Blockchain Distributed Ledger Technologies for Biomedical and Health Care Applications

Tsung-Ting Kuo, PhD, Hyeon-Eui Kim, PhD, Lucila Ohno-Machado, MD, PhD
University of California, San Diego
Objective and Scope

• The Bitcoin and Blockchain
  • Bitcoin is one of the most famous crypto-currencies
  • Blockchain is the underlying technology
  • A new distributed database beyond coin applications

• Objective
  • A preliminary study to investigate the possibility to adopt Blockchain technology on bioCADDIE

• Scope
  • Bitcoin Blockchain features, alternatives, & applications
  • Key benefits and biomedical/healthcare applications
• Challenge 1: **double-spending transactions (TXs)**

Valid TX

![Valid TX Diagram](diagram1)

(suppose Alice only has these 10 coins)

Double-spending (Invalid) TX

![Double-spending Diagram](diagram2)

• We need a **timestamp** to determine order
• **Challenge 2:** single-point-of-failure

![Diagram of Bitcoin Blockchain]

- Node 1
- Node 2
- Node 3
- Node 4

Central Intermediary

Failed Intermediary

Failed TX

Single-point-of-failure

No single-point-of-failure

• **We need a distributed timestamping mechanism**
The Bitcoin Blockchain (3/6)

- **Solution Step 1:** hash-chain timestamping

  Diagram:

  - **Node 1:** Double-spend? → Reject
  - **Node 2:** Double-spend? → Reject
  - **Node 3:** Double-spend? → Reject
  - **Node 4:** Everyone can see everything

  **Notes:**
  - No single-point-of-failure with verified TXs
  - Double-spending node (Alice’s 2 transactions)
  - Distributed timestamping
The Bitcoin Blockchain (4/6)

- Solution Step 1: hash-chain timestamping (cont.)
  - Every node maintains a copy of all TXs
  - Hash-chain (blockchain) decides the order of TXs

- However, we still need to deal with invalid blocks
- Blocks should be hard to create but easy to check
The Bitcoin Blockchain (5/6)

- Solution Step 2: proof-of-work algorithm

Every node starts proof-of-work

Node 1 creates $B_2$ for incentives

Nonce $N_2$ = “7C 4D DB 29” → Hash of $B_2$’s header = “2D F8 8E 32 … 10 9A FE 1C”, NO (T = 10:14:20)
Nonce $N_2$ = “7C 4D DB 30” → Hash of $B_2$’s header = “41 2A B3 DC … 94 29 AB B5”, NO (T= 10:14:25)
Nonce $N_2$ = “7C 4D DB 31” → Hash of $B_2$’s header = “00 00 4F 65 … 2F ED 31 09”, YES (T = 10:14:30)

Node 2

Nonce $N_2$ = “61 0A 3F 3A” → Hash of $B_2$’s header = “A8 C7 08 C9 … 3D F1 A2 F9”, NO (T = 10:14:23)
Nonce $N_2$ = “61 0A 3F 3B” → Hash of $B_2$’s header = “2A E9 84 66 … 91 B4 58 CE”, NO (T = 10:14:28)
Stopped after identifying that Node 1 has completed proof-of-work at time = 10:14:30

Node 3

Nonce $N_2$ = “99 06 10 13” → Hash of $B_2$’s header = “FB 2F 26 D9 … 39 F5 C1 0B”, NO (T = 10:14:21)
Nonce $N_2$ = “99 06 10 14” → Hash of $B_2$’s header = “E2 1C 09 05 … 25 3E AA CF”, NO (T = 10:14:26)
Stopped after identifying that Node 1 has completed proof-of-work at time = 10:14:30

Stopped after identifying that Node 1 has completed proof-of-work at time = 10:14:30
• Solution Step 2: proof-of-work algorithm (cont.)
  • Majority voting (honest CPUs $\triangleright$ malicious CPUs)

"Longest" branch chain

"Honest" Block $H_1$
- Hash
- Nonce
- TXs

"Honest" Block $H_2$
- Hash
- Nonce
- TXs

"Honest" Block $H_3$
- Hash
- Nonce
- TXs

"New" Block $N$
- Hash
- Nonce
- TXs

"Malicious" Block $M_1$
- Hash
- Nonce
- TXs

"Malicious" Block $M_2$
- Hash
- Nonce
- TXs

Not-"Longest" branch chains

8/7/17
Alternatives & Applications

• Alternative crypto-currencies and blockchains
  • **Coins**: Ethereum, Ripple, Dash, Litecoin, Monero, …
  • **Protocols**: proof-of-stake/burn/elapsed-time/…

• Blockchains as distributed ledgers
  • **Metadata of TX**: MultiChain, BigchainDB, …
  • **Smart contract/property**: Ethereum, Hyperledger, …

• Non-financial applications
  • Either **permissioned** or **permission-less** networks

• Use blockchain to help healthcare/research
  • Instead of being harmful (e.g., pay for ransomware)
Key Benefits

• Comparing to traditional distributed databases
  • Decentralized management
  • Immutable audit trail
  • Data provenance
  • Robustness/availability
  • Security/privacy

• Crucial for biomedical and healthcare applications
  • To share, exchange, analyze, record, and validate data
  • One of the most important emerging application area
  • Especially for Health Information Exchange (HIE)
• **Improved medical record management**

<table>
<thead>
<tr>
<th>Key Benefit</th>
<th>Biomedical/Healthcare Use Case</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decentralized Management</td>
<td>Patient-managed healthcare records</td>
</tr>
<tr>
<td>Immutable Audit Trail</td>
<td>Unalterable patient records</td>
</tr>
<tr>
<td>Data Provenance</td>
<td>Source-verifiable medical records</td>
</tr>
<tr>
<td>Robustness/Availability</td>
<td>Reduced burden of patient record keeping</td>
</tr>
<tr>
<td>Security/Privacy</td>
<td>Increased safety of medical records</td>
</tr>
</tbody>
</table>
• Enhanced insurance claim process

<table>
<thead>
<tr>
<th>Key Benefit</th>
<th>Biomedical/Healthcare Use Case</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decentralized Management</td>
<td>Real-time claim processing</td>
</tr>
<tr>
<td>Immutable Audit Trail</td>
<td>Improved claim auditing and fraud detection:</td>
</tr>
<tr>
<td>Data Provenance</td>
<td>Verifiable records for claim qualification:</td>
</tr>
<tr>
<td>Robustness/Availability</td>
<td>Enhanced accessibility of patient data</td>
</tr>
<tr>
<td>Security/Privacy</td>
<td>Increased security of patient medical insurance info</td>
</tr>
</tbody>
</table>
• Advanced biomedical/healthcare data ledger

<table>
<thead>
<tr>
<th>Key Benefit</th>
<th>Biomedical/Healthcare Use Case</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decentralized Management</td>
<td>Decentralized health data backbone</td>
</tr>
<tr>
<td>Immutable Audit Trail</td>
<td>Unchangeable log of clinical research protocols</td>
</tr>
<tr>
<td>Data Provenance</td>
<td>Ensured original manufacturer and ownership transferring in pharmaceutical supply chain</td>
</tr>
<tr>
<td>Robustness/Availability</td>
<td>Improved robustness for counterfeit drug prevention and detection systems in pharmaceutical supply chain</td>
</tr>
<tr>
<td>Security/Privacy</td>
<td>Higher patient confidence for consent recording systems</td>
</tr>
</tbody>
</table>
• **Accelerated clinical/biomedical research**

<table>
<thead>
<tr>
<th>Key Benefit</th>
<th>Biomedical/Healthcare Use Case</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decentralized Management</td>
<td>Improved data sharing/analysis without ceding control</td>
</tr>
<tr>
<td>Immutable Audit Trail</td>
<td>Trackable and timestamped patient-generated data</td>
</tr>
<tr>
<td>Data Provenance</td>
<td>Evidenced provenance for medical research data</td>
</tr>
<tr>
<td>Robustness/Availability</td>
<td>Superior healthcare data availability</td>
</tr>
<tr>
<td>Security/Privacy</td>
<td>Secure and privacy-preserving healthcare data sharing</td>
</tr>
</tbody>
</table>
Potential Activity Area

1. Data Shop
   - Indexing, Linking data to articles
   - Software development
   - New ideas via Pilot Projects, Supplements

2. Data Showcase
   - Search Engine Prototype
   - User Feedback

3. Data Market
   - Incentives to Share, Data Citation
   - Outreach

Blockchain
- Decentralized Management
- Immutable Audit Trail
- Data Provenance
- Robustness/Availability
- Security/Privacy

bioCADDIE is supported by the National Institutes of Health through the Big Data to Knowledge, Grant U24AI117966.
Challenges and Solutions

• Potential problems and challenges
  • Transparency/confidentiality: “pseudonymity”
  • Speed/scalability: max 7 TXs/second for Bitcoin
  • Threat of a 51% attack: honest CPUs < malicious CPUs

• Proposed solutions and implementations
  • Aggregated-data/encrypted-data
  • Index-data/new-implementation
  • Private/VPN/HIPAA-cloud
  • Example: ModelChain

https://healthit.gov/blockchain
Conclusion

• Bitcoin and Blockchain technology
  • Decentralized management, immutable audit trail, data provenance, robustness/availability, security/privacy

• Biomedical/healthcare Blockchain applications
  • Medical record, insurance claim, healthcare ledger, clinical/biomedical research

• We expect many new applications to emerge soon
  • Adoption on bioCADDIE such as Data Market

• Review paper accepted by JAMIA
  • DOI: 10.1093/jamia/ocx068
Thank you

Acknowledgements

NIH U24AI117966
PCORI CDRN-1306-04819
NIH U54HL108460
NIH UL1TR001442
VA IIR12-068