Text based steganography using golay computer science essay

Technology, Information Technology

ABSTRACT

Steganography is the art and science of transmitting hidden messages. In moderncommunications systems, this means hiding information in communication media such asaudio, text, and images. Ideally, except for the sender and receiver, no third party shouldeven suspect the existence of such messages. Digital communications systems require the useof error-correcting codes (ECC) to combat noise, or errors, introduced by the corresponding(communication) channel. Basically, an ECC adds redundancy to a message so that the errorsintroduced by the channel can be corrected. In our context, the code redundancy can be utilized to insert stega bits (that is, bits ofa secret message) masked in the form of artificial errors, which in turn, cannot bedistinguished from genuine channel errors. Therefore, noisy communication channelsprovide a suitable framework for steganography. In this work, we focus on text-basedsteganography. The underlying ECC is the Golay code, which breaks down the informationsequence into blocks of 12 bits. At the end of the encoding process, each 12-bit block istransformed into a 23-bit block, called a codeword. The Golay code is capable of correctingup to three errors in a block of 23 bits and is attractive for combating errors in very noisycommunication channels. Two modes of insertion of stega bits are discussed and compared. The modes represent a trade-off between accuracy and secrecy. In the first, a more accurateversion of the secret message is recovered in comparison with the second; however, it ismore susceptible to being detected by an eavesdropper than the second mode.

INTRODUCTION

Steganography is an art or science of transmitting hidden messages. In moderncommunication system, this means hiding information in communication media such asaudio, text and images. Steganography is supposed to be originated from Greek culturewhere the Greek word steganos means concealed and graphein means to write. Techniques for hiding information have existed for centuries. In Ancient Greece, secret messages were written on wooden plates and wax was used to cover them. Methodsinclude writing hidden messages on paper written in invisible ink in the blank spaces of thepapers. This technique was adopted quite successfully during World War II by the French. Some other techniques were also implemented in the past. Messages were written on theback of postage stamps. Germans used microdots during World War I and World War II. Microdots are nothing but a text or an image substantially reduced in size onto a disc of1 mm in diameter. Special cameras were used to generate microdots attached to letters. Thesemicrodots usually went unnoticed for any intruders and could easily read by the authorizedrecipient microscope. Techniques such as spread spectrum are used these days in digitalcommunication. Electromagnetic or acoustic signals generated for a specific bandwidth arespread over a much wider bandwidth to avoid signal interference or signal jamming. Digital watermarking is also one of the many applications of steganography. Visiblewatermarks are used for copyright protections and source tracking but in case of invisiblewatermarking, the information is difficult to perceive. The secret message is hidden in adigital signal. The spread spectrum mentioned earlier is used for audio

watermarking. SpreadSpectrum is used to embed watermarks which can be implemented easily in any timedomain. After spreading the spectrum the information is hidden in the form of a watermarkand is added to the sender signal as a watermarked signal. The core principal for steganography is that apart from the sender and the receiver, nothird party or the intruder can suspect the presence of any such hidden or covert message. This phenomenon clearly distinguishes steganography from a very renowned technique ofinformation hiding which is cryptography. In cryptography, the information is hidden bydoing encryption of the original message by various encryption algorithms. This encryptionprocess converts the plain text into a cipher text with the help of the encryption key. If ever athird party intrudes and manages to extract the cipher text, this encrypted message is hard todecode without the key. This clearly states that in Cryptography the third party can detect thepresence of secret message easily though it may or may not decrypt the encoded messagewhich is disparate from the principal of steganography which hides the message as well asthe presence of the message. Therefore where cryptography protects the content of themessage, steganography protects both message and communicating parties. So except theauthorized persons, no third party can ever think of any such secret message in thecommunication. Hence steganographic communications do not attract attention since they arenever highlighted or encrypted but always hidden. In computer systems as well, steganography is extensively used. Pictures areembedded in video material. Secure shell connections, remote desktop software such astelnet, virtual host always include some amount of delay before sending the informationpackets over

the network. These delays can be used to encode data. Texts are hidden in webpages. Information is concealed within computer files which can be audio files, jpeg imagesor bit mapped images which are larger in size and contain lot of information in it. Forexample, every nth color bit is replaced with some message bit and sent over the transportnetwork. This change is so minute that it usually goes unnoticed due to highly redundantcode stream. Some tools can be used to transmit valuable data in normal network traffic. InternetControl Message Protocol (ICMP) is an Internet protocol used for networked computers tosend error messages for diagnostic or routing purposes in IP datagram. These ICMPmessages are part of the IP header and transmitted the resulting datagram. Linux has a pingutility which adds 56 bytes of ICMP message to the existing header. Loki is another such toolthat hides data in ICMP traffic. Loki is a client-server program which can be used to transmitdata secretly across the network through back door into a Unix system. A directory startingwith dot (.) is a hidden directory. A directory starting with three dots (...) can be created tostore secret files so they do not come into the file lists. On Windows systems, theC,/winxp/system32 or C,/winnt/system directories are where all the Windows . dll, . dib andset up files are placed. This directory can be used to securely store all the covert filesassuming that no one really dares to tamper or touch the files in those important directories.

REQUIREMENTS

The primary aim behind developing this tool is to understand how steganography canbe achieved using error correcting codes for very noisy

communication channels with covermedia being a text file. Upon completion, this tool will help the students from mathematicsand statistics as well as computer science in learning more about error correcting codes andtheir applications and different techniques used in Information security branch for secure andcovert data communication. This software can also be integrated into different text editorssuch as office word, Kwrite etc. for creating documents with secret messages embedded. Therequirements gathered have been further classified into platform requirements and functionalrequirements.

PLATFORM REQUIREMENTS

• The main objective in choosing the software development kit (software language)was that it should be platform independent so that final product will have thecapability to run on any environment irrespective of the operating system.• The software should run as a stand-alone application rather than a web basedsoftware. Hence Java SDK instead of Java Enterprise Edition has been chosen to bethe appropriate language for writing the code. The operating system is Ubuntu Linuxconsidering in mind the importance of open source software.• There should be a facility to store different versions of code and some repositorywhere code can be checked in and checked out. This is applicable whenever anymodifications are made or any new feature gets added. Keeping this in mind, SVN(subversion) repository has been used.

FUNCTIONAL REQUIREMENTS

• The software should be able to read any large text file as a cover media so that secretmessage can be embedded into it.• The Error correcting code

(ECC) being chosen should not induce additionalcomplexity to the existing bit stream.• The ECC should not increase the bandwidth of the channel by adding too manyredundant bits in such a way that the performance and efficiency gets hampered.• The core principle of steganography should be achieved. That is, message as well asits existence should be concealed.• The transportation of the message blocks over the communication channel should nottake large amounts of time.• The decoder version of the software should be capable enough to join all the receivedand decoded data blocks to get back the original secret message is received.• Since the Golay code is used as an ECC which can correct up to three errors per23 bit codeword, there should be a facility for the encoder to select how manyartificial errors (stega bits) he wants to send for each codeword.• In the Golay code mode2, the detection scheme should be intelligent enough to selectthe erred bits which were not picked by the normal decoding mechanism.• There should be a facility to introduce the genuine channel errors along with theartificial errors so that the communication channel used will look more original.

NOISY CHANNELS AND ERROR CORRECTING CODES

Any message to be transmitted over the communication channel needs some level ofprotection. This is because of many things such as noise, channel error in the communication. These hindrances not only change the message content but also the meaning of the messagevery commonly known as noise.

NOISE

Noise is an unwanted part of any digital or analog signal. It is the factor which isresponsible for degrading the quality of the signal by acting as an interference or blockage inthe communication channel. This entity is naturally present in most of the communicationchannels corrupting the signals passed over. Hence signal to noise ratio should be as high aspossible to ensure the error free communication. However the occurrence of noise is totallyrandom and if proper filter is not used to detect its presence, it usually goes unnoticedcreating disturbances at the receiver's end. To combat the noise, various techniques are usedsuch as increasing the power of the signal, implementing some sort of a modulation such asfrequency modulation (FM), amplitude modulation (AM) etc. or adding a lot of redundantbits to the original signal. Some of these operations are expensive such as increasing thepower of a signal or amplitude modulation. Adding a lot of redundant bits can also be aninefficient method since it increases the channel bandwidth beyond capacity. But if handledin a proper manner, through error correcting codes this method can be used very effectivelyfor to detect the noisy bits or errors at the receiver's end. A very noisy channel is an attractive medium for steganographic communications. This is achieved by having a low transmission rate of concealed messages sent block byblock. This type of communication produces an almost untraceable secret message transfer. Since these noisy channels require error correcting codes to come over the noise, the coderedundancy is utilized very ingeniously to insert the secret message bits to be passed in formof artificial channel errors. The insertion of steganographic bits over a honest communicationchannel is possible only

because of the bit redundancy created by the error correcting codes. Error correcting codes (ECC) provide a technique for data transmission in which afew extra bits are added in each block of data in order to detect the errors and then correctthose errors that may occur in the communication. These redundant bits also known as paritybits make sure that the message is received error free at the receiver's end and once the errorsare corrected, these parity bits are easy to remove from the original content. Parity Bits takecare of the number of checked bits in the code i. e. they are implemented as odd parity whennumber of 1's in a given set of bits are even and similarly for even parity. ECC have been successfully used in physical and data link layer of the OSI modeland also implemented in data disks and computer physical memories in the form ofchecksum. Checksum is an addition of all the codewords in a given set of bits. ECC aremainly divided into two classes viz. convolution and block codes. Convolution codes operateon bit by bit basis and block codes are processed per block basis. This thesis is focused onblock codes. ECC are also known as forward error correction (FEC) since no re-transmissionof the message is performed. There are many types of block codes such as repetition codes, BCH codes, Golay codes, Hamming codes. Some of these are very efficient codes such asGolay, BCH etc. Golay Codes are encoding and decoding techniques are implemented in thisthesis work. The channel used for steganographic communication is a binary symmetricchannel explained below.

BINARY SYMMETRIC CHANNEL

The communication channel used in most of the steganographic communications andencoding-decoding mechanisms of Golay codes in this thesis is a binary symmetric channel(BSC). Let us assume that the communication messages are sent over a very noisy channeland that we can send only two symbols viz. 0 and 1. Also assume that when the sender sendsthe symbol 0, the receiver receives the same symbol 0 with probability p and receives 1 withprobability q. Similarly when the sender sends symbol 1, the probability that 1 is received isp and 0 is received is q. Then for a binary symmetric channel, p + q = 1 also meaningq = p – 1. This type of communication is possible only in binary symmetric channel (BSC). BSC is very frequently used in information and coding theory. It is assumed that in BSC, thesymbol is received with a high probability, but if the symbol or the bit gets flipped, then thatprobability is very small. Figure 4. 1 illustrates this type of communication channel veryclearly. In this diagram, if X is a variable that is randomly transmitted and Y is the variablerandomly received, then the channel is characterized by the conditional probabilities shownbelow: Pr(Y = 0 / X = 0) = p (4. 1)Pr(Y = 1 / X = 0) = 1 – p (4. 2)Pr(Y = 1 / X = 1) = p (4. 3)Pr(Y = 0 / X = 1) = 1 – p (4. 4)

Binary symmetric channel.

LINEAR CODES

A linear code of length n and dimension k is a subspace C of the vector space (F2)n(all n tuples with entries in F2 where F2 ={0, 1} the binary field). Such a code is referred to an(n, k) code. Elements of the code are called

codewords. A generator matrix for an (n, k) codeis a k x n matrix whose rows form a basis for the vector space C. This generator matrix oftendenoted as G, is of the form (Ik | A) where Ik is an identity matrix (k x k) and A is thestandard matrix with dimensions ((n-k) x n). The Hamming distance of a linear code C is equal to the minimum distance betweenany two codewords in C. This is also equal to the minimum weight of the nonzero codewordsin C. These linear codes bear a special property that any two non zero codewords c Є C differ14in at least d positions where d is the Hamming distance between the two codewords andaddition of any these codewords, e. g. c0 and c1, fetch a result say c2 which also is a codewordbelonging to the same set C. In mathematical terms, Hamming distance is the number of coordinatesIn which c0 and c1 disagree.

GOLAY CODES AND APPLICATIONS

Golay codes belong to the class of ECC which are used in mathematics and computerscience fields extensively. Golay codes were invented by Marcel J. E. Golay. Golay was aSwiss mathematician. Golay codes over the years have played an important role in both thetheory and practice of ECC. ECC are generally defined over the finite field namely galoisfield which contains limited number of elements used in digital communications. It isdenoted by GF(n) or Fn where n is the maximum number of elements that can belong to therespective finite field. For every prime integer p and positive integer n, there exists a finitefield with pn elements . With n = 2, the galois field becomes binary and only the two specifiedelements can be present in each of codeword belonging to C. In most of the cases, F2 = {0, 1}were ‘+’ and ‘.’

are addition and multiplication modulo 2. Golay codes are divided into two branches, namely binary and ternary. Binary Golaycodes use 0 and 1 in GF(2) and they are further classified into extended binary Golay codes(EBGC) or perfect binary Golay codes (PBGC). This thesis is designed for PBGC which ismost commonly used in practice and often denoted as just Golay codes. In case of PBGC, the codeword has a length 23. They are denoted as [23, 12, 7]. Thenumbers in the brackets are explained below. A block of twelve binary bits is converted intoa block of 23 bits after passing through a Golay code encoder. Due to the binary nature of a12 bit message block, the total number of codewords this vector space contains is 212 = 4096. Any two codewords belonging to this 12-dimensional subspace W of the vector space Vdiffer in at least 7 positions or equivalently, any non zero codeword has at least seven 1's in itor the Hamming distance between any two codewords is 7. By adding just one parity bit toevery encoded 23 bit codeword, the codeword becomes 24 bits and the perfect Golay codesare transformed into extended Golay codes. Extended binary Golay codes denoted as[24, 12, 8] are very similar to perfect Golay codes except the fact that any Hamming distancebetween any two codewords is 8. Implementing the encoding and decoding mechanisms of Golay codes is simple butvery efficient. The message bits are divided into blocks of 12 bits each. Every such block Mof 12 bits is converted into a block of 23 bits by multiplying it with a generator matrix of(12 x 23). This generator matrix has its rows 12 different and non-zero Golay codewordswhich differ from each other in at least seven positions. An example of one such generatormatrix is shown in Figure

Generator matrix.

The subsequent multiplications and additions which occur in M x G are modulo 2. Hence the resulting row matrix of 23 bits contains only binary numbers 1 and 0. In thedecoding process, the received vector r of length 23 is possibly corrupted in up to threecoordinates. Matrix r is multiplied with the transpose of a matrix A to get the vector s. MatrixA has 506 rows and 23 columns whose rows are dual of Golay codewords. The dual of alinear code C Є (F2)n denoted as $C^{\perp}$ is defined as follows: $C^{\perp}$ = {u Є (F2)n | u . v = 0 for allv in C } where u . v is the usual scalar product between u and v and the dimension of $C^{\perp}$ isequal to n-k. All the additions and multiplications are modulo 2 again. The resulting vector s (1 x 506) is multiplied by matrix A (506 x 23) to get the vectorv (1 x 23). In this case the additions and multiplications are the usual integer additions andmultiplications. Hence 1 + 1 + 1 becomes 3 and not 1 as in the case of modulo 2 operations. The error pattern is determined for each of the 23 bits of v and stored in e where ei is the ithposition of e for i = 1, ..., 23. For each position vi, for i = 1, 2 ...., 23 in v, if vi = 176, 120, or96, then the corresponding bit is in error. This makes ei = 1. In all other cases ei is 0. After determining the error vector e, it is added to the received vector r in modulo 2so that the erred bits are flipped to get back the correct code that was originally sent. Let usconsider an example where the 12 bit binary stream 111000111000 was converted to 23 bitGolay codeword 11100011100011111011111 after multiplying with the generator matrix G. Assume 2 bits were corrupted at 11th and 18th position respectively and hence the streamwhich the receiver received was 11100011101011111111111. When the decoding techniquesare applied, the

error matrix which is generated is nothing but 000000000010000000100000contain 11th and 18th bits as 1 and rest others as 0. This error matrix e when gets added to thereceived vector r yields the original codeword which is 111000111000111111011111. Extended binary Golay codes (EBGC) can correct up to four errors and detect up toseven errors whereas PBGC can correct up to three errors and detect up to six errors in a23 bit codeword. This rate of error correction is very high compared to other ECC hence theywere used extensively in communication channels and spacecraft programs in the late 1980s. Golay Codes with proper combination of codewords reduce the amount of noise in thechannel which magnifies the signal to noise ratio making the code useful for biomedicalDoppler applications. Binary Golay Codes have also been used in NASA spacecraft mission. In the 1980s when the channel bandwidth was very limited, hundreds of thousands of highresolution colorful images of planets such as Jupiter and Saturn were sent using Golay CodesEncoding due to the high probability of error free receipt of these images. EBGC [24, 12, 8]were used in this case since color images required to send 3 times the amount of data andthese codes can correct up to three errors. In high frequency radio systems communications, EBGC have been used for forward error correction according to American governmentstandards.

ENCODING OF STEGANOGRAPHIC MESSAGES

As explained in the previous chapters, error correcting codes add bit redundancy tothe original message block. This redundancy is added to make sure that bits that might getcorrupted during the communication over the

noisy channel are received without any error. This bit redundancy is utilized for steganographic purposes where the secret message bits canbe passed. These secret message bits are known stega bits which are replaced with redundantchannel bits. Also the channel used is binary symmetric channel (BSC). A very noisy channelwill have lot of genuine errors, i. e. errors that have caused just because poor channelreception without any steganographic intervention. To differentiate these errors from theartificially inserted stega bits, the stega bits are known as artificial channel errors. The twobasic modes that have been implemented are discussed be.

FLOW OF ENCODING MECHANISM

In MODE 1, the stega bit is inserted in a fixed position within a codeword c ∈ C asfollows:• If the selected bit in the codeword is 0 and the stega bit is also 0, then the bit is notinserted into the codeword. Same would be the case if both the stega and codewordbits are 1.• If the selected bit in the codeword is 1 and the stega bit is 0, then 1 in the codeword isreplaced by 0 stega bit.• If the selected bit in the codeword is 0 and the stega bit is 1, then 0 in the codeword isreplaced by 1 stega bit. In MODE 2, the stega bit is inserted in a random position within a codeword c ∈ C asfollows:• 0 stega bit is inserted in a codeword as an artificial error in such a way that it replacesany randomly selected 1 in a codeword c ∈ C.• 1 stega bit is inserted in a codeword as an artificial error in such a way that it replacesany randomly selected 0 in a codeword c ∈ C.• If the codeword does not contain any position occupied by 1 and 0 is to be inserted, then no stega bit is inserted. 18• If the codeword does not contain any position occupied by 0 and 1 is to be inserted, then no

stega bit is inserted.• In both modes, code C is used to recognize both the error bit position as well the erredstatus. Thus the decoding mechanism should be intelligent enough to find out the stega bitpositions. Golay Codes, Hamming Codes or Repetition Codes have excellent decodingmechanisms to find out the bits in errors and correct them. However the separation ofgenuine errors from the artificial errors is a difficult task which is handled by separatetechniques in both the modes. In case of Mode 1, the stega information is encoded in a known position hence at thedecoding end, it is quite easy to determine which bit is in the error and then to check the errorstatus. In case of Mode 2, the stega information is carried by a bit which is randomly encodedin a codeword unknown to the receiver, hence it becomes very difficult to identify the erredbit at the receiver end just with the help of normal decoding mechanism. Therefore alongwith this decoding mechanism, special detection criteria must be defined. The entire process of encoding and decoding the stega message to pass over thesteganographic channel (which is BSC) is shown in Figure 5. 1. S is source and U is a user. Eis a binary encoder which uses a liner code C, a BSC with error probability of p and DecoderD. A decoding rule is for the corrupted codewords is described by the full set of coset leadersT and encoding mechanism converts the Codeword into a encoded bit stream K passed overstega-channel.

Flow of steganographic process.

REPETITION CODES

Repetition codes are a type of error correcting codes. They are denoted as (r, 1) wherer is the repetition index, which makes every bit of the secret

message repeat by the repetitionindex r. Hence they are called repetition codes. For repetition codes, each codeword is oflength r. For example if the signal s = 10100 and the scheme used is (3, 1) then every bit of sis repeated thrice and sent over the communication channel. Hence the extended codewordbecomes c = 111000111000000. An encoder for a repetition code is a simple device which repeats every bit of theinformation sequence by the repetition index. This method of encoding is very simple butquite unsophisticated. As a result, it is hardly used in practical applications. But it can give anice insight about the steganographic process for ECC and also which method of ECC is tobe used based on efficiencies. At the receiver end for every bit stream of length r, thecodeword is compressed into a single bit using the majority element rule. For example, consider the codeword received is 110000111000001. In this case since r is 3, every set ofthree bits is compressed into a single bit using this majority rule which simply counts thenumber of occurrences of 1 and 0 and sets the output bit according to the bit in majority. Here:• 1st 3-bit block is 110, hence the majority element is 1.• 2nd 3-bit block is 000, hence the majority element is 0.• 3rd 3-bit block is 111, hence the majority element is 1.• 4th 3-bit block is 000, hence the majority element is 0.• 5th 3-bit block is 001, hence the majority element is 0. All five bits are concatenated to get the decoded message which is 10100. Hence atthe receiving end, the user receives original signal which is 10100. Repetition codes offer a poor solution for the large files datasets since they repeatevery bit by r times and increase the channel bandwidth beyond its capacity. The onlyadvantage of repetition codes is that their execution is fairly simple. Use of repetition codesfor steganographic communication is

explained in the following example. Consider a secret message as 10100010 and cover media which can be a text file oran image be 11010000. We are using repetition codes with scheme (3, 1), i. e. r = 3 whichmakes the encoder repeat every bit of cover media by 3. This operation generates theelongated version of cover media X as 111111000111000000000. Every first bit of the 3 bitblock is ex-or with one bit each of secret message as follows: 20X = 111 111 000 111 000 000 000 000X-OR1 0 1 0 0 0 1 0

---

011 111 100 111 000 000 100 000This new X' which is 011111100111000000100000 is transmitted over a binarysymmetric channel. Lets suppose that this noisy channel introduces few errors in X and themessage received at the User end is Y = 011111100111001000100000 with the marked bit asa corrupted bit due to genuine channel error. A majority element array is calculated for thisreceived codeword considering majority rule for every three bits. The decoding operation isdescribed as follows: Y = 011 111 100 111 001 000 100 000R = 111 111 000 111 000 000 000 000X-OR

Y' = 100 000 100 000 001 000 100 000For Y', every 1st bit of the three bit block is fetched to get the secret message. Thesecret message received is 10100010. The secret message that was sent was 10100010. Hence we have got the exact same secret message despite the fact that there were someartificial errors i. e. stega bits and genuine channel errors. This technique worked for thesmall codeword but looking at the bigger picture, a

color image with RGB as (8 bit *3) 24 bitimage would be converted to a 72 bit stega channel which would be very inefficient. Hencefrom now on, the steganographic communication is achieved using Golay codes which aremore powerful than repetition codes. Due to their simplicity, the repetition codes arecurrently used in following applications:• Some Universal Asynchronous Receivers and Transmitters use majority filters toignore modulations in the noise known as noise spikes. This spike injection filter is arepetition decoder.• Many frequency modulation techniques in the current world transmit a single bit or ablock of few bits over many sinusoidal signal cycles. The low pass filter used at thereceiver's end for the entire bit stream is assumed to be a repetition decoder.

GOLAY CODES ENCODING

Encoding of Golay codes has been explained in brief in the previous chapter. Theprogramming details and implementation techniques that were used in this thesis aredescribed below. Golay codes can detect up to six errors and correct up to three errors in acodeword of 23 bits. The example considered here has a cover media as a text file and secretmessage as a plain text. Both the cover media and the secret message are converted intobinary streams of 1 and 0. This conversion takes place using a function named readFile(). This function has been written in Java and explained below. The readfile() command takes its argument as the filename whose contents aresupposed to be converted and stored into a binary stream. This stream is created by a Javaclass which opens the file, reads it and makes use of Java input-output functions by importingJava. io api. The large stream is first read into a

variable which contains non-binarycharacters. This string is separated into individual characters and each character is stored intoa character array as an array variable using tocharArray() function. Hence if the string isDesktop, the character array, e. g. arr[], is created and Desktop is stored into arr[] asarr[0] = 'D', arr[1] = 'e', arr[2] = 's' etc using the tocharArray() function. Each character has an associated integer value with it. This integer is also known asan ASCII value. For example, 'D' has an ASCII value 68, 'A' has ASCII value 65, 'B' is 66, 'e'is 101, number '0' is 48 and Space is 32. The character array elements are converted into therespective integer values using typecasting technique of Java. Typecasting changes the datatype of a variable to another data type. Here a character is converted to integer by simplyspecifying the character as (int) character, which converts it into an integer value. Thisinteger value is further converted into its respective binary value which ultimately we wantusing a direct toBinaryString() function which Java provides. Now 'A' <=> 65 which when converted to binary fetches 1000001 is a 7 bit binarystream. Another value, e. g. number '6', <=> 54 when converted to binary stream fetches110110 which is a 6 bit binary stream. When entire message stream needs to be convertedinto a binary stream, all these small binary streams are concatenated. At the decoder end, itwould be fairly difficult for the decoder to divide the received bit stream back into ASCIIvalues since the length of individual characters in binary are different as we just saw. Toovercome this problem, all the characters are converted to 8 bit binary streams by appendingzeros at the start. Hence 'A' becomes 01000001 by appending one zero and '6' is converted to00110110 by appending two zeros at the start. The entire process of starting with a

string andfetching an 8 bit binary stream is explained in the following function: public int[] readFile(String passedString)

{

tempArray = new char[passedString. length()], tempArray2 = new char[tempArray. length * 8], intempArray = new int[tempArray. length], longInt = new int[tempArray2. length], String finalString = "", for(int i= 0, i } String str7 = " 0", String str6 = " 00", tempArray = passedString. toCharArray(), for(int i= 0, i } else if(strBinary. length()== 6){strBinary = str6. concat(strBinary),} finalString = finalString. concat(strBinary),

}

}

All the small binary sub-streams are concatenated and stored into a final large streamdeclared as finalString shown above. For storage purposes, the finalString is stored into acharacter array where each character is either 1 or 0. This entire character array is convertedinto an integer array (not into the ASCII integer values but just the regular integer values)where character '1' gets replaced with integer 1 and character '0' is replaced with integer 0. Thus a large binary string of cover media is generated at the sender using the abovetechnique. This binary string is now divided into blocks of 12 bits each. Remember thatGolay codes encode 12 bit block into a 23 bit codeword using encoding mechanism. This12 bit block is first transformed into a row matrix with twelve elements, that is, ifX = 101000101110, then X is converted into a row matrix as shown in Equation 5. 1: M = [1, 0, 1, 0, 0, 0, 1, 0, 1, 1, 1, 0] (5. 1)M is multiplied to the generator matrix G of Golay codes

which have 12 rows and23 columns. The matrix multiplication (M x G) produces a matrix A, a row matrix with23 bits. This row matrix is one of the 4096 codewords of the Golay codes. As explainedbefore, all these subsequent multiplications and additions are modulo 2. Hence matrix A willcontain only binary numbers 1 and 0. The conversion from 12 bit block to 23 bit codeword isprocessed for all the subsequent binary blocks of stream created of cover media text file andall these 23 bit codewords are appended to generate a large stream of bits. This conversionand matrix multiplication is explained below in function matMultiply(): public int[][] matMultiply(int [][] A, int [][] B)

{

int C [][] = new int[A. length][B[0]. length], for(int i= 0, i }} for (int i

=

0, i

<

A. length, i++){for (int j = 0, j < B[i]. length, j++){for (int k = 0, k < A[i]. length, k++){C[i][j] = ((C[i][j]+(A[i][k] * B[k][j]))%2),

}

}

}

return(C),

}

This function takes its arguments as two matrices supposed to get multiplied viz. matrix A and matrix B. The result is stored into matrix C. For matrix multiplication, thenumber of rows of (B) must equal the number of columns of (A). Also the resulting matrix Cshould have dimensions such that there are the same number of rows of (C) as (A), and thesame number of columns of (C) as (B). Hence the matrix C is declared as " int C [][] = new int[A. length] [B[0]. length]," whereA. length specifies number of rows of A and B[0]. length specifies number of columns of B. The matrix C is initialized to zero. The matrix multiplication follows the following rules:• Every element from the row of matrix A is multiplied to its corresponding columnelement of the matrix B. For example, the 1st row, 3rd column element of matrix A ismultiplied by the 3rd row, 1st column of matrix B and so on.• All the intermediate multiplications are added for each row of A and each column of