Uma versão simplificada do Bitcoin, implementada em Agda

Guilherme Horta Alvares da Silva Orientador: Flávio Codeço Coelho

Fundação Getulio Vargas

2019

Uma versão simplificada do Bitcoin, implementada em Agda

- Objetivo
- lustificativa
- Linguagem funcional e Agda
 - Linguagem Funcional
 - Lambda Calculus
 - Agda
- Criptomoedas
 - Definição
 - Funções criptográficas
 - Transações e árvore de transações
 - Blocos e cadeia de blocos
 - Bugs
- Próximos passos
- Referências Bibliográficas



Objetivo

• Programar uma criptomoeda (similar ao Bitcoin) em Agda, que é uma linguagem com tipos dependentes.



Justificativa

- Programar um protocolo de criptomoedas livre de erros (bugs).
- Utilizar Agda permite, além da programação da criptomoeda, especificar como ela deve funcionar (Norell, 2008).

Linguagem Funcional

- Em linguagens funcionais, toda função retorna a mesma saída para a mesma entrada.
- Em Agda, toda função é total e ela sempre termina.

Efeitos Colaterais

- Para Agda ser utilizada como um programa, ela tem que lidar com o sistema operacional. Operação chamada de IO (Input Output).
- Várias pesquisas estão sendo feitas para minimizar os efeitos colaterais do IO.
- Lógica linear é usada para lidar com recursos, como abrir arquivos.
- Efeitos algébricos são utilizados para distinguir os diferentes tipos de efeitos relacionados ao sistema operacional.

Efeitos Colaterais

- Frontend e backend de serviços estão sendo programados na mesma linguagem funcional para minimizar os efeitos colaterais.
- Haxl, usado no Facebook, é uma ferramenta que faz o cache das operações IO para fácil debug, log e teste.
- Facebook também faz Hot Swapping de forma funcional. Ou seja, ele não reinicia o processo para iniciar um novo. Algo que já era comum em Lisp.
- Sistemas distribuídos já estão sendo completamente descritos por meio de linguagens funcionais. Whatsapp utiliza Elixir como linguagem.

Benefícios

- Por causa dos inúmeros benefícios das linguagens funcionais, elas estão sendo cada vez mais utilizadas na indústria.
- Garantem corretude de software. São de fácil reprodutibilidade, testabilidade e mais fáceis de debugar.
- Entretanto, elas são mais difíceis de serem aprendidas.
 Necessitam alto grau de abstração. Portanto, é mais difícil de encontrar mão de obra que saiba utilizá-las.

Empresas



Criptomoedas







- Linguagem funcional com sistema de tipos expressivos capazes de representar as especificações.
- Possibilita especificar e programar em um único lugar.
- O processo de verificação acontece no compilador.

- A linguagem não possui Built-in como em Python.
- Tipos como inteiros, pontos flutuantes, strings e vetores devem ser definidos pelo próprio usuário.
- Tipos em Agda são uma generalização de tipos de dados algébricos encontrados em Haskell e ML.

Identidade

• Agda utiliza alguns conceitos do Lambda Calculus:

$$\mathsf{id'}: \{A:\mathsf{Set}\} \to A \to A \\ \mathsf{id'}\; x = x$$

Booleanos

• Definição de booleanos:

true :
$$\{A: \mathsf{Set}\} \to A \to A \to A$$

true $x \ y = x$
false : $\{A: \mathsf{Set}\} \to A \to A \to A$
false $x \ y = y$

Números naturais

• Definição de números naturais:

zero :
$$\{A : \mathsf{Set}\} \to (A \to A) \to A \to A$$

zero $\mathsf{suc} \ \mathsf{z} = \mathsf{z}$
one : $\{A : \mathsf{Set}\} \to (A \to A) \to A \to A$
one $\mathsf{suc} \ \mathsf{z} = \mathsf{suc} \ \mathsf{z}$
two : $\{A : \mathsf{Set}\} \to (A \to A) \to A \to A$
two $\mathsf{suc} \ \mathsf{z} = \mathsf{suc} \ (\mathsf{suc} \ \mathsf{z})$

Zero

```
isZero : \{A: \mathsf{Set}\} \to ((A \to A) \to A \to A) \to (A \to A \to A) isZero n true false = n (\lambda _- \to false) true isZero-zero : \{A: \mathsf{Set}\} \to \mathsf{Result} (isZero \{A\} zero) isZero-zero = res (\lambda true false \to true) isZero-two : \{A: \mathsf{Set}\} \to \mathsf{Result} (isZero \{A\} two) isZero-two = res (\lambda true false \to false)
```

Soma

plus :
$$\{A : \mathsf{Set}\} \to ((A \to A) \to A \to A)$$

 $\to ((A \to A) \to A \to A)$
 $\to ((A \to A) \to A \to A)$
plus $n = \lambda \ \mathsf{suc} \ z \to n \ \mathsf{suc} \ (m \ \mathsf{suc} \ z)$
 $-+_: \{A : \mathsf{Set}\} \to ((A \to A) \to A \to A)$
 $\to ((A \to A) \to A \to A)$
 $\to ((A \to A) \to A \to A)$
 $-+_ n \ m \ \mathsf{suc} \ z = n \ \mathsf{suc} \ (m \ \mathsf{suc} \ z)$
one+one = res $(\lambda \ \mathsf{suc} \ z \to \mathsf{suc} \ (\mathsf{suc} \ z))$

Lista

```
emptyList : {A List : Set}
    \rightarrow (A \rightarrow List \rightarrow List) \rightarrow List \rightarrow List
emptyList :: nil = nil
natList : {A List : Set}
    A 
ightarrow (((A 
ightarrow A) 
ightarrow A 
ightarrow A) 
ightarrow \mathit{List} 
ightarrow \mathit{List} 
ightarrow \mathit{List} 
ightarrow \mathit{List} 
ightarrow \mathit{List}
natList :: nil = one :: (two :: nil)
sumList : {A List : Set}
    \rightarrow Result (natList \{A\} \{(A \rightarrow A) \rightarrow A \rightarrow A\} + zero)
\mathsf{sumList} = \mathsf{res} \; (\lambda \; \mathsf{suc} \; \mathsf{z} \to \mathsf{suc} \; (\mathsf{suc} \; (\mathsf{suc} \; \mathsf{z})))
```

Soma

• Definição de tipos de soma:

left:
$$\{A \ B \ C : \mathsf{Set}\} \to A \to (A \to C) \to (B \to C) \to C$$

left $x \ f \ g = f \ x$
right: $\{A \ B \ C : \mathsf{Set}\} \to B \to (A \to C) \to (B \to C) \to C$
right $x \ f \ g = g \ x$

Tipo soma

```
zero-isZero : \forall \{A\}
   \rightarrow Result (zero-left \{A\} isZero id)
zero-isZero = res (\lambda true false \rightarrow true)
one-isZero : \forall \{A\}
   \rightarrow Result (one-left \{A\} isZero id)
one-isZero = res (\lambda true false \rightarrow false)
false-id : \forall \{A\}
   \rightarrow Result (false-right \{(A \rightarrow A) \rightarrow A \rightarrow A\} isZero id)
false-id = res (\lambda true false \rightarrow false)
true-id : \forall \{A\}
   \rightarrow Result (true-right \{(A \rightarrow A) \rightarrow A \rightarrow A\} is Zero id)
true-id = res (\lambda true false \rightarrow true)
```

Tuplas

• Definição de tupla:

tuple :
$$\{A \ B \ C : \mathsf{Set}\} \to A \to B \to (A \to B \to C) \to C$$

tuple $x \ y \ f = f \ x \ y$

zero-false :
$$\{A \ B \ C : \mathsf{Set}\} \to (((A \to A) \to A \to A) \to (B \to B \to B) \to C) \to C$$

zero-false = tuple zero false

one-true :
$$\{A \ B \ C : \mathsf{Set}\} \to (((A \to A) \to A \to A) \to (B \to B \to B) \to C) \to C$$

one-true = tuple one true

Adição de tuplas

```
add-true : \{A : \mathsf{Set}\} \to ((A \to A) \to A \to A)
   \rightarrow (A \rightarrow A \rightarrow A) \rightarrow ((A \rightarrow A) \rightarrow A \rightarrow A)
add-true n b suc z = b (suc (n suc z)) (n suc z)
add-zero-false : {A : Set}
   \rightarrow Result (zero-false \{(A \rightarrow A) \rightarrow A \rightarrow A\} add-true)
add-zero-false = res (\lambda suc z \rightarrow z)
add-one-true : \forall \{A\}
   \rightarrow Result (one-true \{(A \rightarrow A) \rightarrow A \rightarrow A\} add-true)
add-one-true = res (\lambda suc z \rightarrow suc (suc z))
```

• Definição de números naturais em Agda:

• data N : Set where

zero : ℕ

 $\mathsf{suc}:\,\mathbb{N}\to\mathbb{N}$

• Eliminação dos números naturais em Agda:

```
    N-elim: (target: N) (motive: (N → Set))
        (base: motive zero)
        (step: (n: N) → motive n → motive (suc n))
        → motive target
        N-elim zero motive base step = base
        N-elim (suc target) motive base step =
        step target (N-elim target motive base step)
```

• Definição de isomorfismo:

```
• record \_\simeq\_ (A \ B : Set) : Set where field

to : A \to B

from : B \to A

fromoto : \forall (x : A) \to from (to x) \equiv x

toofrom : \forall (y : B) \to to (from y) \equiv y
```

• Definição de número binário:

```
data Bit : Set where
     0# : Bit
     1# : Bit
   Bin<sup>+</sup>: Set
   Bin^+ = I ist Bit
  data Bin: Set where
     0# : Bin
     1\# : (bs : Bin^+) \rightarrow Bin
   postulate Bin \simeq \mathbb{N}: Bin \simeq \mathbb{N}
```

• Adição usando a recursão:

• _+'_ :
$$\mathbb{N} \to \mathbb{N} \to \mathbb{N}$$

zero +' $m = m$
suc $n +' m = \text{suc } (n + m)$

• Adição usando a eliminação dos naturais:

$$\bullet \ _+_ : \mathbb{N} \to \mathbb{N} \to \mathbb{N}$$

$$n + m = \mathbb{N} \text{-elim } n \ (\lambda \ _ \to \mathbb{N}) \ m \ \lambda \ _ \ v \to \text{suc } v$$

```
    data ⊥ : Set where
```

```
\perp-elim : \{A : \mathsf{Set}\}\ (bot : \perp) \to A
\perp-elim ()
```

```
data Either {I: Level} (A : Set I) (B : Set I) : Set I where
  left : (I:A) \rightarrow Either A B
  right : (r: B) \rightarrow Either A B
Either-elim : { I I2 : Level } { A B : Set I }
  \{motive : (eab : Either A B) \rightarrow Set 12\}
  (target : Either A B)
  (on-left: (I: A) \rightarrow (motive (left I)))
  (on-right: (r: B) \rightarrow (motive (right r)))
  \rightarrow motive target
Either-elim (left I) onleft onright = onleft I
Either-elim (right r) onleft onright = onright r
```

- Tipo booleano em Agda:
 - Bool : SetBool = Either ⊤ ⊤

• Usando tipos booleanos em casos:

```
• if_then_else_ : {I : Level} {A : Set I}

(b : Bool) (tRes\ fRes : A) \rightarrow A

if b then tRes\ else\ fRes =

Either-elim b (\lambda \rightarrow tRes) \lambda \rightarrow tRes
```

• Eliminação do para todo:

- Tipo multiplicação:
 - data \sum (A : Set) (B : $A \rightarrow$ Set) : Set where $\langle _, _ \rangle$: (x : A) \rightarrow B x \rightarrow \sum A B $\sum \text{-elim} : \forall \{A : \text{Set}\} \{B : A \rightarrow \text{Set}\} \{C : \text{Set}\}$ $\rightarrow (\forall x \rightarrow B x \rightarrow C)$ $\rightarrow \sum A B$ ------ $\rightarrow C$ $\sum \text{-elim} f \langle x, y \rangle = f x y$

Tipo vetor:

```
• data Vec : \mathbb{N} \to \mathsf{Set} where 

[] : Vec zero 

_::_ : \{ \mathit{size} : \mathbb{N} \} \to \mathbb{N} \to \mathsf{Vec} \mathit{size} \to \mathsf{Vec} (suc \mathit{size}) 

nil : Vec zero 

nil = [] 

vec-one : Vec (suc zero) 

vec-one = zero :: nil
```

- Encontrando primeiro elemento do vetor:
 - head : $\{A: \mathsf{Set}\}\ \{n: \mathbb{N}\}\ (\mathit{vec}: \mathsf{Vector}\ A\ (\mathsf{suc}\ n)) \to A$ head $(x:: \mathit{vec}) = x$

• Tipos booleanos:

• data Boolean : Set where

true : Boolean false : Boolean

- Funções que retornam tipo:
 - boolean→Set : (b : Boolean) → Set boolean→Set true = \mathbb{N} boolean→Set false = Bool

Utilizando records:

```
• record Person : Set where constructor person field name : String age : \mathbb{N}

agePerson : (person : Person) \rightarrow \mathbb{N} agePerson (person : name : age) = age
```

- Demonstrando que elementos podem ser encontrados:
 - module WF $\{A : Set\}$ (_ <_ : Rel A) where data Acc (x : A) : Set where acc : $(\forall y \rightarrow y < x \rightarrow Acc y) \rightarrow Acc x$

Well-founded : Set Well-founded = $\forall x \rightarrow Acc x$

- Definindo o que é uma categoria:

Definição

- Uma criptomoeda é um meio de troca descentralizado que se utiliza da tecnologia de blockchain e da criptografia para assegurar a validade das transações e a criação de novas unidades da moeda.
- O Bitcoin é considerado a primeira moeda digital mundial descentralizada, constituindo um sistema econômico alternativo (peer-to-peer electronic cash system) e responsável pelo ressurgimento do sistema bancário livre. (Nakamoto et al., 2008)
- O Bitcoin permite transações financeiras sem intermediários, mas verificadas por todos usuários (nodos da rede). Essas transações são gravadas em um banco de dados distribuídos (uma rede descentralizada), chamado de blockchain.

Função hash

- Uma função hash serve para comprimir dados grandes em dados pequenos.
- Assume-se que dois arquivos diferentes sempre terão hashes diferentes (injetividade), pois a probabilidade de dois arquivos terem o mesmo hash é desprezível.
- Pelo princípio da casa dos pombos, a injetividade do hash é impossível. Porém, nunca foram encontrados dois arquivos que possuem o mesmo hash no SHA-256.
- O Google encontrou dois arquivos que possuíam o mesmo hash no SHA-1. Logo, em seguida, essa função hash parou de ser usada.
- Também não deve ser possível calcular a inversa da função hash.

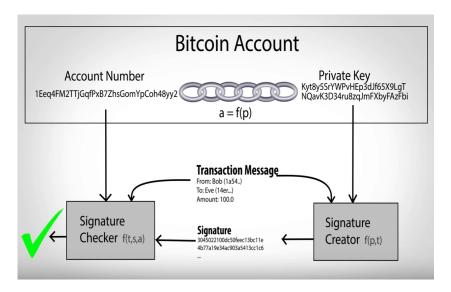
Funções criptográficas

- No Bitcoin, existe a chave privada e a pública.
- A chave privada serve para assinar a transação.
- A chave pública, derivada da chave privada, serve para demonstrar que o usuário é o dono da chave privada.
- Essas funções utilizam como base a função hash SHA-256.

Funções criptográficas em Agda

```
postulate priv\equivpub : PrivateKey \rightarrow PublicKey \rightarrow Set
postulate publicKey2Address : PublicKey → Address
postulate Signed : Msg \rightarrow PublicKey \rightarrow Signature \rightarrow Set
postulate Signed? : (msg : Msg) (pk : PublicKey)
             (sig : Signature) \rightarrow Dec \$ Signed msg pk sig
postulate hashMsg : Msg \rightarrow Hashed
postulate hash-inj : \forall m \ n \rightarrow \text{hashMsg } m \equiv \text{hashMsg } n
             \rightarrow m \equiv n
record SignedWithSigPbk (msg : Msg)(address : Address)
  : Set where
  field
    publicKey: PublicKey
    pbkCorrect: publicKey2Address publicKey \equiv address
    signature : Signature
    signed : Signed msg publicKey signature
```

Conta do Bitcoin



Transações

- A partir de transações, é possível enviar Bitcoins de uma conta para outra.
- Transações são como um cheque. O indivíduo especifica um valor e assina a transação.
- Na transação do Bitcoin, as transações anteriormente não gastas devem ser especificadas. No caso da transação do minerador, isso não deve ser especificado. Além do mais, deve ser especificado quem deve receber o dinheiro dessas transações e também a transação deve ter uma assinatura gerada a partir da chave privada comprovando que o usuário da chave pública aceitou fazer aquela transação.

Transações em Agda

```
record TXSigned
  {time : Time}
  {outSize : Nat}
  {outAmount : Amount}
  (inputs : List TXFieldWithId)
  (outputs : VectorOutput time outSize outAmount) : Set where
  constructor txsig
  field
    nonEmpty: NonNil inputs
    signed: All
      (\lambda input \rightarrow
      SignedWithSigPbk (txEls→MsgVecOut input outputs)
                         (TXFieldWithId.address input))
                         inputs
    in>out: txFieldList \rightarrow TotalAmount inputs \ge outAmount
```

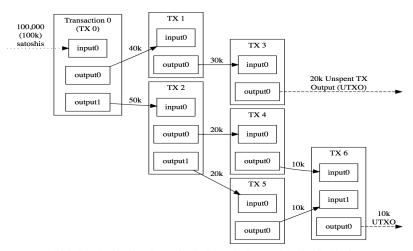
Modelo de transações

- O Bitcoin utiliza o modelo de transações não gastas enquanto o Ethereum utiliza o tradicional modelo bancário.
- No modelo bancario, cada conta possui um saldo. E em toda transação, o saldo da pessoa que enviou é subtraído e o saldo da pessoa que recebeu é incrementado.
- Nesse modelo tradicional, é fácil verificar o saldo de cada um. Porém é difícil saber como se chegou nesse estado final.

Modelo de transações não gastas

- No modelo de transações não gastas, toda transação é adicionada à árvore de transações.
- Para saber o saldo de uma conta é necessário olhar todas as transações não gastas enviadas para essa conta.
- Para transacionar a moeda é necessário utilizar como entrada as saídas das outras transações não gastas.

Bitcoin UTXO

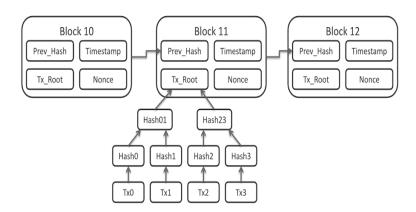


Triple-Entry Bookkeeping (Transaction-To-Transaction Payments) As Used By Bitcoin

Árvores de transação

- A ideia da árvore de transação é reunir todas as transações em uma árvore.
- Dessa forma, é possível sintetizar a informação de todas as transações calculando apenas o hash dela.
- Com a propriedade da injetividade do hash, é possível verificar que duas árvores são iguais em complexidade de tempo constante.

Árvore de transação



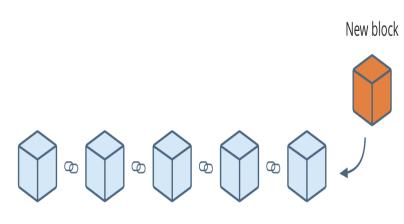
```
mutual
  data TXTree: (time: Time) (block: Nat)
    (outputs : List TXFieldWithId)
    (totalFees : Amount)
    (qtTransactions : tQtTxs) \rightarrow Set where
    genesisTree: TXTree (nat zero) zero [] zero zero
    txtree :
      {block : Nat} {time : Time}
      {outSize : Nat} {amount : Amount}
      {inputs : List TXFieldWithId}
       [outputTX : VectorOutput time outSize amount]
      {totalFees : Amount} {qtTransactions : tQtTxs}
      (tree : TXTree time block inputs totalFees gtTransactions)
      (tx: TX {time} {block} {inputs} {outSize} tree outputTX)
       proofLessQtTX:
         Either
           (IsTrue (lessNat (finToNat qtTransactions) totalQtSub1))
           (isCoinbase tx))
      → TXTree (sucTime time)
         (nextBlock tx)
         (inputsTX tx ++ VectorOutput \rightarrow List outputTX)
         (incFees tx) (incQtTx tx proofLessQtTX)
```

```
data TX {time : Time} {block : Nat} {inputs : List TXFieldWithId}
     {outSize : Nat} {outAmount : Amount}
     {totalFees : Nat} {qtTransactions : tQtTxs}
  : (tr : TXTree time block inputs totalFees qtTransactions)
    (outputs : VectorOutput time outSize outAmount) → Set where
  normalTX:
    (tr : TXTree time block inputs totalFees qtTransactions)
    (SubInputs : SubList inputs)
    (outputs : VectorOutput time outSize outAmount)
    (txSigned : TXSigned (sub→list SubInputs) outputs)
    \rightarrow TX tr outputs
  coinbase :
    (tr: TXTree time block inputs totalFees qtTransactions)
    outputs: VectorOutput time outSize outAmount)
    (pAmountFee : outAmount out≡Fee totalFees +RewardBlock block)
    \rightarrow TX tr outputs
```

Blocos e cadeia de blocos

- As transações são incorporadas em blocos aonde são armazenadas.
- No Bitcoin, existe uma cadeia de blocos chamada de blockchain. O primeiro bloco é chamado de genesis block e a cada 10 minutos em média, um novo bloco é incorporado a essa cadeia de blocos.
- Existe um limite de transações que cada bloco pode possuir (no Bitcoin, é 10 MB de dados). Por isso, é importante pagar uma taxa de transação para que essa transação seja incluída na cadeia de blocos.
- Se existirem duas blockchains em computadores diferentes, considera-se a cadeia de blocos válida aquela que tiver mais blocos.

Blockchain

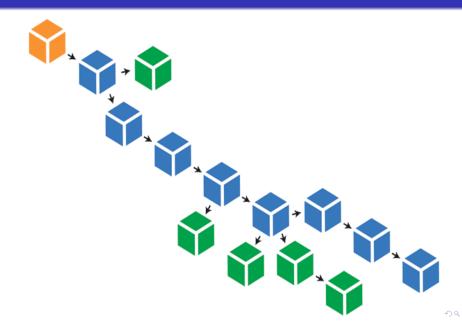


Starting block





Blockchain



Mineração

- Todo bloco possui um problema criptográfico a ser resolvido.
 Ou seja, é preciso calcular um valor nounce tal que o hash do bloco seja menor que o um valor especificado.
- Esse valor especificado é calculado a partir da média dos últimos 2016 blocos (2 semanas) para que a média de mineração de blocos seja de 10 minutos.

```
record Block
  \{block_1 : Nat\}
  {time<sub>1</sub> : Time}
   {outputs1 : List TXFieldWithId}
   {totalFees1 : Amount}
  \{qtTransactions_1 : tQtTxs\}
   (t \times Tree_1 : T \times Tree time_1 block_1 outputs_1 totalFees_1 qtTransactions_1)
  {time<sub>2</sub> : Time}
  {outputs<sub>2</sub> : List TXFieldWithId}
  {totalFees2 : Amount}
  {qtTransactions2 : tQtTxs}
   (txTree<sub>2</sub>: TXTree time<sub>2</sub> block<sub>1</sub> outputs<sub>2</sub> totalFees<sub>2</sub> qtTransactions<sub>2</sub>)
  : Set where
  constructor blocke
  field
     nxTree : nextTXTree txTree<sub>1</sub> txTree<sub>2</sub>
     fstBlock: firstTreesInBlock txTree1
     sndBlockCoinbase : coinbaseTree txTree2
```

DAO Bug

- Bug que aconteceu em um cripto-contrato da rede Ethereum com um prejuízo de mais do que 250 milhões de dólares. (Wood et al., 2014)
- No cripto-contrato, existia uma função recursiva que não terminava. Ou seja, o usuário enviava uma quantidade de ethereum, depois acontecia um loop infinito e só depois era feito a atualização do seu balanço.
- Em Agda, esse tipo de bug seria evitado, pois é necessário provar que a função termina. Logo, loops infinitos não são possíveis em Agda.

Conclusões

- Nesse trabalho, todas as especificações foram colocadas nos tipos.
- Em outros trabalhos, a abordagem pode ser diferente. Ou seja, sem nenhuma informação nos tipos e todas as provas feitas de forma separada.
- Essa abordagem não foi escolhida, pois é muito mais simples de entender o código quando os tipos são mais expressivos.

Detalhes do que falta

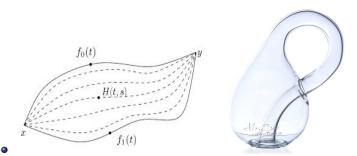
- Fazer o cálculo do hash do bloco e o campo nounce do bloco.
- Mudar o protocolo para receber o bloco inteiro ao invés de transações individuais.
- Adicionar o caso de existir uma blockchain de tamanho maior que a do que a do servidor.

Partes mais complexas de adicionar

- Fazer a modelagem formal da cadeia de blocos como um sistema distribuído.
- Criar código do cliente para interagir com os nós.
- Especificar o protocolo entre cliente e nó e entre nó com nó.
 Entretanto, a parte mais difícil já foi codificada, que é adicionar as provas ao dado puro que o cliente envia.
- Adicionar uma linguagem de script na criptomoeda.

Outras ideias

 Melhorar o sistema de tipos. Por enquanto, está se usando tipos dependentes, mas seria interessante utilizar tipos lineares para evitar o coletor de lixo ou usar *Cubical Type Theory* por ser mais expressivo.



Outras ideias

- Utilizar outros provadores de teoremas. O Lean 4 e CoQ possuem táticas, algo que Agda não possui. O Idris 2 possui tipos lineares.
- Utilizar teorias das categorias para deixar o código mais abstrato e conciso. Além de poder utilizar as bibliotecas já implementadas da linguagem.
- Utilizar bibliotecas para facilitar a transferência de mensagens entre nós diferentes. Cloud Haskell já facilita bastante esse processo.

Referências Bibliográficas

- Nakamoto, S., et al. (2008). Bitcoin: A peer-to-peer electronic cash system.
- Norell, U. (2008). Dependently typed programming in agda. In *International school on advanced functional programming* (pp. 230–266).
- Wood, G., et al. (2014). Ethereum: A secure decentralised generalised transaction ledger. *Ethereum project yellow paper*, *151*, 1–32.