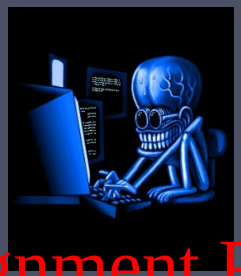
Lec – 3b
Substitution and Transposition, Steganography

Overview

- CIA-AA & Cryptography relationship
- Categories/Dimensions of cryptographic systems
- Substitution techniques:
  - Single-letter
    - Caesar
    - Rot13
  - Multi-letters substitution
    - Playfair
    - Hill Cipher
    - Vigenere
    - Vernam
    - One Time Pad
- Transposition techniques: rail fence and rectangular (route)
- Cryptanalysis: frequency analysis
- Steganography

Confidentiality & Cryptography



Hacker/Bad person

Unencrypted asset/data

Username	Password
Imran	-xcvzuy

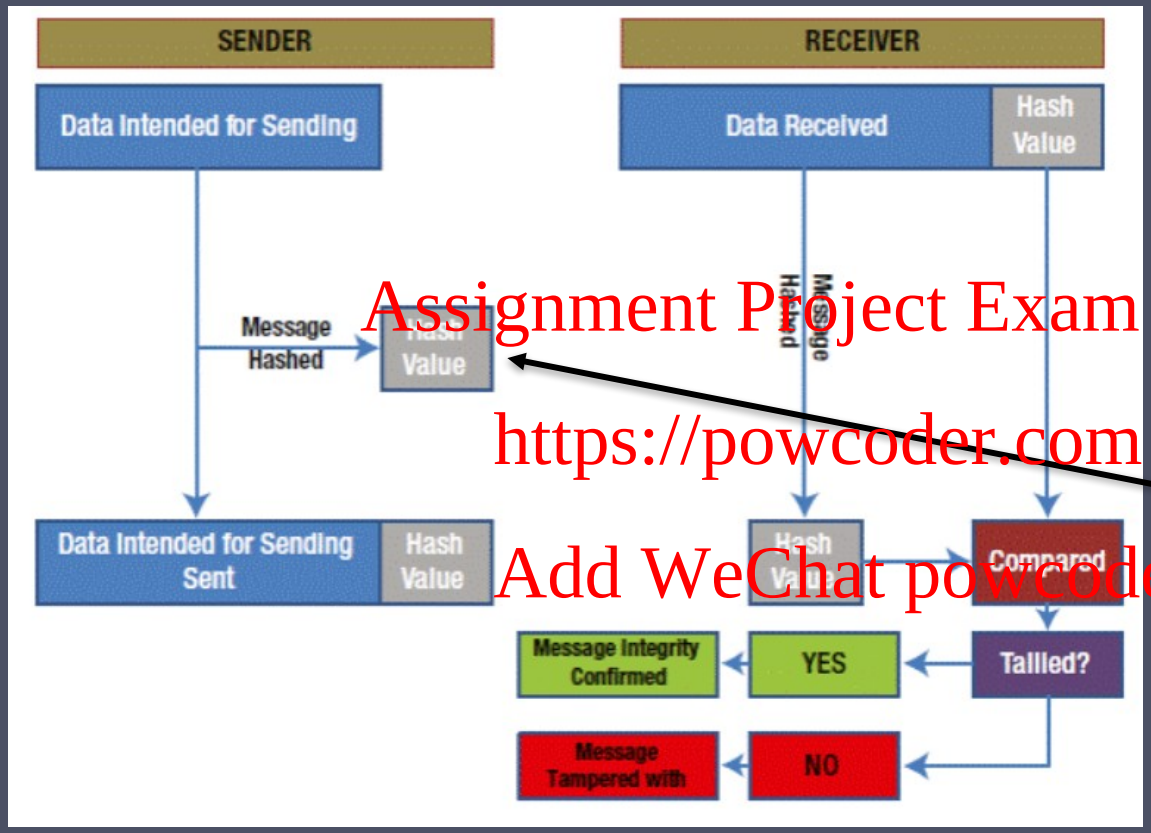
Username	Password
Imran	123456

Encryption
Symmetric
- DES/3DES
- Blowfish
- AES
- ---

Asymmetric
- RSA
- ElGamal
- ECC
- ---

System Developer

Integrity and Hash



SHA- family SHA1-SHA2, SHA3
RC family RC4, RC5..
MD family MD5

Authentication & Cryptography

Asymmetric cryptography is used to confirm authenticity of the sender to the receiver.

This will be discussed when digital certificate and public key cryptography is covered.

Overview of lab work for next two-three weeks

1) OpenSSL – A general purpose cryptographic library [Week 3]

2) Simple web application
- You can use any programming language
  - Java
  - Django

  Please note that I don't know much about those but you are free to use if you are confident
- I will be using Apache/PHP/MySQL, Lab sheet will cover it, so don't worry
  - XAMP local Apache/PHP environment will be set up [Week 3]
  - PHP to create User registration and login [Week 3]
  - MySQL to create DB [Week 3]
- Will use Ciphers to encrypt password [Week 4]
- Will use verities of ciphers using PHP syntax [Week 4]

3) Module is not about web programming
- But I will provide some notes about PHP which will be sufficient for this module

4) This will application that you will developed from next will also be used in exploiting web application vulnerabilities especially injection attacks.

Cryptographic systems

Classified along three independent dimensions:

The type of operations used for transforming plaintext to ciphertext

- Substitution – each element in the plaintext is mapped into another element
- Transposition – elements in plaintext are rearranged

The number of keys used

- Sender and receiver use same key – symmetric
- Sender and receiver each use a different key - asymmetric

The way in which the plaintext is processed

- Block cipher – processes input one block of elements at a time
- Stream cipher – processes the input elements continuously

Cryptographic system dimensions: substitution & transposition

- Recall: secret key ciphers use a secret key for encryption
- Almost all secret key ciphers (no matter how complicated) are essentially a combination of two simple techniques:
- Transposition: rearranging order of the plaintext characters
- Substitution: it is a method of encrypting by which units of plaintext are replaced with ciphertext, according to a fixed system; the "units" may be single letters, pairs of letters, triplets of letters, mixtures of the above, and so forth

Substitution

Substitution

- Monoalphabetic (simple substitution)
  - Simple substitution – each character is replaced
  - e.g. caesar cipher

- Polygraphic (larger groups of letters)

  Uniform **substitution** is performed on blocks of letters

  Plaintext: Welcome to Sussex and Welcome to Introduction to Computer Security

  for example

  if pair of letters moved e.g. we in welcome will be bigraphic

  if three letters moved e.g. wel in welcome will be trigraphic

- Polyalphabetic (a number of substitutions at different positions in the message)
  - For example Playfair cipher, Vigenere
  - In our plaintext listed above, suppose, when wel is substituted by 3 position, come by 4 and so on.

Monoalphabetic: Caesar cipher

- Julius Caesar a roman military general, 2000 years ago
- Earliest known and the simplest substitution cipher
- Involves replacing each letter of the alphabet with the letter standing three places further down the alphabet

  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
- Example

  Plain: meet me after the toga party

  Cipher: PHHW PH DIWHU WKH WRJD SDUWB

  What is the key used in this example?
- Task your turn – Plaintext is: Hello Ciphertext : ?

  Formula can be expressed as…

  C = E(3, p) = (p + 3) mod 26 where p is plaintext

  What is mod?

a	b	c	d	e	f	g	h	i	g	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

a	b	c	d	e	f	g	h	i	g	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

For letter b value is computed as following,
its index is 1
C= (1+3) mod 26 = 4 = e
Mod value can be checked on
https://www.miniwebtool.com/modulo-calculator/

- General Caesar algorithm takes on a value in the range 1 to 25.
  C = E(k, p) = (p + k) mod 26

- The decryption algorithm is simply p = D (k, C) = (C - k) mod 26

- For example for encrypted e p = (C - k) mod 26 = 4-3 mod 26 = 1 mod 26 = 1 = b

  C = E(k, p) = (p + k) mod 26 Where p is index of plaintext character
  P = D (k, c) = (c - k) mod 26 Where c is index of ciphertext character

- Bruteforce against Caesar cipher

  Three important characteristics enabled us to use a brute force:
  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.

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 jnls
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 epgvn
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

ciphertext: PHHW PH DIWHU WKH WRJD SDUWB
plaintext: meet me after the toga party

Large number of keys makes brute-force cryptanalysis impractical largely

The triple DES algorithm makes use of a 168-bit key, giving a key space of or greater than 3.7 * 1050 possible keys.

Problems with monoalphabetic

Key-space

- Total characters in English language are 26, key-space is 26!
- To increase key-space, use **permutation**

Why a need for multi-letter substitution?

- Monoalphabetic easy to break because they reflect the frequency data of the original alphabet.
- A counter measure is to provide multiple substitutes known as **homophones**. For example, the letter e could be assigned a number of different cipher symbols such as 3, 4 and 16.

  Each homophones assigns these symbols in rotation or randomly
- With homophones, each element of plaintext affects only one element of ciphertext, and **multiple letter patterns** will still survive in the ciphertext→ two solutions for this:

  1)Encrypt multiple letters of plaintext

  2)Use multiple cipher alphabets

Monoalphabetic ciphers

- **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 S:
  abc, acb, bac, bca, cab, cba

  Permutations calculator

$$P(n,r) = \frac{n!}{(n-r)!}$$

  Instead of simple cipher (26 keys) there are 26!

  What this symbol ! means?

  Provides bigger key space than DES.

Monoalphabetic ciphers

Key Size (bits)	Number of Alternative Keys	Time Required at 1 Decryption/μs	Time Required at 10^6 Decryptions/μs
32	$2^{32} = 4.3 \times 10^9$	$2^{31}\mu s$ = 35.8 minutes	2.15 milliseconds
56	$2^{56} = 7.2 \times 10^{16}$	$2^{55}\mu s$ = 1142 years	10.01 hours
128	$2^{128} = 3.4 \times 10^{38}$	$2^{127}\mu s = 5.4 \times 10^{24}$ years	5.4×10^{18} years
168	$2^{168} = 3.7 \times 10^{50}$	$2^{167}\mu s = 5.9 \times 10^{36}$ years	5.9×10^{30} years
26 characters (permutation)	$26! = 4 \times 10^{26}$	$2 \times 10^{26}\mu s = 6.4 \times 10^{12}$ years	6.4×10^6 years

William Stallings – Cryptography and Network security

But the problem is:

If language is known, it is still possible to find the plaintext

How? Frequency analysis

Polyalphabetic: playfair cipher

- Best known multiple-letter cipher
- Use a number of substitutions over the entire plaintext
- Treats digrams in the plaintext as single units and translates these units into ciphertext digrams

Use of 5 x 5 matrix

I/J takes on space

- What is a matrix? – If you don't know then research on it yourself.
- Netflix prize – 1 million dollar

Playfair - Example

key: MONARCHY plaintext: FRIEND

1st step: write the key in the matrix

M	O	N	A	R
C	H	Y		

2nd step: along with the key write remaining letters

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

3rd step: make pairs, not possible then write filler

FR	IE	ND
KO	KF	RY

Fill in the letters <u>minus duplicates</u> from left to right and from top to bottom
Fill in the <u>remainder of the matrix</u> with the remaining letters in alphabetic order
Letters <u>I/J</u> count as one letter
Plaintext is encrypted two letters at a time according to the following rules:

Playfair rules

Repeating plaintext letters that are in the same pair are separated with a filler letter, such as x, so that balloon would be treated as ba lx lo on.
Two plaintext letters that fall in the same row of the matrix are each replaced by the letter to the right, with the first element of the row circularly following the last. For example, ar is encrypted as RM.
Two plaintext letters that fall in the same column are each replaced by the letter beneath, with the top element of the column circularly following the last. For example, mu is encrypted as CM.
Otherwise, each plaintext letter in a pair is replaced by the letter that lies in its own row and the column occupied by the other plaintext letter. Thus, hs becomes BP and ea becomes IM (or JM, as the encipherer wishes).

Playfair

- Identification of digrams is more difficult
- Relative frequencies of individual letters exhibit a much greater range than that of digrams

- Used extensively in World war 1 and 2 especially first world war

- There are ways to break this cipher; one method is to find relative frequency.

Substitution multi letter: Hill Cipher

- Research yourself about this cipher
- Look at how encryption and decryption process works.

Polyalphabetic: vigenere cipher

- A set of monoalphabetic ciphers is used
- Vigenère cipher: generalize the rotate cipher by shifting different plaintext letters by different amounts
- Shift amount determined by the letter in the key (a: shift 0, b: shift 1, …, z: shift 25)

  Repeat the key if necessary

Example: [p:We are discovered save yourself ,k:deceptive]


```
Plaintext  wearediscoveredsaveyourself
Key        deceptivedeceptivedeceptive

Ciphertext zicvtwqngrzgvtwavzhcqyglmgj
```

Expressed numerically, we have the following result.

key	3	4	2	4	15	19	8	21	4	3	4	2	4	15
plaintext	22	4	0	17	4	3	8	18	2	14	21	4	17	4
ciphertext	25	8	2	21	19	22	16	13	6	17	25	6	21	19

key	19	8	21	4	3	4	2	4	15	19	8	21	4
plaintext	3	18	0	21	4	24	14	20	17	18	4	11	5
ciphertext	22	0	21	25	7	2	16	24	6	11	12	6	9

Polyalphabetic: Vigenere cipher

- Observation: Repeated plaintext patterns separated by integer multiples of key length results in repeated ciphertext patterns

1) Attacker looks for repeated ciphertext patterns and guess key length

Same example: can guess key length = 3 or 9

```
Plaintext  wearediscoveredsaveyourself
Key        deceptivedeceptivedeceptive
Ciphertext zicvtwqngrzgvtwavzhcqyglmgj
```

2) If (say) key length = 9, the 1st, 10th, 19th, … letter is encrypted using same key letter

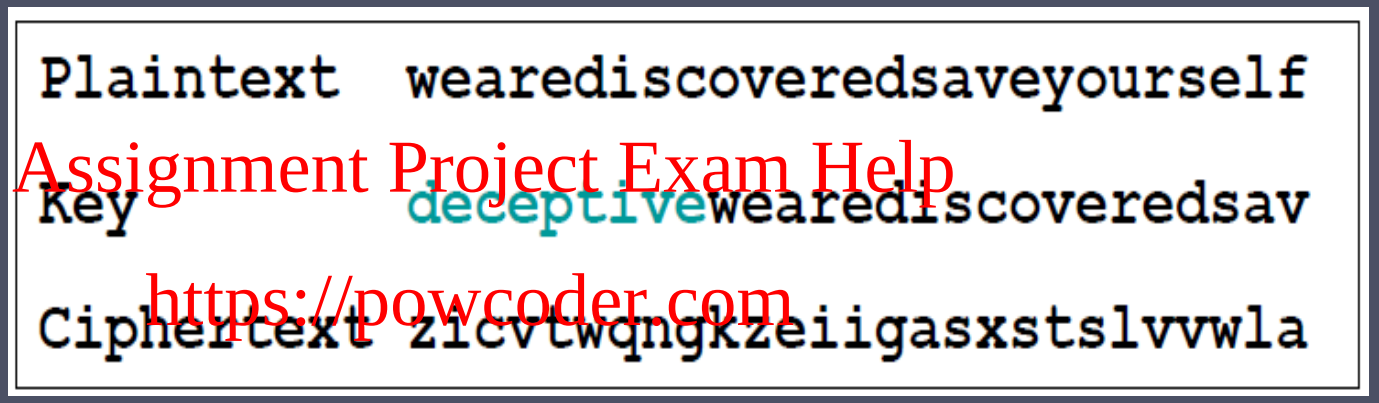
```
Plaintext  wearediscoveredsaveyourself
Key        deceptivedeceptivedeceptive
Ciphertext zicvtwqngrzgvtwavzhcqyglmgj
```

3) The problem reduces to solving a number of monoalphabetic ciphers. Break each such monoalphabetic cipher as before. Combine results

```
Ciphertext zrh...
Works out from frequencies that shift = 3
Plaintext  woe...
```

Vigenere cipher

- Idea: append plaintext as key
- Example:



- Problem with auto-key is that key and plaintext share same frequency

Vernam

- Aim is to choose a keyword that is as long as the plaintext and has not statistical relationship to the plaintext

- The system can be expressed

  Ci = Pi XOR Ki

  Pi = ith binary digit of plaintext

  Ki = ith binary digit of key

  Ci = ith binary digit of ciphertext

  XOR = Exclusive OR operation

- Focus in this technique is on the construction of the key
- Long key is a benefit, makes cryptanalysis difficult
- Can be broken with availability of sufficient cipher texts

For further details and example, look at
https://isaaccomputerscience.org/concepts/data_encrypt_vernam

One Time Pad

- Use a truly random key that is as long as the message
- An unconditionally secure type of cipher
- A brute force attack will fail: you will decrypt into many possible plaintexts
- There is no way to distinguish which is the correct plaintext

One Time Pad

- Why unconditionally secure?
  - There are no statistical relationship between plaintext and ciphertext
  - The ciphertext contains no information about the plaintext
  - In fact, there is always a key to decrypt into whatever plaintext you want

- However, *usually* not feasible in practice
  - Need huge amount of random numbers

How to securely distribute the secret key, which is as long as the message? (If there is such a way, why not use it for the message directly) ☺

Transposition techniques

Transposition techniques

Permutation is a mathematics concept

In permutation order does matter

$P(n,r)=n!/(n-r)!$

e.g. 4 digits code and total digits are 0-9

$P(n,r)=10!/(10-4)!$

$=10\cdot9\cdot8\cdot7\cdot6\cdot5\cdot4\cdot3\cdot2\cdot1/6\cdot5\cdot4\cdot3\cdot2\cdot1=5040$

Transposition ciphers mean that some sort of permutation has been performed on plain text

Transposition: Rail fence technique

Slight variations of it exists

Two steps

- Plaintext is written down as a sequence of diagonals
- Read off as a sequence of rows

Example

Plaintext = meet me after the toga party; rail fence of depth two, **depth is the key**

m	e	m	a	t	r	h	t	g	p	r	y
e	t	e	f	e	t	e	o	a	a	t	

m e m a t r h t g p r y

e t e f e t e o a a t

The encrypted message is : MEMATRHTGPRYETEFETEOAAT

Key/Depth is 3 (Note variation)

mmtetefeear

m				m				t		
	e		t		e		f		e	
		e				a				r

mtaeemfreet

m	t	a	e
e	m	f	r
e	e	t	

Route/Rectangular cipher

Research on this yourself. Find how encryption and decryption works in this particular cipher

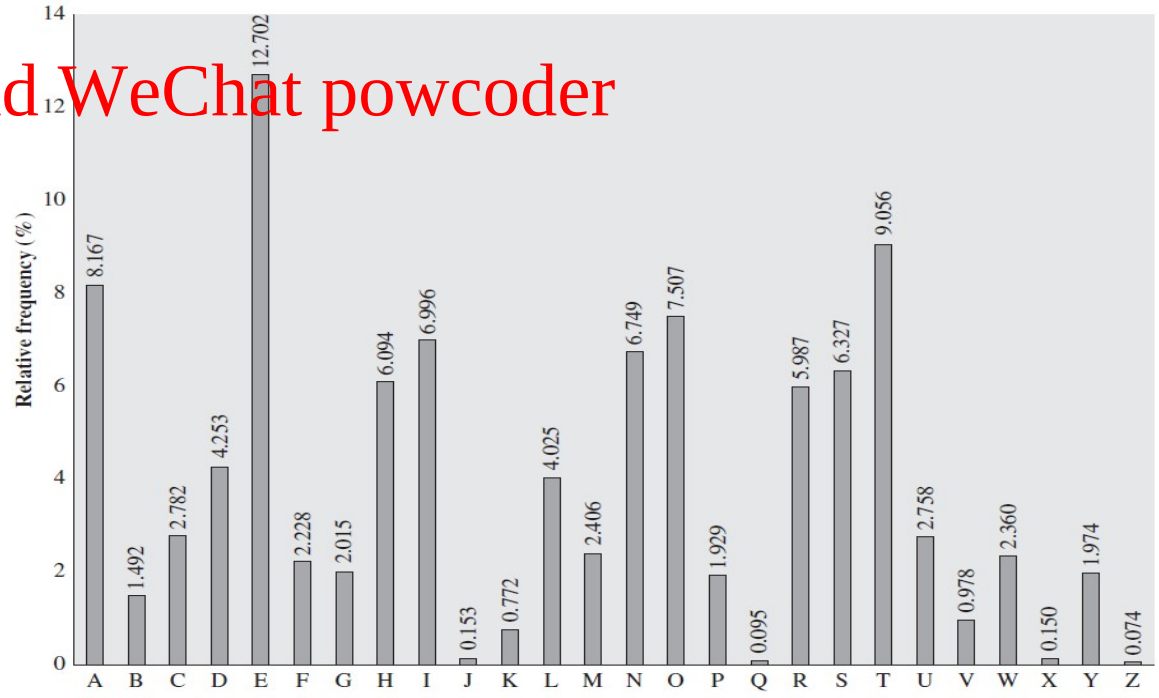
Cryptanalysis using frequency analysis

ciphertext:UZQSOVUOHXMOPVGPOZPEVSGZWSZOPFPESX
UDBMETSXAIZVUEPHZHMDZSHZOWSFPAPPDTSVPQUZWY
MXUZUHSXEPYEPOPDZSZUFPOMBZWPFUPZHMDJUDTMO
HMQ

1st step: Frequency of letters

P 13.33	H 5.83	F 3.33	B 1.67	C 0.00
Z 11.67	D 5.00	W 3.33	G 1.67	K 0.00
S 8.33	E 5.00	Q 2.50	Y 1.67	L 0.00
U 8.33	V 4.17	T 2.50	I 0.83	N 0.00
O 7.50	X 4.17	A 1.67	J 0.83	R 0.00
M 6.67				

2nd step: Compare it with Frequency of letters in English text



P and Z in ciphers are equal to e and t, not sure which is which; high frequency

S,U,O, M and H probably correspond to a,h,i,n,o,r,s ; relatively high frequency

A,B,G,Y,I,J likely included in b,j,k,q,v,x,z ; lower frequency

- 3rd step: Find frequency of two letter combinations and compare with English language two letter frequency. For details on English language, frequency look at this link. Two letters combinations is called **digrams or bigrams.**
  1) th in English language
  2) ZW in cipher text above
  3) Can be concluded that Z correspond with t and W with h
  4) This helps to conclude that P correspond with e
  5) Thus, ZWP appears to be the, most frequent trigram in English, indicate we are on the right track

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.

So far, then, 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
```

4th step: continued analysis of frequencies plus trial and error should easily yield a solution from this point.

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

Steganography

Conceal the existence of the message

Various techniques:

1. Character marking
2. Pin punctures
3. Typewriter correction ribbon
4. Through social media chat
5. Behind images

Does not use a lot as it is not very secure