An Application Tool: To Implement DNA cryptography as a seed of Stream Cipher in the domain of IoT

Sharaban Tahura, Institute of Information Technology (IIT), Jahangirnagar University, Banglasesh, sharabantahura110@gmail.com
Roksana Akter, Department of Computer Science and Engineering, Southeast University, Bangladesh, roksana.akter@seu.edu.bd
Rashed Mazumder, Institute of Information Technology (IIT) Jahangirnagar University, Banglasesh, rmiit@juniv.edu

***Abstract* — Providing security for resource-constrained devices has become an essential necessity in recent years. Our evergreen world is going to be dependent on the Internet of Things (IoT), where resource-constrained devices play a vital role. In the domain of cryptography, stream cipher is one of the encryption applications that provides security to the resource-constrained devices. Because of the stream cipher's bit wise operations, it is suitable for both small and flexible sized messages. In addition, DNA cryptography provides a certain level of better security. In this paper, the main motivation is to apply a DNA cryptography on the stream cipher applications. The value of the initialization vector (IV) or seed is well known for maintaining the secrecy of stream ciphers. However, if an attacker breaks the IV or seed then the security of stream cipher decreases. Under these circumstances, we propose a hybrid DNA sequence technique to make secured initial value or seed of stream cipher which certainly increases its resistance and reduces the time complexity.**

***Index Terms* -- Deoxyribonucleic Acid (DNA), Encryption, Internet of Things (IoT), Initialization Vector (IV), Resource Constrained Devices, Security.**

I. INTRODUCTION

The Internet of Things (IoT) [1] is the first evolution of the internet. IoT technology connects several objects and human beings to communicate with each other without the help of human interactions. It has become the future of the incoming epoch. Healthcare, smart grid, smart home, smart parking every sphere of our day to day lifestyle has gone under IoT applications. For this reason, the IoT systems must provide security as well as privacy and protect user's data from attackers, hackers and vulnerabilities. So, the security in IoT applications should be focused broadly. To prevent unauthorized access, data misuse, data monitoring, modification, security plays the most important role. Usually, IoT system architecture can be classified into three different layers: the physical layer, the commutation layer, and the application layer where the physical layer contains resource-constrained devices with less power and minimal memory. IoT devices are likely to be placed in the physical layer which includes RFID and WSN. So, it's most important to ensure the security of the physical layer [2].

Cryptography is the procedure of providing security of such constrained device and stream cipher is the part of symmetric cryptographic algorithms that consume low power, typically fast and compact, that's the reason behind choosing attractive stream ciphers for resource-constrained devices and low power devices. Researchers found stream cipher useful for such a platform and a milestone event in pertinent research was held in European Network of Excellence for Cryptology during 2004-2008 when efficient and compact stream cipher was developed. The International Organization for Standardization and the International Electrotechnical Commission (ISO/IEC) standardized stream ciphers for LWC in the ISO/IEC 29192-3:2012 [3, 4] in 2012. Since then, based on the new design several eSTREAM structures have been raised. Further Adleman invented a cryptographic field with Deoxyribonucleic acid (DNA) where a relationship has been made between cryptography and molecular biology [5]. It is a newborn field and has become a noticeable and valuable stage of international research on cryptography. Rothemund et al. [6] claimed that XOR computation can be exploited with the DNA cryptosystem which is an unavoidable portion of cryptography. DNA hybridization technique, DNA synthesis technique, Polymerase Chain Reaction (PCR), DNA digital coding, etc. are recent biological technologies of DNA cryptographic algorithms. Although many researchers are researching in this area, it is still in its initial stage and that's why till now only few algorithms have been proposed. Moreover, DNA cryptography, as well as these modern biological technologies are extensively dependent on the laboratory at present therefore there are many problems yet to be solved in this field.

II. Related Work

There are various researches on DNA based cryptography such as DNA-based symmetric cryptographic algorithm, DNA-based asymmetric cryptographic algorithm, elliptic curve-based DNA cryptography technique, quantum cryptography technique and cloud cryptography technique. In [7] Kar et al. have proposed a new cryptography technique based on DNA sequencing on data security and cryptography. The Encryption process used here is based on a symmetric key. Moreover, this technique can be applied to secure the real time distributed systems also. Ibrahim et al. [8] mentioned a data hiding technique to improve security by encrypting secret messages. A double DNA sequence is used to hide the message of plaintext.

In [9], the authors have declared a double layer security method based on vigenere cipher to encrypt data using more secured DNA cryptography but the algorithms' time complexity is high. Anwar et al. [10] introduced a new DNA cryptographic technique combining XOR operation and symmetric key cryptography where DNA hybridization

technique and matrix computations were used to minimize time complexity. Bama et al. [11] established an algorithm selecting DNA sequences out of 55 million possibilities to make data transmission more secured, based on substitution techniques that were already implemented in the Electronic Medical Record System. A DNA based cryptographic algorithm with random key generation and matrix multiplications technique was proposed in [12]. Three types of keys: the initial key, the primer key and the generated key were used to generate different ciphertexts which made it more secure.

Borda et al. [13] invented a secret data writing technique exploiting the DNA hybridization process and one-time pad (OTP) scheme. In [14], the authors proposed a hardware solution usable for wireless communications based on DNA cryptography. DNA encryption, mRNA sequence, amino, etc. biological and mathematical concepts helped this technique to make it secure enough to resist the brute force attack. Sherif et al. [15] invented the DNA cryptographic methodology dependent on the symmetric key where key sequences were obtained from the genetic database. DNA computing-based stream cipher for IoTs using MQTT protocol was proposed by Noor et al. where one-time pad (OTP) and DNA computing performed on LFSR based stream cipher [16]. Kamel et al. [17] proposed a DNA based stream cipher implementation where available online databases are used as a DNA sequencer to generate random key values with different lengths and information. Basim et al. also gave ideas to establish biological improvements on digital platforms [18]. Raj et al. presented a symmetric algorithm-based DNA encryption technique to make secured data transmission [19]. In this technique firstly an initial cipher was generated and then converted into a final cipher. Farah et al. [20] introduced an algorithm to generate encryption keys for block and stream cipher using DNA computing.

TABLE I

Comparison Between Traditional Cryptography and DNA Cryptography [21]

Attributes	Traditional Cryptographic Technology	DNA Cryptographic method
Security	One-Fold with computational trouble	Two-fold with biological complication
Time Complexity	Require few seconds	Needs few hours to calculate
Storage Medium	Uses Silicon Chips of computer	Storage of DNA strands
Storage Capacity	1 gram of silicon chip carries 16 MB [22]	1 gram of DNA carries 10^8 TB [23]
Stability	Dependent on platform, language limitations	Dependent on temperature, pH

III. DNA CRYPTOGRAPHY

It is expected that the number of things connected to the internet is increasing by up to 50 billion in 2020 [24]. Revolutions of the real IoT world have become only possible by merging two types of technologies: Radio Frequency Identification (RFID) [25] and Wireless Sensors (WS) [26]. But these technologies have disadvantages over low computational devices as well as resource-constrained devices. For this reason, these technologies could not be capable of providing security to their applications which have limited battery lifetime and limited computational power. This is the main reason behind focusing on stream cipher in this paper. The stream cipher is more suitable for resource-constrained devices because it uses a small memory size and stream ciphers run faster in hardware than the block cipher. Lightweight stream cipher [27] algorithms use a key of a size equivalent to the data where ciphertext can be obtained by bit by bitsol operations on the plaintext and a keystream is generated from a key and an initialization vector and simply XOR-ed with the plaintext to generate the ciphertext [28]. Because of their bit by bit operations, they are found to be potentially more compact, simpler, lighter and faster schemes which solely rely on confusion concepts.

Adleman published the computational ability of DNA solving *NP*-Complete Hamiltonian path problem of seven vertices [29]. Before that, DNA was thought of as the only biological information carrier. Then DNA was developed into a computational tool as a solution of security as well. The four bases A, C, T, and G are used to deal with DNA computing and building DNA languages and the binding properties of nucleotide bases (A-T, G-C) made DNA more powerful. Exceptional information density and the parallel processing properties of DNA molecules are exploited to analyze several cryptographic techniques. In Table 1, a basic comparison is found between traditional cryptographic techniques and DNA cryptographic techniques. Some properties of these techniques such as level of provided security, time complexity, storage medium, and capacity and stability have appeared there. The traditional cryptography technique provides one-fold security with computational difficulties whereas DNA cryptography showed two-fold security with biological complexities. Traditional cryptographic algorithms take a few seconds to process the techniques, on the other hand, it takes a few hours to calculate DNA computing involving PCR and DNA chip technology. The storage medium of traditional cryptography is silicon chips of computers, whereas DNA cryptography captures the storage of DNA strands manipulated by biological strategy. 1 gram of silicon chip carries 16-megabyte storage capacity, on the contrary, 1 gram of DNA contains 10^8 terabyte storage. DNA cryptography and DNA computing have become more tempting and favorable because of large storage capability. The stability of traditional cryptographic algorithms depends on the implementation conditions, the platform, and language limitations. On the other hand, the stability of DNA cryptography depends on temperature, pH, etc. referred to environmental conditions.

IV. CRYPTOGRAPHY ON INITIALIZATION VECTOR

Stream ciphers are very significant parts of the symmetric encryption algorithm. They work on a single bit of message in a time-varying transformation manner. A stream cipher mainly comprises a pseudo-random generator where the generator

receives a secret key and initialization vector as input to produce a pseudorandom keystream. The encrypted cipher message C_1, C_2, ..., Ci can be obtained by the bitwise XORing of the keystream digit Z_1, Z_2, ..., Z_i and the plain message of digits M_1, M_2, ..., M_i as follows:

$$C(i) = M(i) \oplus Z(i) \quad (1)$$

During decryption the ciphertext need to be decrypted with the same keystream adding it to the encrypted message as follows:

$$M(i) = C(i) \oplus Z(i) = M(i) \oplus Z(i) \oplus Z(i) \quad (2)$$

In a stream cipher, IVs are injected into the keyed internal secret state of the cipher and several cipher rounds are accomplished aiming to produce the ciphertext. Using IV on a stream cipher prevents repetition in encryption which makes it difficult for a hacker to find patterns and break a cipher with the help of a dictionary attack. IV needs to be pseudorandom to make it more secured, that's the reason in this paper to encrypt the initialization vector utilizes a parallel DNA cryptography technique [30]. This method encrypts and decrypts the initial value in a parallel manner and reduces the time complexity.

V. METHODOLOGY

A. *Encryption*

In this paper, a DNA cryptography-based technique [30] will be discussed to encrypt the initialization value of a stream cipher. A randomly generated single strand DNA (ssDNA) string will be needed to encrypt this value. In figure 1, a flow chart of the encryption process has been shown.

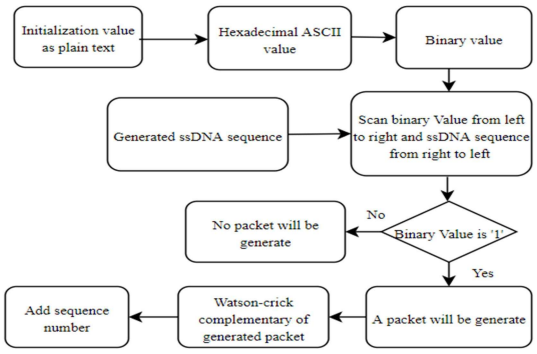

Fig. 1 Process of encryption

The length of the randomly generated DNA string will depend on the length of the binary value of plaintext. If we use n-mer oligoneucleotides then the size of randomly generated single strand DNA will be n* size of binary value. First of all, the IV will be translated into corresponding ASCII code and then this ASCII code will be translated into a binary value. Finally, this binary value will be translated into several packets of encrypted value by using the randomly generated single strand DNA. If we take a value as IV is "KEY". Then its ASCII value is "4B 45 59" and the converted binary value is "1001011 1000101 1011001" = 21 bit.

As the size of the converted binary value is 21 and we use 10-mer oligoneucleotides (it's a variable value, but must be fixed for both encryption and decryption process), then the size

of ssDNA sequence will be: 21*10 = 210. That is suppose,

AGATAGTCATACGTACGACTAACACAGGCATTCACC
ATGGAACAGCGGTTTCCGTAACATCCTGAGTCCAATA
GCGATCCGCTATTCCGTATTTATAAGGGGTCTGATCT
TGCATCCGGCATATAGAGCCGATAGGCAGTATGGGG
AGTAAGCCGGACCAAGGTCAAATAGGGATTCATCGC
CTGATTCCAAAGAATTTGCCGCGCTTCC

Then the ssDNA sequence will be scanned from right to left and the converted binary value will be scanned from left to right in reverse order. Then for every '1' in binary value, a packet will be generated with a sequence number and for '0' no packet will be generated. Here for eleven 1's, we will get 11 packets. To encrypt the value, we take the first 1 bit and last 10-mer oligonucleotides from ssDNA sequences and Watson-crick complementary of this sequence will be added with a sequence number.

AGATAGTCAT-1, ACGTACGACT-0, AACACAGGCA-0, TCACCATGG-1, AACAGCGGTT-1, TCCGTAACAT-0, CCTGAGTCCA-1, ATAGCGATCC-1, GCTATTCCGT-0, ATTTATAAGG-1, GGTCTGATCT-0, TGCATCCGGC-0, ATATAGAGCC-0, GATAGGCAGT-1, ATGGGGAGTA-1, AGCCGGACCA-1, AGGTCAAATA-0, GGGATTCATC-1, GCCTGATTCC-0, AAAGAATTTG-0, CCGCGCTTCC -1.

So, the Watson-crick complementary of the generated 11 packets will be as below:

(CCGCGCTTCC-0 = GGCGCGAAGG-0);
(GGGATTCATC-1 = CCCTAAGTAG-1); (AGCCGGACCA-2 = TCGGCCTGGT-2); (ATGGGGAGTA-3 = TACCCCTCAT-3);
(GATAGGCAGT-4 = CTATCCGTCA-4);
(ATTTATAAGG-5 = TAAATATTCC-5);
(ATAGCGATCC-6 = TATCGCTAGG-6);
(CCTGAGTCCA-7 = GGACTCAGGT-7);
(AACAGCGGTT-8 = TTGTCGCCAA-8);
(TCACCATGG-9 = AGTGGTACC-9);
(AGATAGTCAT-10 = TCTATCAGTA-10).

B. *Decryption*

During the decryption process, after receiving all packets the Watson-crick complementary of the generated packets have to be computed with each 10-mer oligonucleotide string.

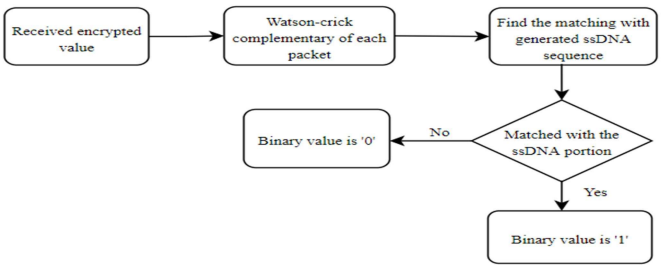

Fig. 2 Process of decryption

Then the attached sequence number will be used to determine the possible position of the packet from the end of randomly generated DNA sequences. After getting all packets each matching position declares a '1' and after completing the evaluation of all packets the unmatched position of the string is referred to as '0'. Thus, the binary value of IV can be obtained and after that, the value is converted into ASCII code to get the plain text message.

VI. Conclusion

Deoxyribonucleic Acid (DNA) cryptography is emerging as a new favorable cryptographic field, where DNA is used to carry the information or to be used as a data encoding approach. Recently, many DNA based algorithms have been developed for data cryptography and cryptographic key generation. This paper proposes an implementation of DNA cryptography as a seed of stream cipher to provide security on the resource-constrained devices of IoT. The implementation uses a hybrid DNA sequence technique to make a secured initialization vector (IV) or seed of stream cipher. Also, the method encrypts and decrypts the initial value in a parallel manner and reduces the time complexity. The encryption/ decryption system has been explained and its cryptanalysis is also discussed here. In the future, DNA encryption will be used in different fields like cloud computing, mobile networks, images, videos, servers, etc. and DNA encryption will take the place of digital signature, authorization and digital timestamps as DNA itself is a unique signature.

VII. References

[1] E. a. o. Dave, "How the next evolution of the internet is changing everything," 2011.

[2] E. R. a. S. H. a. S. M. Naru, "A recent review on lightweight cryptography in IoT," *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC),* pp. 887-890, 2017.

[3] "coordination and support action ECRYPT-CSAresearch network ECRYPT-NET," 09 07 2020. [Online]. Available: https://www.ecrypt.eu.org/?fbclid=IwAR1f8haw3ahwaSyPUsHShzyoUY1wARyh_ckeuw-1SSVc6YzfjUMfdrCzXwu. [Accessed 09 07 2020].

[4] "Security techniques — Lightweight cryptography — Part 3: Stream ciphers," 09 07 2020. [Online]. Available: https://www.iso.org/standard/56426.html. [Accessed 09 07 2020].

[5] L. M. Adleman, "Molecular computation of solutions to combinatorial problems," *Science,* vol. 266, no. 5187, pp. 1021-1024, 1994.

[6] P. W. a. P. N. a. W. E. Rothemund, "Algorithmic self-assembly of DNA Sierpinski triangles," *PLoS Biol,* vol. 2, no. 12, p. e424, 2004.

[7] N. a. M. A. a. S. A. a. D. S. a. P. M. C. Kar, "Data security and cryptography based on DNA sequencing," *International Journal of Information Technology \& Computer Science (IJITCS),* vol. 10, no. 3, 2013.

[8] F. E. a. A. H. a. M. M. Ibrahim, "Enhancing the security of data hiding using double DNA sequences," in *Industry Academia Collaboration Conference (IAC)*, 2015, pp. 6-8.

[9] M. a. K. N. Najaftorkaman, "A method to encrypt information with DNA-based cryptography," *International Journal of Cyber-Security and Digital Forensics (IJCSDF),* vol. 4, no. 3, pp. 417-426, 2015.

[10] T. a. K. A. a. P. S. Anwar, "DNA cryptography based on symmetric key exchange," *International Journal of Engineering and Technology,* vol. 7, no. 3, pp. 938--950, 2015.

[11] R. a. D. S. a. P. K. Bama, "Secure data transmission using DNA sequencing," *IOSR Journal of Computer Engineering (IOSR-JCE) Volume,* vol. 16, 2014.

[12] P. S. a. R. K. G. Varma, "Cryptography based on DNA using random key generation scheme," *International Journal of Science Engineering and Advance Technology (IJSEAT),* vol. 2, no. 7, pp. 168--175, 2014.

[13] M. a. T. O. Borda, "DNA secret writing techniques," in *2010 8th International Conference on Communications*, IEEE, 2010, pp. 451--456.

[14] H. a. C. K. a. D. H. a. V. A. Singh, "DNA based cryptography: An approach to secure mobile networks," *International Journal of Computer Applications,* vol. 1, no. 1, pp. 77--80, 2010.

[15] S. T. a. S. M. a. E.-G. S. Amin, "A DNA-based implementation of YAEA encryption algorithm.," in *Computational intelligence*, 2006, pp. 120--125.

[16] N. A. a. S. M. I. Hussein, "DNA computing based stream cipher for internet of things using MQTT protocol," *International Journal of Electrical and Computer Engineering,* vol. 10, no. 1, p. 1035, 2020.

[17] K. H. a. A. F. M. a. M. L. N. a. H. Y. B. Rahouma, "Design and Implementation of a New DNA Based Stream Cipher Algorithm using Python," *Egyptian Computer Science Journal,* vol. 44, no. 1, 2020.

[18] S. B. a. Y. B. S. Sadkhan-SMIEEE, "DB based DNA Computer to Attack Stream Cipher," in *2019 2nd International Conference on Electrical, Communication, Computer, Power and Control Engineering (ICECCPCE)*, IEEE, 2019, pp. 230--233.

[19] B. B. a. V. J. F. a. M. T. Raj, "Secure data transfer through DNA cryptography using symmetric algorithm," *International Journal of Computer Applications,* vol. 133, no. 2, pp. 19--23, 2016.

[20] F. T. A. E. Hussien, "Proposed Algorithm To Generate Encryption Key For Block And Stream Cipher Using DNA Computing," *Iraqi Journal of Information Technology,* vol. 8, no. 3, 2018.

[21] B. a. S. K. a. H. M. a. D. K. a. o. Anam, "Review on the Advancements of DNA Cryptography," *arXiv preprint arXiv:1010.0186,* 2010.

[22] " Silicon storage technology," 09 07 2020. [Online]. Available: https://rb.gy/tz5uvh. [Accessed 09 07 2020].

[23] G. a. L. C. a. L. H. a. L. X. Cui, "DNA computing and its application to information security field," in *2009 fifth international conference on natural computation*, IEEE, 2009, pp. 148--152.

[24] S. S. a. S. B. a. J. P. Dhanda, "Lightweight cryptography: A solution to secure IoT," *Wireless Personal Communications,* vol. 112, no. 3, pp. 1947--1980, 2020.

[25] "Radio frequency identification," 22 05 2021. [Online]. Available: https://internetofthingsagenda.techtarget.com/definition/RFID-radio-frequency-identification.

[26] J. R. Vacca, "Computer and Information Security," *Computer and Information Security Handbook, Third Edition,* 2017.

[27] S. D. a. B. BHUYAN, "Performance analysis of current lightweight stream ciphers," *A statistical test suite for pseudorandom number generators for cryptographic applications,* p. 45:256, 2020.

[28] C. a. H. G. a. F. K. a. R. K. Manifavas, "Lightweight cryptography for embedded systems--a comparative analysis," in *Data Privacy Management and Autonomous Spontaneous Security*, Springer, 2013, pp. 333--349.

[29] L. M. Adleman, "Molecular computation of solutions to combinatorial problems," *Science,* vol. 266, no. 5187, pp. 1021--1024, 1994.

[30] S. a. S. S. K. Pramanik, "DNA cryptography," in *2012 7th International Conference on Electrical and Computer Engineering*, IEEE, 2012, pp. 551--554.