

ARTIGO

UMA PROPOSTA ALTERNATIVA PARA AVALIAÇÃO DO DESEMPENHO DE SISTEMAS DE TRANSPORTE EMERGENCIAL DE SAÚDE BRASILEIROS

Renata Algisi Takeda

João Alexandre Widmer

Departamento de Transportes
Escola de Engenharia de São Carlos
Universidade de São Paulo

Reinaldo Morabito

Departamento de Engenharia de Produção
Universidade Federal de São Carlos

RESUMO

A rapidez na realização do atendimento às vítimas é uma das maiores exigências dos serviços de atendimento médico de urgência, e o tempo decorrido entre o instante da ocorrência da solicitação pelo serviço e o início do atendimento junto às vítimas é um dos principais fatores que influenciam o desempenho do sistema. Este tempo depende das condições do tráfego, dia e período do dia, número de veículos disponíveis e suas localizações, capacitação profissional da equipe, etc. Este trabalho apresenta uma análise do desempenho atual do serviço oferecido pela cidade de Campinas-SP, tratando o problema por meio do modelo hipercubo de filas, que considera as variações aleatórias dos processos de chegadas e atendimento dos chamados. Sua aplicação produz uma ampla variedade de indicadores de desempenho para o sistema, além de possibilitar, através da investigação de cenários alternativos, a busca de configurações operacionais com melhores níveis de serviço.

ABSTRACT

One of the major concerns in medical emergency systems is the quickness to accomplish the attendance to the victims. Its performance is mainly influenced by the time elapsed between the emergency call and the rescue. This time depends on the traffic features, the day and the hour, the number of available vehicles and their locations, professional team training, etc. This paper presents a performance analysis of the current service offered in Campinas-SP, facing the problem through the hypercube queuing model, which considers stochastic variations of the arrivals and service processes. Its application produces a wide variety of system performance measures, besides making possible, through a careful investigation of alternative sceneries, the indication of a better operational configuration looking for the service level increase.

1. INTRODUÇÃO

A qualidade de vida da população está ligada a uma diversidade de serviços que podem ser classificados em três grandes grupos: serviços de rotina, serviços semi-emergenciais e serviços de emergência, dentre os quais destaca-se o Serviço de Atendimento Médico de Urgência (SAMU). Quando projetado e operado com eficiência, o SAMU pode salvar vítimas; por outro lado, quando ineficiente, é um potencial responsável pelo agravamento clínico dos casos.

A função básica de um SAMU é responder de forma organizada, a fim de evitar o uso excessivo de recursos, a toda situação de urgência que necessite de meios médicos, desde o primeiro contato telefônico até a liberação das vítimas ou seus encaminhamentos hospitalares. O sistema deve determinar e desencadear a resposta mais adequada para o caso, assegurar a disponibilidade dos meios hospitalares, determinar o tipo de transporte exigido e preparar o acolhimento dos pacientes.

Serviços emergenciais como os SAMU's apresentam altos graus de incerteza, e normalmente suas eficiências são medidas através do

tempo médio de resposta a um chamado, ou seja, o tempo que uma vítima espera em média para começar a receber algum tipo de atendimento. Quanto maior o grau de incerteza envolvido e maior a necessidade de se obter respostas rápidas, menor deve ser a taxa de utilização dos operadores e equipamentos do sistema. Caso contrário, o nível de serviço oferecido pode ser deteriorado. Neste contexto, quando bem dimensionados, geralmente ocorrem longos períodos em que os operadores e equipamentos permanecem desocupados (Gonçalves, 1994).

Muitas pesquisas vêm sendo desenvolvidas no sentido de se obter métodos para analisar e dimensionar tais sistemas, de forma a elevar o nível de serviço oferecido e também racionalizar os recursos exigidos. No entanto, uma limitação dos estudos é que muitos deles não consideram a natureza probabilística dos processos de chegada e atendimento dos chamados, e não levam em conta o fato de que as ambulâncias nem sempre estão disponíveis para iniciar um atendimento. Dentre as importantes contribuições, destacam-se trabalhos onde a atenção é concentrada em questões tais como localização de bases, roteirização de veículos, zoneamento da área de atuação do sistema e problemas de congestionamento. A modelagem integrada destas questões é muito complexa, pois geralmente os sistemas reais são compostos por um grande número de veículos, as solicitações por serviço ocorrem temporária e espacialmente, existe cooperação entre veículos de áreas distintas, podem ocorrer múltiplos despachos para atender a um mesmo chamado, o tempo médio de viagem varia de acordo com a região, o dia e o período do dia, e existe a possibilidade de formação de filas de espera.

Alguns pesquisadores dedicaram-se exclusivamente a problemas de localização de bases (Toregas *et al.*, 1971, Anderson e Fontenot, 1992 e Louveaux, 1993), enquanto outros consideraram o problema de projeto das áreas de cobertura específicas (Keeney, 1972 e Larson, 1974). Daskin e Stern (1981) usaram problemas de cobertura com o objetivo de determinar o número necessário de veículos para cobrir cada região individualmente, e também para um conjunto de regiões vizinhas.

Outra maneira de se abordar o problema é por meio de modelos de simulação. Savas (1969) usou um modelo de simulação na cidade de Nova Iorque para mostrar que o tempo médio de resposta a um chamado pode ser reduzido redistribuindo as ambulâncias em suas bases, enquanto Swoveland *et al.* (1973) aplicaram um modelo de simulação para determinar o tempo médio de resposta das ambulâncias em toda a extensão de Vancouver e, usando o algoritmo *branch and bound*, determinou novas localizações e novas configurações das áreas de cobertura do sistema a fim de reduzir o tempo médio de resposta. Fitzsimmons (1973) desenvolveu um modelo baseado em teoria de filas para analisar a localização de ambulâncias na cidade de Los Angeles, que considera como principal fator o tempo médio de viagem de cada veículo a cada chamado.

No presente trabalho utilizou-se o *modelo hipercubo*, desenvolvido por Larson (1974) e estudado por diversos autores (Swersey, 1994), para analisar o desempenho atual do serviço oferecido pelo SAMU da cidade de Campinas-SP. Trata-se de uma ferramenta analítica e descritiva que permite calcular uma ampla variedade de medidas de desempenho, que auxiliam nas decisões operacionais e de configuração do sistema (Brandeau e Larson, 1986). O hipercubo não é um modelo de otimização que determina uma configuração ótima para o sistema, mas fornece uma completa avaliação de desempenho de cada configuração sugerida (Halpern, 1977).

A aplicação original do modelo foi para o problema de patrulhamento policial. Porém, sistemas como bombeiros, ambulâncias, defesa civil, reparos em redes de infra-estruturas básicas, guinchos e entregas domiciliares também podem ser bem representados por esta técnica. No Brasil, alguns exemplos importantes são: o atendimento a interrupções de energia elétrica em Florianópolis, SC (Albino, 1994), a localização de ambulâncias em um trecho da BR 111 - SC (Gonçalves *et al.*, 1994, 1995), e o balanceamento das *workloads* de ambulâncias no sistema Anjos do Asfalto da Rodovia Presidente Dutra (Mendonça e Morabito, 2000). Um exame recente do uso do modelo hipercubo na solução de problemas de localização probabilísticos foi apresentado em Chiyoshi *et al.* (2000).

2. UMA BREVE APRESENTAÇÃO DO MODELO HIPERCUBO

2.1. Descrição do sistema e utilização do modelo

Em um sistema de atendimento de emergência, as principais hipóteses são:

- os chamados chegam em instantes distintos, de diferentes partes de uma região, em uma central telefônica;
- caso exista um veículo disponível, ele é despachado para o local do evento; caso contrário, o primeiro a se tornar “livre” é alocado para realizar o atendimento, ou o chamado é transferido para outro sistema de atendimento;
- ao chegar junto às vítimas, a equipe despende um tempo no local do evento (tempo em cena), as transporta até um hospital (caso necessário), e então retorna a sua base, estando pronta para realizar um novo atendimento.

O modelo hipercubo considera a cidade particionada em um número finito de regiões, denominadas átomos geográficos. O analista coleta informações do processo de chegada das solicitações de cada átomo (intervalos de tempo entre chegadas), processo de atendimento de cada servidor (tempos de atendimento), e tempos de viagem entre todos os pares de átomos, para descrever estatisticamente as operações do sistema e realizar a calibração do modelo. Uma vez calibrado, o hipercubo pode ser usado para determinar indicadores de desempenho para diferentes configurações do sistema. Para cada configuração, o modelo calcula medidas de desempenho para: (i) a extensão total da região, (ii) um veículo específico, (iii) um subconjunto finito de átomos, e (iv) um átomo específico. Baseado nos resultados produzidos e nas exigências do sistema, o tomador de decisões pode pesquisar configurações alternativas que ofereçam níveis de serviço mais elevados aos usuários.

2.2. A matemática do modelo básico

A modelagem pressupõe que a área a ser estudada seja particionada em N_A regiões (átomos geográficos), cada qual gerando uma fração f_j do número total de solicitações de serviços $\left(\sum_{j=1}^{N_A} f_j = 1 \right)$. O tempo médio de viagem entre dois átomos i e j quaisquer é dado por τ_{ij} . Existem N veículos em operação no sistema, e a probabilidade de o veículo n estar localizado no átomo j é l_{nj} $\left(\sum_{j=1}^{N_A} l_{nj} = 1 \right)$. Admite-se que os chamados ocorram no sistema conforme um processo de Poisson, com taxa média λ chamados por unidade de tempo (por exemplo, hora), sendo cada átomo j um gerador de processos de Poisson independentes com taxas médias $\lambda_j = \lambda \cdot f_j$. Caso os servidores sejam homogêneos, com mesma taxa média μ de atendimentos por unidade de tempo, as medidas agregadas do modelo hipercubo equivalem às do modelo clássico $M/M/N$ (com ou sem possibilidade de formação de fila, isto é, $M/M/N/\infty$ ou $M/M/N/0$), sob as seguintes condições:

- apenas um veículo é alocado para atender um chamado;
- o tempo de atendimento de qualquer veículo tem distribuição exponencial negativa, com taxa média μ ;
- o tempo de atendimento é independente da identidade do servidor, da localização das vítimas e da história do sistema;
- para o caso $M/M/N/\infty$, todos os chamados que chegam enquanto os N veículos estão ocupados entram em fila, e são posteriormente atendidos segundo a disciplina *first-come, first-served* (FCFS);

- para o caso M/M/N/0, todos os chamados que chegam enquanto os N veículos estão ocupados são perdidos (ou transferidos para outro sistema de atendimento).

Para cada átomo, a política de despachos de veículos é fixada por meio de uma *ordem de preferências de despachos*, como mostra a tabela 1 (n_j indica o i -ésimo servidor escolhido para atender um chamado do átomo j).

Tabela 1: Matriz de preferências de despachos

Átomo	Veículo				
	1ª preferência	2ª preferência	3ª preferência	...	Nª preferência
⋮	⋮	⋮	⋮	⋮	⋮
j	n_{j1}	n_{j2}	n_{j3}	...	n_{jN}
⋮	⋮	⋮	⋮	⋮	⋮

Consideradas todas estas hipóteses, o sistema pode ser caracterizado como um processo de Markov em tempo contínuo, com $O(2^N)$ possíveis estados, para todas as combinações (livre/ocupado) admissíveis para os veículos. Cada estado do sistema é representado por uma seqüência de N 0's e 1's, onde, em cada posição n da seqüência, o valor 0 corresponde à condição "livre" (ou disponível) do servidor n , e 1 à condição "ocupado". Com isto, as probabilidades de equilíbrio de estado podem ser calculadas, analogamente às probabilidades de estado do modelo M/M/N, resolvendo-se um sistema linear com $O(2^N)$ equações e incógnitas, gerando uma grande variedade de indicadores de desempenho para o sistema.

2.3. Extensão do modelo básico: atendimentos com prioridades

Muitos sistemas reais contêm regras operacionais complexas, que nem sempre são facilmente descritas nos modelos analíticos, como é o caso de sistemas com múltiplas classes de usuários. Serviços que possuem atendimentos diferenciados quanto ao tipo de veículo, formação da equipe, ou até procedimentos tomados junto às vítimas, também podem ser representados pelo modelo hipercubo. Um exemplo é o SAMU de Campinas-SP, onde o sistema é composto por ambulâncias de quatro tipos distintos: VSA (*veículos de suporte avançado*, ou seja, as unidades de tratamento intensivo móveis), VSB

(veículos de suporte básico, semelhante ao veículo-resgate do corpo de bombeiros), VRS (veículos de remoção simples) e PSQ (veículos psiquiátricos).

Para modelar tal situação, Larson e Odoni (1981) recomendaram o “processo de camadas”, onde cada átomo é particionado em “sub-átomos”, um para cada tipo de chamado do átomo original, e então os veículos específicos para realizar o atendimento são considerados como primeiras prioridades de despacho para o átomo. Esta simples adaptação permite o cálculo de todos os indicadores de desempenho para cada classe de usuários e cada tipo de veículo, separadamente, e também para o sistema como um todo.

2.4. Indicadores de desempenho

Determinadas as probabilidades de equilíbrio, o modelo produz algumas medidas importantes para avaliar o comportamento do sistema para uma dada configuração operacional. Dentre elas:

- probabilidade de existência de fila: P_Q ;
- *workload* ou fator de utilização dos veículos: ρ_n , $n = 1, \dots, N$ (na prática, este valor representa a fração do tempo em que o veículo n está ocupado);
- frequência de despachos inter-átomos f_{nj} , isto é, fração de todos os despachos que enviam o veículo n ao átomo j , que pode ser decomposta em:

$$f_{nj} = f^{[1]_{nj}} + f^{[2]_{nj}} \quad (1)$$

onde

$f^{[1]_{nj}}$: fração dos despachos do veículo n ao átomo j , para atender a um chamado que não vem de fila, e

$f^{[2]_{nj}}$: fração dos despachos do veículo n ao átomo j , para atender a um chamado que vem de fila de espera;

- tempo médio de viagem para chamados sujeitos a esperas em fila:

$$\bar{T}_Q = \sum_{i=1}^{N_A} \sum_{j=1}^{N_A} \frac{\lambda_i \cdot \lambda_j}{\lambda^2} \cdot \tau_{ij} \quad (2)$$

- tempo médio de viagem do veículo n ao átomo j:

$$t_{nj} = \sum_{k=1}^{N_A} l_{nk} \cdot \tau_{kj} \quad (3)$$

- tempo médio de viagem para o átomo j:

$$\bar{T}_j = \frac{\sum_{n=1}^N f_{nj}^{[1]} \cdot t_{nj}}{\sum_{n=1}^N f_{nj}^{[1]}} \cdot (1 - P'_Q) + \sum_{i=1}^{N_A} \left(\frac{\lambda_i}{\lambda} \right) \cdot \tau_{ij} \cdot P'_Q \quad (4)$$

onde $P'_Q = P_Q + P_{11\dots 1}$ representa a probabilidade de saturação do sistema ($P_{11\dots 1}$ é a probabilidade do sistema estar com todos os servidores ocupados);

- tempo médio de viagem de cada veículo n:

$$\bar{TU}_n = \frac{\sum_{j=1}^{N_A} f_{nj}^{[1]} \cdot t_{nj} + (T_Q \cdot P'_Q / N)}{\sum_{j=1}^{N_A} f_{nj}^{[1]} + (P'_Q / N)} \quad (5)$$

3. APLICAÇÃO DO MODELO NO SAMU-192 DE CAMPINAS

Baseados no modelo francês de atendimento às urgências, os SAMU's começaram a ser implantados em algumas cidades brasileiras no início da década de 90. Funcionam como centros reguladores das urgências médicas, oferecendo suporte básico de vida, com veículos adequados para tal e operadores preparados para realizar o atendimento de forma segura, não colocando em risco as condições de vida dos seus usuários.

Muitos sistemas que hoje operam em cidades brasileiras surgiram a partir da parceria entre as Secretarias Municipais de Saúde e o Corpo de Bombeiros, com a chegada dos veículos RESGATE. Isto levou a um aumento considerável no nível de serviço oferecido, e os resultados positivos passaram a justificar novos recursos e investimentos no setor, além de resgatar a confiabilidade da população no tipo de atendimento. Atualmente, nota-se um crescente aumento no número de municípios que procuram adequar-se às modernas diretrizes de planejamento e operação do serviço de transporte por ambulância no país, visando alcançar níveis internacionais de qualidade no atendimento às urgências.

3.1. A pesquisa de campo

Dentre diversas cidades de pequeno, médio e grande porte consultadas, apenas oito se dispuseram a mostrar em detalhes os seus sistemas: Jaú, Rio Claro, Araraquara, São Carlos, Limeira, Piracicaba, Ribeirão Preto e Campinas. O principal objetivo desta primeira fase da pesquisa foi conhecer o ambiente e identificar os possíveis problemas a serem estudados, para construir um método de investigação eficaz, de acordo com a realidade envolvida.

Em todos os casos, observou-se a existência de uma base de dados significativa, com registros das principais informações a respeito da operação dos sistemas. Porém, interesses outros acabaram dificultando o acesso a algumas destas informações. Neste contexto, a cidade de Campinas, que dispunha de um sistema considerado modelo no Estado de São Paulo, interessou-se por participar do estudo, contribuindo de forma expressiva para o levantamento dos dados e a interpretação dos resultados obtidos posteriormente com a aplicação do modelo. A equipe de gerentes do SAMU-192 da Campinas julgava possuir um tempo médio de resposta aos chamados insatisfatório, se comparado aos padrões internacionais de excelência.

3.2. O cenário observado

Em 1998, o sistema operava com 18 ambulâncias, sendo 2 veículos de suporte avançado (VSA), 4 veículos de suporte básico (VSB), 11 veículos de remoção simples (VRS) e 1 veículo psiquiátrico (PSQ). Dentre os 11 VRS's, 7 eram dedicados às operações de remoções simples (pacientes agendados) e 4 operavam como sendo VSB's, totalizando 10 veículos dedicados, de fato, às operações de urgência. Portanto, para efeito de análise, foram considerados apenas estes 10 veículos, já que o caso de pacientes agendados e transferências psiquiátricas não são característicos de um sistema de transporte médico de urgência

Todos os veículos permaneciam centralizados na base operacional do sistema, quando disponíveis, local onde se encontra, até o presente momento, a central telefônica 192 para onde convergem todos os chamados. O sistema também permite que seus usuários entrem em uma fila de espera (limitada, em média, por 1 usuário por veículo em operação), caso todas as ambulâncias estejam ocupadas no instante da ocorrência do evento.

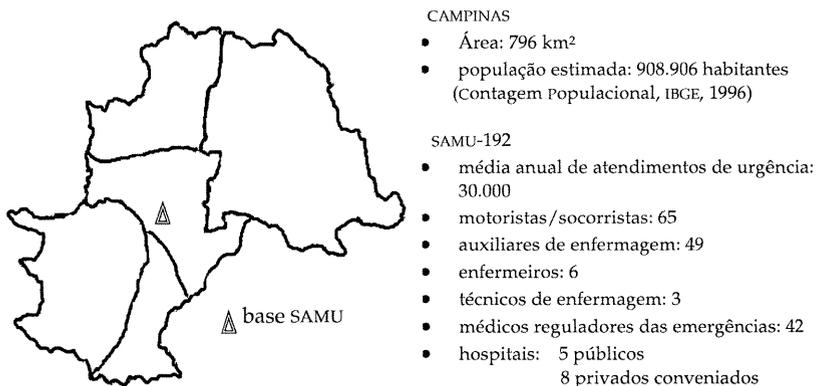


Figura 1: Atuação do SAMU-192 em Campinas-SP

A equipe gerenciadora estava preocupada com o crescimento dos índices de utilização das ambulâncias em determinados períodos do dia, bem como dos tempos de espera das vítimas. Surgiram

discussões sobre uma possível descentralização do serviço, ou seja, quanto uma localização estrategicamente descentralizada das ambulâncias poderia interferir no nível de serviço oferecido, e qual o número ideal de veículos para realizar a operação.

3.3. A coleta de dados

Para implementar o modelo hipercubo no SAMU-192 foi preciso levantar:

- número de regiões de atuação do sistema (átomos geográficos), intervalos de tempo entre chegadas dos chamados, com as respectivas taxas médias de chegadas por região;
- tempo médio de viagem entre todas as regiões;
- tempos de atendimento, com os respectivos tempos médios de atendimento por ambulância, e a fração correspondente do tempo junto ao paciente (tempo em cena);
- localização das ambulâncias.

3.3.1. Átomos geográficos

Há diversas maneiras de se representar os átomos geográficos de um sistema: através da divisão política da cidade, ou dos setores policiais, bairros, etc. Em Campinas, o SAMU obedece as regiões correspondentes às áreas de cobertura dos Centros de Saúde: Norte, Sul, Leste e Oeste, sendo também considerada uma região Central no entorno de sua base.

A coleta de dados determinou taxas médias de solicitação pelo serviço distintas ao longo das 24 horas de operação do sistema. Para fins de modelagem, foi considerado o período de 10 às 14 horas, que representa o período crítico para os gerentes do sistema. Neste período, observou-se uma taxa média $\lambda = 5,4$ chamados/h. Também

foram determinadas as taxas de cada área específica, sendo que as regiões periféricas e central apresentaram os maiores índices.

Com isto, pode-se verificar, através da aplicação do teste não-paramétrico *Kolmogorov-Smirnov*, a hipótese de que o processo de chegada de chamados no sistema, bem como os processos de chegada em cada região, seguem padrão poissoniano, com nível de significância de 5%. Esta verificação foi realizada a partir das observações dos intervalos de tempo entre chegadas de solicitações.

3.3.2. *Tempos de atendimento*

O tempo médio de atendimento observado no sistema foi de 65 minutos, sendo de 63 minutos para os veículos avançados e de 66 minutos para os veículos básicos. Estes valores foram determinados a partir das observações dos atendimentos realizados por todas as ambulâncias. Também foi verificado, através do teste de *Kolmogorov-Smirnov*, que os tempos de atendimento das ambulâncias são exponencialmente distribuídos, com nível de significância de 5%.

3.3.3. *Localização das ambulâncias*

Quando disponíveis, as ambulâncias permaneciam centralizadas na base do SAMU. Uma das propostas para descentralização das mesmas era localizá-las nos pólos geradores de demanda do sistema.

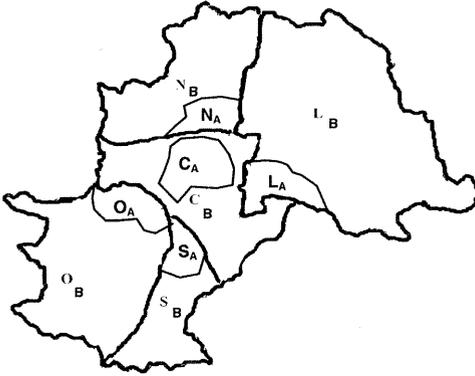
3.4. **Introduzindo classes de usuários na modelagem**

Como mencionado anteriormente, o SAMU-192 de Campinas possui classes diferenciadas de usuários. Neste estudo, foram consideradas apenas duas classes:

- básica – definida pelos chamados atendidos por VSB's;
- avançada – definida pelos chamados atendidos por VSA's.

As cinco regiões foram biparticionadas, gerando um total de 10 átomos geográficos no sistema (Norte B, Norte A, Sul B, Sul A, Leste

B, Leste A, Oeste B, Oeste A, Centro B, Centro A), conforme pode ser observado na figura 2.



Os veículos avançados foram alocados como primeiras preferências de despachos para os átomos Norte A, Sul A, Leste A, Oeste A e Centro A, e últimas opções, para os demais.

A classe avançada representa 10% do total de chamados de emergência do sistema

Figura 2: Redistribuição espacial do sistema

4. VALIDAÇÃO DA TÉCNICA DE MODELAGEM

O modelo foi implementado computacionalmente (linguagem Pascal, estação de trabalho IBM 3CT com 128 MB de memória e sistema operacional IBM AIX) e executado para as condições originais do sistema, ou seja, $N_A = 10$ átomos e $N = 10$ ambulâncias, todas centralizadas na base quando disponíveis.

Os resultados obtidos mostraram a eficiência do modelo hipercubo para avaliar o serviço de transporte médico-emergencial de Campinas, apresentando desvios pouco significativos com relação aos dados reais (desvio médio de 4,84%). A tabela 2 ilustra as *workloads* e os desvios com relação aos tempos médios de resposta de cada ambulância (outros indicadores de desempenho foram comparados, gerando desvios também pouco significativos).

Estes valores confirmam as observações iniciais de que as *workloads* são relativamente altas, em se tratando de um serviço emergencial. Uma das vantagens deste cenário, onde todos os veículos encontram-

se centralizados, é o fato de existir um bom balanceamento das *workloads* dentro de cada classe.

A partir de então, novos cenários foram avaliados com o objetivo de encontrar melhores alternativas operacionais para o sistema. A seguir, serão apresentados apenas dois destes cenários (os demais podem ser encontrados em Takeda, 2000).

Tabela 2: Resultados gerados pelo modelo hipercubo para a configuração original do sistema

			Tempo médio de resposta (minutos)		desvio	
	veículo	Workload	modelo	amostra	minutos	%
VSA	1	0,38	10,88	10,38	0,51	4,91
	2	0,39	10,90	10,70	0,20	1,89
	3	0,63	14,22	13,39	0,83	6,22
VSB	4	0,65	14,24	13,54	0,70	5,19
	5	0,62	14,23	14,30	-0,07	-0,49
	6	0,64	14,24	13,64	0,60	4,43
	7	0,63	14,23	13,17	1,06	8,09
	8	0,62	14,22	13,27	0,95	7,17
	9	0,64	14,24	13,23	1,01	7,61
	10	0,65	14,24	13,76	0,47	3,42

5. DESCENTRALIZAÇÃO DAS AMBULÂNCIAS

Membros da comunidade médica julgam o tempo médio de resposta como sendo o principal indicador do desempenho de um sistema emergencial. Para avaliar o impacto da descentralização dos veículos, foram realizados testes com cenários onde os veículos básicos foram descentralizados. A tabela 3 compara os tempos médios de resposta do sistema original (centralizado) com os de um sistema descentralizado, onde as regiões Norte B, Oeste B e Centro B tornaram-se bases para dois veículos, ficando as regiões Sul B e Leste B (com baixas taxas de solicitação pelo serviço) com apenas uma ambulância, e os veículos avançados ainda posicionados no Centro.

Baseados na equação (4), e calibrando os valores dos tempos médios de atendimento para esta nova configuração, os tempos médios de resposta diminuem sensivelmente quando é implantada a

descentralização das ambulâncias. Observa-se que isto não ocorreu com a região C_B devido ao fato de que esta é solicitada como primeiro *back-up* para todas as outras regiões, dado que a distância média percorrida entre o Centro e os demais átomos é um pouco menor do que entre quaisquer dois outros pares. Também observa-se que as diferenças entre os valores determinados para as regiões geradoras de chamados avançados não é muito significativa, visto que eles permanecem centralizados.

Tabela 3: Tempos médios de resposta para cada região

átomo	Tempo médio de resposta (minutos)		desvio	
	centralizado	descentralizado	Minutos	%
N_B	16,01	11,06	-4,96	-30,95
N_A	13,09	12,48	-0,61	-4,63
S_B	15,93	13,01	-2,92	-18,31
S_A	13,00	12,69	-0,31	-2,42
L_B	15,96	12,68	-3,28	-20,56
L_A	13,03	12,65	-0,38	-2,88
O_B	15,76	11,25	-4,51	-28,63
O_A	12,83	12,26	-0,57	-4,45
C_B	5,50	7,84	2,34	42,59
C_A	4,33	4,58	0,25	5,83

Outro cenário investigado foi com relação ao efeito causado pelo aumento do número de ambulâncias na operação diária. Incentivado pela tendência de redução do tempo médio de resposta a um chamado quando se tem um sistema descentralizado, avaliou-se os impactos causados ao se implantar duas novas ambulâncias em operação, ficando cada região geradora de chamados básicos com duas ambulâncias disponíveis. Os valores dos tempos médios de resposta encontrados diminuíram, em média, 30% com relação ao cenário original (aproximadamente 4 minutos), e em torno de 15% com relação ao cenário descentralizado para as oito ambulâncias básicas avaliado anteriormente. As *workloads* dos veículos básicos reduziram cerca de 30% com relação ao cenário original (de 0,64 para 0,45), resultando em uma sensível redução da probabilidade de um chamado ocorrer quando todas as ambulâncias estão ocupadas (igual a 0,12 no sistema original e 0,05 para o sistema remodelado com os dois novos veículos).

Convém salientar que os resultados aqui apresentados são ainda preliminares. Os resultados completos estão sendo compilados, junto com uma análise dos *trade-off* entre as várias configurações operacionais estudadas, e deverão ser reportados em breve.

6. CONSIDERAÇÕES FINAIS

O modelo hipercubo de filas, desenvolvido por Larson (1974) para redimensionar o patrulhamento policial na cidade de Nova Iorque, foi utilizado para avaliar o desempenho do serviço de transporte médico-emergencial da cidade de Campinas-SP. O modelo mostrou-se uma ferramenta eficaz para auxiliar no planejamento e operação deste sistema.

Observou-se que o processo de camadas é uma técnica eficaz para modelar sistemas com múltiplas classes de usuários, onde prioridades estão embutidas no processo de atendimento, como é o caso do SAMU da cidade de Campinas. A avaliação de cenários alternativos mostrou a importância da descentralização de ambulâncias no sistema, o que reduz consideravelmente o tempo médio de resposta a um chamado em relação ao da configuração atual e, com isso, aumenta as chances de sobrevivência das vítimas.

A aplicação do modelo hipercubo no SAMU-192 contribuiu para difundir para a Rede Brasileira de Cooperação em Emergências (RBCE) a importância de conciliar a experiência administrativa da classe médica com o uso de ferramentas analíticas, para auxiliar nas decisões de planejamento e operação dos sistemas de transporte emergencial de saúde brasileiros.

AGRADECIMENTOS

Os autores agradecem à equipe do SAMU-192 de Campinas, em especial à Dra. Arine Campos Oliveira Assis, coordenadora do sistema, pelo apoio à realização da presente pesquisa.

REFERÊNCIAS BIBLIOGRÁFICAS

- Albino, J. C. C. (1994) *Quantificação e locação de unidades móveis de atendimento de emergência a interrupções em redes de distribuição de energia elétrica: aplicação do modelo hipercubo*. Florianópolis. 66p. Dissertação (Mestrado) – Universidade Federal de Santa Catarina.
- Anderson, L. R. e R. A. Fontenot (1992) Optimal positioning of service units along a coordinate line. *Transportation Science*, v. 26, n. 4, p. 346-351.
- Brandeau, M. L. e R. C. Larson (1986) Extending and applying the hypercube queueing model to deploy ambulances in Boston. *TIMS studies in the Management Science*, v. 22, p. 121-153.
- Chiyoshi, F.; R. D. Galvão e R. Morabito (2000) O uso do modelo hipercubo na solução de problemas de localização probabilísticos. *Gestão & Produção*, v.7, n.2, p.146-174.
- Daskin, M. S. e E. H. Stern (1981) A hierarchical objective set covering model for emergency medical service vehicle deployment. *Transportation Science*, v. 15, p. 137-152.
- Fitzsimmons, J. A. (1973) A methodology for emergency ambulance deployment. *Management Science*, v. 19, n. 6, p. 627-636.
- Gonçalves, M. B. (1994) Métodos de pesquisa operacional em serviços emergenciais. *Anais do XXVIº Simpósio Brasileiro de Pesquisa Operacional*, SOBRAPO, Florianópolis, v. 1, p. 597-601.
- Gonçalves, M. B.; A. G. N. Novaes e J. C. C. Albino (1994) Modelos para localização de serviços emergenciais em rodovias. *Anais do XXVIº Simpósio Brasileiro de Pesquisa Operacional*, SOBRAPO, Florianópolis, v. 1, p. 591-596.
- Gonçalves, M. B.; A. G. N. Novaes e R. Schmitz (1995) Um modelo de otimização para localizar unidades de serviços emergenciais em rodovias. *Anais do IXº Congresso Brasileiro de Pesquisa e Ensino em Transportes*, ANPET, São Carlos, v. 3, p. 962-972.
- Halpern, J. (1977) The accuracy of estimates for the performance criteria of certain emergency service queueing systems. *Transportation Science*, v. 11, n. 3, p. 227-242.
- Keeney, R. L. (1972) A method for districting among facilities. *Operations Research*, v. 20, n. 3, p. 613-618.

- Larson, R. C. (1974) A hypercube queueing model for facility location and redistricting in urban emergency services. *Computer and Operations Research*, v. 1, n. 1, p. 67-95.
- Larson, R. C. e A. R. Odoni (1981) *Urban Operations Research*. New Jersey, Prentice-Hall.
- Louveaux, F. (1993) Stochastic location analysis. *Location Science*, v. 1, p. 127-154.
- Mendonça, F. C. e R. Morabito (2000) Aplicação do modelo hipercubo para análise de um sistema médico-emergencial em rodovia. *Gestão & Produção*, v. 7, n. 1, p. 73-90.
- Savas, E. S. (1969) Simulation and cost-effectiveness analysis of New York's emergency ambulance service. *Management Science*, v. 15, n. 12, p.B-608 – B-627.
- Swersey, A. J. (1994) The deployment of police, fire and emergency medical units. *Handbooks in OR &MS*, v. 6, p. 151-200.
- Swoveland, C.; D. Uyeno; I. Vertinsky e R. Vickson (1973) Ambulance location: a probabilistic enumeration approach. *Management Science*, v. 20, n. 4, Part II, p. 686-698.
- Takeda, R. A. (2000) *Uma contribuição para avaliar o desempenho de sistemas de transporte emergencial de saúde*. São Carlos. Tese (Doutorado) – Escola de Engenharia de São Carlos, Universidade de São Paulo.
- Toregas, C.; R. Swain; C. ReVelle e L. Bergnan (1971) The location of emergency service facilities. *Operations Research*, v. 19, p. 1363-1373.

Endereço dos Autores:

Renata Algisi Takeda
João Alexandre Widmer
Departamento de Transportes - EESC – USP
Avenida Trabalhador São-Carlense, 400
13566-590 - São Carlos-SP - Brasil
E-mail: renata_at@uol.com.br
widmer@usp.br

Reinaldo Morabito
Departamento de Engenharia de Produção – UFSCar
Caixa Postal 676
13565-905 - São Carlos-SP - Brasil
E-mail: morabito@power.ufscar.br