INTRODUCTION TO CRYPTOGRAPHY

Kushal Dhakal | CRYPTOGRAPHY | November 10, 2074

Foreword:

Cryptology is a fascinating discipline at the intersection of computer science,

mathematics and electrical engineering. As cryptology is moving fast, it is hard to keep up with all the developments. During the last 25 years, the theoretical foundations of the area have been strengthened; we now have a solid understanding of security definitions and of ways to prove constructions secure. Also in the area of applied cryptography we witness very fast developments: old algorithms are broken and withdrawn and new algorithms and protocols emerge.

Encryption and Decryption

Encryption is the process of encoding a message so that its meaning is not obvious i.e. converting information from one form to some other unreadable form using some algorithm called ***cipher*** with the help of secret message called ***key***. The converting text is called is ***plaintext*** and the converted text is called ***ciphertext***.

Decryption is the reverse process, transforming an encrypted message back into its normal, original form. In decryption process also the use of key is important.

Alternatively, the terms ***encode*** and ***decode*** or ***encipher*** and ***decipher*** are used instead of ***encrypt*** and ***decrypt***. That is, we say that we encode, encrypt, or encipher the original message to hide its meaning. Then, we decode, decrypt, or decipher it to reveal the original message.

The use of encryption techniques is being used since very long period as it can be noted from the technique called ***Caesar's cipher*** used by Julius Caesar for information passing to his soldiers. Encryption techniques have also been extensively used in military purposes to conceal the information from the enemy. Nowadays to gain the confidentiality encryption is being used in many areas like communication, internet banking, digital right management, etc.

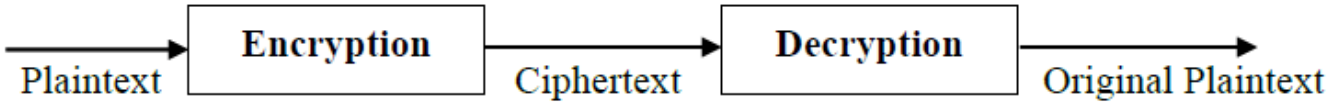

Fig: *Encryption-Decryption*

Key

A ***key*** is a parameter or a piece of information used to determine the output of cryptographic algorithm. While doing the encryption, key determines the transformation of plaintext to the cipher text and vice versa. Keys are also used in other cryptographic processes like message authentication codes and digital signatures. Most of the cryptographic systems depend upon the key and thus the secrecy of the key is very important and is one of the difficult problems in practice. Another important issue for the key is its length. Since key is the sole entity that defines the strength of the security (normally algorithm used is public) we need to select the key in a way such that attacker should take long enough to try all possibilities. To prevent the key from being guessed the choice of the key must be random.

Cipher

A ***cipher*** is an algorithm for performing encryption and decryption. The operation of cipher depends upon the special information called key. Without knowledge of the key, it should be difficult, if not nearly impossible, to decrypt the resulting cipher into readable plaintext. There are many types of encryption techniques that have advanced from history, however the distinction of encryption technique can be broadly categorized in terms of number of key used and way of converting plaintext to the ciphertext.

Cryptosystem

Cryptosystem is a 5-tuple/quintuple (E, D, M, K, C), where M set of plaintexts, K set of keys, C set of ciphertexts, E set of encryption functions e: M * K -> C and D set of decryption functions d: C * K -> M.

Example: *Caesar Cipher*

M = {sequences of letters}

K = {i | i is an integer and $0 \leq i \leq 25$}

E = { E_k | k ε K and for all letters m, $E_k(m) = (m + k) \bmod 26$ }

D = {D_k | k ε K and for all letters c, $D_k(c) = (26 + c – k) \bmod 26$}

C = M

Cryptographic system characteristics :-

Cryptographic systems are characterized along three independent dimensions:

1). The type of operations used for transforming plaintext to ciphertext. All encryption algorithms are based on two general principles: substitution, in which each element in the plaintext (bit, letter, group of bits or letters) is mapped into another element, and transposition, in which elements in the plaintext are rearranged. The fundamental requirement is that no information be lost (that is, that all operations are reversible). Most systems, referred to as product systems, involve multiple stages of substitutions and transpositions.

2).The number of keys used. If both sender and receiver use the same key, the system is referred to as symmetric, single-key, secret-key, or conventional encryption. If the sender and receiver use different keys, the system is referred to as asymmetric, two-key, or public-key encryption.

3).The way in which the plaintext is processed. A block cipher processes the input one block of elements at a time, producing an output block for each input block. A stream cipher processes the input elements continuously, producing output one element at a time, as it goes along

Introduction:

Where Does Cryptography Stands ?

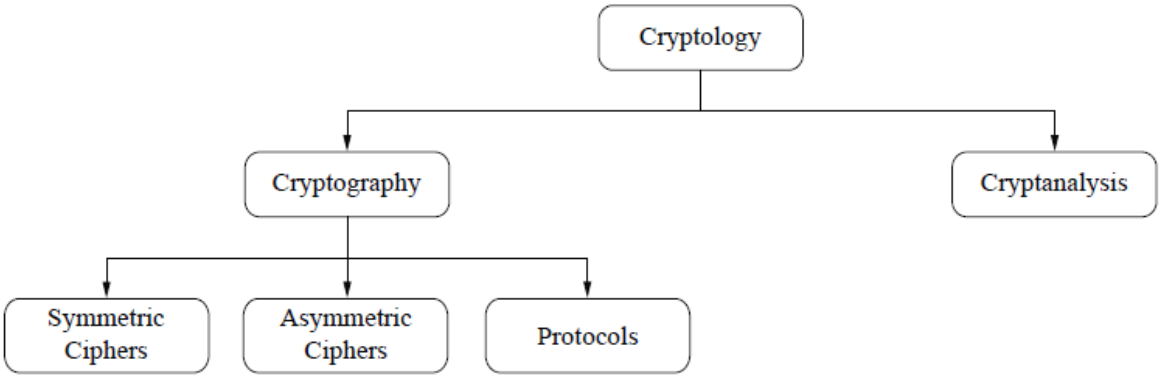

Fig. 1.3 Overview of the field of cryptology

Let's now have a look at the field of *cryptography* (Fig. 1.3). The first thing that we notice is that the most general term is *cryptology* and not *cryptography*. Cryptology splits into two main branches:

Cryptography is the science of secret writing with the goal of hiding the meaning

of a message.

Cryptanalysis is the science and sometimes art of *breaking* cryptosystems. You might think that code breaking is for the intelligence community or perhaps organized crime, and should not be included in a serious classification of a scientific discipline. However, most cryptanalysis is done by respectable researchers in academia nowadays. Cryptanalysis is of central importance for modern cryptosystems: without people who try to break our crypto methods, we will never know whether they are really secure or not.

Because cryptanalysis is the only way to assure that a cryptosystem is secure, it is an integral part of cryptology. Nevertheless, the focus of this CSIT Syllabus is on crypto**graphy**: We introduce most important practical crypto algorithms in detail. These are all crypto algorithms that have withstood cryptanalysis for a long time, in most cases for several decades. In the case of crypt**analysis** we will mainly restrict ourselves to providing state-of-the-art results with respect to breaking the crypto algorithms that are introduced, e.g., the factoring record for breaking the RSA scheme .

Let's now go back to Fig. 1.3. Cryptography itself splits into three main branches:

Symmetric Algorithms are what many people assume cryptography is about: two parties have an encryption and decryption method for which they share a secret key. All cryptography from ancient times until 1976 was exclusively based on symmetric methods. Symmetric ciphers are still in widespread use, especially for data encryption and integrity check of messages.

Asymmetric (or Public-Key) Algorithms In 1976 an entirely different type of cipher was introduced by Whitfield Diffie, Martin Hellman and Ralph Merkle. In public-key cryptography, a user possesses a secret key as in symmetric cryptography but also a public key. Asymmetric algorithms can be used for applications such as digital signatures and key establishment, and also for classical data encryption.

Cryptographic Protocols Roughly speaking, crypto protocols deal with the application

of cryptographic algorithms. Symmetric and asymmetric algorithm

The main focus of this course is on symmetric and asymmetric algorithms, as well as hash functions. However, we will also introduce basic security protocols. In particular, we will introduce several key establishment protocols and what can be achieved with crypto protocols:

1. confidentiality of data

2.integrity of data

3.authentication of data

4.user identification etc.

Symmetric cryptography :

Symmetric cryptographic schemes are also referred to as *symmetric-key*, *secret-key*, and *single-key* schemes or algorithms. Symmetric cryptography is best introduced with an easy to understand problem: There are two users, Alice and Bob, who want to communicate over an insecure *channel* The term channel might sound a bit abstract but it is just a general term for the communication link: This can be the Internet, a stretch of air in the case of mobile phones or wireless LAN communication, or any other communication media you can think of. The actual problem starts with the bad guy, Oscar1, who has access to the channel, for instance, by hacking into an Internet router or by listening to the radio signals of a Wi-Fi communication.

This type of unauthorized listening is called *eavesdropping*.

Classical Cryptosystem

Historical pen and paper ciphers used in the past are sometimes known as classical ciphers. These are the very old or quite old cryptosystem that were used in pre computer age.these crypto system are too weak now days and can be broken easily with computer.But we even studied these cryptosystem because they illustrate basic of the concepts of cryptography.

Substitution Cipher

Shift Cipher (or Caesar Cipher)

We now introduce another historical cipher, the *shift cipher*. It is actually a special case of the substitution cipher and has a very elegant mathematical description. The shift cipher itself is extremely simple: We simply shift every plaintext letter by a fixed number of positions in the alphabet. For instance, if we shift by 3 positions, A would be substituted by d, B by e, etc. The only problem arises towards the end of the alphabet: what should we do with X, Y, Z? As you might have guessed, they should "wrap around". That means X should become a, Y should become b, and Z is replaced by c. Allegedly, Julius Caesar used this cipher with a three-position shift. The shift cipher also has an elegant description using modular arithmetic. For the mathematical statement of the cipher, the letters of the alphabet are encoded as numbers, as depicted in Table 1.3.

Table 1.3 Encoding of letters for the shift cipher

A	B	C	D	E	F	G	H	I	J	K	L	M
0	1	2	3	4	5	6	7	8	9	10	11	12
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
13	14	15	16	17	18	19	20	21	22	23	24	25

Both the plaintext letters and the ciphertext letters are now elements of the ring Z26. Also, the key, i.e., the number of shift positions, is also in Z26 since more than 26 shifts would not make sense (27 shifts would be the same as 1 shift, etc.). The encryption and decryption of the shift cipher follows now as:

Definition 1.4.3 Shift Cipher
Let $x, y, k \in \mathbb{Z}_{26}$.
Encryption: $e_k(x) \equiv x + k \bmod 26$.
Decryption: $d_k(y) \equiv y - k \bmod 26$.

Example 1.11. Let the key be $k = 17$, and the plaintext is:

$$\texttt{ATTACK} = x_1, x_2, \ldots, x_6 = 0, 19, 19, 0, 2, 10.$$

The ciphertext is then computed as

$$y_1, y_2, \ldots, y_6 = 17, 10, 10, 17, 19, 1 = \texttt{rkkrtb}$$

In the above Definition 1.4.3 , we could also write as :- $d_k(y)= (y+26-k) \bmod 26$

As you can guess from the discussion of the substitution cipher earlier in this book, the shift cipher is not secure at all. There are two ways of attacking it:

1. Since there are only 26 different keys (shift positions), one can easily launch a

brute-force attack by trying to decrypt a given ciphertext with all possible 26

keys. If the resulting plaintext is readable text, you have found the key.

2. As for the substitution cipher, one can also use letter frequency analysis.

Attacking the Cipher

Caesar Cipher is quite easily broken even with ciphertext only. One can attack the cipher text using exhaustive search by trying all possible keys until you find the right one. Exhaustive search is best suited if the key space is small and we have only 26 possible keys in Caesar cipher. Another approach of attacking the cipher is statistical analysis where we compare the ciphertext to 1-gram model of English.

Problem with caesar cipher

The main problem with Caesar's Cipher is that the key is too short and can be found by exhaustive search. Again statistical frequencies not concealed well i.e. they look too much like regular English letters. So the solution can be to increase the key length (can be done using multiple letters in key) so that cryptanalysis gets harder.

Transposition Cipher

In transposition ciphers the letters are systematically arranged so that the actual position of letters is gets changed making the text garble.

2. Rail-Fence Cipher

The Rail Fence Cipher is a form of transposition cipher that derives its name from the way in which it is encoded. In the rail fence cipher, the plaintext is written downwards and diagonally on successive "rails" of an imaginary fence, then moving up when we each the bottom rail. When we reach the top rail, the message is written downwards again until the whole plaintext is written out. The message is then read off in rows.

For example, using 3 "rails" and a message of 'WE ARE DISCOVERED FLEE AT ONCE', the cipherer writes out:

W . . . E . . . C . . . R . . . L . . . T . . . E

. E . R . D . S . O . E . E . F . E . A . O . C .

. . A . . . I . . . V . . . D . . . E . . . N . .

Then reads off:

WECRL TEERD SOEEF EAOCA IVDEN

Similarly, if we have 3 "rails" and a message of THIS IS THE PLAINTEXT, the cipherer writes out (we are not showing diagonal move here just write in down rail a step ahead):

T S T P I E

H I H L N X

I S E A T T

The ciphertext is T S T P I E H I H L N X I S E A T T

The problem with Rail Fence Cipher is that the rail fence cipher is not very strong; the number of practical keys is small enough that a cryptanalyst can try them all by hand. To decrypt we get the number of letters to be skipped. For this if the number of rail is n key is so in our e.g. n = 3 and

key is 18/3 = 6 i.e. skip 6 letters from the letter you are reading every time to get plaintext (remember to go circular that is if count ends continue from the starting letter leaving the read letter). See below :

T	S	T	P	I	E	H	I	H	L	N	X	I	S	E	A	T	T
1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6

We, selected letter with index 1 THI. Now choose the letter with index 2, see below

T	S	T	P	I	E	H	I	H	L	N	X	I	S	E	A	T	T
1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6

Continue like this until you read off all the characters.

3. Vigenere Cipher: Substitution Cipher (Polyalphateic)

It is like Caesar cipher, but uses a phrase for e.g. for the message THE BOY HAS THE BALL and the key VIG, encipher using Caesar cipher for each letter:

key VIGVIGVIGVIGVIGV

plain THEBOYHASTHEBALL

cipher OPKWWECIYOPKWIRG

Here, generally, we repeatedly write key above the plaintext and use the Caesar cipher for each letter in the plaintext where key for each letter being processed is taken from the repeated key letter just above it. This process is simplified by using the table as below called Tableau

Key

Plaintext

	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

Fig: Vigenere Tableau

Period: length of key. In example above it is 3.

Tableau: Table used to encipher and decipher. In tableau Vigènere cipher has key letters on top, plaintext letters on the left or vice versa. It is also possible to have key on top (left) plaintexts in middle and ciphertexts in left (top).

Assuming key on top and the plaintext on left, Decryption is performed by finding the position of the ciphertext letter in a column, corresponding to the key letter, of the table, and then taking the label of the row in which it appears as the plaintext letter. For example, in column V (key letter), the ciphertext letter O appears in row T, which taken as the first plaintext letter. The second letter is decrypted by looking up P in column I of the table; it appears in row H, which is taken as the plaintext letter. This process continues until we find the plaintext letters for all the ciphertext letters

Malicious Code

Perhaps the most sophisticated types of threats to computer systems are presented by programs that exploit vulnerabilities in computing systems. Such threats are referred to as **malicious software**, or **malware**. In this context, we are concerned with threats to application programs as well as utility programs, such as editors and compilers, and kernel-level programs.

Example: The following UNIX script is named ls and is placed in a directory.

```
cp /bin/sh /tmp/.xxsh
chmod o+s,w+x /tmp/.xxsh
rm ./ls
ls $*
```

It creates a copy of the UNIX shell that is setuid to the user executing this program. This program is deleted, and then the correct ls command is executed. On most systems, it is against policy to trick someone into creating a shell that is setuid to themselves. If someone is tricked into executing this script, a violation of the (implicit) security policy occurs. This script is an example of malicious logic.

Malicious code refers to a broad category of software threats to our network and systems. Perhaps the most sophisticated types of threats to computer systems are presented by malicious codes that exploit vulnerabilities in computer systems. Any code which *modifies or destroys data, steals data , allows unauthorized access, exploits or damage a system, and does something that user did not intend to do*, is called malicious code. There are various types of malicious code we will encounter, including Viruses, Trojan horses, Logic bombs, and Worms.

A computer program is a sequence of symbols that are caused to achieve a desired functionality; the program is termed malicious when their sequences of instructions are used to intentionally cause adverse affects to the system. In the other words we can't call any "bug" as a Malicious Code. Malicious codes are also called programmed threats. The following figure provides an overall taxonomy of Malicious Code.

Types of malicious code

The terminology in this area presents problems because of a lack of universal agreement on all of the terms and because some of the categories overlap. Table 10.1 is a useful guide.
Malicious software can be divided into two categories: those that need a host program, and those that are independent. The former, referred to as **parasitic**, are essentially fragments of programs that cannot exist independently of some actual application program, utility, or system program. Viruses, logic bombs and backdoors are examples. Independent malware is a self-contained program that can be scheduled and run by the operating system.Worms and bot programs are examples.

We can also differentiate between those software threats that do not replicate and those that do. The former are programs or fragments of programs that are activated by a trigger. Examples are logic bombs, backdoors, and bot programs. The latter consist of either a program fragment or an independent program that, when executed, may produce one or more copies of itself to be

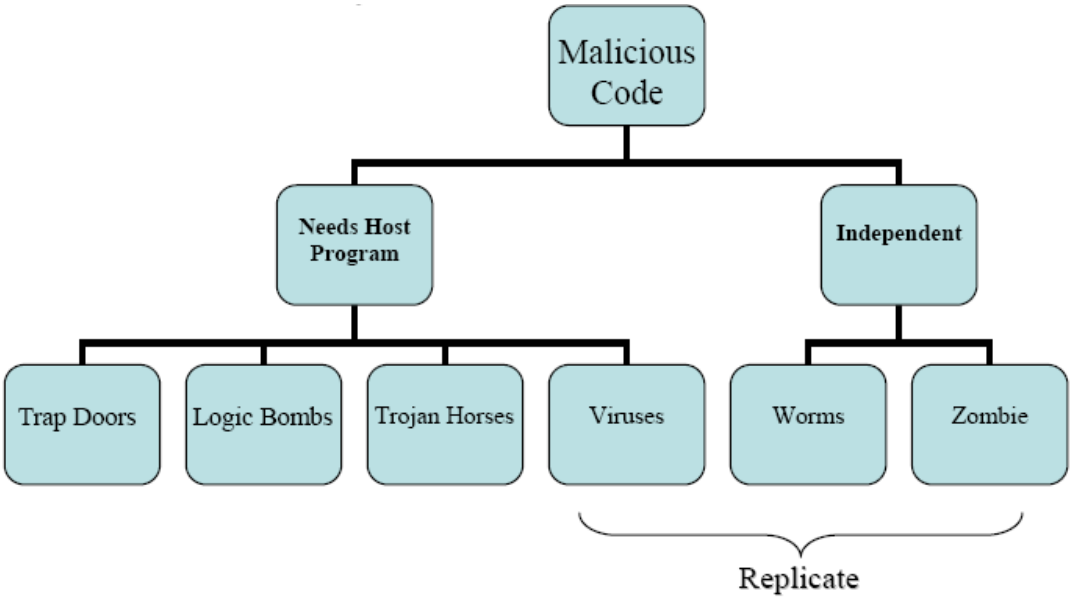

Fig :- Malicious Code and It's Types

Name	Description
Virus	Malware that, when executed, tries to replicate itself into other executable code; when it succeeds the code is said to be infected. When the infected code is executed, the virus also executes.
Worm	A computer program that can run independently and can propagate a complete working version of itself onto other hosts on a network.
Logic bomb	A program inserted into software by an intruder. A logic bomb lies dormant until a predefined condition is met; the program then triggers an unauthorized act.
Trojan horse	A computer program that appears to have a useful function, but also has a hidden and potentially malicious function that evades security mechanisms, sometimes by exploiting legitimate authorizations of a system entity that invokes the Trojan horse program.
Backdoor (trapdoor)	Any mechanism that bypasses a normal security check; it may allow unauthorized access to functionality.
Mobile code	Software (e.g., script, macro, or other portable instruction) that can be shipped unchanged to a heterogeneous collection of platforms and execute with identical semantics.

Fig :- Descriptive version of these terms .