

Secure Electronic Medical Records Transmission using NTRU Cryptosystem and LSB in Audio Steganography

Adamu Abdulkadir, Shafi'i Muhammad Abdulhamid, Oluwafem Osho, Ismaila Idris and John K Alhassan
Department of Cyber Security Science, Federal University of Technology, Minna, Nigeria.

Email: *abdulcybersec2015@gmail.com, shafii.abdulhamid@futminna.edu.ng, femi.osho@futminna.edu.ng, ismi_idris@yahoo.co.uk, jkalthassan@futminna.edu.ng*

Abstract - Electronic medical records (EMR) are vital information, extremely sensitive private data in healthcare, and need to be frequently shared via the internet among peers. One of the major benefits derived from the internet is the ease of sending information from one system to another, irrespective of the location or distance between the nodes. This, and many other related important functionality, over the years, has attracted attackers who dedicate themselves to breaching the integrity, availability and confidentiality of EMR information. Existing literature have proposed cryptographic techniques that are not quantum-safe. In this paper, an audio-based system for hiding EMR information using a quantum-safe cryptographic technique, Nth degree Truncated Polynomial Ring Units (NTRU) cryptosystem, and the Least Significant Bit (LSB) steganographic technique is proposed. The system was evaluated based on embedding capacity (EC), peak signal to noise ratio (PSNR), mean square error (MSE), and histogram plots. Results showed our proposed system is able to securely hide the medical records without causing significant distortions in the original audio.

Keywords - Electronic medical records (EMR), Information hiding, Security, Cryptography, Steganography, LSB, NTRU

I. INTRODUCTION

Electronic medical records (EMR) are very delicate private records for diagnosis and treatment in healthcare, which need to be regularly shared among medical personnel in both rural and urban settings such as healthcare providers, insurance companies, pharmacies, researchers, patient's families, among others. This poses a major challenge on keeping a patient's medical history up-to-date and most at times private. Transmissions of EMR information are mostly done using the cyber space or wide area networks which are prone to attacks.

Since the advent of the internet, its capacity and sophistication have continued to advance. From a platform used primarily for displaying static web pages, it has become a tool for dynamic exchange of EMR data and information. Today, the internet not only serves as a repository of EMR information, but also, among other purposes, provides functionality for exchange of information among different medical personnel.

Unfortunately, the current architecture of the internet does not support security [1]. Attackers exploit this inherent weakness to perpetrate different attacks which target EMR data. Hence, the need for security of EMR data online cannot be over-emphasized. One method of protecting EMR data online is data hiding.

EMR data hiding simply involves embedding EMR data in different media. EMR data, such as text, images, videos, or audio can be concealed in a media, for security purpose. Techniques used for hiding EMR data are watermarking, steganography, and cryptography [2]. Watermarking is essentially used to indicate ownership of an object [3, 4]. Steganography is used to secure EMR data transmission. A message is usually hidden in another message to make it imperceptible to unauthorized entities [5, 6].

One shortcoming, however, with strictly relying on steganography is its vulnerability to steganalysis [7]. With steganalysis hidden messages can be detected [8, 9]. One solution is to encrypt the message before embedding it in a media. Cryptography scrambles messages to render them unintelligible to unauthorized entities. This ensures that even if the attacker discovers the hidden message the actual content of the message is not decoded.

In this study, we propose a security enhancing EMR data hiding system that leverages cryptography and steganography, specifically, the Nth degree TRUncated Polynomial Ring Units (NTRU) and Least Significant Bit (LSB) respectively.

The rest of the paper is organized as follows: in section II, we review some related studies. The methodology used in the study is discussed in the next section. Section IV presents the implementation of the system and results of the evaluation. The study is concluded in section V.

II. RELATED WORKS

Different cryptographic and steganographic techniques have been proposed by authors. In the choice of steganographic techniques, few authors considered the use of transform domain techniques. These include Discrete Wavelet Transform (DWT) [10], and Discrete Cosine Transform (DCT) [11]. However, most studies employed spatial

domain techniques, with LSB as the most common. In the study by [12], a Hash Least Significant Bit (H-LSB) was proposed. This entails the use of hash function to determine the position of insertion in the LSB. [13], in their own work, combined the use of Pixel value differencing (PVD) and LSB to embed messages in truecolor RGB images. PVD helps to determine the size of the secret message embeddable in a pixel, using the difference between two consecutive pixels [14]. With this method, the stego-image can hide much larger information, whilst still maintaining good visual quality [15, 16].

To encrypt the messages before they are embedded into the various media, most studies seem to favor the use of symmetric techniques. Some of the techniques proposed are AES [17-22], DES [23], Blowfish [24], and Affine Cipher [25]. Saraireh [26] proposed the use of the filter bank cipher over Galois field ($GF(2^8)$), to improve the resistivity of the cipher against cryptanalysis attacks, specifically, differential and linear attacks. Usha, Kumar and Boopathybagan [27] proposed an information hiding systems that implements double layer of EMR data encryption. The message is first encrypted using the Playfair cipher. The ciphertext is then encrypted using AES. Hayfaa, Ahmad and Noor [28], in their study, designed a simple substitution cipher which represents each character by five bit. It is essentially based on substituting characters in the English language with a code number.

One of the few studies that involved the use of asymmetric cryptographic scheme is [29]. The study explored the effects of two encryption schemes, RSA and DHA, on time complexity. Their results showed that while the use of RSA increased the time complexity in steganalysis, the Diffie Hellman Algorithm did not. Table 1 presents a summary of some related studies.

The objective of combining steganography with cryptography is to enhance the security of the hidden message, to the effect that even if the hidden message is discovered its true contents remain unreadable to the attacker. The level of security will therefore be dependent on the strength of the cryptographic scheme employed. Most existing studies proposed modern cryptographic techniques. However, these techniques are vulnerable to many attacks including brute-force, known plaintext, chosen ciphertext attacks [30].

One other issue borders the fact that most asymmetric cryptographic schemes are based essentially on either integer factorization or discrete logarithms. These classes of problems, unfortunately, can be solved quickly by quantum computers that employ quantum algorithms. The implication is that, the capacity of the techniques to secure encrypted information cannot be guaranteed [31]. There is therefore need for cryptographic techniques that are

quantum-safe. This is the major contribution of this research.

Table 1. Review of related literatures

No.	Authors (Year)	Steganography		Cryptograph y Technique
		Type	Tech nique	
1.	Abdullah & Aziz (2016)	Image	H-LSB	Affine Cipher
2.	Garg & Kaur (2016)	Image	LSB	AES
3.	Reddy & Kumar (2016)	Image	LSB	AES
4.	Sethi & Kapoor (2016)	Image	LSB	AES
5.	Deshpande, Fusate, Malviya, & Dhyavartiwar (2015)	Audio	LSB	RSA
6.	Hayfaa et al. (2014)	Image	LSB	Substitution cipher
7.	Saraireh (2013).	Image	DWT	Filter bank cipher
8.	Singh & Malik (2013)	Image	LSB	Blowfish
9.	Abikoye, Adewole, & Oladipupo (2012)	Audio	LSB	DES
10.	Gupta et al. (2012)	Image	LSB	RSA and DHA
11.	Phad et al. (2012)	Image	PVD and LSB	AES
12.	Usha et al. (2011)	Image	LSB	Playfair cipher and AES
13.	Sarmah & Bajpai (2010)	Image	DCT	AES

III. METHODOLOGY

EMR data hiding system that leverages on cryptography and steganography to secure message is presented. The message is first encrypted using NTRU. The encrypted message is then embedded into a digital audio media using LSB technique before it is transmitted by the sender. The encryption/decryption process by NTRU and embedding algorithm of the LSB technique are presented in the succeeding sections respectively. Figure 1 presents the embedding process. For extraction of the hidden message by the receiver, the process is essentially reversed.

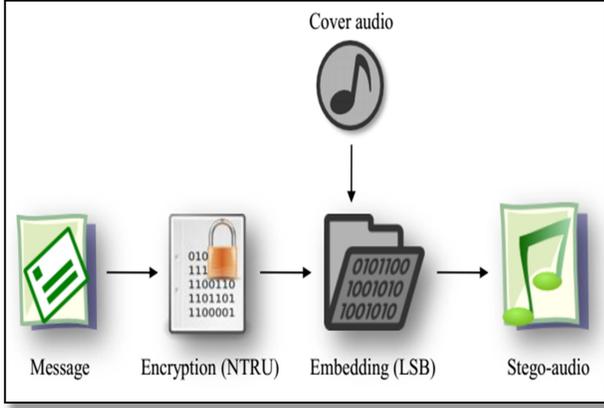


Figure 1. The message embedding process

A. NTRU

NTRU is a ring-based public key cryptosystem proposed by Hoffstein, Pipher and Silverman [32]. It is an efficient and computationally inexpensive cryptosystem, known for its low memory requirement, high speed, moderately, and easily created keys [33]. The technique is secure against brute-force, meet-in-the-middle, multiple transmission, and lattice-based attacks. And it is a quantum-safe cryptosystem.

For an NTRU cryptosystem, we define three integer parameter: (N, p, q) , and four sets of polynomials of degree $N - 1$: $\mathcal{L}_f, \mathcal{L}_g, \mathcal{L}_\phi, \mathcal{L}_m$. We assume that $\gcd(p, q) = 1$, and $q > p$. The notation for the ring is given as: $R = \mathbb{Z}[X] / (X^N - 1)$

We write an element $F \in R$ as:

$$F = \sum_{i=0}^{N-1} F_i x^i \quad (1)$$

We write a star multiplication, denoted by \otimes , which is explicitly a cyclic convolution product.

Key Creation: Bob randomly select polynomials $f, g \in \mathcal{L}_g$. We denote two inverses, F_p and F_q , of polynomial f by:

$$F_p \otimes f \equiv 1(\text{mod } p) \text{ and } F_q \otimes f \equiv 1(\text{mod } q) \quad (2)$$

Bob then computes:

$$h \equiv F_q \otimes g(\text{mod } q) \quad (3)$$

Therefore, the public key of Bob is the polynomial h , while his private key is f .

Encryption: To send a message to Bob, Alice simply selects a message m from the set of plaintexts \mathcal{L}_m , randomly

picks a polynomial $\phi \in \mathcal{L}_\phi$, and then uses the public key of Bob to compute:

$$m_e \equiv p\phi \otimes h + m(\text{mod } q) \quad (4)$$

m_e is the encrypted version of the message Alice finally sends to Bob.

Decryption: For Bob to decrypt the encrypted message m_e from Alice, he computes:

$$a \equiv f \otimes m_e(\text{mod } q) \quad (5)$$

The coefficients of a are chosen from the interval $-q/2$ to $q/2$. He recovers the message via:

$$F_p \otimes a(\text{mod } p)$$

B. LSB

This method is used to hide sequence of binary message in the least significant bit of a digital audio file. This technique capitalizes on the nature of the Human Auditory System (HAS) which does not have the ability to detect slight differences in the audio frequencies, especially when it is more concentrated at the audible spectrum. LSB has the advantage that it allows large amount of information to be hidden without reducing the quality of the audio file. Consequently, it is easy to carry out.

An illustration of how the message 'FUT' is embedded into an audio file using the concept of LSB is presented in Figure 2. Both message and audio file are first converted to bit stream. To convert 'FUT' to binary, we convert the ASCII value equivalent of the different characters in the message to binary. This is presented in Table 2. The least significant (right-most) bit of the audio stream is then replaced with the bit stream of the message.

Table 2 Conversion of message 'FUT' to binary

Character	ASCII Value	Binary Representation
F	70	1000110
U	85	1010101
T	84	1010100

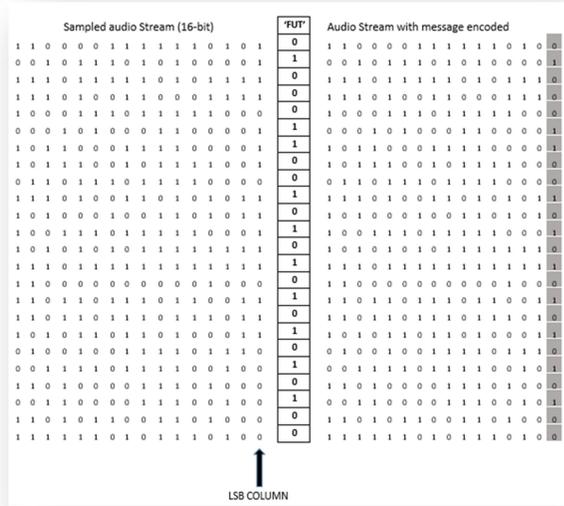


Figure 2. Using LBS technique to hide EMR data in an audio file

C. Performance evaluation metrics

The performance of our proposed system is evaluated by the embedding capacity (EC), mean squared error (MSE), peak signal to noise ratio (PSNR), and histogram plots.

Embedding Capacity (EC): This is the maximum quantity of EMR data that can be embedded into a cover audio without significantly altering the value of the original audio file. It is the fraction of the secret message by the cover audio object.

$$EC = \frac{\text{Secret message size}}{\text{Cover object size}} \quad (6)$$

Mean Squared Error (MSE): Denotes the cumulative square error between the cover audio signal and the stego-audio. When the value of the MSE is low it is better, and therefore the little the error rate between the illustrations which shows little alteration was added. MSE is computed using the formula:

$$MSE = \frac{\sum_{M,N} [I_1(m,n) - I_2(m,n)]^2}{M*N} \quad (7)$$

M and N stand for the rows and columns of the audio samples. I_1 is the stego audio while I_2 is the cover audio

Peak Signal to Noise Ratio (PSNR): It is used to estimate the amount of resemblance that exists between the original audio and the stego-audio. This parameter depends on MSE. It is also referred to as the quality measurement between two or files involved. It is measured in decibels. When the comparison of PSNR is high it means the system is good and this shows that the distortion is low.

$$PSNR = 10 \log_{10} \frac{R^2}{MSE} \quad (8)$$

Where R is the slightest variation of the stego-audio, which is usually 255 in integer EMR data type.

Histogram Plot: This provides a graphical representation of the different amplitude values of the audio signal.

IV. EXPERIMENTAL RESULTS

The proposed NTRU + LSB technique was developed using a PC with the following properties: Pentium (R) CPU T4500 @ 2.30GHz, 4GB RAM, Windows 8.1 pro operating system, and Java programming language (using NetBeans 8.1 platform). Appendices A to D depict some screenshots of system testing.

To evaluate the performance of the proposed technique, five .wav digital audio samples were used. Five messages of different sizes were generated, with each embedded into one audio sample. The corresponding stego-audio files were analyzed on MATLAB 2013a. Details including the audio name, size, and message size, and results of the EC, MSE, and PNSR are presented in Table 3.

Table 3. Performance evaluation of proposed EMR data hiding technique

Audio Name	Audio Size (KB)	Message Size (KB)	Embedding Capacity (%)	MSE (decibel)	PNSR (decibel)
hotstuff	48	3.1	6.5	5.3347e-08	120.8597
Sample Audio	16	3.7	23.1	3.1388e-09	133.1631
mail	20	2.3	11.5	3.6123e-07	112.5530
cello	648	3.9	0.60	2.7825e-09	133.6864
skippy	24	2.8	11.6	5.3586e-08	120.8403

Results showed that the different messages were successfully embedded in the .wav audio covers. The low MSE in each case shows that little alteration was added. Equally, the high PNSR values, which are consequences of the low MSE, confirm high resemblance between the original audio samples and the corresponding stego-audio files. This implies that our proposed system is good, causing no significant distortions in the audios. To further explore the effect of embedding messages in the selected audio files. Figures 3 to 7 present the histogram plots. The results show little or no differences in the audios after steganography.

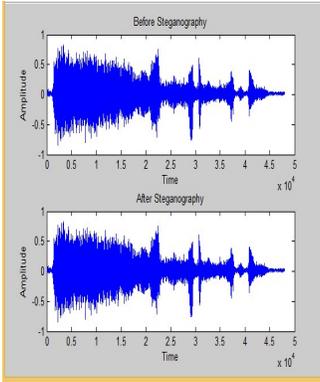


Figure 3. Audio plot (histogram) of *hotstuff.wav* before and after steganography

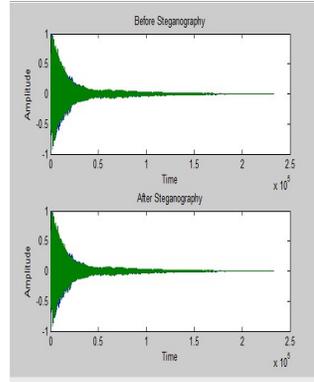


Figure 4. Audio plot (histogram) of *Sample audio.wav* before and after steganography

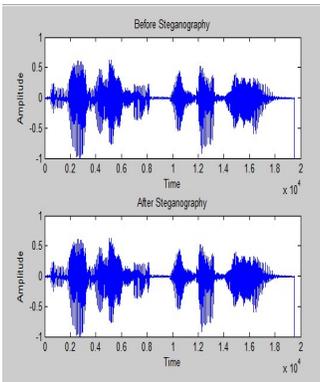


Figure 5. Audio plot (histogram) of *mail.wav* before and after steganography

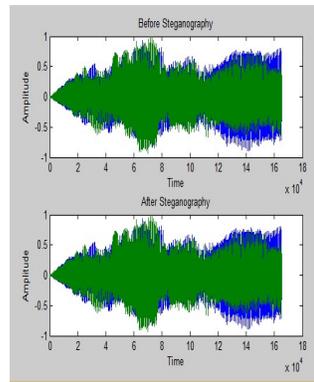


Figure 6. Audio plot (histogram) of *cello.wav* before and after steganography

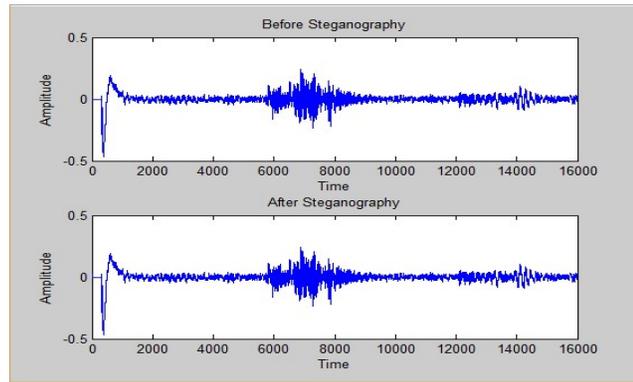


Figure 7. Audio plot (histogram) of *skippy.wav* before and after steganography

V. CONCLUSION

In this study, we proposed and implemented an enhanced EMR information hiding system to protect medical records from unauthorized access in healthcare environment. The system combined NTRU cryptographic algorithm and LSB audio steganography to offer a more robust method for hiding the EMR secret data from unauthorized access. The performance of the crypto-steganographic system was evaluated using Matlab environment. Evaluation of the

performance showed little or no distortion to the sample audios after message embedment. Our system could promote secure communication in healthcare systems, ensuring confidentiality, integrity, and availability.

Our major contribution lies in proposing a secure crypto-steganographic technique. Future studies could consider other quantum-safe cryptographic schemes. This includes techniques based on lattice theory, coding theory, and multivariate quadratic polynomials. Our proposed system implemented the commonly used spatial domain technique, LSB. Frequency domain techniques have been reported to be stronger than those in the spatial domain [34]. Combining highly secure cryptographic and steganographic techniques would no doubt increase the level of security an EMR information hiding system can provide.

VI. REFERENCES

- [1] Abdullah, A.M. & Aziz, R.H.H., 2016. New Approaches to Encrypt and Decrypt Data in Image using Cryptography and Steganography Algorithm. *International Journal of Computer Applications*, 143(4), pp.11–17.
- [2] Abikoye, O.C., Adewole, K.S. & Oladipupo, A.J., 2012. Efficient Data Hiding System using Cryptography and Steganography. *International Journal of Applied Information Systems*, 4(11), pp.6–11.
- [3] Campagna, M. et al., 2015. *Quantum Safe Cryptography and Security: An Introduction, Benefits, Enablers and Challenges*, Available at: <http://www.etsi.org/images/files/ETSIWhitePapers/QuantumSafeWhitepaper.pdf>.
- [4] Das, J., 2014. A Study on Modern Cryptography and their Security Issues. *International Journal of Emerging Technology and Advanced Engineering*, 4(10), pp.320–324.
- [5] Deshpande, N. et al., 2015. Implementation of Audio Steganography Using RSA Algorithm. *International Journal of Technology Enhancements and Emerging Engineering Research*, 3(4), pp.81–84.
- [6] Djebbar, F. et al., 2012. Comparative study of digital audio steganography techniques. *EURASIP Journal on Audio, Speech, and Music Processing*, 25, pp.1–16.
- [7] Garg, N. & Kaur, K., 2016. Hybrid information security model for cloud storage systems using hybrid data security scheme. *International Research Journal of Engineering and Technology*, 3(4), pp.2194–2196.
- [8] Gupta, S., Ankur, G. & Bhushan, B., 2012. Information Hiding Using Least Significant Bit Steganography and Cryptography. *International Journal of Modern Education and Computer Science*, 6, pp.27–34.
- [9] Hayfaa, A., Ahmad, R. & Noor, N.M., 2014. Combining Cryptography and Steganography for Data Hiding in Images. *conference of Applied Computer and Applied Computational Science (ACACOS)*, pp.128–134.
- [10] Latiff, M.S.A., Chiroma, H., Osho, O., Abdul-Salaam, G., Abubakar, A.I. and Herawan, T., 2017. A review on mobile SMS spam filtering techniques. *IEEE Access*, 5, pp.15650–15666.
- [11] Hussain, M. et al., 2015. Pixel value differencing steganography techniques: Analysis and open challenge. In *2015 IEEE International Conference on Consumer Electronics - Taiwan, ICCE-TW 2015*. pp. 21–22.
- [12] Jefferson, D. et al., 2004. *A security analysis of the Secure Electronic Registration and Voting Experiment (SERVE)*, Available at: <http://requiem.googlecode.com/files/paper.pdf>.
- [13] Jefferson, D., Rubin, A. & Simons, B., *A comment on the May 2007 DoD report on Voting Technologies for UOCAVA Citizens*, Available at:

- https://www.verifiedvoting.org/wpcontent/uploads/2014/10/serve_dod_comment_2007.pdf.
- [14] Kaur, A. & Singh, N., 2015. SMS Encryption using NTRU Algorithm. *International Journal of Advanced Research in Computer Science & Technology*, 3(2), pp.96–100.
- [15] Kaur, M. & Kaur, G., 2014. Review of Various Steganalysis Techniques. *International Journal of Computer Science and Information Technologies*, 5(2), pp.1744–1747.
- [16] Phad, V.S., Bhosale, R.S. & Panhalkar, A.R., 2012. A Novel Security Scheme for Secret Data using Cryptography and Steganography. *International Journal of Computer Network and Information Security*, 2, pp.36–42.
- [17] Qian, T. & Manoharan, S., 2016. A comparative review of steganalysis techniques. In *2015 IEEE 2nd International Conference on Information Science and Security, ICISS 2015*. pp. 1–4.
- [18] Rasmi, A. & Mohanapriya, M., 2016. An Extensive Survey of Data Hiding Techniques. *Indian Journal of Science and Technology*, 9(28), pp.1–7. Available at: <http://www.indjst.org/index.php/indjst/article/view/90457>.
- [19] Reddy, M.I.S. & Kumar, A.P.S., 2016. Secured Data Transmission Using Wavelet Based Steganography and Cryptography by Using AES Algorithm. *Procedia Computer Science*, 85, pp.62–69. Available at: <http://dx.doi.org/10.1016/j.procs.2016.05.177>.
- [20] Wu, H.-C. et al., 2005. Image steganographic scheme based on pixel-value differencing and LSB replacement methods. *IEE Proc.-Vis. Image Signal Processing*, 152(5), pp.611–615. Available at: <http://kar.kent.ac.uk/12311/>.
- [21] Sarmah, D.K. & Bajpai, N., 2010. Proposed System for data hiding using Cryptography and Steganography. *International Journal of Computer Applications*, 8(9), pp.1–8. Available at: <http://arxiv.org/abs/1009.2826>.
- [22] Sethi, P. & Kapoor, V., 2016. A Proposed Novel Architecture for Information Hiding in Image Steganography by Using Genetic Algorithm and Cryptography. *Procedia Computer Science*, 87, pp.61–66.
- [23] Singh, A. & Malik, S., 2013. Securing Data by Using Cryptography with Steganography. *International Journal of Advanced Research in Computer Science and Software Engineering*, 3(5), pp.404–409.
- [24] Tsai, Y.Y., Chen, J.T. & Chan, C.S., 2014. Exploring LSB substitution and pixel-value differencing for block-based adaptive data hiding. *International Journal of Network Security*, 16(5), pp.363–368.
- [25] Alhassan, J.K., Abubakar, B., Olalere, M., Abdulhamid, S.I.M. and Ahmad, S., Forensic Acquisition of Data from a Crypt 12 Encrypted Database of Whatsapp.
- [26] Saraireh, S., 2013. A Secure Data Communication System Using Cryptography and Steganography. *International Journal of Computer Networks & Communications*, 5(3), pp.125–137.
- [27] Zachariah, B., Yabuwat, P.N. & Bernard, E., 2016. Application of Steganography and Cryptography for Secured Data Communication – A Review. *International Journal of Engineering Research & Technology*, 5(4), pp.186–190.
- [28] Usha, S., Kumar, G.A.S. & Boopathybagan, K., 2011. A secure triple level encryption method using cryptography and steganography. *Proceedings of 2011 International Conference on Computer Science and Network Technology, ICCSNT 2011*, 2, pp.1017–1020.
- [29] Waziri, V.O., Isah, A., Ochoche, A. and Abulhamid, S.I.M., 2012. Steganography and its applications in information dissimulation on the web using images as securityembedding: a wavelet approach. *Int J Comput Inf Technol*, 1(02), pp.194–202.
- [30] Latiff, M.S.A., 2017. A checkpointed league championship algorithm-based cloud scheduling scheme with secure fault tolerance responsiveness. *Applied Soft Computing*, 61, pp.670–680.
- [31] Hoffstein, J., Pipher, J. & Silverman, J., 1998. NTRU: A ring-based public key cryptosystem. *Algorithmic number theory; Lecture Notes in Computer Science (LNCS)*, pp.267–288. Available at: <http://link.springer.com/chapter/10.1007/BFb0054868>.
- [32] Singla, S. and Bala, A., 2018, April. A Review: Cryptography and Steganography Algorithm for Cloud Computing. In *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)* (pp. 953–957). IEEE.
- [33] Mustafa, G., Ashraf, R., Mirza, M.A. and Jamil, A., 2018, June. A review of data security and cryptographic techniques in IoT based devices. In *Proceedings of the 2nd International Conference on Future Networks and Distributed Systems* (p. 47). ACM.
- [34] Mohsin, A.H., Zaidan, A.A., Zaidan, B.B., bin Ariffin, S.A., Albahri, O.S., Albahri, A.S., Alsalem, M.A., Mohammed, K.I. and Hashim, M., 2018. Real-Time Medical Systems Based on Human Biometric Steganography: a Systematic Review. *Journal of medical systems*, 42(12), p.245.