Reduction from sparse LPN to LPN, Dual Attack 3.0

Kévin Carrier, Thomas Debris-Alazard, Charles Meyer-Hilfiger, Jean-Pierre Tillich

Eurocrypt 2024

Table of Contents

Code-based cryptography and Decoding Problem

Code-based primitives

- **PKE, KEM (NIST):** McEliece, BIKE, HQC, ...
- **Signatures (NIST):** SDitH, Wave, ...

Security of code-based primitives $\rightarrow$ Hardness of decoding linear codes

Decoding Problem at distance t (small)

- **Input:**
  - $\mathscr{C}$ binary linear code of len. n and dim. k (linear subspace of $\mathbb{F}_2^n$ of dimension k)
  - $\mathbf{c} + \mathbf{e}$ with $\mathbf{c} \in \mathscr{C}$ and $|\mathbf{e}| = t$
- **Output:** $\mathbf{e}$

This work: new Decoding Algorithm

Decoding Problem

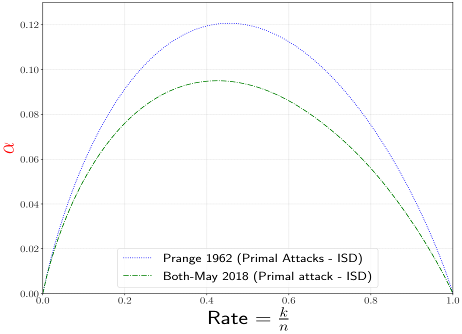


Figure: Complexity $2^{\alpha\, n}$

This work: new Decoding Algorithm

Decoding Problem

$\downarrow$

Reduced to LPN (2.0)

LPN Problem

- **Input:** Many samples $(\mathbf{a}, \langle \mathbf{a}, \mathbf{s} \rangle + e)$
  - $\mathbf{s} \in \mathbb{F}_2^s$ fixed secret
  - $\mathbf{a}$ taken at random in $\mathbb{F}_2^s$
  - $e \sim \mathrm{Bern}(p)$
- **Output:** $\mathbf{s}$

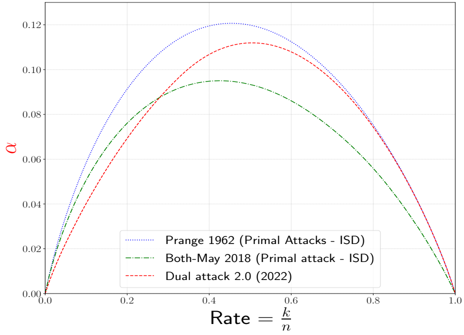


Figure: Complexity $2^{\alpha\, n}$

$\rightarrow$ Big gain for rather small rates

This work: new Decoding Algorithm

Decoding Problem

$\downarrow$

Reduced to **sparse** LPN (2.0)

$\downarrow$

Reduced to plain LPN of smaller dim. (3.0)

LPN Problem

- **Input:** Many samples $(\mathbf{a}, \langle \mathbf{a}, \mathbf{s} \rangle + e)$
  - $\mathbf{s} \in \mathbb{F}_2^s$ fixed secret
  - $\mathbf{a}$ taken at random in $\mathbb{F}_2^s$
  - $e \sim \mathrm{Bern}(p)$
- **Output:** $\mathbf{s}$

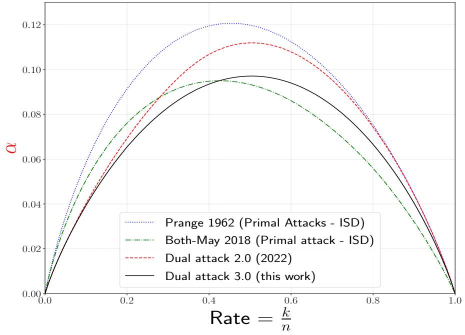


Figure: Complexity $2^{\alpha\, n}$

$\rightarrow$ Big gain for $R < 0.42$

Table of Contents

Setting for Dual Attacks

Dual code

$$\mathscr{C}^{\perp} = \{\mathbf{h} \in \mathbb{F}_2^n \ : \ \langle \mathbf{h}, \mathbf{c} \rangle = 0 \quad \forall \mathbf{c} \in \mathscr{C}\} \qquad \text{with} \qquad \langle \mathbf{x}, \mathbf{y} \rangle = \sum x_i \, y_i \pmod 2$$

Compute dual vector $\mathbf{h} \in \mathscr{C}^{\perp}$

Given $\mathbf{c} + \mathbf{e}$ $\quad \rightarrow \langle \mathbf{c} + \mathbf{e}, \mathbf{h} \rangle = \langle \mathbf{e}, \mathbf{h} \rangle$

How to exploit?

Reducing Decoding to LPN (Dual attack 2.0) [CDMT, 2022]

$$\langle \mathbf{c} + \mathbf{e}, \mathbf{h} \rangle = \langle \mathbf{e}, \mathbf{h} \rangle$$

- Split support in complementary part $\mathscr{P}$ and $\mathscr{N}$ $\rightarrow$ Recover $\mathbf{e}_{\mathscr{P}}$?
- Compute dual vector $\mathbf{h} = \underbrace{\boxed{}}_{\mathscr{P}}\underbrace{\boxed{\quad w \text{ (small)} \quad}}_{\mathscr{N}}$

$$\rightarrow \langle \mathbf{e}, \mathbf{h} \rangle = \langle \underbrace{\mathbf{e}_{\mathscr{P}}}_{\text{secret}}, \mathbf{h}_{\mathscr{P}} \rangle + \underbrace{\langle \mathbf{e}_{\mathscr{N}}, \mathbf{h}_{\mathscr{N}} \rangle}_{\text{noise: biased to 0}}$$

N dual vectors $\rightarrow$ N LPN samples

$$(\mathbf{a}, \langle \mathbf{s}, \mathbf{a} \rangle + e) \text{ w.t } \begin{cases} \mathbf{a} = \mathbf{h}_{\mathscr{P}} \quad \in \mathbb{F}_2^{|\mathscr{P}|} \\ \mathbf{s} = \mathbf{e}_{\mathscr{P}} \\ e = \langle \mathbf{e}_{\mathscr{N}}, \mathbf{h}_{\mathscr{N}} \rangle \end{cases}$$

Hardness of this LPN problem

$e = \langle \mathbf{e}_{\mathscr{N}}, \mathbf{h}_{\mathscr{N}} \rangle \qquad \mathrm{bias}\left(\langle \mathbf{e}_{\mathscr{N}}, \mathbf{h}_{\mathscr{N}} \rangle\right) \stackrel{\triangle}{=} \frac{1}{N} \sum_{\mathbf{h}} (-1)^{\langle \mathbf{e}_{\mathscr{N}}, \mathbf{h}_{\mathscr{N}} \rangle}$

Bias computed theoretically using only $|\mathbf{e}_{\mathscr{N}}|$ and $|\mathbf{h}_{\mathscr{N}}| = w$

$\rightarrow$ is exponentially small

Lower bound

$$N \geq \frac{1}{\mathrm{bias}\left(\langle \mathbf{e}_{\mathscr{N}}, \mathbf{h}_{\mathscr{N}} \rangle\right)^2} \qquad \rightarrow \qquad \text{Can recover secret } \mathbf{e}_{\mathscr{P}}$$

Solving the LPN problem : Score function

LPN samples $(\mathbf{a}, \langle \mathbf{a}, \mathbf{s} \rangle + e) \quad \rightarrow \quad$ Recover $\mathbf{s}$?

Score function for $\mathbf{x} \in \mathbb{F}_2^{|\mathscr{P}|}$

$$F(\mathbf{x}) = \mathrm{bias}\left(\langle \mathbf{a}, \mathbf{s} \rangle + e \;-\; \langle \mathbf{a}, \mathbf{x} \rangle\right) = \frac{1}{N} \sum_{\mathbf{a}} (-1)^{\langle \mathbf{a}, \mathbf{s} \rangle + e \;-\; \langle \mathbf{a}, \mathbf{x} \rangle}$$

When $\mathbf{x} = \mathbf{s}$ then $F(\mathbf{x})$ is high and equal $\mathrm{bias}\,(e)$

Compute $\max F(\mathbf{x}) \quad \rightarrow \quad$ use FFT over $\mathbb{F}_2^{|\mathscr{P}|}$ to compute all values of $F(\mathbf{x})$.

Key remark

$\mathbf{s} = \mathbf{e}_{\mathscr{P}}$ is sparse and yet we compute $F(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{F}_2^{|\mathscr{P}|}$

Table of Contents

Reduction from sparse LPN to plain LPN (1)

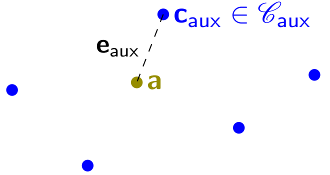
Approach

$$\left(\underset{\mathbb{F}_2^{|\mathscr{P}|}}{\overset{}{\underset{\cap}{\mathbf{a}}}} \;,\; \langle \overset{\text{sparse}}{\overset{\uparrow}{\mathbf{s}}} \;,\; \mathbf{a} \rangle + e \right) \xrightarrow[\text{Increase Noise}]{\text{Lower Dimension}} \left(\underset{\mathbb{F}_2^{\leq |\mathscr{P}|}}{\underset{\cap}{\mathbf{a}'}} \;,\; \langle \overset{\text{plain}}{\overset{\uparrow}{\mathbf{s}'}} \;,\; \mathbf{a}' \rangle + e' \right)$$

Reduction from sparse LPN to plain LPN (2)

Linear code $\mathscr{C}_{\mathsf{aux}} \subset \mathbb{F}_2^{|\mathscr{P}|}$

$$\mathscr{C}_{\mathsf{aux}} = \{\mathbf{m}_{\mathsf{aux}}\mathbf{G}_{\mathsf{aux}} : \mathbf{m}_{\mathsf{aux}} \in \mathbb{F}_2^{\dim(\mathscr{C}_{\mathsf{aux}})}\}$$



$$\mathbf{a} = \mathbf{c}_{\mathsf{aux}} + \underbrace{\mathbf{e}_{\mathsf{aux}}}_{\text{short}}$$

$$\langle \mathbf{s}, \mathbf{a} \rangle + e = \langle \mathbf{s}, \mathbf{c}_{\mathsf{aux}} \rangle + \underbrace{\langle \mathbf{s}, \mathbf{e}_{\mathsf{aux}} \rangle + e}_{e' \text{ new noise}}$$

$$\langle \mathbf{s}, \mathbf{c}_{\mathsf{aux}} \rangle = \langle \mathbf{s}, \mathbf{m}_{\mathsf{aux}}\mathbf{G}_{\mathsf{aux}} \rangle = \langle \mathbf{s}\mathbf{G}_{\mathsf{aux}}^{\top}, \mathbf{m}_{\mathsf{aux}} \rangle$$

Sample space $\mathbb{F}_2^{|\mathscr{P}|} \to \mathbb{F}_2^{\dim(\mathscr{C}_{\mathsf{aux}})}$ is smaller!

Analysis: estimating the number of false candidates?

LPN samples $(\mathbf{a}, \langle \mathbf{a}, \mathbf{s} \rangle + e)$

$$\to \text{Score function } F(\mathbf{x}) = \mathrm{bias}\left(\langle \mathbf{a}, \mathbf{s} \rangle + e \;-\; \langle \mathbf{a}, \mathbf{x} \rangle\right)$$

Key question for complexity analysis

How many $\mathbf{x}$ (apart from the secret $\mathbf{s}$) are such that

$$F(\mathbf{x}) \approx \mathrm{bias}\left(e\right)?$$

Distribution of the score function: a bit of history

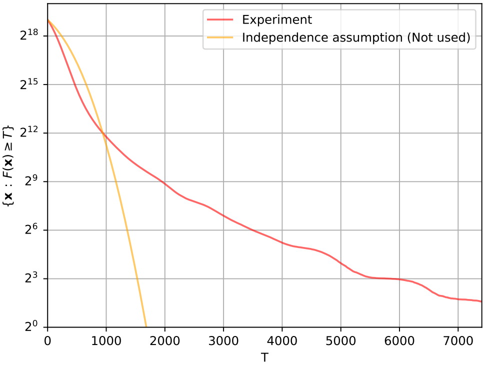


Figure: Distribution score function in Dual Attack 3.0

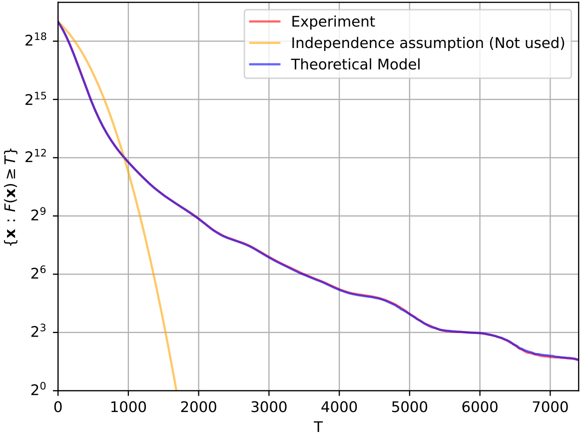
A bit of history about Dual Attacks 2.0:

- [CDMT, 2022] Notice exp. differences
- [M & Tillich, 2023] New model

~~**Independence Assumptions**~~

Prediction of score function

$\rightarrow$ Generalization of [M & Tillich, 2023] to analyze Dual Attacks 3.0



Dual formula

$$F(\mathbf{x}) \approx \sum_{i \in \mathbb{N}} N_i(\mathscr{D})\, K_w(i)$$

- $N_i(\mathscr{D})$ number of codewords of weight i in some code $\mathscr{D}$
- $K_w(i)$ is Krawtchouk polynomial

Proof: Poisson formula $+\ \widehat{1_w} = K_w$

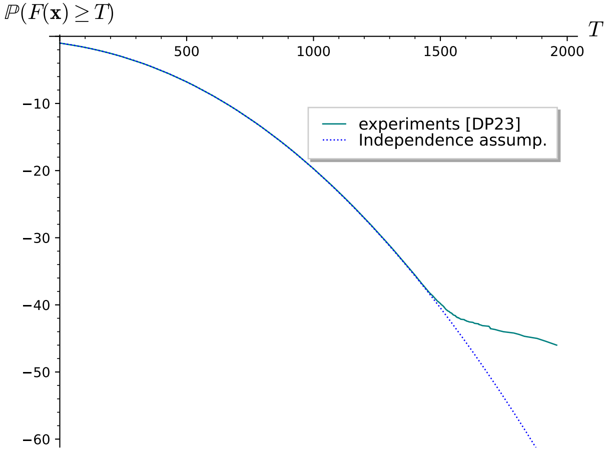
Model

$$N_i(\mathscr{D}) \sim \text{Poisson variable of good expected value}$$

Table of Contents

The problem

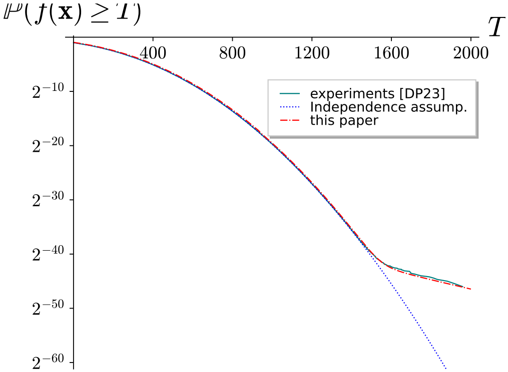
[Ducas & Pulles, 23] $\rightarrow$ Show independence assumption are invalid



$\rightarrow$ Seriously question Dual Attacks in Lattices

Accurate score prediction

$\rightarrow$ We adapt [M & Tillich, 2023] to analyze dual attacks in codes to lattices



Dual formula

$$F(\mathbf{x}) \approx \sum_i N_i(\Lambda) \left(\frac{w}{i}\right)^{n/2} J_{\frac{n}{2}}(2\pi\, w\, i)$$

- $N_i(\Lambda)$ number of lattice points of length i
- J_n Bessel function

Proof : Poisson formula

$$+$$

$$\widehat{1_{\leq w}} = \left(\frac{w}{i}\right)^{n/2} J_{\frac{n}{2}}(2\pi\, w\, i)$$

Conclusion

- New decoding algorithm beat state of art for rates smaller than 0.42
- Analysis not relying on independence assumptions
- Prediction of score function in lattice

Thank you!