

Session 4

Anonymity - Fungibility

BLOC 514: Emerging Topics in Blockchain and Digital Currency

Session Objectives

- Explain (non-) anonymity in blockchain transactions
- Provide practical advice on how to protect the anonymity of a user transacting over a public blockchain
- Introduce methods user anonymity
- Introduce the role of AI towards (algorithmic) de-anonymization
- Explain fungibility



Conversely to (most forms of) fiat currency, digital currency does not offer strong anonymity of a
user's identity, since all transaction data are publicly available in the Blockchain and can be
subject to all sorts of advanced analysis methods that can reveal a user's identity. In this
session, we discuss methods and tools for increasing the anonymity of digital currency
transactions in practice along with related topics including the role of AI techniques. In addition,
we will discuss a series of topics related to fungibility.

Agenda Slide

- How anonymous is a public blockchain?
- Al & de-anonymization
- Fungibility / NFTs
- Toward more anonymous blockchains
- Zero-knowledge (Zk) proofs
- Zk-SNARKS and Zk-STARKS
- Resources
- Conclusions



How anonymous is a public blockchain?

How anonymous is Bitcoin?: Introduction

- Two basic principles of e-cash (from early 90s):
 - 1. Untraceability: "for each incoming transaction all possible senders are equiprobable".
 - 2. Unlinkability: "for any two outgoing transactions it is impossible to prove they were sent to the same person".
- See the work of T. Okamoto and K. Ohta:
 - Universal electronic cash (1991)



How anonymous is Bitcoin?

- One of the most common misconceptions about Bitcoin is that it's anonymous.
- Sure, public addresses are, in theory, anonymous and do not contain any direct links to one's identity.
- However, in practice, users leave lots of traces when transacting in digital currency. Here are some characteristic cases where public addresses are linked to real-life identities:
 - 1. When you buy goods/services online, providing real-life information, such as name and delivery address, and paying through digital currency, you leave a direct link between the public address you use and yourself known, if all goes well, only to the merchant.
 - 2. When you deposit/withdraw digital currency to/from an online exchange, where you have previously undergone through KYC/AML procedures, such as uploading a government ID and/or utility bill known, if all goes well, only to the exchange.
 - 3. When you publish a donation address to your eponymous blog or e-mail signature known to everyone!



Understanding blockchain anonymity

- Digital currencies work with an unprecedented level of transparency
 - Many users find this difficult to deal with time and experience are needed to adopt good practices.
- All transactions are public, traceable and permanently stored in the blockchain
 - Once an address is used, it becomes permanently tainted by the history of all transactions it is involved in.
 - To protect your privacy, you must use a new address for every new transaction.
 - Using multiple wallets also helps protect your privacy.
 - Protecting your IP is also a good idea.
 - Beware of mixing services: further to potential legality issues, they rely on trusting a third party and are, generally, not effective for large transactions.
- Think of public bitcoin addresses as semi-anonymous in the best case.



From addresses to identities

Since all transactions are logged in the blockchain and are publicly reviewable, by design, anyone can see the flow of digital currency from one address to another.

- Alone, this information is insufficient to infer one's real-life identity. But, even if a single address is linked to a real-life identity, reasonable (although not 100% provable) assumptions can be made about the ownership of many other addresses that can be linked to the original.
 - 1. For example: if we know that George owns address A (because, for example, he has posted it as donation address on his web page) and a transaction occurs from A to B, we can reasonably assume that George has made that payment.
 - 2. If the same transaction also used addresses C and D as inputs and created some change to address E, we can now (again, reasonably) assume that George owns these addresses, too (and, hence, other transactions that show them as input, their other inputs/change addresses, etc.).
 - 3. You can see that a relatively large set of addresses can be linked to a real-world person just by one compromised address and a casual inspection of the blockchain. Much more can be inferred by network analysis and surveillance methods.
- Using a new address for every transaction is designed to make this type of inference more difficult (although, of course, not impossible).



From addresses to identities - example

- Block #394,000 (randomly chosen) contains, among others, a transaction (also randomly chosen), which contains 26 inputs and 4 outputs. If we know that just one of the 26 input addresses belongs to Mr Smith, we can make the following inferences, at least:
 - Mr Smith used to own at least 300 BTC (value of this transaction). More likely, he owns much more see points below. Such knowledge can make Mr Smith a potential target for extortionists or hackers after his digital money.
 - The other 25 input addresses in the same transaction also probably belong to Mr Smith. He may be the owner of at least one of the 4 output addresses, too.
 - At least one of his addresses has been seen in 37 other transactions, of a total value of 11,139.5786 BTC. The same address has appeared as input together with >100 other addresses, which presumably also belong to Mr Smith. A readily available taint analysis of the same address can show which addresses have sent money to Mr Smith, while a reverse taint analysis shows which addresses Mr Smith has sent money to.
 - This is information that we could extract from a casual visual inspection of just one transaction. Imagine what is possible by computerized analysis see next slide.

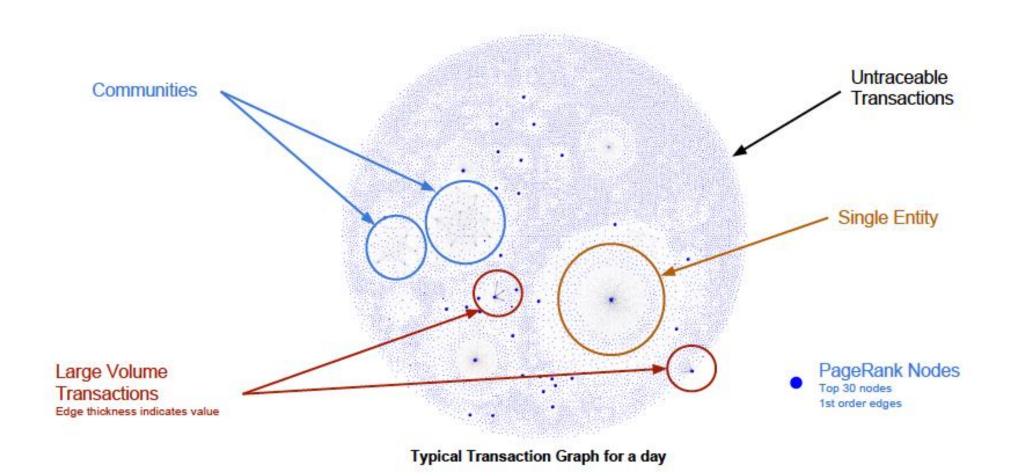


From addresses to identities - research

- Fleder et al. published a <u>paper</u>, in which they demonstrated how far can even simple transaction graph analysis go in connecting public addresses to real-life identities.
 - They annotated the full bitcoin transaction graph by linking public keys to real people either definitively or statistically.
 - They then run the annotated graph to find activity of (known and unknown) users.
 - Just by parsing the bitcointalk.org forum for addresses attached to forum posts, they were able to identify >2,300 users with >2,400 addresses linked to them.
 - Constructing a transaction graph even at a single day's horizon, allowed the authors to identify users engaged in large transactions (amount- or input/output-wise), communities, etc. – see graph on the next slide.
 - Just by combining the previous information, the authors were able to identify: which users had transacted with Silk Road; which users had played SatoshiDICE; which users had transacted with Wikileaks.
- Basically, the only thing limiting analysis capabilities is **volume**: the network graph in the next slide represents a day's worth of data (October 25, 2013, when Bitcoin was much less popular than today) and contains almost 55,000 addresses and 90,000 links between them



One-day transaction graph





5+1 ways to expose your identity

- Here's a list of the most common ways in which a digital currency address can be linked to you:
 - 1. Publishing your address and real name online: Examples include: posting an address for donations on your personal blog; appending an address to your e-mail signature; registering a web page with your real name and posting an address anywhere on it.
 - 2. Exchanging digital/fiat currency in an online exchange. Exchanges are subject to anti-money laundering regulations, so they will ask you to provide government ID, proof of address and other personal information. So, any address you ever use with the exchange (to deposit or withdraw digital currency) is linked to your real-life identity. This is known only to the exchange as well as to regulators, if asked for, and anyone that hacks into the exchange and manages to steal customer data.
 - **3. Buying things with digital currency**. Unless you buy digital goods, which you download directly and anonymously, you will have to provide a name and delivery address. The address you use for payment, plus any change address, will be known to the merchant (for direct payments) or to the payment processor (for indirect payments).
 - **4. Using a lightweight wallet**. All the addresses in your wallet are known to the SPV server (see session 1) that your wallet queries to verify a transaction.
 - **5. Using a web wallet**. All your addresses are known to the wallet operator.
- Unless using a VPN or Tor, your ISP can link your addresses to your IP. Not always true.



Coin tainting

- The expenditure of cryptocurrencies can be tracked since the transactions are formulated as a chain of digital signatures
 - Coin tainting is a way for adding accountability to the network
- Assume a suspicious address
 - The network users can terminate their transactions with this address. As a result the value of their respective coins are demolished.
 - Warnings can be issued: do not accept funds pertaining to the suspicious address.
 - Suspicious addresses can be blacklisted.
- Example: MtGox, a Bitcoin service, traced a large amount (43,000 BTCs) of stolen coin and blocked the accounts that were accepting those coin



Coin tainting

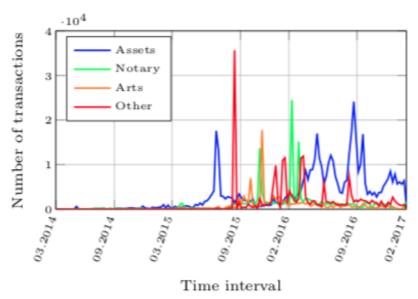
- Disadvantages
 - In principle, coin tainting alters the decentralized aspect
 - Can be abused by opposing parties
- Risks: theoretically, even honest users may become owners of coins of expenditure history
 - Reduced risk for newly mined coins



Al & De-anonymization

Al & De-anonymization: Use case

- You can embed metadata in Bitcoin trnsactions
 - OP_RETURN instruction
- · Research providing an empirical analysis of storing metadata in the Bitcoin blockchain
 - "An analysis of Bitcoin OP_RETURN metadata" by Massimo Bartoletti and Livio Pompianu



Patterns in such metadata may contribute to the discovery of your identity (use of AI)

Al & De-anonymization: Use case

Available in ACM Digital Library

Research article presented at the 2019 International Conference on Big Data and Blockchain (Aug. 2019)

(published under the proceedings of the 3rd International Conference on Vision, Image and Signal – collection of proceedings)

- Basic idea: exploition of various patterns that exist in OP_RETURN's metadata
- Train Machine Learning (ML) models with such patterns
- Experiments: from traditional ML models to deep neural networks → highly accurate identification



AI & De-anonymization: Use case

Research article "Identity Discovery in Bitcoin Blockchain: Leveraging Transactions Metadata via Supervised Learning" (con't)

Findings

- Bitcoin blockchain are pseudo-anonymous
- ML algorithmic approaches can be utilized for discovering the identity of entities involved in blockchain-based transactions
- ML itself does not solve all challenges: to be used in combination with other sources of intelligence, e.g., human intelligence

Potential future directions

 Integration with specific applications (e.g., illegal blockchain-based payments) in collaboration with appropriate authorities



Al & De-anonymization: High-level notes

De-anonymization constitutes a task where AI (specifically Machine Learning) can play a key roles

Two broad approaches

- Classification: data under analysis are automatically assigned a label that describes the respective identity
- Clustering: data under analysis are automatically grouped without the need to know the labels as in the case of classification

In terms of training data

- Classification: for data point, manual annotation (which is costly) is needed
- Clustering: no manual annotation is needed

Models

- Classification: Numerous, spanning from simpler ones up to Deep Neural Networks
- Clustering: Fewer (compared to classification) yet powerful models



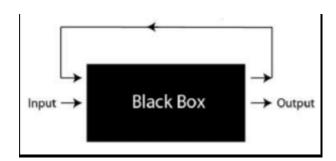
Fungibility / NFTs

Fungibility

- Fungibility is the property of a good or a commodity whose individual units are essentially interchangeable.
- Over the years, it became more clear that bitcoin isn't anonymous. All transactions can be traced
 on the blockchain. A lot of bitcoin tracking companies started to deanonymize users by using this
 data and are actively trying to map the whole bitcoin blockchain.

So how can we avoid this bitcoin tracing?

- Example: Monero uses ring signatures to obfuscate transactions.
- Brief analysis here





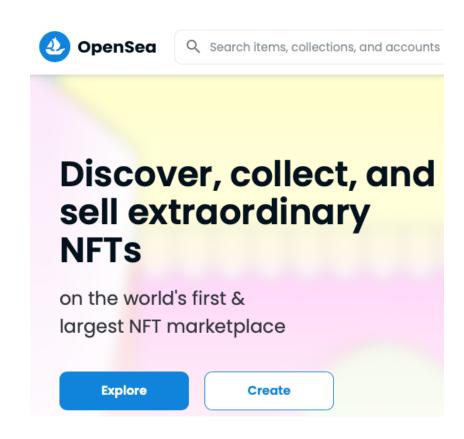
- Blockchain-based assets
- Unique
- Augmented with metadata
- Distinguishable (from each other)
- No trading/exchange at equivalency

Notable use case: **CryptoPunks**



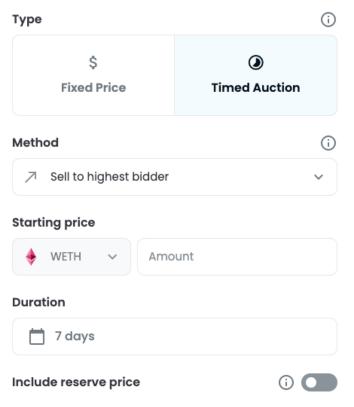
- As expected: Several NFTs market places have appeared
- Indicative example: OpenSea

- Practical issues
 - Centralization
 - Fees + Transaction fees
 - Supported cryptocurrencies + wallet
 - Indexing
 - Etc



Non-Fungible Tokens (NFTs): Selling

- A number of options/parameters are provided when selling
- E.g., see OpenSea's case below





- Powered by smart contracts
- EIP-721: Non-Fungible Token Standard
 - Ethereum Improvement Proposals
 - We will see Bitcoin's Improvement Proposals in this course
- In addition to the standard, we need the implementation(s)
 - See: https://eips.ethereum.org/EIPS/eip-721#implementations

Oxcert ERC721 - a reference implementation

- · MIT licensed, so you can freely use it for your projects
- · Includes test cases
- · Active bug bounty, you will be paid if you find errors

Su Squares – an advertising platform where you can rent space and place images

- Complete the Su Squares Bug Bounty Program to seek problems with this standard or its implementation
- Implements the complete standard and all optional interfaces

ERC721ExampleDeed - an example implementation

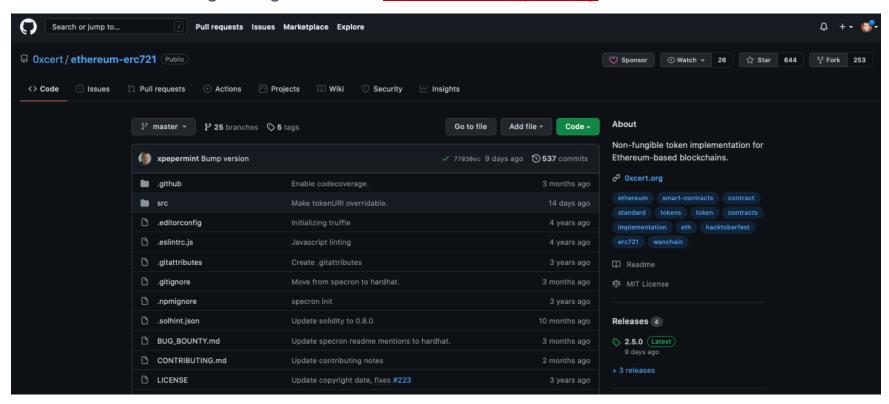
· Implements using the OpenZeppelin project format

XXXXERC721, by William Entriken – a scalable example implementation

• Deployed on testnet with 1 billion assets and supporting all lookups with the metadata extension. This demonstrates that scaling is NOT a problem.



- Oxcert implementation
- Even without a coding background, the <u>source code repository</u> is an excellent info resource



UNIC's NFTs valuation

About the Initiative

The primary project objective is to establish a robust and defensible methodology for calculating Non-Fungible Token (NFT) market capitalizations on an ecosystem basis.

NFTs have grown rapidly in the arts and collectibles space in the last 9 months. As far as we know, our work is the first calculation of the value of all NFTs within a specific NFT ecosystem to create a "market capitalization" for that ecosystem.

The secondary goal is to establish a robust and automated methodology for valuing individual NFTs with relevant confidence intervals.



More info



Toward more anonymous blockchains

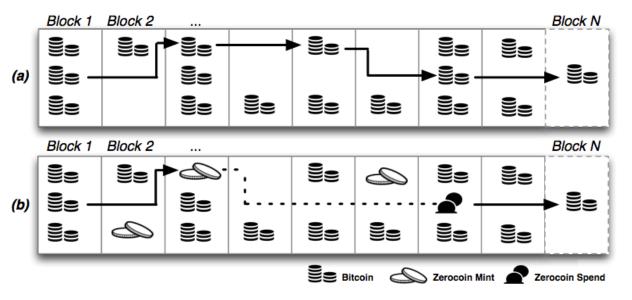
Zerocoin

- Zerocoin (https://en.wikipedia.org/wiki/Zerocoin) is a proposed extension to the Bitcoin protocol that aims at adding true cryptographic anonymity to bitcoin transactions.
- It provides anonymity through the introduction of a **separate crypto-currency** that is stored in the Bitcoin blockchain.
- In essence, Zerocoin implements an online mixing service within the Bitcoin blockchain.
 - Zerocoins are purchased with bitcoin in fixed denominations by a zerocoin mint transaction.
 - Later, these zerocoins can be redeemed for bitcoin to a different bitcoin address by a zerocoin spend transaction.
 - Through the use of cryptographic accumulators and commitments with <u>zero-knowledge proofs</u>, it is not
 possible to link the bitcoin address that was used to mint the original zerocoin to the bitcoin address
 used to redeem the zerocoin.
- The Zerocoin protocol was implemented and released to the public (in testnet) as Moneta in December 2015.



Zerocoin in a nutshell

- Zerocoin operates, roughly, as follows:
 - A user withdraws bitcoin from his wallet and turns them into Zerocoins.
 - Zerocoins from multiple users are then "mixed up".
 - The user can then redeem zerocoins and deposit bitcoins to a new wallet address.
 - Nobody can then link the new bitcoins to the old ones.





Zcash

- **Zcash** is a blockchain that also aims at true cryptographic anonymity.
 - Whereas Zerocoin was constraint to its design by limitations imposed by the Bitcoin protocol, Zcash is not, since it was designed from scratch.
 - It has been developed by the designers of the <u>ZeroCash</u> protocol, Zerocoin developers and leading cryptographers.
 - The principle is similar to Zerocoin, but the cryptography used is the current state-of-the-art, an evolution of zero-knowledge proof constructions, called <u>zk-SNARKs</u>.
 - It has a base non-anonymous currency (basecoin), which can be converted into anonymous coins (zerocoins).
 - It offers both so-called transparent transactions (t-addresses) like Bitcoin, and true anonymous transactions (z-addresses).
 - Transactions using z-addresses are hiding the source, destination and amount involved in the transaction, but they are still verifiable.



Zero-knowledge proofs at a glance

- A zero-knowledge proof can be regarded as a protocol [*]
- Two entities: prover (P) and verifier (V)
- Prover attempts to prove to the verifier that a statement (S) is true without revealing any further info except from the statement itself
- This protocol requires interaction between prover and verifier
- Fundamental properties:
 - 1. Completeness: given a true S, an honest V is convinced
 - 2. Soundness: given a false S, an honest V cannot be convinced
 - 3. Zero-knowledge: given a true S, V knows only S



Zero-knowledge proofs at a glance

Based on the seminal work of S. Goldwasser, S. Micali, and C. Rackoff <u>"The Knowledge Complexity of Interactive Proof-Systems"</u> (1985)

- In summary, this work proposed:
 - Hierarchy of interactive proof systems: computational model in the field of Complexity Theory where the underlying computations are relying on messages exchanged between the involved actors (in this case, prover and verifier)
 - Metric of knowledge complexity: for quantifying the knowledge exchanged between prover and verifier
- Zero-proof knowledge protocols can be evaluated according to the following dimensions:
 - Transparency in terms of trusted configuration
 - Universality
 - Security level wrt attacks triggered from quantum security
 - Ease of respective software implementation
- More info: <u>"Zilch: A Framework for Deploying Transparent Zero-Knowledge Proofs"</u> by Mouris & Tsoutsos



Zk-SNARKS constitute a special case of zero-knowledge proofs

- "Zero-Knowledge Succinct Non-Interactive Argument of Knowledge"
- "S": succinct
 - Practical meaning: small proofs (wrt size), fast proof generation & verification
- "N": non-interactive
 - Practical meaning: communication between prover and verifier takes place only once
- Both "S" and "N" properties are of high importance for blockchain systems
- Interesting reading: <u>"A Review of zk-SNARKs"</u> by Chen et al., especially:
 - Section 1.1: History
 - Section 3.1.2: Zcash Limitations
 - Section 3.2: Tornado Cash (blacklisted by Office of Foreign Assets Control, U.S. Department of the Treasury on Aug 2022)



Zk-STARKS also constitute a special case of zero-knowledge proofs

- New compared to Zk-SNARKS
- "Zero-Knowledge Scalable Transparent Argument of Knowledge"
- "Scalable": Focusing on the enhancement of blockchain scalability
 - Off-chain computation and data storage
- "Transparent": A major difference between Zk-SNARKS and Zk-STARKS
 - Utilization of randomness which is public
 - Note: For Zk-SNARKS, a trusted layer is used
- "Argument of Knowledge": As in the case of Zk-SNARKS, however, a different approach is used regarding the underlying computations which is hash-based and filters out the dependence on trusted layers
- Paper: "Scalable, transparent, and post-quantum secure computational integrity" by E. Ben-Sasson et al.



"Scalable, transparent, and post-quantum secure computational integrity" by E. Ben-Sasson et al.

Scalable, transparent, and post-quantum secure computational integrity

Eli Ben-Sasson* Iddo Bentov[†] Yinon Horesh* Michael Riabzev*

March 6, 2018

Abstract

Human dignity demands that personal information, like medical and forensic data, be hidden from the public. But veils of secrecy designed to preserve privacy may also be abused to cover up lies and deceit by institutions entrusted with Data, unjustly harming citizens and eroding trust in central institutions.

Zero knowledge (ZK) proof systems are an ingenious cryptographic solution to this tension between the ideals of personal *privacy* and institutional *integrity*, enforcing the latter in a way that does not compromise the former. Public trust demands *transparency* from ZK systems, meaning they be set up

Remark: The very first sentence, "Human dignity demands ..."



"Scalable, transparent, and post-quantum secure computational integrity" by E. Ben-Sasson et al.

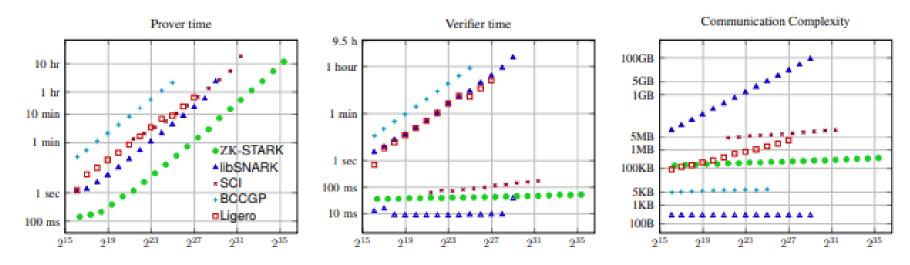


Figure 3: An "apples-to-apples" comparison of different realized proof systems as function of computation size, measured by number of multiplication gates. All systems were tested on the same server (specs below) and executed a computation of size and structure corresponding to the "exhaustive subset-sum" program from [13, Section 3]. The compared systems are SCI (purple x-marks), which lacks ZK, libSNARK (blue triangles), BCGP (cyan +-marks), executed in single-thread mode, Ligero (red squares) and ZK-STARK (green circles). From left to right, we measure prover time, verifier time and communication complexity. For libSNARK, the hollow marks in the middle and right plots measure *only post-processing* verification time and CC, respectively; the full marks measure *total* verification time and CC, and this includes the (non-transparent) key-generation phase. Server specification: 32 AMD cores at clock speed of 3.2GHz, with 512GB of DDR3 RAM. (Each pair of cores shares memory; this roughly corresponds to a machine with 16 cores and hyper-threading.)

Indicative use case: StarkWare

- Development of solutions based on Zk-STARKS
- Two main solutions

1. StarkNet

- L2 network over Ethereum
- Permissionless, decentralized

2. StarkEx

- As a permissioned, centralized version of StarkNet
- 1. Can be customized for decentralized applications



Zero-knowledge proofs: Ethereum rollups

Relation between zero-knowledge proofs and Ethereum

- What: Zero-knowledge rollups
- Why: Increase throughput (Layer 2)
- How: Conduct computation and state storage off-chain
- There are 2 basic technological components
 - 1. Rollup protocol: Implemented as a smart contract, thus, lives on-chain
 - 2. Computations and state storage take place in off-chain virtual machines
- The underlying Layer 1 ensures –for Layer 2 applications- the following properties:
 - 1. Access to secure data
 - 2. Validated transactions in the sense of finality
 - Mechanisms against censorship: transact directly with Layer 1 if censorship at Layer 2 is suspected



Zero-knowledge proofs: Ethereum rollups

More information in this ethreum.org article including the following pros and cons

Pros	Cons
Validity proofs ensure correctness of off-chain transactions and prevent operators from executing invalid state transitions.	The cost associated with computing and verifying validity proofs is substantial and can increase fees for rollup users.
Offers faster transaction finality as state updates are approved once validity proofs are verified on L1.	Building EVM-compatible ZK-rollups is difficult due to complexity of zero-knowledge technology.
Relies on trustless cryptographic mechanisms for security, not the honesty of incentivized actors as with optimistic rollups.	Producing validity proofs requires specialized hardware, which may encourage centralized control of the chain by a few parties.
Stores data needed to recover the off-chain state on L1, which guarantees security, censorship-resistance, and decentralization.	Centralized operators (sequencers) can influence the ordering of transactions.
Users benefit from greater capital efficiency and can withdraw funds from L2 without delays.	Hardware requirements may reduce the number of participants that can force the chain to make progress, increasing the risk of malicious operators freezing the rollup's state and censoring users.
Doesn't depend on liveness assumptions and users don't have to validate the chain to protectheir funds.	Some proving systems (e.g., ZK-SNARK) require a trusted setup twhich, if mishandled, could potentially compromise a ZK-rollup's security model.
Better data compression can help reduce the costs of publishing calldata on Ethereum and	



minimize rollup fees for users.

Topics for discussion

- Given the recent advances in the area of NFTs, a (considerably) re-visited framework with regards to anonymity (+ security).
 - First of all, why (is the above statement true)?
 - If yes, when this is of greater importance?
 - Have two different frameworks (one for cryptos and one for NFTs)?
 - (Etc.)



Conclusions

Conclusions

- Most digital currencies that are based on public blockchains have not been designed for anonymity.
- Instead, users are expected to behave in ways that protect their real-life identities:
 - Not reusing addresses and not publishing address information online are the simplest means of privacy protection.
 - Protecting IPs is also a helpful strategy.
- Advanced network analytics can be used to infer real user identities from publicly available blockchains, especially when combined with knowledge obtained by careless user behavior.
- A number of implementations have been put forward to increase anonymity, and even provide provable anonymity, in digital currency transactions
- Given the recent advances in the area of NFTs, a (considerably) re-visited framework with regards to anonymity as well as security. The exact setup seems to be case-specific depending on factors such how you manage your assets, and the exact type of NFTs.



Bibliography

References

- Karame and Elli Audroulaki (2016). "Bitcoin and Blockchain Security". Artech House, Inc., Norwood, MA, USA.
 - Chapter 5
- Research articles
 - Mauro Conti, Sandeep Kumar E, Chhagan Lal and Sushmita Ruj (2017). "A survey on security and privacy issues of bitcoin" (Section V)
 - https://arxiv.org/pdf/1706.00916.pdf
 - Massimo Bartoletti and Livio Pompianu (2017) "An analysis of Bitcoin OP RETURN metadata" https://arxiv.org/pdf/1702.01024.pdf



Additional Bibliography (optional)

- Reid, F. and Harrigan, M. An Analysis of Anonymity in the Bitcoin system, available here.
- Ron, D. and Shamir, A. *Quantitative Analysis of the Full Bitcoin Transaction Graph*, available here.
- Meiklejohn, S. et al. A Fistful of Bitcoins: Characterizing Payments Among Men with No Names, available <u>here</u>.
- Fleder, M. et al. Bitcoin Transaction Graph Analysis, available here.
- Ober, M. et al. Structure and Anonymity of the Bitcoin Transaction Graph, available here.
- Bartoletti, M. et al. A general framework for blockchain analytics, available <u>here</u>.





Instructor's Email:

iosif.e@unic.ac.cy

