ANÁLISE DE CENÁRIOS EM UM CALL CENTER DE UMA OPERADORA DE PLANO DE SAÚDE UTILIZANDO SIMULAÇÃO

Victor Abu-Marrul Carneiro da Cunha
Pontifícia Universidade Católica do Rio de Janeiro (PUC-RIO)
Departamento de Engenharia Industrial (DEI)
R. Marquês de São Vicente, 225 – Gávea, Rio de Janeiro – RJ, 22451-900
victorabu@gmail.com

Fernando Luiz Cyrino Oliveira
Pontifícia Universidade Católica do Rio de Janeiro (PUC-RIO)
Departamento de Engenharia Industrial (DEI)
R. Marquês de São Vicente, 225 – Gávea, Rio de Janeiro – RJ, 22451-900
cyrino@puc-rio.br

RESUMO

Os *call centers* estão presentes nos mais diversos ramos de negócios, desempenhando papel significativo quanto a satisfação dos clientes de uma companhia. Importantes indicadores como as taxas de abandono, tempos de espera, média de pessoas na fila entre outros, são gerenciados em busca de aperfeiçoamentos. Dessa forma, o presente trabalho visa analisar tais indicadores com o uso da simulação de eventos discretos, através da modelagem de um *call center* real de uma operadora de plano de saúde. O objetivo é descobrir possíveis soluções para a alocação dos atendentes ou para o roteamento das chamadas, com o intuito de propor melhorias aplicáveis à companhia estudada. Foram analisados quatro canais de atendimento agrupados em dois núcleos, com diferentes distribuições de probabilidade associadas à tolerância na espera, por perfil de cliente e aos tempos de atendimento de cada canal. Diversos cenários foram avaliados, a fim de constatar aqueles que apresentavam melhores resultados.

PALAVARAS CHAVE. Central de atendimento, Simulação de eventos discretos, Nível de serviço.

Tópicos: PO em serviços; Simulação; PO na Área de Saúde

ABSTRACT

Call centers are present in various lines of business and play a significant role in customer satisfaction. Important indicators such as abandon rates, waiting times, average number at queue, and others are used in search of improvements. Thus, this study aims to analyze these indicators using the discrete event simulation by modeling a real system of a call center of a health plan operator. The goal is to find possible solutions to the allocation of attendants or the call routings, in order to propose improvements for the studied company. We analyzed four service channels grouped into two pools with different probability distributions of the impatience of the customers and service times for each channel. Several scenarios were evaluated in order to find those with the best results.

KEYWORDS. Call center, Discrete Event Simulation, Service level.

Paper Topics: Or in Services; Simulation; OR in Health

1. Introdução

Um *call center* é um conjunto de recursos (equipamentos de telecomunicação, computadores, funcionários, entre outros), que entrega serviços via telefone aos seus clientes. (Gans et al., 2003). Em geral, o sistema é composto por um número variável de atendentes, que realizam esses serviços, onde os clientes estão virtualmente presentes, sejam sendo atendidos ou aguardando em uma fila telefônica. Um cliente na fila, aguarda até que exista pelo menos um atendente disponível, acumulando impaciência, de forma que o mesmo pode acabar optando por abandonar a fila. (Mandelbaum, 2006)

Em alguns casos, após o contato, os clientes comunicam suas necessidades com a unidade de resposta audível (URA), para que em seguida, sejam então direcionados a um atendente disponível através do distribuidor automático de chamadas (DAC). Os DACs modernos podem ser programados para efetuarem o roteamento das chamadas baseados em inúmeros critérios. (Gans et al., 2003)

Essa distribuição de chamadas é feita baseada na estrutura de roteamento estipulada em cada *call center*. As arquiteturas podem ser do tipo totalmente flexível, onde todos os atendentes são capazes de atender todos os tipos de chamada ou totalmente dedicada, onde cada grupo de atendentes recebe um tipo específico de chamada. Em geral quando os tipos de chamada requerem tempos de duração similares, o modelo totalmente flexível demanda um menor número de funcionários para atingir um nível de serviço pré-definido. (Legros et al., 2015)

Em muitos casos, os diferentes tipos de chamada demandam múltiplas especialidades, tornando inviável o treinamento de atendentes para receberem todos os tipos requeridos. Um exemplo disso é a demanda por *call centers* que atendem em diversas línguas diferentes, devido a globalização, o que torna inviável encontrar atendentes que falem todas as línguas. Porém, podem existir combinações entre as especialidades de cada um, encontrando agentes que falem duas ou três línguas, encontrando a melhor forma de agrupá-los. (Wallace e Whitt, 2005)

Além dos modelos extremos, existem diversos modelos intermediários, alguns deles apresentados por (Garnett e Mandelbaum, 2000), conforme mostra a figura 1.

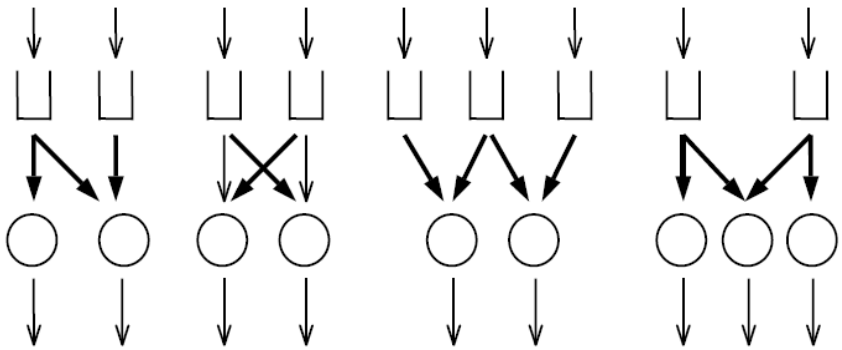

Figura 1: Modelos N, X, W e M (Garnet e Mandelbaum, 2000)

A simulação pode ser usada para tratar esse e outros problemas no gerenciamento de um *call center*. De acordo com Mehrotra e Fama (2003), existem três motivos principais pelos quais a simulação é utilizada em um *call center*, que são: a análise típica usando simulação, onde se avalia uma ou mais operações específicas, a aplicação em problemas de roteamento de chamadas e os problemas de planejamento da escala de trabalho dos atendentes.

Conforme aponta Dijik e Sluis (2008), a simulação de eventos discretos é conhecidamente a ferramenta mais poderosa para a avaliação da manufatura, comunicações e indústria de serviços, como bancos, *call centers*, hospitais entre outros, e destaca que em um modelo de *call center*, a simulação consegue representar uma das mais importantes características presentes no sistema, o abandono de chamadas, avaliando a performance do sistema quanto as taxas de abandono e as medidas de desempenho do serviço.

O básico em um modelo de simulação de *call center*, são as informações das chamadas, dos atendentes e o tempo de funcionamento. O roteamento lógico das chamas é o responsável por conectar a chegada das chamadas com os atendentes durante o tempo de funcionamento estipulado. (Mehrotra e Fama, 2003)

O modelo analítico clássico da teoria de filas frequentemente utilizado para esses casos é o modelo Erlang-C ou M/M/N, no qual assumem-se as chegadas por um processo Poisson, atendimentos distribuídos exponencialmente e desconsidera-se o abandono de chamadas. Porém o abandono não é um aspecto menor e não deve ser negligenciado. (Garnett, et al., 2008)

Ainda de acordo com Garnett et al. (2008), o grande problema dos modelos que ignoram o abandono, é que os mesmos podem distorcer ou falhar na entrega de resultados importantes para o gerenciamento. É importante que o abandono seja considerado na avaliação do nível de serviço prestado aos clientes.

Em geral, os gerentes estão interessados em alguns indicadores de desempenho, como o nível de serviço, definido pelo número de clientes que esperaram menos do que um tempo determinado, as taxas de abandono, o tempo de espera, entre outros. (Avramidis, 2005)

Baseado nisso, o presente trabalho pretende desenvolver um modelo de simulação de um *call center*, que leve em consideração o abandono, a fim de obter medidas de desempenho de foco gerencial, alinhadas com as medidas utilizadas pela empresa estudada.

O artigo está estruturado em 6 seções: após a introdução já apresentada na seção 1, a seção 2 descreve o problema de forma geral, indicando todas as características do problema e os objetivos a serem alcançados com o presente trabalho. A seção 3 detalha os dados fornecidos pela companhia e provê uma análise estatística desses dados a fim de entender o cenário atual da empresa. Na seção 4 é apresentado o modelo de simulação desenvolvido para o tratamento do problema, além de demonstrar resultados preliminares obtidos com testes no modelo com o intuito de validá-lo. Na seção 5 se dá a aplicação do modelo ao caso real, a fim de analisar mudanças na configuração do *call center*, avaliando quais delas podem trazer benefícios à companhia caso venham a ser aplicadas. Por fim a seção 6 apresenta uma conclusão geral sobre o trabalho com o resultados obtidos e propostas para trabalhos futuros.

2. Descrição do Problema e objetivos

O problema estudado é referente a um serviço de *call center* receptivo de uma operadora de plano de saúde situada na cidade do Rio de Janeiro. As chamadas recebidas são classificadas em 6 diferentes tipos, que são: agendamento de consulta, informação, atendimento de empresas, autorizações, atendimento a médicos e intercâmbio, sendo o último deles um canal exclusivo para clientes vinculados a mesma operadora em outras cidades. Para alguns tipos, os atendimentos são consolidados em núcleos, onde alguns atendentes podem receber até dois tipos de chamada. Atualmente existem dois núcleos que agrupam mais de um tipo de chamada, o núcleo que atende os canais de informação e empresas e o outro que atende os canais de autorização e médicos. Os atendimentos de informação e autorização são os de maior demanda e menor complexidade, desta forma, todo novo funcionário de um desses núcleos, é treinado em uma dessas funções de baixa complexidade e só passam a atender os dois tipos de chamada quando ocorre alguma baixa no número de atendentes estipulados pela companhia para realizar tal função.

Os núcleos estudados serão apenas os núcleos combinados, ou seja, aqueles que tratam de dois tipos de chamada, devido a disponibilidade de dados fornecidos pela empresa. Estes núcleos, por possuírem maior volume de chamadas, são também os de mais difícil gestão, muitas vezes não conseguindo atingir as metas estipuladas pela alta direção. Esses núcleos trabalham com modelos do tipo N, onde os atendimentos de um dos tipos de chamada podem ser direcionados pra qualquer atendente, enquanto as chamadas de tipo mais complexo só são atendidos por um grupo específico de funcionários, conforme mostra a figura 2.

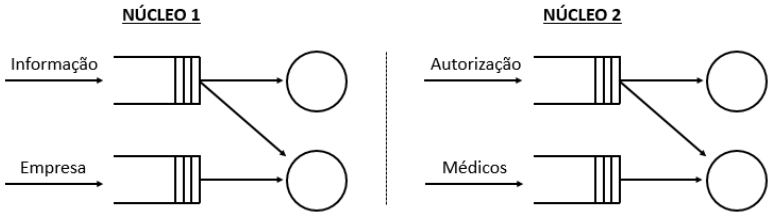

Figura 2: Núcleos de atendimento com modelos do tipo N

Todos os funcionários têm vínculo empregatício com a empresa estudada, sem terceirização de mão-de-obra, o que torna mais fácil a aplicação de possíveis mudanças nas configurações entre núcleos de atendimento e alterações nas políticas de roteamento das chamadas recebidas.

Todas as chamadas passam pela URA, onde a partir da necessidade de cada cliente, são direcionadas a um dos canais existentes. Caso o canal possua atendentes disponíveis no momento em que o cliente é direcionado, o mesmo é atendido automaticamente sem a necessidade de esperar em uma fila, porém, caso o canal ou canais aptos a atendê-lo estejam com todos os atendentes ocupados, o mesmo é colocado em uma fila de espera exclusiva para cada tipo de chamada e aguarda até que um funcionário fique disponível para atendê-lo, contudo os clientes possuem uma tolerância quanto ao tempo de espera, podendo abandonar a chamada, desistindo do atendimento. Os tempos referentes a URA foram desconsiderados no trabalho, dado que a empresa considera o tempo de espera a partir do momento no qual um determinado cliente é colocado na fila referente a sua necessidade.

O *call center* funciona durante 24 horas por dia, 7 dias da semana e possui turnos diferenciados entre os horários e dias da semana. Os atendentes trabalham em turnos de 6 horas e têm 40 minutos de intervalo. Porém, devido à falta de acesso a escala de trabalho dos atendentes definida pela companhia, no trabalho foram considerados números médios de atendentes disponíveis ao longo do dia para cada tipo de chamada. Números esses informados pela própria empresa, desconsiderando o período de intervalo.

Como medida de desempenho, a companhia adota um indicador que avalia o tempo de espera e o abandono de forma conjunta. Chamado pela própria empresa de "nível de serviço", esse indicador de desempenho avalia o percentual de clientes atendidos que não esperaram mais do que 2 minutos. A meta estipulada pela companhia é de que pelo menos 80% dos clientes estejam dentro desse critério. Os clientes que abandonaram a fila, mesmo com tempos de espera inferiores aos 2 minutos estipulados, são contabilizados de forma negativa, dado que o percentual de clientes atendidos abaixo do tempo limite é avaliado em cima do total de clientes que entraram no sistema. Atualmente os dois núcleos a serem avaliados não conseguem cumprir a meta proposta, apesar de inúmeras campanhas motivacionais e institucionais realizadas pela companhia. Desta forma, a simulação pode ser uma ferramenta matemática poderosa, capaz de apontar melhorias para o modelo atual através de estudo embasado nos dados históricos, de forma a propor possíveis mudanças objetivando o atingimento das metas propostas.

Em resumo, a proposta do trabalho é avaliar possíveis mudanças capazes de acarretar melhorias na avaliação de desempenho do serviço de atendimento da empresa. Após a análise dos dados de forma a entender o cenário atual, será construído um modelo de simulação com o uso do software ARENA®, realizada uma validação do modelo em comparação com os dados reais e testado em diversos cenários alternativos ao que é praticado, a fim de verificar se algum desses é capaz de gerar resultados melhores do que os atuais. Serão realizadas alterações tanto no número de atendentes disponíveis em cada canal, como no roteamento das chamadas, de forma que cada alteração ou conjunto de alteração se configura em um novo cenário a ser testado e avaliado. A geração desses cenários deve criar possibilidades reais de mudanças que possam ser aplicadas caso a avaliação aponte bons resultados.

3. Análise dos dados

Foram analisados dados de três relatórios de gestão para cada tipo de chamada, de Junho a Agosto de 2015, contendo os seguintes dados diários: número de chamadas recebidas, atendidas e abandonadas, percentual de abandono, indicador de nível de serviço, tempo médio de espera, duração média das chamadas, média dos tempos de tolerância antes do abandono e tempo máximo de espera. Os relatórios não contemplam os tempos entre chegadas, dado que são relatórios de gestão, utilizados na avaliação de desempenho do serviço de atendimento.

A primeira análise feita foi com relação aos tempos de atendimentos em comparação com os tempos de tolerância conforme mostra a figura 3. Os tempos de espera toleráveis revelaram-se em todos os tipos, menores do que os tempos de duração de um atendimento, sendo assim um potencial gerador de filas, destacando-se o fato de que as chamadas para o canal de médicos são as de menor tolerância para espera, o que pode agravar o acúmulo nas filas e afetar negativamente o nível de serviço desejado, mesmo este sendo um tipo com baixa demanda.

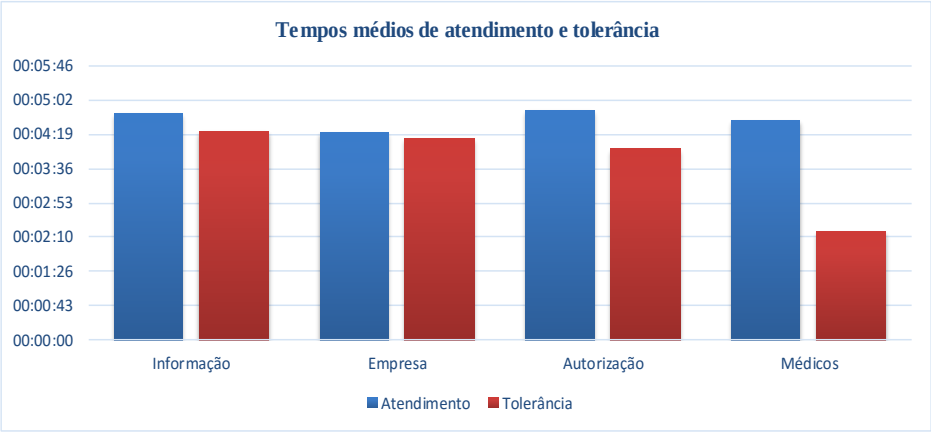

Figura 3 : Comparativo entre os tempos médios de atendimento e de tolerância por tipo de chamada

Posteriormente foram analisados os dados referentes a tolerância dos clientes, para os dados consolidados entre todos os tipos de chamada, indicando uma forte correlação entre os tempos de espera e o percentual de pessoas que abandonam as chamadas, como já era esperado, conforme mostra a figura 4.

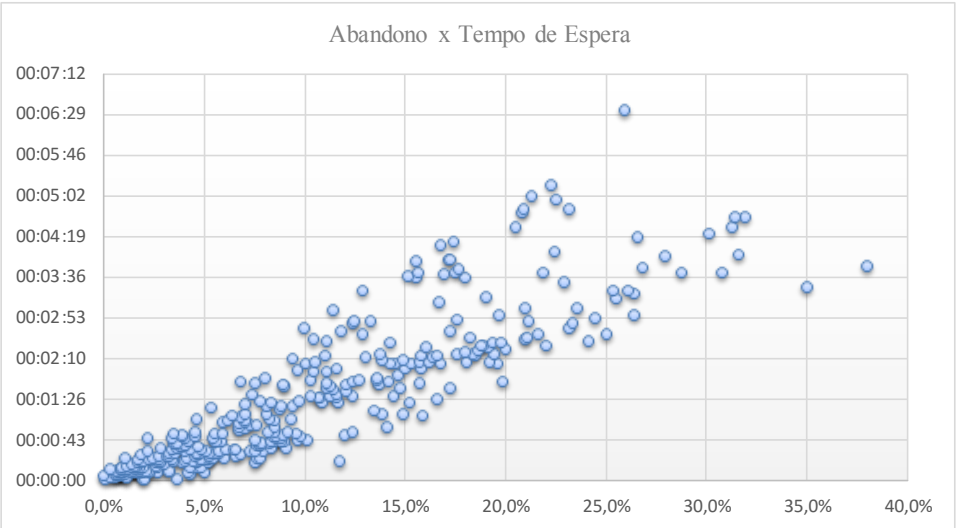

Figura 4: Correlação entre o percentual de abandonos e o tempo de espera

Devido à forte correlação entre os tempos de espera e o percentual de abandono e pela falta de informação relevante capaz de indicar outros motivos de abandono de chamada, o modelo a ser elaborado irá considerar apenas o motivo de espera como responsável pelo abandono.

Devido à baixa demanda nos finais de semana, os canais de atendimento à empresas e médicos não funcionam, enquanto nos demais canais a disponibilidade de atendentes ao longo do dia é reduzida. Em geral esses dias são analisados separadamente nos indicadores por não serem críticos de se gerenciar, quase que sempre cumprindo a meta estabelecida. O número de atendentes aptos não consta nos relatórios por ser um dado variável de difícil controle. Por esses motivos, os dados de entrada levarão em consideração apenas os dias úteis da semana.

Para obter as distribuições de probabilidade a serem utilizadas no modelo, foi feito o uso do *input analyser* do ARENA®, para o tempo médio de atendimento e os tempos médios de tolerância para cada tipo de chamada em segundos, conforme mostra a tabela 1.

Tabela 1: Distribuições para os tempos de atendimento e tolerância por canal

Canal	Atendimento	Tolerância
Informação	TRIA(210, 320, 331)	NORM(264, 113)
Empresa RH	37 + 314 * BETA(0.932, 0.516)	TRIA(87, 329, 385)
Autorização	TRIA(191, 315, 340)	61 + 476 * BETA(1.5, 2.46)
Médicos	NORM(276, 49.1)	NORM(138, 65.8)

4. Desenvolvimento e testes do modelo

O modelo desenvolvido, visa em uma primeira etapa recriar o cenário atual do *call center* da operadora, de forma que os resultados gerados pelo modelo se equiparem com o que ocorre atualmente na empresa, para que então possam ser aplicadas modificações a fim de buscar estratégias de melhoria. Seguindo as premissas estabelecidas e os dados analisados, o modelo foi construído conforme indica a figura 5.

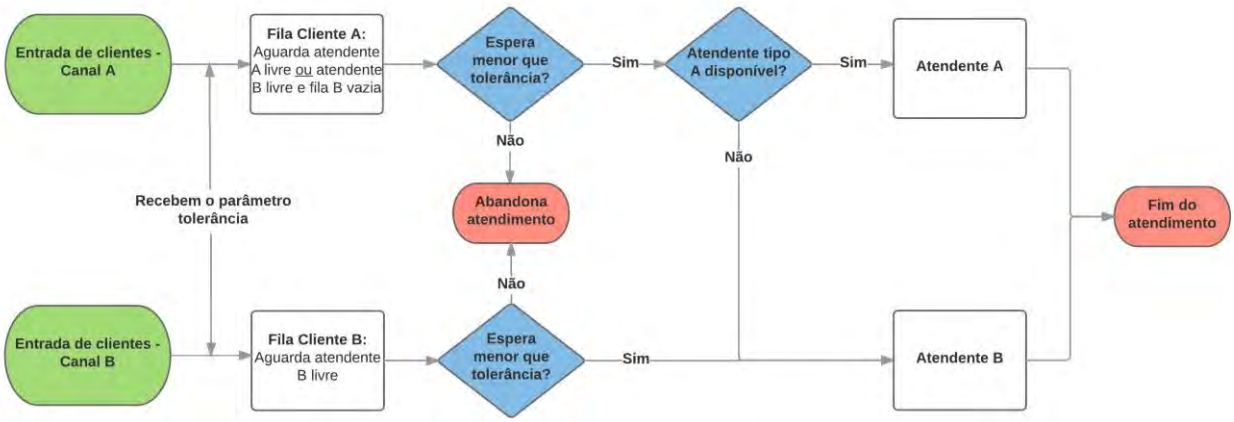

Figura 5: Modelo conceitual de um núcleo

O fluxograma desenvolvido demonstra de forma resumida a estrutura do modelo elaborado. O esquema serve tanto para o núcleo que une os canais de informação e empresas, como para o núcleo de autorização e médicos, de forma que o Canal A assume o papel dos canais de menor complexidade (informação e autorização), enquanto que o canal B assume o papel dos canais de maior complexidade (empresas e médicos). Os clientes entram cada um pelo seu canal exclusivo, e assim que o fazem recebem um tempo referente a sua tolerância. Cada cliente recebe um tempo de tolerância diferente relacionado à distribuição de probabilidade das tolerâncias para

cada tipo de chamada. Após a entrada no sistema, os clientes passam para a fila de espera, onde aguardam determinadas condições para que sejam liberados para atendimento. Os clientes do Canal A podem ser liberados mediante duas condições distintas, a primeira é a disponibilidade de um ou mais atendentes no canal A e a segunda é a conjunção de dois fatores, que são: o Canal B possuir atendentes disponíveis e a não existência de clientes do tipo B aguardando na fila. Isso se deve ao fato de que enquanto existirem clientes do tipo B aguardando na fila, a prioridade para atendimento no canal B é desses clientes. Cabe ressaltar que caso um cliente do tipo A tenha sido direcionado para um atendente do canal B, o mesmo será atendido por esse funcionário até o final, mesmo que entrem na fila um ou mais clientes do tipo B. Já os clientes do canal B, serão liberados sempre que o canal B possuir algum atendente disponível, já que o mesmo não pode ser atendido pelo canal A. Após cada cliente ser liberado para atendimento é feita uma verificação quanto ao tempo de tolerância e o tempo de espera, onde o tempo de espera deve ser menor ou igual ao tempo de tolerância para que o cliente siga para o atendimento, caso contrário o mesmo abandona o sistema. Quando um cliente do tipo A é liberado para atendimento, caso os dois canais estejam com atendentes disponíveis, o mesmo segue para o canal A. Após o atendimento concluído o cliente é liberado e deixa o sistema. Todas as filas obedecem o padrão FIFO (*First in First Out*), onde o primeiro cliente a entrar será o primeiro a ser atendido, respeitando as regras de liberação descritas anteriormente para cada uma das filas.

Seguindo o modelo real, não foi considerado limite para a fila, ou seja, nenhum cliente é impedido de entrar no sistema, mesmo que os tempos de espera estejam muito altos devido as filas geradas.

Baseando-se na teoria das filas, as chegadas foram consideradas exponencialmente distribuídas, tendo como base a média dos atendimentos nos dias úteis para cada tipo de chamada. Dessa forma foi definido um intervalo entre chegadas em segundos para cada tipo de chamada. Os tempos usados para os canais de informação, empresa, autorização e médicos, foram de 15, 81, 17.5 e 173 segundos respectivamente.

Como o *call center* trabalha durante 24 horas ininterruptas, essas mesmas 24 horas foram estipuladas como tempo para a simulação, onde para cada cenário foram feitas 10 replicações. O modelo foi testado com um número maior de replicações, porém mantendo a mesma média nos resultados gerados com 10 replicações e dada a grande quantidade de cenários a serem testados e o tempo de duração para cada rodada, optou-se pela utilização das 10 replicações. Antes de rodar o modelo final, foi verificado o impacto do tempo de aquecimento no resultado final dos quatro canais avaliados, onde mediante aplicação de diversos tempos de aquecimento em segundos, os resultados avaliados para o tempo médio de espera na fila e média de pessoas na fila permaneceram inalterados, conforme mostra a tabela 2, podendo assim ser desconsiderado do modelo.

Tabela 2: Testes no modelo variando o tempo de aquecimento

Tempo de Warm Up	**Informações**		**Empresa**		**Autorização**		**Médicos**	
	Tempo médio de espera	Média de pessoas na fila	Tempo médio de espera	Média de pessoas na fila	Tempo médio de espera	Média de pessoas na fila	Tempo médio de espera	Média de pessoas na fila
0	00:01:33	6,2	00:02:35	1,9	00:01:36	5,7	00:01:57	0,7
7200	00:01:33	6,2	00:02:35	1,9	00:01:36	5,7	00:01:57	0,7
14400	00:01:33	6,2	00:02:35	1,9	00:01:36	5,7	00:01:57	0,7
21600	00:01:33	6,2	00:02:35	1,9	00:01:36	5,7	00:01:57	0,7
28800	00:01:33	6,2	00:02:35	1,9	00:01:36	5,7	00:01:58	0,7

Com o modelo pronto, foi realizada uma rodada inicial com o número médio de atendentes, conforme informação da empresa, para comparação com os indicadores apurados dos meses analisados, a fim de verificar a precisão do modelo. No primeiro núcleo, são 17 atendentes aptos em média no canal de informações e 3 para atendimento a empresas. Já no segundo núcleo, são 14 atendentes no canal de autorização e 2 no canal exclusivo de médicos. A tabela 3, mostra o comparativo entre os dados reais extraídos das bases da empresa e os resultados gerados pelo modelo.

Tabela 3: Comparativo entre os resultados da empresa e o modelo de simulação

Canal	Indicadores Atuais (3 meses)			Resultados do Modelo		
	nível de serviço	% de abandono	Tempo médio de espera	nível de serviço	% de abandono	Tempo médio de espera
Informação	66%	14%	00:01:37	61%	10%	00:01:33
Empresa RH	51%	18%	00:02:28	39%	17%	00:02:35
Autorização	61%	11%	00:02:24	62%	11%	00:01:36
Médicos	54%	20%	00:02:10	46%	42%	00:01:59

Apesar de apresentar algumas discrepâncias, como nos níveis de serviço dos canais de empresas e médicos, a taxa de abandono do canal de médicos e o tempo médio de espera do canal de autorizações, no geral o modelo se mostrou bastante aplicável ao caso real, podendo ser usado para eventuais análises do comportamento do sistema e em como buscar melhorias. Dessa forma, partindo desses resultados, serão propostos alguns cenários para os núcleos a serem avaliados.

5. Aplicação do modelo e avaliação dos cenários

Os cenários foram aplicados separadamente em cada núcleo, onde foram realizadas trocas de funcionários entre os canais, adição de funcionários em um ou mais canais e uma mudança no roteamento das chamadas, impedindo que um tipo de chamada seja atendido pelo outro. Porém, devido a discrepância entre o nível de serviço observado no modelo e os apurados no caso real, em média 5% menores, esses 5% serão abatidos do objetivo, passando assim a meta de 80% para 75%, buscando evitar soluções com medidas drásticas de difícil implementação.

Para o núcleo de informações e empresas, os resultados dos primeiros 7 cenários testados são indicados na tabela 4.

Tabela 4: Simulação de cenários do núcleo 1

Cenário	Canal de atendimento	Atendentes	nível de serviço	% de abandono	% de abandono do cenário	Tempo médio de espera	Média de Pessoas na Fila	Uso do Recurso
Modelo tipo N								
1	Informação	17	61%	10%	11%	00:01:33	6,26	97,8%
	Empresa	3	39%	17%		00:02:35	1,93	98,7%
2	Informação	17	77%	6%	6%	00:01:06	4,35	96,1%
	Empresa	4	63%	6%		00:01:38	1,21	95,9%
3	Informação	18	77%	7%	9%	00:01:06	4,43	96,3%
	Empresa	3	38%	19%		00:02:33	1,89	97,8%
4	Informação	18	86%	4%	4%	00:00:47	3,15	94,2%
	Empresa	4	67%	5%		00:01:30	1,11	94,2%
5	Informação	16	62%	10%	9%	00:01:34	6,26	98,3%
	Empresa	4	64%	7%		00:01:44	1,39	98,3%
Modelo totalmente dedicado								
6	Informação	17	53%	12%	12%	00:01:48	7,19	98,7%
	Empresa	3	55%	14%		00:01:50	1,35	84,5%
7	Informação	18	68%	8%	7%	00:01:22	5,45	97,0%
	Empresa	4	82%	3%		00:00:46	0,57	71,8%

Avaliando os resultados, percebe-se a discrepância entre os resultados dos cenários com modelo N e com modelo totalmente dedicado, evidenciando que para o sistema estudado, o modelo N utilizado atualmente se mostrou uma alternativa melhor. Quanto aos cenários com adição de atendentes, fica evidente a melhoria acarretada quando o acréscimo ocorre no canal empresa, impactando de forma positiva nos dois canais, enquanto que o aumento no canal informações, gerou benefícios apenas para o próprio canal. Dessa forma, a estratégia de trocar um atendente de canal só faz sentido, quando um atendente do canal informações é transferido para o canal empresa. Ao efetuar essa troca, mantendo a disponibilidade média de 20 atendentes, a

melhora apresentada foi significativa tanto no canal de empresa, onde o nível de serviço passou de 39% para 64%, quanto no resultado geral do percentual de abandono do núcleo, além de gerar uma pequena melhora quanto ao nível de serviço do canal informações. Outra vantagem na troca foi um melhor balanceamento no uso dos recursos e a diminuição dos tempos de espera e média de pessoas na fila no canal empresa. Partindo desses resultados, novos cenários foram gerados, sempre transferindo recursos do canal informações para o canal empresa, conforme a tabela 5.

Tabela 5: Segunda rodada de simulação de cenários do núcleo 1

Cenário	Canal de atendimento	Atendentes	nível de serviço	% de abandono	% de abandono do cenário	Tempo médio de espera	Média de Pessoas na Fila	Uso do Recurso
Modelo tipo N								
8	Informação	0	91%	3%	3%	00:00:32	2,12	0,0%
	Empresa	20	96%	0%		00:00:06	0,07	89,9%
9	Informação	15	60%	11%	9%	00:01:36	6,44	98,3%
	Empresa	5	78%	2%		00:01:11	0,87	97,6%
10	Informação	10	75%	7%	7%	00:01:09	4,65	97,9%
	Empresa	10	99%	0%		00:00:23	0,28	94,8%

O primeiro cenário testado, foi o cenário extremo, considerando a possibilidade de ter todos os atendentes no canal empresa a fim de validar se a troca sempre acarretaria melhores resultados. Fica evidente que para esse núcleo essa troca sempre traz benefícios, porém pode ser inviável para a companhia, ter todos os atendentes treinados para atender empresa, dessa forma alguns cenários foram testados, até que fosse encontrado um cenário capaz de atingir a meta estipulada em 75%, de forma a minimizar a quantidade de atendentes necessários no canal empresa. O resultado aponta que para que esse objetivo seja atingido, baseando-se nos tempos atuais de processamento de cada tipo de chamada, são necessários 10 dos 20 atendentes, aptos ao atendimento de empresas.

Os mesmos cenários foram aplicados ao núcleo 2, contendo os canais de autorizações e atendimento exclusivo a médicos, e os resultados são exibidos na tabela 6.

Tabela 6: Simulação de cenários do núcleo 2

Cenário	Núcleo de atendimento	Atendentes	nível de serviço	% de abandono	% de abandono do cenário	Tempo médio de espera	Média de Pessoas na Fila	Uso do Recurso
Modelo tipo N								
1	Autorização	14	61%	11%	14%	00:01:36	5,68	98,4%
	Médicos	2	48%	40%		00:01:57	0,68	98,1%
2	Autorização	14	72%	8%	10%	00:01:17	4,56	97,3%
	Médicos	3	67%	27%		00:01:19	0,46	96,1%
3	Autorização	15	75%	8%	11%	00:01:12	4,28	96,8%
	Médicos	2	46%	41%		00:01:55	0,67	96,2%
4	Autorização	15	84%	5%	7%	00:00:55	3,24	95,3%
	Médicos	3	65%	29%		00:01:17	0,44	92,8%
5	Autorização	13	59%	12%	13%	00:01:39	5,89	98,7%
	Médicos	3	68%	26%		00:01:23	0,47	98,0%
Modelo totalmente dedicado								
6	Autorização	14	45%	18%	18%	00:02:00	7,13	99,2%
	Médicos	2	74%	20%		00:00:55	0,32	64,4%
7	Autorização	15	60%	12%	12%	00:01:36	5,74	98,2%
	Médicos	3	90%	10%		00:00:18	0,1	49,2%

Mais uma vez, os resultados apontaram vantagem na utilização do modelo N para o sistema estudado, onde apesar da melhora de 26% no nível de serviço do canal médicos, no cenário 6 totalmente dedicado, fazendo-o praticamente atingir a meta estipulada, o impacto negativo causado no canal de autorizações foi grande, e por se tratar do canal com maior volume

de chamadas dentre os dois, o resultado negativo impactou diretamente no percentual de abandono do núcleo como um todo. Além de reduzir a utilização do recurso no canal médicos, para 48,2%, deixando os atendentes ociosos em mais da metade do tempo disponível. Para os cenários do modelo N, novamente percebe-se que o aumento de recursos disponíveis no canal de maior complexidade, proporciona melhores resultados. Porém, desta vez a transferência de um atendente do canal de menor complexidade para o de maior, causou um impacto negativo no canal de menor complexidade, ao contrário do ocorrido no núcleo 1. Apesar da piora no nível de serviço do canal de autorização, mediante a troca apresentada no cenário 5, o percentual de abandono geral do cenário diminuiu, cabendo assim uma melhor investigação quanto aos impactos da realização dessa troca. Desta forma, o cenário extremo foi testado, onde os 16 atendentes disponíveis foram colocados no canal de médicos, além de ser testado também o incremento de dois atendentes no canal de médicos, dado o resultado positivo conseguido no cenário 2 onde o mesmo canal foi acrescido em um funcionário, conforme mostra a tabela 7.

Tabela 7: Segunda rodada de simulação de cenários do núcleo 2

Cenário	Núcleo de atendimento	Atendentes	nível de serviço	% de abandono	% de abandono do cenário	Tempo médio de espera	Média de Pessoas na Fila	Uso do Recurso
				Modelo tipo N				
8	Autorização	0	57%	14%	13%	00:01:43	6,11	0,0%
	Médicos	16	92%	4%		00:00:16	0,1	98,6%
9	Autorização	14	83%	5%	6%	00:00:57	3,37	95,9%
	Médicos	4	78%	19%		00:00:57	0,32	92,6%

O cenário extremo, se mostrou ineficiente, prejudicando os resultados do canal autorização, indicando novamente que a troca para esse núcleo não é uma boa opção e a distribuição dentro da mão de obra disponível atualmente é a melhor possível. O cenário 9, no qual foram adicionados dois atendentes na disponibilidade média no núcleo médicos, se mostrou a solução de menor investimento capaz de melhorar os resultados a ponto de atingir-se as metas, onde ambos os canais conseguiram ficar acima dos 75% propostos pelo trabalho, mantendo uma boa distribuição no uso do recurso e diminuindo em mais da metade o percentual de abandono do núcleo, sendo assim uma solução melhor do que o aumento de um atendente em cada canal como demonstrado no cenário 4.

6. Conclusões

O trabalhou apresentou uma proposta de avaliação das medidas de desempenho em um *call center* de uma operadora de plano de saúde, com o uso de simulação. A metodologia, consistiu na elaboração de um modelo de simulação utilizando o software ARENA®, que conseguisse representar o cenário real, obtendo resultados similares aos contidos nos dados fornecidos pela companhia.

A simulação se mostrou uma ferramenta extremamente útil para a avaliação de problemas desse tipo, sendo capaz de indicar se mudanças na configuração de como as chamadas são processadas e no número de atendentes disponível por canal, poderiam acarretar melhores resultados, e quais seriam essas mudanças. Além de indicar de forma detalhada quais seriam os impactos de cada tipo de mudança e quais os benefícios que as melhores soluções podem trazer.

Os modelos analisados, focaram em tentar intervir o mínimo possível para que as metas desejadas fossem estipuladas. A solução encontrada para um dos núcleos analisados, nem se quer envolvia a contratação de novos funcionários, indicando que um reajuste nas quantidades definidas para um dos canais pode por si só proporcionar o atingimento das metas. Apesar de ser uma mudança na configuração, atualmente a empresa realiza treinamentos para habilitar funcionários para os canais de maior complexidade, o que ocorre é que por alguma razão, o número médio estipulado para cada canal não sofre alterações, fazendo com que novos

funcionários só sejam habilitados a serem transferidos para um desses canais, mediante a perda de mão-de-obra. A proposta para o núcleo 1, seria de que de forma gradual, a empresa realizasse treinamentos com os funcionários do canal de informações para que de tempos em tempos, alguns atendentes possam ser habilitados a trabalharem no canal empresa, sem aguardar alguma baixa eventual nesse canal. Já para o segundo núcleo, o estudo apontou que para um completo atingimento das metas, seria necessária a realização de contratações, porém soluções como a realização de campanhas para os médicos a fim de melhorar o tempo de tolerância, indicado pela simples análise dos dados como sendo o cliente mais impaciente, poderia trazer significativas melhorias. Outro fator importante indicado pelo estudo, é de que diferentemente do indicado para o núcleo 1, no núcleo 2 a estratégia de se manter o número estabelecido de atendentes aptos no canal de médicos, trabalhando apenas com reposição em caso de perda de mão-de-obra, se mostrou uma boa estratégia, dessa forma o indicado caso a opção não fosse pela contratação, seria de manter a estruturação conforme é atualmente.

O estudo foi realizado em cima de bases consolidadas de gerenciamento, o que pode gerar resultados pouco específicos. Uma evolução no detalhamento do sistema, aprofundando-se nos dados operacionais, a fim de conhecer a real taxa de chegada dos clientes de cada núcleo, as demandas em cada período do dia e a escala de trabalho em cada canal, pode proporcionar um modelo ainda mais fiel ao sistema real, e acarretar novas e melhores soluções para o problema, mais minuciosa.

Para futuros trabalhos, existem inúmeras possibilidades de modificações no contexto do caso avaliado, seja no roteamento de chamadas, prioridade de filas ou redistribuição de mão-de-obra entre núcleos. Essas mudanças geram resultados diversificados, e o aumento na geração de cenários, produz informação importante para o gerenciamento da companhia, seja na proposição de melhorias ou na indicação de que o modelo utilizado está adequado e que consequentemente deveria ser mantido.

Referências

Avramidis, A.N; L'Ecuyer, P. (2005). Modeling and simulation of call centers. *Simulation Conference, 2005*. Proceedings of the Winter. IEEE.

Aksin, Z.; Armony, M.; Mehrotra, V. (2007). The Modern Call Center: A Multi-Disciplinary Perspective on Operations Management Research. *Production and operations management,* v.16, 665-688.

Dijik, N.M.V.; Sluis, E.V.D. (2008). Practical optimization by OR and simulation. *Simulation Modelling Practice and Theory*, v.16, 1113-1122.

Ertogral, K.; Bamuqabel, B. (2008). Developing staff schedules for a bilingual telecommunication call center with flexible workers. *Computers & industrial engineering*, v.54, 118-127.

Gans, N., Koole, G.; Mandelbaum, A. (2003). Telephone Call Centers: Tutorial, Review and Research Prospects. *Manufacturing & Service Operations Management,* v.5, 79-141.

Garnett, O.; Mandelbaum, A. (2000). An Introduction to Skills-Based Routing and its Operational Complexities. *Teachingo Notes*.

Garnett, O.; Mandelbaum, A.; Reiman, M. (2002). Designing a Call Center with Impatient Customers. *Manufacturing & Service Operations Management*, v.4, 208-227.

Legros, B.; Jouini, O.; Dallery, Y. (2015). A flexible architecture for call centers with skill-based routing. *International Journal of Production Economics*, v.159, 192-207.

Ma, J.; Kim, N.; Rothrock, L. (2011). Performance assessment in an interactive call center workforce simulation. *Modeling and Performance Analysis of Networking and Collaborative Systems*, v.19, 227-238.

Mandelbaum, A. (2006). *Call centers (centres)*: research bibliografy with abstracts.

Mehrotra, V.; Fama, J. (2003). Call center simulation: methods, challenges and opportunities. *Proceedings of the 35th conference on Winter simulation*: driving innovation, 135-143.

Saltzman, R.; Mehrotra, V. (2001). Call Center Uses Simulation to Drive Strategic Change. *Interfaces*, v. 31, 87-101.

Wallace, R. B.; Whitt, W. (2005). A Staffing Algorithm for Call Centers with Skill-Based Routing. *Manufacturing & Service Operations Management*, v.7, 276-294.