Pudim de Pão e Chia

uma blockchain de espaço útil

Mónica Jin, Miguel Matos, and João Barreto

IST U. Lisboa & INESC-ID
monicachenjin@tecnico.ulisboa.pt
miguel.marques.matos@tecnico.ulisboa.pt
joao.barreto@tecnico.ulisboa.pt

Abstract. As blockchains sem permissão, como o Bitcoin, utilizam provas de trabalho (em inglês *proof-of-work*, ou PoW) para garantir a segurança do sistema. Embora eficaz, PoW tem uma pegada energética significativa, o que tem levado à investigação de abordagens alternativas como as provas de espaço (em inglês *proof-of-space*, ou PoSp). As blockchains baseadas em PoSp oferecem uma alternativa promissora, baseando-se em espaço de disco em vez do poder de processamento para garantir a segurança do sistema. As técnicas PoSp atuais geram grandes quantidades de dados aleatórios para a geração de prova, resultando num desperdício de espaço em disco. Este trabalho propõe técnicas que utilizam dados existentes dos participantes das blockchains para a geração da prova de espaço, reduzindo o desperdício de espaço em disco inerente às abordagens existentes. Além disso, esta abordagem alarga a participação no sistema a utilizadores que já possuem grandes quantidades de dados, o que em última instância reforça a segurança do sistema como um todo. Os nossos resultados preliminares mostram que podemos substituir no mínimo 50% dos dados aleatórios por dados úteis.

Keywords: Prova de Espaço · Prova de Armazenamento · Espaço Útil · Blockchain.

1 Introdução

As blockchains são uma das tecnologias mais populares dos últimos anos. Uma blockchain é uma estrutura de dados que visa registar transações numa rede de nós distribuída. Esta permite que os nós cheguem a um consenso sobre uma única versão de um registo, sem a necessidade de uma autoridade central. Uma blockchain sem permissão concede a participação de qualquer nó no sistema, sem exigir permissão para o seu ingresso. Exemplos populares destas blockchains são Bitcoin [**?**] e Ethereum [**?**]. Este tipo de blockchain permite um sistema distribuído totalmente descentralizado, onde não existe uma única entidade central, mas sim várias entidades que seguem um protocolo que permite a colaboração dos participantes. Isto significa que se uma dada entidade falhar, o sistema continua a funcionar como pretendido. No entanto, este tipo de blockchain é mais

suscetível a ataques *Sybil* [**?**], nos quais um nó malicioso cria várias identidades falsas para ganhar influência desproporcional na rede. Um atacante que cria identidades falsas suficientes, é ser capaz de controlar uma parte significativa da rede e influenciar o resultado do processo de consenso. Isto pode levar o atacante a assumir o controlo da blockchain e alterar registos existentes.

Blockchains proeminentes como o Bitcoin [**?**] mitigam o risco destes ataques usando provas de trabalho, que requerem que os participantes gastem uma quantidade significativa de poder de processamento para adicionar um novo bloco à cadeia. Isto dificulta a criação de identidades falsas ao atacante que pretende ganhar influência no sistema. Infelizmente, este mecanismo vem com o custo de um alto consumo energético. O consumo de energia das blockchains PoW é uma fonte de preocupação, pois pode contribuir para as emissões de gases de efeito estufa e outros impactos ambientais [**?**].

Este problema levou à investigação de mecanismos mais eficientes em termos energéticos e à prova de ataques *Sybil*, como a prova de participação (*Proof-of-Stake*, ou PoS) e a prova de espaço (*Proof-of-Space*, ou PoSp). Numa PoS blockchain, o mineiro do próximo bloco é escolhido com base na sua participação (*stake*) na rede, em vez do seu poder computacional. Recentemente, o Ethereum substituiu o seu antigo protocolo PoW por PoS [**?**]. Esta decisão foi tomada para eliminar o alto consumo energético [**?**]. Contudo, uma das maiores desvantagens das blockchains PoS é o problema do enriquecimento dos mais ricos (em inglês *rich-get-richer issue*). Este problema argumenta que se os mineiros *mais ricos* (aqueles que possuem mais participações) obterem mais recompensas, então aumentarão ainda mais sua participação no futuro [**?**].

PoSp é outra abordagem promissora que depende do espaço em disco, em vez do poder de computação, para garantir a segurança do sistema. Numa blockchain PoSp, os mineiros devem provar que têm uma determinada quantidade de espaço de armazenamento disponível, alocando uma parte dele para o sistema. Para isso, as técnicas PoSp existentes dependem da geração de grandes quantidades de dados aleatórios para construir estas provas. O Chia [**?**] é uma blockchain baseada em PoSp que foi lançado em 2021. Chia tem ganho atenção e adoção significativas desde o lançamento devido à sua abordagem energeticamente eficiente para garantir a segurança da rede. Uma preocupação com PoSp é o desperdício de espaço de armazenamento. Para participar nas PoSp blockchains, os mineiros alocam espaço em disco gerando dados aleatórios com determinadas propriedades criptográficas. Isto leva a um desperdício de armazenamento com dados aleatórios. Atualmente, no Chia, todos os mineiros fornecem cerca de 32 bilhões de gigabytes de armazenamento que não têm outra finalidade além de garantir a segurança da blockchain.

Uma maneira de reduzir o desperdício de armazenamento das PoSp blockchains é tornar os dados aleatórios usados para as provas em dados úteis (*i.e.*, têm outra utilidade para o mineiro para além de assegurar a participação no sistema). Neste artigo apresentamos Pudim de Pão e Chia, um protocolo que visa preservar as garantias dos protocolos PoSp tirando proveito dos dados que cada utilizador já dispõe para a geração de provas e diminuindo assim o desperdício de espaço

em disco inerente às abordagens existentes. Os nossos resultados preliminares mostram que podemos substituir no mínimo 50% dos dados aleatórios por dados úteis.

2 Trabalho Relacionado

As provas criptográficas desempenham um papel fundamental na garantia da segurança e integridade das blockchains. Estas verificam se os mineiros investiram num recurso não falsificável, *i.e.*, um recurso que não pode ser replicado ou duplicado sem um custo significativo.

As provas de trabalho (PoW) são provas criptográficas onde o recurso em questão é o poder de processamento computacional. Estas provas permitem que um participante da rede prove que foi gasto um esforço computacional significativo para resolver um problema criptográfico. Esta prova é facilmente verificável por outros participantes. PoW é um mecanismo criptográfico amplamente utilizado em sistemas descentralizados e distribuídos. Embora tenha sido proposto para múltiplos objetivos de segurança, incluindo proteção contra ataques de negação de serviço e *spamming* [**?**,**?**], foi somente com o surgimento do Bitcoin [**?**], a primeira criptomoeda descentralizada, que seu potencial foi reconhecido.

As provas de espaço (PoSp) são provas criptográficas que têm como recurso não falsificável o espaço de armazenamento. Dziembowski et al. [**?**] introduzem pela primeira vez este protocolo como uma alternativa mais ecológica ao PoW. PoSp é um protocolo com duas fases distintas: inicialização e execução. Durante a inicialização, são gerados grandes quantidades de dados aleatórios com determinadas propriedades criptográficas que são armazenados em disco. Na fase de execução, os dados previamente armazenados são usados para a geração de uma prova de PoSp. As PoSp têm um custo computacional inicial que é amortizado ao longo da vida do sistema, uma vez que os dados armazenados podem ser reutilizados para várias provas futuras. A construção das PoSp baseia-se num processo demorado de geração de dados aleatórios. Este processo é significativamente mais demorado do que a busca desses dados armazenados no disco. Portanto, no contexto das blockchains, os mineiros são incentivados a armazenar os dados ao invés dos gerar em tempo real, *i.e.*, quando criam um novo bloco para a cadeia. Existem duas abordagens para PoSp. A primeira apresentada por Dziembowski et al. [**?**] é baseada em grafos *hard-to-pebble*[1] e árvores de Merkle[2]. Dziembowski et al. provam que um atacante utiliza $\Omega(N)$ espaço para armazenar provas ou faz $\Omega(N)$ invocações à função de dispersão. Por outras palavras, o atacante que opta por não armazenar o espaço devido na fase de inicialização, terá de computar várias funções de dispersão durante a criação de um novo bloco para a blockchain. Assim, um participante racional opta sempre

[1] Um grafo acíclico direcionado onde os rótulos de um vértice são o valor de hash dos vértices pais. Um vértice só pode ser rotulado se todos os seus vértices pais também estiverem rotulados.

[2] Estrutura de árvore na qual cada nó folha é o hash de um bloco de dados, e os restantes nós são o hash dos seus filhos.

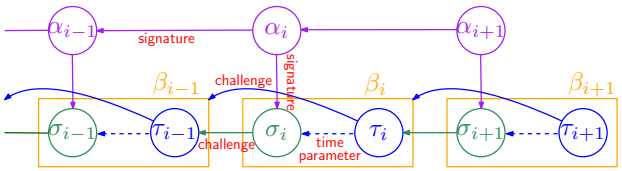

Fig. 1. Ilustração da blockchain do Chia [**?**]. Os dados α de cada bloco estão desacoplados das provas (PoSp σ e PoT τ).

pela primeira opção, já que é mais barata e mais rápida. Esta variante do PoSp oferece as melhores garantias de segurança para esse tipo de prova, mas é difícil de ser implementada. O tamanho das provas é grande (na ordem dos Mbytes), e estas provas não podem ser totalmente não interativas quando adaptadas para as blockchains (*i.e.*, os participantes não podem ingressar a qualquer momento sem preparação prévia). Abusalah et al. [**?**] propõem uma segunda abordagem para o PoSp. Esta abordagem é baseada na inversão de funções de dispersão. Durante a fase de inicialização, colisões das funções de dispersão são computadas e estas são armazenadas em disco. Na fase de execução, a prova é gerada através do conteúdo previamente armazenado. Apesar desta abordagem não ter garantias de segurança tão fortes como a primeira abordagem, tem a vantagem de ter provas com um tamanho pequeno e de ser não interativa.

Chia [**?**] é a primeira blockchain a implementar um protocolo de consenso com base na segunda abordagem de PoSp [**?**]. Este protocolo utiliza provas de espaço e provas de tempo (em inglês *proof of time*, ou PoT). As PoT baseiam-se nas funções de atraso verificáveis (VDF) [**?**,**?**]. Uma VDF é uma função não paralelizável onde o tempo de execução é configurável e predeterminado. Para cada *input* existe somente um *output* que é rapidamente verificável. No Chia, existem dois tipos de mineiros, mineiro PoSp e mineiro PoT. Um mineiro PoSp primeiro dedica uma quantidade de espaço num processo chamado inicialização. Após este processo, o mineiro pode começar a criar blocos. Para estender uma cadeia, o mineiro primeiro calcula o desafio de 256 bits ao obter o resultado da função de dispersão da última PoT. Em seguida, o mineiro procura no seu armazenamento por dados que correspondam ao desafio e gera uma prova através destes. Assim que a prova é gerada, o mineiro estende sua cadeia local com essa prova e transmite-a para a rede. Um mineiro PoT pega nesse bloco não finalizado (ou seja, um bloco com apenas uma prova PoSp) valida a prova PoSp, classifica-a em termos de qualidade e calcula a VDF usando esta qualidade. Após a execução do VDF, este mineiro finaliza o bloco estendendo-o com uma prova PoT. O Chia está ajustado para ter uma prova PoT em cada 9,375 segundos. Assim que cada PoT é partilhada, os mineiros PoSp têm oportunidade para estender a blockchain, recomeçando o processo. A blockchain é estendida alternando entre provas PoSp e PoT (ilustrado na Figura 1). O tempo de execução da prova de tempo é inversamente proporcional à qualidade da prova de espaço. Uma PoSp

com qualidade elevada, resulta num menor tempo de execução para a prova de tempo que a sucede.

O Chia tem sido amplamente adotado como uma alternativa mais acessível e ecológica ao Bitcoin. Um dos principais problemas do Chia é o desperdício de recursos de armazenamento. Os dados gerados no processo de inicialização apenas podem ser usados gerar provas de espaço, o que introduz um desperdício computacional e de armazenamento.

Provas de armazenamento (em inglês *proof of storage*, ou PoSt) [**?**,**?**,**?**,**?**,**?**] são semelhantes aos esquemas PoSp, mas em vez de mostrarem que o espaço está alocado com dados aleatórios, provam que o espaço alocado está a armazenar corretamente dados previamente fornecidos. PoSt são utilizadas nas redes de armazenamento descentralizado, pois permitem aos participantes compartilhar e armazenar dados sem terem de depender de um único fornecedor de armazenamento. Provas de armazenamento têm um grande potencial para arquivamento porque permitem a um cliente armazenar um ficheiro num servidor sem confiar nele e posteriormente verificar a integridade deste.

PoSt são usadas nas blockchains como um mecanismo à prova de ataques *Sybil*. Adicionalmente, o trabalho realizado pelos mineiros de PoSt é considerado *útil*, uma vez que o resultado da computação tem outra utilidade para a rede além de garantir a segurança da blockchain. Exemplos recentes destas blockchains incluem Filecoin [**?**] e Subspace [**?**]. Filecoin [**?**] é uma rede de armazenamento descentralizado que transforma o armazenamento na nuvem num mercado algorítmico. Este mercado funciona com a sua criptomoeda nativa FIL, que os mineiros ganham ao fornecer armazenamento para os clientes. A rede de armazenamento descentralizado baseia-se num tipo de PoSt designado por prova de replicação (*proof of replication*, ou PoRep). A PoRep, além de comprovar que determinados dados foram armazenados por um mineiro, garante que este dedica um espaço em disco único e que os dados são recuperáveis. Os atacantes não podem fingir armazenar várias cópias dos mesmos dados através da deduplicação do armazenamento, garantindo que todos os fornecedores de armazenamento guardam as réplicas de forma independente. Subspace [**?**] é outra PoSt blockchain onde mineiros são incentivados a armazenar blocos da própria blockchain. Cada mineiro estende a blockchain com provas de replicação dos blocos que armazena. Quer o Filecoin, quer o Subspace requerem protocolos para a distribuição dos dados úteis e para assegurar o seu armazenamento durante um longo período de tempo. Estes protocolos aumentam a complexidade do sistema e, como consequência, impactam o desempenho da rede. O nosso trabalho procura uma alternativa mais simples a estes protocolos removendo os requisitos de distribuição e recuperação de dados.

3 Pudim de Pão e Chia

O Pudim de Pão e Chia é uma blockchain onde os mineiros usam ficheiros locais para gerar provas de armazenamento. Blockchains que usam provas de espaço ou provas de armazenamento têm desafios adicionais às blockchains de provas de

trabalho. Estes desafios incluem mecanismos de prevenção de *costless simulation* e *long range attacks*[**?**]. *Costless simulation attacks* comprometem a chegada do sistema a consenso, pois a geração de provas e blocos para a blockchain é deliberadamente rápida. Mineiros podem gerar várias cadeias alternativas que estendem a cadeia mais recente com vários blocos diferentes, o que atrasa o consenso. *Long range attacks* comprometem a integridade da blockchain. Mineiros podem facilmente gerar uma cadeia local mais extensa que a cadeia honesta, uma vez que o custo de gerar provas e blocos é negligenciável. Para resolver estes desafios, o nosso protocolo usa o protocolo do Chia [**?**] como base.

3.1 Desafios e Soluções propostas

O uso de ficheiros locais dos mineiros como provas de espaço não é trivial. De modo a construirmos a blockchain proposta, temos de resolver os seguintes desafios:

1) **estrutura desconhecida**: Provas de espaço [**?**,**?**] demonstram que um determinado espaço está a ser alocado ao guardar dados com determinadas propriedades criptográficas. Estes dados são gerados de acordo com estruturas criptográficas e têm um custo computacional e temporal. Um mineiro que dedica os seus ficheiros à blockchain deve conseguir recuperá-los para outros fins. A estrutura dos ficheiros dos utilizadores é imprevisível e não segue as construções das provas de espaço, não tendo assim as propriedades criptográficas destas. Para além da ausência destas propriedades, é importante que a estrutura do ficheiro seja preservada (i.e. o conteúdo do ficheiro não é alterado e pode ser lido diretamente) ou recuperável (i.e. o conteúdo é alterado mas pode ser recuperado e ser lido). Por outras palavras, apenas é útil usarmos ficheiros dos mineiros se o estado original destes for mantido ou recuperável. Para resolver este problema, propomos o uso de provas de armazenamento [**?**] ao invés de provas de espaço. Dos vários tipos de provas de armazenamento, a prova de replicação [**?**] é a que mais se adequa ao nosso contexto. Esta prova permite codificar os dados e permite a recuperação do seu estado original através da descodificação. Adicionalmente, a prova de replicação pode ser usada como uma prova de espaço, uma vez que o protocolo permite provar que certa quantidade de espaço está a ser alocado com dados úteis.
2) **baixa entropia**: Os dados das provas de espaço são gerados através de funções de dispersão. Caso o input das funções de dispersão seja repetido, obteremos resultados repetidos. A diversidade de resultados é o que permite aos mineiros de provas de espaço obterem várias provas diferentes e aumentar as suas hipóteses de estender a blockchain. Os ficheiros arquivados geralmente têm pouca entropia, uma vez que várias secções destes têm dados repetidos ou redundantes [**?**]. De modo a resolver este desafio, propomos encriptar os ficheiros originais. O uso de uma cifra em modo CBC (*Cipher Block Chaining*) permite aumentar a entropia e assegurar um certo nível de privacidade aos dados dos mineiros. Além disso, a cifra possibilita a encriptação de forma

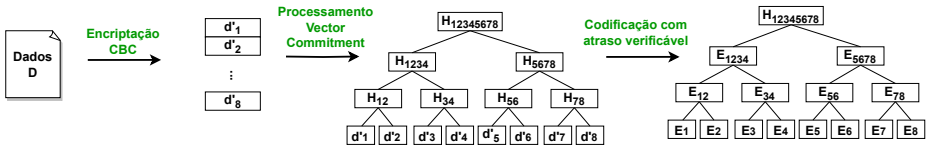

Fig. 2. Processo de Inicialização

sequencial e a desencriptação de forma paralelizável. Assim, o processo de inicialização da prova mantém-se sequencial enquanto a recuperação dos dados originais é eficiente.

3.2 Protocolo proposto

O nosso protocolo substitui a prova de espaço [**?**] do Chia por uma prova de replicação [**?**], para isto modificamos os seguintes algoritmos: 1) inicialização (Figura 2); 2) geração da prova e validação da prova (Figura 3); 3) qualidade da prova.

O processo de inicialização ilustrado na Figura 2 transforma um conjunto de dados numa estrutura com certas propriedades criptográficas. Este processo é constituído por três etapas. A primeira etapa consiste em encriptar os dados com uma cifra CBC e dividi-los em blocos de tamanho constante. Posteriormente, na segunda etapa, estes blocos cifrados são usados para a geração de uma árvore de Merkle, onde cada bloco representa uma folha. Por fim, na terceira etapa, cada nó da árvore de Merkle, à exceção da raiz, é codificado com uma VDF [**?**,**?**], este algoritmo é conhecido como Codificador de Atraso Verificável (em inglês *Verifiable Delay Encoder* ou VDE). Esta codificação é feita em cadeia, onde a codificação de um nó depende do resultado dos nós codificados anteriormente. O processo de inicialização, especificamente a terceira etapa, é demorado e o tempo de execução é proporcional à quantidade de dados e ao tempo da VDF. Este processo pode ser repetido para vários ficheiros e os dados resultantes do processo de inicialização são identificados pela raiz de Merkle. Tal como explicamos posteriormente, quanto mais ficheiros um mineiro usa no processo de inicialização, mais chances tem de estender a blockchain, pois uma maior diversidade de dados armazenados permite a geração de mais provas.

Após a inicialização, o mineiro pode começar a criar provas de replicação para participar na blockchain. O mineiro ao receber um desafio da blockchain (*i.e.*, PoT do último bloco, representado por τ na Figura 1), que consiste num número de 256 bits, escolhe a árvore de Merkle com a raiz mais semelhante ao desafio e cria uma prova através dos nós da árvore. A geração da prova consiste na seleção da folha com o índice indicado pelo resto da divisão do desafio pelo número total de folhas (na Figura 3 é a folha E3), no caminho de Merkle dessa folha (representado com os nós a laranja na Figura 3) e na raiz da árvore (a azul). Para a validação da prova, o mineiro que recebe a prova descodifica cada nó com um Descodificador de Atraso Verificável (i.e. a função inversa do VDE)

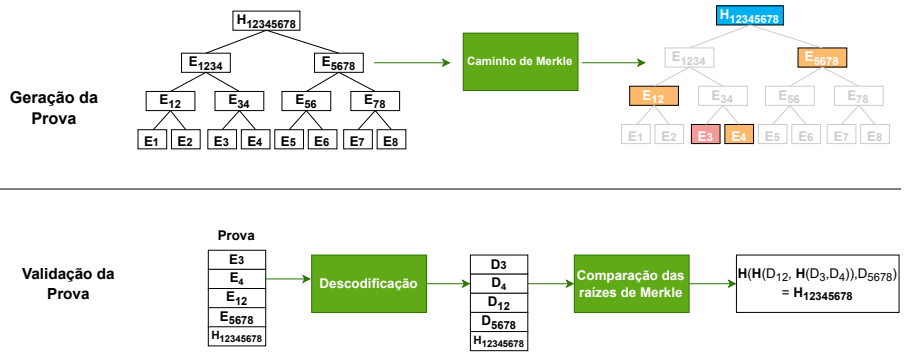

Fig. 3. Geração e validação da prova

e uma nova raiz é computada através destes nós. A prova é válida se a raiz computada é igual à raiz incluída na prova. Adicionalmente, uma prova é válida se e só se o desafio associado é o PoT mais recente. Assim, cada mineiro tem 9.375 segundos para criar e disseminar uma prova PoSp para cada desafio.

Por fim, o nosso protocolo classifica as provas de cada mineiro de modo a selecionar a prova que é incluída na blockchain. Para isto, definimos uma função de qualidade que corresponde à diferença absoluta entre o desafio e um número constituído pelos primeiros 128 bits da raiz da prova e os últimos 128 bits da folha da prova. A prova incluída na blockchain é aquela com menor diferença. Esta definição incentiva os mineiros a alocar mais ficheiros para a blockchain, uma vez que cada ficheiro gera uma raiz nova. Adicionalmente, os mineiros que armazenam honestamente todas as folhas da árvore de Merkle têm um maior domínio de possíveis provas.

3.3 Parâmetros públicos

O protocolo proposto tem os seguintes parâmetros públicos que são ajustáveis:

1. tamanho do bloco e nó m;
2. número de folhas da árvore de Merkle n;
3. fanout da árvore de Merkle f;
4. tempo de execução da função de atraso verificável t.

O ajuste destes parâmetros tem um forte impacto no desempenho do sistema. Na Secção 4.2, apresentamos uma análise aprofundada dos diferentes compromissos entre os parâmetros.

4 Avaliação

Nesta secção avaliamos o Pudim de Pão e Chia em função das seguintes questões de investigação:

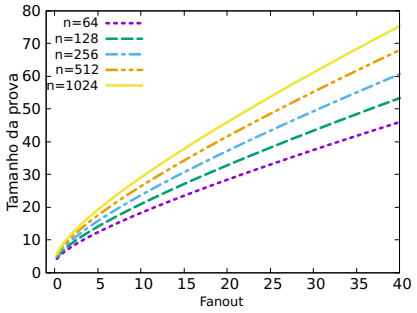

Fig. 4. Tamanho das provas (múltiplo de m) em função do fanout da árvore de Merkle para n = {64, 128, 256, 512, 1024}.

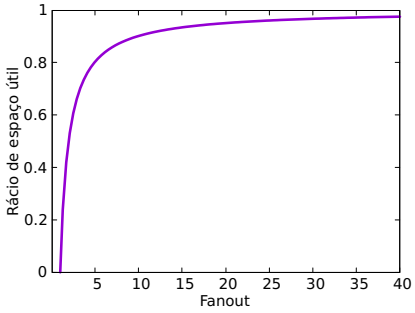

Fig. 5. Rácio de espaço útil em função do fanout da árvore de Merkle para n = {64, 128, 256, 512, 1024}.

Q1: Até que ponto podemos substituir os dados aleatórios das PoSp blockchains por dados úteis?
Q2: Quais são os efeitos de reutilizar os dados dos utilizadores no desempenho das PoSp blockchains?

4.1 Metodologia

Para realizar a avaliação e responder às questões de investigação, focámo-nos nas seguintes métricas:

Rácio de espaço útil: Proporção de espaço útil em relação ao espaço total alocado pelas provas.
Tamanho da prova: : Número de bytes da prova.
Tempo de geração de provas: Tempo de execução do processo de inicialização das provas.
Entropia das provas: Diversidade de provas.

A nossa avaliação é composta por duas partes: avaliação analítica e avaliação experimental. Na avaliação analítica, analisamos o impacto de cada parâmetro público do protocolo de modo a encontrar o melhor ajuste destes. Na avaliação experimental, corremos o protocolo proposto com parâmetros públicos específicos e colhemos os resultados.

4.2 Avaliação Analítica

As métricas de avaliação dependem dos parâmetros públicos do protocolo apresentados na Secção 3.3.

O rácio de espaço útil depende do número de folhas n da árvore de Merkle e do seu *fanout* f. O rácio é dado pela expressão $R(n, f) = \frac{n(f-1)}{fn-1}$ (ilustrada na Figura 5). Para valores mais elevados de *fanout*, obtemos um maior rácio de

Table 1. Tempo de execução em minutos do processo de inicialização para 6, 12, 25, 51, 105 GB.

Protocolo / Tamanho de dados(GB)	6	12	25	51	105
Chia	68.9	130.3	252.5	514.3	1352.2
Pudim de Pão e Chia	72.8	142.5	273.1	528.6	1171.8

espaço útil, uma vez que diminuímos o número de nós intermédios da árvore de Merkle.

O tamanho da prova depende do tamanho de cada bloco, do fanout e do número de folhas. O tamanho da prova é dado pela expressão $T(m, n, f) = m((\log_f n)(f-1)+2)$. O tamanho da prova aumenta quer com o fanout (eixo horizontal da Figura 4), quer com o número de folhas da árvore (funções da Figura 4).

O tempo de execução da fase de inicialização é proporcional ao tempo de execução de cada função de atraso verificável e do comprimento do encadeamento.

Para o desempenho ótimo do sistema, queremos minimizar o tamanho das provas e aumentar o rácio de espaço útil. O tamanho da prova influencia a complexidade do sistema. Uma prova grande tem impacto quer na transmissão dos blocos da cadeia quer no armazenamento da blockchain pelos mineiros. O rácio do espaço útil reflete diretamente a utilidade dos dados que estão a ser armazenados pelos mineiros. Caso este rácio seja muito pequeno, o sistema aproxima-se às blockchains de PoSp atuais onde os dados armazenados não têm outra utilidade para além de garantir a segurança do sistema. Portanto, queremos ter um rácio elevado de modo a dar mais utilidade aos dados que são armazenados e criar uma dinâmica mais sustentável para blockchains.

4.3 Avaliação Experimental

A avaliação experimental foi executada numa máquina contendo um Intel(R) Xeon(R) Silver 4116 CPU 2.10GHz, com 64GB de RAM.

De modo a avaliar experimentalmente o protocolo desenvolvido, executámos a implementação do nosso protocolo e o protocolo original do Chia[3], a nossa referência base. Os dados de utilizadores são criados usando o gerador DEDISbench [**?**] no modo *Personal Files*, que segue a distribuição de sistemas de armazenamento reais de utilizadores.

Os parâmetros públicos usados na avaliação foram os seguintes: $m = 256$bits; $n = 262,144$; $f = 2$; $t = 0.5s$.

A Tabela 1 lista o tempo de execução em minutos da inicialização para dados de vários tamanhos (6, 12, 25, 51 e 105 GB). Estes tamanhos foram selecionados através do parâmetro de espaço k do Chia, que controla o tamanho de cada prova e espaço em disco que lhes é atribuído. Os tempos de execução de cada protocolo são semelhantes, para os parâmetros públicos definidos acima. O rácio

[3] https://github.com/Chia-Network/chiapos

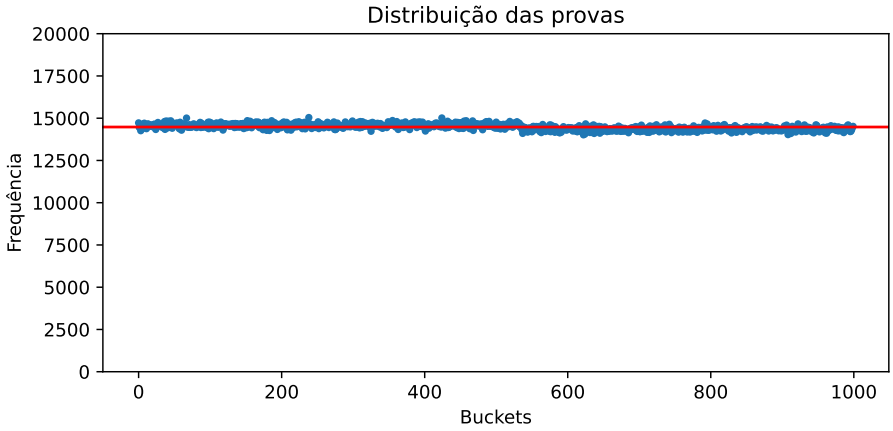

Fig. 6. Distribuição das provas após um processo de inicialização de 105GB (a azul) e a distribuição uniforme (a vermelho).

de espaço útil obtido é cerca de 0.5 incluindo metadados para restaurar os dados originais.

O tamanho da prova do nosso protocolo para os parâmetros definidos é 640 bytes, 2.5 vezes maior que a prova do Chia que é 256 bytes. Tendo em conta que o tamanho máximo de um bloco do Chia é 400KB, a prova do nosso protocolo continua a ter um tamanho prático uma vez que é apenas 0.16% do bloco.

De modo a analisar a entropia das provas, comparamos a distribuição obtida através das provas obtidas através de 105GB com a distribuição uniforme. Para obter a distribuição das provas, considerámos cada prova um número hexadecimal que foi categorizado entre 0 a 1000 e registámos a frequência de cada categoria (Figura 6). Aplicando o teste qui-quadrado de Pearson à distribuição das provas, obtemos um p-value de 6.75^{-57}, o que nos indica que as provas geradas pelo nosso protocolo são uniformemente distribuídas. Isto significa que as provas geradas têm um elevada entropia.

4.4 Discussão

Resposta à **Q1**: Os nossos resultados indicam que os dados aleatórios das PoSp blockchains podem ser substituídas com dados úteis. As limitações da substituição são os metadados que são gerados durante o processo de inicialização dos dados úteis. Estes metadados permitem a reposição dos dados originais e a verificação do espaço alocado. No nosso protocolo, o tamanho dos metadados depende diretamente do valor de *fanout* definido. No mínimo, a nossa construção permite um rácio de 50% de dados úteis. Este rácio pode ser incrementado, com valores mais elevados de fanout como ilustrado na Figura 5.

Resposta à **Q2**: A utilização de dados úteis como provas de espaço permite que mineiros com menos espaço livre participem na blockchain com os

seus próprios ficheiros. Porém, o tamanho das provas de espaço útil tende a ser maiores do que as provas de espaço usadas na prática. Provas de grande tamanho aumentam o tempo de transmissão de blocos no sistema e o espaço de armazenamento ocupado pela blockchain. No nosso protocolo, há um compromisso entre o tamanho das provas e o rácio de espaço útil.

5 Conclusão

Nas PoSp blockchains, os mineiros preenchem o seu espaço de armazenamento gerando dados criptográficos aleatórios com o único propósito de garantir a segurança da blockchain. Estas técnicas levam a um desperdício de armazenamento, uma vez que os dados armazenados não têm outro propósito para o mineiro para além da geração de provas criptográficas.

Este trabalho propõe uma abordagem que reduz a quantidade de desperdício de armazenamento das PoSp blockchains, utilizando dados locais (como ficheiros do utilizador) para gerar as provas. A solução baseia-se no protocolo do Chia e adapta-o substituindo os dados aleatórios das provas por dados úteis.

Resultados preliminares mostram que no mínimos 50% dos dados aleatórios das provas de espaço podem ser substituídas por dados úteis.

5.1 Trabalho Futuro

Como trabalho futuro temos como objetivo correr uma simulação da blockchain proposta de forma a avaliar o desempenho do sistema, tendo como métricas a latência e o débito. Adicionalmente, é importante ter uma análise formal da segurança do nosso protocolo. Assim, tencionamos obter uma demonstração formal da segurança do Pudim de Pão e Chia.

Agradecimentos Este trabalho foi suportado pela FCT – Fundação para a Ciência e a Tecnologia, através dos projectos UIDB/50021/2020 e Ainur (PTDC/CCI-COM/4485/2021), e desenvolvido no âmbito do projeto no. 51 “BLOCKCHAIN.PT - Agenda Descentralizar Portugal com Blockchain”, financiado por Fundos Europeus, nomeadamente “Plano de Recuperação e Resiliência - Componente 5: Agendas Mobilizadoras para a Inovação Empresarial”, incluído no programa de financiamento NextGenerationEU.