Escalabilidade no contexto de Prontuário Eletrônico do Paciente baseado em *Blockchain*: Um Estudo Experimental sobre Armazenamento *Off-chain*

Pamella Soares[1]✉, Allysson Allex Araújo[2], Raphael Saraiva[1] e Jerffeson Souza[1]

[1] Programa de Pós-Graduação em Ciência da Computação (PPGCC)
Universidade Estadual do Ceará (UECE) – Fortaleza, Ceará - Brasil
[2] Grupo de Estudos em Sistemas de Informação e Inovação Digital (GESID)
Universidade Federal do Ceará (UFC) – Crateús, Ceará – Brasil
pamella.soares@aluno.uece.br

Abstract. *The Electronic Patient Record (EPR) is a fundamental structure for maintaining information about the individual's health, whose data demand interoperability and privacy in the provision of care to patients. In this context, blockchain is promising as it allows different actors to form a network that shares reliable information, favoring availability and integrity. However, such integration must still deal with the challenge of blockchain scalability, which can be an impasse in EPRs, given the amount of data intrinsic to the health domain. As part of ongoing research, this article presents a preliminary architectural model and an experimental study that evaluates the use of an off-chain strategy to deal with the scalability challenge in the context of blockchain-based EPR. It contributes, therefore, to the advancement of blockchain adoption in EPRprojects.*

Resumo. *O Prontuário Eletrônico do Paciente (PEP) é uma estrutura fundamental para manutenção da informação sobre a saúde do indivíduo, cujos dados demandam interoperabilidade e privacidade no fornecimento de cuidados aos pacientes. Nesse contexto, o uso de blockchain se demonstra promissor por permitir que diferentes atores formem uma rede que compartilha informações confiáveis, favorecendo a disponibilidade e integridade. Porém, tal integração ainda deve lidar com o desafio da escalabilidade em blockchain, o que pode ser um impasse em PEPs tendo em vista a quantidade de dados intrínseca ao domínio da saúde. Como parte de uma pesquisa em desenvolvimento, este artigo apresenta um modelo arquitetural preliminar e um estudo experimental o qual avalia o uso de uma estratégia off-chain para lidar com o desafio da escalabilidade no contexto de PEP baseado em blockchain. Contribui-se, portanto, no avanço da adoção de blockchain em projetos de PEP.*

1. Introdução

O Prontuário Eletrônico do Paciente (PEP) é uma estrutura para a manutenção de informações eletrônicas sobre a saúde das pessoas e os cuidados que recebem ao longo de suas vidas por meio de equipes multidisciplinares [Massad et al. 2003]. Além de fins de armazenamento, pode promover a interoperabilidade entre diferentes instituições, consequentemente, a continuidade do atendimento ao paciente. Nesse sentido, um requisito necessário é promover uma adequada proteção da privacidade dos dados do paciente, inclusive, embasado pela Lei Geral de Proteção de Dados Pessoais (LGPD) [Brasil 2018].

Dadas as particularidades mencionadas, a tecnologia *blockchain* mostra-se promissora no contexto do PEP, pois partes totalmente anônimas podem formar uma rede que compartilha e armazena informações confiáveis. De modo geral, o *blockchain* pode ser entendido como um estrutura distribuída de armazenamento de dados, protegido por criptografia e regido por um mecanismo de consenso. Tal tecnologia apresenta benefícios intrínsecos como capacidade de auditoria, autenticidade, disponibilidade, não repudiabilidade e integridade das transações. Por sua vez, no "topo" da *blockchain* são executados os Contratos Inteligentes (CIs), que consistem em *scripts* de execução da lógica de negócio da rede que ocorre quando determinadas condições são atendidas [Bashir 2017].

A arquitetura de software é a base de qualquer sistema intensivo de software [Bass et al. 2003]. Portanto, considerar requisitos de qualidade nas decisões de projetos é fundamental para a qualidade desses sistemas. No contexto de PEP, por exemplo, um atributo essencial é a escalabilidade, visto que os sistemas usualmente devem lidar com alto volume de dados [Bacelar and Correia 2015]. Apesar de *blockchain* atrelado ao uso de PEPs ser considerado promissor, a escalabilidade ainda se demonstra um desafio. Considerando um sistema de saúde de larga escala, por exemplo, se a *blockchain* é usada para armazenar dados do paciente, os milhares de registros serão replicados em todos os participantes do *blockchain*, o que pode ocasionar em gargalos devido o alto grau de processamento necessário entre os nós da rede [Agbo et al. 2019]. Isso posto, diferentes abordagens utilizando-se de estratégias *off-chain* têm sido investigadas para mitigar problemas de escalabilidade em *blockchain*, que está relacionada à terceirização do armazenamento dos dados fora da *blockchain* [Shukla and Samet 2020].

Como parte de um estudo em desenvolvimento, este artigo objetiva discutir um estudo experimental, a partir de uma modelagem arquitetural preliminar, o qual busca avaliar o uso de uma estratégia *off-chain* para lidar com o desafio da escalabilidade no contexto de PEP baseada em *blockchain*. Contribui-se, assim, ao discutir possíveis caminhos para adoção de *blockchain* para PEP, sendo estes importantes sistemas intensivos.

2. Modelagem Arquitetural Preliminar

A Figura 1 apresenta uma versão preliminar da arquitetura, na qual tem-se a representação do ativo a ser adicionado e a arquitetura composta pela Camada de Armazenamento *On-chain*, referente aos dados adicionados dentro da *blockchain*; Camada de Armazenamento *Off-chain*, referente ao armazenamento de dados brutos em bancos de dados externos e; a Camada de Aplicação, que realiza a comunicação entre as camadas e usuários.

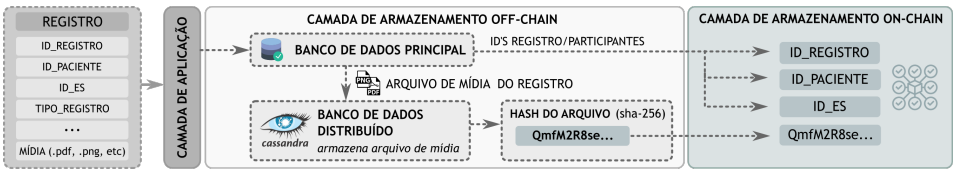


Figura 1. Proposta de Modelo Arquitetural Preliminar.

Em suma, como pode-se observar, não se realiza a inserção dos registros brutos em *blockchain* visto que a Camada de Armazenamento *On-chain* armazena apenas os ponteiros ou identificadores (IDs) advindos dos registros nos bancos de dados que compõem a Camada de Armazenamento *Off-chain*. O ativo em questão é genérico para representar um tipo de registro do paciente o qual contém os identificadores: id do registro, id do

paciente que o detém e o id da Entidade de Saúde (ES) que o inseriu. Adicionalmente, tem-se o campo tipo de registro e o arquivo de mídia associado.

3. Estudo Experimental

Uma Prova de Conceito (PoC) baseada no modelo arquitetural apresentado previamente foi desenvolvida para o estudo experimental. O CI implementado para a Camada de Armazenamento *On-chain* foi implantado na *blockchain* da *Ethereum*. Para tal, utilizou-se o *Ganache*, uma ferramenta especialmente utilizada para simular a execução de uma rede *Ethereum* local e de CIs. Os testes foram aplicados sobre duas funções do CI que foram escolhidas por manipularem o '`Registro`'. Dessa forma, objetivou-se simular as seguintes operações: o armazenamento de dados referentes a um registro do paciente no formato de mídia convertida em `base64` (função `inserirRegistro`) e a busca por registros específicos de dado paciente (função `lerRegisto`), também no formato `base64`. A escolha do arquivo em formato `base64` está sendo utilizada para fins de teste nas camadas da arquitetura implementada. Além disso, é um método que vem sendo utilizado para codificação de arquivo de mídia para transferência na internet.

Com o objetivo de avaliar a arquitetura proposta, foram criados dois tipos de casos, são eles: (i) o **Caso Base** pelo qual o registro no formato `base64` é inserido totalmente na *blockchain*, ou seja, simula todo arquivo de mídia sendo armazenado em *blockchain*; e (ii) o **Caso *Off-chain***, que representa a solução proposta, na qual apenas os IDs do registro e dos participantes são armazenados em *blockchain*, além do *hash* do arquivo, sendo que o arquivo original é armazenado no Camada de Armazenamento *Off-chain*. Por sua vez, a fim de analisar diferentes tamanhos de arquivos armazenados no artefato, em cada caso são simulados registros com os tamanhos: 1000 *bytes* à 2000 bytes; 2000 *bytes* à 4000 bytes; e 4000 *bytes* à 7000 bytes. Apesar dos tamanhos representarem arquivos pequenos, pode-se representar a comparação entre os dois casos. Para cada tamanho de arquivo, os testes foram executados 30 vezes para lidar com possíveis *outliers* dos valores coletados. Além disso, cada execução simulou 10 requisições para as operações realizadas. A configuração do estudo experimental é sintetizada na Figura 2.

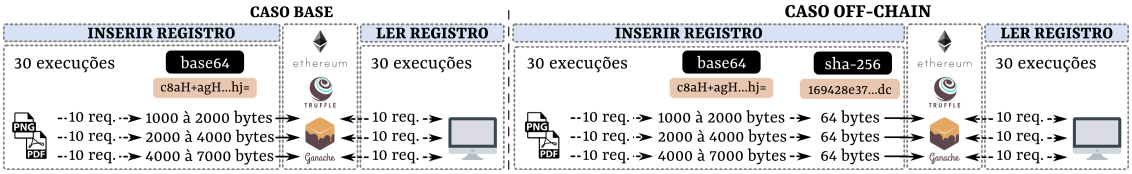


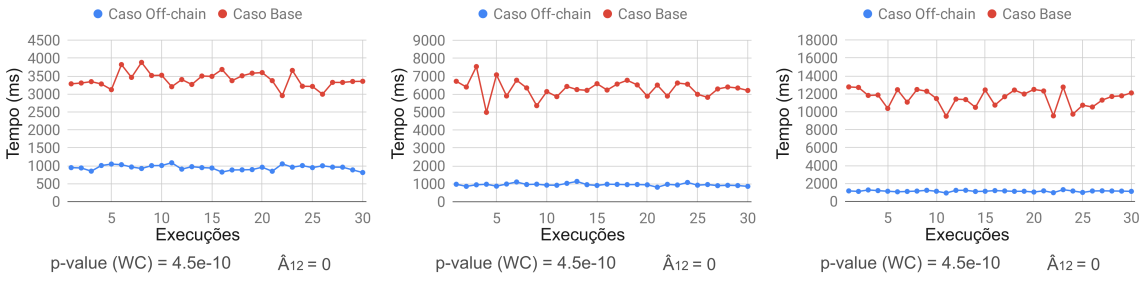
Figura 2. Configurações do Experimento.

As métricas analisadas foram: 1) **tempo de publicação**, calculado pela diferença no tempo entre a requisição de inserção dos dados e a mineração de uma transação em um bloco e; 2) **tempo de busca**, oriundo da diferença entre a requisição de busca de um registro na *blockchain* utilizando seu respectivo id e o retorno das informações. De forma semelhante, o cálculo é realizado ao interagir com a Camada de Armazenamento *Off-chain*. Os cálculos descritos na definição das métricas, anteriormente mencionadas, foram implementados por meio de *scripts* para coleta dos dados de cada transação realizada em *blockchain*. O experimento foi executado em uma computador com 16GB de RAM, SSD de 256GB e processador *Intel Core* i7 (8a geração). Adicionalmente, foram concretizados testes estatísticos de Wilcoxon, que identifica a ocorrência de diferença estatística entre

as amostras considerando um nível de confiança de 95% utilizando-se da correção de Bonferroni; e Vargha-Delaney ($\hat{A}_{12}$), o qual retorna o número relativo de vezes que um tipo de caso (*off-chain* e base) produziu valores superiores ao outro.

3.1. Resultados e Análises Preliminares

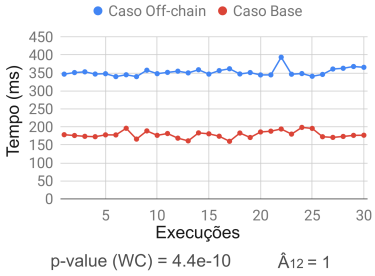
Quanto ao **tempo de publicação**, a Figura 3a, referente à arquivos de tamanho entre 1KB e 2KB, mostra que os tempos vão até aproximadamente 3882 ms, com média das execuções de 3397 ms para a publicação no Caso Base, enquanto para o Caso *Off-chain*, os tempos tem o máximo de até 1082 ms, e a média das execuções é de 947 ms. Já na Figura 3b, referente à arquivos de tamanho entre 2KB e 4KB, o tempo máximo atingiu 7540 ms, com média de 6307 ms no Caso Base.



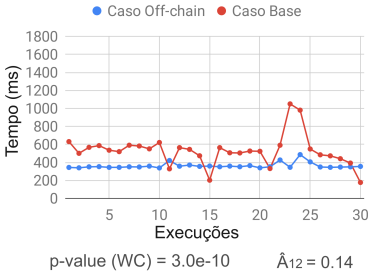
(a) Tamanho entre 1KB à 2KB para publicação.

(b) Tamanho entre 2KB à 4KB para publicação.

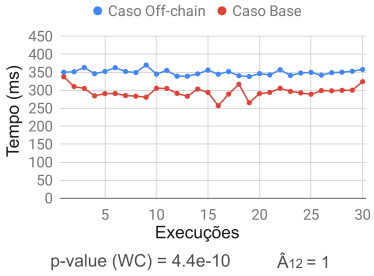
(c) Tamanho entre 4KB à 7KB para publicação.



(d) Tamanho entre 1KB à 2KB para busca.



(e) Tamanho entre 2KB à 4KB para busca.



(f) Tamanho entre 4KB à 7KB para busca.

Figura 3. Tempos de Publicação e Busca para diferentes tamanhos de arquivos.

Por sua vez, no Caso *Off-chain*, o tempo máximo obtido foi 1133 ms e a média foi de 952 ms. Por fim, na Figura 3c, referente a arquivos entre 4KB e 7KB, o tempo máximo atingiu 12770 ms, tendo como média das execuções 11543 ms para a publicação no Caso Base. No Caso *Off-chain*, o tempo máximo obtido foi 1306 ms e a média das execuções 1141 ms. Em todos os tamanhos de arquivo, o Caso *Off-chain* apresenta médias aproximadas, entre 947 ms e 1306 ms, o que era esperado visto que a proposta, em todas as situações, converte o arquivo em *hash*. Por sua vez, no Caso Base, tende a aumentar à medida que o tamanho do arquivo aumenta. Conforme os testes de Wilcoxon e Vargha-Delaney realizados, concluiu-se que, para todas as situações, existem evidências estatísticas que mostram que o tempo de publicação do Caso Base é significantemente maior que o do Caso *Off-chain* em 100% das execuções para todos os tamanhos de arquivos, visto que o *p-value* mostra-se inferior à 0,5 e o $\hat{A}_{12}$ equivale à 0. Em outras palavras, o tempo de publicação/mineração do Caso *Off-chain* foi mais rápido que o Caso Base.

Por sua vez, quanto ao **tempo de busca**, a Figura 3d mostra o tempo com um máximo de 199 ms e uma média de 179 ms no Caso Base, e um máximo de 394 ms e uma

média de 352 ms no Caso *Off-chain*. Constata-se que o tempo de busca por um dado ficou em até 1050 ms, com uma média de 530 ms no Caso Base de arquivos com tamanhos entre 2KB e 4KB (Figura 3e). Por sua vez, no Caso *Off-chain*, obteve-se um máximo de tempo de busca igual à 486 ms, com a média de 362 ms. Observando a Figura 3f, tem-se um tempo máximo de busca de 337 ms, com média de 296 ms no Caso Base e, no Caso *Off-chain*, um máximo de tempo de busca igual a 370 ms, com a média igual a 349 ms. Nas situações das Figuras 3d e 3f, o Caso *Off-chain* demonstra-se com o maior tempo de busca em relação ao Caso Base. A partir disso, supõe-se que o tempo de resposta é menor no Caso Base porque os dados estão sendo buscados diretamente na *blockchain*. Enquanto para os dados do Caso *Off-chain*, estes estão sendo pesquisados na Camada de Armazenamento *Off-chain*, demandando um maior tempo. Os testes estatísticos mostraram que os casos são significativamente diferentes, com o Caso *Off-chain* tendo maiores valores de tempo de busca em 100% das execuções em relação ao Caso Base. Por outro lado, a Figura 3e apresenta um comportamento diferente dos demais, pois o Caso *Off-chain* obteve um tempo maior em 14% das execuções em relação ao Caso Base. Além disso, o Caso Base apresenta uma maior variação nos valores de tempo de resposta.

4. Considerações Finais e Trabalhos Futuros

O presente trabalho discutiu um estudo experimental baseado em um modelo arquitetural preliminar projetado com objetivo de atrelar as vantagens da tecnologia *blockchain* na implementação dos PEPs de forma que o desafio de escalabilidade seja amenizado através do uso de armazenamento *off-chain*. De maneira geral, os tempos de resposta da inserção e da leitura de registros do paciente foram reduzidos consideravelmente. Dessa forma, concluiu-se que o uso de estratégias de armazenamento *off-chain* se demonstra proeminente para lidar com o desafio da escalabilidade no contexto de PEP. Ademais, pretende-se evoluir este trabalho para uma modelagem de simulação detalhada considerando a construção de um modelo mais robusto, ao especificar as camadas e os componentes da arquitetura quanto suas especificidades, retratando com fidedignidade o sistema em estudo. A partir disso, pretende-se configurar e validar uma modelagem computacional, inclusive, para facilitar a experimentação em situações extremas. Por fim, prosseguir com novos experimentos utilizando outras entradas e cenários mais complexos, além do uso de diferentes plataformas de *blockchain*.

Referências

Agbo, C. C., Mahmoud, Q. H., and Eklund, J. M. (2019). Blockchain technology in healthcare: a systematic review. In *Healthcare*, volume 7, page 56. Multidisciplinary Digital Publishing Institute.

Bacelar, G. and Correia, R. (2015). As bases do openEHR. Virtual Care.

Bashir, I. (2017). *Mastering blockchain*. Packt Publishing Ltd.

Bass, L., Clements, P., and Kazman, R. (2003). *Software architecture in practice*. Addison-Wesley Professional.

Brasil (2018). Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD).

Massad, E., Marin, H. F., and Azevedo-Neto, R. S. (2003). O prontuário eletrônico do paciente na assistência, informação e conhecimento médico.

Shukla, P. A. and Samet, S. (2020). Systematization of knowledge on scalability aspect of blockchain systems. In *Future of Information and Communication Conference*, pages 130–138. Springer.