

Teoria de filas e o dimensionamento de servidores: um estudo de caso

Theory of queues and the sizing of servers: a case study

 **Rafaela Zanardi¹**

 **Lori Viali²**

¹Pontifícia Universidade Católica do Rio Grande do Sul
 Possui graduação em Engenharia de Produção pela
 Pontifícia Universidade Católica do Rio Grande do Sul
 (2018). Atua em uma empresa de semicondutores com
 âmbito internacional em atividades de planejamento e
 controle de produção e projetos de melhoria. Suas áreas
 de interesse de pesquisa são Projetos, Processos e
 Análise de Dados.
 Endereço Av. Ipiranga, 6681 - Partenon, Porto Alegre -
 RS, CEP 90610-00.
rafaela.z@hotmail.com

²Pontifícia Universidade Católica do Rio Grande do Sul
 Atualmente é professor adjunto (20 horas) do Instituto
 de Matemática, Departamento de Estatística, da
 Universidade Federal do Rio Grande do Sul e professor
 titular (30 horas) da FAMAT (Faculdade de
 Matemática), Departamento de Estatística, da Pontifícia
 Universidade Católica do Rio Grande do Sul.
 Endereço Av. Ipiranga, 6681 - Partenon, Porto Alegre -
 RS, CEP 90610-00.
viali@pucrs.br

Resumo: Este trabalho apresenta uma aplicação da teoria de filas na área de tecnologia de informação (TI) em uma empresa jornalística cuja perda de produtividade no atendimento de demanda está gerando atrasos na solução de problemas, perda de parceiros comerciais e significativo crescimento de risco de parada da produção. Neste estudo de caso, utilizou uma abordagem quantitativa de formato exploratório e teve por objetivo dimensionar os servidores da subárea de processos de tecnologia de informação em uma empresa jornalística. Como resultado da avaliação de cenários, foi possível verificar que o número de servidores igual a quatro é o mais adequado, porém se for acrescentado um servidor obtém-se uma variação de 8% para 26% em relação à ociosidade, o que se justifica frente a redução do risco da paralisação da produção.

Palavras-chave: Teoria das filas. Processos. Tecnologia de Informação.

Abstract: This paper presents an application of queuing theory in the area of IT in a journalistic company whose loss of productivity in the supply of demand is generating delays in solving problems, loss of commercial partners and significant increase in the risk of production stoppage. In this case study, it used a quantitative approach of exploratory format and had as objective to size the servers of the subarea of information technology processes in a journalistic company. As a result of the scenario evaluation, it was possible to verify that the number of servers equal to four is the most appropriate, but if a server is added a variation of 8% to 26% in relation to the idleness is obtained, which is justified reduction of the risk of production stoppages.

Keywords: Queue theory. Processes. Information technology.

1 Introdução

As empresas jornalísticas, como em qualquer outro tipo de negócio enfrentam dificuldades para se manterem ativas no mercado (Associação Nacional de Jornais [ANJ], 2016). Estas complicações se originam desde estratégias mal definidas, utilização de ferramentas não adequadas e até projetos com objetivos que fogem das reais expectativas do mercado consumidor (Sant'anna, 2007; ANJ, 2016).

O emprego de ferramentas de análise de dados se tornou algo imprescindível para compreensão da forma com que as pessoas trabalham e interagem dentro de suas rotinas e tomam suas decisões (Melo, 2008). No ambiente de prestação de serviços isso se torna mais significativo, visto que melhorias podem ser aplicadas por intermédio da exploração dos dados gerados por um sistema de informação robusto (Farias, Braido, Casalinho & Muniz, 2012; Santos & Campos, 2011).

A utilização de ferramentas de simulação de filas por sua vez, viabiliza a análise dos sistemas de forma mais simples ao aliar a programação matemática, pois desta forma é possível modelar o comportamento do sistema, construindo hipóteses através da observação, obtendo-se resultados mais próximos a realidade e atribuindo-se operações mais complexas (Kleinrock, 1975).

Taha (2008) afirma que ao empregar a modelagem em sistemas que envolvem tempo de espera e determinar medidas de desempenho que viabilizem a análise da produtividade dos mesmos através da teoria das filas, torna possível a tomada de decisão, seja ela a nível de conscientização da condição atual do processo, ou mesmo em projetos de melhorias dentro do sistema observado.

Perdoná, Nunes, das Neves, Naimer e Godoy (2017) afirmam que ao empregar a teoria de filas em processo de otimização de fluxos é possível refinar as restrições com um maior detalhamento e desenvolver diversificadas formas relacionadas a controle.

Desta forma, levanta-se a seguinte questão de pesquisa: como é possível dimensionar o número adequado de servidores e definir os principais parâmetros na área de processos de tecnologia de informação?

Sendo assim, faz-se necessário avaliar as principais características dos modelos de filas e suas relações, explorando em publicações científicas a aplicação em diferentes áreas para confirmar a efetividade de seu uso para atingir o objetivo desta pesquisa que é: dimensionar os servidores da subárea de processos de tecnologia de informação em uma empresa jornalística.

2 Referencial teórico

2.1 Teoria de Filas

A teoria de filas está inserida dentro da área de pesquisa operacional, e consiste basicamente na representação simplificada de um sistema por intermédio de modelos, fazendo uso de problemas de otimização de fluxo (Hillier & Lieberman, 2013).

2.1.1 Caracterização

Os sistemas fazem parte de um grupo de elementos que são capazes de manter a interdependência entre si para garantir o alcance de um objetivo (Novaes, 1975). Esse é o caso dos sistemas probabilísticos que apresentam entrada e saída aleatória e que podem ser descritos apenas em termos probabilísticos (Shannon, 1975).

Um sistema de filas é caracterizado por clientes que aguardam por atendimento de um ou mais servidores que trabalham a uma determinada taxa e pela chegada de clientes ao ambiente a uma taxa a ser determinada (Sampaio & Borschiver, 2008). É comum o surgimento de filas devido ao grande fluxo de pessoas ou processos dentro de um sistema em que a demanda de serviço é consideravelmente maior que a capacidade de atendimento ou quando existe a necessidade de compartilhamento de recursos (Prado, 1999).

A notação mais utilizada para sistemas de filas é a de Kendall, que consiste em representar um modelo de filas por meio de $A/B/C/Z/K/m$ ou mais resumidamente por $A/B/C$, onde em alguns modelos se assume que $Z = \text{FIFO}$, $K = \infty$, $m = \infty$ (Kleinrock, 1975). Além disso, é possível classificar os sistemas de filas de acordo com a quantidade de filas e servidores, tais como: (a) fila única com um servidor; (b) fila única com múltiplos servidores atuando em paralelo; (c) múltiplas filas com vários servidores em paralelo; (d) fila única com servidores atuando em série (Chase, Aquilano & Jacobs, 2006).

As principais relações existentes entre as diversas características de um sistema de filas e seus parâmetros dependem do modelo que se considera. Contudo, algumas relações são válidas para qualquer sistema de filas, como é o caso das Leis de Little, que relacionam o número de clientes em um sistema (L) com o tempo em que ele fica no sistema (W) (Kleinrock, 1975). Pode-se verificar também que:

$$(1) \text{ Número de clientes no sistema: } L = \lambda \times W$$

$$(2) \text{ Número de clientes na fila: } Lq = \lambda \times Wq$$

$$(3) \text{ Número de clientes sendo atendidos: } Ls = \lambda \times Ws$$

$$(4) \text{ Número de clientes no sistema é a soma do número de clientes na fila com o número de clientes sendo atendidos: } L = Lq + Ls$$

$$(5) \text{ Tempo no sistema é a soma do tempo na fila com o tempo sendo atendido: } W = Wq + Ws$$

Além disso, a relação (4) e (5) podem ser estabelecidas para qualquer sistema envolvendo filas.

2.1.2 Aplicações dos modelos de filas

São muitas as aplicações possíveis que a teoria das filas permite, sendo possível nos mais diversos tipos de áreas e problemas como em bancos, logística, hospitais e *call centers* (Perin, 1995).

Taufemback e S. Da Silva (2012), conseguiram diminuir os excessos de reservas bancárias que ocasionavam perda do custo de oportunidade do dinheiro, através de um gerenciamento dos períodos de aplicação associados ao tempo médio das mesmas para corrigir os níveis de saturação. Já Boucher e Piqué (2015), reduziram as desistências de clientes na agência e alcançaram resultados significativos quanto ao uso da teoria das filas em uma seguradora, cuja modelagem de alguns parâmetros facilitou para o segurado perceber o risco associado ao cancelamento das apólices no decorrer do período analisado.

Na logística, V. M. D. Silva, Souza e Bortolotti (2006) afirmam que os retornos são mais evidentes, pois ao descobrirem a distribuição com melhor aderência aos canais de atendimento do Porto de Itajaí, foi possível diagnosticar que acrescentando mais um ponto de atendimento, o tempo médio dos navios no sistema seria reduzido e a taxa de utilização do Porto seria aumentada. Schein (2010) destaca que ao trabalhar com os dados de entradas e saídas dos navios do Porto de Rio Grande dos anos de 2007 até 2015, o mais indicado era aumentar o tamanho da frota de dois para três rebocadores com o intuito de maximizar o nível de serviço e reduzir os custos incorridos no sistema.

Após avaliar a taxa de chegadas de navios em um determinado rio na China, Zhou, Guo e Wu, (2013), constataram que era necessária a intervenção da regulamentação de tráfego de navios da região visto que a capacidade de fluxo estava saturada e havia redução da segurança dos navios que ali navegavam. Liu e Ge (2017) ao utilizarem teoria de filas para modelagem do comportamento de guindastes no processo de descarregamento de contêineres, evidenciaram que era possível minimizar a emissão de CO₂ apenas descobrindo os principais parâmetros daquele sistema e observando quais eram os impactos em relação ao aumento ou redução destes veículos no sistema.

Lu, Tian e Guan (2012) aplicaram a simplificação de modelos dentro de um hospital e ao observar as taxas de chegada e saída dos pacientes enfermos e o tipo de distribuição de probabilidade com melhor aceitação, foi possível identificar que era necessário aumentar o número de médicos especializados em ultrassom e modificar a disposição de alguns equipamentos para otimizar o tempo de atendimento. Já Singer e Donoso (2008) conseguiram melhorar os serviços de operação de ambulâncias com o uso dos indicadores obtidos pela teoria das filas, avaliando o desempenho e possíveis impactos caso fosse inserido um número maior de ambulâncias.

Conceição, Loureiro e Gonçalves (2009) relatam que após análise dos indicadores relacionados à fila em uma empresa de *Call Center*, a escassez de recursos e a falta de qualificação dos servidores eram os principais geradores de congestionamento nas centrais telefônicas e redução de desempenho do sistema, frustrando tanto clientes quanto servidores. Já Yu, Gong e Tang (2016) apresentaram um

estudo com a aplicação de teoria das filas para analisar o número de chamados para a área de *help desk*, com o objetivo de reduzir o tempo de espera do cliente em relação a busca pelo serviço.

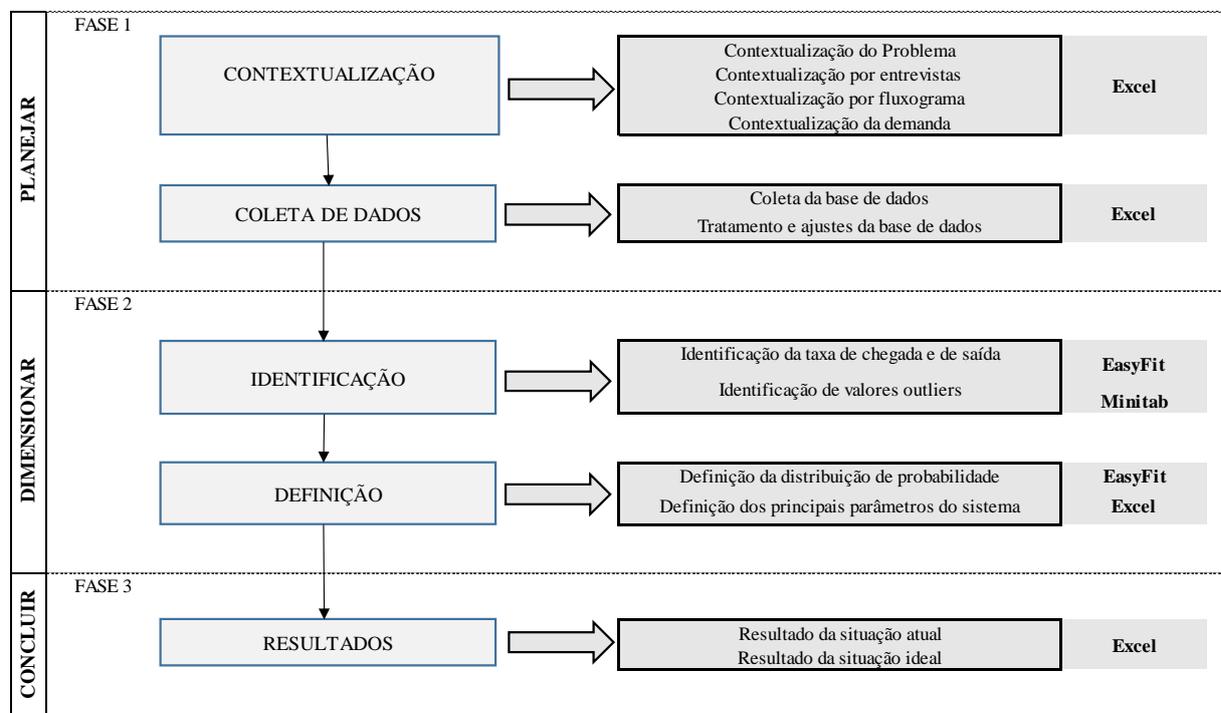
3 Metodologia

A presente pesquisa pode ser caracterizada de natureza aplicada, visto que se utiliza de conhecimentos específicos práticos de aplicação e utilização para sanar problemas, o que se percebe neste estudo (Gil, 2008). Em aspectos de abordagem, o presente estudo tem cunho quantitativo com a aplicação da teoria de filas e determinação de parâmetros do modelo probabilístico para atendimento de demandas (E. L. Silva & Menezes, 2005). Conforme os objetivos apresentados, a pesquisa pode ser classificada como exploratória, com fundamento e compreensão do problema de variabilidade das chamadas e a análise quanto ao tipo de distribuição de probabilidade que se evidencia na área de processos de TI (Bertucci, 2009). Já o procedimento é definido como um estudo de caso, pois possibilita um detalhamento maior em vista da utilização e organização de dados de serviços para interpretação da dinâmica de filas em uma área específica de uma empresa jornalística (Gil, 2008).

A decisão por esse estudo se fez pela necessidade avaliar as condições atuais do sistema observado, e propor sugestões de melhoria para aumento de produtividade do mesmo.

O método de trabalho foi elaborado pelos autores tomando por base o modelo de Chwif e Medina (2006) e o mesmo dividiu-se em quatro fases conforme mostrado na

Figura 1 - Metodologia de trabalho proposta



Fonte: Elaborado pelos autores.

3.1 Análise do ambiente e dos dados

Este estudo ocorreu dentro da área de tecnologia de informação em uma empresa jornalística. Atualmente esta área é dividida em cinco subáreas, e dispõe de 17 funcionários distribuídos conforme apresentado no esquema da

Figura 2 - Disposição do ambiente

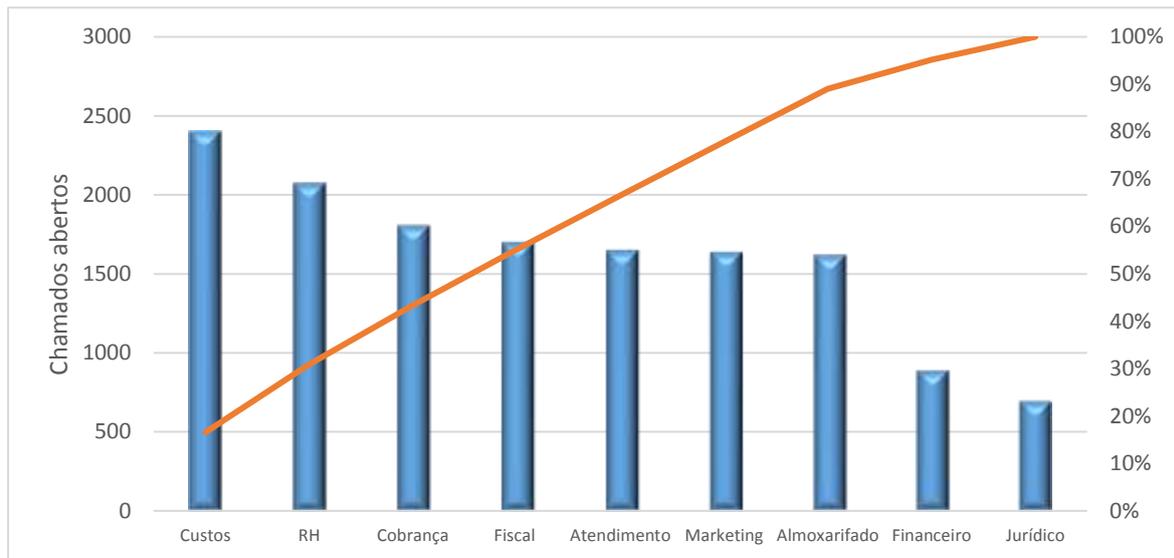
Área	Sub-área	Func.	Função	Demanda	Capacitação
Tecnologia da Informação (Coord. da TI)	Suporte Nível I	5	Help Desk	Interna e Externa	Técnico
	Suporte Nível II	2	Manutenção física de hardware e acessórios	Interna	Técnico
	Processos	5	Suporte para sistema interno da empresa	Interna	Superior
	Infraestrutura	1	Infraestrutura de redes e servidores	Interna	Superior
	Desenvolvimento	3	Desenvolvimento de aplicação interna, relatórios, consultas, gerenciamento de banco de dados	Interna	Superior

Fonte: Elaborado pelos autores.

Cabe ressaltar que a subárea alvo deste estudo foi a de processos por ser a mais crítica em relação a seus impactos de ordem financeira e produção caso ocorra falha no sistema.

Foram feitas entrevistas dentro do ambiente para entender a necessidade da área, identificar qual seria o enfoque, quais eram as atividades realizadas pela equipe, e como estava disponível o histórico da base de dados, tal como sua permissão para uso.

Com o histórico de chamados, foi possível confirmar através do gráfico de Pareto que mais de 50% das demandas tem origem de áreas críticas da empresa, relacionada aos recursos financeiros e humanos, conforme pode ser visto na Figura 3.

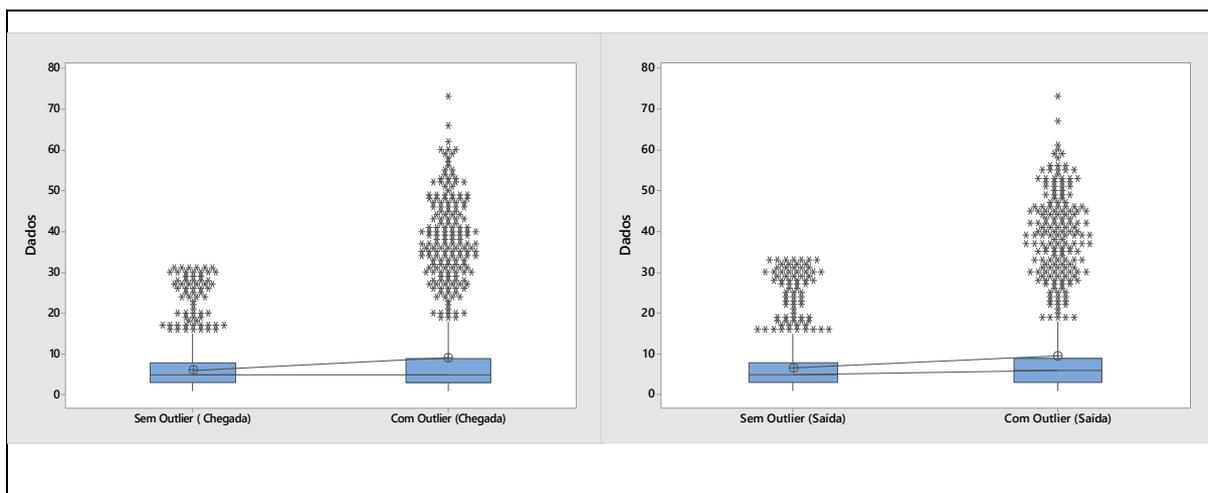
Figura 3 - Pareto das áreas com maior demanda

Fonte: Elaborado pelos autores.

Dentro dos dados observados, utilizou-se apenas chamados menores a 33 dias, ou seja, todos os chamados abertos por mais de 33 dias são finalizados automaticamente, considerando que não há desistência no sistema. Tal decisão foi tomada pela área estratégica da empresa devido a parametrização do sistema ter sido implementada no início do ano de 2017.

Com a retirada dos chamados com mais de 33 dias, houve uma redução de aproximadamente 11% aos dados originais. Desta forma, organizou-se os dados através da contagem de requisições/dia e pelo total de requisições resolvidas/dia. Para a comprovação da existência de valores outliers, empregou-se o gráfico Boxplot do MINITAB®, para os dados de chegada e de saída, para visualização de como os dados se comportavam em relação à média, conforme pode ser visto na Figura 4. Foram retirados dos dados de chegada 1468 valores e 1407 das saídas, eliminando-se um total de 9% e 8% respectivamente do total de dados que foram classificados como outliers.

Figura 4 - Boxplot dos dados de chegada e saída com e sem outliers

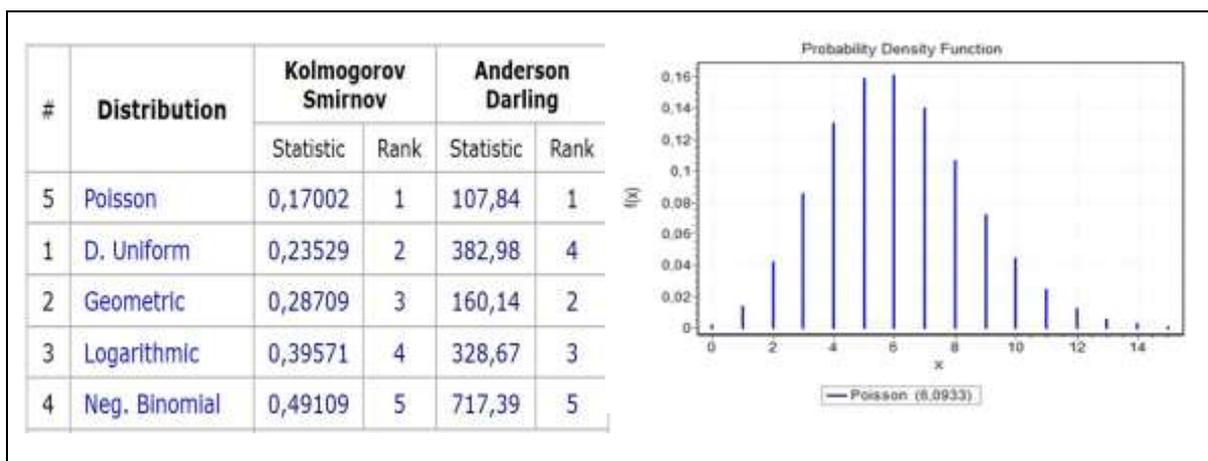


Fonte: Elaborado pelos autores com o auxílio do MINITAB®.

Já para a verificação do modelo, fez-se o input dos dados no software EasyFit®, visando obter as taxas de chegada e de saída, e em conjunto descobrir qual o tipo de distribuição de probabilidade com melhor aderência frente aos dados disponibilizados.

Na Figura 5 tem-se o retorno do software EasyFit® em relação ao teste de aderência Kolmogorov-Smirnov e de Anderson-Darling para os dados das chegadas. Conforme pode ser visto os dois testes resultaram na distribuição de Poisson como a de melhor aderência e média igual a 6,10. Além da Poisson, outras distribuições foram testadas, porém não tão significativas quanto aos resultados expostos.

Figura 5 - Teste de aderência e Distribuição de probabilidade para os dados de chegada

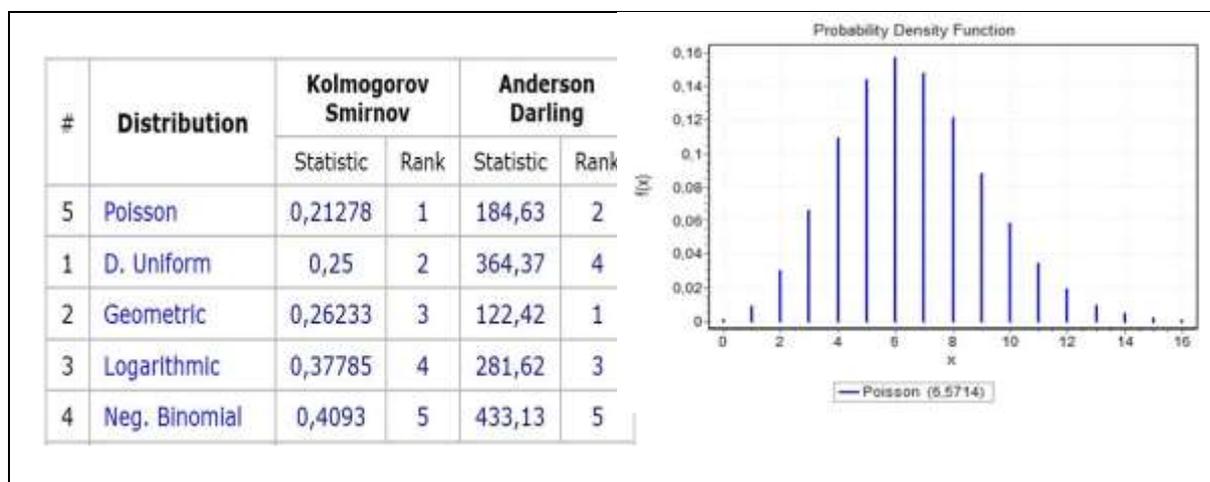


Fonte: Software EASYFIT®.

A Figura 6 apresenta o resultado do teste de aderência obtido no software EasyFit® em relação aos dados das saídas. Conforme pode ser visto, o teste de Kolmogorov-Smirnov apresentou a distribuição de Poisson com a melhor aderência e para o teste de Anderson-Darling a distribuição

geométrica. Nesse caso como não houve concordância entre os dois testes deu-se preferência ao resultado apontado pelo teste K-S, uma vez que a distribuição de Poisson possibilita a resolução analítica do modelo envolvido.

Figura 6 - Teste de aderência e Distribuição de probabilidade para os dados de saída



Fonte: Software EASYFIT®.

Com a utilização desta distribuição, obteve-se uma média de 6,57, e como a área de processos envolve o trabalho de quatro servidores, fez-se necessário dividir a média obtida pelo número de servidores para se obter a distribuição de tempo de serviço de um único servidor. Nesse caso uma distribuição de Poisson com parâmetro (média) 1,63. O modelo mais adequado para esta situação foi o $M/M/s/GD/\infty/\infty$, isto é, com chegadas Poisson e tempos de atendimento exponencial e s servidores atuando em paralelo.

4 Resultados e Discussões

Nesta etapa apresenta-se o conjunto de cenários que foram simulados após a definição do modelo de filas, empregando-se o modelo $M/M/s$ com apoio do software Excel®, no intuito de se ter o resultado dos principais parâmetros de interesse.

Foram simulados cenários com a quantidade atual de servidores (quatro) e posteriormente com mais servidores. Do mesmo modo, fez-se a simulação de um número reduzido de servidores, porém os resultados refletiram em um sistema não ergódico, isto é, se fosse utilizado três servidores ou menos, a fila cresceria indefinidamente. Assim esses resultados não foram considerados no presente estudo. Na Tabela 1 é possível verificar os resultados dos parâmetros da simulação com o número de servidores igual a quatro (condição atual), até nove servidores, visto que as simulações com resultados da inserção de servidores maiores a nove não foram colocados na tabela abaixo, pois apenas mantinham sequencia de aumento ou redução dos parâmetros já esboçados na tabela abaixo.

Tabela 1 - Resultado dos parâmetros atuais e outros cenários

Parâmetros ($\lambda = 6,1$; $\mu = 1,64$)	Atual	Cenários				
Nº de Servidores:	4	5	6	7	8	9
Ocupação:	0,93	0,74	0,62	0,53	0,46	0,41
Ociosidade:	0,07	0,26	0,38	0,47	0,54	0,59
Probabilidade de Iniciar uma fila:	0,84	0,45	0,22	0,10	0,04	0,02
Média de requisições na fila:	11,00	1,00	0,35	0,11	0,03	0,01
Média de requisições sendo atendidas:	4	4	4	4	4	4
Média de requisições no sistema:	14	5	4	4	4	4
Tempo de Espera:	2	0,21	0,06	0,02	0,01	0,00
Tempo de atendimento:	1	1	1	1	1	1

Fonte: Elaborado pelos autores.

Avaliando-se os resultados dos cenários é possível perceber que a taxa de ocupação do sistema sofre uma redução acentuada com o acréscimo do número de servidores, criando um aumento da ociosidade do sistema para 26%, frente aos 8% anteriores. Com cinco servidores a ociosidade salta para 38% e o tempo de espera das requisições passa de dois (atual) para zero dias com o acréscimo de um ou dois servidores, significando que a possibilidade de criar filas dentro do sistema se torna inexistente em aspectos relacionados a condição temporal dia, porém no que se refere as horas daquele dia, pode sim existir uma fila. O tempo de atendimento e o número médio de requisições sendo atendidas não se altera em termos diários, mas apenas valores em termos de horas, lembrando que os chamados abertos pelas áreas podem ter diferentes níveis de dificuldade para resolução; o número médio de requisições no sistema reduz significativamente de 14 para 4 requisições com a variação de quatro para seis servidores. O número médio de requisições na fila se anula em termos de dia, mas apenas em valores em termos de horas, visto que com a adição de servidores, o atendimento das requisições se torna mais efetivo; a probabilidade de o sistema iniciar uma fila cai de 84% para aproximadamente 22% com o aumento do número de servidores. Nota-se, ainda, que com o aumento do número de servidores, a fila passa a se formar apenas com um número maior de requisições, sendo que na condição atual quando são inseridas quatro requisições já existe a incidência imediata de fila.

Ressalta-se ainda que não foram considerados os atributos de prioridades para atendimento de chamadas que já está sendo utilizado pela empresa e, também, não serão analisados os impactos econômicos e financeiros do modelo.

5 Considerações finais

O trabalho desenvolvido teve como objetivo principal propor um dimensionamento de servidores na área de TI por intermédio da utilização da teoria de filas, avaliando as condições atuais do sistema e sugerindo melhorias, e assim o fez.

Através das entrevistas com a equipe compreendeu-se como eram desenvolvidas as atividades, como o setor estava dividido e qual seria a subárea de enfoque. E com o auxílio da base de dados disponibilizada e ferramentas estatísticas, disponibilizou para a empresa o embasamento quanto a origem das demandas e impactos no sistema de forma geral no que se refere a decisão de contratações da área.

A análise da base de dados disponível possibilitou também identificar e classificar por meio de um gráfico de Pareto que as áreas de custos, recursos humanos e cobrança eram as que mais demandavam serviços relacionados à TI, visto sua interação com as demais áreas da empresa.

A metodologia trabalhada neste estudo permitiu ajustar os dados fornecidos para um modelo M/M/s. Já o emprego de softwares facilitou os ajustes das distribuições das bases de dados dos valores das chegadas de requisições e das requisições atendidas, sendo ainda factível a retirada de *outliers* que impactariam de forma negativa os resultados do sistema analisado.

Conseguiu-se através de embasamento teórico identificar o quão importante se faz a caracterização e compreensão das variáveis de teoria das filas e as relações entre elas, junto a conhecimento de aplicações distintas para confirmar a efetividade do uso da mesma em problemas onde se tem a espera como fator principal para correção.

Com uma avaliação de cenários foi verificado que as condições atuais de serviço, isto é, o número atual de servidores é o ideal. Se um número inferior a quatro servidores fosse utilizado o sistema não seria ergódico, isto é, a fila cresceria indefinidamente, bloqueando o sistema. Já um acréscimo do número de servidores de quatro para cinco poderia ser utilizado, pois isso reduziria a fila atual de dois dias para menos de um dia. Nesse caso a ociosidade do sistema que atualmente é de 8% aumentaria para 26%. Isso não é de todo ruim, visto que um dos principais resultados da teoria das filas é justamente mostrar que um sistema deve ter uma relativa ociosidade para evitar gargalos e grandes filas. E desta forma, foi possível compreender que para casos onde o risco de parada na prestação de serviço for significativo junto ao modelo de negócio aplicado pela empresa, é interessante manter certo nível de ociosidade, visto que os riscos factíveis de ocorrer podem impactar de forma negativa no andamento de todo o negócio.

Como propostas de melhoria, sugeriu-se para o coordenador da área elaborar treinamentos mais efetivos para equipe do nível I com o intuito de criar autonomia e aumentar o nível de responsabilidade desta subárea, com melhor utilização do filtro de ações frente à resolução de problemas que chegam diariamente para o primeiro atendimento, deixando para a área de processos

casos de alta complexidade e criticidade. Junto a isso, foi sugerido ainda que existiria a necessidade de implantar um campo obrigatório no canal de atendimento para que o problema fosse melhor descrito pela área demandante, e que ao solucionar o problema, o mesmo fosse caracterizado pelos servidores, com suas principais ações e locais utilizados, sob a forma de gerar um histórico que pudesse ser consultado futuramente.

Para trabalhos futuros, sugere-se a avaliação dos riscos em aspectos de operação, ou seja, na tomada de decisão de atender chamados por grau de prioridade quanto ao tipo de requisição e ao nível de risco para a empresa. Outra sugestão de estudo se aplica em avaliar os custos relacionados ao tempo de espera para resolução dos chamados, levando em consideração a variável escolaridade e nível de qualidade para atendimento das requisições.

Referências

- Associação Nacional de Jornais [ANJ]. Relatório de Atividades 2014 a 2016, 2016, 58 p. Disponível: <https://www.anj.org.br/site/images/pdf/relatividades/RELATO%CC%81RIO-DE-ATIVIDADES-2014-a-2016.pdf>
- Bertucci, J. L. D. O. (2009). Metodologia básica para elaboração de trabalhos de conclusão de cursos (TCC): ênfase na elaboração de TCC de pós-graduação Lato Sensu. *São Paulo: Atlas, 1*.
- Boucher, J. P., & Piqué, G. C. (2015). Modeling the number of insureds' cars using queuing theory [Eletronic version], *Insurance: Mathematics and Economics*, 64, 67-76.
- Chase, R. B., Aquilano, N. J., & Jacobs, F. R. (2006). *Administração Da Produção Para a Vantagem Competitiva*. Bookman.
- Conceição, K.; Loureiro, F. M.; Gonçalves, M. B. (2009). Análise de capacidade de atendimento através de aproximação exponencial de uma central telefônica operando com agentes de múltiplas habilidades. XLI-SBPO-Pesquisa Operacional na Gestão do Conhecimento.
- Melo, D. R. A. de. (2008). A importância da tecnologia da informação nas estratégias das organizações contemporâneas: breve revisão de literatura. V CONVIBRA – Congresso Virtual Brasileiro de Administração.
- Farias, E., Braidó, G. M., Casalinho, G. D. O., & Muniz, R. J. (2012). Os investimentos em TI, além do retorno financeiro: um estudo de caso sobre benefícios intangíveis da implantação de TI no processo de trabalho de uma empresa [Versão eletrônica], *RASM- Revista Acadêmica São Marcos*, 1(1), 9-26.
- Gil, A. C. (2008). *Métodos e técnicas de pesquisa social*. 6. ed. Editora Atlas SA.
- Chwif, L., & Medina, A. C. (2006). *Modelagem e simulação de eventos discretos*.
- Hillier, F. S., & Lieberman, G. J. (2013). *Introdução à pesquisa operacional*. McGraw Hill Brasil.
- Kleinrock, L. (1975). *Queueing Systems: volume 1: Theory*. New York: John Wiley.
- Liu, D., & Ge, Y. E. (2017). Modeling assignment of quay cranes using queueing theory for minimizing CO2 emission at a container terminal [Eletronic version], *Transportation Research Part D: Transport and Environment*, 51 (1), 01-12.
- Lu, X., Tian, R., & Guan, S. (2012). Medical equipment utility analysis based on queuing theory [Eletronic version], *Journal of computers*, 7 (9), 2232-2239.
- Novaes, A. G. N. (1975). *Pesquisa Operacional e Transportes: Modelos Probabilísticos*. Universidade de São Paulo; McGraw-Hill.
- Perdoná, I. I., Nunes, R. V., das Neves, R. M., Naimer, S. C., & Godoy, L. P. (2017). Sistema de manufatura: otimização de processos em uma unidade fabril de cimento através da teoria das filas. *Exacta*, 15(4), 13-25.
- Perin, C., Filho. (1995). *Introdução à simulação de sistemas*. Campinas, SP: Editora Unicamp.
- Prado, D. S. D. (1999). *Teoria das Filas e da Simulação*. Belo Horizonte, MG: Editora de Desenvolvimento Gerencial, 2.

- Sampaio, J. G., & Borschiver, S. (2008). Analysis of Patent Examination Effort Distribution based on the Queuing Theory [Eletronic version], *Journal of Technology Management & Innovation*, 3 (4), 01-16.
- Sant'Anna, L., Filho. (2007). O jornal da sociedade da informação: como a folha, o Globo, e o Estado respondem às inovações tecnológicas, à queda de leitura e à concorrência com novos meios. Tese de Mestrado, Universidade de São Paulo, SP. Disponível: <http://www.teses.usp.br/teses/disponiveis/27/27154/tde-22072009-183602/en.php>
- Santos, G. S., & Campos, F. C. (2011) Operação de serviços de TI: uma abordagem do dimensionamento de recursos e dos incentivos financeiros, [Versão eletrônica], *Revista Produção Online*, 11(4), 1182-207.
- Schein, D. (2010). Uma metodologia para o dimensionamento de frota de rebocadores em terminais portuários: uma aplicação ao Porto do Rio Grande. Tese de mestrado, Universidade Federal do Rio Grande, Rio Grande, RS. Disponível: <http://repositorio.furg.br/handle/1/3359>
- Shannon, R. E. (1975). Simulation: a survey with research suggestions. *AIIE Transactions*, 7(3), 289-301.
- Silva, E. L.; Menezes, E. M. (2005). A pesquisa e suas classificações. _____. *Metodologia da pesquisa e elaboração de dissertação*, 3, 19-23.
- Silva, V. M. D.; Souza, R. A. De; Bortolotti, S. L. V.; Coelho, A. S. (2006). Teoria das filas aplicada ao caso: Porto de Itajaí-SC. *SIMPÓSIO DE ENGENHARIA DE PRODUÇÃO DA UNESP*, 8, 1-9.
- Singer, M., & Donoso, P. (2008). Assessing an ambulance service with queuing theory [Eletronic version], *Computers & Operations Research*, 35 (8), 2549-560.
- Taha, H. A. (2008). *Pesquisa Operacional: uma visão geral* 8ª ed. São Paulo: Pearson Prentice.
- Taufembach, C., & Silva, S. Da (2012). Queuing theory applied to the optimal management of bank excess reserves [Eletronic version], *Physica A: Statistical Mechanics and its Applications*, 391(4), 1381-1387.
- Yu, M., Gong, J., & Tang, J. (2016). Optimal design of a multi-server queuing system with delay information [Eletronic version], *Industrial Management e Data systems*, 116(1), 147-169.
- Zhou, H., Guo, G., & Wu, B. (2013). Nanjing Yangtze River Bridge transit capacity based on queuing theory [Eletronic version], *Procedia Social and Behavioral Sciences*, 96, 2546-2552.

Recebido em: 16 out. 2018 / **Aprovado em:** 01 abr. 2019

Para referenciar este texto

American Psychological Association (APA)

Zanardi, R., & Viali, L. (2020). Teoria de filas e o dimensionamento de servidores: um estudo de caso. *Exacta*, 18(3), 489-501. <https://doi.org/10.5585/ExactaEP.v18n3.10770>.