

# ADAPTAÇÕES DO MODELO DE FILA HIPERCUBO PARA ANÁLISE DE SISTEMAS DE ATENDIMENTO MÉDICO EMERGENCIAIS EM RODOVIAS

**Ana Paula Iannoni**

Departamento de Engenharia de Produção  
Universidade Federal de São Carlos

**Reinaldo Morabito**

Departamento de Engenharia de Produção  
Universidade Federal de São Carlos

## RESUMO

O modelo hipercubo é um modelo descritivo para planejar sistemas emergenciais servidor-para-cliente, baseado em filas espacialmente distribuídas e que permite a avaliação de uma grande variedade de medidas de desempenho do sistema. Neste estudo adaptamos este modelo para tratar sistemas de atendimento médico emergenciais (SAEs) em rodovias com políticas particulares de despacho, em que somente algumas ambulâncias do sistema podem atender chamadas em uma dada região (*backup* parcial), e duas ou mais ambulâncias idênticas ou diferenciadas podem atender uma única chamada (múltiplo despacho). Também consideramos situações em que uma ambulância pode se encontrar ocupada em sua própria base atendendo uma chamada não emergencial, isto é, sem ter que se deslocar na rodovia. Resultados computacionais são analisados aplicando-se os modelos para o estudo de caso de um SAE em rodovias no estado de São Paulo.

## ABSTRACT

The hypercube is a well-known descriptive model for planning server-to-customer systems, which is based on spatially distributed queuing theory and allows the evaluation of a wide variety of system performance measures. In the present study we adapt this model to analyze Emergency Medical Systems (EMS) on highways operating within particular dispatching policies, which considers that only certain ambulances in the system can service calls in a given region (partial *backup*), and two ambulances can simultaneously service a single call (multiple dispatch). We also consider the situations that an ambulance is busy in its base servicing a non-emergency call (i.e. it does not need to travel along the highway). Computational results are analyzed applying the models to the case study of an EMS operating on highways in São Paulo state.

## 1. INTRODUÇÃO

Ao projetar ou modificar a configuração de SAEs, é fundamental a consideração do *trade-off* entre a qualidade de atendimento e os custos de investimento e operação nestes sistemas. Além disso, como os SAEs em geral são serviços do tipo servidor-para-cliente (*server-to-costumer*), em que os servidores precisam se deslocar até o local da solicitação do usuário e há incertezas com relação à localização dos servidores e localização e instante dos chamados, a análise de desempenho destes sistemas deve levar em conta fatores probabilísticos na distribuição espacial e temporal dos chamados e servidores. Importantes contribuições podem ser encontradas na literatura em estudos sobre modelos descritivos e prescritivos (otimização) que incorporam estes aspectos probabilísticos dos SAEs. Nos trabalhos de Swersey (1994), Owen & Daskin (1998), Chiyoshi et al. (2000) e Brotcorne et al. (2003) são revistos os principais modelos de localização para analisar os sistemas de atendimento emergencial, desenvolvidos nas últimas décadas.

O modelo hipercubo (Larson, 1974; Larson & Odoni, 1981), baseado em *teoria de filas espacialmente distribuídas*, tem se mostrado como um dos modelos descritivos mais efetivos para analisar estes sistemas. A idéia básica é expandir o espaço de estado dos servidores de um simples sistema Markoviano de filas com múltiplos servidores (p.e,  $M/M/N$ , onde  $N$  é o número de servidores), de forma a representar cada servidor individualmente e incorporar as complexidades envolvidas na política de despacho. O modelo implica na solução de sistemas

lineares de  $O(2^N)$  equações, cujas variáveis envolvidas correspondem às probabilidades de estado do sistema em equilíbrio. Por meio destas probabilidades podem ser estimadas importantes medidas de desempenho para análise e gerenciamento do sistema, tais como cargas de trabalho dos servidores, tempo médio de resposta aos usuários e frações de despacho de cada servidor para cada região.

Alguns estudos de extensão do modelo hipercubo podem ser encontrados em Larson (1975), Halpern (1977), Chelst & Barlach (1981), Larson & Mcknew (1982), Jarvis (1985), Mendonça & Morabito (2001) e Iannoni (2005). Em particular, Chelst & Barlach (1981) modificaram o modelo hipercubo para considerar duplo despacho de viaturas idênticas em sistemas de patrulhamento policial, e Mendonça & Morabito (2001) modificaram o modelo para considerar simples despacho com *backup* parcial em sistemas de atendimento médico, em que somente algumas ambulâncias podem atender chamadas em uma dada região. Outros estudos propõem a combinação do modelo hipercubo com modelos de otimização, tais como Batta et al. (1989), Saydam & Aytug (2003), Chiyoshi et al. (2003) e Galvão et al. (2005). No Brasil, alguns exemplos de aplicação do modelo hipercubo em sistemas de atendimento médico são: a análise de alguns SAEs urbanos (Takeda et al, 2004; Costa, 2004, Takeda et al., 2005) e alguns SAEs em rodovias do estado de São Paulo e Rio de Janeiro (Mendonça & Morabito, 2001; Iannoni 2005).

No presente artigo, propomos adaptar o modelo hipercubo para analisar os SAEs em rodovias considerando políticas particulares de despacho. Em um primeiro modelo, estendemos o modelo hipercubo múltiplo despacho de Chelst & Barlach (1981) considerando políticas com *backup* parcial e múltiplo despacho de ambulâncias idênticas para uma mesma chamada. No segundo modelo, estendemos o primeiro modelo incorporando um terceiro estado para cada servidor, enquanto realiza atendimentos na sua própria base, que diferem dos atendimentos a acidentes realizados ao longo da rodovia porque não envolvem deslocamento do servidor. Os modelos apresentados no presente estudo são aplicados para analisar duas configurações diferentes do SAE de uma concessionária de rodovias no estado de São Paulo. Este artigo está estruturado da seguinte forma: a seção 2 apresenta uma breve descrição do SAE analisado; a seção 3 descreve as duas adaptações do modelo hipercubo propostas para analisar o SAE; a seção 4 analisa os resultados das principais medidas de desempenho obtidas pelos modelos. Finalmente, a seção 5 apresenta as conclusões deste estudo.

## **2. SISTEMAS DE ATENDIMENTO EMERGENCIAIS EM RODOVIAS**

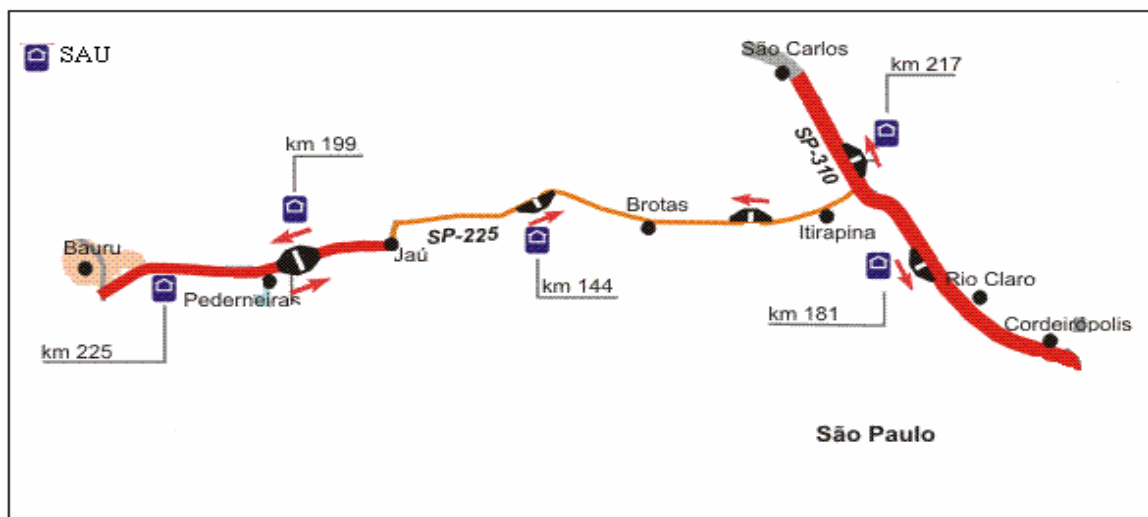
Os SAEs em rodovias são em geral caracterizados por não admitirem filas de espera, pois quando os servidores candidatos estão ocupados, a chamada deve ser transferida a outro sistema. Em geral, a operação dos SAEs em rodovias conta com uma central de informações, localizada em um ponto da rodovia ou cidade estratégica, que é responsável por receber as chamadas por atendimento, despachar as ambulâncias de acordo com a política de despacho do sistema e monitorar os movimentos das mesmas. Alguns SAEs em rodovias brasileiras são gerenciados por organizações privadas como parte do contrato de concessão com o governo do estado. De acordo com este contrato, estas organizações devem realizar todas as obras necessárias de planejamento, ampliação e melhoria das rodovias, assim como prover todo tipo de assistência ao usuário.

O SAE analisado neste estudo é gerenciado por uma concessionária de rodovias no interior do estado de São Paulo. Há  $N = 5$  bases ambulâncias ao longo da rodovia. Neste sistema, a central de informações está localizada em uma das bases na rodovia. A política de despacho

consiste em enviar a ambulância da base mais próxima do local da chamada, se a chamada requer apenas uma ambulância (único despacho). Se a ambulância estiver ocupada, a segunda ambulância mais próxima (chamada *backup*) é despachada. Quando a chamada requer duplo despacho, as duas ambulâncias mais próximas são acionadas. Se alguma delas estiver ocupada, a ambulância disponível é despachada. Caso as duas ambulâncias mais próximas estiverem ocupadas, a chamada (único ou duplo despacho) é considerada perdida para o sistema.

Além disso, considerável parte das chamadas atendidas por cada ambulância ocorre na sua própria base na rodovia. Por exemplo, um usuário da rodovia ou trabalhador de uma instalação próxima de uma base do SAE na rodovia pode solicitar atendimento nesta base. Tais eventos tornam os servidores ocupados mas são caracterizados por tempo de viagem igual a zero, conseqüentemente o tempo médio de atendimento aos usuários (que genericamente compreende um tempo de preparação, o tempo de viagem até o local do acidente, o tempo de atendimento em cena e o tempo de retorno à base) deve ser diferenciado para estas chamadas. Outra característica importante é que quando uma chamada deste tipo chega em uma base cuja ambulância está ocupada realizando um atendimento fora da base, esta é perdida para o sistema, mesmo que ambulâncias em outras bases estejam disponíveis (não há *backup*).

Note que, com esta política particular de despacho, a chamada pode ser atendida por somente duas ambulâncias (preferencial e *backup*) ou por apenas a ambulância preferencial (no caso das chamadas atendidas na base). Portanto, este sistema pode ser definido como um sistema com múltiplos servidores sem filas de espera de chamados, que não satisfaz algumas das principais hipóteses do modelo hipercubo original de Larson (1974). Estas hipóteses admitem que qualquer servidor pode viajar a qualquer átomo, um único servidor pode ser despachado para cada chamada e qualquer chamada pode ser atendida por um servidor *backup*, caso o servidor preferencial esteja ocupado. A Figura 1 ilustra a área de atuação do SAE analisado neste trabalho, envolvendo trechos de rodovias do interior do estado de São Paulo.



**Figura 1:** SAE em trechos de rodovias do interior do estado de São Paulo.

### 3. O MODELO HIPERCUBO PARA SAES EM RODOVIAS

O nome hipercubo é derivado do espaço de estado do sistema, sendo que cada estado do sistema corresponde a um vértice de um hipercubo. Um estado em particular do sistema é representado pela lista de servidores que estão livres e ocupados. Se há dois estados possíveis para cada servidor: livre (0) ou ocupado (1) em certo instante de tempo, temos então  $2^N$  estados para o sistema. As principais hipóteses para aplicação do modelo hipercubo múltiplo despacho para análise dos SAEs em rodovias são:

- A rodovia é particionada em  $N_A$  átomos geográficos (regiões), os quais correspondem a fontes independentes de chamadas. Em cada átomo  $j$  admite-se que as chamadas são geradas de acordo com o processo de Poisson. As chamadas podem ser diferenciadas de acordo, por exemplo: chamadas tipo 1 (com taxa de chegada  $\lambda_j^{[1]}$ ) são chamadas de emergência que requerem despacho de uma única ambulância, e chamadas tipo 2 (com taxa de chegada  $\lambda_j^{[2]}$ ) requerem o despacho simultâneo de duas ambulâncias idênticas.
- Há  $N$  ambulâncias espacialmente distribuídas ao longo da rodovia, que permanecem fixas em suas bases quando disponíveis. Como mencionado anteriormente, de acordo com a política particular dos SAEs em rodovias, cada ambulância pode somente viajar para átomos de suas áreas preferencial e *backup*.
- O despacho dos servidores é realizado de acordo com uma lista de preferência para cada átomo. A ambulância mais próxima está no topo desta lista (primeira a ser despachada). Ocorre o *backup* parcial de ambulâncias, dado que devido às restrições de distância, somente algumas ambulâncias podem atender como *backup*, e a chamada pode não ser atendida mesmo que haja outras ambulâncias disponíveis.
- O tempo médio de atendimento para cada ambulância inclui um tempo de preparação, os tempos de viagem (ida e volta) e o tempo de atendimento em cena. Considera-se que cada ambulância  $i$  do sistema possui um tempo médio de serviço distinto ( $1/\mu_i$ ). O modelo também admite que o tempo de atendimento é representado por uma distribuição exponencial negativa. Porém, dado que os SAEs em rodovias não admitem filas, esta suposição não é tão importante para a acuracidade do modelo, uma vez que modelos M/M/N/N e M/G/N/N têm a mesma distribuição de equilíbrio (Larson & Odoni, 1981; Chiyoshi et al. 2000).
- O tempo de viagem entre cada par de átomos é conhecido ou pode ser estimado utilizando os conceitos de probabilidade geométrica. As variações no tempo de atendimento devido a variações no tempo de viagem são consideradas de segunda ordem quando comparadas a variações no tempo em cena ou no tempo de preparação.

### **3.1 Modelo hipercubo múltiplo despacho com servidores idênticos e backup parcial (Modelo 1)**

Este modelo admite que há dois tipos de chamadas no sistema: chamadas tipo 1 que requerem o despacho de apenas uma ambulância (com taxa de chegada  $\lambda_j^{[1]}$ ) e tipo 2 que requerem o despacho simultâneo de duas ambulâncias (com taxa de chegada  $\lambda_j^{[2]}$ ). Portanto, esta versão do modelo hipercubo trata os SAEs com uma política de despacho particular que envolve: duplo despacho de servidores idênticos para atender um mesmo acidente; *backup* parcial (somente alguns servidores do sistema podem atender um determinado átomo) e não há fila de espera. De acordo com a lista de preferência de despacho em cada átomo, ao ocorrer uma chamada do tipo 1, a primeira ambulância da lista é despachada, e se esta estiver ocupada, a

ambulância *backup* é despachada. No caso de chamadas tipo 2, as duas primeiras ambulâncias da lista são despachadas, e se apenas uma delas estiver disponível, esta deve atender ao chamado sozinha (possivelmente, com a ajuda de outro SAE). Se ambas as ambulâncias estiverem ocupadas, tanto chamadas do tipo 1 quanto do tipo 2 são perdidas, dado que uma terceira ambulância nunca é despachada (*backup* parcial).

Com relação ao processo de atendimento, temos que : chamadas tipo 1 são atendidas por uma única ambulância  $i$  com taxa média de atendimento  $\mu_i$ , e chamadas tipo 2 são atendidas por duas ambulâncias  $i$  e  $k$ , que operam de forma independente entre si com taxas médias de atendimento  $\mu_i$  e  $\mu_k$ , respectivamente. Note que, com relação ao tempo de atendimento, uma chamada tipo 2 é tratada como duas chamadas tipo 1 distintas, sendo atendidas cada uma por uma ambulância.

### 3.2 Modelo hipercubo múltiplo despacho com terceiro estado para cada servidor (Modelo 2)

O modelo 2 modifica o modelo 1 considerando que nos SAEs em rodovias, parte das chamadas são atendidas pelas ambulâncias na sua base e possuem características diferenciadas das que decorrem de acidentes ao longo da rodovia, conforme discussão na seção 2. Tais eventos tornam os servidores ocupados na própria base e são caracterizados por tempo de viagem igual a zero, conseqüentemente o tempo médio de atendimento deve ser diferenciado para estas chamadas. Além disso, se o servidor estiver ocupado quando uma chamada chega na sua base, não há atendimento *backup* e a chamada é perdida para o sistema.

Ao diferenciar as chamadas atendidas na base com relação à política de despacho (não há *backup*) e ao tempo de atendimento (não há tempo de viagem), devemos considerar um terceiro estado para cada ambulância, que corresponde ao estado ocupado atendendo uma chamada na sua base. Extensão similar do modelo hipercubo único despacho foi estudada por Larson & Mcknew (1982) para analisar os sistemas de patrulhamento policial considerando que, uma parte do tempo que permanecem ocupadas, as viaturas de policia estão atendendo eventos que não foram atribuídos por ordem de despacho da central. Neste estudo denominamos as chamadas atendidas na base como tipo 1a, com taxa de chegada  $\lambda_j^{[1a]}$  em cada átomo  $j$ . Como descrito anteriormente, o atendimento deve ser diferenciado para chamadas tipo 1a, e cada ambulância  $i$  possui taxa de atendimento  $\mu_i^{[I]}$  para chamadas tipo 1 (único despacho) e tipo 2 (duplo despacho) e  $\mu_i^{[III]}$  para chamadas tipo 1a em sua base. Uma ambulância pode estar nos estados: (0) livre; (1) ocupada atendendo uma chamada tipo 1 ou tipo 2 ao longo da rodovia; (2) ocupada atendendo uma chamada tipo 1a na sua base, e assim há  $3^N$  estados possíveis para o sistema.

Entre as mais interessantes medidas de desempenho descritas pelo modelo hipercubo com *backup* parcial e múltiplo despacho de ambulâncias, estão: as cargas de trabalho de cada ambulância diferenciando atendimentos de emergência (tipos 1 e 2) e chamadas tipo 1a atendidas na base, o tempo médio de resposta aos usuários (considerando os dois tipos de chamadas), o tempo médio de viagem para chamadas com único e múltiplo despacho (i.e, tipos 1 e 2), o tempo médio de viagem para a primeira e segunda ambulância que chega no local de uma chamada tipo 2, o tempo médio que a primeira ambulância espera até a chegada da segunda ambulância no local de uma chamada tipo 2, a fração de despacho de cada

ambulância para cada região de acordo com o tipo de chamada, a probabilidade de perda para qualquer chamada do sistema ou para cada tipo de chamada.

### 3.3 Algumas medidas de desempenho obtidas pelos modelos hipercubo 1 e 2

#### 3.3.1 Probabilidade de perda:

Esta é uma medida importante para os SAEs em rodovias, dado que uma chamada pode ser perdida para o sistema, mesmo se há ambulâncias disponíveis (*backup* parcial). Podemos calcular a probabilidade de perda para qualquer chamada do sistema ( $P_p$ ) e para cada tipo de chamada. Mais detalhes sobre o cálculo das probabilidades de perda podem ser encontrados em Iannoni (2005).

#### 3.3.2 Fração de despachos :

As diferentes adaptações do modelo hipercubo resultam em medidas adicionais de fração de despacho. Seja  $P_B$  a probabilidade de equilíbrio de um estado  $B$  do sistema. Algumas destas estatísticas são definidas da seguinte forma:

- Fração de despachos tipo 1 em que a ambulância  $i$  é enviada ao átomo  $j$ :

$$f_{ij}^{[1]} = \frac{(\lambda_j^{[1]} / \lambda^{[1]}) \sum_{B \in E_{ij}} P_B}{(1 - P_p^{[1]})} \quad (1)$$

onde  $E_{ij}$  corresponde ao conjunto de estados em que a ambulância  $i$  é a primeira ambulância disponível na lista de preferência do átomo  $j$  e  $\lambda^{[1]} = \sum_{j=1}^{N_A} \lambda_j^{[1]}$ .

- Fração de despachos tipo 2 em que a ambulância  $i$  e  $k$  são enviados ao átomo  $j$ :

$$f_{(i,k)j}^{[2]} = \frac{(\lambda_j^{[2]} / \lambda^{[2]}) \sum_{B \in E_{(i,k)j}} P_B}{(1 - P_p^{[2]})}, \quad (2)$$

onde  $E_{(i,k)j}$  corresponde ao conjunto de estados em que a ambulância  $i$  e  $k$  são as duas primeiras ambulâncias disponíveis na lista de preferência do átomo  $j$  e  $\lambda^{[2]} = \sum_{j=1}^{N_A} \lambda_j^{[2]}$ ;

- Fração de despachos tipo 2 em que a ambulância  $i$  é a única ambulância enviada para o átomo  $j$ , quando ambulância  $k$  está ocupada:

$$f_{ij}^{[2]} = \frac{(\lambda_i^{[2]} / \lambda^{[2]}) \sum_{B \in F_{ij}} P_B}{(1 - P_p^{[2]})}, \quad (3)$$

onde  $F_{ij}$  corresponde ao conjunto de estados em que somente a ambulância  $i$  pode responder chamadas no átomo  $j$ .

Outras estatísticas de fração de despachos podem ser calculadas considerando o total de despachos no sistema, como: a fração de todos os despachos que envia a ambulância  $i$  para o átomo  $j$  para responder chamadas tipos  $m = 1, 2$ , 1a ( $f_{ij}^{[m]}$ ,  $f_{(i,k)j}^{[m]}$ ). Por exemplo, a frequência de duplo despacho considerando todos os despachos do sistema para chamadas tipo 2, é definida da forma:

$$f_{(i,k)j}^{[2]} = \frac{(\lambda_i^{[2]} / \lambda^{[2]}) \sum_{B \in E(i,k)j} P_B}{(1 - P_p)} \quad (4)$$

### 3.3.3 Medidas agregadas de tempo de viagem:

Por meio das medidas de frequência de despacho, podemos obter algumas medidas interessantes de tempo médio de viagem (mais detalhes sobre o cálculo destas e outras medidas de tempo de viagem são encontrados em Iannoni, 2005). Apresentamos a seguir algumas das principais medidas de tempo de viagem, sendo que no modelo 2, as medidas de tempo de viagem são as mesmas do modelo 1, dado que chamadas tipo 1a não requerem tempo de viagem.

- Tempo médio de viagem para chamadas tipo 1:

$$T^{[1]} = \sum_{j=1}^{N_A} \sum_{i=1}^N f_{ij}^{[1]} t_{ij}, \quad (5)$$

onde  $t_{ij}$  é o tempo médio de viagem da ambulância  $i$  à base do átomo  $j$  obtido pela análise da amostra de dados do sistema.

- Tempo médio de viagem para chamadas tipo 2 (considerando a primeira ambulância que chega no local da chamada):

$$\bar{T}^{[2]} = \sum_{j=1}^{N_A} \left[ \sum_{i=1}^{N-1} \sum_{k=i+1}^N f_{(i,k)j}^{[2]} \min(t_{ij}, t_{kj}) + \sum_{i=1}^N f_{ij}^{[2]} t_{ij} \right] \quad (6)$$

- Tempos médio de viagem para a primeira ambulância que chega no local de uma chamada tipo 2, considerando apenas os despachos em que as duas ambulâncias  $i$  e  $k$  são despachadas ;

$$\bar{T}_F = \frac{\sum_{j=1}^{N_A} \sum_{i=1}^{N-1} \sum_{k=i+1}^N f_{(i,k)j}^{[2]} \min(t_{ij}, t_{kj})}{\sum_{j=1}^{N_A} \sum_{i=1}^{N-1} \sum_{k=i+1}^N f_{(i,k)j}^{[2]}} \quad (7)$$

- Tempo médio de viagem no sistema (considerando chamadas tipo 1 e tipo 2):

$$\bar{T} = \left( \sum_{j=1}^{N_A} \sum_{i=1}^N f_{ij}^{[1]} + f_{ji}^{[2]} \right) t_{ji} + \left( \sum_{j=1}^{N_A} \sum_{i=1}^{N-1} \sum_{k=i+1}^N f_{(i,k)j}^{[2]} \right) \min(t_{ij}, t_{kj}), \quad (8)$$

onde  $f_{ij}^{[1]}, f_{ji}^{[2]}, f_{(i,k)j}^{[2]}$  correspondem a frequência de despachos para chamadas do tipo 1 e 2, considerando todos os despachos do sistema.

### 3.3.4 Fração de chamadas atendidas em mais que $\bar{t}_v$ minutos:

Podemos também calcular a probabilidade de uma chamada no sistema ser atendida em um tempo superior a  $\bar{t}_v$  (limite predeterminado), utilizando a distribuição do tempo de viagem da ambulância de sua base a cada átomo (local da chamada):

$$P_{t > \bar{t}_v} = \sum_{i=1}^N \sum_{j=1}^{N_A} f_{ij}^{[v]}, \quad (9)$$

onde  $f_{ij}^{[v]}$  corresponde a fração de todas as chamadas para as quais a ambulância  $i$  é enviada ao átomo  $j$ , e cujo tempo de viagem excede  $\bar{t}_v$ .

#### 4. RESULTADOS COMPUTACIONAIS

Para aplicação dos modelos apresentados na seção 3, o trecho de rodovias estudado foi dividido em 8 átomos (segmentos), de acordo com a área primária de cada base estabelecida pelos operadores e gerentes do sistema. A tabela 2 apresenta os dados do processo de chegada e atendimento e a lista de preferência de despacho para cada átomo da configuração 1 do SAE analisado. Os dados desta configuração foram coletados no período de dezembro de 2001 a junho de 2002, com um total de 945 eventos. Note que,  $\lambda = \sum_{j=1}^8 (\lambda_j^{[1]} + \lambda_j^{[2]}) = 0,1878$  chamadas/h e  $\mu = \sum_{i=1}^5 \mu_i = 5,4321$  chamadas/h.

**Tabela 1-** Dados dos processos de chegada e atendimento

Átomo $j$	Chegada tipo 1 $\lambda_j^{[1]}$ (ch/h)	Chegada tipo 2 $\lambda_j^{[2]}$ (ch/h)	Servidor Preferen- cial	Servidor <i>backup</i>	Servidor $i$	Taxa de serviço $\mu_i$ (ch/h)	Carga de trabalho $\rho_i$
1	0,06181	0,00340	1	2	1	1,3085	0,0587
2	0,04819	0,00476	2	1	2	1,0642	0,0532
3	0,00693	0,00040	1	2	3	0,8247	0,0182
4	0,00996	0,00042	3	1	4	1,0852	0,0252
5	0,00494	-	3	2	5	1,1495	0,0185
6	0,01139	-	4	3			
7	0,01440	-	4	5			
8	0,01937	0,00196	5	4			

Realizamos testes de aderência (*Kolmogorov-Smirnov* e *Qui-quadrado*) para verificar a hipótese de que o processo de chegada é Poisson. Os resultados obtidos desta análise, com nível de significância  $\alpha = 0,05$ , mostraram que, para todos os átomos, não se pode rejeitar esta hipótese. Por outro lado, estes testes de aderência rejeitaram a hipótese de que a distribuição dos tempos de atendimento seja exponencial, similarmente ao sistema analisado por Mendonça & Morabito (2000; 2001). Porém, como comentado em Jarvis (1985) e Chiyoshi et al (2000), este tipo de sistema pode ser analisado pelo modelo hipercubo. Aplicamos também a análise de variância para verificar a variabilidade entre os tempos médio de atendimento dos servidores.

As equações de equilíbrio são definidas considerando que no estado de equilíbrio, o fluxo para dentro de um estado do sistema deve ser igual ao fluxo para fora deste estado. Considerando o modelo 1 aplicado ao SAE analisado com  $N = 5$  ambulâncias, há  $2^5 = 32$  estados possíveis para o sistema. De acordo, com a lista de preferência de despacho na Tabela 1, as equações de equilíbrio podem ser formuladas, por exemplo, como a expressão abaixo para o estado 11001:

$$\begin{aligned}
 P_{11001}((\lambda_4^{[1]} + \lambda_5^{[1]} + \lambda_6^{[1]} + \lambda_7^{[1]} + \lambda_8^{[1]} + \lambda_4^{[2]} + \lambda_5^{[2]} + \lambda_6^{[2]} + \lambda_7^{[2]} + \lambda_8^{[2]}) + \mu_1 + \mu_2 + \mu_5) = \\
 P_{00001}(\lambda_1^{[2]} + \lambda_2^{[2]} + \lambda_3^{[2]}) + P_{10001}(\lambda_1^{[1]} + \lambda_2^{[1]} + \lambda_3^{[1]} + \lambda_1^{[2]} + \lambda_2^{[2]} + \lambda_3^{[2]}) + \\
 P_{01001}(\lambda_1^{[1]} + \lambda_2^{[1]} + \lambda_3^{[1]} + \lambda_1^{[2]} + \lambda_2^{[2]} + \lambda_3^{[2]}) + P_{11000}(\lambda_8^{[1]}) + P_{11101}(\mu_3) + P_{11011}(\mu_4)
 \end{aligned} \quad (10)$$

Os resultados obtidos para as probabilidades de estado de equilíbrio mostraram que na maior

parte do tempo, o sistema permanece ocioso (todos servidores livres), pois  $P_{00000} = 0,8434$  e  $P_{11111} = 0,00000034$ . Calculando a carga de trabalho  $\rho_i$  para cada ambulância  $i$ , os resultados obtidos são:  $\rho_1 = 0,0578$ ,  $\rho_2 = 0,0537$ ,  $\rho_3 = 0,0186$ ,  $\rho_4 = 0,0253$  e  $\rho_5 = 0,0185$ . Observe que  $\rho_1$  (máximo valor) é mais que 200% maior que  $\rho_5$  (mínimo valor). Os desvios dos resultados do modelo para a amostra (tabela 2) são pequenos (desvio médio de apenas 1,0%). A probabilidade de perda para uma chamada de qualquer tipo no sistema é  $P_p = 0,6\%$ . Calculando as medidas de frequências descritas na seção 3, os resultados mostraram que os despachos da ambulância 1 ao átomo 1 e da ambulância 2 ao átomo 2 correspondem aos eventos mais frequentes:  $f_{11}^{[1]} = 33,10\%$ ,  $(f_{(1,2)1}^{[2]} + f_{11}^{[2]}) = 29,48\%$ ,  $f_{11} = 32,89\%$  e  $f_{22}^{[1]} = 25,92\%$ ,  $(f_{(2,1)2}^{[2]} + f_{22}^{[2]}) = 41,46\%$ ,  $f_{22} = 24,73\%$ . Utilizando a expressão da seção 3.4, calculamos a fração de chamadas atendidas em tempo superior a  $\bar{t}_v = 10$  minutos:  $P_{tv > 10}$ . Os resultados obtidos pelo modelo mostraram que para 26,2% de todas as chamadas, 25,7% das chamadas tipo 1 e 33,5% das chamadas tipo 2, o tempo de resposta é maior que 10 minutos.

A tabela 3 apresenta algumas medidas agregadas de tempos de viagem calculadas pelo modelo hipercubo, e compara estes resultados com os resultados obtidos pela análise da amostra de dados e simulação discreta do sistema. Como por meio da análise da amostra não foi possível identificar e considerar os despachos que enviam uma única ambulância para atender uma chamada tipo 2, calculamos  $\bar{T}^{[2]}$  (dados pela expressão abaixo), ao invés de  $\bar{T}^{[2]}$  (conforme definidos na seção 3), respectivamente. Desta forma, contabilizamos somente os despachos que enviam duas ambulâncias para atender uma chamada tipo 2, para comparar os resultados com a amostra de dados.

$$\bar{T}^{[2]} = \frac{\sum_{j=1}^{N_A} \sum_{i=1}^{N-1} \sum_{k=i+1}^N f_{(i,k)j}^{[2]} \min(t_{ij}, t_{kj})}{\sum_{j=1}^{N_A} \sum_{i=1}^{N-1} \sum_{k=i+1}^N f_{(i,k)j}^{[2]}} \quad (11)$$

**Tabela 2** – Medidas agregadas de tempo médio de viagem (em minutos) – modelo 1

Medida	Hipercubo	Amostra	Desvio	Simulação média	Simulação (Intev. Conf.)
$\bar{T}^{[1]}$	6,277	5,962	5,3%	6,279	6,235 – 6,322
$\bar{T}^{[2]}$	8,186			8,133	7,858 – 8,408
$\bar{T}^{[2]}$	7,776	7,201	7,9%	7,821	7,567 – 8,074
$\bar{T}_F$	7,776	7,771	0,06%	7,821	7,567 – 8,074
$\bar{T}_S$	17,373	17,944	3,3%	17,426	17,245 – 17,607

Verificamos na Tabela 2 que os resultados do modelo apresentam desvios relativamente pequenos com relação à amostra. Além disso, o modelo também é validado pelo modelo de simulação discreta, considerando que os resultados obtidos pelo modelo hipercubo estão dentro do intervalo de confiança (nível de confiança  $\alpha = 0,05$ ) dos resultados da simulação (sexta coluna da Tabela 2).

Para aplicação do modelo 2, analisamos separadamente as chamadas tipo 1a (com tempo de viagem diferente de zero) e as chamadas de emergência dos tipos 1 e 2 ao longo da rodovia. A

tabela 3 mostra as taxas de chegada para os três tipos de chamada (1, 2 e 1a) e duas taxas de atendimento:  $\mu_i^{[I]}$  (tipo 1 e 2) e  $\mu_i^{[III]}$  (tipo 1a) para cada ambulância  $i$  do sistema.

**Tabela 3** - Dados dos processos de chegada e atendimento com chamadas tipo 1a

Átomo $j$	Chegada tipo 1 $\lambda_j^{[I]}$ (ch/h)	Chegada tipo 2 $\lambda_j^{[2]}$ (ch/h)	chegada tipo 1a $\lambda_j^{[1a]}$ (ch/h)	Servidor $i$	Taxa de serviço $\mu_i^{[I]}$ (ch/h)	Taxa de serviço $\mu_i^{[III]}$ (ch/h)
1	0,04520	0,00340	0,01660	1	1,1340	2,9911
2	0,04105	0,00476	0,00714	2	1,0387	1,2931
3	0,00693	0,00040	-	3	0,7960	1,0026
4	0,00911	0,00042	0,00085	4	1,0782	1,1219
5	0,00300	-	0,00194	5	1,1213	1,5267
6	0,00932	-	0,00207			
7	0,01148	-	0,00292			
8	0,01663	0,00196	0,00274			

De acordo com discussão do modelo 2 na seção 3.2, temos  $3^5 = 243$  possíveis estados para o sistema. Calculando as probabilidades de estado a partir das equações de estado de equilíbrio, obtemos as principais medidas de desempenho do sistema, incluindo medidas adicionais com relação ao modelo anterior. Por exemplo, calculamos separadamente as cargas de trabalho para chamadas tipo 1 e 2 ( $\rho_i^{[I]}$ ) e tipo 1a ( $\rho_i^{[III]}$ ) para cada ambulância  $i$ , e como esperado os resultados de  $\rho_i$  (modelo 1) são, aproximadamente, a soma dos resultados de  $\rho_i^{[I]}$  e  $\rho_i^{[III]}$  (modelo 2). Os resultados de tempos de viagem para chamadas tipo 2 sofrem poucas alterações com relação aos resultados da análise da aplicação do modelo 1 (apenas as relacionadas às modificações nas probabilidades de equilíbrio). Apresentamos a seguir os tempos de viagem relativos ao único despacho de ambulâncias. Por exemplo, o tempo médio para chamadas tipo 1 no sistema definido na seção 3.4 é:  $\bar{T}^{[I]} = 7,376$  minutos, e o resultado para esta medida obtido pela análise da amostra é 7,196 minutos (desvio de -2,5%). Lembre-se que no modelo 1,  $\bar{T}^{[I]} = 6,277$  (14,89% menor do que o modelo 2). A Tabela 4 apresenta os resultados do tempo médio de viagem de cada ambulância para chamadas tipo 1 ( $\overline{TU}_i^{[I]}$ ), e comparando os com os resultados da análise da amostra (desvio relativo).

**Tabela 4** – Tempo de médio viagem servidor – único despacho  $\overline{TU}_i^{[I]}$  (em minutos)

Servidor $i$	$\overline{TU}_i^{[I]}$ modelo 2	$\overline{TU}_i^{[I]}$ amostra	Desvio	$\overline{TU}_i$ simulação média	Simulação (Intev. Conf.)	$\overline{TU}_i^{[I]}$ modelo 1	Desvio
1	7,481	7,503	0,29%	7,622	7,405 - 7,839	5,993	24,83%
2	8,157	7,871	-3,60%	8,215	8,123 - 8,300	7,342	11,10%
3	7,822	8,068	3,04%	7,689	7,499 - 7,876	6,686	17,00%
4	8,144	8,042	1,27%	8,206	8,079 - 8,332	6,705	21,85%
5	3,724	3,701	-0,62%	3,706	3,627 - 3,785	3,682	1,14%
$\bar{T}^{[I]}$	7,376	7,196	-2,50%			6,277	17,51%

Note nas colunas 7 e 8 da Tabela 4 que o tempo médio de viagem de cada servidor aumenta significativamente com relação aos resultados desta medida obtidos pelo modelo 1. Por

exemplo, para ambulância 1, o desvio do modelo 1 para o modelo 2 é de 24,83%. Isto se deve ao fato de que, diferentemente do modelo 1, a matriz do tempo de viagem calculada no modelo 2 e na análise da amostra com chamadas tipo 1a não contabiliza os tempos de viagem iguais à zero ( $t_{ij} = 0$ ), que correspondem aos atendimentos realizados na base, diferenciando chamadas tipo 1 e tipo 1a com tempos de atendimento  $\mu_i^{[I]}$  e  $\mu_i^{[II]}$ , respectivamente. Desta forma, apesar do modelo atual ser mais caro em termos de esforço computacional, pois o espaço de estados cresce de  $O(2^N)$  para  $O(3^N)$ , o mesmo representa melhor o sistema real analisado.

## 5. CONCLUSÕES

Este estudo mostra como o modelo hipercubo de filas espacialmente distribuídas pode ser modificado e aplicado para análise dos sistemas de atendimento médico emergencial em rodovias, considerando a aleatoriedade envolvida na operação destes sistemas e suas particularidades com relação as políticas de despacho de ambulâncias. Entre as principais particularidades da operação deste sistema, estão: política de múltiplo despacho de ambulâncias e *backup* parcial, servidores ocupados atendendo chamadas que não são de emergência (chamadas atendidas na base do servidor). Os modelos foram aplicados em um estudo de caso de um SAE de uma concessionária no estado de São Paulo. Os resultados obtidos pelos modelos foram validados por meio da análise de amostras e resultados da simulação discreta do sistema. Tais resultados mostraram que os modelos são efetivos para avaliar as principais medidas de desempenho do sistema.

As adaptações e extensões do modelo hipercubo propostas para análise dos SAEs em rodovias podem ser diretamente aplicadas em abordagens que otimizam a localização dos servidores, por exemplo, nos métodos propostos em Batta et al. (1989), Saydam & Aytug (2003), Chiyoshi et al. (2003) e Galvão et al. (2005), ou em abordagens que otimizam o dimensionamento das áreas preferenciais das ambulâncias do sistema como proposto em Iannoni (2005).

### Agradecimentos

Os autores agradecem à concessionária *Centrovias* pela valiosa colaboração durante a realização do estudo de caso. Também agradecemos o apoio financeiro do CNPQ e da CAPES.

### REFERÊNCIAS BIBLIOGRÁFICAS

- Batta R., Dolan J.M., Krishnamurthy N.N. (1989) The maximal expected covering location problem: Revisited. *Transportation Science* v.23, p.277-287.
- Brotcorne L., Laporte G., Semet F. (2003). Ambulance location and relocation models. *European Journal of Operational research* v.147, p. 451-63.
- Burwell, T.H.; Jarvis, J.P.; Mcknew, M.A.(1993) Modeling co-located servers and dispatch ties in the hypercube model. *Computers and Operations Research* v. 20, n.2, p. 113-119.
- Chelst K.; Barlach Z. (1981) Multiple unit dispatches in emergency services: models to estimate system performance. *Management Science* v. 27, n. 12, p. 1390-1409.
- Chiyoshi F., Galvão R. D., Morabito R. (2000) O uso do modelo hipercubo na solução de problemas de localização probabilísticos. *Gestão & Produção* v. 7, n.2, p. 146-174.
- Chiyoshi F., Galvão R. D., Morabito R.. (2001) Modelo hipercubo: análise e resultados para o caso de servidores não-homogêneos. *Pesquisa Operacional* v.21, n. 2, p.199-218.
- Chiyoshi F., Galvão R. D., Morabito R. (2003) A note on solutions to the maximal expected covering location problem. *Computers and Operations Research* v.30, p. 87-96.
- Costa D. M. B.(2004) Uma metodologia iterativa para determinação de zonas de atendimento de serviços

emergenciais. Universidade Federal de Santa Catarina. *Tese* (doutorado em Engenharia de Produção) - Departamento de Engenharia de Produção.

- Galvão R.D., Chiyoshi F., Espejo L.G.A., Rivas M.P.A. (2003) Solução do problema de localização de máxima disponibilidade utilizando o modelo hipercubo. *Pesquisa Operacional* v. 23, p. 61-78.
- Galvão R.D., Chiyoshi F., Morabito R. (2005) Towards unified formulations and extensions of two classical probabilistic location models. *Computers and Operations Research* v. 32, p. 15-33.
- Halpern J. (1977) Accuracy of estimates for the performance criteria in certain emergency service queuing systems. *Transportation Science* v.11, n.3, p. 223-242.
- Iannoni, A. (2005) Otimização da configuração e operação de sistemas médico emergenciais em rodovias utilizando o modelo hipercubo. Universidade Federal de São Carlos. *Tese* (doutorado em Engenharia de Produção) - Departamento de Engenharia de Produção.
- Jarvis J.P. (1985) Approximating the equilibrium behavior of multi-server loss systems. *Management Science* v.31, p. 235 – 239.
- Larson R.C. (1974) Hypercube queuing model for facility location and redistricting in urban emergency services. *Computers and operations research* v.1, p. 67-95.
- Larson R.C. (1975) Approximating the performance of urban emergency service systems. *Operations Research* v.23, p. 845-868.
- Larson R.C., Odoni A.R. (1981) *Urban operations research*. Prentice Hall. New Jersey.
- Larson R., Mcknew M.A. (1982) Police patrol-initiated activities within a systems queuing model. *Management Science* v. 28, n.7, p. 759 – 774.
- Mendonça F.C., Morabito R. (2001) Analyzing emergency service ambulance deployment on a Brazilian highway using the hypercube model. *Journal of the Operation Research Society* 52, 261- 268.
- Owen, S.H., Daskin, M.S. (1998) Strategic facility location: A review. *European Journal of Operational Research* v.111, p. 423 - 447.
- Saydam C., Aytug H. (2003) Accurate estimation of expected coverage: revisited. *Socio-Economic Planning Sciences* v. 37, p. 69-80.
- Swersey A.J. (1994) *Handbooks in OR/MS*. Amsterdam: Elsevier Science B.V., v. 6, 151-200.
- Takeda R. A., Widmer, J. A., Morabito, R. (2004) Aplicação do modelo hipercubo de filas para avaliar a descentralização de ambulâncias em um sistema urbano de atendimento médico de urgência. *Pesquisa Operacional* v. 24, p. 39 – 72.
- Takeda R. A., Widmer, J. A., Morabito, R. (2005) Analysis of ambulance decentralization in an urban medical emergency service using the hypercube queueing model. *Computers and Operations Research* (in press).