LeanIMT: An optimized Incremental Merkle Tree

Privacy & Scaling Explorations

July 20, 2024

1 Abstract

Created: July 20, 2024

Updated: July 20, 2024

Contents

1	Abs	tract			 														1
2	Intr 2.1		on																5
9	Mo	alala Tu																	5
3	3.1		ee																o 5
																			5 5
	3.2	Binary	Tree		 • •	• •	٠	• •	•	• •	• •	•	•	•	•	•	•	•	Э
4	Lea	nIMT			 														5
	4.1	Definit	$tion \dots \dots$		 														5
	4.2	Inserti	on		 														6
		4.2.1	Pseudocode .		 														8
		4.2.2	Correctness .		 														9
		4.2.3	Time complex	xity .	 														10
	4.3	Batch	Insertion	·	 														10
		4.3.1	Pseudocode .																12
		4.3.2	Correctness .																12
		4.3.3	Time complex																12
	4.4	Updat																	14
		4.4.1	Pseudocode																15
		4.4.2	Correctness .																16
		4.4.3	Time complex																16
	4.5		ле																16
	1.0	4.5.1	Pseudocode																17
		4.5.2	Correctness .																17
		4.5.3	Time complex																17
	4.6		ate Merkle Pro																17
	4.0	4.6.1	Pseudocode .																20
		4.6.2	Correctness .																20
		4.6.3	Time complex																20
	4.7		Merkle Proof																$\frac{20}{21}$
	4.1	4.7.1	Pseudocode .																$\frac{21}{23}$
		4.7.1																	23 23
		4.7.3	Correctness .																23 23
		4.7.3	Time complex	xity .	 		•		•			•	•	•	•	٠	•	•	23
5	Imp	lemen	tations		 														23
	5.1	TypeS	cript/JavaScri	pt	 														24
	5.2	Solidit	y		 														24
6	Dom	ahman	ks																24
U	6.1																		
	0.1	0 -	cript/JavaScri	-															24
			Node.js		 		٠		•			٠	٠	•	•	٠	•	•	24

7			nss												
	6.2	2. 2. 2	y			-									
		6.1.5	LeanIMT	: Inser	t Loc	p vs	Batch	Ins	erti	on					27
		6.1.4	Insert Fu	nction	: IMT	vs	LeanII	MT							26
		6.1.3	Node.js v	s Brov	vser .										26

2 Introduction

2.1 Motivation

3 Merkle Tree

3.1 Incremental Merkle Tree

An Incremental Merkle Tree (IMT) is a Merkle Tree (MT) designed to be updated efficiently.

3.2 Binary Tree

A Binary Tree is a tree data structure in which each node has at most two children, referred to as the left child and the right child.

TODO: Explain what is a Merkle tree and an Incremental Merkle Tree.



4 LeanIMT

4.1 Definition

The **LeanIMT** (Lean Incremental Merkle Tree) is a Binary IMT.

The LeanIMT has two properties:

- 1. Every node with two children is the hash of its two child nodes.
- 2. Every node with one child has the same value as its child node.

The tree is always built from the leaves to the root.

The tree will always be balanced by construction.

In a LeanIMT a node is either a leaf or a parent.

Example of a LeanIMT

T - Tree

V - Vertices (Nodes)

E - Edges (Lines connecting Nodes)

$$T = (V, E)$$

$$V = \{a_0, a_1, a_2, H_0, H_1, H_2\}$$

$$E = \{(a_0, H_0), (a_1, H_0), (a_2, a_2), (H_0, H_1), (a_2, a_2)\}\$$



4.2 Insertion

There are two cases:

- 1. When the new node is a left node.
- 2. When the new node is a right node.

We will always see one of these cases in each level when we are inserting a node. It is like, when you insert a node, if that node is left node, the parent node which is in the next level, will be the same node. If it is a right node the parent node, will be the hash of this node with the node in its left. This

algorithm will be the same in each level, not only in level 0.

Case 1: The new node is a left node

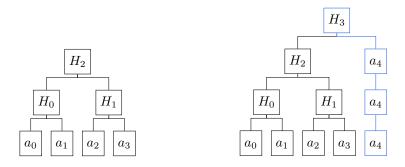
It will not be hashed, it's value will be sent to the next level.

If we add a_4 .

$$T = (V, E)$$

$$V = \{a_0, a_1, a_2, a_3, H_0, H_1, H_2\}$$

$$E = \{(a_0, H_0), (a_1, H_0), (a_2, H_1), (a_3, H_1), (H_0, H_2), (H_1, H_2)\}$$

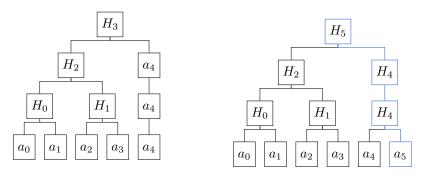


Before inserting a_4

After inserting a_4

Case 2: The new node is a right node

If we add a_5 .



Before inserting a_5

After inserting a₅

4.2.1 Pseudocode

Algorithm 1 LeanIMT Insert algorithm

```
1: procedure Insert(leaf)
        \mathbf{if} \ depth < newDepth \ \mathbf{then}
                                                \triangleright newDepth is the new depth of the tree
    after inserting the new node
             add a new empty array to nodes
                                                                        ▶ Add a new tree level
 3:
        end if
 4:
        node \leftarrow leaf
 5:
        index \leftarrow size \quad \triangleright The index of the new leaf equals the number of leaves
 6:
    in the tree.
 7:
        {\bf for} level from 0 to depth - 1 {\bf do}
             nodes[level][index] \leftarrow node
 8:
             \mathbf{if} \ \mathrm{index} \ \mathbf{is} \ \mathbf{odd} \ \mathbf{then}
                                                                             ▷ It's a right node
 9:
                 sibling \leftarrow nodes[level][index - 1]
10:
                 node \leftarrow hash(sibling, node)
11:
12:
             end if
             index \leftarrow |index/2|
                                               \triangleright Divides the index by 2 and discards the
13:
    remainder.
         end for
14:
        nodes[depth] \leftarrow [node]
                                                     ▷ Store the new root at the top level
15:
16: end procedure
```

4.2.2 Correctness

The idea is to prove that after inserting a new node to the LeanIMT, the tree keeps all the properties.

The Mathematical Induction method will be used to prove the correctness of the algorithm.

n: Number of leaves

T: LeanIMT Tree

n = 0

If the tree is empty, the Insert function returns a new node with the value of the leaf. This satisfies the LeanIMT repoperties because the node does not have children. (Trivial case)

 $n \Rightarrow n+1$

Let's assume that with n nodes, the tree T is a correct LeanIMT and prove that with n + 10 nodes, T is still a correct LeanIMT. (Inductive step)

The new node is always added at the end of the list of nodes.

To insert a new node, there are two possible cases:

- 1. When the new node is a left node
- 2. When the new node is a right node

Case 1

If the new node is a left node, it means that there were an even number of nodes.

Then, sice it's a left node, the parents has only one child and the parent has the same value as the child which is is the new node. Then all the ancestors will be constructed following the two properties of the LeanIMT.

Since T was a correct LeanIMT with n nodes and inserting a new node that is a left node follows the properties of the LeanIMT \Rightarrow the entire tree T is still a correct LeanIMT.

Case 2

If the new node is a right node, it means that there were an odd number of nodes.

Since it is a right node, the value of the parent will be the hash of the left child with the right child which is the new node. Then all the ancestors will be constructed following the two properties of the LeanIMT.

Since T was a correct LeanIMT with n nodes and inserting a new node that is a right node follows the properties of the LeanIMT \Rightarrow the entire tree T is still a correct LeanIMT.

Then for the two cases 1 and 2 the Insert algorithm follows the LeanIMT properties \Rightarrow the Insert algorithm is correct.

4.2.3 Time complexity

n: Number of leaves in the tree.

d: Tree depth.

Every time a new node is added, it is necessary to update or add the ancestors up to the root of the tree.

Number of operations when adding a leaf: [d]

$$\lceil d \rceil \le d + 1$$

$$\le O(\log n) + 1$$

$$\Rightarrow \boxed{O(\log n)}$$

$$d = \lceil \log(n) \rceil \le \log(n) + 1$$

$$\Rightarrow O(\log n)$$

The time complexity of the *Insert* function is $O(\log n)$.

4.3 Batch Insertion

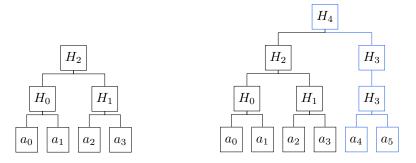
Performing the insertion in bulk rather than individually using a loop can lead to significant performance improvements because the number of hashing operations is reduced.

When inserting n members, all levels will be updated n times if the batch insertion function is not being used.

The core idea behind the batch insertion algorithm is to update each level only once even if there are many members to be inserted.

The algorithm will go through the nodes that are necessary to update the next level of the tree. The other nodes in the tree will not change.

Insert a_4 and a_5 .



Before inserting a_4 and a_5

After inserting a_4 and a_5

4.3.1 Pseudocode

Algorithm 2 LeanIMT InsertMany algorithm

```
1: procedure InsertMany(leaves: List of nodes)
       startIndex \leftarrow |size/2| \triangleright Divides the size of the tree by 2 and discards
    the remainder.
       Add leaves to the tree leaves
3:
       for level from 0 to depth - 1 do
 4:
           numberOfNodes \leftarrow [nodes[level].length/2]
                                                                           ▷ Calculate
    the number of nodes of the next level. numberOfNodes will be the smallest
   integrer which is greater than or equal to the result of dividing the number
    of nodes of the level by 2.
           for index from startIndex to numberOfNodes - 1 do
 6:
               rightNode \leftarrow nodes[level][index * 2 + 1] \triangleright Get the right node if
7:
    exists.
               leftNode \leftarrow nodes[level][index * 2]  \triangleright Get the left node if exists.
8:
               if rightNode exists then
9:
                   parentNode \leftarrow hash(leftNode, rightNode)
10:
               else
11:
                   parentNode \leftarrow leftNode
12:
13:
               nodes[level + 1][index] \leftarrow parentNode \triangleright Add the parent node to
14:
    the tree.
           end for
15:
           startIndex \leftarrow |startIndex/2| \triangleright Divide startIndex by 2 and discards
16:
    the remainder.
17:
       end for
18: end procedure
```

4.3.2 Correctness

4.3.3 Time complexity

n: Number of leaves in the tree.

d: Tree depth.

m: Number of leaves to insert.

Number of operations when inserting elements in batch:

$$\lceil m \rceil + \lceil \frac{m}{2} \rceil + \lceil \frac{m}{4} \rceil + \dots + \lceil \frac{m}{2^d} \rceil$$

That is the same as $\sum_{k=0}^{d} \lceil \frac{m}{2^k} \rceil$

$$\lceil m \rceil \leq m+1 \text{ then } \lceil \frac{m}{2^k} \rceil \leq \frac{m}{2^k} + 1$$

$$\sum_{k=0}^d \lceil \frac{m}{2^k} \rceil \leq \sum_{k=0}^d (\frac{m}{2^k} + 1)$$

$$\leq \sum_{k=0}^d \frac{m}{2^k} + \sum_{k=0}^d 1$$

$$\leq 2m + O(\log(n+m))$$

$$\leq O(m) + O(\log(n+m))$$

$$\Rightarrow \boxed{O(m)}$$

$$\sum_{k=0}^d \frac{m}{2^k} = m \sum_{k=0}^d \frac{1}{2^k} \approx m * 2 \Rightarrow 2m$$

$$\sum_{k=0}^d \frac{1}{2^k} \text{ (Geometric series)}$$

$$|r| < 1; \ r = \frac{1}{2}$$

$$= \frac{1}{2}$$

$$= 2$$

$$\sum_{k=0}^d 1 = \frac{a}{1-r} = d+1$$

$$= O(\log(n+m)) + 1$$

$$\Rightarrow O(\log(n+m))$$

$$d = \lceil \log(n+m) \rceil \leq \log(n+m) + 1$$

$$\Rightarrow O(\log(n+m))$$

Then the time complexity of the InsertMany function is O(m).

Loop Insertion vs Batch Insertion

The time complexity of the Insertion function using a loop is $O(\log(n+m))$.

Going to the root to update or add nodes requires $\log(n+m)$ number of operations.

If we go to the root to update or add nodes m times (one time per leaf to add) then we will have:

$$m * \log(n+m) \Rightarrow O(m\log(n+m))$$

- \Rightarrow Time complexity Loop Insertion is superlinear: $O(m \log(n + m))$
- \Rightarrow Time complexity Batch Insertion (InsertMany function) is linear: O(m) (linear)
- \Rightarrow In terms of time complexity, it is more efficient to use the *InsertMany* function than the *Insert* function in a loop.

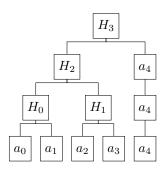
4.4 Update

There are two cases:

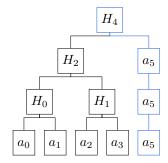
- 1. When there is no right sibling.
- 2. When there is right sibling.

Case 1: There is no right sibling

Update a_4 to a_5



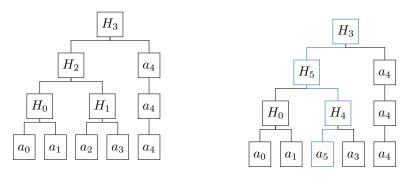
Before updating a_4



After updating a_4

Case 2: There is right sibling

Update a_2 to a_5



Before updating a_2 for a_5

After updating a_2 for a_5

4.4.1 Pseudocode

Algorithm 3 LeanIMT Update algorithm

```
1: procedure UPDATE(index, newLeaf)
 2:
         node \leftarrow newLeaf
         {\bf for} level from 0 to depth - 1 {\bf do}
 3:
             nodes[level][index] \leftarrow node
 4:
             \mathbf{if} \ \mathrm{index} \ \mathbf{is} \ \mathbf{odd} \ \mathbf{then}
                                                                             ▷ It's a right node
 5:
                  sibling \leftarrow nodes[level][index - 1]
 6:
                  node \leftarrow \mathbf{hash}(sibling, node)
 7:
                                                                               \triangleright It's a left node
 8:
             else
                  sibling \leftarrow nodes[level][index + 1]
 9:
                  if sibling exists then
                                                     ▷ It's a left node with a right sibling
10:
                      node \leftarrow hash(node, sibling)
11:
                  end if
12:
             end if
13:
14:
             index \leftarrow |index/2|
                                               ▷ Divides the index by 2 and discards the
    remainder.
         end for
15:
         nodes[depth] \leftarrow [node]
                                                     ▷ Store the new root at the top level
16:
17: end procedure
```

4.4.2 Correctness

4.4.3 Time complexity

n: Number of leaves in the tree.

d: Tree depth.

Every time a leaf is updated, it is necessary to update all the ancestors up to the root of the tree.

Number of operations when updating a leaf: [d]

This proof is the same as the proof of the time complexity of the Insert function.

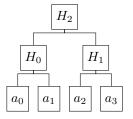
The time complexity of the *Update* function is $O(\log n)$.

4.5 Remove

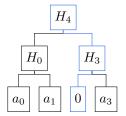
The remove function is the same as the update function but the value used to update is 0.

You can use a value other than 0, the idea is to use a value that is not a possible value for a correct member in the list.

Remove a_2



Before removing a_2



After removing a_2

4.5.1 Pseudocode

Algorithm 4 LeanIMT Remove algorithm

- 1: **procedure** REMOVE(index)
- 2: update(index, 0)
- 3: end procedure

4.5.2 Correctness

4.5.3 Time complexity

4.6 Generate Merkle Proof

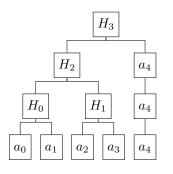
There are two cases:

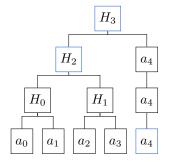
- 1. When the leaf is the last leaf and a left node.
- 2. Other cases
- d: Depth of the tree

We will always see one of these two cases. When you want to generate the proof for the case 1, there will always be one node in the siblings list, for the case 2 there will always be $\lceil d \rceil$ number of nodes.

Case 1: The leaf is the last leaf and a left node

If we want to generate a proof for the node a_4 .





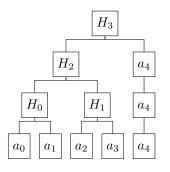
LeanIMT to generate a proof for a_4

Nodes used to generate a proof for a_4

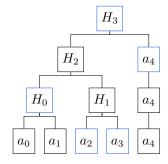
```
path: [1] Merkle Proof: \{ root: H_3 leaf: a_4 index: 1 siblings: [H_2] \}
```

Case 2: Other cases

If we want to generate a proof for the node a_3 .



LeanIMT to generate a proof for a_3



Nodes used to generate a proof for a_3

```
path: [1, 1, 0]
Merkle Proof: \{
root: H_3
leaf: a_3
index: 3
siblings: [a_2, H_0, a_2]
\}
```

4.6.1 Pseudocode

Algorithm 5 LeanIMT generateProof algorithm

```
1: procedure GENERATEPROOF(index)
        siblings \leftarrow empty list \triangleright List to store the nodes necessary to rebuild the
    root.
        path \leftarrow empty list \triangleright List of 0s or 1s to help rebuild the root. 0 if the
3:
    current node is a left node and the sibling is a right node and 1 otherwise.
        for level from 0 to depth - 1 do
 4:
           isRightNode \leftarrow index is odd
5:
           if isRightNode is true then
                                                                     ▷ It's a right node
6:
               siblingIndex \leftarrow index - 1
7:
                                                                      ▷ It's a left node
           else
8:
               siblingIndex \leftarrow index + 1
9:
10:
           end if
           sibling \leftarrow nodes[level][siblingIndex]
11:
           if sibling exists then
12:
               add isRightNode to path
13:
               add sibling to siblings
14:
           end if
15:
16:
           index \leftarrow |index/2|
                                          ▷ Divides the index by 2 and discards the
    remainder.
       end for
17:
       leaf \leftarrow leaves[index]
18:
       index ← reverse path and use the list as a binary number and get the
19:
    decimal representation
20:
        siblings \leftarrow leaves[index]
        proof \leftarrow \{root, leaf, index, siblings\}
21:
       return proof
22:
23: end procedure
```

4.6.2 Correctness

4.6.3 Time complexity

n: Number of leaves in the tree.

d: Tree depth.

To generate a Merkle Proof it is necessary to visit all the ancestors of the leaf up to the root of the tree.

Number of operations to generate a Merkle Proof: [d]

This proof is the same as the proof of the time complexity of the Insert function.

The time complexity of the *generateProof* function is $O(\log n)$.

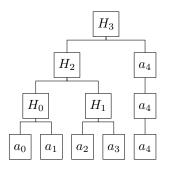
4.7 Verify Merkle Proof

There are two cases:

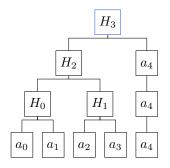
- 1. When the leaf is the last leaf and a left node.
- 2. Other cases

Case 1: The leaf is the last leaf and a left node

If we want to verify a proof for the node a_4 .



LeanIMT to verify a proof for a_4



Nodes rebuilt to verify a proof for a₄

path: [1] Merkle Proof: $\{$ root: H_3

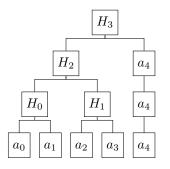
leaf: a_4

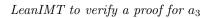
index: 1

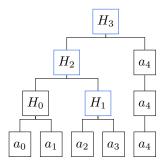
```
siblings: [H_2]
```

Case 2: Other cases

If we want to verify a proof for the node a_3 .







Nodes rebuilt to verify a proof for a_3

```
path: [1, 1, 0]
Merkle Proof: \{
root: H_3
leaf: a_3
index: 3
siblings: [a_2, H_0, a_2]
\}
```

4.7.1 Pseudocode

Algorithm 6 LeanIMT verifyProof algorithm

```
1: procedure VERIFYPROOF(index)
        \{ \text{ root, leaf, siblings, index } \} \leftarrow \text{proof}
                                                                 ▷ Deconstruct the proof
 2:
 3:
        node \leftarrow leaf
        {\bf for}i from 0 to siblings.length - 1 {\bf do}
 4:
            isOdd \leftarrow devide index by 2 i times and check if the result is odd
 5:
            if isOdd is true then
                                                                   \triangleright node is a right child
 6:
 7:
                node \leftarrow hash(siblings[i], node)
            else
                                                                         ▷ It's a left node
 8:
 9:
                node \leftarrow hash(node, siblings[i])
            end if
10:
        end for
11:
        if root is equal node then
12:
13:
            return true
14:
        else
            return false
15:
        end if
16:
17: end procedure
```

4.7.2 Correctness

4.7.3 Time complexity

n: Number of leaves in the tree.

d: Tree depth.

To verify a Merkle Proof it is necessary to visit (rebuild) all the ancestors of the leaf up to the root of the tree.

Number of operations to verify a Merkle Proof: $\lceil d \rceil$

This proof is the same as the proof of the time complexity of the Insert function.

The time complexity of the verifyProof function is $O(\log n)$.

5 Implementations

TODO: Explain TypeScript and Solidity implementations.

5.1 TypeScript/JavaScript

5.2 Solidity

6 Benchmarks

6.1 TypeScript/JavaScript

6.1.1 Node.js

Table 1: All Functions (100 iterations)

Function	ops/sec	Average Time (ms)	Relative to IMT
IMT - Insert	1287	0.77687	
LeanIMT - Insert	2358	0.42391	$1.83 \times faster$
IMT - $InsertMany$	12	77.98467	
LeanIMT - $InsertMany$	144	6.94025	11.24 x faster
IMT - Update	1283	0.77933	
LeanIMT - Update	1223	0.81708	$1.05 \times \text{slower}$
IMT - Remove	1306	0.76554	
LeanIMT - Remove	1301	0.76838	$1.00 \times \text{slower}$
IMT - GenerateProof	300868	0.00332	
LeanIMT - GenerateProof	321586	0.00311	1.07 x faster
IMT - VerifyProof	1331	0.75121	
LeanIMT - VerifyProof	1336	0.74810	1.00 x faster

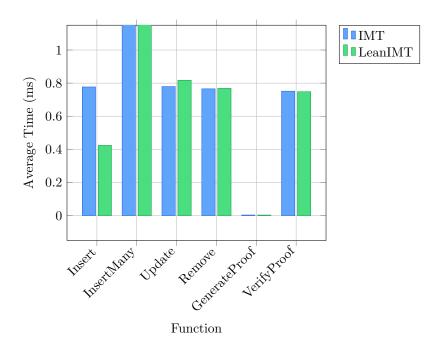


Figure 1: Functions IMT vs LeanIMT (100 iterations)

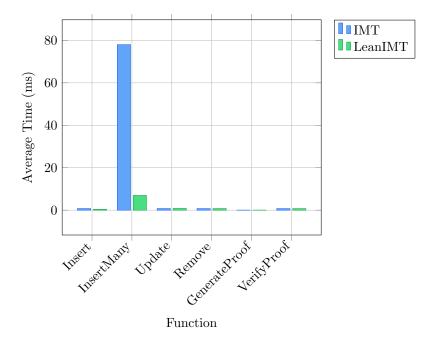


Figure 2: Functions IMT vs LeanIMT (100 iterations)

6.1.2 Browser

6.1.3 Node.js vs Browser

6.1.4 Insert Function: IMT vs LeanIMT

Table 2: Insert Function (1000 iterations)

Function	ops/sec	Average Time (ms)	Relative to IMT
IMT	814	1.22803	
LeanIMT	1453	0.68790	1.79 x faster

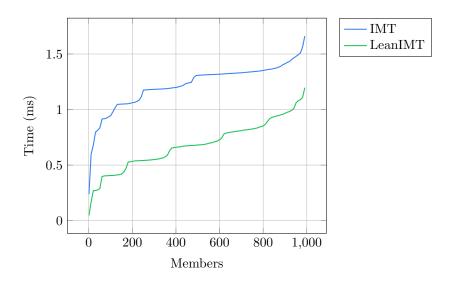


Figure 3: Insert function IMT vs LeanIMT (1000 iterations)

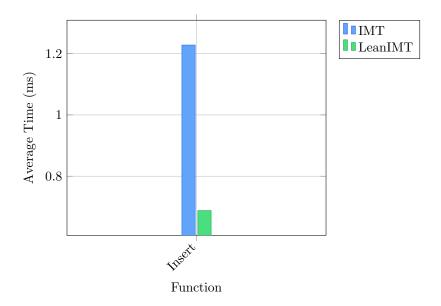


Figure 4: Insert function IMT vs LeanIMT (1000 iterations)

6.1.5 LeanIMT: Insert Loop vs Batch Insertion

Table 3: Insert Function (1000 iterations)

		\	
Function	${\rm ops/sec}$	Average Time (ms)	Relative to Insert
Insert in Loop	47	20.97820	
${\bf InsertMany}$	136	7.31698	$2.87 \times faster$

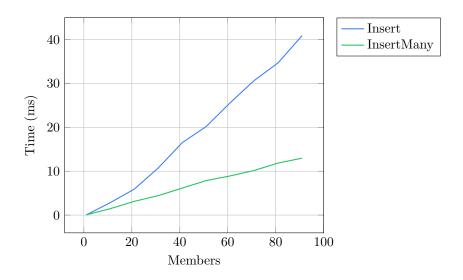


Figure 5: Batch Insertion LeanIMT

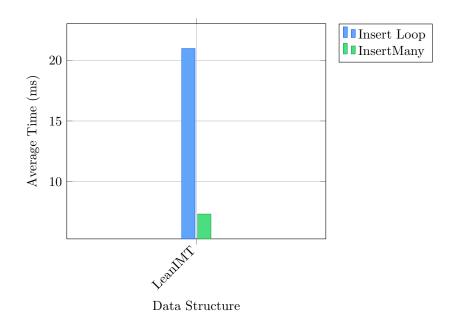


Figure 6: Batch Insertion LeanIMT (1000 iterations)

6.2 Solidity

Solidity and Network Configuration										
Solidity: 0.8.23	Optim: true		viaIR: false	Block: 30,00	00,000 gas					
Methods					İ					
Contracts / Methods	Min	Max	Avg	# calls	usd (avg)					
BinaryIMTTest					İ					
init	105,471	374,307	357,505	16	- I					
initWithDefaultZeroes	91,272	91,870	91,471	3	-					
insert	98,112	2,501,619	560,351	31	-					
remove	471,034	473,216	472,710	7	-					
update	_		474,000	1	-					
Deployments				% of limit	ı					
BinaryIMT	1,237,933	1,238,005	1,237,998	4.1 %	-					
BinaryIMTTest	378,277	378,337	378,329	1.3 %	-					
PoseidonT3	_		3,693,362	12.3 %	- 1					
Key										
\circ Execution gas for this method does not include intrinsic gas overhead										
△ Cost was non-zero but be	\mid $_{ riangle}$ Cost was non-zero but below the precision setting for the currency display (see options)									
Toolchain: hardhat	Toolchain: hardhat									

Figure 7: IMT Gas Report

Solidity and Network	[····· <u>·</u> ·····								
Solidity: 0.8.23	Optim: true	Runs: 200	viaIR: false	Block: 30,00	00,000 gas				
Methods									
Contracts / Methods	Min	Max	Avg	# calls	usd (avg)				
LeanIMTTest		·····		·····					
insert	93,938	163,708	119,051	47					
insertMany	95,891	715,164	322,619	7	-				
remove	104,558	296,279	233,235	13	-				
update	58,909	252,738	197,830	8	-				
Deployments				% of limit					
LeanIMT	1,018,010	1,018,082	1,018,077	3.4 %	-				
LeanIMTTest	455,827	455,911	455,908	1.5 %					
PoseidonT3	_	_	3,693,362	12.3 %					
Key					i				
\circ Execution gas for this method does not include intrinsic gas overhead									
△ Cost was non-zero	but below the prec	ision setting for	the currency disp	olay (see options))				
Toolchain: hardhat					l				

Figure 8: LeanIMT Gas Report

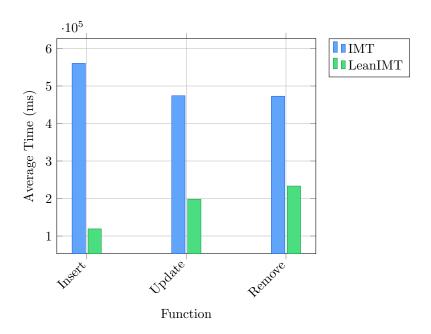


Figure 9: Gas cost of the execution of the Functions IMT vs LeanIMT $\,$

7 Conclusionss

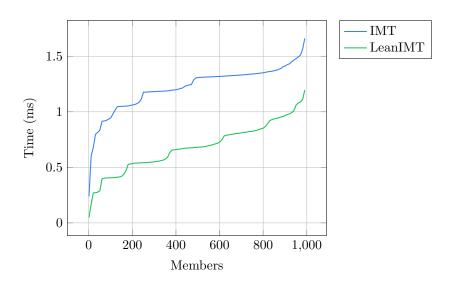


Figure 10: Insert function IMT vs LeanIMT

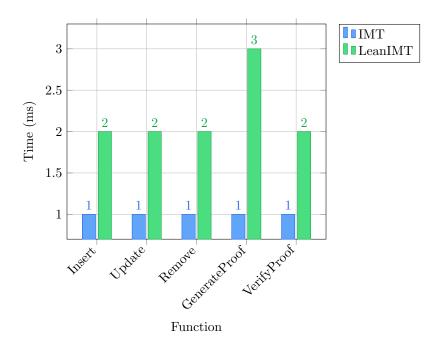


Figure 11: Insert function IMT vs LeanIMT

Table 4: Insert Function (1000 iterations)

Function	ops/sec	Average Time (ms)	Relative to IMT
IMT	814	1.22803	
LeanIMT	1453	0.68790	1.79 x faster

This document is based on the work of [1].

References

[1] Barry Whitehat Kobi Gurkan Koh Wei Jie. "Semaphore: Zero-Knowledge Signaling on Ethereum". In: (2020). URL:

https://semaphore.pse.dev/whitepaper-v1.pdf.