Code-Based Cryptography in Different Metrics II

Anna-Lena Horlemann

Institute of Computer Science, University of St.Gallen, Switzerland

attacc, DLR Oberpfaffenhofen, 2024

1. Lee Metric
2. Subspace Metric
3. Overall Summary and Conclusions

Lee metric and isometries

Definition (Lee distance)

For $u, v \in \mathbb{Z}_m^n$:

$$d_L((u_1, \ldots, u_n), (v_1, \ldots, v_n)) := \sum_{i=1}^{n} \min\{|v_i - u_i|, |m - (v_i - u_i)|)\}$$

Theorem

The linear isometries for the metric space $(\mathbb{Z}_m^n, d_L)$ are represented by the monomial matrices with non-zero entries in $\{1, -1\}$.

A Lee McEliece scheme

- The receiver chooses an alternant code $C \subseteq \mathbb{Z}_p^n$ of dimension k with generator matrix G and minimum Lee distance $2t+1$. Moreover, they choose a random Lee isometry M and some $A \in \mathrm{GL}_k(p)$.
- **Private key:** G and (A, M) (decoder for C given by G)
- **Public key:** $G' = AGM$
- **Encryption:** Choose a random error vector e of Lee weight at most t and encrypt the message m as

$$c = m\,G' + e.$$

- **Decryption:** Compute $c' = cP^{-1}$ and decode this in C to recover mA. Recover m by multiplying with A^{-1}.

Digital signatures

We can do the same as in Hamming or rank metric...

- Hash and sign
  - given a message m, find a random m' such that $Hash(m, H_{pub}, m')$ is a decodable syndrome (error vector is main part of signature)
  - Problem: finding this m' is inefficient

Digital signatures

We can do the same as in Hamming or rank metric...

- Hash and sign
  - given a message m, find a random m' such that $Hash(m, H_{pub}, m')$ is a decodable syndrome (error vector is main part of signature)
  - Problem: finding this m' is inefficient
- Fiat-Shamir transform of zero-knowledge identification (ZKID) schemes
  - public information: random parity check matrix
  - secret: decodable error vector
  - interactive (5-pass) scheme, where one reveals either the weight of the error vector, or that it is a solution to the syndrome equation

Digital signatures

We can do the same as in Hamming or rank metric...

- Hash and sign
  - given a message m, find a random m' such that $Hash(m, H_{pub}, m')$ is a decodable syndrome (error vector is main part of signature)
  - Problem: finding this m' is inefficient
- Fiat-Shamir transform of zero-knowledge identification (ZKID) schemes
  - public information: random parity check matrix
  - secret: decodable error vector
  - interactive (5-pass) scheme, where one reveals either the weight of the error vector, or that it is a solution to the syndrome equation

Problem: Because of the nature of the isometries we reveal a lot information about the whole error vector.

$\Longrightarrow$ restricted error model (and hence no more Lee metric)

Minimum distance of random codes

In the Lee metric there is no one analog of the Singleton bound, and all known bounds are not attainable for most parameters.

Theorem

For $n \to \infty$ a random linear code in $\mathbb{F}_q^n$ gets arbitrarily close to the Gilbert-Varshamov bound by probability 1.

$\implies$ Use these results as an estimate for the minimum distance of the random linear code in an ZKID scheme.

— Attacks —

Generic decoding (message recovery attacks)

Question: What is the right approach in the Lee metric?

- **ISD approach:** ISD (and follow-up approaches) find vectors with a lot of zeros. This is a good indicator for a low Lee weight, as well (but not exactly the same). We have a few Hamming-ISD adaptations to the Lee metric.
- **Lattice approach:** The Lee metric corresponds to the ℓ_1-metric over $\mathbb{Z}_m$. For large m Lee metric decoding in $\mathbb{Z}_m^n$ corresponds to LWE in the corresponding lattice in the ℓ_1-norm. In some cases this can be approximated with ℓ_2-norm.

$$\text{Hamming metric} \xleftarrow{\text{small } m} \text{Lee metric on } \mathbb{Z}_m^n \xrightarrow{\text{large } m} \ell_1\text{-lattice}$$

1. Lee Metric

2. Subspace Metric

3. Overall Summary and Conclusions

Quick reminder

Definition

Denote by $\mathcal{P}_q(n)$ the set of all subspaces of $\mathbb{F}_q^n$ and by $\mathcal{G}_q(k,n)$ the set of all k-dimensional subspaces of $\mathbb{F}_q^n$ ("Grassmannian").

1. A subset $\mathcal{C} \subseteq \mathcal{P}_q(n)$ is called a **subspace code**. If $\mathcal{C} \subseteq \mathcal{G}_q(k,n)$, then it is also called a **constant-dimension code**.
2. The **subspace distance** on $\mathcal{P}_q(n)$ is defined as

$$d_S(\mathcal{U},\mathcal{V}) := \dim(\mathcal{U}) + \dim(\mathcal{V}) - 2\dim(\mathcal{U} \cap \mathcal{V}).$$

Problems to solve

1. Efficient representation of the public key ("linearity"):

2. Distance measure for the correct solution ("weight"):

Problems to solve

1. Efficient representation of the public key ("linearity"):
   - lifted rank-metric codes

2. Distance measure for the correct solution ("weight"):
   - lifted rank-metric codes – rank weight ($\cong$ subsp. dist. to rs$[I \mid 0]$)

Problems to solve

1. Efficient representation of the public key ("linearity"):
   - lifted rank-metric codes
   - orbit codes – generators of the group in GL_n defining the orbit code
2. Distance measure for the correct solution ("weight"):
   - lifted rank-metric codes – rank weight ($\cong$ subsp. dist. to $\mathrm{rs}[I \mid 0]$)
   - orbit codes – subspace distance to a prescribed "zero" codeword

Problems to solve

1. Efficient representation of the public key ("linearity"):
   - lifted rank-metric codes
   - orbit codes – generators of the group in GL_n defining the orbit code
2. Distance measure for the correct solution ("weight"):
   - lifted rank-metric codes – rank weight ($\cong$ subsp. dist. to $\mathrm{rs}[I \mid 0]$)
   - orbit codes – subspace distance to a prescribed "zero" codeword
3. For McEliece/Niederreiter type systems we also need an efficient decoding algorithm:
   - lifted rank-metric codes – rank metric decoders
   - orbit codes – ???

McEliece with lifted Gabidulin codes

- **Secret key:** Gabidulin code $\mathcal{C}_{Gab} \subseteq \mathbb{F}_{q^{n-k}}^{\kappa \times k}$
- **Public key:** Generator matrix G_{pub} of $\mathcal{C}_{pub} := \phi(\mathcal{C}_{Gab})$ [1]
- **Encryption (encoding plus random subspace errors):**

$$\mathcal{D}_\rho(\mathrm{rs}[I_k \mid \underbrace{mG_{pub}}_{\text{expanded over } \mathbb{F}_q}]) \oplus \mathcal{E}$$

  such that $\rho + \dim(\mathcal{E}) \leq t$ (error correction capability)
- **Decryption:** Use lifted Gabidulin decoder with application of ϕ^{-1}

[1] ϕ can be any valid rank-metric disguising function.

Decoding with transformation to secret code

Decoding the ciphertext in the received word can be translated to

$$\underset{X \in \mathcal{C}_{pub}}{\operatorname{argmin}} \operatorname{rk} \begin{pmatrix} \hat{L} & X - R \\ 0 & \hat{E} \end{pmatrix} = \underset{X' \in \mathcal{C}_{Gab}}{\operatorname{argmin}} \operatorname{rk} \begin{pmatrix} \bar{L} & X' - \phi^{-1}(R) \\ 0 & \bar{E} \end{pmatrix}$$

which is in turn equivalent to a rank-metric decoding problem with row and column erasures.
($\hat{L}, \hat{E}, R$ are given by the structure of the cipher vector space; $\bar{L}, \bar{E}$ also depend on ϕ.)

Decoding with transformation to secret code

Decoding the ciphertext in the received word can be translated to

$$\underset{X \in \mathcal{C}_{pub}}{\operatorname{argmin}} \operatorname{rk} \begin{pmatrix} \hat{L} & X - R \\ 0 & \hat{E} \end{pmatrix} = \underset{X' \in \mathcal{C}_{Gab}}{\operatorname{argmin}} \operatorname{rk} \begin{pmatrix} \bar{L} & X' - \phi^{-1}(R) \\ 0 & \bar{E} \end{pmatrix}$$

which is in turn equivalent to a rank-metric decoding problem with row and column erasures.
($\hat{L}, \hat{E}, R$ are given by the structure of the cipher vector space; $\bar{L}, \bar{E}$ also depend on ϕ.)

$\Longrightarrow$ For both the receiver and the attacker it is equivalent to a rank-metric decoding problem with row/column erasures.

Decoding with transformation to secret code

Decoding the ciphertext in the received word can be translated to

$$\underset{X \in \mathcal{C}_{pub}}{\operatorname{argmin}} \operatorname{rk} \begin{pmatrix} \hat{L} & X - R \\ 0 & \hat{E} \end{pmatrix} = \underset{X' \in \mathcal{C}_{Gab}}{\operatorname{argmin}} \operatorname{rk} \begin{pmatrix} \bar{L} & X' - \phi^{-1}(R) \\ 0 & \bar{E} \end{pmatrix}$$

which is in turn equivalent to a rank-metric decoding problem with row and column erasures.
($\hat{L}, \hat{E}, R$ are given by the structure of the cipher vector space; $\bar{L}, \bar{E}$ also depend on ϕ.)

$\Longrightarrow$ For both the receiver and the attacker it is equivalent to a rank-metric decoding problem with row/column erasures.

$\Longrightarrow$ Subspace metric not necessary, can just use rank metric.

Multiplicative analog of linearity via orbit codes

- Group theoretic (multiplicative instead of additive) analog of linear codes: orbit codes in $\mathcal{G}_q(k,n)$

Definition

Let $G \leq \mathrm{GL}_n$ be a group and $\mathcal{U}_0 \in \mathcal{G}_q(k,n)$. Then $\mathcal{U}_0 G$ is an orbit code in $\mathcal{G}_q(k,n)$.

[2] except for the cases where the orbit code is also a lifted MRD or spread code

Multiplicative analog of linearity via orbit codes

- Group theoretic (multiplicative instead of additive) analog of linear codes: orbit codes in $\mathcal{G}_q(k, n)$

Definition

Let $G \leq \mathrm{GL}_n$ be a group and $\mathcal{U}_0 \in \mathcal{G}_q(k, n)$. Then $\mathcal{U}_0 G$ is an orbit code in $\mathcal{G}_q(k, n)$.

- No efficient decoders known (yet)[2] $\implies$ not usable for McEliece
- But for identification scheme?!

[2] except for the cases where the orbit code is also a lifted MRD or spread code

Theoretical setup for McEliece with orbit codes

- **Secret key:** Generators of orbit code $\mathcal{C}_{orb} = \mathcal{U}_0 G \subseteq \mathcal{G}_q(k, n)$
- **Public key:** Generators of disguised code $\mathcal{C}_{pub} := \phi(\mathcal{C}_{orb})$
- **Encryption (encoding plus random subspace errors):**

$$\mathcal{D}_\rho(\mathrm{rs}[I_k \mid \underbrace{mG_{pub}}_{\text{expanded over } \mathbb{F}_q}]) \oplus \mathcal{E}$$

  such that $\rho + \dim(\mathcal{E}) \leq t$ (error correction capability)
- **Decryption:** Use orbit decoder with application of ϕ^{-1}

Theoretical setup for McEliece with orbit codes

- **Secret key:** Generators of orbit code $\mathcal{C}_{orb} = \mathcal{U}_0 G \subseteq \mathcal{G}_q(k, n)$
- **Public key:** Generators of disguised code $\mathcal{C}_{pub} := \phi(\mathcal{C}_{orb})$
- **Encryption (encoding plus random subspace errors):**

$$\mathcal{D}_\rho(\mathrm{rs}[I_k \mid \underbrace{mG_{pub}}_{\text{expanded over } \mathbb{F}_q}]) \oplus \mathcal{E}$$

  such that $\rho + \dim(\mathcal{E}) \leq t$ (error correction capability)
- **Decryption:** Use orbit decoder with application of ϕ^{-1}

Questions:

1. What could ϕ be?
   It should keep the orbit structure (for representability), but hide the structure of the secret code.
2. Do we find orbit codes with an efficient decoder?

Idea for a subspace metric ZK-ID scheme

- **Secret:** coset leader $\mathcal{V} \in \mathcal{G}_q(k, n)$, s.t.
  - $\operatorname{argmin}_{B \in G} d_S(\mathcal{U}_0, \mathcal{V}B) = I_n$
  - $d_S(\mathcal{U}_0, \mathcal{V}) = t$
- **Public information:**
  - group $G \leq \mathrm{GL}_n(q)$ and $\mathcal{U}_0 \in \mathcal{G}_q(k, n)$ (orbit code $\mathcal{C} := \mathcal{U}_0 G$)
  - an identifier S of the orbit $\mathcal{V}G$
  - distance t
- **Interactive protocol:** Prove to the verifier one of the two per round:
  - secret is on the orbit $\mathcal{V}G$
  - secret has subspace distance t to $\mathcal{U}_0$

Idea for a subspace metric ZK-ID scheme

- **Secret:** coset leader $\mathcal{V} \in \mathcal{G}_q(k, n)$, s.t.
  - $\operatorname{argmin}_{B \in G} d_S(\mathcal{U}_0, \mathcal{V}B) = I_n$
  - $d_S(\mathcal{U}_0, \mathcal{V}) = t$
- **Public information:**
  - group $G \leq \mathrm{GL}_n(q)$ and $\mathcal{U}_0 \in \mathcal{G}_q(k, n)$ (orbit code $\mathcal{C} := \mathcal{U}_0 G$)
  - an identifier S of the orbit $\mathcal{V}G$
  - distance t
- **Interactive protocol:** Prove to the verifier one of the two per round:
  - secret is on the orbit $\mathcal{V}G$
  - secret has subspace distance t to $\mathcal{U}_0$

Questions:

- How to implement the interactive protocol (computationally)?
- What could the orbit identifier ($\cong$ syndrome) be?
- How difficult is the general coset leader decoding problem for orbit codes ($\Longrightarrow$ security)?

Essential open problems

1. We need a complexity estimate for a generic orbit decoder in $\mathcal{G}_q(k,n) \implies$ security level
2. We need a syndrome-like identifier for the orbits, and a corresponding map $\odot$ such that we can recover the orbit $\mathcal{U}_0 UG$ from the orbit $\mathcal{U}_0 UEG$.
   (Non-commutativity makes this problem really hard.)
3. We need a "$\mathcal{U}_0$-isometry" σ with $d_S(\mathcal{U}_0 E, \mathcal{U}_0) = d_S(\mathcal{U}_0 \sigma(E), \mathcal{U}_0)$ and $\sigma(UE) = \sigma(U)\sigma(E)$.
4. The maps/operators need to come from large enough sets to make it cryptographically secure.

1. Lee Metric

2. Subspace Metric

3. Overall Summary and Conclusions

Summary and conclusions

- Hamming metric well studied for CBC.
- Other metrics promise better parameters, since generic decoding is more difficult.
- However, lots of open questions/problems in the other metrics.

	Hamming	rank	sum-rank	Lee	subspace
NP-hard	YES	probabilistic	YES/probl.	YES	?
good random codes	YES	YES	YES	YES	?
distinguishers	star product	Frobenius intersection/sum	star + Frobenius (in trivial case)	?	?
generic decoding	x	$> x$	$> x$	$> x$ lattice tech. ?!	?

Summary and conclusions

- Hamming metric well studied for CBC.
- Other metrics promise better parameters, since generic decoding is more difficult.
- However, lots of open questions/problems in the other metrics.

	Hamming	rank	sum-rank	Lee	subspace
NP-hard	YES	probabilistic	YES/probl.	YES	?
good random codes	YES	YES	YES	YES	?
distinguishers	star product	Frobenius intersection/sum	star + Frobenius (in trivial case)	?	?
generic decoding	x	$> x$	$> x$	$> x$ lattice tech. ?!	?

Thank you for your attention!

Questions? – Comments?