International Journal of Engineering Research and Development
e-ISSN : 2278-067X, p-ISSN : 2278-800X, www.ijerd.com
Volume 2, Issue 5 (July 2012), PP. 64-70

Significance of Hash Value Generation in Digital Forensic: A Case Study

Kailash Kumar[1], Sanjeev Sofat[2], S.K.Jain[3], Naveen Aggarwal[4]
[1]*PhD (CSE) Student, PEC University of Technology, Chandigarh, India*
[2]*Prof. and Head CSE Deptt. PEC University of Technology, Chandigarh, India*
[3]*Deputy Director (Ball.) CFSL, Chandigarh, India*
[4]*Assistant Professor CSE Deptt. UIET, Chandigarh, India*

Abstract—Digital forensics tools frequently use to calculate the hash value of digital evidence drive. MD5 and SHA hash function is used in digital forensic tools to calculate and verify that a data set has not been altered, due to the application of various evidence collection and analysis tools and procedures. Additionally, due to the impact on the personal life of the subject of an investigation, verification of the correct operation of tools and procedures is critical. This paper discusses the significance of hash value in digital forensic for the digital evidence. The research conducts six possible different cases as an experiment to generate and verify the hash value of test drive using forensic tool to demonstrate the importance of hash value in digital forensic. Additionally, unreliable results can be obtained because of the improper use of the application of the tools.

Keywords—MD5; SHA; digital forensic; hash function.

I. INTRODUCTION

The field of digital forensic analysis has experienced rapid growth in recent years, as the use of computer forensic analysis proved invaluable in a wide range of legal proceedings. Digital forensics is used not only to investigate computerized crimes, such as network intrusion, fabrication of data and illegitimate material distribution through digital services, but also to investigate crime where evidence is stored in any digital format on any digital device [1]. One of the most important steps performed during a digital forensics investigation is the data acquisition step, which is "the task of collecting digital evidence from electronic media" [6]. During this step, the investigator makes an exact copy of the evidence disk or file to produce a forensics copy. To avoid destroying evidence, the investigation is conducted on the forensic copy, rather than the original evidence data.

Thus, any data corruption that occurs during the investigation process can be rectified by using the evidence disk to create a new forensic copy to use to continue the investigation. Because a digital investigation frequently yields results that are used in a criminal or civil court proceeding which can drastically affect a person life, the investigator must be completely certain that the forensic copy is an exact copy of the evidence.

The hash value plays an important role in forensic investigation for proving accuracy of digital data in front of judiciary. In this research article we are proposing some real time case study to prove the importance of hash value in digital forensic investigation process.

The remainder of this paper is organized as follows. Section 2 gives an overview of digital forensics. Section 3 provides a short background on hash function. Section 4 focuses on some case study of hash value generated in digital hard drive for forensic point of view. Finally section 5 concludes the paper and our future work.

II. DIGITAL FORENSIC SCIENCE

The use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations. In 2001, the Digital Forensics Research Workshop [DFRW] [3] proposed a process for digital investigations that involves the following six steps. In this time we are more concern about the analysis phase, hash analysis is also a part of whole digital analysis that we mention in the original DFRW model illustrates in fig 1.

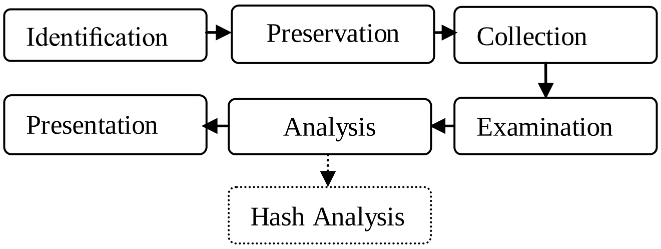


Fig. 1. Digital forensic investigation process

Digital forensics is the science of identifying, extracting, analyzing and presenting the digital evidence that has been stored in the digital electronic storage devices to be used in a court of law [1, 4, 5].

III. HASH FUNCTION

Definition: An algorithm that turns a variable-sized amount of text into a fixed-sized output (hash value). Hash functions are used in creating digital signatures, hash tables and short condensations of text for analysis purposes. Hash functions are also known as "cryptographic hash functions.

A hash function H is a transformation that takes a variable-size input 'm' and returns a fixed-size string, which is called the hash value h (that is, h = H(m)).

Hash functions with just this property have a variety of general computational uses, but when employed in cryptography the hash functions are usually chosen to have some additional properties.

The basic requirements for a cryptographic hash function are:

- the input can be of any length,
- the output has a fixed length,
- H(x) is relatively easy to compute for any given x ,
- H(x) is one-way,
- H(x) is collision-free.

A hash function H is said to be one-way if it is hard to invert, where "hard to invert" means that given a hash value h, it is computationally infeasible to find some input x such that H(x) = h.

If, given a message x, it is computationally infeasible to find a message y not equal to x such that H(x) = H(y) then H is said to be a weakly collision-free hash function.

A strongly collision-free hash function H is one for which it is computationally infeasible to find any two messages x and y such that H(x) = H(y).

The hash value represents concisely the longer message or document from which it was computed; one can think of a message digest as a "digital fingerprint" of the larger document. Perhaps the main role of a cryptographic hash function is in the provision of digital signatures. Since hash functions are generally faster than digital signature algorithms, it is typical to compute the digital signature to some document by computing the signature on the document's hash value, which is small compared to the document itself. Additionally, a digest can be made public without revealing the contents of the document from which it is derived. This is important in digital time stamping where, using hash functions, one can get a document time stamped without revealing its contents to the time stamping service.

A. Practical Hash Function

This section covers hash functions that are most likely used in digital forensic software/tools: MD5 and SHA-1. For their detailed description we refer the reader to the documents issued by standardization bodies.

MD4 and MD5:MD4 was proposed by Ron Rivest in 1990 and MD5 [7] followed shortly thereafter as its stronger version. Their design-had great influence on subsequent constructions of hash function. The letters "MD" stands for "message digest" and the numerals refer to the functions being the fourth and fifth designs from the same hash-function family.

SHA-0 and SHA-1: The Secure Hash Algorithm (SHA) was initially approved for use with the Digital Signature Standard (DSS) in 1993 [2]. Two years later the standard was updated to become what is currently known as SHA-1 [8]. The first version of SHA is referred in the cryptographic literature as SHA-0, although it has never been its official designation. SHA-1 differs from SHA-0 by exactly one additional instruction, which is nonetheless extremely important from the cryptanalytic perspective. Since there were no reasons to prefer the initial version of the standard, SHA-1 replaced SHA-0 in all but most antiquated applications. The details of these hash function is briefly illustrates in table 1below.

Table1. Practical Hash function

Name	Block Size(bits)	Word Size(bits)	Output Size(bits)	Rounds
MD4	512	32	128	48
MD5	512	32	128	64
SHA-0	512	32	160	80
SHA-1	512	32	160	80

B. Hash value generation in digital forensic

Generally hash value is used to check the integrity of any data file but, in digital forensic it is used to check the integrity of evidence disk data. The image of a disk is created in digital forensic for analysis so, it is necessary the image have exactly or replica of evidence disk. The hash value generated during imaging should match when that image of evidence disk is extracted for detail analysis. In digital forensic hash value is generated for whole disk data not only single or multiple files.

The hash value generated using forensic tools in the form of hexadecimal notation. Here we are giving an example to convert it in too two easily understandable form for forensic practitioner who don't have enough knowledge about computer system.

Using the hash value generated of case1: 79EAB87F0D3A3B45954779A72F79AE63

1. Table 2 shows the binary form of the given hexadecimal value:

Table2. Binary code for Hash value

0111	1001	1110	1010	1011	1000	0111
7	9	E	A	B	8	7
1111	0000	1101	0011	1010	0011	1011
F	0	D	3	A	3	B
0100	0101	1001	0101	0100	0111	0111
4	5	9	5	4	7	7
1001	1010	0111	0010	1111	0111	1001
9	A	7	2	F	7	9
1010	1110	0101	0011			
A	E	6	3			

2. The following steps involve to convert given hexadecimal hash value into decimal form:

Step1. Use Hexadecimal to Decimal conversion process as given below:

$7*16^{31} + 9*16^{30} + E*16^{29} + A*16^{28} + \ldots\ldots\ldots\ldots 6*16^{1} + 3*16^{0}$

Step2. Substitutes the equivalent numerical value in place of alphabet in hexadecimal hash value as given below.
A=10, B=11, C=12, D=13, E=14, F=15

After substitution:
$7*16^{31} + 9*16^{30} + \mathit{14}*16^{29} + \mathit{10}*16^{28} + \ldots\ldots 6*16^{1} + 3*16^{0}$

Step3. Calculate hash value in decimal form.

IV. PROPOSED FRAMEWORK FOR HASH VALUE CALCULATION

The experimental model/framework for hash value calculation in digital forensic is proposed in fig 2. Using this model the test data has been generated on digital hard drive. Finding evidence for system tampering, data hiding or deleting utilities, unauthorized system modifications etc. should also be performed. Detecting and recovering hidden or obscured information is a major tedious task involved. Data should be searched thoroughly for recovering passwords, finding unusual hidden files or directories, file extension and signature mismatches etc. while searching the above said information from an evidence disk the forensic software is also create the hash value of the whole drive to check the integrity of the disk. In forensic data acquisition phase the hash value is generated at the time of imaging the evidence disk and compare that hash value at the time of examination or copying the contents of the disk. If both the hash value is identical than the forensic expert assume everything is fine otherwise some kind of tampering is involved with the evidence disk. Here the case study focuses the importance of hash value in forensic examination has been explored.

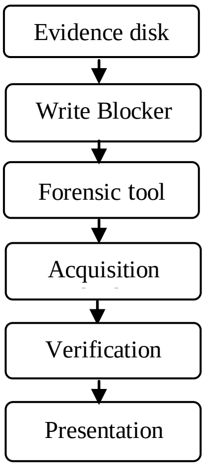
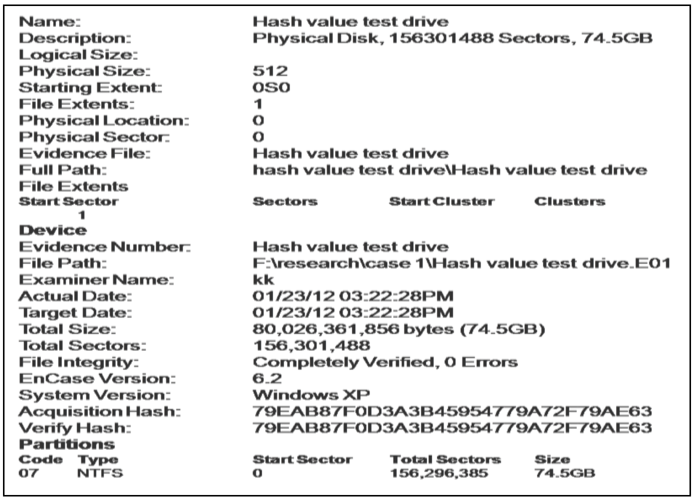


Fig. 2. Hash calculation model

The model is used to create an image of test hard drive connected with write blocker (Ex. Fast Blok), to avoid writing anything by any vulnerable program running on the system. We can use any forensic tool for creating disk image along with hash value. The Encase forensics [9] is a simple but concise tool used in this case study. It saves an image of a hard disk in one file or in segments that may be later on reconstructed. It calculates MD5 hash values and confirms the integrity of the data before closing the files. Now the raw image created by Encase program is used for analysis and examination purpose. While extracting data from raw image the Encase program also verifies previous generated hash value and creates summary/report for forensic validation and presentation that match finds.

The importance of hash value in digital forensic is illustrates in six different cases has been generated and analyzed below.

1. CASE1. Compute hash value in original: Here we generate hash value of original test evidence disk, which contains some suspect files/data for forensic analysis and also verifies the hash value after imaging/acquisition, report shown below in figure 3.

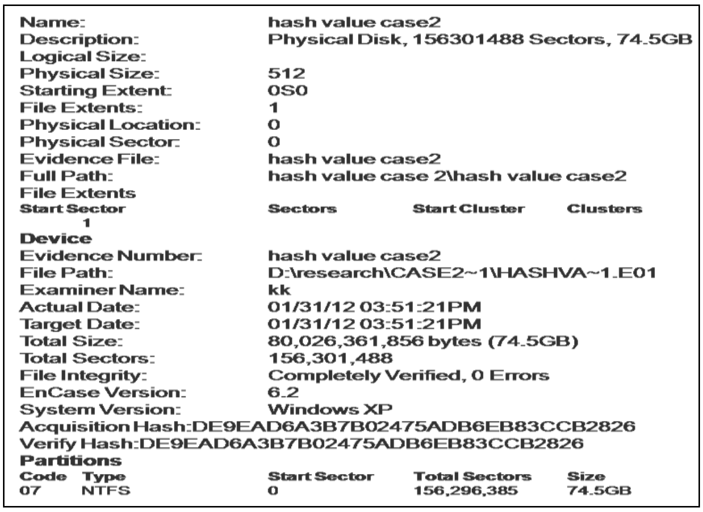


```
Name:                  Hash value test drive
Description:           Physical Disk, 156301488 Sectors, 74.5GB
Logical Size:
Physical Size:         512
Starting Extent:       0S0
File Extents:          1
Physical Location:     0
Physical Sector:       0
Evidence File:         Hash value test drive
Full Path:             hash value test drive\Hash value test drive
File Extents
Start Sector           Sectors        Start Cluster     Clusters
     1
Device
Evidence Number:       Hash value test drive
File Path:             F:\research\case 1\Hash value test drive.E01
Examiner Name:         kk
Actual Date:           01/23/12 03:22:28PM
Target Date:           01/23/12 03:22:28PM
Total Size:            80,026,361,856 bytes (74.5GB)
Total Sectors:         156,301,488
File Integrity:        Completely Verified, 0 Errors
EnCase Version:        6.2
System Version:        Windows XP
Acquisition Hash:      79EAB87F0D3A3B45954779A72F79AE63
Verify Hash:           79EAB87F0D3A3B45954779A72F79AE63
Partitions
Code   Type            Start Sector   Total Sectors   Size
07     NTFS            0              156,296,385     74.5GB
```

Fig. 3. Case1 report

2. CASE2. Add any file in test drive and check hash with original: In this case the experiments generated the report shown in fig 4. The impact of adding any file in the test drive by mistake or by concern, correspondingly the hash value is verified with the original. The hash value differs from the original/actual evidence drive.

Original hash value: 79EAB87F0D3A3B45954779A72F79AE63
New hash value: DE9EAD6A3B7B02475ADB6EB83CCB2826

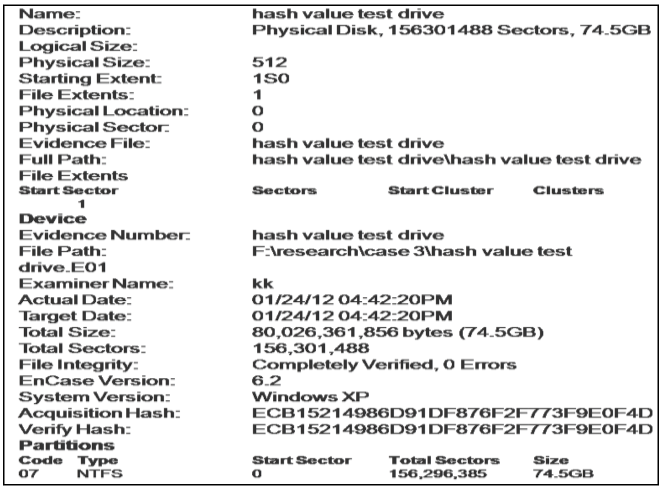


```
Name:                  hash value case2
Description:           Physical Disk, 156301488 Sectors, 74.5GB
Logical Size:
Physical Size:         512
Starting Extent:       0S0
File Extents:          1
Physical Location:     0
Physical Sector:       0
Evidence File:         hash value case2
Full Path:             hash value case 2\hash value case2
File Extents
Start Sector           Sectors        Start Cluster     Clusters
     1
Device
Evidence Number:       hash value case2
File Path:             D:\research\CASE2~1\HASHVA~1.E01
Examiner Name:         kk
Actual Date:           01/31/12 03:51:21PM
Target Date:           01/31/12 03:51:21PM
Total Size:            80,026,361,856 bytes (74.5GB)
Total Sectors:         156,301,488
File Integrity:        Completely Verified, 0 Errors
EnCase Version:        6.2
System Version:        Windows XP
Acquisition Hash:DE9EAD6A3B7B02475ADB6EB83CCB2826
Verify Hash:DE9EAD6A3B7B02475ADB6EB83CCB2826
Partitions
Code   Type            Start Sector   Total Sectors   Size
07     NTFS            0              156,296,385     74.5GB
```

Fig. 4. Case2 report

3. CASE3. Remove any file from the test drive and compare hash with original: In this case the experiments illustrate if any single or multiple files deleted from the evidence disk, correspondingly hash value of the drive is generated and compare with original. The difference of hash value generated report is shown in fig 5.

Original hash value: 79EAB87F0D3A3B45954779A72F79AE63
New hash value: ECB15214986D91DF876F2F773F9E0F4D



```
Name:                  hash value test drive
Description:           Physical Disk, 156301488 Sectors, 74.5GB
Logical Size:
Physical Size:         512
Starting Extent:       1S0
File Extents:          1
Physical Location:     0
Physical Sector:       0
Evidence File:         hash value test drive
Full Path:             hash value test drive\hash value test drive
File Extents
Start Sector           Sectors       Start Cluster    Clusters
      1
Device
Evidence Number:       hash value test drive
File Path:             F:\research\case 3\hash value test
drive.E01
Examiner Name:         kk
Actual Date:           01/24/12 04:42:20PM
Target Date:           01/24/12 04:42:20PM
Total Size:            80,026,361,856 bytes (74.5GB)
Total Sectors:         156,301,488
File Integrity:        Completely Verified, 0 Errors
EnCase Version:        6.2
System Version:        Windows XP
Acquisition Hash:      ECB15214986D91DF876F2F773F9E0F4D
Verify Hash:           ECB15214986D91DF876F2F773F9E0F4D
Partitions
Code   Type            Start Sector  Total Sectors    Size
07     NTFS            0             156,296,385      74.5GB
```

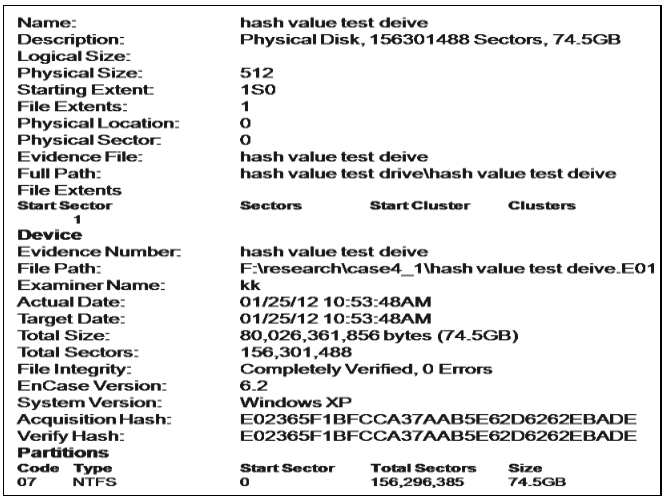
Fig. 5. Case3 report

4 CASE4. Modify any file:

This case is totally differing from the previous two cases of add and remove files from the drive, describes in two sub cases below:

- **Case 4.1. Add some contents in any file and check hash with original:** Here we are demonstrates the case when small amount of data is added in any file. The report generated with hash value shows in fig 6. The comparison of hash value with the original is also mention below.

Original hash value: 79EAB87F0D3A3B45954779A72F79AE63
New hash value: E02365F1BFCCA37AAB5E62D6262EBADE

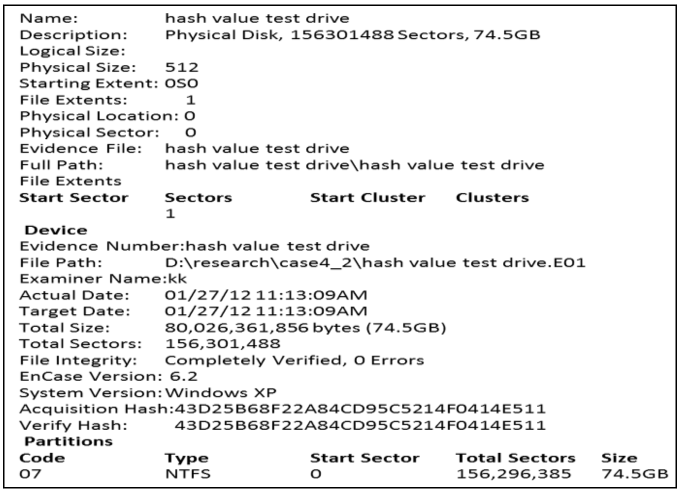


```
Name:                  hash value test deive
Description:           Physical Disk, 156301488 Sectors, 74.5GB
Logical Size:
Physical Size:         512
Starting Extent:       1S0
File Extents:          1
Physical Location:     0
Physical Sector:       0
Evidence File:         hash value test deive
Full Path:             hash value test drive\hash value test deive
File Extents
Start Sector           Sectors       Start Cluster    Clusters
      1
Device
Evidence Number:       hash value test deive
File Path:             F:\research\case4_1\hash value test deive.E01
Examiner Name:         kk
Actual Date:           01/25/12 10:53:48AM
Target Date:           01/25/12 10:53:48AM
Total Size:            80,026,361,856 bytes (74.5GB)
Total Sectors:         156,301,488
File Integrity:        Completely Verified, 0 Errors
EnCase Version:        6.2
System Version:        Windows XP
Acquisition Hash:      E02365F1BFCCA37AAB5E62D6262EBADE
Verify Hash:           E02365F1BFCCA37AAB5E62D6262EBADE
Partitions
Code   Type            Start Sector  Total Sectors    Size
07     NTFS            0             156,296,385      74.5GB
```

Fig. 6. Case4.1 report

- **Case 4.2. Remove some contents from any file and compare hash with original:** Here we are demonstrates the case when some portion of data is erased from any file. The generated report with hash value shown in fig 7 below.

Original hash value: 79EAB87F0D3A3B45954779A72F79AE63
New hash value: 43D25B68F22A84CD95C5214F0414E511

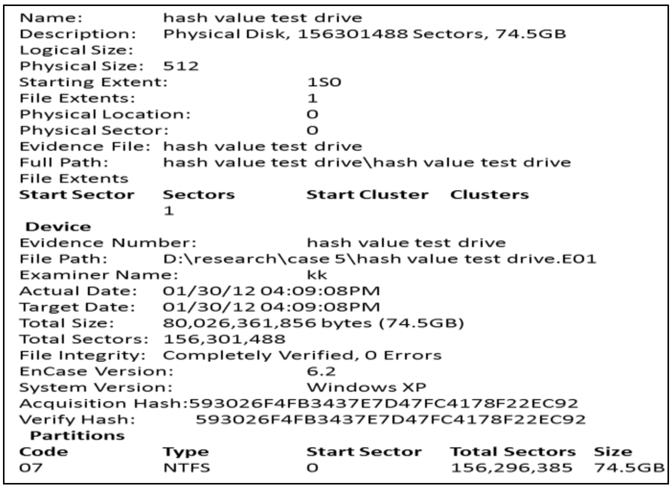


```
Name:            hash value test drive
Description:     Physical Disk, 156301488 Sectors, 74.5GB
Logical Size:
Physical Size:   512
Starting Extent: 0S0
File Extents:      1
Physical Location: 0
Physical Sector:   0
Evidence File:   hash value test drive
Full Path:       hash value test drive\hash value test drive
File Extents
Start Sector     Sectors          Start Cluster    Clusters
                 1
 Device
Evidence Number:hash value test drive
File Path:       D:\research\case4_2\hash value test drive.E01
Examiner Name:kk
Actual Date:     01/27/12 11:13:09AM
Target Date:     01/27/12 11:13:09AM
Total Size:      80,026,361,856 bytes (74.5GB)
Total Sectors:   156,301,488
File Integrity:  Completely Verified, 0 Errors
EnCase Version: 6.2
System Version:Windows XP
Acquisition Hash:43D25B68F22A84CD95C5214F0414E511
Verify Hash:      43D25B68F22A84CD95C5214F0414E511
 Partitions
Code             Type             Start Sector     Total Sectors    Size
07               NTFS             0                156,296,385      74.5GB
```

Fig. 7. Case4.2 report

5. CASE5. Shifting some contents from one file to another and check hash with original: Sometimes the content of one file is shifted to another file rather than the whole file the results of hash value comparison with original shows in figure 8 below.

Original hash value: 79EAB87F0D3A3B45954779A72F79AE63
New hash value: 593026F4FB3437E7D47FC4178F22EC92

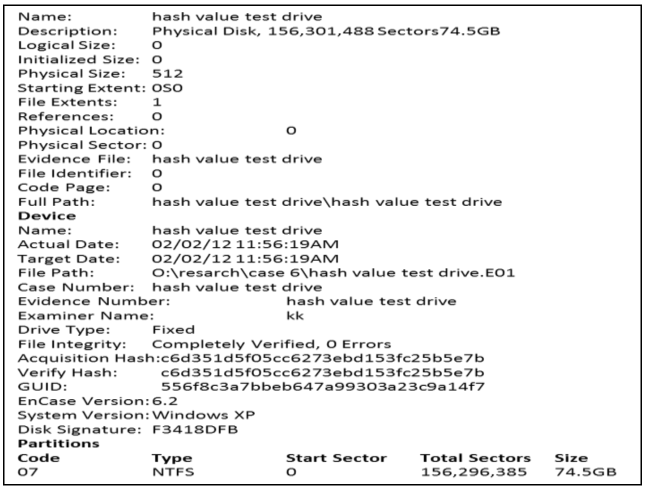


```
Name:            hash value test drive
Description:     Physical Disk, 156301488 Sectors, 74.5GB
Logical Size:
Physical Size:   512
Starting Extent:                  1S0
File Extents:                     1
Physical Location:                0
Physical Sector:                  0
Evidence File:   hash value test drive
Full Path:       hash value test drive\hash value test drive
File Extents
Start Sector     Sectors          Start Cluster    Clusters
                 1
 Device
Evidence Number:                  hash value test drive
File Path:       D:\research\case 5\hash value test drive.E01
Examiner Name:                    kk
Actual Date:     01/30/12 04:09:08PM
Target Date:     01/30/12 04:09:08PM
Total Size:      80,026,361,856 bytes (74.5GB)
Total Sectors:   156,301,488
File Integrity:  Completely Verified, 0 Errors
EnCase Version:                   6.2
System Version:                   Windows XP
Acquisition Hash:593026F4FB3437E7D47FC4178F22EC92
Verify Hash:        593026F4FB3437E7D47FC4178F22EC92
 Partitions
Code             Type             Start Sector     Total Sectors    Size
07               NTFS             0                156,296,385      74.5GB
```

Fig. 8. Case5 report

6. CASE6. Update some contents in existing file and compare hash: The case is related where financial and accounting data is much more valuable than other contents of the disk. Sometimes the suspect modified only numerical contents of the data files. The experiments demonstrate here is to check whether the hash value generated in forensic tool differ or match with any previous case shows in figure 9 below.

Original hash value: 79EAB87F0D3A3B45954779A72F79AE63
New hash value: C6D351D5F05CC6273EBD153FC25B5E7B



```
Name:             hash value test drive
Description:      Physical Disk, 156,301,488 Sectors74.5GB
Logical Size:     0
Initialized Size: 0
Physical Size:    512
Starting Extent: 0S0
File Extents:     1
References:       0
Physical Location:                  0
Physical Sector: 0
Evidence File:    hash value test drive
File Identifier:  0
Code Page:        0
Full Path:        hash value test drive\hash value test drive
Device
Name:             hash value test drive
Actual Date:      02/02/12 11:56:19AM
Target Date:      02/02/12 11:56:19AM
File Path:        O:\resarch\case 6\hash value test drive.E01
Case Number:      hash value test drive
Evidence Number:                    hash value test drive
Examiner Name:                      kk
Drive Type:       Fixed
File Integrity:   Completely Verified, 0 Errors
Acquisition Hash:c6d351d5f05cc6273ebd153fc25b5e7b
Verify Hash:       c6d351d5f05cc6273ebd153fc25b5e7b
GUID:              556f8c3a7bbeb647a99303a23c9a14f7
EnCase Version:6.2
System Version:Windows XP
Disk Signature:  F3418DFB
Partitions
Code              Type              Start Sector      Total Sectors    Size
07                NTFS              0                 156,296,385      74.5GB
```

Fig. 9. Case6 report

V. CONCLUSION & FUTURE WORK

The role of hash value is demonstrated in this research work with the help of different cases involved in data tampering is analyzed and verified. The focus of this research is strongly on the hash value of whole digital drive not a single file. The aim of this work to show that even if a small modification happen in digital evidence, detected in the hash value.

Given a different heuristic, it would be interesting to apply this technique in future to other file systems other than that of Windows and to compare results.

VI. ACKNOWLEDGMENT

The authors thankful to Mr. Vivek Kumar Forensic Analyst Central Forensic Science Laboratory, Chandigarh, for their useful contribution and Software/Hardware support for this case study.

REFERENCES

[1]. S. Ardisson. Producing a Forensic Image of Your Client's Hard Drive? What You Need to Know, Qubit, 1, pp. 1-2, 2007.

[2]. "Digital signature standard," FIPS PUB 186-2, 2000. Available from: www.csrc.nist.gov/publications/fips/fips186-2/fips186-2-change1.pdf

[3]. G. Palmer. A road map for digital forensic research. Report From the First Digital Forensic Research Workshop (DFRWS), August 2001.

[4]. M. Alazab, S. Venkatraman & P. Watters. Digital forensic techniques for static analysis of NTFS images, Proceedings of ICIT2009, Fourth International Conference on Information Technology, IEEE Xplore, 2009.

[5]. M. Reith, C. Carr, & G. Gunsch. An examination of digital forensic models, International Journal of Digital Evidence, 1, pp.1-12, 2002.

[6]. Nelson, B. Phillips, A. Enfinger, F. Steuart, C. Guide to Computer Forensics and Investigations: Third Edition, Course Technology, 2008.

[7]. Ronald L. Rivest. The MD5 message-digest algorithm, IETF RFC 1321, 1992. Available from www.ietf.org/rfc/rfc1321.txt

[8]. Secure hash standard, FIPS PUB 180-1, 1995. Available from: www.itl.nist.gov/fipspubs/fip180-1.htm

[9]. Guidance Software Inc. Encase forensics. http://www.guidancesoftware.com.