

FILE STRUCTURE ANALYSIS OF MEDIA FILES SENT AND RECEIVED OVER
WHATSAPP

by

HENRY LUKEN RISEMBERG

B.F.A., University of Cincinnati, 2017

A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirements for the degree of
Master of Science
Recording Arts Program

2020

This thesis for the Master of Science degree by

Henry Luken Risemberg

has been approved for the

Recording Arts Program

by

Catalin Grigoras, Chair

Jeff M. Smith

Cole Whitecotton

Date: May 16, 2020

Risemberg, Henry Luken (M.S., Recording Arts Program)

File Structure Analysis of Media Files Sent and Received Over WhatsApp

Thesis directed by Associate Professor Catalin Grigoras

ABSTRACT

This research study explores the effects of sending and downloading image and audio files through the WhatsApp platform. A better understanding of how images and audio are affected by WhatsApp is necessary because of its popularity and the prevalence of digital images and audio as evidence in digital forensic investigations.

WhatsApp is a cross-platform communication service that allows the sending of media files and is one of the most popular services of its kind used worldwide. This application can be downloaded for use on mobile phones, iOS and Windows computers. There is also a website application available. The research done here will help investigators understand the process of different uploading and downloading techniques through different devices and the effects these methods have on file structure and metadata. An examination of recompression by WhatsApp, and how the WhatsApp software behaves when interacting with original images and audio will be conducted. The resulting media file's structure, metadata, binary data, quantization table data and other compression characteristics will be examined, and changes between files that are sent and the corresponding files that are downloaded will be analyzed.

The form and content of this abstract are approved. I recommend its publication.

Approved: Catalin Grigoras

DEDICATION

I would like to thank my parents for supporting me and guiding me to where I am today. Thank you for supporting me and motivating me to pursue an education. I would also like to especially dedicate this to my grandfather. I looked up to him more than anything, and how he lived his life further inspired me to work hard.

To those who have kept in contact with me throughout this process, Luke, Katie, Michael, Kevin, Henri, and my sisters, I am grateful.

To my co-workers and management at the Texas Department of Public Safety Crime Lab in Austin Texas who welcomed me with open arms.

And finally, a very special thank you to Bill.

ACKNOWLEDGEMENTS

I want to acknowledge everyone at the NCMF for their knowledge, support, and patience during my studies at the University of Colorado Denver. Especially Jeff and Catalin. If not for our introduction at AES conference in Los Angeles I never would have even known this was an opportunity for me. Thank you for giving me a chance in this field, and for making my time at University of Colorado Denver more than enjoyable.

Thank you, Cole, for always being there for me and for providing me with valuable input throughout the process of developing this thesis. And Leah for helping with every administrative aspect of this program.

I want to acknowledge my co-workers, Deegant, Erin, Jennifer, Marco, Melissa, Traig, and Caitlyn for their valuable input and help throughout this process.

And finally, to the NCMF 2017 – 2019 cohort. I am honored to have shared this time with such a bright group of people. We worked hard and had so much fun along the way.

TABLE OF CONTENTS

CHAPTER

I. INTRODUCTION	1
WhatsApp.....	2
Related Works	3
II. TECHNICAL OVERVIEW	4
JPEG Compression	4
Hashing and Stream Hashing	6
Metadata	6
Exif Data.....	7
Hex Data.....	9
Baseline JPEG vs. Progressive JPEG	11
Quantization Tables	11
Lossy Compression Analysis.....	12
III. MATERIALS.....	14
Device Specifications.....	14
Test Media Preparation	16
IV. METHODOLOGY.....	19
Sending Methods	19
Download Methods	20
Analysis Procedure	21
Analysis Tools	22
V. RESULTS.....	23

File Name	23
Hash Analysis	24
Exif Data.....	27
Hex Data (Image).....	30
Baseline JPEG vs. Progressive JPEG	35
Quantization Tables	36
Lossy Compression Analysis	39
Hex Data (Audio).....	41
VI. DISCUSSION.....	45
Conclusions	45
Further Research.....	47
REFERENCES	48

LIST OF TABLES

TABLE

1: PC Laptop Specifications	14
2: MacBook Pro Laptop Specifications.....	14
3: iPhone 6 Plus Specifications.....	15
4: iPhone 7 Specifications	15
5: Android G5 Specifications.....	15
6: Galaxy S9 Specifications.....	16
7: Web Application Specifications.....	16
8: Image Send Methods	19
9: Image Download Methods.....	20
10: Analysis Tools and Versions	22

LIST OF FIGURES

FIGURE

1: Discrete Cosine Transform.....	5
2: Exif Data.....	8
3: Hex Data.....	10
4: Baseline vs. Progressive Encoding.....	11
5: JPEG Quantization Tables.....	12
6: Lossy Compression Analysis.....	13
7: Table of Hash Values for iPhone Image 1.....	25
8: Table of Hash Values for iPhone Image 2.....	25
9: Table of Hash Values for Android Image 1.....	26
10: Table of Hash Values for Android Image 2.....	26
11: Exif Entries for Image Downloaded from WhatsApp.....	28
12: File Signature for iPhone Image 1.....	30
13: File Signature for Android Image 1.....	31
14: File Header of .jpg File Sent Over WhatsApp via iPhone.....	32
15: File Header of .jpg File Sent Over WhatsApp via Android.....	32
16: File Header of .jpg File Sent Over WhatsApp via Windows Application.....	33
17: File Header of .jpg File Sent Over WhatsApp via MacOS Application.....	34
18: File Header of .jpg File Sent Over WhatsApp via Web Application.....	35
19: Quantization Tables and Quality Factor of Original iPhone Image.....	36
20: Quantization Tables and Quality Factor of Image Sent Over WhatsApp via Windows Application.....	37

21: Quantization Tables and Quality Factor of Image Sent Over the iPhone WhatsApp Application.....	38
22: Quantization Tables and Quality Factor of Image Sent Over the Android WhatsApp Application	38
23: Lossy Compression Analysis Results of an Audio File Recorded With the “Compressed” Setting in the Apple Voice Memos Application Enabled and Sent Through WhatsApp	39
24: Lossy Compression Analysis Results of an Audio File Recorded With the “Lossless” Setting in the Apple Voice Memos Application Enabled and Sent Through WhatsApp	40
25: File Header of Original Lossless Recording on Left, and That Same Audio Recording Compressed by WhatsApp on Right	41
26: File Footer of Original Lossless Recording on Left, and That Same Audio Recording Re-Compressed by WhatsApp on Right.....	42
27: File Header of Original Recording Compressed by Apple Voice Memos on Left, and That Same Recording Compressed by WhatsApp on Right	43
28: File Footer of Original Recording Compressed by Apple Voice Memos on Left, and That Same Recording Compressed by WhatsApp on Right	44
29: Variable Data Table.....	46

CHAPTER I

INTRODUCTION

With the advent of computers, the internet, and cell phone technology, people in today's world can communicate with anyone anywhere in the world practically instantly at any time. As technology has evolved, social media platforms have emerged as an efficient vehicle for many people to share information and keep in touch with each other. Although instant text messaging is a common means of communication, one of the more effective ways to communicate a message is through image or audio messages. With current technology, these types of files can now be easily sent over a variety of social media platforms in virtually an instant.

Given their current popularity and the privacy that many popular social media platforms offer, social media services that offer instant messaging and file sharing are not only being used for legitimate purposes, but also for nefarious criminal activity. A significant number of cybercrimes involve the illicit distribution and possession of media files, and many other cases involve media files that have been sent and received over social media platforms in one way or another. For these types of cases, cell phones and computers are a rich source of evidence to investigators. When evidence media is discovered on these devices, it is important for investigators to understand how sending and receiving media files over social media platforms effects the data associated with those files.

For these social media platforms to facilitate the sending of media files containing image and audio information while providing users with a private and user-friendly environment, many will take original files and re-compress them so that they will work most efficiently within the platform environment. While this re-compression makes it easier for platforms to transmit, load, and store media files, data that could be important from an investigative standpoint is lost.

Depending on how re-compression takes place however, data can also be embedded into media files.

Central to the research conducted in this study is the idea that some of the embedded information, such as when and where a file was created for example, could possibly be useful to investigators who discover media files on a suspect device. The investigative research will hopefully be able to help digital forensic investigators recognize the changes that are made to media files sent and received over the WhatsApp platform. The file structure analysis of media files sent and received over WhatsApp can be added to the growing library of information collected in previous studies on how social media platforms interact with media files.

WhatsApp

WhatsApp is a free messaging service that facilitates the delivery of voice communications and multimedia sessions over internet protocol networks (VoIP). The application was created and launched in 2009 and in 2014 was acquired by Facebook for \$19 Billion [2]. WhatsApp is currently the 3rd most popular social media network, and the most popular communication application worldwide with 1,600,000,000 active users [4]. Within the application, users can interact with each other one-on-one or in group messages. The service supports messaging, media file sharing, voice and video calling all while utilizing the internet via cell phone data plans or Wi-Fi. Because of this feature, users can use the application to connect to other WhatsApp users for free. The app can operate on iPhone, Android and KaiOS mobile devices. It is also available as a desktop application for Microsoft Windows and MacOS, as well as through the web application WhatsApp Web.

One important feature that has contributed to the popularity of WhatsApp is the use of end-to-end encryption with every form of communication on the service including multimedia

messaging and calls. With this encryption in place, it is advertised that WhatsApp employees cannot even comply with court orders for access to information disseminated through WhatsApp. This feature is popular with users because it offers privacy. However, end to end encryption along with the large number of users creates the perfect environment for those with criminal or terroristic intentions to communicate and share files securely. WhatsApp was even used by terrorist organization ISIS to orchestrate the November 2017 Paris Attacks and the April 2017 Stockholm Attack [1] [3].

Related Works

Structure and image re-compression analysis of media file transfers over other social media and messaging platforms such as Twitter [5] and Instagram [6] have been done in recent years. There have also been more general analyses of multimedia file signatures for smart-phone forensics [7] and forensic analyses of WhatsApp messaging [8]. None of this previous research has focused on the file structure of multimedia files sent and received over WhatsApp. These papers provide a foundation for the research done in this study.

CHAPTER II

TECHNICAL OVERVIEW

JPEG Compression

The ability to capture digital images and process, store, transmit, and display them efficiently is something that we take for granted as part of our modern everyday lives. Digital image compression technology is what gives us this ability. Compression is necessary to reduce file size so that computers can handle images that would otherwise take up large amounts of storage space and processing power, ideally while maintaining the visual integrity of images. Lossy compression algorithms such as JPEG seek to discard information that is less easily noticed by the human eye and eliminate redundant information that takes up unnecessary space to achieve this. However, as the compression is applied more aggressively to save more and more space, quality is sacrificed, and compression artifacts start to become visible.

Of the different methods of image compression that are available today, JPEG is by far the most widely used and versatile. JPEG, which is an acronym for Joint Photographic Experts Group, was developed in the late 1980's and officially published in 1992 [9]. The JPEG standard was developed out of the necessity for a universal image compression specification in a time when images were starting to be shared over the internet and computers were not particularly good at processing images efficiently.

In JPEG compression, a digital image is divided into non-overlapping 8-by-8 blocks and the Discrete Cosine Transform (DCT) is computed for each block. This makes a set of 64 DCT Coefficients for each block. These coefficients are then divided by a quantization matrix and rounded off to the nearest integer. This is where data is lost. Many of these coefficients may now become zero and no longer need to be stored. Higher frequency information, which is less

perceptible to the human eye, is also discarded here. Then, a compression algorithm is run on the entire set of integers [14].

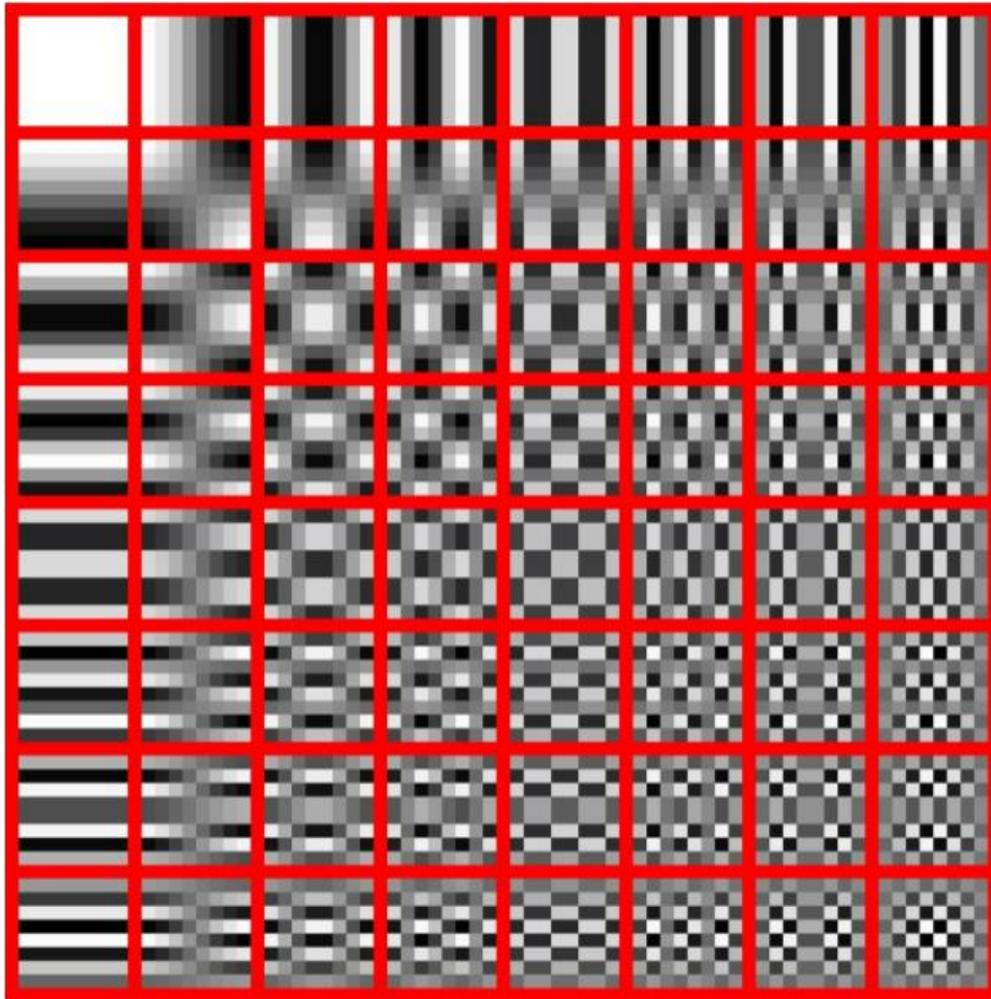


Figure 1: Discrete Cosine Transform

The 64 base cosine waves that can be combined to reproduce any image in 8x8 pixels for one channel

In this study, only JPEG images were analyzed. This is not only because JPEG images are so common, but also the fact that WhatsApp is primarily a mobile application, and as a default setting on most mobile devices images are captured as JPEG.

Hashing and Stream Hashing

Hash algorithms create a value of hexadecimal characters using complex mathematics based on a set of data. If any piece of that data changes, the resulting hash will also change. When two different sets of data produce the same hash value, this is called a collision. A collision can be engineered or accidental. While collisions have been engineered by computer scientist in the past, an accidental collision has never been recorded [10]. The Message Digest 5 (MD5) hash algorithm used in this study has been compromised as a result of an engineered collision before, but MD5 is still accepted as appropriate for digital signatures. Hash generation software such as the one used in this study generate hash values for given data sets very quickly and easily. The user selects a file to hash and the software calculates a hash value for that file. This is useful to identify files that are exact copies of each other as well as to identify files that may look or sound identical but have differences that are imperceptible to the human eye or ear, or have differences in metadata.

Stream hashing operates on the same concept as hashing. However, stream hashing only hashes the decoded data stream of an image file and excludes header and footer or metadata information. With this hashing method, it can be determined if the data streams of two different files are identical even if header and footer information may be different.

Metadata

Metadata, according to the Scientific Working Group on Digital Evidence, is “Data, frequently embedded within a file, that describes a file or directory, which can include the locations where the content is stored, dates and times, application specific information, and permissions” [18]. So, metadata is basically data that provides information about other data. In the case of this study, we are referring to metadata associated with image and audio files. This

data is some of the most important data to digital forensic investigators as it can include unique identifiable information about the who, what, when, where and how associated with a media file.

Exif Data

Exchangeable Image File Format, abbreviated as “Exif”, is a form of metadata and is the format in which data associated with image files captured with digital cameras is stored. The format was developed and is maintained by the Japan Electronic Industries Development Association and was first specified in 1998 [11]. Exif data associated with media files can include, but is not limited to, the GPS coordinates of where the media file was created, time and date information, identifying information of the recording device, camera settings and much more.

The volume of Exif data entries associated with a given file depends on the device used to capture the original image and can be affected by post processing an image file. Some files may have many Exif data entries while others may have very few. Simply observing how many Exif data entries there are may give clues to how the image file was captured and processed, and the content of Exif entries provides important information to digital forensic investigators.

```
ExifTool Version Number      : 11.88
File Name                    : IMG_0007.JPG
Directory                   : C:/Users/hankr/Desktop/Thesis/Test Files/Analysis Images/Originals/iPhone/Oranges
File Size                    : 3.1 MB
File Modification Date/Time  : 2020:02:10 19:59:42-06:00
File Access Date/Time       : 2020:02:20 18:52:54-06:00
File Creation Date/Time     : 2020:02:20 18:52:54-06:00
File Permissions            : rw-rw-rw-
File Type                   : JPEG
File Type Extension         : jpg
MIME Type                   : image/jpeg
Exif Byte Order             : Big-endian (Motorola, MM)
Make                        : Apple
Camera Model Name           : iPhone7
Orientation                 : Rotate 90 CW
X Resolution                : 72
Y Resolution                : 72
Resolution Unit             : inches
Software                    : 13.3.1
Modify Date                 : 2020:02:10 18:19:43
Y Cb Cr Positioning        : Centered
Exposure Time               : 1/15
F Number                    : 1.8
Exposure Program           : Program AE
ISO                         : 50
Exif Version                : 0231
Date/Time Original         : 2020:02:10 18:19:43
Create Date                 : 2020:02:10 18:19:43
Offset Time                 : -06:00
Offset Time Original       : -06:00
Offset Time Digitized      : -06:00
Components Configuration   : Y, Cb, Cr, -
Shutter Speed Value        : 1/15
Aperture Value              : 1.8
Brightness Value           : 2.157272519
Exposure Compensation      : 0
Metering Mode               : Multi-segment
Flash                       : Auto, Did not fire
Focal Length                : 4.0 mm
Subject Area                : 2015 1511 2217 1330
Run Time Flags              : Valid
Run Time Value              : 12246114995833
Run Time Scale              : 1000000000
Run Time Epoch              : 0
Acceleration Vector         : 0.02315517515 -0.9819984435 -0.1617534608
Sub Sec Time Original      : 012
Sub Sec Time Digitized     : 012
Flashpix Version            : 0100
```

Figure 2: Exif Data

This figure is an example of a printout of Exif data that was obtained running the program

ExifTool on a JPEG image file

Hex Data

At the most basic level, all digital information is stored, processed and represented in binary form. That is, the binary numerical system of ones and zeros. Each one or zero represents a binary digit, the smallest unit of data in a computer. Computers can take these bits and process them as instructions to perform tasks that we want the computer to perform. These ones and zeros are incomprehensible to most people, but digital forensic investigators must make sense of this data to represent what it means.

To solve this problem, we have a slightly easier way of representing binary data in the form of hexadecimal data. Hexadecimal uses a base 16 number system to represent data, using the numbers 1 through 9 to represent values zero to nine, and letters A through F to represent values ten to fifteen. Hex reader software can act as a translator to represent the data in Hexadecimal. This way data is more easily understood. Hex readers can further translate some data into meaningful ASCII information, which is an abbreviation for American Standard Code for Information Interchange [12]. ASCII information sometimes contains meaningful words and phrases that give important information about a file.

Baseline JPEG vs. Progressive JPEG

JPEG images can come in two different forms, Baseline or Progressive. A Baseline JPEG image uses an algorithm that starts to display image data as it becomes available, line by line from top to bottom. If you were to see a Baseline JPEG image load slowly, you could see the image showing up on your screen in this manner. Progressive JPEG on the other hand loads images in a different way. This type of image is displayed first as a blurry version of the image in its entirety, becoming clearer and clearer as more image data is made available [13].

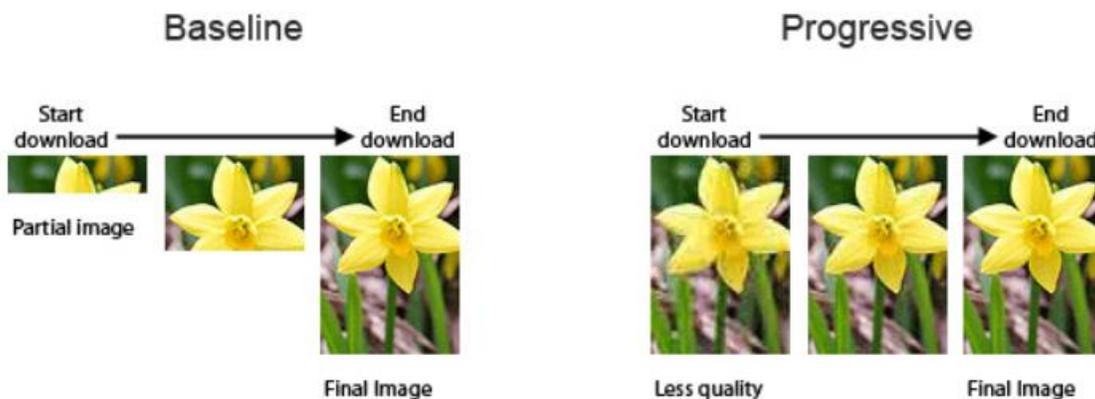


Figure 4: Baseline vs. Progressive Encoding

Quantization Tables

Quantization is part of the JPEG compression process. The JPEG compression algorithm uses one or more quantization tables that dictate the degree to which images will be compressed. This in turn determines the images overall “Quality Factor”. The original image is processed into coefficients, and these coefficients are rounded to the nearest integer. This is the part of jpeg compression where image data is lost and file size is reduced. The resulting integers are divided by the value in the quantization table that corresponds to that integer. If a higher level of compression is used the quantization table has higher values in it. In this case a lower quality image file with a smaller file size is created. If a lower level of compression is used the

quantization table uses lower values. In this case a higher quality image file results but the file size is not reduced as much [15].

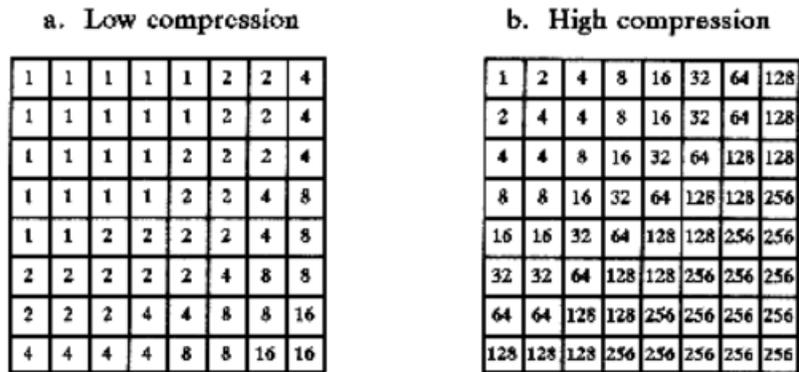


Figure 5: JPEG Quantization Tables

This figure portrays two examples of quantization tables that may be used in JPEG compression.

The table on the left is indicative of a lower level of compression, and the table on the right a higher level of compression.

Lossy Compression Analysis

Lossy compression analysis is a method that can be used to identify what type of software or device was used to record an audio recording by assessing the traces of lossy compression in the signal [17]. In this method, reference recordings from known phones or devices are used to configure a database of compression models for each phone or device. The compression model that is created consists of AAC decoded MDCT coefficients, the Long-Term Average Sorted Spectrum, and the Audio compression level of known audio files. Unknown audio files can be compared against the database automatically to determine if the unknown files compression characteristics match that of any of the models in the database.

Appendix 5. FDK-AAC (upper) and Nero (lower) decoded original Voice Memos iOS 11.0.2 .M4A/AAC recordings

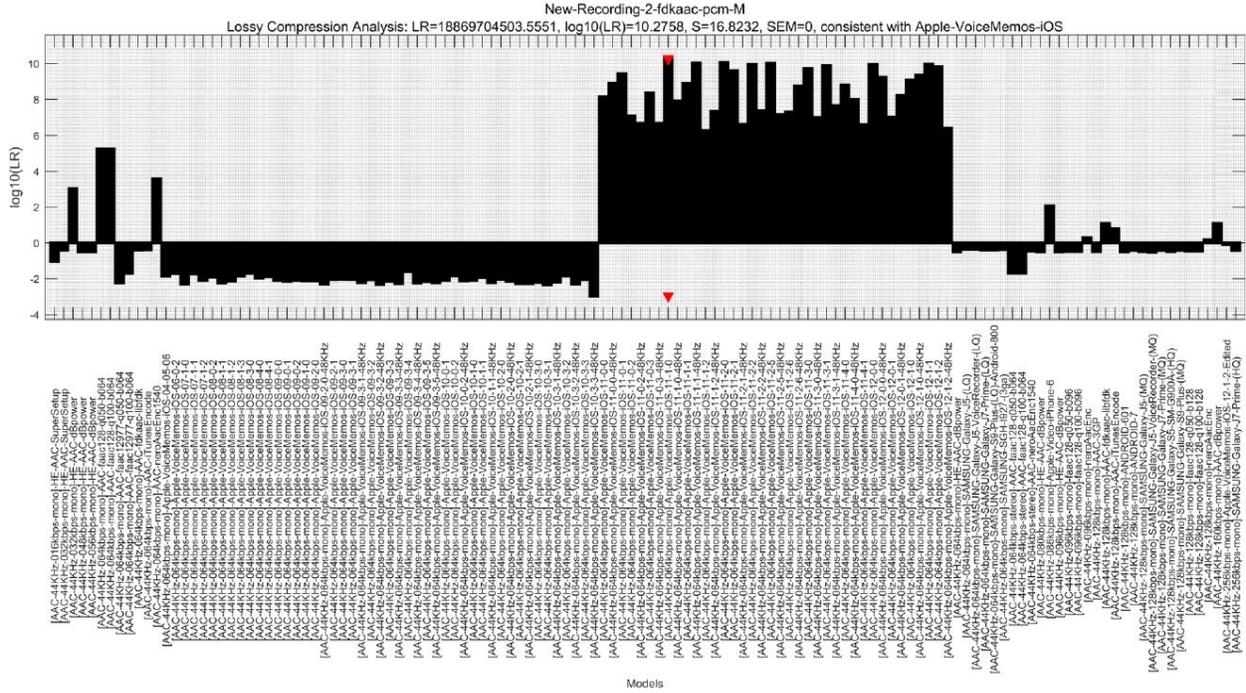


Figure 6: Lossy Compression Analysis

This figure shows the result of lossy compression level analysis against a curated database. The red arrow indicates which compression model the inputted file is most consistent with.

CHAPTER III

MATERIALS

Device Specifications

The research in this study was conducted with six different devices including two laptop computers and four mobile devices. It was determined that two iPhone devices and two Android devices should be used. This was because logging in and out of multiple WhatsApp accounts multiple times on mobile devices caused WhatsApp to lock out the user for increasingly long periods of time. When it came time to send images from Android to Android or iPhone to iPhone another device was employed to avoid this issue. The same issue was not experienced with desktop or web applications. The specifications of the devices used as well as the Google Chrome web application are listed below.

Table 1: PC Laptop Specifications

PC Laptop	
Manufacturer	MSI
Model	GF65 Thin
Processor	Intel Core i5-9300H CPU @ 2.4 GHz
RAM	8.00 GB
Operating System	Windows 10 Build 18362.719
WhatsApp Version	0.4.315

Table 2: MacBook Pro Laptop Specifications

MacBook Pro	
Manufacturer	Apple Inc.
Model	Mid 2015, 15-inch, Retina Display
Processor	2.8 GHz Intel Core i7
RAM	16.00 GB
Operating System	MacOS Mojave Version 10.14.4
WhatsApp Version	0.4.315

Table 3: iPhone 6 Plus Specifications

iPhone 6 Plus	
Manufacturer	Apple Inc.
Model Number	NGAU2LL/A
Operating System	iOS 12.4.5
Serial Number	F9CS706UG5QJ
IMEI Number	35 931906 197647 9
WhatsApp Version	2.20.21

Table 4: iPhone 7 Specifications

iPhone 7	
Manufacturer	Apple Inc.
Model Number	MN9D2LL/A
Operating System	iOS 13.3.1
Serial Number	F17VH2K1HG7F
IMEI Number	35 676008 911620 9
WhatsApp Version	2.20.21

Table 5: Android G5 Specifications

Android G5	
Manufacturer	LG
Model Number	LGLS992
Software Version	LS992ZVF
Build Number	NRD90U
IMEI Number	355602072466687
WhatsApp Version	2.20.21

Table 6: Galaxy S9 Specifications

Samsung Galaxy S9 Plus	
Manufacturer	Samsung
Model Number	SM-G965U
Software Version	G965USQS7DTB1
IMEI Number	356420092963925
WhatsApp Version	2.20.108
Android Version	10

Table 7: Web Application Specifications

In instances where the Web Application was used, it was accessed via the MSI Laptop.

Web Application	
Browser	Google Chrome
Browser Version	79.0.3945.130 (official build) (64-bit)
URL	web.whatsapp.com

Test Media Preparation

It was determined that images captured on mobile phones would be the most relevant type of images to use in this study because the WhatsApp application is primarily used on mobile devices [4]. The goal of the test media preparation in this study was to ensure that the images and audio used for testing were identical before sending regardless of the device they originated from.

A total of four images were used for testing in this study. Two images were captured on the iPhone 7, and two more images were captured on the Android G5. For each device, one image was captured in portrait orientation (with the device held upright) and one image was captured in landscape orientation (with the device held 90 degrees from upright).

- The image captured on the iPhone 7 in portrait orientation will henceforth be referred to as “iPhone Image 1”.
- The image captured on the iPhone 7 in landscape orientation will henceforth be referred to as “iPhone Image 2”
- The image captured on the Android G5 in portrait orientation will henceforth be referred to as “Android Image 1”
- The image captured on the Android G5 in landscape orientation will henceforth be referred to as “Android Image 2”

After the images were captured with each device, the original images were extracted from the iPhone and the Android devices directly to the MSI laptop computer. The hashes of these images were then calculated. Then, the images were loaded onto every other device that the images did not originate from via a data cable transfer. After the images were populated to every device, a hash verification was conducted against every corresponding image on every device to ensure that images were not altered by the process of moving them from device to device in preparation for the study. This verification was successful. At this point, there were 4 images on the MSI Laptop, MacBook Pro, iPhone 6 Plus, iPhone 7, Android G5 and Galaxy S9.

A total of twenty audio recordings were also used for this study. The twenty audio recordings with an average approximate length of one minute were recorded on the iPhone 7 using the Apple Voice Memos application. These audio recordings were only sent from the iPhone 7. Sending the audio recordings from other devices other than the iPhone 7 was determined to be beyond the scope of this study. Ten of these recordings were recorded with the “Lossless” setting enabled in the iPhone Voice Memos settings menu, and ten more recordings were recorded with the “Compressed” setting enabled.

A WhatsApp account with the name “Hank Thesis” was create solely for the purpose of this study to provide a sterile environment for testing. Since WhatsApp requires a phone number to be associated with any account, a Google Voice account was created to associate the WhatsApp account with.

CHAPTER IV
METHODOLOGY

Sending Methods

After the four original images were populated to the six devices to be used in this study, each image was sent using the ten different sending methods listed below. This list of sending methods includes every easily accessible method to send images from within the WhatsApp user interface in a chat between two users. If more than three images are sent without a text message in between, the images are grouped into a collection of images that needs to be expanded to be viewed and downloaded. To avoid the variables that this would introduce, one text message was sent between the sending of each image.

Table 7: Image Send Methods

Send Methods		
Device	Reference in Hash Tables	Method
iPhone	Send Method 1	‘+’ Icon
	Send Method 2	Camera Icon
Android	Send Method 3	Attachment Icon
	Send Method 4	Camera Icon
PC Application via MSI Laptop	Send Method 5	Drag and Drop
	Send Method 6	Attachment Icon
MacOS Application via MacBook Pro	Send Method 7	Drag and Drop
	Send Method 8	Attachment Icon
Google Chrome Application	Send Method 9	Drag and Drop
	Send Method 10	Attachment Icon

The audio files in this study were sent using the “share” function within the Apple Voice Memos application of the iPhone 7. It is possible to select audio files from within the WhatsApp application for sending, however exploring the effects of using different send methods for audio was determined to be beyond the scope of this study.

Download Methods

After the images were sent through WhatsApp, they were downloaded using the following download methods from the four different devices and web application. The list of download methods includes every easily accessible method to download images within the WhatsApp user interface in a chat between two users.

Table 8: Image Download Methods

Download Methods		
Device	Reference in Hash Tables	Method
iPhone	DL M 1	Automatic Download
	DL M 2	Automatic Download Disabled
Android	DL M 3	Automatic Download
	DL M 4	Automatic Download Disabled
PC Application via MSI Laptop	DL M 5	Menu Icon
	DL M 6	Download Arrow
	DL M 7	Expanded Download Arrow
MacOS Application via MacBook Pro	DL M 8	Menu Icon
	DL M 9	Download Arrow
	DL M 10	Expanded Download Arrow
Google Chrome Application	DL M 11	Download Arrow
	DL M 12	Right Click Download
	DL M 13	Expanded Download Arrow
	DL M 14	Expanded Rick Click Download

There is currently no application programming interface for WhatsApp, so images that could have otherwise been sent and downloaded automatically were sent and downloaded manually. After the images were downloaded via all the downloaded methods in table 8, they were extracted from each device to the MSI laptop for analysis.

Audio files were downloaded from the WhatsApp PC application to a device using the Windows 10 operating system for analysis.

Analysis Procedure

The sending and downloading of four images utilizing all ten send methods and all fourteen download methods resulted in the creation of a set of 560 test images to be analyzed. After the images were extracted from all devices and consolidated onto the MSI laptop, the first step in the analysis was to calculate hash values for each individual image.

Calculating the hash values of all resultant images was done to determine what sending and downloading methods produced bit stream duplicate files and if any combination of methods did not change the files at all. Patterns of combinations of sending and downloading methods that resulted in duplicate files could be observed and documented. After this step analysis could be limited to files that were not duplicates of each other. Stream hashes were calculated as a second level of further assessing files that did not have matching hash values. This was done to determine if there were images that may have identical image data and only have differences in metadata.

After unique images were identified for analysis, file characteristics and structure were observed. Analysis of Metadata, Hex data, Exif data, Quantization Tables, and how the images were encoded was conducted and conclusions drawn from this information.

Analysis of the audio files was conducted separately from the image files in this study. Audio files were downloaded from WhatsApp to a desktop computer and a Lossy Compression Analysis was conducted against the files.

Analysis Tools

The following table lists the tools and their respective versions that were used for analysis in this study.

Table 9: Analysis Tools and Versions

Analysis Tools	
Tool	Version
ExactFile	1.0.0.15
MediaInfo	19.09
ffmpeg	4.2.1
JPEG Snoop	1.7.3
ExifTool	11.88
010 Editor	10.0.1

CHAPTER V

RESULTS

File Name

The first file characteristic to be observed in this study was file name. All downloaded images had unique file names. Some useful information was able to be observed in the names of the downloaded image files.

The original images all had a file extension of “.jpg”. After images were sent through WhatsApp and downloaded however, the file extension changed to “.jpeg” in all instances.

The naming convention applied to image files downloaded from WhatsApp is dictated by what type of device or web application was used to download the images. A unique naming convention is applied if the image is received on an iPhone or an Android device. If the image is downloaded through the MacOS application, Windows application, or web application (except for the right click/download method within the web application), the naming convention is the same. The naming conventions with examples are described as follows.

- Download to iPhone device
 - Four uppercase letters followed by four numbers
 - Ex.) “QOJY4018”
- Download to Android device
 - “IMG” followed by “-“ followed by date arranged in “yyyymmdd”, followed by “-“, followed by “WA”, followed by a four digit number that seems to indicate the order in which the images were downloaded to the phones file system
 - Ex.) “IMG-20200215-WA0049”

- Download via Mac Application, Windows Application, or Web Application
 - “WhatsApp Image” followed by date in the format of “yyyy-mm-dd” followed by “at”, followed by time that the message was *sent* “h.mm.ss”, followed by “AM” or “PM”
 - Ex.) “WhatsApp Image 2020-02-15 at 4.43.32 PM”
- Download using the right click / download feature within the WhatsApp Web Application
 - 8 numbers and lower-case letters, followed by 4 numbers and lowercase letters, followed by 4 more numbers and lowercase letters, followed by 12 numbers and lowercase letters. No patterns were determined in this naming convention
 - Ex.) “2d7a76ec-310b-400c-abad-7b7555699584”

Hash Analysis

Hash values were calculated for all downloaded images. Every hash value was entered into a table with sending methods on the y-axis and download methods on the x-axis. A different table was created for each source image. This helped facilitate visually observing patterns of likeness between downloaded images dependent on the method of sending and downloading. In the tables to follow, only a portion of the hash values are shown. Reference Table 7 and 8 for what send methods and download methods are used in each row and column.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1		DL M 1	DL M 2	DL M 3	DL M 4	DL M 5	DL M 6	DL M 7	DL M 8	DL M 9	DL M 10	DL M 11	DL M 12	DL M 13	DL M 14
2	Send Method 1	5e4e04k													
3	Send Method 2	5e4e04k													
4	Send Method 3	630b1ac	630b1ac	630b1ac	630b1ac	0360165	0360165	0360165	0360165	0360165	0360165	0360165	0360165	0360165	0360165
5	Send Method 4	630b1ac	630b1ac	630b1ac	630b1ac	0360165	0360165	0360165	0360165	0360165	0360165	0360165	0360165	0360165	0360165
6	Send Method 5	e267b34													
7	Send Method 6	e267b34													
8	Send Method 7	f5671d5													
9	Send Method 8	f5671d5													
10	Send Method 9	22f31e0													
11	Send Method 10	22f31e0													

Figure 7: Table of Hash Values for iPhone Image 1

After sending and receiving “iPhone Image 1” by means of every combination of send and download methods, five different images were produced. One different image was created for each device or web application that was used to send the image.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1		DL M 1	DL M 2	DL M 3	DL M 4	DL M 5	DL M 6	DL M 7	DL M 8	DL M 9	DL M 10	DL M 11	DL M 12	DL M 13	DL M 14
2	Send Method 1	1567a27													
3	Send Method 2	1567a27													
4	Send Method 3	ebbfa3c	ebbfa3c	ebbfa3c	ebbfa3c	52263fc									
5	Send Method 4	ebbfa3c	ebbfa3c	ebbfa3c	ebbfa3c	52263fc									
6	Send Method 5	4967444	4967444	4967444	4967444	4967444	4967444	4967444	4967444	4967444	4967444	4967444	4967444	4967444	4967444
7	Send Method 6	4967444	4967444	4967444	4967444	4967444	4967444	4967444	4967444	4967444	4967444	4967444	4967444	4967444	4967444
8	Send Method 7	9558354	9558354	9558354	9558354	9558354	9558354	9558354	9558354	9558354	9558354	9558354	9558354	9558354	9558354
9	Send Method 8	9558354	9558354	9558354	9558354	9558354	9558354	9558354	9558354	9558354	9558354	9558354	9558354	9558354	9558354
10	Send Method 9	9558354	9558354	9558354	9558354	9558354	9558354	9558354	9558354	9558354	9558354	9558354	9558354	9558354	9558354
11	Send Method 10	9558354	9558354	9558354	9558354	9558354	9558354	9558354	9558354	9558354	9558354	9558354	9558354	9558354	9558354

Figure 8: Table of Hash Values for iPhone Image 2

After sending and receiving “iPhone Image 2” by means of every combination of send and download methods, four different images were produced. One different image was created for each device or web application that was used to send the image. However, images sent from the WhatsApp application running on Windows 10 matched images that were sent from the WhatsApp application on the MacOS.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1		DL M 1	DL M 2	DL M 3	DL M 4	DL M 5	DL M 6	DL M 7	DL M 8	DL M 9	DL M 10	DL M 11	DL M 12	DL M 13	DL M 14
2	Send Method 1	ead7f31	ead7f31af5a67c7	ead7f31	ead7f31										
3	Send Method 2	ead7f31	ead7f31	ead7f31											
4	Send Method 3	b9b640f	b9b640f	b9b640f											
5	Send Method 4	b9b640f	b9b640f	b9b640f											
6	Send Method 5	a39aafd	a39aafd	a39aafd											
7	Send Method 6	a39aafd	a39aafd	a39aafd											
8	Send Method 7	4ce9ad0	4ce9ad0	4ce9ad0											
9	Send Method 8	4ce9ad0	4ce9ad0	4ce9ad0											
10	Send Method 9	4ce9ad0	4ce9ad0	4ce9ad0											
11	Send Method 10	4ce9ad0	4ce9ad0	4ce9ad0											

Figure 9: Table of Hash Values for Android Image 1

After sending and receiving “Android Image 1” by means of every combination of send and download methods, four different images were produced. This time, the same results were observed as with “iPhone Image 2”. One different image was created for each device or web application that was used to send the image, except images sent from the WhatsApp application running on Windows 10 matched images that were sent from the WhatsApp application on the MacOS.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1		DL M 1	DL M 2	DL M 3	DL M 4	DL M 5	DL M 6	DL M 7	DL M 8	DL M 9	DL M 10	DL M 11	DL M 12	DL M 13	DL M 14
2	Send Method 1	fe518b1													
3	Send Method 2	fe518b1													
4	Send Method 3	0632e42													
5	Send Method 4	0632e42													
6	Send Method 5	b25a2b9													
7	Send Method 6	b25a2b9													
8	Send Method 7	b25a2b9													
9	Send Method 8	b25a2b9													
10	Send Method 9	1f234d6													
11	Send Method 10	1f234d6													

Figure 10: Table of Hash Values for Android Image 2

After sending and receiving “Android Image 2” by means of every combination of send and download methods, four different images were produced. One different image was created for each device or web application that was used to send the image, except this time images sent from the WhatsApp application running on Windows 10 matched images that were sent from the WhatsApp Web application.

After every send and download method was used to process the four different original images, a total of seventeen different downloaded images were produced. Different images were produced dependent on the send method, and different send methods within an application on one device or web application had no effect on the downloaded images. Different download methods had no effect on the output images outside of how the files were named. In some cases, but not all, there is uniformity between images that are sent from different desktop devices or web applications.

Five different versions of “iPhone Image 1” were produced, and four different images were produced for the other three images, “iPhone Image 2”, “Android Image 1”, and “Android Image 2”. From this point forward, analysis was conducted only on one each of those seventeen different images.

Stream hash values were calculated for each of the resulting seventeen different images and there were no matches. This shows that there are differences in the core image data of all seventeen images.

Exif Data

The four original images and seventeen images downloaded from WhatsApp were inputted into ExifTool. The volume of Exif entries observed for images was significantly reduced after those images were sent and then downloaded through WhatsApp. There were a total of 123 Exif entries for both images that were captured on the iPhone, and 79 Exif entries for both images that were captured on Android. After images were sent and then downloaded through WhatsApp however, only 22 Exif entries were observed.

```

C:\Users\hankr\Desktop\Thesis\Test Files\Analysis Images\Originals\iPhone\Oranges>exiftool "C:\Users\hankr\Desktop\Thesis\Test Files\Analysis Images\Oranges\Sent via iPhone\WhatsApp Image 2020-02-17 at 2.47.48 PM.jpeg"
ExifTool Version Number      : 11.88
File Name                    : WhatsApp Image 2020-02-17 at 2.47.48 PM.jpeg
Directory                   : C:/Users/hankr/Desktop/Thesis/Test Files/Analysis Images/Oranges/Sent via iPhone
File Size                    : 209 kB
File Modification Date/Time  : 2020:02:17 14:53:16-06:00
File Access Date/Time       : 2020:02:20 18:34:03-06:00
File Creation Date/Time     : 2020:02:20 18:34:03-06:00
File Permissions             : rw-rw-rw-
File Type                    : JPEG
File Type Extension         : jpg
MIME Type                    : image/jpeg
JFIF Version                 : 1.01
Resolution Unit              : None
X Resolution                 : 1
Y Resolution                 : 1
Image Width                  : 1200
Image Height                 : 1600
Encoding Process             : Progressive DCT, Huffman coding
Bits Per Sample              : 8
Color Components             : 3
Y Cb Cr Sub Sampling        : YCbCr4:2:0 (2 2)
Image Size                   : 1200x1600
Megapixels                   : 1.9

```

Figure 11: Exif Entries for Image Downloaded From WhatsApp

Of the twenty-two Exif entries, the following entries were the same for every image.

- File Permission: rw-rw-rw
- File Type: JPEG
- File Type Extension: jpg
- MIME Type: image/jpeg
- JFIF Version: 1.01
- Resolution Unit: 1.01
- X Resolution: 1
- Y Resolution: 1
- Bits Per Sample: 8
- Color Components: 3
- Y Cb Cr Sub Sampling: YCbCr4:2:0 (2 2)

Of the twenty-two Exif entries, the following entries varied between images.

- File Name
- Directory
- File Size
- File Modified Date/Time
- File Access Date/Time
- File Creation Date/Time
- Image Width
- Image Height
- Encoding Process
- Image Size
- Megapixels

One variable that experienced some interesting and consistent changes was image size and dimensions. Images that were originally captured on an iPhone had their width and height dimensions reduced by a factor of 2.52 by the WhatsApp compression. Images that were originally captured on an Android device had their width and height dimensions reduced by a factor of 3.32 by the WhatsApp compression. Images that were captured on an Android device *and* were sent by an Android device had their width and height dimensions reduced by a factor of exactly 4 by the WhatsApp compression.

The file size of images that originated from the same source image and were sent via desktop applications were all very similar. In cases where images originating from the same source image that were sent over different desktop applications did not have matching hash values, there was only a relatively very small difference (less than a kilobyte) in their file size.

Hex Data (Image)

Hex data analysis and observations were conducted using 010 Editor. The color coding is applied by 010 Editor to highlight logical segments of data. Original images displayed the JPG EXIF file signature FF D8 FF E1 ?? ?? 45 78 69 66 00 00. Shown below is the beginning of the file header for one original iPhone image and one original Android image.

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0000h:	FF	D8	FF	E1	20	06	45	78	69	66	00	00	4D	4D	00	2A	ÿ	ÿ	á	.	Exif	.	MM	*								
0010h:	00	00	00	08	00	0B	01	0F	00	02	00	00	00	06	00	00															
0020h:	00	92	01	10	00	02	00	00	00	09	00	00	00	98	01	12	.'.....~															
0030h:	00	03	00	00	00	01	00	06	00	00	01	1A	00	05	00	00															
0040h:	00	01	00	00	00	A2	01	1B	00	05	00	00	00	01	00	00c.....															
0050h:	00	AA	01	28	00	03	00	00	00	01	00	02	00	00	01	31	..a.(.....1															
0060h:	00	02	00	00	00	07	00	00	00	B2	01	32	00	02	00	00².2....															
0070h:	00	14	00	00	00	BA	02	13	00	03	00	00	00	01	00	01°.....															
0080h:	00	00	87	69	00	04	00	00	00	01	00	00	00	CE	88	25	..+i.....î^%															
0090h:	00	04	00	00	00	01	00	00	07	AA	00	00	08	A4	41	70ª...Ap															
00A0h:	70	6C	65	00	69	50	68	6F	6E	65	20	37	00	00	00	00	ple.iPhone 7....															
00B0h:	00	48	00	00	00	01	00	00	00	48	00	00	00	01	31	33	..H.....H....13															
00C0h:	2E	33	2E	31	00	00	32	30	32	30	3A	30	32	3A	31	30	..3.1..2020:02:10															
00D0h:	20	31	38	3A	31	39	3A	34	33	00	00	23	82	9A	00	05	..18:19:43..#,.š..															
00E0h:	00	00	00	01	00	00	02	78	82	9D	00	05	00	00	00	01x,.....															
00F0h:	00	00	02	80	88	22	00	03	00	00	00	01	00	02	00	00	...e^".....															
0100h:	88	27	00	03	00	00	00	01	00	32	00	00	90	00	00	07	^'.....2.....															
0110h:	00	00	00	04	30	32	33	31	90	03	00	02	00	00	00	140231.....															
0120h:	00	00	02	88	90	04	00	02	00	00	00	14	00	00	02	9C	...^.....ce															
0130h:	90	10	00	02	00	00	00	07	00	00	02	B0	90	11	00	02°.....															
0140h:	00	00	00	07	00	00	02	B8	90	12	00	02	00	00	00	07															
0150h:	00	00	02	C0	91	01	00	07	00	00	00	04	01	02	03	00	...À'.....															
0160h:	92	01	00	0A	00	00	00	01	00	00	02	C8	92	02	00	05	'.....È'...															
0170h:	00	00	00	01	00	00	02	D0	92	03	00	0A	00	00	00	01Đ'.....															
0180h:	00	00	02	D8	92	04	00	0A	00	00	00	01	00	00	02	E0	...ø'.....à															
0190h:	92	07	00	03	00	00	00	01	00	05	00	00	92	09	00	03	'.....'															

Figure 12: File Signature for iPhone Image 1

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0000h:	FF	D8	FF	E1	AF	B2	45	78	69	66	00	00	4D	4D	00	2A	ÿøÿá²Exif..MM.*															
0010h:	00	00	00	08	00	0A	01	32	00	02	00	00	00	14	00	002.....															
0020h:	00	86	88	25	00	04	00	00	00	01	00	00	1A	C2	01	10	.+^%.....Â..															
0030h:	00	02	00	00	00	08	00	00	00	9A	02	13	00	03	00	00š.....															
0040h:	00	01	00	01	00	00	01	28	00	03	00	00	00	01	00	02(.....															
0050h:	00	00	01	1B	00	05	00	00	00	01	00	00	00	A2	01	12ç.....															
0060h:	00	03	00	00	00	01	00	01	00	00	87	69	00	04	00	00+i.....															
0070h:	00	01	00	00	00	C1	01	1A	00	05	00	00	00	01	00	00Á.....															
0080h:	00	AA	01	0F	00	02	00	00	00	0F	00	00	00	B2	00	00	.ª.....².....															
0090h:	1B	83	32	30	32	30	3A	30	32	3A	30	39	20	31	32	3A	.f2020:02:09 12:															
00A0h:	34	32	3A	30	39	00	4C	47	4C	53	39	39	32	00	00	00	42:09.LGLS992...															
00B0h:	00	48	00	00	00	01	00	00	00	48	00	00	00	01	4C	47	.H.....H....LG															
00C0h:	20	45	6C	65	63	74	72	6F	6E	69	63	73	00	00	1E	A0	Electronics...															
00D0h:	01	00	03	00	00	00	01	00	01	00	00	82	9D	00	05	00,.....															
00E0h:	00	00	01	00	00	02	2F	90	04	00	02	00	00	00	14	00/.....															
00F0h:	00	02	37	92	0A	00	05	00	00	00	01	00	00	02	4B	A4	..7'.....Kα															
0100h:	02	00	03	00	00	00	01	00	00	00	00	92	92	00	02	00'.....															
0110h:	00	00	07	00	00	02	53	A0	03	00	04	00	00	00	01	00S.....															
0120h:	00	0B	AC	A4	06	00	03	00	00	00	01	00	00	00	00	A3	..α.....f															
0130h:	01	00	07	00	00	00	01	01	00	00	00	92	91	00	02	00' \.....															
0140h:	00	00	07	00	00	02	5A	A4	04	00	05	00	00	00	01	00Zα.....															
0150h:	00	02	61	88	22	00	03	00	00	00	01	00	00	00	00	A4	..a^".....α															
0160h:	03	00	03	00	00	00	01	00	00	00	00	A0	02	00	04	00															
0170h:	00	00	01	00	00	14	C0	92	90	00	02	00	00	00	07	00À'.....															
0180h:	00	02	69	92	01	00	0A	00	00	00	01	00	00	02	70	92	..i'.....p'															
0190h:	07	00	03	00	00	00	01	00	02	00	00	90	03	00	02	00															

Figure 13: File Signature for Android Image 1

After these images were sent through WhatsApp and downloaded, the Exif tag was no longer displayed in the Hex data. Instead, images downloaded from WhatsApp displayed the JPG JFIF file signature FF D8 FF E0 00 10 4A 46 49 46 00 01. Below are five examples of file headers of images downloaded from WhatsApp.

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0000h:	FF	D8	FF	E0	00	10	4A	46	49	46	00	01	01	00	00	01	ÿÿà..JFIF.....															
0010h:	00	01	00	00	FF	DB	00	43	00	08	06	06	07	06	05	08	...ÿÛ.C.....															
0020h:	07	07	07	09	09	08	0A	0C	14	0D	0C	0B	0B	0C	19	12															
0030h:	13	0F	14	1D	1A	1F	1E	1D	1A	1C	1C	20	24	2E	27	20 \$.'															
0040h:	22	2C	23	1C	1C	28	37	29	2C	30	31	34	34	34	1F	27	","#..(7),01444.'															
0050h:	39	3D	38	32	3C	2E	33	34	32	FF	DB	00	43	01	09	09	9=82<.342ÿÛ.C...															
0060h:	09	0C	0B	0C	18	0D	0D	18	32	21	1C	21	32	32	32	322!..!2222															
0070h:	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	2222222222222222															
0080h:	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	2222222222222222															
0090h:	32	32	32	32	32	32	32	32	32	32	32	32	32	32	FF	C2	222222222222222ÿÄ															
00A0h:	00	11	08	06	40	04	B0	03	01	22	00	02	11	01	03	11	...@.°..".....															
00B0h:	01	FF	C4	00	1A	00	01	01	01	01	01	01	01	00	00	00	..ÿÄ.....															
00C0h:	00	00	00	00	00	00	00	01	02	03	04	05	06	FF	C4	00ÿÄ.															
00D0h:	18	01	01	01	01	01	01	00	00	00	00	00	00	00	00	00															
00E0h:	00	00	00	01	02	03	04	FF	DA	00	0C	03	01	00	02	10ÿÛ.....															
00F0h:	03	10	00	00	02	CB	2C	DE	9C	F5	0C	EA	8C	4D	6C	E7Ë, ðœ.ê€Mlç															
0100h:	B2	B1	AC	41	DB	9E	84	CF	53	37	28	D4	E5	B0	D4	35	±-AÛž,,İs7(ôâ°ôS															

Figure 14: File Header of .jpg File Sent Over WhatsApp via iPhone

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0000h:	FF	D8	FF	E0	00	10	4A	46	49	46	00	01	01	00	00	01	ÿÿà..JFIF.....															
0010h:	00	01	00	00	FF	DB	00	84	00	06	06	06	06	07	06	07	...ÿÛ.....															
0020h:	08	08	07	0A	0B	0A	0B	0A	0F	0E	0C	0C	0E	0F	16	10															
0030h:	11	10	11	10	16	22	15	19	15	15	19	15	22	1E	24	1E".....".\$.															
0040h:	1C	1E	24	1E	36	2A	26	26	2A	36	3E	34	32	34	3E	4C	...\$.6*&&*6>424>L															
0050h:	44	44	4C	5F	5A	5F	7C	7C	A7	01	06	06	06	06	07	06	DDL_z_ s.....															
0060h:	07	08	08	07	0A	0B	0A	0B	0A	0F	0E	0C	0C	0E	0F	16															
0070h:	10	11	10	11	10	16	22	15	19	15	15	19	15	22	1E	24".....".\$.															
0080h:	1E	1C	1E	24	1E	36	2A	26	26	2A	36	3E	34	32	34	3E	...\$.6*&&*6>424>															
0090h:	4C	44	44	4C	5F	5A	5F	7C	7C	A7	FF	C2	00	11	08	06	LDDL z_ sÿÄ....															
00A0h:	40	04	B0	03	01	22	00	02	11	01	03	11	01	FF	C4	00	@.°..".....ÿÄ.															
00B0h:	31	00	00	03	01	01	01	01	00	00	00	00	00	00	00	00	1.....															
00C0h:	00	00	00	01	02	03	04	05	06	01	01	01	01	01	01	01															
00D0h:	00	00	00	00	00	00	00	00	00	00	00	01	02	03	04	05															
00E0h:	FF	DA	00	0C	03	01	00	02	10	03	10	00	00	02	CD	C9	ÿÛ.....íÉ															
00F0h:	96	B9	52	58	B6	19	AA	B3	3B	61	9D	E6	83	7C	E8	14	-¹RX¶.ª³;a.æf è.															

Figure 15: File Header of .jpg File Sent Over WhatsApp via Android

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0000h:	FF	D8	FF	E0	00	10	4A	46	49	46	00	01	01	00	00	01	ÿÿà..JFIF.....															
0010h:	00	01	00	00	FF	DB	00	43	00	03	02	02	02	02	02	03ÿÛ.C.....															
0020h:	02	02	02	03	03	03	04	06	04	04	04	04	04	04	08	06															
0030h:	06	05	06	09	08	0A	0A	09	08	09	09	0A	0C	0F	0C	0A															
0040h:	0B	0E	0B	09	09	0D	11	0D	0E	0F	10	10	11	10	0A	0C															
0050h:	12	13	12	10	13	0F	10	10	10	FF	DB	00	43	01	03	03ÿÛ.C...															
0060h:	03	04	03	04	08	04	04	08	10	0B	09	0B	10	10	10	10															
0070h:	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10															
0080h:	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10															
0090h:	10	10	10	10	10	10	10	10	10	10	10	10	10	10	FF	C0ÿÄ															
00A0h:	00	11	08	06	40	04	B0	03	01	22	00	02	11	01	03	11@.°..".....															
00B0h:	01	FF	C4	00	1D	00	00	01	05	01	01	01	01	00	00	00	..ÿÄ.....															
00C0h:	00	00	00	00	00	00	04	00	01	02	03	05	06	07	08	09															
00D0h:	FF	C4	00	4E	10	00	01	03	03	03	02	04	04	03	07	03	ÿÄ.K.....															
00E0h:	04	01	01	02	0F	01	02	03	11	00	04	05	12	21	31	06!1.															
00F0h:	41	13	22	51	61	14	71	81	91	07	32	A1	15	23	42	B1	A."Qa.q.'2j.#B±															
0100h:	C1	D1	F0	52	E1	F1	08	16	24	62	33	43	72	82	17	25	ÃÑÖRañ..\$b3Cr,.%															
0110h:	53	92	34	09	26	A2	B2	63	35	54	64	83	C2	FF	C4	00	S'4.&ç²c5TdfÿÄ.															
0120h:	1C	01	01	01	01	00	03	01	01	01	00	00	00	00	00	00															
0130h:	00	00	00	01	02	03	04	05	06	07	08	FF	C4	00	41	11ÿÄ.A.															
0140h:	00	02	02	01	03	03	01	05	06	05	03	03	03	03	05	01															
0150h:	00	01	02	11	03	04	21	31	05	12	41	51	06	13	22	61!1..AQ.."a															
0160h:	71	32	81	91	A1	B1	D1	14	23	C1	E1	F0	07	15	42	16	q2.'j;+Ñ.#Ááð..B.															
0170h:	52	F1	24	33	92	43	62	82	34	53	72	A2	C2	E2	FF	DA	Rñ\$3'Cb,4SrcÄÿÛ															
0180h:	00	0C	03	01	00	02	11	03	11	00	3F	00	CA	6B	AC	DC?.Ëk-Û															
0190h:	B3	BC	CE	64	AF	7F	11	AD	AE	2E	2E	ED	94	84	DB	38	³¼îd¯.-@.í",Û8															

Figure 16: File Header of .jpg File Sent Over WhatsApp via Windows Application

	0	1	2	3	4	5	6	7	8	9	Ä	B	C	D	E	F	0123456789ÄBCDEF
0000h:	FF	D8	FF	E0	00	10	4A	46	49	46	00	01	01	00	00	01	ÿÿà..JFIF.....
0010h:	00	01	00	00	FF	DB	00	43	00	03	02	02	02	02	02	03ÿÛ.C.....
0020h:	02	02	02	03	03	03	03	04	06	04	04	04	04	04	08	06
0030h:	06	05	06	09	08	0A	0A	09	08	09	09	0A	0C	0F	0C	0A
0040h:	0B	0E	0B	09	09	0D	11	0D	0E	0F	10	10	11	10	0A	0C
0050h:	12	13	12	10	13	0F	10	10	10	FF	DB	00	43	01	03	03ÿÛ.C...
0060h:	03	04	03	04	08	04	04	08	10	0B	09	0B	10	10	10	10
0070h:	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
0080h:	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
0090h:	10	10	10	10	10	10	10	10	10	10	10	10	10	10	FF	C0ÿÄ
00A0h:	00	11	08	06	40	04	B0	03	01	22	00	02	11	01	03	11@.°..".....
00B0h:	01	FF	C4	00	1D	00	00	01	05	01	01	01	01	00	00	00	..ÿÄ.....
00C0h:	00	00	00	00	00	00	04	00	01	02	03	05	06	07	08	09
00D0h:	FF	C4	00	4A	10	00	01	03	03	03	02	04	04	03	07	03	ÿÄ.J.....
00E0h:	04	01	02	00	0F	01	02	03	11	00	04	05	12	21	31	06!1.
00F0h:	41	13	22	51	61	14	71	81	91	07	32	A1	15	23	42	B1	A."Qa.q.'2;.#B±
0100h:	C1	D1	F0	52	E1	F1	08	16	24	62	33	43	72	82	17	25	ÄÑÖRáñ..\$b3Cr,;%
0110h:	53	09	34	92	B2	26	A2	35	63	64	83	C2	FF	C4	00	1C	S.4'²&¢5cdfÄÿÄ..
0120h:	01	01	01	01	00	03	01	01	01	00	00	00	00	00	00	00
0130h:	00	00	01	02	03	04	05	06	07	08	FF	C4	00	41	11	00ÿÄ.A..
0140h:	02	02	01	03	03	01	05	06	05	03	04	01	04	01	05	00
0150h:	01	02	11	03	04	21	31	05	12	41	51	06	13	22	61	71!1..AQ.."aq
0160h:	32	81	91	A1	B1	D1	14	23	C1	E1	F0	07	15	42	16	33	2.';±Ñ.#Ááð..B.3
0170h:	52	F1	92	24	43	62	82	34	53	72	A2	C2	E2	FF	DA	00	Rñ'şCb,4SrcÄÄÿÛ.
0180h:	0C	03	01	00	02	11	03	11	00	3F	00	CA	6B	AC	DC	B3?.Èk-Û³
0190h:	BC	CE	64	AF	7F	11	AD	AE	2E	2E	ED	94	84	DB	38	E4	¿íd¯...-@.i"„Û8ä

Figure 17: File Header of .jpg File Sent Over WhatsApp via MacOS Application

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0000h:	FF	D8	FF	E0	00	10	4A	46	49	46	00	01	01	00	00	01	ÿÿà..JFIF.....															
0010h:	00	01	00	00	FF	DB	00	43	00	03	02	02	02	02	02	03ÿÛ.C.....															
0020h:	02	02	02	03	03	03	04	06	04	04	04	04	04	08	06																
0030h:	06	05	06	09	08	0A	0A	09	08	09	09	0A	0C	0F	0C	0A															
0040h:	0B	0E	0B	09	09	0D	11	0D	0E	0F	10	10	11	10	0A	0C															
0050h:	12	13	12	10	13	0F	10	10	10	FF	DB	00	43	01	03	03ÿÛ.C...															
0060h:	03	04	03	04	08	04	04	08	10	0B	09	0B	10	10	10	10															
0070h:	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10															
0080h:	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10															
0090h:	10	10	10	10	10	10	10	10	10	10	10	10	10	10	FF	C0ÿÀ															
00A0h:	00	11	08	06	40	04	B0	03	01	22	00	02	11	01	03	11@.°.".....															
00B0h:	01	FF	C4	00	1D	00	00	01	05	01	01	01	01	00	00	00	.ÿÀ.....															
00C0h:	00	00	00	00	00	00	04	00	01	02	03	05	06	07	08	09															
00D0h:	FF	C4	00	4E	10	00	01	03	03	03	02	04	04	04	06	01	ÿÀ.K.....															
00E0h:	03	03	01	01	11	01	02	03	11	00	04	05	12	21	31	06!1.															
00F0h:	41	13	22	51	61	14	71	81	91	07	32	A1	B1	15	23	42	A."Qa.q.'.2;±.#B															
0100h:	C1	D1	F0	E1	24	52	F1	08	16	33	62	72	17	43	82	25	ÃÑóá\$Rñ..3br.C,%															
0110h:	53	92	09	34	A2	B2	26	35	44	54	63	64	83	FF	C4	00	S'.4ç²&5DTcdfÿÀ.															
0120h:	1C	01	01	01	01	00	03	01	01	01	00	00	00	00	00	00															
0130h:	00	00	00	01	02	03	04	05	06	07	08	FF	C4	00	41	11ÿÀ.A.															
0140h:	00	02	02	01	03	03	01	05	06	05	03	03	03	03	05	01															
0150h:	00	01	02	11	03	04	21	31	05	12	41	51	06	13	22	61!1..AQ.."a															
0160h:	71	32	81	91	A1	B1	D1	14	23	C1	E1	F0	07	15	42	16	q2.';±Ñ.#Ááδ..B.															
0170h:	33	52	24	92	F1	43	62	82	34	53	72	A2	C2	E2	FF	DA	3R\$'ñCb,4SrcÃÿÛ															
0180h:	00	0C	03	01	00	02	11	03	11	00	3F	00	CA	6B	AC	DC?.Èk-Û															
0190h:	B3	BC	CE	64	AF	7F	11	AD	AE	2E	2E	ED	94	84	DB	38	³¼îd̄...@.í"„Û8															

Figure 18: File Header of .jpg File Sent Over WhatsApp via Web Application

As visualized, header information is different for images sent via iPhone as opposed to images sent via Android. However, images sent via the WhatsApp application on Windows 10, MacOS, or the WhatsApp web applications have extremely similar header information with practically all the differences between those files residing in the actual image data stream.

Baseline JPEG vs. Progressive JPEG

Images were inputted into ExifTool to observe if they were compressed with Baseline or Progressive encoding. Original images captured on iPhone or Android mobile devices were compressed with Baseline DCT encoding. Images sent over WhatsApp via an iPhone or Android device were compressed and displayed with Progressive DCT encoding. This was apparent in the WhatsApp application. As these images downloaded and were being decoded, a blurry version of

the image was displayed in its entirety before the download was completed. When images were sent via the Windows 10, MacOS, or WhatsApp Web Application, the images were compressed with Baseline DCT encoding.

Quantization Tables

Images were inputted into JPEG Snoop to observe quantization tables and quality factors. Images sent via the Windows, MacOS, or WhatsApp Web Applications all had the same quality factor identified. The level of compression applied by WhatsApp when images were sent using a desktop application is similar to the level of compression applied to the original iPhone images and is a relatively low level of compression when compared to images downloaded from WhatsApp that were sent over mobile devices.

```
Precision=8 bits
Destination ID=0 (Luminance)
DQT, Row #0:  2  2  2  3  5  6  8 10
DQT, Row #1:  2  2  2  3  5  6  8 10
DQT, Row #2:  2  2  3  5  6  8 10 12
DQT, Row #3:  3  3  5  6  8 10 12 14
DQT, Row #4:  5  5  6  8 10 12 14 15
DQT, Row #5:  6  6  8 10 12 14 15 15
DQT, Row #6:  8  8 10 12 14 15 15 15
DQT, Row #7: 10 10 12 14 15 15 15 15
Approx quality factor = 91.94 (scaling=16.12 variance=12.56)
-----
Precision=8 bits
Destination ID=1 (Chrominance)
DQT, Row #0:  2  2  4  7 16 16 16 16
DQT, Row #1:  2  4  4 11 16 16 16 16
DQT, Row #2:  4  4  9 16 16 16 16 16
DQT, Row #3:  7 11 16 16 16 16 16 16
DQT, Row #4: 16 16 16 16 16 16 16 16
DQT, Row #5: 16 16 16 16 16 16 16 16
DQT, Row #6: 16 16 16 16 16 16 16 16
DQT, Row #7: 16 16 16 16 16 16 16 16
Approx quality factor = 92.03 (scaling=15.95 variance=1.27)
```

Figure 19: *Quantization Tables and Quality Factor of Original iPhone Image*

```

Precision=8 bits
Destination ID=0 (Luminance)
DQT, Row #0: 3 2 2 3 4 6 8 10
DQT, Row #1: 2 2 2 3 4 9 10 9
DQT, Row #2: 2 2 3 4 6 9 11 9
DQT, Row #3: 2 3 4 5 8 14 13 10
DQT, Row #4: 3 4 6 9 11 17 16 12
DQT, Row #5: 4 6 9 10 13 17 18 15
DQT, Row #6: 8 10 12 14 16 19 19 16
DQT, Row #7: 12 15 15 16 18 16 16 16
Approx quality factor = 91.86 (scaling=16.28 variance=1.13)

*** Marker: DQT (xFFDB) ***
Define a Quantization Table.
OFFSET: 0x00000059
Table length = 67
----
Precision=8 bits
Destination ID=1 (Chrominance)
DQT, Row #0: 3 3 4 8 16 16 16 16
DQT, Row #1: 3 3 4 11 16 16 16 16
DQT, Row #2: 4 4 9 16 16 16 16 16
DQT, Row #3: 8 11 16 16 16 16 16 16
DQT, Row #4: 16 16 16 16 16 16 16 16
DQT, Row #5: 16 16 16 16 16 16 16 16
DQT, Row #6: 16 16 16 16 16 16 16 16
DQT, Row #7: 16 16 16 16 16 16 16 16
Approx quality factor = 91.90 (scaling=16.20 variance=0.15)

```

Figure 20: *Quantization Tables and Quality Factor of Image Sent Over WhatsApp via Windows*

Application

Images sent over the mobile WhatsApp application had more aggressive compression applied to them as indicated by higher numbers in the quantization tables and lower quality factors identified.

```

Precision=8 bits
Destination ID=0 (Luminance)
DQT, Row #0:  8  6  5  8 12 20 26 31
DQT, Row #1:  6  6  7 10 13 29 30 28
DQT, Row #2:  7  7  8 12 20 29 35 28
DQT, Row #3:  7  9 11 15 26 44 40 31
DQT, Row #4:  9 11 19 28 34 55 52 39
DQT, Row #5: 12 18 28 32 41 52 57 46
DQT, Row #6: 25 32 39 44 52 61 60 51
DQT, Row #7: 36 46 48 49 56 50 52 50
Approx quality factor = 74.75 (scaling=50.51 variance=0.81)

*** Marker: DQT (xFFDB) ***
Define a Quantization Table.
OFFSET: 0x00000059
Table length = 67
----
Precision=8 bits
Destination ID=1 (Chrominance)
DQT, Row #0:  9  9 12 24 50 50 50 50
DQT, Row #1:  9 11 13 33 50 50 50 50
DQT, Row #2: 12 13 28 50 50 50 50 50
DQT, Row #3: 24 33 50 50 50 50 50 50
DQT, Row #4: 50 50 50 50 50 50 50 50
DQT, Row #5: 50 50 50 50 50 50 50 50
DQT, Row #6: 50 50 50 50 50 50 50 50
DQT, Row #7: 50 50 50 50 50 50 50 50
Approx quality factor = 74.74 (scaling=50.52 variance=0.19)

```

Figure 21: Quantization Tables and Quality Factor of Image Sent Over the iPhone WhatsApp Application

```

Precision=8 bits
Destination ID=0 (Luminance)
DQT, Row #0:  6  6  6  7 10 15 22 34
DQT, Row #1:  6  7  8 11 14 16 21 30
DQT, Row #2:  6  8 10 12 17 25 36 54
DQT, Row #3:  7 11 12 16 21 30 42 62
DQT, Row #4: 10 14 17 21 28 38 52 76
DQT, Row #5: 15 16 25 30 38 50 68 95
DQT, Row #6: 22 21 36 42 52 68 90 124
DQT, Row #7: 34 30 54 62 76 95 124 167
Approx quality factor = 71.19 (scaling=57.62 variance=593.35)
----
Precision=8 bits
Destination ID=1 (Chrominance)
DQT, Row #0:  6  6  6  7 10 15 22 34
DQT, Row #1:  6  7  8 11 14 16 21 30
DQT, Row #2:  6  8 10 12 17 25 36 54
DQT, Row #3:  7 11 12 16 21 30 42 62
DQT, Row #4: 10 14 17 21 28 38 52 76
DQT, Row #5: 15 16 25 30 38 50 68 95
DQT, Row #6: 22 21 36 42 52 68 90 124
DQT, Row #7: 34 30 54 62 76 95 124 167
Approx quality factor = 80.24 (scaling=39.51 variance=961.47)

```

Figure 22: Quantization Tables and Quality Factor of Image Sent Over the Android WhatsApp Application

and downloaded to configure the database. Ten new audio recordings were recorded in the Apple Voice Memos application with the “Lossless” setting enabled and sent through WhatsApp. These recordings were downloaded from WhatsApp and used to create a WhatsApp audio compression profile.

After configuring the Lossy Compression Analysis database with the longer files initially recorded with lossless compression and then sent through WhatsApp, the lossy compression applied by WhatsApp was able to be observed and verified. When the ten shorter recordings recorded with the lossless setting in Apple Voice memos enabled were compared to the database, the WhatsApp compression was detected.

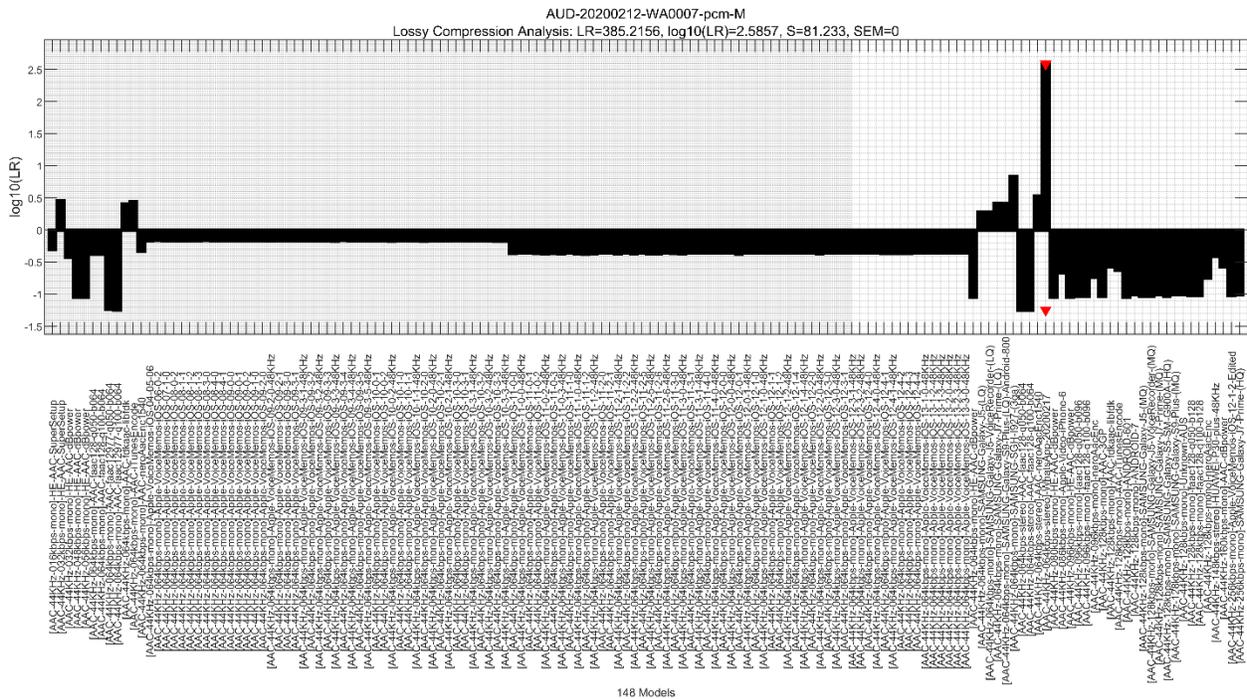


Figure 24: Lossy Compression Analysis Results of an Audio File Recorded With the “Lossless” Setting in the Apple Voice Memos Application Enabled and Sent Through WhatsApp

Hex Data (Audio)

Hex data analysis and observations of the audio files examined in this study was conducted using 010 Editor. Original audio recordings were recorded with the lossless setting enabled. In the original recordings, file metadata is mainly contained in the footer of the file after the core audio data. After the WhatsApp compression, metadata is no longer seen in the footer of the file, but rather written in the header of the file before the core audio data. After the WhatsApp compression, ASCII strings associated with the iPhone are no longer seen. When viewed in ExifTool, metadata identifying Apple and iPhone as the encoder is also no longer visible after the WhatsApp compression.

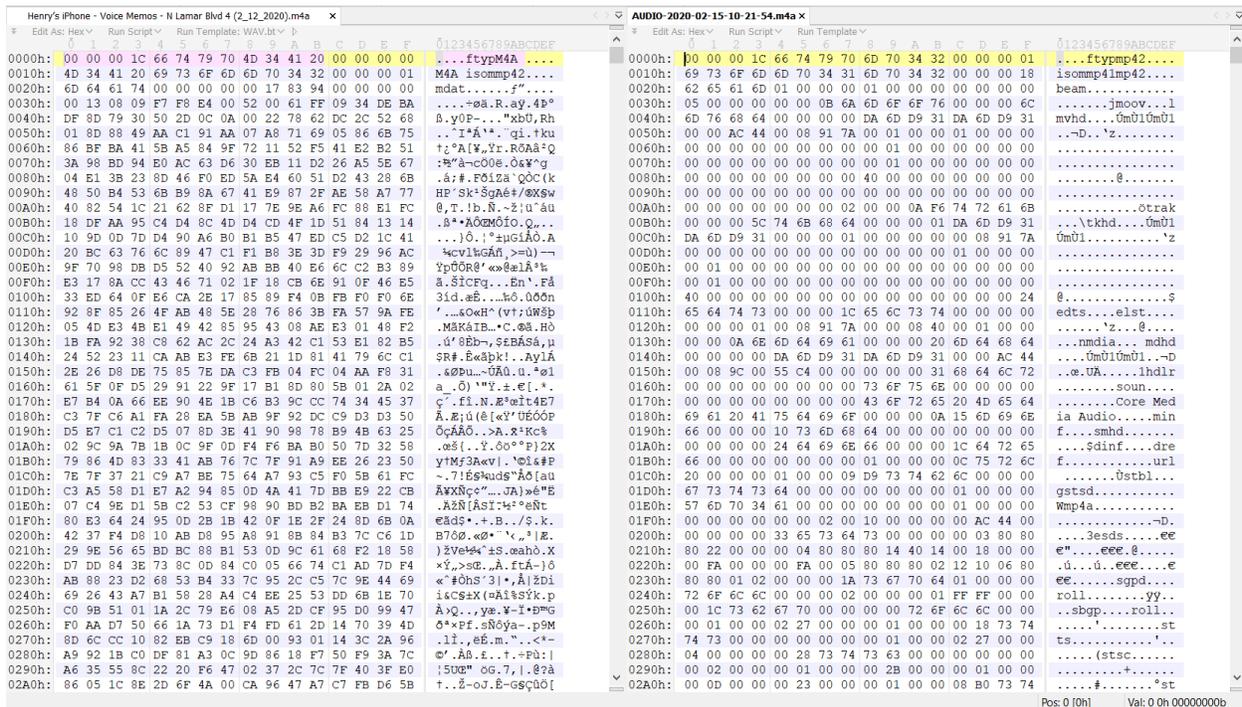


Figure 25: File Header of Original Lossless Recording on Left, and That Same Audio Recording Compressed by WhatsApp on Right

e - Voice Memos - N Lamar Blvd 4 (2_12_2020).m4a																Henry's iPhone - Voice Memos - N Lamar Blvd (2_12_2020).m4a																AUDIO-2020-02-15-10-21-54.m4a																AUDIO-2020-02-15-10-21-18.m4a																																																																																																																																																																																																																																																																																																																			
Edit As: Hex				Run Script				Run Template				0123456789ABCDEF				0123456789ABCDEF				0123456789ABCDEF				0123456789ABCDEF																																																																																																																																																																																																																																																																																																																																											
0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F																																																																																																																																																																																																																																																																																																				
0000h:	00	00	00	1c	66	74	79	70	4d	34	41	20	00	00	00	0020h:	62	65	61	6d	01	00	00	01	00	00	00	00	00	00	00	0000h:	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	0000h:	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00																																																																																																																																																																																																																																																																																																				
0010h:	4d	34	41	20	69	73	6f	6d	6d	70	34	32	00	00	01	0030h:	06	00	00	00	00	0b	46	6d	6f	6f	76	00	00	00	00	0001h:	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	0001h:	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00																																																																																																																																																																																																																																																																																																				
0020h:	6d	64	61	74	00	00	00	00	00	01	39	92	00	00	07	0040h:	6d	76	68	64	00	00	00	00	00	00	00	00	00	00	00	0002h:	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	0002h:	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00																																																																																																																																																																																																																																																																																																				
0030h:	01	06	14	ad	f4	76	22	1c	46	04	10	bf	e6	61	9d	fd	0050h:	00	00	bb	80	00	06	dd	1c	00	01	00	00	01	00	00	0003h:	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	0003h:	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00																																																																																																																																																																																																																																																																																																			
0040h:	63	b9	1b	32	a9	89	18	07	d9	72	a6	c4	c3	ef	10	f5	0060h:	00	00	00	00	00	00	00	00	00	01	00	00	00	00	00	0004h:	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	0004h:	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00																																																																																																																																																																																																																																																																																																			
0050h:	03	1e	5f	82	43	e4	47	ce	52	3c	8f	e5	29	3f	ae	9e	0080h:	00	00	00	00	00	00	00	00	00	40	00	00	00	00	00	0005h:	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	0005h:	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00																																																																																																																																																																																																																																																																																																			
0060h:	ec	27	d0	3e	1d	13	fa	35	f4	c0	9f	ab	fc	28	43	db	0090h:	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	0006h:	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	0006h:	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00																																																																																																																																																																																																																																																																																																			
0070h:	1e	54	23	8d	e4	44	fb	27	7b	27	d5	b5	76	8d	02	7b	00a0h:	00	00	00	00	00	00	02	00	00	09	2c	74	72	61	6e	0007h:	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	0007h:	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00																																																																																																																																																																																																																																																																																																			
0080h:	ad	e6	4e	9e	46	14	b2	7d	fb	b3	13	df	d0	27	04	04	00b0h:	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	0008h:	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	0008h:	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00																																																																																																																																																																																																																																																																																																			
0090h:	2f	6c	48	6f	b5	e4	32	b4	c8	0a	41	0d	17	01	a2	68	00c0h:	da	6a	46	d8	00	00	00	00	00	00	00	00	00	00	00	0009h:	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	0009h:	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00																																																																																																																																																																																																																																																																																																			
00a0h:	12	3b	08	57	41	3b	fa	02	6c	77	e2	bb	94	99	03	d2	00d0h:	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	0010h:	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	0010h:	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00																																																																																																																																																																																																																																																																																																			
00b0h:	d8	f0	3d	71	6f	90	9a	0b	d1	ef	cb	9e	3d	7f	11	ba	00e0h:	00	01	00	00	00	00	00	00	00	00	00	00	00	00	00	0011h:	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	0011h:	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00																																																																																																																																																																																																																																																																																																			
00c0h:	ff	4f	64	80	d8	bf	68	d6	a1	6d	eb	79	dd	3c	82	c0	00f0h:	6d	64	69	61	00	00	00	20	6d	64	68	64	00	00	00	0012h:	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	0012h:	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00																																																																																																																																																																																																																																																																																																			
00d0h:	6c	ef	d1	88	a3	26	af	83	a6	f2	8f	80	db	01	47	63	0100h:	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	0013h:	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	0013h:	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00																																																																																																																																																																																																																																																																																																			
00e0h:	db	cb	1a	c8	b6	4f	17	c7	30	ce	ba	25	b9	84	14	a0	0110h:	88	54	76	b7	02	9c	7d	ab	0f	a6	92	33	e8	d6	fa	0120h:	bb	bb	02	9e	31	16	00	03	c6	b8	1c	12	da	f9	be	5f	0130h:	2e	03	6c	3e	ef	ca	3a	12	80	07	01	36	14	ab	b0	93	0140h:	1b	52	da	a6	9a	9c	6a	d0	9d	ea	69	a5	32	59	31	bd	0150h:	ec	35	60	32	fb	f3	4a	e8	28	6d	4c	1f	93	14	52	27	0160h:	8c	9d	03	2f	a1	9b	5e	6a	35	ee	b1	95	ee	5d	4f	c3	0170h:	af	67	90	3e	ef	de	8a	7c	fe	2d	2d	b7	e2	7d	12	17	0180h:	0f	1f	d7	03	b2	e9	12	2c	a3	93	be	59	c9	7b	ab	db	0190h:	6d	3b	e5	8a	87	ca	44	2a	57	2f	ab	5f	ce	6e	76	cc	0200h:	8e	36	34	96	b8	19	8c	87	d5	1c	a6	65	bc	73	d6	8d	0210h:	0b	0f	85	ca	97	73	da	af	36	8c	cf	99	de	27	f0	17	0220h:	f1	91	ef	0d	f4	5b	d4	7e	5f	1a	a2	73	11	45	58	4b	0230h:	03	1e	5f	82	43	e4	47	ce	52	3c	8f	e5	29	3f	ae	9e	0240h:	7a	be	34	9b	4b	b6	7b	f0	76	d8	fc	19	fa	19	4d	49	0250h:	2e	ee	be	71	77	b3	65	af	7e	7d	e2	8f	12	5d	de	dc	0260h:	6b	72	15	63	8d	bb	5c	94	ee	71	0f	5d	b6	5a	2c	21	0270h:	2e	5b	cf	96	9c	25	51	97	07	c1	7d	0b	f1	f9	74	37	0280h:	79	be	bc	bf	2b	8a	78	1c	58	41	00	01	30	14	a1	4c	0290h:	84	41	14	44	44	12	20	79	59	db	c8	bc	16	5e	1a	c3	02a0h:	ac	00	7c	f3	f3	43	b2	c7	4f	68	e9	92	97	3a	42	85

Figure 27: File Header of Original Recording Compressed by Apple Voice Memos on Left, and That Same Recording Compressed by WhatsApp on Right

CHAPTER VI

DISCUSSION

Conclusions

Based on the results of the analysis, it appears as though different compression schemes are applied to images that are sent over WhatsApp depending on which of three broad methods are used to send them. Those methods being over iPhone, Android, or desktop applications. Although this study explored many different methods of sending media files, only these broader categories of methods influenced how the media files were changed. None of the more specific sending methods within those broader methods or the download methods used influenced how the files were compressed. The first observation made was that the naming convention applied to the downloaded images was dictated by the type of device or web application that was used to download them. The most striking differences in images downloaded from WhatsApp were observed between images sent via cellular devices and those sent via desktop applications. Images that were sent via mobile devices had more aggressive compression applied to them based on observing the quantization tables and quality factors identified, and these images were compressed with Progressive DCT encoding. This is compared to images that were sent via desktop devices that had much less aggressive compression applied to them and were compressed with Baseline DCT encoding.

There were also some commonalities among all images that were sent and then downloaded from WhatsApp compared to their original counterparts. After the transmission, the volume of Exif entries identified was significantly reduced to 22 entries no matter how many entries were observed in the original images. The Exif tag was lost, and instead the JFIF file signature was identified.

An interesting finding was that in some cases, but not all, there is uniformity between images that are sent from different desktop applications. However, based on the Exif data, Hex data, and quantization tables, we can see that the same compression is used by all desktop applications examined in this study. The information gathered here can be used to help identify images that have been sent over the WhatsApp application and the type of device that was used to send them.

Below is a table visualizing the variable data associated with each different downloaded image. Color coded rows indicate images that have matching hash values.

	A	B	C	D	E	F
1		Size	Baseline/Progressive Compression	Dimensions	QF (Luminance)	QF (Chrominance)
2	iPhone Image 1 (Sent via Original)	3125 kB	Baseline DCT	4032x3024	91.94	92.03
3	iPhone Image 1 (Sent via Android)	182 kB	Progressive DCT	1200x1600	71.19	80.24
4	iPhone Image 1 (Sent via iPhone)	209 kB	Progressive DCT	1200x1600	74.75	74.74
5	iPhone Image 1 (Sent via Mac)	443 kB	Baseline DCT	1200x1600	91.86	91.9
6	iPhone Image 1 (Sent via Web App)	443 kB	Baseline DCT	1200x1600	91.86	91.9
7	iPhone Image 1 (Sent via Windows)	443 kB	Baseline DCT	1200x1600	91.86	91.9
8						
9	iPhone Image 2 (Original)	2919 MB	Baseline DCT	4032x3024	91.94	92.03
10	iPhone Image 2 (Sent via Android)	123 kB	Progressive DCT	1600x1200	71.19	80.24
11	iPhone Image 2 (Sent via iPhone)	153 kB	Progressive DCT	1600x1200	74.75	74.74
12	iPhone Image 2 (Sent via Mac)	319 kB	Baseline DCT	1600x1200	91.86	91.9
13	iPhone Image 2 (Sent via Web App)	321 kB	Baseline DCT	1600x1200	91.86	91.9
14	iPhone Image 2 (Sent via Windows)	319 kB	Baseline DCT	1600x1200	91.86	91.9
15						
16	Android Image 1 (Sent via Original)	6775 kB	Baseline DCT	5312x2988	96.95	96.99
17	Android Image 1 (Sent via Android)	93 kB	Progressive DCT	747x1328	71.19	80.24
18	Android Image 1 (Sent via iPhone)	146 kB	Progressive DCT	900x1600	74.75	74.74
19	Android Image 1 (Sent via Mac)	382 kB	Baseline DCT	900x1600	91.86	91.9
20	Android Image 1 (Sent via Web App)	382 kB	Baseline DCT	900x1600	91.86	91.9
21	Android Image 1 (Sent via Windows)	382 kB	Baseline DCT	900x1600	91.86	91.9
22						
23	Android Image 2 (Sent via Original)	9550 kB	Baseline DCT	5312x2988	96.95	96.99
24	Android Image 2 (Sent via Android)	225 kB	Progressive DCT	1328x747	71.19	80.24
25	Android Image 2 (Sent via iPhone)	337 kB	Progressive DCT	1600x900	74.75	74.74
26	Android Image 2 (Sent via Mac)	649 kB	Baseline DCT	1600x900	91.86	91.9
27	Android Image 2 (Sent via Web App)	648 kB	Baseline DCT	1600x900	91.86	91.9
28	Android Image 2 (Sent via Windows)	649 kB	Baseline DCT	1600x900	91.86	91.9

Figure 29: Variable Data Table

For the audio portion of this study, observations of the hex data associated with the audio files that were sent and then downloaded from WhatsApp were recorded. A model of the WhatsApp audio compression was then detected and configured for a Lossy Compression

Analysis database. This information can be used to help identify audio recordings that have been sent over the WhatsApp application.

Further Research

WhatsApp has more capabilities for sending and receiving media files that were not explored in this study. Some of the upload and download methods available within WhatsApp were excluded from this study because they were only available in group chat communications. Some additional downloading methods were also available for images when more than three were sent without a text message in between. However, based on the research done in this study it would seem unlikely that these methods would create different images.

Another capability that was excluded from this study is the ability to send and receive videos. Videos can be recorded within the WhatsApp application, or videos that are stored on a device can be selected and sent through WhatsApp. Within WhatsApp, audio messages and images can also be captured without leaving the app and then immediately sent. How the WhatsApp application captures images, video and audio could most definitely also be an interesting avenue for further research.

In this study, only 4 source images were used. Those images were taken using only two mobile devices, and the image capture settings were not changed. Those two mobile devices were the only mobile devices used in this study. This opens the door for further research utilizing images of different sizes, dimensions, and file types. Using different mobile devices as well as desktop devices to send and receive images could be a valuable topic for further research considering the plethora of different devices available to the consumer that are supported by WhatsApp.

REFERENCES

- [1] Osborne, Samuel. "Stockholm suspect Rakhmat Akilov 'exchanged Whatsapp messages with Isis supporter before and after attack'" *The Independent*. 10 Apr. 2017.
- [2] "WhatsApp." *Wikipedia*. Wikimedia Foundation, 2020.
- [3] Graham, Robert. "How Terrorists Use Encryption" *Combating Terrorism at West Point*, 15 Nov. 2017.
- [4] Clement, J. "Social Media – Statistics and Facts." *Statista*, 4 Sep. 2019.
- [5] Lomboy, Gretchel. "Digital Image Recompression Analysis: Twitter." Thesis. University of Colorado, Denver, 2018. *Research Theses from the Master of Science in Recording Arts Emphasis in Media Forensics Program*. National Center for Media Forensics.
- [6] Douglas, Zachary. "Digital Image Recompression Analysis of Instagram." Thesis. University of Colorado, Denver, 2015. *Research Theses from the Master of Science in Recording Arts Emphasis in Media Forensics Program*. National Center for Media Forensics.
- [7] D. Hamdi, F. Iqbal, T. Baker and B. Shah, "Multimedia File Signature Analysis for Smartphone Forensics," *2016 9th International Conference on Developments in eSystems Engineering (DeSE)*, Liverpool, 2016, pp. 130-137.
- [8] Anglano, Cosimo. "Forensic Analysis of WhatsApp Messenger on Android Smartphones," *Digital Investigation Journal*, Alessandria, Italy. Vol 11. Sep. 2014.
- [9] Hudson, Graham, et al. "JPEG-1 Standard 25 Years: Past, Present, and Future Reasons for a Success." *Journal of Electronic Imaging*, International Society for Optics and Photonics, 2018
- [10] "SWGDE Position on the Use of MD5 and SHA1 Hash Algorithms in Digital and Multimedia Forensics" Version: 1.0. *SWGDE*. Nov 20, 2018.
- [11] AV&IT Standardization Committee. "JEITA CP-3451C Exchangeable image file format for digital still cameras: Exif Version 2.3". *Standards of Japan Electronics and Information and Technology Industries Association*. April, 2010.
- [12] Mackenzie, Charles E. *Coded Character Sets, History and Development*. Reading: Addison-Wesley, 1980. *The Systems Programming Series*.
- [13] "Image Format - Baseline JPEG vs. Progressive JPEG." *Tourwriter Knowledge Base*. N.p., 09 Aug. 2017.
- [14] Rajwade, Ajit. "Image Compression". *CSE, IIT Bombay*. Fall, 2014. Lecture.
- [15] Smith, Steven W., "Chapter 27: Data Compression" *The Scientist and Engineer's Guide to*

Digital Signal Processing (2002).

[16] Craig, William. "JPEG 101: A Crash Course Guide on JPEG." *WebFX Blog*, 21 Dec. 2018,

[17] Grigoras, Catalin. Smith, Jeff. "Forensic Analysis of AAC Encoding on Apple iPhone Voice Memos Recordings. *The Audio Engineering Society*. Conference on Audio Forensics. 2019.

[18] "SWGDE Digital & Multimedia Evidence Glossary Version 3.0." *Scientific Working Group on Digital Evidence*. SWGDE, 23 June 2016.