Hiding in the Plain Text: A Critical Analysis of Whitespace Steganography

Eswara Sai Prasad Chunduru[1], Nagendar Rao Koppolu[2]

[1]Assistant Director, Digital Forensic Division, Central Forensic Science Laboratory, Hyderabad
[2]Inspector of Police (In-charge State Cyber Vertical), Telangana Police Department, Hyderabad

---***---

Abstract: *Steganography is a technical method of sending hidden messages of any nature under a cover. The cover may be of any file. The hidden message may be plain text or any data in the form of cascade of bits. Steganography offers users to protect their privacy while providing security over the normal communication channel. In the present paper, while overviewing the pros and cons of the various text steganography techniques, the advantage of white space steganography, WhiteSteg is emphasized.*

Keywords: WhiteSteg, Steganography, stream of bits, etc.

1. INTRODUCTION

In the current digital world, with more emphasis on the security and privacy over the public communication networks, the use of Cryptography technics, such as Internet Protocol Security (IPsec), Secure Socket Layer (SSL) and End to End Encryption with PKI has become the order of the day. These techniques are, however, the most sought-after ones, involving the highest complex algorithms. Cryptography, by its nature, changes (encrypts) the original data, the plaintext, into an unreadable representation of random scrambling characters, the ciphertext. This can raise the attention of any surveillance mechanism for attempting to access confidential data. Furthermore, since the cryptography algorithms are available to the public and the security is based on the key used, the attacker can decrypt (read) the data successfully if the cryptographic key is compromised. Therefore, steganography can be used as an additional mechanism to provide incremental secrecy to confidential data to overcome this problem.

Steganography hides the data within another data using two functions: embedding (using any embedded algorithm/steganography tools) and extraction (using steganalysis tools) as depicted in the following Fig -1:

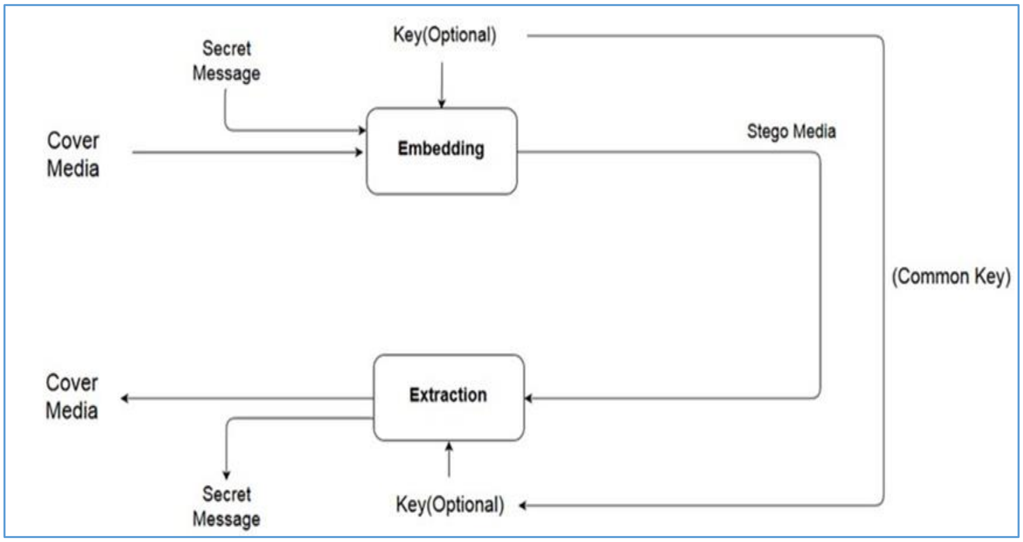


Fig – 1: Steganography [13]

Since the stego media appears to be a normal file in the communication channel, only recipients aware of the steganography technique can retrieve the hidden data from the stego media. While appearing to be normal over the communication channels, this distracts the surveillance techniques. Due to this, steganography can be used in aggregation with cryptography to provide an extra layer of security for privacy and confidentiality.

In general, in cryptography the focal is to protect the contents of a message; while steganography is targeted towards concealing the fact that a secret message is being sent and its contents. Following Table-1 shows a comparison of both the cryptographic and steganographic methods of data security:

Table 01 – Comparison of Cryptography and Steganography

Technique	Cryptography	Steganography
Application	Secret Information	Secret Communication
Principle	Plain Text + Key = Encryption = Cipher Text	Cover File + Secret Message + Key = Embedding = Stego File
Visibility	Visible	Invisible
Requirement	Robustness	Undetectability Capacity

2. Types of Steganography

2.1 Text Steganography:

It comprises concealing information inside the text documents. The mystery information is taken cover behind each nth letter of each expression of a text message. Quantities of methods are accessible for concealing information in a text record.

These methods are 1) Format Based Method 2) Random and Statistical Method 3) Linguistic Method [1].

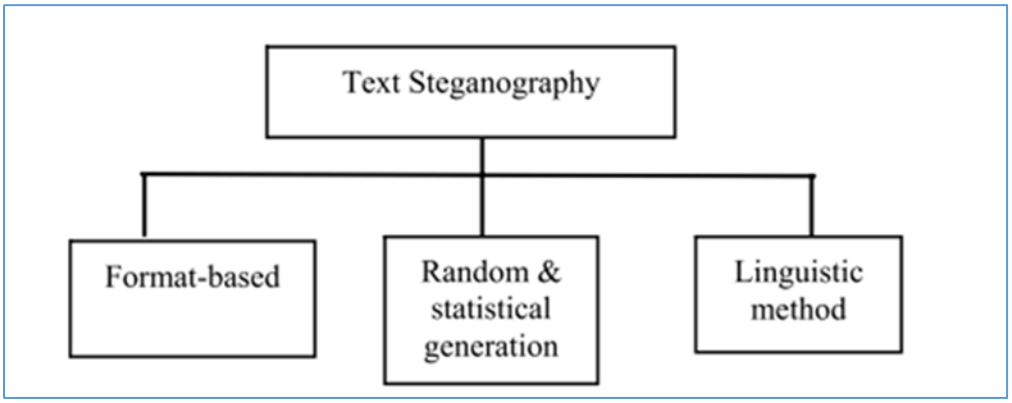


Fig – 2: Three Basic Categories of Text Steganography [12]

2.1.1 Format Based Method:

In Format-based techniques, physical text formatting of text is done at a place in which the information is to be hide. Insertion of spaces at the word spaces and at the last of sentences, deliberate misspellings, and resizing the fonts throughout the text are some of the methods used in this method. However, these format-based methods are not identified with the human visual system, but they can be detected using a computer system (Bennett). White Space Steganography, WhiteSteg falls under this category [2].

2.1.2 Random and Statistical Method:

In Random and statistical generation techniques, a cover text is generated based on statistical properties. This method is based on character and word sequences. Within character sequences the information to be hide is embedded on the information to appear in a random sequence of characters. The sequence is interpreted as a random one while anyone intercepts the message . It can be also done by taking the statistical properties of word length and letter frequency to create "words" (without lexical value) that leverage same statistical count as the original one . While concealing the information a codebook with mapping between lexical items and sequence of bits or dictionary words is used.

2.1.3 Linguistic Method:

The linguistic method, considers the linguistic properties of generated and modified text, frequently uses linguistic structure as a placeholder for hidden messages. Steganographic data can be hidden within the syntactical structure itself.

The following Fig.-3 shows the mechanism of text-based steganography [1][3] :

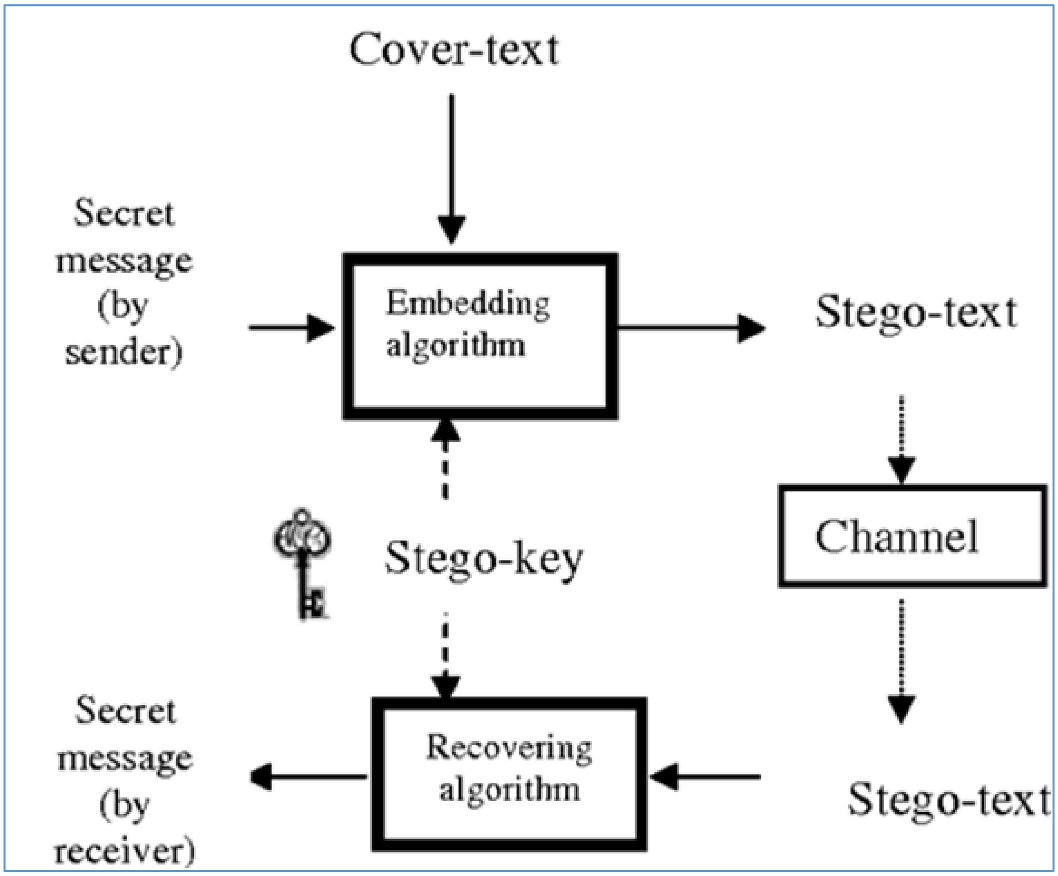


Fig – 3: Mechanism of Text Steganography [11]

2.2 Image Steganography

Hiding the information by taking the spread information as an image is known as image steganography. In image steganography, pixel powers are used to cover the information.

2.3 Audio Steganography

It includes concealing information in audio records. This technique conceals the information in WAV and MP3 sound records. There are diverse methods of audio steganography. These methods are 1) Low Bit Encoding, 2) Phase Coding, 3) Spread Spectrum.

2.4 Video Steganography

It is a system of concealing any records or information into advanced video format. For this, a video which is stack of picture frames is utilized as a bearer for concealing the information. For the most part, Discrete Cosine Change (DCT) adjusts the qualities (e.g., 8.667 to 9) utilized to shroud the information in every one of the images in the video, which is unnoticeable to the human eye. Mp4, MPEG, AVI are the formats utilized by video steganography.

2.5 Network or Protocol Steganography

It includes concealing the information by taking the network convention, for example, TCP, UDP, ICMP, IP and so on, as a spread item. In the OSI layer network display, there exist undercover channels where steganography can be used.

3. Applications of Steganography

- Using steganography, a secret, outline, or other delicate information can be transmitted without cautioning any potential aggressors.
- Tracking the printouts to their origination by adding secret tracking dots is another application of steganography.
- Feature Tagging Elements can be inserted inside an image, as the names of people in a photograph or areas in a guide providing a channel of copyright element.
- Copy the Stego-image likewise duplicates the majority of the installed features and just gatherings that have the disentangling stego-key will most likely concentrate and view the features
- Confidential Communication and Secret Data Storing: Steganography provides us with-
    - Potential capability to hide the existence of confidential data.
    - Hardness of detecting the hidden (i.e., embedded) data.
    - Enhancing the secrecy of the encrypted data.
- In most steganography algorithms, the embedded data is present in an extremely fragile manner. This feature provided a means of Protection from Data Alteration, such as a "Digital Certificate Document System", providing "no authentication bureau requirement."
- Steganography can be used as an Access Control System for Digital Content Distribution over Internet through public channels via an access key to specified or intended uses.
- The other application of steganography is found in Media Database Systems. Without losing the metadata or the annotation data, the media files can be grouped and shared among the users over the communication channels. This, in some scenarios, facilitates in finding the origination and authorship of the media content.

4. White Space Steganography

White space steganography is format-based text steganography. The open space method is one of the first used methods to hide data in white space between words, lines and paragraphs; this method is divided into three stages which are as follows [2][4] :

4.1 In first stage, Encoding of binary message into a text by placing either one or two spaces after each terminating character [4].

4.2 Encoding data by inserting spaces at the end of lines; two spaces encode one bit per line, four encode two, and eight encode three (4).

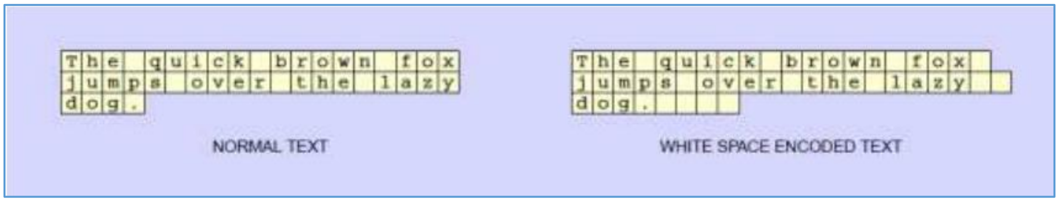
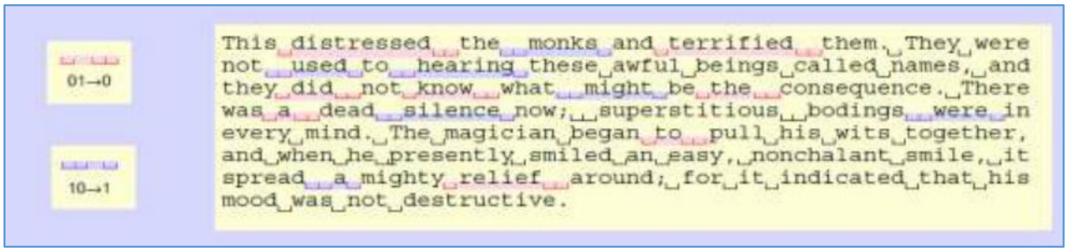


Fig – 4: Data hiding using white space [4]

4.3 Encoding binary data by taking advantage of the justified format of the text by indicating where the extra space is and set one space as "0" and two spaces as "1".



This distressed the monks and terrified them. They were
not used to hearing these awful beings called names, and
they did not know what might be the consequence. There
was a dead silence now; superstitious bodings were in
every mind. The magician began to pull his wits together,
and when he presently smiled an easy, nonchalant smile, it
spread a mighty relief around; for it indicated that his
mood was not destructive.

Fig – 5: Data hidden through justification [4]
(Text from a Connecticut Yankee in King Arthur's Court by Mark Twain)

The issues in this open space method are (4):

- A very large cover text file is required to hide small text as it takes eight spaces available in the cover text for each character in the secret text.
- In the justified format of hiding the data, 1 byte for each character with one space per bit is required and all the spaces are not available and can be used to hide data.

Above issues can be passed by changing the size, font and pixels of the secret text.

5. Implementation

According to Bender, "soft-copy text is in many ways the most difficult place to hide data." This is because the reader can easily notice extra characters or punctuation in a document. Various methods can do the implementation of whitespace steganography. The most popular one is Stegsnow (SNOW) (Steganographic Nature of Whitespace). SNOW is a free program for non-commercial use available on the Internet and is authored by Matthew Kwan. According to him, this program (SNOW) is used for concealing messages in text files by appending tabs and spaces on the end of lines, and for extracting messages from files containing hidden messages. Most of the text viewing applications don't show these hidden characters and hiding the reality that extra information is available. SNOW uses a 64-bit block cipher to encrypt the messages. SNOW uses the elementary Huffman encoding for compression, where the tables are optimized for English text. SNOW has various platform support. The tool can be downloaded from https://sbmlabs.com/notes/snow_whitespace_steganography_tool/.

- Stegsnow conceals messages in ASCII text by appending whitespaces to the end of lines. As tabs and spaces are not visible in text viewer applications the casual observers are unable to sense the hidden content. When built-in encryption is used, such messages cannot be read, even if they are detected.
- SNOW exploits the Steganographic Nature of Whitespace, locating trailing whitespace in text.
- It uses methods of embedding and extraction that are applied on the .txt file. The main embedding approach is based on modifying the whitespaces between the characters to hide the secret message characters that are mapped into binary format, represented by whitespace.
- A common key between the sender and recipient is used to shuffle the place of whitespace in each embedding and provide different character-binary mapping; it can be more difficult for an attacker to guess the hidden data characters.
- The embedding and extraction functions are explained in detail as below:
  - Embedding: Initially, the metadata, i.e., number of words, number of whitespaces and file size of the cover text are identified. This information is important as it helps in determining if the secret message can be embedded within the file or not. Since the whitespace between the words in a line is the main approach, the number of characters that can be used for hiding the data will be computed accordingly.
  - The secret message to be hidden within the carrier file is to be chosen by three conditions for this review:

        - The secret message to contain only English alphabets.
        - The secret message is to be short in length.
        - Only the lower-case English alphabets are to be considered.
    - Extraction: On the other hand, the recipient applies the same procedure but reversely to extract the secret message from the stego text. Furthermore, since using the first five whitespaces in each line is known by both parties through developed software, the correct key is needed to be chosen to decode the hidden data from the stego file. It is important to highlight that the result will be wrong if the key is chosen differently, generating random characters.
    - The key is shared between two parties before using different communication channels such as the Internet, telecommunications in a secure manner. Lastly, based on the modified whitespaces, once the binary values are generated, they will be converted to some values, which are mapped to the proper characters, based on the common key to retrieve the hidden data successfully.
- By appending sequences of up to 7 spaces interspersed with tabs, the data is concealed in the text file. This usually allows 3 bits to be stored for every eight columns. An alternative encoding scheme of using alternating spaces and tabs to represent zeroes and ones was rejected as it requires more columns per bit while uses lesser bytes (4.5 vs 2.67) [14].
- The compression that SNOW exercises is not viable in cases of large data sets and if the data is not a text based one. In these scenarios any external compression application like winzip will help in achieving the task[14].
- SNOW supports encryption using the ICE encryption algorithm in 1-bit cipher feedback (CFB) mode. Passwords of any length up to 1170 characters are supported (since only 7 bits of each character are used, keys up to 1024 bytes are supported). If a message string or message file are specified on the command line, SNOW will attempt to conceal the message in the file if specified [14].
- The resulting file will be written to an out file if specified or to standard output if not specified. If no message string is provided, SNOW attempts to extract a message from the input file. The result is written to the output file or standard output.
- The various options available in SNOW are as below [14]:

OPTIONS :

-C Compress the data if concealing, or uncompress it if extracting.

-f message-file : The contents of this file will be concealed in the input text file.

-l line-len : When appending whitespace, stegsnow will always produce lines shorter than this value. By default, it is set to 80.

-m message-string : The contents of this string will be concealed in the input text file. Unless a newline is somehow included in the string, a newline will not be printed when the message is extracted.

-p password : If this is set, the data will be encrypted with this password during concealment or decrypted during extraction.

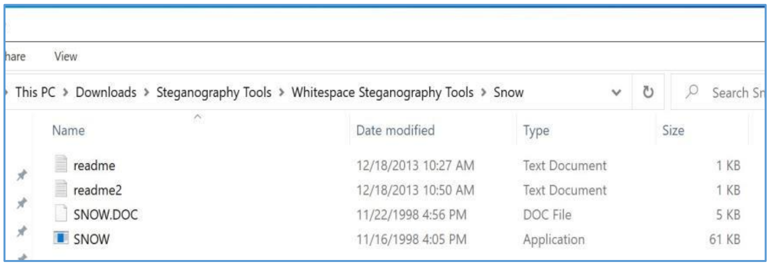
-Q Quiet mode: If not set, the program reports statistics such as compression percentages and the amount of available storage space used.

-S Report: on the approximate amount of space available for a hidden message in the text file. Line length is taken into account, but other options are ignored.

-V, --version : Display version information and exit.

-h, --help : Display usage information and exit.

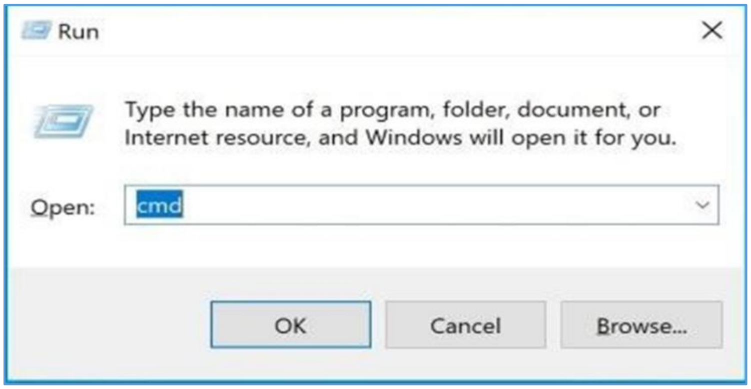
- The various steps in implementing whitespace steganography by SNOW are listed below:
- Step 1: Point to the location of the application on the computer.


hare
View
This PC › Downloads › Steganography Tools › Whitespace Steganography Tools › Snow
Search Sn

Name	Date modified	Type	Size
readme	12/18/2013 10:27 AM	Text Document	1 KB
readme2	12/18/2013 10:50 AM	Text Document	1 KB
SNOW.DOC	11/22/1998 4:56 PM	DOC File	5 KB
SNOW	11/16/1998 4:05 PM	Application	61 KB

Fig – 6: Contents of the Application Folder

- Step 2: Press Window + R to open run and type cmd to open the command prompt.

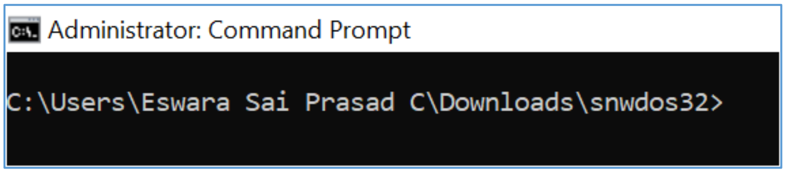


Type the name of a program, folder, document, or Internet resource, and Windows will open it for you.

Open: cmd

Fig – 7: Calling the Command Prompt

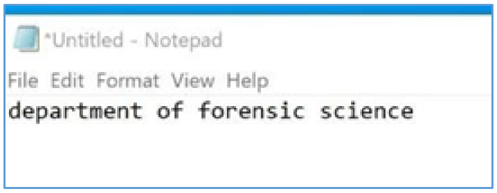
- Step 3: In the command prompt, go to the path where the SNOW application is located.



```
C:\Users\Eswara Sai Prasad C\Downloads\snwdos32>
```

Fig – 8: Folder Path to SNOW

- Step 4: Again, open the Run dialog to open Notepad, create a text file and save the file in the same path as the application.



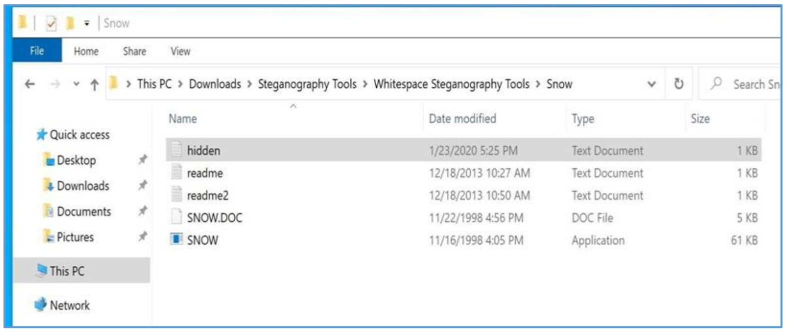
```
department of forensic science
```

Fig – 9: Notepad Native content

- Step 5: In the command prompt, enter the following command to embed the text into hidden.txt.

Snow –C –m "my atm card number is 1234567834567654 and cvv is 000" –p "department" hidden.txt department.txt

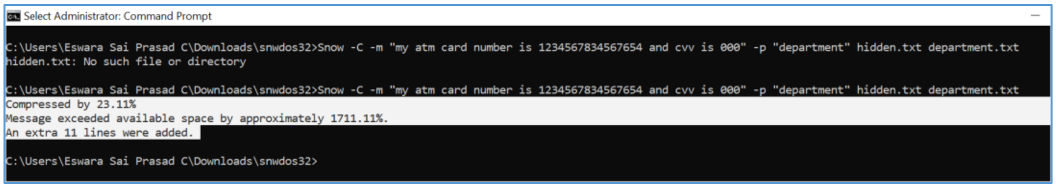
- Step 6: The resulting window will be as below:



Name	Date modified	Type	Size
hidden	1/23/2020 5:25 PM	Text Document	1 KB
readme	12/18/2013 10:27 AM	Text Document	1 KB
readme2	12/18/2013 10:50 AM	Text Document	1 KB
SNOW.DOC	11/22/1998 4:56 PM	DOC File	5 KB
SNOW	11/16/1998 4:05 PM	Application	61 KB

Fig – 10: Contents of the Folder

- Step 7: A new text file with the name department will be saved at the same location as the application.



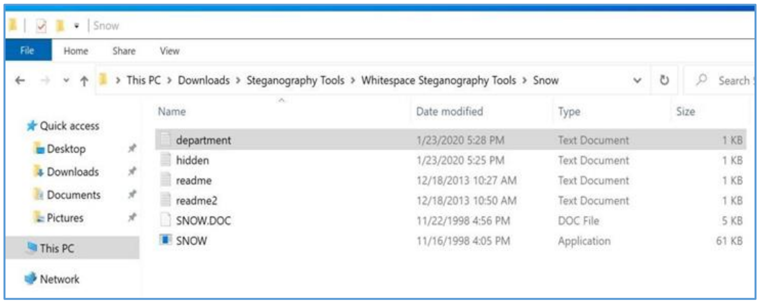
```
C:\Users\Eswara Sai Prasad C\Downloads\snwdos32>Snow -C -m "my atm card number is 1234567834567654 and cvv is 000" -p "department" hidden.txt department.txt
hidden.txt: No such file or directory

C:\Users\Eswara Sai Prasad C\Downloads\snwdos32>Snow -C -m "my atm card number is 1234567834567654 and cvv is 000" -p "department" hidden.txt department.txt
Compressed by 23.11%
Message exceeded available space by approximately 1711.11%.
An extra 11 lines were added.

C:\Users\Eswara Sai Prasad C\Downloads\snwdos32>
```

Fig – 11: Results View

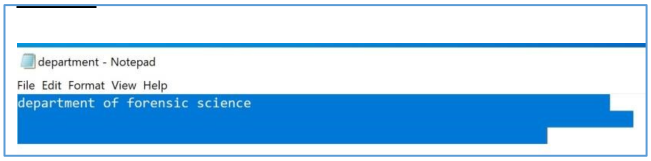
- Step 8: Now, check for the embedded text in the department.txt file.



Name	Date modified	Type	Size
department	1/23/2020 5:28 PM	Text Document	1 KB
hidden	1/23/2020 5:25 PM	Text Document	1 KB
readme	12/18/2013 10:27 AM	Text Document	1 KB
readme2	12/18/2013 10:50 AM	Text Document	1 KB
SNOW.DOC	11/22/1998 4:56 PM	DOC File	5 KB
SNOW	11/16/1998 4:05 PM	Application	61 KB

Fig – 12: New Text File

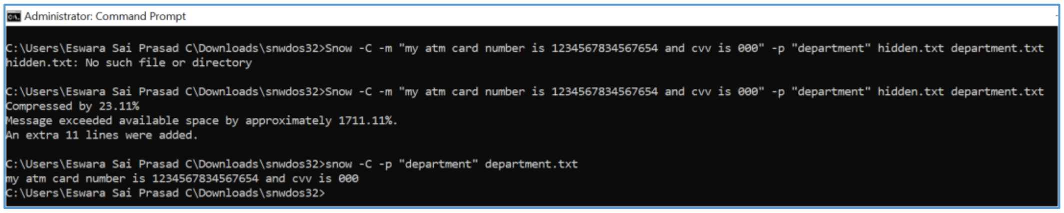
- Step 9: To extract the embedded text, navigate to the file location and give the following command in the command prompt.
  snow –C –p "department" department.txt
- Step 10: The resulting window will be like below:



```
department of forensic science
```

Fig – 13: Contents of the modified file with the extra hidden text

- The embedded text will be displayed within the command prompt.



```
C:\Users\Eswara Sai Prasad C\Downloads\snwdos32>Snow -C -m "my atm card number is 1234567834567654 and cvv is 000" -p "department" hidden.txt department.txt
hidden.txt: No such file or directory

C:\Users\Eswara Sai Prasad C\Downloads\snwdos32>Snow -C -m "my atm card number is 1234567834567654 and cvv is 000" -p "department" hidden.txt department.txt
Compressed by 23.11%
Message exceeded available space by approximately 1711.11%.
An extra 11 lines were added.

C:\Users\Eswara Sai Prasad C\Downloads\snwdos32>snow -C -p "department" department.txt
my atm card number is 1234567834567654 and cvv is 000
C:\Users\Eswara Sai Prasad C\Downloads\snwdos32>
```

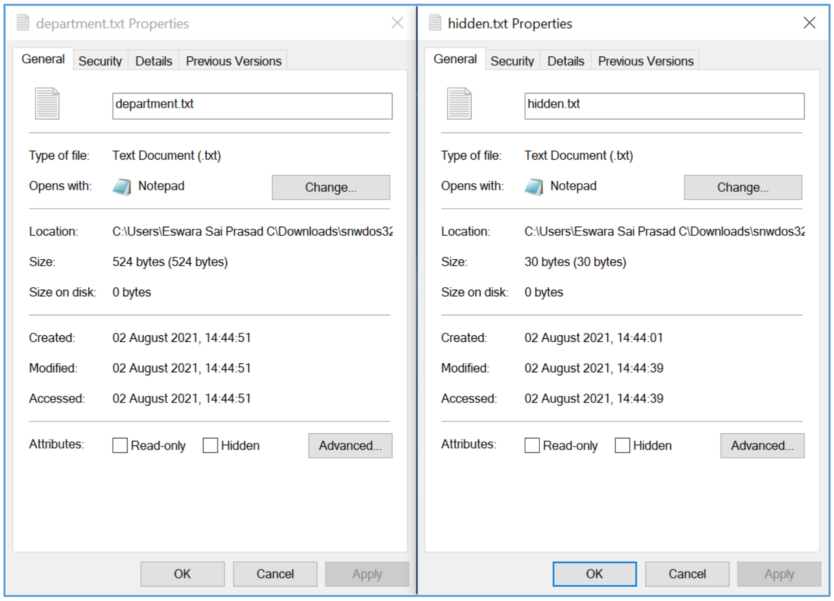
Fig – 14: Final Result window

6. Advantages of Steganography

- It is simple to use and implement.
- The technique can be extended to use in Unicode characters with effectiveness by using invisible characters such as ZERO WIDTH SPACE (U+200B) and ZERO WIDTH JOINER (U+200D) (7).
- It can be used to create a 3-D effect to emphasize certain words within a line spacing (8).
- It can be used to conceal Web Shell in PHP.
- It can be used in hiding the backdoor.
- It is used in information Hiding in Simple Object Access Protocol (SOAP) Messages of Web services (6) (9).
- It is used in Zero Width Space Steganography (ZWSP)
- It can be used in ANiTW: A Novel Intelligent Text Watermarking technique for forensic identification of spurious information on social media (5) (10).
- It isn't easy to detect. Only the receiver can detect.
- It can be done faster with a large number of software applications.

7. Disadvantages of Steganography

- A Crude method can sometimes be easily detected by using statistical analysis techniques. See the size in the below Fig. - 15:



department.txt

Type of file: Text Document (.txt)
Opens with: Notepad
Location: C:\Users\Eswara Sai Prasad C\Downloads\snwdos32
Size: 524 bytes (524 bytes)
Size on disk: 0 bytes
Created: 02 August 2021, 14:44:51
Modified: 02 August 2021, 14:44:51
Accessed: 02 August 2021, 14:44:51
Attributes: □ Read-only □ Hidden

hidden.txt

Type of file: Text Document (.txt)
Opens with: Notepad
Location: C:\Users\Eswara Sai Prasad C\Downloads\snwdos32
Size: 30 bytes (30 bytes)
Size on disk: 0 bytes
Created: 02 August 2021, 14:44:01
Modified: 02 August 2021, 14:44:39
Accessed: 02 August 2021, 14:44:39
Attributes: □ Read-only □ Hidden

Fig. - 15: Size variation between the Modified and Native files

- A huge number of data, huge files size, so someone can suspect about it.
- While steganography can help improve private data sharing, the same methods can be used by nefarious elements.

- The algorithms maintain the confidentiality of information, and if the algorithms are known, then this technique is of no use.
- Unauthorized access of data may occur due to Password leakage.

8. CONCLUSION

The Whitespace Steganography in the current digital age can be attributed to individuals' desire to hide communication through a medium rife with potential listeners, the absolute necessity of maintaining control over one's own, and the integrity of data it passes through this medium. This overview covered some of the more common methods of data hiding using widely used file formats and easily available tools as an introduction to the primary concepts of whitespace steganography. These discussions should serve as a starting point to exploring more complex steganographic techniques involving network packets and unused hard disk space as a cover medium or the more complex methodologies used on text files. To extend this, whitespace steganography can also be implemented on other document file formats using many other available tools. Various steganography techniques can be used for these purposes and can be considered the future scope of this work.

REFERENCES

[1] Banerjee, Dr- Indradip & Bhattacharyya, Dr.Souvik & Sanyal, Prof (Dr.) Goutam. (2011). An approach of quantum steganography through special SSCE code. World Academy of Science, Engineering and Technology. 80. 939-946.

[2] Shirali-Shahreza, Mohammad. (2008). Text Steganography by Changing Words Spelling. 3. 1912 - 1913. 10.1109/ICACT.2008.4494159.

[3] Ali, Maisaa & Khairi, Teaba. (2020). Review: A comparison Steganography Between Texts and Images. Journal of Physics: Conference Series. 1591. 012024. 10.1088/1742-6596/1591/1/012024.

[4] Abdallah, Yaman & H. O. Nasereddin, Hebah. (2013). Proposed Data Hiding Technique – Text under Text. American Educational Research Journal. 5. 243-248.

[5] Taleby Ahvanooey, Milad & Li, Qianmu & Hou, Jun & Dana Mazraeh, Hassan & Zhang, Jing. (2018). AITSteg: An Innovative Text Steganography Technique for Hidden Transmission of Text Message via social media. IEEE Access. 2018. 65981-65995. 10.1109/ACCESS.2018.2866063.

[6] Text to Text Embedding Approach for Information Security System Hsint Hsint Htay, Kaythi Aung San, Phyu Phyu Htun, Chaw Kalyar Than, Saw Zaw Lin, Kyaw Zin Htun, Aung Myint Aye, Moh Moh Aung https://www.ucstgi.edu.mm/storage/2020/10/3.pdf.

[7] S. T. Abaas, Improve Capacity in Text in Text Steganography, Education College, University of Kufa, Iraq, European Academic Research, Vol. II, Issue 12/march 2015.

[8] S. Bhavana, Department of ECE, Bangalore, Text Steganography using LSB Inserting Method Along with Chaos Theory, International Journal of Computer Science, Engineering and Application (IJCSEA), Vol.2, No.2. April 2012.

[9] Information Hiding in SOAP Messages: A Steganographic Method for web Services, Bachar Alrouh, Adel Almohammad, Gheorghita Ghinea, Brunel University, West London, UK, International Journal for Information Security Research (IJISR), Volume 1, Issue 1, March 2011

[10] Milad Taleby Ahvanooey, Qianmu Li, Xuefang Zhu, Mamoun Alazab, Jing Zhang, ANiTW: A Novel Intelligent Text Watermarking technique for forensic identification of spurious information on social media, Computers & Security, Volume 90, 2020, 101702, ISSN 0167-4048, https://doi.org/10.1016/j.cose.2019.101702.

[11] Por, Yee & Ang, T & Beh Mei Yin, Delina. (2008). WhiteSteg: A new scheme in information hiding using text steganography. WSEAS Transactions on Computers.

[12] Novel Text Steganography through Special Code Generation by Indradip Banerjee, Souvik Bhattacharyya and Prof. Gautam Sanyal International Conference on Systemics, Cybernetics and Informatics.

[13] Patiburn, Sivabalan & Iranmanesh, Vahab & Teh, Phoey. (2017). Text Steganography using Daily Emotions Monitoring. International Journal of Education and Management Engineering. 7. 1-14. 10.5815/ijeme.2017.03.01.

[14] Webpage: http://manpages.ubuntu.com/manpages/bionic/man1/stegsnow.1.html

AUTHORS' PROFILES:

Mr. Eswara Sai Prasad Chunduru pursued M.Sc. (Physical Chemistry), M.Sc (IT) and Criminal Justice and Data Analytics from IIT, Kanpur. He is currently working as Assistant Director and Scientist C in Digital Forensic Division of CFS, GoI, Hyderabad. During his tenure of 22 years, he has analyzed more than 1500 cybercrime cases. He is a Guest Faculty at various organizations. He is co-author of the book - "Handbook on Cybercrime Investigation."

Mr. Nagendar Rao Koppolu joined Police Service as Sub-Inspector in the year 1998. He served in Law Enforcement, Bureau of Immigration (IB), Central Bureau of Investigation (CBI) (Anti-Corruption Wing), State Intelligence Department, and State Information Technology Cell. He pursued M.Tech (CSE), M.Sc. (IT), and Criminal Justice Data Analysis (IIT Kanpur). He is a certified Cyber Security Professional and ISO 27001 ISMS Lead Auditor. He co-authored two books on cybercrime. Presently, he is Inspector (in-charge) of State Cyber Vertical, Telangana.