

Usando IoT e Conceitos de Elasticidade em Cloud para Análise de Recursos Humanos em Hospitais Inteligentes

Gabriel Souto Fischer¹, Cristiano André da Costa¹, Rodrigo da Rosa Righi¹

¹Software Innovation Lab - PPG em Computação Aplicada - Unisinos - RS, Brasil

`gabriel.souto.fischer@gmail.com, {cac, rrrighi}@unisinos.br`

***Resumo.** A alocação de profissionais de saúde em ambientes hospitalares muitas vezes não é capaz de se adaptar à demanda de pacientes. Há situações onde existem salas com pouco uso e profissionais ociosos, e salas com muito uso e menos profissionais do que o necessário. Neste contexto, nós propomos uma solução baseada no conceito de elasticidade em Cloud Computing, onde é possível adicionar, migrar ou remover recursos de acordo com as necessidades do serviço. Dessa forma, nós introduzimos o modelo Helastic, baseado em Internet das Coisas, capaz de monitorar o uso das salas pelos pacientes de um hospital e adaptar a capacidade de atendimento dessas salas. Usando uma abordagem preditiva, o Helastic é capaz de antecipar quando uma sala vai ter demanda acima da capacidade e propõe ações para movimentação dos recursos humanos para melhor atender a demanda. Utilizando dados de uma policlínica brasileira, constatou-se a diminuição do tempo de espera em até 96.73%.*

1. Introdução

Internet das Coisas (IoT) é um conceito onde objetos físicos (*coisas*) estão conectados através de uma estrutura de rede e fazem parte das atividades da internet [Singh and Kapoor 2017]. Um cenário relevante para aplicação da IoT é na área da Saúde [Fischer et al. 2020]. Pacientes podem ser supervisionados ininterruptamente por sistemas IoT, permitindo que situações de risco possam ser detectadas e tratadas apropriadamente [Sarhan 2018]. Nesse cenário, a alocação estática dos profissionais da saúde pode ser considerada ineficiente, pois podem surgir situações em que um determinado enfermeiro ou médico esteja ocioso em um setor de atendimento com pouca demanda, enquanto que em outro setor exista um grande volume de pacientes, sobrecarregando os profissionais que o atendem. Nesse contexto, se cada sala possui uma lista de especialidades requeridas e cada profissional possui uma lista com todas as suas especialidades, os atendentes ociosos de uma sala poderiam ser transferidos para a sala com maior necessidade, desde que tenham a especialidade requisitada pela sala de destino. De fato, a alocação de atendentes deveria sempre se *adaptar* às condições dos ambientes de saúde.

Portanto, é necessário encontrar estratégias eficazes para adaptar recursos humanos em tempo real. A elasticidade em computação em nuvem é uma das principais estratégias para adaptar recursos sob demanda [Al-Dhuraibi et al. 2018]. Uma abordagem de elasticidade proativa emprega técnicas de predição para antecipar o comportamento do sistema (sua carga) e, assim, decidir as ações de adaptação de recursos [Al-Dhuraibi et al. 2018]. Essa abordagem permite que a aplicação esteja pronta para lidar com a demanda crescente quando ela ocorrer. Além disso, segundo [Rostirolla et al. 2018], o conceito de elasticidade pode ser estendido a outras áreas além

da computação. Dessa forma, com essas técnicas é possível antecipar a demanda de pacientes e preparar adequadamente a alocação da equipe médica.

Baseado no conceito de elasticidade da computação em nuvem, nós apresentamos um modelo para *Alocação elástica de recursos humanos em ambientes de saúde* (Helastic) como alternativa à alocação estática tradicional da equipe médica. O Helastic funciona ajustando a alocação da equipe médica de hospitais inteligentes com base na abordagem proativa de elasticidade. Em particular, o Helastic usa sensores IoT para acompanhar a demanda de pacientes, modelada como uma série temporal e usada para estimar a demanda futura por atendimento. Essas estimativas permitem que o modelo seja capaz de prever quando a demanda de pacientes pode ser maior que capacidade de atendimento dos funcionários disponíveis. Usando essas informações, o modelo propõe uma alocação eficiente da equipe médica, movendo os profissionais e também alocando novos atendentes para as áreas com maior demanda. A ideia é oferecer sempre um tempo de espera razoável para os pacientes, independentemente da carga de trabalho (número deles no hospital). Para este fim, o Helastic propõe um algoritmo para alocação e migração de recursos humanos com o uso de sensores IoT e técnicas de predição computacional.

2. Trabalhos Relacionados

Algumas abordagens focaram na otimização do fluxo de pacientes para melhor alocar recursos. [Capocci et al. 2017] usaram técnicas de simulação discreta de eventos para propor algumas mudanças para balancear os níveis de ocupação dos profissionais de saúde e alcançar um tempo menor de espera pelos pacientes. [Vieira and Hollmén 2016] investigaram maneiras de minimizar gargalos no fluxo de pacientes devido a consultas, visitas, uso de recursos, etc. Na mesma linha de pensamento, [Graham et al. 2018] tiveram como objetivo prever a chegada de pacientes na emergência de um hospital para preparar adequadamente a alocação da equipe médica. [Weissman et al. 2020] tiveram como objetivo prever quando a demanda será superior a capacidade de atendimento de casos de Covid-19 na Filadélfia (EUA), para fornecer dados para preparar adequadamente os recursos das unidades hospitalares. Em ambos os trabalhos [Vieira and Hollmén 2016, Graham et al. 2018, Weissman et al. 2020], o objetivo era exclusivamente identificar padrões de dados específicos, em vez de propor contra-medidas para melhorar a alocação de recursos em saúde.

Na tentativa de aumentar a cobertura de saúde, alguns estudos propuseram modelos de previsão para entender a evolução da oferta de médicos e a demanda de pacientes para melhor adaptar a oferta à demanda. [Ishikawa et al. 2017] se concentraram em uma abordagem para treinar médicos suficientes para atender à demanda de pacientes em Hokkaido até 2030. [Liu et al. 2017] se focaram em um problema semelhante, mas de uma perspectiva global. Em contraste com o Helastic, a adaptação dos recursos do hospital ao fluxo de pacientes foi deixada de lado nesses trabalhos. Analisando os trabalhos citados, existem várias abordagens para analisar e estimar o uso de recursos humanos nos ambientes de saúde, para que o fluxo de pacientes possa ser melhorado ou para entender a evolução do problema da falta de profissionais.

As abordagens que propõem soluções, como treinamento de médicos [Ishikawa et al. 2017], ou a movimentação de enfermeiros entre duas salas [Capocci et al. 2017], são muito específicas e não podem ser usados ambientes dis-

tintos aos que foram propostos. Nesse contexto, podemos enumerar a principal lacuna na área da seguinte maneira: *Embora vários modelos sejam capazes de identificar a demanda atual e futura em um ambiente hospitalar, esses modelos não possuem soluções para ajudar a resolver o problema da falta de recursos hospitalares.* A insuficiência de recursos humanos nos ambientes de saúde não é nova e, com base nos trabalhos estudados, podemos ver que esse problema ainda existirá no futuro [Ishikawa et al. 2017, Liu et al. 2017]. Portanto, encontrar maneiras de otimizar o uso dos recursos existentes e ajustar a capacidade dos hospitais para atender à demanda dos pacientes são desafios que podem fazer toda a diferença. O uso de técnicas de previsão e da Internet das Coisas contribui para futuras soluções ou automação de processos na área da saúde. No entanto, o potencial das tecnologias está sendo subutilizado, pois é possível propor soluções como otimização e melhor uso dos recursos humanos existentes.

3. Modelo Helastic

Essa seção apresenta o Helastic, um modelo preditivo para alocação eficiente de recursos humanos com base no fluxo de pacientes nos ambientes hospitalares. O Helastic adapta o conceito de elasticidade da computação em nuvem ao contexto de recursos humanos, ajustando a capacidade de atendimento do hospital à demanda de pacientes, onde os profissionais são alocados, desalocados e realocados de acordo com as necessidades do hospital. O modelo agrupa informações de várias fontes: chegadas e necessidades de pacientes (usando sensores IoT e um conjunto de dados do hospital), movimentação de pacientes (usando sensores de IoT) e disponibilidade da equipe médica (a partir de um conjunto de dados). Usando essas informações, medimos a demanda de pacientes em tempo real e empregamos um algoritmo de previsão para antecipar a demanda futura.

3.1. Decisões de Projeto

Baseamos o modelo na premissa de que existem sensores espalhados pelo hospital, nas entradas e saídas de todos os ambientes e salas, capazes de identificar os pacientes que passam por eles. Cada paciente deve ter uma pulseira de identificação e deve carregá-la o tempo todo no ambiente interno do hospital. Assim, é possível identificar quando e onde um determinado paciente está assim que ele entrar no ambiente de saúde, juntamente com o tempo de permanência em cada sala enquanto é atendido. Dessa forma, utilizamos um *Real-Time Location System* (RTLS) [Boulos and Berry 2012] com precisão de localização a nível da sala. RTLS são sistemas para identificar e rastrear a localização de ativos e/ou pessoas em tempo real ou quase em tempo real [Boulos and Berry 2012]. O modelo deve ser transparente para os pacientes, no sentido de que eles não precisam relatar nenhuma condição relacionada a sua movimentação no ambiente hospitalar, sendo uma atividade realizada automaticamente pelo sistema. Com relação à estratégia de previsão de dados, o modelo usa uma abordagem baseada em estatística através da implementação do modelo ARIMA. Como podemos descrever o número de pacientes que aguardam atendimento ao longo do tempo como uma série temporal, optamos por usar a abordagem através do ARIMA por ser um modelo com um bom desempenho preditivo de séries temporais quando comparado com outras abordagens [Nisha and Sreekumar 2017].

3.2. Arquitetura

A arquitetura contempla três serviços: (i) um web service, responsável pela camada de visualização; (ii) um serviço de inferência, responsável pelo processamento de dados, tra-

tamento dos registros de movimentação, previsão da demanda de pacientes, e tomada de decisão da alocação de recursos humanos; e (iii) um serviço de banco de dados. Ainda, o modelo é subdividido em cinco módulos responsáveis pelo processamento da informação desde a sua captura pelos sensores até o resultado disponibilizado na interface web. A Figura 1 apresenta os módulos propostos, detalhando a arquitetura do modelo.

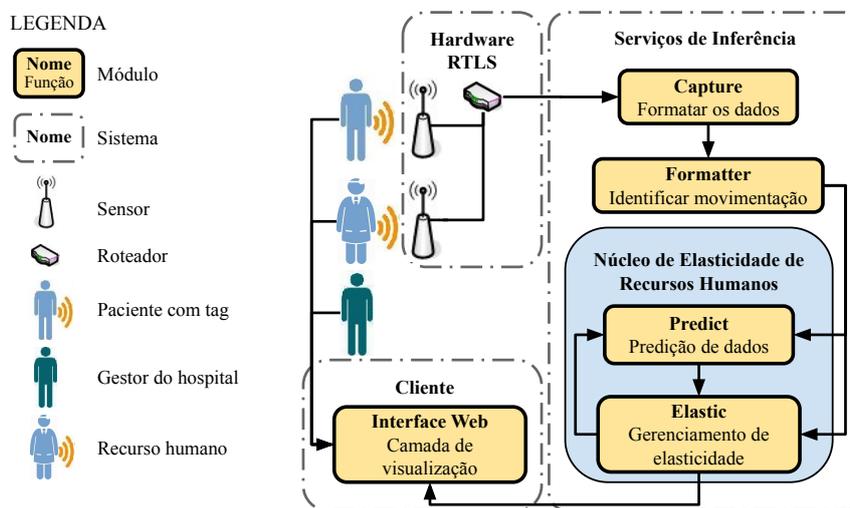


Figura 1. Detalhamento da arquitetura do modelo.

O módulo *Capture* recebe e processa os dados capturados pelos sensores espalhados pelo hospital e encaminha para o módulo *Formatter*, responsável por processar os dados, e identificar a movimentação das pessoas pelos ambientes do hospital. Após, o módulo *Predict*, identifica os padrões relacionados a chegadas de pacientes em cada ambiente do hospital, e os padrões relacionados ao tempo de espera por atendimento, e utiliza essa informação para prever a chegada futura de pacientes em cada sala do hospital. O módulo *Elastic* faz o gerenciamento de elasticidade do sistema. Ele verifica a alocação de recursos humanos em cada ambiente, e analisa as previsões feitas pelo módulo anterior. Este módulo gera uma alocação inteligente e automática dos recursos humanos para atender melhor à demanda dos pacientes. Por fim, o módulo de *Interface Web* exibe para os recursos humanos as notificações de elasticidade geradas anteriormente.

3.3. Elasticidade Preditiva de Recursos Humanos

O modelo emprega o termo *elasticidade* com um significado ligeiramente diferente do usado na computação em nuvem. Aqui, refere-se à capacidade do sistema de alocar/realocar/desalocar recursos humanos capazes de atender pacientes, a fim de se adaptar às diferentes demandas de pacientes em tempo real. Em particular, no contexto deste trabalho, elasticidade refere-se a: **(i) Alocação**, que é a capacidade do sistema de requisitar profissionais de saúde que não estão no hospital para atender a demanda de pacientes; **(ii) Realocação (ou migração)**, que refere-se a capacidade do sistema de migrar profissionais alocados em um ambiente hospitalar específico para outro ambiente em que são necessários mais profissionais; **(iii) Desalocação** que é a capacidade do sistema de liberar recursos humanos não mais necessários para atender a demanda de pacientes.

O modelo usa uma abordagem multinível, onde (i) a *nível de sala*, ele deve identificar o uso futuro de uma determinada sala e verificar se o número de atendentes é

suficiente para atender à demanda de pacientes e, (ii) a *nível de hospital*, ele deve verificar se há atendentes suficientes para atender à demanda de pacientes de todas as salas do ambiente hospitalar, com os atendentes se movendo entre as salas. Usamos essa estratégia em vários níveis, pois salas diferentes podem ter limites de tempo diferentes para o atendimento. Um exemplo desses dois níveis é apresentado na Figura 2. Ainda, a proposta adapta a estratégia de elasticidade proativa usando limites superiores e inferiores para o contexto de pessoas, com base no tempo de espera para atendimento em cada uma das filas de espera de um ambiente hospitalar. A Figura 3 ilustra o uso de limites, onde o modelo prevê que o limite superior será atingido (o que significa que devem ser adicionados recursos humanos para atender a necessidade) e logo após o modelo prevê que o limite inferior será atingido (ou seja, pessoas podem ser liberados para outros setores).

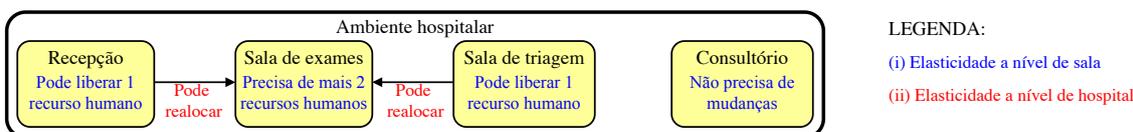


Figura 2. Elasticidade preditiva de recursos humanos.

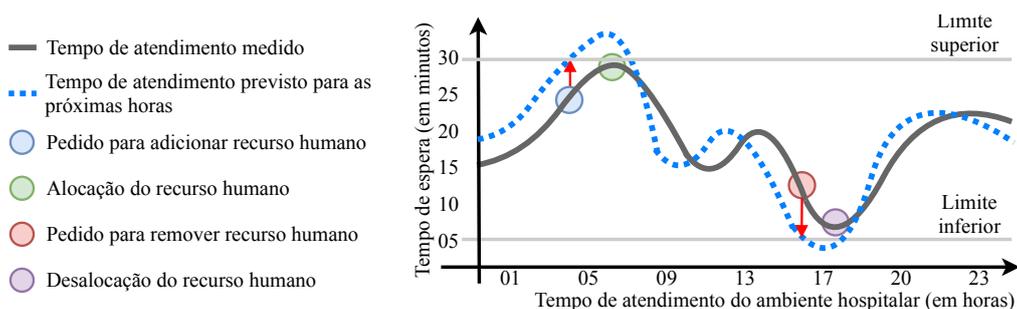


Figura 3. Elasticidade preditiva baseada nos tempos de espera.

A nível de sala, em cada ciclo de monitoramento, o modelo precisa prever a taxa de chegada de pacientes com base nas chegadas atuais e anteriores de cada sala. A previsão é feita usando o modelo ARIMA com base no tempo médio de atendimento com a alocação atual dos atendentes e no tempo estimado de espera para a fila de atendimento. Quando o modelo identifica que o tempo de espera se tornará maior ou menor que os valores limite definidos pelo gerente do hospital, o Helastic deve calcular o número de recursos de saúde necessários para atender à demanda dos pacientes por meio do *Speedup Elástico Preditivo de Recursos Humanos*. O Speedup Elástico Preditivo de Recursos Humanos usa uma abordagem preditiva para determinar a demanda futura de pacientes e definir dinamicamente o número adequado de atendentes, identificando o ganho futuro de tempo de atendimento em um ambiente hospitalar. O Helastic propõe alguns formalismos matemáticos para estimar o Speedup Elástico Preditivo de Recursos Humanos.

$CV(r, t_i, t_f)$ representa o vetor de atendimentos na sala r no intervalo de tempo entre os instantes t_i e t_f . O tamanho de qualquer vetor pode ser definido por $size(x)$. Usando essas duas funções, o tempo médio de atendimento em uma sala r entre os tempos t_i e t_f pode ser formulado através da Equação 1, onde $CDT(x[i])$ refere-se ao tempo de um atendimento $x[i]$ que já ocorreu na sala e $x[] = CV(r, t_i, t_f)$ é o vetor de atendimentos que ocorreram em uma sala r .

$$ACT(r, t_i, t_f) = \frac{1}{size(x)} \sum_{i=0}^{size(x)-1} CDT(x[i]) \quad (1)$$

Devido à elasticidade dos recursos humanos, em instantes de tempo diferentes, há um número diferente de atendentes alocados em cada uma das salas do hospital. A quantidade média de atendentes em uma sala r entre os tempos t_i e t_f é definida na Equação 2, onde $NA(r, t_n)$ representa a quantidade de atendentes alocados em uma sala r no instante de tempo n .

$$ANA(r, t_i, t_f) = \frac{1}{t_f - t_i} \sum_{t_n=t_i}^{t_f-1} NA(r, t_n) \quad (2)$$

A mesma ideia da função anterior é útil para a realidade de pacientes, pois em diferentes momentos de tempo, existem diferentes quantidades de pacientes aguardando por atendimento. Assim, a quantidade estimada de pacientes aguardando por atendimento em uma sala r entre os tempos t_i e t_f é definida pela Equação 3, onde $NWP(r, t_i)$ se refere a quantidade de pacientes aguardando por atendimento na sala r no instante de tempo t_i , e $NIP(r, t_n)$ se refere a quantidade de pacientes entrando na sala r a cada instante de tempo t_n .

$$ENP(r, t_i, t_f) = NWP(r, t_i) + \sum_{t_n=t_i+1}^{t_f-1} NIP(r, t_n) \quad (3)$$

Usando as equações propostas anteriormente, o Helastic calcula o tempo estimado de atendimento de todos os pacientes em espera e estima o tempo que um novo paciente que chega ao hospital precisa esperar para ser atendido. A função $ECT(r, t_i, t_f)$ é definida pela Equação 4, onde $ACT(r, t_i, t_f)$ é o tempo médio de atendimento para a sala r entre os tempos t_i e t_f , e $ENP(r, t_i, t_f)$ é o número estimado de pacientes que estão aguardando por atendimento na sala r entre os instantes de tempos t_i e t_f .

$$ECT(r, t_i, t_f) = ACT(r, t_i, t_f) \cdot ENP(r, t_i, t_f) \quad (4)$$

Conhecendo o $ECT(r, t_i, t_f)$, é possível analisar o tempo de espera de qualquer paciente aguardando em uma sala r entre os tempos t_i e t_f . Entretanto, esses valores representam uma sala com apenas um atendente mas, na maioria dos casos, terão mais do que um atendente alocado em cada sala, tornando necessário identificar o tempo médio de atendimento com diferentes quantidades de atendentes. Nesse contexto, o Helastic usa uma alocação *paralela* de recursos humanos, em contrapartida a alocação paralela de máquinas virtuais usada em sistemas de elasticidade. Assim, nós introduzimos a Equação 5 para *Speedup Elástico de Recursos Humanos*.

$$HRES(r, t_i, t_f) = \frac{ECT(r, t_i, t_f)}{ANA(r, t_i, t_f)} \quad (5)$$

$HRES(r, t_i, t_f)$ retorna o tempo estimado de atendimento em uma sala r entre os tempos t_i e t_f , considerando a alocação paralela de atendentes nesse período de tempo, através do uso da função $ANA(r, t_i, t_f)$. Assim, com o aumento da quantidade média de atendentes alocados, o tempo estimado de atendimento diminui, de forma inversamente proporcional. Antecipando o momento em que o limite superior será atingido, a movimentação de pessoas pode ocorrer com antecedência, minimizando ou evitando as sobrecargas do fluxo de pacientes no hospital. Nesse contexto, propomos a Equação 6 para *Speedup Elástico Preditivo de Recursos Humanos* da seguinte maneira:

$$PHRES(r, a, f_i, f_f) = \frac{ECT(r, f_i, f_f)'}{a}, \quad (6)$$

onde a é a quantidade de atendentes alocados entre os tempos futuros f_i e f_f , e $ECT(r, f_i, f_f)'$ é a predição do tempo futuro de atendimento dessa sala usando ARIMA. Nós podemos definir ECT' como:

$$ECT(r, f_i, f_f)' = ACT(r, f_i, f_f)' \cdot ENP(r, f_i, f_f)', \quad (7)$$

onde $ACT(r, f_i, f_f)'$ e $ENP(r, f_i, f_f)'$ são as predições do tempo médio de atendimento e da quantidade de pacientes futuros na sala r , respectivamente. Assim, para cada sala r sendo calculada, nós geramos uma série temporal para $ACT(r, t_i, t_f)$ que ocorreram no passado, e a usamos para prever $ACT(r, f_i, f_f)'$. Além disso, para cada sala também geramos uma série temporal para $NIP(r, t_i, t_f)$, e a usamos para prever a chegada de pacientes e encontrar $ENP(r, f_i, f_f)'$. Variando o atributo a na equação $PHRES$, com o aumento e a diminuição do número de profissionais de saúde presentes, o modelo pode identificar quantos atendentes seriam necessários para ajustar o tempo de espera de qualquer sala aos limites propostos, conforme definido pelo gestor do hospital.

A nível de hospital, o modelo precisa testar diferentes alocações para os atendentes, para garantir que todas as salas identificadas na etapa anterior (nível local) tenham atendentes suficientes, minimizando a superlotação. O Helastic considera a possibilidade de mover profissionais de saúde entre diferentes ambientes hospitalares, a fim de otimizar o tempo de atendimento médico. Para redistribuir esses atendentes entre diferentes salas, o modelo utiliza algumas estratégias conhecidas de outros contextos da computação e as adapta à Elasticidade Preditiva de Recursos Humanos. O Algoritmo 1 apresenta o pseudocódigo para elasticidade preditiva a nível de hospital. Cada sala possui uma especialidade específica, e o processo de realocação ou alocação de recursos humanos é realizado apenas entre profissionais que possuem a especialidade da sala de destino. As salas que necessitem de uma especialidade que nenhum outro profissional do hospital possui, apenas a alocação de novos recursos humanos é realizada. Isso é necessário porque, em uma sala de exames laboratoriais, é necessário um profissional de enfermagem acostumado a exames de sangue, por exemplo, e mesmo se tivermos técnicos de raio-x disponíveis para realocação, eles não poderão melhorar o atendimento de exames de coleta.

Para obter uma realocação balanceada de recursos humanos, desenvolvemos uma variação do algoritmo de *List Scheduling* [Wang and Sinnen 2018], usado originalmente para agendamento de processos. No algoritmo proposto, todas as salas do hospital estão em uma lista ordenada pelo número de atendentes disponíveis para realocação. Dessa forma, sempre que uma sala r precisar de mais atendentes, o modelo busca por atendentes disponíveis, com a especialidade da sala r , na primeira sala da lista. Se os atendentes estiverem disponíveis, eles serão realocados para a sala que falta, e a lista será ordenada novamente. Se forem necessários mais atendentes, o algoritmo verifica a primeira sala da lista novamente, e assim por diante, até que a sala obtenha todos os atendentes necessários.

A Figura 4 ilustra o processo de realocação, onde a sala 1 precisa de mais três atendentes e as salas 2 e 4 têm alguns atendentes disponíveis. Seguindo a lógica do algoritmo de *List Scheduling* adaptado, no primeiro *round*, a sala 2 é a primeira da lista, com três atendentes disponíveis, e fornece um atendente para a sala 1. No segundo *round*, mesmo que todas as salas tenham o mesmo número de atendentes disponíveis, a sala 2

profissionais de saúde que não estão no hospital, mas estão disponíveis para alocação. Ressaltamos que o hospital deve ter uma estratégia para definir quais recursos humanos estarão disponíveis para alocação. Por fim, se o modelo identificar que a demanda futura de atendimento de todas as salas do hospital é muito baixa e que a desalocação de atendentes de alguma sala não prejudica o todo, o modelo propõe a desalocação de atendentes alocados fora do horário normal de trabalho. Da mesma maneira que a realocação, a alocação e a desalocação também são protegidas pelo *cooldown period*.

4. Metodologia de Avaliação e Resultados encontrados

Nós avaliamos o desempenho do modelo através de simulações em um ambiente hospitalar simulado. Considerando a indisponibilidade de dados, o ambiente hospitalar foi definido com base em cargas de trabalho sintéticas. De acordo com [Rostirolla et al. 2018], as cargas de trabalho sintéticas podem ser consideradas uma forma representativa para avaliar a elasticidade em nuvens computacionais. O modelo foi implementado principalmente em Java, exceto o método ARIMA, que foi implementado em Python. Para simulação de filas hospitalares, foi utilizado um relógio com incrementos discretos de dez segundos. A cada avanço no relógio de simulação, o simulador verifica os pacientes que estão em atendimento e aqueles que devem deixar o atendimento. A cada ciclo de monitoramento, a chegada dos pacientes deve ser verificada.

Para o procedimento de simulação hospitalar, consideramos três cenários diferentes: **C1:** *Hospital sem o modelo (não-elástico)*; **C2:** *Hospital inteligente com realocação de recursos humanos apenas*; e **C3:** *Hospital inteligente com alocação, realocação e desalocação de recursos humanos*. Para realizar a simulação do ambiente hospitalar, usamos os dados coletados no estudo de [Capocci et al. 2017] realizado em uma policlínica de Guarulhos, em São Paulo. Dessa forma, todos os parâmetros usados na simulação, como distribuições de tempos de atendimento, movimentação de pacientes entre as salas e alocação de profissionais podem ser encontrados em [Capocci et al. 2017].

No Brasil, o modelo de trabalho adotado para ambientes hospitalares é o 12x36. De acordo com o Decreto-Lei nº 5.452 [Brasil 1943], sob esse regime de trabalho, um funcionário pode trabalhar por doze horas consecutivas e deve descansar por trinta e seis horas antes de um novo turno de trabalho. Nesse regime, quatro profissionais em turnos alternados são suficientes para garantir o atendimento de um mesmo posto de trabalho de forma ininterrupta. Assim, enquanto um recurso humano está em seu horário de trabalho, outros três funcionários que desempenham a mesma função estão no período de descanso. Dessa forma, definimos que qualquer funcionário em repouso só estará disponível para alocação sob as seguintes regras: **Regra 1:** O período mínimo de descanso para um recurso humano estar disponível para alocação é de onze horas, e um funcionário alocado deve ser desalocado no máximo 11 horas antes do seu próximo turno de trabalho normal em função da determinação que exige esse tempo entre jornadas no Decreto-Lei nº 5.452 [Brasil 1943]; **Regra 2:** Um funcionário alocado não pode trabalhar mais de doze horas, em função desse ser o maior período de trabalho possível permitido no Decreto-Lei nº 5.452 [Brasil 1943]; e **Regra 3:** Cada funcionário deve cumprir um dos períodos de descanso de 36 horas na mesma semana para cumprir uma determinação do Decreto-Lei nº 5.452 [Brasil 1943] que exige que todos os trabalhadores tenham no mínimo um período de descanso remunerado de 24 horas por semana.

Como o estudo de caso foi baseado em dados de um hospital brasileiro, estabelecemos limites adequados à essa realidade. Assim, com base no Projeto de Lei de 14 de junho de 2018 [Goiânia 2018] que propõe um tempo máximo de espera para atendimento em hospitais, clínicas e laboratórios de 30 minutos em dias regulares (de segunda a domingo), definimos a carga máxima do Helastic (ou seja, 100%) em 30 minutos e definimos o limite inferior em 9 minutos (30% do tempo máximo de espera). Para ações de elasticidade, definimos 10 minutos para o processo de realocação (movimentação de recursos humanos entre salas) e 60 minutos para o processo de alocação (para simular a movimentação de um novo recurso humano até o hospital).

Em relação à carga dos pacientes, modelamos duas cargas de trabalho: ascendente e onda. A ideia de usar diferentes comportamentos de carga para uma mesma aplicação é utilizada para observar como a carga de entrada pode afetar pontos de saturação, gargalos e na adição ou remoção de recursos. Esses dois comportamentos de carga de trabalho são baseados nos propostos por [Rostirolla et al. 2018]. Assim, a carga onda é a mais próxima da realidade hospitalar e a carga crescente foi escolhida para identificar o comportamento do modelo em uma situação de aumento da carga de pacientes, que poderia ser causada, por exemplo, por um surto viral. A fim de avaliarmos o modelo proposto, as seguintes métricas foram consideradas: (i) *Tempo máximo de espera*; e (ii) *Quantidade elástica de recursos humanos*. Para a métrica do tempo máximo de espera, esperamos uma adequação dessa métrica aos limites estabelecidos, bem como uma diminuição do tempo médio de espera. Com relação à quantidade elástica de recursos humanos, esperamos que o Helastic utilize os profissionais existentes no hospital de maneira otimizada. Assim, a alocação estática de C1, com onze funcionários trabalhando, pode ser comparada à alocação elástica do modelo, com o número de pessoas variando ao longo do dia.

5. Avaliação e Análise dos Resultados

Com base na metodologia de avaliação, realizamos seis simulações do ambiente hospitalar proposto, a fim de coletar resultados para análise. Para cada cenário proposto, entre C1, C2 e C3, foi realizada uma simulação para cada uma das cargas, Crescente e Onda. Para a primeira métrica, esperávamos uma diminuição no tempo de espera por atendimento. A Figura 5 mostra o tempo máximo de espera identificado para cada carga de trabalho nos cenários propostos durante o período de uma semana simulado. Percebemos uma redução significativa no tempo máximo de espera entre C1 e C2, e uma segunda diminuição ao comparar C2 e C3, independentemente da carga de trabalho utilizada. Após uma análise minuciosa, podemos identificar que no C3 para as salas de recepção, triagem, consultório médico e exames de coleta, em nenhum momento houve um tempo de espera superior a 30 minutos. Quanto às salas de medicação, raio-x e eletrocardiograma, houve alguns momentos em que esse limite foi excedido. Através dos dados coletados, identificamos uma redução significativa no tempo de espera com o uso do Helastic quando comparado ao hospital sem o uso da elasticidade. Graças aos procedimentos de realocação e alocação, o Helastic demonstrou diminuir o tempo de espera em 96,73% e 96,65% para as cargas onda e crescente, respectivamente, em comparação com o hospital não elástico.

Para a quantidade elástica de recursos humanos utilizados, esperamos um aumento no número de profissionais no hospital, bem como uma variação desse número ao longo do período de atendimento. A Figura 6 apresenta a quantidade elástica de recursos humanos utilizados no C3, o único cenário em que o número de funcionários pode variar.

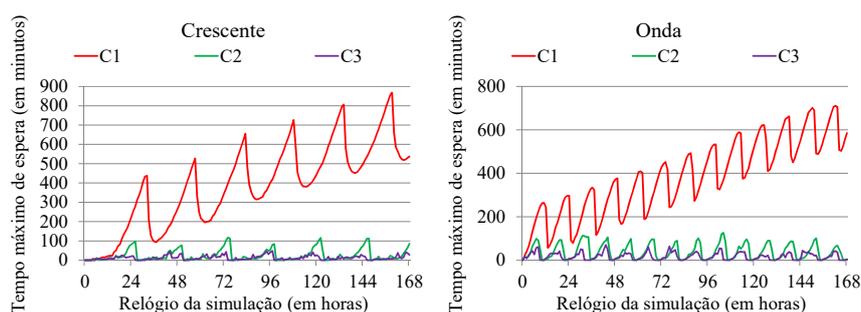


Figura 5. Tempo máximo de espera aferido nas simulações.

Podemos observar que a quantidade elástica de recursos humanos variou entre 11 a 14 por hora. Além disso, conforme exposto na Figura 6, sempre que o modelo realoca ou aloca profissionais para o atendimento, o tempo de espera diminui. Com base nas métricas estabelecidas, podemos observar que o Helastic foi capaz de melhorar o desempenho do ambiente hospitalar simulado em todas as cargas de trabalho utilizadas. Conforme proposto na metodologia de avaliação, esperávamos que o tempo máximo de espera apresentasse uma diminuição gradual entre os cenários C1, C2 e C3, o que ocorreu de fato, cumprindo o objetivo dessa métrica. Para a quantidade elástica de recursos humanos utilizados, era esperado um aumento entre os cenários C2 e C3, e o Helastic mais uma vez conseguiu cumprir a meta proposta. Assim, os resultados esperados foram alcançados.

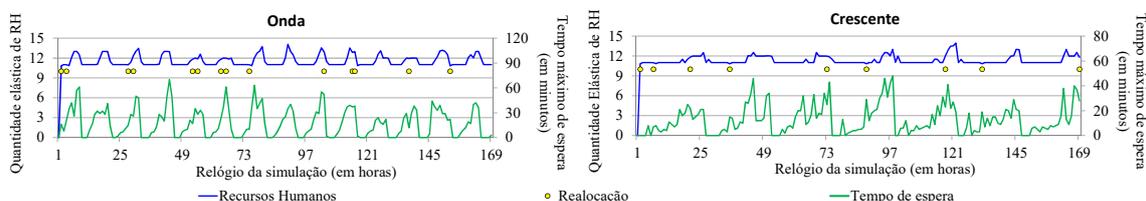


Figura 6. Quantidade elástica de recursos humanos utilizados em comparação com o tempo máximo de espera no hospital no Cenário C3.

6. Conclusão

Este artigo apresentou o modelo Helastic, que trata a alocação elástica de recursos humanos em ambientes de saúde. Diferentemente dos trabalhos relacionados, o modelo não apenas propõe o uso de técnicas de predição para antecipar eventuais problemas, mas também apresenta um modelo para alocar, realocar e desalocar profissionais em hospitais. Dessa forma, estendemos o conceito de elasticidade da computação em nuvem para o contexto de gerenciamento de recursos humanos, propondo novos formalismos matemáticos, algoritmos e definições para fornecer uma alocação dinâmica e elástica de profissionais em ambientes hospitalares. Esperamos que o modelo proposto neste trabalho possa ajudar a diminuir o tempo de espera dos pacientes em atendimento à saúde. No estudo de caso, o tempo de espera diminuiu 95,65% e 96,73% para as cargas crescente e onda, respectivamente. Embora apresentemos resultados animadores, existem algumas limitações que devem ser abordadas na implementação do modelo em um ambiente hospitalar real: (i) funcionários e pacientes devem carregar suas etiquetas de identificação durante todo o tempo no hospital inteligente; (ii) o modelo apenas gera notificações para os recursos humanos, mas a movimentação efetiva da equipe pelos ambientes hospitalares depende da decisão individual dos profissionais de seguir as recomendações; (iii) instalação prévia de sensores RTLS em todos os ambientes do hospital.

Referências

- Al-Dhuraibi, Y., Paraiso, F., Djarallah, N., and Merle, P. (2018). Elasticity in cloud computing: State of the art and research challenges. *IEEE Trans. Services Comput.*, 11(2):430–447.
- Boulos, M. N. K. and Berry, G. (2012). Real-time locating systems (RTLS) in healthcare: a condensed primer. *International Journal of Health Geographics*, 11(25):1–8.
- Brasil (1943). Decreto-Lei No 5.452, de 1 de maio de 1943. Aprova a Consolidação das Leis do Trabalho.
- Capocci, N. R., Nascimento, B. S., Lopes, F. B., Rodrigues, E. F., and Maiellaro, J. R. (2017). Simulation as a hospital management support tool. *Independent Journal of Management & Production*, 8(5):798–811.
- Fischer, G. S., Righi, R. d. R., Rodrigues, V. F., and Costa, C. A. d. (2020). Use of Internet of Things With Data Prediction on Healthcare Environments: a Survey. *International Journal of E-Health and Medical Communications*, 11(2):1–19.
- Goiânia (2018). Projeto de Lei no 2018/000159, de 11 de novembro de 2018. dispõe sobre o tempo máximo de espera para atendimento nos estabelecimentos que especifica.
- Graham, B., Bond, R., Quinn, M., and Mulvenna, M. (2018). Using data mining to predict hospital admissions from the emergency department. *IEEE Access*, 6:10458–10469.
- Ishikawa, T., Fujiwara, K., Ohba, H., Suzuki, T., and Ogasawara, K. (2017). Forecasting the regional distribution and sufficiency of physicians in japan with a coupled system dynamics—geographic information system model. *Hum. Resour. Health*, 15(1):64.
- Liu, J. X., Goryakin, Y., Maeda, A., Bruckner, T., and Scheffler, R. (2017). Global health workforce labor market projections for 2030. *Hum. Resour. Health*, 15(1):11.
- Nisha, K. G. and Sreekumar, K. (2017). A Review and Analysis of machine Learning and Statistical Approaches for Prediction. In *2017 Int. Conf. on Inventive Communication and Computational Technologies (ICICCT)*, pages 135–139. IEEE.
- Rostirolla, G., Righi, R., Barbosa, J., and Costa, C. (2018). Elcity: An elastic multilevel energy saving model for smart cities. *IEEE Trans. Sustain. Comput.*, 3(1):30–43.
- Sarhan, Q. I. (2018). Internet of things: a survey of challenges and issues. *International Journal of Internet of Things and Cyber-Assurance*, 1(1):40–75.
- Singh, K. J. and Kapoor, D. S. (2017). Create your own internet of things: A survey of iot platforms. *IEEE Consumer Electronics Magazine*, 6(2):57–68.
- Vieira, D. and Hollmén, J. (2016). Resource frequency prediction in healthcare: Machine learning approach. In *2016 IEEE 29th Int. Symp. on Computer-Based Medical Systems*, pages 88–93, Dublin, Ireland. IEEE.
- Wang, H. and Sinnen, O. (2018). List-scheduling versus cluster-scheduling. *IEEE Trans. Parallel Distrib. Syst.*, 29(8):1736–1749.
- Weissman et al. (2020). Locally Informed Simulation to Predict Hospital Capacity Needs During the COVID-19 Pandemic. *Annals of Internal Medicine*.