

Blockchain Cheat Sheet

4. Atomicity
- A BC transaction supports sequential actions, which can combine multiple financial operations.
 - Flash loan example.
 - This combination can be enforced to be atomic.
 - While this programmable atomicity property mostly absent from CeFi, (likely costly and slow) legal agreements could enforce atomicity in CeFi as well.
5. Execution Order Malleability
- Users on permissionless blockchains typically share publicly the transactions.
 - No centralized entity ordering transaction execution, peers can perform transaction fee bidding contests to steer the transaction execution order.
 - In CeFi: regulatory bodies impose strict rules on financial institutions and services as in how transaction ordering must be enforced.
6. Transaction Costs
- Transaction fees in DeFi and blockchains in general are essential for the prevention of spam.
 - In CeFi, financial institutions can opt to offer transaction services at no cost (or are mandated by governments to offer certain services for free) because of the ability to rely on KYC/AML verifications of their clients.
7. Anonymous Development and Deployment
- Many DeFi projects are developed and maintained by anonymous teams.
8. Non-stop Market Hours
- It is rare for CeFi markets to operate without downtime.
 - DeFi has no pre- or post-market trading.
 - System outages at CeFi stock happened.

Regulation

- Regulatory uncertainty
- Censoring (Temporarily) Transactions
 - Miners can decide to temporarily censor transactions
 - Nodes in lightning may simply refuse a transaction (forcing the user to fall-back to on-chain payment channels).
- Blacklists, Fungibility and Destruction of Assets
 - Once a service provider is KYC/AML regulated, the freezing and confiscation of financial assets may be requested.

Future?

DeFi could become the underlying infrastructure of future banks, whereas traditional finance / custody adapt.

Exchange Rate

Cefi

ask/bid, sell/buy, the last trade.

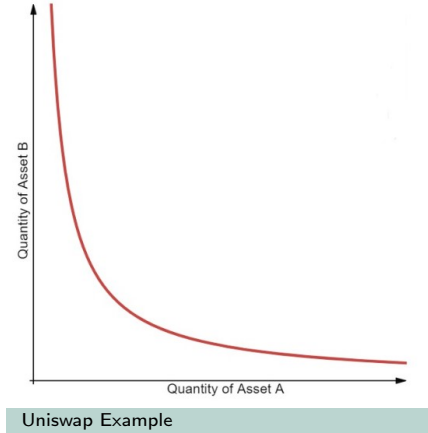
- Price changes if trade happens, ask was same or lower than bid. Ask/bid submitted by users
- Slippage: you see a price, submit, and until its executed, price can change.
- Order/time important → frontrunning, more data stored on chain

Defi

ratio of pairs: function that define price. Also slippage. Large swap can change price (as with CEX).

Decentralized Swaps

Uniswap uses $X * Y = k$, where k is constant, X and Y are asset values. (if you take out X you need to provide Y). Therefore pool can not be drained:



Uniswap Example

Simple exchange price calculation (Uniswap):
Swap for 0.5 ETH, if you send 0.5 ETH to pool:

$200/1.5 \rightarrow 133DAI \rightarrow 133DAI \text{ for } 1.5ETH$

Deduct 66 DAI from pool $\rightarrow 133/1.5 \rightarrow 88$ DAI for 1 ETH. Not draining the pool, but trading with better price than resulting pool.

Algorithms

Bloom-Filter

Variations

Compressed Bloom Filters

When the filter is intended to be passed as a message. False-positive rate is optimized for the compressed bloom filter (uncompressed bit vector m will be larger but sparser)

Generalized Bloom Filter

Two type of hash functions gi (reset bits to 0) and hj (set bits to 1). Start with an arbitrary vector (bits can be either 0 or 1). In case of collisions between gi and hj, bit is reset to 0. Store more info with low false positive. Produces either false positives or false negatives.

Counting Bloom Filters

Entry in the filter not be a single bit but a counter. Delete operation possible (decrementing counter).

Scalable Bloom Filter

Adapt dynamically to number of elements, consist of regular Bloom filters A SBF is made up of a series of one or more (plain) Bloom Filters; when filters get full due to the limit on the fill ratio, a new one is added; querying is made by testing for the presence in each filter.

Usage

Fast search. Check if username is taken: Check bloom filter instead of query database.

Merkle Tree

A Merkle tree is a binary hash tree containing leaf nodes. Constructed bottom-up. Used to summarize all transactions in a block. To prove that a specific transaction is included in a block, a node only needs to produce hashes, constituting a merkle path connecting the specific transaction to the root of the tree.

Location Transparency

Distributed Management/Retrieval of Data

Challenges:

- Location: Where shall the item be stored?
- Location: How can the item be found?
- Scalability: keep the complexity for communication and storage scalable
- Robustness and resilience in case of faults and frequent changes

Strategies for Data Retrieval

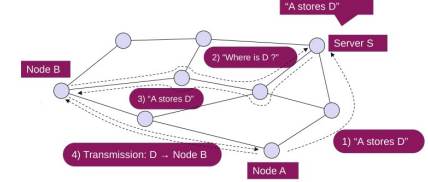
Strategies to store and retrieve data items in distributed systems:

Node State

System	Per Node State	Communication Overhead	Fuzzy Queries	No false negatives	Robustness / horizontal scalable
Central Server	O(N)	O(1)	✓	✓	(X)
Flooding	O(1)	O(N)	✓	(X)	✓
Distributed Hash Tables	O(log N)	O(log N)	(X)	✓	✓

System	Per Node State	Communication Overhead	Fuzzy Queries	No false negatives	Robustness / horizontal scalable
Central Server	O(N)	O(1)	✓	✓	(X)
Flooding	O(1)	O(N)	✓	(X)	✓
Distributed Hash Tables	O(log N)	O(log N)	(X)	✓	✓

- Central server**
- e.g., service registry, reverse proxy - although main use case is load balancing. Simple strategy: Central Server (can be powerful - vertical scaling!). Server stores information about locations:
1. Node A (provider) tells server that it stores item D
 2. Node B (requester) asks server S for the location of D
 3. Server S tells B that node A stores item D
 4. Node B requests item D from node A



Advantages

- Search complexity of O(1) - "just ask the server"
- Complex and fuzzy queries are possible
- Simple and fast

Problems

- No Scalability
 - O(N) node state in server
 - O(N) network and system load of server
- Single point of failure or attack
- (Single) central server not suitable for systems with massive numbers of users

Flooding search

e.g., layer 2 broadcasting, wireless mesh networks, Bitcoin Fully-distributed Approach and opposite approach of central server. No information on location of a content.

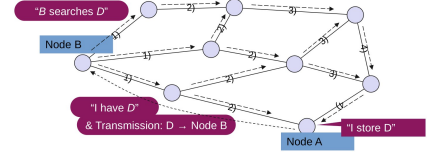
Retrieval of data:

- No routing information for content
- Necessity to ask as much systems as possible / necessary
- No guarantee to reach all nodes
- Flooding: high traffic load on network, scalability issues

How it works

There is an search fee to prevent spamming.

1. Node B (requester) asks neighboring nodes for item D
2. Nodes forward request to further nodes (breadth-first search / flooding)
3. Node A (provider of item D) sends D to requesting node B



For Bitcoin

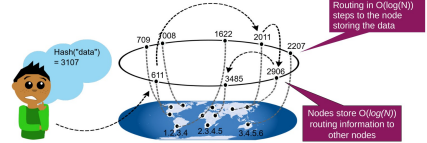
Bitcoin has all data in the block, so it searches itself.

Distributed indexing / Hash Tables

Tor, Bittorrent, IPFS, Apache Cassandra, Dynamo, Atek.

Goal is scalable complexity for:

- Communication effort: O(log(N)) hops
- Node state: O(log(N)) routing entries



1. Hash data
2. Every node knows log(N) neighbors (for 1000, min. 3)
3. First node asks neighbors
4. Neighbors references next neighbor, and so on.
5. Until node is found which is responsible for data item.

Approach

- Data and nodes are mapped into same address space
- Nodes maintain routing information to other nodes

Problems

- Maintenance of routing information required
- Fuzzy queries not primarily supported

Characteristics

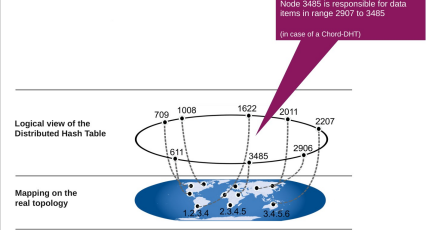
- Reliability / Scalability
- Equal distribution of content among nodes.
- Assignment of responsibilities to new nodes.
- Re-assignment and re-distribution of responsibilities in case of node failure or departure.
- Consistent hashing → nodes responsible for hash value intervals.
- More peers = smaller responsible intervals.
- Hash Table is something different.

Mapping of nodes and data

- Peers and content are addressed using flat identifiers: E.g., Address is public key (256bit) or SHA256 of public key. Content ID = SHA256(content)
- Nodes are responsible for data in certain parts of the address space
- Association of data to nodes may change since nodes may disappear

Storing

- Each node is responsible for part of the value range:
 - Often with redundancy (overlapping of parts)
 - Continuous adaptation
 - Real (underlay) and logical (overlay) topology are uncorrelated

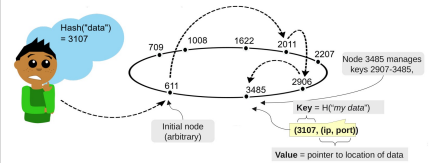


Storing / Looking up data

- Store data = first, search for responsible node (Not necessarily known in advance)
- Search data = first, search for responsible node

Routing to Data Item

- Start lookup at arbitrary node of DHT
- Routing to requested data item (key)
- K/V-pair is delivered to requester
- Requester analyzes K/V-tuple (and downloads data from actual location - in case of indirect storage)



Indirect vs Direct:

Indirect is one step more: e.g.: Download from IP, Port. Good for larger files.

Direct is good for smaller files.

Join

1. Calculation of node ID (normally random / or based on PK)
2. New node contacts DHT via arbitrary node (bootstrap node)
3. Lookup of its node ID (routing)
4. Copying of K/V-pairs of hash range (in case of replication)
5. Notify neighbors

Leave

Failure

- Use of redundant K/V pairs (if a node fails)

Blockchain Cheat Sheet

- Use of redundant / alternative routing paths
- Key-value usually still retrievable if at least one copy remains
- need for keep-alive messages

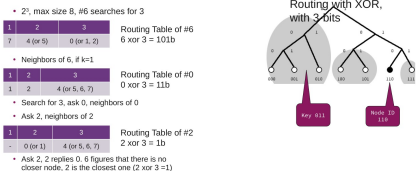
Departure

- Copying of K/V pairs to corresponding nodes
- Friendly unbinding from routing environment

Kademlia

Each Kademlia node and data item has unique identifier. Keys are located on the node whose node ID is closest to the key. Knows neighbors well, further nodes not that much.

Example Search



Sybil Attacks

- Create large number of identities
- Larger than honest nodes
- Isolate nodes

Prevention

- Creation of identities costs money
- Always assume data from other nodes may be missing
- Chain of trust / reputation

Redundancy

Direct Replication

- Originator peer is responsible
- Periodically refresh replicas
- Example: tracker that announces its data
- Problem: Originator offline → replicas disappear. Content has TTL



Indirect Replication

- The closest peer is responsible, originator may go offline vs any close peers are responsible.
- Periodically checks if enough replicas exist.
- Detects if responsibility changes.
- Problem: Requires cooperation between responsible peer and originator
- Problem: Multiple peers may think they are responsible for different versions → eventually solved

Consistency in Replicas

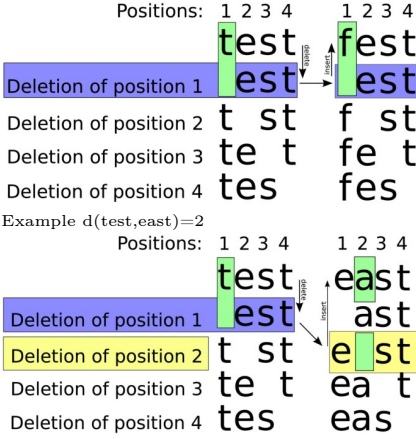
DHTs have weak consistency. If two changes are at the same time, one get overwritten. Requires a coordinator. (Leader Election with tools)

Created with X

Similarity Search					
Levenshtein Distance					
Example d(test,east) = 2 (remove a, insert t). Main idea: pre-calculate errors:					
		T	E	S	T
	0	1	2	3	4
E	1	1	1	2	3
A	2	2	2	2	3
S	3	3	3	2	3
T	4	3	4	3	2

If Letters are the same: Calculate left or top and add one or select diagonal top-left and add 0. Select minimum.
If letters are not the same: select the minimum of either left, top, or diagonal top-left and add 1.

FastSS
FastSS pre-calculates with deletions only. Neighbors for test with ed 2: test, est, st, et, es, tst, tt, ts, te, te, tes. 11 neighbors → 11 more queries, indexed enlarged by 11 entries.
Example d(test,fest)=1

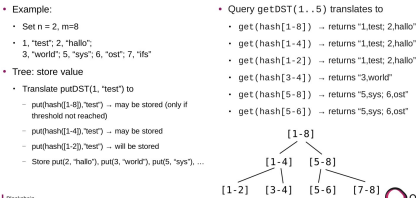


Range Queries
Problem: random insert vs. sequence insert

Solution: Over-DHT

DST: stores data on each level (redundancy) up to a threshold

Example DST



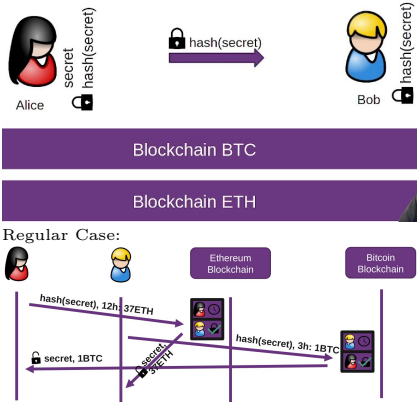
Cross-chain Atomic Swaps

Terminology
• Cross-Chain Swaps: Trade Ethereum for Bitcoin
• Atomic: Ether fully swapped or not at all. Use case: I want to exchange my 1 BTC to 37 ETH.
Obvious approach: use a centralized exchange, such as Binance, Bitstamp, or Kraken.
Good Approach: With Atomic Swaps no trust in a centralized platform is needed.

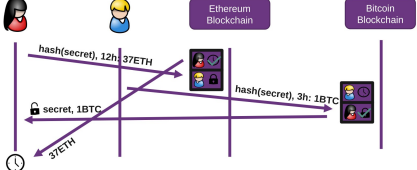
Hashed Time-Locked Contracts
Cryptographic hashing:
• one-way function - computationally efficient in one way, computationally highly expensive the other way
• deterministic - same input - same output
• collision resistant - highly expensive to find two inputs that hash to the same output

Hash lock
Building block for cross-chain atomic swaps and payment ch
• store hashed secret - publicly stored in a smart contract
• unlock - only if secret is provided (publicly)
• OR: unlock - after timeout
Both Chains need to support the same hash function. E.g.: SHA256

Example
Now we are ready to do an atomic swap with Alice and Bob with HTLC. Alice (initiator) creates "secret", shares with Bob, hash(secret). Bob now knows hash(secret).



Das Secret wird veröffentlicht. Wenn das Secret veröffentlicht und eigesetzt wird, bekommt Bob die ETH.
Worst Case:

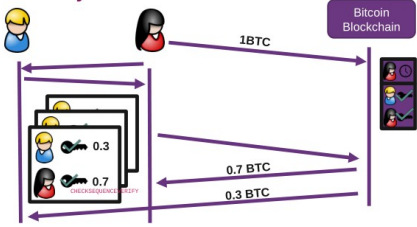


Layer 2 Payment Channels
Blockchains grow linearly: Bitcoin Blockchain is 376.6 GB.

Solutions
• First Layer Scalability Solutions
– Sharding (distribute storage)
– Improve protocol (SegWit, Taproot, Rollups)

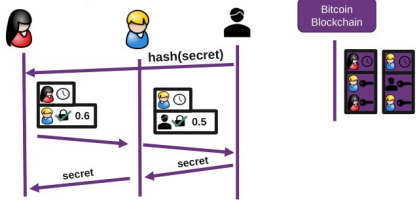
• Second Layer Scalability Solutions (off-chain)
– State Channels (payment channels): Lightning Network
– Sidechains / Blockchain Interoperability
2-of-2 Multisig
Initial offchain TX. Multiple transaction possible before everything is written to the blockchain. Offchain transactions.

Example
1 BTC of Alice to Locked Multisig. Alice can send 0.1BTC, then 0.2BTC, then the rest (or other constellation). Bob can request some amount, then another amount (not exceeding Alice tho) as long as the multisig contract is open.



After all transactions and when everyone's happy, both sign the contract and the total amount is given to the persons and the transaction is written to the blockchain.
For validation: CHECKSEQUENCEVERIFY.

Indirect Payment with HTLC
Idea of Lightning Network.
1 BTC lockup, Alice - Bob, Bob - Charlie.
Alice wants to send 0.5 BTC to Charlie (no direct channel).



WebRTC

- In General**
- WebRTC for browser to browser communication.
 - P2P, no server involved (-mostly. Server is needed to get the communication running).
 - Real time communication (RTC) via API.
 - Supported by Google, Microsoft, Mozilla, Opera, Apple
 - Standard still not finalized

- Use Case**
- Filling gap in the Web-Experience
– Video Chat → Google Hangouts Plugin, Flash, Java
– Multimedia / Conferences → Expensive, proprietary 3rd party apps
– Customer Service → Chat only, 3rd party plugins/apps
– Online Games → Flash
– Real-time Feeds → Proprietary software
– File Sharing → Requires Server / BitTorrent

- WebRTC widely deployed, no client necessary!
- Used in WhatsApp, Facebook Messenger, whereby.com
- WebRTC forbids unencrypted communication

Concerns
• Outdated documentation

- HTML browsers get bloated → Several GB RAM to open couple of tabs?
- WebRTC API could be simplified
- Security Concerns: Private IP / IP behind VPN, Tor? Not anymore!
- Complexity because of Security.

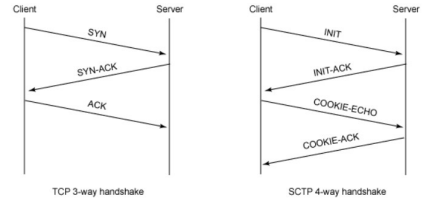
Stream Control Transmission Protocol (SCTP)

Pro

- SCTP Message-based (Like UDP, but reliable).
- Allows data to be divided into multiple streams.
- Syn cookies - SCTP uses a four-way handshake with a signed cookie.
- 32-bit end-to-end checksum (CRC32C).
- Multi-homing multiple IP addresses of endpoints.

Con

- Not widely used.
- If not used, tunneled over UDP.



Introduction

On the bright side: developer does not need to care about NAT. Abstraction using STUN, ICE, TURN.

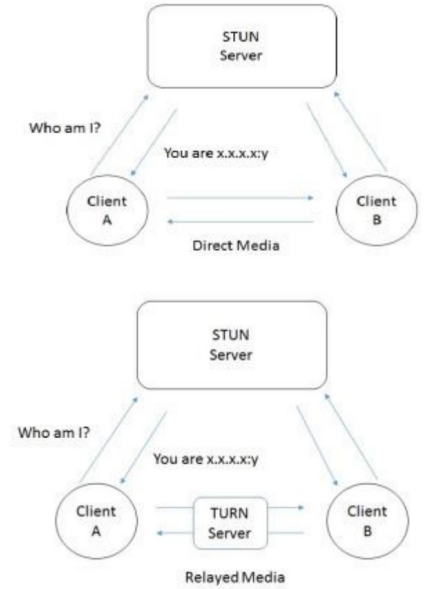
STUN

STUN: session traversal utilities for NAT (detect which kind of NAT). Uses Tests to figured out if NAT is used and which.

TURN

traversal using relays around NAT

- TURN always UDP server and the peer
- TURN client/server UDP, TCP/TLS
- UPnP / NAT-PMP setup by the browser optional?
- TURN is a relay extension for STUN



ICE

Interactive Connectivity Establishment. ICE works by exchanging a multiplicity of IP addresses and ports, which are then tested for connectivity by peer-to-peer connectivity checks.

Connectivity

Encryption is mandatory for all WebRTC components:

- SRTP for Media, DTLS for Data, HTTPS for signaling

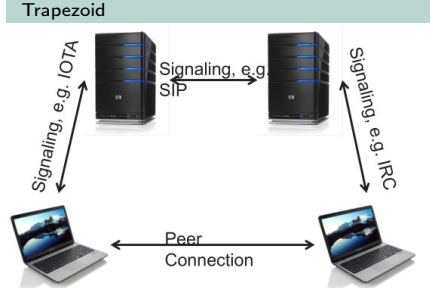
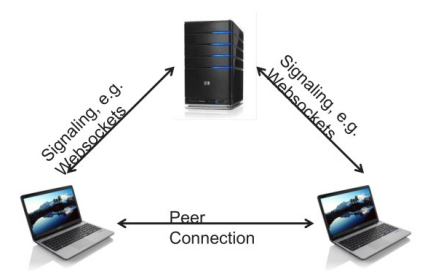
WebRTC is not a plugin.

- Camera and microphone access must be granted explicitly

Once connection is established - easy API

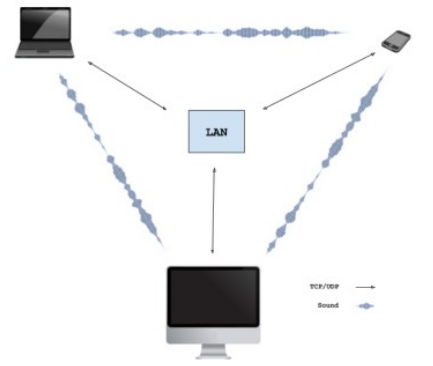
- `RTCPeerConnection.send("hallo")`
- `RTCPeerConnection.onmessage = function ...`

Triangle



Signaling PoC

A proof-of-concept for WebRTC signaling using sound. Works with all devices that have microphone + speakers. Runs in the browser:



Outlook

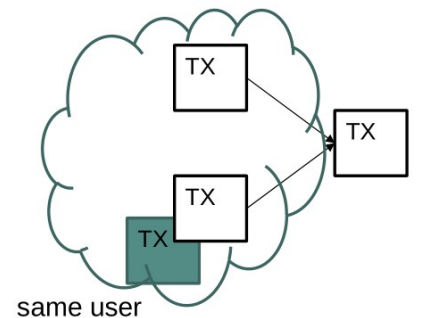
- Strong focus on VoIP.
- Fewer plugins (flash, java), fewer registrations
- New types of applications - Conferencing, gaming, P2P file sharing
- Object Real-Time Communications (ORTC) instead Session Description Protocol (SDP)

Anonymity

ein heuristisches Prinzip (Arbeitshypothese, vorläufige Annahme als Hilfsmittel der Forschung, Untersuchung, Erklärung)

HEURISTIC 1

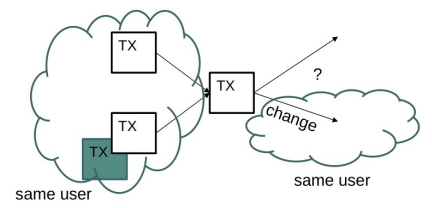
If two (or more) addresses are inputs to the same transaction, they are controlled by the same user; i.e., for any transaction t , all $pk \in \text{inputs}(t)$ are controlled by the same user.



HEURISTIC 2

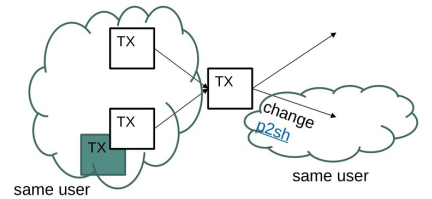
The one-time change address is controlled by the same user as the input addresses; i.e., for any

transaction t , the controller of $\text{inputs}(t)$ also controls the one-time change address $pk \in \text{outputs}(t)$ (if such an address exists).



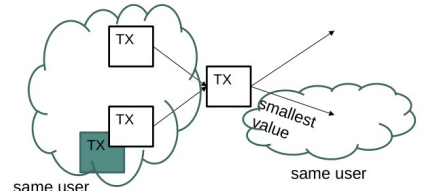
HEURISTIC 3

Multisig wallets usually use p2sh change, but the recipient rarely uses p2sh, which allows to determine the correct change output with high probability.



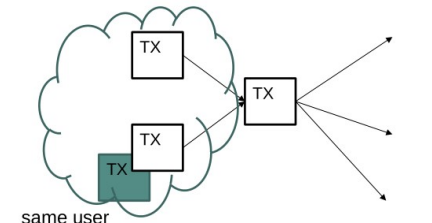
OPTIMAL CHANGE HEURISTIC

The assumption is that wallet software does not spend outputs unnecessarily. Therefore the change value is smaller than any of the spent outputs. Because if the change was larger than one output then this output would be left out and the change would be reduced by the output's value.



CONSUMER HEURISTIC

Consumer wallets only create transactions with two outputs. Therefore, if an output is spent by a transaction with 3 outputs it is not change.



Monero

- Confidential Block explorer (No Output or receiver)
- No scripting (no smart contracts)
- Proof of work algorithm (CPU Mining)
- Due to cryptographic algorithms, unknown: addresses trading monero, amounts, address balances, or transaction histories.
- Malware mining, due to CPU bound Proof of Work