2019

A case study involving creating and detecting steganographic images shared on social media sites

Lindsey Kathryn Trotter
Iowa State University

Recommended Citation

Trotter, Lindsey Kathryn, "A case study involving creating and detecting steganographic images shared on social media sites" (2019). *Graduate Theses and Dissertations*. 17798.
https://lib.dr.iastate.edu/etd/17798

A case study involving creating and detecting steganographic images shared on social media sites

by

Lindsey Kathryn Trotter

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Information Assurance

Program of Study Committee:
Jennifer L. Newman, Major Professor
Thomas E. Daniels
Olga Chyzh

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this thesis. The Graduate College will ensure this thesis is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2019

DEDICATION

I would like to dedicate this thesis to my fiancé, Nick Musel, who's belief in me and encouragement kept me going. Also, to my bearded dragons, Lucy and Drakus who were beautiful models for the photos seen throughout this thesis. They also gave me lots of snuggles during my late nights and long afternoons studying.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

I would like to take this opportunity to express my thanks to those who helped me with various aspects of conducting research and the writing of this thesis. First and foremost, Dr. Jennifer Newman for her guidance, patience and support throughout this research and the writing of this thesis. Her insight and guidance inspired wisdom and focus. Without her, this thesis would not be possible.

I would also like to thank my committee members for their time and contributions to this work: Dr. Thomas Daniels and Dr. Olga Chyzh.

And lastly, thank you to Nicholas Musel for his constant words of encouragement, for believing in me, and for letting me use his computer. Without him, I would not have had the resources to do this research or the confidence that I could get everything done.

ABSTRACT

There are billions of people that use social media as a way to share information and communicate with other users around the world and these social media sites provide easy platforms for sharing images. Digital images can be used to hide messages which may contain incriminating evidence or malware. This project focused on how images with hidden messages, called steganographic (stego) images, can be shared on social media and how to detect the presence of such images. Two case studies were performed with the goal of finding a way to both share images on Facebook and Twitter that contain hidden messages and also to detect the presence of such images. This study found that it is possible to share such images on Facebook and Twitter when following a very specific method. Additionally, Quantization Matrices can be used to gather information about said images which may include whether or not the image in question is a stego image.

CHAPTER 1. INTRODUCTION

Steganography is the art of hiding or concealing a message within another message. It is often described as hiding a message in plain sight, meaning that an outsider looking in would not know there is a message to find. There are many ways to hide messages. For example, messages can be hidden within physical objects, other messages or images. A Steganographic (Stego) Image refers to an image with a stego message hidden inside of it. With the dawn of the technological, it is fairly simple to embed message data inside the bits of an image. In fact, there are many applications available for free on different types of devices and platforms that embed a message into an image. Sending digital images to one another has also become increasingly easier. These images can be shared via Online Social Networks (OSNs) or Social Media Sites such as Facebook, Instagram and Twitter. Sharing images on these sites make the images available to anywhere from a few hundred, to a few billion users across the world. This is an issue because these images could contain a hidden message, and the message could contain potential incriminating evidence, classified data or even malware which can be easily looked over by investigators. With all of this in mind, this paper will explore criminal and legal issues related to both social media and steganography, how existing tools can be used to create stego images to share on different social media sites, and ways stego images can be detected on OSNs.

Although previous work has been done on this topic, my contribution to this topic of study includes finding a method in which stego images can consistently be shared on social media in a way which retains the message. Additionally, I focus on finding a method to detect that an image shared on social media may contain a stego message through analyzing the quantization matrices of such images.

1.1 Important Terminology

The following list defines important terminology that will be used throughout this thesis

- **Steganography (Stego)**: Concealing a message inside of another message or object
- **Image Steganography**: Concealing a message inside of a digital image
- **Cover Image**: Used during image steganography to refer the original image that was used to hide a message
- **Stego Image**: Used during image steganography to specify the image that contains a hidden message
- **Stego Message**: This refers to the hidden message
- **Online Social Networks (OSNs) or Social Media Sites** Network where users can communicate and share information with other users as well as get updates from other users
- **Quantization Matrices (QM)**: Matrix used during the process of compressing an image into the .JPEG format

1.2 Document Overview

Table 1.1: Document Overview

Section	**Name**	**Description**
CHAPTER 1	Introduction	Introduction to the paper, important terminology as well as the Document Overview.
CHAPTER 2	Background	Background information on steganography, image steganography and Online Social Networks
Section 2.1	History of Steganography	The origin of steganography and examples on steganography techniques seen throughout history
Section 2.2	Steganography vs Cryptography	What cryptography is and how it differs from steganography

Table 1.1: (Continued)

Section	**Name**	**Description**
Section 2.3	Image Steganography	A brief overview of image steganography
Section 2.4	Online Social Networks	What OSNs are, how popular they are and how they can be used to share stego images
Section 2.4.1	Why use OSNs to share stego images	Why OSNs are a good medium in which to share stego images (if you do not want them to be found)
CHAPTER 3	Legal Issues	Overview of the legal and criminal issues related to steganography
Section 3.1	Social Media in Criminal Cases	Examples of criminal cases in which OSNs were involved in either the crime or solving the crime
Section 3.2	Social Media Evidence	Details on how and when information gathered on social media can be used as admissible evidence in court
Section 3.2.1	Privacy Issues with Social Media Evidence	Explores when gathering information on social media is a breach of privacy and when it is not
Section 3.2.2	Authentication Issues with Social Media Evidence	Explores how to authenticate social media evidence for court cases and why it is a challenge to do so
Section 3.3	Steganography in Criminal Cases	Examples of crimes which involved steganography
CHAPTER 4	Previous Work	A compilation of previous work done of these topics
Section 4.1	Steganographic Methods	An overview of the previous work that has been done on this topic
Section 4.2	Success of distributing Stego Images on Social media	A compilation of how previous researches used OSNs as a method of distributing stego images and how successful their methods were
Section 4.3	Summary	A summary of the previous work that has been done on this topic
CHAPTER 5	Case Study 1: How to share stego messages on Social Media	A case study on how to create and share stego messages on Facebook and Twitter in a way that retains the message
Section 5.1	Background	Background on what the case study will cover and why
Section 5.2	Method	The method used to create and share stego messages and how to test that the messages are retained

Table 1.1: (Continued)

Section	**Name**	**Description**
Section 5.3	Findings	Findings of the case study on how to create and share stego images on social media
Section 5.3.1	Failures	Findings related to where failures occurred when creating and sharing stego messages on social media and how to avoid them
Section 5.3.2	Images	Findings related to the images that were created and if/how encoding a stego message inside the image changed it
Section 5.4	Takeaways	Takeaways from case study 1
CHAPTER 6	Case Study 2: How to detect stego images on social media using Quantization Matrices	An analysis of quantization matrices on images created in Chapter 5
Section 6.1	QM Overview	An overview of what Quantization Matrices are and how they are used
Section 6.2	QM Method	The method used to analyze Quantization Matrices
Section 6.3	QM Findings	Findings related to analyzing the QM's of the images created in Chapter 5
Section 6.3.1	Table Quality Values Findings	Findings related to analyzing the Table Quality Values of the images created in Chapter 5
Section 6.4	Takeaways	Takeaways from case study 2
Section 7	Conclusion	Conclusion of the findings gathered during this study and why it is important
Section 7.1	Recommended method for creating and sharing stego messages on social media	This section depicts the recommended method for creating stego messages to share on social media in a way that will retain the stego message
Section 7.2	Recommended method for detecting stego messages on social media	This section details the recommended method for determining if an image shared on social media contains a hidden message
Section 7.3	Conclusions and Future Research	This section contains the final takeaways from this study, why it is important and recommendations for future work to do on this topics.

CHAPTER 2. BACKGROUND

2.1 History of Steganography

Steganography comes from the Greek words “steganos” (covered or secret) and -“graphy” (writing or drawing), or in other words, “secret writing” [3]. There are many documented cases of Steganography being used throughout history. The earliest documented use of steganography was in Ancient Greece. Herodotus documented the use of it in the book The Histories [4]. Herodotus describes how Histiaeus, a Greek ruler at the time, shaved a slave’s head, tattooed a message on the slave’s scalp, let the slave’s hair grow back and then sent that slave on his way to deliver a message to Aristagorus who was the regent of the city of Miletus. Once the slave arrived, Aristagorus shaved the slaves head to retrieve the message. The message encouraged Aristagorus to start a revolt against the Persian king. If the slave had been intercepted, the interceptor would likely not have thought to shave the slaves head to retain a message. Another instance of steganography that Herodotus documented was when Demeratus alerted Sparta that the Persian King was planning to invade Greece. In this instance, Demeratus used a wax tablet. He took the wax writing tablet, scraped the wax off, carved a message in the wood and then applied a fresh layer of wax. Then, the tablet just looked like a blank wax tablet, which would not arouse any suspicion. However, upon its arrival to the Sparta, he removed the wax to reveal the secret message.

Steganography has also been seen in more recent history. During the revolutionary war, both sides would relay messages by writing them in invisible ink. The invisible ink would be used to write a message between the lines of a harmless message or on a blank piece of parchment. As [5] explains, invisible ink used during the war usually consisted of a mixture of ferrous sulfate and water. The message would be revealed either by heating the letter or applying a chemical such as sodium bicarbonate. At the time, since both the British and Colonists were using invisible ink, Washington wanted a more complicated ink that couldn’t be revealed with an “ordinary chemical”.

He hired a British doctor, James Jay to create such a solution. This solution, when brushed with a special chemical would appear.

In these instances, if anyone intercepted the message, they most likely would have not have thought that they needed to shave someones head, melt wax or heat a piece of parchment to view a message. This is the benefit of using steganography over cryptography.

2.2 Steganography versus cryptography

Oftentimes, steganography is confused with cryptography. Cryptography comes from the Greek words “kryptos” (hidden) and “graphy” (meaning writing), or in other words “hidden writing”[6]. Whereas steganogrpahy’s origin, as stated in Section 2.1, is “covered writing”. Where steganography is the art of hiding something in plane sight, cryptography is the art of creating and deciphering codes.

If two people are trying to communicate and do not want other people to “listen” in or intercept their message, they can use cryptography or steganography or both. When cryptography is used, the sender will create a message that has been encrypted with some key that both sender and receiver know. An outsider looking in will see that there is a message but will not be able to decipher it without the necessary key. For example, if the sender wants to send the message “MEET ME AT EIGHT TONIGHT”, they could encrypt the message by shifting the letters over two spaces in the alphabet (A becomes C, B becomes D etc.). The message then would become “OGGV OG CV GKIJV VQPKIJV”. The person decrypting the message would be given the shift size (in this case two) in order to decode the message. This simple code, known as the “Caesar shift cipher” is easy to break by brute force [7]. Because there are only 25 possibilities for what the shift size is, someone intercepting the message, could try all 25 to see if the resulting message makes sens.

With steganography on the other hand, the sender might send a plain message, a wax tablet or a blank piece of parchment, for example, that has a secret message embedded in it. An outsider looking in will see a plain message, wax tablet or blank piece of parchment and not know that there was another hidden message to look for. This technique of hiding messages in “plain sight” can be extended to digital images as well, which brings us to todays digital era.

2.3 Image Steganography

Image steganography is the process of hiding a message in a digital image. The goal is to hide the message in such a way that is imperceptible to human eyes. The study of image steganography has led to many different algorithms in which to embed message data with that goal in mind. One such method is called “least significant bit insertion” When using the least significant bit insertion algorithm to encode a message inside of a digital image, the least significant bit of a pixel is flipped (from a zero to a one or from a one to a zero) or remains the same, depending on the message. In [8], the author explains that “Given a message data such as 11010010, the most significant bits (MSB) are those that lie to the far left and the least significant bits (LSB) lie to the far right”. Note that, a digital image can be represented as a stacked array of binary planes, where each binary plane consists of zeros and ones. The author in [1] depicts this representation in figure 2.1. In the case of bit planes, the “least significant bits” would be in bit plane 0. Because of how human vision is limited and how the least significant bitplane of an image affects the color and intensity of that pixel, if a least significant bit value is changed from a one to a zero or vise versa, it will most likely not be noticeable to the naked eye.

The author in [8] provides more details on other algorithms that can be used to create stego images, how these algorithms work and the pros and cons of each algorithm.

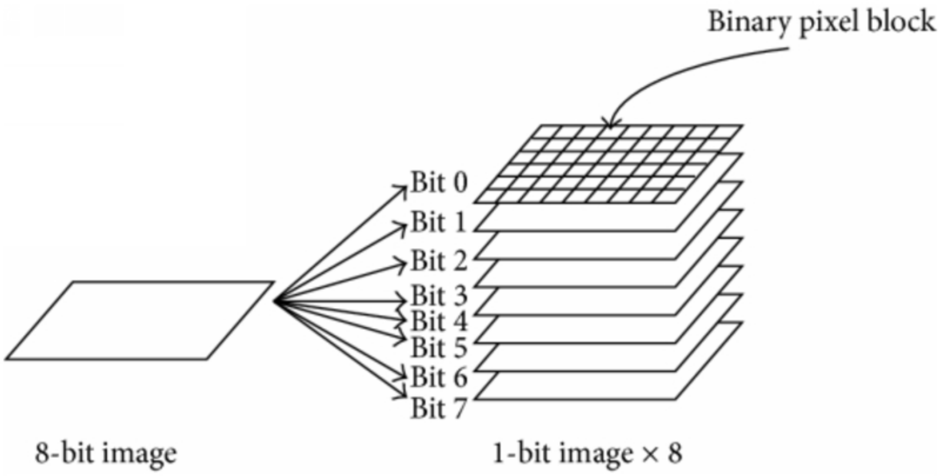
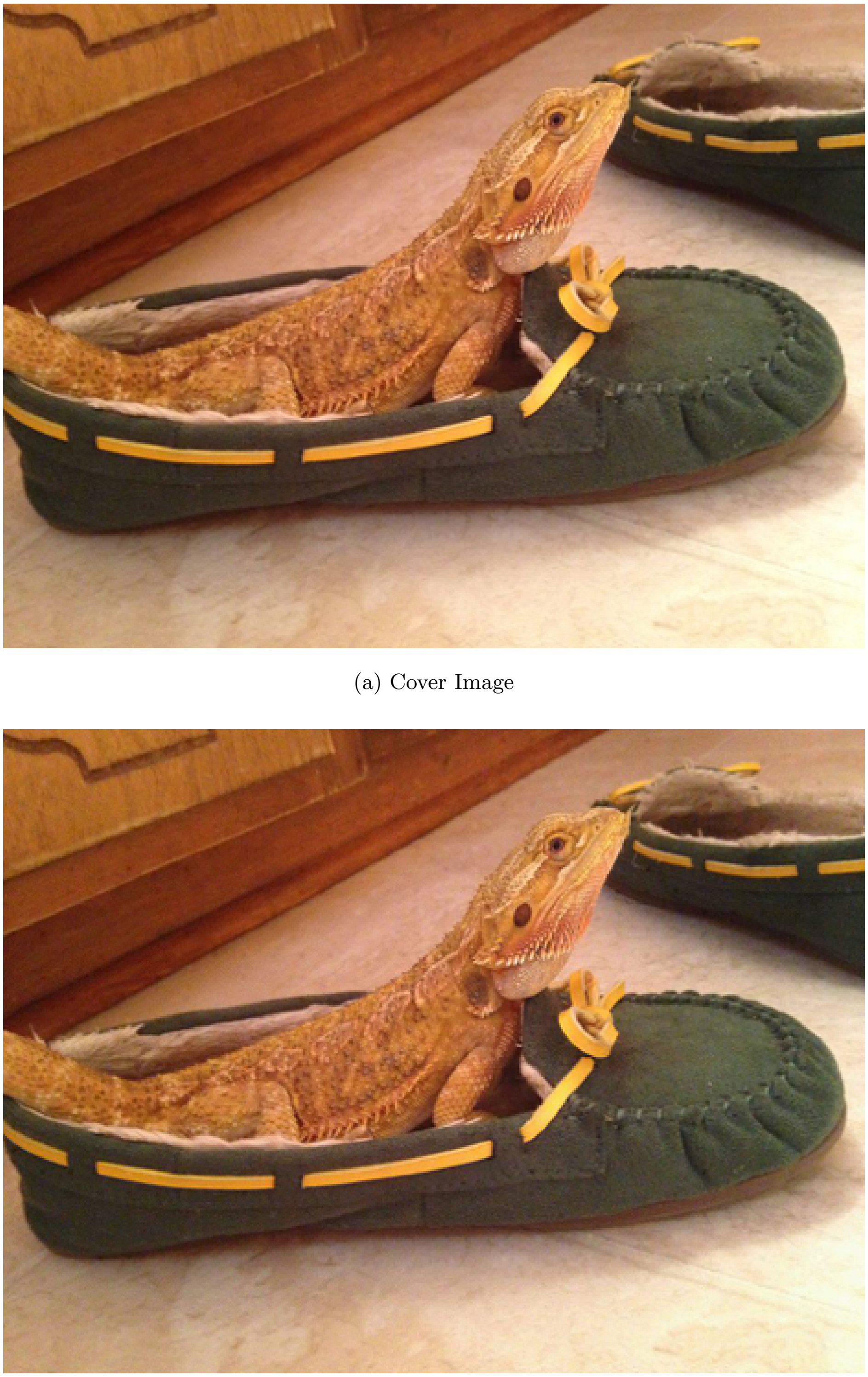


Figure 2.1: Digital Image represented by a bit planes[1]

The images in Figures 2.2 and 2.3 demonstrate how embedding a message in an image changes the image. Figure 2.2a represents the "cover image" (the original image) and figure 2.2b represents the "stego image" (the image that has a message embedded in it). This image was created using the free app SilentEye on a Mac book laptop [9]. By looking at these two images and trying to spot differences, one will most likely conclude that there are no real noticeable differences. However, in order to determine exactly how these images differ, the images were subtracted from each other using Matlab. The subtracted image (i.e. the difference between the cover and stego image) can be seen in Figure 2.3. You can see that almost the whole image changed at least by a small amount. There are also areas througout the image that stand out as having a large difference. This is where the message is hidden.

Because there are so many different methods to embed a message into an image, detecting the presence of a hidden message can be challenging. Each method needs its own detection.. Additionally, people are coming up with more methods for embedding data. This would not be an issue if steganography was not used for criminal purposes but unfortunately it is (see Section 3.3 below on more details related to criminal cases in which steganography was used).



(a) Cover Image

(b) Stego Image

Figure 2.2: Cover Image versus Stego Image

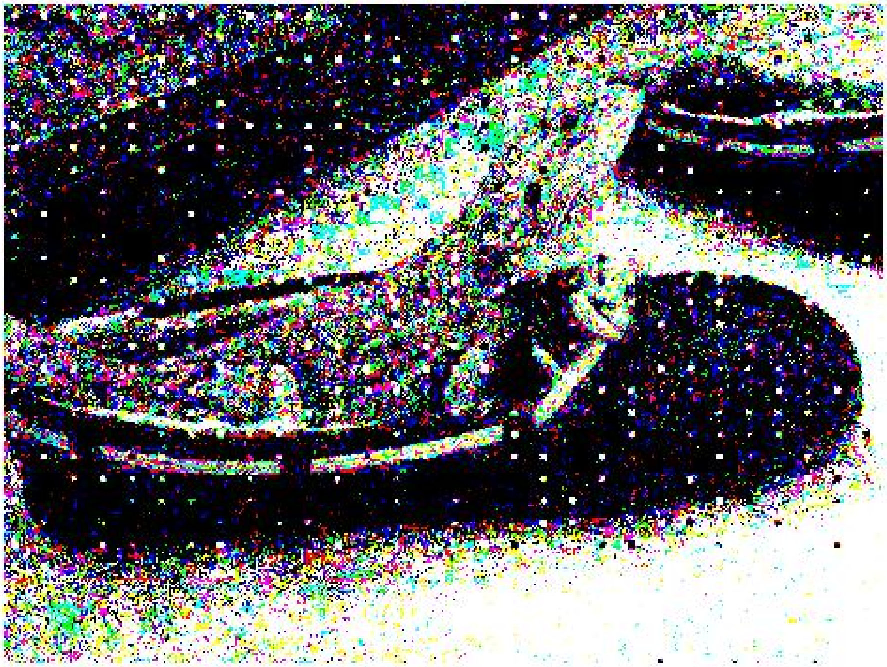


Figure 2.3: Difference between Cover Image and Stego Image in Figure 2.2

2.4 Online Social Networks

Social media is defined in Webster Dictionary as “forms of electronic communication (such as websites for social networking and microblogging) through which users create online communities to share information, ideas, personal messages, and other content (such as videos)” [10]. Online Social Networks (OSNs) are places in which users (including individuals and companies) can post updates on their lives, products etc. and share it with other “friends” or “followers”. A user that has an account on such a network will use the network both as means of sharing updates on their lives as well as “following” their friends, companies or celebrities to see updates from them. This definition of Social Media also covers applications (such as Facebook Messenger or Google Hangouts) where users send private messages to an individual or group of individuals.

Social media was first introduced to the world wide web through a website called “Six Degrees” in 1997 [11]. Social media became more popular in the early 2000s when MySpace was launched. After MySpace came LinkedIn, then Facebook and around 2010 there were dozens of websites that provided social media services. Around 2010 social media became extremely widely used partially due to the fact that businesses started using it. Businesses would have a link to their Facebook page on their website, mention their social media “handles” or addresses in commercials and so on. Social media was also used as a means for clubs and groups to communicate in a group message or invite people to events they were hosting. And as of November 23 2018, of the 7.593 billion people in the world, 3.196 billion of them use social media [12].

Today there are thousands of different social media sites each with slightly different or even very similar purposes [11]. For example, Instagram is used to share images, LinkedIn is for business professionals to network, and Twitter is to share short snippets of information. A study done in October 2018 ranked the popularity of these sites worldwide [12]. Table 2.1 details these rankings. The asterisks denote messaging apps (see next paragraph for importance on these and why they are distinguished from other sites).

Table 2.1: Most Popular Social Media Sites

Rank	**Network**	**Millions of Users**
1	Facebook	2,234
2	YouTube	1,900
3	WhatsApp*	1,500
4	Facebook Messenger*	1,300
5	WeChat*	1,058
6	Instagram	1,000
7	QQ*	803
8	QZone	548
9	Douyin/Tik Tok	500
10	Sina Weibo	431
11	Twitter	335
12	Reddit	330
13	LinkedIn	303
14	Baidu Tieba	300
15	Skype*	300

For purposes of this paper, the "messaging" apps (i.e. networks that are used for private or group messages rather than posting information to the masses) will not be looked at. This is due to the fact that if a person sent a stego image through a messaging app and that message is found, it would be easy to determine the recipient of the message by looking at the person or people included in that group chat. If that same image was posted on a social media site, it is harder to determine the recipient since any one of that persons followers or friends could be the intended recipient. The average number of Facebook "friends" a user has is 338 (as of March 5th, 2018) [13], meaning that if a user posted a stego image on Facebook, someone would have to wade through an average of 338 people to determine who that secret message was intended for.

2.4.1 Why use OSNs to share stego images

Why use social media to share stego images, and does it really work? As stated in Section 2.4, there are billions of Facebook users worldwide. Investigators do not have the time or resources to scan every single image posted on social media for potentially hidden data. According to brand watch, as of June of 2019, **3.2 billion images are shared on social media each day** [14]. Even if investigators or companies did have the resources to analyze 3.2 billion photos each day, once a stego image was found, investigators would have a harder time narrowing down the intended recipient if it was posted on a social media account.

If a suspect involved in a crime sent an e-mail or used a messaging app to send a message to someone prior to the crime, investigators would most likely take the time to analyze that message or at least bring the other person in for questioning. If they found in image in this message and were suspicious that the image contained a hidden message, they may try to find the message. If instead, the suspect posted a picture of their dog on their Facebook page with a hidden message in it (potentially containing information about the crime), any person looking at their account would most likely think it was nothing more than a cute dog picture, except for the intended recipient. Even if the investigators did find out that the dog picture had a hidden message, they would not immediately know who the image was intended for. If the privacy settings are set so the profile is public, anyone with a Facebook account would be able to see this picture. Even if this users profile was private, that still leaves hundreds of friends or followers that have access to image. So not only do investigators have to figure out that this specific picture includes a message, they have to figure out who the message was intended for.

CHAPTER 3. LEGAL ISSUES

Studying OSNs and how steganography can be used on them is important because steganography and social networks can be used for harm. Social media has become increasingly important and useful in solving crimes. Understanding how people might use steganography and social media for harm or even just as a means of communication could lead to breakthroughs in cases.

The following subsections detail examples of when social media was used to solve or commit crimes as well as crimes involving steganography.

3.1 Social Media in Criminal Cases

CBS News has a webpage devoted to “social media related crimes” in which it lists 23 different cases where social media was either used to commit or to solve certain crimes [15]. In many of these cases, the criminals were merely being illogical and bragged about their crimes on social media. This lead their followers to alert authorities and all the evidence they needed to convict the criminals was right there.

In other cases, the police posted photos of the criminal, asking users to come forward with information, similar to an APB you may see on the news. Users on different social media sites will often “share” these posts with their followers and the post will quickly propagate through many users across social media. In many of the cases detailed in the above article, the investigator’s post got around to enough people that someone who did have information saw the post and came forward, allowing the police to catch the criminal.

Other times, crimes were a result of retaliation for things said or done on social networks. For example, in 2011 a woman in Des Moines, Iowa was arrested after she burned her friends house down because the woman “unfriended” her on Facebook. One of the most common crimes committed as a result of social media is burglary. In these cases, a user posted that they were going on vacation and one of their followers, knowing the house would be empty, went over and robbed the house. [15]

3.2 Social Media Evidence

Due to the fact that social media is a more recent means to help solve criminal cases, there are different legal aspects related to when and how social media can be used as evidence in a court case. The two biggest issues related to social media evidence are privacy and authentication.

In general, social media evidence is allowed in court if it is relevant, authentic and gathered by legal means. Relevancy is rather self explanatory. Determining if social media information is authentic is more involved and debated. Authenticity of social media evidence is explored in more detail in section 3.2.2. The other issue then is whether or not information is gathered by legal means. Oftentimes, users want to argue that when investigators or lawyers obtain information from their social media account that it is a breach of their privacy. Therefore, the question of a social media user's right to privacy is explored in section 3.2.1.

3.2.1 Privacy Issues with Social Media Evidence

There are many legal issues related to internet privacy. The Stored Communications Act (SCA) of 1986 was introduced to address that "the Internet presented a host of potential privacy breaches that the Fourth Amendment does not address" [16]. This law addresses the privacy of internet users and defines that it is the internet users personal right or privilege for privacy regarding stored emails and other electronically stored data. The SCA was originally written in relation to Internet Service Providers (ISPs) but does applies to social media sites today. This law details if and when ISP's and social media companies are allowed and/or forced to disclose personal data stored on their networks. In a personal injury case, for example, a person cannot fight a subpoena for information.

One way to acquire social media evidence legally is through the use of a warrant. Additionally, if information was gathered through "passive" means, then accessing that information does not count as a breach of the user's privacy. The author in [17] describes an example of a case in which social media information was gathered passively. The case, Ehling v. Monmouth-Ocean Hospital Service Corp., took place in New Jersey in 2013. In this case, Ehling posted something offensive on her Facebook page, one of her co-workers saw this post and brought it forward to her manager. The hospital then fired her. She tried to sue the hospital, stating that they violated the SCA by accessing her post. However, the hospital did not actively seek out this information, but rather, the information was provided to management without coercion or payment. Therefore, they did have the right to access this information. Although this court case is not an example where social media information was not used as evidence, it does set an important precedence for when information on social media is really considered "private".

Another way in which investigators are allowed to gather information from social media without breaching the user's privacy is in a personal injury case. One case, detailed by [18], was Largent v. Reed. During this case, the court actually ordered Largent to hand over her Facebook login information to the defense counsel so that they could inspect it for evidence. They had two weeks to gather information from it before she was allowed to change her password. This is one of the most intrusive methods of gathering information and is not often popular with the parties in question or even courts. In fact, not all courts have endorse this method.

The author in [18] explains that because the purpose of social media is to share information with people, that users cannot expect that this information will remain private. She states, "Courts generally find that 'private' is not necessarily the same as not public. By sharing the content with others even if only a limited number of specially selected friends the litigant has no reasonable expectation of privacy with respect to the shared content."

3.2.2 Authentication Issues with Social Media Evidence

The main issue with authenticating social media evidence is related to the fact that it is possible for accounts to be hacked, meaning that it is not guaranteed that the post was, in fact, posted by the user in question. Additionally, a user could share their social media log in information with a friend and that friend might post on their behalf.

One author, Jared Staver, details information regarding authentication issues related to social media evidence discussed next [16]. In his paper, he references a Pennsylvania supreme court case, Commonwealth v. Mangel. In this particular case, the social media evidence on a Facebook post was not allowed in court, due to the fact that “the information identifying the defendant was not enough to allow the Facebook posts as evidence”. Staver states, “anyone wishing to use social media evidence in a case must present direct or circumstantial evidence that corroborates the identity of the author”. This can be done, for example, through testimony or an expert witness.

Staver explains that “no aspect of social media content makes it inherently inadmissible in court”. Social media evidence needs to meet the requirements related to the Rules of Evidence. However, the main difficulty with social media evidence (as detailed in the case above) is authentication. Or more specifically, authenticating that the information posted was in fact posted by the user in question. There is currently no consensus across all courts on the best way to do this. Some treat social media evidence authentication the same as a traditional “hard copy” piece of evidence. Others may require more authentication such as the Maryland Supreme Court case Griffin v. State of Maryland in 2011. This court determined that “the party [must] prove no one else was the author of the content”.

There are still many legal issues related to when and how social media can be admissible in court and although there have been court cases related to this in different states supreme courts, there is still no “gold standard” or consensus across all courts in the US. And although social media can be extremely helpful in solving cases, some still argue this is an invasion of privacy. Others are concerned that there is not always a guarantee that a specific user was the one who posted the content on their page.

3.3 Steganography in Criminal Cases

Steganography can be used as a means of communication between criminals or criminal organizations, a way to share classified information or even as a means to hack into peoples computers. This section details specific cases in which steganography was used.

The “first confirmed use of this high-tech form of data concealment in real life” was reported in 2010 by NBC [19]. They reported a case in which Russians were arrested and accused of encoding messages in online pictures. Although the details of this information and how it was hidden are classified, it goes to show that transmitting hidden messages via images can be used to transport secret or malicious data. This article also points out why this method is so difficult to detect and that is that the “sheer numbers of pictures online allow stego images to hide with the safety of numbers”. This article also sites that after 9/11 there were rumors that Al Qaeda hid messages inside pornographic images although it was never confirmed.

In 2012, the Al Qaeda were caught transporting information through the use of steganography. NBC reported that Germany security officials caught a Pakistani Al Qaeda operative with a memory disk containing a pornographic video that embedded over 100 documents outlining plans for terror attacks through Europe [20]. In this case, a video was used because it can hide more data than an image. The authors cite how difficult steganography is to detect since the data is hidden in plain sight also explaining that a pornographic video might be the last place they would look. Professor of Information at UC Berkeley, Steven Weber, explains that a government analyst might be uncomfortable looking at pornographic videos on their laptop screen.

Another case in which steganography was used to hide malware that allowed attackers to hack into people’s computers was detailed in a BBC article from 2015 [21]. In this case, they reported on an instance where hackers posted links to photos on twitter that contained instructions for how to obtain information for the users network. The image and Twitter account was believed to be created from a tool called Hammertoss developed by a Russian group. This tool creates the twitter account, posts a link to a photo on their Twitter account and embeds instructions into that image. In most cases, the instructions are embedded over multiple photos making this difficult to detect and anti-virus software often miss the attack.

In [22], the authors talk about how common it is to use steganography as means of hiding malware. The authors additionally conducted an experiment involving “click bait”. Click bait is essentially some sort of image or link that contains a catchy phrase or interesting information targeted to get users to click on it. In this study, the authors wanted to see how often people would open a link that stated “Is your PC virus-free? Get it infected here!” They were shocked to see that during six months, 409 people opened the link “either by mistake, out of curiosity or stupidity”. This goes to show that people often open up things, sometimes on accident and sometimes by naivety, that can cause harm to their computer.

With steganography it is even easier to get people to “fall for” downloading or opening something that has malware because it is hiding in some sort of image that just by looking at it does not seem harmful. [23] explains that not only is it often easy to get users to open up images with malware hidden using steganography, but most security tools do not even look for data hidden using steganography. This is mostly due to the fact that steganographic data can be difficult to detect and “The performance challenges of scanning almost every file for small, non-impacting anomalies are huge. Its just not practical to check every file coming in and out of an organization at the depth required.”

In all of the cases detailed above, the key to cracking them was knowing that the image or video in question needed to be analyzed further to see if there was something hidden in it. If taken at face value, the pornographic video, did not look any more suspicious than a pornographic video can be, but further investigation revealed it was also a means to communicate terror attacks across a large organization. The importance of the following research, then, is to complete case studies of how steganography can be used in social media and how to detect the presence of such online steganographic images.

CHAPTER 4. PREVIOUS WORK

4.1 Previous Work Overview

The following publications analyzed using social media as a means for sharing secret messages through image steganography:

- Transmitting Hidden Information using Steganography via Facebook by Nathaniel D. Amsden, Lei Chen and Xiaohni Yuan [24]
- Analysis of Facebook Steganographic Capabilities by Nathaniel D. Amsden and Lei Chen [25]
- Pictographic steganography based on social networking websites by Feno Heriniaina R. and Xiaofeng Liao [26]
- Using Facebook for Image Steganography by Jason Hiney, Tejas Dakve, Krzysztof Szczypiorski and Kris Gaj [27]
- The OSN-Tagging Scheme: Recoverable Steganography for Online Social Networks by Tayanna Morkel [28]
- Secret Message Sharing Using Online Social Media by Jianxia Ning, Indrajeet Singh, Harsha V. Madhyastham, Srikanth V. Krishnamurthy, Guohong Caox, and Prasant Mohapatraz [29]
- Steganographic Checks In Digital Forensic Investigation: A Social Networking Case by Brian Cusack and Aimie Chee. [30]

All of these articles focused on the technology of creating and sharing images via social media. They did not, however, consider detecting the presence of a secret message in online images. Most of the articles focused on Facebook, most likely since it is the most popular OSN. This chapter details the findings of these articles and creates a comprehensive look at the work that has already been done on this topic.

4.2 Success of distributing Stego Images on Social media

There are many different methods to embed a message into an image. Some of these methods include least significant bit insertion (described in Section 2.3 above), masking, filtering, transformations and watermarking. In [24], the authors detail how different methods work and their pros and cons. If a user did not want to or was not technically capable of implementing an algorithm themselves, there are a wide variety of applications out there available for free that, given an image and a message, will perform the necessary embedding to create a stego message. A quick google search leads to the following two links, both of which list ten or more free apps for performing steganography. These links list apps that can be downloaded onto an iPhone or Android phone in addition to apps that can be downloaded onto a Windows or Mac PC. So creating a stego image can be as simple as downloading a free app. Many more apps for mobile phones can be found on the Apple store or Google play store.

- https://www.networkworld.com/article/2291708/security/130370-15-FREE-steganography-apps-formobile-devices.html
- http://resources.infosecinstitute.com/steganography-and-tools-to-perform-steganography/#gref

Even though there are tools available for download that create stego messages, not all these tools and methods work effectively with social media. This is due to the fact that many social networking apps compress data in such a way that the message is lost.

In the seven papers listed in Section 5.1 above, they completed case studies that used different apps or algorithms to create stego images, uploaded these images to different social media sites, then downloaded the images to see if they could detect the hidden message. Most of the case studies focused on Facebook but some also included Twitter, Google+ and Flickr. Facebook rarely retained the hidden message, Google+ always retained the hidden message and Twitter and Flickr were somewhere in the middle. Table 4.2 below shows which sites retained the stego message after uploading and downloading the image created from the given app or algorithm. The ✓denotes the message was retained all of the time, the ✓* means the message was retained in some instances, the X denotes the message was not retained and the N/A means the given algorithm or app was not tested with the listed Social Network. The subscripts list which source that particular study was completed under. Table 4.1 is a key of the coloring in Table 4.2. The key reason that the message was not recoverable is due to the way in which social media sites compress images before posting them.

Table 4.1: Key for "Social Networks versus Stego Apps" Table

Key
The stego message created by the listed tool was always retained after uploading it to and then downloading it from the listed Social Network
The stego message created by the listed tool was sometimes retained after uploading it to and then downloading it from the listed Social Network
The stego message created by the listed tool was never retained after uploading it to and then downloading it from the listed Social Network
There is no existing work done on whether or not the listed app works with the listed Social Network

Table 4.2: Social Networks versus Stego Apps

	Facebook	**Google+**	**Twitter**	**Flickr**
EOF	X[30]	✓[30]	N/A	N/A
F5	X[27],[29]	✓[29]	✓[29]	X[29]
Ghost Host	X[29]	✓[30]	X[29]	X[29]
Incognito	X[27]	N/A	N/A	N/A
Invisible Secrets	X[30]	✓[30]	N/A	N/A
JP Hide & Seek	✓*[24],[25],[27],[30]	✓[30]	N/A	N/A
Open Puff	X[27]	N/A	N/A	N/A
Our Secret	X[27]	N/A	N/A	N/A
Outguess Rebirth	X[27],[29]	✓[29]	✓[29]	X[29]
S-Tools	X[30]	✓[30]	N/A	N/A
Silent Eye	X[30]	✓[30]	N/A	N/A
Steganography	X[27]	N/A	N/A	N/A
Steghide	✓*[27],[29],[30]	✓[29],[30]	✓[29]	X[29]
YASS	✓*[29]	✓[29]	✓[29]	✓*[29]

As you can see in Table 4.2, there were no methods used that consistently retained stego images after posting them on Facebook. Therefore, many studies went further and investigated different methods that would consistently retain stego messages after sharing them on social media.

One method proposed to mitigate this issue can be found in Morkel's paper [28]. In this paper, the author proposes using an “OSN Tagging scheme” to embed message data. This method follows the basic algorithm “Caronnis tagging scheme”. This tagging scheme first identifies specific locations based on variance and brightness level that are suitable for inserting tags. These locations are NxN bit rectangles. Then the brightness level of these tags is adjusted to create a message. For example, if the brightness is increased, this could represent a 1 and if the brightness is decreased, this could represent a 0. Morkel implemented this algorithm and tested it by uploading images to Facebook and then re-downloading them. They tested different size tags and found with a tag size of 8x8 pixels there was a 100% recovery rate of the message. In this experiment, they also tested different brightness adjustment levels and recommend an adjustment of 2% to mitigate how Facebook compresses images. The drawback with this method, however, is that in order to extract the message, it must be compared to the original image. This means that the person who is trying to communicate through this secret message has to get the original image to their correspondent(s) first, meaning that a line of communication must be established prior to sharing the image to everyone on social media.

Another alternative method was described in [26]. The authors propose using pictographs as a way of communicating rather than embedding messages. Pictographs are a series of images that, when compared with a dictionary of some sort spell out a message. The example they give in their paper is that a picture of a car could mean “car”, a picture of a heart could mean “love” etc. In this method, it does not matter if any bits are changed during a Social Networks compression because even if its compressed or resized, a car will still look like a car after its posted. Therefore, this algorithm will work 100% of the time. However, like the algorithm listed above, there needs to be an exchange of the dictionary prior to the message extraction. The message is completely meaningless without knowing how to interpret it.

4.3 Summary

Many different studies have been done to see which social networks retain hidden messages using different stego apps. The consensus between these articles is that Facebook and Flickr rarely retain stego messages, Twitter sometimes retains messages and Google+ always does. Additional studies in [28] and [26] develop different methods of sending stego data over Facebook that does retain the message. Although the authors in both papers were able to devise methods that work, there are drawbacks to their methods. Also, although Google+ always retains messages, the site was discontinued in April of 2019, so no further studies will be done on that site. Facebook, being the most popular social media site, would be the ideal network to study. Some methods that other articles found useful in retaining Facebook stego images include posting images as “cover photos” since they are embedded differently, making your image a certain size before it is posted, or using a tagging scheme versus other embedding methods.

The next step aims at finding a combination of stego app and social network that consistently retains the stego message. Chapter 5 below includes details on this case study. After a method of creating and sharing stego messages is found, additional analysis will be done to determine if there is a way to detect that a certain image contains a secret message. This will be done by analyzing quantization matrices (QMs). Details on what QMs are and findings related to analyzing them can be found in Chapter 6.

CHAPTER 5. CASE STUDY 1: HOW TO SHARE STEGO MESSAGES ON SOCIAL MEDIA

5.1 Background

Much of the previous research done on sharing steganographic images on social media found that posting images on social media often strips the message data from the image due to the different compression methods that social media sites use (See Chapter 4 above). The research below focuses on finding consistent methods to share stego messages on social media sites so that the stego message is retained. Table 5.1 shows the top 11 most popular social networking apps why we chose to include or exclude them in our research. No messaging or foreign apps were used. See Section 2.4 above for details on why messaging apps were skipped over.

Table 5.1: Most Popular Social Media Sites and Why to Research or Not To

Rank	**Network**	**Will Research**	**Reasoning behind decision**
1	Facebook	Yes	Most popular app
2	YouTube	No	Shares Vidoes, not images
6	Instagram	No	Does not allow the ability to download images, so there is no way to retrieve stego messages
8	QZone	No	Foreign App
9	Douyin/Tik Tok	No	Foreign App
10	Sina Weibo	No	Foreign App
11	Twitter	Yes	Second most popular app that hasn't been disqualified

Previous work found that they were able to retain message data after uploading to Facebook in some circumstances. One study found that Facebook sometimes retained data when the JP Hide & Seek Stego app was used. Another study had success in retaining stego messages when Facebook photos were uploaded as “Cover Photos” instead of timeline photos. Previous studies had much more success when it came to retaining messages after sharing stego images on Twitter.

We chose the following stego applications to create stego images:

Table 5.2: Steganography Applications to Use in Research

App Name	**Where to Find**	**Compatibility**
JP Hide & Seek	http://linux01.gwdg.de/ alatham/stego.html	Windows
Silent Eye	https://silenteye.v1kings.io/	Windows, Linux and Mac

JP Hide & Seek was chosen because previous studies found that JP Hide & Seek had some success with creating stego images that Facebook retained. Silent Eye was chosen because it can be used on either a Mac or a PC. Both applications can be downloaded for free from the sites listed in Table 5.2.

5.2 Method

Using the information gathered from previous studies we constructed the procedure detailed in Figure 5.1 to create and share stego images. From these images created, we were able to determine the viability of passing a hidden message this way. Note that there were two different methods used. Method 1 uploaded the image directly from the phone to the social media site, whereas method 2 uploaded the photo from the phone to the PC, then from the PC to the social media site. Table 5.3 shows the success rate of these two different methods and if, in fact, these two methods resulted in different findings.

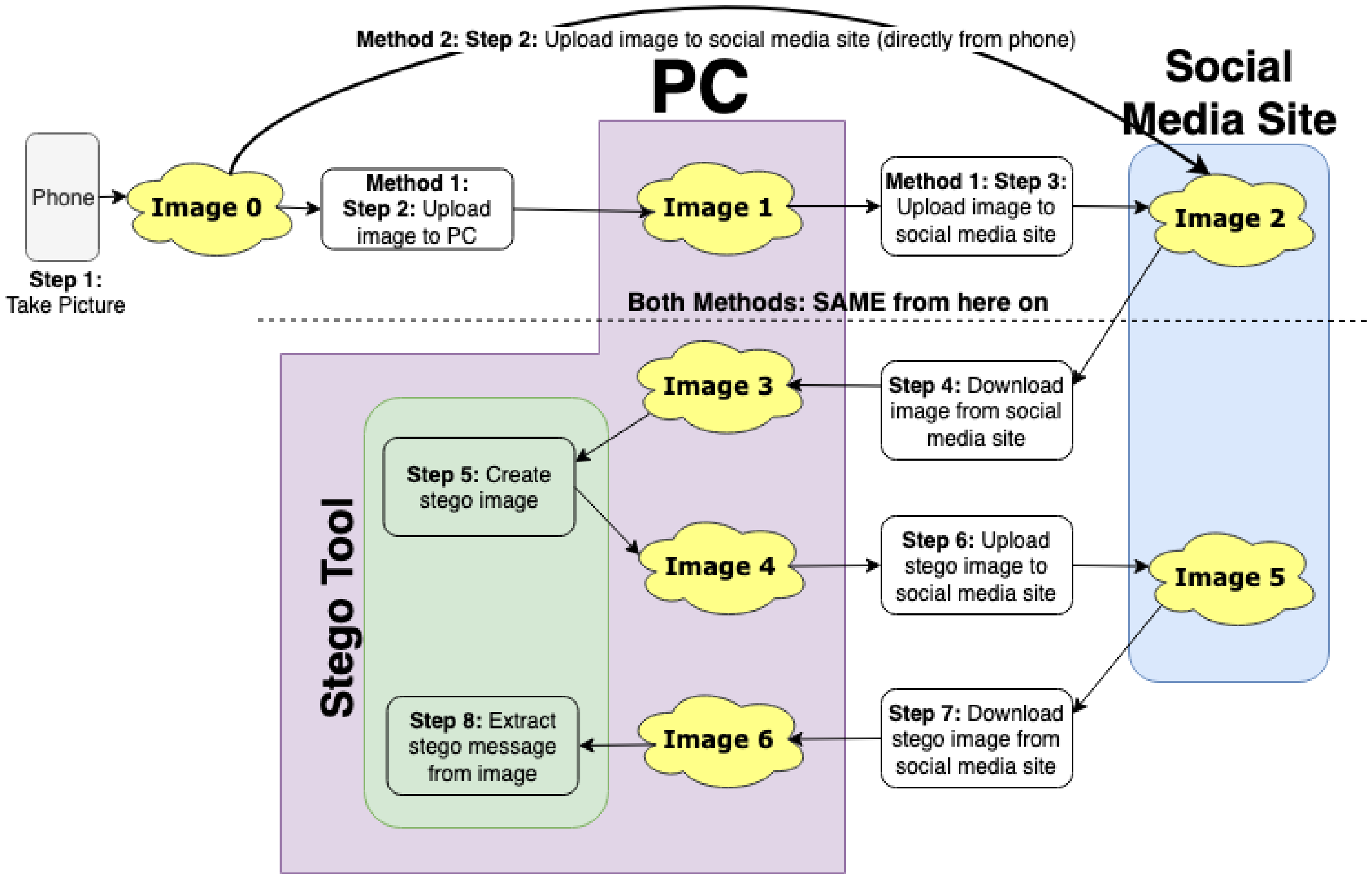
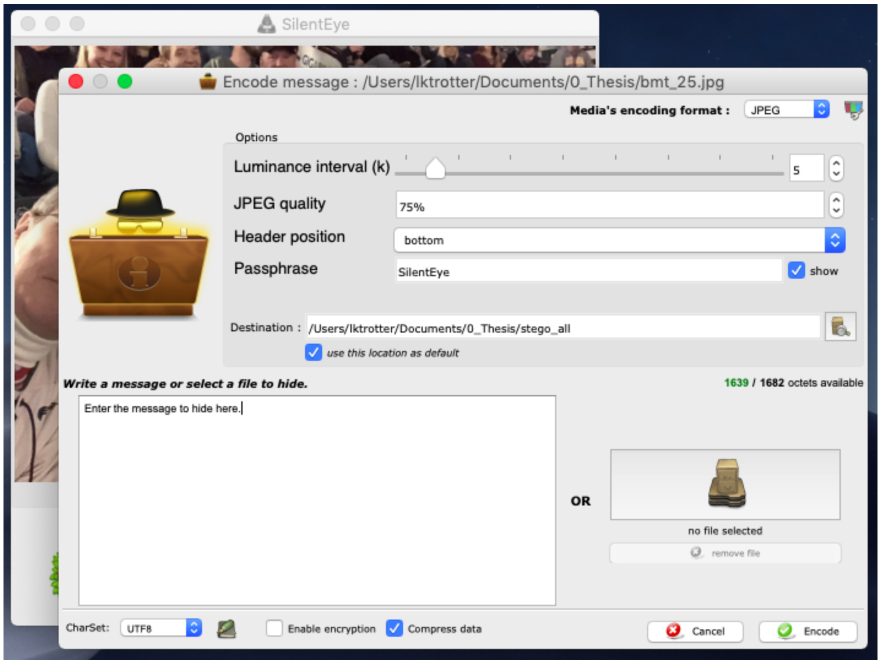

Method 2: Step 2: Upload image to social media site (directly from phone)
PC
Social Media Site
Phone
Step 1: Take Picture
Image 0
Method 1: Step 2: Upload image to PC
Image 1
Method 1: Step 3: Upload image to social media site
Image 2
Both Methods: SAME from here on
Image 3
Step 4: Download image from social media site
Stego Tool
Step 5: Create stego image
Image 4
Step 6: Upload stego image to social media site
Image 5
Step 7: Download stego image from social media site
Image 6
Step 8: Extract stego message from image

Figure 5.1: Method for creating and sharing stego images

In regards to Figure 5.1, all images that were uploaded to Facebook were uploaded as Facebook "Cover Photos" (which is a type of Facebook image that is treated differently than "timeline photos"). JP Hide & Seek was used on a Windows PC. The message hidden using JP Hide & Seek was done so by hiding a text files inside the image. The same text file was used for all of the trials. SilentEye was used on a Mac laptop. SilentEye allows the user to either hide a file inside the image or type a message directly into the application. For trials completed with SilentEye, the message was hidden by typing a message directly into the tool. The same message was used for all trials. SilentEye also allows the user to specify a Luminance interval, JPEG quality, Header position, PassPhrase and whether or not to enable encryption and compress the data. For the purposes of this experiment, the default values were used. Figure 5.2 is a screenshot of the default values used when encoding a message inside an image using the SilentEye application.



Encode message : /Users/lktrotter/Documents/0_Thesis/bmt_25.jpg

Media's encoding format : JPEG

Options

Luminance interval (k) 5

JPEG quality 75%

Header position bottom

Passphrase SilentEye ☑ show

Destination : /Users/lktrotter/Documents/0_Thesis/stego_all

☑ use this location as default

Write a message or select a file to hide.

1639 / 1682 octets available

Enter the message to hide here.

no file selected

CharSet: UTF8

☐ Enable encryption ☑ Compress data

Figure 5.2: SilentEye application default values for encoding

During this method, there are four different images created that are saved to the PC. The images created correspond to Image 1, 3, 4 and 6 in the method depicted in figure 5.1. These images may also be referred to using the following naming convention:

- Image 1: Original Image before [Facebook or Twitter]
- Image 3: Original Image after [Facebook or Twitter]
- Image 4: Stego Image before [Facebook or Twitter]
- Image 6: Stego Image after [Facebook or Twitter]

5.3 Findings

Table 5.3 shows how the social media app/stego app combination fared. The "Method" column refers to if method 1 or method 2 (described above in figure 5.1) was used. Rows that have a method of "both" contain the total of how both methods fared. The "# of Trials" column is the number of trials performed with that particular pairing. The "% success" column is the percentage of trials in which the stego message was successfully retrieved from Image 6. The "Where Failure(s) Occurred" column describes where in the process the image application failed to extract the stego message from the image (if a failure occurred).

5.3.1 Failures

This subsection details findings related to where during the process failures occurred and how these can be avoided. See the last column in table 5.3 for which trials these failures occurred in.

When using Silent Eye, any failures that occurred did not occur after the photo was shared on social media, but rather immediately after the stego tool was used. All stego images that were created successfully did retain the stego message after being uploaded to Twitter or Facebook. So the failure was with Silent Eye creating an image and not in social media stripping the image away. In order to avoid these failures, one would merely need to verify that the image SilentEye created does have the stego message in it prior to sharing it on social media. To do so, they would run their new image through the SilentEye decoding process and verify a message is found. If no message was found, then they know that no message will be found if they share that image on social media. If this happens, they can try using a different image.

Table 5.3: Success of Different Stego Messages on Facebook and Twitter

Social Media App	**Stego App**	**Method Used**	**# of Trials**	**% success**	**Where failure(s) occurred**
Facebook	Silent Eye	1	20	80.0%	All failures occurred during step 5 in the diagram above.
”	”	2	37	91.9%	All failures occurred during step 5 in the diagram above.
”	”	Both	57	87.71%	All failures occurred during step 5 in the diagram above. 50 of the 57 trials were successful.
Facebook	JP Hide & Seek	1	15	100.0%	N/A
”	”	2	40	32.5%	All failures occurred during step 8 in the diagram above.
”	”	Both	55	50.90%	All failures occurred during step 8 in the diagram above. 28 of the 55 trials were successful
Twitter	Silent Eye	1	15	86.7%	All of the failures occurred during step 5 in the diagram above
”	”	2	40	67.5%	4 of the failures occurred during step 5 in the diagram above and **9 of the failures occurred during step 8 in the diagram above.
”	”	Both	55	72.72%	6 of the failures occurred during step 5 in the diagram above and **9 of the failures occurred during step 8 in the diagram above. 40 of the 55 trials were successful
Twitter	JP Hide & Seek	1	15	20.0%	All failures occurred during step 8 in the diagram above.
”	”	2	40	5.0%	All failures occurred during step 8 in the diagram above.
”	”	Both	55	9.09%	All failures occurred during step 8 in the diagram above. 5 of the 55 trials were successful

Using JP Hide & Seek to create stego images to share on Twitter was only successful in 5 of the 55 trials. This method would not be recommended to use for sharing stego images over Twitter. Silent Eye would be recommended instead. The SilentEye tool can be used on a Mac or a Windows PC.

Another interesting finding is that JP Hide & Seek was much more successful when the photo was uploaded from the phone to the PC then the PC to the social media site instead of uploading the photo directly from the phone to the social media site.

5.3.2 Images

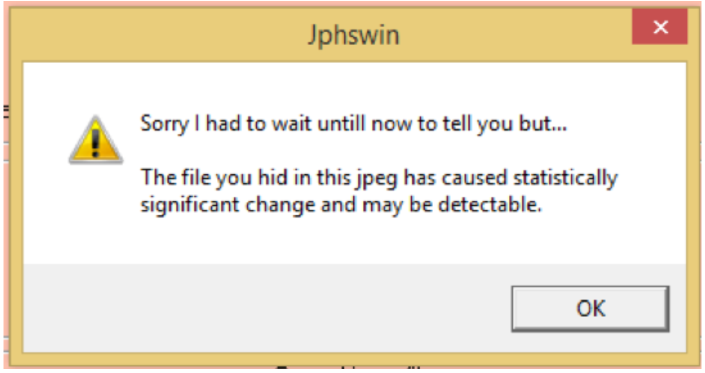
This subsection details findings related to the stego images that were created.

5.3.2.1 JP Hide & Seek Images

When using JP Hide & Seek on images downloaded from social media sites (i.e. images that have gone through some sort of compression algorithm), the tool generated a warning message, saying "The file you hid in this jpeg has caused statistically significant change an may be detectable" (see figure 5.3). If JP Hide & Seek was used on an image prior to it being uploaded to social media (i.e a non-compressed image), the JP Hide & Seek warning message did not pop up. The four images below depict cover and stego images generated using JP Hide & Seek before the social media compression versus the cover and stego images generated using JP Hide & Seek after the social media compression. Even though the tool says the change may be detectable, from a personal observations, it does not appear as if there is an observable difference. The four images below show the progression of the image.

Note, The figures in Figure 5.4 relate to the figures described in Figure 5.1 above in the following ways:

(a) Figure a corresponds to "Image 1" in the method



Jphswin

Sorry I had to wait untill now to tell you but...

The file you hid in this jpeg has caused statistically significant change and may be detectable.

Figure 5.3: JP Hide & Seek warning message

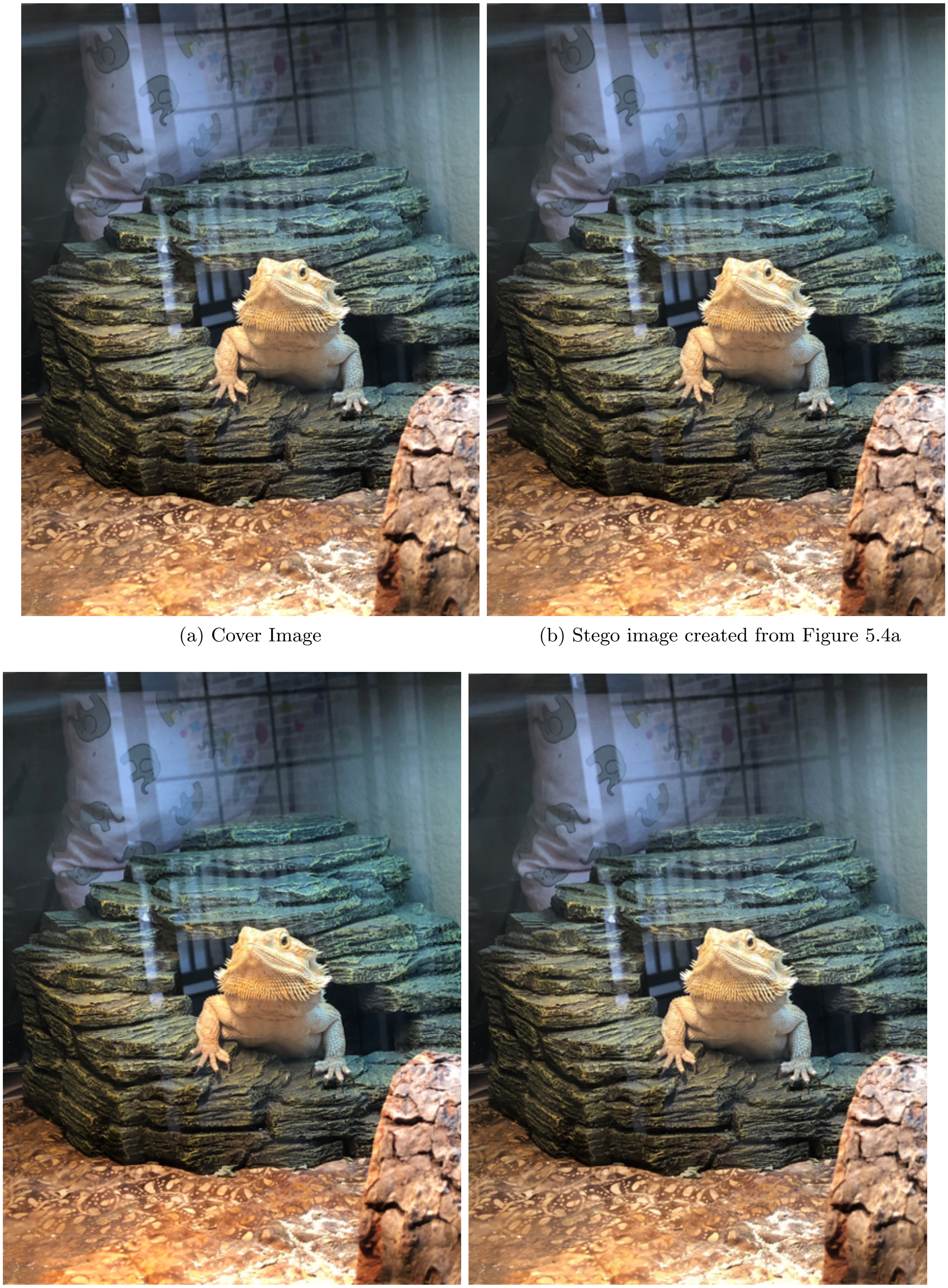
(b) Figure b is "Image 1" in the method ran through the stego app. This image is not used or shown on the diagram above because the only way to successfully maintain stego data was to upload it to the social media site before running the image through a stego app

(c) Figure c corresponds to "Image 3" in the method

(d) Figured corresponds to "Image 4" in the method

To fully understand how different these images are, a subtraction was done between the two images using Matlab. The image differences can be seen in Figure 5.5.

Figure 5.5a shows the difference between Figures 5.4a and 5.4b. This represents the difference between the original image and the stego image created from that. Here there are small, seemingly random dots of differences throughout the image. Figure 5.5b shows the difference between Figures 5.4c and 5.4d. This represents the difference between the image downloaded from Facebook and the stego image created from that. Here that the dots of differences are much larger, because the pixels themselves are larger. This is due to the fact that the image downloaded from Facebook is a lot smaller than the original so there are less pixels in the image and therefore they look larger. Although the differences themselves are bigger, they are still scattered seemingly randomly throughout the image.



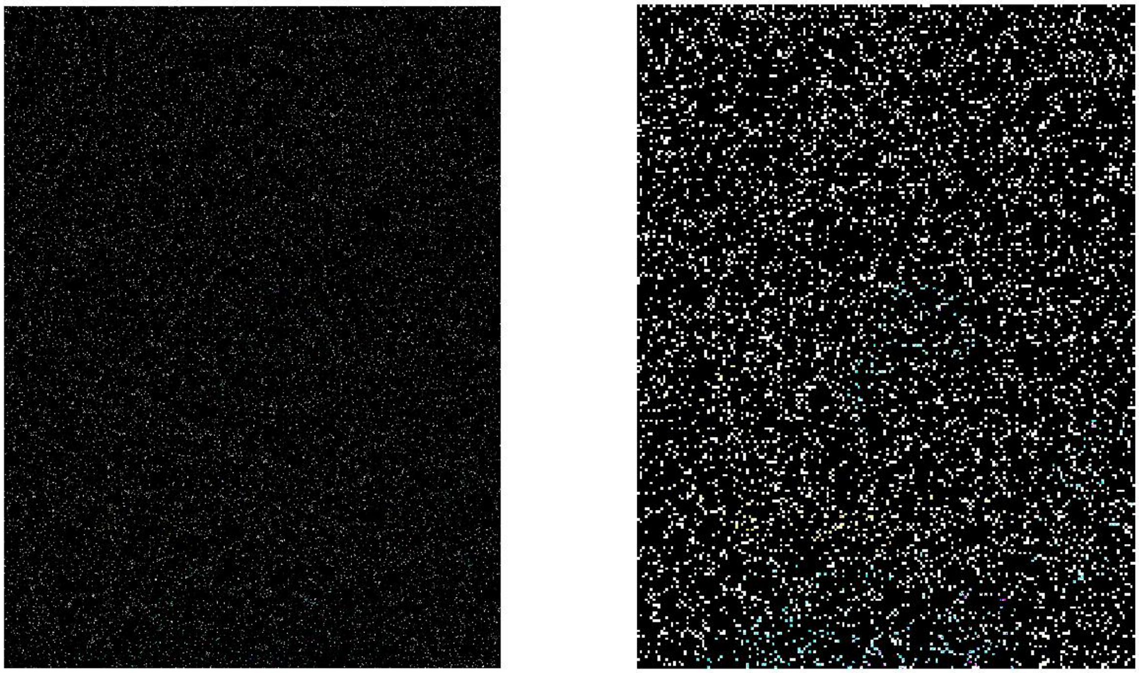
(a) Cover Image

(b) Stego image created from Figure 5.4a

(c) Cover Image after uploading/downloading Image 5.4a to and from from Facebook

(d) Stego image created from Figure 5.4c

Figure 5.4: JP Hide & Seek Comparisons



(a) Difference between Figures 5.4a and 5.4b

(b) Difference between Figures 5.4c and 5.4d

Figure 5.5: JP Hide & Seek Differences

When it comes to using the JP Hide / Seek tool, when Figure 5.4b was created from Figure 5.4a, there was no warning saying that the difference may be noticeable. However, when Figure 5.4d was created from 5.4c, there was a warning saying that the difference may be noticeable. Looking at the actual differences, it makes sense that the warning would be present because there is a more drastic difference when using the compressed image.

5.3.2.2 SilentEye Images

Another interesting finding is that when using SilentEye, if the original image contained areas that were mostly one solid color, the stego image created had little squares throughout it where the color was changed that was visible to the naked eye. Even if someone looking at the photo did not know anything about steganography, they would probably notice that the image had weird squares throughout it and wonder why. This was true both before and after images were put through the

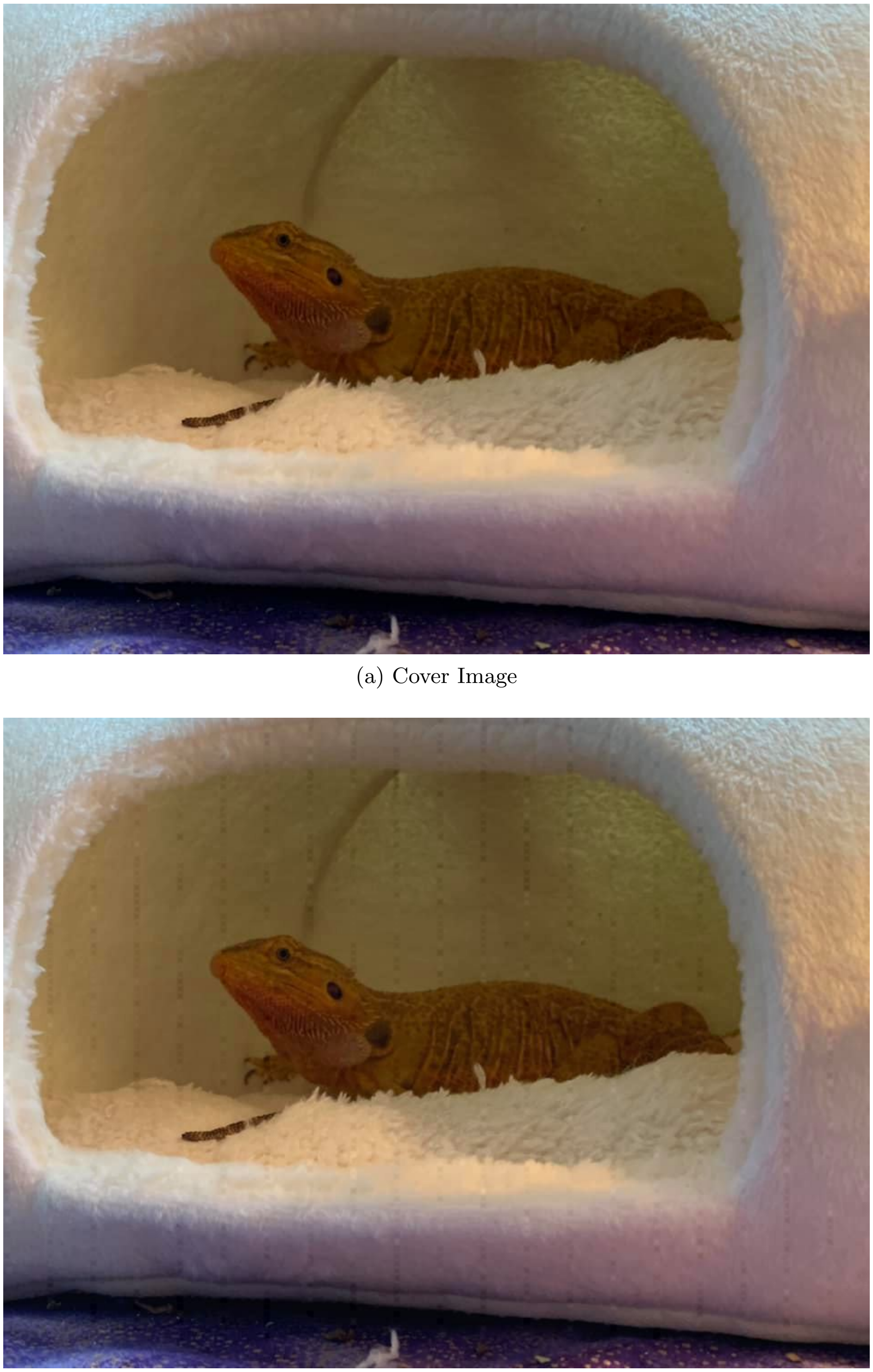
Facebook or Twitter compression method. So, being able to visually detect that the image had been tampered with was not related to the size of the image or if the image went through any compression. When the original image contained a lot of different colors and textures, the stego image created did not have changes that could bee seen with the naked eye. Figure 5.6 is an example of an image with a lot of solid color throughout it before and after the stego message was embedded in it. Figure 5.7 is an example of an image with many different colors and textures throughout before and after a stego message was embedded in it. If a person using Silent Eye did not verify that their image still looked "normal" after the message was embedded, someone might be able to tell that an image has been changed just by looking at it.

To investigate this further, Matlab was again used to find the difference between such images. Figure 5.8a contains the difference between the cover and stego photos when a photo with a almost solid color background was used. Here, the differences stand out and are very obvious. Unlike when JP Hide & Seek is used, here the differences look like they follow a specific pattern rather than random. Additionally, you can see parts of the image itself, as if it is doing some sort of compression to the image and changing almost all of the bits.

Figure 5.8b contains the difference between the cover and stego photos when a photo with more texture is used. Although the differences still look like they form pattern, they do not stand out as much because they are almost swallowed up or hidden in the textures of the background. In the darker, more solid areas of the image (i.e. in the corners), the differences still stand out.

5.4 Takeaways

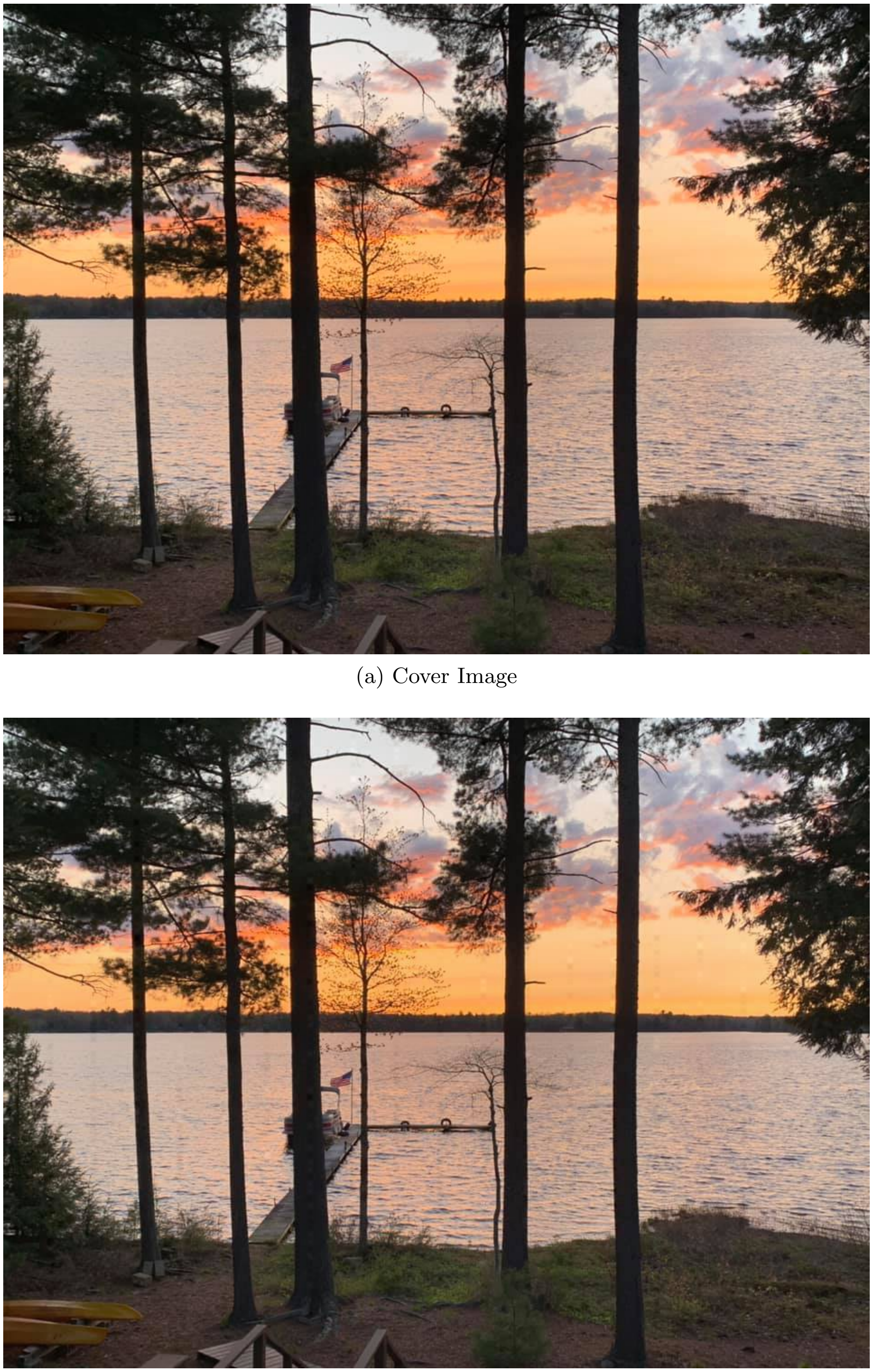
One of the takeaways that can be gathered from the first case study is that it is possible to create and share stego images on Facebook and Twitter. However, to do so, there is a very specific procedure that must be followed. The first thing you have to do is upload and download the image from the social media site in question so that it goes through the compression process prior to hiding a message inside of it. The downside with this is that when a smaller size image is used, there are less bits to hide the message in so the message may be more noticeable.



(a) Cover Image

(b) Stego Image

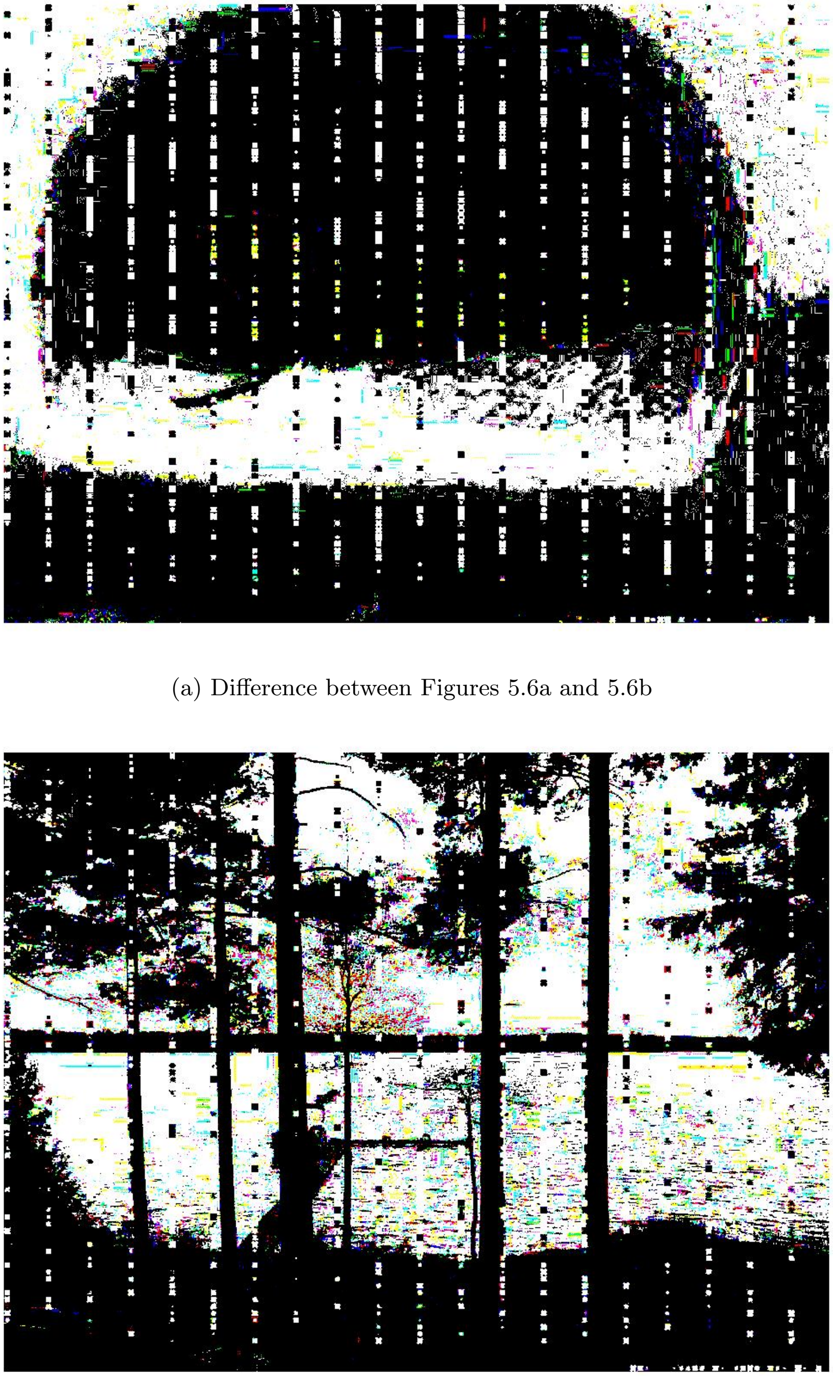
Figure 5.6: Silent Eye Cover Image vs Stego Image 1



(a) Cover Image

(b) Stego Image

Figure 5.7: Silent Eye Cover Image vs Stego Image 2



(a) Difference between Figures 5.6a and 5.6b

(b) Difference between Figures 5.7a and 5.7b

Figure 5.8: Difference between Stego and Cover Photos created using SilentEye

Another finding is that the method only works with Facebook cover photos. From an investigators perspective, the pro with this is that there are a lot less images that need to be scanned for stego messages. The drawback to this, however, is that cover photos are public, meaning that anyone who has a Facebook account can see anyone else's cover photo. So the subject pool of who the message might be intended for if a stego message was found inside of a Facebook cover photo is anyone who has a Facebook account.

Final recommendations for creating stego images to share on social media are to verify the stego message exists before posting the image on social media. This needs to be done because there were many cases in which the message was not successfully embedded into an image even before posting it to the social media site. Lastly, if you do use this method to share a stego message on social media, it is recommended that you delete the original photo. That way, someone would not be confused or suspicious if they saw someone post two of what looked like the exact same photo.

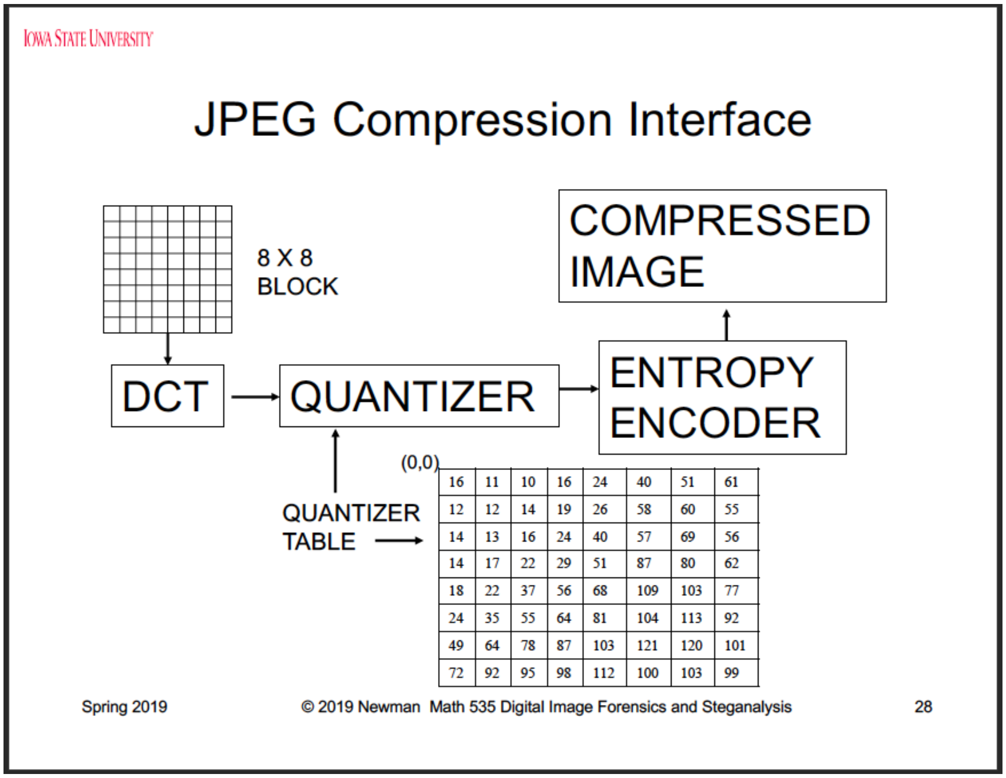
CHAPTER 6. CASE STUDY 2: HOW TO DETECT STEGO IMAGES ON SOCIAL MEDIA USING QUANTIZATION MATRICES

In order to try and find a way to detect the presence of stego images on social media, a second case study was done. This case study focused on analyzing quantization matrices (QMs) to see if they could be used to draw any conclusions about the image in question. To do so, all of the images created in Chapter 5 were all ran through a tool to extract QMs.

6.1 QM Overview

A quantization matrix is something that only JPEG images have. When an image is compressed into a JPEG image, it goes through the process depicted in Figure 6.1. First the image is broken up into 8x8 bit blocks, a Discrete Cosine Transfer is done on those blocks, then that matrix is divided by a quantizer table and put through an entropy encoder. The quantization matrix is essentially a "constant" that everything is divided by during this process. The QM is designed to get rid of unimportant bits and keep the important ones. There is no standard quantization matrix, although there are many specific recommended ones. Since there is no industry standard table to use, many cameras or application that compresses images into JPEGs use their own QM across all of their images. Because of this, sometimes knowing the QM of a JPEG image can give an idea of where the image came from.

There have been previous studies in which QMs were analyzed to determine if information can be gathered from them. Hany Farid, a professor who specializes digital images, conducted such a study in 2008 in which he tried to use QMs to determine what digital camera was used to take that image [31]. He compared images take from different digital cameras and found that 62 of the 204 cameras had a unique QM. He concluded that "This simple observation allows for a rather crude form of digital image ballistics, whereby the source of an image can be confirmed or denied." The study done for this paper, similarly aims at trying to determine if different social media or stego applications have unique QMs. If they do, then one could potentially use QMs to determine the origin of that image.



IOWA STATE UNIVERSITY

JPEG Compression Interface

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

Spring 2019 © 2019 Newman Math 535 Digital Image Forensics and Steganalysis 28

Figure 6.1: How Quantization Matrices are used to compress images [2]

6.2 QM Method

After the images were created using the method above, all of the PC images (i.e. Images 1, 3, 4 and 6 in the Figure 5.1) were run through one of two applications to retrieve the Quantization Matrix (QM) and Table Quality value. This was done to determine if the social media sites and/or the stego apps used have a standard QM.

The following tools were used to generate the QM's:

- On a Mac PC, the following website was used: https://29a.ch/photo-forensics/#jpeg-data

- On a Windows PC, the ”JPEG Snoop” app was used, which can be downloaded from here: https://www.impulseadventure.com/photo/jpeg-snoop.html

6.3 QM Findings

Through analyzing the images created, the following conclusions were drawn.

First and foremost, all images downloaded from twitter prior to creating the stego image (image 3 in figure 5.1) were the same. Table 6.1 and 6.1 show what the standard QM for twitter images is.

Table 6.1: Standard Twitter Quantization Tables

Table 0	5	3	4	4	4	3	5	4	Table 1	5	5	5	7	6	7	14	8
	4	4	5	5	5	6	7	12		8	14	30	20	17	20	30	30
	8	7	7	7	7	15	11	11		30	30	30	30	30	30	30	30
	9	12	17	15	18	18	17	15		30	30	30	30	30	30	30	30
	17	17	19	22	28	23	19	20		30	30	30	30	30	30	30	30
	26	21	17	17	24	33	24	26		30	30	30	30	30	30	30	30
	29	29	31	31	31	19	23	34		30	30	30	30	30	30	30	30
	36	34	30	36	28	30	31	30		30	30	30	30	30	30	30	30

Not all images downloaded from Facebook prior to creating a stego image (image 3 in figure 5.1) had the same Quantization Matrix. But 41 of the 57 images created (about 71.9%) did have the same QM. See Table **??** for the most common quantization matrix seen for images downloaded from Facebook.

Table 6.2: Most Common Facebook Quantization Table

Table 0	9	6	7	8	7	6	9	8	Table 1	10	10	10	14	12	14	27	15
	8	8	10	10	9	11	14	23		15	27	57	38	32	38	57	57
	15	14	13	13	14	28	20	21		57	57	57	57	57	57	57	57
	17	23	34	30	35	35	33	30		57	57	57	57	57	57	57	57
	32	32	37	42	53	45	37	39		57	57	57	57	57	57	57	57
	50	40	32	32	46	63	47	50		57	57	57	57	57	57	57	57
	55	57	60	60	60	36	45	66		57	57	57	57	57	57	57	57
	70	65	58	70	53	59	60	57		57	57	57	57	57	57	57	57

Another important discovery is that all images created using SilentEye had the same quantization matrix. Table 6.3 depicts what this table was. This was true for images created by Silent Eye prior to uploading it to either social media site (Image 4 in Figure 5.1 above) AND after the image was uploaded to then downloaded from either social media site (Image 6 in Figure 5.1 above). Sixty other images not used in this study were analyzed to see if the standard QM for SilentEye images was found in other images. The images tested included images from a PC, from an iPhone, from an Android phone, images downloaded from Facebook, images downloaded from Twitter and screenshots taken on an iPhone, to name a few. None of these sixty images had the same QM that the SilentEye photos had. This is important because if an image downloaded from Facebook or Twitter had the QM listed in 6.3, then we can conclude with a fair amount of certainty that the image in question was created from Silent Eye and therefore has a secret message hidden inside it.

Table 6.3: Standard Silent Eye Quantization Table

Table 0	8	6	6	7	6	5	8	7
	7	7	9	9	8	10	12	20
	13	12	11	11	12	25	18	19
	15	20	29	26	31	30	29	26
	28	28	32	36	46	39	32	34
	44	35	28	28	40	55	41	44
	48	49	52	52	52	31	39	57
	61	56	50	60	46	51	52	50

Table 1	9	9	9	12	11	12	24	13
	13	24	50	33	28	33	50	50
	50	50	50	50	50	50	50	50
	50	50	50	50	50	50	50	50
	50	50	50	50	50	50	50	50
	50	50	50	50	50	50	50	50
	50	50	50	50	50	50	50	50
	50	50	50	50	50	50	50	50

When JP Hide & Seek was used to create stego images, the Quantization Matrix was the same as the cover image. In reference to Figure 5.1, this means that image 3 had the same QM as image 4 when using JP Hide & Seek to create Image 4. Because of this, we can conclude that JP Hide & Seek does not have it's own standard QM and analyzing images QM's will not give us any insight into if that image was created using JP Hide & Seek.

6.3.1 Table Quality Values Findings

The Table Quality value was also looked at for each of the generated images. The section below details the findings related to the Table Quality value.

The chart below shows the different Table Quality values that were found for Facebook images prior to generating a stego image (image 3 in the figure 5.1) and the frequency of those values. A table quality of value 71 was fairly common. More analysis would need to be done to determine if we could conclude that a JPEG image with a Table Quality of 71 came from Facebook. Since only 71.9% of images downloaded from Facebook had a table quality value of 71, we cannot assume that if an image downloaded from Facebook does not have that table quality value that it was tampered with.

Table 6.4: Table Quality Values for Facebook Images

Table Quality Value	**Frequency (in 57 images)**
71	41
72	1
73	1
74	1
77	3
79	3
81	1
82	1
89	1
91	1
92	3

All Twitter photos prior to generating a stego image (image 3 in the figure 5.1) had a Table Quality value of 85. Therefore, if an image downloaded from Twitter did not have a table quality value of 85 we could assume with a fair amount of certainty that that image may have been tampered with.

All stego photos generated using the SilentEye app (for both images 4 and 6 in the figure 5.1) had a Table Quality value of 75. Therefore, if an image downloaded from Facebook or Twitter had a Table Quality value of 75 we can assume that that image was edited using SilentEye.

6.4 Takeaways

Takeaways from the second case study are that we can use QM's as a way to gather information about an image and potentially determine if that image is a stego image. Essentially, if an image contains the standard SilentEye QM, then one could conclude that that image probably contains a stego message. If an image downloaded from Twitter does not contain the standard Twitter QM, then one could conclude that that image was probably tampered with in some way. Since Facebook did not seem to have a standard QM, then the confidence level is lower for being able to draw conclusions based on the QM of an image downloaded from Facebook.

CHAPTER 7. CONCLUSION

7.1 Recommended Method for creating stego images to share on social media

Previous studies on this topic found that when they created stego images and shared them to social media, that the compression the image went through in order to be shared on that site oftentimes stripped the message from the image. This first case study done for this paper aimed at finding a method in which stego images could be shared on social media without the message being lost. This study found that the method depicted in figure 7.1 of sharing social media data worked consistently.

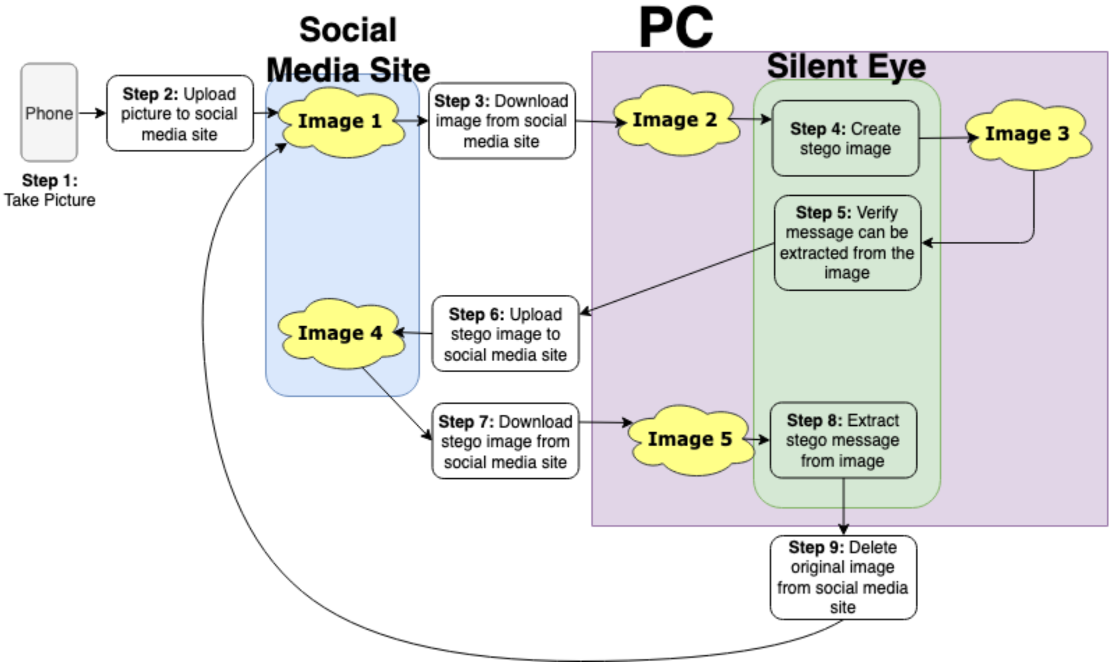

Social Media Site
PC
Silent Eye
Phone
Step 1: Take Picture
Step 2: Upload picture to social media site
Image 1
Step 3: Download image from social media site
Image 2
Step 4: Create stego image
Image 3
Step 5: Verify message can be extracted from the image
Step 6: Upload stego image to social media site
Image 4
Step 7: Download stego image from social media site
Image 5
Step 8: Extract stego message from image
Step 9: Delete original image from social media site

Figure 7.1: Recommended method for creating and sharing stego messages on social media

Although the study has some success using JP Hide & Seek for creating stego messages, that application had more failures than SilentEye. Also, JP Hide & Seek failed more often when the image was uploaded directly to the social media site from a phone rather than uploaded to the PC then to the site. The reasoning that the first method of uploading the image directly to the social media site from the phone is preferred is simply because it is one less step. Another benefit of using SilentEye over JP Hide & Seek is that SilentEye is compatible with Windows, Macs and Linux.

If you compare the method depicted in figure 7.1 to the method shown in figure 5.1, you will see that an additional step was added after creating the stego image and posting it to social media. This Step (step 5 in Figure 7.1) says to ”Verify message can be extracted from the image”. It was added because during the case study in Chapter 5 there were times that the stego message could not be retrieved from the final image shared on social media. Upon further investigation, it was determined that the image could not be retrieved from the stego image created even before posting it to the social media site. This is most likely due to bugs in the SilentEye tool and not in how the social media site compresses the image. The simplest way to avoid this failure is to verify that the stego image created does contain the message prior to positing it to the social media site. If the message cannot be extracted, try the method with a different image (i.e. start back at Step 1 with a new image).

Another difference between the method used in Figure 5.1 an the method used in Figure 7.1 is that a final step was added. This step recommends that once complete, the user should delete the first image from the social media site. This is recommended because if someone was monitoring your social media site and they saw two of the same images posted, they could compare the two images and if they were different be suspicious that the second image could contain a secret message.

7.2 Recommended Method for detecting stego images on social media

A second case study was completed to that aimed at finding a way in which to detect if an image shared on social media contained a secret message in it. The flow chart shown in Figure 7.2 depicts questions to ask in order to determine if an image posted on social media may contain a secret message in it.

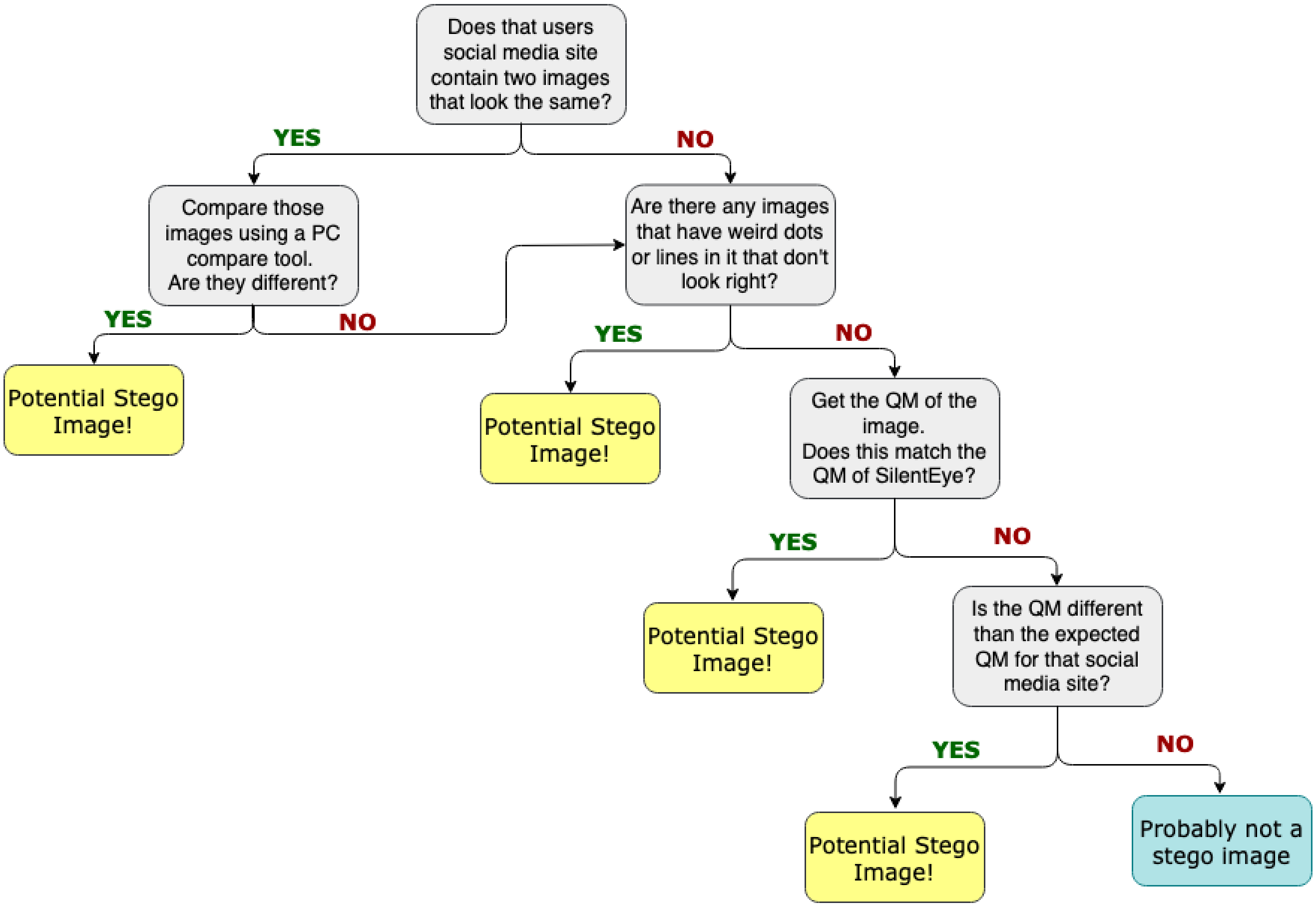

Does that users social media site contain two images that look the same?
YES
NO
Compare those images using a PC compare tool. Are they different?
Are there any images that have weird dots or lines in it that don't look right?
YES
NO
Potential Stego Image!
YES
NO
Potential Stego Image!
Get the QM of the image. Does this match the QM of SilentEye?
YES
NO
Potential Stego Image!
Is the QM different than the expected QM for that social media site?
YES
NO
Potential Stego Image!
Probably not a stego image

Figure 7.2: Recommended method for determining if an image on social media is a stego image

Through analyzing the Quantization Matrices (QMs) of the images created, a couple other important findings were discovered. The first finding is that all images downloaded from twitter have the same QM. This can be seen in Table 6.1 above. This is important because if an investigator was looking at images on twitter and downloaded said images, they could look at the QM of those images and if they did not match the QM seen in Table 6.1, then they could gather that that image had been tampered with in some way.

Additionally, all stego images created by SilentEye had the same QM even after it was posted to and then re-downloaded from a social media site. If an image retrieved from a social media site had this QM, then an investigator could conclude that that image most likely contains a stego message created by running it through SilentEye.

7.3 Conclusions and Future Research

This study and it's findings are important because image and video steganography can be used to share confidential or incriminating evidence or malware. Knowing how steganographic images can be created and shared on social media sites and how investigators can detect the presence of such messages is important to grasp the impact that this may have on our society. This can be used by criminal organizations, such as the Al Qaeda to share data with each other in plain site. Or it can affect your every day social media user who downloads a potential harmless looking image of a puppy which ends up infecting their computer.

Fortunately, the way in which social media compresses images when they are posted to the site limits the ways in which stego images can be shared on social media. Oftentimes this compression method strips the image of its hidden message. This study found that it is possible to share images on social media containing hidden messages. In order to do so, the user must first post an image without the message, download that image, then encrypt it with the hidden message, and then upload that message again. Another limitation of using Facebook in particular as a means of distributing secret messages is that these images must be uploaded as Facebook cover photos. This limits the size of information that can be shared. In one criminal case, noted in Section 3.3 above, the message was embedded inside a video instead of an image because videos can store much more data. If a user wished to share a long message embedded in an image, they would most likely have to break the message up over multiple images. If a user could share stego images as regular "Timeline photos" instead of cover photos, they could share a longer message by creating an entire album of timeline photos and sharing them all at once. But since cover photos must be used, the user would have to share the images one at a time which is more time consuming.

This study also found that one can detect the presence of a stego image if the SilentEye tool was used to create the image by retrieving that image's Quantization Matrix. All images created by SilentEye had the same QM even after that image was shared on a social media site. Because of this, investigators can analyze QM's of photos downloaded from Facebook or Twitter and if the QM of that image matches the one shown in 6.3, then they can conclude that that image was most likely created using the SilentEye app and contains hidden data in it.

In relation to the legal issues with social media, in order for investigators to use potential stego images as evidence in court, they must be able to prove it is relevant and authentic as well as retrieve the information legally. If a tool was developed that analyzed image's quantization matrices for certain values, this could be beneficial to investigators in the above circumstance. However, due to privacy laws and sheer number of images, it may not be practical or feasible to scan every single image posted on social media for a secret message or malware. Because a user's social media information is protected by the SCA, if such a tool was developed, the social media sites themselves would have to deploy and monitor the tool. This is due to the fact that investigators do not have the right to all of a person's social media information unless they have a relevant case in which to investigate that individual. Because stenography is not inherently bad, social media sites may be reluctant to deploy such a tool.

One recommendation for future work to complete on this topic would be to perform more case studies on different stego app/social media combinations. The purpose of those case studies, similar to the purpose of this study, would be to see if that application can be used to create stego images to share on social media. Additionally, that study could also gather information on if that social media site or stego application have a standard QM that could be used to determine the origin of said image.

Another future topic of study would be to determine why stego messages are sometimes lost when posted to social media. One could investigate the specific compression method used for different social media sites, if those methods are public knowledge. They could also investigate if images with different properties (such as a smaller size) do not get compressed when posted. This could potentially remove the necessity for posting images on social media twice. Instead of posting the original image to allow it to go to the compression process before embedding the message, users could change the properties of the image (such as make it a smaller size) before putting the stego message into it, knowing that the image will not be compressed.

From a more political/legal standpoint, it would be interesting to see if social media sites are currently scanning for stego images or if there is any actual data on how many images on social media sites may contain stego data. Also, if one were to implement a program that looked for stego images by analyzing QMs or looking for similar images on someones profile it would be interesting to see if social media sites would be interested in using such a tool or if investigators would use it in the case that they were granted legal access to someones social media account during an investigation.

Although there is more research to do on this topic, this study helped to open up the idea that it is possible to use social media as a means of sharing stego images and there are ways in which to detect stego images created by specific applications through analyzing QMs.

BIBLIOGRAPHY

[1] S. Sun, "A new information hiding method based on improved bpcs steganography," *Research Gate*, 2015. https://www.researchgate.net/publication/277594171_A_New_Information_Hiding_Method_Based_on_Improved_BPCS_Steganography.

[2] J. Newman, "Discrete cosine transform jpeg compression," 2019.

[3] "Steganography," *Merriam-Webster's Collegiate Dictionary.* https://www.merriam-webster.com/dictionary/steganography.

[4] D. Slincourt, Aubrey, and J. Marincola, *The histories.* Penguin Classics, 1996.

[5] "Invisible ink," *Revolutionary War Spy Quest.* https://sites.google.com/site/revolutionarywarspyquest/invisible-ink.

[6] "Cryptography," *Merriam-Webster's collegiate dictionary.* https://www.merriam-webster.com/dictionary/cryptography.

[7] K. Merrell, "Modular math and the shift cipher," *Khan Academy.* https://www.khanacademy.org/computing/compuer-science/cryptography/ciphers/a/shift-cipher.

[8] F. Ansuh, "Steganography: Not just a tool for the bad guys," *Global Information Assurance Certification Paper*, 2000 - 2002. https://www.giac.org/paper/gsec/1910/steganography-tool-bad-guys/103335.

[9] "Silent eye," 2010. https://silenteye.v1kings.io/.

[10] "Social media," *Merriam-Websters collegiate dictionary.* https://www.merriamwebster.com/dictionary/social%20media.

[11] "The history of social media: Social networking evolution!," *History Cooperative.* https://historycooperative.org/the-history-of-socialmedia/.

[12] "Global social media research summary 2018," *Smart Insights*, 21 November, 2018. https://www.smartinsights.com/social-mediamarketing/social-media-strategy/new-global-socialmedia-research/.

[13] K. Smith, "47 incredible facebook statistics and facts," *Brand Watch*, 5 March, 2018. https://www.brandwatch.com/blog/47-facebookstatistics/.

[14] K. Smith, "126 amazing social media statistics and facts," *Brandwatch*, 2019. https://www.brandwatch.com/blog/amazing-social-media-statistics-and-facts/.

[15] "Social media related crimes," *CBS News*. https://www.cbsnews.com/pictures/social-media-related-crimes.

[16] J. Staver, "Beaten by social media: Certainty and social media evidence," *Jurist*, 20 June, 2018. https://www.jurist.org/commentary/2018/06/jared-staver-personal-socialmedia/.

[17] H. Sherrod, "Ehling v. monmouth-ocean: Private facebook posts are protected," *Social Media Law Bulletin*, 2013. https://www.socialmedialawbulletin.com/2013/10/ehling-v-monmouth-ocean-private-facebook-posts-are-protected/.

[18] M. DiBianca, "Discovery and preservation of social media evidence," *American Bar*, 2014. https://www.americanbar.org/groups/business_law/publications/blt/2014/01/02_dibianca/.

[19] "Fbi: Russian spies hid codes in online photos," *NBC News*, 30 June 2010. http://www.nbcnews.com/id/38028696/ns/technology_and_science-science/t/fbi-russian-spies-hidcodesonline-photos/#.WuFA0Mgh03E.

[20] E. Niiler, "How al qaeda hid secrets in a porn video," *NBC News*, 11 July, 2012. http://www.nbcnews.com/id/47254281/ns/technology_and_science-science/t/how-al-qaeda-hidsecrets-porn-video/#.XA3bqydRfNY.

[21] I. Westbrook, "Hackers combine coded photos and twitter to hit targets," *BBC News*, 29 July 2015. http://www.bbc.com/news/technology-33702678.

[22] K. Debattista, "The threats of steganography," *Tech Talk*, 11 January, 2010. https://techtalk.gfi.com/threats-steganography/.

[23] B. Rossi, "How cyber criminals are using hidden messages in image files to infect your computer," *Information Age*, July 27, 2015. https://www.information-age.com/how-cybercriminals-are-using-hidden-messages-image-filesinfect-your-computer-123459881/.

[24] N. D. Amsden, L. Chen, and X. Yuan, "Transmitting hidden information using steganography via facebook," *IEEe - 33044*, 13 July, 2014. https://ieeexplore.ieee.org/document/6963080.

[25] N. D. Amsden and L. Chen, "Analysis of facebook steganography capabilities," *2015 International Conference on Computing, Networking and Communications, Communications and Information Security Symposium*, 2015. http://ieeexplore.ieee.org/document/7069317/.

[26] F. Heriniaina and X. Liao, "Pictographic steganography based on social networking websites," *ACSIJ Advances in Computer Science: an International Journal, Vol. 5, Issue 1, No.19*, Jan-

uary 2016. https://www.academia.edu/21385682/Pictographic_steganography_based_on_social_networking_website.

[27] J. Hiney, T. Dakve, K. Szczypiorski, and K. Gaj, “Using facebook for image steganography,” *Cornell University Library*, 5 June, 2015. https://arxiv.org/abs/1506.02071.

[28] T. Morkel, “The osn-tagging scheme: Recoverable steganography for online social networks,” *2017 1st International Conference on Next Generation Computing Applications (NextComp)*, 2017. http://ieeexplore.ieee.org/document/8016169/.

[29] J. Ning, I. Singh, H. V. Madhyastha, S. V. Krishnamurthy, G. Caox, and P. Mohapatraz, “Secret message sharing using online social media,” *2014 IEEE Conference on Communications and Network Security*, 2014. https://ieeexplore.ieee.org/document/6963080.

[30] B. Cusack and A. Chee, “Steganographic checks in digital forensic investigation: A social networking case,” *Edith Cowan University Research Online*, 2013. https://ro.ecu.edu.au/cgi/viewcontent.cgi?article=1118&context=adf.

[31] H. Farid, “Digital image ballistics from jpeg quantization,” 2008. https://www.cs.dartmouth.edu/~trdata/reports/TR2006-583.pdf.