DNA BASED ADVANCED ENCRYPTION STANDARD ALGORITHM USING VERILOG HDL

P. Roopa Ranjani

1Department of Electronics and Communication Engineering, Assistant Professor,
G.Narayanamma Institute of Technology and Science, Hyderabad.

Abstract
The specific use of mechanical and related procedural protections is a significant obligation of each Federal association in giving satisfactory security to its electronic information frameworks. This distribution determines a cryptographic calculation, the Advanced Encryption Standard (AES) which might be utilized by Federal associations to ensure sensitive information. Assurance of information during transmission or while reception might be important to keep up the privacy and trustworthiness of the data spoken to by the information.

Information encryption (cryptography) is used in different applications and conditions. The particular use of encryption and the usage of the AES is found in numerous elements. Cryptography is utilized to secure information while it is being conveyed between two parties or while it is put away in a medium defenseless against physical burglary. Correspondence security gives assurance to information by enciphering it at the transmitting point and deciphering it at the receiving point.

Document security gives assurance to information by enciphering it when it is recorded on a capacity medium and decoding it when it is used again from the capacity medium. In this, the key must be accessible to the transmitter and recipient at the same time during correspondence.

A field of cryptography based on the combination of AES and DNA computing is employed to secure data as it provides security, a vast range of parallelism, and meager power consumption. An AES algorithm is designed with the inclusion of DNA units at the level of S-Boxes for compression of bit length. The AES algorithm is coded in Verilog HDL and synthesized in the Xilinx ISE Design Suite.

Keywords DNA (Deoxyribonucleic Acid) DNA Mapping, Key-Block- Round Combination, AES (Advanced Encryption standard) , Verilog HDL (Hardware Description Language) , ASCII (American Standard Code for Information Interchange), FPGA (Field Programmable Gate Array).
1 Introduction

Cryptography largely consists of designing and developing ways to convert readable text to complete meaningless (Encryption) and then using a reverse procedure to convert the meaningless text to readable text (Decryption). Nowadays a variety of computational and mathematical methods are used to develop algorithms. A cipher is a pair of algorithms that create encryption and reverse decryption.

The detailed operation of a cipher is controlled both by the algorithm and in each instance by a "key". The key is a secret (ideally known only to the communicants), usually a short string of characters, which is needed to decrypt the ciphertext. Formally, a "cryptosystem" is the ordered list of elements of finite possible plaintexts, finite possible cipher-texts, finite possible keys, and the encryption and decryption algorithms that correspond to each key. Keys are important both formally and in actual practice, as ciphers without variable keys can be trivially broken with only the knowledge of the cipher used and are therefore useless for most purposes. Encryption is a way to protect the information/data from unauthorized access to prevent the classified information from being read by anyone without authority.

In cryptography, encryption is the process of encoding a message or information in such a way that only authorized parties can access it. Encryption does not of itself prevent interference but denies the intelligible content to an intruder. In an encryption scheme, the intended information or message, referred to as plaintext, is encrypted using an encryption algorithm, which utilizes a key to encrypt the plain text, in the end generating cipher-text that can only be read if decrypted. For technical reasons, an encryption scheme usually uses a pseudo-random encryption key generated by an algorithm. An authorized recipient can easily decrypt the message with the key provided by the originator to recipients but not to unauthorized users. Encryption algorithms are broadly classified into two,

**Symmetric key algorithms**

Uses the same key for encryption and decryption processes.

**Asymmetric key algorithms**

Uses two keys, a public encryption key and the other is a private decryption key.

Data manipulation in symmetric systems is faster than asymmetric systems as they generally use shorter key lengths. Use of asymmetric systems enhances the security of communication. Encrypting data using standardized cryptographic algorithms may consume more energy which drastically reduces the lifetime of the components. Two main approaches are followed to design and implement security primitives which are fit with extremely constrained devices. Firstly, designing a new lightweight cryptosystem. For instance, are some recently proposed lightweight cryptographic algorithms. Secondly, modifying the existing standard cryptosystem in a lightweight fashion. Possible examples of the second approach are modification of the Advanced Encryption Standard Algorithm (AES), SHA-256 etc.

With respect to the security aspect and implementation complexity, AES is considered as one of the strongest and efficient algorithms. Despite that, like othersymmetric encryption algorithms, the secret key distribution is still considered as a critical issue. Again to encrypt or decrypt a single block (128-bit) of data, an essential amount of computational processing has to be done which consumes enormous battery power. As components of IoT have resource-constraint characteristics, consuming immense power may cause expiration of such components. Analyzing related work, we come to know that Substitution Layer is the most energy consuming portion of AES in the round based design.
1.1 Literature Survey

The literature survey focuses its attention towards AES, particularly to utilize lowpower consumption, high security, better performance and improved efficiency. The implementation feasibility in the VLSI environment is also studied and analyzed in depth. Ross Anderson, Eli Biham, Lars Knudsen (1999) presented a proposal for the Advanced Encryption Standard. Its design is highly conservative, yet still allows a very efficient implementation. With a 128-bit block size and a 256-bit key, it is as fast as DES on the market leading Intel Pentium/MMX platforms yet we believe it to be more secure than three-key triple-DES. The linear transformations were just bit 27 permutations, which were applied as rotations of the 32-bit double words in the bit slice implementation. The authors also considered replacing the XOR operations by seemingly more complex operations, such as additions. Finally, cognate algorithms with the same structure as Serpent but with block sizes of 64, 256 and 512 bits.

A.J. Elbirt, W. Yip, B. Chetwynd, C. Paar (2000) presented an FPGA implementation and performance evaluation of the AES block cipher candidate algorithm finalists. Reprogrammable devices such as Field Programmable Gate Arrays (FPGAs) are highly attractive options for hardware implementations of encryption algorithms as they provide cryptographic algorithm agility, physical security, and potentially much higher performance than software solutions. The implementations of each algorithm will be compared in an effort to determine the most suitable candidate for hardware implementation within commercially available FPGAs. A design methodology was established which in turn led to the architectural requirements for a target FPGA. The best speed-optimized implementations were identified for each AES finalist in both non-feedback and feedback modes.

Bertoni, G., Breveglieri, L., Koren, I., Maistri, P., and Piuri, V (2002) presented Concurrent fault detection for a hardware implementation of the Advanced Encryption Standard (AES). It is very important not only to protect the encryption/decryption process from random faults. It will also protect the encryption/decryption circuitry from an attacker who may maliciously inject faults in order to find the encryption secret key. They presented a parity prediction algorithm which was designed for each of the four round transformations employed by 30 the Encryption module. Since all byte elements of the output state for each transformation was computed in parallel, they did the same with the output parity bits, and they showed that the proposed scheme leads to very efficient and high coverage fault detection.

The older encryption standard Triple DES has been developed and tested experimentally. Khoa Vu, David Zier (2003) presented FPGA Implementation AES for CCM Mode encryption using Xilinx Spartan-II. This paper discusses a possible FPGA implementation of the AES algorithm specifically for the use in CCM Mode Encryption. It investigates the possibility of creating an off-chip AES system for CCM so that the process can be sped up. The AES implementation on the FPGA is a viable solution for improving the speed and processing power of CCM Mode Encryption. A much larger FPGA or ASIC would be preferred, for both encryption and decryption could be implemented as well as some pipelining of processes. Although the design was implemented, it was intended to be interfaced with a microcontroller/microprocessor running the CCM Mode Encryption.

Xinnmiao Zhang and Keshab K. Parhi (2004) presented high speed VLSI architectures for the AES algorithm. A novel high-speed architecture for the hardware implementation of the Advanced Encryption Standard (AES) algorithm. Composite field arithmetic is employed to reduce the area requirements, and different implementations for the inversion in subfield GF (24) are compared. In order to explore the advantage of sub-pipelining further, the SubBytes/InvSubBytes is implemented by combinational logic to avoid the unbreakable delay of LUTs in the traditional designs. Fully Sub-pipeline encryptor/decryptors using other key lengths can be implemented by adding more copies of round units and modifying the key expansion unit slightly. Expected throughput is slightly lower than the encryptor-only implementations.

David Hwang, Patrick Chaumont, Kristiri, and Ingrid Verbauwhede (2006) discussed securing embedded systems Education, pp. 134-161, 2006. A 29 a top-down, multi abstraction layer approach for embedded security design reduces the risk of security flaws, letting designers maximize security while limiting area, energy, and computation costs. Embedded systems are essentially processor-based devices. Embedded
devices pose severe resource constraints on the security architecture in terms of memory, computational capacity, and energy operating under resource-constrained conditions. At the algorithm level, a designer must select or design both cryptographic and application-specific algorithms for implementation on the embedded device.

Breveglieri, L., Koren, I., and Maistri, P., “An operation-centered approach to fault detection in symmetric cryptography ciphers,” IEEE Transactions on Computers, vol. 56, no. 5, pp. 635-649, 2007. An operation-centered approach to fault detection in symmetric cryptography ciphers is presented. They proposed a general framework for error detection in symmetric ciphers based on an operation-centered approach. They first enumerated the arithmetic and logic operations included in the cipher and analyzed the efficiency and hardware complexity of several error-detecting codes for each such operation. They recommended an error-detecting code for the cipher as a whole based on the operations it employs. The trade-off between the checkpoint frequency and the error coverage was also evaluated.

Yen, C. H. and Wu, B.F., “Simple error detection methods for hardware implementation of advanced encryption standard”. IEEE Transactions on Computers, vol. 55, no. 6, pp. 720-731, 2006. In this they proposed a simple, symmetric, and high-fault-coverage error detection scheme for AES. Although the erroneous bits were diffused in AES, this work used the linear behavior of each operation in AES to design a detection scheme. It is possible to use only one 8-bit register for storing the parties during hardware implementation. This error detection may also be used in encryption-only or decryption-only designs. Because of the symmetry of the proposed detection scheme, the encryption and decryption circuit can share the same error detection hardware. The proposed schemes can be applied in the implementation of AES against differential fault attacks and can be easily implemented in a variety of structures, such as 8-bit, 32-bit, or 128-bit structures.

1.2 Problem Statement

DNA-based Advanced Encryption Standard Algorithm is used for securing the DNA sample by a standard algorithm AES. The AES algorithm encrypts and decrypts the DNA at the transmitting and receiving end hence securing it to the core.

A field of cryptography based on AES and DNA computing is employed. It will ensure a secure transfer of DNA.

The proposed method is based on DNA Encoding in the underlying Galois Field of the AES and unlike other countermeasures does not add any additional operation to the algorithm execution flow, does not decrease the working frequency, does not alter the algorithm and keeps perfect compatibility with the published standard. In addition, it does not require pre-computed tables for storing masked values in ROM. This is specifically important for constrained security tokens such as smart cards. We focused on the optimization strategies for 128-b data path architecture to achieve High Security, High-throughput hardware AES encryption module providing multiple levels of security with small area footprint. The area is saved by reorganizing the encryption data path to minimize the number of data registers and combinational logics.

2 Block Diagram and Working

In the proposed system the DNA based key is used for Encryption and Decryption. As the same key is used for both the operations, this concept works for symmetric key algorithms. DNA sequences can be represented using the DNA binary strands. DNA binary bits are formed by repeated concatenation of the oligonucleotides encoding bits through the complementary sticky ends. The DNA binary bits are the representative of the corresponding digital binary strings.

DNA input is converted to hexadecimal format and then given to the AES encryption algorithm along with a randomly generated key, where the hexadecimal input is encrypted and cipher text is obtained. This cipher text is given to the AES decryption algorithm where the cipher text is decoded with the same key and then resultant is hexadecimal format and then again this obtained result is converted back to the required DNA as shown in Fig. 2.1.
DNA (Deoxyribonucleic Acid)

DNA (Deoxyribonucleic Acid) is a nucleic acid that is the backbone of all the living organisms. The DNA holds the necessary genetic information which can help to build other cells like proteins and RNA (Ribonucleic Acid), DNA is a macromolecule consisting of two strands that twist around a common axis in a shape called a double helix. The double helix looks like a twisted ladder, the rungs of the ladder are composed of pairs of nitrogenous bases (base pairs), and the sides of the ladder are made up of alternating sugar molecules and phosphate groups.

Molecules of DNA range in length from hundreds of thousands to millions of base pairs. The smallest chromosome in the human genome, Chromosome 21, has around 48 million base pairs.

DNA replicates by separating into two single strands, each of which serves as a template for a new strand. The new strands are copied by the same principle of hydrogen-bond pairing between bases that exists in the double helix. Two new double-stranded molecules of DNA are produced, each containing one of the original strands and one new strand. This “semiconservative” replication is the key to the stable inheritance of genetic traits.

DNA is double helix structure in which two strands are coiled to each other and it is made of nucleotides. There are 4 bases in nucleotide which are named as Adenine (A), Thymine (T), Guanine (G) and Cytosine (C) as shown in Fig. 2.2.

2.2 DNA Cryptography

Cryptography is the science of the study of secret writing. It helps in encrypting a plaintext message to make it unreadable. It is a very ancient art; the root of its origin dates back to when Egyptian scribes used non-standard hieroglyphs in an inscription. In DNA Cryptography base pairs forms an information carrier. DNA Cryptosystem encrypts the data in the form of A, T, C, G which is the combination of 0’s and 1’s such as 00-
A, 01-T, 10-C 11- G. Hybridization in which double stranded DNA molecules use single stranded DNA molecules. In this process Adenine always pairs with Thymine while Guanine always pairs with Cytosine. Polymerase Chain Reaction (PCR) is the process of amplifying a single or multiple copies to produce millions of copies of DNA sequence. Primer is a strand of nucleic acid that functions as a beginning point for DNA synthesis. Transcription and Splicing is the process of removing the non-coding areas and rejoining the remaining coding areas and the information of DNA is moved into mRNA. Translation in which the information of mRNA sequence is translated into amino acid which is protein made. Heretofore messages of varying length begin and end with domains termed as s and e respectively which are nothing but predetermined terminators. The coded binary strands are in the form of s{0 | 1}e. Two different types of partially double stranded DNA oligonucleotides with sticky ends are used for representation of 0-DNA bit and 1-DNA bit. Terminator domains also have sticky ends.

DNA binary strands are formed by repeated concatenation of the oligonucleotides encoding bits through the complementary sticky ends. Many biological operations can be done on DNA molecules which will aid us in solving mathematical and computational problems. Some of the arithmetic and logical operations performed on DNA are as follows. Addition and subtraction are the basic arithmetic operations which can be applied on DNA nucleotides. DNA nucleotides have nitrogenous bases adenine-A, thymine-T, cytosine-C and guanine-G represented as 00, 01, 10 and 11 respectively.

Addition and subtraction can be applied on DNA nucleotide’s bases according to the binary rules for example binary addition of 10 and 01 will be 11 similarly addition of C and T will be G. If 00 is subtracted from 10 then the result will be 10. Similarly, A is subtracted from C the result will be different logical operations can be implemented on DNA sequence. When there is a set of A, T, G and C characters, they can assign some decimal numbers based on a gene sequence database that contains reference to DNA sequence. There are millions of sequences that can be used freely, therefore the chance of finding the right sequence is almost non-existent.

DNA Cryptography is a rapid emerging technology which works on concepts of DNA computing. DNA stores a massive amount of information inside the tiny nuclei of living cells. It encodes all the instructions needed to make every living creature on earth. The main advantages of DNA computation are miniaturization and parallelism of conventional silicon-based machines. For example, a square centimeter of silicon can currently support around a million transistors, whereas current manipulation techniques can handle to the order of 1020 strands of DNA. DNA, with its unique data structure and ability to perform many parallel operations, allows one to look at a computational problem from a different point of view.

A simple mechanism of transmitting two related messages by hiding the message is not enough to prevent an attacker from breaking the code. DNA Cryptography can have special advantages for secure data storage, authentication, digital signatures, steganography, and so on. DNA can also be used for producing identification cards and tickets. “Trying to build security that will last 20 to 30 years for a defense program is very, very challenging,” says Benjamin Jun, vice president and chief technology officer at Cryptography Research. Multiple studies have been carried out on a variety of biomolecular methods for encrypting and decrypting data that is stored as DNA. With the right kind of setup, it has the potential to solve huge mathematical problems.

A gram of DNA contains 1021 DNA bases = 108 Terabytes of data.

2.3 DNA Conversion

A DNA sequence is selected as a key and gathered in blocks where each block contains four characters. A table is formed on the basis of how the characters occupy the block positions. Sequence is plotted using DNA Mapping as shown in Table 2.1. Finally, from this table, the randomly selected DNA sequence gets converted into an encrypted form. The cipher sequence and key is transferred to the receiver through the communication medium. The DNA sequences are decoded by the following Table. The reverse steps are applied to get back the same message. This algorithm is somehow different from others since traditional mathematical operations or data manipulation techniques are not used. Hence this method cannot be applied for multilevel security.
Table 2.1 DNA Mapping

<table>
<thead>
<tr>
<th>DNA component</th>
<th>Binary equivalent</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>00</td>
</tr>
<tr>
<td>C</td>
<td>01</td>
</tr>
<tr>
<td>G</td>
<td>10</td>
</tr>
<tr>
<td>T</td>
<td>11</td>
</tr>
</tbody>
</table>

A DNA based symmetric key cryptography for secure data transfer over the communication channel. Firstly, Input data, image or text converted into ASCII value, ASCII value is transformed into its binary form, Binary form transformed into DNA code then this code is randomly allocated based on a private key and is converted to the extended ASCII code.

At last, the input is encrypted using DNA code, and clinical permutation is done with the private key. Java is used to implement this modern symmetric key encryption technique. A DNA chromosome is required for data transmission over the communication medium.

DNA

TACATCAGTGCCACCTAGCTGC

Binary

11000100110100101101110011000010111001001111001

Hexadecimal

3010310231320113021321

2.4 Working

A DNA signal generator is used to generate a sample input of DNA. A DNA format that consists of A, C, G, T are mapped to 00, 01, 10, 11 respectively using python. Then again this obtained sequence is converted to decimal format and then to hexadecimal format.

A random key is also generated using a pseudo random generator for increasing the security of the key. This generated key can be used at both Encryption and Decryption processes. Random key generated is then converted to hexadecimal format to give it as input to the algorithm.

The AES algorithm takes inputs as clk, data i.e., hexadecimal DNA, and key. This DNA input is simulated through a series of operations like Add round key that XORs 128 bit with 128 bit key to give the resultant 128 bit. Then Byte substitution is performed where every byte is replaced with another byte. Next shift row operation is performed and it is given to mix column operation where 128 bits are multiplied to a predefined matrix generated from an algebraic function. Once again, the Add Round key is performed. This is repeated for 9 rounds and the 10th round doesn’t have a mix column operation to increase security. A cipher text is obtained at the end of encryption.

This cipher text is processed through the decryption process which is a whole inverse of the encryption process. To obtain a decipher text i.e., hexadecimal output. This hexadecimal output is given to the python compiler which will perform the inverse operation to obtain the required DNA output.

3 AES

Cryptography is one in all the foremost significant and popular techniques to secure the information from attackers by using two vital processes that are Encryption and Decryption. Encryption is that the method of encoding data to prevent intruders from reading the primary data easily. This stage has the flexibility to convert the initial data (Plaintext) into an unreadable format referred to as Cipher text. The following process that has got to be administered by the authorized person is Decryption. Decryption is contrary to encryption. It’s the
method to convert cipher text into plain text without missing any words within the original text. To perform this process cryptography relies on mathematical calculations together with some substitutions and permutations with or without a key. There are variety of algorithms available to encrypt and decrypt sensitive data which are typically divided into three types. First one is symmetric cryptography, the identical secret’s is used for encryption and decryption data. Second is Asymmetric cryptographic. This kind of cryptography relies on two different keys for encryption and decryption. Finally, a cryptographic hash function using no key instead key is mixed with the information. The symmetric key is rather more effective and faster than Asymmetric. A number of the common symmetric algorithms are Advanced Encryption Standard (AES), Blowfish, Simplified Data Encryption Standard (S-DES) and 3DES.

### 3.1 Basic structure of AES

The Advanced Encryption Standard (AES) algorithm is one of the block cipher encryption algorithms was published by National Institute of Standards and technology (NIST) in 2000. The main aim of this algorithm was to replace the DES algorithm after some vulnerable aspects of it. NIST invited experts who work on encryption and data security all over the world to introduce an innovative block cipher algorithm to encrypt and decrypt data with powerful and complex structure. From around the world many groups submitted their algorithm. NIST accepted five algorithms for evaluation. After performing various criteria and security parameters, they selected one of the five encryption algorithms proposed by two Belgian cryptographers Joan Daeman and Vincent Rijmen. The original name of the AES algorithm is the Rijndael algorithm. However, this name has not become a popular name for this algorithm, instead it is recognized as the Advanced Encryption Standard (AES) algorithm around the world.

AES has the ability to deal with 128 bits (16 bytes) as a fixed plaintext block size. These 16 bytes are represented in a 4x4 matrix and AES operates on a matrix of bytes. In addition, another crucial feature in AES is the number of rounds. The number of rounds are based on the length of the key. There are three different key sizes used by AES algorithm to encrypt and decrypt data such as (128, 192 or 256 bits). The AES parameters are shown in Table 3.1.
AES is an iterative rather than Feistel cipher. It is based on ‘substitution–permutation network’. It comprises of a series of linked operations, some of which involve replacing inputs by specific outputs (substitutions) and others involve shuffling bits around (permutations). AES performs all its computations on bytes rather than bits. Hence, AES treats the 128 bits of a plaintext block as 16 bytes. These 16 bytes are arranged in four columns and four rows for processing as a matrix. Unlike DES, the number of rounds in AES is variable and depends on the length of the key. AES uses 10 rounds for 128-bit keys, 12 rounds for 192-bit keys and 14 rounds for 256-bit keys.

4 Software description

The AES Algorithm for secure is implemented by using Xilinx ISE Design Suite and code is written in Verilog HDL.

Python

The python language is one of the most accessible programming languages available because it has simplified syntax and not complicated, which gives more emphasis on natural language. Due to its ease of learning and usage, python codes can be easily written and executed much faster than other programming languages.

Microsoft Excel

Microsoft Excel is a spreadsheet developed by the Microsoft for Windows, mac OS, Android and iOS. It is a spreadsheet program, which contains columns and rows, the intersection of a column and a row is called cell. It features calculation, graphing tools, pivot tables, and a macro programming language called Visual Basic for Applications (VBA). Scatter plot in MS excel is used to plot the ECG graph from the data.

Verilog HDL

Verilog is a Hardware Description Language (HDL). It is a language used for describing a digital system like a network switch or a microprocessor or a memory or a flip–flop. It means, by using a HDL we can describe any digital hardware at any level. Designs, which are described in HDL are independent of technology, very easy for designing and debugging, and are more useful than schematics, particularly for large circuits.

5 Results

Python compiler is used to generate random DNA signals and convert them to binary by using the DNA Mapping Table 2.1. And then it is converted to hexadecimal as Xilinx accepts only hexadecimal or binary. A random key is also generated using a Pseudorandom generator to increase the security as shown in Fig. 5.1.

![Fig. 5.1 Output of Random DNA Generator](source:image)
Hexadecimal output that is obtained is given to the python compiler to get the required DNA input. The output of decrypted DNA is obtained as shown in Fig. 5.4. The DNA output obtained in Fig. 5.4 is same as the DNA input generated as shown in Fig. 5.1.

![Decrypted DNA](https://example.com/decrypted_dna.png)

**Fig. 5.4 Output of Decrypted DNA**

DNA input that is generated is given to the AES algorithm. Three samples are taken from a random generator and simulated as shown in Fig. 5.2.

The DNA inputs are encrypted to give cipher text and then it is decrypted to give the hexadecimal value.

![Simulated Outputs of AES Algorithm](https://example.com/simulated_outputs.png)

**Fig. 5.2 Simulated Outputs of AES Algorithm**

**Conclusion** Cryptography is utilized to secure information while it is being conveyed between two parties or while it is put away in a medium defenseless against physical burglary. Correspondence security gives assurance to information by enciphering it at the transmitting point and deciphering it at the receiving point. DNA (Deoxyribonucleic Acid) contains sensitive information regarding a person's genetic information hence need to be protected while transmission. This project implementation is based on mathematical properties of AES algorithm and based on DNA cryptography. A Symmetrical AES with 128 bit key which is randomly generated is used for encryption and decryption process during transmission of DNA. Encryption system is designed by using Byte Substitution, Shift rows, Mixed Column, Add Round Key, which produces cipher text which is given as input to the decryption process. Decryption processes are designed by using the inverse of the encryption process. The input passed is accurately decrypted and the given DNA sequence is retrieved successfully.
REFERENCES


