

UNIVERSIDADE FEEVALE

NICOLAS LAND

**POLÍTICA ADAPTATIVA DE ATENDIMENTO DE FILAS: UMA
ABORDAGEM BASEADA EM MULTIAGENTE**

Novo Hamburgo

2016

NICOLAS LAND

POLÍTICA ADAPTATIVA DE ATENDIMENTO DE FILAS: UMA
ABORDAGEM BASEADA EM MULTIAGENTE

Trabalho de Conclusão de Curso
apresentado como requisito parcial
à obtenção do grau de Bacharel em
Ciência da Computação pela
Universidade Feevale

Orientadora: Me. Adriana Neves dos Reis

Novo Hamburgo

2016

AGRADECIMENTOS

Agradeço a todos os que, de alguma maneira, contribuíram para a realização deste trabalho de conclusão.

Em especial, aos abaixo citados.

À professora Me. Adriana Neves dos Reis, pelo apoio e incentivo dado à realização deste trabalho.

A todos os professores da Feevale que de alguma maneira contribuíram em minha formação.

E aos meus pais, Leo e Sandra, e ao meu irmão Lucas, pelo apoio e incentivo em todos os momentos.

RESUMO

A Teoria de Filas, tradicionalmente, é uma abordagem eficaz para modelar o funcionamento de sistemas que possuem de forma característica uma demanda por serviços momentaneamente maior do que a capacidade de trabalho disponível. Porém, em alguns casos, mais particularmente quando as demandas requerem tipos de serviços diferenciados ao longo do tempo, a mesma não trata como esses cenários dinâmicos podem ser organizados. Para estas situações, é comum adotar uma estratégia híbrida, utilizando modelos complementares, de modo a melhorar o desempenho das métricas de satisfação dos clientes e de utilização dos recursos. Entre os modelos que podem ser utilizados em conjunto com a abstração de Teoria de Filas estão os Sistemas Multiagente. Estes possuem competências como funcionamento flexível e desenvolvimento das habilidades, o que sugere um potencial relevante em complementar um sistema de filas no contexto anteriormente descrito. Assim, o objetivo deste trabalho é a proposta de uma Política Adaptativa de Atendimento de Filas baseada em Sistemas Multiagente. Para tanto, o sistema utiliza uma abordagem quantitativa, a partir do emprego de técnicas de modelagem e simulação, tendo como contexto de estudo o serviço de atendimento ao aluno em uma universidade. Através da identificação de situações adversas utilizando modelos da Teoria de Filas, foram desenvolvidos uma disciplina e três agentes capazes de atuar de forma concisa nestes contextos. Com uma análise comparativa dos efeitos dos elementos propostos, em várias situações distintas, foi possível demonstrar que a agregação de Sistemas Multiagente aos modelos da Teoria de Filas contribui para a melhoria do desempenho em ambientes de atendimento proporcionalmente à instabilidade deste, ou seja, quanto mais inconstante forem os padrões do sistema, maior a efetividade dos algoritmos híbridos.

Palavras-chave: Política Adaptativa de Atendimento. Teoria de Filas. Sistemas Multiagente.

ABSTRACT

The Queueing Theory is usually an effective approach in modelling systems that have, as main characteristic, a momentarily higher demand than the system's capacity to supply. However, in some cases, particularly when different types of services are required over time, it does not address the organization these dynamic scenarios may have. In these situations it is often used a hybrid strategy with complementary models, so that the client satisfaction and resources utilization metrics performance can be improved. One of the models that can be used in conjunction with the Queueing Theory's abstraction is the Multiagent System. These models have some qualities that suggest them as being potentially relevant in complementing the systems described before, such as dynamic behavior and ability evolving. Therefore, the objective of this paper is the proposal of a Queueing Attendance Dynamic Policy based on Multiagent System. In order to do so, the system uses a quantitative approach, applying modelling and simulation technics, in the context of a university campus's student attendance system. Through the identification of adverse situations using Queueing Theory, it was developed one service discipline and three agents that were able to act concisely in these cases. After comparing the effects of the proposed elements, in various distinct situations, it was possible to demonstrate that adding Multiagent System's resources to Queueing Theory models contributes to better performances proportionally to the attendance environment's instability, that is, the more unstable the system patterns are, the better the results of the hybrid algorithms.

Key words: Attendance Dynamic Policy. Queueing Theory. Multiagent System.

LISTA DE FIGURAS

| | |
|--|----|
| Figura 1 – Representação de um sistema de atendimento | 14 |
| Figura 2 – Ordenação de fila conforme a disciplina | 19 |
| Figura 3 – Exemplo de distribuição Normal..... | 21 |
| Figura 4 – Exemplo de distribuição Exponencial | 21 |
| Figura 5 – Exemplo de distribuição Uniforme..... | 21 |
| Figura 6 – Exemplo de distribuição Erlang | 22 |
| Figura 7 – Exemplo de distribuição Poisson..... | 22 |
| Figura 8 – Representação de um agente em seu ambiente | 27 |
| Figura 9 – Modelo de agente reativo | 29 |
| Figura 10 – Modelo de agente cognitivo | 30 |
| Figura 11 – Processo de desenvolvimento, simulação e análise de resultados..... | 32 |
| Figura 12 – Guia de <i>interface</i> do NetLogo | 33 |
| Figura 13 – Guia de código fonte do NetLogo | 34 |
| Figura 14 – Pseudocódigo da disciplina ADA..... | 40 |
| Figura 15 – Pseudocódigo do Agente de aprendizagem..... | 42 |
| Figura 16 – Pseudocódigo do Agente de priorização..... | 46 |
| Figura 17 – Pseudocódigo do Agente de dimensionamento..... | 50 |
| Figura 18 – Comparação entre demandas dos cenários <i>Destacado</i> e <i>Variável</i> | 55 |
| Figura 19 – Balanceado: Tempo médio de fila X Algoritmo X Tipo de atendimento | 58 |
| Figura 20 – Balanceado: Tempo total de fila do histórico X Algoritmo | 58 |
| Figura 21 – Destacado: Tempo médio de atendimento X Algoritmo X Tipo de atendimento..... | 59 |
| Figura 22 – Destacado: Tempo total de fila do histórico X Algoritmo X Tipo de atendimento | 60 |
| Figura 23 – Destacado: Tempo total de fila do histórico X Algoritmo..... | 60 |
| Figura 24 – Predominante: Tempo médio de atendimento X Algoritmo X Tipo de atendimento | 61 |
| Figura 25 – Predominante: Tempo total de fila do histórico X Algoritmo X Tipo de atendimento | 62 |
| Figura 26 – Predominante: Tempo total de fila do histórico X Algoritmo | 63 |
| Figura 27 – Estável: Tempo médio de atendimento X Algoritmo X Tipo de atendimento..... | 64 |
| Figura 28 – Estável: Tempo total de fila do histórico X Algoritmo X Tipo de atendimento | 65 |
| Figura 29 – Estável: Tempo total de fila do histórico X Algoritmo..... | 65 |
| Figura 30 – Variável: Tempo total de fila do histórico X Algoritmo X Tipo de atendimento | 66 |
| Figura 31 – Variável: Tempo total de fila do histórico X Algoritmo..... | 67 |
| Figura 32 – Tempo total de fila do histórico X Casos de testes X Algoritmos | 69 |

LISTA DE TABELAS

| | |
|---|----|
| Tabela 1 – Distribuições estatísticas..... | 20 |
| Tabela 2 – Símbolos da Notação Kendall..... | 23 |
| Tabela 3 – Notação Kendall de um exemplo de ambiente desenvolvido | 36 |
| Tabela 4 – Diferença entre as prioridades das disciplinas PRI e ADA | 39 |
| Tabela 5 – Exemplo de cálculo das experiências de um atendente com Lh em 100..... | 43 |
| Tabela 6 – Exemplo de sistema de atendimento com problema de priorização | 45 |
| Tabela 7 – Exemplo de atuação do Agente de priorização com tempo referência de 1 hora | 47 |
| Tabela 8 – Exemplo de sistema de atendimento com <i>stress</i> | 49 |
| Tabela 9 – Exemplo de atuação do Agente de dimensionamento com tempo referência de 1 hora | 50 |
| Tabela 10 – Exemplo de combinações de funcionalidades com a disciplina PRI | 52 |
| Tabela 11 – Partilha de chegadas de clientes entre os tipos de atendimento | 56 |
| Tabela 12 – Combinações de algoritmos escolhidas | 57 |
| Tabela 13 – Melhores algoritmos em cada cenário e seus ganhos de desempenho | 68 |

LISTA DE ABREVIATURAS E SIGLAS

| | |
|------|---|
| ADA | Adaptativa |
| Apr | Agente de aprendizagem |
| Dim | Agente de dimensionamento |
| Eb | Experiência base |
| Em | Experiência mínima do atendimento |
| Et | Experiência total do tipo de atendimento |
| Eu | Experiência unitária por atendimento |
| FIFO | <i>First In, First Out</i> |
| Lh | Limite do histórico de experiências |
| LIFO | <i>Last In, First Out</i> |
| Na | Número de atendentes |
| Pm | Percentual mínimo de experiência para cada tipo de atendimento |
| PRI | <i>Priority</i> |
| Pri | Agente de priorização |
| Qa | Quantidade de atendimentos recentes |
| Qf | Quantidade atual em fila do tipo |
| Qt | Quantidade de atendimentos recentes do tipo de atendimento |
| Qu | Quantidade de clientes no histórico do tipo |
| SIRO | <i>Served In Random Order</i> |
| SMA | Sistema Multiagente |
| Ta | Tipos de atendimento |
| Td | Tempo necessário para decremento |
| Tg | Tempo total estimado em fila para todos os tipos de atendimento |
| Ti | Tempo necessário para incremento |
| Tr | Tempo de referência |
| Tt | Tempo total estimado atualmente em fila do tipo de atendimento |
| Tu | Tempo em fila de clientes no histórico do tipo de atendimento |

SUMÁRIO

| | | |
|----------|--|-----------|
| 1 | INTRODUÇÃO..... | 11 |
| 1.1 | OBJETIVOS..... | 11 |
| 1.1.1 | Objetivo geral | 11 |
| 1.1.2 | Objetivos específicos | 12 |
| 1.2 | JUSTIFICATIVA | 12 |
| 1.3 | ESTRUTURA DO TRABALHO | 12 |
| 2 | SISTEMAS DE ATENDIMENTO | 13 |
| 2.1 | CARACTERÍSTICAS DO AMBIENTE..... | 13 |
| 2.1.1 | Serviços prestados..... | 15 |
| 2.1.2 | Atendentes..... | 16 |
| 2.2 | GERENCIAMENTO DE DEMANDAS | 16 |
| 3 | TEORIA DE FILAS..... | 18 |
| 3.1 | PROPRIEDADES..... | 18 |
| 3.2 | DISTRIBUIÇÕES | 20 |
| 3.3 | NOTAÇÃO KENDALL | 23 |
| 3.4 | APLICAÇÕES..... | 24 |
| 3.5 | LIMITAÇÕES | 25 |
| 4 | AGENTES | 27 |
| 4.1 | CARACTERÍSTICAS | 28 |
| 4.2 | CLASSIFICAÇÕES | 29 |
| 4.2.1 | Agentes reativos | 29 |
| 4.2.2 | Agentes cognitivos..... | 30 |
| 4.3 | DESTAQUES..... | 31 |
| 5 | DESENVOLVIMENTO | 32 |
| 5.1 | FERRAMENTA DE SIMULAÇÃO | 33 |
| 5.2 | ESCOPO DA TEORIA DE FILAS | 34 |
| 5.2.1 | Algoritmos..... | 35 |
| 5.2.2 | Testes e identificação de melhorias | 36 |
| 5.3 | POLÍTICA ADAPTATIVA | 37 |
| 5.3.1 | Disciplina..... | 38 |
| 5.3.2 | Agentes | 40 |
| 5.3.2.1 | Agente de aprendizagem..... | 41 |
| 5.3.2.2 | Agente de priorização | 43 |
| 5.3.2.3 | Agente de dimensionamento..... | 48 |
| 5.4 | AMBIENTE DE SIMULAÇÃO FINAL | 51 |
| 6 | EXPERIMENTOS | 53 |
| 6.1 | CASOS DE TESTE | 53 |

| | | |
|-------|--|-----------|
| 6.1.1 | Chegada de clientes..... | 54 |
| 6.1.2 | Atendimento das demandas..... | 56 |
| 6.2 | SIMULAÇÕES..... | 56 |
| 6.3 | RESULTADOS | 57 |
| 6.3.1 | Caso de teste 1 – Balanceado..... | 57 |
| 6.3.2 | Caso de teste 2 – Destacado | 59 |
| 6.3.3 | Caso de teste 3 – Predominante..... | 61 |
| 6.3.4 | Caso de teste 4 – Estável | 64 |
| 6.3.5 | Caso de teste 5 – Variável..... | 66 |
| 6.4 | ANÁLISE DOS RESULTADOS | 68 |
| 6.4.1 | Disciplina..... | 69 |
| 6.4.2 | Agente de aprendizagem | 70 |
| 6.4.3 | Agente de priorização | 70 |
| 6.4.4 | Agente de dimensionamento..... | 71 |
| 7 | CONCLUSÃO..... | 72 |
| 7.1 | TRABALHOS FUTUROS..... | 73 |
| | REFERÊNCIAS BIBLIOGRÁFICAS..... | 75 |

1 INTRODUÇÃO

Na atualidade, cada vez mais as pessoas estão envolvidas com os ambientes de atendimentos de empresas. Tais sistemas são oferecidos para facilitar o atendimento a clientes, porém nem sempre é pensado neles quando o planejamento de tais ambientes é desenvolvido. Isto traz efeitos indesejados para ambos os lados, como companhias perdendo clientes por insatisfação e clientes perdendo horas por um atendimento relativamente simples.

A computação possui uma área de pesquisa relacionada ao gerenciamento de atendimento de demandas e filas, a Teoria de Filas. Este tipo de modelo não é específico para problemas da computação, sendo aplicável nos problemas de demandas de diversas áreas do conhecimento. Esta generalidade possui como ponto positivo a cobertura da maioria dos problemas de demandas encontrados, porém também gera pontos negativos, sendo o principal destes a simplicidade necessária para a própria generalização. Esta simplicidade não permite implementações necessárias em algumas áreas, como por exemplo o tratamento de eventos críticos e sazonais em um ambiente de atendimento.

Na área de Inteligência Artificial existe uma necessidade de simular o comportamento humano, ou pelo menos uma parte dele. As principais questões que são trabalhadas nestas simulações são o aprendizado, a comunicação, a pró-atividade, a cognição e a heterogeneidade. Englobando estas questões, surgiu a modelagem conhecida como Agentes, os quais são elementos com características que se assemelham ao ser humano em termos de sistematização de inteligência, podendo ser inseridos em diversos ambientes para atuação, principalmente aqueles que requerem adaptação ao longo de sua execução. Assim, o presente trabalho investiga as contribuições de agentes para modelos baseados em filas.

1.1 OBJETIVOS

1.1.1 Objetivo geral

O objetivo deste trabalho é desenvolver uma política de atendimento de filas, utilizando uma abordagem híbrida entre Teoria de Filas e Sistemas Multiagente, com a capacidade de adaptação da política de atendimento conforme as necessidades e competências dos agentes prestadores do serviço.

1.1.2 Objetivos específicos

- Realizar um levantamento bibliográfico sobre Teoria de Filas e Sistemas Multiagente;
- Desenvolver um modelo híbrido para atendimento no contexto do trabalho;
- Desenvolver uma política adaptativa de atendimento de filas utilizando uma abordagem híbrida entre as teorias propostas;
- Verificar o comportamento da política criada;
- Simular cenários com e sem a aplicação da política gerada;
- Avaliar e comparar os resultados gerados;
- Indicar casos de recomendação da política adaptativa proposta.

1.2 JUSTIFICATIVA

O comportamento dinâmico, contudo, pouco variável, que se obtém ao utilizar somente a Teoria de Filas, no tratamento de questões relacionadas a sistemas de atendimento, acaba introduzindo limitações aos mesmos. Estas limitações, muitas vezes, são vistas como deficiências do sistema pelos clientes, gerando reclamações quando ocorrem com frequência considerável. Em virtude disto, o presente trabalho almeja desenvolver uma política de atendimento de filas que se adapte à maioria das situações que geram tais desconfortos neste ambiente em função da aleatoriedade da demanda, tendo como base de desempenho um conjunto de métricas pré-definidas pelos gestores de um determinado processo.

1.3 ESTRUTURA DO TRABALHO

O presente trabalho é constituído de sete capítulos, contendo a introdução, a fundamentação teórica, o desenvolvimento, a experimentação e a conclusão. Ao final são apresentadas as referências bibliográficas utilizadas na elaboração do estudo.

No primeiro capítulo é realizada uma introdução ao assunto abordado no trabalho, são apresentados os objetivos e as justificativas para a elaboração do mesmo. O segundo capítulo engloba a fundamentação teórica dos sistemas de atendimento atuais, enquanto que no terceiro capítulo é explanada a Teoria de Filas e no quarto capítulo são abordados os Agentes e Sistemas Multiagente. No quinto capítulo é descrito o processo de desenvolvimento deste estudo, enquanto que o sexto consiste nos experimentos com os algoritmos desenvolvidos. Por fim, o sétimo capítulo contém as conclusões do estudo e possíveis trabalhos futuros.

2 SISTEMAS DE ATENDIMENTO

Na busca pelo crescimento, tanto organizacional como lucrativo, as empresas têm tendência em investir na melhoria dos seus processos, aperfeiçoando tempos e custos de todas as maneiras possíveis. Segundo Doile (2010), as empresas que compreendem e adequam a sua capacidade produtiva possuem uma vantagem significativa em relação a seus concorrentes que trabalham em níveis não aperfeiçoados, tanto acima como abaixo da capacidade.

Um dos ambientes que costuma sofrer com esta prática de adequação são os sistemas de atendimento a clientes. Nestes casos, as ações realizadas para a melhoria do processo acabam, em sua maioria, interferindo diretamente no funcionamento do sistema, como a diminuição do número de atendentes ou a troca dos mesmos por outros menos experientes, porém mais baratos. Entretanto, as consequências que estas mudanças podem trazer ao sistema devem ser tratadas com cautela, pois os ambientes de atendimento são a representação da qualidade da empresa prestadora do serviço (FIGUEIREDO; ROCHA, 2010).

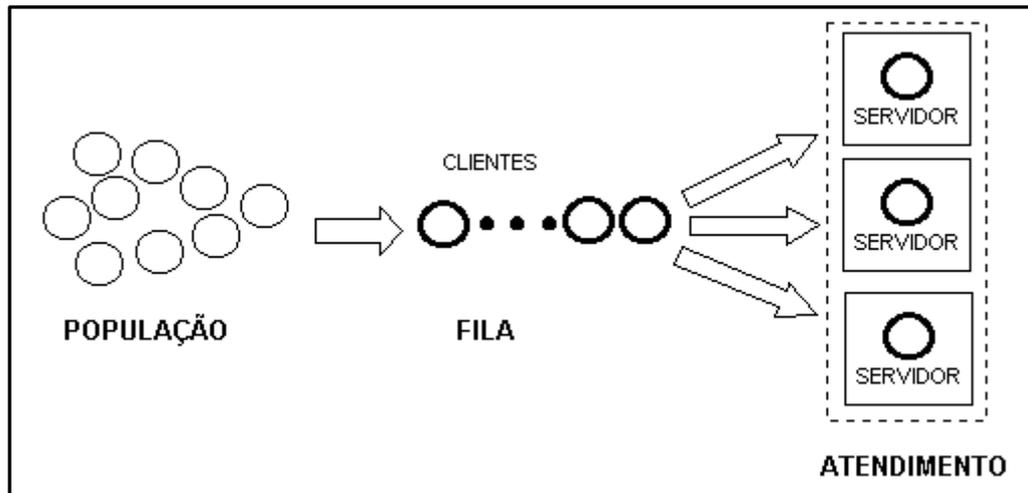
Neste capítulo são apresentados os conceitos básicos de sistemas de atendimento, suas características e os elementos neles envolvidos. Por fim, são abordados os seus métodos de gerenciamento.

2.1 CARACTERÍSTICAS DO AMBIENTE

Um ambiente de atendimento, em meios tradicionais, pode ser descrito como um sistema que possui um ou mais servidores de prestação de serviços e clientes que desejam a aquisição destes serviços, os quais chegam em momentos não específicos (aleatórios). Os ambientes de atendimento podem ter compostos por diferentes estruturas, oferecendo diversos tipos de serviços distintos, com diferentes tamanhos e complexidades (FIGUEIREDO; ROCHA, 2010).

A Figura 1 apresenta um sistema de atendimento tradicional, onde múltiplos clientes demandam o atendimento de suas requisições no mesmo intervalo de tempo, formando uma fila. Para atender estas demandas, o ambiente possui canais de atendimento, os servidores, que atendem os clientes em fila seguindo regras pré-definidas no sistema.

Figura 1 – Representação de um sistema de atendimento



Fonte: <http://www.ebah.com.br>

Bruns, Soncim e Sinay (2001) definem sucintamente o processo de atendimento como sendo a chegada de clientes a um determinado ambiente e o atendimento de sua solicitação de serviço, podendo este ser feito individualmente ou coletivamente. Já Pilares (1989, p.72) indica que “a relação interpessoal que envolve o funcionário de uma organização e o destinatário de produtos, ideias ou serviços desta organização [...] constitui o atendimento ao cliente”. Ainda segundo o autor, normalmente o atendimento também considera o acompanhamento e o encaminhamento dos clientes dentro do ambiente.

Para os clientes, um sistema de atendimento falho deixa uma impressão ruim sobre todo o resto da empresa, gerando queixas, desistências e, em alguns casos, propaganda negativa da empresa pelos consumidores (KOOLE; MANDELBAUM, 2001). Segundo Pilares (1989), o cliente normalmente espera um atendimento diferenciado, sendo reconhecida a importância tanto de sua pessoa como de sua solicitação. Ou seja, clientes gostam de ser tratados como “únicos”, sendo o centro das atenções quando são atendidos e tendo suas requisições sendo atendidas prontamente.

Os prestadores de serviço, também conhecidos como *atendentes*, têm a responsabilidade de satisfazer as necessidades dos clientes da melhor maneira possível. Para obter os resultados desejados, este processo precisa possuir um conjunto de regras que deve ser respeitado por todos os envolvidos. Porém, por cada atendimento ser único, com situações e variáveis diferenciadas, as regras que definem o processo não podem ser “engessadas”, pois se não fica inviável o atendimento correto em todas as situações de demandas (PILARES, 1989).

Os processos de atendimento nos modelos tradicionais de filas não preveem mudanças de comportamento nas variáveis do sistema, tendo assim comportamento não adaptável. Ou seja, a modelagem utilizada considera que os clientes podem chegar ao ambiente em tempos diferentes, porém a política de tratamento é a mesma para todos; ou então não considera que os tempos de atendimento variam momentaneamente conforme a experiência dos atendentes. Este funcionamento faz com que ambientes sejam, muitas vezes, superdimensionados, sendo modelados para suportar os intervalos de pico das demandas e gerando desperdício de recursos nos outros momentos (BRUNS; SONCIM; SINAY, 2001).

2.1.1 Serviços prestados

Os ambientes de atendimento são caracterizados pelo fornecimento de prestação de serviços aos seus clientes. Conforme Oliveira e Favaretto (2013, p. 2), “serviço é a soma de elementos que se juntam para criar uma experiência inesquecível, em um determinado ponto no tempo, para um cliente”. O serviço prestado varia em cada sistema, pois cada um possui um objetivo diferente, como, por exemplo, lojas de produtos, que visam à oferta e venda de produtos através do atendimento ao cliente, e *check-ins* de aeroportos, que têm o intuito de registrar a presença do cliente. Koole e Mandelbaum (2001) afirma que os serviços prestados são de difícil mensuração, ainda mais quando envolvem fatores humanos, pois mesmo os estudos mais simples requerem a análise de diferentes dados complexos.

Segundo Leal (2003), um serviço possui quatro propriedades fundamentais que são necessários para diferenciar os mesmos de produtos:

- Intangibilidade: É algo que não pode ser tocado, somente presenciado;
- Inseparabilidade: Não há como separar ou quebrar o serviço, devendo ele ser prestado por inteiro;
- Variabilidade: Podem ser prestados de forma desigual aos olhos de observadores externos, variando conforme uma série de fatores, como atendente e cliente;
- Perecibilidade: Não há a possibilidade de estocar, manter guardado para uso futuro, pois deve ser realizado no momento da negociação.

Em sua pesquisa, Doile (2010) argumenta que os serviços prestados a um cliente devem atender somente às mínimas exigências de qualidade, diminuindo ao máximo o tempo de sua prestação e, assim, contribuindo para uma percepção positiva do cliente perante o ambiente.

Porém esta afirmação objetiva somente a percepção de qualidade do atendimento pelo cliente, que segundo o estudo é o mesmo que não esperar demais na fila por um atendimento. Esta é uma abordagem diferenciada para a melhoria dos ambientes de atendimento, e não é a mesma utilizada no presente trabalho, o qual envolve a melhoria do sistema como um todo, e não somente para a percepção dos clientes.

2.1.2 Atendentes

Os atendentes em sistemas de atendimento são os responsáveis por prestar o serviço requisitado pelos clientes. Estes atendentes, que são pessoas em ambientes tradicionais, possuem várias características que afetam diretamente o desempenho do sistema, como habilidade, eficiência, desgaste, motivação, experiência, absenteísmo, entre outros (LOPES; FERREIRA; SILVA, 2008).

A modelagem tradicional de filas de atendimento não possui nenhum tratamento especial para as características dos atendentes, considerando-os homogêneos, ou seja, que todos eles são iguais. Porém, na realidade, não é isso que acontece, pois como seres humanos, os atendentes possuem características e habilidades únicas, ou seja, são heterogêneos (MANDELBAUM; SAKOV; ZELTYN, 2001).

Em seu estudo, Lopes, Ferreira e Silva (2007) verificaram que há uma grande diferença de desempenho entre atendentes conforme a quantidade de atendimentos realizados, demonstrando assim a característica de *experiência*. Mandelbaum, Sakov e Zeltyn (2001) também enfatizam esta propriedade, explicando que são esperados de um atendente novo tempos maiores de atendimento do que dos mais experientes, e frisa que esta característica afeta muitas decisões nos sistemas de atendimento.

2.2 GERENCIAMENTO DE DEMANDAS

O gerenciamento das demandas existentes em um ambiente de atendimento é muito importante, e se não realizado ou realizado incorretamente, pode levar o sistema ao caos e trazer sérias consequências à empresa gestora. Segundo Slack, Chambers e Johnston (2002) quando o sistema está em estado de equilíbrio, em que a capacidade consegue atender a demanda, a lucratividade é alta e os clientes ficam satisfeitos, enquanto que em sistemas incapazes de suprir tal demanda tais consequências não são obtidas, pelo menos não com o mesmo êxito.

A Teoria de Filas é a abordagem tradicional para desenvolver modelos em casos em que um serviço prestado está sendo procurado com uma demanda maior do que a capacidade de recebimento momentânea (MEDHI, 2002). Este modelo tem por objetivo identificar alternativas de que o processamento das solicitações seja prestado de modo mais satisfatório, de acordo com um conjunto de métricas que caracterizam o desempenho de um dado sistema. Para isto, é realizado “o desenvolvimento de modelos matemáticos que permitam prever o comportamento de sistemas de prestação de serviços” (MARINS, 2011).

A seguir é apresentado o modelo de tratamento de demandas mais utilizado por estes ambientes, a Teoria de Filas. São descritas as suas principais propriedades e a sua lógica de organização.

3 TEORIA DE FILAS

Em um processo que possui maior demanda do que a sua capacidade de processamento, deve haver um controle para que esta demanda seja atendida assim que possível. Conforme Slack, Chambers e Johnston (2002), a capacidade de atendimento de um sistema é o maior nível de atividade que se pode atingir em condições normais de trabalho durante um determinado período de tempo. Segundo Medhi (2002), a capacidade é extrapolada quando um serviço prestado possui demanda maior que a sua capacidade, gerando assim uma sobrecarga no sistema.

Ao ocorrer uma sobrecarga de demanda no sistema de atendimento, não havendo atendentes o suficiente para suprir a quantidade de clientes, acontece uma fila (BOUZADA, 2009). Estas filas têm, então, a função de organizar as solicitações extras que estão sendo exigidas do processo.

Normalmente é utilizado a Teoria de Filas para realizar os controles de demanda, pois é uma teoria simples, de fácil aplicação, eficaz, e que necessita de poucos recursos para ser bem aproveitada (BOUZADA, 2009). Também é sua função a melhoria do processo de atendimento dos serviços prestados, visando, assim, à diminuição das filas de encaminhamentos.

A seguir são apresentadas as características que formam a Teoria de Filas. Inicia-se com o detalhamento das suas propriedades, seguido das distribuições de tempos trabalhadas pela teoria, a notação que simboliza os diferentes tipos de filas e, por fim, um exemplo de sua aplicação e suas limitações.

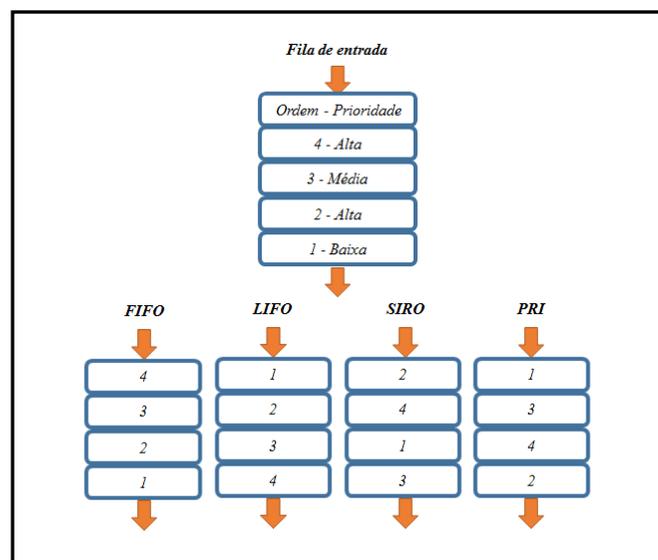
3.1 PROPRIEDADES

Os sistemas de filas são classificados por cinco propriedades, sendo a identificação destas um elemento importante para uma análise correta do ambiente, como definido por Marins (2011), Figueiredo e Rocha (2010), Gross et al. (2008) e Medhi (2002):

- **Chegada:** É a distribuição probabilística de tempo da chegada de clientes à fila. Dependendo do tempo de intervalo entre as chegadas, esta propriedade também pode ser definida como determinística (intervalos fixos) ou aleatória (intervalos diferentes), sendo a segunda definição a mais usual.

- **Atendimento:** Assim como a chegada, também há uma distribuição de tempo de atendimentos realizados, sendo esta contagem também uma propriedade que pode ser considerada como determinística ou aleatória. Em geral esta propriedade é aleatória.
- **Canais:** Representa o número de prestadores de serviços disponíveis simultaneamente para realizar os atendimentos.
- **Capacidade:** Caracteriza o limite do sistema em fornecer atendimento de forma normal, considerando os clientes em atendimento e os que estão na fila. Ao ser extrapolada a capacidade, os novos clientes acabam tendo um tratamento diferenciado, normalmente sendo a recusa da demanda.
- **Disciplina:** Esta propriedade é o conjunto de regras que definem o modelo de atendimento da fila, ou seja, é a maneira de seleção para atendimento dos clientes que estão na fila. Os tipos de disciplina existentes são:
 - **FIFO (*First In, First Out*)** – É conhecida como ordem de chegada, pois é a regra em que o primeiro a entrar é o primeiro a sair.
 - **LIFO (*Last In, First Out*)** – Nesta o primeiro elemento a sair é o último a entrar.
 - **SIRO (*Served In Random Order*)** – O atendimento da fila é realizado através de ordem aleatória, não importando a ordem de entrada.
 - **PRI (*Priority*)** – É realizada uma ordenação para atendimento conforme uma ou mais regras de priorização.

Figura 2 – Ordenação de fila conforme a disciplina



Fonte: do autor

Na Figura 2 podem ser vistas as diferentes saídas das disciplinas de filas, onde uma mesma fila de entrada, com indicação de ordem de chegada e nível de prioridade, é trabalhada pelas quatro disciplinas. Nos modelos FIFO e LIFO, a prioridade do atendimento não altera em nada a ordem de processamento, enquanto que, no PRI, a ordenação considera primeiramente a prioridade dada a cada atendimento, seguindo depois a ordem de entrada com a utilização do modelo FIFO. Já o modelo SIRO não utiliza nenhuma informação sobre a fila de entrada, sorteando o próximo atendimento.

3.2 DISTRIBUIÇÕES

As características de tempo de chegada e tempo de atendimento, em sistemas de atendimento, são fenômenos em sua maioria aleatórios. Chwif e Medina (2010) caracterizam estes modelos de entrada como “modelos probabilísticos responsáveis por representar a natureza aleatória de um dado fenômeno”. Os modelos probabilísticos, também chamados de *distribuições estatísticas*, possuem parâmetros que os caracterizam e diferenciam dos demais, além de domínios com seus limites de atuação.

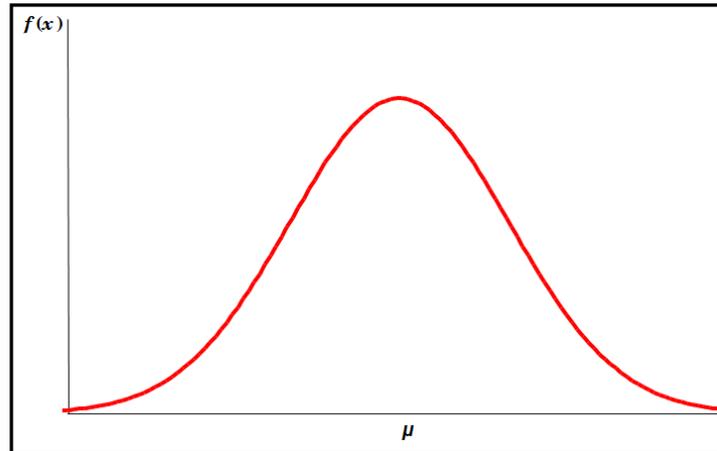
As distribuições mais comuns de serem encontradas em ambientes de atendimento são descritas na Tabela 1, conforme Lopes, Ferreira e Silva (2007), Koole e Mandelbaum (2001) e Mandelbaum, Sakov e Zeltyn (2001). Em seguida, são apresentados os exemplos das distribuições citadas.

Tabela 1 – Distribuições estatísticas

| Distribuição | Parâmetros | Domínio | Exemplo |
|--------------|---|--------------------------------|----------|
| Normal | μ : Média σ : Desvio padrão | Contínua ($0 \leq x$) | Figura 3 |
| Exponencial | λ : Taxa de ocorrências | Contínua ($0 \leq x$) | Figura 4 |
| Uniforme | a : Menor valor b : Maior valor | Contínua ($a \leq x \leq b$) | Figura 5 |
| Erlang | λ : Taxa de ocorrências k : Parâmetro de forma | Contínua ($0 \leq x$) | Figura 6 |
| Poisson | λ : Média | Discreta ($x = 0,1,2,\dots$) | Figura 7 |

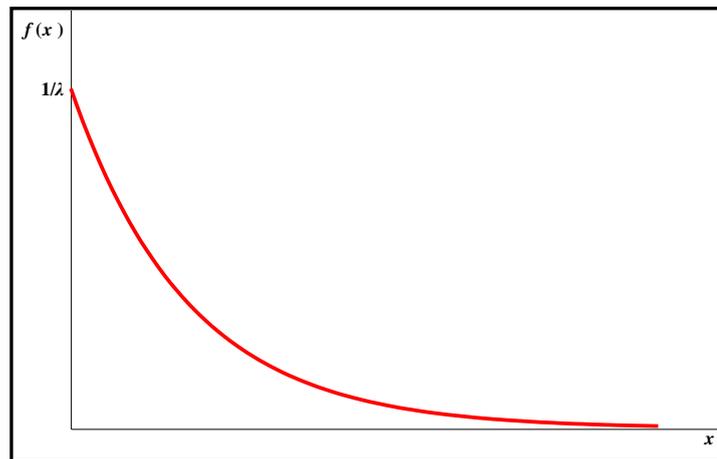
Fonte: do autor, adaptado de Lopes, Ferreira e Silva (2007), Koole e Mandelbaum (2001) e Mandelbaum, Sakov e Zeltyn (2001)

Figura 3 – Exemplo de distribuição Normal



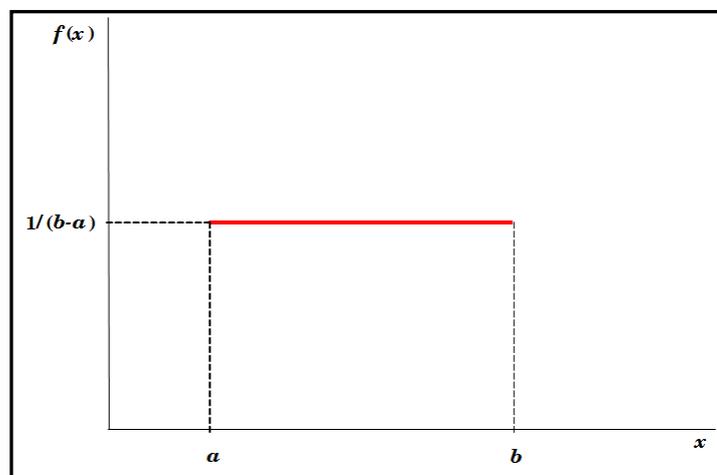
Fonte: Chwif e Medina (2010).

Figura 4 – Exemplo de distribuição Exponencial



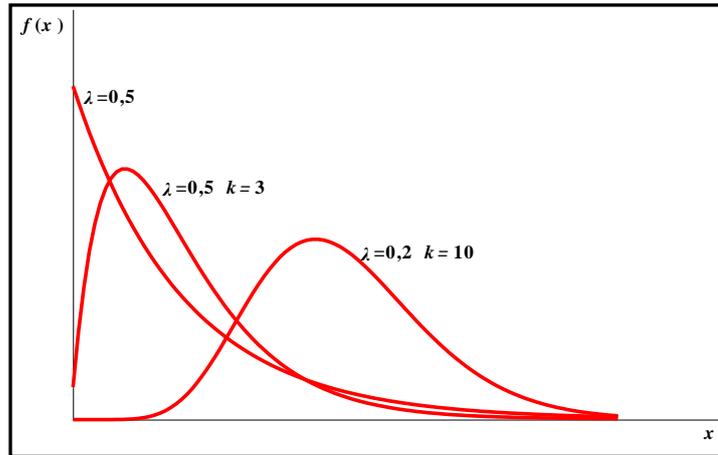
Fonte: Chwif e Medina (2010).

Figura 5 – Exemplo de distribuição Uniforme



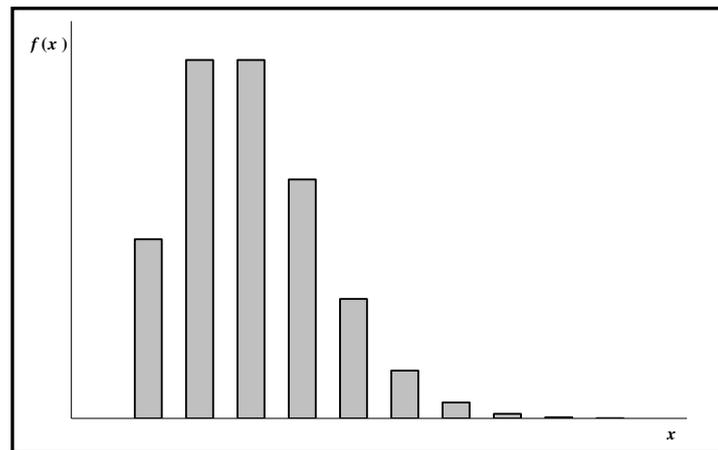
Fonte: Chwif e Medina (2010).

Figura 6 – Exemplo de distribuição Erlang



Fonte: Chwif e Medina (2010).

Figura 7 – Exemplo de distribuição Poisson



Fonte: Chwif e Medina (2010).

A distribuição de tempos de chegada de clientes à fila mais comum é, segundo Doile (2010) e Koole e Mandelbaum (2001), a distribuição Poisson. A distribuição Poisson, segundo Chwif e Medina (2010), modela “o número de eventos independentes que ocorrem em um intervalo de tempo fixo”, e ainda segundo os autores, em uma distribuição Poisson o tempo entre os eventos possui distribuição exponencial.

Para os tempos de atendimento normalmente é utilizada a distribuição exponencial, porém isto ocorre somente por não haver provas que seja outro tipo de distribuição (KOOLE; MANDELBAUM, 2001). Mandelbaum, Sakov e Zeltyn (2001) afirmam que os tempos de serviço não necessariamente são exponenciais, sendo apresentado um caso de tempos de serviço que não se assemelham em nada com uma distribuição exponencial.

3.3 NOTAÇÃO KENDALL

Para a designação de tais propriedades de uma forma mais simplificada, foi desenvolvida uma notação, a qual recebeu o nome de seu principal contribuinte, Kendall (GROSS et al., 2008). Esta notação consiste em simbolizar as propriedades do sistema separando os símbolos por barras, no formato A/B/X/Y/Z. Conforme Medhi (2002), as três características principais dos sistemas são representadas pelos símbolos A/B/X, com a possibilidade de adicionar mais duas, Y/Z, conforme necessidade, sendo descritas como:

- A: Distribuição probabilística dos tempos de chegadas de clientes à fila;
- B: Distribuição probabilística dos tempos de atendimento;
- X: Número de canais prestadores de serviço disponíveis;
- Y: Capacidade do sistema (pode ser omitido);
- Z: Disciplina de filas utilizada (pode ser omitido).

Como exemplo, podemos dizer que um sistema de filas é M/D/3, significando que o intervalo entre chegadas é exponencial, os tempos de atendimento são determinísticos, o sistema possui 3 atendentes simultâneos, não há limite na capacidade da fila e é utilizado o sistema FIFO. Nota-se que as duas últimas notações podem ser omitidas quando possuem os seus valores padrões, o que ocorre no caso exemplificado.

Tabela 2 – Símbolos da Notação Kendall

| Característica | Símbolo | Descrição |
|-----------------------------------|-------------|--|
| | <i>M</i> | Distribuição Exponencial |
| | <i>D</i> | Distribuição Determinística |
| Tempo entre chegadas (A) | <i>Ek</i> | Distribuição Erlang <i>k</i> (<i>k</i> = 1,2,...) |
| Tempo de atendimento (B) | <i>Hk</i> | Mistura de exponenciais <i>k</i> |
| | <i>PH</i> | Tipo fase |
| | <i>N</i> | Normal |
| Número de canais de serviço (X) | 1,2,3,...,∞ | |
| Capacidade do sistema (Y) | 1,2,3,...,∞ | |
| | FIFO | <i>First In, First Out</i> |
| | LIFO | <i>Last In, First Out</i> |
| Disciplina de filas utilizada (Z) | SIRO | <i>Served in Random Order</i> |
| | PRI | <i>Priority</i> |

Fonte: do autor, adaptado de Gross et al. (2008)

A simbologia descrita na Tabela 2 possui alguns elementos já vistos ou auto explanatórios, porém outros não são triviais. É o caso do símbolo N , que indica uma distribuição de probabilidade normal com desvio padrão, enquanto que Erlang k é uma distribuição exponencial da soma de variáveis independentes e aleatórias (GROSS et al., 2008).

3.4 APLICAÇÕES

A Teoria de Filas consiste em trabalhar com as propriedades citadas anteriormente, necessitando do fornecimento de dados sobre o ambiente em que o sistema opera. Segundo Gross et al. (2008) os 3 principais dados que podem ser fornecidos à teoria para melhoria são: (i) uma medida de tempo de espera que um cliente típico aguarda atendimento, (ii) uma indicação de como os consumidores se acumulam, e (iii) uma medição do tempo que os prestadores de serviço ficam ociosos. Ainda conforme os autores, um valor aproximado ou a probabilidade de distribuição destes pode ser utilizado, pois normalmente estes dados são variáveis em sistemas de filas.

A utilização dos dados pelo *analista de filas* acaba tendo normalmente uma de duas finalidades: a modelagem de um sistema “ótimo”, definido através de alguns critérios; ou a determinação de medidas aceitáveis para os dados colhidos, e modelar o sistema para a melhoria destas medidas (GROSS et al., 2008).

Lima e Belderrain (2007) consideram que a aplicação da Teoria de Filas deve ser feita com cuidados, pois dependendo do ambiente e de suas variáveis utilizadas, o sistema poderá ter momentos piores do que o esperado. Isso ocorreria, por exemplo, quando utilizado somente valores médios de tempos em um sistema que possui momentos em que estas médias são muito diferentes, superestimando o sistema para os intervalos de baixa demanda, e subestimando para momentos de pico.

A utilização da Teoria de Filas normalmente se dá com um levantamento do ambiente que a aplicação é desejada, verificando se o mesmo é compatível e montando a Notação Kendall referente ao modelo. Após isso, podem ser elaboradas as equações e os cálculos que serão aplicados à esta configuração do ambiente. A seguir devem ser obtidos os dados que serão trabalhados e aplicados nos cálculos, o que pode ocorrer através de levantamentos no ambiente, simulações, ou outro meio que se achar necessário. Por fim, aplicam-se os dados nas fórmulas e são verificados os resultados que devem ser tratados e/ou obtidos, como mudança no número de atendentes ou a alteração da disciplina de filas.

Um caso de aplicação da Teoria de Filas que segue estes moldes pode ser encontrado no trabalho de Figueiredo e Rocha (2010). Neste estudo, foram levantadas as demandas e os atendimentos realizados em intervalos de 30 minutos, durante 7 semanas de comportamento padrão do ambiente, desconsiderando dias atípicos. Através desta etapa, foi verificado que o ambiente atendia as características da Notação Kendall, sendo então utilizados os dados coletados para calcular o número de canais disponíveis, a cada período de 30 minutos. Os resultados então foram aplicados em modelos da Teoria de Filas e comparados com o sistema já utilizado no ambiente, apresentando um ganho de desempenho sobre o anterior.

3.5 LIMITAÇÕES

A utilização de Teoria de Filas na solução de problemas em sistemas de atendimento é uma solução amplamente empregada atualmente. Porém, conforme as definições da própria teoria, existem algumas limitações intrínsecas a ela na melhoria de algumas métricas (ARAÚJO; ARAÚJO; ADISSI, 2003).

Gross et al. (2008) indicam que o analista de filas pode enfrentar muitos desafios em resolver problemas utilizando métodos analíticos. Como balancear os tempos de espera em fila com os tempos de ociosidade dos prestadores de serviço? Como dimensionar o espaço do ambiente de atendimento, levando em consideração o número de clientes no ambiente (na fila e em atendimento) e o número de canais de serviço? Em alguns casos, o analista acaba tendo que utilizar recursos extras, como simulações.

Bouzada (2009) considera em seu estudo sobre *call centers* que a aplicação de Teoria de Filas nas filas de atendimento “[...] quando bem administradas, podem trazer reduções significativas nos tempos de espera dos clientes que por elas passam”. Porém o autor também releva que existem características que dificultam esta aplicação, como distribuições genéricas do tempo de atendimento, taxas de chegada variantes pelo horário do dia, sobrecargas temporárias e abandonos de demandas.

Já Praia e Gomes (2015) citam em sua pesquisa que a Teoria de Filas, sendo este o modelo analítico de filas mais comum, deve ser usada sempre que possível, porém muitos sistemas de filas atuais são inaptos a serem estudados pelos conceitos contidos nessa teoria. Assim, os autores dão ênfase à evolução de simuladores como formas de obter melhores resultados no melhoramento de desempenho das filas.

Em sua pesquisa de melhorias de atendimento em prontos socorros, Lima e Belderrain (2007) citam que a Teoria de Filas cumpre sua tarefa somente em parte do sistema, necessitando de outras teorias aplicadas em conjunto para a sua mais correta utilização. Além disso, por se tratar de um ambiente com muitas variáveis humano-relacionadas, as soluções propostas pelo modelo devem primeiramente passar por simulações, sendo verificado seu real resultado, pois dificilmente um modelo matemático consegue considerar corretamente o relacionamento interpessoal em tantos contextos.

Figueiredo e Rocha (2010) realizaram, em sua aplicação da Teoria de Filas, uma divisão temporal do ambiente, a fim de aperfeiçoar a aplicação do modelo para cada intervalo de tempo. Isto, segundo eles, diminui a influência de tempos médios obtidos em intervalos longos de tempo, que podem ser distorcidos por comportamentos momentaneamente extremos. Esta situação não precisaria ser abordada caso a modelagem suportasse a mudança dinâmica de comportamento, prevendo que o sistema possui comportamentos diferentes conforme o passar do tempo.

Oliveira e Favaretto (2013) indicam que a Teoria de Filas pode ser aplicada nos problemas relacionados a filas, porém quando os problemas são muito complexos a análise das mesmas acaba sendo algo inviável. Nestes casos, os autores sugerem a utilização de simulações.

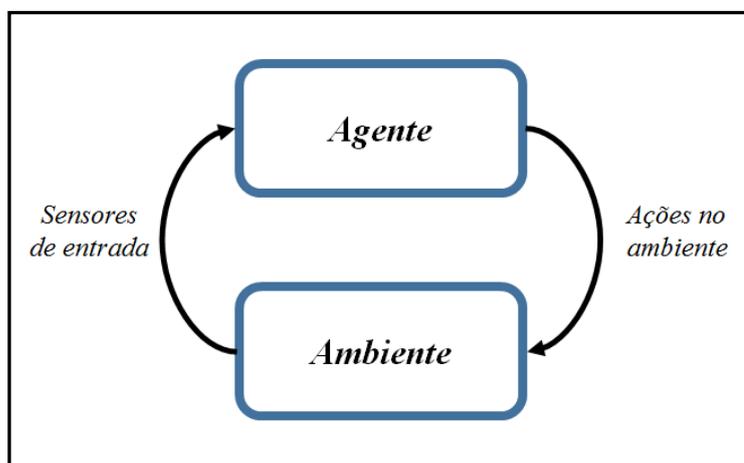
Com este levantamento bibliográfico, é visada neste trabalho a complementação da Teoria de Filas com um modelo que possa diminuir as limitações apresentadas.

4 AGENTES

A definição do conceito de *agente* surgiu na área de Inteligência Artificial na década de 1980, com a necessidade de criar modelos que possam reproduzir, pelo menos parcialmente, a inteligência humana (WOOLDRIDGE; JENNINGS, 1995). Atualmente este termo é utilizado em mais áreas da computação, como banco de dados, sistemas operacionais e redes, tendo em sua definição leves peculiaridades, porém o conceito principal se mantém intacto (FERREIRA, 2008).

Segundo Wooldridge (2002), “agente é um sistema computadorizado que está situado em um ambiente, e é capaz de ações autônomas neste ambiente para cumprir seus objetivos” (tradução nossa). Esta definição é a mais utilizada na literatura, abordada pela maioria das áreas que utilizam agentes, porém o próprio autor sugere que a definição de agentes é muito subjetiva e que pode variar muito em casos específicos.

Figura 8 – Representação de um agente em seu ambiente



Fonte: do autor, adaptado de Wooldridge (2002).

A ideia básica do funcionamento de um agente é, conforme pode ser visto na Figura 8, que ele está situado em um ambiente e interage com o mesmo. Para a interação ocorrer o agente pode receber informações externas, ou então decidir por realizar uma ação. Após isso, de alguma maneira o agente deve agir no ambiente, podendo ser a alteração de alguma informação neste contida, a comunicação com outros agentes, ou a realização de alguma outra tarefa.

Não é usual a aplicação de agentes únicos em sistemas e simulações, sendo a aplicação coletiva, denominada Sistemas Multiagente (SMA) o mais comum. Esta abordagem trabalha essencialmente com uma filosofia de distribuição de problemas, em que estes são particionados

e atribuídos a diferentes agentes, juntando as soluções encontradas para obter a resposta final do problema (SILVA, 2006). Esta sistemática traz um grande benefício na modelagem de agentes que é a simplificação de suas definições, pois um agente não precisa fazer tudo, ele somente precisa fazer uma parte do trabalho, talvez uma parte muito pequena (FERREIRA, 2008).

4.1 CARACTERÍSTICAS

Algumas propriedades são necessárias para que uma entidade possa ser considerada um agente, porém alguns estudiosos consideram algumas destas características opcionais. Bonabeau (2001 apud MACAL; NORTH, 2005) considera qualquer componente independente como agente, enquanto que para Jennings (2000 apud MACAL; NORTH, 2005) a única característica essencial em um agente é a autonomia. Estas opiniões podem gerar certos debates na comunidade científica, porém normalmente são consideradas somente nas áreas em que tais autores exercem suas pesquisas.

As características usualmente consideradas como “requeridas” para a definição de um agente são enumeradas por vários autores, como Macal e North (2005), Lima et al. (2009), Ferreira (2008) e Juchem e Bastos (2001), conforme segue:

- Deve ser identificável, possuir um conjunto de características e regras que governam o seu comportamento no ambiente, e estes devem possuir limites bem definidos e claros.
- Deve estar situado em um ambiente junto com outros agentes, sendo capaz de interagir tanto com o ambiente situado como com os outros agentes. Também deve ser capaz de distinguir os outros agentes uns dos outros.
- Deve ser orientado aos seus objetivos, respeitando as suas limitações comportamentais.
- Deve possuir autonomia para realizar as suas tarefas, ou pelo menos na maioria delas. Não deve depender de eventos e informações do ambiente ou de outros agentes.
- Deve ter a capacidade de aprender com ações passadas, ganhando experiência para futuras decisões. Também deve poder se adaptar a mudanças no ambiente e em outros agentes.

Existem também características que definem os agentes em SMA e que não fazem sentido constar na lista de atributos individuais. Silva (2006) descreve alguns exemplos de

propriedades coletivas, como agentes homogêneos ou heterogêneos, possuidores de objetivos em comum ou não, sendo competidores ou colaboradores.

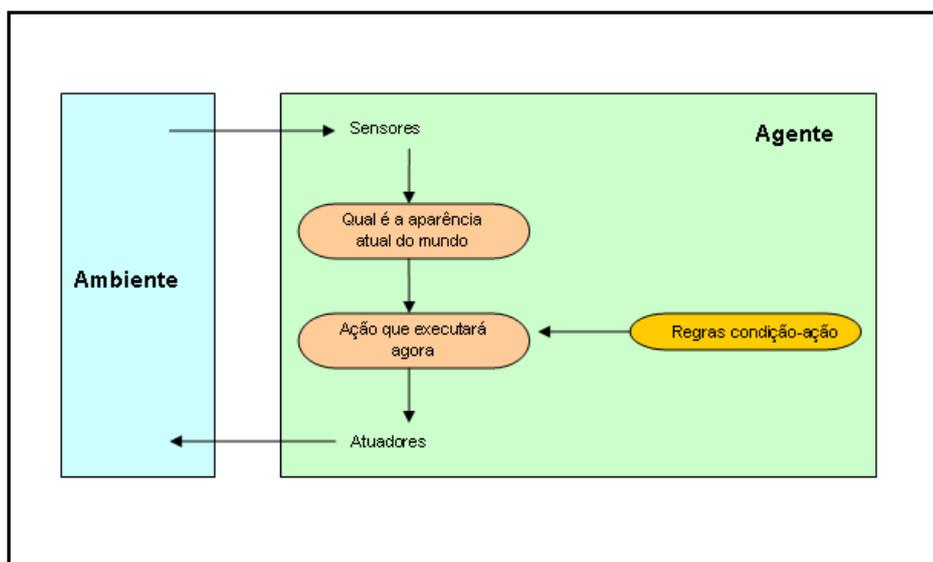
4.2 CLASSIFICAÇÕES

Os agentes podem respeitar todas as características básicas citadas anteriormente e mesmo assim serem elementos muito diferentes. Isto se deve ao fato de não existir uma regra fixa de como devem ser construídos (WOOLDRIDGE, 2002). A partir desta distinção, os agentes também são classificados em dois grupos, conforme os modelos de organização em que são baseados: agentes *reativos* e agentes *cognitivos*.

4.2.1 Agentes reativos

Em geral, agentes reativos são aqueles que possuem modelagem mais simples, funcionamento mais básico e por isso acabam sendo os mais comuns (SILVA, 2006). Juchem e Bastos (2001) afirmam que eles “não têm capacidade de raciocínio sobre suas intenções, reagindo tão somente sobre regras e planos estereotipados”. Assim, os agentes reativos acabam sendo utilizados em ambientes mais simples, que não exigem capacidades e ações complexas.

Figura 9 – Modelo de agente reativo



Fonte: <http://www.inf.ufes.br/~liviaufmt>

A Figura 9 demonstra de forma prática o funcionamento de um agente reativo. As informações são recebidas pelo agente através de sensores, os quais são elementos que capturam dados do ambiente. Após, tais informações são repassadas a um conjunto de regras para que

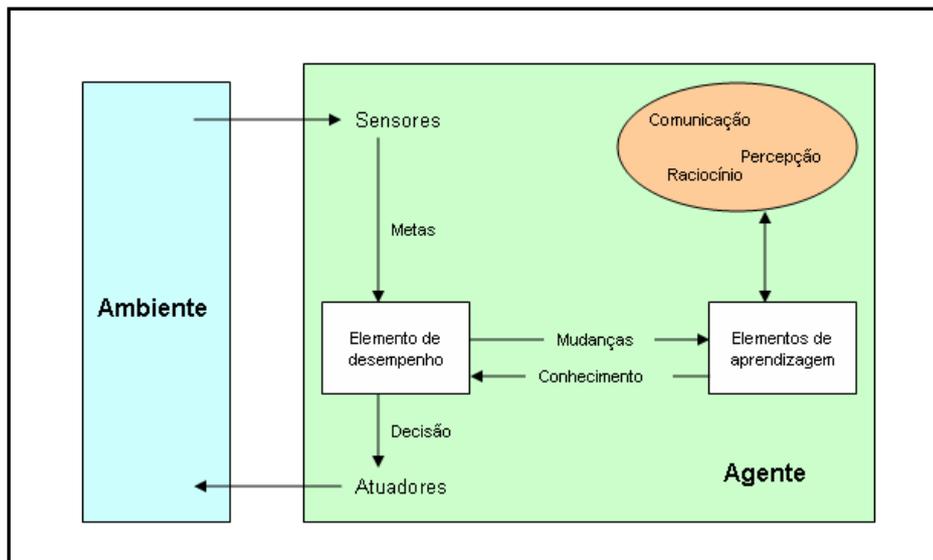
seja tomada a decisão do que fazer, sendo esta aplicada utilizando recursos controláveis presentes no ambiente, os atuadores. A implementação real do agente, e com isso a sua complexidade, acaba sendo a definição do conjunto de regras.

4.2.2 Agentes cognitivos

O conceito de cognição vem da área da psicologia e trabalha com a aquisição de conhecimento através de aspectos humanos e sociais. Na área de agência, a cognição é utilizada em modelos que trabalham com emoções, intenções e crenças, incorporando aspectos sociais que influenciam as tomadas de decisão (MACAL; NORTH, 2005). Com uma representação mais abstrata do mundo, os agentes cognitivos utilizam raciocínio lógico sobre símbolos que podem ser concebidos fisicamente, gerando uma estrutura operável (JUCHEM; BASTOS, 2001).

Esta categoria de agente costuma ser de grande complexidade, pois é aquele que incorpora o maior número de “subsistemas” trabalhando para receber todos os detalhes do ambiente e de outros agentes, aprender, tomar decisões e avaliar de que forma o processo pode ser melhorado nas próximas vezes (SILVA, 2006).

Figura 10 – Modelo de agente cognitivo



Fonte: <http://www.inf.ufes.br/~liviaufmt>

Na Figura 10 pode-se ver um esquema do funcionamento mais complexo do agente cognitivo em relação ao reativo. Ao receber as informações do ambiente e de outros agentes através dos sensores, as mesmas passam primeiramente por um processo de aprendizagem, para então resultar na decisão que é repassada para os atuadores do agente. Nota-se que a grande

mudança é a troca do conjunto de regras fixas, existente no conceito reativo, para um sistema de aprendizado e raciocínio dinâmico, em que praticamente tudo pode mudar a cada situação.

4.3 DESTAQUES

A utilização de agentes na modelagem e simulação de sistemas se destaca pelo comportamento adaptativo e evolutivo que eles possuem, sendo concedido a eles o reconhecimento de *agentes inteligentes*. Estes atributos são cada vez mais necessários na atualidade, pois a complexidade dos sistemas e a interconectividade entre eles é cada vez maior, tornando a utilização de modelos estáticos muito difícil (MACAL; NORTH, 2005).

Segundo Bonabeau (2002 apud LIMA et al., 2009) a utilização de agentes traz alguns benefícios comparado a outras técnicas de modelagem, como flexibilidade, descrição natural de sistemas e monitoramento em tempo real de eventos e fenômenos. Algumas situações em que SMA se sobressaem são aquelas em que há a possibilidade de aprendizado e adaptação, quando há uma heterogeneidade muito forte no ambiente ou a interação entre os elementos é muito complexa, ou ainda quando o sistema é mais bem descrito por atividades ao invés de processos, entre outros casos.

Jennings (1996 apud JUCHEM; BASTOS, 2001) afirma que SMA possuem grandes diferenciais em relação a modelos unitários, tais como maior flexibilidade através da heterogeneidade de agentes, aumento de segurança com agentes colaboradores, comunicação mais eficaz e objetiva ao invés da distribuição de dados brutos e a possibilidade de paralelismo, entre outros.

Em sua dissertação, Silva (2006, p. 19) considera que “tratar sistemas como sendo compostos de agentes autônomos que interagem entre si tem sido considerada uma promissora abordagem para o desenvolvimento de aplicações em domínios complexos”.

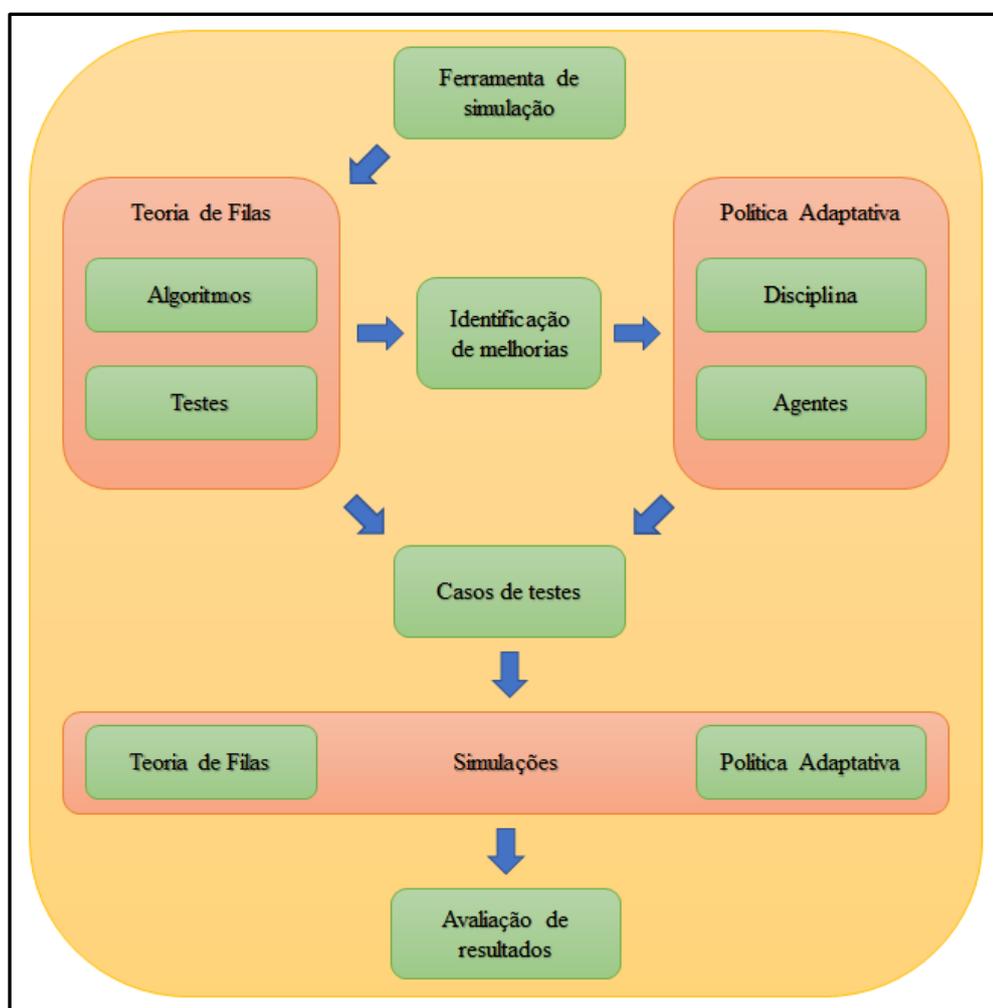
Assim sendo, nota-se que os agentes e SMA apresentam as características necessárias para o complemento da Teoria de Filas nas questões que podem ser melhoradas. Com uma correta implementação, a aplicação de agentes pode trazer melhorias significativas à abordagem atual sem desfigurar o trabalho desta.

5 DESENVOLVIMENTO

A estratégia adotada para o desenvolvimento da política de atendimento consistiu em uma sequência de passos seguidos com rigor, garantindo, assim, a consistência do estudo e dos resultados para a avaliação do seu desempenho.

Para tanto, a Figura 11 representa esquematicamente o processo seguido para execução da modelagem, implementação da simulação, e análise dos resultados.

Figura 11 – Processo de desenvolvimento, simulação e análise de resultados



Fonte: do autor

Este capítulo descreve aspectos relevantes da etapa de modelagem. São contemplados a escolha da ferramenta de simulação, o desenvolvimento e os testes dos algoritmos da Teoria de Filas, a identificação de melhorias e a sua aplicação na Política Adaptativa, em forma de disciplina ou de agente.

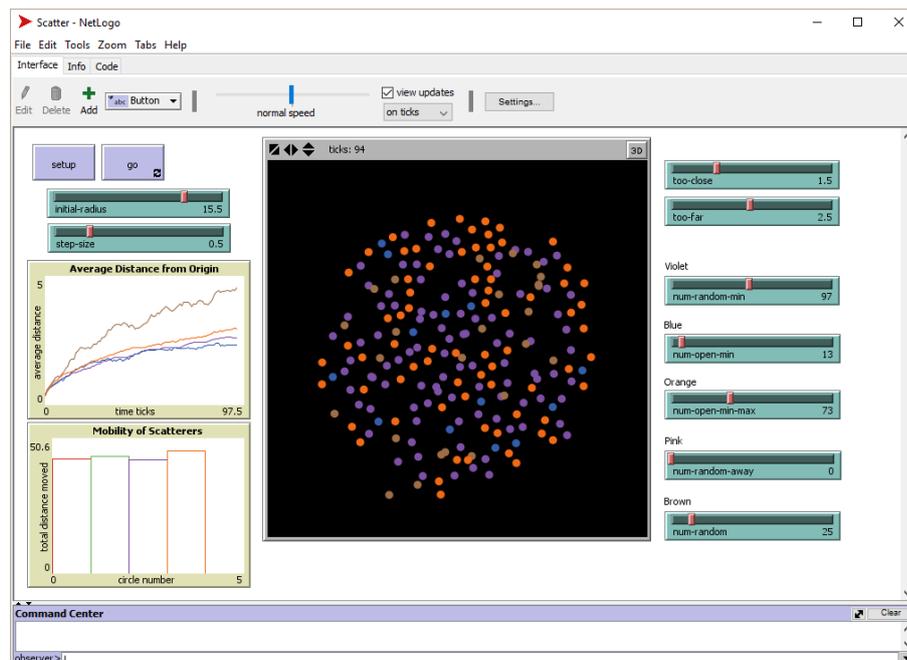
5.1 FERRAMENTA DE SIMULAÇÃO

A ferramenta de desenvolvimento e simulação escolhida para realizar as simulações foi o *NetLogo*. Tisue e Wilensky (2004) o classificam como um ambiente para modelagem e programação multiagente para simulação de fenômenos naturais e sociais, sendo muito eficiente na simulação de sistemas que evoluem conforme o tempo.

O NetLogo oferece algumas características favoráveis para a criação de ambientes simulados, com possibilidade de parametrização e de acompanhamento de execução. Wilensky e Rand (2015) citam entre elas: a fácil legibilidade de sua linguagem de programação, a possibilidade de execução do código fonte para testes, e a grande variedade de modelos multiagente, de diversas áreas do conhecimento, já desenvolvidos no NetLogo e disponíveis para consulta.

No aplicativo, estão disponíveis 3 guias de desenvolvimento: uma é a *interface* gráfica, parte “controlável” da simulação, que pode receber comandos durante a execução do modelo e apresentar resultados ao usuário; a segunda é a parte de informações, utilizada para documentar o modelo; e a terceira e última parte contém o código fonte do simulador, em uma linguagem própria da ferramenta.

Figura 12 – Guia de *interface* do NetLogo



Fonte: Modelo *Scatter* do NetLogo

Figura 13 – Guia de código fonte do NetLogo

```

Scatter - NetLogo
File Edit Tools Zoom Tabs Help
Interface Info Code
Find... Check Procedures Indent automatically

;; set heading in the direction opposite
;; of the closest turtle, without stopping
to move-random-away ;; turtle procedure
set closest-classmate min-one-of other turtles [distance myself]
face closest-classmate
rt 180
avoid-walls
fd step-size
set stopped? false
end

;; set heading towards the largest open space,
;; stopping when all other turtles are at least "too-close" away
to move-open-min ;; turtle procedure
set nearby-classmates other turtles in-radius too-close
ifelse any? nearby-classmates
[ facexy (mean [xcor] of nearby-classmates)
  (mean [ycor] of nearby-classmates)
  rt 180
  avoid-walls
  fd step-size
  set stopped? false ]
[ set stopped? true ]
end

;; when turtles are too close, move to an open space,
;; when turtles get too far, move in towards a more populated area, and otherwise, stop
to move-open-min-max ;; turtle procedure
set close-classmates other turtles in-radius too-close
set far-classmates other turtles in-radius too-far
ifelse any? close-classmates ; move to an open space
[ facexy (mean [xcor] of close-classmates)
  (mean [ycor] of close-classmates)
  rt 180
  avoid-walls
  fd step-size
  set stopped? false ]
[ ifelse not any? far-classmates ; move to a more populated space in the 'room'
  [ facexy mean [xcor] of other turtles
    mean [ycor] of other turtles
    avoid-walls
    fd step-size
    set stopped? false ]
  [ set stopped? true ] ]
end

```

Fonte: Modelo *Scatter* do NetLogo

No atual trabalho foram utilizados a primeira e a terceira guias, sendo desenvolvida uma interface para controle e acompanhamento do processo de execução do simulador, como demonstrado na Figura 12, além de todos os algoritmos utilizados, semelhante ao exemplo da Figura 13.

5.2 ESCOPO DA TEORIA DE FILAS

O desenvolvimento dos algoritmos que remetem puramente à Teoria de Filas era requisito para avaliar os resultados da política desenvolvida como objetivo do trabalho. Por isso, necessariamente deveriam existir definições sobre as propriedades que estariam disponíveis no simulador, incluindo suas restrições.

Para este propósito, inicialmente foram decididos quais algoritmos seriam desenvolvidos. Foi optado por implementar somente as características mais comuns em ambientes reais, com o intuito de servirem como caso base de comparação com o novo algoritmo. Já as dimensões destas não seriam extremas, evitando com isto deixar as simulações muito onerosas, tanto de tempo como de recursos. Caso necessário, o redimensionamento de ambientes reais poderia ser utilizado nas simulações.

As propriedades definidas foram as seguintes:

- Tempos entre chegadas: Distribuições exponencial, determinística e normal;

- Tempos de atendimento: Distribuições exponencial, determinística e normal;
- Número de canais de serviço: 1 a 8 atendentes;
- Capacidade do sistema: 500 clientes;
- Disciplina de filas: FIFO, LIFO e PRI.

As duas primeiras propriedades, que são métricas de tempos, requerem também a determinação de um tempo base e uma forma de variação. O primeiro foi limitado, em ambas as situações, entre 0 a 20.000 segundos, ou 5h:33m:20s, devido à possibilidade de redução em escala dos ambientes reais refletidos nas simulações. Para o tipo de variação foi escolhido o desvio padrão como medida, por ser uma métrica muito eficaz na dispersão de dados em relação a uma média.

O funcionamento da disciplina PRI também requer a designação de uma prioridade de atendimento. Para isso, o sistema possui uma característica de prioridade do atendimento, a qual pode variar de 0 até 100 unidades de prioridade.

Uma peculiaridade que não pertence normalmente ao escopo da Teoria de Filas, porém deve existir em ambientes de atendimento, é a caracterização dos mesmos. Neste estudo são considerados 5 tipos de atendimentos diferenciados (Ta), que podem ter configurações diferentes entre si, salvo casos em que a configuração é própria do ambiente e não do atendimento. Neste caso se enquadram, inicialmente, as definições de tempos entre chegadas e de atendimento e também a prioridade do atendimento.

5.2.1 Algoritmos

As características apresentadas foram desenvolvidas e aplicadas, sendo monitoradas através de gráficos em tempo real para demonstrar o seu funcionamento e permitindo ajustes finos das parametrizações. O ambiente de simulação permite a parametrização completa entre estas características para que, ao realizar as simulações, possa ser construído qualquer caso desejado utilizando tais propriedades. Ainda foi previsto a necessidade da adição de novas características ao ambiente através do desenvolvimento da nova política adaptativa de atendimento e seus agentes.

Assim como a escolha entre as disciplinas FIFO, LIFO e PRI, também foi permitido a configuração do número máximo de clientes e número de atendentes, sendo todas estas configurações únicas, do ambiente. Além delas, também foram disponibilizadas as

propriedades particulares de cada um dos 5 tipos de atendimento (nomeados aqui como Dúvida, Matrícula, Financeiro, Documento e Negociação).

Para fins comparativos, um exemplo em notação Kendall para um ambiente de simulação pode ser visto na Tabela 3.

Tabela 3 – Notação Kendall de um exemplo de ambiente desenvolvido

| Tipo de atendimento | Tempo entre chegadas | Tempo de atendimento | Número de canais | Capacidade do sistema | Disciplina de fila |
|---------------------|----------------------|----------------------|------------------|-----------------------|--------------------|
| Dúvida | $M(200/40)$ | $M(30/5)$ | 4 | 250 | <i>LIFO</i> |
| Matrícula | $M(1000/200)$ | $D(120/20)$ | | | |
| Financeiro | $D(300/30)$ | $N(40/5)$ | | | |
| Documento | $D(3000/500)$ | $M(200/80)$ | | | |
| Negociação | $M(700/200)$ | $N(50/2)$ | | | |

Fonte: do autor

Nota-se que as três últimas características dos atendimentos são fixas entre os tipos (pois são próprias do ambiente). Já as duas primeiras possuem a indicação do tempo base e do desvio padrão configurado na característica, sendo estas diferentes para cada tipo. A prioridade do atendimento acabou não sendo relevante para este caso de exemplo (disciplina LIFO configurada), portanto não foi citada.

5.2.2 Testes e identificação de melhorias

Com os algoritmos que envolvem a Teoria de Filas funcionais, foram realizados vários testes em diferentes situações. Assim, situações passíveis de melhorias, que poderiam ser utilizadas no desenvolvimento da política adaptativa de atendimento, puderam ser identificadas.

As principais situações observadas, porém, não as únicas, foram as seguintes:

- **Habilidade de atendimento:** Comparando dois atendentes, um só realizou atendimentos do tipo Financeiro, e o outro somente do tipo Matrícula. Quando o segundo precisa realizar um atendimento do tipo Financeiro, a habilidade dele é a mesma que o primeiro atendente. Em ambientes que envolvem atores inteligentes como atendentes, como é o caso de muitos ambientes reais, este comportamento pode não estar correto. Através do encaminhamento das demandas para canais com mais experiência na área, o processo estaria melhor representado;

- **Atendimentos sem prioridade:** Existem casos em que uma categoria de serviços possui uma prioridade de atendimento muito baixa, por razões diversas. Devido a isso, em alguns momentos ocorre de clientes com esta classificação estarem na fila há algum tempo e não serem atendidos logo, pois clientes com tipo de atendimento mais prioritário podem ter chegado à fila. Isto ocasiona clientes com altos tempos de fila, por terem sido ignorados enquanto os atendimentos de prioridade mais alta não tivessem sido exauridos;
- **Variações da demanda:** Em ambientes que possuem muita variação na chegada de clientes, ocorre que em alguns momentos o número de clientes cresce muito rapidamente. Como as regras de atendimento dos atendentes são fixas, normalmente feitas para o modo “estável” do sistema, essa variação acaba não sendo tratada. Esta situação pôde ser contornada, em sua maioria, com o aumento do número de canais.

Em alguns destes casos, o próprio ambiente de simulação já oferece os recursos necessários para a solução dos entraves, com intervenções manuais. Utilizando desta facilidade, não somente os problemas foram reconhecidos, como também as soluções para alguns deles já foram testadas.

5.3 POLÍTICA ADAPTATIVA

O princípio deste estudo é melhorar o processo de atendimento em cenários que a Teoria de Filas não abrange corretamente, ou seja, situações não previstas pela mesma. Estas ocasiões são abundantes em nosso dia-a-dia, ocorrendo em diferentes sistemas de filas, como por exemplo os bancos e sua diversidade de demandas, caixas de supermercados com prioridades fixas e recursos não utilizados, entre outros.

Existem alguns casos que possuem maior notoriedade na relação de problemas com a Teoria de Filas, como um ambiente de atendimento de universidade. Tal sistema possui uma grande variação de demandas conforme o período do ano. No começo e no fim de semestres letivos há uma grande demanda relacionada à matrícula dos estudantes. Já no meio dos semestres há poucas requisições deste gênero, sendo outros atendimentos, como a requisição de documentos, os mais exercidos. Além disso, existem tipos de atendimento que variam diariamente, como é o caso da área financeira, possuindo demandas em sua maioria no início dos meses.

Em muitas destas situações, os ambientes são compostos por pessoas, tanto no papel de clientes como de atendentes. Neste contexto, cada pessoa possui características únicas: clientes desejam tipos de atendimento diferentes, possuem mais ou menos problemas, etc.; atendentes podem possuir pré-disposição a determinado assunto, podem ser afetados pelo ambiente externo, possuem a capacidade de aprendizagem, têm habilidades distintas, etc. Estes cenários, variáveis em quase todos os aspectos, oferecem muitas opções de tratamentos específicos conforme a situação.

Para ocorrer o desenvolvimento de melhorias para estes ambientes, é possível que algumas abordagens sejam utilizadas. Podem ser avaliados aspectos que naturalmente oferecem oportunidades de melhoria, como recursos não utilizados e qualificação dos profissionais, assim incrementando o aproveitamento do processo já existente. Ou então podem ser levados em consideração pontos negativos do ambiente, como a priorização excessiva de demandas onerosas e a falta de canais disponíveis, aplicando tratamentos específicos para assim evitá-los.

Após os testes realizados com os algoritmos da Teoria de Filas, foram identificados pontos de atuação da política adaptativa e também o que poderia ser feito. De uma forma automática, a nova política deveria possuir uma disciplina de filas mais eficaz e inteligente, recebendo o suporte de agentes para o tratamento de situações específicas.

5.3.1 Disciplina

A disciplina de atendimento foi o primeiro ponto analisado para a implementação da política adaptativa. Nesta definição, foram levados em consideração os prós e os contras das disciplinas da Teoria de Filas, objetivando com isto aproveitar os pontos positivos e evitar a implementação de mecanismos já existentes e que não trazem o retorno esperado. Através desta análise, foi identificado que a maioria dos “contras” é baseado no fato de que as disciplinas são fixas, ignoram o estado atual do sistema. Esta, então, foi a principal regra a ser evitada na nova disciplina.

Seguindo a definição do termo *adaptação*, o princípio é ser ajustável ao ambiente atual, trabalhar conforme a presente situação do sistema, seja ela qual for. Não deve seguir sempre a mesma regra ou limitar as possibilidades do sistema que a emprega, aceitando alterações de regras de funcionamento em qualquer momento de seu funcionamento.

A proposta de modelagem da disciplina é baseada em um método de escolhas por afinidade. O canal de atendimento escolhe a demanda a atender conforme afinidade com a

mesma, preferindo aquelas em que o seu conhecimento é maior. Este funcionamento é comparável ao de alguns ambientes reais, como os citados anteriormente, em que os atendentes podem escolher suas demandas conforme algum tipo de regra por eles seguida. Estes sistemas são, em sua maioria, aqueles que possuem pessoas nos papéis de canais de atendimento, utilizando seus próprios instintos para definição de prioridades.

A partir desta proposta, a regra de ordenação dos atendimentos poderia seguir o mesmo fluxo que ocorre em tais modelos. J. B. Skinner possui uma teoria em que, segundo Viotto Filho, Ponce e Almeida (2009, p. 29),

[...] o homem age no mundo em decorrência das consequências positivas ou negativas de seu comportamento e, diante disso, não se pode pensar o ser humano simplesmente submetido ao ambiente de forma passiva, pois os mesmos não estão apenas "atentos" ao mundo, mas respondem de forma idiossincrática àquilo que já experienciaram objetiva ou subjetivamente ao longo de sua história. Isso implica afirmar que o ser humano é produto do processo de aprendizagem vivido ao longo de sua vida.

Acompanhando este raciocínio, ao fazer uma escolha, a pessoa tende a escolher aquela à qual possui mais afinidade, conhece melhor, tem mais experiência. Assim, pode ser obtido de alguma forma este dado do nosso ambiente e aplicar nos papéis principais do atendimento, os atendentes.

Seguindo o conceito da disciplina PRI da Teoria de Filas, em que existe uma prioridade “global” de atendimento (todos os atendentes seguem a mesma regra), a disciplina Adaptativa (ADA), desenvolvida neste trabalho, consiste em individualizar esta prioridade entre os atendentes. Deste modo, não existe mais uma regra geral de priorização, mas sim regras pré-definidas e individualizadas para cada atendente.

Tabela 4 – Diferença entre as prioridades das disciplinas PRI e ADA

| Serviço | <i>PRI</i> Atendentes 1 a 4 | <i>ADA</i> Atendente 1 | <i>ADA</i> Atendente 2 | <i>ADA</i> Atendente 3 | <i>ADA</i> Atendente 4 |
|------------|--------------------------------|---------------------------|---------------------------|---------------------------|---------------------------|
| Dúvida | 1 | 1 | 2 | 1 | 4 |
| Matrícula | 2 | 2 | 5 | 2 | 1 |
| Financeiro | 3 | 5 | 3 | 5 | 2 |
| Documento | 4 | 4 | 1 | 4 | 3 |
| Negociação | 5 | 3 | 4 | 3 | 5 |

Fonte: do autor

Conforme demonstrado na Tabela 4, na disciplina ADA, cada atendente possui uma regra de prioridade pessoal, que pode ou não ser igual à de outros atendentes (atendentes 1 e 3 possuem regras iguais), diferentemente da disciplina PRI, que possui uma regra geral.

Figura 14 – Pseudocódigo da disciplina ADA

```
Monta lista de prioridades;  
Ordena lista de prioridades em modo decrescente;  
Para tipo = 1 até 5 fazer  
    Busca cliente do tipo (x);  
    Se x ? 0 então  
        Indica próximo cliente (x);  
        Termina para;  
    Fim se;  
Fim para;
```

Fonte: do autor

A Figura 14 apresenta a implementação do algoritmo da disciplina ADA em pseudocódigo. Inicialmente as prioridades dos tipos de serviços são ordenadas de forma decrescente, resultando em uma lista com o atendimento prioritário no topo. Esta lista é utilizada, de forma sequencial, buscando clientes que possam estar na fila com determinada classe de requisição. Ao encontrar um próximo cliente, o algoritmo finaliza seu processamento.

5.3.2 Agentes

A definição dos agentes considera a utilização da nova disciplina de filas ADA pela política adaptativa, porém sem haver um vínculo de obrigatoriedade, para que os agentes também pudessem ser aplicados nas disciplinas da Teoria de Filas. Esta desvinculação tem por objetivo permitir testar o desempenho dos agentes em situações diversas, sendo este avaliado tanto individualmente como em conjunto com outros agentes.

A análise realizada para definição dos agentes baseia-se em ampliar um modelo construído a partir da Teoria de Filas em contextos em que o sistema perde estabilidade, ou seja, quando o estado atual do ambiente se encontra diferente daquele para qual o modelo atual de atendimento foi proposto. Por exemplo, no início de semestres em uma universidade, em que as demandas da classe Matrícula de atendimento aumentam significativamente, voltando ao normal após um certo período, como já citado anteriormente.

É importante destacar, contudo, que este trabalho não tem por finalidade propor uma otimização do sistema, e sim adaptações no processo que ofereçam melhorias no gerenciamento do atendimento. Portanto, a escolha dos agentes que foram desenvolvidos levou em consideração, principalmente, as necessidades de adaptação da política de atendimento em situações atípicas que um modelo de teoria de filas pode enfrentar na realidade e o possível ganho de desempenho, que é a melhoria dos tempos de aguardo em fila dos clientes, que elas trariam nestes contextos.

5.3.2.1 Agente de aprendizagem

Para abordar o primeiro caso relatado que a Teoria de Filas não contempla, a ideia inicial é controlar o nível de capacidade de cada atendente para cada tipo de atendimento. O conceito deste controle é comparável ao do aprendizado de um ser humano (atores representados pelos agentes de atendimento), sendo esta uma habilidade que possui níveis iniciais e pode evoluir/regredir, conforme o modelo de ganho/perda do mesmo que é utilizado.

A noção mais comum que se tem do aprendizado de uma pessoa sobre algo é de que “quanto mais, melhor”. Ou seja, quanto mais vezes o indivíduo passou por determinada experiência, mais ele ganha afinidade e obtém conhecimento sobre esta. Traduzindo este conceito para um sistema de atendimento, obtém-se a conclusão de que quanto mais atendimentos uma pessoa faz sobre um determinado assunto, maior vai ser o conhecimento na área envolvida.

Para implementar este conceito no ambiente de simulação deste trabalho, a ideia foi contabilizar os últimos atendimentos de um tipo específico feitos por um atendente, para saber o nível de experiência do mesmo sobre o assunto envolvido. Alguns limites foram impostos devido a situações que não espelham a realidade, como, por exemplo, um atendente não pode ser tão bom em um determinado assunto que ele resolve tudo instantaneamente; inversamente, o atendente poderia não conseguir resolver algo por falta de experiência, porém foi decidido por não abordar este lado extremo, pois a complexidade dos modelos envolvidos seria demasiada.

A consulta de dados dos últimos atendimentos requer uma nova propriedade no sistema: o tamanho deste histórico. Para que haja a possibilidade de trabalhar com experiências passadas, precisa haver um registro do que ocorreu, com detalhes como tempo, tipo de atendimento, atores envolvidos, etc. Este histórico, visando não manter o registro completo de todas as atividades (em um ambiente real sempre há um limite de ações memorizadas, variando conforme o ambiente), possui um limite definido de 1000 experiências (Lh). Ao ultrapassar o limite, cada novo registro guardado acaba eliminando o mais antigo salvo.

A lógica para se obter a experiência do atendente em um determinado tipo de atendimento envolve os cálculos

$$Em = Eb * Ta * Pm \quad (1)$$

$$Eu = ((Eb - Em) * Ta) / Qa \quad (2)$$

$$Et = Em + (Qt * Eu) \quad (3)$$

onde Em é a experiência mínima do atendimento, Eb é a experiência base, Pm é o percentual mínimo de experiência para cada tipo de demanda, Eu é a experiência unitária por atendimento, Qa é a quantidade de atendimentos recentes, Et é a experiência total do tipo e Qt é a quantidade de atendimentos recentes do tipo. Ao acumular o Et de todas as classes de serviços, se obtém o limite de experiência do atendente, que é Em sem Pm .

A lógica proposta por tais cálculos oferece algumas garantias necessárias para o correto funcionamento do ambiente, de modo a atender à necessidade de simular um processo de aprendizagem. Com ela, por exemplo, não há possibilidade de um atendente não possuir conhecimento em uma área de atendimento, ou possuir conhecimento tão alto que atenda requisições instantaneamente. Além disso, para fins de abstração adequada, não pode ocorrer conhecimento ilimitado.

Figura 15 – Pseudocódigo do Agente de aprendizagem

```

Calcula experiência mínima (Em);
Calcula experiência unitária (Eu);
Para tipo = 1 até 5 fazer
    Busca quantidade de atendimentos (Qt);
    Calcula experiência do tipo de atendimento (Et);
Fim para;

```

Fonte: do autor

A Figura 15 apresenta um pseudocódigo do algoritmo utilizado pelo Agente de aprendizagem. Utilizando os cálculos já explanados, as experiências de todas as classes de serviços são calculadas sempre juntas, garantindo que a experiência ganha em um tipo será removida de outra. Isto porque a experiência total do atendente, conforme a lógica utilizada, deve se manter no limite de Eb multiplicado por Ta . Portanto, se alguma classe de serviços tiver a experiência alterada, esta mesma modificação deverá ocorrer, inversamente, em outra categoria de atendimento, assim mantendo a estabilidade da experiência total do atendente.

A partir da experiência calculada, quando utilizado a disciplina ADA, a prioridade de atendimento pode basear-se no conhecimento dos atendentes. Como este conhecimento pode ser recalculado em qualquer momento, como por exemplo a cada requisição cumprida ou a cada dia de serviço, a ordem de priorização pode ser reajustada na frequência necessitada pelo sistema, dinamizando o processo de escolha.

Tabela 5 – Exemplo de cálculo das experiências de um atendente com Lh em 100

| Número do atendimento | Atendimentos Dúvida | Atendimentos Matrícula | Atendimentos Financeiro | Atendimentos Documento | Atendimentos Negociação | Experiência total Dúvida (Et) | Experiência total Matrícula (Et) | Experiência total Financeiro (Et) | Experiência total Documento (Et) | Experiência total Negociação (Et) |
|-----------------------|---------------------|------------------------|-------------------------|------------------------|-------------------------|-----------------------------------|--------------------------------------|---------------------------------------|--------------------------------------|---------------------------------------|
| 1.050 | 20 | 20 | 20 | 20 | 20 | 1 | 1 | 1 | 1 | 1 |
| 1.100 | 35 | 12 | 27 | 7 | 19 | 1,5625 | 0,7 | 1,2625 | 0,5125 | 0,9625 |
| 1.150 | 9 | 40 | 12 | 15 | 24 | 0,5875 | 1,75 | 0,7 | 0,8125 | 1,15 |
| 1.200 | 2 | 63 | 8 | 12 | 15 | 0,325 | 2,6125 | 0,55 | 0,7 | 0,8125 |
| 1.250 | 0 | 84 | 6 | 5 | 5 | 0,25 | 3,4 | 0,475 | 0,4375 | 0,4375 |
| 1.300 | 2 | 71 | 10 | 14 | 3 | 0,325 | 2,9125 | 0,625 | 0,775 | 0,3625 |

Fonte: do autor

Na Tabela 5 é exemplificado o funcionamento do cálculo de experiência de um atendente em um determinado intervalo de atendimentos. Através das definições já realizadas (E_b é 1, T_a é 5 e P_m é 5), E_m pode ser calculado como 0,25; E_u então pode ser obtido, verificando que Q_a é 100 (e já atingiu o limite Lh), com o valor de 0,0375; estes dados então permitem calcular cada Et apresentado na tabela, como por exemplo no atendimento 1.100, com 35 Qt , que possui Et de 1,5625. Salienta-se que a ocorrência do atendimento 1.250 demonstra uma grande amplitude entre o número de atendimentos realizados dos tipos “Dúvida” e “Matrícula”, refletindo isso para Et e salientando E_m quando o número de atendimentos é nulo.

Este agente pode ser enquadrado na categoria de Agentes Reativos, pois o mesmo possui comportamento fixo, em que sempre utiliza a mesma quantidade de experiências de atendimento para montar as habilidades do atendente, as quais também são montadas sempre da mesma forma. Além disso, este agente só é utilizado após determinados eventos (término de um atendimento), não existindo motivos de ser usado fora deste contexto. Com estas definições, também pode ser considerado um agente “simples”, com implementação não muito complexa.

5.3.2.2 Agente de priorização

O segundo caso observado em que se pode implementar um agente controlador não é vinculado ao comportamento do ser humano, como foi o primeiro. Porém isso não reduz a sua complexidade, muito pelo contrário, pois não existe uma definição correta, e não é algo presente no cotidiano das pessoas assim como é a implementação do primeiro agente.

Primeiramente podem ser salientadas algumas observações do problema relatado:

- Acontece em casos que o sistema possui fila ativa, e não em momentos de ociosidade;

- Ocorre normalmente com tipos de atendimento com baixa prioridade, porém ocorre com os de alta prioridade também, em menor quantidade;
- Gera diferenças drásticas de filas entre os tipos de atendimento.

Este caso é pertinente a sistemas de filas que utilizam disciplina que envolve priorização da demanda, como são os casos da PRI e ADA. Ao realizar uma priorização, o sistema pode, em alguns momentos, deixar clientes muito tempo nas filas, pois a sua prioridade é mais baixa do que aquelas demandas que estão chegando. Ou seja, há uma troca de tempos em fila entre as maiores e as menores prioridades, diminuindo as primeiras e aumentando as segundas.

Para abordar esta situação, o algoritmo proposto neste estudo foi desenvolvido seguindo a lógica de identificação dos atendimentos com tal problema e a sua correção. A lógica deste não implica uma mudança significativa e complexa no funcionamento dos outros agentes e da própria disciplina.

O processo de identificação dos clientes “esquecidos” na fila (com pouca prioridade) foi feito através da fórmula

$$Tt = \left(Qf * \sum_0^{Lh} Tu \right) / Qu \quad (4)$$

, onde Qf é a quantidade atual em fila do tipo, Qu a quantidade de clientes no histórico do tipo, Tt é o tempo total estimado atualmente em fila do tipo e Tu o tempo em fila de clientes no histórico do tipo. Este tempo é uma estimativa, pois, como é um dado obtido através de um cálculo que utiliza a média do histórico, não se pode dizer que este será de fato o número preciso, somente uma aproximação dele.

Esta aproximação do tempo total em fila permite que o problema de priorização seja antecipado, analisando e comparando os tempos de cada tipo de atendimento. Através deste processo, atendimentos com tempos fora do comum, muito além de seu desvio padrão, podem ser identificados.

Na Tabela 6 pode ser visto um exemplo em que a classe de atendimento “Financeiro” possui um tempo total estimado na fila atual muito superior aos demais tipos, por ter a menor das prioridades. Já as classes “Documento” e “Dúvida” possuem tempos estimados na fila atual zerados, por terem as maiores prioridades e assim não possuindo fila atualmente.

Tabela 6 – Exemplo de sistema de atendimento com problema de priorização

| Tipo de atendimento | Prioridade de atendimento | Quantidade de clientes na fila atual | Tempo médio histórico | Tempo total na fila atual |
|---------------------|---------------------------|--------------------------------------|-----------------------|---------------------------|
| Dúvida | 2º | 0 | 0:03:23 | 0:00:00 |
| Matrícula | 4º | 4 | 0:09:33 | 0:38:12 |
| Financeiro | 5º | 9 | 0:17:01 | 2:33:09 |
| Documento | 1º | 0 | 0:02:55 | 0:00:00 |
| Negociação | 3º | 1 | 0:05:47 | 0:05:47 |

Fonte: do autor

Com estes dados, pode ser desenvolvido um agente que trate corretamente esta questão, não ocorrendo mais situações extremas de filas. A solução proposta no atual estudo, porém não sendo a única, é a alocação fixa de atendentes para as categorias que são identificadas como tendo problemas de priorização, de forma gradativa. Esta reserva de atendentes, para determinados tipos de serviços, pode afetar o funcionamento do resto do sistema. Para que esta interferência não seja muito impactante, existem algumas características e limitações para o seu funcionamento:

- Quantidade de atendentes alocados: O sistema não permite que muitos atendentes sejam alocados, mesmo que uma classe de atendimento esteja com filas exageradas. Isto pode levar à alocação demasiada de atendentes, com revezamento constante de tipos de serviços problemáticos, havendo poucos ou nenhum atendente para atender aquelas classes com maior prioridade. Estes, por sua vez, podem acumular filas e também se tornar problemáticos, e assim que um atendente deixar de ser reservado para outro tipo, o mesmo seria alocado para uma nova categoria.
- Escolha do atendente para alocação: Ao alocar um atendente, o Agente de priorização estuda todos os atendentes ativos e escolhe aquele que se mostra melhor para o tipo de atendimento envolvido. A escolha do agente deve ser muito inteligente para que resulte no melhor resultado possível, afetando o mínimo possível o funcionamento do resto do ambiente. Um atendente não deve ser escolhido, por exemplo, se o mesmo já está priorizando esta categoria de serviços através da lógica da disciplina ADA, pois isto não gera efeito nenhum.
- Tempo referência: Para identificar que um tipo de atendimento está com uma fila muito grande, o agente precisa de alguma referência que indique quando isto ocorre. Este tempo de referência (Tr), que não é fixo e pode ser ajustado conforme necessidade, deve ser alto o suficiente para que não reconheça qualquer situação de variação de chegada

de clientes à fila como um problema. Porém, também deve ser baixo o suficiente para que consiga determinar quando há um problema na fila o mais cedo possível.

- Frequência de ação: A frequência em que o agente pode agir sobre a fila também é algo personalizável conforme o ambiente. Ela pode ser:
 - Constante: Tempo de resposta ótimo, porém não faz muito sentido ser utilizado, pois as mudanças de estado do ambiente não ocorrem com tanta frequência. Como exemplo, um ambiente que acaba de ser reconhecido como instável não irá se tornar estável após a requisição de serviço ser completada;
 - Intervalada: Tempo de resposta variável, porém possui personalização dos intervalos conforme o ambiente e a configuração/disponibilidade das outras características/limitações.

No agente implementado no atual trabalho, a quantidade de atendentes disponíveis para alocação é definida como sendo a metade dos atendentes ativos. Já o tempo referência utilizado é de 1 hora, e a frequência de ação é de 6/hora.

Figura 16 – Pseudocódigo do Agente de priorização

```

Para tipo = 1 até 5 fazer
  Busca quantidade de clientes na fila ( $Qf$ );
  Calcula tempo total em fila ( $Tt$ );
  Se  $Tt > Tr$  então
    Busca quantidade de atendentes fixos ( $x$ );
    Busca quantidade de atendentes fixos do tipo ( $y$ );
    Se  $x < (quantidade\ de\ atendentes/2)$  e
       $y < (quantidade\ de\ atendentes/4)$  então
        Busca atendentes que não prioriza o tipo ( $z$ );
        Aloca  $z$  para o tipo;
    Fim se;
  Fim se;
  Busca quantidade de atendentes fixos do tipo ( $y$ );
  Se  $y \neq 0$  então
    Busca quantidade de clientes na fila ( $Qf$ );
    Calcula tempo total em fila ( $Tt$ );
    Se  $Tt = (Tr/2)$  então
      Busca atendente fixo do tipo ( $z$ );
      Desaloca  $z$  do tipo;
    Fim se;
  Fim se;
Fim para;

```

Fonte: do autor

O algoritmo desenvolvido para o agente é descrito na Figura 16 através do uso de pseudocódigo. Devem ser realizadas várias validações para identificar se realmente pode priorizar, ou mesmo se deve, uma classe de serviços. Ainda é demonstrado que não só é

necessária uma lógica para a alocação de atendentes, como também a desalocação dos mesmos após o momento crítico ter passado.

Tabela 7 – Exemplo de atuação do Agente de priorização com tempo referência de 1 hora

| Atuação | Tempo em fila atual | Atendentes alocados |
|---------|---------------------|---------------------|
| 1º | 0:20:00 | 0 |
| 2º | 1:10:00 | 1 |
| 3º | 1:25:00 | 2 |
| 4º | 0:45:00 | 2 |
| 5º | 0:17:00 | 1 |
| 6º | 0:02:00 | 0 |

Fonte: do autor

Na Tabela 7 pode ser visto um exemplo do efeito da atuação do Agente de priorização. Nesta situação, para que sejam alocados atendentes para o tipo de atendimento - além de atender todos os requisitos descritos anteriormente - o tempo em fila atual deve superar o tempo referência (1 hora). Já para desalocar atendentes é necessário que o tempo em fila atual baixe da metade deste tempo (30 minutos).

Como demonstrado pelo quarto acionamento da Tabela 7, ocorrem casos em que o tempo em fila atual já está abaixo do tempo referência, e ainda possui atendentes alocados. Este intervalo entre a alocação e desalocação serve principalmente para que não ocorram casos de o tempo em fila ficar oscilando perto do tempo referência, sem realmente conseguir baixar a um nível aceitável. A ideia é de que, se chegou ao ponto de reservar um atendente para o tipo, então o atendente deve ficar neste estado não somente até baixar do nível de reserva, porém atuar até que a situação atual esteja “aliviada”.

A construção deste agente, visando a atuação em conjunto com a disciplina ADA e outros agentes, contempla casos específicos gerados por tais peculiaridades. Um caso de exemplo é na escolha do atendente que será alocado quando o Agente de aprendizagem estiver ativo, devendo ser priorizados para escolha os atendentes que não são os melhores do ambiente na sua melhor categoria de serviços. Outro caso, prevendo a mudança no número de atendentes disponíveis, é de que a desalocação de atendentes deve prever a alocação maior do que o limite (pois o limite é imposto pelo número de atendentes atuantes) e desalocar prontamente um atendente, evitando que fique com muitas reservas e assim afete negativamente o ambiente.

A classificação deste agente pela sua categoria é de Agente Cognitivo, pois ele possui uma dinâmica variável de funcionamento, com várias dependências e limitações, as quais podem inclusive serem alteradas a qualquer momento. O seu acionamento não depende de um

evento no sistema, podendo ser feito a qualquer momento, sob demanda ou programado. Como possui muitas características variáveis e dependentes de outros dados, além de limitações em sua aplicação, seu desenvolvimento acaba sendo difícil e complexo.

5.3.2.3 Agente de dimensionamento

Um dos questionamentos que mais frequentemente aparecem em situações de sistemas de atendimento com filas, tanto em mercados, bancos, universidades, etc., é o porquê de não haver mais atendentes para consumir a demanda atual. Como na Teoria de Filas o sistema possui uma configuração fixa, e nela está contida o número de canais disponíveis, os modelos não preveem nem o aumento nem a diminuição deste número (pode haver necessidade de diminuição por parte do ambiente por outros motivos, como por exemplo custos, o que não é tratado no presente trabalho).

Esta questão remete à terceira observação citada anteriormente, em que a variação no recebimento de demandas não é contabilizada nos modelos e acaba gerando uma instabilidade no sistema. Este problema é pertinente a sistemas que possuem modelos com tempos de chegada variáveis.

Conforme Meyer (1974, p. 176, apud ABENSUR et al., 2003), a distribuição de probabilidade exponencial é a mais comum e adequada em situações em que as variações possam ser representadas por um número de ocorrências de um evento durante uma determinada medida de tempo, sendo a chegada de clientes à uma fila de atendimento um exemplo deste processo. Seguindo este padrão, não é difícil um ambiente entrar em uma situação de *stress*, acumulando rapidamente um nível alto de demandas, pois uma grande parte das ocorrências de chegadas de clientes se encontra próxima da medida mínima de tempo.

O tratamento deste problema de *stress* na fila, no atual projeto, parte do mesmo princípio acometido no Agente de priorização: identificar em que ponto o estado do sistema pode ser considerado como “em *stress*” e, a partir daí, agir sobre o número de atendentes. Esta ação pode ser tanto aumentando como diminuindo a capacidade ambiente, mantendo assim um número estável de atendentes em regime de funcionamento normal.

A diferença na obtenção deste ponto de mudança de estados entre os Agentes de priorização e de dimensionamento é a quantidade de tipos de serviços considerados. No primeiro, é levado em consideração somente a fila atual e o histórico de uma classe específica de atendimento, porém no segundo é levado em consideração todas as categorias de requisições.

Para a obtenção do momento de *stress* do sistema, é utilizado o cálculo

$$Tg = \sum_1^{Ta} \left(\left(Qf * \sum_0^{Lh} Tu \right) / Qu \right) \quad (5)$$

que resulta no tempo total estimado em fila para todos os tipos (Tg). Assim como no caso do Agente de priorização, o dado calculado aqui também é somente uma aproximação, provavelmente não refletindo corretamente o que irá ocorrer, mas podendo ser usado para prever casos macros de *stress* no sistema.

Tabela 8 – Exemplo de sistema de atendimento com *stress*

| Tipo de atendimento | Prioridade de atendimento | Quantidade na fila atual | Tempo médio histórico | Tempo total na fila atual |
|---------------------|---------------------------|--------------------------|-----------------------|---------------------------|
| Dúvida | 2° | 13 | 0:03:23 | 0:43:59 |
| Matrícula | 4° | 7 | 0:09:33 | 1:06:51 |
| Financeiro | 5° | 9 | 0:17:01 | 2:33:09 |
| Documento | 1° | 8 | 0:02:55 | 0:23:20 |
| Negociação | 3° | 11 | 0:05:47 | 1:03:37 |
| Total | - | 48 | - | 5:50:56 |

Fonte: do autor

O exemplo da Tabela 8 mostra o mesmo sistema apresentado pela Tabela 6, porém com diferentes quantidades de clientes na fila atual, caracterizando um sistema em estado de *stress*. Nota-se que mesmo a categoria “Financeiro” sendo aquela com última prioridade, a mesma possui menos clientes na fila atualmente dos que outras classes prioritárias, como “Negociação” e “Dúvida”, descaracterizando parcialmente o ambiente para atuação do Agente de priorização, conforme sua definição.

O desenvolvimento do agente pode, então, basear-se nas definições descritas anteriormente para ser realizado. Neste estudo o algoritmo escolhido para o agente proposto trabalha com níveis de *stress*, em que o sistema é classificado em várias situações de instabilidade baseando-se no tempo total estimado na fila atual, atuando no número de atendentes de forma gradual. Este agente também possui algumas características e limitações, assim como o Agente de priorização. Além das características Tempo de referência e Frequência de ação, explanadas no agente anterior, uma outra se destaca no atual agente:

- Quantidade de atendentes ativos: Deve haver sempre um número mínimo de atendentes ativos, para que possa manter o sistema funcional em momentos estáveis. Já o máximo

de atendentes é algo que, no presente trabalho, não é uma questão altamente abordada, pois esta limitação faz mais sentido em ambientes totalmente reais, não em virtuais.

A quantidade de atendentes disponíveis para dimensionamento foi definida para possuir, como limite mínimo, o mesmo número de atendentes utilizados quando o agente não é executado (número padrão do ambiente). Já o limite máximo definido é o mesmo que o limite do ambiente de simulação, ou seja, 8 atendentes. Como tempo referência é utilizado 15 minutos e a frequência de ação é de 6/hora.

Figura 17 – Pseudocódigo do Agente de dimensionamento

```

Busca quantidade de clientes na fila ( $Qf$ );
Calcula tempo total em fila ( $Tg$ );
Se quantidade de atendentes < máximo e
   $Tg > (Tr * \text{quantidade de atendentes})$  então
    Adiciona atendente ao ambiente;
Fim se;
Se quantidade de atendentes > mínimo e
   $Tg < (Tr * \text{quantidade de atendentes}/2)$  então
    Remove atendente ao ambiente;
Fim se;

```

Fonte: do autor

A Figura 17 apresenta um pseudocódigo do algoritmo implementado para o agente. A complexidade desta lógica notavelmente menor, comparado ao Agente de priorização. Ambos algoritmos possuem estrutura parecida de funcionamento: em um primeiro momento, a lógica para a alteração do ambiente a que são propostos; a seguir possuem a lógica para desfazer tal alteração.

Tabela 9 – Exemplo de atuação do Agente de dimensionamento com tempo referência de 1 hora

| Atuação | Tempo em fila atual | Número de atendentes | Tempo necessário para incremento | Tempo necessário para decremento |
|---------|---------------------|----------------------|----------------------------------|----------------------------------|
| 1º | 0:20:00 | 4 | 1:00:00 | - |
| 2º | 1:10:00 | 5 | 1:15:00 | 0:00:23 |
| 3º | 1:50:00 | 6 | 1:30:00 | 0:00:28 |
| 4º | 0:50:00 | 6 | 1:30:00 | 0:00:28 |
| 5º | 0:26:00 | 5 | 1:15:00 | 0:00:23 |
| 6º | 0:02:00 | 4 | 1:00:00 | - |

Fonte: do autor

Na Tabela 9 pode ser visto a aplicação do Agente de dimensionamento em um ambiente no momento em que ocorre um *stress* momentâneo na chegada de clientes. Pode ser visto na terceira atuação que a alteração do número de atendentes se dá gradativamente, nunca ocorrendo em quantidades maiores que 1. Neste agente também há um intervalo de tempo entre

a atuação para lados diferentes, ou seja, o ponto em que o número de agentes é incrementado não é o mesmo para quando é decrementado.

Assim como no Agente de priorização, a meta é deixar ativo o resultado da aplicação do agente por um tempo maior, pois o ponto de atuação passado é visto como um ponto ruim, e o intuito é melhorar o ambiente para que não volte para tal lugar logo em seguida. Para serem calculados os tempos de incremento e decrementos, são utilizadas as fórmulas

$$Ti = (Tr/4) * Na \quad (6)$$

$$Td = (Tr/4) * (Na/2) \quad (7)$$

, em que Na é o número de atendentes, Td o tempo necessário para decremento e Ti o tempo necessário para incremento.

A categoria de enquadramento do Agente de dimensionamento é a de Agente Cognitivo, devido a vários fatores: trabalha com múltiplas variáveis complexas; nem sempre possui o mesmo comportamento para situações idênticas; não depende de eventos para a sua utilização, podendo ser acionado em momentos distintos, sob demanda ou de forma agendada, ou através de eventos no ambiente. Ainda deve ser salientado que existe uma parte muito mais complexa neste agente, que é a parte não tratada no atual trabalho, referente à um ambiente real e todas as características e limitações envolvidas neste. Com todos estes fatores, a sua implementação e seu funcionamento acabam sendo altamente complexos.

5.4 AMBIENTE DE SIMULAÇÃO FINAL

Após o término do desenvolvimento de todos os algoritmos e agentes, o ambiente de simulação pode ser configurado para utilizar qualquer combinação das funcionalidades criadas, a fim de testar e avaliar o seu desempenho. Algumas combinações podem não gerar o efeito desejado, e por isto se torna importante a possibilidade de poder adicionar ou remover recursos de cada algoritmo proposto.

Na Tabela 10 é apresentado um exemplo de combinações que podem ser feitas a partir da utilização da disciplina de filas PRI. Esta mesma quantidade de combinações se repete para as outras disciplinas, gerando um total de 32 combinações únicas.

Tabela 10 – Exemplo de combinações de funcionalidades com a disciplina PRI

| Disciplina | Aprendizagem | Priorizador | Dimensionador |
|------------|--------------|-------------|---------------|
| PRI | Não | Não | Não |
| | Sim | Não | Não |
| | Não | Sim | Não |
| | Não | Não | Sim |
| | Sim | Sim | Não |
| | Sim | Não | Sim |
| | Não | Sim | Sim |
| | Sim | Sim | Sim |

Fonte: do autor

A implementação dos casos de testes, simulações e análise dos resultados é descrita no capítulo seguinte, utilizando os recursos desenvolvidos até então para elaboração dos algoritmos adaptativos.

6 EXPERIMENTOS

Com o ambiente de simulação desenvolvido, torna-se necessário a definição de como o mesmo deve ser utilizado para que os objetivos do estudo possam ser alcançados. Balci et al. (1990 apud CHWIF, 1999) define ambiente de simulação como “um conjunto integrado de ferramentas de *hardware* e *software* que fornecem um suporte automatizado, a um custo eficiente, por todo o ciclo de vida de um modelo de simulação”. Já Harrell et al. (2000) sugere que as simulações são concepções de sistemas reais, utilizando modelos computacionais para a obtenção de melhores resultados através de avaliações comparativas.

Para a comparação de desempenho das combinações de algoritmos desenvolvidos no atual trabalho, devem ser executados os mesmos casos de testes em cada modelo proposto, assim gerando resultados de diferentes tratamentos para uma única situação. Banks (2000) explica que uma simulação deve consistir de uma história artificial, baseada em fatos reais, e as características desta história podem então ser observadas e manipuladas para obterem-se casos próximos aos da realidade. Em sistemas complexos e dinâmicos, que possuem grandes variações de dados e alta imprevisibilidade, a aplicação de simulação computacional é algo natural e propício para tal fim (PIDD 1998 apud XAVIER et al., 2009).

Assim, este capítulo aborda a montagem dos casos de testes, a execução das simulações com diversos algoritmos diferentes, tanto de Teoria de Filas como da Política Adaptativa, e a análise dos resultados destas.

6.1 CASOS DE TESTE

Os casos de teste definidos para serem utilizados neste estudo podem ser divididos em dois grupos:

- Constantes: São aqueles em que os dados de entrada, ou as chegadas de clientes ao sistema, possuem cadência imutável, não havendo alteração de seu comportamento durante a simulação e assim sendo de mais fácil uso;
- Variáveis. Estes não seguem o mesmo comportamento de entrada de dados durante todo o processo, havendo no mínimo uma mudança de comportamento. Por sua vez, tendem a simular com mais fidelidade casos que se espelham em situações reais, que envolvam

variáveis com comportamentos sazonais ou outras características que definam mudanças do padrão básico do ambiente.

A escolha de tais modelos de simulação considera que a nova política adaptativa de atendimento deve ser primeiramente confrontada com as políticas já existentes na Teoria de Filas. Ao utilizar casos de testes constantes, o objetivo é validar o seu comportamento nos casos mais simples possíveis, sem variações. Já a utilização de simulações variáveis implica na imagem de histórias reais, em que os padrões de funcionamento podem mudar a qualquer momento e poucos elementos podem ser considerados “fixos”.

6.1.1 Chegada de clientes

A definição dos dados de entrada dos casos de testes, no presente trabalho, consiste em definir uma quantidade fixa de clientes que chegam ao sistema de atendimento em um tempo fixo, desconsiderando desvios padrões, particular para cada tipo de atendimento. Ou seja, caso definido que 20 clientes do atendimento “Financeiro” devem chegar em 1 hora, ou 1 a cada 3 minutos, não necessariamente isto é seguido à risca devido à natureza variável dessa característica. Porém, o resultado ao final é muito próximo ou até igual a 20 clientes em 1 hora.

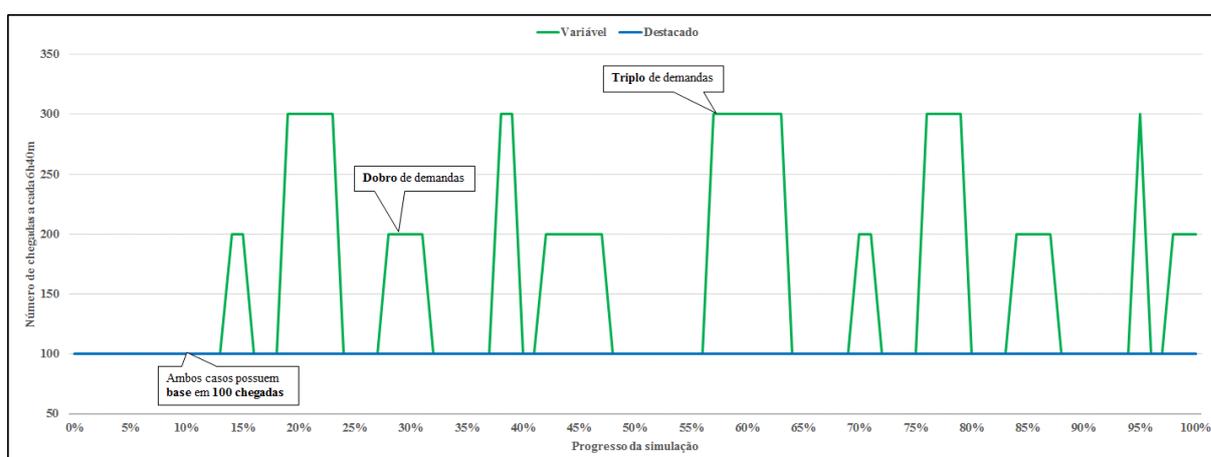
Para fins de padronização em todas as simulações, é considerado que a distribuição utilizada é a exponencial, por representar os cenários reais com maior fidelidade. O número considerado de clientes que chegam ao sistema é de 100 a cada 6 horas e 40 minutos com desvios padrões, sendo uma medida escalável no ambiente e, com isso, é necessário apenas ser definido um padrão. Esta última definição é feita de modo “global” no sistema para casos constantes, sendo que a quantidade de chegadas pode variar entre os tipos de atendimento, em qualquer proporção que resulte no limite proposto. Para os casos variáveis, a última definição é tomada como base, porém pode ser alterada conforme o momento em que se encontra a simulação.

As distribuições de chegadas foram formuladas conforme segue:

- **Balanceado** – Representa a chegada, na mesma quantidade, de todas as demandas de atendimento em um intervalo de tempo. É a representação de um ambiente totalmente igual na questão de entrada de dados, o que também é encontrado em casos que não possuam classes de atendimento diferenciados, pois como todos tem as mesmas características, é como se fosse somente um tipo;

- Destacado – Neste caso existe uma categoria de serviços específico com demanda proeminente em relação aos demais. Esta representação pode ser interpretada de maneiras distintas, como, por exemplo, um momento sazonal de um tipo de atendimento, que possui uma demanda muito maior, ou então simplesmente um ambiente de atendimento que possui demanda seis vezes maior de um atendimento específico do que os outros;
- Predominante – Aqui é representada uma situação semelhante à anterior, porém mais acentuada, sendo ainda mais notável a diferença entre a classe de serviços destacada e as demais. A significância do tipo proeminente é muito grande, comparado ao total de atendimentos realizados no ambiente, sendo a demanda deste definida como doze vezes maior do que as demais;
- Estável – Este caso contempla uma situação em que duas classes de atendimento possuem demanda igual, sendo a dos outros três menores e não necessariamente igual. Representa sistemas de atendimento que possuem duas variedades de requisições “principais”, com outras sendo “auxiliares”.
- Variável – Neste, o número de clientes que chegam ao ambiente recebe variações constantes, porém a distribuição entre os tipos de atendimento segue a mesma. A lógica utilizada para esta mudança de comportamento, utilizando como base o caso *Destacado*, é mostrada na Figura 18.

Figura 18 – Comparação entre demandas dos cenários *Destacado* e *Variável*



Fonte: do autor

Na Figura 18 são destacados os três pontos distintos entre ambos cenários, que são a base de chegadas a cada 6h40m de 100 clientes, e as variações para o dobro e triplo desta demanda no caso *Variável*.

As chegadas de clientes ao ambiente foram repartidas entre os tipos de atendimento, conforme as distribuições formuladas, seguindo Tabela 11.

Tabela 11 – Partilha de chegadas de clientes entre os tipos de atendimento

| Caso de teste | Dúvida | Matrícula | Financeiro | Documento | Negociação | Total |
|---------------------|--------|-----------|------------|-----------|------------|-------|
| <i>Balanceado</i> | 20 | 20 | 20 | 20 | 20 | 100 |
| <i>Destacado</i> | 10 | 60 | 10 | 10 | 10 | 100 |
| <i>Predominante</i> | 5 | 80 | 5 | 5 | 5 | 100 |
| <i>Estável</i> | 10 | 40 | 40 | 5 | 5 | 100 |
| | 10 | 60 | 10 | 10 | 10 | 100 |
| <i>Variável</i> | 20 | 120 | 20 | 20 | 20 | 200 |
| | 30 | 180 | 30 | 30 | 30 | 300 |

Fonte: do autor

6.1.2 Atendimento das demandas

Para manter o foco na variância da entrada de demandas na fila, as características utilizadas no atendimento dos clientes foram distribuídas de forma equilibrada entre os tipos de demandas. Sendo assim, todas as classes possuem a mesma distribuição, mesmo tempo base de atendimento, e mesmo desvio padrão do tempo base.

Neste trabalho as propriedades de atendimento escolhidas foram de distribuição normal, a qual é semelhante às situações reais, e tempo base de 7 minutos com desvio padrão de 1:30 minutos, sendo esta uma característica escalável assim como nas chegadas de demandas.

6.2 SIMULAÇÕES

Para comparações básicas de desempenho, os algoritmos escolhidos devem ser submetidos aos casos de testes propostos e, ao final, emitir um relatório da situação do ambiente. Entretanto, como o ambiente trabalha com processos variáveis, somente uma simulação acaba não sendo suficiente para se obter um resultado conclusivo, pois esta pode ter tido um comportamento aquém do normal do algoritmo, ou um desvio no seu padrão.

Devido a isso, é uma boa prática a realização de várias simulações para cada algoritmo, sendo os seus resultados computados e obtido a média entre eles, evitando assim utilizar dados de simulações que saíram do seu comportamento padrão. No presente projeto é definido que cada caso de teste deve ser executado 10 vezes, para que seja feita a média destes e obtido o resultado do algoritmo.

As simulações foram realizadas 10 vezes para cada algoritmo combinado, utilizando os casos de teste definidos anteriormente, com um tempo total de 200h por simulação. Os resultados foram obtidos através da média das saídas geradas pelas repetições. As combinações de algoritmos escolhidas são demonstradas na Tabela 12.

Tabela 12 – Combinações de algoritmos escolhidas

| Disciplina | Aprendizagem | Priorizador | Dimensionador |
|------------|--------------|-------------|---------------|
| FIFO | Não | Não | Não |
| LIFO | Não | Não | Não |
| PRI | Não | Não | Não |
| FIFO | Sim | Não | Não |
| LIFO | Sim | Não | Não |
| PRI | Sim | Não | Não |
| ADA | Sim | Não | Não |
| ADA | Não | Sim | Não |
| ADA | Não | Não | Sim |
| ADA | Sim | Sim | Não |
| ADA | Sim | Não | Sim |
| ADA | Não | Sim | Sim |
| ADA | Sim | Sim | Sim |

Fonte: do autor

A execução dos testes foi realizada em um computador com sistema operacional Windows 10 64-bits, processador Intel Core i7-3610 2,30Ghz e 12Gb de memória RAM. A média de tempo para completar as 10 repetições foi de 18 minutos, havendo casos mínimos de 6 minutos e casos máximos de 2 horas e 15 minutos, além de simulações que não puderam ser executadas (algoritmos que usam LIFO no cenário *Variável*).

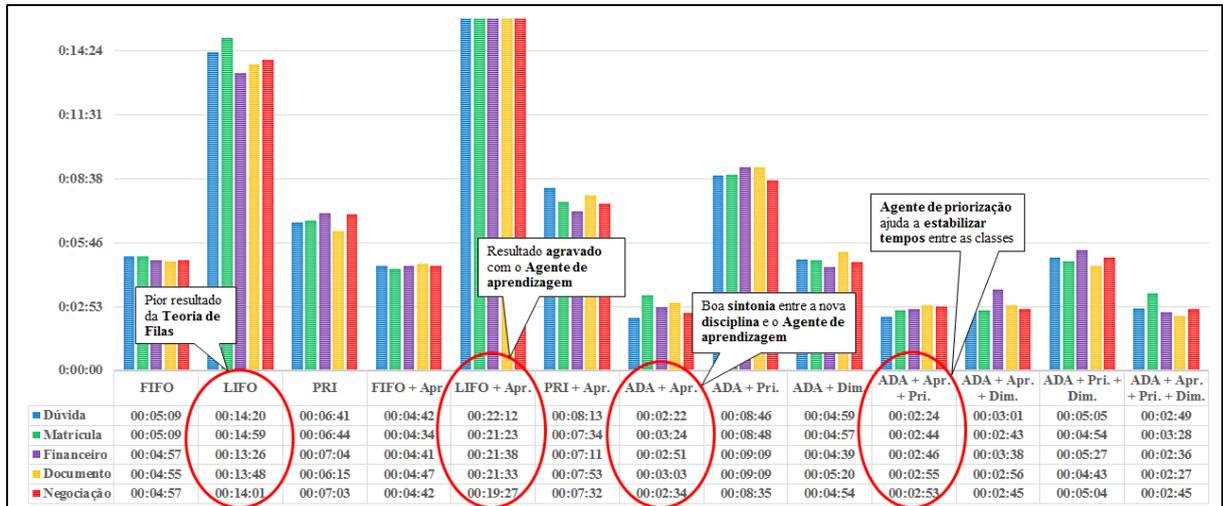
6.3 RESULTADOS

Os tempos citados nos cenários a seguir foram aproximados, assim facilitando as comparações e também a leitura. Para identificar os algoritmos escolhidos, são utilizados termos diminutivos de suas nomenclaturas, salvo as disciplinas que já possuem tal referência.

6.3.1 Caso de teste 1 – Balanceado

Na Figura 19 são comparados os tempos de fila dos atendimentos no histórico do ambiente, para cada algoritmo utilizado, sendo estes tempos separados por tipo de atendimento.

Figura 19 – Balanceado: Tempo médio de fila X Algoritmo X Tipo de atendimento

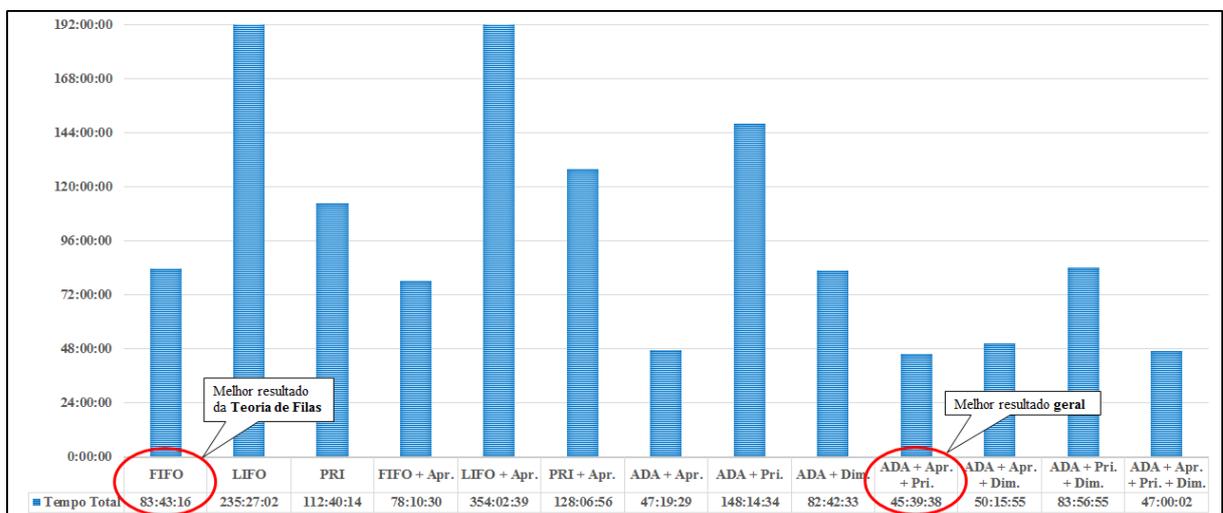


Fonte: do autor

Inicialmente é ressaltada a característica que, entre os tipos de serviços de um mesmo algoritmo, não há variações significativas de tempos, seguindo a lógica proposta no caso *Balanceado*. Existe uma variação notável dos tempos médios de fila entre os algoritmos, sobressaindo-se negativamente aqueles que utilizam a disciplina LIFO (LIFO e LIFO + Apr.). Os resultados, em sua maioria, acabam sendo positivos com a aplicação da Política Adaptativa de Atendimento, em especial aqueles que utilizam o Agente de aprendizagem, como visto nos algoritmos ADA + Apr. e ADA + Apr. + Pri.

Na Figura 20 é mostrado um gráfico com o tempo total de fila do histórico do ambiente, sendo este tempo acumulado entre todas as classes de serviços, para cada algoritmo aplicado.

Figura 20 – Balanceado: Tempo total de fila do histórico X Algoritmo



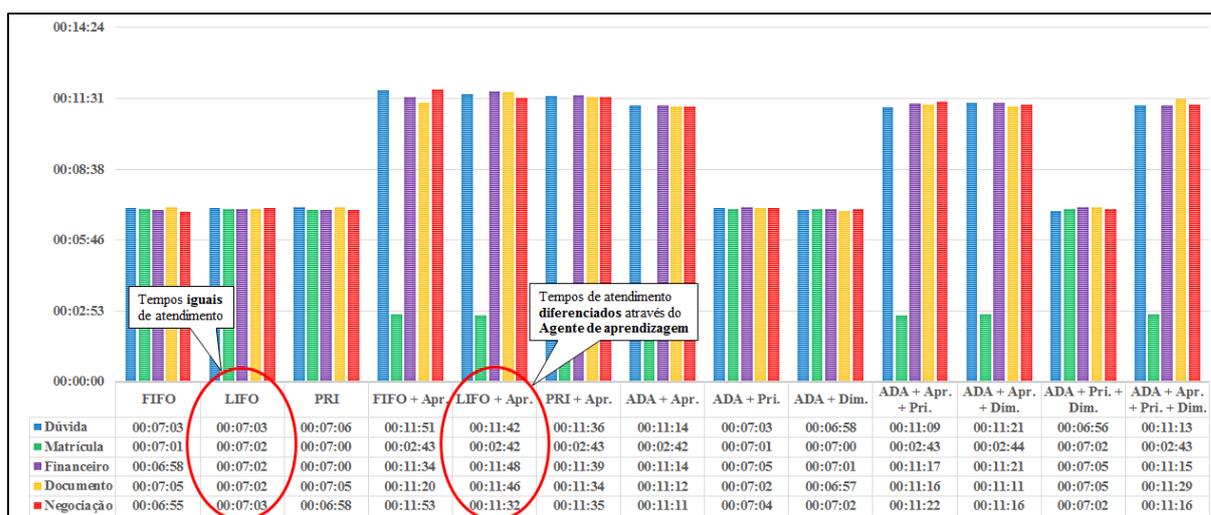
Fonte: do autor

Através deste gráfico, é possível ver que mesmo em ambientes *Balancedos* a Política Adaptativa de Atendimento consegue reduzir o tempo total de fila consideravelmente. Uma comparação prática pode ser feita entre os algoritmos FIFO e ADA + Apr. + Pri., sendo os tempos de **83:43:16** e **45:39:38**, respectivamente. Isso representa uma diminuição de **46%** do tempo em fila de todo o ambiente.

6.3.2 Caso de teste 2 – Destacado

Neste caso de teste pode ser facilmente observado a mudança que ocorre nos tempos de atendimento quando o algoritmo possui o Agente de aprendizagem. Esta alteração é apresentada pela Figura 21.

Figura 21 – Destacado: Tempo médio de atendimento X Algoritmo X Tipo de atendimento

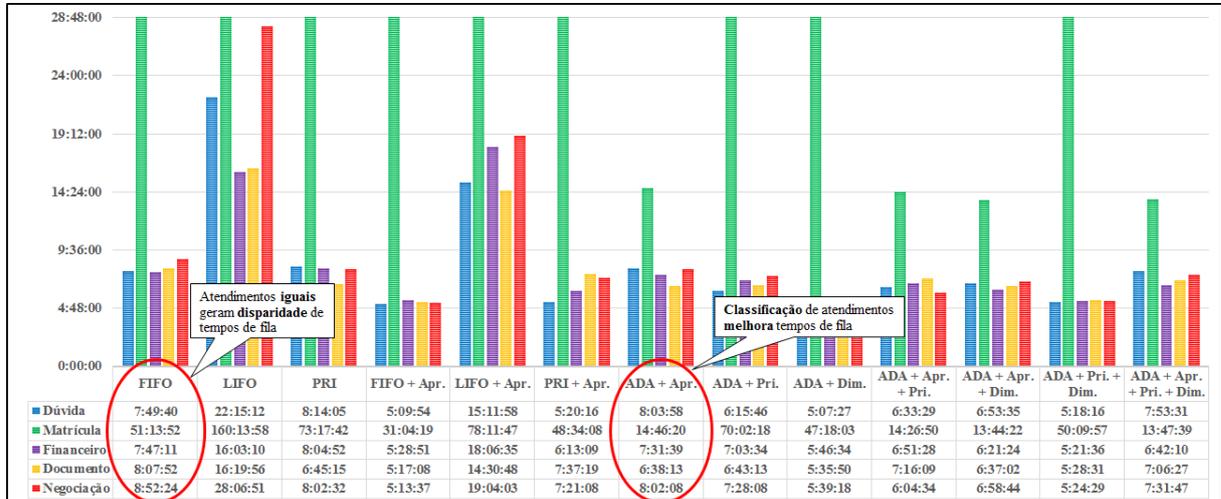


Fonte: do autor

É visível que, quando o Agente de aprendizagem está presente, a classe de atendimento que mais ocorre (neste caso, Matrícula) tem o seu tempo de atendimento muito abaixo dos demais tipos. O gráfico mostra que no algoritmo LIFO + Apr., um destes casos, a média dos atendimentos de Matrícula é de 2:42, enquanto que os outros atendimentos possuem médias de 11:42. Já no algoritmo LIFO, quando não há atuação do Agente de aprendizagem, o tempo médio de atendimento de todas as classes fica na faixa de 7:02.

Este padrão de comportamento, perdendo tempo de atendimento em alguns tipos para ganhar em outro, parece estranho em um primeiro momento, transparecendo uma perda de eficácia nos atendimentos. Porém esta teoria é questionada com os resultados demonstrados no gráfico da Figura 22.

Figura 22 – Destacado: Tempo total de fila do histórico X Algoritmo X Tipo de atendimento

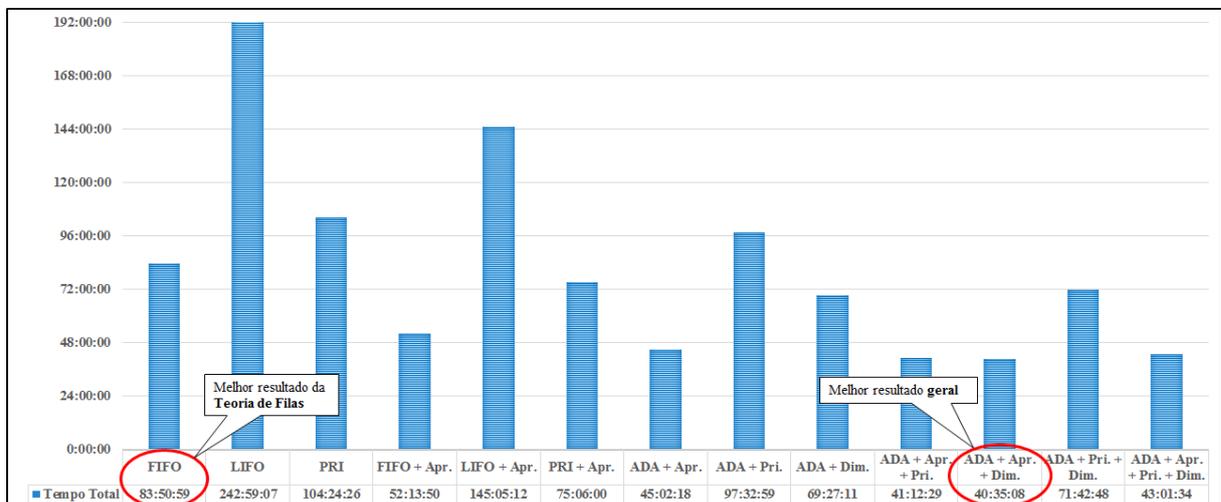


Fonte: do autor

Os tempos totais de fila do histórico mostram que o tipo de atendimento Matrícula é o que possui maior representação nos tempos de fila do sistema, com ou sem a atuação do Agente de aprendizagem. O gráfico apresenta, no algoritmo tradicional FIFO, que o tempo total de fila do tipo Matrícula é de 51:13:52, enquanto que nas outras classes este tempo é em torno de 8:09:17.

Este gráfico também demonstra a efetividade do padrão de comportamento citado anteriormente. Em sua maioria, os algoritmos que aplicam o Agente de aprendizagem diminuem drasticamente o tempo de fila do tipo Matrícula, sem sacrificar os tempos dos outros tipos, como é o caso do algoritmo ADA + Apr.

Figura 23 – Destacado: Tempo total de fila do histórico X Algoritmo



Fonte: do autor

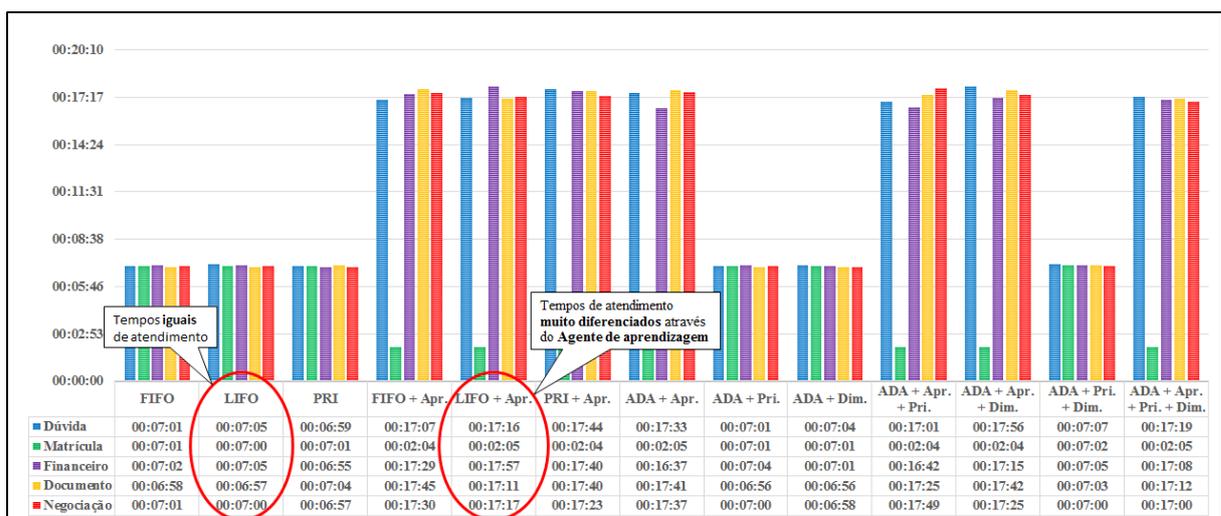
Na Figura 23 é apresentado um gráfico com o tempo total de fila do histórico, para cada algoritmo utilizado na simulação. Nesta pode ser visto que a aplicação da Política Adaptativa de Atendimento também melhora os tempos de fila em casos *Destacados*, usando a especialidade deste caso (a maioria dos atendimentos é de um único tipo) em favor da obtenção de melhores resultados. Isto é possível devido às características de identificação de peculiaridades e utilização das mesmas em favor do sistema, assim como foi descrito no desenvolvimento da disciplina ADA e dos agentes.

Ao comparar os algoritmos FIFO e ADA + Apr. + Dim., com seus tempos totais de fila de **83:50:59** e **40:35:08**, respectivamente, pode ser observado o ganho possível com a aplicação da nova política. A queda do tempo é de **53%**, ou seja, menos da metade em relação à utilização única da disciplina FIFO da Teoria de Filas.

6.3.3 Caso de teste 3 – Predominante

Os padrões dos resultados do caso *Predominante* seguiram as expectativas após o *Destacado*, pois como descrito anteriormente, o primeiro cenário é o segundo de uma forma mais aguda. Porém os números mostram que houve uma leve mudança na proporção entre a demanda acentuada e as demais.

Figura 24 – Predominante: Tempo médio de atendimento X Algoritmo X Tipo de atendimento



Fonte: do autor

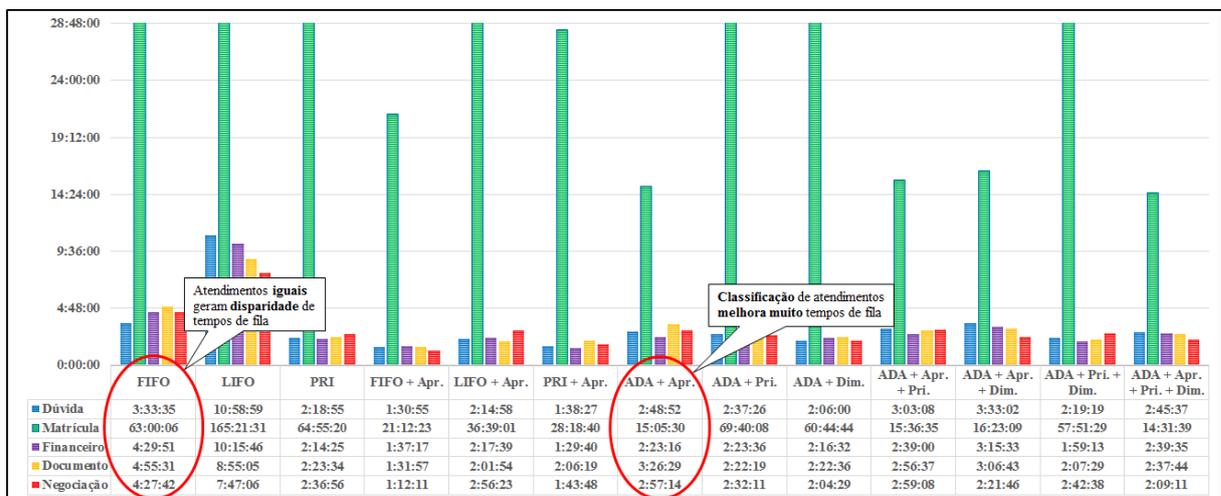
A Figura 24 apresenta a mesma situação que a Figura 21 aborda, com a mesma diferença entre os algoritmos que possuem Agente de aprendizagem e os que não o possuem. Porém,

pode ser visto que a diferença entre o tipo Matrícula e os demais, no atual cenário, é muito mais acentuada nos algoritmos com tal agente.

Comparando diretamente com os resultados do caso anterior, o algoritmo LIFO + Apr. possui média do tempo de atendimento da categoria Matrícula de 2:05, ao mesmo tempo que as outras classes têm esta média em 17:25. Contudo, no algoritmo LIFO, sem a atuação do Agente de aprendizagem, o tempo médio continua em 07:01. Enquanto que este último se mantém igual ao caso *Destacado*, as outras medições mudam bastante: o tipo de serviço crítico do ambiente melhora o tempo de atendimento, em média, 0:37, representando um ganho de 28%; nos outros tipos a piora no tempo médio oscila na casa de 5:43, significando uma perda de 50%.

Este comportamento transparece a mesma perda de efetividade nos atendimentos que o caso *Destacado* possui. Porém, novamente, com mais informações de outras situações, foi visto que tal situação realmente é válida. Isto é demonstrado através de um gráfico na Figura 25.

Figura 25 – Predominante: Tempo total de fila do histórico X Algoritmo X Tipo de atendimento



Fonte: do autor

Neste gráfico é possível ver que a Matrícula continua sendo o serviço distinto dos demais, assim como no caso anterior. Porém os tempos e proporções também estão diferentes, seguindo a mudança observada na Figura 24.

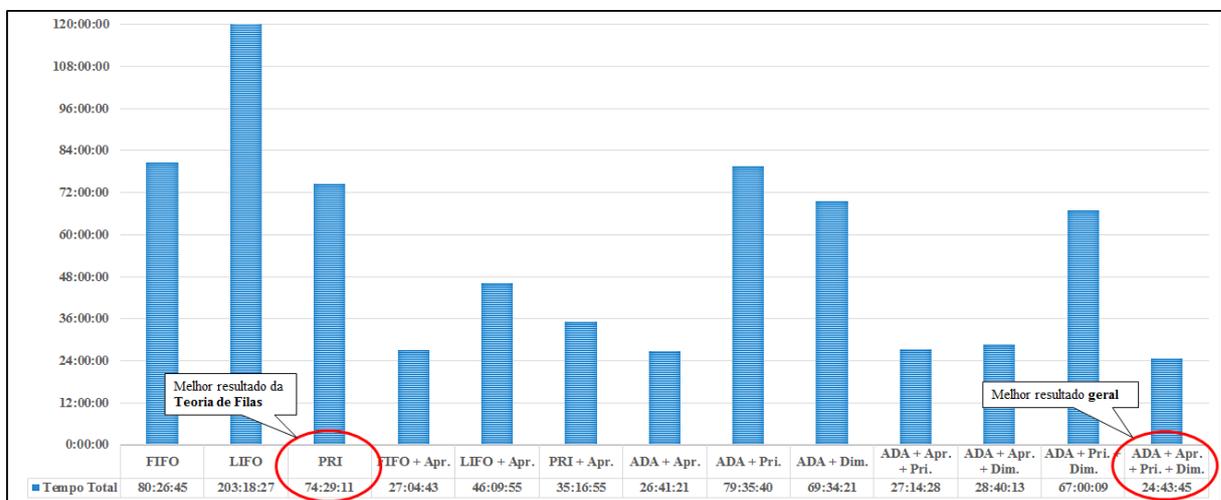
O algoritmo FIFO possui tempo de fila do tipo mais representativo de 63:00:06, enquanto que as outras classes de serviços têm uma média de 4:21:40. Em comparação ao caso de testes anterior, os atendimentos de Matrícula passaram a acumular mais tempo em fila, 11:46:14 para ser mais preciso. Em compensação, os outros quatro obtiveram tempos médios

aproximadamente 4 horas menores cada. Houve um comportamento invertido nas duas situações, piorando o tempo em um e melhorando nos outros.

Já o algoritmo ADA + Apr., também já analisado, possui um comportamento parecido, porém o incremento de tempo médio para a classe mais significativa é menor. No caso de testes anteriores, era de 14:46:20, e no atual é de 15:05:30. Enquanto que no algoritmo FIFO este atendimento teve acréscimo no período de 21%, o ADA + Apr. apresentou acréscimo de somente 2%. Isto demonstra que, ao “piorar” a situação de um tipo de atendimento, deixando-o mais expressivo em relação aos demais, a Política Adaptativa de Atendimento consegue absorver a mudança de maneira mais eficaz do que modelos da Teoria de Filas.

A Figura 26 pontua o resultado compreendido até então, apresentando novamente uma eficácia elevada da nova política. Destacam-se também os resultados dos algoritmos que agregam a Teoria de Filas com o Agente de aprendizagem, e o algoritmo da Política Adaptativa de Atendimento que possui todos os elementos desenvolvidos (ADA + Apr. + Pri. + Dim.), demonstrando a integração no funcionamento entre os mesmos.

Figura 26 – Predominante: Tempo total de fila do histórico X Algoritmo



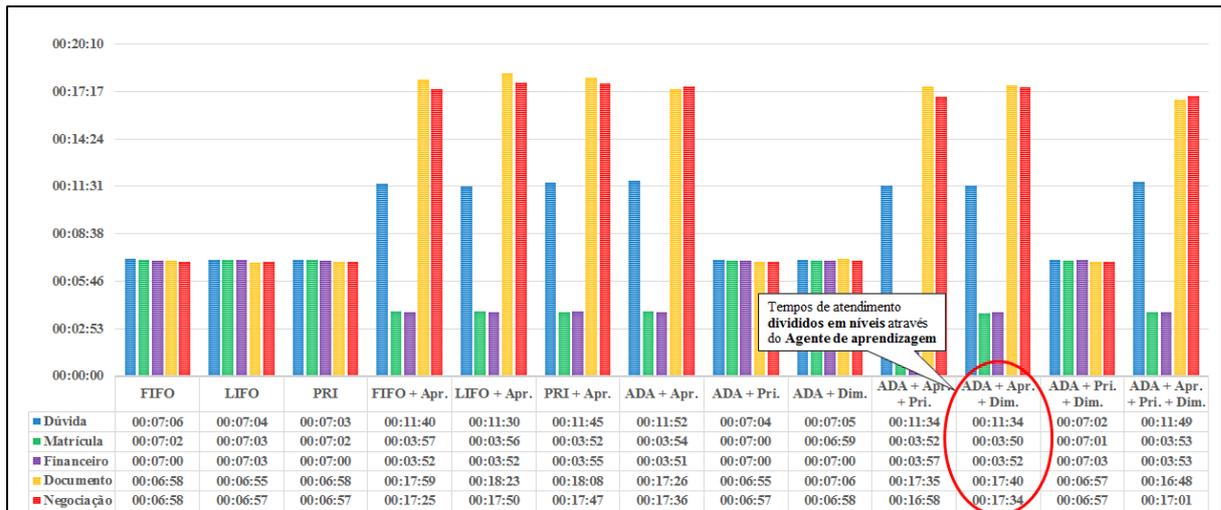
Fonte: do autor

Este caso de testes evidencia que nem sempre o algoritmo FIFO é o que possui melhores resultados dentre os representantes da Teoria de Filas. É visto que o modelo PRI obteve resultado mais significativo, com tempo total de fila de **74:29:11**, porém quando comparado aos algoritmos novos, na maioria dos casos a diferença é grande. Aqui é comparado com o modelo que utiliza todos os módulos desenvolvidos, ADA + Apr. + Pri. + Dim., o qual obteve tempo total de fila de **24:43:45**. A diminuição do tempo de fila, nesta comparação, é de **66%**.

6.3.4 Caso de teste 4 – Estável

Nas simulações realizadas utilizando este caso de teste, pode ser verificado que os novos elementos nem sempre causam um ganho no desempenho do ambiente. Existem situações em que um Agente pode mais atrapalhar do que ajudar no processo de atendimento.

Figura 27 – Estável: Tempo médio de atendimento X Algoritmo X Tipo de atendimento

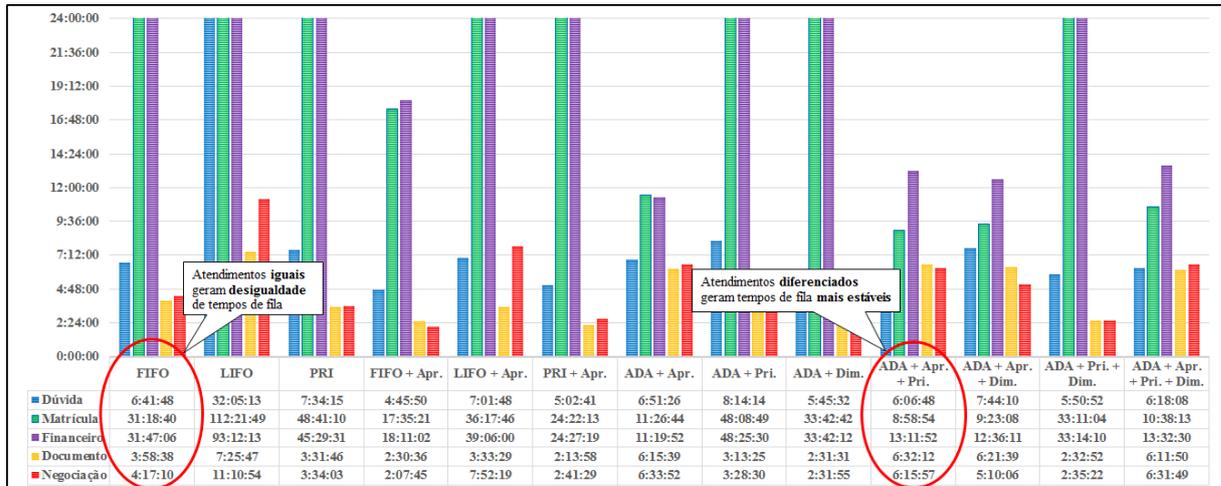


Fonte: do autor

Através do gráfico apresentado na Figura 27 é possível identificar que o comportamento dos algoritmos é comparável aos dois casos anteriores, *Destacado* e *Predominante*. Entretanto, ao invés de uma classe de serviços prevalecer sobre as demais no número de chegadas, esta propriedade é dividida em três níveis: duas classes mais representativas, de mesma preponderância (Matrícula e Financeiro); um tipo intermediário (Dúvida); e outros dois que demandam menos do ambiente, porém em mesma quantidade, que os demais (Documento e Negociação). Esta escala é percebida com mais facilidade observando os algoritmos que aplicam o Agente de aprendizagem, o qual trabalha diretamente com tais variações de demandas.

Com a Figura 28 é possível verificar que o mesmo processo também ocorre com os tempos de fila de cada tipo de atendimento ao aplicar os elementos adaptativos. Nas classes de serviços que possuem mais requisições, os tempos sofrem uma redução, enquanto que as classes com menos requisições sofrem o inverso, ou seja, um aumento.

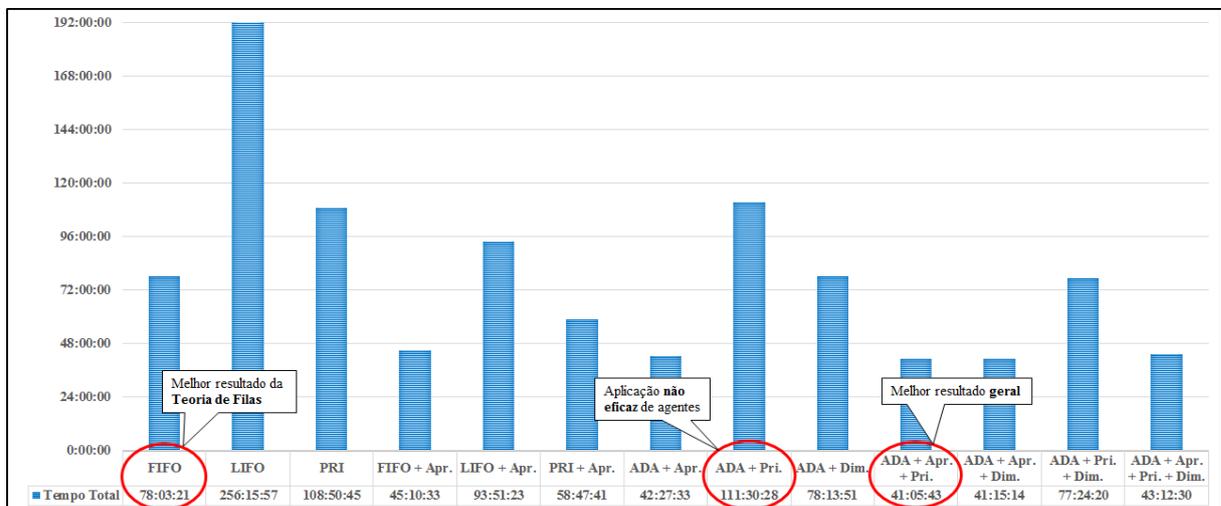
Figura 28 – Estável: Tempo total de fila do histórico X Algoritmo X Tipo de atendimento



Fonte: do autor

Mesmo que o procedimento de melhoria no sistema seja equiparável, o resultado final acabou trazendo dados novos. Este cenário mostrou que, em algumas situações, a aplicação de um determinado recurso pode não ajudar, mas sim piorar o desempenho do sistema.

Figura 29 – Estável: Tempo total de fila do histórico X Algoritmo



Fonte: do autor

A Figura 29 apresenta uma comparação entre o desempenho de três algoritmos que demonstra a opinião formada até então neste cenário. A aplicação da Política Adaptativa continua melhorando o desempenho na maioria dos algoritmos utilizados, contudo há casos em que a situação se inverte, culminando em uma perda de performance em sua execução.

O ganho na utilização de um algoritmo com a Política Adaptativa de Atendimento, ADA + Apr. + Pri., com tempo total de fila de **41:05:43**, ao invés de um puramente da Teoria de Filas, FIFO, com tempo total de fila de **78:03:21**, é de **47%**. Entretanto, a diferença deste último

com um outro algoritmo que também aplica a Política Adaptativa, ADA + Pri., com tempo total de fila de 111:30:28, é uma perda de 43%.

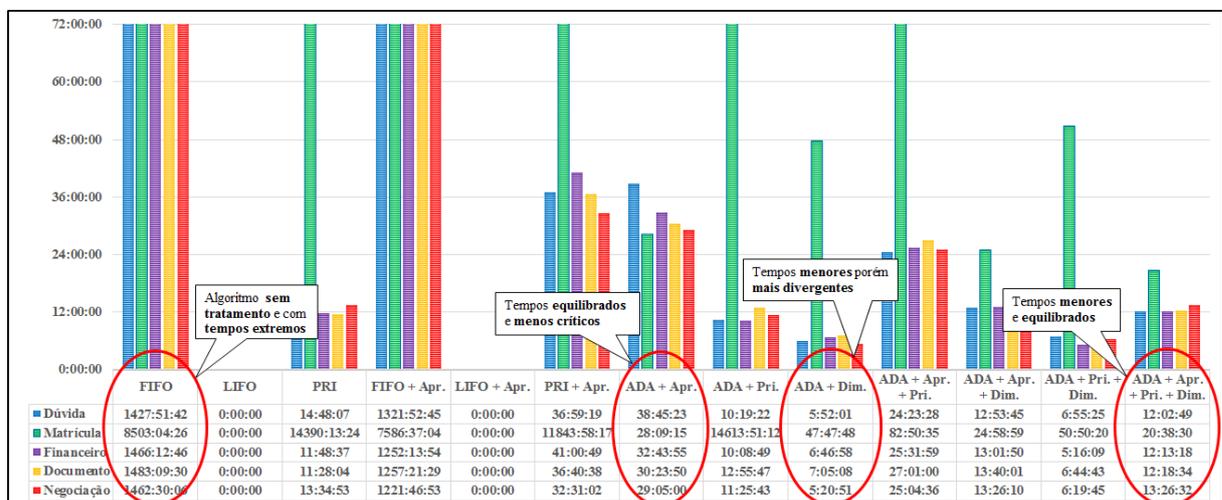
Esta situação ocorre, pois, o Agente de priorização, aplicado neste algoritmo, foi formulado a partir de um problema derivado da Teoria de Filas que não ocorre neste cenário. O problema em que o sistema prioriza muito algumas demandas de forma com que algum tipo de serviço não seja atendido é inexistente aqui, pois não há uma priorização definida no ambiente.

A disciplina ADA possui priorização conforme o ambiente, quando não possui o Agente de aprendizagem, ou conforme a experiência do canal quando este é utilizado. Como o único agente aplicado é o priorizador, e o ambiente não possui priorização definida (igual para todos), o sistema possui funcionamento idêntico à disciplina PRI. Ou seja, mesmo sendo aplicada a disciplina ADA com o Agente de priorização, o funcionamento é o mesmo da disciplina PRI da Teoria de Filas. Isto pode ser identificado também no resultado de todos os cenários anteriores (Figura 20, Figura 23 e Figura 26).

6.3.5 Caso de teste 5 – Variável

Os resultados das simulações que utilizaram este cenário foram bem diferentes dos anteriores, com novas situações: algoritmos que não puderam ser executados, algoritmos que atingiram o limite de demandas do sistema, algoritmos que tiveram resultados muito parecidos com o de situações menos adversas, dentre outros. Os algoritmos que utilizam a disciplina LIFO (combinações LIFO e LIFO + Apr.) não puderam ser simulados neste cenário, pois a situação era extrema a ponto de o ambiente não suportar a realização das simulações até o seu término.

Figura 30 – Variável: Tempo total de fila do histórico X Algoritmo X Tipo de atendimento

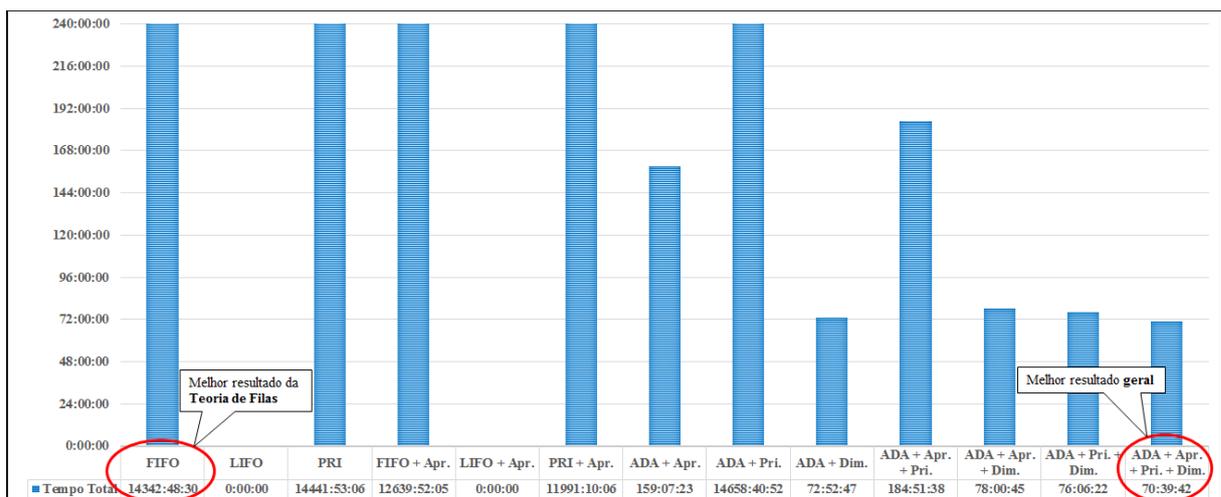


Fonte: do autor

O gráfico apresentado na Figura 30 mostra como a aplicação de cada elemento desenvolvido afeta o padrão de tempos de fila do sistema. Observa-se que o algoritmo FIFO teve tempos extremos de fila, sem qualquer possibilidade de ser um modelo válido para o ambiente. Já o algoritmo ADA + Apr. diminuiu os tempos de fila a níveis menos críticos, comparado aos modelos da Teoria de Filas, e estabilizou os tempos entre as cinco classes de serviços. O modelo ADA + Dim., por sua vez, também diminuiu os tempos de fila, mas manteve a discrepância entre o tipo de atendimento mais requisitado e os demais. Por fim, o algoritmo que utiliza todos os recursos desenvolvidos no estudo, ADA + Apr. + Pri. + Dim., conseguiu o melhor resultado geral, reduzindo todos os tempos de forma conjunta.

Através do gráfico na Figura 31 é possível ver o resultado final dos tempos de fila neste cenário.

Figura 31 – Variável: Tempo total de fila do histórico X Algoritmo



Fonte: Autor

Destaca-se como melhor resultado, assim como visto no gráfico anterior, o algoritmo ADA + Apr. + Pri. + Dim. Este, com tempo total de **70:39:42**, ao ser comparado ao modelo tradicional FIFO, que obteve tempo total de **14342:48:30**, apresentou uma melhora de **99%** no tempo de fila.

Ao comparar os resultados deste com o do cenário base, *Destacado*, que não possui comportamento variável, é verificado que o melhor resultado do caso base, ADA + Apr. + Dim., passa a não ser mais o melhor em ambiente oscilante. Porém ao adicionar o Agente de priorização nesta combinação, o resultado se torna o melhor.

Também pode ser observado que o algoritmo ADA + Pri. + Dim. teve o seu resultado quase inalterado, de 71:42:48 no caso *Destacado* para 76:06:22 no cenário *Variável*. Mesmo

que tais resultados não sejam os melhores de cada caso de teste, o algoritmo demonstra uma grande estabilidade de comportamento, independente da oscilação na quantidade de demandas do ambiente, desde que seja mantida a proporção entre as classes de serviços requisitadas.

Este caso de testes demonstra como os recursos desenvolvidos para a Política Adaptativa de Atendimento podem alterar o resultado de algoritmos quando o ambiente passa por mudanças comportamentais. Ou seja, tais elementos trazem ao algoritmo o comportamento adaptativo, objetivo do atual trabalho.

6.4 ANÁLISE DOS RESULTADOS

Através dos resultados obtidos na experimentação realizada com os cinco cenários distintos, é possível verificar que a aplicação de elementos de Sistemas Multiagente pode melhorar o desempenho de ambientes de atendimento na grande maioria dos casos. Também pode ser visto que, para cada situação, pode haver uma combinação diferente que oferece os melhores resultados, não sendo necessária sempre a utilização de todos os recursos oferecidos.

Tabela 13 – Melhores algoritmos em cada cenário e seus ganhos de desempenho

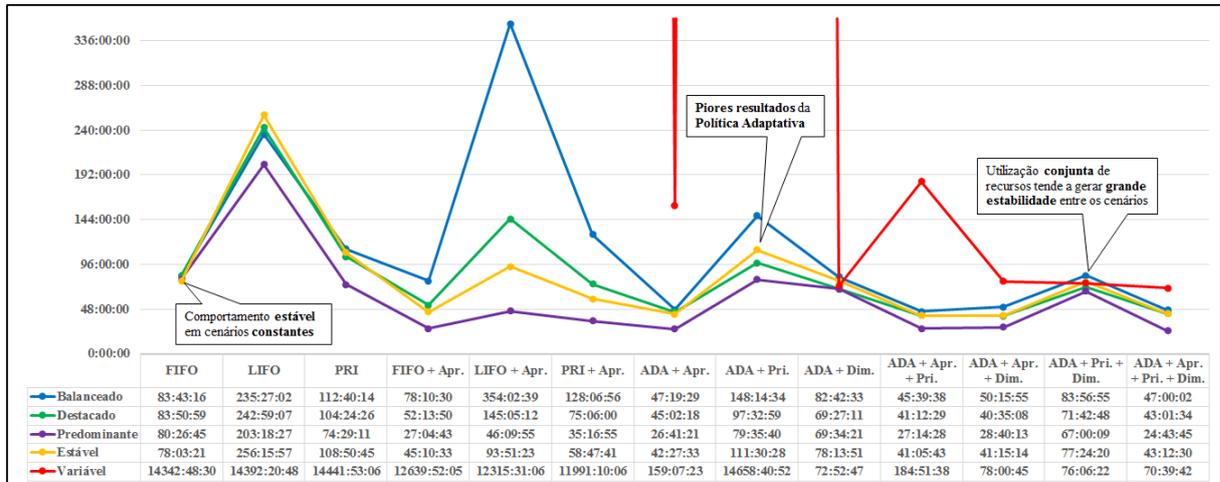
| Caso de teste | Melhor algoritmo da Teoria de Filas | Melhor algoritmo geral | Ganho |
|---------------|-------------------------------------|--------------------------|-------|
| Balanceado | FIFO | ADA + Apr. + Pri. | 46% |
| Destacado | FIFO | ADA + Apr. + Dim. | 53% |
| Predominante | PRI | ADA + Apr. + Pri. + Dim. | 66% |
| Estável | FIFO | ADA + Apr. + Pri. | 47% |
| Variável | FIFO | ADA + Apr. + Pri. + Dim. | 99% |

Fonte: do autor

A Tabela 13 exhibe os melhores algoritmos para cada caso de teste, tanto o melhor modelo geral quanto o melhor baseado na Teoria de Filas, e também o ganho no tempo de fila entre ambos. É possível ver que a melhora no desempenho, obtida através dos elementos da Política Adaptativa de Atendimento, tem tendência crescente conforme a instabilidade do ambiente. Ou seja, quanto mais instável for o ambiente, melhor o resultado provido pelos recursos desenvolvidos.

Na Figura 32 é possível observar a diferença que cada algoritmo induz em cada cenário simulado. Nela, são demonstrados os resultados finais de cada algoritmo aplicado em cada caso de teste proposto nos experimentos.

Figura 32 – Tempo total de fila do histórico X Casos de testes X Algoritmos



Fonte: do autor

Inicialmente é destacado que o modelo mais tradicional da Teoria de Filas, FIFO, possui muita estabilidade quando aplicado a casos constantes, porém não suporta o cenário variável. Após variar muito nos próximos algoritmos que utilizam a Teoria de Filas, é obtido o pior resultado da Política Adaptativa de Atendimento com ADA + Pri., conforme já explicado. Posteriormente, a partir do momento que começam a ser aplicados os recursos novos em conjunto, como no caso do algoritmo ADA + Pri. + Dim, os resultados são estabilizados, inclusive do cenário variável, abaixo daqueles obtidos no modelo mais tradicional e da maioria dos demais.

A seguir são analisados individualmente os elementos desenvolvidos neste trabalho, a disciplina ADA e os três agentes de Sistemas Multiagente. São destacados os seus pontos fortes e pontos fracos de atuação, assim como a sinergia na sua utilização em conjunto.

6.4.1 Disciplina

A disciplina adaptativa desenvolvida neste trabalho se mostrou eficaz, principalmente ao atuar em conjunto com o Agente de aprendizagem. Ela também evidencia sinergia com o Agente de priorização, pois o mesmo utiliza as prioridades dos atendentes como base de suas escolhas, e são estas prioridades que são manipuladas pela disciplina.

Entretanto, quando o Agente de aprendizagem não é utilizado, a disciplina não demonstra um ganho significativo de desempenho. Nestes casos, ela possui um comportamento muito parecido com a disciplina PRI da Teoria de Filas, somente disponibilizando prioridades

individualizadas por atendente. Este comportamento não foi abordado neste estudo, pois o intuito era de utilizar a disciplina em conjunto com os agentes desenvolvidos.

6.4.2 Agente de aprendizagem

Este agente foi o que gerou efeitos mais facilmente destacáveis nos resultados obtidos. A não ser pelo caso *Balanceado*, em que a demanda é toda igual, sempre que este era aplicado os dados possuíam diferenças significativas, tanto em tempos de atendimento como tempos de fila. A importância deste agente pode ser observada facilmente ao final do trabalho, pois, em todos os casos simulados, ele está presente nos algoritmos com melhor resultado. A sua atuação faz com que a disciplina desenvolvida seja mais utilizada, atuando de forma constante para identificar e encaminhar demandas para os melhores atendentes. Já o Agente de priorização reconhece com mais facilidade os casos para o qual foi proposto, assim corrigindo as situações mais rapidamente e também melhorando o seu desempenho.

A sinergia não foi tão evidente com o Agente de dimensionamento, principalmente pelo fato deste não realizar uma “escolha” de qual atendente desativar. Ou seja, para o agente dimensionador, tanto faz a prioridade de atendimento do canal sendo removido, sendo esta a principal característica trabalhada pelo Agente de aprendizagem. Este funcionamento pode, em alguns casos, até implicar perda de desempenho, como visto no cenário *Predominante*.

Este agente também possui uma característica diferente dos outros: ele é baseado em algo que pode variar conforme o ambiente em que for aplicado. Neste estudo, utilizando a base de um sistema de atendimento, que utiliza pessoas como atendentes, foi escolhida uma característica pertencente a este grupo na composição da aprendizagem. Porém em outros ambientes (que atendentes são computadores, por exemplo), este modelo de funcionamento pode não servir, pois tais grupos de atores possuem definição de experiência e aprendizagem diferentes. Isto dá uma liberdade maior para o agente, ao mesmo tempo em que aumenta a sua complexidade.

6.4.3 Agente de priorização

Este agente teve o comportamento mais volátil dos três desenvolvidos. Isto foi devido, em grande parte, à sua atuação em momentos para o qual não foi realmente proposto. Conforme sua proposta de funcionamento, este deveria ser utilizado somente quando há problemas de “esquecimento” de tipos de atendimento na fila. Porém quando aplicado em casos que isto não

ocorre, como por exemplo no algoritmo ADA + Pri. (em todos os cenários), ele acabava criando um ambiente com e sem prioridades, um funcionamento misto não previsto. Isto ocasionava em um tempo de fila menor no tipo priorizado pelo canal, porém as outras classes sofriam aumentos em seus tempos de forma semelhante, sendo assim o efeito negativo maior do que o positivo.

Todavia, ao ser combinado com outros elementos propostos neste estudo, em ambientes que poderiam ocorrer a causa de seu desenvolvimento, este agente conseguiu contribuir na melhoria do desempenho. O cenário *Estável* demonstra muito bem este contexto, pois os algoritmos que utilizam os outros agentes exclusivamente tiveram desempenho pior do que quando adicionado o Agente de priorização neles.

O seu propósito, a partir do problema em que foi baseado, sugere que somente seja utilizado quando há priorização de demandas. Isto foi possível em quase todos os cenários que utilizam a nova disciplina ADA, porém em nenhum cenário que utiliza as disciplinas da Teoria de Filas, sendo este foi o principal motivo para a não inclusão de tais casos de testes.

6.4.4 Agente de dimensionamento

O agente que possivelmente possui o modelo de funcionamento mais conhecido por todos é o Agente de dimensionamento. Isto porque o consenso geral sobre um sistema com filas é de aumentar a capacidade quando a quantidade de demandas se torna muito alta. No atual trabalho foi utilizado justamente esta lógica para o agente, sem outros controles que um ambiente real deve possuir como disponibilidade, custos, entre outros.

A sua aplicação pode ser feita com qualquer combinação de disciplinas e algoritmos, pois o seu efeito é o simples fato de incrementar o número de canais disponíveis. Com isso, aumenta a capacidade do sistema e finalmente a frequência de atendimento das requisições. Porém a sua lógica para retornar o ambiente ao formato original, desativando os atendentes extras, pode ocasionar problemas em algumas situações.

Como o agente não prevê uma “escolha” de qual canal remover, sempre são removidos somente aqueles que foram adicionados, de modo similar a uma pilha. Através disto, há a possibilidade de gerar instabilidade nas prioridades do sistema quando houver outros agentes atuando em conjunto. Um exemplo disto é quando um dos atendentes removidos for o melhor em uma classe de serviços, e esta tiver baixa prioridade em outros atendentes. Isto deixa o tipo de serviço sem um atendente que o priorize, o que, se não tratado em tempo, pode gerar outros efeitos negativos e assim sucessivamente.

7 CONCLUSÃO

Neste trabalho, com o objetivo de melhorar o desempenho de ambientes de atendimento em situações que a Teoria de Filas não abrange, foram desenvolvidos: uma disciplina de filas adaptativa, baseada nas já existentes FIFO e PRI da Teoria de Filas; e agentes provenientes de Sistemas Multiagente, passíveis de utilização avulsa ou em conjunto. Estes elementos, associados à modelagem de Teoria de Filas, permitem adaptações mais genéricas, conforme a necessidade da situação envolvida.

Inicialmente, foi observado que os modelos tradicionais da Teoria de Filas conseguem obter desempenhos razoáveis em ambientes que possuam certo grau de estabilidade em sua aleatoriedade. Entretanto, ao adicionar comportamentos variáveis, simulando um ambiente real e complexo, tais modelos passam a não atender as demandas da melhor forma possível, ou mesmo nem atender. Tais situações acabam gerando quedas de desempenho no tratamento de demandas, piorando o funcionamento do sistema.

Com a aplicação da disciplina e dos agentes desenvolvidos, obteve-se ganhos de desempenho em todos os cenários testados. Tais benefícios normalmente foram obtidos com qualquer combinação de recursos aplicados, porém os melhores algoritmos foram aqueles que utilizavam vários, se não todos, ao mesmo tempo. Estes resultados comprovaram que a utilização de vários agentes simultaneamente, com funcionamento em sinergia, pode gerar melhores resultados do que a aplicação individualizada de cada um.

Após analisar todos os resultados, pôde ser avaliado o ganho produzido por cada complemento desenvolvido. Concluiu-se que a disciplina ADA é vital para o comportamento dinâmico do algoritmo, porém se não for utilizado nenhum artifício para priorizar demandas (deixando todas demandas com a mesma prioridade), ela não traz benefício algum além do que a disciplina PRI já possui.

Já o Agente de aprendizagem foi o responsável pelos maiores ganhos de efetividade, visto que o mesmo foi capaz de identificar os melhores recursos de atendimento para cada demanda, encaminhando de forma inteligente as requisições. Esta realização é devida, em grande parte, à proposta diferenciada que este agente possui em relação aos demais. O seu comportamento não foi aplicável somente em determinadas situações, corrigindo situações críticas do ambiente, mas sim em qualquer momento de atendimento, melhorando até mesmo o funcionamento em casos de baixa demanda.

Com o Agente de priorização, foi observado que seus ganhos requerem situações específicas, e quando aplicado em ambientes com falta destas ocasiões, pode até gerar perdas de desempenho. Contudo, tais casos foram exceções nos casos de testes, sendo combinações de algoritmos usadas somente para comparações de resultados. Quando o agente foi aplicado em conjunto com os demais, trouxe melhora no rendimento do algoritmo, proporcionalmente à instabilidade do cenário. Pode ser concluído, então, que este agente possui certas restrições para utilização, sendo eficaz somente em alguns cenários, e que outros recursos podem ser aplicados em conjunto para melhorar sua eficácia.

A utilização do Agente de dimensionamento trouxe uma melhora significativa nos resultados, pois o mesmo aumenta a capacidade do ambiente na forma de mais canais de atendimento. Em cenários constantes o mesmo não surtiu muito efeito, pois raramente pôde ser acionado, mas em casos extremos de demanda a sua eficácia foi muito grande. Este agente também possibilitou uma melhora de desempenho dos outros agentes, pois quanto mais canais o ambiente possui, maior a utilização e aplicação destes.

É concluído a partir deste estudo que, através do uso de agentes para complementar os modelos da Teoria de Filas, podem ser obtidos melhores resultados no desempenho de filas em sistemas de atendimento. Esta melhoria dos resultados é proporcional à instabilidade do ambiente, situação não contemplada pela Teoria de Filas, porém corrigida com a aplicação de agentes específicos. Assim, o estudo evidencia que os agentes são uma opção interessante para casos de sistemas de filas que possuem algum tipo de volatilidade em seu processo.

O modelo híbrido se mostrou muito eficaz em ambientes teóricos, entretanto os seus benefícios podem não ser iguais em todos os sistemas reais sujeitos à sua aplicação. Isto é devido ao fato de que cenários reais possuem uma complexidade muito maior que os teóricos, alterando a dinâmica de situações abordadas e com isso ocasionando resultados diferenciados, menos significantes. Através desta possível mudança de comportamento e da complexidade inserida no ambiente com o uso de tal modelo, a viabilidade de sua aplicação requer um estudo detalhado no ambiente proposto.

7.1 TRABALHOS FUTUROS

Para trabalhos futuros, podem ser avaliadas outras situações passíveis de melhorias ao utilizar os métodos tradicionais de gerenciamento de filas. Podem ser agregados, aos já existentes neste trabalho, agentes que contribuam para a melhoria de desempenho das filas do

sistema e, assim, formar uma política mais avançada, que contempla mais casos e possui melhor rendimento.

Também podem ser abordadas questões deixadas de fora neste estudo, referentes à aplicação de novos agentes desenvolvidos. Neste contexto se encontram a disponibilidade de recursos e os custos da aplicação destes, entre outros.

Outra linha de estudos seria a complementação do ambiente desenvolvido no presente trabalho, permitindo outras abordagens em novos estudos. Existem questões como a alteração de atendentes (mantendo a mesma quantidade dos mesmos) através de escalas, ou então a compatibilização com a métrica temporal utilizada em ambientes reais, através do controle de dias, meses, expedientes de trabalho, folgas, entre outros.

REFERÊNCIAS BIBLIOGRÁFICAS

ABENSUR, Eder Oliveira; BRUSTEIN, Israel; FISCHMANN, Adalberto A.; HO, Linda Lee. Tendências para o Auto-atendimento Bancário Brasileiro: Um Enfoque Estratégico Baseado na Teoria da Filas. **Revista de Administração Mackenzie**, São Paulo, SP, ano 4, n.2, p. 39-59, 2003.

ARAÚJO, Marcus Augusto Vasconcelos; ARAÚJO, Francisco José Costa; ADISSI, Paulo José. Distribuição da demanda telefônica de um call center através da criação e priorização de filas inteligentes. **Revista Produção Online**, Florianópolis, v. 3, n. 4, 2003. Disponível em: <<http://producaoonline.org.br/rpo/article/view/570>>. Acesso em: 22 ago. 2015.

BANKS, Jerry. *Introduction to Simulation. In: Winter Simulation Conference*, 2000, Orlando. **Proceedings....** Orlando: IEEE, 2000.

BOUZADA, Marco Aurélio Carino. Dimensionamento de um Call Center: Simulação ou Teoria de Filas? In: Simpósio de Administração da Produção, Logística e Operações Internacionais, 12., 2009, São Paulo. **Anais eletrônicos....** São Paulo: EAESP-FGV, 2009. Disponível em: <http://www.simpoi.fgvsp.br/arquivo/2009/artigos/E2009_T00060_PCN54008.pdf>. Acesso em: 25 ago. 2015.

BRUNS, Rafael de; SONCIM, Sérgio Pacífico; SINAY, Maria Cristina Fogliatti de. Pesquisa Operacional: Uma aplicação da Teoria de Filas a um sistema de atendimento. In: Encontro Nacional de Engenharia de Produção, 21., 2001, Salvador. **Anais eletrônicos....** Rio de Janeiro: ABEPRO, 2001. Disponível em: <http://www.abepro.org.br/biblioteca/ENEGEP2001_TR60_0158.pdf>. Acesso em: 14 set. 2015.

CHWIF, Leonardo; MEDINA, Afonso Celso. **Modelagem E Simulação de Eventos Discretos: Teoria e Aplicações**. 3 ed. São Paulo: Bravarte, 2010. 320 p.

CHWIF, Leonardo. **Redução de modelos de simulação de eventos discretos na sua concepção: Uma abordagem causal**. 1999. 139 f. Tese (doutorado) – Departamento de Engenharia Mecânica, Escola Politécnica da Universidade de São Paulo, 1999.

DOILE, Luiz Fernando Pacheco. **Teoria de Filas – Analisando o fluxo de atendimento e o número de atendentes em um supermercado**. 2010. 49 f. Trabalho de Conclusão de Curso (Monografia) – Curso de Administração, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2010.

FERREIRA, Abel Augusto de Carvalho. **Metodologia para análise e projeto de Sistemas Multi-Agentes**. 2008. 62 f. Trabalho de Conclusão de Curso (Monografia) – Curso de Ciência da Computação, Universidade Federal de Santa Catarina, Florianópolis, SC, 2008.

FIGUEIREDO, Danielle Durski; ROCHA, Silvana Heidemann. Aplicação da teoria das filas na otimização do número de caixas: Um estudo de caso. **Iniciação Científica CESUMAR**, Maringá, v. 12, n. 2, p. 175-182, jul./dez. 2010.

GROSS, Donald; SHORTLE, John F.; THOMPSON, James M.; HARRIS, Carl M. *Fundamentals of Queueing Theory*. 4th. ed. Nova Jersey: John Wiley & Sons Inc, 2008. 500 p.

HARREL, Charles; GHOSH, Biman K; BOWDEN, Royce. *Simulation using Promodel*. Nova York: McGraw-Hill, 2000. 603 p.

JUCHEM, Murilo; BASTOS, Ricardo Melo. **Arquitetura de Agentes**. 2001. 32 f. Relatório técnico – Curso de Ciência da Computação, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2001. Disponível em: <<http://www3.pucrs.br/pucrs/files/uni/poa/facin/pos/relatoriostec/tr013.pdf>>. Acesso em: 17 set. 2015.

KOOLE, Ger; MANDELBAUM, Avishai. *Queueing Models of Call Centers - An Introduction*. *Annals of Operations Research*, [S.l.], v. 113, n.1, p. 41-59, 2001.

LEAL, Fabiano. **Um diagnóstico do processo de atendimento a clientes em uma agência bancária através de mapeamento do processo e simulação computacional**. 2003. 224 f. Trabalho de conclusão de curso (Dissertação) – Curso de Pós-Graduação em Engenharia de Produção, Universidade Federal de Itajubá, Itajubá, MG, 2003.

LIMA, Fabiana Xavier Correia; BELDERRAIN, Mischel Carmen Neyra. Propostas de melhorias de atendimento num Pronto Socorro utilizando Teoria de Filas e Teoria de Restrições. In: Encontro de Iniciação Científica e Pós Graduação do ITA, 13., 2007, São José dos Campos. **Anais eletrônicos...** São José dos Campos: ITA, 2007. Disponível em: <<http://www.bibl.ita.br/xiiiencita/MEC15.pdf>>. Acesso em: 07 ago. 2015.

LIMA, Tiago França Melo de; FARIA, Sérgio Donizete; SOARES FILHO, Britaldo Silveira; CARNEIRO, Tiago Garcia de Senna. Modelagem de sistemas baseada em agentes: Alguns conceitos e ferramentas. In: Simpósio Brasileiro de Sensoriamento Remoto, 14., 2008, Natal. **Anais eletrônicos...** São José dos Campos: INPE, 2009. Disponível em: <<http://marte.dpi.inpe.br/col/dpi.inpe.br/sbsr@80/2008/11.17.15.46/doc/5279-5286.pdf>>. Acesso em: 07 out. 2015.

LOPES, Claudio Lúcio do Val; FERREIRA, Ricardo Poley Martins; SILVA, Gustavo Rodrigues Lacerda. Modelo de Simulação Multiagente de uma Central de Teletendimento. In: Encontro Nacional de Inteligência Artificial, 6., 2007, Rio de Janeiro. **Anais eletrônicos...** Rio de Janeiro: IME, 2007. Disponível em: <http://www.cos.ufrj.br/~ines/enia07_html/pdf/27989.pdf>. Acesso em: 03 nov. 2015.

LOPES, Claudio Lúcio do Val; SILVA, Gustavo Rodrigues Lacerda; FERREIRA, Ricardo Poley Martins. Um estudo sobre a influência do desempenho e da escala de trabalho dos atendentes nas Centrais de Teletendimento utilizando um modelo de Simulação Multiagentes. In: Simpósio Brasileiro de Pesquisa Operacional, 40., 2008, João Pessoa. **Anais eletrônicos...** João Pessoa: SBPO, 2008. Disponível em: <<http://www.din.uem.br/sbpo/sbpo2008/pdf/arq0290.pdf>>. Acesso em: 07 nov. 2015.

MACAL, Charles M; NORTH, Michael J. *Tutorial on Agent-based modeling and Simulation*. In: *Winter Simulation Conference*, 37., 2005, Orlando. **Proceedings...** Orlando: IEEE, 2005. Disponível em: <<http://www.informs-sim.org/wsc05papers/002.pdf>>. Acesso em: 26 out. 2015.

MANDELBAUM, Avishai; SAKOV, Anat; ZELTYN, Sergey. *Empirical Analysis of a Call Center*. Tel Aviv: Techion, 2001. 73 p.

MARINS, Fernando Augusto Silva. **Introdução à Pesquisa Operacional**. São Paulo: Cultura Acadêmica, 2011. 176 p.

MEDHI, Jyotiprasad. *Stochastic Models in Queueing Theory*. 2nd. ed. San Diego: Academic Press, 2002. 450 p.

OLIVEIRA, Natacha Maria de Carvalho; FAVARETTO, Fabio. Análise de Filas de um Sistema de Serviços Utilizando Simulação e Eventos Discretos. In: Encontro Nacional de Engenharia de Produção, 33., 2013, Salvador. **Anais eletrônicos...** Rio de Janeiro: ABEPRO, 2013. Disponível em: <http://www.abepro.org.br/biblioteca/enegep2013_TN_STO_177_014_23117.pdf>. Acesso em: 10 out. 2015.

PILARES, Nanci Capel. **Atendimento ao cliente: O recurso esquecido**. 1. ed. São Paulo: Nobel, 1989. 87 p.

PRAIA, Carlos Ribeiro; GOMES, Carlos Francisco Simões. **Modelo de simulação para o dimensionamento de profissionais de um centro de compra**. Niterói, 2015. Disponível em: <<http://cidsid.org.br/sbpo2015/wp-content/uploads/2015/08/142335.pdf>>. Acesso em: 02 set. 2015.

SILVA, Rosaura Espírito Santo da. **Um Estudo sobre Modelagem Conceitual Baseada em Agentes**. 2006. 93 f. Trabalho individual – Curso de Mestrado em Ciência da Computação, Universidade Católica de Pelotas, RS, 2006.

SLACK, Nigel; CHAMBERS, Stuart; JOHNSTON, Robert. **Administração da produção**. 2. ed. São Paulo: Atlas, 2002. 748 p.

TISUE, Seth; WILENSKY, Uri. *NetLogo: A Simple Environment for Modeling Complexity*. In: *International Conference on Complex Systems*, 2004, Boston. **Proceedings....** Boston: NECSI, 2004.

VIOTTO FILHO, Irineu A. Tuim; PONCE, Rosiane de Fátima; ALMEIDA, Sandro Henrique Vieira de. As compreensões do humano para Skinner, Piaget, Vygotski e Wallon: pequena introdução às teorias e suas implicações na escola. **Psicologia da Educação**, São Paulo, SP, n. 29, p. 27-55, dez. 2009. Disponível em: <<http://pepsic.bvsalud.org/pdf/psie/n29/n29a03.pdf>>. Acesso em: 24 abr. 2016.

WILENSKY, Uri; RAND, William. **An Introduction to Agent-Based Modeling: Modeling Natural, Social, and Engineered Complex Systems with NetLogo**. Londres: MIT Press, 2015. 504 p.

WILENSKY, Uri. 1999. NetLogo. <http://ccl.northwestern.edu/netlogo/>. *Center for Connected Learning and Computer-Based Modeling*, Northwestern University. Evanston, IL.

WILENSKY, Uri. 2004. NetLogo Scatter model. <http://ccl.northwestern.edu/netlogo/models/Scatter>. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

WOOLDRIDGE, Michael J. **An Introduction to Multiagent Systems**. Liverpool: John Wiley & Sons, LTD, 2002. 343 p.

WOOLDRIDGE, Michael J.; JENNINGS, Nicholas R. Agent Theories, Architectures, and Languages: A Survey. In: ECAI94 Workshop on Agent Theories Architectures and Languages, 1994, Amsterdam. **Proceedings...** Berlin: Springer Verlag, 1994. 32 p.

XAVIER, Amanda Fernandes; GOMES, José Henrique de Freitas; DELALIBERA, Pedro Henrique Athanasio; PAIVA, Cibele Nogueira; PINHO, Alexandre Ferreira. A simulação a eventos discretos como ferramenta na tomada de decisão na implementação de uma linha de montagem em uma indústria do setor automobilístico. In: Encontro Nacional de Engenharia de Produção, 29., 2009, Salvador. **Anais eletrônicos...** Rio de Janeiro: ABEPRO, 2009. Disponível em: http://www.abepro.org.br/biblioteca/enegep2009_TN_STO_096_652_13413.pdf. Acesso em: 28 abr. 2016.