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)

FILE added as DOC (WhatsApp)

Around 30 Metadata Features

FILE added as IMAGE (WhatsApp)

Around 15 Metadata Features
**FILE added as DOC (WhatsApp)**

**Around 30 Metadata Features**

<table>
<thead>
<tr>
<th>File</th>
<th>EXIF</th>
<th>JFIF</th>
<th>Composite</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filename: IMG_20201015_191111_771.jpg</td>
<td>Image Width: 3303</td>
<td>JFIF Version: 1.01</td>
<td></td>
</tr>
<tr>
<td>File Size: 468 kB</td>
<td>Image Height: 1840</td>
<td>Resolution Unit: None</td>
<td></td>
</tr>
<tr>
<td>File Type: JPEG</td>
<td>Light Source: Unknown</td>
<td>X Resolution: 1</td>
<td></td>
</tr>
<tr>
<td>File Type Extension: jpg</td>
<td>Orientation: Unknown (0)</td>
<td>Y Resolution: 1</td>
<td></td>
</tr>
<tr>
<td>MIME Type: image/jpeg</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exif Byte Order: Big-endian (Motorola, MM)</td>
<td>GPS Latitude: 10 deg 54' 9.00&quot; N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Image Width: 3303</td>
<td>GPS Latitude Ref: North</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Image Height: 1840</td>
<td>GPS Longitude Ref: East</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Encoding Process: Progressive DCT, Huffman coding</td>
<td>GPS Longitude: 76 deg 53' 49.00&quot; E</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bits Per Sample: 8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Color Components: 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YCbCr Sub Sampling: YCbCr4:2:0 (2:2)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**FILE added as IMAGE (WhatsApp)**

Around **15 Metadata Features**
Original IMAGE from Mobile

Around 90 Metadata Features
A hash function is any function that can be used to map data of arbitrary size to data of fixed size

- Uses **Cryptography** Algorithms (MD5, SHA, ...)
- **Variable** length Input = **Constant** length Output
- **Irreversible** (only for checking the Integrity)
“the eagle flies at midnight”
How Are Evidence Copies Verified?

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

<table>
<thead>
<tr>
<th>Input</th>
<th>Hash Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dog</td>
<td>8AEFB06C 426E07AO A671AE1E 488B4858 D694A730</td>
</tr>
<tr>
<td>The brown dog runs down the hill</td>
<td>E192A01E CF8D20AD 0AFFEFDF 32CE934E 32FFC7E2</td>
</tr>
<tr>
<td>The brown dog leaps down the hill</td>
<td>47AB9979 443FB76D 1C193D06 773333BA 7876094F</td>
</tr>
</tbody>
</table>

Original Hard Drive

M75HZV33 9BQRID9B CPBN8251L 0BE5223G FW3CVD89

Cloned Hard Drive

M75HZV33 9BQRID9B CPBN8251L 0BE5223G FW3CVD89

PINPOINT LABORATORIES

www.pinpointlabs.com

©2008 Pivotal Guidance
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
- Acquisition
- Analysis
- Reporting
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

Analysis

Acquisition

Reporting
Acquisition

Authentication
- write blockers
- hashing

Acquisition (owning)
- Creating duplicate/clone of evidence by "Imaging"
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
- Analysis
- Acquisition
- 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.

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

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

These 2 images have the same md5 hash: 
253dd04e87492e4fc3471de5e776bc3d

[https://crypto.stackexchange.com/questions/7434/are-there-two-known-strings-which-have-the-same-md5-hash-value](https://crypto.stackexchange.com/questions/7434/are-there-two-known-strings-which-have-the-same-md5-hash-value)
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:

```plaintext
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
d5: 9054625255F81A266B4B4C221AEF54EB4

file: hash2.bin
d5: 344223C5FC65765168771F3F825A6B15
```

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
One could create collisions using Marc Steven's HashClash on AWS and estimated 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
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.
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)
- **Doc 2** -> SHA-1: 3E2A..AE (Bad doc)
- **Collisions** (same hashes):
  - SHA-1: 3713..42

Normal behavior - **different** hashes
Collision - **same** hashes
SHAttered
The first concrete collision attack against SHA-1
https://shattered.io

Marc Stevens
Pierre Karpman

Elie Bursztein
Ange Albertini
Yarik Markov

CWI
Google
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.

MD5

SHA-x
Digital Evidence Collection & Preservation Procedures

ROLE OF METADATA (Hashing)

Hashing in General
Collision Attack
How to Defend
Future of Hashing
Hashing for Digital Forensics
META DATA

Digital Evidence Collection & Preservation Procedures

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

NULL - OWASP Coimbatore Chapter
6th July 2024
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
Use SHA-256 or SHA-3 as replacement

Use shattered.io to test your PDF

Google products are already protected

Use collision detection code
ROLE OF METADATA (Hashing)

Digital Evidence Collection & Preservation Procedures

Date: 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.
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)
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(=)OK
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.
ASHOK KUMAR MOHAN

Google Search  I'm Feeling Lucky

Food & night shelters available during COVID-19

Google offered in: हिंदी बांग्लা अंग्रेजी मराठी गुजराती हंगारा संस्कृती पंजाबी
What we've covered?

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

META DATA

Hashing in General

Hashing for Digital Forensics

Collision Attack

How to Defend

Future of Hashing
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 non-source (x).

<table>
<thead>
<tr>
<th>Index</th>
<th>Artifact (Evidence)</th>
<th>Source</th>
<th>Field: subject</th>
<th>Field: tags</th>
<th>Field: category</th>
<th>Field: copyright</th>
<th>Field: title</th>
<th>Field: &lt;sparse field&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>pinkie.jpg</td>
<td>Ex1:C2M</td>
<td>pirated</td>
<td>stolen</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X2</td>
<td>birds.jpg</td>
<td>Ex1:C2M</td>
<td>&lt;null&gt;</td>
<td>pirated</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X3</td>
<td>DOC-S1As1.docx</td>
<td>Ex1:P2D</td>
<td>&lt;null&gt;</td>
<td>stolen</td>
<td>pirated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X4</td>
<td>pinkie.jpg</td>
<td>Ex1:L2P</td>
<td>&lt;null&gt;</td>
<td>&lt;null&gt;</td>
<td>stolen</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X5</td>
<td>pinkie.jpg</td>
<td>Ex1:D2C</td>
<td>stolen</td>
<td>&lt;null&gt;</td>
<td></td>
<td>pirated</td>
<td>&lt;null&gt;</td>
<td></td>
</tr>
<tr>
<td>X6</td>
<td>Filename.vbe</td>
<td>Ex2:*</td>
<td>&lt;null&gt;</td>
<td>stolen</td>
<td></td>
<td></td>
<td></td>
<td>amazon</td>
</tr>
<tr>
<td>X7</td>
<td>Filename.xlsm</td>
<td>Ex3:*</td>
<td>&lt;null&gt;</td>
<td>pirated</td>
<td></td>
<td></td>
<td></td>
<td>fighter</td>
</tr>
<tr>
<td>X8</td>
<td>Filename.raw</td>
<td>Ex4:*</td>
<td>&lt;null&gt;</td>
<td>&lt;null&gt;</td>
<td></td>
<td>rao</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

\begin{verbatim}
FF D8 FF E0 00 10 4A 46 49 46 00 01 01 00 04 80 00 48 00 00 FF E1 13 EA 45 78 69 66 00 00 4D 4D 02 2A 00 00 00 00 08 00 0E 01 .. .. .. 28 00 03
\end{verbatim}

\((z')\) pinkie.jpg (S1As6-Facebook)

\begin{verbatim}
FF D8 FF E0 00 10 4A 46 49 46 00 01 01 00 04 80 00 48 00 00 FF E1 13 EA 45 78 69 66 00 00 4D 4D 02 2A 00 00 00 00 08 00 0E 01 .. .. .. 28 00 03
\end{verbatim}

### 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
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
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 \( <\text{field}:\text{value}> \) pair as in (3,4) for the publicly available metadata standards.

\[
M_f - \text{ID}_n \text{ be the identifier for the } 1\text{st tuple }\left\{\text{field} : \right\} \\
\forall f \text{ identifical notation } \exists an \text{ fixed } n \in \left[\text{ASCII}\left(\text{num|char}\right)\right]
\]

\[
M_v - \text{ID}_n \text{ be the identifier for the } 2\text{nd tuple }\left\{ : \text{value} \right\} \\
\forall v \text{ identifical notation } \exists an \text{ viable } n \in \left[\text{ASCII}\left(\text{num|char}\right)\right]
\]

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.
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 $S_n : A_n - M_n - ID_n : M_n - ID_n$ 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
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)

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