American Journal of Applied Sciences 9 (1): 40-46, 2012
ISSN 1546-9239

New Approaches to Ancient Crypto-Steganography Methods

[1]Joseph Raphael, A. and [2]V. Sundaram
[1]Research Scholar, Karpagam University, Coimbatore, India
[2]Department of Computer Applications, Karpagam College of Engineering, Coimbatore, India

Abstract: Problem statement: The requirement and ability to hide information from others has existed in the world for centuries. Cryptography and steganography are the most commonly used techniques for information sharing. The science of securing a data by encryption is cryptography whereas the method of hiding secret messages in other media is Steganography, so that the secret's very existence is concealed. *N*either of them alone is secure enough for sharing information over an unsecure communication channel but at the same time they are vulnerable to intruder attacks. Although these techniques are often combined together to achieve higher levels of security but still there is a need of a highly secure system to transfer information over any communication media minimizing the threat of intrusion. **Approach:** Based on the combination of cryptography and steganography techniques, this study introduces two new methods to encrypt data and to hide the same in another media, in this way the message sent through an unsecured channel is concealed. **Results:** One of the methods shows how to encrypt and hide data in a xy graph and the other method shows a new way of encrypting and hiding data through Unicode symbols. **Conclusion:** Most of the traditional methods available today use the pixel bits of an image to hide information and are limited in terms of hiding capacity. The proposed methods hide information using Unicode symbols and plane xy coordinates which increases the hiding capacity and the message can be transmitted through an unsecured channel without any suspicion. This encoding and decoding scheme of the proposed method is significantly different as compared to the traditional methods.

Key words: Encryption, cipher data, decryption, xy graph, unicode symbols, unicode values, cryptography

INTRODUCTION

The fast development of the Internet and the digital information revolution caused major changes in the global society, ranging from the influence on the world economy to the way that people nowadays communicate. Broadband communication networks and multimedia data available in a digital format opened many challenges and opportunities for innovation. Versatile and simple-to-use software and diminishing prices of digital devices have made it possible for consumers from all over the world to create, edit and exchange multimedia data. Cryptography was created as a technique for securing the secrecy of communication, different methods have been developed to encrypt and decrypt data to keep the message secret. Unfortunately sometimes intruder succeeds in retrieving the secret message and thus cryptography technique fails. This leads the necessity to hide the existence of the secret message in a cover media and the same was achieved using a technique called steganography. But intruder extended their intelligence in retrieving the hidden secret messages from the cover media. Steganography and cryptography are best ways to protect information from third parties but neither technology alone is having its own boundaries. Once the presence of secret message is revealed or even suspected, the purpose of steganography is partially defeated. The strength of steganography can thus be augmented by combining it with cryptography. Most of the traditional steganographic methods alter the pixel values of an image to hide the message which leads to intruders attack and also limits the hiding capacity of the message (Narayana and Prasad, 2010; Cvejic, 2004; Houcque, 2005).

Two new methods are proposed by combining cryptography and steganography to encrypt and hide the message in a cover media. Unlike other methods, encrypted message is hidden in a xy graph and secondly using unicode symbols. In both the methods, messages are hidden in such a way that it will not create any suspect to intruders because hidden message in a media

Corresponding Author: Joseph Raphael, A., Research Scholar, Karpagam University, Coimbatore, India;
Lecturer, Department of Information Technology, Ibra College of Technology, Sultanate of Oman

appears to be a normal graph and unicode symbols to human vision. This combined science of cryptography and steganography could open a new application which leads to more secured communication through an open channel. Both the methods are implemented using MATLAB programming language.

MATLAB: MATLAB (MATrix LABoratory) is a numerical computing environment and fourth-generation programming language. It is a high-performance language for technical computing which integrates computation, visualization and programming environment. Furthermore, MATLAB is a modern programming language environment: it has sophisticated data structures, contains built-in editing and debugging tools and supports object-oriented programming. These factors make MATLAB an excellent tool for teaching and research. MATLAB has many advantages compared to conventional computer languages for solving technical problems. MATLAB is an interactive system whose basic data element is an array that does not require dimensioning. It has powerful built-in routines that enable a very wide variety of computations. It also has easy to use graphics commands that make the visualization of results immediately available.

Unicode: Unicode is a character encoding standard that has widespread acceptance. Unicode defines a large number of characters and assigns each of them a unique number, the Unicode code, by which it can be referenced. This encoding standard provides the capacity to encode all of the characters used for the written languages of the world. The objective of Unicode is to unify all the different encoding schemes so that the confusion between computers can be limited as much as possible. The most common Unicode encodings are called UTF-n, where UTF stands for Unicode Transformation Format and n is a number specifying the number of bits in a basic unit used by the encoding. Two very common encodings are UTF-16 and UTF-8. In UTF-16, which is used by modern Microsoft Windows systems, each character is represented as one or two 16-bit (two-byte) words and provides code point for more than 65000 characters (65536). Unix-like operating systems, including Linux, use another encoding scheme, called UTF-8, where each Unicode character is represented as one or more bytes. The benefit of Unicode is that, it assigns each character a unique value and symbol, no matter what the platform, no matter what the program, no matter what the language.

MATERIALS AND METHODS

Data hiding techniques have been widely used to hide and transmit secret message for long time. The aim of the system is to give new insights and directions on how to improve the existing methods of hiding secret messages, possibly by combining cryptography and steganography. The two new approaches combine cryptography and steganography methods which help us to achieve a higher level of secrecy and security. In the first method, encrypted message is hidden in a xy graph and in the second, the encrypted message is hidden through Unicode symbols. Both the approaches make it harder for any intruder to retrieve the plaintext of a secret message from a stego-object.

Method 1: Using xy graph:

Encryption and hiding message: In the proposed method, each character is extracted from the secret message; special character and blank space are also considered as one character. The extracted characters are combined together in two and their binary equivalent values are found which is further converted into decimal value. The process is repeated until all the characters from the secret message are converted into decimal values. The advantage of such an encryption method is that two characters are encrypted and hidden in one decimal number. The Fig. 1 below shows the original message and its equivalent decimal values (cipher data) obtained from the above said process.

MATLAB supports structure data types. Since all variables in MATLAB are arrays, a more adequate name is “structure array”, where each element of the array has the same field names. In line with the above said concept, the entire encrypted data vector is divided equally into two vectors say x and y as shown in Fig. 2a.

The x and y vector is given as an input to the built-in MATLAB function which generates a graph as an output based on the values of the vectors. The corresponding values from both the vectors form the xy points on the coordinate plane. Two characters are embedded in each vector value and since one point on the coordinate plane is plotted by taking values from x vector and y vector, every point on the polar coordinate holds four characters. Since four characters are hidden in one point, the hiding capacity of the secret message is increased which is the strength of this hiding method. The minimum and maximum values of x-axis and y-axis can be manually supplied to the software or based on the x and y data vector, default values are assigned by the software. Graph labels as well as title are assigned by the user according to his choice to divert the attention of an intruder from extracting the values from the graph.

Securing a data by encryption is cryptography whereas hiding secret message in other media is steganography

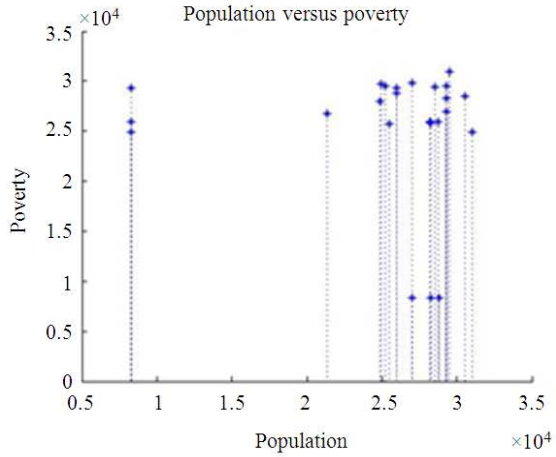
21349 25461 29289 28263 8289 8292 24948 24864 25209 8293 28259 29305 28788 26991 28192 26995 8291 29305 28788 28519 29281 28776 31008 30568 25970 25953 29472 26729 25705 28263 8307 25955 29285 29728 28005 29555 24935 25888 26990 8303 29800 25970 8301 25956 26977 8297 29472 29556 25959 24942 28519 29281 28776 31008

Fig. 1: Original message and its equivalent decimal values

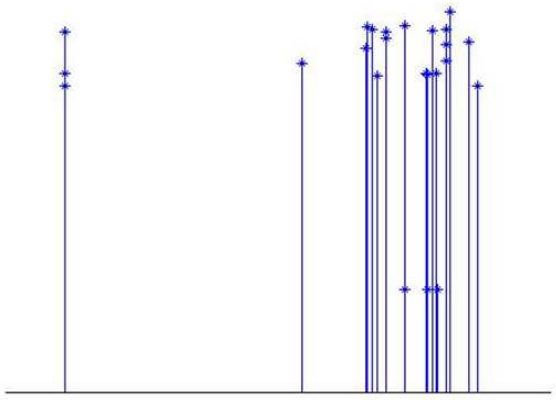
x = [21349 25461 29289 28263 8289 8292 24948 24864 25209 8293 28259 29305 28788 26991 28192 26995 8291 29305 28788 28519 29281 28776 31008 30568 25970 25953 29472]

y = [26729 25705 28263 8307 25955 29285 29728 28005 29555 24935 25888 26990 8303 29800 25970 8301 25956 26977 8297 29472 29556 25959 24942 28519 29281 28776 31008]

(a)



(b)



(c)

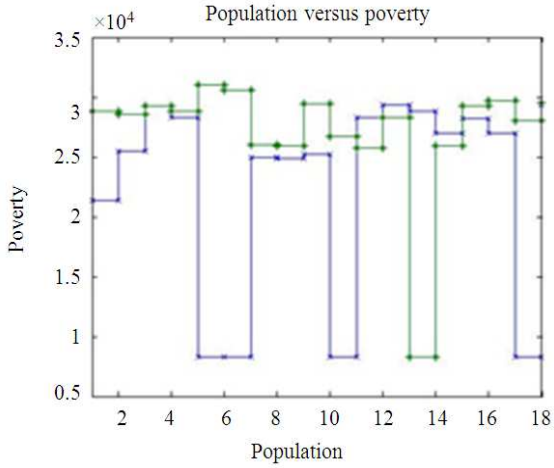
Fig. 2: (a) Encrypted data vector is divided into x vector and y vector (b) Resultant Stem type graph with axes (c) Resultant Stem graph without axes

x = [21349 25461 29289 28263 8289 8292 24948 24864 25209 8293 28259 29305 28788 26991 28192 26995 8291 29305]

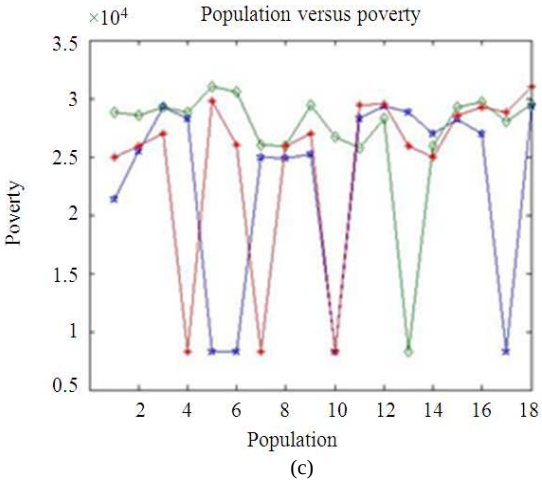
y = [28788 28519 29281 28776 31008 30568 25970 25953 29472 26729 25705 28263 8307 25955 29285 29728 28005 29555]

z = [24935 25888 26990 8303 29800 25970 8301 25956 26977 8297 29472 29556 25959 24942 28519 29281 28776 31008]

(a)



(b)



(c)

Fig. 3: (a) Cipher data assigned to 3 variables (b) Two series stair type graph (c) Three series line type graph

MATLAB supports different types of graph to be displayed such as line, bar, stair and stem, so the resultant graph can be of any said type in any format. The Fig. 2b and c shows the resultant graph in jpeg format; here each xy point holds 4 characters.

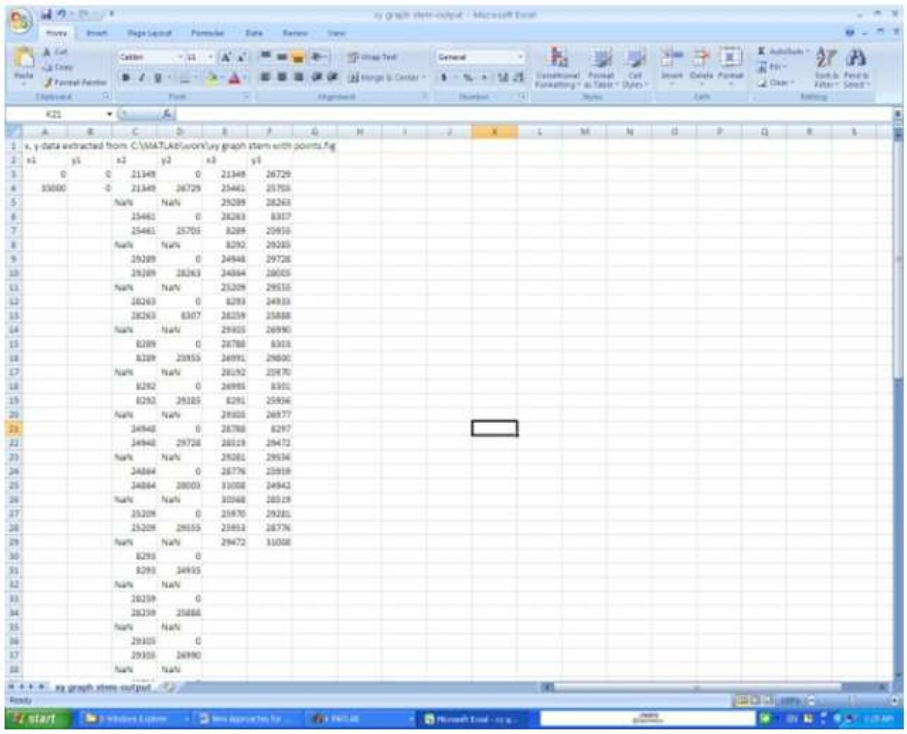


Fig. 4: Cipher data extracted from xy graph (from Fig. 2a)

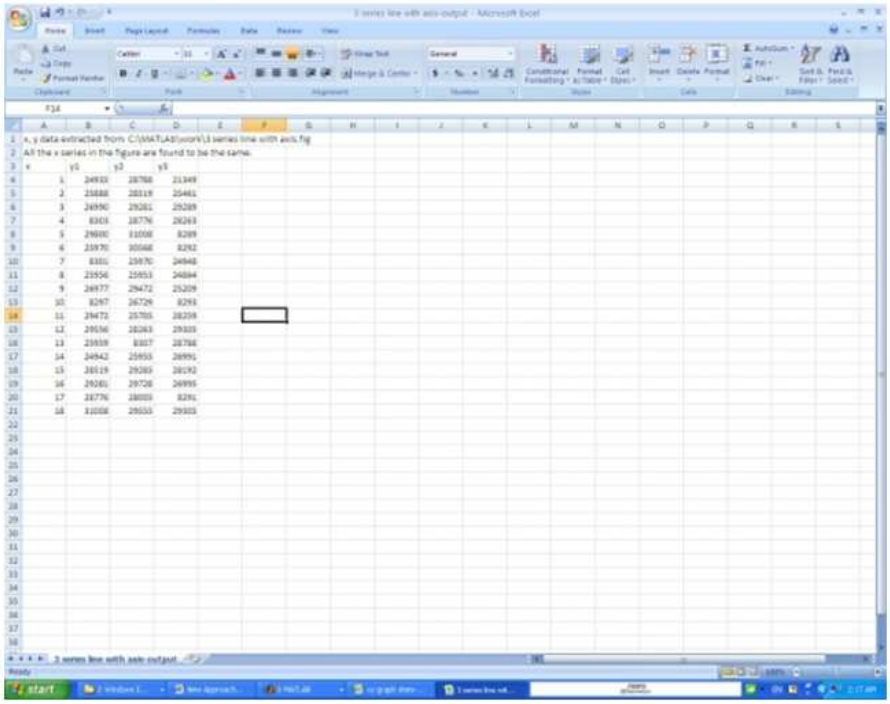


Fig. 5: Cipher data extracted from 3 series graph (from Fig. 3c)

Fig. 6: Text file with unicode symbols

Fig. 7: Method 1, using xy Graph

Instead of the above said xy graph, the available original cipher data can be divided and assigned to two or three variables and the graph can be plotted as two series or three series as shown in Fig. 3.

Decryption and retrieving original message: This module is just a reverse process of above said encryption module which involves two steps, first retrieving the hidden cipher data from the graph and secondly, converting the cipher data back to original message.

The resultant xy graph is given as an input to a MATLAB program which extracts the values of x axis and y axis separately in two vectors in an excel file. The Fig. 4 shows the output in excel file with the extracted xy values in the last two columns.

Using the same procedure as mentioned above the cipher data can be extracted from two series graph and three series graph as shown in Fig. 5.

The extracted value from excel file is read vector wise and its equivalent binary number is found in 16 bits. By dividing the resultant binary value as 8 bit each, its equivalent two characters are found. The process is repeated until all values are read from vectors and the accumulated characters forms the original message.

Method 2: using Unicode symbols:

Encryption and Hiding messages: In the early decades, each character was encoded in ASCII code which occupies one-per-byte in memory. In recent trend, characters are encoded in Unicode symbols which occupy more than one byte in memory. In the proposed method, the characters are extracted from the secret message, each special character and blank space in the secret message is considered as one character in the formation of Unicode symbols. The extracted characters are combined together in two and its binary equivalent values are found which is further converted into Unicode value, the resultant Unicode value is then mapped into its equivalent Unicode symbol. Newly generated Unicode symbols are written into a text file and the process is repeated until all the characters from the secret message are converted into Unicode symbols. The Fig. 6 shows the resultant text file with Unicode symbols generated from the original message (Fig. 1).

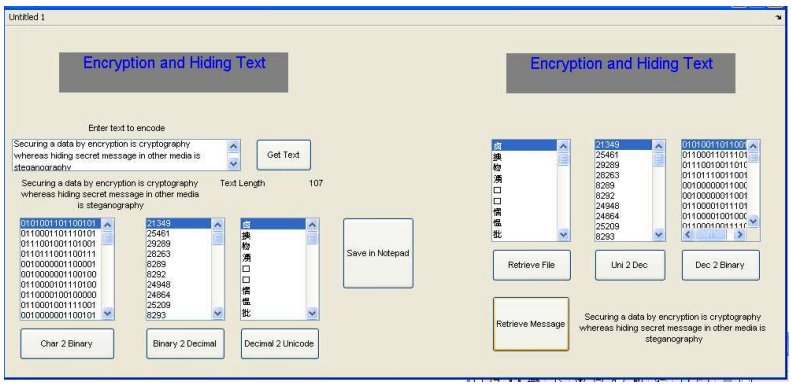


Fig. 8: Method 2, using Unicode Symbols

Advantage of this method is that, two characters are hidden in one Unicode symbol. Here, the science of cryptography and steganography are combined together by encrypting the original message to Unicode values and hiding the same values by mapping to its equivalent Unicode symbols. Security can be provided by rotating the Unicode values by using a key or by changing the sequence of the values in some other order.

In MATLAB, the process of converting the Unicode value from a text file to its equivalent Unicode symbol is achieved by the following built-in routine:

f = load (‘Unicode Decimal Values.txt’); disp (f) char(f)

Decryption and retrieving the original message: Decryption and retrieving the message is as simple as encryption. Each Unicode symbol from a text file is read and converted to its equivalent Unicode value; further the same is converted to binary value of 16 bits. The resultant binary value is divided into two (8 bits each) and its equivalent characters are found. The process is repeated until all the Unicode symbols are converted into normal readable characters. Finally the secret message is derived by combining all accumulated characters. The MATLAB routine to convert the Unicode symbols back to its equivalent Unicode value is shown below:

fid = fopen(‘Unicode Symbols.txt’, ‘rb’);
c = fread(fid, ‘*uint8’); fclose(fid);
str = native2unicode(c,’UTF-16’) double(str)

RESULTS

In the first method, the original message is encrypted and the obtained cipher data is plotted as a point in polar coordinate which allows hiding four characters. Since there is no limitation in plotting the point in a graph, the hiding capacity of the message can be as large as a message to be conveyed in secret. In the second method, two characters are hidden in one Unicode symbol and since the symbols are stored in a text file, there is no limitation for the size of the message to be conveyed in secret.

The implementation of both the methods were carried out successfully using MATLAB. The screeshots of both the methods are displayed in Fig. 7 and Fig. 8.

DISCUSSION

There are a large number of Steganographic methods, which most of us are familiar with ranging from invisible ink and microdots to secreting a hidden message in the second letter of each word of a large body of text etc. With computers and networks there are many other ways of hiding information, such as hiding text within Web pages, Null ciphers etc. This paper describes two methods for integrating together cryptography and Steganography for secure communication using a xy points on a plane and unicode symbols. Methods implemented however are significantly more sophisticated than the examples above.

Most of the combined cryptography and steganographic techniques available today use the pixel bits of an image to hide information and are limited in terms of hiding capacity. Small piece of information can only be embedded in an image carrier because of the limitation of altering more pixels which reduces the intensity of the image and create suspicion to others when passing through an open channel. The proposed method overcomes the above said disadvantages since messages are not hidden by altering the pixel bits of an image. Added advantage is that, to hide n number of characters, only n/2 Unicode symbols are required which increases the hiding capacity. Since text file contains normal Unicode symbols, it is difficult for any

third person to suspect the existence of hidden message and to decode the symbol which makes the system complicated. In summary, in the first method, four characters are hidden in one polar coordinate and in the second method two characters are hidden in one Unicode symbol. These new methods could open new applications for the combined cryptography and steganography which leads to a more secured Internet communication.

CONCLUSION

This study enhances the system security by combining the two techniques of cryptography and steganography. It can also enhance confidentiality of information and provides a means of communicating privately. Here secret message is first encrypted and hidden through media such as graph and Unicode symbols. Since the original message passes through two layers of transformation, the retrieval process is difficult unless the mechanism is known to an intruder. The system security can be further enhanced by using a password at each layer of the system. These methods give the means of hiding data, establishing its authenticity and preventing its undetected modification or illegal use. Furthermore, this presents a method that can transmit large quantities of secret message and provide secure communication between two parties. Any kind of text data can be included as a secret message and is sent over the open channel, the proposed procedure is straightforward and easy to implement. It goes well beyond simply embedding text in an image.

REFERENCES

Cvejic, N., 2004. Algorithms for audio watermarking and steganography.

Houcque, D., 2005. Introduction to MATLAB for engineering students. Northwestern University.

Narayana, S. and G. Prasad, 2010. Two new approaches for secured image steganography using cryptographic techniques and type conversions. Signal Image Process.: Int. J., 1: 60-64.