Data Security and Information Technology

Faez H. Ali[1], Wadi. S. Mahdi[2], Adnan M. Ali[3], Emad B. Abdulkareem[1]

[1] Mathematics Department, College of Science, Mustansiriyah University, Iraq.
[2] Prime Minister Office, Baghdad, Iraq.
[3] Computer Technology Engineering Department, Baghdad College of Economic Sciences University, Iraq.

***Abstract*— In this paper, a general study is introduced about using the information technology in data security. The most famous technologies of data protection, the cryptography and information hiding are introduced. In addition, we describe the cryptanalysis and Steganalysis methods to explain how we can attack data, especially the transmitted data.**

***Keywords*— Cryptography, Cryptanalysis, Stream Cipher, Block Cipher, Genetic Algorithm, Neural Network, Information Hiding, Steganography, Steganalysis, Watermarking.**

I. INTRODUCTION

A message is expressed in plaintext (PT) (cleartext). Encryption is a manner of concealing the contents of a message. Ciphertext (CT) is an encrypted message. Decryption is the procedure of transforming CT back to PT. A digital currencies algorithm, commonly referred to as a cipher, is the mathematical function that allows for encryption and decryption. An effort to use cryptanalyze is referred to as an attack. a theory put forward by Kerckhoffs in the 19th century, secrecy must be entirely encapsulated within the what we called as a key [1].

Swarm Intelligence (SI) is an Artificial Intelligence (AI) technique that studies the collective behavior of a decentralized system composed of a population of simple agents that interact locally with one another and with their surroundings (Xiaodong 2004) [2]. AI is used to investigate distributed problem solving without a centralized control structure. This is considered a better alternative to centralized, rigid, and preprogrammed control. Real-life SI can be found in ant colonies, beehives, bird flocks, and animal herds. The most common SI systems are Ant Colony Optimization (ACO), Genetic Algorithm (GA), Particle Swarm Optimization (PSO), and Marriage in Honey Bees Optimization (HBO) [3].

Neural networks (NN's) are information processing systems that were created as generalization algorithms of human cognition in neural biology. We focus on demonstrating the application of SI,AI,and NN concepts in cryptanalysis [4]. Steganography is the practice of obscuring an existence of correspondence by incorporating an encrypted transmission into an innocuous-looking conceal paperwork, such as a digital image, videos, sound files, or other computer files containing intuitively unnecessary or unnecessary data, in order to serve as covers or carriers for private messages [5].

II. CRYPTOGRAPHY

Cryptography is the study of the rules or we can say some procedures that allow the data or important or classified information to be transform into what we called a ciphertext and then its transmit in public or secure channel and demonstrated by illegal authorized individuals or person using the private key, but making it difficult or computationally impractical for someone who is not authorized to do so. Cryptanalysis is the study (and artwork) of collecting details may be just depending on ciphertexts without knowing the secret key. Cryptography and cryptanalysis terms are belonged to Cryptology. Cryptography is concerned with encryption and decryption procedures [6]. A number algorithms for encryption use keys, so the ciphertext message is established based on the initial plaintext and the key value. Figure (1) illustrates that the encryption and decryption keys are sometimes the same

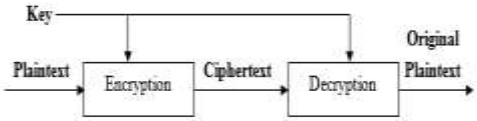

Figure (1) Single-Key Cryptosystem

Sometimes decryption and encryption keys come are reverse to each other's. This case shown in figure (2).

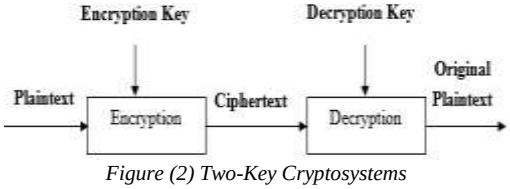

Figure (2) Two-Key Cryptosystems

A key enables multiple encryptions of a single plaintext message simply through modifying the key. The use of a key enhances security. If the interceptor gains access to the encryption algorithm, future messages will remain private due

DOI:http://doi.org/10.24086/cocos2024/paper.1484

to the interceptor person won't have access to the key value. A type of cipher that does not need a key is known as a Keyless Cipher.

III. .CRYPTOGRAPHIC SYSTEMS (CRYPTOSYSTEM)

the systems (that can be constructed software or hardware) which are depending on two processes; the encryption and decryption, is called Cryptosystem which can be classified as in figure (3) [6].

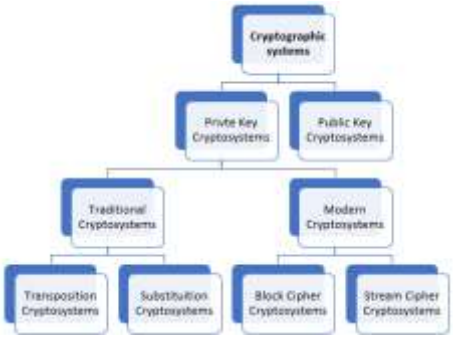

Figure (3) Classification of Cryptosystems

A. *Cryptosystem using Private Key*

It's also known as symmetric cryptosystems. A traditional private-key system of cryptography as shown in figure (4), uses the same key ($e_k = d_k = k \in K$) for both cryptography processors.

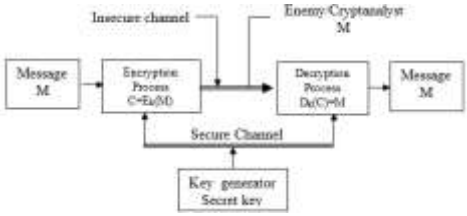

Figure (4) Conventional secret-key cryptosystems $e_k = d_k$.

The sender of the message applies a reversible alteration to generate the ciphertext: $C = (E_k(m)), m \in M$ and $c \in C$, which then gets sent to the receiver via a widely available uncertain channel. The private key (k) (which we called it sometimes Basic key (BK)) must be sent to the valid recipient for decoding, but through a channel that is safe. Because the valid recipient recognizes the key k, he may decode c using alteration: $D_k(c) = D_k(E_k(m)) = m, c \in C$ and $m \in M$, representing the original plaintext message.

Secret key cryptosystems are classified into several types, including monoalphabetic ciphertexts, polyalphabetic ciphertexts, stream (of bits) ciphers and block (set of characters) ciphers [7,8].

B. *Classical Cryptosystems*

The term classical ciphers refers to encryption techniques which have become well-known over time, and generally created prior to the second half of the twentieth century (in some cases, many hundreds of years earlier). Many classical techniques are variations of simple substitution and simple transposition.

1. *Substitution Cipher Systems*

First, we'll imagine that the plaintext is in uppercase and the ciphertext is in lowercase. In monoalphabetic substitutions, the order of letters is permuted or stumbled, with each plaintext letter mapping to a distinct ciphertext letter. A permutation is a reorganizing of elements in a series. In a monoalphabetic substitution, each $c_i = a_\pi(p_i)$, where $a_1, a_2, \dots, a_k$ are the characters of the plaintext alphabet and π is a permutation of the numbers $1,2,\dots,k$.

For example, $\pi(a)$ could be the function $\pi(a) = 25 - a$, where A can be expressed as z, B as y, and Z as a. This permutation is simple to write down from memory and could be used in the field [9].

Let Z denote the alphabet, and K denote the keys, which are all possible variations of Z symbols. For each permutation $\pi \in K$, define the encryption function $E_\pi(x) = \pi(x)$ and the decryption function $D_\pi(y) = \pi^{-1}(y)$, where π^{-1} is the inverse permutation of π.

For example, define Z as the 26-letter English alphabet. A random permutation could result in:

This determines the encryption function, such that E_π (A)=d,E_π (B)= l, and so on. Switching the rows yields the inverse permutation D_π (y), which is used for decryption.

A	B	C	D	E	F	G	H	I	J	K	L	M
d	k	a	c	i	o	m	e	b	g	f	j	q
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
z	x	w	p	t	n	H	u	s	l	r	y	v

A polyalphabetic cipher uses several cipher alphabets. For instance, in order to help with encryption, all of the characters are usually set out in a large table called a tableau. A characteristic tableau size is 26×26, which accommodates 26 full CT characters. The method for filling the tableau and deciding which language to use next establishes that particular polyalphabetic cipher. All of these ciphers are more straightforward to decode than initially thought, because substitution elements are used for sufficiently large PT.

Blaise de Vigenère's work was among the most popular. It was first published in 1585 and was thought to be unbreakable until 1863. In this system, each initial line of the tableau is a replicate of the PT, and afterward segments are purely changed one spot to the left. (A simple tableau known as tubule recta corresponds mathematically to adding the key and PT modulo 26). A keyword is then used to choose the ciphertext. Every letter in the expression is used in consequence, followed by another use of the beginning. For instance, if the key statement is 'BOY', the first letter of PT is encoded as 'B', the second as 'O', the third as 'Y', the fourth as 'B', and so on [10].

There exists an optimal and unbreakable encryption system. Mauborgne [1] created the one-time pad (OTP) in 1917. An OTP is basically a large, nonrepeating set of sincerely random key letters written on layers of paper and bound together to make a pad. In the beginning, it was a one-time tape for the sender. The person sending the message encodes exactly one PT character with each key note on the pad. Module 26 combines the PT and OTP character sets to form encryption.

Each key character appears just once in a single PT. The sender encodes the PT before discarding the pad's used parts or tape section. The other party has a comparable pad and decrypts

DOI:http://doi.org/10.24086/cocos2024/paper.1484

every character of the ciphertext using the keys on the pad in turn. Following decryption of the message, the other party eliminates the same pad pages or tape sections. New message, new key letters.

2. Transposition Cryptosystems [3]

A transposition is a method of encryption whereby the characters of a message are rearranged. The goal of a transposition is diffusion, which involves spreading data from the transmitted data or key throughout the ciphertext. Transpositions attempt to disrupt established patterns. A permutation is another term for a transposition, which is a restructuring of a message's symbols.

Columnar transposition is the rearrangement of the plaintext's characters into columns. The example below shows a five-column transposition. Plaintext characters are divided into segments of five and organized a single line after a different one as shown here.

c1	c2	c3	c4	c5
c6	c7	c8	c9	c10
c11	c12	etc.		

The obtained CT is reformed by changing the columns: c1c6c11….c2c7c12…..c3c8, etc.

As an example, let's have the following PT as:

T	H	I	S	I
S	A	M	E	S
S	A	G	E	T
O	S	H	O	W
H	O	W	A	C
O	L	U	M	N
A	R	T	R	A
N	S	P	O	S
I	T	I	O	N
W	O	R	K	S

The resulting CT is as follows:

tssoh oaniw haaso lrsto imghw
utpir seeoa mrook istwc nasns

C. Modern Cryptosystems

Now let's defined the following important notations:

- Key space (KS): a set of all possible or probability of keys with respect to the language, the KS space includes two important types of keys: encryption key (e_k) and the decryption key (d_k).
- The Encryption Function (algorithm) EF: $E_{e_k}(M) = C$.
- The Decryption Function (algorithm) DF: $D_{d_k}(C) = M$.

The Two main Functions EF and DF must be invertible functions such that**:** $D_{d_k}(C) = D_{d_k}(E_{e_k}(M)) = M$.

There are basically two various kinds of cryptosystems, which are as follows [7].

1. Stream Cipher (SC) Cryptosystems

In SCs, the PT units are represented by less values are: bits. The key of the SC is usual obtained from a pseudorandom keygenerator of bits (see figure (5)). The PT is encrypted on a bit-by-bit basis.

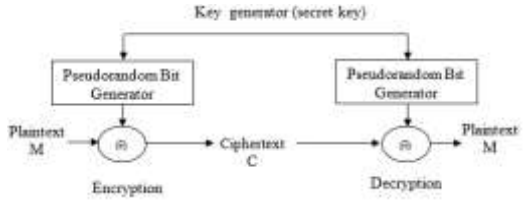

Figure (5) Stream cipher cryptosystem.

The key is fed into a random bit generator, which produces an extended series of binary data. This "key-stream" k is then mixed with plaintext m, typically using a bit wise XOR (Exclusive-OR modulo 2 addition), to produce the ciphertext stream, which uses the same random bit generator and seed.

A feedback shift register (FSR) consists of two components: a feedback function (FF) and shift register (SR). The SR is a series of bits. Every time (t) we need just one bit of the initial of SR, all the bits in the FSR are shifted just one step or we can say one bit to the next cell of FSR. The most basic and common type of FSR is a Linear Feedback Shift Register (LFSR), as illustrated in details of the cells in figure (6).

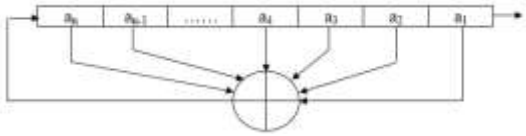

Figure (6) Block Diagram of LFSR.

The FF is simply depend on the adding function (XOR) of specific bits in the SR. Because of the simple FF, a wide range of mathematical formulation, theory with important theorems in algebraic theory and can be used to analyze $LFSRs$. $LFSRs$ are probably the most prevalent form of SR found in cryptography.

The fundamental strategy to building a stream of a keygenerator using $LFSR$ is straightforward. First, you take one or more $LFSRs$, usually of various lengths and FF (If the sizes of LFSR and their $2^L - 1$ are all comparatively prime (Mersenne prime) and the FF are all primitive, then the designed keygenerator has maximal period). Every $LFSR's$ must has an initial state in order to be moved to generates key sequence. Move the $LFSR$ once, the result of a bit is a function, this function ideally must be nonlinear to obtain high linear complexity for the keygenerator. This function is called the combining function (CF) [11,12]. From modern SCC, Ali et. al. introduce a new efficient cryptosystem to protect the image files [34].

2. Block Cipher (BC) Cryptosystems

A BC is a function that converts e-bit plaintext blocks of data to e-bit ciphertext blocks, where n is the blocklength. It could be seen as a simple substitution cipher (SSC) with a large character size. The function is defined by a k-bit key (K), which

DOI:http://doi.org/10.24086/cocos2024/paper.1484

takes information from a portion K (the key space) of the total number of all k-bit transmission vectors V_k. It is generally believed that the key will be selected at random. Data expansion can be avoided by using equal-sized plaintext and ciphertext blocks.

In 1973, the Data Encryption Standard (DES) became the largest and most commonly employed cryptosystem in the world. DES was created at IBM. The Federal Information Processing Standards Publication 46, published in 1977, provides a complete description of DES. DES encodes a normal text bitstring x of dimension 64 with a K of length 56, resulting in a ciphertext bitstring of length 64. We first provide a "high-level" overview of the cryptosystem [10].

B. Public Key (PK) Cryptosystem [10]

The PK Cryptosystem also called asymmetric cryptosystems. In this cryptosystem (see figure (7)), decryption key d_k and the encryption key e_k are different, where $e_k \neq d_k$.

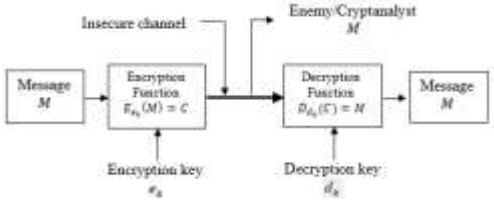

Figure (7) PK cryptosystem $e_k \neq d_k$.

C. Chaos Theory

Chaos theory is a field of study mathematics, with applications in several disciplines including physics, economics, biology, and philosophy. Chaos theory studies the behavior of dynamical systems that are highly sensitive to initial conditions, an effect which is popularly referred to as the butterfly effect. Small differences in initial conditions yield widely diverging outcomes for chaotic systems, rendering long-term prediction impossible in general.

One-dimensional and higher-dimensional chaotic maps are the two categories into which chaotic maps fall. One variable is used throughout distinct time steps in 1-D chaotic maps, which include logistic maps, tent maps, and sine maps. These chaotic maps are simple to use, appear to have structure, and are minimal in complexity [33].

Higher dimension chaotic maps have at least two variables, better performance, and their chaotic orbits are more difficult to predict and they have high computing costs and are difficult to implement in hardware, implying that they are not real-time processing [31]. For example, we will look at the following chaotic maps:

A 1-dimention logistic map with three parameters [32], this logistic map is as follows:

$$x_{n+1} = ax_n(\alpha - x_n)^{\beta} \quad \ldots(1)$$

Where $a = 1.5, \alpha = 3, \beta = 0.5$, and $0 < x_0 < 3$, $n = 0,1,2, \ldots$

D. Multiplicative Cyclic Group (MCG)

This system firstly introduced by Ali, (2006) [29] which is referred to the revolving multiplication group for that, and by taking advantage of his research. MCG unit which generates the sequence S is a function of five independent variables s.t. $S = MCGU(q, \alpha_t, \alpha_r, \gamma, m)$, where $1 \leq \gamma \leq q-1$ is an arbitrarily chosen start point. Before we introduce the MCGU algorithm we have to introduce the following notations:

- q is prime number.
- $\alpha_t \neq \alpha_r$ are two different generators.
- m is the number system where $m = 2$ for binary sequences.
- γ is the start point where $1 \leq \gamma \leq q-1$.
- L is the length of the sequences want to be generated.

The $MCGU$ algorithm steps can be seen in [29].

The highly efficient sequences were generated through the developed generator, we can discuss the following cases [30]:

- Linear Complexity (LC): The high non-linearity $MCGU$ proved that the LC is in secure zone.
- Periodicity ($P(S)$): Given that S has a period of $q-1$ for every two distinct generator elements and that we have $P_2^{g(q)}$ generators, the $MCGU's$ period is $P(S) = P_2^{g(q)} * (q-1)$.
- Randomness ($R(S)$): The randomization of $MCGU$ is proved in [1], so any keygenerator depends on this unit will has good randomness properties.
- General Complexity ($GC(S)$): Let T be the number of primes q in some range or in available data base, we have $g(q)$ ways to choose α to construct $A(q)$, $\mathrm{P}_2^{\mathrm{g(q)}}$ ways to select two various generators in the set $A(q)$, and $q-1$ ways to select γ so we can calculate the GC as follows: $GC(S) = T * \phi(q-1) * \mathrm{P}_2^{\mathrm{g(q)}} * (q-1)$, while the GC of $MCGU$ is $T * \mathrm{P}_2^{\mathrm{g(q)}} * (q-1)$.

IV. Cryptanalysis

The most common assumption is that the cryptanalyst knows every think about the details of the cryptosystem. That is generally main rules of Kerckhoff's. Of course, if the other person is unfamiliar with the cryptosystem in use, his task will be more difficult. However, we do not want to construct a cryptosystem's security on the (perhaps uncertain) assumption that the adversary is unaware of the system in use. The kinds that are most prevalent are explained as follows [9]:

- CT-Only: The opponent (cryptanalyst) has an array of ciphertext.
- Chosen PT: The cryptanalyst has gained limited access to the EF. As a result, he can select a PT and create an equivalent CT.
- Known PT: The cryptanalyst has all or some of the PT and the corresponding CT.

DOI:http://doi.org/10.24086/cocos2024/paper.1484

- Chosen CT: The adversary has gained brief access to the DF. As a result, he can select a CT and create the appropriate PT.

The goal in each case is to identify which key was utilized.

A. Cryptanalysis of Classical Cryptosystems

A cryptanalyst's chore is to Break an encryption; this means that the cryptanalyst will attempt to deduce the meaning of a ciphertext message, or to determine a decrypting algorithm that matches an encrypting algorithm. The analyst can do any or all of three different things:

- Attempt to break a single message.
- Attempt to recognize patterns in encrypted messages, in order to be able to break subsequent ones by applying a straightforward decryption algorithm.
- Attempt to find general weaknesses in an encryption algorithm, without necessarily having intercepted any messages.

An analyst employs messages that have been encrypted, called codes for encryption, captured plaintext, data elements that are suspected or known to be in an encrypted text information, mathematically or statistically significant tools and techniques, language and computer properties, and a great deal of ingenuity and luck. In this section, we will go over the analysis methods for some of the crypto systems mentioned above, including some examples and tables [13].

1. The Distributions of the Characters Frequency

As known the plaintext message which want to be encrypted by some classical cryptosystem, consists of a 27-symbol alphabet-A through Z and the "blank" character or separator between words.

In English there are relatively few small words, such as "AM", "IS", "TO", "BE", "HE", "WE", and "AND", "ARE", "YOU", "SHE", and so on. Therefore, one attack is to substitute known short words at appropriate places in the ciphertext and try substituting for matching characters other places in the ciphertext.

Now we must describe Frequency Distributions (FD); in English, certain characters occur more frequently than other characters. More than on study made in this mannar, we see that the letters E, T, and A are repeated more common than some letters like Z, Q, and J, for example. The piece of writing being analyzed has an impact on the distribution as well.

Table (1) displays the number and relative frequency of characters in some text. These frequencies are very similar to released contributes from other sources.

Table (1) describes the percent frequency of each character in some English text message, where the corresponding frequencies are very comparable to those released from other sources.

Table (1) Letter FD in English.

Letter	Percent	Letter	Percent	Letter	Percent
A	7.49	J	0.27	S	6.95
B	1.29	K	0.47	T	9.85
C	3.45	L	3.57	U	3.00
D	3.62	M	3.39	V	1.16
E	14.00	N	6.74	W	1.69
F	2.18	O	7.37	X	0.28
G	1.74	P	2.43	Y	1.64
H	4.22	Q	0.26	Z	0.04
I	6.65	R	6.14		

2. Index of Coincidence (IC)

The IC is a measure of the variation between frequencies in a distribution, (this description follows the lines of Sinkov [14]). The IC, is an equation to describe the approximation of variance from observed data.

$$IC = \frac{\sum_{i=a}^{z} Freq_i * (Freq_i - 1)}{n * (n-1)}$$

Where $Freq_i$ denotes the observed frequencies of the letter i, and n is the text length.

The index of coincidence ranges from 0.0384, for a polyalphabetic substitution with a perfectly flat distribution, to 0.068, for a monoalphabetic substitution from common English [15].

3. The Kasiski Method for Repeated Patterns

The method of Kasiski, named for its developer, a Prussian military officer, is a way of finding the number of alphabets that were used for encryption. The Kasiski method follows this rule: If a message is encoded with n alphabets in cyclic rotation, and if a particular word or letter group appears k times in a plaintext message, it should be encoded approximately k/n times from the same alphabet. As an example, if a keyword is six characters long, there are only six different ways to position the keyword over the plaintext word. A plaintext word or letter group that appears more than six times must be encrypted at least twice by the same position of the keyword, and those occurrences will all be enciphered identically. For the Kasiski method, the steps are [10]:

1. Identify repeated patterns of three or more characters.
2. For each pattern write down the position at which each instance of the pattern begins.
3. Compute the difference between the starting points of successive instances.
4. Determine all factors of each difference.

If a polyalphabetic substitution cipher was used, the key length will be one of the factors that appear often in step 4.

4. Columnar Method

The starting step in cryptanalysis of the Transposition Cipher (TC) is to calculate the characters frequencies. The simple fact that all characters will appear at their normal frequencies indicates that a transposition has occurred. The idea is to separate a string of text into columns [14].

The process includes compare a block of ciphertext characters against characters successively farther away in the ciphertext. Assume the block being compared is seven characters. The first comparison is c1 to c8, c2 to c9,...,c7 to c14. Then the window of comparison shifts and c1 is compared

DOI:http://doi.org/10.24086/cocos2024/paper.1484

to c9, c2 to c10, and so forth. The window shifts again to c1 against c10. This process is shown in figure (8).

The process involves comparing a block of ciphertext elements to characters located further apart in the ciphertext. Consider the block that is being compared contains (7) characters. The primary comparison is from c_1 to c_8, followed by c_2 to c_9, and so on until c_7 to c_{14}. The scope of comparison then shifts, with c_1 being assessed against c_9, c_2 to c_{10}, and so on. This process is depicted in Figure (8).

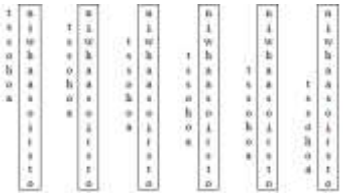

Figure (8) Moving Comparisons.

As another example, consider the following cryptogram:

```
tssoh oaniw haaso lrsto imghw
utpir seeoa mrook istwc nasns
```

The first comparison is of the first seven characters against the next seven: TSSOHOA to NIWHAAS. Some of the diagrams produced that way are likely: -SI-and-SW-, but-TN- and -OH- are unlikely.

The next window places TSSOHOA against IWHAASO. Again, there are some likely diagrams, but -AO- is very suspicious. The third window's comparison also has some suspicious points. We know the next window will be the correct one, since the column length was ten. The comparison there is TSSOHOA against HAASOLR. We see that all the diagrams are reasonable [14].

B. *Cryptanalysis of Modern Cryptosystems*

Now we will discuss the cryptanalysis techniques to attack or break stream cipher (SC) cryptosystems. The classification of these techniques is described in Figure (9).

Figure (9) cryptanalysis techniques of SC.

This categorization is based on the data processing approaches and tools used in cryptoanalysis. The classification separates into two major categories: traditional techniques (TT) and modern techniques (MT). The TT are the most efficient techniques for attacking the SC, but there are others, such as the matrix style technique (MST), the Berlekamp-Massey algorithm (BMA), and correlation techniques (COT) (including fast correlation). However, MT rely on different strategies regarding data processing, particular biological-like processing. We call them modern techniques because they use new tools like Genetic Algorithms (GA) and Neural Networks (NN), which are examples of Artificial Intelligence (AI) [16].

1. *Cryptanalysis of Traditional Techniques (TT)*

Although there is a large body of research on the design of cryptographic systems, there is limited open-access information on cryptanalysis techniques. In this section, we provide an overview of the most efficient techniques for cryptanalysis of SC (linear and non-linear types):

a. Matrix Method

Meyer et al. [17] displayed a technique for breaking a "$N-stage$" $LFSR$ using ($2N$) successive bits corresponding to known PT. The basic technique is to construct the equation of the matrix $K = MC$:

$$\begin{bmatrix} K_{n-1} \\ K_{n-2} \\ K_{n-3} \\ \vdots \\ K_{0} \end{bmatrix} = \begin{bmatrix} M_{n} & M_{n-1} & M_{n-2} & \cdots & M_{1} \\ 0 & M_{n} & M_{n-1} & \cdots & M_{2} \\ 0 & 0 & M_{n} & \cdots & M_{3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & M_{n} \end{bmatrix} \begin{bmatrix} C_{1} \\ C_{2} \\ C_{3} \\ \vdots \\ C_{n} \end{bmatrix}$$

where

K : The vector of the key which want to be calculate.
C : The vector of the CT of switch states.
M : The successive shifts of the first the matrix for n bits of PT.
To obtain the switch values, invert M and solve $C = M^{-1} K$.

As a result, every detail of the key will be fully known. determining the inverse of a matrix is a simple process, despite the fact that it takes $O(N^3)$ time over $GF(2)$. However, this technique is limited because knowing less than $2N$ recurrent bits may not be sufficient for identifying the entirety of the sequence.

b. Algorithm of Berlckamp-Massey for Finding Initial of SR

BerleKamp-Massey iterative algorithm for deciphering Bose-Chaudhuri-Hocquenghem (BCH) codes offers an optimal solution to the problem of creating the smallest $LFSR$ that is able to providing a finite sequence of n bits. The algorithm produces the polynomial with the smallest possible degree (N) by providing ($2N$) bits from a SR of length (N) [18].

c. Correlation and Fast Correlation

In 1985, Siegenthaler [19] suggested the procedure for finding the $LFSRi$ part of the initial key which are loaded to cells separately of the other $LFSRs$ parts using a new basic technique called "divide and conquer". This approach demonstrated that the number of trials required for determining the key will be greatly decreased in instances where a relationship can be established between outcomes of the functioning keygenerator used in a SC and the $LFSRi$ a series with relationship probability ($P \geq 0.75$). The keygenerators have been cryptanalysis for $LFSR$ lengths ($N < 50$).

In 1988, Meier and Staffelbach [20] established two algorithms (A and B) that are significantly quicker than the above attack and revealed to be effective toward SR of a significant N ($N > > 50$), where N represents the number of taping cells is small ($t < 10$ if $P \leq 0.75$).

DOI:http://doi.org/10.24086/cocos2024/paper.1484

2. Modern Techniques (MT) for Cryptanalysis

Modern cryptanalysis techniques rely on novel approaches to reduce the time and cost of an attack or cryptanalysis. Figure (9) illustrates how these techniques can be used, for example; GA and NN concepts.

a. Genetic Algorithm (GA) Technique

GAs are natural selection and genetics-based search algorithms. They use evolutionary theory concepts to "breed" more efficient responses to problems with large solution spaces. GA appears to be an obvious candidate for use in cryptanalysis, given their capacity to efficiently search large solution spaces [21]. Matthews suggested the use of GAs to break or cryptanalysis the classical TC by identifying the transposition sequence used. Spillman et al. [18] suggesting the use of GA for the attack or cryptanalysis of $SCCs$.

However, while some cryptosystems are difficult to be cryptanalysis or attack with GA, numerous cryptosystems, like $SCCs$, appear to be vulnerable to GA attack [21]. Abbas proposed using GA in cryptanalysis of the SCC class in 1998 [22], based on the discovery of a correlation between ciphertext and the results of some $LFSRs$.

b. Neural Networks Method

Artificial Neural Network (ANN) is an information-processing system that has certain performance characteristics in common with biological neural networks. ANNs represent an important area of research, which opens a variety of new possibilities in different fields including classification or pattern recognition or predictions. One of the main advantages of the ANNs is that they infer solutions from data without prior knowledge of the regularities in the data; they extract the regularities empirically [23].

V. INFORMATION HIDING (IH) TECHNIQUE

One of the most recent areas of focus in privacy study involves IH. It was inspired by two of the biggest and most pressing policy issues of the data age: intellectual property rights and government surveillance. IH (directly, wrapped writing) is the concealing oneself of encrypted messages within another seemingly message, such as the host signal, data, or carrier [24]. Figure 10 depicts the categorization of IH.

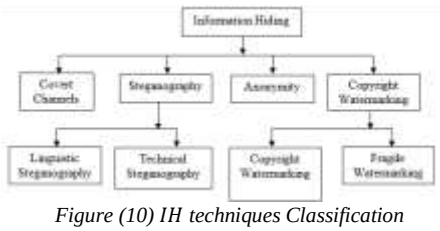

Figure (10) IH techniques Classification

Stenography is the art and science of hiding the fact that communication is taking place. While classical stenographic systems depend on keeping the encoding system secret, modern stenography tries to be undetectable unless secret information is known, namely, secret key.

Anonymity is the ability to conceal the intended content of messages, i.e., the sender and those receiving of a message.

Digital watermarks provide a way to insert a copyright notice into a document or image. The watermark is often a small image or block of text that is repeated frequently through out the document or image.

Covert channels are commonly used by trustworthy programs to leak data to their owners while providing a service to another program [25].

A. Steganography

The word Steganography comes from the Greek Steganos (covered or secret) and Grpahy (writing or drawing) and means literally covered writing. Stenography means encompasses methods of transmitting secret messages through innocuous cover carriers in such a manner that the existence of the embedded messages is undetectable. The hidden message may be plaintext, ciphertext, or anything that can be represented as a bit stream. Cover is an object which considered as an input to the stego-system (SS), this SS is specialized to hides or contains the embedded elements. The specific cover can carriers some innocent-looking transporters (image, audio, video, text, and others) that will carry the hidden data or important and classified information. The information can be embedded in the cover is meant to be hidden. The message or we can say PT is the hidden information, which could be PT, CT, images, or whatever else that can be embedded in a bit stream. Embedding is the procedure, function, method, or technique of hiding an embedded message. Hiding the information need an important factor to complete the embedding process this factor is the Stegokey (SK), which is further secret information, such as a password, required for embedding that data. Extracting means obtaining the hidden CT or PT which is embedded in the stego-cover after using the SK [26].

B. Hiding Techniques

The most common approaches to information hiding in images are [26]:

a. Least Significant Bit (LSB): The LSB insertion method is probably the most well know image Steganography technique. It is a common, sample approach to embedding information in a graphical image file. Unfortunately; it is extremely vulnerable to attacks, such as image manipulation. A simple conversion from GIF or BMP format to a lossy compression format such as JPEG can destroy the hidden information in the image.

b. Masking and Filtering: Masking and filtering techniques hide information by marking an image in a manner similar to paper watermarking. We take advantage of the observation that the human visual system can't recognize slight changes in certain temporal domains of an image by covering, or masking, one pale but obvious indicate by a second to make the first non-perceptible.

c. Algorithms and Transformations: JPEG images use the Discrete Cosine Transform (DCT) to achieve compression. DCT is lossy compression transform,

DOI:http://doi.org/10.24086/cocos2024/paper.1484

because the cosine values cannot be calculated precisely, and rounding errors may be introduced. Variances between the original data and the recovered data depend on the values and methods used to calculate the DCT.

VI. STEGANALYSIS

Steganalysis is the process of detecting and eliminating such covert messages. The term "steganalyst" refers to someone who uses steganalysis to detect hidden information [27]. We assume, as in cryptanalysis, that the steganographic technique is public, with the possible inclusion of a private key. The technique is safe if the stego-images contain no detectable artifacts from message embedding.

Steganalysis is a manner of finding and extracting the covert PT or CT. The expression "steganalyst" means a professional who employs steganalysis for identifying hidden data [27]. We suppose, that the steganosystem is known or public, with the possibility of including a private key. The technique is safe if the stego-cover have no detectable message embedding artifacts.

The attacks on stego-media are classified into two types. The first is the passive attack, which involves exploiting alterations to the cover image of statistical properties that the embedding process ignores and breaking the system. The second is an active attack, which refers to the ability to make minor changes to the cover during the course of communication.

A. Steganalytic Techniques [28]

The most important two Techniques of Steganalytic are:

a. Statistical Attack: the most steganosystems treat LSB's of the cover file as random data and therefore assume that they can overwrite these bits with other random data (the encrypted secret massage), however as the visual attacks have showed the LSB's of image are not random. When a steganographic program a bit through overwrite the LSB of a pixel in the cover file, the color value of this pixel is changed to an adjacent color value in the palette (or in the RGB cub if the cover file is a true color image).

b. Visual Attack: is a stego-only attack that takes advantage of the common assumption among steganosystem authors that the LSBs of a cover file are random. depending on a human to determine whether an image presented by an analyzing algorithm incorporates concealed information or not.

B. Steps of Steganalysis of Images

Steganalysis includes the following parts, all or some of which must be completed depending on the objective of the application: Figure 11 illustrates the detection, extraction, removal, and insertion steps [30].

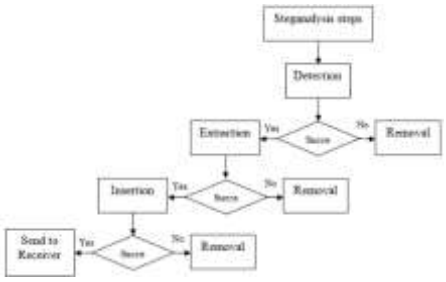

Figure (11) Basic modules in Steganalysis.

C. *Difference between Cryptanalysis and Steganalysis*

Table (2) shows the most important significant differences between Cryptanalysis and Steganalysis [32].

Table (2) The differences between Steganalysis and Cryptanalysis

Steganalysis	Cryptanalysis
try to diagnosis whether there is any of undetected data.	Try out to decipher or crack the message that is encrypted.
Steganography allows for a comparison between the cover-media and possible message portions.	Cryptography compares components of the plaintext (maybe none) to components of the ciphertext.
The steganalyst can attack using known cover (KC), stego-only (SO), chosen stego (CS), known message (KM), known stego attack (KSA), or chosen message (CM).	The cryptanalyst has four attack options: known PT, CT-only, chosen CT, chosen PT.
If the message is encrypted in Steganography, cryptanalysis techniques are used after it has been extracted.	No need for the Steganalysis.

VII. CONCLUTIONS AND FUTER WORK

1. Our conclusion from this paper is that pn-generators should constructed using non-linear function f which have probability that the output z is uncorrelated to all input x_i.
2. The aim of information hiding is to make information invisible to any one during it transmission over network. To make this information more secure, it is possible to encrypt this information before hiding process.
3. GA's and NN's can be used as powerful tools in generating pseudorandom sequences with good statistical properties and a high linear complexity and overcome all problems and difficulties which face designers of cipher systems.
4. GA has proved highly successful in cryptanalysis of most conventional cipher systems (substitution and transposition) and some of modern cipher systems (stream cipher and Knapsack problem).
5. NN's has proved highly successful in cryptanalysis of stream cipher using known plaintext attacks.
6. According to conclusions 4 and 5, we have to design our cipher systems "National systems" which are immune against these types of attacks.

DOI:http://doi.org/10.24086/cocos2024/paper.1484

REFERENCES

[1]. Kahn, D., "The Code breakers: The Story of Secret Writing", New York: Macmillan Publishing Co., 1967.

[2]. Xiaodong L., Gao L. and Gao H. (2003), "Swarm Intelligence and its Applications", Congress on Evolutionary Computation, Canberra, Australia, 8th- 12th December.

[3]. Tariq S. Abdul-Razaq and Faez H. Ali, "Modification of Some Solution Techniques of Combinatorial Optimization Problems to Analyze the Transposition Cipher", Mathematical Theory and Modeling, ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online), Vol.4, No.9, PP 120-141, 2014. www.iiste.org.

[4]. Tariq S. Abdul-Razaq and Faez H. Ali, "Constructing of an Artificial Neural Networks to Minimize Total Completion Time and Total Tardiness", IOSR Journal of Mathematics (IOSR-JM) e-ISSN: 2278-3008, p-ISSN:2319-7676. Volume 10, Issue 2 Ver. VI (Mar-Apr. 2014), PP 25-37 www.iosrjournals.org.

[5]. A. Soria-Lorente and S. Berres, Secure Steganographic Algorithm Based on Frequency Domain for the Transmission of Hidden Information, Hindawi, Security and Communication Networks Volume 2017, Article ID 5397082, 14 pages https://doi.org/10.1155/2017/5397082A.

[6]. F. H. Ali, "Improving Exact and Local Search Algorithms for Solving Some Combinatorial Optimization Problems", Ph.D. Dissertation, Math. Dept, College of Science, Mustansiriyah University, 2015.

[7]. Yan, S. Y., "Number Theory for Computing", Springer-Verlag Berlin Heidelberg, New York, 2000.

[8]. Schneier B., "Applied Cryptography", John Wiley & Sons, 1995.

[9]. Pfleeger, C. P., "Security in Computing", the University of Tennessee, 1989.

[10]. Menezes, A., Van Oorschot, P. & Vanstone, S., "Handbook of Applied Cryptography", Boca Raton: CRC Press, 1997.

[11]. Golomb, S.W., "Shift Register Sequences" San Francisco: Holden Day 1967, (Reprinted by Aegean Park Press in 1982).

[12]. Faez H. Ali and Abdullah A. Naser, "Robust and Efficient Dynamic Stream Cipher Cryptosystem", Iraqi Journal of Science, Vol. 59, No.2C, pp: 1105-1114, DOI:10.24996/ijs.2018.59.2C.15, 2018. 2.10(1).

[13]. Faez H. Ali, Mahammed G. Sabri and Ahmed A. Yousif, "Analyzing Cryptosystems by Using Artificial Intelligence", special Issus: 1st Scientific International Conference, College of Science, Al-Nahrain University, 21-22/11/2017, Part I, pp.100-108, DOI. 10.22401/SIC.I.14, 2017.

[14]. Sinkov, A., "Elementary Cryptanalysis", Mathematical Association of America, 1966.

[15]. Friedman, W. F., "The Index of Coincidence and its Application in Cryptography", Riverbank Publication No.22, 1987.

[16]. Davies D. W., Price W. L., "Security for Computer Networks, an Introduction to Data Security in Teleprocessing and Electronic Funds Transfer", John Wiley and Sons Ltd., 1989.

[17]. Meyer C, Tuchmab W., "Pseudo-Random Codes Can be Cracked", Electronic Design, Vol. 23, pp. 74-76, 1972.

[18]. Spillman R., Janssen M., Nelson B., & Kepner M, "Use of A Genetic Algorithm in the Cryptanalysis of Simple Substitution Ciphers", Cryptologia, vol.16, no. 1, pp. 31-44, 1993.

[19]. Siegenthaler T., "Decrypting A Class of Stream Cipher Using Ciphertext Only", IEEE Transaction on Computers, vol. c-34, no. 1, pp. 81-85, January, 1985.

[20]. Meier W., Staffelbach O., "Fast Correlation Attacks on Stream Cipher", Advances in Cryptology, Eurocrypt, Springer-Verlag, pp. 301-314, 1988.

[21]. Obermeier K. K. & Barron J. J., "Time to Get Fired Up", Byte, pp. 217-224, August 1989.

[22]. Al-Ageelee S. A., "Use of Genetic Algorithms in Cryptanalysis of a Class of Stream Cipher Systems", Ph. D. Thesis, University of Technology, Baghdad, 1998.

[23]. Eskandarian A., Omar T. & Bedewi N. "Vehicle Crash Modeling Using Recurrent Neural Networks", Mathematical and Computer Modeling, Vol. 28, No. 9, pp. 31 - 42, 1998.

[24]. Johnson, N. F., Duric, Z. & Jajodia, S. "Information Hiding: Steganography and Watermarking-Attacks and Countermeasures", Kluwer Academic Publishers, Boston Dodrecht London, 2000.

[25]. Al-Hamami, M., "Information Hiding Attack in Images", M. Sc. Thesis, Iraqi Commission for Computer & Informatics, Informatics Institute for Postgraduate Studies, 2002.

[26]. Sellars, D., "An Introduction to Steganography", http://www.cs.uct.ac.za, 1999.

[27]. Jonathan, W, "Steganography-Messages Hidden in Bits" Multimedia Systems Coursework, Dept. of Electronics and Computer Science, University of Southampton, UK, 2000.

[28]. Provos, N., "Defending Against Statistical Steganalysis", Center for Information Technology Integration, University of Michigan, 2001.

[29]. Faez Hassan Ali, (2006), Use the multiplicative cyclic group to generate pseudo random digital sequences, Al-Rafidain University College For Sciences, Issue 20, pp: 122-135.

[30]. Faez Hassan Ali, (2009), High Efficient Sequences Generate from developed MCG generator, Al- Rafidain university college, Baghdad-Iraq, Issue 25, pp:169-182.

[31]. Ismail, I.M, Amin, M. and Diab, H., (2010), A digital image encryption algorithm based a composition of two chaotic logistic maps, International journal of network security, Vol. 11, No. 1, PP. 1-10.

[32]. Mohammed A. A., (2019), Bifurcation analysis and design of classes of nonlinear dynamical systems with applications, M. Sc. Thesis, Mustansiriyah university, College of Science, Mathematics Science Department.

[33]. Kattan, P.I., (2012), Chaos theory, Simply explained, Basic fractals/ Chaos series.

[34]. Adnan M. Ali, Faez H. Ali, Sabah M. Redha and Nurideen Abubakari, "Image Encryption Using Non-Linear Stream Cipher Cryptosystem", Al-Mustansiriyah Journal of Science, Volume 34, Issue 2, 2023, http://doi.org/10.23851/mjs.v34i2.1294.