

# **COMPLEX HASHING - HASHING FOR IMAGES USING METADATA**

Sivadanam Yaswanth Lingam

Panguluri Sai Deepika

yaswanth0212@gmail.com

saideepika.panguluri@gmail.com

TIFAC CORE in Cyber Security, Amrita Viswa Vidyapeetham, Coimbatore, India

**Abstract.** Cyber forensics is a finding technique that is used to control and expose technical illegal evidence of any Automated Devices. It involves electronic data storage extraction for legal issues to find pieces of evidence that help to reconstruct and report the actual incident. Mobile Forensics is the challenging forensic area as the device is completely different from the General digital devices for forensics. In general, forensics action is based on the hashes which make the evidence unique and if any small change can recognize easily. Proposing a new hashing technique with the help of existing hashing.

**Keywords:** Hashing, Message digest, Padding

## **1.INTRODUCTION**

In cyber forensics, mobile forensics is toughest and hard to identify the evidence as it is more sophisticated than the computer it has a highly advanced operating system which is not having specific log monitoring system which makes a mobile vulnerable to the real world and in this hashing of the file is having the prominent role. A hash that provides the unique integrity of the files in mobile will help to overcome the vulnerability in the mobile environment.

## **2.EXISTING SYSTEM**

In general, if we pass a file/image to the MD5 hash tool, it processes the given input file and gives us hash. This hash is based on the content in the given input file. So, if I pass copied in another device file to hash calculator, you will get the hash which is equal to the hash which generated in other devices as a duplicate copy but provides original hash. So, there will be no difference in the device copy and to find out the origin of the file



Fig. 1. Output for regular MD5 hash

We believe there are many hashing algorithms and techniques proposed and still proposing. we use of metadata in hashing along with data.

Furthermore, in this document, we use the hash md5 and this tool is used to combine the hash of data and metadata to create a unique hash to be used for mobile forensic.

### 3. MOBILE FORENSICS

Mobile forensics is used the recovery of evidence or digital data from a mobile phone in forensic conditions [1]. Mobile device generally refers to a digital device with inner memory & communication capabilities, with PDA devices, Global Positioning System devices, and a tablet. Images are common digital evidence that is easily operated. The mobile forensic process is divided into three major types: crisis analysis, acquisition, and examination and forensic auditors encounter a few difficulties in using a mobile device as evidence.[2] In crime scene, if the mobile device is switched off, the forensic officer use a faraday bag to avoid changes if the device turns on automatically or intentionally. Faraday bags are designed for the phone isolation from the network. Incase phone

is turned on, turn it off. If the PIN or password is locked or encrypted try to unlock it only after taking certain precautions. Mobile phones are network devices that can transmit data through many ways like Wi-Fi and tele communications. So once Mobile is switched on, an adversary can wipe the data within the device by operating an isolated wipe operation. When a mobile is switched on, it must be placed in a Faraday bag. If possible, before keeping any mobile in the Faraday bag, separate it from the network to guard the tests by enabling flight mode and deactivating all network connections (Wi-Fi, GPS, access points, etc.). This will also preserve the battery, which will discharge while in a Faraday bag and protect against leaks in the Faraday bag. Once the mobile device has been successfully confiscated, the examiner may need various forensic tools to acquire and analyze the data stored on the phone [1].

#### **4. CHALLENGES OF MOBILE FORENSICS**

One of the major forensic encounters in the mobile platform is the element that information can be retrieved, stored, and matched across multiple mobile devices. Since data is unstable and can be remotely altered or deleted, further efforts are needed to preserve it. Mobile forensic analysis is unlike mainframe forensic analysis and creates unique encounters for forensic examiners [2].

Here a few of the difficulties to assemble evidence from the mobile devices:

**4.1 Hardware Issues:** Mobile phones can be from different manufacturers. Forensic auditors can face a various number of mobile types, which are in size, hardware, features, and operating system [5]. Also, with a short product development cycle, new models emerge very frequently. The mobile scenery is fluctuating day by day, this is tough for the auditor to adjust to all the tasks and keep on updating on mobile forensic techniques [2].

**4.2 The operating system of Mobile:** In computers Windows is dominant in the world market for years, But mobile devices has lot of operating systems namely iOS, Android, Windows Mobile, Bada OS ,HP's web OS, Symbian Nokia OS and etc., [2].

**4.3 Mobile security:** Modern mobiles are having integrated security options to secure user's data. These features act as an obstacle during acquisition and forensic examination. For example, modern mobile devices have pre-determined encryption mechanisms inbuilt and may have to break these encryption mechanisms to extract data from devices [2].

**4.4 Lack of Resources:** Requirement of tools by a forensic auditor will also rise along with the no of devices. Forensic objects, like USB cables, batteries, and chargers for various mobile phones [4] [6].

**4.5 Anti-forensic techniques:** These are very rare but tough to identify such as data obfuscation, data falsification, and secure cleaning, are setting investigations tough in digital data [2].

**4.6 Unintended reset:** Cell phones offer functions to restore everything. Accidentally resetting the device during the exam can cause data loss [2].

**4.7 Remote Wipe:** This can delete or wipe the devices from any remote location which can lead the data wipeout permanently.

**4.8 Passcode recovery:** If the mobile is secured with a password; the forensic examiner must have entree to the mobile device to get back the passcode by existing techniques[2].

## 5. HASHING

The hash value is the output of the hash algorithms. The hash can be performed on images, strings, electronic files, and even on the entire hard drive contents. The result is therefore called the hash codes or hashes. There are many hash algorithms among which MD5, SHA-1, SHA-256, SHA-512 are popular. These hash functions play a very important role in digital forensics to calculate and authenticate so that data has not been altered [7]. When data is modified within the file it is considered a new file. So, a new hash value shall generate. The hash values are reliable, fast and this is secure to compare the data of each file and media, though it is one file containing a small amount of data or a large number of data i.e., more than Terabytes on the server.

### 5.1 MD5 Hash

MD5 hash algorithm is the fifth version of message digest algorithms technologically advanced and it is a 128-bit digest. It takes the plain text of 512-bit blocks which is divided into 16 blocks, each of 32 bits, and produces 128 message digests. The MD5 hash procedure is a one-side cryptographic mechanism that takes a message of any size as input and gives the output as a fixed size value which can be used for authenticating the original message [9].

This cryptographic mechanism has five steps mainly:

**Step 1** is adding the padding bits to the original message is done.

**Step 2** is post padding 64 bits have to add at the end to find the length of the original input.

**Step 3** is starting the buffer of the message digest.

**Step 4** is processing message within 16-word block

**Step 5** is output.

## 6. DATASETS

FORMAT	COUNT OF IMAGES
Bmp	20
Png	60
Jpg	60
Psd	60
Tif	60

### Before applying proposed model (MD5)

SNo	File name	File size (in KB)	File format	MD5 Hash
1.	1.jpg	158	JPG	2E03C3D6AC7D0E4E59707F1BFA36D64B
2.	2.jpg	169		3FF72695260722D20A148D0745E8AF9D
3.	3.jpg	171		B64F9AE9703E4C7F8CAC9B302A91A8AA
4.	4.jpg	164		5F1A3EF0CA822FC387F7862E65D9321C
5.	5.jpg	170		F8E12554E92EAB42A88DCD88927BFFF6
6.	1.png	30	PNG	EED4C86CDF8DFC360F2EDF33E5720CED
7.	2.png	36		FC793E3A2AF263325A4221D470C7DBBE
8.	3.png	39		7FF8A428A24A7B4AE29A8CA88FEBC442
9.	4.png	32		ECC9DDE59BEDED0B4D47B1A1AFAD448

10.	5.png	37		93894E4A8765309622 A6E04E8FDA9ABB
11.	1.psd	289	PSD	A2EF0A99C44A529C 9352C0367976CD90
12.	2.psd	296		19B688DB9C69D6CD 29141F214ECB3772
13.	3.psd	299		28EA7FA6995AD6A4 DF9F93AE88FB33B4
14.	4.psd	297		76B53C8CCE990EBC A3726D36EADA0C69
15.	5.psd	297		C54C57FB3EA41DC8 9C009669973E0910
16.	1.tif	6515	TIFF	9FEEE1A77FA56D0F 93DF2B509E38D813
17.	2.tif	6516		4734AB673577722A3 F4264FBBB6346B7
18.	3.tif	6517		14660DE11D84557E9 D031DEB8CA17837
19.	4.tif	6516		AD84F2108EC8775E3 505EF3BB82353AB
20.	5.tif	6515		1ECC7FF955AEA8B2 82AB587778439438
21.	1.bmp	6329	BMP	834BA4EF0F0D1AEA 02DC2773A3008785
22.	2.bmp	6329		15B52C89B4608DB7C 23263E5BE559B53
23.	3.bmp	6329		7544F9552E73C1390E 17F0B3BCD08F4C
24.	4.bmp	6329		A6234C1CC51EAC10 13A1C49402A3D311
25.	5.bmp	6329		709A2A837C79F68A6 BF01151592C86EB

## **7. REQUIRED TOOLS & MODULES**

### **7.1 Metadata**

“Metadata is "data that provides information about other data". Metadata has various purposes. It helps users find appropriate data and discover properties. It supports unite electronic resources, make available digital proof of identity, and record and store resources. It can provide the file information regarding the date of creation name and type of that file and the size along with the other details”[10].

### **7.2 EXIF**

“Exchangeable-Image-File-Format (EXIF) is a standard which specific information[metadata] of an image. It can store such important data like camera exposure, date/time the image was captured, and even GPS location”[8].

### **7.3 Python**

Python is a high level and indentation sensitive, a user-friendly programming language by Guido van Rossum. Python language has the object-oriented technique that will support programmers to write pure, logical code for small and large-scale tasks and it will support MD5 hash with help of the hashlib module by encoding the digest into base64.

### **7.4 PyPillow**

PIL abbreviated as Python-Imaging-Library (in latest releases known as Pillow) is a free open-source supplementary library for the Python language which helps to open, manipulate and save the many image file extensions.

### **7.5 Hashlib**

The Hashlib module provides the same interface to many unlike secure hashes & message digest algorithms.hashlib.MD5 () is the constructor for converting the hash digest to md5.



## 7.6 ExifTool

```

D:\cyber_forensics\MOBILE_FORENSICS\FINAL_COMPLEX_HASHING\COMPLEX_HASHING\exiftool(-k).exe
NAME
    exiftool - Read and write meta information in files

RUNNING IN WINDOWS
    Drag and drop files or folders onto the exiftool executable to display
    meta information, or rename to "exiftool.exe" and run from the command
    line to access all exiftool features.

    This stand-alone Windows version allows simple command-line options to
    be added to the name of the executable (in brackets and separated by
    spaces at the end of the name), providing a mechanism to use options
    when launched via the mouse. For example, changing the executable name
    to "exiftool(-a -u -g1 -w txt).exe" gives a drag-and-drop utility which
    generates sidecar ".txt" files with detailed meta information. As
    shipped, the -k option is added to cause exiftool to pause before
    terminating (keeping the command window open). Options may also be added
    to the "Target" property of a Windows shortcut to the executable.

SYNOPSIS
    Reading
        exiftool [*OPTIONS*] [-*TAG*...] [--*TAG*...] *FILE*...

    Writing
        exiftool [*OPTIONS*] -*TAG*[+<]=[*VALUE*]... *FILE*...

    Copying
        exiftool [*OPTIONS*] -tagsFromFile *SRCFILE* [-*SRCTAG*[>]*DSTTAG*...]
        *FILE*...

-- More --

```

Fig. 2. Free Open Source ExifTool

ExifTool is a free open-source tool that is a software program that works in the command line to read, write, and manipulate the given image, audio, video, & PDF metadata. ExifTool forms its specific metadata structure as output. This tool is designed to collect the meta-information from the sources, in binary or text format, and package it together by any kind of file. That can be one file, or existing data, or used as a backup file, carrying.

## 8. PROPOSED SYSTEM

$$N\_H = [MD5(A\_D) + MD5(M\_D)]$$

Fig. 3. Formula for Complex MD5 hash

N\_H is the new hash generated by Actual data (A\_D) and Metadata (M\_D).

- Based on the metadata of the given file the hash will be generated.

This method takes all the metadata information like file name, date of creation, modified date, etc. of the given input file. Here in this proposed model, one part of the hash will contain the MD5 hash, and the second half will contain the metadata information which becomes easy whenever the file gets modified.

### 8.1 Design

1. Install the latest version of python
2. Write a script to md5 hash for an image
3. Extract the metadata from the image
4. Hash the required and unique metadata
5. Join both hashes to form New Proposed hash

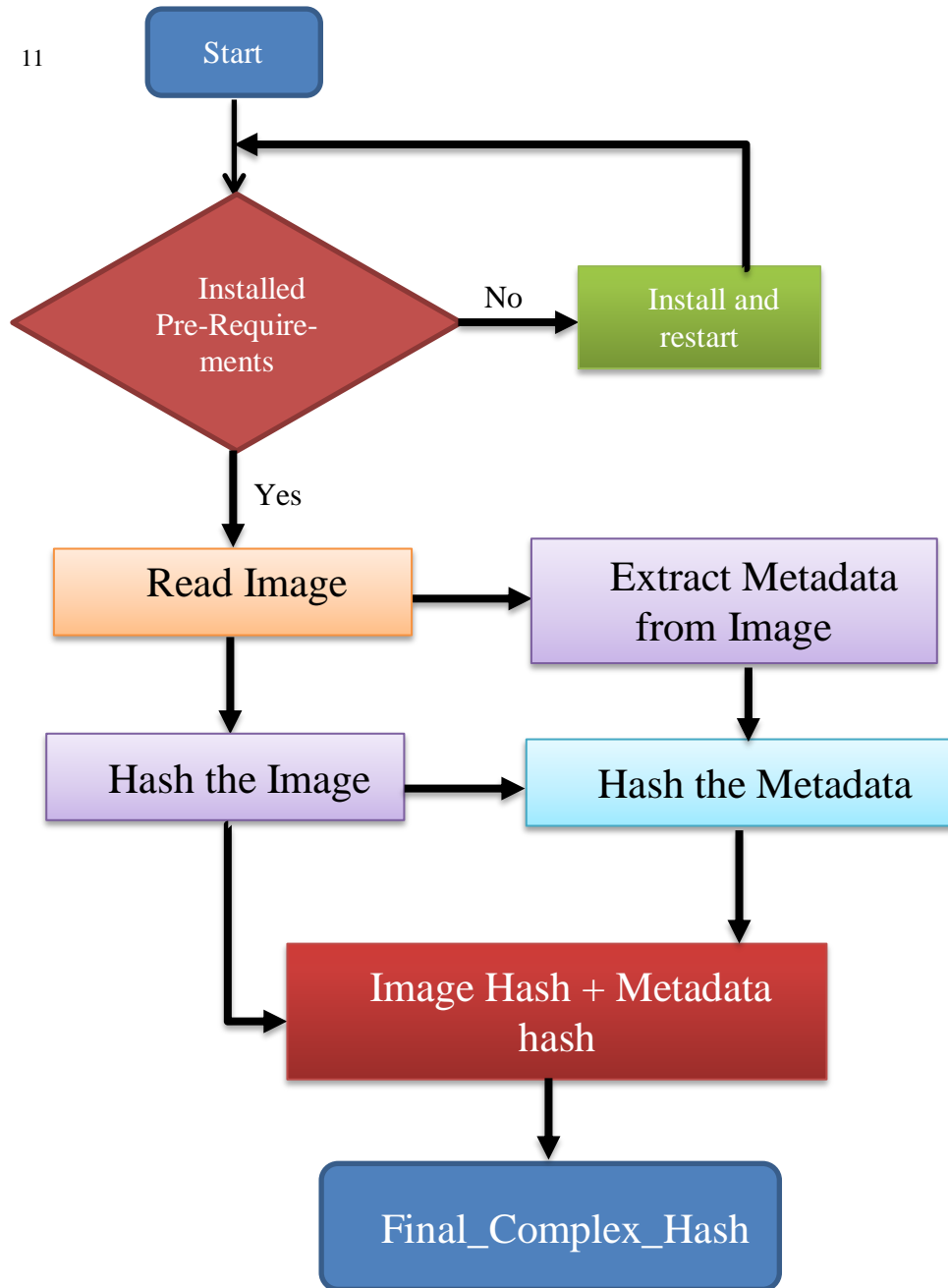
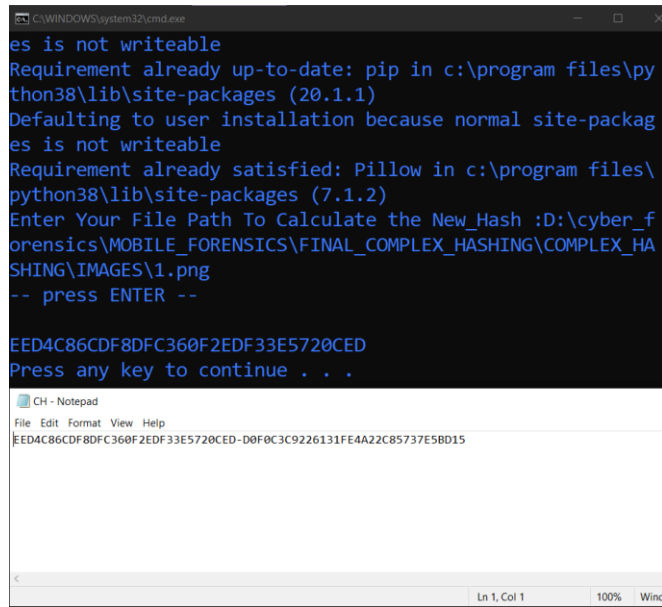


Fig. 4. Flowchart for the proposed hash

## 9. RESULTS



```

C:\WINDOWS\system32\cmd.exe
es is not writeable
Requirement already up-to-date: pip in c:\program files\python38\lib\site-packages (20.1.1)
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: Pillow in c:\program files\python38\lib\site-packages (7.1.2)
Enter Your File Path To Calculate the New_Hash :D:\cyber_forensics\MOBILE_FORENSICS\FINAL_COMPLEX_HASHING\COMPLEX_HASHING\IMAGES\1.png
-- press ENTER --

EED4C86CDF8DFC360F2EDF33E5720CED
Press any key to continue . . .

```

CH - Notepad

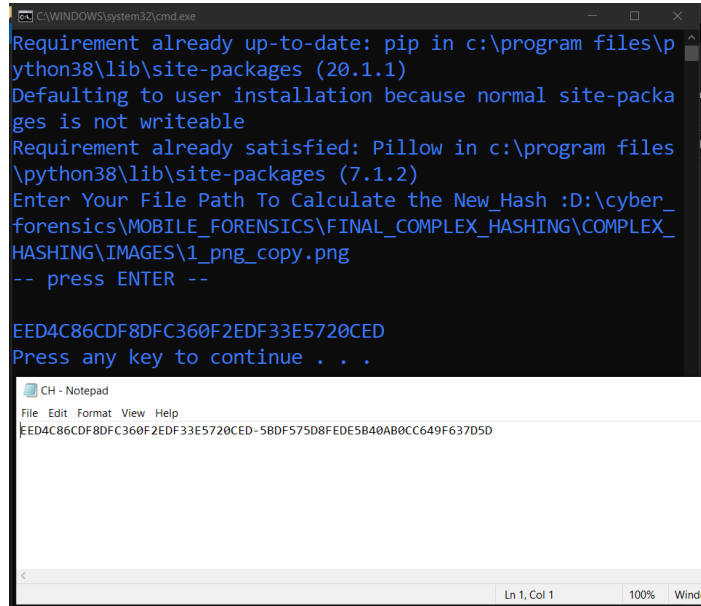
```

File Edit Format View Help
EED4C86CDF8DFC360F2EDF33E5720CED-D0F0C3C9226131FE4A22C85737E5BD15

```

Ln 1, Col 1 100% Wind

Fig. 5. New Hash for image 1



```

C:\WINDOWS\system32\cmd.exe
Requirement already up-to-date: pip in c:\program files\python38\lib\site-packages (20.1.1)
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: Pillow in c:\program files\python38\lib\site-packages (7.1.2)
Enter Your File Path To Calculate the New_Hash :D:\cyber_forensics\MOBILE_FORENSICS\FINAL_COMPLEX_HASHING\COMPLEX_HASHING\IMAGES\1_png_copy.png
-- press ENTER --

EED4C86CDF8DFC360F2EDF33E5720CED
Press any key to continue . . .

```

CH - Notepad

```

File Edit Format View Help
EED4C86CDF8DFC360F2EDF33E5720CED-5BDF575D8FEDE5B40AB0CC649F637D5D

```

Ln 1, Col 1 100% Wind

Fig. 6. New Hash for image 1 copy

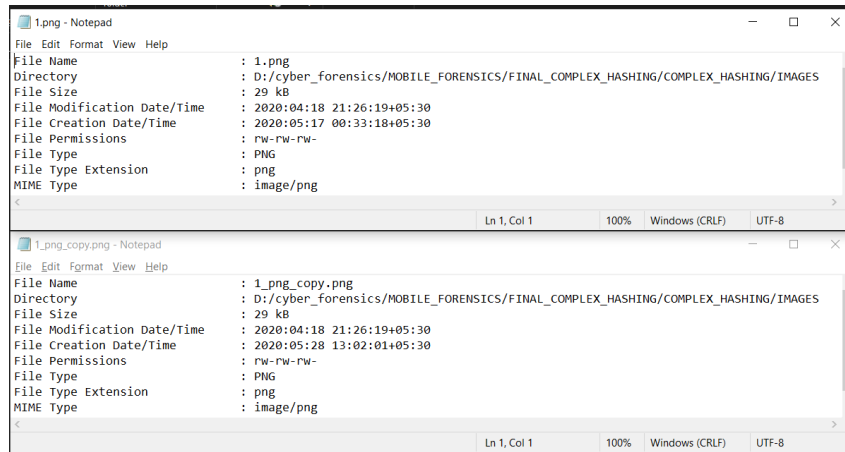


Fig. 7. Comparing Metadata of image1 & image1 copy

#### AFTER APPLYING THE PROPOSED MODEL

- This hash is created using the existing hashing technique but created with a unique mechanism that will make the forensics on the image file easy and it has a lot of uses apart from the old hashing mechanisms
- This hashing is collision free hashing technique
- The hash created by this process will be used as the mobile forensic purpose but not checksum where it will not same even though the data is similar and this can be used to identify the duplicate image files.
- NOTE : Hash varies with the device. This hashing technique will give different hashes for same image in different devices to reach the forensic requirements.

SNo	File name	File size (in KB)	File format	MD5 Hash	PROPOSED HASH
1.	1.jpg	158	JPG	2E03C3D6AC7D0E4E59707F1BFA36D64B	2E03C3D6AC7D0E4E59707F1BFA36D64B-316D9A7DF3D08AECDC9BF00528CF1246
2.	2.jpg	169		3FF72695260722D20A148D0745E8AF9D	3FF72695260722D20A148D0745E8AF9D-50B93ADBDF957F51935E872D87D3600B
3.	3.jpg	171		B64F9AE9703E4C7F8CAC9B302A91A8AA	B64F9AE9703E4C7F8CAC9B302A91A8AA-666F59052F5A2BFE7332F3D8D07C5D42

4.	4.jpg	164		5F1A3EF0CA8 22FC387F7862 E65D9321C	5F1A3EF 0CA822F C387F786 2E65D932 1C- 2BFBC2E CE89DEE 6C25A31 D144E452 953
5.	5.jpg	170		F8E12554E92E AB42A88DCD 88927BFFF6	F8E12554 E92EAB4 2A88DCD 88927BFF F6- 806CEBE 2EF920D6 56B3D1A 8EC33B1 A8A
6.	1.png	30	PNG	EED4C86CDF8 DFC360F2EDF 33E5720CED	EED4C86 CDF8DFC 360F2ED F33E5720 CED- D0F0C3C 9226131F E4A22C8 5737E5B D15
7.	2.png	36		FC793E3A2AF 263325A4221D 470C7DBBE	FC793E3 A2AF263 325A4221 D470C7D BBE- 5AF7F8C 9CBE920 F831DFD

					CE2B4D3 3D30
8.	3.png	39		7FF8A428A24 A7B4AE29A8 CA88FEBC442	7FF8A428 A24A7B4 AE29A8C A88FEBC 442- 9BAF066 EAE06E1 A2B5A14 44868010 53
9.	4.png	32		ECC9DDE59B EDED0B4D47 B1A1AFAD44 8	ECC9DD E59BEDE D0AB4D4 7B1A1AF AD448- DD89F70 B16CD32 B0853F07 0930A6F A83
10.	5.png	37		93894E4A8765 309622A6E04E 8FDA9ABB	93894E4A 87653096 22A6E04 E8FDA9A BB- 6273E1B D9A6414 0251158F 71F5141A C2
11.	1.psd	289	PSD		A2EF0A9 9C44A529 C9352C03 67976CD9 0-



				A2EF0A99C44 A529C9352C03 67976CD90	BABF194 BD83538 FEE662C 7EF496D3 3FA
12.	2.psd	296		19B688DB9C6 9D6CD29141F 214ECB3772	19B688D B9C69D6 CD29141 F214ECB 3772- 55BDD12 08331596 976D58C F3D7E632 46
13.	3.psd	299		28EA7FA6995 AD6A4DF9F93 AE88FB33B4	28EA7FA 6995AD6 A4DF9F9 3AE88FB 33B4- 9C9D35C CBDFB5E 2F6AD62 5608D8A 566F
14.	4.psd	297		76B53C8CCE9 90EBCA3726D 36EADA0C69	76B53C8 CCE990E BCA3726 D36EAD A0C69- 124211F3 937B49A B6476206 0685CE94 4
15.	5.psd	297			C54C57F B3EA41D

				C54C57FB3EA 41DC89C00966 9973E0910	C89C0096 69973E09 10- F108E89E 5349F9F5 33E84C24 BB66FF8 9
16.	1.tif	6515	TIF	9FEEE1A77FA 56D0F93DF2B 509E38D813	9FEEE1A 77FA56D 0F93DF2 B509E38 D813- 11684D7 DAF79EC 8C869FB3 1AF691D E61
17.	2.tif	6516		4734AB673577 722A3F4264FB BB6346B7	4734AB67 3577722A 3F4264FB BB6346B 7- BD7EAD FE1092E4 35834A5 A56676F3 5D0
18.	3.tif	6517		14660DE11D84 557E9D031DE B8CA17837	14660DE1 1D84557E 9D031DE B8CA178 37- 18F2579D DC999ED 9855F2FD

					8BEE724 A8
19.	4.tif	6516		AD84F2108EC 8775E3505EF3 BB82353AB	AD84F21 08EC8775 E3505EF3 BB82353 AB- 0F0F969B 4E75C2C 87E2F36E DDC7601 FD
20.	5.tif	6515		1ECC7FF955A EA8B282AB58 7778439438	1ECC7FF 955AEA8 B282AB5 87778439 438- 92904C0F B40AC4A FEA0B40 983A6BA 0B4
21.	1.bmp	6329		834BA4EF0F0 D1AEA02DC2 773A3008785	834BA4E F0F0D1A EA02DC2 773A3008 785- 5EDC8F4 2793B4F5 E3BF94E F4A1CFF 7C0
22.	2.bmp	6329	BMP		15B52C89 B4608DB 7C23263E 5BE559B 53-

				15B52C89B460 8DB7C23263E 5BE559B53	D4947D9 96E31E55 550601D1 D799F695 A
23.	3.bmp	6329		7544F9552E73 C1390E17F0B3 BCD08F4C	7544F955 2E73C139 0E17F0B3 BCD08F4 C- 4293FDB 5B47CC8 15E3F61A C2ECDD1 8AE
24.	4.bmp	6329		A6234C1CC51 EAC1013A1C4 9402A3D311	A6234C1 CC51EAC 1013A1C4 9402A3D 311- 62DBF36 012E5711 718BFA2 9743618E 4F
25.	5.bmp	6329		709A2A837C7 9F68A6BF0115 1592C86EB	709A2A8 37C79F68 A6BF011 51592C86 EB- FE12DD8 0976A69 D4F1198E 0E9254E DCB

## 10. FUTURE WORK

In this paper we used only image data it can be extended to the other data which may use all extensions in the mobile and using this hash will help to find out the original data to copied data and that can be for any file in a future enhancement.

## 11. CONCLUSION

This image hashing is different to a traditional hash where this can be used for Mobile forensic investigation to identify the original file and along with the location and identify the multiple files and to find original file so that copied files can be easily identified and this technique can also use to avoid hash collisions as it can also hash for the metadata which is not same for every file.

## 11. REFERENCES

- [1] [https://en.wikipedia.org/wiki/Mobile\\_device\\_forensics](https://en.wikipedia.org/wiki/Mobile_device_forensics)
- [2] <https://hub.packtpub.com/introduction-mobile-forensics>
- [3] <https://www.reddit.com/r/ZentaChain>
- [4] [https://subscription.packtpub.com/book/application\\_development/9781783288311/1/ch01lv11sec08/mobile-forensics](https://subscription.packtpub.com/book/application_development/9781783288311/1/ch01lv11sec08/mobile-forensics)
- [5] <https://d3pakblog.wordpress.com/2017/01/07/challenges-in-mobile-forensics>
- [6] [https://subscription.packtpub.com/book/networking\\_and\\_servers/9781786464200/1/ch01lv11sec10/mobile-forensics](https://subscription.packtpub.com/book/networking_and_servers/9781786464200/1/ch01lv11sec10/mobile-forensics)
- [7] A framework to (Im) Prove „Chain of Custody “in Digital Investigation Process
- [8] <https://photographylife.com/what-is-exif-data>
- [9] <https://www.smteto.com/hash/md5-hash-generator-online/>

## **12.APPENDIX**

We present some simplified code which is useful for understanding the paper.The link to access the code is given below:

[https://github.com/yaswanth0212/COMPLEX\\_HASHING\\_WITH\\_METADATA](https://github.com/yaswanth0212/COMPLEX_HASHING_WITH_METADATA)