Transações OP_RETURN: uma oportunidade de análise forense

Luis Benito Sanches Bulhões[1], **Ivan Sendin**[1], **Rodrigo Sanches Miani**[1]

[1]Faculdade de Computação – Universidade Federal de Uberlândia (UFU)
Av. João Naves de Ávila, 2121 – Santa Mônica, 38400-902 – Uberlândia – MG – Brazil

{luis.bulhoes, sendin, miani}@ufu.br

Abstract. *The introduction of the instruction OP_RETURN in Bitcoin made it possible to insert data of any format into transactions. This paper analyzes the instruction content in transactions carried out between 2014 and 2019 and seeks to compare such content with the profile of users responsible for the transactions. The results show that the majority of transactions using the instruction OP_RETURN were mapped with unknown protocols. Regarding the readable content present in such field, we found potential new protocols, URLs related to torrents, memes, and markings regarding events about Bitcoins.*

Resumo. *A introdução da instrução OP_RETURN na Bitcoin possibilitou a inserção de dados de qualquer formato em transações. Este trabalho analisa o conteúdo da instrução em transações realizadas entre 2014 e 2019 e busca traçar um comparativo entre tais conteúdos com o perfil dos usuários responsáveis pelas transações. Os resultados mostram que a maioria das transações que utilizam a instrução OP_RETURN foram mapeadas em protocolos desconhecidos. Acerca do conteúdo legível presente em tal campo, foram encontrados potenciais novos protocolos, URLs relacionadas a torrents, memes e marcações de eventos relacionados à Bitcoins.*

1. Introdução

Bitcoin é uma moeda digital que permite o pagamento instantâneo entre os usuários da sua rede [Nakamoto 2008]. Ela usa uma infraestrutura de rede *peer-to-peer* onde mineradores, de forma coletiva, trabalham para autenticar as transações. Com a autenticação da transação, as informações serão armazenadas na *blockchain*, que é o livro razão da *Bitcoin*. A *blockchain*, por sua vez, é uma estrutura constituída por blocos, sendo que cada bloco será composto por transações autenticadas.

Na atualização do cliente *Bitcoin Core* no ano de 2014, os desenvolvedores tornaram padrão a instrução *OP_RETURN* [Antonopoulos 2014]. Como o próprio nome sugere, esta instrução gera um retorno, encerrando a execução do *script*. Assim, todo o código do *script* que vêm após uma instrução OP_RETURN nunca será executado e pode ser tratado como um campo de dados. O objetivo de tal instrução é permitir a inserção e persistência de dados arbitrários em uma transação. Anteriormente, um método muito utilizado para se inserir dados em uma transação *Bitcoin* era conhecido como *unspendable outputs*, que utiliza o campo de saída de uma transação para a inserção dos dados. Entretanto, tal método necessita da "queima" de *Bitcoins*, ou seja, *Bitcoins* teriam de ser travadas, sem a possibilidade de serem recuperadas. Com o *OP_RETURN*, a inserção

de dados se torna mais simples. A Figura 1 apresenta um exemplo de transação com o *OP_RETURN*.

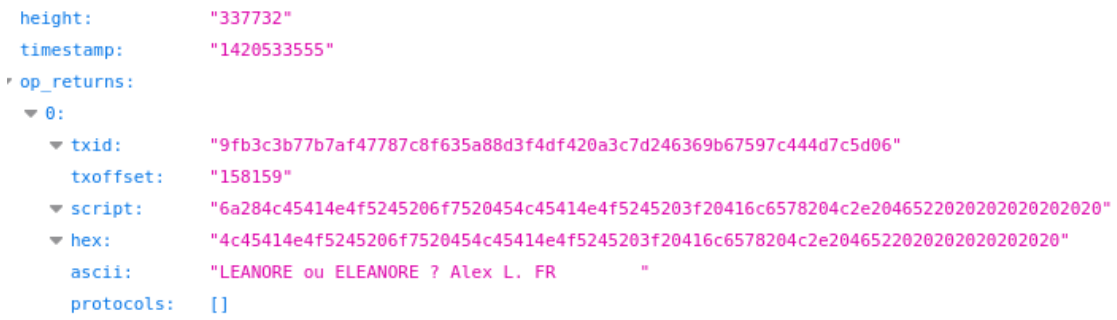

Figura 1. Um exemplo de transação com uso de *OP_RETURN* obtido do site coinsecrets.org. No campo *script* os dois caracteres iniciais, 6a, são para identificar o *OP_RETURN*.

Na Figura 1, o campo *script* possui no início o *opcode 6a* no qual se refere ao *OP_RETURN*. O resto do *script* será formado pelos dados que o usuário deseja inserir. O campo *hex* e *ascii* são apenas para facilitar a leitura do *script* e são disponibilizados pela própria *API* do *Bitcoin*.

Considerando o cenário descrito anteriormente, o objetivo deste artigo é analisar o perfil do uso do *OP_RETURN*, como por exemplo, investigar os valores gastos nas transações e os tipos de conteúdos inseridos nas mesmas. A análise foi feita levando em consideração transações ocorridas entre 2014 e 2019, pois 2014 foi quando tornou-se padrão o uso do *OP_RETURN* [Antonopoulos 2014] e 2019 por conta de ter sido o ano anterior ao inicio da coleta dos dados. Espera-se que com tal análise, seja possível revelar novos comportamentos dos usuários da rede *Bitcoin*.

2. Fundamentação Teórica

O uso do *OP_RETURN* foi discutido em [Strehle and Steinmetz 2020]. Os autores mostraram que diversos tipos de serviços financeiros passaram a adotar tal funcionalidade para o armazenamento de dados. Alguns exemplos de serviços incluem a *Veriblock*, que utiliza o conceito de *Proof of Proof* [Sanchez and Fisher 2019], e o *Omni*, cujo serviço promove uma das *stablecoins*, a *Tether*[Tether Operations 2021].

Outros tipos de investigações foram conduzidas com o intuito de compreender o comportamento das transações que utilizam *OP_RETURN*. [Bartoletti and Pompianu 2017] e [Bartoletti et al. 2019] identificaram e classificaram uma série de protocolos associados a transações usando *OP_RETURN*. [Ali et al. 2018] mostraram como a comunicação entre bots e servidores de comando e controle (C&C) poderiam ser realizadas usando o *OP_RETURN*. Além disso, [Matzutt et al. 2018] identificaram conteúdos sensíveis armazenados no *OP_RETURN* de transações como dados de denunciantes (*whistleblowers*), com informações sigilosas de governos e de entidades privadas comprovado pelos *links* de *backups* dos dados do *WikiLeaks*. Os autores também encontraram violação de direitos autorais com links de *download* de materiais pirateados e um *hidden file* que era um *backup* de listas de *links* com conteúdos de pornografia infantil.

3. Metodologia

As principais etapas seguidas para a realização desta pesquisa são:

1. Extração de dados das transações que usaram *OP_RETURN* usando o serviço *Coinsecrets*[coinsecrets.org] (atualmente indisponível) e no site *Blockchain*[blockchain.com] através de suas *APIs* que enviaram os dados no formato *JSON*;
2. Armazenamento dos dados extraídos em um banco de dados não relacional, o *MongoDB*;
3. Análise dos dados extraídos, como valores, interação entre as transações através de grafos e relacionamento entre serviços e outros tipos de conteúdos inseridos. Os seguintes itens foram desenvolvidos:
    - análise de valores das transações à partir do cálculo de médias e da quantidade de *Bitcoins* usadas nas transações;
    - análise de agrupamento das transações usando grafos;
    - análise de conteúdo usando expressões regulares.

4. Resultados

As transações coletadas na etapa de extração foram divididas em dois tipos: conhecidas e desconhecidas. As conhecidas são transações que usam protocolos que o serviço *Coinsecrets* consegue identificar. A Tabela 2 mostra a quantidade de transações usando protocolos conhecidos e desconhecidos entre 2014 e 2019. Um protocolo (ou serviço), nesse caso, seria uma string legível gravada no *OP_RETURN* e utilizada em outras transações. Portanto, é possível notar a diversidade de conteúdos inseridos usando o *OP_RETURN*. A Tabela 1 ilustra a evolução do uso das transações que utilizaram o *OP_RETURN* desde sua padronização, em 2014, até 2019 [opreturn.org]. É possível notar um aumento expressivo no número de transações OP_RETURN.

Ano	Quantidade
2014	4.005
2015	125.916
2016	443.876
2017	1.285.805
2018	3.115.893
2019	16.716.374

Tabela 1. Tabela com o total de transações realizadas disponíveis no *site coinsecrets.org*

A Figura 2 retrata a porcentagem de transações que usaram o *OP_RETURN* na rede *Bitcoin* entre 2014 e 2019. Observa-se um significativo aumento no número de transações *OP_RETURN* após a sua padronização em 2014.

Com relação aos valores, grande parte das transações se encontram dentro de um intervalo de 0 a 1 *Bitcoin*, em média, como pode ser visto na Figura 3. Apenas o ano de 2014 teve uma média muito maior, por volta de 10 bitcoins. Imagina-se que este fato esteja relacionado a própria introdução do *OP_RETURN* que ocorreu em 2014.

Tipo	Quantidade
Conhecidos	10.279
Desconhecidos	21.681.590
Total	21.691.869

Tabela 2. Quantidade de transações que usaram *OP_RETURN* entre 2014 e 2019, divididos em protocolos conhecidos e não conhecidos pela *API* do site coinsecrets.org.

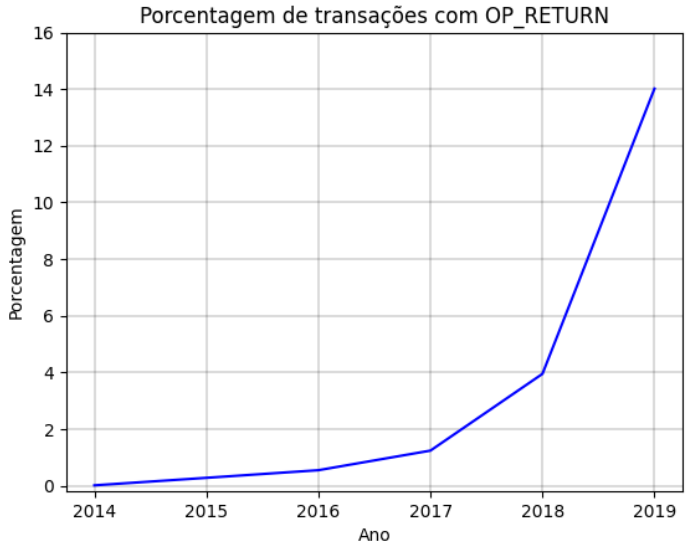

Figura 2. Porcentagem de transações, entre 2014 e 2019, que utilizaram a instrução *OP_RETURN*.

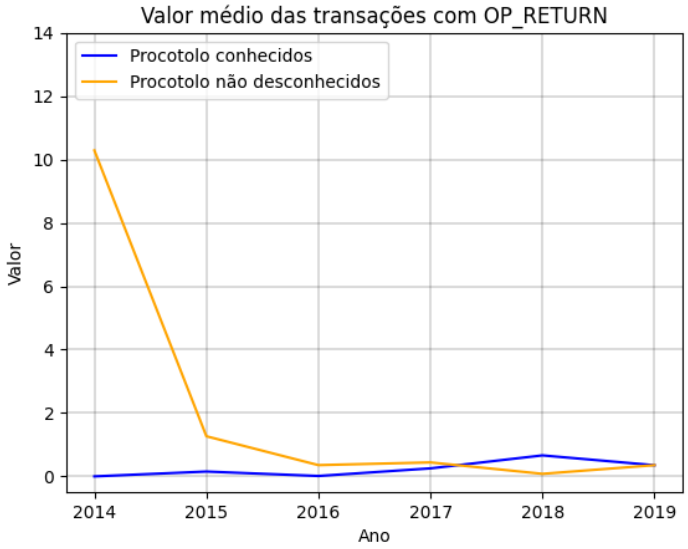

Figura 3. Valores médios das transações realizadas com *OP_RETURN* entre 2014 a 2019.

O próximo passo da análise envolveu identificar conteúdos legíveis presentes no *OP_RETURN*. Com auxílio da função *findall* do *regex* da linguagem Python, foi possível identificar conteúdos legíveis presentes no *OP_RETURN*. Para isso, foi usado como argumento no *findall* uma expressão regular que identifica três caracteres distintos. Escolheu-se esta quantidade de caracteres pois após uma análise preliminar nas transações, notou-se que muitos protocolos são identificados usando, ao menos, três caracteres. Com isso, foi possível encontrar 463.342 potenciais protocolos. Ao realizar a filtragem dos conteúdos, também foi possível contabilizar o número de ocorrência de cada um dos potenciais protocolos. A Tabela 3 contém alguns dos serviços encontrados durante a inspeção dos *OP_RETURN*. Na referida tabela, apenas estão inclusos os protocolos que foram possíveis de serem encontrados em algum tipo de registro ou documentação e que tiveram um alto número de ocorrência, acima de 500 ocorrências. Este número foi escolhido para evidenciar a relevância do serviço sendo utilizado.

Serviço	**Descrição**	**Identificador**
Safex	Usado para *Ecommerce*.	Safex1
Photector (antigo Peirmobile)	Serviço de seguro de transportadoras.	PROTECTOR
Mathwallet	Carteira digital com suporta a diversas cripto moedas.	mathwallet
SolarCoin	É um serviço de recompensa para produtores de energia solar.	SLR
Babel Finance	Plataforma financeira de cripto moedas que disponibiliza serviços diversificados.	BabelBank bitfaith No
Nodeasy	Empresa voltada a análises e monitoramento de *Masternodes*. Além disso, oferecem serviços de desenvolvimento de *Masternode*.	nodeasy
OriginalMy	Serviço voltado ao ramo de autenticação com certificados e assinaturas digitais.	ORIGMY
Omni Layer	Possibilita a criação de ativos através da *Bitcoin*.	omni

Tabela 3. Serviços que utilizam o *OP_RETURN* com a sua funcionalidade e identificador

Além dos potenciais protocolos, também foram encontrados *links* para *downloads* via *torrents* que infringem os direitos autorais de jogos eletrônicos e álbuns de bandas. Além disso, também foram encontradas diversas mensagens arbitrárias como comemorações de ano novo, marcações de algum evento tanto relacionado a *Bitcoins* quanto globais e *memes*. Foi possível encontrar diversas carteiras que, de forma periódica, realizam transações com elas mesmas repassando uma quantidade pequena de *Bitcoin*. O conteúdo encontrado no *OP_RETURN* destas transações não é legível e sempre se encontra quantidades altas de *Bitcoins* repassadas a estas carteiras. A Figura 4 ilustra uma destas carteiras, com uma breve demonstração da sua atividade periódica, que neste caso durou aproximadamente 3 anos.

Para identificar como as carteiras interagem entre si na rede *Bitcoin* foi feito um

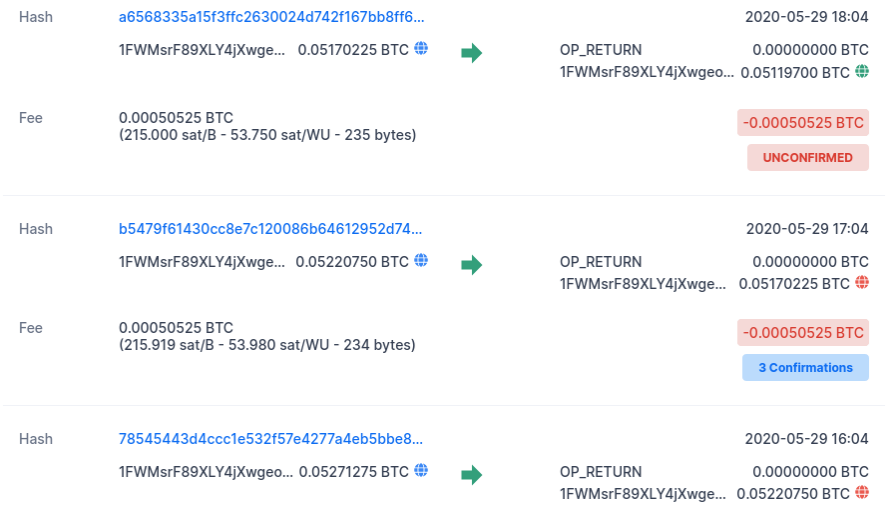

Figura 4. Imagem com atividade de uma carteira. Imagem retirada do site blockchain.com.

agrupamento das carteiras. Carteiras são agrupadas quando um endereço de entrada de uma transação é igual ao de entrada de outra transação e foram apenas incluídas as carteiras que realizaram mais de uma transação. A Figura 5 mostra um exemplo de comportamento das transações que usam *OP_RETURN*. Os vértices destacados em vermelho são as carteiras de entradas e os azuis são os de saída. Na Tabela 4 encontra-se o número de transações que foram agrupadas e a característica deste agrupamento.

Ano	Isolado	uma ou duas saídas	múltiplas saídas	total
2014	402	1143	155	1700
2015	1310	15636	1756	18702
2016	2476	50222	3345	56043
2017	7998	117364	4826	130188

Tabela 4. Informações anuais das clusterizações. Um agrupamento isolado é quando a carteira de entrada é a mesma que a de saída em uma transação, já de uma ou duas e múltiplas saídas será quando os endereços forem diferentes nos campos de entrada e de saída.

Na Tabela 4 observa-se que grande parte das atividades dos agrupamentos está em transações com uma ou duas saídas, mostrando que as atividades dos usuários do *OP_RETURN* são mais isoladas e direcionadas a um ou dois grupos. Além disso, com os dados da Tabela 4 foi possível identificar que apenas uma pequena parcela dos usuários enviou Bitcoins para endereços inválidos ou para nenhum tipo de endereço, fato comprovado ao observar os nós isolados vermelhos, vértices cujos endereços de entradas são os mesmos do que os de saída.

5. Conclusão

Com as análises dos dados foi encontrada uma grande diversidade de utilidades para o *OP_RETURN*, sejam elas utilidades por meios de serviços ou para fins ilegais, vistos nos *links* piratas. Além disso, observou-se como os usuários que utilizam o *OP_RETURN*

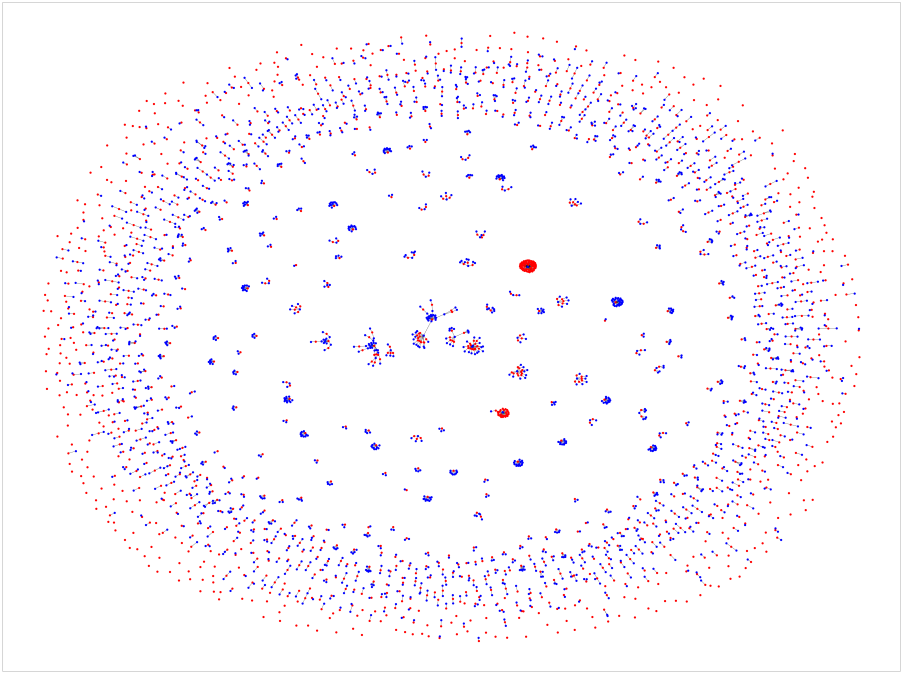

Figura 5. Grafo com a interações das carteiras no ano de 2014. Os vértices vermelhos que não possuem nenhuma conectividade com um vértice azul são classificados como isolados.

interagem entre si, através das análises dos grafos gerados e a quantidade de carteiras pertencentes aos agrupamentos da Tabela 4. Por fim, sugere-se que em estudos futuros sejam feitas análises de conteúdos de outros meios de inserção de dados arbitrários nas transações, como o *unspendable outputs*, que utiliza o campo de *outputs* de uma transação, e traçar um comparativo com os dados inseridos através do *OP_RETURN*, a fim de observar as semelhanças e diferenças entre os tipos de dados encontrados.

Referências

Ali, S. T., McCorry, P., Lee, P. H.-J., and Hao, F. (2018). Zombiecoin 2.0: managing next-generation botnets using bitcoin. *International Journal of Information Security*, 17(4):411–422.

Antonopoulos, A. M. (2014). *Mastering Bitcoin: Unlocking Digital Cryptocurrencies - Andreas M. Antonopoulos - Google Books*.

Bartoletti, M., Bellomy, B., and Pompianu, L. (2019). A journey into bitcoin metadata. *Journal of Grid Computing*, 17(1):3–22.

Bartoletti, M. and Pompianu, L. (2017). An analysis of bitcoin op_return metadata. In *International Conference on Financial Cryptography and Data Security*, pages 218–230. Springer.

blockchain.com. Api blockchain. https://www.blockchain.com/api/blockchain_api.

coinsecrets.org. Api coinsecrets. http://api.coinsecrets.org.

Matzutt, R., Hiller, J., Henze, M., Ziegeldorf, J. H., Müllmann, D., Hohlfeld, O., and Wehrle, K. (2018). A quantitative analysis of the impact of arbitrary blockchain content on bitcoin. In *International Conference on Financial Cryptography and Data Security*, pages 420–438. Springer.

Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system.

opreturn.org. opreturn site. https://opreturn.org/.

Sanchez, M. and Fisher, J. (2019). Proof-of-Proof: A Decentralized, Trustless, Transparent, and Scalable Means of Inheriting Proof-of-Work Security. White Paper, Veriblock.

Strehle, E. and Steinmetz, F. (2020). Dominating op returns: The impact of omni and veriblock on bitcoin. *Blockchain Research Lab*.

Tether Operations (2021). Tether. https://tether.to.