Digital Evidence Collection & Preservation Procedures

ROLE OF METADATA (Hashing)

NULL - OWASP Coimbatore Chapter
6th July 2024

ASHOK KUMAR MOHAN
Founder & Director
@ KRIYAVAN Cyber Forensic Service,
Mdu, TN, IN

Digital Evidence Collection & Preservation Procedures

ROLE OF METADATA (Hashing)

NULL - OWASP Coimbatore Chapter

6th July 2024

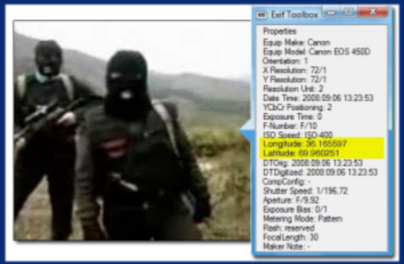
ASHOK KUMAR MOHAN

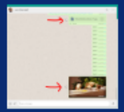
Founder & Director
@ KRIYAVAN Cyber Forensic Service,
Mdu, TN, IN

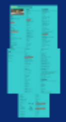
Metadata

“**data about data**” is Metadata

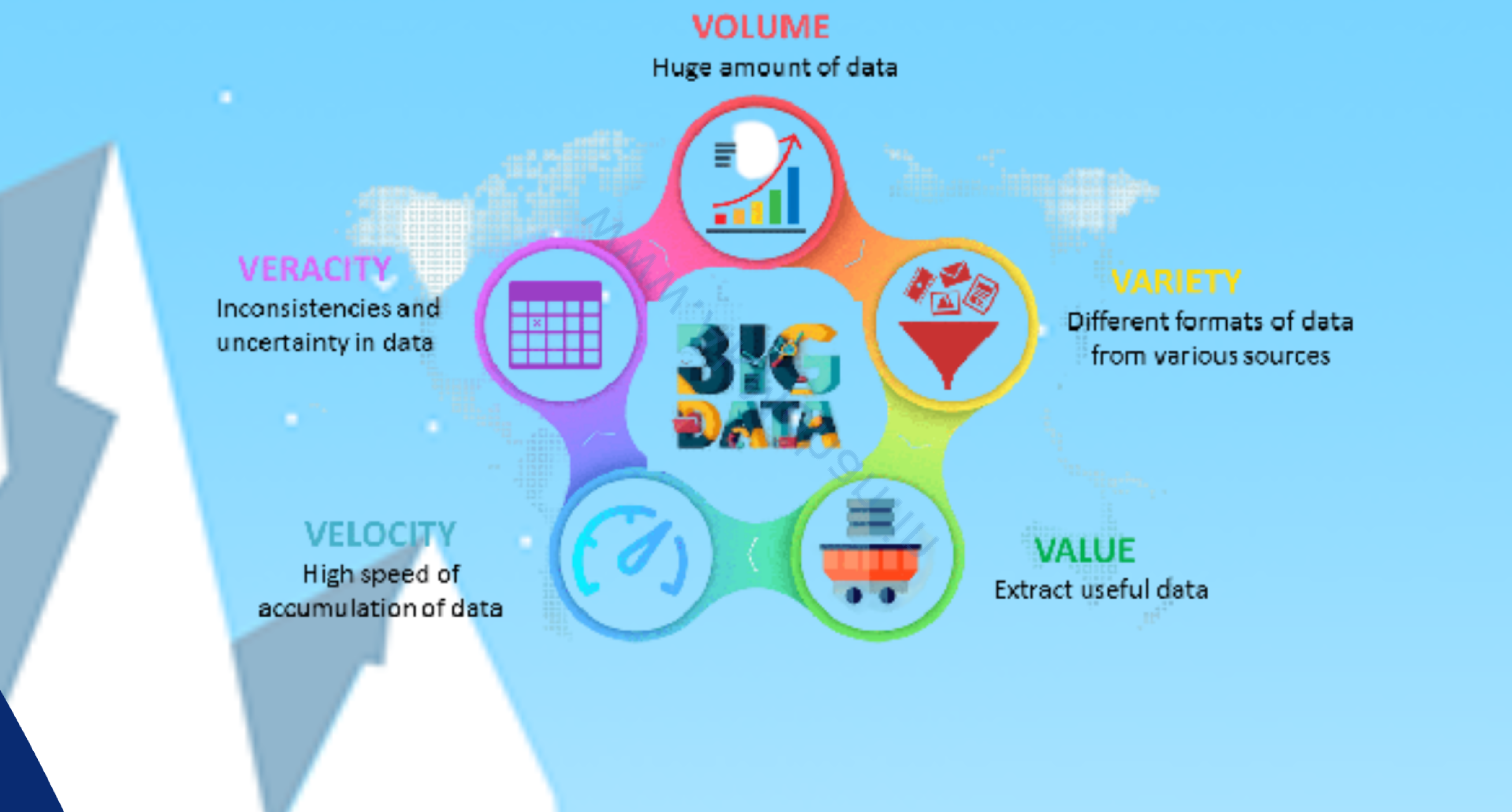
Exif - Exchangeable image file format
(Modified Accessed Created times)

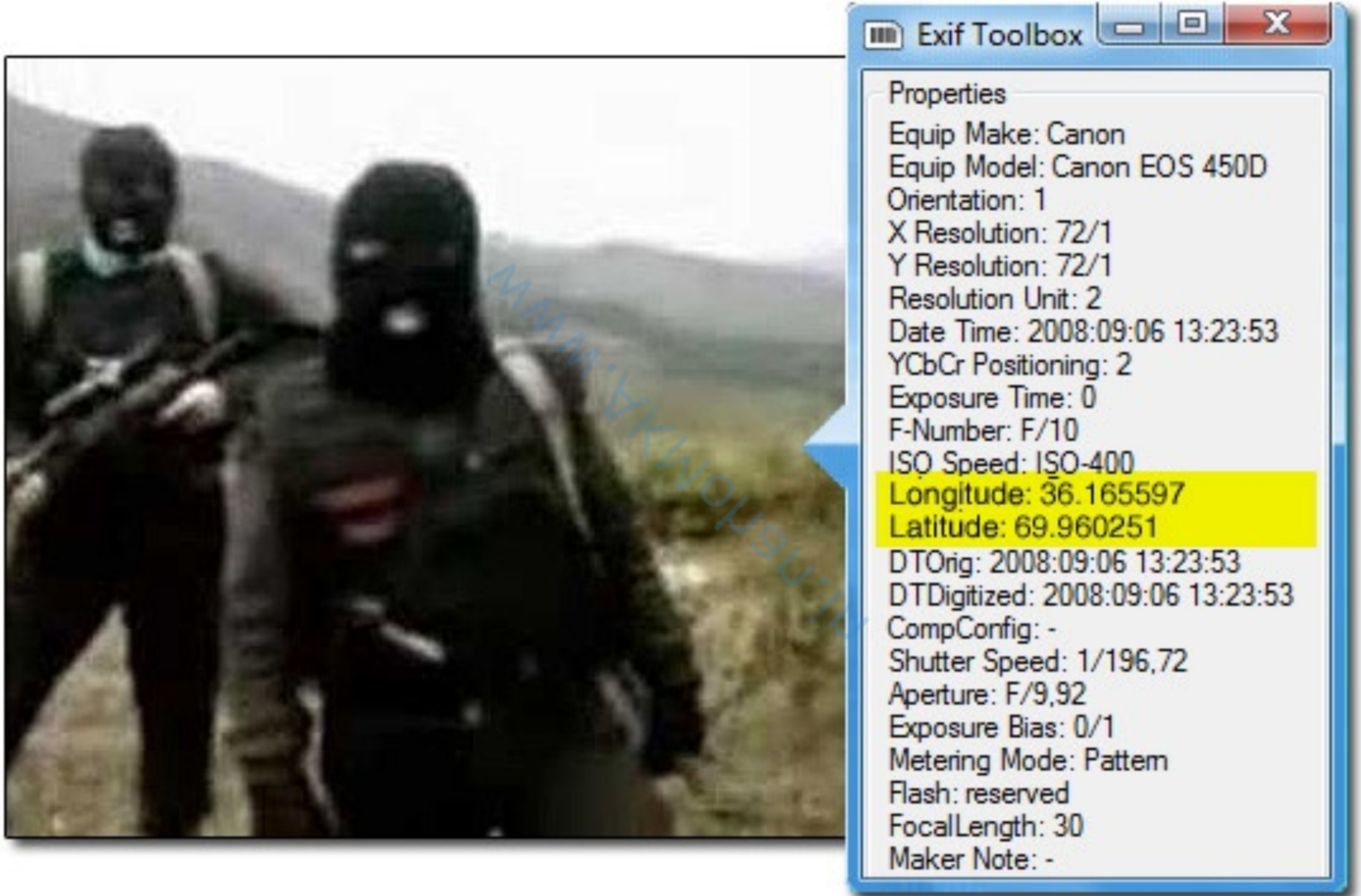


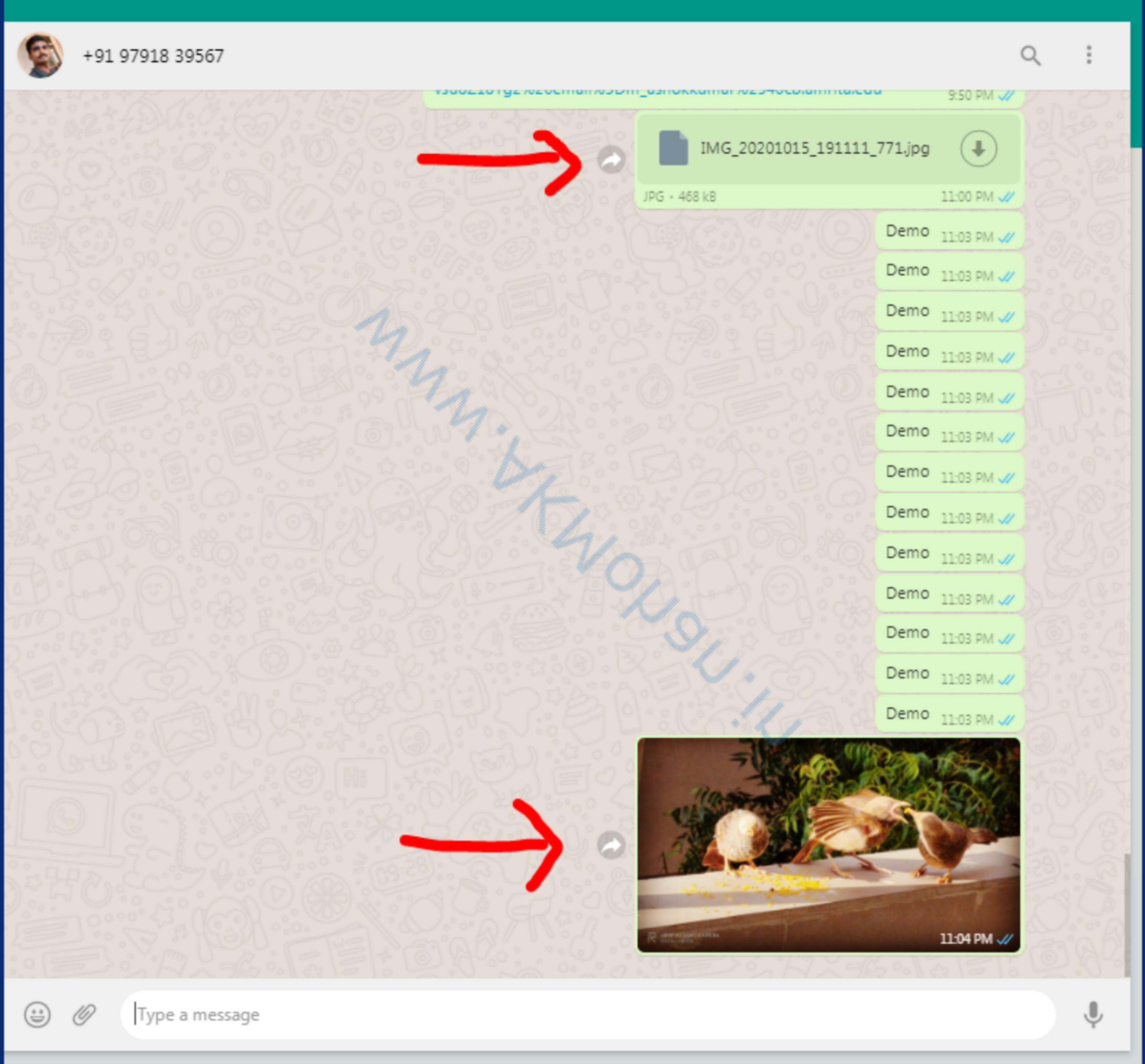




SRC: http://www.forensichandbook.com/catching-criminals-with-digital-photos/


VOLUME
Huge amount of data
VARIETY
Different formats of data from various sources
VALUE
Extract useful data
VELOCITY
High speed of accumulation of data
VERACITY
Inconsistencies and uncertainty in data
BIG DATA


Exif Toolbox
Properties
Equip Make: Canon
Equip Model: Canon EOS 450D
Orientation: 1
X Resolution: 72/1
Y Resolution: 72/1
Resolution Unit: 2
Date Time: 2008:09:06 13:23:53
YCbCr Positioning: 2
Exposure Time: 0
F-Number: F/10
ISO Speed: ISO-400
Longitude: 36.165597
Latitude: 69.960251
DTOrig: 2008:09:06 13:23:53
DTDigitized: 2008:09:06 13:23:53
CompConfig: -
Shutter Speed: 1/196,72
Aperture: F/9,92
Exposure Bias: 0/1
Metering Mode: Pattern
Flash: reserved
FocalLength: 30
Maker Note: -


+91 97918 39567
9:50 PM
IMG_20201015_191111_771.jpg
JPG • 468 kB
11:00 PM
Demo 11:03 PM
Demo 11:03 PM
Demo 11:03 PM
Demo 11:03 PM
Demo 11:03 PM
Demo 11:03 PM
Demo 11:03 PM
Demo 11:03 PM
Demo 11:03 PM
Demo 11:03 PM
Demo 11:03 PM
Demo 11:03 PM
Demo 11:03 PM
11:04 PM
Type a message

Exif Info: IMG_20201015_191111_771.jpg

FILE added as DOC (WhatsApp)

Around **30** Metadata Features

File

Filename
IMG_20201015_191111_771.jpg

File Size
468 kB

File Type
JPEG

File Type Extension
jpg

MIME Type
image/jpeg

Exif Byte Order
Big-endian (Motorola, MM)

Image Width
3303

Image Height
1840

Encoding Process
Progressive DCT, Huffman coding

Bits Per Sample
8

Color Components
3

Y Cb Cr Sub Sampling
YCbCr4:2:0 (2 2)

EXIF

Image Width
3303

Image Height
1840

Light Source
Unknown

Orientation
Unknown (0)

GPS Latitude
10 deg 54' 9.00"

GPS Latitude Ref
North

GPS Longitude Ref
East

GPS Longitude
76 deg 53' 49.00"

JFIF

JFIF Version
1.01

Resolution Unit
None

X Resolution
1

Y Resolution
1

Composite

Image Size
3303x1840

Megapixels
6.1

GPS Latitude
10 deg 54' 9.00" N

GPS Longitude
76 deg 53' 49.00" E

GPS Position
10 deg 54' 9.00" N, 76 deg 53' 49.00" E

Exif Info: WhatsApp Image 2020-10-15 at 11.03.11 PM.jpeg

FILE added as IMAGE (WhatsApp)

Around **15** Metadata Features

File

Filename
WhatsApp Image 2020-10-15 at 11.03.11 PM.jpeg

File Size
103 kB

File Type
JPEG

File Type Extension
jpg

MIME Type
image/jpeg

Image Width
1280

Image Height
712

Encoding Process
Progressive DCT, Huffman coding

Bits Per Sample
8

Color Components
3

Y Cb Cr Sub Sampling
YCbCr4:2:0 (2 2)

JFIF

JFIF Version
1.01

Resolution Unit
None

X Resolution
1

Y Resolution
1

Composite

Image Size
1280x712

Megapixels
0.911

Exif?

Exif Info: IMG_20201015_191111_771.jpg

FILE added as DOC (WhatsApp)

Around **30** Metadata Features

File

Filename
IMG_20201015_191111_771.jpg

File Size
468 kB

File Type
JPEG

File Type Extension
jpg

MIME Type
image/jpeg

Exif Byte Order
Big-endian (Motorola, MM)

Image Width
3303

Image Height
1840

Encoding Process
Progressive DCT, Huffman coding

Bits Per Sample
8

Color Components
3

Y Cb Cr Sub Sampling
YCbCr4:2:0 (2 2)

EXIF

Image Width
3303

Image Height
1840

Light Source
Unknown

Orientation
Unknown (0)

GPS Latitude
10 deg 54' 9.00"

GPS Latitude Ref
North

GPS Longitude Ref
East

GPS Longitude
76 deg 53' 49.00"

JFIF

JFIF Version
1.01

Resolution Unit
None

X Resolution
1

Y Resolution
1

Composite

Image Size
3303x1840

Megapixels
6.1

GPS Latitude
10 deg 54' 9.00" N

GPS Longitude
76 deg 53' 49.00" E

GPS Position
10 deg 54' 9.00" N, 76 deg 53' 49.00" E

Exif Info: WhatsApp Image 2020-10-15 at 11.03.11 PM.jpeg

FILE added as IMAGE (WhatsApp)

Around **15** Metadata Features

File

Filename
WhatsApp Image 2020-10-15 at 11.03.11 PM.jpeg

File Size
103 kB

File Type
JPEG

File Type Extension
jpg

MIME Type
image/jpeg

Image Width
1280

Image Height
712

Encoding Process
Progressive DCT, Huffman coding

Bits Per Sample
8

Color Components
3

Y Cb Cr Sub Sampling
YCbCr4:2:0 (2 2)

JFIF

JFIF Version
1.01

Resolution Unit
None

X Resolution
1

Y Resolution
1

Composite

Image Size
1280x712

Megapixels
0.911

Exif?

53' 49.00" E

Exif Info: IMG20201015073702.jpg

File

Filename
IMG20201015073702.jpg

File Size
1198 kB

File Type
JPEG

File Type Extension
jpg

MIME Type
image/jpeg

Exif Byte Order
Little-endian (Intel, II)

Image Width
4000

Image Height
1840

Encoding Process
Baseline DCT, Huffman coding

Bits Per Sample
8

Color Components
3

Y Cb Cr Sub Sampling
YCbCr4:2:0 (2 2)

JFIF

EXIF

Image Width
4000

Image Height
1840

Make
realme

Camera Model Name
realme 5 Pro

Orientation
Horizontal (normal)

X Resolution
72

Y Resolution
72

Resolution Unit
inches

Modify Date
2020:10:15 07:37:02

Y Cb Cr Positioning
Centered

Interoperability Index
Unknown ()

Interoperability Version

Exposure Time
1/661

F Number
1.8

Exposure Program
Not Defined

ISO
180

ICC_Profile

Profile CMM Type
Apple Computer Inc.

Profile Version
4.0.0

Profile Class
Display Device Profile

Color Space Data
RGB

Profile Connection Space
XYZ

Profile Date Time
2018:06:24 13:22:32

Profile File Signature
acsp

Primary Platform
Apple Computer Inc.

CMM Flags
Not Embedded, Independent

Device Manufacturer
Unknown (OPPO)

Device Model

Device Attributes
Reflective, Glossy, Positive, Color

Rendering Intent
Perceptual

Connection Space Illuminant
0.9642 1 0.82491

Profile Creator
Apple Computer Inc.

Profile ID
0

Y Cb Cr Sub Sampling
YCbCr4:2:0 (2 2)

JFIF

JFIF Version
1.01

Resolution Unit
None

X Resolution
1

Y Resolution
1

Exposure Program
Not Defined

ISO
180

Exif Version
0210

Date/Time Original
2020:10:15 07:37:02

Create Date
2020:10:15 07:37:02

Components Configuration
Y, Cb, Cr, -

Shutter Speed Value
1

Aperture Value
1.5

Brightness Value
undef

Exposure Compensation
0

Max Aperture Value
1.0

Metering Mode
Unknown

Flash
Off, Did not fire

Focal Length
4.7 mm

User Comment
oppo_0

Sub Sec Time
734000

Sub Sec Time Original
734000

Sub Sec Time Digitized
734000

Profile Creator
Apple Computer Inc.

Profile ID
0

Profile Description
Display P3

Profile Copyright
Copyright Apple Inc., 2017

Media White Point
0.95045 1 1.08905

Red Matrix Column
0.51512 0.2412 -0.00105

Green Matrix Column
0.29198 0.69225 0.04189

Blue Matrix Column
0.1571 0.06657 0.78407

Red Tone Reproduction Curve
[binary data]

Chromatic Adaptation
1.04788 0.02292 -0.0502 0.02959 0.99048
-0.01706 -0.00923 0.01508 0.75168

Blue Tone Reproduction Curve
[binary data]

Green Tone Reproduction Curve
[binary data]

Composite

Aperture
1.8

Image Size
4000x1840

Megapixels
7.4

Shutter Speed
1/661

Create Date
2020:10:15 07:37:02.734000

User Comment
oppo_0

Sub Sec Time
734000

Sub Sec Time Original
734000

Sub Sec Time Digitized
734000

Flashpix Version
0100

Color Space
Uncalibrated

Exif Image Width
0

Exif Image Height
0

Sensing Method
Unknown (0)

Scene Type
Unknown (0)

Original IMAGE
from Mobile

Around
90
Metadata
Features

Image Size
4000x1840

Megapixels
7.4

Shutter Speed
1/661

Create Date
2020:10:15 07:37:02.734000

Date/Time Original
2020:10:15 07:37:02.734000

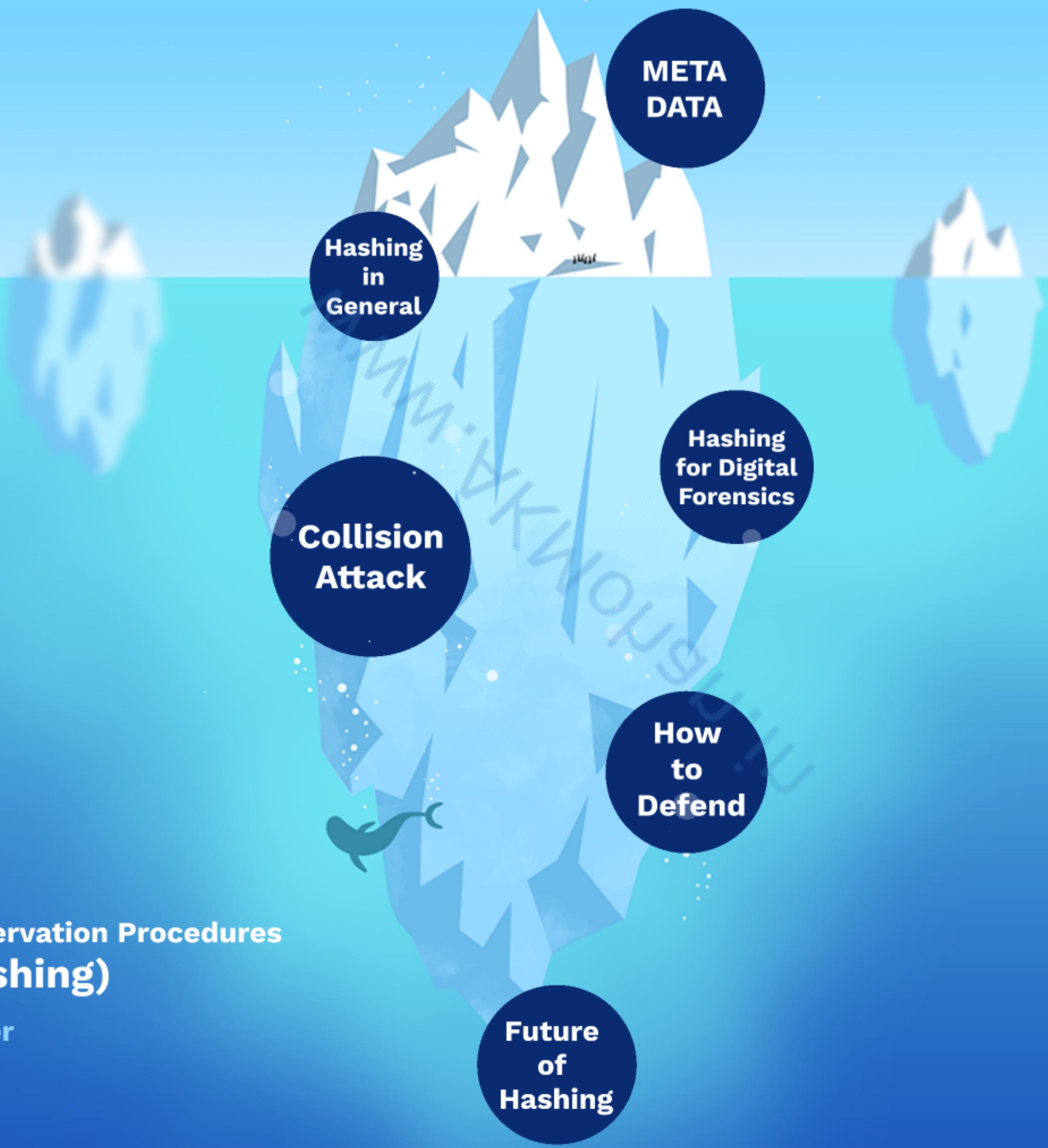
Modify Date
2020:10:15 07:37:02.734000

GPS Latitude

GPS Longitude

Focal Length
4.7 mm

Light Value
10.2

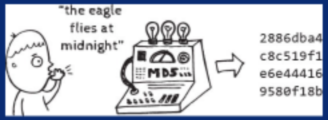


Digital Evidence Collection & Preservation Procedures

ROLE OF METADATA (Hashing)

NULL - OWASP Coimbatore Chapter
6th July 2024

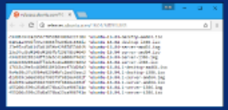
ASHOK KUMAR MOHAN
Founder & Director
@ KRIYAVAN Cyber Forensic Service,
Mdu, TN, IN



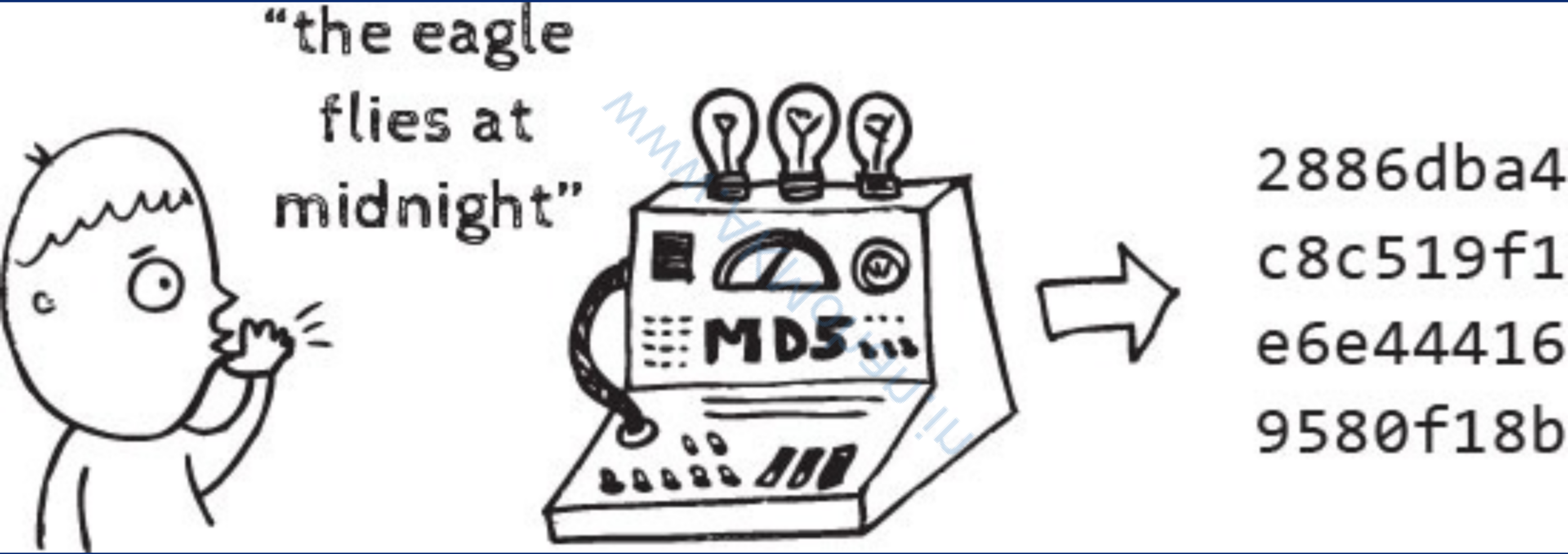
HASHING

A hash function is any function that can be used to **map data of arbitrary size to data of fixed size**

- Uses **Crypto**graphic Algorithms (MD5, SHA, ...)
- **Variable** length Input = **Constant** length Output
- **Irreversible** (only for checking the Integrity)



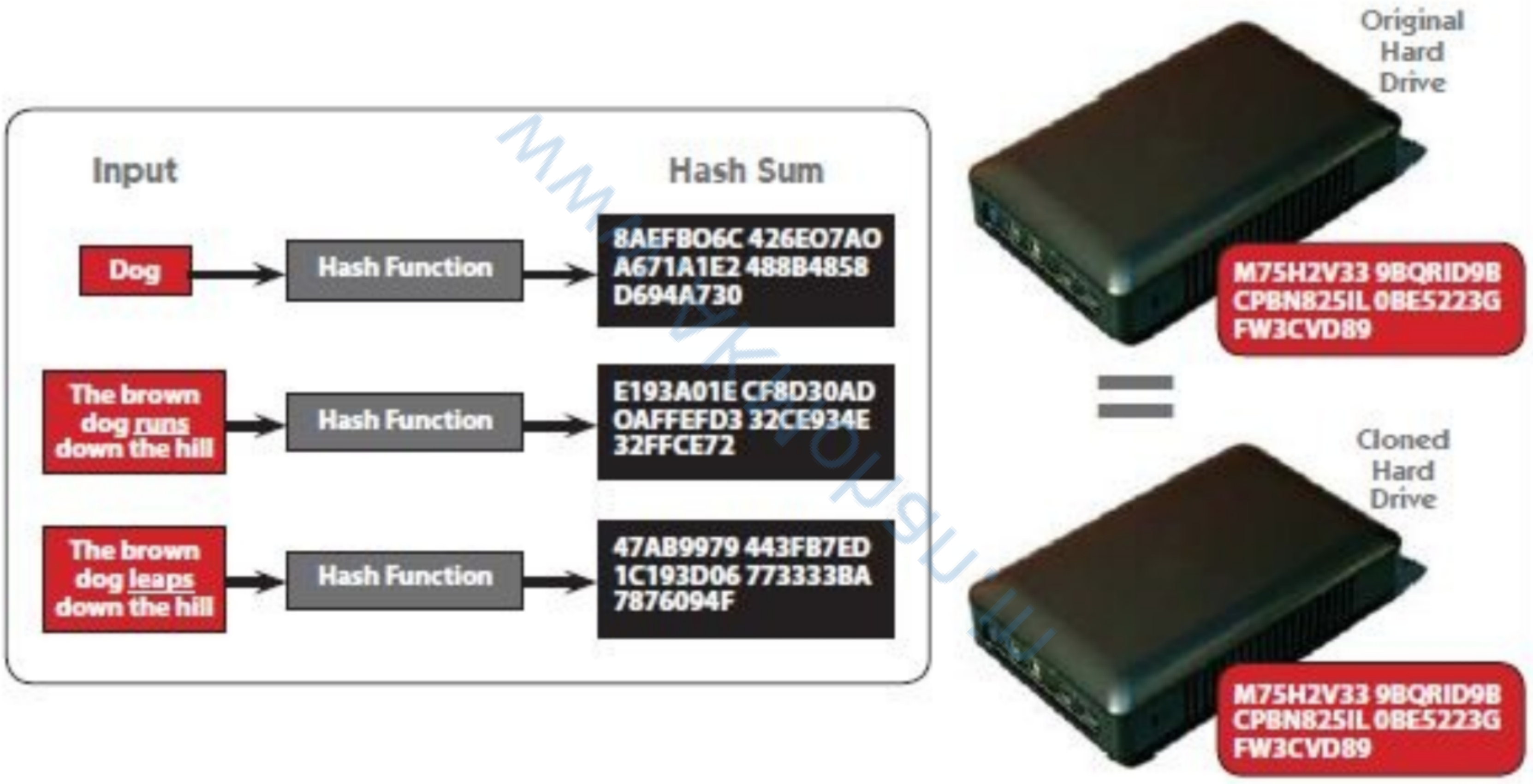
SRC: http://www.pinpointlabs.com/wordpress/tag/hash-verification/
http://www.softpedia.com/get/Security/Security-Related/MD5-and-SHA-1-Checksum-Utility.shtml


"the eagle flies at midnight"
MD5...
2886dba4
c8c519f1
e6e44416
9580f18b

HASHING

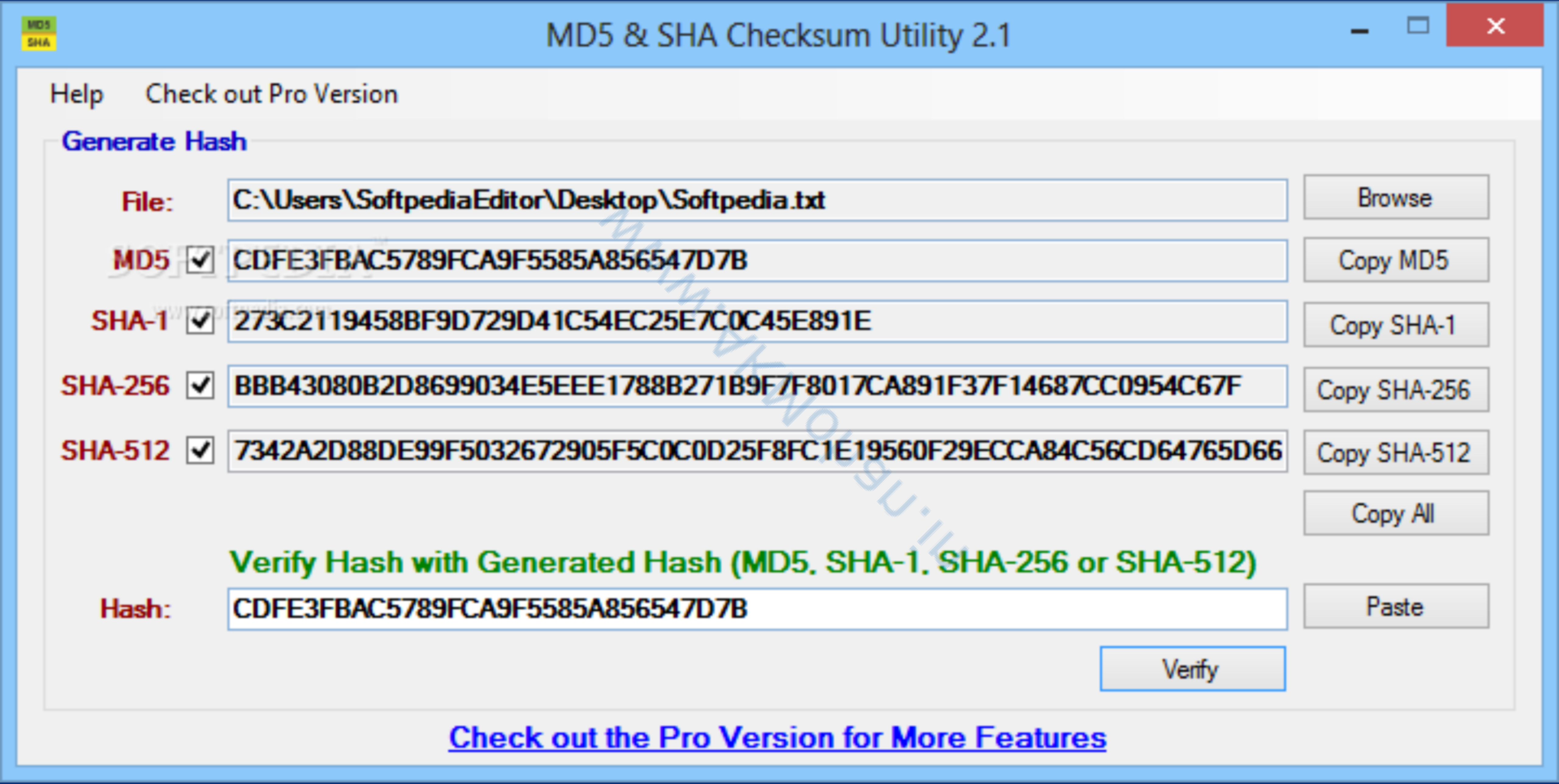
How Are Evidence Copies Verified?

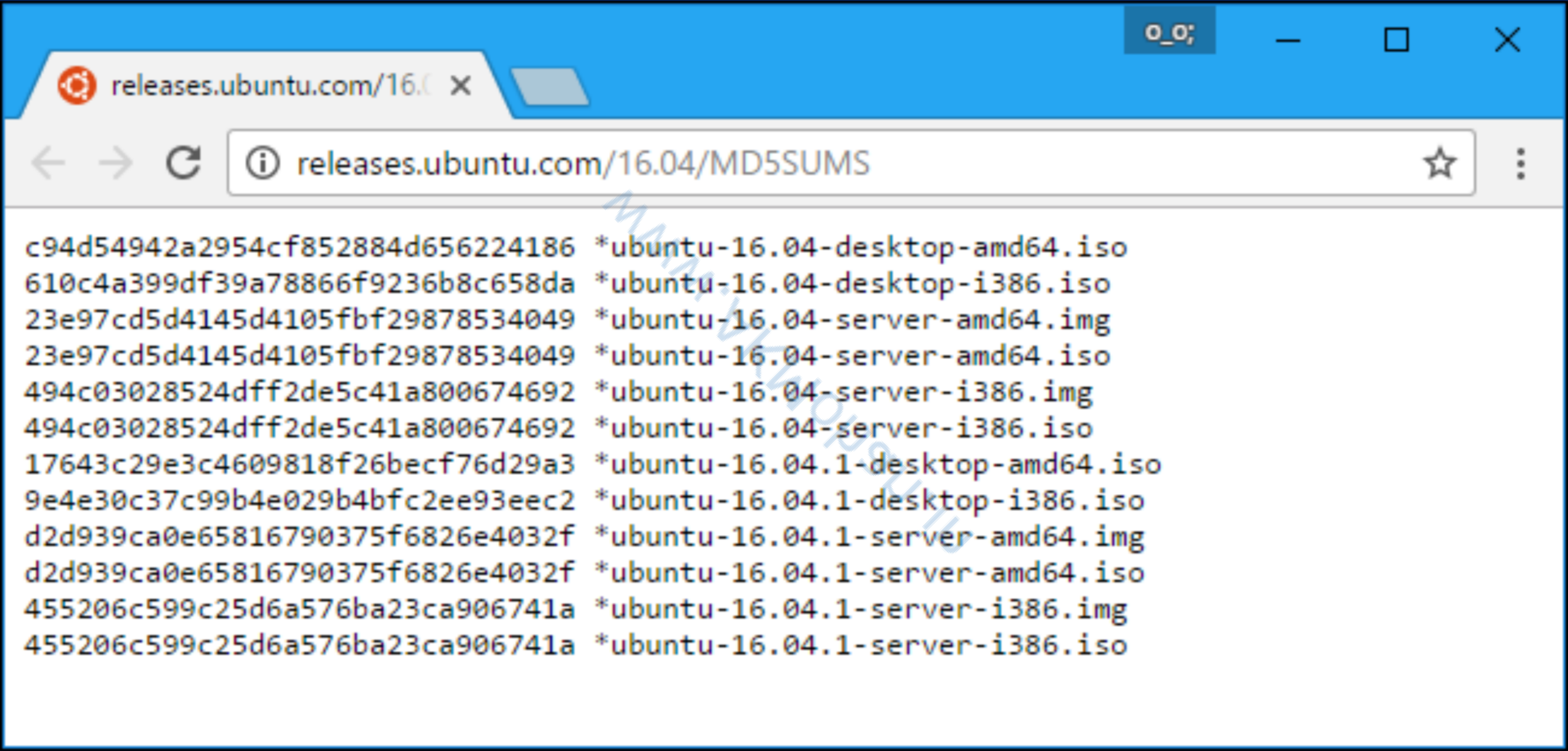
The Hash Value (Thumbprint) of the Source and Copied Data are Compared

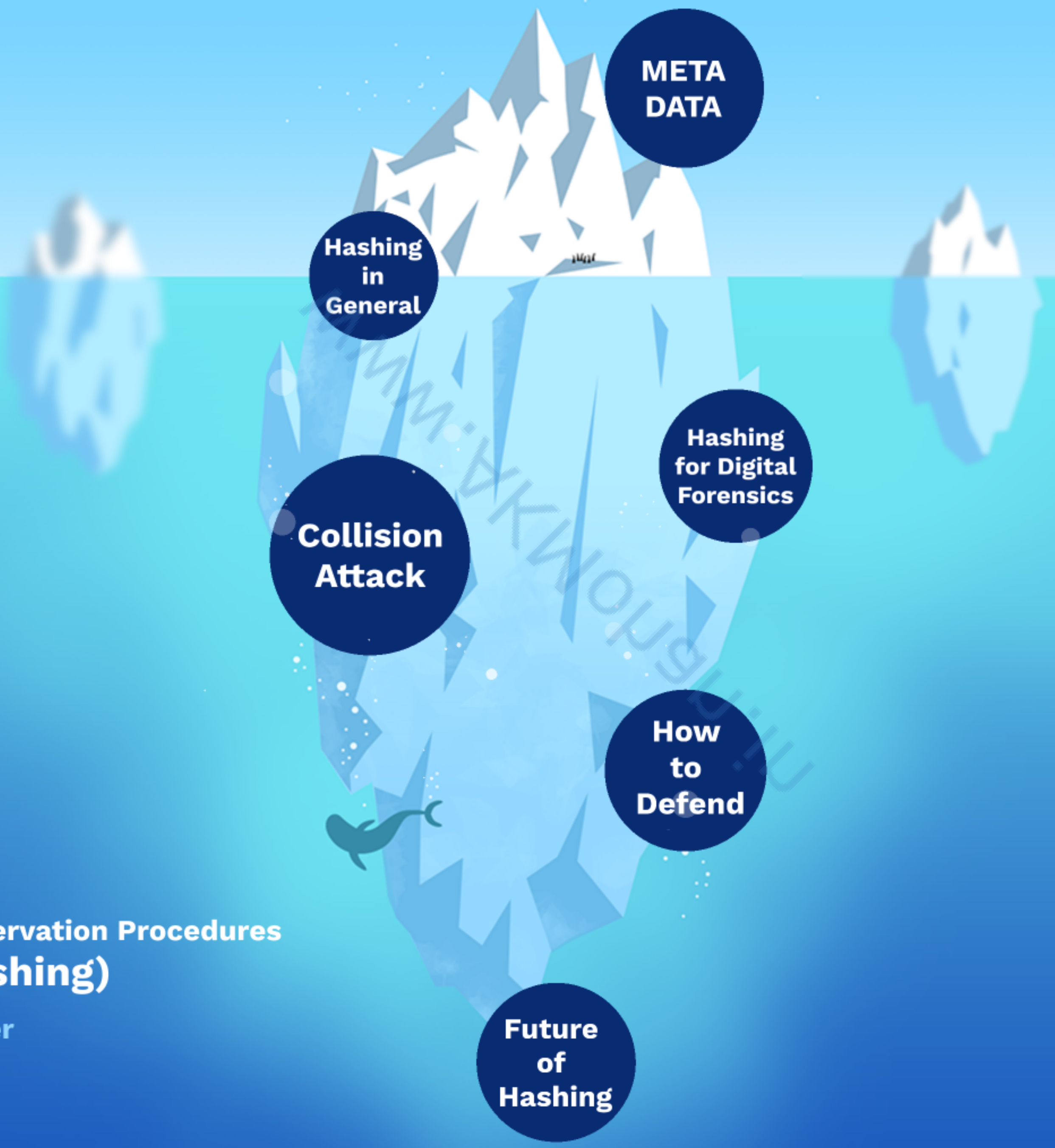


PINPOINT LABORATORIES

www.pinpointlabs.com


MD5 SHA
MD5 & SHA Checksum Utility 2.1
Help
Check out Pro Version
Generate Hash
File:
C:\Users\SoftpediaEditor\Desktop\Softpedia.txt
Browse
MD5
CDFE3FBAC5789FCA9F5585A856547D7B
Copy MD5
SHA-1
273C2119458BF9D729D41C54EC25E7C0C45E891E
Copy SHA-1
SHA-256
BBB43080B2D8699034E5EEE1788B271B9F7F8017CA891F37F14687CC0954C67F
Copy SHA-256
SHA-512
7342A2D88DE99F5032672905F5C0C0D25F8FC1E19560F29ECCA84C56CD64765D66
Copy SHA-512
Copy All
Verify Hash with Generated Hash (MD5, SHA-1, SHA-256 or SHA-512)
Hash:
CDFE3FBAC5789FCA9F5585A856547D7B
Paste
Verify
Check out the Pro Version for More Features


releases.ubuntu.com/16.0
o_o;
releases.ubuntu.com/16.04/MD5SUMS
c94d54942a2954cf852884d656224186 *ubuntu-16.04-desktop-amd64.iso
610c4a399df39a78866f9236b8c658da *ubuntu-16.04-desktop-i386.iso
23e97cd5d4145d4105fbf29878534049 *ubuntu-16.04-server-amd64.img
23e97cd5d4145d4105fbf29878534049 *ubuntu-16.04-server-amd64.iso
494c03028524dff2de5c41a800674692 *ubuntu-16.04-server-i386.img
494c03028524dff2de5c41a800674692 *ubuntu-16.04-server-i386.iso
17643c29e3c4609818f26becf76d29a3 *ubuntu-16.04.1-desktop-amd64.iso
9e4e30c37c99b4e029b4bfc2ee93eec2 *ubuntu-16.04.1-desktop-i386.iso
d2d939ca0e65816790375f6826e4032f *ubuntu-16.04.1-server-amd64.img
d2d939ca0e65816790375f6826e4032f *ubuntu-16.04.1-server-amd64.iso
455206c599c25d6a576ba23ca906741a *ubuntu-16.04.1-server-i386.img
455206c599c25d6a576ba23ca906741a *ubuntu-16.04.1-server-i386.iso



Digital Evidence Collection & Preservation Procedures

ROLE OF METADATA (Hashing)

NULL - OWASP Coimbatore Chapter
6th July 2024

ASHOK KUMAR MOHAN
Founder & Director
@ KRIYAVAN Cyber Forensic Service,
Mdu, TN, IN

Steps in Digital Forensics

Seizure

Identify the probable Evidences

- Acquire appropriate Warrant from Law Enforcement Agencies

Seizure:

- **Mostly capturing forcefully**
- Collection and preservation of evidence

Steps in Digital Forensics

Seizure

Acquisition

Analysis

Reporting

Acquisition

Authentication

- write blockers
- **hashing**

Acquisition (owning)

- Creating duplicate/clone of evidence by "Imaging"

Steps in Digital Forensics

Seizure

Acquisition

Analysis

Reporting

Seizure
Acquisition
Analysis
Reporting

Analysis

- **Verification** (CoC)
- **Validation** (**MD5**)

Analysis = **search** and **recover** all* the hidden/deleted evidence

Steps in Digital Forensics

Seizure

Acquisition

Analysis

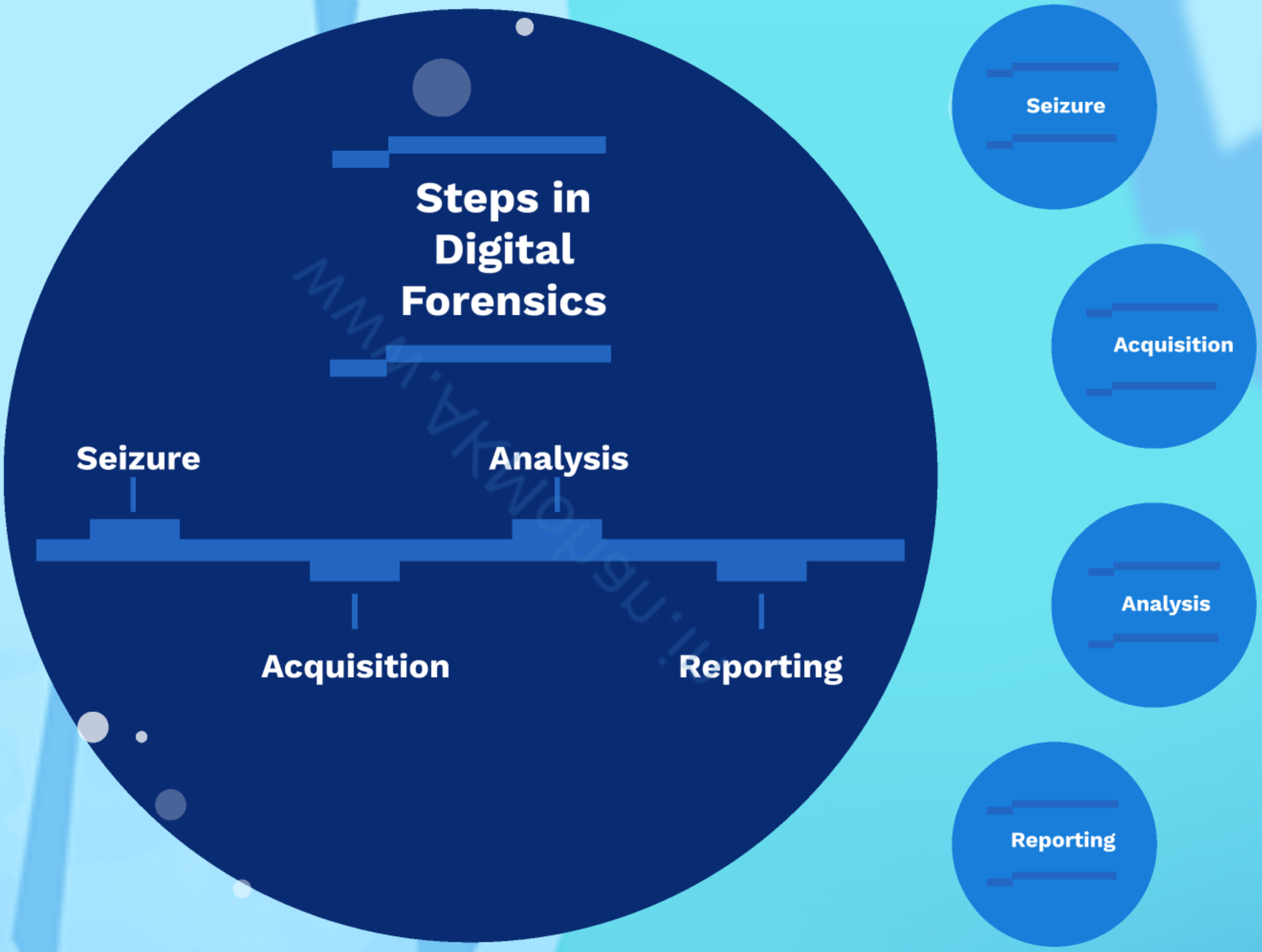
Reporting

Reporting

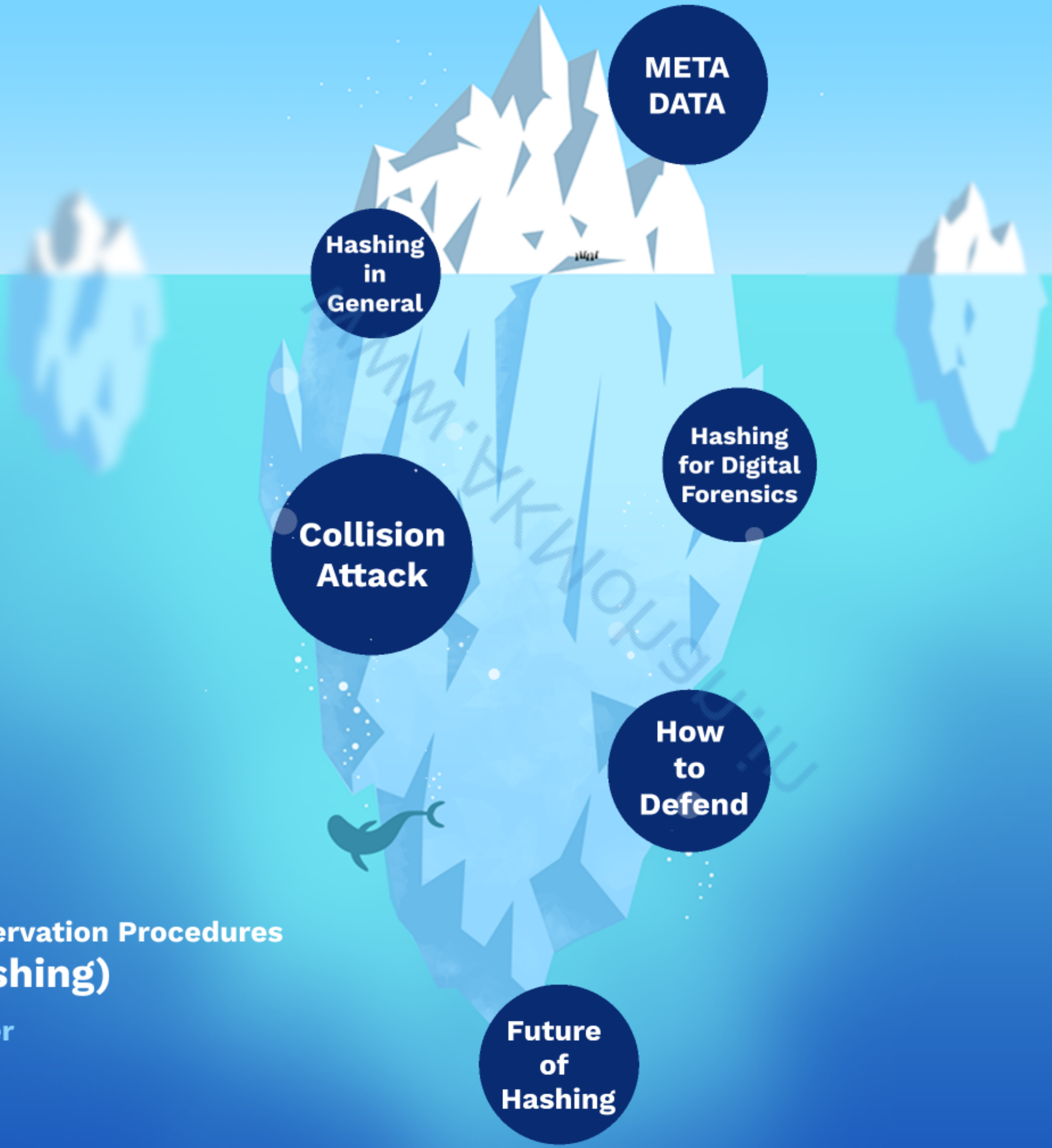
- **Documentation** (CoC)
- **Preservation** (Disposal of Backups)

  Reporting : document all the above process done on the evidence to prove in court

- **Presentation** (Expert Witness)



Steps in Digital Forensics
Seizure
Acquisition
Analysis
Reporting
Seizure
Acquisition
Analysis
Reporting



Digital Evidence Collection & Preservation Procedures

ROLE OF METADATA (Hashing)

NULL - OWASP Coimbatore Chapter
6th July 2024

ASHOK KUMAR MOHAN
Founder & Director
@ KRIYAVAN Cyber Forensic Service,
Mdu, TN, IN

Hash Collisions

- A collision is a condition whereby two different messages (evidences),
  - let say m1 (a.jpg) and m2 (b.jpg),
  - after applying the hash value, then H(m1) = H(m2).
- A collision can always be found using Brute Force algorithm,
  - however it is computationally difficult.

Table 1: List of most widely used Digital Forensic Tool

No.	Digital Forensic Tool	Hash Function	Features
1	EnCase	MD5	Remote Forensic Capability Evidence Processor Manager Smartphone and Table support Case Analyzer Email Review
2	Sun Sift	MD5	Network Forensics Computer Forensics Cloud Forensics Memory Forensics
3	Sleuth Kit	MD5	Contains a collection of unix commands for volume analysis and file systems
4	FTK Imager	SHA1 and MD5	Acquire and Preserve data from different media Forensics for computer and mobile Detect and validate suspected Malicious activities
5	Bulk Extractor	MD5	Forensic Scanner Feature Extraction Files, images and emails

MD5

SHA-x

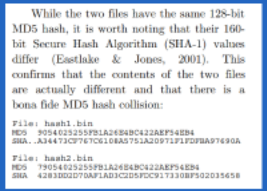
- A collision is a condition whereby two different messages (evidences),
  - let say m1 (a.jpg) and m2 (b.jpg),
  - after applying the hash value, then H(m1) = H(m2).
- A collision can always be found using Brute Force algorithm,
  - however it is computationally difficult.

Table 1: List of most widely used Digital Forensic Tool

No.	Digital Forensic Tool	Hash Function	Features
1	EnCase	MD5	Remote Forensic Capability Evidence Processor Manager Smartphone and Table support Case Analyzer Email Review
2	San Sift	MD5	Network Forensics Computer Forensics Cloud Forensics Memory Forensics
3	Sleuth Kit	MD5	Contains a collection of unix commands for volume analysis and file systems
4	FTK Imager	SHA1 and MD5	Acquire and Preserve data from different media Forensics for computer and mobile Detect and validate suspected Malicious activities
5	Bulk Extractor	MD5	Forensic Scanner Feature Extraction Files, images and emails

The Impact of **MD5** File Hash Collisions On Digital Forensic Imaging

Is there an example of two known strings which have the same MD5 hash value (representing a so-called "MD5 collision")?


While the two files have the same 128-bit MD5 hash, it is worth noting that their 160-bit Secure Hash Algorithm (SHA-1) values differ (Eastlake & Jones, 2001). This confirms that the contents of the two files are actually different and that there is a bona fide MD5 hash collision:

```
File: hash1.bin
MD5  9054025255FB1A26E4BC422AEF54EB4
SHA..A3447JCF767C6108A5751A20971F1FDFBA97690A

File: hash2.bin
MD5  79054025255FB1A26E4BC422AEF54EB4
SHA  4283DD2D70AF1AD3C2D5FDC917330BF502035658
```

One could create collisions using Marc Steven's HashClash on AWS and estimated the the cost of around $0.65 per collision.

These 2 images have the same md5 hash:
253dd04e87492e4fc3471de5e776bc3d

https://crypto.stackexchange.com/questions/1434/are-there-two-known-strings-which-have-the-same-md5-hash-value

Collisions On
Digital Forensic Imaging

Is there an example of two known strings which have the same MD5 hash value (representing a so-called "MD5 collision")?

While the two files have the same 128-bit
ID5 hash, it is worth noting that their 160-
it Secure Hash Algorithm (SHA-1) values
iffer (Eastlake & Jones, 2001). This
onfirms that the contents of the two files
re actually different and that there is a
ona fide MD5 hash collision:

```
ile: hash1.bin
D5  9054025255FB1A26E4BC422AEF54EB4
HA..A34473CF767C6108A5751A20971F1FDFBA97690A
```

One could create collisions using Marc Steven's HashClash on AWS and estimated the the cost of around

While the two files have the same 128-bit MD5 hash, it is worth noting that their 160-bit Secure Hash Algorithm (SHA-1) values differ (Eastlake & Jones, 2001). This confirms that the contents of the two files are actually different and that there is a bona fide MD5 hash collision:

```
File: hash1.bin
MD5  9054025255FB1A26E4BC422AEF54EB4
SHA..A34473CF767C6108A5751A20971F1FDFBA97690A

File: hash2.bin
MD5  79054025255FB1A26E4BC422AEF54EB4
SHA  4283DD2D70AF1AD3C2D5FDC917330BF502035658
```

g a so-called "MD5 collision")?

One could create collisions using Marc Steven's HashClash on AWS and estimated the the cost of around $0.65 per collision.

nages have the same md5 hash:

e87492c4fe3471de5e776be3d

These 2 images have the same md5 hash:

253dd04e87492e4fc3471de5e776bc3d

https://crypto.stackexchange.com/questions/1434/are-there-two-known-strings-which-have-the-same-md5-hash-value

Hash Collisions

- A collision is a condition whereby two different messages (evidences),
  - let say m1 (a.jpg) and m2 (b.jpg),
  - after applying the hash value, then H(m1) = H(m2).
- A collision can always be found using Brute Force algorithm,
  - however it is computationally difficult.

Table 1: List of most widely used Digital Forensic Tool

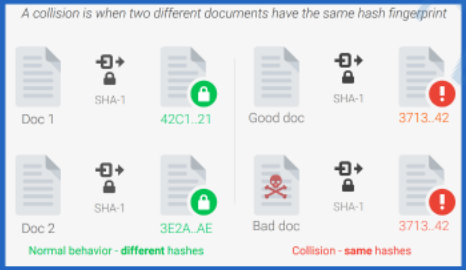
No.	Digital Forensic Tool	Hash Function	Features
1	EnCase	MD5	Remote Forensic Capability Evidence Processor Manager Smartphone and Table support Case Analyzer Email Review
2	Sun Sift	MD5	Network Forensics Computer Forensics Cloud Forensics Memory Forensics
3	Sleuth Kit	MD5	Contains a collection of unix commands for volume analysis and file systems
4	FTK Imager	SHA1 and MD5	Acquire and Preserve data from different media Forensics for computer and mobile Detect and validate suspected Malicious activities
5	Bulk Extractor	MD5	Forensic Scanner Feature Extraction Files, images and emails

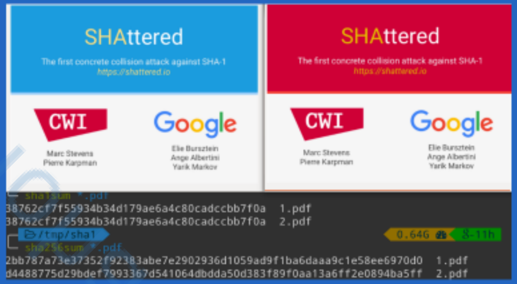
MD5

SHA-x

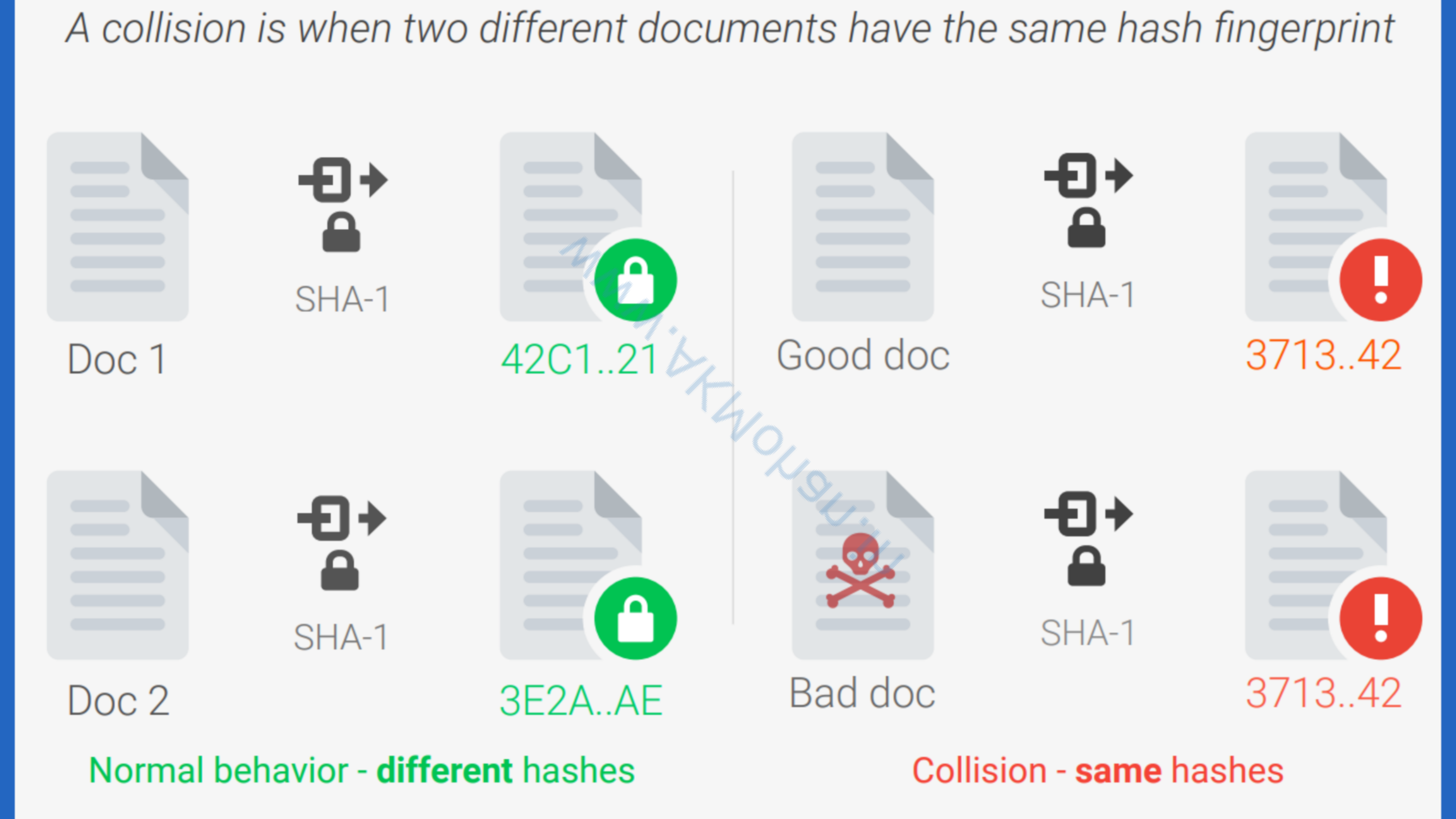
Broken **SHA-1** in practice

https://shattered.io/

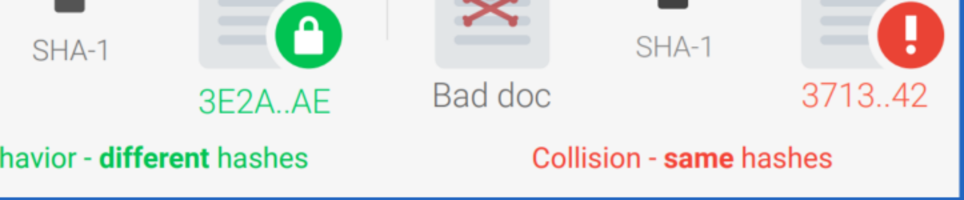


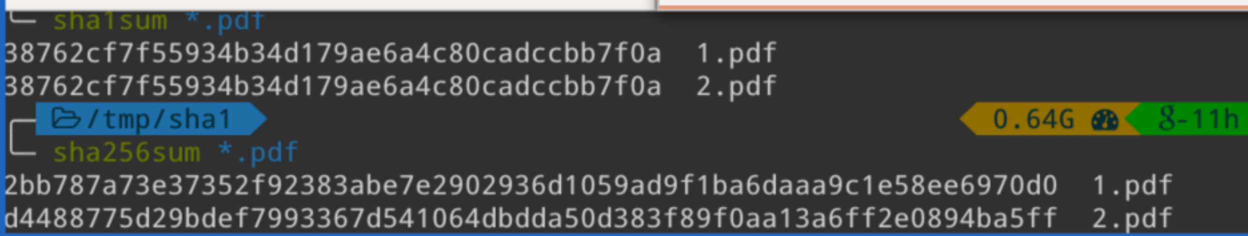


This attack required over **9,223,372,036,854,775,808** SHA1 computations. This took the equivalent processing power as **6,500 years** of single-CPU computations and **110 years** of single-GPU computations.


A collision is when two different documents have the same hash fingerprint
Doc 1
SHA-1
42C1..21
Good doc
SHA-1
3713..42
Doc 2
SHA-1
3E2A..AE
Bad doc
SHA-1
3713..42
Normal behavior - different hashes
Collision - same hashes


```
└ sha1sum *.pdf
38762cf7f55934b34d179ae6a4c80cadccbb7f0a  1.pdf
38762cf7f55934b34d179ae6a4c80cadccbb7f0a  2.pdf
┌ /tmp/sha1                                    0.64G   8-11h
└ sha256sum *.pdf
2bb787a73e37352f92383abe7e2902936d1059ad9f1ba6daaa9c1e58ee6970d0  1.pdf
d4488775d29bdef7993367d541064dbdda50d383f89f0aa13a6ff2e0894ba5ff  2.pdf
```





This attack required over **9,223,372,036,854,775,808** SHA1 computations. This took the equivalent processing power as **6,500 years** of single-CPU computations and **110 years** of single-GPU computations.

Hash Collisions

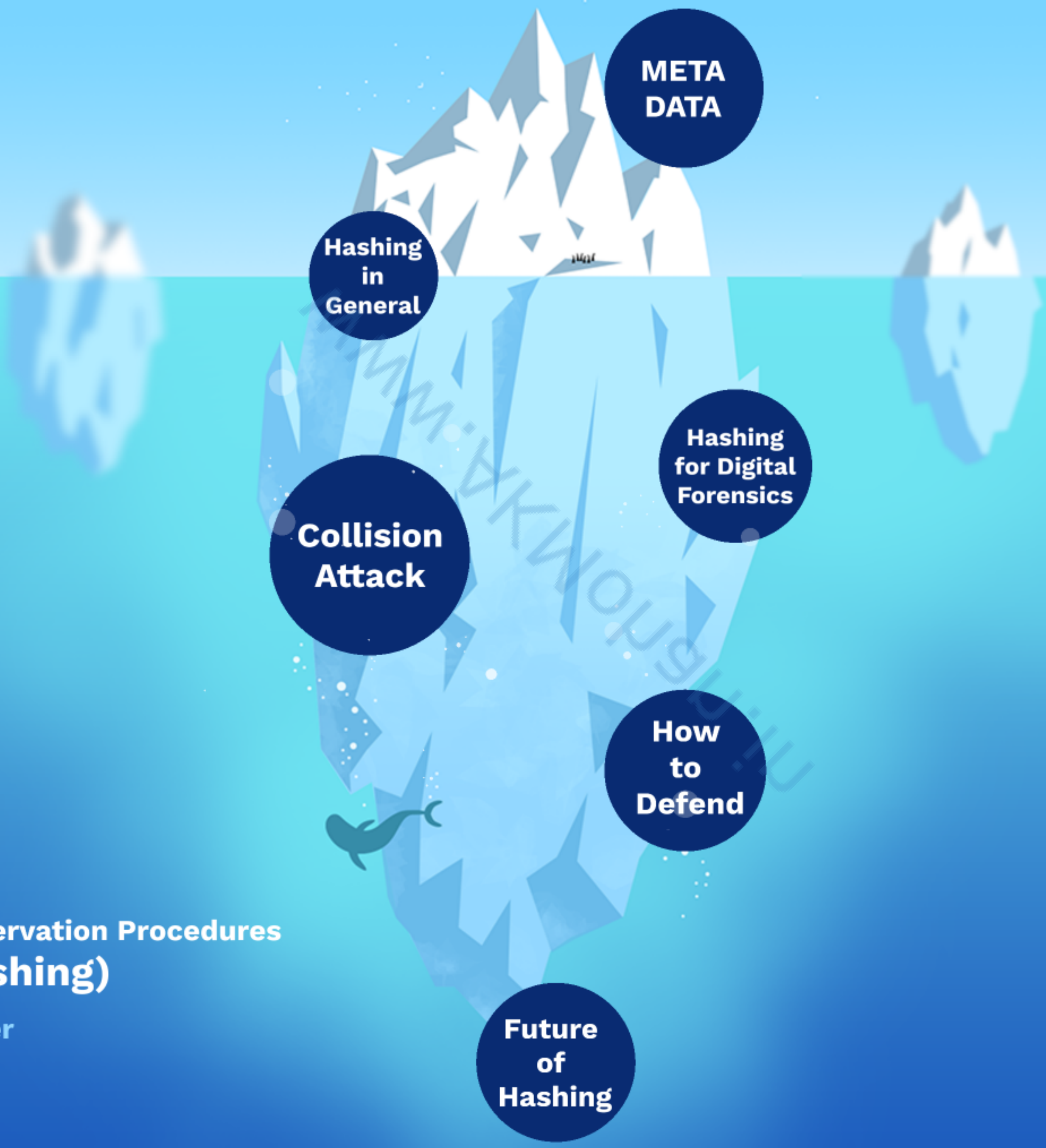
- A collision is a condition whereby two different messages (evidences),
  - let say m1 (a.jpg) and m2 (b.jpg),
  - after applying the hash value, then H(m1) = H(m2).
- A collision can always be found using Brute Force algorithm,
  - however it is computationally difficult.

Table 1: List of most widely used Digital Forensic Tool

No.	Digital Forensic Tool	Hash Function	Features
1	EnCase	MD5	Remote Forensic Capability Evidence Processor Manager Smartphone and Table support Case Analyzer Email Review
2	Sea Sift	MD5	Network Forensics Computer Forensics Cloud Forensics Memory Forensics
3	Sleuth Kit	MD5	Contains a collection of unix commands for volume analysis and file systems
4	FTK Imager	SHA1 and MD5	Acquire and Preserve data from different media Forensics for computer and mobile Detect and validate suspected Malicious activities
5	Bulk Extractor	MD5	Forensic Scanner Feature Extraction Files, images and emails

MD5

SHA-x



Digital Evidence Collection & Preservation Procedures

ROLE OF METADATA (Hashing)

NULL - OWASP Coimbatore Chapter
6th July 2024

ASHOK KUMAR MOHAN
Founder & Director
@ KRIYAVAN Cyber Forensic Service,
Mdu, TN, IN

How to ?

Use the latest hash Algorithms
or multi-tier hashes

Defense

- Use SHA-256 or SHA-3 as replacement
- Use shattered.io to test your PDF
- Google products are already protected
- Use collision detection code

Defense

Use SHA-256 or SHA-3 as replacement

Use shattered.io to test your PDF

Google products are already protected

Use collision detection code

Digital Evidence Collection & Preservation Procedures

ROLE OF METADATA (Hashing)

NULL - OWASP Coimbatore Chapter
6th July 2024

ASHOK KUMAR MOHAN
Founder & Director
@ KRIYAVAN Cyber Forensic Service,
Mdu, TN, IN

Future of Hashing !?!

(ought to be collision free)

- Cuckoo Hashing
- Perfect Hash Function
- Minimal Perfect Hashing
- **Fuzzy Hashing (SSDEEP)**
- Modified Secure Hashing algorithm (MSHA-512)

MD5, once considered really safe, now it's completely compromised.
Then there was **SHA-1**, which is now unsafe.
The same thing will surely happen to the widely used **SHA-2 & 3** someday in near future.

Sharing is Security

- Modified Secure Hashing algorithm (MSHA-51

MD5, once considered really safe, now it's completely compromised.

Then there was **SHA-1**, which is now unsafe.

The same thing will surely happen to the widely used **SHA-2 & 3** someday in near future.

What we've covered ?

Bird's Eye View of **Metadata** **(Hashing)**

Sh@rin9 !s
S3cur!tY

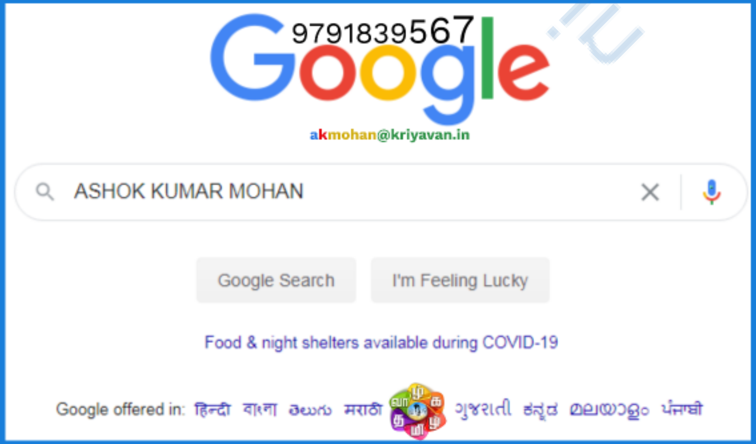
h @ 9 9 y
$ |-| @ 4 ! N 9

$h@Rin9 !5 S3<ur!"]["Y
h @ 9 9 y $ |-| @ 4 ! N 9

Disclaimer

This is a custom-made session which comprises of **my personal opinions, experiences and Bits 'n' Pieces** from all my mistakes accumulated over a decade in learning these stuffs. Every effort is made to keep the concepts authentic; but limited to ever changing information of the context used. Copyrights of the images used corresponds to the sources cited (SRC:) appropriately. All inferences discussed here are communicated at my discretion. By viewing/using this presentation i assume that you understand and will accept to share these concepts at your own risk of defending the same.

\\as(=)0K



$h@Rin9 !5 S3<ur!"]["Y
h @ 9 9 y $ |-| @ 4 ! N 9

Disclaimer

This is a custom-made session which comprises of **my personal opinions, experiences and Bits 'n' Pieces** from all my mistakes accumulated over a decade in learning these stuffs. Every effort is made to keep the concepts authentic; but limited to ever changing information of the context used. Copyrights of the images used corresponds to the sources cited (SRC:) appropriately. All inferences discussed here are communicated at my discretion. By viewing/using this presentation i assume that you understand and will accept to share these concepts at your own risk of defending the same.

\\as(=)0K

9791839567

Google

akmohan@kriyavan.in

ASHOK KUMAR MOHAN

Google Search I'm Feeling Lucky

Food & night shelters available during COVID-19

Google offered in: हिन्दी বাংলা తెలుగు मराठी வாழ்க தமிழ் ગુજરાતી ಕನ್ನಡ മലയാളം ਪੰਜਾਬੀ

What we've covered ?

Bird's Eye View of **Metadata** **(Hashing)**

Sh@rin9 !s
S3cur!tY

h @ 9 9 y
$ |-| @ 4 ! N 9

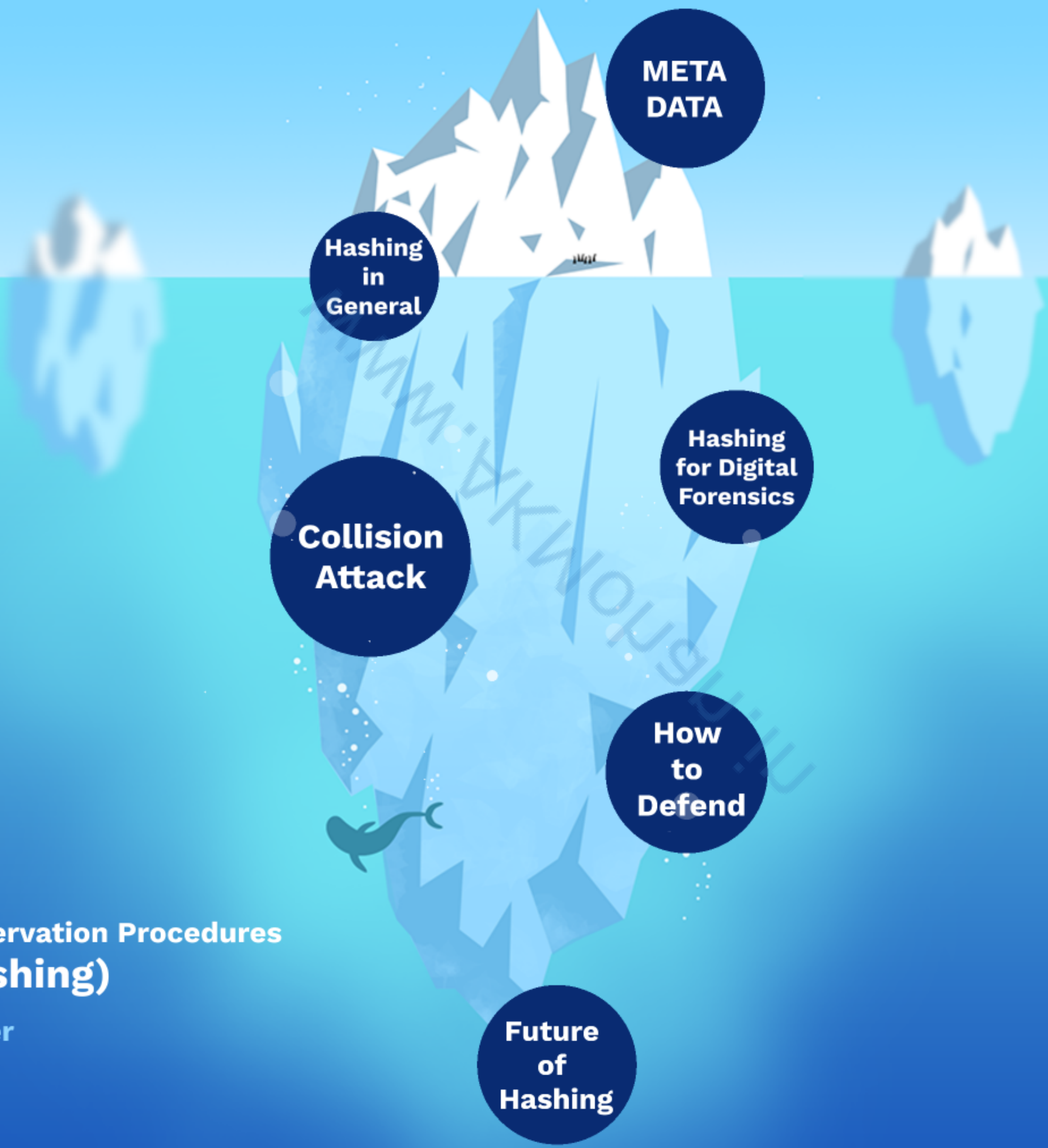
Future of Hashing !?!

(ought to be collision free)

- Cuckoo Hashing
- Perfect Hash Function
- Minimal Perfect Hashing
- **Fuzzy Hashing (SSDEEP)**
- Modified Secure Hashing algorithm (MSHA-512)

MD5, once considered really safe, now it's completely compromised.
Then there was **SHA-1**, which is now unsafe.
The same thing will surely happen to the widely used **SHA-2 & 3** someday in near future.

Sharing is Security



Digital Evidence Collection & Preservation Procedures

ROLE OF METADATA (Hashing)

NULL - OWASP Coimbatore Chapter
6th July 2024

ASHOK KUMAR MOHAN
Founder & Director
@ KRIYAVAN Cyber Forensic Service,
Mdu, TN, IN

Table 2. Demonstration of Assorted and Sparse Metadata (Filed-Value) Combinations

Index	Artifact (Evidence)	Source	Field: subject	Field: tags	Field: category	Field: copyright	Field: title	Field: <sparse field>
X1	pinkie.jpg	Ex1:C2M	pirated	stolen	<null>	<null>	<null>	<null>
X2	birds.jpg	Ex1:C2M	<null>	pirated	<null>	<null>	<null>	<null>
X3	DOC-S1As1.docx	Ex1:P2D	<null>	stolen	pirated	<null>	<null>	<null>
X4	pinkie.jpg	Ex1:L2P	<null>	<null>	<null>	stolen	<null>	<null>
X5	pinkie.jpg	Ex1:D2C	stolen	<null>	<null>	<null>	pirated	<null>
X6	Filename.vbe <random file type>	Ex2:*	<null>	stolen	<null>	<null>	<null>	amazon
X7	Filename.xlsm < unusual file type>	Ex3:*	<null>	<null>	<null>	pirated	<null>	fighter
X8	Filename.raw <corrupt file type>	Ex4:*	<null>	<null>	<null>	<null>	<null>	rao

in place. In existing similarity metadata matches, these sparse occurring file types are ignored totally and are addressed in the proposed unique association models. In the course of this article, the authors explain the unique mapping methodology to achieve the same. As a proof of concept the metadata field values namely amazon, fighter, pirated, rao, and stolen are embedded into the artifact metadata fields for demonstration.

The authors make use of Exiftool(a platform-agnostic CLI application) created and managed by Phil Harvey (2005) for interpretation, marking, and even restricting metadata over a variety of file types. It is powerful, speedy, customizable, and also provisionally processes files based on the value of any metadata taking numerous output formatting options. It also notes down every change in the file to creation, modification, and access date. Also, it's straightforward to create a text output file for each image file and the same can be extended to be stored in json, csv, and xls file formats.

With reference to the standard digital evidence analysis models by Agrawal, N., Bolosky, et al., (2007), the authors have categorized every digital artifacts (Origin O) into six major variety of families namely image (Family 1), file archiver (Family 2), executable (Family 3), document (Family 4), multimedia (Family 5) and forensic image (Family 6) as in Figure 1. The authors demonstrate the raw headers of one of the sample artifacts from the recently generated Amrita-TIFAC-Cyber/Digital-Forensics/UMAM-DF (Unique Metadata Association Model - Digital Forensics) datasets (2020). It shows the shift of metadata identifiers from the source (z) and the same artifact copied to

Figure 1. Families and Groups of Digital Artifacts (Author's Perception)

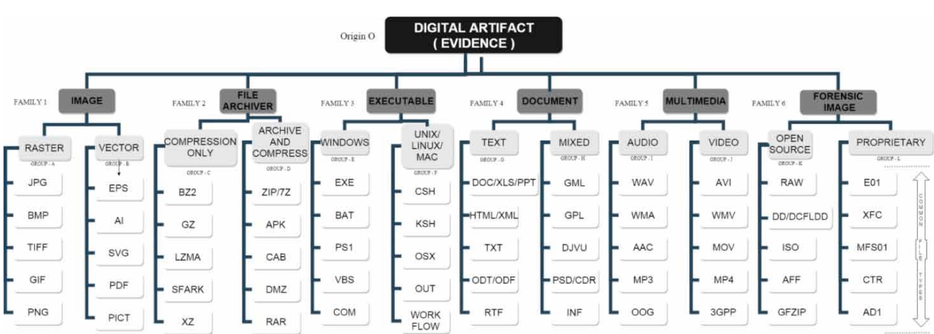


Table 3. Homogeneous and Heterogeneous Artifact Mapping

Artifact Mapping	Same Family-Same Type	Same Family -Different Type	Different Family -Same Type	Different Family -Different Type
File (pair) Nature	Purely Homogeneous	Habitually Homogeneous	Habitually Heterogeneous	Purely Heterogeneous
Example 1	G1: JPG - GIF	G1: JPG - EPS	G1: TIFF - PS1	G1: JPG - MP3
Example 2	G2: PNG - JPG	G2: TIFF - SVG	G2: BZ2 - 3GPP	G2: EXE - ISO
Example 3	G3: JPG - PNG	G3: PNG - PICT	G3: CAB - GFZIP	G3: TXT - E01

social media platform Facebook (z') illustrated around 90% of the actual metadata is modified or removed by the social media platform that possesses a nightmare for digital forensic investigators while proving their hypothesis before the jurisdiction.

(z) pinkie.jpg (S1As1-Mobile)
FF D8 FF E0 00 10 4A 46 49 46 00 01 01 01 0048 00 48 00 00 FF E1 13 EA 45 78 69 66 00 00 4D 4D 00 2A 00 00 00 08 00 0E 01 .. .. .. 28 00 03
(z') pinkie.jpg (S1As6-Facebook)
FF D8 FF E0 00 10 4A 46 49 46 00 01 01 00 0001 00 01 00 00 FF ED 00 84 50 68 6F 74 6F 73 68 6F 70 20 33 2E 30 00 38 42 49 .. .. .. 33 30 30

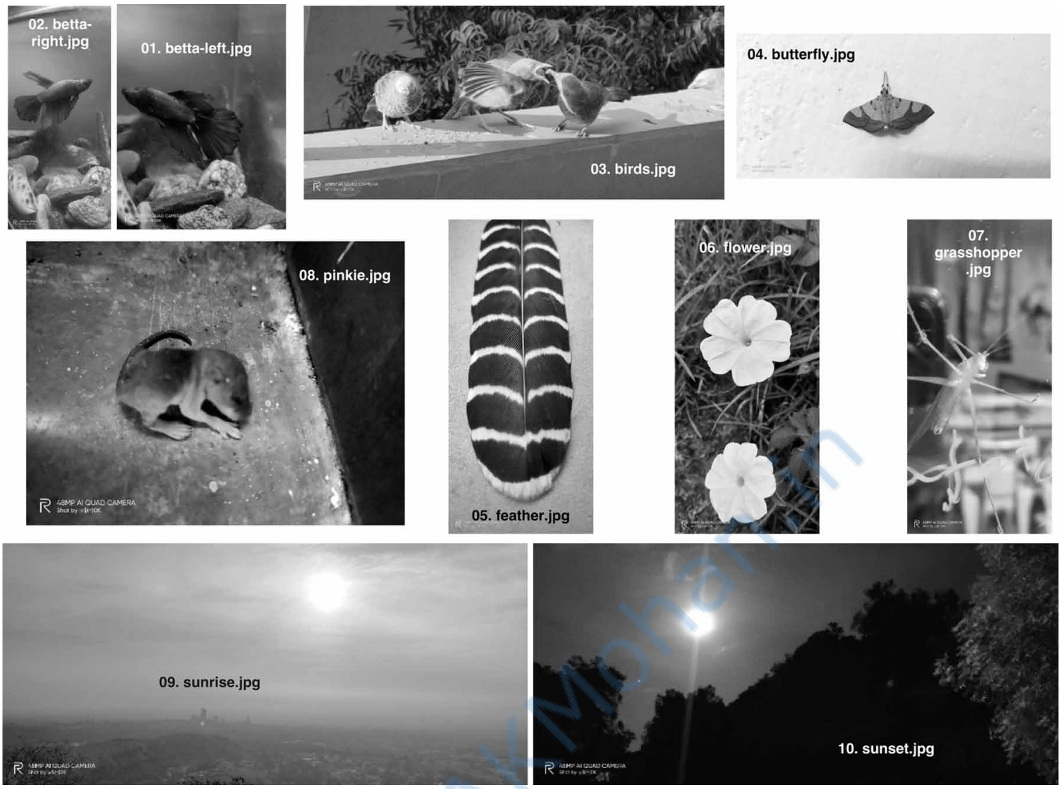
Metadata Association Models

The lemma based theorems on metadata similarity by Raghavan, S., & Raghavan, S. V. (2017) to identify the cause and effect of the relationship between metadata values to derive a grouping artifact on reducing the volume of metadata to be examined is a remarkable work. They gave details about the similarity between metadata in two hierarchies as similarity pockets and similarity groups. Afterward from these two association group is derived to find out the reduction factor and grouping efficiency by performing a lemma based analytics on metadata. Their future works were comprehensible on applying the theoretical proofs to existing datasets and to evaluate the difference between the forthcoming practical results of lemma implementation of their models. They also put forward to broaden the operational metadata association model to heterogeneous data sources and automating the same to be valid for digital evidence stored and processed during big data. This metadata association model is pretty good while handling any evidence with a distinct number of digital artifacts where a set of distinct extensions from a selected source is considered. The authors categorize artifacts into evidence types in various families and distinct file types with the example grouping shown in the following Table 3 with respect to Figure 1.

Determining Sparse Associations Between Metadata

With respect to the demonstration of assorted and sparse metadata (filed-value) combinations from Table 2, being motivated to generate and share the unique metadata-based dataset to the digital forensic research community. After comprehensive literature, on existing digital forensic datasets the authors have taken the following ten unique JPG images from dataset mobile source S1 and these acts as the reference (genesis) artifacts for the proposed unique mapping algorithm. The same set is synthetically recreated across all other sources as shown in Figure 2 keeping in mind each file holds the metadata created from their corresponding source file system and application for the visually similar images as stated by Buchholz, F., & Spafford, E. (2004). The ultimate purpose of this dataset is to recreate

Figure 2. Real-World Images Obtained from S1: Mobile (S1As1) with Complete Metadata



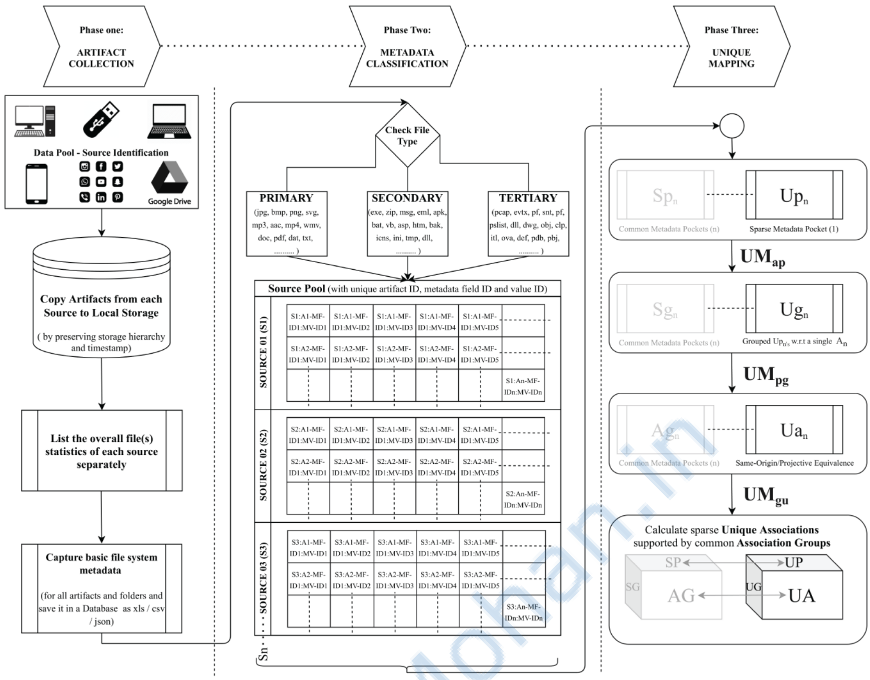
visually similar evidence (images in this case) at all sources and monitor the change or degradation of metadata on each iteration as shown in Figures 5 and 6.

WIDESPREAD SIMILARITY ASSOCIATION(S)

Metadata associations have been discussed in handling the digital forensic investigation for a while and there exist a plethora of syntactical models that roughly match the metadata composition and are not as much of predominant in addressing the explicit semantic behavior of the metadata attributes and their corresponding parameters. Raghavan, S., Clark, A., et al., (2009, January) hypothetically explicate the handling of multiple sources of evidence in a single framework (FIA) classified based upon source, data semantics, and storage file formats with the help of Malcolm Corney case on car theft investigation at Queensland University of Technology. They also emphasize extending this framework to design a suitable contrivance for validating their prototype amid real-world digital forensic datasets.

Raghavan, S., & Raghavan, S. V. (2013b, November) plotted metadata associations to establish a relationship between the artifacts and group the associated artifacts. AssocGEN analysis engine determines the relationship stuck between artifacts from files, logs, and network packet source to group the interrelated artifacts with respect to the circumstance of a digital investigation. Raghavan, S., & Saran, H. (2013, November) put forward the Provenance Information Model (PIM) to deal with the challenges related to timestamp analysis transversely for manifold time zones to precisely take into custody, the time zone in sequence and authenticate time-related affirmation during metadata analysis named after UniTIME timelining tool. Raghavan, S. (2014) thesis on Metadata Association Model

Figure 4. Phase-wise Implementation and Data Flow of Unique Associations



The artifacts are categorized into three distinct classifications namely primary, secondary, and tertiary as shown in Figure 4 for a convincing artifact triaging. The author's scope on this cataloging is to collect each and every metadata from a primary category like images, documents, and multimedia files in a forensically sound manner. Then the necessary metadata is collected in a secondary category based on the combination of EXIF, ICC, IPTC, and XMP metadata standards and lastly, the universally obtainable file system metadata is collected in the tertiary category. Unique metadata mapping aims at collecting all metadata even from tertiary evidence like pcap or evtx that might have a sparse association with any of the primary or secondary evidences.

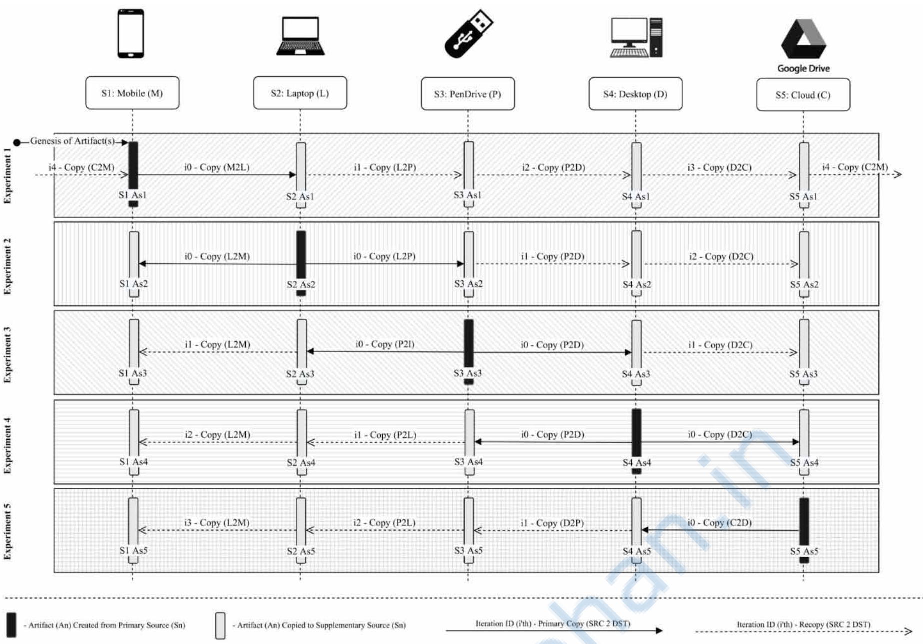
The building blocks for the metadata element for any artifact is represented by a regular 2-tuples notation by the authors throughout the article as <*field:value*> pair as in (3,4) for the publicly available metadata standards.

$$M_f - ID_n\ be\ the\ identifier\ for\ the\ 1st\ tuple \left\langle field: \right\rangle \\ \forall f\ identifcal\ notation\ \exists an\ fixed\ n \in \left[ASCII\left(num|char\right)\right]\} \tag{3}$$

$$M_v - ID_n\ be\ the\ identifier\ for\ the\ 2nd\ tuple \left\langle\ : value \right\rangle \\ \forall v\ identifcal\ notation\ \exists an\ viable\ n \in \left[ASCII\left(num|char\right)\right]\} \tag{4}$$

The combine notation of any metadata value corresponding to a metadata field that belongs to a unique artifact from a selected source is represented via (5) the below distinctive notation.

Figure 5. Iterative and Sequential Mobility (of Artifacts) in UMAM-DF Dataset



considerations for dataset collection are unchanged as the first set of experiments and it results in ten unique datasets. It collects the metadata of the file before and after sharing them between source devices and social media to calculate the final Association Group (AG) and Unique Association(UA) matches are shown in Figure 6.

The authors labeled the following metadata archive as "UMAM-DF" (Unique Metadata Association Model - Digital Forensics) dataset and are made publicly available at Amrita-TIFAC-Cyber/Digital-Forensics/UMAM-DF (Unique Metadata Association Model - Digital Forensics) datasets (2020) for suggestions and recommendations to enhance the same in near future for upcoming research works.

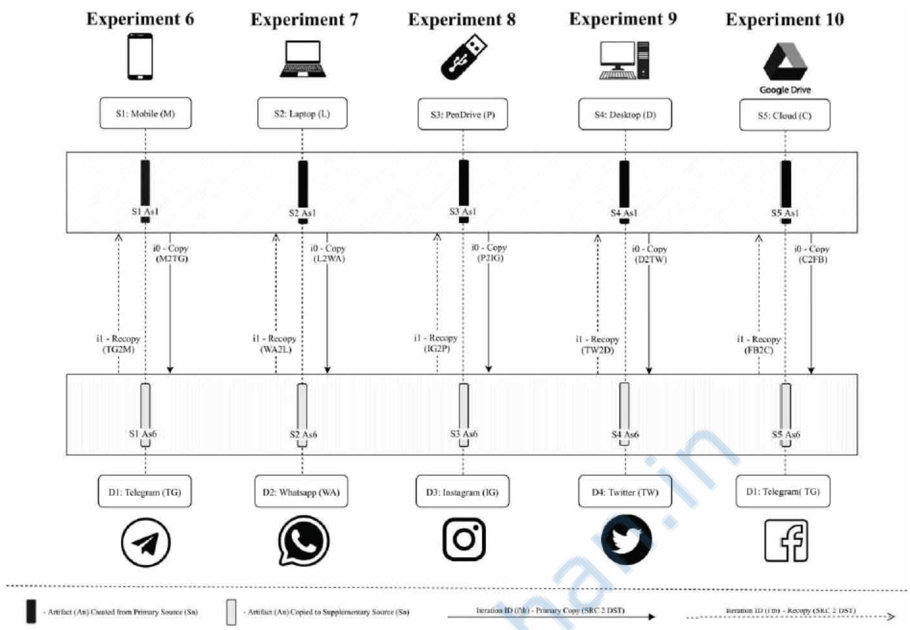
PROTOTYPE IMPLEMENTATION ON UMAM-DF DATASET

The series of sequential experiments collected with UMAM-DF dataset is engaged in testing the availability of metadata field-value pair matches across the sources with the collected set of 36 evidence sets as shared in Amrita-TIFAC-Cyber/Digital-Forensics/UMAM-DF (Unique Metadata Association Model - Digital Forensics) datasets (2020). The statistics of the similarity model and unique model of unaltered datasets are depicted in Table 5 resulting in linear Unique Group (UG) matches and variable Unique Association (UA) matches to adhere with their mathematical proof and algorithmic sequences.

The authors post a disclaimer for the repetitive values in SG produced during the experiment, as it is purely caused due to the availability of multiple identical metadata $S_n : A_n - M_f - ID_n : M_v - ID_n$ field-value pairs. This coherence can be ignored to maintain the integrity of the dataset as it is shared across the forensic community for reproducing the results as expected to verify the proposed model. The extended version of the same with normalized features is tabulated in Table 6.

Experiment 1 as shown in Figure 5 reveals the metadata matches of SP increases from 23(S1AS1) to 26(C2M) concluding that the additional metadata field-value pairs to be 22.5 and shows for every

Figure 6. Social Media Mobility (of Artifacts) in UMAM-DF Dataset



copy/paste at an average ±2 SP is achieved. The UP count reducing from 388 at S1AS1 in step 1 to 341 in step 6 reveals that around 47 unique pockets went missing when the files (namely01.betta-left.jpg to 10.sunset.jpg) went on to a complete round from mobile, back to mobile passing all other four sources as plotted in Figure 7. The experiment 2,3,4&5 expresses a similar shift over 47,21,62&62 unique pockets respectively in UP. The UG for all the experiments varies by ± SP across all experiments.

Unique pockets count of 380, 396, 339, 319 & 319 from source S1As6 drastically got reduced to 95,210, 96, 66 & 66 after passing via Telegram, Whatsapp, Instagram, Twitter, and Facebook

Table 5. Results for UP, UG, UA with respect to SP, SG, AG. (Unaltered UMAM-DF dataset)

UMAM-DF Dataset	Source	SP	UP	SG	UG	AG	UA
Experiment 1	S1As1	23	388	01	20	02	31
Experiment 2	S2As2	23	400	01	20	04	33
Experiment 3	S3As3	23	347	01	20	01	26
Experiment 4	S4As4	21	327	01	20	03	25
Experiment 5	S5As5	23	183	01	24	02	25
Experiment 6	S1As6	22	380	11	20	07	07
Experiment 7	S2As6	24	396	01	20	03	27
Experiment 8	S3As6	23	339	01	20	03	08
Experiment 9	S4As6	21	319	03	20	06	13
Experiment 10	S5As6	22	181	01	24	01	01
Overall Matches in SnAsn		225	3260	22	208	32	196