CHAPTER 1

INTRODUCTION

1.1 PREAMBLE

Advances in computer and communication technologies have resulted in information exchange in almost all fields and the twenty first century is called information age. The expansion of internet and increase in electronic transaction has resulted in exchange of information over the network. This necessitates the protection of data passing through the network. This is a challenging problem and needs more mathematical methods to be used. Cryptology is the science which uses mathematical designs or quantitative methods to protect confidentiality of data from the access of unauthorized intruders, as well as the design of methods by which these intruders may attack the confidentiality of such data. It is contended that this is not simply an interesting problem from the point of view of an applied mathematician, but that it may and should also be addressed by operations researchers. There is ample scope for techniques which are traditionally associated with operations research, such as statistical analyses as well as exact and heuristic optimization procedures, to be applied to the field of information security (Grundlingh 2003).

Substitution cipher and transposition cipher are the two categories of classical cipher, they are still used in the modern cipher like in Advanced Encryption Standard (AES) and International Data Encryption Algorithm (IDEA). The operations of classical cipher are the building blocks of modern ciphers. Hence, in any research for new attack, the first consideration should be taken is with classical cipher.
1.2 CRYPTOGRAPHY

Cryptography is the science of using mathematics to encrypt and decrypt data. Cryptography enables users to store sensitive information or transmit it across insecure networks (like the Internet) so that it cannot be read by anyone except the intended recipient. While cryptography is the science of securing data, cryptanalysis is the science of analyzing and breaking secure communication. Classical cryptanalysis involves an interesting combination of analytical reasoning, application of mathematical tools, pattern finding, patience, determination, and luck. Cryptanalysts are also called attackers. Cryptology embraces both cryptography and cryptanalysis. (Network association, 1990).

Cryptographic systems are characterized along three independent dimensions:

- 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.

- 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.

- 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. (William Stallings 2005).

Cryptography is the science of making communication unintelligible to everyone except the intended receiver(s). A cryptosystem is a set of algorithms, indexed by some key, for encoding messages into ciphertext and decoding them back into plaintext. Cryptographic systems can be divided into those using a secret key, called symmetric cryptosystems, and those using a public/private pair of keys, called public key or asymmetric cryptosystems. The model for a secret key system first proposed by Shannon (1949) is shown in Figure 1.1. The model for a public key system is given in Figure 1.2.

![Diagram of Shannon's model of secret communication](#)

**Figure 1.1 Shannon’s model of secret communication**
1.3 NEED FOR CRYPTOGRAPHY

Cryptography is concerned with keeping communications private, therefore the most important benefit of cryptography is privacy.

The field of cryptology today represents that branch of information theory which deals with the security of information confidentiality. Methods in cryptology may be subdivided into two classes, namely that of cryptography (methods applied by authorized information sharers to design and develop encryption schemes in order to ensure confidentiality of information; the suffix -graphy means “to write”) and that of crypt-analysis (mathematical and statistical attempts by unauthorized persons to break ciphers in order to reveal the meaning of the underlying protected data).
1.4 TYPES OF CRYPTOGRAPHY

There are several ways of classifying cryptographic algorithms. For purposes of this thesis, they will be categorized based on the number of keys that are employed for encryption and decryption. The three types of algorithms will be discussed.

1.4.1 Symmetric-Key Cryptography

Symmetric-key cryptography refers to encryption methods in which both the sender and receiver share the same key (or, less commonly, in which their keys are different, but related in an easily computable way). This was the only kind of encryption publicly known until 1976 (Whitfield Diffie and Martin Hellman 1976).

The modern study of symmetric-key ciphers relates mainly to the study of block ciphers and stream ciphers and to their applications. A block cipher is, in a sense, a modern embodiment of Alberti's polyalphabetic cipher: block ciphers take as input a block of plaintext and a key, and output a block of ciphertext of the same size. Since messages are almost always longer than a single block, some method of knitting together successive blocks is required. Several methods have been developed, some with better security in one aspect or another than others. They are the mode of operations and must be carefully considered when using a block cipher in a cryptosystem.

1.4.2 Public-Key Cryptography

Symmetric-key cryptosystems typically use the same key for encryption and decryption, though this message or group of messages may have a different key than others. A significant disadvantage of symmetric ciphers is the key management necessary to use them securely. Each distinct
pair of communicating parties must, ideally, share a different key, and perhaps each ciphertext exchanged as well. The number of keys required increases as the square of the number of network members, which very quickly requires complex key management schemes to keep them all straight and secret. The difficulty of establishing a secret key between two communicating parties, when a secure channel doesn't already exist between them, also presents a chicken-and-egg problem which is a considerable practical obstacle for cryptography users in the real world. Figure 1.2 shows simple model of public-key cryptosystem. One algorithm is used for encryption with two keys, one for encryption and one for decryption. The sender and receiver must have one of the matched pair of keys (not the same one).

1.4.3 Hash Functions

Hash functions, also called message digests and one-way encryption, are algorithms that, in some sense, use no key. Instead, a fixed-length hash value is computed based upon the plaintext that makes it impossible for either the contents or length of the plaintext to be recovered. Hash algorithms are typically used to provide a digital fingerprint of a file's contents often used to ensure that the file has not been altered by an intruder or virus. Hash functions are also commonly employed by many operating systems to encrypt passwords. Hash functions, then, provide a measure of the integrity of a file, and it has no key since the plaintext is not recoverable from the ciphertext. Figure 1.3 shows simple model of hash function.
1.5 RANDOMNESS IN CRYPTOGRAPHY

Randomness is crucial in cryptographic applications, because it provides a way to create information that an adversary can’t learn or predict. It’s then the task of a good protocol designer to leverage this power in the best possible way to protect data and communication (Rosario Gennaro 2006).

1.6 CONCEPT OF UNPREDICTABILITY

Although it’s useful in some circumstances, the informal definition from the introductory paragraph seems to confuse the concept of randomness with the concept of unpredictability. A random process is unpredictable, but the reverse is not completely true, some unpredictable processes are not fully random. The canonical example of a random process is tossing an unbiased coin that will land on either side with the same chance (1/2). Repeated coin tosses are independent, meaning the outcome of one toss doesn’t affect the outcome of the others. If the bit 0 is associated with the event that the coin lands on its head, then by repeatedly flipping the coin N times, a uniformly distributed N-bit string can be obtained: each possible string has the same chance of being produced ($2^{-N}$). Consider now the process of rolling a die: its output is the binary expansion of the number on which the die rolls, so if the die rolls a 1, the output will be 001, for 2, it is 010, and so on, until for 6, it’s 110. To produce an N-bit string, the die $N/3$ times are needed to be rolled.
Although this string is unpredictable, it isn’t fully random, in particular, it isn’t uniformly distributed. A roll of the die would never produce strings 000 and 111, for example, which means this process can’t produce all possible $N$-bit strings; the ones that are produced will appear with a probability higher than $2^{-N}$.

### 1.7 UNCONDITIONAL AND COMPUTATIONAL SECURITY

A cryptosystem is unconditionally secure, or theoretically secure, if it is secure against an attacker with unlimited computational resources. It is computationally secure, or practically secure, if it is secure against an opponent with a specified limited computational power.

#### 1.7.1 Unconditional Security

A cryptographic system provides perfect secrecy if the ciphertext letters and the plaintext letters are statistically independent.

Suppose $n$ random variables, $X_0, X_1, \ldots, X_{n-1}$, is considered and defined on the plaintext alphabet and $Y_0, Y_1, \ldots, Y_{n-1}$, defined on the ciphertext alphabet.

$X_0, X_1, \ldots, X_{n-1}$ and $Y_0, Y_1, \ldots, Y_{n-1}$ are independent if:

$$P(X_i = x_i \cap Y_j = y_j) = P(X_i = x_i) \cdot P(Y_j = y_j) \quad 0 \leq i, j < n.$$ 

Shannon showed that perfect secrecy is obtainable, although it requires a key space as large as the set of all possible messages. Such a system is the one time pad, which is a substitution system with a number of substitutions equal to the number of letters in the message (usually defined on a binary alphabet). As long as the key (in this case the set of substitutions)
remains unknown to the attacker and is not reused, it is impossible to break the system (Schneier 1995).

For cryptographic systems with less than perfect secrecy (i.e. for all cryptosystems except variants of the one time pad) there is some dependency between ciphertext and plaintext. This dependency is what the cryptanalyst hopes to exploit to recover the plaintext. But this dependency is only useful if there is some redundancy in the language. Unconditional security is also achievable if a language has no redundancy. Shannon defined a cryptographic system to be strongly ideal if, for a given key, the ciphertext gives no information about the key. If the plaintext letters are independent and equally likely, then almost any substitution system will be strongly ideal.

A strongly ideal system is secure against a ciphertext only attack. The redundancy in a language can be reduced by data compression, but it cannot be entirely removed. Although mainly of theoretical interest, the ideas behind the unconditionally secure systems indicate good principles for the utilization of a cryptosystem, namely that,

- The secret key should be changed as frequently as possible.
- The plaintext letters should be as random as possible.

1.7.2 Computational Security

The complexity of an attack means, the average number of operations used in that attack. A cryptosystem is computationally secure if the complexity of the optimum attack exceeds the computational capability of the cryptanalyst. Generally it is not possible to determine whether an attack is optimal. Instead, the best presently known method of attack is used to evaluate the security. For any attack, the data complexity is the average amount of input data in the form of ciphertext or plaintext/ciphertext pairs.
required, and the processing complexity is the average amount of computations needed to process the data to cryptanalyse a cryptosystem. The dominant quantity is called the complexity for the attack.

1.8 CRYPTANALYSIS OF THE CONVENTIONAL ENCRYPTION SCHEME

The objective of cryptanalysis any encryption system is to recover the key, then to recover the plaintext of a single ciphertext. There are two general approaches to attack a conventional encryption scheme: Cryptanalysis and Brute Force attack.

1.8.1 Cryptanalysis

Cryptanalysis can be described as the process of attempting to recover the plaintext and/or key from a ciphertext. A typical cipher takes a clear text message (plaintext) and some secret keying data (key) as its input and produces an encrypted version of the original message (ciphertext). An attack on a cipher can make use of the ciphertext alone or it can make use of some plaintext and its corresponding ciphertext (known plaintext attack). Cryptanalytic attacks 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. This type of attack exploits the characteristics of the algorithm to attempt to deduce a specific plaintext or to deduce the key being used (William Stallings 2005).
1.8.2 Brute-Force Attack

The attacker tries every possible key on a piece of ciphertext until an intelligible translation into plaintext is obtained. On an average, half of all possible keys must be tried to achieve success. In most cases, a cipher is considered secure if it can only be broken by brute force. If either type of attack succeeds in deducing the key, the effect is catastrophic. All future and past messages encrypted with that key are compromised.

The following are the various types of cryptanalytic attacks, based on the amount of information known to the cryptanalyst.

1. Ciphertext only.
2. Known plaintext: In this type one or more plaintext-ciphertext pairs formed with secret key should be known to the cryptanalyst.
3. Chosen plaintext: Plaintext message chosen by cryptanalyst with its corresponding ciphertext.
4. Chosen ciphertext: Purported ciphertext chosen by cryptanalyst, with its corresponding decrypted plaintext.
5. Chosen text: Plaintext message chosen by cryptanalyst, with its corresponding ciphertext generated, purported ciphertext chosen by cryptanalyst, with its corresponding decrypted plaintext.

The most difficult problem is the ciphertext attack. In some cases, not even the encryption algorithm is known, but in general it is assumed that the opponent does know the algorithm used for encryption. One possible attack under these circumstances is the brute-force approach of trying all possible keys. If the key space is very large, this becomes impractical. Thus,
the opponent must rely on an analysis of the ciphertext itself, generally applying various statistical tests to it. To use this approach, the opponent must have some general idea of the type of plaintext that is concealed, such as English or French text, an EXE file, a Java source listing, an accounting file, and so on. Note that for cryptanalysis of these types of attack, the Cryptanalyst should know the Encryption algorithm and Ciphertext.

1.9 TIME REQUIRED FOR EXHAUSTIVE KEY SEARCH

The disadvantage of exhaustive key search is the time required to find the correct key. The time required to find the key in some types of encryption algorithms range from hours to millions of years based on key size, for example if the key size is 56 bits, the time required to search the key is 10 hours if the system speed is $10^6$ decryption/µs. But for key size 128 bits, the time required to search the key is $5.4 \times 10^{18}$ years for the same system speed.

1.10 CLASSICAL CRYPTANALYSIS

Although the actual word “cryptanalysis” is relatively recent (it was coined by William Friedman in 1920), methods for breaking codes and ciphers are much older. The first known recorded explanation of cryptanalysis was given by 9th century Arabic polymath, Abu Yusuf Yaqub ibn Ishaq al-Sabbah Al-Kindi in “A Manuscript on Deciphering Cryptographic Messages” (Ibrahim 1992). Frequency analysis is the basic tool for breaking classical ciphers. In natural languages, certain letters of the alphabet appear more frequently than others; in English, “E” is likely to be the most common letter in any given sample of text. Similarly, the digraph “TH” is the most likely pair of letters, and so on. Frequency analysis relies on a cipher failing to hide these statistics. For example, in a simple substitution cipher (where each letter is simply replaced with another), the most frequent letter in the
ciphertext would be a likely candidate for “E”. Frequency analysis relies as much on linguistic knowledge as it does on statistics, but as ciphers became more complex, mathematics gradually became the main approach to cryptanalysis. To give a concrete example of redundancy in English, single character frequencies (including the apostrophe) are shown for a sample corpus in Figure 1.4.

![Graph of character frequencies](image)

**Figure 1.4 Relative character frequencies from great expectations (1860-1861)**

1.11 CLASSIFYING SUCCESS IN CRYPTANALYSIS

Lars Knudsen (1998) classified various types of attack on block ciphers according to the amount and quality of secret information that was discovered:

- Total break - the attacker deduces the secret key.
- Global deduction - the attacker discovers a functionally equivalent algorithm for encryption and decryption, but without learning the key.
- Instance (local) deduction - the attacker discovers additional plaintexts (or ciphertexts) not previously known.
• Information deduction - the attacker gains some Shannon information about plaintexts (or ciphertexts) not previously known.

• Distinguishing algorithm - the attacker can distinguish the cipher from a random permutation.

Similar considerations apply to attacks on other types of cryptographic algorithm.

1.12 METHODS OF CRYPTANALYSIS OF CLASSICAL CIPHERS

1.12.1 Frequency Analysis

In the field of cryptanalysis, frequency analysis is a method for “breaking” simple substitution ciphers, like the Caesar cipher. These ciphers replace one letter of the plaintext with another to produce the ciphertext, and any particular letter in the plaintext will always, in the simplest and most easily breakable of these ciphers, turn into the same letter in the cipher. For instance, all E’s will turn into X’s. Frequency analysis is based on the fact that certain letters and combinations of letters, appear with characteristic frequency in essentially all texts in a particular language. For instance, in the English language, E is very common, while X is not. Likewise, ST, NG, TH, and QU are common combinations, while XT, NZ, and QJ are exceedingly uncommon, or even “impossible”. Given our example of all E’s turning into X’s, a ciphertext message containing lots of X’s already seems to suggest one pair in the substitution mapping. In practice the use of frequency analysis consists of first counting the frequency of ciphertext letters and then assigning “guessed” plaintext letters to them. Many letters will occur with roughly the same frequency, so a cipher with X’s may indeed map X onto R, but could
also map X onto G or M. More Xs in the ciphertext than anything else suggest E in the plaintext, but T and A are also very common in English, so X might be either of them also. It’s very unlikely to be a plaintext Z or Q which aren’t common in English, though Z is more common in both German and Italian. Thus the cryptanalyst may need to try several combinations of mappings between ciphertext and plaintext letters. Once the common letters are solved, the technique typically moves on to pairs and other patterns. These often have the advantage of linking less commonly used letters in many cases, filling in the gaps in the candidate mapping table being built. For instance, Q and U nearly always travel together in that order in English, but Q is rare. Frequency analysis is extremely effective against the simpler substitution ciphers and will break astonishingly short ciphertexts with ease. Frequency analysis was first discovered in the Arab world, and is known to have been in use by about 1000 CE. It is thought that close textual study of the Koran first brought to light that Arabic has a characteristic letter frequency which can be used in cryptanalysis. Its use spread, and was so widely (though secretly) used by European states, that by the Renaissance several schemes were invented by cryptographers to defeat it.

1.12.2 Index of Coincidence

The index of coincidence is useful both in the analysis of natural-language plaintext and in the analysis of ciphertext (cryptanalysis). Even when only ciphertext is available for testing and plaintext letter identities are disguised, coincidences in ciphertext can be caused by coincidences in the underlying plaintext. This technique is used to cryptanalyze the Vigenère cipher. For example, for a repeating-key polyalphabetic cipher arranged into a matrix, the coincidence rate within each column will usually be highest when the width of the matrix is a multiple of the key length, and this fact can be
used to determine the key length, which is the first step in cracking the system (Menezes et al. 1997).

Coincidence counting can help determine when two texts are written in the same language using the same alphabet. (This technique has been used to examine the purported Bible code.) The causal coincidence count for such texts will be distinctly higher than the accidental coincidence count for texts in different languages, or texts using different alphabets, or gibberish texts.

1.13 METHODS OF CRYPTANALYSIS OF SYMMETRIC ALGORITHMS

1.13.1 Brute Force Method

In cryptanalysis, a brute force method is a way of defeating a cryptographic scheme by trying a large number of possibilities; for example, exhaustively working through all possible keys in order to decrypt a message. In most schemes, the theoretical possibility of brute force method is recognized, but it is set up in such a way that it would be computationally infeasible to carry out. Accordingly, one definition of “breaking” a cryptographic scheme is to find a method faster than a brute force method. The selection of an appropriate key length depends on the practical feasibility of performing a brute force attack. By obfuscating the data to be encoded, brute force attacks are made less effective as it is more difficult to determine when one has succeeded in breaking the code.
1.13.2 Differential Cryptanalysis

Differential cryptanalysis is a general form of cryptanalysis applicable primarily to block ciphers, but also to stream ciphers and cryptographic hash functions. In the broadest sense, it is the study of how differences in an input can affect the resultant difference at the output. In the case of a block cipher, it refers to a set of techniques for tracing differences through the network of transformations, discovering where the cipher exhibits non-random behaviour, and exploiting such properties to recover the secret key (Coppersmith 1994).

1.13.3 Linear Cryptanalysis

In cryptography, linear cryptanalysis is a general form of cryptanalysis based on finding affine approximations to the action of a cipher. Attacks have been developed for block ciphers and stream ciphers. Linear cryptanalysis is one of the two most widely used attacks on block ciphers; the other being differential cryptanalysis. The discovery is attributed to Mitsuru Matsui, who first applied the technique to the Fast Data Encipherment Algorithm (FEAL) cipher (Matsui and Yamagishi 1992). Subsequently, Matsui published, “An attack on the Data Encryption Standard (DES)”, eventually leading to the first experimental cryptanalysis of the cipher reported in the open community (Matsui 1993, 1994). The attack on DES is not generally practical, requiring $2^{13}$ known plaintexts.

A variety of refinements to the attack have been suggested, including using multiple linear approximations or incorporating non-linear expressions. Evidence of security against linear cryptanalysis is usually expected of new cipher designs.
1.14 CLASSICAL CIPHER

Classical ciphers operate on letters or groups of letters and were, in practice, implemented by hand or with simple mechanical devices. By contrast, modern schemes use computers or other digital technology, and operate on bits and bytes. Classical schemes are often breakable in a ciphertext-only attack, and sometimes even without knowledge of the system itself, typically using frequency analysis. Sometimes classed with classical ciphers are the electromechanical rotor machines, such as the Enigma machine. The importance of classical ciphers is that, the modern ciphers in common use today utilize the operations of the classical ciphers as their building blocks.

1.15 BREAKING THE CIPHER

Bruce Schneider sums up this approach: “Breaking a cipher simply means finding a weakness in the cipher that can be exploited with a complexity less than brute force. Never mind that brute-force might require $2^{128}$ encryptions; an attack requiring $2^{110}$ encryptions would be considered a break...simply put, a break can just be a certification weakness: evidence that the cipher does not perform as advertised” (Schneier 2000). Breaking a cipher doesn't necessarily mean finding a practical way for an eavesdropper to recover the plaintext from just the ciphertext.

1.16 COMPLEXITY

Attacks can also be characterized by the amount of resources they require. This can be in the form of:
• Time - the number of “primitive operations” which must be performed. This is quite loose; primitive operations could be basic computer instructions, such as addition, XOR, shift, and so forth, or entire encryption methods.

• Memory - the amount of storage required to perform the attack.

• Data - the quantity of plaintexts and ciphertexts required.

In academic cryptography, a weakness or a break in a scheme is usually defined quite conservatively.

1.17 PASSIVE ATTACKS

In cryptography, general keys are kept secure by the parties using them; in particular, if a key is stored in a computer, by assumption an adversary can’t break into that computer and steal the key. Let’s consider an attacker who can only eavesdrop on the transmissions between Alice and Bob. We might think that his or her goal would be to recover the secret key $K$, so that he could decrypt all further transmissions between Alice and Bob, but this is a very ambitious goal that would correspond to a total break of the encryption scheme. In reality, the attacker’s goal would be something easier - for example, recovering $M$ given $C$, or even obtaining only one specific bit of $M$. Because this bit might carry valuable information, being able to recover it would clearly make the scheme insecure (Jean-Sébastien Coron 2006).

1.18 ACTIVE ATTACKS

In practice, though, an adversary might be able to inject or modify the messages transmitted over a network. Specifically, he or she could inject
ciphertexts and possibly obtain some partial information about their corresponding plaintexts. To deal with active attacks, Rackoff and Simon (1991) introduced the notion of security under an adaptive chosen ciphertext attack. As earlier, the adversary must tell whether a challenge ciphertext \( c \) is an encryption of \( m_1 \) or \( m_2 \), but he or she can also obtain the decryption of any ciphertext, except for the challenge ciphertext \( c \). Ronald Cramer and Victor Shoup invented the first practical public-key encryption scheme to provably achieve this property (Cramer and Shoup 1998) based on a standard hardness assumption. The Optimal Asymmetric Encryption Padding (OAEP) scheme mentioned earlier also achieves this property, but only in the random oracle model, an idealized model of computation in which hash functions are viewed as completely random functions. The field of provable security is the combination of three steps: security definition, scheme, and proof of security. Although this approach is now part of the mainstream in modern cryptography, provable security has some limitations. In the real world, real computation takes time, consumes power, and leaks radiation, and adversaries can exploit these vulnerabilities with timing attacks, differential power analysis, and fault attacks. An interesting future research direction would be to extend the models used in provable security to include such attacks (Jean-Sébastien Coron 2006).

1.19 GENETIC ALGORITHM IN CRYPTANALYSIS

The most fundamental criterion for the design of a cipher is that the key space (the total number of possible keys) should be large enough to prevent it from being searched exhaustively. Due to their complexity, the task of determining weakness in ciphers is generally a difficult manual task and the process of exploiting the discovered weaknesses is rarely quick or simple even when computers are used to implement the exploit (unless the cipher contains a major flaw). This further highlights the importance of the use of
genetic algorithm in cryptanalysis. The possibility of using the genetic algorithm in key search is very attractive due to the ability of genetic algorithm to reduce the complexity of the search problem, and since the cryptanalysis problem is a search problem in principle, and the security of a cipher is based in many cases on the complexity of the type of attacks to this particular cipher system. Many researches have developed in the direction of the reduction of the complexity of the attacks on different cipher systems. One of these elegant techniques developed was the differential cryptanalysis. As Albassal and Wahdan (2003) concluded, the complexity of the attack was reduced in their work, but by using the Genetic Algorithms (GA), the complexity of the attack is reduced by at least 50%.

1.20 GA BASED VIGENERE CIPHER (POLYALPHABETIC)

The first step before starting proposed method to cryptanalysis Vigenere cipher is to find the key length which is guessed by using new method with the benefit of Genetic Algorithm. The main idea in this method, is to employ frequency of bigrams and trigrams as cost function in genetic algorithm with few numbers of parameters. After that, the proposed method will be applied to get the correct key. The attack is implemented by generating 10 independent keys to represent the target key. The first generation is generated randomly using a simple uniform random number generator. The fitness value is incremented and finally normalized to the number of pairs. The genetic algorithm then goes in the normal way to generate new generations. The algorithm is terminated based on the criteria described earlier. The algorithm has been implemented to get fitness; essentially the attack shall continue upward to get the best key.

Two types of cost function are used to calculate the fitness value:
• The fitness function relied on the language statistical characteristic to represent the fitness of the key. For example, the letter "E" is the most common letter in English language, so the fitness of the key can be measured based on how likely it is going to give correct letter frequency in the deciphered text.

• Instead of using all possible bigrams and trigrams a subset of the most common ones are chosen.

1.21 GA BASED SUBSTITUTION CIPHER

Of two different ways to calculate the fitness, first way is frequency analysis by special equation and the second way is by using score of bigrams and trigrams.

The block size \((B=3)\) and this GA’s position within the block, \(j (1<=j<= 3)\) the maximum number of iterations for the GA \((G)\) and the solution pool size \((M=24)\).

The algorithm should be repeated till assumed iteration number, but this number can be increased till getting the best solution.

1.22 GA BASED VERNAM CIPHER

The first step in the proposed approach that is based on Genetic Algorithm for cryptanalysis of Vernam cipher is to build a dictionary which consists of words that show up frequently in a general English document,
such as the following words: Dictionary = {this, it, has, have, do, does, as, been, what, when, how, why, who, such, the, while, when. ...}.

A good chromosome should have many correct keys that decipher the ciphertext to many correct words in the dictionary. Then, a good chromosome should have a high fitness value for reproduction ciphertext to many correct words in the dictionary. Then, a good chromosome should have a high fitness value for reproduction. After reproduction, a simple two-point crossover is performed. Crossover is the process by which two parent chromosomes recombine to create a new offspring chromosome. Over many generations the solutions in the population are improved until the best of the population is near optimal.

The fitness of chromosomes is calculated and then the parents of the next generation are selected based on their fitness value. Mutation operator can be also applied when the chromosome generated is not up to the expected fitness value. Here a key will be changed randomly by some random generate key. The data chromosome coding is performed as sequence of alphabets.

1.23 GA BASED TRANSPOSITION CIPHER

Genetic Algorithms in cryptanalysis of transposition cipher is applied as usual for other types of cipher, the modification has been done in some steps which are the cost function and some of changes in crossover operation.

The following is the proposed algorithm briefly:

- Inputs are the Ciphertext, key size and maximum number of iterations.
• Generate 12 keys randomly.
• Decrypt the ciphertext by the 12 keys.
• Calculate the cost function of each key and sort the keys based on decreased value of cost function
• Choose the 6 pairs for next generation
• Apply the operations of Genetic Algorithm

Best key is solution for cryptanalysis of transposition cipher

The unigram frequencies for a message are unchanged during the encryption process of a transposition cipher and so, they are ignored when evaluating a key. There is another method for calculating the fitness value; this method is called frequency score of common bigrams and trigrams.

1.24 MOTIVATION OF THE RESEARCH

The motivation of this research is to use genetic algorithm to reduce the search space of some classical cipher algorithms. The aim of cryptography is to hide the information from the hacker and the importance in cryptography is the strength of the key, which makes the cryptosystem unconditionally secure, that is the ciphertext does not contain enough information to determine the plaintext, and makes the cryptosystem computationally secure if it satisfies the following two criteria:

• The cost of breaking the cipher exceeds the value of the encrypted information.
• The time required to break the cipher exceeds the useful lifetime of the information.
In order to investigate the use of genetic algorithm in cryptanalysis, the following investigations have been made in this research:

1. Investigating the Cryptanalysis of Polyalphabetic Substitution and Vigenere Ciphers.

2. Investigating the methods of guessing the key in Vigenere Cipher.

3. Investigating the use of Genetic Algorithm in Cryptanalysis of Vernam

4. Investigating the use of Genetic Algorithm in Cryptanalysis of Transposition Ciphers.

1.25 THESIS ORGANIZATION

The Thesis is organized as follows: Chapter 1 deals with background of the Cryptography, need of Cryptography and the types of Cryptography. It describes the Cryptanalysis and methods of Cryptanalysis. Further, it briefs the use of Genetic Algorithm with classical ciphers.

Chapter 2 deals with Literature Survey of some of the existing methods (Substitution Ciphers, Vigenere Cipher, Vernam Cipher and Transposition Cipher).

Chapter 3 deals with the background of Polyalphabetic Cipher and its advantages. It describes the principle of the Vernam Cipher, its security and Transposition Ciphers with their types.

Chapter 4 deals with Genetic Algorithm and its operation. It describes stopping conditions for Genetic Algorithm.
Chapter 5 deals with the proposed methods in Cryptanalysis of Vigenere Cipher and substitution Cipher with the experimental results. It compares the results with existing methods.

Chapter 6 deals with the proposed methods in Cryptanalysis of Vernam Cipher and Transposition Cipher with the experimental results.

Chapter 7 discusses the conclusions of the thesis and future extensions.