Lecture 9

Cryptography is the practice and study of hiding information.

There is a sender-receiver team on one side and a kibitzer on the other side.

Definition 1: *Plaintext* is the original message from the sender.

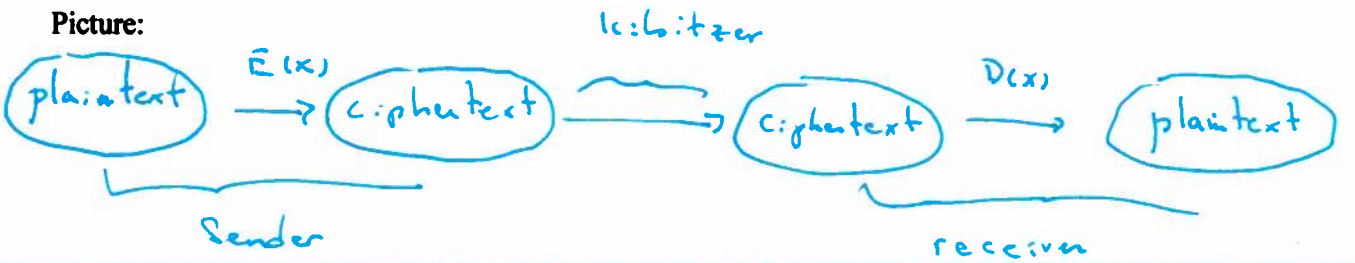
Definition 2: The sender knows an *encoding function* $E(x)$, which turns plaintext into gibberish.

Definition 3: The gibberish from $E(x)$, is called *ciphertext*.

Definition 4: The *kibitzer* is someone trying to obtain the original message from gibberish without $E(x)$.

Definition 5: The receiver knows a *decoding function* $D(x)$, which turns the gibberish back into the original message. So we have $D(E(x)) = x$ for any message x.

Picture:



$D(E(x)) = x$ and also $E(D(x)) = x$

We encode letters as numbers using mod 26:

A	B	C	D	E	F	G	H	I	J	K	L	M
0	1	2	3	4	5	6	7	8	9	10	11	12

N	O	P	Q	R	S	T	U	V	W	X	Y	Z
13	14	15	16	17	18	19	20	21	22	23	24	25

Caesar's Code:

$E(x) \equiv x - 6 \pmod{26}$ ← encoding function

$D(x) \equiv x + 6 \pmod{26}$ ← decoding function

In general, we have

$$E(x) \equiv x + k \pmod{26}$$

$$D(x) \equiv x - k \pmod{26}$$

Called a generalized Caesar's code

Ex: with $k = 3$, generalized Caesar's code to encipher LET

$E(x) \equiv x + 3$

$L \equiv 11$, $E \equiv 4$, $T \equiv 19$

$E(11) \equiv 11 + 3 \pmod{26} \equiv 14 \pmod{26}$

$L \to O$

$E \equiv 4$, $E(4) \equiv 4 + 3 \equiv 7 \pmod{26}$

$E \to H$

$T \equiv 19$, $E(19) \equiv 19 + 3 \equiv 22 \pmod{26}$

$T \to W$

Plaintext LET encoded as OHW ciphertext

Generalized Caesar's Code

a. The sender picks a value k and then shifts each letter in the plaintext with the encoding function:

$$E(x) \equiv x + k \bmod 26$$

b. The receiver can find the plaintext by using the decoding function:

$$D(x) \equiv x - k \bmod 26$$

c. The kibitzer can shift each letter in the ciphertext down the alphabet one by one until a word makes sense

Example 1:

a) Play the role of the sender. With $k \equiv 3$ use the generalized Caesar's code to encipher the word:

LET
MFU
NGV
OHW

b) Play the role of the receiver. With $k \equiv 3$ decipher the following word which was encoded using the generalized Caesar's code.

IT
JU
KV
LW

c) Play the role of the kibitzer. Find the plaintext from the following ciphertext which was encoded using the generalized Caesar's code.

BE
CF
DG
EH
FI
GJ
HK

Note: the answer is not "EH" or "OR" since the message is "Let it BE".

Note: Generalized Caesar code is not very secure. can be easily cracked, since all we are doing is shifting the letters of alphabet (mod 26).

If the encoding function is $E(x) \equiv x + k \pmod{26}$ we don't need to know what k is to crack the code.

Ex: Suppose the ciphertext is

G	C	F	C	N	U	L	S
H	D	G	D	O	V	M	T
I	E	H	E	P	W	N	U
J	F	I	F	Q	X	O	V
K	G	J	G	R	Y	P	W
L	H	K	H	S	Z	Q	X
M	I	L	I	T	A	R	Y

Found what may be a word in the plaintext

from this we get the encoding function:

$$E(x) \equiv x - 6 \pmod{26}$$

and the decoding function:

$$D(x) \equiv x + 6 \pmod{26}$$

Practice: All the groups below were encoded with a generalized Caesar's code.

GROUP 1

Koymncihm uly nby wlyuncpy uwnm iz chnyffcayhwy. - Zluhe Echaxih

GROUP 2

Dolu avsk "Spml pz ohyk." Cvsahpyl hzrlk "Jvtwhylk av doha?"

GROUP 3

Nby vynnyl juln iz ihy'm fczy wihmcmnm iz bcm zlcyhxmbcjm. --Uvlubug Fchwifh

GROUP 4

B phnew ktmaxk ux otznxer kbzam matg ikxvblxer pkhgz. – Dxrgxl

GROUP 5

"Mh uxebxox t mabgz bl bfihllbuex bl mh ftdx bm lh." - Ykxgva ikhoxku

GROUP 6

"N zna pna or qrfgeblrq ohg abg qrsrngrq." - Rearfg Urzvatjnl

GROUP 7

"hmttxj yt gj tuynrnxynh, ny kjjqx gjyyjw." - Ifqn Qfrf

Linear Codes:

- Sender encodes the plaintext with the encoding function

$$E(x) \equiv ax + k \pmod{26}$$

where a is an integer relatively prime to 26.

- Receiver decodes the ciphertext with the decoding function

$$D(x) \equiv \bar{a}(x - k) \pmod{26}$$

where $\bar{a}$ is the inverse of a modulo 26, (exists because a and 26 are relatively prime).

Ex: An investment firm has two prospective employees, Norwood and Abcham. The CEO has a private investigator look into their backgrounds, and tells the investigator to send him, by code, the name of the one to hire, i.e. NOR or ABE. He discovers that ABE has a history of embezzlement, so he sends the message NOR to the CEO.

They use the encoding function

$$E(x) \equiv 2x + 6 \pmod{26}$$

(A = 0, B = 1, C = 2, ..., Z = 25)

N = 13, $E(13) \equiv 2 \cdot 13 + 6 \equiv 32 \equiv 6 \pmod{26}$

O = 14, $E(14) \equiv 2 \cdot 14 + 6 \equiv 34 \equiv 8 \pmod{26}$

R = 17, $E(17) \equiv 2 \cdot 17 + 6 \equiv 40 \equiv 14 \pmod{26}$

N	O	R	← Plaintext
13	14	17	← x
6	8	14	← E(x)
G	I	O	← ciphertext

Investigator sends the ciphertext G I O.

Investment firm decodes G I O using the decoding function

$$D(x) \equiv \frac{x-6}{2} \pmod{26}$$

G I O ← ciphertext

$$D(6) \equiv \frac{6-6}{2} \equiv 0 \pmod{26}$$

6 gets decoded as A

$D(8) \equiv \frac{8-6}{2} \equiv 1 \pmod{26}$

So I gets decoded as B

$D(14) \equiv \frac{14-6}{2} \equiv 4 \pmod{26}$

So O gets decoded as E

The investment firm has decoded the message NOR as ABE and they hire the wrong person.

What went wrong?

NOR $\longrightarrow$ 13 14 17 $\xrightarrow{E(x)}$ 6 8 14 $\longrightarrow$ GIO

ABE $\longrightarrow$ 0 1 4 $\xrightarrow{E(x)}$ 6 8 14 $\longrightarrow$ GIO

So both names get encoded to the same codeword and $D(x) \equiv \frac{x-6}{2}$ won't decode properly for

$E(x) = 2x + 6 \pmod{26}$.

Problem here: All integers 0 – 25 get mapped into just even integers, can never get an odd integer. So can't "undo" the encode.

Linear Codes

a. The sender can encode plaintext with the function

$$E(x) \equiv ax + k \bmod 26$$

where a is relatively prime with 26.

b. The receiver can decode the ciphertext with the function:

$$D(x) \equiv a^{-1}(x - k) \bmod 26$$

where a^{-1} is the multiplicative inverse of a.

c. The kibitzer has to work harder to decipher a linear code than a Caesar's code. Though (s)he can use the following strategies:

1. By using frequency analysis for very long passages of English text, the following letter-use percentages have been observed:

Letter	% Frequency
E	12.02
T	9.1
A	8.12
O	7.68
I	7.31
N	6.95
S	6.28
R	6.02
H	5.92
D	4.32
L	3.98
U	2.88
C	2.71

Letter	% Frequency
M	2.61
F	2.3
Y	2.11
W	2.09
G	2.03
P	1.82
B	1.49
V	1.11
K	0.69
X	0.17
Q	0.11
J	0.1
Z	0.07

2. By using educated guesses/trial and error, the kibitzer may be able to find a common word in the ciphertext.

3. If the kibitzer can figure out two letters in the ciphertext, say $D(x_1) \equiv x_2$ and $D(y_1) \equiv y_2$, (s)he may be able to figure out the decoding function $D(x) \equiv bx + c$ by solving the system:

$$x_2 \equiv bx_1 + c \pmod{26}$$
$$y_2 \equiv by_1 + c \pmod{26}$$

$E(x) \equiv ax + k \pmod{26}$

$D(x) \equiv a^{-1}(x - k)$

Example 2:

a) Play the role of the sender. With $a \equiv 5$, $k \equiv 5$ use the linear code to encode the word:

THIS ← plaintext

$E(T) \equiv 5 \cdot 19 + 5 \equiv -4 \equiv 22 \equiv W$

$E(H) \equiv 5 \cdot 7 + 5 \equiv 40 \equiv 14 \equiv O$

$E(I) \equiv 5 \cdot 8 + 5 \equiv 45 \equiv -7 \equiv 19 \equiv T$

$E(S) \equiv 5 \cdot 18 + 5 \equiv 95 \equiv -9 \equiv 17 \equiv R$

ciphertext

WOTR

b) Play the role of the receiver. With $a \equiv 5, k \equiv 5$ use the linear code to decipher the word:

LTII ← plaintext

$D(L) \equiv D(11) \equiv -5(11-5) \equiv -4 \equiv 22 \equiv W$

$D(T) \equiv D(19) \equiv -5(19-5) \equiv -5 \cdot 14 \equiv 8 \equiv I$

$D(I) \equiv D(8) \equiv -5(8-5) \equiv -15 \equiv 11 \equiv L$

ciphertext

WILL

$5 \cdot 5 \equiv 25 \equiv -1 \pmod{26}$

$5 \cdot (-5) \cdot 5 \equiv 1 \pmod{26}$

so $a^{-1} \equiv -5 \equiv 21 \pmod{26}$

c) Play the role of the kibitzer. The ciphertext below was taken from a very long passage of ciphertext in which the letter Z appeared most often and the letter W appeared the second most often. Given that the very long passage of ciphertext was encoded using a linear code, find the plaintext.

CFRR ← ciphertext

Let $D(x) \equiv bx + c$ and guess that $D(6) \equiv E$, $D(W) \equiv T$

Solve the system: $D(25) \equiv 4$, $D(22) \equiv 19$

$$\left.\begin{aligned} 4 &\equiv b(-1) + c \\ -7 &\equiv b(-4) + c \end{aligned}\right\}$$

Subtract:

$11 \equiv b(-1+4) \equiv 3b$ $\qquad 3^{-1} \equiv 9$

so $b \equiv 9 \cdot 11 \equiv -5 \equiv 21$, so $b \equiv -5$

$4 \equiv -b + c \equiv 5 + c$, so $c \equiv -1$

Check: $D(W) \equiv -5 \times 4 \cdot 1$ so $D(C) \equiv -11 \equiv P$

$D(F) \equiv -26 \equiv A$, $D(R) \equiv -8 \equiv S$

plaintext

PASS

- **Finding the Inverse**

If our encoding function is given by

$$E(x) \equiv ax \pmod{26}$$

then the decoding function is given by

$$D(x) \equiv \bar{a}^{1} \cdot x \pmod{26}$$

provided that $\bar{a}^{1}$ exists, i.e. provided that the integers a and 26 are relatively prime.

There are several methods to find the inverse of a mod 26.

1. Look at the equivalence class or congruence class containing the integer 1, find nonprime integers in [1]

Modulo 26:

$[1] = \{1, 27, 53, 79, 105, 131, 157, 183, 209, \ldots\}$

The nonprimes are

$1, 27, 105, 209, \ldots$

and $1 \cdot 1 = 1$, $3 \cdot 9 = 27$, $3 \cdot 5 \cdot 7 = 105$, $11 \cdot 19 = 209$

So 1 is its own inverse, 3 and 9 are inverses, 15 and 7 are inverses, and 11 and 19 are inverses modulo 26.

2. Test each multiple of a to find the inverse of a (mod 26).

$15 \cdot 1 \equiv 15 \not\equiv 1 \pmod{26}$

$15 \cdot 3 \equiv 45 \not\equiv 1 \pmod{26}$

$15 \cdot 5 \equiv 75 \not\equiv 1 \pmod{26}$

$15 \cdot 7 \equiv 105 \equiv 1 \pmod{26}$

So the inverse of 15 modulo 26 is 7, i.e.

$15^{-1} \equiv 7 \pmod{26}$

(No need to test 2, 4, 6, 8, .. since they don't have inverses modulo 26).

3. Use the Euclidean algorithm.

- Usually, this is used to find the greatest common divisor of two integers.

  Apply the division algorithm repeatedly to find the g.c.d. (26, 7)

At each point in the algorithm, we have an equation of the form

$$\boxed{m = q \cdot n + r}$$

obtained when m is divided by n to get the quotient q and the least nonnegative remainder r

If both m and n are divisible by a positive integer d, then both n and r are divisible by d.

So $$\gcd(m, n) = \gcd(n, r)$$

The next step is to divide n by r to get

$$n = s \cdot r + t$$

so that $\gcd(n, r) = \gcd(r, t)$, ...

Continue until the remainder is 0.

The last nonzero remainder is $\gcd(m, n)$, so we find:

$$26 = 3 \cdot 7 + 5$$
$$7 = 1 \cdot 5 + 2$$
$$5 = 2 \cdot 2 + \textcircled{1}$$
$$2 = 2 \cdot 1 + 0$$

So $\gcd(26, 7) = 1$.

Now start from the bottom up to get

$$1 = 5 \cdot 26 + 19 \cdot 7$$

$$26 = 3 \cdot 7 + 5$$

$$7 = 1 \cdot 5 + 2$$

$$5 = 2 \cdot 2 + 1 \leftarrow$$

$$2 = 2 \cdot 1 + 0$$

Write

$$1 = 5 - 2 \cdot 2 = 5 - 2 \cdot (7 - 1 \cdot 5) = 3 \cdot 5 - 2 \cdot 7$$

$$= 3 \cdot (26 - 3 \cdot 7) - 2 \cdot 7 = 3 \cdot 26 - 11 \cdot 7$$

So

$$3 \cdot 26 - 11 \cdot 7 = 1$$

Get $-11 \cdot 7 \equiv 1 \pmod{26}$

So $7^{-1} \equiv -11 \equiv -11 + 26 \equiv 15 \pmod{26}$

Called the extended Euclidean algorithm.