World of Computer Science and Information Technology Journal (WCSIT)
ISSN: 2221-0741
Vol. 1, No. 4,138-143, 2011

Secured Crypto-Stegano Communication Through Unicode

A. Joseph Raphael
PhD Research Scholar, Karpagam Univesity
Coimbatore, India;
Lecturer in Information Technology,
Ibra College of Technology, Sultanate of Oman
Email: raphaelaj@gmail.com

Dr. V. Sundaram
Director and Head, Department of Computer Applications
Karpagam College of Engineering,
Coimbatore, India
Email: dr.vsundaram@gmail.com

Abstract - Techniques for information hiding are becoming increasingly more sophisticated and widespread. This paper proposes a new alternative method for secured communication to protect digital data from tampering by combining the techniques of cryptography and steganography. Message to be sent is transformed into encrypted form using Unicode symbols and hidden in an image carrier without disturbing its bits. On the other end, extraction algorithm is designed in such a way that the process separates the message and image into two different entities; then reads the extracted message which is in the encrypted form and transforms it from the Unicode symbols to a readable form. The method is defined as undetectable, strong and secured communication of data related to the multimedia image. Thus any confidential message can be send to any target without the knowledge of others through an unsecured communication channel. This encoding and decoding scheme of the proposed new method is significantly different as compared to the traditional schemes.

Keywords: Cipher; Crypto-Stegano object; Unicode Value; Unicode Symbol.

I. INTRODUCTION

Cryptography and Steganography are two popular methods of sending secret information in a secured way. One hides the existence of the message and the other distorts the message itself. These are well known and widely used techniques that manipulate messages in order to cipher or hide their existence respectively. Cryptography is the art and science of protecting information from undesirable individuals by converting it into a form non-recognizable by its attackers while stored and transmitted [1]. Steganography is the art and science of communicating in a way which hides the existence of the communication. Cryptography scrambles a message so it cannot be understood; the Steganography hides the message so it cannot be seen. In this paper we will focus to develop one system, which uses both cryptography and steganography for better confidentiality and security.

A new method is found to hide secret information through Unicode. This is based on programming which generates a Unicode symbol where characters are hidden and appears to be a normal Unicode symbol for human vision. Secret information is encrypted in the form of Unicode and compressed before it is hidden in an image carrier with cryptographic password which makes the communication more secured. This is important because in this way we minimize the size of information to be sent, and it is also easier to hide a random looking message into the carrier than to hide a message with a high degree of regularity. Added strength of the new technique is that, to hide n number of characters n/2 Unicode symbols are required which makes the information hiding capacity large irrespective of the original image data. This combined science of cryptography and steganography could open a new application which leads to more secured communication through an open channel.

II. UNICODE

A. Birth of Unicode

ASCII, a character set based on 7-bit integers is still popular and its provision for 128 characters was sufficient at the time of its birth in the 1960s, the growing popularity of personal computing all over the world made ASCII inadequate for people speaking and writing many different languages with different alphabets. Newer 8-bit character sets, such as the ISO-8859 family, could represent 256 characters. This solution was good enough for many practical uses, but is a bit limiting for all the languages in the world [4]. In the end, the other parts of the world began creating their own encoding schemes and things started to get a little bit confusing. It became apparent that a

new character encoding scheme was needed and the Unicode standard was born.

B. What is 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 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 [4]. 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 [5].

III. PROPOSED SYSTEM

Data hiding techniques have been widely used to hide and transmit secret message for long time. In this system, the plain text is encrypted into a system generated text file with Unicode symbols and then the text file is compressed with the help of cryptographic key. The resultant file is hidden in an image which produces a crypto-stegano object, in the reserve process the resultant object is given as an input to decrypt and to convert into plain text. The system combines the effect of cryptography and steganography to enhance the security of the data. The affine transformation of plain text to crypto-stegano object and the reverse process is depicted in the figure 1 shown below. The system satisfies the requirements for secured data transmission such as confidentiality, integrity and security in an open channel.

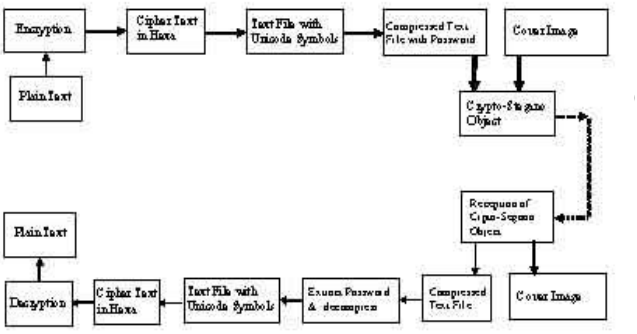

Fig. 1 Plain text to crypto-stegano object and crypto-stegano object to Plain text

A. Message Hiding

The hiding process is done in two stages using the combined science of cryptography and steganography; first, the message is encrypted into Unicode symbols using cryptography and secondly, encrypted text file is hidden in an image carrier using steganography.

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 which occupies more than one byte in memory. The idea of hiding characters in Unicode is derived from the above said scenario of character encoding methods. 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, since Unicode values are in hexadecimal the resultant Unicode value is 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. Below mentioned figure 2 shows the original message entered into the software and hidden through Unicode symbols.

Fig. 2 Text File contains Original message hidden through Unicode Symbols

The resultant text file of any size obtained is compressed by assigning a cryptographic password and an image is selected to hide the compressed text file which acts as a carrier in an open channel.

Fig. 3a Original image before hiding compressed file, image size is 20.9 kb

In recent years, several steganographic programs have been posted on Internet, all those techniques are limited in terms of information hiding capacity. So, in order to hide a large message to be conveyed in secret, an alternative steganographic technique is used to hide the above said compressed file in the image carrier by a copy command which helps to hide the same without disturbing the bits of an image. This produces a crypto-stegano object which appears to be an ordinary image for a human vision as shown in figure 3b.

Fig. 3b. Crypto-Stegano object, image size after hiding file is 21.2 kb

No difference in the image is noticed, but only the size of the image increases to some extent which cannot be identified by a third party because images are available in different sizes depending on the mega pixels as shown in figures 3a and 3b. Therefore, the advantage of this method is that secret messages are encrypted and hidden without disturbing the bits of the carrier which reduces the suspicion from others, whereas several existing methods modify the bits of each pixel to hide the information which disturbs the intensity of an image when more bits are replaced.

B. Algorithm 1

Generating Unicode symbols and hiding message

Input: Secret Message and an image

Output: Crypto-Stegano object contains image as carrier and hidden compressed text file with Unicode symbols.

Steps:

1. Input the secret message to be hidden
2. Extract each character from the message
3. Embed the character in Unicode format
4. Display the equivalent Unicode symbol
5. Save the Unicode symbols in a text file
6. Repeat the process until all characters are converted into Unicode symbols
7. Generate a password
8. Compress the text file to reduce the size of the file
9. Select an image as carrier to hide the file

C. Implementation – Message Hiding

Fig. 4. The process of hiding message is implemented using VB.net

D. Messag Extraction

The hidden message from the crypto-stegano object is extracted using the reverse process of the technique explained in message hiding. The crypto-stegano object is given as an input to the software that allows simple recovery of the innocent image and the compressed text file from which the original encoded message can be extracted, as shown in figure 5. Further the compressed text file is decompressed from which the normal text file is extracted, each Unicode symbol is read from the extracted text file and its equivalent characters are found. The process is repeated until all the Unicode symbols are converted into normal readable characters. Figure 6 shows the secret message derived by combining all accumulated characters.

Fig. 5. Files extracted from crypto-stegano object

E. Algorithm 2 –Message Extraction

Input: Crypto-Stegano object

Output : Secret message in readable form

Steps:

1. Select the crypto-stegano object
2. Discover the password
3. Extract the compressed file from the crypto-stegano object
4. Convert the compressed file to a normal text file
5. Read each Unicode symbol from the text file
6. For each Unicode symbol find the equivalent characters.
7. Repeat step 5 and 6 until all the Unicode symbols are converted into characters
8. Accumulated characters form the secret message

F. Implementation –Message Extraction

Fig. 6. Process of retrieving text

IV. VISUAL BASIC .NET AND UNICODE

Nowadays most of the applications developed are based on globalization; on the other hand all internet application has to support all the languages with respect to the customer's locality.

A. Globalization in Visual Basic.Net

One of the big changes in software, and especially in the move up from VB 6 to VB.NET, is the ability to build globalization into code [6]. Globalization refers to the process with which an application or software will be designed and developed so as to make it run across all platforms and all sites with minimum or no modification to the software application. Under any normal circumstance, there will be two processes in Globalization and they are customization or localization of the application and internationalizing the application codes so as to meet the standards of the local culture and other related matters [7].

There is a whole list of things that need to be changed in programs to make it acceptable to other cultures. They include

- language differences such as different fonts and right-to-left instead of left-to-right writing styles
- different calendars
- format patterns for dates, currency, and numbers
- the sort order for strings

All of these can be customized using dozens of different objects in the System.Globalization namespace.

Another key technology for globalization is the Unicode standard. This standard was originally created by a worldwide industry group and has now been recognized by the World Wide Web Consortium (W3C). VB.NET is fully Unicode compliant. The key advantage of Unicode is that you can display virtually all characters in all languages.

B. Unicode Encoding and Decoding routines

The implementation part of hiding text and retrieving text is done using vb.net because of the globalization to support and display all characters in all languages of the world which is provided by Unicode standard. After hiding the text, the process of converting Unicode value into Unicode symbol is attained by the following code snippets.

```
Private Sub btnunisymbol_Click(ByVal sender As
System.Object, ByVal e As System.EventArgs)
Handles btnunisymbol.Click
Dim i As Integer
Dim intCode As Integer
For i = 0 To n - 1
intCode = CInt("&H" & unihexa(i))
txtunisymbol.Text &= ChrW(intCode)
Next
End Sub
```

Fig. 7. Code to convert from Unicode value to Unicode symbol

On the other end, the message received in a text file is in the form of Unicode symbols which has to be converted into a readable form. The process of converting the Unicode symbol into Unicode Value is achieved through the following function:

```
Private Sub Button1_Click_1(ByVal sender As
System.Object, ByVal e As System.EventArgs) Handles
btnuni2hexa.Click
  txthexavalue.Text = ConvUnihex(txtdisplay.Text)
End Sub
Function ConvUnihex(ByVal inx As String) As String
  Dim i As Integer
  Dim Conv_Unihex As String = ""
  For i = 0 To inx.Length - 1
  Conv_Unihex +=
Xformat(Hex(Microsoft.VisualBasic.AscW(inx.Chars(i))))
  Next
  Return (Conv_Unihex )
End Function
```

V. SECURITY CONSIDERATIONS

A cryptographic system uses two keys - a *public key* known to everyone and a *private* or *secret key* known only to the recipient of the message. In the proposed new method, an important element to the public key system is that the public and private keys are related in such a way that only the public key can be used to derive private key and only the corresponding private key can be used to extract and decrypt the hidden text.

A text is agreed upon between the sending and receiving parties beforehand to generate password. The agreed upon text is rotated randomly to generate password, based on the random number generated its equivalent ASCII code is displayed as key which acts as a public key (primary key) in retrieval process. Since the public key is generated randomly by the software, to the receiving end, the primary key is sent along with the filename of the crypto-stegano object. In the retrieval process, agreed upon text and the equivalent ASCII code of public key is entered into the software to extract the private key (password). Since private key is generated randomly on every occasion by the software the complexity rate of discovering the same is amplified which is an added security to the system.

VI. ADVANTAGE

Most of the steganographic techniques available today use the pixel bits of an image to hide information and are limited in terms of information 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. In this new method, since information is stored in a text file and hidden in an image, its hiding capacity can be as large as a message to be conveyed in secret. 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 contain Unicode symbols, it is difficult to decode the symbol which makes the system complicated in retrieval of message for an unknown person. The text file is hidden in an image without modifying the bits of the pixel, hence the image intensity is not disturbed and avoids suspicion while transmitted over an open channel. This could open new applications for steganography leading to a more secure Internet communication age.

VII. CONCLUSION

As cryptography and steganography becomes more widely used in computing, there are issues that need to be resolved. There are a wide variety of different techniques with their own advantages and disadvantages. Many currently used techniques are not robust enough to prevent detection and removal of embedded data. The use of benchmarking to evaluate techniques should become more common and a more standard definition of robustness is required to help overcome this. For a system to be considered as robust it should have the following properties [2] such as (i) the quality of the carrier should not visibly degrade upon addition of a secret message (ii) secret message should be undetectable without secret knowledge, typically the key (iii) the secret data should survive attacks that don't degrade the perceived quality of the work.

This work presents a scheme that can transmit large quantities of secret information and provide secure communication between two parties. Steganography with cryptographic password can be woven into this scheme to make the detection more complicated. Any kind of text data can be employed as secret message and is sent over the open channel. In addition, the proposed procedure is simple and easy to implement.

REFERENCES

[1] B.B.Zaidan, A.A. Zaidan, A.K. Al-Frajat and H.A. Jalab,"On the Differences between Hiding Information and Cryptography Techniques: An Overview", Journal of Applied Sciences, 2010, ISSN 1812-5654.

[2] Shashikala Channalli and Ajay Jadhav: Steganography An Art of Hiding Data, "International Journal on Computer Science and Engineering", Vol.1 (3), 2009, 137-141.

[3] Venkata Sai Manoj, "Cryptography and Steganography", International Journal of Computer Applications, Volume 1 – No. 12, 2010, 0975-8887.

[4] Michal Kosmulski, "Introduction to Unicode", http://www.linux.com/archive/articles/39911

[5] The Unicode Consortium, http://www.unicode.org

[6] Dan Mabbutt, "Globalization in Visual Basic .NET", http://visualbasic.about.com/od/usingvbnet/a/globvbnet1.htm

[7] Understanding Globalization in .NET, http://www.dotnet-guide.com/globalization.html

AUTHOR PROFILE

A. Joseph Raphael obtained his Master degree in Computer Science from St. Joseph's College, Tiruchirapalli and Master of Philosophy from Alagappa University, Karaikudi. Currently, he is a PhD research scholar at Karpagam University, Coimbatore, India and also working as a lecturer in the department of Information Technology, Ibra College of Technology, Sultanate of Oman.

Dr. V. Sundaram earned his PhD in mathematics from Madras University. He is a research guide of Anna University, Coimbatore and Karpagam University in the field of computer science and computer applications. He is currently guiding several PhD students in the areas of theoretical computer science, network security, cryptography and data mining. He has published several papers in national and international journals and organized 5 national conferences. He is a life member of ISTE and member of Computer Society of India.