

CHAPTER I: INTRODUCTION

Nowadays, as modern technology advances and the usage of the Internet becoming more and more common to the world, the need to secure information arises. At most of the time transmitting sensitive information through the internet can be secured using cryptography protocols such as Internet Protocol Security (IPsec) and Secure Socket Layer (SSL). But when it comes to transmitting confidential data and information, we must be more cautious as leaking of the confidential data might cause lost for individuals, organisation, and government. In general, cryptography and steganography can be used as two security mechanisms in order to secure the confidential data. In fact, cryptography changes (encrypts) the original data, plaintext, into an unreadable representation of random scrambling characters, cipher text that leads to generate an obvious data that can raise the attention of an attacker in order to access to the confidential data. Furthermore, since the cryptography algorithms are available to the public and the security of them is relied on the key, in a case that the cryptographic key is compromised, the attacker can decrypt (read) the data successfully. Therefore, to overcome this problem, steganography can be used as an additional mechanism to provide secrecy to the confidential data. As a matter of fact, steganography hides the data within another data using two functions, namely embedding (using any embedded algorithm/steganography tools) and extraction (using steganalysis tools) from following fig. In addition, since the stego media that contains a secret message, which is appeared to be a normal file in the channel, only recipients who are aware of the steganography technique can retrieve the hidden data from the stego media. Since such this hiding method allows the data to be concealed from the attackers, it can be used in conjunction with cryptography in order to provide more security for the confidential data.

Steganography is the practice of concealing a file, message, image, or video within another file, message, image, or video. The word *steganography* comes from New Latin *steganographia*, which combines the Greek words *steganós*, meaning "covered or concealed", and Latin *-graphia* meaning "writing".

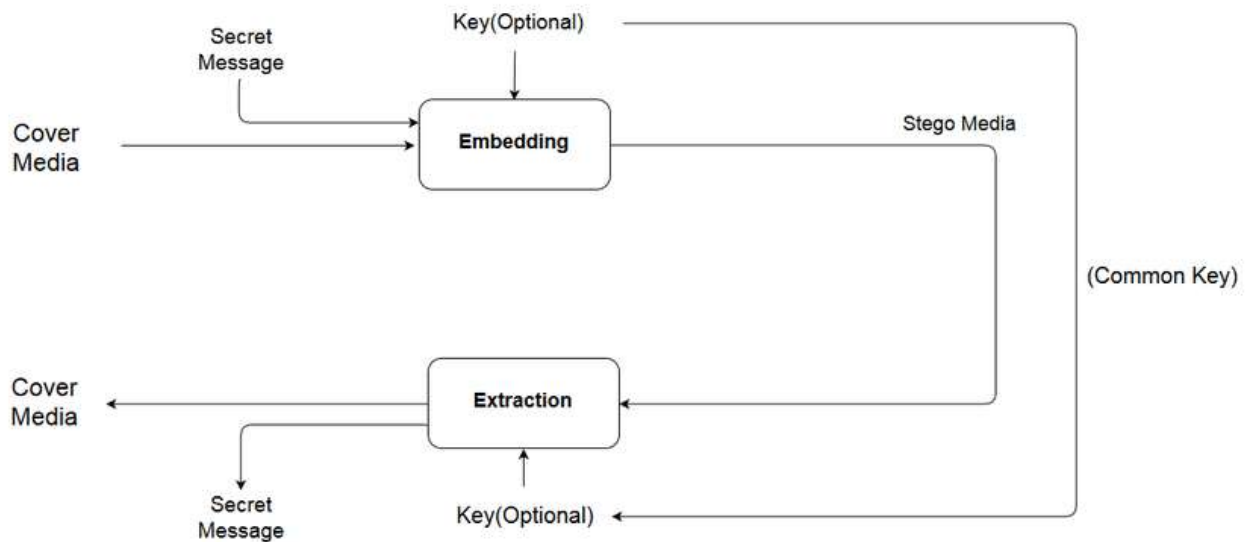


Fig 1: Flowchart of Text Steganography

Generally, the hidden messages appear to be (or to be part of) something else: images, articles, shopping lists, or some other cover text. For example, the hidden message may be in invisible ink between the visible lines of a private letter. Some implementations of steganography that lacks a shared secret are forms of security through obscurity, and key-dependent steganographic schemes.

The advantage of steganography over cryptography alone is that the intended secret message does not attract attention to itself as an object of scrutiny. Plainly visible encrypted messages, no matter how unbreakable they are, arouse interest and may in themselves be incriminating in countries in which encryption is illegal. Whereas cryptography is the practice of protecting the contents of a message alone, steganography is concerned both with concealing the fact that a secret message is being sent and its contents.

Steganography includes the concealment of information within computer files. In digital steganography, electronic communications may include steganographic coding inside of a transport layer, such as a document file, image file, program or protocol. Media files are ideal for steganographic transmission because of their large size. For example, a sender might start with an innocuous image file and adjust the colour of every hundredth pixel to correspond to a letter in the alphabet. The change is so subtle that someone who is not specifically looking for it is unlikely to notice the change.

White space steganography is used to conceal messages in ASCII text by appending **whitespace** to the end of lines, spaces or tabs. If encryption is used the message cannot be read even if detected.

White space text steganography is a common information hiding method, which utilise the spaces within the text file that can be found in different locations such as the end of each sentence, each line, between words or after each paragraph. A fig shows an example of adding white spaces at the end of each line, which can be used to embed any hidden data within the text file. However, in a case that the number of whitespace within the text is not sufficient, few characters as the secret message can be hidden respectively. In addition, by comparing both modified and non-modified text file using any kind of text processor such as Microsoft word, the difference as whitespaces can be leaked as the hidden data.

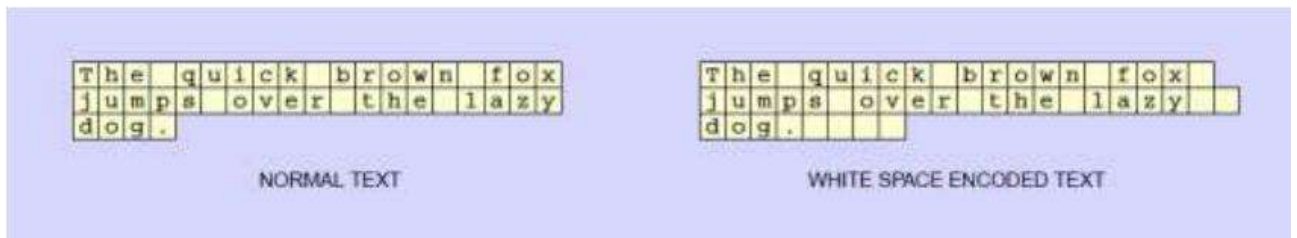


Fig 2: Example of difference in Text Steganography

We are proposing a new approach on hiding information through manipulation of whitespaces between words and paragraph as a hybrid method, namely WhiteSteg. WhiteSteg is able to provide more capacity for hiding more bits of data into a cover-text. Hiding information within spaces appears to have potential as people can hardly identify the existence of the hidden bits which appear in the whitespaces. Bender et al. had shown that one space is interpreted as “0” whereas two spaces are interpreted as “1”. This embedding scheme was applied in the space which appears between the words. The major drawback of Bender’s method is that it requires a great deal of space to encode few bits. For example, a character is equivalent of 8 bits, and it requires approximately 8 inter-spaces to encode one character. However, this problem can be resolved by compressing the character in the secret message from 8 bits to 3 bits in WhiteSteg. With the inter paragraph spacing, more whitespaces in the cover text document could be utilised effectively. In the embedding method, spaces can be inserted between a newline character and another newline character. Thus, WhiteSteg is hybrid schemes of both inter word and inter-paragraph spacing. Currently, manipulation of whitespaces seems beneficial and has its potential in information hiding because whitespaces appear in a text documents more than the appearance of words. Thus, this can be an advantage when no one will know that a blank piece of document is actually vital secret information which can be retrieved after a decoding process. The syntax that involve in whitespace manipulation are space (ASCII char 32), tab (ASCII 9) and line feed (ASCII 10). Even though this syntax appears to be invisible, but it is an advantage to

utilise this syntax in text steganography, especially appending spaces and manipulating them to hide information. The cover-text will be dynamically generated according to the length of the secret message. The maximum capacity of hidden bits is determined by the system whether the length of the secret message can be accommodated in the particular capacity. A minimum input of text by a user is considered where one character or a few characters as a secret message. Thus, the cover text will be generated in 1 byte, 2 bytes, 4 bytes, 8 bytes, 16 bytes, 32 bytes, 64 bytes, 128 bytes, 256 bytes and 512 bytes. This range of maximum capacity is used for the dynamic generated cover text to obtain the preliminary experimental results. The results will be analysed in order to determine to what extent of the payload can be assigned to accommodate the secret message so that the frequent occurrences of extra spaces may not alert the potential adversaries. By adapting the source of the generated text from any lyrics of the nursery rhymes, the stego text will definitely present an innocuous and naive appearance. Besides, the chorus of the lyrics can be duplicated and reused to generate a longer cover text. There is a greater advantage in using centered alignment because the appearance of the spacing occurs naturally without arising even the slightest suspicion. The significance of this research is by merging the two concepts and creating an algorithm for the data embedding system. The capacity of the cover text depends on the length of the secret message. A user is required to enter the secret message as provided in the text field or by selecting a file that contains secret message. The system will calculate the length of the message and generate a cover-text which is suitable to encode the secret message. The size of the payload (amount of hidden information) which is calculated by the system and assigned to the most appropriate cover-text is depended upon the indication.

CHAPTER II: LITERATURE REVIEW

The visual attacks (Westfeld et al.) detect the steganography by making use of the ability of human eyes to inspect the images for the corruption caused by the embedding.

Pairs analysis was proposed (Fridrich et al.). This approach is well suited for the embedding archetype that randomly embeds messages in LSBs of indices to palette colours of palette image.

The F5 algorithm was introduced by German researchers (Westfeld). It embeds message bits into non-zero AC coefficients and adopts matrix encoding to achieve the minimal number of changes in quantised coefficients during embedding process. The matrix encoding is the core of the F5 algorithm. It is determined by the message length and the number of non-zero AC coefficients. It can be represented as the form $(c, n, \text{ and } k)$. The parameter c tells how many coefficients at most will be modified during embedding, and n is the number of coefficients involved in embedding the k -bit message. In the embedding process, the message is divided into segments of k bits to embed into a group of n randomly chosen coefficients. F5 algorithm manipulates the quantised coefficients when the hash of that group does not match the message bits, thus the histogram values of DCT coefficients are modified. For example, if the shrinkage occurs, the number of zero AC coefficients will increase and the number of remaining non-zero coefficients decreases with embedding. The changes in the histogram of DCT coefficients may be utilised to detect the presence of hidden message.

Fridrich et al. developed a steganalytic technique based on this for detection of LSB embedding in colour and grayscale images. They analyse the capacity for embedding lossless data in LSBs. Randomising the LSBs decreases this capacity. To examine an image, they define Regular groups (R) and Singular groups (S) of pixels depending upon some properties. Then with the help of relative frequencies of these groups in the given image, in the image obtained from the original image with LSBs flipped and an image obtained by randomising LSBs of the original image, they try to predict the levels of embedding.

Many steganalysis researchers such as Neil et al. attempt to categorise steganalysis attacks to recover modify or remove the message, based on information available. The steganalysis technique developed can detect several variants of spread-spectrum data hiding techniques (Marvel et al.). The first steganalysis technique using Wavel et decomposition was developed (Farid). Fridrich et al. , have shown that this change is proportional to the

level of embedding. They also showed that, if an image is cropped by 4 rows and 4 columns, then original DCT histogram can be obtained.

The basic assumption here is that the quantised DCT coefficients are robust to small distortions and after cropping the newly calculated DCT coefficients will not exhibit clusters due to quantisation. Also, because the cropped stego image is visually similar to the cover image, many macroscopic characteristics of cover image will be approximately preserved. After predicting DCT coefficient's histogram in the original image and comparing with that of a stegoed image, the hidden message length can be calculated. Sullivan et al. use an empirical matrix as the feature set to construct a steganalysis. Chen et al. enhanced and applied the statistical moments on JPEG image steganalysis.

An early method used to detect LSB hiding is the χ^2 (chi-squared) technique later successfully used to steg detect for detection of LSB hiding in JPEG coefficients. Another LSB detection scheme was proposed by (Avcibas et al.), using binary similarity measures between the 7th bit plane and the 8th (least significant) bit plane. It is assumed that there is a natural correlation between the bit planes that is disrupted by LSB hiding. This scheme does not auto-calibrate on a per image basis, and instead calibrates on a training set of cover and stego images. The scheme works better than a generic steganalysis scheme, but not as well as state-of-the-art LSB steganalysis.

Another LSB detection scheme was proposed using binary similarity measures between the 7th bit plane and the 8th (least significant) bit plane. It is assumed that there is a natural correlation between the bit planes that is disrupted by LSB hiding. This scheme does not auto-calibrate on a per image basis, and instead calibrates on a training set of cover and stego images. The scheme works better than a generic steganalysis scheme, but not as well as state-of-the-art LSB steganalysis.

Scheme, proposed by Fridrich et al. is a specific steganalysis method for detecting LSB data hiding in images. Sample pair analysis is a more rigorous analysis due to (Dumitrescu et al.) of the basis of the RS method, explaining why and when it works. Roue et al. uses estimates of the joint probability mass function (PMF) to increase the detection rate of RS/sample pair analysis. Fridrich et al. uses local estimators based on pixel neighbourhoods to slightly improve LSB detection over RS.

Harmsen et al. proposed steganalysis of additive hiding schemes such as spread spectrum. Their decision statistic is based initially on a PMF estimate called histogram. Since additive hiding is an addition of two random variables: the cover and the message sequence, the PMF of cover and message sequences are involved. In the Fourier domain, this is equivalent to multiplication. Therefore the DFT of the histogram, termed the histogram characteristic function (HCF), is taken. It is shown for typical cover distributions that the expected value or centre of mass (COM), of the HCF does not increase after hiding, and in practice typically decreases. The authors choose then to use the COM as a feature to train a Bayesian multivariate classifier to discriminate between cover and stego. They perform tests on RGB images, using a combined COM of each colour plane, with reasonable success in detecting additive hiding.

Fridrich et al. content-independent stochastic modulation is statistically identical to spread spectrum and Celik et al. proposed using rate-distortion curves for detection of LSB hiding. They observe that data embedding typically increases the image entropy, while attempting to avoid introducing perceptual distortion to the image. On the other hand, compression is designed to reduce the entropy of an image while also not inducing any perceptual changes.

It is expected therefore that the difference between a stego image and its compressed version is greater than the difference between a cover and its compressed form. Distortion metrics such as MSE, mean absolute error, and weighted MSE are used to measure the difference between an image and compressed version of the image. A feature vector consisting of these distortion metrics for several different compression rates (using JPEG2000) is used to train a classifier. False alarm and missed detection rates are each about 18%.

The following schemes are designed to detect any arbitrary scheme. Instead of classifying cover images and images with LSB hiding, they discriminate between cover images and stego images with any hiding scheme, or class of hiding schemes. The underlying assumption is that cover images possess some measurable naturalness that is disrupted by adding data. In some respects this assumption lies at the heart of all steganalysis. To calibrate the features of chosen to measure “naturalness” the systems learn using some form of supervised training.

An early approach was proposed by (Avcibas et al.) to detect arbitrary hiding schemes. He design a feature set based on image quality metrics (IQM), metrics designed to mimic the human visual system (HVS). In particular they measure the difference between a received image and a filtered (weighted sum of 3×3 neighbourhood) version of the image. This is very similar in spirit to the work by (Celik et al.) except with filtering instead of compression. The key observation is that filtering an image without hidden data changes the IQMs differently than an image with hidden data. The reasoning here is that the embedding is done locally (either pixel-wise or block wise), causing localised discrepancies.

A supervised learning has been used to detect general steganalysis (Lyu et al.). Lyu et al.] use a feature set based on higher-order statistics of wavelet sub band coefficients for generic detection. The earlier work used a two-class classifier to discriminate between cover and stego images made with one specific hiding scheme. Later work however uses a one class, multiple hyper spheres, SVM classifier. The single class is trained to cluster clean cover images. Any image with a feature set falling outside of this class is classified as stego. In this way, the same classifier can be used for many different embedding schemes. The one-class cluster of feature vectors can be said to capture a “natural” image feature set. As with Avcibas et al., the general applicability leads to a performance hit in detection power compared with detectors tuned to a specific embedding scheme. However the results are acceptable for many applications.

Martin et al. attempts to directly use the notion of the naturalness of images to detect hidden data. Though they found that data hidden certainly caused shifts from the natural set, knowledge of the specific data hiding scheme provides far better detection performance. Fridrich et al. presented supervised learning method tuned to JPEG hiding schemes. The feature vector is based on a variety of statistics of both spatial and DCT values. The performance seems to improve over previous generic detection schemes by focusing on a class of hiding schemes (Kharrazi et al.).

Another steganographic scheme has been based on LSB hiding, but designed to evade the chi square test (Provos). Here, LSB hiding is done as usual (again in JPEG coefficients), but only half the available coefficients are used. The remaining coefficients are used to compensate for the hiding, by repairing the histogram to match the cover. Although the rate is lower than F5 hiding, since half the coefficients are not used, but by Fridrich et al. F5 detector, and in fact by any detector using histogram statistics. However, because the embedding is done in the block wise transform domain, there are changes in the

spatial domain at the block borders. Specifically, the change to the spatial joint statistics, i.e. the dependencies between pixels, is different than for standard JPEG compression.

Due to the success of steganalysis in detecting early schemes, new steganographic methods have been invented in an attempt to evade detection. F5 by (Westfeld) is a hiding scheme that changes the LSB of JPEG coefficients, but not by simple overwriting. By increasing and decreasing coefficients by one, the frequency equalisation noted in standard LSB hiding is avoided. That is, instead of standard LSB hiding, where an even number is either unchanged or increased by one and an odd is either unchanged or decreased by one, both odd and even numbers are increased and decreased. This method does indeed prevent detection by the 2 test.

However, (Fridrich et al.) note that although F5 hiding eliminates the characteristic “step-like” histogram of standard LSB hiding, it still changes the histogram enough to be detectable. A key element in their detection of F5 is the ability to estimate the cover histogram. As mentioned above, the 2 test only estimates the likelihood of an image being stego, providing no idea of how close it is to cover. By estimating the cover histogram, an unknown image can be compared to both an estimate of the cover, and the expected stego, and whichever is closest is chosen. Additionally, by comparing the relative position of the unknown histogram to estimates of cover and stego, an estimate of the amount of data hidden, the hiding rate can be determined. The method of estimating the cover histogram is to decompress, crop the image by 4 pixels (half a JPEG block), and recompress with the same quantisation matrix (quality level) as before.

Fridrich et al. were able to exploit these changes at the JPEG block boundaries again using a decompress crop recompress method of estimating the cover (joint) statistics; they are able to detect OutGuess and estimate the message size with reasonable accuracy. Eggers et al. suggest a method of data-mappings that preserve the first order statistics, called histogram-preserving data-mapping (HPDM). As with the method proposed by Franz, the distribution of the message is designed to match the cover, resulting in a loss of rate.

Fridrich et al. find this cropped and recompressed image is statistically very close to the original, and generalise this method to detection of other JPEG hiding schemes. Tzschoppe et al. suggest a minor modification to avoid detection: basically not hiding in perceptually significant values. Fridrich et al. propose the stochastic modulation hiding scheme designed to mimic noise expected in an image. The non-content dependent version

allows arbitrarily distributed noise to be used for carrying the message. If Gaussian noise is used, the hiding is statistically the same as spread spectrum, though with a higher rate than typical implementations. The content dependent version adapts the strength of the hiding to the image region.

An example of a detection-theoretic approach to steganalysis is (Cachin et al.). The steganalysis problem is framed as a hypothesis test between cover and stego hypotheses. Cachin suggests a bound on the Kullback-Leibler (KL) divergence (relative entropy) between the cover and stego distributions as a measure of the security between cover and stego. Another information theoretic derivation is done for a slightly different model by (Zolner et al.). They first assume that the steg analyst has access to the exact cover, and prove the intuition that this can never be made secure. They modify the model so that the detector has some, but not complete information on the cover. From this model they find constraints on conditional entropy similar to Cachin though more abstract and hence more difficult to evaluate in practice.

Westfeld et.al proposed raw image steganalysis based on the assumption that the message length should be comparable to the pixel count in the cover image. Detection theory is well developed and has been applied to a variety of fields and applications (Provos). Its key advantage for steganalysis is the availability of results prescribing optimal (error minimising) detection.

Chandramouli et al. use a detection-theoretic framework to analyse LSB detection. Guillon et al. analyse the detecting ability of QIM steganalysis, and observe that QIM hiding in a uniformly distributed cover does not change the statistics. Since typical cover data is not in fact uniformly distributed, they suggest using a non-linear compressor to convert the cover data to a uniformly distributed intermediate cover. The data is hidden into this intermediate cover with standard QIM, and then the inverse of the function is used to convert to final stego data. Farid explained about the usage of higher order statistics for generic steganalysis techniques and the first order statistics for the specific steganalysis techniques. Fridrich explained a technique for estimating the unaltered histogram to find the number of changes and length of secret message.

Sidorov presented work done on using hidden Markov model (HMM) theory for the study of steganalysis. He presents analysis on using Markov chain and Markov random field models, specifically for detection of LSB. Though the framework has great potential, the

results reported are sparse. He found that a Markov chain (MC) model provided poor results for LSB hiding in all but high-quality or synthetic images, and suggested a Markov random field (MRF) model, citing the effectiveness of the RS/sample pair scheme.

Sallee proposed a means of evading optimal detection. The basic idea is to create stego data with the same distribution model as the cover data. That is, rather than attempting to mimic the exact cover distribution, mimic a parameterised model. The justification for this is that the steg analyst does not have access to the original cover distribution, but must instead use a model. A specific method for hiding in JPEG coefficients using a Cauchy distribution model is proposed.

Detection theory to steganalysis is Hogan et al. QIM (quantisation index modulation) steganalysis. Hernandez et al. proposed a global steganalysis methodology by comparing some of the steganalysis methods. Using stego images generated by typical data hiding algorithms, the secret message detection capacities of these steganalysis methods are evaluated. The evaluation of steganalysis methods is represented in terms of false negative and false positive error rates using 100 images. Chao et al. proposed a method based on the good property of fractional Fourier transform (FRFT) coefficients of image histogram for extracting two kinds of features of an image. SVM is used as a classifier.

Mei et al. introduced an alpha-trimmed method as an image estimation technique for distinguishing cover and stego images. This method estimates steganographic messages within images in the spatial domain that provides flexibility for classifying various steganalysis methods in the JPEG compression domain. Wang et al. used a new kind of transition probability matrix is constructed to describe correlations of the quantised DCT coefficients in the multi-directions. Subsequently, 96-dimensional feature vector is extracted by merging two different calibrations. SVM is trained to build the steg analyser.

Zhiping Zhou et al. developed zigzag scanning pattern to arrange both DCT blocks and coefficients in each block. The computational complexity of the proposed method is manageable with the help of Threshold and truncation techniques. Bidirectional Markov matrix is exploited to capture the correlations between the adjacent coefficients in both intra-block and inter-block senses, which have been changed during data embedding. Features for steganalysis are derived from intra-block and inter-block Markov transition matrixes.

Qian-lan et al. proposed an image steganalysis scheme based on the differential image histogram in frequency domain. The difference is calculated in three directions, horizontal, vertical and diagonal towards adjacent pixels to obtain three-directional differential images for a natural image. The features for steganalysis are extracted from the DFT of the histogram of differential images and divided into low and high frequency bands. SVM with RBF kernel is applied as classifier.

Xiaoyuan et al. used Wavelet based Markov Chain (WBMC) model for nature images. It presents statistic divergence between cover image and steg image prominently. Based on Markov chain empirical matrix, the difference between low frequency domain and high frequency domain generalised by steg process is discussed. It also defined two models: WBMC_L model and WBMC_H model respective to construct WBMC model. Wenqiong et al. constructed nine statistical models from the DCT and decompressed spatial domain for a JPEG image. Feature set is measured by calculating the histogram characteristic function (HCF) and the centre of mass (COM). SVM are used as classifiers.

Seongho Cho et al. classify the image blocks into multiple classes on steganalysis that provides decomposed image blocks. Also it uses a classifier for each class to decide whether a block is from a cover or stego image. Consequently, the steganalysis of the whole image can be performed by fusing steganalysis. Jingwei Wang et al. design a multi-classifier which classifies stego images depending on their steganographic algorithms. Based on steganalysis results of decomposed image blocks stego image is distinguished from cover images.

Yamini et al. calculated the length of embedded message using SVM as a classifier. Zhi-Min et al. [138] proposed a RBF Neural Network (RBFNN) optimised by the Localised Generalisation Error Model (L-GEM) for steganography detection. Discrete cosine transforms (DCT) features and the Markov features are given as inputs of neural networks. They enhance the generalisation capability of the RBFNN and the performance of detecting steganalysis in future images. The architecture of the RBFNN is selected by minimising the L-GEM.

Ramezani et al. compared Fisher linear discriminant (FLD), Gaussian naïve Bayes, multilayer perceptron, and k nearest neighbour for steganalysis of suspicious images. The method exploits statistics of the histogram, wavelet statistics, and amplitudes of local extrema from the 1D and 2D adjacency histograms, centre of mass of the histogram

characteristic function and co-occurrence matrices for feature extraction process. In order to reduce the proposed features dimension and select the best subset, genetic algorithm is used and the results are compared through principle component analysis and linear discriminant analysis.

Gireesh Kumar et al. compared the efficiency of two embedding algorithms using the image features that are consistent over a wide range of cover images, but are distributed by the presence of embedded data. Image features were extracted after wavelet decomposition of the given image. These features were then given to a SVM classifier to identify. Holoska et al. compared universal neural network classification and a linear classification tool (Steg detect). Based on the results it is concluded that neural networks were better than the linear classification tool. Sheikhan et al. extracted the features from Contourlet coefficients and co-occurrence metrics of sub band images. Analysis of Variance (ANOVA) method is used to reduce the number of features. The selected features are fed to nonlinear SVM for classification.

Ke Ke et al. explore Bhattacharya Distance principle to recognise stego algorithms that are being used. The most important features are selected by the means of applying Bhattacharya distance. BPA is used to classify cover and stego images. Chen Qunjie et al. proposed a steganographic detection method for JPEG image which is based on the data-dependent concept. The initial classifier is obtained by SVM training. Then the kernel function is modified with conformal transformation by using the information of Support Vectors and retrain with the new kernel to enlarge the spacing around classification boundary. Repeat this until the best result is obtained.

Li Hui et al. proposed the scheme based on the characteristic function (CF) moments of three-level wavelet sub bands as well as the further decomposition coefficients of the first scale diagonal sub band. The first three statistical moments of each wavelet band of test image and prediction-error image are selected to form 102 dimensional features for steganalysis. Principal Components Analysis (PCA) is utilised to reduce the features. SVM is adopted as the classifier.

Ping et al. proposed a novel method for universal steganalysis on frequency domain to detect hidden message. The detection is achieved based on the spectrum analysis of difference histogram of frequency coefficients according to evident spectrum difference between cover images and stego images. Experimental results from detecting

steganographic images of DCT domain and DWT domain show that the detection performance is satisfied.

Although there are some techniques that can detect steganography there are major problems that steg analysts face. Even if there are noticeable distortions and noise, predictable patterns cannot always be detected. Some steganographic techniques are particularly difficult to detect without the original image. And in most cases, it is highly unlikely that a forensic investigator will be conveniently presented with the steganographic and original image. Even until today, most steganalysis techniques are based on visual attacks and methods beyond this are being explored. Unfortunately a general steganalysis technique has not been devised (Johnson et al.).

While visual attacks are more prominent, JPEG images, which is one of the most commonly distributed type of image format; the steganographic modifications take place in the frequency domain. This means that this type of steganography is not susceptible to visual attacks unlike in image formats such as GIF images where the modifications happen in the spatial domain Provos et al.; Niel Provos et al. created a cluster that scans images from newsgroups to detect steganographic content in order to verify the claims about terrorists with the help of Internet to distribute secrets using steganography. For reasons that no hidden messages were discovered, it raises the question of the practicality of such detection systems (Krenn).

Xiaochuan Chen et al. used statistical analysis of empirical matrix (EM) to detect the hidden message in an image. With the help of projection histogram of EM, moments of PH and the moments of the characteristic function of PH features are extracted. To enhance the performance, features extracted from prediction-error image are also included. SVM is used as a classifier.

Yuan Liu et al. proposed three methods for deriving the feature vector such as Robert gradient energy in pixel domain, variance of Laplacian parameter in DCT domain and higher-order statistics extracted from wavelet coefficients. BPA neural network is applied as the classifier.

Xiangyang Luo et al. used WPT to decompose image into three scales and obtained 85 coefficient sub bands together. Multi-order absolute characteristic function moments of histogram are extracted from these sub bands as features. Finally these features are

normalised and combined to a 255-D feature vector for each image. Back propagation neural network is used as a classifier.

Yuan -Tu et al. proposed a method for feature extraction by calculating the features from the luminance and chrominance components of the images. Features are extracted both in DCT and DWT domains. Wavelet high-order statistics is substituted with the moments of wavelet characteristic function. Non-linear SVM classification is implemented.

Jing-Qu Lin et al. proposed Binary Similarity Method (BSM) for capturing the seventh and eighth bit planes of the non-zero DCT coefficients from JPEG images and 14 features of each image are computed. SVM is used as a classifier. Zhi-Min He et al. used RBFNN for steganalysis. DCT features and the Markov features are used as inputs of neural networks.

Sheikhan et al. proposed a method for extracting features from Contourlet coefficients and co-occurrence metrics of sub band images. Analysis of Variance (ANOVA) method is used and hence the number of features is reduced. Non-linear SVM is used as a classifier. Lie et al used the gradient energy and statistical variance as two features for detecting the presence of hidden messages in spatial or DCT domain. Shi et al. proposed a method that uses statistical moments of characteristic functions of the prediction-error image, the test image, and their wavelet sub bands as selected features. ANN issued as classifier.

CHAPTER III: AIM and OBJECTIVES

Aim:

To embed the text message in a white space of a text file using SNOW forensic tool.

Objectives:

- To install the application of the SNOW.
- To embed a message in a whitespace of a text file.
- To retrieve the embedded message from a particular text file.

CHAPTER IV: METHODOLOGY

WhiteSteg is proposed for text steganography by creating a hybrid method in utilising whitespaces between words and paragraphs. This method could be an improvement of open space method because it is not solely using a method of encoding data as what has been mentioned in. By integrating both methods which are inter-word spacing and inter-paragraph spacing into an embedding algorithm, a larger capacity for embedding hidden bits is provided. Text Steganography is defined as Random & statistical generation Linguistic method Format-based. The proposed scheme is inspired by Bender's open space method and also a non-commercial used program namely SNOW by Matthew Kwan. Instead of using one method for every embedding mechanism, we propose to create a hybrid method in manipulation of whitespaces so that it is able to hide the secret bits in a dynamic generated cover text to produce a seemingly innocent stego-text.

STEGANOGRAPHY TYPES

Text Steganography: It comprises of concealing information inside the text documents. In this strategy, the mystery information is taken cover behind each nth letter of each expression of text message. Quantities of methods are accessible for concealing information in text record. These methods are i) Format Based Method; ii) Random and Statistical Method; iii) Linguistics Method.

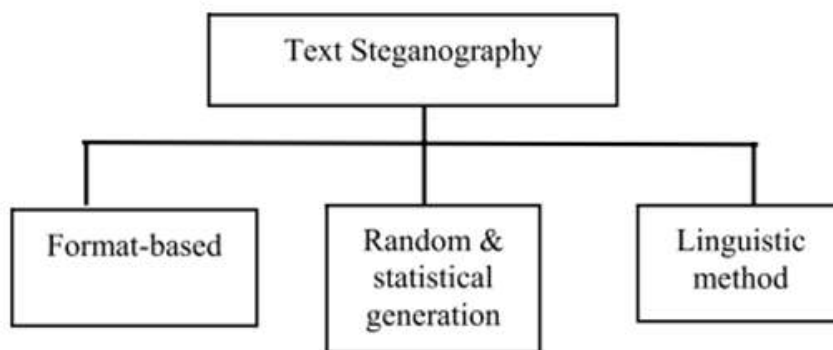


Fig 3: Three Basic Categories of Text Steganography

Format-based methods use physical text formatting of text as a place in which to hide information. Generally, this method modifies existing text in order to hide the steganographic text. Insertion of spaces in between words or end of the sentence, deliberate misspellings and resizing of the fonts throughout the text are some of the many format-

based methods being used in text steganography. However, Bennett has stated that those format-based methods could not be seen with the human visual system but it is possible to detect with the computer system.

Random and statistical generation is generating cover text according to the statistical properties. This method is based on character sequences and word sequences. The hiding of information within character sequences is embedding the information to be appeared in random sequence of characters. This sequence must appear to be random to anyone who intercepts the message. A second approach to character generation is to take the statistical properties of word-length and letter frequency in order to create “words” (without lexical value) which will appear to have the same statistical properties as actual words in a given language. The hiding of information within word sequences, the actual dictionary items can be used to encode one or more bits of information per word using a code-book of mappings between lexical items and bit sequences, or words themselves can encode the hidden information.

The final category is **linguistic method** which specifically considers the linguistic properties of generated and modified text, frequently uses linguistic structure as a place for hidden messages. In fact, steganographic data can be hidden within the syntactic structure itself.

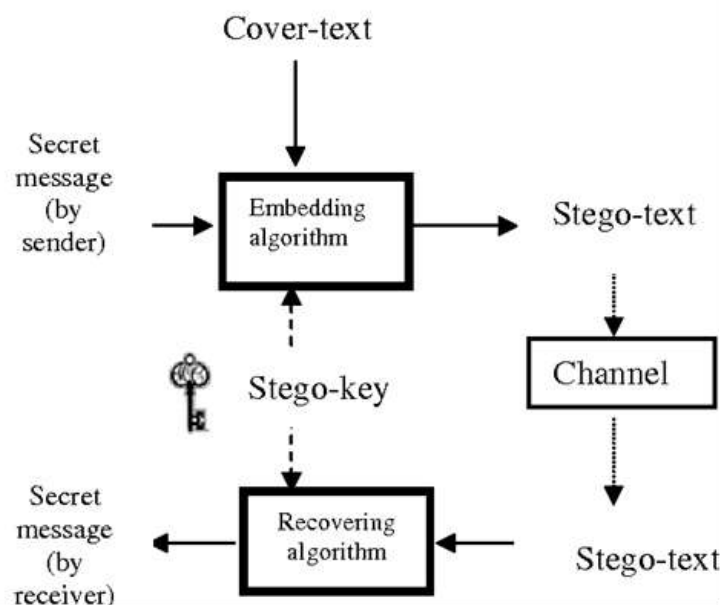


Fig 4: Mechanism of Text Steganography

Image Steganography: Hiding the information by taking the spread information as image is known as image steganography. In image steganography pixel powers are utilised to shroud the information. In computerised steganography, images are generally utilised spread source on the grounds that there are number of bits shows in advanced portrayal of an image.

Audio Steganography: It includes concealing information in audio records. This technique conceals the information in WAV and MP3 sound records. There are diverse methods of audio steganography. These methods are i) Low Bit Encoding ii) Phase Coding iii) Spread Spectrum.

Video Steganography: It is a system of concealing any sort of records or information into advanced video format. For this situation video (blend of pictures) is utilised as bearer for concealing the information. For the most part discrete cosine change (DCT) adjusts the qualities (e.g., 8.667 to 9) which are utilised to shroud the information in every one of the images in the video, which is unnoticeable by the human eye. Mp4, MPEG, AVI are the formats utilised by video steganography.

Network or Protocol Steganography: It includes concealing the information by taking the network convention, for example, TCP, UDP, ICMP, IP and so on, as spread item. In the OSI layer network display there exist undercover channels where steganography can be used.

APPLICATIONS OF STEGANOGRAPHY

Secret Communications the utilisation of steganography does not promote secret correspondence and along these lines keeps away from examination of the sender side, message, and beneficiary. A secret, outline, or other delicate information can be transmitted without cautioning potential aggressors. Feature Tagging Elements can be inserted inside an image, as the names of people in a photograph or areas in a guide. Copy the stego-image likewise duplicates the majority of the installed features and just gatherings that have the disentangling stego-key will most likely concentrate and view the features. Copyright Protection, Copy protection components that avert information, for the most part computerised information, from being copied.

Confidential Communication & Secret Data Storing: The "secrecy" of the embedded data is essential in this area. Historically, steganography have been addressed in

this area. Steganography provides us with: (A) Potential capability to hide the existence of confidential data. (B) Hardness of detecting the hidden (i.e., embedded) data. (C) Enhancing the secrecy of the encrypted data.

In practice, when someone use steganography, first select a vessel data according to the size of the embedding data. The vessel should be innocuous. Then, embed the confidential data by using an embedding program (which is one component of the steganography software) together with some key. When extracting on the other side use an extracting program (another component) to restore the embedded data by the same key ("common key" in terms of cryptography). In this case a "key negotiation" is required with other side before starting confidential communication.

Protection of Data Alteration: The take advantage of the fragility of the embedded data in this application area. "The embedded data can rather be fragile than be very robust." Actually, embedded data are fragile in most steganography programs. Especially, Qtech Hide & View program embeds data in an extremely fragile manner. However, this fragility opens a new direction toward an information-alteration protective system such as a "Digital Certificate Document System." The most novel point among others is that "no authentication bureau is needed." If it is implemented, people can send their "digital certificate data" to any place in the world through Internet. No one can forge, alter, nor tamper such certificate data. If forged, altered, or tampered, it is easily detected by the extraction program.

Access Control System for Digital Contents Distribution: In this area embedded data is "hidden", but is "explained" to publicise the content. Today, digital contents are getting more and more commonly distributed over Internet than before. For example, music companies release new albums on their Webpage in a free or charged manner. However, in this case, all the contents are equally distributed to the people who can make access to the Webpage. So, an ordinary Web distribution scheme is not suited for a "case-by-case" and "selective" distribution. Of course it is always possible to attach digital contents to e-mail messages and send them to the customers. But it will take a lot of cost in time and labor.

If there is some valuable content which is distributable and who really needs it, then it is possible to upload that content on internet in some covert manner. And issue a special "access key" to extract the content selectively, then it will be very happy about it. A steganographic scheme can help realise this type of system. Develop a prototype of an

"Access Control System" for digital content distribution through Internet. The following steps explain the scheme. (i) A content owner classify his/her digital contents in a folder-by-folder manner, and embed the whole folders in some large vessel according to a steganographic method using folder access keys, and upload the embedded vessel (stego data) on his/her own Webpage. (ii) On that Webpage the owner explains the contents in depth and publicise worldwide. The contact information to the owner (post mail address, e-mail address, phone number, etc.) will be posted there. (iii) The owner may receive an access-request from a customer who watched that Webpage. In that case, the owner may (or may not) creates an access key and provide it to the customer (free or charged). In this mechanism the most important point is, a "selective extraction" is possible or not. In fact it is already developed such a selective extraction program to implement the system. However, we will not go down to the details about this.

Media Database Systems: In this application area of steganography secrecy is not important, but unifying two types of data into one is the most important. Media data (photo picture, movie, music, etc.) have some association with other information. A photo picture, for instance, may have the following. (1) The title of the picture and some physical object information. (2) The date and the time when the picture was taken. (3) The camera and the photographer's information. Formerly, these are annotated beside the each picture in the album.

Recently, almost all cameras are digitalised. They are cheap in price, easy to use, quick to shoot. They eventually made people feel reluctant to work on annotating each picture. Now, most homes PC's are stuck with the huge amount of photo files. In this situation it is very hard to find a specific shot in the piles of pictures. "Photo album software" may help a little. Sort the pictures and put a couple of annotation words to each photo. When required to find a specific picture, then make a search by keywords for the target picture. However, the annotation data in such software are not unified with the target pictures. Each annotation only has a link to the picture. Therefore, when the pictures are transferred to different album software, all the annotation data are lost.

This problem is technically referred to as "Metadata (e.g., annotation data) in a media database system (photo album software) are separated from the media data (photo data) in the database managing system (DBMS)." This is a big problem. Steganography can solve this problem because a steganography program unifies two types of data into one by way of embedding operation. So, metadata can easily be transferred from one system to

another without hitch. Specifically, it can embed all your good/bad memory (of your sight-seeing trip) in each snap shot of the digital photo. It can either send the embedded picture to other friend to extract memory on his/her PC, or keep it silent in own PC to enjoy extracting the memory ten years after. Q tech Hide & View v02 may be a good program for such purposes. If a "motion picture steganography system" has been developed in the near future, a keyword based movie-scene retrieving system will be implemented. It will be a step to a "semantic movie retrieval system."

STEGANOGRAPHY TECHNIQUES

In audio steganography, secret message is inserted into digitised audio flag bringing about slight modification of double arrangement of the comparing audio document. There are different methods are accessible for audio steganography.

LSB Coding: In LSB both the spread record and the secret message will be changed over into their constituent parallel format. At that point the LSB of certain bytes of secured document will be supplanted with the grouping of bytes secret message. Generally the privilege most pieces are considered as LSB as it has minimal effect over the nature of spread file.

Parity Coding: In equality coding the equality bit of cover file is checked and in the event that they are same, at that point do nothing and on the off chance that they are unique, at that point it changes the LSB of anyone (cover record or secret message) to make the equality equal.

Phase Coding: In stage encoding the period of an underlying audio fragment is substituted with a reference stage that speaks to the shrouded information. It encodes the secret message bits as stage moves in the stage range of an advanced flag, accomplishing an imperceptible encoding as far as flag to-clam or ratio.

Echo Data Hiding: In reverberation information concealing the secret information is embedded by adding reverberation to the spread audio file. Data covering up is communicated by three varieties of the parameters, rot rate, sufficiency beginning e, and delay. The introductory abundance is helpful to decide unique information sound plentifulness. Rot rate is utilised to decide the reverberation capacity to be made. The balance work is utilised to decide the separation between the first discourse signals with the reverberation that has been made.

Installing Extraction Video Steganography: In video steganography, video signals are utilised to shroud secret information. The goal is to conceal substantial measure of secret information in video files. In this strategy, AVI record is utilised as bearer. Video documents containing audio are isolated into video and audio outlines. Video outlines are as images, and consequently image steganography is utilised on video outlines. At the point when audio is isolated from or separated from video documents, it resembles an audio record and henceforth audio steganography is utilised on audio documents. Since both audio and video outlines utilised as bearer, limit of steganography is expanded. The secret information can be image and audio or text. In this technique, secret image and audio signals are covered up in the video documents. Favourable position of this strategy is its strength. It opposes activities, for example, sifting, trimming, turn and pressure. The concealed information isn't distinguished by outsider, consequently the framework is secure.

CHAPTER V: IMPLEMENTATION

The entire process that is used by proposed technique for hiding the data within text file using two main functions that are called embedding and extraction are applied on the .txt file. In fact, the main embedding approach is based on modifying the whitespaces between the characters for hiding the secret message characters that are mapped into binary format, which are represented by whitespace. In this way, since a common key between the sender and recipient is used to shuffle the place of whitespace in each embedding as well as providing different character-binary mapping, it can be more difficult for attacker to guess the hidden data characters. In the following, two embedding and extraction functions are explained in details.

Embedding: First of all, in order to embed the secret message, it is necessary to read the cover text. In this way, several useful information regarding to the selected text file such as the number of words, number of whitespaces and file size will be retrieved from the file. In fact, the attained information is important as they help to determine if the secret message can be embedded within the file or not. Since the whitespace between the words in line is the main approach, the number of characters that can be used for hiding the data will be computed.

Secondly, the secret message that must be hidden within the selected file is given to the technique, but three conditions are required to be considered. Firstly, it is assumed that the secret message can be only English alphabets in this study. Secondly, the secret message must be in the form of compressed message to save more whitespaces for hiding. In this case, if a user mentions any whitespaces accidentally as part of the hidden data, the developed software will remove it automatically and concatenate all the characters together (compress message). Lastly, since there is not a difference between upper case and lower case characters in term of meaning, the compressed message is considered to be in lowercase format.

Extraction: On the other hand, the recipient needs to do the same procedure but in a reverse way in order to extract the secret message from the stego text. In this case, he needs to read the stego text. Furthermore, since using first five whitespace in each line is known by both parties through developed software, the correct key is needed to be chosen in order to decode the hidden data from the stego file. It is important to highlight that if the key will be chosen differently, the result will be wrong, which provides random characters.

Therefore, the key is shared between two parties who transmitted it in prior using different channels such as the Internet in a secure manner. Lastly, once the binary values, based on the modified whitespaces are generated, they will be converted to some values, which are mapped to the proper characters, based on the common key to retrieve the hidden data successfully.

The entire process that is going to be implemented by using the application named as SNOW which is also called as Steganographic Nature Of Whitespace known for its Whitespace steganography. Snow is a free for non-commercial use program available in internet and is authored by Matthew Kwan. According to Bender, et.al, “soft-copy text is in many ways the most difficult place to hide data.” This is due to the fact that extra characters or punctuation in a document can easily be noticed by the reader. The Snow program utilises an open space method of embedding data in a text file – it adjusts the trailing spaces of each line of text to encode its message. The author of Snow describes it as “a program for concealing messages in text files by appending tabs and spaces on the end of lines, and for extracting messages from files containing hidden messages” (Kwan). Snow relies on the fact that most text viewers and editors do not by default show these characters when viewing, thus masking the fact that there is extra data in the file. Snow is run from the windows command line and is available for Unix/Linux platforms, and there is a java version available as well.

Snow is a program for concealing messages in text files by appending tabs and spaces on the end of lines, and for extracting messages from files containing hidden messages. Tabs and spaces are invisible to most text viewers, hence the steganographic nature of this encoding scheme. The data is concealed in the text file by appending sequences of up to 7 spaces, interspersed with tabs. This usually allows 3 bits to be stored every 8 columns. An alternative encoding scheme, using alternating spaces and tabs to represent zeroes and ones, was rejected because, although it used fewer bytes, it required more columns per bit (4.5 vs 2.67).

The start of the data is indicated by an appended tab character, which allows the insertion of mail and news headers without corrupting the data. Snow provides rudimentary compression, using Huffman tables optimized for English text. However, if the data is not text, or if there is a lot of data, the use of the built-in compression is not recommended, since an external compression program such as compress or gzip will do a much better job.

Encryption is also provided, using the ICE encryption algorithm in 1-bit cipher-feedback (CFB) mode. Because of ICE's arbitrary key size, passwords of any length up to 1170 characters are supported (since only 7 bits of each character are used, keys up to 1024-bytes are supported). If a message string or message file are specified on the command-line, snow will attempt to conceal the message in the file if specified, or standard input otherwise.

The resulting file will be written to out file if specified, or standard output if not. If no message string is provided, snow attempts to extract a message from the input file. The result is written to the output file or standard output.

OPTIONS

-C Compress the data if concealing, or uncompressed it if extracting.

-Q Quiet mode. If not set, the program reports statistics such as compression percentages and amount of available storage space used.

-S Report on the approximate amount of space available for hidden message in the text file. Line length is taken into account, but other options are ignored.

-p password

If this is set, the data will be encrypted with this password during concealment, or decrypted during extraction.

-l line -len

When appending whitespace, snow will always produce lines shorter than this value. By default it is set to 80.

-f message-file

The contents of this file will be concealed in the input text file.

-m message-string

The contents of this string will be concealed in the input text file. Note that, unless a newline is somehow included in the string, a newline will not be printed when the message is extracted.

Advantages of steganography

It is used in the way of hiding not the information but the password to reach that information.

Difficult to detect. Only receiver can detect.

Can be applied differently in digital image, audio and video file.

It can be done faster with the large number of soft wares.

- The advantage of steganography over cryptography is that message doesn't attract the attention to them.

Disadvantages of steganography

Huge number of data, huge files size, so someone can suspect about it.

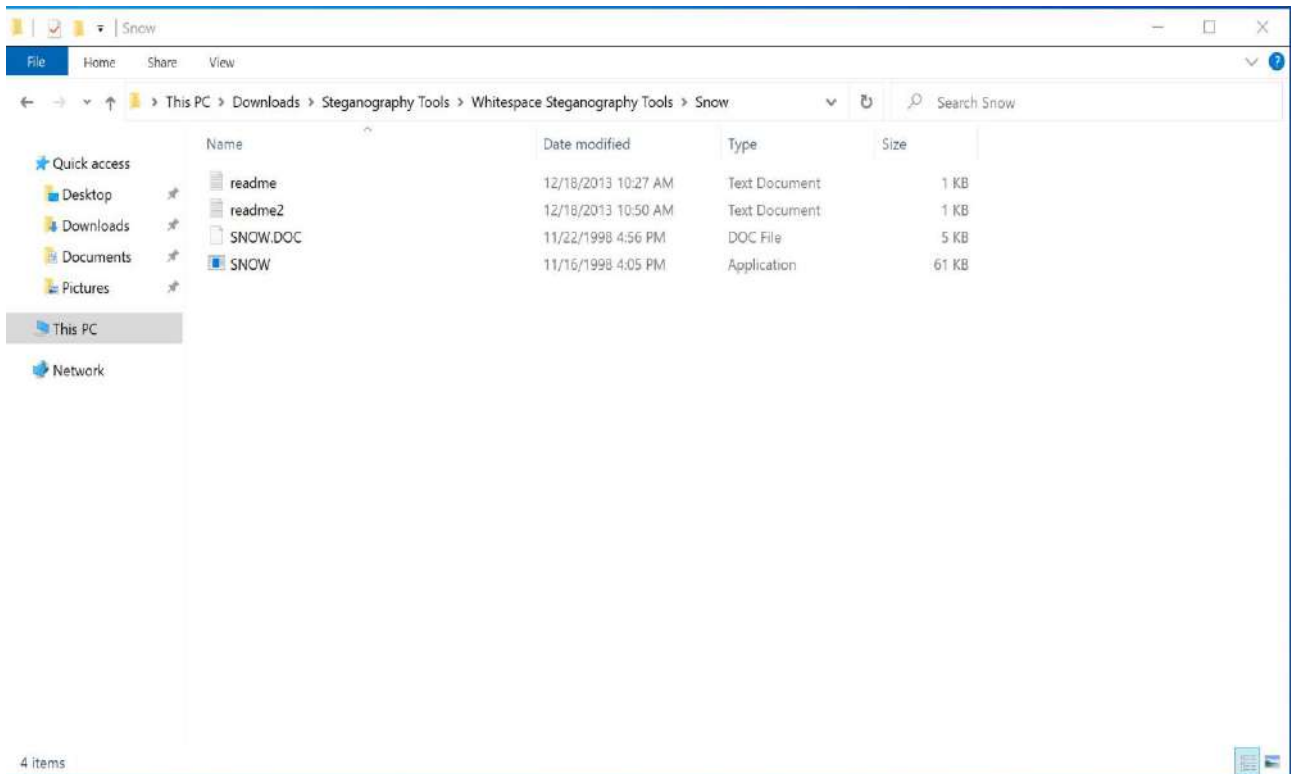
If this technique is gone in wrong hands like hackers, terrorists, criminals then this can be very much dangerous for all.

The confidentiality of information is maintained by the algorithms, and if the algorithms are known then this technique is of no use.

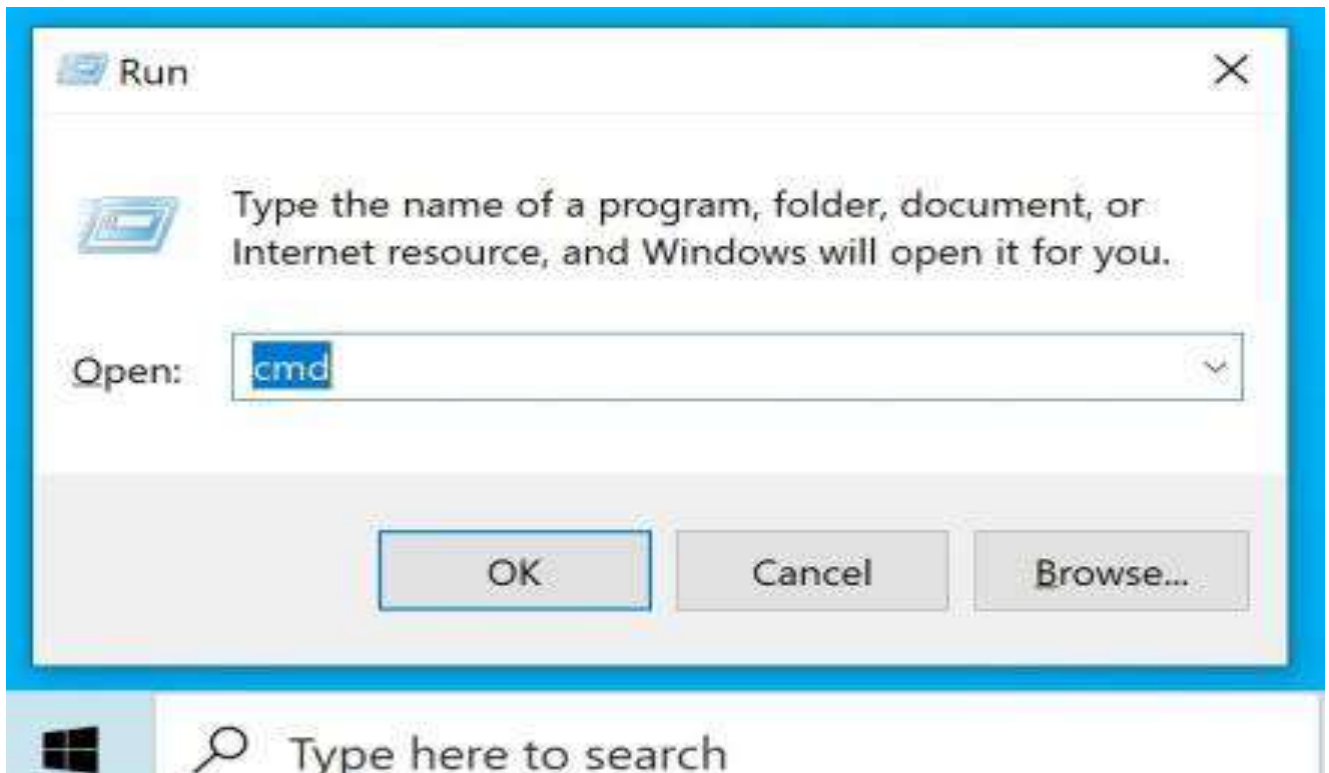
Password leakage may occur and it leads to the unauthorised access of data.

Note: *The purpose of steganography is covert communication to hide a message from a third party. This differs from cryptography, the art of secret writing, which is intended to make a message unreadable by a third party but does not hide the existence of the secret communication.*

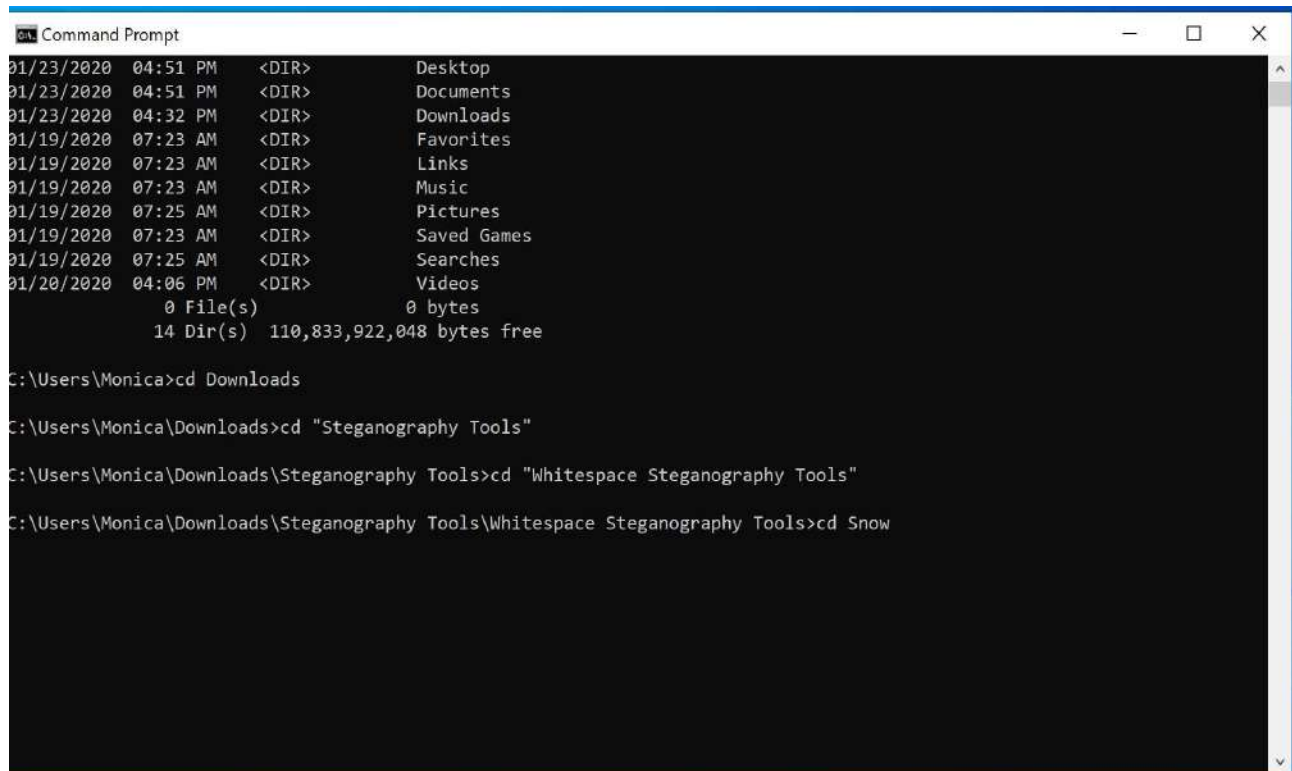
Step: 1 Firstly point to the location of the application in the computer.



Step: 2 Press Window + R to open run and type cmd to open command prompt.



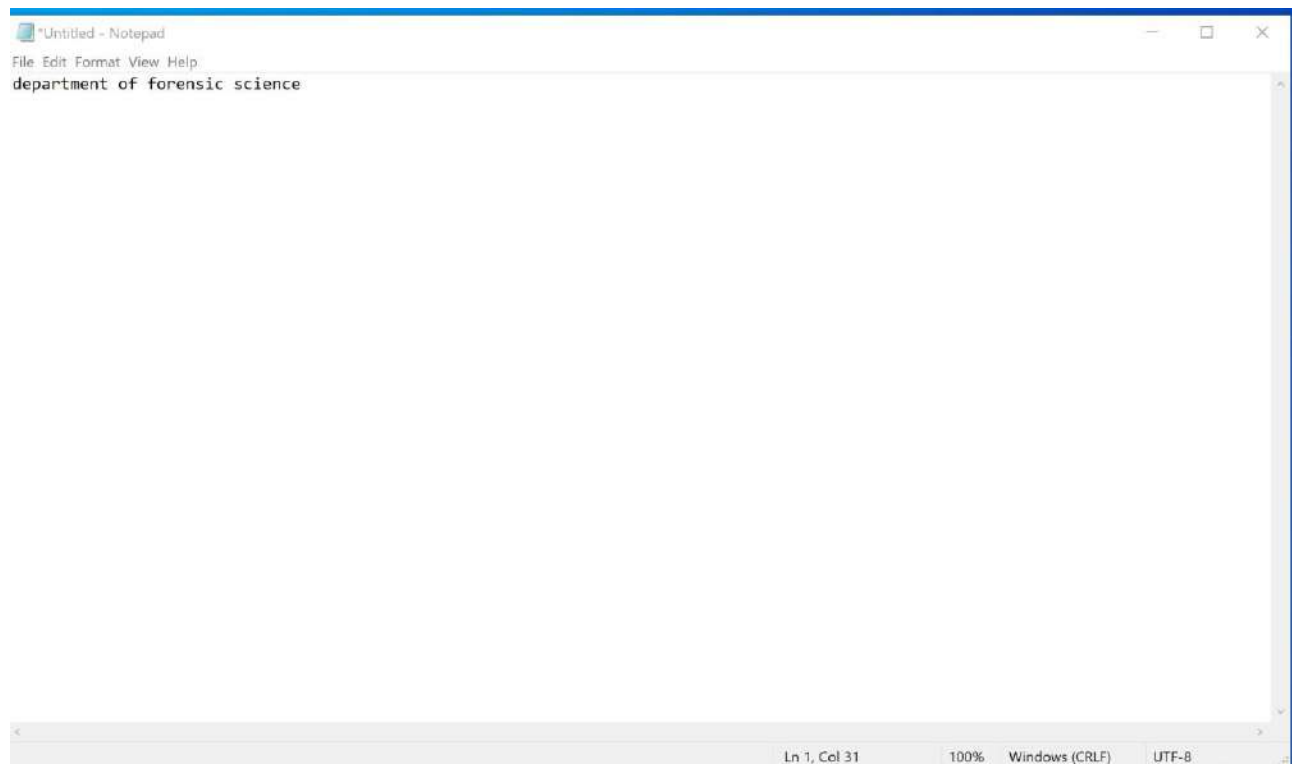
Step: 3 In command prompt goto the path where SNOW application is located.
(C:\Users\Monica\Downloads\Steganography Tools\Whitespace Steganography Tools\Snow)



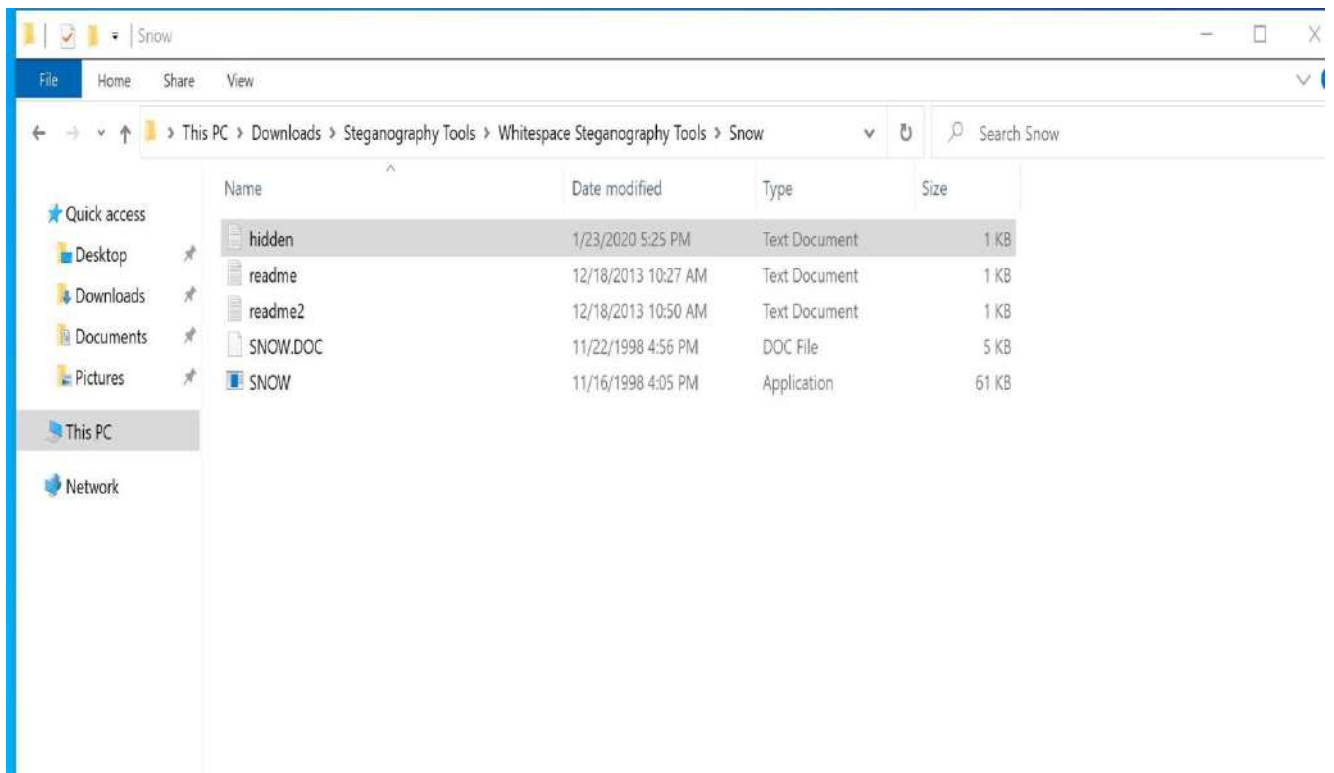
```
01/23/2020 04:51 PM <DIR> Desktop
01/23/2020 04:51 PM <DIR> Documents
01/23/2020 04:32 PM <DIR> Downloads
01/19/2020 07:23 AM <DIR> Favorites
01/19/2020 07:23 AM <DIR> Links
01/19/2020 07:23 AM <DIR> Music
01/19/2020 07:25 AM <DIR> Pictures
01/19/2020 07:23 AM <DIR> Saved Games
01/19/2020 07:25 AM <DIR> Searches
01/20/2020 04:06 PM <DIR> Videos
0 File(s) 0 bytes
14 Dir(s) 110,833,922,048 bytes free

C:\Users\Monica>cd Downloads
C:\Users\Monica\Downloads>cd "Steganography Tools"
C:\Users\Monica\Downloads\Steganography Tools>cd "Whitespace Steganography Tools"
C:\Users\Monica\Downloads\Steganography Tools\Whitespace Steganography Tools>cd Snow
```

Step: 4 Again open run and open notepad and create a text file and save in the same path of application.

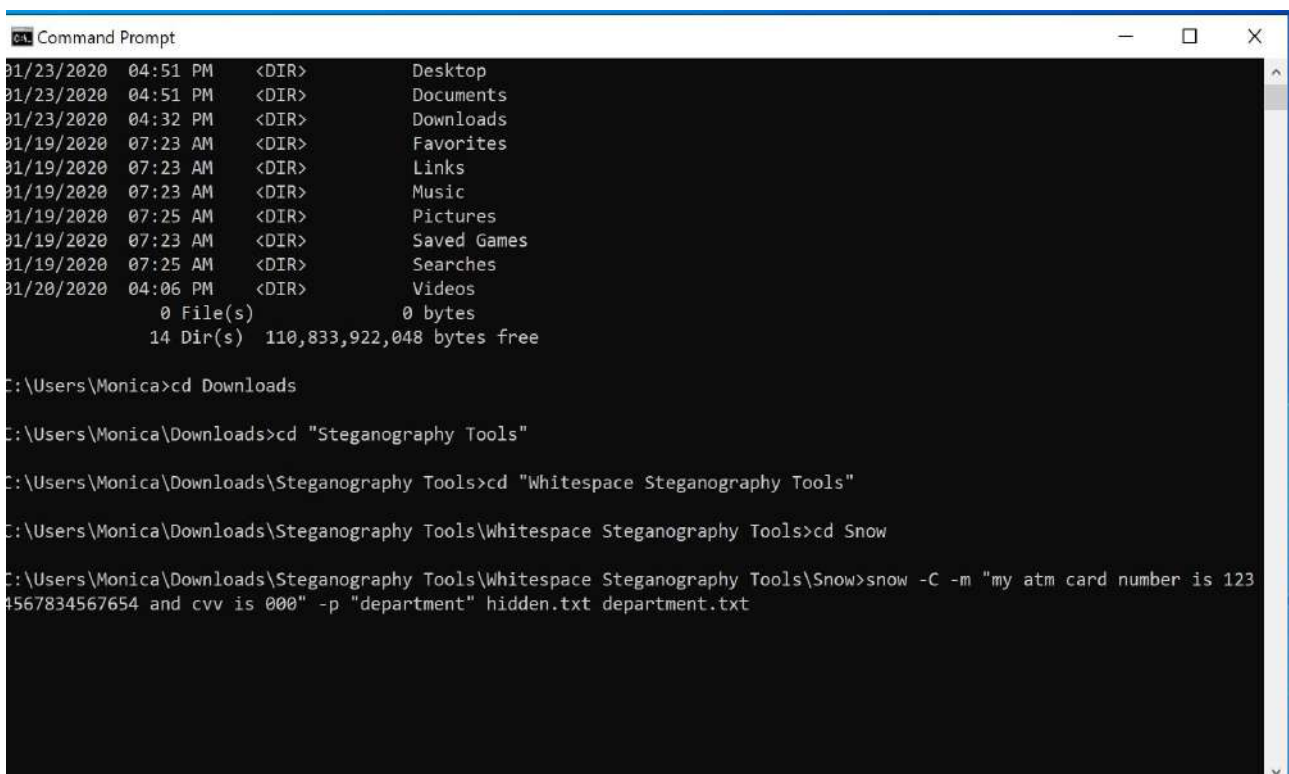


```
department of forensic science
```



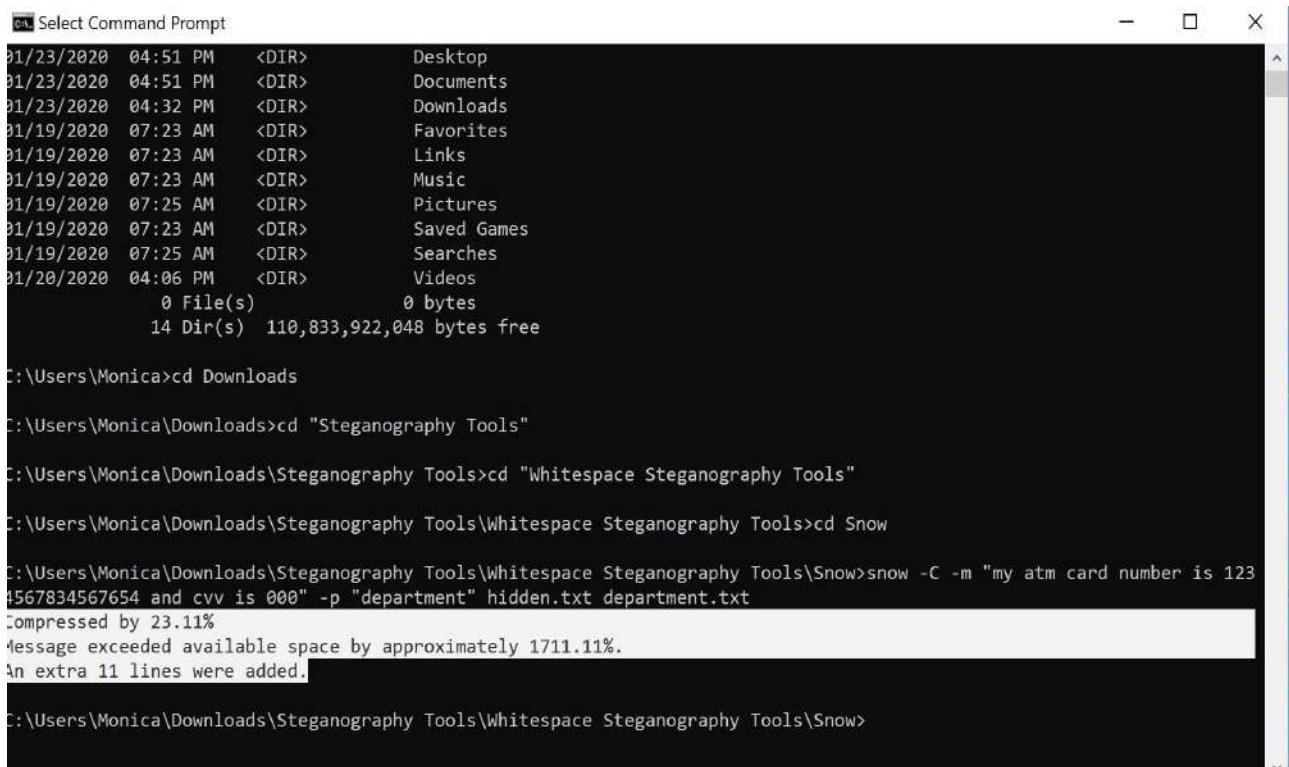
Step: 5 Now in command prompt enter the following command to embed the text into hidden.txt.

```
Snow -C -m "my atm card number is 1234567834567654 and cvv is 000" -p "department"
hidden.txt department.txt
```



Step: 5a

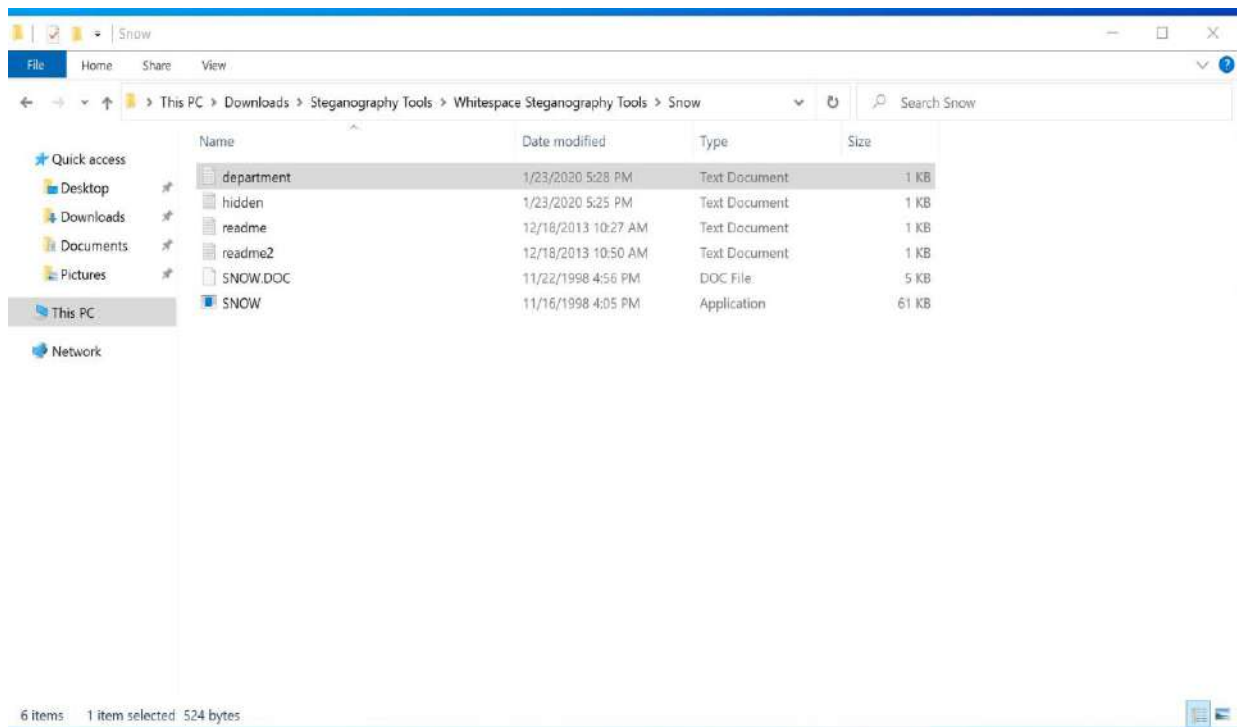
A prompt of the amount of embedded text details will be shown on successful message concealing.



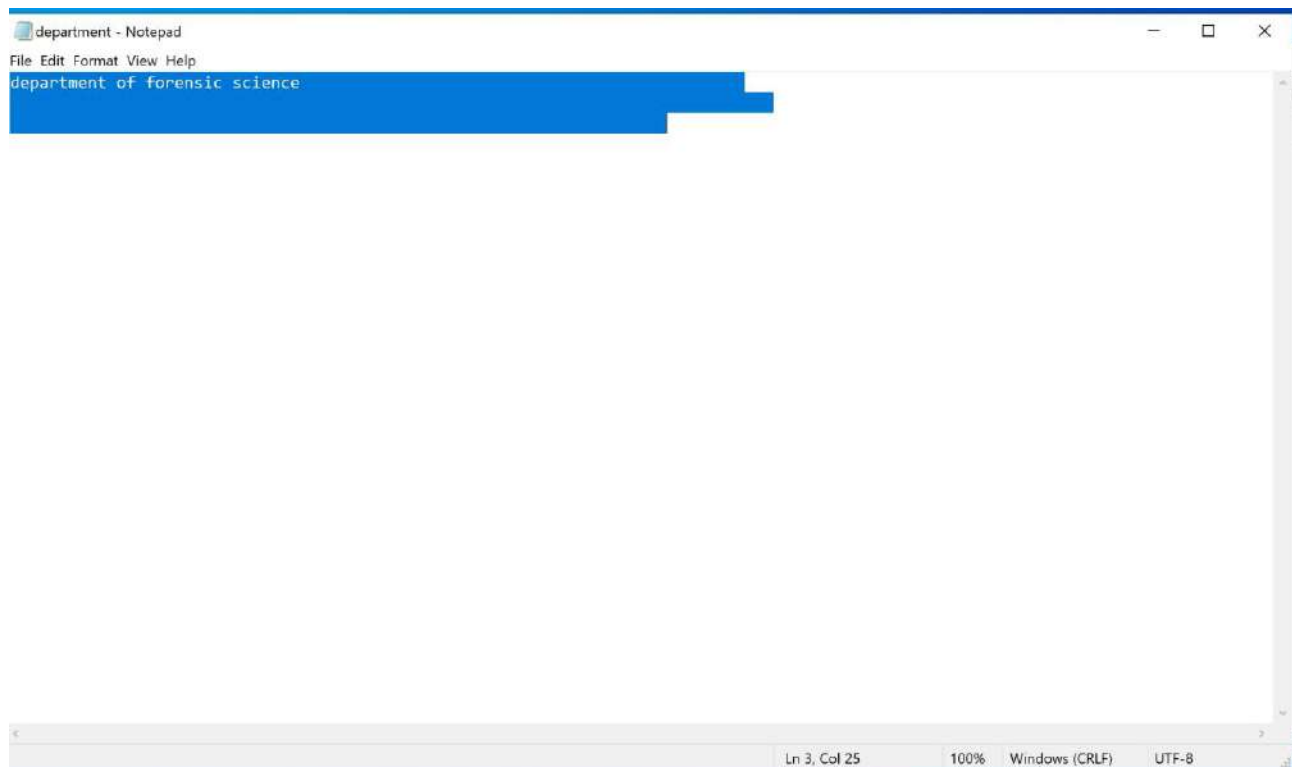
```
01/23/2020 04:51 PM <DIR> Desktop
01/23/2020 04:51 PM <DIR> Documents
01/23/2020 04:32 PM <DIR> Downloads
01/19/2020 07:23 AM <DIR> Favorites
01/19/2020 07:23 AM <DIR> Links
01/19/2020 07:23 AM <DIR> Music
01/19/2020 07:25 AM <DIR> Pictures
01/19/2020 07:23 AM <DIR> Saved Games
01/19/2020 07:25 AM <DIR> Searches
01/20/2020 04:06 PM <DIR> Videos
0 File(s) 0 bytes
14 Dir(s) 110,833,922,048 bytes free

C:\Users\Monica>cd Downloads
C:\Users\Monica\Downloads>cd "Steganography Tools"
C:\Users\Monica\Downloads\Steganography Tools>cd "Whitespace Steganography Tools"
C:\Users\Monica\Downloads\Steganography Tools\Whitespace Steganography Tools>cd Snow
C:\Users\Monica\Downloads\Steganography Tools\Whitespace Steganography Tools\Snow>snow -C -m "my atm card number is 123
4567834567654 and cvv is 000" -p "department" hidden.txt department.txt
Compressed by 23.11%
Message exceeded available space by approximately 1711.11%.
An extra 11 lines were added.
C:\Users\Monica\Downloads\Steganography Tools\Whitespace Steganography Tools\Snow>
```

Step: 6 A new text file with name department will be saved at same location of application.



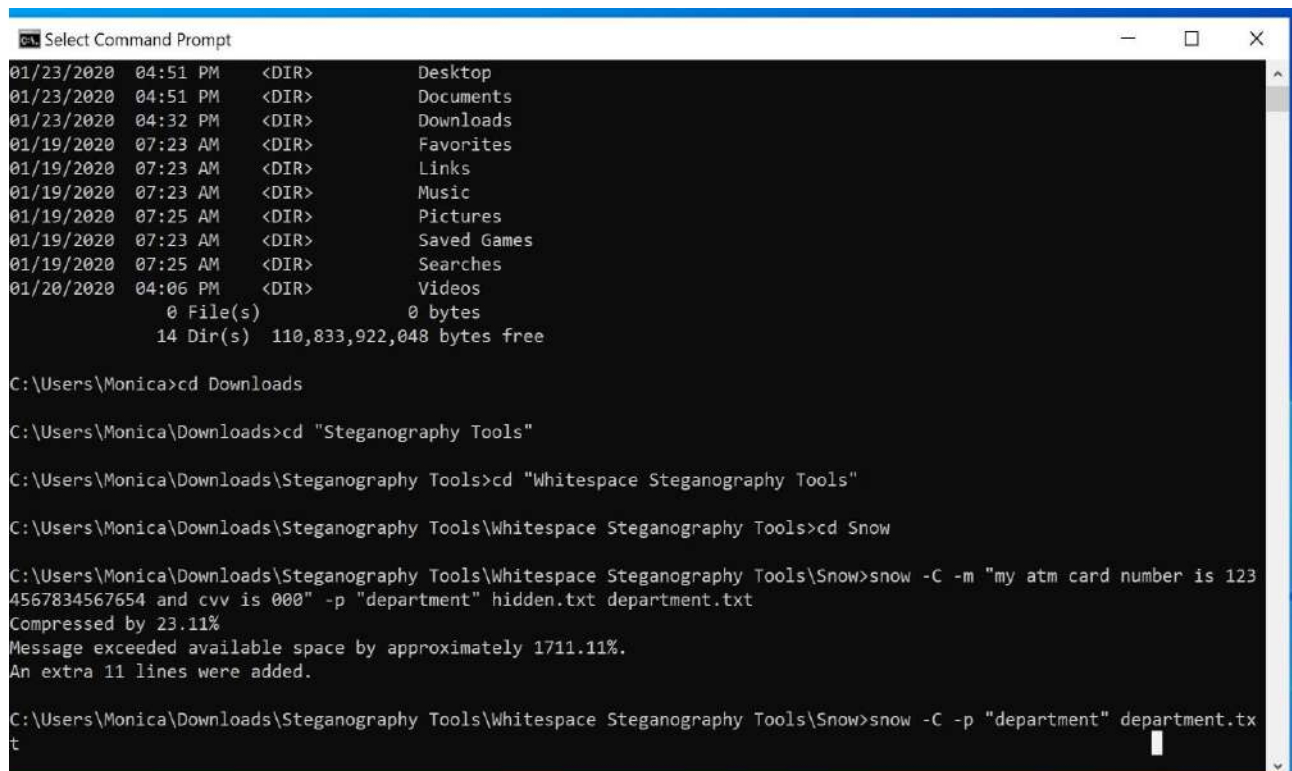
Step: 7 Now check for the embedded text in the department.txt file.



Step: 8

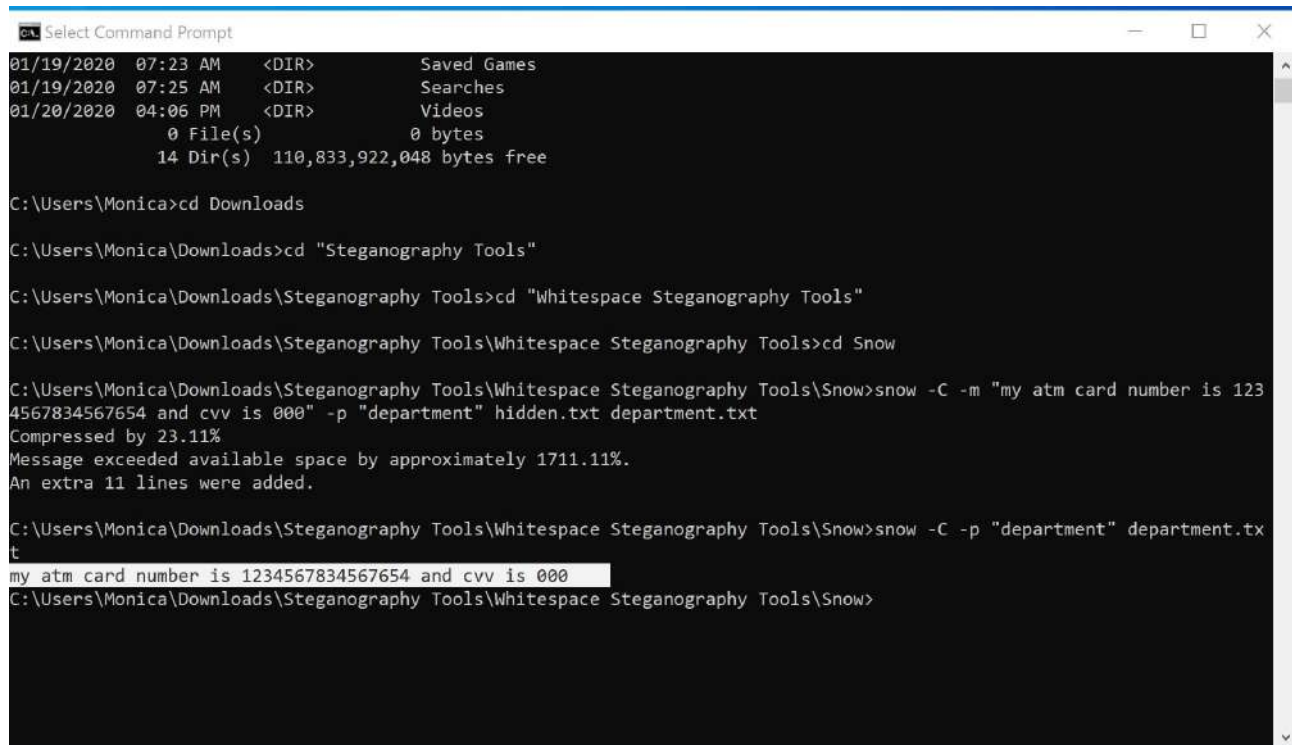
In command prompt goto file location and give following command to extract embedded text.

snow -C -p "department" department.txt



Step: 9

The embedded text will be explored there itself in the command prompt.



```
01/19/2020 07:23 AM <DIR> Saved Games
01/19/2020 07:25 AM <DIR> Searches
01/20/2020 04:06 PM <DIR> Videos
0 File(s) 0 bytes
14 Dir(s) 110,833,922,048 bytes free

C:\Users\Monica>cd Downloads
C:\Users\Monica\Downloads>cd "Steganography Tools"
C:\Users\Monica\Downloads\Steganography Tools>cd "Whitespace Steganography Tools"
C:\Users\Monica\Downloads\Steganography Tools\Whitespace Steganography Tools>cd Snow
C:\Users\Monica\Downloads\Steganography Tools\Whitespace Steganography Tools\Snow>snow -C -m "my atm card number is 123
4567834567654 and cvv is 000" -p "department" hidden.txt department.txt
Compressed by 23.11%
Message exceeded available space by approximately 1711.11%.
An extra 11 lines were added.
C:\Users\Monica\Downloads\Steganography Tools\Whitespace Steganography Tools\Snow>snow -C -p "department" department.tx
t
my atm card number is 1234567834567654 and cvv is 000
C:\Users\Monica\Downloads\Steganography Tools\Whitespace Steganography Tools\Snow>
```

STEGANALYSIS

Steganalysis is the way toward distinguishing steganography by reviewing different parameter of a stego media. The essential advance of this procedure is to distinguish a suspected stego media. After that steganalysis procedure decides if that media contains shrouded message or not and afterward endeavour to recuperate the message from it. In the cryptanalysis unmistakably the caught message is encoded and it surely contains the concealed message in light of the fact that the message is mixed. However, on account of steganalysis this may not be valid. The speculated media could possibly be with shrouded message. The steganalysis procedure begins with a lot of suspected information streams. At that point the set is diminished with the assistance of development factual methods.

Steganalysis Techniques

The properties of electronic media are being changed in the wake of concealing any article into that. This can result as debasement as far as quality or surprising attributes of the media: Steganalysis methods based on abnormal example in the media or Visual Detection of the equivalent. For instance on account of Network Steganography uncommon example is

presented in the TCP/IP bundle header. In the event that the bundle examination strategy of Intrusion Detection System of a network is based on white rundown design (common example), at that point this technique for network steganography can be crushed. On account of Visual recognition steganalysis strategy a lot of stego images are contrasted and unique spread images and note the obvious distinction. Mark of the shrouded message can be determined by looking at various images. Trimming or cushioning of image likewise is a visual intimation of concealed message since some stego instrument is editing or cushioning clear spaces to fit the stego image into fixed size. Contrast in record estimate between spread image and stego images, increment or decline of novel hues in stego images can likewise be utilised in the Visual Detection steganalysis technique.

Steganography Tools

- Xiao Steganography
- Image Steganography
- Steghide
- Crypture
- RSteg
- SSuite Piscal
- Our Secret
- OpenStego
- SteganPEG
- Hide'N'Send

The detection of steganography software on a suspect computer is important to the subsequent forensic analysis. As the research shows, many steganography detection programs work best when there are clues as to the type of steganography that was employed in the first place. Finding steganography software on a computer would give rise to the suspicion that there are actually steganography files with hidden messages on the suspect computer. Furthermore, the type of steganography software found will directly impact any subsequent steg analysis (e.g., S-Tools might direct attention to GIF, BMP, and WAV files, whereas JP Hide-&-Seek might direct the analyst to look more closely at JPEG files).

One commonly used detection program is Niels Provos' steg detects. Steg detect can find hidden information in JPEG images using such steganography schemes as F5, Invisible Secrets, JPHide, and JSteg (OutGuess 2003).

WetStone Technologies' Stego Watch (WetStone Technologies 2004B) analyses a set of files and provides a probability about which are steganography media and the likely algorithm used for the hiding (which, in turn, provides clues as to the most likely software employed). The analysis uses a variety of user-selectable statistical tests based on the carrier file characteristics that might be altered by the different steganography methods. Knowing the steganography software that is available on the suspect computer will help the analyst select the most likely statistical tests.

CHAPTER VI: CONCLUSION

The implementation of the Whitespace Steganography interest in the use of steganography in our current digital age can be attributed to both the desire of individuals to hide communication through a medium rife with potential listeners, the absolute necessity of maintaining control over one's ownership and the integrity of data as it passes through this medium. This increased interest is evidenced in the sheer number of available tools to provide easy steganographic techniques to the end user, as well as the proliferation of research and press on the topic.

The intent of this work was to cover some of the more common methods of data hiding using widespread file format and easily available tools as an introduction to the primary concepts of white space steganography. These discussions should serve as a starting point to the exploration of more complex steganographic techniques involving, for example, the use of network packets and unused hard disk space as cover medium, or the more complex methodologies used on text files.

In extend to this the white space steganography can also be implemented on the other formats of the document files using many other available tools. The steganography techniques can be used for these purposes and can be considered the future scope of this work.

REFERENCE

https://www.researchgate.net/profile/Yee_Por/publication/228672143_WhiteSteg_A_new_scheme_in_information_hiding_using_text_steganography/links/0deec531ae4e4dd4a9000000.pdf

https://www.garykessler.net/library/fsc_stego.html

<http://datahide.org/BPCSe/applications-e.html>

<http://www.darkside.com.au/snow/>

<https://en.wikipedia.org/wiki/Steganography#Techniques>

International Journal For Technological Research In Engineering Volume 6, Issue 7, March-2019 ISSN (Online): 2347 – 4718

Rakhi, Suresh Gawande, "A REVIEW ON STEGANOGRAPHY METHODS", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, 2015

Jasleen Kour, Deepankar Verma, "Steganography Techniques –A Review Paper", International Journal of Emerging Research in Management & Technology, 2014

Prashant Johri, Arun Kumar, Amba, "Review Paper On Text And Audio Steganography Using GA", International Conference on Computing, Communication and Automation, 2015

Rehana Begum R.D , Sharayu Pradeep, "Best Approach for LSB Based Steganography Using Genetic Algorithm and Visual Cryptography for Secured Data Hiding and Transmission over Networks", ISSN: 2277 128X , Volume 4, Issue 6, June 2014.

Abhishek Tripathy, Dinesh Kumar, "Genetic Algorithm Based Image Steganography for Enhancement of Concealing Capacity and Security", International Journal of Advanced Research in Computer Science and Software Engineering, ISSN: 2277 128X, Volume 4, Issue 4, April 2014

Ms. Pratidnya Sapate, Ms. Varsha Patil, Ms. Mayuri Pardeshi, Prof. Arjun Nichal, "A Review Paper on Video Steganography", International Advanced Research Journal in Science, Engineering and Technology, 2016

Pritam Kumari, Chetna Kumar, Preeyanshi and Jaya Bhushan, "Data Security Using Image Steganography And Weighing Its Techniques,"International Journal Of Scientific & Technology Research Volume 2, Issue 11, November 2013, pp. 238-241

Soumyendu Das, Subhendu Das,"Steganography and Steganalysis: Different Approaches",IEEE,2013