Algorithms of Information Security: Error-correcting codes II

Faculty of Information Technology
Czech Technical University in Prague

October 6, 2021

Cyclic code

Definition

A code C is cyclic if every cyclic shift of a codeword in C is also a codeword. That is, $(c_0, c_1, \ldots, c_{n-1}) \in C$ implies that $(c_{n-1}, c_0, \ldots, c_{n-2}) \in C$.

In the notation of representing polynomials, a code C is cyclic if and only if $c(x) \in C$ implies

$$x \cdot c(x) \text{ mod } (x^n - 1) \in C.$$

If a code is linear, then equivalently we can say that $c(x) \in C$ implies

$$u(x) \cdot c(x) \text{ mod } (x^n - 1) \in C$$

for every $u(x) \in F_q[x]$. Hence, C is a linear cyclic code if and only if C is an ideal in the ring $R = F_q[x]/(x^n - 1)$. R is principal ideal domain (every ideal is principal) and hence we can express the code C using generator $g(x)$ as follows: $C = (g) = \{g \cdot h | h \in R\}$.

Cyclic code

Theorem

Let C be a cyclic code over F_q and g the monic polynomial in C of minimal positive degree (prove that it is unique!). Then g generates C, i.e., $c \in C$ iff $g \mid c$.

Theorem

A polynomial code is cyclic if and only if its generator polynomial divides $x^n - 1$, where n is length of the code.

Dual Codes of Cyclic Codes

Let C be a cyclic $[n,k]$-code with a generator $g(x) = \sum_{i=0}^{n-k} g_i x^i$.
We know that g divides $x^n - 1$, and therefore, there exists
$h(x) = \sum_{i=0}^{k} h_i x^i$ such that $gh = x^n - 1$.
Let $c \in C$. As g generates C we have $c = ga$ for some $a \in F_q[x]$.
Therefore

$$hc \text{ mod } (x^n - 1) = hga \text{ mod } (x^n - 1) = 0$$

This translates to the constraints:

$$c_0 h_i + c_1 h_{i-1} + \ldots + c_{n-k} h_{i-n+k} = 0,$$

for every $0 \leq i \leq n-1$, where the indices are modulo n.

Dual Codes of Cyclic Codes

It follows that

$$H = \begin{pmatrix} h_k & h_{k-1} & \dots & h_0 & & & \\ & h_k & h_{k-1} & \dots & h_0 & & \\ \vdots & \ddots & & & & & \\ & & & h_k & h_{k-1} & \dots & h_0 \end{pmatrix}$$

is a $(n-k) \times n$ matrix and since it has the correct rank $n-k$ it is a parity check matrix of C.

Theorem

Let C be an $[n,k]$ cyclic code generated by $g(x)$ and let $h(x) = \frac{x^n-1}{g(x)}$. Then, the dualcode of C is a cyclic $[n, n-k]$code whose generator polynomial is $x^k h(x^{-1})$. The polynomial $h(x)$ is called the check polynomial of C.

The Binary Hamming Code

Definition

Let $r \geq 2$ and let C be a binary linear code with $n = 2^r - 1$ whose parity check matrix H is such that the columns are all of the non-zero vectors in F_2^r. This code C is called a binary Hamming code of length $2^r - 1$, denoted $\mathrm{Ham}(r, 2)$.

Propositions.

1. All binary Hamming codes of a given length are equivalent.
2. For every $r \in \mathbb{N}$, the dimension of $\mathrm{Ham}(r, 2)$ is $k = 2^r - r - 1$.
3. For every $r \in \mathbb{N}$, the distance of $\mathrm{Ham}(r, 2)$ is $d = 3$ and so the code can correct exactly one error.

Example. A generator matrix for Ham$(r, 2)$, where $r = 3$, is as follows:

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

Size of the code is
$M =\mid C \mid=\mid \{\sum_{i=1}^{4} u_i v_i, u_i \in \{0, 1\}\} \mid= 2^4 = 16.$
The parity check matrix for Ham$(r, 2)$ is as follows:

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

As can be seen, for Ham$(r, 2)$ we have $n = 7, k = 4$ and H is the matrix of type (3×7) over F_2.

The Hamming code is cyclic

Any binary Hamming code is equivalent to a cyclic code.

Theorem

Fix a field F_{2^r} and let $n = 2^r - 1$. Then, there exists a $[n, k = n - r, 3]_2$ cyclic code. Since the only code with such length, dimension and distance is the Hamming code, the Hamming code is cyclic.

The Reed-Solomon code

Reed-Solomon codes were invented by Irving S. Reed and Gustave Solomon in 1960. In 1977 RS codes have been implemented in Voyager space program. The first commercial application of RS codes in mass-consumer products was in 1982.

Reed-Solomon codes were used in digital television, satellite communication, wireless communication, bar-codes, compact discs, DVD, ...

Fix a field F_q of size q with a generator α of F_q^*. Element α is called primitive element. The code $RS : F_q^k \to F_q^n$ corresponds to evaluating all polynomials of degree atmost $k-1$ (coefficients are given to us as the input message) on all nonzero field elements. That is, $n = q - 1$ and

$$RS(a_0, \ldots, a_{k-1}) = (p_a(\alpha^0), p_a(\alpha^1), \ldots, p_a(\alpha^{n-1})),$$

where $p_a(x) = \sum_{i=0}^{k-1} a_i x^i$.

Every nonzero polynomial of degree $k-1$ can have at most $k-1$ zeros in F_q, so the weight of every nonzero codeword is at least $d = n-(k-1) = n-k+1$. Thus, the RS code is an $[n,k,n-k+1]_q$ code. By inspection, the $n \times k$ transpose of generating matrix is given by

$$G^T = \begin{pmatrix} (\alpha^0)^0 & (\alpha^0)^1 & \dots & (\alpha^0)^{k-1} \\ (\alpha^1)^0 & (\alpha^1)^1 & \dots & (\alpha^1)^{k-1} \\ \vdots & \vdots & \ddots & \vdots \\ (\alpha^{n-1})^0 & (\alpha^{n-1})^1 & \dots & (\alpha^{n-1})^{k-1} \end{pmatrix}$$

so for $0 \le i \le n-1$ and $0 \le j \le k-1$, $G^T[i,j] = \alpha^{ij}$.

Now, consider the $n \times (n-k)$ matrix

$$H^T = \begin{pmatrix} (\alpha^0)^1 & (\alpha^0)^2 & \dots & (\alpha^0)^{n-k} \\ (\alpha^1)^1 & (\alpha^1)^2 & \dots & (\alpha^1)^{n-k} \\ \vdots & \vdots & \ddots & \vdots \\ (\alpha^{n-1})^1 & (\alpha^{n-1})^2 & \dots & (\alpha^{n-1})^{n-k} \end{pmatrix}$$

so for $0 \leq i \leq n-1$ and $1 \leq j \leq n-k, H^T[i,j] = \alpha^{ij}$. H^T is a Vandermonde matrix, so in order to prove that H^T is indeed the parity-check matrix of the RS code, it is sufficient to prove that $H \cdot G^T = 0_{(n-k)\times k}$.

We have that

$$(H \cdot G^T)[a,b] = \sum_{k=0}^{n-1} H[a,k]G[b,k] = \sum_{k=0}^{n-1} (\alpha^{a+b})^k.$$

a ranges from 1 to $n-k$ and b ranges from 0 to $k-1$, so $1 \leq a+b \leq n-1$ and the above sums to zero.

Theorem

Reed-Solomon codes are linear codes.

Theorem

Reed-Solomon codes are polynomial codes.

Note. Reed-Solomon (RS) codes are non-binary cyclic codes.

An interesting property of Reed-Solomon codes

$$RS(k,q)^{\perp} = RS(q-k,q).$$

Singleton bound

Theorem

The minimum distance for a linear $[n, k]$-code is bounded by

$$d \leq n - k + 1.$$

For Reed-Solomon codes $d \geq n - k + 1$, so $d = n - k + 1$ and all Reed-Solomon codes meet the Singleton bound – they are optimal $[n, k, n - k + 1]$-codes, $n = q - 1$.

Definition

Codes that meet the Singleton bound are called Maximum Distance Separable codes (MDS).

Note. The dual code of an Reed-Solomon code is also MDS code.

Example

Let $n = 8 - 1 = 7$. We want form F_8 from $x^3 + x + 1$.

α^0	1
α^1	α
α^2	α^2
α^3	$\alpha + 1$
α^4	$\alpha^2 + \alpha$
α^5	$\alpha^2 + \alpha + 1$
α^6	$\alpha^2 + 1$

Let $k - 1 = 3$, i.e., $k = 4$.

Then

$$H^T = \begin{pmatrix} 1 & \alpha & \alpha^2 & \alpha^3 & \alpha^4 & \alpha^5 & \alpha^6 \\ 1 & \alpha^2 & \alpha^4 & \alpha^6 & \alpha & \alpha^3 & \alpha^5 \\ 1 & \alpha^3 & \alpha^6 & \alpha^2 & \alpha^5 & \alpha & \alpha^4 \\ 1 & \alpha^4 & \alpha & \alpha^5 & \alpha^2 & \alpha^6 & \alpha^3 \end{pmatrix}$$

Then

- RS: $n = q - 1 = 7$, where $q = 8$
- RS: we have $[7, 3, 5]_8$ code.
- RS: $q^k = 8^3 = 2^9 = 512$ codewords.

BCH codes

BCH codes were discovered by independently by Bose and Ray-Chaudhuri and by Hocquenghem in the late 1950s.
BCH codes can be defined over any field, first we will focus on binary BCH codes:

Definition

For a length $n = 2^m - 1$, a distance d, and a primitive element $\alpha \in F^*_{2^m}$, we define the binary BCH code

$$BCH[n,d] = \Big\{(c_0, c_1, \ldots, c_{n-1}) \in F_2^n \mid c(X) = c_0 + c_1X + \ldots + c_{n-1}X^{n-1}$$

$$\text{satisfies } c(\alpha) = c(\alpha^2) = \ldots = c(\alpha^{d-1}) = 0\Big\}.$$

Lemma

The BCH codes form linear spaces.

Definition

For prime power q, integer m, and integer d, the BCH code $BCH_{q,m,d}$ is obtained as follows: Let $n = q^m - 1$ and let F_{q^m} be an extension of F_q and let C' be the (extended) $[n, n-(d-1), d]_{q^m}$ Reed-Solomon code obtained by evaluating polynomials of degree at most $n-1$ over F_{q^m} at all the points of F_{q^m}. Then the code $BCH_{q,m,d} = C' \cap F_q^n$.

If we have $q = 2$ then $BCH_{2,m,d} = C \cap F_2^{2^n}$ where C is given as a Reed-Solomon code $C = RS[n = 2^m, n-(d-1), d]_{2^m}$.

The BCH code could be constructed in the following manner: Look at the Reed-Solomon code and only pick up the codewords that are in $F_2^{2^m}$.
Conjecture. Dimension of BCH code is at least $n - m(d-1)$.
The general idea of a BCH code is to identify its generating polynomial by the roots(instead of in terms of the coeficients).

Theorem

For prime power q, integers m and d, the $BCH_{q,m,d}$ is an $[n, n-1-m\lceil\frac{(d-2)(q-1)}{q}\rceil, d]_q$ code, for $n = q^m$.

In the case of $q = 2$ (binary codes) we have the following corollary:

Corollary

For every integer m and t, the code $BCH_{2,m,2t}$ is an $[n, n-1-(t-1)\log n, 2t]$ code, for $n = 2^m$.

Note. Reed-Solomon codes are nonbinary BCH codes.