Steganography of Encrypted Messages inside Valid QR Codes Using Wavelet Transforms

Rafath Fatima Patel[1], Dr Y.V.S.Sai Pragathi[2]
[1]M.Tech, Scholar, [2]Professor,
Stanley College of Engineering and Technology for Women
Email: rafathkohinoor@gmail.com

ABSTRACT
Steganography is the initial layer of protection in information security because it conceals a malicious payload (information) behind a seemingly innocuous file (container), allowing it to be sent without being detected by the adversary. Steganographic methods rely only on the envelope to conceal the payload. In this research, we provide a steganographic scheme in which the container is used to both conceal the payload and mislead the attacker. In order to do this, we use QR codes as a storage medium. Aside from the payload, the QR codes created by our suggested system may also convey the system's normal message. The message can be read by anybody, but the payload is secured by a private key. Because the message and the payload are independent, any one may be created without the other. Using it to our advantage, we may craft a message that offers our opponent false information. By conducting extensive tests, we demonstrate that the produced QR code is (valid), or indistinguishable from a standard QR code. This makes it seem to be harmless and less vulnerable to assault from an adversary. In addition, it is vulnerable to steganography assaults, takes up little space, and has passable noise immunity.

INTRODUCTION:

Sensitive data, such as email passwords and financial account numbers, are routinely transmitted over the internet. It's crucial that this data is secured so that it doesn't end up in the wrong hands. If you need to safeguard sensitive data from an attacker, you can use both cryptography and steganography. Cryptography refers to the practice of encoding and decoding data to ensure its safety in transit. That part is unreadable because the sender is processing it in a way that makes it so. The generated cipher text is normally distributed, meaning that each bit has an equal chance of being either 1 or 0. This information is encrypted and can only be read by using a special key. Unfortunately, there are times when information security measures are insufficient and its very existence must be concealed. Steganography is the practice of concealing malicious code (payload) within a seemingly innocuous, non-secret file (container). The words "stegos," meaning "cover," and "grafia," meaning "writing," are the etymological roots of the term. Images and other forms of multimedia are favored as containers to conceal the payload because of their tolerance for distortion. Discrepancies in encryption systems inspire new study of steganographic methods. In contrast to encryption, which makes the cipher text vulnerable to attacks, steganographic techniques allow the secret information to be transmitted undetected. When fully developed, quantum computing, for instance, can be used to decrypt encrypted data. Also, the use of cryptographic methods may be restricted or outright forbidden by the laws of some nations.

Due to this, people and businesses started looking for other options. Businesses are beginning to see the potential of steganography, which allows them to conceal information about new products within a seemingly innocuous family photo, rather than more suspiciously transferring an encrypted file. To combine the benefits of steganography and cryptography, it is recommended to encrypt the payload before embedding it in a container. Since the encrypted payload is distributed uniformly, it can be mistaken for noise if it is ever detected. What's more, even if the payload is recovered, it will still be unreadable. The container is an untapped potential in steganography. For example, an image is only used to conceal the malware's payload. More efficient utilization of the container is a matter we must address. For instance, the adversary could be tricked by providing false information via the container. The use of QR codes allows for this to happen. QR codes, or quick response codes, are a type of two-dimensional bar code that can be read by scanning the image with a camera. They're versatile enough to encode anything from alphanumeric data to control sequences. QR codes can store a variety of data, including text, URLs, and IDs. Denso-Wave, a Toyota subsidiary in Japan, introduced QR codes in 1994. Quick Response (QR) codes were utilised as a means of speedily monitoring automobile components. Thereafter, they went global, especially within the fields of advertising and commerce. A quick and simple method of disseminating data, QR codes are widely popular. Simply use your mobile device to take a picture of the QR code. In the field of marketing and promotion, they achieved remarkable success. They are ubiquitous, appearing on everything from product packaging to billboards.

An example of a QR code with its structure is shown in FIGURE 1

(a) QR code structure

Pattern	Definition
	Quiet Zone
	Format
	Timing
	Version
	Position
	Alignment
	Data

(b) Legend

A potential of steganography that has not yet been fully utilized is the container. An image, for example, is just used to hide the payload and that is it. We need to address how

to use the container more. For example, using the container to give misleading information to the adversary. This can be achieved by using Quick Response (QR) codes. QR codes are machine-readable two-dimensional bar codes that can be read through cameras [3]. They can encode various types of information, such as alphanumeric and control codes. A phone number, a URL and an ID are examples of information that can be embedded in QR codes [3]. QR codes were firstly presented by Denso-Wave, a Japanese company which is part of Toyota in 1994. QR codes were used as a tool to quickly track vehicle parts. From there, they spread to the whole world especially in the marketing and business industries [3]. QR codes are convenient, easy to use way to share information. All you have to do is taking your cellphone and snap a picture of the QR code. They made a huge success in the advertising and marketing businesses. They can be found everywhere; from food labels to big advertising signs.

A QR code consists of the following items [5].

1. A quiet zone which is a border of empty space required for reading the QR code. The quiet zone is used to ease the symbol detection.
2. Formatting information which determines the mask pattern and the error correction level used in the QR code.
3. A timing pattern which recognizes the central coordination of each cell with alternating black and white patterns. It corrects the central coordination of the cell when it is distorted or when there is an error in the cell area.
4. The versions of QR code which range from Version 1 to Version 40. Each version has different cell numbers and configurations.
5. A position pattern which is a pattern for finding the correct position of the QR code. By positioning this pattern at the three corners of a QR code, the position, the size, and the angle can be discovered. It can be sensed in all directions (360^o).
6. An alignment pattern which is a pattern used for adjusting the distortion of the QR code, especially nonlinear distortions.
7. The data which are encoded as '1' or '0' and is converted to black and white cells inside the QR code. A Reed- Solomon code [5] is inserted with the data for error correction.

The QR code generation process, based on the latest QR code standard (ISO/IEC 18004:2015) can be summarized as follows [5], [6].

1. Choose the data to be encoded.
2. Select one of the 4 levels of error correction (ECL). A higher ECL can withstand more damage to the QR code, but this comes at the cost of increased QR code size.
3. Divide the encoded data into groups or segments, each group consists of a header and a payload.
4. Select an appropriate QR size (version) based on the data groups and ECL.
5. Concatenate the data groups and add a terminator in the end.
6. Represent the resulted data bits as a sequence of bytes. Add padding bits to the data sequence such that it can be divided into sequence of equally-sized blocks. Each error correction code (ECC) is calculated and added to its corresponding block. The

resulted data produce the final 8-bit codewords. These codewords are drawn in the QR code.

7. Start with an empty square grid based on the QR code size.
8. Add the QR code overheads (patterns) (alignment, version info, timing, finders etc.). These overheads do not include any user data.
9. Add the codewords into the QR code beginning from the bottom-right, then the data are added in zigzag mode
10. going upwards and downwards. Two QR code columns are used at a time.
11. Determine a mask pattern to process the data patterns. The mask can be chosen manually or automatically. If the mask is chosen automatically, then the mask with the least penalty is applied.
12. Render the resulted QR code as an image which can be printed or stored in digital image format.

The main features of QR codes are summarized as follows [5].

1. Advanced error correction. QR codes can withstand up to 50% damaged areas without affecting the message.
2. 360^o direction scanning. QR codes can be scanned from any direction.
3. Small printout size. QR codes can store messages in 2D, thus making it space-efficient compared to 1D codes.
4. High data capacity. A QR code can encode 4,296 alphanumeric characters, 7,089 numeric characters and 1,817 kanji characters.
5. High speed scanning. A cellphone with camera can easily and quickly acquire the message from a QR code.

Our Motivation: The motivation behind our work is taking advantage of QR codes as containers to hide the payload. These advantages can be summarized as follows.

1. A QR code can be presented as a coded message as well as an image. If we can hide the payload in the QR code in a way that only the intended user can read it (i.e. encrypted) without affecting the message, then we can use the message to mislead the adversary.
2. Usually, the payload degrades the container. But thank- fully, the QR code can be up to 50% damaged without affecting the message. This is a very good property for a container as it can hide a large payload size compared to the QR code size.
3. A QR code is presented as a binary image, which makes it more size-efficient compared to other containers such as colored images. Moreover, being a binary image makes QR codes very compressible compared to other images. This is because there are a lot of adjacent pixels with the same value (‘0’ or ‘1’) unlike other images. Lossless compression algorithms such as run length encoding [7] can be used effectively to compress QR codes. We note here that lossy compression of images, if used, must be handled with care because it can negatively affect the payload [7].

RELATED WORK

We first review some related work to steganography then we review some work related to

the application of QR codes in information security.

STEGANOGRAPHY

Sahu and Swain [8] presented a data hiding system which is built using pixel value differencing and modulus function (PVDMF). Their system is implemented in two versions. The first version (PVDMF v1) increases the peak signal-to-noise ratio (PSNR) and the second version (PVDMF v2) increases the embedding capacity (EC). PVDMF v1 and v2 are embed- ding the payload by calculating the difference between two successive pixels depending on adaptive range table [8].

Sahu and Swain [9] also presented two enhanced reversible data hiding (RDH) based systems. Their first system is an enhanced reversible dual images least significant bit (LSB). While the second system uses four duplicate cover images for hiding the payload using n-rightmost bit replacement (n-RBR) and modified pixel value differencing (MPVD). Experimental results show that the presented sys- tem provides high noise immunity.

Swain and Sahu [10] proposed an n-rightmost bit (n-LSB) substitution image steganographic system to embed the pay- load in, where 1 n 4. Their system hides the payload by calculating the difference between the n-LSB for every pixel in the image and n-LSB of the payload. Their system enhances PSNR for low value of n, increases EC for high value of n, evades the fall of boundary problem (FOBP) and provides immunity against salt and pepper noise.

Swain and Sahu [11] presented a multi steganographic sys- tem that is based on LSB which increases EC and enhances image quality. In their system, each pixel in the image is used to compute four new pixels in which the payload is hidden. Then these pixels are processed to enhance the image qual- ity. Their system improves PSNR and endures steganalysis attacks.

Sahu and Swain [12] presented a steganographic system based on pixel value differencing (PVD) and LSB. it is focused on the error block problem (EBP) and FOBP. The image is segmented into blocks with two adjacent pixels. The blocks are split into three levels subject to the differ- ence of pixel values. The block level and the pixel differ- ence determine the embedding capacity of the block. Their system improves PSNR and embedding capacity. In addi- tion, the system is immune to pixel difference histogram (PDH) analysis.

Marvel et.al. presented a steganographic system that depends on spread spectrum in images [13]. Their sys- tem conceals the payload inside an image without affect- ing the image dynamic range and size. They used spread spectrum techniques for image restoration and error control.

Liao et.al. proposed a medical joint photographic experts group (JPEG) image steganographic system based on the dependencies of inter-block coefficients [14]. Their system depends on keeping the deferences between the discrete cousin transform (DCT) in the same place in neighboring DCT blocks as close as possible. In the embedding process, the cost values are assigned dynamically consistent with the alterations of inter-block neighbors.

Rachmawanto et.al. implemented a security system that merges between cryptography by using one-time pass- word (OTP), vernam encryption and steganography by using DCT in a digital image [15]. Their system is tested using PSNR and normalized cross correlation

(NCC) to test the quality of the decrypted message. Their sys- tem is immune against JPEG compression and median filters.

Muhammad et.al. presented an image steganographic sys- tem which is based on stego key-directed adaptive least significant bit (SKA-LSB) [16]. In their system, a key is encrypted using a two-level encryption algorithm (TLEA). After that, the message is encrypted using a multi-level encryption algorithm (MLEA), then the encrypted payload is concealed in the host image using an adaptive LSB substi- tution. Their system achieves a reasonable balance between quality and security.

Zhou et.al. presented a steganographic system based on coverless images without embedding [17]. In their system, a database is constructed by a group of chosen images, and these images are indexed in the database by generating hash- ing sequences for these images. The payload is divided into segments. An image is sent if its hashing sequence is match- ing with the segment. Their system is robust to luminance changing and noise.

QR CODES IN INFORMATION SECURITY

Rani at.al. proposed a secure system in which they combined between steganography and QR codes [3]. Their system con- sists of two parts. The first one is creating a QR code of the encrypted payload. The second one is hiding the generated QR code inside a colored image. The hiding process does not generate a visible image distortion and generates a very minimal bit error rate (BER).

Barrera et.al. implemented a system that uses QR codes in optical encryption as containers [18]. They choose QR codes as containers in their system because of their tolerance to pollutant speckle noise. In addition, QR codes are easy to read using cellphones' cameras. The results show that their system is more prone to noise compared to normal optical encryption. Dey et. al. presented a steganographic system which is based on a randomized intermediate QR host that is embed- ded with an encrypted payload [19]. First, the payload is encrypted then concealed in a QR code. Then, the QR code is hidden inside an image. Using double encryption and embedding techniques makes the system hard to break. But on the other hand, it makes the system less time-efficient.

PROPOSED SYSTEM

The proposed system consists of two parts: the payload embedding and the payload extraction. The flowchart of embedding the payload in the QR code is shown in Fig. 3.

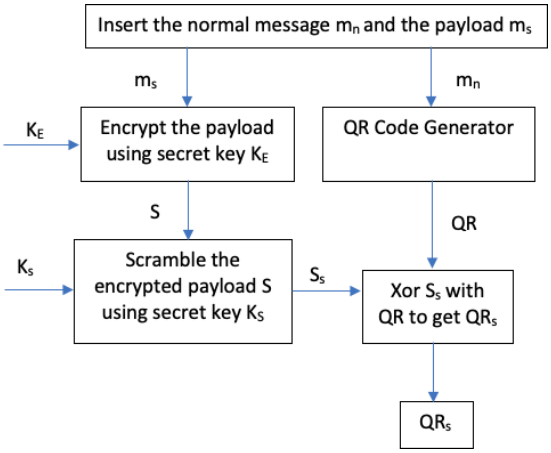


FIGURE 2 Generating QR code embedded with the payload.

The message embedding in the QR code is explained as follows.

1. Choose N 17 $4x$ for any arbitrary integer value x where N is the dimension (width or height) of the QR code image. This is a prerequisite for QR code genera- tion [4]. Note here that QR codes are square images.
2. Choose any message m_n and generate a QR code using N, m_n.
3. Generate the payload m_s and use an encryption algo- rithm such as the advanced encryption standard (AES), OTP or RSA [20], [21] to encrypt m_s to get the ciphertext S.
4. Scramble S using a scrambling algorithm such as the baker map and tent map [22] to get S_s as follows. First, generate an empty image with size equal to the QR code. Then insert S in that empty image. Scramble the image using baker map to get S_s.
5. Embed S_s in the QR code by xoring S_s with pixels in the QR code $QR_s = QR \oplus S_s$.

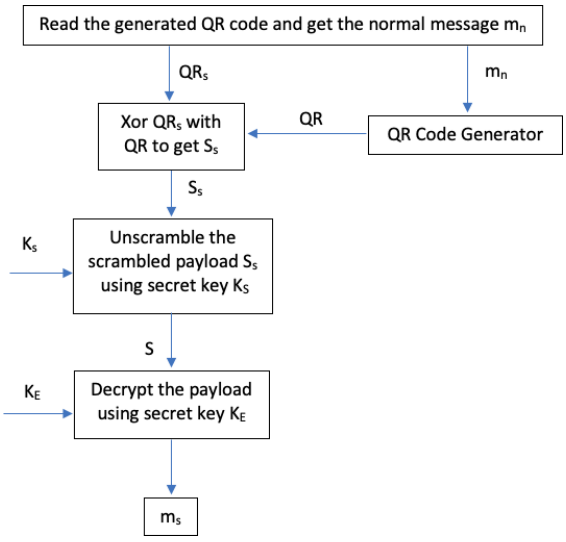


FIGURE 3. Extraction of the payload.

RESULT AND DISCUSSIONS

Here, we visually compare the created QR codes to standard QR codes to ensure a high level of accuracy. In addition, we scanned both QR codes using QR code reading app on a mobile phone. Through the use of correlation coefficients and the number of pixels changing per second (NPCR), we also examine the parallels between the two QR codes.

NPCR calculates the proportion of total pixels that are unique between two photos. The lower the NPCR, the more similar the two pictures are.

Let x and y be two pictures with horizontal and vertical pixel indices, respectively. Create a new array where Di,j=0 and 1 if xi,j=yi,j and -1 otherwise. What is meant by the NPCR is:

To calculate NPCR, just multiply the percentage of DNA present in a sample by the number of copies of each gene in the sample (HWi,.....(1)

where W and H are the image's width and height.

Here, we point you that besides the UACI and the histogram analysis, there are many additional testing methods available. Due to QR codes being binary pictures, however, the findings of these tests will be identical to NPCR (i.e. black and white, no gray).

Then, we take into account the dimensions of the QR code and the message size that does not distort the QR code or deface the message to determine the maximum allowable payload size. Using the aforementioned tools, we construct various QR codes with varying size ratios between (1) the payload and QR code, and (2) the payload and the message. The degree to which these QR codes differ is also measured.

We also evaluate the proposed system, the 1-LSB, 2-LSB, and 3-LSB systems using the PSNR, SSIM, and embedding capacity (EC).

The Peak-to-Surface Noise Ratio (PSNR) is defined as the ratio of the highest value of a picture to the difference between the original and steganographic versions of the image. In general, the higher the ratio, the greater the quality of the steganographic picture. The expression looks like this:

PSNR=10×log10W×H×255×255/∑Wi=1∑Hj=1

(xi,j−yi,j)2.....(2)

where xi,j are the values of the pixels in the original picture and yi,j are the values of the pixels in the stego-image at locations i,j.

You may also use the structural similarity index (SSIM) to find out how much the original and stego-image seem alike. It may take on values between 1 and 1, with 1 denoting perfect congruence between the two pictures [10, 11]. Specifically, it may be written as.

SSIM=(2moms+c1)(2σos+c2)/(m2om2s+c1)(σ2o+σ2s+c2).....(3)

where the image's mean, variance, and standard deviation are denoted by m, m2, and, respectively. o and s indicate the authentic and steganographic representations, respectively. the correlation between the two pictures is denoted by os. For a grayscale picture with a maximum value of 255, the constants c1=k1L and c2=k2L are respectively equal to L=0.01 and L=0.03, respectively. When referring to containers, the embedding capacity is the largest possible payload size that may be safely concealed within.

Finally, we put the suggested system through its paces in terms of its ability to filter out both the noise provided by

(1) QR code symbol noise, such as QR code bending or improper reading angle, and

(2) message noise.

In these simulations, we build the system using the SHA3-256 hash function and use the advanced encryption standard (AES) for encryption and the baker map for scrambling. Here, we emphasise that any method for encrypting data, scrambling data, or creating a hash value may be employed.

It is important to emphasise that all of the paper's research is relevant to both scanned and printed QR codes. The same message can be gleaned from reading either type of QR code, the payload can be blocked in the same way (by generating the QR code first, and then printing it), and the payload can be extracted in the same way (by scanning the printed QR code and then extracting it) using steganography for both types of QR codes.

Figure 4 Embedding the Secret Data

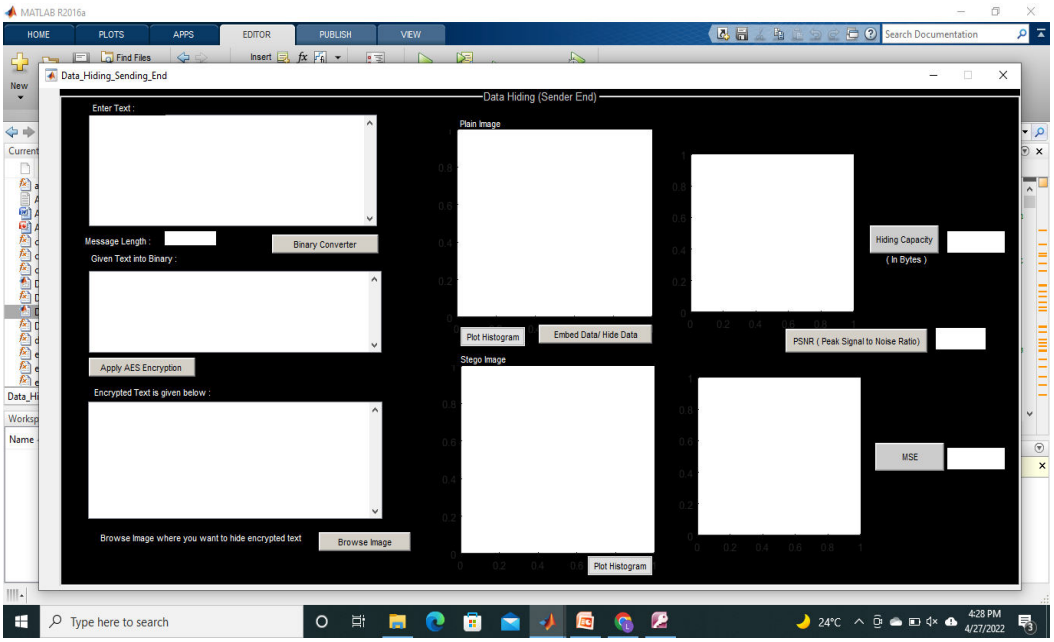


Figure 5 Extraction of secret data

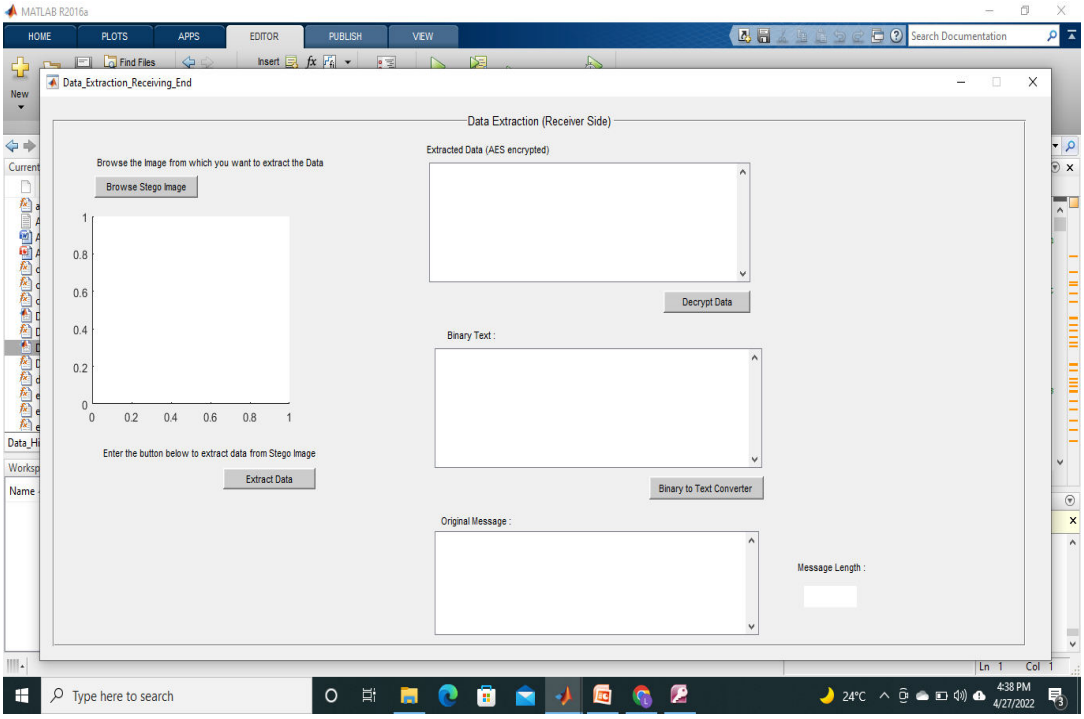
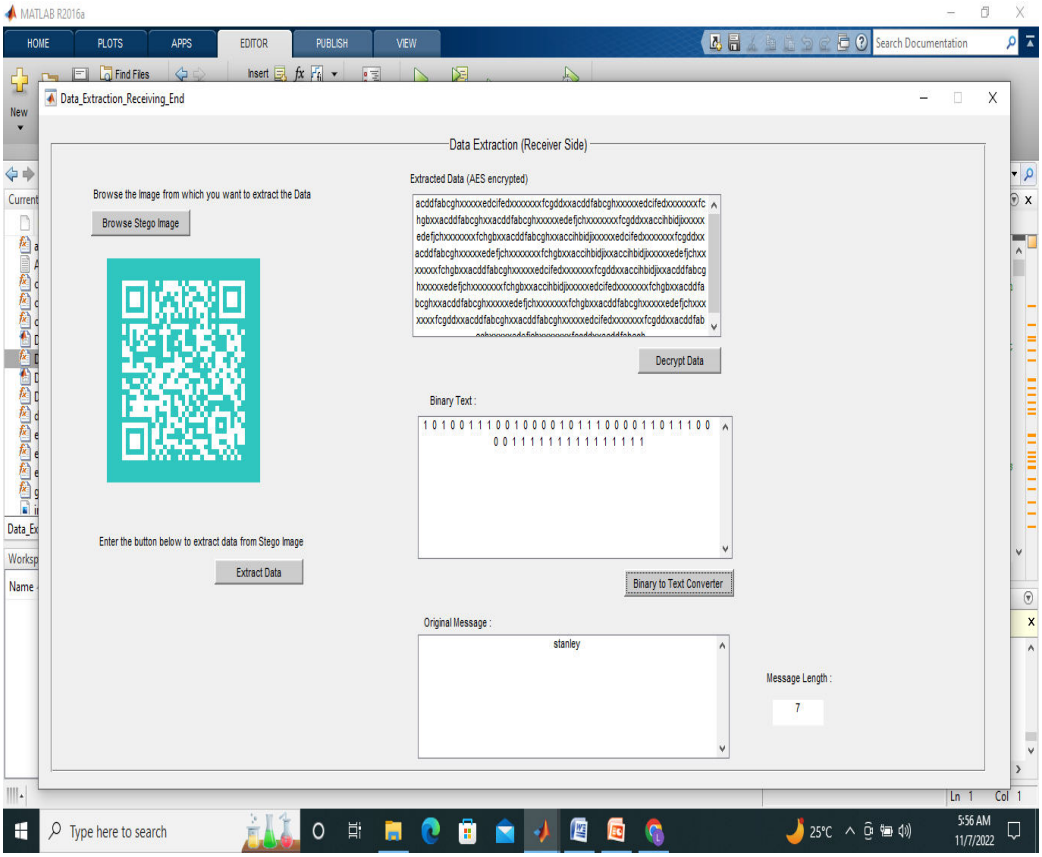


Figure 5 Encryption of message inside a valid qr code

Figure 6 Decryption of message inside a qr code Stegno qr code



CONCLUSION

The suggested approach successfully protects sensitive data. First, no one will suspect that sensitive information is being sent, and second, encryption will keep it safe. The given system successfully conceals the data via image-steganography with the help of LSB, and it adds extra protection for the data with the help of the cryptographic method AES-256, and it fortifies the vulnerable spot that a hacker could target, which is the key, by using the hashing method SHA 256. We now have a closed, unbreakable system. The experimental work done guarantees the practicability of the suggested system. The outcomes take on a disguised appearance, a subtle distortion that makes it nearly impossible for the attention of attackers to be drawn to it; even after having it tranced, no one was able to access the data without the genuine key.

REFERENCES

1. Provos, Niels, and Peter Honeyman. "Hide and seek: An introduction to steganography." IEEE security & privacy 99.3 (2003): 32-44.
2. Henri Gilbert and Helena Handschuh, “Security Analysis of SHA-256 and Sisters*”, 2003.
3. Selent, Douglas. "Advanced encryption standard." Rivier Academic Journal 6.2,

ISSN (Online): 2319-7064 Index Copernicus Value (2015): 78.96 ,(2010): 1-14.

4. United States National Security Agency (NSA), U.S. Department of Commerce, National Institute of Standards and Technology (NIST), Information Technology Laboratory (ITL), "Secure Hash Signature Standard (SHS) (FIPS PUB 180-2)", National Institute of Standards and Technology (NIST), 2001 : 9-22.
5. Charles G. Boncelet, Jr., Newark, DE (US); Lisa M. Marvel, Churchville, MD (US); Charles T. Retter, Belcamp, MD (US). "Spread spectrum and image stenographer", 2003.
6. Po-Yueh Chen* and Hung-Ju Lin, "A DWT based approach for image stenographer", DOI:10.6703/IJASE.2006. 4(3).275, 2006.
7. Domenico Bloisi and Luca Iocchi, "Image based steganography and cryptography",2007.
8. Ali Al-Ataby and Fawzi Al-Naima, "A modified high capacity image steganography technique based on wavelet transform", Vol. 7, No. 4, October 2010.
9. Shailender Gupta, Ankur Goyal, Bharat Bhushan, "Information hiding using least significant bit steganography and cryptography",
10. Saiful Islam*, Mangat R Modi and Phalguni Gupta "Edge-based image steganography"
11. *Khan Muhammad, Jamil Ahmad, Haleem Farman, Muhammad Zubair, "A novel image steganographic approach for hiding text in color images using HSI color model",
12. Khan Muhammad, Jamil Ahmad, Muhammad Sajjad, Muhammad Zubair, "Secure image steganography using cryptography and image transposition", October2015.
13. Mwaffaq Abu-Alhaija "Crypto-Steganographic LSB-based System for AES Encrypted Data", *(IJACSA)* International Journal of Advanced Computer Science and Applications, Vol. 10, No. 10, 2019.
14. Katzenbeisser, S. and Petitcolas, F., 2000. Information Hiding Techniques for Steganography and Digital Watermarking, Artech House. Ker, A., 2004.
15. Improved Detection of LSB Steganography in Grayscale Images, Proc. 6th International Workshop. Toronto (Canada), Springer LNCS, 3200, pp. 97–115. Ker, A., 2005.
16. Steganalysis of LSB Matching in Grey scale Images, IEEE Signal Process Letter, 12(6), pp. 441– 444. Kipper, G., 2003.
17. Investigator's guide to steganography. Auerbach Publishers. Koch, E. and Zhao, J., 1995.
18. Towards Robust And Hidden Image Copyright Labelling, Proceedings of 1995 IEEE Workshop on Nonlinear Signal and Image Processing, Halkidiki, Greece, pp. 452–455. Kuhn, T., 1996.
19. The Structure of Scientific Revolutions, Chicago: University of Chicago Press. Lawless, H. T., and Heymann, H., 1998.
20. Sensory Evaluation of Food: Principles and Practices. Chapman and Hall, New 20. York, NY, pp. 606–608. Lee, Y. K., and Chen, L. H., 2000.
21. . High Capacity Image Steganographic Model, IEEE Proc., Vis. Image Signal Process, 147(3), pp. 288-294.

22. Morkel, J. H. Eloff and M. S. Olivier, "An overview of image steganography", *Proc. ISSA*, pp. 1-11, 2005.
23. S. R. M. Mary and E. K. Rosemary, "Data security through Qr code encryption and steganography", *Adv. Comput. Int. J.*, vol. 7, no. 2, pp. 1-7, Mar. 2016.
24. J. Waleed, H. D. Jun, S. Saadoon, S. Hameed and H. Hatem, "An immune secret QR-code sharing based on a twofold zero-watermarking scheme", *Int. J. Multimedia Ubiquitous Eng.*, vol. 10, no. 4, pp. 399-412, Apr. 2015.
25. T. J. Soon, "QR code", *Synth. J.*, vol. 2008, no. 3, pp. 59-78, 2008.
26. ISO/IEC, 18004:2015, "Information Technology—Automatic Identification and Data Capture Techniques—QR Code bar Code Symbology Specification", 2015.
27. "Lossless compression of fragmented image data", Jun. 2019.