Downsampling Blockchain Algorithm

Qin Huang，Li Quan，Shengli Zhang
Beihang University

2019.04.29

Content

Content

Background

◆ Definition

Blockchain technology is a distributed ledger that cryptographically secures records of transactions [1]. It provides a secure distributed database.

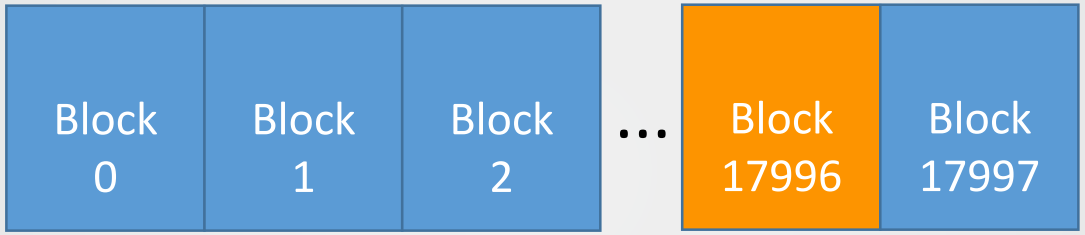

Fig. 1. The structure of blockchain.

◆ Characteristics

Highly-redundant storage, time-series, non-falsification, non-forgery, distributed credit, smart contract and privacy protection.

[1] A. Narayanan, J. Bonneau, E. Felten, A. Miller, and S. Goldfeder, Bitcoin and Cryptocurrency Technologies: A Comprehensive Introduction. Princeton University Press, 2016.

Background

◆ Problems

Storage bloating:

- About 250,000 transactions per day;
- About 50GB per year;
- More than 190GB data storage now.

Network routing:

In order to verify transactions and broadcast, each node needs to save all the data of the blockchain.

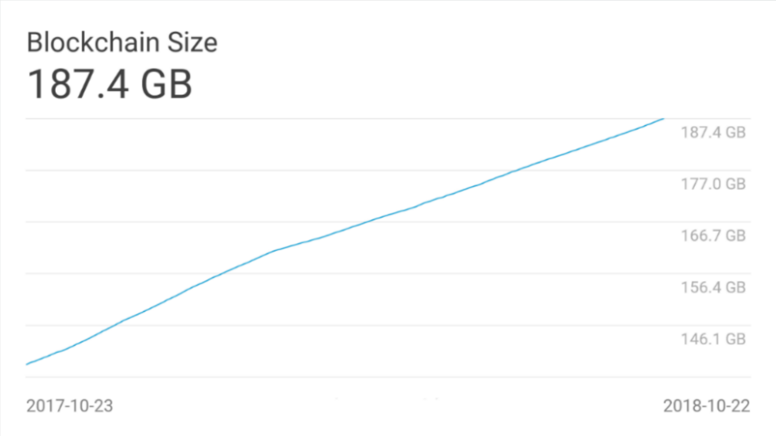

Fig. 2. The blockchain size of Bitcoin [2].

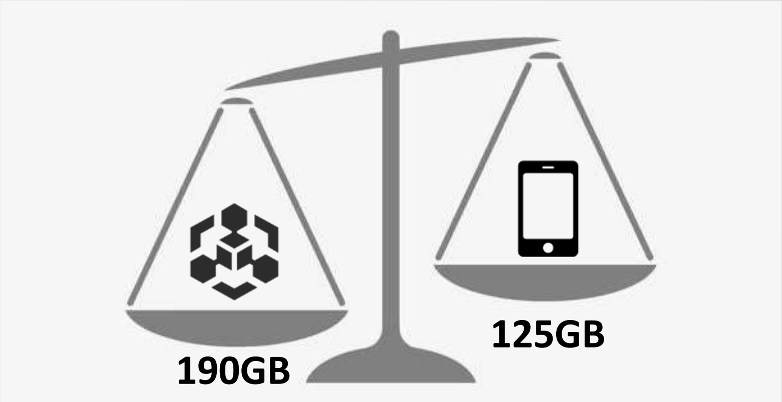

Fig. 3. Comparison of blockchain size and mobile phone storage capacity.

[2] Blockchain Monitoring Website [Online], available: https://blockchain.info/, October 23, 2018.

Requirement: reduce the storage redundancy of nodes for the use of mobile devices and the IoT.

Background

◆ Contributions

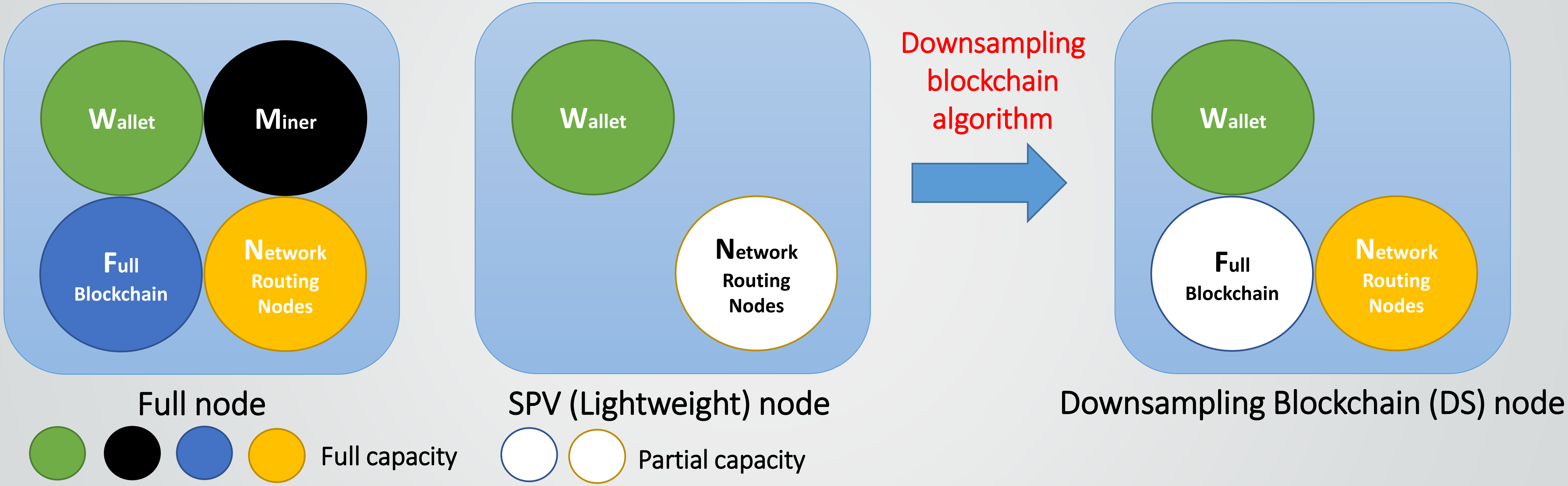

Fig. 4. Different types of nodes on the extended bitcoin network [3].

[3] A. M. Antonopoulos, Mastering Bitcoin: unlocking digital cryptocurrencies. O'Reilly Media, Inc., 2014.

DS node: broadcast transactions, be more secure, reduce the workload of the full node.

Content

Downsampling blockchain algorithm

1. Verify and broadcast transactions
2. Estimate the block where the most recent state is located
3. Get elastic storage size and broadcast accuracy

Downsampling blockchain algorithm

1. Verify and broadcast new transactions

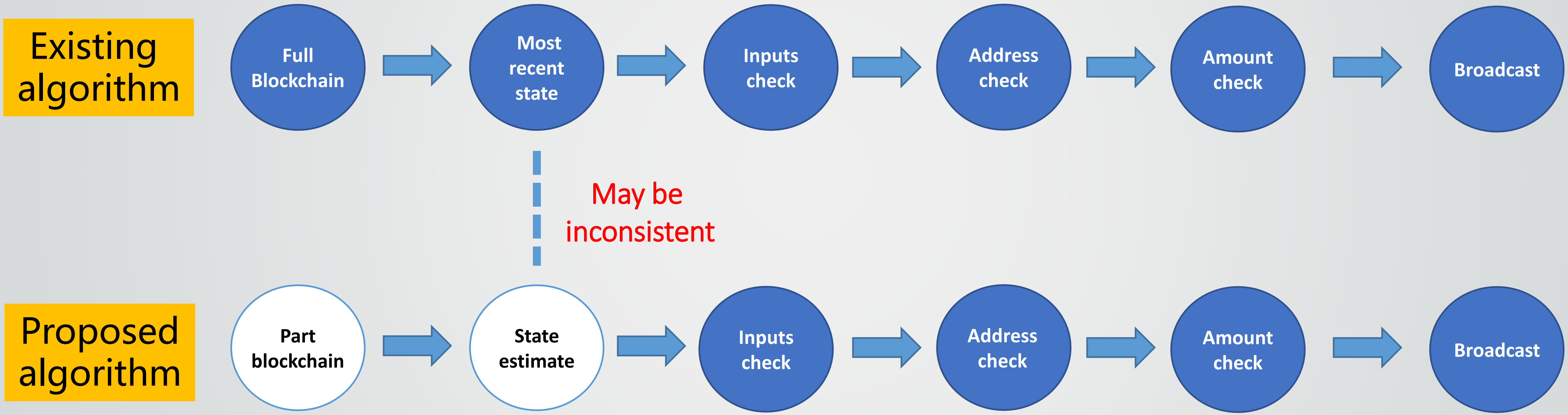

Fig. 5. Process for verifying and broadcasting new transactions.

Downsampling blockchain algorithm

2. Estimate the block where the most recent state is located

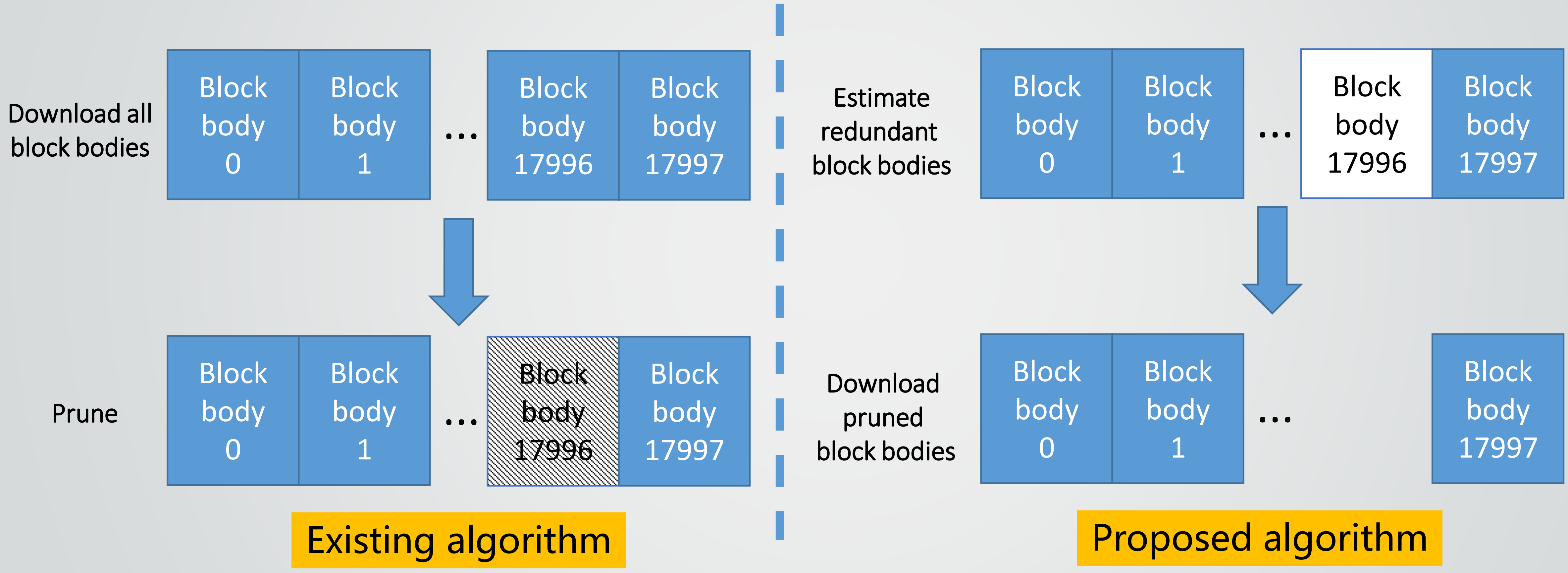

Fig. 6. Process for downsampling blockchain.

Predicting the most recent state distribution based on block information entropy estimation.

Downsampling blockchain algorithm

3. Get elastic storage size and broadcast accuracy

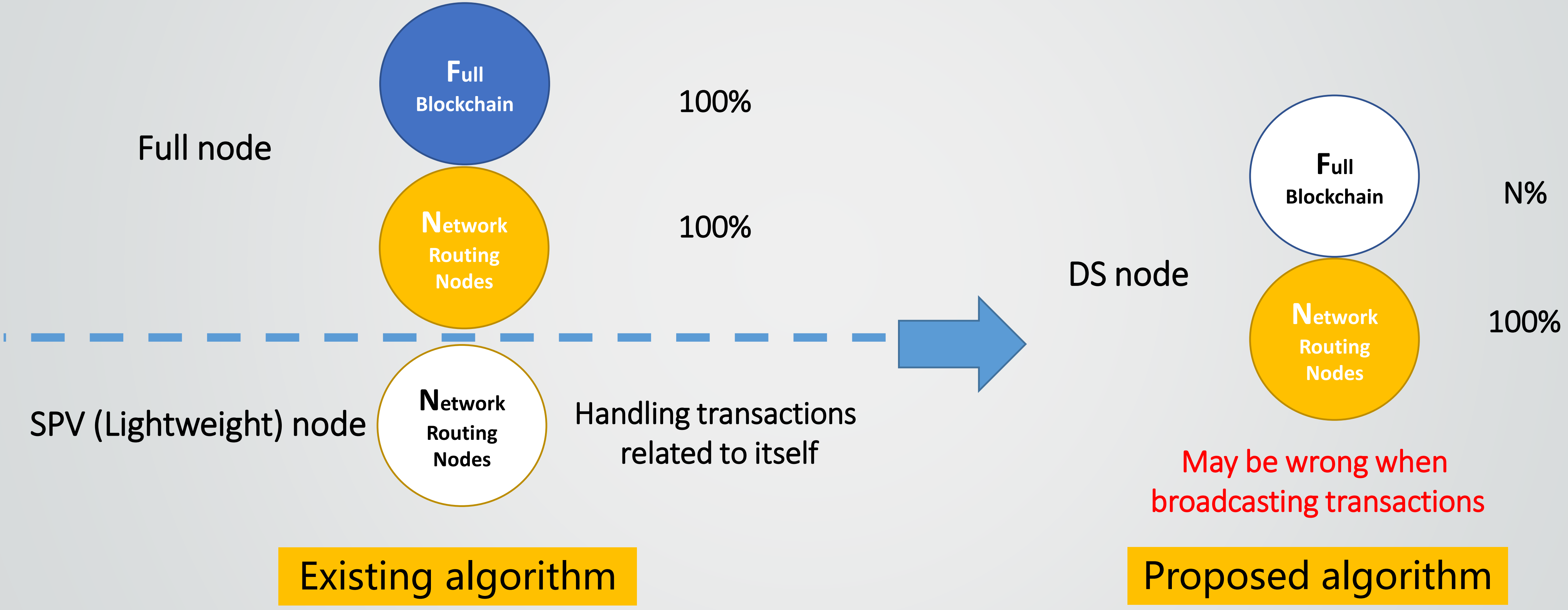

Fig. 7. Routing capability of different nodes.

Partial blockchain, full routing capability, trade-off between broadcast accuracy and storage.

Downsampling blockchain algorithm

Determine downsampling factor → Calculate base → Obtain information entropy → Choose reserved set → Download block bodies

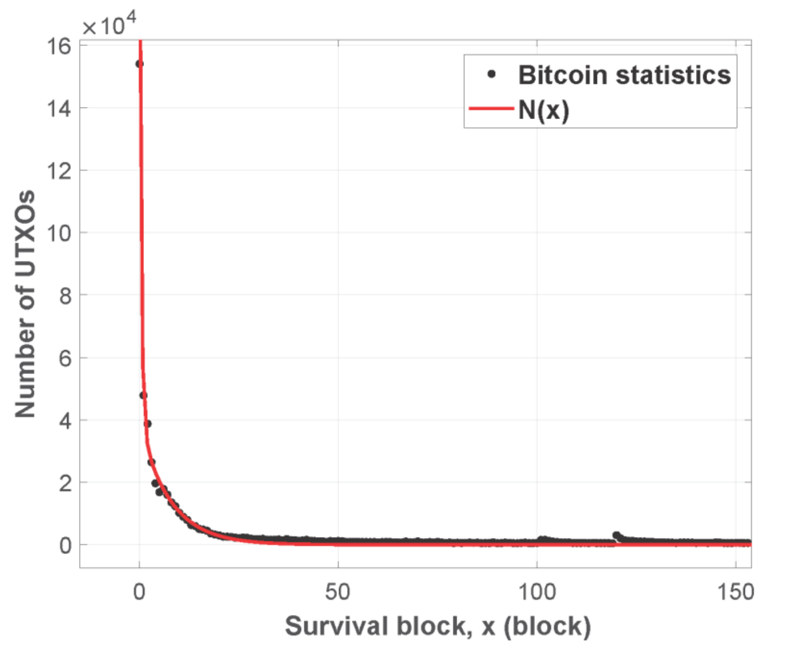

Fig. 8. UTXOs' survival block.

- **Definition 1.** Reserved set is the set of δ blocks with the largest information entropy.

- **Definition 2.** Survival block is the number of blocks that states have been sustained. The survival block of the most recent state reflects the inherent rules of the most recent state.

Cumulative distribution → Probability density → Information entropy

Content

Analysis and simulation

1. Performance analysis

- **Definition 3.** The broadcast accuracy, denoted by φ, is the probability that a node broadcasts valid transactions.

- **Definition 4.** The storage efficiency, denoted by R, is broadcast accuracy storage data size ratio.

  For a DS node,

$$\varphi_{\mathcal{D}} = \frac{N_{su}}{N_u} \qquad R_{\mathcal{D}} = \frac{\varphi_{\mathcal{D}}}{S_{\mathcal{D}}},$$

where $\mathcal{D} \subseteq \{d_1, d_2, \dots, d_\delta\}$ is reserved set, δ is the number of reserved block bodies and $S_{\mathcal{D}}$ is the total block size of $\mathcal{D}$.

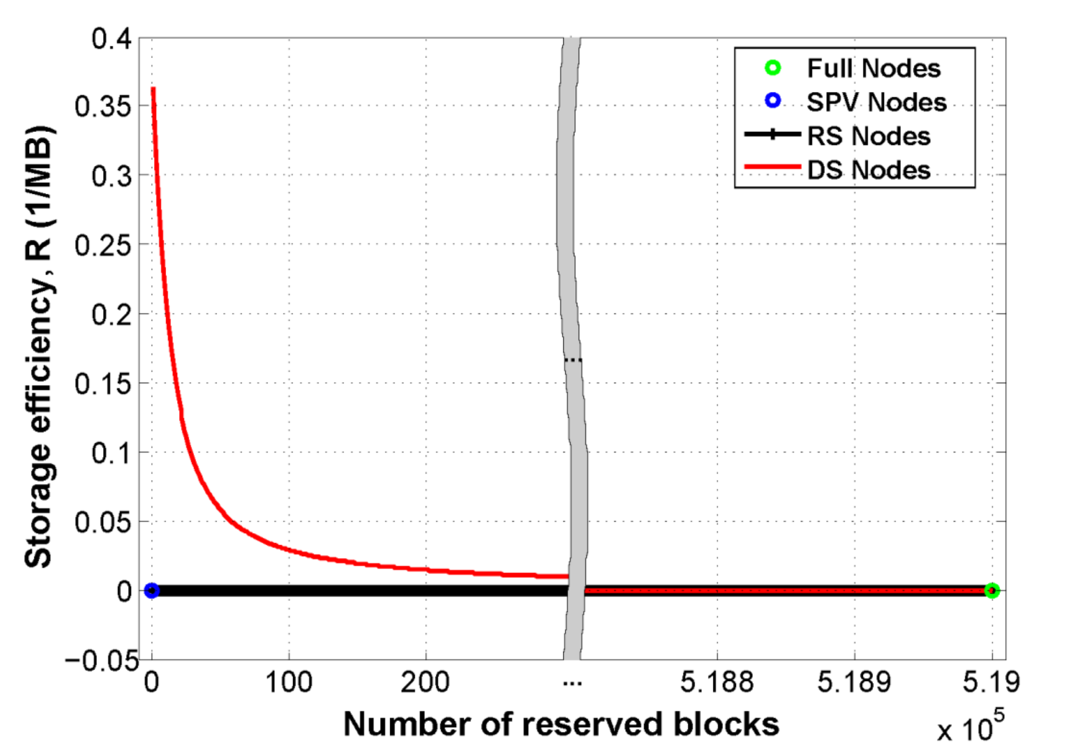

Fig. 9. Storage efficiency of full nodes, Simplified Payment Verification nodes, random sampling nodes, and downsampling nodes.

DS nodes have a better storage efficiency than full nodes, SPV nodes and RS nodes.

Analysis and simulation

2. Complexity analysis

- The selection of blocks in the downsampling blockchain algorithm is a one-time job. Although the block depth is constantly changing, the distribution of the reserved block depth is stable. Thus, the complexity of DS nodes is determined by the number of downloaded block bodies.

TABLE I
AVERAGE DOWNLOADED BLOCKS NUMBER OF FULL NODES , RANDOM SAMPLING NODES , AND DOWNSAMPLING NODES

M	1	16	256	1024	4096
Full Nodes	519000	-	-	-	-
RS Nodes	519000	32438	2028	507	127
DS Nodes	519000	32438	2028	507	127

DS nodes have about $1/M$ average number of downloaded blocks of full nodes.

Analysis and simulation

3. Security analysis

- The blockchain can be viewed as a transaction-based state machine, which begins with a genesis state and incrementally executes transactions to morph it into some final state. Formally $\sigma_{t+1} \equiv \Upsilon(\sigma_t, T)$, where σ_t is the world state at slot t, Υ is the state transition function and T is a transaction. The t_{max} is current largest t, and $\sigma_{t_{max}}$ is the most recent state. We can combine state transition function and transactions, denoted as Υ_t, then $\sigma_{t+1} \equiv \Upsilon_t(\sigma_t)$.
- For an Unspent Transaction Outputs (UTXOs) based blockchain, σ_t can be expressed as $\{UTXO_t^1, UTXO_t^2, \ldots, UTXO_t^n\}$.
- **Lemma 1.** For a Transaction Output ($TXO_{t_0}^x$), we can know that it is UTXO if there are not any $\Upsilon_t s$ changing its state, where $t_0 \le t \le t_{max}$.
- **Lemma 2.** If we know all recent transactions, we can get a set of UTXOs, which is a subset of $\sigma_{t_{max}}$.

The expectation that a DS node is deceived can approach zero, though some valid Ts cannot pass.

Content

Conclusion

DS node

Reduce the storage requirement of nodes with downsampling

1. Verify and broadcast transactions;
2. Estimate the block where the most recent state is located;
3. Get elastic storage size and broadcast accuracy.

Routing capability

Safer than SPV node

Reduce the workload of full node

More flexible and stable network

Thank you!