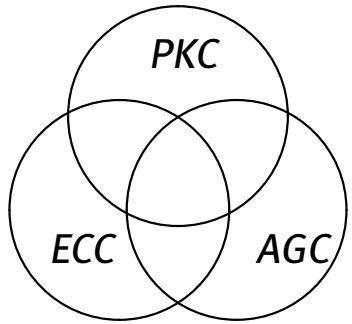
Attacking public-key cryptosystems that use algebraic geometry codes

Ruud Pellikaan
g.r.pellikaan@tue.nl

RISC seminar, CWI Amsterdam,
November 18, 2010

TU/e Technische Universiteit Eindhoven University of Technology

Introduction



- ECC = Error-correcting codes
- AGC = Algebraic geometry curves
- PKC = Public-key cryptosystems

Error-correcting codes

$\mathbb{F}_q$ finite field with q elements
Hamming distance on $\mathbb{F}_q^n$:

$$d(\mathbf{x}, \mathbf{y}) = |\{i \mid x_i \neq y_i\}|$$

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

parameters $[n, k, d]$:
n = length
$k = k(C)$ = dimension of C
$d = d(C)$ = minimum distance of C

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

Generator and parity check matrix

C an $\mathbb{F}_q$-linear code of dimension k

G generator matrix of C: a $k \times n$ matrix with entries in $\mathbb{F}_q$ and

$$C = \{ \mathbf{x}G \mid \mathbf{x} \in \mathbb{F}_q^k \}$$

Code is nondegenerate if G has no zero column

H parity check matrix of C: a $(n-k) \times n$ matrix with entries in $\mathbb{F}_q$ and

$$C = \{ \mathbf{c} \in \mathbb{F}_q^n \mid \mathbf{c}H^T = 0 \}$$

Projective systems and codes

Let $\mathbb{F}$ be a field
Tsfasman-Vladut: A projective system $\mathcal{P} = (P_1, \ldots, P_n)$ in $\mathbb{P}^r(\mathbb{F})$
is an n-tuple of points P_j in this projective space
such that not all these points lie in a hyperplane

Let $P_j = (p_{0j} : p_{1j} : \ldots : p_{rj})$
Let $G_{\mathcal{P}}$ be the $(r+1) \times n$ matrix with $(p_{0j}, p_{1j}, \ldots, p_{rj})^T$ as j-th column

Then $G_{\mathcal{P}}$ has rank $r+1$, since not all points lie in a hyperplane.

If $\mathbb{F}$ is a finite field, then $G_{\mathcal{P}}$ is the generator matrix
of a nondegenerate $[n, r+1, d]$ code over $\mathbb{F}$ where
$n - d$ is the maximal number of points of $\mathcal{P}_G$
that lie in a hyperplane of $\mathbb{P}^{k-1}(\mathbb{F})$

Codes and projective systems

Conversely:
Let G be a generator matrix of a nondegenerate $[n, k, d]$ code over $\mathbb{F}_q$
Then G has no zero columns
Take the columns of G as homogeneous coordinates
of points in $\mathbb{P}^{k-1}(\mathbb{F}_q)$

This gives the projective system $\mathcal{P}_G$ over $\mathbb{F}_q$ of G

One-to-one correspondence between:
generalized equivalence classes of nondegenerate $[n, k]$ codes over $\mathbb{F}_q$
and
equivalence classes of projective systems of n points in $\mathbb{P}^{k-1}(\mathbb{F}_q)$

Generalized Reed-Solomon codes

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

GRS code:
$GRS_k(\mathbf{a}, \mathbf{b}) = \{ (f(a_1)b_1, \ldots, f(a_n)b_n) \mid f(X) \in \mathbb{F}_q[X], \deg(f(X) < k \}$

parameters: $[n, k, n-k+1]$ if $k \leq n$

generator matrix:

$$G_k(\mathbf{a}, \mathbf{b}) = \begin{pmatrix} b_1 & \cdots & b_j & \cdots & b_n \\ a_1 b_1 & \cdots & a_j b_j & \cdots & a_n b_n \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ a_1^{k-1} b_1 & \cdots & a_j^{k-1} b_j & \cdots & a_n^{k-1} b_n \end{pmatrix}$$

Normal rational curves and GRS codes

The projective system of the the code $GRS_k(\mathbf{a}, \mathbf{b})$ with generator matrix $G_k(\mathbf{a}, \mathbf{b})$ is

$$\mathcal{P}_k(\mathbf{a}) = ((1 : a_j : \cdots : a_j^i : \cdots : a_j^{k-1}) \mid j = 1, \ldots, n)$$

Consider the embedding $\mathbb{P}^1 \to \mathbb{P}^r$ by the degree r map given by

$$(y_0 : y_1) \mapsto (y_0^r : y_0^{r-1} y_1 : \cdots : y_0^{r-i} y_1^i : \cdots : y_0 y_1^{r-1} : y_1^r)$$

The image of this map in $\mathbb{P}^r$ is the NRC (*normal rational curve*) $\mathcal{X}_r$

Every hyperplane intersects $\mathcal{X}_r$ in at most r points and

$$\mathcal{P}_k(\mathbf{a}) \subseteq \mathcal{X}_{k-1}(\mathbb{F}_q)$$

Vanishing ideal of rational normal curve

The vanishing ideal $I(\mathcal{X}_r)$ of $\mathcal{X}_r$ is generated by the elements:

$$X_i X_{r-i} - X_j X_{r-j}, \quad \text{for } 0 \leq i < j \leq r$$

that is the determinantal ideal of the 2×2 minors of the $2 \times r$ matrix

$$\begin{pmatrix} X_0 & X_1 & \cdots & X_i & \cdots & X_{r-1} \\ X_1 & X_2 & \cdots & X_{i+1} & \cdots & X_r \end{pmatrix}$$

since the rows of the matrix

$$\begin{pmatrix} 1 & y & \cdots & y^i & \cdots & y^{r-1} \\ y & y^2 & \cdots & y^{i+1} & \cdots & y^r \end{pmatrix}$$

are dependent for all y and ....

Algebraic geometry codes

Let $\mathcal{X}$ be an algebraic variety over $\mathbb{F}_q$
with a subset $\mathcal{P}$ of $\mathcal{X}(\mathbb{F}_q)$ enumerated by $P_1, \ldots, P_n$

Suppose that we have a vector space L over $\mathbb{F}_q$
of functions on $\mathcal{X}$ with values in $\mathbb{F}_q$
So $f(P_i) \in \mathbb{F}_q$ for all i and $f \in L$
In this way we have an evaluation map

$$ev_{\mathcal{P}} : L \longrightarrow \mathbb{F}_q^n$$

defined by $ev_{\mathcal{P}}(f) = (f(P_1), \ldots, f(P_n))$

This evaluation map is linear, so its image is a linear code

Codes on the affine line

The classical example: Generalized Reed-Solomon codes

The geometric object $\mathcal{X}$ is the affine line over $\mathbb{F}_q$
The points are n distinct elements of $\mathbb{F}_q$
L is the vector space of polynomials of degree at most $k-1$
and with coefficients in $\mathbb{F}_q$

This vector space has dimension k
Such polynomials have at most $k-1$ zeros
so nonzero codewords have at least $n-k+1$ nonzeros

This code has parameters $[n, k, n-k+1]$ if $k \leq n$

Codes on curves

Let $\mathcal{X}$ be an algebraic curve over $\mathbb{F}_q$ of genus g
$\mathbb{F}_q(\mathcal{X})$ is the function field of the curve $\mathcal{X}$ with field of constants $\mathbb{F}_q$

Let f be a nonzero rational function on the curve
The divisor of zeros and poles of f is denoted by (f)

Let E be a divisor of $\mathcal{X}$ of degree m. Then

$$L(E) = \{ f \in \mathbb{F}_q(\mathcal{X}) \mid f = 0 \text{ or } (f) \geq -E \}$$

The dimension of the space $L(E)$ is denoted by $l(E)$
Then $l(E) \geq m + 1 - g$ and equality holds if $m > 2g - 2$
by the Theorem of Riemann-Roch

Codes on curves

Let $D = (P_1, \ldots, P_n)$ an n-tuple of mutual distinct points of $\mathcal{X}(\mathbb{F}_q)$
The divisor $P_1 + \cdots + P_n$ will also be denoted by D

If the support of E is disjoint from D, then the evaluation map

$$\mathrm{ev}_D : L(E) \to \mathbb{F}_q^n$$

where $\mathrm{ev}_D(f) = (f(P_1), \ldots, f(P_n))$, is well defined.

The algebraic geometry code $C_L(\mathcal{X}, D, E)$
is the image of $L(E)$ under the evaluation map ev_D
If $m < n$, then $C_L(\mathcal{X}, D, E)$ is an $[n, k, d]$ code with

$$k \geq m + 1 - g \text{ and } d \geq n - m$$

Codes on curves

Embedding of $\mathcal{X}$ in linear system of E of degree m
Let $f_1, f_2, \ldots, f_k$ be a basis of $L(E)$

$$\varphi : \mathcal{X} \longrightarrow \mathbb{P}^{k-1}$$

$$P \mapsto (f_1(P), f_2(P), \ldots, f_k(P))$$

$\mathcal{P} = (\varphi(P_1), \ldots, \varphi(P_n)$ projective system

$$G = \begin{pmatrix} f_1(P_1) & \cdots & f_1(P_j) & \cdots & f_1(P_n) \\ f_2(P_1) & \cdots & f_2(P_j) & \cdots & f_2(P_n) \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ f_k(P_1) & \cdots & f_k(P_j) & \cdots & f_k(P_n) \end{pmatrix}$$ generator matrix

minimum distance $\geq n - m$

Decoding linear codes

Decoding problem

Input: $(G, \mathbf{y})$
where G is a $k \times n$ a matrix G over $\mathbb{F}_q$ of rank k, and $\mathbf{y}$ in $\mathbb{F}_q^n$

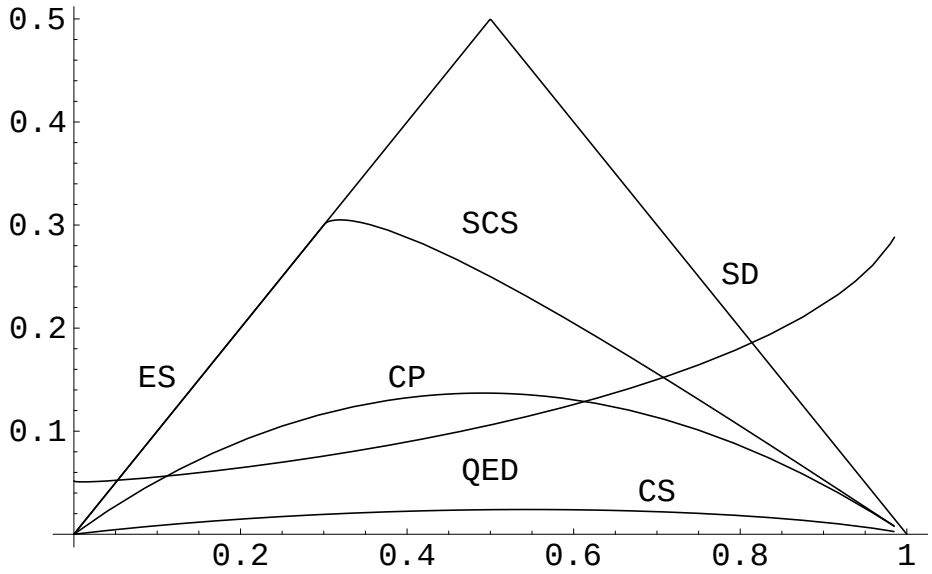
Output: A closest codeword $\mathbf{c}$
so $d(\mathbf{c}, \mathbf{y})$ is minimal for all $\mathbf{c}$ in the code C with generator matrix G

This problem is NP-hard
Berlekamp-McEliece-Van Tilborg

Decoding up to half the minimum distance

Decoding arbitrary linear codes

Exponential complexity $\approx q^{e(R)n}$



x-axis: information rate $R = k/n$

y-axis: complexity exponent $e(R)$

Decoding special classes of codes

Efficient decoding algorithms up to half the minimum distance for:

– Generalized Reed-Solomon codes
– Goppa codes
– Algebraic geometry codes

Polynomial complexity $\mathcal{O}(n^3)$

– Peterson, Arimoto 1960
– Berlekamp-Massey 1963
– Justesen-Larsen-Havemose-Jensen-Hoeholdt 1989
– Skorobogatov-Vladut 1990
– Sakata 1990
– Feng-Rao, Duursma 1993
– Sudan, Guruswami 1997

Public-key cryptosystems

PKC systems use trapdoor one-way functions

by mathematical problems that are (supposedly) hard

RSA, factoring integers: given $n = pq$ find (p, q)
Diffie-Hellman, discrete-log problem in $\mathbb{F}_q$: given $b = a^n$ find n
Elliptic curve PKC, addition on elliptic curve: given $Q = nP$, find n

Code based PKC systems: decoding of codes

McEliece (Goppa codes)
Niederreiter (Generalized Reed-Solomon codes)
Janwa-Moreno (Algebraic geometry codes)

Code based PKC systems

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: $G' = SGP$

Message: $\mathbf{m}$ in $\mathbb{F}_q^k$
Encryption: $\mathbf{y} = \mathbf{m}G' + \mathbf{e}$ with random chosen $\mathbf{e}$ in $\mathbb{F}_q^n$ of weight t
Decryption: $\mathbf{y}P^{-1} = \mathbf{m}SG + \mathbf{e}P^{-1}$ and $\mathbf{e}P^{-1}$ has weight t
Decoder gives $\mathbf{c} = \mathbf{m}SG$ as closest codeword

McEliece PKC system with Goppa codes

Binary Goppa codes with parameters $[n, k, d]$ where

$$n = 2^m, \quad k \geq 2^m - mr, \quad d \geq 2r + 1$$

Are subfield subcodes of GRS codes over $\mathbb{F}_{2^m}$ with parameters

$$[2^m, 2^m - mr, r + 1]$$

Testcase: binary Goppa code with $m = 10$, $r = 50$

$$[1024, 524, 101]$$

Corrects 50 errors

Attacking McEliece PKC system

– McEliece 1978
– Brickell-Lee 1988
– Leon 1988
– van Tilburg 1988
– Stern 1989
– Canteaut-Chabaud-Sendrier 1998
– Bernstein-Lange-Peters 2008

McEliece PKC system using binary Goppa [1024, 524, 101] can be broken in:

– 1400 days by a single CPU or
– 7 days by a cluster of 200 CPU's

Attack by Sidelnikov-Shestakov

Suppose $\mathcal{C}$ is the class of Generalized Reed-Solomon codes

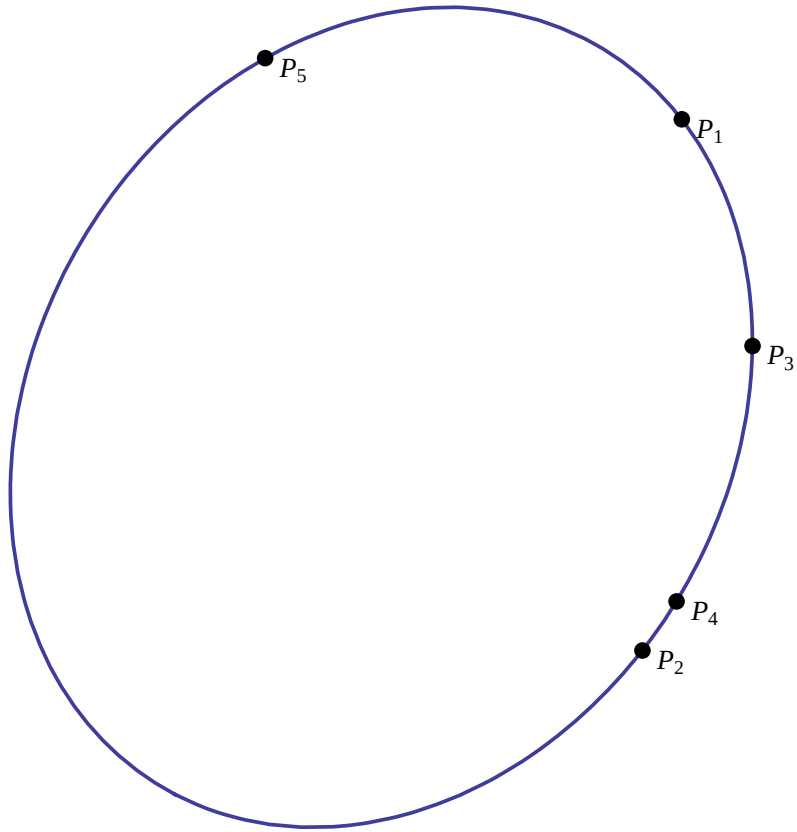
A GRS code of length n and dimension $k = r + 1$
gives a projective system of n points in general position
on a NRC of degree r in projective space of dimension r

Special case: $k = 3$ and $r = 2$:
a NRC of degree 2 in the projective plane is a conic
5 points in general position determine this conic

Steiner: parametrization of this conic in the plane given these 5 points

Algorithm of Sidelnikov-Shestakov for arbitrary k
Complexity: linear algebra $\mathcal{O}(n^3)$

Conic determined by 5 points



NRC of degree r determined by $r + 3$ points

Veronese 1882, Bordiga 1885, Castelnuovo 1885:

Let $\mathcal{P}$ be a collection of $r + 3$ points in general position in $\mathbb{P}^r$
Then there is a unique NRC of degree r passing through the points of $\mathcal{P}$

Twisted cubic, r=3:

(spiraal)

PKC system using subcodes of GRS codes

Berger-Loidreau: use subcodes of GRS codes

Goppa codes are specific subfield subcodes of GRS codes

Projection of NRC to linear subspace

Singular rational curve in projective space

Attack: Wieschebrink

Interpolation problem

Rational plane curves

Example: $k = 3$, rational plane curve of degree r
Suppose that the $P_j = (x_j : y_j : z_j)$ are not lying on the line $Z = 0$.
So we may suppose that $z_j = 1$ for all j.

$$G = \begin{pmatrix} x_1 & \cdots & x_j & \cdots & x_n \\ y_1 & \cdots & y_j & \cdots & y_n \\ 1 & \cdots & 1 & \cdots & 1 \end{pmatrix}$$

Interpolation problem
Input: n mutually distinct points $P_i = (x_i, y_i)$, $i = 1, \ldots, n$
Output: an irreducible rational plane curve of degree at most r
going through the n given points

Bézout: If $n > r^2$, then this curve is unique
Remark: Curve is rational iff parameterizable

Projection of twisted cubic to the plane

(spiraal)

Plane curves

An (irreducible) projective plane curve of degree r has genus

$$g = \frac{1}{2}(r^2 - 3r + 2) - \sum_P \delta_P$$

Variety of projective plane curve of degree r has dimension

$$\frac{1}{2}(r^2 + 3r)$$

Expected dimension of variety of
rational projective plane curves of degree r

$$\frac{1}{2}(r^2 + 3r) - \frac{1}{2}(r^2 - 3r + 2) = 3r - 1$$

Rational plane curves

Kontsevich-Manin 1994:
The variety of rational projective plane curves of degree r
has dimension $3r-1$

Given $3r-1$ points in the projective plane in general position,
Then there are N_r rational projective plane curves of degree r
going through these points

$$N_r = \sum_{r_1+r_2=r} N_{r_1} N_{r_2} \left[r_1^2 r_2^2 \binom{3r-4}{3r_1-2} - r_1^3 r_2 \binom{3r-4}{3r_1-1} \right].$$

$N_1 = 1$: there is a 1 line going through 2 given points
$N_2 = 1$: there is a 1 conic going through 5 given points (no 4 on a line)
$N_3 = 12$: there are 12 rational cubics going through 8 points
in general position

Rational interpolation problem in the plane

Input: n mutually distinct points $P_i = (x_i, y_i)$, $i = 1, \ldots, n$
Output: polynomials f, g and h of degrees at most r such that f, g and h are relatively prime and $h \neq 0$ and there are $t_1, \ldots, t_n$ with

$$x_i = \frac{f(t_i)}{h(t_i)} \text{ and } y_i = \frac{g(t_i)}{h(t_i)} \text{ for all } i = 1, \ldots, n.$$

System of $2n$ equations in $n + 3(r + 1)$ variables $t_1, \ldots, t_n$ and the coefficients of f, g and h

Notice: the points $t_1, \ldots, t_n$ are not given as input as in the Lagrange interpolation problem

PKC system using algebraic geometry codes

Janwa-Moreno: use algebraic geometry codes

Problem for the attacker:
Input: a generator matrix of an AG code $C_L(\mathcal{X}, D, E)$
Output: $(\mathcal{X}, D, E)$ if this triple is unique or $(\mathcal{X}', D', E')$ otherwise

Every code C has a representation as $C = C_L(\mathcal{X}, D, E)$

Automorphisms of the curve $\mathcal{X}$ that leave the divisors D and E invariant give the same code $C_L(\mathcal{X}, D, E)$ but permute the order of the points in D

Fractional transformations of projective line: fix 0, 1 and ∞

Retrieving E from $C_L(\mathcal{X}, D, E)$ given $(\mathcal{X}, D)$

Let $\mathcal{X}$ be a curve of genus g and $D = P_1 + \cdots + P_n$

Let E and E' be divisors of degree m with $2g - 1 < m < n - 1$

Then

$$C_L(\mathcal{X}, D, E) = C_L(\mathcal{X}, D, E')$$

if and only if

$$E' \equiv_D E \text{ that is } E' = E + (f) \text{ and } f(P_j) = 1 \text{ for all } j$$

Curves defined by quadrics

Normal rational normal curve is defined by quadratic equations.

The canonical model of a non-hyperelliptic projective curve
of genus at least three is the intersection of quadrics and cubics,
and of quadrics only except in case of
a trigonal curve and a plane quintic
Enriques 1919, Petri 1923 and Babbage 1939

This result for the canonical divisor was generalized for
arbitrary divisors E under certain constraints on the degree
Mumford 1970, Saint-Donat 1972 and Arbarello 1978

Retrieving $(\mathcal{X}, D, E)$ from $C_L(\mathcal{X}, D, E)$

Let $\mathcal{X}$ be a curve embedded in projective r-space of degree m
Let $I(\mathcal{X})$ be the vanishing ideal of $\mathcal{X}$
Let $\mathcal{P}$ be a subset of $\mathcal{X}$ of n points
Then

$$I(\mathcal{X}) \subseteq I(\mathcal{P})$$

Let $I_2(\mathcal{X})$ be the ideal generated by
the homogeneous elements of degree two in $I(\mathcal{X})$

Suppose $I_2(\mathcal{X}) = I(\mathcal{X})$

$$\text{If } n > 2m, \text{ then } I(\mathcal{X}) = I_2(\mathcal{P})$$

By Bézout

Questions?