Code-based Cryptography

Christiane Peters

Technische Universiteit Eindhoven

Workshop on Computer Security and Cryptography
CRM Montréal

April 12, 2010

Bad news

Quantum computers will break the most popular public-key cryptosystems:

- RSA,
- DSA,
- ECDSA,
- ECC,
- HECC
- ...

can be attacked in polynomial time using Shor's algorithm.

Good news

Post-quantum cryptography deals with cryptosystems that

- run on conventional computers and
- are secure against attacks by quantum computers.

Examples:

- Hash-based cryptography.
- Code-based cryptography.
- Lattice-based cryptography.
- Multivariate-quadratic-equations cryptography.
- Secret-key cryptography.

Overview:
Bernstein, Buchmann, and Dahmen, eds., Post-Quantum Cryptography. Springer, 2009.

Today's talk

Code-based cryptography.

1. Background
2. The McEliece cryptosystem
3. Attacks on the McEliece PKC
4. Recent results

1. Background

2. The McEliece cryptosystem

3. Attacks on the McEliece PKC

4. Recent results

Linear codes

A binary linear code C of length n and dimension k is a k-dimensional subspace of $\mathbf{F}_2^n$.

A generator matrix for C is a $k \times n$ matrix G such that $C = \{\mathbf{m}\,G : \mathbf{m} \in \mathbf{F}_2^k\}$.

The matrix G corresponds to a map $\mathbf{F}_2^k \to \mathbf{F}_2^n$ sending a message $\mathbf{m}$ of length k to an n-bit string.

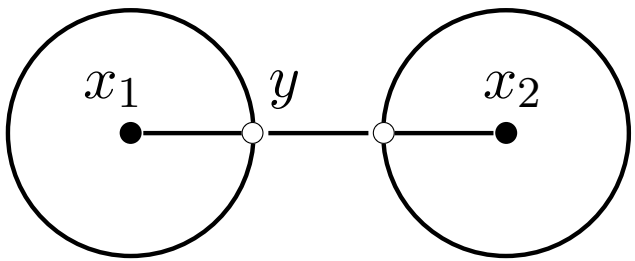
Example: The matrix

$$G = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

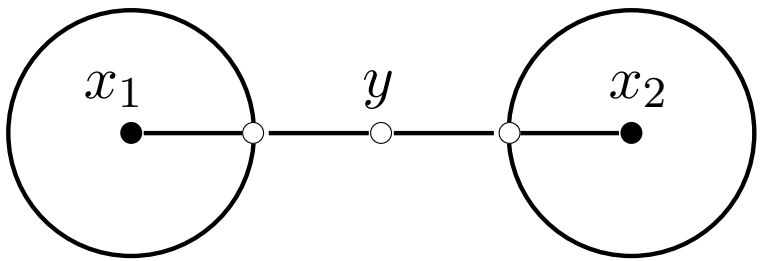
generates a code of length $n = 8$ and dimension $k = 4$.
Example of a codeword: $\mathbf{c} = (0110)G = (11110111)$.

Hamming distance

- The Hamming distance between two words in $\mathbf{F}_2^n$ is the number of coordinates where they differ.
- The Hamming weight of a word is the number of non-zero coordinates.
- The minimum distance of a linear code C is the smallest Hamming weight of a non-zero codeword in C.



code with minimum distance 3



code with minimum distance 4

Decoding problem

Classical decoding problem: find the closest codeword $\mathbf{c} \in C$ to a given $\mathbf{y} \in \mathbf{F}_2^n$, assuming that there is a unique closest codeword.

There are lots of code families with fast decoding algorithms

- E.g., Goppa codes/alternant codes, Reed-Solomon codes, Gabidulin codes, Reed-Muller codes, Algebraic-geometric codes, BCH codes etc.

However, given a binary linear code with no obvious structure.

Berlekamp, McEliece, van Tilborg (1978) showed that the general decoding problem for linear codes is NP-complete.

- About $2^{(0.5+o(1))n/\log_2(n)}$ binary operations required for a code of length n and dimension $\approx 0.5n$.

1. Background

2. The McEliece cryptosystem

3. Attacks on the McEliece PKC

4. Recent results

Setup of the McEliece cryptosystem

R. J. McEliece in 1978:

- Given a 50-error correcting classical binary Goppa code Γ of length 1024 and dimension 524 which is kept secret.

- The McEliece secret key is a triple (G, S, P) consisting of a generator matrix G for the code Γ;
  a 1024×1024 permutation matrix P, and
  an invertible 524×524 matrix S.

- The sizes $1024, 524, 50$ are public system parameters.

- The McEliece public key is the 524×1024 matrix
  $\hat{G} = SGP$.

Encryption and Decryption

McEliece encryption of a message $\mathbf{m} \in \{0,1\}^{524}$:

- Compute $\mathbf{m}\hat{G}$ and hide the message by adding a random length-1024 error vector $\mathbf{e}$ of weight 50.
- Send $\mathbf{y} = \mathbf{m}\hat{G} + \mathbf{e}$.

McEliece decryption:

- Compute $\mathbf{y}P^{-1} = \mathbf{m}SG + \mathbf{e}P^{-1}$.
- Note that $\mathbf{m}SG$ is a codeword in the secret code Γ and that the permuted error vector $\mathbf{e}P^{-1}$ has weight 50.
- Use the decoding algorithm for Γ to find $\mathbf{m}S$ and thereby $\mathbf{m}$.

In practice (1)

Biswas and Sendrier. McEliece Cryptosystem Implementation: Theory and Practice. PQCrypto 2008.

- 3.0GHz Intel Core 2 Duo E6850 CPU (single-core implementation)

n	k	w	encryption (cycles/byte)	decryption (cycles/byte)	key size	sec level
1024	524	50	243	7938	32 kB	60
2048	1696	32	178	1848	74 kB	87
8192	7958	18	119	312	232 kB	91

Comparison (EBATS preliminary report 2007):

	encryption (cycles/byte)	decryption (cycles/byte)
RSA 1024	800	23100
RSA 2048	834	55922
NTRU	4753	8445

In practice (2)

Eisenbarth, Güneysu, Heyse, and Paar. MicroEliece: McEliece for Embedded Devices. CHES 2009.

Linear binary code with $(n, k, w) = (2048, 1751, 27)$ providing 80-bit security.

1. ATxMega192A1 μC (16 kB of SRAM, 192 kB internal Flash memory) (clocked at 32 MHz)
   - generator matrix 448 kB does not fit into the 192 kB internal Flash memory
   - about $14 \cdot 10^6$ cycles for encryption of one message
   - about $20 \cdot 10^6$ cycles for decryption of one message
2. Xilinx Spartan-3AN XC3S1400AN-5 FPGA

1. Background

2. The McEliece cryptosystem

3. Attacks on the McEliece PKC

4. Recent results

Motivation

Goal:

- Strengthen confidence in the system.
- Find best parameters.

Note:

- All known attacks have exponential complexity.

The McEliece PKC from an attacker's point of view

An attacker who got hold of an encrypted message $\mathbf{y} = \mathbf{m}\hat{G} + \mathbf{e}$ has two possibilities in order to retrieve the original message $\mathbf{m}$.

- Find out the secret code; i.e., find G given $\hat{G}$, or
- Decode $\mathbf{y}$ without knowing an efficient decoding algorithm for the public code given by $\hat{G}$.

Attacks of the first type are called structural attacks.

- Goppa codes: no subexponentional time algorithm known to retrieve G.

We will deal with attacks of the second kind.

Attacks on the McEliece PKC

Most effective attack against the McEliece cryptosystem is information-set decoding.

Many variants: McEliece (1978), Leon (1988), Lee and Brickell (1988), Stern (1989), van Tilburg (1990), Canteaut and Chabanne (1994), Canteaut and Chabaud (1998), and Canteaut and Sendrier (1998), Bernstein-Lange-P. (2008), Finiasz-Sendrier (2009), Bernstein-Lange-P. (2010).

Information-set decoding (1)

Given a generator matrix G of a binary linear code C in systematic form, i.e., $G = (I_k \mid Q)$ for some $k \times (n-k)$-matrix Q.

Let $\mathbf{c} = \mathbf{m}G = \mathbf{m}(I_k|Q)$ for some $\mathbf{m} \in \mathbf{F}_2^k$.

Note that the first k positions of $\mathbf{c}$ equal $\mathbf{m}$.

Let $\mathbf{y} \in \mathbf{F}_2^n$ and let $\mathbf{c}$ be the closest codeword in C at distance w, i.e., $\mathbf{y} = \mathbf{c} + \mathbf{e}$ for a vector $\mathbf{e}$ of weight w.

- If the first k positions of $\mathbf{y}$ are error-free, then the first k positions of $\mathbf{y}$ are the original message $\mathbf{m}$, and $\mathbf{c} = \mathbf{m}G = \mathbf{y}|_{(1\ldots k)}G$.
- Probability $\frac{\binom{n-k}{w}}{\binom{n}{w}}$.

Information-set decoding (2)

Given a generator matrix G of a binary linear code C in systematic form, i.e., $G = (I_k \mid Q)$ for some $k \times (n-k)$-matrix Q.

Let $\mathbf{c} = \mathbf{m}G = \mathbf{m}(I_k|Q)$ for some $\mathbf{m} \in \mathbf{F}_2^k$.

Note that the first k positions of $\mathbf{c}$ equal $\mathbf{m}$.

Let $\mathbf{y} \in \mathbf{F}_2^n$ and let $\mathbf{c}$ be the closest codeword in C at distance w, i.e., $\mathbf{y} = \mathbf{c} + \mathbf{e}$ for a vector $\mathbf{e}$ of weight w.

- If the first k positions of $\mathbf{y}$ contain 1 error, $\mathbf{m}G = \mathbf{y}|_{(1\ldots k)}G + \mathbf{g}_i$ for some row $\mathbf{g}_i$ of G.
- Find the row of G corresponding to this error position. Additional cost: check all rows of G.
- Probability $\frac{\binom{k}{1}\binom{n-k}{w-1}}{\binom{n}{w}} = k\frac{\binom{n-k}{w-1}}{\binom{n}{w}}$.

Information-set decoding (3)

Given a generator matrix G of a binary linear code C in systematic form, i.e., $G = (I_k \mid Q)$ for some $k \times (n-k)$-matrix Q.

Let $\mathbf{c} = \mathbf{m}G = \mathbf{m}(I_k|Q)$ for some $\mathbf{m} \in \mathbf{F}_2^k$.

Note that the first k positions of $\mathbf{c}$ equal $\mathbf{m}$.

Let $\mathbf{y} \in \mathbf{F}_2^n$ and let $\mathbf{c}$ be the closest codeword in C at distance w, i.e., $\mathbf{y} = \mathbf{c} + \mathbf{e}$ for a vector $\mathbf{e}$ of weight w.

- If the first k positions of $\mathbf{y}$ contain 2 errors, $\mathbf{m}G = \mathbf{y}|_{(1\ldots k)}G + \mathbf{g}_i + \mathbf{g}_j$ for two rows $\mathbf{g}_i$ and $\mathbf{g}_j$ of G.
- Find the rows of G corresponding to those two error positions.
  Additional cost: check all combinations of two rows of G.
- Probability $\frac{\binom{k}{2}\binom{n-k}{w-2}}{\binom{n}{w}}$.

Information-set decoding algorithms

Error distribution among the columns of G.

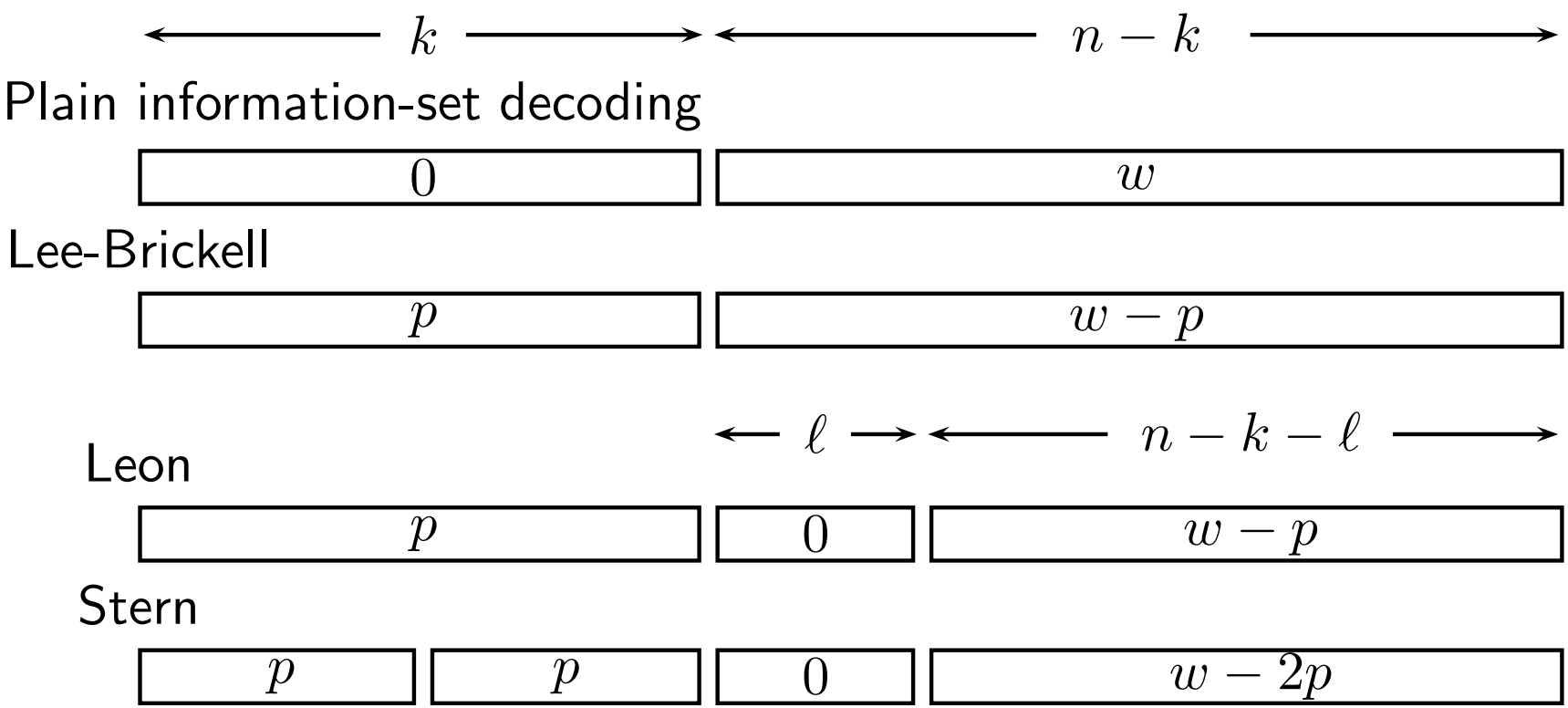


Figure from Overbeck and Sendrier: *Code-based Cryptography*, in *Post-Quantum Cryptography* (eds.: Bernstein, Buchmann, and Dahmen)

1. Background

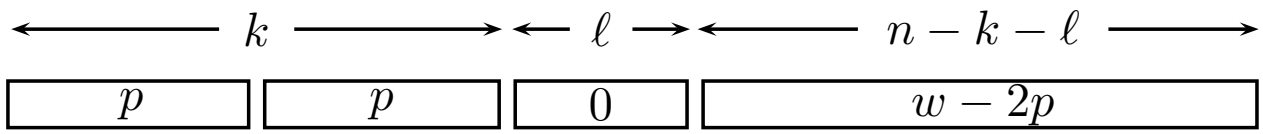
2. The McEliece cryptosystem

3. Attacks on the McEliece PKC

4. Recent results

Bernstein, Lange, P. at PQCrypto 2008

- Attack is most easily understood as a variant of Stern's attack.



- Attack software extracts a plaintext from a ciphertext by decoding 50 errors in a $[1024, 524]$ binary code.
- Faster by a factor of more than 150 than previous attacks; now within reach of a moderate cluster of computers.
- Attack on a single computer with a 2.4GHz Intel Core 2 Quad Q6600 CPU would need approximately 1400 days (2^{58} CPU cycles) (or on 200 such computers $\approx$ one week)

Actual attack

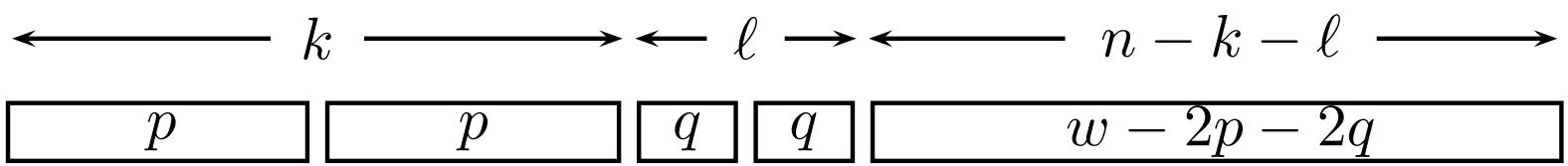
- About 200 computers involved, with about 300 cores; computation finished in under 90 days; used about 8000 core-days

Bounds

Finiasz and Sendrier. Security bounds for the design of code-based cryptosystems. Asiacrypt 2009.

- Lower bound on cost of information-set decoding.
- Birthday-decoding trick increasing the probability of an iteration to succeed in Stern's algorithm.

Bernstein, Lange, P. Ball-collision decoding. To appear (2010).



- Asymptotically beating the Finiasz-Sendrier "lower bound".
- Proposing a new safer bound for information-set-decoding algorithms.

Thank you for your attention!