Privacy in Blockchains: Course Summary

Giulia Fanti

Foundations of Privacy

Fall 2019

Some slides from Joe Bonneau, Eran Tromer

Administrative

- All HW4 submissions should be in by midnight tonight!

- Project presentations start next Monday, Dec. 2
  - You MUST upload your slides to Canvas by **9 am EST** on the morning of your presentation (M or W)
  - 10 minutes each!

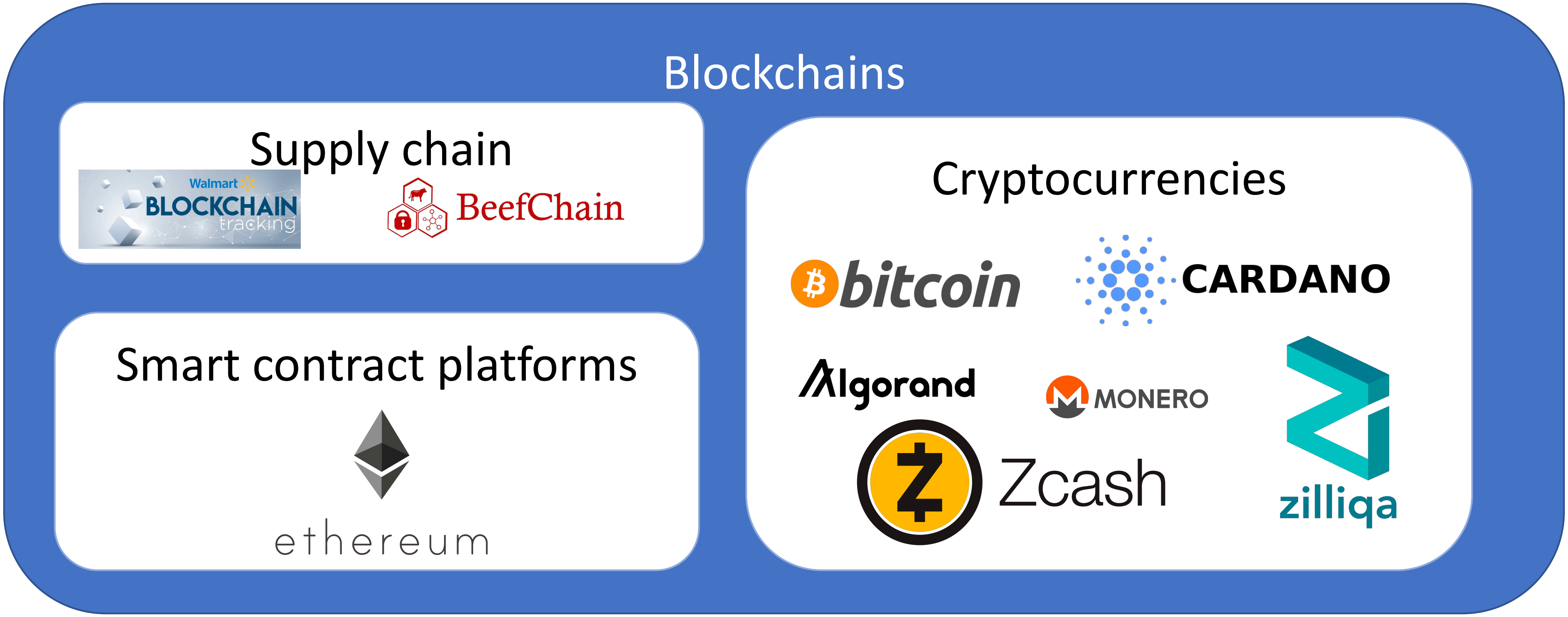
- Presentations will be in the same category as mid-semester presentations
  - Overall grade will be the average of your two presentations

- Rubric for presentations and writeup are on Canvas/Gradescope

Next: Blockchains

- How do they work?
- What privacy problems can arise?
- How can the tools from class be used to fix these problems?

Blockchains



Blockchains

Supply chain

Smart contract platforms

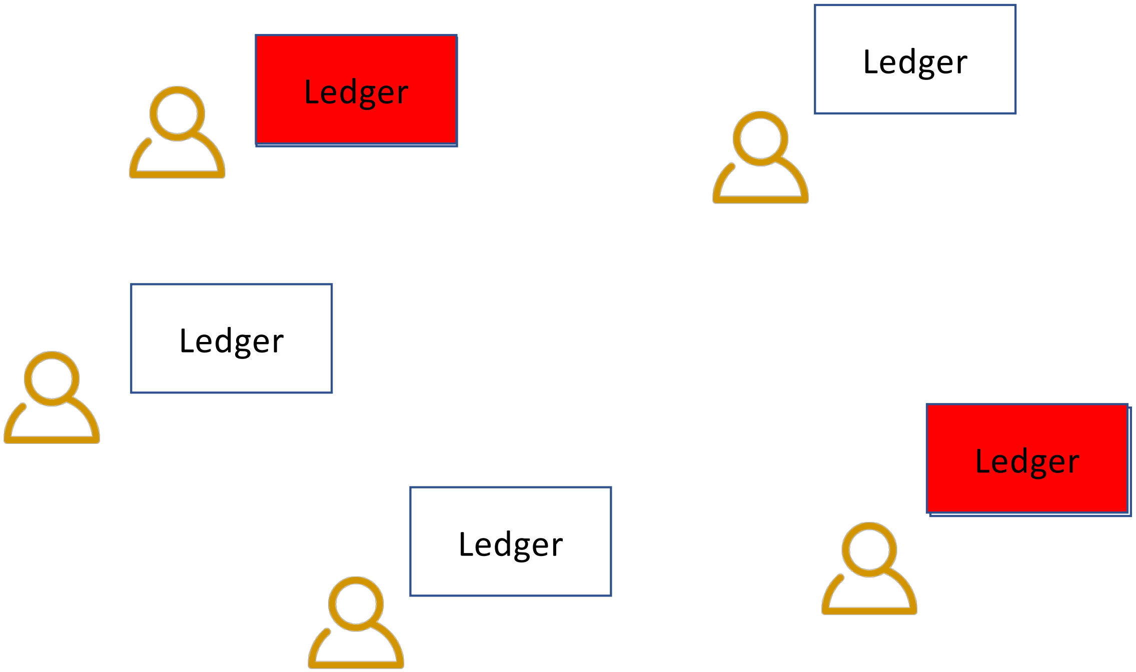
Cryptocurrencies

Core Problem

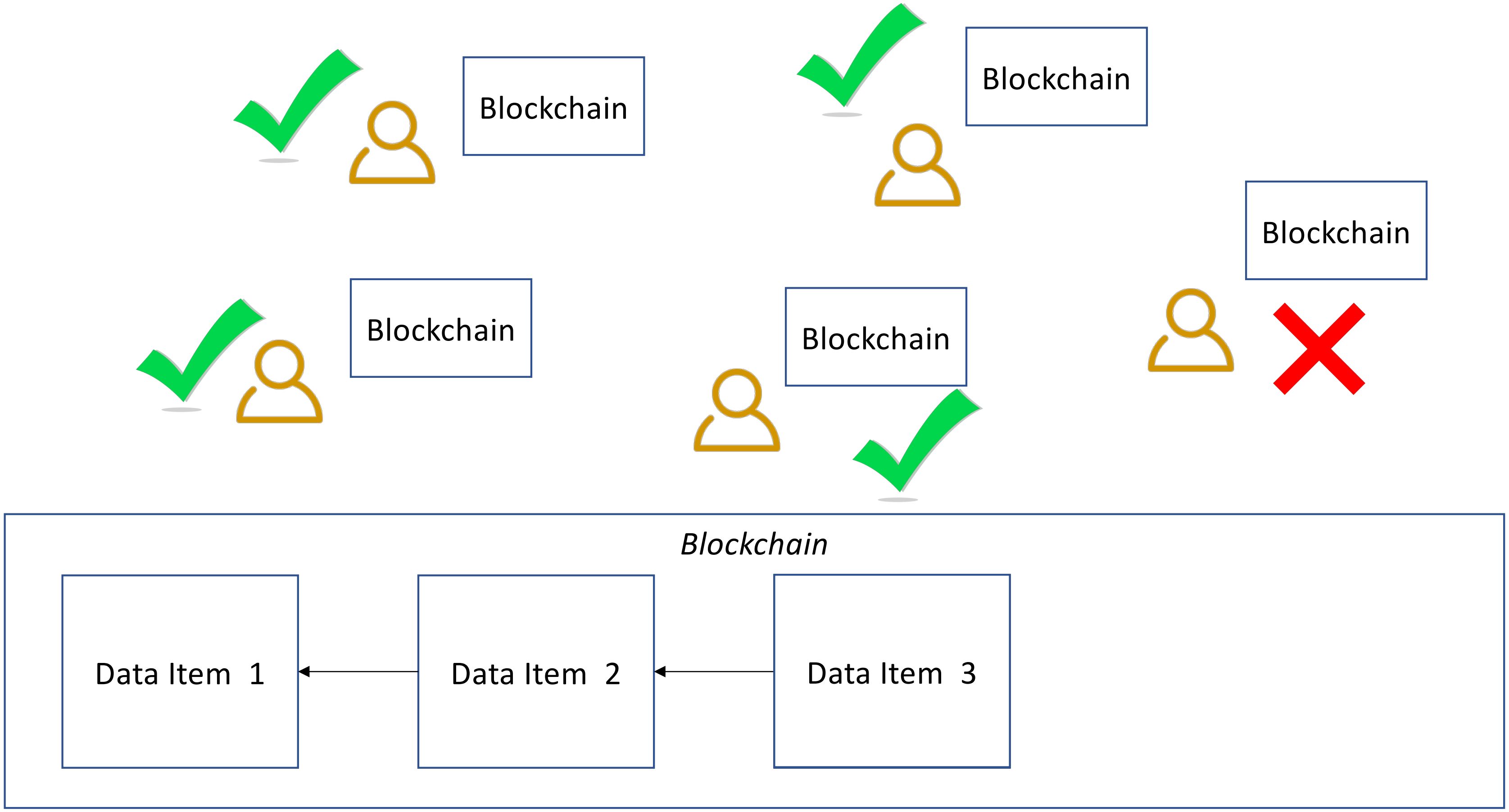
How do we maintain data…

… among many nodes…

…when some of them could be corrupt?



The basic approach



Blockchain

Blockchain

Blockchain

Blockchain

Blockchain

Blockchain

Data Item 1

Data Item 2

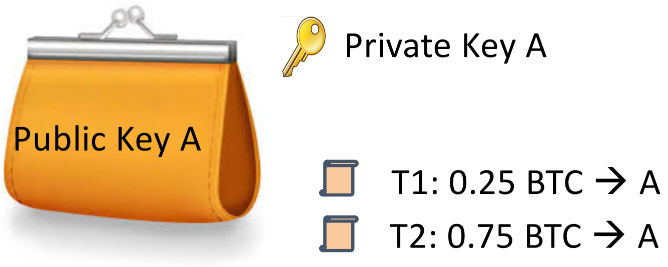
Data Item 3

Bitcoin primer (1/2)

- A peer-to-peer digital payment system
- Completely decentralized digital currency
  - **No central mint** to produce currency
  - **No central bank** to verify transactions
  - Once confirmed, transactions are **irreversible**
  - Predictable, capped, currency supply
- Key innovation in Bitcoin: coin production and verification is done by **network consensus**

Bitcoin primer (2/2)

- **No notion of a "coin"**



Public Key A

Private Key A

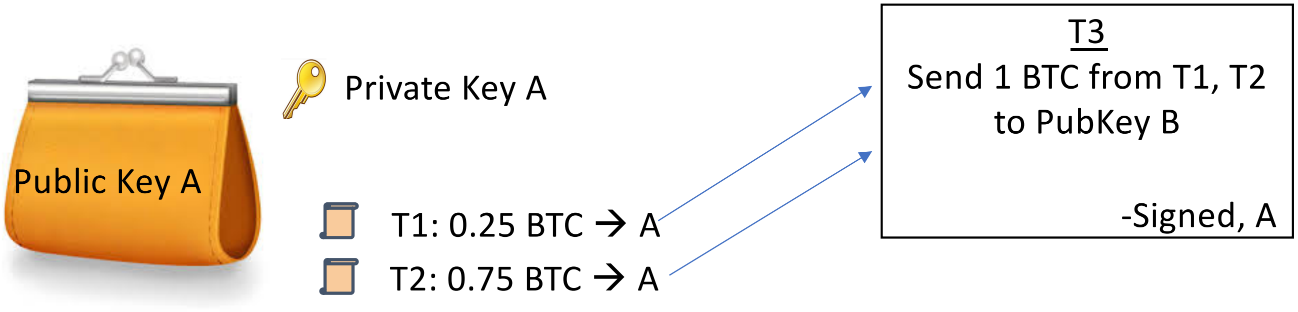
T1: 0.25 BTC → A

T2: 0.75 BTC → A

- **Wallets** are addressed by public keys.
- Owner of wallet knows the private key.
- **Transactions** are at the heart of the protocol.
  - Value of a wallet = the bitcoins transferred to it in the past.

Bitcoin transactions

- Alice wants to send 1 BTC to Bob
    - She picks a transaction (or a group of transactions) that she has previously been the recipient of and that cumulatively contain at least 1 BTC
    - She then appends Bob's wallet address to the transaction and digitally signs it
- When Bob subsequently wants to spend the 1 BTC, all he has to do is to repeat the operation



Public Key A

Private Key A

T1: 0.25 BTC → A

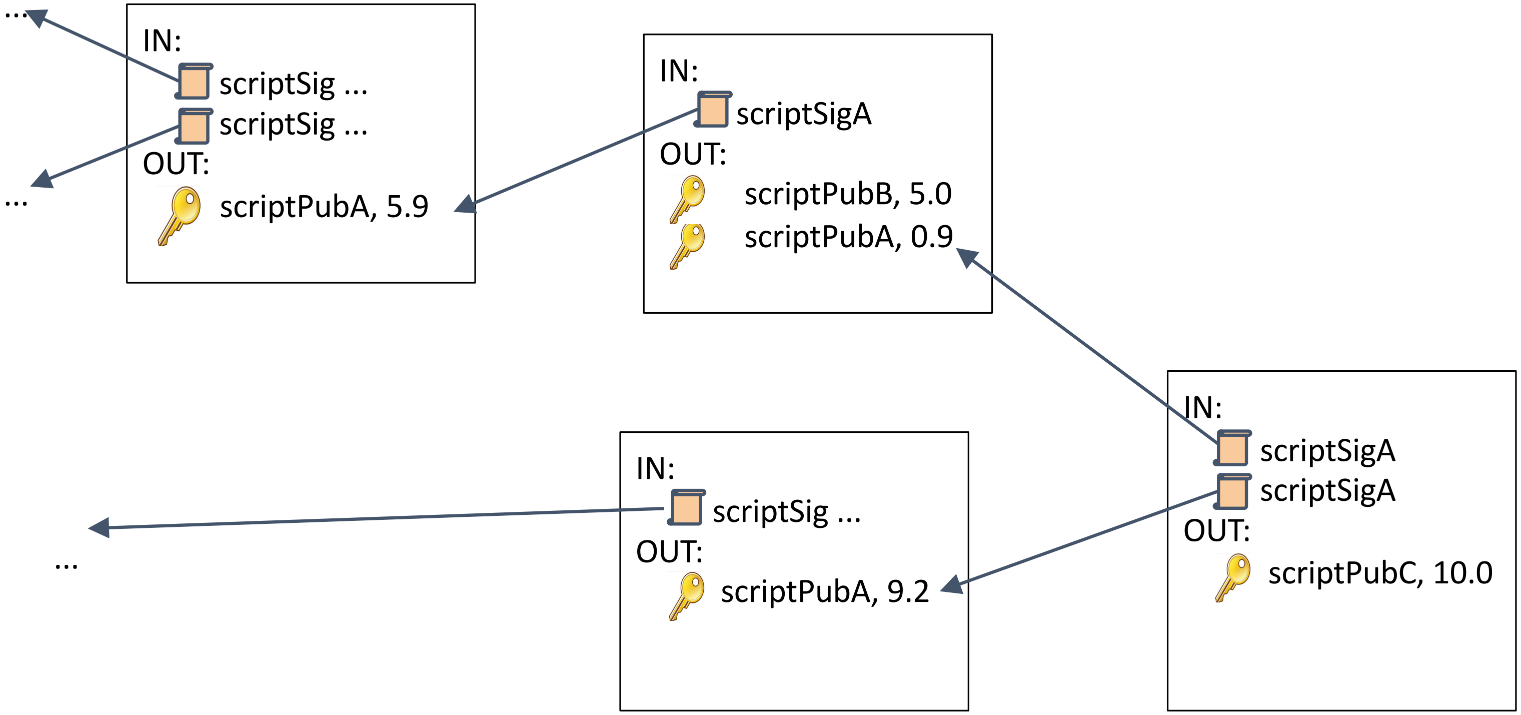
T2: 0.75 BTC → A

T3

Send 1 BTC from T1, T2
to PubKey B

-Signed, A

Bitcoin is *transaction-based*



Slide credit: Joe Bonneau

Preventing double-spending

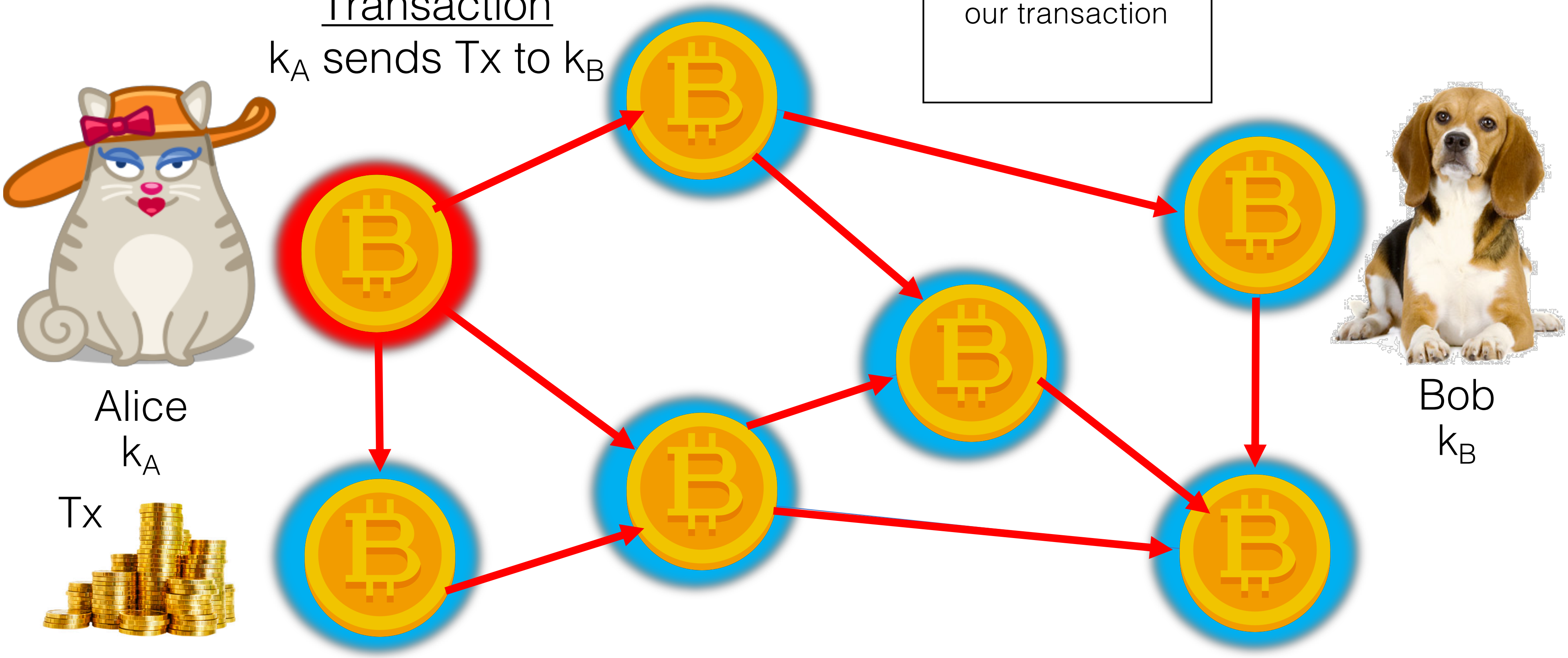
- Bob now has 1 BTC
  - He wants to send it to Charlie…
  - … while keeping it for himself at the same time

- To prevent this Bob (and Alice before him) **has to broadcast** the transaction to everybody in the Bitcoin network

- Then other peers can verify that the transaction is not a double-spend

- Once this is done, the transaction is **embedded forever in a public ledger**

Bitcoin Primer



Transaction

k_A sends Tx to k_B

Blockchain

sd93fjj2

pckrn29

…

our transaction

Alice

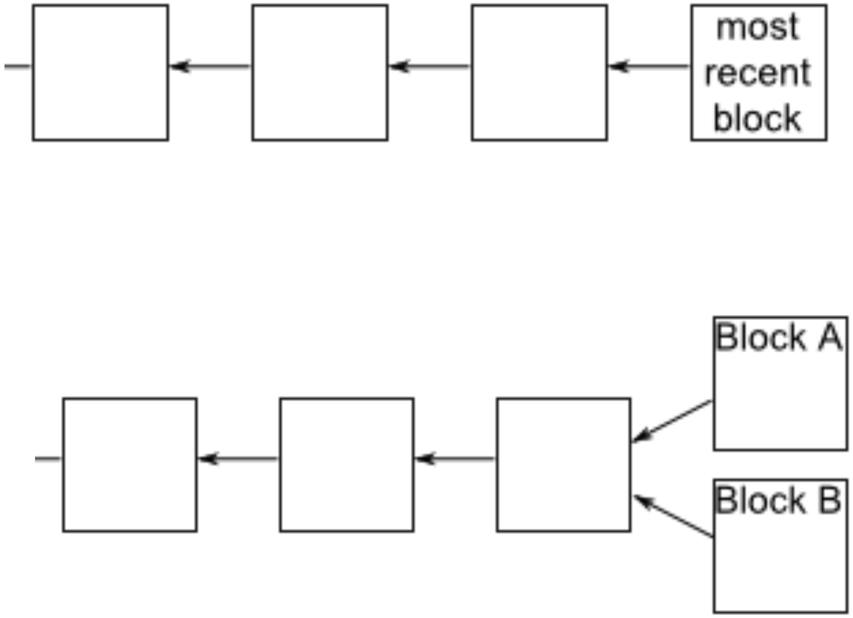
k_A

Tx

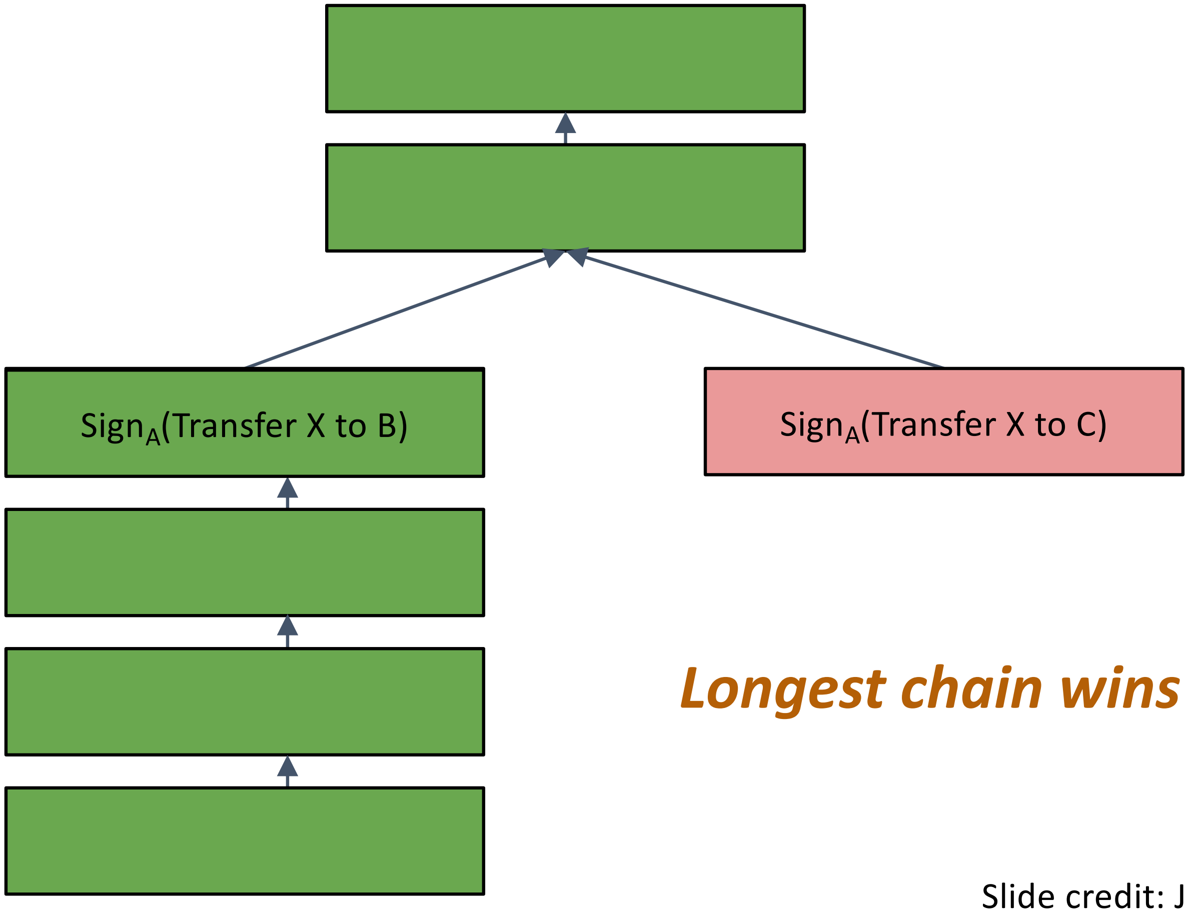
Bob

k_B

The Blockchain, or Ledger



Preventing double spending

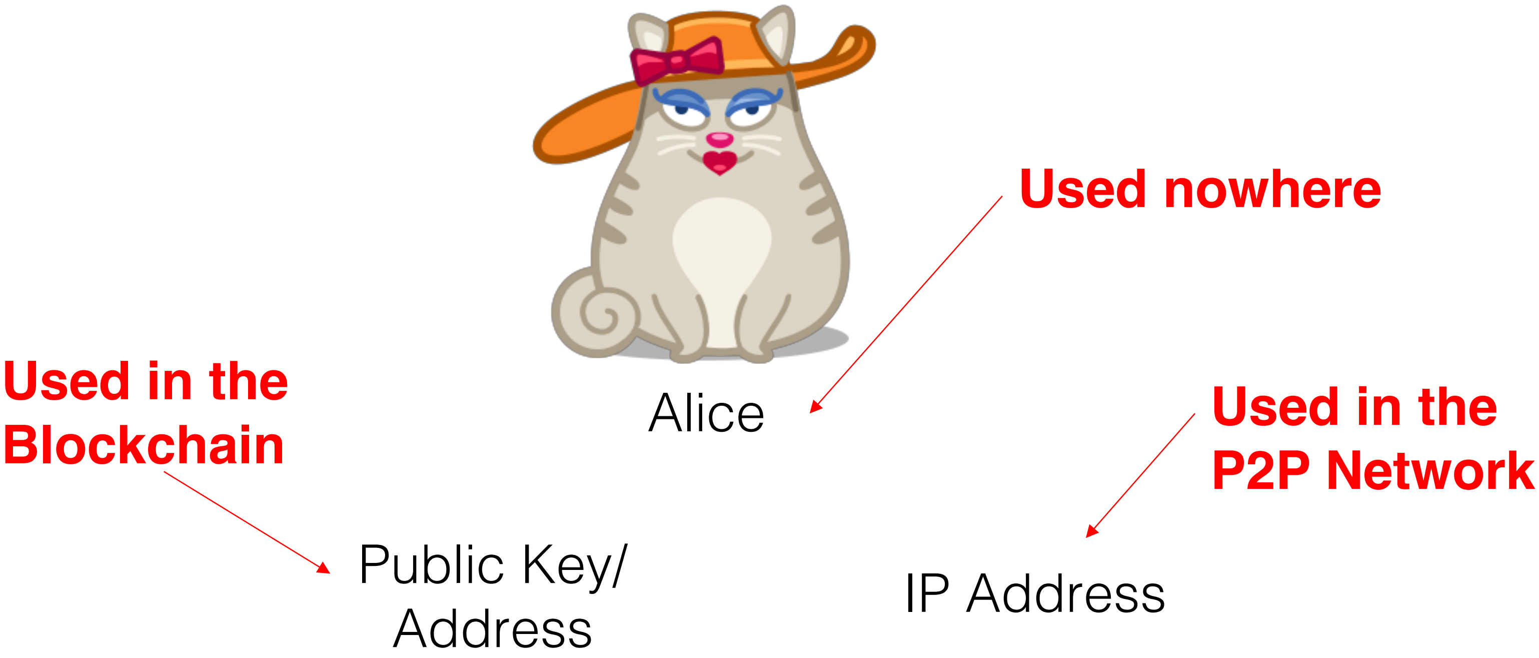

Sign_A(Transfer X to B)
Sign_A(Transfer X to C)

Longest chain wins

Slide credit: Joe Bonneau

What privacy problems can arise?

Multiple Identities



Alice

Used nowhere

Used in the Blockchain

Public Key/ Address

Used in the P2P Network

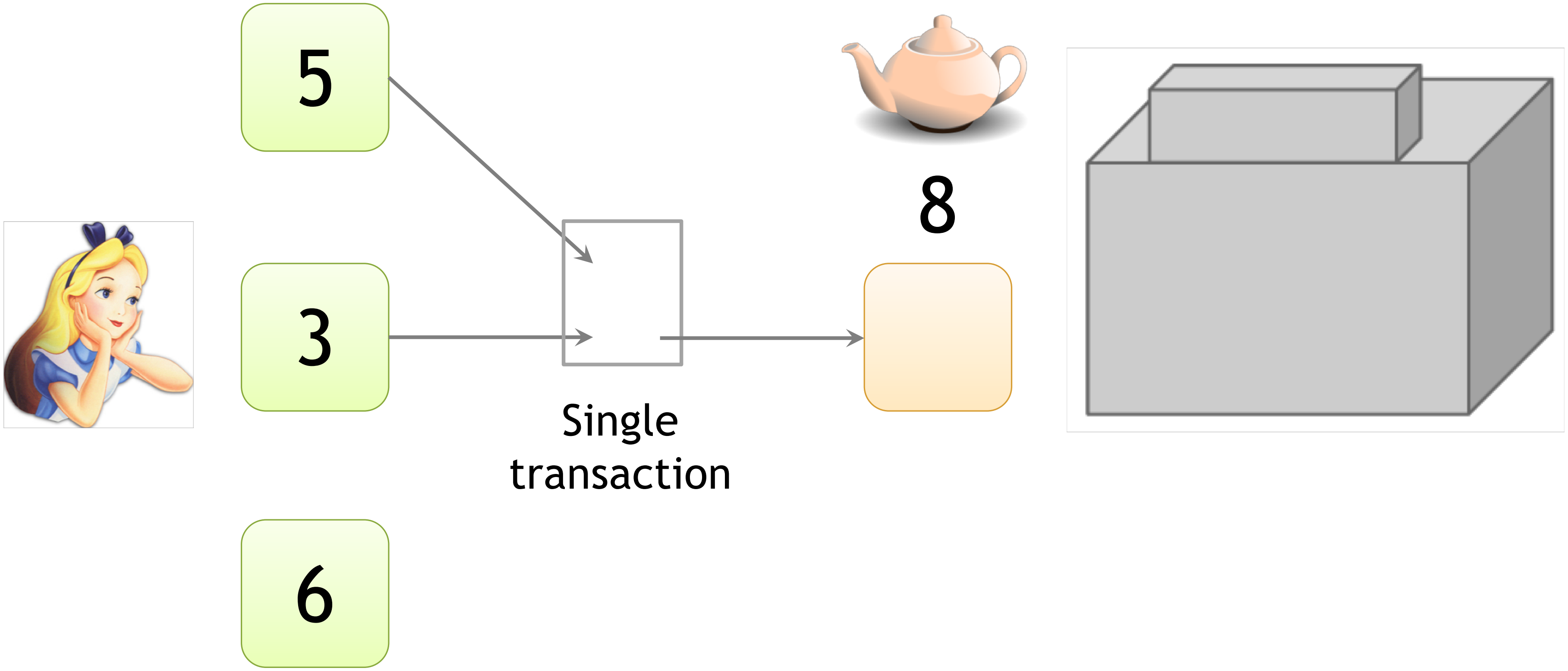
IP Address

Trivial to create new address

Best practice: always receive at fresh address

So, unlinkable?

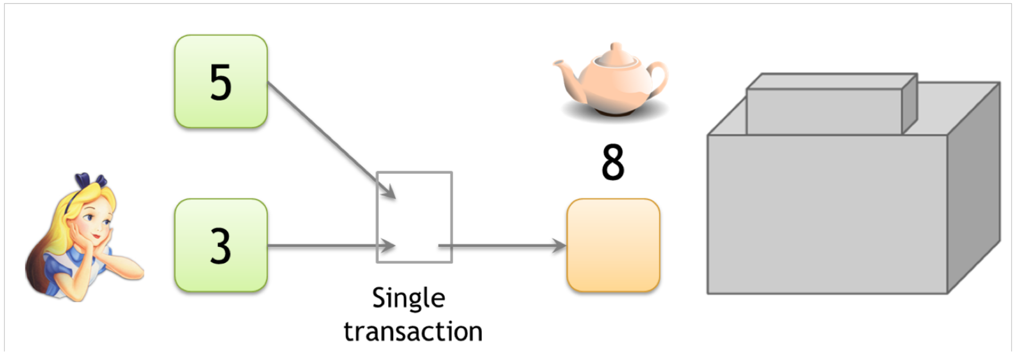
Alice buys a teapot at Big box store



Single transaction

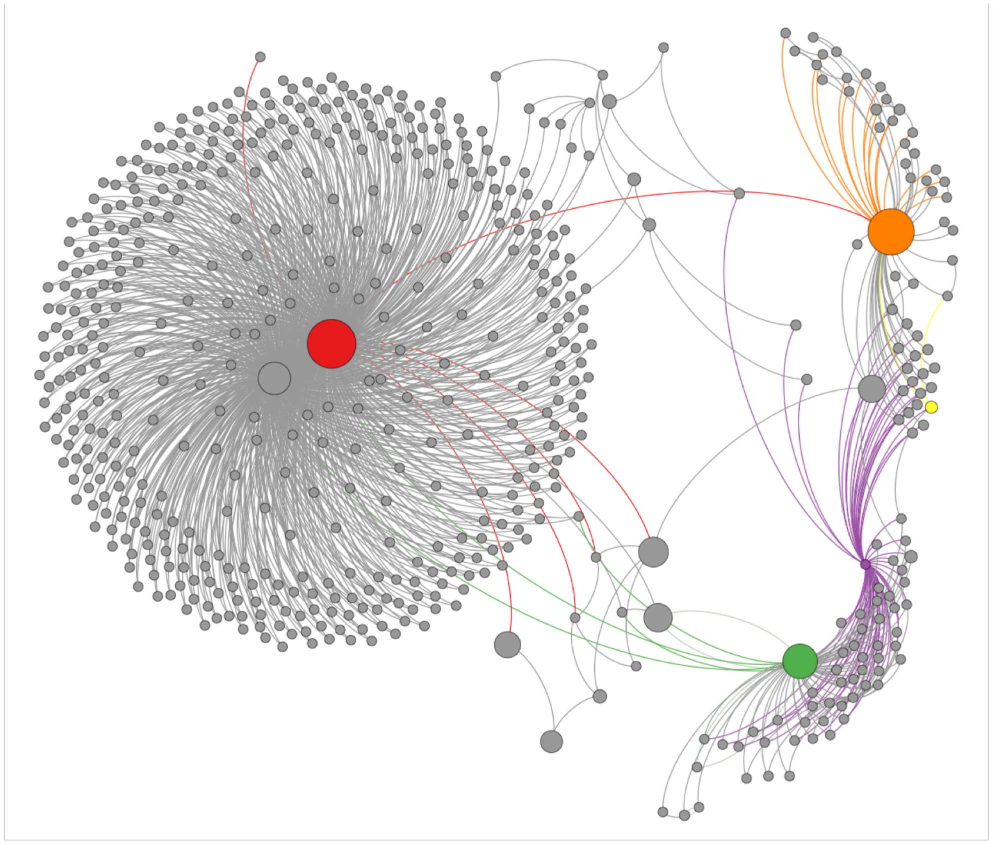
Linking addresses

Shared spending is evidence of joint control



Addresses can be linked transitively

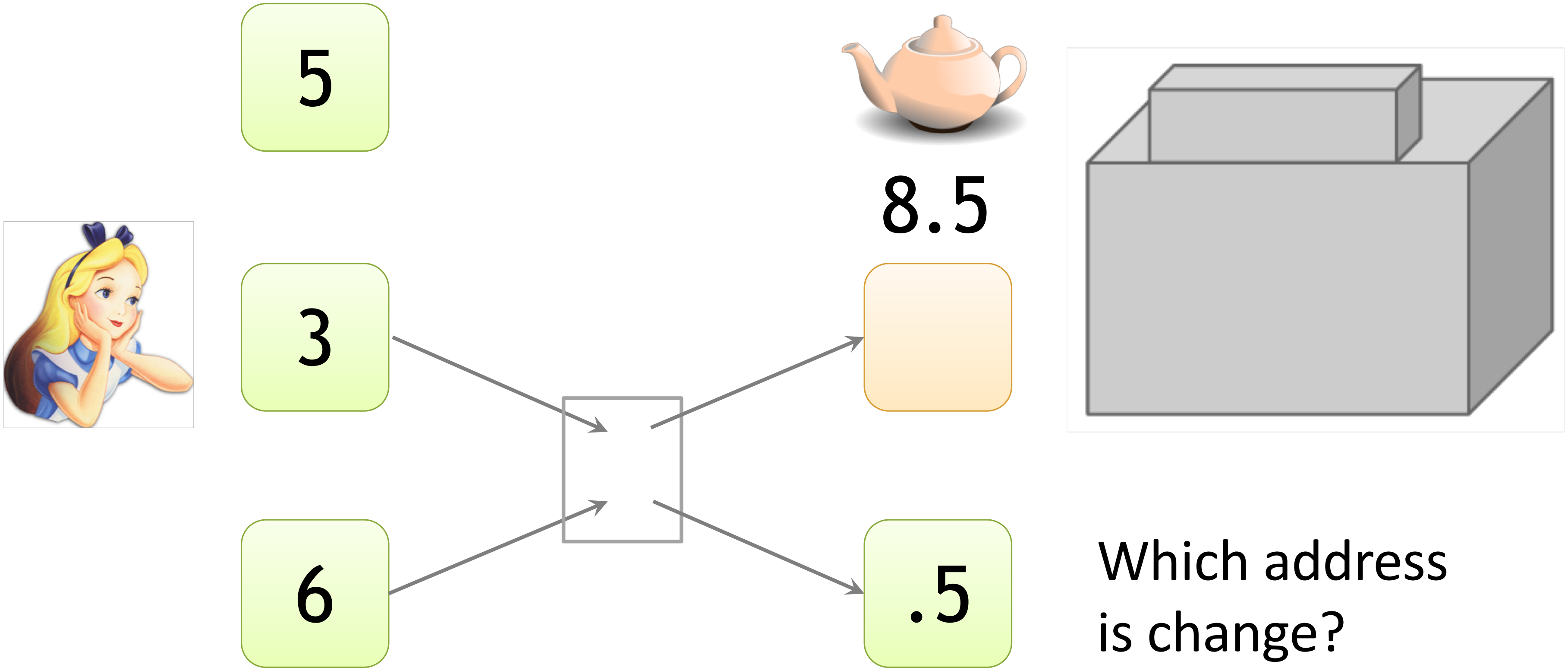
Clustering of addresses



An Analysis of Anonymity in the Bitcoin System

F. Reid and M. Harrigan
PASSAT 2011

Change addresses

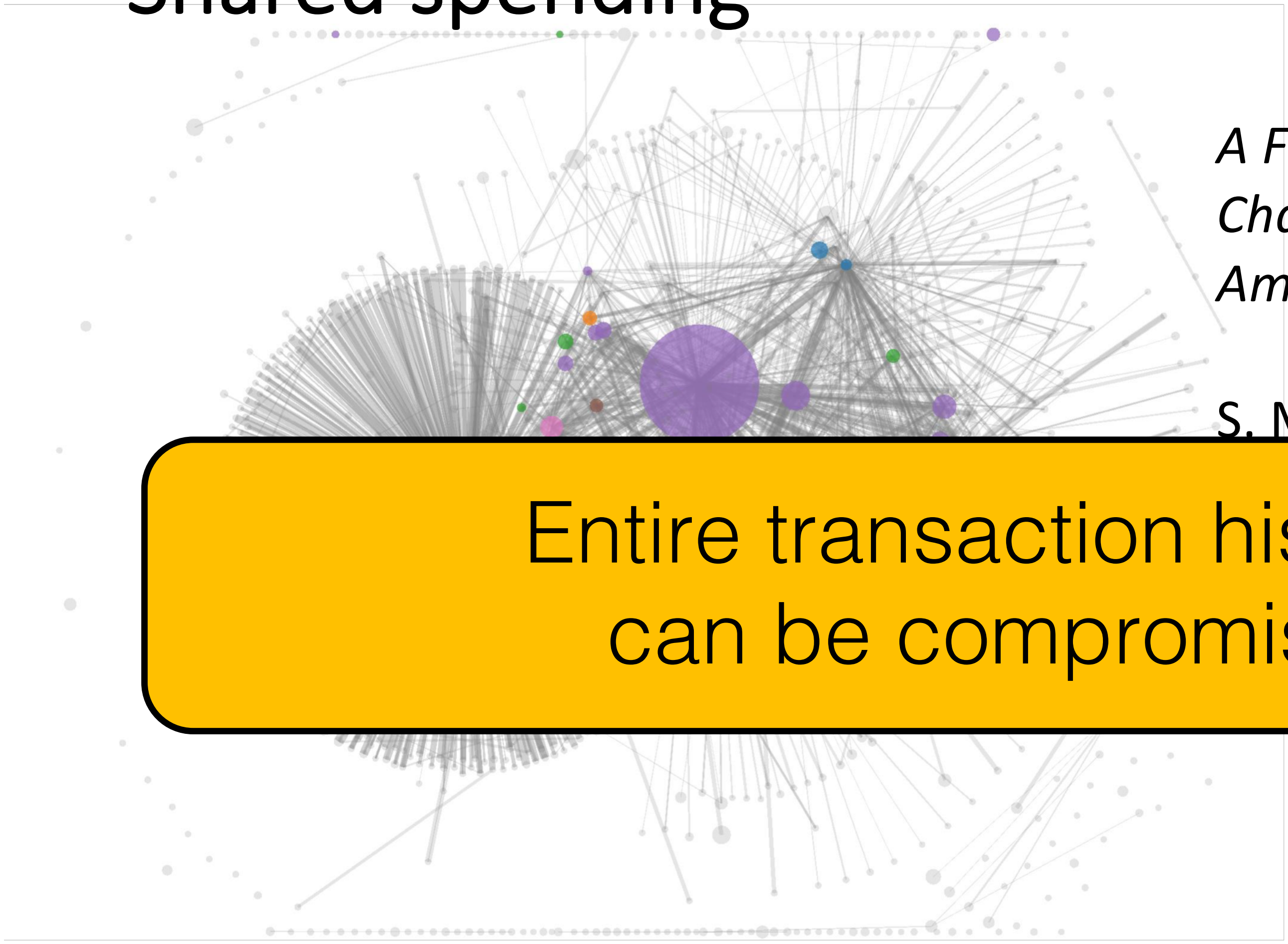


Which address
is change?

“Idioms of use”

- Implementation details or Idiosyncratic features of wallet software that can be used to design heuristics for identifying (linking) change addresses
- For example, most wallets generate a fresh address (never appeared on the Blockchain) when a change address is required
- Bitcoin-Qt library bug → change output always first output in a transaction

Shared spending

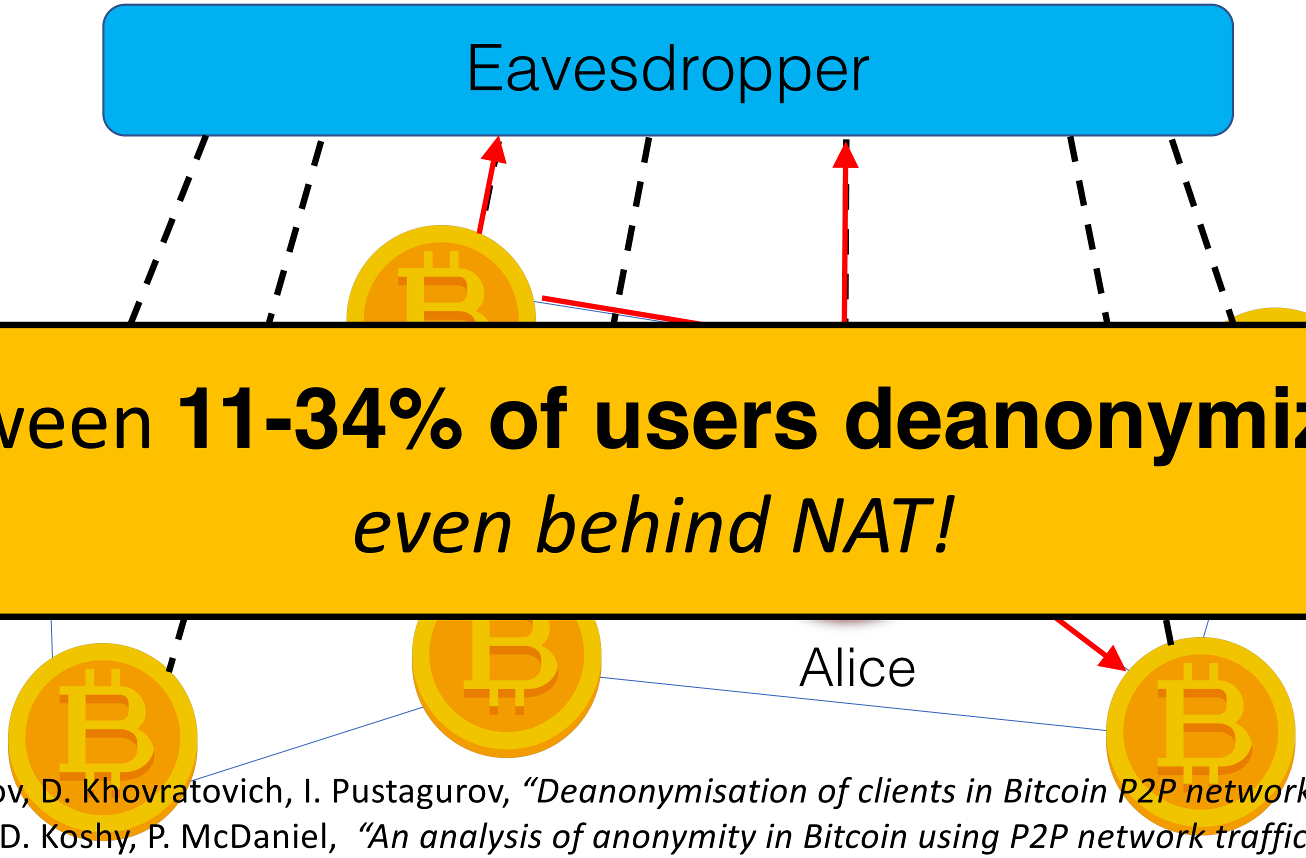


A Fistful of Bitcoins: Characterizing Payments Among Men with No Names

S. Meiklejohn et al.

Entire transaction histories
can be compromised.

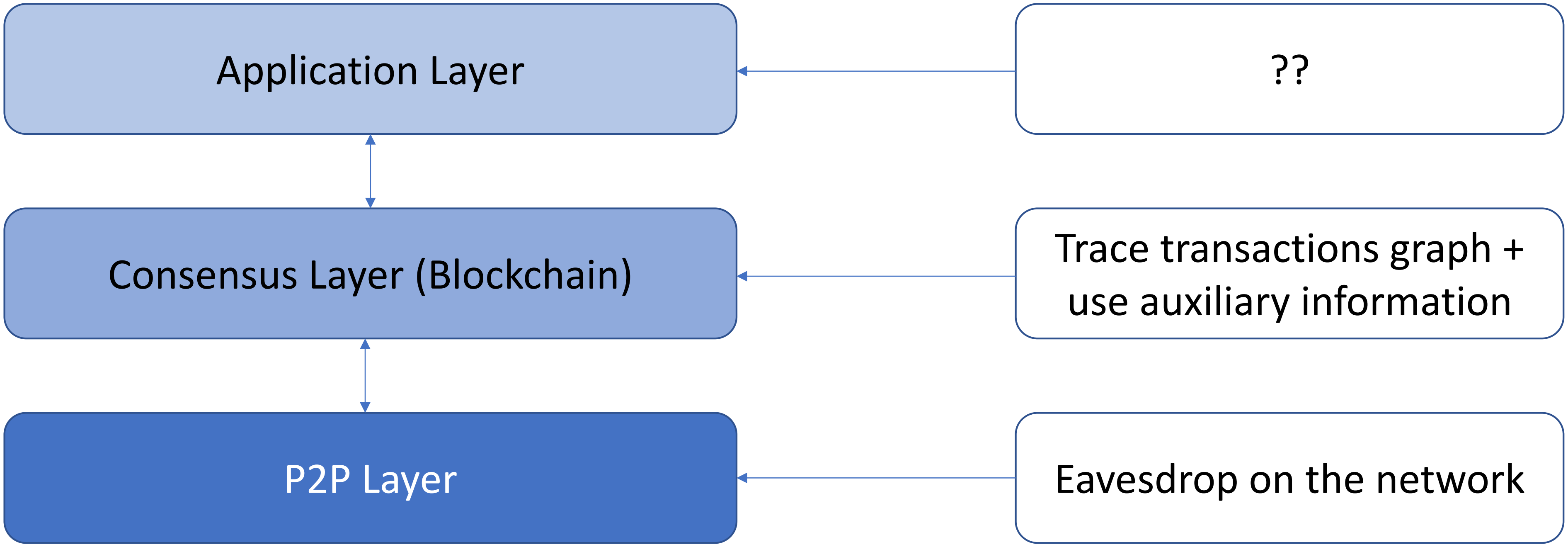
How do we link public keys to IP addresses?



Between **11-34% of users deanonymized,**
even behind NAT!

A. Biryukov, D. Khovratovich, I. Pustagurov, *"Deanonymisation of clients in Bitcoin P2P network"*, CCS 2014
P. Koshy, D. Koshy, P. McDaniel, *"An analysis of anonymity in Bitcoin using P2P network traffic"*, Financial Crypto 201

How to deanonymize blockchain transactions



Application Layer

Consensus Layer (Blockchain)

P2P Layer

??

Trace transactions graph + use auxiliary information

Eavesdrop on the network

How can we fix these issues?

Network Layer

- Which of the techniques we have learned about would protect against a network adversary?

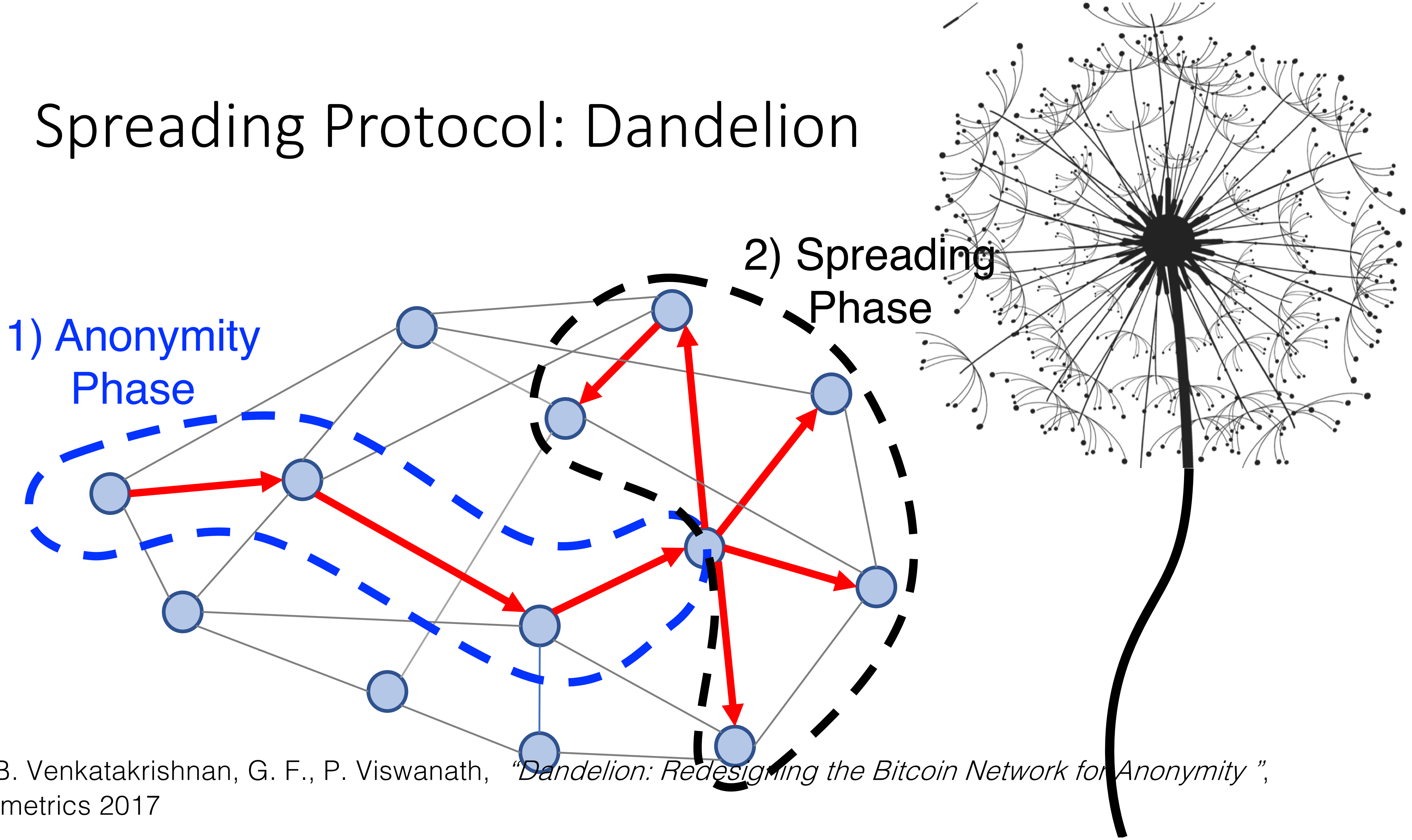
Onion routing (use Tor, I2P)

Mix networks (implement one in the P2P network)

DC Nets

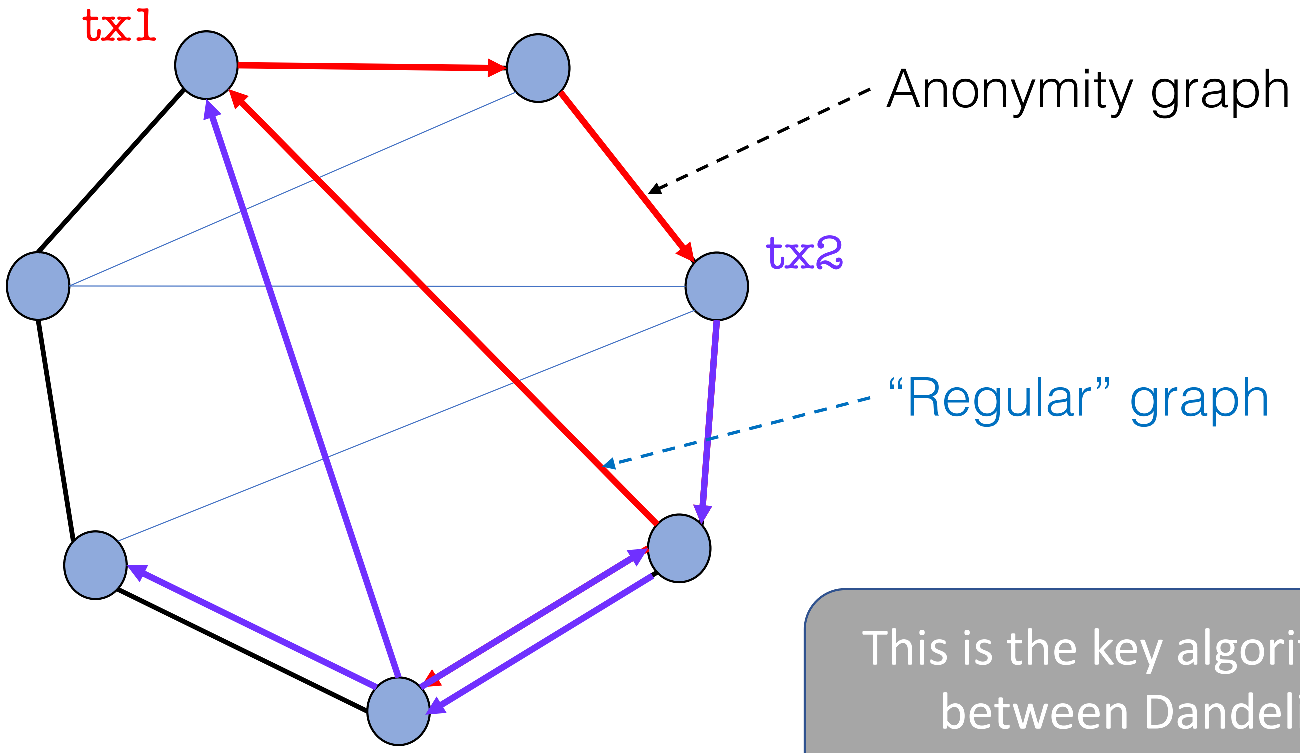
Dandelion routing

Spreading Protocol: Dandelion



S. B. Venkatakrishnan, G. F., P. Viswanath, *"Dandelion: Redesigning the Bitcoin Network for Anonymity"*,
Sigmetrics 2017

Graph Topology: Line



This is the key algorithmic difference between Dandelion and prior approaches, e.g. Crowds

Why Dandelion spreading?

Theorem: Dandelion spreading has an optimally low maximum probability of detection of $p + O\left(\frac{1}{n}\right)$.

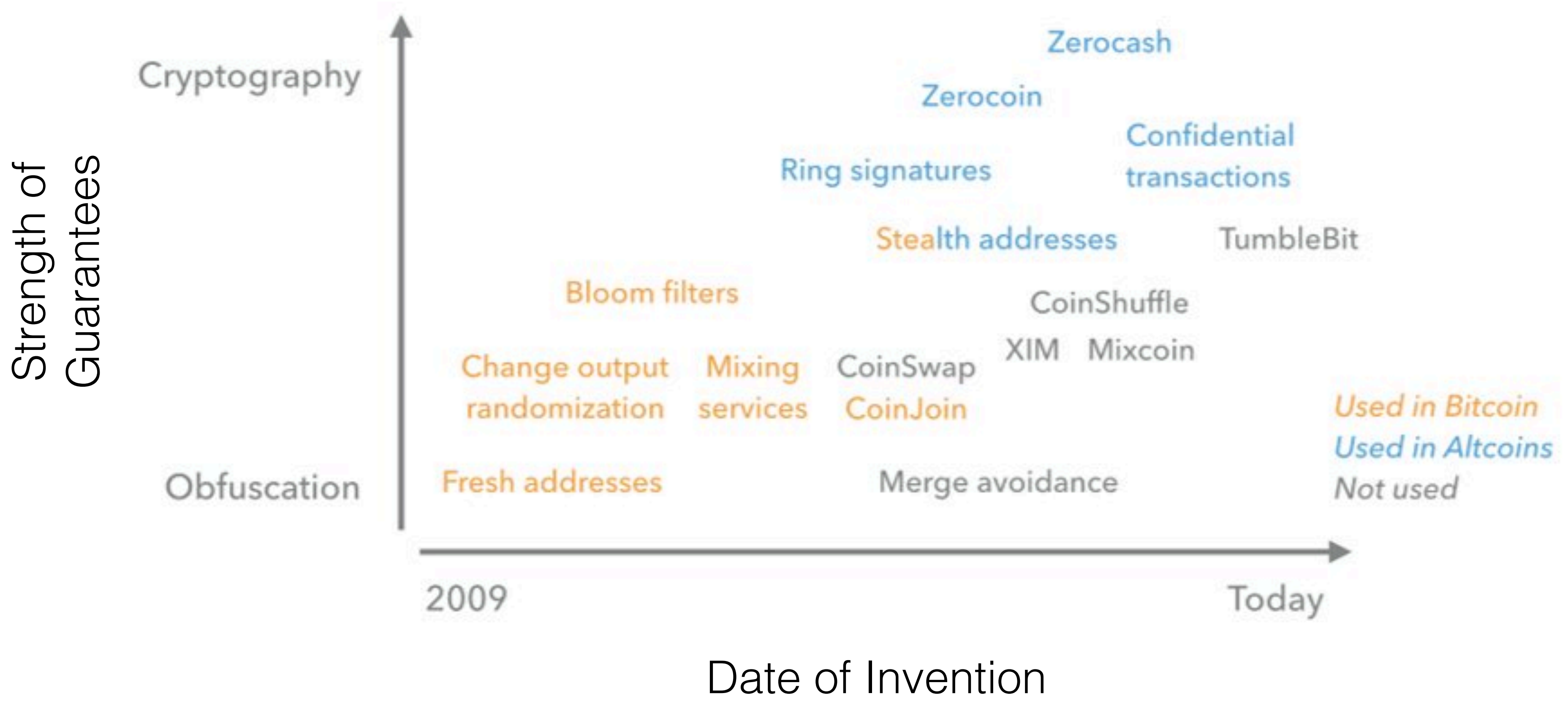
Theorem: Fundamental lower bound = p

fraction of spy nodes

number of total nodes

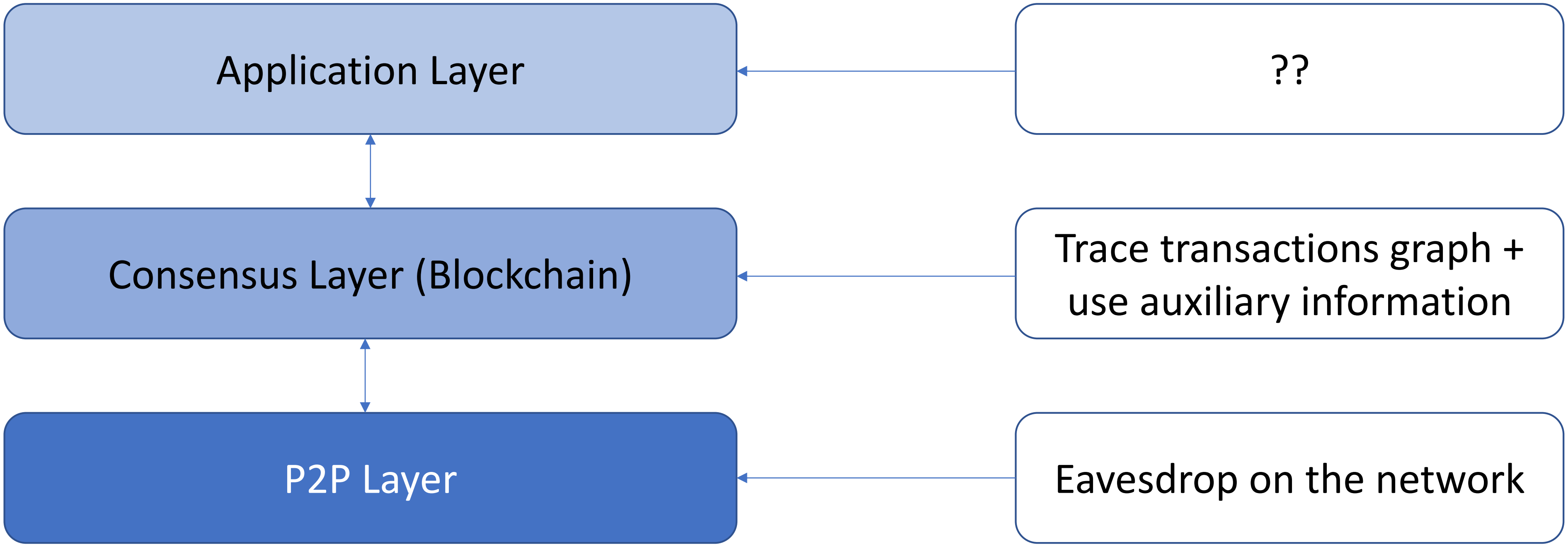
Tradeoffs

	Privacy Guarantees	Latency	Implementation Overhead	Biggest weakness
Onion routing	Statistical (timing-based)	Low	Moderate	Susceptible to timing attacks, sybil attacks
Mix networks	Anonymity within fixed set	High	Moderate	Requires trust in at least some mix nodes, high latency
DC Nets	Information-theoretic	High	High	Completely impractical
Dandelion	Statistical (topology based)	Low	Low	Susceptible to sybil attacks



Narayanan and Möser, 2017

How to deanonymize blockchain transactions



Application Layer

Consensus Layer (Blockchain)

P2P Layer

??

Trace transactions graph + use auxiliary information

Eavesdrop on the network

Consensus Layer: Zerocoin and Zerocash

Privacy-preserving protocol for digital currency

Masks link between public keys using zero-knowledge proofs

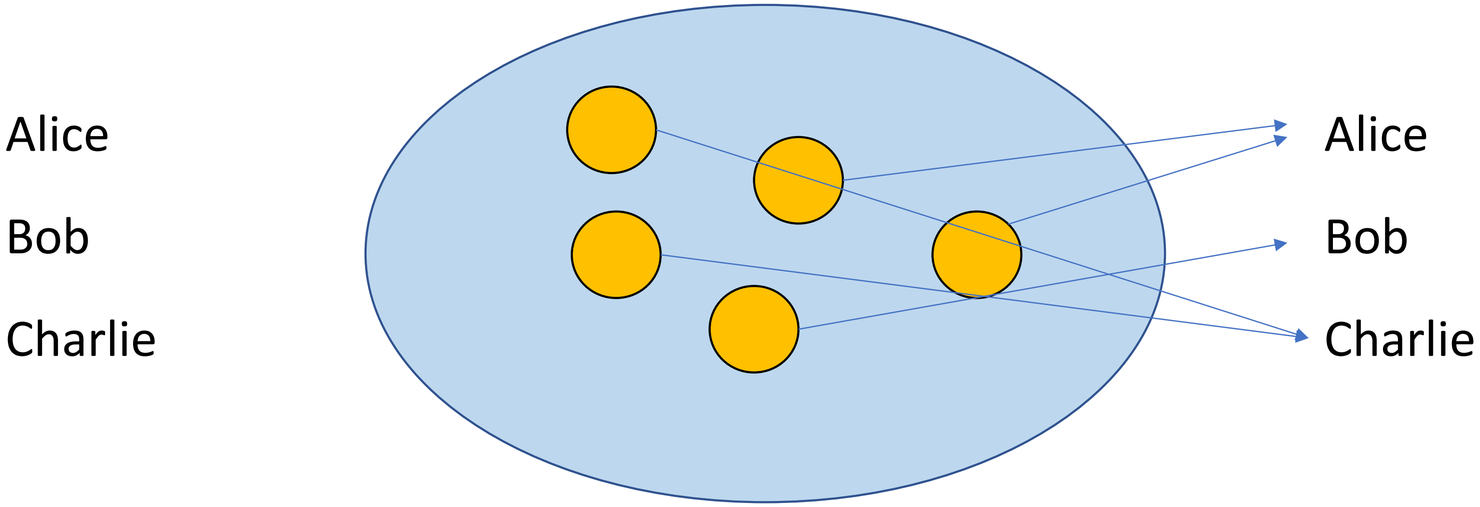
Basis of Zcash cryptocurrency

Zerocash: Decentralized Anonymous Payments from Bitcoin. Ben-Sasson, Chiesa, Garman, Green, Miers, Tromer, Virza. IEEE S&P 2014

Zerocoin: Anonymous Distributed E-Cash from Bitcoin. I. Miers et al. IEEE S&P 2013

Basic idea

- Create a pool of interchangeable tokens
- Put them on the blockchain
- Users add money to the pool ahead of time
- Withdraw from the pool to spend money



Alice

Bob

Charlie

Alice

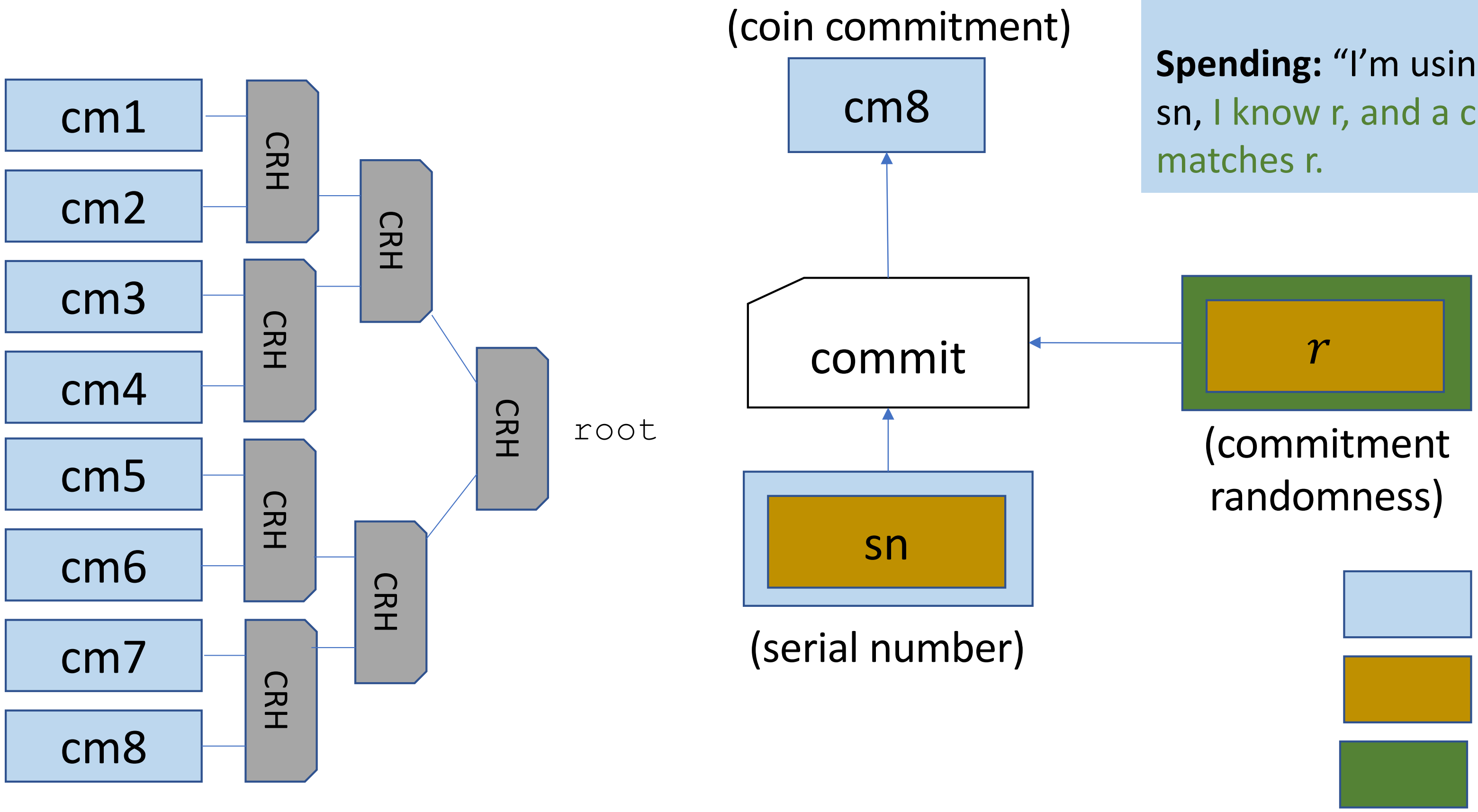
Bob

Charlie

Basic anonymous e-cash

Minting: “I spend 1 BTC to create cm”

Spending: “I’m using a coin with unique sn, I know r, and a cm in the tree that matches r.



🟦 In public ledger

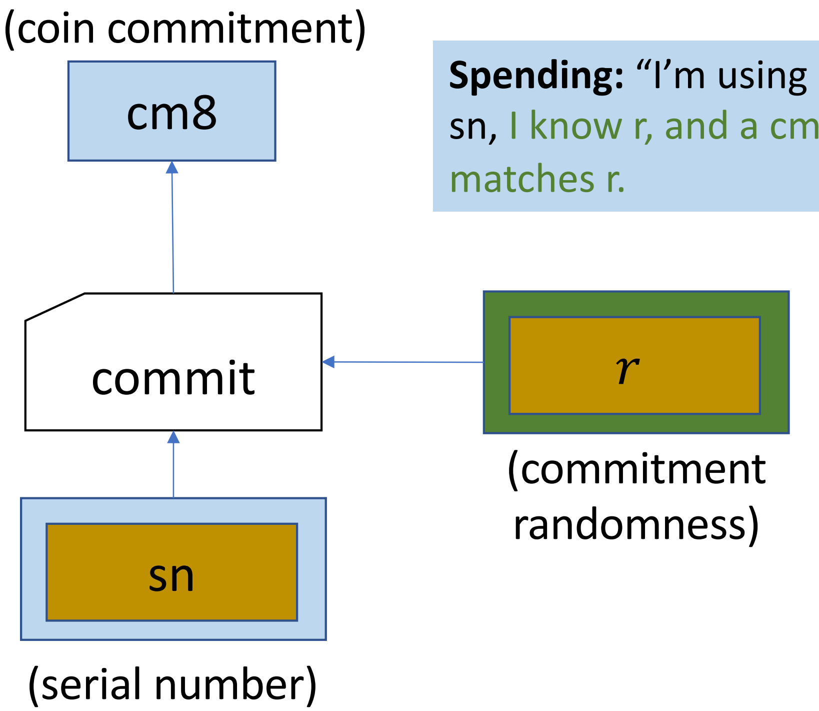
🟧 In private wallet

🟩 Proved to be known

Sander, Ta-Shma. Auditable, “Anonymous Electronic Cash”, CRYPTO 1999.

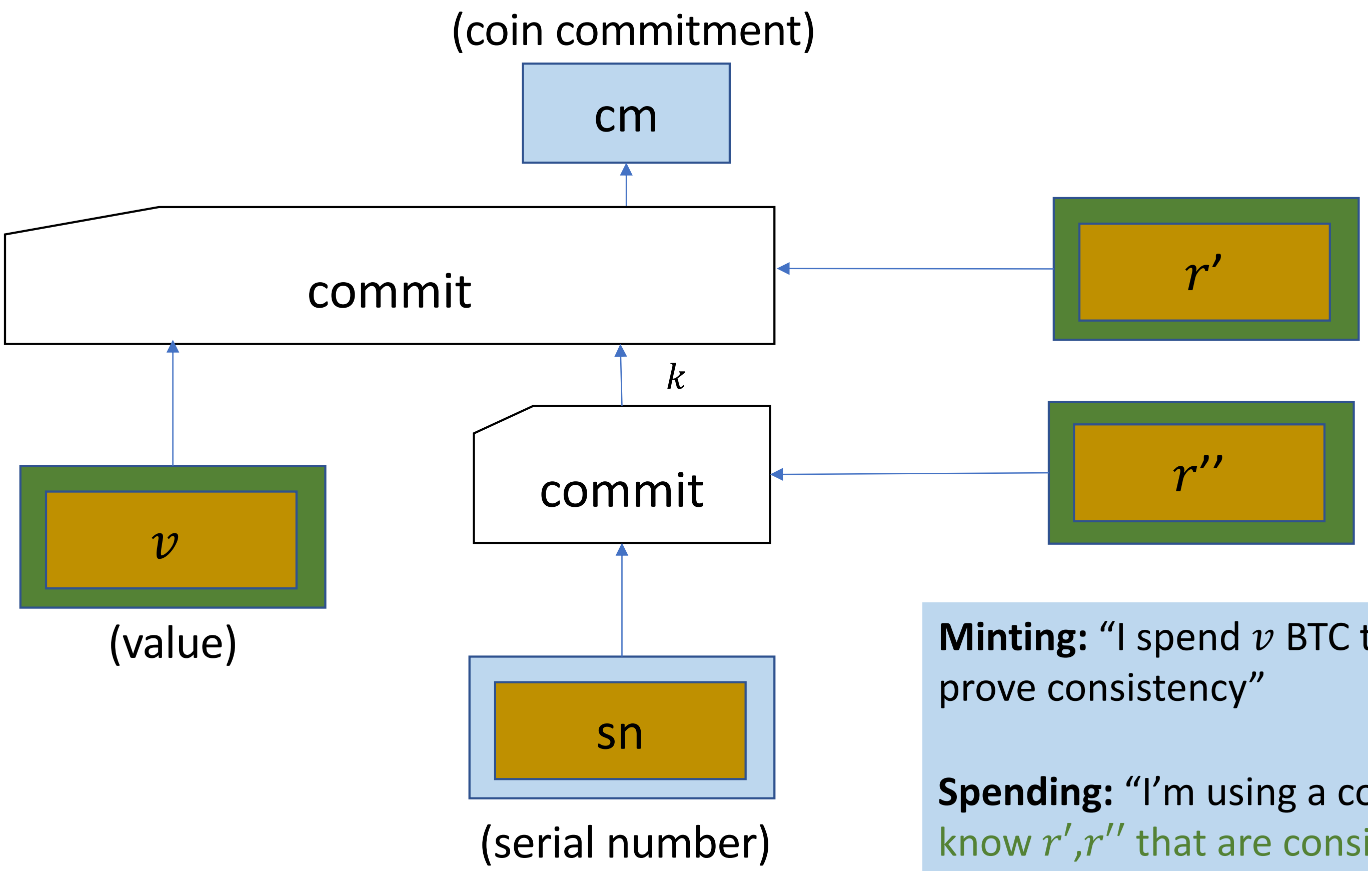
Key ingredient: zkSNARK

- Spending requires:
  - **z**ero **k**nowledge
  - **s**uccinct
  - **n**oninteractive
  - **a**rgument
  - of **k**nowledge



Spending: "I'm using a coin with unique sn, I know r, and a cm in the tree that matches r.

Adding variable denomination



Minting: “I spend v BTC to create cm, and here is k, r' to prove consistency”

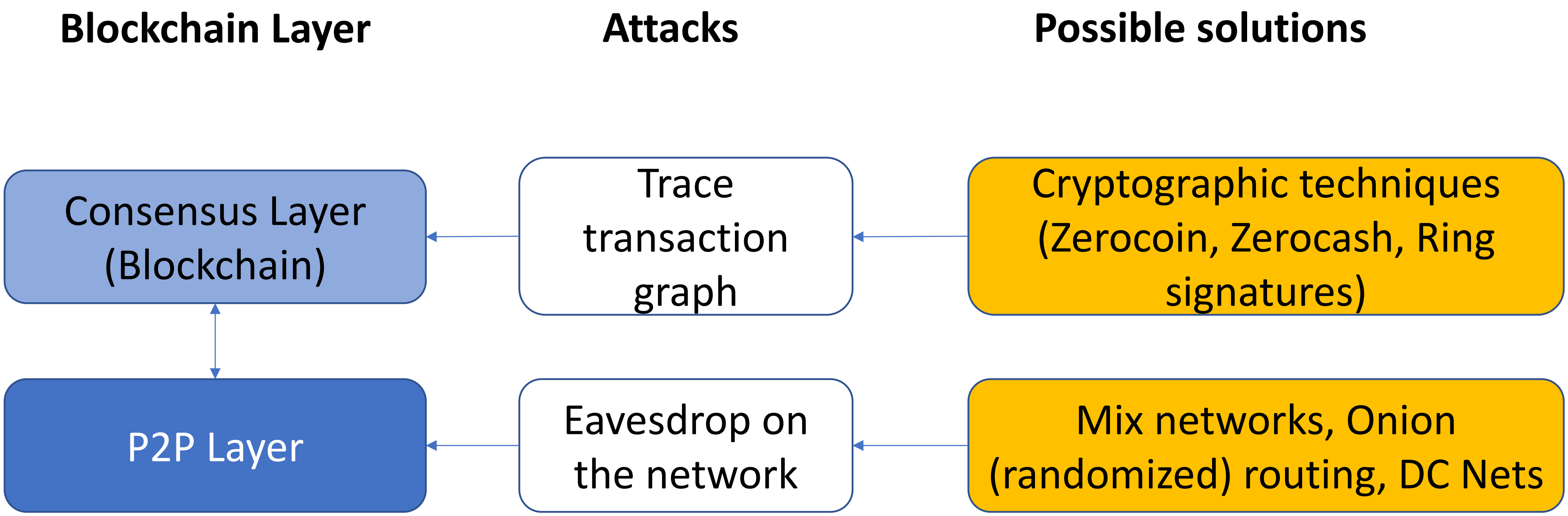
Spending: “I’m using a coin with value v, unique sn, and I know r',r'' that are consistent with cm

Performance

- Efficiency:
  - – 288 proof bytes/spend at 128-bit security level
  - <6 ms to verify a proof
  - <1 min to create a proof (for 264 coins, asymptotically log(#coins))
- Trust in initial generation of system parameters (once)

- This performance has been improved since

Blockchain Privacy Challenges



Blockchain Layer

Attacks

Possible solutions

Consensus Layer (Blockchain)

Trace transaction graph

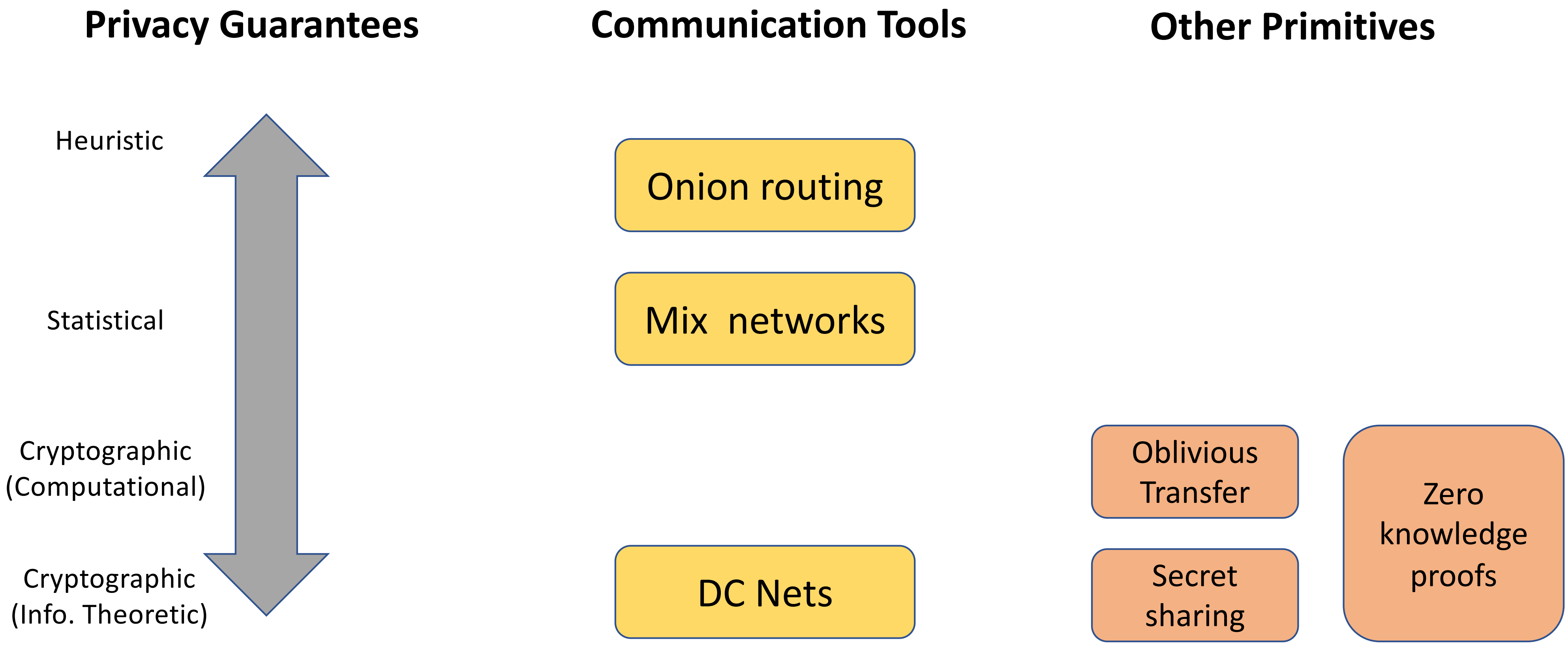
Cryptographic techniques (Zerocoin, Zerocash, Ring signatures)

P2P Layer

Eavesdrop on the network

Mix networks, Onion (randomized) routing, DC Nets

Summary of Unit III



Privacy Guarantees

Heuristic

Statistical

Cryptographic (Computational)

Cryptographic (Info. Theoretic)

Communication Tools

Onion routing

Mix networks

DC Nets

Other Primitives

Oblivious Transfer

Secret sharing

Zero knowledge proofs

What you should be able to do: Unit III

- Identify potential privacy leaks in a full system
  - Communication and/or data storage+processing
  - E.g. blockchain example

- Identify what tools can be applied to various privacy problems

- Design/understand rudimentary systems using these tools
  - reason about their tradeoffs

Course Summary

Units	Main challenges	Main tools
1) Enforcing Privacy Policies	- Translating English policies into binary - Detecting non-obvious violations (e.g. data use)	- First-order Logic
2) Privacy + Fairness in Big Data	- Identifying attacks - Designing useful + PP systems - Defining/measuring fairness - Implementing fair ML pipelines	- K-anonymity - Differential privacy - Fair ML
3) Special Topics	- Identifying attacks - Designing PP systems that are **efficient** (latency, comm)	- Cryptographic tools - DC Nets, Onion routing, secret sharing, zk proofs, oblivious transfer

Wrap-Up Game: Good fit or bad?

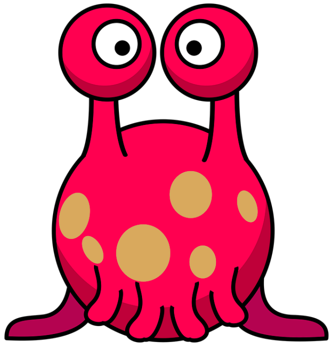
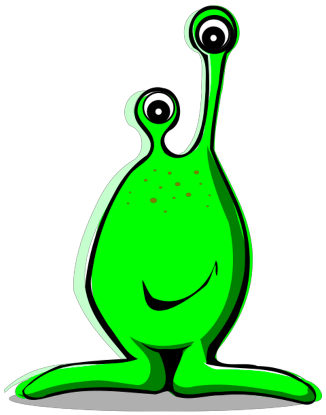
Check your understanding

Rules

- I will give you a scenario where a tech company wants to use a privacy/fairness technology
- You tell me if the technology makes sense for this problem
- If not, what tool(s) would be a better fit?

Spacebook

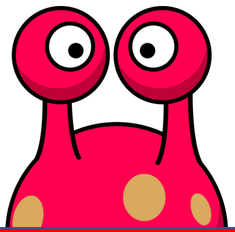
- Collects lots of data about aliens—community is getting worried

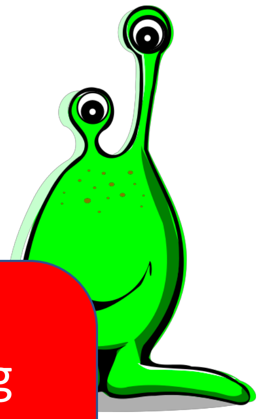



- Announces initiative to use global differential privacy to protect user privacy

Spacebook

- Collects lots of data about aliens—community is getting worried





- Announces otect user privacy

Bad!

Aliens are worried about Spacebook holding their data. Global DP doesn’t prevent that.

Instead: Use local DP.

Hooli

- Realized they are spending millions of dollars annually on privacy policy compliance
- They decide to invest in setting up an automated system (e.g., Grok + Legalease) to automate the checking process

Hooli

- Realized they are spending millions of dollars annually on privacy policy compliance
- They decide to invest in setting up an automated system (e.g., Grok + Legalease) to automate the checking process

Good fit

Tools like Grok and Legalease are designed precisely to automate privacy policy checking.

Anonymous Internet

- Blueseed decides to build protocols and infrastructure for an anonymous Internet

- Decides to use onion routing

Anonymous Internet

- Blueseed decides to build protocols and infrastructure for an anonymous Internet

- Decid

Good fit

Onion routing is designed to provide privacy for low-latency applications like web browsing.

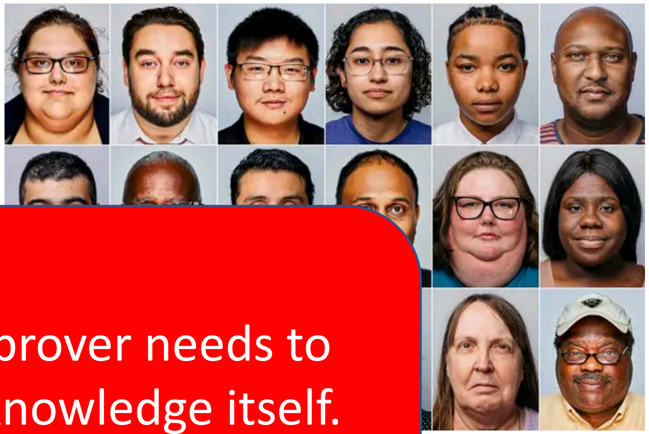
Building a completely anonymous-by-default Internet may be problematic for other reasons (spam, abuse, etc.)

ABC: Health Data

- A 3-letter agency (ABC) has a dataset of health records for all of its citizens
- Developing algorithms to predict who is at risk of depression
- Want to share this data with companies (e.g., Spacebook), without revealing too much data
- A consultant advises them to use zero-knowledge proofs

ABC: Health Data

- A 3-letter agency (ABC) has a dataset of health records for all of its citizens
- Developing algorithms to predict who is at risk of depression
- Want to share this data with companies (e.g., much
- A con know



Bad!

Zero-knowledge proofs are useful if the prover needs to prove knowledge without revealing the knowledge itself.

Instead: Reveal the classification without other data.

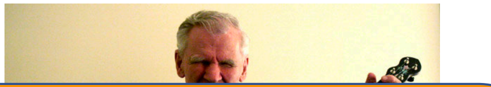
Scrapple

- Scrapple offers loans to aspiring musicians to launch their careers.
- Recently, a newspaper revealed that they are 3x more likely to offer loans to EDM musicians as they are folk musicians. The music world is in an uproar. Scrapple decides to do something about it.

- Scrapple announces a new initiative where they will use differentially-private machine learning to determine loan status.

Scrapple

- Scrapple offers loans to aspiring musicians to launch their careers.
- Recently, a newspaper revealed that they are 3x more likely to offer loans to EDM musicians as they are folk musicians. The music world is in an uproar. Scrapple decides to do something about it.



Might accidentally work

Differentially private classifiers are related to fair classifiers, and in some cases, fairness implies DP and vice versa. However, DP machine learning is not designed for fair ML.

Instead: Use a fair ML algorithm.

- Scrap ntially- private