

AVALIAÇÃO DO DIMENSIONAMENTO DE ATENDENTES DO *CALL CENTER* EM UMA EMPRESA TELEFÔNICA NACIONAL



Danielle Magalhães de Almeida (UFRJ/ COPPE)
danielle@pep.ufrj.br

Samir Jorge Guedes Sias Thompson (UFRJ/ COPPE)
samir@pep.ufrj.br

Glaydston Mattos Ribeiro (UFRJ/ COPPE)
glaydston@pet.coppe.ufrj.br

Edilson Fernandes de Arruda (UFRJ/COPPE)
efarruda@pep.ufrj.br

O tempo de espera do consumidor por atendimento é uma das métricas utilizadas para avaliar a qualidade do serviço prestado por empresas que possuem uma central de chamadas. Utilizou-se o software Arena® para modelar 3 cenários que simularam a dinâmica de atendimento de uma célula da maior operadora de telefonia fixa do Brasil. Comparou-se os dados a serem utilizadas no modelo através dos softwares Arena® e R® mediante análises gráficas e testes de hipóteses, aplicando no modelo o melhor resultado de ajuste de acordo com critérios estatísticos. Ratificou-se, por fim, o quanto a Simulação pode auxiliar no processo de dimensionamento adequado do número de atendentes visando, além de proporcionar uma boa experiência para o cliente, redução dos custos operacionais.

Palavras-chave: Simulação, Otimização, Call center, Arena.

1. Introdução

A central de chamadas de uma concessionária de serviços de telecomunicações representa o processo mais significativo na etapa do planejamento estratégico que engloba o atendimento ao cliente. Também chamada de *call center*, a central de atendimento tem como objetivo principal criar uma relação direta entre o cliente e a empresa. Para a empresa utilizada no estudo de caso deste trabalho, o desempenho do time de *call center* é determinante para o aumento da qualidade do serviço prestado que, por sua vez, ocasiona aumento de receita e redução de custos. A empresa citada é a maior operadora de telefonia fixa, a quarta maior provedora de telefonia móvel do Brasil, e a terceira maior empresa do setor de telecomunicações na América do Sul. A qualidade na prestação do serviço de atendimento telefônico é medida a partir de informações como tempo de espera, número de transferências, chamadas repetidas e avaliação do atendimento. Evitar que o cliente permaneça por muito tempo na fila de atendimento é um dos principais objetivos dos gerentes responsáveis pelo departamento de *call center*. De acordo com Legros (2019), os administradores de muitos sistemas de prestação de serviço definem um tempo limite para controlar o tempo gasto pelos clientes no sistema. Alguns *call centers* utilizam *softwares* para gerenciar uma política de rejeição de chamada após um determinado limite de tempo de espera. Para a empresa de energia Primagaz, por exemplo, são seis minutos, enquanto são três minutos para a operadora de telecomunicações Keyyo. Para a empresa utilizada neste estudo de caso, no entanto, não existe uma política de rejeição de chamadas e o tempo de espera é uma métrica que, além de indicar a qualidade do serviço prestado, é inversamente proporcional à satisfação do cliente, visto que a demora no atendimento é classificada como o principal motivo para o abandono de chamadas e posteriormente, o pedido de cancelamento dos contratos de serviços prestados pela empresa. Considerando o impacto que o número de atendentes causa no tempo de espera por atendimento, propõe-se um modelo de simulação de *call center* com o intuito de proporcionar uma visualização do comportamento do sistema diante da modificação desse parâmetro. Busca-se, assim, o número ideal de atendentes em uma operação de *call center*, considerando o tempo de espera dos clientes na fila por atendimento e os custos operacionais da companhia. Avaliou-se também um cenário com aumento na taxa de chegadas de ligações para avaliar se o número de atendentes seria suficiente para um eventual aumento de demanda. O presente trabalho está organizado em cinco seções: introdução ao tema, referencial teórico, metodologia aplicada, resultados obtidos e considerações finais.

1.1. A empresa

O presente trabalho modela e avalia o *call center* de uma concessionária de serviços de telecomunicações presente em 25 estados brasileiros e no Distrito Federal. Seus aproximadamente 61 milhões de clientes estão divididos em quatro grupos de serviços: telefonia fixa, telefonia móvel, banda larga e televisão. Cada grupo de serviço conta com clientes de contratos individuais ou agrupados (combos). Devido a dificuldades de acesso à base de dados, este trabalho trata apenas desses últimos tipos de contratos relativos aos clientes que não possuem serviço de telefonia móvel. Dois perfis de clientes são utilizados na simulação para auxiliar a tomada de decisões: os que possuem e os que escolheram não utilizar o serviço de televisão (TV).

A concessionária conta com sete centrais de atendimento telefônico especializado, capazes de atender todos os perfis de clientes. Existe uma célula para clientes ingressantes, uma segunda para venda de novos serviços para quem já é cliente e ainda uma terceira, responsável por reter os clientes que desejam o cancelamento parcial ou total dos serviços. A última, chamada de célula de retenção, foi a que motivou a realização deste trabalho. As interações com os clientes são registradas em sua totalidade e analisadas a fim de melhorar processos, aumentar receita e/ou reduzir custos. Existe uma gerência de inteligência responsável por consolidar todas as informações de atendimento para posterior utilização no processo de tomada de decisões. Todavia, esse processo ainda prescinde de análises estatísticas e simulações de cenários que avaliem o impacto da variação do número de atendentes na qualidade do atendimento e, indiretamente, na receita e nos custos. É exatamente essa lacuna que motivou a realização do presente trabalho.

Quando finalizada, a chamada é classificada de acordo com a decisão final tomada pelo cliente, o qual pode tomar sete decisões diferentes. Os mnemônicos dos grupos descritos abaixo são funções da decisão tomada:

- *Cancelado*: quando o cliente solicita o cancelamento de todos os produtos oferecidos pela empresa;
- *Cancelado individual*: quando o cliente solicita que o serviço seja prestado individualmente para cada produto;
- *Oferta Down*: quando o cliente solicita o cancelamento de algum produto, porém a manutenção dos outros;
- *Oferta Lateral*: quando o cliente deseja manter os produtos, porém alterar seu perfil;

- *Oferta Ret Inc*: quando o cliente deseja retirar um produto e incluir outro;
- *Oferta Up*: quando o cliente deseja adquirir um novo produto;
- *Retido Arg*: quando o cliente não altera em nada os seus produtos e é retido pela companhia através da argumentação do atendente.

2. Referencial teórico

2.1. Teoria de Filas

As filas podem ser caracterizadas como um conjunto de entidades que aguardam atendimento de um determinado servidor. A Teoria de Filas é um método analítico que aborda a modelagem de sistemas através de formulação matemática. De acordo com Medeiros (2010), o objetivo da teoria de filas é avaliar a formação de filas em um sistema por meio de elementos como o tempo médio de permanência do cliente, seu tempo médio de espera, a distribuição de probabilidade do tempo de chegada dos clientes e de atendimento dos servidores, o número de servidores, a capacidade do sistema e a disciplina de atendimento. As filas podem refletir o nível de insatisfação do cliente, daí a importância de se entender o motivo da formação das filas, avaliar sua dinâmica e tomar medidas buscando sua redução. Em geral, a formação de fila de atendimento em empresas de *call center* está relacionada, dentre outras coisas, ao tamanho da equipe de atendentes.

A disciplina de filas a ser avaliada neste trabalho é FIFO (*first in, first out*), segundo a qual os clientes são atendidos em ordem de chegada. O modelo utilizado para compor tal sistema é o M/M/C, cujas distribuições dos tempos entre chegadas e de atendimento são exponenciais e representadas pelas letras M. Além disso, C é um número inteiro que representa o número de atendentes. Utiliza-se *Simulação Computacional* como ferramenta de análise dos sistemas de filas avaliados. De acordo com Hillier & Lieberman (2010), a técnica de simulação envolve o uso de um computador para simular a operação de um processo ou sistema como um todo. A simulação imita repetidamente a evolução das movimentações envolvidas no sistema para gerar um perfil de possíveis resultados. O computador então, gera e registra de forma aleatória, as ocorrências dos vários eventos que dirigem o sistema como se ele estivesse operando fisicamente. É possível simular um período longo de operação em questão de segundos. Registrar o desempenho da operação simulada do sistema para procedimentos operacionais alternativos ou futuros projetos, torna possível a avaliação e a comparação de alternativas disponíveis antes de escolher uma.

2.2. Simulação

A Simulação pode ser entendida como uma imitação do comportamento aleatório de um sistema com a finalidade de estimar as suas medidas de desempenho a partir de um modelo computacional. Sendo uma das ferramentas mais gerais em Pesquisa Operacional, a Simulação possibilita a modelagem de situações reais de forma mais acurada que os modelos analíticos de Teoria de Filas. Enquanto nos métodos analíticos faz-se necessário ajustar a realidade a um padrão pré-estabelecido, a Simulação é capaz de representar as peculiaridades de uma operação complexa. Na Simulação os resultados variam a cada execução, o que acontece na realidade. De acordo com Sargent (2011), modelos de simulação estão cada vez mais sendo usados para auxiliar na tomada de decisões porque possibilitam a visualização do impacto que mudanças em determinados parâmetros podem causar no sistema a ser avaliado. Experimentos computacionais permitem a avaliação do impacto de possíveis alterações antes de sua implementação. Tipicamente, um projeto de simulação fornece os resultados adequados para a tomada de decisão apenas se três etapas forem seguidas (Chwif e Medina, 2010): formulação do modelo, implementação do modelo e análise dos resultados. Tais passos foram seguidos para a realização deste estudo de caso.

A crescente importância dos serviços de *call center* aumenta a necessidade de utilização de métodos científicos para a tomada de decisões em detrimento do uso da intuição. Mehrotra e Fama (2003) explicam que a Simulação é capaz de modelar a aleatoriedade de chegada das chamadas e tempos de atendimentos, além da interação entre elas, as rotas das ligações e a capacidade de utilização dos atendentes.

3. Metodologia

Os dados utilizados para a realização deste trabalho foram obtidos através da central de armazenamento de informações da companhia. Os registros de chamadas são salvos individualmente com os dados de horário de início e término da chamada, tempo de atendimento na URA (Unidade de resposta audível), tempo de utilização do atendente pessoal, informações de quais serviços o cliente tem contratados ao ligar e ao final da ligação. Analisou-se então a dinâmica de atendimento dos nove primeiros meses de 2019. As chamadas abandonadas foram removidas da base de dados porque o sistema automaticamente zera o tempo dessas ligações. Utilizou-se o programa Microsoft Excel® para higienizar os dados e calcular o intervalo de tempo entre chamadas. Utilizou-se as seguintes variáveis: chamadas, tempo de atendimento na URA, serviços que o cliente possui, duração do primeiro atendimento, opções de atendimento, duração do segundo atendimento, resultado da chamada e número de

atendentes. O primeiro atendimento é feito pela URA e engloba a avaliação dos serviços que o cliente possui para direcionamento da chamada. O segundo atendimento engloba a avaliação da movimentação do cliente dentre as opções oferecidas pelo atendente. Avaliou-se um volume de 704.933 chamadas e todos os tempos foram medidos em segundos. Valores atípicos (*outliers*) não foram desconsiderados neste trabalho porque em um sistema de *call center* esses dados impactam diretamente os indicadores e são potenciais indicadores de alta insatisfação dos clientes. Um dos motivos que levaram a essa decisão é o fato de chamadas de cancelamento demandarem mais tempo de atendimento. Naturalmente, essas chamadas não poderiam ser desconsideradas. A Tabela 1 mostra o percentual de chamadas de acordo com o resultado de atendimento.

Tabela 1 – Percentual das chamadas de acordo com o resultado de atendimento

	Com TV	Sem TV
Total Chamadas de clientes	51,11%	48,89%
Chamada Cancelada	14,41%	25,96%
Chamada Cancelada individualmente	5,01%	2,60%
Chamada Oferta Down	4,12%	0,33%
Chamada Oferta Lateral	25,27%	19,10%
Chamada Oferta com Retirada e Inclusão	0,16%	0,02%
Chamada Oferta Up	0,60%	0,18%
Chamada Retido com argumentação	50,43%	51,82%

Fonte: Elaboração própria

A operação simulada é composta por 31 atendentes, sendo 16 para atender clientes que optam pelo serviço de TV e 15 para os que não optam. A taxa de chegada de chamadas é, em média, de 33,5 ligações por segundo. E o tempo de atendimento (TMA) é 17,15 minutos, em média. Para encontrar a melhor distribuição de atendimento para as entidades do modelo, comparou-se os resultados a serem utilizadas através dos softwares Arena® versão 14 e R® versão 3.6.0 implementados em um computador com sistema operacional Windows 10 home 64 bits e 8,00GB de memória RAM e processador AMD A6-6310 APU.

Para todos os ajustes, no Arena®, a ferramenta utilizada foi o *Input Analyzer* com o comando *fit all*. Tais resultados foram comparados com os obtidos no R® através de análises gráficas e testes de hipóteses como o Teste de Qui-Quadrado de Pearson e o Teste de Kolmogorov-Smirnov, que são os mais utilizados dentro da literatura estatística, de acordo com Rosemberg (1976) e compatíveis com os dois softwares. Aplicou-se no modelo o melhor resultado de ajuste de acordo com critérios estatísticos. A seleção de um tempo de aquecimento adequado para um modelo de simulação é uma escolha de difícil determinação de acordo com cada modelagem. No presente trabalho, optamos por não utilizar um tempo de aquecimento já que o serviço de atendimento telefônico funciona 24h por dia. Definiu-se por fim, que para elaborar esse modelo de simulação, seriam necessárias 30 replicações.

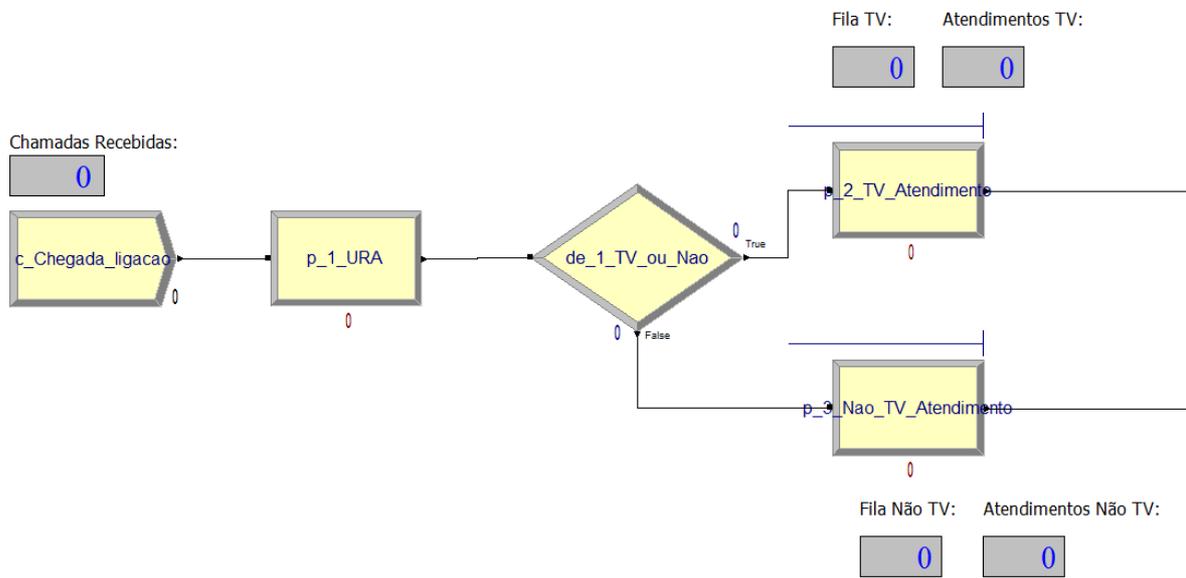
A Tabela 2 apresenta os módulos lógicos do Arena® utilizados na simulação do cenário de atendimento da célula de retenção e suas respectivas distribuições. A entidade foi definida como sendo as chamadas recebidas que representam os clientes da empresa a ser avaliada. As figuras 1, 2 e 3 apresentam a modelagem configurada no Arena®.

Tabela 2 - Módulos lógicos utilizados para a simulação no Arena® e seus parâmetros de distribuição

Parâmetro	Distribuição	P-valor Teste de Qui-Quadrado de Pearson	P-Valor Teste de Kolmogorov-Smirnov
Chamadas que chegam no atendimento.	Exponencial: $1/\lambda = 33,46018$.	$2,2 \times 10^{-16}$	$2,2 \times 10^{-16}$
Atendimento na URA.	Normal: $\mu = 54,87737$ e $\sigma = 30,20862$.	$2,2 \times 10^{-16}$	0,5848
Atendimento para clientes com serviço de TV.	Normal: $\mu = 525,9001$ e $\sigma = 17,78516$.	0,7775	0,9999
Atendimento para clientes sem serviço de TV.	Normal: $\mu = 503,0572$ e $\sigma = 17,01265$.	0,7991	0,9999
Atendente do process com TV - Chamada Cancelada.	Normal: $\mu = 75,83416$ e $\sigma = 2,564598$.	0,9995	0,9999
Atendente do process com TV - Chamada Cancelada individualmente.	Normal: $\mu = 26,34131$ e $\sigma = 0,8908238$.	0,9999	0,9999

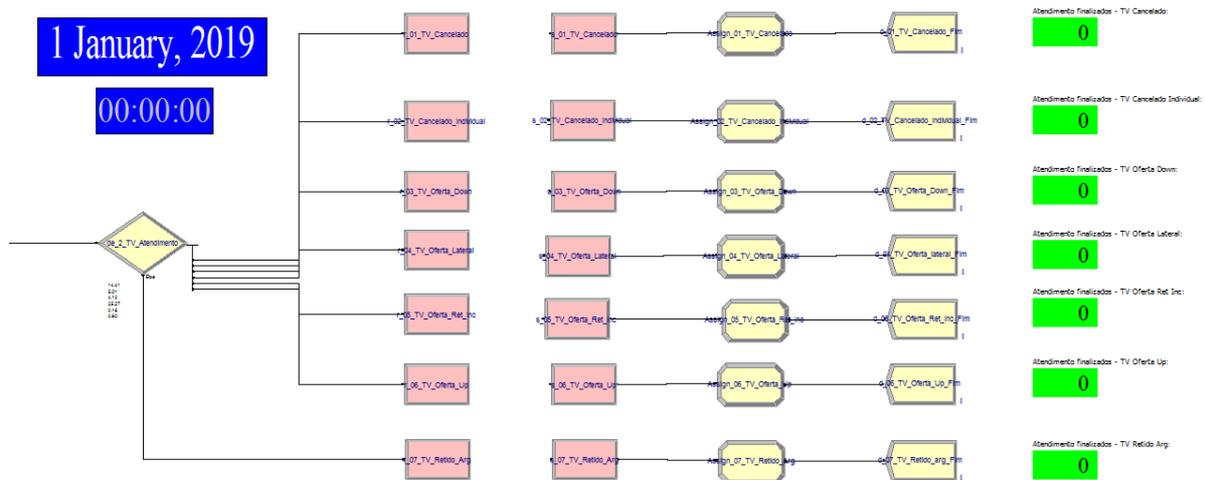
Atendente do process com TV - Chamada Oferta Down.	Normal: $\mu = 21,711$ e $\sigma = 0,7342336$.	0,9999	0,9999
Atendente do process com TV - Chamada Oferta Lateral.	Normal: $\mu = 132,8384$ e $\sigma = 4,492396$.	0,9965	0,9999
Atendente do process com TV - Chamada Oferta com Retirada e Inclusão.	Normal: $\mu = 0,8231659$ e $\sigma = 0,02783824$.	0,9999	0,9999
Atendente do process com TV - Chamada Oferta Up.	Normal: $\mu = 3,086872$ e $\sigma = 0,1043934$.	0,9999	0,9999
Atendente do process com TV - Chamada Retido com argumentação.	Normal: $\mu = 265,2652$ e $\sigma = 8,970874$.	0,965	0,9999
Atendente do process sem TV - Chamada Cancelada.	Normal: $\mu = 130,5747$ e $\sigma = 4,415841$.	0,9967	0,9999
Atendente do process sem TV - Chamada Cancelada individualmente.	Normal: $\mu = 13,06776$ e $\sigma = 0,4419321$.	0,9999	0,9999
Atendente do process sem TV - Chamada Oferta Down.	Normal: $\mu = 1,646332$ e $\sigma = 0,05567648$.	0,9999	0,9999
Atendente do process sem TV - Chamada Oferta Lateral.	Normal: $\mu = 96,10462$ e $\sigma = 3,250115$.	0,9989	0,9999
Atendente do process sem TV - Chamada Oferta com Retirada e Inclusão.	Normal: $\mu = 0,1028957$ e $\sigma = 0,00347978$.	0,9999	0,9999
Atendente do process sem TV - Chamada Oferta Up.	Normal: $\mu = 0,9260616$ e $\sigma = 0,03131802$.	0,9999	0,9999
Atendente do process sem TV - Chamada Retido com argumentação.	Normal: $\mu = 260,7378$ e $\sigma = 8,817763$.	0,9669	0,9999

Figura 1 - Primeira dinâmica de atendimento na célula de retenção



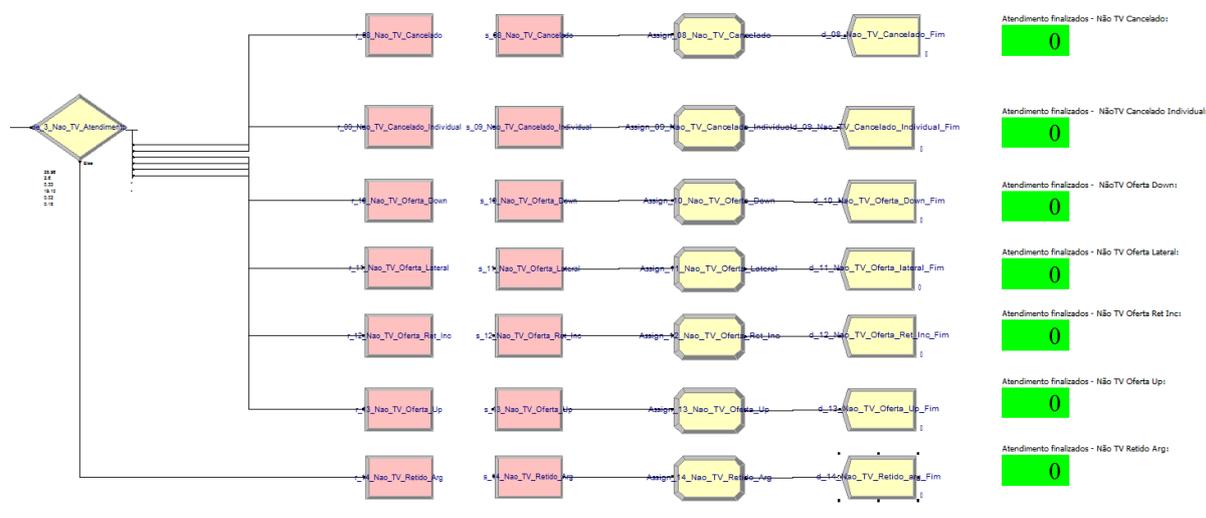
Fonte: Elaboração própria

Figura 2 -Atendimento para clientes que possuem o serviço de TV



Fonte: Elaboração própria

Figura 3 -Atendimento para clientes que não possuem o serviço de TV



Fonte: Elaboração própria

4. Resultados

A primeira simulação, cujo objetivo era avaliar se o modelo estava adequado o suficiente para representar a realidade, teve um tempo de processamento de 4 horas e 03 minutos. A taxa de chegada de chamadas é de 33,46018 por segundo. O tempo médio de espera na fila para clientes com serviço de TV foi de 0,4544 segundos enquanto para os clientes sem serviço de TV foi de 0,4536. O número de chamadas totais finalizadas de acordo com a simulação foi de 79.941. Para manter os 31 atendentes, a companhia tem gastos operacionais mensais de R\$85.442,00. Elaborou-se um segundo cenário para simular o aumento (em média é de 15%) na taxa de chegada de chamadas, típico dos meses posteriores a reajustes de mensalidade ou do encerramento do período de promoções que envolvem gratuidade. A taxa de chegada de chamadas por segundo foi então ajustada para 38,47921. Diante dessa alteração, o tempo de espera na fila subiu para 1,9475 segundos para clientes com serviço de TV e 1,8455, para clientes sem serviço de TV. O número de chamadas totais finalizadas, de acordo com a simulação do Arena®, foi de 69.503, o que representa uma queda de 13%. O tempo de processamento caiu para 3 horas e 35 minutos. Elaborou-se um terceiro cenário alterando o parâmetro número de atendentes considerando agora 12 para a célula que atende os clientes com serviço de TV e 10 para os sem esse serviço. O tempo de processamento computacional da simulação subiu para 5 horas e 20 minutos. O tempo de espera na fila para clientes com serviço de TV foi de 11,0666 segundos e para os clientes sem serviço de TV, 29,5291 segundos. A redução nos custos operacionais foi de 28%, o que representa uma economia mensal de

R\$24.224,00. A simulação do Arena®, dessa vez, mostrou que o volume de chamadas totais finalizadas foi de 79.964, praticamente o mesmo que no primeiro cenário.

A Anatel (Agência Nacional de Telecomunicações) é o órgão responsável por regulamentar e fiscalizar os serviços de telecomunicações garantindo os direitos do consumidor. Desde julho de 2008, por determinação do Ministério da Justiça no Decreto nº 6.523, o prazo máximo de espera do consumidor para o atendimento por telefone após digitar a opção de falar com o atendente, é de 45 segundos. O não cumprimento dessas normas pode incidir em multas entre R\$200 e R\$3 milhões. A empresa analisada neste trabalho conseguiria reduzir custos mantendo o tempo de espera por atendimento dentro da obrigatoriedade da Lei se adotasse o cenário 3 para períodos de menor demanda. A Tabela 3 abaixo apresenta um sumário dos resultados das simulações considerando o cenário 1 como a realidade atual da empresa, o cenário 2 alterando o parâmetro taxa de chegada e o cenário 3 alterando o número de atendentes.

Tabela 3 – Resultados para os 3 cenários de simulação

Cenários	Tempo de processamento computacional	Taxa de chegada	Tempo na fila		Chamadas Finalizadas	Número de atendentes	Custos operacionais
			Com TV	Sem TV			
1	4:03h	33.46018/s	0.4544s	0.4536s	79.941	31	85.422K
2	3:35h	38.47921/s	1.9475s	1.8455s	69.503	31	85.422K
3	5:20h	33.46018/s	11.0666s	29.5291s	79.964	22	61.198K

Fonte: Elaboração própria

5. Considerações finais

Dentre os desafios encontrados ao longo da elaboração deste estudo de caso, pode-se citar como mais significativos aqueles relacionados à limpeza de base. Verificou-se a importância de se trabalhar com dados coerentes e fidedignos à realidade de uma célula de *call center* porque apenas com a apuração e definição correta dos KPIs (*Key Performance Indicators*) é possível elaborar métricas que fomentam a tomada de decisões mais complexas. Diante de parâmetros distintos como o volume de chamadas e número de atendentes, assim como a sensibilidade dos mesmos, o processo inicial de modelagem demorou mais tempo que o esperado. Ocorreram mudanças nos objetivos traçados previamente em decorrência da avaliação do tempo de aquecimento que acabou sendo desconsiderado na simulação. Além desse fator, como sugestão

para trabalhos futuros seria interessante avaliar a taxa de abandono das chamadas (que atualmente é ausente na base de dados da empresa) e fazer a divisão de atendentes por turno de trabalho.

REFERÊNCIAS

BRASIL, Decreto nº 6.523, de 31 de julho de 2008. **Diário Oficial da República Federativa do Brasil**, Poder Executivo, Brasília, DF, 2008.

Chwif, L; Medina, A. C. (2010) **Modelagem e Simulação de Eventos Discretos – Teoria e Aplicação**. Editora do Autor, São Paulo.

Hillier, Frederick S. & Lieberman, Gerald J. **Introdução à Pesquisa Operacional**. 8ª Edição. São Paulo: McGraw-Hill, 2010.

Legros, B. (2019) **Late-rejection, a strategy to perform an overflow policy**. European Journal of Operational Research. Elsevier.

Medeiros, E. S. (2010) **Pesquisa operacional para os cursos de administração e engenharia: programação linear, simulação**. Atlas, São Paulo.

Mehrotra, V., J. Fama. 2003. **Call center simulation modeling: Methods, challenges, and opportunities**. In Proceedings of the 2003 Winter Simulation Conference, 135–143: IEEE Press.

Rosemberg, M. (1976) **A Lógica da Análise do Levantamento de Dados**. Ed. Cultrix/EDUSP, São Paulo.

Sargent, R. G. **Verification and validation of simulation models**. In. Winter simulation conference. Proceedings... Phoenix, AZ, USA, 2011.