Revista Produção On Line
Universidade Federal de Santa Catarina
www.producaoonline.inf.br
ISSN 1676 - 1901 / Vol. 4/ Num. 2/ Maio de 2003

METODOLOGIA PARA AVALIAR CENTRO DE EMERGÊNCIA:
Aplicação ao Centro de Emergência da Polícia Militar de Santa Catarina.

Leise Kelli de Oliveira
Pontifícia Universidade Católica do Paraná
Mirian Buss Gonçalves
Universidade Federal de Santa Catarina – UFSC
Departamento de Matemática

RESUMO

Este artigo apresenta o Modelo Hipercubo de Filas, uma metodologia de avaliação que fornece indicadores de desempenho de sistemas emergenciais, auxiliar na tomada de decisões estratégicas e operacionais. Esta metodologia é aplicada ao Centro de Emergência da Polícia Militar de Santa Catarina, localizado em Florianópolis, buscando validar as hipóteses de aplicação do modelo.

ABSTRACT

This article presents the Hypercube Queueing Model, a methodology of the evaluation, which supplies performance indicators of emergency systems, its helps in strategic and operations decision-making. This methodology is apply in the Emergency Center of the Santa Catarina Military Police located in Florianópolis, its search validate the hypothesis of the application of the model.

1. INTRODUÇÃO

Segundo Takeda (2000), quando o problema consiste em quantas unidades de serviço alocar em uma região e onde localizar estas unidades, este se torna complexo por incorporar características individuais destas unidades, como por exemplo, seu estado. Uma maneira de resolver problemas que apresentam estas características é utilizar o Modelo Hipercubo de Filas desenvolvido por Larson (1974) que conserva a identidade das unidades de serviço.

Data de Submissão: Set/03 Data de Aprovação: Abr/04

O modelo hipercubo apresenta-se como uma importante ferramenta para o planejamento de serviços e possibilita a utilização de teoria das filas em modelos de localização probabilísticos, permitindo o cálculo de medidas de desempenho que caracterizam um dado sistema com base em filas espacialmente distribuídas. A solução do modelo implica em resolver um sistema de equações lineares que fornece as probabilidades de equilíbrio dos possíveis estados do sistema (probabilidade de estado). Este artigo apresenta a modelagem do Modelo Hipercubo de Filas e resultados da aplicação do mesmo para o Centro de Emergência da Polícia Militar de Santa Catarina.

2. Modelo Hipercubo de Filas

A escolha de um modelo de fila é motivada pela representação do sistema espacial urbano, ilustrando métodos e aproximações que envolvem cenários espaciais e os modelos de avaliação são utilizados há muito tempo, buscando justificar as medidas de desempenho dos sistemas, criando cenários alternativos, se necessário, para melhorar as medidas de desempenho.

Larson (1974) formulou o Modelo Hipercubo de Filas para este fim, que constitui um modelo de avaliação e baseia-se na divisão da região em um conjunto de áreas de demanda ou *átomos geográficos*, onde cada átomo representa, ao longo do tempo, uma fonte independente solicitadora de serviços. O modelo considera as chamadas com distribuição espacial e temporal pela região.

Tal modelo calcula, a partir da distribuição de probabilidade dos estados do sistema, medidas de desempenho do mesmo. Este foi aplicado originalmente para o patrulhamento policial de Nova York, apresentando-se como importante ferramenta para localização de ambulâncias, como se pode verificar em Brandeau e Larson (1986) e Takeda (2000).

Segundo Larson e Odoni (1981), o modelo original tem as seguintes conjecturas:

1. A região é dividida em sub-áreas ou átomos (nós de um grafo);

2. De cada átomo, ocorrem chamadas para atendimento, segundo um processo de Poisson;
3. O tempo de atendimento é exponencial, com taxa média igual em todos os servidores;
4. Exatamente uma unidade é despachada.

O modelo hipercubo foi estudado por diversos autores pois se trata de uma ferramenta específica para o planejamento e avaliação de sistemas de atendimento que possuam demanda aleatória espacialmente distribuída.

Chelst e Jarvis (1979) modificaram este modelo para calcular a distribuição de probabilidade do tempo de viagem. Larson e Mcknew (1982) expandiram o modelo original para chamadas em que, inicialmente, não são atribuídas patrulhas para despacho.

A partir do Modelo Hipercubo, foram desenvolvidos modelos alternativos, denominados de híbridos, que buscam, também, otimizar a localização das unidades de serviços emergenciais com a preocupação de atender à demanda. Destacam-se o trabalho de Batta et al. (1989), que utilizaram o Modelo Hipercubo, juntamente com o Modelo de Cobertura Máxima Esperada proposto por Daskin (1983). Gonçalves et al. (1994) utilizaram o modelo hipercubo e também o modelo de p-mediana para a localização de unidades emergenciais em rodovias.

3. Considerações Iniciais do Modelo

O atendimento às chamadas é realizado por *servidores* tais como patrulhas, ambulâncias, veículos de entrega e bombeiros que se encontram distribuídos na região. Estes servidores podem estar fixos, posicionados em um ponto estratégico, ou móveis, sendo o caso do patrulhamento da polícia, onde a localização é obtida pela probabilidade do carro de patrulha estar localizado em um dos átomos geográficos. Importante ressaltar que a área de patrulhamento não precisa necessariamente coincidir com a área principal de resposta.

Para um servidor, *a área de cobertura primária* é composta pelos átomos aos quais o servidor, quando disponível, tem preferência para despacho, podendo esta área possuir qualquer formato, inclusive formas não adjacentes.

O nome *hipercubo* deriva do espaço de estados que descreve a disponibilidade dos servidores. Cada servidor pode apresentar-se em dois estados: *livre* ou *ocupado*. Um estado particular do sistema é dado por toda lista de servidores que estão ocupados ou livres. Como exemplo, o estado 110 corresponde ao sistema de 3 servidores em que a unidade 1 está livre e as unidades 2 e 3 estão ocupadas. Observe que a configuração 110 descreve os estados dos servidores da direita para a esquerda. O espaço do estado para um sistema com três servidores é dado pelos vértices de um cubo. Para um número de servidores maior que três, a noção de cubo é transportada para o espaço n-dimensional, chamado hiper-espaço, e assim o nome hipercubo. Este modelo admite formação de filas de pedidos solicitados, podendo o espaço de estado ser acrescido de uma *cauda infinita*. A Figura 1 mostra o espaço de estados para sistemas com três servidores.

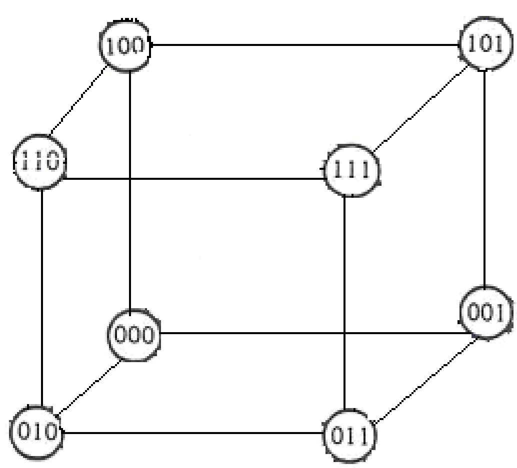

Figura 1 – Espaço dos estados para sistemas com três servidores

O modelo trata sistemas em que não é permitida a formação de fila, quanto de sistemas em que as chamadas quando chegam e não há servidor disponível, aguardam para receber atendimento. Neste caso, o espaço de estados deve ser acrescido de uma cauda junto ao estado totalmente ocupado, representando os usuários que aguardam serviço. Estes usuários são atendidos na medida em que os servidores tornam-se livres, segundo disciplina FCFS (*first-come, first-served*). A Figura 2 descreve o espaço de estados para um sistema com 3 servidores, considerando a formação de fila de espera.

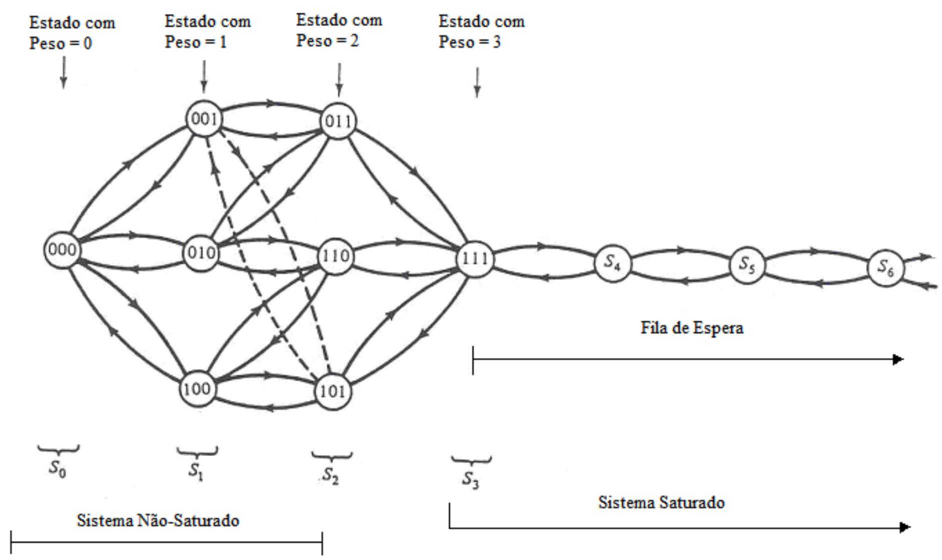

Figura 2 – Espaço de estados para sistemas com três servidores e cauda infinita

4. Hipóteses do Modelo

Segundo Larson e Odoni (1981), o modelo apresenta nove hipóteses que devem ser verificadas para a validação do modelo.

1. ***Átomos Geográficos:*** A área em que o sistema presta o serviço pode ser dividida em N_A átomos geográficos, sendo cada átomo modelado como um único ponto localizado próximo ao centro do átomo.

2. ***Processo de chegada como processo de Poisson independente:*** As chamadas são geradas de forma independentemente em cada átomo *j* segundo processo de Poisson, com taxa média λ_j (j=1, 2, ..., N_A) constantes no tempo, possíveis de medição ou estimação. Embora esta hipótese parece restritiva, é comum ser satisfeita em vários sistemas reais.

3. ***Tempo de viagem:*** Os tempos médios de viagem τ_{ij} entre o átomo *i* até o átomo *j* (i,j=1, 2,...,N_A) deverão ser conhecidos. Caso estes tempos não sejam conhecidos, eles devem ser estimados através de conceitos de probabilidade geométrica.

4. ***Servidores:*** O sistema é composto por N servidores espacialmente distribuídos ao longo da região, que podem se deslocar e atender qualquer um dos átomos. Em certos casos, esta hipótese pode ser relaxada para representar políticas de despacho particulares. Mendonça e Morabito (1999) apresentam uma aplicação para esta hipótese relaxada.

5. ***Localização dos Servidores:*** A localização de cada servidor, quando disponível, é conhecida. Para servidores fixos, a localização é a base; no caso de servidores móveis dentro de uma determinada área (por exemplo, viaturas de patrulhamento) a localização é conhecida de forma probabilística.

6. ***Despacho de um Servidor:*** Em resposta a cada chamada, exatamente uma unidade de resposta é despachada para o local da solicitação. O modelo não representa de maneira adequada, situações em que mais de uma unidade é despachada para a mesma chamada. Caso um conjunto de servidores seja despachado para atender uma catástrofe tem-se o conjunto como um único servidor. Se não houver unidades disponíveis para realizar o atendimento, a chamada poderá entrar em fila em sistemas que permitem a formação de filas ou perder a chamada nos casos em que não se admite fila. Para o último caso, as chamadas poderão ser transferidas para outro sistema de atendimento.

7. ***Política de despacho de servidores:*** O modelo pode operar com qualquer política de despacho de preferência fixa, havendo uma lista da referida preferência para cada átomo. Se o primeiro servidor desta lista estiver disponível, ele é despachado para atender a chamada no átomo, caso contrário o próximo servidor disponível na lista é despachado (backup). A lista de preferência é fixada inicialmente e permanece inalterada durante a operação do sistema. Podem ocorrer adaptações para casos em que um átomo tem mais de

um servidor, podendo ser feito através da introdução da distribuição de freqüência de despachos de cada servidor para cada átomo nas equações de balanço do sistema.

8. ***Tempo de Atendimento:*** O tempo total de atendimento de uma chamada inclui o tempo de preparo do servidor, tempo de viagem até o local da ocorrência, tempo de execução do serviço junto ao usuário e o tempo de retorno à base. Os servidores podem ter, em geral, um tempo médio de atendimento, μ_n (n=1,2,...,N) distintos. Para sistemas que permitem formação de filas, o modelo torna-se mais eficiente na medida em que os tempos médios de atendimento aproximam-se dos respectivos desvios padrões, isto é, dos tempos de atendimento exponencialmente distribuídos.

9. ***Dependência do Tempo de Atendimento em relação ao Tempo de Viagem:*** Variações no tempo de atendimento causadas por variações no tempo de viagem são de ordem secundária, quando comparadas com as variações de tempo de execução e/ou tempo de preparação. Esta hipótese é geralmente satisfeita nos serviços urbanos, porém limita a aplicabilidade do modelo.

5. Transição entre Estados

Considere os servidores como patrulhas, por exemplo. No modelo hipercubo, as transições entre estados ocorrem de modo idêntico às transições de outros modelos de fila admitindo que apenas um servidor seja designado para atender uma chamada. Admite-se como nula, a probabilidade de chegar duas chamadas simultâneas, como também é impossível dois servidores ocupados tornarem-se livres simultaneamente. Em síntese, qualquer transição de um passo é permitida e não são permitidas as transições de mais de um passo.

Dado que todas as chamadas são imediatamente atendidas existindo pelo menos um servidor disponível, todos os estados que estão a um servidor do estado de saturação devem apresentar taxa de transição ascendente igual a taxa global de chamadas da região atendida pelo sistema, λ. Assim, todas as transições ascendentes para o estado de saturação devem ser iguais a λ.

As taxas das transições ascendentes restantes são dadas pela soma da taxa de chamadas da área de cobertura primária do servidor que fica ocupado com as taxas de chamadas dos átomos que têm o servidor em questão com primeiro backup e cujo servidor primário esteja ocupado.

Considerando um sistema com 3 servidores, os possíveis estados em que o sistema pode se encontrar, em dado instante de tempo são: (000), (001), (010), (100), (011), (101), (110), (111), (S_4), (S_5), (S_6), ..., onde S_i, $i \geq 4$ corresponde ao estado em que *i* usuários estão no sistema aguardando atendimento. Tais estados podem ser representados em forma de diagrama, como um cubo para o caso de três servidores, onde os vértices correspondem aos estados, possuindo uma cauda infinita a partir do estado (111).

6. Equações de Equilíbrio

A solução do modelo é dada a partir da construção das equações de equilíbrio para o sistema, e os resultados baseiam-se nos valores das probabilidades de estado dos modelos clássicos de filas. As equações de equilíbrio são definidas supondo-se que o sistema atinja equilíbrio. Para cada estado do sistema, o fluxo que entra neste estado deve ser igual ao fluxo que sai do estado.

Para exemplificar as equações de equilíbrio, considere o sistema com três átomos e 3 servidores, onde os servidores 1, 2 e 3 atendem, respectivamente, os átomos 1, 2 e 3. Iniciando com o estado (000) em que todos os servidores estão livres, tem-se que o sistema passa deste para (001) quando ocorre uma chamada originada do átomo 1. A taxa de ocorrência deste evento é λ_1. O mesmo ocorre para os outros estados. Em conseqüência, a taxa total de transição do estado (000) para outros estados será $\lambda=\lambda_1+\lambda_2+\lambda_3$.

Considerando as transições no sentido inverso, tem-se em primeiro lugar, que o estado (000) pode ser alcançado a partir do estado (001) quando ocorre a conclusão do atendimento que o servidor 1 está realizando, evento que ocorre com taxa μ_1. O mesmo podendo ocorrer com os outros estados. Assim, a equação de equilíbrio em torno deste estado será:

(000) $\lambda P_{000} = \mu_1 P_{001} + \mu_2 P_{010} + \mu_3 P_{100}$, (1)

onde p_i é a probabilidade de equilíbrio do estado i, $i = \{(000), (001), ..., (111)\}$.

Como o modelo admite que apenas um servidor seja alocado para atender cada chamada (hipótese 6) e que sendo impossível à chegada de duas chamadas simultâneas (hipótese 2), nunca teremos a transição do estado (000) para o estado (011). Da mesma forma, dado que os servidores operam independentemente sendo impossível que dois servidores terminem seus serviços simultaneamente (hipóteses 6 e 8), não teremos transição do estado (011) para o estado (000).

Quando o sistema encontra-se no estado (001), além do fato de que qualquer chamada levará o sistema para fora deste estado, é possível a transição para o estado (000) via conclusão do atendimento do servidor 1, de modo que a taxa total de transição do estado (001) para outros estados será de $\lambda+\mu_1$ As transições de outros estados para o estado (001) são:

- do estado (000) pela chegada de uma chamada do átomo 1;
- do estado (011) pela conclusão do atendimento do servidor 2;
- do estado (101) pela conclusão do atendimento do servidor 3.

Estas possíveis transições levam à equação:

(001) $(\lambda+\mu_1)P_{001} = \lambda_1 P_{000} + \mu_2 P_{011} + \mu_3 P_{101.}$ (2)

As equações relativas aos estados (010) e (100) têm derivações e estruturas análogas:

(010) $(\lambda+\mu_2)P_{010} = \lambda_2 P_{000} + \mu_1 P_{011} + \mu_3 P_{110}$ (3)

(100) $(\lambda+\mu_3)P_{100} = \lambda_3 P_{000} + \mu_1 P_{101} + \mu_2 P_{110.}$ (4)

Considerado o sistema no estado (011), além do fato de que qualquer chamada levará o sistema para estado (111), é também possível à transição para o estado (001) e o estado (010), onde os servidores terminam atendimento e tornam-se disponíveis. A taxa total de transição do estado (011) para outros estados será de $\lambda+\mu 1+\mu 2$. Transições de outros estados para o estado (011) também são possíveis, como para o estado (001) onde a chegada de uma chamada do átomo 2 provoca tal transição. Outra possível transição para o estado (011) é a partir do estado (010), pela chegada de uma chamada originada no átomo 1, e, a partir do estado (111), pela conclusão do atendimento do servidor 3. Tudo considerado, a equação para este estado é a seguinte:

(011) $(\lambda+\mu_1+\mu_2)P_{011} = (\lambda_1+\lambda_2)P_{001} + \lambda_1 P_{010} + \mu_3 P_{111}$ (5)

De forma análoga, tem-se:

(101) $(\lambda+\mu_3+\mu_1)P_{101} = \lambda_3 P_{001} + (\lambda_1+\lambda_3)P_{100} + \mu_2 P_{111}$ (6)

(110) $(\lambda+\mu_3+\mu_2)P_{110} = \lambda_3 P_{010} + \lambda_2 P_{100} + \mu 1 P_{111}$ (7)

Chega-se finalmente ao estado (111) em que todos os servidores estão ocupados e não há nenhuma chamada esperando atendimento. Neste estado, qualquer chegada ou conclusão de atendimento provocará a transição do sistema para fora deste estado. Tem-se a equação:

(111) $(\lambda+\mu)P_{111} = \lambda P_{011} + \lambda P_{101} + \lambda P_{110}$ (8)

As equações de equilíbrio (1), (2), (3), (4), (5), (6), (7), (8) representam um sistema finito de equações lineares em termos das variáveis p_{000}, $p_{001,}$ p_{010}, p_{100}, p_{011}, p_{101}, p_{110}, p_{111}.

7. Insuficiência das Equações de Equilíbrio

Observando o sistema de equações (1)-(8) na forma Ax=b, percebe-se que estas conduzem a um sistema com b = 0. Tal sistema é homogêneo com solução trivial para $P_{ijk} = 0$, com i, j, k =0,1. Além do mais, tal sistema é indeterminado. A forma natural de levantar a indeterminação é introduzir uma equação adicional de normalização, com a seguinte forma:

$$P_{000} + P_{001} + P_{010} + P_{100} + P_{011} + P_{101} + P_{110} + \frac{P_{111}}{1-\rho} = 1 \qquad (9)$$

Assim, o conjunto das equações de equilíbrio forma um sistema de 2^N equações lineares, que pode ser resolvido com algum pacote matemático disponível, obtendo as probabilidades de cada estado do sistema, que são variáveis das equações.

8. Desempenho do Sistema

Com a distribuição de equilíbrio de estados, podemos computar diversas medidas de desempenho interessantes para o gerenciamento do sistema, como carga média de trabalho, freqüência de despacho, tempos de viagem, dentre outros. Mais informações sobre medidas de desempenho de sistema pode ser obtidas em Larson e Odoni (1981).

9. Aplicação do Modelo e Resultados

O Modelo Hipercubo de Filas foi utilizado como ferramenta para avaliar o Centro de Emergência da Polícia Militar de Santa Catarina, localizado na Ilha de Santa Catarina. O Centro de Emergência consiste em uma central de chamadas, através do número 190, com vários operadores que atendem ligações provenientes do 4º Batalhão da Polícia Militar.

A recepção das chamadas ocorre pelos números 190/193 e são identificadas através de um equipamento de identificação de chamadas, que se encontra conectado a um computador em rede. De posse do telefone identificado, o sistema busca em sua base de dados, a localização precisa do local da ocorrência, bem como, listará os meios disponíveis para o atendimento. De posse da localização da chamada, é despachada a guarnição responsável pela região que solicitou o atendimento. Caso esta guarnição esteja ocupada, outra será despachada, levando-se em conta a proximidade geográfica.

O sistema proporciona acompanhamento completo ao atendimento da chamada, monitorando os horários e meios empregados durante o atendimento. Após o encerramento do atendimento, o comandante da guarnição passará ao sistema, informações complementares para finalizar o mesmo, gerando um boletim de ocorrência que será armazenado pelo sistema. É importante salientar que uma guarnição é um conjunto de instrumentos necessários a realizar o atendimento à chamada, sendo composto de uma viatura, policiais e armamento necessário. Estatísticas são elaboradas periodicamente com o objetivo de proporcionar aos comandantes destas organizações, condições de analisar os níveis operacionais, tanto no aspecto preventivo, quanto em situações de emergência.

Para este estudo, são consideradas as chamadas oriundas da 3º Companhia, do 4º Batalhão da Polícia Militar. Isto se deve ao fato de que cada Companhia, que compõe a Ilha de Santa Catarina, atua com certa independência, possuindo estrutura própria. Optou-se também pela 3º Companhia por esta apresentar pouca variação quanto ao número e a natureza das chamadas nos períodos de sazonalidade.

Os resultados da modelagem mostram que as atividades do Centro de Emergência da Polícia Militar de Santa Catarina, para a 3º Companhia do 4º Batalhão da Polícia Militar, podem ser representadas por esta ferramenta por produzir desvios poucos significativos dos valores gerados pelo modelo com relação aos dados reais do sistema. Cenários alternativos foram criados para análises e estes evidenciaram que existem outras configurações que poderiam ser adotadas para elevar o nível de serviço do sistema.

10. Conclusão

O Modelo Hipercubo de Filas é uma importante ferramenta para o planejamento e avaliação de sistemas de atendimento que possuem demanda aleatória espacialmente distribuída. Desta forma, esta ferramenta quando utilizada para analisar configurações que podem elevar o nível de serviço oferecido por sistemas, estabelece uma relação entre as características operacionais de serviço e os níveis de atendimento proporcionados aos usuários.

Da aplicação do modelo aos cenários alternativos, conclui-se que existem alternativas operacionais que podem elevar o nível do serviço oferecido como, por exemplo, o tempo de viagem dos cenários alternativos, que proporcionam um ganho em relação ao tempo de viagem do sistema em estudo, no que se refere ao deslocamento até o átomo. Levando-se em conta, ser este o tempo significativo para os gerentes do sistema, estas são alternativas a serem adotadas, sem custos, para proporcionar o policiamento ostensivo, assegurando a ordem pública.

Portanto, o Modelo Hipercubo de Filas é uma importante ferramenta para análise de sistemas emergenciais. Para futuros trabalhos, a comparação de método é um interessante exercício, de forma analisar se existem melhorias em relação ao nível de serviço. Outra área de pesquisa é associar a este modelo estocástico, um modelo de dimensionamento de zonas. Esta análise permite um ajuste entre o tamanho da zona e o número de viaturas, proporcionando um patrulhamento mais ostensivo.

11. REFERÊNCIAS BIBLIOGRÁFICAS

BATTA, R.; DOLAN, J.; KRISHNAMURTHY, N. (1989). The maximal expected covering location problem: Revisited. Transportation Science. v.23, 277-287.

BRANDEAU, M.; LARSON, R.C. (1986) Extending and applying the hypercube queuing model to deploy ambulances in Boston. In: SWERSEY, A.J. and INGALL, E.J., (eds) Delivery of Urban Services. **TIMS studies in the Management Science**. v.22, Elsevier, 121-153.

CHELST, K.R.; JARVIS, J.P. (1979). Estimating unit dispatches in emergency services: Models to estimate system performance. **Management Science**. v.27, n.1, 199-204.

DASKIN, M.S. (1983). A maximum expected covering location model: Formulation, properties and heuristic solution. **Transportation Science**. v.17, n.1, 48-70.

GONÇALVES, M. B.; NOVAES, A. G.; ALBINO, Jean C. (1994) **Modelo para localização de serviços emergenciais em rodovias.** XXVI Simpósio Brasileiro de Pesquisa Operacional, Florianópolis.591-601.

LARSON, R.C. (1974) Illustrative police sector redesign in district 4 in Boston. **Urban Analysis**. v.2, 51-91.

LARSON, R.; MCKNEW, M.A. (1982). Police patrol-initiated activities within a systems queuing model. **Management Science.** v.28, n.7, 759-774.

LARSON, R. C.; ODONI, A. R. (1981). **Urban Operations Research.** New Jersey, Prentice-Hall.

OLIVEIRA, L.K. (2003) **Uma aplicação do Modelo Hipercubop de Filas para Avaliação do Centro de Emergência da Polícia Militar de Santa Catarina.** Florianópolis. Dissertação (Mestrado): Universidade Federal de Santa Catarina.

TAKEDA, R. A. (2000) **Uma contribuição para avaliar o desempenho de sistemas de transportes de saúde.** Tese de Doutorado, Departamento de Engenharia de Transportes, Escola de Engenharia de São Carlos, USP.

Leise Kelli de Oliveira

leisekelli@yahoo.com.br

Mirian Buss Gonçalves

mirian@mtm.ufsc.br