

Modelo Estocástico para Avaliação de Desempenho de Hospitais Inteligentes

Laécio Andrade Rodrigues[‡], Patricia Takako Endo[◇] e Francisco Airton Pereira da Silva[‡]

¹ [‡] Universidade Federal do Piauí (UFPI)
[◇] Universidade de Pernambuco (UPE)
Brasil

faps@ufpi.edu.br, patricia.endo@upe.br

Abstract. *This paper performs analytical modeling to evaluate the performance of an smart hospital architecture, avoiding investment in real equipment. The results show that this work has the potential to assist hospital system administrators in planning more optimized architectures according to their needs.*

Resumo. *Este trabalho adota a abordagem de modelagem analítica para avaliar o desempenho de um sistema hospitalar inteligente, evitando o investimento em equipamentos reais. Os resultados mostram que este trabalho tem o potencial de ajudar os administradores de sistemas hospitalares a planejar arquiteturas mais otimizadas de acordo com suas necessidades.*

1. Introdução

A Internet das Coisas (IoT) está inovando em todas as áreas da nossa sociedade, conectando milhões de aparelhos e facilitando o cotidiano das pessoas. O setor da saúde pode se beneficiar bastante com IoT. O uso de dispositivos e sensores inteligentes podem revolucionar a saúde das pessoas em qualquer ambiente, como em casa ou na rua. Em um hospital não é diferente, sinais vitais de pacientes captados por sensores podem ser transmitidos para Unidades de Tratamento Intensivo (UTI), taxa de ocupação dos leitos, índice de produtividade dos funcionários, entre outros.

Neste trabalho, decidimos aplicar uma abordagem de modelagem analítica para representar e avaliar o desempenho de hospitais inteligentes; e para isso usamos as redes de Petri Estocásticas (SPNs). Os SPNs são modelos analíticos que podem representar sistemas complexos com diversas características, incluindo paralelismo e simultaneidade.

Em resumo, as principais contribuições deste trabalho são: (i) Um modelo de SPN para representar um hospital inteligente, incluindo um terminal local para monitorar os pacientes, armazenamento de dados em dois locais diferentes: data center local e nuvem. (ii) Um modelo flexível com múltiplos parâmetros, permitindo a avaliação de um grande número de diferentes cenários. (iii) Estudos de caso usando o modelo de SPN proposto que servem como guias para administradores de sistemas planejarem suas infraestruturas hospitalares específicas.

2. Trabalhos Relacionados

Esta seção apresenta alguns trabalhos relacionados a IoT na área da saúde e a utilização de recursos. Na apresentação dos trabalhos foi destacado as correlações e diferenças com

o trabalho proposto. A Tabela 1 exibe um resumo comparativo dos trabalhos relacionados com o trabalho atual. Como pode ser observado na tabela, nenhum trabalho relacionado explora o cálculo do tempo médio de resposta (MRT). Quanto às métricas, nosso estudo foi o mais abrangente. Alguns trabalhos também não apresentam avaliação de desempenho que no contexto de hospitais inteligentes é de importância crucial.

Tabela 1. Trabalhos Relacionados

Trabalho Relacionado	Parâmetro de Utilização de Recursos	MRT	Redes de Petri Estocástica	Parâmetro do Número de Usuários	Avaliação de Desempenho
[Rahmani et al. 2015]	Não	Não	Não	Não	Sim
[Tata et al. 2017]	Não	Não	Não	Não	Não
[Oueida et al. 2018]	Sim	Não	Não	Sim	Sim
[Chen et al. 2018]	Sim	Não	Não	Sim	Sim
[Zhang et al. 2018]	Não	Não	Não	Não	Sim
[Rahmani et al. 2018]	Não	Não	Não	Não	Sim
[Akmandor and Jha 2018]	Não	Não	Não	Não	Não
[Sigwele et al. 2018]	Não	Não	Não	Não	Não
[Greco et al. 2019]	Sim	Não	Não	Não	Sim
Este trabalho	Sim	Sim	Sim	Sim	Sim

3. Arquitetura e Modelo

A Figura 1 apresenta uma arquitetura de um sistema de monitoramento de saúde baseado em IoT que pode ser usado em hospitais inteligentes. Nestes sistemas, as informações de saúde do paciente são coletadas por sensores colocados no corpo do paciente. Esses dados de saúde também podem ser complementados com informações de contexto, como data, hora, local, temperatura etc. Conhecer o contexto nos permite identificar padrões incomuns e fazer inferências mais precisas sobre a situação. Outros sensores e atuadores (equipamentos médicos) também podem ser conectados a sistemas de transmissão de dados para o pessoal médico, como imagens de alta resolução, como tomografias computadorizadas.

Essa arquitetura consiste em uma rede de sensores sem fio (WSN - Wireless Sensor Network), um gateway e um servidor supervisor, em que médicos e enfermeiros (supervisores) podem monitorar pacientes em tempo real. O WSN é responsável por detectar e coletar os sinais biomédicos e de contexto que são capturados do corpo e do ambiente. Estes dados serão utilizados para tratamento e diagnóstico dos pacientes. Os dados são transmitidos para o gateway através de protocolos de comunicação sem fio, como Bluetooth, Wi-Fi ou IEEE 802.15.4. O gateway suporta diferentes protocolos de comunicação e atua como um ponto de contato entre a rede de sensores e o supervisor. O gateway recebe dados de diferentes sub-redes, realiza a conversão de protocolos e fornece outros serviços de nível superior, como agregação de dados, filtragem e assim por diante. O servidor supervisor é composto por nós de processamento; Esses nós, dependendo da demanda, podem ser virtualizados (contêineres) ou máquinas físicas com redundância.

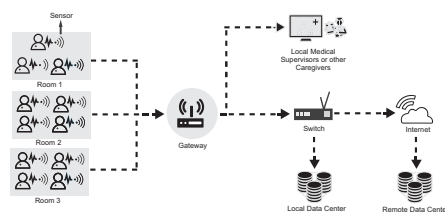


Figura 1. Arquitetura de um Hospital Inteligente

A Figura 2 representa um modelo de SPN para a arquitetura apresentada com as seguintes funções: (i) *Admission* que lida com a chegada dos dados; (ii) *Gateway* que encaminha os dados para o supervisor e para o switch; (iii) *Supervisor* que recebem os dados para realizar o monitoramento do paciente. Os componentes são representados

por gráficos, como locais (círculos), transições temporizadas (barras vazias) e marcações (pequenas bolas pretas). (iv) *Switch* é o principal responsável por gerenciar a rota para o servidor local. (v) *Local Server* armazena os dados do paciente para análise futura e para evitar a perda de dados. (vi) *Remote Server* processa e armazena os dados na nuvem.

Dada a visão geral do modelo, descrevemos agora o fluxo de processamento de dados entre seus componentes. A sub-rede Admission é composta de dois lugares $P_Arrival$ e $P_InputQueue$, que representam a espera entre de dados e aceitação destes dados na fila. Os tokens gerados em $P_Arrival$ representam qualquer tipo de requisição que envolva entrada de dados a ser processada e armazenada. A transição TO representa o recebimento da requisição, note que é uma transição imediata, não possuindo atraso associado. TO dispara assim que existir um token em $P_InputQueue$ e pelo menos um token em $P_GatewayCapacity$. A capacidade do gateway pode ser interpretada como canais de distribuição disponíveis, dada pela marcação GC.

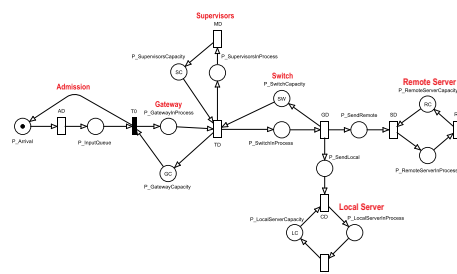


Figura 2. Modelo SPN para Hospitais Inteligentes

Quando TO dispara, a sub-rede do gateway é alcançada. Um token é retirado de $P_InputQueue$ e $P_GatewayCapacity$. Um token é retornado ao $P_Arrival$, permitindo um novo disparo. Um token é adicionado então ao lugar $P_GatewayInProcess$. A quantidade de tokens em $P_GatewayInProcess$ representa o enfileiramento de requisições no Gateway. O enfileiramento ocorre quando não há capacidade disponível para servir a requisição recém-chegada. Se houver capacidade disponível no servidor supervisor e no switch (lugar $P_SupervisorsCapacity$ e $P_SwitchCapacity$), a transição TD é disparada para ambos simultaneamente e posteriormente a requisição segue em frente para ser processada. TD representa o tempo necessário para o gateway enviar uma requisição ao servidor supervisor.

Caso haja recurso disponível no servidor supervisor, o lugar $P_SupervisorsInProcess$ conterá a quantidade de requisições na fila para processamento. O tempo que os trabalhos permanecem em processamento em um nó depende da transição MD . Tais transições possuem a semântica *infinite server*, então, cada dado é processado independentemente. É importante notar que o tempo de processamento depende muito da capacidade computacional dos nós do servidor supervisor. Assim, a transição MD deve ser configurada com um tempo de processamento para um único trabalho em um tipo de recurso específico. Após o disparo da transição GD a requisição segue em frente para ser processada tanto no servidor local quanto no servidor remoto simultaneamente.

Em seguida, é necessário aguardar a hora de chegada dos dados no servidor local, esse tempo é representado pela transição CD . Após a chegada da requisição, esta segue em frente para ser processada. O tempo que os trabalhos permanecem em processamento no servidor local depende da transição LD . LD representa o tempo de serviço do servidor local. Tal transição possui a semântica *infinite server*, então, cada requisição é processada independentemente caso haja recurso disponível em $P_LocalServerCapacity$. Novamente, é importante notar que o tempo de processamento depende muito da capacidade compu-

tacional do nó (VM ou container por exemplo) bem como do banco de dados utilizado para armazenar os dados. Assim, a transição *LD* deve ser configurada com um tempo de processamento para um único trabalho em um tipo de recurso específico.

Se não houver nenhum recurso disponível no servidor remoto (*P_RemoteServerCapacity*), haverá enfileiramento no lugar *P_SendRemote*. Em caso afirmativo, há um tempo associado ao envio de solicitação para o servidor remoto (transição *SD*) e um tempo de serviço no servidor remoto (transição *RD*). Essas transições possuem a semântica *infinite server*, portanto, cada solicitação é processada independentemente.

O tempo entre as chegadas é atribuído à transição *AD*. Consideramos que os tempos entre os disparos são exponencialmente distribuídos, essa suposição pode ser modificada pela alteração dessa distribuição. A transição *AD* leva em conta apenas o tempo que os pedidos entraram no sistema, ou seja, as perdas por conta da rede não são levadas em consideração.

O modelo proposto permite avaliar diversos cenários pois o avaliador pode configurar até 13 parâmetros (8 transições e 5 marcações). Qualquer alteração em um destes parâmetros pode impactar significativamente no tempo médio de resposta do sistema e conseqüentemente no custo de infraestrutura. A variação das possibilidades de cenários considerando um grande número de fatores da arquitetura do Smart Hospital é o que torna este modelo uma das principais contribuições deste trabalho.

4. Métricas

Esta seção define métricas para avaliar a arquitetura de um Hospital Inteligente com base no modelo proposto. Neste trabalho calculamos quatro métricas: tempo médio de resposta (*MRT*), probabilidade de descarte (*Discard_Probability*), número de requisições descartadas em função do tempo (*Discard_Number*) e probabilidade de utilização de recursos (*Utilization<ServerName>*). O *MRT* pode ser obtido pela Lei de Little [Little and Graves 2008] que relaciona o número médio de solicitações em andamento em um sistema (*RequestsInProgress*), a taxa de chegada de novas solicitações (*ArrivalRate*). O *MRT* é obtido pela equação: $MRT = \frac{RequestsInProgress}{ArrivalRate}$. A taxa de chegada é o inverso do atraso de chegada: $ArrivalRate = \frac{1}{AD}$. Portanto, é possível também calcular o *MRT* através da equação: $MRT = RequestsInProgress \times AD$. A Lei de Little requer um sistema estável, o que significa que a taxa de chegada é menor que o tempo de serviço.

Para calcular o número de requisições em progresso no sistema, precisa-se somar a quantidade de tokens em cada um dos lugares que representam uma requisição em andamento. Esse número de requisições é obtido pela equação: $RequestsInProgress = Esp(P_GatewayInProgress) + Esp(P_SupervisorsInProgress)$. Na equação $RequestsInProgress$, $Esp(Lugar)$ representa a esperança estatística de existir tokens em “Lugar”, onde $Esp(Lugar) = (\sum_{i=1}^n P(m(Lugar) = i) \times i)$. Em outras palavras, $Esp(Lugar)$ indica quantos tokens ocupam aquele Lugar.

Para calcular o descarte é necessário existir token na fila de entrada (*P_ArrivalQueue*) e não restar mais nenhum recurso disponível no gateway. A probabilidade de ocorrer descarte é dada pela equação: $Discard_Probability = (P((P_InputQueue = 1) \wedge (P_GatewayCapacity = 0))) \times 100$. Onde $P(Lugar = n)$ calcula a probabilidade de existirem *n* tokens em “Lugar”.

Para obter o número de descarte em um período de tempo, basta multiplicar a probabilidade de descarte por *ArrivalRate* e o tempo *T*. A equação que define o número de descartes em um determinado período de tempo *T* é dado por: $Discard_Number = Discard_Probability \times ArrivalRate \times T$. Por fim, nós calculamos também a probabilidade de utilização dos recursos, onde a equação é dada pelo número de tokens do lugar correspondente ao momento da execução dividido pelo total da capacidade daqueles recursos. Fazemos isso para cada recurso do modelo. Observe que a capacidade em questão é dada pela marcação do lugar correspondente àquele recurso.

A equação para Utilização do Gateway é dada por: $UtilizationGateway = \frac{Esp(P_GatewayInProcess)}{GC} \times 100$. A equação para Utilização do servidor supervisor é dada por: $UtilizationSupervisors = \frac{Esp(P_SupervisorsInProcess)}{SC} \times 100$. A equação para utilização do switch é dada por: $UtilizationSwitch = \frac{Esp(P_SwitchInProcess)}{SW} \times 100$. A equação para a utilização do servidor local é dada por: $UtilizationLocalServer = \frac{Esp(P_LocalServerInProcess)}{LC} \times 100$. A Equação para Utilização do Servidor Remoto é dada por: $UtilizationRemoteServer = \frac{Esp(P_RemoteServerInProcess)}{RC} \times 100$.

5. Análises Numéricas

Esta seção apresenta oito análises numéricas do modelo. Os valores atribuídos às transições temporizadas e às marcações do modelo SPN para realizar uma avaliação de desempenho baseada em análise numérica, são listados a seguir. Transições temporizadas: AD = 1.0-10.0(incremento de 0.5), TD = 3.5, MD = 20.3, GD = 2.5, CD = 1.0, LD = 20.3, SD = 4.0, RD = 20.3. Marcações: GC = 8, SC = 8, SW = 8, LC = 16, RC = 16. Nós novamente variamos o valor da transição correspondente ao arrival delay (AD) de 1.0ms até 10.0ms com incrementos de 0.5ms. Os outros parâmetros permaneceram fixos. Todos os resultados da análise com o modelo são apresentados nos gráficos das Figuras 3(a), 3(b), 3(c), 3(d) e 3(e).

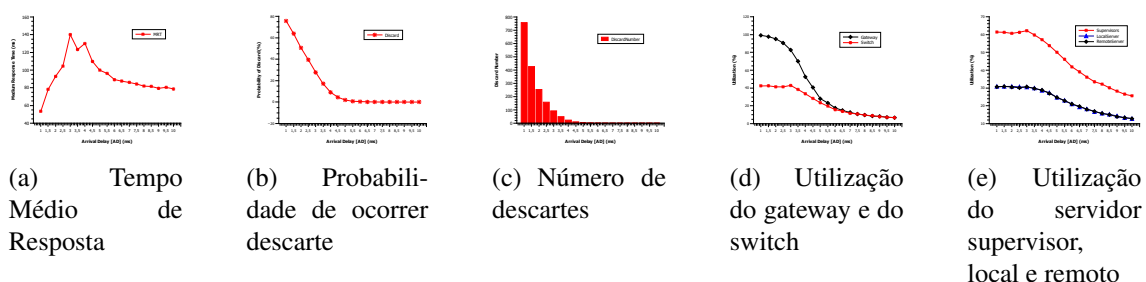


Figura 3. Resultados da avaliação de desempenho do modelo

A Figura 3(a) apresenta os resultados para MRT que aumenta até AD=3.0ms, com MRT=140ms para logo depois diminuir. A Figura 3(b) apresenta a probabilidade de ocorrer descarte de requisições. É possível notar que essa probabilidade diminui à medida que o AD aumenta (entre 1.0ms e 5.5ms), até estagnar em zero quando o AD chega a 6.0ms. A Figura 3(c) exhibe o número de descartes em um determinado período de tempo T (10.0ms). Conforme o AD aumenta, o número de dados descartados diminui e a partir de 6.5 ms o número de descartes chega a zero. A Figura 3(d) apresenta o nível de utilização do Gateway e do Switch, onde o Switch possui uma menor porcentagem de utilização. Isso ocorre devido o gargalo no Gateway. A Figura 3(e) apresenta o nível de

utilização do servidor supervisor, do servidor local e do servidor remoto. A diferença de utilização se dá por conta da capacidade utilizada no cenário: servidor local (16 núcleos), servidor remoto (16 núcleos) e servidor supervisor (8 núcleos).

6. Conclusão

Este trabalho adotou um modelo analítico para avaliar o desempenho de sistemas hospitalares inteligentes sem ter que investir em equipamentos reais de antemão. Usando SPNs, representamos e avaliamos arquiteturas hospitalares inteligentes. Nosso modelo permite a configuração de 13 parâmetros, possibilitando avaliar um grande número de diferentes cenários. Em diferentes cenários, foi possível observar a estreita relação entre MRT, utilização e descarte, principalmente para altas taxas de chegada.

Referências

- Akmandor, A. O. and Jha, N. K. (2018). Smart health care: An edge-side computing perspective. *IEEE Consumer Electronics Magazine*, 7(1):29–37.
- Chen, M., Li, W., Hao, Y., Qian, Y., and Humar, I. (2018). Edge cognitive computing based smart healthcare system. *Future Generation Computer Systems*, 86:403–411.
- Greco, L., Ritrovato, P., and Xhafa, F. (2019). An edge-stream computing infrastructure for real-time analysis of wearable sensors data. *Future Generation Computer Systems*, 93:515–528.
- Little, J. D. and Graves, S. C. (2008). Little’s law. In *Building intuition*, pages 81–100. Springer.
- Oueida, S., Kotb, Y., Aloqaily, M., Jararweh, Y., and Baker, T. (2018). An edge computing based smart healthcare framework for resource management. *Sensors*, 18(12):4307.
- Rahmani, A. M., Gia, T. N., Negash, B., Anzanpour, A., Azimi, I., Jiang, M., and Liljeberg, P. (2018). Exploiting smart e-health gateways at the edge of healthcare internet-of-things: A fog computing approach. *Future Generation Computer Systems*, 78:641–658.
- Rahmani, A.-M., Thanigaivelan, N. K., Gia, T. N., Granados, J., Negash, B., Liljeberg, P., and Tenhunen, H. (2015). Smart e-health gateway: Bringing intelligence to internet-of-things based ubiquitous healthcare systems. In *2015 12th Annual IEEE Consumer Communications and Networking Conference (CCNC)*, pages 826–834. IEEE.
- Sigwele, T., Hu, Y. F., Ali, M., Hou, J., Susanto, M., and Fitriawan, H. (2018). An intelligent edge computing based semantic gateway for healthcare systems interoperability and collaboration. In *2018 IEEE 6th International Conference on Future Internet of Things and Cloud (FiCloud)*, pages 370–376. IEEE.
- Tata, S., Jain, R., Ludwig, H., and Gopisetty, S. (2017). Living in the cloud or on the edge: Opportunities and challenges of iot application architecture. In *2017 IEEE International Conference on Services Computing (SCC)*, pages 220–224. IEEE.
- Zhang, H., Li, J., Wen, B., Xun, Y., and Liu, J. (2018). Connecting intelligent things in smart hospitals using nb-iot. *IEEE Internet of Things Journal*, 5(3):1550–1560.