

## APLICAÇÃO DO MODELO HIPERCUBO DE FILAS PARA AVALIAR A DESCENTRALIZAÇÃO DE AMBULÂNCIAS EM UM SISTEMA URBANO DE ATENDIMENTO MÉDICO DE URGÊNCIA

**Renata Algisi Takeda**

**João Alexandre Widmer**

Departamento de Transportes – EESC

Universidade de São Paulo

São Carlos – SP

**Reinaldo Morabito \***

Departamento de Engenharia de Produção

Universidade Federal de São Carlos

São Carlos – SP

[morabito@power.ufscar.br](mailto:morabito@power.ufscar.br)

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

*Recebido em 06/2003; aceito em 10/2003*

*Received June 2003; accepted October 2003*

### Resumo

Uma das maiores preocupações de sistemas urbanos de atendimento médico-emergencial é a rapidez no atendimento às vítimas. Os tempos de resposta dependem de diversos fatores como condições de tráfego local, dia da semana e período do dia, tipo e número de veículos disponíveis, localização destes veículos, políticas de despacho, etc. Neste trabalho apresenta-se uma análise dos efeitos da descentralização de ambulâncias que operam no sistema de atendimento médico-emergencial (SAMU-192) de Campinas, SP. O problema é tratado por meio do modelo hipercubo de filas espacialmente distribuídas, que considera variações aleatórias dos processos de chegada e atendimento dos chamados emergenciais. A aplicação do modelo produz uma ampla variedade de indicadores de desempenho para o sistema, que são comparados com os valores reais observados. Os resultados obtidos com a descentralização das ambulâncias mostraram uma significativa elevação do nível de serviço oferecido aos usuários.

**Palavras-chave:** modelo hipercubo de filas; serviços emergenciais de saúde; localização probabilística.

### Abstract

One of the major concerns of urban emergency medical systems is to provide the fastest possible first care medical assistance to the victims. The response times depend on different aspects such as: local traffic conditions, weekday and time, type and number of available vehicles, location of these vehicles, dispatching policies, etc. This work analyzes the effects of decentralizing ambulances in the emergency medical system (SAMU-192) of Campinas, SP. The problem is dealt with the hypercube queuing model, which considers stochastic variations of the arrival and service processes of the emergency calls. The application of the model produces a wide variety of system performance indicators, which are compared to the actual observed values. The results obtained with ambulance decentralization showed a significant increase in the service level offered to the users.

**Keywords:** hypercube queuing model; emergency medical services; probabilistic location.

## 1. Introdução

A qualidade de vida da população, especialmente das áreas urbanas, está ligada a uma variedade de serviços públicos 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 vidas; por outro lado, quando ineficiente, é um potencial responsável pelo agravamento dos casos. A rapidez no atendimento às vítimas é uma das maiores exigências destes sistemas, e o tempo decorrido entre o instante da solicitação pelo serviço e o início do atendimento, denominado *tempo de resposta*, é um dos principais fatores que influenciam o nível de serviço aos usuários.

Este tempo é resultante de um conjunto de fatores, como por exemplo, experiência e capacitação profissional da equipe, condições de tráfego local, dia da semana e período do dia, tipo e número de veículos disponíveis, localização destes veículos, políticas de despacho, etc. A regulamentação americana para os serviços médicos de urgência estabelece que 95% das solicitações em área urbana devem ser atendidas em, no máximo, 10 minutos, sendo este período estendido para 30 minutos para a área rural (Ball & Lin, 1993). Nos casos dos serviços de ambulâncias das cidades de Londres e Montreal, 95% das solicitações devem ser servidas em 14 e 10 minutos, respectivamente, e 50% e 70% das solicitações devem ser servidas em 8 e 7 minutos, respectivamente (Geandreau *et al.*, 2001; Galvão *et al.*, 2003a). No Brasil, entretanto, não há legislação específica que determine limitações para os tempos de resposta.

Serviços emergenciais como os SAMU's apresentam alto grau de incerteza em suas características operacionais e, quanto maior o grau de incerteza envolvido e maior a necessidade de respostas rápidas, menor deve ser a taxa de utilização dos operadores e equipamentos do sistema. Caso contrário, o nível de serviço oferecido aos usuários pode ser deteriorado. Neste contexto, quando os sistemas são bem dimensionados, geralmente ocorrem longos períodos em que os operadores e equipamentos permanecem desocupados (Gonçalves, 1994). Estudos recentes mostram elevados índices de ocupação dos veículos e suas equipes (da ordem de 60%) em observações realizadas em algumas cidades do interior paulista (Takeda, 2000).

Diversas 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 o uso dos recursos disponíveis. No entanto, uma limitação dos estudos é que muitos deles não consideram a natureza probabilística dos processos de chegada e atendimento dos chamados emergenciais, e não levam em conta o fato de que as ambulâncias nem sempre estão disponíveis para iniciar um atendimento. Dentre as contribuições, destacam-se trabalhos onde a atenção é concentrada em questões tais como *localização de bases* (Toregas *et al.*, 1971; Daskin & Stern, 1981; Anderson & Fontenot, 1992; Galvão, 1993; Grendreau *et al.*, 1997 e Grendreau *et al.*, 2001), *programação de equipes* (Kaplan, 1979; Green & Kolesar, 1984; Taylor & Huxley, 1989 e Aubin, 1992) e *problemas de congestionamento* (Berman & Larson, 1982; Brandeau & Larson, 1986; Marianov & Serra, 1998). Revisões dos estudos desenvolvidos nas últimas décadas podem ser encontradas em Kolesar & Swersey (1986), Louveaux (1993), Swersey (1994), Chiyoshi *et al.* (2000) e Brotcorne *et al.* (2003).

A modelagem integrada destas questões é complexa, pois geralmente os sistemas reais são compostos por vários veículos; as solicitações por serviço ocorrem distribuídas no tempo e

no espaço; 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 ainda existe a possibilidade de formação de filas de espera. O *modelo hipercubo de filas distribuídas*, desenvolvido por Larson (1974) e estendido por outros autores (Swersey, 1994), é uma importante ferramenta neste sentido, dando condições de se avaliar não só o nível de serviço oferecido, como também os impactos de decisões tomadas para o sistema. 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 & 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).

Vários estudos são encontrados na literatura com o objetivo de avaliar o uso deste modelo na solução de problemas de localização probabilística, em particular, para sistemas de atendimento emergencial. Por exemplo, Batta *et al.* (1989) sugeriram o uso do hipercubo como alternativa para relaxar a hipótese de independência entre os servidores dos problemas de localização com máxima cobertura, desenvolvido por Daskin (1983). Chiyoshi *et al.* (2001, 2002) analisaram o caso de servidores não homogêneos e apresentaram comparações entre os modelos de localização com máxima cobertura e máxima cobertura ajustada (Batta *et al.*, 1989). Saydam & Aytug (2003) aplicaram um algoritmo genético incorporando a metodologia do hipercubo desenvolvida por Jarvis (1975) no cálculo da cobertura esperada. Gendreau *et al.* (1997, 2001) aplicaram busca tabu em contextos similares. Galvão *et al.* (2003a, 2003b) investigaram o uso de *simulated annealing* com o modelo hipercubo na solução de problemas de localização com máxima cobertura e máxima disponibilidade (ReVelle & Hogan, 1989).

A aplicação original do modelo hipercubo foi para o problema de patrulhamento policial. Porém, sistemas como bombeiros, ambulâncias, defesa civil, reparos em redes de infra-estruturas básicas (água, energia elétrica e telecomunicações), guinchos e entregas domiciliares também podem ser representados por esta técnica. Nos Estados Unidos várias aplicações podem ser citadas: a localização de ambulâncias em Boston (Brandeau & Larson, 1986) e Greenville (Burwell *et al.*, 1993), o patrulhamento policial em New Haven (Chelst & Barlach, 1981) e Orlando (Sacks & Grief, 1994), programas de visitas de serviço social (Larson & Odoni, 1981), dentre outros. No Brasil, alguns exemplos 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), o balanceamento do fator de utilização das ambulâncias do sistema Anjos do Asfalto da Rodovia Presidente Dutra (Mendonça & Morabito, 2000, 2001), e a análise de desempenho do sistema SAMU-192 de Campinas-SP (Takeda *et al.*, 2001).

O presente trabalho apresenta uma extensão dos estudos em Takeda *et al.* (2001), partindo de cenários simples, como é o caso da descentralização de uma única ambulância, e suas conseqüências, até situações em que vários veículos são reposicionados. Também são mostrados os impactos do aumento do número de ambulâncias, aliado à descentralização, sobre alguns importantes indicadores de desempenho do sistema. O artigo está organizado da seguinte maneira: na seção 2 é feita uma descrição resumida do sistema SAMU-192, evidenciando seus aspectos operacionais importantes para a aplicação do modelo. Na seção 3 é realizada uma breve revisão sobre o modelo hipercubo e sua aplicação no sistema estudado; também são calculadas as principais medidas de desempenho para a configuração original do SAMU-192. Na seção 4 o modelo é aplicado para explorar cenários alternativos;

em particular, descentralizando ambulâncias na configuração original e avaliando seus efeitos sobre alguns indicadores de desempenho. Também são analisados os efeitos do aumento do número de ambulâncias no sistema. Finalmente, na seção 5, são apresentadas as conclusões do estudo e algumas perspectivas para futuras pesquisas.

## 2. Os Sistemas Emergenciais e o SAMU-192

A Organização Mundial de Saúde estabelece, como função principal de um serviço de atendimento médico-emergencial, dar suporte básico de vida a toda situação de risco envolvendo pessoas e bens (*World Health Organization*, 2000). No Brasil, desde o início dos anos 90, esforços têm sido direcionados no sentido de impulsionar a organização dos atendimentos médicos de urgência. Por se tratar de um serviço público de atendimento, o sistema brasileiro assemelha-se ao modelo francês, em operação há mais de 20 anos. Graças à ativa cooperação entre algumas secretarias de saúde e o governo francês, as populações de Araraquara, Belo Horizonte, Campinas, Fortaleza, Goiânia, Limeira, Porto Alegre, Recife, Ribeirão Preto, Salvador, São Paulo e Vale do Ribeira já estão recebendo um atendimento de qualidade superior, com um melhor aproveitamento dos recursos disponíveis.

Operacionalmente, os SAMU's são sistemas que englobam uma determinada cidade ou até uma região constituída de várias cidades, e contém uma central telefônica de recepção de chamados, equipes de socorristas centralizados ou distribuídos pela área de atuação do sistema, ambulâncias com equipamentos de suporte básico de vida, e em alguns casos, unidades de reanimação pré-hospitalar (UTI's móveis) com equipe qualificada. Esta organização está coordenada com uma rede hospitalar municipal ou regional em que as estruturas hospitalares de atendimento são escolhidas em função de sua complexidade.

### 2.1 Caracterização da pesquisa de campo

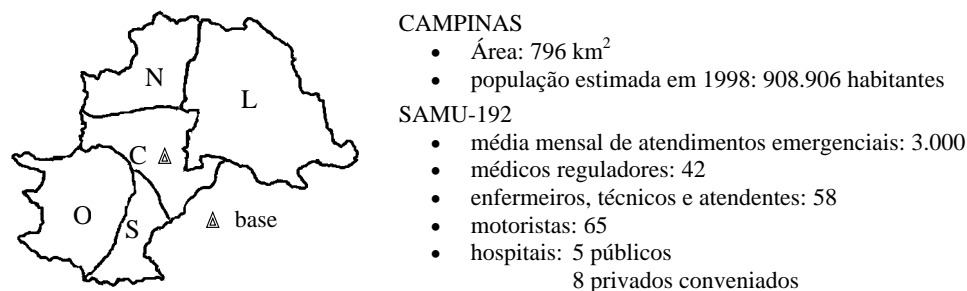
No início do estudo tentou-se contatar os serviços de cinquenta municípios de pequeno, médio e grande porte do interior paulista, dentre os quais apenas oito se dispuseram a divulgar informações sobre os atendimentos realizados. Neste primeiro contato observou-se que os serviços operavam 24 horas por dia com uma central telefônica para onde convergiam todos os chamados. Todas as cidades possuíam uma base manuscrita de dados relacionados aos atendimentos. Também foi observado que não havia um programa de avaliações periódicas que indicasse o desempenho dos sistemas. Considerando-se então a caracterização destes oito sistemas, o SAMU-192 foi escolhido como cenário para estudo.

A cidade de Campinas passou por uma reorganização do seu sistema de atendimento médico de urgência em 1994, implantando o SAMU-192 com a função de ser o centro regulador das urgências médicas do município. O SAMU-192 de Campinas é considerado um sistema modelo no Estado de São Paulo, com uma equipe qualificada em todos os seus níveis. Sua coordenação mostrou-se interessada em participar do estudo, contribuindo de forma importante para o levantamento dos dados e a interpretação dos resultados obtidos posteriormente com a aplicação do modelo.

No ano de 1998, quando a pesquisa de campo foi iniciada, o sistema operava com dezoito ambulâncias, sendo elas: onze VRS's (veículos de remoção simples), utilizados para o transporte de pacientes com casos clínicos simples, e principalmente para serviços agendados (sua equipe é formada apenas pelo motorista, eventualmente acompanhado de um auxiliar de

enfermagem); quatro VSB's (veículos de suporte básico), utilizados para o transporte das urgências, com equipamentos básicos para o atendimento domiciliar das vítimas (necessariamente sua equipe é formada pelo motorista e um enfermeiro); dois VSA's (veículos de suporte avançado), utilizado para os casos graves, também denominados UTI's móveis (equipe composta pelo motorista, um enfermeiro e um médico) e um PSQ (veículo psiquiátrico) totalmente adaptado para suportar o transporte de pacientes com comportamento agressivo, alcoolizados e drogados (com profissionais especialmente treinados para tratar estes casos). Dentre os onze VRS's, sete eram dedicados às operações de remoções simples (pacientes agendados) e quatro operavam adaptados com equipe e equipamentos de VSB's. No total, dez ambulâncias eram dedicadas, de fato, às operações de urgência (dois VSA's e oito VSB's).

A região de Campinas estava subdividida em cinco áreas distintas – Norte, Sul, Leste Oeste e Central – de acordo com a distribuição de áreas de cobertura dos principais Centros de Saúde do município. Todos os veículos, quando disponíveis, permaneciam centralizados na base operacional do sistema, local onde se encontra a central telefônica 192 para onde convergem todos os chamados. O sistema também permitia que seus usuários entrassem em uma fila de espera caso todas as ambulâncias estivessem ocupadas no instante da ocorrência do chamado (esta fila era limitada, em média, por um usuário para cada veículo em operação). A Figura 1 mostra algumas características da cidade de Campinas e do SAMU-192 no período de desenvolvimento da pesquisa.



**Figura 1** – Atuação do SAMU-192 em Campinas-SP

É importante ressaltar que a equipe que gerenciava o sistema estava preocupada com os altos índices de utilização das ambulâncias em determinados períodos do dia, o que podia provocar esperas para os usuários. Neste sentido, reflexões sobre uma possível mudança na estrutura operacional do sistema foram levantadas. Questões como o reposicionamento das ambulâncias na área de atuação do SAMU-192, e mesmo o número ideal de veículos para atender a demanda, predominavam.

## 2.2 Coleta de dados

Para implementar o modelo hipercubo no SAMU-192, a coleta de dados foi separada em duas etapas: a primeira, com a descrição dos tipos e quantidade de atendimentos realizados no sistema, e a segunda com a seleção de uma amostra significativa para determinar as estatísticas necessárias para validar as hipóteses do modelo.

### Descrição dos atendimentos

A coordenação do SAMU-192 informou que não existiam grandes diferenças no número de solicitações de serviço ao longo dos meses do ano; meses de férias escolares, entretanto, apresentavam uma leve redução no número de atendimentos realizados. Também observaram que os finais de semana (sábados e domingos) apresentavam menores índices de solicitação de serviço, e a experiência da equipe contribuiu para confirmar que estas são características típicas de sistemas urbanos de atendimento médico de urgência (estas informações foram verificadas nas fichas de atendimentos realizados, desde a implantação do serviço). Para constatar a primeira afirmação, foi considerada uma amostra aleatória de cada mês do ano de 1997. A Figura 2 mostra a distribuição deste conjunto de dados. Tendo comprovado que não havia diferenças significativas no número de atendimentos/mês, o próximo passo foi examinar uma outra amostra aleatória de um mês do ano, para verificar se os dias úteis da semana se comportam de maneira semelhante. O mês sorteado foi novembro e dele, selecionou-se uma amostra de dez dias úteis de suas quatro semanas (Figura 3).

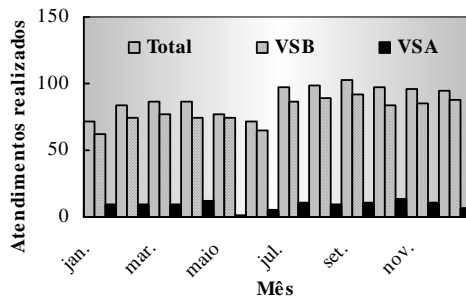


Figura 2 – Atendimentos emergenciais realizados em 1997

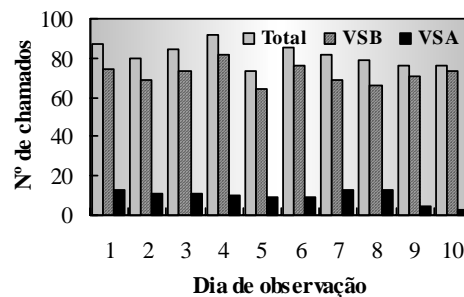


Figura 3 – Atendimentos realizados no mês de novembro (1997)

Outra importante observação foi com relação aos períodos do dia em que se constatava o maior índice de chamados no sistema. Para a equipe que coordenava o SAMU-192, o período entre 10 e 21 horas era considerado crítico. Com base nesta informação, construiu-se a distribuição dos chamados ao longo do dia, para a amostra do mês de novembro (Figura 4). Na amostra selecionada, foram observados 814 atendimentos, dos quais 17,1% tiveram origem na região Norte, 16,5% na região Sul, 17,6% na Leste, 32,2% na Oeste e 16,7% na região Central.

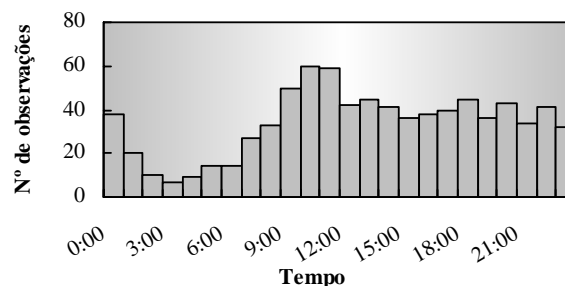


Figura 4 – Distribuição dos atendimentos ao longo de 24 horas

### Identificação de períodos de pico

Conforme observado na Figura 4, o SAMU-192 apresentava um período de pico bem caracterizado (das 9 às 14 horas, aproximadamente). Entretanto, foi necessário realizar uma análise mais detalhada para determinar a amplitude do intervalo que define o pico. Sendo assim, realizou-se um estudo baseado nas taxas de chegada de chamados dentro de possíveis combinações de períodos entre 8 e 15 horas e 16 e 22 horas, para verificar qual é o intervalo de tempo que, em geral, apresentava taxa média de chegadas significativa, com menor variabilidade. A Tabela 1 mostra os valores das taxas médias dos períodos mais prováveis e suas respectivas variações (a última coluna corresponde ao desvio-padrão do período, d-p).

**Tabela 1** – Taxas médias de chegadas para possíveis períodos de pico (chegadas/minuto)

Período	DIA										média	d-p
	1	2	3	4	5	6	7	8	9	10		
9-14	0,11	0,08	0,06	0,10	0,07	0,08	0,10	0,09	0,08	0,07	0,09	0,02
10-14	0,11	0,08	0,07	0,10	0,08	0,08	0,10	0,09	0,09	0,07	0,09	0,01
9-13	0,13	0,08	0,06	0,11	0,07	0,08	0,10	0,10	0,08	0,08	0,09	0,02
10-13	0,13	0,08	0,07	0,11	0,07	0,07	0,12	0,09	0,09	0,08	0,09	0,02

Os demais intervalos analisados apresentaram taxas menores, não caracterizando, de fato, o período de congestionamento do sistema. O intervalo escolhido, portanto, inicia-se às 10 e termina às 14 horas, pois foi o que apresentou maiores frequências durante os dez dias amostrados, com menor variação em torno do valor médio. Neste período, ocorreram 209 observações, dentre as quais, 190 foram atendidas por equipes básicas e 19 por equipes avançadas. A distribuição percentual para as cinco regiões se manteve. Portanto, as avaliações que seguem refletem algumas características do SAMU-192 durante o seu período de pico. A Tabela 2 apresenta medidas de desempenho da amostra de dez dias, para valores observados nas 24 horas do dia e para valores observados apenas no período de pico (10 às 14 horas).

**Tabela 2** – Medidas de desempenho obtidas no SAMU-192

	24 horas		10-14 horas	
	VSB	VSA	VSB	VSA
número de chamados	70	10	20	2
intervalos entre chegadas sucessivas (min)	14	106	10	92
tempo médio de resposta (min)	12	10	13	11
tempo em cena (min)	37	39	37	39
tempo total de atendimento (min)	64	62	66	63

### 3. O Modelo Hipercubo Aplicado ao SAMU-192

O modelo hipercubo baseia-se na partição da região de atuação do sistema em um conjunto finito de áreas geradoras de demanda, denominadas *átomos geográficos*. Cada átomo é considerado como uma fonte pontual independente de solicitação de serviço ao longo do tempo. Portanto, são consideradas as distribuições espacial e temporal dos chamados. O atendimento é realizado por servidores distribuídos na região e que, quando disponíveis,

podem estar fixos em alguns pontos (bases), ou em movimento (neste caso, sua localização deve ser conhecida ao menos probabilisticamente). A área de cobertura primária de um servidor é definida pelo conjunto dos átomos que este servidor atende prioritariamente, ou seja, é o primeiro a ser chamado. Nos casos onde o servidor preferencial está ocupado, um outro servidor é chamado para atender a solicitação, de acordo com lista de preferência de despachos pré-definida pelo sistema.

O nome hipercubo deriva do espaço de estados que descreve o *status* dos servidores. Cada servidor pode estar livre (0) ou ocupado (1) num certo instante. Um estado particular do sistema é dado pela lista dos servidores que estão livres e ocupados. Por exemplo, para um sistema com três servidores, o estado 011 corresponde à situação onde o servidor número 1 está ocupado, o servidor número 2 está ocupado e o servidor número 3 está livre (note que os servidores estão representados da direita para a esquerda). Desta maneira, um sistema com três servidores tem seu espaço de estados em  $\mathbb{R}^3$ , e é dado pelo conjunto de vértices do cubo de arestas iguais a 1. À medida que cresce o número de servidores, a representação se dá no espaço  $N$ -dimensional, pelo conjunto de vértices de um *hipercubo*.

O analista deve coletar informações sobre os processos de chegadas e de atendimento 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, ele calcula medidas para: (i) a extensão total da região; (ii) um veículo específico; (iii) um subconjunto finito de átomos, e (iv) um átomo específico.

### 3.1 Hipóteses da modelagem

Para se aplicar o modelo, o sistema deve atender a nove hipóteses básicas:

#### *Átomos geográficos*

A área na qual o sistema opera é particionada em  $N_A$  áreas menores, cada qual representando uma fonte independente de solicitações por serviço. Vários sistemas do tipo *server-to-customer* admitem implícita ou explicitamente critérios de prioridade para os chamados que recebem. Porém, nem sempre é uma tarefa simples representar tais critérios no modelo. Entretanto, existem situações onde estes critérios devem ser introduzidos, como é o caso dos sistemas de atendimento médico-emergencial, em que existem servidores dedicados a serviços especializados. Em Campinas, por exemplo, VSA's e VSB's realizam a operação do sistema.

Neste caso, o modelo pode ser adaptado para refletir os critérios de prioridade, bem como as regras de despacho dos servidores. Particiona-se cada átomo do sistema em duas áreas distintas e independentes: uma gerando apenas chamados do tipo avançado, e outra gerando apenas chamados do tipo básico – processo de camadas (Larson & Odoni, 1981). Assim, todas as regiões mostradas na Figura 1 foram biparticionadas em básicas (B) e avançadas (A), gerando um total de 10 átomos geográficos no sistema: Norte B (NB), Norte A (NA), Sul B (SB), Sul A (SA), Leste B (LB), Leste A (LA), Oeste B (OB), Oeste A (OA), Centro B (CB), Centro A (CA). No SAMU-192, as áreas geradoras de chamados avançados são atendidas, preferencialmente pelos VSA's; neste caso, os VSB's ficam como *backups*. De maneira semelhante, as áreas geradoras de chamados básicos são atendidas preferencialmente pelos VSB's e os VSA's ficam como *backups*.



### Processo de chegadas poissoniano

Considera-se que os chamados chegam independentemente em cada átomo  $j$ , segundo processos de Poisson com taxas médias  $\lambda_j$  ( $j=1,2,\dots,N_A$ ), possíveis de serem medidas ou estimadas. Embora esta hipótese pareça muito restritiva, ela é frequentemente satisfeita em diversos sistemas reais. No caso do SAMU-192, tal hipótese foi verificada examinando se o número de chegadas em cada átomo tem distribuição de Poisson. Foram assim determinadas as taxas médias de solicitação por serviço durante o período de observação (equivalentemente, os intervalos médios entre chegadas sucessivas para os dez átomos do sistema). Constatou-se que os chamados ocorrem independentemente átomo a átomo, e o número de chamados de cada átomo representa uma proporção do total do sistema. Assim, foram determinadas as probabilidades  $p_j$  que representam as chances de um chamado qualquer ter origem no átomo  $j$ .

Desta maneira, a hipótese pode ser testada utilizando a propriedade de decomposição de processo de Poisson (Wolff, 1989). Para determinar a taxa média de chegadas ao sistema, foram considerados os intervalos médios entre chegadas sucessivas para todos os dias de observação, no período das 10 às 14 horas, considerando todos os átomos do sistema. A Tabela 3 apresenta o tempo médio, desvio-padrão e coeficiente de variação (desvio-padrão/tempo médio) para cada dia. Nota-se que os desvios-padrão são, em geral, da mesma ordem de grandeza das médias, o que sugere que os intervalos de tempo entre chegadas sucessivas ao sistema podem ser considerados exponencialmente distribuídos.

**Tabela 3** – Intervalos médios entre chegadas sucessivas ao sistema

<b>Dia</b>	<b>Intervalo médio (min)</b>	<b>Desvio-padrão</b>	<b>Coefficiente de variação</b>
<b>1</b>	9	7	0,8
<b>2</b>	11	12	1,1
<b>3</b>	14	12	0,9
<b>4</b>	9	9	1,0
<b>5</b>	11	17	1,6
<b>6</b>	13	15	1,2
<b>7</b>	9	7	0,8
<b>8</b>	10	14	1,4
<b>9</b>	10	9	0,9
<b>10</b>	15	19	1,3
<b>Média</b>	<b>11,10 minutos (0,1850 hora)</b>	<b>12,10</b>	1,1
<b><math>\lambda = 1/0,1850 = 5,4054</math> chamados/hora</b>			

Para verificar estatisticamente tal hipótese, foi aplicado o *Método de Kolmogorov-Smirnov*, ao nível de significância  $\alpha = 0,05$ . O resultado indicou que não se pode rejeitar a hipótese de que os intervalos entre chegadas sucessivas ao sistema têm distribuição exponencial e, portanto, foi admitido que o número de chegadas por intervalo de tempo tenha distribuição de Poisson. Com este resultado, e considerando as devidas proporções que representam as chegadas em cada átomo do sistema, admitiu-se que os chamados chegam em cada átomo segundo Processos de Poisson independentes, com taxas médias  $\lambda_j = \lambda p_j$  ( $j = 1, 2, \dots, 10$ ). A Tabela 4 resume as informações sobre as taxas médias de chegadas.

**Tabela 4** – Taxas médias de chegadas para cada átomo

Átomo	Nº de chamados	$p_j$	$\lambda_j$ (chamados/hora)
1 - NB	33	0,1579	0,8535
2 - NA	4	0,0191	0,1035
3 - SB	33	0,1579	0,8535
4 - SA	3	0,0144	0,0776
5 - LB	33	0,1579	0,8535
6 - LA	3	0,0144	0,0776
7 - OB	58	0,2750	1,5001
8 - OA	5	0,0239	0,1293
9 - CB	32	0,1531	0,8276
10 - CA	5	0,0239	0,1293
<b>Total</b>	<b>209</b>	<b>1,0000</b>	<b>5,4054</b>

**Tempos de viagem entre átomos**

Os tempos médios de viagem  $\tau_{ij}$  ( $i, j = 1, 2, \dots, N_A$ ) entre o átomo  $i$  e o átomo  $j$  devem ser conhecidos ou estimados através dos conceitos de probabilidade geométrica. A matriz dos tempos médios de viagem entre átomos foi fornecida pelo próprio SAMU-192, admitidas as dez regiões geradoras de chamados (Tabela 5). Os veículos de suporte avançado geralmente desempenham tempos médios de viagem entre átomos menores que os básicos, devido ao grau de urgência envolvido nos atendimentos. Os tempos de viagem dentro do próprio átomo são distintos de zero, pois é preciso representar as reais distâncias percorridas por uma ambulância e os tempos necessários para tal, dentro de uma determinada área.

**Tabela 5** – Tempos médios de viagem entre átomos (minutos)

Átomo	NB	NA	SB	SA	LB	LA	OB	OA	CB	CA	Média
1 - NB	6	6	28	28	16	16	24	24	16	16	<b>18</b>
2 - NA	4	4	24	24	12	12	20	20	12	12	<b>14</b>
3 - SB	28	28	6	6	24	24	16	16	16	16	<b>18</b>
4 - SA	24	24	4	4	20	20	12	12	12	12	<b>14</b>
5 - LB	16	16	24	24	6	6	24	24	16	16	<b>17</b>
6 - LA	12	12	20	20	4	4	20	20	12	12	<b>14</b>
7 - OB	24	24	16	16	24	24	6	6	16	16	<b>17</b>
8 - OA	20	20	12	12	20	20	4	4	12	12	<b>14</b>
9 - CB	16	16	16	16	16	16	16	16	5	5	<b>14</b>
10 - CA	12	12	12	12	12	12	12	12	3	3	<b>10</b>
<b>Média do sistema = 14,5</b>											

**Servidores**

O sistema é composto por  $N$  servidores (distintos ou não) espacialmente distribuídos, que podem deslocar-se e atender qualquer um dos átomos. No cenário considerado, o SAMU-192 trabalha com uma frota de 10 veículos distintos: oito VSB's e dois VSA's. Todas as ambulâncias permanecem centralizadas na base, quando disponíveis, podendo deslocar-se para qualquer um dos átomos para realizar um atendimento.

### Despacho dos servidores

Apenas um servidor é despachado para atender um chamado. Se não houver veículo disponível para realizar o atendimento, o chamado poderá entrar em uma fila ou então ocorrer a perda do chamado, sendo este, por exemplo, transferido para outro sistema de atendimento. O SAMU-192 admite a política de enviar apenas uma equipe para atender a um chamado emergencial e em sua política de atendimento é permitida a formação de fila. O chamado que chega quando todas as ambulâncias estão ocupadas aguardam em fila até que um veículo desocupe para atendê-lo, desde que esta fila de espera não ultrapasse o limite de dez usuários (isto é, um usuário por ambulância em operação).

### Localização dos servidores

A localização de cada servidor, quando disponível, deve ser conhecida. No caso de servidores fixos, a localização é a sua base; no caso de servidores que se movem dentro de uma determinada área (por exemplo, viaturas de patrulhamento policial), esta localização deve ser conhecida ao menos probabilisticamente. A representação destas localizações é bastante genérica: ela se apresenta de forma matricial ( $L = [l_{nj}]$ ), onde os elementos da matriz representam a probabilidade de um servidor estar localizado em um determinado átomo, quando disponível.  $L$  é, necessariamente, uma matriz estocástica, ou seja  $\sum l_{nj} = 1$ , com  $l_{nj} = 1$  se o servidor está localizado no átomo  $j$ , e  $l_{nk} = 0$  para todo  $k \neq j$ . A matriz  $L$  é facilmente determinada para a configuração original do SAMU-192, pois todos permanecem na base central quando disponíveis. Para simplificar a nomenclatura, a seguir os veículos de suporte avançado passam a ser chamados de veículos 1 e 2, e os veículos de suporte básico de 3, 4, 5, 6, 7, 8, 9 e 10.

$$L = \begin{array}{cccccccccc|c} \text{NB NA SB SA LB LA OB OA CB CA} & & & & & & & & & & & \\ \left[ \begin{array}{cccccccccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \mathbf{1} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \mathbf{2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \mathbf{3} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \mathbf{4} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \mathbf{5} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \mathbf{6} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \mathbf{7} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \mathbf{8} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \mathbf{9} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \mathbf{10} \end{array} \right. \end{array}$$

### Política de despacho dos servidores

Há uma lista fixa de preferências de despacho para cada átomo. Se o primeiro servidor desta lista estiver disponível, ele é despachado para atender o chamado; caso contrário, o próximo servidor disponível da lista é despachado. Devido à política de prioridade entre veículos VSA's e VSB's, a lista de preferência de despacho do SAMU-192 foi obtida segundo o procedimento de geração aleatória de matrizes de Burwell *et al.* (1993). A Tabela 6 apresenta um exemplo de uma matriz gerada aleatoriamente para o SAMU-192. Considerando a configuração original do sistema, a lista de preferências de despachos apresenta, nas linhas

correspondentes aos átomos geradores de chamados da classe avançada, os veículos 1 e 2 (seqüenciados aleatoriamente) como servidores primários. As demais posições são ocupadas pelos oito VSB's (também seqüenciados aleatoriamente), que assumem a função de *backup*. Similarmente, nas linhas que representam as áreas geradoras de chamados que necessitam de atendimento básico, as oito primeiras posições são ocupadas pelos VSB's (seqüenciados aleatoriamente), e as duas últimas posições restantes são ocupadas pelos VSA's (também seqüenciados aleatoriamente), que atuam como *backup*.

**Tabela 6** – Matriz de preferência de despachos (cenário original)

Átomo	Preferência de despacho									
	1 <sup>a</sup>	2 <sup>a</sup>	3 <sup>a</sup>	4 <sup>a</sup>	5 <sup>a</sup>	6 <sup>a</sup>	7 <sup>a</sup>	8 <sup>a</sup>	9 <sup>a</sup>	10 <sup>a</sup>
1 - NB	9	10	3	4	5	6	7	8	2	1
2 - NA	1	2	10	7	3	4	9	5	8	6
3 - SB	10	3	4	5	6	7	8	9	1	2
4 - SA	2	1	5	3	10	9	4	6	7	8
5 - LB	3	4	7	10	8	5	6	9	2	1
6 - LA	1	2	6	9	4	10	5	3	8	7
7 - OB	4	5	8	6	7	3	9	10	2	1
8 - OA	2	1	9	8	3	5	10	4	6	7
9 - CB	5	6	8	10	9	4	3	7	1	2
10 - CA	2	1	10	5	6	8	7	9	3	4

**Veículos**

### *Tempos de atendimento*

O tempo total de atendimento de um chamado é composto por um tempo de preparo do servidor (*setup time*), tempo de viagem do servidor até o local da ocorrência, tempo de execução do serviço junto ao usuário (*tempo em cena*) e o tempo de retorno à base. Os servidores podem ter taxas médias de atendimento  $\mu_n$  distintas, para cada servidor  $n = 1, 2, \dots, N$ . No caso de sistemas que permitem a formação de fila, o modelo hipercubo é ainda mais preciso à medida que os tempos médios de atendimento aproximam-se dos respectivos desvios-padrão (Larson & Odoni, 1981).

*Tempo de preparação da equipe:* Os valores dos tempos de preparação tanto para equipes de VSA's quanto para equipes de VSB's são pequenos (menos de 1 minuto). Este tempo é praticamente desprezível uma vez que, estando a ambulância disponível no momento da ocorrência do chamado, basta que sua equipe se dirija ao veículo, para que partir para o local do evento.

*Tempo de viagem de ida:* Na amostra considerada, o tempo médio de viagem observado (incluindo possíveis tempos de espera em fila) foi de 11 minutos para os VSA's e de 13 minutos para os VSB's. Estes valores são utilizados posteriormente para a validação do modelo.

*Tempo em cena:* No caso dos VSA's, o tempo médio observado na amostra foi de 39 minutos, e para os VSB's, 37 minutos.

*Tempo de viagem de volta:* Os tempos médios de viagem de retorno à base, são, em geral diferentes do tempo de viagem de ida, pois em sua grande maioria, este tempo é medido a

partir do local para onde as vítimas tenham sido removidas (unidade hospitalar ou domicílio), e não a partir do local do evento. Para a amostra considerada, este tempo foi em média de 13 minutos para VSA's e de 16 minutos para VSB's. Ou seja, o tempo médio de atendimento para um chamado de urgência no SAMU-192 é de 63 minutos para chamados atendidos por VSA's e de 66 minutos para chamados atendidos por VSB's (Tabela 2). A Tabela 7 mostra os tempos médios de atendimento de cada ambulância, seus desvios-padrão, coeficientes de variação e as taxas médias correspondentes.

Observa-se nos resultados da Tabela 7 que os desvios-padrão dos tempos médios de atendimento são da ordem de grandeza das médias. Estes valores são, portanto, indicações de que os tempos de atendimento são aproximadamente exponenciais. De fato, verificou-se, através do *Método de Kolmogorov-Smirnov*, que não se pode rejeitar esta hipótese, ao nível de significância  $\alpha = 0,05$ . Também foram testadas hipóteses de igualdade entre os tempos médios de atendimento de VSA's e VSB's. Para isto foi utilizada a técnica de análise de variância, a fim de identificar diferenças entre as médias populacionais devidas às possíveis causas atuando simultaneamente sobre os elementos da população. O teste verificou com nível de significância  $\alpha = 0,05$  diferenças entre os tempos médios de atendimento para os dez veículos. Para efeito de modelagem, optou-se então por considerar servidores distintos no sistema.

**Tabela 7** – Taxas médias de atendimento para cada ambulância

Ambulância	Tempo médio de atendimento (*)	Desvio padrão (*)	Coefficiente de variação	$\mu$ (**)
1	61	40	0,7	0,9836
2	64	44	0,7	0,9375
3	63	52	0,8	0,9524
4	70	48	0,7	0,8571
5	61	44	0,7	0,9836
6	66	58	0,9	0,9091
7	64	50	0,8	0,9375
8	61	40	0,7	0,9836
9	70	58	0,8	0,8571
10	70	56	0,8	0,8571
<b>VSA</b>	<b>63</b>	<b>42</b>	<b>0,7</b>	<b>0,9606</b>
<b>VSB</b>	<b>66</b>	<b>51</b>	<b>0,8</b>	<b>0,9172</b>
<b>Total</b>	<b>65</b>	<b>49</b>	<b>0,8</b>	<b>0,9259</b>

(\*) minutos      (\*\*) chamados/hora

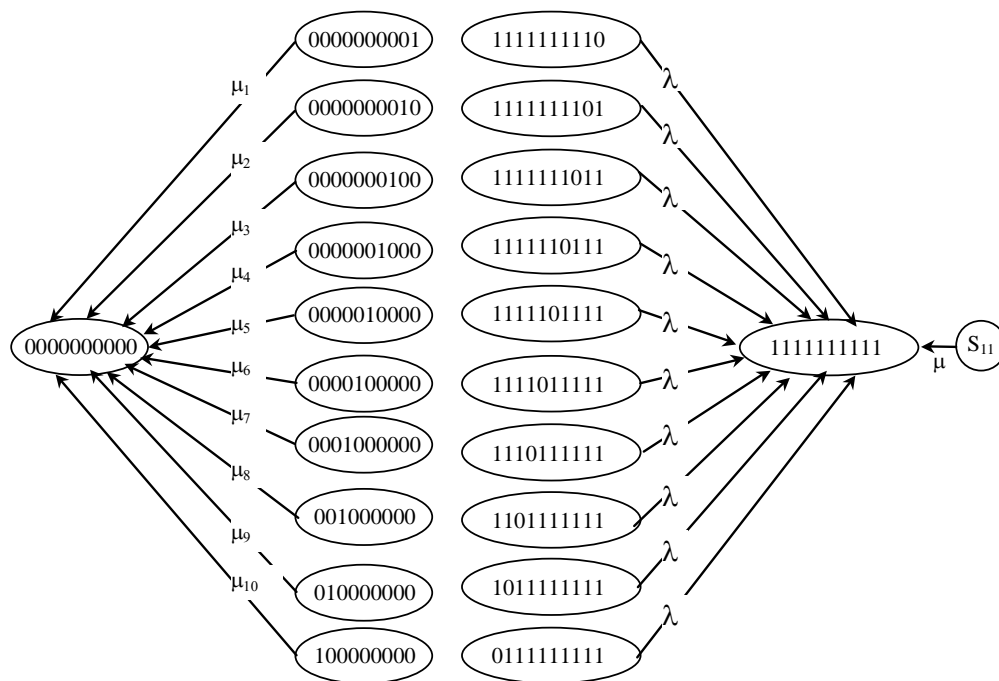
### **Relação entre o tempo de atendimento e o tempo de viagem**

Variações no tempo total de atendimento devidas às variações no tempo de viagem são admitidas como sendo de segunda ordem, quando comparadas às variações dos tempos junto ao paciente (em cena) e/ou tempo de preparação da equipe (*setup*). Esta hipótese, que limita a aplicabilidade do modelo, é frequentemente verificada em serviços urbanos, e menos observada em serviços em rodovia. Pode-se observar que os tempos médios de viagem são pequenos em relação aos tempos médios de atendimento, para cada ambulância (em média, cerca de 18%), e que variações nos tempos médios de viagem contribuem pouco para as variações nos tempos médios de atendimento (compare os valores apresentados nas Tabelas 7 e 8).

### 3.2 Possíveis estados do sistema

No modelo hipercubo, as transições entre estados ocorrem de modo idêntico aos modelos Markovianos clássicos de filas. Admite-se que apenas um servidor é despachado para atender um chamado, e que a probabilidade de chegarem dois chamados exatamente no mesmo instante, e a probabilidade de dois atendimentos terminarem exatamente no mesmo instante, são nulas. Em síntese, qualquer transição de um passo é permitida, enquanto todas as transições com mais de um passo não são (processos estocásticos de nascimento e morte).

O SAMU-192 possui 10 átomos e 10 ambulâncias; assim, o modelo é aplicado para uma situação em que  $N_A = 10$  e  $N = 10$ . O serviço em operação na cidade de Campinas permite que seus usuários entrem em uma fila, caso todas as ambulâncias estejam ocupadas no instante da ocorrência do chamado. Cada estado particular do sistema é representado por um vetor com  $N=10$  elementos, onde cada um pode assumir valores 0 ou 1, representando o estado de cada ambulância (0 se estiver disponível e 1, caso contrário). Logo, existem  $2^{10} = 1024$  estados possíveis no sistema, além dos 10 estados saturados,  $S_{11}, S_{12}, \dots, S_{20}$ , que definem a fila. A Figura 5 mostra os diagramas de fluxos de entrada nos estados 0000000000 (sistema vazio) e 1111111111 (sistema cheio), de acordo com a matriz de preferência de despachos apresentada na Tabela 6.



**Figura 5** – Diagramas de fluxos de entrada nos estados 0000000000 e 1111111111

De maneira análoga, podem ser considerados os fluxos para todos os estados intermediários do sistema. A Figura 6 ilustra para o estado intermediário 1000000101.

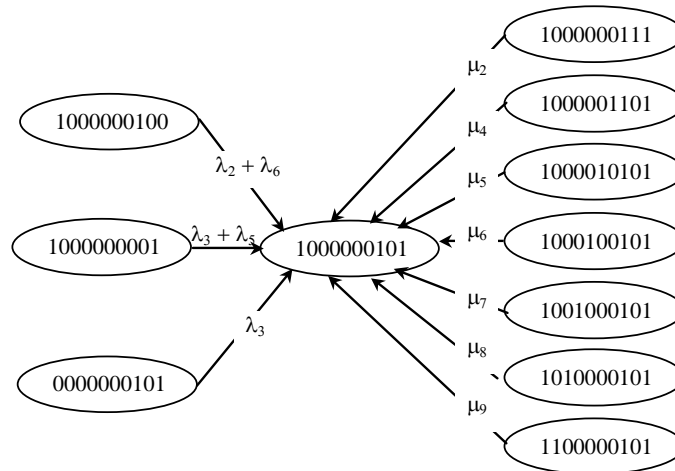


Figura 6 – Diagrama de fluxos de entrada no estado 1000000101

### 3.3 Equações de balanço

Para se descrever as equações de balanço dos estados, é preciso admitir a condição de equilíbrio para o sistema, ou seja, para todos os possíveis estados, o fluxo com que o sistema entra em tal estado deve ser igual ao fluxo com que ele sai deste estado. Considera-se a matriz de preferência de despacho dos servidores, uma vez que os veículos são despachados na ordem de preferência que aparecem nesta lista (pois as taxas de transição dependem da ordem em que os veículos são enviados para os átomos). O sistema de equações produzido para este estudo de caso resulta em 1024 equações lineares (ou seja, uma para cada estado), cujas incógnitas são suas probabilidades de equilíbrio do sistema. Por ilustração, seguem abaixo as equações de equilíbrio dos estados 0000000000, 1111111111 e 1000000101 (Figuras 5 e 6).

- Estado 0000000000:

$$\begin{aligned} \lambda \cdot p_{0000000000} &= \mu_1 \cdot p_{0000000001} + \mu_2 \cdot p_{0000000010} + \mu_3 \cdot p_{0000000100} + \mu_4 \cdot p_{0000001000} + \\ &+ \mu_5 \cdot p_{0000010000} + \mu_6 \cdot p_{0000100000} + \mu_7 \cdot p_{0001000000} + \mu_8 \cdot p_{0010000000} + \mu_9 \cdot p_{0100000000} + \\ &+ \mu_{10} \cdot p_{1000000000} \end{aligned}$$

- Estado 1111111111:

$$\begin{aligned} (\lambda + \mu) \cdot p_{1111111111} &= \lambda \cdot p_{1111111110} + \lambda \cdot p_{1111111101} + \lambda \cdot p_{1111111011} + \lambda \cdot p_{1111110111} + \\ &+ \lambda \cdot p_{1111101111} + \lambda \cdot p_{1111011111} + \lambda \cdot p_{1110111111} + \lambda \cdot p_{1101111111} + \lambda \cdot p_{1011111111} + \\ &+ \lambda \cdot p_{0111111111} + \mu p_{S_{11}} \end{aligned}$$

- Estado 1000000101:

$$\begin{aligned} &(\lambda + \mu_1 + \mu_3 + \mu_{10}) \cdot p_{1000000101} = \\ &= \mu_2 \cdot p_{1000000111} + \mu_4 \cdot p_{1000001101} + \mu_5 \cdot p_{1000010101} + \mu_6 \cdot p_{1000100101} + \mu_7 \cdot p_{1001000101} + \\ &+ \mu_8 \cdot p_{1010000101} + \mu_9 \cdot p_{1100000101} + (\lambda_2 + \lambda_6) \cdot p_{1000000100} + (\lambda_3 + \lambda_5) \cdot p_{1000000001} + \\ &+ \lambda_3 \cdot p_{0000000101} \end{aligned}$$

Da teoria de filas sabemos que o sistema acima é indeterminado. Uma forma de eliminar esta indeterminação é introduzindo uma equação de normalização, ou seja, considerando que a soma das probabilidades de todos os possíveis estados do sistema deve ser igual a 1. Devido à analogia com os modelos de filas clássicos M/M/N, considerando-se a política operacional onde se permite a formação de filas, facilmente pode-se estabelecer as seguintes relações para os estados que compõem a fila (maiores detalhes podem ser encontrados em Takeda, 2000):

$$\begin{cases} \lambda \cdot p_{1111111111} = \mu \cdot P(S_{11}) \\ \lambda \cdot P(S_{11}) = \mu \cdot P(S_{12}) \\ \vdots \\ \lambda \cdot P(S_{19}) = \mu \cdot P(S_{20}) \end{cases}$$

$$\text{com } p_{0000000000} + p_{0000000001} + \dots + \sum_{n=0}^{10} p^n \cdot p_{1111111111} = 1.$$

### 3.4 Solução do modelo

Sistemas que operam com  $N$  servidores possuem um sistema linear com  $2^N$  equações, cujo resultado são os valores das probabilidades de equilíbrio do sistema. Portanto, mesmo em casos onde  $N$  é pequeno, é essencial assegurar a viabilidade computacional de se resolver o problema. Tal sistema pode ser resolvido, por exemplo, pelo método de Gauss-Jordan, ou pelos métodos iterativos de Gauss-Seidel ou Gauss-Jacobi. Os métodos iterativos são métodos aproximados por natureza, pois sua implementação exige um critério de parada baseado no erro absoluto máximo tolerável. Além disso, existe a possibilidade de não convergirem. No presente trabalho, o sistema de equações que representa os possíveis estados do SAMU-192 foi solucionado utilizando-se o método de Gauss-Jordan que, apesar de ser mais custoso computacionalmente (se comparado com o método de Gauss-Seidel), não envolve condições adicionais para a obtenção da solução.

Larson (1975) propôs um método aproximado baseado na solução de um sistema não linear com apenas  $N$  equações. Este método pode ser usado em casos onde  $N$  é grande, e alternativamente aos métodos exatos, que podem ficar impraticáveis computacionalmente. O método aproximado de Larson difere do método exato, basicamente em dois aspectos: o primeiro é a suposição de que todos os servidores são homogêneos; o segundo é que a escolha do servidor a ser despachado para realizar o atendimento se dá de forma aleatória em um sistema M/M/N, enquanto que no método exato esta seleção se dá na seqüência estabelecida pela política de despachos do sistema (Chiyoshi *et al.*, 2001).

### 3.5 Resultados

Os resultados obtidos com a aplicação do modelo hipercubo mostraram sua eficácia e precisão para avaliar o SAMU-192. A probabilidade de se encontrar o sistema vazio ( $p_{0000000000}$ ), ou seja, com todas as ambulâncias disponíveis é pequena, em torno de 0,003, e a probabilidade do sistema estar saturado ( $p_s$ ) é de 0,09, isto é, cerca de 9% dos chamados encontram o sistema com todos os servidores ocupados mais 0, 1, 2, ..., ou 10 usuários em fila. A probabilidade de perda do sistema é de apenas 0,0001. Com a distribuição de



equilíbrio de estados computada, pode-se estimar diversas medidas de desempenho interessantes para o gerenciamento do sistema. Os resultados que seguem condizem com as condições operacionais originais do SAMU-192.

### **Carga média de trabalho**

A carga de trabalho (*workload*) de um servidor  $n$  ( $\rho_n$ ), representa a fração de tempo em que o servidor permanece ocupado, e pode ser facilmente obtida através da soma das probabilidades dos estados em que o servidor está ocupado, ou seja:

$$\begin{cases} \rho_1 = p_{00\dots01} + p_{00\dots011} + \dots + p_{11\dots11} + p_Q \\ \rho_2 = p_{00\dots010} + p_{00\dots011} + \dots + p_{11\dots11} + p_Q \\ \vdots \\ \rho_N = p_{100\dots000} + p_{100\dots001} + \dots + p_{11\dots11} + p_Q \end{cases} \quad (1)$$

onde  $p_Q$  é a probabilidade de formação de fila, e pode ser calculada como:  $p_Q = P(S_1) + P(S_2) + P(S_3) + \dots = 1 - (p_{00\dots00} + p_{00\dots10} + \dots + p_{11\dots11})$ .

No caso do SAMU-192, os resultados produzidos pelo modelo confirmaram as observações iniciais feitas pela coordenação. Os VSB's ficam, em média, 63% do tempo ocupados, e os VSA's, 39% do tempo ocupados. Quanto ao sistema como um todo, o valor obtido foi de 58% (compatível com as informações colhidas no sistema, de 60%). A Tabela 8 apresenta uma comparação entre as cargas médias de trabalho produzidas pelo modelo com as informações fornecidas pelo sistema para os VSA's e VSB's. Uma das vantagens deste cenário, onde todos os veículos permanecem centralizados, é o fato de existir um bom balanceamento da carga de trabalho das ambulâncias.

**Tabela 8** – Carga média de trabalho de cada ambulância

Ambulância	Carga média de trabalho		Desvio	
	SAMU-192	modelo	absolutos	%
1	0,40	0,39	-0,01	-0,03
2	0,40	0,39	-0,01	-0,03
3	0,65	0,63	-0,02	-0,03
4	0,65	0,63	-0,02	-0,03
5	0,65	0,63	-0,02	-0,03
6	0,65	0,63	-0,02	-0,03
7	0,65	0,63	-0,02	-0,03
8	0,65	0,63	-0,02	-0,03
9	0,65	0,63	-0,02	-0,03
10	0,65	0,63	-0,02	-0,03
<b>Média VSA</b>	<b>0,40</b>	<b>0,39</b>	<b>-0,01</b>	<b>-0,03</b>
<b>Média VSB</b>	<b>0,65</b>	<b>0,63</b>	<b>-0,02</b>	<b>-0,03</b>
<b>Sistema</b>	<b>0,60</b>	<b>0,58</b>	<b>-0,02</b>	<b>-0,03</b>

### Tempos médios de viagem

Estes valores são obtidos a partir da matriz origem-destino dos tempos de viagem entre todos os pares de átomos ( $\tau_{ij}$ ), apresentados na Tabela 5. Em geral, refletem a influência de fatores como configurações do tráfego local, presença de barreiras, horário, etc. Desta forma, o tempo gasto por um servidor qualquer para ir do átomo  $i$  até o átomo  $j$  não necessariamente coincide com o tempo gasto para ir de  $j$  para  $i$ ; isto é, nem sempre  $\tau_{ij} = \tau_{ji}$ .

#### a) Tempo médio de viagem para o sistema

Para calcular o tempo médio de viagem do sistema ( $\bar{T}$ ), é necessário conhecer a localização de um servidor, quando despachado para atender a um chamado, o tempo médio necessário para um servidor  $n$ , quando disponível, viajar até o átomo  $j$ , e o tempo médio de viagem para chamados sujeitos a espera em fila.

*Tempo médio de viagem para um servidor  $n$  deslocar-se até o átomo  $j$ :*

$$t_{nj} = \sum_{k=1}^{N_A} l_{nk} \cdot \tau_{kj} \quad n = 1, 2, \dots, N \quad e \quad j = 1, 2, \dots, N_A \quad (2)$$

*Tempo médio de viagem para chamados sujeitos a espera em fila (aproximado por):*

$$\bar{T}_Q = \sum_{i=1}^{N_A} \sum_{j=1}^{N_A} \frac{\lambda_i \cdot \lambda_j}{\lambda^2} \cdot \tau_{ij} \quad (3)$$

onde as razões  $\lambda_j/\lambda$  e  $\lambda_i/\lambda$  correspondem à probabilidade de um chamado que está em fila ter sido gerado no átomo  $j$ , e à probabilidade deste chamado ser atendido por um servidor localizado no átomo  $i$ , respectivamente.

*Frequências de despachos:*

$$f_{nj} = f_{nj}^{[1]} + f_{nj}^{[2]} = \underbrace{\frac{\lambda_j}{\lambda} \cdot \sum_{B \in E_{nj}} p_B}_{f_{nj}^{[1]}} + \underbrace{\frac{\lambda_j}{\lambda} \cdot p_S \cdot \frac{\mu_n}{\mu}}_{f_{nj}^{[2]}} \quad (4)$$

onde  $f_{nj}$  representa a fração de todos os despachos do servidor  $n$  ao átomo  $j$ ;  $f_{nj}^{[1]}$  corresponde à fração de todos os despachos do servidor  $n$  para o átomo  $j$  que não implica em tempo de espera para o usuário;  $f_{nj}^{[2]}$  representa a fração de todos os despachos do servidor  $n$  para o átomo  $j$  sujeitos a esperas pelo atendimento;  $E_{nj}$  é o conjunto dos estados nos quais um chamado do átomo  $j$  é atendido pelo servidor  $n$ , e  $p_s = p_Q + p_{11\dots 1}$  é a probabilidade de saturação do sistema. Assim, o tempo médio de viagem do sistema pode ser obtido por:

$$\bar{T} = \sum_{n=1}^N \sum_{j=1}^{N_A} f_{nj}^{[1]} \cdot t_{nj} + p_S \cdot \bar{T}_Q \quad (5)$$

Para o SAMU-192, este tempo foi calculado em 13,8 minutos, enquanto o valor observado na amostra foi de 14,5 minutos (diferença de 0,7 minutos, conforme pode ser observado na Tabela 5). O desvio com relação ao valor amostrado é de apenas 4,5%, o que é aceitável para validar o modelo hipercubo nesta aplicação, considerando, por exemplo, possíveis imprecisões nos dados de entrada e na operação do sistema.

b) *Tempo médio de viagem de cada ambulâncias*

Outro importante indicador de desempenho do sistema é o tempo médio de viagem de cada servidor  $n$ ,  $\overline{TU}_n$ , ou seja, quanto tempo, em média, cada equipe gasta para se deslocar até o local do chamado e iniciar o atendimento junto às vítimas, independentemente da região que solicitou o serviço.

$$\overline{TU}_n = \frac{\sum_{j=1}^{N_A} f_{nj}^{[1]} \cdot t_{nj} + \frac{\mu_n}{\mu} \cdot \overline{T_Q} \cdot p_S}{\sum_{j=1}^{N_A} f_{nj}^{[1]} + \frac{\mu_n}{\mu} \cdot p_S} \quad (6)$$

A Tabela 9 mostra estes tempos comparados aos obtidos na amostra. Os valores apontam para respostas mais rápidas quando o chamado é do tipo avançado. Isto se explica devido ao grau de urgência envolvido. Os desvios com relação aos tempos observados no sistema são da ordem de 3,3% para os VSA's e de 5,1% para os VSB's, com um desvio médio de 4,8%, se consideradas todas as ambulâncias. Estes desvios são relativamente pequenos, o que corrobora para a validação do hipercubo para analisar o presente sistema.

**Tabela 9** – Tempos médios de viagem das ambulâncias (cenário original)

Ambulância	Tempo médio de viagem (minutos)		Desvio	
	Modelo	Amostra	minutos	%
1	10,9	10,4	0,5	4,8
2	10,9	10,7	0,2	1,9
3	14,2	13,4	0,8	6,20
4	14,2	13,5	0,7	5,2
5	14,2	14,3	-0,1	-0,5
6	14,2	13,6	0,6	4,40
7	14,2	13,2	1,1	8,1
8	14,2	13,3	1,0	7,6
9	14,2	13,2	1,0	7,6
10	14,2	13,6	0,5	3,5
<b>Média VSA</b>	10,9	10,5	0,4	3,3
<b>Média VSB</b>	14,2	13,5	0,7	5,1

c) *Tempo médio de viagem aos átomos*

Outra maneira de olhar para o desempenho do sistema é através do tempo médio de viagem para cada átomo  $j$ ,  $\overline{T}_j$ , ou seja, em função da região onde estão as vítimas (em média, quanto tempo uma equipe qualquer do sistema gasta para se deslocar e iniciar os procedimentos médicos em cada região).

$$\overline{T}_j = \frac{\sum_{n=1}^N f_{nj}^{[1]} \cdot t_{nj}}{\sum_{n=1}^N f_{nj}^{[1]}} \cdot (1 - p_S) + \sum_{i=1}^{N_A} \left( \frac{\lambda_i}{\lambda} \right) \cdot \tau_{ij} \cdot p_S \quad (7)$$

Desconsiderando-se a região central, onde se localiza a base das ambulâncias, observa-se que as regiões geradoras de chamados básicos têm tempos muito semelhantes; o mesmo acontece para as regiões geradoras de chamados avançados. Isto se deve ao fato da localização da base do SAMU-192, na região Central, na prática, ser equidistante de todos os demais átomos do sistema. Os desvios relativos com relação aos valores observados no próprio sistema foram pouco significativos, em termos práticos, da ordem de 0,1% no caso dos átomos da classe avançada e 5,5% para a classe básica e um desvio médio para o sistema de apenas 3,0% (Tabela 10). Os desvios absolutos foram da ordem de 6,6% para os átomos da classe avançada, 9,0% para os átomos da classe básica e 7,9% para o sistema.

**Tabela 10** – Tempos médios de viagem para cada átomo (cenário original)

Átomo	Tempo médio de viagem (minutos)		Desvio	
	Modelo	Amostra	minutos	%
1 - NB	16,0	15,5	0,6	3,6
2 - NA	13,1	13,6	-0,5	-3,8
3 - SB	15,9	14,5	1,4	9,7
4 - SA	13,0	11,7	1,3	11,4
5 - LB	16,0	14,8	1,2	8,0
6 - LA	13,0	12,5	0,5	4,2
7 - OB	15,8	14,2	1,6	11,2
8 - OA	12,8	14,0	-1,2	-8,4
9 - CB	5,5	6,6	-1,1	-17,0
10 - CA	4,3	4,5	-0,2	-3,8
Média A	11,3	11,2	0,0	0,1
Média B	13,8	13,1	0,7	5,5

Outras medidas de desempenho foram computadas e podem ser encontradas em Takeda (2000). Os resultados mostraram que o modelo está razoavelmente calibrado, pois produziu desvios pouco significativos com relação aos dados reais, validando a hipótese inicial de que o hipercubo constitui uma ferramenta eficaz para auxiliar nas análises de planejamento e operação deste sistema.

### 3.6 Calibração dos tempos de atendimento

É comum observar em determinados serviços de emergência que a parcela relativa ao tempo de viagem dos servidores representa uma fração não desprezível do tempo total de atendimento, como pode ser o caso dos serviços de ambulância. Este tempo é composto pelo tempo de viagem até o local da ocorrência, tempo de viagem do local da ocorrência até um hospital ou mesmo domicílio das vítimas e tempo de viagem de retorno à base. Nestes casos, é necessário que o tempo de atendimento de cada servidor seja ajustado, a fim de refletir os vários fatores que o influenciam. Isto é feito através do processo de calibração de  $\mu_n^{-1}$ , utilizando para isto, a expressão do tempo médio de viagem do servidor  $n$ ,  $\overline{TU}_n$ .

O processo consiste em verificar se existe diferença significativa entre os tempos médios de atendimento admitidos como parâmetros de entrada no modelo e os respectivos tempos médios de atendimento gerados. Se a soma de todos os tempos que compõem o tempo médio de atendimento do servidor  $n$  for diferente de  $\mu_n^{-1}$ , admitido no início da modelagem, então, o modelo deve ser executado novamente, utilizando como parâmetro este novo tempo médio

de atendimento. Este procedimento deve ser repetido até que a diferença entre os tempos médios de atendimento admitidos como parâmetros e os tempos médios de atendimento gerados pelo modelo sejam suficientemente próximos. Experiências de diversos autores mostram que este procedimento costuma convergir em duas ou três iterações, para uma precisão razoável dos valores de  $\mu_n^{-1}$  (Larson & Odoni, 1981). Este procedimento é utilizado nos resultados da próxima seção.

#### 4. Avaliação de Cenários Alternativos

Em serviços de urgência como o SAMU-192, é óbvia a necessidade de se operar com tempos de resposta pequenos. Este tempo é importante do ponto de vista técnico, pois quanto mais cedo as vítimas começam a receber atendimento, maiores são suas chances de recuperação, e menores as probabilidades de agravamento dos casos. Na prática, o tempo de resposta inclui o tempo de preparação da equipe, o tempo de viagem até o local onde estão as vítimas, mais um eventual tempo de espera em fila (caso todas as ambulâncias estejam ocupadas). Como o tempo médio de preparação da equipe é muito pequeno (conforme discutido na seção 3), para sistemas onde a política operacional não admite a formação de filas, o tempo médio de resposta praticamente coincide com o tempo médio de viagem do sistema ( $\bar{T}$ ). Ao se permitir a formação de fila, o tempo médio de resposta passa a ser calculado considerando-se  $\bar{T}_Q$  e as probabilidades dos estados que representam a fila, por exemplo, para um único usuário em fila,  $\bar{T}_Q$  com probabilidade  $P(S_{11})$ , para dois usuários em fila,  $\bar{T}_Q$  com probabilidade  $P(S_{12})$ , e assim sucessivamente. Na configuração original do sistema, estas probabilidades são relativamente baixas (da ordem de  $10^{-3}$ ). Por simplicidade, nas análises que seguem, os impactos da descentralização e do aumento do número de ambulâncias são avaliados em termos dos tempos médios de viagem (ao invés de tempos médios de resposta).

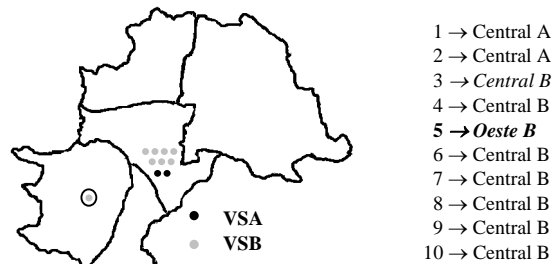
Vários são os fatores que podem influenciar no tempo de resposta a um chamado, dentre eles: dia da semana, período do dia, condições do tráfego local, etc, que muitas vezes fogem do controle do sistema, e outros, como capacitação profissional da equipe, número de veículos disponíveis e suas localizações, que podem ser planejados e dimensionados pela equipe de gerência. Uma das alternativas a ser avaliada é com relação aos efeitos provocados por um aumento no número de ambulâncias para realizar as operações. Esta opção nem sempre é a mais viável para ser implantada, pois envolve a disponibilidade de recursos financeiros para a aquisição de novos veículos, contratação de equipe especializada e manutenção de veículo e equipe. Por outro lado, opções menos custosas como, por exemplo, uma reconfiguração da estrutura já existente, podem causar efeitos importantes sobre o nível de serviço oferecido, sem exigir que sejam feitos grandes investimentos. Neste sentido, as primeiras alternativas operacionais apresentadas a seguir refletem os impactos da descentralização das ambulâncias já existentes no SAMU-192. Intuitivamente, as distâncias médias de viagem tornam-se menores; logo, os tempos médios de viagem também devem diminuir, implicando em respostas mais rápidas.

##### 4.1 Descentralização das ambulâncias

A legislação brasileira não impõe que a base de um SAMU deva ser junto a um hospital, pronto-socorro, ou qualquer outro tipo de unidade de saúde do município. Entretanto, ao se concentrar as equipes em uma unidade deste tipo, há o aproveitamento da infraestrutura existente, gerando uma economia que pode vir a ser aplicada para a manutenção e

melhoramento do sistema, além da possível colaboração de profissionais da unidade com as equipes de atendimento de emergência. Neste sentido, os cenários descentralizados apresentados a seguir consideram como bases as principais unidades de saúde do município, localizados em cada uma das regiões consideradas (Norte, Sul Leste Oeste e Central). Várias alternativas operacionais de descentralização das ambulâncias do sistema foram estudadas, desde o simples reposicionamento de uma única ambulância, até cenários com descentralização total. O primeiro cenário é descrito em detalhes; os demais são similares e, portanto, são descritos mais resumidamente por motivo de economia de espaço. Para mais detalhes deste cenário, o leitor pode consultar Takeda (2000).

**Cenário 1: descentralização de uma ambulância**



**Figura 7** – Primeiro cenário (1 VSB em OB)

Neste primeiro cenário estudou-se a alternativa mais simples de descentralização. Nele, um VSB é localizado na região de maior demanda do sistema (Tabela 4): o VSB 5 foi escolhido arbitrariamente (dentro os oito VSB's do sistema) e posicionado na região OB (Figura 7). Os demais VSB's e os dois VSA's permanecem centralizados na base. No caso de um chamado básico do átomo OB, o VSB 5 é o primeiro a ser chamado para realizar o atendimento, seguido dos demais VSB's (escolhidos aleatoriamente), por sua vez seguidos dos dois VSA's (escolhidos aleatoriamente). No caso de um chamado avançado do átomo OA, os seus servidores primários continuam sendo os veículos 1 e 2 (escolhidos aleatoriamente); caso estes dois VSA's estejam ocupados, o VSB 5 é o primeiro *backup* a ser chamado, seguido dos demais VSB's (escolhidos aleatoriamente). Nos demais átomos a política de prioridade continua igual à configuração original do sistema. Devido à nova organização do sistema, as hipóteses (5) e (7) da modelagem são modificadas. Para exemplificar, serão apresentadas as modificações ocorridas nestas hipóteses apenas para este cenário. A matriz de localização dos servidores é obtida facilmente para esta configuração:

$$L = \begin{matrix} & \begin{matrix} \text{NB} & \text{NA} & \text{SB} & \text{SA} & \text{LB} & \text{LA} & \text{OB} & \text{OA} & \text{CB} & \text{CA} \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \end{matrix} & \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \end{matrix}$$

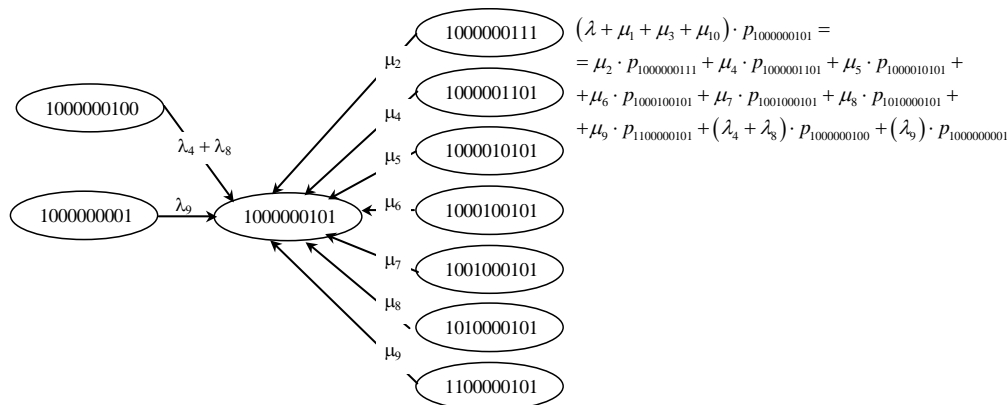
Para implementar a política de despacho descrita acima, optou-se por continuar usando o método de geração aleatória de matrizes de Burwell *et al.* (1993), agora levando-se em consideração a localização descentralizada do VSB 5. A seguir, apresenta-se, na Tabela 11, um exemplo de lista de preferência de despachos (gerada aleatoriamente), que reflete os efeitos da nova estrutura operacional, onde o VSB 5 é alocado como servidor primário para o átomo 7 (OB), e primeiro *backup* básico para o átomo OA. Os demais átomos continuam sendo atendidos, com prioridade, pelos veículos posicionados na região Central.

**Tabela 11** – Matriz de preferência de despachos (cenário 1)

Átomo	Preferência de despacho									
	1 <sup>a</sup>	2 <sup>a</sup>	3 <sup>a</sup>	4 <sup>a</sup>	5 <sup>a</sup>	6 <sup>a</sup>	7 <sup>a</sup>	8 <sup>a</sup>	9 <sup>a</sup>	10 <sup>a</sup>
1 - N <sub>B</sub>	7	3	4	8	6	10	9	5	1	2
2 - N <sub>A</sub>	2	1	7	4	8	3	6	10	9	5
3 - S <sub>B</sub>	4	3	8	10	9	6	7	5	2	1
4 - S <sub>A</sub>	1	2	3	4	8	10	9	6	7	5
5 - L <sub>B</sub>	6	8	4	3	7	9	10	5	1	2
6 - L <sub>A</sub>	2	1	6	4	3	8	7	10	9	5
7 - O <sub>B</sub>	5	10	9	8	4	3	6	7	2	1
8 - O <sub>A</sub>	1	2	5	10	9	8	3	4	6	7
9 - C <sub>B</sub>	3	4	8	6	7	9	10	5	1	2
10 - C <sub>A</sub>	2	1	3	8	4	6	7	9	10	5

**veículos**

A Figura 8 exemplifica o diagrama de fluxos considerando o estado intermediário 1000000101 nesta nova configuração operacional, e sua equação de balanço. Os demais estados e suas respectivas equações podem ser obtidos de maneira semelhante. As modificações sobre as hipóteses para os demais cenários adiante e as adaptações sobre seus diagramas de fluxos e suas equações de equilíbrio também são semelhantes.



**Figura 8** – Diagrama de fluxos para o estado 1000000101 e sua correspondente equação de balanço para o cenário 1

A Tabela 12 apresenta algumas medidas de desempenho para esta configuração e os compara aos valores obtidos pelo modelo para a configuração original do sistema. Observou-se valores semelhantes de cargas médias de trabalho nos dois cenários. O VSB 5 apresentou um fator de utilização um pouco mais elevado que os demais (66%) em virtude de ter sido posicionado na região Oeste, que representa o maior índice de solicitações de chamados no sistema. O tempo médio de viagem das ambulâncias diminuiu suavemente, em particular no caso do VSB 5 localizado em OB, cuja redução é significativa (38,0%), respondendo de forma mais rápida às solicitações desta região. O tempo médio de viagem do sistema sofreu uma redução de 4,4% (de 13,6 para 13,0 minutos), correspondendo a 0,9% para os VSA's e 5,0% para os VSB's (Tabela 14). O tempo médio de viagem aos átomos sofreu uma suave redução: nas áreas geradoras de chamados básicos a redução média foi de 3,6%, não sendo observadas reduções nas demais regiões (Tabela 15).

**Tabela 12** – Cargas médias de trabalho e tempos médios de viagem para cada ambulância

Ambulância	Cenário original		Cenário 1		Desvio	
	Carga média de trabalho	Tempo médio de viagem (min)	Carga média de trabalho	Tempo médio de viagem (min)	minutos	%
1	0,39	10,9	0,39	11,3	0,4	3,7
2	0,39	10,9	0,41	10,6	-0,3	-2,8
3	0,63	14,2	0,64	14,2	0,0	0,0
4	0,63	14,2	0,63	14,2	0,0	0,0
5	0,63	14,2	0,66	8,8	-5,4	-38,0
6	0,63	14,2	0,64	14,2	0,0	0,0
7	0,63	14,2	0,65	14,2	0,0	0,0
8	0,63	14,2	0,65	14,2	0,0	0,0
9	0,63	14,2	0,64	14,2	0,0	0,0
10	0,63	14,2	0,63	14,2	0,0	0,0
<b>Média VSA</b>	<b>0,39</b>	<b>10,9</b>	<b>0,40</b>	<b>11,0</b>	<b>0,1</b>	<b>0,9</b>
<b>Média VSB</b>	<b>0,63</b>	<b>14,2</b>	<b>0,64</b>	<b>13,5</b>	<b>-0,7</b>	<b>-5,0</b>
<b>Sistema</b>	<b>0,58</b>	<b>13,6</b>	<b>0,58</b>	<b>13,0</b>	<b>-0,6</b>	<b>-4,4</b>

### *Cenário 2: descentralização de duas ambulâncias*

Considerando os resultados obtidos deslocando-se uma única ambulância, neste cenário avaliou-se a descentralização de dois VSB's. Várias são as alternativas de localização para estas ambulâncias, porém, a experiência da equipe do SAMU-192, aliada aos altos índices de demanda apresentados por OB, posicionou os dois VSB's também à região Oeste. As cargas médias de trabalho dos veículos continuaram balanceadas, e com valores altos, sendo que os veículos que foram deslocados para OB (veículos 3 e 5) apresentaram uma suave redução, causada pela prioridade em atender chamados desta região específica. O tempo médio de viagem dos VSB's descentralizados foi menor (redução média de 32,4% com relação ao cenário original), o que implica em respostas mais rápidas nesta região particular. Para o sistema como um todo, o tempo médio de viagem das ambulâncias reduziu de 13,6 para 12,5 minutos. Reduções semelhantes foram observadas para os tempos médios de viagem aos átomos. As Tabelas 14 e 15 resumem estes resultados.



### ***Cenário 3: descentralização de três ambulâncias***

Este cenário considera a estrutura proposta no cenário 2 e desloca uma nova ambulância para a região NB (VSB 3 em NB, e VSB's 4 e 6 em OB). Embora o número de solicitações de NB seja praticamente igual ao observado em LB (conforme apresentado na seção 3), o posicionamento deste novo VSB em NB é justificado pelo fato de que, nesta região, encontra-se localizado um grande centro hospitalar de referência do município. Os resultados obtidos foram próximos dos obtidos nos cenários 1 e 2 (Figura 11 e Tabelas 14 e 15). Os tempos médios de viagem das ambulâncias e para os átomos foram menores, se comparados ao cenário original. O tempo médio de viagem para o sistema como um todo diminuiu aproximadamente 12% com relação ao cenário original.

### ***Cenário 4: descentralização de quatro ambulâncias***

Estendendo as análises, este cenário apresenta os resultados de uma estrutura operacional onde cada região passou a ser base de ao menos um VSB. Os átomos NB, SB, LB e OB tornaram-se bases de um VSB (VSB 3 em NB; VSB 5 em SB; VSB 6 em OB; VSB 7 em LB), e os outros quatro veículos permaneceram centralizados na base do SAMU-192. As cargas médias de trabalho dos VSB's sofreram certo desbalanceamento, que pode ser explicado pelas diferenças entre as demandas de cada região, especificamente. Quanto aos tempos médios de viagem das ambulâncias, a redução foi sensível em todos os casos, em relação ao cenário original. O tempo médio de viagem para o sistema passou de 13,6 para 11,7 minutos (Tabela 14). Os tempos médios de viagem aos átomos reduziram-se em proporções semelhantes (Tabela 15). Esta opção mostrou-se atraente, ficando um pouco prejudicada apenas por causa do desbalanceamento das cargas médias de trabalho das ambulâncias.

### ***Cenário 5: descentralização de cinco ambulâncias***

O cenário 4 apresentou bons resultados nas análises de descentralização. Entretanto, as principais medidas de desempenho para a região OB foram um pouco inferiores se comparadas às apresentadas pelas demais regiões. Isto se explica principalmente pelo fato de ser a região de maior demanda do sistema. Assim, o cenário 5 amplia a estrutura do cenário 4, posicionando um outro veículo (VSB 9) em OB. Houve, então, um melhor balanceamento das cargas médias de trabalho (Tabela 13), e o tempo médio de viagem do sistema passou de 11,7 (cenário 4) para 11,5 minutos (Tabelas 14 e 15).

### ***Cenário 6: descentralização de seis ambulâncias***

Os resultados até aqui obtidos motivaram a configuração operacional do cenário 6, que combina as configurações dos cenários 3 e 5, respeitando-se as variações da demanda em cada região e a viabilidade de se constituir a região NB como base de mais uma ambulância. Neste caso, a nova distribuição dos VSB's foi: VSB's 3 e 10 em NB; VSB's 6 e 7 em OB; VSB 4 em LB; VSB 5 em SB. Nesta configuração, houve um bom balanceamento das cargas médias de trabalho (Tabela 13) e o tempo médio de viagem para o sistema reduziu cerca de 16,2% com relação ao cenário original (Tabelas 14 e 15).

Nota-se que os diversos cenários de descentralização provocaram melhorias em algumas medidas de desempenho, que refletem diretamente a qualidade do serviço oferecido. Este fato justificou outras investigações nesta mesma direção. A descentralização de sete ambulâncias, deixando a região Central somente com um VSB e, por fim, a descentralização

dos oito VSB's, ficando apenas os dois VSA's na base do SAMU-192, conforme sugerido inicialmente pela equipe do SAMU-192. Estes cenários geraram resultados pouco satisfatórios, sempre inferiores aos dos cenários 1 a 6 (isto se explica pelo fato de que a região Central é responsável pela segunda maior taxa de solicitações de serviço neste sistema). Questões sobre o reposicionamento de VSA's também foram levantadas. Entretanto, são condições pouco viáveis na prática do SAMU-192, pois todo atendimento realizado por uma equipe avançada possui um médico acompanhando, e devido à escassez de recursos, este profissional também desempenha a função de médico regulador das urgências na base Central do SAMU-192.

#### **4.2 Aumento do número de ambulâncias**

Verificadas as diferenças nos tempos médios de viagem por meio de descentralização, foi também desenvolvido um estudo sobre os impactos do aumento do número de ambulâncias, partindo-se da configuração operacional proposta no cenário 6, que apresentou bons resultados.

##### ***Cenário 7: sistema operando com um novo VSB***

Este cenário posiciona um novo VSB (número 11) na região Central. Esta alternativa operacional se baseou nas variações da demanda do sistema. As cargas médias de trabalho diminuíram, em média, 17,9% em relação ao cenário 6 (Tabela 13); os tempos médios de viagem das ambulâncias reduziram em torno de 14% para os VSB's 9% para os VSA's, se comparados aos valores observados no cenário 6 (Tabela 14). Os tempos médios de viagem aos átomos também diminuíram: 3,6% para as regiões geradoras de chamados avançados e 9,5% para as regiões geradoras de chamados básicos (Tabela 15). O tempo médio de viagem do sistema passou para 10,3 minutos.

##### ***Cenário 8: sistema operando com dois novos VSB's***

Considerando que as regiões Leste e Sul ainda operavam como bases de apenas uma equipe básica, este cenário trabalhou com a alocação de um novo VSB (número 12) em LB. Esta escolha foi feita levando-se em conta as demandas de LB e SB. A região SB é, em síntese, de característica industrial, com baixa densidade demográfica. Observou-se, neste caso, redução e balanceamento das cargas médias de trabalho das ambulâncias. Comparadas ao cenário 6, as cargas médias de trabalho apresentaram-se 30% inferiores (Tabela 13). Os tempos médios de viagem das ambulâncias e para os átomos reduzem de maneira significativa (Tabelas 14 e 15).

##### ***Cenário 9: sistema operando com dois novos VSB's e um novo VSA***

O cenário 9 estende o cenário 8, considerando mais um novo VSA (veículo 13). Discussões com a coordenação do sistema indicaram a região Oeste como a receptora desta nova ambulância. Apesar das exigências para a operação de um VSA (conforme comentário na seção 4), a região Oeste também representa o maior índice de solicitação de chamados do tipo avançado (Tabela 4). Portanto, com base nos resultados obtidos no cenário 8 com relação aos tempos de viagem (e conseqüentes tempos de resposta), a melhor estratégia operacional seria localizar o novo VSA nesta região. As Tabelas 13, 14 e 15 (e as Figuras 9, 10 e 11) apresentam os resultados obtidos e as respectivas reduções nas medidas de desempenho.

### 4.3 Comparação entre as alternativas operacionais

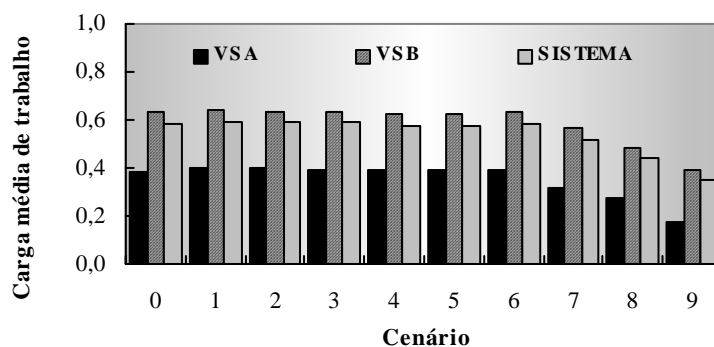
Para sintetizar as análises acima descritas e facilitar as comparações entre os diversos cenários investigados, as tabelas e figuras seguintes apresentam as principais medidas de desempenho obtidas.

#### *Carga média de trabalho*

A Tabela 13 e a Figura 9 mostram as variações das cargas médias de trabalho para os nove cenários estudados; apresentam também os resultados referentes à configuração original do sistema (denominado “Cenário 0”). Como esperado, as cargas médias de trabalho mostram-se sensíveis principalmente ao aumento do número de ambulâncias no sistema (Cenários 7, 8 e 9).

**Tabela 13** – Cargas médias de trabalho das ambulâncias para os nove cenários

Ambulância	Carga média de trabalho									
	Cenários									
	0	1	2	3	4	5	6	7	8	9
1	0,39	0,39	0,39	0,39	0,39	0,39	0,39	0,31	0,27	0,17
2	0,39	0,41	0,40	0,40	0,39	0,39	0,39	0,34	0,27	0,18
3	0,63	0,64	0,60	0,61	0,64	0,64	0,58	0,51	0,44	0,41
4	0,63	0,63	0,62	0,64	0,60	0,64	0,65	0,58	0,49	0,42
5	0,63	0,66	0,60	0,62	0,59	0,60	0,62	0,56	0,50	0,42
6	0,63	0,64	0,64	0,64	0,70	0,60	0,64	0,62	0,45	0,45
7	0,63	0,65	0,65	0,65	0,64	0,60	0,62	0,65	0,43	0,44
8	0,63	0,65	0,64	0,65	0,60	0,65	0,65	0,55	0,54	0,40
9	0,63	0,64	0,63	0,65	0,62	0,61	0,67	0,57	0,54	0,42
10	0,63	0,63	0,64	0,64	0,61	0,65	0,61	0,50	0,45	0,42
11								0,55	0,44	0,41
12									0,48	0,42
13										0,17
<b>Média VSA</b>	<b>0,39</b>	<b>0,40</b>	<b>0,40</b>	<b>0,40</b>	<b>0,39</b>	<b>0,39</b>	<b>0,39</b>	<b>0,32</b>	<b>0,27</b>	<b>0,17</b>
<b>Média VSB</b>	<b>0,63</b>	<b>0,64</b>	<b>0,63</b>	<b>0,64</b>	<b>0,63</b>	<b>0,63</b>	<b>0,63</b>	<b>0,56</b>	<b>0,48</b>	<b>0,42</b>
<b>Sistema</b>	<b>0,58</b>	<b>0,59</b>	<b>0,58</b>	<b>0,59</b>	<b>0,58</b>	<b>0,58</b>	<b>0,58</b>	<b>0,51</b>	<b>0,44</b>	<b>0,33</b>



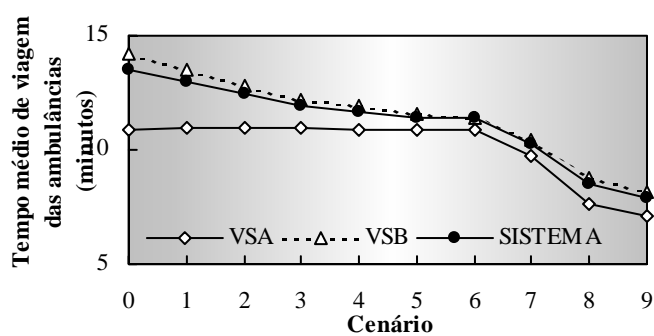
**Figura 9** – Comparação entre as cargas médias de trabalho dos diversos cenários

**Tempo médio de viagem de cada ambulância**

Sistemas descentralizados produzem tempos de viagem menores, e na prática, isto representa níveis de serviços mais elevados (Tabela 14 e Figura 10). Estas observações são também verificadas se o número de ambulâncias aumenta.

**Tabela 14** – Tempos médios de viagem das ambulâncias para os nove cenários

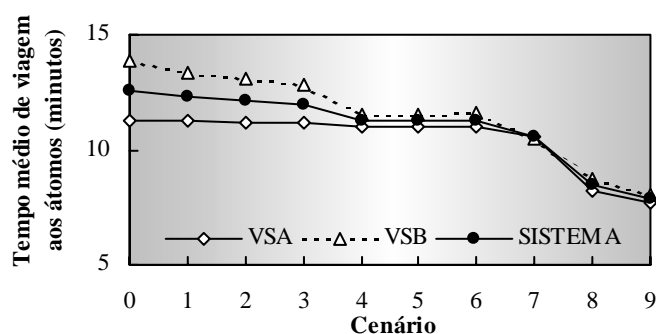
Ambulância	Tempos médios de viagem (minutos)									
	Cenários									
	0	1	2	3	4	5	6	7	8	9
1	10,9	11,3	11,4	11,3	10,9	10,9	10,9	10,0	7,7	7,5
2	10,9	10,6	10,5	10,6	10,9	10,9	10,8	9,4	7,6	7,1
3	14,2	14,2	10,0	9,0	11,7	11,7	11,6	11,0	8,9	8,8
4	14,2	14,2	13,9	9,7	12,4	12,8	11,3	9,9	9,2	8,4
5	14,2	8,8	9,2	13,8	10,7	10,9	11,4	9,5	9,1	7,7
6	14,2	14,2	14,0	9,7	12,0	10,5	10,5	9,6	8,9	7,9
7	14,2	14,2	14,1	13,9	9,9	12,0	10,4	9,3	8,9	7,8
8	14,2	14,2	14,0	14,0	12,7	12,0	12,7	11,4	8,1	8,1
9	14,2	14,2	13,8	13,8	13,0	10,5	12,6	11,2	8,2	8,2
10	14,2	14,2	13,8	13,9	12,9	12,1	11,6	11,2	8,8	8,7
11								10,5	8,5	8,2
12									8,7	8,2
13										6,6
Média VSA	10,9	11,0	11,0	11,0	10,9	10,9	10,9	9,7	7,7	7,1
Média VSB	14,2	13,5	12,9	12,2	11,9	11,6	11,5	10,4	8,7	8,2
Sistema	13,6	13,0	12,5	12,0	11,7	11,5	11,4	10,3	8,6	7,9

**Figura 10** – Tempos médios de viagem das ambulâncias**Tempo médio de viagem aos átomos**

Outra medida de tempo de resposta importante é o tempo médio de viagem aos átomos. No SAMU-192, esta medida reflete quanto tempo, em média, um chamado numa região qualquer do sistema espera para receber o atendimento (Tabela 15 e Figura 11).

**Tabela 15** – Tempos médios de viagem aos átomos para os nove cenários

Átomo	Tempo médio de viagem (minutos)									
	Cenário									
	0	1	2	3	4	5	6	7	8	9
1 - NB	16,0	16,2	16,5	16,2	13,0	13,3	11,0	10,0	7,7	7,5
2 - NA	13,1	13,2	13,2	15,2	12,7	12,8	12,5	12,1	8,7	8,2
3 - SB	15,9	15,9	15,9	15,8	12,4	12,5	13,2	12,2	9,3	9,1
4 - SA	13,0	13,1	13,1	12,9	12,6	12,6	12,7	12,3	9,5	8,7
5 - LB	15,9	16,1	16,4	16,2	12,9	13,3	13,1	12,3	10,5	8,9
6 - LA	13,0	13,1	13,2	13,1	12,7	12,7	12,7	12,3	10,8	9,5
7 - OB	15,8	12,5	10,2	9,8	13,3	11,1	11,5	10,8	9,8	8,6
8 - OA	12,8	12,4	12,1	12,6	12,5	12,2	12,3	12,1	8,6	8,6
9 - CB	5,5	5,8	6,2	6,1	5,9	7,6	9,0	7,1	6,3	6,3
10 - CA	4,3	4,5	4,5	4,0	4,3	4,7	4,9	4,1	3,4	3,4
Média VSA	11,3	11,3	11,2	11,2	11,0	11,0	11,0	10,6	8,2	7,7
Média VSB	13,8	13,3	13,0	12,8	11,5	11,5	11,6	10,5	8,7	8,1
Sistema	12,5	12,3	12,1	12,0	11,3	11,3	11,3	10,5	8,5	7,9

**Figura 11** – Tempos médios de viagem aos átomos

Outras medidas de desempenho também foram consideradas na análise. Por exemplo, a proporção de despachos de ambulâncias inter-átomos e a fração de chamados avançados que foram atendidos por equipes básicas (*backup*). À medida que esta fração aumenta, os usuários que necessitam de suporte avançado de atendimento estão recebendo assistência médico-emergencial de qualidade inferior. Analogamente às conclusões referentes aos tempos médios de viagem, sistemas parcialmente descentralizados apresentam frações menores que as observadas na configuração original do sistema (cenário 0). Por exemplo, para a configuração original esta fração resultou em 19%, enquanto para o cenário 9, este valor reduziu para 5%. Para mais detalhes destas análises, o leitor pode consultar Takeda (2000).

## 5. Conclusões e Perspectivas

O modelo hipercubo de filas espacialmente distribuídas é uma ferramenta precisa e eficaz para modelar sistemas urbanos de atendimento médico de urgência, como é o caso do SAMU-192 de Campinas. Os resultados por ele produzidos podem auxiliar tanto no planejamento quanto em ações diretas sobre a operação do sistema. A aplicação do modelo no SAMU-192 mostrou que a descentralização de ambulâncias é uma alternativa eficaz para se elevar o nível de serviço oferecido por este sistema. Se, por um lado, é óbvio que melhorias são obtidas com ações que reduzem os tempos de viagem, e conseqüentemente o tempo de resposta a um chamado, quantificar com precisão tais melhorias (por meio de diversos índices de desempenho) não é um processo simples, e o modelo hipercubo de filas é um instrumento eficiente para isso. Os resultados obtidos com os cenários aqui estudados mostraram principalmente que:

i) O simples reposicionamento de uma das ambulâncias para a região de maior índice de demanda (Cenário 1) provocou reduções sensíveis nos tempos médios de viagem. Os impactos foram particularmente sentidos na região que passou a ser a nova base da ambulância (conforme mostrado na Seção 4);

ii) Aumentando-se o nível de descentralização, os resultados tornaram-se cada vez melhores (por exemplo, cenários 2 e 3). Alternativas operacionais onde cada região tornou-se base de ao menos uma ambulância também produziram impactos significativos nos tempos de viagem (cenários 4, 5 e 6). Os seis primeiros cenários oferecem condições para o tomador de decisões optar por configurações operacionais não tão complexas, economicamente viáveis, e que produzem melhores níveis de serviços aos usuários do que a configuração original. Ainda neste contexto, pode-se afirmar que os indicadores de desempenho do sistema foram melhorando, gradativamente, porém até certo ponto, conforme discutido no cenário 6. Alternativas de descentralização de sete ou oito equipes básicas não produziram resultados satisfatórios, o que mostra que a melhor configuração não necessariamente é a descentralização total, como sugerido inicialmente pela equipe do SAMU-192.

iii) O aumento do número de ambulâncias no sistema, como esperado, provocou impactos significativos sobre as medidas de desempenho, tornando-as próximas dos padrões internacionais de atendimento (Volz, 1971 e Fujiwara *et al.*, 1987). Entretanto, tratam-se de alternativas que exigem investimentos financeiros não só para suas implantações, mas também para a manutenção das novas equipes.

A experiência da aplicação do modelo hipercubo no sistema SAMU-192 de Campinas contribuiu para difundir, para a Rede Brasileira de Cooperação em Emergências (RBCE), a importância de conciliar a experiência da classe médica com o uso de ferramentas analíticas para auxiliar nas decisões estratégicas do atendimento emergencial de saúde à população brasileira. Este trabalho também mostrou a importância de pesquisas e métodos para a análise quantitativa do desempenho operacional dos serviços de atendimento médico de urgência. Neste sentido, algumas propostas para estudos futuros surgiram.

Uma delas é a remodelagem das regiões do sistema SAMU-192, uma vez que a subdivisão utilizada pela Secretaria Municipal de Saúde (como sendo de cobertura dos Centros de Saúde) pode não refletir as reais necessidades de demanda de um serviço de atendimento de urgência. Considerando as variações de demanda observadas em Campinas, um estudo particular sobre a partição da região Oeste (Oeste 1 e Oeste 2, por exemplo) também poderia ser interessante. Em nível estadual, seriam interessantes pesquisas com a aplicação da

modelagem em outros serviços, considerando cenários de cidades médias e pequenas, e também estudos caracterizando serviços de outras regiões brasileiras. Em outra direção ainda pouco explorada na literatura, seriam interessantes pesquisas sobre a variabilidade das medidas de desempenho geradas pelo modelo hipercubo, visto que em geral estas medidas são caracterizadas apenas pelas médias das distribuições de equilíbrio.

### **Agradecimentos**

Os autores agradecem aos dois revisores anônimos pelos úteis comentários e sugestões, e à equipe do SAMU-192 de Campinas, em especial à Dra. Arine Campos O. Assis (coordenadora do sistema no período em que este estudo foi desenvolvido), pelo apoio à realização deste trabalho. Esta pesquisa contou com financiamento do CNPq (Processos 800773/91-8, 522973/95-7).

### **Referências Bibliográficas**

- (1) Anderson, L.R. & Fontenot, R.A. (1992). Optimal positioning of service units along a coordinate line. *Transportation Science*, **26**(4), 346-51.
- (2) 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.
- (3) Aubin, J. (1992). Scheduling ambulances. *Interfaces*, **22**(2), 1-10.
- (4) Ball, M.O. & Lin, F.L. (1993). A reliability model applied to emergency service vehicle location. *Operations Research*, **41**(1), 18-36.
- (5) Batta, R.; Dolan, J.M. & Krishnamurthy, N.N. (1989). The maximal expected covering location problem: revisited. *Transportation Science*, **23**(4), 277-87.
- (6) Berman, O. & Larson, R.C. (1982). The median problem with congestion. *Computers and Operations Research*, **9**(2), 119-26.
- (7) Brandeau, M. & Larson, R.C. (1986). Extending and applying the hypercube queueing model to deploy ambulances in Boston. **In:** *Delivery of Urban Services* [edited by A.J. Swersey and E.J. Ignall], **22**, TIMS Studies in the Management Science, Elsevier, 121-53.
- (8) Brotcorne, L.; Laporte, G. & Semet, F. (2003). Ambulance location and relocation models. *European Journal of Operational research*, **147**, 451-63.
- (9) Burwell, T.H.; Jarvis J.P. & McKnew, M.A. (1993). Modeling co-located servers and dispatch ties in the hypercube model. *Journal of Computers and Operations Research*, **20**(2), 113-120.
- (10) Chelst, K.R. & Barlach, Z. (1981). Multiple unit dispatches in emergency services: models to estimate system performance. *Management Science*, **27**(12), 1390-409.
- (11) Chiyoshi, F.; Galvão, R.D. & Morabito, R. (2000). O uso do modelo hipercubo na solução de problemas de localização probabilísticos. *Gestão & Produção*, **7**(2), 146-74.

- (12) Chiyoshi, F.; Galvão, R.D. & Morabito, R. (2001). Modelo hipercubo: análise e resultados para o caso de servidores não-homogêneos. *Pesquisa Operacional*, **21**(2), 199-218.
- (13) Chiyoshi, F.; Galvão, R.D. & Morabito, R. (2002). A note on solutions to the maximal expected covering location problem. *Computers & Operations research*, **30**, 87-96.
- (14) Daskin, M.S. (1983). A maximum expected covering location model: formulation, properties and heuristic solution. *Transportation Science*, **17**(1), 48-70.
- (15) Daskin, M.S. & Stern, E.H. (1981). A hierarchical objective set covering model for emergency medical service vehicle deployment. *Transportation Science*, **15**, 137-152.
- (16) Fujiwara, O.; Makjamroen, T. & Gupta, K.K. (1987). Ambulance deployment analysis: a case study of Bangkok. *European Journal of Operational Research*, **31**, 9-18.
- (17) Galvão, R.D. (1993). The use of Lagrangean relaxation in the solution of uncapacitated facility location problems. *Location Science*, **1**, 57-79.
- (18) Galvão, R.D.; Chiyoshi, F. & Morabito, R. (2003a). Towards unified formulations and extensions of two classical probabilistic location models. *Computers & Operations Research*. /no prelo/
- (19) Galvão, R.D.; Chiyoshi, F.Y.; Espejo, L.G.A. & Rivas, M.P.A. (2003b). Solução do problema de localização de máxima disponibilidade utilizando o modelo hipercubo. *Pesquisa Operacional*, **23**(1), 61-78.
- (20) Gendreau, M.; Laporte, G. & Semet, F. (1997). Solving an ambulance location model by Tabu search. *Location Science*, **5**, 75-88.
- (21) Gendreau, M.; Laporte, G. & Semet, F. (2001). A dynamic model and parallel Tabu search heuristic for real-time ambulance relocation. *Parallel Computing*, **27**, 1641-53.
- (22) Gonçalves, M.B. (1994). Métodos de P.O. em serviços emergenciais. **In:** SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL, 26, Florianópolis, SC, 1994. *Anais*. Florianópolis, **1**, 597-601.
- (23) Gonçalves, M.B.; Novaes, A.G. & Albino, J.C.C. (1994). Modelos para localização de serviços emergenciais em rodovias. **In:** SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL, 26, Florianópolis, SC, 1994. *Anais*. Florianópolis, **1**, 591-596.
- (24) Gonçalves, M.B.; Novaes, A.G. & Schmitz, R. (1995). Um modelo de otimização para localizar unidades de serviços emergenciais em rodovias. **In:** CONGRESSO DE PESQUISA E ENSINO EM TRANSPORTES, 9, São Carlos, SP, 1995. *Anais*. São Carlos, **3**, 962-972.
- (25) Green, L. & Kolesar, P. (1984). A comparison of the multiple dispatch and M/M/C priority queueing models of police patrol. *Management Science*, **30**(6), 665-670.
- (26) Halpern, J. (1977). The accuracy of estimates for the performance criteria in certain emergency service queueing systems. *Transportation Science*, **11**(3), 223-242.
- (27) Jarvis, J. (1975). Optimization in stochastic systems with distinguishable servers. *Technical Report*, 19-75.
- (28) Kaplan, E.H. (1979). Evaluating the effectiveness of one-officer versus two officer patrol units. *Journal of Criminal Justice*, **7**, 325-55.



- (29) Kolesar, P. & Swersey, A.J. (1986). The deployment of urban emergency units: a survey. **In: *Delivery of Urban Services*** [edited by A.J. Swersey and E.J. Ignall], **22**, TIMS Studies in the Management Science, Elsevier, 87-119.
- (30) Larson, R.C. (1974). A hypercube queueing model for facility location and redistricting in urban emergency services. *Computers and Operations Research*, **1**, 67-95.
- (31) Larson, R.C. & Odoni, A.R. (1981). *Urban Operations Research*. Prentice Hall, New Jersey.
- (32) Louveaux, F. (1993). Stochastic location analysis. *Location Science*, **1**, 127-154.
- (33) Marianov, V. & Serra, D. (1998). Probabilistic, maximal covering location-allocation models for congested systems. *Journal of Regional Science*, **38**(3), 401-424.
- (34) Mendonça, F.C. & Morabito, R. (2000). Aplicação do modelo hipercubo para análise de um sistema médico-emergencial em rodovia. *Gestão & Produção*, **7**(1), 73-91.
- (35) Mendonça, F.C. & Morabito, R. (2001). Analysing emergency medical service ambulance deployment on a Brazilian highway using hypercube queueing model. *Journal of the Operational Research Society*, **52**, 261-70.
- (36) ReVelle, C. & Hogan, K. (1989). The maximum availability location problem. *Transportation Science*, **23**(3), 192-200.
- (37) Sacks, S.R. & Grief, S. (1994). Orlando Police Department uses OR/MS methodology new software to design patrol district. *OR/MS Today*, 30-32.
- (38) Saydam, C. & Aytug, H. (2003). Accurate estimation of expected coverage: revisited. *Socio-Economic Planning Sciences*, **37**, 69-80.
- (39) Swersey, A.J. (1994). The deployment of police, fire, and emergency medical units. **In: *Handbooks in OR & MS*** [edited by S.M. Pollock *et al.*], **6**, 151-200.
- (40) Takeda, R.A. (2000). Uma contribuição para avaliar o desempenho de sistemas de transporte emergencial de saúde. São Carlos. 176p. Tese (Doutorado) – Escola de Engenharia de São Carlos, Universidade de São Paulo.
- (41) Takeda, R.A.; Widmer, J.A. & Morabito, R. (2001). Uma proposta alternativa para avaliação do desempenho de sistemas de transporte emergencial de saúde brasileiros. *Transportes*, **9**(2), 9-27.
- (42) Taylor, P.E. & Huxley, S.J. (1989). A break from tradition for San Francisco police: patrol officer scheduling using an optimization-based decision support system. *Interfaces*, **19**(1), 4-24.
- (43) Toregas, C.; Swain, R.; Revalle, C. & Bergman, L. (1971). The location of emergency service facilities. *Operations Research*, **19**, 1363-1373.
- (44) Volz, R.A. (1971). Optimum ambulance location in semi-rural areas. *Transportation Science*, **5**, 193-203.
- (45) World Health Organization (2000). *The world health report 2000: health systems – improving performance*. Geneva, Switzerland. World Health Organization, 1211.
- (46) Wolff, R.W. (1989). *Stochastic modeling and theory of queues*. Prentice-Hall, New Jersey.