Crypto 2024 Presentation

Polymath: Groth16 Is Not The Limit Helger Lipmaa, University of Tartu, Estonia

Computation: f Public input (statement) \mathbb{X} $\begin{array}{l|l} \text{Computation:} \ f & \text{Computation:} \ \text{Public input (statement)} \mathbin{\times} & \text{Public input (s)} \ \text{Private input (witness)} \mathbin{\mathbb{W}} & \text{F} & \text{F} \end{array}$

Computation: f
Public input (statement) \mathbb{X}

Computation: f Public input (statement) \mathbb{X} Private input (witness) W

f Computation: f Public input (statement) \mathbb{X}

Computation: f Public input (statement) X Private input (witness) W

Computation: f Public input (statement) \mathbb{X} Private input (witness) W

Computation: f Public input (statement) \mathbb{X} Private input (witness) W

f Computation: f Public input (statement) X srs srs • Knowledge-soundness

• Completeness

Computation: f Public input (statement) \mathbb{X} Private input (witness) W

- Completeness
-
- Zero-knowledge

f Computation: f Public input (statement) X srs srs

• Knowledge-soundness

Computation: f Public input (statement) X Private input (witness) W

- Completeness
-
- Zero-knowledge
-

• Knowledge-soundness • Succinct arguments

srs srs

f Computation: f Public input (statement) X

Landscape

Landscape

Composition: • GKR + **Groth16** • Brakedown + Groth16 • FRI + **Groth16**

 \bullet

…

FRI

Tensor-code-based (Brakedown, Binius,

…)

Landscape Huge progress in zk-SNARK land in last 5 years Groth16 still lands supreme after 8 years • Shortest argument • Fastest verifier

GKR

Good for Both

Composition: • …

• $pp = (p, \mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T, [1], [1]_2, \hat{e})$

- $pp = (p, \mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T, [1], [1]_2, \hat{e})$
	- \mathbb{G}_i are additive abelian groups of large prime order p

- pp = $(p, \mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T, [1]_1, [1]_2, \hat{e})$
	- \mathbb{G}_i are additive abelian groups of large prime order p
	- $[1]_i$ is a generator of \mathbb{G}_i

- pp = $(p, \mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T, [1], [1], \hat{e})$
	- \mathbb{G}_i are additive abelian groups of large prime order p
	- $[1]_i$ is a generator of \mathbb{G}_i
	- $\hat{e}: \mathbb{G}_1 \times \mathbb{G}_2 \to \mathbb{G}_T$ is a bilinear map

- pp = $(p, \mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T, [1]_1, [1]_2, \hat{e})$
	- \mathbb{G}_i are additive abelian groups of large prime order p
	- $[1]$ is a generator of \mathbb{G}
	- $\hat{e}: \mathbb{G}_1 \times \mathbb{G}_2 \to \mathbb{G}_T$ is a bilinear map
	- $\hat{e}([a]_1, [b]_2) = [ab]_T$ for $a, b \in F_p$

- pp = $(p, \mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T, [1]_1, [1]_2, \hat{e})$
	- \mathbb{G}_i are additive abelian groups of large prime order p
	- $[1]$ is a generator of \mathbb{G} .
	- $\hat{e}: \mathbb{G}_1 \times \mathbb{G}_2 \to \mathbb{G}_T$ is a bilinear map
	- $\hat{e}([a]_1, [b]_2) = [ab]_T$ for $a, b \in F_p$
- "Standard" curve for 128-bit security level: BLS12-381

- pp = $(p, \mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T, [1]_1, [1], \hat{e})$
	- \mathbb{G}_i are additive abelian groups of large prime order p
	- $[1]$ is a generator of \mathbb{G}
	- $\hat{e}: \mathbb{G}_1 \times \mathbb{G}_2 \to \mathbb{G}_T$ is a bilinear map
	- $\hat{e}([a]_1, [b]_2) = [ab]_T$ for $a, b \in F_p$
- "Standard" curve for 128-bit security level: BLS12-381

• $\ell(\mathbb{F}_p) = 256$, $\ell(\mathbb{G}_1) = 384$, $\ell(\mathbb{G}_2) = 768$ (bits)

- pp = $(p, \mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T, [1]_1, [1]_2, \hat{e})$
	- \mathbb{G}_i are additive abelian groups of large prime order p
	- $[1]$ is a generator of \mathbb{G}
	- $\hat{e}: \mathbb{G}_1 \times \mathbb{G}_2 \to \mathbb{G}_T$ is a bilinear map
	- $\hat{e}([a]_1, [b]_2) = [ab]_T$ for $a, b \in F_p$
- "Standard" curve for 128-bit security level: BLS12-381
	- $\ell(\mathbb{F}_p) = 256$, $\ell(\mathbb{G}_1) = 384$, $\ell(\mathbb{G}_2) = 768$ (bits)
- Curves for 192-bit security level:

- pp = $(p, \mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T, [1]_1, [1]_2, \hat{e})$
	- \mathbb{G}_i are additive abelian groups of large prime order p
	- $[1]$ is a generator of \mathbb{G}
	- $\hat{e}: \mathbb{G}_1 \times \mathbb{G}_2 \to \mathbb{G}_T$ is a bilinear map
	- $\hat{e}([a]_1, [b]_2) = [ab]_T$ for $a, b \in F_p$
- "Standard" curve for 128-bit security level: BLS12-381

• $\ell(\mathbb{F}_p) = 256$, $\ell(\mathbb{G}_1) = 384$, $\ell(\mathbb{G}_2) = 768$ (bits)

• Curves for 192-bit security level:

• $\ell(\mathbb{F}_p) = 256$, $\ell(\mathbb{G}_1) = 512$, $\ell(\mathbb{G}_2) = 2048$ (bits)

Groth16: Bird's-Eye View

srs(*f*) srs(*f*)

Computation: f Public input (statement) \mathbb{X} Private input (witness) W

• SRS depends on the circuit

Computation: f Public input (statement) X Private input (witness) W

• SRS depends on the circuit • Argument length: only 3 group elements

Computation: f Public input (statement) X Private input (witness) W

- SRS depends on the circuit
- Argument length: only 3 group elements
- Verifier executes three pairings and $|\mathbb{X}|$ group ops

• Groth16 has **three** group elements

- Groth16 has **three** group elements
- Lower bound [Groth, EC16]:

- Groth16 has **three** group elements
- Lower bound [Groth, EC16]:
	- **At least two group elements** needed

- Groth16 has **three** group elements
- Lower bound [Groth, EC16]:
	- **At least two group elements** needed
	- One of them has to be in \mathbb{G}_2

- Groth16 has **three** group elements
- Lower bound [Groth, EC16]:
	- **At least two group elements** needed
	- One of them has to be in \mathbb{G}_2
- Using a different arithmetization, one can have **two** group elements

- Groth16 has **three** group elements
- Lower bound [Groth, EC16]:
	- **At least two group elements** needed
	- One of them has to be in \mathbb{G}_2
- Using a different arithmetization, one can have **two** group elements
	- **SAP** (Square Arithmetic Programming) instead of **R1CS**

- Groth16 has **three** group elements
- Lower bound [Groth, EC16]:
	- **At least two group elements** needed
	- One of them has to be in \mathbb{G}_2
- Using a different arithmetization, one can have **two** group elements
	- **SAP** (Square Arithmetic Programming) instead of **R1CS**
	- However, then one relies on **symmetric** pairings

- Groth16 has **three** group elements
- Lower bound [Groth, EC16]:
	- **At least two group elements** needed
	- One of them has to be in \mathbb{G}_2
- Using a different arithmetization, one can have **two** group elements • **SAP** (Square Arithmetic Programming) instead of **R1CS** • However, then one relies on **symmetric** pairings • Group elements will be considerably longer => **worse in practice**
	-
	-
	-

- Groth16 has **three** group elements
- Lower bound [Groth, EC16]:
	- **At least two group elements** needed
	- One of them has to be in \mathbb{G}_2
- Using a different arithmetization, one can have **two** group elements
	- **SAP** (Square Arithmetic Programming) instead of **R1CS**
	- However, then one relies on **symmetric** pairings
	- Group elements will be considerably longer => **worse in practice**

• The lower bound works in the **standard model** (no RO)

- Groth16 has **three** group elements
- Lower bound [Groth, EC16]:
	- **At least two group elements** needed
	- One of them has to be in \mathbb{G}_2
- Using a different arithmetization, one can have **two** group elements
	- **SAP** (Square Arithmetic Programming) instead of **R1CS**
	- However, then one relies on **symmetric** pairings
	- Group elements will be considerably longer => **worse in practice**

• The lower bound works in the **standard model** (no RO)

• It talks about **#group elements**, not **bit-length**

Scenic Route to Polymath For non-muggles

• Problem: \mathbb{G}_2 elements are long

- Problem: \mathbb{G}_2 elements are long
- $[b]_2 \longrightarrow [b]_1$, but how?

- Problem: \mathbb{G}_2 elements are long
- $[b]_2 \longrightarrow [b]_1$, but how?
- Groth16 uses pairings to do quadratic checks

- Problem: \mathbb{G}_2 elements are long
- $[b]_2 \longrightarrow [b]_1$, but how?
- Groth16 uses pairings to do quadratic checks
- We can KZG-open the polynomial commitment $[b]_1$ to some b and do quadratic checks by using \bar{b}

$[b]_1$ to some \overline{b}

- Problem: \mathbb{G}_2 elements are long
- $[b]_2 \longrightarrow [b]_1$, but how?
- Groth16 uses pairings to do quadratic checks
- We can KZG-open the polynomial commitment $[b]_1$ to some b and do quadratic checks by using \bar{b}
- KZG opening is shorter than a \mathbb{G}_2 element

$[b]_1$ to some \overline{b}

- Problem: \mathbb{G}_2 elements are long
- $[b]_2 \longrightarrow [b]_1$, but how?
- Groth16 uses pairings to do quadratic checks • We can KZG-open the polynomial commitment $[b]_1$ to some b and do quadratic checks by using $[b]_1$ to some \overline{b} \bar{b}
	- KZG opening is shorter than a \mathbb{G}_2 element
	- (a field element \overline{b} and a \mathbb{G}_1 element $[h]_1$)

- Problem: \mathbb{G}_2 elements are long
- $[b]_2 \longrightarrow [b]_1$, but how?
- Groth16 uses pairings to do quadratic checks • We can KZG-open the polynomial commitment $[b]_1$ to some b and do quadratic checks by using \bar{b}
	- KZG opening is shorter than a \mathbb{G}_2 element
	- (a field element \overline{b} and a \mathbb{G}_1 element $[h]_1$)

Problem:

- we still have $[b]_1$ in the argument!
- $\ell([b]_2) < \ell([b]_1) + \ell(\bar{b}) + \ell([h]_1)$

$[b]_1$ to some \overline{b}

• $\ell([b]_2) < \ell([b]_1) + \ell(b) + \ell([h]_1)$ in 128-bit level

If we use SAP instead of R1CS, we get $[b]_1 = [a]_1$

If we use SAP instead of R1CS, we get $[b]_1 = [a]_1$

• No need to send it!

If we use SAP instead of R1CS, we get $[b]_1 = [a]_1$

- No need to send it!
- **Cost:** circuit ≈ 2 longer => slower prover

If we use SAP instead of R1CS, we get $[b]_1 = [a]_1$

- No need to send it!
- Cost: circuit ≈ 2 longer => slower prover
	- Multiplication gates => squaring gates

If we use SAP instead of R1CS, we get $[b]_1 = [a]_1$

- No need to send it!
- Cost: circuit ≈ 2 longer => slower prover
	- Multiplication gates => squaring gates

Problem:

-
- Not clear how to use KZG

• Groth16 has five trapdoors, KZG is univariate

• Univariatization:

- Univariatization:
	- Replace each trapdoor with x^i for some i and a single trapdoor x

- Univariatization:
	- Replace each trapdoor with $xⁱ$ for some *i* and a single trapdoor x
- Also done in [Lipmaa, PKC 2022] who used exhaustive search to find *i*'s

- Univariatization:
	- Replace each trapdoor with $xⁱ$ for some *i* and a single trapdoor x
- Also done in [Lipmaa, PKC 2022] who used exhaustive search to find *i*'s

Problem:

- => Results in high prover complexity
- $\bullet\,$ even after exhaustive search, the exponents i are quite large • KZG prover time Ω(polynomial degree)
	-

• **Observation 1:** Groth16 for SAP has -1 trapdoor

• **Observation 1:** Groth16 for SAP has -1 trapdoor **• Observation 2:**

- **Observation 1:** Groth16 for SAP has -1 trapdoor **• Observation 2:**
	- One trapdoor in Groth16 is only needed to mask SRS elements corresponding to the statement

- **Observation 1:** Groth16 for SAP has -1 trapdoor **• Observation 2:**
	- One trapdoor in Groth16 is only needed to mask SRS elements corresponding to the statement
- We use a different verification algorithm, getting rid of "statement" trapdoor

- **Observation 1:** Groth16 for SAP has -1 trapdoor **• Observation 2:**
	- One trapdoor in Groth16 is only needed to mask SRS elements corresponding to the statement
	- We use a different verification algorithm, getting rid of "statement" trapdoor
		- Verifier becomes faster

- **Observation 1:** Groth16 for SAP has -1 trapdoor **• Observation 2:**
	- One trapdoor in Groth16 is only needed to mask SRS elements corresponding to the statement

• Verifier becomes faster

• In Polymath, V interpolates a polynomial in \mathbb{F}_p of degree $|\mathbb{X}|$

We use a different verification algorithm, getting rid of "statement" trapdoor

- **Observation 1:** Groth16 for SAP has -1 trapdoor **• Observation 2:**
	- One trapdoor in Groth16 is only needed to mask SRS elements corresponding to the statement

• Verifier becomes faster

- In Polymath, V interpolates a polynomial in \mathbb{F}_p of degree $|\mathbb{X}|$
	- Instead of doing $|X|$ -long MSM in Groth16

We use a different verification algorithm, getting rid of "statement" trapdoor

• We only have three trapdoors

- We only have three trapdoors
- It is easier to choose "good" exponents!

- We only have three trapdoors
- It is easier to choose "good" exponents! We replace two trapdoors with y^{α} and y^{γ} , where $y = x^{\sigma}$

- We only have three trapdoors
- It is easier to choose "good" exponents! We replace two trapdoors with y^{α} and y^{γ} , where $y = x^{\sigma}$
	- *x* is real trapdoor, *α*, *γ*, *σ* are field elements

- We only have three trapdoors
- It is easier to choose "good" exponents! We replace two trapdoors with y^{α} and y^{γ} , where $y = x^{\sigma}$
	- *x* is real trapdoor, *α*, *γ*, *σ* are field elements
- Exhaustive search to find small exponents that result in knowledgesoundness

- We only have three trapdoors
- It is easier to choose "good" exponents! We replace two trapdoors with y^{α} and y^{γ} , where $y = x^{\sigma}$
	- *x* is real trapdoor, *α*, *γ*, *σ* are field elements
- Exhaustive search to find small exponents that result in knowledgesoundness
- Kamek saith: $\alpha = 3$, $\gamma = -5$, $\sigma = n + 3$

- We only have three trapdoors
- It is easier to choose "good" exponents! We replace two trapdoors with y^{α} and y^{γ} , where $y = x^{\sigma}$
	- *x* is real trapdoor, *α*, *γ*, *σ* are field elements
- Exhaustive search to find small exponents that result in knowledgesoundness
- Kamek saith: $\alpha = 3$, $\gamma = -5$, $\sigma = n + 3$

Problem:

- SRS is circuit-dependent
- It does not contain enough elements to compute $|h|_1|$

• This adds elements to the SRS

We add another trapdoor $[z]_1$ that is only used to compute KZG opening

• This adds elements to the SRS

We add another trapdoor $[z]_1$ that is only used to compute KZG opening

• Completeness and zero-knowledge proofs are short and standard

- Completeness and zero-knowledge proofs are short and standard
- Soundness proof is seven pages

- Completeness and zero-knowledge proofs are short and standard
- Soundness proof is seven pages
- 1. Groth16 itself has a complicated proof

- Completeness and zero-knowledge proofs are short and standard
- Soundness proof is seven pages
- 1. Groth16 itself has a complicated proof
- 2. Using virtual trapdoors x^l instead of independent trapdoors makes proof more complicated *xi*

- Completeness and zero-knowledge proofs are short and standard
- Soundness proof is seven pages
- 1. Groth16 itself has a complicated proof
- 2. Using virtual trapdoors x^l instead of independent trapdoors makes proof more complicated *xi*
- 3. To get tight security after Fiat-Shamir, we prove special-soundness
- Completeness and zero-knowledge proofs are short and standard
- Soundness proof is seven pages
- 1. Groth16 itself has a complicated proof
- more complicated *xi*
- 3. To get tight security after Fiat-Shamir, we prove special-soundness
-

2. Using virtual trapdoors x^l instead of independent trapdoors makes proof

4. Proof is in **AGM** with **Oblivious Sampling** [Lipmaa-Parisella-Siim, TCC 2023]

- Completeness and zero-knowledge proofs are short and standard
- Soundness proof is seven pages
- 1. Groth16 itself has a complicated proof
- 2. Using virtual trapdoors x^l instead of independent trapdoors makes proof more complicated *xi*
- 3. To get tight security after Fiat-Shamir, we prove special-soundness 4. Proof is in **AGM** with **Oblivious Sampling** [Lipmaa-Parisella-Siim, TCC 2023] • LPS23 noted that KZG is often used wrongly
-
-

- Completeness and zero-knowledge proofs are short and standard
- Soundness proof is seven pages
- 1. Groth16 itself has a complicated proof
- 2. Using virtual trapdoors x^l instead of independent trapdoors makes proof more complicated *xi*
- 3. To get tight security after Fiat-Shamir, we prove special-soundness 4. Proof is in **AGM** with **Oblivious Sampling** [Lipmaa-Parisella-Siim, TCC 2023] • LPS23 noted that KZG is often used wrongly
-
- - Constructions are secure in AGM but not when adversary can do o.s.

- Completeness and zero-knowledge proofs are short and standard
- Soundness proof is seven pages
- 1. Groth16 itself has a complicated proof
- 2. Using virtual trapdoors x^l instead of independent trapdoors makes proof more complicated *xi*
- 3. To get tight security after Fiat-Shamir, we prove special-soundness 4. Proof is in **AGM** with **Oblivious Sampling** [Lipmaa-Parisella-Siim, TCC 2023] • LPS23 noted that KZG is often used wrongly
-
- - Constructions are secure in AGM but not when adversary can do o.s.
	- And o.s. is for free!

- Completeness and zero-knowledge proofs are short and standard
- Soundness proof is seven pages
- 1. Groth16 itself has a complicated proof
- 2. Using virtual trapdoors x^l instead of independent trapdoors makes proof more complicated *xi*
- 3. To get tight security after Fiat-Shamir, we prove special-soundness 4. Proof is in **AGM** with **Oblivious Sampling** [Lipmaa-Parisella-Siim, TCC 2023] • LPS23 noted that KZG is often used wrongly
-
- - Constructions are secure in AGM but not when adversary can do o.s.
	- And o.s. is for free!

Part of Polymath's proof is machine-checked

• First improvement on Groth16 since 2016

- First improvement on Groth16 since 2016
	-

• Hopefully encourages other researchers to improve Groth16 even further

- First improvement on Groth16 since 2016
	-
- **• Pros:**

• Hopefully encourages other researchers to improve Groth16 even further

- First improvement on Groth16 since 2016
	-
- **• Pros:**
	- Smaller arguments (more prominent in 192/256-bit security level)

- First improvement on Groth16 since 2016
	-
- **• Pros:**
	-
	- Smaller arguments (more prominent in 192/256-bit security level) • No adversarially created \mathbb{G}_2 elements $-$ good for batching

- First improvement on Groth16 since 2016
	-
- **• Pros:**
	-
	-
	- Smaller arguments (more prominent in 192/256-bit security level) • No adversarially created \mathbb{G}_2 elements $-$ good for batching • Verifier is potentially faster for a short public input

- First improvement on Groth16 since 2016
	-
- **• Pros:**
	-
	- Smaller arguments (more prominent in 192/256-bit security level) • No adversarially created \mathbb{G}_2 elements $-$ good for batching • Verifier is potentially faster for a short public input
	-
- And definitely faster for a long public input

• Hopefully encourages other researchers to improve Groth16 even further

- First improvement on Groth16 since 2016
	-
- **• Pros:**
	-
	- Smaller arguments (more prominent in 192/256-bit security level) • No adversarially created \mathbb{G}_2 elements $-$ good for batching
	- Verifier is potentially faster for a short public input
		- And definitely faster for a long public input
- **• Cons:**

• Mitigated in SNARK composition when used as a final SNARK

• Hopefully encourages other researchers to improve Groth16 even further

• Mitigated in SNARK composition when used as a final SNARK • Prover's input is shorter => prover speed less important

- First improvement on Groth16 since 2016
	-
- **• Pros:**
	-
	- Smaller arguments (more prominent in 192/256-bit security level) • No adversarially created \mathbb{G}_2 elements $-$ good for batching • Verifier is potentially faster for a short public input
	-
- And definitely faster for a long public input
- **• Cons:**
	- Prover is slower

• Hopefully encourages other researchers to improve Groth16 even further

- First improvement on Groth16 since 2016
	-
- **• Pros:**
	-
	- Smaller arguments (more prominent in 192/256-bit security level) • No adversarially created \mathbb{G}_2 elements $-$ good for batching
	- Verifier is potentially faster for a short public input
		- And definitely faster for a long public input
- **• Cons:**
	- Prover is slower
	- Uses random oracle model on top of AGM(OS)

• Mitigated in SNARK composition when used as a final SNARK • Prover's input is shorter => prover speed less important • Any known reasonable initial SNARK candidate uses ROM

