Algebraic Techniques to solve the Regular Syndrome Decoding Problem

Pierre Briaud[1], joint work with Morten Øygarden[2].

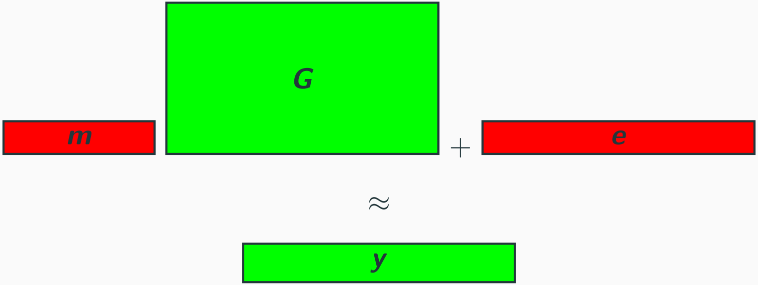
Rennes, December 16

[1]Inria Paris & Sorbonne Université
[2]Simula UiB

Decoding Problem (DP) over $\mathbb{F}_q$ (aka Primal LPN)

Given full-rank $\boldsymbol{G} \in \mathbb{F}_q^{k \times n}$, distinguish

- $\boldsymbol{y} = \boldsymbol{m}\boldsymbol{G} + \boldsymbol{e}$, $\boldsymbol{m} \in \mathbb{F}_q^k$, error $\boldsymbol{e} \sim \chi$
- $\boldsymbol{y} \sim \mathcal{U}(\mathbb{F}_q^n)$



Our setting

Bounded number of samples $n = k^{1+\alpha}$, $0 < \alpha < 1$

Error $\boldsymbol{e}$ of low Hamming weight, $|\boldsymbol{e}| = t$

$\rightarrow$ Coding theory point of view ! Length n, dim. k, code rate $R \stackrel{def}{=} k/n$

Underlying code $\mathcal{C}$

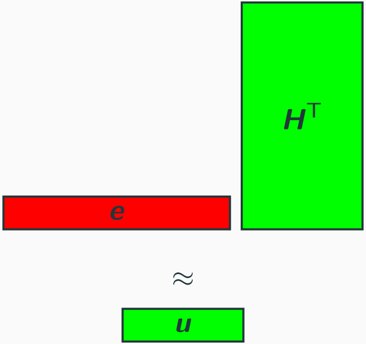
$$\mathcal{C} \stackrel{def}{=} \left\{\boldsymbol{mG},\ \boldsymbol{m} \in \mathbb{F}_q^k\right\} = \left\{\boldsymbol{x} \in \mathbb{F}_q^n,\ \boldsymbol{xH}^\intercal = \boldsymbol{0}\right\},\ \boldsymbol{H} \in \mathbb{F}_q^{(n-k)\times n}$$

"Philosophical" difference ? Error distribution χ

Syndrome Decoding (SD) Problem (aka Dual LPN)

Given full-rank $\boldsymbol{H} \in \mathbb{F}_q^{(n-k)\times n}$, distinguish

- $\boldsymbol{u} = \boldsymbol{e}\boldsymbol{H}^\top \in \mathbb{F}_q^{n-k}, \ \boldsymbol{e} \sim \chi$
- $\boldsymbol{u} \sim \mathcal{U}(\mathbb{F}_q^{n-k})$



Some use cases

- Symmetric crypto [HB01]
- PKE: Alekhnovich scheme [Ale03]

Pseudorandom correlation generators (PCGs): correlated randomness [Boy+19]

- PRG $(\boldsymbol{m}, \boldsymbol{e}) \mapsto \boldsymbol{mG} + \boldsymbol{e}$ or $\boldsymbol{e} \mapsto \boldsymbol{eH}^\intercal$ (correlated seeds) + Function Secret Sharing
- used to build secure MPC, ZK proofs ...

Non-standard parameters

LOW noise (inverse poly, not constant) $\to$ Very large sizes

Possibly large field (typically $\mathbb{F}_{2^{128}}$)

ex: $\lambda = 128$ over $\mathbb{F}_2$, $(n = 2^{22},\ k = 67440,\ t = 4788)$ [Boy+19]; [Liu+22]

Regular SD (RSD)

Assume $n = N \times t$ for some $N \in \mathbb{N}$ (blocksize)

Regular distribution [AFS05]

- For $1 \leq i \leq t$, sample $\boldsymbol{e}_i \in \mathbb{F}_q^N$ **random of weight 1**
- Final error is $\boldsymbol{e} \stackrel{def}{=} (\boldsymbol{e}_1, \ldots, \boldsymbol{e}_t) \in \mathbb{F}_q^n$

Introduction in Secure Computation [Haz+18]

Now used in many protocols [Boy+19]; [Wen+20]; [Yan+20] ...

$\rightarrow$ Reduce Function Secret Sharing cost

[AFS05] Augot, Finiasz, and Sendrier. "A Family of Fast Syndrome Based Cryptographic Hash Functions". *MYCRYPT 2005*.

[Haz+18] Hazay et al. *TinyKeys: A New Approach to Efficient Multi-Party Computation*.

Known attacks on RSD

Attacks on Plain SD ! Do NOT exploit regular distribution:

- "Folklore attack" and ISD algorithms [Pra62]; [MMT11]; [MO15]...
- Statistical Decoding [Jab01]
  (recently improved by [Car+22])

What about algebraic techniques ?

[Pra62] Prange. "The use of information sets in decoding cyclic codes".

[Jab01] Jabri. "A Statistical Decoding Algorithm for General Linear Block Codes".

[Car+22] Carrier et al. *Statistical Decoding 2.0: Reducing Decoding to LPN*.

Algebraic attacks

Generic technique in cryptanalysis:

- Model scheme or hard problem as polynomial system
- Solve it ! (Gröbner Bases, linearization)

This talk

1^{st} algebraic attack on RSD

- competitive for very small code rates $\leftrightarrow$ enough samples
- algebraic system + detailed analysis

(Naive) algebraic system

Modeling regular structure

Polynomial ring $R \stackrel{def}{=} \mathbb{F}_q[(e_{i,j})_{i,j}]$ in n variables, block $\boldsymbol{e}_i \stackrel{def}{=} (e_{i,1}, \ldots, e_{i,N}) \in \mathbb{F}_q^N$

Coordinates $\in \mathbb{F}_q$ (field equations)

$$\forall i,\ \forall j,\ e_{i,j}^q - e_{i,j} = 0. \quad (1)$$

One $\neq 0$ coordinate per block

$$\forall i,\ \forall j_1 \neq j_2,\ e_{i,j_1} e_{i,j_2} = 0. \quad (2)$$

Over $\mathbb{F}_2$, this coordinate is 1

$$\forall i,\ \sum_{j=1}^{N} e_{i,j} = 1. \quad (3)$$

We consider quadratic system $\mathcal{Q} \stackrel{def}{=} (1) \cup (2) \cup (3)$

Adding parity-check equations

Linear equations in the $e_{i,j}$'s from $\boldsymbol{e}\boldsymbol{H}^{\intercal} = \boldsymbol{u}$:

Parity-checks

$$\mathcal{P} \stackrel{def}{=} \{\forall i \in \{1..n-k\},\ \langle \boldsymbol{e}, \boldsymbol{h}_i \rangle - u_i = 0\}\,.$$

Final system $\mathcal{S} \stackrel{def}{=} \mathcal{P} \cup \mathcal{Q}$.

Set of solutions to $\mathcal{S}$ = Set of solutions to RSD (let's say 1)

Recurrent issue

Cost of System Solving !

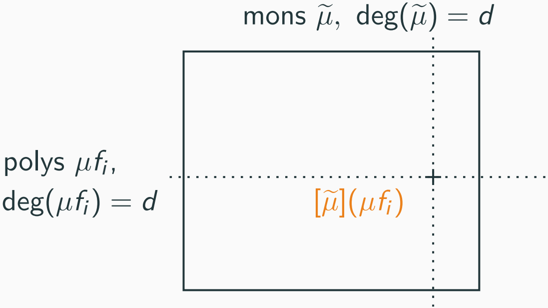
- Known for random systems (or at least for well-studied ones)
- $\mathcal{S}$ neither random nor well-studied ...

Solving Algorithms

1. multiply eqs by all monomials μ:
   $\rightarrow$ polys μf_i, f_i initial eq
2. store them in matrix, fixed degree $d = \deg(\mu f_i)$
3. do linear algebra

Macaulay matrix M_d

Matrix of size $\exp(d)$



What do we need ?

Highest degree $d = D$ for a Macaulay matrix

Analyzing $\mathcal{S}$

Approach

Recall that $\mathcal{S} = \underbrace{\{\text{parity-checks}\}}_{\mathcal{P}} \cup \underbrace{\{\text{regular structure}\}}_{\mathcal{Q}}$

$$\mathcal{P} = \{\forall i \in \{1..n-k\},\ \langle \boldsymbol{e}, \boldsymbol{h}_i \rangle - u_i\}$$

$$\mathcal{Q} = \{\forall i \in \{1..t\}, \forall j \in \{1..N\},\ e_{i,j}^2 - e_{i,j}\} \cup \{\forall i, \forall j_1 \neq j_2,\ e_{i,j_1} e_{i,j_2}\} \cup \{\forall i,\ \textstyle\sum_{j=1}^{N} e_{i,j} - 1\}$$

- To keep internal structure, treat $\mathcal{P}$ and $\mathcal{Q}$ separately
- Focus on homogeneous parts: $\langle \mathcal{S}^{(h)} \rangle = \langle \mathcal{P}^{(h)} \rangle + \langle \mathcal{Q}^{(h)} \rangle$

Highest degree from Hilbert Series

Polynomial ring $R \stackrel{def}{=} \mathbb{F}_q[(e_{i,j})_{i,j}]$, $R = \bigoplus_{d\in\mathbb{N}} R_d$ hom. components

Hom. ideal $I \stackrel{def}{=} \langle f_1, \ldots, f_m \rangle$, $I_d \stackrel{def}{=} I \cap R_d$

Hilbert Series (HS) of I

Contains properties of I we need (in particular highest degree D)

$\rightarrow$ Find Hilbert Series for $\langle \mathcal{S}^{(h)} \rangle$ then deduce D

Formal definition:

$$\mathcal{H}_{R/I}(z) \stackrel{def}{=} \sum_{d\in\mathbb{N}} \dim\left(R_d/I_d\right) z^d$$

0-dimensional ideal ($\mathcal{H}_{R/I}(z)$ is a polynomial): $H(I) \stackrel{def}{=} \min\{\delta \in \mathbb{N}, I_\delta = R_\delta\}$ (index)

Structural part $\mathcal{Q}$

Only depends on regular distribution. We analyze $q = 2$ (*e.g.* we can use (3))

$$\mathcal{Q}^{(h)} = \underbrace{\{\forall i \in \{1..t\}, \forall j \in \{1..N\},\ e_{i,j}^2\}}_{(1)} \cup \underbrace{\{\forall i, \forall j_1 \neq j_2,\ e_{i,j_1} e_{i,j_2}\}}_{(2)} \cup \underbrace{\{\forall i,\ \sum_{j=1}^{N} e_{i,j}\}}_{(3)}$$

HS 1

We have $\dim(R_d/\langle \mathcal{Q}^{(h)} \rangle_d) = \binom{t}{d}(N-1)^d$. Thus,

$$\boxed{\mathcal{H}_{R/\langle \mathcal{Q}^{(h)} \rangle}(z) = (1 + (N-1)z)^t}$$

Proof (monomial counting).

Using (1) and (2), squarefree + at most one variable per $\boldsymbol{e}_i$ block

Using (3), we get rid of one variable per $\boldsymbol{e}_i$ block □

Random part $\mathcal{P}$

We have $\mathcal{P}^{(h)} = \{\boldsymbol{e}\boldsymbol{H}^\mathsf{T}\}$. By assumption on $\boldsymbol{H}$, "random" linear equations

- but we want "randomness" in $R/\langle \mathcal{Q}^{(h)} \rangle$
- here, randomness means (semi)-regularity:

Semi-regularity over $\mathbb{F}_2$ [Bar04]

Let $S \stackrel{def}{=} \mathbb{F}_2[\mathbf{e}]/\langle \mathbf{e}^2 \rangle$, $\mathcal{F} = \{f_1, \ldots, f_m\}$ homogeneous, 0-dim, index $d_{\langle \mathcal{F} \rangle}$
System $\mathcal{F}$ is semi-regular over $\mathbb{F}_2$ if $\langle \mathcal{F} \rangle \neq S$ and if

$$\forall i,\ \deg(g_i f_i) < d_{\langle \mathcal{F} \rangle},\ g_i f_i = 0 \in S/\langle f_1, \ldots, f_{i-1} \rangle \Rightarrow g_i = 0 \in S/\langle f_1, \ldots, f_i \rangle \quad (4)$$

In this paper, we adapt it to $R/\langle \mathcal{Q}^{(h)} \rangle$ instead of $R/\langle \mathbf{e}^2 \rangle$

[Bar04] Bardet. "Étude des systèmes algébriques surdéterminés. Applications aux codes correcteurs et à la cryptographie".

Combining everything

Semi-regular HS are known ! (write exact sequences from (4))

Assumption

We assume semi-regularity of $\mathcal{P}^{(h)}$ with our new definition

We have $\langle \mathcal{S}^{(h)} \rangle = \langle \mathcal{P}^{(h)} \rangle + \langle \mathcal{Q}^{(h)} \rangle$, we know $\mathcal{H}_{R/\langle \mathcal{Q}^{(h)} \rangle}$. We want $\mathcal{H}_{R/\langle \mathcal{S}^{(h)} \rangle}$
Under Assumption, we get

$$\mathcal{H}_{R/\langle \mathcal{S}^{(h)} \rangle}(z) = \frac{\mathcal{H}_{R/\langle \mathcal{Q}^{(h)} \rangle}(z)}{(1+z)^{n-k}}$$

HS for $\mathcal{S}^{(h)}$ (under Assumption + using HS 1)

$$\mathcal{H}_{R/\langle \mathcal{S}^{(h)} \rangle}(z) = \frac{(1+(N-1)z)^t}{(1+z)^{n-k}}$$

Solving $\mathcal{S}$ (more concretely)

Cost of Gröbner Basis

- **Dense** linear algebra on Macaulay matrix $\boldsymbol{M}_D \to$ row ech. form
- Cost exponential in D, $2 \leq \omega < 3$:

$$T_{\text{solve}}(\mathcal{S}) = \mathcal{O}(\#\text{cols}(\boldsymbol{M}_D)^{\omega}) = \mathcal{O}\left(\binom{t}{D}^{\omega}(N-1)^{\omega D}\right)$$

Highest degree D from HS

Index of first < 0 coef. in $\mathcal{H}_{R/\langle \mathcal{S}^{(h)} \rangle}$
$+$ "Degree fall assumption": same D for $\mathcal{S}^{(h)}$ and $\mathcal{S}$

Hybrid approach I

Conjectured D may be too high to be practical

Hybrid approach (folklore & [BFP10])

Fix f variables + solve specialized system $\mathcal{S}_{\text{spec},f}$

Hope: smaller D for $\mathcal{S}_{\text{spec},f}$

$\rightarrow$ Guess $f \geq 0$ zero positions in $\boldsymbol{e}$ (as Prange but $f \ll k$)

- Simplest way: $u \overset{def}{=} f/t$ per block, success proba $\left(\frac{\binom{N-1}{u}}{\binom{N}{u}}\right)^t = (1 - u/N)^t$

[BFP10] Bettale, Faugère, and Perret. "Hybrid approach for solving multivariate systems over finite fields".

Hybrid approach II

Cost of solving $\mathcal{S}_{\text{spec},f}$? Same assumptions as for $\mathcal{S}$, same analysis:

$$\mathcal{H}_{R/\langle \mathcal{S}^{(h)}_{\text{spec},f}\rangle}(z) = \frac{(1+(N-1-u)z)^t}{(1+z)^{n-k}}$$

Final complexity:

$$\mathcal{O}\left(\min_{0\le u\le N-1}\left\{(1-u/N)^{-t}\times T_{\text{solve}}(\mathcal{S}_{\text{spec},u\cdot t})\right\}\right)$$

- Other ways to fix zeroes (inspired by ISDs ?). We analyze one more in the paper.

From Dense to Sparse (under the carpet)

Use **sparse** linear algebra: $\searrow T_{\text{solve}}(.)$?

- Need XL-Wiedemann instead of Gröbner Basis
- Kernel of affine Macaulay matrix

XL at conjectured D may fail !

(need other parameter attached to affine systems: *witness degree*)

Experimental Verification

We relied on Magma

- Check Assumption: compute HS for both $\mathcal{S}^{(h)}$ and $\mathcal{S}^{(h)}_{\text{spec},f}$ (various f)
- Check Degree Fall assumption: steps of Magma's F4 on affine system
- To do: show that XL can work

Conclusion

Conjectured cost with Wiedemann

Parameters from Boyle *et al.* [Boy+19], updated analysis by Liu *et al.* [Liu+22]

Large field: no more $\{\forall i,\ \sum_{j=1}^{N} e_{i,j} = 1\}$, fields eqs of high degree (that's ok)

n	k	t	$\mathbb{F}_2$ [Liu+22]	This work $\mathbb{F}_2$	$\mathbb{F}_{2^{128}}$ [Liu+22]	This work $\mathbb{F}_{2^{128}}$
2^{22}	64770	4788	147	**104**	156	**111**
2^{20}	32771	2467	143	$\underline{\underline{126}}$	155	$\underline{\underline{131}}$
2^{18}	15336	1312	139	$\underline{\underline{123}}$	153	$\underline{\underline{133}}$
2^{16}	7391	667	135	141	151	151
2^{14}	3482	338	132	140	150	152
2^{12}	1589	172	131	136	155	$\underline{152}$
2^{10}	652	106	176	**146**	194	$\underline{\underline{180}}$

[Liu+22] Liu et al. *The Hardness of LPN over Any Integer Ring and Field for PCG Applications*.

More on the results

- Sometimes beats Gauss/ISDs for low rates (Primal LPN)
- Zone with "constant" deg. $D \to$ polynomial algorithm ?

Similar to Arora-Gê modeling on LWE [AG11]
(Polynomial for sufficiently many samples)

[AG11] Arora and Ge. "New Algorithms for Learning in Presence of Errors". *Automata, Languages and Programming*.