Uma Arquitetura Baseada em Blockchain para Auditoria de Conformidade com Regulamentos de Proteção de Dados

Marcos Maciel de Castro[1], **Marciel Barros Pereira**[1], **Miguel Franklin de Castro**[1]

[1]Mestrado e Doutorado em Ciência da Computação (MDCC)
Universidade Federal do Ceará (UFC)
Fortaleza – CE – Brasil

marcosmc@gmail.com, marciel@crateus.ufc.br, miguel@ufc.br

Abstract. *The emergence of personal data protection regulations, such as the GDPR in the European Union and the LGPD in Brazil, motivates organizations that process this data to seek an auditing solution to demonstrate that they follow applicable legal practices. However, it is still not clear which is the best technique to provide such a solution. This work presents an ongoing research to design a blockchain-based architecture to audit operations performed on personal data in a reliable, transparent, and cost-effective way.*

Resumo. *O surgimento de regulamentos de proteção de dados pessoais, como a GDPR na União Europeia e a LGPD no Brasil, motiva organizações que processam esses dados a buscarem uma solução de auditoria para evidenciar que seguem as práticas legais aplicáveis. Entretanto, ainda não está claro qual a melhor técnica para prover tal solução. Este trabalho apresenta uma pesquisa em andamento para concepção de uma arquitetura, baseada em blockchain, para auditar operações de tratamento realizadas sobre dados pessoais de modo confiável, transparente e com boa relação custo-benefício.*

1. Introdução

Nos últimos anos, vazamentos de dados pessoais tornaram-se frequentes, causando danos morais ou materiais de difícil reparação. Nesse cenário, a União Europeia (UE) decidiu aprovar, em 2016, o seu *regulamento para proteção de pessoas naturais em relação ao processamento de dados pessoais e ao livre movimento de tais dados*, intitulado General Protection Data Regulation (GDPR) [GDPR 2016], em vigor desde 2018. No Brasil, por sua vez, foi sancionada a Lei Geral de Proteção de Dados Pessoais (LGPD) [LGPD 2018] em 2018, influenciada pelas diretrizes presentes na GDPR. Tais dispositivos visam esclarecer regras, requisitos e instruções para implantação de controles pelas organizações, porém, para proteger a privacidade das pessoas, as novas obrigações exigem mudanças consideráveis em processos e sistemas [Tikkinen-Piri et al. 2018].

Por outro lado, a adoção crescente e massiva do paradigma de computação em nuvem tem favorecido o armazenamento de uma enorme quantidade de dados pessoais de usuários em *data centers* espalhados pelo mundo [Bilal et al. 2018]. A grande quantidade de mecanismos de coleta de dados pessoais e sua transferência para processamento ou armazenamento em nuvem também preocupa os usuários em relação à segurança dos procedimentos realizados com seus dados [Hossein et al. 2019].

Nesse âmbito, os negócios impactados pela GDPR passaram a ter de se preocupar tanto com conformidade quanto com processamento, armazenamento e compartilhamento

dos dados de usuários em nuvem [Russo et al. 2018]. Uma vez disponibilizados pelos usuários, passa a ser responsabilidade dos provedores de serviço demonstrar a conformidade legal das práticas sobre os dados pessoais sob sua custódia. Entretanto, ainda não está claro qual a técnica mais adequada para que a indústria forneça soluções transparentes, eficientes e com uma boa relação custo-benefício [Wu et al. 2019]. Além disso, boa parte das soluções utilizadas atualmente para auditoria se baseiam em arquitetura cliente-servidor centralizada, com mecanismos de demonstração de conformidade que acabam deixando dúvidas quanto à lisura e transparência de suas operações [Truong et al. 2020].

O presente trabalho descreve uma pesquisa em andamento para propor uma arquitetura capaz de auditar as operações de tratamento realizadas sobre dados pessoais de modo confiável, transparente e com custo razoável. A GDPR e a LGPD serão utilizadas como referência, e um bom glossário de termos comuns a esses normativos pode ser encontrado no artigo 5º da [LGPD 2018]. Além disso, os mecanismos abordados na arquitetura apresentada podem ser estendidos a outros dispositivos legais.

2. Trabalhos Relacionados

É proposta em [Wu et al. 2019] uma arquitetura baseada em *blockchain* pública, em que um mecanismo baseado em *state channels* e *sticky policies* é utilizado para melhorar o *throughput* e a escalabilidade da plataforma e possibilitar a verificação de informações *off-chain*. Todavia, não é definido o mecanismo de punição por comportamento errático dos pares em *state channels* e apresenta maior *overhead* devido à exigência da execução de agentes para interpretar as *sticky policies* nos nós por onde transitam os dados.

[Truong et al. 2020] propõe uma arquitetura descentralizada em duas *blockchains*: uma para um sistema de autenticação, autorização e controle de acesso, e outra para um sistema de *logging* e validação de *tokens* de acesso. Os dados pessoais são armazenados *off-chain*, para conformidade com a GDPR e melhor escalabilidade e eficiência do sistema. A implementação é realizada sobre uma *blockchain* com autorização (*permissioned*), mas não é especificado o mecanismo de consenso e o *log* das operações é realizado diretamente na *blockchain*, podendo causar problemas de *throughput* caso se utilize uma *blockchain* pública e sem autorização (*permitionless*), como é o caso da *Ethereum*.

[Barati et al. 2022] propõe o uso de agentes implantados em contêineres para registrar as operações com os dados dos usuários através de requisições a um serviço *Representational State Transfer* (REST), enviando-as, em seguida, a um motor de coleta. Para aliviar a carga do sistema e melhorar a escalabilidade da solução, essas operações são filtradas de acordo com um conjunto de preferências definidas pelo titular dos dados através de contratos inteligentes, que também são utilizados para encaminhar os registros a uma *blockchain* e executar as verificações de conformidade, permitindo a detecção de violações à legislação de proteção de dados sob demanda. Todavia, o armazenamento de operações diretamente na *blockchain* aumenta o custo e diminui o *throughput* do sistema.

No contexto de Internet das Coisas (IoT) em *e-health*, [Hossein et al. 2019] propõe uma arquitetura de controle de acesso a dados para resguardar a privacidade de pacientes de um sistema de saúde. Visando escalabilidade, é proposta uma abordagem de clusterização de mineradores e a redução do tamanho das transações, para diminuir o *overhead* sobre a rede. São utilizadas duas *blockchains*: uma para registrar os *hashes* dos dados e outra para armazenar as políticas de acesso. A estratégia de clusterização de

mineradores, no entanto, não é adequada para uso em *blockchains* públicas.

Uma arquitetura não baseada *blockchain* é proposta por [Kunz et al. 2020], cuja prioridade é tratar a questão da localização geográfica dos dados e de suas réplicas. É apresentado um sistema para rastreamento de fluxo de dados baseado na descoberta de rótulos previamente atribuídos e metadados que são utilizados para executar validações em relação a políticas pré-estabelecidas. No entanto, a solução não apresenta um mecanismo de auditoria e é pouco adequada ao tratamento do que não seja arquivo.

3. Concepção da Arquitetura

Com base nos trabalhos relacionados, esta pesquisa propõe uma arquitetura em alto nível para auditoria de conformidade com normas de uso de dados pessoais, adequada ao uso com *blockchains* públicas e sem autorização (*permitionless*) e visando custo razoável sem comprometer o *throughput* da solução. Uma visão geral é representada na Figura 1, cujos elementos estão descritos a seguir.

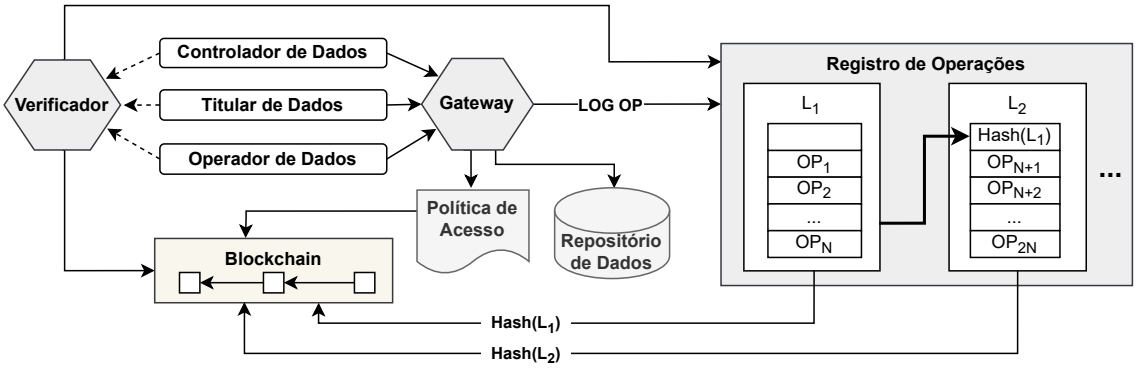

Figura 1. Representação da arquitetura proposta nesta pesquisa.

Titular, Controlador e Operador de Dados: Correspondem aos mesmos papéis definidos na GDPR e na LGPD, e representam as entidades que fornecem os seus dados pessoais ou os manipulam. Controlador e Operador são também chamados Agentes de Tratamento.
Política de Acesso: Um conjunto de regras indicando que um certo Agente de Tratamento – *quem* – pode acessar um dado pessoal de um certo titular – *o quê* –, para certa finalidade – *porquê* – e durante certo período – *quando*, além da base legal. Na ausência de uma regra, o dado simplesmente não pode ser acessado pelo Agente de Tratamento.
***Blockchain*:** Os Agentes de Tratamento acordam com o Titular a Política de Acesso, que é, então, armazenada neste registro distribuído e imutável. Além disso, armazena atestados de não conformidades nas operações com dados pessoais.
Repositório de Dados: Repositório no qual os dados pessoais são armazenados de forma independente dos Agentes de Tratamento. O acesso aos dados então armazenados é feito apenas através do *Gateway* e mediante autenticação e autorização de acordo a Política de Acesso. Esse repositório deve permitir a alteração e a exclusão de informações, prevendo a necessidade de retificação ou de eliminação de dados do Titular.
Registro de Operações: Serviço de armazenamento das operações realizadas sobre os dados pessoais dos titulares. Cada item indica a operação realizada, a sua finalidade, o *timestamp*, a Política de Acesso e a identificação do agente de tratamento. Esse registro não armazena diretamente nenhum dado pessoal e nem permite alterações ou exclusões,

sendo que apenas o *Gateway* pode solicitar a inclusão de novos itens. Periodicamente ou após certo número de inserções, é calculado um *hash* que será armazenado na *Blockchain*.
***Gateway*:** Componente responsável por receber os dados pessoais, encaminhá-los para armazenamento no Repositório de Dados, conceder ou negar acesso dos agentes de tratamento aos dados com base na Política de Acesso armazenada na *Blockchain* e registrar todas essas operações no Registro de Operações. Pode ser visto como o intermediário entre o Titular, os Agentes de Tratamento e o Repositório de Dados.
Verificador: Agente que realiza, periodicamente ou sob demanda, uma verificação de conformidade dos itens do Registro de Operações do Agente de Tratamento sobre os dados do titular, de acordo com a correspondente Política de Acesso vigente para o momento da operação. Um registro é, então, gravado na *Blockchain*, indicando a presença ou ausência de não conformidades nessas operações. O trabalho do verificador é atestar que o mecanismo de autenticação e autorização do *Gateway* e as operações dos Agentes de Tratamento não violaram as regras da Política de Acesso.

Para realizar o tratamento de dados pessoais, um agente deve satisfazer aos requisitos determinados pela legislação aplicável. Por exemplo, a LGPD apresenta dez bases legais em seu artigo 7º, dentre as quais apenas o inciso I exige o consentimento expresso do titular. Assim, a arquitetura proposta tem o funcionamento básico descrito a seguir:

1. Se requerido pela base legal aplicável, o Controlador e o Operador solicitam, através do *Gateway*, consentimento ao Titular para realizar tratamento sobre seus dados pessoais;
2. O *Gateway* realiza o registro da Política de Acesso correspondente na *Blockchain*;
3. O Titular encaminha os dados solicitados ao *Gateway*, que irá armazená-los no Repositório de Dados;
4. O Controlador e o Operador solicitam o tratamento dos dados[1] ao *Gateway*. Após seu retorno, a operação OP_i é salva no Registro de Operações;
5. Após um tempo *T*, haverá as operações $L_k = \{OP_1, \cdots, OP_N\}$ no Registro de Operações.
6. O Registro de Operações calcula $hash(L_k)$, armazenando-o na *Blockchain* através da execução de um contrato inteligente;
7. Após um tempo preestabelecido ou se solicitada uma verificação, o Verificador realiza a validação da integridade dos novos itens do Registro de Operações, efetuando a conferência com os valores dos *hashes* armazenados na *Blockchain*.
8. Caso seja detectada alguma não conformidade, o Verificador executa um outro contrato inteligente para registrá-la, e o Titular e o Controlador são informados dessa violação.

4. Discussão

A imutabilidade dos registros em uma *blockchain* torna-a ideal para armazenar registros de auditoria, de preferência em uma rede pública com alta disponibilidade e sem necessidade de permissão para participação – *permitionless*, promovendo, assim, a transparência de seu conteúdo. Por outro lado, esse tipo de rede apresenta baixo *throughput*. Como exemplo, a rede *Ethereum* realiza em média de cerca de 13 Transações por Segundo (TPS)[2], com um tempo de produção de blocos de 13 segundos[3].

[1] Conforme definido no Art. 5º, X da [LGPD 2018].
[2] https://ethtps.info/ . Acessado em 20/06/2022.
[3] https://etherscan.io/chart/blocktime . Acessado em 20/06/2022.

O Registro de Operações é empregado para contornar o custo financeiro e a baixa escalabilidade da rede *blockchain*, cujo uso é inviável para armazenamento de dados em cenários de alto volume de operações. A princípio, esse componente não é capaz de garantir, por si só, a imutabilidade dos dados, e, por isso, deve calcular, periodicamente, o valor $h_k = hash(L_k)$ da k-ésima lista L_k dos N novos itens nele inseridos (vide Figura 1). Em seguida, deve-se encaminhar h_k para inserção na *Blockchain*.

Utilizando ideias do conceito de *blockchain*, h_k é incluído em L_{k+1} e, assim, h_{k+1} também será calculado sobre o valor h_k. Logo, é possível, a partir do Registro de Operações e do último *hash* inserido na *Blockchain*, garantir a imutabilidade dos itens. A conferência dos *hashes* é feita pelo Verificador, bem como a obervação da conformidade das operações com as regras acordadas na Política de Acesso. Por se basear em uma *blockchain* pública e *permissionless*, não há restrição quanto à quantidade de organizações que podem ou devem fazer uso da solução para gerar seus rastros de auditoria.

Os dados pessoais propriamente ditos não fazem parte do escopo do processo de auditoria, mas somente as operações realizadas sobre eles. Conforme descrito em [Zemler and Westner 2019], o armazenamento de dados pessoais em uma *blockchain* é incompatível com os direitos do Titular de Dados garantidos pela GDPR, e, por consequência, incompatível com a LGPD, sendo infringidos os princípios da exatidão de dados e da limitação de armazenamento, como descrito nos artigos 5º da GDPR e 18º da LGPD, assim como o direito ao esquecimento do titular. Por esse motivo, o Repositório de Dados precisa utilizar uma tecnologia que permita alteração e exclusão dos dados.

Finalmente, é necessário que, para o correto funcionamento do modelo proposto, o *Gateway*, o Repositório de Dados, o Registro de Operações e o Verificador se comportem de maneira honesta, dentro das regras da arquitetura. Por exemplo, não se espera que, maliciosamente, o *Gateway* deixe de enviar registros de acesso ou que o Registro de Operações altere determinada operação após sua gravação. Além disso, é necessária uma infraestrutura de chaves públicas que possibilitem aos elementos da arquitetura a identificação mútua entre si.

5. Conclusão e Trabalhos Futuros

Este artigo apresentou uma pesquisa em andamento para conceber uma arquitetura para auditoria de uso de dados pessoais por agentes de tratamento. Em sua continuação, pretende-se especificar o modelo de dados para a *Blockchain* e os contratos inteligentes, realizar uma avaliação de segurança da proposta com a descrição de um modelo de ameaças e desenvolver um experimento prático usando *Ethereum*, com a finalidade de avaliar o seu desempenho sobre uma rede pública e estimar a precificação de sua operação, avaliando a relação de equilíbrio entre o período de envio dos *hashes* para a *Blockchain* e o custo financeiro e a segurança do Registro de Operações, considerando, ainda, a definição de tecnologia para a sua implementação.

Entre oportunidades de pesquisas futuras estão a investigação de aplicabilidade de novas tecnologias de camada 1 e 2 de *blockchains*, como *sharding* e *state channels*, respectivamente, e de bancos de dados imutáveis baseados em provas criptográficas, como o *immudb*[4].

[4]https://codenotary.com/technologies/immudb/

Referências

[Barati et al. 2022] Barati, M., Aujla, G. S., Llanos, J. T., Duodu, K. A., Rana, O. F., Carr, M., and Ranjan, R. (2022). Privacy-aware cloud auditing for GDPR compliance verification in online healthcare. *IEEE Transactions on Industrial Informatics*, 18(7):4808–4819.

[Bilal et al. 2018] Bilal, K., Khalid, O., Erbad, A., and Khan, S. U. (2018). Potentials, trends, and prospects in edge technologies: Fog, cloudlet, mobile edge, and micro data centers. *Computer Networks*, 130:94–120.

[GDPR 2016] GDPR (2016). European Parliament, Council of the European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union*, L 119:1–88.

[Hossein et al. 2019] Hossein, K. M., Esmaeili, M. E., Dargahi, T., and khonsari, A. (2019). Blockchain-based privacy-preserving healthcare architecture. In *2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)*, pages 1–4.

[Kunz et al. 2020] Kunz, I., Casola, V., Schneider, A., Banse, C., and Schütte, J. (2020). Towards tracking data flows in cloud architectures. In *2020 IEEE 13th International Conference on Cloud Computing (CLOUD)*, pages 445–452.

[LGPD 2018] LGPD (2018). Governo Brasileiro. Lei N° 13.709, de 14 de agosto de 2018 - Lei geral de Proteção de Dados Pessoais (LGPD). *Diário Oficial da União*, 15/08/2018, Edição 157, Seção 1:59–64.

[Russo et al. 2018] Russo, B., Valle, L., Bonzagni, G., Locatello, D., Pancaldi, M., and Tosi, D. (2018). Cloud computing and the new eu general data protection regulation. *IEEE Cloud Computing*, 5(6):58–68.

[Tikkinen-Piri et al. 2018] Tikkinen-Piri, C., Rohunen, A., and Markkula, J. (2018). Eu general data protection regulation: Changes and implications for personal data collecting companies. *Computer Law & Security Review*, 34(1):134–153.

[Truong et al. 2020] Truong, N. B., Sun, K., Lee, G. M., and Guo, Y. (2020). GDPR-compliant personal data management: A blockchain-based solution. *IEEE Transactions on Information Forensics and Security*, 15:1746–1761.

[Wu et al. 2019] Wu, Z., Williams, A. B., and Perouli, D. (2019). Dependable public ledger for policy compliance, a blockchain based approach. In *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*, pages 1891–1900.

[Zemler and Westner 2019] Zemler, F. and Westner, M. (2019). Blockchain and GDPR: Application scenarios and compliance requirements. In *2019 Portland International Conference on Management of Engineering and Technology (PICMET)*, pages 1–8.