

Data Security and Privacy

Chapter 8:

Using Encryption:

Overview of Cryptographic tools for Data Security.

Introduction

- **Cryptology:**
 - science of encryption; combines cryptography and cryptanalysis.
- **Cryptography: (kryptos, meaning hidden; graphein, meaning to write)**
 - process of making and using codes to secure transmission of information
- **Cryptanalysis:**
 - process of obtaining original message (or plaintext) from encrypted message (or ciphertext) without knowing algorithms and keys used to perform the encryption.
- **Encryption:**
 - converting original message into a form unreadable by **unauthorized individuals**.
- **Decryption:**
 - the process of converting the ciphertext message back into plaintext

Foundations of Cryptology

- Cryptology has a long and multicultural history
- With emergence of technology, need for encryption in information technology environment greatly increased
- All popular Web browsers use built-in encryption features for secure e-commerce applications.

Cryptography

- characterize cryptographic system by:
 - type of **encryption operations** used
 - substitution / transposition / product
 - **number of keys** used
 - single-key or private / two-key or public
 - way in which plaintext is processed
 - **block / stream**

Some Basic Terminology

- **Algorithm:** mathematical formula used to convert an unencrypted message into an encrypted message.
- **Cipher:** The transformation of the individual components (characters, bytes, or bits) of an unencrypted message into encrypted components.
- **Ciphertext or cryptogram:** The encrypted or encoded message resulting from an encryption.
- **Decipher:** To decrypt or convert ciphertext to plaintext.
- **Encipher:** To encrypt or convert plaintext to ciphertext.
- **Key or cryptovariable:** The information used **in conjunction** with the algorithm to create ciphertext from plaintext.
- **Plaintext:** The original **unencrypted message** that is encrypted and results from successful decryption.
- **Steganography:** The process of hiding messages in a picture or graphic.

Cipher Methods

- Plaintext can be encrypted through **bit stream** or **block cipher** method
- **Bit stream:**
 - each plaintext bit transformed into cipher bit one bit at a time
- **Block cipher:**
 - message divided into blocks (e.g., sets of 8- or 16-bit blocks) and each is transformed into encrypted block of cipher bits using algorithm and key.
- **M:** represent the original message
- **C:** represent the ending ciphertext
- **E:** represent the encryption process: $E(M) = C$
- **K:** represent the key, thus $E(M, K) = C$
- **D:** represents the decryption or deciphering process, $D[E(M)] = M$.

Encryption Operations

- The most commonly used algorithms include two functions: **substitution and transposition**.

Substitution Cipher:

- Substitute one value for another (*very powerful if combined with other operations*).
- **Monoalphabetic substitution:**
 - uses only one alphabet
- **Polyalphabetic substitution:**
 - more advanced; uses two or more alphabets
- **Vigenère cipher:**
 - advanced cipher type that uses simple polyalphabetic code; made up of 26 distinct cipher alphabets
- **Caesar Cipher:** used a three-value shift to the right.(used in military)
 - replaces each letter by 3rd letter on
 - ex: meet me after the toga party
PHHW PH DIWHU WKH WRJD SDUWB

It has the alphabet written out 26 times in different rows.
 Each row starts with a key letter. The rest of the row holds the letters A to Z (in shifted order).

The person sending the message chooses a **keyword** and repeats it until it matches the length of the plaintext.

Ex:
 text to be encrypted:
 ATTACKATDAWN.

Key:LEMON

ATTACKATDAWN
 LEMONLEMONLE

LXFOPVEFRNHR(Ciphertext)

	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

Table 8-2 The Vigenère Square

Ceasar Cipher

- can define transformation as:

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
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

- mathematically give each letter a number

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

- then have Caesar cipher as:

$$c = E(p) = (p + k) \bmod (26)$$

$$p = D(c) = (c - k) \bmod (26)$$

- Example:

– howdy (7,14,22,3,24) encrypted using key f (ie a shift of 5) is MTBID

Transposition Cipher(permutation cipher)

- Easy to understand, but if properly used, produces ciphertext that is [difficult to decipher](#).
- Rearranges values within a block to create ciphertext.
- move these bits or bytes to another location in the block, without altering the actual letters used
 - so that bit 1 becomes bit 4, bit 2 becomes bit, 7 etc.
- Can be done at the bit level or at the byte (character) level
- To make the encryption even stronger, the keys and block sizes can be made much larger
- A pure transposition cipher is easily recognized because it has the same letter frequencies as the original plaintext.

Rail Fence cipher

- write message letters **out diagonally** over a number of rows
- then read off cipher row by row
- Message: meet me after the toga party
- eg. write message out as:

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

- giving ciphertext

```
MEMATRHTGPRYETEFETEOAAT
```

Row Transposition Ciphers

- A more complex transposition
- write letters of message out in rows over a specified number of columns
- then reorder the columns according to some key before reading off the rows

Plaintext:

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

Key: 3 4 2 1 5 6 7

Ciphertext: TTNAAPTMTSUOAODWCOIXKNLYPETZ

Transposition Ciphers

Key: 3 4 1 2

M=COUNT THE COST $\Rightarrow$

1	2	3	4
C	N	H	O
O	T	E	S
U	T	C	T

 $\xrightarrow{\text{行换位}}$

1	2	3	4
H	O	C	N
E	S	O	T
C	T	U	T

$\Rightarrow$ C= HOCN ESOT CTUT(横取)

解密

C $\Rightarrow$

1	2	3	4
H	O	C	N
E	S	O	T
C	T	U	T

 $\xrightarrow{\text{取 } 3412}$

1	2	3	4
C	N	H	O
O	T	E	S
U	T	C	T

 $\Rightarrow$ M=COUNT THE COST (直取)

Product Ciphers

- ciphers using **substitutions or transpositions** are not secure because of language characteristics
- hence consider using several ciphers in succession to make harder, but:
 - two substitutions make a more complex substitution
 - two transpositions make more complex transposition
 - but a substitution followed by a transposition makes a new much harder cipher
- this is bridge from classical to modern ciphers

Exclusive OR (XOR)

- Function of Boolean algebra; two bits are compared
 - If two bits are **identical**, result is binary **0**
 - If two **bits not identical**, result is binary **1**
- A very simple symmetric cipher that is used in many applications where security is not a defined requirement

First Bit	Second Bit	Result
0	0	0
0	1	1
1	0	1
1	1	0

Table 8-3 XOR Truth Table

Vernam Cipher(one-time pad)

- Uses set of characters once per encryption process(use of a **one-time pre-shared key** same size as the message being sent)
- To perform:
 - The pad values are added to numeric values that represent the plaintext that needs to be encrypted
 - Each character of the plaintext is turned into a number ($a=0, b=1, \dots, z=25$) and a pad value for that position is added
 - The resulting sum for that character is then converted back to a ciphertext letter for transmission
 - If the sum of the two values exceeds 26, then 26 is subtracted from the total(Modulo 26).

Vernam Cipher(one-time pad)

Plain-Text: RAMSWARUPK

Key: RANCHOBABA

PT: R A M S W A R U P K

NO: 17 0 12 18 22 0 17 20 15 10

KEY: R A N C H O B A B A

NO: 17 0 13 2 7 14 1 0 1 0

CT-NO: 34 0 25 20 29 14 18 20 16 10

CT-NO: 8 0 25 20 3 14 18 20 16 10

CIPHER-TEXT: I A Z U D O S U Q K

Book or Running Key Cipher

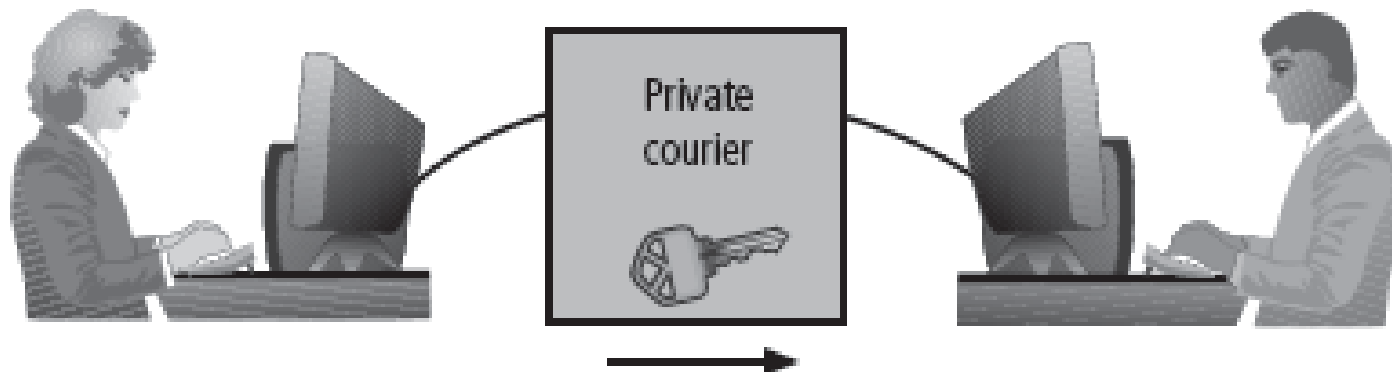
- Uses **text in book as key** to encrypt a message
- same internal workings as Vigenere cipher.
- Difference in how the key is chosen;
 - Vigenere cipher uses a short key that repeats.
 - Running key cipher uses a long key such as an **excerpt from a book**.
 - key does not repeat, making cryptanalysis more difficult.
- Typical sources are dictionaries and thesauruses

Cryptographic Algorithms

- Often grouped into two broad categories, **symmetric and asymmetric**
 - Today's popular cryptosystems use hybrid **combination** of **symmetric and asymmetric** algorithms.
- Symmetric and asymmetric algorithms distinguished by **types of keys** used for encryption and decryption operations.

Symmetric Encryption

- Uses same “secret key” to encipher and decipher message.
 - Encryption methods can be extremely efficient, requiring minimal processing.
 - Both sender and receiver must possess encryption key.
 - If either copy of key is compromised, an intermediate can decrypt and read messages.



Rachel at ABC Corp. generates a secret key. She must somehow get it to Alex at XYZ Corp. out of band. Once Alex has it, Rachel can use it to encrypt messages, and Alex can use it to decrypt and read them.

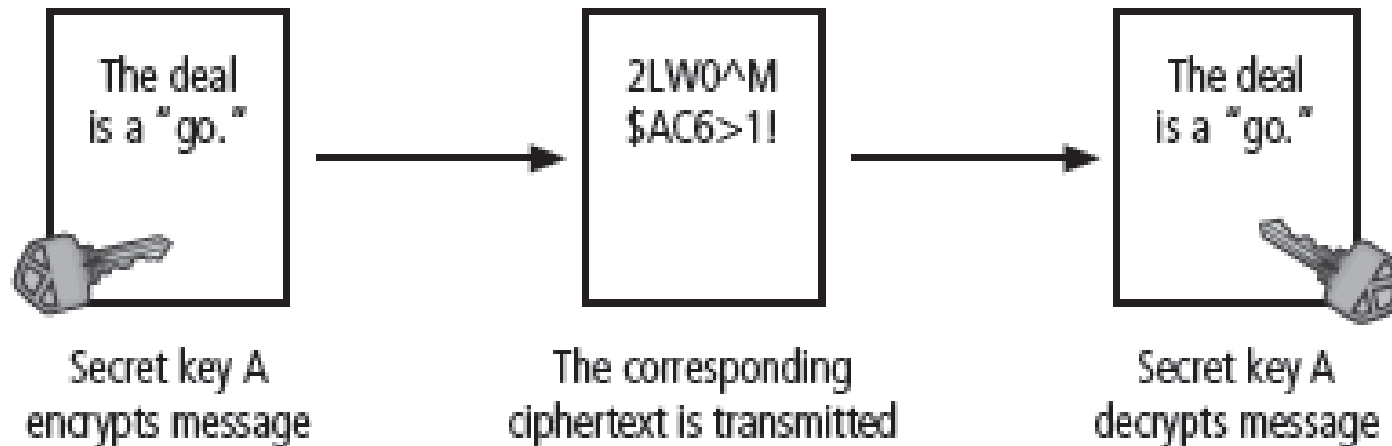


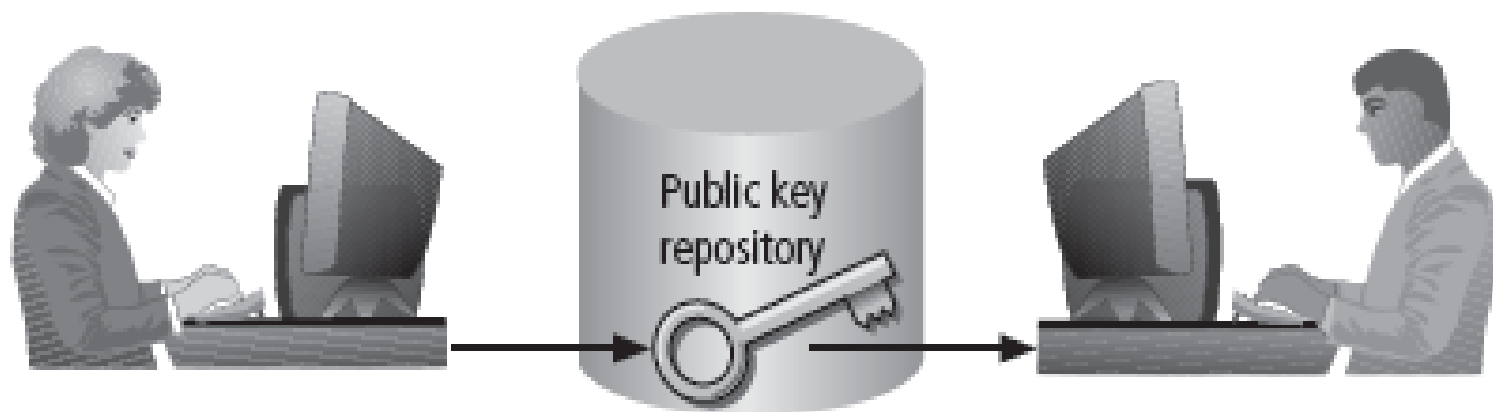
Figure 8-5 Example of Symmetric Encryption

Symmetric Encryption (cont'd.)

- **Data Encryption Standard (DES):** one of most popular symmetric encryption cryptosystems
 - 64-bit block size; 56-bit key
 - Adopted by NIST in 1976 as federal standard for encrypting non-classified information
 - Due to limitations of DES (small key and block sizes) NIST started a open process to select a new block cipher.
- **Triple DES (3DES):**
 - created to provide security far beyond DES
- **Advanced Encryption Standard (AES):**
 - developed to replace both DES and 3DES

Asymmetric Encryption(public-key encryption)

- Uses two different but related keys
 - Either key can encrypt or decrypt message.
 - If Key A encrypts message, only Key B can decrypt.
 - One key serves as private key(known only to the owner of the key pair) and the other serves as public key(stored in a public location).
- The problem with asymmetric encryption is that it requires four keys to hold a single conversation between two parties.
- Asymmetric encryption is not as efficient as symmetric encryptions in terms of CPU computations.
- Ex:
 - **RSA**(Rivest–Shamir–Adleman) algorithm, used by modern computers to encrypt and decrypt messages.



Alex at XYZ Corp. wants to send a message to Rachel at ABC Corp. Rachel stores her public key where it can be accessed by anyone. Alex retrieves Rachel's key and uses it to create ciphertext that can be decrypted only by Rachel's private key, which only she has. To respond, Rachel gets Alex's public key to encrypt her message.

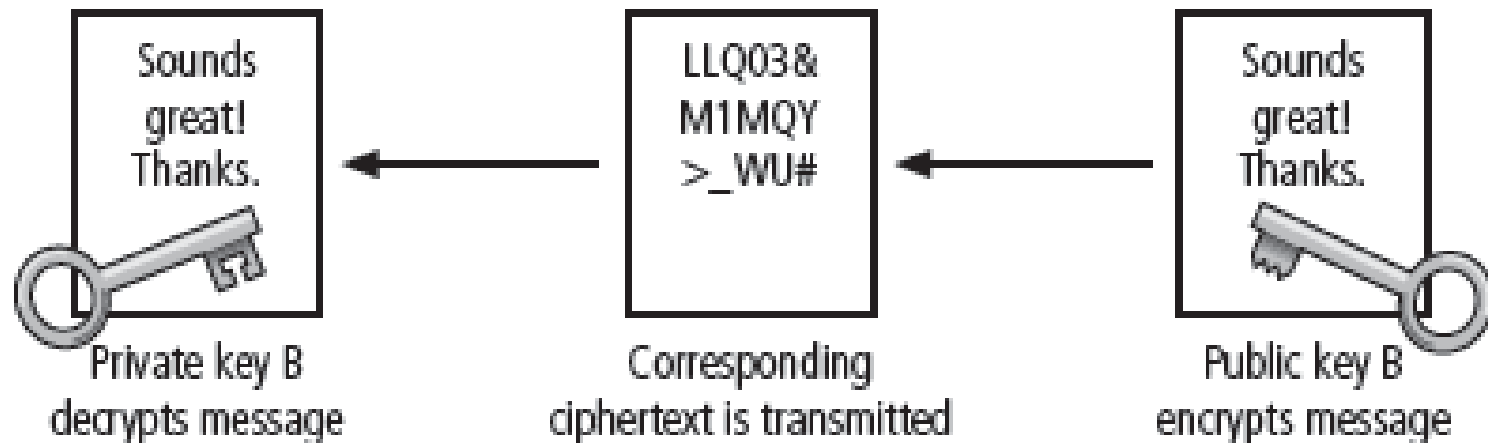


Figure 8-6 Example of Asymmetric Encryption

Encryption Key Size

- When using ciphers, size of key is very important.
- Strength of many encryption applications and cryptosystems measured by **key size**.
 - The **key** must be long enough so that an attacker **cannot try all possible combinations**.
- For cryptosystems, security of encrypted data is not dependent on keeping encrypting algorithm secret.
 - algorithms are often published, so that research to uncover their weaknesses can be done.
- Cryptosystem security depends on keeping some or all key(s) secret.

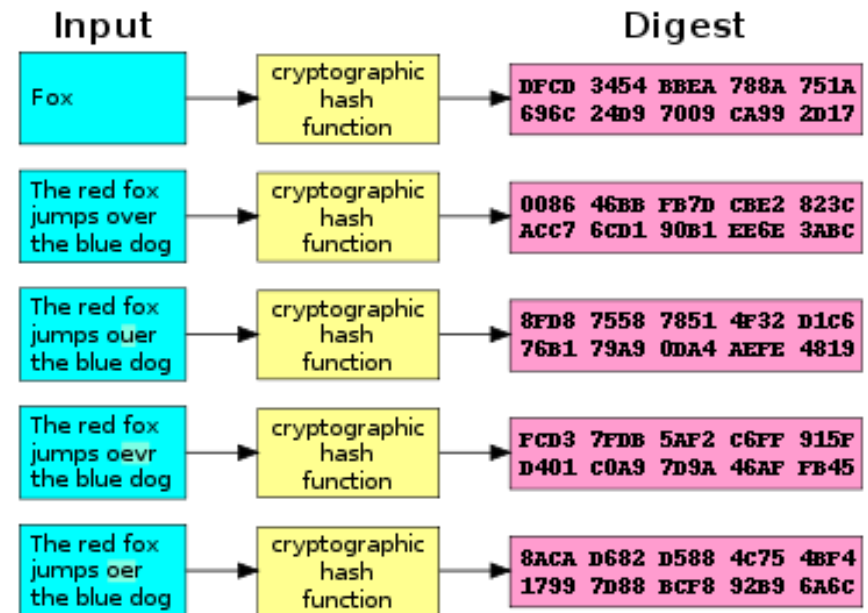
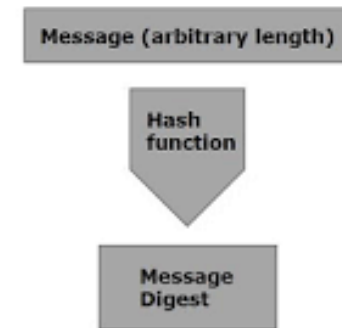
Hash Functions

Data Integrity and Source Authentication

- Encryption does not protect data from modification by another party.
- Need a way to ensure that data arrives at destination **in its original form** as sent by the sender and it is coming from an **authenticated source**.

Hash Functions

- Mathematical algorithms that generate **message summary** known as **message digest**.
 - publicly known functions that create hash value (message digest).
 - Converting **variable-length messages** into a single **fixed-length value**.
 - Hashing functions do not require the use of keys.
- Ex:
 - used in password verification systems to confirm identity of user.

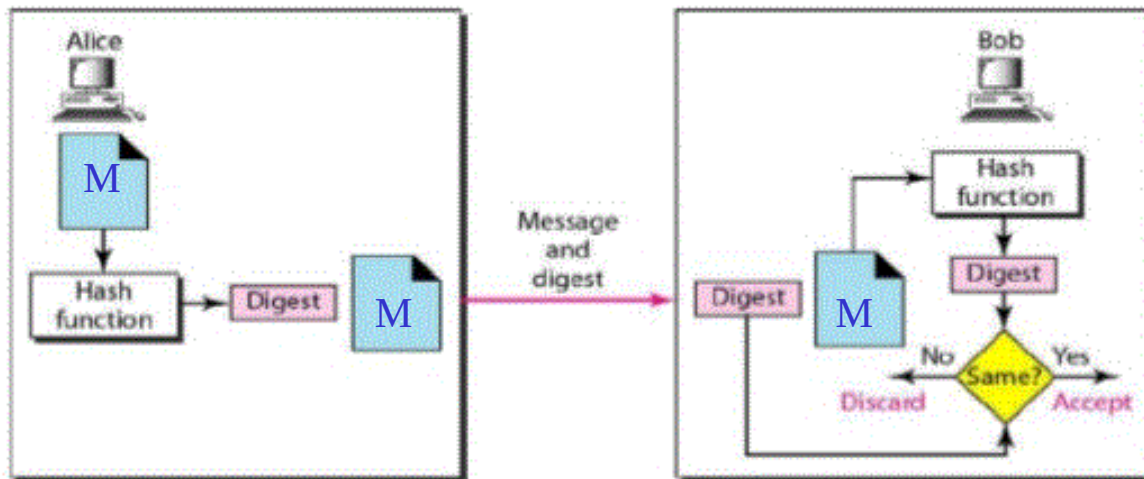


Hash Functions

- The message digest is :
 - a *fingerprint* of the author's message that is to be compared with the receiver's locally calculated hash of the same message.

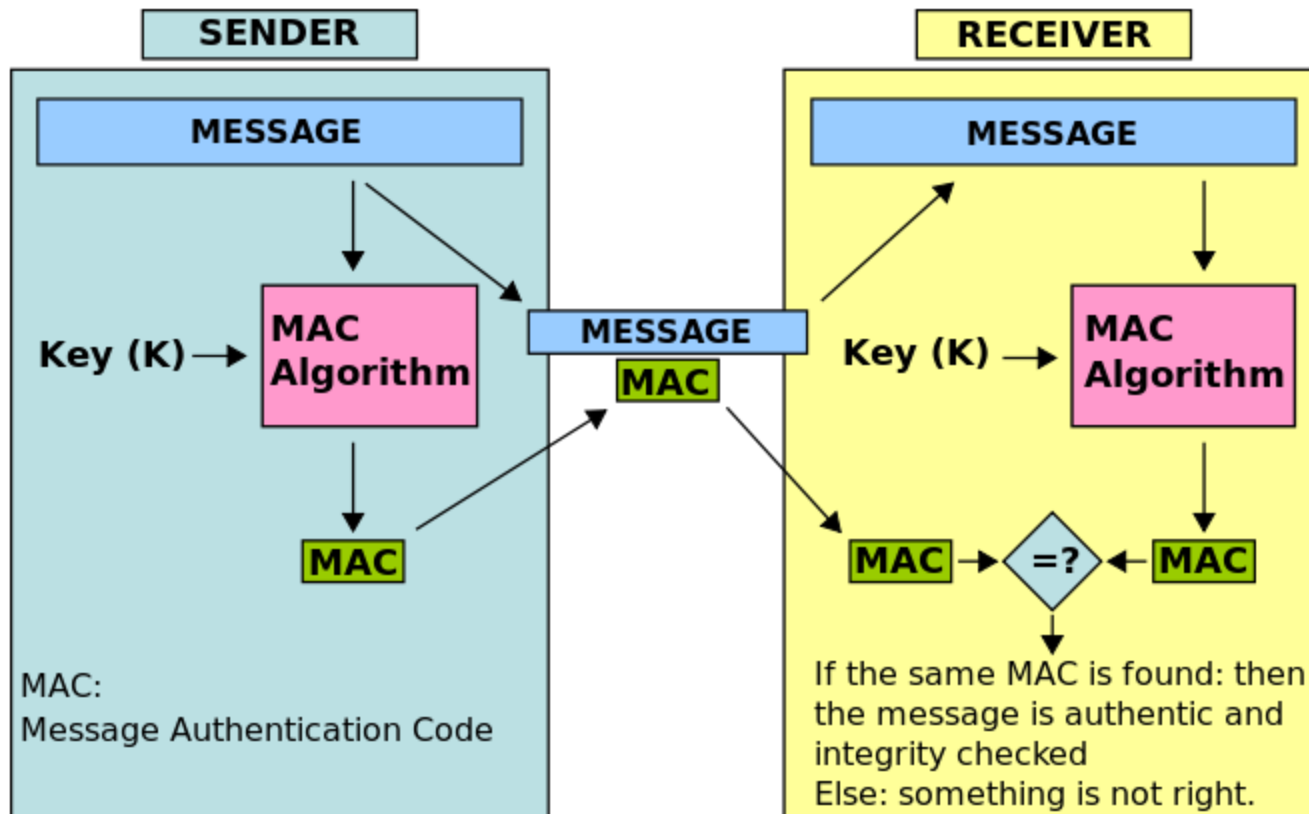


- confirm no content has changed (message integrity).



Message Authentication Code(MAC)

- A MAC scheme is a **hash family**, used for message authentication.
 - The sender and the receiver **share a key(K)**.



Cryptographic Tools

- Potential areas of use include:
 - Ability to conceal the contents of sensitive messages.
 - Verify the **contents of messages** and the **identities** of their senders.
- Tools must **embody** cryptographic capabilities so that they can be applied to the everyday world of computing.

Public-Key Infrastructure (PKI)

- **IS** an Integrated system of:
 - software,
 - encryption methodologies,
 - protocols,
 - legal agreements,
 - and third-party services enabling users to communicate securely.
- PKI systems based on **public-key** cryptosystems.
- PKI **protects information** assets in several ways:
 - Authentication
 - Integrity
 - Privacy
 - Authorization
 - Nonrepudiation

Public-Key Infrastructure (PKI)

- **Authentication:**

- Digital certificates in a PKI system permit parties to **validate the identity of other** parties in an Internet transaction.

- **Integrity:**

- A digital certificate demonstrates that the **content signed by the certificate has not been altered** while being moved from server to client.

- **Privacy:**

- Digital certificates **keep information from being intercepted** during transmission over the Internet.

- **Authorization:**

- Digital certificates issued in a PKI environment can replace user IDs and passwords, **enhance security**, and reduce some of the overhead required for **authorization processes and controlling access privileges**.

- **Non repudiation:**

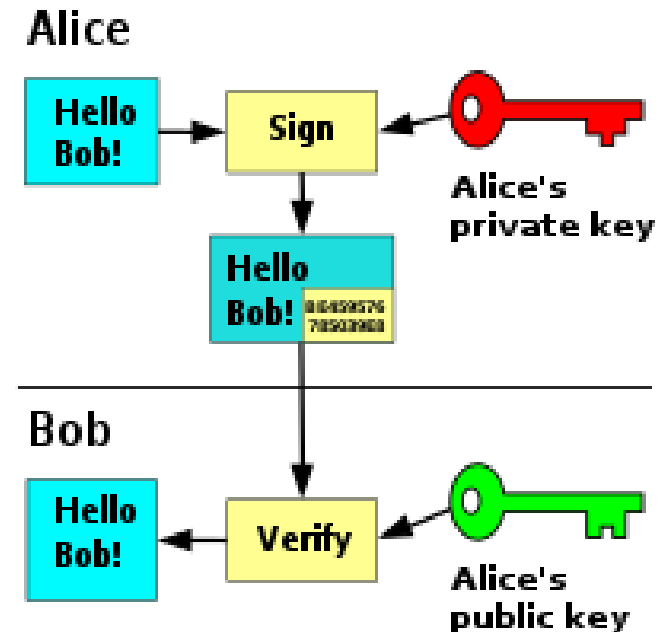
- Digital certificates can validate actions, making it less likely that customers or partners **can later repudiate a digitally signed transaction**.

Public-Key Infrastructure (PKI) (cont'd.)

- PKI solution protects the transmission and reception of secure information by integrating:
 - A certificate authority (CA)
 - A registration authority (RA)
 - Certificate directories
 - Management protocols
 - Policies and procedures

Digital Signatures

- Created in response to rising need to **verify information transferred using electronic systems.**
- A digital signature is a mathematical scheme for **verifying the authenticity** of digital messages or documents.
- Asymmetric encryption** processes used to **create digital signatures.**
- digital signature = authentication tag of a message.
- IT can only be computed by **one user**, but can be **verified by many.**



Digital Signatures

- A valid digital signature gives a recipient very strong reason to believe that the message was:
 - created by a known sender (authentication)
 - was not altered in transit(integrity).
- Non-repudiation:
 - the process that verifies the message was sent by the sender and thus cannot be refuted.
- Ex:
 - Digital Signature Standard (DSS):
 - suite of algorithms that the US federal government has approved for generating and verifying a **digital signature**.
 - **ECDSA**: Elliptic Curve Digital Signature Algorithm

Digital Certificates(public key certificate)

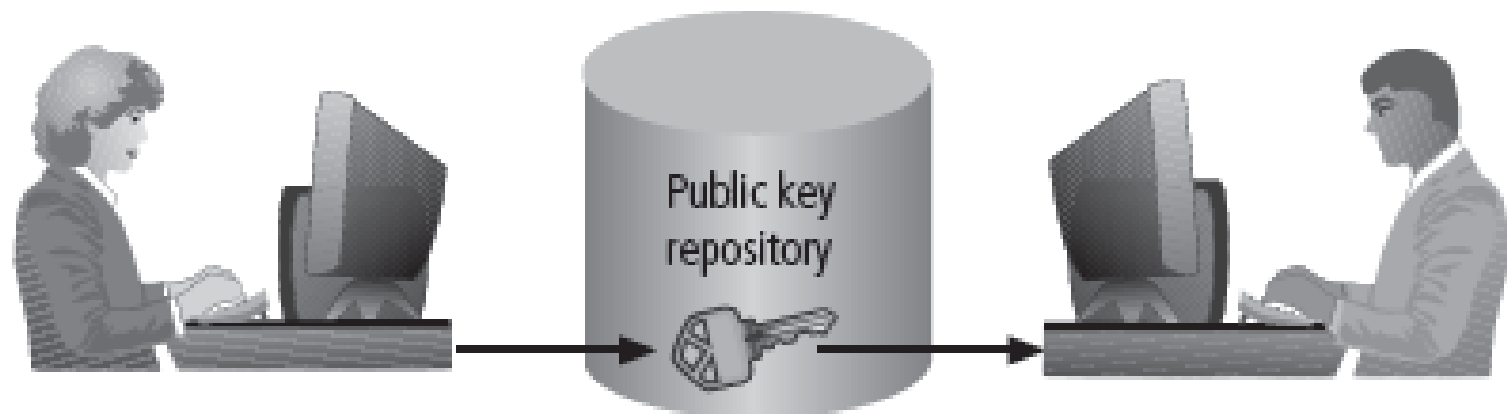
- Electronic document containing key value and identifying information about entity that controls key.
- A *Digital Certificate* is an electronic "password" that **allows** a person, organization to exchange data securely over the Internet using the public key infrastructure.
 - an electronic document, **similar to a digital signature, attached to a file** certifying that this file is from the organization it claims to be from and has not been modified from the originating format.
- **Certificate authority:**
 - is an agency that manages the issuance of certificates and verify their worth and integrity.
- Digital signature attached to certificate's container file to certify file is from entity it claims to be from.

Version
Certificate Serial Number
Algorithm ID • Algorithm ID • Parameters
Issuer Name
Validity • Not Before • Not After
Subject Name
Subject Public Key Info • Public Key Algorithm • Parameters • Subject Public Key
Issuer Unique Identifier (Optional)
Subject Unique Identifier (Optional)
Extensions (Optional) • Type • Criticality • Value
Certificate Signature Algorithm
Certificate Signature

Table 8-8 X.509 v3 Certificate Structure¹¹

Hybrid Cryptography Systems

- Except with digital certificates, **pure asymmetric key encryption not widely used**.
- Asymmetric encryption more often used with symmetric key encryption, creating hybrid system.
- **Diffie-Hellman Key Exchange** method:
 - is a **way to exchange keys** without exposure to any third parties using public key encryption:
 - **asymmetric encryption** is used as a method to exchange **symmetric keys**.



Rachel at ABC Corp. stores her public key where it can be accessed. Alex at XYZ Corp. retrieves it and uses it to encrypt his session (symmetric) key. He sends it to Rachel, who decrypts Alex's session key with her private key, and then uses Alex's session key for short-term private communications.

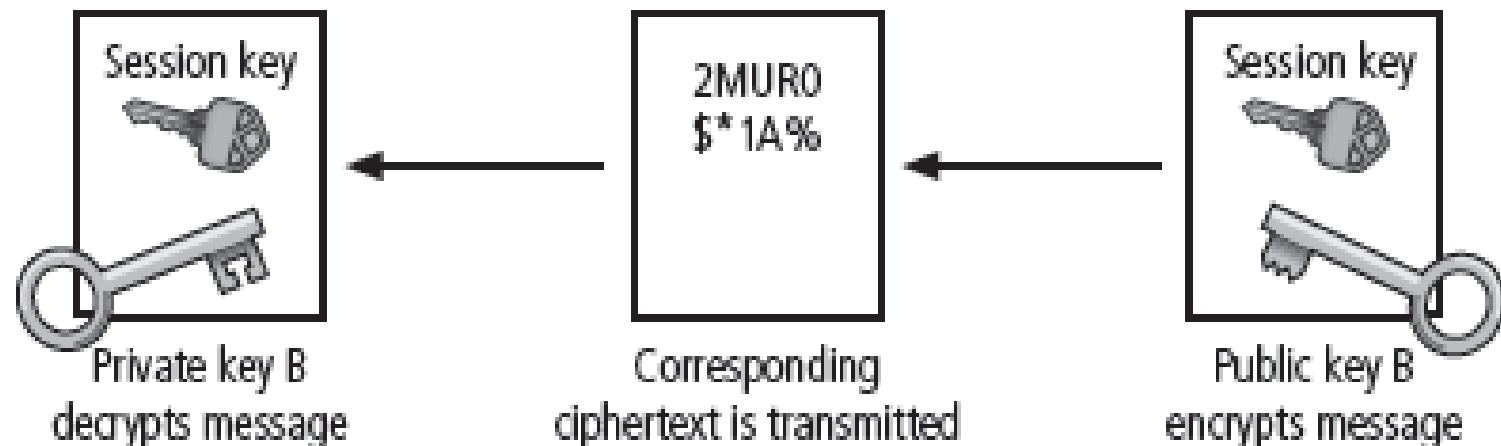


Figure 8-9 Example of Hybrid Encryption

Steganography

- Process of hiding information (steganos meaning “covered” and graphein meaning “to write.”)
- Has been in use for a long time
- Most popular modern version hides information within files appearing to contain digital pictures or other images.
- Some applications hide messages in .bmp, .wav, .mp3, and .au files, as well as in unused space on CDs and DVDs.
- Ex:
 - computer graphics standards use a combination of three color values (red, blue, and green (RGB)) to represent a picture element, or pixel.
 - Each of the three color values requires 8-bit code for color’s intensity (e.g., 00000000 for no red and 11111111 for maximum red).
 - This inability to perceive difference on the part of humans provides the steganographer with one bit per color to use for encoding data into an image file.

Cryptanalysis

- objective to **recover key** not just message
- general approaches:
 - **cryptanalytic attack**:
 - rely on the nature of the algorithm plus perhaps some knowledge of the **general characteristics** of the plaintext or even some sample plaintext-ciphertext pairs
 - **brute-force attack**:
 - try every possible key on a piece of ciphertext until an intelligible translation into plaintext is obtained.
 - On average, half of all possible keys must be tried to achieve success.

Brute Force Search

- always possible to simply try every key
- most basic attack, proportional to key size
- assume either know / recognise plaintext

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