

SOLUÇÃO DO PROBLEMA DE LOCALIZAÇÃO DE MÁXIMA DISPONIBILIDADE UTILIZANDO O MODELO HIPERCUBO

Roberto Diéguez Galvão *

Fernando Y. Chiyoshi

Luis Gonzalo Acosta Espejo

Maria del Pilar Alvarez Rivas

Programa de Engenharia de Produção / COPPE

Universidade Federal do Rio de Janeiro

Rio de Janeiro – RJ

galvao@pep.ufrj.br

chiyoshi@uol.com.br

luis@pep.ufrj.br

* *Corresponding author*/autor para quem as correspondências devem ser encaminhadas

Recebido em 02/2002, aceito em 11/2002 após 1 revisão

Resumo

O Problema de Localização de Máxima Disponibilidade (PLMD) busca localizar m servidores tal que o máximo número de chamadas a um serviço de emergência tenha um servidor disponível a menos de uma distância crítica S predefinida, com confiabilidade α . Neste artigo o PLMD é estendido para o caso em que as taxas de ocupação são calculadas individualmente, por servidor. Isto é conseguido através do uso do Modelo Hipercubo, embutido em uma heurística de substituição de vértices. Resultados computacionais são mostrados para redes de até 150 vértices disponíveis na literatura.

Palavras-chave: localização probabilística; modelo de localização de máxima disponibilidade; modelo hipercubo.

Abstract

The Maximum Availability Location Problem (MALP) seeks to locate m servers so that the maximum number of calls for an emergency service has a server available within a predefined critical distance S , with reliability α . In this paper MALP is extended for the case where the busy fractions are calculated individually, for each server. This is achieved through the use of the Hypercube Model, embedded into a single vertex substitution heuristic. Computational results are given for networks of up to 150 vertices available in the literature.

Keywords: probabilistic location; maximum availability location problem; hypercube model.

1. Introdução

Problemas de localização no setor público podem ser classificados em duas categorias: localização de serviços não-emergenciais e localização de serviços de emergência. Na primeira categoria estão incluídos a localização de escolas, agências de correio, de alguns serviços de saúde pública e mesmo de serviços relacionados ao meio-ambiente, como suprimento de água e instalações para o depósito de lixo. A categoria de serviços de emergência inclui por exemplo a localização de hospitais, de serviços de atendimento de emergência por ambulâncias e de estações do corpo de bombeiros. As medidas de utilidade a serem otimizadas são diferentes para as duas categorias.

Na localização de serviços de emergência busca-se em geral prover *cobertura* a áreas de demanda. A noção de *cobertura* implica na definição de uma *distância (tempo) de serviço*, que é a *distância (tempo) crítica* além da qual a área de demanda é considerada *não coberta*. Uma área de demanda é portanto considerada *coberta* se está a menos da distância crítica de pelo menos uma das facilidades (servidores) existentes.

Os primeiros modelos a serem desenvolvidos para a localização de serviços de emergência foram modelos determinísticos. O mais simples dos modelos matemáticos existentes para problemas de localização com restrições de cobertura determinísticos é o correspondente ao *Problema de Localização para a Cobertura de Conjuntos (PLCC)*, que consiste na determinação do número mínimo de facilidades e de sua localização, de tal forma que cada área de demanda esteja a menos da distância crítica de pelo menos uma das facilidades localizadas.

Muitas vezes atingir a cobertura total para uma dada região, como no *PLCC*, pode tornar-se inviável do ponto de vista econômico, no sentido que o número de facilidades necessárias pode não ser compatível com os recursos disponíveis. Usualmente busca-se uma solução de compromisso, que proporcione níveis de cobertura aceitáveis e seja financeiramente mais acessível. O *Problema de Localização de Máxima Cobertura (PLMC)* foi desenvolvido por Church & ReVelle (1974) com este propósito. Neste caso o objetivo é localizar um número pré-especificado de facilidades (digamos m facilidades), compatível com os recursos disponíveis, tal que a máxima população possível de uma dada região seja coberta a menos de uma distância crítica S predefinida.

Uma desvantagem dos modelos determinísticos descritos acima é que eles partem da hipótese que os servidores estão disponíveis quando solicitados, o que nem sempre é razoável em aplicações práticas. Em sistemas não congestionados, com pouca demanda, a hipótese é razoável, mas em sistemas congestionados, nos quais chamadas frequentes mantêm por exemplo ambulâncias na rua 20% a 30% do tempo, a hipótese é totalmente injustificada. O congestionamento em serviços de atendimento de emergência, que pode causar a indisponibilidade de um servidor quando solicitado, foi um dos fatores que motivou o desenvolvimento dos modelos de localização probabilísticos.

Uma revisão bibliográfica detalhada de problemas de localização probabilísticos é apresentada por Owen & Daskin (1998). Swersey (1994) examina também alguns modelos probabilísticos. Chiyoshi *et al.* (2000) analisam o uso do modelo hipercubo em modelos probabilísticos. Na modelagem probabilística de serviços de emergência algumas hipóteses simplificadoras permitem o uso da programação matemática. Situações em que hipóteses simplificadoras não são aplicáveis conduzem, no entanto, ao tratamento desses problemas através da teoria das filas.

O modelo hipercubo de Larson (1974, 1975), examinado na Seção 2 deste artigo, possibilita a utilização da teoria das filas em modelos de localização probabilísticos. O modelo hipercubo é um modelo descritivo, que permite o cálculo de medidas de desempenho que caracterizam um dado sistema com base em filas espacialmente distribuídas. Aplicado de forma isolada não assegura a solução direta de problemas de localização, sendo no máximo uma ferramenta que pode ser aplicada ao estudo de diversos cenários. Este modelo tem sido no entanto embutido por alguns autores em métodos heurísticos desenvolvidos para resolver problemas de localização probabilísticos, ver por exemplo Batta *et al.* (1989).

Modelos de localização probabilísticos definidos na década de 80 incluem o *Problema de Máxima Cobertura Esperada (PMCE)*, definido por Daskin (1983) e o *Problema de Localização de Máxima Disponibilidade (PLMD)*, que é derivado do *PLMC*, seu equivalente determinístico. O *PLMD*, introduzido por ReVelle & Hogan (1989), busca localizar m servidores tal que o máximo número de chamadas a um serviço de emergência tenha um servidor disponível a menos da distância crítica S com confiabilidade α . Ou seja, deseja-se que uma chamada a um serviço de emergência encontre um servidor disponível a menos da distância S com probabilidade $P \geq \alpha$, sendo α um valor predefinido.

O *PLMD* foi apresentado por ReVelle e Hogan em duas versões, ambas incorporando hipóteses simplificadoras, que permitem definir uma formulação determinística equivalente para o mesmo. As duas versões diferem na maneira pela qual a fração do tempo em que os servidores estão ocupados é calculada. A primeira versão, *PLMDI*, parte da hipótese que todos os servidores têm a mesma taxa de ocupação, igual à taxa de ocupação média do sistema, ρ . Já na segunda versão, *PLMDII*, calcula-se de forma diferenciada a fração do tempo em que os servidores estão ocupados, com valores específicos para diferentes setores da região geográfica em consideração.

Neste trabalho apresentamos uma extensão do *PLMD* definido por ReVelle & Hogan (1989), na qual as taxas de ocupação são calculadas por servidor, utilizando o modelo hipercubo. A solução do modelo é obtida por um método heurístico de substituição de vértices, desenvolvido originalmente por Teitz & Bart (1968) para o problema das p -medianas. Por este método busca-se achar configurações alternativas que melhorem o valor da função objetivo do problema. Para cada configuração considerada é necessário resolver o modelo hipercubo, para que o valor da máxima cobertura com confiabilidade α correspondente possa ser calculado. A heurística de substituição de vértices prossegue até que nenhuma troca simples de localização de servidores (mudança de localização de um único servidor) produza uma melhora no valor da máxima cobertura com confiabilidade α .

O trabalho está organizado da seguinte forma. Na Seção 2 são apresentados o *PLMD* desenvolvido por ReVelle & Hogan (1989), algumas considerações sobre o modelo hipercubo e a relaxação de algumas hipóteses básicas do *PLMD*, que gera uma extensão do *PLMD* incorporando o modelo hipercubo. Na Seção 3 é apresentado o método heurístico de substituição de vértices para a solução do *PLMD* estendido, com taxas de ocupação calculadas por servidor através do modelo hipercubo. Resultados computacionais são apresentados na Seção 4, o que é seguido pelas Conclusões (Seção 5).

2. Problema de Localização de Máxima Disponibilidade

Conforme já dito, o *PLMD* definido por ReVelle & Hogan (1989) foi apresentado em duas versões, que diferem na maneira pela qual a fração do tempo em que os servidores estão ocupados é calculada. A primeira versão, *PLMDI*, parte da hipótese que todos os servidores têm a mesma taxa de ocupação, igual à taxa de ocupação média do sistema, ρ . Já na segunda versão, *PLMDII*, calcula-se de forma diferenciada a fração de tempo em que os servidores estão ocupados, com valores específicos para diferentes setores da região geográfica em consideração.

Note que a definição de taxas de ocupação específicas para diferentes regiões geográficas em *PLMDII* não é equivalente a definir taxas de ocupação por servidor, nem permite o cálculo de tais taxas, o que exigiria que o modelo hipercubo fizesse parte do método de solução; isto não foi considerado por ReVelle e Hogan em seu desenvolvimento. Assim, embora a utilização de taxas de ocupação específicas para cada região represente um avanço em relação à hipótese simplificadora de *PLMDI*, *PLMDII* também incorpora hipóteses simplificadoras. No que se segue mostraremos a obtenção de uma formulação de programação matemática para *PLMDI*. O modelo matemático correspondente a *PLMDII* não será analisado neste trabalho.

2.1 Formulação Matemática de *PLMDI*

Em *PLMDI*, ρ , a fração do tempo em que os servidores estão ocupados, é estimada através de uma fórmula desenvolvida por Daskin (1982). Sejam ϕ_j o número de chamadas por dia originadas na área de demanda $j \in J$ e \bar{t} a duração média do atendimento a uma chamada, medida em horas. Então ρ pode ser calculada por

$$\rho = \frac{\bar{t} \sum_{j \in J} \phi_j}{24 \sum_{i \in I} y_i} = \frac{\bar{t} \sum_{j \in J} \phi_j}{24m},$$

isto é, a fração ocupada de cada servidor é calculada como sendo a divisão do número médio de horas de serviço necessárias por dia no sistema pelo número de horas diárias disponíveis, levando-se em conta que serão localizados m servidores.

Sejam I ($|I|=n$) o conjunto de locais onde servidores podem ser localizados; $a_{ij}=1$ se a área de demanda j está a menos da distância crítica S de um local $i \in I$ ($a_{ij}=0$ caso contrário); $y_i=1$ se um servidor for localizado em $i \in I$ ($y_i=0$ caso contrário). A restrição de que pelo menos um servidor deve estar disponível a menos de S de uma área de demanda $j \in J$, com probabilidade maior ou igual a α , pode ser escrita da seguinte forma:

$$\begin{aligned} P[\text{um ou mais servidores disponíveis a menos de } S] &\geq \alpha \equiv \\ &\equiv (1 - P[\text{nenhum servidor disponível a menos de } S]) \geq \alpha = \\ &= 1 - \rho^{\sum_{i \in I} a_{ij} y_i} \geq \alpha, \end{aligned} \quad (1)$$

onde $\sum_{i \in I} a_{ij} y_i$ é o número de servidores disponíveis a menos de S da área de demanda $j \in J$.

Ou, tomando logaritmos, $\sum_{i \in I} a_{ij} y_i \geq d$, onde $d = \left\lceil \frac{\log(1-\alpha)}{\log \rho} \right\rceil$, $\lceil t \rceil$ denotando o menor inteiro maior ou igual a t .

Da expressão linear equivalente obtida para a restrição probabilística é possível notar que cada área de demanda $j \in J$ exige a presença de pelo menos d servidores disponíveis a menos da distância crítica S para que seja assegurada uma cobertura da mesma com confiabilidade α . Portanto, para maximizar o número de chamadas atendidas por um serviço de confiabilidade α , é necessário maximizar o número de chamadas com pelo menos d servidores disponíveis a menos da distância S .

Seja a variável $x_{jk} \in \{0,1\}$ tal que $x_{jk} = 1$ se a área de demanda j tem pelo menos k servidores a menos S , $x_{jk} = 0$ caso contrário. A expressão $\sum_{k=1}^n x_{jk}$ representa o número de vezes que a área de demanda $j \in J$ é coberta a menos de S . É finalmente possível escrever a seguinte expressão, que conta o número de coberturas disponíveis para cada área de demanda $j \in J$ com confiabilidade α : $\sum_{k=1}^d x_{jk} \leq \sum_{i \in I} a_{ij} y_i, j \in J$. Para maximizar o número de chamadas cobertas com confiabilidade α devemos maximizar $\sum_{j \in J} \phi_j x_{jd}$. A formulação matemática de *PLMDI* pode ser finalmente escrita da seguinte forma:

$$\text{Maximizar } Z = \sum_{j \in J} \phi_j x_{jd}, \quad (2)$$

$$\text{sujeito a } \sum_{k=1}^d x_{jk} \leq \sum_{i \in I} a_{ij} y_i, j \in J, \quad (3)$$

$$x_{jk} \leq x_{j(k-1)}, j \in J, k = 2, \dots, d, \quad (4)$$

$$\sum_{i \in I} y_i = m, \quad (5)$$

$$y_i, x_{jk} \in \{0,1\}, i \in I, j \in J, k=2, \dots, d. \quad (6)$$

As restrições (3) asseguram que uma área de demanda $j \in J$ tem cobertura com confiabilidade α se existirem pelo menos d servidores a menos de S da mesma. Já as restrições (4) expressam o fato que para uma área de demanda ser coberta por k servidores ela tem que ser coberta por pelo menos $(k-1)$ servidores, para $2 \leq k \leq d$. As restrições (5) estabelecem que m servidores devem ser localizados, enquanto as restrições (6) definem a natureza binária das variáveis de decisão.

ReVelle & Hogan (1989) resolveram o *PLMD* para dados do sistema de combate a incêndios da cidade de Baltimore, nos Estados Unidos. Esses dados consistem de 207 áreas de planejamento, com a frequência média de chamadas por unidade de tempo conhecida para cada área. Eles analisaram a localização de ambulâncias nesse sistema, com uma ou mais ambulâncias podendo ser localizadas em cada uma das 31 estações do corpo de bombeiros existentes na região em estudo. Para resolver o problema os autores utilizaram o “pacote” de programação matemática MPSX. Infelizmente, no entanto, não são fornecidos detalhes dos resultados computacionais obtidos, o que impede uma avaliação mais acurada da metodologia utilizada.

2.2 O Modelo Hipercubo

O modelo hipercubo é um modelo desenvolvido por Larson (1974) para descrever sistemas de filas em que as demandas por serviço ocorrem distribuídas no espaço e os atendimentos requerem os deslocamentos das unidades prestadoras de serviço até o local da demanda.

Dentre várias características que lhe são peculiares, duas são mais importantes para distinguir o modelo hipercubo dos modelos convencionais de filas. A primeira refere-se ao fato de que se supõe que a área de demanda está dividida em átomos, cada um com demanda específica. A segunda refere-se a que a operação do sistema faz-se através de uma política de despachos em que se define a priori, para cada átomo de demanda, uma ordem de prioridades do uso das unidades de serviço, de modo que o atendimento de uma chamada originada em um dado átomo é feito pela unidade de maior prioridade que estiver livre.

Para descrever o modelo hipercubo não basta especificar o número de servidores ocupados como nos modelos convencionais de fila: é necessário conhecer a situação de cada servidor individualmente. Para tanto utiliza-se uma variável binária associada a cada servidor, com os valores 0 e 1 representando os estados “livre” e “ocupado” do servidor, respectivamente. Desta forma, no caso mais simples de um modelo hipercubo com m servidores que não comporta fila de espera, os estados do sistema seriam representados por um vetor de variáveis binárias $\{b_1, b_2, \dots, b_m\}$, onde b_e indica o estado atual do servidor e . Observe-se que os estados do sistema de m servidores são representados pelos vértices de um hipercubo unitário no espaço m -dimensional, sendo este o fato que dá o nome ao modelo.

Para resolver o modelo hipercubo parte-se das hipóteses que as demandas (chegadas a cada átomo do sistema) têm distribuições de Poisson e que os tempos de serviço têm distribuições exponenciais. Sob tais hipóteses constrói-se as equações de equilíbrio em torno de cada estado do sistema, obtendo-se um sistema de 2^m equações lineares cuja solução produz as probabilidades de equilíbrio associadas aos estados do sistema. Essas probabilidades por sua vez permitem determinar as principais características operacionais do modelo, em particular as taxas de ocupação dos servidores. Esta forma de resolver o modelo hipercubo, conhecida na literatura como “método exato”, apresenta o inconveniente de envolver a solução de um sistema linear de 2^m equações, com um custo computacional que pode tornar-se proibitivo mesmo para valores moderados de m .

Como alternativa ao método exato, de questionável eficiência computacional, Larson (1975) desenvolveu um método aproximado computacionalmente mais eficiente, por envolver um sistema não-linear de apenas m equações em que as incógnitas são as taxas de ocupação dos servidores. As equações são construídas a partir da relação que existe entre a taxa total de despachos de um servidor e sua taxa de ocupação. Na construção dessas equações há a necessidade de avaliar as probabilidades associadas a eventos da forma $\{B_1 B_2 \dots B_l F_{l+1}\}$, de que uma chamada encontre seus primeiros l servidores preferenciais ocupados e o $(l+1)$ -ésimo servidor preferencial livre. Se os servidores operassem independentemente uns dos outros, essas probabilidades seriam dadas por $\rho_1 \rho_2 \dots \rho_l (1 - \rho_{l+1})$. Como trata-se de um sistema que pressupõe cooperação entre servidores, Larson propõe o uso de um fator de correção Q para levar em conta a não independência dos servidores, escrevendo:

$$\text{Prob}\{B_1 B_2 \dots B_l F_{l+1}\} = Q(m, \rho, l) \rho_1 \rho_2 \dots \rho_l (1 - \rho_{l+1}).$$

Para derivar o fator de correção Q , Larson reinterpreta o evento $\{B_1 B_2 \dots B_l F_{l+1}\}$ como o evento de que, em uma amostragem sem reposição de servidores de um modelo M/M/m de fila, os primeiros l servidores selecionados estejam ocupados e o $(l+1)$ -ésimo servidor selecionado esteja livre. Determina então a probabilidade deste evento sob a nova interpretação, e nela identifica o componente que incorpora os efeitos da não independência dos servidores, tomando-o como fator de correção a ser usado no modelo hipercubo.

2.3 Relaxação das Hipóteses do Problema de Máxima Disponibilidade (PLMDI)

Conforme visto em 2.1, em *PLMDI* considera-se que a demanda originada de um átomo j está coberta se a probabilidade de que exista pelo menos um servidor a menos da distância crítica é igual ou maior do que α . A restrição correspondente é dada por (1).

Esta formulação assume que os servidores têm a mesma taxa de ocupação e operam independentemente uns dos outros. Consideradas no contexto de um modelo hipercubo, cada uma das hipóteses subjacentes poderia ser aceitável como aproximação. Por exemplo, se o sistema operasse com taxa de ocupação global elevada, a cooperação entre servidores tenderia a homogeneizar as taxas de ocupação dos mesmos e o uso de uma taxa comum de ocupação para todos os servidores tenderia a não envolver graves distorções. Por outro lado, se a taxa global de ocupação fosse baixa, a cooperação entre servidores não seria importante, tornando aceitável a hipótese de independência dos mesmos. Percebe-se no entanto que, em conjunto, as hipóteses do modelo de máxima disponibilidade atuam de forma tal que uma fragiliza a outra: quando há pouca cooperação, a relativa independência dos servidores torna importante a heterogeneidade das taxas de ocupação; quando há muita cooperação, a hipótese de independência dos servidores torna-se questionável.

As hipóteses do problema de máxima disponibilidade podem ser relaxadas com o uso dos resultados do modelo hipercubo. Para isso é necessário inicialmente redefinir a variável de localização y_i uma vez que, para introduzir as taxas de ocupação específicas de cada servidor, não é mais suficiente saber que há um servidor localizado na base potencial i : é necessário identificar o servidor ali localizado. Pode-se então definir uma variável de localização y_{ik} , tal que $y_{ik}=1$ se o servidor k está localizado em i , $y_{ik}=0$ caso contrário.

Usando as taxas de ocupação específicas dos servidores e introduzindo o fator de correção de não independência dos mesmos, a restrição de cobertura pode ser escrita como:

$$\left\{ 1 - \prod_{k=1}^m \rho_k^{\sum_{i \in I} a_{ij} y_{ik}} Q \left(m, \rho, \sum_{i \in I} \sum_{k=1}^m a_{ij} y_{ik} - 1 \right) \right\} \geq \alpha.$$

Quando uma área de demanda j não é coberta por ao menos um servidor o terceiro parâmetro de Q torna-se -1. Como Q não é definido para este valor, convencionamos $Q(m, \rho, -1) = 1$.

Note-se que, diferentemente da formulação original de ReVelle & Hogan (1989), o uso da variável de localização mais detalhada y_{ik} permite que mais de um servidor sejam localizados em uma mesma base. Para remover esta flexibilidade seriam necessárias restrições adicionais da forma

$$\sum_{k=1}^m y_{ik} \leq 1, i \in I.$$

Quando as hipóteses originais do problema de máxima disponibilidade são relaxadas, obtemos o que chamamos de *Problema de Máxima Disponibilidade Estendido (PLMDE)*. Nesse caso a restrição de cobertura passa a incorporar variáveis (no caso as taxas de ocupação dos servidores) que dependem das variáveis de decisão. Em consequência, a restrição de cobertura em sua formulação mais geral não pode ser usada em um modelo de programação matemática. Ela permite no entanto determinar a cobertura correspondente a uma dada configuração (localização dos servidores), podendo ser útil na construção de métodos heurísticos. É disso que tratamos a seguir na Seção 3.

3. A Solução do Problema de Máxima Disponibilidade Estendido

Uma formulação matemática do *PLMDE* poderia ser dada definindo-se a variável de decisão z_j como $z_j=1$ se a área de demanda j tem pelo menos um servidor disponível a menos da distância crítica S com confiabilidade α ($z_j=0$ caso contrário):

$$\text{Maximizar } Z = \sum_{j \in J} \phi_j z_j, \quad (7)$$

$$\text{Sujeito a } \left[\left\{ 1 - \prod_{k=1}^m \rho_k^{\sum_{i \in I} a_{ij} y_{ik}} Q \left(m, \rho, \sum_{i \in I} \sum_{k=1}^m a_{ij} y_{ik} - 1 \right) \right\} - \alpha \right] z_j \geq 0, j \in J, \quad (8)$$

$$\sum_{i \in I} \sum_{k=1}^m y_{ik} = m, \quad (9)$$

$$y_{ik}, z_j \in \{0,1\}, i \in I, j \in J, k = 1, \dots, m. \quad (10)$$

O modelo matemático dado por (7)-(10) é de pouca utilidade prática dado que, conforme já observado, as restrições de cobertura (8) passam a incorporar parâmetros (ρ_k , Q) que dependem das variáveis de decisão. A estrutura do modelo possibilita, no entanto, o desenvolvimento de métodos heurísticos para solucioná-lo. Neste artigo propomos um método heurístico para o *PLMDE* que possui os seguintes componentes: (i) um mecanismo que gera as localizações dos servidores; (ii) um procedimento para calcular as taxas de ocupação ρ_k correspondentes a estas localizações; (iii) uma função que calcula o valor do fator de correção Q e (iv) um procedimento para calcular a cobertura esperada correspondente.

A heurística que desenvolvemos tem por base o método de substituição de vértices de Teitz & Bart (1968). Existem na literatura versões melhoradas da heurística de Teitz e Bart, por exemplo a heurística *Fast Interchange* desenvolvida inicialmente por Whitaker (1983), aprimorada por Hansen & Mladenovic (1998) e usada recentemente por Galvão *et al.* (2002) e Espejo *et al.* (2003). A principal característica desta heurística é sua rapidez para avaliar soluções. Devido à estrutura do *PLMDE* esta característica do *Fast Interchange* não pode infelizmente ser aproveitada.

A heurística desenvolvida busca achar configurações alternativas que melhorem o valor da função objetivo do problema (máxima cobertura com confiabilidade α). Para cada configuração considerada é necessário resolver o modelo hipercubo e calcular o fator de correção Q , para que o valor da função objetivo correspondente possa ser calculado. A heurística prossegue até atingir a *Condição de Parada*, que acontece quando nenhuma troca simples de localização de servidores (mudança de localização de um único servidor) produz uma melhora no valor da função objetivo. Para otimizar o código computacional correspondente o procedimento para calcular o valor de Q é executado uma única vez, quando é criada uma tabela com os valores de Q que serão necessários durante a execução do algoritmo. O procedimento da heurística é descrito abaixo em pseudo-código.

Procedure com base na Substituição de Vértices para o PLMDE
{ inicialização: lê dados e obtém as taxas de chegada }
calcular fatorQ;
construir localização Inicial(LocIni);
rodar Hipercubo; { calcula taxas de ocupação utilizando o modelo Hipercubo }
calcular cobertura (CobPop);
MelhorSol = CobPop; { guarda a melhor solução }
LocAux = LocIni; { guarda a solução corrente para desfazer a troca }
MelhorLoc = LocIni; { guarda melhor localização }
While (Condição de parada = false) **do**
 i=0
 While (i < NumÁreasDeDemanda) **do**
 i = i+1;
 For índice = 1 **to** NumServidores **do**
 iEntra = i;
 iSai = índice;
 If (LocIni[iSai] ≠ iEntra) **then**
 LocIni[iSai] = iEntra; { tenta alternativa de troca }
 rodar Hipercubo;
 calcular cobertura(CobPop); { cobertura para a localização corrente }
 If (CobPop > MelhorSol) **Then**
 MelhorSol = CobPop; { atualiza a melhor solução }
 Meli = iEntra; { guarda a melhor alternativa para entrar na solução }
 Piori = iSai; { guarda o índice da alternativa que sairá da solução }
 End_if
 LocIni[iSai] = LocAux[iSai]; { restitui a solução corrente }
 End_if
 End_for
 If (houve uma melhora na cobertura) **then**
 MelhorLoc[Piori] = Meli; { atualiza a melhor localização }
 LocIni[Piori] = Meli; { atualiza a localização corrente }
 LocAux = LocIni; { atualiza localização }
 i=0; { possibilita localizar mais de um servidor por área de demanda }
 End_if
 End_while
 fazer teste de parada;
End_while
Escrever resultados;
End_procedure.

Os fatores que influenciam o desempenho da heurística são: a localização inicial dos servidores, o método usado para resolver o modelo Hipercubo e a estratégia usada para realizar a substituição de vértices. A seguir discutiremos cada um desses pontos.

Localização Inicial dos servidores – Ao iniciar a heurística restringimos a localização a um único servidor por área de demanda. Depois de obter a localização inicial dos servidores esta restrição é relaxada. Foram testadas duas estratégias para escolher a localização inicial dos servidores: (i) selecionar aleatoriamente m locais do conjunto I ; (ii) seja $c_i = \sum_{j \in J} a_{ij} \phi_j$ o

número total de chamadas que um servidor poderia atender se estivesse localizado em $i \in I$ (assumindo que o servidor está sempre disponível); selecionar então os m locais com os maiores c_i (empates resolvidos arbitrariamente). A estratégia (ii) foi a que deu melhores resultados. Os resultados computacionais mostrados na Seção 4 usam esta estratégia.

Método Usado para Resolver o Modelo Hipercubo – Testamos o uso dos métodos exato e aproximado para resolver o modelo Hipercubo. Na Seção 4 mostramos resultados computacionais obtidos com os dois métodos de solução; esses resultados justificam a escolha do método aproximado para resolver este modelo.

Estratégia Usada para a Substituição de Vértices – Por hipótese todas as áreas de demanda são candidatas à localização de um ou mais servidores, e um servidor pode ser localizado em qualquer das áreas de demanda. Durante o processo iterativo da heurística temos portanto dois conjuntos: um conjunto com as localizações correntes (o conjunto *LocIni*, percorrido usando a variável *índice*) e um conjunto com as áreas de demanda, que são bases em potencial para os servidores. Observe-se que não usamos uma variável específica para representar este conjunto de bases em potencial, e sim o índice i que indica a área que está sendo considerada.

Testamos duas estratégias para a substituição de vértices: troca na melhor de todas as possíveis substituições e troca na primeira melhor substituição. Trocar na melhor de todas as substituições implica em selecionar uma a uma cada base em potencial, e avaliar a melhora na função objetivo ao se substituir tentativamente cada uma das localizações correntes pela base em potencial que está sendo avaliada. Depois de avaliar todas as possibilidades se realiza uma troca sempre e quando se produz uma melhora no valor da máxima cobertura com confiabilidade α . Trocar na primeira melhor substituição implica em selecionar uma base em potencial e avaliar a melhora na função objetivo ao se substituir tentativamente cada uma das localizações correntes pela base em potencial que está sendo avaliada. Caso se produza uma melhora na função objetivo se realiza a troca, caso contrário passa-se a avaliar outra base em potencial. Evidentemente a segunda possibilidade reduz o esforço computacional necessário. Alguns resultados computacionais obtidos com estas duas estratégias são mostrados na Seção 4. A versão da heurística apresentada acima (pseudo-código) corresponde à versão com a troca na primeira melhor substituição.

4. Resultados Computacionais

Nesta seção mostramos um resumo dos resultados computacionais que obtivemos ao resolver algumas instâncias do *PLMDE* utilizando a heurística apresentada na Seção 3. Os experimentos computacionais estão estruturados da seguinte maneira. Inicialmente determinamos duas componentes principais da heurística para o *PLMDE*: o método de

solução a ser utilizado para resolver o modelo hipercubo e a estratégia utilizada na troca de vértices. A seguir analisamos os resultados computacionais obtidos com a heurística para redes de 55, 100 e 150 vértices.

Os testes computacionais foram realizados utilizando um microcomputador com processador AMD Athlon de 1.1 Ghz, barramento de 200 Mhz e memória Ram de 128 Mb. A heurística foi codificada em Fortran PowerStation 4.0 da Microsoft.

4.1 Problemas Testados

A heurística foi testada em uma variedade de redes disponíveis na literatura: a rede de 55 vértices de Swain (1971) e as redes de 100 e 150 vértices geradas aleatoriamente por Galvão & ReVelle (1996). Todos os vértices de demanda são locais potenciais para a localização dos servidores e se permite localizar mais de um servidor por vértice. No nosso caso a demanda associada a cada vértice dessas redes corresponde ao número de chamadas por dia originadas em cada área de demanda.

As instâncias criadas consideram possíveis combinações dos seguintes parâmetros: nível de confiabilidade (α); taxa de ocupação do sistema (ρ); número de servidores (m); distância crítica (S). A taxa de ocupação ρ é a taxa média das taxas de ocupação dos servidores individuais e varia de $\rho = 0,10$ a $\rho = 0,50$, com incrementos de 0,10. Os valores utilizados para esses parâmetros e o número de problemas gerados para cada rede são mostradas na Tabela 1; no total foram gerados 240 problemas teste.

Tabela 1 – Resumo dos problemas teste

Rede	Nível de Confiabilidade α	Taxa de Ocupação do Sistema ρ	Número de Servidores m	Distância Crítica S	Número de Problemas
55	0,90; 0,95	0,10 - 0,50	8; 10; 12	10; 15	60
100	0,90; 0,95	0,10 - 0,50	8; 10; 12	70; 80; 90	90
150	0,90; 0,95	0,10 - 0,50	8; 10; 12	80; 100; 120	90

Neste trabalho tratamos de sistemas com servidores homogêneos, adotando como tempo médio de serviço a unidade de tempo; assim, a capacidade total de atendimento de um sistema de m servidores é dada por $\sum_{k=1}^m \mu_k = m$. Para obter uma dada taxa de ocupação ρ (relação entre demanda e capacidade de atendimento), a demanda total deve ser ajustada de forma que $\sum_{j \in J} \lambda_j = \rho m$. As taxas individualizadas por área podem ser obtidas rateando a

demanda total proporcionalmente à demanda diária de cada área: $\lambda_j = \frac{\phi_j \rho m}{\sum_{j \in J} \phi_j}$.

4.2 Determinação da Heurística para o *PLMDE*

Na determinação da heurística para o *PLMDE* avaliamos dois fatores que influenciam o desempenho da mesma: o método de solução a ser utilizado para resolver o modelo hipercubo e a estratégia utilizada na heurística de substituição de vértices. A avaliação desses fatores é realizada com base na qualidade da solução obtida (porcentagem da população coberta com confiabilidade α) e no tempo computacional consumido (em segundos de CPU). Para esta avaliação utilizamos os 60 problemas teste obtidos a partir da rede de 55 vértices de Swain (1971).

Conforme visto na seção 2.2, o modelo hipercubo pode ser resolvido de duas formas: por meio do método exato ou por meio do método aproximado. O método exato envolve resolver um sistema de 2^m equações lineares, enquanto que o método aproximado implica em resolver um sistema não-linear de apenas m equações. Observe-se na Tabela 2 que, no que diz respeito à qualidade da solução, os resultados obtidos pelos dois métodos de solução do modelo hipercubo estão muito próximos. Já em relação ao esforço computacional, o método aproximado consegue reduzir o tempo utilizado pelo método exato em até duas ordens de grandeza. Esses resultados justificam a escolha do método aproximado para resolver o modelo hipercubo.

Para determinar que estratégia usar para a troca de vértices no método de substituição foram testadas duas estratégias: troca na primeira melhor substituição e troca na melhor de todas as substituições. Na maioria dos problemas as porcentagens das coberturas obtidas estão próximas, mas a troca na primeira melhor substituição reduz o esforço computacional necessário. Nos testes realizados para determinar o método para resolver o modelo hipercubo, foi utilizada a estratégia da troca na melhor substituição. Nos testes para determinar a estratégia para a troca de vértices, o modelo hipercubo foi resolvido de maneira aproximada. Isto explica os resultados iguais das colunas *método aproximado* e *melhor substituição* da Tabela 2.

Com base nos resultados da Tabela 2 decidimos que a heurística para o *PLMDE* deve utilizar o método aproximado para solucionar o modelo hipercubo e que a estratégia de troca de vértices deve utilizar a primeira melhor substituição.

Tabela 2 – Média da porcentagem de cobertura e do tempo computacional.
Rede de 55 vértices de Swain (1971)

	Solução do Modelo Hipercubo		Troca de Vértices	
	Método Exato	Método Aproximado	Primeira Melhor Substituição	Melhor Substituição
% Cobertura	86,72%	86,97%	86,46%	86,97%
Tempo (Seg. CPU)	2715,87	16,18	9,50	16,18

4.3 Análise dos Resultados Computacionais Obtidos com a Heurística para o *PLMDE*

A análise dos resultados obtidos com a heurística para o *PLMDE* consiste em mostrar:

- O comportamento da cobertura em relação aos parâmetros do problema (nível de confiabilidade, taxa de ocupação dos servidores, número de servidores e distância crítica).
- O comportamento das taxas de ocupação dos servidores para algumas instâncias de interesse da rede de 100 vértices.

Na Tabela 3 apresentamos um resumo dos tempos computacionais consumidos pela heurística para resolver os 240 problemas teste. Deve-se destacar a rapidez da heurística para resolver esses problemas.

Tabela 3 – Resumo dos tempos computacionais para as redes testadas

Rede	Média (segs. CPU)	Desvio Padrão	Mínimo (segs. CPU)	Máximo (segs. CPU)
55	16,18	10,40	2,00	43,00
100	21,66	16,30	3,00	92,00
150	41,29	38,48	8,00	210,00

Comportamento da cobertura em relação aos parâmetros do problema

Mostramos a seguir, nas tabelas 4, 5 e 6, o comportamento da cobertura obtida com a heurística para o *PLMDE* para as instâncias de 55, 100 e 150 vértices, respectivamente. Essas tabelas possibilitam observar o nível de cobertura atingido para diferentes combinações dos parâmetros do problema. Por exemplo, uma análise por linhas nessas tabelas permite observar que mantendo fixos os valores do nível de confiabilidade, taxa de ocupação do sistema e distância crítica e aumentando o número de servidores, o nível de cobertura atingido em geral aumenta. Realizando uma análise por colunas das mesmas tabelas pode-se observar que para um mesmo valor da distância crítica, número de servidores e nível de confiabilidade, a cobertura diminui quando se aumenta a taxa de ocupação do sistema.

Tabela 4 – Cobertura atingida com a heurística para o PLMDE.
Rede de 55 vértices de Swain (1971)

α	ρ	$S = 10$ Número de Servidores			$S = 15$ Número de Servidores		
		8	10	12	8	10	12
0,90	0,10	98,44	99,69	99,53	98,75	100,00	100,00
	0,20	85,47	90,31	87,50	94,53	94,53	100,00
	0,30	80,78	82,66	77,03	96,72	98,44	97,50
	0,40	74,84	78,91	83,75	91,41	90,00	92,03
	0,50	70,16	77,66	72,66	83,59	87,81	90,00
0,95	0,10	93,75	91,09	87,50	95,47	99,69	100,00
	0,20	79,38	80,00	77,03	96,25	91,09	96,25
	0,30	71,72	78,91	80,16	96,41	90,00	89,84
	0,40	77,81	71,72	72,97	83,59	90,00	94,53
	0,50	66,41	70,31	69,84	80,47	81,88	89,69

Tabela 5 – Cobertura atingida com a heurística para o PLMDE.
Rede de 100 vértices de Galvão & ReVelle (1996)

α	ρ	$S = 70$			$S = 80$			$S = 90$		
		Número de Servidores			Número de Servidores			Número de Servidores		
		8	10	12	8	10	12	8	10	12
0,90	0,10	73,71	80,76	87,89	73,99	85,69	91,60	96,00	99,10	100,00
	0,20	46,76	65,55	65,06	50,63	65,35	71,55	87,77	92,34	95,60
	0,30	40,28	55,03	59,11	42,19	53,97	63,68	80,27	88,75	93,19
	0,40	36,77	44,31	43,46	35,34	31,63	43,86	74,07	80,31	83,73
	0,50	17,08	18,63	36,04	24,95	29,23	35,79	67,02	73,62	76,36
0,95	0,10	55,32	59,48	67,31	54,38	67,22	70,28	87,61	93,23	93,56
	0,20	37,95	60,05	59,89	55,73	60,42	66,82	79,62	89,44	93,44
	0,30	32,08	30,70	39,26	33,63	39,62	45,82	75,66	79,74	83,65
	0,40	17,57	22,14	35,63	19,77	26,82	30,86	66,20	72,60	75,91
	0,50	9,95	18,96	20,06	15,00	21,32	26,58	58,62	63,11	65,31

Tabela 6 – Cobertura atingida com a heurística para o PLMDE.
Rede de 150 vértices de Galvão & ReVelle (1996)

α	ρ	$S = 80$			$S = 100$			$S = 120$		
		Número de Servidores			Número de Servidores			Número de Servidores		
		8	10	12	8	10	12	8	10	12
0,90	0,10	52,48	62,28	67,36	97,99	99,48	100,00	100,00	100,00	100,00
	0,20	28,89	34,55	44,13	89,92	96,78	97,54	98,57	100,00	100,00
	0,30	29,11	34,71	40,09	87,26	93,35	96,80	95,74	100,00	100,00
	0,40	24,75	22,79	29,20	76,49	84,67	88,30	89,80	94,72	98,68
	0,50	17,86	18,44	18,06	73,32	81,64	83,35	82,08	89,80	95,41
0,95	0,10	30,88	43,11	42,31	89,92	96,78	99,31	98,57	100,00	100,00
	0,20	29,92	33,85	38,68	86,81	94,69	97,74	97,93	99,35	100,00
	0,30	24,75	25,36	30,56	79,95	87,45	90,44	88,29	94,81	97,47
	0,40	17,86	24,47	20,88	73,44	76,84	83,17	82,08	88,65	94,89
	0,50	14,25	18,83	18,83	63,52	68,73	73,45	69,04	72,83	86,12

Comportamento das taxas de ocupação dos servidores para um conjunto de problemas de 100 vértices

Uma das características principais do *PLMDE* está na possibilidade de se trabalhar com taxas de ocupação diferenciadas por servidor. Nesta seção analisamos o comportamento das taxas de ocupação dos servidores para um conjunto de 30 instâncias da rede de 100 vértices de Galvão & ReVelle (1996), com o número de servidores fixado em 8. Na Figura 1 pode-se visualizar o grau de variabilidade ($\rho_k - \rho$) das taxas de ocupação dos servidores, para quatro problemas da rede de 100 vértices. Os números indicados no eixo horizontal dessas figuras referem-se aos vértices onde os servidores estão localizados. Quando o valor da taxa de

ocupação do sistema é baixo ($\rho = 0,10$), pode-se assumir que há pouca cooperação entre os servidores; os resultados indicam no entanto que existe heterogeneidade nas taxas de ocupação dos mesmos. Quando o valor da taxa de ocupação do sistema é elevado (por exemplo $\rho = 0,50$) assume-se que há muita cooperação entre os servidores; os servidores no entanto ainda funcionam com taxas de ocupação diferenciadas. Observe-se que inclusive, para servidores localizados no mesmo local, as taxas de ocupação são diferenciadas. Por exemplo, no problema com $\alpha = 0,90$, $\rho = 0,50$ e $S = 90$ foram localizados dois servidores no local 17, com taxas de ocupação bastante diferenciadas: 0,493 para um deles e 0,329 para o outro.

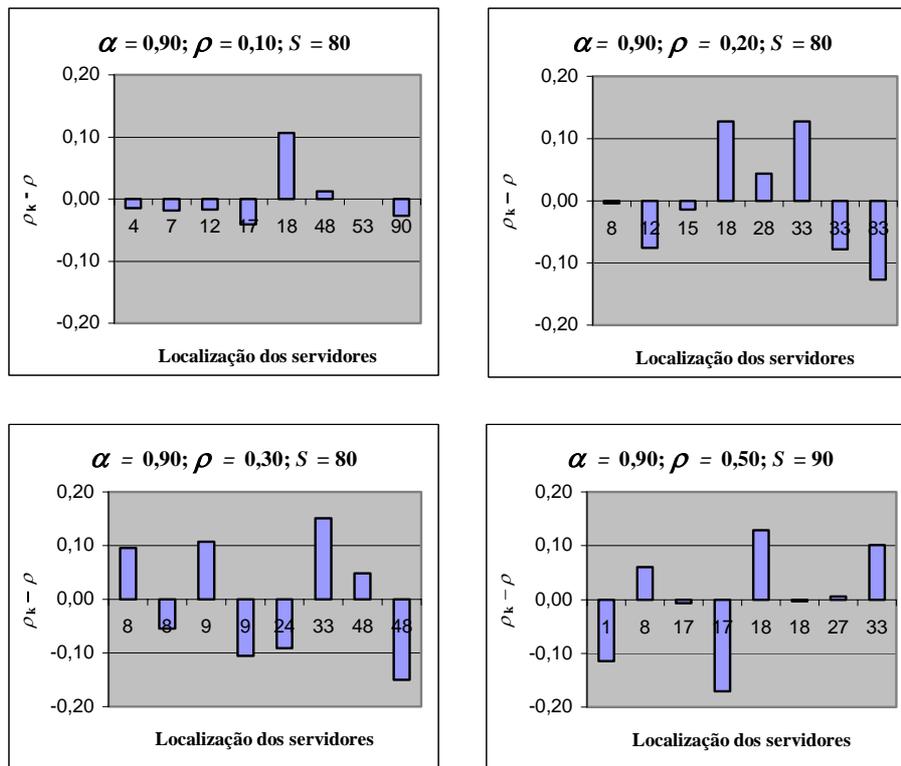


Figura 1 – Taxa de ocupação de 8 servidores: desvio em relação à taxa de ocupação do sistema. Rede de 100 vértices de Galvão & ReVelle (1996)

O desvio padrão das taxas de ocupação dos servidores pode ser usado para evidenciar o grau de heterogeneidade dessas taxas. A Tabela 7 mostra o desvio padrão das taxas de ocupação para o conjunto dos 30 problemas da rede de 100 vértices que estamos analisando. Pode-se observar nessa tabela que a heterogeneidade das taxas de ocupação tende a aumentar quando se aumenta o nível de confiabilidade (α) e a taxa de ocupação do sistema (ρ). A heterogeneidade das taxas de ocupação tende a diminuir quando se aumenta a distância crítica.

Tabela 7 – Desvio padrão das taxas de ocupação de 8 servidores.
Conjunto de 30 problemas da rede de 100 vértices de Galvão & ReVelle (1996)

α	ρ	Distância Crítica (S)		
		70	80	90
0,90	0,10	0,054	0,046	0,027
	0,20	0,083	0,095	0,099
	0,30	0,098	0,114	0,063
	0,40	0,114	0,113	0,140
	0,50	0,076	0,122	0,102
0,95	0,10	0,073	0,057	0,057
	0,20	0,086	0,100	0,054
	0,30	0,132	0,122	0,104
	0,40	0,127	0,111	0,113
	0,50	0,125	0,138	0,109

4.4 Resultados Disponíveis na Literatura

Não há na literatura resultados disponíveis para o *PLMDE* com os quais pudéssemos comparar os resultados obtidos com nosso algoritmo pois, segundo nos consta, este é o primeiro artigo em que o modelo hipercubo é utilizado para resolver o *Problema de Localização de Máxima Disponibilidade* definido por ReVelle & Hogan (1989), o que permite a utilização de taxas de ocupação calculadas por servidor. Por outro lado, existem poucos resultados computacionais disponíveis para os modelos *PLMDI* e *PLMDII* definidos pelos dois autores acima.

Em seu trabalho (1989) esses autores apresentam alguns resultados comparativos da aplicação dos modelos *PLMDI* e *PLMDII* objetivando a localização de ambulâncias em 31 estações do Corpo de Bombeiros na cidade de Baltimore, nos Estados Unidos. Os dois modelos são comparados em termos da previsão que fazem da cobertura obtida pela população, em função do nível de recursos disponíveis e da distribuição espacial das localizações escolhidas pelos modelos, para diferentes níveis de confiabilidade. Conforme esperado o *PLMDII* fornece resultados mais precisos, possibilitando por exemplo o uso de um número menor de ambulâncias para o mesmo nível de cobertura, para um dado nível de confiabilidade escolhido.

Seria interessante poder comparar os resultados obtidos por ReVelle & Hogan (1989) para a cidade de Baltimore com os que seriam obtidos utilizando taxas de ocupação calculadas por servidor, utilizando o algoritmo por nós desenvolvido para o *PLMDE*. Mesmo que todos os dados do estudo original estivessem disponíveis (*átomos* ou regiões em que a cidade foi dividida, taxa de chegada de pedidos por serviço de ambulância para cada *átomo*, tempos de serviço por ambulância, entre outros), tal comparação mostrar-se-ia inviável pela indisponibilidade dos dados adicionais que seriam necessários para a utilização do modelo hipercubo (dados esses não necessários para aplicação de modelos simplificados como o *PLMDI* e o *PLMDII*), como por exemplo a lista de preferências de despacho dos servidores.

5. Conclusões

O problema de localização de máxima disponibilidade desenvolvido por ReVelle & Hogan (1989) foi estendido neste trabalho para o caso em que as taxas de ocupação são calculadas individualmente, por servidor. O modelo hipercubo, embutido em uma heurística de substituição de vértices, produziu resultados em tempo computacional reduzido para redes de até 150 vértices disponíveis na literatura.

Os resultados computacionais obtidos com a heurística de substituição de vértices em geral confirmam o comportamento que seria esperado de ρ , taxa média de ocupação do sistema (e dos ρ_k 's, taxas individuais de cada servidor), em relação aos diversos parâmetros considerados (m , α , S). O uso de metaheurísticas na solução do *PLMDE*, e a comparação dos resultados obtidos com os correspondentes ao modelo de máxima cobertura esperada de Daskin (1983), são extensões possíveis do presente trabalho.

Agradecimentos

Esta pesquisa foi financiada pelo CNPq (Auxílio CNPq 471919/01-2) e pela FAPERJ (Auxílio FAPERJ E26/151.934/2000).

Referências Bibliográficas

- (1) Batta, R.; Dolan, J.M. & Krishnamurthy, N.N. (1989). The maximal expected covering location problem: Revisited. *Transportation Science*, **23**, 277-287.
- (2) Chiyoshi, F.Y.; Galvão, R.D. & Morabito, R. (2000). O uso do modelo hipercubo na solução de problemas de localização probabilísticos. *Gestão & Produção*, **7**, 146-174.
- (3) Church, R. & ReVelle, C.S. (1974). The maximal covering location problem. *Papers of the Regional Science Association*, **32**, 101-118.
- (4) Daskin, M.S. (1982). Application of an expected covering model to emergency medical service system design. *Decision Sciences*, **13**, 416-439.
- (5) Daskin, M.S. (1983). A maximum expected covering location model: formulation, properties and heuristic solution. *Transportation Science*, **17**, 48-70.
- (6) Espejo, L.G.A.; Galvão, R.D. & Boffey, B. (2003). Dual-based heuristics for a hierarchical covering location problem. *Computers & Operations Research*, **30**, 165-180.
- (7) Galvão, R.D.; Espejo, L.G.A. & Boffey, B. (2002). A hierarchical model for the location of perinatal facilities in the municipality of Rio de Janeiro. *European Journal of Operational Research*, **138**, 495-517.
- (8) Galvão, R.D. & ReVelle, C.S. (1996). A Lagrangean heuristic for the maximal covering location problem. *European Journal of Operational Research*, **88**, 114-123.

- (9) Hansen, P. & Mladenovic, N. (1998). Variable neighborhood search for the p -median. *Location Science*, **5**, 207-226.
- (10) Larson, R.C. (1974). A hypercube queuing model for facility location and redistricting in urban emergency services. *Computers & Operations Research*, **1**, 67-95.
- (11) Larson, R.C. (1975). Approximating the performance of urban emergency service systems. *Operations Research*, **23**, 845-868.
- (12) Owen, S.H. & Daskin, M.S. (1998). Strategic facility location: A review. *European Journal of Operational Research*, **111**, 423-447.
- (13) ReVelle, C.S. & Hogan, K. (1989). The maximum availability location problem. *Transportation Science*, **23**, 192-200.
- (14) Swain, R. (1971). A decomposition algorithm for a class of facility location problems. *Ph.D. Thesis*, Cornell University, Ithaca, NY.
- (15) Swersey, A.J. (1994). The deployment of police, fire and emergency medical units. **In: Handbooks in OR & MS** [edited by Pollock, S.M. *et al.*], Elsevier Science B. V., **6**, 151-200.
- (16) Teitz, M.B. & Bart, P. (1968). Heuristic methods for estimating the generalized vertex median of a weighted graph. *Operations Research*, **16**, 955-961.
- (17) Whitaker, R. (1983). A fast algorithm for the greedy interchange for large-scale clustering and median location problems. *INFOR*, **21**, 95-108.