Privacy in Cryptocurrencies: An Overview

Yi Sun and Yan Zhang

Yi Sun · Follow
11 min read · Oct 25, 2018

Listen Share

(This article is part of a series whose later parts can be found here and here.)

Cryptocurrencies are often portrayed as anonymous in the media, yet other articles suggest that they can be traced more easily than ordinary fiat currencies like the dollar. To reconcile these two narratives, it is important to understand: *What does it mean for a cryptocurrency to be **private***? The question is harder to answer than it appears, because privacy can mean many different things in blockchains. To be a more informed developer, investor, or participant in cryptocurrencies, it is important to understand **what privacy actually means** in a cryptoeconomic system. We wrote this post to share our perspective on this skill.

Imagine that Alice starts a Venmo account, which requires her to provide and verify her real name. Because Venmo knows her real name and can potentially share it with others, Alice loses some privacy of identity. If Bob sends Alice $20 on Venmo and shares it on her feed, then her transaction information is public, but she has some privacy over her balance to the extent that only Venmo knows it. On the other hand, suppose that Alice creates a Bitcoin address and asks Bob to instead send her $20 worth of BTC. Compared to Venmo, Alice gains some privacy of identity because her real name is not linked to her Bitcoin address. However, the fact that BTC was transferred from Bob's address to Alice's address as well as the final balance of Alice's address will both be publicly visible to anyone in the world. As we can see, by using Bitcoin, Alice has gained privacy in some sense but lost it in others.

Tradeoffs like these are common when using different cryptocurrencies. We understand them by addressing three main aspects of privacy in the context of

cryptocurrencies, corresponding to:

- the **identity** of the user performing an operation using the cryptocurrency;
- the **transaction data** specific to the operation the user is performing; and
- the **total blockchain state** formed by combining the knowledge of all transactions.

The protocol can use cryptography to make different parts of each of these aspects impossible (or theoretically extremely difficult) to know or compute from the perspective of an external observer. At the same time, attackers who want to discover an attribute of the blockchain can combine disparate pieces of information to conclude or guess what they want to know. The goal of privacy is then to design the protocol to reveal as little information as possible about specific attributes to potential attackers.

Importantly, whether a specific attribute is private may not be black and white. For example, it might be known to some observers but not others, or observers might be able to guess it with some probability, but not with total certainty. This ambiguity means general statements such as "MyFavoriteCoin guarantees privacy" or "MyFavoriteCoin is more private than YourFavoriteCoin" often do not make sense. When used carelessly, such statements can lead to confusion and misinterpretation, and they can be used to mislead for precisely this reason. We believe it makes more sense to make more nuanced statements like "transaction amounts are private in Monero" or even "sender addresses are private up to some anonymity set in ZCash". As we will see later, in some cases cryptographic tools such as zero-knowledge proofs can help us quantify such statements and even provide rigorous proofs for them. We now dive into each aspect of privacy as it relates to cryptocurrency.

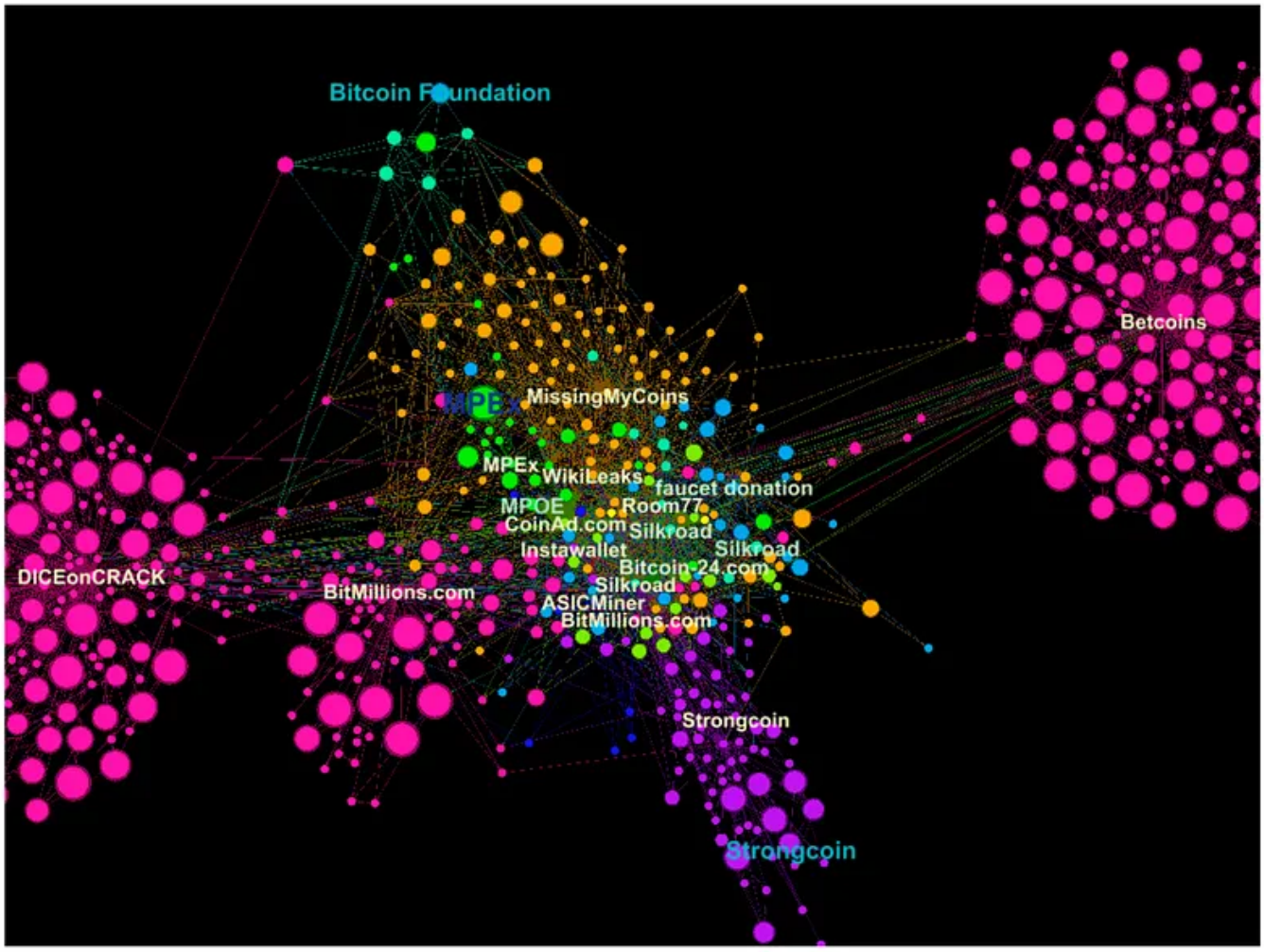
Privacy of Identity (Anonymity)

One of the first things people think of when they hear "privacy" is anonymity, meaning that users' actions are not linked to their **real-world identities**. One version of this type of privacy called **pseudonymity** is fairly easy to achieve; in fact, we already do it in most of our interactions with online services by providing a pseudonym such as bitcoinlover2008@gmail.com instead of an actual name. In this situation, the physical / legal identity of bitcoinlover2008@gmail.com (say Alice Jones) would not appear in most interactions with everyone else in the protocol.

In most cryptocurrencies such as Bitcoin, users are given a public/private key signature pair, where the **public key** is the analogue of username and the **private key** is the analogue of password. The important property is that only someone who knows the correct private key (legitimately or by theft) can generate messages that are "signed" by you, in the sense that anyone can check (with your public key) that those messages came from someone with the private key (with extremely high probability). This property allows users to receive e.g. Bitcoin at one of several public keys, or **addresses**, they control and to send bitcoin using their **private key,** all without the intervention of a centralized authority. These concepts form a bedrock of modern mathematical cryptography and are interesting in their own right. However, for our purposes, having a private/public key pair is just a way to implement pseudonymity in a decentralized context.

Pseudonymity is usually intrinsic to cryptocurrency protocols, leading to the mistaken impression among the media and general public that all cryptocurrencies are anonymous, or at least a lot more anonymous than just having a pseudonym. Not surprisingly, this misunderstanding has driven users to leverage cryptocurrencies for various illicit purposes such as online gambling or transactions on the dark web. However, these users may be sorely disappointed by the level of anonymity they actually receive. While it is true that users send and receive coins using public key addresses, meaning that no real names are seen during Bitcoin transactions, certain actions that users take can link their public addresses to their real identities in other ways.

First, most users purchase bitcoin using fiat currency at an exchange. Transacting in fiat usually requires interacting with the ordinary banking system and therefore proving real world identity. Because all transaction data in Bitcoin is totally public (as discussed in the next section), this allows anyone with access to the exchange database to link addresses to real names. For instance, if Alice withdraws 0.1 BTC from Coinbase to an address she controls, e.g. 36n452uGq1x4mK7bfyZR8wgE47AnBb2pzi, then Coinbase can link her real name to that address. If she then sends the 0.2 BTC to the address of a known illicit online sports betting site, then an external observer may be able to deduce (and provide *immutable public evidence*) that Alice is participating in illegal gambling activity. Techniques such as these are called **blockchain analysis**, and they have been used by companies like Chainalysis to cluster public keys by ownership, link public keys to real identities, and analyze transaction flows.



An early example (2009–2012) of Bitcoin blockchain analysis. Source: https://doi.org/10.3390/fi8010007

Second, making a cryptocurrency transaction requires sending some information over the internet. In some circumstances, metadata from that interaction can be used to trace the IP address of the user initiating the transaction, even if an anonymization service like Tor is used. These two reasons combine to make transacting anonymously using the pseudonymity property of cryptocurrencies alone a difficult task.

Privacy of Transaction Data

When people talk about private cryptocurrencies, they usually mean that *some aspects* of the **transactions** are private. Loosely speaking, a **transaction** is an action a

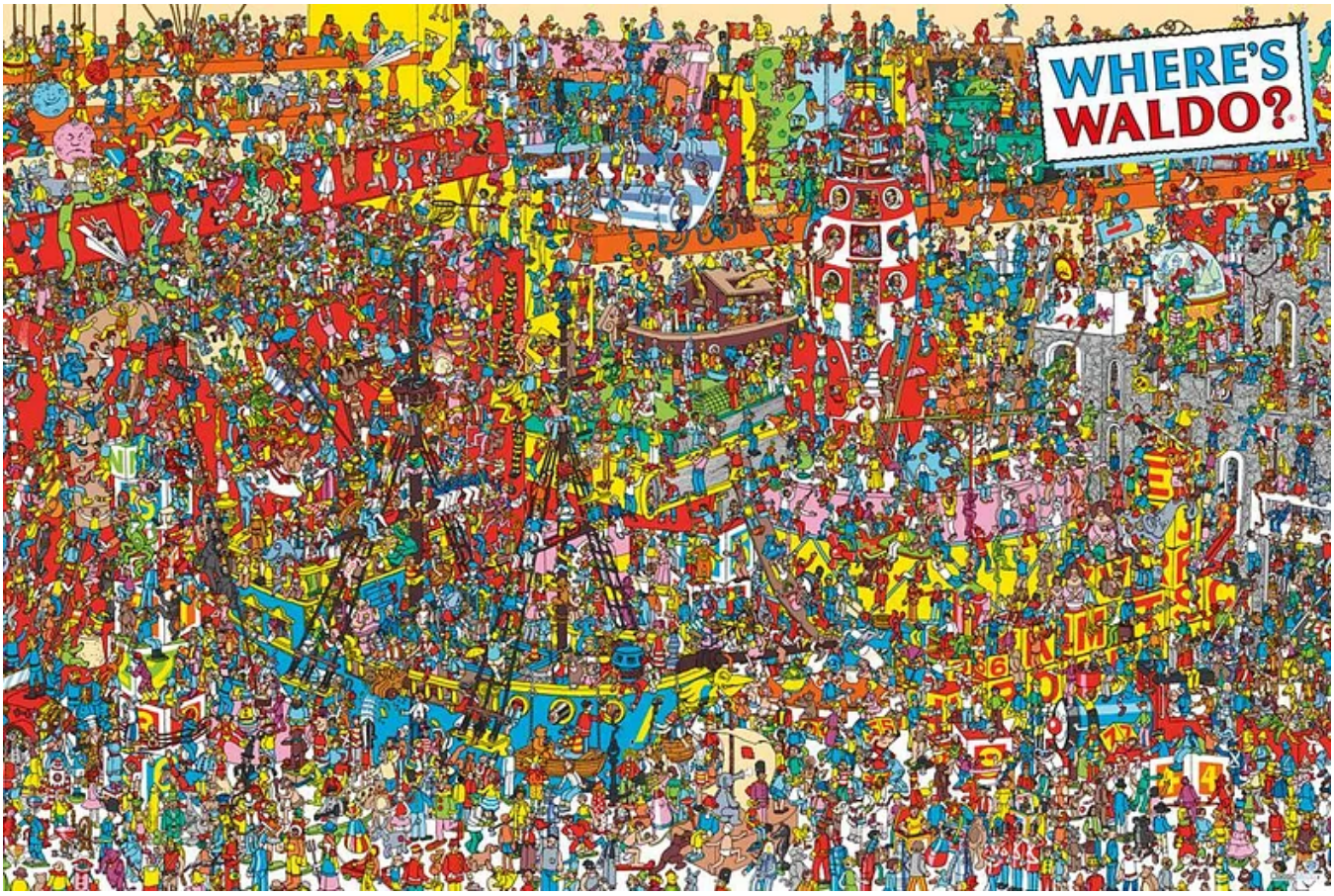
a God-eyed-view would look like:

- one of Alice's addresses, e.g. 36n452uGq1x4mK7bfyZR8wgE47AnBb2pzi
- the linkage between Alice and Bob's address

- one of Bob's addresses
- the amount of coins sent

More complicated transactions can contain other types of information, such as smart contract code in Ethereum. Different blockchains represent all of this transaction data in different ways, and some of these representations allow for different aspects to be private from a third-party observer who sees only the raw data on the blockchain. This is why this section is named "Privacy of Transaction Data" instead of "Privacy of Transactions", since different types of transaction data can be private to different extents.

The most important attributes which can be made private are the addresses of Alice and Bob. If these are private, the sender and recipient of a transaction cannot be identified from the transaction itself. This can help to stymie the type of blockchain analysis we described earlier. That is, if Alice purchases Monero, a cryptocurrency with this feature, from Binance, a popular exchange, and withdraws it, Binance will not be able to link that withdrawal to any further transactions Alice makes with those coins. Conversely, if Bob receives Monero from Alice, he will not be able to know that it was purchased by Alice on Binance.



A cartoon anonymity set. Source: https://www.globalprints.com/products/wheres-waldo-toys-24×36-flm01425

To further complicate things, whether a piece of transaction data is private is not binary. In the example of Alice's address, this can be measured by the size of the **anonymity set**, which is defined to be the smallest collection of addresses the sender of a transaction can be narrowed down to based on blockchain data alone. The larger the anonymity set, the less information about the sender there is in the blockchain data. For example, in Bitcoin, the anonymity set of the most straightforward type of transaction has size 1, since the sender's address is in the transaction itself, while Monero can offer significantly larger anonymity sets.

Privacy of State

In Bitcoin, all transaction data is public, meaning that an outside observer who sees all blocks in the Bitcoin chain is able to reconstruct the ledger of balances belonging to addresses (though these balances may be divided into different UTXO's), which we refer to as **the total state** of the blockchain. However, if some aspects of transactions are private, then knowledge of the entire blockchain does not give a user knowledge of the total state. Instead, this knowledge is shared between different users, with the blockchain ensuring consistency between their knowledge.

Though a user's knowledge of a specific attribute of the blockchain state depends only on the protocol and her knowledge of the transactions which led to the creation of that state, the link between the two can create complex interactions. As a result, different attributes of the state can be private to different extents, with a few examples including:

- the list of all addresses
- the balance of a specific address, e.g. 0x2569C92345013F55CFb47C633c57F2f5756B9acA has 1 ETH
- the smart contract code at a specific address, e.g. the code for the CryptoKitties contract at 0x06012c8cf97BEaD5deAe237070F9587f8E7A266d
- contract-specific state, e.g. data stored by the CryptoKitties contract

As an easy example of the possible range of deductions, in ZCoin the amount of each transaction is public but the sender and recipient are private, which means that user balances still remain private. On the other hand, in Mimblewimble, the amount of

each transaction is private, but the sender and recipient are public, providing another way to ensure privacy of user balances. In fact, users in Mimblewimble must store their own balances, with the blockchain only storing enough information to ensure that no user can spend a balance that exceeds what they truly own.

While in most circumstances making additional attributes of a transaction private is beneficial for individual users, this is not always true of the blockchain state. For example, if the total number of coins in a cryptocurrency is private, it becomes impossible for users to verify properties of the protocol like the total supply schedule. In particular, it would be difficult to detect an attacker who uses a cryptographic vulnerability or backdoor in the protocol to mint new coins in an unauthorized way.

Privacy Properties of Some Existing Protocols

	Bitcoin	Ethereum	Lightning Network / State Channels ** = to participants outside channel	Plasma ** = to participants outside the chain	Bitcoin sidechains ** = to participants outside the chain	Monero ** = up to finite anonymity set	Mimblewimble	ZCoin ** = for private tx	ZCash ** = for shielded tx	Dash	Verge
Blockchain analysis possible?	yes	yes	for opening / closing states	for entry / exit tx	for entry / exit tx	partially	yes	among public tx	among unshielded tx	yes	yes
Sender address of transaction	public	public	private outside channel; public within channel	entry / exit public; up to operator within chain	entry / exit public; up to sidechain protocol within chain	private **	public	private **	private **	public	public
Recipient address of transaction	public, optionally unlinkable	public, optionally unlinkable	private outside channel; public within channel	entry / exit public; up to operator within chain	entry / exit public; up to sidechain protocol within chain	public, unlinkable	public, unlinkable	private **	private **	public, optionally unlinkable	public, unlinkable
Linkage of sender and recipient address	public (private up to finite anonymity set with CoinJoin)	public	private outside channel; public within channel	entry / exit public; up to operator within chain	entry / exit public; up to sidechain protocol within chain	private **	private up to finite anonymity set	private **	private **	public (private up to finite anonymity set with PrivateSend)	public
Transaction amount / data	public	public	opening / closing states public; intermediate states private **	entry / exit public; intermediate states private **	entry / exit public; intermediate states private **	private	private	public	private **	public	public
List of addresses	public	public	public	public	public	private **	public	private **	private **	public	public
Balances / smart contract code	public	public	opening / closing states public; intermediate states private **	entry / exit public; intermediate states private **	entry / exit public; intermediate states private **	private	private	public	private **	public	public

Different Approaches to Privacy

We have focused our attention thus far on whether particular pieces of information are private or public. It is also helpful to organize approaches to privacy by what techniques they use. We end with a rough overview of these approaches. In future posts in this series, we will discuss the mixing and zero-knowledge based approaches in more technical detail. *[Part 2 on mixing is here.]*

- **Second-layer protocols** like lightning network, state channels, or Plasma on top of a "base layer" cryptocurrency allow small groups of users to transact among themselves "off-chain". This means **all intermediate state is stored between**

those users and only periodic summaries of state changes are written to the main blockchain. As a result, the intermediate states are invisible to outside observers because they never appear on the main blockchain at all. Of course, the second-layer protocol itself can have (or choose not to have!) different levels of privacy for off-chain states among its participants, so this is more of a meta-idea than a privacy technique. As such, we will not focus on second-layer protocols further, although a vast body of work lies behind them for the interested reader.

- **Mixing-based approaches** take inputs and outputs of different transactions and combine them into a single large transaction to obscure links between addresses of senders and recipients. In a followup post, we will discuss privacy protocols motivated by mixing and its extensions. These include some of the earliest approaches to privacy such as tumblers, CoinJoin, Mimblewimble, and Monero.

- **Zero-knowledge based privacy** comes when users of the protocol supply **zero-knowledge proofs (ZKP's),** i.e. data which demonstrates knowledge of a piece of information without revealing the information itself. When used correctly, this cryptographic technique can ensure both privacy of transactions/state and soundness of the blockchain. In a followup post, we will focus on the main ideas behind zero-knowledge proofs and the flavors of ZKP applicable to cryptocurrencies such as zk-SNARK's and zk-STARK's.

We mention also a few other techniques which have some impact on privacy.

- **User Best Practices.** Even when using a cryptocurrency without any additional privacy features, users can mitigate some effects of network and blockchain analysis. To combat the use of network metadata to deanonymize users, users can use Tor or I2P to hide the origin of their transactions. To combat blockchain analysis, users are generally advised to use a new address for every payment they receive. Cryptocurrencies like Monero and Verge offer this feature as a native option (though in some cryptocurrencies these addresses can still be linked by later user behavior).

- A **Trusted Execution Environment** (TEE) is a processor (such as Intel SGX) which claims to cryptographically protect the integrity and confidentiality of data and code within it. Several protocols including Ekiden (commercialized by Oasis Labs) are proposing to use TEE. For example, account balances can be encrypted with a private key stored in the TEE so that they are only decrypted

and modified in the TEE. This offloads privacy guarantees to the TEE, which comes with its own vulnerabilities. For example, side channel attacks may be able to extract keys (such attacks have already occurred on Intel SGX), and existing TEE's may require a license from the manufacturer or allow the manufacturer to break confidentiality (though alternatives like Keystone and Gradient attempt to address these).

Conclusion

When thinking about privacy in cryptocurrencies, instead of vague general statements such as "MyCoin is more private than HerCoin," we recommend trying to answer the following questions: Certain pieces of knowledge (*which*?) about the state of the world (*when*?) are private (*to what extent?*) to particular people (*whom*?). This mentality allows us to more objectively analyze privacy technologies and the tradeoffs they make. In future posts, we dig deeper and more quantitatively at specific approaches, while still being motivated by qualitative understanding. *[Part 2 on mixing is here, and part 3 on zero-knowledge is here.]*

Thanks to Brian Gu, Mihaly Barasz, and Sizhao Yang for reviewing drafts of this post.

Bitcoin Cryptocurrency Privacy Ethereum Blockchain

Follow

Written by Yi Sun

179 Followers

More from Yi Sun