An Eccentric Scheme for Oblivious Communication

Khan Farhan Rafat and Muhammad Sher

Department of Computer Science, International Islamic University
Islamabad, 44000, Pakistan

Abstract
Trust is the foremost requirement in any form of communication. Building trust through physical contact, gesture etc. is easy but hard to establish in electronic data communication as one can’t be sure of the presence of the intended recipient at the other end. This necessitated the need to devise and suggest covert schemes for oblivious communication that only the intended recipient may unveil. Steganography is one such technique where information to be sent is seamlessly superimposed on the carrier such that it easily crosses the information confronting barriers without detection. This paper endeavors at evolving a data hiding scheme envisaged by Shannon and in adherence to Kerchoff’s principle that hides secret information inside statistically random English alphabets appearing as cryptogram. Effect of bit embedding is computed through contrasting probability distribution of cover text and stego object.

Keywords: Information Theoretic Security, Cryptogram, Trust, Oblivious Communication, Steganography.

1. Introduction

400 B.C. saw Greeks articulating Steganography for covert communication. Herodotus while accounting on the quarrel of 5th Century B.C. between Greece and Persia elaborated in [1] on how Demaratus; exiled from Greek; informed his natives in Persia of Xerxes plan of invading Greece by engraving secret message on wooden tablet and coating it with wax where recipient melted wax to read the carved message.

In another incident, as narrated by Herodotus; another Greek named Histaiacus; called for an insurgency against Persian King by shaving and tattooing a secret message of Persian plan of revolt on the head of one of his slaves. Ample time was then given for his hair to grow so as to hide / cover the tattooed message after which he was sent to the enemy territory. The intended resident in Persia read the hidden message after shaving the head of the messenger.

Ancient Chinese are known to write their message on a piece of silk which they crunched into a little ball. The ball is then covered with wax and gets swallowed by the messenger [2].

Æneas, the author of the book titled “On the Defense of Fortified Places” developed a Steganographic System by blurring holes into wood that corresponds to Greek alphabets and then threading the yarn through these holes in an order which when detached spelt out letters of hidden message in its correct sequence [3]:

“padielaporsymesarpomeuaspeludydynamalpreaxo”

“Ave Maria” is yet another smart method for encoding secret message where list of distinct words for each unique alphabet are grouped in a series of Tables. For every alphabet constituting the message, word corresponding to that letter from the set of Tables gets substituted and appeared as innocent rhyme whenever the Tables are used in sequence.

Cardano introduced the “Grille System” [4] where every recipient was given a piece of paper with a number of punched holes in it which when placed over an innocuous Stego Object (Text Message) reveals letters of hidden secret message.

In his book “Mercury” often referred to as “The Secret and Swift Messenger”, author Bishop John Wikims described ways such as use of alum, ammonia salt, onion juice besides “distilled juice of Glowworms” for hidden writing that glows when exposed to light [5].

In an attempt to dethrone protestant queen Elizabeth, Mary a Catholic, made use of information hiding techniques to secretly communicate with Catholic Noblemen of England [6].

George Washington (General and first United States President) also made use of Steganography and Cryptography for cover communication and ordered for the use of invisible inks on regular messages in an attempt to divert suspicion [7].

Franco-Prussian war (1870-71) saw Pigeons brought in France, 1000 in numbers, being used as message carriers for delivering innumerable Military and more than 95,000 secretive dialogues from and into France [8].

Tobin et al. in [9] elaborated on the use of Quilts, hung outside to dry, with special symbols / patterns sewed into its body thereby innocuously pointing towards directions to help prisoners escape.

World War – I (WW-I) saw improvement in Cardano’s Grill that gets rotated 90 degrees each time a series of letters get encoded when Germans introduced different type of Grills based on varying message lengths. This
technique, however, was cracked within four months since its inception. In another incident a lady alleged as German spy was found in possession of blank sheet of paper which when inspected exposed a hidden message written with invisible ink [10].

Radio communication was also used for the purposes of message transmission, for the first time, also during WW-I. Cable censor and Cigar movements conforming to ships movements are also some popular examples of the same retro [11].

World War II also witnessed extensive usage of Steganography as explicated through text messages written in quotes using technique called concealment cipher, extracted from [12] as under:

“Apparently neutral’s protest is thoroughly discounted and ignored. Isman hard hit. Blockade issue affects pretext for embargo on by products, ejecting suets and vegetable oils.”

The stealth message “Pershing sails from NY June 1.” can be extracted from afore mentioned quoted string by taking second letter of each word.

Captivated crew members of U.S. Naval spy ship “Pueblo” used body gestures to secretly communicate the message “Snow Job” when the ship got captured by North Koreans on January 23, 1968 [13].

During Vietnam War (1954) Commander Jeremiah Deuton of United States Navy secretly communicated the plight “T-O-R-T-U-R-E” through his eyes using Morse code [14].

Former British Prime Minister Margret Thatcher was so aggravated with information degeneracies that she ordered for Word Processors to secretly hide identity of its user in word spaces to identify the disloyal amongst her Ministers and Staff [15].

Aligned with context of this paper there exist a web-based text generating Steganographic tool called “spammimic” known for using “spam” grammar and a mimicry algorithm named after Wayner to generate text cover [1].

The most recent case of hidden communication is the famous verdict by Judge in the “Devince code” case where first few lines of his verdict contains his hidden nickname as “Smithy Code” i.e., a statement given by Smith [16].

Prior to proceeding further it may not be out of place to mention here some personalities who presented revolutionary ideas often emphasized in context of cryptography but surely are rudimentary to ensure security of any cryptographic or Steganographic system:

a. Born in 1535 Giovanni Porta gave the concept of substitution and transposition that earned him a name in the history of Cryptography [17].

b. Auto Key concept given by Blaise de Vigenere that got re-invented in the nineteenth century [18].

c. According to Kerckhoff security of a System must be a feature of its Key when the algorithmic details are made public [19].

Rest of the paper is premeditated as follows: Section 2 builds the foundation for our proposed solution which is explained in detail in Section 3. Section 4 illustrates test results. Advantages and recommendations are given in Section 5 and 6 respectively. Section 7 concludes the discussion.

2. Building Blocks

Text Steganography uses “text” as cover for secret communication. Operations such as text formatting, text insertion, text substitution, generating text through context-free grammars etc. are in use for text Steganography. Further “text format” stands alone when it comes to saving extra information (meta data) associated with specific file format i.e., text files are saved, retrieved and viewed as these appear before human eye, and it is this peculiar trait that makes it difficult to devise ways for hiding information inside text cover.

Images (due to high data capacity) by far are the most preferred choice when it comes to cover selection for the purpose of information hiding. However, as in [20], [21] and [22] major limitations of Image Steganography include:

a. Data/Information Loss when compressed using lossy compression.

b. Choice of cover image.

c. Original cover or its related information is required at receiver’s end for information retrieval.

The popular method for image Steganography is inserting secret message bits at least significant bit positions in the cover image, called as LSB (Least Significant Bit) Steganography. Figure 1 illustrates hard to distinguish (from perceptibility point of view) cover and Stego images along with their graphical representation respectively.

2.1 Difficulty in using LSB scheme for ASCII Text Steganography

To highlight the difficulty associated with hiding data inside text cover, we chose the cover text (extracted from http://ijcsi.org/about.php) shown in Figure 2 and substituted “0” and “1” at LSB position of each of its words the outcome of which are illustrated in Figure 3 and 4 respectively.
It is evident from above illustrations that even a single bit change at LSB position in cover text raises perceptibility concern and hence may not be used for information hiding.

2.2 Frequency Count Reveals Data Hiding Scheme

Samuel Morse (1791-1872) contribution of encoding English alphabets based on their frequency of occurrence [23] in Standard English text paved the way for American Standard Code for Information Interchange referred to as ASCII character set and pronounced as /æski/ASS-kee [24] for digital information exchange / electronic communication. The concept of use of frequency of occurrence, however, has been browbeaten by cryptanalysts in solving cryptogram i.e., counting frequency of occurrence of English alphabets helps in revealing the type of encryption technique being used as either substitution or transposition [25]. Case insensitive analysis of English alphabets for cover text gives the frequency distribution shown in Table 1.

Frequency count of English alphabets shown in Table 1 seems in agreement (to a great extent) with that given in [12] and illustrated in Figure 5, from where cryptanalyst may easily infer that the text under examination is the outcome of substitution scheme.

Assuming the under discussion text be a result of monoalphabetic substitution, even then the frequency of substituted letters would correlate with that as in Figure 5, thereby enabling the cryptologist to infer that before him is a case of substitution cipher.

2.3 How to Impede Prediction?

Aforesaid revealing attribute of English alphabets resulted in use of cryptographically secure randomly generated numbers through hard to predict processes like radioactive decay, thermal, acoustic sources [26] etc. Randomness
adds uncertainty that elevate Wendy’s effort towards determining the underlying information hiding scheme – a concept also supported by Information Theory.

Table 1: Frequency Count for English Alphabets in Cover Text

<table>
<thead>
<tr>
<th>Alphabets</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>54</td>
</tr>
<tr>
<td>B</td>
<td>9</td>
</tr>
<tr>
<td>C</td>
<td>34</td>
</tr>
<tr>
<td>D</td>
<td>15</td>
</tr>
<tr>
<td>E</td>
<td>78</td>
</tr>
<tr>
<td>F</td>
<td>17</td>
</tr>
<tr>
<td>G</td>
<td>9</td>
</tr>
<tr>
<td>H</td>
<td>17</td>
</tr>
<tr>
<td>I</td>
<td>69</td>
</tr>
<tr>
<td>J</td>
<td>7</td>
</tr>
<tr>
<td>K</td>
<td>0</td>
</tr>
<tr>
<td>L</td>
<td>24</td>
</tr>
<tr>
<td>M</td>
<td>8</td>
</tr>
<tr>
<td>N</td>
<td>51</td>
</tr>
<tr>
<td>O</td>
<td>43</td>
</tr>
<tr>
<td>P</td>
<td>23</td>
</tr>
<tr>
<td>Q</td>
<td>1</td>
</tr>
<tr>
<td>R</td>
<td>44</td>
</tr>
<tr>
<td>S</td>
<td>61</td>
</tr>
<tr>
<td>T</td>
<td>48</td>
</tr>
<tr>
<td>U</td>
<td>17</td>
</tr>
<tr>
<td>V</td>
<td>8</td>
</tr>
<tr>
<td>W</td>
<td>2</td>
</tr>
<tr>
<td>X</td>
<td>1</td>
</tr>
<tr>
<td>Y</td>
<td>5</td>
</tr>
<tr>
<td>Z</td>
<td>0</td>
</tr>
</tbody>
</table>

2.4 Cryptogram

Electronically generated output of encryption process called cryptogram can take any form but are best expressed in terms of alphabets, numbers or blend of two and takes care of the secrecy aspect of communication security amongst confidentiality, integrity, and availability (C.I.A.). Figure 6 is an illustration as on p - 99 in [12]. For more recent example [27] is referred.

The desirous attribute of Cryptogram is to exhibit unpredictability/randomness [28], and if possessed is then ascertained to pass statistical tests evolved to verify on the randomness of generated data [29].

This section suggests a cover generation scheme envisioned from Shannon’s deliberation on types of secrecy systems [30] where he conferred that furtive systems may exercise methods such as “…, message in fake cryptogram, …”, and discussed as follows:

3. Proposed Atypical Steganographic Scheme

Perceptibility issue of Stego objects (Figure 3 and 4 refers) as a result of LSB substitution of words in Cover text (Figure 2 refers) ensued our analysis of various combinations of 7-bit binary strings for ASCII codes (with zero prefix) so as to arrive at the best possible scheme where LSB substitution does not lead to non-alphabetic code thereby avoiding perceptibility concerns together with least programming constraints in its subsequent implementation. The study, however, remained confined only to upper case English alphabets, where Table 2 illustrates our findings.

Additionally we wanted our proposed scheme to observe the following:

b. Does not require original cover text at receiver’s end for hidden bits extraction.
c. Provisioning for scanning of Stego object.
d. Ease in erroneous code detection and its subsequent correction.
e. Transmission of Cryptogram in printed form.
3.1 Bit Embedding

We generated random numbers using a PRNG the discussion of which is beyond the scope of this paper, in range 0 ~ 65535. The same were reduced to modulo 26 and the value 65 (ASCII code for alphabet “A”) was added to that yielded an English alphabet in range “A” to “Z” using equation:

\[ r \leftarrow \text{random MOD 26} + 65 \]  

(1)

Secret message was translated into equivalent binary bits. A 256-bit stego key was fed as input to 256-bit SHA-2 [31] HASH algorithm, the outcome of which was translated into equivalent binary bits and number of ON binary bits (i.e. 1’s) counted. Output from HASH algorithm was fed-back as next input (in place of stego key) as long as the number of ON bits remained less than the number of message bits. Next, random alphabets greater than or equal to the number of output bits from HASH algorithm (in multiple of 5) were generated and binary bit 1 or 0 from that HASH output was placed sequentially over each alphabet thus generated. Finally the random alphabets were traversed taking each binary message it, searching for alphabets A and Z. Upon finding the desired random alphabet, secret message bit got replaced with its LSB and substituted with alphabet corresponding to new binary equivalent. Whenever number of random alphabets fell short of secret message bits new alphabets got generated using aforesaid process. The process of bit embedding is illustrated in Figure 8.

Table 2: Considerations before Contriving LSB Based Cover Generation Steganographic Scheme

<table>
<thead>
<tr>
<th>ASCII code</th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>Observation</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>65</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>64 &lt; 65</td>
<td>( r \text{ random MOD 26} + 65 )</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>64 &lt; 65</td>
<td>Changing LSB to “O” results in “@” &lt; “A”</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>96 &lt; 90</td>
<td>String having its 0th bit ON are not suitable</td>
</tr>
<tr>
<td>90</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>92 &lt; 90</td>
<td>No dispute</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>98 &lt; 90</td>
<td>Perceptibility concern</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>90 &lt; 90</td>
<td>No dispute</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>93 &lt; 90</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>94 &lt; 90</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>91 &lt; 90</td>
</tr>
</tbody>
</table>

Output:

\[
z :: \text{Output bits; /* Stego Object */}
\]

Begin

1. \( q \leftarrow \text{Length(x)} \)
2. \( \text{lm} \leftarrow \text{Length(y)} \)
3. \( m \leftarrow 0 \)
4. \( n \leftarrow 0 \)
5. \( \text{for } w \leftarrow 1 \text{ to } q \)

Begin

6. \( \text{if Stego-Key(m)} = 1 \text{ then} \)
7. \( \text{if Mid(x,w,1)} <> “A” \text{ and Mid(x,w,1)} <> “Z” \text{ then} \)
8. \( \text{Mid(x,w,1)} \leftarrow \text{Character (ASC(left_7(Mid(x,w,1))))} \text{ || } \text{(Midy, n, 1))} \)
9. \( \text{if(n} \geq \text{lm)} \text{ then exit For Loop} \)
10. \( m \leftarrow (m+1) \text{ MOD 256} \)

End
End

The complexity of the algorithm 1.0 for worst case is worked out as \( O(q) \); while for best case it is \( O(\text{lm}) \).

3.2 Bit Extraction

Stego Key was fed into 256-bit HASH algorithm as input and the output got translated into bits. The output was also used as next input to the HASH algorithm for subsequent iterations– the process remained unremitting till the output equals alphabets in Stego Object. The ones and zeros thus obtained from Stego Key’s HASH were placed over every Stego Object’s alphabet.

The LSB of alphabets immediately beneath binary bit 1 of HASH bits less alphabets A and Z were extracted, and segregated into chunks of eight bits. The 8-bit binary strings were translated into equivalent ASCII codes (of English alphabets) that comprised hidden message. Figure 9 illustrates bit extraction process.

Table 3 succinctly exemplifies on bit embedding and extraction process.
4. Test Results

Test results speaks high on our proposed logic where these were quantified using Hamming [32], Levenshtein (Edit) [33], and Jaro-Winkler [34] distance given in Table 4, mean, variance and standard deviation shown in Table 5 and graphical illustration of Figure 7 using MiniTab 16 [35] by contrasting cover text vs. stego object as follows:

5. Advantages

Following justifies to opt for letter based cryptogram solution which, however, may be protracted to all ASCII character codes to take care of specific situation-based requirements:

a. Operational ease.

b. Minimized error propagation.

c. For transmitting as Hard copy.

d. Original cover text not required at receiving end.

6. Recommendations

Following are recommended for future enhancement of the proposed scheme:

a. Preference be given to use of True Random Number Generator (TRNG).

b. With oblivious Stego Key selection mechanism, even originator won’t be able to embark on locations used for embedding secret message bits in cover text.

7. Conclusion

This paper presented a secure steganographic cover generation scheme for ASCII text document by first deliberating on traits concerning ASCII character codes such as revelation of underlying data hiding technique e.g. substitution or transposition, by just counting the number of occurrences of English alphabets and contrasting those against their predetermined values in Standard English text. 256-bit Stego key dependent bit embedding is in
adherence to Kerchoff’s principle. The statistical characteristics of the stego object (randomness) makes it hard for Warden Wendy not to differentiate it from a cryptogram thereby providing information theoretically secure communication for a time sufficient enough to protect vital information constraint only by her effort towards breaking it. Moreover, the stego object is less prone to errors and can also be transmitted in printed form. Further, availability of cover text at receiving end is not required.

References


