Generalization of the Ball-Collision Algorithm

Violetta Weger

joint work with Carmelo Interlando, Karan Khathuria, Nicole Rohrer and Joachim Rosenthal

University of Zurich

7th Code-Based Cryptography Workshop

19 May 2019

Outline

1. Motivation
2. Introduction
3. Prange's Algorithm
4. Improvements overview
5. Ball-collision Algorithm
6. New directions
7. Comparison of Complexities
8. Open questions

Motivation

Proposing a code-based cryptosystem

- Structural Attacks
- Nonstructural Attacks

Have to consider Information Set Decoding (ISD)

Motivation

Proposing a code-based cryptosystem

- Structural Attacks
- Nonstructural Attacks

Have to consider Information Set Decoding (ISD)

ISD algorithms and syndome decoding problem

1978 Berlekamp, McEliece and van Tilborg: Decoding a random linear code is NP-complete

Problem (Syndrome decoding problem)

Given a parity check matrix H of a (binary) code of length n and dimension k and a syndrome s:

$$s = Hx^{\mathsf{T}} \in \mathbb{F}_2^{n-k}$$

and the error correction capacity t, we want to find $e \in \mathbb{F}_2^n$ of weight t such that

$$s = He^{\mathsf{T}}.$$

ISD algorithms and syndome decoding problem

- Syndrome decoding problem is equivalent to the decoding problem and

Problem (Decoding problem)

Given a generator matrix G of a (binary) code of length n and dimension k and a corrupted codeword c:

$$c = mG + e \in \mathbb{F}_2^n$$

and the error correction capacity t, we want to find $e \in \mathbb{F}_2^n$ of weight t.

- equivalent to finding a minimum weight codeword, since in $\mathcal{C} + \{0, c\}$ the error vector e is now the minimum weight codeword.

Informationset

Notation

Let $c \in \mathbb{F}_q^n$ and $A \in \mathbb{F}_q^{k \times n}$, let $S \subset \{1, \ldots, n\}$, then we denote by c_S the restriction of c to the entries indexed by S and by A_S the columns of A indexed by S. For a code $\mathcal{C} \subset \mathbb{F}_q^n$, we denote by

$$\mathcal{C}_S = \{c_S \mid c \in \mathcal{C}\}.$$

Definition (Informationset)

Let $\mathcal{C} \subset \mathbb{F}_q^n$ be a code of dimension k. If $I \subset \{1, \ldots, n\}$ of size k is such that

$$\mid \mathcal{C} \mid = \mid \mathcal{C}_I \mid,$$

then we call I an information set of $\mathcal{C}$.

Informationset

Notation

Let $c \in \mathbb{F}_q^n$ *and* $A \in \mathbb{F}_q^{k \times n}$, *let* $S \subset \{1, \dots, n\}$, *then we denote by* c_S *the restriction of* c *to the entries indexed by* S *and by* A_S *the columns of* A *indexed by* S. *For a code* $\mathcal{C} \subset \mathbb{F}_q^n$, *we denote by*

$$\mathcal{C}_S = \{c_S \mid c \in \mathcal{C}\}.$$

Definition (Informationset)

Let $\mathcal{C} \subset \mathbb{F}_q^n$ *be a code of dimension* k. *If* $I \subset \{1, \dots, n\}$ *of size* k *is such that*

$$\mid \mathcal{C} \mid = \mid \mathcal{C}_I \mid,$$

then we call I *an information set of* $\mathcal{C}$.

Informationset

Definition (Informationset)

Let G be the $k \times n$ generator matrix of $\mathcal{C}$. If $I \subset \{1, \ldots, n\}$ of size k is such that G_I is invertible, then I is an informationset of $\mathcal{C}$.

Definition (Informationset)

Let H be the $n - k \times n$ parity check matrix of $\mathcal{C}$. If $I \subset \{1, \ldots, n\}$ of size k is such that H_{I^c} is invertible, then I is an informationset of $\mathcal{C}$.

Informationset

Definition (Informationset)

Let G be the $k \times n$ generator matrix of $\mathcal{C}$. If $I \subset \{1, \ldots, n\}$ of size k is such that G_I is invertible, then I is an informationset of $\mathcal{C}$.

Definition (Informationset)

Let H be the $n-k \times n$ parity check matrix of $\mathcal{C}$. If $I \subset \{1, \ldots, n\}$ of size k is such that H_{I^c} is invertible, then I is an informationset of $\mathcal{C}$.

Prange's algorithm

1962 Prange proposes the first ISD algorithm.
Assumption: All t errors occur outside of the information set.

Input: $H \in \mathbb{F}_2^{n-k \times n}, s \in \mathbb{F}_2^{n-k}, t \in \mathbb{N}$
Output: $e \in \mathbb{F}_2^n, wt(e) = t$ and $He^\intercal = s$.

1 Choose an information set $I \subset \{1, \ldots, n\}$ of size k.

2 Find an invertible matrix $U \in \mathbb{F}_2^{n-k \times n-k}$ such that $(UH)_I = A$ and $(UH)_{I^c} = \mathrm{Id}_{n-k}$.

3 If $wt(Us) = t$, then $e_I = 0$ and $e_{I^c} = Us$.

4 Else start over.

Prange's algorithm

1. Choose an information set $I \subset \{1, \ldots, n\}$ of size k.

Let us assume for simplicity that $I = \{1, \ldots, k\}$.

Prange's algorithm

1. Choose an information set $I \subset \{1, \dots, n\}$ of size k.
2. Find an invertible matrix $U \in \mathbb{F}_2^{n-k \times n-k}$ such that $(UH)_I = A$ and $(UH)_{I^c} = \mathrm{Id}_{n-k}$.

Let us assume for simplicity that $I = \{1, \dots, k\}$.

$$UH = \begin{pmatrix} A & \mathrm{Id}_{n-k} \end{pmatrix},$$

hence

$$UHe^\intercal = \begin{pmatrix} A & \mathrm{Id}_{n-k} \end{pmatrix} \begin{pmatrix} 0 \\ e_{I^c} \end{pmatrix} = Us.$$

Prange's algorithm

1. Choose an information set $I \subset \{1, \ldots, n\}$ of size k.
2. Find an invertible matrix $U \in \mathbb{F}_2^{n-k \times n-k}$ such that $(UH)_I = A$ and $(UH)_{I^c} = \mathrm{Id}_{n-k}$.
3. If $wt(Us) = t$, then $e_I = 0$ and $e_{I^c} = Us$.

Let us assume for simplicity that $I = \{1, \ldots, k\}$.

$$UH = \begin{pmatrix} A & \mathrm{Id}_{n-k} \end{pmatrix},$$

hence

$$UHe^{\intercal} = \begin{pmatrix} A & \mathrm{Id}_{n-k} \end{pmatrix} \begin{pmatrix} 0 \\ e_{I^c} \end{pmatrix} = Us.$$

From which we get the condition $e_{I^c} = Us$.

The cost of an ISD algorithm is given by the product of

- the cost of one iteration,
- inverted success probability = average number of iterations needed.

The success probability is given by the weight distribution of the error vector.

Example (Success probability of Prange's algorithm)

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

Prange's algorithm

The cost of an ISD algorithm is given by the product of

- the cost of one iteration,
- inverted success probability = average number of iterations needed.

The success probability is given by the weight distribution of the error vector.

Example (Success probability of Prange's algorithm)

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

Prange's algorithm

The cost of an ISD algorithm is given by the product of

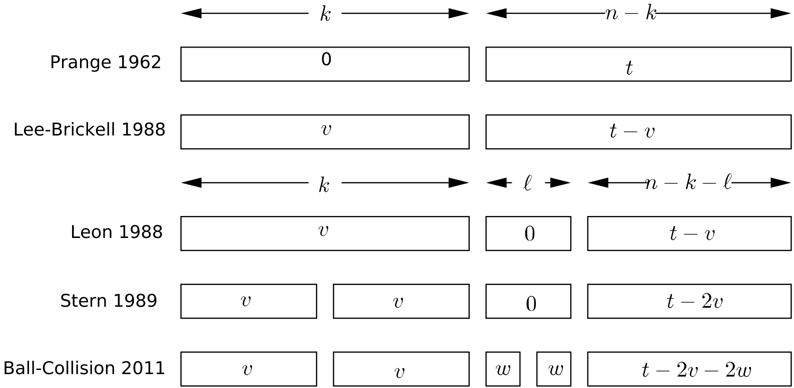
- the cost of one iteration,
- inverted success probability = average number of iterations needed.

The success probability is given by the weight distribution of the error vector.

Example (Success probability of Prange's algorithm)

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

Improvements Overview



Ball-collision Algorithm

Algorithm 1 Ball-collision over $\mathbb{F}_q$

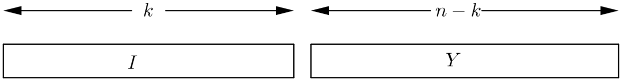
Input: The $(n-k)\times n$ parity check matrix H, the syndrome $s\in\mathbb{F}_q^{n-k}$ and the positive integers $p_1, p_2, q_1, q_2, k_1, k_2, \ell_1, \ell_2 \in \mathbb{Z}$, such that $k = k_1 + k_2$, $p_i \leq k_i$, $q_i \leq \ell_i$ and $t - p_1 - p_2 - q_1 - q_2 \leq n - k - \ell_1 - \ell_2$.
Output: $e \in \mathbb{F}_q^n$ with $He^\top = s$ and $w(e) = t$.

1: Choose an information set $I \subseteq \{1, ..., n\}$ of H of size k.
2: Partition I into two disjoint subsets X_1 and X_2 of size k_1 and $k_2 = k - k_1$ respectively.
3: Partition $Y = \{1, ..., n\}\backslash I$ into disjoint subsets Y_1 of size ℓ_1, Y_2 of size ℓ_2 and Y_3 of size $\ell_3 = n - k - \ell_1 - \ell_2$.
4: Find an invertible matrix $U \in \mathbb{F}_q^{(n-k)\times(n-k)}$, such that $(UH)_Y = \mathrm{Id}_{n-k}$ and $(UH)_I = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}$, where $A_1 \in \mathbb{F}_q^{(\ell_1+\ell_2)\times k}$ and $A_2 \in \mathbb{F}_q^{\ell_3 \times k}$.
5: Compute $Us = \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$, where $s_1 \in \mathbb{F}_q^{\ell_1+\ell_2}$ and $s_2 \in \mathbb{F}_q^{\ell_3}$.
6: Compute $S = \{(A_1(\pi_I(x_1)) + \pi_{Y_1\cup Y_2}(y_1), x_1, y_1) \mid x_1 \in \mathbb{F}_q^n(X_1), wt(x_1) = v_1, y_1 \in \mathbb{F}_q^n(Y_1), wt(y_1) = w_1\}$,
7: Compute $T = \{(-A_1(\pi_I(x_2)) + s_1 - \pi_{Y_1\cup Y_2}(y_2), x_2, y_2) \mid x_2 \in \mathbb{F}_q^n(X_2), wt(x_2) = v_2, y_2 \in \mathbb{F}_q^n(Y_2), wt(y_2) = w_2\}$.
8: **for** $(v, x_1, y_1) \in S$ **do**
9:     **for** $(v, x_2, y_2) \in T$ **do**
10:         **if** $w(-A_2(\pi_I(x_1 + x_2)) + s_2) = t - p_1 - p_2 - q_1 - q_2$ **then**
            Output: $e = x_1 + x_2 + y_1 + y_2 + \sigma_{Y_3}(-A_2(\pi_I(x_1 + x_2)) + s_2)$
11:         **else** go to Step 1 and choose new information set I.

Ball-collision Algorithm

1 Choose an information set I.

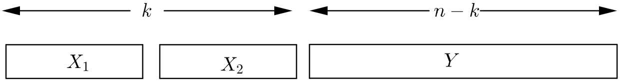
Let us assume for simplicity that $I = \{1, \ldots, k\}$.



Ball-collision Algorithm

1. Choose an information set I.
2. Partition I into X_1 and X_2.

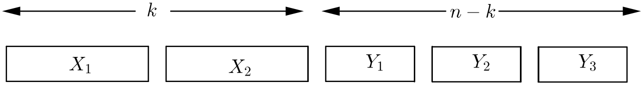
Let us assume for simplicity that $I = \{1, \ldots, k\}$.



Ball-collision Algorithm

1. Choose an information set I.
2. Partition I into X_1 and X_2.
3. Partition Y into Y_1, Y_2, Y_3.

Let us assume for simplicity that $I = \{1, \ldots, k\}$.



Ball-collision Algorithm

1. Choose an information set I.
2. Partition I into X_1 and X_2.
3. Partition Y into Y_1, Y_2, Y_3.
4. Bring H in systematic form.

$$UHe^{\intercal} = \begin{pmatrix} A_1 & \mathrm{Id}_{\ell_1+\ell_2} & 0 \\ A_2 & 0 & \mathrm{Id}_{\ell_3} \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix} = \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} = Us.$$

We get the conditions

$$A_1 e_1 + e_2 = s_1,$$
$$A_2 e_1 + e_3 = s_2.$$

1. Choose an information set I.
2. Partition I into X_1 and X_2.
3. Partition Y into Y_1, Y_2, Y_3.
4. Bring H in systematic form.

$$UHe^\intercal = \begin{pmatrix} A_1 & \mathrm{Id}_{\ell_1+\ell_2} & 0 \\ A_2 & 0 & \mathrm{Id}_{\ell_3} \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix} = \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} = Us.$$

We get the conditions

$$A_1 e_1 + e_2 = s_1,$$
$$A_2 e_1 + e_3 = s_2.$$

Ball-collision Algorithm

Conditions:

$$A_1 e_1 + e_2 = s_1,$$
$$A_2 e_1 + e_3 = s_2.$$

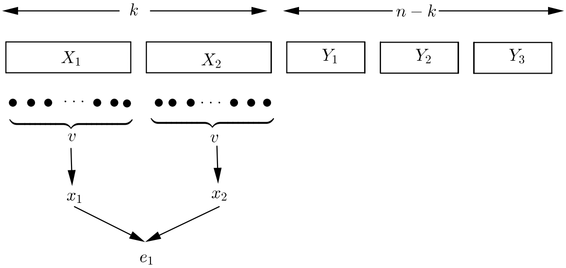
Assumptions:

a e_1 has support in $I = X_1 \cup X_2$ and weight $2v$

b e_2 has support in $Y_1 \cup Y_2$ and weight $2w$

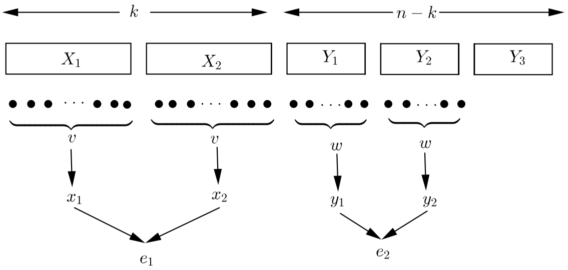
c e_3 has support in Y_3 and weight $t - 2v - 2w$

a e_1 has support in $I = X_1 \cup X_2$ and weight $2v$



a e_1 has support in $I = X_1 \cup X_2$ and weight $2v$

b e_2 has support in $Y_1 \cup Y_2$ and weight $2w$



Ball-collision Algorithm

$$A_1 e_1 + e_2 = s_1, \tag{1}$$
$$A_2 e_1 + e_3 = s_2. \tag{2}$$

For condition (1):
go through all choices of e_1 and e_2 and check with collision if (1) is satisfied.
For condition (2):
define $e_3 = s_2 - A_2 e_1$ and check if e_3 has weight $t - 2v - 2w$.

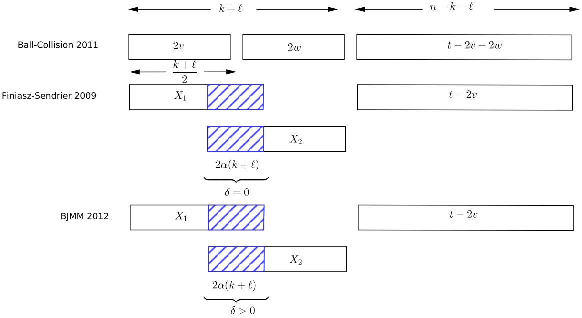
Success probability:

$$\binom{\lfloor k/2 \rfloor}{v}\binom{\lceil k/2 \rceil}{v}\binom{\lfloor \ell/2 \rfloor}{w}\binom{\lceil \ell/2 \rceil}{w}\binom{n-k-\ell}{n-2v-2w}\binom{n}{t}^{-1}.$$

Idea of overlapping sets:

2009 Finiasz and Sendrier: X_1 and X_2 can overlap

2012 Becker, Joux, May and Meurer: can add redundant errors in the overlap



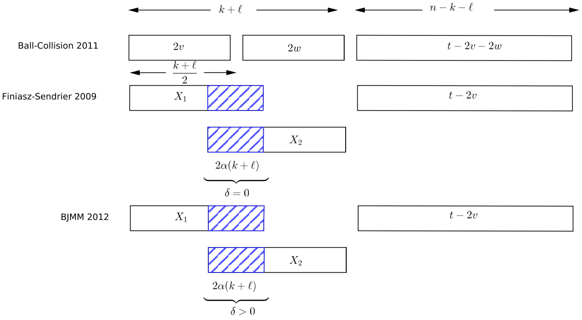
New directions

New parameters:

- α overlap-ratio
- δ amount of redundant errors

2009 Finiasz and Sendrier: $\alpha = 1/2, \delta = 0$

2012 BJMM: $\alpha = 1/2, \delta > 0$



Comparison of Complexities

Let $F(q, R)$ be the exponent of the optimized asymptotic complexity. The asymptotic complexity of half-distance decoding at rate R over $\mathbb{F}_q$ is then given by $q^{F(q,R)n+o(n)}$.

q	$q-$Stern	$q-$Stern-MO	$q-$Ball-collision	$q-$BJMM-MO
2	0.05563	0.05498	0.055573	0.04730
3	0.05217	0.05242	0.052145	0.04427
4	0.04987	0.05032	0.049846	0.04294
5	0.04815	0.04864	0.048140	0.03955
7	0.04571	0.04614	0.045697	0.03706
8	0.04478	0.04519	0.044770	0.03593
11	0.04266	0.04299	0.042656	0.03335

Open questions

- Is partitioning into more sets giving us better asymptotic complexities?
- With new code-based cryptographic schemes, e.g. using rank-metric codes, can we adapt these ideas to these metrics?
- Can we use some structure, e.g. of cyclic codes, to improve the ISD algorithms in these cases?

Thank you!