Data Protection Tool for Privacy

Astha Jain[1], Prof. Kathireshan V.[2]

[1]*Master Student, Department of MCA, Jain Deemed-to-be University, Bengaluru, Karnataka, India*
[2]*Associate Professor, Department of MCA, Jain Deemed-to-be University, Bengaluru, Karnataka, India*

---***---

Abstract – *Data Privacy is something that everyone worries about from an individual to a whole organization. Data Security is constantly a key topic of concern for the data scientists, organizations, individuals, etc. and it has now even become a more serious issue with the use of cloud computing in which the data is stored all over the globe anywhere without the owner's knowledge. Everyday many new techniques are being researched and developed to solve the issue of data privacy. It is becoming more unremarkably important as new data storage and processing techniques are being developed every day and data security and privacy is very important in respect to future advancement of these techniques. Data is a significant entity in today's world whether its individual, government, corporate, trade, industries, etc. There are various techniques and tools for safeguarding the data from in cooperation with hardware and software facets. In this paper, we will discuss about a new tool that safeguards our data by secluding it from all the other data in the system and hiding it in plain sight.*

***Key Words*: Security, Cryptography, Data Hiding, Authentication, Data Protection, Steganography, Data hiding**

1. INTRODUCTION

Data is a small term but the most important thing at the same time. Data privacy is very important in today's world and therefore many researchers are working day and night to develop new techniques to provide more advance and better security and privacy measures to secure our data as data is the fuel to our digital engines. Data is an important component in all digital aspects whether it's a simple word file, a basic software or a complex AI based Tool. Privacy of data is a major concern and needs to be solved. Data is produced every single minute in abundance whether it's important or not, it's still is data and can be manipulated as per need. Data starts producing as soon as we start our digital device such as mobile phones; when we call someone, when we browse the internet, when we watch videos, movies, etc., when we play games, when we chat, when we perform online transactions, etc. In every single click we are producing data which is used in different ways by service providers, large technology organizations, developers to provide us with better user experience all at the cost of our data[1].

2. INFORMATION SECURITY SYSTEMS

The information Security Systems also denoted as INFOSEC refers to mainly technologies and methodologies developed over time to secure data both physically and digitally from hackers, crackers, infiltrators, masquerades and deceivers. INFOSEC not only means protecting already stored data but it means protecting the data and information as it is produced like telephonic conversation [3]. The main aim of INFOSEC is to fulfil its basic principle that are:

a. Confidentiality: Measures taken to make sure that the information is only accessed by an authorized personal and that the information and data is safe and can only be accessed by those who have a right to access it like all data in an organization is not required by all its employees they just need the data that they work on.
b. Integrity: Measures taken to make sure that the data or information has not been manipulated or changed in the time of transit that is no unauthorized modification has taken place and that the data can be trusted and is accurate.
c. Availability: Measures taken to ensure that the support systems are always working and can be used whenever required that is, data is available at all times to its authorized users to process or modify it [2].

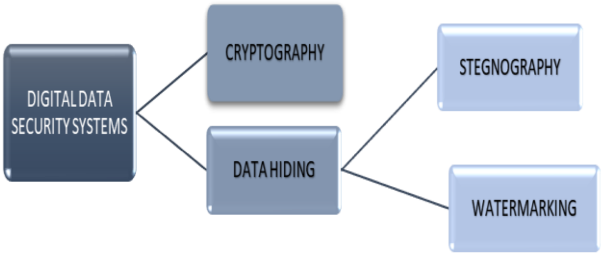

Fig 1: Digital Data Security System Classification

The information security systems are off many types like application security, cloud security, digital data security, infrastructure security, incident response and vulnerability[3]. In here we are discussing about digital data security which we are using in the proposed system. The digital data security systems are particularly divided into two categories that are Cryptography and Information Hiding. These techniques are quite different from each other as Cryptography refers to the encoding of the text or data in such a manner that it is hard to understand it without decoding it using proper algorithm and keys whereas Information Hiding simply refers to walloping the information in plain sight without the invader's knowledge[4].

3. CRYPTOGRAPHIC APPROACHES

Cryptography is one of the most famous and widely used techiques to safeguard the data. It is made of two Greek words that are "Kryptos" and "Graphein" where 'Kryptos' was converted to "Crypto" which means "hidden" and 'Graphein' was converted to "Graphy" which means "to write". So, it completely means hidden writing that is altering the way the data is existing already using some algorithms so it becomes difficult for the intruder to understand and derive a conclusion from it[5]. It is a way of ensuring that only the people who are intended to see, use and process the data can access it and others can't that is unauthorized users. It makes use of a set of codes and algorithms; the algorithms are mathematical concepts and rule-based concepts. These codes and algorithms together are used to convert the plain text into cipher text. The cipher text is the encoded form of plain text which is hard to understand and can only be decoded by a set of keys(codes) known to authorize users [6].

There are various cryptographic approaches developed over time that are mainly categorized into four that are Homomorphic encryption, secret sharing, Garbled Circuits and Secure Processors. But the most famous is Homomorphic encryption. These approaches are used for different purposed like producing keys for digital endorsement, digital signatures, for performing digital transactions using debit card/credit card/online banking, etc[7]. Cryptography is of three types:

3.1. Symmetric Key Cryptography: This makes use of only a single key for both encryption and decryption purpose. This key needs to be shared in a secure manner. This key is mostly used to maintain confidentiality of data. The most famous symmetric key algorithm is DES (Data Encryption Standard).

3.2. Asymmetric Key Cryptography: In this, two different keys are used; one for encryption and anther for decryption. For encryption a public key is used but for decryption a private key is used. RSA is the most popular Asymmetric key encryption technique. It is mostly use for non-repudiation and authentication purpose. In this even if the public key is known to everyone, knowledge of the private key is only held by the authorize user.

3.3. Hash Function: In this, a digital fingerprint is created of the encrypted data which cannot be reversed. This encrypted form is obtained using a mathematical function known as hash function. It does not make use of any key and the once the data is encrypted using this the contents can never be recovered completely.

Homomorphic encryption is what is used in this tool. This can be both symmetric and asymmetric. Homomorphic encryption refers to the encryption in which the plain data and the ending decoded data are completely same without any modifications or alterations. This can be put in an equation in the following way: (Where f(p) is plain text function and f(k) is key function) [6] according to eq. 1.

$$f(p * k) = f(p) * f(k) \ (1)$$

When a function is performed on the plain text and the key it results in cipher text which when decoded results in the original text. The function can be both multiplicative or additive. Subtraction and division cannot be used as functions as they lead to loss of data which is not a property of homomorphic encryption.

This tool uses two of the most famous encryption algorithms that are AES and RSA.

AES refers to Advanced encryption standard. This is a very famous symmetric type cryptography which means it makes use of only one key that is "public key" for both encryption and decryption. It makes use of block ciphers rather than stream cipher. AES was selected by NIST (National Institute of Standards and Technology) for safeguarding all type of top-secret information all around the world. The block ciphers used in AES can be of three levels which means key lengths that are AES-128, AES-192 and AES-256. The data goes through different rounds in AES encryption. The number of rounds in the encryption process are determined by the level of key like AES-128 has 10 rounds whereas AES-192 has 12 rounds.

RSA refers to Rivest, Shamir, Adleman. This is very famous asymmetric type cryptography which means that it makes use of two different keys, a public key for encryption and a private key for decryption. The encrypted message can only be decrypted using the private key which is known only by the receiver or authorized user and this makes RSA more secured. This behaviour of RSA is a result of distinct mathematical properties. RSA also uses block ciphers but in RSA the block size are not fixed that is the user can make use of variable length block ciphers as per need. This encryption algorithm was developed based on the idea that factoring of two large numbers is problematic but multiplication of the same is easy. It is used in all modern computers for everyday transactions like online transactions, password exchange, cable televisions, online video streaming, etc.

4. DATA HIDING

It is a technique of hiding the data in plain sight without anyone's knowledge so that the data is safe and secured and in a place no one can find easily. It is not something that can replace cryptography as it is different form cryptography but still serves the same purpose that is data security. It is a technique developed to enhance security that is by firstly encrypting the data and then embedding it into some other media or hide it in plain sight so that even if the intruder somehow gains the algorithm and key for decryption, they still need to locate the data which is hidden[8]. Data Hiding is the art of concealing the confidential or important data so that it is not accessible by the intruders and they cannot use it against us. In data hiding, the information or data is stored in some media like photos, videos, text documents, etc. who serves as a coverage for the original information leaving no trace of data existence and embedding; to human eyes the data is not there and therefore it becomes difficult to even locate it. When we embed the data into another media, they both become a single entity which means they are not connected by any link and thus, this reduce the chance of separation at the time of its transfer. Not only this, but by using this system we can ensure integrity as tamper detection and authentication can also be embedded along with the data to provide assurance and proof when there is a need for assurance that the data was not modified by an unauthorized user. It can be achieved using two different techniques that are Steganography and Watermarking.

4.1. Steganography: In this technique, the original text is embedded into some other media like image or video and that media is then transmitted or stored. The original text is hidden securely behind the cover medium. This word was derived from two Greek words "steganós" which was converted to stegno which means "concealed" and "graphia" which was converted to graphy which means "writing", hence concealed writing. For performing the stegnography we may require a stegno key [9]. This key is something additional we are embedding with the message like password [8]. So this can be equated as:

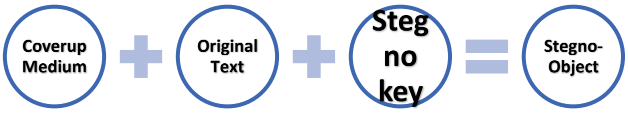

Fig 4.1: Steganography Equation

With the use of Steganography, we can embed any type of data in the cover medium. The resultant object that is stegno object does not look any different or work in any different way than the original cover medium. The stegno object is similar to the coverup medium even though it has embedded confidential data inside it. A stegnographic application requires at least the following elements [10]:

Coverup Medium (C): It is the media that is being used for embedding the secret data.

Secret Message(M): It is the secret data that needs to be stored or transmitted securely. This data can be encrypted or in plain text as well as it can be in any form video, audio, text or image.

Stegno File/Object(s): It is the object that is obtained when steganography is performed. This object looks and works similar to the coverup medium.

F_e: A steganographic function that is used to create the stegno object. This takes coverup medium and secret message as parameters and sometimes a stegno key as well.

F_e^{-1}: A steganographic function that is used to retrieve the secret message when it is accessed by an authorized user. This takes Stegno object/file as input and produces the secret message as output.

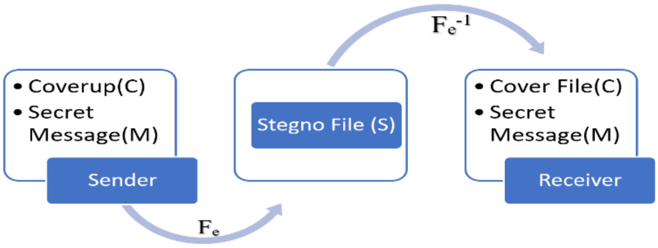

Fig 4.2: A Steganographic System

4.2. Watermarking:

It is similar to steganography as data is embedded into some media whether its video, audio, text or image but watermarking is done not with the sole purpose of information protection but also for copyright purpose. Watermarking is a famous technique used to hide secret information by embedding it into other media by using some mathematical algorithms which results in watermarked media. If this media is shared anywhere, the embedded information also shared with it but is of no use because the same algorithm only can retrieve the secret information/data embedded in it. Therefore, this technique is used for authentication purpose as well as to protect copyright. The data is embedded into the media in such a way that it is very difficult to eradicate the original data without harming it [11]. Watermarking is of two types that are visible and invisible. In visible watermarking, the data that is embedded into the media is visible. It is mainly use to show ownership like logo of a company behind its documents or logo of a broadcaster in the corner of the channel. In invisible watermarking, the data that is embedded into the media is of great importance and is therefore not visible to anyone. The user needs to use the algorithm that was used to embed the data into media to retrieve the data. It is used for transferring of confidential information and for authentication purposes [12].

The watermarking comprises of two different phases; One Embedding and Second Extracting.

The embedding operation requires a unique watermark that can be embedded into the media. Here, the watermark can be a logo of a company or any confidential information. This watermark is embedded into the media using an Embedding algorithm E. Thus, the algorithm E accepts Image I (for example) and Watermark W as its parameters and thus gives Watermarked Image I' as output[13]. This is shown by the following diagram (Fig 4.3.1).

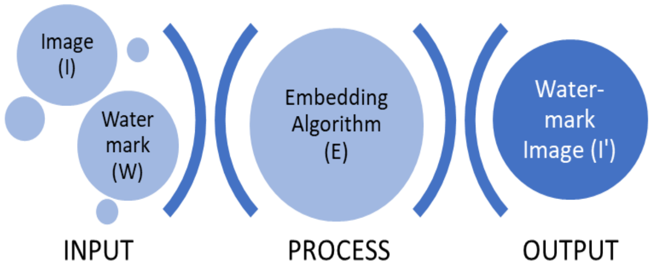

Fig 4.3.1: Embedding Operation

The Extracting Operation requires the watermarked image I' as input and sometimes the original media for authentication purpose, this depends on the technique used and the type of watermarking as it is mainly required in Visible watermarking. The Extracting algorithm E' takes Watermarked image I' as input and sometimes original image I and gives us the watermark W as output (Fig 4.3.2)[13].

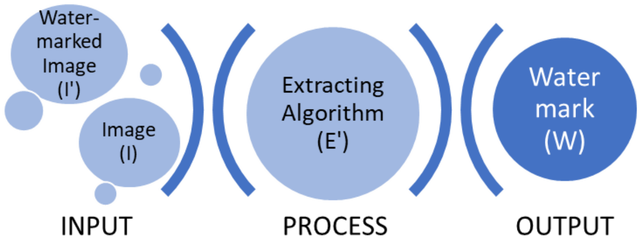

Fig 4.3.2: Extracting Operation

5. PROPOSED SYSTEM

A tool created for providing data privacy to individual users on their systems; these users can belong to an organization, can be freelancers, can be researchers, etc. In this, the tool creates a folder that is hidden in plain sight and can only be accessed by running the tool that will authenticate the operator who is trying to gain access to that folder. Once the operator gets access to the folder, the folder contents itself are encrypted using different cryptographic approaches. The user of the tool decides which security level he/she requires and according to the requirement the security is decided and implemented. The user can choose from two different homomorphic encryption algorithms and a secret sharing algorithm or a combination of them.

6. PROCEDURE

The tool works in the following way:

- Firstly, the tool is been executed and at runtime it asks for a password to be provided that will be used to access it and hence a folder gets created.
- Then, a folder appears on the desired location. We can put files or more folders in this folder.
- When moving a file in this folder, the user gets an option to encrypt the contents of the file.
- The user can choose from 2 different cryptographic approaches to secure there data or the user can choose a combination of them.
- When user selects an approach, he/she needs to enter different details required to implement that security level.
  - ➢ Once you choose the security level, it let you choose the key also as per your need in RSA and AES.
  - ➢ AES makes use of 128 bit key which means 10 rounds for encryption and 10 rounds for decryption process.
  - ➢ RSA makes use of 1024 bit key for encryption and decryption.
  - ➢ The encrypted and decrypted data both are stored in different files that are created during the process.
  - ➢ Combining both AES and RSA results in high level security.
- Once the security procedure completes the user can exit the folder.
- The user needs to re-execute the tool and put the password so that the folder gets concealed.
- This folder can be accessed when the used runs the program again and puts the password.

7. CONCLUSION

In this paper we discussed about data privacy and hiding. Data is an important component in the digital world and needs the highest level of security that can be provided to it. With time many different techniques came into existence to improve data security and still researches are going on. A simple tool is created using those existing techniques to provide a simple and effective way for individual users to protect their personal data. The tool simply creates a folder that is hidden and secluded in plain sight and provides data security using cryptographic approaches. This tool can be used by individuals to store there files in the system hiding it from others so that the user can stores its personal and confidential files in it.

8. REFERENCES

[1]. Nils Gruschka, Vasileios Mavroeidis, Kamer Vishi , Meiko Jensen ,"Privacy Issues and Data Protection in Big Data: A Case Study Analysis under GDPR", University of Oslo, Norway, 20 Nov 2018

[2]. Jaap Wieringaa , P.K. Kannanb , Xiao Mac , Thomas Reuttererd , Hans Risseladaa , Bernd Skiera "Data analytics in a privacy-concerned world" 3 May 2019

[3]. Prachi Kohale, Sheetal Girase "Privacy Preservation of Data in Data Mining" June 2014

[4]. Astha Jain, Ishita Popli, "Data Privacy Using Cryptographic Approaches", May 2020

[5]. B.B. Zaidan, A.A. Zaidan, A.K. Al-Frajat and H.A. Jalab ,"On the Differences between Hiding Information and Cryptography Techniques: An Overview", June 2010

[6]. Astha Jain, Ishita Popli, Dr. C. Umarani. "PML: COMPARATIVE STUDY ON CRYPTOGRAPHIC APPROACHES IN MACHINE LEARNING" 2020

[7]. Varsha, Dr. Rajender Singh Chhillar, "Data Hiding Using Steganography and Cryptography", April 2015

[8]. Lisa M. Marvel, "Information Hiding: Steganography and Watermarking", 2002

[9]. Arvind Kumar Km. Pooja , "Steganography- A Data Hiding Technique", November 2010

[10]. Minati Mishra, Priyadarsini Mishra and Flt. Lt. Dr. M.C. Adhikary. "DIGITAL IMAGE DATA HIDING TECHNIQUES: A COMPARATIVE STUDY", 2012

[11]. Sonam Tyagi; Harsh Vikram Singh; Raghav Agarwal; Sandeep Kumar Gangwar "Digital watermarking techniques for security applications", 2016

[12]. Michael Agbaje, Awodele Oludele, Chibueze Ogbonna "Applications of Digital Watermarking to Cyber Security (Cyber Watermarking)", January 2015

[13]. Jobin Abraham, "Digital Image Watermarking: An Overview", February 2011