Enhancing Data Security using Video Steganography, RSA and Huffman Code Algorithms with LSB Insertion

Richard Apau
Department Of Computer Science
Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

J. B. Hayfron-Acquah
Department Of Computer Science
Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

Frimpong Twum
(Corresponding Author)
Department Of Computer Science
Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

ABSTRACT
Security (i.e. Confidentiality, Integrity, Authentication, Non-Repudiation, and Availability) in the field of data communication have remained a subject matter of discussion over the years. The internet as well as computer technology have made significant strive in data communication existence. Transferring data securely and safely amidst vulnerabilities of computer networks remain a source of worry to many in the field of data communication. Without security there is no need for data communication. The main objective of this study was to ensure security of data transmitted over the internet. The study proposed a novel approach of data security using video steganography, Huffman Code compression and asymmetric cryptography. In the proposed system, messages are encrypted with RSA and encrypted messages are compressed using Huffman code algorithm. The compressed encrypted messages are hidden using Least Significant Bit (LSB) algorithms. This research brings to light the concept of effectively combining steganography, compression and asymmetric cryptographic algorithm. The preference of RSA over any other cryptographic algorithm is due to its ability to provide better security for large file size thereby reducing computational complexity. The use LSB for video embedding is also good for larger file sizes due to its low computational complexity. Huffman code compression is a lossless compression algorithm which allows reduction in size of data without loss of data. From the results obtained in this research, it was observed that when video steganography is combined with Huffman code compression and asymmetric cryptography, a higher level of security, robustness and capacity are achieved. The distortion experienced in this study is negligible; therefore the study achieved increased security by the high PSNR values and low MSE and BER values.

General Terms
Cryptography, Steganography, Steganalysis, Data Compression, Least Significant Bits, RSA

Keywords
Video Steganography, Asymmetric Cryptography, Huffman Code, LSB, RSA, PSNR, MSE, BER.

1. INTRODUCTION
The high increased in internet penetration has led to many computer related crimes. Privacy and secrecy of data still remain a major challenge in today's technological world. Though stringent measures have been put in place to ensure the security and privacy of data transmitted over the computer networks, hackers and eavesdroppers also continue to device a more sophisticated and complex way of accessing such information. The internet as well as computer technology have made significant strive in data communication existence [1]. According to Wajgade and Kumar [2], the effect of combining steganography and cryptography is more beneficial in terms of obtaining the security and privacy of data. Various approaches and techniques of data security and information have been implemented by researchers to achieve secret communication. In today's digital world, steganography is one of the most safest forms of data communication [3]**.** Due to the techniques of steganalysis tools, that have the capacity of detecting hidden messages in covert media [4], cryptography has been employed as another means of securing data transmission. Although a number of studies have been done on video steganography, many focused on the use of symmetric cryptographic algorithm. Few others also used asymmetric cryptographic algorithm, those done in the area of asymmetric cryptographic algorithm places little or no emphasis on the use of RSA encryption and compression algorithms and LSB insersiton. This study is therefore designed to enhance video steganography using RSA encryption algorithm and Huffman code with LSB insertion algorithm.

2. REVIEW OF LITERATURE
2.1 Steganography and Cryptography
Many until modern times believed that steganography is an alternative to cryptography. According to Ramalingam [1], steganography is rather the dark cousin of cryptography and not an alternate. Whereas steganography provides secrecy cryptography is intended to provide privacy. The very essence of steganography is to hide or mask the existence of the message while cryptography concerns itself with the masking of the content of a message. The seemly exploitation of steganalysis which detect the presence of hidden message in covert media [5], made the combined effect of steganography and cryptography crucial. Whereas steganography can use cryptography, cryptography cannot use steganography [6]. The word "Steganography" is derived from the Greek words "Stegano" or "Stegos" meaning covered or hidden and "Graphia" or "Graptos" meaning writing [7]. For steganography to provide the needed security and concealment from eavesdroppers and attackers, it must be based on appropriate techniques [8]. Cryptography can simply be defined as the process of data storing and transmitting in a particular way such that, those whom the message is intended for can read and process it [9]. Cryptographic technique plays an essential role in protecting data communication, and also ensures that only intended recipient receives the message. According to Kundalakesi et al.[9], cryptography is very important in data communication especially when the data is being transferred over an untrusted medium particularly the internet.

2.2 Video Steganography Basics

The main objective of securely hiding data in video is to achieve better confidentiality and data recovery. Video files are composition of images and sounds or series of frames. Using video stream to hide data should remain undetected by the human eye. If a steganography algorithm based on video is detected then it is invalid. By far, video steganography hiding technique is the best since it overcame the capacity problem of image steganography and alteration problem of text steganography. Using video as a cover object didn't overcome the capacity problem only, but it also enhanced the security of the embedded data [10].

2.3 Asymmetric Cryptography

The asymmetric cryptography has the comparative advantage of increased security and convenience. Jain [11] proposed public key steganography using LSB with Deffie-Hellman key exchange protocol. A shared stego-key was found between two communicating people through the application of Deffie-Hellman algorithm. Kour and Kaur [12] proposed a data hiding system that is applicable in images and videos. DSA was combined with multiple LSB for the steganography. The carrier file file being video or image is selected, and then the raw data to be hidden is also selected with the implementation of the Digital Signature Algorithm(DSA) to generate the key. Vegh and Miclea [13] proposed a system for ensuring security in physical cyber system using LSB for the steganography and ElGamal for cryptography in addition to digital signature. Mohanta [14] proposed a secure video steganography using Elliptic Curve Cryptography (ECC) and LSB insertion algorithm. The study first encrypted a secret message with the ECC algorithm. The encrypted file is then inserted into an image using LSB. The image is subsequently embedded in the cover video using the same LSB technique. Kaur and Singh [6] however proposed an improved version of the system proposed by Mohanta [14]. In this approach, an image to be transmitted was first encrypted with ECC encryption algorithm and inserted into the cover video using LSB insertion. Huffman compression was applied on the hidden image for bandwidth optimization. Shukla and Singh [15] proposed a secured data communication application based on RSA encryption with LSB insertion. A raw data was embedded in an AVI video file after performing RSA encryption. The proposed system combines RSA and Huffman Code Algorithm with LSB to achieve high security, high robustness and high embedding capacity.

2.4 Least Significant Bit (LSB)

For the purpose of concealing message in video two video steganographic algorithms; Least Significant Bit (LSB) and Discrete Cosine Transform (DCT) have been proposed [16]. In LSB algorithm, the least significant bits of the carrier file's individual pixels are modified; hence the hidden data is encoded. Ahmed [17] proposed a system of comparing data hiding techniques in video file. Experimental results from the system showed that 75% of the hidden message using LSB showed higher PSNR values than the one using DCT.

2.5 Huffman Code Compression Algorithm

In order that, the application is able to embed a larger amount of data, a compression technique was used. There are many compression techniques available for use. The technique used in this study was the Huffman code compression. The Huffman algorithm is easy to implement and produces lossless compression. When this compression is applied on text, it increases the volume of text to be hidden indirectly.

Huffman compression belongs into a family of algorithms with a variable codeword length. The basic idea behind the algorithm is to build the tree bottom-up whose leaves are labelled with the weights.

For example. Let us consider this ASCII fixed length code.
A 000 B 001 C 010 D 011 E 100 F 101 G 110 H 111 .With this code, the message BACADAEAFABBAAAGAH is encoded as the string of 54 bits as

001000010000011000100000101000001001000000000011000 0111.

Now let us consider the variable length code below:

A 0 B 100 C 1010 D 1011 E 1100 F 1101 G 1110 H 1111, with this code, the same message as above is encoded as the string of 42 bits.

100010100101101100011010100100000111001111. In comparison, 20% of space is saved.

3. METHOD

The proposed model makes use of covert media, specifically video as a carrier for the secret data. The data to be sent secretly over the communication channel is the input secret file. At the destination, the legitimate recipient has to undergo some required steps to process the message so as to reveal the data; else the very existence of the secret file is indiscernible. The proposed technique fundamentally ensures that quality information is hidden to differentiate this system from a typical data hiding application currently in existence. This is so, due to the larger payloads capacity that this system is significantly required to provide. There are three main phases that composed of the proposed model. The first is the encryption phase that converts the message to be sent into binary data to get the secret file. Compression is applied as a second phase to reduce the size of the encrypted file in order to accommodate larger payloads. The third phase is the embedding technique which deals with the process of hiding the encrypted file into the cover video.

3.1 Proposed Model

The entire process of the proposed model is composed of ten main consecutive steps: The first step is the encryption phase in which a standard encryption algorithm is employed to convert the secret message into an encrypted file normally in the form of binary data for utmost security. In the second step, the encrypted file is compressed using a lossless data compression algorithm in order that a large amount of the file would be embedded and also to ensure bandwidth optimization during the transmission process. The third step is a process in which the video is converted into frames for the techniques of embedding to be made possible. The fourth step is the application of the embedding technique to hide the compressed encrypted file into the appropriate selected frames. The fifth step is the reconstruction of the converted frames to obtain the stego video. In the sixth step, the stego video is sent over the communication channel, normally untrusted medium like the internet to the intended recipient. At the destination, the intended recipient undergoes the seventh process in which the cover video or the stego video is separated again into frames in order to get the frames holding the compressed encrypted secret file. In the eighth step secret file is extracted from the holding frames. In the ninth step the secret file is uncompressed to get the encrypted file, whereas the tenth step decrypts the encrypted file using the intended receiver's private key to obtain the original input file. The proposed model however does not concern with securing the

communication channel. Figure 1 depicts the proposed model.

3.2 Parameters Used

Steganography applications are evaluated on some basic views. The criteria upon which the performance of such applications is evaluated are Robustness, Capacity, and Security. These performance evaluation criteria are independent of each other. Fig 2 illustrates the three performance evaluation criteria.

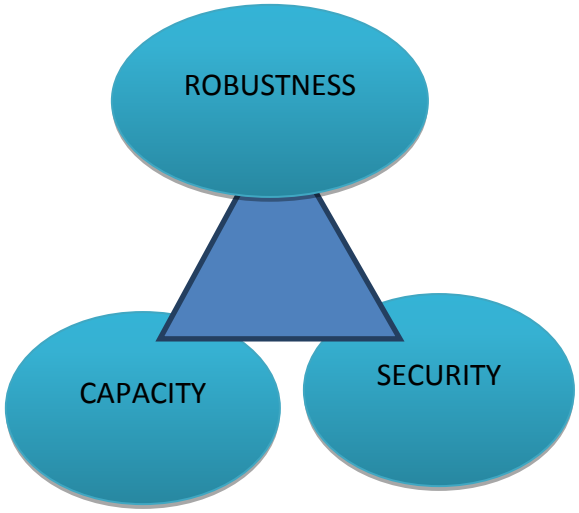

Fig 2: Performance Evaluation

Robustness of a steganographic application is the ability of the system to withstand various attacks and manipulation. It clearly demonstrates the quality of the system developed. In this regard, the quality of the original video is compared with the quality of the stego video. The Parameter used to demonstrate the robustness performance of the application is Peak-Signal- to -Noise –Ratio (PSNR). PSNR is a video quality measure by comparing the original video to the stego video. The unit of measurement of PSNR is decibels (dB). The higher the PSNR value, the quality the video is. PSNR is calculated using:

$$PSNR = 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right) = 20 \cdot \log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right) = 20 \cdot \log_{10}(MAX_I) - 10 \cdot \log_{10}(MSE)$$

MSE is the Mean Square Error which is the measure to determine the distortion between the original and the stego video. MSE is calculated using:

$$MSE = \frac{1}{m\,n}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2$$

where M and N are the Height and the Width of the video respectively

Capacity which is mostly referred as embedding capacity or payload capacity is the amount of data that can be embedded or hidden in a cover object without the quality of the video deteriorating statistically or without causing statistically significant modification. Capacity is expressed in terms of bits per pixel whereas the maximum hiding capacity is expressed in term of percentage.

$$Capacity = \frac{number\ of\ bits\ used\ to\ hide\ data}{total\ number\ of\ bits\ in\ image} \times 100\%$$

Security is the ability of an unauthorized person or a third party to discover or detect a hidden information or message in a video. This criterion is purely demonstrated by the embedding algorithm and the encryption algorithm used. In this case, the RSA and LSB algorithm should be able to make it difficult for a third party to detect the hidden information.

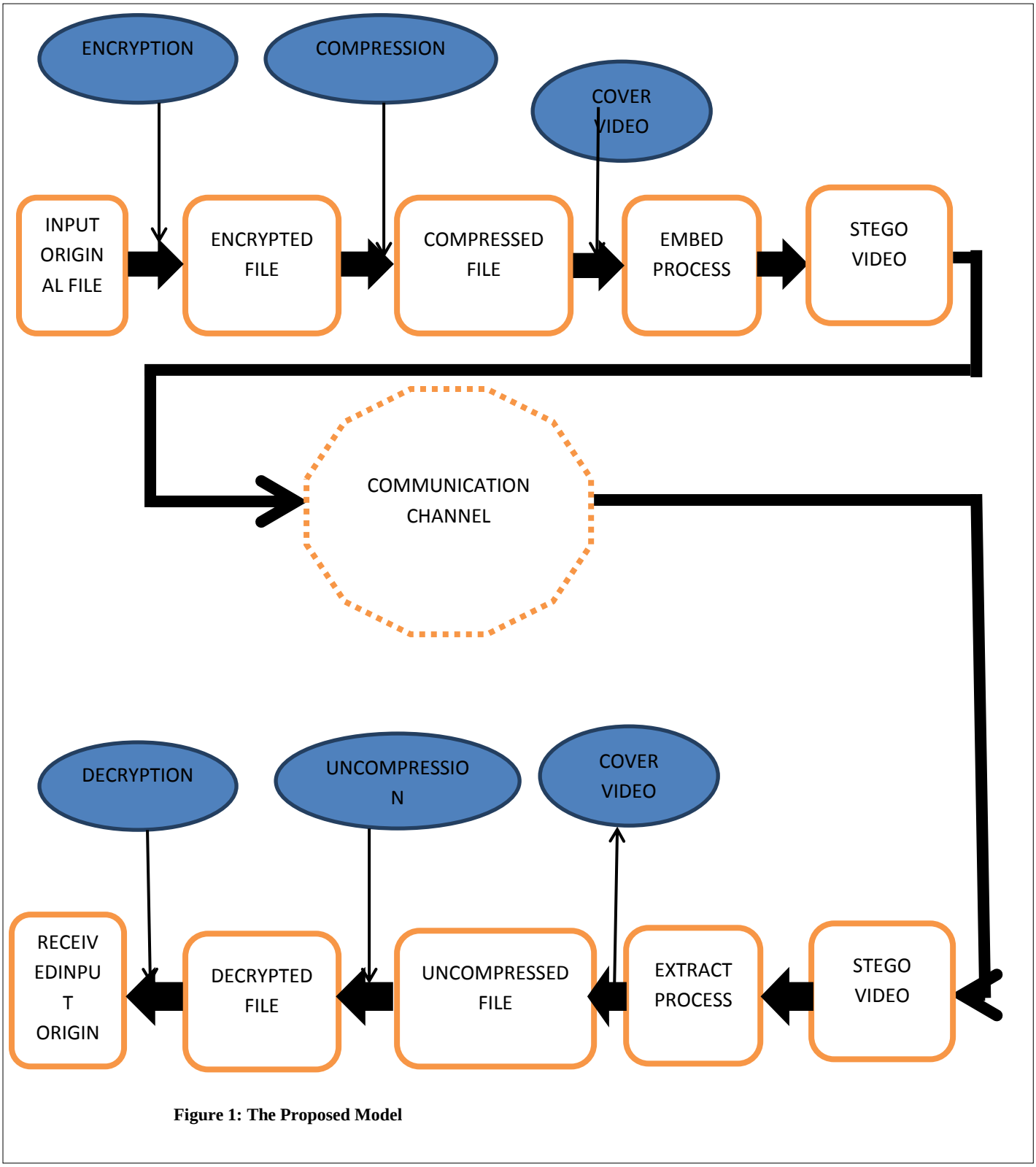

Figure 1: The Proposed Model

4. IMPLEMENTATION OF THE PROPOSED SYSTEM

The strategy which was adopted and appropriate to implement this empirical study is experimental research. Experimental research has the objective of evaluating and testing existing system or new system with the expectation of how the said system works or behaves. This is an important element of engineering and scientific process. Therefore, if the experiment is properly and well executed, the outcome would solve the identified problem. The procedure followed in this study was the development of Test Suite using JAVA SDK and NetBeans IDE in a lab. Fig. 3 shows the home screen when the application is run. The home screen shows the various processes that have to be followed to encrypt, compressed, embed and send a message over the internet to the intended recipient. It also shows the necessary steps the recipient has to follow in other to retrieve the message. This consists of security which comprises of encrypting the message with the public key and the subsequent decryption of the message with the private key at the recipient end. Stego Utility is the insertion of the secret message using the LSB and the extraction of the secret message at the recipient side. Compression deals with the Huffman code compression of the encrypted file and its resultant decompression at the recipient end. The file is then sent to the recipient through an IP address if on same network or via the email.

5. ANALYSIS, RESULTS, AND DISCUSSIONS

In this study, the stego video was passed through MATLAB software to generate its properties. The application was applied on a wide range of files, including images, database files, word files, pdf file and many more. In order to ensure that the file transferred reach its destination unaltered, a bit error rate (BER) analysis was performed using the proposed model. The results were compared to a previous work of [12], and the proposed system proved far better. The GRAPH results of figure 4 shows that the system performed better in terms of the number of bit errors recorded. The technique proposed in this study was analysed on seven different cover videos. The videos are SITI.mp4, SOC 360.mp4, SOC 366.mp4, GEOG 366.mp4, Man.avi, Saddest.avi, and Youtube.avi. These seven cover videos were analysed with a common text file. The resultant invisibility of the hidden file is shown by comparing the stego video with the cover video. There was virtually no loss in quality and also the presence of a hidden message in the video was proven to be undetected. Table 1 and Table 2 show the results obtained from the video properties. From the results of table 2, it can be observed that when RSA algorithm is combined with Huffman code using LSB insertion algorithm in video steganography by employing JAVA as the development tool, the PSNR values are higher. It also gives high embedding capacity without the video losing it quality. The high embedding capacity explains the reasons behind the low MSE and the high PSNR values. From the results, it can be realised the proposed system works better than previous works conducted as its average embedding capacity of 765979 is far higher than that of [12] which was 230400, and also that of [18] which was 497670. Also, the PSNR average value of 61 is an indication the proposed system is very good and that the security of the stego video is very high. Furthermore, the

MSE values obtained show that, the bit error rate is minimal and the difference between the pixels of the received video and the original video is insignificant.

Figure 5 demonstrates the differences in PSNR with respect to the video capacity. Different values of PSNR are given to show the video quality. Differences in the video resolutions resulted in the PSNR variations. Figure 6 shows the variation in MSE with different video resolutions. The MSE is a mean square error between the cover frame of the original video and the cover frame of the stego video.

The proposed system was analysed with different file sizes in the same cover video. This was done to test the quality and the security of the system with small and large file sizes. In table 3, text files of different sizes were embedded in the same cover video to determine the variations in PSNR and MSE. The values in table 3 show that, as the file size of the secret embedded message increases, there is variation in the PSNR values. The PSNR values decreases as the file size increases. The values in table 3 show that, as the file size of the secret embedded message increases, there is variation in the MSE values. The MSE values increase as the file size increases.

Table 1: Video Properties

Video File	Resolution	Frames	Total Bits Rate (kbps)	Length (seconds)	Size (MB)
SITI.mp4	640*360	40	954	711	81
SOC 360.mp4	1280*720	45	909	1168	126
SOC 366.mp4	720*360	44	751	1203	108
CSM 366.mp4	540*297	39	1215	990	143
MAN.avi	720*480	57	1806	532	48.2
SADDEST.avi	426*240	38	379	319	61.4
YOUTUBE.avi	426*212	32	369	822	99.1

Fig 3: Home Screen Showing the Proposed System Processes

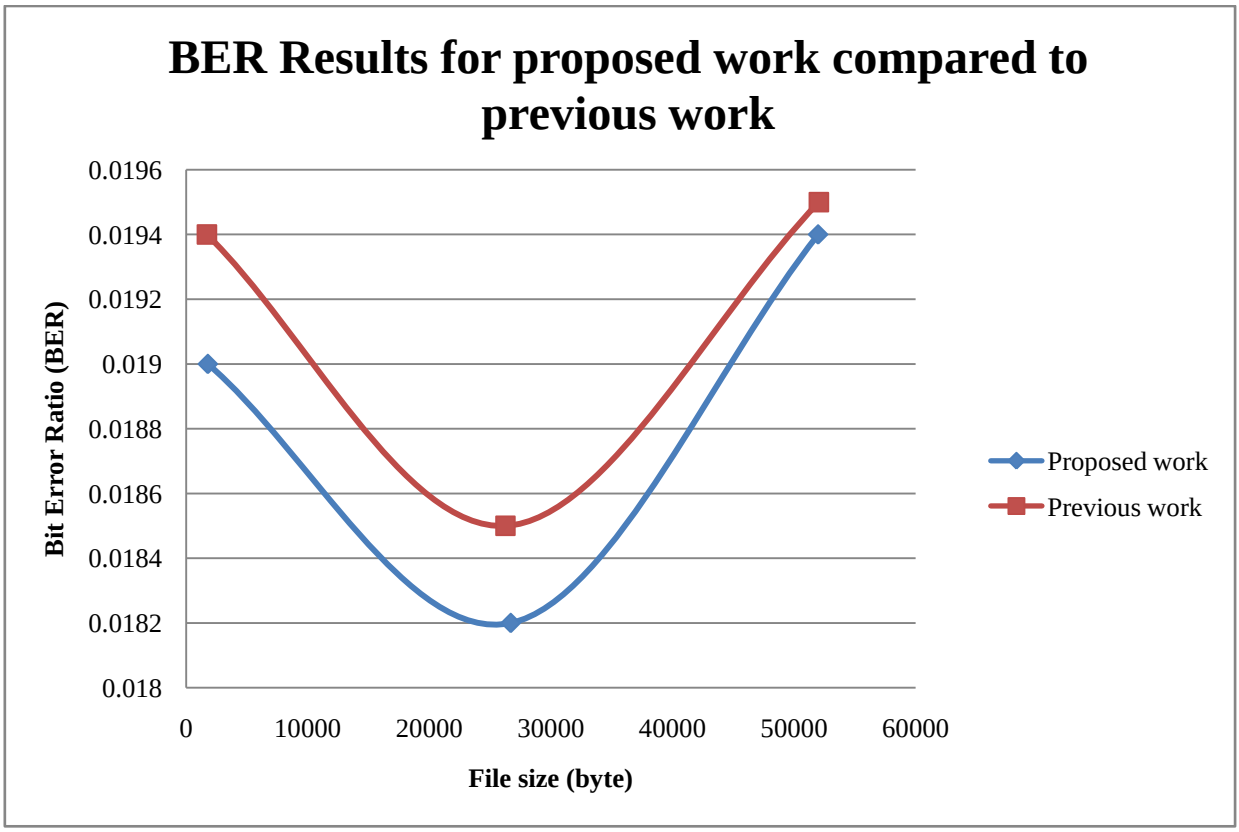

Fig 4: BER results of proposed work and previous work

Table 2: Obtained Results

Video File	Number Of Characters in Text File	Embedding capacity	PSNR	MSE
SITI.mp4	1200	856360	60.3565	0.0599
SOC 360.mp4	1200	998576	63.3886	0.0298
SOC 366.mp4	1200	902600	62.0030	0.0410
CSM 366.mp4	1200	699400	59.6923	0.0698
MAN.avi	1200	987580	66.1997	0.0156
SADDEST.avi	1200	504780	58.5932	0.0899
YOUTUBE.avi	1200	412560	58.1351	0.0999

Table 3: PSNR and MSE variations with varied file sizes

Video File	File Size	PSNR	MSE
SOC 360 .mp4 1280*720 45 frames	123 chaars (A)	68.2185	0.0098
	633 chars (B)	68.0448	0.0102
	867 chars (C)	65.1423	0.0199
	1200 chars (D)	63.3886	0.0298

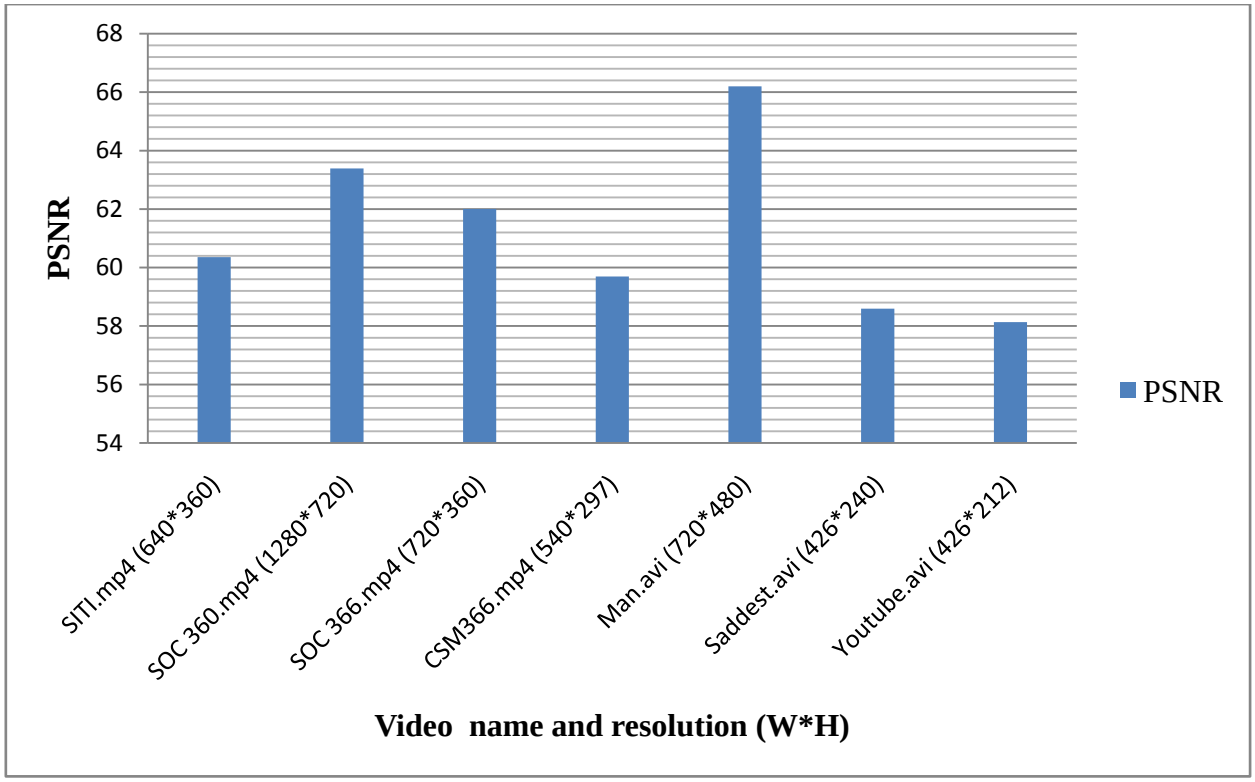

Figure 5: PSNR vs. Video Resolution

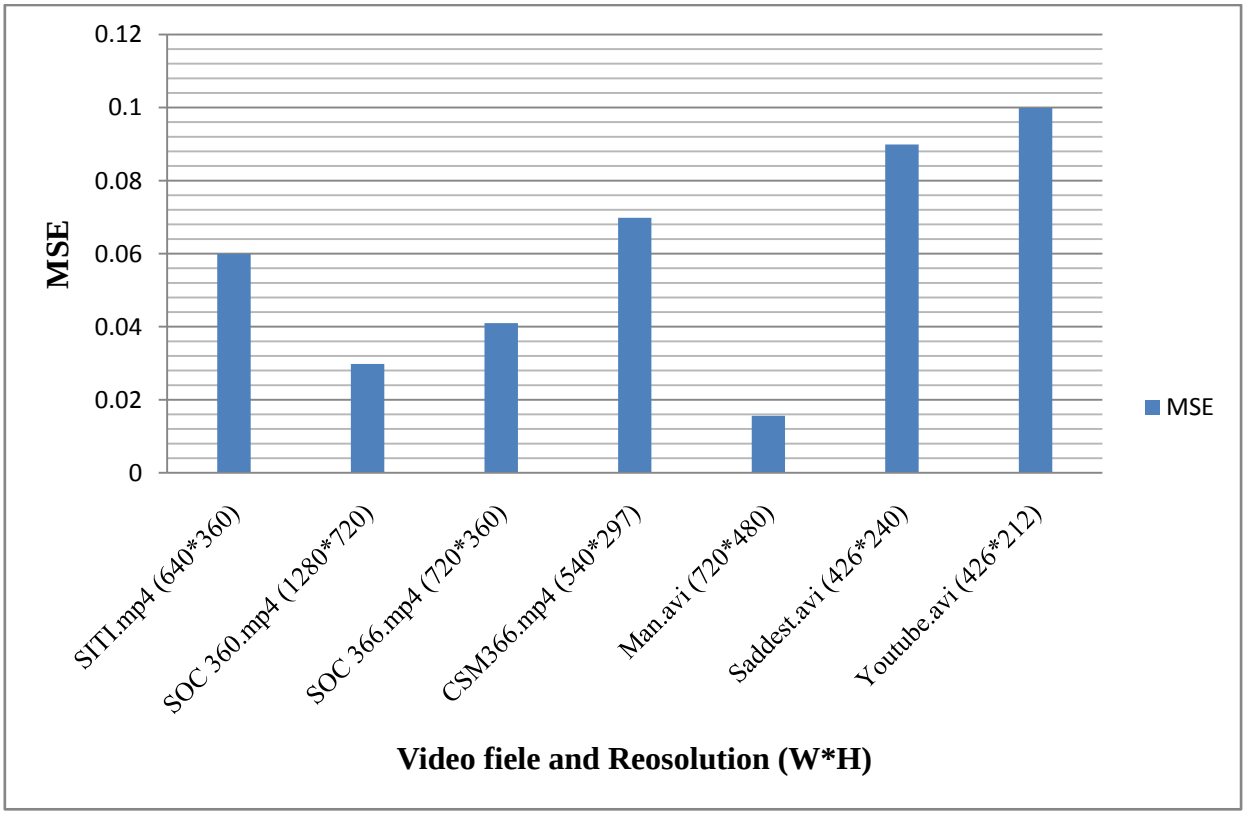

Figure 6: MSE vs. Video Resolution

6. CONCLUSIONS

As the internet revolution progresses steadily, hackers and attackers may reveal the content of secret or sensitive information of individuals or organizations. The most appropriate scientific remedy to the problems is steganography, data compression and cryptography. The role of steganography is to hide the existence of a message by employing some communication techniques such that the hidden message is not seen or discovered. Cryptography however, hides the content of the message whereas compression reduces the size of the hidden data. This research

brings to light the concept of effectively combining steganography, compression and cryptography specifically RSA cryptographic algorithm, Huffman code compression with LSB insertion. The preference of RSA over any other cryptographic algorithm is due to its ability to provide better security for large file size thereby reducing computational complexity. From the results obtained in this research, it can be concluded that when steganography is combined with cryptography and compression, higher levels of security, capacity and robustness are achieved. The distortion experienced in this study is negligible; therefore the study achieved increased security by the high PSNR values and low MSE and BER values. Again, in order not to send files of enormous or large size, a compression algorithm was introduced and implemented in this study. A lossless compression algorithm popularly called Huffman coding was used on the message to be hidden to increase capacity. Hence, for a higher security the message is encrypted and for more embedding capacity appropriate compression technique is used to get encrypted compressed message for embedding. Lastly, most of the steganographic applications already in existence can hardly handle all type of video file sizes. The most commonly one normally used is the .AVI. The proposed system deals with different type of video files including .AVI, MP4, MPEG and .flv. The use of RSA encryption and Huffman code with LSB in video steganography is contemporary field and has stupendous purview of research study.

7. REFERENCES

[1] Ramalingam, M. (2011). Stego Machine–Video Steganography using Modified LSB Algorithm. *World Academy of Science, Engineering and Technology*, *74*, 502-505.

[2] Wajgade, V. M., & Kumar, D. S. (2013). Enhancing Data Security Using Video Steganography. *International Journal of Emerging Technology and Advanced Engineering*, *3*(4), 549-552.

[3] Dengre , A. R., Gawande, A. D. Deshmukh, , A. B.(June, 2013). Effect of Audio Steganography based on LSB insertion with Image Watermarking using AVI video . International Journal of Application or Innovation in Engineering & Management (IJAIEM), 2(6), 2319 – 4847.

[4] Budhia, U., Kundur, D., & Zourntos, T. (2006). Digital video steganalysis exploiting statistical visibility in the temporal domain. *Information Forensics and Security, IEEE Transactions on*, *1*(4), 502-516.

[5] Das, S., Das, S., Bandyopadhyay, B., & Sanyal, S. (2011). Steganography and Steganalysis: different approaches. *arXiv preprint arXiv:1111.3758*.

[6] Kaur, R. and Singh, T. (2015). Hiding Data in Video Sequences using LSB with Elliptic Curve Cryptography. International Journal of Computer Applications (0975 – 8887) Volume 117 – No. 18

[7] Odeh, A., & Elleithy, K. (2012). Steganography in Arabic Text Using Zero Width and Kashidha Letters. *International Journal of Computer Science & Information Technology (IJCSIT)*, *4*(3), 1-11.

[8] Sumathi, C. P., Santanam, T., & Umamaheswari, G. (2014). A Study of Various Steganographic Techniques Used for Information Hiding. *arXiv preprint arXiv:1401.5561*.

[9] Kundalakesi,M., Sharmathi.R, Akshaya.R (2015) Overview of Modern Cryptography. International Journal of Computer Science and Information Technologies(IJCSIT), Vol. 6 (1), 350-353.

[10] Basheer, R & Safiya M.K (2014).Video data hiding in selective pixels of forbidden zone using mapping function. International Journal of Advanced Computer Technology (IJACT).ISSN:2319-7900.

[11] Jain, V. (2012). Public-Key Steganography Based On Modified LSB Method. *Journal of Global Research in Computer Science*, *3*(4), 26-29.

[12] Kour, H. & Kaur, S. (2015).Data Hiding Using MLSB Steganography. International Journal of Advanced Research in Computer Science and Software Engineering. Page |994, Volume5, Issue 1,ISSN: 2277 128X

[13] Vegh, L., & Miclea, L. (2015) Improving the Security of a Cyber-Physical System using Cryptography, Steganography and Digital Signatures.International Journal of Computer and Information Technology (ISSN: 2279–0764),Volume 04,–Issue 02,

[14] Mohanta, H. K. (2014). Secure Data Hiding using Elliptical Curve Cryptography and Steganography. *International Journal of Computer Applications*, *108*(3).

[15] Shukla, C. P., & Singh, A. K. (2014). Secure Communication with the help of Encryption in Video Steganography.Current Trends in Technology and Sciences.ISSN: 2279-0535. Volume: 3, Issue: 6

[16] Bodhak, P. V., & Gunjal, B. L. (2012). Improved protection in video steganography using dct & lsb. *International journal of engineering and innovative technology (IJEIT)*, *1*(4).

[17] Ahmed, Z. H. (2014). *Comparison of data hiding using LSB and DCT for image* (Doctoral dissertation, Universiti Tun Hussein Onn Malaysia).

[18] Deshmukh, P. R. & Rahangdale, B. (2014).Data Hiding using Video Steganography. International Journal of Engineering Research & Technology (IJERT).ISSN: 2278-0181.Vol. 3 Issue 4.