Secure Storage and Sharing of COVID-19 Data in Health Facilities using AES-Cryptography and Audio Steganography

Gabriel Junior Arome
Department of Cyber Security
Federal University of Technology, Akure, Nigeria

Abstract

Well over two years after the breakout and subsequent declaration of the dreaded COrona VIrus Disease 2019 (COVID-19) as a pandemic by the World Health Organisation (WHO), the nations and continents of the world are still suffering from the incredible damage the disease has done to human life and the global economy at large. As the world still grapples with all these, most of the strategies that have been proposed and adopted for eliminating or ameliorating the effect of COVID-19 involves real time data collation, storage (in hospital databases) and even sharing (among hospitals) across untrusted/unprotected open channels. A very significant concern that needs serious and urgent attention than it is currently getting is the security of the COVID-19 data while in store or while being exchanged, as well as the privacy of persons involved. Existing solutions still suffer susceptibility to fraudulent security/privacy breaches. This study thus presents an efficient/robust audio steganography system for secure storage and sharing of COVID-19 related data. In the developed system, sensitive/private COVID-19 patients' data is first encrypted using the Advanced Encryption Standard (AES) crypto-scheme, then the resultant cipher-text is compressed using the Lempel-Ziv Welch and Huffman's Compression Algorithms. Thereafter, the compressed file is embedded in a suitable cover (audio-file) to obtain an output which is not distinguishable from the original audio cover file. This output can then be safely stored or shared as necessary. The Signal-to-Ratio (SNR) obtained from our experiments reveal that, the new system has little or no distortion in its outputs.

1. Introduction

The outbreak of the COVID-19 epidemic has caused untold hardship on individuals, businesses, and even the economies of most nations. As the world grapple with this pandemic, several global initiatives and/or proposals geared towards treatment and limiting its spread have emerged. Indeed, with technological advancements, several mechanisms and applications have been developed for tracing the physical contact made by an individual with someone who has been tested COVID-19 positive. In fact, existing literature reveals that, over 32 countries have designed apps as a means to reducing or eliminating the scourge of this dreaded pandemic. However, many of these apps suffer susceptibility to security and privacy related attacks, while some of the apps were developed with little or no consideration for privacy of patients and their personal records. These security/privacy concerns have greatly slowed down the rate of adoption of these technologies by the masses. Besides, cyber criminals and other individuals with malicious intents now take COVID-19 as an opportunity to gain unauthorized access to peoples' personal data and perpetuate other forms of cybercrimes especially for monetary gains. Healthcare systems are being attacked with ransom-ware and resources such as patient's records confidentiality, and integrity are being compromised even on daily basis.

Existing literature holds it that, as individuals and stakeholders across the globe battle with questions that pertains to the spread pattern of the epidemic, how long it will take before an infected person recovers or die, when the epidemic will subside or disappear completely, or even, places of highest concentration of infected persons, other strategies like, real time data collation, modelling and predictions became very relevant. One very important concern that needs serious and urgent attention is the cyber security implications of COVID-19. One issue is however common, and requires much attention than it is currently getting. That is the cyber security and privacy issues that are creeping up [1].

Cybercrime is one of the prime or leading threat to all Internet-based transactions on the planet, and has now become the most serious issue with humanity. In today's reality, cybercrime seem to be more lucrative than even the trade of illegal drugs. Cybercrime costs consist of so much including, financial frauds, theft/loss of data, Intellectual property theft, reduced productivity, disruption of business, loss of time, loss of money, loss of convenience, reduced availability of systems, and even misappropriation of funds [1-2].

The collection of patients' information, its storage and sharing or exchange across the open enterprise/public network will greatly expose some of these patients or users to the dangerous activities

 | DOI: 10.20533/jitst.2046.3723.2021.0090

of cybercriminals. One potent strategy to mitigating these challenges is the use of cryptography or even steganography.

The security (privacy, confidentiality and integrity) of information while in storage and even in transit during their respective e-service implementation is essential to the success rate of all these application areas [3-4]. The information overload that characterizes today's e-society (Internet) offers several benefits. Yet, it can make life miserable and pretty difficult for stakeholders.

Enormous amount of personal information is involved especially when we talk about contact tracing. This has heightened the concerns and other e-commerce stakeholders, especially as it concerns the protection of personal data. People give away a lot of personal information that includes but are not limited to their names, addresses, credit card numbers, products purchased, location and time of transactions, and even customers' behavioral patterns. This information can be harnessed for further analysis by unauthorized third parties (attackers or adversaries) with malicious intentions that could ultimately result in various forms of cybercrimes like blackmail, financial fraud, theft, and even business secret leakage to competitors. Indeed, in some regards, it is compulsory for businesses that process customers' personal data, to state and uphold strict compliance with data protection regulations. Consequently, countering these security and privacy concerns has therefore become of serious interest to both customers and merchants [1, 5].

Most of the existing solutions are still susceptible to common classical and quantum attacks. Indeed, most of the existing systems are either based on cryptography or on steganography. While cryptography has the advantage of completely encoding information before it is transmitted and only the intended recipient can access the encoded/enciphered message, its major drawback lies in the fact that cryptography arouses suspicion. Steganography on the other hand seek to hide the mere existence of the secret communication in the first place. Its limitation however lies in the fact that a curios and intelligent criminal/fraudster could easily carryout steganalysis and recover the hidden message. There is therefore a major motivation for studies into ways of developing more efficient and robust solutions. A combination of both steganography and cryptography could be a more robust and efficient solution to the security concerns of this age. This paper therefore proposes a combination of AES, DCT, LZW as well as the Huffman's algorithm to develop an audio steganography system for securing COVID-19 related data storage and transmission.

2. Related works

The literature reviews show that their numbers of related research work that has been done and most of them exhibits limitations that serve as vulnerabilities or drawbacks. Akinyede et al., in [6] presented an electronic payment system whose security is based on the AES encryption scheme. Their system was however not evaluated. Besides, the use of only AES, arouses suspicion from malicious persons. There is need for equipping such system with ability to hide the existence of secret communication. The authors of [7] and [8] contributed their quota in the quest to finding solution to the security concerns. Their steganography system however are not so efficient, as some of their outputs are distorted and could lead to arousing suspicion. Besides, the work in [8] supports the use of only one audio file format (.wav) that is a limitation on its own. Although the authors in [9] presented a research work geared towards increasing the capacity of low bit encoding audio steganography, their system incurs high computational overhead especially with increasing size of input secret message.

The authors of the work reported in [10] developed a system for encrypting text files and hiding the resultant cipher-text in a digital object. Their goal was to harness the strengths of both cryptography and steganography for a robust security system. The mp3 audio file format was used as cover file. As a drawback, their system yields poor quality stego-files as outputs. Furthermore, they could have also tried other file formats apart from the MP3 format.

The authors of [11] carried out an analysis and design of three LSB techniques. The major goal of their work was to devise a way of embedding image files in audio cover files using 3rd LSB technique. There was, however, a huge variance in the quality of the cover audio file before and after the embedding procedure. This implies their system has low robustness and maybe susceptible to statistical analysis. Besides, only the .WAV file format was used.

A survey of audio steganography methods was done by the authors of [12]. They reached a conclusion that in terms of simplicity, LSB method is the best. However, a better method for increased robustness of steganography systems is the spread spectrum technique.

Adeboje et al. in [13], developed an audio steganography system that supports two (2) file formats being used as cover files. The authors used the MP3 and the MP4 file formats for evaluating their system. There is need to explore other audio file formats to determine their suitability.

In order to cater for the protection of information against common classical and even quantum attacks on the cloud [4], and even electronic voting systems

 | DOI: 10.20533/jitst.2046.3723.2021.0090

[5], Gabriel et al., proposed more attack-resilient information systems that combine post quantum cryptography with steganography.

Olomo et al. 2020 in [14], carried out a study on image steganography and steganalysis. They reported impressive results. Even Gabriel et al. in [15], carried out a study that developed a two-layer steganography system for covet communication. However only images were considered as cover files in these two independent studies. Further studies are required to achieve a similar system for audio or even video files.

3. The proposed system

The proposed system design is discussed under four sub-titles; the Data Sharing Network Model, the threat model, the specific objectives, and the proposed system architecture/Design

3.1. The Data Sharing Network Model

The proposed system is made up of four (4) major entities or actors. Exchange of information is between these entities. These entities are; the health service provider, facility or system *S*, who is usually the expert or health authority, the Customer or patient *D*, whose COVID-19 status is the subject of focus in the entire system, the government representative or ministry of health *B*, which handles control aspect of the system, as well as the policy enforcement agencies *DA*, which handles policy enforcement aspect of the setup. The communication between these entities is usually done across the open Internet.

3.2. The Threat Model

In this paper, we consider the major threat to be an adversary who is a malicious entity/individual with the ability to; 1) compromise any of the entities involved in this e-health data transmission scenario and/or 2) eavesdrop messages exchanged between the entities involved in the communication, with the aim of deducing and retrieving information on users' or patients' private life, towards blasphemy or financial gains. There is therefore need to detect and block the threats posed by the adversary. More specifically; users' privacy preservation must be ensured. That implies that, it should be impossible for the adversary to gain any knowledge of the COVID-19 or even general health status of individual users or patients. None of the other entities should have a fine-grained knowledge of customers personal information. Also, Integrity of messages exchanged must be completely ensured. That implies that, every receiver-entity must verify the authenticity of the sender-entity. Attempts at modification, FDI, MMA, replay and other common cyber-attacks must be detected and/or rebuffed.

3.3. The Proposed System Design

In this paper, the problem of secure COVID-19 Data storage and sharing is illustrated in terms of the famous *prisoners' problem*, where Alice (the sender) and Bob (the receiver) are two inmates that have a goal of discussing to hatch an escape strategy. Both Alice and Bob must do everything possible not to arouse the suspicion of Wendy the prison warden (while discussing/communicating) who, will place them on solitary confinement.

In our proposed secure e-health data sharing framework, we have Bob (the *sender*) representing client systems used by both the customers and other stakeholders. A client system wishes to securely send a secret message *M* (that is, patients' personal information such as their COVID-19 status, their location/address, phone numbers etc.) to Alice (the *receiver*), representing the server system belonging to the government-designated Trusted Agent (TA) or Trusted Third Party (TTP). In order to allow for covert exchange of information between these entities or stake holders (health facilities, patients, and the TA), Bob first encrypts the personal or private information using the Advanced Encryption Standard (AES) algorithm to produce a cipher- text, *Y*.

Then Bob passes the resultant cipher-text through the Huffman's compression algorithm. That output is further passed through the Lempel Ziv Welch compression algorithm. These help in reducing the size of the cipher-text considerably. Ultimately, the robustness of the steganography system to statistical analysis is enhanced, and the perceptibility of distortion in its final output (stego-file) by the human auditory system is greatly reduced or eliminated.

Bob the sender (merchant/customer) then picks a suitable audio file (in this case M4A audio file format) as a cover file *C*. In this research work, the Spread Spectrum technique, (specifically, the frequency-hopping spread spectrum technique) was used to embed the encrypted and compressed secret message (Text file) into the digital .M4A audio signal. The frequency-hopping spread spectrum technique used in this work, alters the audio cover file's frequency spectrum so that it hops rapidly between frequencies. This way, it identifies the low frequencies of the audio signal frames and creates a final output stego-file *S*, by hiding/embedding bits of the cipher-text in them without causing audible distortions (that is, without arousing Wendy's suspicion).

The resultant stego-image, *S* is stored in the local hospital's database or transmitted over a public channel (monitored by Wendy/cyber-criminals) and is received by Alice only if Wendy has no suspicion on it. Once Alice receives *S*, she can retrieve the original plain-text message via the extraction process. as shown at the receivers' segment of the architecture of the system the right-hand side of the diagram in Figure 1.

 | DOI: 10.20533/jitst.2046.3723.2021.0090

The embedding process denotes the serious job for a steganography system. This is because, the final stego-file output must be as similar as possible to the original selected audio cover file. No distortion must be observed by Wendy the unauthorized eavesdropper.

As shown in Figure 1, at both the patients' as well as the Trusted Agents' sides of the secure e-health data storage/sharing system, two major processes are carried out; the Cryptography procedure as well as the Steganography procedure.

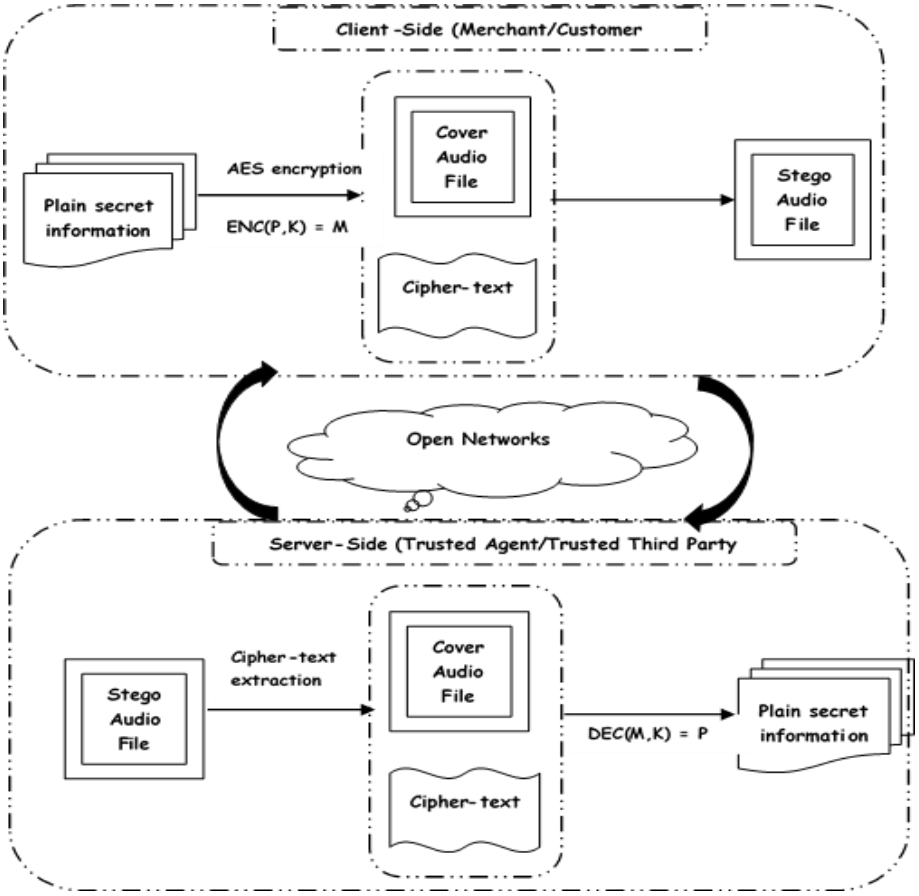

Figure 1. Architecture of the proposed system

4. Results and Discussions

In order to cater for the implementation and subsequent evaluation of the proposed system, MATLAB programming language was used on Windows 10 Operating System platform with hardware configuration of 4 Gigabytes RAM, 2.16 Gigahertz Intel processor speed and 1 Terabyte of hard disk capacity.

The M4A audio file format was chosen as the cover media in this experiment. Secret text files of diverse characteristics were embedded in the chosen audio cover files to allow for the evaluation of our system. The resultant stego files obtained as outputs of the proposed system were examined. It was observed that, the stego files (audio file) retained its initial size, and the amount of information that the developed system can hide is very high (500KB).

In order to cater for the evaluation of the research work, standard performance metrics as highlighted in section 4.1 were used.

4.1. System Evaluation Metrics

Standard metrics such as, computational time, bit per character, compression ratio and signal to noise ratio was used in the evaluation of our secure e-commerce transactions system. Compression ratio refers to the ratio of the size of the cover medium before and after the secret financial information is implanted into it. This evaluation metric is as captured in equation 1.

$$Compression\ Ratio = \frac{output\ file\ size}{input\ file\ size} \quad (1)$$

The compression ratio achieved from the result is 1. This implies that there is no difference in the size of the audio cover file before and after the steganography experiment.

 | DOI: 10.20533/jitst.2046.3723.2021.0090

The second metric used for evaluating this e-commerce transaction system is the Signal to Noise Ratio (SNR), which is measured in decibel. This metric given by the equation in 2, reveals the level of corruption introduced by the noise. It is defined as the ratio of the signal power to the noise power.

$$\text{SNR (db)} = 10\log \frac{\sum_n I_n}{\sum n(E_n - I_n)^2} \quad (2)$$

where E_n= Stego file and I_n= Original Audio Signal

The third metric of evaluation is the *computation time*. This refers to the time taken for the system to execute its function.
These times were recorded for the two (2) audio files used as cover-media in this research work.

4.2. Proposed System Evaluation Results

We present the results of the performance of the new system on given inputs (plain texts, audio cover files). M4A Audio file was used as cover file, in which COVID-19 patient records of varying sizes were embedded.

Table 1. Results of experiment when .M4A audio file was used as cover

Plain text size (kb)	Compressed cipher text size (kb)	Compression Ratio (%)	Proposed System Operation Time (seconds)	SNR (db)	Output or Stego Audio File Size	Extraction Time
50	43	14.2	9	59.7	4.81	7
100	89	11.1	8	55.9	4.81	6
250	132	12.2	8	51.6	4.81	6
200	163	18.3	8	49.8	4.81	6
250	225	10.1	8	42.6	4.81	6

It was observed from results that, the SNR values exceeds 50db whenever the size of the text file to be embedded lies between 50kb to 150kb, but the value of SNR for 200kb text file is 49.8 which is approximately 50db can also be accepted as good SNR value. This indicates that there will be no distortion in the audio cover file for sizes of the text file embedded up to about 200kb. Anything beyond 200kb, the SNR value decreases below 50db, which is too much, and will yield observable distortions in the final output. A graph of the SNR values against the respective text file sizes is presented in Figure 1.

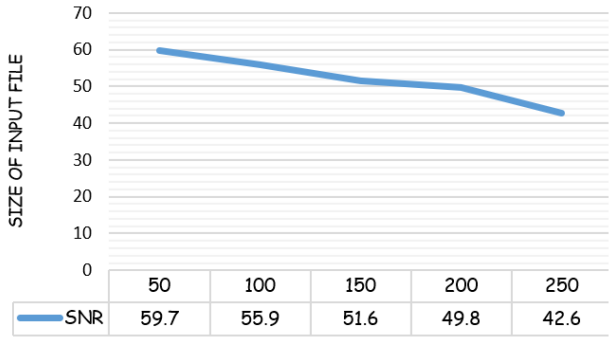

Figure 2. Graph of Signal-to-Noise Ratio

5. Conclusion and Future Works

The frequent and frightening rate of occurrence of financial crime and fraud in today's electronic society is worrisome. Information leakage, loss of privacy and monumental loss of money and other valuables have been reported almost daily. Existing methods suffer several limitations and are grossly weak. This paper has presented the development of an audio steganography security framework for ensuring secure and privacy preserving electronic commerce transactions over open enterprise networks. The M4A audio file format was used as cover medium/file in this work.

The .M4A audio file have the SNR ratio that's greater than or approximately equal to 50db at 200kb. This implies that large file of size up to about 200kb can be hidden without causing audible distortion in the cover medium. Also, for real-time applications that

 | DOI: 10.20533/jitst.2046.3723.2021.0090

requires speed of execution to be very fast, the new system performed well on both audio file formats.

The developed system would be very useful in securing and sharing large amount of sensitive e-health data between health business transacting entities without arousing suspicion of fraudsters and other unauthorized individuals who have malicious intentions. Future work will consider the combination of Post-Quantum Cryptography and DCT Steganography for images, audio and video file format covers.

6. References

[1] Gabriel A.J., Darwsih A., Hassanien A.E. (2021). Cyber Security in the Age of COVID-19. In: Hassanien A.E., Darwish A. (eds) Digital Transformation and Emerging Technologies for Fighting COVID-19 Pandemic: Innovative Approaches. Studies in Systems, Decision and Control, vol 322. Springer, Cham. DOI: 10.1007/978-3-030-63307-3_18.

[2] Gabriel, A. J., Alese, B. K., Adetunmbi, A. O. and Adewale. O. S. (2013). Post-Quantum Crystography; A combination of post-quantum cryptography and steganography. The 8th International Conference for Internet Technology and Secured Transactions (ICITST-2013), Technically Co-Sponsored by IEEE UK/RI Computer Chapter. 9th-12th December. London, UK, 454-457.

[3] S. Gorbunov, V. Vaikuntanathan, D. Wichs, (2015). Leveled Fully Homomorphic Signatures from Standard Lattices. STOC'15, June 14-17. ACM 978-1-4503-3536-2/15/06. DOI: 10.1145/2746539.2746576.

[4] Gabriel, A. J., Alese, B. K., Adetunmbi, A. O. and Adewale, O. S. (2015). Post-Quantum Crystography based Security Framework for Cloud Computing,Journal of Internet Technology and Secured Transactions (JITST). 4(1), 351-357.

[5] Gabriel, A. J., Alese, B. K., Adetunmbi, A. O., Adewale, O. S. and Sarumi O. A. (2019). Post-Quantum Crystography System for Secure Electronic Voting, Open Computer Science. DeGruyter 9:292-298.

[6] Akinyede, R. O., Adewale, O. S. and Alese, B. K. (2014). Building a Secure Environment for Client-Side E-Commerce Payment System using Encryption System. Proceedings of the World Congress on Engineering. Vol I, WCE 2014, July 2 - 4. London, U.K.

[7] Wheeler, D., Johnson, D., Yuan, B. and Lutz, P. (2012). Audio Steganography Using High Frequency Noise Introduction. Thomas Golisano College of Computing & Information Sciences Rochester Institute of Technology, Rochester NY, RIT Scholar Works, http://scholarworks.rit.edu/other/302. (Access Date: 11 December 2021).

[8] Ghanwat, D. and Rajan, R. S. (2013). Spread Spectrum-based Audio Steganography in the Transformation Domain. Global Journal of Advanced Engineering Technologies, 2(4):66-77. 2013.

[9] Olanrewaju, R. F., Othman, H. A. and Suliman, K. R. (2013). Increasing the Hiding Capacity of Low-bit encoding Audio Steganography using a Novel Embedding Technique, World Applied Sciences Journal, 2013, 21(26): 79-83.

[10] Kresnha, P. E. and Mukaromah, A. (2014). A Robust Method of Encryption and Steganography using ElGamal and Spread Spectrum Technique on MP3 Audio File. In Proceeding of Conference on Application of Electromagnetic Technology, 2014, 3(9):11-15.

[11] Kumari, L., Goyal, D. and Gyan, S. (2013). Analysis and Design of Three LSB Techniques for Secure Audio Steganography. International Journal of Advanced Research in Computer Engineering and Technology (IJARCET), 2013, 2(2): 44-55.

[12] Balgurgi, P. P. and Jagtap, S. K. Audio Steganography Used for Secure Data Transmission. In Kumar M. et al., (eds) Proceedings of International Conference on Advances in Computing. Advances in Intelligent Systems and Computing, vol 174. Springer, New Delhi. DOI:10.1007/978-81-322-0740-5_83.

[13] Adeboje, O. T., Adetunmbi, A. O. and Gabriel, A. J. (2020). Embedding Text in Audio Steganography System using Advanced Encryption Standard and Spread Spectrum. International Journal of Computer Applications (0975-8887). DOI:10.5120/ijca2020919. Volume 177, Number 41. Pp 46-51.

[14] Olomo, R., Misra, S., Ahuja, R., Adewumi, A., Ayeni, F. and Mmaskeliunas, R. (2020). Image Steganography and Steganalysis Based on Least Significant Bit (LSB). In: Singh P., Panigrahi B., Suryadevara N., Sharma S., Singh A. (eds) Proceedings of ICETIT 2019. Lecture Notes in Electrical Engineering, vol 605. Springer, Cham.

[15] Gabriel, A. J., Adetunmbi, A. O. and Obaila, P. (2020). A Two-Layer Image-Steganography System for Covert Communication Over Enterprise Network. In: Gervasi O. et al. (eds) Computational Science and Its Applications – ICCSA 2020. Lecture Notes in Computer Science, vol 12254. Springer, Cham. DOI: 10.1007/978-3-030-58817-5_34.

7. Acknowledgements

Federal University of Technology, Akure, Nigeria, for providing the environment for the study.