International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 24 (2017) pp. 15363-15377

Security Issues in DNA Based on Data Hiding: A Review

[1]Dilovan Asaad Zebari, [2] Habibollah Haron and [3]Subhi R. M. Zeebaree

[1,2] Faculty of Computing, Universiti Teknologi Malaysia (UTM), 81310 Johor Bahru, Johor, Malaysia.
[3] Department of Computer and Communications Engineering, Nawroz University, Tenahi, Duhok, Kurdistan Region, Iraq.

[1]Orcid: 0000-0002-7643-6359

Abstract

Information security and confidentiality are a key concern, particularly with the rampant growth and use of the internet. Along with the growth comes the incidents of unauthorized information access which are countered by the use of varied secure communication techniques, namely; cryptography and data hiding. More recent trends are concerned with the application of DNA cryptography and data hiding by using it as a carrier thereby making use of its bio-molecular computational properties. This paper provides a survey of recently published DNA based data hiding algorithms which make use of DNA to safeguard critical data being transmitted over an insecure communication channel. Several DNA-based data hiding techniques will be discussed with particular emphasis on strength and weaknesses of the algorithm in question; algorithms are compared based on the cracking probability, double layer of security, single or double hiding layer, blindness, and much more. This will be useful for future research in the design of more efficient and reliable of secure DNA data hiding techniques.

Keywords: Data security, Cryptography, Data hiding, DNA.

INTRODUCTION

In general, the notion of security is the protection of information from being accessed by unauthorized people. The main purpose of security in the modern perspective of computer science is to keep sensitive data from being changed, destroyed, and stolen or being damaged by a third party [1]. The most commonly used techniques in communication and computer security fields are cryptography and data hiding. Usually, communications are kept secure through the use of cryptography, and data hiding (which are related concepts) [2, 3]. Even though both methods have a similar goal, their implementation and usage differ greatly; cryptography changes the meaning of secret writing while data hiding is a hidden form of writing which hides the existence of the embedded message. Thus, data hiding is more secure, sufficient, and often preferred to cryptography in the transmission of data across an insecure, public channel [4, 5].

DATA HIDING REQUIREMENTS

In literature, several data hiding schemes are discussed in this paper, along with their parameters. Different schemes use various data embedding approaches, each of which have common requirements which are used in measuring performance and identifies their advantages and disadvantages. One of the three common requirements is vulnerability to attacks by intruders. It is imperative that the embedded data is kept undetectable perceptually as well as statistically so as not to arouse any suspicion. In addition, care should be taken to ensure that the carrier files' attributes and properties are not tampered with in the course of embedding the message thereby providing a perfectly secure system where statistics of both carrier and stego files are identical [5, 6]. The second common requirement is capacity, defined as the quantity of data hidden within the carrier. The design of a data hiding technique should be done in a manner that allows for more secret data to be hidden within the carrier while retaining the properties of the stego file [1, 5]. A good data hiding technique should have sufficient embedding capacity for holding the information [6]. The third common requirement is imperceptibility, defined as the possession of a high embedding capacity as well as the ability to resist intruders. The stego carrier should preferably be free from visual artefacts; the higher the fidelity of the stego carrier, the better [2].

FUNDAMENTAL CONCEPTS OF DATA HIDING

The concept of data hiding is usually modelled by a pair of algorithms which are embedding and extracting as shown in Figure 1. An embedding algorithm refers to the combination of two files together, the secret data and carrier with an optional key in order to obtain a stego file which holds the secret data. However, the extraction algorithm serves the purpose of extracting the secret data from the stego file [7]. Pure data hiding is considered to be a category of data hiding type which does not utilize any key; its security is based on the privacy of the algorithm. Thus, it is considered to be a less secure method [8, 9]. Secret data hiding is another data hiding type which uses a single key for both processes (embedding and extraction). One of the greatest advantages of this type is the provision of a fast process in during both procedures [10, 11]. Unlike previous types, public data hiding uses a pair of keys in both processes:

one for embedding and another one for extracting. The main advantage of this type is the robustness of the system; if one key is known by a third party, it is quite hard to find the other key [10, 12]. However, this type is slower than private data hiding by about 100-1000 times [13].

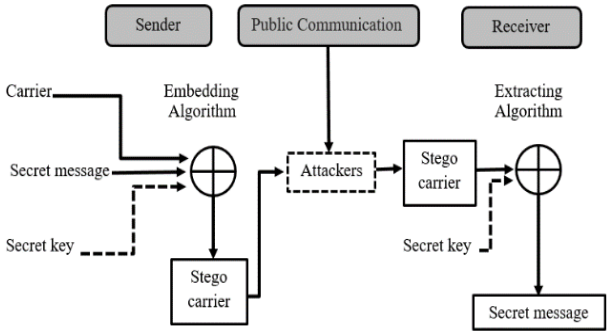

Figure 1: General Principle of Data Hiding System

Several applications are used to represent a container for sensitive information. These applications are used as cover objects or carriers in the data hiding systems. Every carrier has its own characteristics which serve the data hiding technology. The amount of secret information required in order to conceal data within each carrier depends on the availability of the region of the specific carrier. Therefore, carriers are an essential ingredient of any data hiding system. Multimedia that serves as data hiding carriers include text, audio, video, and images. Text can be hidden by modifying the layout of the text, using an *n*th character from text or by the alteration of some of the rules such as spaces, etc. Text can also be hidden by using a code made up of characters, lines and page numbers. This approach, however, is not secure [2]. The main advantage of this carrier is that it does not require large volumes of memory and it is also easy to transfer. While it has very small amount redundant data compared with other carriers [10, 14]. Data can be hidden in audio files through the use of inaudible frequencies and by slightly changing the binary sequence of an audio file [2, 15]. Hiding data in video files is much more effective and successful due to the large capacity available thus making it possible to hide the data within different frames of a video [16]. Generally, data hiding in video is classified into two main types which are uncompressed and compressed video. Digital images have also become popular carriers for hiding secret information since they possess a high degree of redundancy, high capacity in images, low impact on the visibility and simplicity in their manipulation [15, 17]. DNA is a recent carrier have been used in data hiding area. In this paper, we focus only on data hiding in DNA.

DNA

In biology, a Deoxyribonucleic Acid (DNA) is the leading molecular structure for encoding the data required to create and direct all chemical elements in the human body. For this reason, DNA has been put forward as a viable option for use in computational applications [18].

DNA Structure

DNA is defined as the genetic drawing of every living creature. Each body cell has a unique complete set of DNAs; a polymer comprised of monomers referred to as deoxyribose nucleotides made up of three components as shown in Figure 2 [19].

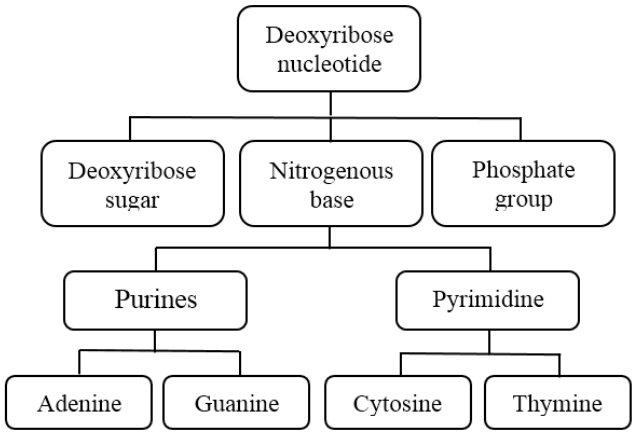

Figure 2: The Structure of Deoxyribose Nucleic DNA [19]

The human body consists of trillions of cells each of which serves different functions. Each cell contains a nucleus that has a number of chromosomes as illustrated in Figure 3. Most of the DNA contents are found in a nucleus called nuclear DNA, and the rest of it is contents are found in mitochondria which are called mitochondria DNA (mtDNA). The function of each cell is controlled by DNA. Each DNA's chromosome consists of a DNA molecule which holds genes. The gene is the entire genetic makeup, which essentially contains information from all the chromosomes [20].

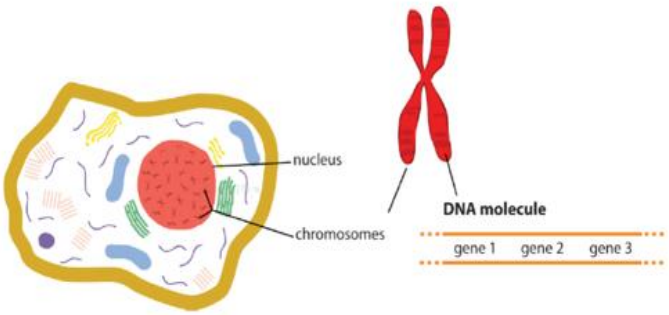

Figure 3: Cell and it is Genetic Material [20]

The genetic material known as DNA structure was found out by Watson and Crick in (1953). DNA is a long molecule located in all living organism's body cells. The DNA is a germ form of

plasm that contains all lifestyle; it is formed by two backbone strands twisted around each other, called a Double-helix as shown in Figure 4. Each DNA strand is composed of many tiny subunits called nucleotides. Adenine (A), thymine (T), guanine (G), and cytosine (C) are the four chemical bases held in the DNA sequence which are stuck onto sugar and phosphate on the backbone to complete the nucleotide. Biologically, there are two pairs of DNA bases: Purine (A and G) and Pyrimidine (T and C). Constantly (A) linked to (T) within two hydrogen bonds and (C) linked to (G) within three hydrogen bonds [19, 21].

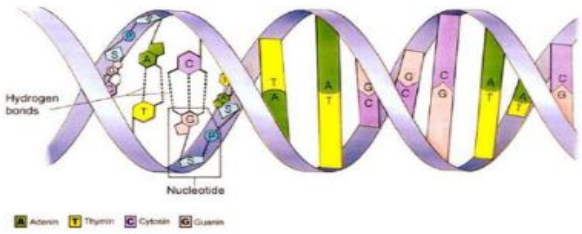

Figure 4: Helical Structure of DNA [20]

Every three adjacent nucleotides make up a codon. Given that each nucleotide could have any of the four chemical bases and each codon is comprised of three nucleotides, then there is a sum of 4^3 = 64 different possible combinations. These combinations determine the amino acids to be used by living organisms, whose arrangement determines the structure and function of the resultant protein [10]. Transcription is the process through which RNA, an intermediary copy of the instructions contained in DNA, is created. The RNA is also made up of four bases: adenine (A), cytosine (C), uracil (U) and guanine (G). Figure 5, shows all of the 64 codons. Three of which, the ones labelled STOP, do not actually code for any amino acid but instead, they signal the end of the protein chain. The remaining 61 codons are responsible for specifying the 20 amino acids. However, some of the amino acids are coded for by more than one codon through the feature referred to as degeneracy [11]. It follows that this codon redundancy can be used to alter the genetic sequence while retaining its functionality [11, 22, 23].

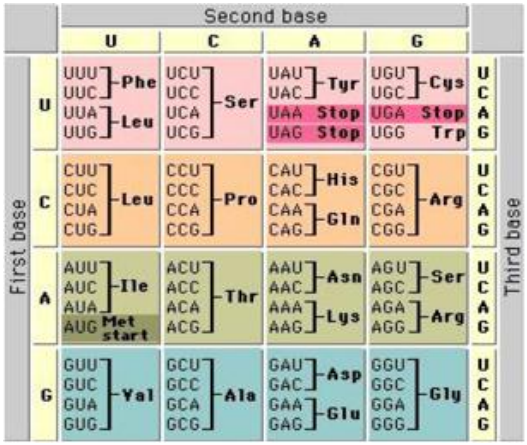

Figure 5: The Amino Acid and Codons Table [11]

DNA Computing

For the being time, biology techniques diffuse in many different areas. DNA is the most recent biological technique being used in various applications [24]. This is because DNA computing can potentially be used to solve a wide range of computational problems whose computation time grows exponentially; `NP-complete' or non-deterministic polynomial time complete problems. A significant amount of research in this area was done, with great strides being made on the subject of DNA and the immune system [19]. In 1994 the first experiment of DNA computing (bio-molecular computing) had done by Leonard Adelman, in this method tools of molecular biology had been used for solving a part of standard path of Hamiltonian problem. At that time, computing with molecule directly was discovered a field that considered as new in term of science security [25]. The research study by Lipton in 1995 used DNA computing to solve the satisfaction problem (SAT), which is a NP-complete problem. The solution provided made use of the parallelism property of DNA as well as it is computing and storage capabilities [19]. Later, in 1997 Ogihara and Ray discovered that DNA can simulate Boolean AND and OR gates [26]. The first experiment of data hiding technique in DNA has been proposed successfully, using DNA microdots, for concealing secret data by Clelland [27].

DNA Binary Coding

Each DNA strand has four chemical bases: A, C, G, and T. Biologically, A is associated to T and C is associated to G. In the binary computing area, the synthesis of DNA bases can be changed by the input decisions, by assuming that T is associated with C or T is associated with G, and so on [28]. In order to store data in DNA molecules, researchers have to encode a secret message into DNA bases using a binary coding rule to combine with the DNA sequence. Researchers have the option of selecting any equivalent binary form for each base (A): the binary forms can be '00', '01', '10', or '11', and so on. This coding along with the randomness properties makes DNA a befitting application for both computing and cryptography. Therefore, the coding of DNA to binary form can give 4! = 24 different encoding ways [29]. Which in turn makes it possible to carry out logical operations such as Addition, Subtraction, XOR, AND, OR, and NOT over the DNA bases.

F(X): X → Y, where X = {A, C, G, T} and Y = {00, 01, 10, 11}. This can be express as in Table 1.

Table 1: DNA Binary Coding

DNA Base	Binary code
A	00
C	01
G	10
T	11

DNA Data Hiding Techniques

Several carriers have been used in data hiding algorithms where each carrier has it is own characteristics. Depending on the specific carrier, different techniques have been used to hide secret information. The process of hiding and the mechanism of each technique is different from another. According to the techniques, different changes will happen during the hiding process. In DNA data hiding, three techniques were proposed in 2010 by [23], all of which were considered to be the main techniques of hiding data in DNA sequences. The three main techniques can be defined as follow:

Insertion Technique

This technique depends on the merging among the reference of DNA sequence (*S*) which use as a carrier and secret message. During the process of this technique, both of them are translated into a binary system according to any binary coding rule. After that, the DNA reference is separated into equal sized segments in order to insert each bit of the secret message after each segment of DNA reference and then converting back into a DNA sequence resulting in a stego DNA. Furthermore, less modification rate is considered as a great feature of this technique because it depends on inserting secret data in the DNA reference not replacing the contents of DNA reference [30]. However, the main disadvantage of this technique is the increase in redundancy during the process, and the stego DNA length will be higher than that of the DNA reference. This implies that the use of this technique will attract the attention of unauthorized users [30, 31, 32, 33, 34].

Complementary Pair Rule Techniques

In this technique, the procedure begins with the selection of a DNA sequence in which the longest existing complementary pair is contained. This is followed by the random generation of two complementary string pairs whose length is one more than that existing in the sequence, after which these pairs are padded with a 'T' at the posterior and anterior. Afterwards, they are inserted one at a time into *S* while ensuring that there is no overlapping. The message is then divided into segments, each containing an even number of bits after which the data is coded back into nucleotides using the binary coding rule. For each pair of a complementary substring in the converted sequence, a message bit is inserted before T*aj*T, where *aj* represents the pair of longest complementary substrings. A resultant sequence containing message *S*' is then obtained. This scheme results in a significantly alters the length of the DNA sequence which rouses the suspicion of a hacker to the existence of an embedded message [30]. Legally there are six main complementary rules as it is shown in Figure 6.

AT	TC	CG	GA
AT	TG	GC	CA
AC	CT	TG	GA
AC	CG	GT	TA
AG	GT	TC	CA
AG	GC	CT	TA

Figure 6: The Six Complementary Rules

Substitution Technique

Regarding this technique there is no merge between reference DNA sequence and the secret information. In this scheme, specific positions in the DNA reference are selected randomly as determined by the algorithm. After that, at least one complementary rule should be selected to replace each letter of the message with the DNA contents in particular locations. Depending on the contents of the message, the process will be carried out to obtain the stego DNA. Hence, the DNA length is maintained following the embedding of the message only the replacement has done between secret message and DNA reference. This, in turn, means that in an effort to conceal the secret data, the resulting stego DNA is highly modified [30]. As a result, this technique is considered as a more efficient technique than the previous techniques because it provides more complexity and better performance [36].

COMPARATIVE STUDY

In this subsection some of security issues will be compared among forty-one recent algorithms such as cracking probability, double layer of security, single or double hiding layer, blindness, and more in Table II relative to strong and weak points of each one. The objective of the comparison provided in this study is to ensure that researchers are armed with the knowledge of the disadvantages presented by existing data hiding schemes thereby providing motivation for future advances in this field.

Table 2: Comparison of Strong and Weak Points between Various DNA Based on Data Hiding Techniques

No	Author	Year	Strong Points	Weak Points
1	Ref [23]	2010	***Insertion technique*** . High capacity. . Easy to implement. . Low modification rate.	**.** Length of stego DNA is longer than DNA reference. . Payload not equal zero. . Need multiple data in extraction process. . Does not preserve the functionality of the amino acid. . Un-blind algorithm. . Increase the redundancy. . Pure data hiding method.
			Complementary Pair rules technique . Easy to implement. . Attackers need to know multiple data to crack the secret data.	. Payload not equal zero. . High modification rate. . Does not preserve the functionality of the amino acid. . Un-blind algorithm. . Changing the length of DNA after embedding process. . Pure data hiding method.
			Substitution technique . High capacity. . Payload is equal to zero. . Easy to implement. . More efficient, more complexity and better performance than previous methods.	. High modification rate. . Does not preserve the functionality of the amino acid. . Un-blind algorithm. . Pure data hiding method.
2	Ref [21]	2011	. Carrying out the result of hiding data in cloud to increase the level of confidentiality and complexity. . Simple algorithm and high capacity. . Payload equal to zero. . Preserve the functionality of the biological DNA.	. Un-blind algorithm. . Increase the message size. . Security depends on the DNA reference. . Pure data hiding algorithm.
3	Ref [24]	2011	. Construct a reversible data hiding method. . Preserve the functionality of DNA. . Blind algorithm. . Secret key is used.	. Does not encrypt secret data before hiding.
4	Ref [18]	2012	. Mapped between DNA codons and amino to provide security. . Encrypt secret message by playfair cipher before hiding. . Improve playfair cipher by modifying to 5*5 to avoid its drawbacks where after encryption the diagraphs and the structure of secret text still exists. . Providing double layer of security. . Blind Algorithm. . Hight capacity and better time performance. . Provide high probability cracking. . Secret key is used.	. Expand the length of stego DNA. . Not preserve the functionality of the biological DNA. . In order to extract the secret message from stego DNA, it needs to send multiple data to receiver. . Payload not equal to zero.
5	Ref [36]	2012	. Enhanced the effectiveness original substitution method. . Transmission efficiency can be increased for data hiding system in internet. . Providing more performance in term of capacity and security.	. Not preserve the functionality of the biological DNA. . To extract the secret message from stego DNA, it needs to send multiple data including DNA reference, stego DNA, secret message location set, and the table rule.

			. TLSM is extended which can hide secret data in any sequences of letters or symbols. . To improve the performance of the TLSM two approaches are presented which are the Base-t TLSM and the Extended TLSM (ETLSM). . Capacity increased.	. Un-Blind algorithm. . High modification rate. . Pure data hiding algorithm.
6	Ref [13]	2012	. Propose a cryptographic-data hiding protocol reduce the using of public key as well as for best security. . Payload is equal to zero. . High capacity. . Utilizing the innovative of DNA data hiding to hide the secret key within DNA reference for more security.	. Un- blind algorithm. . Does not preserve the functionality of the biological DNA.
7	Ref [31]	2012	. Does not expand the length of stego DNA means payload equal to zero. . Simple algorithm. . Improved the hiding capacity. . Minimize the modification rate. . Used substitution method in hiding.	. Does not preserve the functionality of the biological DNA. . High modification rate if DNA reference contains a lot of repeated nucleotides. . Un- blind algorithm also injective mapping and complementary rule should be known by both sender and receiver. . Pure data hiding algorithm. . Security depends on the DNA reference.
8	Ref [37]	2012	. Flexible algorithm. . Easy to implement. . Encrypt secret message by modified playfair algorithm which using DNA and amino acid. . Does not expand the length of DNA after hiding process. . Secret key is used. . Used substitution method in hiding.	. Does not preserve the functionality of the biological DNA. . Un- blind algorithm. . Security depends on the DNA reference.
9	Ref [38]	2012	. Using three keys in the algorithm. . Improved first and third techniques of Ref [23] in term of modification rate. . Does not expand the stego DNA. . Blind algorithm.	. Does not apply any encryption technique. . Hiding secret data only in nucleotides which their marks are equal to zeros after converting to binary.
10	Ref [39]	2013	. Easy to implement. . Low modification rate. . Does not expand the length of stego DNA. . One of the powerful encryption technique (RSA) is used to encrypt secret data before hiding. . Public key is used.	. Does not preserve the functionality of the biological DNA. . Preserve the position of each DNA base which hold the secret data then send to the receiver for extraction. . The size of secret data is increased. . Un- blind algorithm. . Low probability cracking.
11	Ref [40]	2013	. Simple algorithm. . High capacity.	. Does not preserve the functionality of the biological DNA. . Un- blind algorithm. . Pure data hiding algorithm. . Low probability cracking.
12	Ref [11]	2013	. Easy to implement. . Preserve the functionality of the biological DNA. . Low modification rate. . Blind algorithm. . Secret key is hiding in DNA reference which	. Low capacity due of using LSB in hiding process. . No encryption on secret data before hiding. . Low probability cracking.

			provide more security. . Does not expand the DNA reference after hiding confidential data.	
13	Ref [22]	2013	. Encrypt secret data by exhibiting DNA amino acids. . Encrypt the secret key by RSA algorithm before hiding within DNA reference. . High capacity. . public key is used. . High probability cracking.	. Un- blind algorithm. . High modification rate. . Payload not equal to zero. . Does not preserve the functionality of amino acids.
14	Ref [41]	2013	. Preserve the translation of protein in protein coding DNA (PcDNA). . Encoding the data in robust and near to optimal way. . Preserve the codon statistics. . Embedding data achieved near to optimum. . Embedded data in DNA in an efficient and robust way. . Secret key is used.	. Complex calculation. . Intruders can estimate the key for unconstrained ncDNA hiding.
15	Ref [42]	2013	. Encrypt secret data before hiding by playfair algorithm. . High security. . Since hiding at last step in an audio will not attract attackers. . Hide secret data and interpret into audio file so it is not easy to prove both data are existing inside audio. . Provide double hiding layers. . Secret key is used.	. Multiple data need to extract the secret data. . Un- blind algorithm.
16	Ref [43]	2013	. The key space is big enough to resist brute force passive intruders. . Encrypt the secret data before hiding in host document. . Blind algorithm. . The ratio of embedding capacity is 100%. . Provide double hiding layers. . DNA reference is construct by Chebyshev maps. . Substitution method is used in hiding.	. Complex calculation.
17	Ref [44]	2013	. Modified the secret key of Ref [41] algorithm to incorporate the secret key. . As well as retaining all strong points of Ref [41].	. Complex calculation. . Pure data hiding algorithm.
18	Ref [35]	2014	. Increased the capacity and security of the original substitution technique. . Blind algorithm. . Improved substitution method.	. Does not preserve the functionality of the biological DNA. . Increased the size of secret binary in case of multiplying by 6 if not equal zero will add extra zeros. . Expanding the length of stego DNA. . Pure data hiding algorithm.
19	Ref [45]	2014	. High capacity. . Simple algorithm. . Sending secret data as (ABCD) format.	. Un- blind algorithm. . Random DNA sequence and selected complementary pair rule should send to the receiver. . No encryption on data before embedding.

				. Low probability cracking. . Pure data hiding algorithm.
20	Ref [46]	2014	. Blind algorithm. . Sending only the integer value of stego DNA to receiver. . High security. . It is hard for intruders to know the generated random number seeds. . It is hard for intruders to know how many packets are divided, also how many message bits and DNA binary in a packet. . Randomly merged secret message bits and DNA reference bits. . Secret key is used. . High probability cracking.	. Increased the redundancy. . Increased the message size. . Does not preserve the DNA functionality. . Expand the length of stego DNA.
21	Ref [47]	2014	. Encrypt secret message before hiding using RC4. . High capacity. . Providing good security. . Provide double hiding layers. . Construct DNA from image. . Secret key is used.	. The algorithm needs multiple keys during extraction process.
22	Ref [48]	2014	. Security of secret data is improved in better extent. . Unequal size of extra grids can be used to store additional data. . Encrypt secret data based upon BASE64 encoding. . Provide double hiding layers. . DNA is construct from secret text. . Secret key is used.	----------
23	Ref [49]	2014	. High security. . High capacity. . The possible prime range is between 420- 440 because key of prime length is between 20-40. . Increased payload amount and decreased the image distortion. . RC4 encryption is used to encrypt confidential data before hiding. . Provide double hiding layers. . Construct DNA from cover image. . Secret key is used.	. The algorithm has two parts of extraction header and data extractions.
24	Ref [50]	2014	. Provide double layers of security. . Encrypt secret data by AES-128. . Good amount of security has provided by AES. . Applied different operations on confidential data before and after encrypting such as XOR and HASH-512 operations. . To enhance the security DNA is embedded within microdot. . Secret key is used.	. Multiple data needs during extraction. . Does not preserve the DNA functionality.
25	Ref [33]	2015	. Low modification rate. . No expand of DNA reference after embedding confidential data. . Uses double DNA references.	. Replacement rules should be send to receiver. . Plain text must be containing only (capital, letters, small letters, 0, …, 9, period, and dot). It cannot contain other punctuation marks.

			. Preserved the functionality of the original DNA reference. . Blind algorithm. . Does not change in the non-labelled nucleotides. . High capacity. . Encrypt plain text before embedding. . High probability cracking.	. Pure data hiding algorithm.
26	Ref [34]	2015	. Uses three DNA reference in the proposed algorithm. . Blind algorithm. . Encrypt the plain text before hiding. . Secret key is used. . High probability cracking.	. High modification rate. . Does not preserve the functionality of the biological DNA.
27	Ref [51]	2015	. Easy to implement using any programming language. . Create random codon table to convert secret message to DNA format. . High redundancy because of using insertion technique.	. No encryption. . Does not Preserving the information of organism's life. . Expand the length of DNA reference after embedding. . Un- blind algorithm. . Pure data hiding algorithm.
28	Ref [32]	2015	. Low modification rate. . Preserved the functionality of the original DNA reference. . Blind algorithm. . Using playfire algorithm to encrypt the secret message. . No expansion in the DNA reference after hiding secret data. . Substitution method is used in hiding. . Secret key is used.	. If the DNA reference contains many repeated bases the modification rate will be high. . Low probability cracking.
29	Ref [52]	2015	. Encrypt secret message by modified playfair algorithm. . Does not expand the length of DNA after hiding process. . High capacity. . Simple, low execution time, and better performance than Ref [37]. . Enhanced the hiding process of Ref [37]. . Substitution method is used in hiding. . Secret key is used.	. Does not preserve the functionality of the biological DNA. . Un- blind algorithm. . Probability cracking is very low.
30	Ref [53]	2015	. Imperceptible technique. . Encrypt secret message before hiding. . Provide double hiding layers.	. Un- blind algorithm. . Hiding DNA message in only one component of cover image. . Pure data hiding algorithm.
31	Ref [54]	2015	. Hiding two secret images within image without distortion. . Provide double hiding layers. . Secret key is used.	. No encryption technique applied on secret data. . Un- blind algorithm.
32	Ref [55]	2015	. Improved better security. . Improved double carrier by reducing noise bits in image. . Provide a reasonable capacity. . Using several parameters of two dimensional 2D logistic map.	. Require multiple data during the whole embedding and extraction Processes.

			. RC4 used to encrypt secret data. . Provide double hiding layers. . DNA is constructed from image. . Secret key is used. . Substitution method is used in hiding.	
33	Ref [56]	2015	. Imperceptible technique. . An efficient technique. . Providing security by hiding in random video frame. . Provide double hiding layers.	. Un- blind technique. . Multiple data need in extraction process. . Pure data hiding algorithm.
34	Ref [57]	2015	. Highly secure and efficient method. . Encrypt secret data using RSA algorithm before embedding. . Blind algorithm. . Provide double hiding layers. . Public key is used.	. Does not preserve the functionality of the biological DNA.
35	Ref [58]	2016	. Preserving the information of organism's life. . Blind algorithm. . High capacity. . Does not expand the length of stego DNA. . XOR and PRBG is used to encrypt the secret data. . Errors derived and corrected by Reed-Solomon (RS) code. . Secret key is used. . High probability cracking.	. Not easy to implement. . High modification rate.
36	Ref [59]	2016	. Any type and any size of the secret data and the key can be used. . Applying different encryption techniques and analysing them to select the good one before hiding. . Normal key is used to select English letters to generate the playfair cipher grid which can be more secure. . High hiding capacity. . No redundancy in the process. . More simplicity. . Good performance with low time execution. . Substitution method is used in hiding.	. Un- blind algorithm. . Does not preserve the amino acid functionality. . High modification rate. . Low probability cracking.
37	Ref [60]	2016	. Using the vigenere notion and a complementary pair rule to construct a table for hiding secret data. . Simple algorithm. . Secret key is used.	. Un- blind algorithm. . Does not preserve the amino acid functionality.
38	Ref [61]	2016	. Blind algorithm. . Encrypting secret message before hiding using playfair. . High embedding capacity. . High robustness method. . Secret key is used. . High probability cracking.	. Does not preserve the amino acid functionality. . Payload not equal zero. . High modification rate. . High redundancy.
39	Ref [62]	2016	. High payload capacity. . It is a fragile reversible technique. . Transparency is very high. . The reliability of secret information is	----------

			provided. . High quality. . DNA is construct from secret text. . Provide double hiding layers. . Secret key is used.	
40	Ref [63]	2017	. Encrypt secret data by vigener or playfair cipher. . Double amount of data hiding. . High security. . After hiding DNA reference will be send in a microdot in a Paper before sending to the receiver. . Regenerate another key and DNA sequence then the process of hiding will happen again if the paper is contaminated. . Preserve the functionality of DNA sequence and avoid any mutations.	. Un- blind algorithm. . Sending multiple data to the receiver for extraction process. . High modification rate in non-coding region.
41	Ref [64]	2017	. High security. . Double random key generator. . Secret key is used. . Probability cracking is very high.	. Un- blind algorithm. . Does not preserve the functionality of DNA. . Payload not equal to zero.

The conclusion behind the derived comparison in Table II is to explain the strength and weakness points of the proposed DNA based on data hiding algorithms. The strength is fostered by encrypting the confidential data into cipher data before embedding instead of embedding the original data format thus providing more security as achieved by [13,18, 22, 32, 33, 34, 35, 37, 39, 42, 47, 48, 50, 52, 55, 57, 58, 59, 60, 61, 63]. The most promising encryption technique to combine with DNA based data hiding technique is playfair technique as it is approved in [59]. In their work a comprehensive comparative of some encryption techniques has been done which are vigenere and playfair, AES, and RSA ciphers. Each of them combined with a data hiding in DNA based on substitution method. The results showed that the playfair cipher is not only fast and simple, but it can also provide more security and large capacity.

The main feature which supported by DNA based data hiding technique is the blindness property which does not need to send the original DNA reference to the receiver. The main objective behind the blindness property is to maximize the security degree as possible and to prevent any way can be discovered by an attacker to be obtained through minimizing the required data that's sent to the receiver as much as possible as it is achieved in [11, 18, 24, 32, 33, 34, 35, 38, 43, 46, 57, 58, 61].

Biologically conserving the original functionality of the DNA reference after embedding process is considered as one of the strength points while maintaining a reasonable data payload. DNA reference could be used to hide secret data in such a way that it is functionality of producing proteins is not affected. Exploiting some of DNA properties such as silent mutation and codon redundancy can be used to hide data and change the genetic sequence without changing the protein chain it codes for, as achieved in [11, 21, 24, 32, 33, 41, 52, 58, 63].

In most data hiding algorithms, the carrier will face some distortion after hiding process. Techniques of hiding data are concerned with embedding the existence of the embedded data that's why it can be unobtrusively communicated. Thus, carrier distortion is very important to minimize. In the DNA sequence, the length of the DNA sequence and the extent of modification are considered as the measures of the quality of a stego DNA once data has been embedded. A low modification rate and no expansion rate result in a better quality stego DNA which draws less attention from potential attackers. A low modification rate is achieved in [11, 31, 32, 33, 38, 39]. Furthermore, no expansion rate attribute of stego DNA which means payload is equal to zero is achieved in [11, 13, 21, 31, 32, 33, 37, 38, 39, 58].

Two phase data hiding approach is proposed to hide secret data in more depth than the original data hiding approaches. Using two different carriers in the same approach will provide more security and more difficulties to understand or retrieve the secret data by attackers. Several approaches used DNA reference with another multimedia carrier to hide secret data as achieved in [42, 43, 47, 48, 49, 54, 55, 56, 57, 62]. Some of them constructed the DNA from a cover image or from secret data as in [43, 47, 48, 49, 55], while some others either carried out a random selection or selected from the online database as in [42, 54, 56, 57, 62].

One of the most essential elements in data hiding approaches is the key. Data hiding systems can be divided into three types based on the key used. The first type is called pure data hiding which is less secure because it does not use any type of key as in [21, 23, 31, 33, 35, 36, 40, 45, 51, 53, 56]. Thus, to provide more security for the system it is important to use a key which

makes an attack on the data hiding system much more difficult because even if the attackers identify the data hiding scheme used, they still cannot extract the secret information from the carrier because they do not have the key. Only the sender and the receiver have the key. So, in data hiding, it is preferable to use a strong key thereby ensuring a more secure system. The second type is called secret key as achieved in [11, 13, 18, 24, 32, 34, 37, 38, 41, 42, 43, 44, 46, 47, 48, 49, 50, 52, 54, 55, 58, 59, 60, 61, 62, 63, 64]. The third type is called public key as achieved in [22, 39, 57]. In general, the public key is more secure but slower than secret key.

The algorithm's cracking probability is the probability of breaking the algorithm and obtaining the hidden confidential data. The reason for analyzing the probability of the scheme being cracked is so as to distinguish the factors which ensure a minimum cracking probability. The cracking probability is not based on the number of attempts made before an intruder gains access to the secret hidden data, but is based on the inclusion of every unknown variable applied in the algorithm so as to conceal the confidential data. Therefore, high probability cracking leads for high security for the data hiding approach as achieved in [18, 22, 33, 34, 46, 58, 61, 64].

The substitution method is considered as a more efficient method for hiding data in DNA. This method can preserve the length of the DNA sequence, always keeping the payload at zero. Also, it has a more efficient capacity because it substitutes some of DNA nucleotides with secret data bits or other nucleotides based on the secret data as achieved in [31, 32, 35, 36, 37, 43, 52, 55, 59].

In any data hiding technique, capacity is a critical part which considered as a one of the main requirements of data hiding techniques. It is imperative that a data hiding technique possesses a sizable hiding capacity. This capacity can be described in terms of absolute measures such as the size of the secret message or using a relative value relative (for instance, data embedding rate, bits per pixel, bits per non-zero discrete cosine transform coefficient, or the ratio of the secret message to the cover-medium, etc.). The DNA capacity can be measured in terms of bit per nucleotide (pbn). Thus, improving the capacity of the hidden data is one of the major concern for researchers in this field, which has previously been achieved in [13, 18, 21, 22, 31, 33, 35, 36, 40, 45, 47, 49, 52, 55, 58, 59, 62, 63].

Thus, it can be concluded that the main objective for proposed DNA based double layer hiding algorithms, double layer security which are encrypt confidential data then to hide high capacity, blind, biologically preserved, low modification rate, a zero-payload algorithm, not a pure system, and with high probability cracking. In [32, 33, and 58], a low modification rate, conserving the expansion length of DNA reference, blindness, conserving the DNA functionality, double layer of security, high capacity, and the not a pure algorithm are proposed.

CONCLUSION

A surge in the demand for storage has created a tremendous demand for the development of new and emerging techniques for powerful data storage. Presently, DNA has been identified as an effective data carrier which also has the added advantage of reliable data storage. The bio-molecular computational abilities of DNA are being applied in cryptography and data hiding. This paper lays out a comparison of several recent DNA based data hiding algorithms highlighting their security issues. The pros and cons of each algorithm are also laid out in. Some critical issues are discussed in term of probability cracking, double layer of security, single and double hiding layers, blindness, biologically preserved DNA, modification rate, expansion of DNA reference, not pure algorithm, applying substitution method, and capacity also. The objective of the comparison provided in this study is to equip the researchers with the knowledge to conduct future research on improved secure DNA data hiding techniques which are both efficient and reliable.

ACKNOWLEDGEMENT

The authors greatly acknowledge the Research Management Centre (RMC) Universiti Teknologi Malaysia and Ministry of Higher Education (MOHE) for the financial support through the Fundamental Research Grant Scheme (FRGS) No. R.J130000.7828.4F860.

REFERENCES

[1] Singh, G., *A study of encryption algorithms (RSA, DES, 3DES and AES) for information security.* International Journal of Computer Applications, 2013. 67(19).

[2] Subhedar, M.S. and V.H. Mankar, *Current status and key issues in image steganography: A survey.* Computer science review, 2014. 13: p. 95-113.

[3] Hamed, G., et al. *Comparative study for various DNA based steganography techniques with the essential conclusions about the future research*. in *Computer Engineering & Systems (ICCES), 2016 11th International Conference on*. 2016. IEEE.

[4] Amin, M.M., et al. *Information hiding using steganography*. in *Telecommunication Technology, 2003. NCTT 2003 Proceedings. 4th National Conference on*. 2003. IEEE.

[5] Al-Mohammad, A., *Steganography-based secret and reliable communications: Improving steganographic capacity and imperceptibility*. 2010, Brunel University, School of Information Systems, Computing and Mathematics Theses.

[6] Santoso, K.N., et al., *Information Hiding in*

Noncoding DNA for DNA Steganography. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, 2015. 98(7): p. 1529-1536.

[7] Kumari, P. and R. Kapoor, *Image Steganography for Data Embedding & Extraction using LSB Technique.* International Journal of Computer Applications & Information Technology, 2016. 9(2): p. 192.

[8] Ashok, J., et al., *Steganography: an overview.* International Journal of Engineering Science and Technology, 2010. 2(10): p. 5985-5992.

[9] Nickfarjam, A.M. and Z. Azimifar. *Image steganography based on pixel ranking and Particle Swarm Optimization*. in *Artificial Intelligence and Signal Processing (AISP), 2012 16th CSI International Symposium on*. 2012. IEEE.

[10] Sheelu, A., B., *An Overview of Steganography.* IOSR Journal of Computer Engineering (IOSR-JCE), 2013. 11(1): p. 15-19.

[11] Khalifa, A. *LSBase: A key encapsulation scheme to improve hybrid crypto-systems using DNA steganography*. in *Computer Engineering & Systems (ICCES), 2013 8th International Conference on*. 2013. IEEE.

[12] Jain, S. and V. Bhatnagar. *Analogy of various DNA based security algorithms using cryptography and steganography*. in *Issues and Challenges in Intelligent Computing Techniques (ICICT), 2014 International Conference on*. 2014. IEEE.

[13] Torkaman, M.R.N., N.S. Kazazi, and A. Rouddini, *Innovative approach to improve hybrid cryptography by using DNA steganography.* International Journal of New Computer Architectures and their Applications (IJNCAA), 2012. 2(1): p. 224-235.

[14] Bansod, S. and G. Bhure, *Data encryption by image steganography.* Int. J. Inform. Comput. Technol. Int. Res. Publ. House, 2014. 4: p. 453-458.

[15] Singh, K.U., *Video steganography: text hiding in video by LSB substitution.* International Journal of Engineering Research and Applications, 2014. 4(5): p. 105-108.

[16] Chandel, B., Jain, S., *Video Steganography: A Survey.* IOSR Journal of Computer Engineering (IOSR-JCE) 2016. 18(1): p. 11-17.

[17] Yang, Y., *Information analysis for steganography and steganalysis in 3D polygonal meshes*. 2013, Durham University.

[18] Atito, A., A. Khalifa, and S. Rida, *DNA-based data encryption and hiding using playfair and insertion techniques.* Journal of Communications and Computer Engineering, 2012. 2(3): p. 44.

[19] AL-WATTAR, A.H.S., MAHMOD, R., ZUKARNAIN, Z. A., UDZIR, N.,, *REVIEW OF DNA AND PSEUDO DNA CRYPTOGRAPHY.* International Journal of Computer Science and Engineering (IJCSE), 2015. 4(4): p. 65-76.

[20] Tornea, O., *Contributions to DNA cryptography: applications to text and image secure transmission*. 2013, Université Nice Sophia Antipolis.

[21] Abbasy, M.R., et al., *DNA base data hiding algorithm.* International Journal of New Computer Architectures and their Applications (IJNCAA), 2012. 2(1): p. 183-192.

[22] Skariya, M. and M. Varghese, *Enhanced double layer security using RSA over DNA based data encryption system.* International Journal of Computer Science& Engineering Technology (IJCSET), 2013. 4(06): p. 746-750.

[23] Shiu, H., et al., *Data hiding methods based upon DNA sequences.* Information Sciences, 2010. 180(11): p. 2196-2208.

[24] Mousa, H., et al., *Data hiding based on contrast mapping using DNA medium.* Int. Arab J. Inf. Technol., 2011. 8(2): p. 147-154.

[25] Adleman, L.M., *Molecular computation of solutions to combinatorial problems.* Nature, 1994. 369: p. 40.

[26] OGIWARA, M. *Simulating Boolean Circuits on DNA Computers*. in *Proceedings of the 1st International Conference on Computational Molecular Biology, 1997*. 1997. ACM Press.

[27] Clelland, C.T., V. Risca, and C. Bancroft, *Hiding messages in DNA microdots.* Nature, 1999. 399(6736): p. 533-534.

[28] Sureshraj, D. and V.M. Bhaskaran, *Automatic DNA sequence generation for secured cost-effective multi-cloud storage.* 2012.

[29] Singh, A. and R. Singh. *Information hiding techniques based on DNA inconsistency: An overview*. in *Computing for Sustainable Global Development (INDIACom), 2015 2nd International Conference on*. 2015. IEEE.

[30] Bhateja, A. and K. Mittal, *DNA Steganography: Literature Survey on its Viability as a Novel Cryptosystem.* Journal of Computer Science and Engineering, 2015. 2(1): p. 8-14.

[31] Guo, C., C.-C. Chang, and Z.-H. Wang, *A new data hiding scheme based on DNA sequence.* Int. J. Innov. Comput. Inf. Control, 2012. 8(1): p. 139-149.

[32] Hamed, G., et al. *Hybrid technique for*

steganography-based on DNA with n-bits binary coding rule. in *Soft Computing and Pattern Recognition (SoCPaR), 2015 7th International Conference of*. 2015. IEEE.

[33] 33 Ibrahim, F.E., H. Abdalkader, and M. Moussa. *Enhancing the Security of Data Hiding Using Double DNA Sequences*. in *Industry Academia Collaboration Conference (IAC)*.

[34] El-Latif, E.I.A. and M.I. Moussa, *Chaotic Information-hiding Algorithm based on DNA.* International Journal of Computer Applications (0975–8887) Volume, 2015. 122(10).

[35] Agrawal, R., M. Srivastava, and A. Sharma. *Data hiding using dictionary based substitution method in DNA sequences*. in *Industrial and Information Systems (ICIIS), 2014 9th International Conference on*. 2014. IEEE.

[36] Taur, J.-S., et al., *Data hiding in DNA sequences based on table lookup substitution.* International Journal of Innovative Computing, Information and Control, 2012. 8(10): p. 6585-6598.

[37] Khalifa, A. and A. Atito. *High-capacity DNA-based steganography*. in *Informatics and Systems (INFOS), 2012 8th International Conference on*. 2012. IEEE.

[38] Huang, Y.-H., C.-C. Chang, and C.-Y. Wu, *A DNA-based data hiding technique with low modification rates.* Multimedia tools and applications, 2014. 70(3): p. 1439-1451.

[39] Mitras, B.A. and A. Abo, *Proposed steganography approach using DNA properties.* International Journal of Information Technology and Business Management, 2013. 14(1): p. 96-102.

[40] Bhattacharyya, D. and S.K. Bandyopadhyay, *Hiding secret data in dna sequence.* International Journal of Scientific &Engineering Research, 2013. 4(2).

[41] Haughton, D. and F. Balado, *Biocode: Two biologically compatible algorithms for embedding data in non-coding and coding regions of dna.* BMC bioinformatics, 2013. 14(1): p. 121.

[42] Shyamasree, C. and S. Anees. *Highly secure DNA-based audio steganography*. in *Recent Trends in Information Technology (ICRTIT), 2013 International Conference on*. 2013. IEEE.

[43] Liu, H., D. Lin, and A. Kadir, *A novel data hiding method based on deoxyribonucleic acid coding.* Computers & Electrical Engineering, 2013. 39(4): p. 1164-1173.

[44] Haughton, D. and F. Balado. *Security study of keyed DNA data embedding*. in *Global Conference on Signal and Information Processing (GlobalSIP), 2013 IEEE*. 2013. IEEE.

[45] Menaka, K. *Message encryption using DNA sequences*. in *Computing and Communication Technologies (WCCCT), 2014 World Congress on*. 2014. IEEE.

[46] Manna, S., et al. *Modified technique of insertion methods for data hiding using DNA sequences*. in *Automation, Control, Energy and Systems (ACES), 2014 First International Conference on*. 2014. IEEE.

[47] Das, P. and N. Kar. *A DNA based image steganography using 2d chaotic map*. in *Electronics and Communication Systems (ICECS), 2014 International Conference on*. 2014. IEEE.

[48] Majumdar, A., M. Sharma, and N. Kar, *An Improved Approach to Steganography using DNA Characteristics.* 2014: p. 138-145.

[49] Das, P. and N. Kar. *A highly secure DNA based image steganography*. in *Green Computing Communication and Electrical Engineering (ICGCCEE), 2014 International Conference on*. 2014. IEEE.

[50] Chaudhary, H. and V. Bhatnagar. *Hybrid approach for secure communication of data using chemical DNA*. in *Confluence the Next Generation Information Technology Summit (Confluence), 2014 5th International Conference-*. 2014. IEEE.

[51] Yamuna, M., Elakkiya, A., *Codons in Data Safe Transfer.* International Journal of Engineering Issues, 2015(2): p. 85-90.

[52] Marwan, S., A. Shawish, and K. Nagaty. *An Enhanced DNA-based Steganography Technique with a Higher Hiding Capacity*. in *BIOINFORMATICS*. 2015.

[53] Manisha, B., P., Mohit, *DOUBLE LAYERED DNA BASED CRYPTOGRAPHY.* IJRET: International Journal of Research in Engineering and Technology, 2015. 4(4): p. 2321-7308.

[54] Chakraborty, S., Bandyopadhyay, K. S., *Data Hiding by Image Steganography Appling DNA Sequence Arithmetic.* International Journal of Advanced Information Science and Technology (IJAIST) 2015. 44(44).

[55] Das, P., et al., *An Improved DNA based dual cover steganography.* Procedia Computer Science, 2015. 46: p. 604-611.

[56] Indora, S. *Cascaded DNA cryptography and steganography*. in *Green Computing and Internet of Things (ICGCIoT), 2015 International Conference on*. 2015. IEEE.

[57] Tank, R.M., H.D. Vasava, and V. Agrawal, *DNA-Based Audio Steganography.* Oriental journal of Computer Science and Technology, 2015. 8: p. 43-48.

[58] Santoso, K., et al., *Sector- based DNA information hiding method.* Security and Communication Networks, 2016. 9(17): p. 4210-4226.

[59] Marwan, S., A. Shawish, and K. Nagaty, *DNA-based cryptographic methods for data hiding in DNA media.* Biosystems, 2016. 150: p. 110-118.

[60] Meena, K., Menaka, K., Subramanian, R. K., *Secure Message Transfer Using DNA Sequences.* Journal of Computational Intelligence in Bioinformatics, 2016. 9(1): p. 1-6.

[61] Khalifa, A., A. Elhadad, and S. Hamad, *Secure Blind Data Hiding into Pseudo DNA Sequences Using Playfair Ciphering and Generic Complementary Substitution.* Appl. Math, 2016. 10(4): p. 1483-1492.

[62] Tuncer, T. and E. Avci, *A reversible data hiding algorithm based on probabilistic DNA-XOR secret sharing scheme for color images.* Displays, 2016. 41: p. 1-8.

[63] Marwan, S., A. Shawish, and K. Nagaty, *Utilizing DNA Strands for Secured Data-Hiding with High Capacity.* International Journal of Interactive Mobile Technologies, 2017. 11(2).

[64] Malathi, P., et al., *Highly Improved DNA Based Steganography.* Procedia Computer Science, 2017. 115: p. 651-659.