Information Set Decoding in the Lee Metric

Violetta Weger

joint work with Franco Chiaraluce, Marco Baldi, Massimo Battaglioni, Anna-Lena Horlemann-Trautmann, Edoardo Persichetti and Paolo Santini

University of Zurich

University of Zurich[UZH]

CBCrypto 2020

Changing the Metric

- The original McEliece cryptosystem using Goppa codes remains unbroken but suffers from large key sizes.
- Many attempts of fixing this issue by exchanging the family of codes.
- Example: Niederreiter proposed to use GRS codes, which have the highest error correction capacity, hence promise low key sizes, but are vulnerable to algebraic attacks.
- Within the 7 code-based cryptosystems in the NIST round 2, the ones that are achieving the lowest key sizes are based on the rank metric.

Rank Metric

Definition (Rank Metric)

For $A, B \in \mathbb{F}_q^{m \times n}$ we define the rank weight to be $wt_R(A) = rk(A)$ and the rank distance between A and B to be $d_R(A, b) = wt_R(A - B)$.

Definition ($\mathbb{F}_q$-linear Rank Metric Code)

$\mathcal{C}$ is a $\mathbb{F}_q$-linear rank metric code of length n and dimension k, if $\mathcal{C}$ is a k -dimensional linear subspace of $Mat_{m \times n}(\mathbb{F}_q)$ equipped with the rank metric.

Rank Metric

Definition (Rank Metric)

For $x, y \in \mathbb{F}_{q^m}^n$ we define the rank weight to be $wt_R(x) = dim(\langle x_1, \ldots, x_n \rangle_{\mathbb{F}_q})$ and the distance between x and y to be $d_R(x, y) = wt_R(x - y)$.

Definition ($\mathbb{F}_{q^m}$-linear Rank Metric Code)

$\mathcal{C}$ is a $\mathbb{F}_{q^m}$-linear rank metric code of length n and dimension k, if $\mathcal{C}$ is a k -dimensional linear subspace of $\mathbb{F}_{q^m}^n$ equipped with the rank metric.

Note: all $\mathbb{F}_{q^m}$-linear rank metric codes are also $\mathbb{F}_q$-linear rank metric codes.

Difference between Rank and Hamming Metric

Let $x \in \mathbb{F}_{q^m}^n$.

	Hamming	Rank
$\mathrm{Supp}(x)$	$\{1 \leq i \leq n \mid x_i \neq 0\}$	$\langle x_1, \ldots, x_n \rangle_{\mathbb{F}_q}$
$\mathrm{wt}(x)$	$\mid \mathrm{Supp}(x) \mid$	$\dim(\mathrm{Supp}(x))$
Bruteforce cost	$\binom{n}{t}(q^m-1)^t$	$\begin{bmatrix} m \\ t \end{bmatrix}_q = \prod_{i=0}^{t-1} \frac{q^m - q^i}{q^t - q^i} \sim q^{(m-t)t}$

Difference between Rank and Hamming Metric

	Hamming	Rank
Advantages	NP-complete studied thoroughly	SDP more costly low key sizes
Disadvantages	large key sizes	not studied thoroughly only randomized reduction

Properties

Definition (Lee Weight)

Let $x \in \mathbb{Z}/m\mathbb{Z}$, *then* $wt_L(x) = \min\{x, \mid m - x \mid\}$.

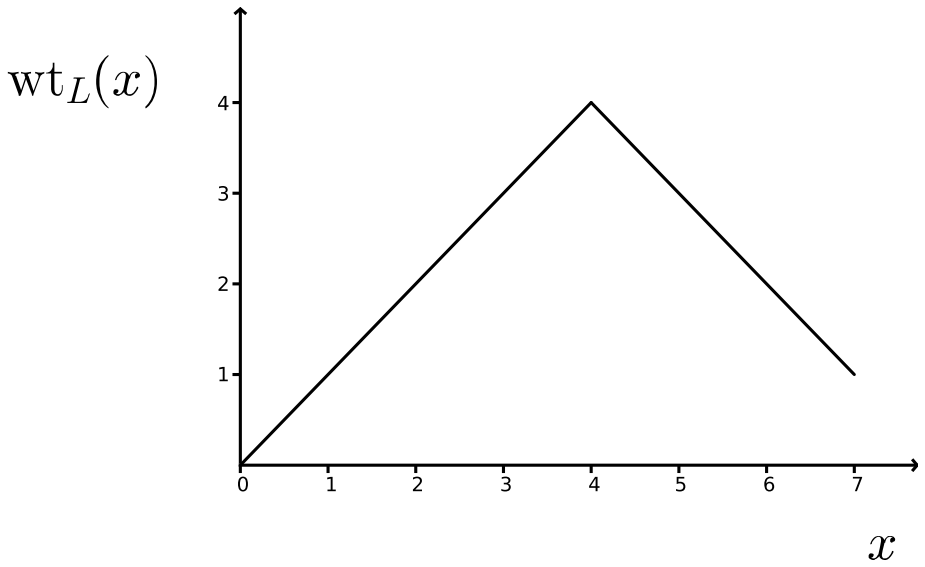
Example ($\mathbb{Z}/8\mathbb{Z}$)

$$\begin{aligned} \mathrm{wt}_L(0) &= 0 \\ \mathrm{wt}_L(1) = \mathrm{wt}_L(7) &= 1 \\ \mathrm{wt}_L(2) = \mathrm{wt}_L(6) &= 2 \\ \mathrm{wt}_L(3) = \mathrm{wt}_L(5) &= 3 \\ \mathrm{wt}_L(4) &= 4 \end{aligned}$$

Properties

Definition (Lee Weight)

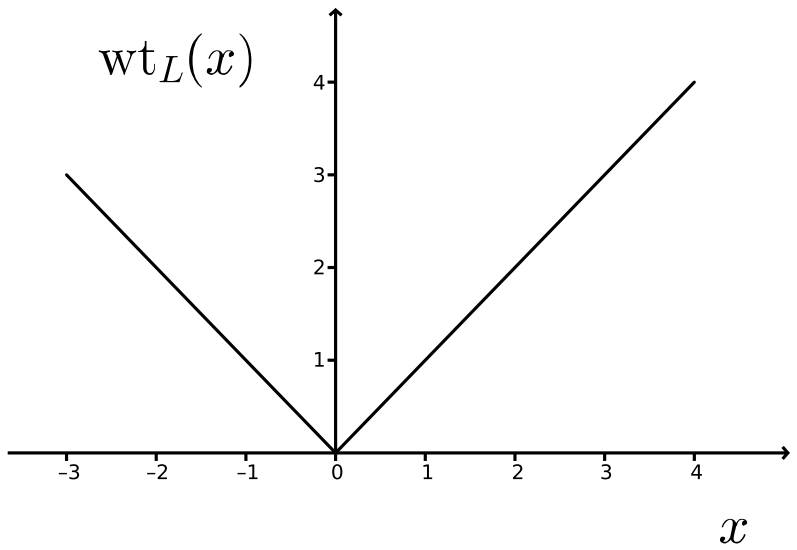
Let $x \in \mathbb{Z}/m\mathbb{Z}$, *then* $wt_L(x) = \min\{x, \mid m - x \mid\}$.



Properties

Definition (Lee Weight)

Let $x \in \mathbb{Z}/m\mathbb{Z}$, *then* $wt_L(x) = \min\{x, \mid m - x \mid\}$.



Properties

Definition (Lee Metric)

Let $x, y \in (\mathbb{Z}/m\mathbb{Z})^n$, *then the Lee weight is defined as* $wt_L(x) = \sum_{i=1}^{n} wt_L(x_i)$ *and the Lee distance between* x *and* y *is* $d_L(x, y) = wt_L(x - y)$.

Clearly: For all $x \in (\mathbb{Z}/m\mathbb{Z})^n : \mathrm{wt}_H(x) \leq \mathrm{wt}_L(x)$.

Definition (Lee Metric Code)

$\mathcal{C}$ *is a linear Lee metric code of length* n *and type* $|\mathcal{C}|$, *if* $\mathcal{C}$ *is an additive subgroup of* $(\mathbb{Z}/m\mathbb{Z})^n$ *equipped with the Lee metric.*

Quaternary Codes

Definition (Quaternary Code)

$\mathcal{C}$ is a quaternary code of length n and type $4^{k_1}2^{k_2}$, if $\mathcal{C}$ is an additive subgroup of $(\mathbb{Z}/4\mathbb{Z})^n$ equipped with the Lee metric.

Definition (Gray Isometry)

$$\varphi : (\mathbb{Z}/4\mathbb{Z}, wt_L) \to (\mathbb{F}_2^2, wt_H)$$
$$0 \mapsto (0,0)$$
$$1 \mapsto (0,1)$$
$$2 \mapsto (1,1)$$
$$3 \mapsto (1,0)$$

We can extend $\varphi_n : (\mathbb{Z}/4\mathbb{Z})^n \to \mathbb{F}_2^{2n}$.

Quaternary Codes

Definition (Quaternary Code)

$\mathcal{C}$ is a quaternary code of length n and type $4^{k_1}2^{k_2}$, if $\mathcal{C}$ is an additive subgroup of $(\mathbb{Z}/4\mathbb{Z})^n$ equipped with the Lee metric.

Definition (Gray Isometry)

$$\begin{aligned}\varphi : (\mathbb{Z}/4\mathbb{Z}, wt_L) &\to (\mathbb{F}_2^2, wt_H)\\ 0 &\mapsto (0,0)\\ 1 &\mapsto (0,1)\\ 2 &\mapsto (1,1)\\ 3 &\mapsto (1,0)\end{aligned}$$

We can extend $\varphi_n : (\mathbb{Z}/4\mathbb{Z})^n \to \mathbb{F}_2^{2n}$.

Differences

Let $\mathcal{C}$ be a quaternary code of length n and type $4^{k_1}2^{k_2}$, then the systematic form of the generator matrix is given by

$$G = \begin{pmatrix} \mathrm{Id}_{k_1} & A & B \\ 0 & 2\mathrm{Id}_{k_2} & 2C \end{pmatrix},$$

where $A \in \mathbb{Z}_2^{k_1 \times k_2}$, $B \in \mathbb{Z}_4^{k_1 \times (n-k_1-k_2)}$, $C \in \mathbb{Z}_2^{k_2 \times (n-k_1-k_2)}$.
The systematic form of the parity check matrix is given by

$$H = \begin{pmatrix} D & E & \mathrm{Id}_{n-k_1-k_2} \\ 2F & 2\mathrm{Id}_{k_2} & 0 \end{pmatrix},$$

where $D \in \mathbb{Z}_4^{(n-k_1-k_2)\times k_1}$, $E \in \mathbb{Z}_4^{(n-k_1-k_2)\times k_2}$, $F \in \mathbb{Z}_2^{k_2 \times k_1}$.

ISD over the Hamming Metric

Prange's algorithm:
Given: $H \in \mathbb{F}_q^{(n-k)\times n}$, $s \in \mathbb{F}_q^{n-k}$, $t \in \mathbb{N}$.
Find: $e \in \mathbb{F}_q^n$, such that $He^\top = s^\top$ and $\mathrm{wt}_H(e) = t$.
Main idea: Assume no error happen in the information set.

$$UHe^\top = \begin{pmatrix} A & \mathrm{Id}_{n-k} \end{pmatrix} \begin{pmatrix} 0 \\ e'^\top \end{pmatrix} = Us^\top.$$

Thus we get the condition $e'^\top = Us^\top$.

Structure of ISD Algorithms

1. Choose an information set.
2. Bring the parity check matrix into systematic form and perform the same row operations on the syndrome.
3. By assuming a certain weight distribution of the error vector we get conditions on the error vector.
4. Go through all possible vectors and check if conditions are satisfied, if they are output the error vector.
5. If not, start over with a new information set.

Cost of ISD Algorithms

The cost of an ISD algorithm is given by

$$\text{number of iterations} \cdot \text{cost of one iteration}.$$

number of iterations = reciprocal of the success probability of one iteration.
Example:
Prange in the Hamming metric has a success probability of

$$\binom{n-k}{t}\binom{n}{t}^{-1}.$$

Quaternary Prange

Given: $H \in \mathbb{Z}_4^{(n-k_1)\times n}, s \in \mathbb{Z}_4^{n-k_1}, t \in \mathbb{N}$.
Find: $e \in \mathbb{Z}_4^n$ with $He^\top = s^\top$ and $\mathrm{wt}_L(e) = t$.

$$UHe^\top = \begin{pmatrix} A & \mathrm{Id}_{n-k_1-k_2} \\ 2C & 0 \end{pmatrix} \begin{pmatrix} 0 \\ e'^\top \end{pmatrix} = \begin{pmatrix} s_1^\top \\ 2s_2^\top \end{pmatrix}.$$

From this we get the conditions $e' = s_1$ and $s_2 = 0$.
New success probability:

$$\binom{2(n-k_1-k_2)}{t} \binom{2n}{t}^{-1}.$$

GV - Bounds

Proposition (Gilbert-Varshamov Bound)

Let n and d be positive integers. There exists a linear binary code $\mathcal{C}$ of length n and minimum Hamming distance d, such that

$$|\mathcal{C}| \geq \frac{2^n}{\sum_{j=0}^{d-1} \binom{n}{j}}.$$

Furthermore there exists a linear quaternary code $\mathcal{C}$ of length n and minimum Lee distance d, such that

$$|\mathcal{C}| \geq \frac{4^n}{(\sum_{j=0}^{d-1} \binom{2n}{j} - 1)3 + 1}.$$

Performance for theoretical Parameters

In the Lee metric:

n	k_1	k_2	d_L	t_L	cost Prange	Key Size
101	5	90	25	12	83.42	1050
463	230	3	105	52	80.29	107180
173	9	154	41	20	129.96	3106
863	430	3	193	96	128.82	372380
375	20	334	85	42	256.03	14534
1943	970	3	431	215	256.33	1887620

In the Hamming metric:

n	k	d_H	t_H	cost Prange	Key Size
903	451	103	51	80.53	203852
1683	841	189	94	128.03	708122
3863	1931	429	214	256.68	3730692

Disclaimer

These are only theoretical parameters, since we are not actually proposing a code to be used within the quaternary McEliece cryptosystem!

Difficulties of Generalizing

Let $\mathcal{C}$ be a linear Lee metric code over $\mathbb{Z}_{p^s}$ of length n and type $(p^s)^{k_1}(p^{s-1})^{k_2}\dots p^{k_s}$. Then the systematic form of the generator matrix is

$$G = \begin{pmatrix} \mathrm{Id}_{k_1} & A_{1,2} & \dots & A_{1,s} & A_{1,s+1} \\ 0 & p\mathrm{Id}_{k_2} & \dots & pA_{2,s} & pA_{2,s+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & p^{s-1}\mathrm{Id}_{k_s} & p^{s-1}A_{s,s+1} \end{pmatrix},$$

and the systematic form of the parity check matrix is

$$H = \begin{pmatrix} B_{1,1} & B_{1,2} & \dots & B_{1,s} & \mathrm{Id}_{n-K} \\ pB_{2,1} & pB_{2,2} & \dots & p\mathrm{Id}_{k_s} & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ p^{s-1}B_{s,1} & p^{s-1}\mathrm{Id}_{k_2} & \dots & 0 & 0 \end{pmatrix},$$

where $K = \sum_{i=1}^{s} k_i$.

Simplification for ISD

For the purpose of ISD algorithms we can choose the following form

$$G = \begin{pmatrix} \mathrm{Id}_{k_1} & A \\ 0 & pB \end{pmatrix}, \quad H = \begin{pmatrix} C & \mathrm{Id}_{n-K} \\ pD & 0 \end{pmatrix},$$

with $A \in \mathbb{Z}_{p^s}^{k_1 \times (n-k_1)}$, $B \in \mathbb{Z}_{p^{s-1}}^{(K-k_1)\times(n-k_1)}$, $C \in \mathbb{Z}_{p^s}^{(n-K)\times K}$ and $D \in \mathbb{Z}_{p^{s-1}}^{(K-k_1)\times K}$.

This way we are putting all the zero-divisors together, only considering k_1.

Simplification for ISD

Example: Lee-Brickell
We assume that the error vector has weight v in the information set and $t - v$ outside the information set.

$$UHe^\top = \begin{pmatrix} C & \mathrm{Id}_{n-K} \\ pD & 0 \end{pmatrix} \begin{pmatrix} e_1^\top \\ e_2^\top \end{pmatrix} = \begin{pmatrix} s_1^\top \\ ps_2^\top \end{pmatrix} = Us^\top.$$

From this we get the conditions

$$\begin{aligned} Ce_1^\top + e_2^\top &= s_1^\top \\ pDe_1^\top &= ps_2^\top \end{aligned}$$

Note that the second condition is again a syndrome decoding problem, but over a smaller ring and of smaller size.

Open Problems

- Find quaternary code with the properties from code-based cryptography: large error correction capacity, efficient decoding algorithm and a large family of codes.
- Find applications of the Lee metric for code-based cryptography, ongoing work: identification scheme, signature scheme.
- Computation of the cost of the iterative ISD algorithm.
- Is there a faster way to solve the SDP using tools from lattice-based cryptography?

Thank you!