Solução de Nodos de Baixo Armazenamento para o Futuro da Internet

Ramon Cordeiro[1], **Leonardo da Costa**[1], **Antônio Abelém**[1]

[1]Faculdade de Computação – Universidade Federal do Pará (UFPA)
Belém – PA – Brazil

{ramoncord,lbc,abelem}@ufpa.br

Abstract. *In blockchain, full nodes (FNs) store all existing transactions and are responsible for validating new blocks. The amount of data stored by FNs has been increasing significantly in popular blockchains, such as that of Bitcoin. The excessive amount of data from blockchains increases the storage and processing overhead in FNs, which may cause a reduction in the number of devices that validate and store blockchain data. Also, it may put the decentralization principle of blockchain in risk. This paper proposes a less expensive data storage mechanism for blockchain FNs. The proposed mechanism aims to reduce the storage and processing overhead in FNs, and to guarantee the decentralization characteristic of the network.*

Resumo. *Em blockchain, os nodos completos (NCs) armazenam todas as transações existentes e são responsáveis por validar novos blocos. A quantidade de dados armazenados por NCs vem aumentando significativamente nas principais blockchains, como a do Bitcoin. O excesso de dados de blockchains aumenta a sobrecarga de armazenamento e processamento nos NCs, podendo causar a redução de validadores e armazenadores dos dados e colocando em risco o princípio de descentralização em blockchain. Este artigo propõe um mecanismo de armazenamento dos dados menos custoso para NCs de blockchain. O mecanismo visa diminuir a sobrecarga de armazenamento e processamento nos NCs, e garantir a característica de descentralização da rede.*

1. Introdução

Blockchain tem recebido grande atenção nos últimos anos por conta do seu sucesso na aplicação em diversas áreas, e pela garantia de confiabilidade entre partes não confiáveis [Xu 2018, Zheng et al. 2018]. A tecnologia tem sido amplamente utilizada em, por exemplo, soluções de suprimentos, governos eletrônicos e segurança de dados.

Blockchain consiste em um banco de dados descentralizado, no qual os dados são criptografados e armazenados em blocos encadeados mantidos por uma rede P2P (*peer-to-peer*) [Xu et al. 2018]. A tecnologia, vista por muitos como base para a infraestrutura da Internet em vários cenários no futuro, tem atraído grande interesse da academia e da indústria devido às características inerentes da mesma [da Costa et al. 2018]. Dentre elas, *blockchain* permite a troca segura de dados (transações) entre entes que não confiam uns nos outros. Os dados ficam armazenados permanentemente na cadeia de blocos, sendo computacionalmente muito custoso comprometer a integridade dos mesmos.

A rede P2P de uma *blockchain* é composta por uma série de nodos completos (NCs), onde cada NC armazena e verifica a validade de todos os dados já contidos na cadeia, que é gradualmente formada através de um protocolo de consenso [Kim et al. 2019]. Essa forma de armazenamento, com o passar tempo e aumento de volume de dados, causa sobrecarga de armazenamento e processamento nos NCs [Kim et al. 2019].

A necessidade de mais espaço de armazenamento para novos blocos gerados aumenta o custo de permanência ou entrada de novos NCs na rede. Como consequência, a rede pode sofrer centralização, pois somente NCs com condições para aumentar o espaço de armazenamento poderão continuar operando [Xu 2018]. Tal tendência de centralização pode também causar sobrecarga de processamento nos NCs remanescentes, visto que eles precisarão gerar novos blocos e verificar todas as transações [Xu 2018].

O objetivo deste trabalho é propor um mecanismo alternativo de armazenamento dos dados de uma *blockchain* em NCs. O mecanismo visa diminuir a sobrecarga de armazenamento nos NCs, mantendo a descentralização da rede e permitindo que dispositivos com baixa capacidade de armazenamento tornem-se NCs validadores e armazenadores da rede.

O restante desse trabalho está organizado da seguinte forma. A Seção 2 aborda os trabalhos relacionados. Em seguida, a Seção 3 apresenta a proposta, enquanto a Seção 4 realiza uma discussão sobre o mecanismo proposto. Por fim, a Seção 5 conclui o trabalho e aponta trabalhos futuros.

2. Trabalhos Relacionados

Os trabalhos da literatura que apresentam propostas para mitigar o problema de sobrecarga de armazenamento e processamento de NCs em *blockchain* normalmente utilizam algum mecanismo de compressão ou codificação de dados com o intuito de comprimir as transações já inseridas na cadeia através de técnicas advindas da teoria dos códigos [Milies 2009] ou criptografia [Shamir 1979].

Os trabalhos apresentados por [Dai et al. 2018], [Kim et al. 2019] e [Perard et al. 2018] utilizam a abordagem de codificação dos dados com o intuito de gerar economia de armazenamento. Essas propostas tem em comum a necessidade de realizar processamento na codificação e decodificação dos blocos e/ou transações. A diferença entre as propostas está na maneira com que cada mecanismo realiza o processamento.

[Dai et al. 2018] propõem um mecanismo de economia no armazenamento das transações da *blockchain* através da codificação e fatiamento dos blocos. Cada bloco gerado é particionado em sub-blocos de tamanhos iguais, que são codificados, gerando outros blocos. Um exemplo descrito por [Dai et al. 2018] em que um bloco gerado é particionado em K blocos (ex:A,B e C), em seguida é codificado em outros N blocos (p.ex:A, B, C, $A + B + C$, $A + 2B + 3C$, $A + 4B + 2C$).

O trabalho de [Kim et al. 2019] cria domínios de pares (par no conceito do autor é um nodo validador de *blockchain*) em que cada domínio é responsável por blocos específicos. Em um determinado domínio, um novo bloco é partilhado utilizando segredo de Shamir na divisão do bloco, no qual cada NC possui uma parte encriptada do bloco e uma parte de uma chave privada que é segregada entre os participantes do domínio. Im-

portante salientar que existe uma chave privada geral que é dividida (utilizando Shamir) entre todos os nodos da rede para que seja possível recuperar a chave de um domínio e consequentemente decodificar um bloco.

O trabalho de [Perard et al. 2018] propõe um novo tipo de NC, diferente de NCs tradicionais. Este novo tipo de NC armazena apenas parte da cadeia, mas atua também como validador. O armazenamento de fragmentos codificados de bloco por cada *low storage node* (LS) é obtido primeiramente particionando o bloco em partições de tamanhos iguais e gerando combinação linear dos fragmentos. A combinação linear é específica de cada bloco e cada *node*, isto é, os NCs armazenam fragmentos codificados diferentes, mesmo que aqueles sejam originários do mesmo bloco.

Os trabalhos apresentam redução da sobrecarga de armazenamento nos NCs. Entretanto, a utilização de codificação dos dados aumenta a necessidade de processamento tanto na inserção dos dados na *blockchain* quanto no processo de recuperação dos mesmos. Esta etapa dificulta que NCs com baixa capacidade computacional tornem-se validadores de transações em redes de *blockchain*. Tendo como motivação este problema, a próxima seção apresenta um mecanismo para tratar a sobrecarga de armazenamento e processamento em NCs de *blockchain*.

3. Mecanismo Proposto

Nesta seção, apresenta-se o mecanismo de armazenamento de dados da *blockchain* em NCs, objetivando reduzir a sobrecarga de armazenamento e processamento nos mesmos. *Blockchain* possui a premissa básica de descentralização, especialmente dos dados. Outra premissa básica de *blockchain* é a redundância dos dados em entes da rede que estejam interessados em contribuir com a *blockchain*. Desta forma, propor uma maneira de armazenar os dados da *blockchain* de maneira menos dispendiosa deve seguir estas premissas.

Um dos problemas das *blockchains* tradicionais é que os NCs armazenam toda a cadeia de blocos. Embora que inicialmente esta abordagem seja vantajosa, à medida que a cadeia de blocos aumenta, é necessário muito mais espaço de memória secundária disponível em cada NC. Isto pode implicar no desligamento de NCs da *blockchain* ou impossibilidade de dispositivos com menor capacidade adentrarem na rede para tornar-se NCs. A saída de NCs da rede e a impossibilidade de novos dispositivos adentrarem para desempenhar o papel de NCs pode acarretar em centralização da rede. Uma solução possível é reduzir a quantidade de dados que um NC armazena sem perder as características de elemento validador da rede e replicador dos dados (características essas que diferenciam nodos completos de clientes leves [da Costa et al. 2018]).

No mecanismo introduzido neste trabalho, propõe-se que sejam criados domínios de armazenamento, nos quais os NCs mantém redundância dos dados a serem mantidos pelo domínio. Os domínios são escaláveis de acordo com o comportamento da *blockchain*. Novos domínios podem ser criados, a partir do desmembramento de domínios que atingem o limiar de teto, ou podem ser eliminados quando este atinge o limiar inferior. Cada domínio mantém uma parte da *blockchain* e seus membros possuem as informações completas que o domínio armazena, desta forma a redundância é garantida, como mostra a figura 3.

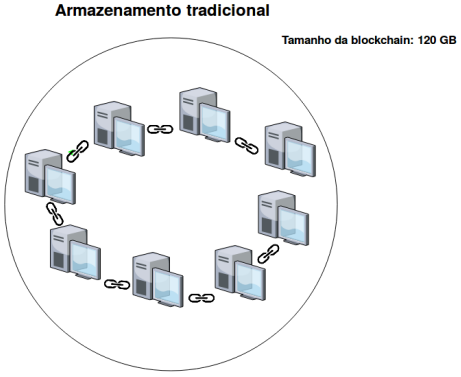
Os limiares de teto e limiar inferior, são valores previamente definidos através do consenso, em que quando o limiar de teto é atingido, um novo domínio pode ser criado, e

alguns dos membros do atual domínio se tornarão membros do novo domínio. Também, neste processo os dados do domínio anterior são divididos de maneira que parte dos dados é armazenada pelos NCs do primeiro domínio e o restante dos dados pelos membros do segundo domínio. Como inicialmente os membros do novo domínio possuem os dados de ambos os domínios, os dados do primeiro domínio podem ser apagados nestes membros.

Existe a possibilidade de um domínio perder membros de tal forma que seja atingido o limiar inferior, quando isto acontecer os membros devem ser alocados em outro domínio existente, receber os dados que os membros deste domínio possuem e replicar os seus dados com os outros NCs. Os domínios são utilizados para manter a replicação dos dados na rede, sem sobrecarregar os NCs, permitindo que NCs de menor capacidade participem da rede e esta seja mantida com a sua característica de descentralização.

Pode-se utilizar a rede do Bitcoin como exemplo de caso de uso. Atualmente esta *blockchain* possui um tamanho pouco superior a 200 Gigabytes (GB), e a rede possui em torno de 10 mil NCs. No modelo atual, cada NC do Bitcoin armazena inteiramente os mais de 200 GB, como mostra a Figura 1. Suponha-se uma situação teórica, em que no consenso seja implantado que cada domínio deve possuir pelo menos 50 NCs como limiar inferior e 200 NCs de limiar superior. Além disso, com o rearranjo da rede, suponha-se que sejam criados 100 domínios com 100 NCs cada. Neste cenário, cada domínio necessita armazenar 2 GB da *blockchain*, isto significa que cada NC armazena apenas os 2 GB delegados ao domínio à que este pertence. Neste exemplo, é perceptível a economia de espaço de armazenamento em cada um dos NCs. A Figura 2, mostra o exemplo de economia de armazenamento numa *blockchain* de 120 Gb.

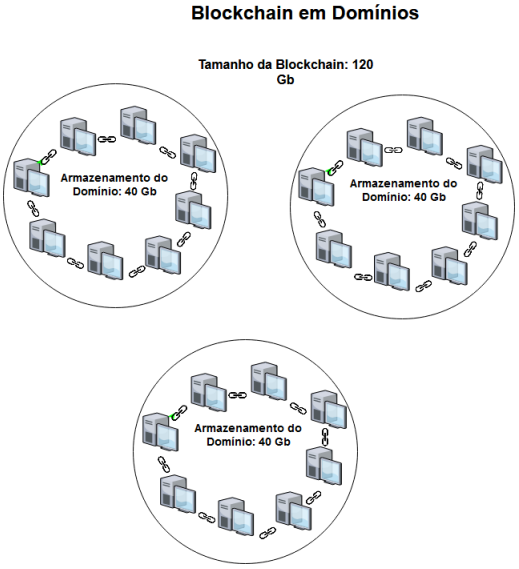
Figura 1. Armazenamento tradicional em *blockchain*.



4. Discussão

A proposta apresentada na Seção 3 possui alguns benefícios. Primeiramente, devido a redução da sobrecarga de armazenamento em cada NC, o tempo de sincronização de novos NCs deverá ser menor, pois não será necessário baixar toda a *blockchain*. O volume de dados trocados também será menor. Outra vantagem significativa é a possibilidade de utilizar dispositivos com menor capacidade de armazenamento como NC validador e mantenedor da *blockchain*, garantindo a descentralização. Isto garante também a possibilidade de utilizar dispositivos de *IoT* na *blockchain* desempenhando o papel de NCs, o que não é possível na maioria dos trabalhos apresentados na Seção 2, tendo em vista que

Figura 2. Armazenamento em forma de domínios *blockchain*.



em muitos dos trabalhos relacionados, são utilizadas técnicas de compressão de dados, normalmente baseadas em teoria dos códigos.

A utilização de compressão/descompressão no âmbito de redução de armazenamento é positiva, mas aumenta o processamento necessário em cada NC para armazenar ou recuperar informações, em alguns casos, o adicional de processamento pode inviabilizar ou dificultar a utilização de dispositivos de baixa capacidade como NC armazenador da *blockchain*. Neste aspecto, a proposta apresenta vantagem. A desvantagem das propostas apresentada na Seção 3 pode ser o tamanho dos dados armazenados por cada NC em relação a alguns trabalhos que utilizam compressão de dados no armazenamento [Dai et al. 2018, Kim et al. 2019].

5. Conclusões e Trabalhos Futuros

O mecanismo proposto é uma alternativa possível para reduzir a sobrecarga de armazenamento em *blockchain*, especialmente porque possui o foco, além do armazenamento, em manter a descentralização da rede, uma das premissas básicas de *blockchain*. A preocupação em manter a descentralização, neste caso garante a replicação dos dados, visto que o mecanismo utiliza-se da descentralização para manter os dados replicados.

Como trabalhos futuros, deverá ser implementado a proposta numa *blockchain* (preferencialmente não-permissionada e pública), realizar os testes com NCs, verificar a redução no armazenamento de dados por NC e em seguida comparar a redução com resultados obtidos em outras propostas.

Agradecimentos

Essa pesquisa teve suporte parcial da CAPES e da chamada conjunta RNP-NSF para pesquisa e desenvolvimento em cibersegurançaa (INSaNE - Improving Network Security at the Network Edge), financiada pela National Science Foundation e pelo Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTIC) através da RNP e do CTIC.

Referências

da Costa, L., Neto, A., Pinheiro, B., Araújo, R., and Abelém, A. (2018). Dlcp: Um protocolo para a operação segura de clientes leves em blockchains. In *Anais do IX Workshop de Pesquisa Experimental da Internet do Futuro (WPEIF - SBRC 2018)*, volume 9, Porto Alegre, RS, Brasil. SBC.

Dai, M., Zhang, S., Wang, H., and Jin, S. (2018). A low storage room requirement framework for distributed ledger in blockchain. *IEEE Access*, 6:22970–22975.

Kim, Y., Raman, R. K., Kim, Y.-S., Varshney, L. R., and Shanbhag, N. R. (2019). Efficient local secret sharing for distributed blockchain systems. *IEEE Communications Letters*, 23(2):282–285.

Milies, C. P. (2009). Breve introdução a teoria dos códigos corretores de erros. *Departamento de Matemática, UFMS*.

Perard, D., Lacan, J., Bachy, Y., and Detchart, J. (2018). Erasure code-based low storage blockchain node. *arXiv preprint arXiv:1805.00860*.

Shamir, A. (1979). How to share a secret. *Communications of the ACM*, 22(11):612–613.

Xu, Y. (2018). Section-blockchain: A storage reduced blockchain protocol, the foundation of an autotrophic decentralized storage architecture. In *2018 23rd International Conference on Engineering of Complex Computer Systems (ICECCS)*, pages 115–125. IEEE.

Xu, Z., Han, S., and Chen, L. (2018). Cub, a consensus unit-based storage scheme for blockchain system. In *2018 IEEE 34th International Conference on Data Engineering (ICDE)*, pages 173–184. IEEE.

Zheng, Q., Li, Y., Chen, P., and Dong, X. (2018). An innovative ipfs-based storage model for blockchain. In *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pages 704–708. IEEE.