

MODELO DE FILA HIPERCUBO COM MÚLTIPLO DESPACHO E *BACKUP* PARCIAL PARA ANÁLISE DE SISTEMAS DE ATENDIMENTO MÉDICO EMERGENCIAIS EM RODOVIAS

Ana Paula Iannoni

Reinaldo Morabito *

Departamento de Engenharia de Produção
Universidade Federal de São Carlos (UFSCar)
São Carlos – SP
papi@iris.ufscar.br
morabito@power.ufscar.br

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

Recebido em 06/2005; aceito em 05/2006
Received June 2005; accepted May 2006

Resumo

O modelo hipercubo é um modelo de filas espacialmente distribuídas baseado em aproximações Markovianas para analisar a configuração e operação de sistemas emergenciais servidor-para-cliente, em que os servidores se deslocam até os clientes. 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 da aplicação do modelo num estudo de caso de um SAE de uma concessionária de rodovias no interior de São Paulo são apresentados e analisados.

Palavras-chave: sistemas emergenciais; modelo hipercubo de filas; despacho de ambulâncias; rodovias.

Abstract

The hypercube is a spatially distributed queuing model based on Markovian analysis approximations to analyze the configuration and operation of server-to-customer emergency systems, in which servers travel to the locations of the customers. In the present study we adapt the model to analyze emergency medical systems (EMS) on highways operating with particular dispatching policies, which considers that only certain ambulances in the system can service calls in a given region (partial *backup*), and two or more identical or distinct ambulances can simultaneously service a single call (multiple dispatch). We also consider the situations in which 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 of the model application to a case study of an EMS operating on highways in São Paulo state are presented and analyzed.

Keywords: emergency medical systems; hypercube queuing model; ambulance deployment; highways.

1. Introdução

Nos sistemas de atendimento médico emergenciais (SAEs), o tempo médio de resposta é uma das medidas mais importantes para o nível de serviço ao usuário. 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-customer*), 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. No entanto, a maioria destes modelos considera apenas a aleatoriedade associada à disponibilidade dos servidores e não admitem que há outros aspectos probabilísticos que devem ser considerados na análise. 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.

Como o modelo hipercubo original de Larson (1974) admite hipóteses que podem limitar sua aplicação direta para análise dos SAEs, diversas adaptações e extensões vêm sendo estudadas, principalmente em patrulhamento policial e despacho de ambulâncias. Alguns estudos podem ser encontrados em Larson (1975), Halpern (1977), Chelst & Barlach (1981), Larson & Mcknew (1982), Jarvis (1985), Burwell *et al.* (1993), Mendonça & Morabito (2000, 2001) e Iannoni *et al.* (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 (2000, 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 *et al.* (1994), Saydam & Aytug (2003), Chiyoshi *et al.* (2003), Galvão *et al.* (2005) e Iannoni & Morabito (2006b).

Exemplos de aplicações do modelo hipercubo em SAEs nos Estados Unidos podem ser encontrados em Larson & Odoni (1981), Chelst & Barlach (1981), Brandeau & Larson (1986), Burwell *et al.* (1993) e Sacks & Grief (1994). Recentemente, o modelo hipercubo também vem sendo estudado para aplicação em sistemas de emergência que atuam em caso de ataques terroristas e catástrofes naturais de grande escala (Larson, 2004). 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, 2007; Costa, 2004) e alguns SAEs em rodovias do estado de São Paulo e Rio de Janeiro (Mendonça & Morabito, 2000, 2001; Iannoni, 2005; Iannoni *et al.*, 2005).

No presente artigo, estudamos modificações no modelo hipercubo para analisar os SAEs em rodovias considerando políticas particulares de despacho. No 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. No terceiro modelo, estendemos o primeiro modelo considerando que os servidores são diferenciados e em cada região são geradas chamadas que podem requer único ou múltiplo despacho de diferentes tipos de servidores.

Estas variações do modelo hipercubo podem também ser incorporadas em procedimentos de otimização que utilizam o modelo hipercubo, por exemplo, em heurísticas de substituição de vértices de forma a determinar a localização ótima das ambulâncias que maximiza a cobertura esperada dos usuários, tais como os estudos de Batta *et al.* (1989), Saydam & Aytug (2003), Chiyoshi *et al.* (2003) e Galvão *et al.* (2005). Para sistemas com moderado número de servidores, também podem ser utilizados simples métodos de enumeração, conforme estudado em Iannoni *et al.* (2005). 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 organizado da seguinte forma: a seção 2 apresenta uma breve descrição de duas diferentes configurações do SAE analisado; a seção 3 descreve cada uma das três variações do modelo hipercubo propostas para analisar o SAE; a seção 4 analisa os resultados das principais medidas de desempenho obtidas pelos modelos, validando estes resultados por meio de análise da amostra de dados e simulação discreta do sistema. Finalmente, a seção 5 apresenta as conclusões deste estudo e perspectivas para pesquisas futuras.

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 (nem sempre capaz de oferecer a mesma qualidade de serviço). 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. As ambulâncias destes SAEs têm a função de socorrer as vítimas de acidentes nas rodovias e, se necessário, realizar o transporte das mesmas ao hospital da cidade mais próxima.

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. Neste estudo,

analisamos duas configurações diferentes de um SAE de uma concessionária do estado de São Paulo. Os dados analisados das duas configurações foram obtidos em duas pesquisas de campo do mesmo SAE em épocas diferentes: A primeira configuração operava até 2004 e a segunda, operando desde então, corresponde à configuração anterior com certas modificações na política operacional, como descrito a seguir.

Configuração 1:

Na primeira configuração há cinco bases de ambulâncias (chamadas SAU – Serviço de Atendimento ao Usuário) ao longo do trecho de concessão, sendo que cada base possui uma ambulância. As ambulâncias são todas idênticas, e 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 transferida para outro sistema, e é considerada perdida para o sistema. O tempo de resposta em um duplo despacho é considerado igual ao da primeira ambulância que chega no local, que deve iniciar imediatamente o atendimento.

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, não há atendimento *backup*. Esta chamada é perdida para o sistema, mesmo que ambulâncias em outras bases estejam disponíveis.

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), e a terceira ambulância nunca é despachada. 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.

Na Figura 1 as localizações dos cinco SAUs (bases das ambulâncias) são: SAU 1 (km 217), SAU 2 (km 182), SAU 3 (km 144), SAU 4 (km 199) e SAU 5 (km 225). Estes correspondem a trechos de três rodovias paulistas: Rodovia SP-310 Washington Luiz (Cordeirópolis a São Carlos – SAU 1 e SAU 2), Rodovia SP-225 Eng. Paulo Nilo Romano (Itirapina a Jaú – SAU 3) e Rodovia SP-225 Comandante João Ribeiro de Barros (Jaú a Bauru – SAU 4 e SAU 5).

Configuração 2:

A configuração 2 apresenta algumas modificações com relação à configuração 1 do SAE analisado. A partir de 2004, o sistema passou a operar com seis veículos, sendo cinco ambulâncias idênticas (ou veículos resgates) e um veículo médico. O veículo médico diferencia-se do veículo resgate por ser um veículo mais leve que não permite transporte de vítimas, mas que transporta o médico, o enfermeiro, medicamentos, instrumentos básicos de socorro e operações emergenciais (p.e., equipamentos para realização de um parto). Os veículos resgates são mais pesados que o veículo médico dado que transportam, além das vítimas do acidente e equipamentos de primeiros socorros, equipamentos para combate a incêndio e quebra de ferragens para remoção das vítimas. A localização e o número de bases de emergência (cinco SAUs) ao longo da rodovia não sofreu alterações, sendo que o veículo médico está localizado no SAU 1 (km 217) (veja Figura 1).

Dado que os veículos resgates diferenciam-se do veículo médico nos aspectos descritos acima, a política de despacho do sistema também sofreu alterações. Por meio desta nova política, as ambulâncias passaram a ser despachadas de acordo com o tipo de chamada de emergência ao longo da rodovia. De acordo com a descrição dos gerentes e operadores do sistema, o despacho dos veículos pode ocorrer da seguinte forma:

- Despacho do veículo médico como único despacho: que ocorre para determinados tipos de chamadas que não requerem o uso de equipamentos mais especializados e/ou transporte de vítimas;

- Despacho de veículos resgates como único ou múltiplo despacho, tal como no sistema configuração 1. Em geral atendem, por exemplo, acidentes que envolvam transporte de vítimas, quebra de ferragens, combate a incêndio, remoção de vítimas ou animais, entre outros;
- Despacho do veículo médico e de um veículo resgate: diversas chamadas no sistema exigem a presença do médico no local do acidente para realizar ou orientar os procedimentos de emergência necessários, além da operação de um veículo resgate, por exemplo, para resgatar as vítimas e realizar o transporte das mesmas;
- Despacho de até três veículos (veículo médico e dois veículos resgates, ou três veículos resgates): que ocorre apenas quando o número de vítimas envolvidas no acidente está acima da capacidade de atendimentos de um ou dois veículos resgates.

Uma diferenciação importante da política de despacho da configuração 2 em relação à da configuração 1 é que pode ocorrer o despacho do terceiro veículo mais próximo. Ou seja, quando ocorre uma chamada que requer o despacho do veículo médico e de um resgate, e um destes está ocupado, o terceiro veículo mais próximo pode ser despachado. Assim, além de violar as mesmas hipóteses do modelo hipercubo original (da mesma forma que a configuração 1), na configuração 2 há chamadas que requerem único ou múltiplo despacho de diferentes tipos de veículos em cada região.

Nas seções seguintes, mostramos três extensões do modelo hipercubo múltiplo despacho para estudar os SAEs em rodovias:

- O primeiro modelo (modelo 1) é aplicado à configuração 1, cuja política de despacho implica em múltiplo despacho de ambulâncias idênticas e *backup* parcial. No entanto, esta extensão do modelo não analisa o sistema considerando as diferenças de atendimento entre chamadas atendidas na base e chamadas atendidas ao longo da rodovia;
- O segundo modelo (modelo 2) também estuda a configuração 1 do SAE analisado, mas trata as chamadas atendidas na base (que não requerem tempo de viagem) como um novo tipo de chamada, com tempo de atendimento diferenciado das chamadas de emergência com único e duplo despacho de ambulância;
- O terceiro modelo (modelo 3) analisa a configuração 2 do sistema, considerando que a operação deste SAE envolve único e múltiplo despacho de ambulâncias idênticas e diferenciadas (veículo médico e veículos resgates) e *backup* parcial.

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 (vértice) 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 (vértices) para o sistema. Por exemplo, em um sistema com $N = 3$ servidores, o estado 101 corresponde ao estado (vértice de um cubo) em que os servidores 1 e 3 estão ocupados e o servidor 2 está livre.

As principais hipóteses que são comuns as três extensões 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 um processo de Poisson. As chamadas podem ser diferenciadas de acordo com o tipo de chamada (modelos 1, 2 e 3), tipo de atendimento (modelo 2) e tipo de veículo despachado de acordo com o tipo de chamada (modelo 3). Por exemplo, no modelo 1, as chamadas podem ser de dois tipos: 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*. Uma área preferencial de uma ambulância corresponde aos átomos para os quais a ambulância é despachada se disponível, mesmo que todas as outras ambulâncias estejam disponíveis.
- 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. Descrevemos a seguir como a política de despacho é definida em cada uma das configurações e como cada modelo é definido de forma a considerá-la.
- 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 admite que o tempo de atendimento é representado por uma distribuição exponencial negativa, mas tem sido observado que desvios razoáveis desta suposição não alteram significativamente a precisão do modelo. Dado que os SAEs em rodovias não admitem filas, esta suposição é ainda menos 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).
- 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)

Como mencionado anteriormente, o modelo 1 pode ser aplicado aos SAEs em rodovias cuja política de despacho é similar a da configuração 1 do estudo de caso (seção 2). 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 μ_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 μ_i e μ_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. Como discutido em Chelst & Barlach (1981), tal consideração mantém o mesmo número de estados do sistema do modelo hipercubo original (2^N), pois se diferenciássemos o tempo de atendimento para chamadas tipo 2, teríamos que considerar um estado adicional para cada ambulância (quando ocupada atendendo uma chamada tipo 2), e o número de estados possíveis para o sistema passaria para 3^N .

Equações de equilíbrio no modelo 1:

Para descrever brevemente como as equações de equilíbrio são definidas no modelo 1, utilizamos um sistema exemplo com $N_A = 4$ átomos e $N = 3$ servidores. Neste exemplo, em cada átomo j , pode ocorrer chamadas tipo 1 ($\lambda_j^{[1]}$) e tipo 2 ($\lambda_j^{[2]}$), conforme política de despacho descrita acima. A lista de preferência de despacho para cada átomo com duas ambulâncias (preferencial e *backup*) é composta da seguinte forma: átomo 1 – ambulâncias 1 e 2; átomo 2 – ambulâncias 2 e 1; átomo 3 – ambulâncias 2 e 3 e átomo 4 – ambulâncias 3 e 2, sendo que, a lista de preferência de despacho do átomo é a mesma para chamadas tipo 1 e tipo 2.

Analisando um dos $2^3 = 8$ estados possíveis do sistema, por exemplo, o estado $\{110\}$ (onde os servidores 1 e 2 estão ocupados e o servidor 3 está livre), temos a equação de equilíbrio:

$$(\lambda_3^{[1]} + \lambda_4^{[1]} + \lambda_3^{[2]} + \lambda_4^{[2]} + \mu_1 + \mu_2).P_{110} = (\lambda_1^{[2]} + \lambda_2^{[2]}).P_{000} + (\lambda_1^{[1]} + \lambda_2^{[1]} + \lambda_1^{[2]} + \lambda_2^{[2]}).P_{010} + (\lambda_1^{[1]} + \lambda_2^{[1]} + \lambda_3^{[1]} + \lambda_4^{[1]} + \lambda_3^{[2]} + \lambda_4^{[2]}).P_{100} + \mu_3.P_{111}$$

Note que esta equação é formulada considerando que, em equilíbrio, o fluxo para dentro do estado deve ser igual ao fluxo para fora do estado. Assim, as possíveis transições para fora deste estado são:

$\{110\} \rightarrow \{111\}$ – ocorre com a chegada de uma chamada tipo 1 ou tipo 2 nos átomos 3 e 4 (taxa total $\lambda_3^{[1]} + \lambda_4^{[1]} + \lambda_3^{[2]} + \lambda_4^{[2]}$), pois o servidor 3 (único disponível), só atende chamadas nestes átomos;

$\{110\} \rightarrow \{010\}$ e $\{110\} \rightarrow \{100\}$ – ocorre com o término de serviço do servidor 1 (taxa μ_1) e servidor 2 (taxa μ_2), respectivamente.

As transições para dentro do estado $\{110\}$ são:

$\{111\} \rightarrow \{110\}$ – término de serviço do servidor 3 (taxa μ_3);

$\{010\} \rightarrow \{110\}$ – chegada de uma chamada tipo 1 ou tipo 2 no átomo 1 (taxa $\lambda_1^{[1]} + \lambda_1^{[2]}$) ou no átomo 2 (taxa $\lambda_2^{[1]} + \lambda_2^{[2]}$). Note que uma chamada tipo 2 é atendida por apenas um servidor disponível da sua lista de despacho (servidor 1);

$\{100\} \rightarrow \{110\}$ – chegada de uma chamada do tipo 1 nos átomos 1, 2 ou 3 (taxa total $\lambda_1^{[1]} + \lambda_2^{[1]} + \lambda_3^{[1]}$) e do tipo 2 nos átomos 1 e 2 (taxa total $\lambda_1^{[2]} + \lambda_2^{[2]}$). Note que uma chamada tipo 2 nos átomos 1 ou 2 só pode ser atendida pelo servidor 2, mesmo que o servidor 3 também esteja disponível;

$\{000\} \rightarrow \{110\}$ – chegada de uma chamada tipo 2 nos átomos 1 ou 2 (taxa total $\lambda_1^{[2]} + \lambda_2^{[2]}$).

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. Os autores chamam estes eventos de PIAs (*Patrol Initiated Activities*). 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 $\mu_i^{[1]}$ para chamadas tipo 1 (único despacho) e tipo 2 (duplo despacho) e $\mu_i^{[1a]}$ 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.

Equações de equilíbrio no modelo 2:

Utilizamos um exemplo com $N_A = 4$ átomos e $N = 3$ servidores, similar ao da seção 3.1, para descrever brevemente como as equações de equilíbrio são definidas no modelo 2. No entanto, neste caso consideramos que além de chamadas tipo 1 ($\lambda_j^{[1]}$) e tipo 2 ($\lambda_j^{[2]}$), ocorrem também chamadas do tipo 1a ($\lambda_j^{[1a]}$), que são diferenciadas e ocorrem na base dos servidores. De acordo com a mesma lista de despacho descrita para o exemplo da seção 3.1, chamadas tipo 1a em cada átomo só podem ser atendidas pelo servidor preferencial. Como discutido anteriormente, cada servidor passa a ter 3 estados e há $3^3 = 27$ estados possíveis para sistema. Analisando, por exemplo, o estado $\{210\}$, temos a equação de equilíbrio:

$$(\lambda_3^{[1]} + \lambda_4^{[1]} + \lambda_3^{[2]} + \lambda_4^{[2]} + \lambda_4^{[1a]} + \mu_1^{[II]} + \mu_2^{[I]}).P_{210} = (\lambda_4^{[1a]}).P_{010} + (\lambda_1^{[1]} + \lambda_2^{[1]} + \lambda_3^{[1]} + \lambda_1^{[2]} + \lambda_2^{[2]}).P_{200} + \mu_3^{[I]}.P_{211} + \mu_3^{[II]}.P_{212}$$

As transições para fora do estado $\{210\}$ (lado esquerdo da equação) são:

$\{210\} \rightarrow \{211\}$ – chegada de uma chamada tipo 1 ou tipo 2 nos átomos 3 e 4 (taxa total $\lambda_3^{[1]} + \lambda_4^{[1]} + \lambda_3^{[2]} + \lambda_4^{[2]}$);

$\{210\} \rightarrow \{212\}$ – chegada de uma chamada tipo 1a no átomo 4 (taxa $\lambda_4^{[1a]}$), atendida pelo servidor 3;

$\{210\} \rightarrow \{010\}$ – término de serviço do servidor 1 ao atender uma chamada tipo 1a (taxa $\mu_1^{[II]}$);

$\{210\} \rightarrow \{200\}$ – término de serviço do servidor 2 ao atender uma chamada tipo 1 ou tipo 2 – chamadas de emergência (taxa $\mu_2^{[I]}$).

As transições para dentro do estado $\{210\}$ são:

$\{211\} \rightarrow \{210\}$ – término de serviço do servidor 3 ao atender uma chamada tipo 1 ou tipo 2 (taxa $\mu_3^{[I]}$);

$\{212\} \rightarrow \{210\}$ – término de serviço do servidor 3 ao atender uma chamada tipo 1a no átomo 4 (taxa $\mu_3^{[II]}$);

$\{010\} \rightarrow \{210\}$ – ocorre com a chegada de uma chamada tipo 1a no átomo 1 (taxa $\lambda_1^{[1a]}$);

$\{200\} \rightarrow \{210\}$ – ocorre com a chegada de uma chamada tipo 1 nos átomos 1, 2 ou 3 (taxa total $\lambda_1^{[1]} + \lambda_2^{[1]} + \lambda_3^{[1]}$) e tipo 2 nos átomos 1 e 2 (taxa total $\lambda_1^{[2]} + \lambda_2^{[2]}$).

3.3 Modelo hipercubo múltiplo despacho com servidores diferenciados (modelo 3)

Alguns SAEs em rodovias operam com servidores diferenciados (além de chamadas diferenciadas do tipo 1 e 2), sendo que o despacho do tipo de servidor depende do tipo de chamado. Por exemplo, alguns destes SAEs possuem veículos resgates, que são veículos em geral mais pesados que transportam, além de equipamentos de primeiros socorros,

equipamentos para combate a incêndio e quebra de ferragens para remoção das vítimas de acidente. Outro veículo comum em alguns SAEs é o veículo de atendimento avançado (também chamado de UTI móvel) que possui equipamentos mais especializados para tratar o paciente durante o transporte, mas que não transporta outros equipamentos pesados (p.e, de combate a incêndio e remoção de ferragens) como o veículo resgate. Além disso, outra possível diferença entre estes dois tipos de veículos é a equipe de profissionais transportada (resgatistas, médicos, enfermeiros). Esta situação é semelhante a dos sistemas urbanos SAMU, descrita em Takeda *et al.* (2004; 2007), onde as chamadas podem ser do tipo avançada (muito urgentes) ou básicas (urgentes) e, dependendo do tipo de chamada, é despachado prioritariamente um veículo de suporte avançado (VSA) ou um veículo de suporte básico (VSB).

Os SAEs similares à configuração 2 do estudo de caso (seção 2) operam com veículos que transportam apenas o médico ao local do acidente (veículo médico), despachado simultaneamente com o veículo resgate. O veículo médico não transporta pacientes, medicamentos ou equipamentos. Em algumas das mais congestionadas rodovias do Brasil, os SAEs utilizam também helicópteros para socorro das vítimas para determinados tipos de acidente.

Além das demais características de SAEs com múltiplo despacho, *backup* parcial e servidores idênticos (configuração 1), devemos também considerar que nestes sistemas:

- As chamadas em cada átomo são diferenciadas (p.e., tipo 1, 2, 3) e requerem diferentes tipos de servidores, portanto cada tipo de chamada em cada átomo possui uma lista de preferência de despacho. Assim, um dado átomo possui mais de uma lista de preferência. Note que esta situação é diferente das anteriores, todas com mesma lista de despacho por átomo.
- A política de despacho torna-se mais complexa dado que podemos ter chamadas que requerem: 1 servidor, 2 servidores idênticos, 2 servidores diferenciados, 3 servidores idênticos, 3 servidores diferenciados, etc.
- Além disso, a política de *backup* parcial também torna-se mais complexa dado que podemos ter chamadas com mais de 2 servidores candidatos em sua lista de despacho. Por exemplo, nos casos em que uma das bases mais próximas abriga dois servidores diferenciados.

O modelo 3 é uma extensão do modelo hipercubo múltiplo despacho para tratar SAEs em rodovias similares a configuração 2 do estudo de caso. Neste sistema há dois tipos de veículos: veículo resgate e veículo médico. Cada átomo j do sistema possui chamadas de 4 tipos:

- Chamadas tipo 1 (taxa de chegada $\lambda_j^{[1]}$) que requerem um único despacho (um veículo resgate ou um veículo médico), dispondo para isso de até 2 possíveis servidores (como na configuração 1);
- Chamadas tipo 2a (taxa de chegada $\lambda_j^{[2a]}$) que requerem duplo despacho (dois veículos resgates idênticos), dispondo de até 2 possíveis servidores (como na configuração 1);
- Chamadas tipo 2b (taxa de chegada $\lambda_j^{[2b]}$) que requerem duplo despacho (dois veículos diferenciados: 1 veículo resgate e o veículo médico), dispondo de até 3 possíveis servidores (há um terceiro veículo resgate que é *backup* de um dos dois primeiros, caso esteja ocupado);

- Chamadas tipo 3 (taxa de chegada $\lambda_j^{[3]}$) que requerem triplo despacho (dois veículos resgate e o veículo médico, ou três veículos resgates), dispondo de até 3 possíveis servidores (não há backup para este tipo de chamada).

Note que consideramos neste sistema que o veículo médico pode ter como backup um veículo resgate (chamadas tipo 2b), o que ocorre de fato no sistema real analisado. Esta hipótese também torna mais simples a análise. Caso contrário, teríamos que considerar para cada tipo de chamada, uma lista de servidores backup para o veículo médico e o veículo resgate, separadamente.

Para descrever o modelo 3 utilizamos, como nas seções 3.1 e 3.2, um sistema exemplo com operação similar à configuração 2 do SAE do estudo de caso. Este sistema, representado na Figura 2, possui $N_A = 3$ átomos e $N = 4$ servidores, sendo três veículos resgates (R1, R2 e R3) e um veículo médico (CM). Similarmente a configuração 2 do sistema real, o veículo médico divide sua base com o resgate 1 no átomo 1, o resgate 2 está localizado no átomo 2 e o resgate 3 no átomo 3. Em cada átomo j , chamadas podem ser do tipo 1, 2a, 2b e 3, com taxas $\lambda_j^{[1]}$, $\lambda_j^{[2a]}$, $\lambda_j^{[2b]}$ e $\lambda_j^{[3]}$.

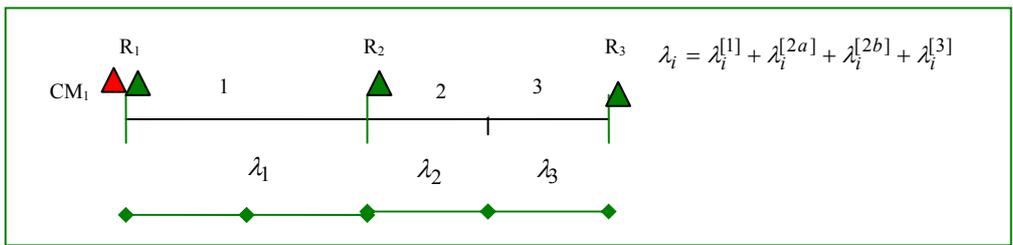


Figura 2 – SAE em rodovia – múltiplo despacho com veículo médico.

Para representar os diferentes tipos de chamadas que possuem diferentes listas de preferência de despacho, subdividimos cada átomo de acordo com o número de listas de preferências. No caso deste exemplo, cada átomo é subdividido em duas camadas (“layering”): camada a para chamadas que são atendidas somente por veículos resgate e camada b para chamadas que são atendidas por veículos médico e resgate. A lista de preferência de despacho de cada sub-átomo (camada do átomo) é dada a seguir na Tabela 1, considerando servidor 1 (veículo médico), servidor 2 (resgate 1), servidor 3 (resgate 2) e servidor 4 (resgate 3). Note que chamadas do tipo 3 ocorrem nos átomos 1b, 2b e 3a (ao invés de 3b).

Tabela 1 – Subdivisão de átomos de acordo com o tipo de chamada e lista de despacho.

Sub-átomo	Chamadas	Primeiro	Segundo	Terceiro
1a	1, 2a	2	3	-
1b	2b,3	1	2	3
2a	1,2a	3	2	
2b	2b,3	1	3	2
3a	1,2a,3	4	3	2
3b	2b	1	4	3

A política de despacho para chamadas tipo 1 e tipo 2a ocorre de forma similar ao modelo 1. Considerando a política de múltiplo despacho de servidores diferenciados, no caso do exemplo da Figura 2, temos:

- Chamadas tipo 2b são atendidas pelo primeiro servidor preferencial (veículo médico) e o segundo servidor preferencial (veículo resgate). Quando um destes estiver ocupado, o terceiro servidor da lista de preferência é despachado. Quando apenas um dos três servidores está livre, o mesmo é despachado sozinho, podendo receber auxílio de servidores de outro SAE mais próximo. Se os três servidores estiverem ocupados, a chamada é perdida para o sistema (átomos 1b, 2b e 3b);
- Chamadas tipo 3 são atendidas pelo veículo médico e dois veículos resgates (átomos 1b e 2b), ou por três veículos resgates (átomo 3a). Se apenas dois dos três servidores estiverem livres, os mesmos devem atender a chamada como duplo despacho. Se há apenas um servidor disponível, o mesmo deve também ser despachado sozinho como descrito acima. Se os três servidores estiverem ocupados, a chamada é perdida para o sistema.

Equações de equilíbrio no modelo 3:

Como nos casos anteriores, analisamos um dos $2^4 = 16$ possíveis estados do sistema exemplo, para ilustrar como as equações de equilíbrio do modelo 3 são definidas. Por exemplo, a equação de equilíbrio do estado {1110}:

$$\begin{aligned}
 (\lambda_{3a}^{[1]} + \lambda_{3a}^{[2a]} + \lambda_{3b}^{[2b]} + \lambda_{3a}^{[3]} + \mu_1 + \mu_2 + \mu_3).P_{1110} = & (\lambda_{1b}^{[3]} + \lambda_{2b}^{[3]}).P_{0000} + (\lambda_{1a}^{[1]} + \lambda_{1a}^{[2a]} + \lambda_{1b}^{[2b]} + \lambda_{1b}^{[3]} + \\
 & \lambda_{2a}^{[1]} + \lambda_{2a}^{[2a]} + \lambda_{2b}^{[2b]} + \lambda_{2b}^{[3]}).P_{1000} + (\lambda_{1b}^{[2b]} + \lambda_{1b}^{[3]} + \lambda_{2b}^{[2b]} + \lambda_{2b}^{[3]}).P_{0100} + (\lambda_{1b}^{[2b]} + \lambda_{1b}^{[3]} + \lambda_{2b}^{[2b]} + \lambda_{2b}^{[3]}).P_{0010} + \\
 & (\lambda_{1a}^{[1]} + \lambda_{1a}^{[2a]} + \lambda_{1b}^{[2b]} + \lambda_{1b}^{[3]} + \lambda_{2a}^{[1]} + \lambda_{2a}^{[2a]} + \lambda_{2b}^{[2b]} + \lambda_{2b}^{[3]}).P_{1100} + (\lambda_{1a}^{[1]} + \lambda_{1a}^{[2a]} + \lambda_{1b}^{[2b]} + \lambda_{1b}^{[3]} + \lambda_{2a}^{[1]} + \lambda_{2a}^{[2a]} + \\
 & \lambda_{2b}^{[2b]} + \lambda_{2b}^{[3]}).P_{1010} + (\lambda_{1b}^{[2b]} + \lambda_{1b}^{[3]} + \lambda_{2b}^{[2b]} + \lambda_{2b}^{[3]}).P_{0110} + \mu_4.P_{1111}
 \end{aligned}$$

Note no lado esquerdo da equação acima (fluxo para fora do estado {1110}) que somente chamadas geradas nos átomos 3a e 3b podem ser atendidas pelo sistema, quando no estado {1110}, pois somente o servidor 4 está livre e atende chamadas sozinho.

No lado direito da equação (fluxo para dentro do estado {1110}), verificamos que três servidores (1, 2 e 3) podem passar de livre a ocupados simultaneamente quando chamadas tipo 3 são geradas nos átomos 1 e 2 (transição {0000} → {1110}). Além disso, note que as transições {0100} → {1110} e {0010} → {1110} podem ocorrer com a chegada de chamadas tipo 2b nos átomos 1 e 2, dado que se um dos dois servidores preferenciais estiver ocupado, o terceiro servidor também é despachado. Por outro lado, chamadas tipos 2a, 2b e 3 que encontram um único servidor disponível são atendidas por este servidor, por exemplo, nas transições: {0110} → {1110}, {1010} → {1110} e {1100} → {1110}.

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 as duas ambulâncias despachadas para uma chamada tipo 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.

Por meio do modelo 3, podemos calcular medidas adicionais considerando cada tipo m de chamada ($m = 1, 2a, 2b$ e 3) e o número de servidores enviados (único, duplo ou triplo despacho), tais como: a probabilidade de perda para qualquer chamada do sistema e para cada tipo de chamada m , as frequências de despacho de cada ambulância a cada átomo considerando cada tipo de chamada m , a frequência de despacho e o tempo médio de viagem no sistema para chamadas tipo m quando ocorre único despacho, a frequência de despacho e o tempo médio de viagem para chamadas com duplo despacho para chamadas tipos $2a$ e $2b$ e tipo 3 (quando somente dois servidores estão disponíveis), o tempo médio de viagem para a primeira, segunda e terceira (no caso de triplo despacho) ambulância a chegar no local da chamada, entre outras.

3.4 Outras medidas de desempenho para modelos 1, 2 e 3

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 (p.e. no modelo 3, a probabilidade de perda para chamadas tipo m ($P_p^{[m]}$)). Mais detalhes sobre o cálculo das probabilidades de perda nos modelos 1, 2 e 3 podem ser encontrados em Iannoni (2005).

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 :

$$f_{ij}^{[1]} = \frac{(\lambda_j^{[1]} / \lambda^{[1]}) \sum_{B \in E_{ij}} P_B}{(1 - P_p^{[1]})}, \text{ onde } E_{ij} \text{ corresponde ao conjunto de estados em que a}$$

ambulância i é a primeira ambulância disponível na lista de preferência do átomo j e

$$\lambda^{[1]} = \sum_{j=1}^{N_A} \lambda_j^{[1]}. \text{ Esta expressão é válida para os três modelos descritos acima.}$$

- Fração de despachos tipo 2 (modelos 1 e 2) e tipo 2a (modelo 3) em que as ambulâncias i e k são enviadas ao átomo j :

$$f_{(i,k)j}^{[2]} = \frac{(\lambda_j^{[2]} / \lambda^{[2]}) \sum_{B \in E_{(i,k)j}} P_B}{(1 - P_p^{[2]})}, \text{ onde } E_{(i,k)j} \text{ corresponde ao conjunto de estados em que}$$

as ambulâncias 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 (modelo 1 e 2) em que a ambulância i é a única ambulância enviada para o átomo j , quando a ambulância k está ocupada:

$$f_{ij}^{[2]} = \frac{(\lambda_i^{[2]} / \lambda^{[2]}) \sum_{B \in F_{ij}} P_B}{(1 - P_p^{[2]})}, \text{ onde } F_{ij} \text{ corresponde ao conjunto de estados em que}$$

somente a ambulância i pode responder chamadas no átomo j . No modelo 3 podem ocorrer despachos de uma única ambulância para uma chamada do tipo 2a (se uma das duas ambulâncias da lista estiver ocupada); tipo 2b e 3 (se duas das três ambulâncias da lista estiverem ocupadas).

Note que $\sum_{i=1}^N \sum_{j=1}^{N_A} f_{ij}^{[1]} = 1$ e $\sum_{j=1}^{N_A} \left[\sum_{i=1}^N f_{ij}^{[2]} + \sum_{i=1}^{N-1} \sum_{k=i+1}^N f_{(i,k)j}^{[2]} \right] = 1$. 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 tipo m ($f_{ij}^{[m]}$, $f_{(i,k)j}^{[m]}$ e $f_{(i,k,l)j}^{[m]}$, em caso de triplo despacho no modelo 3) e fração de todos os despachos que envia a ambulância i para o átomo j para responder qualquer tipo m de chamadas, entre outras. Lembre-se que no modelo 1: $m = 1$ e 2, no modelo 2: $m = 1, 2$ e 1a, e no modelo 3: $m = 1, 2a, 2b$ e 3. Por exemplo, as frequências de duplo despacho considerando todos os despachos do sistema para chamadas tipo 2a, 2b e 3, são definidas da seguinte forma:

$$(i) \text{ Chamadas tipo 2a e 2b: } f_{(i,k)j}^{[2a]} = \frac{\lambda_j^{[2a]} \sum_{B \in E(i,k)j} P_B}{(1 - P_p)} \text{ e } f_{(i,k)j}^{[2b]} = \frac{\lambda_j^{[2b]} \sum_{B \in E(i,k)j} P_B}{(1 - P_p)}$$

Note que, diferentemente das chamadas tipo 2a, no caso de chamadas tipo 2b, a ambulância i ou k pode ser o terceiro servidor da lista de preferência de despacho do átomo j , se a primeira ou segunda ambulância estiver ocupada.

- (ii) Chamadas tipo 3 (atendidas por duas ambulâncias i e k – únicos disponíveis na lista de preferência de despacho):

$$f_{(i,k)j}^{[3]} = \frac{\lambda_j^{[3]} \sum_{B \in F(i,k)j} P_B}{(1 - P_p)}, \text{ onde } F_{(i,k)j} \text{ corresponde ao conjunto de estados em que as}$$

ambulâncias i e k são as únicas disponíveis na lista de preferência de despacho do átomo j .

Diversas medidas adicionais de frequência podem ser definidas tais como: frequências de único despacho para chamadas tipo m , frequências de triplo despacho para chamadas tipo 3 (considerando todos os despachos do sistema ou todos os despachos tipo 3), entre outras.

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 para o modelo 1. Lembre-se 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:

$T^{[1]} = \sum_{j=1}^{N_A} \sum_{i=1}^N f_{ij}^{[1]} \bar{t}_{ij}$, onde \bar{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):

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

- Tempo médio de viagem para as duas ambulâncias enviadas para atender uma chamada tipo 2 (modelos 1 e 2):

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

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

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

- Tempo médio de viagem para a ambulância i – único despacho e duplo despacho:

$$\overline{TU}_i^{[1]} = \frac{\sum_{j=1}^{N_A} f_{ij}^{[1]} \bar{t}_{ij}}{\sum_{j=1}^{N_A} f_{ij}^{[1]}} \quad \overline{TU}_i^{[2]} = \frac{\sum_{j=1}^{N_A} \left[\sum_{k \neq i}^N f_{(i,k)j}^{[2]} \bar{t}_{ij} + f_{ij}^{[2]} \bar{t}_{ij} \right]}{\sum_{j=1}^{N_A} \left[\sum_{k \neq j}^{N-1} f_{(i,k)j}^{[2]} + f_{ij}^{[2]} \right]}$$

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

$\bar{T} = \left(\sum_{j=1}^{N_A} \sum_{i=1}^N f_{ij}^{[1]} + f_{ij}^{[2]} \right) \bar{x}_{ji} + \left(\sum_{j=1}^{N_A} \sum_{i=1}^{N-1} \sum_{k=i+1}^N f_{(i,k)j}^{[2]} \right) \cdot \min(\bar{t}_{ij}, \bar{t}_{kj})$, onde $f_{ij}^{[1]}$, $f_{ji}^{[2]}$, $f_{(i,l)j}^{[2]}$ correspondem a frequência de despachos para chamadas do tipo 1 e 2, considerando todos os despachos do sistema.

No modelo 3, além das medidas agregadas de tempo de viagem descritas pelos modelos 1 e 2, podemos definir outras medidas adicionais de acordo com o tipo de chamada e o número de ambulâncias enviadas (único, duplo ou triplo despacho). Entre estas estão: tempo médio de viagem no sistema para chamadas tipo m quando ocorre único despacho, tempo médio de viagem para chamadas com duplo despacho para chamadas tipos 2a, 2b e tipo 3 (quando somente dois servidores estão disponíveis), tempo médio de viagem para a primeira, segunda e terceira (no caso de triplo despacho) ambulância a chegar no local da chamada, entre outras.

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):

$P_{t > \bar{t}_v} = \sum_{i=1}^N \sum_{j=1}^{N_A} f_{ij}^{[v]}$, 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 . No modelo 1, por exemplo, esta medida pode ser definida como: $f_{ij}^{[v]} = p(t_{ij} > \bar{t}_v) \left(f_{ij}^{[1]} + f_{ij}^{[2]} + \sum_{k \in L_j^{[v]}} f_{(i,k)j}^{[2]} \right)$, onde t_{ij} corresponde a variável aleatória do tempo de viagem da ambulância i ao átomo j ; $L_j^{[v]}$ corresponde ao conjunto de ambulâncias da lista de despacho do átomo j tal que $t_{ij} < t_{kj}$, e o termo $p(t_{ij} > \bar{t}_v)$ corresponde a fração de chamadas no átomo j que esperam mais que \bar{t}_v minutos pela chegada da ambulância i (calculado a partir dos dados de tempo de viagem da amostra).

4. Resultados Computacionais

4.1 Análise da configuração 1 usando os modelos 1 e 2

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. Como mencionado anteriormente, os dados de entrada das duas configurações do sistema foram coletados em épocas diferentes na central de operações do SAE. 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 2 – 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 backup	Servidor i	Taxa de serviço μ_i (ch/h)	Carga de trabalho ρ_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, para todos os servidores, a distribuição dos tempos de atendimento seja exponencial, similarmente ao sistema analisado por Mendonça & Morabito (2000; 2001). Porém, como comentado em Jarvis (1985) e na seção 3, dado que o sistema não admite filas, desvios razoáveis desta suposição não alteraram significativamente a precisão do modelo. Aplicamos também a análise de variância para verificar a variabilidade entre os tempos médios de atendimento dos servidores. Esta análise mostrou que, com nível de significância $\alpha = 0,05$, as diferenças são significantes e os servidores devem ser considerados não homogêneos na aplicação do modelo hipercubo (i.e., as taxas de atendimento μ_i são distintas).

As equações de equilíbrio são definidas como no exemplo da seção 3.1. Como a configuração 1 do SAE analisado tem $N = 5$ ambulâncias, há $2^5 = 32$ estados possíveis. De acordo, com a lista de preferência de despacho na Tabela 2, as equações de equilíbrio podem ser formuladas, por exemplo, como a expressão abaixo para o estado $\{11001\}$:

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

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 ρ_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 ρ_1 (máximo valor) é mais que 200% maior que ρ_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: $P_{t_v > 10}$. Esta medida foi calculada utilizando os dados da amostra da distribuição de chamadas atendidas em tempo maior que 10 minutos para cada servidor em cada átomo. 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. Note que, apesar do sistema permanecer a maior parte do tempo ocioso (dada a relação demanda/capacidade da Tabela 2), estas frações de chamadas atendidas em tempo superior a $\bar{t}_v = 10$ minutos são relativamente altas, se comparadas com outros sistemas com boa qualidade de serviço. De acordo com Gendreau *et al.* (2001) e Galvão *et al.* (2005), a legislação americana estabelece que 95% de chamadas sejam atendidas dentro de 10 minutos. De forma similar, nas cidades de Montreal e Londres, 70% e

50% das chamadas devem ser atendidas dentro de 7 e 8 minutos, respectivamente. No Brasil, apesar de não haver uma legislação que estabeleça limites para o tempo de resposta nos SAEs urbanos, as concessionárias de rodovias têm procurado reduzir as frações $P_{n>10}$ nos seus SAEs para melhorar a qualidade do serviço e satisfazer cláusulas contratuais com os governos dos estados.

Para calcular algumas medidas de tempo de viagem da seção 3, utilizamos a matriz dos tempos de viagem de cada base a cada átomo, obtida pela análise da amostra de dados. 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. 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 $\bar{T}^{[2]}$ e \bar{T}'_t (dados pelas expressões abaixo), ao invés de $\bar{T}^{[2]}$ e \bar{T}_t (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.

$$\bar{T}^{[2]} = \frac{\sum_{j=1}^{N_A} \sum_{i=1}^{N-1} \sum_{k=i+1}^N f_{(i,k)j}^{[2]} \min(\bar{t}_{ij}, \bar{t}_{kj})}{\sum_{j=1}^{N_A} \sum_{i=1}^{N-1} \sum_{k=i+1}^N f_{(i,k)j}^{[2]}} \quad \bar{T}'_t = \frac{\sum_{j=1}^{N_A} \sum_{i=1}^{N-1} \sum_{k=i+1}^N f_{(i,k)j}^{[2]} (\bar{t}_{ij} + \bar{t}_{kj})}{\sum_{j=1}^{N_A} \sum_{i=1}^{N-1} \sum_{k=i+1}^N f_{(i,k)j}^{[2]}}$$

Algumas medidas obtidas pelo modelo 1 da seção 3.1 foram também validadas por um modelo de simulação, por meio da análise do intervalo de confiança (nível de confiança $\alpha = 0,05$) dos resultados. Para construir o modelo de simulação realizou-se a análise estatística dos tempos de atendimento sem os tempos de viagem da base ao local da chamada, dado que a simulação utiliza estes dados como um dado de entrada independente do tempo atendimento (lembre-se que no modelo hipercubo o tempo médio de atendimento inclui os tempos de viagem). O modelo foi construído e analisado por meio do *software* Arena. Utilizamos um procedimento descrito em Iannoni & Morabito (2002, 2006a), baseado no método de loteamento (Law & Kelton, 1991; Pedgen *et al.*, 1995; Kelton *et al.*, 2002), para calcular o tempo de fase transiente e o tempo total de simulação necessário para análise dos resultados. Na quinta e sexta colunas da Tabela 3, encontram-se o valor médio e o intervalo de confiança para cada medida obtida pela simulação.

Tabela 3 – 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}'_t	25,130	25,144	0,06%		
\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 3 que os resultados do modelo apresentam desvios relativamente pequenos com relação à amostra. Além disso, o modelo também é validado pelo modelo de simulação, considerando que os resultados obtidos pelo modelo hipercubo estão dentro do intervalo de confiança dos resultados da simulação (sexta coluna da Tabela 3).

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 4 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 $\mu_i^{[II]}$ (tipo 1a) para cada ambulância i do sistema.

Tabela 4 – Dados dos processos de chegada e atendimento com chamadas tipo 1a.

Átomo j	Chegada tipo 1 $\lambda_j^{[1]}$ (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 $\mu_i^{[II]}$ (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 a discussão do modelo 2 na seção 3.2, temos $3^5 = 243$ possíveis estados para o sistema. Calculando as probabilidades de estado, 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 ($\rho_i^{[II]}$) para cada ambulância i , e como esperado os resultados de ρ_i (modelo 1) são, aproximadamente, a soma dos resultados de $\rho_i^{[I]}$ e $\rho_i^{[II]}$ (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: $\bar{T}^{[1]} = 7,376$ 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, $\bar{T}^{[1]} = 6,277$ (14,89% menor do que o modelo 2). A Tabela 5 apresenta os resultados do tempo médio de viagem de cada ambulância para chamadas tipo 1 ($\bar{TU}_i^{[1]}$), comparando-os com os resultados da análise da amostra (desvio relativo).

Tabela 5 – Tempo médio de viagem servidor – único despacho $\overline{TU}_i^{[1]}$ (em minutos).

Servidor <i>i</i>	$\overline{TU}_i^{[1]}$ modelo 2	$\overline{TU}_i^{[1]}$ amostra	Desvio	\overline{TU}_i simulação (média)	Simulação (Intev. Conf.)	$\overline{TU}_i^{[1]}$ 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	-3,6%	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%
$\overline{T}^{[1]}$	7,376	7,196	-2,50%			6,277	17,51%

Note nas colunas 7 e 8 da Tabela 5 que o tempo médio de viagem de cada servidor aumenta significativamente com relação aos resultados desta medida obtidos pelo modelo 1. Por exemplo, para a 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 $\mu_i^{[III]}$, 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. As colunas 5 e 6 da Tabela 5 comparam os resultados do modelo 2 com os resultados do modelo de simulação considerando chamadas tipo 1a (via intervalo de confiança, de forma similar à análise do modelo 1 descrita acima).

4.2 Análise da configuração 2 usando o modelo 3

Para aplicação do modelo 3, utilizamos os dados coletados da configuração 2 do SAE analisado, sendo mantida a divisão do trecho total em 8 átomos. Os dados foram coletados durante o período de janeiro a setembro de 2004, com um total de 1498 eventos. Como discutido na seção 3.3, a aplicação do modelo hipercubo múltiplo despacho para um sistema com servidores diferenciados requer que os átomos do sistema sejam subdivididos em camadas, de acordo com o tipo de chamada que possui uma lista de despacho particular. A Tabela 6 apresenta como os 8 átomos são subdivididos de acordo com o tipo de chamada e a lista de preferência de despacho para cada sub-átomo.

Note que, chamadas tipo 1 e 2a possuem 2 servidores em sua lista de despacho, e chamadas tipo 2b e 3 possuem até 3 servidores possíveis. Não foi necessário subdividir os átomos 5, 7 e 8 em camadas, dado que todos os tipos de chamada nestes átomos possuem uma mesma lista de despacho. Similarmente às aplicações dos modelos 1 e 2 na configuração 1 do SAE, realizamos análises estatísticas dos dados dos processos de chegada e atendimento utilizando os testes de aderência e análise de variância. Os resultados destas análises foram similares as anteriores, dado que os testes não rejeitaram a hipótese de chegada Poisson para todos os átomos do sistema, mas rejeitaram a hipótese de atendimento exponencial para alguns servidores do sistema. Por meio do teste de variância verificamos que os servidores devem

ser considerados não-homogêneos na análise. A Tabela 7 apresenta a taxa de chegada total λ_j em cada átomo j (chamadas/hora), igual à $\sum_{m=1} \lambda_j^m$, onde λ_j^m corresponde a taxa de chegada de cada tipo de chamada $m = 1, 2a, 2b$ e 3 no átomo j . A tabela mostra também as taxas de atendimento μ_i (chamadas/hora) para ambulância i do sistema. Note que $\lambda = \sum_{j=1}^8 \lambda_j = 0,2251$ chamadas/h e $\mu = \sum_{i=1}^6 \mu_i = 6,3714$ chamadas/h.

Tabela 6 – Tipo de chamada em cada sub-átomo e lista de preferência de cada átomo.

Átomo e sub-átomos	Chamadas	Primeiro	Segundo	Terceiro
1a	1, 2a	2	3	-
1b	1, 2b, 3	1	2	3
2a	1, 2a	3	2	-
2b	2b, 3	1	3	2
3a	1, 2a	2	4	-
3b	2b, 3	1	2	4
4a	1, 2a	4	2	-
4b	2b	1	4	2
5	1, 2a	4	5	-
6a	1, 2a	5	4	-
6b	2b	4	5	6
7	1, 2a, 2b	5	6	4
8	1, 2a	6	5	-

Tabela 7 – Dados do processo de chegada e atendimento para a configuração 2 do SAE.

Átomo j	Taxa de chegada λ_j (ch/h)	Sub atôm	Tipo 1 $\lambda_j^{[1]}$ (ch/h)	Tipo 2a $\lambda_j^{[2a]}$ (ch/h)	Tipo 2b $\lambda_j^{[2b]}$ (ch/h)	Tipo 3 $\lambda_j^{[3]}$ (ch/h)	Serv i	Taxa atend. μ_i (ch/h)	Carga de trabalho ρ_i
1	0,0848	1a 1b	0,0328 0,0308	0,00090	0,01967	0,00060	1	1,2136	0.0442
2	0,0579	2a 2b	0,0528	0,00145	0,00252	0,00118	2	1,0196	0.0613
3	0,0062	3a 3b	0,0036	0,00019	0,00228	0,00017	3	1,0323	0.0561
4	0,0100	4a 4b	0,0091	0,00029	0,00062		4	0,9082	0.0229
5	0,0085		0,0084	0,00015			5	1,0733	0.0335
6	0,0194	6a 6b	0,0185	0,00073	0,00016		6	1,1244	0.0215
7	0,0153		0,0146	0,00059	0,00015				
8	0,0230		0,0213	0,00164	-				

As equações de equilíbrio foram formuladas como descrito na seção 3, e como na configuração 2 do SAE analisado há 6 servidores (1 veículo médico e 5 veículos resgates), então há $2^6 = 64$ estados possíveis do sistema. Analisando, por exemplo, o estado $B = \{11001\}$, com base na lista de preferência de despacho da Tabela 6, temos a seguinte equação de equilíbrio:

$$\begin{aligned} p_{111000}(\lambda - [\lambda_{1a}^{[1]} + \lambda_{1a}^{[2a]} + \lambda_{1b}^{[1]} + \lambda_{1b}^{[2b]} + \lambda_{1b}^{[3]} + \lambda_{2a}^{[1]} + \lambda_{2a}^{[2a]} + \lambda_{2b}^{[2b]} + \lambda_{2b}^{[3]}]) + \mu_1 + \mu_2 + \mu_3 = \\ p_{000000}(\lambda_{1b}^{[3]} + \lambda_{2b}^{[3]}) + p_{001000}(\lambda_{1b}^{[2b]} + \lambda_{1b}^{[3]} + \lambda_{2b}^{[2b]} + \lambda_{2b}^{[3]} + \lambda_{3b}^{[2b]}) + p_{010000}(\lambda_{1b}^{[2b]} + \lambda_{1b}^{[3]} + \lambda_{2b}^{[2b]} + \lambda_{2b}^{[3]}) + \\ p_{100000}(\lambda_{1a}^{[2a]} + \lambda_{1b}^{[2b]} + \lambda_{1b}^{[3]} + \lambda_{2a}^{[2a]} + \lambda_{2b}^{[2b]} + \lambda_{2b}^{[3]}) + p_{011000}(\lambda_{1b}^{[1]} + \lambda_{1b}^{[2b]} + \lambda_{1b}^{[3]} + \lambda_{2b}^{[2b]} + \lambda_{2b}^{[3]}) \\ + p_{101000}(\lambda_{1a}^{[1]} + \lambda_{1a}^{[2a]} + \lambda_{1b}^{[1]} + \lambda_{1b}^{[2b]} + \lambda_{1b}^{[3]} + \lambda_{2a}^{[1]} + \lambda_{2a}^{[2a]} + \lambda_{2b}^{[2b]} + \lambda_{2b}^{[3]} + \lambda_{3a}^{[1]}) \\ p_{110000}(\lambda_{1a}^{[1]} + \lambda_{1a}^{[2a]} + \lambda_{1b}^{[1]} + \lambda_{1b}^{[2b]} + \lambda_{1b}^{[3]} + \lambda_{2a}^{[1]} + \lambda_{2a}^{[2a]} + \lambda_{2b}^{[2b]} + \lambda_{2b}^{[3]}) + p_{111010}(\mu_5) + p_{111001}(\mu_6) \end{aligned}$$

Similarmente à configuração 1, os resultados obtidos para as probabilidades de estados do sistema mostraram que o mesmo tem probabilidade muito pequena de estar saturado (todos servidores ocupados), dado que $p_{111111} = 0,0000002$ e $p_{000000} = 0,7964$. Calculando a carga de trabalho ρ_i para cada ambulância i , os resultados obtidos são: $\rho_1 = 0,0454$, $\rho_2 = 0,0621$, $\rho_3 = 0,0576$, $\rho_4 = 0,0226$, $\rho_5 = 0,0336$ e $\rho_6 = 0,011$. Ao comparar estes resultados com a análise da amostra os desvios são pequenos (desvio médio de apenas 1,0%, conforme Tabela 7). A probabilidade de perda para uma chamada de qualquer tipo no sistema é $P_p = 0,6\%$ (como na configuração 1). Como discutido na seção 3, podemos definir diversas medidas de frequência de despacho de acordo com cada tipo de chamada no sistema. Além das frequências de duplo despacho para chamadas tipo 2a, 2b e 3 (se um dos três servidores estiver ocupado) e único despacho para chamadas tipo 2a (se um dos dois possíveis servidores estiver ocupado) e tipos 2b e 3 (se dois dos três servidores estiverem ocupados), outras medidas também podem ser calculadas como, por exemplo, as frequências de triplo despacho no sistema $f_{(i,k,l)j}^{[3]}$, onde i, k e l correspondem respectivamente, aos servidores 1, 2 e 3 da lista de preferência de despacho do átomo j , como descrito no capítulo 4. Considerando os despachos triplos do sistema em que $i = 1$, temos: $f_{(1,2,3)1b}^{[3]} = 0,2645$, $f_{(1,3,2)2b}^{[3]} = 0,5183$ e $f_{(1,2,4)3b}^{[3]} = 0,0796$.

Medidas adicionais de tempo de viagem comentadas na seção 3.3 também podem ser calculadas para descrever este sistema. A seguir, mostramos apenas os resultados para o tempo de viagem de cada ambulância i para chamadas do tipo 1, e comparamos os resultados com a análise da amostra (Tabela 8). Note que os desvios em relação aos resultados obtidos da amostra são suficientemente pequenos.

Tabela 8 – Tempo médio de viagem para cada servidor – único despacho $\overline{TU}_i^{[1]}$ (em minutos).

Servidor i	$\overline{TU}_i^{[1]}$ modelo	$\overline{TU}_i^{[1]}$ amostra	Desvio
1	2,674	2,674	0,0%
2	8,060	7,868	2,44%
3	7,729	7,562	2,21%
4	7,771	7,895	-1,57%
5	8,707	8,898	-2,15%
6	6,356	6,602	-3,73%
$\overline{T}^{[1]}$	7,033	7,053	-0,28%

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 às 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 na própria base, e servidores diferenciados (p.e, veículo médico e veículos resgates).

Três modelos foram propostos: no primeiro modelo, estendemos o modelo hipercubo para analisar os SAEs em rodovias com uma política envolvendo *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, considerando que os atendimentos realizados na sua base, por não envolverem tempo de viagem, possuem taxa de atendimento distinta da taxa de atendimentos a acidentes realizados ao longo da rodovia. No terceiro modelo, estendemos o primeiro modelo considerando que, em alguns SAEs em rodovias, os servidores são diferenciados e em cada região são geradas chamadas que requerem único ou múltiplo despacho de diferentes tipos de servidores.

Entre as principais medidas adicionais de desempenho que podem ser estimadas pelos modelos 1, 2 e 3 estão: probabilidade de perda para cada tipo de chamada, fração de único, duplo e triplo despacho para os diferentes tipos de chamada de cada ambulância a cada átomo do sistema e medidas agregadas de tempo de viagem diferenciando tipo de chamada e número de ambulâncias enviadas. 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 *et al.* (2005) e Iannoni & Morabito (2006b). Além disso, os métodos que integram o modelo hipercubo em procedimentos de otimização podem ser utilizados para avaliar o *trade-off* entre as medidas de desempenho conflitantes do sistema. Outra perspectiva deste estudo é a aplicação dos modelos e métodos propostos para análise de outros SAEs em rodovias, com características similares ao SAE analisado no presente estudo, em particular, sistemas com múltiplo despacho de ambulâncias idênticas e/ou diferenciadas e *backup* parcial.

Agradecimentos

Agradecemos aos revisores anônimos pelos úteis comentários e sugestões. Também agradecemos à concessionária *Centrovias* pela valiosa colaboração durante a realização do estudo de caso, e ao CNPq (processos 140178/01-5 e 522973/95-4) e à CAPES (processo BEX2468/02-6) pelo apoio financeiro.

Referências Bibliográficas

- (1) Batta, R.; Dolan, J.M. & Krishnamurthy, N.N. (1989). The maximal expected covering location problem: Revisited. *Transportation Science*, **23**, 277-287.
- (2) Brandeau, M. & Larson, R.C. (1986). Extending and applying the hypercube queuing model to deploy ambulances in Boston. In: *Delivery of Urban Services. TIMS Studies in the Management Science* [edited by A.J. Swersey and E.J. Ingball], Elsevier, **22**, 121-153.
- (3) Brotcorne, L.; Laporte, G. & Semet, F. (2003). Ambulance location and relocation models. *European Journal of Operational Research*, **147**, 451-63.
- (4) 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*, **20**(2), 113-119.
- (5) Chelst, K. & Barlach, Z. (1981). Multiple unit dispatches in emergency services: models to estimate system performance. *Management Science*, **27**(12), 1390-1409.
- (6) Chiyoshi, F.; Galvão, R.D. & Morabito, R. (2000). O uso do modelo hipercubo na solução de problemas de localização probabilísticos. *Gestão & Produção*, **7**(2), 146-174.
- (7) 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.
- (8) Chiyoshi, F.; Galvão, R.D. & Morabito, R. (2003). A note on solutions to the maximal expected covering location problem. *Computers and Operations Research*, **30**(1), 87-96.
- (9) Costa, D.M. (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.
- (10) 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*, **23**(1), 61-78.
- (11) Galvão, R.D.; Chiyoshi, F. & Morabito, R. (2005). Towards unified formulations and extensions of two classical probabilistic location models. *Computers and Operations Research*, **32**, 15-33.
- (12) Gendreau, M.; Laporte, G. & Semet, F. (2001). A dynamic model and parallel tabu search heuristic for real-time ambulance relocation. *Parallel Computing*, **27**, 1641-1653.
- (13) Halpern, J. (1977). Accuracy of estimates for the performance criteria in certain emergency service queuing systems. *Transportation Science*, **11**(3), 223-242.
- (14) Iannoni, A.P. (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.

- (15) Iannoni, A.P. & Morabito, R. (2002). Análise do sistema logístico de recepção de cana de açúcar: Um estudo de caso utilizando simulação discreta. *Gestão & Produção*, **9**(2), 107-128.
- (16) Iannoni, A.P. & Morabito, R. (2006a). A discrete simulation analysis of a logistics supply system. *Transportation Research*, E **42**(3), 191-210.
- (17) Iannoni, A.P. & Morabito, R. (2006b). Modelo hipercubo integrado a um algoritmo genético para análise de sistemas médicos emergenciais em rodovias. *Gestão & Produção*, **13**(1), 93-104.
- (18) Iannoni, A.P.; Morabito, R. & Saydam, C. (2005). Analyzing the configuration and operation of emergency medical systems on highways using the hypercube model. submetido para publicação no *Annals of Operations Research*.
- (19) Jarvis, J.P. (1985). Approximating the equilibrium behavior of multi-server loss systems. *Management Science*, **31**, 235-239.
- (20) Kelton, W.D.; Sadowski, R.P. & Sadowski, D.A. (2002). *Simulation with Arena*. 2.ed. McGrawHill, New York.
- (21) Larson, R.C. (1974). Hypercube queuing model for facility location and redistricting in urban emergency services. *Computers and Operations Research*, **1**, 67-95.
- (22) Larson, R.C. (1975). Approximating the performance of urban emergency service systems. *Operations Research*, **23**, 845-868.
- (23) Larson, R.C. (2004). OR models for homeland security. *OR/MS Today*, **31**, 22-29.
- (24) Larson, R.C. & Odoni, A.R. (1981). *Urban Operations Research*. Prentice Hall, New Jersey.
- (25) Larson, R. & Mcknew, M.A. (1982). Police patrol-initiated activities within a systems queuing model. *Management Science*, **28**(7), 759-774.
- (26) Law, A.M. & Kelton, W.D. (1991). *Simulation Modeling and Analysis*. 2.ed. McGraw-Hill, New York.
- (27) 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.
- (28) 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.
- (29) Owen, S.H. & Daskin, M.S. (1998). Strategic facility location: A review. *European Journal of Operational Research*, **111**, 423-447.
- (30) Pegden, C.D.; Shannon, R.E. & Sadowski, R.P. (1995). *Introduction to Simulation Using SIMAN*. 2.ed. McGraw-Hill, New York.
- (31) Sacks, S.R. & Grief, S. (1994). Orlando Police Department uses OR/MS methodology, new software to design patrol districts. *OR/MS Today*, Baltimore, 30-32.
- (32) Saydam, C.; Repede, J. & Burwell, T. (1994). Accurate estimation of expected coverage: a comparative study. *Socio-Economic Planning Sciences*, **28**(2), 113-120.

- (33) Saydam, C. & Aytug, H. (2003). Accurate estimation of expected coverage: revisited. *Socio-Economic Planning Sciences*, **37**, 69-80.
- (34) Swersey, A.J. (1994). *Handbooks in OR/MS*. Elsevier Science B.V., Amsterdam, **6**, 151-200.
- (35) 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*, **24**(1), 39-72.
- (36) Takeda, R.A.; Widmer, J.A. & Morabito, R. (2007). Analysis of ambulance decentralization in an urban medical emergency service using the hypercube queueing model. *Computers and Operations Research*, **34**(3), 727-741.