A Survey on Cryptography, Encryption and Compression Techniques.

Shiv Suraj Oberoi[1], Yashika Varyani[1], Dr. Deepak Chahal[2]

[1]MCA Student, Department of IT, Jagan Institute of Management Studies, Sector -05, Rohini, New Delhi. India
[2]Professor , Department of IT, Jagan Institute of Management Studies, Sector -05, Rohini, New Delhi. India

***

Abstract: Data is any type of information that can be stored digitally. Security of data is an aspect that comes into play when the data which is digitally stored is to be protected. Data security is one of the major factors of this century, it refers to the protective measures taken to prevent unauthorized access to the data stored in personal computers, databases, and online websites or cloud services. Cryptography is a concept which helps the user to protect their data by giving them the necessary tools and algorithms for encryption and decryption of data.

Cryptography is a very popular way of sending/receiving data in a very secretive way. Compression can be defined as reducing the number of bytes required to represent a set of data. Various techniques are used to reduce the size of data depending on the requirement of the user.

Keywords: Data Encryption and decryption, Compression, Cryptography Concept, Security, Integrity.

1. Introduction

Data can be defined as a collection of characters that is organized to serve a purpose. It can be anything ranging from text, number to images, videos. Raw data can be defined as data in its most basic form. Raw Data becomes "information" when it is processed and converted into a form which is convenient for movement and processing. In computing terms, Data can be described as a collection of 1's and 0's which are called "bits". A bit is the smallest unit of data and a collection of eight bits is called a "Byte". Database can be defined as an organized collection of data and Database Management System where created to organize and manipulate the information stored on it our main problem is the lack of standardization regarding procedures and techniques. Coming out of education and moving into the industry you can find yourself with knowledge of various methods and approaches, but no clear guide on best practices [1].

Cryptography is the practise of generating ciphers which are used to encrypt a message to ensure secure communication between the sender and the receiver. The encrypted message can only be decrypted by the person for whom the information was intended. Cryptography comprises of two very basic steps: Encryption and Decryption. Encryption is performed at the sender's side and Decryption is done at the receiver's side.

Encryption is the process of converting information into non-human readable format using an algorithm. It is the process of applying cryptography. It is the process through which plaintext can be converted into ciphertext where the plaintext acts an input to the encryption process and the ciphertext acts as an output to the encryption process.

Decryption is a converse process of encryption; it decrypts the encoded data into meaningful information. It is the process through which ciphertext is converted back to plaintext where the ciphertext acts an input to the decryption process and the plaintext acts as an output to the decryption process.

Data compression is a technique through which the size of the data set is reduced by removing excessive information and modifying the bit structure of data in such a way that it acquires less space on the disk. Once data compression is done, file maybe fully recovered without any loss of actual information. Compressed data files can be easily uploaded and downloaded from the internet thus making the transfer of large sized files easier.

Data compression techniques can broadly be divided into two categories: Lossy techniques and Loss-less techniques. In Lossy data compression, some part of the data is removed while compressing it but in Loss-less datacompression the same is achieved without the loss of data.

In short, Data compression is the technique of encoding data to lesser number of bits than the original representation so that it requires less storage space and can be easily shared between networks. A compression tool is used to compress the data from and easy-to-use format to one optimized of compactness.

Similarly, a decompression tool is used to restore compressed data to its original form. Data decompression is needed nearly in all cases of compressed data including, lossy and lossless compression.

2. Cryptography

Cryptography is usually referred to as "the study of secret". It is the transformation of understandable and readable information into a form which cannot be understood in-order to prevent data leakswith the helpof a key and an algorithm. Cipher is the algorithm which is used to convert plaintext to ciphertext, this method is called "encryption". Cryptograph covers three core areas that protect us and our data from theft, unauthorized use and possible fraud, these functions are usually referred as the goals of the security system. These goals can be listed under the following categories: Integrity, Authentication, Confidentiality.

2.1 Integrity

Integrity of data refers to protecting information from falsely being modified by an unauthorized party. Information is valuable only if it is correct, tampered information could prove costly to both the sender and the receiver party. When sensitive data is being exchanged between the sender and receiver, the receiver must have the assurance that the message received has not been tampered with and has been sent by the Sender itself. There primarily two of threats in data integrity: Passive and Active.

2.1.1 Passive Threat:

This type of threat happens when unintentional changes happen in the data. The possibility of such errors increases when there is noise in the communication channel. Error correcting codes and methods like Cyclic Redundancy Checks are used to detect loss of data integrity. The victim remains unaware of the attack. In passive attacks, attackers majorly observe the transmission and tries to make use of the information from the system but does not affect system resources.

2.1.2 Active Threats: In this type of threat, the attacker attempts to alter system resources or effect their operations. Active attacks involve some level of modification to the data stream or tries to falsely tamper with the data.The victim gets informed about the attack, such attacks always cause some level of damage to the system.

2.2 Confidentiality

The main aim of confidentiality is to ensure that the information is not disclosed to unauthorized parties. It is used to make sure that nobody between sender and the receiver is able to what data or information is being transmitted. This can only be achieved through Encryption. There are two types of encryption algorithms, Symmetric and Asymmetric. Symmetric algorithms allow encrypting and decrypting data with the same single key. Asymmetric algorithms have two kinds of keys: Public and a Private key. The public key is commonly available to the public and while the private key is just available for each specific user. Everything that is encrypted using the public key can only be decrypted using the private key and vice versa.

2.3 Authentication

Authentication is the process of proving one's identity or it can be defined as the process of determining someone is in fact who he claims to be. It is necessary for the communication participants to prove their identities as what they have claimed using some techniques so as to ensure the authenticity [2]. This means that before sending data or receiving data the identity of the sender and the receiver must be verified.The authentication process beings at the start of the application, the incoming request is associated with a set of identifying credentials. The credentials are often in the form of a password, which is a secret and is known only to specific individuals and the system. The process of authentication can be distinctively be divided into two phases: identification and actual authentication.Identification phase provides a user an identity on the security system in the form of an ID. Identification and authentication are very similar the only difference between them is, in identification the user claims an identity by providing a usernameand in authentication the user proves his identity by providing a password.

3. Encryption

Encryption is a process in which an algorithm is used to convert information into an unreadable format so that it becomes useless for unauthorized parties. In an encryption process, the information which is referred as the "plain-text" is encrypted using and encryption algorithm and a key, the encryption generates a "cipher text" from the information which can only be read if it is decrypted. This method protects useful information about the user like credit card details and passwords by encoding them and converting them into a cipher text. Encryption is essential for safe and trusted exchange of sensitive data. The process of building applications has been a journey and it varies depending on one's application requirements and purpose [3].

3.1 Types

Encryption algorithms can be categorized into two categories: Symmetric encryption and Asymmetric encryption.

Symmetric encryption algorithm, is also known as "Private-key cryptography". In this process there is only one key which is used for both encryption and decryption. Both the sender and the receiver each have a copy of the key before the process of encryption and decryption. The secret key could be a password/code or a random string. At the sender side the plaintext is encrypted and converted to ciphertext using the "key" and the reverse of this process is done at the receiver's side where the ciphertext is decrypted and converted back to plaintext using the "key" again. There are two types of Symmetric encryption algorithm:*Block* *algorithms and Stream algorithms*. In *Block algorithms*, definite lengths of bits are encrypted in blocks of data with use of a secret key. As the data is being encrypted, the system holds the data in its memory. Examples of Block algorithms are: Advanced Encryption standard (AES), Data encryption standard (DES), International data encryption algorithm(IDEA), BLOWFISH, Rivest Cipher 5(RC5) and Rivest Cipher 6 (RC6).

In *Stream algorithms*, data is encrypted directly as it streams instead of being retained in the system's memory. Example of Stream algorithms is: Rivest Cipher 4 (RC4).

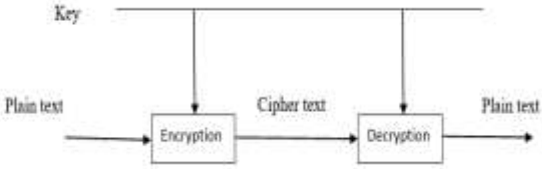

Fig 1. Symmetric Cryptosystem

Asymmetric encryption algorithm, is also known as "Public-key cryptography", it uses two types of keys: Public key and Private key. Public key is used to encrypt the data and generate a ciphertext. The ciphertext is decrypted by the receiver whenever it receives the encrypted data by using it its own Private Key. Private key is a secret key and is only known to the person who the data is intended for and is unknown to all. Public keys are stored in a database for anyone to see. There are two types of Symmetric encryption algorithm: *RSA algorithms and Digital signatures*.

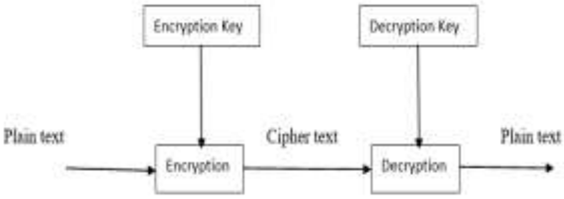

Fig 2 Asymmetric cryptosystem

4. Data Compression

Data compression is a technique through which a file (Text, audio or video) can be compressed, i.e. the size of the file is reduced by removing the excessive information and redundancy. Data compression is only considered successful if the original file can be retrieved back with any loss of information.

Data compression reduces file size considerably thus saving storage space and making the transfer of such files over the internet much easier as compressed files can be uploaded or downloaded from the internet much faster. Data compression has important in areas such as File storage and Distributed system. Shorted data sequences are more suitable for sending/ receiving because it reduces the overall cost of transmission. Data compression is particularly useful in communications because it enables devices to transmit or store the same amount of data in fewer bits. Data compression has two components: *Encoding Algorithm* and *Decoding Algorithm*. The Encoding Algorithm takes a message and generates a compressed version of it and conversely the Decoding algorithm reconstruct the original message from the compressed version. A focus is made on machines as machines cannot be understood by verbal communication it forms abstractions and concepts [4].

4.1 Compression and Decompression

Data Compression techniques can be categorized into two categories: *Lossy Technique* and *Lossless Technique*. The algorithms which removes some part of data during compression is called "Lossy Data compression" and the algorithm which does not lose data while compressing and is able to retrieve the same file after decompression is called "Lossless Data compression".

4.2 Types of Data Compression

In some applications loss of some amount of data is acceptable, in such cases *Lossy Compression* technique is used. The loss of data may be in the form of colour depth or graphic detail. The best example for Lossy compression technique is video conferencing. In video conferencing considerable amount of data is lost in transmission in order to deliver the image in real time. Lossy compression targets redundant pixel information and discards it. Lossy compression is majorly used with media elements which can still work without all their original data such as images, videos, audios and details graphics which are used in screen designing.Lossy Compression is not used for documents and software because they need all of their data to be intact during transmission.

Original Text Data
Compressed Data
Restored Data

Fig 3 Lossy Compression

In *Lossless compression* the data is compressed without losing any amount of data. It reduces the size of file without degrading the image quality. When the file is decompressed, original data is retrieved. The data of the file is only temporarily thrown away so that the transmission of the file becomes easier. This type of compression can be applied to not only just graphics but to spreadsheets, documents and software application as well because no amount of data will be lost in transmission. While the advantage of this technique is that the quality is maintained but the disadvantage of this method is that it doesn't reduce the size of the file considerable.

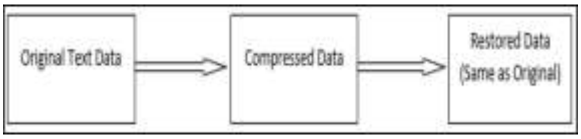

Fig 4 Lossless Compression

5. Conclusions

This paper focuses on what data is, different encryption and decryption techniques, Data compression and various compression techniques.

Cryptography is used to encrypt data so that when the data is sent over a network, so that it can be securely transmitted and the contents of that message cannot be altered with.

The main aim of data compression techniques is to reduce the size of a file stored on the disk so that even large sized files can be easily transferred on through a digital network.

6. References

[1] Chahal D. et al. Data Science Applications, Challenges Related Future Technology, International Journal of Trend in Scientific Research and Development, Volume 3 Issue 1, Nov-Dec 2018

[2] Chahal D. et al. Security Concepts Underlying MANET, International Journal of Emerging Technologies in Engineering Research (IJETER), Volume 5 Issue 3, March (2017).

[3] Kharb, L. (2018, January). A Perspective View on Commercialization of Cognitive Computing. In 2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 829-832). IEEE

[4] L. Kharb et al (2019) "Brain Emulation Machine Model for Communication" in International Journal of Scientific & Technology Research (IJSTR). pp 1410-1418