Software Engineering Challenges in Blockchain-Based Decentralized Systems

LECTURE: Davor Svetinovic
My Research Background

I have participated in a wide variety of complex multi-disciplinary research projects, e.g.,:

- Sustainable cities modelling and development (systems)
- Smart grid cyber security (systems)
- Data mining of software repositories (analytics)
- Software system evolution and cloning (analytics)
- Blockchain engineering (systems / analytics)
- Decentralized AI (systems / analytics)
My Blockchain Technology Background

- Interested in Bitcoin / Blockchain Technology since 2009
- Various teaching and research projects since 2010
  - Bitcoin C port/library, Bitmessage, Electrum, Multibit, Dogecoin, Ethereum
  - Cryptocurrency for CO2 emissions trading
  - Solar mining cryptocurrency (solar energy signature)
  - Automated Bitcoin energy trade (IoT, smart cars, smart meters, etc.)
  - Intelligent trading challenges, no central price signal, exchanges, GCC market
  - Blockchain analytics (community detection, forensics)
Money

Functions

- Unit of account
- Medium of exchange
- Store of value

Properties:
- Fungible
- Durable
- Divisible
- Portable
- Limited in supply
- Acceptable
"Physical" Transfer of Assets (and Value) in the Cyberspace

Source: https://coinmarketcap.com/currencies/bitcoin/

Institute for Distributed Ledgers and Token economy
# Second Contribution: “Anonymous”

## Decentralized Entities

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Price</th>
<th>24h %</th>
<th>7d %</th>
<th>Market Cap</th>
<th>Volume (24h)</th>
<th>Circulating Supply</th>
<th>Last 7 Days</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bitcoin BTC</td>
<td>$61,808.96</td>
<td>-4.11%</td>
<td>-12.41%</td>
<td>$1,162,884,586,025</td>
<td>$45,471,142,388</td>
<td>18,845,000 BTC</td>
<td><img src="https://example.com/graph1.png" alt="Graph" /></td>
</tr>
<tr>
<td>2</td>
<td>Ethereum ETH</td>
<td>$3,859.44</td>
<td>-1.40%</td>
<td>-7.03%</td>
<td>$454,236,879,655</td>
<td>$18,594,942,813</td>
<td>117,945,259 ETH</td>
<td><img src="https://example.com/graph2.png" alt="Graph" /></td>
</tr>
<tr>
<td>3</td>
<td>Binance Coin</td>
<td>$477.92</td>
<td>-2.45%</td>
<td>-13.00%</td>
<td>$80,470,824,269</td>
<td>$2,478,539,066</td>
<td>168,137,036 BNB</td>
<td><img src="https://example.com/graph3.png" alt="Graph" /></td>
</tr>
<tr>
<td>4</td>
<td>Cardano ADA</td>
<td>$2.20</td>
<td>-0.62%</td>
<td>2.65%</td>
<td>$72,480,998,935</td>
<td>$3,154,570,061</td>
<td>32,904,527,669 ADA</td>
<td><img src="https://example.com/graph4.png" alt="Graph" /></td>
</tr>
<tr>
<td>5</td>
<td>Tether USDT</td>
<td>$0.9993</td>
<td>-0.03%</td>
<td>0.10%</td>
<td>$68,544,240,857</td>
<td>$81,807,542,217</td>
<td>68,589,070,064 USDT</td>
<td><img src="https://example.com/graph5.png" alt="Graph" /></td>
</tr>
<tr>
<td>6</td>
<td>XRP XRP</td>
<td>$1.14</td>
<td>-0.08%</td>
<td>-6.25%</td>
<td>$53,371,231,689</td>
<td>$4,476,133,429</td>
<td>46,878,114,887 XRP</td>
<td><img src="https://example.com/graph6.png" alt="Graph" /></td>
</tr>
<tr>
<td>7</td>
<td>Solana SOL</td>
<td>$158.13</td>
<td>2.43%</td>
<td>0.82%</td>
<td>$47,498,124,872</td>
<td>$2,828,989,571</td>
<td>300,272,746 SOL</td>
<td><img src="https://example.com/graph7.png" alt="Graph" /></td>
</tr>
</tbody>
</table>

Source: [coinmarketcap.com](http://coinmarketcap.com/)

Institute for Distributed Ledgers and Token economy
Bitcoin

- **Protocol** – a decentralized peer-to-peer network
- **Blockchain** – a public transaction ledger
- **Consensus mechanism** – a decentralized mechanism of providing trust through proof of work (mining)
- Currency issuance and payment verification system
Hash functions (any string $\rightarrow$ fixed length string) with security properties: collision free, hiding, and puzzle-friendly

- Hash pointers
- Hash pointer linked lists (blockchains)
- Hash pointer binary trees (Merkle trees)
- Digital signatures and public identities
Presentation-Related Publications


*corresponding author
Project #1: Need for Ever More Privacy!


Addr: 1LDNL...
Tag: SilkRoad

Addr: 1933p...
Tag: SilkRoad

Over 50 addresses which belong to Tumbler Mixer

Addr: 1EG9H...
Tag: Tumbler Mixer

Addr: 1AHki...
Tag: Tumbler Mixer

Sweep transaction

Relay transactions

TX:afeec...

TX:fd059...

TX:f3B60...

TX:70d46...

TX:758b7...

◆ : Address (Different colors represent different tags)  ○ : Transaction
**Project #1: Users Get Creative to Protect Privacy**

- **Peel Transaction:** The most common type of transaction. No matter what the number of inputs in a transaction is, there will always be two output addresses.

- **Sweep Transaction:** A transaction combines multiple inputs into one output. This type of transaction makes it easier for entities to process and control their bitcoins.

- **Distribution Transaction:** Distributing transactions usually have any number of inputs and more than three output addresses. This kind of transaction is commonly related to one single party or organization trying to pay multiple parties, e.g., gambling rewards or pooled mining rewards.

- **Relay Transaction:** This kind of pattern only includes one input and one output. It can be used to move bitcoins from one party to another without leaving any correlated information behind, such as the change address or other input addresses.

- **Self-Spending Transaction:** An input address also appears as an output in the same transaction, the number of self-spending addresses in one transaction can be many.

- **Peeling Chain Transaction:** A series of peel transactions. The change address is used as the input to a subsequent peel transaction.
Project #1: Contributions

- The **largest study so far** conducted on the clustering of Bitcoin addresses (at the publishing time)—the dataset used in this paper contains over 46 million Bitcoin addresses and 46.5 million transactions.

- A **new clustering approach** was developed based on a novel combination of known-address clustering with transaction patterns.

- The proposed approach was compared to the existing known-address clustering, and the results showed that our approach produced **significantly more homogeneous clusters** (lower “impurity”), which better predict shared transactions validated against the test set.

- **Scalable approach** that can handle large datasets.
Project #2: It’s all a bit Social!

Despite the different purposes of the various social networks and Bitcoin transaction network, one can observe universal dynamics such as the densification power law, shrinking diameter, and modular structure.

However, we also found that Bitcoin deviates in important ways due to anonymity-seeking behavioral patterns of its users. As a result, the network exhibits a two-orders-of-magnitude larger diameter, sparse tree-like communities, and an overwhelming majority of transitional or intermediate accounts with incoming and outgoing edges but zero cumulative balances.
• From the systems perspective, it was observed that some of these (anti-)social incentives are affecting critical system properties, such as security and privacy.

• The social incentives are leading to formation or disintegration of certain network communities. This in turn is leading to the improper use or the intentional misuse of the overall system.

• These community alterations present serious threats to a subset of the system properties that we identified: decentralization, longevity (of the system that’s supposed to evolve over next hundreds of years), trust, participation incentive, privacy, security, and usage ethics.
We realized what we knew all along:

it is all not that much about the blockchain

as it is about the consensus!

And not just the consensus at the technical level of adding new transactions to the blockchain, but about the general consensus what we are using all these systems for.

And the system-level consensus is not just about privacy and security but about an even more expensive construct: trust.
And Trust is Very Expensive


<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bitcoin's current estimated annual electricity consumption (TWh)</td>
<td>175.10 TWh</td>
</tr>
<tr>
<td>Annualized global mining revenues</td>
<td>$21,156,404,372 (Total value of mining rewards (including fees) per year.)</td>
</tr>
<tr>
<td>Annualized estimated global mining costs</td>
<td>$8,755,116,628 (Assuming a fixed rate of 5 cents per kilowatt-hour.)</td>
</tr>
<tr>
<td>Current cost percentage</td>
<td>41.38% (Estimated ratio of electricity costs to total miner income.)</td>
</tr>
</tbody>
</table>

vs. Austria has the total consumption of 64.60 billion kWh (TWh)

of electric energy per year. Per capita this is an average of 7,227 kWh. (vs. 1765.42 kWh / Bitcoin transaction)

Austria can partly provide itself with self-produced energy. The total production of all electric energy producing facilities is 61 bn kWh. That is 94% of the country's own usage. The rest of the needed energy is imported from foreign countries. Along with pure consumptions the production, imports and exports play an important role. Other energy sources such as natural gas or crude oil are also used. (source: https://www.worlddata.info/europe/austria/energy-consumption.php)

Institute for Distributed Ledgers and Token economy
Project #3: Trust in Cryptocurrency Ecosystems / Blockchain Oracles

Project #4: Trust RE in Blockchain Systems

Project #4: Trust RE in Blockchain Systems


Trust-related Requirements Model for Blockchain Systems

System Quality Characteristics
- Safety
  - Feedback
  - Protection
  - Resilience
  - Integrity
- Security
  - Confidentiality
  - Integrity
  - Non-repudiation
  - Authenticity
- Reliability
  - Maturity
  - Availability
  - Fault tolerance
  - Recoverability
- Privacy
  - Anonymity
  - Pseudonymity
  - Undetectability
  - Unlinkability
- Decentralization
  - Governance
  - Accountability
  - Consensus
  - Immutability
  - Confidentiality

Data Quality Characteristics
- Inherent
  - Accuracy
  - Completeness
  - Consistency
  - Credibility
  - Currentness
- Common
  - Accessibility
  - Compliance
  - Confidentiality
  - Efficiency
  - Precision
- System Dependent
  - Traceability
  - Understandability
  - Availability
  - Portability
  - Recoverability

Institute for Distributed Ledgers and Token economy
• Operating correctly is critical for security/privacy/trust.

• **Energy and effort** invested in building the blockchain must be there to preserve security/privacy/trust (avoid alternative versions of the blockchain)

• (Random) work to build a chain + network selection of the longest (most difficult) chain = Nakamoto’s (practical) solution to Bysantine Generals’ Problem => core blockchain security/trust mechanism

• Given how expensive all of this is... what about the various applications of blockchain technology?
Many are nothing more but standard centralized systems: “Private” or “money-less” blockchains -> have “fun” with standard software/hardware security/privacy issues

However, for fully public, decentralized, and incentivized blockchains -> we are not really sure what we are dealing with yet in terms of trust: Economics (?) ++ Software ++ Hardware ++ ???

Consensus is implemented in software and ensured by hardware. Economic incentives must be done right, and “absorb” all the software/hardware “bugs.” We believe ??? unknown might be the missing “social” constructs which we were not able to identify so far beyond...
The Important Social Blockchain Trust Threats

- Social threats
- Planting rogue entities
- Creating dissent
- Spreading wrong information
- Creating arguments
- Legalization
- Regulation, etc.

- And at the same time ongoing blockchain convergence with AI (with its own trust, ethical, fairness issues) and IoT (with its own privacy, security, and trust issues).

So, what should we do?
Blockchain-AI Infrastructures for Trust in Socio-Technical Systems

• We are already implementing AI-Blockchain-IoT platforms (e.g., singularitynet), which is good, but we don’t have even basic trust concepts and issues resolved.

• What is trust? Unfortunately, from the engineering perspective there is no even a well-defined technical requirements model. (We worked on it)

• Critical to tackle all components in parallel: hardware security/trust, software security/trust, social security/trust, explainable AI and emerging behavior security/trust.

• We must develop a trust infrastructure that can absorb the large costs and enable integration of complex AI applications that will (at least in theory) replace traditional social trust-related services.
Decentralized Blockchain-Based Federated Machine Learning Applications

• AI is creating many new exciting socio-technical systems engineering challenges to facilitate development of new emergent, adaptable, AI-based systems. Lots of bias and fairness issues.

• In particular, ultra-long term nature of such systems must be taken into account. We lack software engineering techniques for development of such systems.

• Make social science findings and results engineering ready. Learn how to implement them using trust-focused decentralized Federated Learning Applications.

Institute for Distributed Ledgers and Token Economy
https://www.wu.ac.at/en/tecon

The Institute focuses on the cutting-edge research in the fundamentals of blockchain technology and the applications to economics, law, business, and social sciences.

Research Institute for Cryptoeconomics
https://www.wu.ac.at/en/cryptoeconomics/team

The institute focuses on interdisciplinary analysis and applications of cryptoeconomics and blockchain technology.
Prof. Davor Svetinovic
Department of Information Systems and Operations Management
Institute for Distributed Ledgers and Token Economy
Vienna University of Economics and Business
T +43-1-313 36-4255
davor.svetinovic@wu.ac.at
wu.ac.at/tecon
davors.com