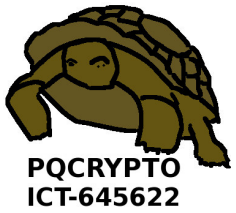
Code-based Cryptography

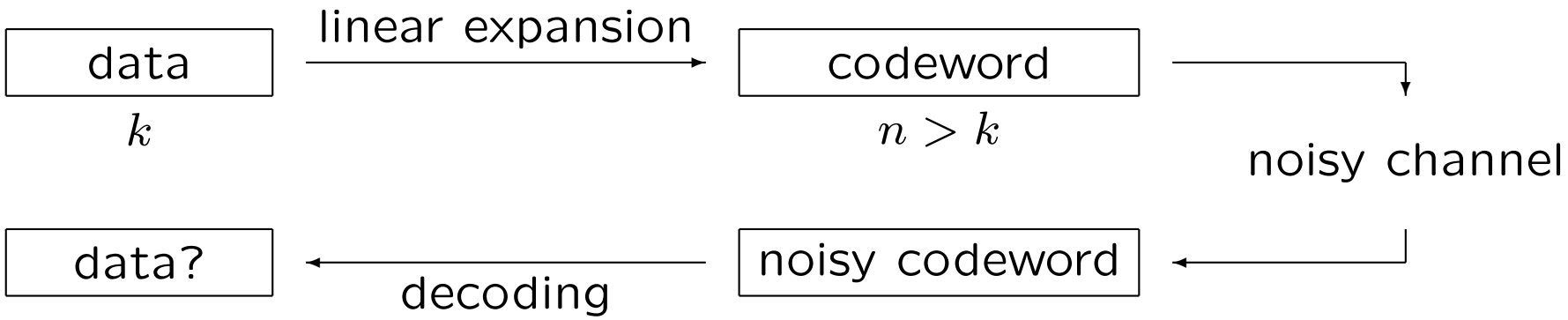
ECRYPT-CSA Executive School on Post-Quantum Cryptography

2017
TU Eindhoven

Nicolas Sendrier



Linear Codes for Telecommunication



[Shannon, 1948] (for a binary symmetric channel of error rate p):
Decoding probability $\longrightarrow 1$ if $\frac{k}{n} = R < 1 - h(p)$

($h(p) = -p \log_2 p - (1-p) \log_2(1-p)$ the binary entropy function)

Codes of rate R can correct up to λn errors ($\lambda = h^{-1}(1-R)$)

For instance 11% of errors for $R = 0.5$

Non constructive $\longrightarrow$ no poly-time algorithm for decoding in general

Random Codes Are Hard to Decode

When the linear expansion is random:

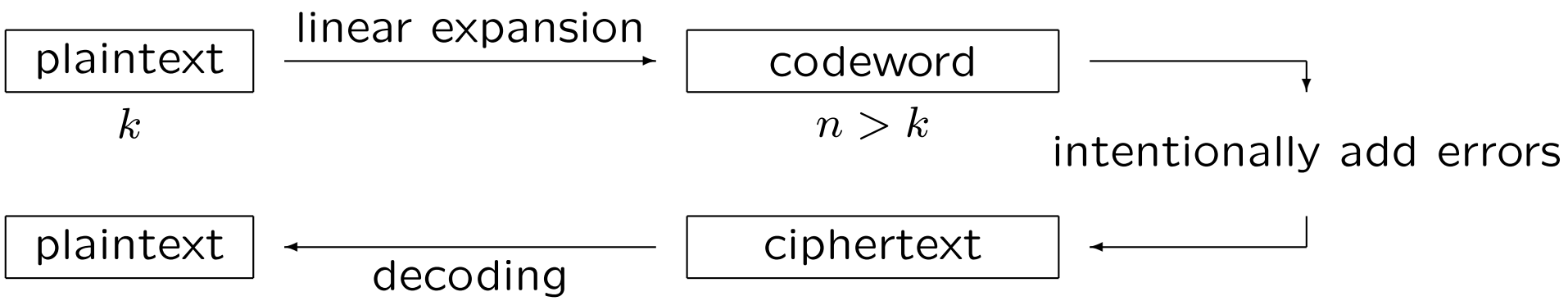
- Decoding is NP-complete [Berlekamp, McEliece & van Tilborg, 78]
- Even the tiniest amount of error is (believed to be) hard to remove. Decoding n^{ε} errors is conjectured difficult on average for any $\varepsilon > 0$ [Alekhnovich, 2003].

Codes with Good Decoders Exist

Coding theory is about finding "good" codes (i.e. linear expansions)

- alternant codes have a poly-time decoder for $\Theta\left(\frac{n}{\log n}\right)$ errors
- some classes of codes have a poly-time decoder for $\Theta(n)$ errors (algebraic geometry, expander graphs, concatenation, . . .)

Linear Codes for Cryptography



- If a random linear code is used, no one can decode efficiently
- If a "good" code is used, anyone who knows the structure has access to a fast decoder

Assuming that the knowledge of the linear expansion does not reveal the code structure:

- The linear expansion is public and anyone can encrypt
- The decoder is known to the legitimate user who can decrypt
- For anyone else, the code looks random

Why Consider Code-Based Cryptography?

Because

- some nice features
  - efficient (algorithmic coding theory)
  - secure (relies on well studied algorithmic problems)

- cryptography needs diversity
  - quantum computing
  - algorithmic progress

Outline

I. Introduction to Codes and Code-based Cryptography

II. Instantiating McEliece

III. Security Reduction to Difficult Problems

IV. Practical Security - The Attacks

V. Other Public Key Systems

I. Introduction to Codes and Code-based Cryptography

Linear Error Correcting Codes

$\mathbf{F}_q$ the finite field with q elements

Hamming weight: $x = (x_1, \ldots, x_n) \in \mathbf{F}_q^n$,

$$|x| = |\{i \in \{1, \ldots, n\} \mid x_i \neq 0\}|$$

A generator matrix $G \in \mathbf{F}_q^{k \times n}$ of $\mathcal{C}$ is such that $\mathcal{C} = \left\{xG \mid x \in \mathbf{F}_q^k\right\}$

A parity check matrix $H \in \mathbf{F}_q^{r \times n}$ of $\mathcal{C}$ is such that $\mathcal{C} = \left\{x \in \mathbf{F}_q^n \mid xH^T = 0\right\}$

t-bounded decoder

$$\text{for all } x \in \mathcal{C} \text{ and all } e \in \mathbf{F}_q^n,\ |e| \leq t \Rightarrow \Phi_{\mathcal{C}}(x + e) = x$$

t-bounded H-syndrome decoder

$$\text{for all } e \in \mathbf{F}_q^n,\ |e| \leq t \Rightarrow \Psi_H(eH^T) = e$$

McEliece Public-key Encryption Scheme – Overview

Let $\mathcal{F}$ be a family of t-error correcting q-ary linear $[n, k]$ codes

Key generation:

pick $\mathcal{C} \in \mathcal{F} \rightarrow$ $\begin{cases} \textbf{Public Key:} & G \in \mathbf{F}_q^{k \times n}, \text{ a generator matrix} \\ \textbf{Secret Key:} & \Phi : \mathbf{F}_q^n \rightarrow \mathcal{C}, \text{ a } t\text{-bounded decoder} \end{cases}$

Encryption: $\left[\begin{array}{rccc} E_G : & \mathbf{F}_q^k & \rightarrow & \mathbf{F}_q^n \\ & x & \mapsto & xG + e \end{array} \right]$ with e random of weight t

Decryption: $\left[\begin{array}{rccc} D_\Phi : & \mathbf{F}_q^n & \rightarrow & \mathbf{F}_q^k \\ & y & \mapsto & \Phi(y)G^* \end{array} \right]$ where $GG^* = 1$

Proof: $D_\Phi(E_G(x)) = D_\Phi(xG + e) = \Phi(xG + e)G^* = xGG^* = x$

Niederreiter Public-key Encryption Scheme – Overview

Let $\mathcal{F}$ be a family of t-error correcting q-ary $[n,k]$ codes, $r = n - k$

Let $\mathcal{S}_n(\mathbf{0}, t) = \{e \in \mathbf{F}_q^n \mid |e| = t\}$

Key generation: pick $\mathcal{C} \in \mathcal{F}$

$\rightarrow$ { **Public Key:** $H \in \mathbf{F}_q^{r \times n}$, a parity check matrix
Secret Key: $\Psi : \mathbf{F}_q^r \rightarrow \mathbf{F}_q^n$, a t-bounded H-syndrome decoder

Encryption:
$$\left[\begin{array}{rccc} E_H : & \mathcal{S}_n(\mathbf{0}, t) & \rightarrow & \mathbf{F}_q^r \\ & e & \mapsto & eH^T \end{array}\right]$$

Decryption:
$$\left[\begin{array}{rccc} D_\Psi : & \mathbf{F}_q^r & \rightarrow & \mathcal{S}_n(\mathbf{0}, t) \\ & s & \mapsto & \Psi(s) \end{array}\right]$$

Proof: $D_\Psi(E_H(e)) = D_\Psi(eH^T) = e$

McEliece/Niederreiter Security

The following two problems must be difficult enough:

1. Retrieve an efficient t-bounded decoder from the public key (*i.e.* a generator matrix or a parity check matrix)

   The legitimate user must be able to decode thus some structure exists, it must remain hidden to the adversary

2. Decode t errors in a random q-ary $[n, k]$ code

   Without knowledge of the trapdoor the adversary is reduced to use generic decoding techniques

   The parameters n, k and t must be chosen large enough

In Practice

[McEliece, 1978]
"A public-key cryptosystem based on algebraic coding theory"

The secret code family consisted of irreducible binary Goppa codes of length 1024, dimension 524, and correcting up to 50 errors

- public key size: 536 576 bits
- cleartext size: 524 bits
- ciphertext size: 1024 bits

A bit undersized today (attacked in [Bernstein, Lange, & Peters, 08] with $\approx 2^{60}$ CPU cycles)

[Niederreiter, 1986]
"Knapsack-type cryptosystems and algebraic coding theory"

Several families of secret codes were proposed, among them Reed-Solomon codes, concatenated codes and Goppa codes. Only Goppa codes are secure today.

II. Instantiating McEliece

Which Code Family ?

Finding families of codes whose structure cannot be recognized seems to be a difficult task

Family	Proposed by	Broken by
Goppa	McEliece (78)	-
Reed-Solomon	Niederreiter (86)	Sidelnikov & Chestakov (92)
Concatenated	Niederreiter (86)	Sendrier (98)
Reed-Muller	Sidelnikov (94)	Minder & Shokrollahi (07)
AG codes	Janwa & Moreno (96)	Faure & Minder (08)
		Couvreur, Márquez-Corbella. & Pellikaan (14)
LDPC	Monico, Rosenthal, & Shokrollahi (00)	
Convolutional codes	Löndahl & Johansson (12)	Landais & Tillich (13)

[Faugère, Gauthier, Otmani, Perret, & Tillich, 11] distinguisher for binary Goppa codes of rate $\to 1$

More on Goppa Codes

Goppa codes are not limited to the binary case. It is possible to define q-ary Goppa codes with a support in $\mathbf{F}_{q^m}$.

[Bernstein, Lange, & Peters, 10]: Wild McEliece. The key size can be reduced in some case. There are limits:

- [Couvreur, Otmani, & Tillich, 14] Choose $m > 2$
- [Faugère, Perret, & Portzamparc, 14] Caution if q not prime

Reducing the Public Key Size

In a block-circulant matrix, each (square) block is completely defined by its first row $\rightarrow$ public key size is linear instead of quadratic

$$G = \begin{array}{|c|c|c|} \hline g_{0,0} & g_{0,1} & g_{0,2} \\ \hline g_{1,0} & g_{1,1} & g_{1,2} \\ \hline \end{array}$$

- Quasi-cyclic [Gaborit, 05] or quasi-dyadic [Misoczki & Barreto, 09] alternant (Goppa) codes. Structure + structure must be used with great care [Faugère, Otmani, Perret, & Tillich, 10]

- Disguised QC-LDPC codes [Baldi & Chiaraluce, 07]. New promising trend.

- QC-MDPC [Misoczki, Tillich, Sendrier, & Barreto, 13]. As above with a stronger security reduction.

Some Sets of Parameters for Goppa Codes

	text size in bits					
$(n = 2^m)$	McEliece		Niederreiter		key	message
m, t	cipher	clear	cipher	clear	size	security*
$10, 50$	1024	524	500	284	32 kB	52
$11, 40$	2048	1608	440	280	88 kB	81
$12, 50$	4096	3496	600	385	277 kB	120

* logarithm in base 2 of the cost of the best known attack
lower bound derived from ISD, BJMM variant (generic decoder)

the key security is always higher ($\approx mt$)

key size is given for a key in systematic form

Encryption/Decryption Speed

	sizes		cycles/byte		cycles/block		
m, t	cipher	clear	encrypt	decrypt	encrypt	decrypt	security
$11, 40$	2048	1888	105	800	25K	189K	81
$12, 50$	4096	3881	98	618	47K	300K	120

(Intel Xeon 3.4Ghz, single processor) 100 Kcycle ≈ 30 μs
AES: 10-20 cycles/byte

McBits [Berstein, Chou, & Schwabe] gains a factor ≈ 5 on decoding (bit-sliced field arithmetic + algorithmic innovations for decoding).
Targets key exchange mechanism based on Niederreiter.

Some Sets of Parameters for QC-MDPC-McEliece

Binary QC-MDPC $[n, k]$ code with parity check equations of weight w correcting t errors

	size in bits			security*	
(n, k, w, t)	cipher	clear	key	message	key
$(9602, 4801, 90, 84)$	9602	4801	4801	80	79
$(19714, 9857, 142, 134)$	19714	9857	9857	128	129

* logarithm in base 2 of the cost of the best known attack
lower bound derived from ISD, BJMM variant

The best key attack and the best message attack are both based on generic decoding

III. Security Reduction to Difficult Problems

Hard Decoding Problems

[Berlekamp, McEliece, & van Tilborg, 78]

Syndrome Decoding NP-complete

Instance: $H \in \mathbf{F}_2^{r \times n}$, $s \in \mathbf{F}_2^r$, w integer

Question: Is there $e \in \mathbf{F}_2^n$ such that $|e| \leq w$ and $eH^T = s$?

Computational Syndrome Decoding NP-hard

Instance: $H \in \mathbf{F}_2^{r \times n}$, $s \in \mathbf{F}_2^r$, w integer

Output: $e \in \mathbf{F}_2^n$ such that $|e| \leq w$ and $eH^T = s$

[Finiasz, 04]

Goppa Bounded Decoding NP-hard

Instance: $H \in \mathbf{F}_2^{r \times n}$, $s \in \mathbf{F}_2^r$

Output: $e \in \mathbf{F}_2^n$ such that $|e| \leq \dfrac{r}{\log_2 n}$ and $eH^T = s$

Open problem: average case complexity (Conjectured difficult)

Hard Structural Problems

Goppa code Distinguishing NP

Instance: $G \in \mathbf{F}_2^{k \times n}$

Question: Does G span a binary Goppa code?

- NP: the property is easy to check given (L, g)
- Completeness status is unknown
- Easy when the information rate $\to 1$
  (Faugère, Gauthier, Otmani, Perret, & Tillich, 11)

Goppa code Reconstruction

Instance: $G \in \mathbf{F}_2^{k \times n}$

Output: (L, g) such that $\Gamma(L, g) = \left\{ xG \mid x \in \mathbf{F}_q^k \right\}$

- Tightness: gap between decisional and computational problems

Security Reduction for McEliece

We consider an instance of McEliece using t-error correcting binary $[n, k]$ Goppa codes

Theorem

If there exists a (T, ε)-adversary against McEliece then there exists either

- a $(T, \varepsilon/2)$-decoder for t errors in a random $[n, k]$ codes,
- or a $(T + O(n^2), \varepsilon/2)$-distinguisher for Goppa codes.

This theorem says essentially that if McEliece can be broken then either "Syndrome Decoding" or "Goppa Code Distinguishing" can be efficiently solved.

This assumes that the key pair and the cleartext were chosen uniformly at random

$\rightarrow$ McEliece is an OWE (One Way Encryption) scheme

Malleability Attacks

Create New Ciphertext. folklore

If y is a ciphertext and a is a codeword then $y + a$ is a ciphertext

Not a desirable feature *a priori...*

Resend-message Attack. [Berson, 97]

The same message x is sent twice with the same public key G

$\rightarrow$ the message can be recovered

Reaction Attack. [Kobara & Imai, 00] ??

We assume the decryption system can be used as an oracle and behaves differently when

- its input is at distance $> t$ from the code,
- its input is at distance $\leq t$ from the code.

$\rightarrow$ the oracle can be tranformed into a decoder

Semantically Secure Conversions

Being OWE is a very weak notion of security. In the case of code-based systems, it does not encompass attacks such that the "resend-message attack", the "reaction attack" or, more generally, attacks related to malleability.

Fortunately, using the proper semantically secure conversion any deterministic OWE scheme can become IND-CCA, the strongest security notion.

McEliece is not deterministic but IND-CCA conversion are possible nevertheless, see [Kobara & Imai, 01] for the first one.

An IND-CPA conversion without random oracle also exists [Nojima, Imai, Kobara & Morozov, 08].

IV. Practical Security - The Attacks

Best Known Attacks

Decoding attacks. For the public-key encryption schemes the best attack is always Information Set Decoding (ISD), this will change for other cryptosystems

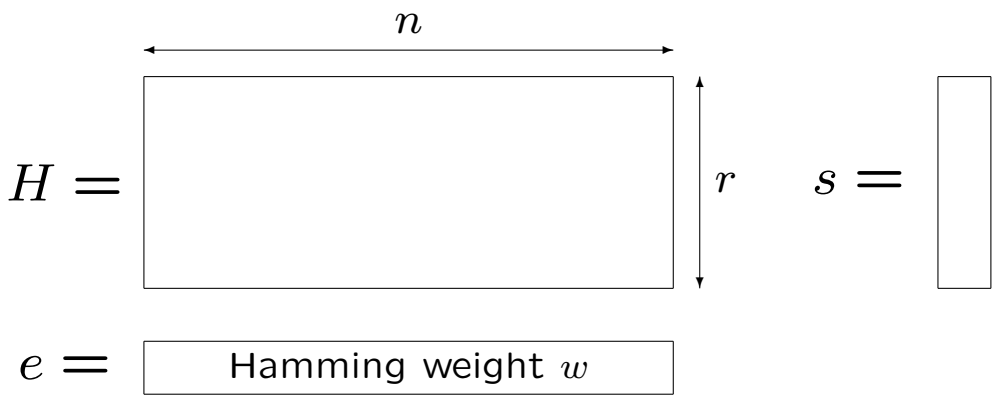
Key attacks. Most proposals using families other than binary Goppa codes have been broken

For binary Goppa codes there are only exhaustive attacks enumerating either generator polynomials either supports (that is permutations)

Syndrome Decoding – Problem Statement

Computational Syndrome Decoding

CSD(n, r, w) Given $H \in \mathbf{F}_2^{r \times n}$ and $s \in \mathbf{F}_2^r$, solve $eH^T = s$ with $\lvert e \rvert \leq w$

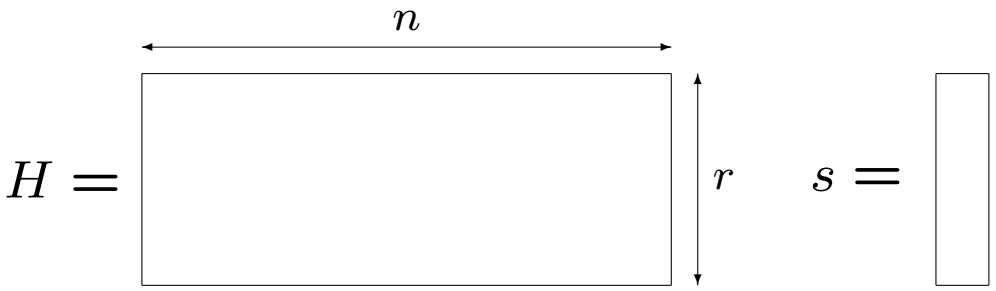


Find w columns of H adding to s

Very close to a subset sum problem

For instance $\begin{cases} n = 2048 \\ r = 352 \\ w = 32 \end{cases}$ $\rightarrow$ computing effort $> 2^{80}$

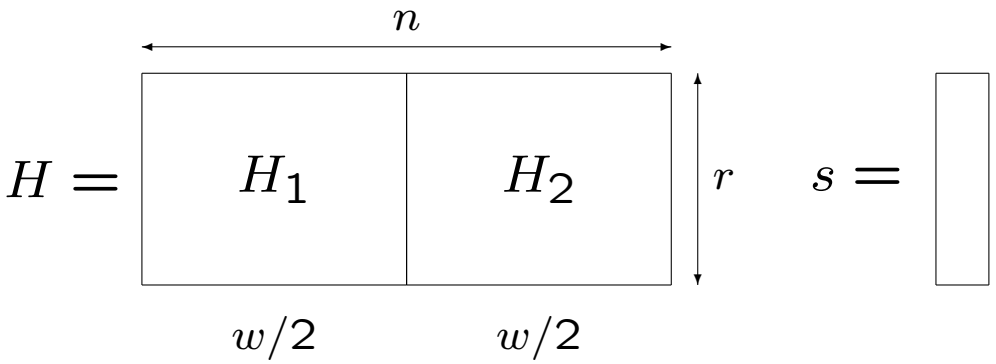
Algorithm 0



Compute every sum of w columns $\rightarrow$ complexity $\binom{n}{w}$ column ops.

1 column operation $\begin{cases} \text{1 read or write} \\ \quad \text{and} \\ \text{1 test} \\ \quad \text{and} \\ \text{1 addition or weight computation} \end{cases}$

Algorithm 1: Birthday Decoding



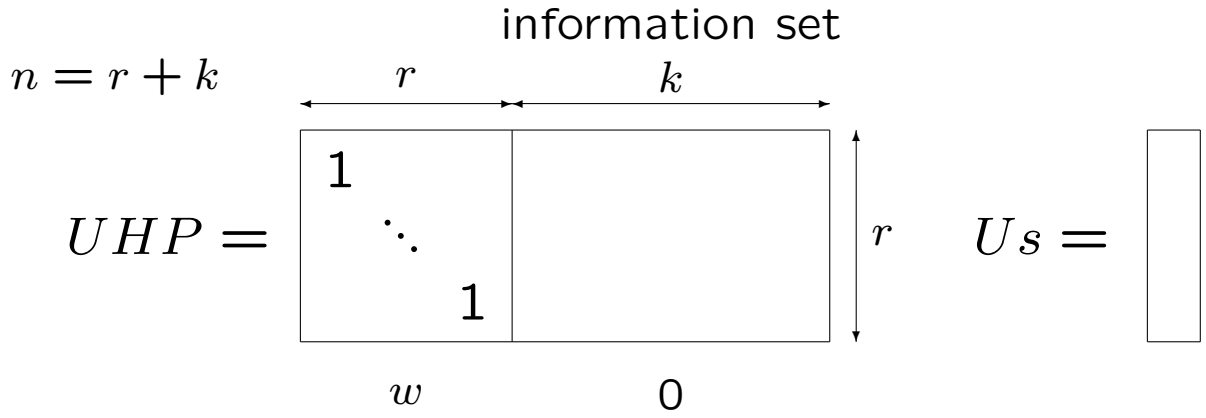
Compute $\{H_1 e \mid |e| = w/2\} \cap \{s + H_2 e \mid |e| = w/2\}$

Complexity $2\binom{n/2}{w/2}$ and non-empty with probability $\dfrac{\binom{n/2}{w/2}^2}{\binom{n}{w}}$

$\rightarrow$ average cost $2\dfrac{\binom{n}{w}}{\binom{n/2}{w/2}} \approx \sqrt[4]{8\pi w}\sqrt{\binom{n}{w}}$

Algorithm 2: Information Set Decoding [Prange, 1962]

Big difference with subset sums: one can use linear algebra



Repeat for several permutation matrices P

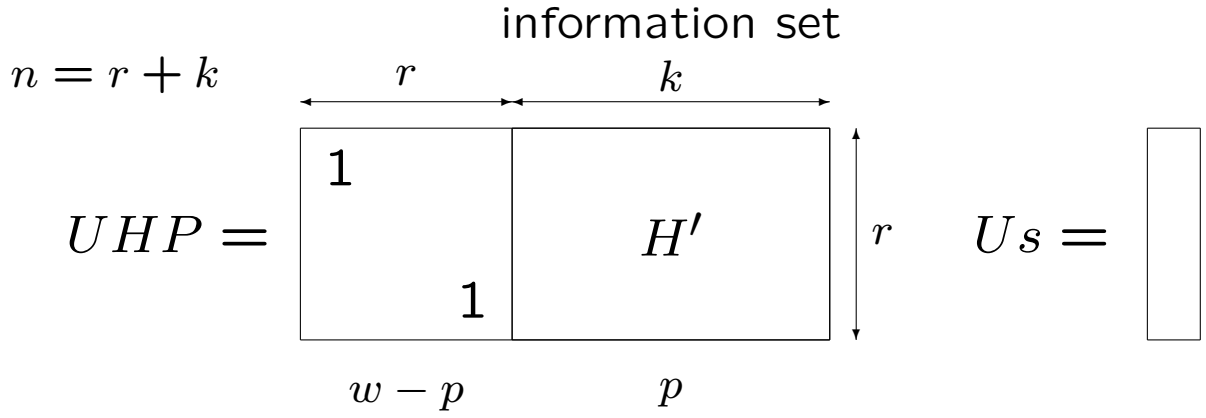
Claim: if $|Us| \leq w$, I win!

Success probability: $\binom{r}{w} / \binom{n}{w} \approx (r/n)^w$

Total cost: $\approx rn(n/r)^w$ column operations

Algorithm 2': ISD [Lee & Brickell, 1988]

Idea: amortize the Gaussian elimination



Repeat for several permutation matrices P

Claim: if $\exists e$ with $|e| = p$ and $|Us + H'e| = w - p$, I win!

Success probability: $\dfrac{\binom{r}{w-p}\binom{k}{p}}{\binom{n}{w}}$ Iteration cost: $rn + \binom{k}{p}$

Total cost: $\dfrac{\binom{n}{w}}{\binom{r}{w-p}}\left(1 + \dfrac{rn}{\binom{k}{p}}\right)$, only a polynomial gain

Generalized Information Set Decoding

[Stern, 89] ; [Dumer, 91]

$$UHP = \begin{array}{|c|c|} \hline 1 \quad\quad\; & H'' \\ \quad\quad\; 1 & \\ \hline 0 & H' \\ \hline \end{array} \quad \begin{array}{l} \\ \\ r-\ell \\ \ell \end{array} \qquad Us = \begin{array}{|c|} \hline s'' \\ \hline s' \\ \hline \end{array}$$

$k+\ell$ above the matrix; $w-p$ and p below the columns.

Repeat:
1. Permutation + partial Gaussian elimination
2. Find many e' of weight p such that $H'e' = s'$
3. For all good e', test $|s'' + H''e'| \leq w - p$

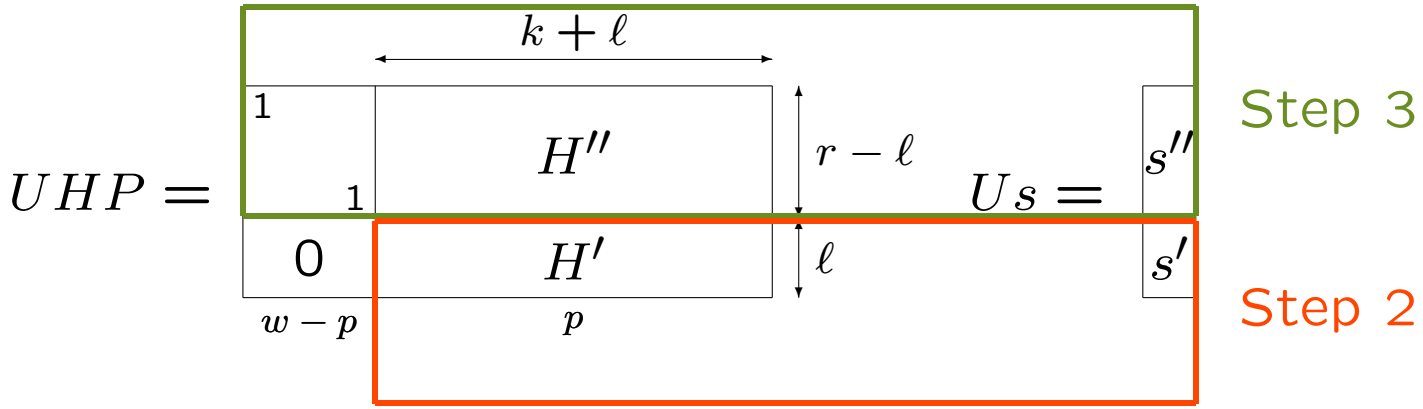
Step 3. is (a kind of) Lee & Brickell which embeds Step 2

Step 2. is Birthday Decoding (or whatever is best)

Total cost is minimized over ℓ and p

Generalized Information Set Decoding

[Stern, 89] ; [Dumer, 91]



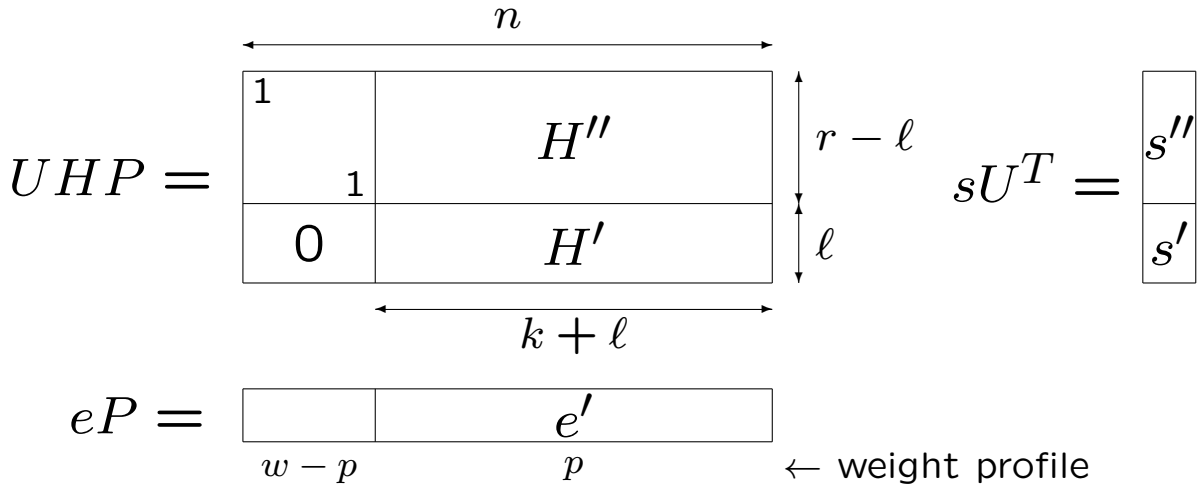
Repeat:
1. Permutation + partial Gaussian elimination
2. Find many e' of weight p such that $H'e' = s'$
3. For all good e', test $|s'' + H''e'| \leq w - p$

Step 3. is (a kind of) Lee & Brickell which embeds Step 2

Step 2. is Birthday Decoding (or whatever is best)

Total cost is minimized over ℓ and p

Generalized Information Set Decoding – Workfactor



Assuming the Gaussian elimination cost is not significant

$$\mathrm{WF}_{\mathrm{ISD}} = \min_{p,\ell} \frac{\binom{n}{w}}{\binom{r-\ell}{w-p}\binom{k+\ell}{p}} \left(\sqrt{\binom{k+\ell}{p}} + \frac{\binom{k+\ell}{p}}{2^\ell} \right)$$

column operations up to a small constant factor. Simplifies to

$$\mathrm{WF}_{\mathrm{ISD}} = \min_{p} \frac{\binom{n}{w}}{\binom{r-\ell}{w-p}\sqrt{\binom{k+\ell}{p}}} \text{ with } \ell = \log\left(\sqrt{\binom{k+\ell}{p}}\right)$$

Information Set Decoding – Timeline

- Information Set Decoding: [Prange, 62]
- Relax the weight profile: [Lee & Brickell, 88]
- Compute sums on partial columns first: [Leon, 88]
- Use the birthday attack: [Stern, 89], [Dumer, 91]
- First "real" implementation: [Canteaut & Chabaud, 98]
- Initial McEliece parameters broken: [Bernstein, Lange, & Peters, 08]
- Lower bounds: [Finiasz & Sendrier, 09]
- Ball-collision decoding [Bernstein, Lange, & Peters, 11]
- Asymptotic exponent improved [May, Meurer, & Thomae, 11]
- Decoding one out of many [Sendrier, 11]
- Even better asymptotic exponent [Becker, Joux, May, & Meurer, 12]
- "Nearest Neighbor" variant [May & Ozerov, 15]
- Sublinear error weight [Canto Torres & Sendrier, 16]

Information Set Decoding – Asymptotic Exponent

For a binary linear $[n, k]$ code and for t errors

We may express the cost of generic decoder as

$$\text{WF} = 2^{cn}$$

(maximized over k and with t at Gilbert-Varshamov bound)

- [Prange, 1962]: $c = 0.012$
- ...
- [May & Ozerov, 2015]: $c = 0.0097$ ($\approx 25\%$ gain)

Worse,

- [Canto Torres & Sendrier, 2016]:
  when $t = o(n)$, no asymptotic improvement since Prange

Key Security

This is the main security issue in code based cryptography

- Find families of codes whose generator matrices are indistinguishable from random matrices
- Goppa codes: excluding a few extremal cases, Goppa codes (binary or not) seem to be pseudorandom $\rightarrow$ best attack is essentially an exhaustive search

  We assume it is true, do we have better arguments?
- Can we find quasi-cyclic families which are indistinguishable?

  QC-MDPC is an answer to some extent. Can we do better?

Conclusion for Public Key Encryption

- Good security reduction

  partly heuristic though:
  - nothing proven on the average case complexity of decoding
  - indistinguishability assumptions need more attention

- The best attacks are decoding attacks

  $\rightarrow$ generic decoding is an essential long term research topic (including with quantum algorithms)

- Open problems are mainly related to the key security
  - find other good families of codes
  - safely reduce the public key size

V. Other Public Key Systems

Other Public Key Systems

- Digital Signature, [Courtois, Finiasz & Sendrier, 01]

  Same kind security reduction:
  Hardness of decoding & Indistinguishability of Goppa codes

- Zero Knowledge identification
  [Stern, 93], [Véron, 95], [Gaborit & Girault, 07]

  Much stronger security reduction: Hardness of decoding only

- And also...

  ID based signature [Cayrel, Gaborit & Girault, 07]

  Threshold ring signature [Aguilar-Melchor, Cayrel & Gaborit, 08],

CFS Digital Signature

$H \in \mathbf{F}_2^{r \times n}$ a parity check matrix of a t-error correcting Goppa code

Signing: the message M is given

- Hash the text M into a binary word $h(M) = s \in \mathbf{F}_2^r$
- Find e of minimal weight such that $eH^T = s$
- Use e as a signature

Verifying: M and e are given

- Hash the text M into a binary word $h(M) = s \in \mathbf{F}_2^r$
- Check $eH^T = s$

CFS Digital Signature – Not so Easy

In practice $n = 2^m = 2^{16}$, $t = 9$ and $r = n - k = tm = 144$
The public key H has size 144×65536 ($\approx 1.2\,\text{MB}$)

Let $s \in_R \mathbf{F}_2^{144}$, let w be the minimal weight of e such that $s = eH^T$

- $w \leq 9$ with probability $\approx 3\,10^{-6}$ (in general $w \leq t$ with prob. $1/t!$)
- $w = 10$ with probability $\approx 10^{-2}$
- $w = 11$ with probability $\approx 1 - 10^{-46}$

$w = 11$ is the smallest number such that $\binom{2^{16}}{11} > 2^{144}$

Problem:

- the trapdoor only allows the correction of $t = 9$ errors
- we need to decode 11 errors $\rightarrow$ we have to guess 2 error positions
- requires $t! = 362880$ decoding attempts on average

The legitimate user has to pay $\approx 2^{33}$ while the attacker has to pay $> 2^{77}$

CFS Digital Signature – Scalability

Binary Goppa code of length $n = 2^m$ correcting t errors

The public key $H \in \mathbf{F}_2^{r \times n}$ (where $r = tm$ is the codimension)

Signature cost	$t!O(m^2t^2)$
Signature length	$tm - \log_2(t!)$
Verification cost	$O(mt^2)$
Public key size	$tm2^m$
Security bits	$\frac{1}{2}tm$

- The signature cost is exponential in t
- The key size is exponential in m
- The security is exponential in tm

CFS Digital Signature – Decoding One Out of Many

Bleichenbacher's "Decoding One Out of Many"-type attack (2003 or 2004, unpublished) reduces the security to $\frac{1}{3}tm$

[Finiasz, 10] Parallel-CFS: sign several related syndrome.

- take a (λ times) longer hash of the message $h(M) = (s_1, ..., s_\lambda)$
- sign all λ syndromes $\rightarrow$ security back to $\frac{1}{2}tm$
- λ must be 3 or 4 (do not need to grow with the security parameter)

Signature length & cost and verification cost all multiplied by λ

CFS Digital Signature – Implementation

- [Landais & Sendrier, 12] Software implementation of parallel-CFS

  $(m, t) = (20, 8)$, $\lambda = 3 \rightarrow 80$ bits security

  Key size: 20 MB, one signature in ≈ 1.5 seconds

- [+ Schwabe] bit-sliced field arithmetic $\rightarrow$ 100 milliseconds for one signature

An important security issue: binary Goppa codes of rate $\rightarrow 1$ are not pseudorandom (no attack, but no security reduction either)

Stern ZK Authentication Protocol

Parameters: $H \in \mathbf{F}_2^{r \times n}$, weight $w > 0$, commitment scheme $c(\cdot)$
Secret: some word e of weight w ($w \approx$ Gilbert-Varshamov distance)
Public: the syndrome $s = eH^T$

	Prover		Verifier
Commitment	$\sigma \leftarrow \mathcal{S}_n$ $y \leftarrow \mathbf{F}_2^n$	$\xrightarrow{c_0, c_1, c_2}$	
Challenge		$\xleftarrow{b}$	$b \leftarrow \{0, 1, 2\}$
Answer		$\xrightarrow{A_b}$	check commitments

$$\begin{cases} c_0 = c(\sigma(y+e)) \\ c_1 = c(yH^T, \sigma) \\ c_2 = c(\sigma(y)) \end{cases} \qquad \begin{cases} A_0 = y, \sigma \\ A_1 = \sigma(y), \sigma(e) \\ A_2 = (y+e), \sigma \end{cases}$$

$$\text{Check: } \begin{cases} \text{if } b = 0 \text{ check } c_1 \text{ and } c_2 \\ \text{if } b = 1 \text{ check } c_0 \text{ and } c_2 \text{ (and } |\sigma(e)| = w) \\ \text{if } b = 2 \text{ check } c_0 \text{ and } c_1 \end{cases}$$

Stern ZK Authentication Protocol – Security

- An honest prover always succeeds (**completeness**)
- A dishonest prover succeeds for one round with probability 2/3 at most (eventually leading to **soundness**)
- No information on the secret leaks (**zero-knowledge**)

$\rightarrow$ For a security level S, $S/\log_2(3/2) \approx 1.7S$ rounds are needed (80 bits security $\rightarrow$ 137 rounds, 128 bits security $\rightarrow$ 219 rounds)

$\rightarrow$ Can be transformed into a signature (Fiat-Shamir NIZK)

$\rightarrow$ A tight security reduction to syndrome decoding

Signing with Stern ZK Protocol

	Prover		Verifier
Commitment	$\sigma_i \leftarrow \mathcal{S}_n$ $y_i \leftarrow \mathbf{F}_2^n$	$\xrightarrow{c_{0,i},c_{1,i},c_{2,i}}$	
Challenge		$\xleftarrow{b_i}$	$b_i \leftarrow \{0,1,2\}$
Answer		$\xrightarrow{A_{b_i,i}}$	check commitments

- Draw σ_i, y_i, and compute $c_{0,i}, c_{1,i}, c_{2,i}$ for all i, $1 \leq i \leq R$
- Compute $x = Hash((c_{0,i}, c_{1,i}, c_{2,i})_{1 \leq i \leq R})$
- Draw b_i, $1 \leq i \leq R$, using a PRNG with seed x
- The signature is $(A_{b_i,i}, c_{0,i}, c_{1,i}, c_{2,i})_{1 \leq i \leq R}$

80 bits security $\rightarrow$ signature of 174 Kbits

128 bits security $\rightarrow$ signature of 445 Kbits

[Aguilar-Melchor, Gaborit, & Schrek, 11] reduced to 79 and 202 Kbits

General Conclusions

- Code-based cryptosystems are practical, efficient, secure, versatile

  . . . some of them at least

- Also symmetric schemes (hash function, stream ciphers,. . .)

- Strong features
  - Hardness of decoding, tight security reductions in that respect
  - Efficient algorithms: fast public key encryption

- Not so strong features
  - Public key size (not necessarily a problem)
  - Few code families: biodiversity would be welcome

- Main open problems
  - Key security (security assumptions, families of codes, . . .)
  - Key size reduction: what gain for what cost?
  - Improve the digital signature

Thank you for your attention