Extending Blockchains with AI for Risk Management

Raj Jain
Barbara J. and Jerome R. Cox, Jr. Professor
Washington University in St. Louis
jain@wustl.edu

Scan this to download these slides

ACM Student Chapter Meeting, Washington University in St. Louis
October 19, 2021

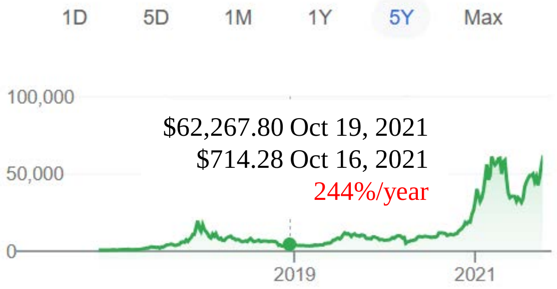
Slides and Video Recording of this talk are at:
http://www.cse.wustl.edu/~jain/talks/pbc_acm.htm

Overview

1. Strength of Blockchains
2. Weaknesses of the blockchains
3. Extending Blockchains: Converting data to knowledge
4. Applications of Knowledge Chains

What is a Blockchain?

1. Satoshi Nakamoto invented Bitcoin
2. He used blockchains to make it decentralized
3. Since then blockchains have found numerous other applications
4. Blockchains allow two complete strangers to enter into a smart contract without a trusted third party.
5. This talk is about blockchains, <u>not</u> about Bitcoin.



$62,267.80 Oct 19, 2021
$714.28 Oct 16, 2021
244%/year

Example of a Contract: Wedding

Example of a Contract: Wedding

Centralized Trust

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

Decentralized Trust

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

Examples of Centralized Systems

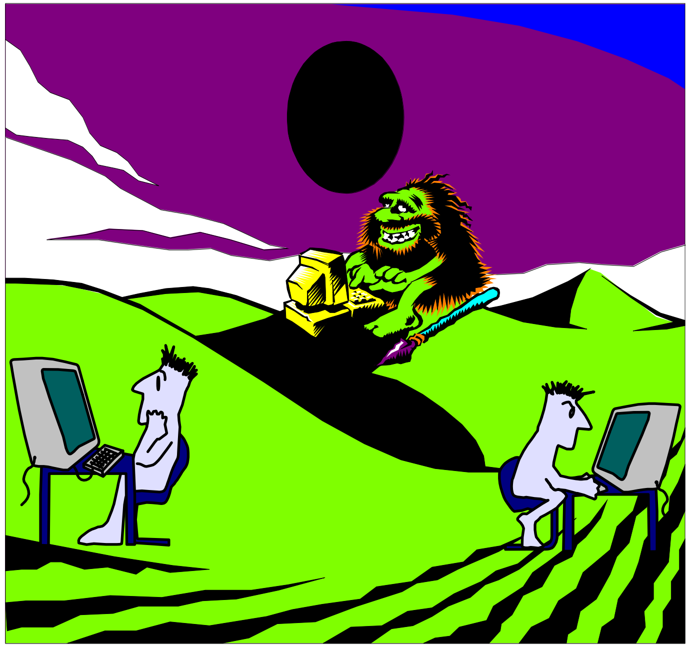
- **Banks**: Allow money transfer between two accounts
- **City Records**: Wedding registers, Property ownership
- **Networks**: Certificate Authorities, DNS

- In all cases:
  - There is a central third party to be trusted
  - Central party maintains a large database ⇒ Attracts Hackers
  - Central party may be hacked ⇒ Affects millions
  - Central party is a single point of failure.
    Can malfunction or be bribed

Trend: Decentralized ⇒ Centralized ⇒ Decentralized

Decentralized

Industrialization ⇒ Centralized



COVID ⇒ Decentralized

Time is a cycle: Decentralized vs. Centralized debate

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

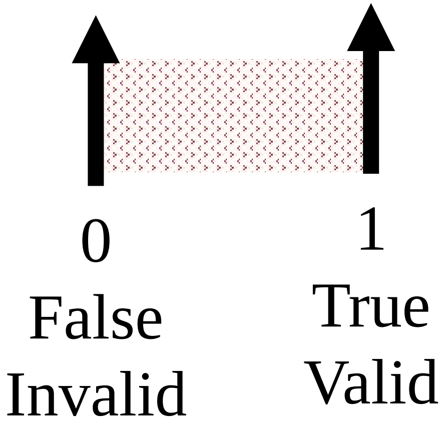
Can the Blockchains be Enhanced?

Limitation 1: Only facts are recorded

- Alice is married to Bob
- Alice gave 20 coins to Bob
- Alice signed a contract with Bob to pay 10 coins for 1 kg of xx.

Limitation 2: Binary Validity

- All transactions 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.

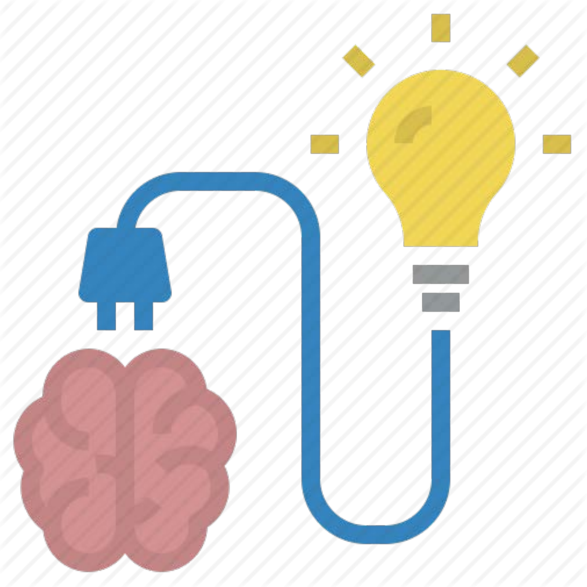


Limitation 3: Deterministic Events only

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

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
  - ➢ We are in big data and machine learning age
  - ➢ Real life is probabilistic
  - ➢ Most to the decisions we make are probabilistic
    ⇒ All decisions have some risk



Decisions with Risk

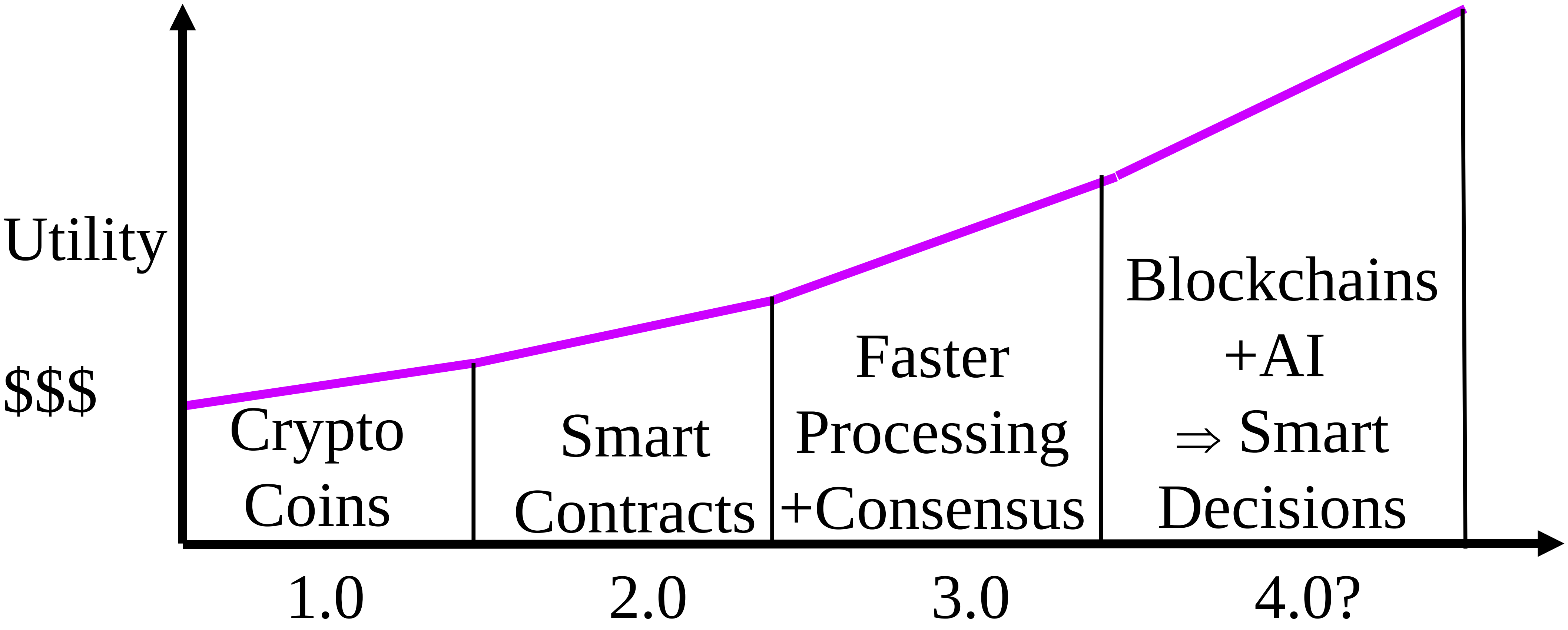
- Sell insurance
- Buy insurance
- Sell a stock
- Buy a stock
- Download a software application on your computer
- Update your computer
- Marry someone

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

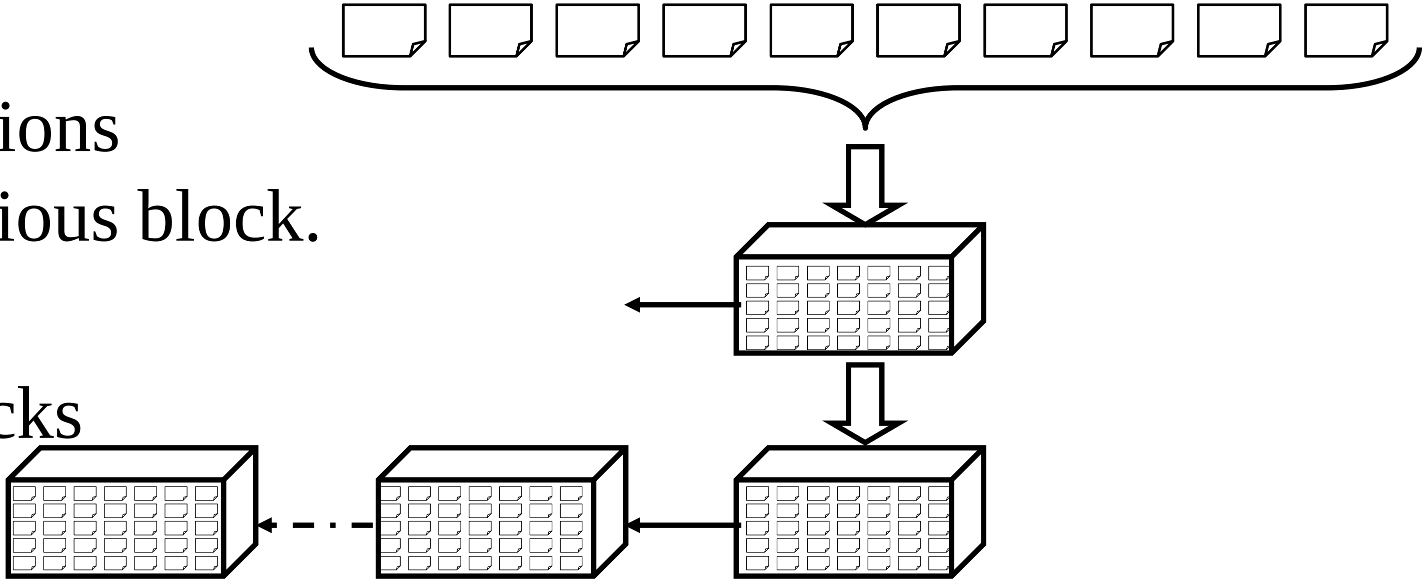


Blockchain Process

1. **Users** broadcast signed transactions or smart contracts

2. **Mining nodes** validate transactions and create blocks. Point to previous block.

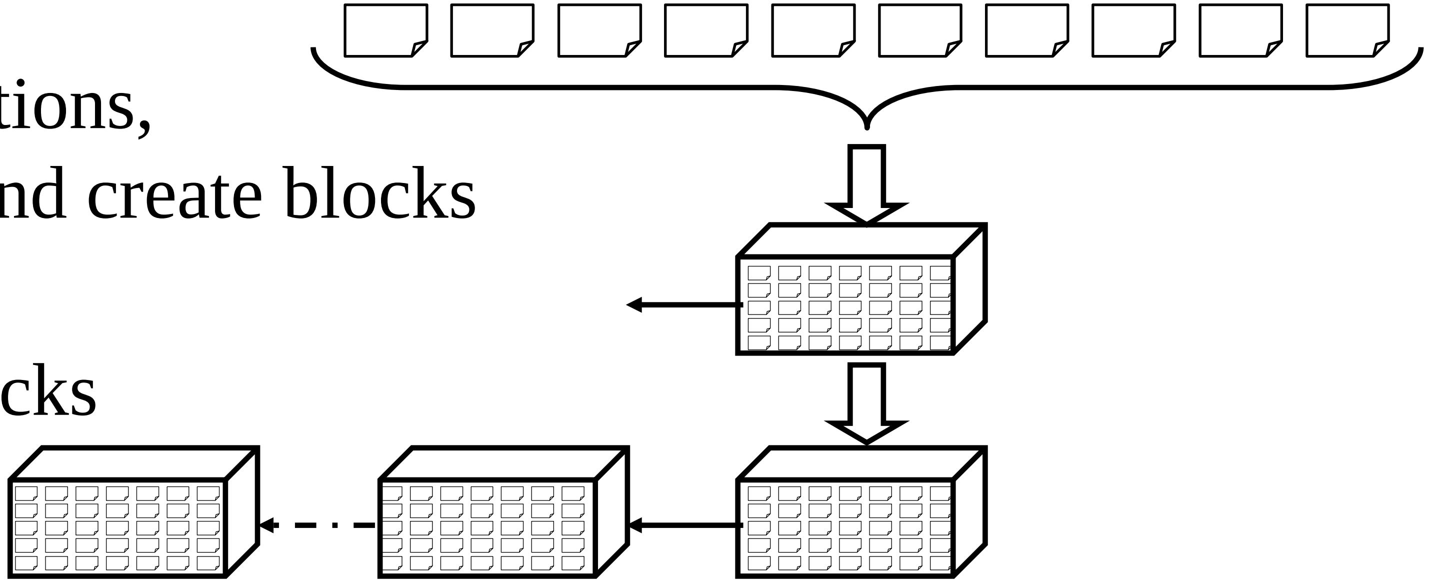
3. **Blockchain nodes** validate blocks and construct a chain



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

Knowledge Chain /Probabilistic Blockchain

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



- ❑ Two types of users:
  - ➢ **Agent nodes** provide their probabilistic opinions/decisions
  - ➢ **Management nodes** that inquire the blockchain and use it for group decisions

Knowledge Chain Example

- ❑ **Issue**: Whether Cisco stock will go up tomorrow?
- ❑ i^{th} Agent says that the probability that it will go up is p_i
- ❑ Summary of all opinions related to this issue is:

$$\text{P[Stock will rise]} = \text{G}(\{p_1, p_2, \ldots, p_n\})$$

Here, G = Machine Learning Algorithm = Summarizing function

Ref: T. Salman, R. Jain, and L. Gupta, "**Probabilistic Blockchains: A Blockchain Paradigm for Collaborative Decision-Making**," 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON 2018), New York, NY, Nov. 8-10, 2018, 9 pp., http://www.cse.wustl.edu/~jain/papers/pbc_uem.htm

Generalizing the Summary Function

- Summary can be any other reasonable function of individual decisions:
  - 90-percentile
  - Median
  - Mode
  - 2nd Moment
- Summary can be a vector: {1st moment, 2nd moment, …, nth moment}
- Summary can be the result of any **statistical** algorithm
- Summary can be the result of a **data mining** algorithm
- Summary can be the result of a **machine learning algorithm**

Empirical Validation

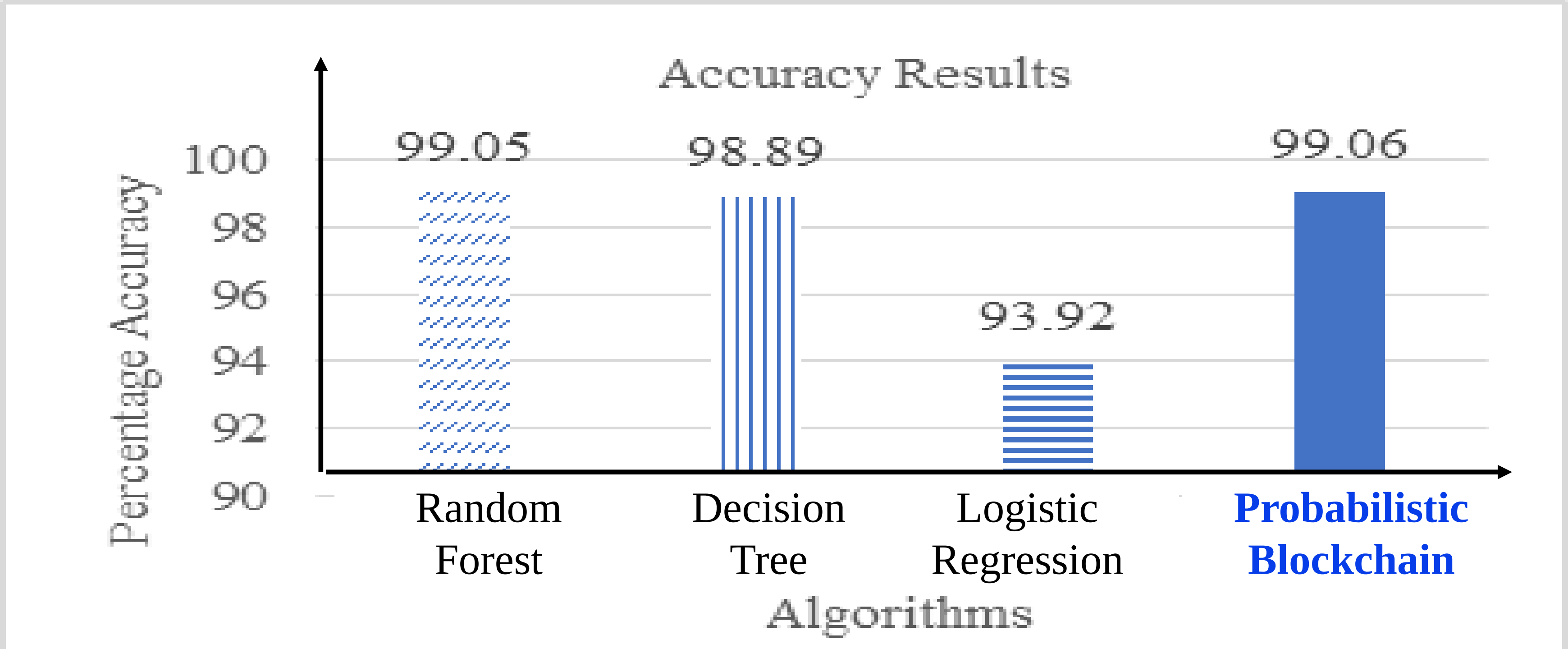
- Issue: Whether a network traffic pattern represents intrusion
- 1000 Agents* using different machine learning algorithms give their decisions: Yes or No
- Mining nodes summarize these decisions using the majority function

$$P = \frac{1}{n}\sum p_i$$

*In our simulation, agent modules randomly pick one of the 3 algorithms: Random Forest, Decision Tree, Logistic Regression

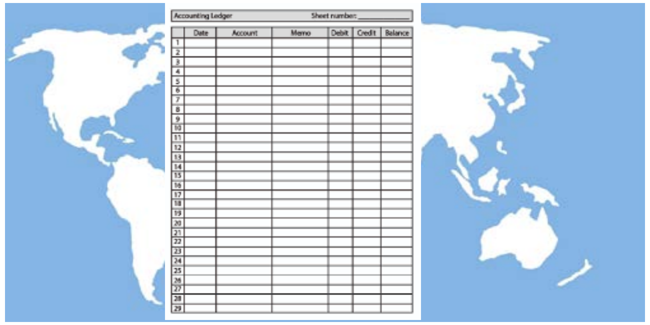
Results

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



Distributed decision making is better than any individual decision

Blockchain 4.0: Database to Knowledge Base



- ❑ Blockchain = Distributed ledger/database
- ❑ 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?
  - ➢ Is this a fake news? Spam? Fraud?

Knowledge Chain

- ❑ Customer query to blockchain network:
  How is the Cisco stock doing today?
- ❑ Blockchain to Customer: With 60% confidence, the probability of stock rising is 90%, …
- ❑ Ideal for **large** distributed systems with **no national boundaries**, no exchange limitations, no brokers in between
- ❑ **Crowd-sourced knowledge**, crowd-sourced decisions

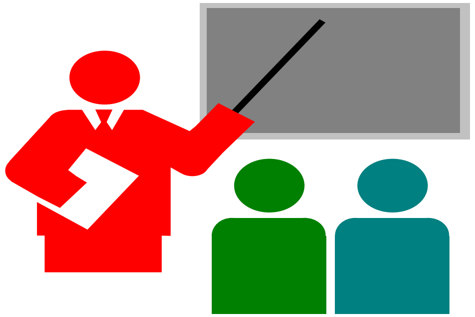
Application Examples

1. Spam from Email/IP Addresses/Cloud providers/source/public IP
2. Intrusions/attacks from IP Addresses. Anonymously share attack information.
3. Gray domains: Share gray list among agents.
4. Reliability/Issues with recent software updates
5. Error/reliability statistics of network/IoT devices
6. Virus in software

Issues to Resolve

1. **Summary functions**
2. **Overhead of consensus mechanisms**: Proof of Work, Proof of Stake, …
3. **Reputation of Experts and Bad Actors**:
   - Some agents are better than others
   - Group decisions should give more weight to them
   - How to incentivize better agents
   - How to penalize bad actors

Summary



1. Blockchains provide an immutable, secure, distributed database
2. Three generations: Crypto currency, Smart contract, faster performance
3. All three generations are deterministic and **only provide storage**
4. The next generation needs to **connect computation and AI** to make knowledge/decisions in addition to data storage
5. Consensus can be probabilistic result of any statistical algorithm, data mining, or machine learning ⇒ **Knowledge Chain**

Related Papers

- 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
- T. Salman, R. Jain, and L. Gupta, "**A Reputation Management Framework for Knowledge-Based and Probabilistic Blockchains**," 2019 IEEE International Conference on Blockchain, Atlanta, July 14, 2019, http://www.cse.wustl.edu/~jain/papers/rpmcewa.htm

List of Acronyms

AI	Artificial Intelligence
DNS	Domain Name Service
IEEE	Institution of Electrical and Electronics Engineers
IoT	Internet of Things
IP	Internet Protocol
PKI	Public Key Infrastructure
SSL	Secure Socket Layer

Scan This to Download These Slides

Raj Jain

rajjain.com

Slides and Video Recording at:

http://www.cse.wustl.edu/~jain/talks/pbc_acm.htm