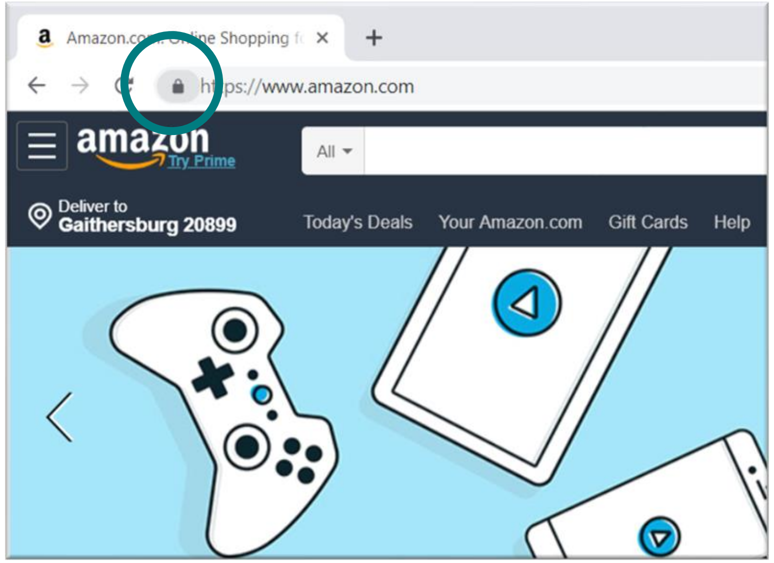
Code-based Cryptography

Angela Robinson

BRIDGES Conference, June 7, 2022

Motivation

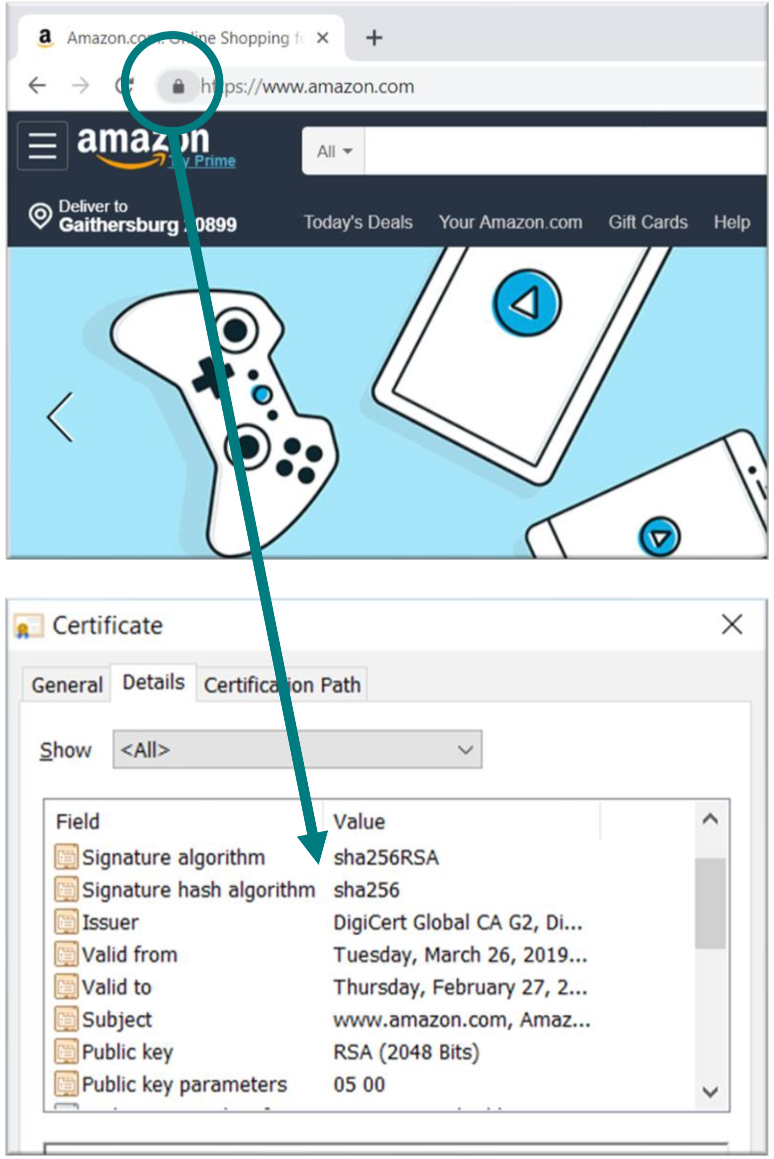
Cryptography sightings



Cryptography sightings

Secure websites are protected using cryptography

- Encryption – confidentiality of messages
- Digital signature – authentication
- Certificates – verify identity



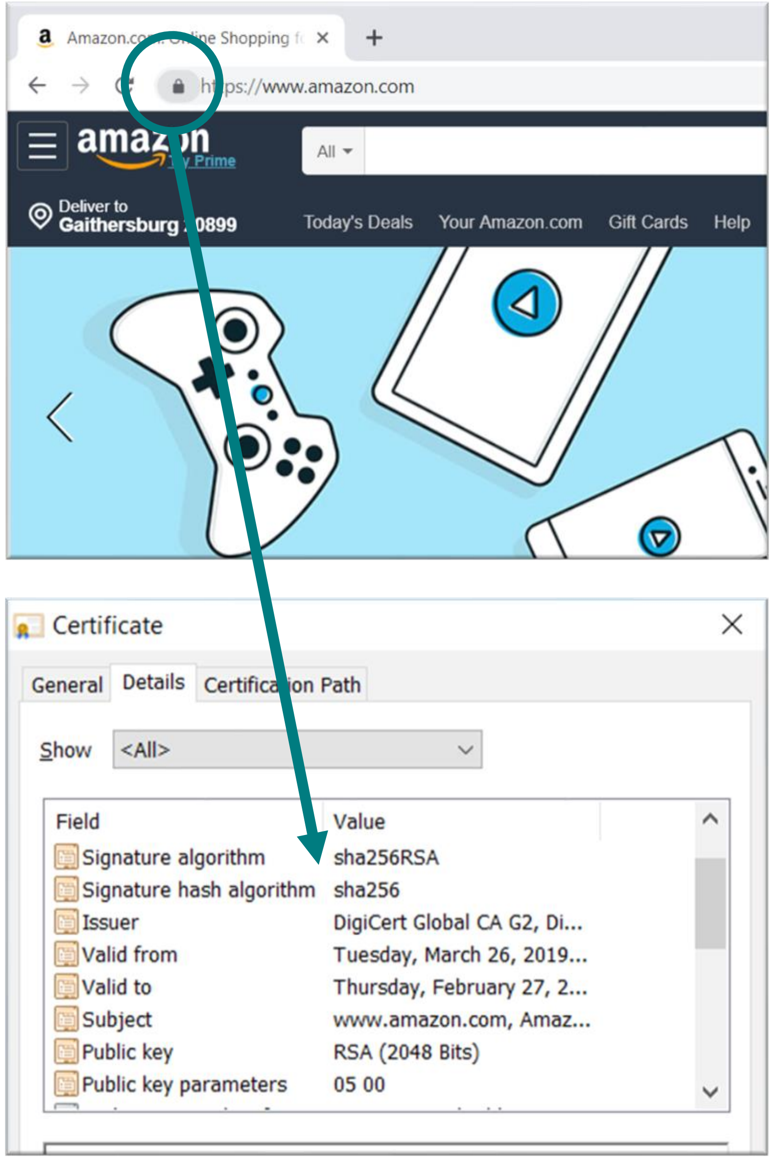
Cryptography sightings

Secure websites are protected using cryptography

- Encryption – confidentiality of messages
- Digital signature – authentication
- Certificates – verify identity

Security is quantified by the resources it takes to break a cryptosystem

- Best known cryptanalysis
- Cost of implementing the cryptanalysis



Cryptography at NIST

National Institute of Standards and Technology
U.S. Department of Commerce

Cryptographic Standards

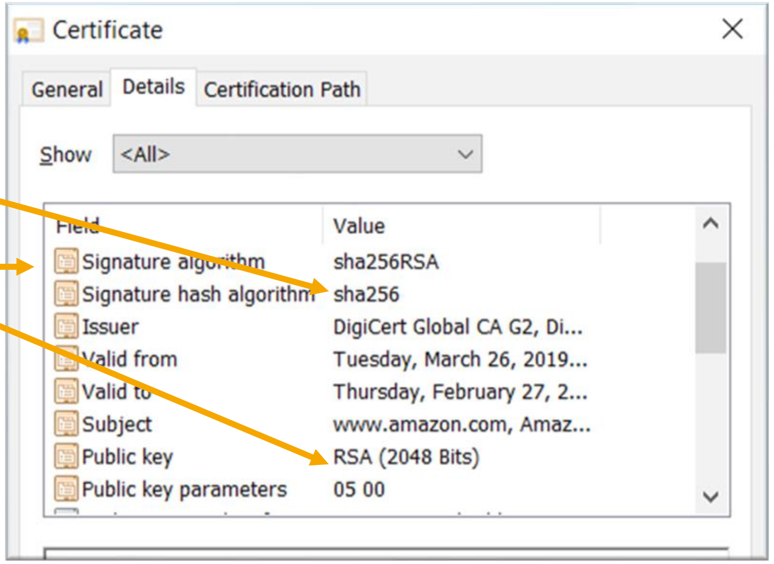
- Hash functions
- Encryption schemes
- Digital signatures
- …

Cryptography at NIST

Cryptographic Standards

- Hash functions
- Encryption schemes
- Digital signatures
- …

Example



Present threat

Some current NIST standards are vulnerable to quantum threat.

Peter Shor (1994): polynomial-time quantum algorithm that breaks

- Integer factorization problem (RSA)
- Discrete logarithm problem (Diffie-Hellman Key Exchange, Elliptic Curve DH, …)
- Impact: a full-scale quantum computer can break today's public key crypto

Options for mitigating the threat

- ~~Stop using public key crypto~~ not practical
- Find quantum-safe public key crypto

NIST PQC Standardization effort

Call for public key cryptographic schemes believed to be quantum-resistant (2016)

- Received 80+ submissions (2017)
- Only 15 submissions are still under consideration (2022)
- Code-based algorithms
  - Round 2: BIKE, Classic McEliece*, HQC, LEDAcrypt**, NTS-KEM*
  - Round 3: BIKE, Classic McEliece, HQC

*merged during Round 2

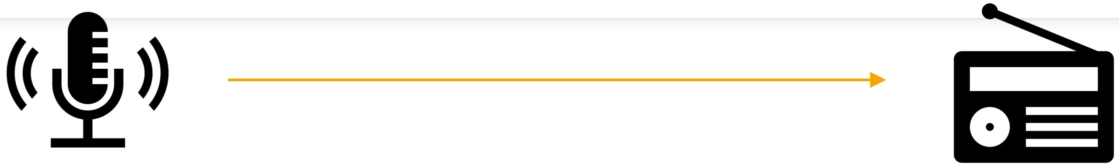
** broken [APRS2020]

Background

Error-correcting codes

Noisy channels

Messages are sent over various channels



- Analog
  - Compact disks, DVDs
  - Radio
  - Telephone
- Digital

Environmental noise can distort or alter the message before it is received

Error-correcting codes

Error-detecting and error-correcting codes are designed to locate and remove noise from messages received over noisy channels

Error-correcting codes

Error-detecting and error-correcting codes are designed to locate and remove noise from messages received over noisy channels

This is accomplished by adding some **extra bits** to the message before transmission that will enable error-detection and error-correction

Error-correcting codes

Error-detecting and error-correcting codes are designed to locate and remove noise from messages received over noisy channels

This is accomplished by adding some **extra bits** to the message before transmission that will enable error-detection and error-correction

Example: Repetition code. Consider message 1001001

Repetition code

Example: Repetition code. Consider message 1001001

1001001 1001001 1001001 → Noisy channel → 1001101 1001001 0001001

1. Sender sends 3 copies of the message
2. Receiver decodes by taking most frequent bit for each position

Repetition code

Example: Repetition code. Consider message 1001001

1001001 1001001 1001001 → Noisy channel → 1001101 1001001 0001001

1. Sender sends 3 copies of the message
2. Receiver decodes by taking most frequent bit for each position

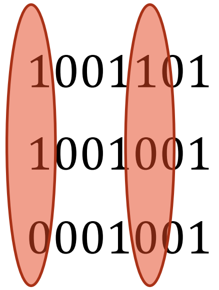
1001101
1001001
0001001

Repetition code

Example: Repetition code. Consider message 1001001

1001001 1001001 1001001 → Noisy channel → 1001101 1001001 0001001

1. Sender sends 3 copies of the message
2. Receiver decodes by taking most frequent bit for each position
3. Receiver recovers 1001001

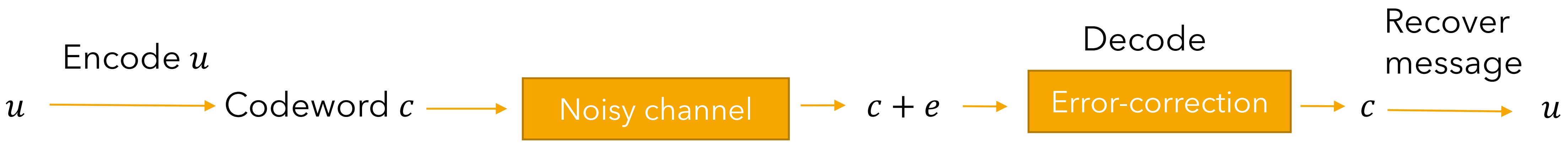


Disadvantages?

Error-correcting codes

Error-detecting and error-correcting codes are designed to locate and remove noise from messages received over noisy channels

This is accomplished by adding some **extra bits** to the message before transmission that will enable error-detection and error-correction



Definitions

Definition: a **vector space** over a field $\mathbb{F}$ consists of a set V (of vectors) and a set $\mathbb{F}$ (of scalars) along with operations $+$ and $\cdot$ such that

- If $x, y \in V$, then $x + y \in V$
- If $x \in V$ and $\alpha \in \mathbb{F}$, then $\alpha \cdot x \in V$

Definitions

Definition: a **vector space** over a field $\mathbb{F}$ consists of a set V (of vectors) and a set $\mathbb{F}$ (of scalars) along with operations $+$ and $\cdot$ such that

- If $x, y \in V$, then $x + y \in V$
- If $x \in V$ and $\alpha \in \mathbb{F}$, then $\alpha \cdot x \in V$

Definition: Let V be a vector space. A linearly independent spanning set B for V is called a **basis.**

Definition: The **dimension** of a vector space is the cardinality of its bases

Definitions

Definition: a **vector space** over a field $\mathbb{F}$ consists of a set V (of vectors) and a set $\mathbb{F}$ (of scalars) along with operations $+$ and $\cdot$ such that

- If $x, y \in V$, then $x + y \in V$
- If $x \in V$ and $\alpha \in \mathbb{F}$, then $\alpha \cdot x \in V$

Definition: Let V be a vector space. A linearly independent spanning set B for V is called a **basis.**

Definition: The **dimension** of a vector space is the cardinality of its bases

Example: $\mathbb{R}^3$ is a vector space, $B = \{1 \quad 0 \quad 0, \quad 0 \quad 1 \quad 0\ , \quad 0 \quad 0 \quad 1\}$ is the standard basis for $\mathbb{R}^3$

$\dim(\mathbb{R}^3) = 3.$

Definitions

$\mathbb{F}_2$ - finite field of two elements

denote the additive identity by $\mathbf{0}$

denote the multiplicative identity by $\mathbf{1}$

$\mathbb{F}_2^n$ - vector space over $\mathbb{F}_2$

elements are vectors of length $\boldsymbol{n}$ whose components are from $\mathbb{F}_2$

standard basis: $\begin{cases} 1\ 0\ 0\ 0\ \ldots 0 \\ 0\ 1\ 0\ 0\ \ldots 0 \\ \quad\quad\vdots \\ 0\ 0\ 0\ 0\ \ldots 1 \end{cases}$

scalars $\{0, 1\}$

Binary linear code

Definition: a **binary linear code** $C(n, k)$ is a k-dimensional subspace of $\mathbb{F}_2^n$.

The code $C: \mathbb{F}_2^k \to \mathbb{F}_2^n$ maps information vectors to codewords

Binary linear code

Definition: a **binary linear code** $C(n, k)$ is a k-dimensional subspace of $\mathbb{F}_2^n$.

The code $C: \mathbb{F}_2^k \to \mathbb{F}_2^n$ maps information vectors to codewords

How do we describe a code?

Binary linear code

Definition: a **binary linear code** $C(n,k)$ is a k-dimensional subspace of $\mathbb{F}_2^n$.

The code $C: \mathbb{F}_2^k \to \mathbb{F}_2^n$ maps information vectors to codewords

How do we describe a code?

1. Select a basis of the k-dim vector space $\{g_0, g_1, \dots, g_{k-1}\}$
2. Basis forms a **generator matrix** $\boldsymbol{G_{k\times n}}$ of the code

$$G = \begin{bmatrix} g_{0,0} & \cdots & g_{0,n-1} \\ \vdots & \ddots & \vdots \\ g_{k-1,0} & \cdots & g_{k-1,n-1} \end{bmatrix}$$

Descriptions of a code $C(n, k)$

Two equivalent descriptions of $C(n, k)$

- Generator matrix
  - Encoding: multiply k-bit information word u by G
  - codewords are x such that there's a solution u to $uG = x$

Encode u

u → uG is codeword c

Descriptions of a code $C(n, k)$

Two equivalent descriptions of $C(n, k)$

- Generator matrix
  - Encoding: multiply k-bit information word u by G
  - codewords are x such that there's a solution u to $uG = x$
- Parity-check matrix H (dimension $(n-k)\ x\ n$)
  - $GH^T = 0$
  - codewords are x such that $Hx^T = 0$
  - Product of generic n-bit vector with H^T is called a syndrome

Encode u

$u \longrightarrow uG$ is codeword c

Descriptions of a code $C(n, k)$

Parity-check matrix H (dimension $(n-k)\ x\ n$)

- $GH^T = 0$
- codewords are $\boldsymbol{x}$ such that $Hx^T = 0$
- Product of generic $\boldsymbol{n}$-bit vector with H^T is called a syndrome

Example: Let H, x_1, x_2 be as follows.

$$H = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

$$x_1 = [0 \quad 0 \quad 1 \quad 0 \quad 0 \quad 1]$$

$$x_2 = [1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0]$$

Descriptions of a code $C(n,k)$

Parity-check matrix H (dimension $(n-k)\ x\ n$)

- $GH^T = 0$
- codewords are $\boldsymbol{x}$ such that $H\boldsymbol{x}^T = 0$
- Product of generic $\boldsymbol{n}$-bit vector with H^T is called a syndrome

Example: Let H, x_1, x_2 be as follows.

$$H = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

$$x_1 = [0 \quad 0 \quad 1 \quad 0 \quad 0 \quad 1]$$

$$x_2 = [1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0]$$

$$Hx_1^T = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

Descriptions of a code $C(n,k)$

Parity-check matrix H (dimension $(n-k)\ x\ n$)

- $GH^T = 0$
- codewords are $\boldsymbol{x}$ such that $Hx^T = 0$
- Product of generic $\boldsymbol{n}$-bit vector with H^T is called a syndrome

Example: Let H, x_1, x_2 be as follows.

$$H = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

$$x_1 = [0 \quad 0 \quad 1 \quad 0 \quad 0 \quad 1]$$

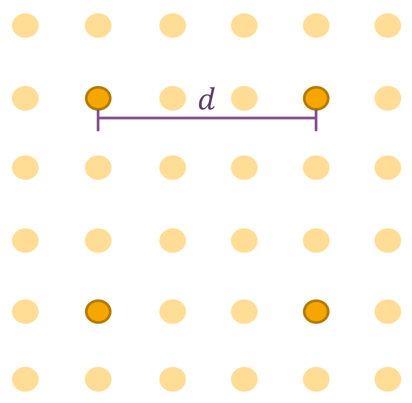
$$x_2 = [1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0]$$

$$Hx_1^T = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

Syndrome is nonzero, so x_1 is not in the code defined by H.

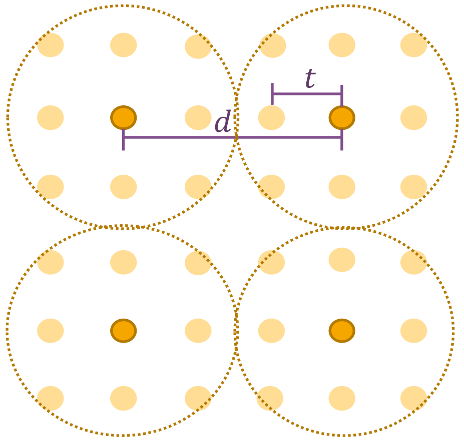
Error correction

Definition: A linear (n, k, d)-code C over a finite field $\mathbb{F}$ is a k-dimensional subspace of $\mathbb{F}^n$ with **minimum distance** $d = min_{x \neq y \epsilon C} dist(x, y)$, where $dist$ is the Hamming distance.



Error correction

Definition: A linear (n, k, d)-code C over a finite field $\mathbb{F}$ is a k-dimensional subspace of $\mathbb{F}^n$ with **minimum distance** $d = min_{x \neq y \epsilon C} dist(x, y)$, where $dist$ is the Hamming distance.



Please excuse visual imperfections

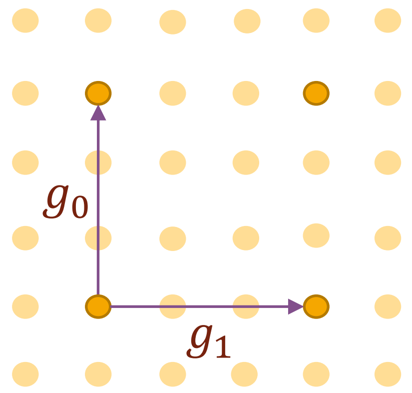
Theorem.

A linear (n, k, d)-code C can correct up to $t = \left\lfloor \frac{d-1}{2} \right\rfloor$ errors.

Visual recap

Generator matrix formed by basis vectors

Code is closed under addition, scalar multiplication

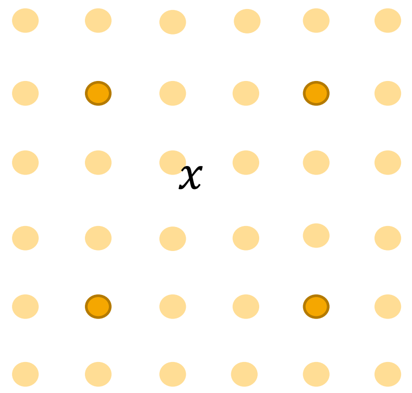


Hard problems

Decoding problems

General Decoding Problem

Given $x \epsilon \mathbb{F}^n$, find $c \epsilon C$ such that $dist(x, c)$ is minimal.

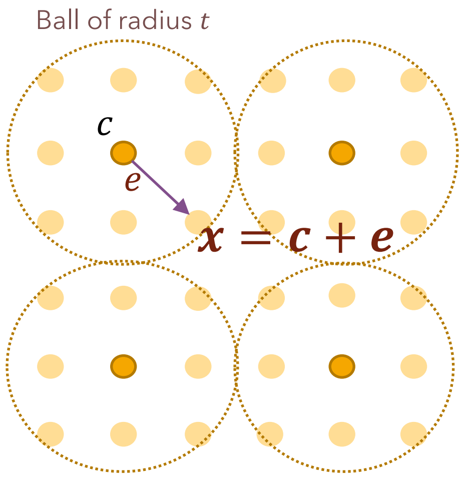


Decoding problems

General Decoding Problem: Given an $[n, k, d]$ linear code C, $t = \left\lfloor \frac{d-1}{2} \right\rfloor$, and a vector $x \epsilon \mathbb{F}^n$, find a codeword $c \epsilon C$ such that $dist(x, c) \leq t$.

Note: If $x = c + e$, and e is a vector with $|e| \leq t$, then x is uniquely determined.

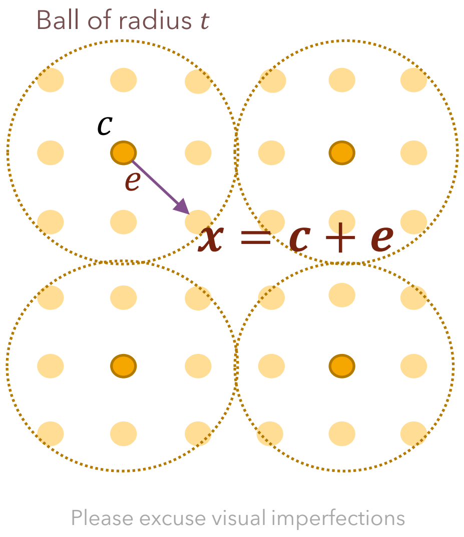
Shown to be NP-complete for **general linear codes** in 1978 (Berlekamp, McEliece, Tilborg) by reducing the three-dimensional matching problem to these problems.



Please excuse visual imperfections

Decoding problems

General Decoding Problem: Given an $[n, k, d]$ linear code C, $t = \left\lfloor \frac{d-1}{2} \right\rfloor$, and a vector $x \epsilon \mathbb{F}^n$, find a codeword $c \epsilon C$ such that $dist(x, c) \leq t$.



Note: Not all codes have a minimum distance d.
Rewrite problems in terms of linear (n, k) codes.

Shown to be NP-complete for **general linear codes** in 1978 (Berlekamp, McEliece, Tilborg) by reducing the three-dimensional matching problem to these problems.

Decoding problems

Let $C(n, k)$ be a linear code over finite field $\mathbb{F}$.

General decoding problem

Given a vector $\mathbf{x} \in \mathbb{F}^n$, a target weight $t > 0$,

find a codeword $\mathbf{c} \in \mathbb{F}^n$ such that $\mathbf{dist}(x, c) \leq t$.

Decoding problems

Let $C(n,k)$ be a linear code over finite field $\mathbb{F}$.

General decoding problem

Given a vector $\mathbf{x} \in \mathbb{F}^n$, a target weight $t > 0$,
find a codeword $\mathbf{c} \in \mathbb{F}^n$ such that $\mathbf{dist}(x,c) \leq t$.

Syndrome-decoding problem.

Given a parity check matrix $\mathrm{H} \in \mathbb{F}^{(n-k)\times n}$, a syndrome $\mathbf{s} \in \mathbb{F}^{n-k}$, a target weight $t > 0$,
find a vector $\mathbf{e} \in \mathbb{F}^n$ such that $wt(e) = t$ and $H \cdot e^T = s$.

Codeword-finding problem

Given a parity check matrix $\mathrm{H} \in \mathbb{F}^{(n-k)\times n}$ and a target weight $\mathrm{w} > 0$
find a vector $\mathbf{e} \in GF_2^n$ such that $wt(e) = w$ and $H \cdot e^T = 0$.

Relevance

In general, code-based cryptosystems rely upon this property:

- Encryption (some sort of matrix-vector product) is easy to compute
- Decryption is difficult without the trapdoor (the secret key which enables efficient decoding)

McEliece Cryptosystem

McEliece cryptosystem

First code-based cryptosystem.

Designed by Robert McEliece, presented in 1978.

McEliece cryptosystem

First code-based cryptosystem.

Designed by Robert McEliece, presented in 1978.

Idea: “hide” a message by converting it into a codeword, then add as many errors as the code is capable of correcting

Let $C[n, k, d]$ be a linear code with a fast decoding algorithm that can correct t or fewer errors

- Let G' be a generator matrix for C
- Let S be a $k \times k$ invertible matrix
- Let P be an $n \times n$ permutation matrix

McEliece cryptosystem

Let $C[n, k, d]$ be a linear code with a fast decoding algorithm that can correct t or fewer errors

- Let G' be a generator matrix for C
- Let S be a $k \times k$ invertible matrix
- Let P be an $n \times n$ permutation matrix

Define public key $G = SG'P$ with private key S, G', P

- Encrypt: $m \rightarrow mG + e$, $wt(e) \leq t$
- Decrypt:

1. Multiply $(mG + e)P^{-1} = mSG' + e'$

$wt(e) = wt(e')$

McEliece cryptosystem

Let $C[n, k, d]$ be a linear code with a fast decoding algorithm that can correct t or fewer errors

- Let G' be a generator matrix for C
- Let S be a $k \times k$ invertible matrix
- Let P be an $n \times n$ permutation matrix

Define public key $G = SG'P$ with private key S, G', P

Encrypt: $m \rightarrow mG + e, wt(e) \leq t$

Decrypt:

1. Multiply $(mG + e)P^{-1} = mSG' + e'$ $wt(e) = wt(e')$
2. $mSG' + e'$ → Fast decoding algorithm → mSG'
3. Multiply on the right by G'^{-1}, then by $S^{-1}$to recover m

Example

McEliece using (7,4) Hamming Code

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

http://www-math.ucdenver.edu/~wcherowi/courses/m5410/ctcmcel.html

Illustrate McEliece cryptosystem using (7,4) Hamming Code

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Secret scrambler and permutation matrices S, P chosen as

$$S = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix} \text{and } P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

http://www-math.ucdenver.edu/~wcherowi/courses/m5410/ctcmcel.html

Illustrate McEliece cryptosystem using (7,4) Hamming Code

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Secret scrambler and permutation matrices S, P chosen as

$$S = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix} \text{ and } P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

Then the public generator matrix $G' = SGP = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}$

http://www-math.ucdenver.edu/~wcherowi/courses/m5410/ctcmcel.html

Encrypt

Suppose Alice wishes to send message $u = 1\,1\,0\,1$ to Bob

1. Alice constructs a weight 1 error vector, say $e = 0\,0\,0\,0\,1\,0\,0$
2. Alice computes $uG' + e = 0\,1\,1\,0\,0\,1\,0 + 0\,0\,0\,0\,1\,0\,0$

$$= 0\,1\,1\,0\,1\,1\,0$$

Alice sends ciphertext $\mathbf{0\,1\,1\,0\,1\,1\,0}$ to Bob

http://www-math.ucdenver.edu/~wcherowi/courses/m5410/ctcmcel.html

Decrypt

1. Bob multiplies the ciphertext on the right by P^{-1}: $\mathbf{0\ 1\ 1\ 0\ 1\ 1\ 0}\begin{bmatrix}0&0&0&1&0&0&0\\1&0&0&0&0&0&0\\0&0&0&0&1&0&0\\0&1&0&0&0&0&0\\0&0&0&0&0&0&1\\0&0&0&0&0&1&0\\0&0&1&0&0&0&0\end{bmatrix}$

2. Bob takes the result 1 0 0 0 1 1 1 and uses fast decoding algorithm to remove the single bit of error

3. Bob takes the resulting codeword 1 0 0 0 1 1 0

- Knows that there is some x that satisfies x$G = x\begin{bmatrix}1&0&0&0&1&1&0\\0&1&0&0&1&0&1\\0&0&1&0&0&1&1\\0&0&0&1&1&1&1\end{bmatrix} =$ 1 0 0 0 1 1 0
- Equivalently knows that x$S = 1\ 0\ 0\ 0$, so multiplying on the right by S^{-1} yields 1 1 0 1

http://www-math.ucdenver.edu/~wcherowi/courses/m5410/ctcmcel.html

McEliece cryptosystem

Idea: “hide” a message by converting it into a codeword, then adding as many errors as the code is capable of correcting

Underlying code: McEliece used Goppa codes

- Efficient decoding
- Scrambled public key $G = SG'P$ is indistinguishable from random codes
- Public key ≈ a few megabits

McEliece cryptosystem

Idea: “hide” a message by converting it into a codeword, then adding as many errors as the code is capable of correcting

Underlying code: McEliece used Goppa codes

- Efficient decoding
- Scrambled public key $G = SG'P$ is indistinguishable from random codes
- Public key ≈ a few megabits (2^{19})
  - Typical RSA key sizes are 1,024 or 2,048 or 4,096 bits
  - ECDH key sizes are roughly 256 or 512 bits

Trapdoor

NP-completeness of decoding problem does not indicate cryptographic security for concrete instances

Private key S, G', P turn out to be trapdoors ($G = SG'P$)

Encryption: $mG + e$ easy to compute

Decryption difficult without S, G', P

Best known algorithm to solve decoding problems: **Information Set Decoding (Prange, 1962)**