto^{4†}**

Resumo: Este artigo apresenta uma discussão sobre os resultados obtidos por meio de um algoritmo multiobjetivo recentemente desenvolvido para otimizar, simultaneamente, o tamanho total das áreas de espera, a taxa total de serviço e a taxa de saída de uma rede de filas finitas com serviço geral. Como tais objetivos eram conflitantes, utilizou-se uma versão de um algoritmo genético projetado para encontrar soluções ótimas a mais de um deles. Foram obtidas algumas propriedades que podem auxiliar na análise e no projeto de tais importantes sistemas estocásticos.

Palavras-chave: Redes de filas. Algoritmos genéticos. Otimização multiobjetivo.

Abstract: This paper presents a discussion about results obtained from a multi-objective algorithm that was recently developed to simultaneously optimize the total number of buffers, the overall service rate, and the throughput of a general-service finite queuing network. These conflicting objectives were optimized by means of a multi-objective genetic algorithm, designed to produce solutions for more than one of them. Some properties were identified that may help the analysis and design of these important stochastic systems.

Keywords: Network of queues. Genetic algorithms. Multiobjective optimization.

1. INTRODUÇÃO

Desde o trabalho de Hillier (1963), um dos pioneiros na área de filas, sistemas de máquinas (ou centros de trabalho) separados por áreas de espera finitas e organizados em redes têm sido o foco de interesse de muitos pesquisadores, principalmente pelo seu potencial como modelo capaz de descrever diversas situações do mundo real, tais como sistemas: baseados na *web* (CHAUDHURI et al., 2007), no projeto e no planejamento de *job-shops* (BITRAN; MORÁBITO, 2009); do tráfego de pessoas em ambientes exígios (CRUZ

et al., 2010); de comunicação (GONTIJO et al., 2011); de filas de prioridade (HE et al., 2012) e redes de filas finitas (CRUZ et al., 2012).

Enfocam-se neste trabalho as filas de um único servidor com tempos entre chegadas exponencialmente distribuídos e durações de serviço com distribuição geral, configuradas em redes em uma topologia acíclica arbitrária. Mais especificamente, o foco está nas redes de filas $M/G/1/k$, as quais, na notação de Kendall, possuem chegadas Markovianas, tempos de serviço gerais, um único servidor e uma capacidade total de k itens, incluindo um item em serviço (ou seja, área de espera

1. Professor Associado, Doutor, Departamento de Estatística, Universidade Federal de Minas Gerais – Belo Horizonte, MG – Brasil. E-mail: fcruz@est.ufmg.br

2. Professor Adjunto, Doutor, Departamento de Estatística, Universidade Federal de Ouro Preto – Ouro Preto, MG – Brasil. E-mail: fernandoluz@iceb.ufop.br

3. Professor Associado, Doutor, Departamento de Engenharia Elétrica, Centro Federal de Educação Tecnológica de Minas Gerais – Belo Horizonte, MG – Brasil. E-mail: jhissa@des.cefetmg.br

4. Professor Associado, Ph.D, Departamento de Engenharia Elétrica, Universidade Federal de Minas Gerais – Belo Horizonte, MG – Brasil (*in memoriam*).

com capacidade para $b=k-1$ item). Um exemplo deste tipo de rede é apresentado na Figura 1. O objetivo disso foi alcançar, simultaneamente, uma taxa de saída máxima (θ), utilizando-se a menor capacidade possível ($K=[k_1, k_2, \dots, k_n]^T$), ou a área de espera possível inferior ($B=[b_1, b_2, \dots, b_n]^T$), em que $b_i=k_i-1$, e com as menores taxas de serviço possíveis ($\mu=[\mu_1, \mu_2, \dots, \mu_n]^T$), após serem conhecidas a topologia da rede e as taxas de chegada externas ($\lambda=[\lambda_1, \lambda_2, \dots, \lambda_n]^T$).

Há um compromisso entre a capacidade total, as taxas de serviço e a taxa de saída resultante. Devido ao elevado custo representado pelas áreas de espera e pelos serviços, o tamanho global da área de espera e a capacidade total de serviço alocado devem ser restritos. Do ponto de vista do usuário da rede, por outro lado, deve-se alcançar a maior taxa de saída possível, que é diretamente afetada pelas áreas de espera e pelas taxas de serviço, tornando tais objetivos conflitantes. De fato, se a área de espera e a taxa de serviço são reduzidos, é possível que ocorra uma redução indesejável na taxa de saída,

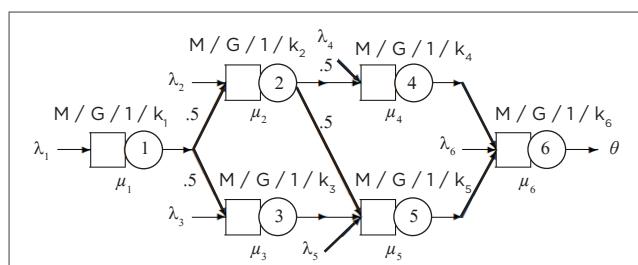


Figura 1. Uma rede de filas $M/G/1/K$ (adaptada de Brito et al., 2012).

como pode ser observado na Figura 2. Esta ilustra uma taxa de saída, θ , para uma fila única $M/G/1/k$ com $c\sigma^2=1,5$ (quadrado do coeficiente de variação do tempo de serviço) e $\lambda=5$ usuários por unidade de tempo (taxa de chegada externa), representada como uma função da capacidade total, k , e da taxa de serviço, μ . As respectivas curvas de nível também são mostradas em tal figura.

Comportamento semelhante também será observado para a taxa de saída das filas configuradas em redes. A suavidade da superfície do gráfico mostrado na Figura 2 parece sugerir uma função côncava. Outros resultados similares foram relatados para as redes de filas simples (MEESTER; SHANTHIKUMAR, 1990). No entanto, a característica plana na parte superior da superfície se apresenta como um problema para os métodos tradicionais de otimização. Em Smith e Cruz (2005), foi proposto um algoritmo de otimização que combina, com sucesso, o método clássico de Powell com múltiplos pontos iniciais, evitando-se uma convergência prematura para soluções ótimas locais.

Neste artigo, apresenta-se uma abordagem multiobjetivo capaz de otimizar, simultaneamente, o tamanho total das áreas de espera, a taxa global de serviço alocada e a taxa de saída para as redes de filas $M/G/1/k$. O método sugerido produz um conjunto de soluções eficientes, denominadas Pareto ótimo, para mais de um objetivo (CHANKONG; HAIMES, 2008). Com tal abordagem, o decisor tem condições de avaliar o resultado final devido à cada solução escolhida. Além disso, a abordagem multiobjetivo também permite ao usuário avaliar

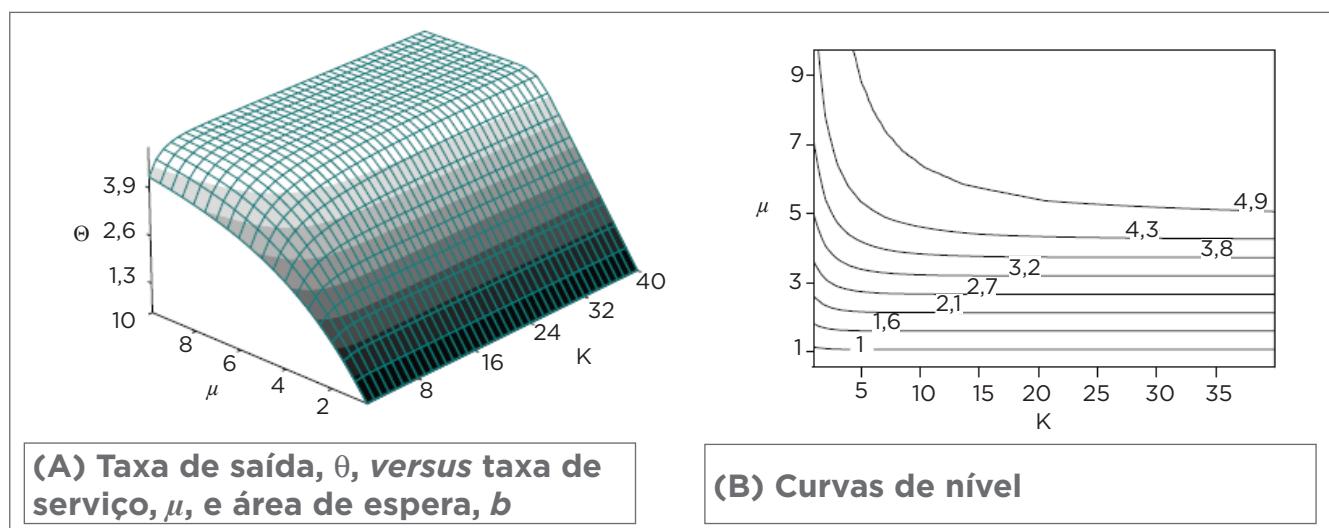


Figura 2. Resultado para uma única fila $M/G/1/k$, para $\lambda=5,0$ (Cruz et al., 2012).

os efeitos do aumento de um dos objetivos (por exemplo, a taxa de saída), enquanto reduz simultaneamente os outros (por exemplo, as áreas de espera e as taxas de serviço alocadas).

Neste artigo, um algoritmo multiobjetivo evolucionário (em inglês, MOEA) foi usado em combinação com o método de expansão generalizado (em inglês, GEM), que é uma ferramenta bem conhecida e eficaz na obtenção de aproximações precisas para o desempenho das filas de espera configuradas em redes (KERBACHE; SMITH, 1987). Os MOEAs são particularmente adequados para problemas com vários objetivos simultâneos, devido ao bom desempenho demonstrado em problemas multiobjetivos semelhantes de otimização em redes, como visto em Carrano et al. (2006).

Contribuições deste artigo incluem a confirmação de que o efeito do cv^2 do tempo de serviço pode ser significativo nas alocações, evidenciando a importância do desenvolvimento de uma metodologia baseada em tempos de serviço gerais, isto é, filas $M/G/1/k$; portanto, com cv^2 arbitrário, em oposição às $M/M/1/k$ (que têm $cv^2=1$). Outra importante contribuição é a demonstração da utilidade das curvas de nível aproximadas, na identificação dos pontos críticos de *tradeoff* entre a taxa de saída (correspondente à receita gerada pelo sistema) e as alocações de áreas de espera e de taxas de serviço (correspondentes aos insumos necessários para funcionamento do sistema neste patamar de atendimento). Finalmente, apresenta-se uma análise em detalhe de soluções Pareto eficientes, de modo a possibilitar a identificação de pontos críticos de *tradeoff* e de oportunidades de projetos eficientes.

Este artigo está organizado da seguinte forma. A Seção 2 trata da apresentação de um algoritmo evolutivo multiobjetivo especificamente desenvolvido, juntamente com o método de expansão generalizado, que é uma ferramenta de avaliação de desempenho utilizada para estimar a taxa de saída, θ . Na Seção 3, são abordados os resultados dos experimentos computacionais realizados com o algoritmo. Finalmente, a Seção 4 conclui o trabalho com considerações finais e sugestões para futuras pesquisas na área.

2. ALGORITMOS

2.1 FORMULAÇÃO DO PROBLEMA

O problema da otimização de redes de filas finitas que está sendo tratado é definido por uma formulação multiobjetivo,

na qual os custos da área de espera total e a taxa global de serviço são minimizados, enquanto que a taxa de saída é maximizada. A restrição é que a atribuição das áreas de espera seja um número inteiro e as taxas de serviço, um real não negativo. Definindo-se uma rede de filas como um dígrafo $G(N, A)$, em que N é um conjunto finito de nós (filas) e A é um de arcos (par de filas conectadas), uma formulação possível pode ser observada nas Equações 1 a 3:

$$\text{minimize } \mathbf{F}(\mathbf{K}, \boldsymbol{\mu}), \quad (1)$$

sujeito a:

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

$$\mu_i \geq 0, \forall i \in N, \quad (3)$$

em que as variáveis de decisão k_i e μ_i indicam a capacidade total e a taxa de saída para a i -ésima fila $M/G/1/k$, respectivamente. As funções objetivo,

$$\mathbf{F}(\mathbf{K}, \boldsymbol{\mu}) = [f_1(\mathbf{K}), f_2(\boldsymbol{\mu}), -f_3(\mathbf{K}, \boldsymbol{\mu})],$$

representam a capacidade total alocada,

$$f_1(\mathbf{K}) = \sum_{i \in N} k_i,$$

a taxa global de serviço,

$$f_2(\boldsymbol{\mu}) = \sum_{i \in N} \mu_i,$$

e a taxa de saída,

$$f_3(\mathbf{K}, \boldsymbol{\mu}) = \theta(\mathbf{K}, \boldsymbol{\mu}),$$

respectivamente.

É importante ressaltar que, na literatura, a taxa de saída é frequentemente modelada como uma restrição que deve ser maior do que um valor limiar, ao invés de ser um objetivo que deve ser maximizado, conforme considerado neste artigo. O problema é que, para se resolver a versão mono-objetivo do problema, a restrição da taxa de saída deve ser relaxada, e um valor arbitrário deve ser estabelecido para o limiar, o que não é uma tarefa trivial. Além disso, muitas vezes, um pequeno

decréscimo em tal quantia produz uma redução significativa na atribuição da capacidade total e das taxas de serviço. O compromisso entre a taxa de saída, a capacidade total e as taxa de serviço não ficará visível em uma formulação mono-objetivo equivalente, que geralmente combina os objetivos múltiplos e um único por meio de um vetor de pesos, ω . Ainda, a determinação do vetor de pesos, ω , é difícil e, frequentemente, conduz a formulações mono-objetivo arbitrárias.

Os algoritmos serão apresentados em duas partes. Inicialmente, o de avaliação de desempenho será descrito. Em seguida, será detalhado aquele proposto para otimizar o problema.

2.2 AVALIAÇÃO DE DESEMPENHOS EM FILAS FINITAS INDIVIDUAIS

Nas filas únicas (não exatamente o caso de interesse aqui), a taxa de saída, $\theta(k, \mu)$, é dada pela Equação 4:

$$\theta(k, \mu) = \lambda(1 - p_k), \quad (4)$$

em que λ é a taxa de chegada externa e p_k é a probabilidade de bloqueio, que é a chance de encontrar o sistema cheio (isto é, com um número de itens igual à capacidade total, k). Desse modo, a resolução do problema de determinar $\theta(k, \mu)$ passa por aquela do valor p_k .

Para o caso especial de sistemas Markovianos puros, ou seja, filas $M/M/1/k$, a expressão da probabilidade de bloqueio pode ser facilmente deduzida, conforme mostrado na literatura da teoria de filas – Equação 5 (GROSS et al., 2009):

$$p_k = \frac{(1-\rho)\rho^k}{1-\rho^{k+1}} \quad (5)$$

válida para $\rho < 1$, em que $\rho \equiv \lambda/\mu$ é a utilização do sistema. Relaxando-se a restrição de integralidade de k é possível expressar, em forma fechada, a alocação ótima de capacidade total para filas $M/M/1/k$, em termos de ρ e p_k , como visto na Equação 6:

$$k_M = \left\lceil \frac{\ln\left(\frac{p_k}{1-\rho + p_k\rho}\right)}{\ln(\rho)} \right\rceil \quad (6)$$

em que $[x]$ é o menor inteiro não inferior a x . Consequentemente, é possível mostrar que a alocação ótima da área de espera

(excluídos o item em serviço) para filas $M/M/1/k$ é vista na Equação 7:

$$b_M = k_M - 1. \quad (7)$$

Para filas gerais $M/G/c/k$, a probabilidade de bloqueio pode ser determinada apenas por técnicas de aproximação. Em particular, Smith e Cruz (2005) usaram uma aproximação por dois momentos, baseada na expressão Markoviana, Equação 7, que é bem efetiva (Equação 8):

$$b_e(cv^2) = b_M + \text{INT} \left[\frac{(cv^2 - 1)\sqrt{\rho}}{2} b_M \right] \quad (8)$$

em que $\text{INT}[x]$ é a parte inteira de x . Em particular, para filas gerais com servidor único, $M/G/1/k$, conhecido o ρ e cv^2 , a alocação ótima de áreas de espera pode ser escrita como a Equação 9:

$$b_e(cv^2) = \frac{\ln\left(\frac{p_k}{1-\rho + p_k\rho}\right) + \ln(\rho)(2 + \sqrt{\rho}cv^2 - \sqrt{\rho})}{2\ln(\rho)} \quad (9)$$

Finalmente, pode-se isolar p_k a fim de se estabelecer uma expressão fechada para a probabilidade de bloqueio em filas $M/G/1/k$, em função de k (note que, para filas $M/G/1/k$, $k=1+b_e$), como observado na Equação 10:

$$p_k = \frac{(1-\rho)\rho^{\left(\frac{2+\sqrt{\rho}cv^2-\sqrt{\rho}+2(k-1)}{2+\sqrt{\rho}cv^2-\sqrt{\rho}}\right)}}{1-\rho^{\left(\frac{2+\sqrt{\rho}cv^2-\sqrt{\rho}+(k-1)}{2+\sqrt{\rho}cv^2-\sqrt{\rho}}\right)}}. \quad (10)$$

Como última observação, vale ressaltar a similaridade existente entre as expressões para p_k para filas $M/M/1/k$ e $M/G/1/k$, respectivamente, nas Equações 5 e 10.

2.3 AVALIAÇÃO DE DESEMPENHOS EM FILAS FINITAS CONFIGURADAS EM REDE

Para as redes de filas, a estimativa da taxa de saída é realizada por meio do GEM, que é utilizado com sucesso para estimar o desempenho das redes acíclicas, de filas finitas, arbitrariamente configuradas (KERBACHE; SMITH, 1987). O GEM é uma combinação de uma decomposição nó a nó e de tentativas repetidas, em que cada fila é analisada

separadamente, e são realizadas modificações que representam os efeitos entre as filas da rede.

Como descrito detalhadamente por Kerbache e Smith (1987), o GEM cria, para cada fila finita j , uma auxiliar, h_j , que é modelada como uma $M/G/\infty$, conforme ilustrado na Figura 3. Cada entidade que se dirige à j pode ser bloqueada (com probabilidade p_{kj}), ou não (com probabilidade $1-p_{kj}$). Quando isso ocorre, ela é encaminhada à fila h_j , onde espera até que um espaço seja liberado na j . Assim, o papel de h_j é registrar o tempo que uma entidade tem de esperar antes de entrar na fila j . O objetivo final do GEM é atualizar as taxas de serviço de cada uma das filas i que são seguidas por finitas, como apresentado na Equação 11:

$$\tilde{\mu}_i^{-1} = \mu_i^{-1} + p_{kj} (\mu'_h)^{-1} \quad (11)$$

e, assim, possibilitar a obtenção de uma aproximação acurada para a taxa de saída da rede θ_j .

O processo de avaliação de desempenho deve ser conduzido em uma ordem específica, e análise da rede em estudo, definida como dígrafo $G(N, A)$, é apresentada no Algoritmo 1. O algoritmo calcula os bloqueios nos nós de serviços a montante, resultando em taxas de serviço corrigidas que são reduzidas de acordo com a Equação 11. Note-se que o algoritmo de avaliação de desempenho é uma variante daquele de Dijkstra para a determinação de caminhos mínimos. Por exemplo, na rede ilustrada na Figura 1, uma sequência de avaliação válida é $1 \rightarrow 2 \rightarrow 4 \rightarrow 3 \rightarrow 5 \rightarrow 6$, uma vez que, no algoritmo de Dijkstra, um nó j pode apenas ser avaliado se todos os seus predecessores i já tiverem sido avaliados. Especificamente, a sequência deve certificar que um nó somente será acessado após todos os seus antecessores. Assumindo que os circuitos

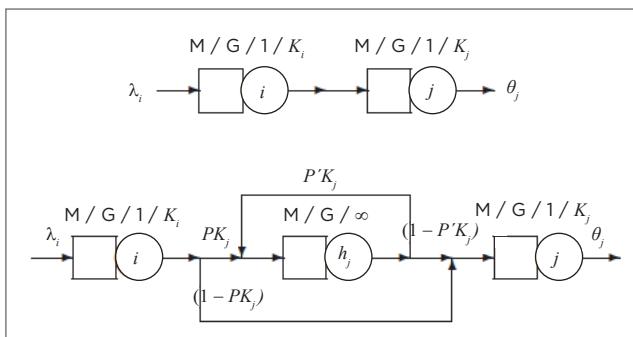


Figura 3. Método de expansão generalizado.

não estão presentes em $G(N, A)$, o GEM tem uma complexidade de tempo de execução de $O(N^2)$, que está de acordo com o algoritmo de Dijkstra.

2.4 ALGORITMO DE OTIMIZAÇÃO

Para o problema de otimização considerado neste artigo, definido pelas Equações 1 a 3, um algoritmo genético (AG) multiobjetivo parece ser uma escolha adequada. Os AG multiobjetivos pertencem à classe dos MOEA e são algoritmos de otimização que realizam uma busca global aproximada, baseada em informações obtidas a partir da avaliação de várias soluções do espaço de busca (DEB, 2001). As soluções candidatas, que convergem para um valor ótimo, são obtidas pela aplicação dos operadores genéticos de mutação, cruzamento, seleção e elitismo.

Cada um destes operadores define um tipo de AG, que pode ser implementado de várias maneiras. Além disso, a convergência do AG é garantida pela atribuição de um valor de aptidão para cada solução candidata, preservando-se a diversidade. De fato, aplicações recentes, bem-sucedidas de AGs, foram relatadas em estudos mono- (LIN, 2008) e multiobjetivos (CARRANO et al., 2006). O AG utilizado neste estudo se baseia no AG com elitismo e ordenação não dominada (NSGA-II), desenvolvido por Deb et al. (2002), mostrado no Algoritmo 2. Na aplicação de AGs para otimização

algoritmo

```

leia grafo,  $G(N, A)$ 
leia probabilidades de roteamento,  $r_{[ij]}$ ,  $\forall (i, j) \in A$ 
leia taxas de chegada externas e de serviço,  $\Lambda_i$ ,  $\mu_i$ ,  $\forall i \in N$ 
inicialize conjunto de nós rotulados,  $P \leftarrow \emptyset$ 
enquanto  $P \neq V$ 
    escolha  $j$  tal que  $(j \in N) \text{ e } (j \notin P)$ 
    se  $\{i \mid (i, j) \in A\} \subseteq P$  então
        /* calcule medidas de desempenho */
        calcule  $P_{kj}$ ,  $\theta_j$ 
        /* repasse informação aos nós sucessores */
        para  $\forall k \in \{k' \mid (j, k') \in A\}$  então
             $\lambda_k \leftarrow \lambda_k + \theta_j \times r_{[jk]}$ 
        fim para
        /* rotule nó como pré avaliado */
         $P \leftarrow P \cup \{j\}$ 
    fim se
    fim enquanto
fim algoritmo

```

Algoritmo 1. Algoritmo de rotulação para a análise de desempenho.

algoritmo

```

leia grafo, taxas de chegada e de serviço,  $G(N, A)$ ,  $\Lambda_i \forall i \in N$ 
 $P_1 \leftarrow \text{GeraPopulaçãoInicial}(\text{popSize})$ 
para  $i = 1$  até numGen faça
    /* gera novas soluções por cruzamento e mutação */
     $Q_i \leftarrow \text{FaçaPopulaçãoNova}(P_i)$ 
    /* combina soluções */
     $R_i \leftarrow P_i \cup Q_i$ 
    /* encontre fronteiras não-dominadas  $F = (F_1, F_2, \dots)$  */
     $F \leftarrow \text{OrdenaçãoNãoDominante}(R_i)$ 
    /* encontre nova população pela distância de aglomeração*/
     $P_{i+1} \leftarrow \text{GeraPopulaçãoNova}(R_i)$ 
fim para
 $P_{\text{numGen}+1} \leftarrow \text{ExtraiPareto}(P_{\text{numGen}})$ 
escreva  $P_{\text{numGen}+1}$ 
fim algoritmo

```

Algoritmo 2. Algoritmo genético multiobjetivo com elitismo NSGA-II (Deb et al., 2002).

multiobjetivo, os operadores de seleção e elitismo devem ser especificamente ajustados para identificar corretamente as condições ótimas, como será mostrado a seguir.

Elitismo se baseia no conceito de dominância. Um ponto $x_i = (x_{i1}, x_{i2}, \dots, x_{in})$ domina um $x_j = (x_{j1}, x_{j2}, \dots, x_{jn})$, caso x_i seja melhor que x_j em um dos objetivos k ($p.e., f_k(x_i) < f_k(x_j)$, para minimização, e não seja pior em nenhum outro objetivo l ($p.e., f_l(x_i) \neq f_l(x_j)$). Para incluir o operador de elitismo, utilizou-se um procedimento conhecido como algoritmo rápido de ordenação não dominante (DEB et al., 2002), que isola os indivíduos na população em várias camadas (fronteiras) F_i , de modo que suas soluções sejam não dominadas e toda solução em uma $F_i, i > 1$ seja dominada por pelo menos uma solução na camada F_{i-1} e por nenhuma em $F_j, j \geq i$. Como mostrado em Deb et al. (2002), isto pode ser realizado com complexidade de tempo da ordem de $O(n \log n)$.

A seleção é realizada a partir da escolha sequencial de soluções de cada fronteira não dominada (F_1, F_2, \dots) até que se obtenha o número máximo que foi estabelecido para seguir à próxima geração. Algumas decisões devem ser tomadas caso este valor seja excedido após a adição das soluções da fronteira F_i . Uma possibilidade é calcular uma medida de diversidade (como, por exemplo, a distância entre os aglomerados da população), de forma a assegurar a maior diversidade do grupo de soluções, como definido em Deb et al. (2002). Dessa forma, apenas as soluções com a maior distância entre si nos aglomerados são mantidas para as próximas iterações.

É conhecido que os operadores de cruzamento e mutação dependem da aplicação. Para o problema aqui tratado, foi escolhido um mecanismo de cruzamento uniforme (BÄCK et al., 1997) muito utilizado em codificações multivariáveis. Neste mecanismo, o cruzamento é realizado para cada variável de decisão com uma probabilidade rateCro, conforme o operador de cruzamento. O operador de cruzamento binário simulado (em inglês, SBX) foi utilizado no algoritmo, sendo (DEB, 2001) bastante conveniente para AGs com variáveis reais, em razão da sua capacidade de simular os operadores de cruzamento binário e evitar a recodificação das variáveis. As novas soluções candidatas são calculadas a partir das soluções candidatas anteriores, de acordo com as Equações 12 e 13:

$$x_{i,(1,t+1)} = 0,5((1 + \beta)x_{i,(1,t)} + (1 - \beta)x_{i,(2,t)}) \quad (12)$$

$$x_{i,(2,t+1)} = 0,5((1 - \beta)x_{i,(1,t)} + (1 + \beta)x_{i,(2,t)}) \quad (13)$$

em que β é uma variável aleatória obtida da função de distribuição de probabilidade da Equação 14:

$$f(\beta) = \begin{cases} 0.5(\eta + 1)\beta^\eta, & \text{se } \beta \leq 1, \\ 0.5(\eta + 1)\frac{1}{\beta^{\eta+2}}, & \text{caso contrário,} \end{cases} \quad (14)$$

de acordo com as técnicas clássicas para geração de variáveis aleatórias (veja detalhes em ROSS, 2012).

Uma vez que um valor da variável aleatória β tenha sido obtido, as Equações 12 e 13 podem ser utilizadas para gerar novas soluções candidatas, as quais possuem uma grande aptidão de busca, similar à de um cruzamento codificado em binário para os AGs (DEB, 2001). Pelo ajuste de η , podem ser gerados diferentes pesos β capazes de criarem soluções candidatas que sejam mais (η pequeno) ou menos semelhantes (η elevado) às descobertas.

Para cada uma das variáveis de decisão (k_i ou μ_i), o operador de mutação atua com uma probabilidade específica rate-Mut. Como sugerido por Deb (2001), perturbações gaussianas foram adicionadas às variáveis de decisão, $k_i + \epsilon_i$ e $\mu_i + \epsilon_{N+i}$, para todo $i \in N$, com $\epsilon_i \sim \text{Normal}(0, 1)$, $i \in \{1, 2, \dots, 2N\}$.

Finalmente, para garantir a factibilidade das restrições (Equações 2 e 3), após a aplicação dos operadores de cruzamento e mutação, os valores das variáveis reais devem ser arredondados adequadamente e todas estas são reajustadas

pela aplicação dos seguintes operadores de reflexão (Equações 15 e 16):

$$k_{rfl_i} = k_{\text{lowlim}} + |k_i - k_{\text{lowlim}}|, \quad (15)$$

$$\mu_{rfl_i} = \mu_{\text{lowlim}_i} + |\mu_i - \mu_{\text{lowlim}}|, \quad (16)$$

em que k_{lowlim} é o limite inferior da capacidade total alocada (ou seja, $k_{\text{lowlim}}=1$) e μ_{lowlim} é aquele da taxa de serviço alocada, de modo a assegurar que $\rho < 1$. Vale ressaltar que k_i e μ_i são os valores resultantes após o cruzamento e a mutação, e que k_{rfl_i} e μ_{rfl_i} foram obtidos após a operação de reflexão. O esquema proposto garantidamente gera soluções viáveis, sem evitar ou favorecer qualquer uma em particular.

2.5. PROBLEMAS DE CONVERGÊNCIA

Recentemente, o critério de parada dos algoritmos de otimização multiobjetivo evolutivos foi analisado em detalhes. Evidentemente, o número máximo de gerações numGen desempenha um papel importante na qualidade das soluções. No entanto, aumentar a quantidade de gerações pode não ser ideal, pois o tempo computacional é desperdiçado com muitas iterações que não levam a uma melhora significativa do resultado alcançado. Assim, Rudenko e Schoenauer (2004) sugeriram que seja critério de parada quando um número fixado de iterações for realizado sem nenhuma melhora. Para demonstrar a complexidade do tema, Rudenko e Schoenauer (2004) realizaram um estudo abrangente de experimentos computacionais. Os resultados revelaram que um critério de parada óbvio, quando toda a população está na fronteira F , não é adequado. Rudenko e Schoenauer (2004) propuseram, então, que se calculasse uma medida da estabilidade da solução não dominada após cada iteração, com base na convergência da distância máxima do aglomerado, d_l , medida ao longo de L gerações e calculada pelo seguinte desvio padrão (Equação 17):

$$\sigma_L = \sqrt{\frac{1}{L} \sum_{l=1}^L (d_l - \bar{d}_L)^2}, \quad (17)$$

em que \bar{d}_L é a média de d_l ao longo de L gerações e o critério $\sigma_L < \delta_{\text{lim}}$ deve indicar quando se deve encerrar a execução do MOEA. Rudenko e Schoenauer (2004) também sugerem que L e δ_{lim} devem ser ajustados para 40 e 0,02, respectivamente, o que leva ao critério de parada $\sigma_{40} < 0,02$.

3. RESULTADOS E DISCUSSÃO

Para aproveitar uma implementação disponível do GEM (SMITH; CRUZ, 2005), baseada na biblioteca *International Mathematics and Statistics Library* (IMSL), o algoritmo de otimização foi codificado na linguagem de programação FORTRAN. O código está disponível, mediante solicitação aos autores, para fins de ensino e pesquisa. Em primeiro lugar, os experimentos computacionais foram realizados a fim de se descobrir um conjunto subótimo de parâmetros que garantam uma rápida convergência. Por fim, uma análise detalhada de uma rede de filas foi realizada.

3.1 AJUSTE DOS PARÂMETROS

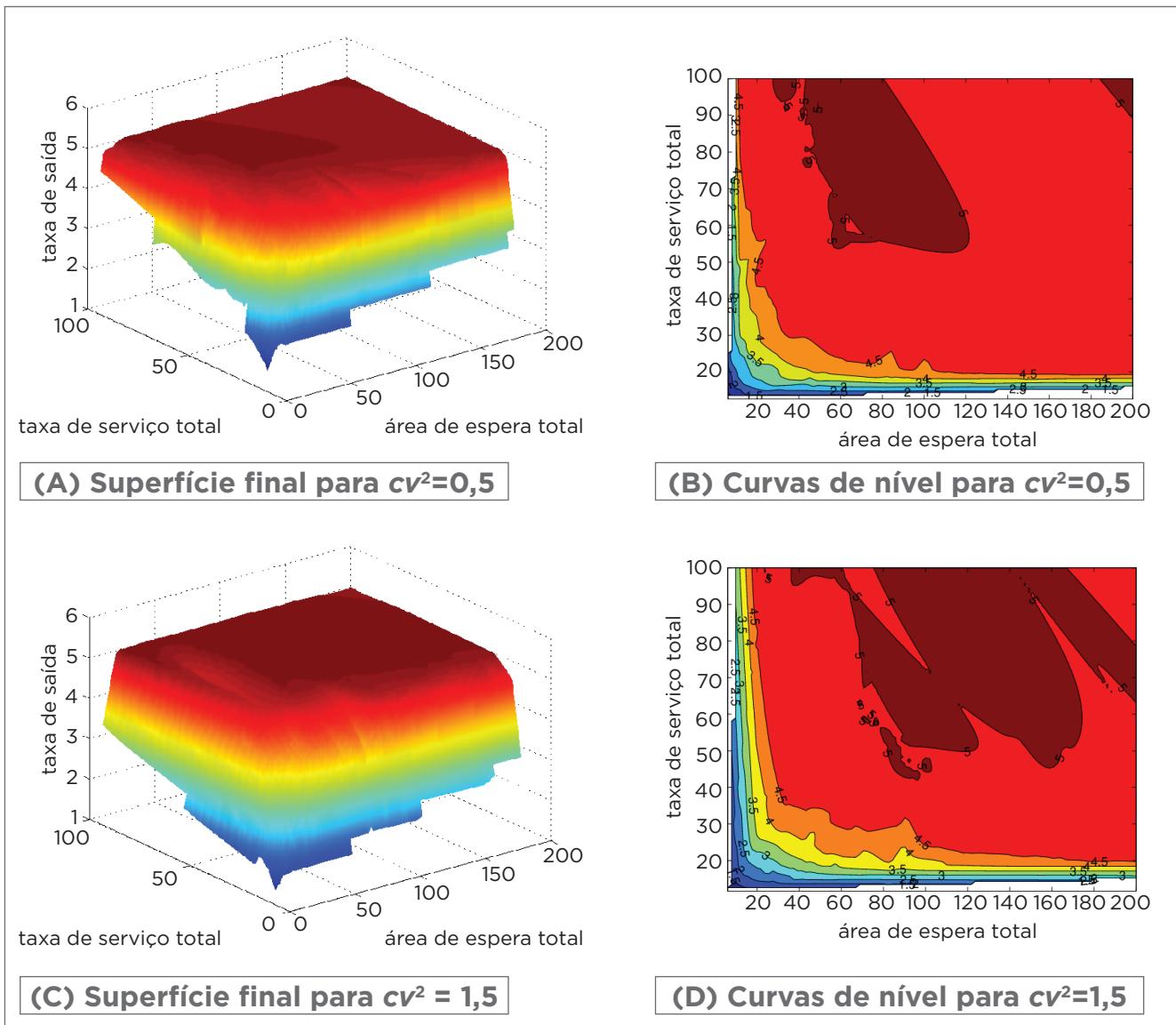
Tal como indicado por estudos anteriores sobre o AG, o conjunto subótimo de parâmetros que assegurem uma rápida convergência, com uma quantidade mínima de esforço computacional, deve ser determinado por tentativa e erro. Diferentes topologias de redes acíclicas de tamanhos variados foram testadas, e os resultados foram semelhantes.

O melhor grupo de parâmetros para o algoritmo é a seguinte combinação: (i) a utilização combinada do SBX e da mutação, com (ii) uma taxa de mutação inferior a 2%, (iii) embora quanto maior, melhor, uma população de 400 soluções candidatas parece ser suficiente, e (iv) o parâmetro de dispersão, η , não deve ser maior do que 8. Para assegurar um tempo de computação finito, o número máximo de gerações, numGen, foi ajustado para 4.000. O MOEA é suficientemente robusto e apresenta um bom desempenho para uma ampla variedade de problemas, como confirmado pela variedade de experimentos realizados (não mostrados).

3.2 ANÁLISE DE UMA REDE DE FILAS

A rede mostrada na Figura 1 foi analisada com o método proposto. Dois diferentes valores para o quadrado do coeficiente de variação do tempo de serviço foram analisados, $c\nu^2 = 0,5$ e $1,5$, com taxa de chegada $\lambda_1 = 5,0$. O ajuste do conjunto experimental foi idêntico ao da análise anterior. No entanto, os resultados indicaram que a convergência era estável a 2.000 iterações.

A Figura 4 apresenta as conclusões de tal ajuste. É possível observar o conjunto de soluções finais, representado em

**Figura 4.** Resultados finais para a rede da Figura 1.

superfícies aproximadas, e suas respectivas curvas de nível. É importante mencionar que as soluções eficientes obtidas são igualmente ótimas para a abordagem multiobjetivo. Ainda, é notável a semelhança entre as curvas de nível correspondentes às superfícies aproximadas apresentadas e o gráfico exato para uma fila única, Figura 2b. Os resultados sugerem que redes de filas parecem se comportar como uma única equivalente. É desconhecido se seria ou não possível obter algum tipo de algoritmo capaz de prever os parâmetros equivalentes para uma fila única. Adicionalmente, nota-se que, para $cv^2=0,5$, têm-se curvas de nível mais próximas da origem do que para $cv^2=1,5$. Esse

é um comportamento esperado, visto que um menor cv^2 indica uma menor variabilidade no tempo entre serviço. Os resultados obtidos mostram-se, portanto, consistentes. Outro ponto interessante é que as curvas de nível ajudam a identificar os pontos a partir dos quais não compensa mais aumentar as áreas de espera (ou as taxas de serviço), devido ao fato de serem irrisórios os ganhos nas taxas de saída, θ .

A Tabela 1 apresenta algumas soluções Pareto eficientes para uma análise mais detalhada. Nota-se que, com a abordagem multiobjetivo, é possível identificar pontos a partir dos quais não interessa aumentar o gasto

Tabela 1. Soluções Pareto eficientes selecionadas a partir de experimentos computacionais.

cv^2	Σ^k	$\Delta\%$	$\Sigma\mu$	$\Delta\%$	θ	$\Delta\%$
0,5	316		20,3		4,850	
	480	52	20,3	0	4,935	2
	340		20,5		4,899	
	343	1	26,0	27	5,000	2
1,5	118		23,6		4,656	
	378	220	23,6	0	4,956	6
	520		20,5		4,914	
	520	0	22,7	11	4,991	2

$\Delta\%$: variação percentual entre dois valores consecutivos na coluna anterior, por exemplo, $\Delta\% = 100\% \times (4,935 - 4,850) / 4,850 \approx 2$.

em recursos (áreas de espera e taxa de serviço), uma vez que o ganho obtido na taxa de saída será muito pequeno. Por exemplo, para um $cv^2=0,5$, temos que, mantida fixa a taxa total de serviço, um incremento de 52% na área de espera total produziria um ganho de apenas 2% na saída. De maneira similar, pode haver um valor para a taxa de serviço em que isso também ocorre. De fato, um aumento de 27% na taxa de serviço alocada pode produzir outro de 2%, o que pode ser considerado pouco significativo. Além disso, com um $cv^2=1,5$, tal fenômeno pode ocorrer de maneira ainda mais pronunciada. Observa-se que pode ser necessário um aumento de até 220% na área total alocada, para um de apenas 6% na taxa de serviço. É, portanto, mais vantajoso manter um sistema com uma alocação que produza, na saída 98%, da entrada ($4,914/5,000$) do que gastar 11% a mais em (taxa total alocada de) serviço, a fim de elevar a saída em 2% (isto é, aumentá-la para 99,8% da taxa de entrada). Esses são apenas alguns exemplos de análises que podem ser realizadas em filas finitas configuradas em redes, via metodologia multiobjetivo.

4. CONCLUSÕES

Com o propósito de otimizar a taxa de saída, os tamanhos das áreas de espera e as taxas de serviço das redes de filas com um único servidor e distribuição geral dos tempos de serviço, uma abordagem multiobjetivo foi apresentada. O GEM foi acoplado a um AG multiobjetivo,

o que tornou possível melhorar a compreensão das redes de filas gerais finitas. De fato, curvas de Pareto (fronteira eficiente) foram obtidas, a partir das quais foi possível a identificação de pontos notáveis para a análise de *tradeoff* entre os objetivos otimizados, quais sejam a taxa de saída, a área de espera total alocada e a taxa de serviço global alocada.

Tópicos para futuras investigações em tal área incluem extensões para redes de filas gerais multisservidoras, possivelmente por meio de núcleo estimadores (GONTIJO et al., 2011). Também é interessante considerar as diferentes medidas de desempenho, tais como o trabalho em processo (em inglês, WIP), o tempo de permanência etc. Esses são apenas alguns exemplos de possíveis temas para pesquisas.

5. AGRADECIMENTOS

Este trabalho foi parcialmente financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq (projetos 201046/1994-6, 301809/1996-8, 307702/2004-9, 472066/2004-8, 304944/2007-6, 561259/2008-9, 553019/2009-0, 550207/2010-4, 501532/2010-2 e 303388/2010-2), pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES (projeto BEX-0522/07-4) e pela Fundação de Amparo à Pesquisa do estado de Minas Gerais – FAPEMIG (projetos CEX-289/98, CEX-855/98, TEC-875/07, CEX-PPM-00401/08, CEX-PPM-00390-10, CEX-PPM-00071-12 e APQ-00613-12).

REFERÊNCIAS

- BÄCK, T.; FOGEL, D.; MICHALEWICZ, Z. (eds.) *Handbook of Evolutionary Computation*. Bristol, UK: Institute of Physics Publishing and Oxford University Press, 1997.
- BITRAN, G.R.; MORÁBITO, R. Modelos de otimização de redes de filas abertas para projeto e planejamento de job-shops. *Pesquisa Operacional para o Desenvolvimento*, v. 1, n. 1, p. 63-83, 2009.
- BRITO, N.L.C.; DUARTE, A.R.; FERREIRA, J.H.; CRUZ, F.R.B. Multiobjective Optimization of Finite Queueing Networks. In: 3rd International Conference on Engineering Optimization - EngOpt 2012, *Anais...*, p. 1-10, Rio de Janeiro, Brazil, 2012.
- CARRANO, E.G.; SOARES, L.A.E.; TAKAHASHI, R.H.C.; SALDANHA, R.R.; NETO, O.M. Electric distribution network multiobjective design using a problem-specific genetic algorithm. *IEEE Transactions on Power Delivery*, v. 21, n. 2, p. 995-1005, 2006.
- CHANKONG, V.; HAIMES, Y.Y. *Multiobjective Decision Making: Theory and Methodology*. UK: Dover Publications, 2008.
- CHAUDHURI, K.; KOTHARI, A.; PENDAVINGH, R.; SWAMINATHAN, R.; TARJAN, R.; ZHOU, Y. Server allocation algorithms for tiered systems. *Algorithmica*, v. 48, n. 2, p. 129-146, 2007.
- CRUZ, F.R.B.; KENDALL, G.; WHILE, L.; DUARTE, A.R.; BRITO, N.L.C. Throughput Maximization of Queueing Networks with Simultaneous Minimization of Service Rates and Buffers. *Mathematical Problems in Engineering*, v. 2012, Article ID 692593, 19 p, 2012.
- CRUZ, F.R.B.; VAN WOENSEL, T.; SMITH, J.M.; LIECKENS, K. On the system optimum of traffic assignment in M/G/c/c state-dependent queueing networks. *European Journal of Operational Research*, v. 201, n. 1, p. 183-193, 2010.
- DEB, K. *Multi-objective Optimisation using Evolutionary Algorithms*. Chichester, England: John Wiley & Sons, Ltd., 2001.
- DEB, K.; PRATAP, A.; AGARWAL, S.; MEYARIVAN, T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, v. 6, n. 2, p. 182-197, 2002.
- GONTIJO, G.M.; ATUNCAR, G.S.; CRUZ, F.R.B.; KERBACHE, L. Performance evaluation and dimensioning of GIX/M/c/N systems through kernel estimation. *Mathematical Problems in Engineering*, v. 2011, Article ID 348262, 20 p, 2011.
- GROSS, D.; SHORTLE, J.F.; THOMPSON, J.M.; HARRIS, C.M. *Fundamentals of Queueing Theory*. 4 ed. New York, USA: Wiley-Interscience, 2009.
- HE, Q.M.; XIE, J.; ZHAO, X. Priority queue with customer upgrades. *Naval Research Logistics*, v. 59, n. 5, p. 1520-16750, 2012.
- HILLIER, F.S. Economic models for industrial waiting line problems. *Management Science*, v. 10, n. 1, p. 119-130, 1963.
- KERBACHE, L.; SMITH, J.M. The generalized expansion method for open finite queueing networks. *European Journal of Operational Research*, v. 32, p. 448-461, 1987.
- LIN, F.T. Solving the knapsack problem with imprecise weight coefficients using genetic algorithms. *European Journal of Operational Research*, v. 185, n. 1, p. 133-145, 2008.
- MEESTER, L.E.; SHANTHIKUMAR, J.G. Concavity of the throughput of tandem queueing systems with finite buffer storage space. *Advances in Applied Probability*, v. 22, n. 3, p. 764-767, 1990.
- ROSS, S.M. *Simulation*. 5 ed. San Diego, CA: Academic Press, 2012.
- RUDENKO, O.; SCHÖNAUER, M. A steady performance stopping criterion for Pareto-based evolutionary algorithms. *Proceedings of the 6th International Multi-Objective Programming and Goal Programming Conference*. Hammamet, Tunisia, 2004.
- SMITH, J.M.; CRUZ, F.R.B. The buffer allocation problem for general finite buffer queueing networks. *IIE Transactions*, v. 37, n. 4, p. 343-365, 2005.