Is it hard to retrieve an error-correcting pair?

Ruud Pellikaan
and
Irene Márquez-Corbella

Applications of Computer Algebra (ACA-2016)
Computer Algebra in Coding Theory and Cryptography (CACTC-2016)
Kassel, Germany

1 August 2016

Faculteit Wiskunde & Informatica

TU/e Technische Universiteit Eindhoven University of Technology

Content

- Decoding algorithm with an Error-Correcting Pair
- Existence of Error-Correcting Pairs
- Code Based Public Key Cryptosystem
- Possible attacks
- ECP one-way function

Decoding algorithm with an Error-Correcting Pair

Error-correcting codes

C linear block code: $\mathbb{F}_q$-linear subspace of $\mathbb{F}_q^n$

parameters $[n, k, d]$:

n = length
k = dimension of C
d = minimum distance of C

$$d = \min |\{ d(\mathbf{x}, \mathbf{y}) \mid \mathbf{x}, \mathbf{y} \in C, \mathbf{x} \neq \mathbf{y} \}|$$

t = error-correcting capacity of C

$$t = \lfloor \frac{d-1}{2} \rfloor$$

Inner product

Let $\mathbf{a}$ and $\mathbf{b}$ in $\mathbb{F}_q^n$
The standard inner product is defined by

$$\mathbf{a} \cdot \mathbf{b} = a_1 b_1 + \cdots + a_n b_n$$

Two subsets A and B of $\mathbb{F}_q^n$ are perpendicular:
$A \perp B$ if and only if $\mathbf{a} \cdot \mathbf{b} = 0$ for all $\mathbf{a} \in A$ and $\mathbf{b} \in B$

Th dual code of C is defined by

$$C^{\perp} = \{ \mathbf{x} \,|\mathbf{x} \cdot \mathbf{c} = 0 \text{ for all } \mathbf{c} \in C \}$$

Star product

The star product is defined by coordinatewise multiplication:

$$\mathbf{a} * \mathbf{b} = (a_1 b_1, \ldots, a_n b_n)$$

Then

$$(\mathbf{a} * \mathbf{b}) \cdot \mathbf{c} = \sum a_i b_i c_i = (\mathbf{a} * \mathbf{c}) \cdot \mathbf{b}$$

For two subsets A and B of $\mathbb{F}_q^n$

$$A * B = \langle \mathbf{a} * \mathbf{b} \mid \mathbf{a} \in A \text{ and } \mathbf{b} \in B \rangle$$

Efficient decoding algorithms

The following classes of codes:

- Generalized Reed-Solomon codes
- Cyclic codes
- Alternant codes
- Goppa codes
- Algebraic geometry codes

have efficient decoding algorithms:

- Arimoto, Peterson, Gorenstein, Zierler
- Berlekamp, Massey, Sakata
- Justesen et al., Vladut-Skorobogatov, ...
- Error-correcting pairs

Star product and evaluation of polynomials

Notice that multiplying polynomials first and than evaluating gives the same answer as first evaluating and than multiplying
If $f(X), g(X) \in \mathbb{F}_q[X]$ and $h(X) = f(X)g(X)$ then

$$h(a) = f(a)g(a) \text{ for all } a \in \mathbb{F}_q$$

So

$$\mathrm{ev}(f(X)g(X)) = \mathrm{ev}(f(X)) * \mathrm{ev}(g(X)) \text{ and}$$

$$\mathrm{ev}_a(f(X)g(X)) = \mathrm{ev}_a(f(X)) * \mathrm{ev}_a(g(X))$$

Generalized Reed-Solomon codes

Let $\mathbf{a} = (a_1, \ldots, a_n)$ be an n-tuple of distinct elements of $\mathbb{F}_q$
Let $\mathbf{b} = (b_1, \ldots, b_n)$ be an n-tuple of nonzero elements of $\mathbb{F}_q$
Evaluation map:

$$\mathrm{ev}_{\mathbf{a},\mathbf{b}}(f(X)) = (f(a_1)b_1, \ldots, f(a_n)b_n)$$

$$\mathrm{GRS}_k(\mathbf{a}, \mathbf{b}) = \{\, \mathrm{ev}_{\mathbf{a},\mathbf{b}}(f(X)) \mid f(X) \in \mathbb{F}_q[X], \deg(f(X) < k \,\}$$

Parameters:

$$[n, k, n-k+1] \text{ if } k \leq n$$

Since a polynomial of degree $k-1$ has at most $k-1$ zeros

PROPOSITION

$$\mathrm{GRS}_k(\mathbf{a},\mathbf{b}) * \mathrm{GRS}_l(\mathbf{a},\mathbf{c}) = \mathrm{GRS}_{k+l-1}(\mathbf{a},\mathbf{b}*\mathbf{c})$$

and

$$RS_k(n,b) * RS_l(n,c) = RS_{k+l-1}(n,b+c-1) \text{ if } n = q-1$$

Kernel of a received word

Let A and B be linear subspaces of $\mathbb{F}_{q^m}^n$
and $\mathbf{r} \in \mathbb{F}_q^n$ a received word

Define the kernel

$$K(\mathbf{r}) = \{ \mathbf{a} \in A \mid (\mathbf{a} * \mathbf{b}) \cdot \mathbf{r} = 0 \text{ for all } \mathbf{b} \in B \}$$

Let $\mathbf{a}_1, \ldots, \mathbf{a}_l$ and $\mathbf{b}_1, \ldots, \mathbf{b}_m$ be bases of A and B

Then the kernel $K\mathbf{r}$ can be computed by means of the right null space of the $m \times l$ syndrome matrix

$$((\mathbf{b}_i * \mathbf{a}_j) \cdot \mathbf{r} | 1 \leq j \leq l, 1 \leq i \leq m)$$

Kernel for a GRS code

Let $A = \mathrm{GRS}_{t+1}(\mathbf{a}, \mathbf{1})$ and $B = \mathrm{GRS}_t(\mathbf{a}, \mathbf{1})$ and $C = \langle A * B \rangle^{\perp}$

Let
$\mathbf{a}_i = \mathrm{ev}_{\mathbf{a},1}(X^{i-1})$ for $i = 1, \ldots, t+1$
$\mathbf{b}_j = \mathrm{ev}_{\mathbf{a},1}(X^{j})$ for $j = 1, \ldots, t$
$\mathbf{h}_l = \mathrm{ev}_{\mathbf{a},1}(X^{l})$ for $l = 1, \ldots, 2t$

Then
$\mathbf{a}_1, \ldots, \mathbf{a}_{t+1}$ is a basis of A
$\mathbf{b}_1, \ldots, \mathbf{b}_t$ is a basis of B
$\mathbf{h}_1, \ldots, \mathbf{h}_{2t}$ is a basis of $C^{\perp}$

Furthermore

$$\mathbf{a}_i * \mathbf{b}_j = \mathrm{ev}_{\mathbf{a},1}(X^{i+j-1}) = \mathbf{h}_{i+j-1}$$

Matrix of syndromes for a GRS code

Let $\mathbf{r}$ be a received word and its syndrome

$$(s_1, \ldots, s_{2t}) = \mathbf{r}H^T$$

Then

$$(\mathbf{b}_j * \mathbf{a}_i) \cdot \mathbf{r} = s_{i+j-1}$$

To compute the kernel $K(\mathbf{r})$ we have to compute the null space of the matrix of syndromes

$$\begin{pmatrix} s_1 & s_2 & \cdots & s_t & s_{t+1} \\ s_2 & s_3 & \cdots & s_{t+1} & s_{t+2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ s_t & s_{t+1} & \cdots & s_{2t-1} & s_{2t} \end{pmatrix}$$

LEMMA - 1

Let C be an $\mathbb{F}_q$-linear code of length n

Let $\mathbf{r}$ be a received word with error vector $\mathbf{e}$
So $\mathbf{r} = \mathbf{c} + \mathbf{e}$ for some $\mathbf{c} \in C$

If

$$(A * B) \subseteq C^{\perp}$$

then

$$K(\mathbf{r}) = K(\mathbf{e})$$

The space $A(J)$

Let J be a subset of $\{1, \ldots n\}$

We defined before the subspace $A(J)$ of A by

$$A(J) = \{\, \mathbf{a} \in A \mid a_j = 0 \text{ for all } j \in J \,\}$$

LEMMA - 2

Let $\mathbf{e}$ be an error vector of the received word $\mathbf{r}$

If

$$(A * B) \subseteq C^{\perp}$$

and

$$I = \mathrm{supp}(\mathbf{e}) = \{\, i \mid e_i \neq 0 \,\}$$

then

$$A(I) \subseteq K(\mathbf{r})$$

LEMMA - 3

Let **e** be an error vector of the received word **r**

If

$$(A * B) \subseteq C^{\perp}$$

and

$$I = \mathrm{supp}(\mathbf{e}) = \{\, i \mid e_i \neq 0 \,\}$$

and

$$d(B^{\perp}) > \mathrm{wt}(\mathbf{e}) = t$$

then

$$A(I) = K(\mathbf{r})$$

COROLLARY

Let $\mathbf{e}$ be an error vector of the received word $\mathbf{r}$
If

$$(A * B) \subseteq C^{\perp}$$

and

$$I = \mathrm{supp}(\mathbf{e}) = \{ i \mid e_i \neq 0 \}$$

and

$$d(B^{\perp}) > \mathrm{wt}(\mathbf{e}) = t$$

and $\mathbf{a}$ an element of $K(\mathbf{r})$ with

$$J = \{ j \mid a_j = 0 \}$$

Then

$$I \subseteq J$$

Error-locator

Let $I = \mathrm{supp}(\mathbf{e})$ be the set of error positions
The set of zero coordinates of $\mathbf{a} \in A(I)$ contains the set of error positions by the Corollary

The elements of $A(I)$ are called error-locator vectors or functions

The space $A(I)$ is not known to the receiver
The space $K(\mathbf{r})$ can be computed after receiving the word $\mathbf{r}$
The equality $A(I) = K(\mathbf{r})$ implies that
all elements of $K(\mathbf{r})$ are error-locator functions

Error-correcting pair

Let C be a linear code in $\mathbb{F}_q^n$

The pair (A, B) of linear subcodes of $\mathbb{F}_{q^m}^n$ is a called a t-error correcting pair (ECP) over $\mathbb{F}_{q^m}$ for C if

- E.1 $(A * B) \perp C$
- E.2 $k(A) > t$
- E.3 $d(B^{\perp}) > t$
- E.4 $d(A) + d(C) > n$

Basic algorithm

Let $A * B \subseteq C^{\perp}$
Let (A, B) be a t-ECP for C with $d(C) \geq 2t + 1$
Suppose that $\mathbf{c} \in C$ is the code word sent and $\mathbf{r} = \mathbf{c} + \mathbf{e}$ is the received word for some error vector $\mathbf{e}$ with $\mathrm{wt}(\mathbf{e}) \leq t$

The basic algorithm for the code C:

- Compute the kernel $K(\mathbf{r})$
  This kernel is nonzero since $k(A) > t$
- Take a nonzero element $\mathbf{a}$ of $K(\mathbf{r})$
  $K(\mathbf{r}) = K(\mathbf{e})$ since $(A * B) \perp C$
- Determine the set J of zero positions of $\mathbf{a}$
  $\mathrm{supp}(\mathbf{e}) \subseteq J$ since $d(B^{\perp}) > t$
- Compute the error values by erasure decoding
  $|J| < d(C)$ since $n - d(A) < d(C)$

t-ECP corrects t errors efficiently

Theorem

Let C be an $\mathbb{F}_q$-linear code of length n
Let (A, B) be a t-error-correcting pair over $\mathbb{F}_{q^m}$ for C

Then the basic algorithm corrects t errors
for the code C with complexity $\mathcal{O}((mn)^3)$

Existence of Error-Correcting Pairs

t-ECP for $\mathrm{GRS}_{n-2t}(\mathbf{a}, \mathbf{b})$

Let $C^{\perp} = \mathrm{GRS}_{2t}(\mathbf{a}, \mathbf{1})$
Then $C = \mathrm{GRS}_{n-2t}(\mathbf{a}, \mathbf{b})$ for some $\mathbf{b}$
has parameters: $[n, n-2t, 2t+1]$

Let $A = \mathrm{GRS}_{t+1}(\mathbf{a}, \mathbf{1})$ and $B = \mathrm{GRS}_{t}(\mathbf{a}, \mathbf{1})$
Then $(A * B) \subseteq C^{\perp}$

A has parameters $[n, t+1, n-t]$
B has parameters $[n, t, n-t+1]$
So $B^{\perp}$ has parameters $[n, n-t, t+1]$

Hence (A, B) is a t-error-correcting pair for C

Converse is true

Theorem

Let C be an $[n, n-2t, 2t+1]$ code that has a t-ECP
Then C is a GRS code

t-ECP for AG codes

Let C be an algebraic geometry code over $\mathbb{F}_q$ coming from a curve of genus g and of designed minimum distance d^*

Then C has a

$$\lfloor \frac{d^* - 1 - g}{2} \rfloor$$

ECP over $\mathbb{F}_q$
and a

$$\lfloor \frac{d^* - 1}{2} \rfloor$$

ECP over an extension of $\mathbb{F}_q$

Existence of ECPs

Error-Correcting Pairs exist for:

- Generalized Reed-Solomon codes
- Cyclic codes
- Alternant codes
- Goppa codes
- Algebraic geometry codes

Code Based Public Key Cryptosystems

Code based PKC systems - 1

McEliece:
Let $\mathcal{C}$ be a class of codes that have
efficient decoding algorithms correcting t errors with $t \leq (d-1)/2$

Secret key: (S, G, P)

– S an invertible $k \times k$ matrix
– G a $k \times n$ generator matrix of a code C in $\mathcal{C}$.
– P an $n \times n$ permutation matrix

Public key: $G' = SGP$

Code based PKC systems - 2

McEliece:

Encryption with public key $G' = SGP$ and message $\mathbf{m}$ in $\mathbb{F}_q^k$:

$$\mathbf{y} = \mathbf{m}G' + \mathbf{e}$$

with random chosen $\mathbf{e}$ in $\mathbb{F}_q^n$ of weight t

Decryption with secret key (S, G, P):

$$\mathbf{y}P^{-1} = (\mathbf{m}G' + \mathbf{e})P^{-1} = \mathbf{m}SG + \mathbf{e}P^{-1}$$

SG and G are generator matrices of the same code C
$\mathbf{e}P^{-1}$ has weight t
Decoder gives $\mathbf{c} = \mathbf{m}SG$ as closest codeword

Code based PKC systems - 3

Minimum distance decoding is NP-hard
(Berlekamp-McEliece-Van Tilborg)

It is assumed that:

1. $P \neq NP$
2. Decoding up to half the minimum distance is hard
3. One cannot distinguish nor retrieve the original code by disguising it with S and P

Possible attacks

Attacks on code based PKC systems - 1

Generic attack – decoding algorithms:

– McEliece 1978

.....

– Brickell, Lee 1988
– Leon 1988
– van Tilburg 1988
– Stern 1989
– Canteaut, Chabaud, Sendrier 1998
– Finiasz-Sendrier 2009
– Bernstein-Lange-Peters 2008-2011
– Becker-Joux-May-Meurer Eurocrypt 2012

Attacks on code based PKC systems - 3

Structural attacks:

– GRS codes (Sidelnikov-Shestakov)
– subcodes of GRS codes (Wieschebrink, Márquez-Martínez-P)
– Alternant codes: open
– Goppa codes: open
– Algebraic geometry codes: (Faure-Minder, genus $g \leq 2$)
– VSAG codes: (Márquez-Martínez-P-Ruano, arbitrary g)
– Polynomial attack on AG codes and certain subcodes
(Couvreur-Márquez-P, using ECP's)

ECP one-way function

Error-Correcting Pair

Let (A, B) be a pair of $\mathbb{F}_{q^m}$-linear subcodes of $\mathbb{F}_{q^m}^n$
Consider the following conditions

- E.1 $(A * B) \perp C$
- E.2 $k(A) > t$
- E.3 $d(B^{\perp}) > t$
- E.4 $d(A) + d(C) > n$
- E.5 $d(A^{\perp}) > 1$ that means A is a non-degenerated code
- E.6 $d(A) + 2t > n$

If conditions E.2, E.3, E.5 and E.6 hold and $C := \mathbb{F}_q^n \cap (A * B)^{\perp}$
Then $d(C) \geq 2t + 1$ and conditions E.1 and E.4 hold
Therefore (A, B) is a t-ECP for C

The ECP one-way function - 1

Let $\mathcal{P}(n, t, q)$ be the collection of pairs (A, B) such that there exist a positive integer m and a pair (A, B) of $\mathbb{F}_{q^m}$-linear codes of length n that satisfy the conditions E.2, E.3, E.5 and E.6

Let C be the $\mathbb{F}_q$-linear code of length n that is the subfield subcode that has the elements of $A * B$ as parity checks

$$C := \mathbb{F}_q^n \cap (A * B)^{\perp}$$

Then the minimum distance of C is at least $2t + 1$ and (A, B) is a t-ECP for C

The ECP one-way function - 2

Let $\mathcal{F}(n,t,q)$ be the collection of $\mathbb{F}_q$-linear codes of length n and minimum distance $d \geq 2t+1$

Consider the following map

$$\begin{array}{rccc} \varphi_{(n,t,q)}: & \mathcal{P}(n,t,q) & \longrightarrow & \mathcal{F}(n,t,q) \\ & (A,B) & \longmapsto & C \end{array}$$

The question is whether this map is a one-way function

Bilinear equations - 1

One step is finding a pair of codes (A, B) such that $(A * B) \perp C$ for a given code C

Let G be a generator matrix of C with entries g_{ij}, $1 \leq i \leq k,\ 1 \leq j \leq n$
Generator matrix for A with variables X_{ij} with $1 \leq i \leq t+1,\ 1 \leq j \leq n$
and similarly for B by Y_{ij}, $1 \leq i \leq t,\ 1 \leq j \leq n$

Finding a pair of codes (A, B) such that $(A * B) \perp C$
is equivalent to finding a solution of the following system
of $kt(t+1)$ quadratic equations in $n(2t+1)$ variables:

$$\sum_{j=1}^{n} g_{wj} X_{uj} Y_{vj} = 0, \quad \text{for all} \quad 1 \leq u \leq t+1,\ 1 \leq v \leq t,\ 1 \leq w \leq k$$

Bilinear equations - 2

This is a bilinear homogeneous system of equations

Such systems are studied by Faugère et al. using
Groebner bases theory and Buchbergers algorithm
and have improved complexity

Use puncturing/shortening to reduce the number of varaibles

Conclusion

- Decoding algorithm with an Error-Correcting Pair
- Existence of Error-Correcting Pairs
- Code Based Public Key Cryptosystem
- Possible attacks
- ECP one-way function

Thanks
for your attention!