

ANÁLISE DAS MEDIDAS DE DESEMPENHO DO SERVIÇO DE ATENDIMENTO MÓVEL DE URGÊNCIA E EMERGÊNCIA EM CURITIBA-PARANÁ ATRAVÉS DO MODELO MATEMÁTICO HIPERCUBO DE FILAS COM PRIORIDADES.

Alessandra Memari Pavanelli

Universidade Federal Paraná (UFPR)
CP19081 – Curitiba,PR; CEP 81531-990
alessandracmc@bol.com.br

Deise Maria Bertholdi Costa

Universidade Federal Paraná (UFPR)
Departamento de Expressão Gráfica
CP19081 – Curitiba,PR; CEP 81531-990
deise@ufpr.br

Genival Pavanelli

Universidade Federal Paraná (UFPR)
CP19081 – Curitiba,PR; CEP 81531-990
pavanellc@bol.com.br

Paulo Henrique Siqueira

Universidade Federal Paraná (UFPR)
Departamento de Expressão Gráfica
CP19081 – Curitiba,PR; CEP 81531-990
paulohs@ufpr.br

Reginaldo Santa Rosa

Universidade Federal Paraná (UFPR)
CP19081 – Curitiba,PR; CEP 81531-990
Reginaldo_00@hotmail.com

RESUMO

O presente trabalho utilizou o modelo hipercubo de filas com prioridades a fim de analisar as medidas de desempenho do sistema de atendimento de urgência e emergência na cidade de Curitiba-PR, bem como compará-las com os dados obtidos no sistema. Com o objetivo de analisar as configurações dos sistemas de atendimento médicos, foram calculadas as medidas de desempenho tais como *workload* dos veículos, tempo de atendimento às vítimas, tempos de viagens dos veículos, dentre outras. Analisando essas medidas de desempenho e comparando com o cenário original, poderemos propor novas distribuições dos veículos a fim de viabilizar um menor tempo de atendimento ao chamado de um usuário do sistema. Para ilustrar a aplicação do modelo hipercubo de filas com prioridades, analisamos os resultados de um estudo de caso de uma empresa de atendimento de urgência e emergência da cidade de Curitiba-PR.

PALAVRAS CHAVE. Modelo Hipercubo de Filas e Medidas de Desempenho e Atendimento de Urgência e Emergência.

Área principal: MP – Modelos Probabilísticos, OA – Outras Aplicações em PO

ABSTRACT

This study used the hypercube queuing model with priorities in order to analyze the performance measures of the care and emergency system in the city of Curitiba-PR, and also to compare them with data obtained in the system. In order to analyze the configurations of medical care systems, we calculated the performance measures such as vehicle's workload, attendance time to the victims, travel times of vehicles, among others. Analyzing these performance measures and by comparing the original scenario, we propose new distributions of vehicles in order to provide a shorter service time at the call of the system's users. To illustrate the application of the hypercube queuing model with priorities, we analyze the results of a case study of an emergency care company in the city of Curitiba-PR.

KEYWORDS. Hypercube Queuing Model, Performance Measures and Care and Emergency System.

Main área: PM – Probabilistic Models, OA – Other Applications in Operations Research

1. Introdução

No Brasil, o Ministério da Saúde foi instituído em 25 de julho de 1953 com a Lei nº 1.920, embora sua história tenha iniciado em 1808. O ministério limitava-se apenas a atender os problemas mais importantes da saúde pública, e tratava-se da principal unidade administrativa de ação sanitária do governo. Em 25 de fevereiro de 1967, com a implantação da Reforma Administrativa Federal, o Ministério da Saúde passou a ser o responsável pela formulação e coordenação da Política Nacional de Saúde: atividade médica e paramédica, ação preventiva em geral, vigilância sanitária de fronteira, controle de drogas, entre outras. O Ministério da Saúde passou por diversas reformas ao longo desses anos, destaca-se a determinação, instituída pela Constituição Federal de 1988, da obrigatoriedade do estado garantir saúde a toda a população, nesse momento criou-se o sistema único de saúde SUS. Outro marco importante foi à aprovação da Lei Orgânica da Saúde pelo Congresso Nacional no ano de 1990.

É sabido que a qualidade de vida de uma população está diretamente relacionada ao acesso à saúde de boa qualidade. Apesar dos avanços, os serviços de saúde no Brasil tanto em áreas urbanas como os prestados nas rodovias, necessitam de novas propostas de hierarquização e regionalização para uma assistência de melhor qualidade, maior rapidez e custos menores. Buscando atender essas necessidades foram criados os sistemas de atendimento emergencial (SAE) das cidades e das rodovias, os quais têm como objetivo socorrer as vítimas, prestando os devidos atendimentos no menor tempo possível. Desta forma, o SAE não poderá admitir fila de espera, pois tem a função de socorrer as vítimas de acidente e transportá-las até o hospital mais próximo, o mais rápido possível.

Neste trabalho serão analisados dados dos serviços de atendimento de urgência e emergência prestada por uma empresa, que iniciou suas atividades em Curitiba-PR no ano de 1992. Sua finalidade é realizar atendimentos as solicitações de socorro fora do ambiente hospitalar com o principal objetivo de reduzir qualquer tipo de sequelas decorrentes da falta de socorro precoce. Seus sistemas operam com políticas de despacho pré-determinadas, como por exemplo, somente um veículo pode ser despachado para atender a um chamado emergencial. Devido a essas características primordiais, modelos matemáticos de localização de facilidades tais como o modelo hipercubo de filas, podem vir a ser uma ferramenta muito útil a fim de direcionar o planejamento de sistemas de saúde emergencial.

Listam-se abaixo trabalhos correlatos ao tema aqui apresentado bem como a abordagem das técnicas com as quais se pretende trabalhar, dentre as quais se ressalta modelo hipercubo de filas, algoritmo genético e simulação.

No trabalho de TAKEDA (2000), foi realizada uma análise no sistema de atendimento de urgência e emergência (SAMU) em Campinas, SP. O problema foi tratado com o modelo hipercubo de filas que se mostrou muito eficiente nesse tipo de abordagem, onde o serviço

prestado é realizado com uma demanda aleatória e espacialmente distribuída. O modelo gera vários indicadores de desempenho para o sistema, podendo estes ser comparado com o sistema atual. Criaram-se cenários alternativos, novas localizações das ambulâncias, a fim de avaliar novas medidas de desempenho e verificar a melhor solução em termos de tempo de serviço prestado, já que o tempo é de suma importância nesse tipo de atendimento.

ALBINO (1994) trata do dimensionamento de um sistema de atendimento as emergências geradas pelas interrupções no fornecimento de energia elétrica. As concessionárias de energia elétrica, em caso de interrupções de energia, necessitam atender a população no menor tempo possível. Esse tempo que têm comportamento aleatório é dado pela soma do tempo de espera, tempo de deslocamento e tempo de execução do serviço. O modelo hipercubo de filas se enquadrou perfeitamente devido sua complexidade geográfica e temporal.

No trabalho de MORABITO *et al.* (2001), o objetivo foi analisar o uso e a solução do modelo hipercubo para o caso de servidores não homogêneos (servidores com diferentes tempos médios de serviço). Sistemas com servidores não homogêneos podem ser encontrados em diversas situações reais, como por exemplo, nos Serviços de Atendimento Móvel de Urgência (SAMU) de algumas cidades brasileiras. A importância de se considerar explicitamente a não homogeneidade dos servidores no modelo hipercubo é mostrada inicialmente através de um exemplo ilustrativo. É em seguida demonstrado que a solução para o caso não homogêneo pode ser obtida vantajosamente pelo método de Gauss-Siedel. Este método foi testado para uma rede de 55 vértices disponível na literatura, em modelos com 10 a 17 servidores, variando-se a taxa de ocupação do sistema de 0,1 a 0,9. Finalmente, propõe-se um modelo de regressão para estimar o tempo de processamento necessário para resolver um dado problema.

MORABITO *et al.* (2008) apresentam extensões do modelo hipercubo de filas com algoritmo genético a fim de otimizar a configuração e operação de sistemas de atendimento emergenciais em rodovias. O objetivo de tais procedimentos é inicialmente utilizar um método para localizar as bases de ambulância ao longo da rodovia, e em seguida estender a abordagem para apoiar as duas decisões combinadas, ou seja, localizar as bases das ambulâncias e dimensionar as regiões de cobertura de cada base na rodovia. Foram analisados os resultados de dois estudos de caso em rodovias brasileiras.

O trabalho de COSTA (2003) foi realizado junto ao corpo de bombeiros na cidade de Curitiba, PR. O objetivo principal foi propor uma metodologia para determinação de zonas de atendimento para unidades de serviços emergenciais, estabelecendo assim novas posições de localização para as viaturas, de modo que suas áreas de atendimento fossem homogêneas de acordo com algum critério pré-estabelecido. Foram coletadas algumas medidas de desempenho tais como tempo médio de deslocamento até o local do acidente, *workload* média das viaturas entre outras, as quais foram utilizadas a fim de gerar uma estimativa entre o Atendimento Simulado e o Modelo Hipercubo. Foram estabelecidas, através da metodologia proposta, zonas de atendimento para as viaturas que diminuiriam os tempos de deslocamento bem como os tempos em fila de espera.

SOUZA (2010) apresenta um estudo de caso no SAMU de Ribeirão Preto, SP, cujo objetivo foi o de aplicar o modelo hipercubo para considerar fila com prioridade e propor uma abordagem para múltiplas configurações de localização das ambulâncias, explorando variações importantes da demanda e do serviço ao longo do dia. Vários cenários alternativos foram analisados e os resultados foram muito interessantes, ajudando na tomada de decisão por parte dos operadores do sistema de atendimento da saúde.

A literatura apresenta diversos problemas de localização probabilísticos dos serviços urbanos de emergência, principalmente a partir da década de 60, quando foram definidos os primeiros modelos matemáticos nesse assunto. O modelo hipercubo proposto em LARSON (1974), baseado em teoria de filas espacialmente distribuídas, tem se mostrado muito eficaz para analisar todos os tipos de serviços de localização, incluindo os SAEs em rodovias.

O principal objetivo desse trabalho é o de avaliar as medidas de desempenho dos atuais sistemas de atendimentos de urgências e emergências na cidade de Curitiba, através do modelo hipercubo de filas espacialmente distribuídos com prioridade, que considera variações aleatórias

dos processos de chegadas e atendimento das chamadas. O modelo matemático fornecerá indicadores de desempenho, os quais serão comparados com os dados reais do sistema a fim de otimizar a operação e configuração desse sistema.

2. Descrição do problema, levantamento e tratamento dos dados

A empresa que forneceu os dados para o presente trabalho é a primeira empresa de emergência médica do Brasil, surgiu em 1992 em Curitiba PR, uma das pioneiras da América latina e tem por finalidade prestar socorro emergencial às pessoas bem como garantir a qualidade no atendimento. Assim, torna-se um desafio constante oferecer um serviço de boa qualidade para a população, com o menor custo possível. O serviço funciona 24 horas por dia e é composto por profissionais da saúde: médicos, enfermeiros, auxiliares de enfermagem e socorristas. Esse serviço presta o atendimento em qualquer local, residências, locais de trabalho e vias públicas. O socorro é feito após uma ligação telefônica. A empresa tem a responsabilidade de socorrer as vítimas e, nos casos mais urgentes, fazer ainda um pré-atendimento, no sentido de estabilizar a vítima e depois encaminhá-la a um hospital próximo. Uma Central de Regulação Médica recebe os chamados e encaminha para o médico regulador, que julga e decide sobre a gravidade do caso, orientando o atendimento a ser realizado e disponibilizando o envio específico do veículo.

O sistema conta com cinco veículos para casos graves (UTIs), cinco veículos para casos não tão graves e nove veículos para casos leves (remoção). A diferença básica entre os tipos de veículos é dada pelos equipamentos que cada um possui. O veículo a ser despachado depende da gravidade do chamado e da sua localização. Os veículos estão distribuídos em três bases de atendimento localizadas em Curitiba e na região metropolitana.

2.1 Pesquisa de campo

Os dados que nortearam o presente trabalho foram obtidos, conforme dito anteriormente, junto à uma empresa de urgência e emergência de Curitiba, que além de disponibilizar tais dados, elucidou acerca do funcionamento do sistema, contribuindo de maneira significativa para o levantamento dos dados bem como a interpretação dos resultados obtidos após a aplicação da ferramenta utilizada nesse estudo de caso. O sistema possui uma informatização que agiliza o processamento das chamadas, mas não existem avaliações periódicas que indicam o desempenho do sistema.

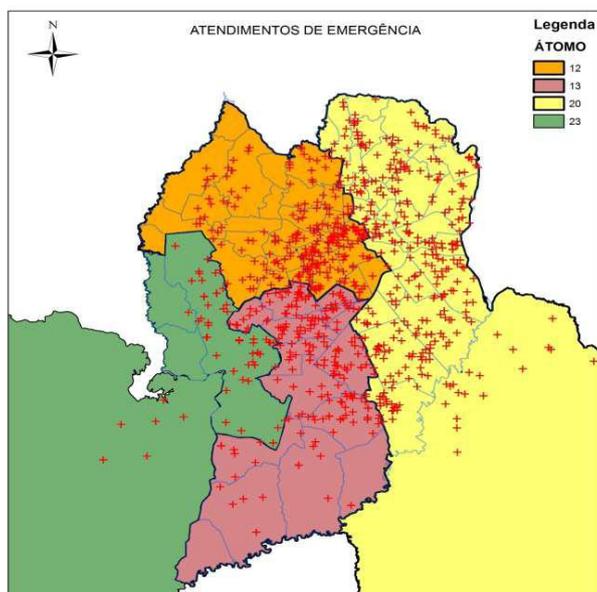


Figura 2.1 Divisão de Curitiba (BPM)

A pesquisa de campo foi iniciada no ano de 2012, o sistema opera com 19 veículos, porém para essa pesquisa apenas dez foram considerados pois serviam casos emergenciais.

A fim de estipular regiões de atendimento a grande Curitiba foi dividida em quatro áreas distintas segundo os Batalhões da Polícia Militar (BPM) de Curitiba. São elas: 12° BPM, 13° BPM, 20° BPM, 23° BPM, sendo que o município de Araucária foi adicionado ao polígono do 23° BPM e o de São José dos Pinhais ao do 20° BPM.

Os veículos, quando disponíveis, ficam localizados em apenas três das quatro áreas, ou seja, na área 23° BPM não há base. O sistema admite fila de espera de um chamado. A Figura 2.1 mostra a divisão da grande Curitiba em seus quatro átomos e a quantidade de chamados em sete dias da amostra analisados.

2.2 Coleta dos dados

A coleta de dados foi realizada em dois momentos distintos. No primeiro momento analisou-se a quantidade e tipos de atendimentos ocorridos ao longo do ano e dos meses. Em seguida coletou-se uma amostra de sete dias para validar as hipóteses do modelo hipercubo de filas.

De acordo com as estatísticas apresentadas pela empresa ao longo dos anos de 2012 e 2013 verificou-se que não existem diferenças significativas entre os meses com relação ao número de chamadas, exceto em meses de férias. Existem também diferenças entre dias úteis e finais de semanas, bem como em feriados prolongados, entretanto, essas diferenças são aceitáveis para esse tipo de serviço de emergência.

Quanto à segunda parte da coleta de dados, o objetivo foi o de analisar uma amostra aleatória de um determinado mês do ano, a fim de verificar se os dias se comportam de maneira semelhante. O mês escolhido foi fevereiro do ano de 2013 e a amostra contém sete dias.

Ao longo do dia verificou-se que em determinados períodos existia um aumento do número de chamados em todos os átomos conforme Gráfico 2.1, esse horário variava entre 13 horas e 14 horas para cada dia, conforme apresentado no Gráfico 2.2.

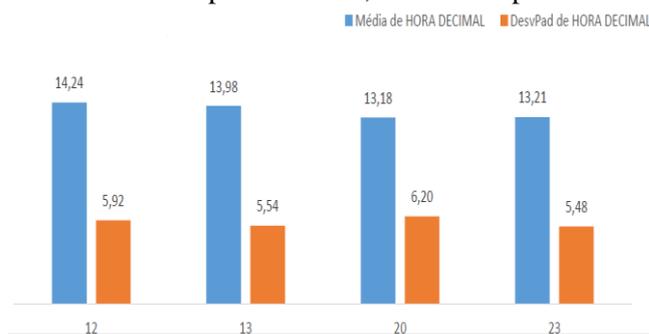


Gráfico 2.1 Média do horário de chegada de ocorrências

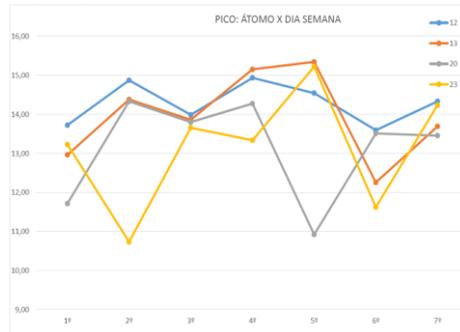


Gráfico 2.2 Horário de "pico"

Buscando encontrar o período de pico com o qual se pretendia trabalhar, dividiu-se o dia em intervalos aleatórios de horas, calculando a média e o desvio padrão para cada um destes intervalos.

Conforme Tabela 2.1 o horário de pico ficou caracterizado pelo período compreendido entre as 10 e as 14 horas, uma vez que este apresentou maior taxa de chegada dos chamados, conseqüentemente a maior média. Vale lembrar que todos os intervalos entre as 9 e 16 horas foram analisados e constatou-se em todos eles que as taxas de chegada foram inferiores ao intervalo mencionado como pico. Na tabela 2.1 apresentamos apenas alguns dos intervalos com suas médias e seus respectivos desvios padrão.

	1º dia	2º dia	3º dia	4º dia	5º dia	6º dia	7º dia	Média	Desv. Pad
9-14h	0,17	0,14	0,1	0,123	0,123	0,156	0,123	0,125	0,023572
10-14h	0,18	0,13	0,12	0,1416	0,123	0,15	0,1208	0,13108	0,021717
11-15h	0,17	0,12	0,1	0,1116	0,096	0,138	0,1458	0,11828	0,026744
13-16h	0,13	0,1	0,14	0,122	0,116	0,133	0,127	0,1276	0,013077

Tabela 2.1 Taxas médias de chegadas para possíveis períodos de pico (chegadas por minutos)

A amostra coletada conta com 1142 chamados, dos quais 3,32% não puderam ser georeferenciadas por motivo de endereço incompletos. Trabalhou-se então com 1104 atendimentos no total, dos quais 30,45% provem da região do 12º BPM, 22,5% da região do 13º BPM, 39,98% da região do 20º BPM e 7% da região do 23º BPM. A proporção dos chamados se manteve para o período de pico, onde ocorreram 239 chamados dos quais, 71 foram casos graves e o restante, 168 chamados, casos não tão graves, porém emergenciais.

3. Metodologia

3.1 Teoria de filas

Teoria de filas é uma das técnicas da pesquisa operacional que trata de assuntos relacionados a problemas de congestionamento de sistemas. Sistemas que de alguma forma podem gerar filas devido a alguma(s) restrição(ões) do atendimento.

A teoria de filas pode ser aplicada nos mais diversos serviços básicos de atendimento que a população utiliza, entre eles podemos citar: atendimento médico emergencial, policiamento, transportes coletivos, coleta de lixo, entre outros. À medida que a população aumenta, o que não ocorre de maneira linear, faz-se necessário acompanhar as melhorias nesse tipo de atendimento de serviço básico.

Para melhor entendimento de um sistema de filas e seus componentes pode-se visualizar a Figura 3.1 a seguir.

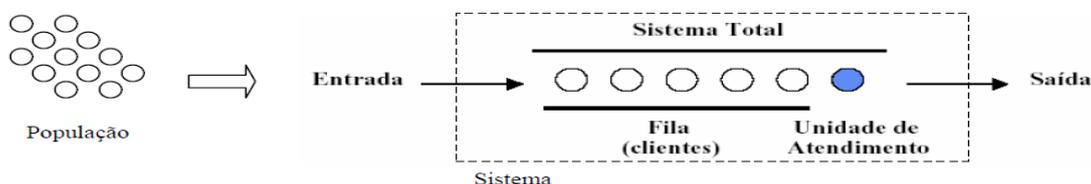


Figura 3.1 - Exemplo de fila com seus componentes

Fonte: Introdução à pesquisa operacional

Se considerarmos um problema de filas onde a taxa de atendimento dos clientes seja maior que a taxa de chegada dos clientes ($\mu \geq \lambda$), poderíamos concluir erroneamente que não existe fila no sistema. Porém, isso só seria correto se os processos de chegada e de atendimento fossem regulares. No entanto, nos problemas reais, as chegadas são irregulares e aleatórias. A abordagem matemática da teoria de filas exige que exista estabilidade no sistema e devem-se manter constantes no decorrer do tempo, caso contrário, deverá utilizar técnicas de simulação por computador para realizar tais análises.

3.2 Modelo Hipercubo de Filas

O modelo hipercubo de filas, desenvolvido por LARSON (1974), tem por essência a Teoria de Filas e baseia-se na teoria de Markov em tempo contínuo. Inicialmente foi aplicado para resolver um problema de patrulhamento policial e sua aplicabilidade se adéqua em situações onde servidores se deslocam para oferecer algum tipo de serviço (*server-to-customer*). Essa poderosa ferramenta pode ser muito bem utilizada em problemas relacionados a diversas situações, como por exemplo, corpo bombeiros (COSTA, 2003), serviços de urgência e emergência (SOUZA, 2010), defesa civil, reparos em redes de infraestruturas básicas (água, energia elétrica e telecomunicações) (ALBINO,1994), guinchos e entregas dentre outros.

É um modelo basicamente descritivo e considera toda a complexidade geográfica e temporal do sistema. Baseia-se na partição da região de atuação do sistema em um conjunto finito de áreas geradoras de demanda, denominadas átomos geográficos (MENDONÇA e MORABITO, 2001). Cada átomo é considerado como uma fonte pontual independente de solicitação de serviço ao longo do tempo. Portanto, são consideradas as distribuições espacial e temporal dos chamados. A área de cobertura primária de um servidor é definida pelo conjunto dos átomos que este

servidor atende prioritariamente.

O modelo hipercubo de filas é preditivo, não determina uma configuração ótima para o sistema, pois, não é um modelo de otimização, mas pelo fato de calcular diversos indicadores de desempenho, poderá fornecer uma completa avaliação de desempenho de cada configuração sugerida, a cada reposicionamento dos servidores, o modelo hipercubo permite a avaliação de diferentes cenários, e com isso auxilia nas decisões operacionais do sistema.

O atendimento poderá ser realizado por diversos servidores (veículos) distribuídos ao longo de uma região. A localização dos servidores deverá ser conhecida, ao menos probabilisticamente.

Para esse modelo não basta apenas especificar a quantidade de servidores ocupados, mas deve-se conhecer a disponibilidade de cada um. Para isso, utiliza-se uma variável binária associada a cada servidor, com os valores 0 ou 1 representando os estados, livre ou ocupado, do servidor, num determinado instante. O atendimento será realizado preferencialmente pelo servidor principal, caso esteja ocupado, o atendimento será realizado pelo servidor que estiver disponível. Caso nenhum servidor possa atender, o chamado é perdido. O sistema pode ter filas ou não, pois o modelo atende perfeitamente os dois casos, neste trabalho o sistema admite fila de espera com capacidade limitada a fim de facilitar o entendimento do modelo, porém a empresa trabalha com fila ilimitada de usuários.

3.2.1 Hipóteses da aplicação do modelo

Para que o modelo hipercubo possa ser aplicado a um problema real, ele deve satisfazer a algumas hipóteses, são elas:

- a. A área deverá ser dividida em N_A átomos geográficos independentes:

Existem diversas maneiras de fazer essa divisão, entre elas, setores policiais, bairros, divisão política. Nesse trabalho, optou-se por fazer divisão por setores do Batalhão da Polícia Militar de Curitiba, que são quatro: 12º BPM, 13º BPM, 20º BPM e 23º BPM, e dentro dessa divisão devido aos tipos de atendimentos diferenciados que são os casos graves e não tão graves, cada átomo foi biparticionado, totalizando oito subátomos no sistema;

- b. As solicitações por serviço em cada átomo j ($j = 1, \dots, N_A$), devem chegar independentemente de acordo com uma distribuição de Poisson. O que geralmente é verificado neste tipo de serviço de atendimento.

É necessário verificar estatisticamente se o processo de chegada dos chamados em cada átomo j é independente, com taxas λ_j , possíveis de serem medidas ou estimadas. Essa hipótese foi comprovada através do teste de Qui-Quadrado e Kolmogorov-Smirnov, realizou-se um teste de aderência na amostra e não se pode rejeitar a hipótese de que os intervalos entre chegadas sucessivas ao sistema têm distribuição exponencial, portanto, foi admitido que o número de chegadas por intervalo de tempo tem distribuição de Poisson.

Através do número de chegada de chamado em cada átomo foram calculadas as proporções dos chamados ρ_j que representam as probabilidades de um chamado qualquer ter origem num átomo j . O intervalo analisado para os sete dias da amostra foi entre as 10 e 14 horas, horário que apresentou maior taxa de chegada dos chamados. Com isso considerando as devidas proporções que representam as chegadas em cada átomo pode-se admitir que os chamados cheguem em cada átomo segundo processo de Poisson, com taxas médias $\lambda_j = \lambda \times \rho_j$ ($j = 1, 2, \dots, 8$), conforme Tabela 3.1.

Átomos	n° chamados	ρ_j	λ_j
12A	21	0,088608	0,5316
12B	54	0,227848	1,3671
13A	17	0,07173	0,4304
13B	37	0,156118	0,9367
20A	26	0,109705	0,6582
20B	64	0,270042	1,6202
23A	6	0,025316	0,1519
23B	12	0,050633	0,3038
Total	237	1	5,99994

Tabela 2.1 Proporção de chamados e taxas médias dos chamados (por hora) em cada subátomo

- c. Os tempos de viagem do átomo i para o átomo j ($i, j = 1, \dots, N_A$) devem ser conhecidos ou estimados.

No caso deste trabalho, os tempos de viagem foram fornecidos pelo próprio sistema conforme Tabela 3.2.

Átomo	12A	12B	13A	13B	20A	20B	23A	23B	Média
12A	15,1	15,1	17,3	17,3	19,6	19,6	16,1	16,1	17,025
12B	15,1	15,1	17,3	17,3	19,6	19,6	16,1	16,1	17,025
13A	17,3	17,3	14,4	14,4	16,1	16	14	14	15,4375
13B	17,3	17,3	14,4	14,4	16	16	14	14	15,425
20A	19,6	19,6	16,1	16,1	17,5	17,5	20,7	20,7	18,475
20B	19,6	19,6	16,1	16,1	17,5	17,5	20,7	20,7	18,475
23A	16,1	16,1	14	14	14	14	16	16	15,025
23B	16,1	16,1	14	14	14	14	16	16	15,025

Tabela 3.2 Tempos médio de viagem entre subátomos obtidos a partir de dados da empresa

Cabe ressaltar que os tempos de viagem dentro de um mesmo átomo são diferentes de zero, conforme se pode observar na tabela 3.2, pois os átomos possuem dimensões consideráveis.

- d. O sistema opera com servidores espacialmente distribuídos, que podem se deslocar e atender a qualquer um dos átomos.

Nesse caso, o sistema opera com dez veículos distribuídos em seis dos oito átomos, exceto nos átomos 23A e 23B. Os veículos podem atender a qualquer átomo e a qualquer tipo de chamado.

- e. A localização dos servidores deve ser conhecida, ao menos probabilisticamente.

A representação destas localizações dos veículos se dá de forma matricial $L = [l_{ij}]$, onde os elementos da matriz representam a probabilidade de um veículo estar localizado em determinado átomo no momento em que se encontrar disponível. No caso da empresa, essa localização tornou-se bem simples, pois, os veículos possuem bases definidas em seus átomos. A fim de facilitar a nomenclatura, os veículos 1, 2, 3, 4 e 5 são os veículos que atendem prioritariamente casos graves, os veículos 6, 7, 8, 9 e 10 atendem os casos não tão graves.

	12A	12B	13A	13B	20A	20B	23A	23B
1	0	0	0,5	0,5	0	0	0	0
2	0	0	0	0	0,5	0,5	0	0
3	0	0	0	0	0,5	0,5	0	0
4	0,5	0,5	0	0	0	0	0	0
5	0,5	0,5	0	0	0	0	0	0
6	0,5	0,5	0	0	0	0	0	0
7	0,5	0,5	0	0	0	0	0	0
8	0,5	0,5	0	0	0	0	0	0
9	0,5	0,5	0	0	0	0	0	0
10	0,5	0,5	0	0	0	0	0	0

Tabela 3.3 Matriz de localização dos veículos

- f. Apenas um servidor é despachado para atender um chamado.

Essa é a política de despacho da empresa, apenas um veículo é despachado para o atendimento, no caso de todos os servidores encontrarem-se ocupados, o chamado entrará em uma fila de espera.

- g. Há uma lista de preferências de despacho de servidores para cada átomo.

Nesse caso em particular, como existem chamados graves e não tão graves, os veículos preparados com UTIs, que são os veículos 1, 2, 3, 4 e 5, são os primeiros a serem enviados num chamado grave (quando disponíveis), os demais veículos servem como *backup*. O mesmo procedimento serve para os outros veículos, são os primeiros a serem chamados para casos não tão graves e os veículos com UTIs são seus *backups*. A lista de preferência de despacho foi gerada de maneira aleatória, mas respeitando essas particularidades do sistema. A Tabela 3.4 apresenta a matriz de preferência de despacho, onde as linhas representam os átomos geradores

dos chamados.

	1 ^a	2 ^a	3 ^a	4 ^a	5 ^a	6 ^a	7 ^a	8 ^a	9 ^a	10 ^a
12A	3	4	0	1	2	5	6	7	8	9
12B	5	6	7	8	9	3	4	0	1	2
13A	0	3	4	2	1	6	7	8	9	5
13B	6	7	8	9	5	0	3	4	2	1
20A	1	2	3	4	0	7	8	9	5	6
20B	7	8	9	5	6	1	2	3	4	0
23A	4	1	2	0	3	8	9	5	6	7
23B	8	9	5	6	7	4	1	2	0	3

Tabela 3.4 Matriz preferência de despacho

- h. Tempo total de atendimento de um chamado é composto pelo somatório dos seguintes tempos: tempo de preparo do servidor (*setup time*), tempo de viagem do servidor até o local da ocorrência, tempo de execução do serviço junto ao usuário (tempo em cena) e o tempo de retorno à base. A tabela 3.5 apresenta os veículos bem como as respectivas taxas médias de atendimento.

Veículos	T	μ (por hora)	Veículos	T	μ (por hora)
1	63,9757	1,0663	6	35,9886	0,5998
2	40,2614	0,6710	7	42,5160	0,7086
3	56,6900	0,9448	8	39,9433	0,6657
4	48,2800	0,8047	9	49,1000	0,8183
5	39,1671	0,6528	10	37,9633	0,6327

Tabela 3.3 Tempos e taxas médias para cada veículo

- i. Dependência de atendimento em relação ao tempo de viagem:

Faz-se necessário verificar se os tempos médios de viagem são pequenos em relação ao tempo de viagem. Essa condição é normalmente atendida nesse tipo de serviço em áreas urbanas. No estudo em questão essa condição foi satisfeita conforme Tabela 3.6.

Veículos	T médio serviço	T médio de viagem	T viagem:Tserviço
1	63,9757	10,6	0,165
2	40,2614	16,1	0,399
3	56,6900	10,9	0,193
4	48,2800	6,8	0,140
5	39,1671	13,4	0,341
6	35,9886	5,0	0,139
7	42,5160	14,2	0,333
8	39,9433	15,0	0,376
9	49,1000	11,1	0,226
10	37,9633	11,0	0,290

Tabela 3.6 Relação do tempo de viagem com o tempo de serviço

Na prática, poucos são os sistemas reais que aderem completamente a todas as hipóteses mencionadas acima. Nesse momento deverá ser feita uma análise da complexidade do problema em questão para verificar a aplicabilidade do modelo.

3.2.2 Transições entre estados e equações de equilíbrio

O modelo hipercubo baseia-se na resolução de um sistema de equações lineares. São essas equações que fornecem as probabilidades de equilíbrio dos possíveis estados do sistema e automaticamente as medidas de desempenho. Porém, com o aumento do número de servidores há um aumento exponencial de estados do modelo. O cálculo das medidas de desempenho para um sistema que opera com N servidores e não permite filas envolve uma solução de um sistema de 2^N equações lineares, ou seja, quanto maior o número de servidores, maior o número de equações e com isso o tempo computacional poderá tornar-se inviável.

A probabilidade de que cheguem ao sistema dois chamados simultaneamente é considerada nula, também é considerada nula a probabilidade de dois servidores ficarem livres simultaneamente. Com isso, em qualquer transição é permitida apenas um passo por vez.

As taxas de transições de um estado para outro são chamadas de ascendentes ou

descendentes. A solução do modelo é dada pela construção das equações de equilíbrio do sistema, que são definidas supondo-se que o sistema atinja o equilíbrio, o fluxo que entra neste estado deve ser igual ao fluxo que sai deste estado. Em um sistema não saturado, com capacidade de fila infinita, as probabilidades de estado do modelo hipercubo são calculadas a partir das equações de balanço, construídas a partir dos possíveis estados.

Num exemplo ilustrativo com apenas três servidores teríamos oito possíveis estados.

Neste estudo de caso temos oito átomos e 10 veículos, o que nos fornecerá um número de equações lineares igual a 2^{10} , ou seja, 1024 equações, isto sem considerarmos a fila de espera, ou seja, uma para cada estado, cujas incógnitas são suas probabilidades de equilíbrio do sistema. Esse é um sistema indeterminado e uma maneira de eliminar esta indeterminação é introduzindo uma equação de normalização, ou seja, considerando que a soma das probabilidades de todos os possíveis estados do sistema deve ser igual a um.

$$p_{00...0} + p_{00...1} + p_{00...10} + \dots + p_{1...11} = 1 - p_Q$$

O sistema de equações poderá ser resolvido por vários métodos, nesse trabalho, o sistema de equações que representam os possíveis estados, foi solucionado pelo método Gauss.

4. Implementação computacional e análise dos resultados

O modelo foi implementado computacionalmente em linguagem Visual Basic e executado em um computador com processador Intel Core I5-2400 CPU 3.10GHz, memória RAM 8GB e sistema operacional Windows 7 de 64 Bits.

O modelo hipercubo de filas com prioridades mostrou-se muito eficiente para analisar as medidas de desempenho da empresa. Forneceu a probabilidade de encontrar o sistema vazio ($p_{0000000000}$), sem que nenhum veículo esteja ocupado, igual a 0,01229, ou seja, 1,2%. A probabilidade de encontrar o sistema saturado, com usuários aguardando para serem atendidos é de aproximadamente 0,072, isto é 7,2% de probabilidade.

A seguir serão apresentadas algumas medidas importantes de desempenho fornecidas pelo modelo:

- a. *Workload*: carga média de trabalho de cada veículo representa a fração de tempo que um veículo permanece ocupado. É obtida através das somas das probabilidades dos estados em que o veículo está ocupado.

$$p_1 = p_{0000000001} + p_{0000000010} + \dots + p_{1111111111} + p_Q$$

$$p_2 = p_{0000000010} + p_{0000000011} + \dots + p_{1111111111} + p_Q$$

⋮

$$p_n = p_{1000000000} + p_{1000000001} + \dots + p_{1111111111} + p_Q$$

Onde p_Q é a probabilidade de formação de fila e p_n é a soma das probabilidades do servidor n estar ocupado.

Essa medida é muito importante para avaliar o quanto cada veículo está sendo solicitado ao longo do dia, o que possibilita a elaboração de propostas de melhor distribuição dos veículos.

Os resultados obtidos através do modelo e na amostra encontram-se na Tabela 4.1, onde o VSG refere-se aos veículos que atendem serviços graves, e VSL são veículos que atendem serviços leves.

Veículos	Modelo	Amostra	Desv.Pad
1	0,22	0,27	0,03
2	0,36	0,17	0,14
3	0,13	0,24	0,08
4	0,31	0,20	0,08
5	0,16	0,16	0,00
6	0,48	0,15	0,23
7	0,55	0,18	0,26
8	0,67	0,17	0,36
9	0,46	0,20	0,18
10	0,13	0,16	0,02
VSG	0,24	0,21	0,07
VSL	0,45	0,17	0,22

Tabela 4.1 *Workload* dos veículos do modelo e da amostra

Pode-se verificar que alguns veículos apresentam altos valores nos desvios, isso pode ser devido à pequena quantidade de dados obtidos para esses veículos.

- b. A frequência de despacho f_{ij} , apresentada na equação abaixo, é a fração de todos os despachos do servidor i ao átomo j e pode ser decomposta em duas partes: $f_{ij}^{[nq]}$, que corresponde à fração de despachos de um servidor i para um átomo j que não implica tempo de espera em fila para o usuário; e $f_{ij}^{[q]}$, que corresponde à fração de despachos de um servidor i para um átomo j sujeito a espera em fila.

$$f_{ij} = f_{ij}^{[nq]} + f_{ij}^{[q]} = \frac{\lambda_j}{\lambda} \sum_{B \in E_{ij}} P_B + \frac{\lambda_j}{\lambda} P_Q \frac{\mu_i}{\mu}$$

Quanto à frequência de despachos interátomos, o modelo forneceu um valor de 0,48405 para a fração dentre todos os despachos que são interáreas de cobertura.

Os despachos interáreas de cada um dos veículos e a fração dos chamados da área N que é atendida por outros veículos que não pertencem à mesma área encontram-se na Tabela 4.2.

Veículos	F. Interareas	F. Int. outro veículo	Veículos	F. Interareas	F. Int. outro veículo
1	0,126	0,219	6	0,095	0,474
2	0,089	0,361	7	0,397	0,549
3	1,000	0,000	8	0,344	0,666
4	0,241	0,308	9	0,834	0,463
5	0,579	0,157	10	1,000	0,000

Tabela 4.2 Fração dos despachos de cada veículo

- c. Tempos médios de viagem são obtidos através da matriz dos tempos de viagem no próprio sistema. Porém os tempos de viagem podem sofrer influência de diversos fatores no dia-a-dia, portanto o tempo deslocamento de um átomo i até um átomo j não será necessariamente igual ao caminho inverso.

O tempo médio de viagem é obtido através da equação:

$$\bar{T} = \sum_{i=1}^m \sum_{j=1}^{NA} f_{ij}^{[nq]} t_{ij} + P_Q \bar{T}_Q$$

Onde: t_{ij} , é o tempo médio de viagem do servidor i ao átomo j ;

$$\bar{T}_Q = \sum_{p=1}^{NA} \sum_{j=1}^{NA} \frac{\lambda_p \lambda_j}{\lambda^2} t_{pji}, \text{ é o tempo médio de viagem para cada chamado em fila.}$$

O modelo forneceu um tempo médio de deslocamentos para chamadas na fila de espera em minutos de 17,1265 e o tempo médio global de deslocamento é de 15,74.

A Tabela 4.3 mostra os resultados obtidos no modelo dos tempos médios aproximado de viagem para cada um dos dez veículos do sistema (T_1) e na próxima coluna o tempo médio de deslocamentos para área de cobertura primária (T_2).

Veículos	T_1	T_2	Veículos	T_1	T_2
1	14,68	15,86	6	15,68	15,02
2	17,62	17,97	7	15,66	14,52
3	17,58	0,000	8	16,40	14,82
4	23,5	20,97	9	16,32	13,00
5	15,87	16,69	10	9,09	0,000

Tabela 4.3 – Tempo médio de viagem dos veículos

O tempo médio de viagem de cada veículo ao átomo é outro indicador importante, pois reflete o nível de serviço oferecido pelo sistema.

$$\bar{T}_j = \frac{\sum_{i=1}^m \sum_{k \in D} f_{i,jk}^{[nq]} t_{i,jk}^{[nq]}}{\sum_{i=1}^m \sum_{k \in D} f_{ij}^{[nq]}} (1 - P_Q) + \frac{\sum_{i=1}^m \sum_{k \in D} f_{i,jk}^{[q]} t_{i,jk}^{[q]}}{\sum_{i=1}^m \sum_{k \in D} f_{i,jk}^{[q]}} P_Q$$

A Tabela 4.4 apresenta os tempos médios de viagem de cada veículo ao átomo do problema abordado no modelo e na amostra, apresentando um pequeno desvio padrão total de 0,26.

Átomos	Modelo	Amostra	Desv.Pad	Átomos	Modelo	Amostra	Desv.Pad
1	20,98	17,025	2,79	5	17,97	18,475	0,35
2	15,03	17,025	1,41	6	14,83	18,475	2,58
3	15,86	15,437	0,29	7	16,69	15,025	1,17
4	14,53	15,425	0,63	8	13,00	15,025	1,43

Tabela 4.4 Tempos médios de viagem de cada veículo para cada átomo

O modelo hipercubo de filas constitui uma ferramenta adequada ao problema, uma vez que os desvios foram pequenos quando comparados a algumas imprecisões dos dados fornecidos pelo sistema.

5. Conclusões

O objetivo desse trabalho foi analisar as configurações dos sistemas de atendimento médicos móvel de uma empresa de urgência e emergência de Curitiba-PR, a partir do modelo hipercubo de filas com prioridades. A região foi dividida em quatro átomos segundo o critério do BPM de Curitiba, dentro de cada átomo houve uma subdivisão de acordo com as prioridades da empresa, atendimentos de casos graves e não tão graves, cabe ressaltar que essa prioridade foi atendida perfeitamente pelo modelo hipercubo de filas. Após a coleta e análise de dados foram calculadas suas medidas de desempenho. Algumas medidas obtidas pelo modelo puderam ser comparadas com os resultados obtidos no sistema. Um exemplo foi o tempo médio global de deslocamento dos veículos que no modelo apresentou um valor igual a 15,74 minutos, enquanto que no sistema este resultado é de 16,48 minutos, o que caracteriza um desvio padrão de 0,52.

Sugere-se para trabalhos futuros a união desse modelo descritivo com modelos de otimização a fim de se criar cenários alternativos que possam se analisados a partir do Modelo Hipercubo de Filas, gerando cenários que reduzam os tempos de atendimento, fator primordial neste tipo de problema.

Referências

- Albino, J. C. C.**, (1994), Quantificação e localização de unidades móveis de atendimento de emergência a interrupções em redes de distribuição de energia elétrica: Aplicação do Modelo Hipercubo, (mestrado em engenharia da produção) UFSCar.
- Costa D. M. B.**, (2004) Uma metodologia iterativa para determinação de zonas de atendimento de serviços emergenciais. Tese (doutorado em engenharia de produção) - Departamento de Engenharia de Produção, UFSCar.
- Larson, R.C.**, (1974) Hypercube queuing model for facility location and redistricting in urban emergency services. *Computers and operations research* 1, 67-95.
- 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.
- Morabito, R., Chiyoshi, F., Galvão, R. D.**, (2001), Modelo Hipercubo: Análises e resultados para o caso de servidores não homogêneos. *Pesquisa Operacional*, v21, n.2, p 199-218.
- Morabito, R., Iannoni, A. P.**, (2008), Otimização da localização das bases de ambulâncias e do dimensionamento das suas regiões de cobertura em rodovias. *Produção* V 18, p.47-63.
- Souza, R. M.**, (2010), Análise da configuração de SAMU utilizando Modelo Hipercubo de filas com prioridades e múltiplas alternativas de localização de ambulâncias. Tese (doutorado em engenharia de produção) – UFSC.
- Takeda, R. A.**, (2000) Uma contribuição para avaliar o desempenho de sistemas de transporte emergencial de saúde. Universidade de São Paulo. Tese (doutorado em Transportes) – UFSC.