Introduction to Blockchains for Computer Networking

Raj Jain
Washington University in Saint Louis
Saint Louis, MO 63130
Jain@wustl.edu

Audio/Video recordings of this lecture are available at:

http://www.cse.wustl.edu/~jain/cse570-19/

Overview

1. Trend: Centralized to Decentralized
2. Importance of Blockchain
3. Technical Innovations of Bitcoin
4. Blockchain Applications

Example of a Contract: Wedding

Wedding (Cont)

- Centralized

- Centralized registry
- Single point of failure
- Easier to hacked

- Decentralized

- Decentralized
- No single point of failure
- Very difficult to hack

Blockchains

- **What** it allows:
  - Two complete strangers can complete a transaction without a third party
  - 1st Generation: Transaction = Money transaction
  - 2nd Generation: Contracts, Agreements, Property, …
  - Revolutionizing and changing the way we do banking, manufacturing, education, computer networking, …
- **How** is it done?
  - A singly linked chain of blocks of verified signed transactions is replicated globally on millions of nodes
  - You will have to change millions of nodes to attack/change
- **Who** is interested in it: Banks, ISPs, Venture Capitalists, …
  ⇒ Researchers, students, …

Examples of Centralized Systems

- **Banks**: Allow money transfer between two accounts
- **Currency**: Printed and controlled by the government
- **Stocks**: Need brokers and clearing house (NY stock exchange, Bombay Stock Exchange, …)
- **Credit Card companies**
- In all cases:

1. There is a central third party to be trusted
2. Central party maintains a large database of information
   ⇒ Attracts Hackers
3. Central party may be hacked
   ⇒ affects millions
4. Central party is a single point of failure.
   Can malfunction or be bribed.

Trend: Centralized to Decentralized

- **Trend**: Make everything decentralized with no central point of control
- You can send money to your friends in Russia, China without their governments knowing it
- You can make a wedding contract, Property contract
- Decentralized systems are
  1. More reliable: Fault tolerant
  2. More secure: Attack tolerant
  3. No single bottleneck ⇒ Fast
  4. No single point of control ⇒ No monopoly ⇒ Cheaper
- Libertarians decided to build a totally decentralized system with no central authority. Blockchain is one way to do this.

Bitcoin

- First Successful Virtual Currency
- Has survived 9 years and has become legal in several jurisdiction
- Decentralized: No one company or government controls it
  - Decentralized Transaction Verification
  - Decentralized Ledger (accounting book)
  - Decentralized Mint to make new coins
  - Decentralized peer-to-peer network
- Has been designed to control over-minting, double-spending, counterfeiting
- 1 BTC = 8473.34 USD (November 17, 2019)
- 10^{-8} BTC = 1 Satoshi = 0.00008 cents
- 18 Million BTC (November 17, 2019)
- Total 21 Million BTC will ever be generated.

Ref: https://coinmarketcap.com/

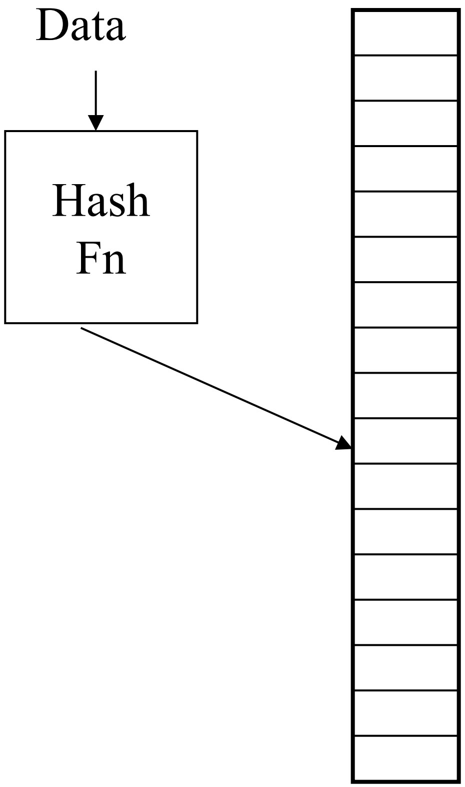
Bitcoin History

- Satoshi Nakamoto published a *whitepaper* in 2008. How to do direct transfer of money without involving a 3rd party.
- He also published complete reference code to transact, store, and mint Bitcoins. Made the software open source.
- He supported the software and answered all questions for 3 years and then disappeared
  (may be because he was rich or fearful)
- P2P Network:
  - Nodes come up and leave at random
  - Packets are delayed, lost, duplicated
  - Some nodes are malicious
- As long as a majority of CPU power is not with attackers, the system works ⇒ Proof of Work

Ref: Satoshi Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," https://Bitcoin.org/Bitcoin.pdf

Hash Function

- Hash tables used in data searches
- The hash function should
- Take variable size input
- Produce fixed output size (Size of the table)
- Be easy to compute
- Be pseudorandom so that it distributes uniformly over the table ⇒ Minimizes collisions
- Deterministic: Same input always produces the same hash
- Example: h(M) = M mod 9; M=13 ⇒ h(M)=4

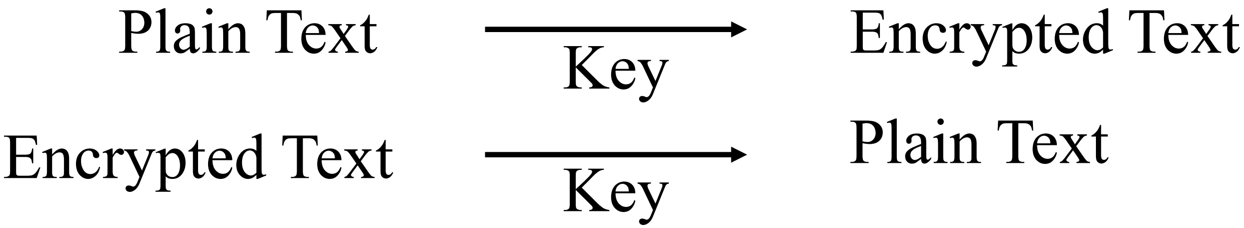


Cryptographic Hash Functions

- One-way
  It is not possible to find any M, given h.
- Very Very difficult to compute M given h(M)
- SHA-2: Secure Hash Algorithm standardized by National Institute of Standards and Technology (NIST).
  - SHA-256 produces a 256-bit hash of any number
- RIPEMD: RACE Integrity Primitive Evaluation developed in EU
  - RIPEMD160 produces 160-bit hash

Secret Key Cryptography

- Secret Key Cryptography:



Plain Text $\xrightarrow[\text{Key}]{}$ Encrypted Text

Encrypted Text $\xrightarrow[\text{Key}]{}$ Plain Text

- Key must be kept secret.
  Anyone with a key can read/write/change messages

Public Key Encryption

- Invented in 1975 by Diffie and Hellman at Stanford
- Encrypted_Message = Encrypt(Key1, Message)
- Message = Decrypt(Key2, Encrypted_Message)

$$\text{Text} \xrightarrow{\text{Key1}} \text{Ciphertext} \xrightarrow{\text{Key2}} \text{Text}$$

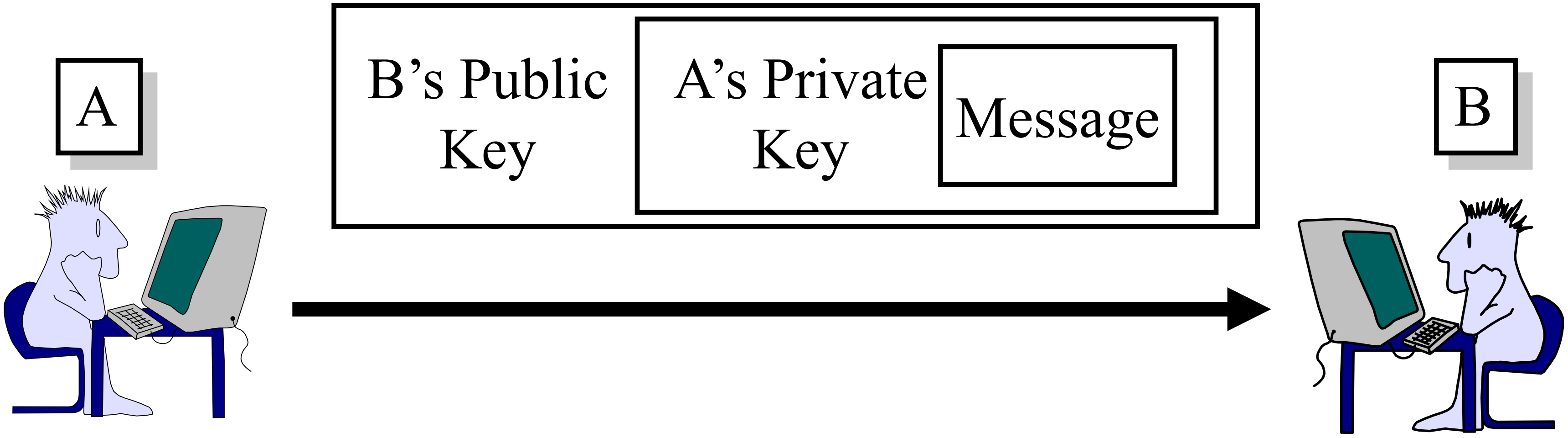
- Keys are **interchangeable**:

$$\text{Text} \xrightarrow{\text{Key2}} \text{Ciphertext} \xrightarrow{\text{Key1}} \text{Text}$$

- One key is made **public** while the other is kept **private**
- Sender knows only public key of the receiver ⇒ **Asymmetric**

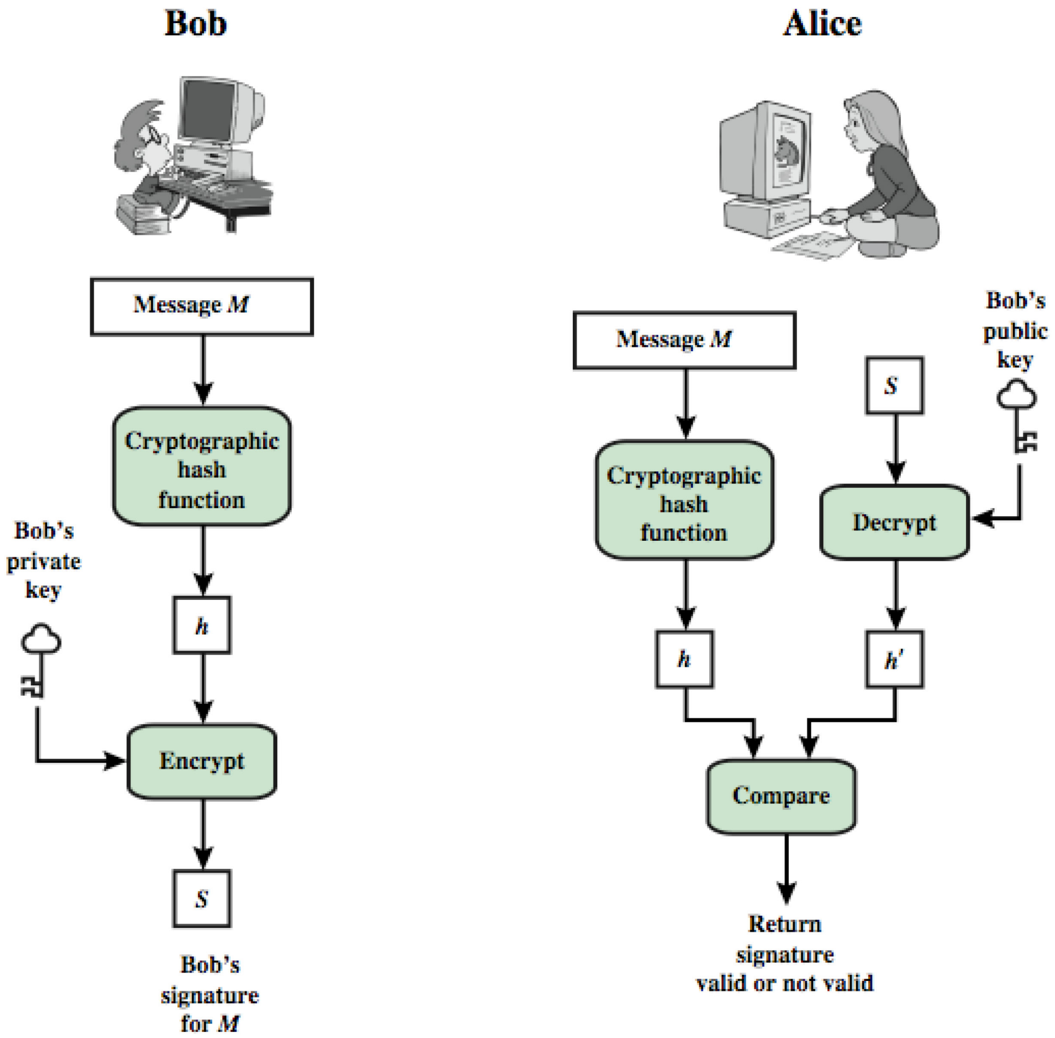
Ref: http://en.wikipedia.org/wiki/Public-key_cryptography

Public-Key Authentication and Secrecy



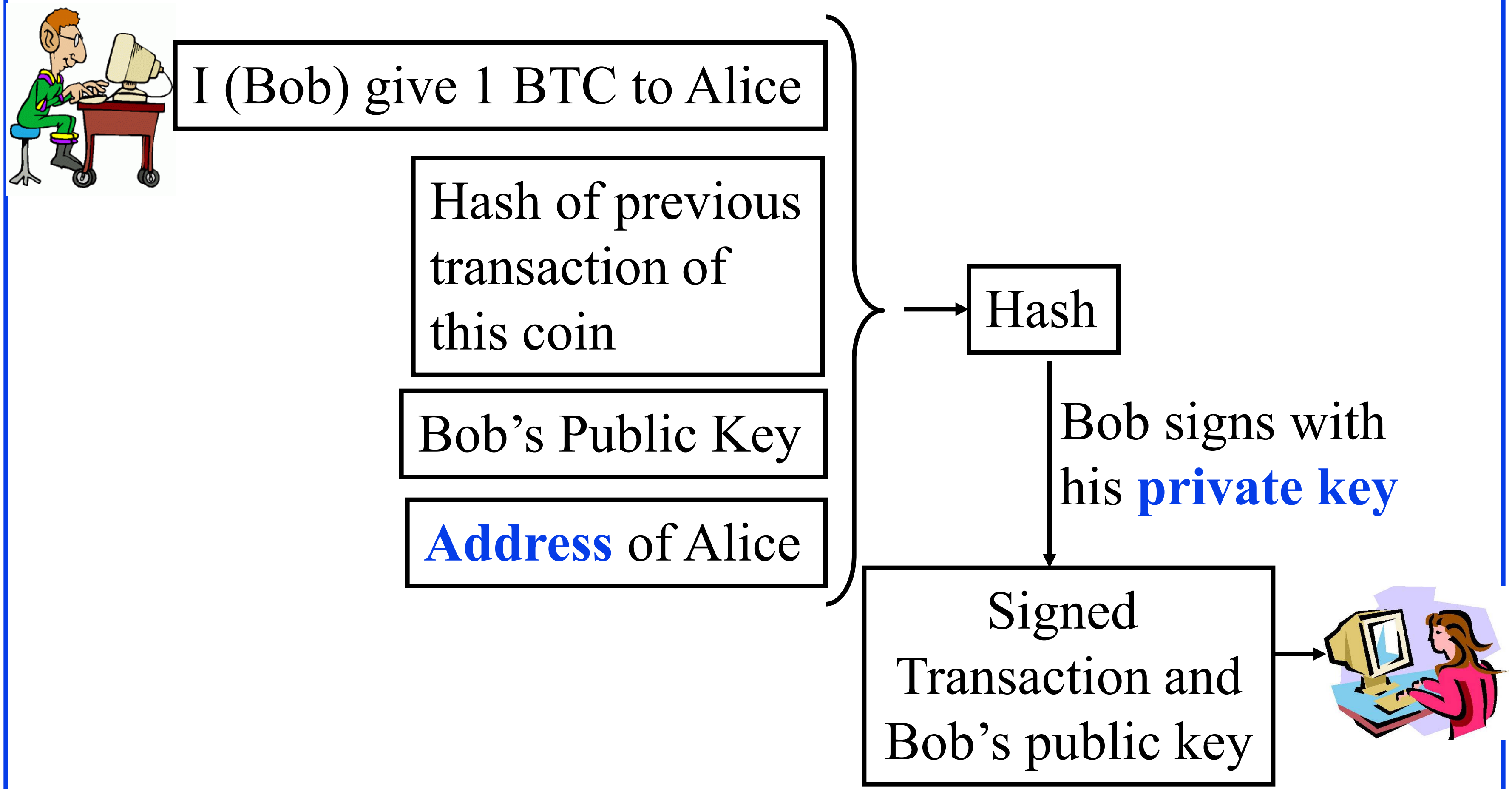
- A encrypts the message with its private key and then with B's public key
- B can decrypt it with its private key and A's public key
- No one else can decrypt ⇒ Secrecy
- No one else can send such a message
  ⇒ B is assured that the message was sent by A
  ⇒ Authentication

Digital Signature



Transaction

- ❑ Bob gives 1 BTC to Alice



I (Bob) give 1 BTC to Alice

Hash of previous transaction of this coin

Bob's Public Key

Address of Alice

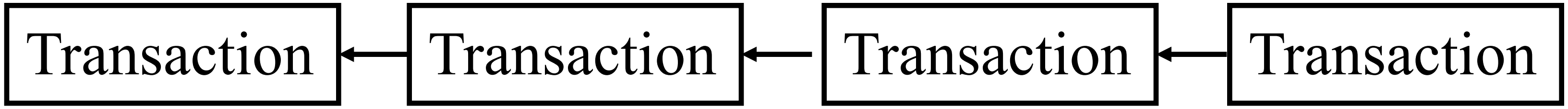
Hash

Bob signs with his **private key**

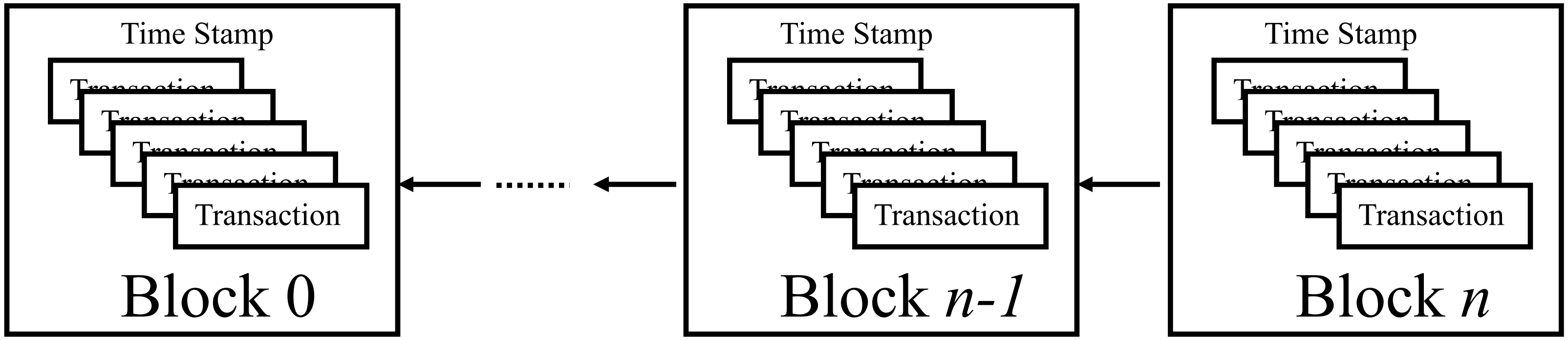
Signed Transaction and Bob's public key

Blocks

- Transaction Chain:

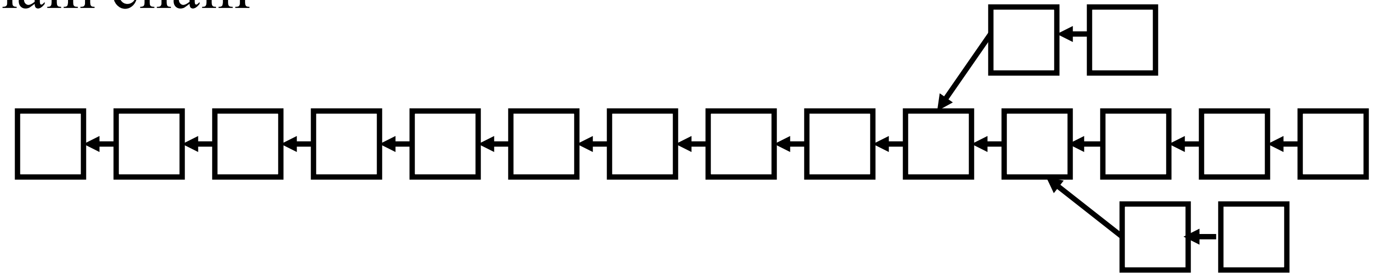


- Problem:
  - Too many transactions ⇒ Chain too long
  - Takes too long to find and verify a transaction
- Solution: Combine several transactions into blocks of verified transactions



Blockchains

- Block maker (Miners) ensures that all transactions in the block are valid
- Miners have significant computing power
- Miner with the highest computer power wins.
  His/her block is added to the end of the chain
- Miner is rewarded.
  He/She is allowed to mint a few new coins and keep them
- Proof of computing power ⇒ **Proof of work**
  ⇒ Solve a puzzle
- Chain with the highest cumulative difficulty is selected as the main chain



Bitcoin Address

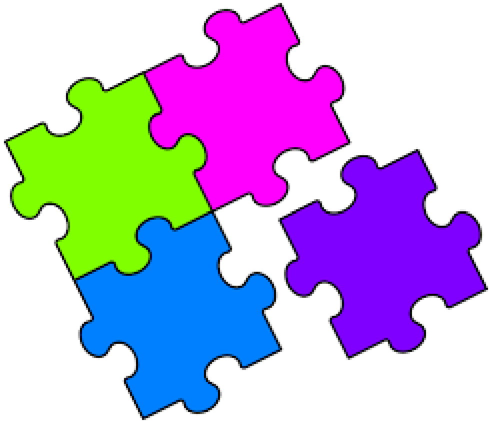
- Addresses=RIPMD160(SHA-256(Public Key))
- Addresses are encoded with Base-58 encoding
  (10 digits + 26 uppercase + 26 lower case – 4 (0, O, 1, I)
  that is, lower case L, and upper case I
- Base58 Check Encoding: 4-byte checksum is appended.
  Checksum=First 4 bytes of SHA256(SHA256(Prefix+Data)
- Prefix is 0x00 = version
- After encoding a 1 is added to indicate that it is an address
- Always start with 1
- Generally presented as a QR Code

Pseudo-anonymous

- Using a nonce, you can generate a new public/private key pair
- RIPMD160 of SHA-256 hash of the public key is your address
- All transactions are between two addresses
- You can have as many addresses as you like
- You do not need to disclose your name, ID, or physical address ⇒ Pseudo anonymous
- If a transaction touches the physical world, your identity is disclosed, e.g., when buying your first Bitcoin with your credit card

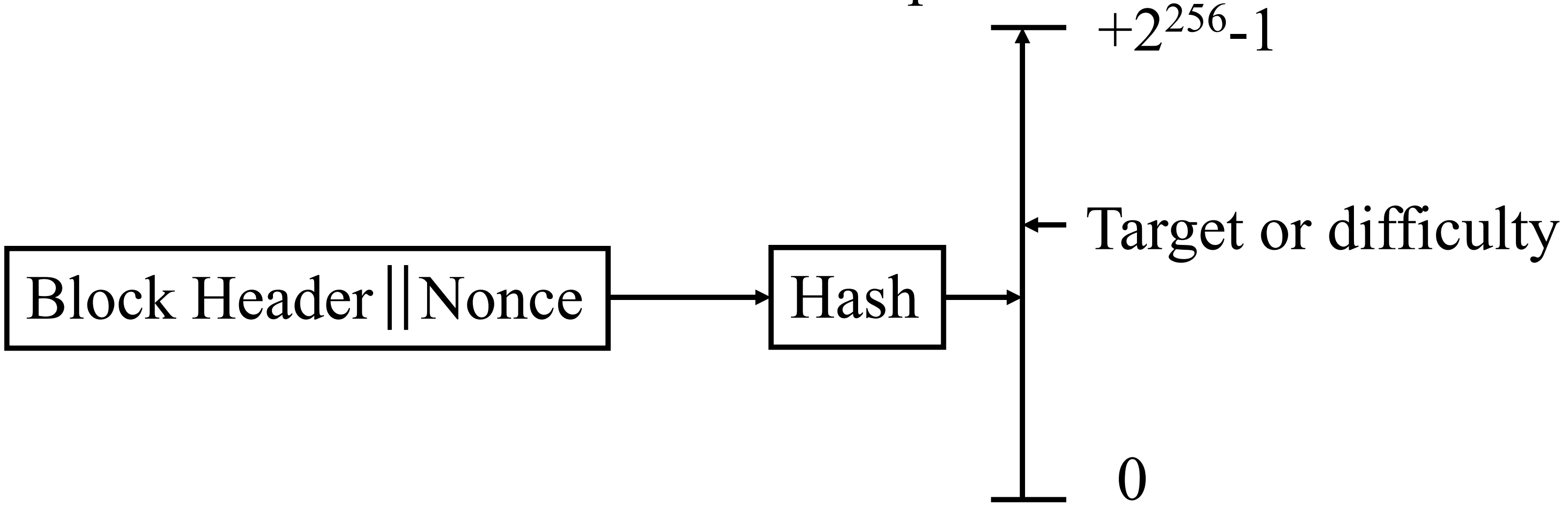
Proof-of-Work

- When someone requests a service, ask them to do something that is difficult for the requester but easy to verify for the server. Captcha is one example
- Bitcoin requires a proof that you can compute faster than others
- A puzzle is given and the node that solves it first wins
- Puzzle is such that it can be solved in ~ 10 minutes
  ⇒ Puzzles are being made harder as the computing power is increasing with Moore's Law



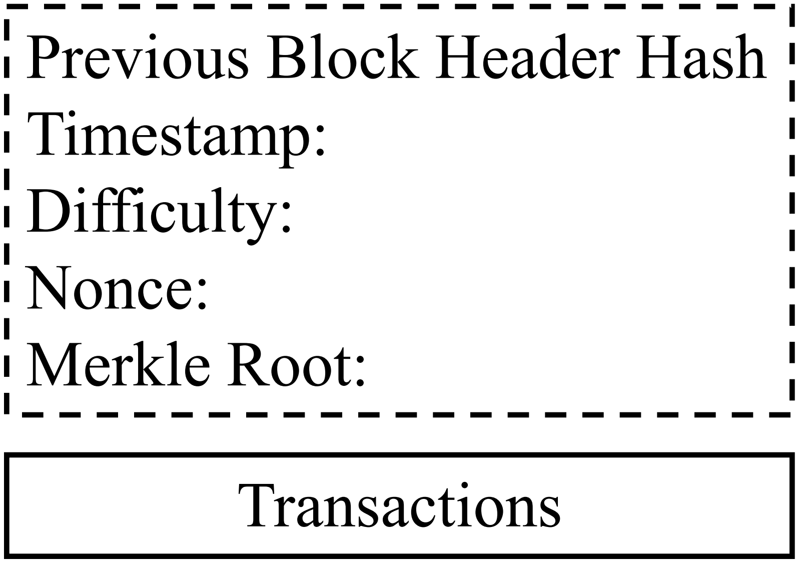
Puzzle

- Find a nonce that will make the hash of the block header less than a specified target
- Lower target ⇒ More difficult to find
- Puzzle can be made harder/easier by specifying a higher or lower target
- Target is adjusted by all miners every 2 weeks (2016 blocks) so that it takes 10 minutes to solve the puzzle.



Block Structure

- Block header contains a double-hash of the previous block header, a hash of the root of the Merkle tree of transactions in the block, a time stamp, difficulty, nonce



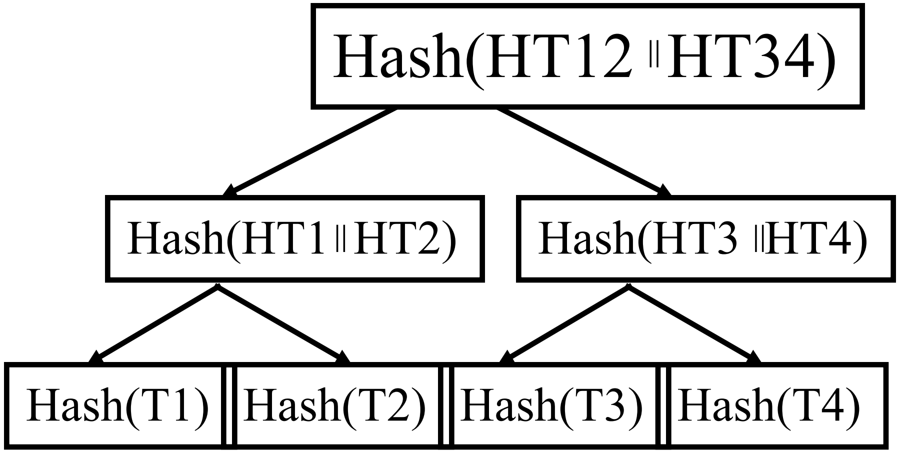
Previous Block Header Hash
Timestamp:
Difficulty:
Nonce:
Merkle Root:

Transactions

Ref: A. M. Antonopoulos, "Mastering Bitcoin," O'Reilly, 2015, 274 pp.

Merkle Tree

- A Binary hash tree to efficiently summarize and verify the integrity of large sets of data
- Hashes of the transactions are stored in the tree
- Parents contain hash of the concatenation of children
- Takes $\log_2(n)$ comparisons to find the transaction among n



Ref: A. M. Antonopoulos, "Mastering Bitcoin," O'Reilly, 2015, 274 pp.

Smart Property

- ❑ Bob: I give $100 to Alice if IBM stock goes below $5
  - ➢ Locking script: if IBM stock < $5 Return True
  - ➢ Unlocking script: IBM stock price is $4
- ❑ Property exchange happens if certain conditions are satisfied. Conditions can be checked automatically ⇒ Allows trustless exchanges
- ❑ **Smart Contracts**: Not just buy/Sell. Any agreement.

Potential Blockchain Applications

- **Financial**: Currency, Private equities, Public equities, Bonds, Derivatives, Commodities, Mortgage records, Crowd-funding, Micro-finance, Micro-charity
- **Public Records**: Land titles, Vehicle registries, Business license, Criminal records, Passports, Birth certificates, Death certificates, Building permits, Gun permits
- **Private Records**: Contracts, Signatures, Wills, Trusts, Escrows
- **Other Semi-Public Records**: Degree, Certifications, Grades, HR records, Medical records, Accounting records
- **Physical Asset Keys**: Apartment keys, Vacation home keys, Hotel room keys, Car keys, Rental car keys, Locker keys
- **Intangibles**: Patents, Copyrights, Trademarks

Ref: http://ledracapital.com/blog/2014/3/11/Bitcoin-series-24-the-mega-master-blockchain-list

Networking Applications of Blockchains

- Multi-Domain Systems:
  - Multiple Cloud Service Providers
  - Multiple cellular providers
  - Multi-Interface devices: WiFi, Cell, Bluetooth, …
  - BGP: BGP Authentication
- Globally Centralized Systems:
  - DNS
  - Certificate Authorities

Explore blockchains for multi-domain/centralized systems

Networking Applications (Cont)

- Public Key Infrastructure
  - Certificate Authorities issue certificates
  - Single Point of Failure
  - Diginotar – Dutch certificate authority was compromised in 2011)
- **NameCoin**: A decentralized key-value registration and transfer platform using blockchains.
  - A decentralized **Domain Names Registry**
  - .bit domain names

Ref: T. Salman, et al, "**Security Services Using Blockchains:A State of the Art Survey**" IEEE Communications Surveys and Tutorials, First Quarter 2019, Volume 21, Issue 1, 858-880 pp., http://www.cse.wustl.edu/~jain/papers/bcs.htm

Public Key Infrastructure

- Certificate Authorities issue certificates
  - Single Point of Failure
  - CA Keys are often compromised (Diginotar – Dutch certificate authority was compromised in 2011)
- Web of Trust: Anyone can issue a certificate
- Blockchain solution: Store user ID and public key
  - Blockstack
  - Certcoin

Data Provenance

- Keeping track of origin and history of movement of data among the databases or documents
- Traditional solution: Logging and auditing
- In a distributed cloud environment, centralized logging is required and is difficult
- Blockchains can be used to log the changes Miners verify the changes
  - ProvChain
  - SMARTDATA
- Also used in supply chains

Data Privacy

- Facebook and Google have massive amounts of personal information
- Who can access this information?
- Can someone do statistics on the database without having rights to personal information of all?
- Can the user hide its identity?
- Traditional Method: Access Control Lists (ACL) managed centrally (by Facebook and Google)
- Blockchains can be used to keep ACL and data stored in a distributed manner with no central control

Data Integrity

- Data has not been corrupted
- Traditional techniques: Digital Signatures and PKI, Replication
- In blockchains, data can not be tempered once committed to a block.
- Ericson provides a blockchain based integrity assurance service

Blockchain Challenges

- **Selfish mining**: Some one creating a large number of bad blocks keeping the miners busy with discards
- **Sybil Attacks**: Some one creating a large number of transactions denying service to legitimate users
- **51% Attack**: One entity owns the majority of miners
- Communication overhead
- Solving the puzzles for "Proof of Work" wastes computing resources

Alternatives to "Proof of Work"

- **Proof of Space**: Computation is replaced by storage
- **Measure of Trust**: Most trustworthy miner wins
- **Minimum Block Hash** (rather than fastest) miner wins ⇒ More random
- **Proof of Importance**
- **Proof of Stake**

Blockchain Implementations

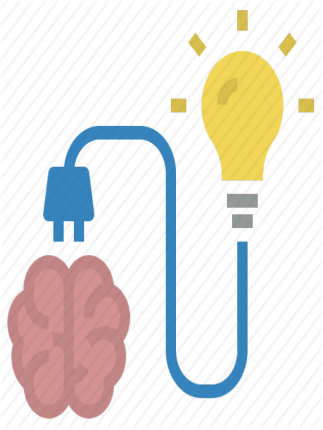
- **Open Source Implementations**:
  - Bitcoin
  - Etherum
  - Hyper Ledger
- **Commercial Implementations**: Block Chain as a Service from
  - IBM
  - Microsoft Azure
  - SAP
  - Deloitte

Key Strengths of Blockchains

1. **Distributed**: No single point of failure
2. **Decentralized Consensus**: Transactions valid only if agreed by majority
3. **Trustless**: Transacting or processing parties do not need to trust
4. **Cryptographic Security**: Elliptic Curve Cryptography
5. **Non-Repudiation Guarantee**: All transactions are signed

Ideas to Enhance Blockchains

- Blockchain is just a distributed **data storage** of valid transactions
- All transactions are ***deterministic***
- What's Wrong?
  - Need to convert data to knowledge
  - Real life is probabilistic
  - Most decisions we make are probabilistic ⇒ All decisions have some risk



Risk Propels Progress

- Banks take money from risk-averse savers and give them interest
- Banks invest the money in corporations ⇒ Takes the country forward
- Venture capitalists take risk by investing in half-cooked ideas
- Startups take risk by working in unchartered territories

Decisions with Risk

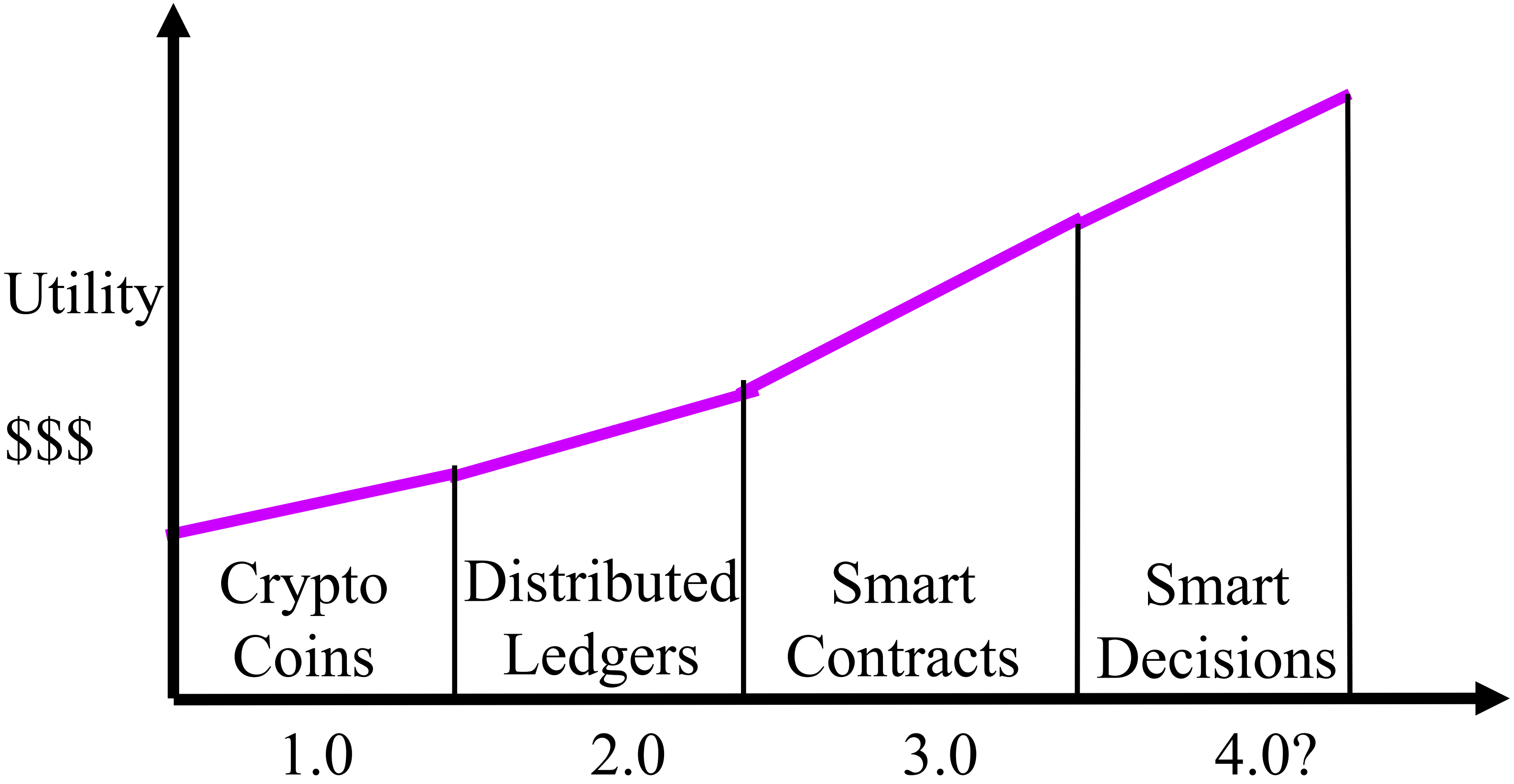
- Sell insurance
- Buy insurance
- Sell a stock
- Buy a stock
- Download a software application on your computer
- Update software

Our Goal

- Moving the chain from deterministic to **probabilistic**
- Moving the chain from storage to **computation**
- Moving the chain from data to **knowledge**
- Moving the chain from information to **decision making**

- Google is moving from "Search" to "Suggest" using AI
- A blockchain that provides knowledge
  - A **knowledge chain** would be more useful

Blockchain Generations



Can the Blockchains be Enhanced?

Limitation 1: Only facts are recorded

- Alice gave 20 coins to Bob

Limitation 2: Binary Validity

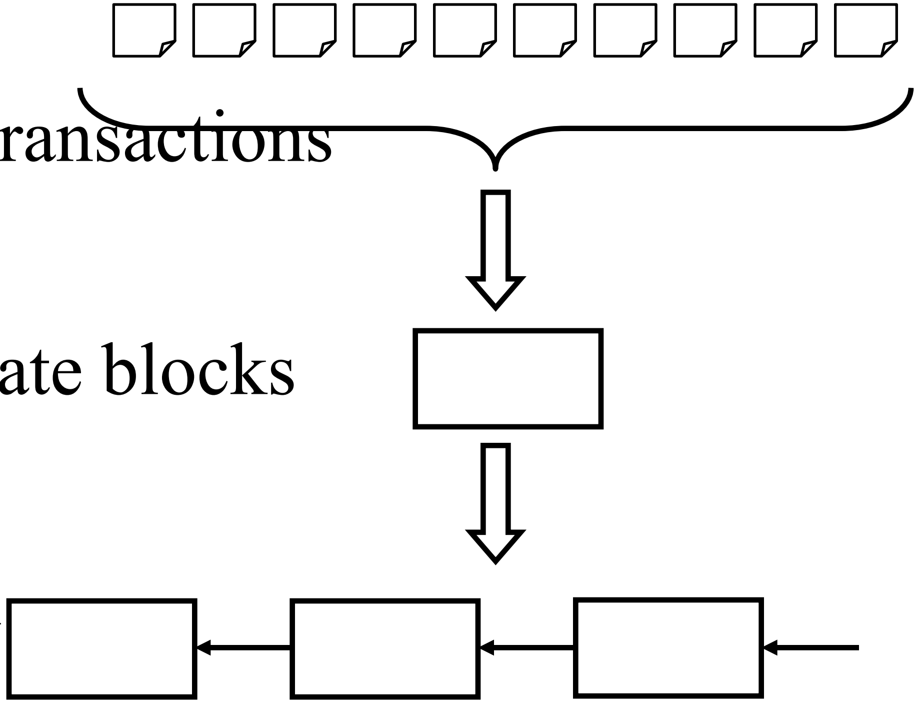
- All transactions/contracts recorded on the blocks that are committed are valid
- Those not on the committed blocks and old are invalid
- So the recording is binary: only 0 or 1.

Limitation 3: Deterministic Events only

- Can not record that I am only 90% sure that Alice gave 20 coins to Bob.

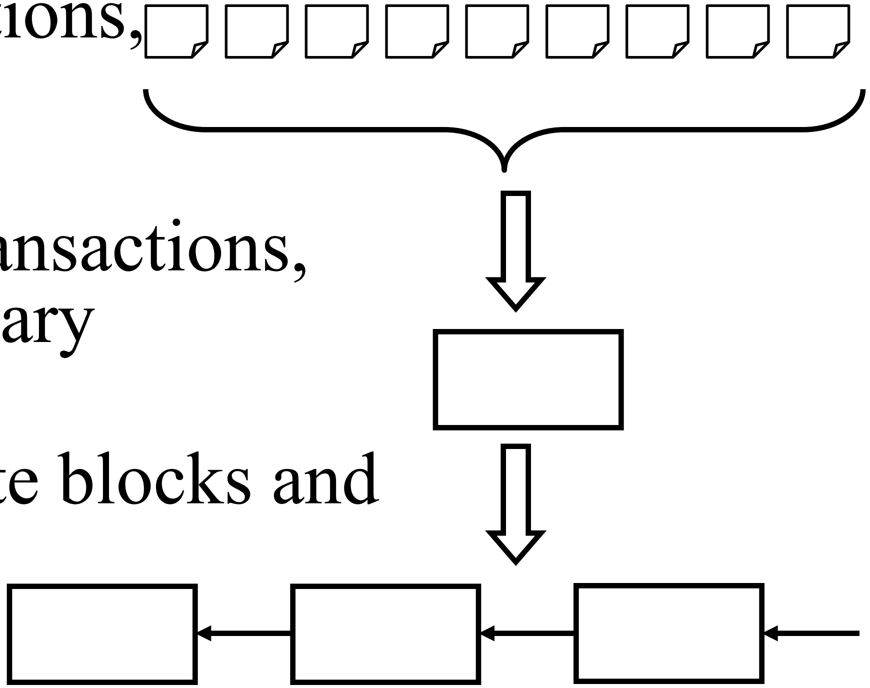
Current Blockchain Process

1. **Users** broadcast transactions or smart contracts
2. **Mining nodes** validate transactions and create blocks
3. **Blockchain nodes** validate blocks and construct a chain



- There are many users, many mining nodes, and many blockchain nodes.
- More nodes ⇒ Better.
  Less ⇒ Blockchain not required/useful.

Probabilistic Blockchain Process



1. **Agents** broadcast transactions, Transactions = Opinions/decisions
2. **Mining nodes** validate transactions, create a knowledge summary and create blocks
3. **Blockchain nodes** validate blocks and construct a chain
4. Two types of users:
   - **Agent nodes** provide their probabilistic decisions
   - **Management nodes** that inquire the blockchain and use it for group decisions

Blockchain 4.0: Database to Knowledge Base

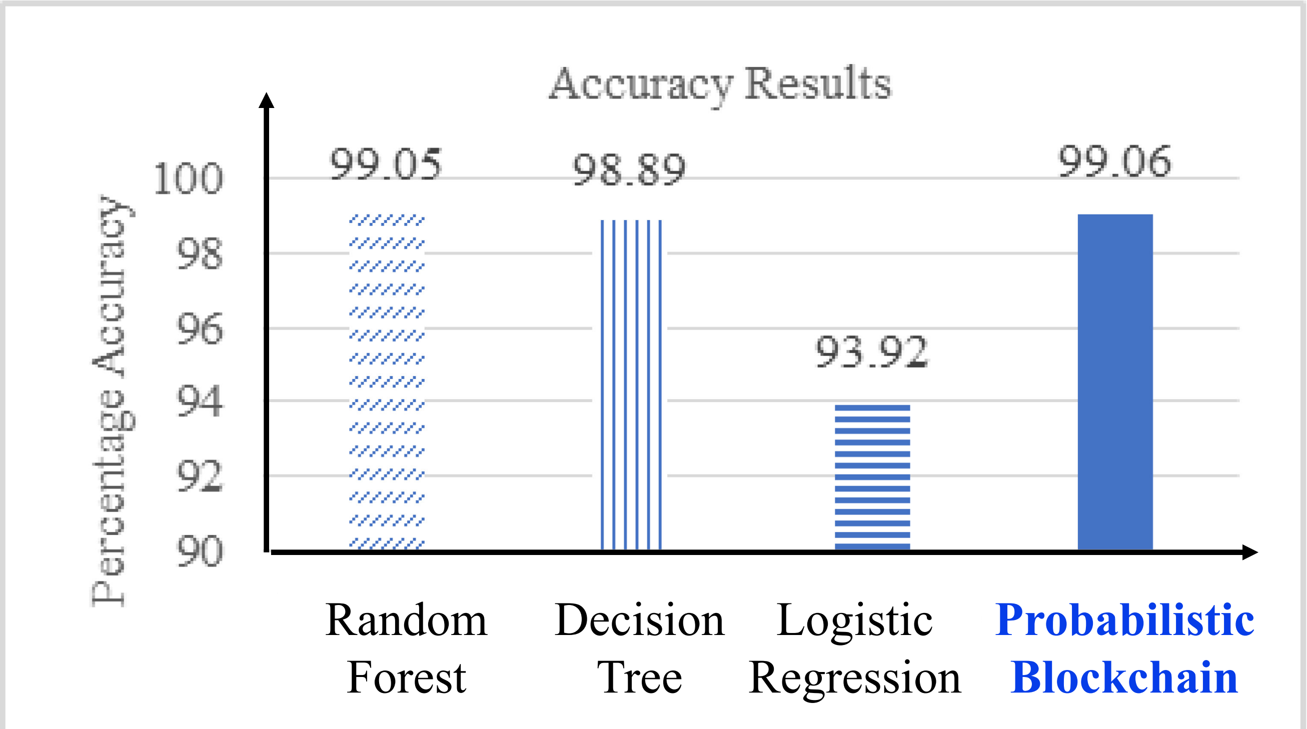
- Blockchain = Distributed database of smart contracts
- Probabilistic blockchain = Knowledge + database
- Database = Who bought, who sold, what quantity, what price, what time
- Knowledge =
  - Where the market is going?
  - Whether we should buy, sell, or hold?

Empirical Validation

- Issue: Whether a network traffic pattern represents intrusion
- 1000 Agents using different machine learning algorithms give their decisions: Yes or No
  - Agents randomly pick one of the 3 algorithms:
    - Random Forest, Decision Tree, Logistic Regression
- Mining nodes summarize these decisions using the majority function

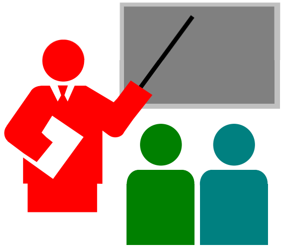
Results

$$\text{Accuracy} = \frac{\text{Correct Predictions}}{\text{Overall Samples}} \times 100\%$$



Distributed decision making is better than any individual decision

Summary



1. Current trend is to make everything decentralized
2. Bitcoin is a decentralized currency.
3. Blockchain 1.0 is used to global consensus on Bitcoin transactions.
4. Blockchain 3.0 allow sophisticated contracts making it useful for many network and security applications
5. Probabilistic Blockchains allow probabilistic statements to make decisions under risk.

Reading List

- Koshik Raj, "Foundations of Blockchain," Packt Publishing, January 2019, ISBN: 9781789139396 (Safari Book)
- Tara Salman, Raj Jain, and Lav Gupta, "**Probabilistic Blockchains: A Blockchain Paradigm for Collaborative Decision-Making**," 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON 2018), New York, NY, November 8-10, 2018, 9 pp., http://www.cse.wustl.edu/~jain/papers/pbc_uem.htm
- Tara Salman, Maede Zolanvari, Aiman Erbad, Raj Jain, and Mohammed Samaka, "**Security Services Using Blockchains:A State of the Art Survey**" IEEE Communications Surveys and Tutorials, Accepted September 2018, 28 pp., http://www.cse.wustl.edu/~jain/papers/bcs.htm

Other Readings

- A. M. Antonopoulos, “Mastering Bitcoin,” Oreilly, 2014, 272 pp. (Safari Book)
- A. Lewis, “The Basics of Bitcoins and Blockchains,” Mango Publishing, 2018, 408 pp.
- A. Narayanan, J. Bonneau, E. Felten, A. Miller, S. Goldfeder, “Bitcoin and Cryptocurrency Technology: A Comprehensive Introduction,” Princeton University Press, 2016, 304 pp.

Online Resources

- CoinDesk: Bitcoin News, Prices, Charts, Guides & Analysis, http://www.coindesk.com/
- Bitcoin magazine, https://bitcoinmagazine.com/
- CCN: Bitcoin, Blockchain, FinTech, & Cryptocurrency News, https://www.cryptocoinsnews.com/
- CoinTelegraph, https://cointelegraph.com/
- Bitcoin Stack Exchange, http://bitcoin.stackexchange.com/
- Let's talk Bitcoin, https://letstalkbitcoin.com/
- Epicenter - Weekly Podcast on Blockchain, Ethereum, Bitcoin and ..., https://epicenter.tv/
- Epicenter Bitcoin, https://epicenter.tv/
- Ethercasts, https://www.youtube.com/user/EtherCasts

Acronyms

API	Application Programming Interface
BTC	Bitcoin
CCN	Crypto Coin News
DARPA	Defense Advanced Research Project Agency
HR	Human Resources
ICANN	Internet Committee for Assigned Names and Numbers
ID	Identifier
IoT	Internet of Things
IPFS	Internet Protocol File System
ISP	Internet Service Provider
QR	Quick Response Code
RFP	Request for Proposal
RIPEMD	RACE Integrity Primitives Evaluation Message Digest
SHA	Secure Hash Algorithm
USD	United States Dollar
VC	Venture Captial

Scan This to Download These Slides

Raj Jain
http://rajjain.com

Related Modules

CSE571S: Network Security (Spring 2017),
http://www.cse.wustl.edu/~jain/cse571-17/index.html

CSE473S: Introduction to Computer Networks (Fall 2016),
http://www.cse.wustl.edu/~jain/cse473-16/index.html

Wireless and Mobile Networking (Spring 2016),
http://www.cse.wustl.edu/~jain/cse574-16/index.html

CSE571S: Network Security (Fall 2014),
http://www.cse.wustl.edu/~jain/cse571-14/index.html

Audio/Video Recordings and Podcasts of
Professor Raj Jain's Lectures,
https://www.youtube.com/channel/UCN4-5wzNP9-ruOzQMs-8NUw