Off-Shor

Christiane Peters

Technische Universiteit Eindhoven

JQI/NIST Quantum Information Workshop

October 28, 2010

Another advertisement

Christiane Peters

Technische Universiteit Eindhoven

JQI/NIST Quantum Information Workshop

October 28, 2010

Code-based cryptography: the state of the art

Christiane Peters

Technische Universiteit Eindhoven

JQI/NIST Quantum Information Workshop

October 28, 2010

1. Public-key cryptography

2. Linear error-correcting codes

3. The McEliece cryptosystem

4. Attacks on the McEliece PKC

5. Designs

Public-key cryptography

- Introduced by Diffie and Hellman in 1976
- Use a key pair consisting of a public key and a private key.
- Each key has only one task: the public key is used for encryption and the private key for decryption.
- Should be infeasible to derive the private key from the public key.

Post-quantum cryptography

- Shor strikes: Quantum computers will break the most popular public-key cryptosystems such as RSA, DSA, ECDSA etc.

- Off-Shor: This talk's example for post-quantum cryptography:
  - Code-based cryptography — started by McEliece in 1978.

Background on McEliece's work

- Robert J. McEliece: Allen E. Puckett Professor and Professor of Electrical Engineering at CalTech.

- Work on information theory and theory of error-correcting codes.

- Robert J. McEliece. A public-key cryptosystem based on algebraic coding theory, 1978. Jet Propulsion Laboratory DSN Progress Report 42–44.
  URL: `http://ipnpr.jpl.nasa.gov/progress_report2/42-44/44N.pdf`.

1. Public-key cryptography

2. Linear error-correcting codes

3. The McEliece cryptosystem

4. Attacks on the McEliece PKC

5. Designs

Linear codes

A 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

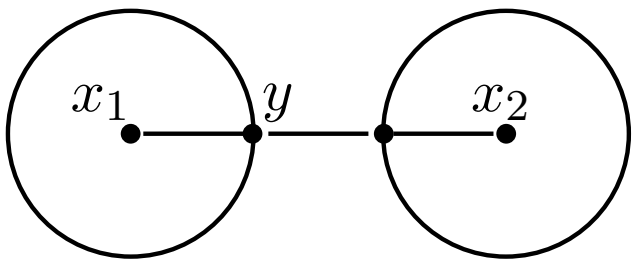
$$G = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

generates a code over $\mathbf{F}_2$ of length $n = 8$ and dimension $k = 4$.

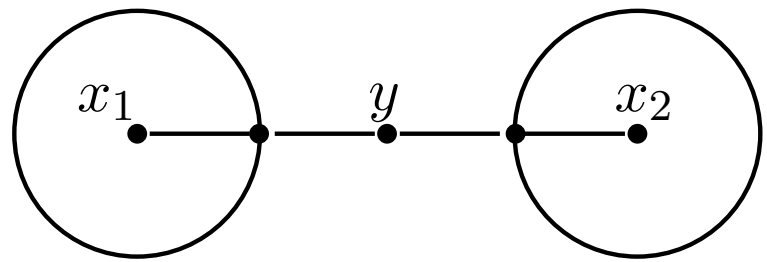
Example of a codeword: $\mathbf{c} = (0110)G = (11111011)$.

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.

- Best known attack: 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. Public-key cryptography

2. Linear error-correcting codes

3. The McEliece cryptosystem

4. Attacks on the McEliece PKC

5. Designs

Encryption

- The public key is a random-looking 524×1024 matrix $\hat{G}$ with entries in $\{0, 1\}$.
- Encrypt a message $m \in \{0, 1\}^{524}$ as

$$m\hat{G} + e$$

  where e is a random error vector of length 1024 and weight 50.
- The sizes $1024, 524, 50$ are public system parameters.
- Need to correct 50 errors to find m.
- Decoding is not easy without knowing the structure of the code generated by $\hat{G}$.

Secret key

The public key $\hat{G}$ has a hidden Goppa-code structure allowing fast decoding:

$$\hat{G} = SGP$$

where

- G is the generator matrix of a Goppa code Γ of length 1024 and dimension 524 and error-correcting capability 50;
- S is a random 524×524 invertible matrix; and
- P is a random 1024×1024 permutation matrix.

The triple (G, S, P) together with a decoding algorithm for Γ form the secret key.

Note: Detecting this structure, i.e., finding G given $\hat{G}$, seems even more difficult than attacking a random $\hat{G}$.

Decryption

The legitimate receiver knows S, G and P with $\hat{G} = SGP$ and a decoding algorithm for Γ.

How to decrypt $y = m\hat{G} + e$.

1. Compute $yP^{-1} = mSG + eP^{-1}$.
2. Apply the decoding algorithm of Γ to correct 50 errors and thereby finding mSG which is a codeword in Γ from which one obtains 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 2^{80} 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

Also: Heyse. Low-Reiter: Niederreiter Encryption Scheme for Embedded Microcontrollers. PQCrypto 2010.

1. Public-key cryptography

2. Linear error-correcting codes

3. The McEliece cryptosystem

4. Attacks on the McEliece PKC

5. Designs

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:

Prange (1962), Leon (1988), Lee and Brickell (1988), Krouk (1989), Stern (1989), Dumer (1989), van Tilburg (1990) (independently from Omura in the 1960's), Dumer (1991), Canteaut and Chabanne (1994), Canteaut and Chabaud (1998), Canteaut and Sendrier (1998), Bernstein-Lange-P. (2008), Finiasz-Sendrier (2009), P. (2010), Bernstein-Lange-P. (2010).

Information-set decoding is the best non-structural attack to decrypt a message without knowing the McEliece secret key.

Information sets

Given a generator matrix G of a code of length n and dimension k.

- An information set is a size-k subset $I \subseteq \{1, 2, \ldots, n\}$ such that the I-indexed columns of G are invertible.
- Denote the matrix formed by the I-indexed columns of G by G_I. The I-indexed columns of $G_I^{-1}G$ are the $k \times k$ identity matrix.

Let $\mathbf{y} \in \mathbf{F}_2^n$ have distance w to a codeword in $\mathbf{F}_2^k G$, i.e., $\mathbf{y} = \mathbf{c} + \mathbf{e}$ for a codeword $\mathbf{c} \in \mathbf{F}_2^k G$ and a vector $\mathbf{e}$ of weight w.

- Denote the I-indexed positions of $\mathbf{y}$ by $\mathbf{y}_I$.
- If $\mathbf{y}_I$ is error-free, $\mathbf{y}_I G_I^{-1}$ is the original message and $\mathbf{c} = (\mathbf{y}_I G_I^{-1})G$. Thus, $\mathbf{e} = \mathbf{y} - (\mathbf{y}_I G_I^{-1})G$.

Information-set-decoding algorithms

Error distribution among the columns of G.

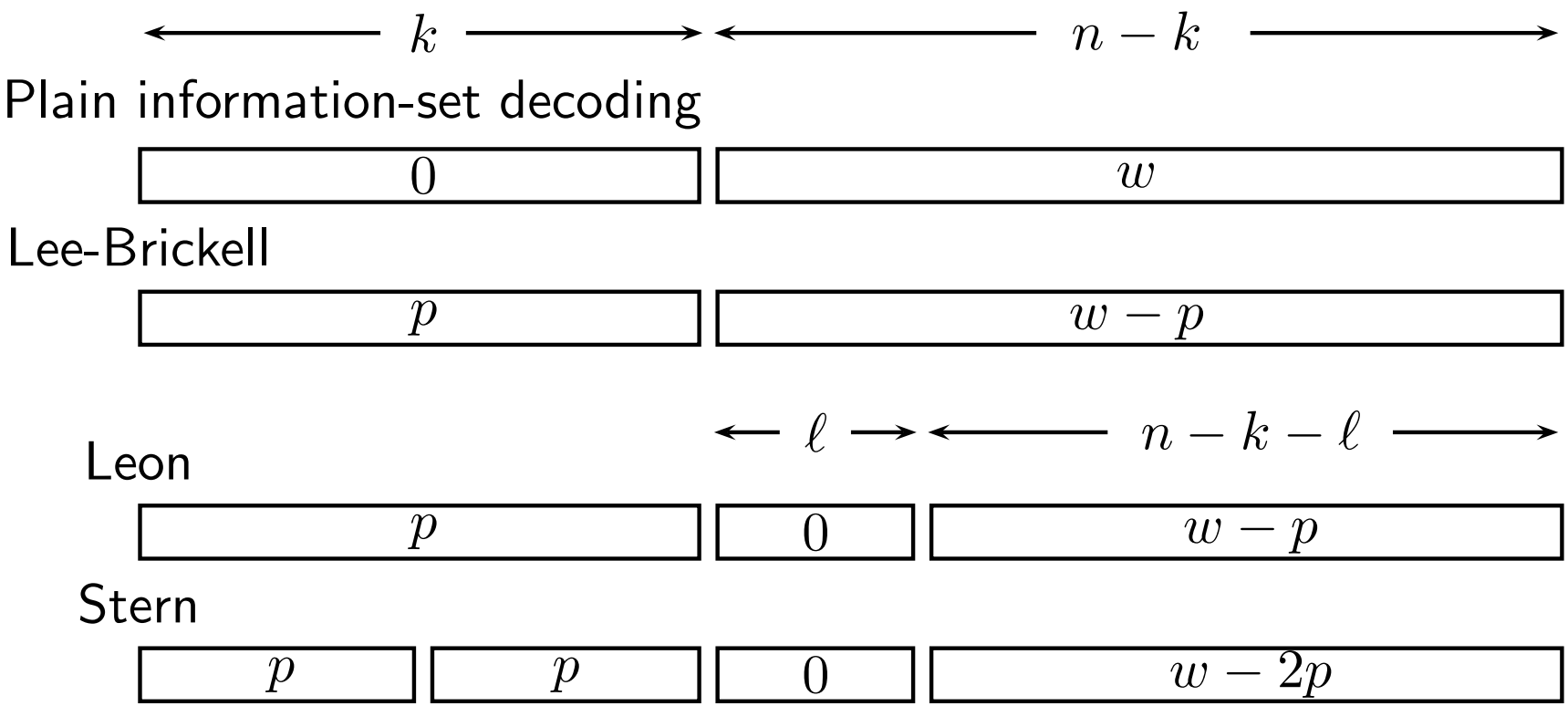
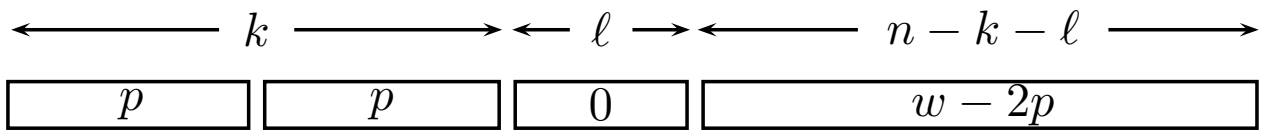


Figure from Overbeck and Sendrier: *Code-based Cryptography*, in *Post-Quantum Cryptography* (eds.: Bernstein, Buchmann, and Dahmen)

Bernstein, Lange, P. at PQCrypto 2008

- Actual break of McEliece's original parameters (used variant of Stern's algorithm).



- 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

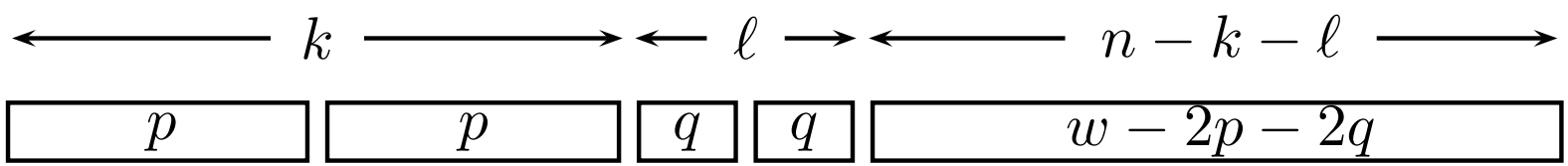
- About 200 computers involved, with about 300 cores; computation finished in under 90 days; used about 8000 core-days

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. Preprint 2010.



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

Quantum information-set decoding

- Conventional wisdom: Grover's algorithm (square-root complexity) forces key size to double.

Bernstein. Grover vs. McEliece. PQCrypto 2010.
`http://cr.yp.to/papers.html#grovercode`

- Describes how to attack the McEliece PKC using Grover's quantum root-finding algorithm.

- Analyzes the cost of this quantum information-set-decoding algorithm.

Consequence: asymptotically one needs to quadruple the McEliece key size to thwart Bernstein's attack.

1. Public-key cryptography

2. Linear error-correcting codes

3. The McEliece cryptosystem

4. Attacks on the McEliece PKC

5. Designs

Reducing the key size (1)

- Bernstein, Lange, P., PQCrypto 2008: binary-Goppa-code parameters achieving 2^{128} security produce a 1537536-bit key.

- Smaller-key variants use other codes such as Reed-Solomon codes, generalized Reed-Solomon codes, quasi-cyclic codes, quasi-dyadic codes or geometric Goppa codes.

Quasi-dyadic codes

Misoczki-Barreto. Compact McEliece Keys from Goppa Codes. SAC 2009.

- Hide quasi-dyadic Goppa code as quasi-dyadic public key.
- Certain instances broken (Faugere et al, Eurocrypt 2010).
- Binary quasi-dyadic Goppa codes still hold up. `http://eprint.iacr.org/2009/187`
- For 2^{128} security the dyadic public key has only 32768 key bits.

Reducing the key size (2)

- Goppa codes are the most confidence-inspiring choice.
- Using Goppa codes over larger fields decreases the key size at the same security level against information-set decoding (P., PQCrypto 2010).
- A Goppa code over $\mathbf{F}_{31}$ leads to a 725741-bit key for 2^{128} security.
- Drawback: can correct only $\lfloor (t+1)/2 \rfloor$ errors if $q > 2$ (vs. t in the binary case).

Wild McEliece

Bernstein, Lange, P. Wild McEliece. SAC 2010.

Use the McEliece cryptosystem with Goppa codes of the form

$$\Gamma_q(a_1, \ldots, a_n, g^{q-1})$$

where g is an irreducible monic polynomial in $\mathbf{F}_{q^m}[x]$ of degree t.

- Note the exponent $q-1$ in g^{q-1}.
- We refer to these codes as wild Goppa codes.

Minimum distance of wild Goppa codes

Theorem (Sugiyama-Kasahara-Hirasawa-Namekawa, 1976)

$$\Gamma_q(a_1, \ldots, a_n, g^{q-1}) = \Gamma_q(a_1, \ldots, a_n, g^q)$$

for a monic squarefree polynomial $g(x)$ in $\mathbf{F}_{q^m}[x]$ of degree t.

- Our paper contains a streamlined proof.
- The case $q = 2$ of this theorem is due to Goppa, using a different proof that can be found in many textbooks.

Error-correcting capability

- Since $\Gamma_q(\ldots, g^{q-1}) = \Gamma_q(\ldots, g^q)$ the minimum distance of $\Gamma_q(\ldots, g^{q-1})$ equals the one of $\Gamma_q(\ldots, g^q)$ and is thus $\geq \deg g^q + 1 = qt + 1$.

- We present an alternant decoder that allows efficient correction of $\lfloor qt/2 \rfloor$ errors for $\Gamma_q(\ldots, g^{q-1})$.

- Note that the number of efficiently decodable errors increases by a factor of $q/(q-1)$ while the dimension $n - m(q-1)t$ of $\Gamma_q(\ldots, g^{q-1})$ stays the same.

Attacks on Wild McEliece

- The wild McEliece cryptosystem includes, as a special case, the original McEliece cryptosystem.

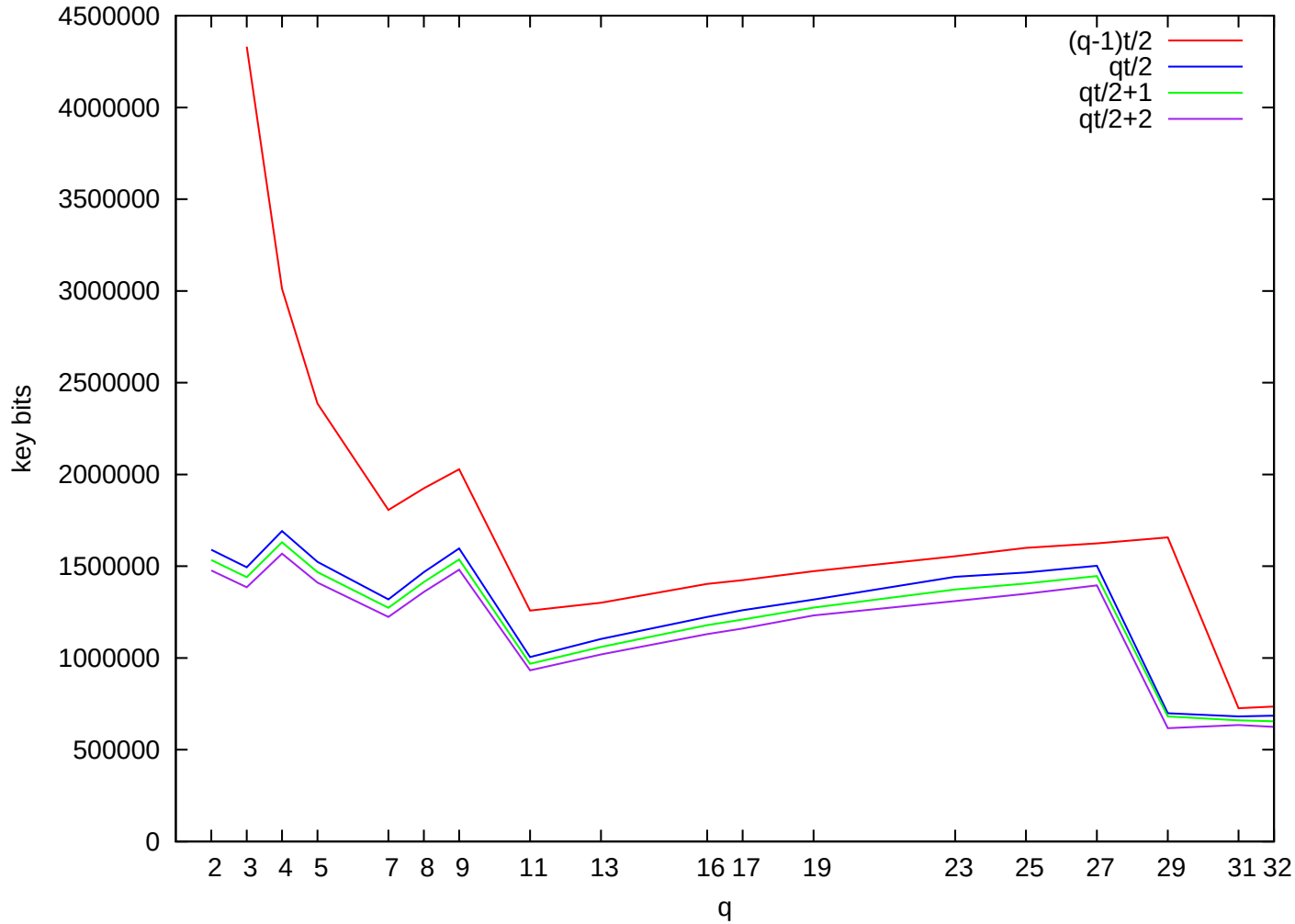
- A complete break of the wild McEliece cryptosystem would therefore imply a complete break of the original McEliece cryptosystem.

Information-set decoding vs. Wild McEliece

- The top threat against the original McEliece cryptosystem is information-set decoding.
- The same attack also appears to be the top threat against the wild McEliece cryptosystem for $\mathbf{F}_3$, $\mathbf{F}_4$, etc.
- Use complexity analysis of state-of-the-art information-set decoding for linear codes over $\mathbf{F}_q$ from [P. 2010] to find parameters (q, n, k, t) for Wild McEliece.

Key sizes for various q at a 2^{128} security level

McEliece with $\Gamma_q(a_1, \ldots, a_n, g^{q-1})$ and $\lfloor (q-1)t/2 \rfloor$, $\lfloor qt/2 \rfloor$, $\lfloor qt/2 \rfloor + 1$, or $\lfloor qt/2 \rfloor + 2$ added errors.



PQCrypto 2011

Nov 29 – Dec 2, Taipei

`http://pq.crypto.tw/pqc11/`

Thank you for your attention!