



An approach of image steganography by combine application of DNA sequence and arithmetic encoding

Suman Chakraborty¹, Prof. Samir K. Bandyopadhyay²

¹B.P.Poddar Institute of Management & Technology,
137, VIP Road, Kolkata-700081, India.

Sumanc75@gmail.com.

²Department of Computer Science and Engineer, University of Calcutta,
92 A.P.C Road, Kolkata-700081.

Skb1@vsnl.com.

ABSTRACT

In this paper an attempt has been made to hide an image with in a cover image by using DNA sequence and Arithmetic encoding. DNA sequence used to represent an image and arithmetic encoding convert this DNA sequence in decimal form within the range of 0 to 1. In this procedure encoding has been done in 2 level, 1st Secret-Image converted into decimal form and 2nd decimal form to DNA sequence. So, in this approach secret information is hidden in more depth. Increase depth provides more degree of security from external attack.

Keywords

DNA sequence, mRNA, Cover-Image, Secret-Image, Arithmetic Encoding, Nucleotide.



Council for Innovative Research

Peer Review Research Publishing System

Journal: [International Journal of Management & Information Technology](#)

Vol. 5, No. 3

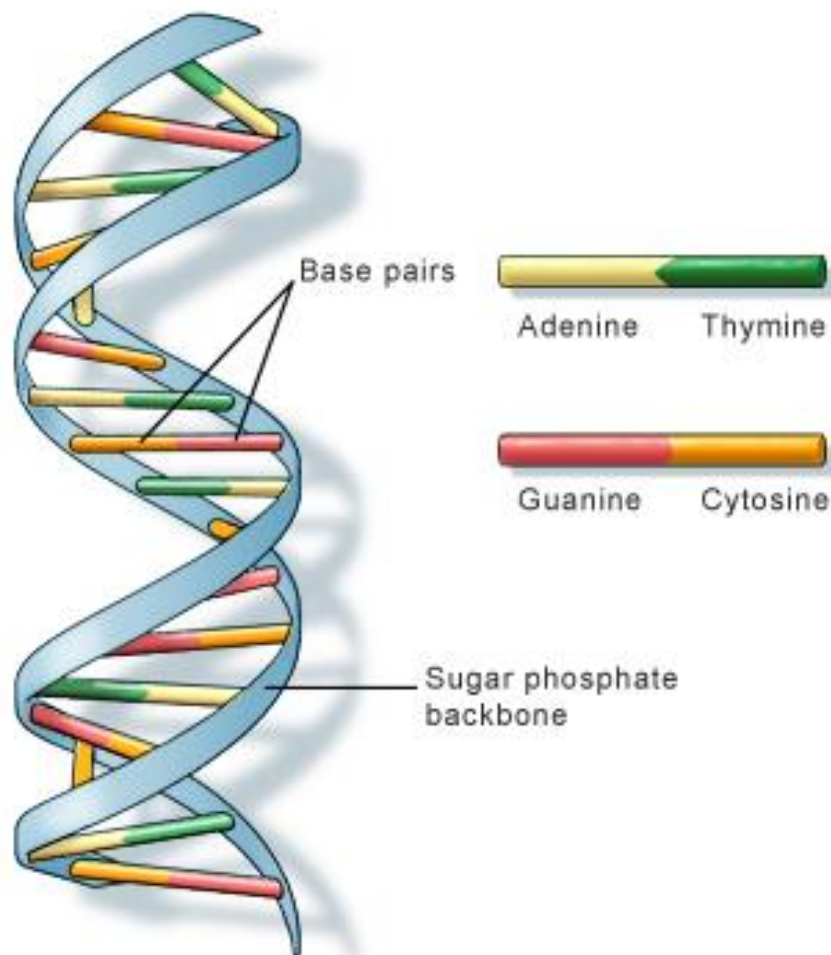
editor@cirworld.com

www.cirworld.com, member.cirworld.com

1. INTRODUCTION

Steganography is the art of hiding information imperceptibly in a cover medium. The word "Steganography" is of Greek origin and means "covered or hidden writing". The main aim in steganography is to hide the very existence of the message in the cover medium.[1]

The information in DNA is stored as a code made up of four chemical bases: adenine (A), guanine (G), cytosine (C), and thymine (T). Human DNA consists of about 3 billion bases, and more than 99 percent of those bases are the same in all people. The order, or sequence, of these bases determines the information available for building and maintaining an organism, similar to the way in which letters of the alphabet appear in a certain order to form words and sentences. DNA bases pair up with each other, A with T and C with G, to form units called base pairs. Each base is also attached to a sugar molecule and a phosphate molecule. Together, a base, sugar, and phosphate are called a nucleotide. Nucleotides are arranged in two long strands that form a spiral called a double helix. The structure of the double helix is somewhat like a ladder, with the base pairs forming the ladder's rungs and the sugar and phosphate molecules forming the vertical sidepieces of the ladder. An important property of DNA is that it can replicate, or make copies of itself. Each strand of DNA in the double helix can serve as a pattern for duplicating the sequence of bases. This is critical when cells divide because each new cell needs to have an exact copy of the DNA present in the old cell. This is shown in figure 1.



U.S. National Library of Medicine

Figure-1 (Figure is taken from U.S. National Library of Medicine)

mRNA is transcribed from DNA, carrying information for protein synthesis. Three consecutive nucleotides in mRNA encode an amino acid or a stop signal for protein synthesis. The trinucleotide is known as a codon. The sequence of mRNA is complementary to DNA's template strand, and thus the same as DNA's coding strand, except that T is replaced by U.[2-5]



Table-1.mRNA to amino acid mapping

UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA Stop UAG Stop	UGU } Cys UGC } UGA Stop UGG Trp	U C A G
CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }	U C A G
AUU } AUC } Ile AUA } AUG Met	ACU } ACC } Trp ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	U C A G
GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }	U C A G

Third letter

In this article, we proposed a new image hiding scheme where Secret-image hiding in more depth in the form of DNA sequence. DNA sequence is unfamiliar concept with the respect of seganography. This makes staganography process more secure.

The remainder of the paper is organized as follows. In Section 2 Related work is given. In Section 3, flow diagram is presented. In Section 4, proposed algorithms are given. Section 5 illustrates Example. In Section 6 conclusion is expressed in terms of future activities.

2. RELATED WORK

Arithmetic coding is a form of entropy encoding used in lossless data compression. Normally, a string of characters such as the words "hello there" is represented using a fixed number of bits per character, as in the ASCII code. When a string is converted to arithmetic encoding, frequently used characters will be stored with fewer bits and not-so-frequently occurring characters will be stored with more bits, resulting in fewer bits used in total. Arithmetic coding differs from other forms of entropy encoding such as Huffman coding in that rather than separating the input into component symbols and replacing each with a code, arithmetic coding encodes the entire message into a single number, a fraction n where $(0.0 \leq n < 1.0)$. [6-7]

For example, if we are going to encode the random message "BILL GATES", we would have a probability distribution that looks like this:



Character	Probability
-----	-----
SPACE	1/10
A	1/10
B	1/10
E	1/10
G	1/10
I	1/10
L	2/10
S	1/10
T	1/10

Once the character probabilities are known, the individual symbols need to be assigned a range along a probability line, which is nominally 0 to 1. It doesn't matter which characters are assigned to the segment of the range, as long as it is done in the same manner by both the encoder and the decoder. The nine character symbol set use here would look like this:

Character	Probability	Range
-----	-----	-----
SPACE	1/10	0.00 - 0.10
A	1/10	0.10 - 0.20
B	1/10	0.20 - 0.30
E	1/10	0.30 - 0.40
G	1/10	0.40 - 0.50
I	1/10	0.50 - 0.60
L	2/10	0.60 - 0.80
S	1/10	0.80 - 0.90
T	1/10	0.90 - 1.00

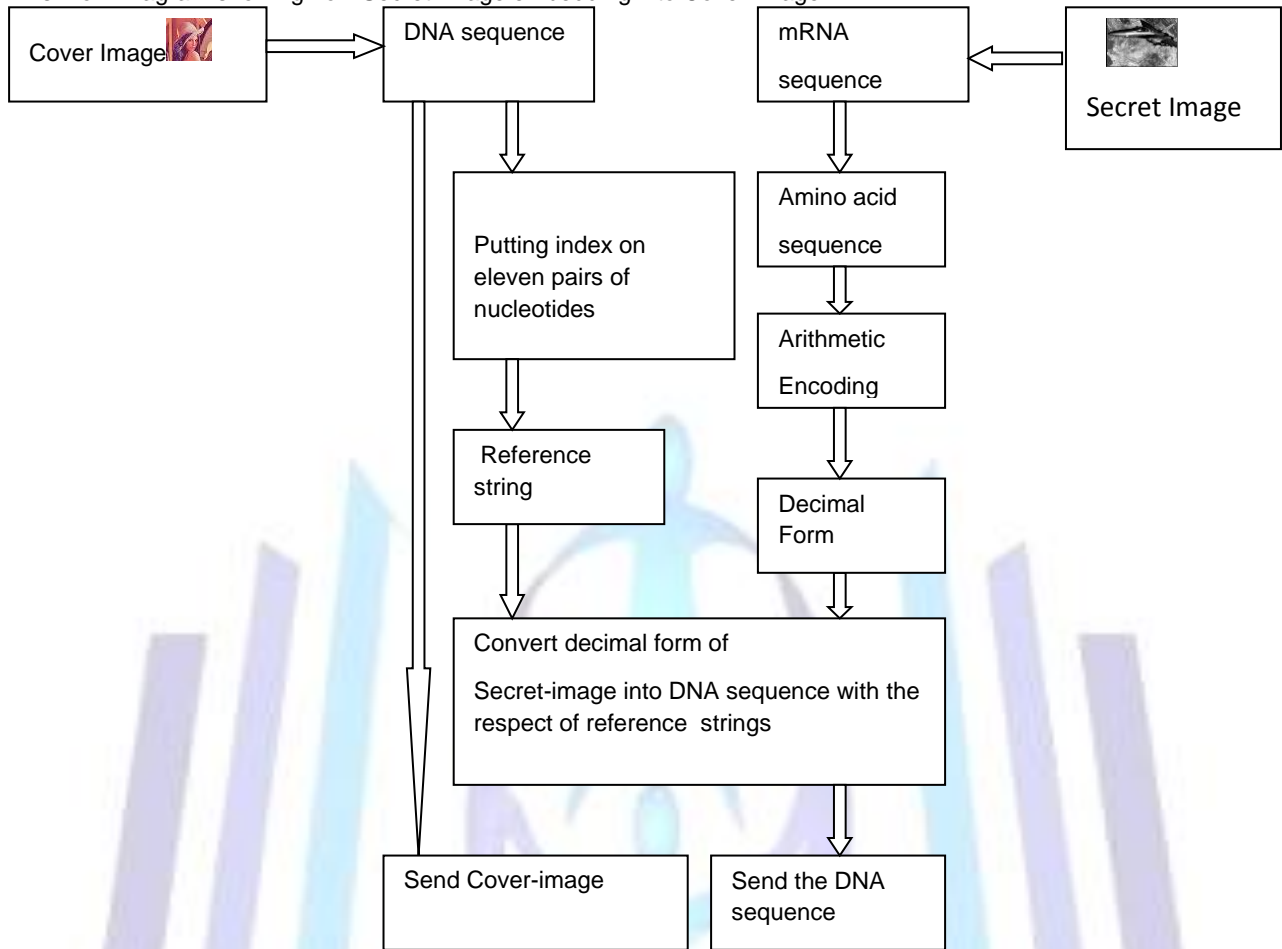
Each character is assigned the portion of the 0-1 range that corresponds to its probability of appearance. Note also that the character "owns" everything up to, but not including the

higher number. So the letter 'T' in fact has the range 0.90 – 0.9999....The most significant portion of an arithmetic coded message belongs to the first symbol to be encoded. When encoding the message "BILL GATES", the first symbol is "B". In order for the first character to be decoded properly, the final coded message has to be a number greater than or equal to 0.20 and less than 0.30. What we do to encode this number is keep track of the range that this number could fall in. So after the first character is encoded, the low end for this range is 0.20 and the high end of the range is 0.30. After the first character is encoded, we know that our range for our output number is now bounded by the low number and the high number. What happens during the rest of the encoding process is that each new symbol to be encoded will further restrict the possible range of the output number. The next character to be encoded, 'I', owns the range 0.50 through 0.60. If it was the first number in our message, we would set our low and high range values directly to those values. But 'I' is the second character. So what we do instead is say that 'I' owns the range that corresponds to 0.50-0.60 in the new sub range of 0.2 – 0.3. This means that the new encoded number will have to fall somewhere in the 50th to 60th percentile of the currently established range. Applying this logic will further restrict our number to the range 0.25 to 0.26.

3. FLOW DIAGRAM

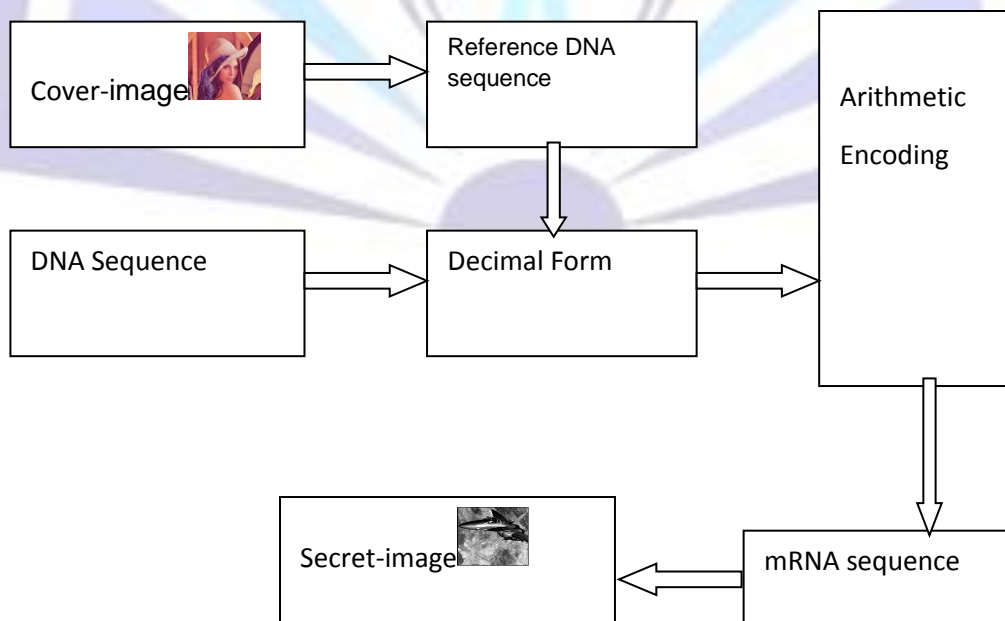
3.1 Staganography Process

This Flow-Diagram showing how Secret-Image embedding into Cover-image-



3.2 Extraction Process

Flowing pictorial representation showing decoding procedure at receiver side-



4. PROPOSED ALGORITHM

4.1 Encoding Algorithm

Step-1: Chose arbitrarily a mRNA sequence of Cover-image. Covert it to corresponding amino acid sequence.

Step-2: Use Arithmetic encoding to get decimal (D) form of Secret-image.

Step-3: Assign a DNA sequence to Cover-image.

Step-4: By applying DNA base pairing rules, makes a DNA sequence of pair nucleotides. For ten decimal value Putting index on only ten pairs. Another pair will indicate decimal point. Because out come of Arithmetic encoding in between 0 to 1. Which pair will contain what index value, it will decide randomly. This is the reference string(R).

Step-5: Covert D into DNA sequence(S) by using R.

Step-6: Send Cover-image and S.

Step-7: End.

4.2 Extraction Algorithm

Step1: Received Cover-image and DNA sequence(S).

Step-2: Convert S into decimal form (D) by R.

Step-3: Find out equivalent mRNA sequence by using Arithmetic encoding.

Step-4: End.

5. EXAMPLE



Secret-image



Cover-image

For the above Secret- image corresponding mRNA sequence is "CUU CCG UGC GAU GUA GCC GGU AUC UUU GGA CAU UGG UAU AUU UCA UGC" which is chosen arbitrarily.

Converting this mRNA sequence into amino acid sequence with the help of Table-1. The amino acid sequence is " Leu Pro Cys Asp Val Ala Gly Ile Phe Gly His Trp Tyr Ile Ser Cys' .

By using Arithmetic encoding equivalent decimal form (D) is 0.272850788.

DNA sequence of the above Cover-image is ATCGAATTCGCGCTGAGTCACAATTCGCGCTGAGTGAACC.

After indexing reference string(R) is AT **CG**₁ **AA**₂ **TT**₃ **CG**₃ **CG**₄ **CTGA**₇ **GT**₆ **CA**_{CA}₂ **ATTC**₈ **GC**_{GC}₅ **TGAGT**₉ **AACC**₀.

Equivalent DNA string(S) of D is CCCGGACATCGCCCGATCTC.

6. CONCLUSION

Here an attempt has been made to hide an image with in a cover image by using DNA sequence and Arithmetic encoding. DNA sequence used to represent an image and arithmetic encoding convert this DNA sequence in decimal form within the range of 0 to 1. The result is quite encouraging.

REFERENCE 7

- [1] K. Solanki, K. Sullivan, U. Madhow, B. S. Manjunath, and S. Chandrasekaran, "Provably secure steganography: Achieving zero K-L divergence using statistical restoration", in Proc. IEEE International Conference on Image Processing (ICIP06), Atlanta, GA, USA, Oct. 2006.

- [2] Peterson P., "Hiding in DNA," in Proceedings of Muse, pp. 22, 2001.
- [3] Rijmen P., "Advanced Encryption Standard," in Proceedings of Federal Information Processing Standards Publications, National Institute of Standards and Technology, pp. 19-22, 2001.
- [4] Rivest L., Shamir A., and Adleman L., "A Method for Obtaining Digital Signature and Public Key Cryptosystem," Computer Journal of Communications of the ACM, vol. 21, no. 2, pp. 120-126, 1978.
- [5] Saeb M., El-abd E., and El-Zanaty M., "On Covert Data Communication Channels"
- [6] Shimanovsky B., Feng J., and Potkonjak M., Hiding Data in DNA, Springer, UK, 2003.
- [7] Smid E. and Branstad M., "Data Encryption Standard," in Proceedings of Federal Information Processing Standards Publications, National Institute of Standards and Technology, pp. 550-559, 1988
- [8] Prof. Samir Kumar Bandyopadhyay and S Chakraborty, "Image Hiding in DNA Sequence Using Arithmetic Encoding", Journal of Global Research in Computer Science, Volume 2, No.4, April 2011.
- [9] Prof. Samir Kumar Bandyopadhyay and S Chakraborty, "Image Compression using DNA sequence", International Journal of Computer Science & Engineering Technology, Vol 1, Issue 11, December 2011.
- [10] Prof. Samir Kumar Bandyopadhyay and S Chakraborty, "Image Steganography Using DNA Sequence", Asian Journal of Computer Science And Information Technology, Vol 1, No 2, September 2011.

Author' biography



Research Fellow of West Bengal University of Technology, India and Asst. Prof, BP Poddar Institute of Management and Technology . Author of more than Ten publications in International Journal . Field of Specialization Image Processing, Data Structure, Programming Language etc.



Dr. Samir Kumar Bandyopadhyay, B.E., M.Tech., Ph. D (Computer Science & Engineering), C.Engg., D.Engg., FIE, FIETE, Sr. Member IEEE, currently, Professor of Computer Science & Engineering, University of Calcutta, Kolkata, India. Visiting Faculty, Dept. of Comp. Sc., Southern Illinois University, USA, MIT, California Institute of Technology, etc. His research interests include Bio-medical Engg, Mobile Computing, Pattern Recognition, Graph Theory, Software Engg., etc. He has 25 Years of experience at the Post-graduate and under-graduate Teaching & Research experience in the University of Calcutta. He has already got several Academic Distinctions in Degree level/Recognition/Awards from various prestigious Institutes and Organizations. He has published 300 Research papers in International & Indian Journals and 5 leading text books for Computer Science and Engineering. Dr. Bandyopadhyay is the former Registrar of University of Calcutta and West Bengal University of Technology, Kolkata, and former Acting-VC, West Bengal University of Technology, Kolkata, India.