International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2019): 7.583

An Empirical Study Illustrating Effects on Hash Value Changes in Forensic Evidence Appreciation

Akash A. Thakar[1], Prof. B. V. Patel[2], Dr. Kapil Kumar[3]

[1]Research Scholar, Gujarat University, India
thakarakash[at]gmail.com

[2]Professor, Gujarat University, India
patelbaldev65[at]yahoo.in

[3]Associate Professor, Gujarat University, India
kkforensic[at]gmail.com

Abstract: ***The process of digital forensics consists of the collection, preservation, analysis, and presentation of digital evidence in the court of law. During this process integrity of evidence plays a major role in the admissibility of evidence. A hash value is the only way to determine the integrity of evidence. MD-5 and SHA-1 are used commonly functions in the forensic fraternity. Many types of research have been done on the collision of MD-5 and SHA-1 hash function. This illustrates that two different files having different content can have the same hash value. It is demonstrated that without manipulating any content of a file, the hash value may change. Considering this fact author has identified parameters that affect changes in the hash value. The author set-up different scenarios and elucidated in what manner hash value may vary without alteration of content and also compared resultant hash value in .doc and .txt files respectively in the same condition.***

Keywords: Hash Value, Hash Function, MD-5, SHA-1, Data Integrity

1. Introduction

The cryptographic hash function converts any arbitrary size of data into a fixed-length. The resultant value is called as Hash Value [1]. A hash value is an important aspect in the forensic field pretentiously to demonstrate the integrity of any digital evidence. Presently widely used hash functions for data integrity in the forensic fraternity are MD-5 and SHA family [2]. MD-5 was published by Ronald Rivest which is 128 bits in 1991 and was specified in 1992 [3]. SHA was developed by US National Security Agency (NSA) and published by the National Institute of Standard and Technology (NIST) [4]. The Hash value can be thought of as a digital fingerprint of any file. It changes even after a small change happens in any file. The only way to determine the data integrity of digital evidence is its hash value. Numerous times, it is observed that changes could be unintended nonetheless evidence could be eliminated. In this paper an illustrative attempt is made to detect the parameters that affects the changes in resulting hash value of any given forensic exhibit. The objective of the study is to determine the role of metadata in the changes of hash value and establish technical and scientific purpose behind the changes in different document file format irrespective of any alteration in content of any of the file.

2. Literature Review

Rasjid et. al [2] has worked on collision of the different hash function. They have taken MD-5 and SHA family because of their wide use in general. They found that two different files with different contents can have the same hash value. It is called a hash collision. They conclude that because of hash collision, the question is raised about the integrity of evidence. In a blog, the author has practically presented a demo of MD-5 hash collision. He took 2 files with different contents and had the same MD-5 hash value. He also developed a tool by which the same MD-5 hash value of different files can be generated. The tool works on Wang and Yu's attack known as the chosen prefix collision method [5].

Kessler [6] has identified the impact of collision in SHA-1 in the digital forensic image. He took 2 different files with different content but having similar SHA-1 hash value. He then copied it in external media with other similar files and imaged it. It is observed that two files having similar hash values give different hash values when the source is imaged using a forensic imaging tool.

Wang et al [7] proposed a modified MD-5 hash function which is of variable length. The computing procedure of the modified hash is almost the same as MD-5. The proposed algorithm of modified MD-5 is claimed as more secure and flexible.

Umesh et al [8] has done a comparative study on different hash algorithms. They have taken 3 different hash functions. MD-5, modified MD-5, and SHA-1. They compared all three hash functions with different parameters like timing, security, space, and hash code. They concluded that there is a need for the development of a secure hash algorithm that is efficient too.

Raychaudhuri et al [9] has worked on how file system affects changes in hash value after the source is imaged with forensic tools. They took an image of 4GB USB thumb drive by FTK imager where write blocker device is used. In another scenario, they took the same image without a write blocker. A change in hash value is noted. In conclusion, they

compared system files and metadata files of both images.

Ball [10] in his blog has tried to explain how hash value is changed in many circumstances. He suggested that there are two metadata, one is application metadata and the other is system metadata. If application metadata is changed, the hash value of the file will also change even though the content is not altered. He also gave some examples where hash value will change without altering actual content.

3. Material and Methods

This section shows different conditions of files and their effect on the hash value. This experiment is conducted under the NTFS file system. Windows 10 is used as the host operating system. MS-office 2019 is used for .doc files. Notepad application is used for creating .txt documents. The Hash calculator tool is used to calculate the hash value of a given sample. The following parameters are considered for the testing.

- Twenty .doc files and twenty .txt files are created in different directories of the same computer. Then same content is added in 10 files and the hash value is taken.
- Content of a .doc file when copied and pasted in another .doc file. Likewise, the content of one .txt document is copied and pasted into another .txt document. In both these scenarios, the Hash values were obtained and compared with.
- As it's a known fact that the Hash value changes after a minor change happen in the content of a file. When the content of a .doc and .txt document is changed, the hash value is changed as per the above rule. The same content is added whatever is removed from the .doc and .txt file. The hash value is compared with its original file.
- A .doc file and .txt document are printed. To prove the integrity of that file is very important for forensic purposes Hash value is measured before printing and after the printing of .doc and .txt file. No content is changed in both files.
- A .doc file is encrypted with a password and saved. Hash value noted before and after password protection.
- The hash value of only content and a hash value of .doc and .txt file is compared.
- All files are copied in the thumb drive and each file is pasted in different computers.

4. Results and Discussion

Ball [10] explained in his blog that there are two types of metadata. One is system metadata which resides outside of the file and the other is application metadata which resides inside the file. Considering that fact, the below observations can be concluded. Where he mentioned that the hash value will change after printing the .doc file. But in the experiment, it is observed that the hash value does not change after printing a .doc file

- When twenty .doc files and twenty .txt files are created and their hash value is noted without adding any content, the hash value of all files will be the same. Moreover, the hash value of the .doc file and .txt file is also the same. When the same content is added in .doc files and .txt files, the hash value of .doc files is different but the hash value of .txt files is the same. The reason for that is when content is added in a .doc file, application metadata is also changed which resides inside the .doc file and it affects a change of hash value. While notepad does not store any application metadata. So, the hash value of the .txt file would not change.
- When the content of a .doc file is copied and pasted into another .doc file, the hash value of both files will be different because it affects application metadata. When the content of a .txt file is copied and pasted into other .txt files, the hash value is found the same.
- When the content of the file is changed, the hash value is also changed. When the content of the .doc file is again changed with the original data, the hash value is different. When the same experiment is conducted for a .txt file, the hash value is the same.
- Forensic reports are generally sent in the form of printed documents. When a .doc file is printed, the hash value of the files does not change. .Doc file must store last printing information in application metadata, but it is observed that last printing time is not stored in .doc file. While the .txt document is printed, the hash value remains the same.
- When the password is given to the .doc file, its hash value will change and when the password is removed, the hash value will be again different because it resides inside the file.
- When the hash value of the content typed in .doc and .txt file is measured and compared hash value with respective file, it is observed that hash value of .txt file is found same and whereas the hash value of .doc file is found different.
- When all files are copied and pasted in different locations or even different machines, its system metadata will change which does not affect on resultant hash value, so the hash value is found the same for all files.

Table 1 Results of parameters considered for testing Hash Value variations

Compared files	*variation in the hash value*
The Hash value of all newly created .doc file	No
The Hash value of all newly created .txt file	No
The Hash value of a newly created .doc and .txt file	No
The same content is added in the .doc file	Yes
The same content is added in a .txt file	No
Content of a .doc file is copied and pasted in another .doc file	Yes
Content of a .txt file is copied and pasted into another .txt file	No
Changing in content and again restored in .doc file	Yes
Changing in content and again restored in a .txt file	No
Printing a .doc file	No
Printing a .txt file	No
Encrypting .doc file with password	Yes
Comparing content with .doc file	Yes
Comparing content with a .txt file	No
Files are copied in a thumb drive and pasted in other machines	No

5. Conclusion

This experiment concludes that there are two metadata for .doc files. One is system metadata and the other is

Application metadata. When application metadata is changed, the hash value of the respective file will also change. But system metadata like timestamp or name and location of the file will not affect the hash value. That is the reason when any content is modified in a .doc file and saved and original content is again restored, the hash value will be different. In the .txt file, such application metadata is not found. That is the reason when content is modified and again original content is restored, the hash value will not change for a .txt file. Ball in his blog writes that after printing a .doc file, the hash value will change, but in this experiment, it is observed that the hash value does not change even after a .doc file is printed. So forensically hash value is the only mean by which data integrity is measured. This paper suggests that how hash value can be changed even though the content is not altered.

6. Future Scope

This experiment is conducted on the NTFS file system. The same experiment can be conducted for the various file system. This experiment can also be conducted for mobile phones which can be crucial nowadays. In this paper, the author has worked on .doc and .txt files. The same experiment can be done with images, audio, pdf, video, or other files. Other parameters can also be considered to carry forward this work.

References

[1] "Hash_function @ en.wikipedia.org." [Online]. Available: https://en.wikipedia.org/wiki/Hash_function.
[2] Z. E. Rasjid, B. Soewito, G. Witjaksono, and E. Abdurachman, "A review of collisions in cryptographic hash function used in digital forensic tools," *Procedia Computer Science*, vol. 116. pp. 381–392, 2017, doi: 10.1016/j.procs.2017.10.072.
[3] "MD5 @ en.wikipedia.org." [Online]. Available: https://en.wikipedia.org/wiki/MD5.
[4] T. Fisher, "what-is-sha-1-2626011 @ www.lifewire.com." [Online]. Available: https://www.lifewire.com/what-is-sha-1-2626011.
[5] X. Wang, "MD5 Collision Demo," *Program*, no. March 2005, pp. 1–4, 2009.
[6] G. Kessler, "The Impact of SHA-1 File Hash Collisions On Digital Forensic Imaging: A Follow-up Experiment," *J. Digit. Forensics, Secur. Law*, vol. 11, no. 4, 2016, doi: 10.15394/jdfsl.2016.1433.
[7] M. J. Wang and Y. Z. Li, "Hash Function with Variable Output Length," *Proc. - 2015 Int. Conf. Netw. Inf. Syst. Comput. ICNISC 2015*, pp. 190–193, 2015, doi: 10.1109/ICNISC.2015.22.
[8] U. Gandhi, "a Review Towards Various Hash Algorithms and Their Comparative Analysis," *Int. Res. J. Eng. Technol.*, vol. 4, no. 2, pp. 1316–1319, 2017, [Online]. Available: https://irjet.net/archives/V4/i2/IRJET-V4I2257.pdf.
[9] K. Raychaudhuri and M. G. Christopher, "An Empirical study to determine the role of file-system in modification of hash value An Empirical study to determine the role of file-system in modification of hash value," vol. 3, no. 1, pp. 24–41, 2020.
[10] C. Ball, "A Hash of It," pp. 1–9, 2012, [Online]. Available: https://craigball.net/2012/03/05/a-hash-of-it/.

Author Profile

Akash Thakar is a research scholar at Gujarat University and Certified Ethical Hacker, having sound knowledge of Digital Forensics and VAPT. He has completed his master's in Forensic Science from Gujarat University and at present pursuing a Doctorate of Philosophy in the field of Digital Forensics from Gujarat University. He has taught various subjects to the students of UG and PG courses like Computer Forensics, Digital Evidence, Network Forensics, Malware Analysis, Advance Digital Forensics, OS Security and Forensics etc. His research area is Digital Forensic Investigation Process and Memory Forensics. He has published papers in various conferences.

Prof. B. V. Patel is a senior professor at Gujarat University. He has guided various students in his research work and published various research articles in various journals.

Dr. Kapil Kumar is an Associate Professor at Gujarat University, Ahmedabad. He is an experienced professor with a demonstrated history of working in the education field. His expertise is in Research, Forensic Analysis, E-Learning, Criminal Justice, and Teaching. Strong education professional with a doctorate and a Master's Degree focused in Forensic Science, specialized in cyber forensics, questioned documents examination and Forensic ballistics.