Blockchains Beyond Bitcoin: Towards Optimal Level of Decentralization in Storing Financial Data

Thach Ngoc Nguyen[1], Olga Kosheleva[2]
Vladik Kreinovich[2], and Hoang Phuong Nguyen[3]

[1]Banking University of Ho Chi Minh City, Vietnam,
Thachnn@buh.edu.vn
[2]University of Texas at El Paso, El Paso, Texas 79968, USA
olgak@utep.edu, vladik@utep.edu
[3]Division Informatics, Math-Informatics Faculty, Thang Long University,
Hanoi, Vietnam, nhphuong2008@gmail.com

1. How Financial Information Is Currently Stored

- At present, usually, the information about each financial transaction is stored in three places:
  - with the buyer,
  - with the seller, and
  - with the bank.
- In many real-life financial transactions, a problem later appears.
- So it becomes necessary to recover the information about the sale.
- From this viewpoint, the current system of storing information is not fully reliable.

2. Limitations of the Current Scheme

- If a buyer has a problem, and his/her computer crashes and deletes the original record,
  - the only neutral source of information is then the bank,
  - but the bank may have gone bankrupt since then.
- It is thus desirable to have more duplication, to increase the reliability of storing financial records.

3. Blockchain as an Absolutely Reliable – but Somewhat Wasteful – Scheme for Storing Data

- The known reliable alternative to the usual scheme of storing financial data is the *blockchain* scheme.
- It was originally designed for bitcoin transactions.
- In this scheme, the record of each transaction is stored at the location of every single participant.
- This extreme duplication makes blockchains a very reliable way of storing financial data.
- On the other hand, in this scheme:
  - every time anyone performs a financial transaction,
  - this information needs to be transmitted to all the nodes.
- This takes a lot of computation time.
- So, from this viewpoint, this scheme is wasteful.

4. Formulation of the Problem

- What scheme should we select to store the financial data?
- It would be nice to have our data stored in an absolutely reliable way.
- Thus, it may seem reasonable to use blockchain for all financial transactions, not just for bitcoins.
- However, already for bitcoins, each world-wide update corresponding to each transaction takes about 10 seconds.
- And bitcoins participate in only a small percentage of financial transactions.
- If we apply the same technique to all financial transactions, this delay would increase drastically.

5. Formulation of the Problem (cont-d)

- The resulting hours of delay will make the system completely impractical.
- So, it is desirable to find the *optimal* level of duplication for each financial transaction.
- This level may be different for different transactions.
- When a customer buys a relatively cheap product, too much duplication probably does not make sense.
- The risk is small but the need for additional storage would increase the cost.
- On the other hand, for an expensive purchase, we may want to spend a little more to decrease the risk.
- This is similar to how we buy insurance when we buy a house or a car.

6. Formulation of the Problem (cont-d)

- Good news is that the blockchain scheme itself – with its encryptions etc. – does not depend on
  - whether we store each transaction at every node
  - or only in some selected nodes.
- In this sense, the technology is there, no matter what level of duplication we choose.
- The only problem is to find the optimal duplication level.
- In this paper, we show how to find the optimal level of duplication for each type of financial transaction.

7. Notations

- Let d denote the level of duplication.
- So d is the number of copies of the original transaction record that will be independently stored.
- Let p be the probability that each copy can be lost.
- This probability can be estimated based on experience.
- Let c denote the total cost of storing one copy of the transaction record.
- Finally, let L be the expected financial loss that will happen if:
  - a problem emerges related to the original sale, and
  - all the copies of the corresponding record have disappeared.

8. Notations (cont-d)

- This expected financial loss L can estimated by multiplying:
  - the cost of the transaction and
  - the probability that the bought item will turn out to be faulty.
- The cost c of storing a copy is about the same for all the transactions, whether they are small or large.
- On the other hand, the potential loss L depends on the size of the transaction – and on the corresponding risk.

9. Analysis of the Problem

- Since the cost of storing one copy of the financial transaction is c, the cost of storing d copies is equal to $d \cdot c$.
- To this cost, we need to add the expected loss if all copies of the transaction are accidentally deleted.
- For each copy, the probability that it will be accidentally deleted is p.
- The copies are assumed to be independent. Since we have d copies:
  - the probability that all d of them will be accidentally deleted is therefore equal to
  - the product p^d of the d probabilities p corresponding to each copy.
- So, we have the loss L with probability p^d.

10. Analysis of the Problem (cont-d)

- We have the loss L with probability p^d.
- Thus, the expected loss from losing all the copies of the record is equal to the product $p^d \cdot L$.
- Hence, once we have selected the number d of copies, the overall expected loss E is equal to $E = d \cdot c + p^d \cdot L$.
- We need to find the value d for which this overall loss is the smallest possible.

11. Let Us Find the Optimal Level of Duplication

- To find the optimal d, we equate the derivative of the above expression with respect to 0; then:

$$d = \frac{1}{|\ln(p)|} \cdot \ln(L) + \frac{\ln|\ln(p)| - \ln(c)}{|\ln(p)|}.$$

- As one can easily see, the larger the expected loss L, the more duplications we need.

- In general, the number of duplications is proportional to the logarithm of the expected loss.

12. Discussion

- The value d computed by the above formula may be not an integer.
- Due to the loss formula, the derivative of the overall loss E is first decreasing then increasing.
- Thus, to find the optimal integer value d, it is sufficient to consider two integers on the two sides of the above real value:
  - its floor $\lfloor d \rfloor$ and
  - its ceiling $\lceil d \rceil$.
- Out of these two values, we need to find the one for which the overall loss E is the smallest.

13. Acknowledgments

- This work was supported in part by the US National Science Foundation via grant HRD-1242122.
- The authors are thankful to Professor Hung T. Nguyen for valuable discussions.