Journal of Engineering and Applied Sciences 14 (Special Issue 8): 10240-10244, 2019
ISSN: 1816-949X

ECC Based Blind Steganography-DNA for Hidden Information

Zena Ahmed, Saher Adil and Saif M. Kh. Al-Alak
Department of Computer Science, College of Science for Women
University of Babylon, Babylon, Iraq

Abstract: The strong need to secure valuable information against unauthorized attacks. Steganography one of the methods that provide a solution against these attacks by hiding the information in digital media such as images, video, sound. Further more, steganography took a step further to hide the information by exploiting DNA characteristics. In this study, proposed an Elliptic Curve (ECC) protocol based blind hiding approach to hide data in a DNA, blind hiding provides a solution to the problem that the sender and the receiver have to secretly communicate both the stego-DNA and the reference sequence. The proposed protocol implemented in two stages: the first stage, encrypt the message using play fairs. Second stage, hide the encrypted message into DNA reference in a random location using ECC and over lapping the stego-DNA with the reference cover. In this way, the extraction process can be done blindly and the communicating parties do not really have to trade anything in progress. The proposed strategy appeared an extraordinary execution giving high security and more speed of time performance.

Key words: Information hiding, DNA, blind extraction, random location, communicate, receiver

INTRODUCTION

The valuable data must keep safe from attacker; the term steganography is the science and art for hiding message inside a cover media (Cheddad *et al*., 2010). Furthermore, it includes cover and message to be hidden inside the cover, many types of digital media are considered as a cover like image (Cheddad *et al*., 2010), video (Cheddad and Lewis, 2010), audio (Yan *et al*., 2012), network (Murdoch and Lewis, 2005), even text (Anderson *et al*., 1998). In fact, before steganography information must encrypted, cryptography is a science to protect information from eavesdrops by ciphering in various technique (Cherian *et al*., 2013), many type like Symmetric key (Abbasy and Shanmugam, 2011), asymmetric key (Lu *et al*., 2007). Khalifa *et al*. (2016) proposed a protocol. That hide information inside the DNA media by Generic Complementary Based Substation (GCBS) in first stage, in second stage the hiding is done by insertion method, the algorithm most first finds cipher message and hide by GCBS, second stage is sent random number for segment the stego-cover with reference one that happen to improve blindness. The drawback of this method is hiding by GCBS is done in not random locations and the random number for insertion method is not random improves. Furthermore, the encryption in (Khalifa *et al*., 2016) before hiding (play fair) is take a lot of time. In this study, suggested protocols for DNA-based Steganography, the protocol improves the randomness for hiding process by an Elliptic Curve Cryptography (ECC) and implementation ciphering stage in parallel. The objectives of this research are: expanding the security of the system and decrement the execution of time by parallel ciphering. Furthermore, the extraction prepare can be done "blindly" without required to the reference grouping of DNA, since, the ECC worked.

Literature review: The new media of the steganography techniques are used to keep information hard to detect by any people. That new media is Deoxyribo Nucleic Acid (DNA). For using DNA as a cover to hide information, many researcher worked on this field: Shiu *et al*. (2010) suggested three methods (insertion method, substitution method, complementary method) and compare between these methods in each method reseracher represent the capacity of carrying the information and hardly for attacker for extract information inside cover. (Shimanovsky *et al*., 2003; Sabry *et al*., 2010) suggests, "Arithmetic encoding" for hidden information into DNA chain. The thought of the algorithm was based on the feature of codon redundancy. Numerous codons were interpreting to the same amino acids of the central dogma. In this way, it begins by change over message that to be covered up in double arrangement into a decimal number between and arithmetic encoding is utilized to parse through the different codon tables. The length of the resultant stego-DNA depends on the accuracy of the embedded division that clearly influences the precision of the blind recovery handle.

Corresponding Author: Zena Ahmed, Department of Computer Science, College of Science for Women, University of Babylon, Babylon, Iraq

By Sabry *et al.* (2010) proposed a critical alteration to the ancient play fair cipher by presenting DNA-based and amino acids-based structure to the center of the ciphering prepare. In this think about, a parallel frame of information such as plain text messages or pictures are changed into groupings of DNA nucleotides. Hence, these nucleotides pass through a play fair encryption prepare based on amino-acids structure.

Amal Khalifa and Safwat Hamad (Khalifa *et al.*, 2016) proposed a unique thought of covering up information in DNA or RNA called LSBase (Least Significant Base substitution). It employments a momentous property of codon excess to present noiseless changes into DNA groupings. In this way, the DNA grouping can be modified without influencing the sort or the structure of protein it produces. The researchers in (Khalifa *et al.*, 2016) suggest hiding in in two stage first by novel hiding of substitution and second hide in insertion method each codon is translated to some amino acid, since, there is 22 amino acid with 64 codons, after convert message to DNA sequences start cipher by play fair (Sabry *et al.*, 2010) and hiding by GCBS and then hide by insertion method to be blind extraction.

MATERIALS AND METHODS

Proposed scheme

Before explained the proposed work must know important material

Genetic material of DNA: The Deoxoraibo Nucleic Acid (DNA) is a double helix strand form of building blocks called nucleotides. Each nucleotide is continued of collection of base, the base four: Adenine (A), Thiamin (T), Guanine (G) and Cytosine (C). Ghosh and Bansal (2003) they contact with each other by nitrogen bases. Therefore, DNA is viewed as the sequence of base pairs: AATAAGTAATATCGAATCGAT. Adjacent of three nucleotides become a unit known as the codon that codes for an amino acid and these amino acid is transfer into protein of molecular biology of the human beings (Ghosh and Bansal, 2003). That process are called the central dogma.

Steganography technique: Steganography is the science for hide's message inside the digital media (Cheddad *et al.*, 2010) two types of steganography includes blind (Khalifa *et al.*, 2016) and non-blind (Shiu *et al.*, 2010). Non-blind means the extraction process is done by comparing the cover media with stego-media to find the secret message, for that way we need the cover DNA to extract the message. On the other hand, blind steganography means extract message without need for cover media only stego this method is used in this proposal.

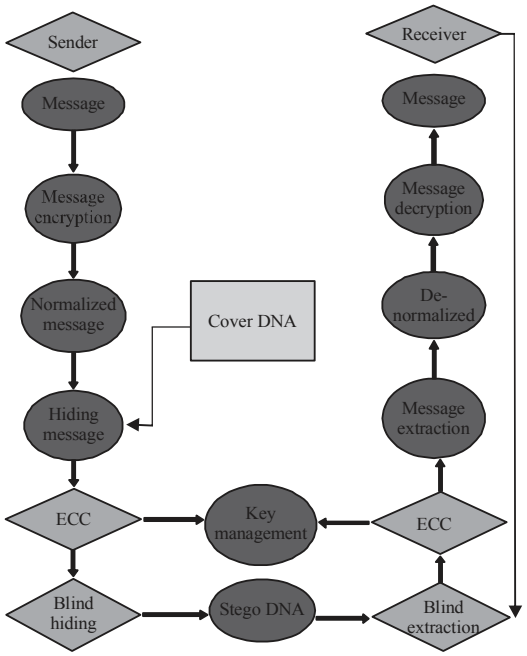

Fig. 1: General stages of system

Table 1: Khalifa *et al.* (2016) mapped character to DNA form

Characters	Codon	Characters	Codon
A	GCT, GCC, GCA, GCG	O	AAT, AAC
B	TAA, TGA, TAG	P	CCT, CCG, CCC, CCA
C	TGT, TGC	Q	CAA, CAG
D	GAT, GAC	R	CGT, CGC, CGC, CGA
E	GAA, GAG	S	TCT, TCC, TCA, TCG
F	TTT, TTC	T	ACT, ACC, ACA, ACG
G	GGT, GGC, GGA, GGG	U	AGA, AGG
H	CAT, CAC	V	GTT, GCT, GTA, GTG
I/J	ATT, ATC, ATA	W	TGG
K	AAG	X	AGT, AGC
L	CTT, CTC, CTG, CTA	Y	TAT
M	ATG	Z	TAC

The proposal: The proposed scheme done by two stages parallel stage (in ciphering stage) and random location generator (in hiding stage) before explained these stages must show in Fig. 1. General system.

Parallel stage: In this stage, the process of (Message encryption) by play fair by Khalifa *et al.* (2016) is done here in parallel. The message is encrypted by play fair (Sabry *et al.*, 2010) play fair is working by ordering the letters of the English language in the matrix dimensions 5*5 after placing the secret key of play fair characters in the matrix. The original work of algorithm is same but in this research, proposed the division of the matrix into two parts. Worker 1 and 2 executes the cods simultaneously. Each core is take a copy of whole matrix and execute a half of it the second core is implement the rest of the matrix after the implementation of execution is done the result is gathering. The output of that algorithm is

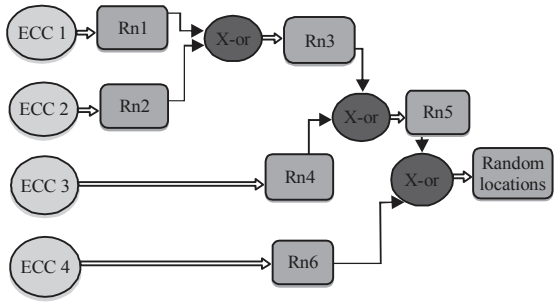

Fig. 2: Proposed ECC protocol

Table 2: Khalifa *et al*. (2016) complementary rules

Base	Complement
Watson-Crick base pairing	
A	T
C	G
G	C
T	A
A generic base pairing	
A	C
C	T
G	A
T	G

character and then convert to DNA form as illustrated in Table 1 (Khalifa *et al*., 2016)the output entered to second stage (hiding) Table 1.

Hiding stages: In this stage the hiding is perform by Generic Complementary Base Substitution (GCBS) (Khalifa *et al*., 2016). GCBS is proposed by Khalifa *et al*. (2016) but here altered the algorithm by hiding in a random location using Elliptic Curve Cryptography (ECC).

Elliptic curve cryptography: Elliptic curve is a mathematical way to find the point on its curve into different way according to Eq. 1 cryptography use this kind of ciphering (block cipher system that) to make the message hard to detect by intruder. Figure 2 show the proposed ECC:

$$y^2 = x^3 + ax + b \quad (1)$$

Here we proposed four series of ECC. One elliptic is (x-or) with series two and the result is hashed with series three. The results of these series is (x-or) with series four. In this way, this is hard to detect the location of hiding process.

Elliptic Curve Cryptography (ECC) used in two stages of hiding:

- Generic Complementary Based Substitution (GCBS) method
- Insertion method

GCBS method: In this method substitute the cover sequence with its GCBS in order the location of (ECC), ECC (Elliptic Curve Cryptography) is cryptography method used for ciphering but here used for random location generator for hiding, GCBS (Khalifa *et al*., 2016) is research on Table 2 (Khalifa *et al*., 2016) under this rule:

Message base	A →sj	Stego-base
	C→C (sj)	
	G→CC (sj)	
	T →CCC (sj)	

ECC is used to generate a number of series and merge between them in order to get a number of random locations that locations used for hidden the message by using (GCBS) these series are generated according to equation below:

$$Pi = ECC\ (Si\text{-}1)i\ 0> \quad (2)$$

$$Pi = H\ (Si, Sj)i,j > 0 \quad (3)$$

$$Pi = H\ (Si, Sj, Sk)i, j, k > 0 \quad (4)$$

$$Pi = H\ (Si, Sj, Sk, Sm)\ i, j, k, m > 0 \quad (5)$$

Where:

Pi : Random Point on a curve
H : Hash function
Si : Series one of ECC
Sj : Series two of ECC
Sk : Series three of ECC
Sm : Series four of ECC

Insertion method: In second stage use (ECC) to create a location between stego-DNA and cover DNA to start overlapping between them that algorithm is used to make the extraction be blind. The randomness of these locations can be tested by using randomness tests. That is used to measure the robustness of ECC.

Randomness tests: The quality or state of missing a design or rule of organization; unpredictability. It tests by statistical powerful test called (diehard). Diehard is a statistical test to measurement the randomness, it includes 18 different tests to test the p-value Saif:

Birthday spacing test, overlapping 5-Permutation (OPERM5) test, binary rank test (for 31×31 matrices), binary rank test (for 32×32 matrices), binary rank test (for 6×8 matrices), Monkey (Bit-stream) tests, Overlapping-Pairs-Sparse-Occupancy (OPSO) test, Overlapping-Quadruples-Sparse-Occupancy (OQSO) test, DNA test, count-the-1's test on a stream of bytes, count-the-1's test for specific bytes, parking lot test, minimum distance test, 3D spheres test, sqeeze test, overlapping sum test, runs test, craps test.

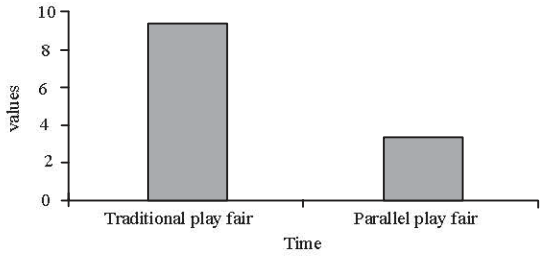

Fig. 3: Diagram for time of ciphering

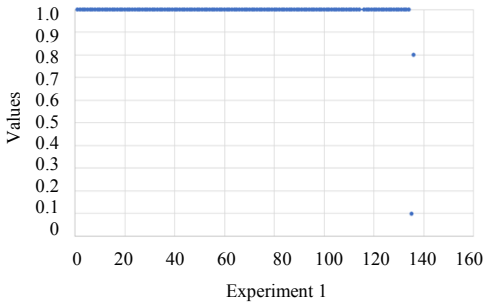

Fig. 4: Experiment one

RESULTS AND DISCUSSION

The proposed research improves system performance and robustness.

System performance:

Experiment 1: Is implemented ciphering in MATLAB the result of whole ciphering without parallel is take 9.294777 sec (Fig. 3).

Experiment 2: Is implemented ciphering in MATLAB the result of parallel ciphering is take a time 3.337362 sec.

Here notice that the performance is prove by decreasing the time of execution process in ciphering process.

Robustness: The randomness of the locations used to hide the secret information is hard to detect which gives the robustness of system. These locations generated by ECC as series.

One series of ECC generated to find the random number of locations on a curve of ECC as illustrated in Fig. 4 the size of file that use for testing is 10.5 MB.

This figure shows that the number of points in fail aria is too many that is ranged from (0-0.1 and 0.9-1) while the doubt area from 0.1-0.25 and 0.75-0.9 is only one location by using the first generated series by the ECC. There are no exist for the safe points which is ranged from (0.25-0.75). When ECC is used to generate two series these two series are hashed to maximize the number of random locations on the hashed series as illustrated in Fig. 5.

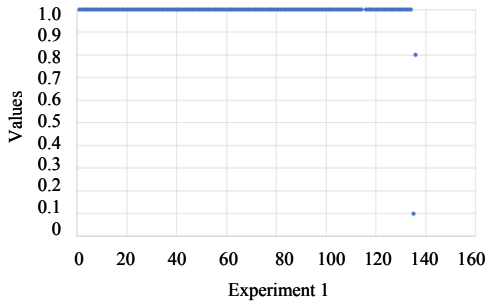

Fig. 5: Experiment 2

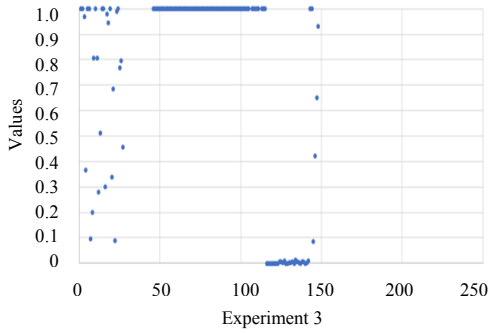

Fig. 6: Experiment 3

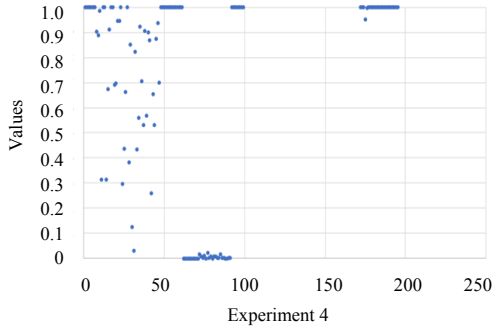

Fig. 7: Experiment 4

This figure shows the number of points in fail area is less than experiment one and the number of points in safe area is more than one point while keeping the doubt area is almost constant. When ECC is used to generate three series and hashed them as shown in Fig. 6.

In this experiment shows the number of points in fail area is less than Fig. 4 and 5 the number of point in safe area is more than Fig. 4 and 5 while the doubt area is more than Fig 4 and 5. When ECC is used to generate four series and hashed them as shown in Fig. 7.

This figure shows the number of points in fail area is less than experiment one and two and three and the number of points in safe area is more than one points while the doubt area is more than experiment one and two and three.

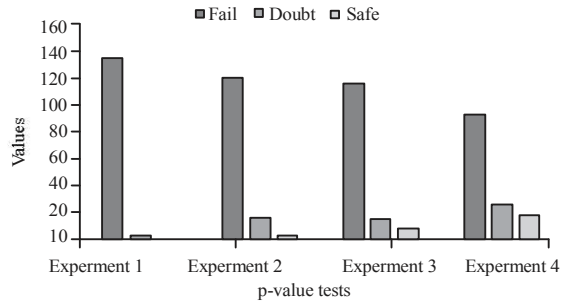

Fig. 8: Results of experiments

The abstract of four experiment above lets that whenever using more than one elliptic curve and hashed them the randomness increases. This increasing means the random of generated locations are safe and unpredictable. Fig. 8 illustrate this.

CONCLUSION

In this study, show a strong hiding by using protocol of elliptic curve for generating random location in hiding stage. In this way hides inside DNA is not recognize by attacker because the Information must be ciphering before hiding and hiding is random improves. Process of encryption called play fair here using it but work in parallel to reduce the time of ciphering to start hiding stage.

REFERENCES

Abbasy, M.R. and B. Shanmugam, 2011. Enabling data hiding for resource sharing in cloud computing environments based on DNA sequences. Proceedings of the 2011 IEEE World Congress on Services, July 4-9, 2011, IEEE, Washington, DC, USA., ISBN:978-1-4577-0879-4, pp: 385-390.

Anderson, R., R. Needham and A. Shamir, 1998. The steganographic file system. Proceedings of the 2nd International Workshop on Information Hiding, Apr. 14-17, Portland, Oregon, USA., pp: 73-82.

Cheddad, A., J. Condell, K. Curran and P. McKevitt, 2010. Digital image steganography: Survey and analysis of current methods. Signal Process., 90: 727-752.

Cherian, A., S.R. Raj and A. Abraham, 2013. A survey on different DNA cryptographic methods. Intl. J. Sci. Res., 2: 167-169.

Ghosh, A. and M. Bansal, 2003. A glossary of DNA structures from A to Z. Acta Crystallogr. Sect. D. Biol. Crystallogr., 59: 620-626.

Khalifa, A., A. Elhadad and S. Hamad, 2016. Secure blind data hiding into pseudo DNA sequences using playfair ciphering and generic complementary substitution. Appl. Math. Inf. Sci., 10: 1483-1492.

Lu, M., X. Lai, G. Xiao and L. Qin, 2007. Symmetric-key cryptosystem with DNA technology. Sci. China Ser. F: Inform. Sci., 50: 324-333.

Murdoch, S.J. and S. Lewis, 2005. Embedding covert channels into TCP/IP. Proceedings of the 7th International Workshop on Information Hiding, June 6-8, 2005, Springer, Berlin, Germany, ISBN:978-3-540-29039-1, pp: 247-261.

Sabry, M., M. Hashem, T. Nazmy and M.E. Khalifa, 2010. A DNA and amino acids-based implementation of playfair cipher. Intl. J. Comput. Sci. Inf. Secur., 8: 129-136.

Shimanovsky, B., J. Feng and M. Potkonjak, 2003. Hiding Data in DNA. In: Information Hiding, Petitcolas, F.A.P. (Ed.). Springer, New York, ISBN: 9783540364153, pp: 373-386.

Shiu, H.J., K.L. Ng, J.F. Fang, R.C.T. Lee and C.H. Huang, 2010. Data hiding methods based upon DNA sequences. Inform. Sci., 180: 2196-2208.

Yan, D., R. Wang, X. Yu and J. Zhu, 2012. Steganography for MP3 audio by exploiting the rule of window switching. Comput. Secur., 31: 704-716.