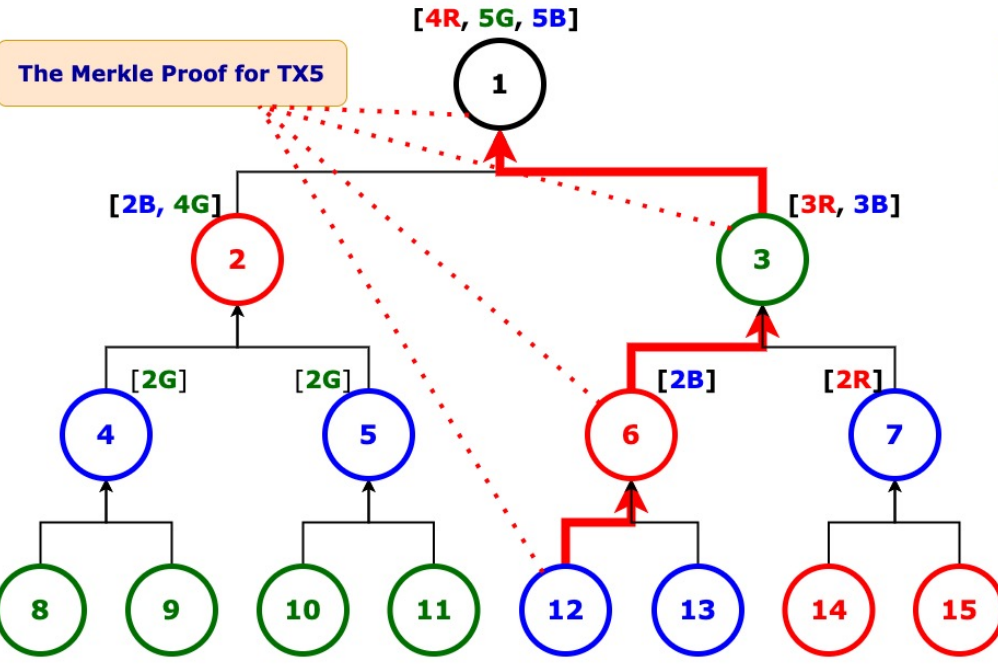


# Private Retrieval of Merkle Proofs for Light Clients with Balanced Sets in Perfect Merkle Tree

## Abstract

Protecting the privacy of light clients (e.g., Simple Payment Verification (SPV) clients in Bitcoin) is crucial to developing any blockchain system. Privacy exists in many forms, but in this work, we are interested in **privacy mechanisms that allow a light client to retrieve a Merkle proof for a particular transaction** (allowing the client to verify if that transaction is indeed included in a block) **without revealing the transaction itself to the full node(s)**. We propose a **parallel scheme that allows private retrieval of Merkle proofs for light clients based on a novel concept called balanced ancestral coloring of binary trees**. This scheme provides completed privacy for SPV clients and is much more efficient than the baseline solutions. The results of our study are well suited to empowering clients' privacy for any database relying on Merkle trees and Merkle proofs such as Bitcoin, Ethereum, DynamoDB, and Certificate Transparency.

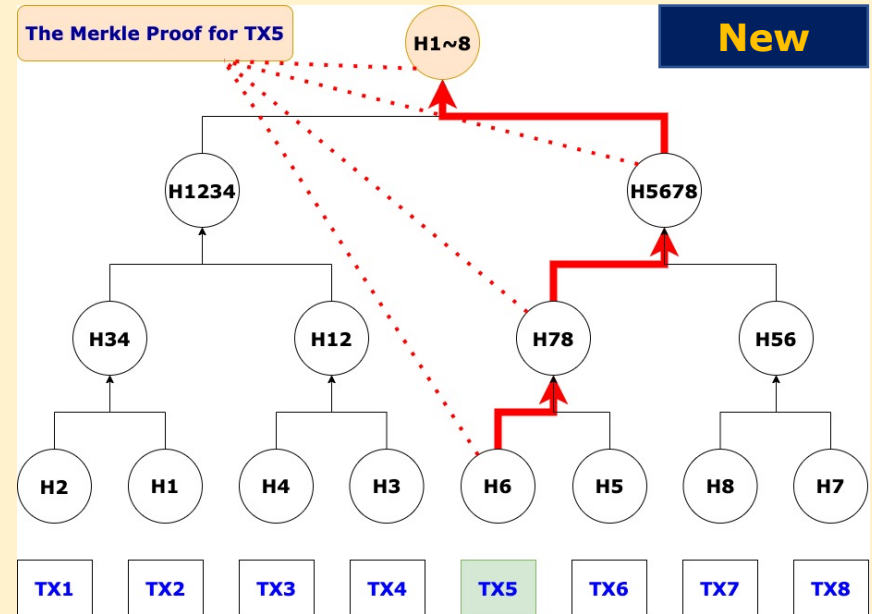
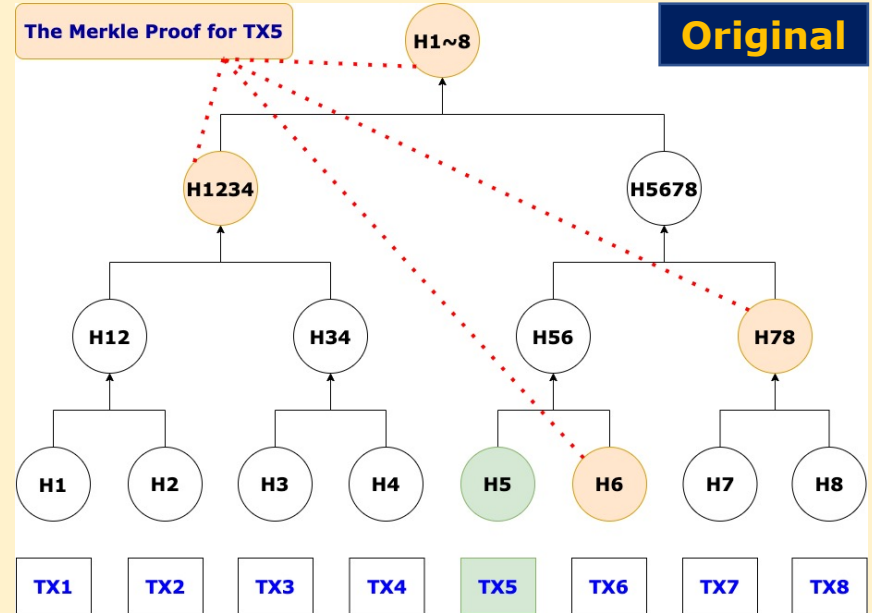
## A Divide-and-Conquer Algorithm Called Color-Splitting Algorithm



Balanced Sets
Red: {2, 6, 14, 15}
Green: {3, 8, 9, 10, 11}
Blue: {4, 5, 7, 12, 13}

- Conditions:**
- Any two nodes that are ancestor and descendant of each other must have different colors.
  - The numbers of nodes in any two distinct color classes differ by at most one.

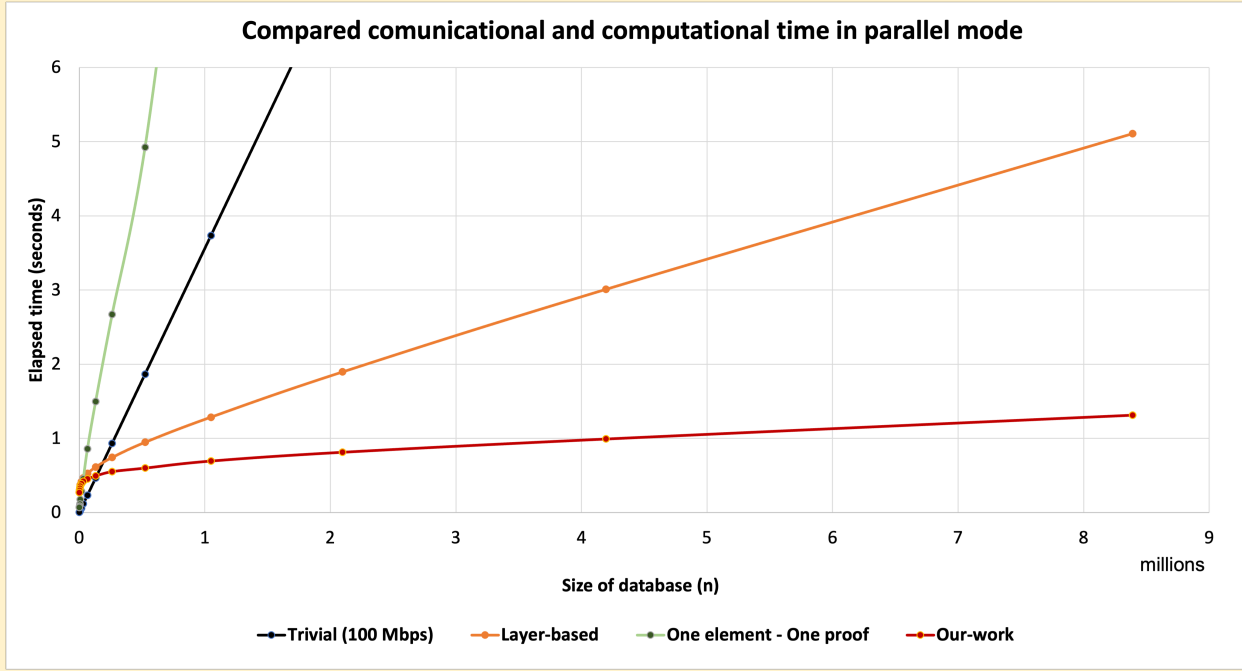
An illustration of the Color-Splitting Algorithm being applied to T(3) and the initial color configuration  $c = [4, 5, 5]$ . A client only needs to privately download one item from each set to retrieve any Merkle Proof privately.



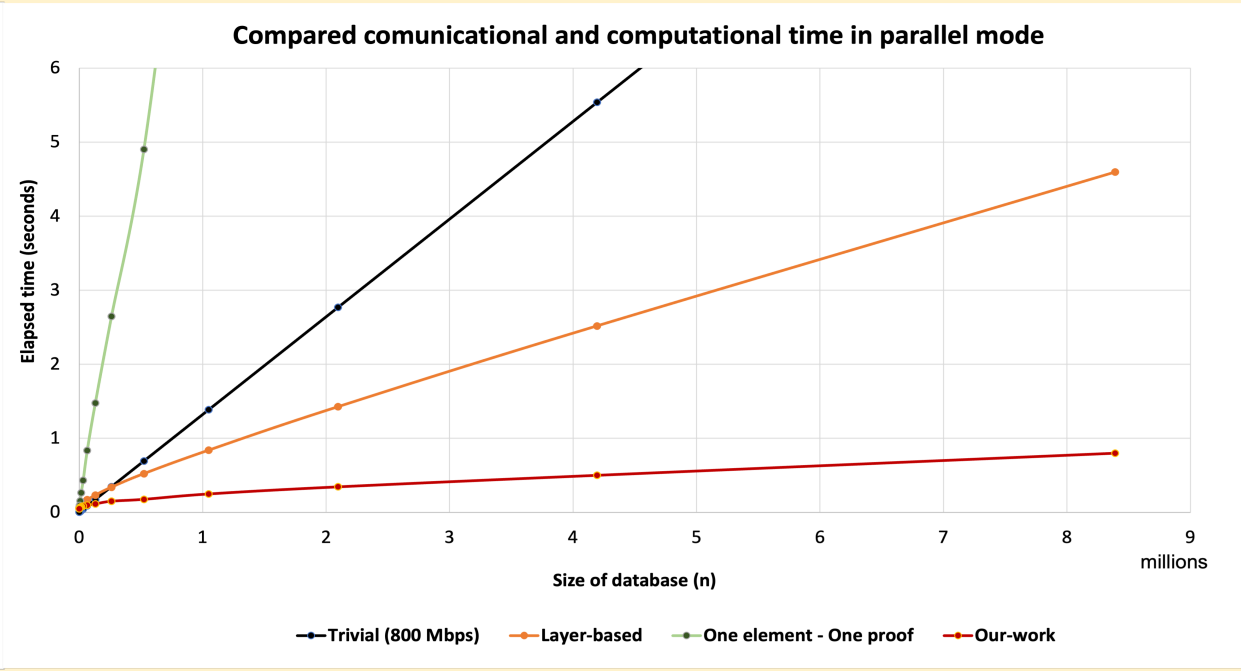
Transformation

# Private Retrieval of Merkle Proofs for Light Clients with Balanced Sets in Perfect Merkle Tree

## An Efficient Parallel Implementation of SealPIR Based on Balanced Sets



The comparison between four types of layer-based, balanced partition based, trivial solution and proof as item in the network bandwidth **100 Mbps**. **From  $n = 2^{17}$ , our solution beats the trivial solution.**



The comparison between four types of layer-based, balanced partition based, trivial solution and proof as item in the network bandwidth **800 Mbps**. **From  $n = 2^{18}$ , our solution beats the trivial solution.**