Introduction to Classical Encryption Techniques

Dr. Arup Kumar Pal

Department of Computer Science & Engineering
Indian Institute of Technology (Indian School of Mines),
Dhanbad, Jharkhand-826004
E-mail: arupkrpal@iitism.ac.in

Symmetric Encryption

Mathematically:

Encryption: $C = E_K(P)$

Decryption: $P = D_K(C)$

Where P = plaintext and C = ciphertext

K = secret key

E = encryption algorithm

D = decryption algorithm

Both E and D are known to public

Traditional Ciphers

- Substitution Ciphers : A substitution cipher replaces one symbol with another.
  - Monoalphabetic Ciphers
  - Polyalphabetic Ciphers
- Transposition Ciphers: A transposition cipher reorders (permutes) symbols in a block of symbols

Caesar/Additive Cipher

- Invented by Julius Caesar and the earliest known substitution cipher.
- Each letter is replaced by the letter three positions further down the alphabet.

Plaintext: a b c d e f g h i j k l m n o p q r s t u v w x y z

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

- Example: ISM Dhanbad → OVP GKDQEDG

Contd...

- Mathematically, map letters to numbers:

  a, b, c, ..., x, y, z

  0, 1, 2, ..., 23, 24, 25

- Then the general Caesar cipher is:

  $C = \mathrm{E}_K(P) = (P + k) \bmod 26$

  $P = \mathrm{D}_K(C) = (C - k) \bmod 26$

Cryptanalysis

Objective: to recover the plaintext of a ciphertext or, more typically, to recover the secret key.

Kerckhoff's principle: the adversary knows all details about a cryptosystem except the secret key.

Two general approaches:

- brute-force attack
- non-brute-force attack (cryptanalytic attack)

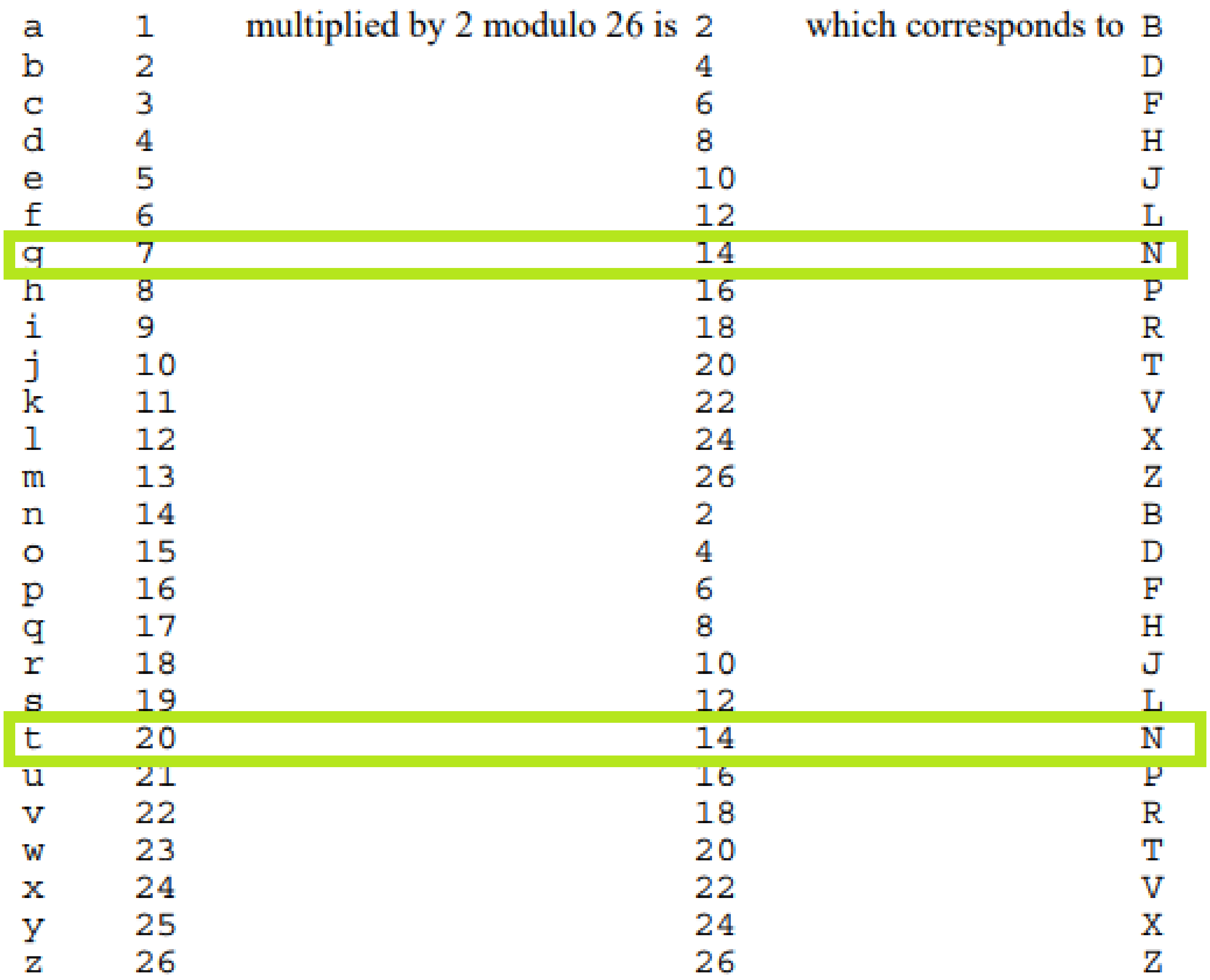
Cryptanalysis of Additive Cipher

- Not enough keys.
  - If we shift a letter 26 times, we get the same letter back
  - A shift of 27 is the same as a shift of 1, etc.
  - So we only have 25 keys (1 to 25)

    Key space: {0, 1, ..., 25}
- Eve just tries every key until she finds the right one so vulnerable to brute-force attacks.

Multiplicative Cipher

- The plaintext is encrypted by multiplication operation.
- The decryption algorithm specifies division of the ciphertext by the key.
- Decryption is not always possible for any key.



a	1	multiplied by 2 modulo 26 is	2	which corresponds to	B
b	2		4		D
c	3		6		F
d	4		8		H
e	5		10		J
f	6		12		L
g	7		14		N
h	8		16		P
i	9		18		R
j	10		20		T
k	11		22		V
l	12		24		X
m	13		26		Z
n	14		2		B
o	15		4		D
p	16		6		F
q	17		8		H
r	18		10		J
s	19		12		L
t	20		14		N
u	21		16		P
v	22		18		R
w	23		20		T
x	24		22		V
y	25		24		X
z	26		26		Z

- The encryption is

$$C = (K \times P) \bmod 26$$

- The Decryption will be

$$P = (K^{-1} \times C) \bmod 26$$

- One-to-one mapping is possible when key is coprime with modulo i.e. gcd(K,26)=1

Modular Arithmetic

The set of integers, denoted by Z, contains all integral numbers (with no fraction) from negative infinity to positive infinity

$$\boxed{\mathbf{Z} = \{\, \ldots, -2, -1, 0, 1, 2, \ldots \,\}}$$

The modulo operation creates a set, which in modular arithmetic is referred to as the set of **least residues modulo n, or Zn**.

$$\boxed{\mathbf{Z}_n = \{\, 0, 1, 2, 3, \ldots, (n-1) \,\}}$$

$$\boxed{\mathbf{Z}_2 = \{\, 0, 1 \,\}} \quad \boxed{\mathbf{Z}_6 = \{\, 0, 1, 2, 3, 4, 5 \,\}} \quad \boxed{\mathbf{Z}_{11} = \{\, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 \,\}}$$

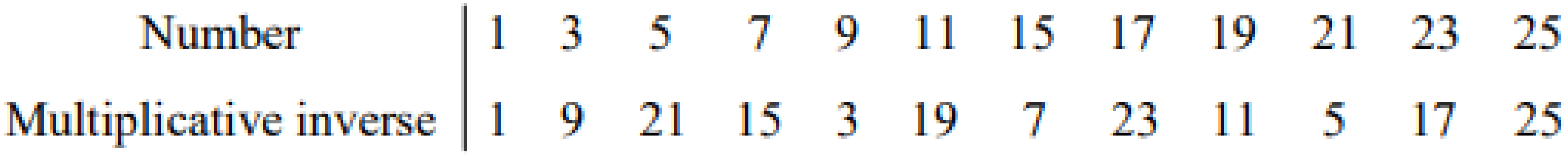
In Zn, two numbers a and b are the multiplicative inverse of each other if

$$a \times b \equiv 1 \pmod{n}$$

Multiplicative Inverse

- The key domain for any multiplicative cipher under modulo 26:

1,3,5,7,9,11,15,17,19,21,23,25



Number	1	3	5	7	9	11	15	17	19	21	23	25
Multiplicative inverse	1	9	21	15	3	19	7	23	11	5	17	25

Affine Cipher

- In additive shift cipher and multiplicative cipher are not secure.
- There are only 25 possible shifts for English alphabet and the letters are not mixed enough.
- The Affine cipher is done as follow:

$$C = (\alpha P + \beta) \bmod 26$$

With condition $\gcd(\alpha, 26) = 1$

The decryption is done as follow

$$P = (\alpha^{-1}(C - \beta)) \bmod 26$$

Cryptanalysis of Affine Cipher

- The key for this encryption method is the pair (α,β).
- There are 12 possible choice for α with $\gcd(\alpha,26)=1$
- 26 choices for β .
- Total choice=$12 \times 26=312$.

Contd...

- Ciphertext attack:
  - An exhaustive search through all 312 keys will take longer than additive shift cipher.
- Known Plaintext:

  Suppose the plaintext is *if* and ciphertext is PQ

  8(=i) maps to 15(P) will yields the equation

  $$8\alpha + \beta \equiv 15 \pmod{26}$$

  5(=f) maps to 15(Q) will yields the equation

  $$5\alpha + \beta \equiv 16 \pmod{26}$$

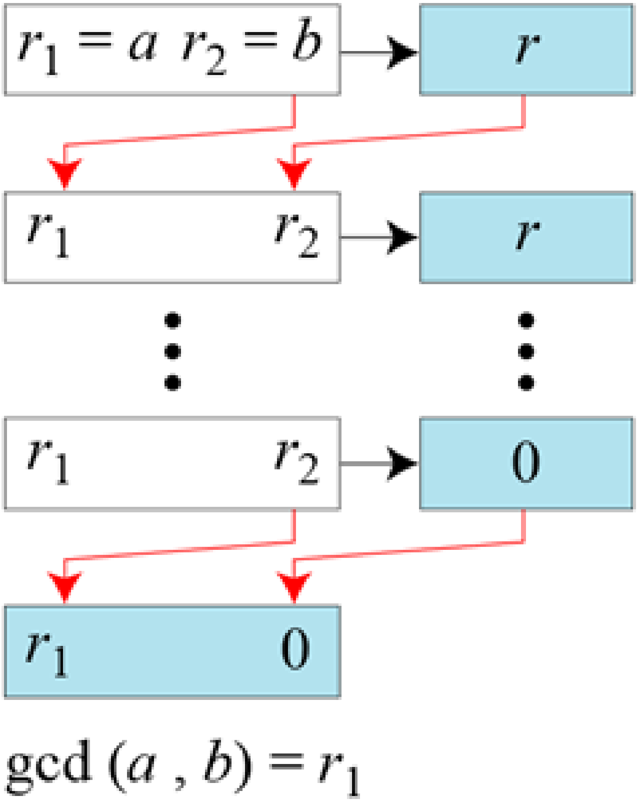
- Chosen Plaintext:
  - If ab is the plaintext then what will be the key.
- Chosen the Ciphertext:
  - Solve the plaintext and ciphertext pair with decryption algorithm.

Euclidean Algorithm

Rule 1: gcd (a, 0) = a

Rule 2: gcd (a, b) = gcd (b, r)

where r is the remainder of dividing a by b



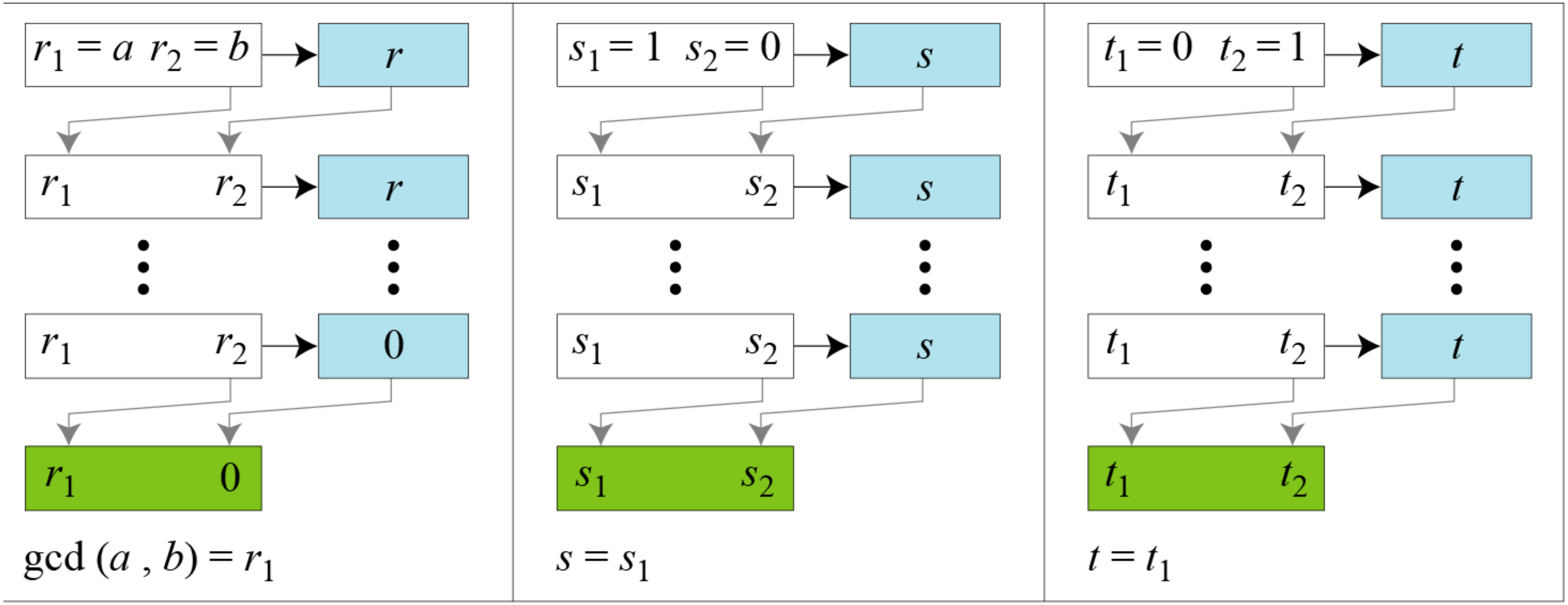
Extended Euclidean Algorithm

Given two integers a and b, we often need to find other two integers, s and t, such that

$$s \times a + t \times b = \gcd(a, b)$$

The extended Euclidean algorithm can calculate the gcd(a,b) and at the same time calculate the value of s and t.

Contd...



where

$r \leftarrow r_1 - q \times r_2;$ (Updating r's)

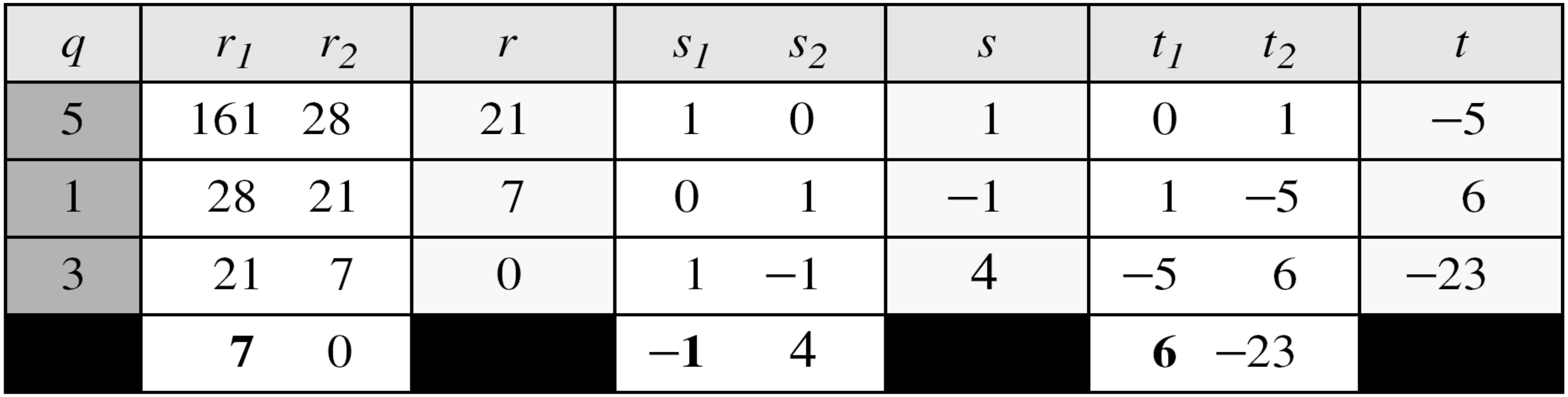
$s \leftarrow s_1 - q \times s_2;$ (Updating s's)

$t \leftarrow t_1 - q \times t_2;$ (Updating t's)

Example

Given a = 161 and b = 28, find gcd (a, b) and the values of s and t.

We get gcd (161, 28) = 7, s = −1 and t = 6.



q	r_1	r_2	r	s_1	s_2	s	t_1	t_2	t
5	161	28	21	1	0	1	0	1	−5
1	28	21	7	0	1	−1	1	−5	6
3	21	7	0	1	−1	4	−5	6	−23
	7	0		**−1**	4		**6**	−23	

Multiplicative inverse

$$s \times a + t \times b = \gcd(a, b)$$

$(s\times n)+(t\times a)$=gcd(n,a)=1

$((s\times n)+(t\times a))$modn=1mod n

(0+$(t\times a)$mod n=1

$(t\times a)$mod n=1

This means t is the multiplicative inverse of a

Congruence

- Two numbers ***a*** and ***b*** are said to be "*congruent modulo **n***" if

  a mod n = b mod n

  implies ***a ≡ b(mod n)***

- The difference between ***a*** and ***b*** will be a multiple of ***n***

  So ***a-b = kn*** for some value of ***k***

If $a \equiv 0 \pmod{n}$*, then* $n|a$*.*

Properties of Congruences

1. $a \equiv b \pmod{n}$ *if* $n|(a-b)$
2. $a \equiv b \pmod{n}$ *implies* $b \equiv a \pmod{n}$
3. $a \equiv b \pmod{n}$ *and* $b \equiv c \pmod{n}$ **imply** $a \equiv c \pmod{n}$

Proof of 1.

If $n|(a-b)$, then $(a-b) = kn$ for some k. Thus, we can write

$a = b + kn$. Therefore,

$(a \bmod n)$ = (remainder when $b + kn$ is divided by n) =
(remainder when b is divided by n) = $(b \bmod n)$.

Modular Arithmetic

<u>**Claim:**</u> If $a \equiv b \bmod m$ and $c \equiv d \bmod m$

$$ac \equiv bd \bmod m$$

<u>**Proof:**</u> $a - b = k_1 m$ and $c - d = k_2 m$

$$ac = (b + k_1 m)(d + k_2 m) =$$

$$= bd + (k_1 d + k_2 b)m + k_1 k_2 m^2 =$$

$$= bd + (k_1 d + k_2 b + k_1 k_2 m)m$$

$$ac - bd = (k_1 d + k_2 b + k_1 k_2 m)m$$

Monoalphabetic Substitution Cipher

- Rather than having a fixed shift change, every plaintext letter to an arbitrary ciphertext letter
- To decrypt we just look up the ciphertext letter in the table and then write down the matching plaintext letter

Plaintext	Ciphertext
a	G
b	X
c	N
d	S
e	D
…	…
z	Q

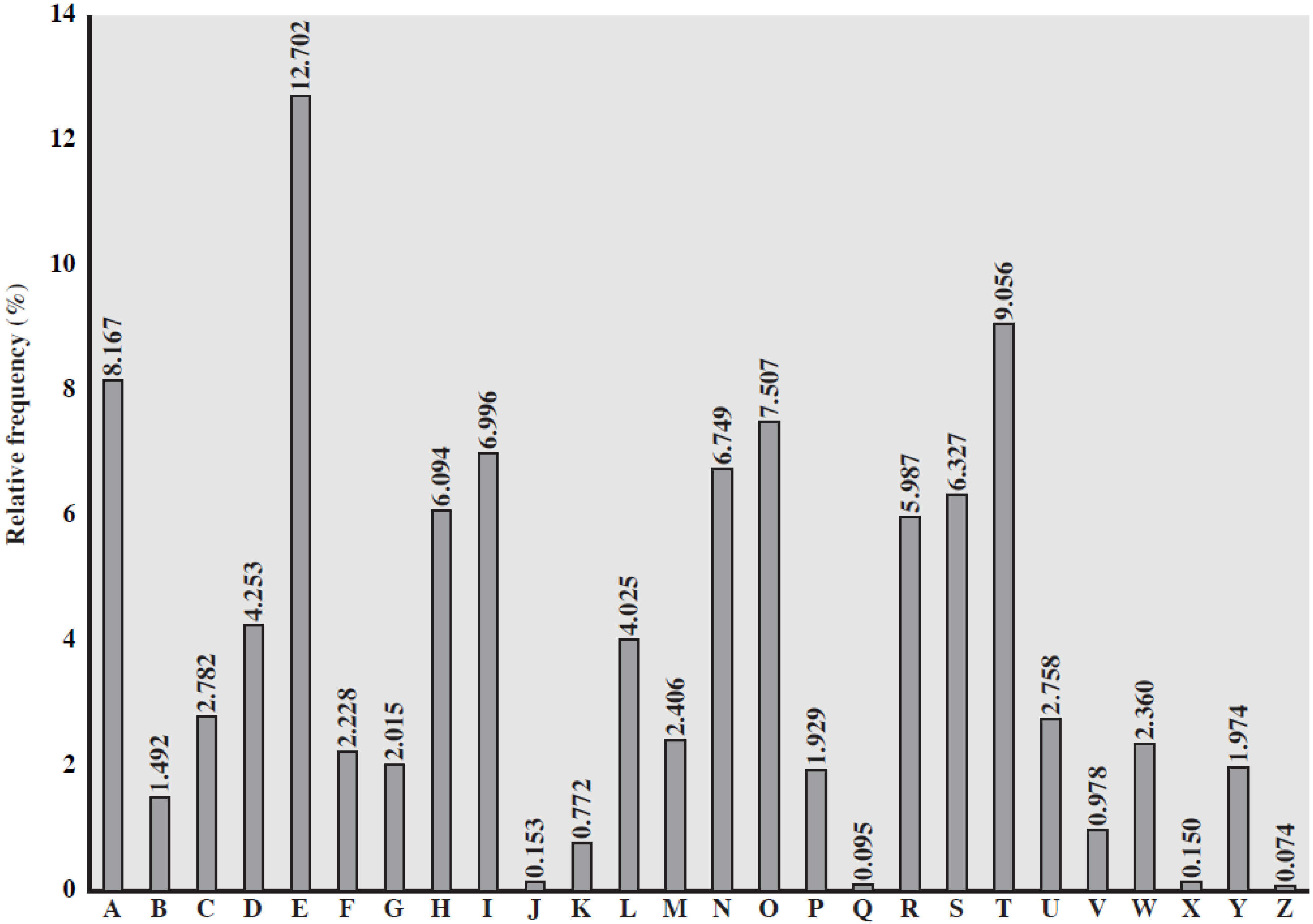
Contd…

- Now we have a total of 26! = 4 x 10^{26} keys.
- With so many keys, it is secure against brute-force attacks.
- But not secure against Frequency Analysis.

Frequency Analysis

- Human languages are not random.
- In English (or any language) certain letters are used more often than others.
- In English, E is by far the most common letter, followed by T, R, N, I, O, A, S.
- If we look at a ciphertext, certain ciphertext letters are going to appear more often than others.
- It would be a good guess that the letters that occur most often in the ciphertext are actually the most common English letters.

English Letter Frequencies



Statistics for double & triple letters

- In decreasing order of frequency

Double letters:

th he an in er re es on, …

Triple letters:

the and ent ion tio for nde, …

Frequency Analysis in Practice

- Suppose this ciphertext

  dq lqwurgxfwlrq wr frpsxwlqj surylglqj d eurdg vxuyhb ri wkh glvflsolqh dqg dq lqwurgxfwlrq wr surjudpplqj. vxuyhb wrslfv zloo eh fkrvhq iurp: ruljlqv ri frpsxwhuv, gdwd uhsuhvhqwdwlrq dqg vwrudjh, errohdq dojheud, gljlwdo orjlf jdwhv, frpsxwhu dufklwhfwxuh, dvvhpeohuv dqg frpslohuv, rshudwlqj vbvwhpv, qhwzrunv dqg wkh lqwhuqhw, wkhrulhv ri frpsxwdwlrq, dqg duwlilfldo lqwhooljhqfh.

Frequency Analysis in Practice

- Count relative letter frequencies in ciphertext.
- Compare this distribution against the known one.
- For example, there is a good chance that the most frequently occurred character from ciphertext will map corresponds to e, and so on
- Proceeding with trial and error to finally get the probable plain text.

Playfair Cipher

- Not even the large number of keys in a monoalphabetic cipher provides security.
- One approach to improving security is to encrypt multiple letters at a time.
- The **Playfair Cipher** is the best known such cipher.

Playfair Key Matrix

- Use a 5 x 5 matrix.
- Fill in letters of the key (without duplicates).
- Fill the rest of matrix with other letters.
- E.g., key = *playfair*

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

Encrypting and Decrypting

Plaintext is encrypted two letters at a time.

1. If a pair is a repeated letter, insert filler like 'X’.
2. If both letters fall in the same row, replace each with the letter to its right (circularly).
3. If both letters fall in the same column, replace each with the letter below it (circularly).
4. Otherwise, each letter is replaced by the letter in the same row but in the column of the other letter of the pair.

Example

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

et becomes MN

me becomes EG

qa becomes WB

Security of Playfair Cipher

- Equivalent to a monoalphabetic cipher with an alphabet of 26 x 26 = 676 characters.
- Security is much improved over the simple monoalphabetic cipher.
- Actually, it **can** be broken, because it still leaves some structure of plaintext intact.
- Also, unless the keyword is long, the last few rows of the matrix are predictable.

Polyalphabetic Ciphers

- The mono alphabetic ciphers do not change the frequency of characters in the cipher text.
- It improve security using multiple cipher alphabets
- Make cryptanalysis harder with more alphabets to guess and flatter frequency distribution
- Use a key to select which alphabet is used for each letter of the message

Vigenère Cipher

- Simplest polyalphabetic substitution cipher
- Consider the set of all Caesar ciphers:

  $\{ C_a, C_b, C_c, \ldots, C_z \}$

- Key: e.g. security
- Encrypt each letter using C_s, C_e, C_c, C_u, C_r, C_i, C_t, C_y in turn.
- Repeat from start after C_y.
- Decryption simply works in reverse.

Example of Vigenère Cipher

- Keyword: *deceptive*

```
key:         deceptivedeceptivedeceptive
plaintext: wearediscoveredsaveyourself
ciphertext: ZICVTWQNGRZGVTWAVZHCQYGLMGJ
```

Security of Vigenère Ciphers

- There are multiple (how many?) ciphertext letters corresponding to each plaintext letter.
- So, letter frequencies are obscured but not totally lost.
- To break Vigenere cipher:

1. Try to guess the key length. How?
2. If key length is N, the cipher consists of N Caesar ciphers.

   Plaintext letters at positions k, N+k, 2N+k, 3N+k, etc., are encoded by the same cipher.
3. Attack each individual cipher as before.

Guessing the Key Length

- Main idea: Plaintext words separated by multiples of the key length are encoded in the same way.
- In our example, if plaintext = "…thexxxxxxthe…" then "the" will be encrypted to the same ciphertext words.
- So look at the ciphertext for repeated patterns.
- E.g. repeated "VTW" in the previous example suggests a key length of 3 or 9:

```
ciphertext:   ZICVTWQNGRZGVTWAVZHCQYGLMGJ
```

- Of course, the repetition could be a random fluke.

Hill Cipher

- In the shift, affine, and substitution ciphers, a given letter in the ciphertext always comes from exactly one letter in the plain text.
- This greatly facilitates finding the key using frequency analysis.
- In Vigenère Ciphers, the use of blocks of letters, corresponding to the length of the key, made the frequency analysis more difficult, but still possible, since **there was no interaction among the various letters in each block**.

Contd…

- **Block ciphers** avoid these problems by encrypting blocks of several letters or numbers simultaneously.
- A change of one character in a plaintext block should change potentially all the characters in the corresponding ciphertext block.
- Although Playfair is an example of block cipher, but blocks of two letters are too small to be secure.

Contd…

- Many modern cryptosystems are type of block ciphers.
- Symmetric cryptosystems like DES (Block of 64 bits), and AES (Block of 128 bits)
- Asymmetric cryptosystems like RSA uses blocks several bits long, depending on the modulus used.
- Hill Cipher is an example of block cipher.

Hill Cipher

- The sender and receiver must first agree upon a key matrix *A* of size *n* x *n*.
- The key must be invertible under mod 26.
- Encryption: $C = \left(P \times K_{n\times n}\right) \bmod 26$
- Decryption: $P = \left(C \times K_{n\times n}{}^{-1}\right) \bmod 26$

Example

Encryption

$$K_{3\times3}=\begin{pmatrix}1 & 2 & 3\\ 4 & 5 & 6\\ 11 & 9 & 8\end{pmatrix}$$

$$P=abc=\begin{pmatrix}0 & 1 & 2\end{pmatrix}$$

$$\begin{aligned}C &\equiv (P\times K)\bmod 26\\ &\equiv \begin{pmatrix}0 & 1 & 2\end{pmatrix}K \bmod 26\\ &\equiv \begin{pmatrix}0 & 23 & 22\end{pmatrix}\bmod 26\end{aligned}$$

$$C=\begin{pmatrix}0 & 23 & 22\end{pmatrix}=AXW$$

Decryption

- In order to decrypt, we need decryption key K^{-1} i.e. $KK^{-1}\equiv 1 \bmod 26$
- The determinant of K to satisfy such $\gcd(\det(K),26)=1$

Transposition Ciphers

- Also called **permutation** ciphers.
- Shuffle the plaintext, without altering the actual letters used.
- Example: Row Transposition Ciphers

Row Transposition Ciphers

- Plaintext is written row by row in a rectangle.
- Ciphertext: write out the columns in an order specified by a key.

Key: 3 4 2 1 5 6 7

Plaintext:

a	t	t	a	c	k	p
o	s	t	p	o	n	e
d	u	n	t	i	l	t
w	o	a	m	x	y	z

Ciphertext: TTNAAPTMTSUOAODWCOIXKNLYPETZ

Diffusion and Confusion

- Shannon introduced two fundamental properties of a effective cryptosystems.
- In order to thwart statistical analysis: **diffusion** and **confusion** are used.
- Diffusion mean that is we change a character of the plaintext, then several characters of the ciphertext should change, and , similarly, if we change a character of the ciphertext, then several characters of the plaintext should change.
- Hill Cipher has this property.

Contd...

- Confusion means that the key does not relate in a simple way to the ciphertext. In particular, each character of the ciphertext should depend on several parts of the key.
- The Vigenère and substitution ciphers do not have the properties of diffusion and confusion, which is why they are vulnerable to frequency analysis attack.
- The concepts of diffusion and confusion play an important role in any well-designed block cipher.
- A disadvantage of diffusion is error propagation.

Conclusions

- These algorithms are simpler than modern ciphers and easier to understand.
- These are no longer highly secure in current era.
- These are the basic foundation of cryptography to understand the modern cryptography.