UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Câmpus de São José do Rio Preto

INSTITUTO DE BIOCIÊNCIAS, LETRAS E CIÊNCIAS EXATAS
IBILCE / UNESP - CÂMPUS DE SÃO JOSÉ DO RIO PRETO

GABRIEL HENRIQUE DA SILVA GAVA

Framework para tratamento de prontuários médicos com suporte ao compartilhamento e privacidade de dados

São José do Rio Preto
2021

GABRIEL HENRIQUE DA SILVA GAVA

Framework para tratamento de prontuários médicos com suporte ao compartilhamento e privacidade de dados

Trabalho de Conclusão de Curso apresentado ao Departamento de Ciências de Computação e Estatística do Instituto de Biociências Letras e Ciências Exatas da Universidade Estadual Paulista "Júlio de Mesquita Filho", como parte dos requisitos necessários para obtenção do grau de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Carlos Roberto Valêncio

São José do Rio Preto
2021

G279f

Gava, Gabriel Henrique Silva

Framework para tratamento de prontuários médicos com suporte ao compartilhamento e privacidade de dados / Gabriel Henrique Silva Gava. -- São José do Rio Preto, 2021

51 p. : il., tabs.

Trabalho de conclusão de curso (Bacharelado - Ciência da Computação) - Universidade Estadual Paulista (Unesp), Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto

Orientador: Carlos Roberto Valêncio

1. Banco de Dados. 2. Registros médicos. 3. Blockchain. 4. Framework. 5. Segurança. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

GABRIEL HENRIQUE DA SILVA GAVA

Framework para tratamento de prontuários médicos com suporte ao compartilhamento e privacidade de dados

Trabalho de Conclusão de Curso apresentado ao Departamento de Ciências de Computação e Estatística do Instituto de Biociências Letras e Ciências Exatas da Universidade Estadual Paulista "Júlio de Mesquita Filho", como parte dos requisitos necessários para obtenção do grau de Bacharel em Ciência da Computação.

Banca examinadora:

Prof. Dr. Carlos Roberto Valêncio
UNESP – São José do Rio Preto
Orientador

Prof. Dr. Geraldo Francisco Donega Zafalon
UNESP – São José do Rio Preto

Profa. Dr. Valeriano Antunes de Oliveira
UNESP – São José do Rio Preto

São José do Rio Preto
2021

Agradecimentos

Agradeço primeiramente aos meus pais, em especial minha mãe Gislene, por todo amor e apoio incondicional em todo meu trajeto acadêmico, desde os primeiros passos no ensino fundamental até o último dia de minha graduação.

A todos os meus amigos do curso de graduação que compartilharam dos inúmeros desafios que enfrentamos juntos, ao Douglas Canevarollo por toda vez que dedicou do seu tempo para me ajudar e ensinar, ao Thiago Leal por me incentivar e auxiliar várias vezes quando acreditava que não conseguiria fazer algo, ao Douglas Brandão por compartilhar de diversas aventuras e desafios que vão além da graduação, ao Rafael pelo auxílio e conselhos para confecção deste trabalho, ao David por sempre se mostrar presente para ajudar, ao Douglas Honda e Pedro Afonso pelos momentos mais diversos que compartilhamos e, finalmente, ao Gustavo Villar por ser meu grande ombro amigo e minha dupla preferida para todos os momentos da vida.

Ao meu namorado Maikel Oliveira, pelo carinho, apoio e incentivo durante todo processo de realização deste trabalho.

A todos os integrantes do Grupo de Banco de Dados – GBD, sem os quais esse trabalho não seria possível.

A todos os professores que tive durante esses anos, em especial ao Prof. Dr. Carlos Roberto Valêncio por ter me orientado nesse trabalho e pelas valiosas contribuições dadas durante todo o processo de confecção deste.

“Treinar na solidão serve apenas para perpetuar erros.”

- Vesemir (The Witcher)

Resumo

Os prontuários eletrônicos, assim como outros dados médicos, são fontes de informações valiosas devido à sua natureza e seu elevado potencial na extração de conhecimento e na tomada de decisão no ramo da saúde. Tais características implicam em uma alta demanda no compartilhamento dessas informações junto a instituições de pesquisas e hospitais. Entretanto, estes dados, em sua grande parte, possuem informações pessoais ou sensíveis, de tal modo que seu compartilhamento pode ferir os direitos de privacidade individual. Dessa forma, a proteção e privacidade no armazenamento e compartilhamento de dados médicos é um grande desafio no ramo da saúde. Diante deste problema, a tecnologia *blockchain*, complementada por algoritmos de anonimização aplicados sobre as informações pessoais do paciente, pode ser usada como uma alternativa na segurança no compartilhamento dos dados, assim como na preservação da privacidade do paciente e na prevenção de possíveis ataques. A literatura disponibiliza trabalhos que apresentam alguns desses componentes, mas não de maneira agrupada e de fácil implementação. Deste modo, este trabalho propõe como principal contribuição o desenvolvimento de um *framework* capaz de oferecer recursos para criação de sistemas gerenciadores de prontuários eletrônicos seguros, com o uso da tecnologia *blockchain*, bem como diferentes métodos e algoritmos de anonimização de dados para preservação da privacidade individual dos pacientes. Os experimentos realizados mostraram que o *framework* é capaz de ser aplicado em um ambiente real sem gerar carga demasiada para os sistemas de saúde, permitindo que seja possível o compartilhamento de dados médicos anonimizados para instituições de pesquisa e colaborando para o desenvolvimento da ciência.

Palavras-chave: Banco de dados. Registros médicos. *Blockchain*. Anonimização. Compartilhamento de dados. *Framework*. Privacidade. Segurança.

Abstract

Electronic medical records, like other medical data, are valuable sources of information due to their nature and their high potential in knowledge extraction and decision-making in the healthcare field. Such characteristics imply a high demand for sharing this information with research institutions and hospitals. However, this data, for the most part, contains personal or sensitive information, in such a way that its sharing may violate individual privacy rights. Thus, the protection and privacy in the storage and sharing of medical data is a major challenge in the healthcare field. Faced with this problem, the blockchain technology, complemented by anonymization algorithms applied to the patient's personal information, can be used as an alternative to secure data sharing, as well as to preserve the patient's privacy and prevent possible attacks. The literature provides works that present some of these components, but not in a grouped and easy-to-implement way. Thus, this work proposes as its main contribution the development of a framework capable of offering resources for the creation of secure electronic medical records management systems, using blockchain technology, as well as different methods and algorithms for data anonymization to preserve individual privacy of patients. The experiments carried out showed that the framework is capable of being applied in a real environment without generating too much load for health systems, allowing the sharing of anonymized medical data to research institutions and contributing to the development of science.

Keywords: Database. Medical records. Blockchain. Anonymization. Data sharing. Framework. Privacy. Safety.

Lista de Ilustrações

Lista de Tabelas

Lista de Abreviaturas e Siglas

API – *Application Programming Interface*

CFM – Conselho Federal de Medicina

CID – Classificação Estatística Internacional de Doenças e Problemas Relacionados com a Saúde

EUA – Estados Unidos da América

GBD – Grupo de Banco de Dados

GPE – Gerenciador de Prontuários Eletrônicos

HTML – *Hypertext Markup Language*

HTTPS – *Hypertext Transfer Protocol Secure*

IBM - *International Business Machines*

LGPD – Lei Geral de Proteção de Dados Pessoais

NONCE – *Number Only Used Once*

ORM – *Object-Relational Mapping*

PEP – Prontuário Eletrônico do Paciente

P2P – *Peer to Peer*

PoW - *Proof-of-Work*

PoS - *Proof-of-Stake*

PBFT - *Pratical Byzantine Fault Tolerance*

PLN – Processamento de Linguagem Natural

REN – Reconhecimento de Entidades Nomeadas

SDK – *Software Development Kit*

SGBD – Sistema Gerenciador de Banco de Dados

TPS – Transação por segundo

UDT – *User-Defined Types*

Sumário

1 Introdução

Hospitais e clínicas de saúde possuem um grande volume de dados médicos, com destaque aos Prontuários Eletrônico de Paciente (PEP). O PEP se trata de uma ferramenta utilizada nessas instituições para armazenamento e controle digital das informações dos pacientes. Tal ferramenta permite reduzir erros, otimizar recursos, ampliar a segurança e aperfeiçoar o atendimento nas clínicas e demais organizações médicas (NEW et al., 2018). Este documento tem se mostrado um recurso fundamental nos avanços médicos, visto que podem ajudar pesquisadores a estimar progressões de doenças, prever potenciais epidemias e muito mais (NEW et al., 2018).

Com o surgimento da pandemia de COVID-19, a utilização de dados médicos para pesquisas e predições se tornou ainda mais relevante e necessária. Harrison et al. (2020) com uso de PEPs de 24 organizações de saúde nos Estados Unidos da América (EUA), encontraram associações entre comorbidades em pacientes e a taxa de mortalidade. Estiri et al. (2021) com o uso de prontuários eletrônicos criaram um método preditivo de mortalidade, no qual analisa características como comorbidades, sexo, etnia, entre outros. Portanto, é perceptível o proveito que os dados médicos, auxiliados com a pesquisa científica, podem trazer para a sociedade, desde contribuições científicas ao auxílio nas tomadas de decisões em instituições de saúde.

Contudo, a presença de informações sensíveis e pessoais nestes documentos dificultam o compartilhamento entre instituições devido à preocupação com a segurança em torno desses dados, assim como questões legais de privacidade do paciente, visto que a divulgação de seus dados pode ferir o direito à privacidade individual prevista na nova Lei Geral de Dados Pessoais (BRASIL, 2018a).

Segundo um levantamento da IBM-X, dados na área da saúde são uma das que mais sofrem com ataques, resultando em vazamentos, *phishing* ou até re-identificação (CHEN Y, 2019). No início de 2020, por exemplo, ocorreu o vazamento de cerca de 50 mil registros de pacientes no Minnesota Hospital. De acordo com SEH et al. (2020), no ano de 2019 foram 41 milhões de

usuários expostos de provedores de saúde, número que cresceu em 80% em comparação com 2015. Diante disso, pode ser observado que os sistemas de saúde estão suscetíveis a possíveis ataques de re-identificação, além de vazamentos de dados em geral.

1.1 Motivação e Escopo

Os registros referentes à área da saúde se tornaram objetos de muitos trabalhos científicos devido à importância e natureza de seus dados. Todavia, esses dados armazenados são sigilosos e de posse do paciente, portanto os sistemas de informação devem estar protegidos de invasores e intrusos, onde considera-se sempre as questões legais. (AL OMAR et al., 2017).

Motivado por riscos à segurança dos sistemas de saúde, o uso da tecnologia *Blockchain* se tornou recorrente na literatura e amplamente explorada devido ao seu alto nível de segurança na transação de informações. O uso dessa tecnologia permite a imutabilidade dos dados assim como a transferência segura entre as partes, onde apenas pessoas ou instituições autorizadas podem registrar ou acessar os dados armazenados (PETERSON et al., 2016).

Peterson et al. (2016) se aprofundaram no uso de uma estrutura descentralizada da *Blockchain* de modo a criar um sistema capaz de compartilhar informações entre instituições de forma segura. Em 2018, Yi Chen et al. (2018) desenvolveram uma arquitetura de armazenamento e gerenciamento de dados médicos pessoais com uso da tecnologia *Blockchain* junto ao armazenamento em nuvem. Tal arquitetura possui uma estrutura de serviço para compartilhamento onde nenhuma parte tem poder absoluto nos processos envolvidos.

Entretanto, mesmo que as transações em si sejam seguras, a identidade dos pacientes ainda pode ser transmitida entre os interessados, ferindo os princípios de privacidade e sigilo, o que se faz necessária a autorização individual de cada paciente. Porém, ainda que essa autorização pudesse ser realizada facilmente, representaria apenas uma pequena parcela do grande volume de dados mantidos pelas instituições de saúde (New et al. 2018). Os desafios da *Blockchain* com privacidade são classificados em dois tipos: (i) privacidade de transação; e (ii) privacidade de identidade. Sendo assim, é preciso estabelecer técnicas para minimizar os riscos com transações e identidade dos usuários que partilham seus dados (FENG et al. 2019).

A Comissão Nacional de Ética em Pesquisa (CONEP) prevê, a nível de conduta ética, a possibilidade de execução de pesquisas em dados de prontuários médicos desde que seja garantida a anonimização dos pacientes (CONSELHO NACIONAL DE SAÚDE, 2011). Nos termos do artigo 12 da Lei de Proteção dos Dados Pessoais (LGPD), os dados anonimizados estão excluídos do escopo de aplicação da lei (BRASIL, 2019). Portanto, uma forma de garantir a

privacidade do paciente e permitir o compartilhamento é utilizar técnicas e algoritmos de anonimização nos dados médicos.

A tarefa de anonimização é um procedimento complexo e requer a aplicação de algoritmos eficientes para processamento de textos, visto que grande parte dos dados médicos são apresentados como textos não-estruturados (MOHAMMED et al., 2009). Hassan et al. (2018) desenvolveram um reconhecedor de entidades nomeadas com base no aprendizado de máquina, capaz de detectar atributos que tenham implicações de privacidade, conhecidos como atributos sensíveis. CANEVAROLLO (2021) desenvolveu uma arquitetura para sistemas de prontuários eletrônicos de modo a utilizar algoritmos de Processamento de Linguagem Natural (PLN), mais especificamente a engenharia de características do Reconhecimento de Entidade Nomeadas (REN), para anonimizar dados não-estruturados antes de serem armazenados em uma rede *Blockchain*. Porém, em ambos os casos, a implementação desses algoritmos em bases existentes é complexa, visto que em alguns cenários é necessário modificar a estrutura do banco original para adição de novos atributos e/ou adaptação nas ferramentas de PLN para cada instituição.

1.2 Objetivo

Tendo em vista que para compartilhar dados médicos é necessário garantir a segurança nas transações, assim como preservar a privacidade dos pacientes, este trabalho tem como objetivo principal desenvolver um *framework* baseado na tecnologia *Blockchain* que preserve a privacidade dos pacientes no compartilhamento de dados médicos, de tal forma que reduza os riscos de vazamento de dados e ataques de re-identificação. Além disso, neste trabalho será implementado uma interface para que pacientes e instituições consigam visualizar e controlar o compartilhamento de seus dados na rede. Com isto, pretende-se como principal contribuição científica para este trabalho a elaboração e construção de um *framework* que permita a transferência e compartilhamento seguro de dados PEP com base na tecnologia Blockchain, visto que os trabalhos da literatura apenas propõem partes separada do *framework* sugerido, tornando totalmente dependente do desenvolvedor unir todos os elementos para construção de um ambiente de compartilhamento de dados médicos seguro.

Para este trabalho foi elaborado um estudo piloto para o compartilhamento de Prontuários Eletrônico de Paciente entre instituições de saúde e órgãos de pesquisa. Assim, os dados não-estruturados são armazenados na base de dados dos provedores de saúde, sendo transferidos, anonimizados e por fim, armazenados na rede *Blockchain*.

1.3 Metodologia

Para a realização deste trabalho, inicialmente, foi realizado um levantamento bibliográfico a respeito dos conceitos envolvidos no tema, como: Prontuário Eletrônico do Paciente, compartilhamento de dados médicos, *Blockchain*, algoritmos de anonimização, entre outros. Além disso, foi analisado também a Legislação Brasileira, no que diz respeito ao prontuário do paciente e leis de proteção de dados. Para tal, foram utilizadas como fontes artigos científicos disponibilizados na internet, trabalhos desenvolvidos no campus por meio de repositórios públicos e o site do Planalto, que fornece toda legislação de forma online.

Quanto à implementação, esta será iniciada por meio da elaboração e definição da arquitetura a ser utilizada, bem como os componentes necessários e suas relações. Em seguida, será desenvolvida uma rede baseada na tecnologia *Blockchain*, registrando as transações de dados entre os nós conectados. Nesta etapa será levado em consideração aspectos relacionados à segurança, como privacidade de transação e identidade, e ainda prevenção à ataques de re-identificação e vazamentos. Além disso, será avaliado também a utilização de protocolos de consenso e contratos inteligentes.

Posteriormente, será analisada a estrutura de uma base de dados de prontuários eletrônicos, disponibilizados por um hospital parceiro ao Grupo de Banco de Dados (GBD). A partir desta análise será possível verificar os requisitos de anonimização que o *framework* precisa atender.

Por fim, o *framework* será testado e avaliado de acordo com aspectos relacionados à segurança no compartilhamento e à capacidade de anonimizar dados sensíveis dos pacientes.

1.4 Organização da Monografia

Neste capítulo foram apresentados os problemas e dificuldades relacionados ao compartilhamento de dados médicos, além da motivação e o objetivo do trabalho.

Os próximos capítulos estão organizados da seguinte forma:

a) Capítulo 2 – Revisão bibliográfica: Conceitos relevantes sobre prontuários eletrônicos do paciente, compartilhamento de dados, privacidade, *Blockchain*, algoritmos e técnicas de anonimização e trabalhos correlatos;
b) Capítulo 3 – Desenvolvimento: Apresenta detalhes do desenvolvimento do *framework*, bem como sua arquitetura, fluxo e tecnologias;

c) Capítulo 4 – Avaliação experimental: Define como o trabalho foi avaliado e as discussões segundo os resultados obtidos dos experimentos;
d) Capítulo 5 – Conclusão: Neste, são apresentadas as conclusões do trabalho, as principais contribuições científicas e o direcionamento para possíveis trabalhos futuros relacionados;
e) Referências: Lista com todas as publicações científicas que foram utilizadas para embasar a confecção deste trabalho;

2 Fundamentação Teórica

Este capítulo tem como propósito apresentar os conceitos, técnicas e ferramentas que serão abordados nesta pesquisa e estarão presente nas próximas seções do trabalho.

2.1 Prontuário eletrônico do paciente

O Prontuário Eletrônico do Paciente (PEP), de acordo com o Conselho Federal de Medicina (CFM), no Artigo 1º da Resolução nº 1.638/2002, é um documento único constituído de um conjunto de informações, sinais e imagens registradas, geradas a partir de fatos, acontecimentos e situações sobre a saúde do paciente e a assistência a ele prestada, de caráter legal, sigiloso e científico, que possibilita a comunicação entre membros da equipe multiprofissional e a continuidade da assistência prestada ao indivíduo (CONSELHO FEDERAL DE MEDICINA, 2002).

De modo geral, os PEPs são documentos eletrônicos que fornecem dados abrangentes sobre históricos médicos, prescrições, assistências prestadas e tratamentos. Essas informações permitem que pesquisadores e cientistas estimem progressões de doenças, avaliem a eficácia de tratamentos e analisem o uso de recursos nos hospitais, além de gerar praticidade e aprimoramento no atendimento do paciente. É justamente na transmissão dessas informações que os PEPs se mostram valiosos, o que possibilita que diferentes médicos possam acompanhar o andamento de um paciente, beneficiando em tratamentos complexos ou prolongados.

No Brasil, o uso do PEP é regulamentado pela Lei 13.787, de 2018. A norma trata da digitalização e a utilização de sistemas informatizados para a guarda, o armazenamento e o manuseio de prontuário de paciente. A legislação diz ainda que os meios de armazenamento de prontuários eletrônicos devem proteger o documento do acesso, uso, reprodução e destruição não autorizadas (BRASIL, 2019). Ainda, como previsto na LGPD (BRASIL, 2018a), os responsáveis pelo PEP devem respeitar os direitos de privacidade dos pacientes, de modo a garantir que não haja

acesso por partes não autorizadas. Caso haja vazamento desses dados, os responsáveis pelo prontuário podem ser punidos com multas e outras sanções prevista na lei.

2.2 Compartilhamento de dados médicos

O compartilhamento de dados médicos, como os PEPs, possibilita a geração de conhecimento para o meio acadêmico, visto que instituições de pesquisas, laboratórios de análise ou até mesmo universidades, com uso desses dados, podem realizar pesquisas que tragam valor à sociedade como a elaboração de vacinas e medicamentos, prevenção de epidemias e a redução no uso de recursos médicos (NEW et al., 2018). Além disso, permite que o paciente ganhe economia de tempo e de burocracia, na medida em que não terá a incumbência de relatar seu histórico clínico todas as vezes em que se consultar com novos profissionais ou se submeter a novos procedimentos, podendo-se, inclusive, evitar que sejam repetidos exames já realizados anteriormente.

Entretanto, existem questões de suma importância e que devem ser observadas com cautela em relação ao compartilhamento destes dados. Tais documentos reúnem uma série de informações pessoais do paciente, a maioria considerados de caráter sensível pela LGPD (Lei Federal nº13.709/2018), os quais devem ser utilizados dentro das hipóteses autorizadas estabelecidas pela legislação, apenas para finalidades legítimas e informadas ao paciente, e de forma alguma sendo tratados de maneira discriminatória ou abusiva, em prejuízo do titular dos dados pessoais (BRASIL, 2018a).

A LGPD deve ser entendida como importante instrumento de controle dos titulares sobre o tratamento de seus dados pessoais, sendo fundamental que as organizações e os profissionais da saúde deem a devida transparência com relação às operações de tratamento de dados pessoais realizadas e as medidas de segurança adotadas para a proteção das informações. O PEP é sigiloso e pertence exclusivamente ao paciente, de modo que o seu compartilhamento entre profissionais ou instituições da saúde deve ocorrer sempre com todas as diligências necessárias para garantir a privacidade e proteção dos dados pessoais do titular envolvido, sendo observadas também as normas e regulações setoriais já existentes e que versam sobre o tema (BRASIL, 2018b).

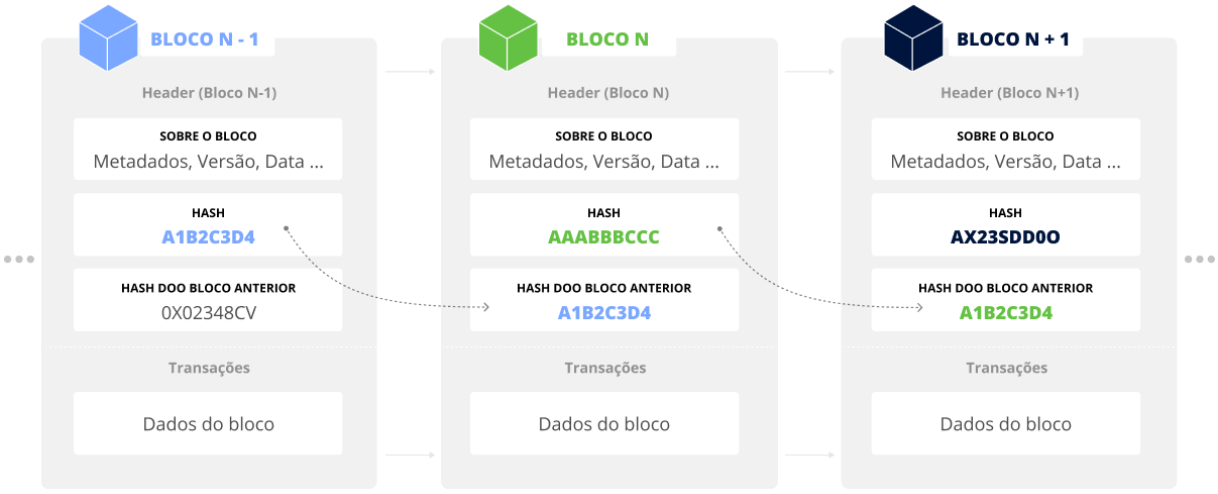
2.3 Tecnologia Blockchain

Tendo em vista as dificuldades com a segurança na transferência de dados médicos, a literatura científica apresenta uma tendência em utilizar a tecnologia *Blockchain* para a garantia de segurança e confiabilidade das informações médicas compartilhadas entre instituições.

A primeira arquitetura *Blockchain* surgiu com a criptomoeda Bitcoin, que possibilita aos usuários efetuar transações financeiras de forma pseudoanônima na internet, sem a necessidade de uma agência intermediadora. Porém, o uso desta tecnologia se ampliou para diferentes áreas e finalidades, sendo utilizada atualmente em sistemas de saúde, sistemas governamentais e até em jogos virtuais (TASATANATTAKOOL, 2018).

A tecnologia *Blockchain* se trata de uma estrutura de dados descentralizada que permite criar uma espécie de livro-razão digital que é compartilhado em uma rede *Peer to Peer* (P2P), onde os nós são independentes (LAURENCE, 2019). Para compor a estrutura, os blocos são ligados entre si, a partir de um ponteiro com o *hash* do bloco anterior, o que impede que os dados do bloco sejam adulterados, já que ao modificá-lo, ocorre a alteração do seu *hash* e, consequentemente, a alteração efetuada causa uma diferença no endereçamento para o *hash* do sucessor, que então acusa a modificação na estrutura. Os nós conectados à rede podem então facilmente detectar tal violação, podendo recusar e apagar esta alteração.

Figura 1 – Estrutura da Blockchain.



Fonte: Elaborado pelo autor

Na Figura 1 está representado a estrutura da *Blockchain.* Ao aplicar *hash* no bloco N, é obtido o valor “AAABBBCCC”. Quando um novo bloco, o N+1, entra na rede, ele é conectado ao

hash do bloco anterior (N) a partir de um ponteiro, que neste caso é igual ao *hash* do Bloco N ("AAABBBCCC"). Caso o conteúdo do bloco N seja alterado, um novo valor *hash* será gerado, tendo em vista que o ponteiro do Bloco N+1 continuará sendo "AAABBBCCC", a ligação entre os blocos será perdida, evidenciando a adulteração dos dados de um bloco.

2.3.1 Definições

Na *Blockchain*, o uso da técnica *hash* é fundamental para o fornecimento da segurança e conexão dos blocos. O *hash* é um valor gerado por uma função que recebe uma mensagem arbitrária de tamanho m e produz um valor $h = H(m)$ de tamanho fixo (STALLINGS, 2014). Além disso, pode ser uma função criptográfica de via única, ou seja, não existe uma função inversa. Dessa forma, se o valor original é alterado, o valor de saída da função *hash* muda completamente. Esta propriedade torna difícil encontrar um valor arbitrário que seja idêntico a um *hash* que já foi gerado (STALLINGS, 2014). O *hash* de um bloco é usualmente formado por todas as informações contidas em seu cabeçalho.

O bloco é o elemento principal da estrutura de dados utilizada na *Blockchain*, é nele que são armazenados o cabeçalho e as transações. O cabeçalho contém informações relevantes ao bloco como: a hora de criação, versão, metadados, *hash*, *hash* do bloco anterior e o nonce, que será explicado adiante. Já as transações, são registros da sequência de operações realizadas entre os clientes (ou usuários) de um sistema, como a troca de um ativo importante entre as partes (ZHENG et al. 2018). Todo novo bloco é ligado ao anterior por um ponteiro *hash* para tornar os dados imutáveis, impedindo possíveis fraudes.

O *Number Only Used Once (Nonce),* é um valor armazenado no cabeçalho de um bloco utilizado para a composição do seu valor de entrada na função *hash*. Esse valor é incrementado até que se encontre uma saída que represente o *hash* do bloco concatenado com bits zeros no início (WALDMAN, 2018). Em redes como o Bitcoin, o nonce é o valor que os mineradores procuram solucionar para que um novo bloco seja adicionado.

Em geral, sistemas que utilizam a tecnologia *Blockchain* são compostos por uma rede *Peer-to-Peer* (P2P) e um banco de dados descentralizados. Uma rede P2P é uma arquitetura de sistemas de computação distribuída, em que os nós (máquinas) conectados cooperam entre si, exercendo os papéis tanto de cliente quanto de servidor.

2.3.2 Características

A *Blockchain* pode ser formalmente definida como uma tecnologia de livro-razão distribuído, alicerçada em uma rede P2P e tendo como principal componente um livro que registra as transações de forma pública, ou privada, em ordem cronológica (MUKHOPADHYAY et al., 2016). Como apresentado na Figura 1, a *Blockchain* representa uma estrutura de dados em que cada bloco é ligado ao anterior por ponteiros previamente criptografados (*hash*) que seguem até o bloco gênesis (o primeiro bloco criado na rede). É dentro desses blocos que todas as transações são agrupadas.

Esta tecnologia possui algumas características que podem contribuir para potencializar os sistemas convencionais de saúde, sendo elas (GREVE et al., 2018):

- **Descentralização**: é uma importante característica, sendo um diferencial comparado a outras tecnologias na indústria. Aplicações convencionais necessitam que uma terceira parte (usualmente um servidor) faça a interação entre as transações dos usuários. Na *Blockchain* isso não acontece pois não há a necessidade de um nó central e confiável para realizar as operações, tudo é feito de maneira distribuída e descentralizada. Essa característica reduz problemas com a indisponibilidade de um nó central, uma vez que cópias dos dados estão totalmente disseminadas por toda a rede.
- **Transparência**: Todas as transações são registradas em um livro-razão público e são auditáveis pelos nós da rede. Entretanto, o nível de transparências depende totalmente da classificação de rede utilizada, podendo ser permissionada ou não-permissionada, e do desempenho da aplicação. Esta propriedade está diretamente associada ao ponteiro *hash* que interliga os blocos, tornando-os livre de fraude.
- **Imutabilidade**: As transações registradas no livro-razão não podem ser adulteradas, uma vez registradas não podem ser refutadas. Atualizações só são possíveis a partir da geração de novas transações e realizando um novo consenso.
- **Consenso**: é um mecanismo estabelecido entre os nós da rede para estabelecer um acordo sobre a condição de uma operação, para que seja validada ou rejeitada. Na *Blockchain* um dos algoritmos mais conhecidos e utilizados é o *Proof-of-Work* (PoW), baseado em falhas bizantino.
- **Confiança**: a *Blockchain* fornece aos usuários, maior confiança na troca de ativos, já que é imutável, auditável, transparente e faz uso de um protocolo para validar os blocos.
- **Contrato Inteligente**: os contratos inteligentes partem da ideia de propriedades inteligentes para a *Blockchain*. O objetivo é possibilitar aos sistemas algo a mais do

que as simples transações de trocas, estabelecendo acordos dinâmicos entras as partes. Este assunto será abordado com mais detalhes posteriormente.

2.3.3 Modelo de Rede

O modelo de rede da *Blockchain* pode ser caracterizado em três grupos, dependendo de como seus dados são protegidos. O modelo permissionado público consiste no controle do processo de ingresso na rede, de tal forma que apenas os nós previamente autorizados poderão ingressar na mesma. Para que seja estabelecido um consenso, os membros autenticados podem fazer parte do conjunto de nós validadores. A leitura dos dados pode ser feita pelos membros autenticados, ou ser restrita somente para nós especificados (ALHADHRAMI et al., 2017). No modelo permissionado privado os nós também necessitam ser autorizados e autenticados para ingressar na rede, a diferença se dá na responsabilidade de uma ou várias organizações confiáveis a função de adicionar novos blocos à cadeia, após a validação. O processo de leitura pode ser feito de forma aberta para os nós autenticados ou ser restrito para um conjunto confiável definido. Em contrapartida, no modelo não-permissionado não há a necessidade de autenticação, ou seja, o acesso a rede é totalmente aberto. Este modelo é composto por um livro-razão transparente, em que qualquer par da rede pode se tornar um minerador e tentar validar os blocos, como no caso do Bitcoin e do Ethereum (ALHADHRAMI et al., 2017).

Esta divisão em categorias da *Blockchain* possibilita a customização das aplicações, oferecendo diferentes regras de acesso de tal forma a sistematizar a organização dos modelos de rede com base nas necessidades de cada aplicação.

2.3.4 Mecanismo de Consenso

Visto que a *Blockchain* é um sistema computacional distribuído, é necessário existir confiança entre os pares da rede. Para isso são aplicados protocolos de consenso de forma a estabelecer a confiança na validação dos blocos que serão adicionados na cadeia. Estes protocolos utilizam dos próprios nós da rede como validadores e seguem uma regra pré-estabelecida para verificar se as informações trocadas são verdadeiras (CACHIN; VUKOLIC, 2017).

Alguns deste mecanismos de consenso utilizados nas plataformas são: *Proof-of-Work* (PoW), utilizado pelo Bitcoin; *Proof-of-Stake* (PoS), empregado no Cardano; *Tangle*, utilizado pelo IOTA. Entre os protocolos disponíveis, para esta pesquisa será empregado o protocolo baseado em Tolerâncias a Falhas Bizantinas Práticas (do inglês, *Pratical Byzantine Fault Tolerance*)

(PBFT), que visa mitigar o recebimento ou envio de mensagens não confiáveis (CACHIN; VUKOLIC, 2017).

Em termos gerais, este protocolo inicialmente elege um nó líder a partir do conjunto de pares da rede. Em seguida, uma solicitação é enviada ao líder para que ele registre a data e hora de entrada do cliente, então, uma mensagem é pré-preparada pelo cliente, na qual o nó líder registra a solicitação e passa para um próximo nó verificar se é válida, para posteriormente ser enviada aos pares da rede ou não. Se o nó aceitar a mensagem, é feita a transmissão para os outros pares verificarem a veracidade que, se aprovada, entra em estado de confirmação. Sendo f o limiar aceitável de nós que podem falhar, n é o número total de nós da rede, e $f < \frac{n}{3}$, o líder aguarda a confirmação dos outros pares, e se o número de mensagens confirmadas for $2 * f + 1$, o líder julga as mensagens como autênticas. Por fim, uma resposta é encaminhada para o cliente a partir do nó líder, indicando que a mensagem foi aprovada e está em estado de operação. Caso haja alguma falha na mensagem é feito o seu reenvio (MINGXIAO et al., 2017).

2.3.5 Contrato Inteligente

Os contratos inteligentes podem ser definidos como programas instalados na *Blockchain* que operam como um participante do sistema, facilitando, verificando e fortificando transações e acordos firmados entre duas ou mais partes (WHANG et at. 2019). Nesse sentido, um contrato inteligente é um código que pode definir regras estritas e consequências, da mesma forma que um documento tradicional, estabelecendo as obrigações, benefícios e penalidades que podem ser devidas a qualquer das partes, proporcionando confiabilidade nas relações entre os nós da rede (SZABO, 1997).

Os ativos e os termos do contrato são codificados e inseridos no bloco de uma rede *Blockchain*, este contrato é então distribuído e copiado várias vezes entre os nós da plataforma. Após o desencadeamento do processo, o contrato é executado de acordo com os termos nele contido, o programa verifica a implementação dos compromissos de forma automática (SZABO, 1997).

Na área da saúde, o prontuário médico é de propriedade do paciente, que tem total direito de acesso e pode solicitar cópia. Ao médico e ao estabelecimento de saúde cabe sua a elaboração e a guarda do documento (Conselho Federal de Medicina, 2009). Portanto, apesar de os direitos de propriedade ser do paciente, quem faz o uso e armazena são as instituições de saúde. Dessa forma, com uso de contratos inteligentes, é possível que a instituição compartilhe tal informação com outro médico ou entidade desde que haja a permissão e autorização prévia do paciente.

2.3.6 Plataforma Hyperledger Fabric

Junto aos avanços da tecnologia *Blockchain* surgem plataformas e ferramentas que facilitam a construção dessas redes, auxiliando no desenvolvimento de redes customizadas e com contratos inteligentes. Umas das primeiras que surgiu foi a Ethereum, que é baseada na infraestrutura da moeda Ether e possibilita a construção de contratos inteligentes de forma simplificada, interligando-os a uma aplicação específica. A *International Business Machines* (IBM), em parceria com *a Linux Foundantion*, criou o *Hyperledger Fabric* (HLF), viabilizando o desenvolvimento de redes permissionadas e contratos customizados.

O HLF provê uma *Blockchain* permissionada que trabalha com diversos protocolos de consenso, como o PBFT. A ferramenta também apresenta uma estrutura com *broadcast* atômico, execução determinística e estado persistente em todos os pares da rede de forma sincronizada.

2.4 Anonimização

Por mais que a *Blockchain* e seus contratos inteligentes possam garantir a confiabilidade e segurança no compartilhamento, essas informações podem ser repassadas para outras instituições sem o conhecimento do proprietário desses dados, o que resultaria em um descaso a privacidade do paciente, visto que os dados contidos nos prontuários carregam informações sensíveis do indivíduo. Como discutido anteriormente, uma forma de solucionar este problema consiste no processo de anonimização de qualquer tipo de informação pessoal presente nestes documentos. Dessa forma, caso não haja o consentimento prévio do paciente, os dados serão anonimizados antes de serem transmitidos.

2.4.1 Análise de texto não-estruturado

Prontuários médicos, assim como muitos outros documentos da área da saúde, são, em sua grande parte, textuais e não-padronizados, o que dificulta a análise probabilística feita pelos algoritmos de anonimização. Com objetivo de tornar os computadores capazes de entender a linguagem informal e natural de comunicação do ser humano, ferramentas de PLN foram desenvolvidas, tornando possível desidentificar um indivíduo com base em entidades de interesse de um texto, como nome próprio, número de documento e outros (CHOWDHURY, 2003).

Os sistemas de PLN permitem que a tecnologia usada não apenas entenda o significado literal de cada palavra, mas também seja capaz de considerar aspectos como: contexto, significado

sintático e semântico, interpretação de texto e análise de sentimentos. Para lidar com toda a complexidade de interpretação de texto livre, sistemas de PLN podem fazer uso de aprendizado de máquina, ou *deep learning*. Por meio dessa funcionalidade, os sistemas aprendem a cada iteração e refinam sua capacidade de reconhecimento e interpretação.

2.4.2 Reconhecimento de entidades nomeadas

O REN é uma subárea de estudo do campo de extração de informação, cujo objetivo é identificar entidades nomeadas, bem como classificá-las dentro de um conjunto de categorias pré-definidas, tais como nome de pessoas, organizações ou local (AMARAL, 2013). Este é uma técnica amplamente utilizada no PLN e consiste na identificação de nomes de entidades-chave, presentes na forma livre de dados textuais. Nesse sentido, a entrada para um sistema de extração de entidades nomeadas é um texto em sua forma livre, e sua saída é um conjunto de textos anotados, ou seja, uma representação estruturada a partir de uma entrada de texto não-estruturada.

2.5 Trabalhos Correlatos

Esta seção tem o propósito de investigar a literatura e discutir a respeito do compartilhamento de dados médico. Foram coletados trabalhos com dois focos principais, sendo eles o compartilhamento seguro de informações da saúde e a preservação da privacidade nos ambientes baseados em *Blockchain*.

2.5.1 Compartilhamento seguro de informações da saúde

O MedRec é uma estrutura baseada em *Blockchain* para armazenamento de registros médicos eletrônicos, projetado por AZARIA et al. (2016). Este trabalho tem como finalidade resolver alguns problemas como o acesso lento aos dados e interoperabilidade, além de promover uma melhor disponibilização dos dados em pesquisas médicas. A proposta tem como base uma rede P2P privada, bem como a integração com contratos inteligentes por meios da plataforma Ethereum, a partir do consenso não permissionado. Desta forma, é possível gerenciar e rastrear as transições de estados dos ativos na rede. Essa abordagem fornece ao paciente uma agência para consulta de todo seu histórico médico, deixando-o informado sobre as decisões médicas. Porém, uma preocupação da proposta é sobre a privacidade dos pacientes, visto que os autores não consideram de forma sistêmica esse aspecto.

ZHENG et al. (2018), propuseram um *framework* de compartilhamento de dados pessoais de saúde baseado em *Blockchain*, armazenamento em nuvem e aprendizado de máquina. O sistema permite que os usuários possam controlar e compartilhar seus dados pessoais de saúde com facilidade e segurança. O foco do trabalho é permitir que dados de saúde gerados pelos usuários, de forma contínua (como dispositivos gerados a partir do advento da Internet das Coisas), possam ser compartilhados entre os provedores, tornando o acesso e divulgação de controle total do usuário. Entretanto, os autores não consideram o compartilhamento desses dados de forma anonimizada com instituições de pesquisa, tornando o uso desses dados apenas comercial.

2.5.2 Preservação da privacidade nos ambientes baseados em Blockchain

O trabalho de CANEVAROLLO (2021), propôs uma arquitetura para desenvolvimento de sistemas médicos seguros e robustos. A arquitetura leva em consideração a interoperabilidade entre plataformas conectadas a uma rede segura e à prova de fraudes com uso da tecnologia *Blockchain*. Além disso, o trabalho leva em consideração a privacidade dos indivíduos envolvidos nas transações, aplicando algoritmos de anonimização, tanto para dados estruturados quanto para não-estruturados. Entretanto, o desenvolvimento das estratégias propostas fica totalmente dependente dos desenvolvedores, visto que o trabalho não oferece um *framework* capaz de encapsular todos os itens comuns da arquitetura.

MediBChain é uma plataforma que emprega criptografia de Curvas Elípticas para garantir a privacidade dos pacientes. O foco da proposta é o gerenciamento de dados centrado no paciente, no qual somente ele consegue disponibilizá-los para outros provedores. Todos os registros médicos transitados pela rede são armazenados seguindo o esquema de nuvem, mas em um banco de dados baseado na *Blockchain* (AL OMAR et al., 2019). Este trabalho traz como principais contribuições: (I) a garantia de segurança e privacidade fundamentada em responsabilidade, pseudônimo, autenticidade e integridade; (II) uma *Blockchain* permissionada para restringir usuários desconhecidos; (III) dados centrados no controle do paciente; (IV) esquema de curvas elípticas para prover maior segurança e pseudônimos aos pacientes. Apesar da proposta oferecer diversas vantagens, ainda possui limitações. A plataforma não é interoperável entre outras organizações de saúde, além disso, está suscetível ao roubo ou perda de chaves, que pode causar a inutilidade de dados por conta do paciente ou médico. Por fim, com o uso de pseudonimização, os pacientes ainda estão sujeitos a ataques de ligação para inferir atributos reais com base em um conhecimento prévio sobre o paciente (AL OMAR et al., 2019).

2.5.3 Comparação dos trabalhos correlatos

Na Tabela 1 é apresentado um comparativo entre os principais trabalhos correlatos anteriormente apresentados, sendo avaliadas as características necessárias para criação de sistemas gerenciadores de prontuários eletrônicos seguros e confiáveis.

Apesar da literatura oferecer soluções viáveis para compartilhamento seguro de dados da saúde, poucos se atentam a fatores relacionados à privacidade individual do paciente. Além disso, alguns não consideram a possibilidade de compartilhar dados anonimizados junto a instituições de pesquisa. CANEVAROLLO (2021), apesar de propor uma solução próxima ao desejado, não oferece um *framework* capaz de encapsular todos os componentes em comum da arquitetura, tornando as estratégias de implementação totalmente dependente dos desenvolvedores. Além disso, também não oferece nenhum tipo de interface para o usuário final interagir com a plataforma e se comunicar com os outros membros da rede.

Tabela 1 - Comparativo entre os trabalhos correlatos

	(AZARIA et al., 2016)	(ZHENG et al., 2018)	(AL OMAR et al., 2019)	(CANEVAROLLO, 2021)
Prove o compartilhamento seguro de dados	✔	✔	✔	✔
Implementa uma rede Blockchain	✔	✔	✔	✔
Gerenciador de prontuários eletrônicos	✔	✘	✔	✔
Anonimiza dados estruturados e não-estruturados	✘	✘	✘	✔
Disponibiliza um framework	✘	✔	✘	✘
Implementa interface para usuário	✘	✘	✘	✘

Fonte: Elaborado pelo autor.

2.6 Considerações finais

Neste capítulo foram apresentados os principais conceitos que fundamentam este trabalho. Entre os temas abordados, podemos destacar a importância na preservação da privacidade dos pacientes, no compartilhamento de dados médicos, assim como a garantia de segurança na transferência e armazenamento das informações. Além disso, foi discutido as características que tornam a *Blockchain* uma tecnologia capaz de sanar grande parte dos problemas existentes na transmissão de dados da saúde. Por fim, foi abordado o uso de métodos de anonimização de dados para permitir que instituições de saúde possam fazer proveito dessas informações, possibilitando uma maior facilidade e confiança nos resultados de pesquisas científicas, bem como benefícios diretos a sociedade.

3 Desenvolvimento

Neste capítulo é apresentado o processo de desenvolvimento do *framework baseado na tecnologia Blockchain,* capaz de criar sistemas gerenciadores de prontuários com suporte ao compartilhamento seguro e preservação da privacidade dos pacientes.

3.1 Framework proposto

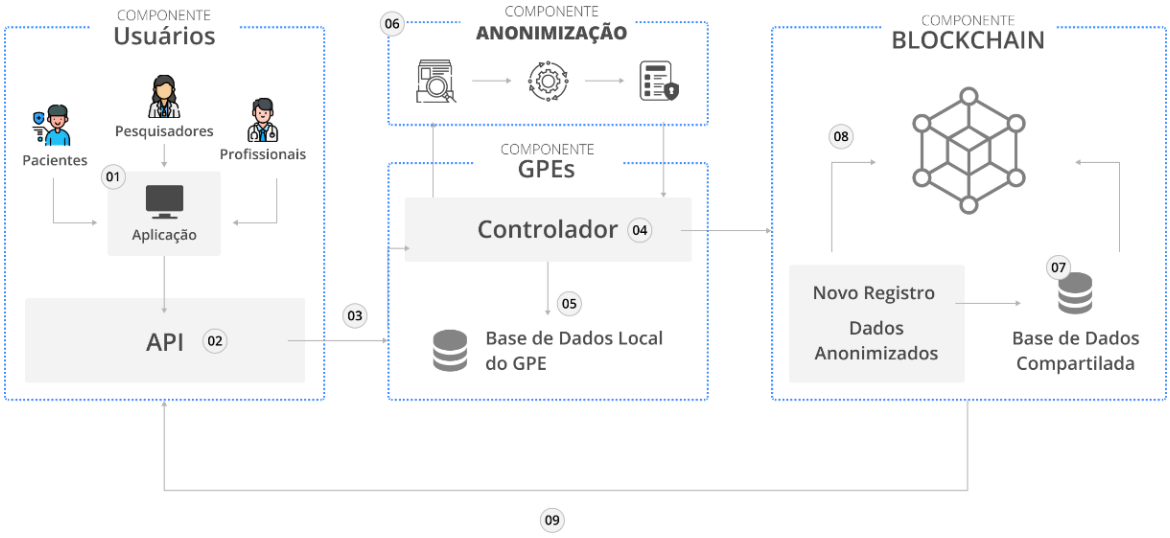
Esta pesquisa propõe um *framework* que suporta a tecnologia *Blockchain*, ou seja, oferece uma estrutura de rede de compartilhamento de dados médicos com o propósito de oferecer a privacidade dos pacientes e protegê-los de ataques e vazamentos. Ao se considerar o escopo do trabalho, não é viável que qualquer nó possa participar da rede sem uma autorização prévia. Sendo assim, a rede será desenvolvida com base em uma *Blockchain* permissionada, dessa maneira apenas computadores de organizações autorizadas podem participar das validações das transações. O *framework* é dividido em 4 principais componentes: componente do usuário; componente da rede *blockchain;* componente de privacidade; e o componente do Gerenciador de Prontuários Eletrônicos (GPE). A Figura 2 descreve a visão geral do *framework*, como seus componentes e seu fluxo. Cada um dos componentes do *framework* atua para fornecer privacidade e rastreabilidade nos dados compartilhados, assim, nas próximas seções, tais componentes serão apresentados de forma detalhada.

3.2 Descrição dos componentes do framework

No componente do usuário, o objetivo é fornecer uma área de interação entre os usuários da rede, que seja de fácil integração em diferentes aplicações, que favorece a utilização do *framework*. Sua construção será feita com base na linguagem *JavaScript* com o *Express*, a partir de métodos baseados no padrão *Representational State Transfer* (REST), que facilita o desenvolvimento de outras aplicações que pretendem consumir dados da rede. Além disso, a *Application Programming Interface* (API) também facilitará o acesso à rede *Blockchain*, uma vez que utiliza o SDK do *Hyperledger Fabric*, que proporciona métodos para acessar os atributos dos contratos inteligentes. Ainda, para este trabalho, será desenvolvido uma aplicação que se

comunica com a API e disponha de uma interface amigável para a interação dos nós da rede, denominada HealthSHARE.

Figura 2 – Componentes do Framework



Fonte: Elaborado pelo autor

O componente do provedor de saúde, os GPEs, representam o principal componente da arquitetura, são eles que possuem a responsabilidade dos fluxos de dados na rede, além de possuírem uma base própria para armazenamento dos dados gerados por seus usuários e pacientes. O objetivo deste componente é estabelecer a comunicação dos usuários finais com a rede *Blockchain* e os provedores de saúde que armazenam os dados dos pacientes. Este componente possui três principais fragmentos: (I) base de dados local que armazena os PEPs dos pacientes; (II) o controlador, que gerencia o tráfego das requisições e transações; e o (III) conector para o componente de privacidade. Este componente verifica novas requisições, captura os dados das respectivas bases locais e os envia para serem anonimizados através do conector, que comunica com o componente de privacidade. Após anonimizados, os dados são enviados de volta para o controlador, que os encaminha pelo canal de comunicação para a rede *Blockchain*. Para implementação deste componente será utilizado a linguagem *JavaScript*, que será utilizado para criação do controlador de requisições do GPE.

O componente de privacidade é responsável por prover a privacidade aos usuários quando seus dados são compartilhados, visto que ele anonimiza os dados localmente antes de disponibilizá-los para pesquisadores ou profissionais da saúde. Para manter os dados protegidos serão aplicados os mesmos métodos que foram propostos por CANEVAROLLO (2021), com a

utilização de processamento de linguagem natural e reconhecimento de entidades nomeadas. Além disso, será mantido o padrão de definição dos tipos de dados da base do GPE estabelecidos pelo autor. O componente será desenvolvido tanto em *JavaScript* como em *Python*, sendo utilizados, respectivamente, para criação da API de comunicação e para o desenvolvimento e execução dos *scripts* de anonimização.

Por fim, o componente da *Blockchain* será responsável por armazenar os logs, bem como o registro das transações na rede, fazendo intersecção com a base de dados compartilhada, que armazena os dados anonimizados. Para construção da rede será utilizado o *Hyperledger Fabric*. Também será definido contratos inteligentes que contém as regras de negócio para armazenar e recuperar os dados, seguindo as políticas de acesso e de privacidade.

3.3 Descrição do fluxo de dados

Os fluxos de dados trafegados no *framework* estão representados na Figura 2 pelos números de 1 a 9. A seguir será apresentado como cada fluxo faz a interação entre os componentes da rede:

1. **Aplicação do usuário**: este fragmento faz parte do componente do usuário e tem como principal objetivo fornecer um meio de comunicação amigável entre os usuários da rede, como profissionais de saúde, pesquisadores e pacientes.
2. **Interface de comunicação**: é uma API REST que é responsável por fazer a comunicação com os outros componentes da rede.
3. **Canal de envio de dados**: é o canal em que os dados são enviados e transmitidos dos usuários aos demais componentes. Esse fragmento observa a chegada de novas requisições pelo canal de comunicação e notifica os demais componentes.
4. **Controlador**: faz parte do componente do provedor de saúde e é executado localmente. Este fragmento tem o objetivo de observar as requisições encaminhadas pelo canal de comunicação, além de coletar os dados solicitados da base local. Por fim, este realiza o envio e recebimento de dados que o componente de privacidade necessita tratar.
5. **Base de dados local**: representa a base de dados, na qual o provedor de saúde armazena os dados de seus pacientes, bem como seus prontuários eletrônicos.
6. **Privacidade dos dados**: é o componente que aplica privacidade aos dados dos prontuários antes de serem compartilhados, utilizando os modelos proposto e desenvolvidos por CANEVAROLLO (2021). Vale ressaltar que este processo é

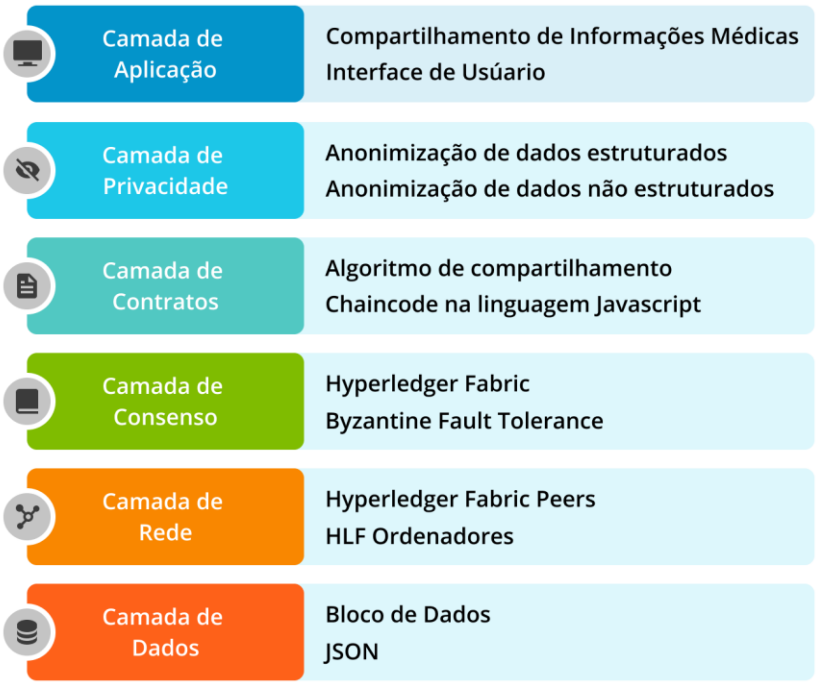
realizado localmente em segundo plano, de tal forma que o processo de anonimização não interfira no funcionamento do sistema principal do provedor de saúde.

7. **Banco de dados compartilhado**: é a rede que compõe os repositórios para os dados anonimizados compartilhados. A rede *off-chain* também permite que os dados possam ser removidos no futuro caso o paciente deseje, essa característica auxilia no cumprimento dos requisitos de controle de acesso da LGPD (BRASIL, 2018a).
8. **Registro na *Blockchain*:** os registros das transações e logs de acesso são inseridos na rede *Blockchain*. Esses fragmentos são imutáveis e auxiliam na prestação de contas por parte das organizações pelos dados acessados.
9. **Canal de troca de informações compartilhadas:** aguarda os dados cujo processo foi solicitado, para assim, enviá-los para a aplicação que os usuários interagem. Caso o dado solicitado já esteja anonimizado, o processo de entrega é mais rápido, visto que grande parte do custo computacional se faz pelo passo de anonimização.

3.4 Arquitetura

A arquitetura da *Blockchain* pode ser organizada por camadas, como apresentado na Figura 3, sendo elas: (I) a camada de aplicação, interage diretamente com os usuários a partir da API, que se comunica com uma aplicação, no caso deste trabalho será desenvolvido uma interface para a área da saúde; (II) a camada de privacidade, descreve o processo de anonimizar os dados e a identidade dos usuários que se comunicam com a camada de contrato; (III) a camada de contrato representa as regras de negócio definidas nos contratos inteligentes para moldar a lógica das transações; (IV) camada de consenso define o protocolo de consenso utilizado para validar as transações na rede, que neste trabalho, foi adotado o método BFT; (V) camada de rede representa a rede *Blockchain*, que define como os dados serão encaminhados e verificados na rede P2P. Em conjunto com a camada de rede, porém fora da *Blockchain*, está a base de dados compartilhada; (VI) camada de dados é composta pelos atributos do bloco que serão incluídos no livro-razão.

Figura 3 –Arquitetura do framework



Fonte: Elaborado pelo autor

3.5 Processo de anonimização

O processo de anonimização adotado para o *framework* tem como referência o modelo sugerido por CANEVAROLLO (2021). A primeira etapa deste processo consiste no pré-processamento do texto, onde algumas tratativas são feitas, como a remoção de *stopwords*, palavras que geralmente não agregam valor semântico ao texto. Além disso é removido também *tags* de linguagem de Marcação de Hipertexto (HTML, do inglês *Hypertext Markup Language*). Posteriormente, os dados sensíveis são identificados utilizando ferramentas de PLN e técnicas de marcações, que identificam entidades em um texto e sua respectiva função em uma sentença. Uma vez identificadas, as entidades podem ser removidas ou substituídas por dados falsos que mantenham o valor semântico.

3.6 Controle de acesso

Com o propósito de proteger os dados dos pacientes, políticas de privacidade e níveis de acesso foram estabelecidos na construção da rede *Blockchain* privada. Tais políticas possibilitam

mitigar o acesso indevido aos dados e controlar quem acessa determinado ativo. Na Tabela 2 são descritos os métodos de escrita e leitura aos componentes do *framework* que apresentam uma visão geral das políticas de acesso, visando na rede, banco compartilhado e banco local. O controle de acesso é definido para os usuários do tipo paciente, pesquisador, profissionais da saúde e provedores de saúde.

Tabela 2 – Políticas de controle de acesso

	Banco Compartilhado	Banco Local	Blockchain
Paciente	Leitura	-	Leitura e Atualização
Pesquisador	Leitura	-	Leitura
Profissional	Leitura	Leitura e Escrita	Leitura
Provedor	Leitura e Escrita	Leitura e Escrita	Leitura e Escrita

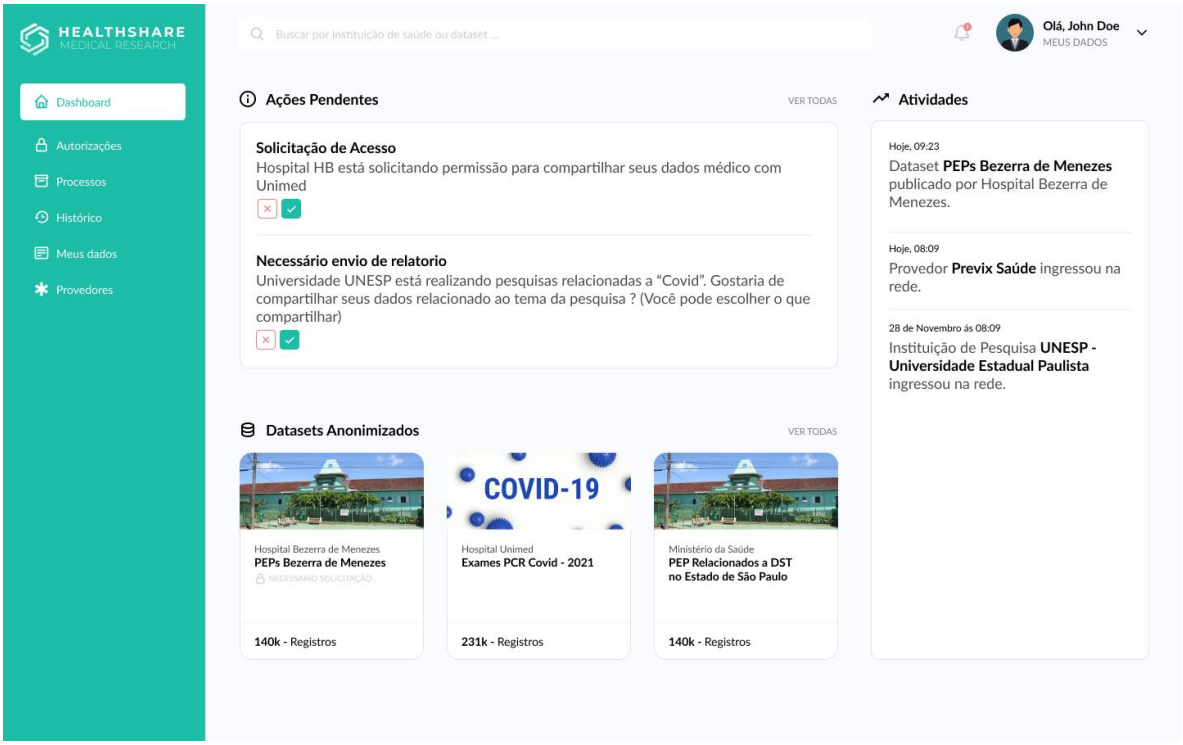
Fonte: Elaborado pelo autor

Os provedores de saúde podem ser considerados como usuários, visto que são organizações confiáveis que trocam dados no *framework*. Os profissionais de saúde e pesquisadores podem acessar os dados anonimizados na base compartilhada, porém o profissional da saúde pode também ler na base local, dependendo da relação e acordos com o provedor. A *Blockchain* é acessada pelos provedores de saúde que são portadores dos dados dos pacientes, podendo então executar ações de leitura e escrita. O paciente pode fazer a leitura e atualização dos dados na *Blockchain*, com o objetivo de aprovar o compartilhamento de seus dados. A base de dados compartilhada é acessada por todos os usuários do *framework*. Por fim, o provedor de saúde pode acessar a base de dados local, que representa os PEPs dos pacientes, armazenado em um hospital, e que posteriormente, poderá ser compartilhado com um pesquisador ou profissional da saúde.

3.7 Interface do usuário

Com o intuito de facilitar a interação entre os pares da rede, foi desenvolvido uma aplicação que fornece uma interface moderna para os usuários, nomeada de HealthSHARE. Esta aplicação faz comunicação direta com a API, que por sua vez, se relaciona com os demais componentes do *framework*. Na Figura 4 é apresentado a tela inicial da aplicação, nesta tela o usuário pode verificar ações pendentes, dados médicos disponibilizados recentemente e as atividades da rede.

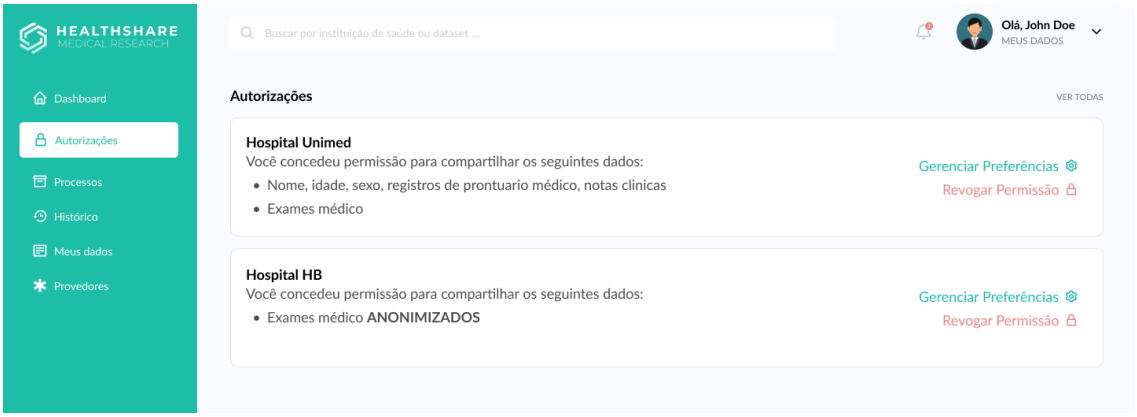
Figura 4 – Página inicial da aplicação HealthSHARE



Fonte: Elaborado pelo autor

Na Figura 5 é ilustrado a página de autorizações e permissões do usuário, nessa tela são listadas todas a permissões concedidas aos provedores de saúde, além de ser exibido também quais dados estão sendo coletados e a forma de coleta. Ainda nessa página é possível revogar o acesso aos dados compartilhado e alterar quais dados o usuário deseja compartilhar.

Figura 5 – Página de autorizações da aplicação HealthSHARE



Fonte: Elaborado pelo autor

3.8 Considerações Finais

Neste capítulo foi abordado sobre o desenvolvimento do *framework* proposto, abordando sobre seus componentes, divididos em usuário, GPE, anonimização e *Blockchain*. Além disso, foi ilustrado e explicado a arquitetura do *framework*, dividida em 6 camadas, descrevendo as propriedades e funções de cada uma delas. Também, foi discorrido sobre as características do controle de acesso aos dados da *Blockchain* e do processo de anonimização. Por fim, foi apresentado a aplicação desenvolvida para que os usuários possam interagir na rede, onde fornece uma interface direta e simples, que permite a visualização e controle sobre os dados fornecidos a provedores de saúde, bem como a possibilidade de solicitar dados e consultar bases públicas já anonimizadas.

4 Avaliação experimental

Neste capítulo é apresentado a aplicação do *framework* proposto em um cenário de simulação e define os experimentos que foram conduzidos, bem como seus respectivos resultados. Nestes experimentos foram analisados o custo computacional para execução da *Blockchain*, assim como seu impacto na rede. Além disso, também foi avaliado o impacto causado para o módulo do provedor de saúde. Por fim, é descrito o cenário de aplicação, a avaliação dos requisitos de privacidade e as discussões sobre os resultados.

4.1 Metodologia e configurações dos experimentos

Estes experimentos têm como objetivo simular um ambiente real de aplicação do *framework*, de tal modo que seja possível validar seu funcionamento como um todo, desde o compartilhamento seguro dos dados na *Blockchain* até a correta anonimização dos dados originados do GPE.

Para realização dos testes com a *Blockchain*, foram utilizadas máquinas com as seguintes configurações: 16 GB de memória RAM, processador Intel® Core ™ i5-11400F, 480 GB de SDD e sistema operacional Linux Ubuntu 20.0.4. O processo de avaliação foi seguido utilizando o Hyperledger Caliper, que é uma ferramenta de *Benchmark* de *Blockchain*, que permite medir o desempenho de uma implementação com um conjunto de casos de uso pré-definidos.

4.2 Descrição dos dados

Os dados manipulados nos experimentos foram extraídos de um hospital parceiro do Grupo de Banco de Dados (GBD), de tal modo que a simulação pudesse ser mais próxima de um cenário real. No estudo foi considerado tanto dados estruturados, referente a dados de pacientes e usuários cadastrados no GEP, como dados não estruturados, que compõem dados textuais livre como notas clínicas e documentos. A base de dados utilizada possui 140.113 notas clínicas e 48.110 cadastros de pacientes.

4.3 Avaliação do framework

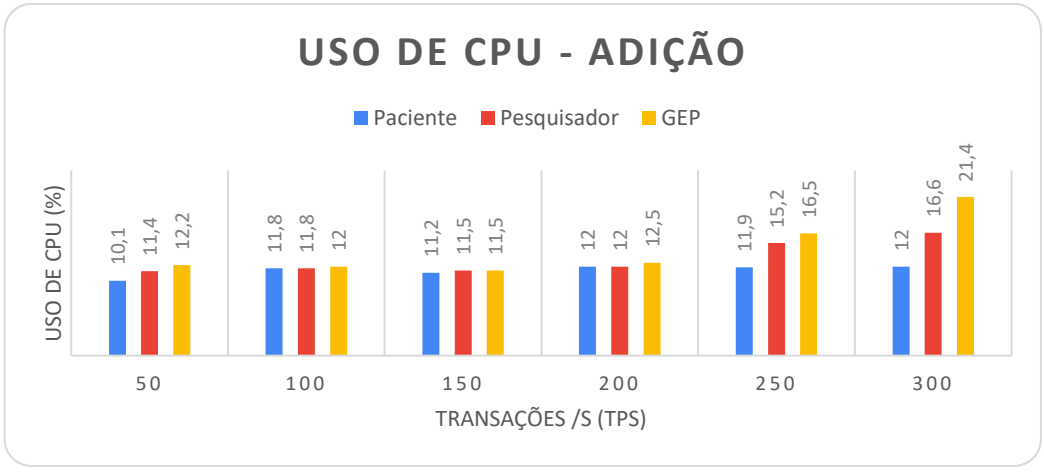
A avaliação do desempenho do *framework* foi realizada utilizando como base duas perspectivas, a da *Blockchain* e a do sistema gerenciador de prontuários, analisando e mensurando o uso de hardware e rede de cada um. Para esta simulação, cada uma das rodadas de experimento foi executada 10 vezes, com 5 rodadas no total. Os gráficos e tabelas a seguir descrevem e expressam os valores das simulações.

4.3.1 Desempenho da blockchain

A avaliação de desempenho do componente da *Blockchain* tem como principal finalidade analisar a capacidade das máquinas no emprego desta tecnologia, podendo então, verificar a viabilidade e possibilidade de aplicação em um ambiente real de saúde. Para auxiliar neste processo de análise, foi utilizado a ferramenta Hyperledger Caliper, capaz de mensurar a *Blockchain* por completo ou por partes. Para esta análise foram consideradas e levantadas as seguintes métricas: latência, vazão, transação por segundo (TPS), uso de CPU e uso de memória. Durante a simulação, os valores de TPS foram variados em: 50, 100, 150, 200, 250 e 300.

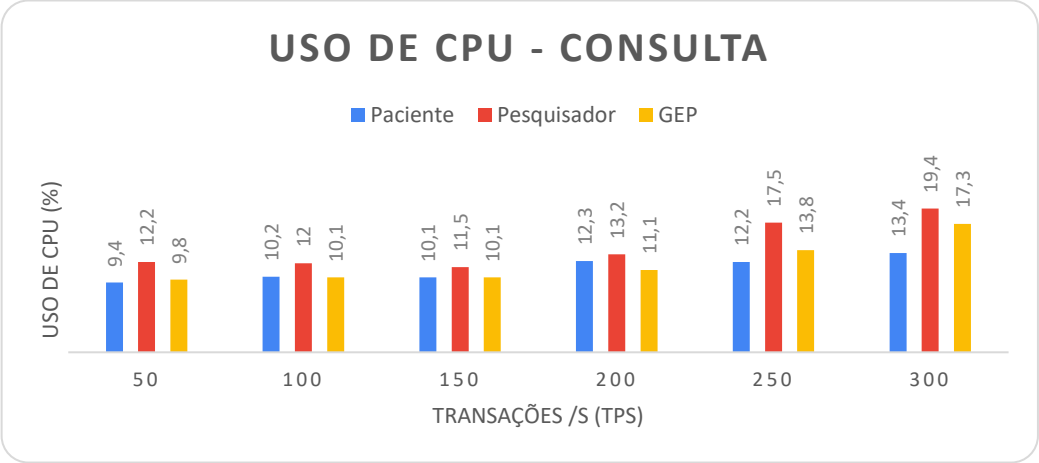
Inicialmente, foi analisado o consumo de hardware das máquinas durante a execução dos experimentos, considerando o uso de CPU e o uso de memória. Neste processo, foi realizado simulações de adição e consulta de registros na rede *Blockchain*, utilizando o Hyperledger Caliper, apoiado pelos métodos aplicados nos contratos inteligentes. As Figuras 6 e 7 descrevem o uso de CPU médio de cada operação durante as execuções das simulações, onde o eixo x representa as transações por segundo (TPS) e o eixo y descreve o percentual de uso da CPU.

Figura 6 – Uso de CPU com adição de registros na blockchain



Fonte: Elaborado pelo autor

Figura 7 – Uso de CPU com consultas na blockchain

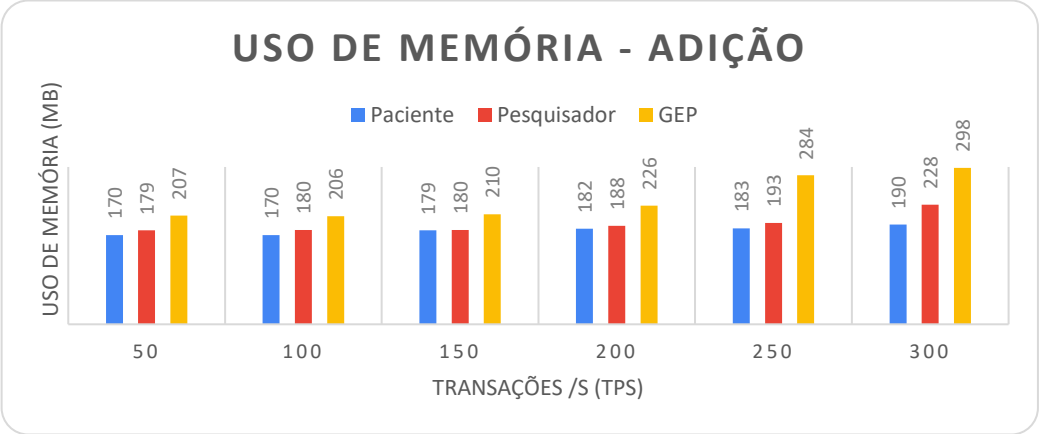


Fonte: Elaborado pelo autor

Na operação de adição, representada na Figura 6, os resultados revelam que o GEP é a organização que consome mais CPU. Já na operação de consulta, o maior custo de consumo de CPU passa para o pesquisador, tendo em vista que interagem constantemente com a *Blockchain* para recuperar dados e solicitar permissões. Como complemento para a análise de desempenho da *Blockchain*, foi avaliado o consumo de memória pelos componentes da estrutura, representados nas Figura 8 e 9. Podemos observar que em ambos os casos, adição e consulta, o GEP é a organização que mais consome memória, uma vez que essa organização faz paralelo com o componente de anonimização e com o controlador. De forma geral, as métricas que correspondem ao uso de CPU e memória consumidos pela *Blockchain* retratam um comportamento esperado, visto que existe um aumento proporcional no consumo de CPU ou memória em relação ao aumento do TPS.

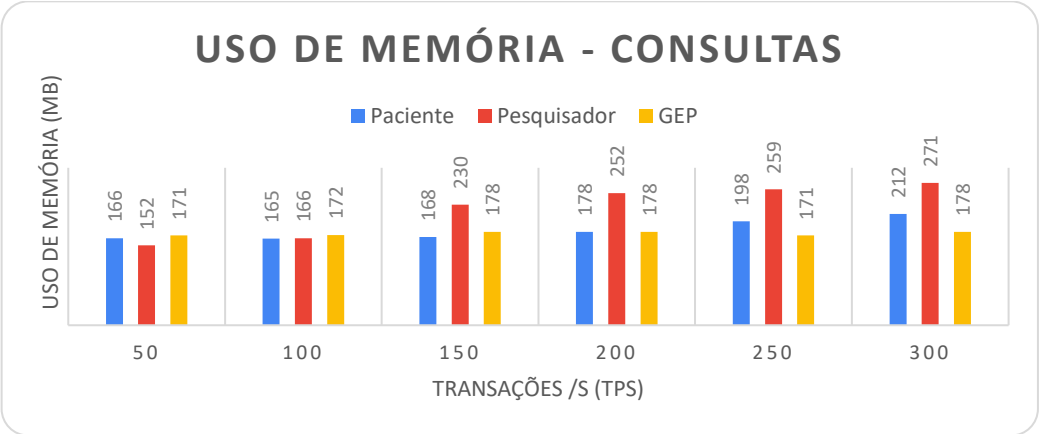
Em seguida foi avaliado o desempenho da *Blockchain* analisando a carga gerada na rede, mensurado com base nas métricas de latência e vazão (*Throughput*), utilizando também a ferramenta Hyperledger Caliper. Na Figura 10 é ilustrado os resultados das simulações para cada uma das métricas, nas operações de adição e consulta dos dados, que são representados pelas barras no gráfico.

Figura 8 – Uso de memória com adição de registros na blockchain



Fonte: Elaborado pelo autor

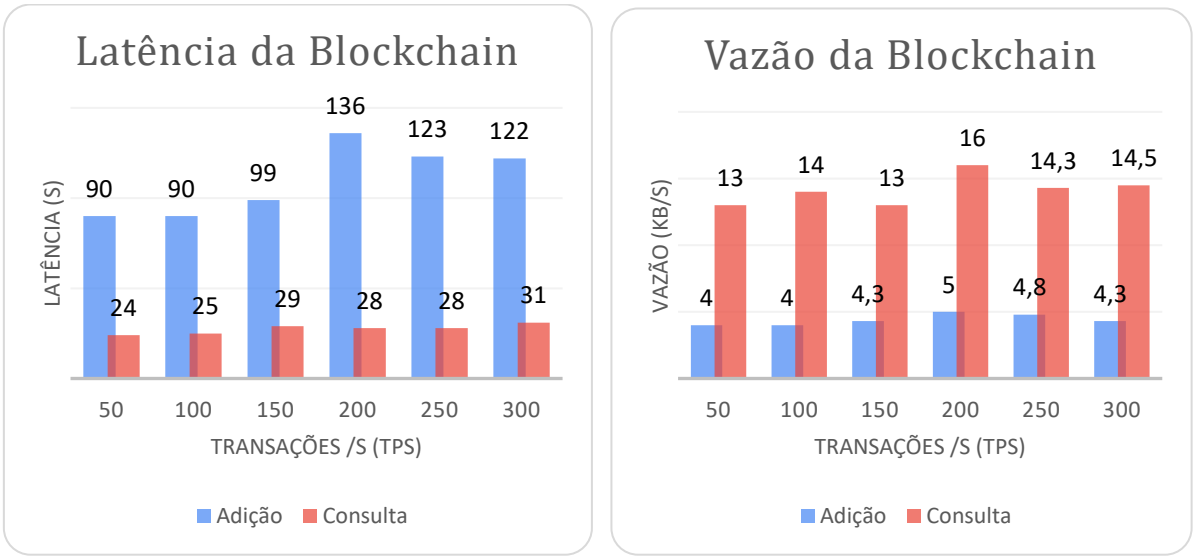
Figura 9 – Uso de memória com consultas na blockchain



Fonte: Elaborado pelo autor

Em suma, as métricas de latência e vazão apresentam um comportamento anormal. Primeiramente, é esperado que os valores de vazão fossem diretamente proporcionais aos valores de TPS, contudo, os valores variam, por exemplo, para a consulta a maior vazão tem TPS igual a 200. Este problema está relacionado com a definição das políticas de endosso, pois impactam no desempenho da rede. Essas políticas definem quais nós executam as transações e devem ser projetadas com as regras de negócios para não causarem gargalo na rede. Portanto, para esta pesquisa não foi considerado o projeto das políticas, assim, gerando um gargalo maior nos experimentos.

Figura 10 –Vazão e latência da blockchain



Fonte: Elaborado pelo autor

4.3.2 Desempenho do módulo de privacidade

Como descrito no Capítulo 3, o *framework* proposto possui um módulo de anonimização, que se relaciona com a *Blockchain*. A comunicação entre esses módulos acontece na rede do provedor de saúde que recupera os dados de suas bases e os enviam para anonimização. Desse modo, com a finalidade de avaliar a carga que o módulo de anonimização gera ao sistema de saúde, foram executadas simulações e testes analisando o uso de CPU, consumo de memória, tempo de execução e acurácia dos resultados. Para esta simulação, foram utilizadas três bases diferentes contendo 50.000, 100.000 e 140.000 notas clínicas. O processo foi executado 10 vezes para cada uma das bases. Como pode ser observado na Tabela 3, de forma geral, o consumo de recursos e o tempo de anonimização aumenta proporcionalmente à quantidade de notas clínicas processadas pelo módulo. Analisando do ponto de vista do pior caso, o processo de anonimização levou uma média aproximada de 364 segundos para sua finalização, apresentando um consumo médio de memória de 489 Mb e uma acurácia média de 94,33%.

Tabela 3 – Análise de desempenho do algoritmo de anonimização

Notas Clínicas	Uso médio de CPU (%)	Uso médio de Memória (Mb)	Tempo para finalização (s)	Acurácia do resultado (%)
50.000	13	348	96	97,11
100.000	14	421	252	95,58
140.000	14	489	364	94,33

Fonte: Elaborado pelo autor

4.3.3 Análise de privacidade

Para esta análise, foi tomada como foco a variável condição do sistema - essa variável foi examinada a partir dos princípios definidos pela LGPD, na qual estipula uma lista com os 10 princípios que organizações e sistemas devem atender para estar em conformidade com a lei (BRASIL, 2018a). Com o levantamento desses princípios é discutido como o *framework* proposto auxilia a aplicar cada um deles, sendo:

1. **Finalidade e Adequação:** este princípio estipula que um sistema ou organização não pode recolher dados sem apresentar uma finalidade específica ao titular dos dados. Quando o paciente da organização é registrado no sistema GPE, todos os dados coletados e as finalidades são apresentados.
2. **Necessidade:** a coleta e utilização dos dados devem se restringir ao mínimo necessário para a realização das finalidades apresentadas. O *framework* proposto apenas compartilha dados anonimizados com pesquisadores. Quando o compartilhamento é feito entre organizações, na sua forma original, o consentimento do titular é necessário. O *framework* não se estende a nenhuma tarefa que não seja o compartilhamento seguro e consensual dos dados.
3. **Acesso livre:** facilidade de acesso do titular à forma que os dados são tratados. O *framework* possui uma aplicação que permite os usuários interagirem na *Blockchain*. Nessa aplicação é possível identificar quais dados foram coletados por quais organizações.
4. **Qualidade:** os dados devem ser atualizados e condizentes, segundo a real necessidade no tratamento. As informações podem ser atualizadas a qualquer momento pelo paciente na aplicação, mantendo a qualidade e consistência dos dados.
5. **Transparência:** este princípio visa a garantia, aos titulares, de informações claras, precisas e facilmente acessíveis sobre a realização do tratamento e os respectivos agentes de tratamento. Todas as transações e processos na rede são registrados na *Blockchain*, que podem ser futuramente consultadas pelo titular através da aplicação ou API.
6. **Segurança:** necessário coibir situações acidentais ou ilícitas como invasão, destruição, perda ou difusão. A *Blockchain* impede que qualquer registro seja deletado ou modificado. Além disso, o *framework* implementa protocolos de segurança na comunicação entre os pares da rede, dificultando possíveis interceptações. Por fim, visto que para ingressar na rede é necessário consentimento dos pares, o processo de invasão se torna muito mais complexo e difícil.

7. **Prevenção:** este princípio versa sobre o ato de estar preparado para lidar com eventuais problemas envolvendo o tratamento de dados pessoais antes mesmo que eles surjam. Como todas as transações na rede são registradas na *Blockchain*, é possível identificar e responsabilizar organizações ou pesquisadores que descumpram a lei.
8. **Não discriminação:** impossibilidade de realização do tratamento para fins discriminatórios ilícitos ou abusivos. Como já discutido, todos dados coletados e tratados pelo *framework* são para fins puramente científico. O uso indevido pode ser facilmente identificado.
9. **Responsabilização:** este princípio dispõe sobre o cumprimento da lei tendo em vista provas e evidências de que medidas e procedimentos foram tomados pela empresa a fim de garantir a proteção dos dados. O *framework* proposto apresenta diversos elementos que asseguram a privacidade e proteção dos dados armazenados pelo GPE.

4.4 Discussão

Os resultados correspondentes ao desempenho da *Blockchain* se mostraram aceitáveis, indicando a possibilidade da adição prática da *Blockchain* a sistemas gerenciadores de prontuários médicos, visto que o *framework* foi capaz de executar a simulação com um custo máximo de 21% de CPU e 298 MB de memória. Já em relação ao desempenho de rede, deve se considerar que possui uma carga considerável para latência e uma baixa vazão, tais características ocorrem em virtude do processo de consenso e da aplicação dos contratos inteligentes. No entanto, essa carga ainda é aceitável para a maioria dos casos, principalmente para sistemas que não necessitam do processamento em tempo real. O custo computacional gerado pelo módulo de privacidade é considerável, porém estratégias podem ser adotadas para amenizar problemas de concorrência de recursos do computador. Uma das soluções possíveis é realizar as tarefas de anonimização apenas quando existe folga nos recursos disponível, visto que o *framework* possibilita implementar e configurar uma fila de tarefas. Outra possível solução seria isolar o módulo de anonimização do sistema gerenciador de prontuário, de tal forma que o provedor de saúde tenha uma máquina exclusiva para realização destes tratamentos.

A LGPD estipula 10 princípios que as organizações e sistemas devem obedecer quanto ao tratamento dos dados. Com a finalidade de validar se o *framework* atendia os aspectos de segurança e privacidade impostos pela LGPD, foi analisado como cada um desses princípios se comportam no *framework* proposto. Após essa análise, é possível verificar que o *framework*

auxilia por completo ou parcialmente no cumprimento de todos os princípios estabelecidos pela LGPD, possibilitando tornar os sistemas de saúde mais seguros e aptos a seguir as novas políticas de segurança e privacidade.

5 Conclusão

As informações médicas derivadas de Prontuários Eletrônicos do Paciente (PEP) se mostraram um recurso essencial para os avanços da ciência nas mais amplas áreas. Entretanto, devido a características pessoais e sensíveis que esses dados possuem, estratégias de segurança digital devem ser levadas em consideração para garantir a proteção dos sistemas de invasores ou intrusos, além da preocupação em preservar a privacidade dos indivíduos.

Neste trabalho foi proposto e apresentado um *framework* baseado na tecnologia *Blockchain* para preservação da privacidade no compartilhamento de dados de saúde, em que foi abordado um estudo piloto para o compartilhamento de prontuários eletrônicos. O *framework* disponibiliza um componente de privacidade capaz de identificar e desassociar dados sensíveis do paciente presentes nos prontuários e notas clínicas, possibilitando o compartilhamento com instituições de pesquisa e favorecendo no avanço da ciência. O *framework* permite a fácil integração de outros módulos, além das modificações dos já existentes, visto que são baseados em componentes e utilizam a rede do Hyperledger Fabric, que permite a construção de contratos inteligentes customizados.

Na prática, o *framework* proposto pode ser integrado a um sistema real de saúde, além de formar uma rede confiável para o compartilhamento desses dados, assegurando a privacidade dos pacientes que partilham suas informações para pesquisa. No entanto, vale ressaltar que o desenvolvimento do *framework* se encontra em um estado inicial, necessitando de refinamentos, adaptações e mais testes para garantir a viabilidade em mais cenários.

5.1 Contribuições

Os resultados obtidos por meio dos experimentos realizados indicam que o *framework* é capaz de trazer segurança e privacidade no compartilhamento de dados médicos. Na Tabela 4 são apresentados os atributos da proposta desenvolvida em comparação com os principais trabalhos correlatos encontrados na literatura.

Tabela 4 - Comparativo entre os trabalhos correlatos e este trabalho

	(AZARIA et al., 2016)	(ZHENG et al., 2018)	(AL OMAR et al., 2019)	(CANEVAROLLO, 2021)	Este trabalho
Prove o compartilhamento seguro de dados	✔	✔	✔	✔	✔
Implementa uma rede Blockchain	✔	✔	✔	✔	✔
Gerenciador de prontuários eletrônicos	✔	✘	✔	✔	✔
Anonimiza dados estruturados e não-estruturados	✘	✘	✘	✔	✔
Disponibiliza um framework	✘	✔	✘	✘	✔
Implementa interface para usuário	✘	✘	✘	✘	✔

Fonte: Elaborado pelo autor.

Como pode ser observado, o *framework*, em conjunto com os componentes apresentados no Capítulo 3, é a principal contribuição científica deste trabalho, permitindo preservar a privacidade dos pacientes em ambientes de saúde com base na tecnologia *Blockchain*, para troca segura de informações sobre dados médicos. Além disso, este trabalho implementa uma interface para que o usuário posso interagir com a rede *Blockchain*, permitindo verificar as transações, acompanhar solicitações e atualizar ou remover documentos, tornando a rede mais interativa e transparente para os usuários.

5.2 Trabalhos futuros

Com o desenvolvimento deste trabalho, algumas possíveis melhorias foram avaliadas e que podem ser aplicadas ao *framework* de modo a contribuir com o estado da arte.

A primeira a destacar é a extensão do *framework* para possibilitar o compartilhamento de outros formatos de dados, tais como imagens, vídeos e sons, permitindo a transferência de todos os tipos de exames médicos.

Além disso, outro passo importante é analisar o cumprimento do *framework* em frente a *General Data Protection Regulation* (GDPR), visto que foi levado em consideração apenas a LGPD para a construção do *framework*.

Outro ponto, é a adição de uma extensão mais robusta e confiável das ferramentas de PLN, para que seja possível que estas consigam identificar mais termos sensíveis relacionadas à área da medicina.

Também, com a finalidade de testar e validar o *framework*, experimentos em diferentes cenários e configurações de hardware podem ser aplicados, possibilitando a coleta de informações mais concretas sobre a viabilidade em diferentes ambientes de saúde, assim como à aplicação de uma avaliação heurística para análise de usabilidade da interface desenvolvida.

Por fim, com o intuito de amenizar problemas de latência ou vazão, faz-se necessário a definição das políticas de endosso na rede *Blockchain*.

Referências Bibliográficas

ALHADHRAMI, Zainab et al. Introducing blockchains for healthcare. In: **2017 international conference on electrical and computing technologies and applications (ICECTA)**. IEEE, 2017. p. 1-4.

AL OMAR, Abdullah et al. MediBChain: A blockchain based privacy preserving platform for healthcare data. In: **International conference on security, privacy and anonymity in computation, communication and storage**. 2017. p. 534-543.

AL OMAR, Abdullah et al. Privacy-friendly platform for healthcare data in cloud based on blockchain environment. **Future generation computer systems**, v. 95, p. 511-521, 2019.

AMARAL, Daniela Oliveira Ferreira do. **O reconhecimento de entidades nomeadas por meio de conditional random fields para a língua portuguesa**. 2013. Dissertação de Mestrado. Pontifícia Universidade Católica do Rio Grande do Sul.

AZARIA, Asaph et al. Medrec: Using blockchain for medical data access and permission management. In: **2016 2nd international conference on open and big data** (OBD). IEEE, 2016. p. 25-30.

BRASIL G. F. do, Lei Geral de Proteção de Dados Pessoais. 2018a. **Diário Oficial da União. 15 ago 2018, Seção I**. Disponível em <https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm> Acesso em 20 de junho de 2021.

BRASIL, G. F. do, Digitalização e a utilização de sistemas informatizados para a guarda, o armazenamento e o manuseio de prontuário de paciente. 2018b. **Diário Oficial da União. 28 dez 2018, Seção I, Página 3**. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13787.htm> Acesso em 09 novembro 2021.

CACHIN, Christian; VUKOLIĆ, Marko. Blockchain consensus protocols in the wild. **arXiv preprint arXiv:1707.01873**, 2017.

CANEVAROLLO, Douglas Armando Cabrelli. **Blockchain como tecnologia para o gerenciamento de prontuários médicos eletrônicos com foco ao compartilhamento de dados e garantia de privacidade**. TCC (Bacharelado em Ciência da Computação) – Universidade Estadual Paulista (Unesp), Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto. 2021.

CHEN, Qian. Toward realizing self-protecting healthcare information systems: Design and security challenges. In: **Advances in Computers**. Elsevier, 2019. p. 113-149.

CHEN, Yi et al. Blockchain-based medical records secure storage and medical service framework. **Journal of medical systems**, v. 43, n. 1, p. 1-9, 2019.

CHOWDHURY, Gobinda G. Natural language processing. **Annual review of information science and technology**, v. 37, n. 1, p. 51-89, 2003.

CONSELHO FEDERAL DE MEDICINA, Resolução N.1.638. **Define prontuário médico e torna obrigatória a criação da Comissão de Revisão de Prontuários nas instituições de saúde**. 2002. Disponível em <https://sistemas.cfm.org.br/normas/visualizar/resolucoes/BR/2002/1638> Acesso em 18 setembro 2021.

CONSELHO NACIONAL DE SAÚDE. Comissão Nacional de Ética em Pesquisa. **Circular nº 093/2011, de 30 de setembro de 2011. Uso de dados de prontuários para fins de Pesquisa.** Brasília, DF, 30 set. 2011

ESTIRI, Hossein et al. Predicting COVID-19 mortality with electronic medical records. **NPJ digital medicine**, v. 4, n. 1, p. 1-10, 2021.

FENG, Qi et al. A survey on privacy protection in blockchain system. **Journal of Network and Computer Applications**, v. 126, p. 45-58, 2019.Garfinkel, S. L. De-identification of personal information. [S.I], 2015

GREVE, Fabíola et al. Blockchain e a Revolução do Consenso sob Demanda. **Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC)-Minicursos**, 2018.

HARRISON, Stephanie L. et al. Comorbidities associated with mortality in 31,461 adults with COVID-19 in the United States: A federated electronic medical record analysis. **PLoS medicine**, v. 17, n. 9, p. e1003321, 2020.

HASSAN, Fadi; DOMINGO-FERRER, Josep; SORIA-COMAS, Jordi. Anonymization of unstructured data via named-entity recognition. In: **International conference on modeling decisions for artificial intelligence**. Springer, Cham, 2018. p. 296-305.

JANI, Shailak. Smart Contracts: Building Blocks for Digital Transformation. **Indira Gandhi National Open University**, 2020.

LAURENCE, T. **Blockchain for dummies**. 2ª ed. São Paulo: John Wiley & Sons, 2019. 248 p

MINGXIAO, Du et al. A review on consensus algorithm of blockchain. In: **2017 IEEE international conference on systems, man, and cybernetics (SMC)**. IEEE, 2017. p. 2567-2572.

MOHAMMED, Noman et al. Anonymizing healthcare data: a case study on the blood transfusion service. In: **Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining**. 2009. p. 1285-1294.

MUKHOPADHYAY, Ujan et al. A brief survey of cryptocurrency systems. In: **2016 14th annual conference on privacy, security and trust (PST)**. IEEE, 2016. p. 745-752.

PETERSON, K. et al. A blockchain-based approach to health information exchange networks. Paper presented at: Proceedings of the NIST Workshop Blockchain Healthcare; vol. 1, 2016: 1-10.

SEH, Adil Hussain et al. Healthcare data breaches: Insights and implications. In: **Healthcare**. Multidisciplinary Digital Publishing Institute, 2020. p. 133.

STALLINGS, William; BRESSAN, Graça; BARBOSA, Akio. **Criptografia e segurança de redes.** Pearson Educacion, 2008.

SZABO, Nick. Formalizing and securing relationships on public networks. **First monday**, 1997.

TASATANATTAKOOL, Pinyaphat; TECHAPANUPREEDA, Chian. Blockchain: Challenges and applications. In: **2018 International Conference on Information Networking (ICOIN)**. IEEE, 2018. p. 473-475.

WALDMAN, Jonathan. Blockchain – Conceitos básicos do Blockchain. **MICROSOFT**. Disponível em: < https://docs.microsoft.com/pt-br/archive/msdn-magazine/2018/march/blockchain-blockchain-fundamentals >. Acesso em 20 de jun. de 2018.

WANG, Xiaonan et al. Blockchain-based smart contract for energy demand management. **Energy Procedia**, v. 158, p. 2719-2724, 2019.

ZHENG, Zibin et al. Blockchain challenges and opportunities: A survey. **International Journal of Web and Grid Services**, v. 14, n. 4, p. 352-375, 2018.

ZHENG, Xiaochen et al. Blockchain-based personal health data sharing system using cloud storage. In: **2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom)**. IEEE, 2018. p. 1-6.