Steganography– Information Hiding in Source Code Langauge

Veenu Arora[#1],Raninder dhillon[#2]
Computer Science Department, Punjab Technical University
[1]*House No.15/16 Street No.8 Ferozepur Cantt Ferozepur ,Punjab India*
[2]*Guru Nanak Dev Engineering CollegeLudhiana,Punjab,India*

Abstract-In the research area of text steganography,algorithm based on case insensitive langauge has great advantage of great capacity but in case of case sensitive langauges less work has been done.Based on the fact that there is change in case sensitive langauge to embed secret bits inside source code langauge.As in case of html secrets bits are embed in html tags not in data which is resprented on web page.So,in this research paper some secrets bits are embed in c++ source code.So this reasearch basically work on source code langauge.So this consider embedding in comments only.It aso consider the position where embedding can take place. Like data is embed in comments at different positions.

Keywords:Encryption,Decryption,,InformationHiding,Steganography,Text Steganography

I. INTRODUCTION

The Technique for information hidng has been widely applied on various fields during recent years [12] and two major branches viz. Digital watermarking and text stegnography have been derived.[13][14] **Steganography** is the art and science of encoding hidden messages in such a way that no one, apart from the sender and intended recipient, suspects the existence of the message. It is a form of security through obscurity. The word *steganography* is of Greek origin and means "concealed writing." It combines the Greek words *steganos* (στεγανός), meaning "covered or protected," and *graphei* (γραφή) meaning "writing." The first recorded use of the term was in 1499 by Johannes Trithemius in his *Steganographia*, a treatise on cryptography and steganography, disguised as a book on magic. Generally, the hidden messages will appear to be (or 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.This paper reprsent the data hiding in case sensitive langauge. Section II throws light on text steganography Literature Review SectionIII describes the proposed approaches. Section IV explain problem formulation the proposed methods. Section V discusses the methodology and implementation. Section VI draws the result and discussion. Section VII draws the conclusion.

II. LITERATURE REVIEW:

There are various techniques are used in text steganography Each has some strengths and weakness.
SandipanDey et all[3], explain about Embedding Secret Data in Html Web Paget his basically work on web document how data is embed in web pages.

Hameed Al-Qaheriet all[9] explain about Hiding inside Html and Other Source Codes in which how data is embed in html tags and it consider both case sensitive and case-insensitive languages. How it works in case insensitive languages.
L. Y. POR1 et all[8] it explains about Information Hiding: A New Approach in Text Steganography this research paper explain about the various methods of text steganography and how text is embed in various documents. Our method offers dynamic generated stego-text with six options of maximum capacity according to the length of the secret message.
H Singh et all[10] explain about a survey on text steganography various techniques are used for text hiding in steganography.
M Garg[11] research on A novel text steganography technique based on html document In this paper he explain about how html data is embed by consider width and height as same thing.How we embed secret bits in tags of html document
In this paper some text is embed in C++ source code because less work is done in case sensitive languages. So in new algorithm works in embedding secret bits in C++ source code languages.

III . HIDING DATA INSIDE C++
I. Exploiting the Case-sensitivity

As we know that C++ is a case sensitive language so consider the text of CPP source code file. In this research paper we embed the some secret bits inside source code which is not visible to any other intruders. Like we have to embed secret message in html in similar way we embed some secret message in C++ source code. We have to embed secret bits in the source of C++ by changing some bits and it will convert combination of upper alphabets to lower alphabets.Embedding can be done in comments. Input is in lower form but output in upper bits by embeds some bits.

The Algorithm for Hiding Data

The algorithm for embedding the secret message inside the C++ cover text is very simple and straight forward. First, we need to separate out the characters from the cover text that will be candidates for embedding; these are the case sensitive text characters inside C++. Figure II shows how html shows very simplified automata for this purpose. Also, let us define the following functions before describing the algorithm:

- $1 : \sum^* \rightarrow \sum$ as:

$l(c) = \{$ ToLower(c) $c \in \{`A'...`Z'\}$
c   otherwise$\}$
whereToLower(c) = c + d

- u: $\sum* \rightarrow \sum$ as:

$u(c) = \{$ ToUpper(c)   c $\in \{`a'...`z'\}$
c   otherwise$\}$
whereToUpper(c) = c-d

- Here d = 'A' - 'a'.

The ascii value of 'A' = 65 and the ascii value of 'a' = 97, with a difference d = 32.

It's easy to see that if the domain$\sum* = \{`a'...`z'\}$ U $\{`A'...`Z'\}$, then l :$\{`A'...`Z'\} \rightarrow \{`a'...`z'\}$ and u : $\{`a'...`z'\} \rightarrow \{`A'...`Z'\}$, implies that l(.) = u(.) = $\sum*$- l(.)

Now, we want to embed secret data bits $b_1 b_2..b_k$ inside the case-insensitive text inside the Comments in C++. If c_1 c_2 ....c_n denotes the sequence of characters inside the C++ source code in cover text (input C++). A character c_i is a candidate for hiding a secret message bit iff it is an alphabet. If we want to hide the j_{th} secret message bit b_j inside the cover text character c_i, the corresponding stego-text will be defined by the following function fstego:

V$c_i \in \{`a'...`z'\}$U$\{`A'...`Z'\}$,i.e. ifIsAlphabet(c_i) is true,
$f_{stego}(c_i) = \{$ $l(c_i)$ $b_j = 0$
$u(c_i)$ $b_j = 1\}$,
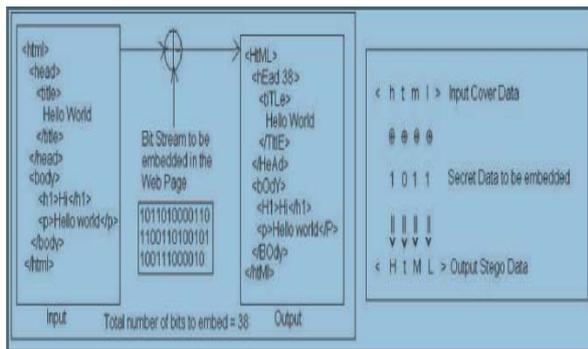
FigureI.how html change in tags by embedding secret bits

Hence, we have the following:
$c_i \in \{`a'...`z'\}$U$\{`A'...`Z'\} => f_{stego}(c_i) = l(c_i)^- b_j + u(c_i)b_j$,V$_i$ ....(1)

The number of bits (k) of the secret message embedded into the html cover text must also be embedded inside C++ source file.

III.The Algorithm for Hidden Data Extraction:

Again, the algorithm for extraction of the secret message bits will be even simpler. As in the embedding process, we must first separate out the candidate text that were chosen in the earlier step for embedding secret message bits. Also, we must extract the number of bits (k) embedded into this page. In order to find out the stego-text, one has to use 'view source'.Now, we have d_i = $f_{stego}(c_i)$; V$_i$ $\in$ $\{1,2,....,n\}$.If $d_i \in \{`a'...`z'\}$ U $\{`A'...`Z'\}$ i.e., an alphabet, then only it is a candidate for decoding and to extract b_i, from d_i, we use the following logic:
bi =$\{0$ $d_i \in \{`a' ... `z'\}$
     1 $d_i \in \{`A' ... `Z'\}\}$

Repeat the above algorithm V$_i$< k, to extract all the hidden bits.
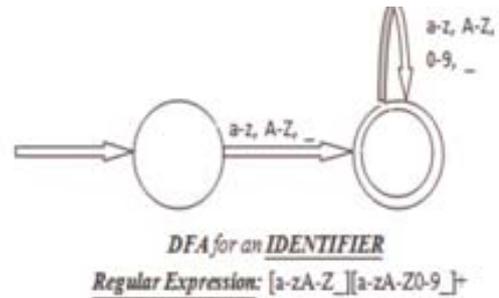
FigureII. DFA for accepting identifier

IV. PROBLEM FORMULATION

Web based communication has a great amount of bandwidth and hence can be used for secret communication. This research works on case sensitive languages like C++. A lot of work is done in case insensitive language like html and cascading style sheet and also in xml document but less work is done in case sensitive languagelike c++. This research aims to propose a new technique on hiding data in case sensitive languages like C++. By using the techniques of case-insensitive language this research work is done in hiding textual data for case sensitive languages like C++ and java. This research need some advantage that how to hide inside case sensitive language because if it can happen in case sensitive language it can also hide in case sensitive language. We implement a technique that can hide text inside source code. This can happen in either in variable and comments in source code. This research paper deals in hiding textual data or some secret bits in source code like in C++.

V. METHODOLOGY AND IMPLEMENTATION

This research paper proposes a new method of hiding data in source code language, C++. In addition to data hiding, security is also taken into consideration by encrypting the data before hiding. With this approach, even if somehow person retrieves the hidden data, he will be unable to understand it. So procedure followed is

1. First this method consider source file to load and embed the secret message in the source file.
2. Then it will encrypt the secret message.
3. Then collect the data be in ASCII values and after that convert it into binary form and it will hide in C++ source code bit by bit.
4. Combination of 8 bits is taken and embeds accordingly by changing some bits and by adding some bits of message.
5. On decryption side, convert these bits in ASCII value and then change accordingly by upper and lower alphabets according to the size of the length given by the user.
6. Output is the message which is embedded in source code file.

The output is same as the input given by the user. But some data is changed in case of comments. If comments are there in file then do the processing of bits to embed secret data. If not there do nothing.

Encryption Module:

1. Search for all the comments present in theC++ cover text and extract all the characters c1c2....cn from inside those source code using the DFA
2. Embed the secret message length k inside C++ source code in the stego text.
3. J← 0.
4. for ci €comments i = 1 …..n do
5. if ci € {'a' . . . 'z'}U {'A' . . . 'Z'} then
6. fstego(ci) = l(c_i)⁻b_j+u(c_i)b_j .
7. j←j + 1.
8. else
9. fstego(ci) = ci.
10. end if
11. if j == k then
12. break.
13. end if
14. end for

Decryption Module:

1. Search for all the comments present in the C++ stego text and extract all the characters d_1d_2....$d_n$from inside those comments using the DFA described in the figure2.
2. Extract the secret message length k from inside Comments in the stego text.
3. j← 0.
4. for di €Comments; i = 1 …..n do
5. if d_i€{'a' . . . 'z'} then
6. b_j = 0.
7. j <-- j + 1.
8. else if d_i€ {'A' . . . 'Z'} then
9. b_j = 1.
10. j←j + 1.
11. end if
12. if j == k then
13. break.
14. end if
15. end for

VI. RESULT AND DISCUSSION:

In this research paper the result is based on case sensitive languages that is C++ source code. The result is that the data in comments in C++ is changed by inserting some bits in comments the comments are changed from lower to upper alphabets in C++ source code. The results are as follows:

1. In the first case the user have to enter the message to encrypt the data in ADD.C file after that we embed the data in the source code and the output in combination of lower and upper bits and the source code is represented as follows in comments these are in combination of lower and upper bits and after decryption output is as same as input given at the time at start of program and decrypted is same as encrypted message :
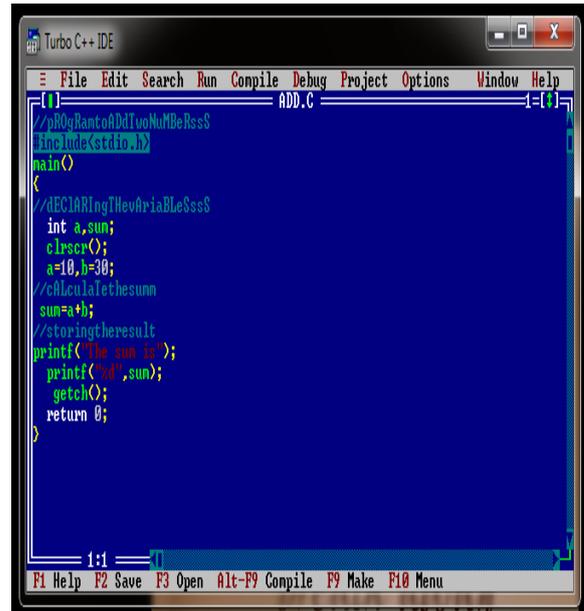
Figure3: comments changes in upper and lower case.

2. In the second case the data is more as compare to first source code data is embedded in some large source code and with encryption and decryption data is embedded in source code of given data some input is given and output is same as the give input which is as shown in figures :
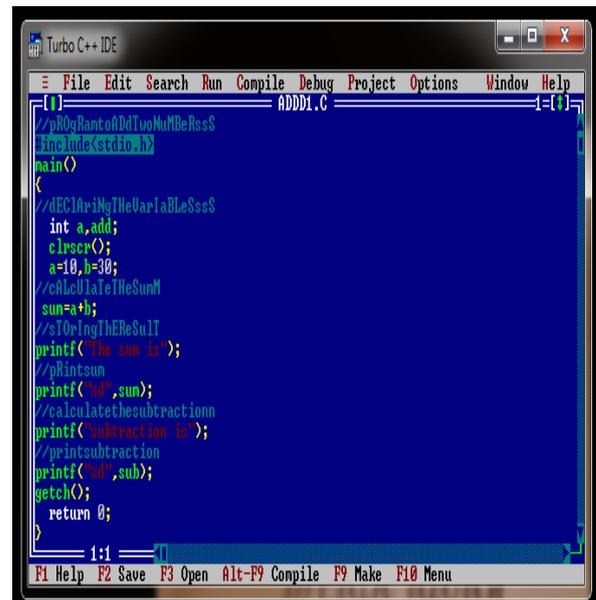
Figure4: comments change from lower to upper some bits are changed.

3. In the third case the user has to enter the text i.e. small program is there for which data is embed in small source code in comments hiding inside only in comments. This image is shown as follows:
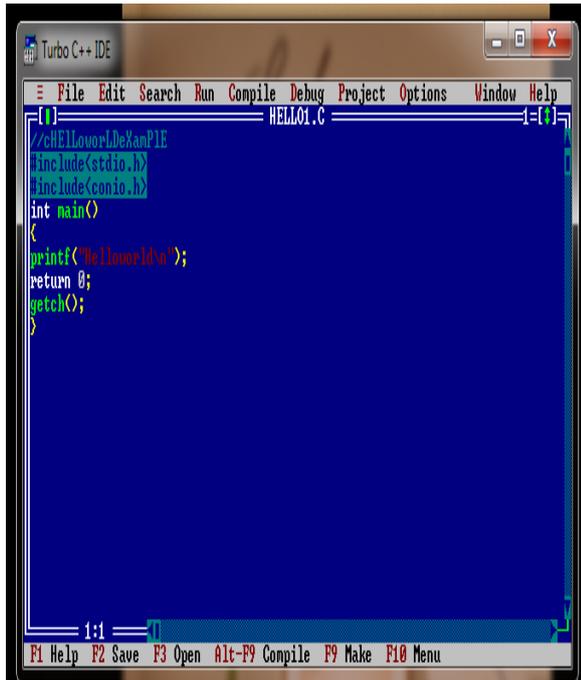
Figure5.Comments in small programs

4. In the last case the user has to enter some large program in which embed the data in matrix multiplication program we get the following results and encryption and decryption at the end of this give output as in the multiple comment in the given source code is as follows :
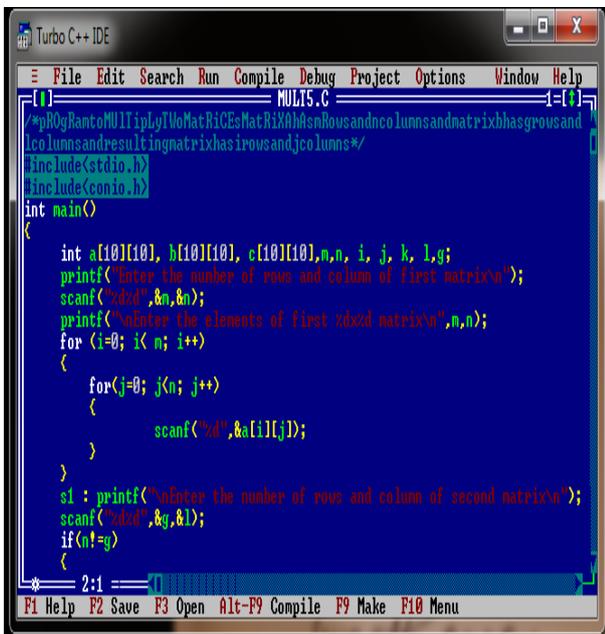
Figure 6. Comments in big program

VII. CONCLUSION:

As less work has been done in case sensitive languages so this research has been done in case sensitive languages in C++. I have research on case sensitive language ie I have work on C++ language in which I hide the text inside comments in source code This is a simple technique made to hide inside source code language .it will help in other languages which are case sensitive.

Future work : This work can be extended to any source code language which are in case sensitive language. This can be implemented in other source languages. This can embed in other languages. Data hiding in case of images is somewhat distorted images but in case of text these problem does not occurred.

REFERENCES:

[1] Chi-Kwong Chan, L.M. Cbeng, "Hiding data in images by simple LSB substitution," Pattern Recognition, Vol. 37, pp. 469-474, 2004.

[2] F. Battisti, M. Carli, A. Neri, K. Egiaziarian, "A Generalized Fibonacci LSB Data Hiding Technique," 3rd International Conference on Computers and Devices for Communication (CODEC- 06) TEA, Institute of Radio Physics and Electronics, University of Calcutta, December 18-20, 2006.

[3] SandipanDey, Hameed Al-Qaheri, SugataSanyal,"Embedding Secret Data in Html Web Page," 1st International Conference on Image Processing and Communications, September 16-18, 2009, Bydgoszcz, Poland. Published in Book: Image Processing & Communications Challenges, Ed. Ryszard S. Choras and AntoniZabtudowski, Academy Publishing House EXIT, Warsaw 2009, pp. 474-481. ISBN: 978- 83-60434-62-8.

[4] SandipanDey, Ajith Abraham, BijoyBandyopadhyay and SugataSanyal, "Data Hiding Techniques Using Prime and Natural Numbers," Journal of Digital Information Management, ISSN 0972-7272, Volume 6, No 3, 2008,pp. 463-485.

[5] SandipanDey, Ajith Abraham and SugataSanyal, "An LSB Data Hiding Technique Using Natural Numbers," IEEE Third International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IIHMSP 2007, Nov 26-28, 2007, Kaohsiung City, Taiwan, IEEE Computer Society press, USA, ISBN 0-7695-2994-1, pp. 473-476, 2007.

[6] SandipanDey, Ajith Abraham and SugataSanyal, "An LSB Data Hiding Technique Using Prime Numbers," Third International Symposium on Information Assurance and Security,August 29-31, 2007, Manchester, UnitedKingdom, IEEE Computer Society press, USA,ISBN 0-7695-2876-7, pp. 101-106, 2007.

[7] K. Bennett, "Linguistic Steganography:Survey, Analysis, and Robustness Concernsfor Hiding Information in Text", PurdueUniversity, CERIAS Tech. Report, 2004.

[8]L. Y. POR1, B. Delina2 "Information Hiding: A New Approach in Text Steganography" Faculty of Computer Science and Information Technology, University of Malaya, 50603, Kuala Lumpur, MALAYSIA

[9] Hiding inside Html and Other Source CodesHAMEED AL-QAHERI ,SANDIPAN DEY ,SUGATA SANYALDepartment of Quantitative Methods and Information Systems,College of Business Administration Kuwait Universityalqaheri@cba.edu.kwCognizant Technology Solutionssandipan.dey@gmail.com School of Technology and Computer Science,Tata Institute of Fundamental Research,HomiBhabha Road, Mumbai - 400005, India sanyal@tifr.res.in

[10]H Singh et all explain"A survey on text steganography"

[11]M Garg[11] research on "A novel text steganography technique based on html document"

[12] R. J. Anderson F. A. P. Petitcolas and M. G. Kuhnl.

[13] S. Katzenbeisser and F. A. P. Petitcolas.Information Hiding Techniques for Steganography and Digital Watermarking. Artech House, London,2000

[14] Z. Duric N. F. Johnson and S. Jajodia.Information Hiding: Steganogra-phy and Digital Watermarking - Attacks and Countermeasures. KluwerAcademic, 2001