§ Coding Theory and Cryptography

Lecture 6

Warm up problem: Wrong Number. (Similar to page 17 and 5.7 of Ecco)

In a city called Five people are getting very upset. In Five, phone numbers are 5 digits long made with the numbers 0 to 4, but lately many phone calls were resulting in wrong numbers. After a correct number has been called, in transmission, one pair of adjacent digits gets swapped ("the switch bug"). For example the number ABCDE could be called but whoever is at AB**DC**E receives the phone call.

After a heated city hall discussion, going against all cultural beliefs the city has decided to add a sixth number to their phone system. This was decided even though the sixth digit will still be prone to swapping with the fifth digit. The scientists of Five are going to add a sixth digit called a *check digit*. After doing so, called numbers that experience "the switch bug" will result in a nonfunctioning number.

How can the check digit be chosen successfully?

Coding Theory

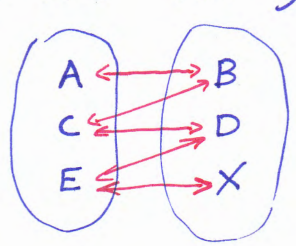
Definition 1: A *codeword* is a word (or string) of digits or letters or other symbols.

Definition 2: A *code* is a collection of codewords.

Example 1: Find a solution to the Wrong Number problem:

A B C D E X

put the digits into two groups:



Notice digits do not get swapped within each group.

Now pick X so:

$$B + D + X \equiv 0 \mod 5$$

Suppose there is an error $C \longleftrightarrow D$:

CASE 1 $\quad B + C + X \not\equiv 0 \mod 5 \quad \therefore$ we report an error.

CASE 2 $\quad B + C + X \equiv 0 \mod 5$

$\Longrightarrow \quad B + C + X \equiv B + D + X \mod 5$

$\Longrightarrow \quad C \equiv D \mod 5$

$\Longrightarrow \quad C = D$ since $C, D \in \{0, 1, 2, 3, 4\}$

$\therefore$ there was no error.

Note: In this solution to the wrong number problem, a code with 5^5 codewords was used.

$$\underbrace{A\,B\,C\,D\,E}_{\text{Freely chosen from } \{0,1,2,3,4\}}\;\underbrace{X}_{\text{determined by } ABCDE}$$

Definition 3: Given two codewords x, y of the same length the *Hamming distance* $H(x, y)$ is the number of places in which the components of the strings differ.

- if $x = 1010101010$, $y = 1111111111$ $\Longrightarrow$ $H(x, y) = 5$

- if $x = AB\$12$, $y = AC\#12$ $\Longrightarrow$ $H(x, y) = 2$

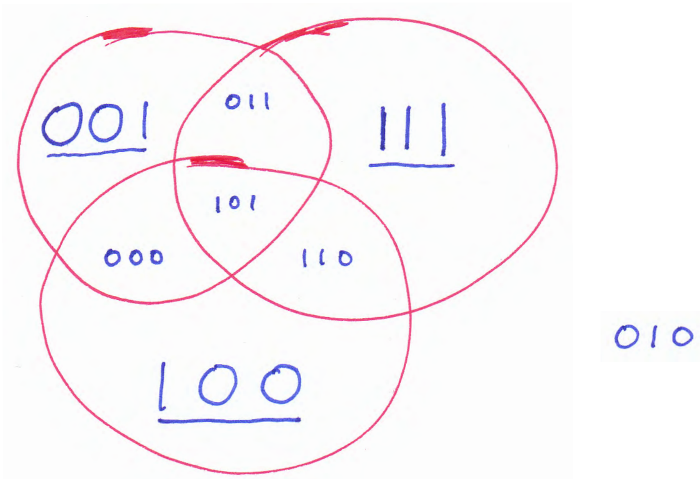
- if $x = 12345$, $y = 54321$ $\Longrightarrow$ $H(x, y) = 4$

Error Detecting

Example 2: In the code:

001
111
100

consider the possible error of one's flipping to zeros and zeros flipping to ones. Draw a Venn diagram where the circles represent a hamming distance of one from each code word.

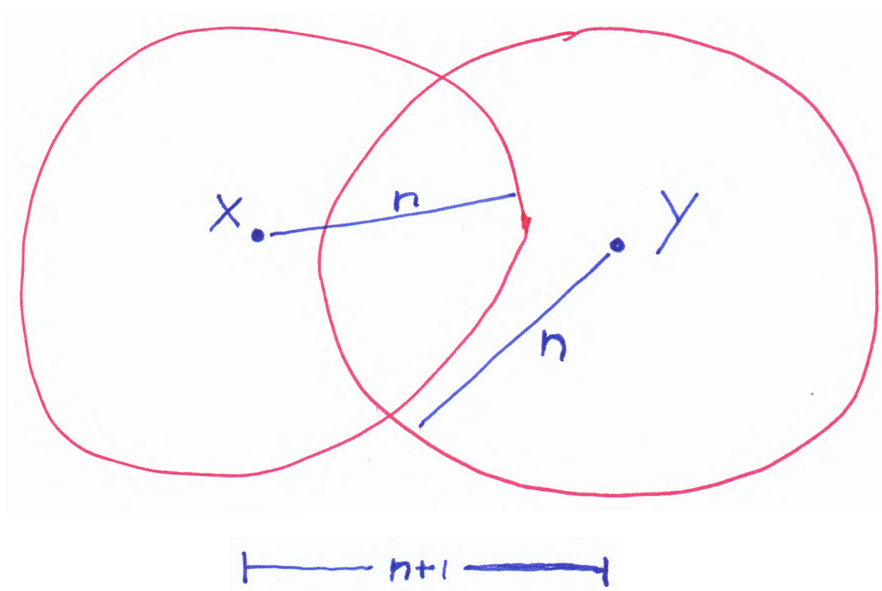


Since in the above code each pair of codewords has a Hamming distance of 2

we can detect 1 error.

Theorem 1: If up to n characters of a codeword can be corrupted, and if the Hamming distance between every pair of codewords is at least $n + 1$, then an error can be detected.

Picture a pair of codewords:

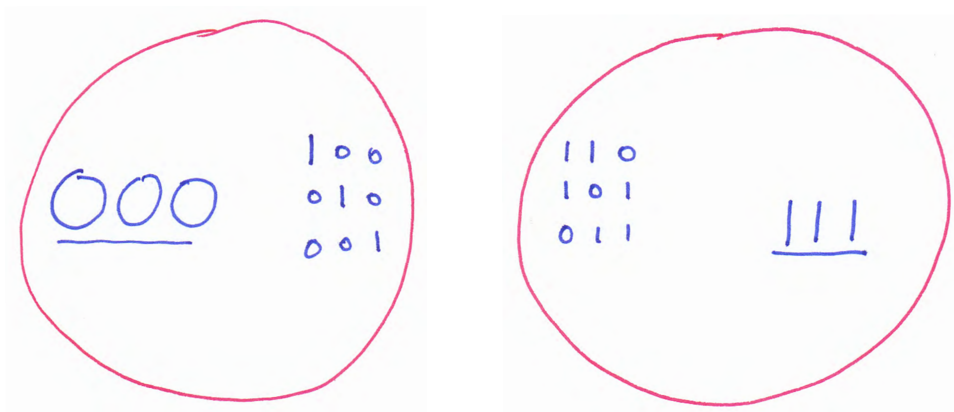


Error Correcting

Example 3: In the code:

000
111

consider the possible error of one's flipping to zeros and zeros flipping to ones. Draw a Venn diagram where the circles represent a hamming distance of one from each code word.

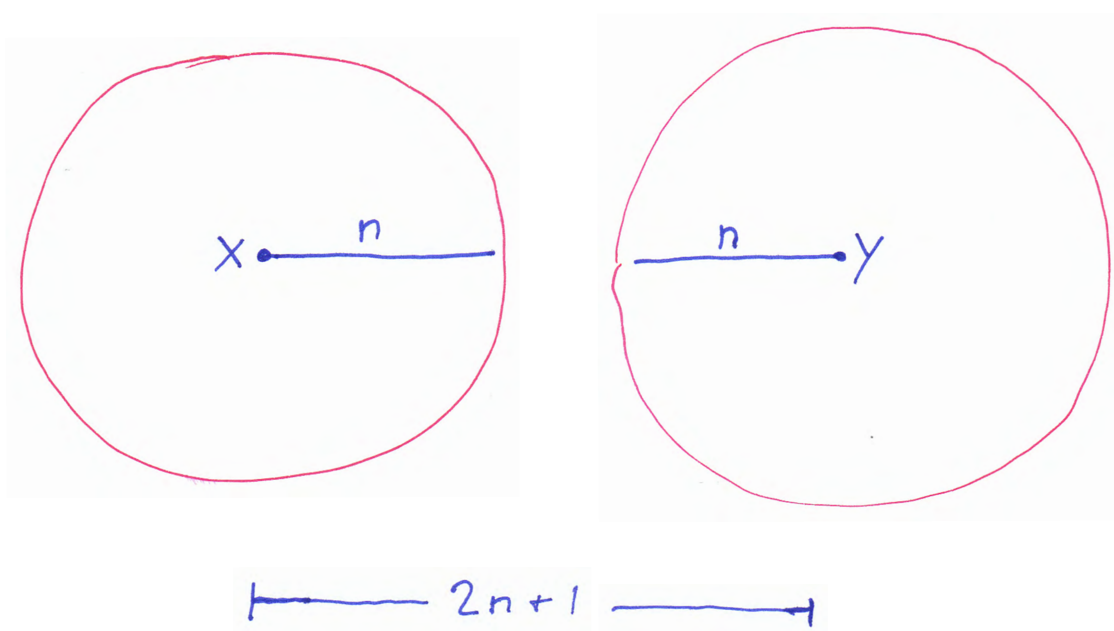


Since in the above code each pair of codewords has a Hamming distance of _3_

we can correct _1_ error.

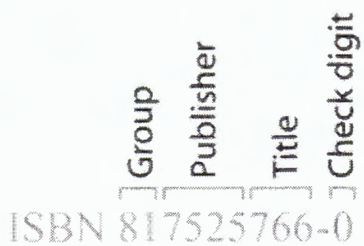
Theorem 2: If up to n characters of a codeword can be corrupted, and if the Hamming distance between every pair of codewords is at least $2n + 1$, then an error can be corrected.

Picture a pair of codewords:



How good is an ISBN code?

The ISBN code has 9 information digits and one check digit for example:



If the first nine digits of an ISBN are $abcdefghi$ then the tenth digit j is chosen so:

$$10a + 9b + 8c + 7d + 6e + 5f + 4g + 3h + 2i + j \equiv 0 \pmod{11}$$

When $j \equiv 10 \pmod{11}$, an X is used instead.

Example 4: The ISBN code can detect a single digit error:

Suppose $a \leftrightarrow y$:

CASE 1 $\quad 10y + 9b + \cdots + j \not\equiv 0 \mod 11$ report an error.

CASE 2

$$\begin{aligned} & 10y + 9b + \cdots + j \equiv 0 \mod 11 \\ - \; & 10a + 9b + \cdots + j \equiv 0 \mod 11 \\ \hline & 10(y-a) \equiv 0 \mod 11 \end{aligned}$$

$\Rightarrow \quad y - a \equiv 0 \mod 11$ since $10, 11$ are r. p. we can use L2 T3.

$\Rightarrow \quad y \equiv a \mod 11$

$\Rightarrow \quad y = a$ since $y, a \in \{0, 1, \ldots, 9\}$

$\therefore$ there was no error

Example 5: This means the ISBN code is a single error detecting code, so every pair of codewords has a Hamming distance of at least 2. For example:

x: 00 0000 000 0 OR x: 11 1111 111 1

y: 00 0000 005 1 y: 11 1111 112 X

$$\Rightarrow H(x,y) = 2$$

Example 6: The ISBN code can detect when two adjacent symbols are interchanged (the switch bug):

Suppose $a \leftrightarrow b$:

CASE 1 $10b + 9a + \cdots + j \not\equiv 0 \mod 11$ report an error.

CASE 2
$$\begin{aligned} 10b + 9a + \cdots + j &\equiv 0 \mod 11 \\ -\ \ 10a + 9b + \cdots + j &\equiv 0 \mod 11 \\ \hline b - a &\equiv 0 \mod 11 \\ \Rightarrow b &\equiv a \mod 11 \\ \Rightarrow b &= a \quad \text{since } a, b \in \{0, 1, \ldots, 9\} \end{aligned}$$

$\therefore$ there was no error

Example 7: The ISBN code can correct a single digit error when the location of the error is known. For example:

Ecco's ISBN: 0—486—2961?—6

$$10 \cdot 0 + 9 \cdot 4 + 8 \cdot 8 + 7 \cdot 6 + 6 \cdot 2 + 5 \cdot 9 + 4 \cdot 6 + 3 \cdot 1 + 2 \cdot y + 6$$

$$\equiv 0 + 3 - 2 - 2 + 1 + 1 + 2 + 3 + 2y + 6$$

$$\equiv 2y + 1$$

$$\equiv 0 \mod 11 \quad \Rightarrow \quad y = 5$$

Example 8: Can The ISBN code correct a single digit error in general?

No in Ex 5 $H(x,y) = 2 < 2(1) + 1 = 3$