Steganography

P2-project
Group A325a
Software
Aalborg University
May 24, 2016

AALBORG UNIVERSITY
STUDENT REPORT

Første Studieår v/ Det Teknisk-Naturvidenskabelige Fakultet
Byggeri og Anlæg
Strandvejen 12-14
9000 Aalborg
http://www.tnb.aau.dk

Title:

Steganography

Project:

P2-project

Project period:

February 2016 - May 2016

Project group:

A325a

Authors:

Anders L. Jakobsen
Andreas N. Jensen
Matias R. Jensen
Rasmus Jespersen
Simon N. Linnebjerg
Theis E. Jendal
Thomas B. Andersen

Supervisor:

Søren Enevoldsen

Pagecount: TODO
Appendix: TODO
Finished 24-05-2016

Abstract:

Synopsis

Preface

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam neque augue, tincidunt id augue at, mollis bibendum felis. Vestibulum ultrices nisi at tortor venenatis, nec ultrices justo pulvinar. Nam non iaculis metus, et consequat lectus. Sed vestibulum dui dolor, quis auctor orci posuere non. Aenean commodo pulvinar augue, eu aliquam sapien iaculis at. Sed eget vestibulum risus, vel viverra justo. Pellentesque ut lacinia dui, vel vestibulum nisi. Sed imperdiet consectetur turpis, id lobortis mi viverra sit amet. Ut quis aliquet nisi. Quisque dolor quam, efficitur quis aliquet eget, commodo consequat tellus. Morbi tincidunt ipsum sit amet velit pretium blandit varius non diam. Nam vitae purus tempus, auctor justo in, malesuada tortor. Donec nulla mauris, faucibus id diam et, finibus maximus orci. Aenean euismod maximus odio, a molestie urna cursus id. Maecenas congue ut sapien blandit tristique.

Duis est purus, lobortis eu ex scelerisque, viverra rhoncus risus. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Donec dictum augue et arcu volutpat ullamcorper. Nulla suscipit eget sem sit amet porta. Praesent consectetur tellus in eros sagittis, id rutrum augue molestie. Nam lobortis massa eu dapibus malesuada. Etiam id velit commodo neque maximus sollicitudin. Maecenas finibus cursus orci, eu fermentum lectus pharetra at. Pellentesque ornare, orci sed ultrices maximus, risus nisi posuere orci, in finibus urna nisl sit amet mi. Ut id accumsan quam. Vivamus porta sem sit amet aliquam hendrerit. Vivamus sagittis magna porttitor hendrerit volutpat. Integer dolor massa, ullamcorper vel arcu euismod, viverra dapibus lectus. Sed sodales porta magna, non pretium risus aliquet vitae.

Contents

1 Project description

Part I

Background

Introduction 2

Introduction to the subject.

Users of steganography and/or steganalysis 3

To enable a proper definition of a problem to be solved, it is necessary to examine the stakeholders in question, in regards to steganography. As such the following analysis will cover two major areas of interest. The interest in using steganography, and the interest in discovering steganography using steganalysis. These can be seen as opposing, but not mutually exclusive interests.

Steganography users have an interest in hiding information, this information can be applied in many areas, and has a usefulness for not only personal or criminal use, but also for government or corporate offices, where secret information can be sent with reduced risk of unauthorized interception. This information could range from classified information such as state secrets to simply private information such as citizens taxes.

Steganalysis users have an interest in discovering hidden data, this is mostly done with statistical analysis of the data structure in question, and is mainly used by governments or corporate offices. This may heighten a nation's or corporate office's security, and make them able to preemptively make decisions to help their own interests such as increasing security based on intercepted messages and thus prevent an attack.

The stakeholder analysis will be a basis for further research, as different parties have different requirements to a potential steganography program.

3.1 Government

The government is divided into a variety of agencies that, in the course of fulfilling their obligations, use steganography and steganalysis. These government departments have different needs, because of the different nature of their overall assignments, and is divided into their own segment of steganography and steganalysis.

Agencies like NSA, and the danish PET, can hypothetically use steganography to hide messages from agents operating undercover in foreign countries. This will enable the agents to hide in plain sight, as they spy for the interests of the particular nation. That nations spy on each other is seen in many examples. fx. the infiltration of the Syrien main internet router in 2012. However not much is known on how governments apply steganography and steganalysis, as this information is closely protected in what each nation defines as a matter of "national security", and the disclosed information here is only known because of the actions of Edward Snowden, who conducted the biggest national security breach in American history.[3] However, this information is mostly about information gathering,

and country on country espionage, and doesn't cover the area of steganography.

It is however proven by FireEye, that an alleged Russian hacker group is responsible for a malware named "Hammertoss".[5] This malware, is directly connected to twitter, where the group can send Steganographic images to a specific account, and by doing that, activate the malware on the infected computers. The malware will then send information, documents and files to a cloud server, where the perpetrators can download the data. This method could theoretically be used in a government perspective, where data can be gathered from personal computers or corporation server systems, if this data is not encrypted. At the same time, this example, can also determine that national security agencies have to increase their efforts to detect steganography on pages like twitter, and ultimately, this means that the national security agencies have to take part in both steganography and steganalysis.

Law enforcement forensic methods are equally hard to come by. This may be because of the nature of IT, and the abilities of IT-criminals to turn that information to their own advantage. It is however, safe to assume that government authorities like the police, are able to use steganalysis to uncover criminals, as these information can be crucial for a court proceeding. As such, the government has an interest in steganalysis, and therefore in steganography, because of a need to uphold the law of the nation in question.

3.2 Journalists

A journalist wants to spread a story to multiple people and wants the sources to be credible. They therefore often have to search for people involved in the subject they are writing about, to have a primary source, which is a person that is directly evolved in the subject or for example written a document about it. The contact or source, can live in a dangerous environment, where secrecy is a vital part of communication. Often the best option is to keep offline, away from the internet and meet the source in person, but this is not always an option or very impractical.[6] This is where steganography or similar methods come in handy. As described in section ?? , Victims of human rights violation, there are governments and groups that want to retain information from public knowledge and disclosing the information could lead to prison or death. A case, Goodwin v. United Kingdom, in 1996, the European Court of Human Rights stated that,

"Protection of journalistic sources is one of the basic conditions for press freedom ... Without such protection, sources may be deterred from assisting the press in informing the public on matters of public interest. As a result the vital public-watchdog role of the press may be undermined and the ability of the press to provide accurate and reliable information may be adversely affected"[1]

meaning that the public should not be afraid of sharing their information to the media. Because of the danger of disclosing information in some countries, journalists can have a hard time finding sources, unless they can communicate in a safe circumstances. The problem is that most of the sources have a minimal amount technological knowledge. As said in "Investigating the Computer Security Practices and Needs of Journalists", the sources rarely have access to security technology and/or do not understand how to use the software.

This leads to steganography, encryption and similar forms of concealment, that can ensure safety or drastically reduce the risk of getting uncovered. If the people feel safe, they are more likely to share their stories and the journalists have the possibility to pass on the information to the public. Programs that can keep the communication secret, would be a great asset to the journalist.

3.3 Privacy advocates

Some steganography users are difficult to place in a specific group. Privacy advocates, for lack of a better name, are those who want to protect themselves from prying eyes, whether those are government bodies, criminals with ill intentions or simply other common people. As such, a subset of these advocates may be referred to as IT-activists, as their sole intention with steganography and cryptography is to counter mass surveillance. With that said, their interest in steganography is obviously broad. Some of these people may have an interest in hiding certain files on their file system, while others may simply use it to send messages to their friends. The motivation for these groups may seem vague or difficult to understand for somebody who does not care about privacy, but the matter is obviously quite important for the privacy advocates.

An argument often used against this group is the 'nothing to hide' argument. People who support this argument will argue that mass surveillance does not threaten the privacy of individuals, unless government agencies uncover illegal activities by this individual, in which case they don't have the right for privacy. On the other hand, privacy advocates will argue that the argument implies full trust in the state, which they argue can be hard to assume, even in fully developed nations. Nonetheless, some people want to hide from criminals trying to misuse their personal information, and tools like steganography and cryptography can be useful in this scenario.

3.4 Corporations

Steganography is also used by many corporations. A common use is in watermarking their products. In this way they hide information in the product which can be used to identify the specific item and possibly the rightful possessor. This can help prevent unauthorized sharing of videos by identifying who is responsible for the leak and thus punish the responsible party and possibly blacklisting said person to prevent future leaks.

However, other uses of steganography pose significant risks for corporate espionage as it enables the exchange of covert messages in otherwise harmless media. Many companies monitor employee emails, and would thus be able to identify obvious encrypted data but not data obscured in harmless media such as family photos.

3.5 Conclusion

Which user is the focus for this project?

Relevant technologies 4

Analysis of relevant technologies.

4.1 Carrier media

This section will serve as an overview of some potential carrier/cover media for the steganography tool. For every type of media, some of the most common file formats will be explained.

4.1.1 Images

There are many different image file types to choose from these days, some easier to implement stenography on than others. In this section we will describe a few image file types along with a general explanation of why images are subject to stenography.

JPEG File Interchange Format (JFIF) is an image format that is a common standard in most modern digital capture devices such as digital cameras and image software alike. JFIF uses the Joint Photographic Experts Group (JPEG) compression method (for more info about compression see chapter **??**).

Portable Network Group (PNG) is another image format widely used eight-bit paletted images because of the support for transparency for every palette color and 24-48 bit truecolor with and without alpha channels.

Windows bitmap (BMP) is a format that is more focused around files within the Microsoft Windows OS. While BMP files are usually larger in file size due to commonly being uncompressed, their advantage lies in simplicity and ease of use.

Graphical Interchange Format (GIF) is widely used to make small animations often to express humor, for example through the use of "memes". GIFs have limited support for colors and transparency compared to other mentioned image file formats. For example GIFs can only use an 8-bit palette or 256 colors.

All these image file types share some common principles, for example the use of pixels. Pixels are essential when talking about images and these pixels are also what is used to hide messages in the images. The difficult thing when hiding messages comes when altering the pixels because the pixels are essentially the visual representation of the image and therefore a pixel cannot be altered too much in order to avoid detection.

Pixels are the smallest element in any digital image. They can be subdivided to make larger resolutions which means having more pixels within the same area. This would mean

that each pixels would be divided into smaller pixels with the same value.[4]

4.2 Least Significant Bit method

A commonly used method in digital steganography is the Least Significant Bit (LSB) method. The basic principle of this method is to make small changes to a large amount of data. LSB is useful in data formats where small change don't cause big differences. Embedding hidden messages in text would for example be a bad idea, as changes would be apparent, but on an image such changes may not be apparent. In a lossless image format like PNG (see **??**), the decompressed data is represented as a 2D matrix of pixels. Depending on the pixel format (sRGB, aRGB or perhaps even grayscale) each pixel has a set amount of color channels, where each value is a byte (8 bits), which determines the intensity of the color channel on that specific pixel. Of course the bit depth is not necessarily always 8 bits, but this is the case in most images shared over the internet. The purpose of the LSB method is to change the least significant bit in each of these bytes, and perhaps even the second least significant bit as well. This means that the decimal value may only change by 1 if modifying one bit and by 3 if modifying two least significant bits.

For example, say we want to embed the character '*****' in a message. This ASCII character has a decimal value of 42, which can be represented as an 8-bit binary number:

$$00101010 \tag{4.1}$$

Since this message is 8 bits long, we would need 8 bytes when modifying one bit and 4 bytes when modifying two bits. In the former case, this means that we would need two aRGB pixels, as this gives us 2 * 4 bytes to hide our message in. For example, let's say we want to hide our image in the two pixels:

$$\begin{array}{cccc} 11111111 & 01001111 & 10110000 & 10000111 \\ 11111111 & 10010110 & 10001110 & 10000001 \end{array} \tag{4.2}$$

With the LSB method, we now need to look at the least significant bit in each of these bytes, and negate the bit if it does not correspond to the bit we want to insert. When the message is embedded in the original pixels, the produced result is:

$$\begin{array}{cccc} 1111111\mathbf{0} & 0100111\mathbf{0} & 1011000\mathbf{1} & 1000011\mathbf{0} \\ 1111111\mathbf{1} & 1001011\mathbf{0} & 1000111\mathbf{1} & 1000000\mathbf{0} \end{array} \tag{4.3}$$

The embedded message can then be extracted by simply combining each least significant bit of the 8 bytes. After doing this, we end up with the message seen in Equation 4.1.

When comparing the pixel values of a modified image (now referred to as a stego-image) and the original image, the change is clear on the bit/byte level, but not be apparent to a human comparing the two images. This means that LSB is prone to steganalysis (see **??**). With the stego-image and original image in possession, it is trivial for a piece of software to detect that one of the images have been edited. In addition, the LSB method will typically increase the size of the image, since adding noise to the least significant bits increases the amount of colors and can affect lossless compression used in formats such as PNG.

4.2.1 Graph theoretic approach

While the method described in the previous section is a sufficient steganography method, one might want a method that is less prone to steganalysis. After all, the point of steganography is to hide the fact that communication is taking place. In the method previously described, pixels are overwritten and the color frequencies are changed, but with a graph-theoretic approach it is possible to exchange pixels, so that the color frequencies are similar when comparing the original and stego-image. Depending on the size of the input message and the picture, it might be possible to hide a message in a picture without overwriting any pixels at all.

4.3 Compression of hidden messages

An obvious issue with digital steganography is the bandwidth available in the information carriers, especially static data formats like images. For example, let's say we have a 100x100 image with 3 color channels (aRGB). If we use the least significant bit method and use the two least significant bits, the amount of bits available in this image is:

$$100 * 100 * 8 * 2 = 160000 \tag{4.4}$$

This effectively means that we can fit about 20 kilobytes of data in this image. Of course this is not a problem if we are simply hiding text messages, since this image would bit about 20.000 UTF8 characters, but one might want to hide files within the image. Whether it's text or files being hidden in the file, compression is a very useful technology, as it allows us to fit more data inside the image.

We can split this technology into two groups: lossy and lossless compression. In lossy compression, an algorithm finds redundant information and removes it. This is the same method used in the JPEG image format, where images are of noticeably worse quality after compression. Lossless compression also finds redundant information, but does not eliminate it. It utilizes statistical redudancy by checking for repeating patterns in the information. For example, images often have repeating pixels. If we have 100 green pixels written as "green pixel, green pixel, green pixel, ..." in the image, the compression algorithm can rewrite it to "100xgreen pixel". This is what the DEFLATE algorithm does, which is implemented in PNG among other file formats.[2]

In conclusion, there is no reason to not use data compression in a steganography tool, as long as a lossless algorithm like DEFLATE is utilized.

Bibliography

[1] *Case of Goodwin v. the United Kingdom.* http://worldlii.org/eu/cases/ECHR/1996/16.html, 1996.

[2] *DEFLATE Compressed Data Format Specification version 1.3.* https://tools.ietf.org/html/rfc1951, 1996.

[3] *What's Next in Government Surveillance.* http://www.theatlantic.com/international/archive/2015/03/whats-next-in-government-surveillance/385667/, 2015.

[4] P. Fisher. The pixel: A snare and a delusion. *http://www.tandfonline.com/*, Volume 18, Issue 3:8, 2010.

[5] S. Jones. Spioner boltrer sig på twitter. 2015.

[6] B. Winegarner. *6 ways journalists can keep their reporting materials private and off-the-record.* http://www.poynter.org/2013/6-tips-for-keeping-your-secret-communications-secret-and-your-anonymous-sources-anon 215242/, 2013.