# Introduction to the Blockchain

Mathematical Modeling Seminar Series

Daniel A. Nagy <daniel@ethereum.org>

# Economic Properties of Gold

- Scarce and desired
- Easy to verify, difficult to counterfeit
- Can only be obtained at cost
  - In exchange for goods and services
  - By extraction from nature (mining)
- Easily transferable
- Easily divisible and joinable
- Lasting value, stable price

# Rai stones of Yap

- PoW + information
- Tx history recited as oral tradition
- Tx by broadcast
- Stable currency for approx. 500 years
- Still in use
- Direct inspiration for Bitcoin

# Cryptographic Proof of Work

- PoW is $x$: $H(x) \in \mathbf{T}$, where $\mathbf{T} \Subset \text{range}(H)$
  - H is a one way fn (cryptographic hash)
- Properties of H
  - $|\mathbf{H}|=|\text{range}(H)|\cong 2^{256}$
  - domain(H)={0;1}*
  - For any $x$, easy to compute $H(x)$
  - For any value $h$ difficult to find $x$: $H(x)=h$
  - Difficult to find $x,y$: $x \neq y \wedge H(x)=H(y)$
- E(number of trials) = $|\mathbf{H}| / |\mathbf{T}|$ a.k.a. **difficulty**

# Bitcoin transactions

- Structure
  - Outputs (beneficiary definitions)
    - Address, i.e. H(*script*), *amount*
  - Inputs (if no inputs – “coinbase”)
    - H(*tx*), output *number*, *witness*
- Properties
  - $\sum amount_{\text{in}} \geq \sum amount_{\text{out}}$ (difference: *tx fee*)
  - One output can be spent **at most once**.

# Bitcoin Network, Mempool

- P2P broadcast network
  - Full nodes have complete tx history
  - Implicit state: UTXO set
  - Finite propagation time
  - Approx. 2500 full nodes and shrinking!
- Mempool
  - The first line of defense against cheating
  - No consensus; different realities possible

# The Blockchain

- A chain of block headers
- Block header $B_i$ structure
  - $H(B_{i-1})$ i.e. “chaining”
  - *Merkle root* of tx’s (with **one** *coinbase*)
  - Timestamp
  - Block number *i*
  - Difficulty (set for 10 min. exp. block time)
  - PoW “salt”

# Merkle Tree

- Authenticates a *sequence* of tx's: $t_1$, $t_2$, ... $t_n$
  - $H(t_1, t_2)$, $H(t_3, t_4)$, ...
    - $H(H(t_1, t_2), H(t_3, t_4))$, ...
      - ...
- Allows for $O(\log n)$ Merkle proofs of inclusion, given the *Merkle root*.
- Allows for authenticated streaming

# Block validity, consensus

- Block size < 1 Megabyte
- All transactions valid, given past
- 1 coinbase tx in each block
- Coinbase $\sum amount_{\text{out}} = \sum fee + reward_i$
- $reward_i = 50 / 2^{\lfloor i/131072 \rfloor}$
- Heaviest chain considered canonical
- Finality after 6 confirmations (1h on average)

# Implications

- Max. 3-4 tx/s (humanity: 100 000 tx/s)
- Tx fee charged for tx size (i.e. satoshi/byte)
- Monotonous miner revenue
- Optimization: NP-complete Packing Problem
- “Child pays for parent” heuristic policy
- Orphaned blocks, non-linear miner advantage
- Mining and validation separable
- Creeping centralization

# Evolution

- Forks
  - Soft forks (e.g. SegWit)
  - UASF (e.g. block size increase)
  - Hard forks
- Off-chain transactions
  - Payment channels
  - Lightning network
- Cross-chain transactions
  - Atomic swap

# Satoshi Nakamoto Missed

- Pooled mining
  - Slush pool
  - P2P pool
  - Pool hopping
- ASIC mining
- Full node extinction
- No trustless light clients
- Selfish mining, “34% attack”

# Successors

- Litecoin (50 tx/s)
- Monero (ring signatures, privacy)
- zCash (zkSNARKs, privacy)
- Ethereum (Turing-complete, ERC20, 15 tx/s)
  - MakerDAO (DAI stablecoin)
  - Serenity, a.k.a. Ethereum 2.0 (PoS, >1000 tx/s)

**Thank you!**

Questions?