Bitcoin & Anonymous Transactions

Dr George Danezis

University College London

<gdanezis@ucl.ac.uk>

Is cash really anonymous?

- Anonymity: the property of not having your "long term" identity linked with your actions.

- Anonymous cash actions?
  - Withdrawing some coins / cash.
  - Paying with some coins.
  - Depositing some coins.
  - Using coins for further transactions.

- Is real cash anonymous? [Kug04]
  - Notes have serial numbers.
  - Transaction chains may be short.
  - Linked banknotes on withdrawal.

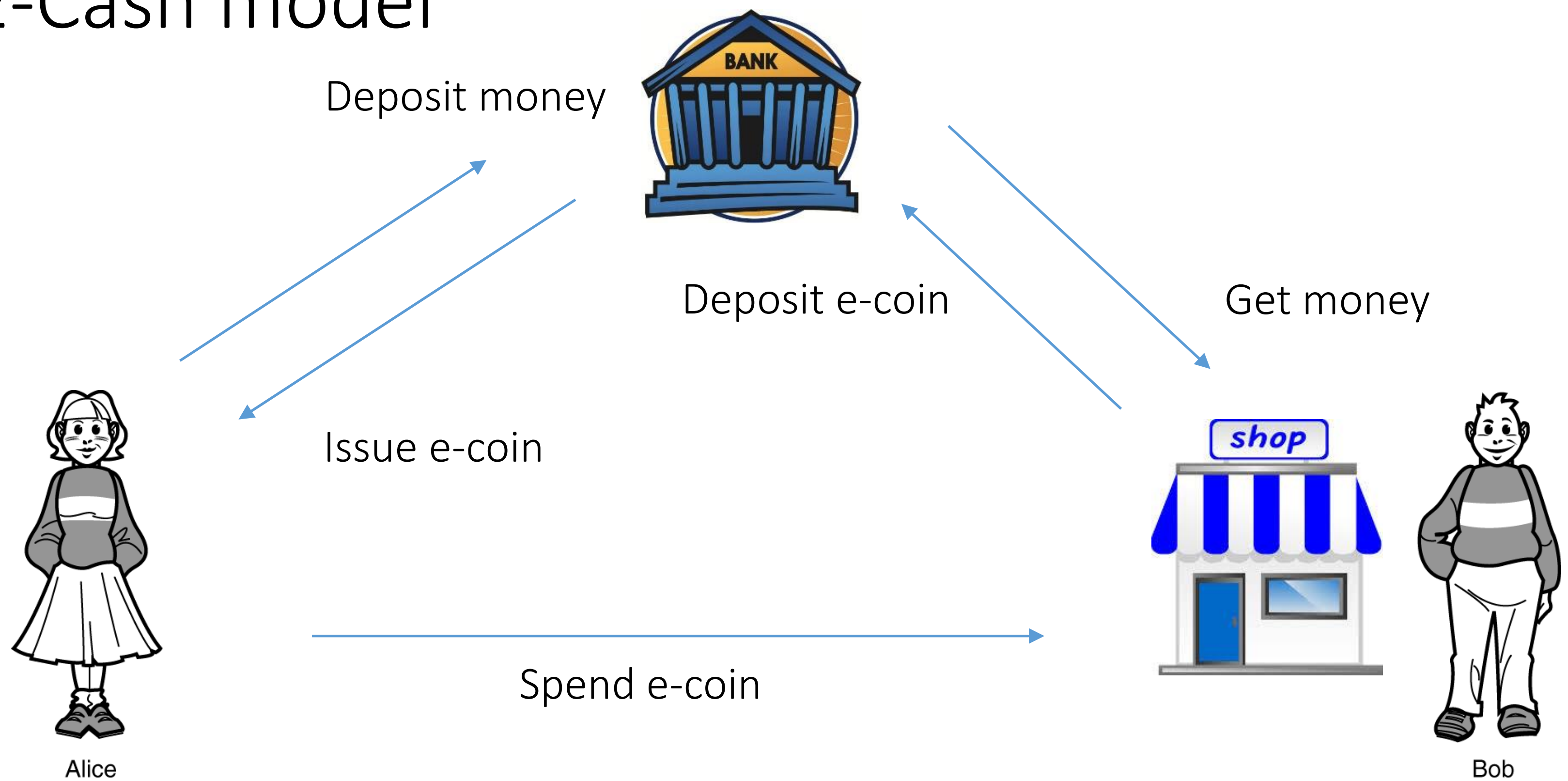
Dennis Kügler: On the Anonymity of Banknotes. [Privacy Enhancing Technologies 2004](): 108-120

Cryptographic Payments

- Mainstream banking:
  - Europay, MasterCard and Visa (EMV) protocols.
  - Interoperation of Cards, Point of Sale terminals (PoS), Automatic teller machines (ATM).
  - First standard EMV 2.0 in 1995.
  - Uses tamper-resistant hardware, symmetric crypto and (maybe) digital signatures.
  - No anonymity or privacy.

- Research & Development:
  - Digicash: Start-up of David Chaum (started 1990, bankrupt 1998).
    - Inventor or selective disclosure credentials.
    - Anonymous cash using cryptography – double spending prevention.
    - Long line of research on efficient e-cash: **we know how to do this**.
    - Model: central issuer of coins, in national currency denominations.

E-Cash model



- Key message:
  - We know how to do this extremely efficiently.
    - Jan Camenisch, Susan Hohenberger, Anna Lysyanskaya: **Compact E-Cash**. EUROCRYPT 2005: 302-321
  - Properties:
    - High authenticity– no double spending.
    - Privacy: Shop and Bank cannot tell who customer was.
  - However:
    - **Not a new currency – This is not what this talk is about!**
    - Centralized "Bank" service to issue and deposit (and hold real value)

E-cash assumptions

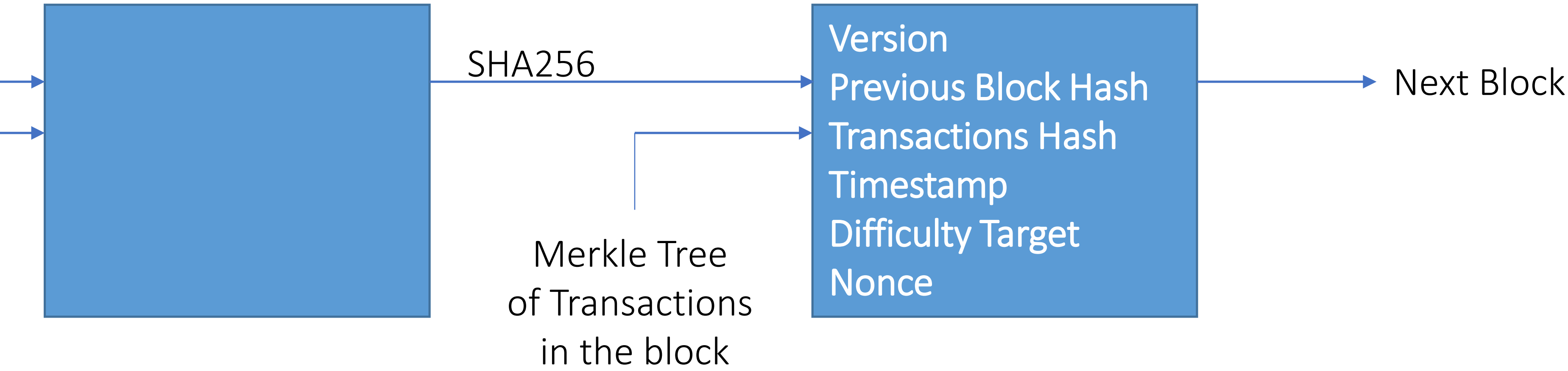
- Bank is a **trusted third party**.
  - **NOT for privacy** – a colluding bank and merchant cannot trace transactions.
  - Money supply: Bank has a signature key that allows the issuance of coins.
  - A rogue bank can “print money”.
  - Anonymity property makes auditability difficult.
- What is the problem with this model?
  - **Power balance**: bank “controls” the e-cash scheme.
  - Transactions chains: bank -> client -> merchant -> bank.
  - Transferable e-cash possible but inefficient for practical purposes.
  - Result: whoever controls the bank can create money or deprive any client / merchant of their money.
- Is that necessary? Is there a model where no central party or parties are necessary?

Bitcoin (BTC)

- Paper in late October **2008**.
  - Released as open source software in 2009
  - **Pseudonymous** developer Satoshi Nakamoto.
  - Disappears in mid-2010.
  - He is estimated to have about 1M BTC.

- Bitcoin features (as in the original email):
  - `Double-spending is prevented with a peer-to-peer network.`
  - `No mint or other trusted parties.`
  - `Participants can be anonymous.`
  - `New coins are made from Hashcash style proof-of-work.`
  - `The proof-of-work for new coin generation also powers the network to prevent double-spending.`

Memory: the block chain



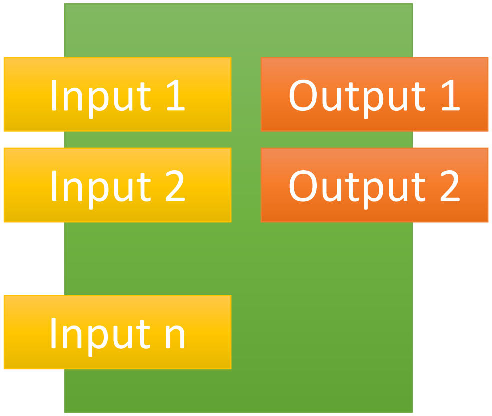
- A **block chain** storing all **transactions** is maintained by all.
- The last block is sufficient to guarantee the integrity of the full chain.
  - They form a hash chain of other blocks and transactions.
- The longest valid chain is recognized by all as the authoritative chain.
  - Blocks have some validity constraints that make them acceptable to all.

Transactions

Each input address signs the transaction.

The address and key must previously be in the block chain.

All value is input.



Transaction

Specify an output value and public key to transfer funds to.

Typical: Transfer and change

(Remaining go to miner as transaction fees to be included.)

- Bitcoins are transferred between addresses.
  - Address is identified by hash of public key
  - Private key used to sign transactions to spend coin.
  - Security property: authorization!
- Special transactions …

Where BTC money lives?

- Money lives in a wallet.
  - Each wallet has an "**address**"
  - Wallet – stores the secret key for all user BTC addresses.
  - Secret keys are just bit sting.
  - If seen by an adversary they can transfer coins away from you.
  - Theft!

- Where do you put the wallet?
  - On client software. Downside: you get hacked – "bye bye" BTC.
  - On services. Exchanges and wallet services.
    - The service gets hacked – everyone's money is stolen.
  - In hardware: a market in its infancy but growing
    - Parallel to Hardware Security Modules.

- Key insight: Hacking now allow you to steal money!
  - So are bad random number generators for the addresses.

Money supply: hashcash

How to avoid a centralized party controlling the money supply?

- Hashcash (Adam Back):
  - Make users find hash collision to rate limit supply in distributed manner.
  - Original use: DoS prevention.

- Who controls the money supply?
  - Convention in code.
  - **Mining**: Take all advertised transactions and try to make a block.
  - A block is made using the previous block, transactions and nonce.
  - Hash of valid blocks need to be smaller than a target difficulty agreed by all.
    - Lottery
  - Difficulty level – tuned for 1 block every 10 minutes.

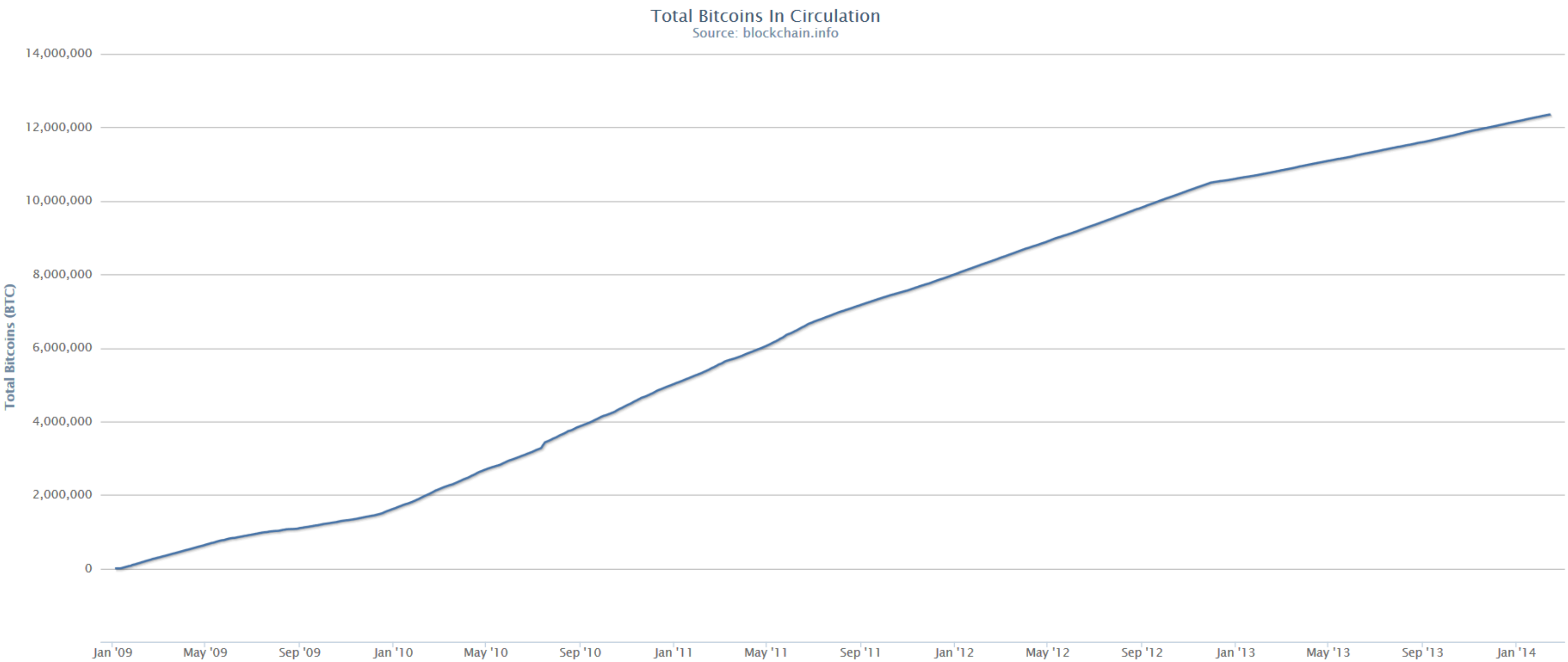
- Details
  - A single special transaction is within each block to create new Bitcoins.
  - How many depends on the length of the block chain.
  - Bitcoins in existence will never exceed 21 million.
  - After that? Transaction fees should kick in to provide incentives to mine.

Double spending prevention

- Each transaction is **broadcast** to all miners in the network.
  - Massive peer-to-peer broadcast network.
- Miners only include, in the new block calculation, transactions that do not have inputs already spent.
- Other miners check blocks for double-spending, otherwise block is invalid.
- After a transaction has been included in a mined block it has received one confirmation.
  - Usually clients wait for 6 confirmations to consider a transaction confirmed.
  - 1 block = 10 min means 1 hour wait.

How much is a BTC worth?

Total BTC in circulation



About 12M BTC out of a maximum of 21M BTC have been mined in 5 years.
Rate of BTC / mining will slow down as more BTC are mined:
25BTC per block today, 6BTC per block by 2020.

How big is the BTC market

A year ago it was small. Now it is big.

Is Bitcoin really anonymous?

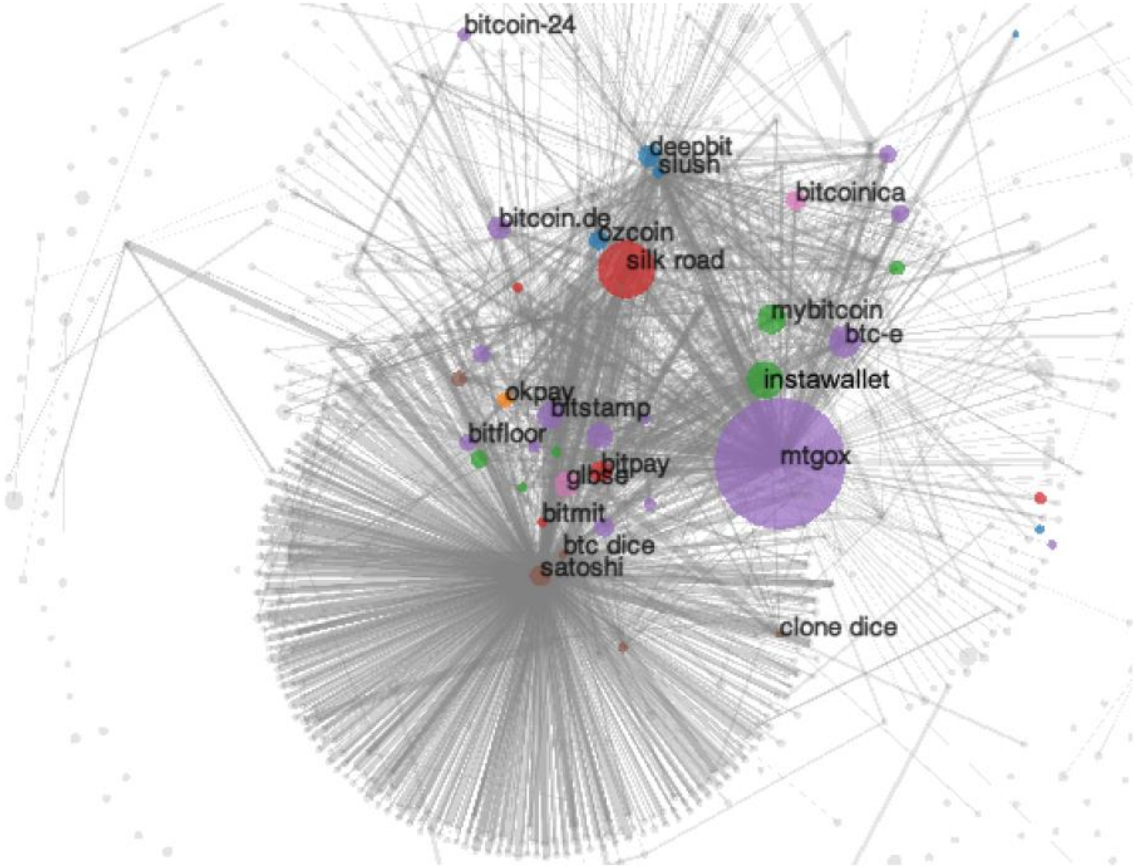
- BTC flows from “address” to “address”.
  - Pseudonymous – not tied to a human, just a secret key.

- However:
  - Exchanges accept national money and provide BTC.
  - Those nowadays implement “know your customer” policies
    (Or payments can be traced if done via conventional banking)
  - Once money is in BTC you can follow money flow chains.
  - It goes into banking system when it leaves.

- Forensic accountancy tricks:
  - Each transaction has many inputs, but two outputs:
    The recipient.
    The change address – this is the same as the sender.
  - Many small change addresses are consolidated to buy big things.
  - Result: can trace, and group, addresses per owner over time.

- In fact: **everyone can do investigations on the public graph.**

Mapping the Bitcoin network



Tracking thefts (1)

- From: UCSD

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

- Case Study 1:
  - The Betcoin gambling site was hacked in April 2012
  - 3,171 BTC were stolen in total (2902, 165, 17, and 87 BTC).
  - Did not move until March 15 2013 (bitcoin goes up)
  - Aggregated with other small addresses into one large address
  - Then began a peeling chain.
  - After 10 hops, a peel went to Bitcoin-24,
  - And in another 10 hops a peel went to Mt. Gox;
    in total, 374.49 BTC go to known exchanges, all directly off the main peeling chain, which originated directly from the addresses known to belong to the thief.

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

Tracking thefts (2)

- Case Study 2: Bitfloor theft
  - Large peels off; several initial peeling chains were then aggregated, and the peeling process was repeated.
  - Nevertheless, by manually following this peel-and-aggregate process to the point that the later peeling chains began, systematically followed these later chains and again observed peels to multiple known exchanges.
  - The third peel off one such chain was 191.09 BTC to Mt. Gox, and in total we saw 661.12 BTC sent three popular exchanges (Mt. Gox, BTC-e, and Bitstamp).
- Case Study 3:
  - Thief stole bitcoins by installing a trojan on the computers of individual users
  - Unable to confidently track the flow of the stolen money
  - Most of the stolen money did not in fact move at all
  - Of the 3,257 BTC stolen to date, 2,857 BTC was still sitting in the thief's address, and has been since November 2012.
- Conclusion: It is very **hard to exfiltrate the proceeds of crime** at scale.

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

Zerocoin: How to make Bitcoin anonymous?

http://zerocoin.org/

- Key idea:
  - Each Address has a hidden serial number and a key.
  - When spending, you have to release the serial number and sign with the key.
  - You also prove that the serial number and key are in the block chain.
  - Without revealing where!

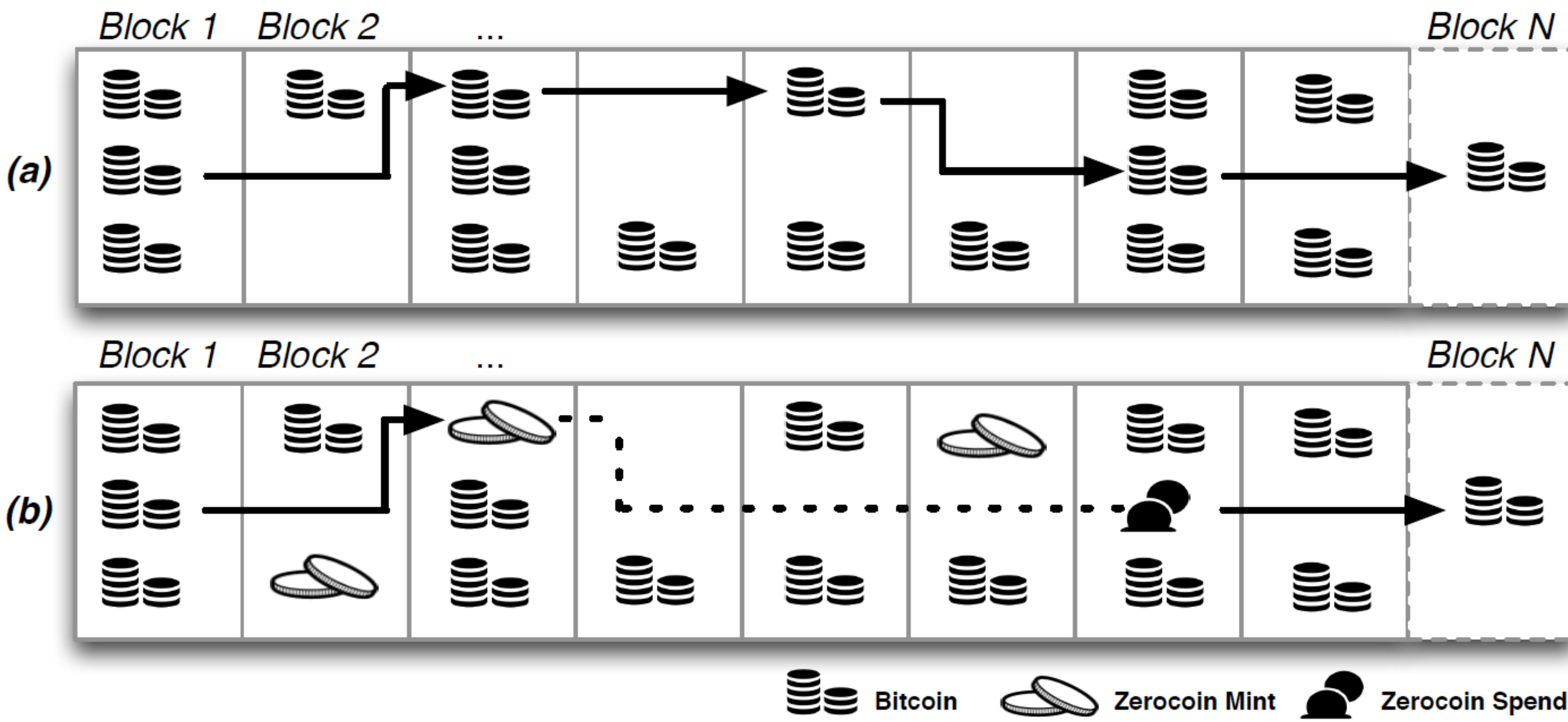
- Security properties:
  - Integrity: Zero-knowledge proof that serial and key are in the block chain.
  - Double-spending prevention: check that the serial is not already used.

What are the challenges?

- Privacy and integrity:
  - Before a transaction is accepted as genuine and included in the block chain.
  - You need to prove: (Integrity)
    - (a) The coin was indeed in the block chain.
    - (b) The coin has a certain value.
    - (c) The coin has not already been spent.
  - BUT You must not reveal: (privacy)
    - (a) Who transferred the coin to the spender.
    - (b) Who the coin is being transferred to.

- How to combine integrity and privacy:
  - Zero-knowledge proofs of knowledge.
  - Allow you to prove that a hidden value is known and satisfies some conditions.

Bitcoin vs. Zerocoin



Intuition behind Zerocoin cryptography

- A commitment scheme allows Alice to Commit to a value S.
  - c = Commit(S, r) for a random large r.
  - Can “open” the commitment by revealing r and S.
  - Can prove she knows S and r (without revealing them).
  - Not a unicorn: $c = g^s h^r \bmod p$ (Pedersen Commitment)

- Zerocoin **minting**:
  - Commits to a serial number S as c.
  - She transfers a fixed amount of BTC to “c”, and places the transaction in the block chain.
- Zerocoin **spending**:
  - Alice reveals an S (but not r!)
  - Alice prove she knows an r such that S and r are valid openings of one of the previous minted Zerocoins.

- Note: the proof is in zero-knowledge and does not leak exactly which Zerocoin was spent.
  - Challenges: how to make this proof efficient? (Linear restricts performance).
  - May have to limit validity of Zerocoins to epoch.
  - The larger the set of valid Zerocoins, the more anonymity.

A mix based approach to anonymity

- Zero-coin: allows a number of coins to have their depositors confused with each other.
- Simpler solutions: “mix” services for coins.
  - Send your coin to a mix, it waits to for other coins, and then pays them out.

What problems do you foresee?

The Mix is a TTP

- TTP: trusted third party.
  - Receives the coin, and a private instruction to pay it to another wallet.
  - This is what **silk road accounts** achieved!
- Privacy: a corrupt mix may store the link between input and output coins.
  - Impossible to audit forgetfulness.
  - However: may use many mixes – rely on one of them being honest and forgetful.
- Integrity: a corrupt mix may never pay out to the output wallets.
  - Bad. Difficult to prevent this.
  - Can build reputation over time.

Traffic analysis applied to money flows

- Is a mix a perfect anonymization tactic?

- Scenario:
  - you are a thief; you just stole 10000 BTC.
  - You wish to “launder” them through a mix.

- Option 1: you send an instruction to pay an “anonymous” output address 10000 BTC.
  - Problem: it is very unlikely that any other large trade goes into the mix.
  - Therefore the amount allows you to trace the input and output address.
  - Secure mixing requires uniform amounts! (Say 1 BTC per mixing round)

- Option 2: you send 1 BTC / round over 10000 rounds.
  - Its slooooooooooooooooooooooooooooooooooooooooooooooooow
  - Is it really secure?
  - No! You can trace the transactions by solving a set of linear equations!

Statistical disclosure attacks on Bitcoin mixes

- Model each output address amount as a linear combinations of input addresses:
  - $Cout_i$ is the amount going to the i^{th} account.
  - Cin_j is the amount going in from j^{th} account
  - $Cout_i = Sum_j\ w_j\ Cin_j$

- Observe a number of repeated transactions:
  - Gather many instances of $Cout_i$
  - They form an over-determined set of linear a equations.
  - Determine the set of w_j
  - To minimize the square error $(Cout_i - Sum_j\ w_j\ Cin_j)^2$

- Result: input accounts with high w_j are the likely input accounts.
  - Well known long term traffic analysis attack: “Statistical disclosure attack”.
  - Can be extended to accounts and mixes with memory.

Danezis, George. "Statistical disclosure attacks." *Security and Privacy in the Age of Uncertainty*. Springer US, 2003. 421-426.

Proper money laundering will cost you

- Ideas – do not try those at home – anti-money laundering laws probably do apply to BTC transactions.
  - Mixed blessing of recognizing BTC as "money".

- How to launder money:

  - Run a "fake" high-street business? (And pretend it is really popular)
    Not so easy since all inwards "cash" flows are traceable.

  - Play in a Casino? Play poker against yourself? Lose.
    Low return rates. Not cash – traceable.

  - Manipulate a market?
    Buy a certain resource to the point its price goes up, sell it from another account. Acts like a mix, and involves others. Is there such a good in the BTC economy?

Opportunities

- The transcation graph is public – should we mine it?