Code-Based Cryptography

Message Attacks (ISD)

Nicolas Sendrier

Code-Based Cryptography

1. Error-Correcting Codes and Cryptography
2. McEliece Cryptosystem
3. **Message Attacks (ISD)**
4. Key Attacks
5. Other Cryptographic Constructions Relying on Coding Theory

3. Message Attack (ISD)

1. **From Generic Decoding to Syndrome Decoding**
2. Combinatorial Solutions: Exhaustive Search and Birthday Decoding
3. Information Set Decoding: the Power of Linear Algebra
4. Complexity Analysis
5. Lee and Brickell Algorithm
6. Stern/Dumer Algorithm
7. May, Meurer, and Thomae Algorithm
8. Becker, Joux, May, and Meurer Algorithm
9. Generalized Birthday Algorithm for Decoding
10. Decoding One Out of Many

Message Attack

A cryptogram for the McEliece encryption scheme has the following form

$$y = xG + e$$

Message Attack

A cryptogram for the McEliece encryption scheme has the following form

$$y = xG + e$$

cryptogram
(or ciphertext)

Message Attack

A cryptogram for the McEliece encryption scheme has the following form

$$y = xG + e$$

cryptogram
(or ciphertext)

message
(or cleartext)

Message Attack

A cryptogram for the McEliece encryption scheme has the following form

$$y = xG + e$$

cryptogram (or ciphertext) → y

message (or cleartext) → x

public key → G

Message Attack

A cryptogram for the McEliece encryption scheme has the following form

$$y = xG + e$$

cryptogram (or ciphertext) — message (or cleartext) — public key — random error

Message Attack

A cryptogram for the McEliece encryption scheme has the following form

$$y = xG + e$$

- y: cryptogram (or ciphertext)
- x: message (or cleartext)
- G: public key
- e: random error

The adversary knows the cryptogram and the public key
and wishes to recover the message (or equivalently the error)

Message Attack

A cryptogram for the McEliece encryption scheme has the following form

$$y = xG + e$$

- y: cryptogram (or ciphertext)
- x: message (or cleartext)
- G: public key
- e: random error

The adversary knows the cryptogram and the public key
and wishes to recover the message (or equivalently the error)

Only an arbitrary generator matrix is known

$\rightarrow$ **generic decoding problem**

Generic Decoding

In contrast with the usual situation where the code is known in advance, a generic decoder takes a q-ary linear $[n, k]$ code as argument

Generic Decoding

In contrast with the usual situation where the code is known in advance, a generic decoder takes a q-ary linear $[n, k]$ code as argument

$G \in \mathbf{F}_q^{k \times n}$ a generator matrix

$\mathcal{C} = \langle G \rangle = \{xG \mid x \in \mathbf{F}_q^k\}$

Generic Decoding

In contrast with the usual situation where the code is known in advance, a generic decoder takes a q-ary linear $[n, k]$ code as argument

$G \in \mathbf{F}_q^{k \times n}$ a generator matrix

$\mathcal{C} = \langle G \rangle = \{xG \mid x \in \mathbf{F}_q^k\}$

$H \in \mathbf{F}_q^{(n-k) \times n}$ a parity check matrix

$\mathcal{C} = \langle H \rangle^\perp = \{c \in \mathbf{F}_q^n \mid cH^T = 0\}$

Generic Decoding

In contrast with the usual situation where the code is known in advance, a generic decoder takes a q-ary linear $[n, k]$ code as argument

$G \in \mathbf{F}_q^{k \times n}$ a generator matrix

$\mathcal{C} = \langle G \rangle = \{xG \mid x \in \mathbf{F}_q^k\}$

$H \in \mathbf{F}_q^{(n-k) \times n}$ a parity check matrix

$\mathcal{C} = \langle H \rangle^\perp = \{c \in \mathbf{F}_q^n \mid cH^T = 0\}$

Generic Decoder:

Generic Decoding

In contrast with the usual situation where the code is known in advance, a generic decoder takes a q-ary linear $[n, k]$ code as argument

$G \in \mathbf{F}_q^{k \times n}$ a generator matrix

$\mathcal{C} = \langle G \rangle = \{xG \mid x \in \mathbf{F}_q^k\}$

$H \in \mathbf{F}_q^{(n-k) \times n}$ a parity check matrix

$\mathcal{C} = \langle H \rangle^\perp = \{c \in \mathbf{F}_q^n \mid cH^T = 0\}$

Generic Decoder:

$$\begin{array}{rccc} \Phi : & \mathbf{F}_q^n \times \mathbf{F}_q^{k \times n} & \to & \mathbf{F}_q^k \\ & (y, G) & \mapsto & x \end{array}$$

Generic Decoding

In contrast with the usual situation where the code is known in advance, a generic decoder takes a q-ary linear $[n, k]$ code as argument

$G \in \mathbf{F}_q^{k \times n}$ a generator matrix

$\mathcal{C} = \langle G \rangle = \{xG \mid x \in \mathbf{F}_q^k\}$

$H \in \mathbf{F}_q^{(n-k) \times n}$ a parity check matrix

$\mathcal{C} = \langle H \rangle^{\perp} = \{c \in \mathbf{F}_q^n \mid cH^T = 0\}$

Generic Decoder:

$$\Phi : \quad \mathbf{F}_q^n \times \mathbf{F}_q^{k \times n} \quad \to \quad \mathbf{F}_q^k$$

$\Phi(xG + e, G) = x$ if e is “small”

“small” = of small Hamming weight

Generic Decoding

In contrast with the usual situation where the code is known in advance, a generic decoder takes a q-ary linear $[n, k]$ code as argument

$G \in \mathbf{F}_q^{k \times n}$ a generator matrix

$\mathcal{C} = \langle G \rangle = \{xG \mid x \in \mathbf{F}_q^k\}$

$H \in \mathbf{F}_q^{(n-k) \times n}$ a parity check matrix

$\mathcal{C} = \langle H \rangle^\perp = \{c \in \mathbf{F}_q^n \mid cH^T = 0\}$

Generic Decoder:

$$\Phi : \quad \mathbf{F}_q^n \times \mathbf{F}_q^{k \times n} \quad \to \quad \mathbf{F}_q^k$$

$\Phi(xG + e, G) = x$ if e is "small"

Generic Syndrome Decoder:

$$\begin{array}{rccc} \Psi : & \mathbf{F}_q^{n-k} \times \mathbf{F}_q^{(n-k) \times n} & \to & \mathbf{F}_q^n \\ & (s, H) & \mapsto & e \end{array}$$

Generic Decoding

In contrast with the usual situation where the code is known in advance, a generic decoder takes a q-ary linear $[n, k]$ code as argument

$G \in \mathbf{F}_q^{k\times n}$ a generator matrix

$\mathcal{C} = \langle G \rangle = \{xG \mid x \in \mathbf{F}_q^k\}$

$H \in \mathbf{F}_q^{(n-k)\times n}$ a parity check matrix

$\mathcal{C} = \langle H \rangle^\perp = \{c \in \mathbf{F}_q^n \mid cH^T = 0\}$

Generic Decoder:

$$\Phi : \quad \mathbf{F}_q^n \times \mathbf{F}_q^{k\times n} \quad \rightarrow \quad \mathbf{F}_q^k$$

$\Phi(xG + e, G) = x$ if e is “small”

Generic Syndrome Decoder:

$$\Psi : \quad \mathbf{F}_q^{n-k} \times \mathbf{F}_q^{(n-k)\times n} \quad \rightarrow \quad \mathbf{F}_q^n$$

$\Psi(eH^T, H) = e$ if e is “small”

Generic Decoding

In contrast with the usual situation where the code is known in advance, a generic decoder takes a q-ary linear $[n, k]$ code as argument

$G \in \mathbf{F}_q^{k \times n}$ a generator matrix

$\mathcal{C} = \langle G \rangle = \{xG \mid x \in \mathbf{F}_q^k\}$

$H \in \mathbf{F}_q^{(n-k) \times n}$ a parity check matrix

$\mathcal{C} = \langle H \rangle^{\perp} = \{c \in \mathbf{F}_q^n \mid cH^T = 0\}$

Generic Decoder:

$$\Phi : \quad \mathbf{F}_q^n \times \mathbf{F}_q^{k \times n} \quad \rightarrow \quad \mathbf{F}_q^k$$

$\Phi(xG + e, G) = x$ if e is “small”

Generic Syndrome Decoder:

$$\Psi : \quad \mathbf{F}_q^{n-k} \times \mathbf{F}_q^{(n-k) \times n} \quad \rightarrow \quad \mathbf{F}_q^n$$

$\Psi(eH^T, H) = e$ if e is “small”

Those two kinds of decoders are equivalent

$\rightarrow$ **we will consider only syndrome decoding**

The Syndrome Decoding Problem

Syndrome Decoding Problem

Instance: $H \in \{0,1\}^{(n-k)\times n}$, $s \in \{0,1\}^{n-k}$, an integer $w > 0$

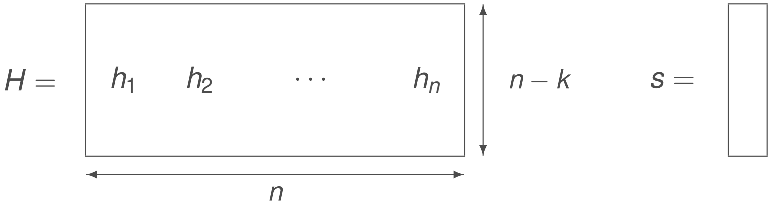
Answer: $e \in \{0,1\}^n$ such that $eH^T = s$ and $\mathrm{wt}(e) \leq w$

The Syndrome Decoding Problem

Syndrome Decoding Problem NP-hard

Instance: $H \in \{0,1\}^{(n-k)\times n}$, $s \in \{0,1\}^{n-k}$, an integer $w > 0$
Answer: $e \in \{0,1\}^n$ such that $eH^T = s$ and $\mathrm{wt}(e) \le w$

Find w columns of H adding to s (modulo 2)



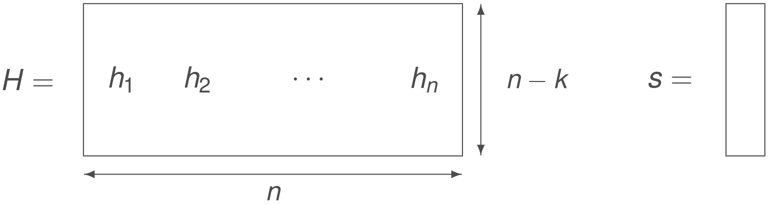
The Syndrome Decoding Problem

Syndrome Decoding Problem — NP-hard

Instance: $H \in \{0,1\}^{(n-k)\times n}$, $s \in \{0,1\}^{n-k}$, an integer $w > 0$

Answer: $e \in \{0,1\}^n$ such that $eH^T = s$ and $\mathrm{wt}(e) \leq w$

Find w columns of H adding to s (modulo 2)



Find $1 \leq j_1 < j_2 < \cdots < j_w \leq n$ such that

$$h_{j_1} + h_{j_2} + \cdots + h_{j_w} = s$$

Single Solution *versus* Multiple Solution

Syndrome Decoding Problem

Instance: $H \in \{0,1\}^{(n-k)\times n}$, $s \in \{0,1\}^{n-k}$, an integer $w > 0$
Answer: $e \in \{0,1\}^n$ such that $eH^T = s$ and $\mathrm{wt}(e) \le w$

Single Solution *versus* Multiple Solution

Syndrome Decoding Problem

Instance: $H \in \{0,1\}^{(n-k)\times n}$, $s \in \{0,1\}^{n-k}$, an integer $w > 0$
Answer: $e \in \{0,1\}^n$ such that $eH^T = s$ and $\mathrm{wt}(e) \le w$

We denote $\mathrm{CSD}(H, s, w)$ the set of all solutions to the above problem

Single Solution *versus* Multiple Solution

Syndrome Decoding Problem

Instance: $H \in \{0,1\}^{(n-k)\times n}$, $s \in \{0,1\}^{n-k}$, an integer $w > 0$
Answer: $e \in \{0,1\}^n$ such that $eH^T = s$ and $\mathrm{wt}(e) \le w$

We denote $\mathrm{CSD}(H, s, w)$ the set of all solutions to the above problem
Fix n and k and let w grow

$\longrightarrow$ $\dfrac{\binom{n}{w}}{2^{n-k}}$ solutions on average

Single Solution *versus* Multiple Solution

Syndrome Decoding Problem

Instance:	$H \in \{0,1\}^{(n-k)\times n}$, $s \in \{0,1\}^{n-k}$, an integer $w > 0$
Answer:	$e \in \{0,1\}^n$ such that $eH^T = s$ and $\mathrm{wt}(e) \leq w$

We denote $\mathrm{CSD}(H, s, w)$ the set of all solutions to the above problem
Fix n and k and let w grow

$\longrightarrow \dfrac{\binom{n}{w}}{2^{n-k}}$ solutions on average

Gilbert-Varshamov radius $\binom{n}{\tau_{\mathrm{GV}}} = 2^{n-k}$

Single Solution *versus* Multiple Solution

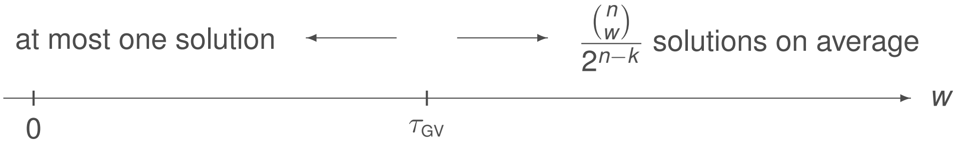
Syndrome Decoding Problem

Instance: $H \in \{0,1\}^{(n-k)\times n}$, $s \in \{0,1\}^{n-k}$, an integer $w > 0$

Answer: $e \in \{0,1\}^n$ such that $eH^T = s$ and $\mathrm{wt}(e) \leq w$

We denote $\mathrm{CSD}(H,s,w)$ the set of all solutions to the above problem

Fix n and k and let w grow



Gilbert-Varshamov radius $\binom{n}{\tau_{\mathrm{GV}}} = 2^{n-k}$

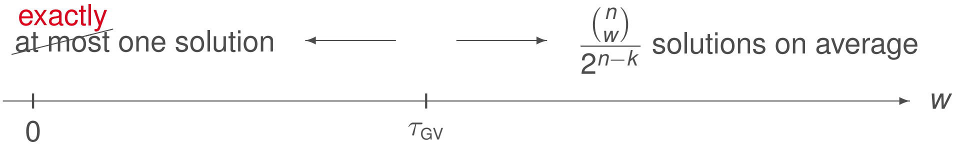
Single Solution *versus* Multiple Solution

Syndrome Decoding Problem

Instance: $H \in \{0,1\}^{(n-k)\times n}$, $s \in \{0,1\}^{n-k}$, an integer $w > 0$
Answer: $e \in \{0,1\}^n$ such that $eH^T = s$ and $\mathrm{wt}(e) \le w$

We denote $\mathrm{CSD}(H, s, w)$ the set of all solutions to the above problem

Fix n and k and let w grow



Gilbert-Varshamov radius $\binom{n}{\tau_{\mathrm{GV}}} = 2^{n-k}$

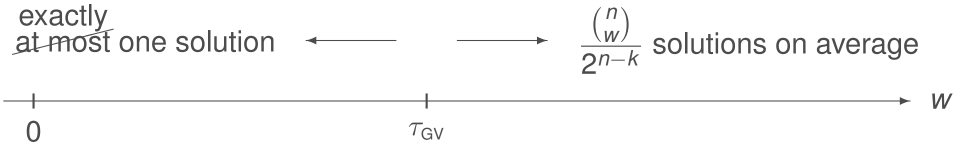
In cryptanalysis, we only consider situations where $\mathrm{CSD}(H, s, w) \neq \emptyset$

Single Solution *versus* Multiple Solution

Syndrome Decoding Problem

Instance: $H \in \{0,1\}^{(n-k)\times n}$, $s \in \{0,1\}^{n-k}$, an integer $w > 0$
Answer: $e \in \{0,1\}^n$ such that $eH^T = s$ and $\mathrm{wt}(e) \leq w$

We denote $\mathrm{CSD}(H, s, w)$ the set of all solutions to the above problem
Fix n and k and let w grow



Gilbert-Varshamov radius $\binom{n}{\tau_{\mathrm{GV}}} = 2^{n-k}$

In cryptanalysis, we only consider situations where $\mathrm{CSD}(H, s, w) \neq \emptyset$

We expect $\approx \max\left(1, \binom{n}{w}/2^{n-k}\right)$ solutions

3. Message Attack (ISD)

1. From Generic Decoding to Syndrome Decoding
2. **Combinatorial Solutions: Exhaustive Search and Birthday Decoding**
3. Information Set Decoding: the Power of Linear Algebra
4. Complexity Analysis
5. Lee and Brickell Algorithm
6. Stern/Dumer Algorithm
7. May, Meurer, and Thomae Algorithm
8. Becker, Joux, May, and Meurer Algorithm
9. Generalized Birthday Algorithm for Decoding
10. Decoding One Out of Many