



Posicionamento de ambulâncias do SAMU através de Programação Inteira e Teoria de Filas

Positioning of ambulances of the SAMU system by Integer Programming and Queueing Theory

Bruno Barreto¹
Fernando Alexandrino¹
Ormeu Coelho¹

Resumo: A configuração de redes logísticas para serviços de emergência é questão estratégica de imensa importância, visto que pequenas variações no tempo de resposta podem implicar na morte do solicitante. Partindo dessa premissa, o trabalho propõe novas alternativas de posicionamento para as ambulâncias do sistema SAMU na cidade de Duque de Caxias, RJ, capazes de reduzir o tempo de resposta do serviço. Essas propostas de reposicionamento das ambulâncias foram construídas em duas etapas: na primeira, dois modelos de Programação Inteira foram aplicados para se obter soluções que provejam maior cobertura à população. Posteriormente, o Modelo do Hipercubo foi empregado para avaliar a disponibilidade dos servidores, dentre outros indicadores de desempenho relevantes, como o tempo médio de resposta e a taxa de ocupação das ambulâncias.

Palavras-chave: Localização de facilidades; Modelo do Hipercubo; Serviços de emergência.

Abstract: *The configuration of emergency services logistic networks for is a paramount strategic issue since small deviations may lead to death of users. From this premise, the work proposes new alternatives for positioning the SAMU ambulance system in the city of Duque de Caxias / RJ, which are able to reduce the service response time. These ambulances repositioning proposals were built in two stages: at first two models of Integer Programming were used in order to obtaining solutions that maximizes coverage. Then, the Hypercube Model was applied for evaluating the server's availability under randomness, as well as other relevant performance indicators, such as average time response, and servers' workloads.*

Keywords: *Facility location; Hypercube Model; Emergency services.*

1 Introdução

Ao projetar uma rede logística é imprescindível atentar para as particularidades inerentes à operação nos setores privado e público. Enquanto no primeiro se busca uma rede na qual os produtos fluam entre pontos de suprimento e demanda, baseada nos custos/lucros correspondentes, no setor público a questão central é otimizar alguma função que meça a disponibilidade do serviço para uma dada população (Ghiani et al., 2004).

No âmbito da administração pública, os três objetivos principais a serem levados em consideração no projeto da rede logística são: redução de capital; redução de custos; e aumento do nível do serviço (Ghiani et al., 2004). Particularmente, na área de saúde, um dos problemas logísticos de maior interesse é o posicionamento de ambulâncias que fazem os atendimentos de emergência. A localização dessas facilidades é muito sensível ao nível de serviço

requerido, caracterizado principalmente pelo tempo de resposta do atendimento. Uma cobertura de má qualidade pode implicar na morte do usuário.

O tempo de resposta ou, como será tratado posteriormente, limite de cobertura definido pela legislação norte-americana é de no máximo 10 minutos para áreas urbanas com nível de serviço de 95%, podendo ser estendido para 30 minutos para áreas rurais (Ball & Lin, 1993). Em Londres, 95% das solicitações devem ser atendidas em até 14 minutos (Galvão et al., 2003a). Em Montreal, o tempo máximo de resposta deve ser menor ou igual a 10 minutos para 70% das chamadas (Gendreau et al., 2001). Porém, de acordo com Takeda et al. (2004), no Brasil não há legislação que determine um limite superior para o tempo de resposta desse tipo de serviço.

No presente trabalho analisa-se o reposicionamento das ambulâncias do Serviço de Atendimento Móvel

¹ Departamento de Engenharia de Produção, Centro Federal de Educação Tecnológica Celso Suckow da Fonseca – CEFET/RJ, CEP 20271-110, Rio de Janeiro, RJ, Brasil, e-mail: brunobarreto7@gmail.com; fernando.hnd@gmail.com; ormeucoelho@gmail.com

Recebido em Mar. 1, 2015 - Aceito em Jun. 12, 2015

Suporte financeiro: Nenhum.

de Urgência (SAMU) em Duque de Caxias, RJ, pela combinação de técnicas de Programação Inteira (PI) e Teoria de Filas. Embora haja vasto uso de modelos determinísticos de otimização em problemas dessa natureza, como, por exemplo, em Schmid & Doerner (2010), Iannoni et al. (2009) e Gendreau et al. (1997), o presente trabalho se caracteriza pelo emprego de um modelo de filas com o objetivo de avaliar as soluções geradas por modelos de PI utilizados para localizar as ambulâncias do SAMU. O uso do Modelo do Hipercubo (MH) em sistemas congestionados permite representar aspectos que fogem às formulações determinísticas.

Espera-se com este estudo avaliar a possibilidade de expandir o número de usuários atendidos dentro de um limite de tempo aceitável apenas reposicionando os servidores em locais do município que possuam infraestrutura mínima. Essa ideia é embasada pelo conceito de Base Descentralizada apresentado na portaria n. 2.657 do Ministério da Saúde, de 16 de dezembro de 2004, a qual permite a existência de bases operacionais descentralizadas que funcionem como postos avançados para ambulâncias e respectivas equipes, para garantir um tempo de resposta de qualidade aos usuários do SAMU (Brasil, 2013).

O trabalho foi dividido em cinco etapas. A seção 2 traz um breve levantamento bibliográfico acerca de modelos de localização de facilidades para cobertura de serviços emergenciais. O desempenho atual do sistema foi analisado através de indicadores de cobertura, conforme mostrado na seção 3. A seção 4 discute a aplicação de dois modelos PI pesquisados aos dados do SAMU Duque de Caxias, RJ, e os resultados obtidos são apresentados na seção 5. Ainda nessa seção, com o intuito de chegar a arranjo logístico que traga a maior extensão da cobertura do serviço, realiza-se uma análise por meio do MH, a fim de avaliar parte das melhores soluções obtidas por um pacote computacional para PI. Conclusões e perspectivas para extensão da pesquisa são apresentadas na seção 6.

2 Revisão bibliográfica

O Problema da Cobertura de Conjuntos foi um dos primeiros modelos discretos de localização de facilidades usados em serviços de emergência (Toregas et al., 1971). Considera-se uma restrição de cobertura, expressa pelo máximo tempo de viagem ou pela distância entre facilidade-servidor e cliente em uma rede logística. Tal medida de separação é muitas vezes chamada de distância crítica, S . O problema é definido em uma rede na qual I é o conjunto dos nós de demanda e J , o conjunto dos pontos candidatos à alocação dos servidores. O nó i ($i \in I$) é considerado coberto se e somente se a distância (em unidades de comprimento/tempo) entre um cliente no nó i e a

ambulância mais próxima, localizada em j ($j \in J$), for menor ou igual a S .

Diferentemente do PCC, o Problema de Localização de Máxima Cobertura proposto por Church & Revelle (1974) procura maximizar a população coberta por determinado serviço dentro da distância crítica S , com um número predefinido de facilidades p . Nesse modelo, o número de instalações-servidores é determinado exogenamente pela existência de orçamento limitado ou restrições gerenciais.

O modelo TEAM, do inglês Tandem Equipment Allocation Model, assume a existência de dois servidores distintos, cada qual com seu respectivo limite de cobertura. Essa premissa é muito comum nos serviços de emergência nos quais existem servidores com equipamentos distintos, capazes de atender diferentes ocorrências (Schilling et al., 1979).

Em situações nas quais o serviço prestado pelas unidades básicas possa também ser provido pelas avançadas, como é o caso dos serviços de emergência, um nível maior de cobertura pode ser alcançado ao se permitir que os servidores sejam posicionados independentemente. Tal adaptação foi apresentada por Schilling et al. (1979) no modelo conhecido como FLEET, Facility-location, Equipment-emplacement Technique. Esse modelo exige que cada nó do cliente seja simultaneamente abrangido por servidores primários (ou básicos) e especiais (ou avançados). Os conjuntos $N_j^p = \{j \in J | d_{ji} \leq S^p\}$ e $N_j^s = \{j \in J | d_{ji} \leq S^s\}$ correspondem aos nós j nos quais a alocação de um servidor, básico e avançado, nessa ordem, permite a cobertura do nó i . d_{ji} é a distância entre j e i , através de algum caminho mínimo na rede, e S^p e S^s são as distâncias críticas para garantir a cobertura por um servidor básico ou avançado, respectivamente. A formulação também considera os parâmetros e variáveis abaixo:

- a_i = população do nó i ;
- p^p = número de facilidades básicas;
- p^s = número de facilidades avançadas;
- p^z = número de novas facilidades a serem criadas.

As variáveis são:

- $x_j^p = \begin{cases} 1, & \text{se uma ambulância básica for alocada em } j; \\ 0, & \text{caso contrário.} \end{cases}$
- $x_j^s = \begin{cases} 1, & \text{se uma ambulância avançada for alocada em } j; \\ 0, & \text{caso contrário.} \end{cases}$
- $z_j = \begin{cases} 1, & \text{se uma facilidade for localizada em } j; \\ 0, & \text{caso contrário.} \end{cases}$
- $y_i = \begin{cases} 1, & \text{se o nó } i \text{ for coberto;} \\ 0, & \text{caso contrário.} \end{cases}$

Sua formulação matemática fica assim definida nas Equações 1-11:

$$\begin{aligned} \max Z &= \sum_{i \in I} a_i y_i & (1) \\ \text{s.a.:} & \\ \sum_{j \in N_i^p} x_j^p &\geq y_i, \quad \forall i \in I & (2) \\ \sum_{j \in N_i^s} x_j^s &\geq y_i, \quad \forall i \in I & (3) \\ \sum_{j \in J} x_j^p &= P^p, & (4) \\ \sum_{j \in J} x_j^s &= P^s, & (5) \\ \sum_{j \in J_N} z_j &= P^z, & (6) \\ x_j^p &\leq z_j, \quad \forall j \in J_N & (7) \\ x_j^s &\leq z_j, \quad \forall j \in J_N & (8) \\ y_i &\in \{0, 1\}, \quad \forall i \in I & (9) \\ z_j &\in \{0, 1\}, \quad \forall j \in J_N & (10) \\ x_j^p, x_j^s &\in \{0, 1\}, \quad \forall j \in J & (11) \end{aligned}$$

A Função Objetivo 1 busca maximizar a população coberta, enquanto as Restrições 2 e 3 computam a cobertura do nó i apenas quando ele for coberto por ao menos uma ambulância básica e uma avançada, respectivamente. As Expressões 4 e 5 definem as quantidades de servidores básicos e avançados. O conjunto de nós disponíveis que podem receber uma facilidade é J_N , $J_N \subset J$, e exatamente P^z facilidades devem ser instaladas, como expresso em 6. As Restrições 7 e 8 garantem que uma ambulância só seja instalada em nós em que há uma facilidade instalada.

Uma das primeiras abordagens estocásticas para a localização de facilidades-servidores de atendimento de emergência foi o Problema de Localização de Máxima Disponibilidade (PLMD) proposto por Revelle & Hogan (1989). O PLMD modela a incerteza inerente às demandas através de hipóteses simplificadoras. Deve-se localizar P servidores para maximizar a população coberta dentro de S , com confiabilidade θ (Galvão et al., 2003b). Assume-se que todos os servidores trabalham com igual taxa de ocupação, ρ , formalmente definida em 12.

$$\rho = \frac{\bar{t} \sum_{i \in I} \lambda_i}{24 \sum_{j \in J} y_j} = \frac{\bar{t} \sum_{i \in I} \lambda_i}{24P} \quad (12)$$

tal que:

- λ_i = taxa de chamadas no vértice $i \in I$;
- \bar{t} = duração média para atendimento das chamadas (em horas);
- P = número de servidores.

O único grupo de variáveis de decisão é y_j , tal que:

$$y_j = \begin{cases} 1, & \text{se um servidor foi localizado em } j; \\ 0, & \text{caso contrário.} \end{cases}$$

O número mínimo de servidores, b , necessário para cobrir um determinado nó ao nível de confiança θ , pode ser obtido de 12. Isso pode ser feito calculando-se a probabilidade de se ter ao menos uma ambulância disponível para atender a uma ocorrência dentro da distância crítica S , de acordo com a taxa de chegada dos chamados (Revelle & Hogan, 1989). A Expressão 13 deriva tal probabilidade, sendo c_{ji} os coeficientes da matriz binária, cujo valor é 1 se $d_{ji} \leq S$, e 0 caso contrário.

$$\begin{aligned} P(\text{pelo menos um servidor está a uma distância crítica } S) &\geq \theta \\ &= [1 - P(\text{nenhum servidor a uma distância crítica } S)] \geq \theta & (13) \\ &= 1 - \rho^{\sum_{j \in J} c_{ji} y_j} \geq \theta \end{aligned}$$

O somatório $\sum_{j \in J} c_{ji} y_j$ define o número de servidores disponíveis a uma distância máxima S de um determinado nó de demanda $i \in I$. Para que um vértice seja coberto com confiabilidade θ , deve-se ter ao menos b servidores capazes de atender um chamado originado nele. Calculando-se os logaritmos dos dois membros em 13 tem-se que $\sum_{j \in J} c_{ji} y_j \geq b$, em que $b = \frac{\log(1-\theta)}{\log \rho}$. Ou seja, dada a confiabilidade esperada e a taxa de ocupação ρ , tem-se o número de facilidades necessárias.

As variáveis no PLMD são:

$$\begin{aligned} y_{ik} &= \begin{cases} 1, & \text{se a demanda em } i \text{ é coberta por no mínimo } k \text{ servidores;} \\ 0, & \text{caso contrário.} \end{cases} \\ x_j &= \begin{cases} 1, & \text{se uma ambulância for alocada em } j; \\ 0, & \text{caso contrário.} \end{cases} \end{aligned}$$

A formulação matemática do PLMD é definida nas Expressões 14-18:

$$\max Z = \sum_{i \in I} \lambda_i y_{ib} \quad (14)$$

s.a.:

$$\sum_{k=1}^b y_{ik} \leq \sum_{j \in J} c_{ji} x_j, \quad \forall i \in I \quad (15)$$

$$y_{ik} \leq y_{i(k-1)}, \quad \forall i \in I, k = 2, \dots, b \quad (16)$$

$$\sum_{j \in J} x_j = P \quad (17)$$

$$x_j, y_{ik} \in \{0, 1\} \quad \forall i \in I, \forall j \in J, k = 2, \dots, b \quad (18)$$

A função objetivo 14 maximiza a população coberta, dentro da distância crítica S , com confiabilidade θ , isso é, apenas os nós cobertos por $k = b$ ambulâncias. O lado esquerdo, em 15, conta o número de servidores a menos da distância crítica S do nó de demanda i ,

assegurando a cobertura quando existem b servidores. Ademais, se k ambulâncias cobrem o vértice i , então é verdade que i também é coberto por $k - 1$ servidores, como expresso pela Restrição 16. Finalmente, a Restrição 17 define o número de ambulâncias.

Embora o PLMD considere a probabilidade de o servidor estar disponível na ocorrência de um chamado, não se pode garantir que a incerteza associada ao processo de chegadas tenha sido bem modelada, o que pode afetar significativamente a chance de atendimento dos chamados, ocasionando a formação de filas de espera e o aumento do tempo de atendimento. Além disso, a premissa de que os servidores possuam a mesma taxa de ocupação é pouco provável no problema real. Batta et al. (1989) também apoiam essa afirmação, apontando causas como a distribuição desproporcional de demanda em toda a região servida e a política de despacho, as quais podem priorizar determinados servidores, desequilibrando assim a fração de tempo em que eles estão ocupados.

Apesar da existência de modelos que consideram taxas de ocupação específica para cada servidor, é difícil inferir seus valores, uma vez que eles são produzidos a partir do posicionamento calculado pelo modelo de localização. Brotcorne et al. (2003) sugerem o uso de simulação ou Teoria de Filas para obtê-los. Brotcorne et al. (2003) apontam que, embora existam modelos que considerem taxas de ocupação específicas para cada servidor, essas são difíceis de se conhecer a priori, já que são *output* do próprio posicionamento efetuado pelo modelo de localização, e sugerem técnicas como simulação ou modelos de filas para obtê-las.

Portanto, é interessante avaliar solução obtida por um dos modelos discutidos anteriormente segundo indicadores influenciáveis pela incerteza (tempo

médio de atendimento, número médio de clientes em espera, dentre outros). No presente trabalho, isso é feito através do Modelo do Hipercubo, proposto por Larson (1974) e comumente usado para modelar sistemas nos quais os servidores se deslocam até os clientes para prestar um serviço, sendo as demandas geograficamente distribuídas em átomos discretos.

A região a ser estudada é dividida em I átomos geográficos e a chegada de chamados oriundos do átomo i é um processo Markoviano com taxa λ_i . Para atendê-las, o sistema conta com N servidores distribuídos ao longo dos átomos, cujos tempos de atendimento são exponencialmente distribuídos, com taxa de atendimento μ_n (Chiyoshi et al., 2000). Em cada instante, um servidor pode estar livre (0) ou ocupado (1), e a combinação dos estados de todos os servidores resulta em um estado do sistema. Por exemplo, para um sistema com três servidores, o estado {001} indica que o servidor #1 está ocupado, enquanto os servidores #2 e #3 estão livres. Logo, o número de possíveis estados é 2^N .

Considera-se que apenas um servidor é despachado para uma dada chamada, e que há uma ordem de prioridade para atendimento de uma chamada originada no átomo i . Se o servidor de maior prioridade estiver ocupado, o segundo servidor é despachado, e assim sucessivamente até o último. No estado {11...1} todos os servidores estão ocupados, e qualquer nova chamada deve esperar em uma fila, seguindo-se disciplina do tipo FCFS (First Come, First Served). Assim, além dos 2^N estados acima referidos há ainda aqueles nos quais existem l usuários no sistema, tais que $l \geq N + 1: \{S_{N+1}\}, \{S_{N+2}\}, \{S_{N+3}\}, \dots$. Os estados do MH podem ser representados pelos vértices de um hipercubo unitário, de onde deriva o nome do modelo (Chiyoshi et al., 2000). A Figura 1 exemplifica um sistema com $N = 3$ e fila de capacidade κ .

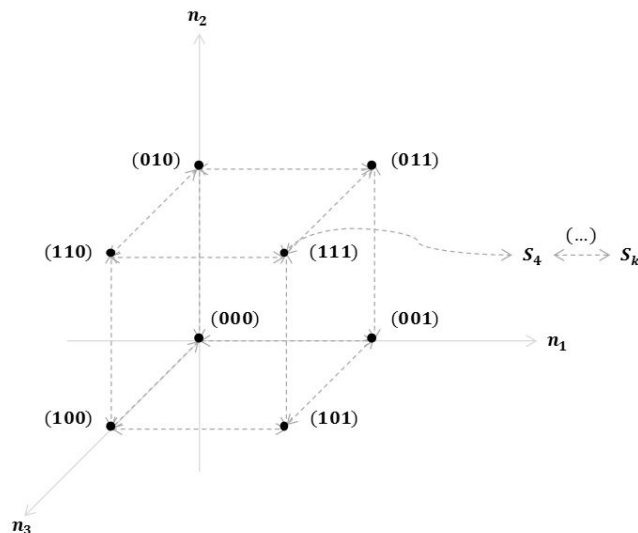


Figura 1. Estados de um sistema com três servidores e fila finita. Fonte: Elaborado pelos autores.

Larson (1974) define duas classes de transição em um Hipercubo: *upward*, na qual o servidor passa de livre para ocupado, e *downward*, quando o servidor muda do estado ocupado para o estado livre. Admitindo-se que as transições ocorram apenas entre os vértices adjacentes do hipercubo e que o fluxo com que o sistema entra em um determinado estado é igual ao fluxo com que ele sai desse estado podem ser construídas as equações de equilíbrio do modelo. Tome-se como exemplo o caso com $N = 3$ servidores e $I = 3$ átomos, assumindo p_B como a probabilidade de o sistema estar no estado B . A equação de equilíbrio em torno de $B = \{000\}$ é então expressa na Equação 19, na qual λ é a taxa total de chamados do sistema:

$$\lambda p_{000} = \mu_1 p_{001} + \mu_2 p_{010} + \mu_3 p_{100} \tag{19}$$

Na Equação 19, $\lambda = \lambda_1 + \lambda_2 + \lambda_3$ já que a partir do estado $\{000\}$ é possível chegar aos estados $\{001\}$, $\{010\}$ ou $\{100\}$ pelo recebimento de uma chamada originária do átomo #1, #2 ou #3, respectivamente. O lado direito da equação indica a possibilidade de se alcançar o estado $\{000\}$ a partir da conclusão dos chamados dos estados $\{001\}$, $\{010\}$ ou $\{100\}$, que ocorrem, respectivamente, a taxas μ_i , $i = 1, 2, 3$. Considerando-se que o servidor n está localizado no átomo $i = n$ e é o preferencial para atendê-lo, a equação do estado $\{001\}$ é construída de maneira análoga:

$$(\lambda + \mu_1) p_{001} = \lambda_1 p_{000} + \mu_2 p_{011} + \mu_3 p_{101} \tag{20}$$

Já o estado $\{111\}$ pode ser obtido através do recebimento de um chamado nos estados $\{011\}$, $\{101\}$ ou $\{110\}$, já que, independentemente da ordem de preferência, só há um servidor disponível para despacho, mas também através da conclusão de qualquer atendimento, quando o sistema se encontra no estado S_4 , no qual há três usuários sendo atendidos e um em espera. Assim que um servidor concluir o atendimento, ele tornar-se-á disponível e será despachado para atender ao chamado que estava em fila, levando o sistema novamente ao estado $\{111\}$. Dessa forma, a equação para o estado $\{111\}$ se encontra em 21:

$$(\lambda + \mu) p_{111} = \lambda p_{011} + \lambda p_{101} + \lambda p_{110} + \mu p_4 \tag{21}$$

Face à condição de equilíbrio do sistema, as taxas de transição entre os estados $\{111\}$ e S_4 devem ser iguais, ou seja, $\lambda p_{111} = \mu p_4$ (Chiyoshi et al., 2000). De forma semelhante, as transições entre os estados $\{S_\kappa\}$ e $\{S_{\kappa+1}\}$, para $\kappa \geq N$, são iguais e equivalentes a $\rho = \frac{\lambda}{\mu}$. As equações de transição nos estados onde todos os servidores estão ocupados formam uma progressão geométrica, e, logo, $p_{111} + p_4 + p_5 + \dots = p_{111} / (1 - \rho)$. Como a soma das probabilidades de todos os estados do sistema é igual a 1, é possível obter-se a equação de normalização 21, resultando em um sistema compatível e determinado de 2^N equações.

$$p_{000} + p_{001} + p_{010} + \dots + p_{111} / (1 - \rho) = 1 \tag{22}$$

A partir da resolução desse sistema, diversas métricas de desempenho podem ser computadas, como o tempo médio de resposta, a taxa de ocupação dos servidores e a probabilidade de formação de filas, permitindo analisar como o posicionamento dos servidores reage quando submetido a demandas estocásticas.

Dentre os estudos que empregam o Modelo do Hipercubo é importante destacar os trabalhos de Larson (1975), Brandeau & Larson (1986), Galvão et al. (2003b), Takeda et al. (2004) e Souza et al. (2013).

3 Caracterização da pesquisa de campo

Duque de Caxias é um município da região metropolitana do estado do Rio de Janeiro que, de acordo com o censo demográfico do IBGE, tinha 855.048 habitantes em 2010 (IBGE, 2014). Para atender essa população, o SAMU dispõe de sete unidades de suporte básico (USB) e duas unidades de suporte avançado (USA). Assim como acontece em muitos municípios brasileiros, sua Secretaria de Saúde não possui qualquer ferramenta computacional para posicionar as ambulâncias, o que então é feito de maneira empírica, na maioria das vezes. A coordenação operacional do SAMU está localizada no Hospital Municipal Doutor Moacir Rodrigues do Carmo, onde duas unidades avançadas foram posicionadas. Na Tabela 1 se observa a atual distribuição das ambulâncias no município.

Tabela 1. Distribuição das ambulâncias do SAMU vigente no momento de realização da pesquisa.

Base de localização	Endereço	USA	USB
Hospital Municipal Dr. Moacir R. do Carmo	Rod. Washington Luiz, 3200 – Beira Mar	2	2
Posto Médico Sanitário de Campos Eliseos	Av. Actura, 333 – Campos Eliseos	0	1
Posto Médico Sanitário Parque Equitativa	Av. Automóvel Clube, s/n – Parque Equitativa	0	1
Posto Médico Sanitário do Pilar	R. Carlos Alves, s/n – Pilar	0	1
Posto Médico Sanitário Saracuruna	Av. Presidente Roosevelt, s/n – Saracuruna	0	1
Posto Médico Sanitário de Xerém	R. Nóbrega Ribeiro, s/n – Xerém	0	1

Fonte: Elaborado pelos autores.

3.1 Construção da rede do problema

Inicialmente, todos os hospitais públicos, postos de saúde e UPAs (unidades de pronto atendimento) do município foram considerados locais candidatos para a instalação de uma facilidade, resultando no conjunto J , definido nas Formulações 1-11 e 14-18. Ademais, a área geográfica do município foi dividida em sub-regiões tão pequenas quanto a disponibilidade de dados permitiu. Em cada uma delas, escolheu-se arbitrariamente um “centroide”, assumindo-se que toda a demanda pelo serviço está nele concentrada. Tem-se assim o conjunto I dos nós de demanda, construído de tal maneira que $J \subset I$. A Tabela 2 descreve os vértices em $V=I$ considerados, na qual os nós sombreados são aqueles pertencentes ao conjunto J , enquanto os demais pertencem unicamente ao conjunto I . A localização de cada átomo geográfico, em termos de latitude (LAT) e longitude (LONG), pode ser consultada na Tabela 3, assim como a respectiva população a_i , obtida a partir do Censo Demográfico 2010 do IBGE (IBGE, 2014).

Por fim, uma rede direcionada foi construída, na qual o conjunto A dos arcos representa as possibilidades de movimentação entre pares de vértices (através de ruas, avenidas ou vielas). Esses arcos são valorados com o tempos de viagem entre pares de vértices em $V, V=I$. Assume-se que, ao realizar um atendimento,

a ambulância se desloca pela rota “mais curta” até o vértice de onde ocorreu o chamado. Os tempo de viagem entre uma facilidade e os demais vértices foram estimados pelos valores em horário de pico, entre 17h e 19h, através de consulta à API (Application Programming Interface) do GoogleMaps. Na Tabela 4 estão dispostos os tempos de viagem, em minutos, na qual os valores sombreados são aqueles para os quais $c_{ji} = 1$, enquanto que, para os demais, $c_{ji} = 0$.

3.2 Análise da situação atual

De acordo com Bertelli et al. (1999), a maior frequência de sobrevida em vítimas de parada cardíaca ocorre quando as manobras de reanimação são realizadas em até 8 minutos. Esse parâmetro foi utilizado como limite do tempo de resposta dos servidores do tipo avançado, ou seja, $s^s = 8$ minutos. Para as ambulâncias básicas, adotou-se $s^p = 12$ minutos. Considerando-se o posicionamento das ambulâncias e os tempos de viagem estimados, empregou-se o MH para avaliar os parâmetros de desempenho no arranjo logístico atual.

Segundo a coordenação operacional do SAMU de Duque de Caxias, foram registradas 17.862 ocorrências entre janeiro e junho de 2013. Pelas hipóteses adotadas no MH, o número de chegadas no intervalo t segue

Tabela 2. Descrição dos nós do problema.

Descrição	Nó	Descrição	Nó
Pam 404 Doutor Fernando Gil	01	Duque de Caxias, CEP 25250-400	25
Posto de Saúde Alaide Cunha	02	Duque de Caxias, CEP 25271-350	26
Duque de Caxias, CEP 25235-460	03	Duque de Caxias, CEP 25036-600	27
Duque de Caxias, CEP 25015-415	04	Duque de Caxias, CEP 25272-410	28
Duque de Caxias, CEP 25267-390	05	Duque de Caxias, CEP 25265-232	29
Posto Médico Sanitário de Campos Elíseos	06	Hospital Municipal Dr. Moacir R. do Carmo	30
Duque de Caxias, CEP 25220-570	07	Duque de Caxias, CEP 25240-650	31
Duque de Caxias, CEP 25245-230	08	Duque de Caxias, CEP 25046-380	32
Centro Municipal de Saúde de Duque de Caxias	09	Posto de Saúde Sarapuí	33
Hospital Infantil Ismélia Silveira	10	UPA Sarapuí	34
UPA Infantil Walter Garcia	11	Duque de Caxias, CEP 25025-300	35
UPA Duque de Caxias	12	Posto Médico Sanitário do Pilar	36
Duque de Caxias, CEP 25251-100	13	Posto Médico Sanitário Santa Cruz da Serra	37
Duque de Caxias, CEP 25243-150	14	Duque de Caxias, CEP 25271-430	38
Duque de Caxias, CEP 25237-030	15	Duque de Caxias, CEP 25040-060	39
Posto Médico Sanitário Parque Equitativa	16	Duque de Caxias, CEP 25045-040	40
Duque de Caxias, CEP 25060-190	17	Posto Médico Sanitário Saracuruna	41
Duque de Caxias, CEP 25231-180	18	Duque de Caxias, CEP 25270-450	42
Posto Médico Sanitário Dr. Jorge R. Pereira	19	Duque de Caxias, CEP 25030-180	43
Posto de Saúde Doutor José de Freitas	20	Duque de Caxias, CEP 25040-610	44
Posto de Saúde Edna Salles	21	Duque de Caxias, CEP 25065-162	45
Posto de Saúde José Camilo dos Santos	22	Hospital Municipal Maternidade de Xerém	46
Hospital Estadual Adão Pereira Nunes	23	Unidade Pré-Hospitalar Álvaro Figueira	47
Duque de Caxias, CEP 25250-130	24	Posto Médico Sanitário de Xerém	48

Fonte: Elaborado pelos autores.

Tabela 3. Localização e distribuição populacional.

Nó	Bairro	LAT	LONG	a_i	Nó	Bairro	LAT	LONG	a_i
01	25 de Agosto	-22,793	-43,299	7.071	25	Mantiquira	-22,596	-43,302	10.616
02	25 de Agosto	-22,786	-43,297	7.071	26	Meio da Serra	-22,626	-43,206	2.344
03	Amapá	-22,676	-43,357	6.477	27	Olavo Bilac	-22,766	-43,328	34.770
04	Bar dos Cavaleiros	-22,795	-43,326	41.209	28	Parada Angélica	-22,629	-43,210	14.458
05	Barro Branco	-22,638	-43,244	15.700	29	Parada Morabi	-22,657	-43,230	4.444
06	Campos Elíseos	-22,660	-43,250	19.622	30	Parque Duque	-22,799	-43,289	44.983
07	Cangulo	-22,688	-43,236	13.053	31	Parque Eldorado	-22,635	-43,307	8.161
08	Capivari	-22,647	-43,328	1.489	32	Parque Fluminense	-22,725	-43,319	34.969
09	Centro	-22,787	-43,308	6.756	33	Parque Sarapuí	-22,751	-43,296	1.009
10	Centro	-22,788	-43,311	6.756	34	Parque Sarapuí	-22,751	-43,299	1.009
11	Centro	-22,793	-43,307	6.756	35	Periquitos	-22,779	-43,324	17.898
12	Centro	-22,786	-43,325	6.756	36	Pilar	-22,711	-43,306	33.525
13	Chácaras Arcampo	-22,657	-43,274	14.120	37	Santa Cruz da Serra	-22,645	-43,274	25.698
14	Chácaras Rio-Petrópolis	-22,665	-43,315	14.085	38	Santa Lúcia	-22,625	-43,210	16.732
15	Cidade dos Meninos	-22,630	-43,222	2.460	39	Santo Antônio	-22,745	-43,317	11.420
16	Cidade Parque Paulista	-22,635	-43,263	33.501	40	São Bento	-22,728	-43,305	22.062
17	Doutor Laureano	-22,764	-43,299	43.996	41	Saracuruna	-22,676	-43,254	46.660
18	Figueira	-22,680	-43,298	16.520	42	Taquara	-22,627	-43,236	12.191
19	Imbariê	-22,636	-43,217	34.332	43	Vila Centenário	-22,774	-43,313	21.922
20	Jardim Anhangá	-22,637	-43,231	12.867	44	Vila São José	-22,742	-43,317	31.009
21	Jardim Gramacho	-22,761	-43,278	53.731	45	Vila São Luís	-22,773	-43,298	30.420
22	Jardim Primavera	-22,695	-43,261	20.915	46	Xerém	-22,599	-43,302	7.466
23	Jardim Primavera	-22,670	-43,279	20.915	47	Xerém	-22,600	-43,292	7.466
24	Lamarão	-22,598	-43,293	192	48	Xerém	-22,601	-43,292	7.466

Fonte: Elaborado pelos autores.

distribuição de Poisson, com média λ se e somente se o tempo entre chegadas é exponencialmente distribuído com média $1/\lambda$ (Taha, 2008). Para verificar a adequação a essa premissa foram tomadas as médias dos intervalos entre chamadas sucessivas ao longo de 21 dias, obtidas a partir de base de dados fornecida pelo SAMU, mostrada na Tabela 5.

Através do teste Kolmogorov-Smirnov para a média com nível de significância $\alpha = 0,05$, obteve-se um p -valor de 0,9, indicando que não se pode rejeitar a hipótese nula de que o intervalo entre chegadas do sistema segue distribuição exponencial negativa. Consequentemente, o número de chamados no intervalo t ocorre segundo um processo de Poisson, com média $\hat{\lambda} = 1/2,39 = 0,42$ chamados/hora. Frente à indisponibilidade de dados desagregados por átomos geográficos, a taxa de chamados por hora para cada átomo (λ_i) foi estimada como em Takeda et al. (2004). Esses autores sugerem aproximar λ_i pelo produto entre p_i (probabilidade de uma ocorrência ter origem no átomo i , isto é, o percentual relativo de chamadas deste átomo) e $\hat{\lambda}$ (a taxa total de chamadas do sistema). A Tabela 6 apresenta as estimativas obtidas.

Os tempos totais de atendimento são definidos como a soma dos tempos de preparo do veículo, viagem ao local da ocorrência, atendimento à vítima e retorno. A média ao longo dos 21 dias analisados

e os desvios-padrão são expostos em minutos na Tabela 7 e distintos por servidor, dos quais os dois primeiros são de suporte avançado. De acordo com Takeda et al. (2004), quando os desvios possuem a mesma ordem de grandeza que as médias, como se verificou neste estudo, pode-se inferir que a distribuição é aproximadamente exponencial. Tal hipótese foi confirmada pelo teste de Kolmogorov-Smirnov, com $\alpha = 0,05$, a partir do qual se obteve p -valor superior a 0,7. Sendo assim, adotou-se a taxa de atendimento $\mu^s = 60/77 = 0,78$ para as ambulâncias do tipo USA, em chamados/hora, e $\mu^p = 60/75 = 0,80$ para as do tipo USB. Finalmente, a matriz de preferência de despachos foi gerada, tomando-se para cada átomo i os servidores mais próximos (em termos do tempo de viagem), em ordem crescente, sem distinção entre tipos de veículo. Em resumo, o servidor na primeira coluna da linha i é o preferido, sendo que os demais atuam como *backups*.

O sistema permite formação de filas e, embora essa não seja capacitada, adotou-se um limite de nove usuários (o número de servidores), a fim de se calcular a probabilidade de chegada de um chamado quando a fila é tão grande que ele seria perdido. A probabilidade de receber um 10º chamado quando a fila já comporta sua capacidade máxima de usuários em espera pode ser calculada como

Tabela 4. Tempos de viagem t_{ji} (em minutos).

Nó	01	02	06	09	10	11	12	16	19	20	21	22	23	30	33	34	36	37	41	46	47	48
01	2	4	24	3	8	5	13	23	24	23	12	20	16	7	11	12	16	22	23	23	22	22
02	4	4	22	5	8	7	10	21	22	21	10	18	14	9	9	10	14	20	21	22	20	20
03	27	26	28	25	23	24	27	23	28	27	27	29	20	28	19	19	15	22	29	16	16	16
04	10	11	27	8	6	8	4	26	27	26	16	23	19	13	15	16	19	25	26	26	25	25
05	27	26	9	29	29	30	33	6	10	5	22	22	13	23	24	25	23	8	14	17	15	15
06	18	18	29	20	20	21	25	13	18	18	14	11	10	15	15	16	15	12	5	13	12	11
07	27	26	30	29	29	30	34	24	30	29	22	10	23	23	24	25	23	23	10	25	23	23
08	26	26	20	28	29	29	33	15	20	19	21	21	13	23	23	25	19	14	21	13	13	12
09	4	5	26	3	8	7	10	25	26	25	14	22	19	9	13	13	19	25	25	26	25	24
10	7	8	28	5	1	4	6	27	28	27	16	24	21	12	13	13	19	26	27	28	26	26
11	9	11	25	7	4	2	9	24	25	24	14	21	18	10	13	14	17	23	24	25	23	23
12	9	9	29	8	5	8	7	28	29	28	17	25	21	14	15	15	22	27	28	28	27	27
13	18	18	12	20	20	21	25	7	12	11	14	14	4	15	15	17	15	3	13	12	10	10
14	25	25	19	27	27	28	32	14	19	18	21	20	11	22	22	23	18	13	20	15	13	13
15	26	26	8	28	28	29	33	9	2	6	22	20	16	23	23	24	23	11	13	20	18	18
16	25	24	11	27	27	28	32	5	10	6	20	20	12	21	22	23	21	6	16	15	14	13
17	12	8	21	9	11	11	13	20	21	20	9	17	14	10	6	6	14	19	20	21	19	19
18	21	21	15	23	25	26	29	14	15	14	17	16	7	18	18	19	13	13	16	15	13	13
19	25	24	6	27	26	27	31	11	4	5	20	18	18	21	22	23	21	13	11	19	18	18
20	25	24	6	27	26	27	31	7	5	4	20	18	14	21	22	23	21	9	11	18	16	16
21	12	11	21	14	14	14	19	20	22	21	3	17	14	8	8	9	14	20	21	21	20	20
22	21	20	24	23	22	23	27	20	24	23	16	9	16	17	18	19	16	19	8	21	19	19
23	17	16	9	19	19	20	24	11	9	8	13	13	16	14	14	15	13	7	11	13	11	11
24	25	25	19	27	27	28	32	13	19	19	21	20	12	22	22	23	22	13	20	3	2	2
25	26	26	21	29	28	29	33	14	21	20	22	22	13	23	24	25	23	14	21	2	3	3
26	27	26	8	29	28	29	33	12	4	8	22	20	19	23	24	25	23	14	13	21	20	20
27	15	14	33	14	11	14	8	32	33	32	22	29	25	20	14	14	21	31	32	32	31	31
28	26	25	8	28	28	29	33	12	3	7	21	19	19	22	23	24	22	14	13	20	19	19
29	23	22	5	25	25	26	30	13	6	7	18	16	15	20	20	21	19	15	10	18	16	16
30	5	8	22	7	8	8	13	21	22	21	11	18	15	6	10	11	14	20	21	22	20	20
31	27	27	22	29	29	30	34	13	22	18	23	23	14	24	25	26	25	12	22	8	7	7
32	15	15	22	14	12	13	16	21	23	22	16	23	15	17	8	8	8	20	23	21	20	20
33	13	13	18	11	10	10	14	17	18	17	8	14	11	9	2	2	11	17	18	18	17	16
34	13	13	20	11	9	10	14	19	21	20	9	16	13	10	2	1	13	19	20	20	19	19
35	11	10	30	10	7	10	4	29	30	29	18	26	23	16	16	16	24	29	30	30	29	28
36	20	20	18	19	17	17	21	16	18	17	20	20	11	21	13	13	5	16	19	17	16	16
37	21	20	13	23	22	24	27	5	13	11	16	16	7	17	18	19	17	2	15	12	11	10
38	27	27	9	29	29	30	34	11	4	9	22	20	19	24	24	25	23	14	14	22	21	21
39	14	13	25	13	10	11	15	24	25	24	14	21	18	15	7	7	14	24	24	25	24	23
40	16	15	25	14	13	13	16	23	25	24	16	23	17	17	8	8	10	22	25	23	22	22
41	21	20	21	23	23	24	28	15	21	20	16	8	13	17	18	19	17	14	6	15	14	14
42	28	28	13	30	30	31	35	7	8	7	24	23	15	25	25	26	25	10	17	18	17	17
43	9	6	27	6	5	6	7	26	27	26	15	23	20	13	11	11	18	25	26	27	25	25
44	15	15	28	13	12	13	16	26	28	27	15	22	20	16	8	8	13	25	26	27	25	25
45	9	5	21	6	8	9	11	20	21	20	9	17	13	9	8	9	13	19	20	21	19	19
46	26	26	20	28	28	29	33	14	20	20	21	21	13	23	23	24	22	13	21	1	3	3
47	25	24	19	27	27	28	32	13	19	18	20	20	11	21	22	23	21	13	19	3	4	1
48	24	24	18	26	26	27	31	12	18	18	20	19	11	21	21	22	20	11	19	3	1	4

Fonte: Elaborado pelos autores.

Tabela 5. Intervalo entre chamadas sucessivas.

Dia	Intervalo (horas)	Dia	Intervalo (horas)
1	0,56	12	5,92
2	1,84	13	1,08
3	2,10	14	1,65
4	3,21	15	0,18
5	6,40	16	2,63
6	0,36	17	5,03
7	0,59	18	5,84
8	1,75	19	0,14
9	0,15	20	2,59
10	2,96	21	1,31
11	3,86	Média	2,39

Fonte: Elaborado pelos autores.

Tabela 6. Taxa de chegada por átomo λ_i (em chamados/hora).

Nó	Chamados	p_i	λ_i	Nó	Chamados	p_i	λ_i
01	278	1,56%	0,007	25	274	1,53%	0,006
02	278	1,56%	0,007	26	274	1,53%	0,006
03	236	1,32%	0,006	27	706	3,95%	0,017
04	396	2,22%	0,009	28	336	1,88%	0,008
05	166	0,93%	0,004	29	256	1,43%	0,006
06	406	2,27%	0,010	30	376	2,11%	0,009
07	166	0,93%	0,004	31	136	0,76%	0,003
08	256	1,43%	0,006	32	506	2,83%	0,012
09	334	1,87%	0,008	33	248	1,39%	0,006
10	334	1,87%	0,008	34	248	1,39%	0,006
11	334	1,87%	0,008	35	146	0,82%	0,003
12	334	1,87%	0,008	36	496	2,78%	0,012
13	286	1,60%	0,007	37	746	4,18%	0,017
14	166	0,93%	0,004	38	386	2,16%	0,009
15	406	2,27%	0,010	39	186	1,04%	0,004
16	316	1,77%	0,007	40	356	1,99%	0,008
17	526	2,94%	0,012	41	666	3,73%	0,016
18	396	2,22%	0,009	42	316	1,77%	0,007
19	1186	6,64%	0,028	43	261	1,46%	0,006
20	686	3,84%	0,016	44	261	1,46%	0,006
21	536	3,00%	0,013	45	1126	6,30%	0,026
22	493	2,76%	0,012	46	91	0,51%	0,002
23	493	2,76%	0,012	47	91	0,51%	0,002
24	274	1,53%	0,006	48	91	0,51%	0,002

Fonte: Elaborado pelos autores.

Tabela 7. Tempo médio de atendimento por servidor (em minutos).

Servidor	Tempo médio de atendimento	Desvio-padrão
1	59	41
2	95	67
3	81	65
4	156	94
5	18	14
6	86	77
7	43	30
8	114	68
9	29	23
Média USA	77	54
Média USB	75	53
Média	76	53

Fonte: Elaborado pelos autores.

$\rho^{10} p_{1...1}$. Na simulação da distribuição geográfica em que as ambulâncias estavam localizadas no momento da pesquisa, a probabilidade de ocorrência desse evento foi de 0,03%, enquanto que, nos cenários que serão tratados, foi inferior a 10^6 , indicando que essa limitação não traz alterações significativas à análise dos indicadores de desempenho do MH.

Os tempos médios de viagem no sistema (\bar{T}) e para cada átomo (\bar{T}_i) são dados pelas Expressões 23 e 24, respectivamente (Chiyoshi et al., 2000):

$$\bar{T} = \sum_{n=1}^N \sum_{i=1}^I f_{ni}^{[1]} \tau_{ni} + p_s \bar{T}_Q \tag{23}$$

$$\bar{T}_i = \left(\frac{\sum_{n=1}^N f_{ni}^{[1]} \tau_{ni}}{\sum_{n=1}^N f_{ni}^{[1]}} \right) (1 - p_s) + \sum_{j=1}^J \left(\frac{\lambda_j}{\lambda} \right) t_{ji} p_s \tag{24}$$

Nas equações anteriores, $f_{ni}^{[1]}$ é a fração dos despachos do servidor n ao átomo i que não implicam em tempo de espera, tal que $f_{ni}^{[1]} = (\lambda_i / \lambda) \sum_{B \in E_{ni}} p_B$, sendo λ_i / λ a probabilidade de uma chamada que incorre em fila ser originada no átomo i e E_{ni} , o conjunto dos estados nos quais uma ocorrência gerada no átomo i é atendida pelo servidor n . O tempo de viagem para um servidor n se deslocar até o átomo i é então dado por:

$$\tau_{ni} = \sum_{m=1}^I g_{nm} t_{mi} \quad (25)$$

em que $g_{nm} = 1$ quando o servidor n está alocado na facilidade $j = i$, e 0, caso contrário. Já p_s é a probabilidade de saturação do sistema, isto é, $p_s = p_0 + p_{11...1}$, tais que $p_0 = 1 - (p_{00...0} + p_{00...10} + \dots + p_{11...1})$ é a probabilidade de formação de fila. Finalmente, \bar{T}_Q é o tempo médio de viagem para uma chamada sujeita a espera, obtido por:

$$\bar{T}_Q = \sum_{j=1}^I \sum_{i=1}^I \frac{\lambda_j \lambda_i}{\lambda^2} t_{ji} \quad (26)$$

Aplicando-se o MH aos dados da configuração logística atual, obtém-se um tempo médio de 13 minutos. Com isso, observa-se que para apenas 42,4% da demanda o tempo de resposta do serviço é inferior à distância crítica de 12 minutos.

4 Desenvolvimento do modelo de localização

Tomando-se por base os modelos de localização encontrados na revisão bibliográfica, foram empregadas duas formulações de PI para localizar as bases e posicionar as ambulâncias do SAMU, maximizando a demanda coberta pelo serviço. As seguintes premissas foram consideradas:

- i. Da abertura de facilidades e alocação dos servidores: Exatamente uma ambulância deve ser alocada a cada facilidade aberta; embora não haja esse requisito no sistema real, tal abordagem toma respaldo no conceito de descentralização apresentado anteriormente e tende a aumentar a população coberta;
- ii. Das restrições de alocação: Tanto ambulâncias básicas como avançadas podem ser alocadas em qualquer vértice da rede e de maneira independente;
- iii. Da disponibilidade de recursos: Número de servidores limitado à disponibilidade atual da Secretaria de Saúde, sendo sete ambulâncias básicas e duas avançadas.

Adotando-se $P^c = P^p + P^s$ na formulação do FLEET, atende-se a todos as premissas supracitadas.

Sua formulação, no entanto, não considera o uso de *backups*, isto é, o modelo não leva em consideração as redundâncias que o sistema deve ter para evitar a formação de filas e minimizar o tempo de atendimento em áreas com demanda maior.

Por essa razão, o PLMD foi usado como uma abordagem alternativa, uma vez que também atende às premissas estabelecidas. Conforme foi visto, o PLMD trata a natureza estocástica do problema requerendo um determinado nível de confiança θ , garantido com o uso de *backups*. A desvantagem nesse caso é a possibilidade de considerar apenas um tipo de servidor, exigindo, portanto, uma hipótese simplificadora: unidades avançadas devem ser tratadas como unidades básicas.

Dado o porte dos programas matemáticos correspondentes aos cenários estudados, todos eles puderam ser otimizados com baixo custo computacional. Para um subconjunto das melhores soluções encontradas com os modelos de PI, o MH foi usado para avaliar os parâmetros de desempenho do sistema. Obviamente, tal abordagem consiste apenas na análise de alguns cenários específicos e o uso de Otimização Estocástica e Otimização Robusta tende a oferecer melhores soluções. Porém, através dessa estratégia binível foram identificados alguns aspectos que muito provavelmente farão parte de uma solução estocástica ótima. Mais precisamente, facilidades que têm grande chance de serem abertas, assim como os tipos de ambulâncias a elas alocadas.

5 Experimentos computacionais

Todos os testes foram executados em um *notebook* Dell Inspiron 14R 3350, com processador Intel Core™ i5, sistema operacional Windows 7 Ultimate 64 bits e 6GB de memória RAM. Empregou-se o AIMMS 3.13 para programar os modelos de PI, e a otimização dos programas matemáticos foi feita através do CPLEX 12.6. Para cada modelo foram escolhidas as melhores 200 soluções obtidas pelo CPLEX, sendo elas repassadas uma a uma ao Microsoft Excel 2010, onde um código em VBA foi usado para implementar as equações do MH, gerando, assim, para cada solução do modelo de PI, os respectivos indicadores de desempenho.

Inicialmente aplicou-se o modelo FLEET com a disponibilidade de servidores existente, ou seja, considerando dois servidores avançados ($P^s = 2$) e sete do tipo básico ($P^p = 7$) e nove facilidades ($P^c = 9$). Quando implementada no MH, a melhor solução cobriu 66,3% da demanda e resultou em um tempo médio de viagem de 11,4 minutos. Os veículos foram posicionados da seguinte maneira: as ambulâncias de suporte avançado ficaram nos nós 1 e 34, enquanto as de suporte básico, nos nós 6, 12, 19, 20, 33, 41 e 46, sendo uma por facilidade. Nota-se que apenas os postos médicos de Campos Elíseos e de

Saracuruna (nós 6 e 41, respectivamente) utilizados no cenário no momento da pesquisa eram comuns ao posicionamento provido pelo FLEET. Isso indica o quão diferente aquele posicionamento estava do proposto pelo modelo, ratificando a importância do uso de métodos computacionais para solução desse problema.

O PLMD foi resolvido com $P = 9$ servidores, ou seja, não se estabeleceu distinção entre as ambulâncias básicas e avançadas, a fim de manter a homogeneidade prevista pelo modelo. Para um nível de confiança de $\theta = 93\%$, o que corresponde à cobertura de pelo menos $b = 5$ servidores, obtém-se uma solução na qual se cobre 43,8% da demanda da rede dentro do limite de tempo adotado, através do MH. Esse resultado pode ser explicado pelo nível mais alto de confiabilidade exigido, que tendeu a concentrar os servidores em regiões mais populosas, em detrimento de grande parte dos nós da rede. O tempo médio de viagem do sistema foi de 14,8 minutos e as ambulâncias foram alocadas em 1, 2, 9, 10, 11, 12, 30, 33 e 34. Com $\theta = 88\%$ de confiabilidade, exige-se uma cobertura de no mínimo quatro servidores em cada área de demanda. A partir da solução ótima, o MH sugeriu uma cobertura de 62,6%, com tempo médio para início do atendimento de 10,9 minutos, sendo as ambulâncias alocadas nos nós 1, 2, 6, 16, 19, 20, 23, 33 e 34. Finalmente, para $\theta = 80\%$, são necessários três servidores, que possibilitam que 72% da população seja atendida em até 12 minutos, com servidores em 2, 6, 9, 10, 16, 19, 20, 33 e 34. É interessante notar que as ambulâncias foram geograficamente agrupadas em dois *clusters* ao longo da rede e que o tempo médio de viagem do sistema foi de 11 minutos. A Tabela 8 resume as métricas discutidas para cada um dos cenários analisados.

Nesse ponto cabe destacar uma observação feita por Chiyoshi et al. (2003) quanto à possibilidade de comparar métricas de cobertura calculadas por modelos distintos. As mesmas nem sempre são equiparáveis, especialmente em formulações estocásticas, uma vez que as premissas nelas adotadas têm implicações práticas relevantes. O mesmo ocorre entre os modelos de PI empregados neste trabalho, que diferem no conjunto de suas hipóteses subjacentes. No presente trabalho, o que viabiliza a comparação das soluções dos dois modelos (e também da configuração existente no

momento de realização da pesquisa) é que a métrica de cobertura usada não provém dos modelos de PI, FLEET e PLMD, mas do MH aplicado às localizações propostas por cada um.

Nota-se que, apesar da proximidade entre os tempos médios, o percentual da população coberta varia de maneira considerável, já que poucas trocas entre as facilidades abertas podem afetar a média. Por exemplo, quando $\theta = 93\%$ no PLMD, o modelo concentra os servidores para reduzir o tempo de resposta em regiões específicas da rede, o que, em contrapartida, eleva o tempo de viagem para outros nós da rede, deixando-os descobertos. Todavia, a média do tempo até o início do atendimento permanece equilibrada. O mesmo ocorreu na configuração do sistema no momento de realização da pesquisa, que possuía tempo médio do sistema muito próximo àqueles dos modelos de Otimização, mas resultado significativamente inferior quando analisada a cobertura.

O Gráfico 1 mostra o percentual da população coberta em função do tempo de resposta. É possível verificar como as soluções propostas pelo PLMD fornecem menores tempos de resposta do que a configuração existente no momento de realização da pesquisa e que o FLEET. Além disso, para $\theta = 80\%$ e $\theta = 88\%$, os resultados são sempre melhores do que a configuração atual, e se tem, praticamente, os mesmos tempos críticos para cobertura entre eles, sugerindo que a localização tem um efeito tão importante quanto o número de servidores usados como *backup*. Vê-se ainda que, para um tempo crítico superior a 14 minutos, o FLEET apresenta percentuais de cobertura superiores aos obtidos pelo PLMD. Entretanto, deve-se atentar para o fato de que o primeiro não considera os efeitos advindos do *backup*, o que tende a reduzir a chance de atendimento.

A Figura 2 apresenta um resumo comparativo da distribuição espacial das ambulâncias no qual: **2a** mostra o arranjo logístico atual; **2b**, o resultado obtido pela aplicação do modelo FLEET; **2c**, **2d** e **2e** ilustram o resultado obtido através do PLMD, para $\theta = 88\%$, $\theta = 93\%$ e $\theta = 80\%$, respectivamente. Nos casos 2a e 2b, o marcador com um asterisco indica o posicionamento de ambulâncias do tipo avançada. Nos casos restantes não há marcas, uma vez que

Tabela 8. Comparação dos resultados obtidos pelo Modelo do Hipercubo.

Cenário	Demanda coberta (%)	Tempo médio de viagem (min)
Situação Atual	42,4	13,0
FLEET	66,3	11,4
PLMD ($\theta = 93\%$)	43,8	14,8
PLMD ($\theta = 88\%$)	62,6	10,9
PLMD ($\theta = 80\%$)	72,0	11,0

Fonte: Elaborado pelos autores.

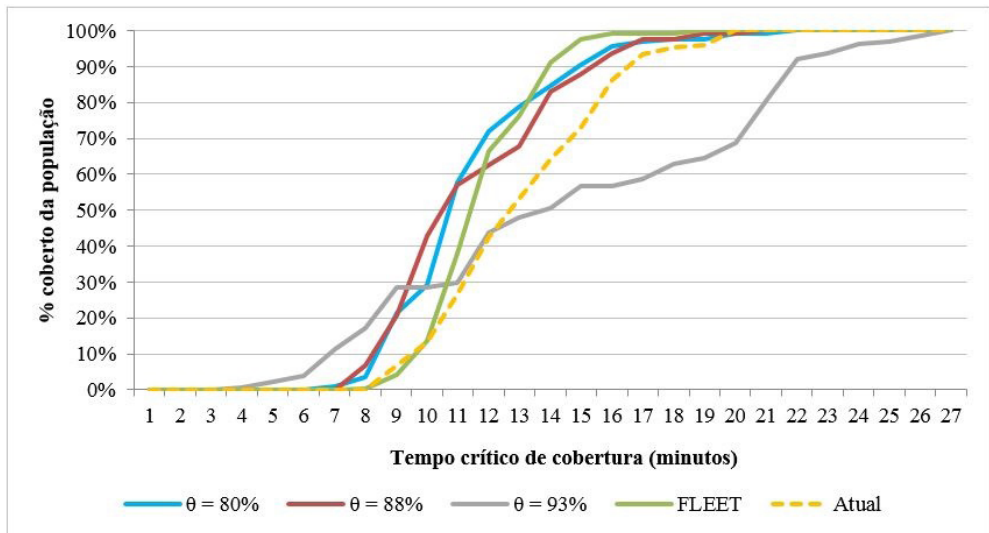


Gráfico 1. Crescimento da população coberta para variações no tempo crítico. Fonte: Elaborado pelos autores.

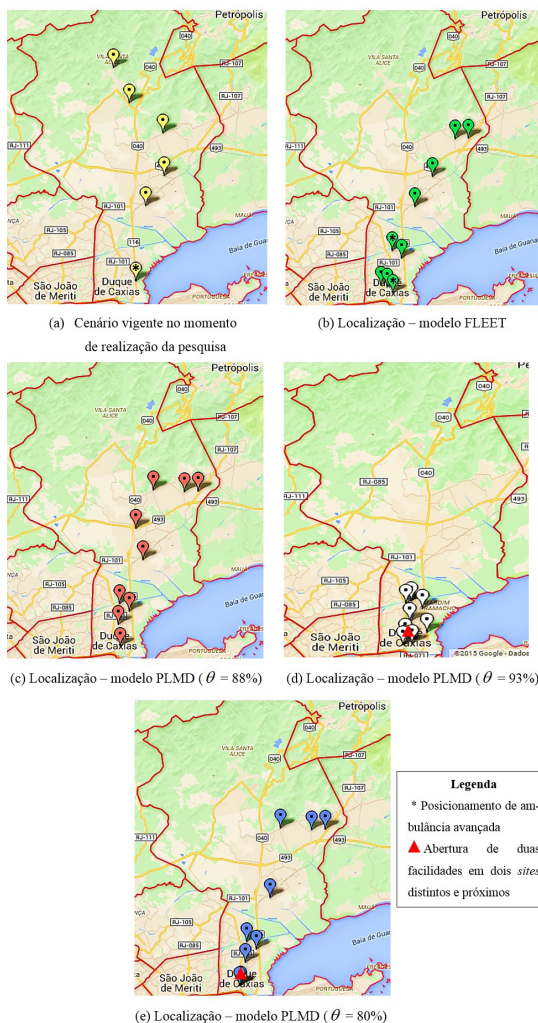


Figura 2. Representação geográfica da localização dos servidores. Fonte: Elaborado pelos autores.

ilustram soluções do PLMD, cujos pressupostos não fazem distinção entre os servidores.

Enquanto a configuração existente no momento de realização da pesquisa possuía somente seis pontos de localização, uma vez que havia quatro servidores concentrados na mesma facilidade, os demais modelos distribuem as ambulâncias ao longo da rede, de acordo com a demanda, garantido que mais usuários sejam cobertos. Outro ponto importante é a concentração de ambulâncias ao sul da cidade, como se vê nas Figuras 2b e 2d. Esse é o centro de Duque de Caxias, região mais populosa e de onde provém o maior número de chamados. Por fim, é importante destacar que as Figuras 2d e 2e utilizam exatamente nove facilidades, o que não pode ser visualizado devido à escala do mapa, uma vez que existem dois locais que, por serem geograficamente próximos, acabam se sobrepondo. No entanto, a escala foi mantida para facilitar as comparações entre as imagens, sendo as facilidades que se sobrepõem marcadas com o símbolo ▲.

Também é possível notar como o PLMD varia a distribuição geográfica de acordo com o crescimento do nível de confiança θ . Nas Figuras 2c e 2e, nas quais $\theta = 88\%$ e $\theta = 80\%$, respectivamente, percebe-se uma tendência de formação de dois *clusters*, um ao sul e outro a nordeste. Na Figura 2d observa-se que o maior nível de confiabilidade requerido ($\theta = 93\%$) produz um único *cluster* ao sul. Muito provavelmente, a disponibilidade de apenas nove ambulâncias não estimula o modelo a criar outros *clusters*, uma vez que não haveria ganhos em termos de cobertura, porque os nós restantes não teriam cinco ou mais servidores remotos para responder às suas chamadas dentro de 12 minutos.

6 Conclusão

Neste trabalho foram combinadas técnicas de Otimização e Teoria de Filas para analisar o posicionamento das ambulâncias do SAMU de Duque de Caxias, RJ, tendo como principais resultados um aumento significativo na cobertura e um tempo de resposta médio inferior ao tempo referente à configuração no momento de realização da pesquisa, considerando o mesmo número de servidores existentes. As soluções propostas foram significativamente diferentes daquelas usadas na configuração vigente no momento de realização da pesquisa, reiterando a importância de métodos matemático-computacionais em estudos de localização.

As soluções obtidas a partir de problemas de Programação Inteira foram usadas no Modelo do Hipercubo com intuito de avaliar o despacho das ambulâncias e o desempenho do sistema sob congestionamento. As análises indicam que PLMD fornece uma melhor modelagem do comportamento estocástico do problema, conduzindo assim a soluções com maior nível de serviço para os mesmos valores de tempo de serviço. Os resultados também sugerem uma tendência ao equilíbrio entre dispersão dos servidores e formação de *clusters* na rede, como tentativa de maximizar a cobertura, ao passo que aumenta a chance de atendimento pela presença de *backups*.

Recomenda-se que estudos futuros façam uma coleta de dados mais ampla e precisa, uma vez que essa foi uma das principais dificuldades encontradas no decorrer desta pesquisa. Outra direção importante para continuação da pesquisa aponta para o emprego de modelos estocásticos, tendo em face o desenvolvimento dessa área nos últimos anos, particularmente das técnicas de Otimização Estocástica e Robusta. Nesse sentido, seria válido comparar a qualidade e a complexidade de se obter as soluções por tais modelos com aquelas propostas por uma abordagem binível como a que se aplicou neste trabalho.

Referências

- Ball, M. O., & Lin, F. L. (1993). A reliability model applied to emergency service vehicle location. *Operations Research*, 41(1), 18-36. <http://dx.doi.org/10.1287/opre.41.1.18>.
- Batta, R., Dolan, J. M., & Krishnamurthy, N. P. (1989). The maximal expected covering location problem: revisited. *Transportation Science*, 23(4), 277-287. <http://dx.doi.org/10.1287/trsc.23.4.277>.
- Bertelli, A., Bueno, M. R., & Sousa, R. M. C. (1999). Estudo preliminar das relações entre duração da parada cardiorrespiratória e suas consequências nas vítimas de trauma. *Revista da Escola de Enfermagem da U S P*, 33(2), 130-141. <http://dx.doi.org/10.1590/S0080-62341999000200004>.
- Brandeau, M., & Larson, R. C. (1986). Extending and applying the hypercube queueing model to deploy ambulances in Boston. In A. J. Swersey & E. J. Ingwall (Eds.), *Delivery of urban services* (TIMS Studies in the Management Science, Vol. 22, pp. 121-153). London: Elsevier.
- Brasil. Ministério da Saúde. (2013). *Portaria GM/MS n.º 2.657, 16 de dezembro de 2004. Estabelece as atribuições das centrais de regulação médica de urgências e o dimensionamento técnico para a estruturação e operacionalização das Centrais Samu-192*. Brasília, DF: Diário Oficial da República Federativa do Brasil. Recuperado em 20 de agosto de 2013, de: http://bvsms.saude.gov.br/bvs/saudelegis/gm/2011/prt2026_24_08_2011.html.
- Brotcorne, L., Laporte, G., & Semet, F. (2003). Ambulance location and relocation models. *European Journal of Operational Research*, 147(3), 451-463. [http://dx.doi.org/10.1016/S0377-2217\(02\)00364-8](http://dx.doi.org/10.1016/S0377-2217(02)00364-8).
- 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. <http://dx.doi.org/10.1590/S0104-530X2000000200005>.
- Chiyoshi, F., Galvão, R. D., & Morabito, R. (2003). A note on solutions to the maximal expected covering location problem. *Computers & Operations Research*, 30(1), 87-96. [http://dx.doi.org/10.1016/S0305-0548\(01\)00083-1](http://dx.doi.org/10.1016/S0305-0548(01)00083-1).
- Church, R. L., & Revelle, C. S. (1974). The maximal covering location problem. *Papers / Regional Science Association. Regional Science Association. Meeting*, 32(1), 101-118. <http://dx.doi.org/10.1007/BF01942293>.
- Galvão, R. D., Chiyoshi, F., & Morabito, R. (2003a). Towards unified formulations and extensions of two classical probabilistic location models. *Computers & Operations Research*, 32(1), 15-33. [http://dx.doi.org/10.1016/S0305-0548\(03\)00200-4](http://dx.doi.org/10.1016/S0305-0548(03)00200-4).
- Galvão, R., Chiyoshi, F., Espejo, L., & Rivas, M. A. (2003b). Solução do problema de localização de máxima disponibilidade utilizando o modelo Hipercubo. *Pesquisa Operacional*, 23(1), 61-78. <http://dx.doi.org/10.1590/S0101-74382003000100006>.
- Gendreau, M., Laporte, G., & Semet, F. (1997). Solving an ambulance location model by tabu search. *Location Science*, 5(2), 75-88. [http://dx.doi.org/10.1016/S0966-8349\(97\)00015-6](http://dx.doi.org/10.1016/S0966-8349(97)00015-6).
- Gendreau, M., Laporte, G., & Semet, F. (2001). A dynamic model and parallel Tabu search heuristic for real-time ambulance relocation. *Parallel Computing*, 27(12), 1641-1653. [http://dx.doi.org/10.1016/S0167-8191\(01\)00103-X](http://dx.doi.org/10.1016/S0167-8191(01)00103-X).
- Ghiani, G., Laporte, G., & Musmanno, R. (2004). *Introduction to logistics systems planning and control*. West Sussex: John Wiley & Sons Ltd.
- Iannoni, A. P., Morabito, R., & Saydam, C. (2009). An optimization approach for ambulance location and the districting of the response segments on highways.

- European Journal of Operational Research*, 195(2), 528-542. <http://dx.doi.org/10.1016/j.ejor.2008.02.003>.
- Instituto Brasileiro de Geografia e Estatística – IBGE. (2014). *Censo demográfico 2010*. Rio de Janeiro: IBGE. Recuperado em 26 de março de 2014, de <http://cidades.ibge.gov.br/xtras/perfil.php?codmun=330170>
- Larson, R. C. (1974). A hypercube queuing model for facility location and redistricting in urban emergency services. *Computers & Operations Research*, 1(1), 67-95. [http://dx.doi.org/10.1016/0305-0548\(74\)90076-8](http://dx.doi.org/10.1016/0305-0548(74)90076-8).
- Larson, R. C. (1975). Approximating the performance of urban emergency service systems. *Operations Research*, 23(5), 845-868. <http://dx.doi.org/10.1287/opre.23.5.845>.
- Revelle, C. S., & Hogan, K. (1989). The maximum availability location problem. *Transportation Science*, 23(3), 192-200. <http://dx.doi.org/10.1287/trsc.23.3.192>.
- Schilling, D. A., Elzinga, D. J., Cohon, J., Church, R. L., & Revelle, C. S. (1979). The TEAM/FLEET models for simultaneous facility and equipment siting. *Transportation Science*, 13(2), 163-175. <http://dx.doi.org/10.1287/trsc.13.2.163>.
- Schmid, V., & Doerner, K. F. (2010). Ambulance location and relocation problems with time-dependent travel times. *European Journal of Operational Research*, 207(3), 1293-1303. PMID:21151327. <http://dx.doi.org/10.1016/j.ejor.2010.06.033>.
- Souza, R., Morabito, R., Chiyoshi, F., & Iannoni, A. (2013). Análise da configuração de SAMU utilizando múltiplas alternativas de localização de ambulâncias. *Gestão & Produção*, 20(2), 287-302. <http://dx.doi.org/10.1590/S0104-530X2013000200004>.
- Taha, H. A. (2008). *Pesquisa operacional* (8. ed.). São Paulo: Pearson Prentice Hall. 326 p.
- Takeda, R., Widmer, J., & Morabito, R. (2004). Aplicação do modelo Hiper cubo de filas para avaliar a descentralização de ambulâncias em um sistema urbano de atendimento médico de urgência. *Pesquisa Operacional*, 24(1), 39-71. <http://dx.doi.org/10.1590/S0101-74382004000100004>.
- Toregas, C. R., Swain, R., Revelle, C. S., & Bergman, L. (1971). The location of emergency service facilities. *Operations Research*, 19(6), 1363-1373. <http://dx.doi.org/10.1287/opre.19.6.1363>.