A Reaction Attack on the QC-LDPC McEliece Cryptosystem

Tomas Fabsic [1], Viliam Hromada [1], Paul Stankovski [2], Pavol Zajac [1], Qian Guo [2], Thomas Johansson [2]

[1]Slovak University of Technology in Bratislava, Slovakia

[2]Lund University, Sweden

PQCrypto 2017

Contents

Contents

Contents

Contents

Contents

Contents

Definitions

Definition

Low-density parity-check (LDPC) code = a binary linear code which admits a parity-check matrix H with a low number of 1s.

Definition

Moderate-density parity-check (MDPC) code - admits a parity-check matrix H with a slightly higher number of 1s than an LDPC code.

Decoding

- Soft-decision decoding (belief propagation algorithms)
- Hard-decision decoding (bit-flipping algorithms)
- Both methods fail with some probability.

Contents

Circulant matrices - definition

Definition

An $n \times n$ matrix C is circulant if it is of the form:

$$C = \begin{pmatrix} c_0 & c_1 & c_2 & \cdots & c_{n-1} \\ c_{n-1} & c_0 & c_1 & \cdots & c_{n-2} \\ c_{n-2} & c_{n-1} & c_0 & \cdots & c_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_1 & c_2 & c_3 & \cdots & c_0 \end{pmatrix}$$

Private Key in QC-MDPC McEliece

- H is a parity-check matrix of an MDPC code.

$$H = (H_0|H_1|\ldots|H_{n_0-1}),$$

where each H_i is a circulant matrix with a low weight. (i.e. H is quasi-cyclic (QC))

How QC-MDPC McEliece works?

- H is randomly generated.
- A generator matrix G is computed.
- G is the public key.
- Encryption of a message x:

$$y = x \cdot G + e,$$

  where e is an error vector.
- Decryption: by a decoding algorithm (uses H).

Contents

- Presented in
  Guo, Johansson and Stankovski: A key recovery attack on MDPC with CCA security using decoding errors, ASIACRYPT 2016.

Distances

Definition

We say that a distance d is present in a vector v of length p if there exist two 1s in v in positions p_1 and p_2 such that

$$d = \min\{p_1 - p_2 \mod p, \quad p_2 - p_1 \mod p\}.$$

E.g., the distance between the 1s in

$$(0, 1, 0, 0, 0, 0, 0, 1, 0)$$

is 3.

Definition

We say that a distance d is present in a $p \times p$ circulant matrix C if the distance d is present in the first row of C.

Key Observation of Guo et al.

- Suppose that the circulant blocks in H are of size $p \times p$.
- Let e be the error vector added to a message during the encryption.
- Let $e = (e^0, e^1, \ldots, e^{n/p-1})$, where each e^i has length p.

Observation

Suppose that e^i contains a distance d. If the distance d is present in the corresponding block H_i in H, then the probability that a bit-flipping algorithm fails to decode the message is lower!

How the attack on QC-MDPC McEliece works?

1. Send a large number of encrypted messages with a randomly generated error vector e.
2. Observe when the recipient requests a message to be resend. (This means that the recipient experienced a decoding error.)
3. Group the encrypted messages into groups Σ_d according to the rule: A message belongs to Σ_d if its error vector contains the distance d in e^0.
4. For each Σ_d estimate the probability of the decoding error.
5. Select the distances with low estimates of the probability of the decoding error. (These are the distances present in H_0.)
6. Reconstruct candidates for H_0.

Contents

Private key in QC-LDPC McEliece

- Private key consists of matrices: H, S, Q.
- All matrices are quasi-cyclic.
- Circulant blocks in all three matrices have the same size $p \times p$.

Private key in QC-LDPC McEliece - matrix H

- H is as in QC-MDPC McEliece but sparser,i.e.

$$H = (H_0|H_1|\dots|H_{n_0-1}),$$

where each H_i is a circulant matrix with a fixed weight.

Private key in QC-LDPC McEliece - matrix Q

- Q is a sparse invertible $n \times n$ matrix.

$$Q = \begin{pmatrix} Q_{00} & \cdots & Q_{0,n_0-1} \\ \vdots & \ddots & \vdots \\ Q_{n_0-1,0} & \cdots & Q_{n_0-1,n_0-1} \end{pmatrix},$$

where each Q_{ij} is a sparse circulant matrix.

Private key in QC-LDPC McEliece - matrix S

- S is a dense invertible $k \times k$ matrix.

$$S = \begin{pmatrix} S_{00} & \cdots & S_{0,k_0-1} \\ \vdots & \ddots & \vdots \\ S_{k_0-1,0} & \cdots & S_{k_0-1,k_0-1} \end{pmatrix},$$

where each S_{ij} is a dense circulant matrix.

Public Key in QC-LDPC McEliece

- H, S, Q are randomly generated.
- A generator matrix G is computed from H.
- Public key G' is computed as:

$$G' = S^{-1} \cdot G \cdot Q^{-1}.$$

Encryption in QC-LDPC McEliece

- Message x is encrypted as:

$$y = x \cdot G' + e,$$

where e is an error vector.

Decryption in QC-LDPC McEliece

1. Compute
$$y' = y \cdot Q.$$
2. Apply an LDPC decoding algorithm (using H) to y'. Denote the result by x'.
3. Compute x as
$$x = x' \cdot S.$$

Contents

Distances in H

- In QC-LDPC McEliece the decoding algorithm is not applied to e, but to $v = eQ$!
- In the QC-MDPC attack the attacker needed to know the distances in the vector to which the decoding algorithm was applied.
- Can the attacker for a given distance d know whether d is present in v?

Distances in H

- Let $e = (e^0, e^1, \ldots, e^{n/p-1})$, where each e^i has length p.
- Let $v = (v^0, v^1, \ldots, v^{n/p-1})$, where each v^i has length p.

Observation

If a distance d is present in e^i, then with a very high probability it will be present in v^j $\forall j$. (Since Q is quasi-cyclic and sparse.)

- Hence, proceeding similarly as in the attack by Guo et al., we can hope to reconstruct candidates for H.
- But the private key also contains Q and S!

Distances in Q

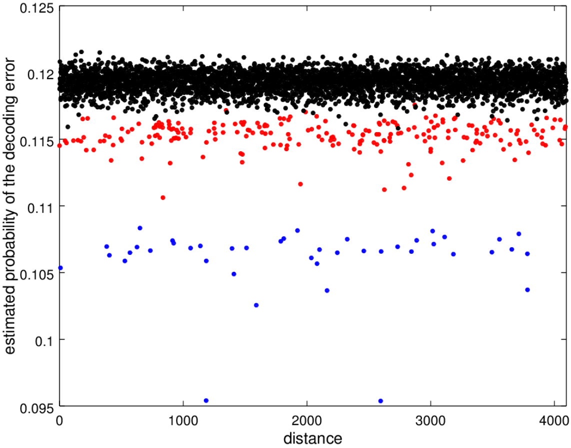
Observation

- *We can learn distances in Q as well!*
- *If a distance d is present in e^i and at the same time it is present in one of the blocks $Q_{i,0}, \ldots, Q_{i,n_0-1}$ in the i-th block-row of Q, then $v = eQ$ has smaller hamming weight then normal.*
- *Smaller hamming weight of $v \Rightarrow$ lower probability of the decoding error.*

Experiment

1. We decrypted a large number of encrypted messages with a randomly generated error vector e.
2. We observed when the decoding error occurred.
3. We grouped the encrypted messages into groups Σ_d according to the rule: A message belongs to Σ_d if its error vector contains the distance d in some e^i.
4. For each Σ_d we estimated the probability of the decoding error.

Experiment results



Learning to decrypt

- If we have candidates for blocks in H and candidates for blocks in Q, we can compute candidates for $\tilde{H}$

$$\tilde{H} = H \times Q^T.$$

- $\tilde{H}$ is a sparse parity check matrix for the public code and can be used for decrypting ciphertexts!

Parameters of the Attacked Cryptosystem

- We used a cryptosystem with parameters for 80-bit security.
- We used messages with a very high number of errors (higher than recommended in the cryptosystem).
- This was done to increase the probability of the decoding error from 10^{-5} to 10^{-1} and thus make it easier to estimate.
- The cryptosystem employed soft-decision decoding. (In Guo et al. hard-decision decoding was used.)

Performance of the Attack

- We considered 2 scenarios:
  - Scenario 1: attacker can choose the error vector.
  - Scenario 2: the error vector was chosen randomly.
- In Scenario 1, we needed 4M decryptions.
- In Scenario 2, we needed 103M decryptions.
- If messages with the recommended number of errors were used, we expect that 10^4 times more decryptions would be needed in each scenario.

Conclusions

1. QC-LDPC McEliece is vulnerable.
2. Soft-decision decoding algorithms are vulnerable.

The End

Thank you for your attention!