Uma Proposta de Algoritmo para a Detecção de Mixers na Ethereum

Pedro Leale[1], **Ivan da Silva Sendin**[1]

[1]Faculdade de Computação – Universidade Federal de Uberlândia (UFU)
Av. João Naves de Ávila, n^o 2121, CEP 38.400-902 – Uberlândia – MG – Brasil.

{pedro.leale,sendin}@ufu.br

Resumo. *Este trabalho apresenta uma metodologia de detecção de contratos inteligentes do tipo mixers na rede Ethereum. Utilizou-se um modelo de aprendizado de máquina baseado em Random Forest, treinado com transações do Tornado Cash e balanceado com amostras de 100 endereços aleatórios não relacionados a mixers. O modelo foi treinado com dados de março de 2025 e validado em 29/10/2020, dia de alto volume de transações, identificando corretamente 3 endereços do Tornado Cash.*

Abstract. *This work presents a methodology for detecting smart contracts of the mixer type on the Ethereum network. A machine learning model based on Random Forest was employed, trained with transactions from Tornado Cash and balanced with samples from 100 randomly selected addresses unrelated to mixers. The model was trained using data from March 2025 and validated on October 29, 2020—a day with high transaction volume—successfully identifying 3 Tornado Cash addresses.*

1. Introdução

A transparência das blockchains públicas, como Ethereum e Bitcoin, é uma de suas características fundamentais, permitindo que qualquer usuário audite transações em tempo real [Nakamoto 2008, Wood et al. 2014]. No entanto, essa mesma transparência representa um risco à privacidade de indivíduos e organizações, já que endereços e movimentações financeiras podem ser vinculados a identidades reais. Para mitigar esse problema, surgiram os *mixers*, serviços projetados para dificultar o rastreamento de criptomoedas ao misturar transações de múltiplos usuários antes de redistribuí-las [Ziegeldorf et al. 2015].

Embora os *mixers* atendam a demandas legítimas de privacidade, como proteger segredos comerciais ou evitar vigilância excessiva, seu uso também está associado a atividades ilícitas, como lavagem de dinheiro e evasão fiscal. Um exemplo emblemático é o Tornado Cash, *mixer* baseado em Ethereum, cujos desenvolvedores foram presos em 2022 sob acusações de facilitar operações criminosas, levantando debates sobre a regulamentação de tecnologias de privacidade [Brownworth et al. 2024].

A detecção de *mixers* na blockchain ainda é um campo incipiente, com poucas ferramentas disponíveis além da análise manual dos contratos inteligentes ou de heurísticas que correlacionam endereços a serviços conhecidos como Tornado Cash [Tang et al. 2022a]. No entanto, a natureza *open-source* dos contratos de mixers (disponíveis publicamente no

GitHub [1]) permite que agentes maliciosos criem variações não catalogadas, operando sem serem percebidos.

Assim, este trabalho tem como objetivo propor um método de detecção de contratos inteligentes do tipo *mixer* na blockchain Ethereum, com ênfase em transações associadas ao Tornado Cash. A metodologia utiliza aprendizado supervisionado com características estatísticas transacionais extraídas da blockchain.

2. Fundamentação teórica

2.1. Análise Forense de Blockchain

Em geral, as criptomoedas oferecem um mecanismo de privacidade conhecido como pseudo-anonimato: os participantes da rede são identificados por endereços, que correspondem a chaves públicas [Ermilov et al. 2017]. Embora uma mesma entidade possa usar múltiplos endereços para dificultar sua rastreabilidade, todas as transações são publicamente registradas na blockchain. Essa transparência permite que relações entre endereços sejam analisadas, viabilizando a identificação de atividades ilícitas. Por exemplo, quando um endereço associado a crimes envia fundos para uma corretora, as autoridades podem rastrear e identificar o usuário responsável [Team 2023].

2.2. Mixers

Os *mixers* são plataformas que possibilitam ao usuário depositar criptomoedas e combiná-las às criptomoedas de outros participantes, permitindo resgate posterior por meio de outros endereços. O objetivo central destes sistemas é dificultar, ou até mesmo impossibilitar, a rastreabilidade das criptomoedas. Empregando protocolos criptográficos sofisticados - como *Zero-Knowledge-Proofs (ZKPs)* - quando usados corretamente, podem eliminar a maioria dos métodos de rastreamento [Wang et al. 2023a].

Diversos *mixers* foram implementados no ecossistema Ethereum ao longo dos anos. Entre os exemplos mais notáveis estão o *Tornado Cash*, considerado o mais famoso, porém foi envolvido com sanções e problemas jurídicos; o *Mobius* foi uma proposta pioneira de *mixer trustless* baseado em *zk-SNARKs*. Esta solução de mixer diferencia-se por seu mecanismo de compromissos criptográficos [Meiklejohn and Mercer 2018]; já o *Typhoon.Network* [Team 2021] opera na Binance Smart Chain, implementando um sistema de recompensas por liquidez similar ao *anonymity mining*. Cada solução adota mecanismos distintos para ofuscar a origem e o destino dos fundos, variando em eficiência, custo e nível de anonimato proporcionado [Wang et al. 2023a].

No funcionamento do *Tornado Cash*, o usuário realiza um depósito gerando um *commitment* a partir de um segredo, utilizando um hash de Pedersen, e envia-o com uma quantia fixa de ETH ao contrato inteligente. Esse compromisso é adicionado como uma folha em uma árvore de Merkle binária de altura 20, cujos nós internos são calculados com o hash MiMC para otimizar as operações *zk-SNARKs*. Para realizar uma retirada, o usuário — ou, um *relayer* — submete uma prova *zk-SNARKs* que comprova a posse do segredo, atestando que o compromisso está presente na árvore, que o caminho de autenticação é válido e que o *nullifier* derivado ainda não foi utilizado, garantindo que cada depósito só possa ser sacado uma única vez. O contrato verifica a

[1] `https://github.com/tornadocash/tornado-core`

prova e, se válida, transfere os fundos para um novo endereço, dissociando o saque do endereço original. O anonimato é reforçado com o uso de *relayers*, que podem ser utilizados para submeter a transação em nome do usuário, mediante o pagamento de uma taxa, impedindo assim o vínculo direto entre o depositante e o endereço do destinatário [Tor 2019, Youn et al. 2023a].

3. Trabalhos relacionados

Os *mixers*, devido a sua estreita relação com atividades ilegais, atraem um grande interesse e diversos trabalhos abordam o assunto. A detecção de *mixers* no Bitcoin é abordada principalmente com uso de estatística e aprendizado de máquina [Shojaeenasab et al. 2022, Xu et al. 2023, Zola et al. 2025], usando esta abordagem transações suspeitas são detectadas e os endereços participantes são identificados e analisados [Tironsakkul et al. 2020, Gomez et al. 2022]. Nas plataformas baseadas em Contratos Inteligentes, a detecção das transações é tida como mais fácil, pois acredita-se que o endereço do contrato de *mixers* é amplamente divulgado então a observação das transações de depósitos, saques e dos relayers sejam suficiente para identificar os principais participantes. Desta forma, os trabalhos sobre *mixers* em contratos inteligentes atuam sobre a análise da blockchain visando a identificação do participantes [Du et al. 2024, Béres et al. 2020, Tang et al. 2022b, Youn et al. 2023b, Wang et al. 2023b] . Os estudos sobre *mixers* e aumento de privacidade em criptomoedas vão além dos aspectos já citados e envolvem outras criptmoedas[Kappos and Piotrowska 2019]; outras técnicas para ocultar os dados na blockchain[Victor and Weintraud 2021]; e serviços como os de corretoras e apostas e atividades ilegais [Patsakis et al. 2024]. Aspectos gerais sobre sobre o tema podem ser encontrados em [Mariani and Homoliak 2025] e em [Nadler and Schär 2023].

4. Proposta

O algoritmo para detecção de *mixers* proposto utiliza o classificador *Random Forest*, escolhido por sua robustez, interpretabilidade e bom desempenho em cenários heterogêneos [Breiman 2001]. O conjunto de dados original, obtido na plataforma Blockchair [Blo 2023], inclui atributos on-chain como número e posição da transação no bloco (block_id, index), hash único (hash), timestamp (time), status (failed), tipo de chamada (type), endereços de emissor e destinatário (sender, recipient), número de chamadas internas (call_count), valores transferidos (value, value_usd), valores internos (internal_value, internal_value_usd), taxas (fee, fee_usd), consumo de gás (gas_used, gas_limit, gas_price), dados hexadecimais (input_hex), nonce (nonce) e componentes de assinatura (v, r, s).

Para treinar o modelo, derivaram-se quatro métricas de privacidade e atividade: entropia dos valores transferidos (value_entropy), das taxas de gás (fee_entropy), dos valores internos (internal_value_entropy) e média de chamadas internas por transação (call_count_mean). Em seguida, rotularam-se como "possível_mixer" todas as transações de um endereço conhecido do Tornado Cash (depósitos de 10ETH) [Tor 2024] e, para a classe "não_mixer", amostraram-se 100 endereços aleatórios (com pelo menos 10 transações cada), previamente inspecionados por meio da plataforma Etherscan, a fim de garantir que não se tratavam de mixers [Etherscan Blockchain Explorer].

Os registros rotulados foram combinados e balanceados por reamostragem antes de alimentar o `RandomForestClassifier` do `scikit-learn`, configurado com

100 árvores, profundidade máxima de cinco níveis e pesos de classe automáticos. Por fim, implementou-se uma função de inferência que recebe o vetor de métricas de cada transação e calcula a probabilidade de "possível_mixer": se $p > 0{,}5$, classifica-se como mixer; caso contrário, como não_mixer. Essa abordagem fornece previsões interpretáveis e consistentes com os padrões estatísticos observados.

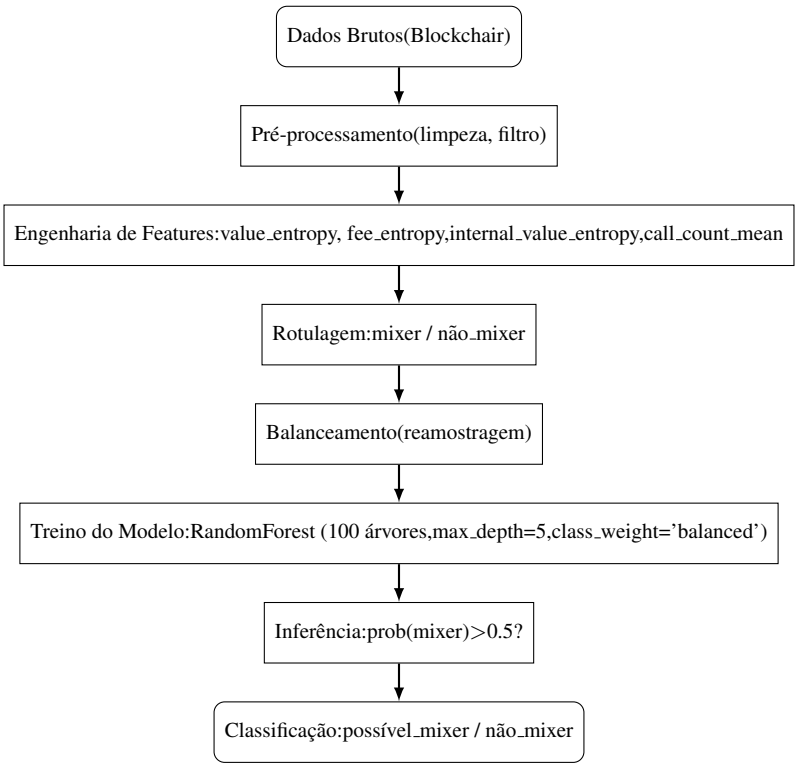

Figura 1. Visão geral do algoritmo

5. Resultados preliminares

O algoritmo de classificação foi aplicado a dados da blockchain Ethereum de março de 2025, mês que apresentou, segundo o Etherscan, picos de transações do Tornado Cash não vistos desde 2022. Embora seja comum treinar modelos em dados passados e validá-los em dados futuros, neste caso o objetivo não é prever, mas identificar padrões estruturais de *mixers*, similares aos do Tornado Cash. Para validação qualitativa, também foram analisados dias com alta atividade do contrato de 10 ETH, como **11/09/2022** e **07/08/2022**. Ressalta-se que o modelo foi treinado exclusivamente com transações do contrato de 10 ETH do Tornado Cash.

Para o teste principal do modelo, foi selecionado o dia **29/10/2020**, data que apresentou um volume elevado de transações envolvendo os contratos do Tornado Cash nas denominações de 0.1 ETH, 1 ETH e 10 ETH. Nesse dia, foram registradas **1.114.702 transações** na blockchain Ethereum segundo os dados fornecidos pela plataforma Blockchair, abrangendo um total de **230.658 endereços distintos**. Especificamente, **239 transações** destinadas ao Tornado Cash, seguindo a distribuição: **73 transações** para o contrato de **0.1 ETH**, **55 transações** para o contrato de **1 ETH** e **111 transações** para o contrato de **10 ETH**.

Ao aplicar o modelo de classificação treinado, obtiveram-se **63 endereços** rotulados como prováveis *mixers*. Destes, **3 pertenciam efetivamente ao Tornado Cash** — e foram precisamente os **3 endereços** associados às 239 transações para esse protocolo naquele dia. Os demais **60 endereços** estavam associados majoritariamente a contratos de trocas de *NFTs*, *liquidity pools*, exchanges descentralizadas e aplicações voltadas a jogos (*GameFi*). A confusão do algoritmo com esse tipo de contrato é compreensível, dado que esses serviços frequentemente exibem alto volume de transações internas e baixa entropia de valores, padrões também observados em mixers.

Curiosamente, ainda que o Tornado Cash opere com denominações fixas por contrato, a entropia dos valores transacionais observados em sua movimentação não é nula, contrariando a expectativa de uniformidade. Isso ocorre porque, nas operações de retirada, o campo externo de valor da transação ('value') é registrado como zero ETH — toda a transferência efetiva de Ether ao beneficiário é executada via chamadas internas ao contrato — e apenas o campo 'internal_value' reflete o montante efetivamente enviado.

6. Discussão

A aplicação do modelo de *Random Forest* demonstrou potencial na identificação de padrões associados a *mixers*, mas revelou uma elevada taxa de falsos positivos em contratos de bolsas de troca de ativos e plataformas de apostas, cujas transações apresentam entropia e múltiplas chamadas internas semelhantes às de serviços de ofuscação. Para mitigar essa limitação, propõe-se a integração de um repositório off-chain de contratos verificados (exchanges, *marketplaces* de NFTs, jogos, etc.), permitindo filtrar esses endereços antes da inferência e concentrar-se em comportamentos transacionais genuinamente ambíguos.

Além disso, sistemas de conversão de criptoativos em moedas fiduciárias (exchanges centralizadas, serviços de *cash-out*, *on/off ramps*) poderiam adotar modelos semelhantes para avaliar se um endereço interagiu com *mixers*, apoiando políticas de conformidade e prevenção de lavagem de dinheiro. Futuramente, a incorporação de features temporais — como a frequência e repetição de uso de *mixers* por um mesmo endereço — poderá aprimorar a acurácia, reduzindo ambiguidades e fortalecendo a detecção de padrões maliciosos.

7. Conclusão

Os resultados demonstram que abordagens de aprendizado de máquina, mesmo com conjunto de treinamento limitado a uma única denominação de *mixer*, podem captar padrões de uso de serviços de privacidade na Ethereum. Contudo, a similaridade de comportamento com contratos legítimos de alto volume exige refinamentos, como a exclusão prévia de contratos conhecidos e a inclusão de atributos temporais — frequência e repetição de uso de *mixers* — para reduzir falsos positivos. Além disso, a integração de sistemas de compliance em corretoras e serviços de *cash-out* pode se beneficiar desta solução para mitigar riscos de lavagem de dinheiro. Futuras pesquisas devem explorar arquiteturas híbridas e bases de dados off-chain para aprimorar a robustez e a aplicabilidade prática da detecção de *mixers* na blockchain.

Referências

(2019). Tornado cash: Privacy solution for ethereum. `https://github.com/tornadocash`.

(2023). Blockchair: Universal blockchain explorer and api. Accessado: 2 de Maio, 2025.

(2024). Tornado.cash: 10 eth mixer contract. `https://etherscan.io/address/0x910cbd523d972eb0a6f4cae4618ad62622b39dbf`. Acesso em: 15-Mar-2025.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Brownworth, A., Durfee, J., Lee, M. J., and Martin, A. (2024). Regulating decentralized systems: Evidence from sanctions on tornado cash. Technical report, Federal Reserve Bank of New York.

Béres, F., Seres, I. A., Benczúr, A. A., and Quintyne-Collins, M. (2020). Blockchain is watching you: Profiling and deanonymizing ethereum users.

Du, H., Che, Z., Shen, M., Zhu, L., and Hu, J. (2024). Breaking the anonymity of ethereum mixing services using graph feature learning. *IEEE Transactions on Information Forensics and Security*, 19:616–631.

Ermilov, D., Panov, M., and Yanovich, Y. (2017). Automatic bitcoin address clustering. In *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 461–466.

Etherscan Blockchain Explorer. Etherscan blockchain explorer. Acesso em: maio de 2025.

Gomez, G., Moreno-Sanchez, P., and Caballero, J. (2022). Detecting cybercriminal bitcoin relationships through backwards exploration.

Kappos, G. and Piotrowska, A. M. (2019). Extending the anonymity of zcash. pages 5–6.

Mariani, J. and Homoliak, I. (2025). Sok: A survey of mixing techniques and mixers for cryptocurrencies.

Meiklejohn, S. and Mercer, R. (2018). Möbius: Trustless tumbling for transaction privacy. *Proceedings on Privacy Enhancing Technologies*, 2018(2):105–121.

Nadler, M. and Schär, F. (2023). Tornado cash and blockchain privacy: A primer for economists and policymakers. *Federal Reserve Bank of St. Louis Review*, 105:122–136.

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

Patsakis, C., Politou, E., Alepis, E., and Hernandez-Castro, J. (2024). Cashing out crypto: state of practice in ransom payments. *International Journal of Information Security*, 23:699–712.

Shojaeenasab, A., Motamed, A. P., and Bahrak, B. (2022). Mixing detection on bitcoin transactions using statistical patterns.

Tang, Y., Xu, C., Zhang, C., Wu, Y., and Zhu, L. (2022a). *Analysis of Address Linkability in Tornado Cash on Ethereum*, pages 39–50.

Tang, Y., Xu, C., Zhang, C., Wu, Y., and Zhu, L. (2022b). Analysis of address linkability in tornado cash on ethereum. In *Communications in Computer and Information Science*, volume 1506 CCIS, pages 39–50. Springer Science and Business Media Deutschland GmbH.

Team, C. (2023). 2023 crypto crime trends: Illicit cryptocurrency volumes reach all-time highs amid surge in sanctions designations and hacking-chainalysis. *Chainalysis Crime Report*, pages 1–8.

Team, T. (2021). Typhoon.network documentation. `https://typhoon-cash.gitbook.io`.

Tironsakkul, T., Maarek, M., Eross, A., and Just, M. (2020). Tracking mixed bitcoins.

Victor, F. and Weintraud, A. M. (2021). Detecting and quantifying wash trading on decentralized cryptocurrency exchanges. In *The Web Conference 2021 - Proceedings of the World Wide Web Conference, WWW 2021*, pages 23–32. Association for Computing Machinery, Inc.

Wang, Z., Chaliasos, S., Qin, K., Zhou, L., Gao, L., Berrang, P., Livshits, B., and Gervais, A. (2023a). On how zero-knowledge proof blockchain mixers improve, and worsen user privacy. In *Proceedings of the ACM Web Conference 2023*, WWW '23, page 2022–2032, New York, NY, USA. Association for Computing Machinery.

Wang, Z., Chaliasos, S., Qin, K., Zhou, L., Gao, L., Berrang, P., Livshits, B., and Gervais, A. (2023b). On how zero-knowledge proof blockchain mixers improve, and worsen user privacy. In *ACM Web Conference 2023 - Proceedings of the World Wide Web Conference, WWW 2023*, pages 2022–2032. Association for Computing Machinery, Inc.

Wood, G. et al. (2014). Ethereum: A secure decentralised generalised transaction ledger. *Ethereum project yellow paper*, 151(2014):1–32.

Xu, C., Xiong, R., Shen, X., Zhu, L., and Zhang, X. (2023). How to find a bitcoin mixer: A dual ensemble model for bitcoin mixing service detection. *IEEE Internet of Things Journal*, 10:17220–17230.

Youn, M., Chin, K., and Omote, K. (2023a). Empirical analysis of cryptocurrency mixer: Tornado cash. In *2023 Congress in Computer Science, Computer Engineering, Applied Computing (CSCE)*, pages 2324–2331.

Youn, M., Chin, K., and Omote, K. (2023b). Empirical analysis of cryptocurrency mixer: Tornado cash. In *Proceedings - 2023 Congress in Computer Science, Computer Engineering, and Applied Computing, CSCE 2023*, pages 2324–2331. Institute of Electrical and Electronics Engineers Inc.

Ziegeldorf, J. H., Grossmann, F., Henze, M., Inden, N., and Wehrle, K. (2015). Coinparty: Secure multi-party mixing of bitcoins. In *Proceedings of the 5th ACM Conference on Data and Application Security and Privacy*, CODASPY '15, page 75–86, New York, NY, USA. Association for Computing Machinery.

Zola, F., Medina, J. A., Venturi, A., and Orduna, R. (2025). Topological analysis of mixer activities in the bitcoin network.