Two Shuffles Make a RAM: Improved Constant Overhead Zero Knowledge RAM

Yibin Yang, Georgia Tech David Heath, UIUC

Witness

Witness

This puzzle has a solution

Witness

This puzzle
has a solution
contract that a solution has a solution

Completeness

V always accepts valid statements

Zero-Knowledge Proof [GMR85]

 $5|3$

6

 $\overline{8}$

Completeness

V always accepts valid statements

Soundness

If P does not have a witness, V rejects statements with high probability

Completeness

V always accepts valid statements

Soundness

If P does not have a witness, V rejects statements with high probability

Zero Knowledge

V learns nothing except that the statement is valid

Generic Zero-Knowledge Proof

Generic Zero-Knowledge Proof

Generic Zero-Knowledge Proof

\bullet mem[i <- x]

mem[i <- x]

ACCESS mem o**i** O x o y r/w if $r/w = 0$: mem else: mem[i <- x]

ACCESS mem o l Q $x \circ y$ mem[i] r/w if $r/w = 0$: mem else: mem[i <- x]

 m [$i < x$]

Our ZK RAM in Recent Works ZK CPU [**Y**HHKV, CCS '24] ZK ML [HCLWZYZ, USENIX Security '24]

completeness soundness zero-knowledge

+

Over some large field (i.e. $|F| = \lambda^{\omega(1)}$)

completeness soundness zero-knowledge

+

Over some large field $(i.e. |F| = \lambda^{\omega(1)})$

completeness soundness zero-knowledge

Nonlinear gates: MUL, INPUT **Linear gates:** ADD, SCALE

+

Over some large field $(i.e. |F| = \lambda^{\omega(1)})$

completeness soundness zero-knowledge

Nonlinear gates: MUL, INPUT **Linear gates:** ADD, SCALE

+

Over some large field $(i.e. |F| = \lambda^{\omega(1)})$

completeness soundness zero-knowledge

Nonlinear gates: MUL, INPUT **Linear gates:** ADD, SCALE

memory slots: n **# accesses:** T

Naïve linear scan: $O(Tn)$

Our Result

Naïve linear scan: *O*(*Tn*)

Our construction: *O*(*T* + *n*)

v.s.

Naïve linear scan: *O*(*Tn*)

Our construction: *O*(*T* + *n*)

Hidden constant: 10

v.s.

Naïve linear scan: *O*(*Tn*)

Our construction: *O*(*T* + *n*)

Hidden constant: 10

Read/Write Memory

Constant-Overhead Zero-Knowledge for RAM Programs

Nicholas Franzese* Jonathan Katz[†] Steve Lu[‡] Rafail Ostrovsky[§] Xiao Wang* Chenkai Weng*

Abstract

We show a constant-overhead interactive zero-knowledge (ZK) proof system for RAM programs, that is, a ZK proof in which the communication complexity as well as the running times of the prover and verifier scale linearly in the size of the memory N and the running time T of the underlying RAM program. Besides yielding an asymptotic improvement of prior work, our implementation gives concrete performance improvements for RAM-based ZK proofs. In particular, our implementation supports ZK proofs of private read/write accesses to 64 MB of memory $(2^{24}$ 32-bit words) using only 34 bytes of communication per access, a more than $80\times$ improvement compared to the recent BubbleRAM protocol. We also design a lightweight RISC CPU that can efficiently emulate the MIPS-I instruction set, and for which our ZK proof communicates only ≈ 320 bytes per cycle, more than $10\times$ less than the BubbleRAM CPU. In a 100 Mbps network, we can perform zero-knowledge executions of our CPU (with 64 MB of main memory and 4 MB of program memory) at a clock rate of 6.6 KHz.

1 Introduction

Zero-knowledge (ZK) proofs enable a prover P to convince a verifier V that the prover knows a witness w on which a particular program P evaluates to 1 without revealing anything additional about w. A series of works over the past several years (e.g., [GGPR13, JKO13, BCC+16, Gro16, AHIV17, KKW18, BBB+18, XZZ+19, TS20, WYKW21]) has shown several highly efficient ZK protocols, however for the most part these improved protocols have focused on the case where the program P is represented as a boolean or arithmetic circuit. This makes such proof systems somewhat difficult to apply to the arguably more natural setting where P is a program intended to be run on a general-purpose CPU, that is, when P is represented as a program in the random-access machine (RAM) model of computation. Although any such program can be converted to a circuit, doing so can be challenging and time consuming; more importantly, it can lead to sub-optimal performance as a general RAM program running in time T and using memory of size N requires a circuit of size $\Theta(TN)$ to verify its execution.

Some prior work has shown ZK proofs in the RAM model of computation. Hu, Mohassel, and Rosulek [HMR15], and subsequently Mohassel, Rosulek and Scafuro [MRS17], proposed an approach in which the addresses of memory accesses are revealed to the verifier; to account for that, they use oblivious RAM to make the original RAM program oblivious. Their focus was on asymptotic performance, and to the best of our knowledge their protocols have never been implemented. The TinyRAM framework $[BCG⁺¹³, BCTV14, WSR⁺¹⁵]$ avoids the use of oblivious

 $\overline{1}$

*Northwestern University, {nicholasfranzese2026@u, wangxiao@cs, ckweng@u}.northwestern.edu

[†]University of Maryland, jkatz2@gmail.com

[‡]Stealth Software Technologies, Inc., steve@stealthsoftwareinc.com

 $$UCLA,$ rafail@cs.ucla.edu

Efficient Proof of RAM Programs from Any Public-Coin Zero-Knowledge System

Cyprien Delpech de Saint Guilhem Emmanuela Orsini Titouan Tanguy Michiel Verbauwhede

> imec-COSIC, KU Leuven, Belgium firstname.lastname@kuleuven.be

Abstract

We show a compiler that allows to prove the correct execution of RAM programs using any zero-knowledge system for circuit satisfiability. At the core of this work is an arithmetic circuit which verifies the consistency of a list of memory access tuples in zero-knowledge. Using such a circuit, we obtain the first constant-round and concretely efficient zero-knowledge

proof protocol for RAM programs using any stateless zero-knowledge proof system for Boolean or arithmetic circuits. Both the communication complexity and the prover and verifier run times asymptotically scale linearly in the size of the memory and the run time of the RAM program; we demonstrate concrete efficiency with performance results of our C++ implementation.

We concretely instantiate our construction with an efficient MPC-in-the-Head proof system, Limbo (ACM CCS 2021). The C++ implementation of our access protocol extends that of Limbo and provides interactive proofs with 40 bits of statistical security with an amortized cost of 0.42ms of prover time and 2.8KB of communication per memory access, independently of the size of the memory; with multi-threading, this cost is reduced to 0.12ms and 1.8KB respectively. This performance of our *public-coin* protocol approaches that of *private-coin* protocol BubbleRAM $(ACMCCS 2020, 0.15ms and 1.5KB per access).$

1 Introduction

A zero-knowledge (ZK) proof is a fundamental cryptographic tool which proves that a statement is true without revealing any other information. Since their introduction by Goldwasser, Micali and Rackoff [GMR85], ZK proofs have had a significant impact on cryptography and have been the object of intense research work due to their theoretical importance and varied applicability.

Many types of ZK proof systems exist, each presenting different trade-offs between several efficiency measures. While in blockchain applications, the main focus is on succinct proofs of small statements [GGPR13, Gro16, Set20], another line of research has focused on prover efficiency [JKO13, AHIV17, KKW18, BCR+19, DIO21, dOT21], while other works have successfully constructed ZK proof systems for very large statements with good concrete efficiency [WYKW21, YSWW21, WYX⁺21, BMRS21].

Unfortunately, these works focus mostly on statements represented as circuits, either Boolean or arithmetic, which can incur a significant overhead to prove properties of large statements that are more naturally represented as random-access machine (RAM) programs. Many interesting functions and applications, such as private database search or verification of program execution,

Constant-Overhead Zero-Knowledge for RAM Programs

Nicholas Franzese* Jonathan Katz[†] Steve Lu[‡] Rafail Ostrovsky[§] Xiao Wang* Chenkai Weng*

Abstract

We show a constant-overhead interactive zero-knowledge (ZK) proof system for RAM programs, that is, a ZK proof in which the communication complexity as well as the running times of the prover and verifier scale linearly in the size of the memory N and the running time T of the underlying RAM program. Besides yielding an asymptotic improvement of prior work, our implementation gives concrete performance improvements for RAM-based ZK proofs. In particular, our implementation supports ZK proofs of private read/write accesses to 64 MB of memory $(2^{24}$ 32-bit words) using only 34 bytes of communication per access, a more than 80x improvement compared to the recent BubbleRAM protocol. We also design a lightweight RISC CPU that can efficiently emulate the MIPS-I instruction set, and for which our ZK proof communicates only ≈ 320 bytes per cycle, more than $10\times$ less than the BubbleRAM CPU. In a 100 Mbps network, we can perform zero-knowledge executions of our CPU (with 64 MB of main memory and 4 MB of program memory) at a clock rate of 6.6 KHz.

1 Introduction

Zero-knowledge (ZK) proofs enable a prover P to convince a verifier V that the prover knows a witness w on which a particular program P evaluates to 1 without revealing anything additional about w. A series of works over the past several years (e.g., [GGPR13, JKO13, BCC+16, Gro16, AHIV17, KKW18, BBB+18, XZZ+19, TS20, WYKW21]) has shown several highly efficient ZK protocols, however for the most part these improved protocols have focused on the case where the program P is represented as a boolean or arithmetic circuit. This makes such proof systems somewhat difficult to apply to the arguably more natural setting where P is a program intended to be run on a general-purpose CPU, that is, when P is represented as a program in the random-access machine (RAM) model of computation. Although any such program can be converted to a circuit, doing so can be challenging and time consuming; more importantly, it can lead to sub-optimal performance as a general RAM program running in time T and using memory of size N requires a circuit of size $\Theta(TN)$ to verify its execution.

Some prior work has shown ZK proofs in the RAM model of computation. Hu, Mohassel, and Rosulek [HMR15], and subsequently Mohassel, Rosulek and Scafuro [MRS17], proposed an approach in which the addresses of memory accesses are revealed to the verifier; to account for that, they use oblivious RAM to make the original RAM program oblivious. Their focus was on asymptotic performance, and to the best of our knowledge their protocols have never been implemented. The TinyRAM framework $[BCG⁺13, BCTV14, WSR⁺15]$ avoids the use of oblivious

*Northwestern University, {nicholasfranzese2026@u, wangxiao@cs, ckweng@u}.northwestern.edu [†]University of Maryland, jkatz2@gmail.com

[‡]Stealth Software Technologies, Inc., steve@stealthsoftwareinc.com

 $$UCLA,$ rafail@cs.ucla.edu

Up to $\approx 20 \times$ improvement

Efficient Proof of RAM Programs from Any Public-Coin Zero-Knowledge System

Cyprien Delpech de Saint Guilhem Emmanuela Orsini Titouan Tanguy Michiel Verbauwhede

> imec-COSIC, KU Leuven, Belgium firstname.lastname@kuleuven.be

Abstract

We show a compiler that allows to prove the correct execution of RAM programs using any zero-knowledge system for circuit satisfiability. At the core of this work is an arithmetic circuit which verifies the consistency of a list of memory access tuples in zero-knowledge. Using such a circuit, we obtain the first constant-round and concretely efficient zero-knowledge

proof protocol for RAM programs using any stateless zero-knowledge proof system for Boolean or arithmetic circuits. Both the communication complexity and the prover and verifier run times asymptotically scale linearly in the size of the memory and the run time of the RAM program; we demonstrate concrete efficiency with performance results of our C++ implementation.

We concretely instantiate our construction with an efficient MPC-in-the-Head proof system, Limbo (ACM CCS 2021). The C++ implementation of our access protocol extends that of Limbo and provides interactive proofs with 40 bits of statistical security with an amortized cost of 0.42ms of prover time and 2.8KB of communication per memory access, independently of the size of the memory; with multi-threading, this cost is reduced to 0.12ms and 1.8KB respectively. This performance of our *public-coin* protocol approaches that of *private-coin* protocol BubbleRAM (ACM CCS 2020, 0.15ms and 1.5KB per access).

1 Introduction

A zero-knowledge (ZK) proof is a fundamental cryptographic tool which proves that a statement is true without revealing any other information. Since their introduction by Goldwasser, Micali and Rackoff [GMR85], ZK proofs have had a significant impact on cryptography and have been the object of intense research work due to their theoretical importance and varied applicability.

Many types of ZK proof systems exist, each presenting different trade-offs between several efficiency measures. While in blockchain applications, the main focus is on succinct proofs of small statements [GGPR13, Gro16, Set20], another line of research has focused on prover efficiency [JKO13, AHIV17, KKW18, BCR+19, DIO21, dOT21], while other works have successfully constructed ZK proof systems for very large statements with good concrete efficiency [WYKW21, YSWW21, WYX⁺21, BMRS21]

Unfortunately, these works focus mostly on statements represented as circuits, either Boolean or arithmetic, which can incur a significant overhead to prove properties of large statements that are more naturally represented as random-access machine (RAM) programs. Many interesting functions and applications, such as private database search or verification of program execution,

\approx 3 \times improvement

Two Shuffles Make a RAM: Improved Constant Overhead Zero Knowledge RAM

Yibin Yang* David Heath[†]

July 17, 2023

Abstract

We optimize Zero Knowledge (ZK) proofs of statements expressed as RAM programs over arithmetic values. Our arithmetic-circuit-based read/write memory uses only 4 input gates and 6 multiplication gates per memory access. This is an almost $3 \times$ total gate improvement over prior state of the art (Delpech de Saint Guilhem et al., SCN'22).

We implemented our memory in the context of ZK proofs based on vector oblivious linear evaluation (VOLE), and we further optimize based on techniques available in the VOLE setting. Our experiments show that (1) our total runtime improves over that of the prior best VOLE-ZK RAM (Franzese et al., CCS'21) by up to $20 \times$ and (2) on a typical hardware setup, we can achieve $\approx 600K$ RAM accesses per second.

We also develop improved read-only memory and set ZK data structures. These are used internally in our read/write memory and improve over prior work.

*Georgia Institute of Technology, yyang811@gatech.edu. [†]University of Illinois Urbana-Champaign, daheath@illinois.edu

mem

i

"One shuffle makes a ROM"

Consider vectors

$$
; \text{ of } \mathbb{F} \text{ elements } x \text{, } y
$$

Are \overrightarrow{x} and \overrightarrow{y} related by a permutation?

Consider vectors

Are \overrightarrow{x} and \overrightarrow{y} related by a permutation?

of
$$
F
$$
 elements \vec{x} , \vec{y}

i $q(X) = \prod (X - y_i)$ *i*

Consider vectors

Are \overrightarrow{x} and \overrightarrow{y} related by a permutation?

$$
; \text{ of } \mathbb{F} \text{ elements } x \text{, } y
$$

i $q(X) = \prod (X - y_i)$ *i*

Proving $\overrightarrow{x} \sim \overrightarrow{y}$ uses only $2n - 2$ MUL gates

$p(X) = \prod (X - x_i)$ *i* $q(X) = \prod (X - y_i)$ *i*

Based on the SZDL Lemma [DL78,Sch80,Zip89]

Proof: $p(\alpha) = q(\alpha)$

Proving $\overrightarrow{x} \sim \overrightarrow{y}$ uses only $2n - 2$ MUL gates

$$
\vec{x} \sim \vec{y} \iff p = q
$$

Generalizes to vectors of tuples [DdSGOTV22]

 $LOAD(i)$

LOAD(i)

Key insight: *We just need to check P does not cheat*

LOAD(i)

Key insight: *We just need to check P does not cheat*

How? *Construct the "one-time" readable entries, and create a "reading history log"*

Make vectors of triples: (index, value, version)

P

Make vectors of triples: (index, value, version)

 $2T-2$ MUL gates to check reads \sim writes

Make vectors of triples: (index, value, version)

 $2T-2$ MUL gates to check reads \sim writes

Make vectors of triples: (index, value, version)

 $2T-2$ MUL gates to check reads \sim writes

Make vectors of triples: (index, value, version)

 $2T-2$ MUL gates to check reads \sim writes

Make vectors of triples: (index, value, version)

 $2T-2$ MUL gates to check reads \sim writes

Make vectors of triples: (index, value, version)

 $2T-2$ MUL gates to check reads \sim writes

Make vectors of triples: (index, value, version)

 $2T-2$ MUL gates to check reads \sim writes

Make vectors of triples: (index, value, version)

 $2T-2$ MUL gates to check reads \sim writes

Difference between ROM and RAM:

RAM must prevent P from reading from the future

Difference between ROM and RAM:

RAM must prevent P from reading from the future

How?

Di fference between ROM and RAM:

RAM must prevent P from reading from the future

How?

Instead of "version", consider "timestamp". Then, prove the "time di fference" is positive.

Di fference between ROM and RAM:

RAM must prevent P from reading from the future

How?

How?

Instead of "version", consider "timestamp". Then, prove the "time di fference" is positive.
Key

Di fference between ROM and RAM:

RAM must prevent P from reading from the future

How? Hint: A ZK ROM with 1,2,…,T indexes

How?

Instead of "version", consider "timestamp". Then, prove the "time di fference" is positive.

Read-Only Memory

On an access, P gives two inputs: value and version

After T accesses, permutation check on length n+T vectors

2 INPUTs, 2 MULs per access

We implemented in the VOLE-based ZK setting [DIO21, YSWW21]

~600K random accesses per second on a 1 Gbps LAN

Evaluation

Figure 12: Speedup of our RAM over $[FKL+21]$'s RAM.

Figure 15: Our RAM's speedup over [DdSGOTV22]'s RAM (we optimize [DdSGOTV22]'s RAM).

yyang811@gatech.edu

Thank you! Q&A

- **• ePrint: <https://eprint.iacr.org/2023/1115>**
-

• GitHub:<https://github.com/gconeice/improved-zk-ram>