International Journal of Security and Its Applications
Vol.8, No.4 (2014), pp.113-122
http://dx.doi.org/10.14257/ijsia.2014.8.4.11

A Sticker-Based Model Using DNA Computing for Generating Real Random Numbers

Saman Hedayatpour, Nazri Kama and Suriayati Chuprat

Advanced Informations School, Universiti Teknologi Malaysia, International Campus, Kuala Lumpur, 54100, Malaysia
saman.hedayatpour@gmail.com, nazrikama@ic.utm.my, suria@ic.utm.my

Abstract

Real random values have wide range of application in different field of computer science such as cryptography, network security and communication, computer simulation, statistical sampling, etc. In purpose of generating real random values, need for a natural noisy source refers to the main challenge where a source of noise may be reliable for using in random number generator if and only if be derived from physical environment. In this work, we address this requirement by using DNA computing concepts where the molecular motion behavior of DNA molecular provides a pure source of physical noise that may be used for generating high quality real random values. Since one of the main factor for evaluating quality of real random values refer to expectation for generating approximately same amount of 0s and 1s, in this article we model a DNA-based random number generator in sticker mode with ability of generating equal numbers of 0 and 1. After using molecular motion behavior of DNA molecular as the natural source of noise into the proposed DNA-based random number generator, the generated value were subjected to frequency, run, and serial tests which are proposed by National Institute of Standards and Technology (NIST) for randomness evaluation. Obtained result from this evaluation shows that beside the achieving high scores in run and serial tests, the values generated by our DNA-based random number generator pass frequency test with 100% success.

Keywords: *Random Number Generator, DNA computing, Randomness test, DNA molecular motion, Sticker-based model of DNA computing*

1. Introduction

Nowadays, real random values play an extremely important role in different field of computer science such as cryptography, network communication, computer simulation, statistical sampling, *etc.*, [1]. The most famous cryptographic algorithms need real random values either directly as the key such as *One Time Pad* algorithm or as a seed for expanding the key like *DES* and *AES*. For embedding security into the network protocols, real random values are used to protect the communication channel by using security mechanism such as TCP three-way handshake [2, 3].

Since all software algorithms for generating random values work based on mathematical formulas, their output will be in form of pseudorandom where even the strongest formulas will repeat the output after some period of time [2, 4]. In purpose of generating real random values, having access to a natural and physical source of noise such as nuclear decay, Brownian motion, analog circuit and quantum mechanics refers to the main requirement in all real random number generators (RNG).

ISSN: 1738-9976 IJSIA

In this work for proposing an extremely fast RNG with ability to generate high quality random, we use DNA computing concepts that refer to one of the main categories of Nano-computing and there is belief that the future of computing will face a revolution by using this amazing technology [3]. The power of DNA computing lies on the unbelievable ability of DNA molecules for processing data in parallel via the huge number DNA molecules that may check huge number of possibilities at once. This vast parallelism computation along with extraordinary density for data storage in DNA molecules will be cause for lots of new achievements in different areas of computer science such as its usage for breaking the modern cryptographic algorithms where none of them is resistance against conducted DNA-based *brute force*.

Beside the mentioned properties and abilities of DNA computing, the main behavior of DNA molecules that makes it unbelievably suitable for using in random number generator refers to the molecular motion that provides a pure source of physical noise. For using this behavior of DNA in process of generating real random value, this paper proposes a DNA-based random number generator which works based on DNA molecular operations such as *Denaturation, Ligation, Annealing(Renaturation), Restriction enzyme,* and *Polymerase Chain Reaction (PCR)* [4, 5].

Once the random values were generated in form of final DNA strand using DNA molecular operations, the final random value (in binary form) will be extracted from target strand. In the final step, the generated real random value will be subjected to three levels of randomness tests proposed by National Institute of Standards and Technology [6, 7]. We organized the rest of this paper as follow: Section 2 is a brief introduction on similar works. Section 3 describes overall structure of the proposed DNA-based random number generator. The evaluation of the proposed RNG is the content of Section 4. In Section 5 we discuss on contributions of this work and finally Section 6 is conclusion and future work.

2. Related Works

There exist a number of works attempt to generate real random values using different natural physical noisy sources such as *Random Number Generator Based on Transformed Image Data Source* [1], *Disk drive generates high speed real random numbers* [8], and *Quantum random number generator based on the photon number decision of weak laser pulses* [9] but in specific area of using DNA computing for generating random value there are just two similar works.

The first one which is *DNA based random number generation in security circuitry* introduced by Christy Gearheart, Benjamin Arazi and Eric Rouchka [10]. That is a prototype schema which uses oligonucleotides' relations for generating random values from DNA sequences while plasmid vectors are used for temporary storage and retrieval of random values.

The *On Pseudorandom Number Generation from Programmable and Computable Biomolecules: Deoxyribonucleic (DNA) as a Novel Pseudorandom Number Generator* refers to the second work in this area which was proposed by Okunoye Babatunde [11]. This work discusses on generating pseudorandom values using Watson-Crick relations.

3. The DNA-Based Random Number Generator

For describing how the DNA-based random number generator works, firstly we have to know some fundamentals on DNA computing concepts. The DNA computing works based on chemical reactions and specific relations between four nucleotides of Guanine, Adenine, Thymine, and Cytosine into the DNA strands. *Watson Crick* refers to the most famous role in

deal with DNA strands which defines DNA structure as: The number of cytosine is equal to guanine and the number of thymine is equal to adenine (The pairs of A: T and C: G are structurally similar). The length of each pair is the same and the two phosphate backbones fit them equally. These pairs are held to each other by hydrogen bonds (a type of chemical attraction that is easy to reform and break). All of the DNA strands are in form of double helix with the hydrogen bonds at the core that provide a way to unzip the two strands for easy replication.

As is demonstrated in Figure 1, a DNA molecular may have randomly vibrational, rotational, and translational motions at the same time [12] where this molecular motion behavior of DNA molecular provides a pure source of physical noise that may be used for generating high quality real random values.

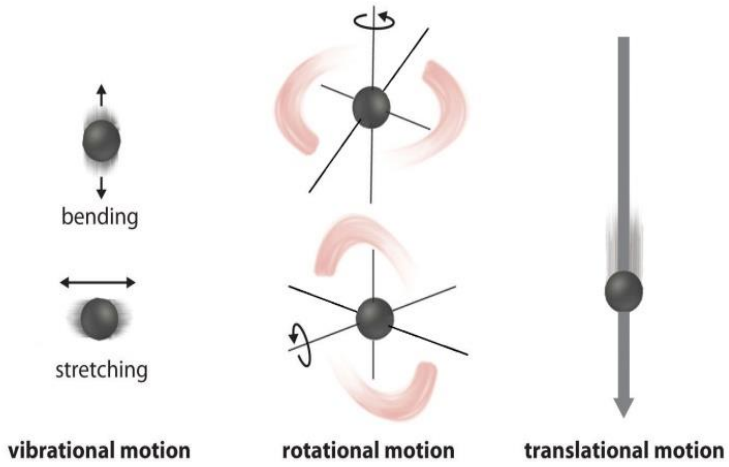

Figure 1. DNA Molecular Motions

For making this pure source of noise available on our DNA-based random number generator we need to use some biological operations on DNA strands such as Denaturation, Renaturation (Annealing), Restriction enzyme, Ligation, and Polymerase Chain Reaction (PCR). These key notions are briefly described are follow:

Denaturation: By increasing the temperature of solution slowly (standard heating condition), the DNA double helix will be broken to two single strands.

Renaturation (Annealing): This operation is reverse of Denaturation that means if the temperature of solution is decreased slowly the complementarily strands are linked together by hydrogen bind.

Restriction enzyme: Enzymes can change stored information in DNA strands. Restriction Endonucleases is one of the most famous enzymes which is able to recognizes the particular strand in the DNA strands and cleaves DNA strand at specific side. As an example for restriction enzyme, Sau3AI will cut the double strands in restriction side which sequence of GATC.

Ligation: This operation is beneficial to link a DNA double helix to other DNA double strands. It bounds the 5‟ phosphate end of one strand with 3‟ hydroxyl end of another strand.

Polymerase Chain Reaction (PCR): Generally PCR is using to increase the number of DNA double helix in such a way that in each cycle of PCR the target DNA double strand will be double.

3.1. Platform

In this work, the entire process of generating real random value built based on a very long double helix strand. This very long double helix strand will be in certain format which must be created using ligation operation on crated strands in certain cycle of PCR. In this work we simulate this process but in practice the shorter strand (but in the same format) must be ordered from worldwide DNA banks. The rest of process will be as already explained in using ligation and PCR operations. In purpose of generating this very long double helix strand, Figure 2 shows the initial double helix strand and how this initial double helix strand will be duplicated in operating one PCR cycle and finally a longer strand in format of the initial double helix strand will be created.

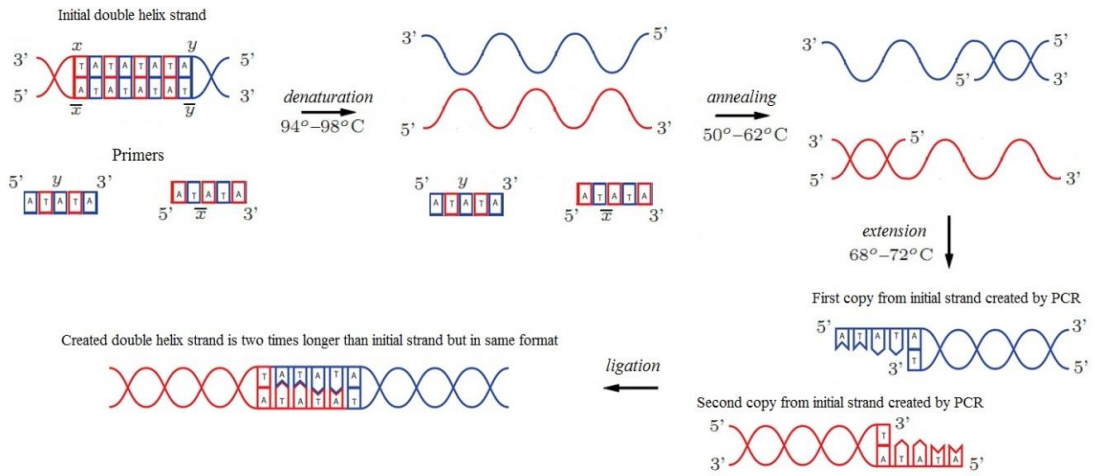

Figure 2. Process of Creating Long Double Helix Strand in Format of Initial Strand

The pairs of nucleotides into the target long strand must be either in pairs of Adenine and Thymine or pairs of Guanine and Cytosine. In this work as is shown in Figure 2 we use pairs of Adenine and Thymine which based on *"Watson Crick"* roles each nucleotides of Adenine in one side is in pair with one Thymine in opposite side (based on Watson Crick roles, Adenine and Thymine only could be in each other pair where Cytosine and Guanine also only could be in each other pair).

The requiring length for creating strand is directly depend on the length of asked random values in such a way that creating strand shall be as long as four times longer than length of asked random values. For instance in a case that generating 10000 bits of random value required, a DNA double helix strand in length of 40000 nucleotides must be (20000 pairs of T A or A T sequences).

3.2. Implementation in Sticker Mode

Once required double helix strand was created, we will slowly increase the temperature of sample to 94-98 Celsius degree. In this condition, denaturation will happen and the long double helix strand will be divided into two long single strands with sequences of Adenine and Thymine.

In this step and after dividing target long double helix strand into two single strands, one of the single strands will be removed from the sample (there are number of biological and

electronically techniques for performing such an operation). Now, short sequences of "AT" will be generated as much as length of required random values by using PCR operation. It means if 10000 bits of random values is required (having long strand in length of 20000 pairs of "TA" sequences) thus, 10000 separate sequence of "AT" shall be generated. Besides using PCR for generating 10000 separate sequence of "AT" there is possibility for using restriction enzymes for dividing the second side of target long strand into these sequences. Figure 3 demonstrates condition of our temples in end of this step.

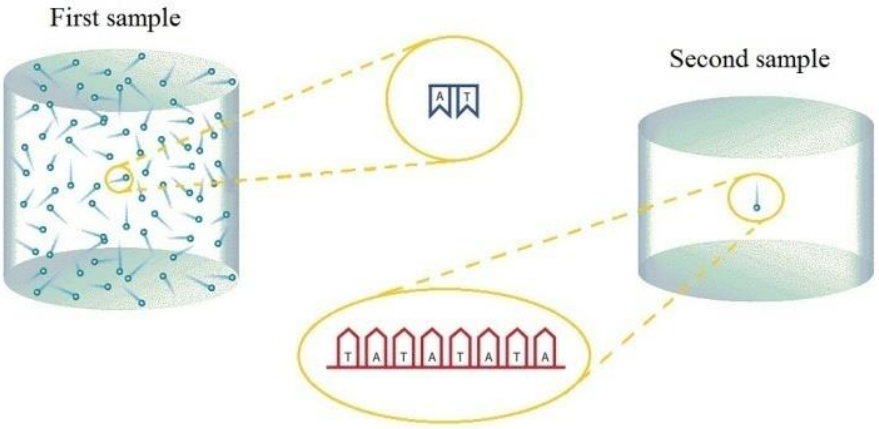

Figure 3. Contents of Initial and Secondary Samples

As is shown in Figure 3, first sample contains 10000 pairs of "AT" sequences while second sample contain just one long strand in length of 40000 nucleotides (20000 pairs of "TA" sequences). Into the last step of the sticker mode, both these samples shall be mixed and riled frequently in specific conditions where temperature must be decreased slowly to 50-62 Celsius degree. This is the required condition for operating Annealing (Renaturation) where based on *"Watson Crick"* roles; shorts strands in pair of "AT" attempt to find and make connection with their match palaces of "TA" into the long strand. Figure 4a shows the final sample immediately after mixture and Figure 4b demonstrates the final sample after a short period of time (required time for annealing operation).

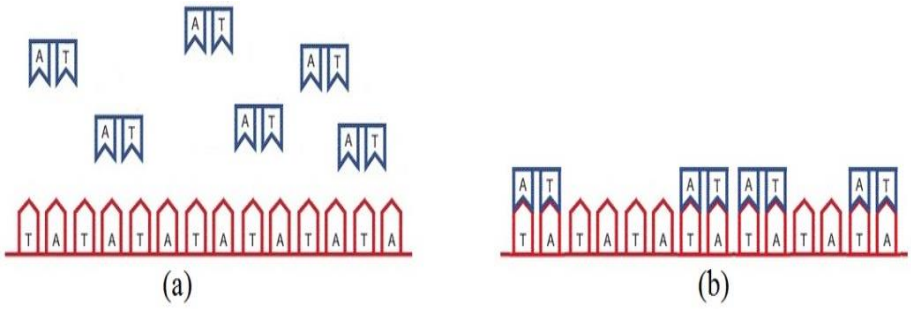

Figure 4. Final Sample Before and After Annealing Operation

Since there is molecular random motion in time of the making connection between different nucleotides, the short sequences of "AT" will be connected to their matched pairs of "TA" in the long strand without any patterns and in form of random seating place.

3.3. Extracting Random Values

In purpose of making final strand readable in binary format, the places in long strand which were matched with pairs of "TA" will be assumed as 1 and the places which are still unmatched will be assumed as 0. Based on this definition, Figure 5 represents the method of translating sticker mode of the DNA strand into the binary format.

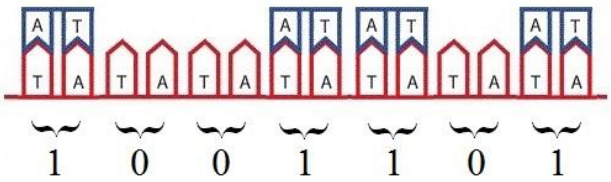

Figure 5. Conversion Sticker Mode of the DNA Strand to the Binary Format

As in shown in Figure 5, this example of DNA strand was translated to "1001101" in binary format. In practice, process of translating random value stored in DNA strand is not as simple as is shown in this simulation. For making a DNA strand in sticker mode readable, each nucleotide shall be fluorescently labelled with a colour then it must be passed to a laser to make it visible and readable in chromatogram.

A chromatogram is a plot that for each nucleotide into the sequence of target strand, one fluorescent colour will be in high intensity while the rest of fluorescent colours for three other nucleotides will be low intensity. For instance, a chromatogram may represents different nucleotides using different colours such as black for Guanine, green for Adenine, red for Thymine, and blue for Cytosine.

Since the final strand in this DNA-based RNG contains either Adenine and Thymine or Guanine and Cytosine, the output of chromatogram will show just either red and green or blue and black.

4. Evaluation

Among the different tests proposed for evaluating randomness, the National Institute of Standards and Technology (NIST) proposes the most famous set of checks where sixteen separate tests provided for evaluating randomness. In this paper we gain three of these tests to evaluate the performance of our DNA-based random number generator [6]. In this purpose, the translated random values from DNA strand were subjected to Frequency, Run, and Serial tests (randomness tests show, whether generated values are in form of true random or not) and the result shows in significances level of α 0.05 the generated random values successfully pass these three tests. These three tests are briefly described as follow [7]:

Frequency test (mono-bit test): The main purpose of this test refers to find whether the amount of 1's and 0's are approximately the same or not. Frequency test is defined as Eq. (1) where X1 is approximately follows a X^2 distribution with 1 degree of freedom if n ≥10.

$$X1 = (n_0 - n_1)^2 / n \quad (1)$$

Serial Test (two-bit-test): The main purpose of this test refers to find whether the amount of appearances of 11, 10, 00 and 01 are approximately the same or not. Serial test is defined as Eq. (2) where X2 approximately follows a X^2 distribution with 2 degrees of freedom if n ≥ 21.

$$X_2 = \frac{4}{n-1}\left(n_{00}^2 + n_{01}^2 + n_{10}^2 + n_{11}^2\right) - \frac{2}{n}\left(n_0^2 + n_1^2\right) + 1 \tag{2}$$

Run Test: This test focuses on the total number of one and zero runs (run refers to the sequence of bits that are appeared in uninterrupted form) of various lengths *i* in the sequence *S* is as expected or not. This test defined as Eq. (3) where X3 approximately follows X^2 distribution with 2k - 2 degrees of freedom.

$$e_i = (n - I + 3)/2^{i+2}$$

$$X_3 = \sum_{i=1}^{k} \frac{(B_i - e_i)^2}{e_i} + \sum_{i=1}^{k} \frac{(G_i - e_i)^2}{e_i} \tag{3}$$

To evaluate the randomness and performance of the proposed DNA-based random number generator, we prepared three samples that the first one includes 100 sequences in length of 1000-bits, second sample contains 100 sequences in length of 10000-bits, and third sample includes 100 sequences in length of 100000-bits. Based on chi-square distribution table by using a significances level of α 0.05, the threshold values for X1, X2 and X3 become 3.84, 5.99 and 9.49 respectively.

The value of X1, X2 and X3 were calculated for all generated sequences in each of three samples. Since this DNA-based RNG is designed in sticker mode in such a way that always number of generating 0s and 1s will be equal thus, 100% of generated sequences pass frequency test. The rest of result shows 98% of the generated values sequences passed the serial test and 94% of them passed run test with I = 3.

5. Discussion

The result of applying three levels of randomness tests on random values generated by this DNA-based random number generator shows that the generated random values meet high level of randomness. In practice there is no more DNA-based RNG with ability to generate equal number of 1s and 0s and consequently passing frequency test with 100% success. Besides the great achievement in deal with frequency test, the random value generated in this work passed run and serial tests with remarkable rates. Two factors of high quality of random value and vast capability in generating long sequences of random value; make this random number generator suitable for using in key parts of huge computer systems.

6. Conclusion and Future Work

By eliminating some serious limitations of current nation of silicon based computers, DNA computing is going to change expectations' borders in processing speed and capability of data storage. In deal with DNA computing, researchers must be familiar with chemical reactions and specific relations between nucleotides into the DNA strands and use this knowledge as the techniques and tools to design, implement, and simulate the computer systems. In this work, we used one of these special behaviors of DNA molecular which is molecular motion as the required source of physical noise to design a DNA-based random number generator. In this work, we design the DNA-based RNG in sticker mode in such a way that always generate equal numbers of 0s and 1s and consequently passing frequency test with 100% of success. Besides the frequency test, generated random values were subjected to run and serial tests

(proposed by NIST for randomness evaluation) where the achieved result prove the high quality of generated random values.

References

[1] S. Hedayatpour and S. Chuprat, "Hash Functions-based Random Number Generator with Image Data Source", IEEE Conference on Open Systems (ICOS2011), Langkawi, Malaysia, **(2011)** September 25-28.
[2] S. Hedayatpour and S. Chuprat, "Random Number Generator Based on Transformed Image Data Source", Advances in Computer, Communication, Control & Automation, Springer-Verlag, Berlin Heidelberg, **(2011)**, pp. 457–464.
[3] T. Schneider and P. N. Hengen, "Molecular Computing Elements, Gates and Flip-Flops", USA, **(2004)**, pp. 37.
[4] A. Fujiwara, K. Matsumoto and W. Chen, "Procedures for logic and arithmetic operations with DNA molecules", International Journal of Foundations of Computer Science, vol. 15, **(2004)**, pp. 461–474.
[5] G. Seelig, D. Soloveichik, D. Y. Zhang and E. Winfree, "Enzyme-free nucleic acid logic circuits", Science, vol. 314, **(2006)**, pp. 1585–1588.
[6] National institute of standards and technology (NIST). Guide to the Statistical Tests. United States **(2008)**.
[7] National institute of standards and technology (NIST). A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications. United States, 800-22, **(2010)**.
[8] S. Erhard and E. Wolfgang, Disk drive generates high speed real random numbers. Microsyst Technol, **(2005)**, pp. 616–622.
[9] W. Wei and H. Guo, "Quantum random number generator based on the photon number decision of weak laser pulses", Lasers & Electro Optics & The Pacific Rim Conference on Lasers and Electro-Optics, Shanghai, China, **(2009)** August 3-30.
[10] M. Christy, A. Benjamin and C. Eric, "DNA based random number generation in security circuitry", Biosystem, pp. 100, (2009), pp. 208-214 (2010).
[11] B. Okunoye, "On Pseudorandom Number Generation from Programmable and Computable Biomolecules: Deoxyribonucleic (DNA) as a Novel Pseudorandom Number Generator", World Applied Programming, vol. 1, no. 3, **(2011)**, pp. 215-227.
[12] M. Silberberg, "Principles of General Chemistry", 3th. McGraw-Hill Science/Engineering/Math, New York. USA, **(2012)** January 17.

Authors

Saman Hedayatpour, he received his Bachelor Degree in Software Engineering from the Jahad Daneshgahi Institute of Higher Education Iran 2008 and Master Degree of Information Security from the Universiti Teknologi Malaysia 2012. He is currently a PhD student and works on Secure Software Development under supervision of Dr. Mohammad Nazri Kama and Dr. Suriayati Chuprat at the Advanced Informations School, Universiti Teknologi Malaysia.

Nazri Kama, he received his Master Degree in Real-time Software Engineering and Bachelor Degree in Management Information System from Universiti Teknologi Malaysia in 2002 and 2000 respectively. He then obtained his PhD degree at The University of Western Australia in 2011. His research interests are in software development, software maintenance, impact analysis, traceability, and requirement interactions.

Suriayati Chuprat, she received her BSc Computer Science, MSc Computer Science (Real-time Software Engineering) from the Universiti Teknologi Malaysia in 1995 and 2000 respectively. In 2009, she received her PhD in Mathematics from Universiti Teknologi Malaysia and completed several research attachments at the University of North Carolina, Chapel Hill, USA. She is currently a senior lecturer at Advanced Informatics School of Universiti Teknologi Malaysia. Her research interests include real-time scheduling theory, scheduling and resource allocation, parallel and distributed computing, information security and cloud computing.