Rsync Block Matching Complexity Proof

Shyamal Suhana Chandra

1 Theorem: Block Matching Complexity

Statement: Finding matching blocks using rolling checksum in splay tree takes $O(m \log n)$ time where m is number of blocks to match and n is number of stored blocks.

2 Proof

For each of m blocks:

1. Compute rolling checksum: $O(1)$
2. Search in splay tree: $O(\log n)$ amortized
3. Verify strong hash: $O(1)$

Total: $m \times O(\log n) = O(m \log n)$

2.1 With Splay Optimization

Frequently matched blocks move to root, reducing average search time for common patterns.

Conclusion: Block matching has $O(m \log n)$ time complexity, with better average case for common patterns.