INTRODUCTION

Syllabus: UNIT- I: Basic Principles

Security Goals, Cryptographic Attacks, Services and Mechanisms, Mathematics of Cryptography

Information Security plays a vital role in today's world of communication, military operations, and in financial organizations, .. etc. Information in the wrong hands can lead to a major loss of country, business or any organization. Before the wide spread use of electronic communication, the security of information is valuable to an organization and it is provided by means of physical and administrative. Physical security of information is by providing lockers to store the sensitive documents. Information can also be secured by screening the employee while hiring. A disgruntled employee can leak company's sensitive information to the outside world.

The requirements of information security in an organization have undergone two major changes in the last several decades.

1. The first change occurred with the introduction of computers. Due to this, there is a need to develop automated tools and s/w for protecting information stored in the computer. This is required more for shared systems, and for systems which can be accessed over a public telephone network. The collection of tools designed to protect data and to thwart (prevent) hackers is **computer security.**
2. The second change occurred with the introduction of distributed computers and the use of networks and communications facilities (routers, gateways) to transport data from one user to another. Therefore Network Security measures are needed to protect data during the transmission.

There are no clear boundaries between these two forms of security. The most popular attack on information systems is the computer virus. A virus may be introduced into a system physically using a diskette (any storage device), or it may also arrive over an internet. In either case, once the virus is identified, internal computer security tools are required to detect and recover from the virus.

Security Violations:

1. User A transmits a confidential file to user B. this file contains some sensitive information that are to be protected from disclosure. An unauthorized user C may capture a file during it's transmission.
2. A network manager, D, transmits a message to a computer, E who is working under him. The message instructs E to update the authorization file and their access privileges,
3. Instead of intercepting a message, a user F constructs his own message and transmits to E, as if the message come from the network manager D.
4. In case of employee firing without any warning or notice, the personal manager sends a message to a server system to invalidate the employee's account. After the invalidation, server posts a notice to the employee for confirmation of the action. Employee intentionally delays to give confirmation so that he can get final acces to the server to retrieve sensitive information.
5. A message is sent from a customer to a stock broker to sell all his shares. Consequently, suppose if shares price goes down, customer denies sending the message.

These are the possible types of security violations, these illustrates the range of concerns of network security. Internetwork security is both interesting and complex.

The Three Aspects of Security:

To assess the security needs of an organization effectively, and to evaluate and choose various security products and policies, the following three aspects of information security are considered

1. **Security Attack:** An action which compromises the security of information owned by an organization. This is an assault against a computer or network infrastructure
2. **Security Mechanism**: a mechanism that is designed to detect, prevent and recover from security attacks.
3. **Security service:** A service that enhances the security of information systems and data transmission of any organization

OSI Security Architecture: ITU-T (International Telecommunication Union Telecommunication Standardization Sec Recommendation X.800, Security Architecture for OSI defines systematic way to

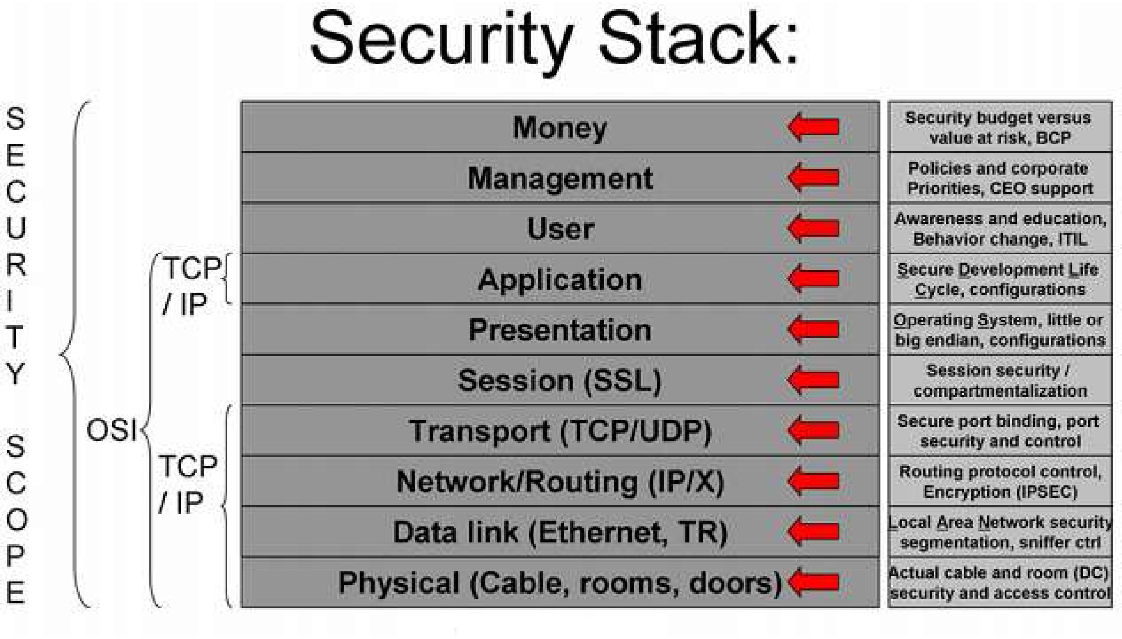
- Defining the requirements for security
- Characterizing the approaches to satisfying those requirements.

The OSI security architecture is useful to managers as a way of organizing the task of providing security. Furthermore, because this architecture was developed as an international standard, computer and communications vendors have developed security features for their products and services that relate to this structured definition of services and mechanisms. These can be defined briefly as follows:

Threats and Attacks (RFC 2828)

Threat : A potential for violation of security, which exists when there is a circumstance, capability, action, or event that could breach security and cause harm. That is, a threat is a possible danger that might exploit a vulnerability.

Attack: An assault on system security that derives from an intelligent threat; that is, an intelligent act that is a deliberate attempt (especially in the sense of a method or technique) to evade security services and violate the security policy of a system.



Security Stack:

Money	Security budget versus value at risk, BCP
Management	Policies and corporate Priorities, CEO support
User	Awareness and education, Behavior change, ITIL
Application	Secure Development Life Cycle, configurations
Presentation	Operating System, little or big endian, configurations
Session (SSL)	Session security / compartmentalization
Transport (TCP/UDP)	Secure port binding, port security and control
Network/Routing (IP/X)	Routing protocol control, Encryption (IPSEC)
Data link (Ethernet, TR)	Local Area Network security segmentation, sniffer ctrl
Physical (Cable, rooms, doors)	Actual cable and room (DC) security and access control

Security Services: Computer and network security research and development have focused on general security services that encompass the various functions required for information security. ITU-T has defined 6 types of security services

1. **Confidentiality:** The general meaning of confidentiality is the state of keeping or being kept secret or private. Ensures that the information in a computer system and transmitted information are accessible only for authorized users.

   Ex: Access types includes: read, print, display and other forms of **disclosure**.

   Disclosure: Revealing the contents of a message

2. **Authentication:** This is a process or action of verifying the identity of a user or person. Ensures that the origin (source) of a message or electronic document is correctly identified with an assurance that the identity is not false.

3. **Integrity:** The general meaning of it is the quality of being honest and having strong moral principles. Ensures that only authorized parties are able to modify computer system assets and transmitted data. And also ensure that the data has not been modified by anyone (third person, intruder) and anywhere else in the network.

   Ex: Modification includes: writing, inserting, changing, deleting, creating, appending and delaying or replaying of transmitted messages.

   The various types of Integrity is:

4. **Nonrepudiation:** It is the assurance that some one cannot deny something. It refers to the ability that a party to a contract or a communication cannot deny the authenticity of their signature on a document or the sending of a message that they originated. In other words, neither the sender nor the receiver of a transmitted message be able to deny the transmission or reception

   The various forms of Nonrepudiation are :

   Nonrepudiation, Origin: Proof that the message was sent by the specified party.

   Nonrepudiation, Destination: Proof that the message was received by intended receiver.

5. **Access Control**: It is the selective restriction of access to an information resource or information. Permission to access a resource is known as authorization. The prevention of unauthorized use of a resource (i.e., this service controls who can have access to a resource, under what conditions access can occur, and what those

accessing the resource are allowed to do). Access privileges to the information system are:

Ex: Write, Read, Execute, Save

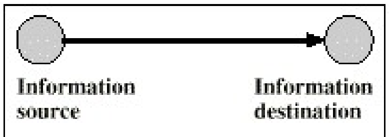
6. **Availability:** It requires that computer system assets be available to authorized users whenever needed.

Security Mechanisms: There is no single mechanism that will provide all the security services. One of the most specific security mechanisms in use is cryptographic techniques. Encryption or encryption-like transformations of information are the most common means of providing security. Some of the mechanisms are:

1. Enciphering and Deciphering
2. Digital Signature
3. Access Control

Security Attacks: According to G.J. Simmons information security deals with how to prevent cheating, or failing that, to detect cheating in information based systems where information has no meaningful physical existence. Attacks on the security of a computer system or network are best characterized by viewing the function of the computer system as providing information.

In general there is a flow of information from a source to destination. This normal flow is depicted in the following figure.

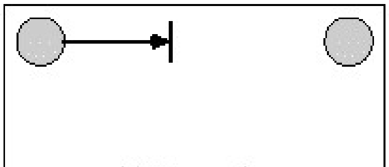


(a) Normal flow

Security attacks are categorized into four types:

1. **Interruption**: Information assets may be destroyed or becomes unavailable or unusable. This attack happens on the lack of **availability** service.

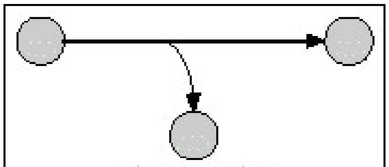
   Ex: destruction of a piece of hardware like hard disk, cutting the communication line, or disabling file management system.



(b) Interruption

2. **Interception:** An unauthorized party gains access to the resources. This is an attack on confidentiality. The unauthorized party may be a person, a program or a computer.

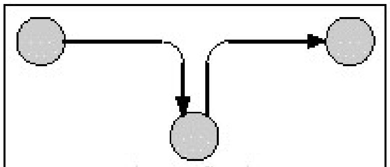
   Ex: Wiretapping, air tapping, illicit copying of files or programs.



(c) Interception

3. **Modification:** An unauthorized party gains access and tampers information assets. This attack is on **Integrity** service.

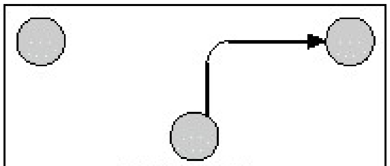
   Ex: changing values in a database, altering a program, modifying the contents of a message.



(d) Modification

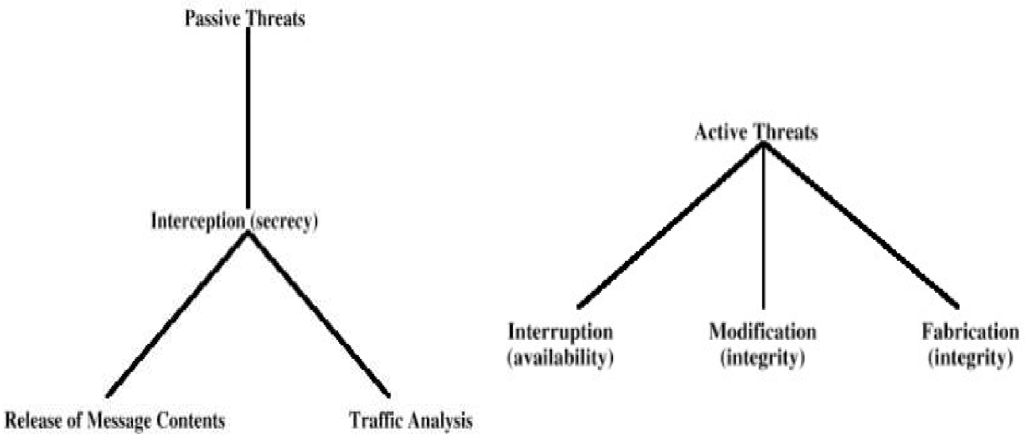
4. **Fabrication:** An unauthorized party inserts forged objects into the system. This attack is on authenticity.

   Ex: inserting fake messages in a network or addition of records in a file.



(e) Fabrication

These four types of attacks are further categorized into passive and active attacks.



Passive Attacks: These attacks comes under the category of eavesdropping and monitoring of transmissions. In this case the goal of opponent is to obtain transmitted information. There are two types of passive attacks:

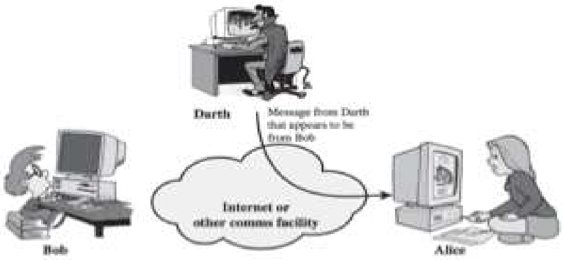
i) **Release of Message Contents:** Any telephone conversation, a transmitted file, an e-mail message may contain confidential information. The opponent will try to capture the message and reveal the contents of transmitted message

ii) **Traffic Analysis:** In case of a protected message, opponents, even if they capture, they could not extract the contents of a message. But still the opponent can observe the traffic these messages. Opponent can also determine the location and identity of communicating hosts (means source and destination) and could observe the frequency and length of messages being exchanged.

These passive attacks are very difficult to detect because they do not alter the contents of message. But measures are available to prevent these attacks.

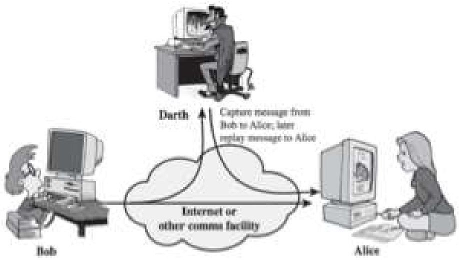
Active attacks: These attacks involve modification data stream or the creation of a fraudulent data. These are further classified into masquerade, replaying, modification of messages, denial of service.

Masquerade: This attack takes place when one user pretends like another user. In terms of communications security issues, a masquerade is a type of attack where the attacker pretends to be an authorized user of a system in order to gain access to it or to gain greater privileges than they are authorized for.


Darth
Message from Darth that appears to be from Bob
Internet or other comms facility
Bob
Alice

Replay: A replay attack (also known as playback attack) is a form of network attack in which a valid data transmission is maliciously or fraudulently repeated or delayed. Replay attacks are the network attacks in which an attacker spies the conversation between the sender and receiver and takes the authenticated information e.g. sharing key and then contact to the receiver with that key. In Replay attack the attacker gives the proof of his identity and authenticity.

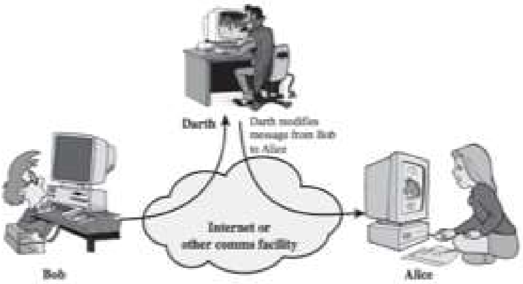
Ex: Suppose in the communication of two parties Bob and Alice; Bob is sharing his secret key to Alice to prove his identity but in the meanwhile Attacker Darth eavesdrop the conversation between them and keeps the information which are needed to prove his identity to Alice. Later Darth contacts to Alice and prove its authenticity.


Darth
Capture message from Bob to Alice; later replay message to Alice
Internet or other comms facility
Bob
Alice

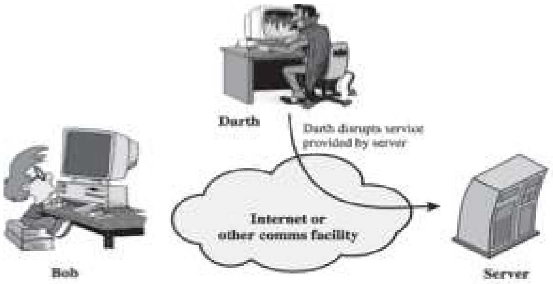
Modification of messages: In a message modification attack, an intruder alters packet header addresses to direct a message to a different destination or modify the data on a target machine.

Ex: Allow Ram to read confidential file is modified to

Allow Shyam to read confidential file.

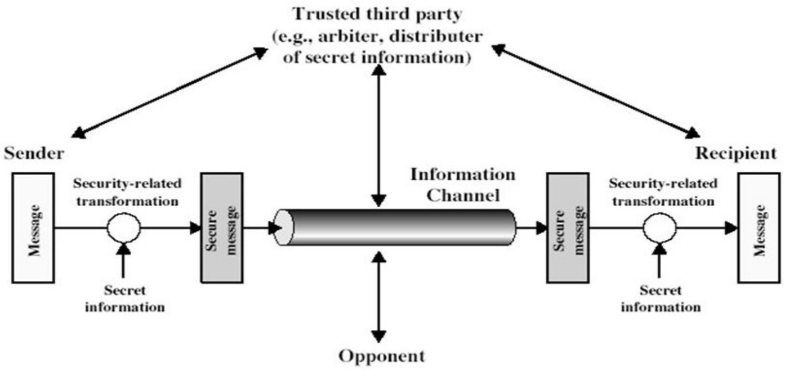

Darth
Darth modifies message from Bob to Alice
Internet or other comms facility
Bob
Alice

Denial of Service: In a denial of service (DoS) attack, users are underprivileged of access to a network or web resource. It prevents the normal use or management of communications facilities. Another form of denial of service is disruption of an entire network, either by disabling the network or be overloading it with huge number of messages so as to degrade performance.


Darth
Darth disrupts service provided by server
Internet or other comms facility
Bob
Server

Network Security Model: A message is transmitted from one party to another across the Internet. The two parties, who are principal communication parties in this scenarios, must cooperate each other to transmit data. A logical channel is established between the two parties using TCP/IP.

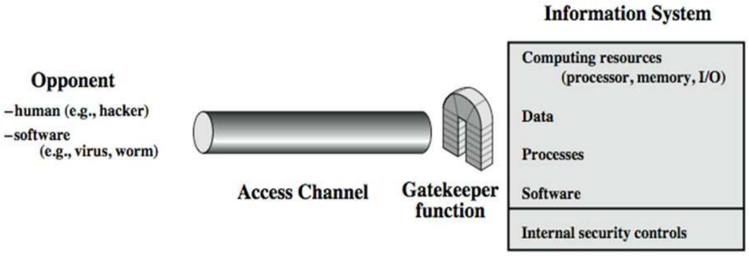
Security aspects are applied when it is necessary to protect information transmission from an opponent (third party) who may attack on confidentiality, authenticity, and any service.



A trusted third party is required to achieve secure transmission. A third party is responsible for distributing the secret information to the two principals. This general model performs four basic tasks:

1. Design an algorithm for performing secure transmission
2. Generate the secret information to be used with the designed algorithm
3. Develop methods for distribution and sharing of secret information
4. Specify protocols which are used by two principals for secure transmission

To protect information systems from unwanted access, the following network access security model is used:



An opponent may be a human or software. Hackers attempt to penetrate a systems that can be accessed over a network. Hacker is a person who uses computers to gain unauthorized access. In computing, a hacker is any skilled computer expert (programmer) that uses their technical knowledge, uses bugs or exploits to break computer systems.

A software program, can present two types of threats:

1. Information access threats: Interception and modification of messages on behalf of unauthorized users who do not have access to that data.

2. Service threats: These threats exploit service fa in computer to prevent use by legitimate user

Virus and worms are two examples of service threats.

Virus attacks can be introduced into a system by copying or downloading a file from a diskette or from any web server. Whereas worms can be inserted into a computer across a network. Computer worms are similar to viruses in that they replicate functional copies of themselves and can cause the same type of damage. In contrast to viruses, which require the spreading of an infected host file, worms are standalone software and do not require a host program or human help to propagate. A worm is a special kind of computer virus that propagates by self-replication over a computer network. This propagation can be either via e-mail or other means such as files being copied over a network.

	Computer Virus	**Computer Worm**
How does it infect a computer system?	It inserts itself into a file or executable program.	It exploits a weakness in an application or operating system by replicating itself.
How can it spread?	It has to rely on users transferring infected files/programs to other computer systems.	It can use a network to replicate itself to other computer systems without user intervention.
Does it infect files?	Yes, it deletes or modifies files. Sometimes a virus also changes the location of files.	Usually not. Worms usually only monopolize the CPU and memory.
whose speed is more?	virus is slower than worm.	worm is faster than virus. E.g. The code red worm affected 3 lack PCs in just 14 Hrs.

	Computer Virus	**Computer Worm**
Definition	The virus is the program code that attaches itself to application program and when application program run it runs along with it.	The worm is code that replicate itself in order to consume resources to bring it down.

The security mechanism required to prevent above two threats is to use gatekeeper function and internal security controls. The first, gatekeeper function use password based login procedures to allow only authorized users. The second one internal security controls monitors the activity and analyse the stored information to detect the presence of unwanted intruders.

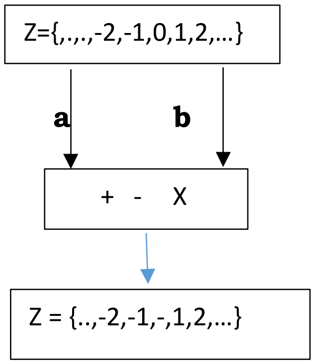
MATHEMATICS OF CRYPTOGRAPHY

Integer Arithmetic :

Set of Integers: The set of integers, denoted by **Z,** contains all integral numbers (with no fractions) from negative infinity to positive infinity.

Z={, , , -2,-1,0,1,2,.,.,.,}

Binary Operations: In Cryptography three binary operations are applied on set of integers. A binary operation, takes two inputs and produces one output. Three common binary operations are : addition, subtraction and multiplication. The two inputs come from set of integers; the output goes into the set of integers.



Ex:

Add : 5+9=14 (-5)+9=4 5+(-9)=-4 (-5)+(-9)=-14

Sub : 5-9=-4 (-5)-9=-14 5-(-9)=14 (-5)-(-9)=4

Mul : 5x9=45 (-5)x9=-45 5x(-9)=-45 (-5)x(-9)=45

Integer Division: IF we divide a by n we get q and r. The relationship between them is:

A=qxn+r

Ex: a=255 n=11, q is 23 and r=2

Two restrictions: when we use division in cryptography, we impose two restrictions:

1. Divisor must be a positive integer (n>0)
2. Remainder must be a nonnegative integer(r>=0)

Ex: a=255 n=11

To apply the restriction that r needs to be positive, we decrement the value of q by 1 and we add the value of n to r to make it positive

Therefore, -255=(-24x11)+(-2+9)=(-24x11)+9

Divisibility : If a is not zero and we let r=0 in the division relation , we get a=qxn. This is known as n divides a and can also be treated as a is divisible by n. This can be shown as a|n. If the remainder is not zero, then n does not provide a and this is represented as a ∤n.

Properties of Divisibility:
1.If a|1 , then a=+_1
2. If a|b and b|a, then a=+_b
3. if a|b and b|c, then a|c
4. If a|b and a|c, then a|(mxb+nxc), where m and n are arbitrary integers
Ex: 3|15 and 15|45 then according to 3rd property 3|45
3|15 and 3|9, according to 4th property,3|(15x2+9x4)=3|66
All Divisors: A positive integer can have more than one divisor. There are two facts about divisors of positive integers.

1. The integer 1 has exactly one divisor itself.
2. Any positive integer has at least two divisors, 1 and itself(but it can have more)

Greatest Common Divisor(GCD): GCD is very much useful in cryptography. Two positive integers may have many common divisors, but there is only one greatest common divisor.
Ex: The common divisors of 12 and 40 are 1,2 and 4. GCD is 4.

Euclidean Algorithm : Finding the GCD of two positive integers by listing all common divisors is not practical when the two integers are large. A famous mathematician Eucild developed an algorithm 2000 years before itself. This algorithm is based on the following two facts:

1. Gcd(a,0)=a
2. 2. Gcd(a,b)=gcd(b,r), where r is the remainder of dividing a by b

```
r1=a;
r2=b;
while(r2>0)
{
q=r1/r2;
r=r1 - qxr2;
r1=r2;
r2=r;
}
gcd(a,b)=r1
```

Ex: find the gcd of 2740 and 1760

q	r1	r2	r
1	2740	1760	980
1	1760	980	780
1	980	780	200
3	780	200	180
1	200	180	20
9	180	20	0
	20	0	

Extended Euclidean Algorithm: Given two integers and b, we often need to find other two integers, s and t, such that sxa + txb = gcd(a,b) . The extended Euclidean algorithm can calculate gcd(a,b) and at the same time calculate the value of s and t.

```
r1=a;
r2=b;
s1=1;
s2=0;
t1=0;
t2=1;
while(r2>0)
{
q=r1/r2;
r=r1 - qxr2;
r1=r2;
r2=r;

s=s1 - qxs2;
s1=s2;
s2=s;

t=t1 – qxt2;
t1=t2;
t2=t;

}
gcd(a,b)=r1;
s=s1;
t=t1;
```

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

q	r1	r2	r	s1	s2	s	t1	t2	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	

Linear Diophantine Equations: This is the immediate application to find the solution to the linear Diophantine equations of two variables is extended Euclidean algorithm.

Ex:ax+by=c we need to find integer values x and y that satisfy the equation. This type of equation has either no solution or an infinite number of solutions.

Let d=gcd(a,b) if d does not divide c then the equation has no solution. If d divides c then we have infinite number of solutions. One of them is called the particular, the rest are general solutions.

Finding a particular solution:
If d divides c

1. Reduce the equation to a1x+b1y=c1 by dividing both sides of the equation by d.
2. Solve for s and t in the relation a1s+b1t=1 using the extended Euclidean algorithm.
3. The particular solution is x0=(c/d)s and y0=(c/d)t

General Solutions:
x=x0+k(b/d) and y=y0-k(a/d)

Ex: Find the particular and general solutions of 21x+14y=5
Sol: d=gcd(a,b)=gcd(21,14)=7
Since 7 divides 35
Particular solution

1. 3x+2y=5
2. 3s+2t=1 using extended Euclidean, s=1 and t=-1
3. X0=5x1=5 y0=5x(-1)=-5

General solutions:
X=5+kx2 and y=-5 –kx3 where k=1,2,3,….
(5,-5),(7,-8),(9,-11)

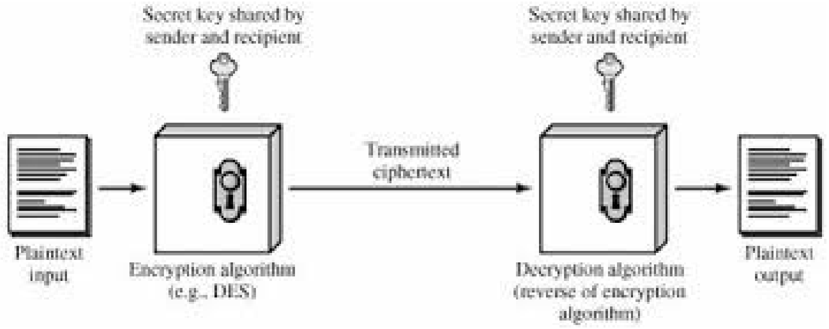
Modular Arithmetic:

CONVENTIONAL ENCRYPTION: CLASSICAL TECHNIQUES

Conventional encryption is also known as symmetric encryption, single key encryption, same key encryption and secret key encryption.

Simplified model of Symmetric Encryption: In conventional encryption,

- The original intelligible message, is considered as plain text.
- This plain text is converted into random nonsense (unintelligible), and is considered as cipher text.
- The ingredients of any encryption algorithm consists of plain text and key.
- The key value is an independent of the plain text.
- Once the cipher text is generated, it is transmitted in the network.
- Upon reception, the receiver transforms the cipher text back to the original plain text by using a decryption algorithm and the same key that was used in encryption.


Secret key shared by sender and recipient
Secret key shared by sender and recipient
Transmitted ciphertext
Plaintext input
Encryption algorithm (e.g., DES)
Decryption algorithm (reverse of encryption algorithm)
Plaintext output

The security of conventional encryption depends on the following factors:

- Encryption algorithm must be powerful and it must be impractical to decrypt a message based on the cipher text
- Secrecy of the key

Means, it is impossible to decrypt a message on the basis of cipher text plus the knowledge of encryption/decryption algorithm. The essential elements of a conventional encryption scheme are :

- A source inputs a plain text message $X=[X_1,X_2,X_3,..,X_M]$.
- The M elements of X are letters in some finite alphabet set
- For encryption, a key scheme $K=[K_1,K_2,K_3,..,K_M]$ is generated
- If the key is generated by a source then it must be shared or distributed to the destination by using secure channel, (or)
- A third party generates a key and deliver it securely to both source and destination.
- With the message X, and encryption key K, the encryption algorithm produces a cipher text Y where $Y= [Y_1,Y_2,Y_3,..,Y_M]$.

$$Y=E_K(X)$$

- The intended receiver, by using the same key decrypts the cipher text using

$$X=D_K(Y)$$

The three independent dimensions of Cryptography (or) Classifications of Cryptography:

1. The type of operations to be performed for transforming plain text to cipher text

a) Substitution: Each element of the plain text is substituted or replaced with another elements.

Ex: Caesar Cipher, Mono alphabetic cipher, Playfair cipher, Hill Cipher, Vigenre, Vernam and One-time pad cipher

b) Transposition: The plain text letters are rearranged to form a cipher text.

Ex: Railfence, Columnar Transposition, Row Transposition Rotor Machines

2. The number of keys used:

a) Single Key : If both the sender and receiver uses the same key then it is known as Single-key, same-key, symmetric key, conventional key or secret key cryptography

Ex: DES, IDEA, CAST128, AES, BlowFish,

b) Multiple Key: If both the sender and receiver uses different keys for encryption and decryption then it is known as two-key, multiple-key, public-key, private-key, asymmetric key cryptography.

Ex: RSA, Diffie-Hellman, ECC

3. The way in which the plain text is processed :

a) Stream Cipher: stream ciphers processes input elements continuously and outputs one element at a time.

Ex: Caesar cipher, mono alphabetic cipher, vigenere, vernam, one-time pad

b) Block Cipher: Block ciphers processes the input one block of elements at a time and produces on output block for each input block

Ex: play fair cipher, hill cipher, DES,CAST 128, IDEA,Blowfish, AES

Classical Encryption Techniques:

Substitution Techniques: The letters of the plain text are replaced by other letters, numbers, or symbols.

1. Caesar Cipher: The earliest and simplest by Julies Caesar. Caesar cipher replaces each letter with the other alphabet. This cipher allocates a numerical code to each alphabet. All cipher algorithms use reversible functions as cryptographic operations.

i.e,

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	26
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

Ex: Plain Text: meet me after the toga party

Key: a numeric value

The formula for Encryption is:

$C=E_K(P)$

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

The formula for Decryption is:

$P= D_K(P)$

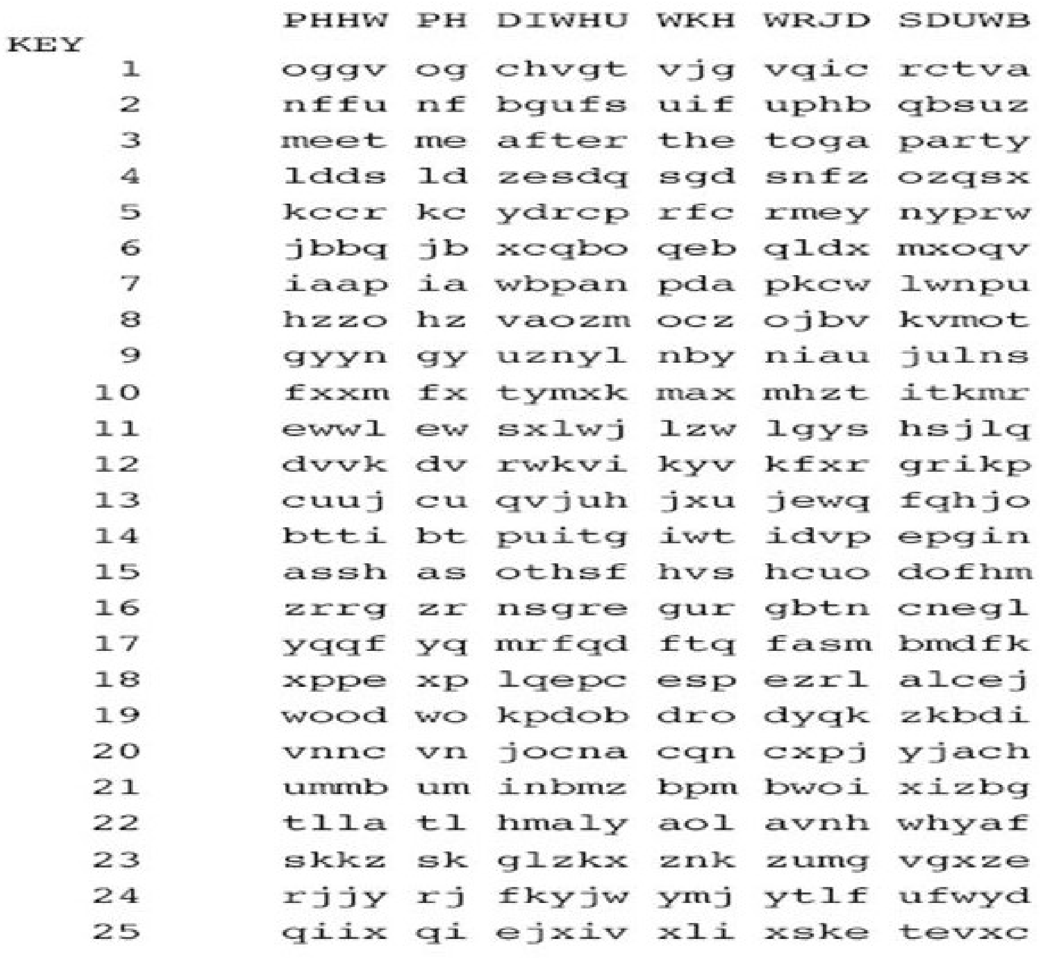
$P=(C-K) \bmod 26$

If K=3, the above plain text can be transformed as

Plain : Meet me after the toga party

Cipher: PHHW PH DIWHU WKH WRJD SDUWB

Brute Force Crypt Analysis: If the crypt analyst knows that this cipher text uses Caesar cipher, then he can apply a brute force attack (Trying with all possible keys) on the cipher text. There are 25 possible keys. Crypt analyst continue this operation till he gets a meaningful content. With only 25 possible keys Caesar cipher is not secure

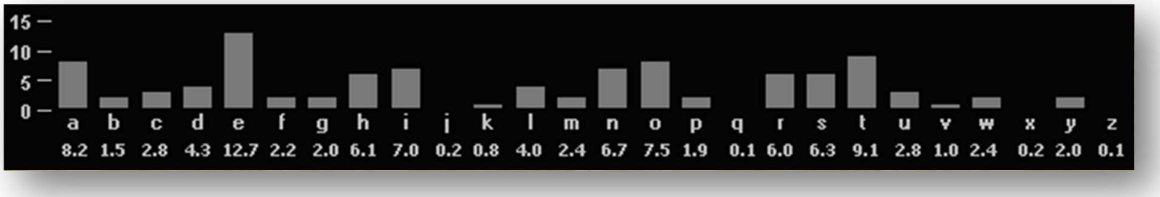


KEY	PHHW	PH	DIWHU	WKH	WRJD	SDUWB
1	oggv	og	chvgt	vjg	vqic	rctva
2	nffu	nf	bgufs	uif	uphb	qbsuz
3	meet	me	after	the	toga	party
4	ldds	ld	zesdq	sgd	snfz	ozqsx
5	kccr	kc	ydrcp	rfc	rmey	nyprw
6	jbbq	jb	xcqbo	qeb	qldx	mxoqv
7	iaap	ia	wbpan	pda	pkcw	lwnpu
8	hzzo	hz	vaozm	ocz	ojbv	kvmot
9	gyyn	gy	uznyl	nby	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

Frequency Analysis: This algorithm is structured based on the regularities of the language. If you have got a message encrypted using the substitution cipher that you want to crack, you can use frequency analysis. In other words, if the sender has tried to disguise a letter by replacing with a different letter, you can still recognise the original letter because the frequency characteristics of the original letter will be passed on to the new letters.

To apply frequency analysis, we need to know the frequency of every letter in the English alphabet, or the frequency of letters in whichever language the sender is using.

Below is a list of average frequencies for letters in the English language. So, for example, the letter E accounts for 12.7% of all letters in English, whereas Z accounts for 0.1 %. All the frequencies are tabulated and plotted below.



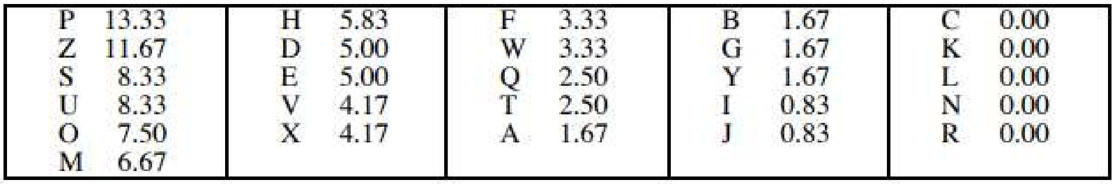
Please note, these frequencies are averages, and E will not always constitute 12.7 % of all the letters in a text, and may not even be the most common letter. The longer the message, the more likely it is that will obey the average distribution shown above. However, there are exceptions to this rule. In 1969, the French author Georges Perec managed to write a 200-page book called 'La Disparition' without using any words containing the letter E. Amazingly, the book was later translated into English by Gilbert Adair, again avoiding the use of the letter E. along with letter frequencies, this cipher also uses the analysis of most frequently occurring digrams, trigrams, and four letter …etc. Consider the following cipher text:

UZQSOVUOHXMOPVGPOZPEVSGZWSZOPFPESXUDBMETSXAIZ

VUEPHZHMDZSHZOWSFPAPPDTSVPQUZWYMXUZUHSX

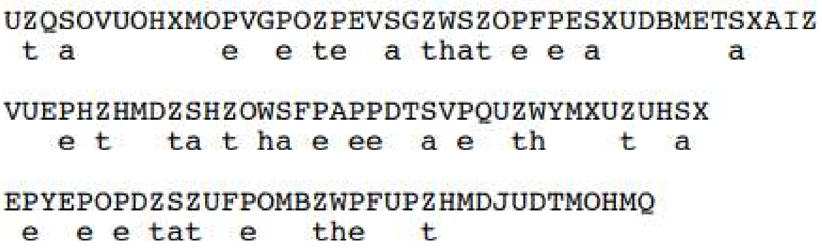
EPYEPOPDZSZUFPOMBZWPFUPZHMDJUDTMOHMQ

At first, the relative frequency of the letters can be determined and compared to a standard frequency distribution for English. If the message were long enough, this technique alone might be sufficient, but because this is a relatively short message, we cannot expect an exact match. In any case, the relative frequency of the letters in the cipher text (in percentages) are as follows:



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								

Comparing these frequencies, it seems that cipher letters P and Z are the equivalents of plain letters e and t. There are a number of ways to proceed at this point. We could make some tentative assignments and start to fill in the plaintext to see if it looks like a reasonable "skeleton" of a message. A more systematic approach is to look for other regularities. For example, certain words may be known to be in the text. Or we could look for repeating sequences of cipher letters and try to deduce their plaintext equivalents. A powerful tool is to look at the frequency of two-letter combinations, known as digrams. 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. Then, by our earlier hypothesis, we can equate P with e. Now notice that the sequence ZWP appears in the ciphertext, and we can now translate that sequence as "the." This is the most frequent trigram (three-letter combination) in English, which seems to indicate that 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. Therefore, 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
```

Only four letters have been identified, but already we have quite a bit of the message. Continued analysis of frequencies plus trial and error should easily yield a solution from this point. The complete plaintext, with spaces added between words, follows:

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

Play fair Cipher: This cipher was actually invented by British scientist Sir Charles Wheatstone in 1854, but it bears the name of his friend Baron Play fair of St. Andrews. This is the best-known multiple-letter encryption cipher, which treats digrams in the plaintext as single units and translates these units into cipher text digrams.

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

In this case, the keyword is Computer. The matrix is constructed by filling with the letters of the keyword from left to right and from top to bottom, and then fill the remaining matrix with the remaining letters in alphabetic order. Since the size of matrix is 5X5 we combine the letters I and J in a single block. Plaintext is encrypted two letters at a time, according to the following rules:

1. Repeating plaintext letters that would fall in the same pair are separated with a filler letter, such as x,

Ex: balloon would be enciphered as ba lx lo on.

2. 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.

Ex: mu is encrypted as PC.

3. Plaintext letters that fall in the same column are each replaced by the letter beneath, with the top element of the row circularly following the last.

Ex: cv is encrypted as TC.

4. Otherwise, each plaintext letter is replaced by the letter that lies in its own row and the column occupied by the other plaintext letter.

Ex: hl becomes FQ and pd becomes CH

Security Analysis: The Play fair cipher is a great advance over simple mono alphabetic ciphers. For one thing, Whereas there are only 26 letters, there are 26 ¥ 26 = 676 digrams, so that identification of individual digrams is more difficult. Furthermore, the relative frequencies of individual letters exhibit a much greater range than that of digrams, making frequency analysis much more difficult. For these reasons, the Play fair cipher was for a long time considered unbreakable

Hill Cipher: Another interesting multiletter cipher is the Hill cipher, developed by the mathematician Lester Hill in 1929. The encryption algorithm takes m successive plaintext letters and substitutes for them m ciphertext letters. The substitution is determined by m linear equations in which each character is assigned a numerical value (a = 0, b = 1, … z = 25). For m = 3, the system can be described as follows:

$$C_1 = (k_{11}p_1 + k_{12}p_2 + k_{13}p_3) \text{ mod } 26$$

$$C_2 = (k_{21}p_1 + k_{22}p_2 + k_{23}p_3) \text{ mod } 26$$

$$C_3= (k_{31}p_1 + k_{32}p_2 + k_{33}p_3) \text{ mod } 26$$

This can be expressed in terms of column vectors and matrices:

$$\begin{pmatrix} C_1 \\ C_2 \\ C_3 \end{pmatrix} = \begin{pmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} \bmod 26$$

Or C=KP mod 26

Here C and P are column vectors of length 3, representing the plaintext and cipher text, and K is a 3¥3 encryption matrix. All operations are performed on mod 26.

Ex: consider the plaintext "paymoremoney",

And the encryption key

$$\mathbf{K} = \begin{pmatrix} 17 & 17 & 5 \\ 21 & 18 & 21 \\ 2 & 2 & 19 \end{pmatrix}$$

Hill cipher encrypts one block at a time. Here the block size is 3. So the first three letters of the plain text are represented in a column matrix (15 0 24).

Then K(15 0 24) = (375 819 486) mod 26

= (11 13 18)

= LNS.

Continuing this process till the length of plain text which yields the following cipher text LNSHDLEWMTRW.

Decryption requires using the inverse of the matrix K. The inverse K^{-1} of a matrix K is defined by the equation $KK^{-1} = K^{-1}K = I$, where I is the matrix that is all zeros except for ones along the main diagonal from upper left to lower right. In this case, the inverse is:

$$\mathbf{K}^{-1} = \begin{pmatrix} 4 & 9 & 15 \\ 15 & 17 & 6 \\ 24 & 0 & 17 \end{pmatrix}$$

C = EK(P)

= KP mod 26

P = DK(C)

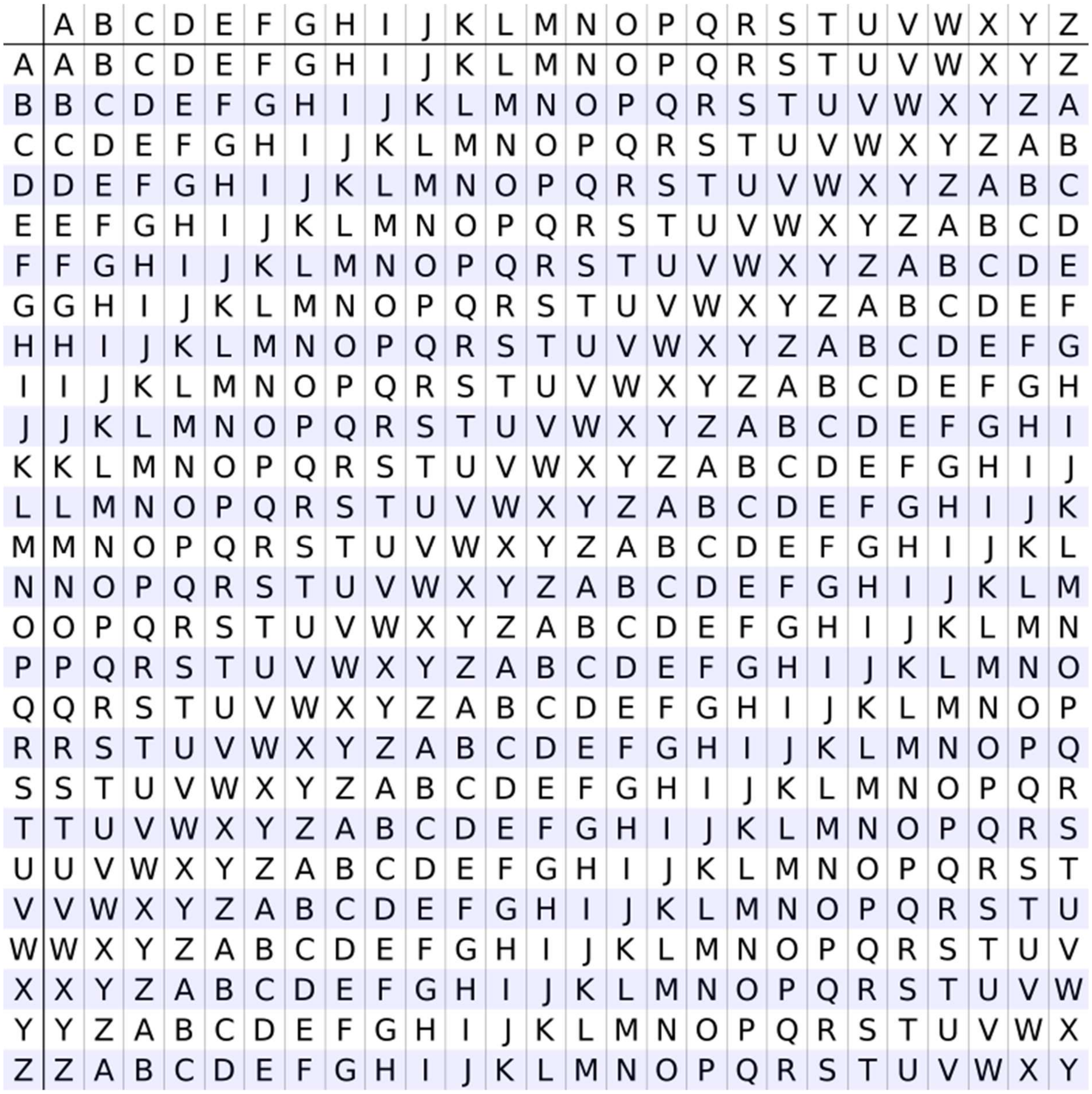
= K^{-1}C mod 26

= K^{-1}KP

= P

Polyalphabetic Ciphers

Another way to improve mono alphabetic technique is to use different mono alphabetic substitutions. The general name for this approach is polyalphabetic substitution cipher. The best-known, and one of the simplest, such algorithms is referred to as the Vigenère cipher. In this scheme, the set of related mono alphabetic substitution rules consists of the 26 Caesar ciphers, with shifts of 0 through 25. Each cipher is denoted by a key letter, which is the cipher text letter that substitutes for the plaintext letter a. To aid in understanding the scheme and to aid in its use, a matrix known as the Vigenère tableau is constructed.



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

The process of encryption is simple: Always plain text characters signify column labels and key characters signify row labels. Now, the cipher text character corresponds to the element which is in plain text character column and key character row. To encrypt a message, a key is needed that is as long as the message. User can select any key word of any length. It may be a repeating keyword.

Ex: Plain Text: we are discovered save your self

Key : deceptive

d e c e p t I v e d e c e p t I v e d e c e p t i v e

w e a r e d I s c o v e r e d s a v e y o u r s e l f

ciphertext: Z I C V T W Q N G R Z G V T W A V Z H C Q Y G L M G J

Decryption is equally simple. The key letter again identifies the row. The position of the cipher text letter in that row determines the column, and the plaintext letter is at the top of that column. The strength of this cipher is that there are multiple cipher text letters for each plaintext letter, one for each unique letter of the keyword. A Vigenère cipher is suspected, then progress depends on determining the length of the keyword, as will be seen in a moment. For now, let us concentrate on how the keyword length can be determined. If two identical sequences of plaintext letters occur at a distance that is an integer multiple of the keyword length, they will generate identical cipher text sequences. In this foregoing example, two instances of the sequence "red" are separated by 9 character positions. Consequently, in both cases, r is encrypted using key letter e, e is encrypted using key letter p, and d is encrypted using key letter t. Thus, in both cases the cipher text sequence is VTW. An analyst looking at only the cipher text would detect the repeated sequences VTW at a displacement of 9 and make the assumption that the keyword is either 3 or 9 letters in length. The periodic nature of the keyword can be eliminated by using a nonrepeating keyword that is as long as the message itself. Vigenère proposed what is referred to as an autokey system, in which a keyword is concatenated with the plaintext itself to provide a running key.

Ex:

key: deceptivewearediscoveredsav

plain text: wearediscoveredsaveyourself

ciphertext: ZICVTWQNGKZEIIGASXSTSLVVWLA

Even this scheme is vulnerable to cryptanalysis. Because the key and the plaintext share the same frequency distribution of letters, a statistical technique can be applied. The ultimate defense against such a cryptanalysis is to choose a keyword that is as long as the plaintext and has no statistical relationship to it. Such a system was introduced by an AT&T engineer named Gilbert Vernam in 1918. His system works on binary data rather than letters. The system can be expressed succinctly as follows:

Ci = $p_i + k_i$ where

p_i = ith binary digit of plaintext

k_i = ith binary digit of key

C_i = ith binary digit of ciphertext

+ = exclusive-or (XOR) operation

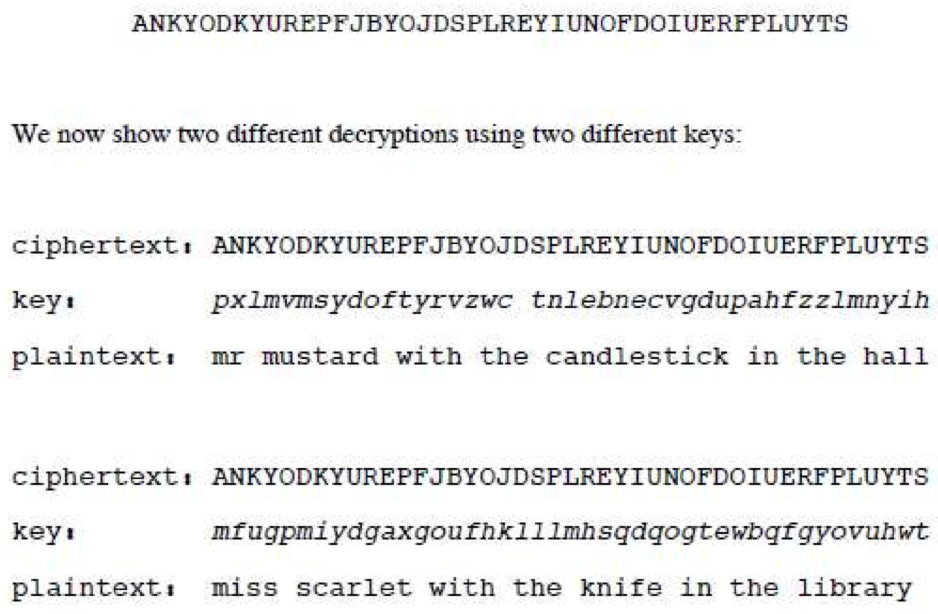
Thus, the ciphertext is generated by performing the bitwise XOR of the plaintext and the key. Because of the properties of the XOR, decryption simply involves the same bitwise operation:

$P_i = C_i + k_i$

One-Time Pad

An Army Signal Corp officer, Joseph Mauborgne, proposed an improvement to the Vernam cipher that yields the ultimate in security. Mauborgne suggested using a random key that was truly as long as the message, with no repetitions. Such a scheme, known as a one-time pad, is unbreakable. It produces random output that bears no statistical relationship to the plaintext. Because the cipher text contains no information whatsoever about the plaintext, there is simply no way to break the code.

Ex: Consider the following cipher text:



ANKYODKYUREPFJBYOJDSPLREYIUNOFDOIUERFPLUYTS

We now show two different decryptions using two different keys:

```
ciphertext: ANKYODKYUREPFJBYOJDSPLREYIUNOFDOIUERFPLUYTS
key:        pxlmvmsydoftyrvzwc tnlebnecvgdupahfzzlmnyih
plaintext:  mr mustard with the candlestick in the hall

ciphertext: ANKYODKYUREPFJBYOJDSPLREYIUNOFDOIUERFPLUYTS
key:        mfugpmiydgaxgoufhklllmhsqdqogtewbgfgyovuhwt
plaintext:  miss scarlet with the knife in the library
```

Transposition Techniques: Substitution techniques substitute or replace each plain text character with another plain text character whereas transposition cipher rearrange the letters of the plain text to get cipher text.

a) Rail fence Cipher: In this cipher plain text is written down as a sequence of diagonals(columns) and then read off as a sequence of rows.

   Ex: meet me after the toga party

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

e t e f e t e o a a t

MEMATRHTGPRYETEFETEOAAT

Columnar Transposition Cipher: This is a somewhat more complex scheme than the above where the message (plain text) is written row by row in a matrix and read off cipher text column by column. The secret key which is used in this cipher is a non repeatable letter key word. Where each letter is given an index based on its alphabetical order and the same is treated as a column index.

Ex: Plain Text : attack postponed until two am

And Key : computer

1	4	3	5	8	7	2	6
C	O	M	P	U	T	E	R
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	u	v	w	x	y	z	a

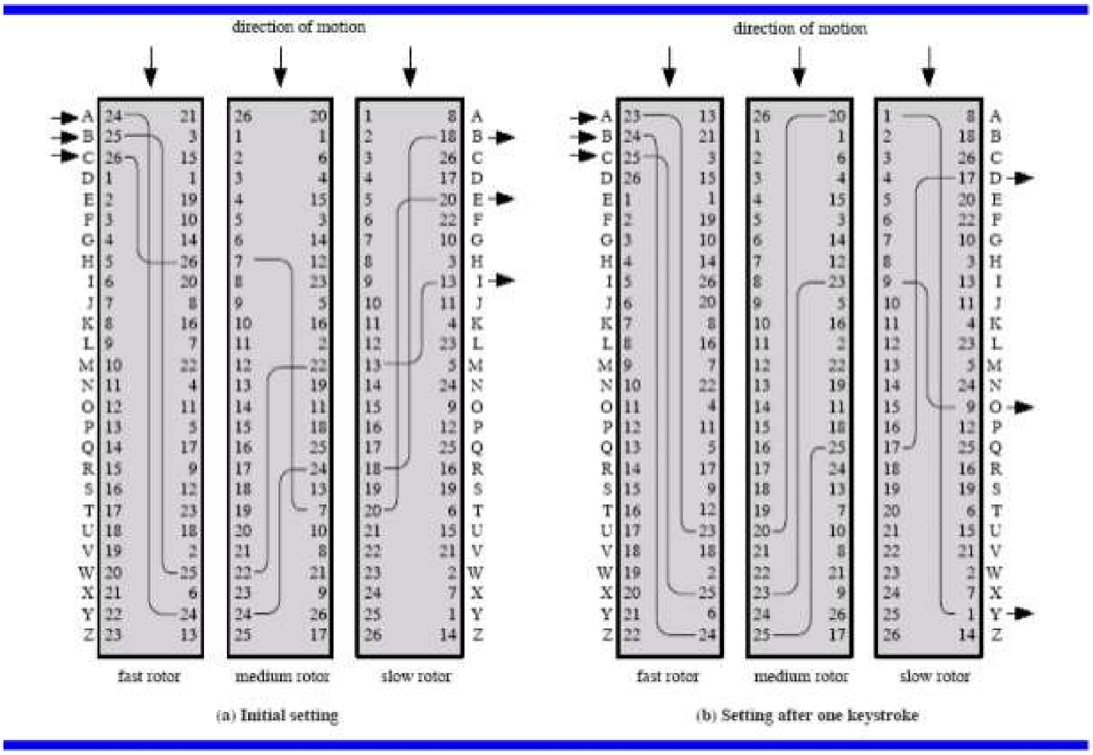
Cipher Text: ASNMPDOZTPIVTTTUAOLWOUAAKEWYCNTX

The transposition cipher can be made significantly more secure by performing more than one stage of transposition. The result is a more complex permutation that is not easily reconstructed.

Rotor Machines: The example just given suggests that multiple stages of encryption can produce an algorithm that is significantly more difficult to cryptanalyze. Before the introduction of DES, the most important application of the principle of multiple stages of encryption was a class of systems known as rotor machines. The basic principle of the rotor machine is illustrated in the following figure. The machine consists of a set of independently rotating cylinders through which electrical pulses can flow. Each cylinder has 26 input pins and 26 output pins, with internal wiring that connects each input pin to a unique output pin. For simplicity, only three of the internal connections in each cylinder are shown.

If we associate each input and output pin with a letter of the alphabet, then a single cylinder defines a monoalphabetic substitution. For example, if an operator depresses the key for the letter A, an electric signal is applied to the first pin of the first cylinder and flows through the internal connection to the twenty-fifth output pin. Consider a machine with a single cylinder.

THREE-ROTOR MACHINES



(a) Initial setting

(b) Setting after one keystroke

After each input key is depressed, the cylinder rotates one position, so that the internal connections are shifted accordingly. Thus, a different mono alphabetic substitution cipher is defined. After 26 letters of plaintext, the cylinder would be back to the initial position. Thus, we have a polyalphabetic substitution algorithm with a period of 26.

A single-cylinder system is trivial and does not present a formidable cryptanalytic task. The power of the rotor machine is in the use of multiple cylinders, in which the output pins of one cylinder are connected to the input pins of the next. Figure 2.7 shows a three-cylinder system. The left half of the figure shows a position in which the input from the operator to the first pin (plaintext letter a) is routed through the three cylinders to appear at the output of the second pin (cipher text letter B. With multiple cylinders, the one farthest from the operator input rotates one pin position with each keystroke. The right half of Figure 2.7 shows the system's configuration after a single keystroke. For every complete rotation of the outer cylinder, the middle cylinder rotates one pin position. Finally, for every complete rotation of the middle cylinder, the inner cylinder rotates one pin position. This is the same type of operation seen with an odometer. The result is that there are 26 ¥ 26 ¥ 26 = 17,576 different substitution alphabets used before the system repeats. The addition of fourth and fifth rotors results in periods of 456,976 and 11,881,376 letters, respectively.

Steganography: A plaintext message may be hidden in one of two ways. The methods of steganography conceal the existence of the message, whereas the methods of cryptography render the message unintelligible to outsiders by various transformations of the text. A simple form of steganography, is one in which an arrangement of words or letters within an apparently innocuous text spells out the real message. For example, the sequence of first letters of each word of the overall message spells out the hidden message. Various other techniques have been used historically; some examples are the following:

- Character marking: Selected letters of printed or typewritten text are overwritten in pencil. The marks are ordinarily not visible unless the paper is held at an angle to bright light.
- Invisible ink: A number of substances can be used for writing but leave no visible trace until heat or some chemical is applied to the paper.
- Pin punctures: Small pin punctures on selected letters are ordinarily not visible unless the paper is held up in front of a light.
- Typewriter correction ribbon: Used between lines typed with a black ribbon, the results of typing with the correction tape are visible only under a strong light.

Some other modern techniques available are:

1. Text Steganography
2. Image Steganography
3. Video Steganography
4. Audio Steganography
5. Linguistic Steganography