Noise can modify messages

Solution: encoding with some redundancy for correct some errors.

We can see a message as a sequence of symbols from a finite set $\mathcal{F}$. If there are errors some symbols will be changed.

To detect errors we recieve only code words, if there are errors in trasmission we can see the message is not a code word

> Suppose $\mathcal{F}$ a finite set of symbols. A n-code over $\mathcal{F}$ is a subset of $\mathcal{F}^n$.
> If $\mathcal{F} = \mathbb{F}_2 = \{0,1\}$ we call the code binary. If $|\mathcal{F}| = q$ it's q-ary

eg. $C := \{00000, 11110, 01011, 10101\}$ is a binary 5-code

> an error occurres when the ith entry of a code word changes

eg. $\underset{\Downarrow}{11110} \longrightarrow \underset{\Downarrow}{01110}$

> The Hamming distance $d(x,y)$ between two sequences x, y is the number of places where they differ eg. $d(1101, 0011) = 3$
>
> properties
>
> - $\forall x, y \in C: d(x,y) = d(y,x)$
> - $\forall x, y \in C: d(x,y) = 0 \iff x = y$
> - $\forall x, y \in C: d(x,y) \leq d(x,z) + d(z,y)$

when i errors occur the Hamming distance $= i$.

To correct the message we use the smallest distance.

example: $C := \{00000, 11110, 01011, 10101\}$

if we recieve 11010 the distances are:

$d(11010, 00000) = 3$

$d(11010, 11110) = 1$ ← we decode it as 11110

$d(11010, 01011) = 2$ it's most likely only one error occured.

$d(11010, 10101) = 4$

The distance of a code is the minimum of the distances beetween all code words

$$d(C) = \min\{d(x,y) \mid x,y \in C, x \neq y\}$$

the parameters of a code are:

- n -> number of symbols useal for a code word
- $|C|$ -> size of the code
- d -> minimal distance

So we speak of a $(n, |C|, d)$- code

Lemma ①

A code C detects (up to) s errors if and only if $d(C) > s$.

proof: $\Rightarrow$

Suppose $d(C) > s$. $c \in C$, x word obtained from c with up to s errors. now $d(x,c) \leq s$, so $s < d(C)$.

So C detects up to s errors

$\Leftarrow$

if $s \geq d(C)$ then exist $c_1, c_2 \in C$ with $d(c_1, c_2) = d(C) \leq s$.

So c_1 can be converted into c_2 with no more than s errors, and this will not be detected.

Lemma ②

A code C corrects (up to) t errors if and only if $d(C) \geq 2t+1$

proof: $\Rightarrow$

suppose $d(C) \geq 2t+1$, $c \in C$ x obtained from c with at most t errors

we have to show that x is closer to c than any other word $r \in C$.

$$2t+1 \leq d(C) \quad \text{(since } d(C) \leq \text{all other distances, } 2t+1 \leq d(C) \leq d(c,x)+d(x,r)\text{)}$$
$$\leq d(c,x) + d(x,r)$$
$$\leq t + d(x,r) \Rightarrow t+1 \leq d(x,r)$$

$\Leftarrow$ if $2t+1 > d(C)$ then $\exists c_1, c_2 \in C \mid d(c_1, c_2) = 2t+1 > d(C) > d(c,x)$

Examples

Repetition codes

$\{000, 111\}$ is a $[3,1]$ repetition code (K|n to make sense)

(K: length of unit being repeated; n: length of each code word)

the number of times each unit is repeated is $r = n/K$

Another example is $[4,2]$. Original message units of length 2 $\{00, 01, 10, 11\}$

the codeword are created by repeating them, so $\{0000, 0101, 1010, 1111\}$

$[n,K] \rightarrow \underbrace{d_1 d_2 \cdots d_K d_1 d_2 \cdots d_K \cdots d_1 d_2 \ldots d_K}_{r}$, where r_i are digits $\in \mathbb{F}$

$q = |\mathbb{F}|$ n. of codewords $= q^K$.

- the minimal distance of the code is (n/K). $(q, q^K, \frac{n}{K})$

Belgium bank accounts

12 digits. Grouped in 3-7-2 eg. 103-0202707-45

digits $[0\_10] \pmod{97} = [11\_12]$

EAN-13 $(13, 10^{12}, 2)$ (12 digits (there are 10 digits))

13 characters

- first 2/3 country code of manufacturer
- 9/10 digits
- 1 digit checksum

the checksum is calculated as follows

a) sum values of digits in even position

b) Multiply this by 3

c) sum values of digits in odd positions

d) sum the results of (b) and (c)

e) the check character is the smallest number which when added to the result of (d) gives a multiple of 10

example: data = 001234567890

a) $0+2+4+6+8+0 = 20$

b) $20 \cdot 3 = 60$

c) $0+1+3+5+7+9 = 25$

d) $60+25 = 85$

e) $X = 5$

code word $\rightarrow$ 0012345678905

Code 39 $(9, \binom{9}{3}, 2)$ *

tutte combinazioni di 3 spazi in 9 caratteri

wide / narrow

Alphabet of two letters $\{w, n\}$

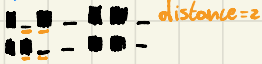
45 code words corresponding to:

- A-Z
- 0-9
- space, (-), (+), (.), ($), (/), (%), special start/stop character

Each codeword is 9 elements and has exactly 3 wides.

Usually the word is printed with black strips

* se sposto uno spazio i punti in cui differisce son due



Parity code $(n, 2^{n-1}, 2)$

Represent alphabet with 5 bits

Character	Code	Character	Code
A	00000	N	01101
B	00001	O	01110
C	00010	P	01111
D	00011	Q	10000
E	00100	R	10001
F	00101	S	10010
G	00110	T	10011
H	00111	U	10100
I	01000	V	10101
J	01001	W	10110
K	01010	X	10111
L	01011	Y	11000
M	01100	Z	11001

add another bit to make the number of 1s even

the sum of their digits is even

If a bit changes the parity does.

Any single error will be detected, but even number of errors will not

no error correction

Hamming square code $(9, 2^4, 4)$

allows single errors to be corrected.

message units = 4 bits write in rows and sum mod 2

M: 1101 → 11 01 | 0 / 01 | 1 / 1 0 | 1 → code word: 11 01 01 1 01

we detect errors by checking parity in the rows

if one error occurs we can identify it because it will cause an error in one row and one column

so we actually can correct it.

If more happen (up to 3) we can only detect them

Equivalence of codes

$C_1 = \{0011100, 0111011, 1100111\}$

$C_2 = \{000110, 101101, 111001\}$ (shift) — positional permutation

$C_3 = \{1100001, 1101110, 0111100\}$ Hamming distance is the same

Lemma: Performing positional permutation preserves the minimum distance of the code

$C_1 = \{0110111, 0101101, 0001111\}$

$C_2 = \{0100011, 0111101, 0011111\}$ change 3rd digit — symbolic permutation

$C_3 = \{0111100, 0100010, 0000000\}$ change last 3 digits

Lemma: performing symbolic permutation at some or all positions preserves the minimum distance

> two codes are equivalent if one can be obtained from the other via symbolic and positional permutations

consider distance of pair of codewords of the codes X and Y

$$X = \{x_1, x_2, x_3, x_4\} = \{0000000, 0011001, 0101011, 0110111\}$$

$$Y = \{y_1, y_2, y_3, y_4\} = \{1011001, 1110111, 1101011, 0111000\}$$

same numbers different order

X	x_1	x_2	x_3	x_4
x_1		3	4	5
x_2			3	4
x_3				3

Y	y_1	y_2	y_3	y_4
y_1		4	3	3
y_2			3	5
y_3				4

if we rearrange the order of words of Y then we get the same table.

We say x and y are distance isomorphic

Two p-ary codes C_1 and C_2 of the same length and size are distance isomorphic if their words can be ordered in such way that $\forall x_i, x_j \in C_1 \wedge \forall y_i, y_j \in C_2 \quad d(x_i, x_j) = d(y_i, y_j)$

exercise bank account

12 digits, grouped in 3-7-2 eg 103-0202707-45

the first 10 digits must equal the last 2 modulo 97

$1030202707 = 10620646 \cdot 37 + 45$

it's 1 error detecting but not 1 error correcting (because $d < 3$)

$n = 12 \quad |C| = 10^{10}$ ↳ since $d < 2t+1 \ \forall t \in \{1, 2, \dots\}$

$d = 2$

prove that we can detect the swap of 2 digits

$7331041885-38$ → it's going to differ after calculation, and since
$1337041885-38$ it's only 2 digits $d' = 2 \leq d$ ✓

Hamming Square $(9, 2^4, 4)$

Parity Code $(n, 2^{n-1}, 2)$

EAN 13 ? $(13, 10^{12}, 2)$

39 code (BARCODE) $(9, \binom{9}{3}, 2)$
$= 84$