A NEW DNA BASED APPROACH OF GENERATING KEY-DEPENDENTMIXCOLUMNS TRANSFORMATION

Auday H. Al-Wattar[1], Ramlan Mahmod[2],Zuriati Ahmad Zukarnain[3] and NurIzura Udzir[4]

[1]Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor.

ABSTRACT

The use of key-dependent MixColumns can be regarded as one of the applied techniques for changing the quality of a cryptographic algorithm. This article explains one approach for altering the MixColumns transformation engaged in the AES algorithm. The approach employed methods inspired from DNA processes and structure, which relied on the key.The parameters of the proposedMixCloumns have characteristics identical to those of the original algorithm AES besides increasing its resistance against attack.The original transformation uses single static MDS matrix while the proposed methods used dynamic MDS. The security of the new MixColumns was analyzed, and the NIST Test Suite tests were used to test the randomness for the block cipher that used the new transformation.

KEYWORDS

MixColumns, Block cipher, DNA, AES, NIST, MDS

1.INTRODUCTION

Recently, the need for anextremelyefficientapproach to attaininformation security is critical and vital. Cryptography has been and stays behind the most proficient approach employed to achieve security. Rijndael is a symmetric block cipher that was selected by (NIST) (National Institute of Standards and Technology), [1] in 2001 as (advanced Encryption Standard, FIPS 197) AES.

In general, it is based on repeated rounds of transformation that alter input plaintext to encryptedtext or ciphertext output. Each round consists of several functions and always includes a depending on the roundsecretkey. Multiple roundsestablish inverse transforming ciphertext into the original, using the same secret key. AES has 128 –bit block size, and a key length is 128, 192 or 256 bits, relying on to the number of rounds for the algorithm. It is using a byte array known asthe state of (4x4)size in each cycle of the encryption / decryption process. The majority of the algorithm calculations is accomplished in finite fields [2,3].

At AES, the MixColumns transformation is the most important function within the linear unit of symmetric encryption algorithms.Itis themajor source of diffusion in the AES block cipher. Each column is dealt with as a polynomial through GF (2^8), and then modulo$x^4 + 1$is multiplied by a fixed polynomial c (x) =$3x^3 + x^2 + x + 2$.The inverse of this polynomial is $c^{-1}(x) = 11x^3 + 13x^2 + 9x + 14$. The MixColumnsprocedure can be executed by

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 2 & 3 & 1 & 1 \\ 1 & 2 & 3 & 1 \\ 1 & 1 & 2 & 3 \\ 3 & 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

multiplying a coordinate vector of four numbers in Rijndael's Galois field by the following circulated MDS matrix:

The multiplication operation is performed as complicated operation using the multiplied while the addition operation is a simple XOR operation, as the math of the operation is madein Rijndael's Galois field. Figure 1 shows the AES MixColumntransformation[4].

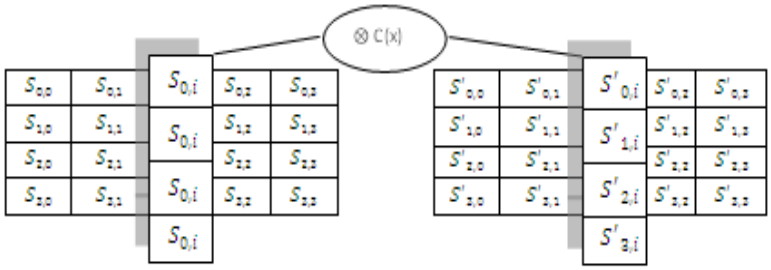

Figure 1: AES MixColumn transformation

The linear and differential cryptanalysis isthecriticall,demandingg for AES algorithm, where, it can beissuedswitho standard techniques of differential and linear cryptanalysis. From the analysis of resistance to differential and linear cryptanalysis, it was deduced that the random, unknown and key-dependent permutation transformation is considered as a good feature in enhancing the resistance of block cipher against the differential and linear attacks, since these attacks are need known transformations. The detailed properties of substitution and permutation functions, specifically the structure that is completely dynamic and unknown to the cryptanalyst, assist and support the block cipher to be resistance to attacks. For the attacker, differential and linear trail over numerous rounds is regarded as afundamental requirement, and in the reality of key-dependent transformations output differences rely on additional key used. In addition, for any additional different key values, there are unrelated differentials over multiple rounds and therefore diverse linear differential trails. As a result, it is difficult for the attacker to utilize the current linear and differential techniques of cryptanalysis.

Although many previous works on enhancing the security of the AES block cipher against attacks, no singlework has proposed a MixColumns that is designed or created using DNA bio-inspired techniques.

This paper proposed a new technique for obtaining a powerful key-dependent MixColumnsdepending on operations that have been inspired from really biological DNA processes. Subsequently, it tested the new transformationusing the NIST randomness tests for the

cryptosystem that with the DNA-based MixColumn and performed a security analysisof the proposed transformation.

2. DNA BACKGROUND

DNA (Deoxyribo Nucleic Acid) is a molecule that characterizes the genetic material for every living organism. It is assumed as the genetic drawing of living or existing creatures. A single-strand of DNA consists of a chain of molecules known as bases, defined as four letters {A, C, G, and T}[5,6]. One of the most fundamentalcharacteristics of the DNA strand series is that it is oriented; as a result, ATCGTACT is distinct from TCATGCTA.

The DNA strands exist as pairs as (G) associated with (C) and (A) associated with (T) figuringcomponents named base pairs. The reverse DNA strands representing the opposite of the strand bases; for example, AGCTAGGCATAA becomes AATACGGATCGA, while the complement of the strands can be represented as $A^- \equiv T$ and $C^- \equiv G$. Therefore, GCATAA will become TTATGC[7,9]. The DNA segment consists of two parts named as Exon, and Entron respectively.

CENTRAL DOGMA

The central dogma process of the DNA strand bases is one of the main methods that characterize the biology DNA system. It comprises of operations on DNA and RNA (Ribonucleic acid), including transcription, translation, and replication. The DNA segment consists of parts: Coded parts (Exon) and non-coding parts (Intron).

The process of the central dogma of molecular biology can be described in Figure 2, in which the genes' information glides into proteins [10].

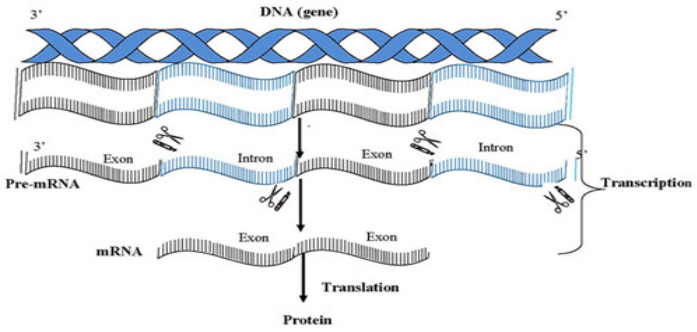

FIGURE 2. DNA BASES REVERSE COMPLEMENT TECHNIQUE

- According to Figure 2, there are two methods within the central dogma as transcription and translation which achieved to obtain the Protein. The transcription methods are performed by removing the non-coding parts (Introns) and keeping the coding part (Entrons), while the translation methods including converting the RNA as MRNA into Proteins.

In this paper, the creation of a key-dependent MixColums is based on and inspired by transcription process.

3. THE PROPOSED METHOD

In this paper the proposed MixColumns transformation will refer to as a key-dependent MixColumns transformation (KdM_T_r).

This (KdM_T_r), willrely on the key cipher (K_r), which is differentin each round (r).

In the original AES, the MDS matrix of the MixColumns transformation remainspermanent for all rounds, whilethe proposed method the MDS matrix will be changed for each round according to the key cipher.

To explain the proposed method, we first used the AES original MDS matrix as a base matrixas shown in Figure 3.

2	3	1	1
1	2	3	1
1	1	2	3
3	1	1	2

Figure 3: A 4×4 MDS

There are two ways to use the key *Kr* for the (KdM_T_r), The first one is to use four bytes of the key such that have taken the values ranging between 0, and 3, while the second way is to use only one byte of the key. In the both ways the key values are stored within*MOK* matrix.

The applying of *MOK* on the MDS matrix to generate a new MDS matrix for each round is inspired by the real DNA transcription process as shown in Figure 2. According to this procedure the skipped columns will be regarded as Introns while the left behind columns will be considered as *Exon*. The values of *MOK* will state the number of columns that will be skipped representing the number of *Intron*. The generating method which called columns transcription is performed by skipping a number of columns of the MDS matrix according to the amount exist in the *MOK*

First:*MOK* with four bytes (values):

For the first way (four bytes), the following procedure will be performed to obtain the (KdM_T_r),

The work of this process will be done according to the key stored in MOK, and the values of the MDS matrix‘s first row.

The columns of this matrixwill be shifted according to the key values as following:

If the key value = 1, then the row’s values of theorigin MDS matrix will be shifted one positionto the right.

If the value of the key = 0, then the row’svalues of MDS will not change.

If the key value = 2, thenthe valuesof MDS rows willbe shifted to the right 2 positions.

Finally, if the value of the key = 3, then the MDSrow values will be shifted 3 positions.
Figure 4, demonstrates the whole process. Since the values of the key are changingat each round, this will lead generating new different MDS for MixColumns transformation for every round.

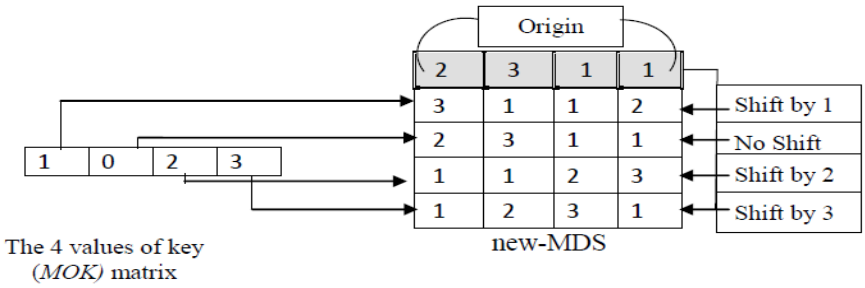

Figure 4: Generating new key-dependent MDS matrix *(MOK)* matrix

Second:*MOK* with one byte (value):

In this case thevalue within the *MOK*(which could be 0, 3) as one byte of the key will specify a certain *MDS* matrix as showing in Figure 5,

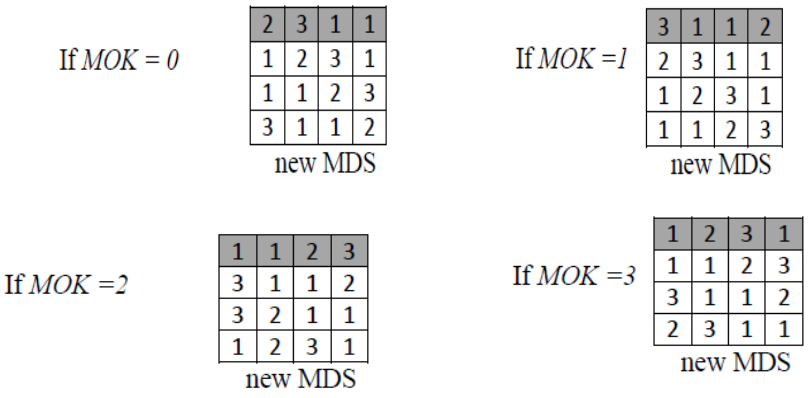

Figure 5: Generating new key-dependent MDS matrix (MOK) matrix

The value of the key in MOK matrix indicates the number of columns of the first row of MDS that should be skipped, referring to biology, DNA concept the skipped columns are considered as (Intron) which are removed, while the remaining columns are considered as (Exon) through the transcription/ splice process

4. TEST AND RESULT

4.1 Correlation coefficient

The correlation coefficient involves values ranging from (-1 and +1). According to[11] the following values are in a good range for explicating the correlation coefficient as stated in Table 1.Note that this paper considers the two variables of plaintext *(p)* and ciphertext *(c)*

Table 1. Accepted range values for interpreting the correlation coefficient

Value	Meaning (state)
-1	Perfect negative linear relationship : When *p* increases in its values, *c* decreases in its values by an exact linear rule
0	Non-Linear relationship
+1	Perfect positive linear relationship
(0,0.3) ,(0,-3)	Weak positive (negative) linear relationship by unstable linear rule.
(0.3,0.7) ,(0.3,-7)	Moderate positive (negative) linear relationship.
(0.7, 1), (-0.7,-1.)	strong positive (negative) linear relationship

Equation 1illustrates the applying of the correlation coefficient functions:

$$E(c) = \frac{1}{s}\sum_{i=1}^{s} p_i \qquad (1)$$

where: s is the entire number of bits, p_i,c_i are the chains of s measurements for p and c, p is bits value of plaintext, c is bits value of ciphertext, E(c) is mathematic anticipation of c.

The variance of p can be expressed by equation 2,

$$D(p) = \frac{1}{s}\sum_{i=1}^{s} [p_i - E(p)]^2 \qquad (2)$$

Lastly, the related coefficients r_{pc} can be expressed by equation 3,

$$r_{pc} = \frac{E\{[p - E(c)][c - E(c)]\}}{\sqrt{D(p)}\sqrt{D(c)}} \qquad (3)$$

The experiment examined theAES block cipher using the new (KdM_T_r). The Scatter chart of the results is presented in Figure 6. It illustrates that the majority of correlation values, at different rounds through the whole block cipher are near to 0, which indicates a strong positive (or negative) non-linear relationship. Only 0.007% are near to +1 or -1, in round 1, representing a weak positive (or negative) non-linear relationship. From the results, it can be inferred that the block cipher has an improved confusion performance.

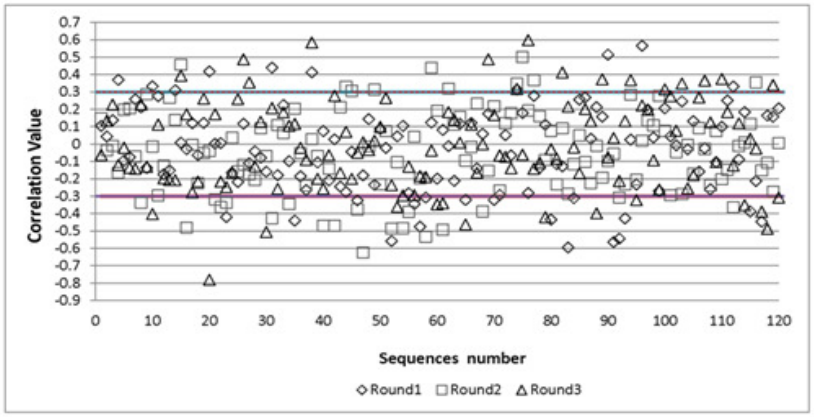

Figure 6. Scatter chart of the correlation test results on whole block cipher

4.2 NIST Suite Randomness test

The randomness test is one of the security analysis to measure the confusion and the diffusion properties of the new encryption algorithm, as carried out in [12-14][15] and[16].

NIST Suite [17] is a statistical test suite for randomness by NIST used to assess the cryptographic algorithm. The suite test assesses whether the outputs of the algorithms under certain test condition exhibit properties that would be expected randomly generated outputs.

In order to evaluate the randomness of ciphertext, an experiment that included a set of data as random plaintext and random 16 byte key in the ECB mode was conducted. The 100 sequences of 1,059,584 bits were constructed and examined. The list of statistical tests applied throughout the experiments is illustrated in Table 2.

Table 2. Breakdown of 15 statistical tests applied during experimentation

Test ID	NIST Statistical Test	Number of p-
1	Frequency	1
2	Frequency-Within-Block	1
3	Runs	1
4	Longest Runs of Ones	1
5	Binary Matrix Rank	1
6	Discrete Fourier	1
7	Non-Over Lapping Template	1
8	Overlapping Template	1
9	Maurer's Universal	1
10	Linear Complexity	1
11-12	Serial	2
13	Approximating Entropy	1
14-15	Cumulative Sums (Cusums)	2
16-23	Random Excursions	8
24-41	Random Excursions Variant	18

The proportion of sequences that passed a specific statistical test should lie above the proportion value p, defined in Equation (4).

$$p_{\propto} = (1-\propto) - 3\sqrt{\frac{\propto(1-\propto)}{n}} = \quad (4)$$

Where$\propto$ is the significant value, n is the number of testing sequences.

For this experiment, the proportion value is:

$$p_{\propto} = (1-0.01) - 3\sqrt{\frac{0.01(1-0.01)}{100}} = 0.960150 \quad (5)$$

Where n= 100, and $\propto$=0. 01.

The p-value readings for each round constructed is illustrated in Figure 7. This figure demonstrates the randomness test for 15 statistical tests of block cipher that used the proposed MixColumnsfor three rounds. From this figure, at the end of the second and third rounds, all of the 41 statistical tests fall over 96.0150%, which is evidence that the output of the algorithm is completely random.

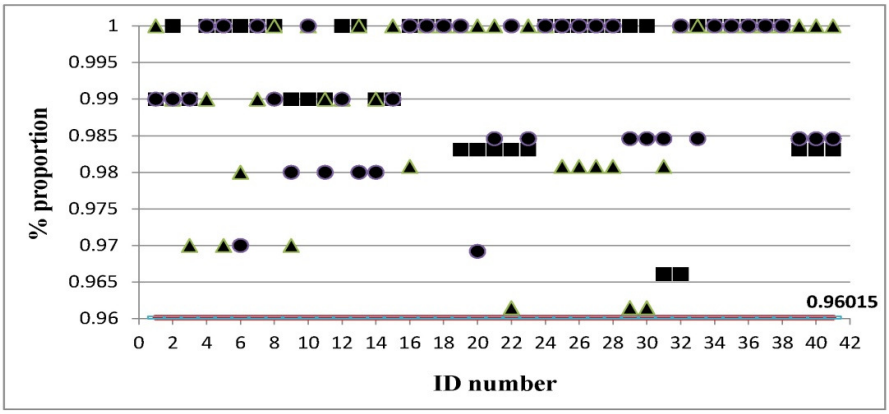

Figure 7. Randomness tests results of whole block cipher for 3 rounds

4.3 CRYPTANALYSIS

The new MixColumns transformation is dynamic and its changing at each round according to round key-value this feature make the job harder for the attackers since the analysis of dynamic unit is more difficult than the static one. For the dynamic MixColumns, the attacker has the possibility of 2^{n},which consider a high number, that increasing the resistance of the algorithm against the attacks.The AES security factors plus the new security features applied by the proposed (KdM_T_r), are worked to enhance security of the block cipher.

5. CONCLUSION

In this paper a new dynamic key-dependent MixColumns transformation for AES block cipher was proposed with chosen byte or four bytes of the private key. The new transformation is not fixed, but changeable at each round according to the round key values. The MDS matrix of the MixColumns is changed according to this key. This unit was tested with the correlation coefficient and the 15 statistical randomness tests of NIST Test Suite. The analyzing of the obtained results showed that, this new transformation is more secure and resistance against the attacks, on which concluded that it is potential to employ it for an encryption. This will increases the stability of AES against linear and differential cryptanalysis.

REFERENCES

[1] J. Daemen and V. Rijmen, "AES proposal: Rijndael," in First Advanced Encryption Standard (AES) Conference, 1998.

[2] V. Rijmen and J. Daemen, "Advanced Encryption Standard," Proceedings of Federal Information Processing Standards Publications, National Institute of Standards and Technology, pp. 19-22, 2001.

[3] L. Information Technology, Announcing the Advanced Encryption Standard (AES) [electronic resource]. Gaithersburg, MD :: Computer Security Division, Information Technology Laboratory, National Institute of Standards and Technology, 2001.

[4] P. FIPS, "197: Advanced encryption standard (AES)," National Institute of Standards and Technology, 2001.

[5] M. Zhang, M. X. Cheng, and T.-J.Tarn, "A mathematical formulation of DNA computation," NanoBioscience, IEEE Transactions on, vol. 5, pp. 32-40, 2006.

[6] M. Zhang, C. L. Sabharwal, W. Tao, T.-J.Tarn, N. Xi, and G. Li, "Interactive DNA sequence and structure design for DNA nanoapplications," NanoBioscience, IEEE Transactions on, vol. 3, pp. 286-292, 2004.

[7] E. R. Kandel, "The molecular biology of memory storage: a dialogue between genes and synapses," Science, vol. 294, pp. 1030-1038, 2001.

[8] G. I. Bell and D. C. Torney, "Repetitive DNA sequences: some considerations for simple sequence repeats," Computers & chemistry, vol. 17, pp. 185-190, 1993.

[9] A. G. D'yachkov, P. L. Erdös, A. J. Macula, V. V. Rykov, D. C. Torney, C.-S. Tung, et al., "Exordium for DNA codes," Journal of Combinatorial Optimization, vol. 7, pp. 369-379, 2003.

[10] F. Crick, "Central dogma of molecular biology," Nature, vol. 227, pp. 561-563, 1970.

[11] K. Wong, "Interpretation of correlation coefficients," Hong Kong Medical Journal, vol. 16, p. 237, 2010.

[12] F. Sulak, A. Doğanaksoy, B. Ege, and O. Koçak, "Evaluation of randomness test results for short sequences," in Sequences and Their Applications–SETA 2010, ed: Springer, 2010, pp. 309-319.

[13] Q. Zhou, X. Liao, K.-w.Wong, Y. Hu, and D. Xiao, "True random number generator based on mouse movement and chaotic hash function," information sciences, vol. 179, pp. 3442-3450, 2009.

[14] V. Patidar, K. K. Sud, and N. K. Pareek, "A Pseudo Random Bit Generator Based on Chaotic Logistic Map and its Statistical Testing," Informatica (03505596), vol. 33, 2009.

[15] J. Soto and L. Bassham, "Randomness testing of the advanced encryption standard finalist candidates," DTIC Document2000.

[16] V. Katos, "A randomness test for block ciphers," Applied mathematics and computation, vol. 162, pp. 29-35, 2005.

[17] E. B. Smid, S. Leigh, M. Levenson, M. Vangel, A. DavidBanks, and S. JamesDray, "A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications."

Authors

AudayH.Al-Wattar obtained his B.Sc. degree in Computer Science from Mosul University and his M.Sc. Degree from Technology University-Baghdad in 2005 .He presently pursing his Ph.D. From Universiti Putra Malaysia, Malaysia under the guidance of Prof. Dr. Ramaln bin Mahmod at Computer Science and Information Technology Faculty. He works as lecturer at Mosul University (since 2005), at Computer Science and Mathematics Faculty - Computer Science Department. His area of interest includes Computer Security, Programming languages.

Ramlan B Mahmodobtained his B.Sc. in computer Science, from Western Michigan, University, U.S.A. in 1983, and M.Sc. degree in Computer science, from Central Michigan, University, U.S.A. and Ph.D. degree in Artificial Intelligence from, Bradford University, UK in 1994. His previous workings Experience/Position are as:

System Analyst, PETRONAS,1979-1980. **Lecturer,**Mathematic Department, Faculty of Science, UPM, 1985-1994. **Lecturer,**Department of Multimedia, Faculty of Computer Science & Information Technology, UPM, 1994 – 2002. **Associate Professor,**Department of Multimedia, Faculty of Computer Science & Information Technology, UPM, 2002 – 2010.

Deputy Dean,Faculty of Computer Science & Information Technology, UPM, 1998 – Nov 2006.**Dean,**Faculty of Computer Science & Information Technology, UPM, 2010-2013. Professor, Faculty of Computer Science & Information Technology, UPM, 2012 - now. His research interest includes Neural Network, Artificial Intelligence, Computer Security and Image Processing.

NurIzurabintiUdzirobtained her B.Sc. in computer Science, from UniversitiPertanian Malaysia, in 1995, and M.Sc. degree in Computer science, from Universiti Putra Malaysia, in 1998, and Ph.D. degree in Computer Science from, University of York, UK in 2006.

Associate Professor,Head Department of Computer Science, Faculty of Computer Science & Information Technology, UPM. Her research interest includes Computer security, secure operating systems, Access control, Distributed systems, Intrusion detection systems.

ZuriatiBinti Ahmad Zukarnain obtained her B.Sc. in Physics,, from Universiti Putra Malaysia, in 1997, and M.Sc. degree in Information Technology, from Universiti Putra Malaysia, in 2000, and Ph.D. degree in Quantum Computation and Quantum Information from, University of Bradford, UK, 2006,UK in 2006.

Associate Professor, Department Of Communication Technology and Networking, Faculty of Computer Science & Information Technology, UPM. Her research interest includes information, computer and communication technology (ICT), Quantum Information Systems and Distributed Systems, Quantum Computing, Computer Networks and Distributed Computing.