UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Eric Patrick Silva dos Santos

Uso do algoritmo H1 na *Forense* da *Blockchain*

Uberlândia, Brasil

2022

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Eric Patrick Silva dos Santos

Uso do algoritmo H1 na *Forense* da *Blockchain*

Trabalho de conclusão de curso apresentado à Faculdade de Computação da Universidade Federal de Uberlândia, Minas Gerais, como requisito exigido parcial à obtenção do grau de Bacharel em Sistemas de Informação.

Orientador: Ivan da Silva Sendin

Universidade Federal de Uberlândia – UFU

Faculdade de Computação

Bacharelado em Sistemas de Informação

Uberlândia, Brasil

2022

Eric Patrick Silva dos Santos

Uso do algoritmo H1 na *Forense* da *Blockchain*

Trabalho de conclusão de curso apresentado à Faculdade de Computação da Universidade Federal de Uberlândia, Minas Gerais, como requisito exigido parcial à obtenção do grau de Bacharel em Sistemas de Informação.

Trabalho aprovado. Uberlândia, Brasil, 18 de agosto de 2022:

Ivan da Silva Sendin
Orientador

Rodrigo Sanches Miani
Professor

Silvio Ereno Quincozes
Professor

Uberlândia, Brasil
2022

Agradecimentos

Agradeço, primeiramente a Deus e a todas as pessoas que me ajudaram nessa caminhada.

Aos meus familiares em especial meus pais, a minha madrinha Marilene e meu tio Rovaldo por me proporcionar essas oportunidades.

Agradeço a meu orientador Ivan Sendin por me auxiliar nesse trabalho.

Gostaria de agradecer aos meus amigos que me ajudaram na minha caminhada na faculdade.

"O insucesso é apenas uma oportunidade para recomeçar com mais inteligência."

- Henry Ford

Resumo

Nos últimos anos, as criptomoedas estão cada vez mais populares entre os usuários e estão sendo utilizadas em atividades lícitas e ilícitas. Sendo assim, é necessário desenvolver ferramentas e técnicas para analisar esse ecossistema das criptomoedas, a fim de fiscalizar as atividades realizadas nesse contexto. Nesse trabalho, serão apresentados alguns conceitos da *Blockchain* e da criptomoeda *Bitcoin*. Além disso, serão citadas algumas atividades ilegais que utilizam o *Bitcoin* como forma de pagamento. Em seguida, serão apresentados algumas ferramentas de análise da *Blockchain* e como são realizadas essas análises. Posteriormente, será apresentado o algoritmo H1 que é um algoritmo de agrupamento que utiliza os endereços de entrada de uma transação para produzir *clusters* de endereços que muito provavelmente pertencem a mesma entidade no mundo real. Este trabalho é finalizado com a implementação e aplicação do algoritmo H1 em um caso real de forense. Além disso, com os resultados obtidos nesse trabalho foram adicionados novos endereços aos endereços iniciais extraídos do caso analisado.

Palavras-chave: Blockchain, Bitcoin, Forense, Algoritmo H1, Transações, Endereços.

Lista de ilustrações

Lista de tabelas

Lista de abreviaturas e siglas

BTC	Bitcoin
KYC	Know Your Customer
API	Application Programming Interface
UTXO	Unspent Transaction Outputs
JSON	JavaScript Object Notation

Sumário

1 Introdução

Desde a popularização do capitalismo como modelo sócio-político-econômico, os povos utilizam várias formas diferentes de realizar pagamentos. Para facilitar esses serviços financeiros, foram criadas instituições financeiras. Nos dias de hoje, os bancos oferecem diversos serviços financeiros, inclusive transações de pagamento, todavia, os custos das operações costumam ser altos para os clientes (Júlia Lewgoy, 2021).

Os bancos funcionam como uma entidade mediadora de todas as transações entre os usuários. Os clientes dos bancos, nesse contexto, não têm a liberdade desejada para movimentação de suas finanças, pois o processo de realização das operações é, frequentemente, altamente burocrático.

Nos últimos anos, foram criadas várias soluções digitais para resolver esse problema de uma entidade mediadora para validar operações de transações financeiras. Uma delas é a *Blockchain*: uma cadeia de blocos dependentes cuja alteração do seu conteúdo é impraticável. Mais detalhes técnicos sobre a *Blockchain* serão vistos na Seção 2.1.

A seguir, foram definidas algumas características usadas para exemplificar uma aplicação da *Blockchain*:

- Através do histórico, é possível descobrir o saldo disponível de cada usuário da rede e as transações bancárias para um determinado usuário;
- Essa transferência ocorrerá digitalmente;
- Usaremos um histórico para registrar as transações e todo usuário que entrar na rede baixará a última versão do histórico;
- Se a maior parte da rede aceita a transação, essa será registrada no histórico.

Diante das regras, temos o seguinte contexto: Ana, Jade, Luís, Eric e Augusto são os usuários dessa rede. Os valores disponíveis nas suas carteiras digitais são, respectivamente, 100, 0, 0, 0, 0. Supondo que Ana transfira 100 para Jade, logo os usuários da rede validam essa transação e o histórico é atualizado.

Sendo assim, Ana tem 0 no saldo disponível e Jade tem 100. Se Eric tentar transferir 10 para Luís, a transação não será validada, pois os usuários da rede sabem, através do histórico, que esse cliente não tem o saldo disponível e, mesmo que Eric e Luís mudem seus históricos, não terá efeito sobre a rede, porque será validado somente o que a maioria dos integrantes da rede definirem em consenso, por meio da validação pelos clientes da rede.

Para isso, será utilizado o histórico atualizado que cada usuário tem disponível no seu computador. Assim, podemos definir a *Blockchain* como um tipo de livro razão de transações comerciais.

Uma rede *Blockchain* é um sistema descentralizado para a troca de ativos. Nele, utiliza-se de um livro razão público, que é um registro de lançamento com a finalidade de coletar dados cronológicos de todas as transações que são realizadas (Carin Tom, 2020). Além disso, é compartilhado para registrar o histórico de transações comerciais eletrônicas que ocorrem em uma rede de negócios *peer-to-peer* (P2P). A *Blockchain* pode usar prova de trabalho ou outro mecanismo de consenso para obter confiança e transparência. Dessa maneira, não é necessária a dependência de uma instituição financeira ou *Trusted Third Party* (Terceira Parte Confiável) (Sloane Brakeville with Bhargav Perepa, 2020).

A *Blockchain* pode ser aplicada em diversos campos como saúde (Arlindo F. da Conceição Vladimir Moreira Rocha, 2019), instituições financeiras e governamentais, contratos diversos e todas as áreas onde existe a necessidade de troca de informações rápidas e seguras (UFRJ, 2018a). Existe uma confiança, pois indivíduos maliciosos têm maiores dificuldades em comprometer o sistema. Além disso, essa ferramenta possui inúmeras vantagens como segurança, rapidez, menor custo de tranferência, transparência e privacidade.

A aplicação mais famosa da *Blockchain* é o *Bitcoin*, uma moeda digital descentralizada que não necessita de terceiros para funcionar (Satoshi Nakamoto, 2008). Seu maior objetivo é a recuperação do controle do dinheiro de grandes instituições financeiras e governos. A princípio, essas tecnologias foram pensadas para o bem das pessoas e da sociedade, mas essas ferramentas estão sendo usadas para uma ampla variedade de atividades financeiras duvidosas, como lavagem de dinheiro.

Nesse trabalho foi desenvolvido uma aplicação para analisar a *Blockchain* do *Bitcoin* visando detectar endereços que pertençam à mesma entidade. Essa aplicação utiliza o algoritmo H1 que será apresentado na seção 2.4.2. Esse algoritmo produz agrupamentos, identificando possíveis endereços que estiverem relacionados. Esse algoritmo será aplicado nos endereços que foram disponibilizados pela investigação promovida pelo Ministério Público e Polícia Federal conhecida como “Faraó dos Bitcoins”. Esses endereços disponibilizados pela notícia são nomeados nesse trabalho como os endereços em análise. Sendo assim, são utilizados esses endereços em análise como ponto de partida para identificarmos outros endereços que possivelmente estão relacionados aos endereços em análise. O objetivo desse trabalho é encontrar novos endereços suspeitos de pertencer a uma entidade que será apresentada na seção 3.

2 Revisão Bibliográfica

2.1 *Blockchain*

O *Bitcoin* (Satoshi Nakamoto, 2008) é a criptomoeda mais popular em todo o mundo. O *Bitcoin* usa tecnologia *peer-to-peer* para operar sem um orgão responsável por controlar, intermediar e autorizar emissões de moedas, transferências e outras operações (Bitcoin Project, 2009). O gerenciamento de transações e a emissão de *Bitcoin* são realizados coletivamente pela rede. Ele fornece um pseudoanonimato aos seus usuários estabelecendo uma identidade que usa chaves públicas como pontos finais de uma transação. Essas transações são registradas em um livro razão público denominado *Blockchain*, que é uma estrutura de dados somente anexada.

A estrutura de dados da *Blockchain* é uma lista ordenada de blocos de transações, nesse contexto, cada bloco é "ligado" ao seu antecessor por uma função de *hashing* criptográfica. A *Blockchain* pode ser armazenada como um arquivo ou em banco de dados (Andreas M. Antonopoulos, 2020). É frequentemente visualizada como um empilhamento vertical, assim, os blocos são empilhados uns sobre os outros onde o primeiro bloco serve como fundação, suportando a pilha. O uso da função de *hashing* criptográfica faz com que o bloco de altura n 'valide' os seus antecessores, tornado a sua alteração - para todos os efeitos práticos - impossível.

2.2 Formação da *Blockchain*

A *Blockchain* é formada por uma cadeia de blocos dependentes. Esses blocos contém dados que reúnem as transações para a inclusão no registro público. A Figura 2 descreve a cadeia de blocos.

Tamanho	Campo	Descrição
4 bytes	Tamanho do bloco	O tamanho do bloco, em bytes, após esse campo
80 bytes	Cabeçalho do bloco	Vários campos formam o cabeçalho do bloco
1-9 bytes	Contador de transações	Quantidade de transações
Variável	Transações	Transações registradas nesse bloco

Tabela 1 – A estrutura de um bloco. Fonte: próprio autor

A estrutura do bloco consiste em 4 campos, sendo tamanho do bloco, cabeçalho do bloco, contador de transações e transações. A Tabela 1 descreve a estrutura de um bloco.

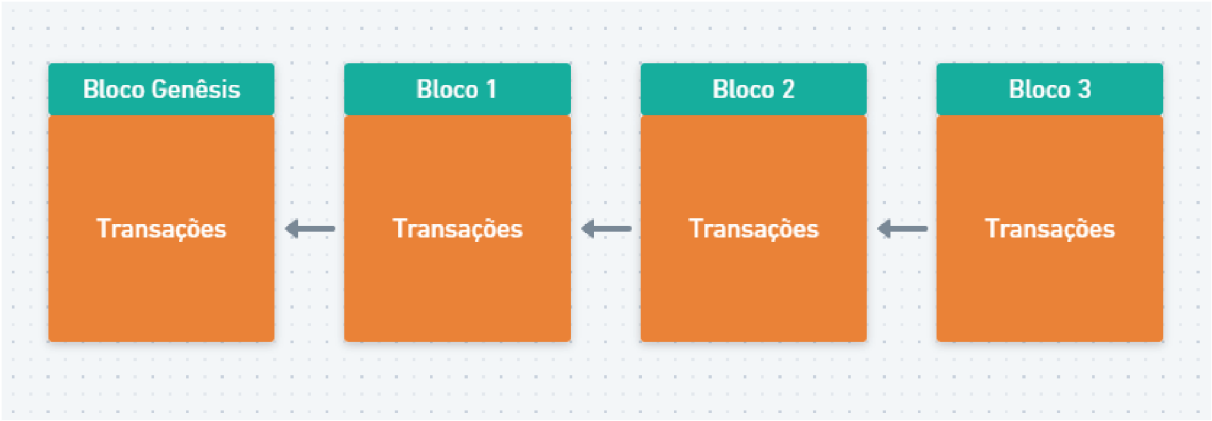

Figura 1 – Cadeia de blocos na *Blockchain*. Fonte: próprio autor

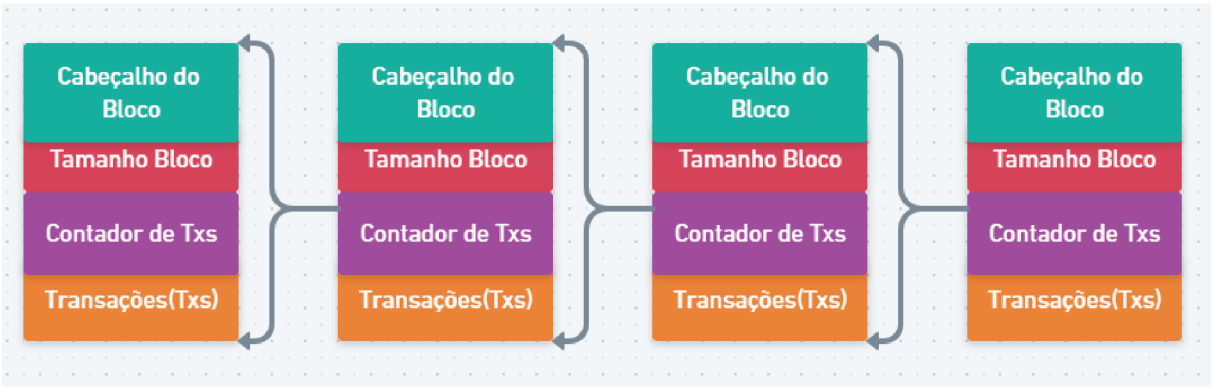

Figura 2 – Cadeia de blocos na *Blockchain* detalhada. Fonte: próprio autor

O cabeçalho do bloco consiste em três conjuntos de metadados de bloco. Primeiro, existe um campo que faz referência ao *hash* do bloco anterior, que conecta esse bloco ao bloco anterior na *Blockchain*. Então, temos o conjunto de metadados, a dificuldade, a data e hora *(timestamp)* e o *nonce*, relacionados à competição da mineração. A terceira parte dos metadados consiste na raiz da árvore de Merkle, uma estrutura de dados usada para resumir de maneira eficiente todas as transações contidas no bloco (Andreas M. Antonopoulos, 2020). A Tabela 2 descreve o cabeçalho de um bloco.

Para construir o cabeçalho do bloco, o nó minerador precisa preenchê-lo em seis campos, como listado na Tabela 2. Primeiramente, é necessário preencher o número da versão do bloco. A seguir, o nó de mineração precisa adicionar o 'Hash do Bloco Anterior' ao bloco anteriormente recebido pela rede. O próximo passo é fazer um resumo de todas as transações com uma árvore de Merkle para adicionar a raiz de Merkle ao cabeçalho do bloco (UFRJ, 2018b). As árvores de Merkle são usadas no *Bitcoin* para resumir todas as transações em um bloco, produzindo uma impressão digital eletrônica geral de todo o conjunto de transações. Esse método fornece um processo muito eficiente para verificar se

Tamanho	Campo	Descrição
4 bytes	Versão	Um número de versão para servir como referência nas atualizações de software/protocolo
32 bytes	Hash do bloco anterior	Uma referência ao hash do bloco anterior na *Blockchain*
32 bytes	Raiz de Merkle	Um hash da raiz da árvore de merkle das transações desse bloco
4 bytes	Data e Hora (Timestamp)	O momento aproximado em que este bloco foi criado (em segundos)
4 bytes	Dificuldade Alvo	A dificuldade do algoritmo de prova de trabalho deste bloco
4 bytes	Nonce	Um número arbitrário usado para o algoritmo de prova de trabalho

Tabela 2 – A estrutura do cabeçalho do bloco. Fonte: próprio autor

uma transação foi incluída em um bloco.

Uma árvore de Merkle é construída através do *hashing* recursivo de pares de nós até que exista apenas um *hash*, conhecido como a "raiz" ou "raiz de Merkle". O algoritmo de *hash* criptográfico usado nas árvores de Merkle do *Bitcoin* é o SHA256 aplicado duas vezes, também conhecido como SHA256-duplo.

A primeira transação adicionada ao bloco é uma transação especial, chamada de "transação de geração" ou "transação *coinbase*". Essa transação é construída no nó minerador e é a recompensa pelos esforços de mineração. O nó minerador cria a transação *coinbase* como um pagamento para a sua própria carteira.

A quantia total da recompensa que o minerador coleta por minerar um bloco é a soma da recompensa da transação *coinbase* com as taxas de transação de todas as transações incluídas no bloco.

O nó de mineração irá, então, adicionar um *timestamp* de 4 *bytes*, codificada como uma *timestamp* Unix "Epoch", que é baseada no número de segundos decorridos desde à meia noite de 1º de janeiro de 1970 UTC/GMT. A seguir, o nó de mineração preenche a dificuldade do alvo que define a dificuldade da prova de trabalho exigida para tornar esse bloco válido. A dificuldade é armazenada no bloco utilizando a métrica de "bits de dificuldade" ("difficulty bits"), que é uma codificação mantissa-expoente do alvo. A codificação tem um expoente de 1 byte, seguido por uma mantissa de 3 bytes.

O campo final é o *nonce* criptográfico, que começa com zero. No momento em que todos os outros campos estiverem preenchidos, o cabeçalho do bloco estará completo e o processo de mineração pode começar. O objetivo, agora, é encontrar um valor para o *nonce* que resulta em um *hash* de cabeçalho de bloco que é menor do que a dificuldade

alvo. O nó de mineração terá que testar bilhões ou trilhões de valores de *nonce* até que encontre um que satisfaça as exigências.

Os resultados da função de *hash* não podem ser determinados previamente, nem é possível criar um padrão que irá produzir um valor de *hash* específico. Essa propriedade das funções de *hash* indicam que a única maneira de se produzir um resultado de *hash* que corresponde a um alvo específico é através de sucessivas tentativas, modificando aleatoriamente o *input* até que resulte em um *hash* ideal.

A prova de trabalho do *Bitcoin* faz o minerador construir um bloco candidato. A seguir, o minerador calcula o *hash* do cabeçalho desse bloco e verifica se ele é menor do que o alvo atual. Se o *hash* não for menor do que o alvo, o minerador irá modificar o *nonce* (geralmente somando o número 1) e tentar novamente. Na dificuldade atual da rede *Bitcoin*, os mineradores tem que fazer várias tentativas até achar um *nonce* que resulte em um *hash* de cabeçalho de bloco baixo o suficiente (Andreas M. Antonopoulos, 2020).

Para obter mais informações relacionadas a esse tópico acesse o livro Antonopoulos, o *whitepaper* e o site oficial do *Bitcoin*(Bitcoin Project, 2009).

2.3 Transações

As transações permitem que os usuários realizem as transferências do *Bitcoin* para outros usuários da rede. Além disso, cada transação é construída a partir de várias partes que permitem pagamentos simples e diretos. Uma vez que a transação é validada, o registro dessa transação é adicionado a *Blockchain*, que armazena o registro de todas as transações já feitas na rede *Bitcoin*. Toda transação envolvendo *Bitcoin* que está registrada na *Blockchain* passou pelo processo de mineração.

Um endereço *Bitcoin* é uma string de dígitos e caracteres que pode ser compartilhada com qualquer pessoa que queira lhe enviar dinheiro. Os endereços são produzidos a partir de chaves públicas, através do uso de *hashing* criptográfico.

De forma simplificada, podemos ver um endereços *Bitcoin* como o *hash code* de uma chave pública, este *hash code* pode ser compartilhado com qualquer pessoa que queira lhe enviar dinheiro. O endereço *Bitcoin* é o que mais comumente aparece como ‘destinatário‘ dos fundos em uma transação.

Em uma transação existem uma lista de entrada que representam os valores em *Bitcoin* que os endereços receberam valores em transações anteriores e uma lista de saída que representam os valores que estão sendo enviados aos endereços. O *Bitcoin* é pseudoanônimo, pois todas as transações são visíveis para o mundo e dificilmente serão encontrados os usuários através de um endereço *Bitcoin*.

O número de usuários da *Blockchain* cresceram tremendamente nos últimos anos.

Devido aos benefícios ao se utilizarem esse ecossistema surgiram alguns problemas como as transações de crimes eletrônicos. Consequentemente, as *Blockchains* públicas tornaram-se um foco de pesquisa para o desenvolvimento de ferramentas para detectar e rastrear usuários e transações relacionadas ao crime eletrônico (Cuneyt Gurcan Akcora and Sudhanva Purusotham and Yulia R. Gel and Mitchell Krawiec-Thayer and Murat Kantarcioglu, 2020).

Apesar do nível de anonimato, dependendo da quantidade de recursos e dos cuidados dos usuários, o rastreio de usuários é facilitada pela *Blockchain*. Sendo assim, é possível utilizar aprendizagem de máquina para detectar a lavagem de dinheiro na *Blockchain* (Joana Lorenz and Maria Inês Silva and David Aparício and João Tiago Ascensão and Pedro Bizarro, 2020).

Além disso, é possível realizar uma análise de dados para conectar os endereços uns aos outros, visando formar uma visão geral dos hábitos de consumo de um indivíduo (Andreas M. Antonopoulos, 2020).

As transações podem ser criadas *online* ou *offline* por qualquer pessoa, mesmo se o indivíduo que estiver criando a transação não seja o que assinará a autorização para gastar os fundos. Uma transação é uma estrutura de dados que codifica uma transferência de valor a partir de uma fonte de fundos, chamada de entrada, para um destino, chamado de saída.

As entradas e as saídas de transações não são relacionados a contas ou identidades. Na verdade, é necessário imaginá-los como quantidades ou pedaços de *Bitcoin*, que são bloqueados com uma senha específica que somente o dono, ou a pessoa que conhece a senha, pode desbloquear.

A matéria-prima principal de uma transação *Bitcoin* é um *output* de transação "não-gasto", ou *UTXO*. Os UTXOs são pedaços indivisíveis da moeda *Bitcoin* vinculados a um dono específico, registrados na *Blockchain*, e reconhecidos como unidades de moeda pela rede.

A rede *Bitcoin* rastreia todos os *UTXOs* disponíveis. Quando um usuário recebe um *Bitcoin* , a quantia é registrada na *Blockchain* como um *UTXO*. Portanto, o *Bitcoin* de um usuário pode estar disperso como UTXOs entre centenas de transações (Andreas M. Antonopoulos, 2020). A Figura 3 representa uma transação. Os campos *inputs* e *out* representam as entradas e as saídas das transações, respectivamente.

2.4 Ferramentas de Análise

As ferramentas de análise da *Blockchain* são essenciais, pois auxiliam as instituições governamentais a identificarem o uso dessas criptomoedas nos comércios ilegais de

```
{
  "hash": "01e7825d78b88abc8a12e0b221d43632a857a09ce4536c21c8fe56c6de5e39ed",
  "size": 296,
  "weight": 1076,
  "tx_index": 8346581138459359,
  "block_height": 712783,
  "inputs": [
    {
      "sequence": 0,
      "witness": "0120000000000000000000000000000000000000000000000000000000000000000",
      "script": "034fe00afabe6d6dbb445a818abbe9fe4aee5d5a747d7b73bc527238c96628ad33e975f194bbc74d0100000000000000008650512e139c1412900000000
      "index": 0,
      "prev_out": {
        "spent": false,
        "tx_index": 0,
        "value": 0,
        "addr": "1CK6KHY6MHgYvmRQ4PAafKYDrglejbHlcE",
        "n": 0,
        "type": 0
      }
    }
  ],
  "out": [
    {
      "type": 0,
      "spent": false,
      "value": 650860285,
      "spending_outpoints": [],
      "n": 0,
      "tx_index": 8346581138459359,
      "script": "76a9147c154ed1dc59609e3d26abb2df2ea3d587cd8c4188ac",
      "addr": "1CK6KHY6MHgYvmRQ4PAafKYDrglejbHlcE"
    }
  ]
}
```


Figura 3 – Um exemplo de transação *Bitcoin*. Nesta transação, o endereço 1CK6KHY6MHgYvmRQ4PAafKYDrglejbHlcE esta enviando 6.50860285 *Bitcoin* para ele mesmo. Provavelmente, o endereço tinha interesse de registrar algum informação na *Blockchain*. Fonte: www.blockchain.com/pt/api/blockchain$_api$

drogas, armas, cartões de crédito, identidades roubadas e material de abuso sexual infantil. Alem disso, identificam crimes financeiros como fraude, lavagem de dinheiro e evasão fiscal.

As criptomoedas apresentam oportunidades substanciais de crescimento. Assim, empresas de análise auxiliam as instituições financeiras a entender como seus clientes estão usando criptomoedas, mitigar riscos potenciais e abraçar uma nova economia baseada em *Blockchain*. Alguns exemplos dessas empresas são a *Chainalysis* e a *Elliptic*.

A *Chainalysis* é a empresa de análise de *Blockchain*. Essas companhias fornecem dados, ferramentas e análises de *Blockchain* para agências governamentais, bolsas e instituições financeiras em 40 países. Alguns exemplos de ferramentas são citados abaixo:

- *Chainalysis* reactor é o software de investigação que conectam transações de criptoativos a entidades do mundo real (Chainalysis Reactor, 2022);
- *Chainalysis Kryptos* fornecem perfis completos de negócios de criptomoedas com base em detalhes KYC e os dados de *Blockchain* mais confiáveis da indústria.

A *Elliptic*, empresa de *compliance* com criptomoedas sediada em Londres, permitem que empresas de criptografia e instituições financeiras automatizem a detecção e

investigação de transações criptográficas arriscadas, economizando tempo e custos. Alguns exemplos de ferramentas são:

Elliptic Discovery fornece visibilidade de centenas de negócios de criptografia. Além disso, fornecem informações detalhadas sobre perfis de risco de conformidade;

Elliptic Lens foi desenvolvido especificamente para detectar riscos e obter conformidade estelar desde a pré até a pós-transação.

Cryptocurrency Forensics cria transparência para uma economia global construída em *Blockchains*, permitindo que bancos, empresas e governos tenham um entendimento comum de como pessoas e empresas usam criptomoedas.

Para mais informações visitem os sites dessas empresas [1,2].

2.4.1 Como são realizadas essas análises

O software *Chainalysis Reactor* disponibilizado pela empresa *Chainalysis* examinam atividades criminosas, como mercados da *darknet*, golpes e *ransomware*, e atividades legítimas, como serviços comerciais. Esse *software* realiza uma investigação que conectam os endereços para entidades do mundo real. Para realizar esse mapeamento esse *software* realiza a análise de várias transações utilizando diversos algoritmos para conseguir um resultado satisfatório.

Além disso, esse *software* rastreia automaticamente qualquer número de transferências para identificar onde os fundos mudaram de usuários. Os resultados da pesquisa incluem informações de usuários da rede *Bitcoin* de código aberto extraídos de milhares de sites, fóruns da *darknet* e bancos de dados de fraudes (Chainalysis Reactor, 2022).

O *Chainalysis Kryptos* disponibilizado pela empresa *Chainalysis* está voltado para a cadeia de negócios de criptomoedas com base nos dados da *Blockchain*. Esse *software* monitora a exposição de um serviço a entidades de risco, como *ransomware*, mercados *darknet* e jurisdições de alto risco (Chainalysis Kryptos , 2022).

Os *softwares* disponibilizados pela empresa *Elliptic* auxiliam as empresas a detectar e prevenir crimes financeiros em criptografia. Além disso, é possível definir regras de risco ao rastrear transações e carteiras (Elliptic, 2022).

2.4.2 Algoritmo H1

O algoritmo H1 é mais usado no contexto do *Bitcoin* e aplicações que imitam essa criptomoeda. O principal objetivo desse algoritmo é o agrupamento por endereços

[1] https://www.chainalysis.com/
[2] https://www.elliptic.co/

idênticos.

O algoritmo H1 é baseado nos endereços que estão nas entradas das transações. A Figura 3 representa uma transação no *Bitcoin* e a lista de entrada da transação é representada pelos *inputs*. Para um usuário do *Bitcoin* utilizar um recurso de um determinado endereço é necessário provar que ele é o proprietário do recurso. Sendo assim, no contexto do *Bitcoin* é necessário assinar a transação utilizando a chave privada. Por isso, quando múltiplos endereços são usados como entrada de uma transação, acredita-se que todos os endereços de entrada pertençam à mesma entidade no mundo real. Desta forma, esses endereços de entrada podem ser agrupados em uma lista de endereços que é representado na Figura 4. Sendo assim, todos os endereços do *cluster* são controlados pela mesma entidade. No artigo (Xi He and Ketai He and Shenwen Lin and Jinglin Yang and Hongliang Mao, 2021), é apresentado que a precisão de acerto do algoritmo pode chegar a 100% sem considerar o fato de que os usuários usam *mixing service* para evitar essa análise.

```
{
    "identificador": "3",
    "cluster": ["addr1", "addr2", "addr3", "addr4", "addr4", "addr5", "addr6"]
}
```

Figura 4 – Objeto que representa um *Cluster*. Fonte: próprio autor

Na Figura 3, pode ser observado um campo chamado *addr* esse é o endereço das entradas das transações que será usado para criar o *cluster*. O *cluster* será criado da junção de todos os endereços que estão nas entradas de cada transação.

Algorithm 1 Algoritmo H1

Entrada: lista de Clusters $L_{cluster}$
Saída: lista de Clusters Resultante $L_{clusterR}$
$L_{clusterR} := []$;
for all *elem* in $L_{cluster}$ **do** ▷ *elem* é representado pelo objeto que está na Figura 4
    tempClusters ← lista de clusters de $L_{cluster}$ com intersecção com *elem*;
    if $|tempClusters| > 0$ **then**
        Cria um cluster C com os elementos de tempCluster e *elem*;
        Insere C em $L_{clusterR}$;
        Remove tempClusters de $L_{cluster}$;
    else
        Insere *elem* em $L_{clusterR}$;
        Remove *elem* de $L_{cluster}$;
    end if
end for
return $L_{clusterR}$;

O algoritmo H1 é definido no Algoritmo 1. O objetivo do algoritmo é unir *clusters* que possuem endereços em comum. Por exemplo, o *cluster* "1" contém os endereços "A" e "B", o *cluster* "2" contém os endereços "B" e "C" e o *cluster* "3" contém os endereços "B" e "W". Para cada endereço do *cluster* "1" será verificado se esse endereço existe nos outros *clusters*. Como esses três *clusters* tem o endereço "B" em comum será criado um *cluster* composto pela união desses três *clusters* que será ["A","B","C","W"].

2.4.3 Atividades Ilegais

Umas das principais características do *Bitcoin* é a segurança nas transações realizadas, pois na *Blockchain* as validações das transações são realizadas por diversos computadores. Sendo assim, essa criptomoeda é usada em várias atividades ilegais, por exemplo, em perda e sequestro de dados de um equipamento, lavagem de dinheiro, pirâmides financeiras e furto de criptomoedas (Arthur Mendes, 2020).

A perda e sequestro de dados é realizada através de um vírus chamado *Ransonware*. O *Ransomware* é um tipo de código malicioso que torna inacessíveis os dados armazenados em um equipamento. O pagamento pedido pelo malfeitor para o usuário ter acesso aos seus dados normalmente é em *Bitcoin*.

O *Bitcoin* garante uma maior praticidade para os indíviduos que utilizam ele para lavagem de dinheiro, pois essa criptomoeda permite transferir muito dinheiro em pouco tempo e é muito difícil rastrear essas transações, pois os *mixing service* dificultam esse rastreamento.

Normalmente, as pirâmides financeiras utilizam o *Bitcoin* da seguinte maneira uma empresa cobra um valor minímo ao novo cliente utilizando esse recurso para comprar *Bitcoin* na promessa de devolver com lucros. No entanto, os investidores mais antigos são financiados pelo dinheiro dos clientes novos e não com o lucro dos investimentos.

Além disso, alguns dos malfeitores utilizam *Darkmarkets* para intermediar essas transações ilícitas. As *Darkmarkets* são sites comerciais que usam *Bitcoin* para vender ou intermediar transações envolvendo drogas, armas e outros bens ilícitos (Abeer ElBahrawy and Laura Alessandretti and Leonid Rusnac and Daniel Goldsmith and Alexander Teytelboym and Andrea Baronchelli, 2019). Alguns exemplos desses *Darkmarkets* são citados abaixo:

- Evolution;
- Flugsvamp;
- Silk road 2.0;
- Welcome to Video (WTV);

- OpenBazaar;
- Particl.io.

Atualmente, alguns mercados de *darkmarket* estão adotando novas infraestruturas para evitar paralisações pela aplicação da lei.

O *OpenBazaar*, por exemplo, tem uma estrutura totalmente descentralizada, semelhante ao próprio *Blockchain* ou o navegador *Tor*, que tornaria impossível removê-lo. Os usuários podem, simplesmente, baixar e executar um programa que os permite conectar diretamente, ao invés de um site.

O *Particl.io* oferece um mercado semelhante com sua própria moeda e infraestrutura de carteira. Nenhum desses mercados alcançaram uma adoção generalizada, mas todos representariam novos desafios para a aplicação da lei se ganhassem popularidade.

2.5 Trabalhos relacionados

No artigo (Nicolas T. Courtois; Kacper T. Gradon; Klaus Schmeh, 2021) foram analisadas várias questões de uso indevido das criptomoedas, especialmente nas áreas de lavagem de dinheiro ou retirada de lucros provenientes de atividades ilícitas. Além disso, foi proposto uma classificação de crimes relacionados à criptomoeda.

O pseudoanonimato é uma das qualidades mais importantes da criptomoeda *Bitcoin*. Normalmente, o proprietário dos endereços da criptomoeda não é conhecida. No entanto, para descobrimos os possíveis donos dos endereços podemos agrupá-los por meio da análise de padrões de comportamento. Esse processo permite que aqueles com atribuição conhecida recebam rótulos que podem ser usados posteriormente para fins legais e de conformidade, auxiliando nas investigações e aplicação da lei (Mengjiao Wang and Hikaru Ichijo and Bob Xiao, 2020).

Apesar de existirem vários algoritmos e ferramentas para ‘quebrarem‘ esse pseudoanonimato das criptomoedas, novos métodos e serviços são desenvolvidos para potencializar esse pseudoanonimato. O *mixing service* é um desses serviços que visam fortalecer o anonimato da rede do *Bitcoin*. Algumas ferramentas permitem fazer o rastreamento das transações e identificar seus autores. Sendo assim, esses serviços dificultam o rastreamento das transações, por meio do embaralhamento do *Bitcoin*. No artigo (L. Wu and Yufeng Hu and Y. Zhou and Haoyu Wang and Xiaopu Luo and Zongguo Wang and Fan Zhang and Kui Ren, 2020), foi proposto um modelo genérico para entender o *mixing service* do *Bitcoin* e um método para identificar transações que alavancam os mecanismos identificados no modelo genérico proposto.

No artigo (Shojaeenasab Ardeshir and Motamed, 2022) foi abordado a importância de rastrear as transações financeiras nas redes de criptomoedas e os problemas em perder essa capacidade causados pelo *mixing service*. Além disso, foi desenvolvido métodos para rastrear as transações e os endereços das atividades lícitas e ilícitas desses serviços.

No ecossistema do *Bitcoin* é necessário desenvolver métodos e ferramentas para analisar as transações dessa criptomoeda, pois ela pode ser utilizada em vários cenários. No artigo (Masarah Paquet-Clouston and Bernhard Haslhofer and Benoit Dupont, 2018), foi apresentado um método baseado em dados para identificar e reunir informações sobre transações de *Bitcoin* relacionadas a atividades ilícitas com base em pegadas deixadas na *Blockchain*. Esse método foi implementado no topo da plataforma de código aberto *GraphSense* e aplicado para analisar empiricamente as transações relacionadas a 35 famílias de *ransomware*.

No artigo (Liao et al., 2016), foi realizado uma análise de medição do *CryptoLocker*. Uma família de *ransomwares* que criptografa os arquivos da vítima até que o resgate seja pago. Essas informações foram coletadas de fóruns online, como *Reddit* e *BitcoinTalk* e foram utilizadas como ponto de partida. Além disso, foi construída uma topologia de rede para detalhar a infraestrutura financeira do *CryptoLocker* e obter informações auxiliares sobre a operação do mesmo. Mais notavelmente, foram encontradas evidências que sugerem conexões com serviços populares de *Bitcoin*, como *Bitcoin Fog* e *BTC-e* e *links* sutis para outros crimes cibernéticos em torno de *Bitcoin*, como o golpe do *Sheep Marketplace* de 2013 . O *Sheep Marketplace*, um mercado online de drogas que se tornou proeminente depois que o *Silk Road* foi fechado pelos promotores dos EUA, também ficou offline e até US$ 44 milhões em Bitcoins podem ter desaparecido com ele (Jim Edwards, 2013).

Embora as tecnologias apresentadas possam desencadear novas mudanças tecnológicas. Elas podem ser utilizadas em outros golpes que ameaçam a segurança financeira, como os esquemas *Ponzi*. O esquema *Ponzi* promete retornos altos e de forma rápida aos seus investidores. Esse retorno alto e de forma rápida depende da entrada de novos investidores de forma recorrente, porque os golpistas pegam o dinheiro das novas aplicações e repassam para quem está há mais tempo no ‘negócio‘ (REDAÇÃO XPEED, 2022). O artigo (JIN et al., 2022), foca-se na detecção do esquema *Pozi* e foi proposto um modelo genérico que pode capturar as informações associadas aos padrões de comportamento da conta e pode ser combinado com os métodos de detecção de esquemas *Ponzi* existentes.

Nesse trabalho foi desenvolvido uma implementação do algoritmo H1 utilizando ferramentas recentes como o *Go* e *MongoDB*. O algoritmo H1 foi aplicado nos endereços disponibilizados pela investigação promovida pelo Ministério Público e Polícia Federal conhecida como “Faraó dos Bitcoins” que foi realizada recentemente. Além disso, os resultados obtidos nessa análise acrescentaram 3 novos endereços que podem pertencer a

entidade que está sendo investigada.

3 Desenvolvimento

Neste Capítulo, são apresentados os passos para a implementação do algoritmo H1 descrito em 2.4.2.

Primeiramente, foram definidas as ferramentas que seriam utilizadas para implementação do objetivo principal desse trabalho. Para serem definidos os métodos e as ferramentas que seriam usadas nesse projeto a princípio definimos os requisitos que essa implementação deveria atender.

Por exemplo, a busca pelas transações serão manuais ou é necessário criar um implementação que faz requisição em uma *API* em busca das transações, como deverá ser construído a base de dados dessas transações, qual será o banco de dados escolhido para armazenar os dados, a linguagem de programação e os possíveis problemas e limitações que poderiam surgir.

Para consultarmos os dados necessários utilizamos uma *API* que é um serviço fornecido pela empresa Blockchain.com[1]. Essa *API* disponibiliza diversos *end-points* que podemos utilizar para consultar blocos, transações, altura de um bloco, endereços e suas transações, *OutPut Unspent*, *Unconfirmed Transactions*, entre outros.

O banco de dados escolhido para armazenar os dados retornados dessa *API* foi o *MongoDB*. O *MongoDB* proporciona várias vantagens ao nosso projeto, pois possibilita uma alta escalabilidade, flexibilidade, manipulação de uma quantidade massiva de dados, bom desempenho e facilidade para consultas. Para aplicarmos o algoritmo H1 e obtermos resultados satisfatórios é necessário armazenar muitas transações e processar vários dados em pouco tempo para que o algoritmo seja o mais eficiente possível, por isso o *MongoDB* é uma excelente escolha. Além disso, os dados obtidos pelo uso da *API* de acesso à *Blockchain* estão no formato *JSON*, e o banco de dados escolhido suporta esse formato de dados.

A linguagem escolhida para implementação dos algoritmos foi o *Go* (Google, 2020), porque o desempenho das aplicações criadas por essa linguagem é bem alto e a simplicidade no desenvolvimento é um ponto forte. Além disso, a eficiência dessa linguagem de programação permite que os serviços desenvolvidos processem volumes maiores de informação. Sendo assim, nesse projeto será processado muitas transações, por isso é necessário uma linguagem com alto desempenho para consultar os dados adquiridos da *API*, processar e armazenar esses valores no *MongoDB*.

Apesar das vantagens do banco de dados e da linguagem de programação é neces-

[1] <https://www.blockchain.com/pt/api/blockchain_api>

sário um alto poder computacional para conseguir processar vários dados e realizar uma análise mais precisa e satisfatória. No entanto, esse projeto pode ser usado como referência para projetos futuros que envolvam a utilização de algoritimos de clusterização.

Antes de explicar como foi implementado o algoritmo H1 e como está organizado os *Database* e as *Collection* no *MongoDB* será apresentado o cenário que foi aplicado o algoritmo H1.

O cenário em que foi aplicado o algoritmo H1 está relacionado aos endereços disponibilizados pela investigação promovida pelo Ministério Público e Polícia Federal conhecida como 'Faraó dos Bitcoins' (Vinícius Golveia, 2021). Os endereços disponibilizados por essa investigação serão nomeados nesse trabalho como endereço em análise. São disponibilizados 9 endereços em análise que são o ponto de partida desse trabalho. Na Tabela 3, são apresentados os endereços em análise.

1JawWE56G5NmnB5iuYbFikbdETs88Fxkwo
bc1qluuy04mjxqj8yc44lgnez8eml4pwulvukfatak
bc1qt7jjpqdfvqhtqkadlnuhzzem2tateg7mm0y95w
bc1q8kmtzc0a43w0cjrzwzwqsa9frxaseyzcg6mq3d
bc1qqgjmxevtn3cyg8cvxfg7yyk6a7n3zudt4hw85t
bc1qn9k6s0lyxgw5mdndta3780md23z9kmu4980clv
bc1qu9tj6kcusrncvm7wm06n2mqq0jtfg26vk9mynm
bc1qnanyweuswqm9sz3d93ag0vrc69mpq4v40g9acq
bc1qehzj8sulj3plzuarzzmdm77d6rd8chvc5hzull

Tabela 3 – Endereços citados na notícia (Vinícius Golveia, 2021). Fonte: próprio autor

A princípio, buscaremos as transações dos 9 endereços.[2]. Em seguida, utilizaremos os endereços que estão nas entradas dessas transações dos 9 endereços para buscarmos mais transações. Sendo assim, estamos buscando transações de endereços que estão em outras transações e essse processo tornará muito longo, pois sempre haverá novos endereços para buscarmos suas transações, por isso, é necessário estipular a distância dos endereços em análise que serão realizados as buscas pelas transações[3].

Definimos que buscaremos as transações até duas distâncias dos endereços em análise. Por exemplo, buscamos as transações de um endereço em análise, os endereços que estão nas transações dos endereços em análise estão na distância 1. Em seguida, buscamos as transações dos endereços que estão na distância 1. Os endereços que estão na distância 2 são os endereços que estão nas transações dos endereços que estão na distância 1. Na Figura 5 é apresentado esse cenário, o endereço analisado é o 'Endereço A' e os endereços Endereço1 e Endereço2 estão na distância 1 em comparação ao endereço

[2] Nessa seção, por simplicidade, o termo transação refere somente às entradas

[3] O termo distância está sendo utilizado para representar a distância que um determinado endereço está do endereço em análise

analisado e o Endereço3, Endereço4 e o Endereço5 estão na distância 2 em comparação ao endereço analisado.

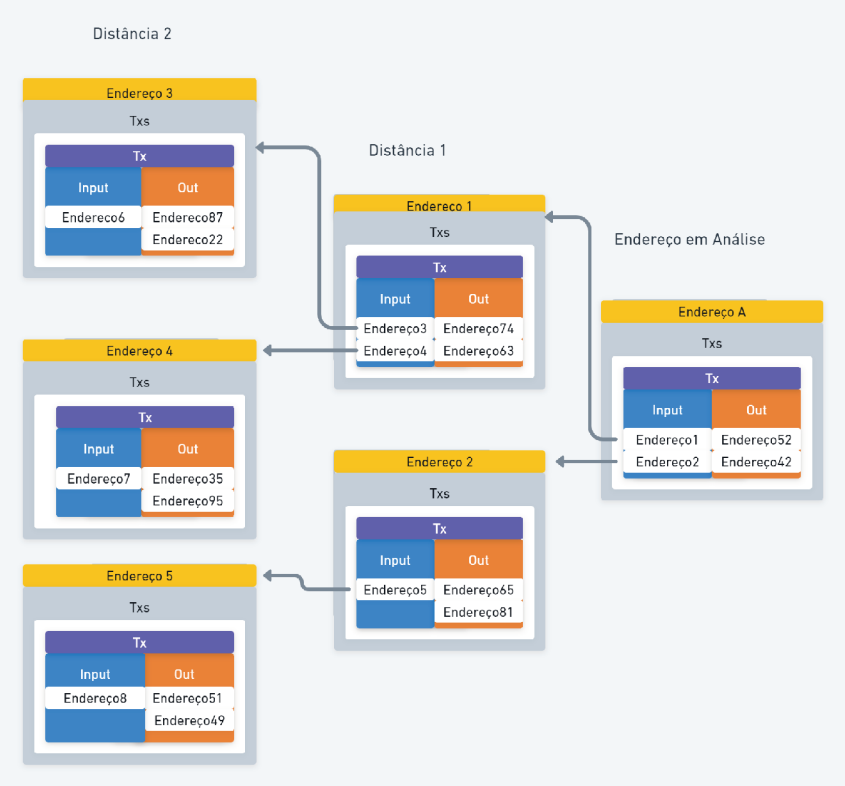

Figura 5 – Ilustração do Cenário. Fonte: próprio autor

Os *Databases* no *MongoDB* contêm uma ou mais *Collections* de documentos. As *Collections* são análogas às Tabelas em bancos de dados relacionais (MongoDB, Inc., 2021). Na Tabela 4 está sendo apresentado os *Databases* e as *Collections* criadas para armazenar as transações, os endereços em análise e os *clusters*.

DataBase	Collection
Endereco	FaraoEnderecos, Farao, Distancia1, Distancia1Informacoes; Distancia2Informacoes,
Cluster	Clusters Identificadores

Tabela 4 – *Database* e *Collection* criadas no *MongoDB*. Fonte: próprio autor

A *Collection* FaraoEnderecos do *Database* Endereco contém os 9 endereços disponibilizados pela investigação citada anteriormente. As *Collections* do *Database Endereco* contém as transações de todos os endereços. As *Collections* do *Database Cluster* contém os *clusters* formados pelas transações de cada endereço.

Em seguida, serão apresentadas as funções principais do sistema. As funções principais da aplicação são *saveTxs*, *saveClusters* e *h1*. Na Figura 6 é ilustrada a ordem que as funções principais devem ser executadas.

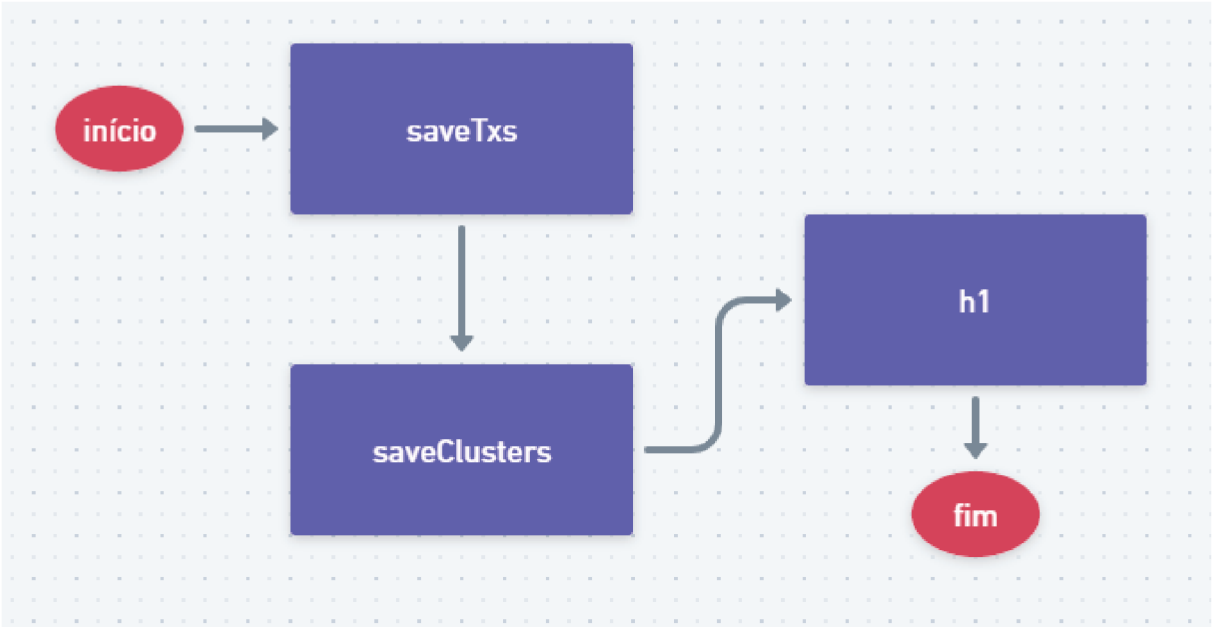

Figura 6 – Ordem de execução das funções principais. Fonte: próprio autor


```
{
  "hash160": "660d4ef3a743e3e696ad990364e555c271ad504b",
  "address": "1AJbsFZ64EpEfS5UAjAfcUG8pH8Jn3rn1F",
  "n_tx": 17,
  "n_unredeemed": 2,
  "total_received": 1031350000,
  "total_sent": 931250000,
  "final_balance": 100100000,
  "txs": [
    "--Array of Transactions--"
  ]
}
```


Figura 7 – Endereço retornado pelo *end-point rawaddr*. Fonte:blockchain.com

A primeira função a ser executada é a *saveTxs*. Essa função realiza uma requisição na *API* citada anteriormente no *end-point rawaddr* que retorna as transações de um endereço. Esse *end-point* possui uma limitação de transações que podemos obter por endereço que é aproximadamente 50 mil. A Figura 7 representa o dado que essa *API* retorna. No algoritmo abaixo, é apresentado a lógica da implementação da função *saveTxs*.

A segunda função a ser executada é a *saveClusters*, pois ela busca as transações que estão salvas na *Collection Farao* e na *Collection Distancia1* do *Database Endereco*

Algorithm 2 Algoritmo SaveTxs

1: **Entrada:** lista de enderecos $L_{enderecos}$
2: **Saída:** lista de Transações Resultante L_{Txs}
3: $L_{Txs} := []$;
4: $offset := 0$;
5: **for all** *endereco* in $L_{enderecos}$ **do**
6: transações ← requisita as transações de *endereco* no *endpoint rawaddr*; ▷ transação é representado pelo objeto que está na Figura 7
7: numeroTransacoes ← recebe $|transações.n\_tx|$;
8: tamanhoTxs ← recebe $|transações.txs|$;
9: **if** $|n\_tx|$ ==|tamanhoTxs| **then**
10: Insere $transações$ em L_{Txs};
11: **else**
12: $offset$ ← tamanhoTxs;
13: Insere $transações$ em L_{Txs};
14: **while** $offset <$ numeroTransacoes **do**
15: transações ← requisita as transações de *endereco* no *endpoint rawaddr* passando o parametro offset = $offset$;
16: Insere $transações$ em L_{Txs};
17: $offset := offset +$ tamanho da lista txs em $transações$;
18: **end while**
19: **end if**
20: **end for**
21: **return** L_{Txs};

e cria *clusters* com essas transações. Esses *clusters* são salvos na *Collection Clusters* do *Database Cluster*.

Por fim, a função *h1* é a última a ser executada, essa função aplica o algoritmo H1 nos dados salvos na *Collection Clusters* do *Database Cluster*. Inicialmente, esse método busca os *clusters* que foram salvos na *Collection Clusters* do *Database Cluster*. Posteriormente, essa função tenta encontrar *clusters* que tenham endereços em comum. Se forem encontrados *clusters* que tenham endereços em comum esses *clusters* são unidos e os dados da *Collection Clusters* do *Database Cluster* são atualizados. Além disso, a próxima execução desse algoritmo é sempre realizada com os valores atualizados da clusterização anterior.

4 Resultados

Neste capítulo, apresentamos os resultados da aplicação do algoritmo H1 em um caso real que ocorreu recentemente. Os dados foram coletados a partir de 28 de março de 2022 até 17 de abril de 2022. A seguir, serão apresentados os resultados descritos nesse trabalho.

4.1 Endereços coletados do 'Faraó dos Bitcoins'

Nas distâncias 1 e 2, foram salvas, respectivamente, 278 e 633314 transações. O total de transações salvas dessas duas distâncias resultam em 633592 transações. Para cada transação será criado um *cluster*. Esses *clusters* serão criados pelo agrupamento de endereços que estão nessas transações. Sendo assim, foram criados 633592 *clusters*. A partir do momento que foi utilizado o algoritmo H1 nesses 633592 *clusters* reduziu a quantidade de *clusters* para 316137 *clusters*, pois foram agrupados por esse algoritmo.

4.2 Resumo

As Tabelas abaixo apresentam um resumo dos resultados obtidos na análise:

Clusters	Endereços	*Clusters* Resultantes
633592	1530531	316137

Tabela 5 – Resumo das informações do processo de clusterização pelo método H1. Fonte: próprio autor

Nº	Endereços Iniciais	Quantidade de endereços do *Cluster*
1	1JawWE56G5NmnB5iuYbFikbdETs88Fxkwo	88592
2	bc1qluuy04mjxqj8yc44lgnez8eml4pwulvukfatak	1
3	bc1qt7jjpqdfvqhtqkadlnuhzzem2tateg7mm0y95w	-
4	bc1q8kmtzc0a43w0cjrzwzwqsa9frxaseyzcg6mq3d	2
5	bc1qqgjmxevtn3cyg8cvxfg7yyk6a7n3zudt4hw85t	2
6	bc1qn9k6s0lyxgw5mdndta3780md23z9kmu4980clv	2
7	bc1qu9tj6kcusrncvm7wm06n2mqq0jtfg26vk9mynm	3
8	bc1qnanyweuswqm9sz3d93ag0vrc69mpq4v40g9acq	2
9	bc1qehzj8sulj3plzuarzzmdm77d6rd8chvc5hzull	3

Tabela 6 – Tamanho dos *Clusters* que os endereços iniciais aparecem Fonte: próprio autor.

Na Tabela 6 são mostrados a quantidade de endereços dos *clusters* que os endereços iniciais estão contidos. Uma análise criteriosa permite chegar às seguintes observações:

- O endereço '1JawWE56G5NmnB5iuYbFikbdETs88Fxkwo' está contido em um *cluster* muito grande. Esse *cluster* é um indício que essa carteira está em uma *exchange* ou passou por um processo de *mixing* (HE et al., 2022). Desta forma, as demais carteiras que estão neste *cluster* não tem relevância imediata no processo de análise forense;
- O endereço 'bc1qt7jjpqdfvqhtqkadlnuhzzem2tateg7mm0y95w' não está contido em nenhum *cluster* que foi formado com as transações que foram obtidas. Como a *API* consultada limita a quantidade de transações que podemos obter de cada endereço não foi possivel consultar as transações de alguns endereços em sua totalidade, sendo assim nas transações obtidas, o endereço representado pelo número 3 da Tabela 6 não foi encontrado nos endereços das transações de entrada;
- Os *clusters* 4 e 6 colapsam e formam um único *cluster*, sem impacto na análise forense.
- Já os *clusters* números 7 e 9 também colapsam, mas agregam um novo endereço;
- Finalmente, os *clusters* 5 e 8, trazem, cada um deles, um novo endereço ao grupo.

Nº	Endereços dos *Cluster*
1	88592
2	bc1qluuy04mjxqj8yc44lgnez8eml4pwulvukfatak
3	-
4	bc1q8kmtzc0a43w0cjrzwzwqsa9frxaseyzcg6mq3d bc1qn9k6s0lyxgw5mdndta3780md23z9kmu4980clv
5	bc1qqgjmxevtn3cyg8cvxfg7yyk6a7n3zudt4hw85t bc1q6p5pn7l9n0vs4v4rc5vet3zs9hfhylwh0fqm35
6	bc1qn9k6s0lyxgw5mdndta3780md23z9kmu4980clv bc1q8kmtzc0a43w0cjrzwzwqsa9frxaseyzcg6mq3d
7	bc1qu9tj6kcusrncvm7wm06n2mqq0jtfg26vk9mynm bc1qehzj8sulj3plzuarzzmdm77d6rd8chvc5hzull bc1qz30fyctnylenx584w483ekxd9tyds07h7pyexq
8	bc1qnanyweuswqm9sz3d93ag0vrc69mpq4v40g9acq bc1qp2nx27jln57va4cw55kmpu34h30s6plzylkmtx
9	bc1qehzj8sulj3plzuarzzmdm77d6rd8chvc5hzull bc1qu9tj6kcusrncvm7wm06n2mqq0jtfg26vk9mynm bc1qz30fyctnylenx584w483ekxd9tyds07h7pyexq

Tabela 7 – Endereços que contém os *Clusters* que os endereços iniciais aparecem Fonte: próprio autor.

A Tabela 7 apresenta os endereços que estão contidos nos *clusters* dos endereços iniciais. Esses endereços que estão no mesmo *cluster* são um forte indício que essas carteiras são controladas pelo Faraó dos Bitcoins. O capital do endereço que tem o *cluster*

maior se diluiu e os endereços dos outros *clusters* também devem ser objeto de estudo pela perícia.

4.3 Resultado Final

Esse trabalho foi iniciado com 9 endereços que foram disponibilizados pela investigação feita pelo Ministério Público. Foi realizado o processo de busca dos endereços que estão nas transações dos endereços em análise, ou seja, obtém o vizinho 1 desses endereços em análise e depois o vizinho do vizinho 1. Posteriormente, foi realizado a clusterização dessas transações. Por fim, utilizando o algoritmo H1 nós temos a inferência de que existem mais endereços que são controlados pelo 'Faraó dos Bitcoins'. A relação dos novos endereços é apresentada na Tabela 8.

Endereços Novos
bc1qz30fyctnylenx584w483ekxd9tyds07h7pyexq
bc1q6p5pn7l9n0vs4v4rc5vet3zs9hfhylwh0fqm35
bc1qp2nx27jln57va4cw55kmpu34h30s6plzylkmtx

Tabela 8 – Endereços suspeitos de pertencer ao 'Faraó dos Bitcoins' Fonte: próprio autor.

5 Conclusão

Nesse trabalho, foram apresentados o ecossistema da *Blockchain* e a criptomoeda *Bitcoin*. Foram esclarecidas algumas características da *Blockchain* e do *Bitcoin*: a *Blockchain* é uma estrutura de dados que registra dados imutáveis, aplicada ao *Bitcoin* tem finalidade de coletar dados cronológicos de todas as transações que são realizadas. Um mecanismo de consenso baseado em Prova de Trabalho é responsável por proporcionar confiabilidade e transparência ao sistema, tornando desnecessária a existência de uma instituição financeira centralizada para controlar as transações. O *Bitcoin* fornece um pseudoanonimato aos seus usuários estabelecendo uma identidade que usam chaves públicas.

Em consequência de algumas características do *Bitcoin*, como a manutenção do anonimato torna-se necessário o rastreamento dessas transações. Sendo assim, foi apresentado uma implementação do algoritmo de clusterização H1 utilizando a linguagem de programação *Go* juntamente com o banco de dados não relacional *MongoDB* que visa identificar os possíveis endereços que possam pertencer a mesma entidade. Além disso, também foram apresentados algumas ferramentas de análise que auxiliam as instituições governamentais a identificarem o uso dessas criptomoedas nos comércios ilegais. A seguir, foi apresentado e detalhado o algoritmo H1 desvendando como é funcionamento do mesmo.

Por meio dessa solução conseguimos demonstrar uma análise forense no âmbito do *Bitcoin*. Usamos a clusterização H1 em um conjunto de endereços do 'Faraó dos Bitcoins', atualmente em investigação pelo Ministério Público e Polícia Federal. Com base nessas análises conseguimos identificar endereços novos que podem estar relacionados a mesma entidade.

Para trabalhos futuros sugerimos uma análise mais complexa dos *clusters*, por exemplo, aplicar outros algoritmos de clusterização em conjunto com o H1 para aumentar a acurácia da operação. Além disso, é necessário aumentar a eficiência do programa desenvolvido na linguagem *Go*. Para aumentar o desempenho do programa pode ser utilizado *Goroutines* que é uma função ou método que executa independentemente e simultaneamente em conexão com qualquer outra *Goroutine* presente no programa. Para ter um maior poder computacional e ser criada uma base de dados com distâncias maiores dos endereços iniciais é necessário executar essa aplicação em nuvem com as melhorias citadas acima.

Uma análise que poderia ser feita em trabalhos futuros seria aplicar os mesmos

passos descritos nesse trabalho usando como base os novos endereços encontrados nesse trabalho.

Referências

Abeer ElBahrawy and Laura Alessandretti and Leonid Rusnac and Daniel Goldsmith and Alexander Teytelboym and Andrea Baronchelli. Collective dynamics of dark web marketplaces. 2019. Citado na página 20.

Andreas M. Antonopoulos. Mastering bitcoin: Programming the open blockchain. 2020. Citado 4 vezes nas páginas 12, 13, 15 e 16.

Arlindo F. da Conceição Vladimir Moreira Rocha, R. Blockchain e aplicações em saúde. *SBCOPENLIB*, 2019. Disponível em: <https://sol.sbc.org.br/livros/index.php/sbc/catalog/view/29/96/246-1>. Citado na página 11.

Arthur Mendes. O uso das criptomoedas na criminalidade cibernética. *jusbrasil*, 2020. Disponível em: <https://atrmendes.jusbrasil.com.br/artigos/838270331/o-uso-das-criptomoedas-na-criminalidade-cibernetica>. Citado na página 20.

Bitcoin Project. Bitcoin é uma rede de pagamento inovadora e um novo tipo de dinheiro. 2009. Disponível em: <https://bitcoin.org/en/>. Citado 2 vezes nas páginas 12 e 15.

Carin Tom. *O que é o Livro Razão na contabilidade?* 2020. <https://contadores.contaazul.com/blog/livro-razao-contabilidade>. Acesso em: 04 November 2020. Citado na página 11.

Chainalysis Kryptos . Chainalysis inc. 2022. Disponível em: <https://www.chainalysis.com/chainalysis-kryptos/>. Citado na página 18.

Chainalysis Reactor. Chainalysis inc. 2022. Disponível em: <https://www.chainalysis.com/chainalysis-reactor/>. Citado 2 vezes nas páginas 17 e 18.

Cuneyt Gurcan Akcora and Sudhanva Purusotham and Yulia R. Gel and Mitchell Krawiec-Thayer and Murat Kantarcioglu. How to not get caught when you launder money on blockchain? *CoRR*, abs/2010.15082, 2020. Disponível em: <https://arxiv.org/abs/2010.15082>. Citado na página 16.

Elliptic. Crypto transaction monitoring. 2022. Disponível em: <https://https://www.elliptic.co/solutions/crypto-transaction-monitoring>. Citado na página 18.

Google. *Site oficial da Linguagem Go*. 2020. <https://go.dev/>. Acesso em: 21 February 2022. Citado na página 24.

HE, X. et al. Bitcoin address clustering method based on multiple heuristic conditions. *IET Blockchain*, v. 2, n. 2, p. 44–56, 2022. Disponível em: <https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/blc2.12014>. Citado na página 30.

Jim Edwards. *'Sheep Marketplace' Goes Offline And Up To $44 Million In Bitcoins Disappears*. 2013. <https://www.businessinsider.com/sheep-marketplace-goes-offline-and-up-to-44-million-in-bitcoins-disappears-2013-12>. Acesso em: 15 February 2022. Citado na página 22.

JIN, C. et al. *Heterogeneous Feature Augmentation for Ponzi Detection in Ethereum*. arXiv, 2022. Disponível em: <https://arxiv.org/abs/2204.08916>. Citado na página 22.

Joana Lorenz and Maria Inês Silva and David Aparício and João Tiago Ascensão and Pedro Bizarro. Machine learning methods to detect money laundering in the bitcoin blockchain in the presence of label scarcity. *CoRR*, abs/2005.14635, 2020. Disponível em: <https://arxiv.org/abs/2005.14635>. Citado na página 16.

Júlia Lewgoy. Tarifas dos grandes bancos saltam acima da inflação durante a pandemia. *https://valorinveste.globo.com/*, 2021. Disponível em: <https://valorinveste.globo.com/produtos/servicos-financeiros/noticia/2021/09/16/tarifas-dos-grandes-bancos-saltam-acima-da-inflacao-durante-a-pandemia.ghtml>. Citado na página 10.

L. Wu and Yufeng Hu and Y. Zhou and Haoyu Wang and Xiaopu Luo and Zongguo Wang and Fan Zhang and Kui Ren. Towards understanding and demystifying bitcoin mixing services. *ArXiv*, abs/2010.16274, 2020. Disponível em: <https://arxiv.org/abs/2010.16274>. Citado na página 21.

Liao, K. et al. Behind closed doors: measurement and analysis of cryptolocker ransoms in bitcoin. p. 1–13, 2016. Citado na página 22.

Masarah Paquet-Clouston and Bernhard Haslhofer and Benoit Dupont. Ransomware payments in the bitcoin ecosystem. *CoRR*, abs/1804.04080, 2018. Disponível em: <http://arxiv.org/abs/1804.04080>. Citado na página 22.

Mengjiao Wang and Hikaru Ichijo and Bob Xiao. Cryptocurrency address clustering and labeling. 2020. Disponível em: <https://arxiv.org/abs/2003.13399v1>. Citado na página 21.

MongoDB, Inc. *Site oficial do MongoDB*. 2021. <https://docs.mongodb.com/manual/core/databases-and-collections/>. Acesso em: 20 March 2022. Citado na página 26.

Nicolas T. Courtois; Kacper T. Gradon; Klaus Schmeh. Crypto currency regulation and law enforcement perspectives. *CoRR*, abs/2109.01047, 2021. Disponível em: <https://arxiv.org/abs/2109.01047>. Citado na página 21.

REDAÇÃO XPEED. O que é esquema ponzi? conheça as características e casos famosos. 2022. Disponível em: <https://xpeedschool.com.br/blog/o-que-e-esquema-ponzi/>. Citado na página 22.

Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. *Www.Bitcoin.Org*, p. 9, 2008. ISSN 09254560. Disponível em: <https://bitcoin.org/bitcoin.pdf>. Citado 2 vezes nas páginas 11 e 12.

Shojaeenasab Ardeshir and Motamed, A. Mixing detection on bitcoin transactions using statistical patterns. arXiv, 2022. Disponível em: <https://arxiv.org/abs/2204.02019>. Citado na página 22.

Sloane Brakeville with Bhargav Perepa. *Blockchain basics: Introduction to business ledgers*. 2020. <https://www.finyear.com/Blockchain-basics-Introduction-to-business-ledgers_a36159.html>. Acesso em: 03 November 2020. Citado na página 11.

UFRJ. Blockchain e aplicações. 2018. Disponível em: <https://www.gta.ufrj.br/ensino/eel878/redes1-2018-1/trabalhos-vf/blockchain/aplications.html>. Citado na página 11.

UFRJ. Segurança e criptografia. 2018. Disponível em: <https://www.gta.ufrj.br/ensino/eel878/redes1-2018-1/trabalhos-vf/blockchain/security.html>. Citado na página 13.

Vinícius Golveia. Carteira utilizada pelo "faraó dos bitcoins" é revelada. *live coins*, 2021. Disponível em: <https://livecoins.com.br/carteira-utilizada-pelo-farao-dos-bitcoins-revelada/>. Citado 2 vezes nas páginas 7 e 25.

Xi He and Ketai He and Shenwen Lin and Jinglin Yang and Hongliang Mao. Bitcoin address clustering method based on multiple heuristic conditions. *CoRR*, abs/2104.09979, 2021. Disponível em: <https://arxiv.org/abs/2104.09979>. Citado na página 19.

APÊNDICE A – Código Fonte

Para mais informações acesse o link https://github.com/ericpatricksantos/Tcc.