Recursive Kate proofs

August 8, 2019

1 Two layer Libert/Kate

Think of our file as a vector $V \in \mathbb{F}_q^N$. We wish to have a "two-layer" version of Kate for proof of retrievability, where the prover's work is only $O(\sqrt{N})$.

Notation/terminology Let $\mathbb{G}$ be the source group of a pairing friendly curve over a field $\mathbb{F}_r$; such that $|\mathbb{G}| = q$. Let $\mathbb{H}$ be a subgroup of a pairing friendly curve of size r, over a field K.

Constructing $\mathbb{G}, \mathbb{H}$ using the zexe curves we have $\log q = 255, \log r = 377, \log |K| = 768$.

and $\mathbb{H}$ be a pairing friendly curve of order r. We use the notation $[n] = \{1, \ldots, n\}$. For a vector $x \in \mathbb{F}_q^n$ and subset $T \subset [n]$ we denote $x_T \in \mathbb{F}_q^{|T|}$ to be x restricted to the indices in T

Libert Commitments We will use the Libert vector commitment for vectors in $\mathbb{F}_q^n$ as a black box. It has the following properties

- A commitment to a vector $(x_1, \ldots, x_n) \in \mathbb{F}_q^n$ is one element of $\mathbb{G}$.
- After a preprocessing taking $O(n^2)$ operations in $\mathbb{G}$, for any subset $S \subset [n]$ of size $t \leq n$, and vector of coefficients $(r_1, \ldots, r_t) \in \mathbb{F}_q^t$, $\sum_{i=1}^t r_i \cdot x_{T(i)}$ can be opened with t group operations and the proof of correct opening is a single element of $\mathbb{G}$.

An important observation about Libert commitments, is

Claim 1.1. *Let $C = \mathsf{Lib}(x_1, \ldots, x_n)$, and $T \subset [n]$ be of size t. Fix $C_T \in \mathbb{G}$. One can construct in $O(t)$ group operations a proof that $C_T = \mathsf{Lib}(x_T)$. The proof consists merely of two $\mathbb{G}$ elements and one $\mathbb{F}_q$ element.*

Proof. We describe the proof. Assume for simplicity of notation that $T = \{1, \ldots, t\}$. The verifier (or using hash of C_T) chooses random $r_1, \ldots, r_t \in \mathbb{F}_q^t$. The prover than opens C with $(r_1, \ldots, r_t, 0, \ldots, 0)$ and C_T with $(r_1, \ldots, r_t)$. If C_T is *not* a commitment to x_T the probability that both open to the same value is $1/q$ (or at most t/q if using the derandomized choice $r_i = r^i$ for the coefficients). $\square$

1.1 The two-layer scheme

The two-layer scheme proceeds as follows.

Commitment to file:

1. Let $n = m := \sqrt{N}$. Split the file into n blocks of m elements. And compute for each block a Kate commitment $g_i := [f_i(s)]_1 \in \mathbb{G}$ where $f_i \in \mathbb{F}_q[X]$ has degree smaller than m.

2. Let $v = (x_1, \ldots, x_n, y_1, \ldots, y_n)$ be the coordinates of $g_1, \ldots, g_n$. Compute a libert commitment C to v. And output C as the commitment to the file.

Remark 1.2. *For simplicity we comit in Step 2 separately to all x, y coordinates. It should be enough to commit just to the sign of y and thus pack many y's into a single field element, thus cutting down the length of v*

Proof of retrievability

1. We have a challenge which is a subset $T \subset [n]$ of blocks, we think of $t := |T|$ as constant. For simplicity of notation assume $T = \{1, \ldots, t\}$. The prover P now wishes to show he knows the contents of the first t blocks.

2. Assume the challenge also includes a uniform elements $\lambda, r \in \mathbb{F}_q$. Denote for $i \in [t], \lambda_i := \lambda^i$. P sends the element $g \in \mathbb{G}$, which is allegedly $\sum_{i \in T} \lambda^i \cdot g_i$. Note that this is exactly the Kate commitment to $f(X) := \sum_{i \in T} \lambda^i \cdot f_i(X)$. P will send a commitment $C_T = \mathsf{Lib}(x_1, \ldots, x_t, y_1, \ldots, y_t)$ and prove it is correct using the protocol from Claim 1.1.

3. Now, let C be a circuit over $\mathbb{F}_r$ whose public inputs are $C_T, g, \lambda_1, \ldots, \lambda_t$; and private inputs are $x_1, \ldots, x_t$,$y_1, \ldots, y_t$ that checks that

   (a) $C_T = \mathsf{Lib}(x_1, \ldots, x_t, y_1, \ldots, y_t)$

   (b) $g = \sum_{i \in T} \lambda_i \cdot (x_i, y_i) = \sum_{i \in T} \lambda_i \cdot g_i$

   P will provide a SNARK proof using the curve $\mathbb{H}$ that C accepts - thus proving that indeed g is the Kate commitment to f.

4. P will now open $f(r)$. Note that if P indeed has the blocks of T stored he can compute the coefficeints of f in $t \cdot n = O(\sqrt{N})$ operations. However, if he is missing any of the coefficients of $\{f_i\}_{i \in T}$ his success probability is negligible.

One important thing to notice is that in the third step the circuit size is $O(t \cdot \log q)$, as we are combining t group elements in it; i.e. there is no dependence on n.

costs and comparissons We estimate the cost of this method in terms of $\mathbb{G}1$ exponentations vs straightforward Kate, and Merkle inclusion proofs. To account also for field operations, and operatios on $\mathbb{H}$, We use the following conversion rates, the first two derived from https://github.com/arielgabizon/pairing/blob/benchresults/benchresults.txt

1. $\mathbb{G}_2$ exp = 3.5 $\mathbb{G}_1$ exp

2. $\mathbb{F}_r$ mult = 1/4300 $\mathbb{G}_1$ exp

We estimate the number of constraints for C will be roughly $1000t$. A rough estimate seems to be a prover work of $7000t$ exponentiations for the proof of C (This is counting a $\mathbb{G}_2$ exponentation as 3.5 $\mathbb{G}_1$ exponentations in the Groth16 scheme. As these are on a curve with twice as many bits, let's counting them as $14000t$. plus another $\sqrt{N}$ for computing the proof for the opening of f.

Thinking of say $t = 200, N = 200000^2 = 40Mil$. we get roughly a seven hundred thoundand exponentiations per opening rather than 40 million using straightforward Kate

2 A version of Kate with $\sqrt{n}$ opening time and $n^{1.5}$ size CRS

The idea is to just sum up independent Kate commitments, but have a CRS that allows us to verifably "narrow down" to one of the commitments during opening.

1. **CRS:** Let $t := \sqrt{n}$. Sample uniform $x_1, \ldots, x_t, \alpha_1, \ldots, \alpha_t \in \mathbb{F}$. Output for each $i \in [t], j \in \{0, \ldots, t-1\}$ x_i^j, and for each $i \in [t] j \in \{0, \ldots, t-1\}, \ell \in [t] \setminus \{i\}$ the element $\alpha_\ell x_i^j$.

2. **Commit:** $\mathsf{cm}(a_1, \ldots, a_n)$: Split the input to t blocks of size t and let $f_i \in \mathbb{F}_{<t}[X]$ be the polynomial whose values/coefficient are the i'th block. Define

$$f(X_1, \ldots, X_t) := \sum_{i \in [t]} f_i(X_i)$$

   The commitment to $\mathsf{cm}(a)$ is $f(x_1, \ldots, x_n)$.

3. $\mathsf{open}(\mathsf{cm}, i, r)$**:** Prover computes $H \in \mathbb{F}_{<t}[X]$ similarly to Kate

$$H(X_i) := (f_i(X_i) - f_i(r))/(X_i - r)$$

   and

$$f_{\neq i}(X_1, \ldots, X_n) := \sum_{i' \neq i} f_i(X_i)$$

   Sends three proof elements:

$$H = [H(x_i)]_1, W = [f_{\neq i}(x_1, \ldots, x_n)]_1, W' = \alpha_i \cdot W$$

4. $\mathsf{ver}(\mathsf{cm}, H, W, W', i, r, s)$: The intiution of verification is to do the regular Kate check on the i'th polynomial with $\mathsf{cm} - W$ as the commitment; and to use a "knowledge check" to verify that $\mathsf{cm} - W$ is really the commitment to f_i - by checking W doesn't have any terms in X_i (and thus $\mathsf{cm} - W$ contains all X_i terms from W).

   To verify that $f_i(r) = s$ check that

$$e(W, \alpha_i) = e(W', [1]_2),$$

   and similarly to Kate:

$$e(\mathsf{cm} - s, [1]_2) = e(H, [X_i - r]_2) \cdot e(W, [1]_2)$$

References