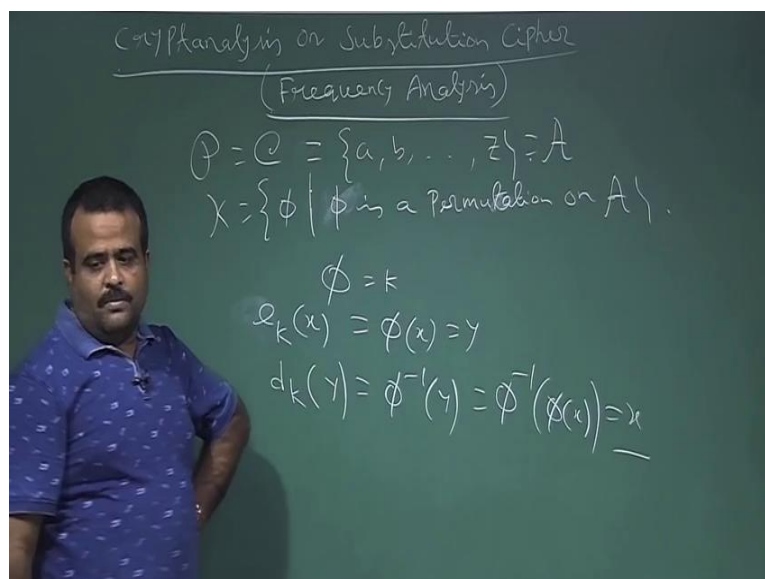


Internetwork Security
Prof. Sourav Mukhopadhyay
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture – 03
Cryptanalysis on Substitution Cipher (Frequency Analysis)

(Refer Slide Time: 00:28)

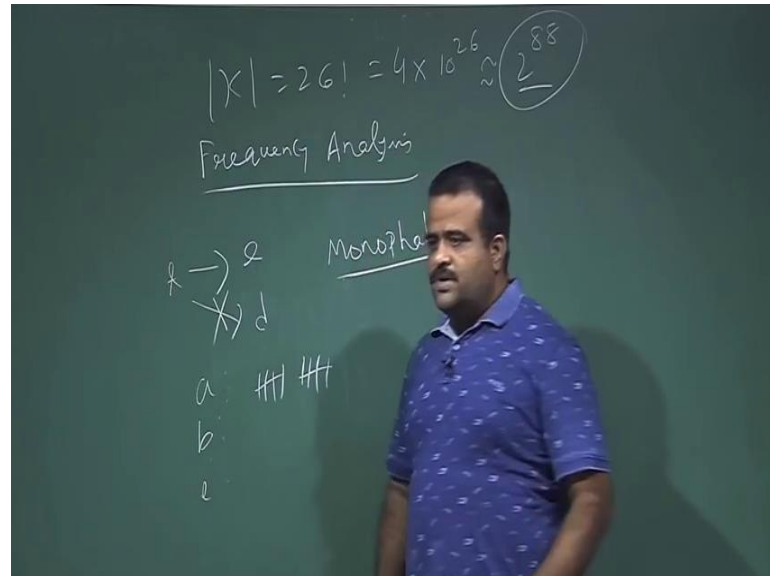


Hello. So, will talk about cryptanalysis on substitution cipher, we have seen the substitution cipher, it is basically the plaintext ciphertext or set of alphabets and key space is basically upon its set of all alphabets which denote this phi, A. So, this is basically set of all permutation power phi is a permutation on A. So, it is basically (Refer Time: 01:10) from A to A, then we have seen if we take a permutation and then the encryption is basically this is the key k. So, encryption e of k on x, x is an alphabet is basically phi of x and this is the encryption and decryption is d of, this is y, then y is basically phi inverse from y. So, which is basically phi inverse of phi of x which should give us x. So, this is the substitute cipher and you have seen some example, 1 example on substrate on earlier class.

Now, the question is how secure is this cipher is? Now this is the key space is the set of all permutation from set of all permutation on this set of alphabets, now the question is how secure this cipher is? So, let me to do the crypt analysis, especially we want to do the frequency analysis on this cipher. So, let us talk about whether we can have a boot

force attack or exhaustive search attack on these. So, let us talk about the key space. So, what is the key space of this cipher? Key is the set of all permutation.

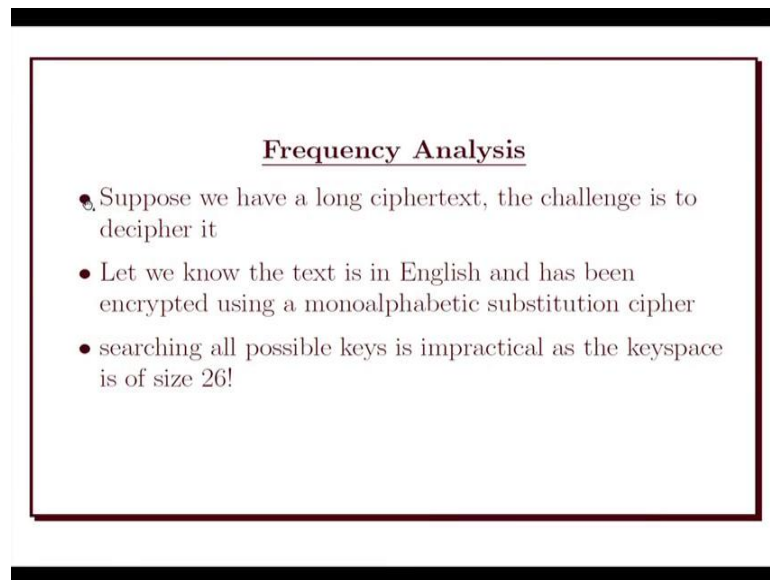
(Refer Slide Time: 02:47)



So, there are factorial 26 permutations. So, this number is basically 4 into 10 to the power 26, which is approximately 2 to the power 88, which is huge very large number. So, these many permutation are possible. So, if one has to go for the boot force method or searching in a key space in the Oscar, in Oscar has to search try for all possible permutation and there are this many permutation. So, Oscar will choose one permutation and check whether getting a meaningful text or not, then choose another formulation and check whether getting a meaningful text or not. So, these Oscar has to talk about try for all possible permutation and which is very expensive in the since that with the computationally it is very hard because 2 to the power 88 is a large number.

So but this way boot force method is not possible to attack this now, but there is attack model which is called frequency analysis.

(Refer Slide Time: 04:13)



The slide is titled "Frequency Analysis" in a bold, black, serif font. Below the title, there are three bullet points, each preceded by a small black circle. The text is in a black, serif font. The slide has a white background with a thin black border. The entire slide is set against a black background.

Frequency Analysis

- Suppose we have a long ciphertext, the challenge is to decipher it
- Let us know the text is in English and has been encrypted using a monoalphabetic substitution cipher
- searching all possible keys is impractical as the keyspace is of size $26!$

This can be a model for this attack, because our text is the English text, we know the plaintext and ciphertext are basically English. So, in the English text there is a pattern of the frequency of the alphabets. So, if we analyze the English big text then we can see this phenomenon is the e, e is the most frequent letter occurring in English text, and this way so if we know this pattern that e is the most frequent letter. So, this is called frequency analysis then we can have a crypt analysis on this substitution cipher.

So, this is suppose we have a long cipher text, which is a long, which is outcome of a long English text and the challenge is to decipher it, and it is encrypted using monoalphabetic substitution cipher; a monoalphabetic will come to that what is the monoalphabetic. In monoalphabetic, alphabet is mapping to a fixed alphabet. So, if t is mapping to e then t cannot be mapping to d this is not possible. So, this is called mono alphabetic, every alphabet is mapping to a fixed alphabet it is not mapping to a different alphabet.

(Refer Slide Time: 05:50)

- In English, *e* is the most common letter, followed by *t*, then *a*, and so on, as shown in the Figure 3

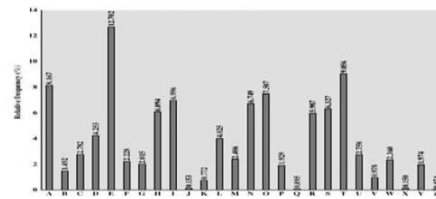


Figure 3: Relative Frequency of letters in the English Language.

So, will we have this phenomenon, this for English; text people have observed that *e* is the most common letter. So, if you take any English big text and if we just do the frequency count on this of the later then we can see this picture that we just do the frequency count and if the letter coming we just put like this, like this, this, this, this way to the frequency count from, how many times *a* is coming, how many times *b* is coming, how many times *e* is coming like this. So, it has observed if you take a big text and *e* is having high frequency than *t* then *t* then *a*. So, *e* is having high frequency than *t*, then *a* like this. So, this is observation on the English text. So, we want to apply this on our crypt analysis of substitution cipher. So, this how we can apply this? So, we have a long cipher text which is basically we know it is coming from English.

(Refer Slide Time: 06:57)

- examine the ciphertext in question, and work out the frequency of each letter
- if most common letter in the ciphertext is, for example, *J* then it would seem likely that this is a substitution for *e*
- if the second most common letter in the ciphertext is *P*, then this is probably a substitution for *t*, and so on
- however, regularities of the language may be exploited, e.g. relative frequency
- frequency analysis requires logical thinking, intuition, flexibility and guesswork

Now if we see now we count the frequency of that cipher text and we count the letter of in that cipher text a letter, if you observe J is coming more frequent, J is having more frequency in that cipher text, then we have to see we have to guess that J is basically e was encrypted to J; like this and then J is likely to the substitution for e and then P if we see P is the most frequent next most frequent word, then we can see the t is the next high frequency letter in the English text, then we can say that t is substitute to P. So, like this will do this frequency analysis and we will just there will be some gap and those gap we need to fill up by our intuition or by some eternal error method.

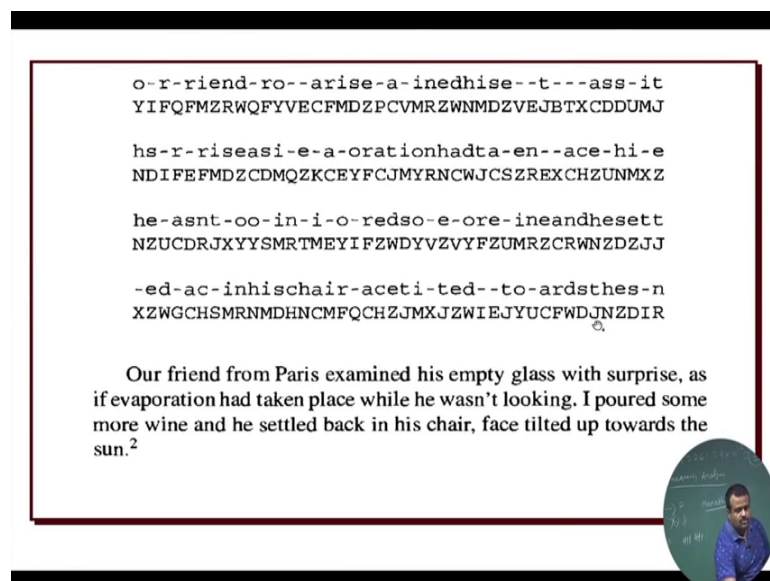
(Refer Slide Time: 08:10)

YIFQFMZRWQFYVECFMDZPCVMRZWNMZVEJBTXCDDUMJ
NDIFEFMZCDMQZKCEYFCJMYRNCWJCSZREXCHZUNMXZ
NZUCDRJXYYSMRTMEYIFZWDYVZVYFZUMRZCRWNZDZJJ
XZWGCHSMRMDHNCMFQCHZJMXJZWIEJYUCFWDJNZDIR

letter	frequency	letter	frequency
A	0	N	9
B	1	O	0
C	15	P	1
D	13	Q	4
E	7	R	10
F	11	S	3
G	1	T	2
H	4	U	5
I	5	V	5
J	11	W	8
K	1	X	6
L	0	Y	10
M	16	Z	20

So, we will take an English text. So, let us take the ciphertext this is a ciphertext, Alice is sending to Bob, now what will do? So, on the cipher text we will just do the frequency count of the letter. So, if you do the frequency count of the letter just A is coming how many times? No A is not coming, B is coming 1 time like this, C is coming this like this. So, we observe that Z is coming 20 times so; that means, wherever Z is, it is intuition that Z is basically E was substitute to Z; because E is the most frequent letter in the English alphabet. So, wherever Z is coming will replace by E. So, that was in the original plaintext because E. So, if you convert the corresponding plaintext if you see that is the English text, so in the English text E is the most frequent letter, now here Z is here encountering as most frequent alphabet so that means, it is quite clear that the E was map to Z and then which the most frequent then this is 16 then m so that means, T was mapped to N like this.

(Refer Slide Time: 09:39)



o-r-riend-ro--arise-a-inedhise--t---ass-it
YIFQFMZRWQFYVECFMDZPCVMRZWNMDZVEJBTXCDUMJ

hs-r-riseasi-e-a-orationhadta-en--ace-hi-e
NDIFEFMZCDMQZKCEYFCJMYRNCWJCSZREXCHZUNMXZ

he-asnt-oo-in-i-o-redso-e-ore-ineandhesett
NZUCDRJXYYSMRTMEYIFZWDYVZVYFZUMRZCRWNZDZJJ

-ed-ac-inhischair-aceti-ted--to-ardsthes-n
XZWGCHSMRNMHDHNCMFQCHZJMXJZWIEJYUCFWDJNZDIR

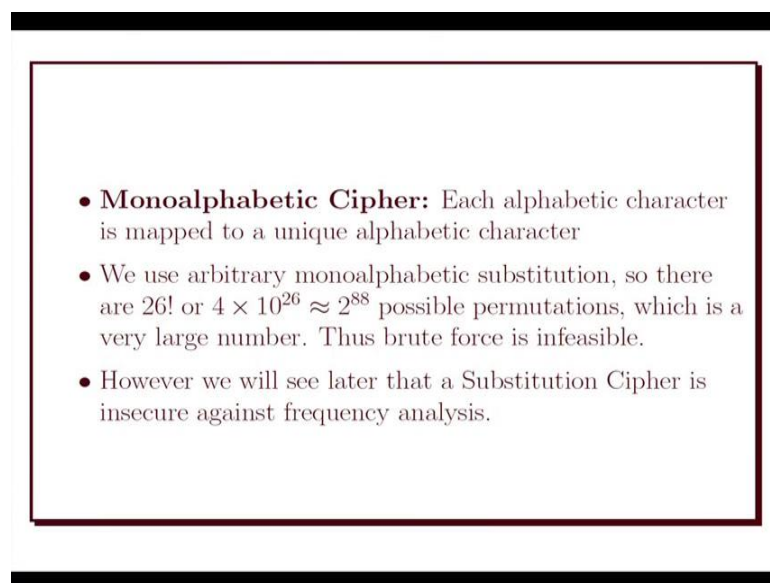
Our friend from Paris examined his empty glass with surprise, as if evaporation had taken place while he wasn't looking. I poured some more wine and he settled back in his chair, face tilted up towards the sun.²

So, if you do that, so this will be e. So, this will be the corresponding letter we place e, d, n like this. So, this is the first part of this first sentence of the ciphertext, this is the second, third, four. So, this we just got some of the letters of the plaintext and then we will try for together I mean by letters, if we try that also then we can fill up something more and then this way if you continue we can fill up and we will use our intuition to fill up the remaining thing. So, if we got some blank, now it is the questioner filling the blanks. We know this is the English text, so it should have a meaningful thing so; that means finally, once we got this from this frequency analysis then now we can guess

because this should be a meaningful English text, so this is this should be our, so our, this should be f is missing friend, this should be from pairs. So, this way one can get the plain text.

So, this is the frequency attack on frequency analysis on substitution cipher and this is possible because substitution cipher is a monoalphabetic cipher, monoalphabetic means each letter is mapping to a unique alphabet, each alphabet is mapping to a unique alphabet.

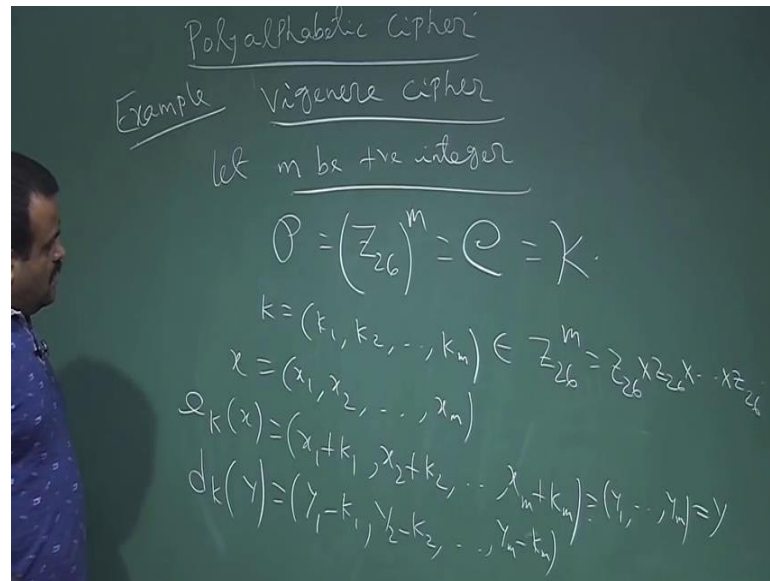
(Refer Slide Time: 11:14)



- **Monoalphabetic Cipher:** Each alphabetic character is mapped to a unique alphabetic character
- We use arbitrary monoalphabetic substitution, so there are $26!$ or $4 \times 10^{26} \approx 2^{88}$ possible permutations, which is a very large number. Thus brute force is infeasible.
- However we will see later that a Substitution Cipher is insecure against frequency analysis.

So, each alphabet character is mapped a unique alphabetic character. So, that is why it is possible to do the frequency analysis on this, on the other hand to prevent this, what we will do we will talk about what is called polyalphabetic cipher.

(Refer Slide Time: 11:32)



(Refer Slide Time: 11:45)

Vigenère Cipher

- **Polyalphabetic cipher:** use different monoalphabetic substitutions while moving through the plaintext.
- Let m be a positive integer
- $\mathcal{P} = \mathcal{C} = \mathcal{K} = (\mathbb{Z}_{26})^m$
- For $k = (k_1, k_2, \dots, k_m) \in \mathcal{K}$,

$$e_k(x_1, x_2, \dots, x_m) = (x_1 + k_1, x_2 + k_2, \dots, x_m + k_m)$$

$$d_k(y_1, y_2, \dots, y_m) = (y_1 - k_1, y_2 - k_2, \dots, y_m - k_m)$$
- All above operations are performed in \mathbb{Z}_{26}

So, in Polyalphabetic cipher it is basically we use the idea is to use the different monoalphabetic substitution while moving through the plaintext. So, we just use different monoalphabetic cipher while we move to the plaintext for the encryption purpose and then it will give us a polyalphabetic; in a sense that we will see in a plaintext. So, e is not mapping to a fixed alphabet, in the different position e is sometimes e is mapping to m , e is mapping to n something like that. So, we will come to an example.

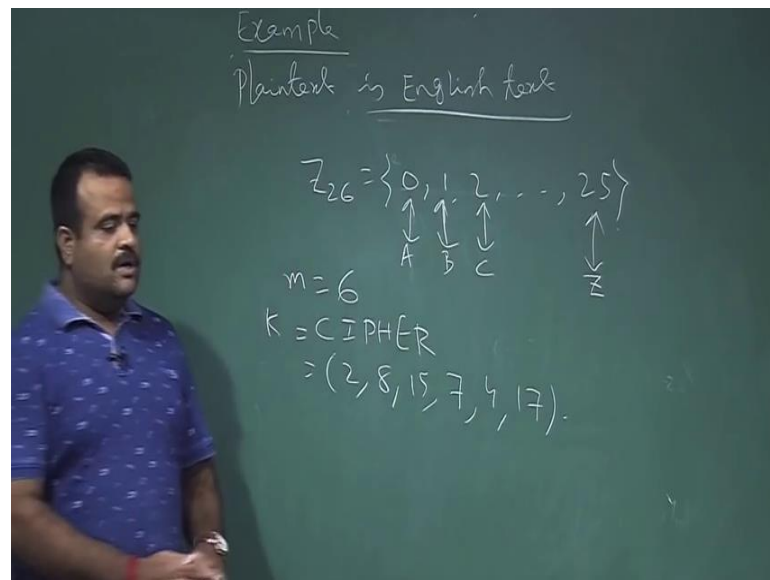
So, this is called visionary cipher first example of polyalphabetic cipher we will talk about, this is visionary cipher. So, this is the idea is to use the substitution cipher basically a sip cipher, while we are moving through the plaintext. So, here let m be this is the size of the window. So, we are blocking the play we are breaking the plaintext into blocks, each block size is just m bit, m be integer positive integer and then we block the plaintext into m bit say. So, our plaintext is basically \mathbb{Z}_{26} to the power m which is same as the ciphertext, space which is same as the key space. So, each key is a, this many bits. So, a key is basically k_1, k_2, k_n . So, this is basically coming from \mathbb{Z}_{26} to the power m ; to the power m means it is a Cartesian product \mathbb{Z}_{26}^m .

So, we will use the substitution cipher while moving through the plaintext m bit blocks wise. So, this is the key now we take a plaintext, plaintexts are also m bit. So, we take a plaintext a or x , x_1, x_2, x_n this is the plaintext, it is also m bit vector. Now the encryption used on this basically. So, it will also be a m bit because ciphertext space is also m bit. So, it is basically be twice; so, x_1 plus k_1 comma x_2 plus k_2 , comma x_n , x_n plus k_n and decryption is basically. So, this is a y , this is the $y_1, y_2 y_m$. So, this is Y vector.

Now, decryption on y vector is basically y_1 minus k_1 , comma y_2 minus y_2 comma dot dot dot y_m minus k_m . So, this is the decryption and each of this plus and this subtraction are under mod 26 because we want this guy should be \mathbb{Z}_{26} ; so each operation here whether addition or subtraction each operation is under mod 26.

So, this is the visionary cipher now we want to apply on some plain text. So, let us take an example of a plaintext.

(Refer Slide Time: 16:09)



So, suppose if plaintext is the English text. So, plain text is the English text and so we have the correspondence between English text and \mathbb{Z}_{26} . So, that is basically this is \mathbb{Z}_{26} , 0, 1, 2 up to 25 and English alphabet. So, this is basically A this is B. So, this is the correspondence we have used between \mathbb{Z}_{26} and the set of all alphabets. So, this is the 1 to 1 correspondence between English alphabet set with the set \mathbb{Z}_{26} .

Now we have to choose the m ; suppose m is 6 and suppose our key is say cipher; cipher means 4; 6 letter now we can replace this by this 1 to 1 correspondence. So, it will be 2, 8, 15, 7, 4, 17. So, this is our key $K = (2, 8, 15, 7, 4, 17)$; 17 is the basically this under this correspondence it is basically R.

(Refer Slide Time: 17:57)

- Plaintext : "thiscryptosystemisnotsecure".
- Encryption: add modulo 26

19	7	8	18	2	17	24	15	19	14	18	24	18	19	4	12
2	8	15	7	4	17	2	8	15	7	4	17	2	8	15	7
21	15	23	25	6	8	0	23	8	21	22	15	20	1	19	19

8	18	13	14	19	18	4	2	20	17	4
4	17	2	8	15	7	4	17	2	8	15
12	9	15	22	8	25	8	19	22	25	19

- Ciphertext:
"VPXZGIAXIVWPU~~BT~~TMJPWIZITWZT".

So, now suppose we want to encrypt a plaintext this; this is our plaintext a, this crypto system is not secure suppose this is the plaintext. So, what we do? We just convert this English alphabet to digit I mean the integer, we know the correspondence a is going to 0, b is 1, c is 2 like this we know this 1 to 1 correspondence here. So, this correspondence will apply and we will get the corresponding digit integer bits. So, these are the integer bits we are having.

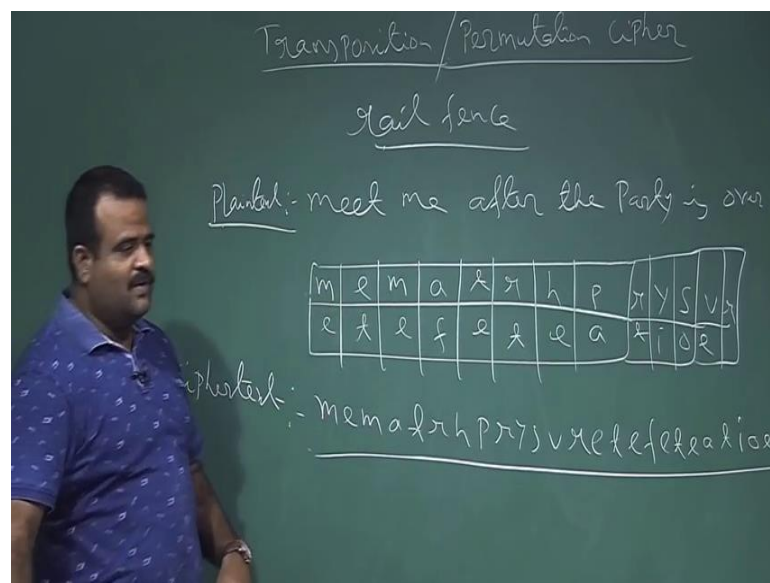
Now, these we have to break it into 6 bits block; bits means each bits is a integer, because our we are moving the through the plaintext and we are applying it monoalphabetic cipher that is the sip cipher we are using. So, this is the first block, this is the second block like this; last block we have only 3, 3 numbers, a 3 digit, 3 integers that is fine. So, now, we know the key is basically cipher. Cipher is basically our key and key is 2, 8, 15 this. So, these we are going to add with these blocks. So, if you do that, here 2, 8, 15, 7, 4, 17 2 8 this is the key. So, on each block we apply this and we get the corresponding ciphertext block. So, this is basically we are adding mod 26.

So, this is 17 plus 17 34; 34 mod 26 is basically 8. So, this is the cipher text. So, this is first block of the ciphertext, second block of the ciphertext. So, finally, we convert this integer again into the alphabet by that 1 to 1 correspondence and so 21 is basically V, 15 is P, 23 is X, 25 is Z like this. So, this is the ciphertext corresponding to this plaintext and why this is polyalphabetic? In the polyalphabetic we know that 1 alphabet can map

to the different different alphabet. So, if you come here if we observe this S; S is mapping to Z, this S is mapping to W, this S is mapping to U, this S is mapping to Z, this S is mapping to again Z. So, J and this is Z so that means, it is not a fixed number. So, S is not mapping to a fixed alphabet, S is mapping to different different alphabet over here. So, frequency analysis could be difficult in this case and this way it is called polyalphabetic cipher, because it is not giving us a unique correspondence. So, a letter is not going to a fixed letter, it may go into a different different alphabet.

So, now we will take another example of polyalphabetic cipher. So, polyalphabetic cipher is difficult to break by the frequency analysis because we know this e - who we know the most frequent letter, but if e is going to some different different alphabet. So, then in the cipher text frequency analysis will not give us the proper result like proper frequency that e is, this letter should come from e like this.

(Refer Slide Time: 21:50)



So next one is what is called transportation cipher, this is basically also called the permutation cipher also transportation cipher.

(Refer Slide Time: 22:09)

Transposition/Permutation Cipher

- Let m be a positive integer
- $\mathcal{P} = \mathcal{C} = (Z_{26})^m$
- \mathcal{K} = set of all possible permutations of $\{1, 2, \dots, m\}$
- For each permutation $\pi \in \mathcal{K}$,
$$e_{\pi}(x_1, x_2, \dots, x_m) = (x_{\pi(1)}, x_{\pi(2)}, \dots, x_{\pi(m)})$$
$$d_{\pi}(y_1, y_2, \dots, y_m) = (y_{\pi^{-1}(1)}, y_{\pi^{-1}(2)}, \dots, y_{\pi^{-1}(m)})$$
- π^{-1} being the inverse permutation of π

So, this is also called I think this is called permutation cipher also. So, transportation of permutation ciphers.

(Refer Slide Time: 22:12)

- **Transposition techniques:** So far all the ciphers we have looked at involved only substitution. A very different kind of mapping is achieved using transposition.
- In its simplest form, the **rail fence** technique involves writing down the plaintext as a sequence of columns and the ciphertext is read off as a sequence of rows. For example, if we use a rail fence of depth 2 with the plaintext *meet me after the party is over* we get:

m	e	m	a	t	r	h	p	r	y	s	v	r
e	t	e	f	e	t	e	a	t	i	o	e	
- Ciphertext is *mematrhpysvretetefeatioe* which is simply the first row concatenated with the second.

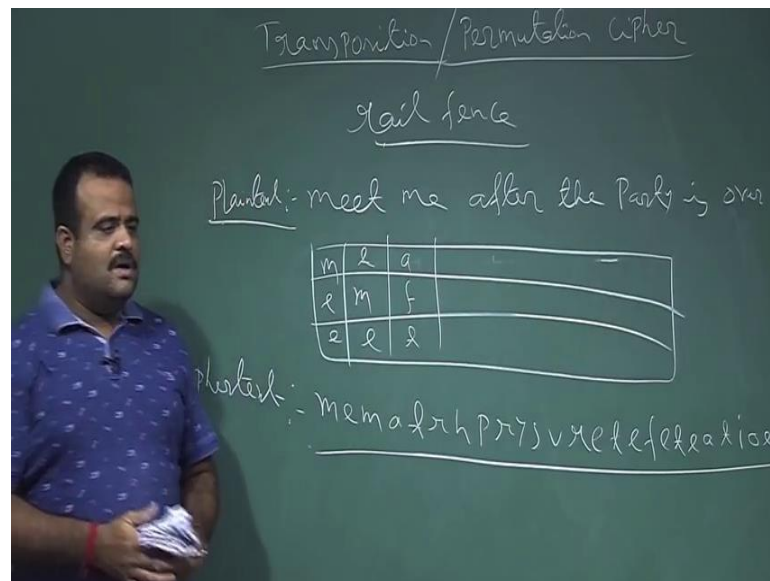
So far we have seen the substitution method, now this is coming from the transportation technique which is basically the rail fence technique. So, what is rail fence technique? Suppose we have a plaintext like this, meet me after the party is over. So, what we do? We suppose you have a 2 by 2 matrix this is rail fence this is called rail fence; this could be 3 by 3, 4 by 4 like this. So, we will put this in this way like this rail track. So, m, e, e,

t, m, e, a, f, t, e, r, t, h, e, e, a, p, a, r, t, y, i, s, o, v, e, r; so we have this plaintext this is the plaintext. So, this is rail fence technique basically.

So, what we are doing? We are just reading this in a column wise. So, this is the aim meet like this, this way we are reading and then these way we are placing in a matrix 2 by 2 matrix and then to get the cipher text, what we will do? Will read it by row wise; so the cipher text will be coming from read it by row wise, so if we read it row wise this is m, e, m, a, t, r, h, p, r, y, s, v, r, e, t, e, f, e, t, e, a, t, i, o, e. So, this is the cipher text and from here it is quite I mean it is not so obvious to guess what was the plaintext. So, this technique is called rail fence technique.

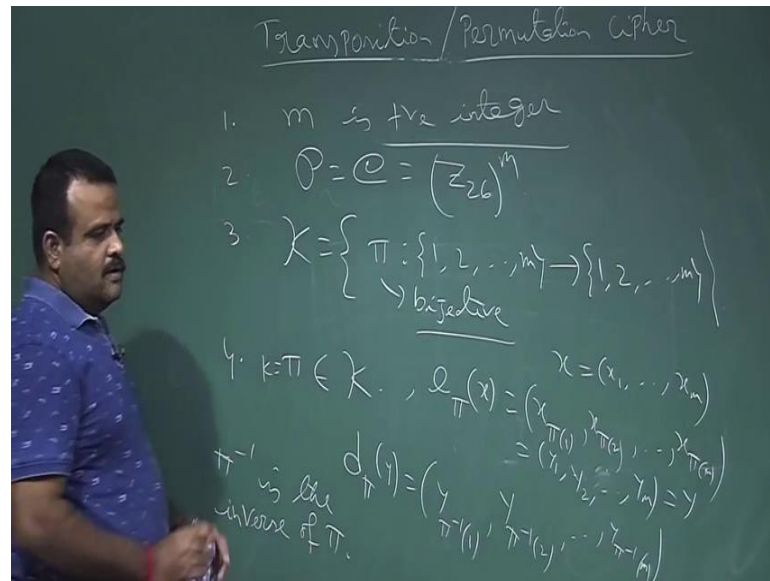
This is this transportation cipher is coming from these ideas. So, this could be instead of 2 by 2 order, we can have 3 by 3 also like this could be 3 by 3 or 4 by 4 like this.

(Refer Slide Time: 25:22)



So, if it is 3 by 3 then we have will put it like this m, e, e, e, m, e, a, f, t like this and we will read row wise. So, that technique is called transported rail fence technique and this technique idea will use for this what is called transportation cipher or the permutation cipher. So, let us formally define this cipher.

(Refer Slide Time: 26:08)



So, this is basically we also have m we choose. So, this is step 1; m is a positive number positive integer. So, our window size is m and here also similar to visionary cipher here also our P ciphertext place is basically Z_{26} to the power m and here our key space is basically set of all permutation from 1 to m to on this set. So, it is basically set of all bijective mapping form 1 to m to itself. So, this is the bijective mapping. So, this is a very set of all permutation. So, how many permutation are there? So, there are m numbers. So, there are basically factorial m permutations.

Now, what is now the encryption? For encryption I need to choose a key. So, we take a permutation which is a π from this key space, this is our K key, now we define this encryption by this e of. So, we need to take a x , x is a plaintext; x is coming from x_1, x_2, \dots, x_n . So, each plaintext is a m bit number. So, this is basically we are randomly we; not random word, I should not use it is basically a sampling of the positions. So, e of x is basically $x_{\pi(1)}, x_{\pi(2)}, \dots, x_{\pi(m)}$. So, basically we are this, this, this numbers we are just linearly permuting.

So, maybe π of 1 is 3, π of 2 maybe depending on the π permutation it could be that way. So, this is the encryption and the decryption is, so d of π y now say this is y ; so this is the y_1 this is a ciphertext; y_1, y_2, \dots, y_n and say this is the cipher text. So, the decryption will be on this is a Y vectra cipher text. So, d of y is basically $y_{\pi^{-1}(1)}, y_{\pi^{-1}(2)}, \dots, y_{\pi^{-1(m)}}$. So, basically we are

shuffling these numbers and then we are applying the we are reshuffling it to make it original one, I mean the original position we are making that, where π^{-1} is the inverse of this π , inverse formulation is the inverse formulation of π . So, this is the way. So, here we are just taking a n bit m blocks like this. So, this is our plaintext space and this is a set of keys set of all permutation basically. So, we choose a particular key we choose a key k , and this e key of this is basically the permutation this and decay of this is this π^{-1} is this.

(Refer Slide Time: 30:06)

Example

- $m = 6$.
- key is the following permutation π :

x	1	2	3	4	5	6
$\pi(x)$	3	5	1	6	4	2
- inverse permutation π^{-1} :

x	1	2	3	4	5	6
$\pi^{-1}(x)$	3	6	1	5	2	4
- Plaintext : "defendthehilltopatsunset"

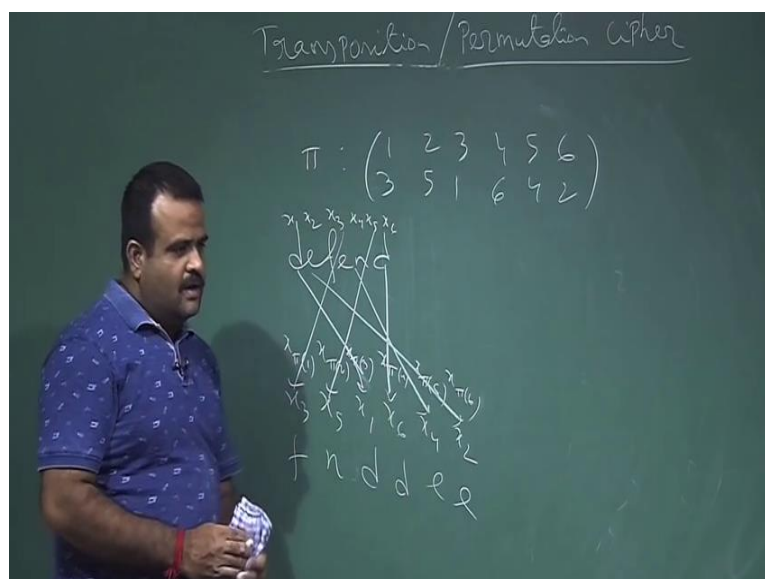
Now, let us take an example. So, suppose m is 6 and we choose a permutation which is basically 1 to m if this 1 is going to here, 3 is coming here, 5 is coming here. So, whatever x 5 it will come here, whatever is x 1 it will come here. So, bits basically it is a sapling I mean this linear permutation on the x 1, x 2, x n . So, this is one permutation we are choosing and then the corresponding inverse permutation is this. Now suppose this is our plaintext defend the hill top at sunset. So, this is our intents. So, if we have to apply this permutation cipher, we have to break it into 6 blocks, I mean the blocks of each 6 bit 6 digits, so d e f e n d. So, this way we just break it into 6 bits. So, last one if it is 6 it is ok, otherwise.

(Refer Slide Time: 31:08)

- partition the plaintext into group of six letters:
defend | thehil | ltopat | sunset
- rearrange according to π :
fnddee | eitlhh | oaltpt | nestsu
- Ciphertext: "FNDDEEEITLHHOALTPTNESTSU"
- Decryption can be done using π^{-1}

So, now, we have to apply this on this each of this block we have to apply this permutation. So, this permutation is basically. So, let us take this example.

(Refer Slide Time: 31:32)



So, if we just come back to our permutation; a permutation is. So, 1 2 3 4 5 6 and this is basically 3 5 1 6 4 2. So, now, this is our permutation.

Now if we take this defend - d e f e n d this is one block. So, now, this is $x_1, x_2, x_3, x_4, x_5, x_6$; now this will go to the this is a plaintext block first plaintext block. So, this will go to the $x_{\pi 1}, x_{\pi 2}, x_{\pi 3}, x_{\pi 4}, x_{\pi 5}, x_{\pi 6}$. Now $x_{\pi 1}$ is basically x_3 this is

$x_{\pi 2}$, $x_{\pi 2}$ is basically x_5 , and $x_{\pi 3}$ is basically x_1 , $x_{\pi 4}$ is basically x_6 , this is x_4 , this is x_2 so; that means, x_3 will come here, x_5 will come here, x_1 will go there, here and x_6 will come here, x_4 will go here and x_2 will go there.

So, we can just write. So, this is basically f , this is n , this is $f \circ n \circ d$, this is also d and this is e , e . So, this is just a linear shuffling on this digit. So, if we just apply that we are getting this the first block, similarly we apply π on this we get this, similarly this like this. So, finally, this is our ciphertext corresponding to that plaintext and for decryption you have to use the inverse permutation; so we take this block and we apply π inverse, so we will get depend similarly we have to apply from this both, we apply π inverse on each this we get this. So, this is called permutation cipher or transportation cipher.

Thank you.