Chapter 1
Introduction

1.1. Overview

An information system is protected from unauthorized access, modification and destruction by the help of information security. There are three prime concepts for information security which are integrity, confidentiality and availability. Integrity deals with the assurance of that the provided information is trustworthy as well as accurate. Confidentiality refers to the restriction of access or privacy of data though it is not same as privacy. Availability ensures the access of infromation by authorized user depending upon their needs.

Cryptography is the process of transmitting information by using other forms of original information for secure communication in the presence of adversaries. Private Key cryptography applies the same key for encryption and decryption. Encrypted text and private key both are shared between sender and receiver. The keys used in receiver's and sender's ends may be identical or easy to be derived from one to another.

Authentication is the process to validate and ensure the identity of the user. Different user authentication processes are implemented in the last two decades. CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) and OTP (One Time Password) are widely used for user authentication. CAPTCHA is a visual test on the basis of challenge and response to determine whether the user is an object or human. OTP authenticates the user for a single session or transaction. Generally, the server automatically generates the OTP which is a numeric or alphanumeric string of characters.

Secret sharing scheme is applied for the color image where the color image is divided into N (where N is a positive integer) number of shares. The image is encrypted using N number of shares and decryption is carried out by secretly selecting K number of shares out of N number of shares. Steganography is the procedure for hiding of the information into other cover objects. Original information is hidden into another cover carrier like image, text, video, and audio.

In the last two decades, a considerable amount of research has been carried out in information hiding, cryptography, authentication and steganography domain. An attempt has been made to develop a security toolkit, where newly implemented independent algorithms are combined together in a cascaded manner from the domain of cryptography, authentication, and steganography.

1.2. CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart)

CAPTCHA [8][17] is normally an abbreviation. CAPTCHA stands for "Completely Automated Public Turing-test to tell Computers and Humans Apart". It is normally a test based on challenge and response to determining whether a user is human or not. Robots, automated programs are restricted by CAPTCHA from unauthorized access, cracking passwords and sign-up E-mail accounts. The human can crack CAPTCHA very easily, which is some visual test or puzzle where a robot or automated program will not be able to crack it. Thus the unauthorized access is blocked from the robots or automated program.

Different forms of CAPTCHA [27][66] tests are available. The mostly used CAPTCHA test represents the codes in the form of letters, images, numbers, and alphabets in an overlapped or intersected manner. To gain the authorized access of the system, the users must have to understand and rewrite code in a specified pattern with proper sequence mentioned in the CAPTCHA code. Human being easily understand and write the

CAPTCHA code where it is impossible for robots or automated program to understand the code. Thus the unauthorized access is restricted from various automated programs, robots to crack passwords, sign-up E-mail accounts, privacy violation, and spam sending. Figure 1.1 represents a CAPTCHA, where the letters are twisted and the background color is applied behind the text.

Figure 1.1. Example of CAPTCHA Code

Different types of CAPTCHAs [112] are present. Some of them are discussed in the next section.

1.2.1. Text CAPTCHA: This type of CAPTCHA is widely used for authentication. Letters and/or characters are represented in a twisted and overlapped manner in text CAPTCHA. A user has to understand and input the text as it is displayed in the CAPTCHA.

1.2.2. Audio CAPTCHA: Text and numbers are represented as an audio clip where the audio clip is distorted by adding some noise and supplied to the user. A user has to input the proper content in text format to pass the security test.

1.2.3. Image CAPTCHA: A user has to select single or a group of images from a provided list of images depending upon the criteria specified by the CAPTCHA program. Image CAPTCHA is normally easier as basic images are used for identification.

1.2.4. Mathematical CAPTCHA: The basic mathematical problem is provided along with the operator and inputs. A user has to solve the problem and input the correct result for passing the security test.

1.2.5. 3D CAPTCHA: This type of CAPTCHA contains 3D images, which are incorporated with both words and images. As this type of CAPTCHA is very hard to solve, thus it is treated as 'Super CAPTCHA'.

1.2.6. Ad Injected CAPTCHA: Advertisement of objects containing brand name, logo are provided as CAPTCHA. A user has to input the brand name to pass the security test.

1.2.7. Drag and Drop CAPTCHA: The user has to shape or drag the mentioned object into its proper position. This type of CAPTCHA is normally very easy to solve and it is implemented by jQuery.

1.3. OTP (One Time Password)

One Time Password (OTP) [18][40][45][118] is a type of password which contains alphanumeric or numeric character string. OTP is generated automatically by a server. OTP is valid for a single transaction or login session for a digital device. OTP provides an additional layer of security which is considered as two factor authentication for user authentication but a device and distribution channel are needed to share the OTP between server and user. Figure 1.2 represents an OTP based authentication scheme.

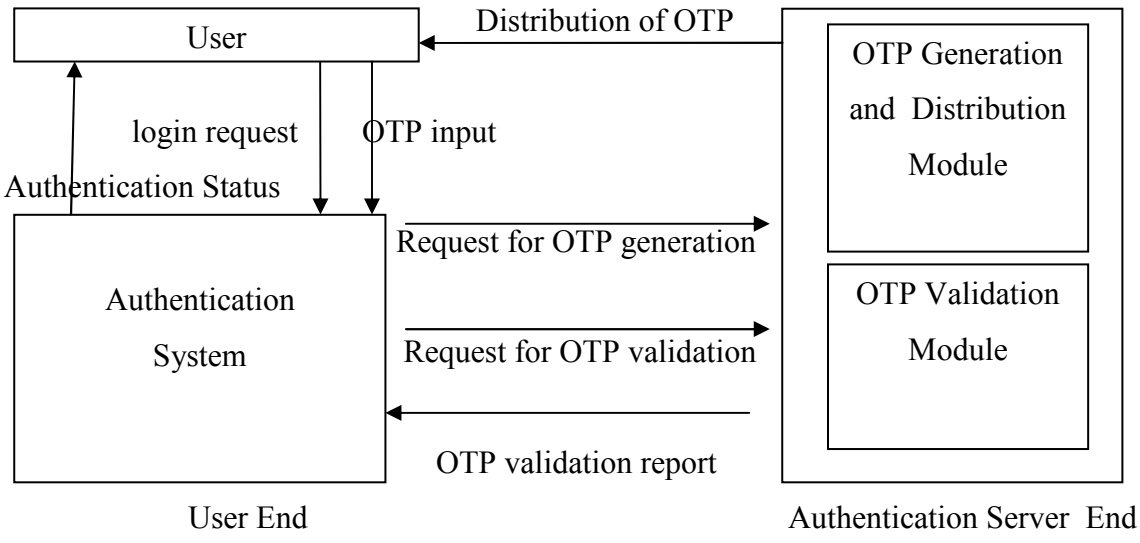

Figure 1.2: OTP based Authentication Scheme

OTP overcomes different shortcomings related to traditional authentication system based on a static password. Reply attacks are protected by OTP as OTP is valid for only one login session or transaction. Thus if an OTP is recorded by an intruder, it does not work as it has already been used for previous session or transaction. OTP provides security for those users who use the same password for login in multiple systems. If the password has been compromised still the system is secured as OTP is needed for authentication.

OTP generation is based on randomness. Different approaches of OTP generation are available like time synchronization between client and server, using mathematical algorithms, algorithms to generate new OTP based on previous OTP, challenge response based OTP implementation, using of a hash function and so on.

OTP is distributed between an authentication server and a user through different modes of communication. Approaches used for OTP delivery are text messaging, web based OTP distribution, delivery based on proprietary tokens and distribution of OTP using hardcopy.

OTP provides more security compared to a static password but still, OTP is vulnerable to man-in-middle attacks and social engineering attacks. New concepts have to be implemented for protecting those mentioned attacks.

1.4. Cryptography

Cryptography is the process of transforming information into other forms to hide it from unauthorized access. An original message and coded message are known as plaintext and ciphertext respectively. Encryption is the process where plaintext is converted into ciphertext. Restoring of plain text from the ciphertext is known as decryption. The entire process is called Cryptography [10][22][39][50][77][90][92]. Figure 1.3 represents a cryptography scheme.

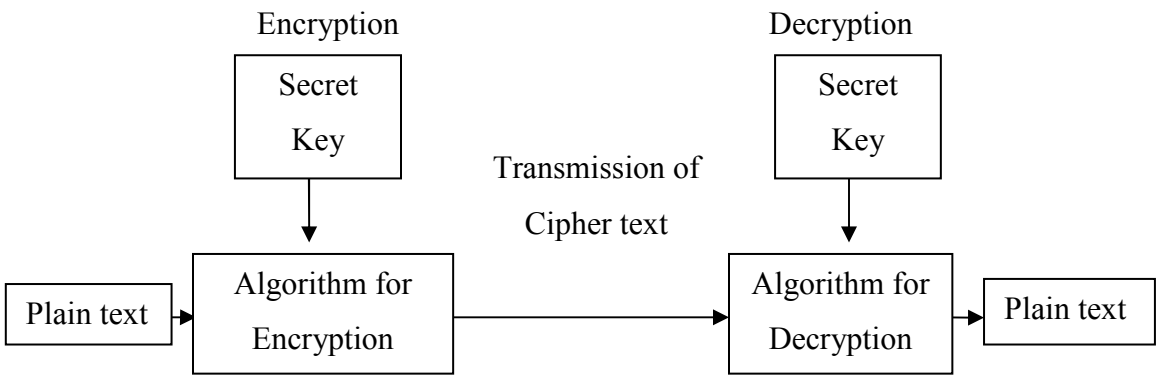

Figure 1.3: Cryptography Scheme

Plaintext: This is the original message that has to be sent to the receiver and it has to be supplied into the algorithm as input.

Encryption Algorithm: It is an algorithm to convert plain text into cipher text by using some key.

Secret Key: The secret key is an independent value and it is inputted by the user to the encryption algorithm. The key value is applied on plaintext to generate the ciphertext. Different output (ciphertext) depending on the specific key is being generated by the Encryption algorithm. The exact substitution and transformation performed by the algorithm depend on the key.

Ciphertext: This is an encrypted message generated from the encryption algorithm as output by using the key. It depends on the plaintext and the secret key. For a specific plaintext, different keys will produce different ciphertext. The ciphertext is an apparently random stream of data and, as it stands, is unintelligible.

Decryption Algorithm: The decryption algorithm runs in reverse of the encryption algorithm. It takes the ciphertext and secret key as inputs and produces the original plaintext as output.

Basically, there are three types of cryptography. They are private key cryptography, public key cryptography, and hash functions.

1.4.1. Private Key Cryptography

A single secret key is used by both the sender and receiver for encryption and decryption. The sender uses the private key for encryption and shares the ciphertext and the private key to the receiver. The receiver uses the same private key for decryption and generates the plain text. The Private Key Encryption [78] procedure is represented in figure 1.4.

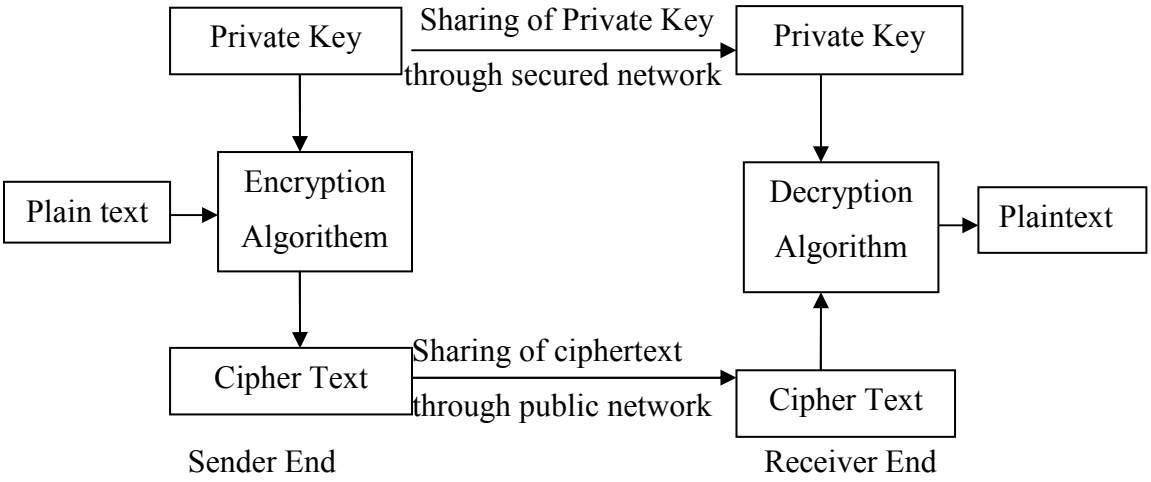

Figure 1.4: Private Key Encryption Process

1.4.2. Public Key Cryptography

Public Key Cryptography is carried out by applying two related keys (public and private key). The public key may be freely distributed while its paired private key remains secret and only being shared between sender and receiver. The public key is applied for encryption while the private key is used for decryption [107]. The Public Key Encryption procedure is represented in figure 1.5.

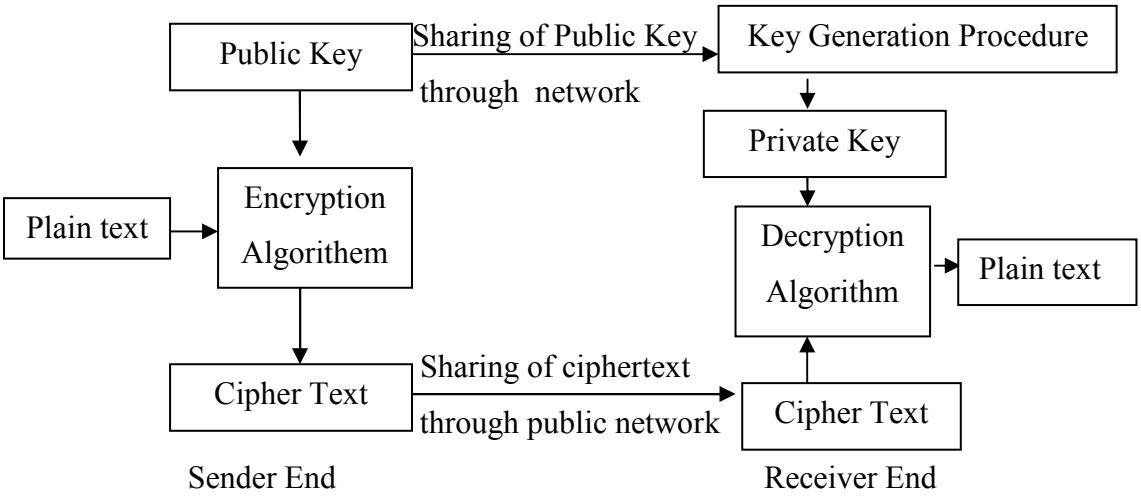

Figure 1.5: Public Key Encryption Process

1.4.3. Hash Functions

Hash Function [122] does not use any key for encryption. Depending upon the plain text, a fixed length hash value is computed, which makes it impossible for the contents of the plain text to be recovered. Normally many operating systems use Hash Functions to encrypt passwords.

1.5. Different Number Types

Different types of special numbers which are used for this current research are mentioned here.

1.5.1. Prime Numbers

A prime number is a number having only factors are itself and 1 with a value greater than 1. By multiplying two smaller natural numbers, a prime number cannot be generated. For example, 3 is a prime number as 3 is greater than 1 and only factors are 1 and itself

(i.e. 3). Similarly, 7 is also a prime number as 7 is greater than 1 and having only two factors 1 and 7 itself.

1.5.2. Palindrome Number

A palindrome number is a number that is read same after its digits are reversed. If a number is even the same both for reversed order as well as the original order is called a palindrome number. As for example 121 is a palindrome number as 121 is read as same both from its reverse order as well as its original order. Similarly, 202 is also palindrome as the reverse of that number is 202 which is the same with the original number 202.

1.5.3. Armstrong Number

An Armstrong number is a number having N number of digits (where N is a positive integer) such that the sum of digits of the number raised to the power N is equal to the number itself. For example 153 is an Armstrong number as 153 having 3 numbers of digits so N=3 and $1^3 + 5^3 + 3^3 = 1 + 125 + 27 = 153$. Similarly 370 is also an Armstrong number as N=3 and $3^3+7^3+0^3= 27+343+0 = 370$.

1.5.4. Amicable Number

A pair of numbers are considered as amicable numbers or amicable pair if the sum of all proper divisors (excluding the number itself) of the first number of the pairs equals with the second number of the pair and vice versa. For example, 1184 and 1210 are amicable pair or amicable numbers. The sum of proper divisor of 1184 excluding 1184 is 1+2+4+8+16+32+37+74+148+296+592=1210. Sum of proper divisor of 1210 excluding 1210 is 1+2+5+10+11+22+55+110+121+242+605=1184. So 1184 and 1210 are amicable numbers.

1.5.5. Perfect Number

A positive integer is considered as a perfect number if the sum of that number's proper positive divisors (excluding the number itself) is equal with that number. Example of the perfect number is 28, because 1, 2, 4, 7, 14 are its proper positive divisors, and 1 + 2 + 4 + 7 + 14 = 28.

1.6. Color Management and Pixel Structure of Digital Image

Brief descriptions of different color models and structure of the pixel of color image are described in this current section.

1.6.1. Color Models

Additive and subtractive color models are widely being used for describing the constitutions of color for the digital image. There are three primary colors red, green and blue (RGB) in the additive model and any desired colors being obtained by mixing different RGB channels. The user can modulate the amount of red, green, blue by controlling the intensity of red, green, blue in the compound light. The brightness of light will be increased as more the colors are mixed. When all red, green, blue will be mixed in extreme intensity, it generates black and in equal intensity, it generates white. The additive color model [28] is also being used for computer display unit. Figure 1.6 represents the RGB color model.

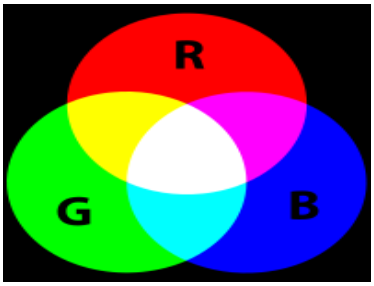

Figure 1.6: RGB Color Model

The subtractive model represents the color by applying the combination of colored-lights reflected from the surface of an object. By mixing cyan (C) with magenta (M) and yellow (Y) pigments, the user can produce a wide range of colors. The more the pigment is added, the lower is the intensity of light, and thus the darker is the light. This is why it is called the subtractive model. C, M, and Y are the three primitive colors of pigment, which cannot be composed of other colors. Figure 1.7 represents the CMY color model.

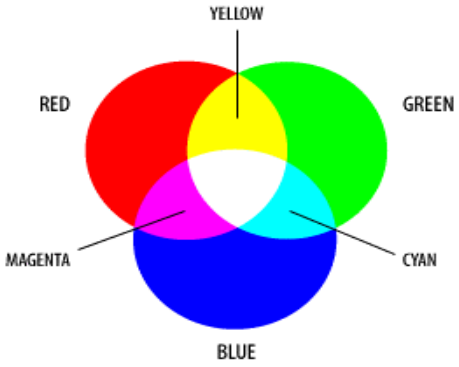

Figure 1.7: CMY Color Model

1.6.2. Pixel Structure

A natural color digital image is composed of a finite number of elements called pixels [49][80]. In a 32 bit digital image each pixel consists of 32 bits, which includes four parts, namely Alpha, Red, Green and Blue; each with 8 bits. Alpha part represents the degree of transparency. If all bits of the Alpha part is '0', then the image is fully transparent where Red, Green and Blue parts represent the density of the corresponding colors. A 32-bit sample pixel is represented in Figure 1.8.

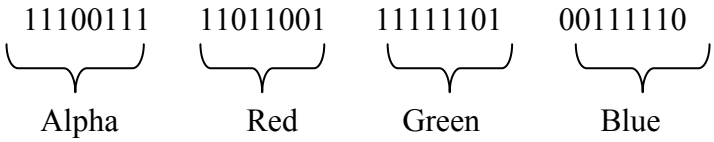

Figure 1.8: Structure of 32-bit Pixel

1.7. Digital Watermarking

Digital watermarking [16][75][79][120] is the procedure for embedding data into other digital multimedia content like text, video, image, audio. The signal of cover media is modified by using a key and thus generates a composite signal. Nature of the watermarking process defines the existence of information present in cover media. Figure 1.9 represents the idea of the basic digital watermarking process.

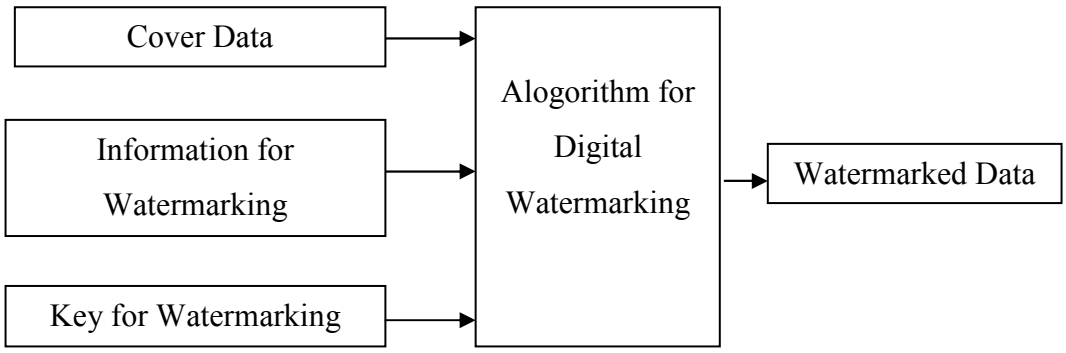

Figure 1.9: Basic Digital Watermarking Techniques

Digital watermarking is used generally for copyright protection. Beside this, it is also applied for other purposes like tracking of the source of digital content, carrying out hidden communications, broadcast tracking for watermarked videos and so on.

Various digital watermarking algorithms are implemented mainly from two domains namely spatial and frequency domain. Techniques based on Least Significant Bit, Additive Watermarking, Patchwork Algorithm, Texture Mapping coding were developed for the spatial domain where frequency domain includes Discrete Wavelet Transforms, Discrete Fourier Transform, and Discrete Cosine Transform techniques.

Digital watermarking can be grouped into visible watermarking and invisible watermarking.

1.7.1. Visible Watermarking

In the visible digital watermarking image, the semi-transparent text is embedded with the host object. The process allows viewing the original image, where the watermark is perceptible to the human eye intentionally. Robustness of visible watermarks is more comparing to image transformation. It is preferable to use visible watermarking for copyright protection of intellectual property which is represented in digital format.

1.7.2. Invisible Watermarking

In the invisible watermarking, the embedded data cannot be perceived using human's eyes. The hidden information may be detected or extracted by using electronic devices or specialized computer software to identify the copyright owner. Invisible watermarking [85] consists of encoding and decoding processes. Specialized digital content like images, text and audio content is been marked using invisible watermarking to prove its authenticity.

1.8. Steganography

Steganography [25][59][87] is the process of hiding and transmitting of confidential information within any carrier media in such a manner that the existence of the data is concealed. The main objective is to hide the secret information into a cover media in such a manner that its presence cannot be discerned. Digital multimedia files, network packet are considered as cover media. Figure 1.10 signifies the view of basic steganographic scheme.

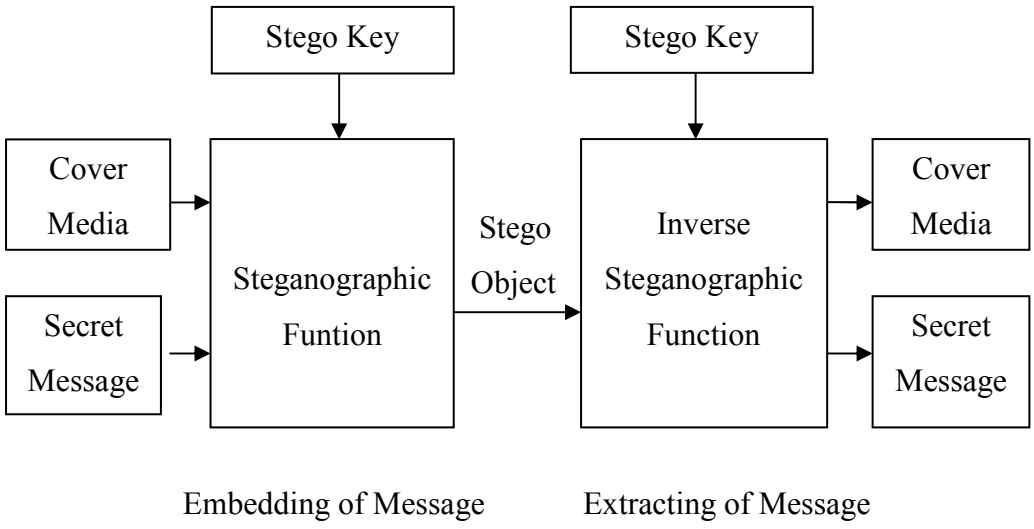

Figure 1.10: Basic Steganographic Scheme

1.8.1. Applications and Comparison with Cryptography and Watermarking

As the output of the steganography is invisible, so steganalysis is being applied to detect the presence of steganography. Steganography is used for intelligence services, modern printing technology, and military services. Cryptography hides the message content by changing its form but not being able to hide the existence of message content where steganography hides the entire message content into the cover media in such a manner that the hacker is not able to trace its existence in cover media. Watermarking hides the content of the message in such a way that an intruder cannot be able to tamper the message embedded in carrier media.

1.8.2. Steganography Approaches on Data Set

Steganography approaches depend upon the nature of the carrier media and the following media namely image, plaintext, IP datagram, audio, and video are being used as the cover objects. In plain text steganography, different techniques like applying of selected

characters, inserting of extra blank spaces are used. Discrete Cosine Transform, Fast Fourier Transform, and other techniques are applied for image steganography. For audio and video steganography, LSB Coding, Spread Spectrum, Phase Coding, Echo Hiding, and other approaches are applied.

1.8.3. Steganography Attacks

Different types of steganography attacks like known carrier attack (cover and stego media are available), known steganography attack (cover, stego media, and algorithm are available) known message attack (availability of hidden message) and steganography only attack (stego media is available) are applied by the hackers for unauthorized access of data.

1.9. Problem Domain

The focus is to build a 'security toolkit' consisting of 'building blocks' (i.e. independent algorithms) from the domain of user authentication, text encryption, and image encryption. The efforts have been made to design and implementation of user authentication systems, which provide better securities over different attacks and block unauthorized access compared to standard authentication schemes. As the distribution of keys over public communication channels is very unsafe from unauthorized access, so the focus has been given to implement some text or image encryption system that resolve this problem and provides a scheme with better security compared to existing encryption schemes in respect of standard parameters. An attempt has been made to implement new image encryption schemes, that provide better security in respect of image encryption and to compare experimental results with standard algorithms to measure the performance of implemented schemes. Focus has also been imposed to define the parameters, methodologies and standard algorithms to measure and compare the performance of newly developed schemes. Finally, attempt has been made to design and implementation of a 'security toolkit' by incorporating all the algorithms together relating to user authentication, text encryption, and image encryption in a cascading manner.

Performance of this newly developed toolkit has been examined by comparing it with other standard algorithms to validate its acceptability.

1.10. Thesis Objective

The basic objective of the work is to design an integrated ciphering system or a so called, 'security toolkit'. Such a system has been formed by combining a set of newly developed independent ciphering protocols which has acted as building blocks related to user authentication, text encryption, and image encryption. Attempt has been made to design and Implementation of user authentication building blocks based on CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) code and OTP (One Time Password) where focus has been imposed on randomness and how they provide security over different network attacks and so on as standard parameters for assessing their performances. In text encryption building block, as the distribution of the private-key without interpretation is very hard to achieve, so an attempt has been made to design a predefined secret procedure to retrieve the secret value from the private key and use it for encryption. Performance of these protocols has been examined in respect of different parameters like execution time, Chi-Square value and degree of freedom. New Steganographic scheme based image encryption building blocks has been introduced where encryption ratio, Noise ratio, structural similarity, and cryptographic security have been considered as parameters for measuring the performances. Based on those parameters, the performance of each building block has been compared with standard existing protocols and such blocks have been cascaded to form security toolkit. After experiencing a satisfactory outcome, these chains of actions have reached to the end of research work.

1.11. Contributions of Thesis

A considerable amount of research work has been carried out in data encryption and information hiding domain in last few decades. In the current work, new text, image, and authentication building blocks (i.e. algorithms) have been implemented and combined

together to form security toolkit with better security as compared to the existing standard schemes. The contribution of the abstract of this research is threefold.

In the first part, user authentication 'building blocks' based on CAPTCHA code and OTP have been implemented. Three authentication schemes namely "CAPTCHA Code based on User Personal information and Likings (CCUPL)", "Secret Value based on Randomize One Time Password (SVROTP)" and "Numeric and biometric Image based One Time Password (NIOTP)" have been designed for 'authentication building blocks'. CAPTCHA or OTP generation time, randomness and their security over different network and web attacks have been considered as standard parameters for assessing their performances.

In the second part, seven text encryption algorithms namely "Prime number with Alphabetic Group based text encryption (PAG)", "Palindrome number with Alphabetic Group and Operator based text encryption (PAGO)", "Multiple Operator and Even Odd position based text encryption (MOEO)", "Multiple Operator and ASCII Value based text encryption (MOAV)", "Multiple Operator and number of Zeros and Ones based text encryption (MOZO)", "Armstrong and Perfect number with Cipher Sequencing based text encryption (APCS)" and "Amicable number with Cipher Sequencing based text encryption (ACS)" schemes have been implemented for 'text encryption building blocks'. The parameters e.g. frequency distribution of characters in plaintext and ciphertext, the time required for encryption and decryption, Pearsonian chi-square value, formulation of the structure of private key is used for measuring the performance of 'text encryption building blocks'. Satisfactory outcomes of the 'building blocks' have been examined as compared with standard AES (Advanced Encryption Standard), Triple DES (Data Encryption Standard), Twofish, Blowfish and Serpent algorithms.

In the third part, key based encryption, digital enveloping, image partitioning, and user defined steganographic scheme based new 'image encryption building blocks' have been introduced. Three image encryption schemes namely "Even Odd block based Digital Enveloping scheme (EODE)", "Cumulative Image encryption using Digital enveloping,

Key based encryption with image Partitioning (CIDKP)" and "Cumulative Image encryption using Steganographic scheme with Pixel repositioning (CISP)" have been implemented for 'image encryption building blocks'. Cryptographic security, the speed of encryption, Peak Signal to Noise Ratio (PSNR), Mean Square Error (MSE), Structural Similarity Index Measure (SSIM), Universal Image Quality Index, Bit Error Rate (BER), Correlation Coefficient (CC) and Normalized Cross Correlation (NCC) have been considered as parameters for assessing the performances.

1.12. Thesis Organization

The thesis has been organized into nine chapters. Brief information of these chapters has been mentioned here.

Chapter 1: Introduction

Chapter 1 contains basic ideas related to different topics such as cryptography, special numbers, encryption, decryption, structure of color image, key based image encryption, steganography, digital enveloping and user authentication based on CAPTCHA and OTP which are very closely relating to this thesis.

Chapter 2: Early Work related to Current Research in User Authentication and Data Encryption

Chapter 2 represents the detailed literature surveys of private key cryptography, text encryption, user authentication based on CAPTCHA and OTP, key based image encryption and steganography. The attempt has been made for carrying out detailed background study on text encryption, user authentication, and image encryption domain.

Chapter 3: Performance Assessment Metrics

Chapter 3 of this thesis defines and elaborates the ideas of different network and web attacks, existing standard parameters, standard methodologies, algorithms and security tools for measuring the performances of newly implemented security toolkit with its building blocks (i.e. algorithms).

Chapter 4: User Authentication Building Blocks

Chapter 4 describes user authentication building block which contains three authentication schemes namely "CAPTCHA Code based on User Personal information and Likings (CCUPL)", "Secret Value based on Randomize One Time Password (SVROTP)" and "Numeric and biometric Image based One Time Password (NIOTP)". Performances of the authentication schemes are accessed with respect to provided protection over different cryptographic attacks.

Chapter 5: Text Encryption Building Blocks using Special Numbers with Alphabetic Group or Cipher Sequencing

Chapter 5 discusses four newly implemented text encryption schemes. The schemes are "Prime number with Alphabetic Group based text encryption (PAG)", "Palindrome number with Alphabetic Group and Operator based text encryption (PAGO)", "Armstrong and Perfect number with Cipher Sequencing based text encryption (APCS)" and "Amicable number with Cipher Sequencing (ACS)". All these text encryption schemes are based on special numbers, alphabetic group, operators and cipher sequencing and satisfactory performances have been measured as compared to standard algorithms like Twofish, AES, Triple DES, Blowfish, and Serpent.

Chapter 6: Text Encryption Building Blocks using Operators and Position of Bits in Plain Text

Chapter 6 includes newly introduced three text encryption schemes based on operators, even odd terms and numbers of '0','1' present in binary representation of plain text's character where the separate private key is applied for each character of a plain text file at the time of encryption. The schemes are "Multiple Operator and Even Odd position based text encryption (MOEO)", "Multiple Operator and ASCII Value based text encryption (MOAV)", "Multiple Operator and number of Zeros and Ones based text encryption (MOZO)". Appreciable performances are accessed with respect to standardizing parameters like encryption and decryption time, the frequency distribution of characters in plain text and ciphertext and chi-square value.

Chapter 7: Image Encryption Building Blocks

Chapter 7 contains newly implemented image encryption schemes namely "Even Odd block based Digital Enveloping scheme (EODE)", "Cumulative Image encryption using Digital Enveloping, Key based encryption with image Partitioning (CIDKP)" and "Cumulative Image encryption using Steganographic scheme with Pixel repositioning (CISP)". User defined pixels repositions procedure and Bitwise Masking for Alternate Sequence (BWMAS) operation have been introduced for the first time. Satisfactory outcomes have been examined as compared with other standard parameters for all the implemented image encryption schemes.

Chapter 8: Security Toolkit and Methodologies

Chapter 8 contains a brief description of newly developed 'security toolkit' which is constructed by combining the building blocks. Newly implemented user defined methodologies and operator used for building blocks are discussed in this chapter.

Chapter 9: Performance Analysis, Conclusions and Future Scope

Chapter 9 deals with the analysis of the outcomes and comparisons between the implemented building blocks in respect of standard parameters. Summary of research work and future scope of work have been included in the later section of this chapter.

1.13. Conclusion

This chapter contains a brief description of different concepts from cryptography, user authentication, steganography, key based text encryption, and image encryption domains. Special numbers, structure of color image have been discussed also in this chapter. The newly implemented algorithms have also been described and assessed with the standard techniques, existing parameters in subsequent chapters.