Anonymity and Privacy

50.037 Blockchain Technology
Paweł Szałachowski

Anonymity

- “*Anonymity ensures that a user may use a resource or service without disclosing the user’s identity. The requirements for anonymity provide protection of the user identity. [...] Anonymity requires that other users or subjects are unable to determine the identity of a user bound to a subject or operation.*”

Privacy

- “*Ability of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others.*”

Anonymity and Privacy

- Anonymity is about hiding identity
- Privacy is about hiding information/actions
- Why we need these properties?
    - Social, political, work, economical, …
- Anonymity and privacy in the context of cryptocurrencies
    - Is good for users/stakeholders/society?
        - Negative vs positive sides

Bitcoin & Anonymity

Is Bitcoin anonymous?

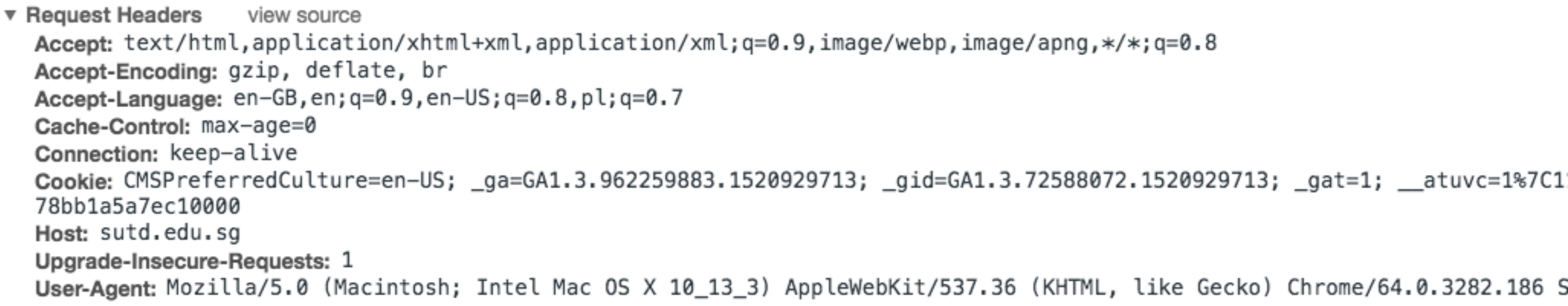
- Anonymity
  - Pseudonymity: use pseudo-identity instead of real identity
    - Anonymity vs Pseudonymity: 4Chan vs Reddit
  - Unlinkability: users interactions cannot be linked by an adversary
- Blockchain is publicly available
  - Each transaction is visible and signed (non-repudiation)
- Linking addresses to identities is sometimes easy
  - Wallets, services, exchanges, merchants, shops, side-channels, …

Unlinkability & Bitcoin

- It should be hard to link:
    - together different addresses of the same user
    - together different transactions made by the same user
    - the sender of a payment to its recipient
        - doesn’t have to be a single transaction
- The third requirements looks pretty challenging
    - Adversary model
    - Anonymity set of your transactions is the set of transactions which the adversary cannot distinguish from your transaction

Anonymity set

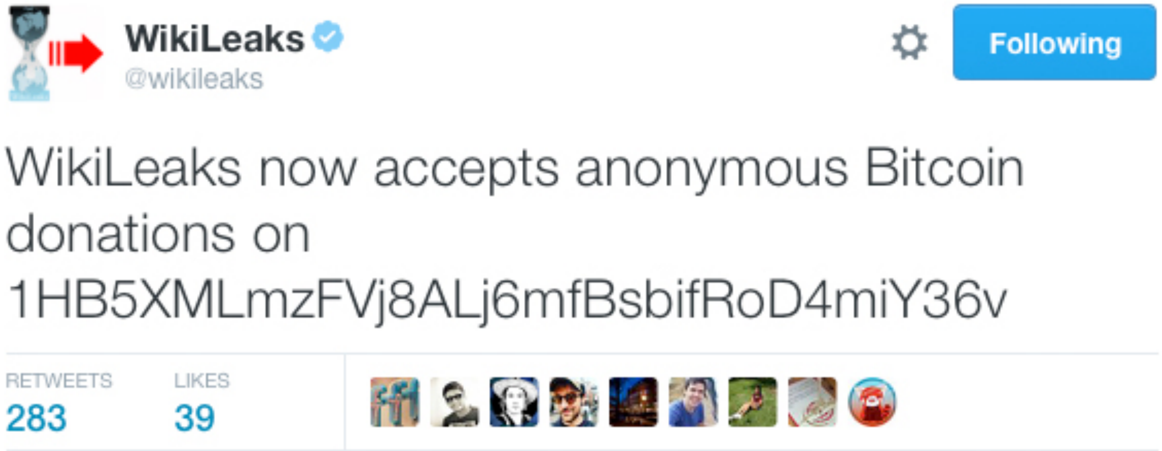
- Other examples of anonymity sets
  - Voting
  - Web browsing
    - Can sutd.edu.sg minimize my anonymity set when I connect to it locally (even with the incognito mode)?



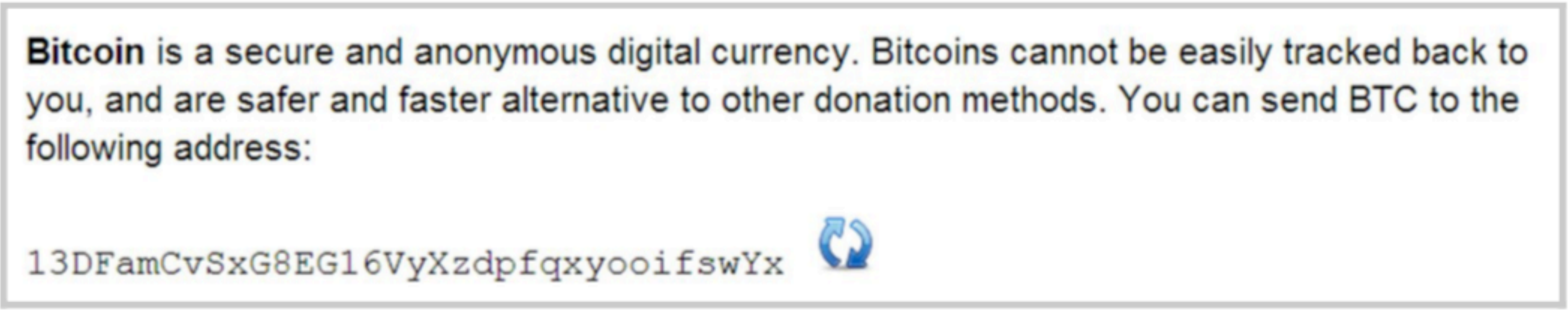
```
▼ Request Headers     view source
   Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,image/apng,*/*;q=0.8
   Accept-Encoding: gzip, deflate, br
   Accept-Language: en-GB,en;q=0.9,en-US;q=0.8,pl;q=0.7
   Cache-Control: max-age=0
   Connection: keep-alive
   Cookie: CMSPreferredCulture=en-US; _ga=GA1.3.962259883.1520929713; _gid=GA1.3.72588072.1520929713; _gat=1; __atuvc=1%7C1
   78bb1a5a7ec10000
   Host: sutd.edu.sg
   Upgrade-Insecure-Requests: 1
   User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10_13_3) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/64.0.3282.186 S
```

Deanonymization

- Track addresses and payments
  - Transaction graph analysis
- Easy to create new random addresses (best practice)
- Good enough?



WikiLeaks now accepts anonymous Bitcoin donations on 1HB5XMLmzFVj8ALj6mfBsbifRoD4miY36v

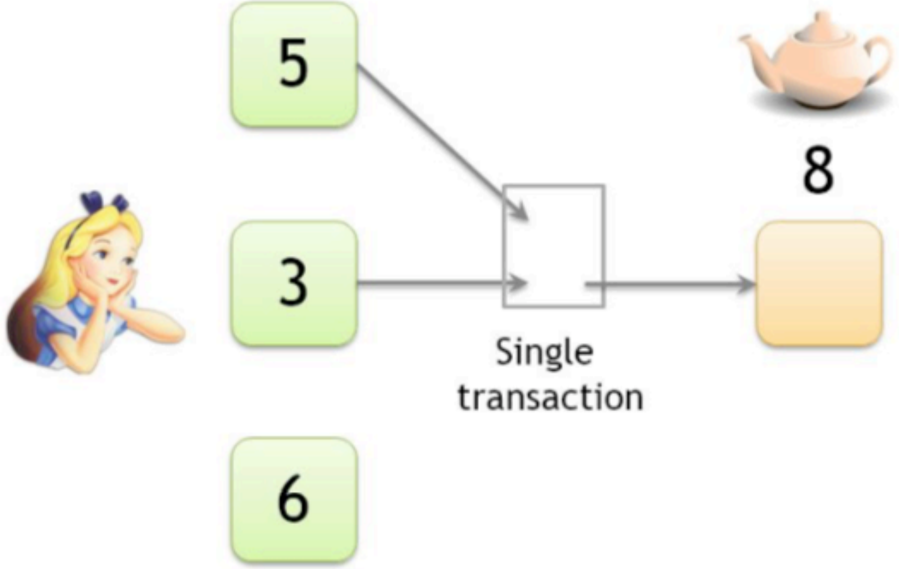


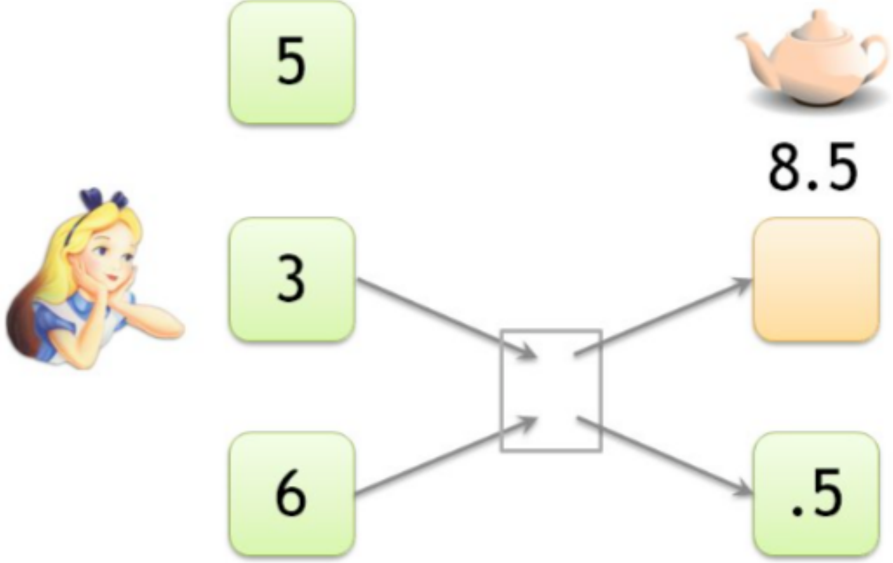
Bitcoin is a secure and anonymous digital currency. Bitcoins cannot be easily tracked back to you, and are safer and faster alternative to other donation methods. You can send BTC to the following address:

13DFamCvSxG8EG16VyXzdpfqxyooifswYx

Linking addresses

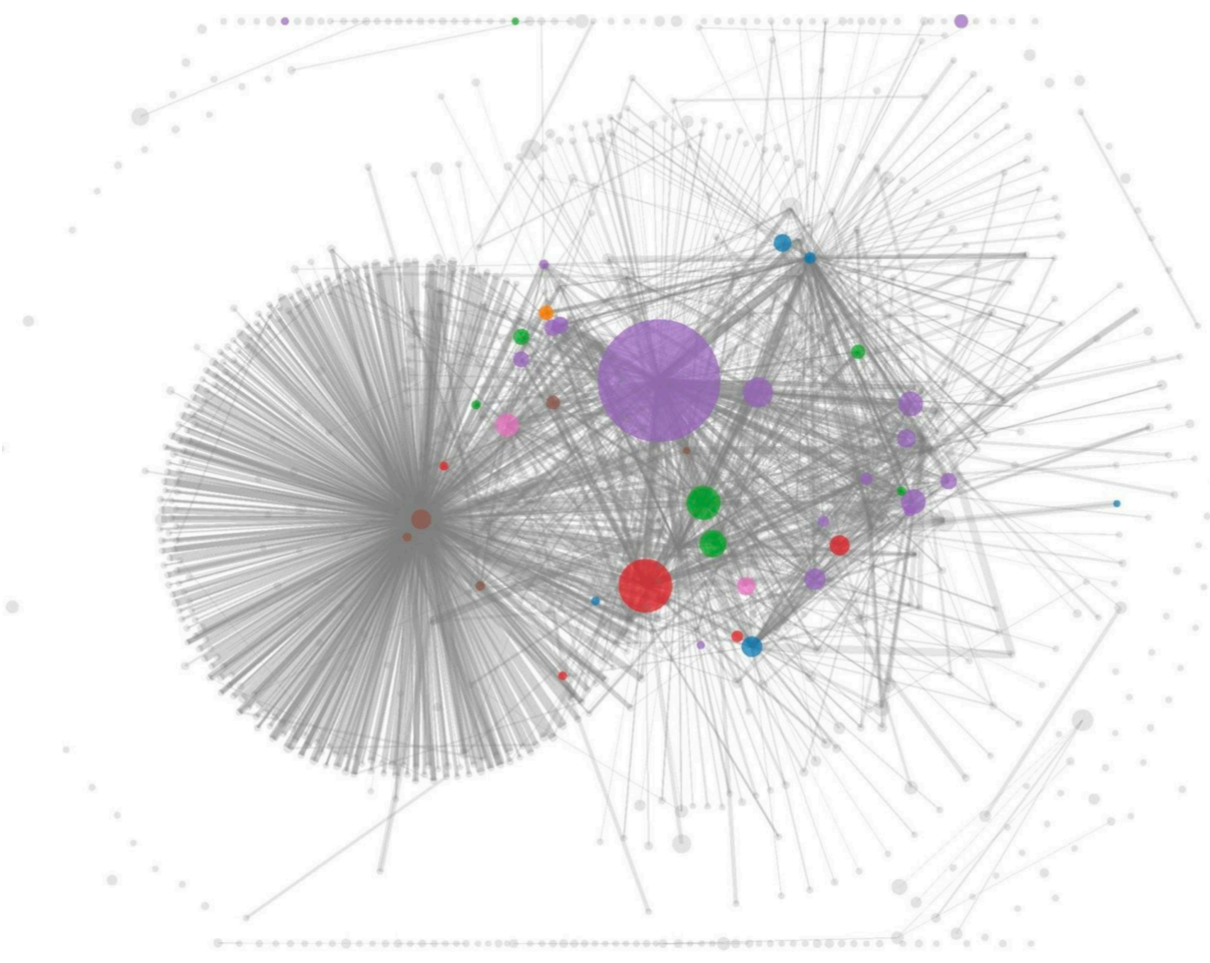
- Alice buys a teapot
  - The price is 8, so needed two unspent outputs
- This is evidence of control over these outputs
- Address can be linked transitively
  - Adversary can be adaptive (provoking transactions, repeating them…)
  - What if the price is 8.5 ? Can you identify the change address (idioms of use)?

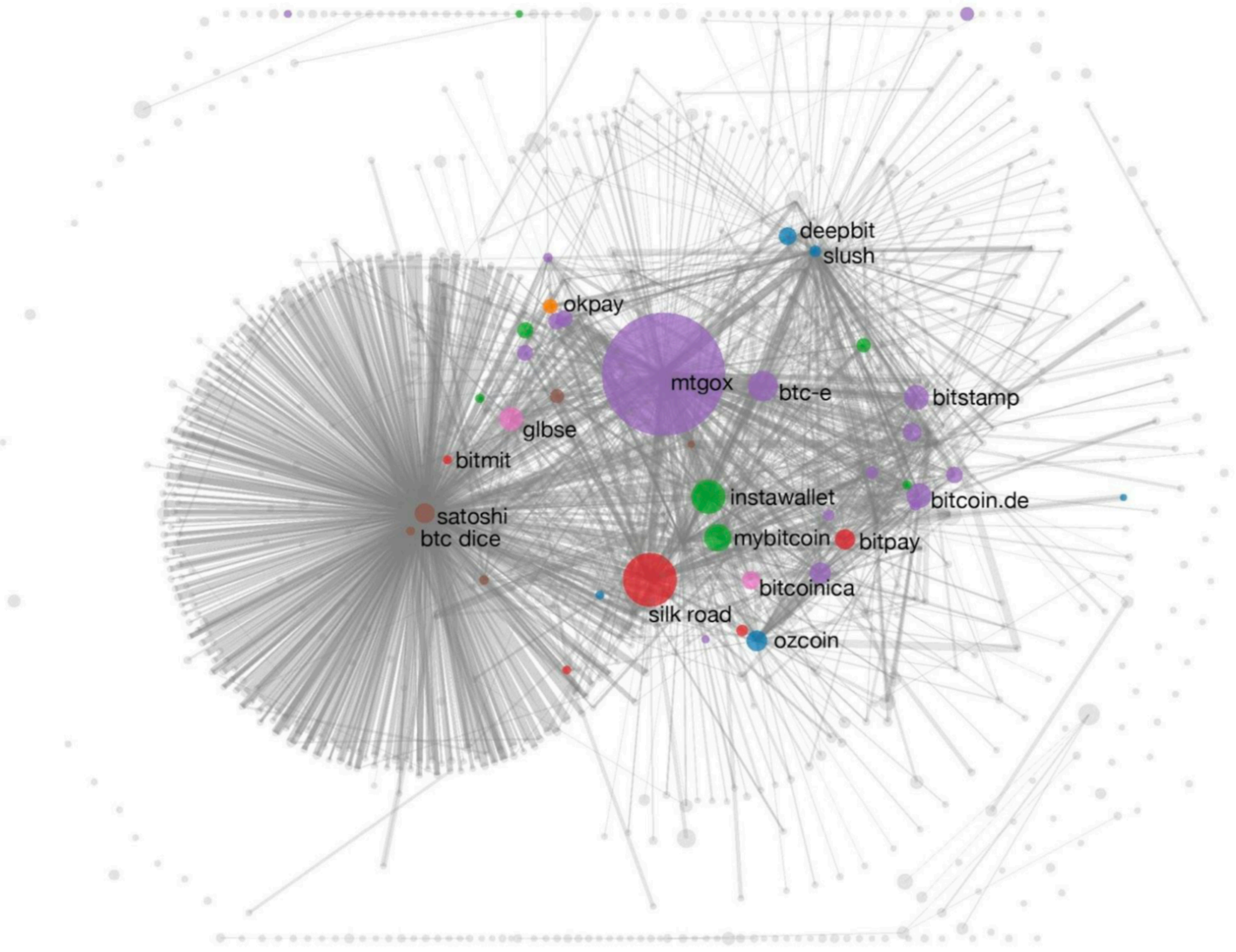




Clustering Addresses

- https://arxiv.org/pdf/1107.4524.pdf
- https://cseweb.ucsd.edu/~smeiklejohn/files/imc13.pdf
    - Real world identities to clusters?





Identifying individuals

- Can we link “little” clusters (of individuals) to real-life identities?

- Directly transacting: a transaction party often knows at least one address belonging to other party(ies)

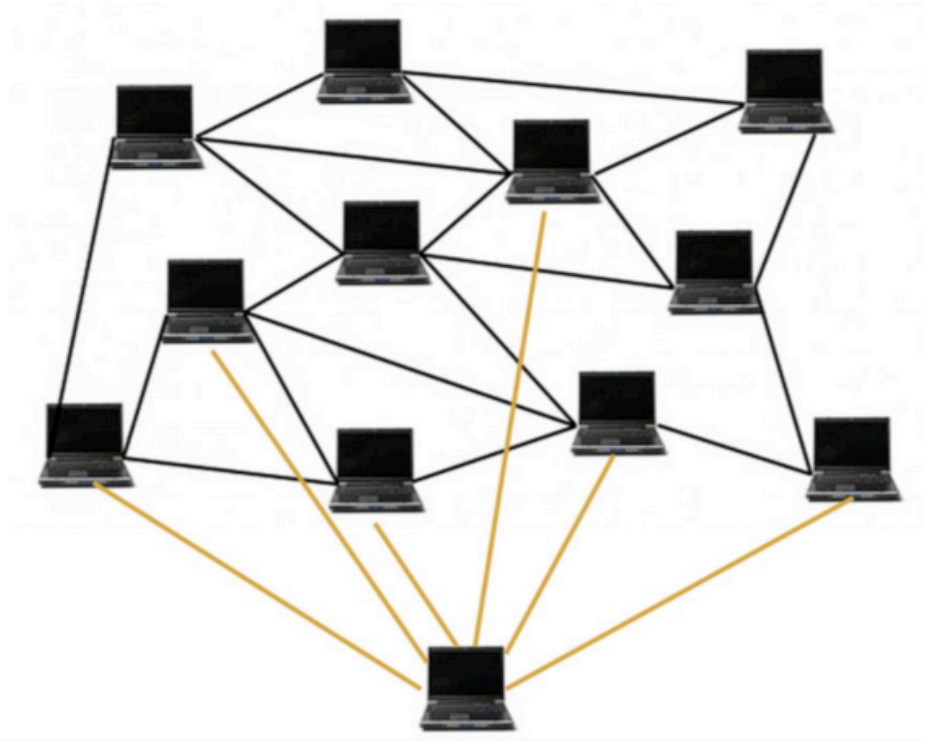
- Via service providers: exchanges or another centralized service provide typically ask users for their identities (Know Your Customer — KYC)

- Carelessness: people often post their Bitcoin addresses in public forums (donations, payments, …)

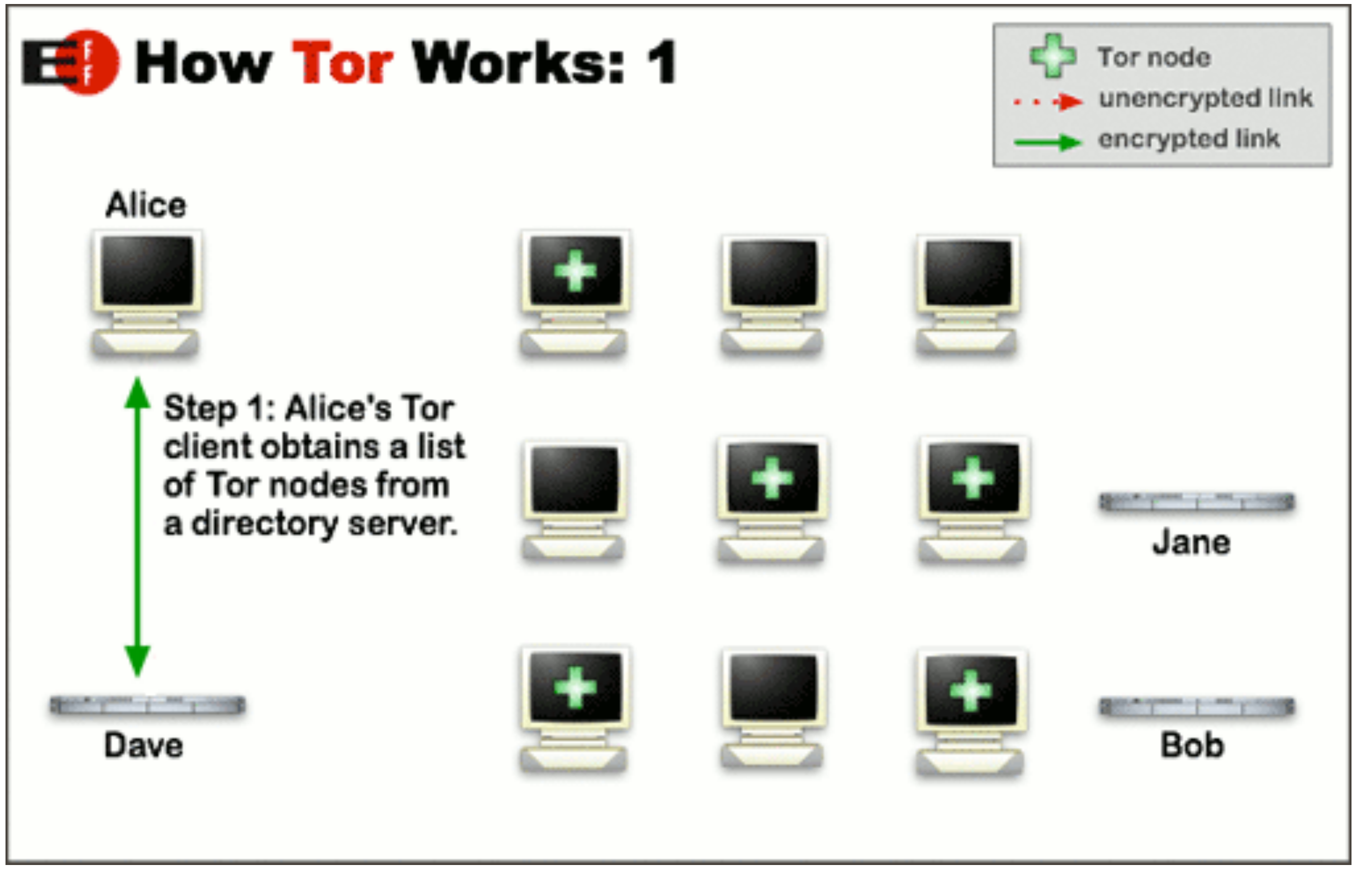
- Things get worse over time: deanonymization algorithms usually improve over time (data is publicly available!)

Network-layer Deanonymization

- Can we use the underlying p2p network?
  - Network identities (IP addresses) can be very precise
- Tracking source
  - Observation: “the first node to inform you of a transaction is probably the source of it.” - D. Kaminsky
- No encryption
- Prevention: network-layer anonymization

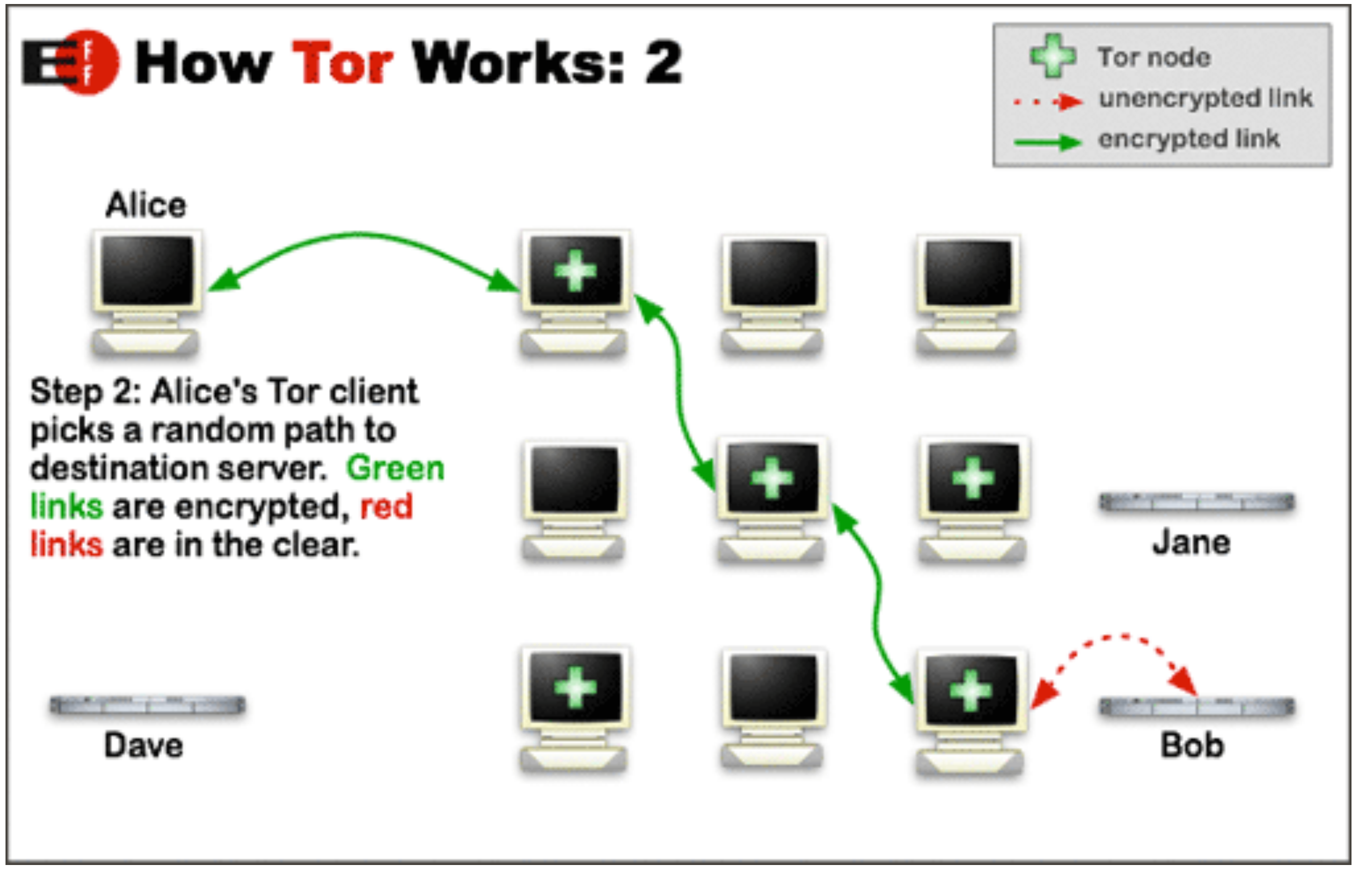


Tor Network



How Tor Works: 1

Step 1: Alice's Tor client obtains a list of Tor nodes from a directory server.



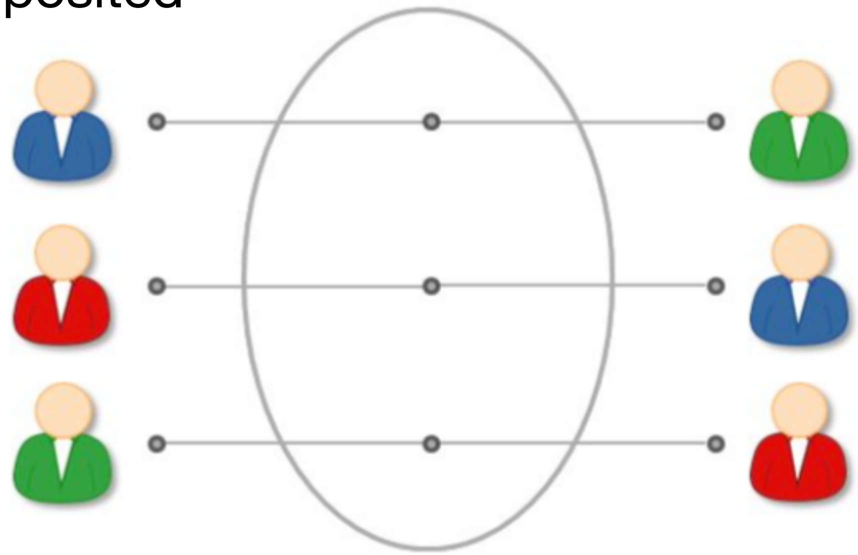
How Tor Works: 2

Step 2: Alice's Tor client picks a random path to destination server. Green links are encrypted, red links are in the clear.

Mixing

Mixing

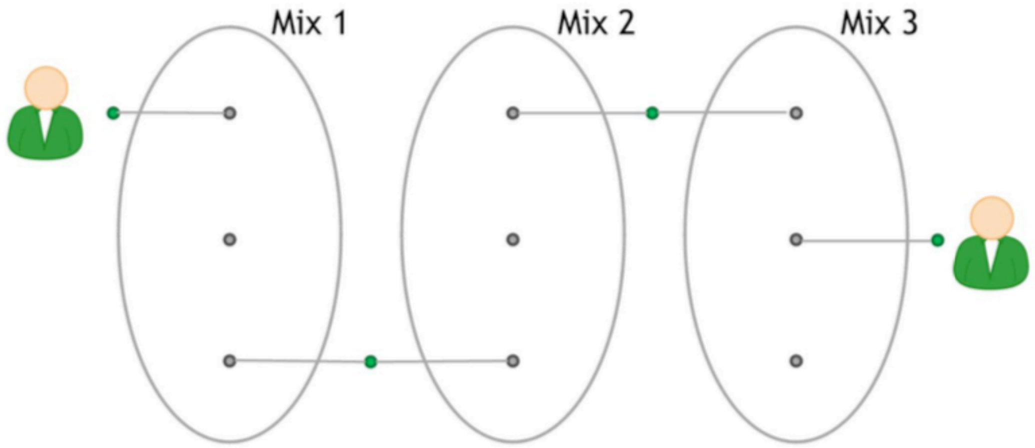
- Tor is a mix-net at the network layer, why not reuse a similar idea on the application (Bitcoin transactions) layer?
  - Use intermediary for mixing
- Users send coins to an intermediary and get back coins that were deposited by other users.
  - Harder to track as it increases anonymity set



- Online wallets or exchanges
  - Deposit coins, withdraw later (not necessary the same coins)
  - Usually not anonymity-specialized (mixing is just a side-effect)
  - Similar to banks
    - KYC, knows in/out TXs, users don't control coins, …

Mixing Services

- Dedicated mixing services
  - promise not to keep records, nor require your identity
  - send coins to a mix address and tell the mix a destination address to send bitcoins to
    - You need trust the mix



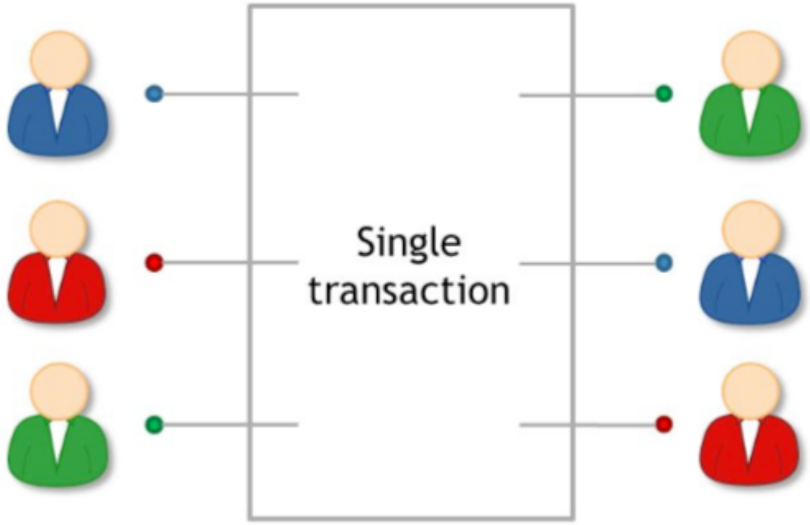
- Series of mixes
  - use a series of mixes, one after the other (similar to Tor)
    - as long as one mix is honest, some level of anonymity can be provided

Mixing Services

- Uniform transactions
    - Mixes should agree on a tx chunk size, as o/w trivial to deanonymize
- Client side should be automated
    - The process should be implemented in a wallet software
- Mixing fees
    - Mixes should be paid, but that could influence chunk size

Decentralized Mixing

- Can we replace mixing services with a peer-to-peer protocol?
  - Users mix by themselves with impossible theft (ideally)
- CoinJoin (high level)
  - Users jointly create a single Bitcoin transaction that combines all of their inputs and spends it to some outputs
    - Inputs/outputs order is randomized
  - They can create signatures independently
  - Adversary cannot determine the in-out mapping
  - Can do multiple such rounds, chunk size important too



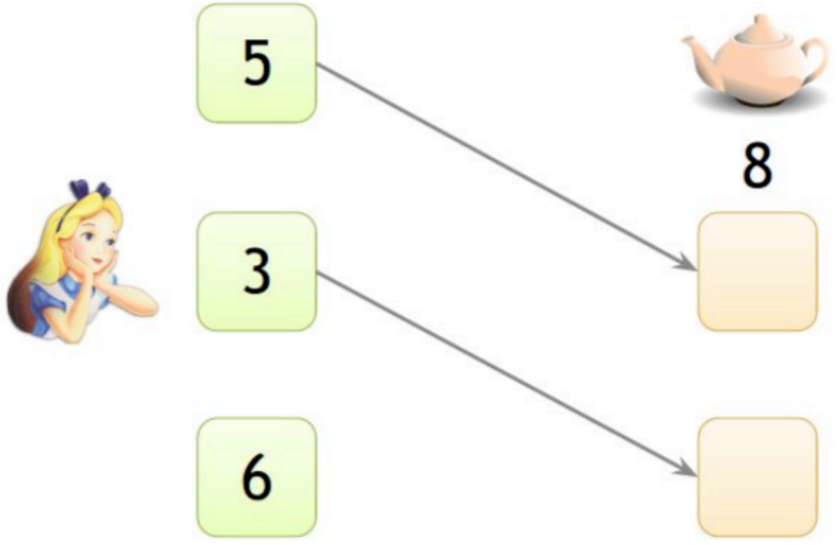
CoinJoin

1. Finding peers
   - A peer discovery service (only helps to group peers and cannot steal coins)
2. Exchanging addresses
   - Peers exchange anonymously (e.g., Tor) input and output addresses with each other
3. One peer creates a transaction including all inputs and outputs and passes it around
4. Each peer verifies its inputs and outputs, and signs it
5. The transaction is finally broadcast

- DoS: a malicious peer can refuse to sign or spend its input.
  - Prevented with PoW or Proof-of-Burn.

High-level flows

- Some scenarios are challenging to anonymize due to patterns and timing
  - Periodic payments
  - Very specific payment amounts



- Merge avoidance
  - The receiver of a payment to provide multiple output addresses
  - The sender and receiver agree upon a set of denominations to break up the payment into multiple transactions
  - The receiver should not recombine these payments
- Need to rely on receivers, some patterns can be leaked. Can we do better?

Zerocoin and Zerocash

Zerocoin & Zerocash

- http://zerocoin.org/media/pdf/ZerocoinOakland.pdf
- http://zerocash-project.org/media/pdf/zerocash-oakland2014.pdf
- Goal: incorporate anonymity at the protocol level
- Cryptographic guarantees
    - Much stronger than mixing etc…
- Incompatible with Bitcoin

Zerocoin

- Basecoin: Bitcoin-like altcoin
    - Zercoin is an extension of Basecoin
- Basecoins can be converted into zerocoins and back
    - That breaks link between original and new basecoin
- Basecoin is the currency for making transactions
- Zerocoin provides unlinkability for these transactions

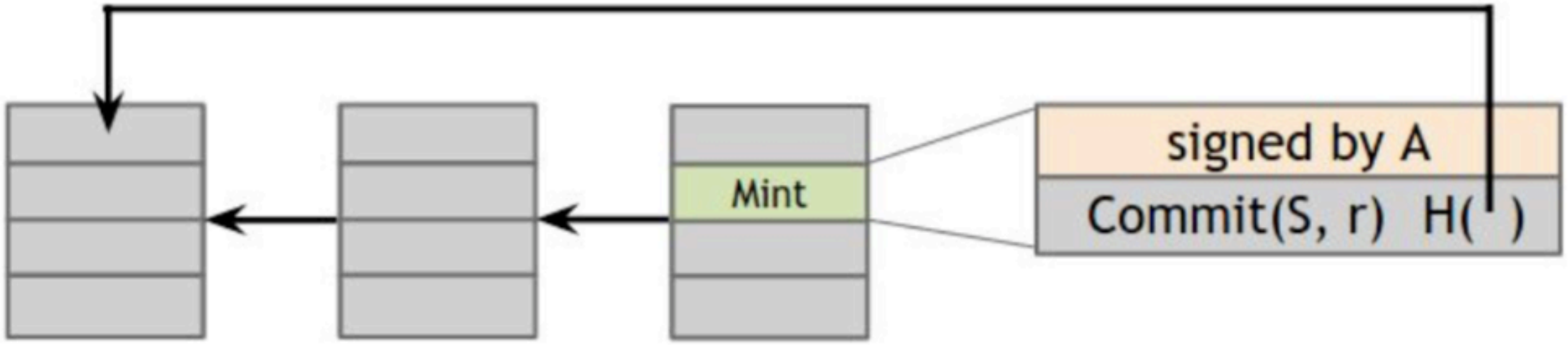
Zerocoin

- A Zerocoin is a cryptographic proof that you owned a Basecoin and made it unspendable
- Miners can verify these proofs
- How?
- Zero-knowledge (zk) proofs: proof on a statement w/o revealing any other information that leads to that statement being true
  - Examples:
    - I known x s.t. H(x) = f730ae…
    - I know x s.t. H(x) is in {0349e…, 330ea…, f850ed…, …}

Minting zerocoins

Anyone can mint a zerocoin (let's assume 1 basecoin denomination)

1. Generate serial number S and a random secret r
2. Compute Commit(S,r), the commitment to the serial number
3. Publish the commitment onto the block chain. This burns a basecoin, making it unspendable, and creates a Zerocoin. Keep S and r secret for now.



Spending zerocoins

At any point, there will be many commitments on the blockchain $c_1, c_2, \ldots, c_n$

1. Create a special “spend” transaction that contains S, along with a zk proof of the statement:

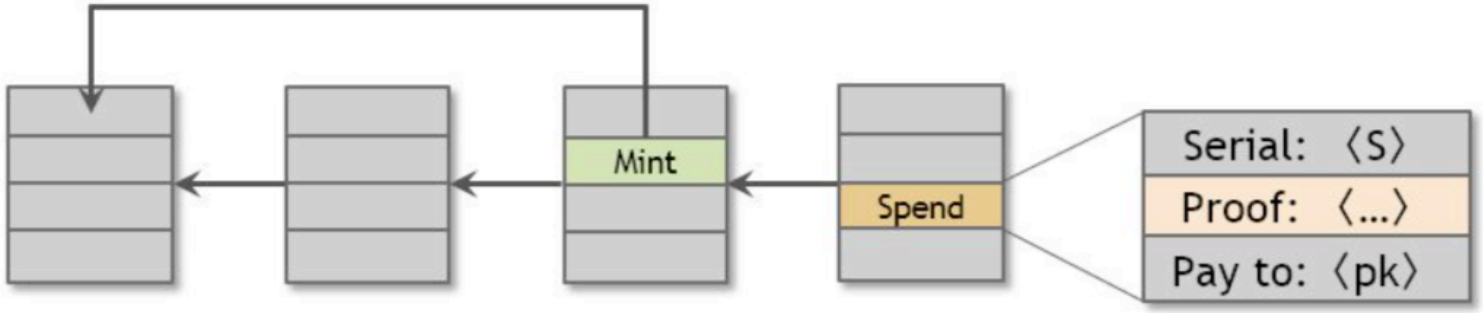
   “I know r such that Commit(S, r) is in the set $\{c_1, c_2, \ldots, c_n\}$”.

2. Miners will verify the zk proof which establishes an ability to open one of the zerocoin commitments on the blockchain, without actually opening it

3. Miners will also check that the serial number S has never been used in any previous spend transaction (since that would be a double-spend)

4. The output of the spend transaction will now act as a new basecoin. For the output address, you should use an address that you own.

- Once a zerocoin is spent, the serial number becomes public, and no one will ever be able to redeem this serial number again



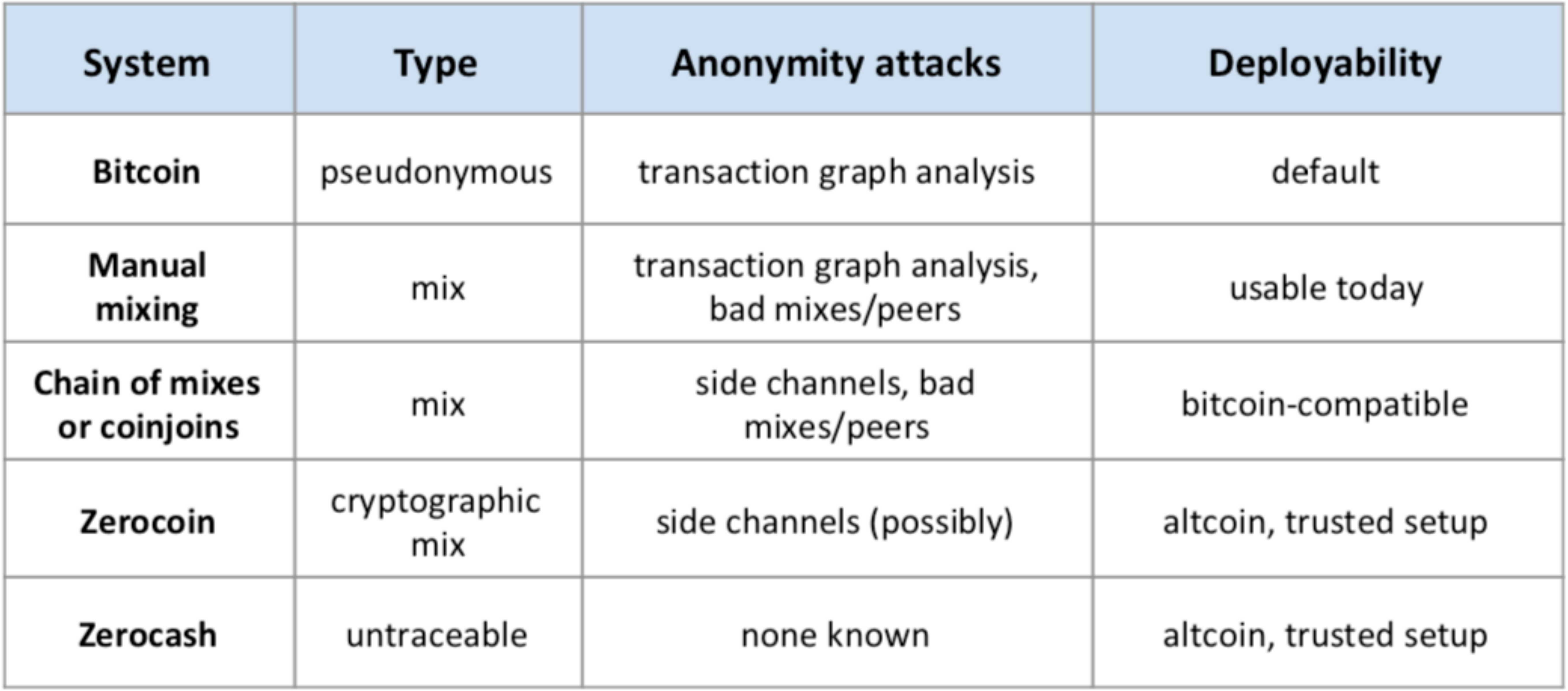
Zerocoin Properties

- Anonymity
  - r is kept secret, nobody knows which serial number corresponds to which zerocoin
  - No link between mint and spend transactions
- Efficiency
  - Size of zk proofs is logarithmic in n (but still ~50kB)
- Trusted setup
  - Some security-critical parameters have to be generated prior deployment (can be removed after the setup)

Zerocash

- zkSNARKS
  - more compact and faster to verify zk proofs
- No need of Basecoin
- Hidden transaction amounts and splitting/merging possible
- Blockchain contains only existence of transactions
  - Neither addresses nor values are revealed (sender and receiver only know it + miners know tx fees)
- Trusted setup

Summary



System	Type	Anonymity attacks	Deployability
Bitcoin	pseudonymous	transaction graph analysis	default
Manual mixing	mix	transaction graph analysis, bad mixes/peers	usable today
Chain of mixes or coinjoins	mix	side channels, bad mixes/peers	bitcoin-compatible
Zerocoin	cryptographic mix	side channels (possibly)	altcoin, trusted setup
Zerocash	untraceable	none known	altcoin, trusted setup

Reading

- Textbook 6
  - (most of the figures is from the textbook)
- … and inline references