Code-based cryptography II

Niederreiter system and schoolbook attaccks

Tanja Lange
with some slides by Tung Chou and Christiane Peters

Eindhoven University of Technology

SAC – Post-quantum cryptography

Systematic form

- A systematic generator matrix is a generator matrix of the form $(I_k|Q)$ where I_k is the $k \times k$ identity matrix and Q is a $k \times (n-k)$ matrix (redundant part).
- Classical decoding is about recovering m from $c = mG$; without errors m equals the first k positions of c.

Systematic form

- A systematic generator matrix is a generator matrix of the form $(I_k|Q)$ where I_k is the $k \times k$ identity matrix and Q is a $k \times (n-k)$ matrix (redundant part).
- Classical decoding is about recovering m from $c = mG$; without errors m equals the first k positions of c.
- Easy to get parity-check matrix from systematic generator matrix, use $H = (Q^{\intercal}|I_{n-k})$.

Systematic form

- A systematic generator matrix is a generator matrix of the form $(I_k|Q)$ where I_k is the $k \times k$ identity matrix and Q is a $k \times (n-k)$ matrix (redundant part).
- Classical decoding is about recovering m from $c = mG$; without errors m equals the first k positions of c.
- Easy to get parity-check matrix from systematic generator matrix, use $H = (Q^\intercal|I_{n-k})$.
  Then
  $$H(\mathbf{m}G)^\intercal = HG^\intercal\mathbf{m}^\intercal = (Q^\intercal|I_{n-k})(I_k|Q)^\intercal\mathbf{m}^\intercal = 0.$$
- Can reduce storage / transmission bandwidth by leaving out the identity matrix part. E.g. for the parity-check matrix:
  $$H = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix}$$
  Any use of H just includes the matrix in the computations.

Different views on decoding

- The syndrome of $\mathbf{x} \in \mathbb{F}_2^n$ is $\mathbf{s} = H\mathbf{x}$.
  Note $H\mathbf{x} = H(\mathbf{c} + \mathbf{e}) = H\mathbf{c} + H\mathbf{e} = H\mathbf{e}$ depends only on $\mathbf{e}$.
- The syndrome decoding problem is to compute $\mathbf{e} \in \mathbb{F}_2^n$ given $\mathbf{s} \in \mathbb{F}_2^{n-k}$ so that $H\mathbf{e} = \mathbf{s}$ and $\mathbf{e}$ has minimal weight.
- Syndrome decoding and (regular) decoding are equivalent:

Different views on decoding

- The syndrome of $\mathbf{x} \in \mathbb{F}_2^n$ is $\mathbf{s} = H\mathbf{x}$.
  Note $H\mathbf{x} = H(\mathbf{c} + \mathbf{e}) = H\mathbf{c} + H\mathbf{e} = H\mathbf{e}$ depends only on $\mathbf{e}$.
- The syndrome decoding problem is to compute $\mathbf{e} \in \mathbb{F}_2^n$ given $\mathbf{s} \in \mathbb{F}_2^{n-k}$ so that $H\mathbf{e} = \mathbf{s}$ and $\mathbf{e}$ has minimal weight.
- Syndrome decoding and (regular) decoding are equivalent:
  To decode $\mathbf{x}$ with syndrome decoder, compute $\mathbf{e}$ from $H\mathbf{x}$, then $\mathbf{c} = \mathbf{x} + \mathbf{e}$.
  To expand syndrome, assume $H = (Q^\intercal | I_{n-k})$.

Different views on decoding

- The syndrome of $\mathbf{x} \in \mathbb{F}_2^n$ is $\mathbf{s} = H\mathbf{x}$.
  Note $H\mathbf{x} = H(\mathbf{c} + \mathbf{e}) = H\mathbf{c} + H\mathbf{e} = H\mathbf{e}$ depends only on $\mathbf{e}$.
- The syndrome decoding problem is to compute $\mathbf{e} \in \mathbb{F}_2^n$ given $\mathbf{s} \in \mathbb{F}_2^{n-k}$ so that $H\mathbf{e} = \mathbf{s}$ and $\mathbf{e}$ has minimal weight.
- Syndrome decoding and (regular) decoding are equivalent:
  To decode $\mathbf{x}$ with syndrome decoder, compute $\mathbf{e}$ from $H\mathbf{x}$, then $\mathbf{c} = \mathbf{x} + \mathbf{e}$.
  To expand syndrome, assume $H = (Q^\intercal | I_{n-k})$.
  Then $\mathbf{x} = (00\ldots0)||\mathbf{s}$ satisfies $\mathbf{s} = H\mathbf{x}$.
- Note that this $\mathbf{x}$ is not a solution to the syndrome decoding problem, unless it has very low weight.

The Niederreiter cryptosystem I

Developed in 1986 by Harald Niederreiter as a variant of the McEliece cryptosystem. This is the schoolbook version.

- Use $n \times n$ permutation matrix P and $n-k \times n-k$ invertible matrix S.
- Public Key: a scrambled parity-check matrix $K = SHP \in \mathbb{F}_2^{(n-k)\times n}$.
- Encryption: The plaintext $\mathbf{e}$ is an n-bit vector of weight t. The ciphertext $\mathbf{s}$ is the $(n-k)$-bit vector

$$\mathbf{s} = K\mathbf{e}.$$

- Decryption: Find a n-bit vector $\mathbf{e}$ with $\mathrm{wt}(\mathbf{e}) = t$ such that $\mathbf{s} = K\mathbf{e}$.
- The passive attacker is facing a t-error correcting problem for the public key, which seems to be random.

The Niederreiter cryptosystem II

- Public Key: a scrambled parity-check matrix $K = SHP$.
- Encryption: The plaintext $\mathbf{e}$ is an n-bit vector of weight t. The ciphertext $\mathbf{s}$ is the $(n-k)$-bit vector

$$\mathbf{s} = K\mathbf{e}.$$

- Decryption using secret key: Compute

$$\begin{aligned} S^{-1}\mathbf{s} &= S^{-1}K\mathbf{e} = S^{-1}(SHP)\mathbf{e} \\ &= H(P\mathbf{e}) \end{aligned}$$

and observe that $\mathrm{wt}(P\mathbf{e}) = t$, because P permutes.
Use efficient syndrome decoder for H to find $\mathbf{e}' = P\mathbf{e}$ and thus $\mathbf{e} = P^{-1}\mathbf{e}'$.

Note on codes

- McEliece proposed to use binary Goppa codes.
  These are still used today.
- Niederreiter described his scheme using Reed-Solomon codes.
  These were broken in 1992 by Sidelnikov and Chestakov.
- More corpses on the way: concatenated codes, Reed-Muller codes, several Algebraic Geometry (AG) codes, Gabidulin codes, several LDPC codes, cyclic codes.
- Some other constructions look OK (for now).
  NIST competition has several entries on QCMDPC codes.

Do not use the schoolbook versions!

Sloppy Alice attacks! 1998 Verheul, Doumen, van Tilborg

- Assume that the decoding algorithm decodes up to t errors, i. e. it decodes $\mathbf{y} = \mathbf{c} + \mathbf{e}$ to $\mathbf{c}$ if $\mathrm{wt}(\mathbf{e}) \leq t$.
- Eve intercepts ciphertext $\mathbf{y} = \mathbf{m}G' + \mathbf{e}$.
  Eve poses as Alice towards Bob and sends him tweaks of $\mathbf{y}$.
  She uses Bob's reactions (success of failure to decrypt) to recover $\mathbf{m}$.
- Assume $\mathrm{wt}(\mathbf{e}) = t$. (Else flip more bits till Bob fails).
- Eve sends $\mathbf{y}_i = \mathbf{y} + \mathbf{e}_i$ for $\mathbf{e}_i$ the i-th unit vector.
  If Bob returns error, position i in $\mathbf{e}$ is 0 (so the number of errors has increased to $t + 1$ and Bob fails).
  Else position i in $\mathbf{e}$ is 1.
- After k steps Eve knows the first k positions of $\mathbf{m}G'$ without error.
  Invert the $k \times k$ submatrix of G' to get $\mathbf{m}$

Sloppy Alice attacks! 1998 Verheul, Doumen, van Tilborg

- Assume that the decoding algorithm decodes up to t errors, i. e. it decodes $\mathbf{y} = \mathbf{c} + \mathbf{e}$ to $\mathbf{c}$ if $\mathrm{wt}(\mathbf{e}) \leq t$.
- Eve intercepts ciphertext $\mathbf{y} = \mathbf{m}G' + \mathbf{e}$.
  Eve poses as Alice towards Bob and sends him tweaks of $\mathbf{y}$.
  She uses Bob's reactions (success of failure to decrypt) to recover $\mathbf{m}$.
- Assume $\mathrm{wt}(\mathbf{e}) = t$. (Else flip more bits till Bob fails).
- Eve sends $\mathbf{y}_i = \mathbf{y} + \mathbf{e}_i$ for $\mathbf{e}_i$ the i-th unit vector.
  If Bob returns error, position i in $\mathbf{e}$ is 0 (so the number of errors has increased to $t + 1$ and Bob fails).
  Else position i in $\mathbf{e}$ is 1.
- After k steps Eve knows the first k positions of $\mathbf{m}G'$ without error.
  Invert the $k \times k$ submatrix of G' to get $\mathbf{m}$ assuming it is invertible.
- Proper attack: figure out invertible submatrix of G' at beginning; recover matching k coordinates.

More on sloppy Alice

- This attack has Eve send Bob variations of the same ciphertext; so Bob will think that Alice is sloppy.
- Note, this is more complicated if $\mathbb{F}_q$ instead of $\mathbb{F}_2$ is used.
- Other name: reaction attack.
  (1999 Hall, Goldberg, and Schneier)
- Attack also works on Niederreiter version:

More on sloppy Alice

- This attack has Eve send Bob variations of the same ciphertext; so Bob will think that Alice is sloppy.
- Note, this is more complicated if $\mathbb{F}_q$ instead of $\mathbb{F}_2$ is used.
- Other name: reaction attack.
  (1999 Hall, Goldberg, and Schneier)
- Attack also works on Niederreiter version:
  Bitflip corresponds to sending $\mathbf{s}_i = \mathbf{s} + K_i$,
  where K_i is the i-th column of K.
- More involved but doable (for McEliece and Niederreiter) if decryption requires exactly t errors.

Berson's attack

- Eve knows $\mathbf{y}_1 = \mathbf{m}G' + \mathbf{e}_1$ and $\mathbf{y}_2 = \mathbf{m}G' + \mathbf{e}_2$; these have the same $\mathbf{m}$.

Berson's attack

- Eve knows $\mathbf{y}_1 = \mathbf{m}G' + \mathbf{e}_1$ and $\mathbf{y}_2 = \mathbf{m}G' + \mathbf{e}_2$;
  these have the same $\mathbf{m}$.
- Then $\mathbf{y}_1 + \mathbf{y}_2 = \mathbf{e}_1 + \mathbf{e}_2 = \bar{\mathbf{e}}$. This has weight in $[0, 2t]$.
- If $\mathrm{wt}(\bar{\mathbf{e}}) = 2t$:

Berson's attack

- Eve knows $\mathbf{y}_1 = \mathbf{m}G' + \mathbf{e}_1$ and $\mathbf{y}_2 = \mathbf{m}G' + \mathbf{e}_2$;
  these have the same $\mathbf{m}$.
- Then $\mathbf{y}_1 + \mathbf{y}_2 = \mathbf{e}_1 + \mathbf{e}_2 = \bar{\mathbf{e}}$. This has weight in $[0, 2t]$.
- If $\mathrm{wt}(\bar{\mathbf{e}}) = 2t$:
  All zero positions in $\bar{\mathbf{e}}$ are error free in both ciphertexts.
  Invert G' in those columns to recover $\mathbf{m}$ as in previous attack.
- Else:

Berson's attack

- Eve knows $\mathbf{y}_1 = \mathbf{m}G' + \mathbf{e}_1$ and $\mathbf{y}_2 = \mathbf{m}G' + \mathbf{e}_2$; these have the same $\mathbf{m}$.
- Then $\mathbf{y}_1 + \mathbf{y}_2 = \mathbf{e}_1 + \mathbf{e}_2 = \bar{\mathbf{e}}$. This has weight in $[0, 2t]$.
- If $\mathrm{wt}(\bar{\mathbf{e}}) = 2t$:
  All zero positions in $\bar{\mathbf{e}}$ are error free in both ciphertexts.
  Invert G' in those columns to recover $\mathbf{m}$ as in previous attack.
- Else: ignore the $2w = \mathrm{wt}(\bar{\mathbf{e}}) < 2t$ positions in G' and $\mathbf{y}_1$.
  Solve decoding problem for $k \times (n - 2w)$ generator matrix G'' and vector $\mathbf{y}_1'$ with $t - w$ errors; typically much easier.

Formal security notions

- McEliece/Niederreiter are One-Way Encryption (OWE) schemes.
- However, the schemes as presented are not CCA–II secure:
  - Given challenge $\mathbf{y} = \mathbf{m}G' + \mathbf{e}$, Eve can ask for decryptions of anything but $\mathbf{y}$.

Formal security notions

- McEliece/Niederreiter are One-Way Encryption (OWE) schemes.
- However, the schemes as presented are not CCA–II secure:
  - Given challenge $\mathbf{y} = \mathbf{m}G' + \mathbf{e}$, Eve can ask for decryptions of anything but $\mathbf{y}$.
  - Eve picks a random code word $\mathbf{c} = \bar{\mathbf{m}}G'$, asks for decryption of $\mathbf{y} + \mathbf{c}$.
  - This is different from challenge $\mathbf{y}$, so Bob answers.

Formal security notions

- McEliece/Niederreiter are One-Way Encryption (OWE) schemes.
- However, the schemes as presented are not CCA–II secure:
  - Given challenge $\mathbf{y} = \mathbf{m}G' + \mathbf{e}$, Eve can ask for decryptions of anything but $\mathbf{y}$.
  - Eve picks a random code word $\mathbf{c} = \bar{\mathbf{m}}G'$,
    asks for decryption of $\mathbf{y} + \mathbf{c}$.
  - This is different from challenge $\mathbf{y}$, so Bob answers.
  - Answer is $\mathbf{m} + \bar{\mathbf{m}}$.
- Fix by using CCA2 transformation (e.g. Fujisaki-Okamoto transform) or (easier) KEM/DEM version:
  pick random $\mathbf{e}$ of weight t, use hash($\mathbf{e}$) as secret key to encrypt and authenticate (for McEliece or Niederreiter).