LDPC Codes in the McEliece Cryptosystem: attacks and countermeasures

Marco Baldi

DIBET, Università Politecnica delle Marche, Ancona, Italy
m.baldi@univpm.it

NATO ARW 2008
Veliko Tarnovo, Bulgaria
October 7$^{\mathrm{th}}$

The McEliece cryptosystem

- Public-key cryptosystem based on algebraic coding theory [McEliece1978].
- It adopts generator matrices as private and public keys.
- Security lies in the difficulty of decoding a large linear code with no visible structure, that is an NP complete problem [Berlekamp1978].

Advantages

The system is faster than competing solutions, like RSA.

Drawbacks

It has large public keys and low transmission rate.

- McEliece, R.J., "A public-key cryptosystem based on algebraic coding theory." DSN Progress Report (1978) 114–116
- Berlekamp, E., McEliece, R., van Tilborg, H., "On the inherent intractability of certain coding problems." IEEE Trans. Inform. Theory **24** (May 1978) 384–386

The McEliece cryptosystem (2)

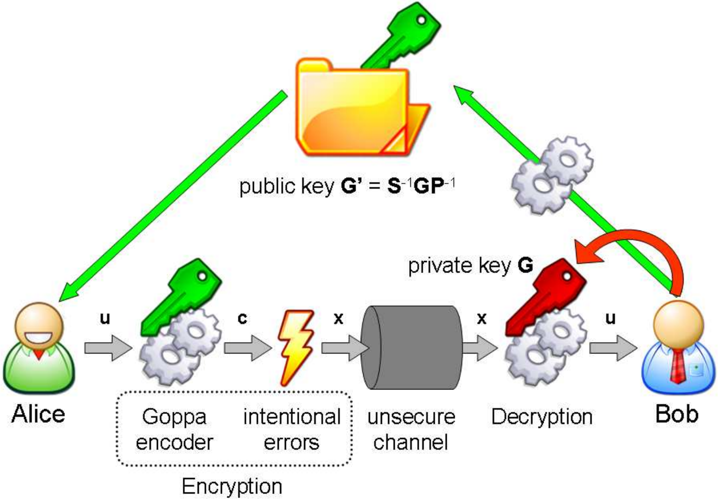
- Private key:
  - $\mathbf{G}$: generator matrix of a t-error correcting Goppa code
  - $\mathbf{S}$: $k \times k$ non-singular scrambling matrix
  - $\mathbf{P}$: $n \times n$ permutation matrix
- Public key: $\mathbf{G}' = \mathbf{S}^{-1} \cdot \mathbf{G} \cdot \mathbf{P}^{-1}$
- Encryption map:

$$\mathbf{x} = \mathbf{u} \cdot \mathbf{G}' + \mathbf{e}$$

(Goppa encoding and addition of t intentional errors)

- Decryption map:
  - $\mathbf{x}' = \mathbf{x} \cdot \mathbf{P} = \mathbf{u} \cdot \mathbf{S}^{-1} \cdot \mathbf{G} + \mathbf{e} \cdot \mathbf{P}$ (inversion of the permutation)
  - $\mathbf{x}' \Rightarrow \mathbf{u}' = \mathbf{u} \cdot \mathbf{S}^{-1}$ (Goppa decoding to correct t errors)
  - $\mathbf{u} = \mathbf{u}' \cdot \mathbf{S}$ (inversion of the scrambling)

The McEliece cryptosystem (3)



The McEliece cryptosystem (4)

- The original version adopts Goppa codes with length $n = 1024$, dimension $k = 524$, and minimum distance $d_{\min}$ of at least 101.
- The key size is $n \times k$ bits $=$ 67072 bytes.
- The transmission rate is $k/n \approx 0.5$.
- Several attempts have been made for adopting other codes, able to overcome the system's drawbacks...
- ...but they always compromised the system security [Niederreiter1986], [Monico2000], [Gaborit2005].

- Niederreiter, H., "Knapsack-type cryptosystems and algebraic coding theory." Probl. Contr. and Inform. Theory **15** (1986) 159–166
- Monico, C. and Rosenthal, J. and Shokrollahi, A. "Using low density parity check codes in the McEliece cryptosystem." Proc. IEEE ISIT 2000, Sorrento, Italy, (June 2000) 215
- Gaborit, P., "Shorter keys for code based cryptography." Proc. WCC 2005, Bergen, Norway (March 2005) 81–90

The McEliece cryptosystem (5)

LDPC Codes in the McEliece Cryptosystem

Marco Baldi

McEliece cryptosystem based on QC-LDPC codes

Preliminaries

The previous proposal and the OTD attacks

The new proposals

Performance

Complexity Assessment

Comparisons

Conclusion

- The original version is unbroken up to now.
- Information Set Decoding (ISD) attacks are among the most dangerous ones.
- Lee and Brickell's method [Lee1988] exploits the random choice of $\mathbf{e}$.
- Alternatively, $\mathbf{e}$ could be searched as the lowest weight codeword in the extended code generated by $\mathbf{G}'' = \begin{bmatrix} \mathbf{G}' \\ \mathbf{x} \end{bmatrix}$.
- Stern's algorithm is effective for searching for low weight codewords in a linear block code [Stern1989].
- It requires $2^{63.5}$ binary operations with the original parameters.

- Lee, P. and Brickell, E., "An observation on the security of McEliece's public-key cryptosystem." Advances in Cryptology - EUROCRYPT 88, Springer-Verlag (1988) 275–280
- Stern, J., "A method for finding codewords of small weight." In Cohen, G., Wolfmann, J., eds.: Coding Theory and Applications. Volume 388 of LNCS., Springer (1989) 106–113

Low-Density Parity-Check Codes

- LDPC codes are state-of-art forward error correcting codes.
- They approach the theoretical Shannon limit [Richardson2001].
- Each code is defined as the kernel of a sparse $(n-k) \times n$ parity-check matrix $\mathbf{H}$.
- Belief Propagation decoding exploits the sparse nature of these matrices to implement very efficient and low-complexity decoding.
- Quasi-cyclic (QC) LDPC codes are a particular class of LDPC codes, characterized by structured $\mathbf{H}$ matrices that allow low-complexity encoding as well.

▶ Richardson, T., Urbanke, R., "The capacity of low-density parity-check codes under message-passing decoding." IEEE Trans. Inform. Theory **47** (February 2001) 599–618

QC-LDPC Codes

We consider a particular class of QC-LDPC codes, for which

The parity-check matrix $\mathbf{H}$

is formed by a row $\{\mathbf{H}_0, \ldots, \mathbf{H}_{n_0-1}\}$ of n_0 binary circulant blocks with size p and row/column weight d_v.

The generator matrix $\mathbf{G}$

is formed by a $k \times k$ identity matrix $\mathbf{I}$ ($k = k_0 \cdot p$), followed by a column of k_0 binary circulant blocks with size p. If $\mathbf{H}_{n_0-1}$ is non-singular,

$$\mathbf{G} = \begin{bmatrix} \mathbf{I} & \begin{matrix} \left(\mathbf{H}_{n_0-1}^{-1} \cdot \mathbf{H}_0\right)^T \\ \left(\mathbf{H}_{n_0-1}^{-1} \cdot \mathbf{H}_1\right)^T \\ \vdots \\ \left(\mathbf{H}_{n_0-1}^{-1} \cdot \mathbf{H}_{n_0-2}\right)^T \end{matrix} \end{bmatrix}.$$

QC-LDPC Codes based on RDF

- "Random Difference Families" (RDFs) are random multi-sets with the properties of difference families.
- They can be used to design **H** matrices for the considered QC-LDPC codes.
- For fixed parameters, the number of different codes is very high.
- Moreover, the generated codes have the same:
  - code length and rate
  - parity-check matrix density
  - nodes degree distributions
  - girth length distribution
- The properties above yield equivalent error correction performance under message passing decoding.

McEliece cryptosystem based on QC-LDPC Codes

- In the original system (adopting Goppa codes) it suffices to hide the secret code through a permutation (**P**).
- When adopting LDPC codes, the sparse nature of the **H** matrix must be hidden to avoid attacks based on it.
- We have recently proposed a QC-LDPC based variant that adopts a more dense transformation [Baldi2007] (**Q**).
- This causes an "error spreading" phenomenon during decryption...
- ...but it is compensated by the high correction capability of LDPC codes.
- This version is able to counter all the classic attacks.

▶ Baldi, M., Chiaraluce, F., "Cryptanalysis of a new instance of McEliece cryptosystem based on QC-LDPC codes." Proc. IEEE ISIT 2007, Nice, France (June 2007) 2591–2595

McEliece cryptosystem based on QC-LDPC Codes (2)

- Bob randomly chooses a code in a family of (n_0, d_v, p) QC-LDPC codes by selecting its parity-check matrix $\mathbf{H}$.
- Private key:
  - $\mathbf{G}$: generator matrix in reduced echelon form
  - $\mathbf{S}$: $k \times k$ non-singular scrambling matrix
  - $\mathbf{Q}$: $n \times n$ sparse transformation matrix (row/col weight m)
- Public key: $\mathbf{G}' = \mathbf{S}^{-1} \cdot \mathbf{G} \cdot \mathbf{Q}^{-1}$
- $\mathbf{G}'$ can be seen as a $k_0 \times n_0$ matrix with entries in the ring of polynomials $\mathbb{R} = \mathrm{GF}(2)[x]/(x^p + 1)$.
- It can be simply described by the set of polynomial coefficients.

McEliece cryptosystem based on QC-LDPC Codes (3)

- Encryption map (same as in standard McEliece PKC):

$$\mathbf{x} = \mathbf{u} \cdot \mathbf{G}' + \mathbf{e}$$

(LDPC encoding and addition of t' intentional errors)

- Decryption map:
  - $\mathbf{x}' = \mathbf{x} \cdot \mathbf{Q} = \mathbf{u} \cdot \mathbf{S}^{-1} \cdot \mathbf{G} + \mathbf{e} \cdot \mathbf{Q}$ (inversion of the transformation and error spreading)
  - $\mathbf{x}' \Rightarrow \mathbf{u}' = \mathbf{u} \cdot \mathbf{S}^{-1}$ (LDPC decoding to correct up to $t = t' \cdot m$ errors)
  - $\mathbf{u} = \mathbf{u}' \cdot \mathbf{S}$ (inversion of the scrambling)

System Parameters

- In the previous version of the cryptosystem we fixed $n_0 = 4$, $d_v = 13$, $p = 4032$, $m = 7$ and $t' = 27$.
- Such choice allows to resist all the standard attacks.
- Both $\mathbf{S}$ and $\mathbf{Q}$ were chosen sparse, with non-null blocks having row/column weight m, and

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_0 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q}_{n_0-1} \end{bmatrix}.$$

- This, together with its low density, gave raise to a new attack formulated by Otmani et al. (OTD) [Otmani2008].

▶ Otmani, A., Tillich, J.P., Dallot, L., "Cryptanalysis of two McEliece cryptosystems based on quasi-cyclic codes." Proc. SCC 2008, Beijing, China (April 2008)

Rationale of the OTD attacks

- By selecting the first k columns of $\mathbf{G}'$, an eavesdropper can obtain

$$\mathbf{G}'_{\leq k} = \mathbf{S}^{-1} \cdot \begin{bmatrix} \mathbf{Q}_0^{-1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_1^{-1} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{Q}_{n_0-2}^{-1} \end{bmatrix}.$$

- By inverting $\mathbf{G}'_{\leq k}$ and considering its block at position (i, j), he can obtain $\mathbf{Q}_i \mathbf{S}_{i,j}$, that corresponds to the polynomial

$$g_{i,j}(x) = q_i(x) \cdot s_{i,j}(x) \mod (x^p + 1).$$

- Both $\mathbf{Q}_i$ and $\mathbf{S}_{i,j}$ are sparse, so it is highly probable that $g_{i,j}(x)$ has exactly m^2 non-null coefficients and its support contains at least one shift $x^{l_a} \cdot q_i(x)$, $0 \leq l_a \leq p - 1$.

OTD attack strategies

First strategy

- The attacker can enumerate all the m-tuples belonging to the support of $g_{i,j}(x)$.
- Each m-tuple is then validated through inversion of its corresponding polynomial and multiplication by $g_{i,j}(x)$.
- If the resulting polynomial has exactly m non-null coefficients, the m-tuple is a shifted version of $q_i(x)$ with very high probability.
- This attack requires a work factor of $2^{50.3}$ binary operations for the specified parameters.

OTD attack strategies

Second strategy

- It is highly probable that the Hadamard product of the polynomial $g_{i,j}(x)$ with a d-shifted version of itself, $g_{i,j}^{d}(x) * g_{i,j}(x)$, gives a shifted version of $q_i(x)$, for a specific value of d.
- The eavesdropper can calculate all the possible $g_{i,j}^{d}(x) * g_{i,j}(x)$ and check whether the resulting polynomial has support with weight m.
- This attack requires a work factor of 2^{36} binary operations.
- The work factor could be even reduced by calculating the periodic autocorrelation of the coefficients of $g_{i,j}(x)$ through efficient algorithms for finding d.

OTD attack strategies

LDPC Codes in the McEliece Cryptosystem

Marco Baldi

McEliece cryptosystem based on QC-LDPC codes

Preliminaries

The previous proposal and the OTD attacks

The new proposals

Performance

Complexity Assessment

Comparisons

Conclusion

Third strategy

- The attacker can consider the i-th row of the inverse of $\mathbf{G}'_{\leq k}$:

$$\mathbf{R}_i = [\mathbf{Q}_i \mathbf{S}_{i,0} | \mathbf{Q}_i \mathbf{S}_{i,1} | \ldots | \mathbf{Q}_i \mathbf{S}_{i,n_0-2}]$$

and the linear code generated by

$$\mathbf{G}_{OTD3} = (\mathbf{Q}_i \mathbf{S}_{i,0})^{-1} \cdot \mathbf{R}_i = \left[\mathbf{I} | \mathbf{S}_{i,0}^{-1} \mathbf{S}_{i,1} | \ldots | \mathbf{S}_{i,0}^{-1} \mathbf{S}_{i,n_0-2} \right].$$

- It admits an alternative generator matrix:

$$\mathbf{G}'_{OTD3} = \mathbf{S}_{i,0} \mathbf{G}_{OTD3} = [\mathbf{S}_{i,0} | \mathbf{S}_{i,1} | \ldots | \mathbf{S}_{i,n_0-2}]$$

that coincides with a block row of matrix $\mathbf{S}$.

OTD attack strategies

Third strategy (2)

- Since matrix $\mathbf{S}$ has been chosen sparse, the code defined by $\mathbf{G}'_{OTD3}$ contains low weight codewords.
- Such codewords coincide with the rows of $\mathbf{G}'_{OTD3}$ and can be effectively searched through Stern's algorithm.
- Once having found matrix $\mathbf{S}$, a significant part of the secret key can be revealed through the knowledge of $\mathbf{G}'_{\leq k}$.
- Searching for low weight codewords in $\mathbf{G}_{OTD3}$ would require, on average, 2^{32} binary operations.

OTD attack strategies

1. OTD1: $2^{50.3}$
2. OTD2: 2^{36}
3. OTD3: 2^{32}

Flaw

The OTD attack strategies rely on the fact that both $\mathbf{S}$ and $\mathbf{Q}$ are sparse and that $\mathbf{Q}$ has block-diagonal form.

Countermeasure

- They can be countered by adopting dense $\mathbf{S}$ matrices, without altering the remaining system parameters.
- For example, $\mathbf{S}$ could have density about 0.5, with odd weight blocks along the main diagonal, and even weight blocks elsewhere, to ensure non-singularity.

New proposals [Baldi2008]

- With dense $\mathbf{S}$ matrices the eavesdropper cannot obtain $\mathbf{Q}_i$ and $\mathbf{S}_{i,j}$, even knowing $\mathbf{Q}_i\mathbf{S}_{i,j}$.
- To preserve the ability of correcting all the intentional errors, $\mathbf{Q}$ is kept sparse (with row/column weight m).
- The choice of a dense $\mathbf{S}$ influences complexity of the decoding stage, that, however, can be reduced by resorting to efficient computation algorithms for circulant matrices.
- The OTD attacks demonstrate that the choice of $\mathbf{Q}$ in block-diagonal form is weak, so we avoid it in the new versions of the cryptosystem.

Baldi, M., Bodrato, M. and Chiaraluce, F., "A New Analysis of the McEliece Cryptosystem based on QC-LDPC Codes." Proc. SCN 2008, Amalfi, Italy (September 2008) Volume 5229 of LNCS., Springer (2008) 246–262

First new variant

The first new variant adopts almost the same parameters of the previous one:

- $p = 4096$
- $n_0 = 4 \Rightarrow n = 16384$
- $k_0 = 3 \Rightarrow k = 12288 \Rightarrow R = 0.75$
- $d_v = 13$, $m = 7$ and $t' = 27$

- **Q** is obtained by randomly permuting the block rows and columns of a matrix of 4×4 circulant blocks with weight 2, except those along the main diagonal, that have weight 1.
- The absence of the block-diagonal structure in **Q** prevents from attacking each single block, and attacking a whole row or column would require $p\binom{p}{2}^3 \approx 2^{81}$ attempts.
- **S** is dense, with row/column weight $\approx k/2$.

Second new variant

The second new variant adopts an alternative choice of the parameters that ensures increased security:

- $p = 8192$
- $n_0 = 3 \Rightarrow n = 24576$
- $k_0 = 2 \Rightarrow k = 16384 \Rightarrow R = 0.67$
- $d_v = 13,\ m = 11$ and $t' = 40$

- Increased security is achieved at the cost of a slightly decreased transmission rate.
- **Q** is obtained by randomly permuting the block rows and columns of a matrix of 3×3 circulant blocks with weight 4, except those along the main diagonal, that have weight 3.
- Attacking a whole row or column of **Q** would require $\binom{p}{4}^2\binom{p}{3} \approx 2^{131}$ attempts.
- **S** is dense, with row/column weight $\approx k/2$.

Encryption and decryption complexity

- Encryption complexity:

$$C_{enc} = C_{mul}\left(\mathbf{u} \cdot \mathbf{G}'\right) + n.$$

- C_{mul} is reduced through efficient algorithms for polynomial multiplication over $\mathrm{GF}(2)[x]$(Toom-Cook method).
- Decryption complexity:

$$C_{dec} = C_{mul}\left(\mathbf{x} \cdot \mathbf{Q}\right) + C_{SPA} + C_{mul}\left(\mathbf{u}' \cdot \mathbf{S}\right).$$

- C_{SPA} is the number of operations required for LDPC decoding through the sum-product algorithm [Hu2001]:

$$C_{SPA} = I_{ave} \cdot n\left[q\left(8d_v + 12R - 11\right) + d_v\right]$$

with I_{ave} average number of decoding iterations and q number of quantization bits used inside the decoder.

▶ Hu, X.-Y., Eleftheriou, E., Arnold, D.-M., Dholakia, A., "Efficient implementations of the sum-product algorithm for decoding LDPC codes." Proc. IEEE GLOBECOM '01, San Antonio, TX (Nov. 2001) 1036–1036E

Performance and Complexity Parameters

	McEliece (1024, 524) [Canteaut1998]	Niederreiter (1024, 524) [Canteaut1998]	RSA [Canteaut1998] 1024-bit mod. public exp. 17	QC-LDPC McEliece 1	QC-LDPC McEliece 2
Key (bytes)	67072	32750	256	6144	6144
Inform. Bits	524	276	1024	12288	16384
Rate	0.5117	0.5681	1	0.75	0.6667
Enc Ops per bit	514	50	2402	658	776
Dec Ops per bit	5140	7863	738112	4678	8901

- The new versions are secure against the known attacks.
- The lowest work factor is achieved by ISD attacks.
- Such attacks require more than 2^{70} and 2^{80} operations for the two new variants, respectively.
- The McEliece and Niederreiter cryptosystems with their standard parameters reach lower security levels.

Canteaut, A. and Chabaud, F., "A new algorithm for finding minimum-weight words in a linear code: application to McElieces cryptosystem and to narrow-sense BCH codes of length 511." IEEE Trans. Inform. Theory **44** (January 1998) 367-378

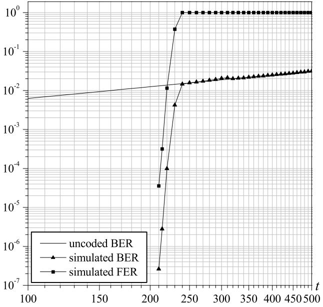
Conclusive remarks

- The adoption of QC-LDPC codes in the McEliece cryptosystem can help overcoming its drawbacks...
- ...but the misuse of sparse transformation matrices can expose the system to total break attacks.
- We have described two new variants of the cryptosystem secure against such attacks.
- They can be seen as a trade-off between the original McEliece cryptosystem and other widespread solutions, like RSA.

An open problem

- The error correction capability of LDPC codes over the "McEliece" channel has been assessed numerically (no guarantee of total error correction).

QC-LDPC(n = 16384, k = 12288)



QC-LDPC(n = 24576, k = 16384)