HAETAE, a Post-Quantum Signature Scheme

Jung Hee Cheon[1,2], **Hyeongmin Choe**[1], Julien Devevey[3], Tim Güneysu[4], Dongyeon Hong[2], Markus Krausz[4], Georg Land[4], Junbum Shin[2], Damien Stehlé[2], MinJune Yi[1]

[1]Seoul National University, [2]CryptoLab Inc., [3]École Normale Supérieure de Lyon, [4]Ruhr Universität Bochum,

Korea University
July 24, 2023

HAETAE
HEAAN
CRYPTO LAB

Table of Contents

1. Brief Introduction to HAETAE:
   - HAETAE Intro
   - HAETAE Recap

2. Security Proof:
   - Security Sketch
   - Underlying ID Protocol

3. HAETAE Details:
   - Secret Key Rejection Sampling
   - Uniform Hyperball Sampling

HAETAE Intro

- Digital signature scheme, submitted to KpqC competition and NIST Round 4.
- Secure against quantum attacks
  - based on **lattice hard problems**, MLWE and MSIS
  - follows **Fiat-Shamir with aborts** framework, secure in QROM
- Goal:

Push Fiat-Shamir Signatures to the Limits!

Scheme	Lvl.	Sig.	vk	Const.-T.	Maskable
Falcon-512	1	666B	897B	✓ [Por19]	✗ [Pre23]
Dilithium-2	2	2,420B	1,312B	✓ [DKL+18]	✓ [MGTF19]
HAETAE-120	2	1,463B	992B	✓	✓

Table: NIST security level, signature size, verification key size, and implementation security, with respect to constant-time and masking of selected signature schemes.

HAETAE Intro

- Simple but short
  - simpler than Falcon[1] & shorter than Dilithium[1]
  - optimal rejection rate with simple rejection condition
- Design rationale: We combine the recent approaches,
  - **Fiat-Shamir with Aborts** framework
  - **Bimodal** rejection sampling
  - randomness sampling from **Hyperball** distribution

  with the NEW techniques,
  - secret key rejection sampling: efficient and easily maskable
  - verification key truncation: in bimodal setting
  - signature compression: in hyperball setting
  - discretized hyperball sampling: a fixed-point implementation

[1] NIST 2022 PQC signature standards

1. Brief Introduction to HAETAE:
   - HAETAE Intro
   - HAETAE Recap

2. Security Proof:
   - Security Sketch
   - Underlying ID Protocol

3. HAETAE Details:
   - Secret Key Rejection Sampling
   - Uniform Hyperball Sampling

HAETAE Recap: Sign

- In “Fiat-Shamir with Aborts” signatures, the signing procedure is given as:

  1. $\mathbf{y} \leftarrow Q_0$
  2. $c \leftarrow H(\mathbf{Ay}, m)$ — $b = 0$: unimodal setting,
  3. $\mathbf{z} \leftarrow \mathbf{y} + (-1)^b c\mathbf{s}$ — $b \leftarrow U(\{0,1\})$: bimodal setting
  4. with probability $\min\left(1, \frac{P(c,\mathbf{z})}{M \cdot Q(c,\mathbf{z})}\right)$, return $\sigma = (c, \mathbf{z})$
  5. if it is not returned, go to step 1

  where Q is the probability distribution of $(c, \mathbf{z})$ output from 3.

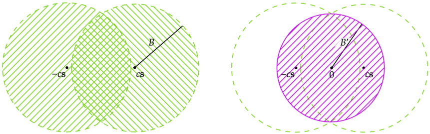
- Rejection sampling:
  - Assume that the Rényi divergence between P and Q are bounded by $M > 0$, i.e., $R_\infty(P \| Q) \leq M$ for some $M > 0$.
  - Then, the distribution Q of the signature (output from 3) turns into a distribution P at the end.

HAETAE Recap: Rejection Sampling

- Rejection sampling guarantees that if $R_\infty(P\|Q) \leq M < \infty$, the following two games are indistinguishable:

$\mathcal{A}^{\text{real}}$:	$\mathcal{A}^{\text{ideal}}$:
1: $\mathbf{x} \leftarrow Q$	1: $\mathbf{x} \leftarrow P$
2: Return $\mathbf{x}$ with probability $\frac{P(\mathbf{x})}{M \cdot Q(\mathbf{x})}$	2: Return $\mathbf{x}$ with probability $\frac{1}{M}$
3: Else repeat 1–2	3: Else repeat 1–2

- In HAETAE, we use the uniform hyperballs for those distributions
  - $Q_0 = U(\mathcal{HB}_0(B))$ and thus $Q = U(\mathcal{HB}_{-c\mathbf{s}}(B) \cup \mathcal{HB}_{c\mathbf{s}}(B))$
  - $P = U(\mathcal{HB}_0(B'))$



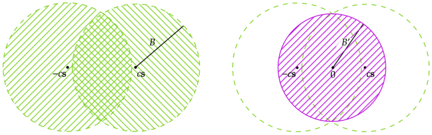
Distribution of Q and P for HAETAE.

HAETAE Recap: Rejection Sampling

- Rejection sampling guarantees that if $R_\infty(P\|Q) \leq M < \infty$, the following two games are indistinguishable:

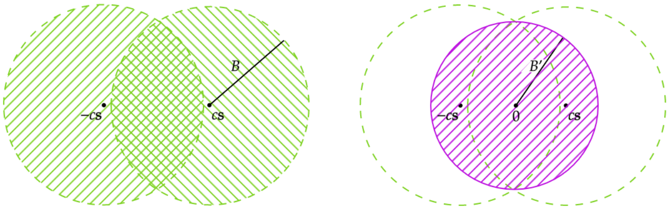
$\mathcal{A}^{\text{real}}$:	$\mathcal{A}^{\text{ideal}}$:
1: $\mathbf{x} \leftarrow Q$	1: $\mathbf{x} \leftarrow P$
2: Return $\mathbf{x}$ with probability $\frac{P(\mathbf{x})}{M \cdot Q(\mathbf{x})}$	2: Return $\mathbf{x}$ with probability $\frac{1}{M}$
3: Else repeat 1–2	3: Else repeat 1–2

- In HAETAE, we use the uniform hyperballs for those distributions
  - $Q_0 = U(\mathcal{HB}_0(B))$ and thus $Q = \frac{1}{2}\chi_{\mathcal{HB}_{-c\mathbf{s}}(B)} + \frac{1}{2}\chi_{\mathcal{HB}_{c\mathbf{s}}(B)}$
  - $P = U(\mathcal{HB}_0(B'))$



Distribution of Q and P for HAETAE.

HAETAE Recap: Rejection Sampling



Distribution of Q and P for HAETAE.

Remark 1. The purple hyperball should be included in every *green-HAETAE-eyes* $\mathcal{HB}_{-c\mathbf{s}}(B) \cup \mathcal{HB}_{c\mathbf{s}}(B)$ for the perfect rejection. Therefore, we have a constraint on B and B' that if $\|c\mathbf{s}\| < S$, then $B' < \sqrt{B^2 - S^2}$.

Remark 2. The expected run time (expected number of rejections $+1$) is M.

HAETAE Recap: Rejection Sampling

In "Fiat-Shamir with Aborts" signatures, the signing procedure is given as:

1. $\mathbf{y} \leftarrow Q_0$
2. $c \leftarrow H(\mathbf{A}\mathbf{y}, m)$ $b = 0$: unimodal setting,
3. $\mathbf{z} \leftarrow \mathbf{y} + (-1)^b c\mathbf{s}$ $b \leftarrow U(\{0,1\})$: bimodal setting
4. with probability $\min\left(1, \frac{P(c,\mathbf{z})}{M \cdot Q(c,\mathbf{z})}\right)$, return $\sigma = (c, \mathbf{z})$,

where Q is the probability distribution of $(c, \mathbf{z})$.

Remark 3. The distributions should be easy to implement since it is related to the signatures, for e.g. uniform distributions.
In HAETAE, the probability can be represented as

- 0: if $\|\mathbf{z}\| \geq B'$,
- $1/2$: else if $\|\mathbf{z} - c\mathbf{s}\| < B$ and $\|\mathbf{z} + c\mathbf{s}\| < B$,
- 1: otherwise.

HAETAE Recap: High-level description (w.o. compression)

KeyGen(1^λ)

1: $\mathbf{A}_{\text{gen}} \leftarrow \mathcal{R}_q^{k\times(\ell-1)}$ and $(\mathbf{s}_{\text{gen}}, \mathbf{e}_{\text{gen}}) \leftarrow S_\eta^{\ell-1} \times S_\eta^k$
2: $\mathbf{b} = \mathbf{A}_{\text{gen}} \cdot \mathbf{s}_{\text{gen}} + \mathbf{e}_{\text{gen}} \in \mathcal{R}_q^k$
3: $\mathbf{A} = (-2\mathbf{b} + q\mathbf{j} |\ 2\mathbf{A}_{\text{gen}} |\ 2\mathbf{Id}_k) \bmod 2q$
4: $\mathbf{s} = (1, \mathbf{s}_{\text{gen}}, \mathbf{e}_{\text{gen}})$
5: **if** $f(\mathbf{s}) > nS^2/\tau^2$, then restart ▷ sk rejection
6: Return sk $= \mathbf{s}$, vk $= (\mathbf{A}, \mathbf{b})$ ▷ $\mathbf{As} = q\mathbf{j} \mod 2q$

Sign(sk, M)

1: $\mathbf{y} \leftarrow U(\mathcal{HB}_0(B))$ ▷ hyperball sampling
2: $c = H(\mathbf{A}\lfloor\mathbf{y}\rceil, M) \in \mathcal{R}_2$
3: $\mathbf{z} = \mathbf{y} + (-1)^b c \cdot \mathbf{s}$ for $b \leftarrow U(\{0,1\})$
4: **if** $\|\mathbf{z}\|_2 \geq B'$, then restart ▷ signature rejection
5: **if** $\|2\mathbf{z} - \mathbf{y}\|_2 < B$, then restart with probability $1/2$
6: Return $\sigma = (c, \lfloor\mathbf{z}\rceil)$

Verify(vk, $M, \sigma = (c, \mathbf{z})$)

1: $\mathbf{w} = \mathbf{Az} - qc\mathbf{j}$
2: Return $(\ c = H(\mathbf{w}, M)\ ) \wedge (\ \|\mathbf{z}\|_2 < B''\ )$

Proof Sketch

- In the ROM, we have a well-known reduction from (S)UF-CMA security to standard MSIS and MLWE problems using the forking lemma.
    - The use of the forking lemma makes the reduction non-tight and non-applicable to the QROM proof.
- To make this reduction tight, the line of works introduced a problem that can be viewed as a "convolution" of lattice and hash, e.g., SelfTargetMSIS.
- In both ROM and QROM, UF-CMA security can be reduced to UF-NMA.
- Then, the UF-NMA security is reduced to the hardness of the "convolution" problem.

Proof Sketch

- In both ROM and QROM, UF-CMA security can be reduced to UF-NMA.
  - Specifically, we follow [DFPS23].
  - It requires the **zero-knowledge** property of the **underlying identification protocol** along with a high enough **commitment min-entropy**.

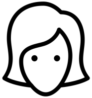
Theorem ([DFPS23], Theorem 10: UF-CMA to UF-NMA)

Assuming that a hash function H is modeled as a random oracle, the underlying ID protocol Σ is Honest-Verifier Zero-Knowledge (HVZK), and the commitment message of the prover has enough min-entropy, then for any quantum adversary $\mathcal{A}$ against UF-CMA security of $\mathsf{FS}(\Sigma, \mathsf{H})$ there exists a UF-NMA adversary $\mathcal{B}$ having a similar run-time with $\mathcal{A}$ and bounding the advantage of $\mathcal{A}$ by the advantage of $\mathcal{B}$ plus some additive constants.

Underlying Identification Protocol

$(\mathsf{sk}, \mathsf{vk}) \leftarrow \mathsf{KeyGen}$ and broadcast vk

Alice (knows sk)



Bob (knows vk)

$\mathbf{y} \leftarrow \mathcal{HB}_0(B)$

$\mathbf{w} = \mathbf{A}\mathbf{y}$

$\xrightarrow{\quad \mathbf{w} \quad}$

$\xleftarrow{\quad c \quad}$ $\qquad c \leftarrow \mathcal{C}$

$\mathbf{z} \leftarrow \mathbf{y} + (-1)^b c\mathbf{s}$

Reject with $p(c, \mathbf{z})$ and restart until succeed

$\xrightarrow{\quad \mathbf{z} \quad}$

check if $\mathbf{A}\mathbf{z} - qc\mathbf{j} = \mathbf{w}$,
and $\|\mathbf{z}\|$ small

Underlying Identification Protocol

$(\mathsf{sk}, \mathsf{vk}) \leftarrow$ KeyGen and broadcast vk

Alice (knows $\mathsf{sk} = \mathbf{s}$)

Bob (knows $\mathsf{vk} = \mathbf{A}$)

$\mathbf{y} \leftarrow \mathcal{HB}_0(B)$

$\mathbf{w} = \mathbf{A}\mathbf{y}$

$\xrightarrow{\mathbf{w}}$

$c \leftarrow \mathcal{C}$

$\xleftarrow{c}$

$\mathbf{z} \leftarrow \mathbf{y} + (-1)^b c\mathbf{s}$

Rejected! $\Rightarrow$ Restart..

$\vdots \qquad \vdots \qquad \vdots$

$\mathbf{y}' \leftarrow \mathcal{HB}_0(B)$

$\mathbf{w}' = \mathbf{A}\mathbf{y}'$

$\xrightarrow{\mathbf{w}'}$

$c' \leftarrow \mathcal{C}$

$\xleftarrow{c'}$

$\mathbf{z}' \leftarrow \mathbf{y}' + (-1)^b c'\mathbf{s}$

Accepted!

$\xrightarrow{\mathbf{z}'}$

check if $\mathbf{A}\mathbf{z}' - qc'\mathbf{j} = \mathbf{w}'$,
and $\|\mathbf{z}'\|$ small

Zero-Knowlege Property

(sk, vk) $\leftarrow$ KeyGen and broadcast vk

Alice (knows sk $= \mathbf{s}$) Bob (knows vk $= \mathbf{A}$)

Alice		Bob
$\mathbf{y} \leftarrow \mathcal{HB}_0(B)$ $\mathbf{w} = \mathbf{A}\mathbf{y}$	$\xrightarrow{\mathbf{w}}$	
	$\xleftarrow{c}$	$c \leftarrow \mathcal{C}$
$\mathbf{z} \leftarrow \mathbf{y} + (-1)^b c\mathbf{s}$ Rejected! $\Rightarrow$ Restart..		
$\vdots$	$\vdots$	$\vdots$
$\mathbf{y}' \leftarrow \mathcal{HB}_0(B)$ $\mathbf{w}' = \mathbf{A}\mathbf{y}'$	$\xrightarrow{\mathbf{w}'}$	$c' \leftarrow \mathcal{C}$
	$\xleftarrow{c'}$	
$\mathbf{z}' \leftarrow \mathbf{y}' + (-1)^b c'\mathbf{s}$ Accepted!		
	$\xrightarrow{\mathbf{z}'}$	check if $\mathbf{A}\mathbf{z}' - qc'\mathbf{j} = \mathbf{w}'$, and $\|\mathbf{z}'\|$ small

- (Statistical) Honest-Verifier Zero-Knowledge (HVZK) requires the existence of an efficient simulator Sim, that outputs the transcripts $(\mathbf{w}, c, \cdots, \mathbf{w}', c', \mathbf{z}')$ such that the distribution of the transcripts has a negligible statistical distance from an honestly generated transcript.

Zero-Knowlege Property

To prove this property, we introduce a simulator Sim$(\mathbf{A}, \mathbf{b}, c)$:

1. $\mathbf{y} \leftarrow U(\mathcal{HB}_0(B))$
2. $\mathbf{w} \leftarrow \mathbf{Ay} - qc\mathbf{j}$
3. $\mathbf{z} \leftarrow \mathbf{y}$
4. $\tilde{\mathbf{w}} \leftarrow U(\mathcal{R}_{2q})$
5. Return $(\mathbf{w}, c, \mathbf{z})$ with probability $p(\mathbf{z}) = \frac{1}{2}\chi_{\mathcal{HB}_0(B')}$, **else** $(\tilde{\mathbf{w}}, c, \perp)$ (i.e. reject).

In fact, this is identical to the case of sk $= 0$, except that $\mathbf{w}$ is sampled differently. It is random over $\mathcal{R}_{2q}$ thanks to decision-MLWE$_{n,k,\ell,2q,\mathsf{Proj}(\mathcal{HB}_0(B))}$ (Proj: a projection map outputting only the first nk coordinates).

Additionally, note that,

- run time: the expected number of rejections does not depend on sk,
- each (aborted) pair $(\tilde{\mathbf{w}}, c)$: the same as before,
- the final distribution of $\mathbf{z}$: uniform in the centered B'-hyperball.

Commitment Min-entropy

- The other condition requiring for the reduction is that the commitments of the protocol have a large min-entropy, at least 256 bits of entropy.
- The min-entropy of the commitments is given as

$$-\log_2\left[\max_{(\mathbf{w},\mathbf{z})}\left[\Pr_{\mathbf{y}}\left[(\mathbf{A}\mathbf{y},\mathbf{y}+(-1)^b c\mathbf{s})=(\mathbf{w},\mathbf{z})\right]\right]\right],$$

  for any $(\mathsf{pk},\mathsf{sk}) \leftarrow \mathsf{KeyGen}$ and $\mathbf{y} \leftarrow U(\mathcal{HB}_0(B))$.
- We easily obtain at least 256 bits of min-entropy in all of our parameter sets.

Secret Key Sampling

As the MSIS bound is given as

$$\|\mathbf{z}\| = \|\mathbf{y} + (-1)^b c\mathbf{s}\| \le \|\mathbf{y}\| + \|c\mathbf{s}\|,$$

we should compute a tight bound for $\|c\mathbf{s}\|$ to achieve efficiency.

1) An easy bound is $\eta \cdot \tau$, where $\mathbf{s} \in S_\eta$ and $\tau = \mathsf{wt}(c)$.
    - This is very easy to compute but gives a much larger bound than the real value. The huge gap between the real value and the computed bound gives inefficiency in choosing the parameters.
    - It is well-known that

      $$\|c\mathbf{s}\| = \|\mathsf{rot}(\mathbf{s}) \cdot \vec{c}\| \le \sigma_{\mathsf{max}}(\mathsf{rot}(\mathbf{s})) \cdot \|\vec{c}\|,$$

      holds over the real numbers, where $\mathsf{rot}(\mathbf{s})$ is the rotational matrix of $\vec{\mathbf{s}}$.

Secret Key Sampling

2) The new bound also has a gap with the actual values since we are dealing with the integer vectors, not the real values. It can be represented as:

$$\begin{aligned}\|c\mathbf{s}\|^2/\|c\|^2 &= \tfrac{1}{n\tau}\textstyle\sum_i |c(\omega_i)|^2 \cdot \|\mathbf{s}(\omega_i)\|^2 \\ &\leq \tfrac{1}{n\tau}\textstyle\sum_i |c(\omega_i)|^2 \cdot \max_i(\|\mathbf{s}(\omega_i)\|^2) = \sigma_{\mathsf{max}}(\mathsf{rot}(\mathbf{s}))^2,\end{aligned}$$

3) With the k-largest $\mathsf{rot}(\mathbf{s})$ instead of the maximum $\mathsf{rot}(\mathbf{s})$, we can bound it more tightly with the similar computation cost, as $\|c\mathbf{s}\|^2/\|c\|^2 \leq f(\mathbf{s})/n$ with

$$f(\mathbf{s}) = \tau \cdot \sum_{i=1}^{m} \max_j^{i\text{-th}} \|\mathbf{s}(\omega_j)\|_2^2 + r \cdot \max_j^{(m+1)\text{-th}} \|\mathbf{s}(\omega_j)\|_2^2.$$

After all, what we do is reject $\mathsf{sk} = \mathrm{s}$ if $f(\mathbf{s}) > \frac{nS^2}{\tau^2}$ (i.e. the bound for $\|c\mathbf{s}\|$ exceeds S). The value S is taken to have 10% to 25% of accepting probability.

Uniform Hyperball Sampling

For sampling $\mathbf{y}$, we need to uniformly sample from a n-dimensional hyperball with radius B, i.e., $\mathcal{HB}_0(B) = \{(y_1, \cdots, y_n) : \sum_i y_i^2 \leq B^2\}$.

- Known method:
  1. $y_i \leftarrow \mathcal{N}(0, 1)$ for $i = 1, \cdots, n+2$
  2. $L \leftarrow \|(y_1, \cdots, y_{n+2})^\top\|_2$
  3. $\mathbf{y} \leftarrow B/L \cdot (y_1, \cdots, y_n)$
  4. return $\mathbf{y}$

- Problem:
  - The floating point arithmetic is not secure.
  - The fixed point arithmetic has an inherent error and also introduces rounding errors, thus inaccurate near the boundary.
  - E.g. the computed value of $y \in \mathcal{HB}_0(B)$ may not be in the hyperball.

Uniform Hyperball Sampling

- So we use a discretized hyperball as,

$$\mathcal{HB}_0(B) \cap (\frac{1}{N}\mathbb{Z})^n = \frac{1}{N}\left(\mathcal{HB}_0(BN) \cap \mathbb{Z}^n\right),$$

  and all the aforementioned analysis is done with this distribution.
  - e.g. M is computed using this distribution, not the continuous hyperball uniform distribution.
- Also, to deal with the inaccuracy near the boundary, we sample in a larger radius and then reject them to the BN-hyperball:
  1. $\mathbf{y} \leftarrow \mathcal{HB}_0(BN+\epsilon)$,
  2. if $\lfloor \mathbf{y} \rceil \leq BN$, output $\lfloor \mathbf{y} \rceil / N$, else restart,

  resulting in a uniform sample in $\mathcal{HB}_0(B) \cap (\frac{1}{N}\mathbb{Z})^n$.
- The continuous Gaussian used for the sampling $\mathcal{HB}_0(BN+\epsilon)$ is replaced by a high-precision discrete Gaussian scaled up by a factor of 2^{72}. The resulting effect on the rejection in 2 is set to be negligible.

Thanks!

Any question?

References I

[DFPS23] Julien Devevey, Pouria Fallahpour, Alain Passelègue, and Damien Stehlé.
A detailed analysis of fiat-shamir with aborts.
Cryptology ePrint Archive, Paper 2023/245, 2023.
https://eprint.iacr.org/2023/245.

[DKL+18] Léo Ducas, Eike Kiltz, Tancrède Lepoint, Vadim Lyubashevsky, Peter Schwabe, Gregor Seiler, and Damien Stehlé.
CRYSTALS-Dilithium: A lattice-based digital signature scheme.
IACR TCHES, 2018(1):238–268, 2018.
https://tches.iacr.org/index.php/TCHES/article/view/839.

[MGTF19] Vincent Migliore, Benoît Gérard, Mehdi Tibouchi, and Pierre-Alain Fouque.
Masking Dilithium - efficient implementation and side-channel evaluation.
In Robert H. Deng, Valérie Gauthier-Umaña, Martín Ochoa, and Moti Yung, editors, ACNS 19, volume 11464 of LNCS, pages 344–362. Springer, Heidelberg, June 2019.

[Por19] Thomas Pornin.
New efficient, constant-time implementations of falcon.
Cryptology ePrint Archive, Paper 2019/893, 2019.

References II

[Pre23] Thomas Prest.
A key-recovery attack against mitaka in the t-probing model.
Cryptology ePrint Archive, Report 2023/157, 2023.
https://eprint.iacr.org/2023/157.