Compressing Bitcoin Blockchain

Abhishek Kumar Shukla, Alexandra McSween, Evan MacNeil, Ivan Lau, Shang Li, Yanhong Xu

Supervised by Germán Luna and Cuneyt Akcora

What is a cryptocurrency?

https://vpnhaus.ncp-e.com/2019/us-moves-to-track-cryptocurrency-transactions/

The blockchain is

- **THE ledger**
- Decentralized
- Immutable

www.austinchronicle.com/books/2008-03-14/601515

The blockchain is

- THE ledger
- **Decentralized**
- Immutable

Figure 3. Centralized money versus decentralized money

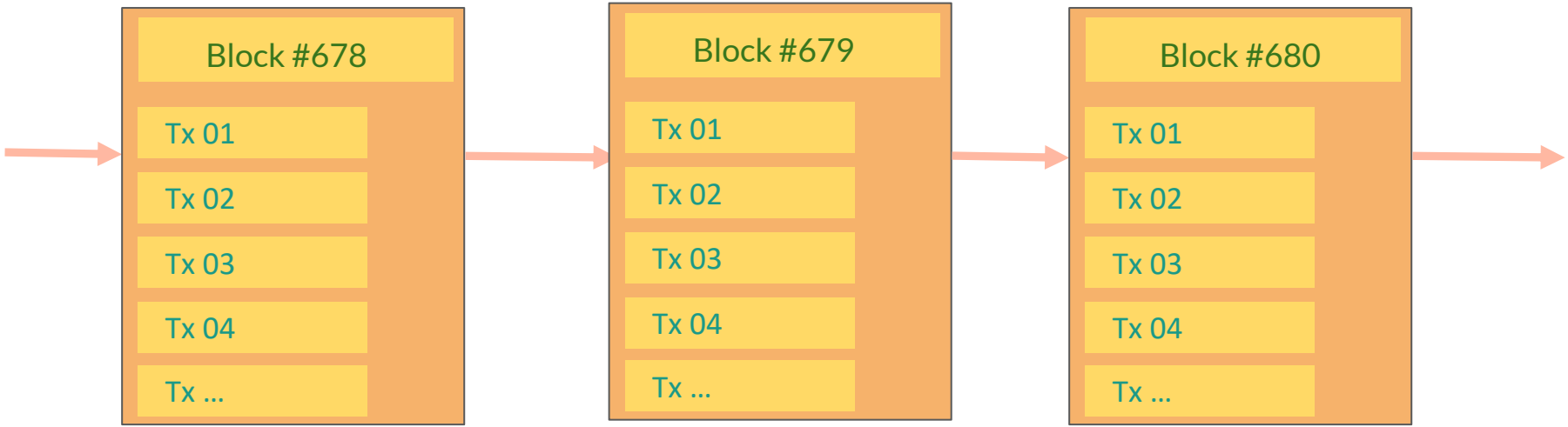
What Is Blockchain?By Sir John Hargrave and Evan Karnoupakis

The blockchain is

- THE ledger
- Decentralized
- **Immutable**

https://courses.lumenlearning.com/suny-hccc-worldcivilization/chapter/the-sumerians/

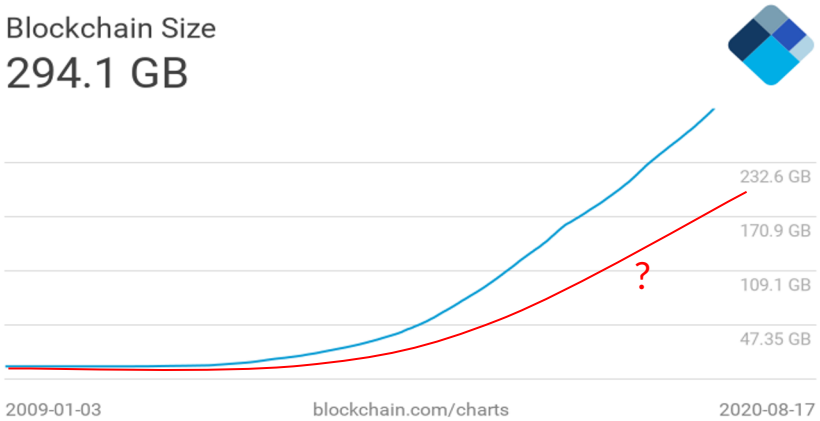
Blockchain: chain of blocks



Problem(s)

Scalability: Blockchain length and size keeps on growing

Playing field: Initial setup requires user to download the blockchain and verify the entire history(takes 12 hours).



Solution

Stare hard at Bitcoin transactions and see where we can penny pinch.

Example transaction

```
0200000001a2b8481a2538389bac1d07402f0db8a8e5e9152cf62f01861
30b23f4fd34dcac010000006b483045022100ba01d0e61f91e55ccf5e08
87f9ca268ea5b6c8208bd1685c23d7a259ecf51ca902202b19ce3dcc743
bdf6a66bdcf7eb56ea6ace829e01279817cbb624ccf45c6949701210323
bfb50086263c1de30efe9ce5b1e3e2f74e928d9e349410f22fe62dd5b74
c74feffffff0278d3180000000000017a9146c83df9bd5b763af48efc5
d41430da09e9c113e78781313c01000000001976a9147caafd48f0ffaf9
a16a1362dba5fd492a8dcff5588acecd50900
```

Solution

- Stare hard at Bitcoin transactions and see where we can penny pinch.

Example transaction

```
0200000001a2b8481a2538389bac1d07402f0db8a8e5e9152cf62f01861
30b23f4fd34dcac010000006b483045022100ba01d0e61f91e55ccf5e08
87f9ca268ea5b6c8208bd1685c23d7a259ecf51ca902202b19ce3dcc743
bdf6a66bdcf7eb56ea6ace829e01279817cbb624ccf45c6949701210323
bfb50086263c1de30efe9ce5b1e3e2f74e928d9e349410f22fe62dd5b74
c74feffffff0278d3180000000000017a9146c83df9bd5b763af48efc5
d41430da09e9c113e78781313c0100000000001976a9147caafd48f0ffaf9
a16a1362dba5fd492a8dcff5588acecd50900
```

Example transaction

Version #	02000000
# of inputs	01
Tx hash	a2b8481a2538389bac1d07402f0db8a8e5e9152cf62f0186130b23f4fd34dcac
Input script index	01000000
Input script	6b483045022100ba01d0e61f91e55ccf5e0887f9ca268ea5b6c8208bd1685c23d7a259ecf 51ca902202b19ce3dcc743bdf6a66bdcf7eb56ea6ace829e01279817cbb624ccf45c69497 01210323bfb50086263c1de30efe9ce5b1e3e2f74e928d9e349410f22fe62dd5b74c74
Seq #	fefffff
# of inputs	02
Value	78d3180000000000
output script	17a9146c83df9bd5b763af48efc5d41430da09e9c113e787
Value	81313c0100000000
output script	1976a9147caafd48f0ffaf9a16a1362dba5fd492a8dcff5588ac ecd50900

Observation

There is way more space allocated to certain parts of the transaction than that is seen in practice.

Compression ideas

Version #	02000000
# of inputs	01
Tx hash	a2b8481a2538389bac1d07402f0db8a8e5e9152cf62f0186130b23f4fd34dcac
Input script index	01000000
Input script	6b483045022100ba01d0e61f91e55ccf5e0887f9ca268ea5b6c8208bd1685c23d7a259ecf 51ca902202b19ce3dcc743bdf6a66bdcf7eb56ea6ace829e01279817cbb624ccf45c69497 01210323bfb50086263c1de30efe9ce5b1e3e2f74e928d9e349410f22fe62dd5b74c74
Seq #	fefffff
# of inputs	02
Value	78d3180000000000
output script	17a9146c83df9bd5b763af48efc5d41430da09e9c113e787
Value	81313c0100000000
output script	1976a9147caafd48f0ffaf9a16a1362dba5fd492a8dcff5588ac ecd50900

Example transaction

Version #	02000000
# of inputs	01
Tx hash	a2b8481a2538389bac1d07402f0db8a8e5e9152cf62f0186130b23f4fd34dcac
Input script index	01000000
Input script	6b483045022100ba01d0e61f91e55ccf5e0887f9ca268ea5b6c8208bd1685c23d7a259ecf 51ca902202b19ce3dcc743bdf6a66bdcf7eb56ea6ace829e01279817cbb624ccf45c69497 01210323bfb50086263c1de30efe9ce5b1e3e2f74e928d9e349410f22fe62dd5b74c74
Seq #	fefffff
# of inputs	02
Tx value	78d3180000000000
output script	17a9146c83df9bd5b763af48efc5d41430da09e9c113e787
Tx value	81313c0100000000
output script	1976a9147caafd48f0ffaf9a16a1362dba5fd492a8dcff5588ac ecd50900

Compressing values

- Values are the amount of **Satoshis** being transferred (1BTC= 100 million Satoshis).
- Currently this uses **8 Byte** of space but **4 Byte** is enough to represent **95%** of values in transactions
- Instead of using fixed length data type, we can use variable-length data types to store the values.
  - Use **4 bytes** to represent those 95% of the values, and leave the remaining as they were (**8 bytes**).
  - A **1 bit flag** to indicates the data type (0 for 8 bytes, 1 for 4 bytes).

Savings: 5.7 GB in Bitcoin blockchain

Example transaction

Version #	02000000
# of inputs	01
Tx hash	a2b8481a2538389bac1d07402f0db8a8e5e9152cf62f0186130b23f4fd34dcac
Input script index	01000000
Input script	6b483045022100ba01d0e61f91e55ccf5e0887f9ca268ea5b6c8208bd1685c23d7a259ecf 51ca902202b19ce3dcc743bdf6a66bdcf7eb56ea6ace829e01279817cbb624ccf45c69497 01210323bfb50086263c1de30efe9ce5b1e3e2f74e928d9e349410f22fe62dd5b74c74
Seq #	fefffff
# of inputs	02
Value	78d3180000000000
output script	17a9146c83df9bd5b763af48efc5d41430da09e9c113e787
Value	81313c0100000000
output script	1976a9147caafd48f0ffaf9a16a1362dba5fd492a8dcff5588ac
	ecd50900

Compressing scripts

- A script is a list of instructions recorded with each transaction to authenticate the transaction
- The most common script is P2PKH. It looks like

`1976a914<address>88ac`

- We can replace this wrapping code with a 1B index into a lookup table

`01<address>`

Savings: 4.5 GB in Bitcoin blockchain

Compressing addresses

- Users' hold and spend Bitcoins through their addresses.
- Each user can have multiple addresses; in fact one for each transaction.

Fact: Single address (20B) is used multiple times

Solution: Put all addresses in a table and refer to them using a 4B index.

Savings: 24 GB in Bitcoin blockchain

Conclusion

Method	Saving
Compressing addresses	24 GB
Compressing scripts	4.5 GB
Compressing values	5.7 GB
Other methods	21.1 GB
Total	55.3 GB

Compression rate ~ 18.7%

Thank you!

Contact: *shukla2@ualberta.ca*