An Approach of Quantum Steganography through Special SSCE Code

Indradip Banerjee, Souvik Bhattacharyya and Gautam Sanyal

Abstract—Encrypted messages sending frequently draws the attention of third parties, perhaps causing attempts to break and reveal the original messages. Steganography is introduced to hide the existence of the communication by concealing a secret message in an appropriate carrier like text, image, audio or video. Quantum steganography where the sender (Alice) embeds her steganographic information into the cover and sends it to the receiver (Bob) over a communication channel. Alice and Bob share an algorithm and hide quantum information in the cover. An eavesdropper (Eve) without access to the algorithm can’t find out the existence of the quantum message. In this paper, a new quantum steganography technique based on the use of indefinite articles (a) or (an) in conjunction with the non-specific or non-particular nouns in English language and quantum gate truth table have been proposed. The authors also introduced a new code representation technique (SSCE - Secret Steganography Code for Embedding) at both ends in order to achieve high level of security. Before the embedding operation each character of the secret message has been converted to SSCE Value and then embeds to cover text. Finally stego text is formed and transmits to the receiver side. At the receiver side different reverse operation has been carried out to get back the original information.

Keywords—Quantum Steganography, SSCE (Secret Steganography Code for Embedding), Security, Cover Text, Stego Text.

I. INTRODUCTION

INFORMATION hiding is a general term encompassing many sub disciplines. One of the most important sub disciplines is steganography [3], [14] as shown in Figure 1. Steganography, is derived from a work by Johannes Trithemius (1462-1516) entitled "Steganographia" and comes from the Greek language defined as "covered writing" [5]. It is an ancient art of hiding information in ways a message is hidden in an innocent-looking cover media so that will not arouse an eavesdroppers suspicion. Steganography differs from cryptography in the sense that where cryptography focuses on keeping the contents of a message secret, steganography focuses on keeping the existence of a message secret [4],[36].

Another form of information hiding is digital watermarking, which is the process that embeds data called a watermark, tag or label into a multimedia object such that watermark can be detected or extracted later to make an assertion about the object. The object may be an image, audio, video or text only [35],[25]. A covert channel could be defined as a communications channel that transfers some kind of information using a method originally not intended to transfer this kind of information. Observers are unaware that a covert message is being communicated. Only the sender and recipient of the message notice it. Steganography works have been carried out on different media like images, video clips, text, music and sound [7],[36].

In Image Steganography method the secret message is embedded into an image as noise to it, which is nearly impossible to differentiate by human eyes [20],[19],[24]. In video steganography, same method may be used to embed a message [11],[12]. Audio steganography embeds the message into a cover audio file as noise at a frequency out of human hearing range [24]. One major category, perhaps the most difficult kind of steganography is text teganography or linguistic steganography because due to the lack of redundant human language to conceal a secret message as defined by Chapman et al. [6].

Quantum Steganography: Comparatively very little research work has been done in quantum steganography also. The idea of hiding secret messages as the error syndromes of a quantum error-correcting code (QECC) was introduced by Julio Gea-Banacloche in [16]. In his work Alice and Bob use the three-bit repetition code to transmit messages to each other using a shared secret key. All the noise in the channel that Eve perceives is because of these deliberate errors that Alice applies. In his model he assumes that Alice and Bob share a binary-symmetric channel. This work does
not address the issue of whether the errors would resemble a plausible channel, nor does it consider the case where the channel contains intrinsic noise. Natori gives a simple treatment of quantum steganography which is a modification of super-dense coding [23]. Martin introduced a notion of quantum steganographic communication based on a variation of Bennett and Brassard’s quantum-key distribution (QKD), hiding a steganographic channel in the QKD protocol [21]. Curty e.al proposed three different quantum steganographic protocols [8].

Quantum gate[22]: Quantum circuit model of computation in quantum computing[15],[10],[9], a quantum gate or quantum logic gate is a basic quantum circuit which operates on a small number of qubits. They are the building blocks of quantum circuits, like classical logic gates are basically for conventional digital circuits. Quantum logic gates are reversible like other classical logic gates. However, classical computing can be performed by the help of only reversible gates. Quantum gates are represented as matrices. A gate which acts on k qubits is represented by a $2^k \times 2^k$ unitary matrix. The number of qubits in the input and output of the gate is equal.

Reversible Classical Logic: The first concept of the reversibility of computation were raised in the 1970s. There were two issues which are logical reversibility and physical reversibility, both were intimately connected. Logical reversibility reconstruct the input from the output of a computation or gate function. The NAND gate is explicitly irreversible, it has two inputs and one output, while the NOT gate is reversible (its own inverse). In case of Physical reversibility the NAND gate has only one output, one of it’s inputs has effectively been erased in the process, whose information has been irretrievably lost. The change in entropy that we would associate with the loss of one bit of information is $\ln 2$, which, thermodynamically, corresponds to an energy increase of $kT \ln 2$, where $k$ is Boltzmann’s constant and $T$ is the temperature. The heat dissipated during a process is usually taken to be a sign of physical irreversibility, that the microscopic physical state of the system cannot be restored exactly as it was before the process took place. Reversible logic gates are symmetric with respect to the number of inputs and outputs. The reversible NOT gate, whose truth table is given in Figure 2. It can also write this in the form of a matrix, or as a graphic. The matrix form lists the lines in the truth table in the form $\langle 0 \rangle$, $\langle 1 \rangle$. The matrix field with 1’s and 0’s such that each horizontal or vertical line has exactly one 1, which is to be interpreted as a one-to-one mapping of the input to the output. A two-bit gate closely related to the NOT gate is the two-bit Controlled-NOT (or C-NOT) gate. Controlled-NOT gate shows in Figure 3, performs a NOT on the second bit if the first bit is $\langle 1 \rangle$, but otherwise has no effect. The C-NOT is sometimes also called XOR, since it performs an exclusive OR operation on the two input bits and writes the output to the second bit.

Figure 4 shows the basic text steganography mechanism. Firstly, a secret message (or an embedded data) will be concealed in a cover-text by applying an embedding algorithm to produce a stego-text. The stego-text will then be transmitted by a communication channel, e.g. Internet or mobile device to a receiver. For recovering the secret which sent by the sender, the receiver needs to use a recovering algorithm which is parameterised by a stego-key to extract the secret message. A stego-key is used to control the hiding process so as to restrict detection and/or recovery of the embedded data to parties who know it [2],[14].

Text steganography can be classified in three basic categories [5] - format-based, random and statistical generation and linguistic method.

Format-based methods used 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, deliberate misspellings distributed throughout the text, resizing the fonts are some of the many format-based methods being used in text steganography.
II. THE PROPOSED MODEL

Figure 6 shows the block diagram of the proposed secretkey text steganographic model. The input messages can be in any digital form and are often treated as a bit stream. The input message is first encrypted using a new code generation technique SSCE. This encrypted message generates the secret key, (which may be called a message enabled key). Then a matrix formed with the help of message length and map the C-NOT truth table (shown in Figure 9) from left most corner in a sequence (vertically or horizontally), after that put the secret key one by one by replacing ‘0’ value. Traverse the matrix and form the secret key with all values and ready for embedding. Before embedding a checking has been done to find out whether the vowels and consonants are placed in the cover text as per the grammatical order, if not place it in proper order. For the improvement of security level, the SSCE code representation has been used to encrypt the message and then secret message has been embed to the cover text by inserting indefinite articles a or an in conjunction with the non-specific or non-particular nouns in English language based on the mapping information given in Figure 7 to form the stego text. At the receiver side other different reverse operation has been carried out to get back the original information.

However, Bennett has stated that those format based methods managed to trick most of the human eyes but it cannot trick once computer systems have been used.

Random and statistical generation is generating cover text according to the statistical properties. This method is based on character sequences and words 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 codebook 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.

In this paper, a new approach of text quantum steganography have been proposed based on the use of indefinite articles a or an in conjunction with the non-specific or non-particular nouns in English language. A new code representation method SSCE also have been proposed here to achieve high level of security. Before the embedding operation each character of the secret message has been encoded using SSCE Value and then embeds into cover text by the proposed text steganography method to form the stego text. This method is an integrated approach of new secret code generation along with a text based steganography method using quantum truth table. Incorporating these two approaches in an embedding algorithm, a high embedding capacity of secret message can be achieved.

The proposed scheme has been inspired by the authors previous work [32],[30],[33],[26],[34],[29],[28],[27] on a new approach of text steganography method by inserting extra blank space between the words of odd or even size of the cover according to the embedding sequence and Introducing SSCE Value in [31],[29],[28].

This paper is organized into the following sections. Section II describes the proposed model. Algorithms of various processes like embedding, extracting, encryption, decryption and GUI are discussed in Section III. Mathematical Analysis are in section IV. Analysis of the processes and results are discussed in Section V and work is concluded in Section VI.

III. ALGORITHM

In this section, algorithms for different processes used both in the sender side and receiver side are described.
A. Algorithm for Message Encryption / Decryption
- Select the message and pick one by one character.
- Convert to its ASCII equivalent.
- Change ASCII code to our generated code from SSCE Table (Figure 10).
- Convert to its character equivalent.

B. Algorithm for Message Embedding
- Select the message and encrypt the message with SSCE table.
- Declare a MATMSG(N x N) matrix, where N is total length of message.
- Map the quantum C-NOT gate to the matrix vertically or horizontally.
- Put the message value by replacing '0' in the matrix MATMSG.
- Pick values one by one from MATMSG and create MSG.
- Select the cover text matrix to embed the message. Check whether the selected text is capable of embedding. If not possible repeat this step otherwise continue.
- Check the message sequence and pick first two bit sequence (MSG).

C. Algorithm for Message Extracting
- Select the newly generated text (stego text) after message embedding and their positions.
- Select the embedding position in TX.
- If there is word (an) and next words first character is vowel, then MSG=11.
- Else If there is word (a) and next words first character is vowel, then MSG=10.
- Else If there is word (an) and next words first character is consonant, then MSG=01.
- Else If there is word (a) and next words first character is consonant, then MSG=00.

D. Algorithm for GUI
In this section the two algorithmic approach is described one for the function of the Sender Side and another for the Receiver Side.

1) Sender side:
- Select the Cover Text from the set of Text files.
- Check whether the selected text is capable to do the embedding or not. If not possible then error.
- Select the message in text form.
- Encode the message through SSCE value.
- Embed the encrypted message in the cover text to form the stego text.
- End.

2) Receiver side:
- Receive the text with embedded message along with positions.
- Extract the encrypt form of message from the Stego Text.
- Decrypt the message with the help of the SSCE value.
- End.


A. Similarity Measure

We are using two methods for comparing the similarity between cover text and stego text, which are described below.

1) Correlation: The most familiar measure of dependence between two quantities is the Pearson product-moment correlation coefficient [113], or "Pearson’s correlation." It is obtained by dividing the covariance of the two variables by the product of their standard deviations. Karl Pearson developed the coefficient from a similar but slightly different idea by Francis Galton. The Pearson correlation is +1 in the case of a perfect positive (increasing) linear relationship (correlation), -1 in the case of a perfect decreasing (negative) linear relationship (anticorrelation) [13], and some value between -1 and 1 in all other cases, indicating the degree of linear dependence between the variables. As it approaches zero there is less of a relationship (closer to uncorrelated). The closer the coefficient is to either -1 or 1, the stronger the correlation between the variables. If the variables are independent, Pearson’s correlation coefficient is 0, but the converse is not true because the correlation coefficient detects only linear dependencies between two variables.

If we have a series of n measurements of X and Y written as $x_i$ and $y_i$, where $i = 1, 2, ..., n$, then the sample correlation coefficient can be used in Pearson correlation r between X and Y. The sample correlation coefficient is written as:

$$ r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} $$

where $\bar{x}$ and $\bar{y}$ are the sample means of X and Y, $s_x$ and $s_y$ are the sample standard deviations of X and Y. The number of matching (but different sequence order) characters divided by two defines the number of transpositions. The Correlation score of comparing cover text and stego text is 5.5460e+003 (in case of long message), -611.7406 (in case of too short message), which means this method is not possible in this work.

2) JaroWinkler Distance: The Jaro-Winkler distance for measuring similarity between two strings has been computed. The Jaro-Winkler distance [37] is a measure of similarity between two strings. It is a variant of the Jaro distance metric [17], [18] and mainly used in the area of record linkage [17] (duplicate detection). The higher the Jaro-Winkler distance for two strings is, the more similar the strings are. The score is normalized such that 0 equates to no similarity and 1 is an exact match. The Jaro distance metric states that given two strings $S_1$ and $S_2$ their distance $d_j$ is:

$$ d_j = \frac{1}{3} \left[ \frac{m}{|S_1|} + \frac{m}{|S_2|} + \frac{m - t}{m} \right] $$

where $m$ is the number of matching characters between the strings, $|S_1|$ and $|S_2|$ are the lengths of the strings, and $t$ is the number of transpositions.
where \( m \) is the number of matching characters and \( t \) is the number of transpositions. Two characters from \( S_1 \) and \( S_2 \) respectively are considered matching only if they are not farther than \( \left\lfloor \frac{\max\{|S_1|,|S_2|\}}{2} \right\rfloor - 1 \). Each character of \( S_1 \) is compared with all its matching characters in \( S_2 \). The number of matching (but different sequence order) characters divided by two defines the number of transpositions. The Jaro score of comparing cover text and stego text is 0.9022, which means they are closely similar. Besides comparison through histogram technique has been done. It has been observed that the histogram of the cover text and the stego text is almost identical.

VI. CONCLUSION

In this paper the authors presented a new approach of text quantum steganography method by changing the (a), (an), vowel and consonant rule of the cover text according to the embedding sequence and also in some cases only the (a)-consonant and (an)-vowel between the words of the original cover text may be used as mapping the embedding sequence. This property generates the stego text with minimum degradation using quantum truth table. This property enables the method to avoid the steganalysis also. The new BASE Value (SSCE - Secret Steganography Code for Embedding) has been used to generate the encrypted form of the message in order to achieve high level of security. This approach is capable of secure transfer of the message compared to earlier techniques. The future work should be focused to improve the capacity of the embedding scheme by incorporating some compression technique on the secret message.
REFERENCES


Gautam Sanyal has received his B.E and M.Tech degree from Regional Engineering College (REC), Durgapur, now, National Institute of Technology (NIT), Durgapur, West Bengal, India. He has received Ph.D (Engg.) from Jadavpur University, Kolkata, West Bengal, India, in the area of Robot Vision. He possesses an experience of more than 25 years in the field of teaching and research. He has published nearly 40 research papers in International and National Journals / Conferences. His current research interests include Natural Language Processing, Stochastic modeling of network traffic, High Performance Computing, Computer Vision. He is presently working as a Professor in the department of Computer Science and Engineering and also holding the post of Dean (Student’s Welfare) at National Institute of Technology, Durgapur, West Bengal, India.