Coding with Cyclic PAM and Vector Quantization for the RLWE/MLWE Channel

Irina E. Bocharova, Henk D.L. Hollmann, **Karan Khathuria**,
Boris D. Kudryashov, Vitaly Skachek

University of Tartu, Estonia

IEEE International Symposium on Information Theory (ISIT)

Aalto University, Espoo, Finland

June 27, 2022

Outline

RLWE/MLWE Based Cryptography

RLWE Channel

Previous Coding Schemes

New Coding Scheme
- Example using BCH codes

Numerical Results and Comparisons

RLWE/MLWE Based Cryptography

Kyber public-key encryption scheme

Setup:

q	prime number
n, k, ℓ	integers
$\mathcal{R}_q$	$\mathbb{Z}_q[X]/(X^n + 1)$
β	distribution on $\mathcal{R}_q$
$\phi : \mathbb{Z}_2^k \to \mathcal{R}_q$	Encoder

Key generation:

- Sample $A \in \mathcal{R}_q^{\ell \times \ell}$ randomly
- Sample $\mathbf{e}, \mathbf{s} \in \mathcal{R}_q^{\ell}$ w.r.t. β

  Public Key: $(A, \mathbf{b} = A\mathbf{s} + \mathbf{e})$

  Secret Key: $\mathbf{s}$

Encryption: Message = $\mathbf{m} \in \mathbb{Z}_2^k$

- Sample $\mathbf{s}', \mathbf{e}' \in \mathcal{R}_q^{\ell}$ w.r.t. β
- Sample $e'' \in \mathcal{R}_q$ w.r.t β
- $\mathbf{u} := A^\intercal \mathbf{s}' + \mathbf{e}'$
- $v := \mathbf{b}^\intercal \mathbf{s}' + e'' + \phi(\mathbf{m})$
- $\mathbf{c} := (\mathbf{u}, v) \in \mathcal{R}_q^{\ell} \times \mathcal{R}_q$

Decryption:

- Compute $v - \mathbf{s}^\top \mathbf{u}$

  $$= \mathbf{b}^\intercal \mathbf{s}' + e'' + \phi(\mathbf{m}) - \mathbf{s}^\top (A^\intercal \mathbf{s}' + \mathbf{e}')$$

  $$= \phi(\mathbf{m}) + \underbrace{\mathbf{e}^\intercal \mathbf{s}' - \mathbf{s}^\intercal \mathbf{e}' + e''}_{\textit{small noise}}$$

- Remove the noise using a decoder $\mathcal{D}$, and restore $\overline{\mathbf{m}}$.

RLWE Channel

$$\underset{\in \mathbb{Z}_2^k}{\mathbf{m}} \longrightarrow \boxed{\begin{matrix}\text{Encoder}\\ \phi\end{matrix}} \xrightarrow[\in \mathcal{R}_q]{\phi(\mathbf{m})} \boxed{\begin{matrix}\text{RLWE/MLWE}\\ \text{Channel}\end{matrix}} \xrightarrow[\in \mathcal{R}_q]{\phi(\mathbf{m}) + \text{noise}} \boxed{\begin{matrix}\text{Decoder}\\ \mathcal{D}\end{matrix}} \longrightarrow \underset{\in \mathbb{Z}_2^k}{\overline{\mathbf{m}}}$$

Lemma (Noise distribution of MLWE [1])

The distribution of one coefficient of the MLWE noise element $\mathbf{e}^\top \mathbf{s}' - \mathbf{s}^\top \mathbf{e}' + e''$ *is given by*

$$\psi = \circledast_{\ell-1}(\circledast_{n-1}\xi) * \circledast_{\ell-1}(\circledast_{n-1}\xi) * \beta, \tag{1}$$

where $\circledast$ *denoted the convolution of distributions.*

1 Georg Maringer, Sven Puchinger, and Antonia Wachter-Zeh. "Higher Rates and Information-Theoretic Analysis for the RLWE Channel". In: *2020 IEEE Information Theory Workshop (ITW)*. 2021, pp. 1–5. DOI: 10.1109/ITW46852.2021.9457596

Previous Coding Schemes

$$\mathbf{m} \in \mathbb{Z}_2^k \longrightarrow \boxed{\begin{matrix}\text{Encoder}\\ \phi\end{matrix}} \longrightarrow \phi(\mathbf{m}) \in \mathcal{R}_q$$

$$\begin{tikzcd} \mathbb{Z}_2^k \arrow[r, "\mathcal{C}"] \arrow[rr, bend left, "\phi"] & \mathbb{Z}_2^n \arrow[r, "\text{Map}"] & \mathcal{R}_q \end{tikzcd}$$

- $\mathcal{C}$ is a linear code,
- $\text{Map} : \mathbb{Z}_2^n \to \mathcal{R}_q$ is defined coordinate-wise:

$$0 \mapsto 0$$
$$1 \mapsto \lfloor q/2 \rfloor$$

$$\phi(\mathbf{m})\text{+noise} \in \mathcal{R}_q \longrightarrow \boxed{\begin{matrix}\text{Decoder}\\ \mathcal{D}\end{matrix}} \longrightarrow \overline{\mathbf{m}} \in \mathbb{Z}_2^k$$

$$\begin{tikzcd} \mathcal{R}_q \arrow[r, "\text{Demap}"] \arrow[rr, bend left, "\mathcal{D}"] & \mathbb{Z}_2^n \arrow[r, "\text{Dec}"] & \mathbb{Z}_2^k \end{tikzcd}$$

- Demap: $\mathcal{R}_q \to \mathbb{Z}_2^n$ is defined coordinate-wise:

$$y \mapsto \begin{cases} 0 & \text{if } |y| < \lfloor q/4 \rfloor \\ 1 & \text{otherwise} \end{cases}$$

- Dec refers to decoding of the code $\mathcal{C}$.

Advantages of Coding schemes

- **Lower decryption failure rate (DFR):**

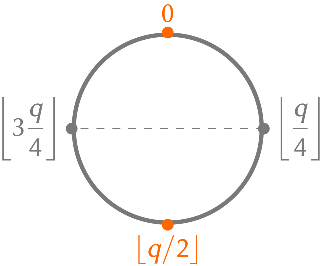

Figure: $\mathbb{Z}_q$

$\mathfrak{p}$ = probability that the magnitude of noise is $< q/4$

Uncoded	Coded
$\text{DFR} \leq 1 - \mathfrak{p}^n$	$\text{DFR} \leq 1 - \sum_{i=0}^{t} \binom{n}{i} \mathfrak{p}^i (1-\mathfrak{p})^{n-i}$

- Equivalently, we can have **lower communication rate**.

New Coding Scheme

Using PAM and vector quantization

Encoding: Fix $Q = 2^L$, $N = nL$ and $h = \lfloor q/Q \rfloor$ (here we assume $L = 2$)

$$\begin{tikzcd}
\mathbb{Z}_2^k \arrow[r, "\mathcal{C}"] \arrow[rrrr, bend left, "\phi"] & \mathbb{Z}_2^N \arrow[r, "\text{Lift}"] & \mathbb{Z}_Q^n \arrow[r, "\text{Gray}"] & \mathbb{Z}_Q^n \arrow[r, "h"] & \mathbb{Z}_q^n
\end{tikzcd}$$

$$\mathbf{m} \longmapsto \mathbf{c} := \mathbf{m}G \longmapsto \mathbf{b} \longmapsto \mathbf{a} \longmapsto h\mathbf{a}$$

- $\mathcal{C}$ is a linear code, generated by $G \in \mathbb{F}_2^{k \times N}$.
- Lift: $\mathbb{Z}_2^{2n} \to \mathbb{Z}_4^n$ given by $(c_1, \ldots, c_n, c_{n+1}, \ldots, c_{2n}) \mapsto (2c_1 + c_{n+1}, \ldots, 2c_n + c_{2n})$
- Gray: $\mathbb{Z}_4^n \to \mathbb{Z}_4^n$ is defined coordinate-wise: $2b_2 + b_1 \mapsto 2b_2 + (b_1 \oplus b_2)$
- $h : \mathbb{Z}_4^n \to \mathbb{Z}_q^n$ is scalar multiplication by h, i.e., $\mathbf{a} \mapsto h\mathbf{a}$.

New Coding Scheme

Using PAM and vector quantization

Decoding: Let $\mathbf{y} = \mathbf{x} + \mathbf{z} \in \mathbb{Z}_q^n$ be the received vector, where $\mathbf{z}$ is the noise from RLWE Channel.

Vector quantization = Maximum-likelihood demodulation + HDD/SDD

1. For each code-symbol $c_1, \ldots, c_{2n}$, compute the log-likelihood ratios (LLRs) $\mathbf{w} = (w_1, \ldots, w_{2n})$:

$$w_{i+jn} := \log \frac{\sum_{\mathbf{c} \in \mathbb{Z}_2^2 : c_j = 1} \psi(y_i - x_{\mathbf{c}} \mod q)}{\sum_{\mathbf{c} \in \mathbb{Z}_2^2 : c_j = 0} \psi(y_i - x_{\mathbf{c}} \mod q)}, \tag{2}$$

   for each $i = 1, \ldots, n$ and $j = 0, 1$.

2. Use the LLRs to perform hard-decision decoding (HDD) or soft-decision decoding (SDD).

Example

Using BCH codes and soft-decision decoding

Kyber-512 parameters: $n = 256, \ell = 2, q = 3329$

Let $\mathcal{C}$ be the primitive binary [512, 256, 62] BCH code.

- **Encoding:** Message $\mathbf{m} \in \mathbb{Z}_2^{256}$

$$\mathbf{m} \overset{\mathcal{C}}{\mapsto} \mathbf{c} := \mathbf{m}G \mapsto \mathbf{x} := h\,\mathrm{Gray}(2c_1 + c_{n+1}, \ldots, 2c_n + c_{2n})$$

- **Decoding:** Received vector $\mathbf{y} = \mathbf{x} + \mathbf{z} \in \mathbb{Z}_q^n$
  1. Compute the LLR's $\mathbf{w} = (w_1, \ldots, w_{2n})$
  2. Perform either HDD or SDD:
     - HDD: Berlekamp-Massey algorithm
     - SDD: Ordered statistic decoding (OSD)

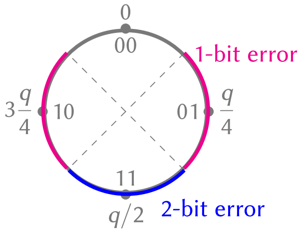

Figure: Gray coded 4-PAM symbols

Numerical Results and Comparisons

Comparisons for Kyber-512 scheme with **fixed communication rate**

Coding scheme	$d_{\min}$	DFR
Uncoded[4]	1	2^{-174}
Ternary BCH code[5]	26	2^{-989}
Binary BCH code with HDD	62	2^{-1325}
Binary BCH code with OSD-8	62	2^{-1414}

4 Joppe Bos et al. “CRYSTALS-Kyber: a CCA-secure module-lattice-based KEM”. In: *IEEE European Symposium on Security and Privacy (EuroS&P).* 2018, pp. 353–367.

5 Georg Maringer, Sven Puchinger, and Antonia Wachter-Zeh. “Higher Rates and Information-Theoretic Analysis for the RLWE Channel”. In: *2020 IEEE Information Theory Workshop (ITW).* 2021, pp. 1–5. DOI: 10.1109/ITW46852.2021.9457596.

Conclusion

- New coding scheme for RLWE channel:
  - Encoding: Linear code + pulse amplitude modulation (PAM)
  - Decoding: Vector quantization + error correction using hard/soft decisions.
- Advantages: Error-correction codes reduces communication rate and/or DFR in the RLWE-based encryption schemes.

Potential future work

- Comparing the performance of other codes, e.g. LDPC codes, Turbo codes
- Constant-time decoding?
- Coding scheme for other LWE-based cryptosystems, e.g. homomorphic encryption schemes, PIR schemes.

Thank you