# Code Based Cryptology at TU/e

Ruud Pellikaan
g.r.pellikaan@tue.nl

University Gadjah Mada
Yogyakarta
9 November 2015



Technische Universiteit
Eindhoven
University of Technology

# Content



1. Ambassador of TU/e
2. Introduction on Coding, Crypto and Security
3. Public-key crypto systems
4. One-way functions and
5. Code based public-key crypto system
6. Error-correcting codes
7. Error-correcting pairs

# Coding



- correct transmission of data
- error-correction
- no secrecy involved
- communication: internet, telephone, ...
- fault tolerant computing
- memory: computer compact disc, DVD, USB stick ...

# Crypto



- private transmission of data
- secrecy involved
- privacy
- eaves dropping
- insert false messages
- authentication
- electronic signature
- identity fraud

# Security



- secure transmission of data
- secrecy involved
- electronic voting
- electronic commerce
- money transfer
- databases of patients

# Public-key cryptography (PKC)



- Diffie and Hellman 1976 in the public domain in
- Ellis in 1970 for secret service, not made public until 1997
- advantage with respect to symmetric-key cryptography
- no exchange of secret key between sender and receiver

# One-way function



- At the heart of any public-key cryptosystem is a
- one-way function
- a function $y = f(x)$ that is
- easy to evaluate but
- for which it is computationally infeasible **(one hopes)**
- to find the inverse $x = f^{-1}(y)$

# Examples of one-way function



- Example 1
- differentiation a function is easy
- integrating a function is difficult

- Example 2
- checking whether a given proof is correct is easy
- finding the proof of a proposition is difficult

# Integer factorization



- $x = (p, q)$ is a pair of distinct prime numbers
- $y = pq$ is its product
- proposed by Cocks in 1973 in secret service
- Rivest-Shamir-Adleman (RSA) in 1978 in public domain
- based on the hardness of factorizing integers

# Discrete logarithm

- $G$ is a group (written multiplicatively)
- with $a \in G$ and $x$ an integer
- $y = a^x$
- Diffie-Hellman in 1974 and 1976 in public domain
- proposed by Williamson in 1974 in secret service
- based on difficulty of finding discrete logarithms in a finite field

# Elliptic curve discrete logarithm



- $G$ is an elliptic curve group (written additively) over a finite field
- $P$ is a point on the curve
- $x = k$ a positive integer $k$
- $y = kP$ is another point on the curve
- obtained by the multiplication of $P$ with a positive integer $k$
- proposed by Koblitz and Miller in 1985
- based on the difficulty of inverting this function in $G$

# Code based cryptography



- $H$ is a given $r \times n$ matrix with entries in $\mathbb{F}_q$
- $\mathbf{x}$ is in $\mathbb{F}_q^n$ of weight at most $t$
- $\mathbf{y} = \mathbf{x}H^T$
- proposed by McEliece in 1978 and later by Niederreiter
- based on the difficulty of decoding error-correcting codes
- it is NP complete

# NP complete problems



- NP = nondeterministic polynomial time
- given a problem with yes/no answer
- if answer is yes and the solution is given
- then one can check it in polynomial time

- Input: integer n
- Query: can one factorize $n$ in $n = pq$ with $p$ and $q > 1$?
- if answer is yes and someone gives $p$ and $q$
- then one easily checks that $n = pq$
- otherwise it is difficult to find $p$ and $q$

# Abstract



- error-correcting codes
- error-correcting pairs correct errors efficiently
- applies to many known codes
- prime example Generalized Reed-Solomon codes
- can be explained in a short time
- is a distinguisher of certain classes of codes
- McEliece public-key cryptosystem
- polynomial attack if algebraic geometry codes are used
- ECP map is a one-way function

$Q$ alphabet of $q$ elements
Hamming distance between $\mathbf{x}$, $\mathbf{y}$ in $Q^n$

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

$C$ block code is a subset of $Q^n$

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

minimum distance of $C$

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

error-correcting capacity of $C$

# Linear codes



$\mathbb{F}_q$ the finite field with $q$ elements, $q = p^e$ and $p$ prime
$\mathbb{F}_q^n$ is an $\mathbb{F}_q$-linear vector space of dimension $n$

$C$ linear code is an $\mathbb{F}_q$-linear subspace of $\mathbb{F}_q^n$

parameters $[n, k, d]_q$

$q$ = size finite field
$n$ = length of $C$
$k$ = dimension of $C$
$d$ = minimum distance of $C$

Singleton bound $d \leq n - k + 1$
Maximum Distance Separable (MDS) if $d = n - k + 1$

# Inner product and dual space



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 $\mathbf{b} \in B$

Let $C$ be a subspace of $\mathbb{F}_q^n$
The dual space is defined by

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

If $C$ has dimension $k$, then $C^{\perp}$ has dimension $n - k$

# Star product



The star product is defined by coordinatewise multiplication

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

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-Skrobogatov, ...........
- Error-correcting pairs

# 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 $d(A) + d(C) > n$

# Generalized Reed-Solomon codes



Let $\mathbf{a} = (a_1, \ldots, a_n)$ be an $n$-tuple of mutually 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:

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

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

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

$$\mathrm{ev}_{\mathbf{a},\mathbf{b}}(f(X)) * \mathrm{ev}_{\mathbf{a},\mathbf{c}}(g(X)) = \mathrm{ev}_{\mathbf{a},\mathbf{b}*\mathbf{c}}(f(X)g(X))$$

$$GRS_k(\mathbf{a}, \mathbf{b}) * GRS_l(\mathbf{a}, \mathbf{c}) = GRS_{k+l-1}(\mathbf{a}, \mathbf{b} * \mathbf{c})$$

# $t$-ECP for $GRS_{n-2t}(\mathbf{a}, \mathbf{b})$



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

Let $A = GRS_{t+1}(\mathbf{a}, \mathbf{1})$ and $B = GRS_{t}(\mathbf{a}, \mathbf{1})$
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 $C$

# 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

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

Lemma
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

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

# Kernel for a GRS code



Let $A = GRS_{t+1}(\mathbf{a}, \mathbf{1})$ and $B = GRS_t(\mathbf{a}, \mathbf{1})$ and $C = \langle A * B \rangle^\perp$

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

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

# Matrix of syndromes for a GRS code



Let r be a received word and
$(s_1, \ldots, s_{2t}) = \mathrm{r}H^T$ its syndrome
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

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

Let $(A, B)$ be a $t$-ECP for $C$

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

Define the subspace of $A$ of error-locating vectors:

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

Lemma

Let $(A * B) \perp C$

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

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

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

**Lemma**

Let $(A * B) \perp C$

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

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

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

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

If $\mathbf{a}$ is a nonzero element of $K(\mathbf{r})$

$J$ zero positions of $\mathbf{a}$

Then

$$I \subseteq J$$

# Basic algorithm



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 $\mathcal{O}((mn)^3)$

# 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 by $S$ and $P$

# 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 - 2



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: (Couvreur-Márquez-P, using ECP's)

# Codes with $t$-ECP



$\mathscr{P}(n, t, q)$ is the collection of pairs $(A, B)$ that satisfy

$$\begin{array}{ll} E.2 & k(A) > t \\ E.3 & d(B^{\perp}) > t \\ E.5 & d(A^{\perp}) > 1 \\ E.6 & d(A) + 2t > n \end{array}$$

Let

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

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

# ECP one-way function



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

Consider the following map

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

Question:
Is this a one-way function?

# Conclusion



- Many known classes of codes
- that have decoding algorithm correcting $t$-errors
- have a $t$-ECP
- and are not suitable for a code based PKC

Question for future research
Is the ECP map a one-way function?

# Thank you for your attention!