AADHAAR data encryption using Steganography Technique

Shreyas ML
Department of Computer Science And Engineering
Manipal Academy of Higher Education
Manipal University, Manipal,
Karnataka, India
shreyasml95@gmail.com

***Abstract*— According to Wikipedia "The Aadhaar Central Database is one of the largest government databases in the world, a 12 digit unique-identity number has been assigned to all enrolled Indian citizens"[1]. This database contains both the biometric and demographic data of the citizens. With the large amount of confidential and private data stored in one singular database, it essential that Aadhaar and Unique Identification Authority of India (UIDAI), the authority that maintains the database, continue to focus their attention wherever there are any security issues. This security issue can be solved by using Steganography techniques of data hiding. "The word steganography is combination of two Greek words steganos meaning 'covered, concealed, or protected,' and graphein meaning 'writing'". In this paper we see the application of Steganography on Aadhaar database to hide the demographic and biometric data of the enrolled Indian citizen. In the proposed method the demographic data is stored in a file and the biometric data is compressed using DCT compression and both are hidden inside the profile image. The unique Aadhaar identification number is used as the primary key for the database.**

Keywords—Steganography, Aadhaar, Encryption, DCT Compression

I. Introduction

"The Aadhaar project is the world's largest biometric identification system launched by the Indian government in 2009"[2].Till date 121 crore Aadhaar cards have been issued by the UIDAI. The aadhaar project has been linked to many government and public schemes such as MGNREGA and domestic LPG scheme for subsidy to people who are below poverty line. Aadhaar card is also a proof of identification and proof of residence. Government has made it compulsory to link aadhaar card to Bank accounts, mobile number and other government related subsidy plans[2]. For all these benefit schemes and to provide authentication to linked domains UIDAI collects basic demographic details like Name, Address, Gender, Date of Birth, mobile number and email ID, and biometric details such as all finger prints, two IRIS scans and facial photograph[3]. However, in recent days there is a considerable deliberation over the security related issues of aadhaar database. There are multiple incidents where aadhaar data has been leaked by the hackers by hacking government websites or third party leaks[4]. Multiple steps are taken by the UIDAI to secure the confidential data of people like usage of AES to encrypt the data while saving in database and communicating with third party authorities like Banking or telecom etc. for authentication purposes. However, there is still risk that when data is leaked it is vulnerable for decryption. We can overcome this risk by making use of steganography technique.

Steganography is a technique of hiding data such as plain text messages, image files, audio and video files in a cover object, invisible to naked eye and store or transmit the cover object so that even if the cover object is leaked it is hard to figure out what data is stored and where the data is stored by an unauthorized person. There are 2 types of steganography Linguistic steganography which deals with the text-based steganography and Technical Steganography is a technique used to manipulate images, audios and videos Nowadays with. The dependability of a steganographic system is evaluated according to multiple criteria. These are the quantity of information cover object replacement, the amount of information which will be hidden within the environment, and how much capacity it has. The endurance power is measured by how very little impacts are affected by visual and statistical attacks on the environment. With developing technology, many data hiding ways have emerged. the rise in the diversity of data concealing ways has necessitated the development and improvement of many steganaliz strategies. as a result of every data-hiding methodology uses its own method, the ways that to discover them are completely different.

II. Literarutre Survey

"In 2009 the government of India established a Unique Identification Authority of India (UIDAI), to generate and assign a 12 digit unique identification number that can be obtained voluntarily by the Indian citizens by collecting their demographic and biometric details"[5]. The demographic and biometric data of every enrolled Indian citizen is stored in Aadhaar's central database protected by AES encryption. The stored data will be used for authentication by government authorized third party sites like Banks, Telecom companies, LPG suppliers etc. The protection of data stored and the data communicated is of great importance. To safeguard the data from leaks and misuse[6].

Image steganography is used to hide data inside cover objects like images, audio and video. There are different techniques used to hide the payload in cover object[7] There are basically 2 methods of image steganography Spatial domain techniques and Transform domain techniques.

1. In Spatial domain technique we perform the data hiding process on the pixel values of the cover object so that the message hidden is invisible to the human eye on the cover object and also we have to make sure that the cover image before and after the data hiding process should look similar to human eye. To perform this type of data embedding different methods are used. Most common methods are analyzed this survey.

*A. **LSB Technique**:* LSB stands for Least Significant Bit. This is the method of data hiding technique in steganography. The idea behing this method is that if we alter the last bit value of the pixel of the image there wont be much visible difference in the color[8]. For example 1 is black ,changing the value to 0 will not make any visible difference since both 0 and 1 represent black. Before we start embedding data we convert the image to grayscale, resize the image if required and hide the message converted into binary bit by bit the LSB value of each pixel using XOR.

*B. **PVD Technique:*** PVD stand for Pixel Value Differnecing In this technique we make use of the fact that human eye can identify slight changes in the flat regions of the object while it is hard to identify changes in the edge regions.The cover image is divided into disjoint sets of two continuous pixels. The difference between two pixels are calculated and Each set difference values are calculated and divided into number of ranges.Larger difference means more bits can be used to embed the data. This method use to hide more data inside the larger range pixel blocks compared to LSB technique[7][9].The other advantage of this method is that the hidden data can be extracted without using the original cover image..

*C. **EBE Technique**:* EBE stands for Edge Based Embedding, in this technique we make use of the fact that human eye can identify minor changes in the flat regions of the image while it is hard to identify the changes in the edges[10]. In this technique the the algorithm applied will identify the edges if the cover image and data can be embedded in three LSB's.The attacker will have less suspicions when data is hidden in the edges of the cover image.The only drawback in this techniques that since we are using only pixels from detected edges the payload that can be hidden is very less compared to other methods.

D. Pixel Mapping Technique: In this spatial domain technique the pixels are selected from the cover image depending on a arithmetic function which relies on the pixel intensity value of the original pixel and its eight neighbors are selected in anti-clockwise direction[11]. Before concealing the data in the cover image we have to make sure that the original or seed pixel and its eight neighbors are inside the cover image frame or not. Data hiding is done by mapping every two and four bits of the secret data in each neighbor pixel based on specific features of the pixels.

E. Pixel Intensity or Gray Level Value (GVL) Technique: "In this technique we map the data by altering the gray level of the cover image pixels"[12]. This technique uses even and odd values to map data inside the image frame. It is a one on one function where each selected pixel is mapped with binary data which are choosen based on some arithmetic calculations.

The above mentioned are few Spatial Domain Techniques in image steganography

2. In Transform Domain technique we use specific mathematical functions on the pixels of the cover image to hide the data on it. Before performing the embedding function, the image has to converted domain such as frequency domain like DCT, DFT, wavelet domain like DWT, IWT or curvelet domain like DCVT etc.

*A. **DCT Technique:*** DCT stands for Discrete Cosine Transform. In this technique we divide the main object into multiple frequency spectrums such as, low, high and middle, making it easier to select the frequency band in which data can be embedded[13]. Hiding the data in middle frequency doesn't not scatter data to important parts of the cover image hence almost always middle frequency bands are selected. and also there is a less chance that high frequency components are focused targets in case of attacks or compressions

*B. **DFT Technique:*** DFT stands for Discrete Fourier Transform. This technique follows DCT technique but instead of cosine transform it uses foruier transform technique to transform the cover image[11]. This makes it harder to change with strong geometric distortion. Complexity is large compared to DCT technique.

*C. **DWT Technique:*** DWT stands for Discrete Wavelet Transform. A wavelet is a less valued wave which swings and decays in unit time. This method analyses the wavelet and uses its advantages as it performs "local" and "multi-resolution analysis(MRA)"[14]. MRA is used to compute a signal at distinct frequencies. This method converts the cover image into wavelet domain. Process the coefficient, hide the data. Then we perform an operation called "Inverse Wavelet Transform" to represent the stego image in the original image format.

*D. **IWT Technique:*** IWT stands for Integer Wavelet Transform. The existing methods as discussed above are robust when both cover image and secret image are grayscale. To hide color image, we make use of IWT technique[15]. This technique is applicable when both cover and secret images are color images. The secret image is concealed by taking three color components separately. Keys are generated using the color components of the cover image and the secret image. Instead of the image, keys are hidden in the cover image.

In DCT and DFT we independently process the coefficients without major interactions between them. The wavelet filters have floating (point) value coefficient, when the data is integer values the filtered data will be corresponding floating point values[11]. During reconstruction of the original image it will not be a perfect reconstruction. However, with IWT that maps integer to integer the output can be calculated as integer only.

In the Literature Survey we have discussed multiple methods of steganography. These are the major differences among them.

Table 1. Major differences between spatial and transform domain techniques.

Feature	**Spatial Domain Technique**	**Transform Domain Technique**
Image Integrity	There are less chances that the cover image will be altered	There is a risk of degradation of image if the coefficients are not chosen properly.
Concealing Capacity	High concealing capacity	Low concealing capacity
Complexity	Less mathematical complexity	Higher complexity in mathematical functions
Implementation	Easier to implement	Greater knowledge of embedding domain is required.
Robustness	Vulnerable to attacks and compressions	High tolerance for attacks and compressions
Flexibility	Rigidity based on image format	Has higher flexibility for concealing
Format	Format dependent	Format Dependent

The suitable technique among the above described techniques will be used to hide the demographic details. However, when it comes to biometric data we have 12 images (10 fingerprints and 2 IRIS scans). So we need to compress these images in order to embed them in the profile image. For compression we will be using DCT compression technique.

In this paper, it is aimed to use image steganography technique to hide Aadhaar data by compressing biometric data. The process will be explained in the next stage.

III. METHADOLOGY

As mentioned in the earlier sections Aadhaar contains two types of data. The demographic data of Aadhaar of an individual will be stored in a file. and the biometric data has 12 images all the 12 images will be compressed using DCT image compression technique and both will be hidden in the profile image of the respective individual by using DCT image steganography. For additional security the demographic data will be encrypted using RSA algorithm with public and private key before hiding them in the file.

After hiding all the details and data in the profile image the image and Aadhaar number are added to the database where the unique 12 digit Aadhaar number is stored as the primary key and the profile image of the respective card holder as the second field.

The following figure Fig 1 will show the flow of process.

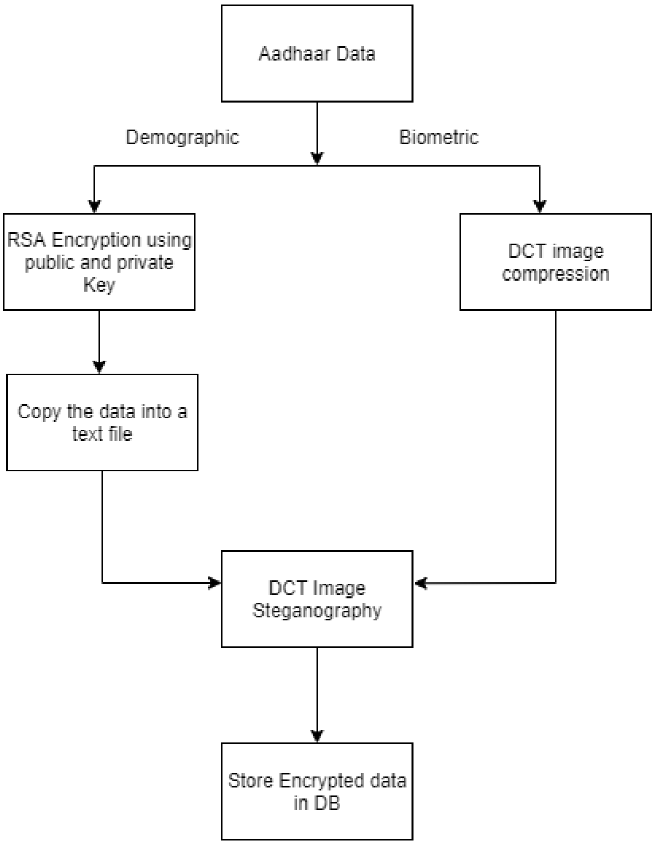
Aadhaar Data
Demographic
Biometric
RSA Encryption using public and private Key
DCT image compression
Copy the data into a text file
DCT Image Steganography
Store Encrypted data in DB

Fig 1. Proposed process Flowchart

In 2012 Maneesha Gupta and Dr.Amit Kumar Garg have explained DCT compression technique in their thesis "Analysis Of Image Compression Algorithm Using DCT". Image compression is a process of compressing digital images into lower pixel resolution. The DCT compression technique separates two parts and transforms it from spatial domain to frequency domain.

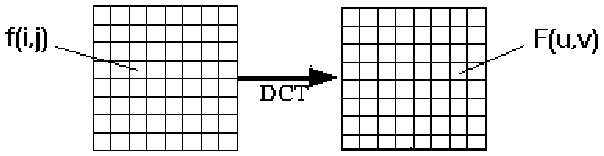

Fig 2. Transformation function in DCT.

In Fig 2. We observe that the pixel values of corresponding ith row and jth column are converted to frequency domain using the mathematical formula (1).

To compress an image into lower resolution we make use of the following algorithm.

1.Divide the source image into 8x8 blocks.

2.Apply 2-D DCT transformation on 2D (NxM image). 2-D DCT transformation can be applied by using the following formula.

$$F(u,v) = \left(\frac{2}{N}\right)^{\frac{1}{2}} \left(\frac{2}{M}\right)^{\frac{1}{2}} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} \Lambda(i).\Lambda(j).cos\left[\frac{\pi.u}{2.N}(2i+1)\right] cos\left[\frac{\pi.v}{2.M}(2j+1)\right].f(i,j) \tag{1}$$

3.Perform quantization minimizing the number of bits needed.

4.Applying entropy encoding to compress image with representation of color with less bits.

The compressed image can be decompressed and brought back to original resolution and size by performing decompression algorithm.

1.Entropy Decoding where we decompress the image to its original resolution.

2.Perform inverse quantization to restore lost bits.

3.Applying formula (2) we get the inverse of the 2-D transformed image.

$$\Lambda(\xi) = \begin{cases} \frac{1}{\sqrt{2}} & \text{for} \xi = 0 \\ 1 & \text{otherwise} \end{cases} \quad (2)$$

By this method the compressed image will be reconstructed to its original form.

In 2011 Blossom Kaur, Amandeep Kaur and Jasdeep Singh had proposed hiding watermark using DCT in “STEGANOGRAPHIC APPROACH FOR HIDING IMAGE IN DCT DOMAIN”. Using the same approach, we are going to hide the data in the profile image.

As mentioned in the DCT compression DCT image steganography of 2-D image also follows formula (1) for converting spatial domain to frequency.

After conversion into frequency domain the image will be divided into three frequency ranges high, low and medium. As mentioned in the earlier section medium range frequencies are selected for advantages. The example of 3 divisions is showed in fig 3.

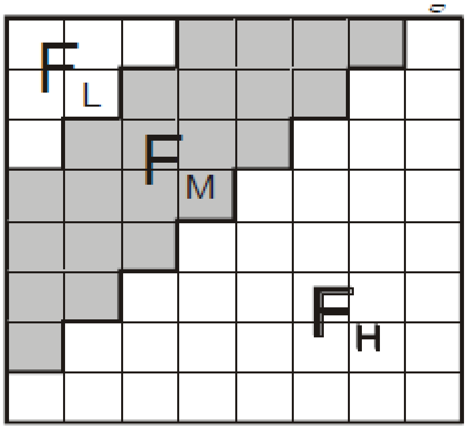

Fig 3. DCT regions(Courtesy: [16])

Two consecutive locations are selected in the medium frequency range based on the quantization value. After selecting the appropriate locations.

(Work in progress)

IV. Results.

From the proposed method we get the resultant database as shown in Fig 4. Which consists of only the unique 12 digit Aadhaar number as the primary key and the profile image of the respective card holder as the second field. All other data are hidden inside the profile image to increase the security of the data. (Work in progress)

Aadhaar number	**Profile Image**
(Sample data has to entered after implementation)	

Fig 4. Schema of the resultant database.

V. Conclusion

(Work in progress)

VI. Future Scope

The Aadhaar system is not only used for identification but also to provide subsidy for deserving people in government plans like LPG subsidy, Scholarships etc. Also government insists on linking Aadhaar number with bank accounts , mobile phones etc. For this aadhaar validation has to done on the third party side. The aadhaar data has to be communicated between UIDAI and the third party. By using stegnographic approach discussed in this paper we can securely transmit the aadhaar data for authentication.(Work in progess).

VII. REFERENCES

[1] "Aadhaar", En.wikipedia.org, 2019. [Online]. Available: https://en.wikipedia.org/wiki/Aadhaar. [Accessed: 23- Sep- 2019].J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] "Steganography", En.wikipedia.org, 2019. [Online]. Available: https://en.wikipedia.org/wiki/Steganography. [Accessed: 23- Sep- 2019]

[3] "Home - Unique Identification Authority of India | Government of India", Unique Identification Authority of India | Government of India, 2019. [Online]. Available: https://uidai.gov.in/. [Accessed: 23- Sep- 2019]

[4] "11 questions on Aadhaar and its misuse, answered by the UIDAI", The Economic Times, 2019. [Online]. Available: https://economictimes.indiatimes.com/news/economy/policy/11-questions-on-aadhaar-and-its-misuse-answered-by-the-uidai/articleshow/62538926.cms. [Accessed: 23- Sep- 2019].

[5] "Aadhaar security breaches: Here are the major untoward incidents that have happened with Aadhaar and what was actually affected- Technology News, Firstpost", Tech2, 2019. [Online]. Available: https://www.firstpost.com/tech/news-analysis/aadhaar-security-breaches-here-are-the-major-untoward-incidents-that-have-happened-with-aadhaar-and-what-was-actually-affected-4300349.html. [Accessed: 23- Sep- 2019].

[6] N. Christopher, "Security experts say need to secure Aadhaar ecosystem, warn about third party leaks", The Economic Times, 2019. [Online].Available:https://economictimes.indiatimes.com/news/politics-and-nation/there-is-a-need-to-secure-full-aadhaar-ecosystem-experts/articleshow/63459367.cms?from=mdr#targetText=While%20the%20Unique%20Identification%20Authority,measures%20to%20keep%20data%20safe. [Accessed: 23- Sep- 2019].

[7] Tiwari, Anjali, Seema Rani Yadav, and N. K. Mittal. ”A review on different image steganography techniques.” International Journal of Engineering and Innovative Technology (IJEIT) Volume 3,Issue 7 (2014): 121-124.

[8] K.Thangadurai, G. Devi, “An analysis of LSB based image steganography techniques,” International Conference on Computer Communication and Informatics (ICCCI -2014), Coimbatore, INDIA

[9] D. Wu and W. Tsai, "A steganographic method for images by pixel-value differencing", Pattern Recognition Letters, vol. 24, no. 9-10, pp. 1613-1626, 2003. Available: 10.1016/s0167-8655(02)00402-6.

[10] K. Naveen BrahmaTeja, Dr.G. L. Madhumati, K. Rama KoteswaraRao, “Data Hiding Using EDGE Based Steganography”, International Journal of Emerging Technology and Advanced Engineering, Volume 2, Issue 11, November 2012

[11] Bhattacharyya, S., Sanyal, G., Study and analysis of quality of service in different image based steganography using Pixel Mapping Method (PMM), International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868 Foundation of Computer Science FCS, New York, USA Volume 2– No.7, May 2012.

[12] S. M and M. R, "EFFICIENT METHOD FOR HIDING DATA BY PIXEL INTENSITY", *International Journal of Engineering and Technology (IJET)*, vol. 5, no. 1, pp. 74-80, 2013. [Accessed 23 September 2019].

[13] B. Kaur, A. Kaur, and J. Singh, "Steganographic approach for hiding image in DCT domain," International Journal of Advances in Engineering & Technology, 2011.

[14] P. Chen, and H. Lin, "A DWT Approach for bnage Steganography," International Journal of Applied Science and Engineering 2006. 4, 3: 275:290.

[15] Hemalatha S., Acharya, U.D., Renuka A. and Priya R. K., An Integer Wavelet Transform Based Steganography Technique for Color Images, International Journal of Information & Computation Technology. ISSN 0974-2239, Volume 3, Number 1, pp. 13-24, 2013.

[16] Maneesha Gupta, Dr.Amit Kumar Garg "Analysis Of Image Compression Algorithm Using DCT",1nternational Journal of Engineering Research and Applications (l.IERA) ISSN: 2248-9622 www.ijera.com Vol. 2, Issue 1, Jan-Feb 2012,pp.515-521

[17] Blossom Kaur, Amandeep Kaur, Jasdeep Singh, “Steganographic Approach for Hiding Image in DCT Domain”, in International Journal of Advances in Engineering & Technology, vol. 1, pp. 72-78, July 2011.