AG codes for Code-based Cryptography

Christiane Peters

Trends in Coding Theory

Ascona – November 1, 2012

joint work with Daniel J. Bernstein and Tanja Lange

AG codes for Code-based Cryptography, or How to End a Bad Reputation

Christiane Peters

Trends in Coding Theory

Ascona – November 1, 2012

joint work with Daniel J. Bernstein and Tanja Lange

“How to end a bad reputation”

“How to end a bad reputation”

- Isn’t such a title a bit too cocky?

“How to end a bad reputation”

- Isn’t such a title a bit too cocky?
- Well, perhaps.

“How to end a bad reputation”

- Isn’t such a title a bit too cocky?

- Well, perhaps.

On the other hand:

- Is it really true that the Janwa–Moreno proposal is broken as claimed in recent papers by Pellikaan et al?

“How to end a bad reputation”

- Isn’t such a title a bit too cocky?

- Well, perhaps.

On the other hand:

- Is it really true that the Janwa–Moreno proposal is broken as claimed in recent papers by Pellikaan et al?

- No it’s not. So let’s be confident and revisit Janwa–Moreno.

"How to end a bad reputation"

- Isn't such a title a bit too cocky?

- Well, perhaps.

On the other hand:

- Is it really true that the Janwa–Moreno proposal is broken as claimed in recent papers by Pellikaan et al?

- No it's not. So let's be confident and revisit Janwa–Moreno.

Disclaimer: this is work in progress.

McEliece encryption

- Given public system parameters n, k, w.
- The public key is a random-looking $k \times n$ matrix G with entries in $\mathbb{F}_q$.
- Encrypt a message $m \in \mathbb{F}_q^k$ as

$$mG + e$$

where $e \in \mathbb{F}_q^n$ is a random error vector of weight w.

Secret key

The public key G has a hidden algebraic structure allowing fast decoding:

$$G = SG'P$$

- G' generates an algebraic code C of length n and dimension k and error-correction capability w,
- S is a random $k \times k$ invertible matrix, and
- P is a random $n \times n$ permutation matrix.

The triple (G', S, P) forms the secret key.

- Choose C so that detecting this structure, i.e., finding G' given G is difficult.

McEliece decryption

The legitimate receiver knows S, G' and P with $G = SG'P$ and an efficient decoding algorithm for the hidden code C.

How to decrypt $y = mG + e$.

1. Compute $yP^{-1} = mSG' + eP^{-1}$.
2. Apply the decoding algorithm of C to find mSG' which is a codeword in C from which one obtains m.

Attacks

There are basically two types of attacks in code-based cryptography.

1. Structural attacks
   - Find the secret code given a public generator matrix.
2. Decrypt a single ciphertext
   - Use a generic decoding algorithm (best known algorithms rely on information-set decoding).

Design goals

Public-key size

- Store redundancy part of a generator matrix in systematic form: $(n-k)k$ bits for an $[n, k]$ code.

Design goals

Public-key size

- Store redundancy part of a generator matrix in systematic form: $(n-k)k$ bits for an $[n,k]$ code.

Thwart structural attacks

- By choosing your hidden code carefully (we'll see more about that later)

Design goals

Public-key size

- Store redundancy part of a generator matrix in systematic form: $(n-k)k$ bits for an $[n, k]$ code.

Thwart structural attacks

- By choosing your hidden code carefully (we'll see more about that later)

Assuming that a structural attack is infeasible

- choose parameters n, k and w so that information-set decoding takes at least 2^b bit operations to correct w errors in one single ciphertext (b-bit security).

Decrease key sizes

Correct more errors while keeping the same code length and the same code dimension.

- The sender, knowing this, can introduce correspondingly more errors;
- the attacker is then faced with a more difficult problem of decoding the additional errors.

Gain

- Decreases the key size at the same security level against information-set decoding.

“Evaluation” of AG codes in code-based crypto

Quote from the conclusion of:

- Marquez-Corbella, Martinez-Moro, Pellikaan: Evaluation of public-key cryptosystems based on algebraic geometry codes (2011).

“Many attempts to replace Goppa codes with different families of codes have been proven to be insecure as for example using GRS codes such as the original Niederreiter system [15] which was broken by Sidelnikov and Shestakov [20] in 1992. Later Janwa and Moreno [9] proposed to use the collection of AG codes on curves for the McEliece cryptosystem. This system was broken for codes on curves of genus $g \leq 2$ by Faure and Minder [5].”

More “evaluation” of AG codes in code-based crypto

- “The security status of this proposal for higher genus was not known. Theorem 12 implies that one should not use VSAG codes for the McEliece PKC system in the range [...]”

From the abstract

- “[...] These two results imply that certain algebraic geometry codes are not secure if used in the McEliece public-key cryptosystem.”

Similar in a more recent conference: Marquez-Corbella, Martinez-Moro, Pellikaan, Ruano (2012): Computational aspects of retrieving a representation of an algebraic geometry code.

- “Indeed, decoding the VSAG representation implies decoding the original code, i.e. breaking the cryptosystem.”

Recap

- Genus 0: Sidelnikov–Shestakov break GRS codes (1992).
- Genus 1: Minder in his PhD thesis (EPFL 2007) breaks genus-1 evaluation codes.
- Genus 2: Minder and Faure break genus-2 evaluation codes.
- Marquez et al break certain higher-genus evaluation codes.

None of these attacks are against subfield subcodes.

Recap

- Genus 0: Sidelnikov–Shestakov break GRS codes (1992).
- Genus 1: Minder in his PhD thesis (EPFL 2007) breaks genus-1 evaluation codes.
- Genus 2: Minder and Faure break genus-2 evaluation codes.
- Marquez et al break certain higher-genus evaluation codes.

None of these attacks are against subfield subcodes.

- We don't use GRS codes in code-based crypto (in the classical setup $G_{\text{pub}} = SG_\Gamma P$; no offense, Mr. Bianchi).
- We use alternant codes, i.e, subfield subcodes.

Subfield subcodes in Code-based Cryptography

McEliece (1978): public-key cryptosystem

- use as public key a hidden algebraic code
- in particular, use a Goppa code,
  i.e., an alternant code = a subfield subcode of a GRS code

Janwa–Moreno (1996):

- "PKS from Subfield Subcodes of Algebraic Geometric Codes"

Subfield subcodes

Definition

Let C be an $[n, k]$ code over $\mathbb{F}_{q^m}$. The subfield subcode $C|_{\mathbb{F}_q}$ of C is the restriction of C to $\mathbb{F}_q$:

$$C|_{\mathbb{F}_q} = \{(c_1, \ldots, c_n) \in C \mid c_i \in \mathbb{F}_q \text{ for } i = 1, \ldots, n\}.$$

Properties of $C|_{\mathbb{F}_q}$

- Minimum distance: $d(C|_{\mathbb{F}_q}) \geq d(C)$.
- Dimension: $k(C|_{\mathbb{F}_q}) \geq n - m(n - k)$.

A family of GRS codes

Let $a_1, \ldots, a_n \in \mathbb{F}_{2^m}$ and g a degree-t element in $\mathbb{F}_{2^m}[x]$ so that $g(a_i) \neq 0$.

- The words $c = (c_1, \ldots, c_n)$ in $\mathbb{F}_{2^m}^n$ with

$$\sum_{i=1}^{n} \frac{c_i}{x - a_i} \equiv 0 \pmod{g(x)}$$

  form a linear code $\Gamma_{2^m}(g) = \Gamma_{2^m}(a_1, \ldots, a_n, g)$ over $\mathbb{F}_{2^m}$ of length n and dimension $n - t$ over $\mathbb{F}_{2^m}$.

Properties of $\Gamma_{2^m}(g)$

- Minimum distance $d(\Gamma_{2^m}(g)) \geq t + 1$.
- Use Berlekamp's algorithm for decoding up to half the minimum distance.

Goppa codes

Definition

The restriction $\Gamma_2(a_1, \ldots, a_n, g)$ of $\Gamma_{2^m}(a_1, \ldots, a_n, g)$ to the field $\mathbb{F}_2$ is called a Goppa code.

Properties of Goppa codes

- $\Gamma_2(a_1, \ldots, a_n, g)$ has length n and dimension $k \geq n - mt$.

If g is squarefree then:

- $\Gamma_2(g) = \Gamma_2(g^2)$,
- minimum distance at least $2t + 1$, and
- Patterson's algorithm efficiently decodes t errors.

AG codes

- Let X be an absolutely irreducible projective non-singular curve over $\mathbb{F}_q$ of genus g,
- $P_1, \ldots, P_n$ distinct rational points on X, and $D = \sum_{i=1}^{n} P_i$,
- G an effective divisor so that *supp* $G \cap$ *supp* $D = \emptyset$.

Definition

The AG code $C_{\mathcal{L}}(D, G)$ associated to D and G is defined as

$$C_{\mathcal{L}}(D, G) = \{(f(P_1), \ldots, f(P_n)) \,|\, f \in \mathcal{L}(G)\} \subseteq \mathbb{F}_q^n$$

where $\mathcal{L}(G)$ denotes the Riemann–Roch space of G.

- Dimension $k = \dim \mathcal{L}(G) - \dim \mathcal{L}(G - D)$.
- Minimum Distance $d \geq n - \deg G$.

Better bounds on the dimension of subfield subcodes

From now on let X be an absolutely irreducible projective non-singular curve over $\mathbb{F}_q$ of genus g for $q = 2^m$.

Theorem (Stichtenoth, 1990)

Let $G \in Div(X)$ so that $G \geq 0$ and $\deg 2G < n$. Consider $C = C_{\mathcal{L}}(2G, D)^{\perp}$. Then

$$\dim(C|_{\mathbb{F}_2}) \geq n - 1 - m\left(\dim \mathcal{L}(2G) - \dim \mathcal{L}(G)\right).$$

Compare to the trivial bound

$$\dim C|_{\mathbb{F}_2} \geq n - m(n - \dim C).$$

Consider $\Gamma_2(g^2)$

Consider the code $\Gamma_2(g^2)$ where $\deg g = t$.

Decoding

- Use Berlekamp's algorithm to correct t errors.

We get higher dimension for $\Gamma_2(g^2)$ than indicated by the trivial bound:

- $\dim \Gamma_2(g^2) = \dim \Gamma_2(g) \geq n - mt$.

Consider $\Gamma_2(g)$

Consider the code $\Gamma_2(g)$ where $\deg g = t$.

Dimension

- Trivial bound $\dim \Gamma_2(g) >= n - mt$.

Decoding

- Use Patterson's algorithm to correct t errors, or
- correct Berlekamp for correcting t errors since $\Gamma_2(g) = \Gamma_2(g^2)$.

Strategy

Imagine that

- we don't know Patterson **and**
- we don't know $\Gamma_2(g) = \Gamma_2(g^2)$.

Strategy

Imagine that

- we don't know Patterson **and**
- we don't know $\Gamma_2(g) = \Gamma_2(g^2)$.

This still leaves one approach with the same results:

- Consider the code $\Gamma_2(g^2)$ where $\deg g = t$.

Decode

- Use Berlekamp's algorithm to decode t errors.

Dimension

- apply Stichtenoth's dimension bound to $\Gamma_2(g^2)$.

Generalize strategy

The strategy can be generalized to AG codes.

- Choose curve X, a divisor $G \geq 0$ so that $G \geq 0$ and $2G < n$.
- Consider $(C_{\mathcal{L}}(2G, D)^{\perp})|_{\mathbb{F}_2}$ for the McEliece cryptosystem.

Decode

- Use your favorite AG decoder for $C_{\mathcal{L}}(2G, D)$.

Dimension

- apply Stichtenoth's dimension bound to get a higher dimension than indicated by the trivial bound.

Proposal

- The best AG decoders seem to want evaluation codes
- Stichtenoth's dimension bound wants duals of evaluation codes.

So start with Hermitian codes of the form $C_{\mathcal{L}}(D, sP_\infty)$ where we can easily write down duals in evaluation form

$$C = C_{\mathcal{L}}(D, sP_\infty)^\perp = C_{\mathcal{L}}(D, (q^3 + q^2 - q - 2 - s)P_\infty).$$

Use McEliece with $C|_{\mathbb{F}_2}$:

- Puncture this code to hide its structure, apply usual defenses (permutations etc).
- thanks to Stichtenoth's bound we have a much better understanding of the code parameters of this subfield subcode.

Ongoing work

- Generalizing to multi-point codes, so that we can use a secret divisor G as in traditional McEliece.
- Speeding up Hermitian decoding algorithms.
- Analyzing the impact of list decoding.
- Parameter optimization.
- Other code families: e.g., asymptotically good codes.

Thank you for your attention!