Substitution Techniques

Dr. E. Suresh Babu

Assistant Professor

Department of CSE

National Institute of Technology, Warangal

Outline

- Substitution Techniques
- Types of Substitution Cipher
- Mono-alphabetic Cipher
  - ✓ Caesar Cipher
  - ✓ Analysis of Caesar Cipher
  - ✓ Monoalphabetic Substitution Ciphers

Operations on Encryption Algorithm

- All **Encryption Algorithms** are based on **two general principles:**
  - ✓ **Substitution**
  - ✓ **Transposition**
- **Substitution** : Each element in the **Plaintext (Bit, Letter, Group of Bits Or Letters)** is mapped into **Another Element**,
- **Transposition** : Each elements in the **plaintext** are **rearranged.**

Substitution Techniques

Introduction

- Before Computers, **Cryptography** consisted of **Character-based Algorithms.**
- The **better algorithms** always use both **Substitution** and **Transposition**

Substitution Cipher

- A **Substitution Cipher** is one in which each **character in the plaintext** is **SUBSTITUTED** for **another character in the ciphertext**.
- The **receiver inverts** the **substitution** on the **ciphertext** to **recover the plaintext.**

Types of Substitution Cipher

Types of Substitution Cipher

❖ In **Classical Cryptography**, there are **Four Types of Substitution Ciphers**:

1. **A Simple Substitution Cipher or Mono-alphabetic Cipher.**
2. **A Homophonic Substitution Cipher.**
3. **A Polygram Substitution Cipher.**
4. **A Polyalphabetic Substitution Cipher.**

Mono-alphabetic Cipher

Mono-alphabetic Cipher

- A **Mono-alphabetic Cipher**.,
  - ✓ Each **character of the plaintext** is **replaced uniquely** with a **corresponding character** of **cipher text.**
- The **famous Caesar Cipher** is falls under **simple substitution cipher**

Types of Mono-alphabetic Cipher

Types of Mono-alphabetic Cipher

- A **Mono-alphabetic Cipher** falls under **three categories**
  - ✓ **Caesar Cipher**
  - ✓ **Substitution Cipher**
  - ✓ **The Affine Cipher**

Caesar Cipher

Caesar Cipher

- **Caesar Cipher** is also called as **Shift Cipher** which is Primitive(basic) Cipher
- **Caesar substitution Cipher** was introduced by **Julius Caesar.**

Working Model of Caesar Cipher

- The **Caesar cipher** involves **replacing each letter of the alphabet** with the **Unique letter** standing **THREE places further down the alphabet.**

Working Model of Caesar Cipher

- Let us Consider the set

  Z_{26} = {0,1,2,3,........} usually represented with Alphabets **A-Z**

- Let us also consider the tuples

  P=C=K= Z_{26} Here K is the Piece of Information called **KEY**

  which also contains 26 Character(0<K<25)

Simplified the Caesar Cipher

- We can define the **transformation by listing all possibilities**, as follows:

Plain Text																									
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
Cipher Text																									
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**

- Let us assign a **numerical equivalent** to each letter:

Plain Text																									
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
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

Working Model of Caesar Cipher

- We define **Encryption Function** with help of **KEY**

$$E_K(x) = x + K \bmod 26$$

- We define **Decryption Function** with help of **KEY**

$$D_K(y) = y - K \bmod 26$$

- Finally We define **Caesar Cipher** with help of **KEY**

$$P=D_K(E_k(x))$$

Example of Working Model

- **P = {A=0,B=1,C=2,D=3.....................}**
- **K = {A=0,B=1,C=2,D=3.....................}**
- We define **Encryption Function** with help of **KEY**

$E_K(x) = x + K \bmod 26$

$E_K(x) = 0(A) + 3(D) \bmod 26 = 3(D)$ is a Cipher text

Example of Working Model

- **We define Encryption Function with help of KEY**

$D_K(y) = y - K \bmod 26$

$E_K(x) = 3(D) - 3(D) \bmod 26 = 0(A)$ is a Plain text

Analysis of Caesar Cipher

Analysis of Caesar Cipher

- **Plain Text** : **meet me after the toga party**
- **Cipher Text** : **PHHW PH DIWHU WKH WRJD SDUWB**

Analysis of Caesar Cipher

If it is **Cryptanalyst** knows that a given **Cipher text** is a **Caesar Cipher**,

- ✓ **Brute-force Cryptanalysis** is easily performed
- ✓ Simply try all the **25 possible keys**

Characteristics of Caesar Cipher

- **Three important characteristics** of this **Caesar Cipher** enabled to use a **brute force cryptanalysis**:

    1. The **encryption and decryption algorithms** are **known.**
    2. There are **only 25 keys** to try.
    3. The language of the **plaintext is known** and **easily recognizable**

What makes brute-force cryptanalyst impractical

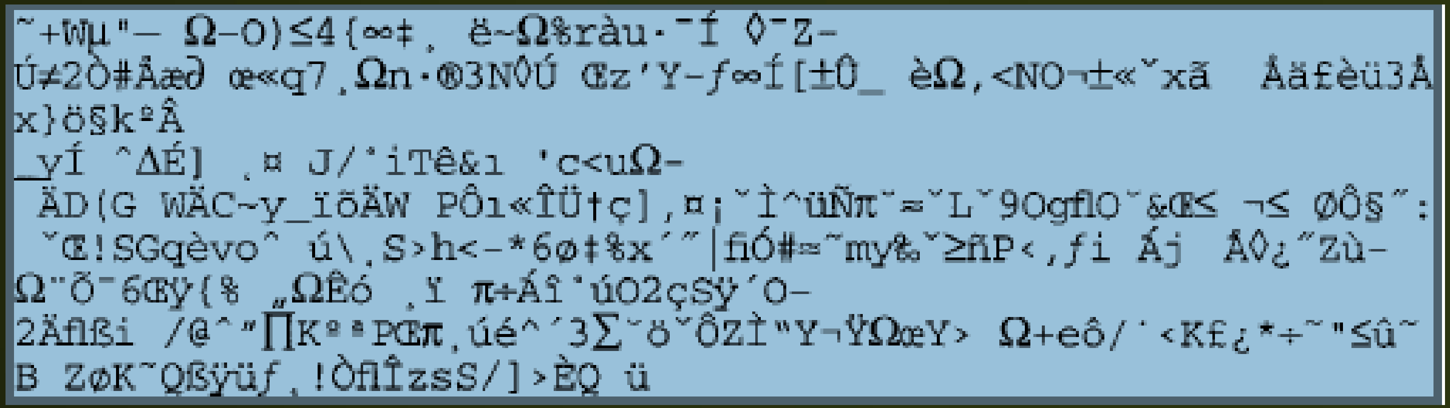
- **Caesar Cipher Algorithm** should employs a **large number of keys**.
- For example, the **triple DES algorithm,** makes use of a **168-bit key,**
- The key space of **2^{168} or 3.7 * 10^{50} possible keys.**

Brute-Force Cryptanalysis of Caesar Cipher

KEY	PHHW PH DIWHU WKH WRJD SDUWB
1	oggv og chvgt vjg vqic rctva
2	nffu nf bgufs uif uphb qbsuz
3	meet me after the toga party
4	ldds ld zesdq sgd snfz ozqsx
5	kccr kc ydrcp rfc rmey nyprw
6	jbbq jb xcqbo qeb qldx mxoqv
7	iaap ia wbpan pda pkcw lwnpu
8	hzzo hz vaozm ocz ojbv kvmot
9	gyyn gy uznyl nby niau julns
10	fxxm fx tymxk max mhzt itkmr
11	ewwl ew sxlwj lzw lgys hsjlq
12	dvvk dv rwkvi kyv kfxr grikp
13	cuuj cu qvjuh jxu jewq fqhjo
14	btti bt puitg iwt idvp epgin
15	assh as othsf hvs hcuo dofhm
16	zrrg zr nsgre gur gbtn cnegl
17	yqqf yq mrfqd ftq fasm bmdfk
18	xppe xp lqepc esp ezrl alcej
19	wood wo kpdob dro dyqk zkbdi
20	vnnc vn jocna cqn cxpj yjach
21	ummb um inbmz bpm bwoi xizbg
22	tlla tl hmaly aol avnh whyaf
23	skkz sk glzkx znk zumg vgxze
24	rjjy rj fkyjw ymj ytlf ufwyd
25	qiix qi ejxiv xli xske tevxc

Observation

- The third characteristic is also **significant.** If the language of the **plaintext is unknown**, then plaintext **output may not be recognizable**



Sample of Compressed Text

Limitations of Caesar Ciphers

- **Caesar cipher** is far from **secure**, With **only 25 possible keys** are used, as **Cryptanalyst** can easily **deduce** the key by **Bruce-force attack.**

Monoalphabetic Substitution Ciphers

Substitution Ciphers

- One way to **increase the key space** and **improve the security of the cipher** is to allow **arbitrary substitution**.
- In this case, the **"cipher"** can be any **PERMUTATION** of the **26 alphabetic characters**.

PERMUTATION

- A **Permutation** of a **finite set of elements** is an **ordered sequence** of all the **elements of S**, with each element **appearing exactly once.**
- **For Example**
  - ✓ if S= {a, b, c} there are six permutations of abc, acb, bac, bca, cab, cba

Monoalphabetic Substitution Cipher

❖ In a **monoalphabetic cipher**, our substitution characters are a **random permutation** of the **26 letters of the alphabet**

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

❖ The **key** now is the **sequence of substitution letters**. In other words, the **key** in this case is the **actual random permutation** of the alphabet used.

Monoalphabetic Substitution Cipher

- Arbitrary **substitution of letters**
- Number of keys **$26 \times 25 \times \ldots \times 1$ = 26!** (Over 4×1026)
- Note that there are **26! permutations** of the alphabet. That is a **number larger than 4×1026.**

Advantages of Substitution Cipher

- **Substitution Cipher** will **eliminate brute-force techniques** for cryptanalysis.
  - ✓ Requires >2^{88} **Possible Keys** for **brute-force attacks** which take **zillions of years** to try out even half the keys

Limitation of Substitution Cipher

- Any **Substitution Cipher**, regardless of the **size of the key space**, can be **broken easily** with a **Statistical Attack.**

Frequency Analysis of Substitution Ciphers

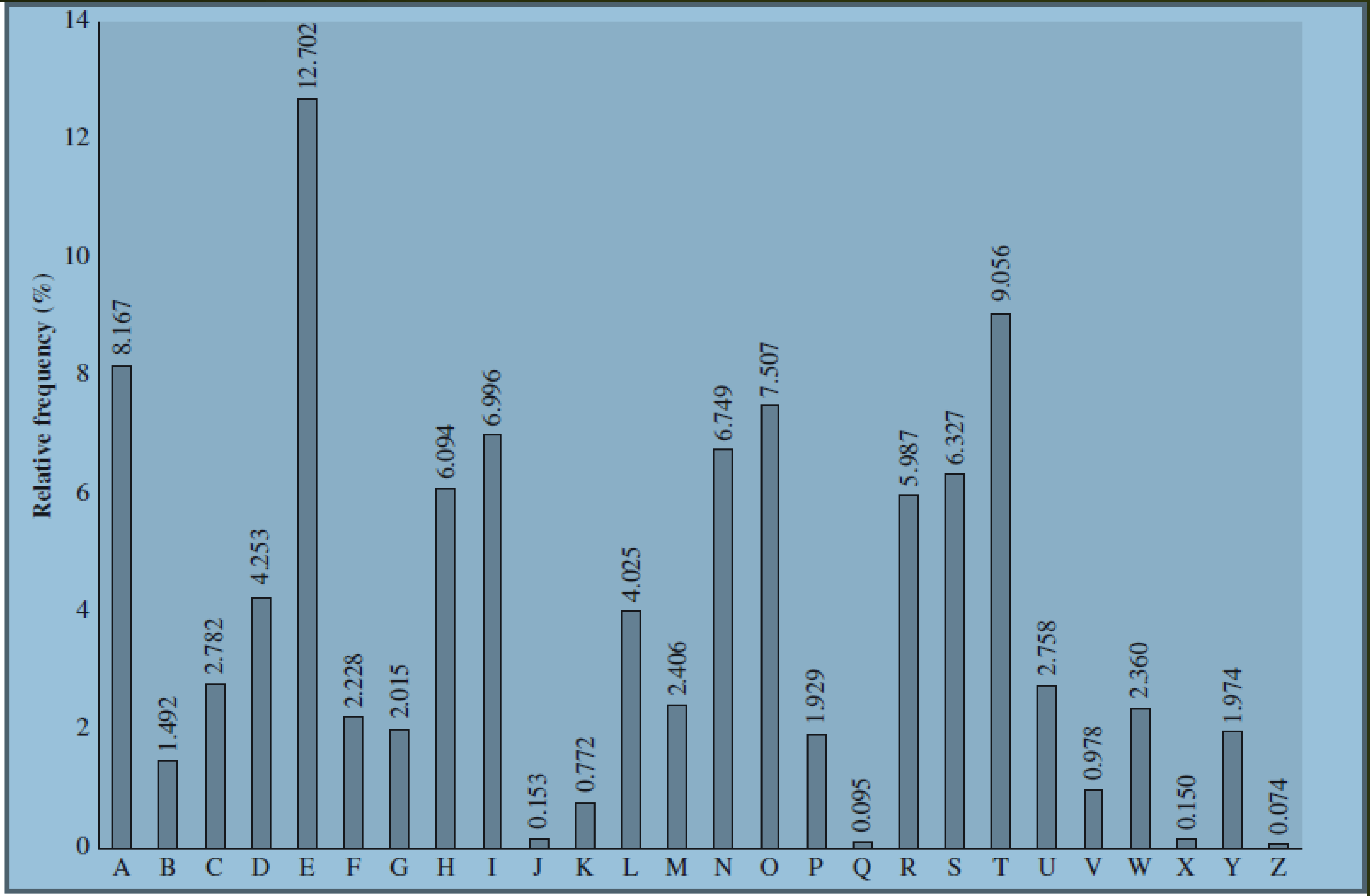
Frequency Analysis

- **Substitution Cipher** suffers **Statistics attack**,
  - ✓ If the **cryptanalyst** knows the **nature of the plaintext**, then the **analyst can exploit** the **regularity of the language**, called **FREQUENCY ANALYSIS**.
- **Frequency Analysis** studies the **frequency of letters or groups of letters** in a **ciphertext.**

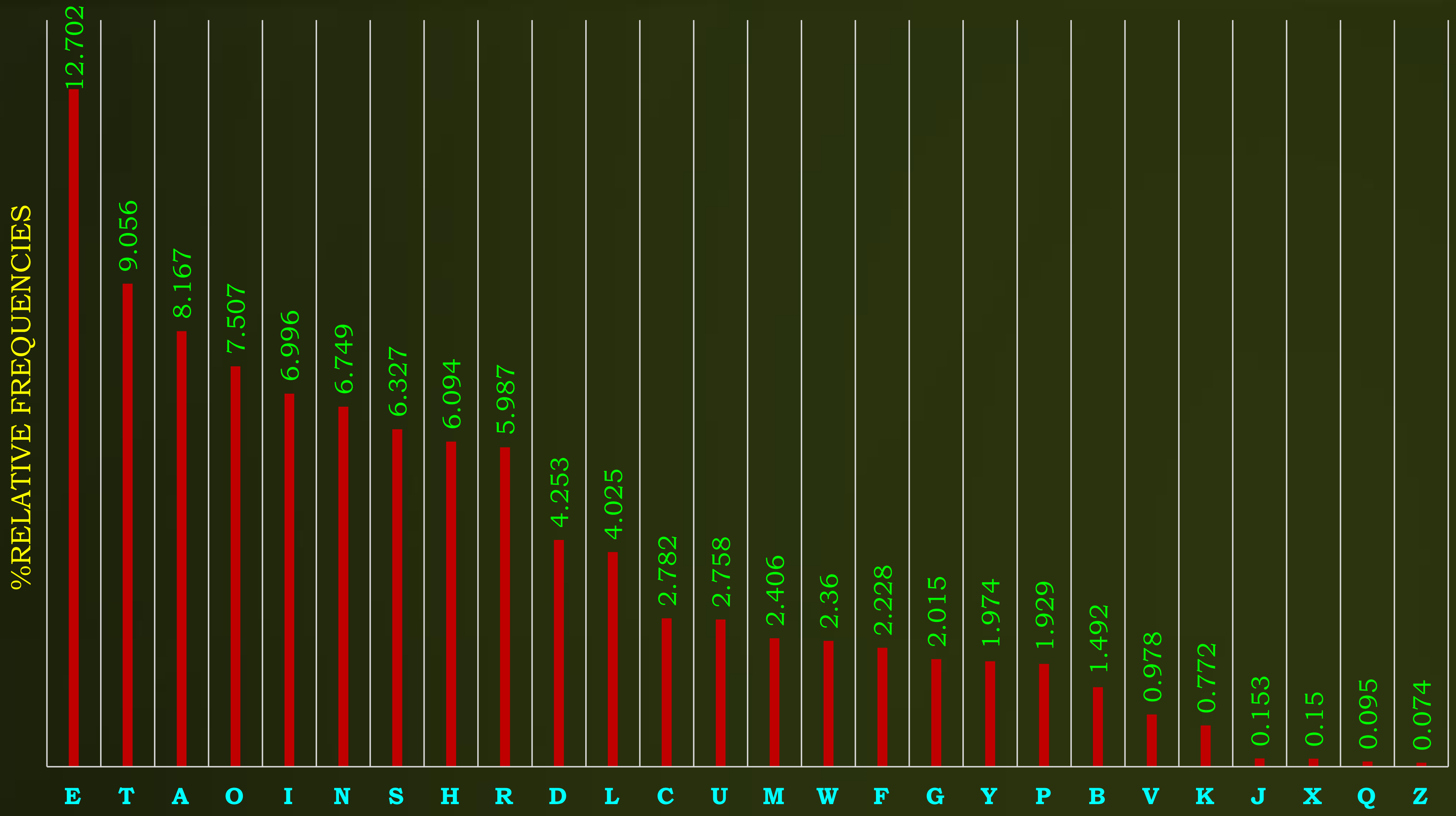
Frequency Analysis

- When the **plaintext is plain English**, a simple form of **statistical attack** consists **measuring the frequency distribution** for
  - ✓ **Single Characters,**
  - ✓ **Pairs of Characters,**
  - ✓ **Triples of Characters**, etc.,
- Comparing **all character** with **similar statistics** for English

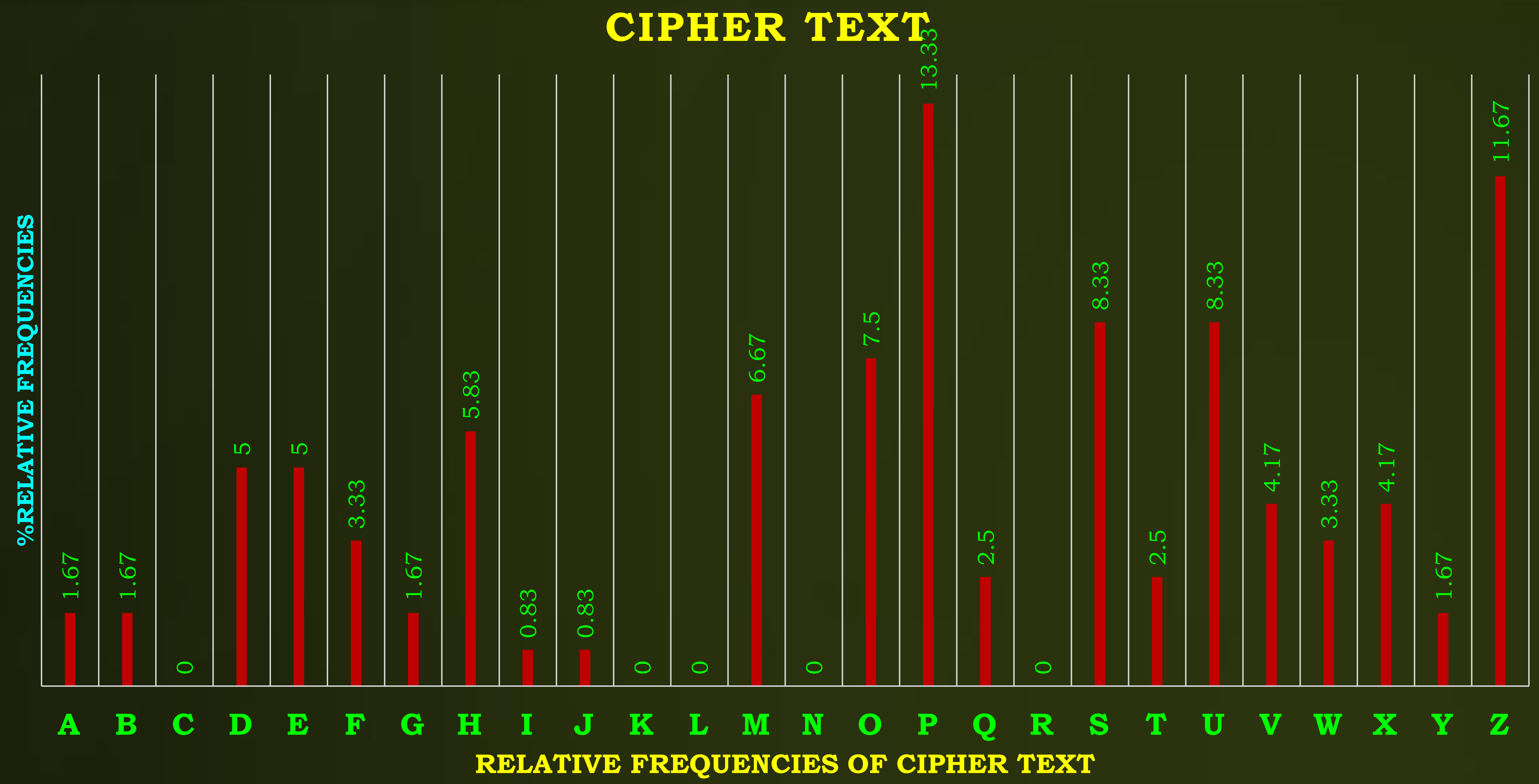
Relative Frequency of the letters of English text



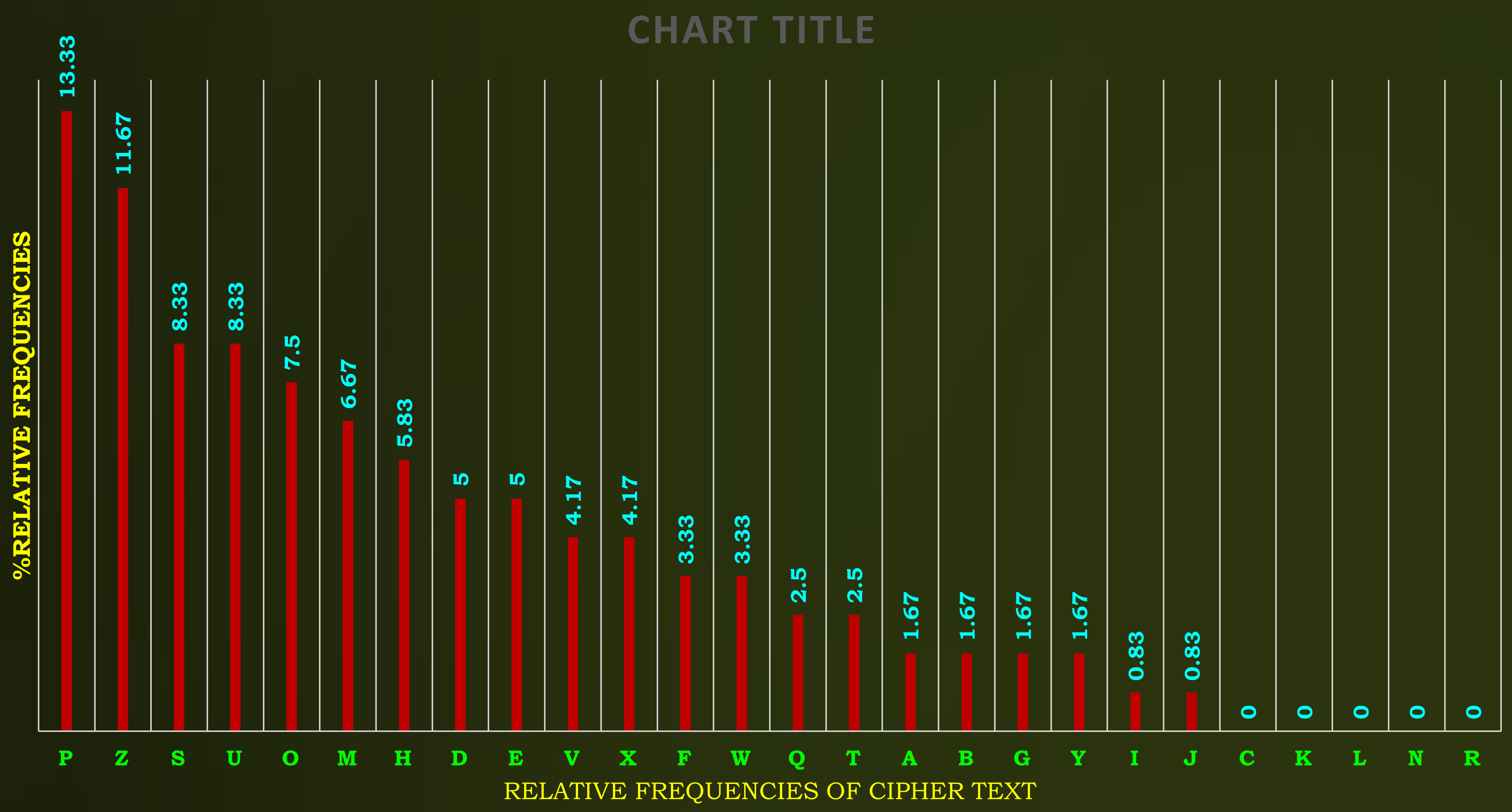
Relative Frequency of the letters of English text



The Relative Frequencies of the Letters in the Ciphertext (In Percentages)

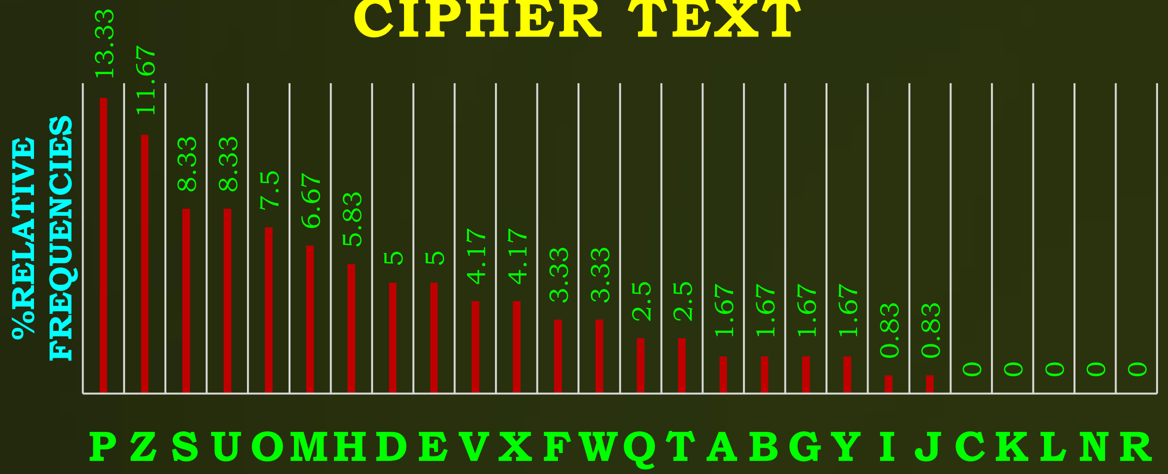


The Relative Frequencies of the Letters in the Ciphertext (In Percentages) in Decreasing Order



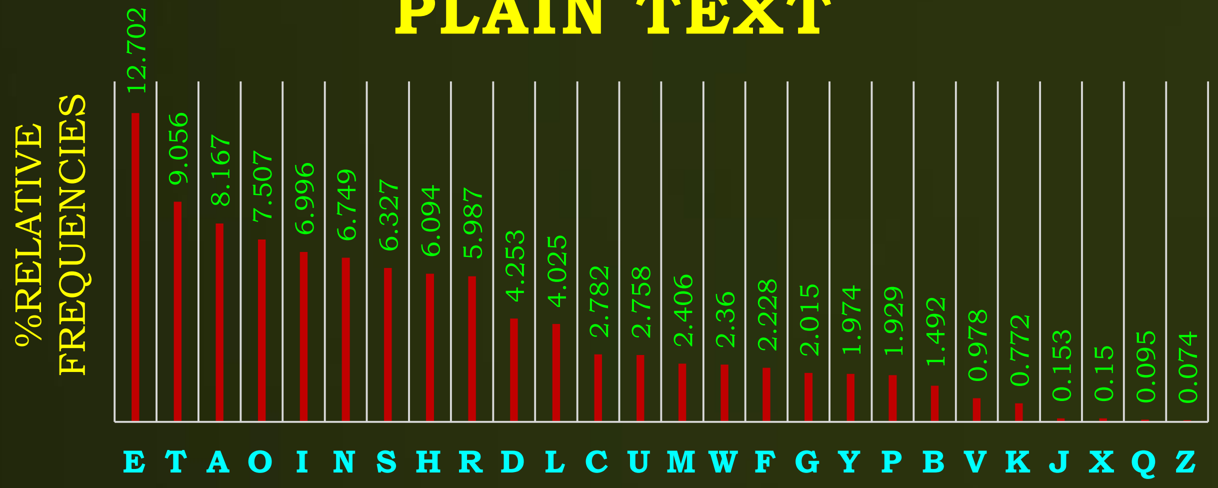
Comparing The Relative Frequencies of the Plain Text & Ciphertext

CIPHER TEXT



RELATIVE FREQUENCIES OF CIPHER TEXT

PLAIN TEXT



Digram Frequencies

digram	frequency	digram	frequency	digram	frequency	digram	frequency
th	3.15	to	1.11	sa	0.75	ma	0.56
he	2.51	nt	1.10	hi	0.72	ta	0.56
an	1.72	ed	1.07	le	0.72	ce	0.55
in	1.69	is	1.06	so	0.71	ic	0.55
er	1.54	ar	1.01	as	0.67	ll	0.55
re	1.48	ou	0.96	no	0.65	na	0.54
es	1.45	te	0.94	ne	0.64	ro	0.54
on	1.45	of	0.94	ec	0.64	ot	0.53
ea	1.31	it	0.88	io	0.63	tt	0.53
ti	1.28	ha	0.84	rt	0.63	ve	0.53
at	1.24	se	0.84	co	0.59	ns	0.51
st	1.21	et	0.80	be	0.58	ur	0.49
en	1.20	al	0.77	di	0.57	me	0.48
nd	1.18	ri	0.77	li	0.57	wh	0.48
or	1.13	ng	0.75	ra	0.57	ly	0.47

Trigrams Frequencies

- A powerful tool is to look at the **frequency of THREE-letter combinations**, known as **Trigrams.**
- The most **frequently occurring trigrams** ordered by decreasing frequency are:

    the, and, ent, ion, tio, for, nde *….. are some of the Trigram*

For Example

The Ciphertext to be solved is

**UZQSOVUOHXMOPVGPOZPEVSGZWSZOPFPESXUDBMETSXAIZVU
EPHZHMDZSHZOWSFPAPPDTSVPQUZWYMXUZUHSXEPYEPOPDZS
ZUFPOMBZWPFUPZHMDJUDTMOHMQ**

For Example

- The most common such **digram is th.** In our ciphertext, the most common **digram is ZW** which appears three times. So we make the correspondence of **Z with t and W with h.** we can equate **P with e.**

UZQSOVUOHXMOPVGPOZPEVSGZWSZOPFPESXUDBMETSXAIZVUEPHZHMDZSHZO WSFPAPPDTSVPQUZWYMXUZUHSXEPYEPOPDZSZUFPOMBZWPFUPZHMDJUDTMO HMQ

UZQSOVUOHXMOPVGPOZPEVSGthSZOPFPESXUDBETSXAIZVUEPHZ HMDZSHZOWSFPAPPDTSVPQUthYMXUZUHSXEPYEPOPDZSZUFPOM BthPFUPZHMDJUDTMOHMQ

For Example

- Now notice that the **sequence ZWP** appears in the **ciphertext**, and we can translate that **sequence as "the."**

UZQSOVUOHXMOPVGPOZPEVSGZWSZOPFPESXUDBMETSXAIZVUEPHZHMDZSHZ
OWSFPAPPDTSVPQUZWYMXUZUHSXEPYEPOPDZSZUFPOMBZWPFUPZHMDJUDTM
OHMQ

UZQSOVUOHXMOPVGPOZPEVSGthSZOPFPESXUDBETSXAIZVUEPHZ
HMDZSHZOWSFPAPPDTSVPQUthYMXUZUHSXEPYEPOPDZSZUFPOM
BtheFUPZHMDJUDTMOHMQ

For Example

- Next, notice the **sequence ZWSZ** in the first line. We do not know **that these four letters form a complete word**, but if they do, it is of the form **th_t**. If so, **S equates with ‘a’.**

UZQSOVUOHXMOPVGPOZPEVSGZWSZOPFPESXUDBMETSXAIZVUEPHZHMDZSHZ OWSFPAPPDTSVPQUZWYMXUZUHSXEPYEPOPDZSZUFPOMBZWPFUPZHMDJUDTM OHMQ

UZQSOVUOHXMOPVGPOZPEVSGthatOPFPESXUDBETSXAIZVUEPHZ HMDZSHZOWSFPAPPDTSVPQUthYMXUZUHSXEPYEPOPDZSZUFPOM BtheFUPZHMDJUDTMOHMQ

For Example

- we have

```
UZQSOVUOHXMOPVGPOZPEVSGZWSZOPFPESXUDBMETSXAIZ
 t a        e  e te  a that e e a       a
VUEPHZHMDZSHZOWSFPAPPDTSVPQUZWYMXUZUHSX
   e t   ta t ha e ee  a e  th    t  a
EPYEPOPDZSZUFPOMBZWPFUPZHMDJUDTMOHMQ
 e  e e tat  e   the   t
```

For Example

- Finally, The **complete plaintext, with spaces** added between words as follows:

```
it was disclosed yesterday that several informal but
direct contacts have been made with political
representatives of the viet cong in moscow
```

Limitations of Monoalphabetic Substitution ciphers

- **Monoalphabetic Substitution ciphers** are easy to **break** because they **reflect the frequency data** of the **original alphabet**

Outline

- ❖ Substitution Techniques
- ❖ Types of Substitution Cipher
- ❖ Mono-alphabetic Cipher
  - ✓ Caesar Cipher
  - ✓ Analysis of Caesar Cipher
  - ✓ Monoalphabetic Substitution Ciphers

Thank U