Pertanika J. Sci. & Technol. 25 (4): 1369 - 1380 (2017)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

DNA and LCG Based Security Key Generation Algorithm

Sodhi, G. K.[1], Monga, H.[2] and Gaba, G. S.[1]*

[1]*Discipline of Electronics & Communication Engineering, Lovely Professional University, Jalandhar, India*
[2]*Jan Nayak Ch. Devi Lal Vidyapeeth, Haryana*

ABSTRACT

To ensure reliable and efficient operations of encryption and hash codes, a unique approach of formulating a security key from Deoxyribonucleic acid (DNA) of an individual is presented in this paper. The fusion of DNA sequence with Linear Congruential Generator (LCG) sequence ensures uniqueness in the keys generated and eradicates the problem of duplicate keys. The obtained key is significant due to its optimum length and robust algorithm. Simulation results reveal that keys produced thus pass the criteria of being random, by a significant coefficient value. Uniqueness is verified through avalanche test, which assures generation of a unique key every time.

Keywords: Authentication, Biometrics, Confidentiality, DNA, Linear Congruential Generator

INTRODUCTION

Communication in today's world focuses on obtaining the data at the desired receiver end, unaltered and retaining its confidentiality from intruders. Security involves authentication, confidentiality and integrity. Integrity means maintaining the trust between two communication ends. As stated by Hao, Anderson, and Daugman, (2006) biometrics is gaining importance these days; biometric features are not only unique but also serves as an authentic representation of an individual. The concept of developing a system which uses a combination of biometrics with factitious intelligence systems to provide high efficiency can be seen in the integration of human iris features with cryptography in Hao et al. (2006). A system that works on audio fingerprint is also proposed by several studies (Covell, & Baluja, 2007; Baluja, & Covell, 2007; Ying, Shu, Jing, & Xiao, 2010). Electrocardiogram (ECG) signals are also used in various studies (Brown

ARTICLE INFO

Article history:
Received: 05 June 2017
Accepted: 23 September 2017

E-mail addresses:
gurpreetsodhi123@gmail.com (Sodhi, G. K.),
himanshumonga@gmail.com (Monga, H.),
er.gurjotgaba@gmail.com (Gaba, G. S.)
*Corresponding Author

ISSN: 0128-7680

& Seberry 1989; Chouakri, Bereksi-Reguig, Ahmaldi, & Fokapu, 2005; Khokher, & Singh, 2015; Ktata, Ouni, & Ellouze, 2009; Garcia-Baleon, Alarcon-Aquino, & Starostenko, 2009). A technique is proposed by Covell, & Baluja (2007) to create signatures for authentication. Identification based on facial features is also reported in the past by Chen, & Chandran (2007); Wei & Jun (2013).

DNA has been used in many cryptographic algorithms by Chang, Kuo, Lo, & Lv (2012). Linear Congruential Generator (LCG) is used to make the technique more efficient and effective compared to traditional generators (Hedayatpour, & Chuprat, 2011). This generator works on a secret seed value which ensures the generation of a different sequence for every input seed value provided. The work reported in this paper is based on the idea of blending the unique and random characteristics of DNA with the sequences generated using Linear Congruential Generator.

The key generating algorithm is tested using NIST tests of randomness as well as evaluated on the basis of avalanche criterion, the results of which are formulated in Table 4 and Table 5 respectively. The proposed technique has outperformed in comparison to the traditional ones, thus, making it well suited in applications where security key is the major concern.

This paper is organized as follows; characteristics of DNA and Linear Congruential Generator are described in Section 2. In Section 3, the method for generating the 256-bit key is presented, where the DNA values are taken from MIT-BIH database by Goldberger et al. (2000), followed by results in Section 4. In the last section, a summary of the main points is presented.

Characteristics of DNA and L.C.G

Progress in the field of forensics biotechnology has made deoxyribonucleic acid (DNA) sequencing more efficient. The DNA sequences of various organisms have been successfully sequenced with accuracy by Goldberger et al. (2000). However, the analysis of DNA sequences using biological methods is a slow process. Therefore, the assistance of computers is crucial.

On the other hand, many distributed databases providing DNA data have been constructed and can be easily accessed from the World Wide Web such as from National Centre for Biotechnology Information (http://www.ncbi.nlm.nih.gov). Most of the techniques involved treat DNA sequences as the symbolic data, a composition of four characters A, G, C, and T corresponding to the four types of nucleic acids: Adenine, Guanine, Cytosine, and Thymine, respectively. However, the bimolecular structures of genomic sequences can be represented as not only the symbolic data but also in a numeric form. DNA is made up of two polymeric strands composed of monomers that include a nitrogenous base (A-adenine, C-cytosine, G-guanine, and T-thymine), deoxyribose sugar and a phosphate group. The sugar and phosphate groups, which form the backbone of the strands, are located on the surface of DNA while the bases are on the inside of the structure. According to studies by Chang et al. (2012), weak hydrogen bonds between complementary bases of each strand (i.e. between A and T and between C and G) give rise to pairing of bases which hold the two strands together. DNA sequences are unique for each individual, even in the case of identical twins. The pattern formed by a DNA sequence specifically represents an individual and its characteristics. Hence, there is no chance of duplicity.

To strengthen the bond of security, a random sequence is generated by LCG. This sequence is generated using a seed value which is kept secret by the user. LCG uses an algorithm that produces a sequence of pseudo-randomized numbers calculated through a linear equation. It's a robust and efficient method of generating pseudo-random numbers.

The working principle of the LCG can be understood through the given equation:

$$X_{n+1} = (aX_n + c) \bmod m \tag{1}$$

Where, X: the sequence of pseudorandom values

X: the sequence of pseudorandom values
m: $0 < m$ the modulus
a: $0 < a < m$, the multiplier
c: $0 \leq c < m$, the increment
X_n: $0 \leq X_n < m$, the seed or start value

This sequence along with the DNA sequence forms a very strong 256-bit key which is not only less susceptible to attacks but also provides a higher level of security.

Security Key Generation

The suggested key is prepared by integrating the DNA sequence of an individual and LCG random sequence. The working principle of the suggested algorithm is explained in three subsequent subsections:

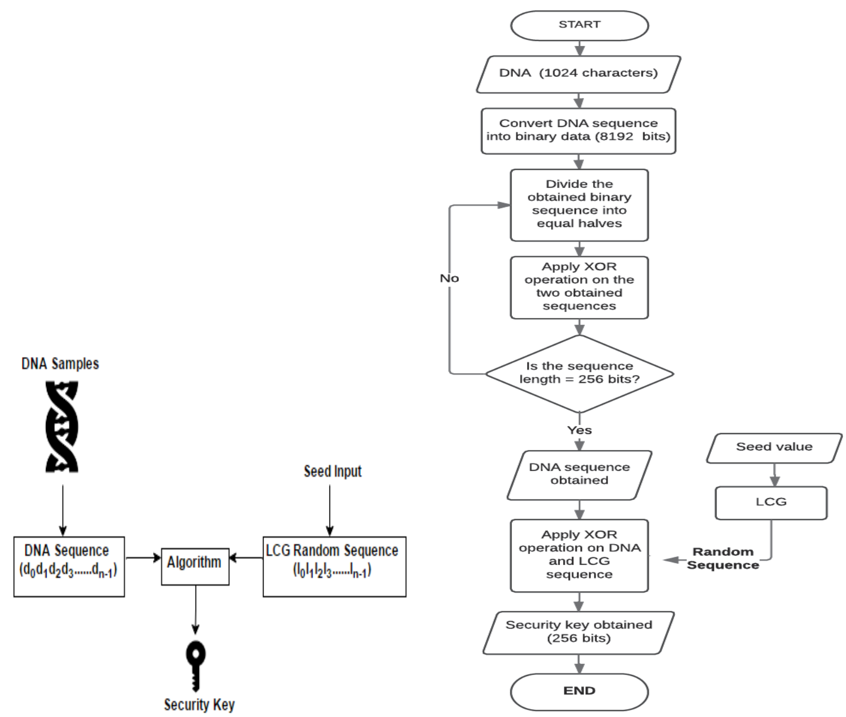
DNA sequence formulation

1. Obtaining a DNA sequence of 1024 characters from the DNA database from Ensembl website (http://www.ensembl.org).

The DNA sequence consists of base pairs 'agct'.

2. Obtaining the binary sequence from DNA characters:
   Each character of the DNA sequence is represented by 8-bit ASCII code. Hence, resulting in a DNA sequence of length 8192 bits.
3. Framing a DNA sequence of 256 bits:
   (i) The DNA sequence is then divided into equal halves.
   (ii) Apply exclusive-or operation on the obtained sequences.
   (iii) The result of modulo-2 summation is further divided into two equal parts and exclusive-or operation is applied again.

The step (iii) is repeated until a sequence of 256 bits is obtained. The whole procedure is summarized in the flow chart (Figure 1).



(a) (b)

Figure 1. (a) Key Generation Model (b) Key Generation Process

The algorithm is repeated three times for three different DNA sequences, the results are tabulated in Table 1 below:

Table 1
DNA sequence formulation

	DNA sequence (1024 characters)	DNA sequence (8192 bits)	DNA sequence](256 bits)
D_1	gcacaatcagaagcaggcgga ggagacggcggccttcgagga ggtcatgaaggacctgagcct gac................................ ....gcacagaggcaaggcgtc agcaggcatcgcccaccctgtc tccgctgtcacccatcactcag gctgtagccatg .	01100111011000110 11000010110001101 10000101100001011 1010001100011…… ……………………… ……………………… ……………………… ……………………… …0111010001100001 01100111011000110 11000110110000101 11010001100111	00000100000000000010 10100010111000001000 00011000010101000010 00000011000000110000 01100000011000010011 0 00000000000000000100 11000000000000010000 00110000000000000110 00000110000001000000 0 01000010011000001100 0 00001000000000000001 0 00000000000010011000 0 0000
D_2	ccacgcgtccgggcgagaaga tggcgacttcgaacaatccgcg gaaattcagcgagaagatcgc gct……….…………… ……………………........ ……………………........ ....ggc gtcagccccctgtccctgagca cagaggcaaggcgtcagcagg catcgcccgccctgtccccgct gtcacccat.	01100011011000110 11000010110001101 10011101100011011 0011101110100........ .....………………..… .....………………..… …………0111010001 10001101100001011 00010110001101100 01101100001011101 00	00000010000000100000 0 01000000010000001100 0 01011100000100000101 1 10000010000000000000 0 00000000000000000110 0 00000100000011000010 1 01000100110000001000 0 00000000000000000100 0 01001100000110000000 0 00000001000000110000 0 01100000001000000000 1 01110000010000000100
D_3	agcccttaggggaagagtcct gctctggctgttgatgctccagc tccagaaatcccagtacctgca actg………………… ……………………… ……………………… …………………tctgg agcagcagctgccctacgcctt cttcacccaggcgggctcccag cagccaccgccgcagccccag cccccgccg	01100001011001110 11000110110001101 10001101110100011 1010001100001… …………………… …………………… …………………… 10001101100011011 00011011000110110 01110110001101100 01101100111	00010101000000000000 0 01000010111000001000 0 01011100000000000000 1 00000010000000110000 0 00000000010000010101 0 00000100000001000000 00 00000000000001010100 0 00010000000100000000 0 00000100000000000001 0 11100000001000010011 00 01001100000110000000 0 00001011100010011000 1 0011

*D_1, D_2, D_3: DNA sequences

Random sequence generation through LCG

LCG generates the random sequence on the basis of equation (1). The values assigned to the variables in the equation are:

a=23; c=0; m=(10^8+1),

Table 2
LCG produced random sequences

	Seed value	Generated random sequences (256 bits)
L_1	47594118	01010000010001100011011010011011010110111100111110101010 01001001111011100000111111110011000001011011100101000011 11010110111111000111011101111100101111111111010111010111 10011011000100100010010001010011100000111011001011110100 10000001001110100001110000000010
L_2	47594122	01010001010001010011001101000111011111101011101000110001 00001110110010101001110100110100101101001111111000000101 10011100000011010000100010010101011101110101011011100000 11110110000001111001010011100011110101101011000001100011 00010011000001101111000110101101
L_3	3435973836	1001000000010101100110000000111000101000110011010011011010 111011000100000010101001100010110000111010100100000110011 00011011001111000000100001110011111001001101101101100010 00001000101001000011010000101100010111111010010100111001 00110110001110010111110011000000

*L_1, L_2, L_3: LCG sequences

Three sequences are generated using three different seed values. The obtained random sequences are summarized in Table 2. Fusion of DNA sequence and LCG random sequences:

Table 3
Security keys

KEY_1	010101000100011000100011100011000101100111001001101111110100 110111101000000010011111010100000011101010100100001111010110 111011110111011101111110101110011111010111010001100111010001 0110001001100100000001000010110110000111101001000010100111010 0000111100000010
KEY_2	010100110100011100110001010001010111100010101101001101010001 1001110011101001110100110100101101001111100000000011110011010 000110000001101110010111011101110101011011100010111001010000 000110010100111000011101000010110110011000010001001100010001 1111010110101001
KEY_3	100001010010101100110010001011111010011100100011110110101110 111001000101010101010001011000011001010111010011000100011111 0011110000000100001100110111001101101100101100010000011001010 010000100011001011100100110010110110001111110011011000101110 0110111111010011

Fusion of DNA sequence and LCG random sequences: Further to escalate the impact of randomness Exclusive-or logic is applied between each DNA and LCG sequence. This is repeated for two other DNA and LCG sequences. Finally, all three 256-bit keys are obtained as shown in Table 3. The three keys are represented as KEY_1, KEY_2, KEY_3.

The obtained keys are unique and random and thus can play a significant role in high tech security systems.

RESULTS AND DISCUSSION

The efficiency of a security key is analysed by inspecting its random characteristics and uniqueness. The National Institute of Standards and Technology (NIST) mentions some aspects for selecting and testing random number generators (Rukhin et al., 2001). The outputs of such generators can be used in many security applications to design security keys. The generators to be used for security applications need to be robust enough to handle attacks. In particular, their outputs should be unpredictable if there is no knowledge about the seed. These tests determine whether or not a generator is suitable for particular security applications. The randomness of a key is evaluated on the basis of its P-value, which should be greater than 0.01 for a random sequence.

The efficiency of the proposed technique is evaluated by comparing it with other traditional techniques used in the field of authentication and security key generation (Garcia et al., 2009; Hedayatpour et al., 2011; Wei & Jun, 2013; Ying et al., 2010). The tests have been performed on KEY_1 and the results are presented in Table 4.

Table 4
Security keys

S. No.	Input Source of random number generator	Key length (bits)	Runs Test	Frequency Test	Approximate Entropy Test	Binary Derivative Test	Maurer's Test	DFT Test	Random Excursion Variant Test
			P-value	P-value	P-value	P- value	P- value	P- value	
1	ECG	128	0.1262	0.2487	0.5468	0.5039	0.9428	0.0294	Random
2	Image	256	0.0809	0.8026	0.9759	0.4887	0.9780	0.4220	Random
3	Iris sequence	128	0.1254	0.3768	0.9409	0.5021	0.9062	0.3304	Random
4	Finger print	128	0.3345	0.3041	0.3345	-	0.2757	0.7597	Random
5	**DNA & LCG**	**256**	**0.0809**	**0.8026**	**0.9497**	**0.0608**	**0.9667**	**0.4220**	**Random**

It is observed that the P-value generated by the proposed algorithm for r all the seven tests is significantly greater than 0.01, ensuring they satisfy the criteria required as efficient security keys.

Avalanche test was also performed on the obtained keys. The purpose of this test is to check the avalanche effect, a desirable property for security keys. Where if the input is changed slightly the output changes significantly. It gives the percentage of bits flipped with a change in input. This is a significant property of security keys.

The test is performed on three sets of DNAs and LCG sequences:

Case 1: In the initial set, two security keys are generated through two DNA sequences while keeping the same LCG sequence.

Case 2: The second set involves generation of two security keys through the same DNA sequence and two LCG sequences.

Case 3: In the third set, two security keys are generated through two DNA and LCG sequences.

Results of the avalanche effect is calculated for each of the three sets are tabulated in Table 5, Table 6 and Table 7 respectively.

Table 5
Avalanche test analysis: Case 1

DNA Sequences (D_n)	Seed Value	LCG Sequence (L_n)	Key Generated K= D_n xor L_n	Avalanche result of Key (K)	
				No. of Bits Flipped	Avalanche Effect
D_1	7594118	L_1	0101010001000110001000111 0001100010110011100100110 1111110100110111101000000 0100111110101000000111010 1010010000111101011011101 1110111011101111110101110 0111110101110100011001110 10001011000100110010000000 1000010110110000111101001 0000101001110100000111100 000010	58	22.65 %
D_2			01010010010001000011010001 0011001010111011101100010 1011100101111011101010000 0111111110011000001011011 1111010000011101000011101 0010110010001111110101111 1111110101110101011000100 00001010000100100010100001 1000010110110100111101101 00000010010110100011000000 000110		

*Refer Table 1 for D_1, D_2, D_3 and Table 2 for L_1, L_2, L_3
**Different DNA sequences - Same LCG sequence

Table 6
Avalanche test analysis: Case 2

DNA Sequences (D_n)	Seed Value	LCG Sequence (L_n)	Key Generated K= D_n xor L_n	Avalanche result of Key (K)	
				No. of Bits Flipped	Avalanche Effect
D_1	47594118	L_1	0101010001000110001000111 0001100010110011100100110 1111110100110111101000000 0100111110101000000111010 1010010000111101011011101 1110111011101111110101110 0111110101110100011001110 1000101100010011001000000 1000010110110000111101001 0000101001110100000111100 000010	123	48.04 %
	4759412	L_2	0101010101000101001001100 1010000011111001011110000 1001000000101011001100100 1101100110010101100101110 1101000001011001110000011 1100000100010010111011100 0101010110111001101111000 0000000111001011011110000 1101000010110010011000110 0010111000001101110001010 101101		

*Refer Table 1 for D_1, D_2, D_3 and Table 2 for L_1, L_2, L_3
**Same DNA Sequence - Different LCG sequence

Table 7
Avalanche test analysis: Case 3

DNA Sequences (D_n)	Seed Value	LCG Sequence (L_n)	Key Generated K= D_n xor L_n	Avalanche result of Key (K)	
				No. of Bits Flipped	Avalanche Effect
D_1	47594118	L_1	0101010001000110001000111 0011000101100111001001101 1 111101001101111010000000 10 011111010100000001110101010 0100001111010110111011 1101 1101110111111010111001 1111 0101110100011001110100010 1 100010011001000000010000101 101100001111010010000 10100 11101000000111100000010 010100110100011100110001 01	117	45.70 %
D_2	4759412	L_2	000101011110001010110100 11 010100011001110011101001 11 010011010010110100111110 00 000000111100110100001100000 011011100101110111011101 01 011011100010111001010000 00 011001010011100001110100 00 101101100110000010001001100 0100011111010110101001		

*Refer Table 1 for D_1, D_2, D_3 and Table 2 for L_1, L_2, L_3
**Different DNA Sequences - Different LCG sequence

CONCLUSION

This paper presents a unique approach to generate security key for cryptography using DNA and LCG sequence. The suggested technique uses the unique biological characteristics along with pseudo-random generator to build a novel key generator. DNA when used in collaboration with the LCG sequence yields better results in terms of security. If used separately, biometrics may prove to be a weak authentication technique as the DNA of an individual can be obtained unaware. Thus, the integration of LCG sequence with biometric features makes the security key a powerful tool with least possibility of being stolen or duplicated. Many researchers in the past generated the key using various biometric inputs such as fingerprints, facial attributes, iris and voice, whereas fewer studies have been reported using DNA as an input for security purposes. The proposed algorithm is used to compute three 256-bit biometric security keys and the performance is evaluated on the basis of NIST Tests. The results revealed that the technique is highly efficient for security key generation. As a future work, other signals like audio, video etc. can be used as inputs for this algorithm other than DNA. The algorithm can also be extended for longer biometric security keys to enhance the strength of security.

REFERENCES

Baluja, S., & Covell, M. (2007). Audio fingerprinting: Combining computer vision & data stream processing. *Proceeding of International Conference on Acoustics, Speech and Signal Processing. Honolulu,* HI: IEEE. Retrieved from http://ieeexplore.ieee.org/document/4217383/

Brown, L., & Seberry, J. (1989). *On the design of permutation P in DES type cryptosystems*. Retrieved from https://link.springer.com/chapter/10.1007/3-540-46885-4_71

Chen, B., & Chandran, V. (2007). Biometric based cryptographic key generation from faces. *Proceedings of 9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications*. Glenelg, Australia: IEEE. Retrieved from http://ieeexplore.ieee.org/document/4426824/

Chang, H. T., Kuo, C. J., Lo, N. W., & Lv, W. Z. (2012). DNA sequence representation and comparison based on quaternion number system. *DNA Sequence, 3*(11), 39-46.

Chouakri, S. A., Bereksi-Reguig, F., Ahmaldi, S., & Fokapu, O. (2005). Wavelet denoising of the electrocardiogram signal based on the corrupted noise estimation. *Proceedings of Computers in Cardiology.* Lyon, France: IEEE. Retrieved from http://ieeexplore.ieee.org/document/1588284/

Covell, M., & Baluja, S. (2007). Known-audio detection using waveprint: spectrogram fingerprinting by wavelet hashing. *Proceeding of International Conference on Acoustics, Speech and Signal Processing.* Honolulu, HI: IEEE. Retrieved from http://ieeexplore.ieee.org/document/4217060/

Garcia-Baleon, H. A., Alarcon-Aquino, V., & Starostenko, O. (2009). A wavelet-based 128-bit key generator using electrocardiogram signals. *Proceeding of 52nd International Midwest Symposium on Circuits and Systems*. Cancun, Mexico: IEEE. Retrieved from http://ieeexplore.ieee.org/document/5236010/

Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., & Stanley, H. E. (2000). Physiobank, physio toolkit, and physio net. *Circulation, 101*(23), (215- 220).

Hao, F., Anderson, R., & Daugman, J. (2006). Combining crypto with biometrics effectively. *IEEE Transactions on Computers, 55*(9), (1081-1088).

Hedayatpour, S., & Chuprat, S. (2011). Hash functions-based random number generator with image data source. *Proceedings of Conference on Open Systems*. Langkawi, Malaysia: IEEE. Retrieved from http://ieeexplore.ieee.org/document/6079248/

Khokher, R., & Singh, R. C. (2015). Generation of security key using ECG signal. *Proceedings of International Conference on Computing, Communication and Automation*. Noida, India: IEEE. Retrieved from http://ieeexplore.ieee.org/document/7148503/

Ktata, S., Ouni, K., & Ellouze, N. (2009). A novel compression algorithm for electrocardiogram signals based on wavelet transform and SPIHT. *International Journal of Signal Processing, 5*(4), 32-37.

Rukhin, A., Soto, J., Nechvatal, J., Smid, M., Barker, E., Leigh, S., Levenson, M., Vangel, M., Banks, D., Heckert, A., Dray, J., & Vo, S. (2001). *A statistical test suite for random and pseudorandom number generators for cryptographic applications*. Retrieved from National Institute of Standards and Technology website: http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-22.pdf

Wei, W., & Jun, Z. (2013). Image encryption algorithm Based on the key extracted from iris characteristics. *Proceedings of 14th International Symposium on Computational Intelligence and Informatics*. Budapest, Hungary: IEEE. Retrieved from http://ieeexplore.ieee.org/document/6705185/

Ying, L., Shu, W., Jing, Y., & Xiao, L. (2010). Design of a Random Number Generator from Fingerprint. *Proceedings of International Conference on Computational and Information Sciences.* Chengdu, China: IEEE. Retrieved from http://ieeexplore.ieee.org/document/5709056/