DATA EXFILTRATION USING STEGANOGRAPHY

Sonali Singh
Student, NCRD's Sterling Institute of Management Studies, Nerul, Navi Mumbai
Sonali.s1096@gmail.com

Sugandha Sharma
Student, NCRD's Sterling Institute of Management Studies, Nerul, Navi Mumbai
sugandhavsharma@gmail.com

Dr. Murlidhar Dhanawade
Professor, M.C.A. Dept. NCRD's Sterling Institute of Management Studies, Nerul, Navi Mumbai
dr.murlidhar.dhanawade@gmail.com

ABSTRACT:

Data Security is the main concern nowadays. Attackers uses various ways to exfiltrate data from a network, one of the methods includes using steganography, considering the company's risk of leaking information. In this paper, researcher is proposing a model by which a company can detect the data exfiltration using steganography in a network. The model will be between the firewall and the network organizations so that if a file is considered suspect then the model will flag the file and give the stegnalysis team an alert for false positive validation. Once these suspicious data are transferred to the team of stegnalysis, the data will be

monitored.The team will pass through the alert messages provided by the model and find out the hidden data. Researcher has used different tools like exiftool, binwalk tool, etc to monitor the data. These tools can be used within the proposed model and will help to perform Stegnalysis so that the important and secret data can be retrieved easily that are hidden within different media files. Since the organization's confidential data must not go out, these methods will help us to protect the data. These tools can go through each and every media files as most of the steganography takes place inside the media files. This paper focuses specifically on protecting data inside the organization, so that sensitive data leakage must not take

place.In turn, it will allow organizations data to be secure within the company so that no othe r entity outside the organization will manipulate the data.

Keywords: *Data, Exfiltration, Information, Malware, Security.*

1. INTRODUCTION:

Now days, data is a major concern for any organization. Whenever an attacker takes foothold in the organization network or any network device, attackers' main objective is to steal the

information / data from the network. An attacker's challenging task here is to find a way to exfiltrate the data in such a way that no organization's firewalls should be able to detect the exfiltration operation, for this an attacker uses multiple ways, including exfiltrating the data using steganography. Most of the attackers uses steganography for exfiltration, even the latest malware available on the market uses this strategy, the reason being that it is easy to hide the data in any multimedia file or text file, the company usually white list the multimedia files, and this makes it easy for an intruder to send data outside the organization using steganography. With all these aspects in mind, researcher is demonstrating a model for any organization by which suspicious exfiltration using steganography can be detected.

A. Steganography

The term steganography is a combination of the Greek words (steganos), meaning "hidden, concealed, or secured" and (graphein) meaning "writing".

Steganography is the process of hiding secret messages inside the image so that the data cannot be recognized by anyone else, except the sender and receiver. It is mainly used in places where confidential data plays the main role.

B. Types of steganography:

Steganography is of different types that are:

a) **Text Steganography**: Hides information inside the text files. The secret data in this system is concealed behind every nth letter of every word in the text message. There are numbers of methods for hiding data in text files. Such approaches are: Format-based method, Random and Statistical method and Method of Linguistics.

b) **Image Steganography**: Hiding the data in an image is referred to as steganography of the image by taking the cover item as background. Pixel intensities in animage are used to hide the details. For digital steganography, images are widely used source of cover because in digital representation of an image there are several bits present.

c) **Audio Steganography**: It involves hiding the info with audio files.

d) **Video Steganography**: It is a technique of encoding files or data of any kind into digital video format.

e) **Network or Protocol Steganography**: This includes covering the information as a cover item by taking the network protocol like TCP, UDP, ICMP, IP etc.

C. How it is used in real world?

The basic example of data exfiltration in the real world can be observed in the following scenario:

If the intruder already has a foothold within the network, he / she now wants to send confiden tial information outside the company, Figure 1 steps illustrate, how a confidential file can be hide inside an image.

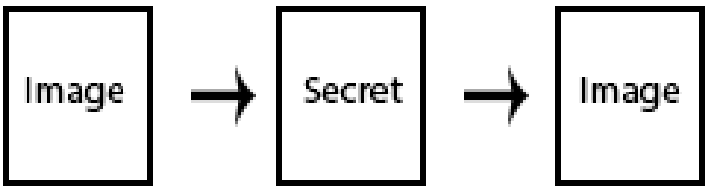

Fig. 1 Steganography [prepared by researcher]

2. OBJECTIVE:

The objectives of this paper are:

- The researcher will present steganography and steganalysis by using various tools such as exif and binwalk.
- Both techniques are used to conceal the secret data inside an image and to get the hidde n data from the image as well.
- These techniques can be used inside a model that will be placed between the firewall and organization network to go through each and every data file that will be going inside and outside the organization so that no secret data must be leaked outside the organization.

3. LITERATURE REVIEW

Gurubaran S (2017), in his article, 'Risk with Steganography and Importance of running Steganalysis with Network Systems' [1] published in 'GB Hackers on Security', reveals the importance of running Steganalysis with network systems. He clarified the use of steganography to hide the documents in an image to overcome filters and other security tools for Data Loss prevention (DLP).He has discussed an idea of a model that will run the steganalysis suite and avoid the attack so that no data loss in the companies can be prevented.

Ahmed Hesham (2019), in his article 'Steganography- A list of useful tools and resources' [2] published in 'OXRICK' has revealed different tools and resources that can be used for steganography and steganalysis: to exfiltrate data. He has explained various tools like

Exiftool, Binwalk, Stegsolve, Steghide, etc with their commands. He has given different ideas of using those commands so that they can be used for data hiding and even for data exfiltration.

Taylor Gibb (2016) in his article 'Embedding Zip Files Inside Image Files'[3] published in 'How-To-Geek' has revealed the. techniques to hide a zip file inside an image without using any extra software. He explains that a zip file can easily be concealed inside an image, whether it's png, jpeg or gif.

Previously, anybody could use steganography to hide the data or confidential information inside the image to bypass the organization's Data Loss Prevention (DLP) filters and other security tools.The company networks filter generally white list the media file so the data leakage is possible. There are various tools that are used for steganography with image and can extract data from it.

The researcher has explored in the present model how the organization or a business can create a model that is nothing but a script that tests whether or not any image file has any information or file embedded in it. If the image does not contain any information or data file, the image will be forwarded and if not then the file will be reviewed and the data retrieved through human contact and accordingly the measures will be taken against the person who attempted to leak the information outside of the organization.

4. RESEARCH METHODOLGY:

Researcher is choosing three popular ways to show how an intruder can conceal a file inside a real file.

- Case 1 – Hiding information within the exif data of an image.

  Usually, each image file has exif data; the exif data is like geolocation, author name, desc ription, etc. An attacker can misuse exif functionality. Storing any data within these fields is simple and this same picture can be transferred to a different end over the network.

  In the Figure 2, researcher has stored "This is Secret Data" data in image description field using a tool called exif pilot.

  Once the data is inserted and saved, this file can be exchanged with another person and on ce the file is outside the premises of the company, the intruder can get the confidential dat a from the field "Image Description" of the data exif.

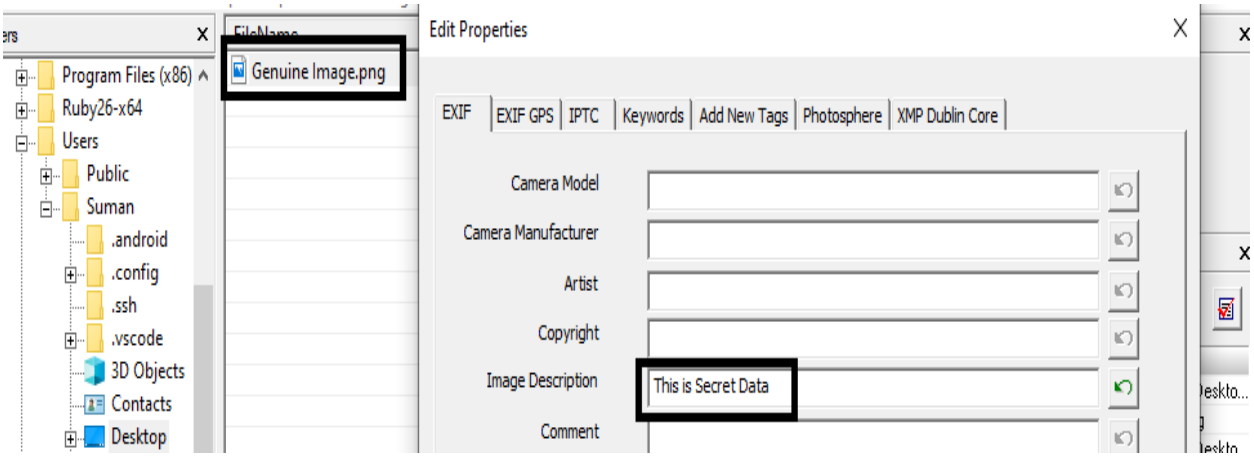

Fig. 2 Data hiding using exif tool

- Case 2 – Hiding a Zip File inside a Image.

Here in this scenario, it has been shown that Copy command has been used to ship with all the new windows by default. In general, the system reads the image file from the top of the header and the zip files are read from the top of the footer, the copy command can be used to combine 2 binary files into 1 file and 1 file is an image file. Now whenever the system wants to locate the file, the file will be read from the top and therefore the file will be treated as an image file. Now this file can be send to the attacker server. Once it has been received, the attacker can extract the confidential zip from the image.

In Figure 3, it can be observed that 2 files: "Sample.jpg" and "Secret.zip" are combined into a single "New.jpg" file, and now the system shows the "New.jpg" image without any problems.

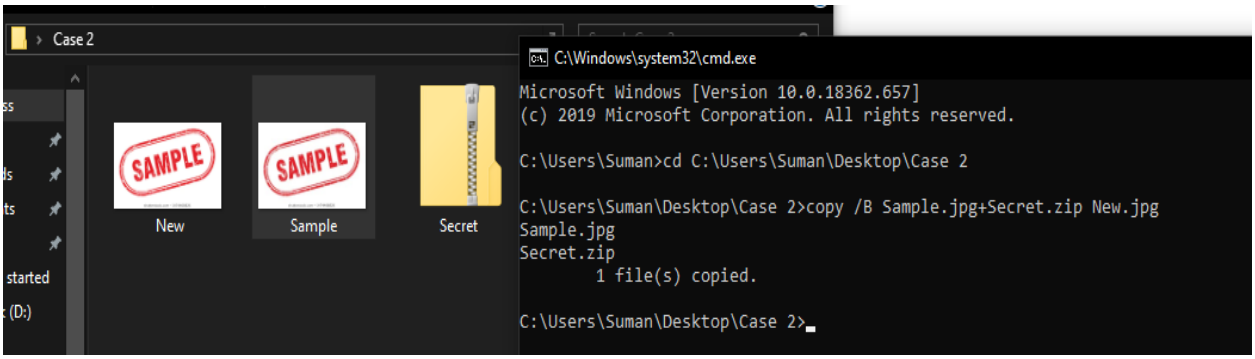

Fig. 3 Hiding zip file inside image

- Case 3 – Hiding text Data inside an ICMP Packet.

Intruder can also send the data in network protocol format. In the Figure 4,it can be observed that intruder is trying to send data "This is secret data" in the form of ICMP packets.

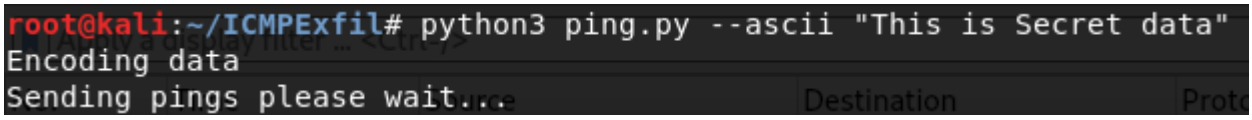

Fig. 4 Sending data in form of ICMP packets

In the Figure 5, it can be seen that the data is being received on the server end. This is one of the most interesting ways to exfiltrate the data from the organization.

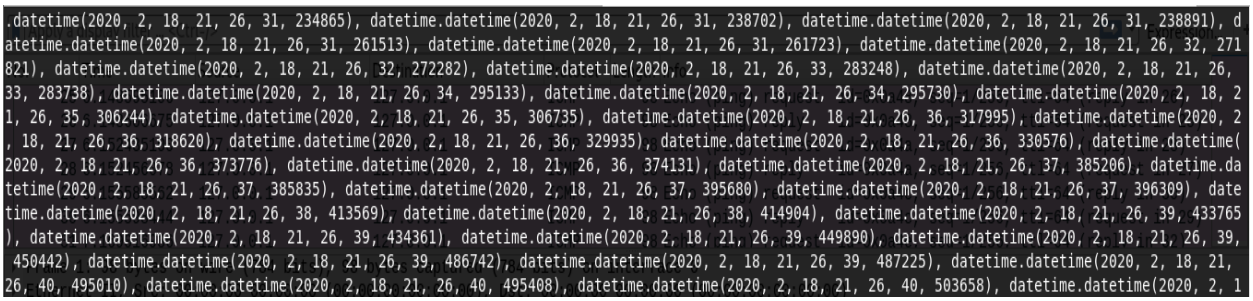

Fig. 5 Data receiving on server end

STEGANALYSIS:

Steganalysis is the method of retrieving hidden data, concentrating the encoded hidden message and, if possible, retrieving the message. By looking at changes between bit designs and unusually large file sizes, the message can be identified. This is considered to be an attack on the information covered. Network steganography detection methods have been developed somewhat independently of IDS / IPS systems. Some steganalysis methods focus on trying to detect the presence of secret communication and then restrict its transmitting capabilities. So, removing all network steganography opportunities is virtually impossible. Below it will be observed that how the secret data can be manually identified, which are hidden inside the images or packets and same can be automated using any programming languages.

Below by 3 ways, it can be identified whether any data is been hidden inside an image or packets.

- Case 1 – Analyzing exif data of an image file.

In this scenario, the exif data of an image file can be analyzed.The tool called "exiftool" is used to grep the description tag and get the details of the same tag as shown in Figure 6.

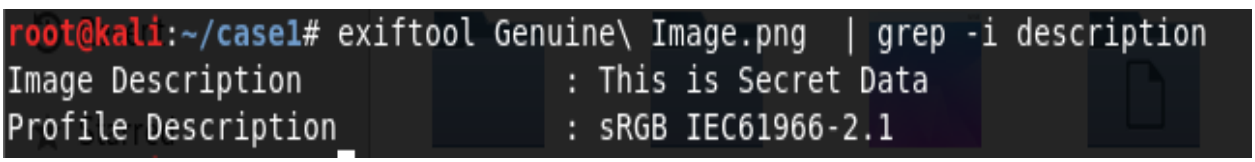

Fig. 6 Using exif tool

- Case 2 – Performing Binary walk on an image file.

In this scenario, binwalk is performed on an image file.A tool called"binwalk" (Binary Walk) is used to walk over the "New.jpg" to everyheader in the binary file. Here in the Figure 7,it can beobserved that "binwalk" was able to find a zip archive data file within the "New.jpg" file.

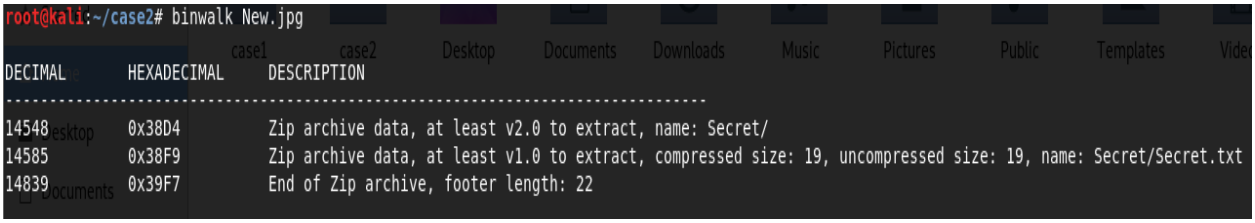

Fig. 7 Binwalk to find archive data

Now "binwalk –e" option is used to extract the files from the "New.jpg", and the hidden file have been found from the "New.jpg" file as shown in Figure 8.

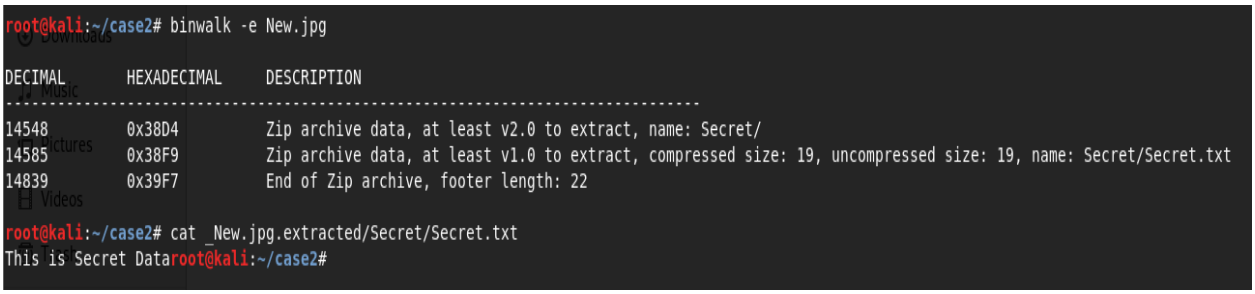

Fig. 8 Binwalk to find hidden data

- Case 3 – Analyzing the ICMP packets.

Generally, in these kinds of analysis, the analyst requires a control over the in and out of packets of the organization just like a firewall has. The analyst collects enough packets on which they can perform some operation like calculating the offset to get the details from it as shown in Figure 9 and Figure 10.

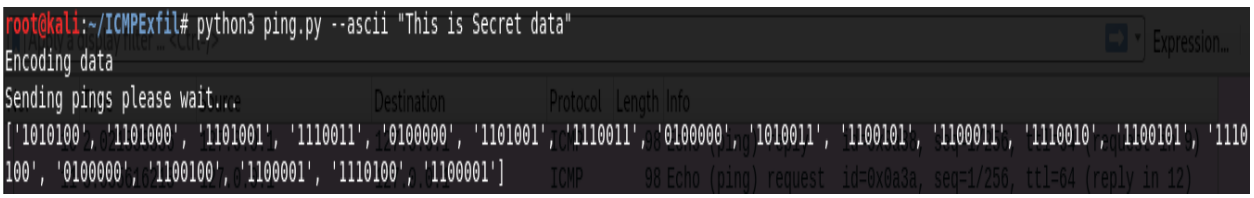

Fig. 9 Calculating offsets

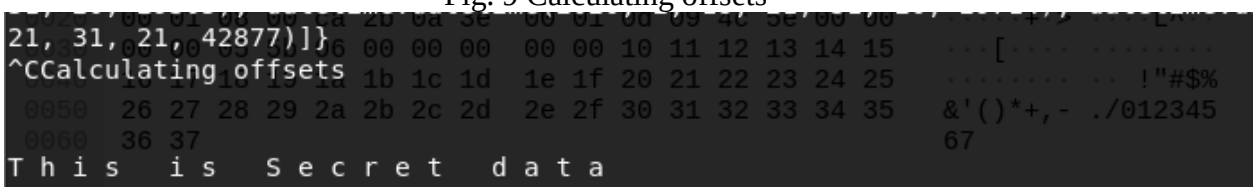

Fig. 10 Calculating offsets

5. PROPOSED MODEL

In each organization the data flow from the organization to the firewall (IDS / IPS) and then the no malicious packets are forwarded to the organization gateway after filtering the malicious activity detection and the same packets are released over the network. Since the IDS / IPS do not have the capability to detect exfiltration of the steganography, what we can do is create a model before Firewall (IDS/IPS) as shown in Figure 11. This model will have a set of tools/techniques which will check all the files and packets. Example like exiftool and binwalk, exiftool can be used to check the image exif tags and binwalk can be used to check that, what all different header files are present in the same file. If a file is found to be suspicious, the model will flag the file and send warning for the false positive validation to the steganalysis team. If a file is found to be suspicious, the model will flag the file and send warning for the false positive validation to the steganalysis team.

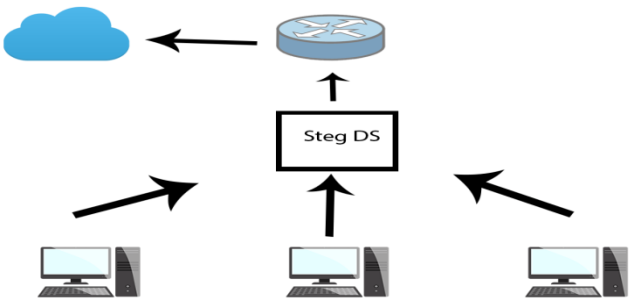

Fig. 11 Creating a model before firewall [prepared by researcher]

6. LIMITATIONS:

Although researcher have shown that using the above techniques, it can be determined whether or not a file / packet is suspicious, but for any information system on the market it is not easy to analyze any data once it has been flagged as suspicious, so once the file is flagged as suspicious, the next step is to do a false positive analysis manually. These things can be automated, which will be analyzed during the real time but not everything can be automated, as it's a cat and mouse game where the attacker tries to find ways to exfiltrate the data and defender tries to find those possible ways by analyzing manually and base on that the IOC's (Indicator Of Compromise) can be created and which can be shared to SOC (Security Operations Center) team for updating the signature database of the firewalls. Within information technology, however, things are changing rapidly; there is a possibility that using ML / AI a system can be built to reduce manual effort to a degree.

7. FUTURE ENHANCEMENT:

- There is a need for various tools to create a script for steganalysis, so that the job does not have to be done manually.
- Script must scan immediately, when the data is entered inside the network.

8. CONCLUSION:

This paper presents essentially the idea of steganography within an image or a network protocol.

Here the researcher has used different tools like exiftools and binwalk tool to introduce steganography so that the exfiltration of data can be recognized by stegnalysis. The exif tool will work upon the exif data so that the data hidden inside the exif part of the image can be recognized. Binwalk tool will help to check whether the two different media files are joined together to hide a secret data. These tools will prevent the confidential data to be leaked outside the organization. These tools will be embedded inside the model. The model will be placed between the firewall and the organizations network. It will basically scan the entire data that will be processing inside and outside of the network. If the model finds any data suspicious then it will give an alert message. This alert message will go to the stegnalysis team. The team will go through the alert messages given by the model and will find out the hidden data. It will basically help the organizations data to be safe within the organization so that no other person outside the organization can misuse the data.

Data is an important asset for any organization and therefore, safeguarding it from online cri minals is crucial.A fear of financial loss is the single most important explanation for adopting data protection strategies. Data loss can result in direct financial losses, such as loss of sales, fine or monetary judgments. This paper plays a major role for adapting the data security through stegnalysis inside an organization.

9. REFERENCES:

[1] "Risk with Steganography and Importance of running Steganalysis with Network Systems", Gurubaran S.

[2] "Steganography- A list of useful tools and resources", Ahmed Hesham.

[3] "Embedding Zip Files inside Image Files", Taylor Gibb.