Aggregation of Digital Forensics Evidences

Premal C. Patel
Department of Computer Science,
IT Systems and Network Security,
Gujarat Technological University,
India

***Abstract*— Cyber forensics tools are storing digital evidences in different formats like Raw format, Proprietary formats, Advanced Forensics Format. Each tool has its own techniques of storing evidences. As Raw format is considered to be best storage technique. Therefore further analysis of these evidences would be bounded to specific group of tools only. But due to some limitations of these formats there is need to create a standard structure for storing digital evidence. Forensic data is most important things in case of during cyber crime investigation therefore it must be secure and preserve that evidence is also challenge. In this type of cases the proper chain of custody must be maintained. The collector of data is responsible for gathering all information related to evidence.**

***Keywords*— advance Forensic Format, Raw Format, Proprietary Formats, File Metadata, Hash, Forensic Data, Aggregation, Storage Format .**

I. Introduction

Computers have become an important part of our lives. This does not exclude criminals who have the technical knowledge how to hack computer network systems. Electronic evidence has played an important role in court but obtaining can be difficult. Computer forensics is beneficial but it also has some disadvantages. Like, Securing digital evidences being tempered, Storing all digital evidences etc. Computers are the most dominant form of technology. It has been used in variety of purposes which has made digital and electronic evidence important.

In modern generation various types of data are available and it is difficult to analyze. Aggregation of Big data is also difficult for forensic analysis. It occupies large storage and requires many storage devices with high space. Even proper storage structure should also be used in order to faster access and further analyze of all data by different tools. If proper structure is not used to store all digital evidence then it requires too much time to gathering and analyzing data. So, we need a proper structure which stores all types of data in proper format which makes easy to analysis base approaches.

Interoperability of forensic data between Forensic Tools is need for common structure to store digital forensic data and evidence. There is one approach is also available called AFF(Advance Forensic Format) which define proper structure to store image of forensic data and dfxml (digital forensic xml) is approach for interoperability.

II. Limitations of different storage format

There are three storage Formats for Digital Evidence 1. Raw format 2. Proprietary formats 3. Advanced Forensics Format (AFF). All these formats have some advantages as well as disadvantages.

A. Raw Format

Makes it possible to write bit-stream data to files

Advantages

- Fast data transfers
- Can ignore minor data read errors on source drive
- Most computer forensics tools can read raw format

Disadvantages

- Requires as much storage as original disk or data
- Tools might not collect marginal (bad) sectors

B. Proprietary Formats

Features offered

- Option to compress or not compress image files
- Can split an image into smaller segmented files
- Can integrate metadata into the image file

Disadvantages

- Inability to share an image between different tools
- File size limitation for each segmented volume

C. Advanced Forensics Format

Developed by Dr. Simson L. Garfinkel of Basis Technology Corporation

The need to easily share digital evidence between different organizations and analysis tools is increasing as crimes and security incidents involve a diverse range of digital devices and administrative domains. Converting between proprietary formats may result in incorrect data, missing metadata, and lost productivity. There is a need for a standard format for storing and transmitting digital evidence and its associated metadata. A standard format would have the following benefits:

- Digital evidence can be effortlessly imported into multiple analysis tools thereby reducing the time needed to convert formats or use a format that carries no metadata or integrity information.
- Metadata are stored with the digital evidence thereby reducing the possibility of introducing errors while converting storage formats and losing information because it is stored in multiple locations (including different storage formats files, proprietary case files, and notebooks).[1]

A standard format would decrease the time needed to complete an investigation, increase the amount of information available to the investigator, and increase the reliability of the evidence.

A standard digital evidence storage format will be analogous to the evidence bags used at physical crime scenes, where the evidence is placed in the bag and the outside of the bag has related information in a standard language, such as the acquisition location and time written in English. The current state of digital evidence storage formats is similar to having no bag, bags with information written in different languages, or bags with different types of locking mechanisms.

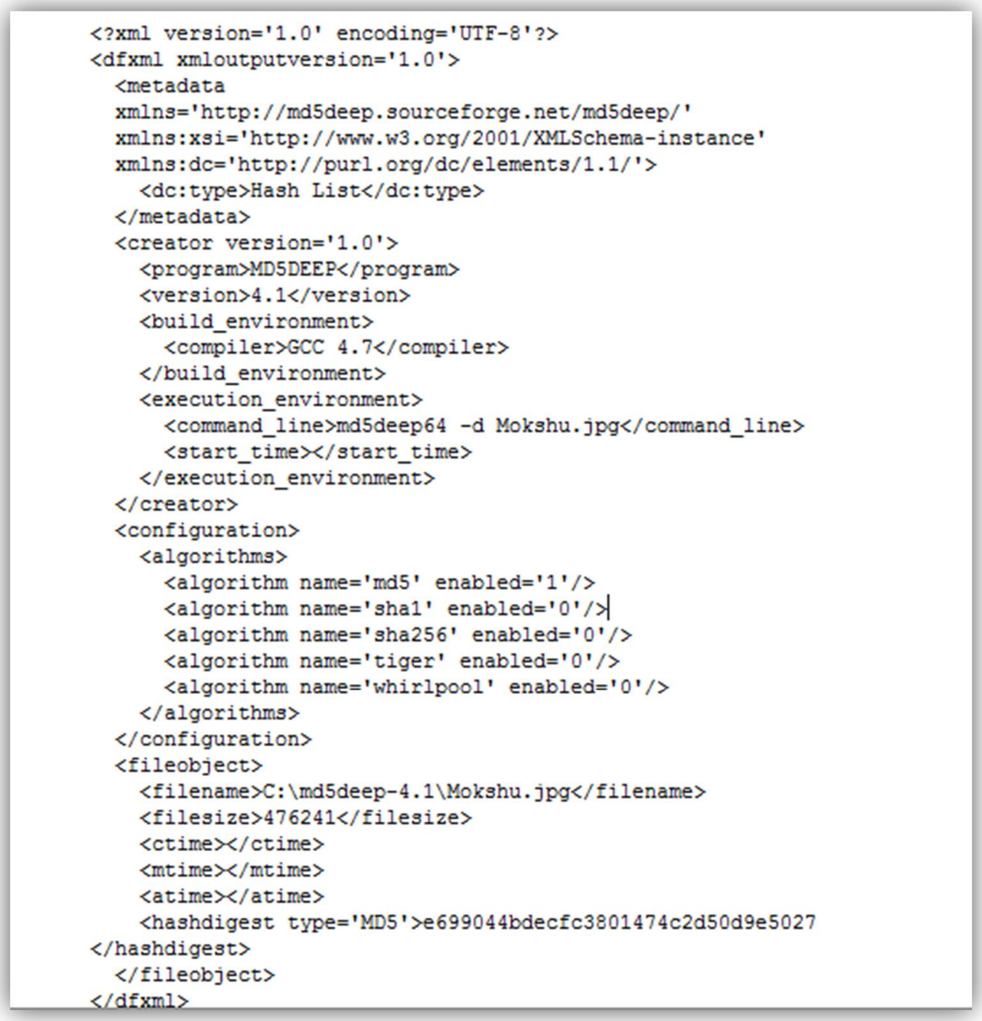
III. Design goals

- Provide compressed or uncompressed image files
- No size restriction for disk-to-image files
- Provide space in the image file or segmented files for metadata
- Simple design with extensibility
- Open source for multiple platforms and Operating Systems
- Internal consistency checks for self-authentication
- File extensions include .afd for segmented image files and .afm for AFF metadata
- AFF is open source[2]

IV. Previous work

Some tools are producing dfxml format output like MD5Deep which has many features for aggregate the information about data but it has also some limitation. It cannot define all information in network based environment also not define system information.

In MD5DEEP tools we can gather information about file, file type, metadata like created, modified, access date and time. we can also get its hash value and file size, its physical location and version of DFXML but we can't get it's system and network information like MAC & IP address. This information is also very valuable for forensic investigation.



```
<?xml version='1.0' encoding='UTF-8'?>
<dfxml xmloutputversion='1.0'>
  <metadata
  xmlns='http://md5deep.sourceforge.net/md5deep/'
  xmlns:xsi='http://www.w3.org/2001/XMLSchema-instance'
  xmlns:dc='http://purl.org/dc/elements/1.1/'>
    <dc:type>Hash List</dc:type>
  </metadata>
  <creator version='1.0'>
    <program>MD5DEEP</program>
    <version>4.1</version>
    <build_environment>
      <compiler>GCC 4.7</compiler>
    </build_environment>
    <execution_environment>
      <command_line>md5deep64 -d Mokshu.jpg</command_line>
      <start_time></start_time>
    </execution_environment>
  </creator>
  <configuration>
    <algorithms>
      <algorithm name='md5' enabled='1'/>
      <algorithm name='sha1' enabled='0'/>
      <algorithm name='sha256' enabled='0'/>
      <algorithm name='tiger' enabled='0'/>
      <algorithm name='whirlpool' enabled='0'/>
    </algorithms>
  </configuration>
  <fileobject>
    <filename>C:\md5deep-4.1\Mokshu.jpg</filename>
    <filesize>476241</filesize>
    <ctime></ctime>
    <mtime></mtime>
    <atime></atime>
    <hashdigest type='MD5'>e699044bdecfc3801474c2d50d9e5027
</hashdigest>
  </fileobject>
</dfxml>
```

Fig.1 md5deep generated dfxml file

The PhotoRec carver and the MD5DEEP hashing application were both modified to produce DFXML files. The DFXML output contains the files identified, their physical location within the disk image, and their cryptographic hashes. md5deep is a suite of cross platform tools to compute and audit hashes for any number of input files.md5deep currently supports MD5, SHA-1, SHA-256, Tiger, and Whirlpool. DFXML improves composability by providing a language for describing common forensic processes (e.g., cryptographic hashing), forensic work products (e.g., the location of files on a hard drive), and metadata (e.g., filenames and time stamps). [3]

Now current strategy on dfxml needs some improvement for analysing in network based environment. Most of organization are working in their own domain and also use their own intranet so collecting and storing information in proper manner makes easy to analysis. Another benefit is interoperability between the tools which support dfxml to store the information.

V. Concept of DFXML

Digital Forensics XML(DFXML) is a tool for describing file systems and metadata.

Today most forensic tools report metadata in human-readable form.

- Location of partitions.
- Location of a file.
- File owner, MAC times, etc.
- Microsoft Office permissions.

This leads to problems:

- Each tool processing a disk image must re-interpret the file system.
- One tool cannot be easily validated against another.
- DFXML allows tools to interoperate.

The basic is to use an XML as an intermediate format. XML allows us to separate file extraction from forensic analysis.[4]

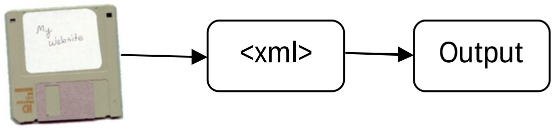


Fig.2 Extraction of data from storage media and using xml as an intermediate

Most of digital forensic tools use XML to store data for digital forensic environment because there is very large amount of data has been stored in other storage media. XML is very flexible for store the data and we can easily produce and consume those data for analysis.

DFXML is an approach for produce and consume those data in such a common format for all tools that define a common method to store data in XML which also call as DFXML. by using this method of storing digital forensic evidence common structure we can use interoperability between those all tools.[5]

We can use any tools to fetch digital forensic data and also use any tools to analysis for that data so it is very easy to use by any tool at any time. Here we can get many benefits by using this type of system to built forensic tools like only one time aggregate the data for all tools.

The rapid pace of innovation in digital technologies presents substantial challenges to digital forensics. New memory and storage devices and refinements in existing ones provide constant challenges for the acquisition of digital evidence. The proliferation of competing file formats and communications protocols challenges one’s ability to extract meaning from the arrangement of ones and zeros within. Overarching these challenges are the concerns of maintaining the integrity of any evidence found, and reliably explaining any conclusions drawn. [6]

VI. Goals for DFXML

- Complement existing forensic formats.
- Be easy to generate
- Be easy to ingest
- Provide for human readability
- Be free, open and extensible
- Provide for scalability
- Adhere to existing practices and standards[5]

VII. Future work

DFXML performs better to store the forensics data and it needs more reliable by preserving the information to prevent unauthorized access. In future work Use of DFXML in a cluster/HPC environment. Use of DFXML for digital archives and in other forensic communities. To provide integrity & confidentiality to forensics data. Data should be archived in well defined manner so that it can be available any time in its original form.

The structure needs implementation for gathering more information about system and network for analysis make the structure simple and sober.

VIII. Conclusion

DFXML helps making the forensic data investigation easy by preserving the integrity of data and also by making it interoperable with other tools and applications. This research describes Digital Forensics XML (DFXML), an XML language for digital forensics research and interchange. DFXML is designed to be an interchange format between forensic tools. The abstractions represented in DFXML have been specifically chosen to represent digital forensic processing steps, allowing for ease of generating and ingesting DFXML objects.

Acknowledgement

I would like to express my gratefulness towards many people whose contribution had made this research work in this area.

I would like to express my heartful gratitude towards my guide Kaushal Bhavsar for this guidance on various aspects of this document and his contribution on the betterment of my work.

I would like to thank Mr. Naresh Gardas and Miss. Kiran Bhagiya for their thoughts and feedbacks of this draft.
Under these responsible people I was able to complete the paper of Aggregation of digital forensics evidence in-time successfully.

Lastly I would also like to give a special thanks to my friends for playing an important role in strengthening my approaches in this field.

References

[1] www.dfrws.org/CDESF/index.shtml

[2] www.cps.brockport.edu/~shen/cps301/Chapter4.ppt

[3] http://md5deep.sourceforge.net

[4] http://faculty.nps.edu/sgarfin

[5] http://simson.net/clips/academic/2012.DI.dfxml.pdf

[6] *Digital Forensics XML and the DFXML Toolset* by Simson Garfinkel in Naval Postgraduate School, 900 N. Glebe, Arlington, VA 22203