Hash Equations and Cryptographic Applications on Forensic

M.T. GÜNEŞER[1] and H.H. OKUYUCU[2]

[1] Karabuk University, Karabuk/Turkey, mtguneser@karabuk.edu.tr
[2]Forensic Medicine Institution Ankara/Turkey, okuyucuhacihasan@gmail.com

***Abstract* - Forensic computing is a new branch of science created to facilitate the decision of the investigator in the case by examining the evidence obtained in the information systems. In forensic computing, all devices in hand have a data summary value (hash). Hash is a numerical value and unique given by the investigator so that evidence can be considered as evidence in the court by preventing the integrity of the evidence. In accordance with Article 134 of the Criminal Procedure Law (CMK), this numerical value, which is unique to the evidence, should not be altered in any way. If it changes, the evidence in question is no longer evidence, and it will not be taken into consideration by the investigating authority even if it gives a clue about the suspect.**

In this study, we expressed hash applications on forensic and the calculation methods of this numerical value which has great importance while judgement.

***Keywords* – Hash functions, Forensic.**

I. INTRODUCTION

In recent years, interest in hash functions has been increasing due to the fact that they are used as infrastructure in virtual money applications. Thanks to steep rise on information technologies and ability of faster processors, the use of Hash equations is diversified like forensic applications. A Hash value obtained by Hash equations can also be referred to as a data summary value in forensic computing [1-3].

When the world has become a globally widespread use of technology, most of the crimes are mostly used in technology and information equipment, or criminals leave behind information-based evidence. Developments in the area of forensic informatics have gained great importance especially for combating cyber-crime, proving crimes and ensuring justice [4-5].

In the investigation processes, to determine whether there is an offense on the evidence, a copy of the data is taken firstly by considering the software and hardware status of the data. This process is called as image acquisition [6-7]. All researches and evaluations are carried out on the image obtained in order to prevent any deterioration and loss on the actual evidence. So, it must be ensured that the image does not deteriorate with the actual evidence. Therefore, at least one sample is taken from almost all parts of the raw data and a numerical hash function specific to that evidence is obtained after applying mathematical and logical algorithms [8-10].

Not to discuss the decisions of the judges, while research process, protecting the health of the evidences and being sure any change deliberately or accidentally has utmost importance. The use of hash functions to ensure that there is no change between the copies and originals of digital evidence is still being discussed. A reproducible or replicable hash function will cause to be discussed the reliability of the digital evidence. The studies about approving whether the designed hash equations are reproducible. Regarding some of these studies, the reliability of some algorithms has been eliminated. So, the use of that algorithms is prohibited in forensic informatics applications [10-11].

Forensic informatics is seen as a rapidly developing science in the last few decades in the world and in our country. In Turkey, the teams of experts on forensics has been working in Forensics Specialized Office, where is in the Institute of Forensic Medicine within the Ministry of Justice. Thanks to the technical equipment and software available in the office, expert reports are prepared on the evidence from the courts. In order to maintain the reliability of the evidences before starting these processes, the identification number is defined by the hash functions. This value is calculated by using non-native programs such as Ditto DX, FTKImager, Multi-Hasher [12].

In this study, the hash functions used in forensic computing are examined and an algorithm has been proposed for use in forensic computing.

II. DESIGNING HASH EQUATIONS

Various Hash functions can be described as seen on Table 1., but basically two types of hash functions are used on forensic informatics, called Message Digest (MD) and Secure Hash Algorithm (SHA). In the process, efforts were made to close the security gaps in order to increase the sensitivity to prove that digital evidence has not changed. Actual active versions are known as MD5 and SHA1 [13-15].

The use of Hash functions is not limited to checking whether data is changed only in physical copying. It can also be controlled by the same method whether the information delivered over a remote client is changed during the transfer. Especially, changes in the content of e-mail and add-ons are also in the interest of forensic informatics. Today, the importance of reliability of data is better understood by considering many important data are being used in common scientific studies overseas under the big data concept. In addition to the accidental change of data, manipulation-oriented interventions should be considered in this context [13-16].

As Hash functions are single-sided, it is not possible to reach the data again with reverse engineering after being created mathematically [16].

Table 1: This caption is centered.

Type	Code	Bit-qty
DES (Unix)	IvS7aeT4NzQPM	13
MD5 (Unix)	$1$12345678$XM4P3PrKB gKNnTaq G9P0Tk	34
MD5 (APR)	$apr1$12345678$auQSX8Mvzt.tdBi4y 6Xgj5	37
MD5(phpBB3)	$H$9123456785DAERgALpsri.D9z3ht 120	34
MySQL	606717496665bcba	16
MySQL5	E6CC90B878B948C35E92B003C79C 46C58C4AF40	40
MD5	c4ca4238a0b923820dcc509a6f75849b	32
MD%x2	28csedde3d61a041511d3b1866f0636	32
SHA1	356a192b7913b04c545574d18c28d46e 6395428ab	40
SHA256 (Unix)	$5$12345678$jBWLgeYZbSvREnuBr5 s3gp13vqiKSNK1rkTk9zYE1v0	55

III. Methods of Hash Calculations and Obtaining Hash Values

SHA1 is one of well-known Hash equation. If 32 bit of SHA1 value is wanted to be generated, the procedure will be followed by using HEXEDECIMAL bites. While the image acquisition process, the evidence file is divided in 16 independent parts. And the sample data of these parts are examined with the algorithm for 16 times to get first bit of Hash value. As seen on Table 1., SHA1 has 40 bit for the Hash value, so this process is repeated 40 times to achieve the whole Hash value [17-19].

MD5 is used as another method to calculate Hash value in forensic informatics. Obtaining Hash value via MD5 method is represented on Figure 1. MD5 has a total of 64 operations, applied 16 times from this cycle consisting of 4 rounds. F in this cycle is a non-linear function. M_i represents a 32-bits message, and K_i represents a constant generated for each process. MD5 processes a variable length message as a fixed length output of 128 bits. The input message is divided into 512-bits block pieces of sixteen 32-bits words. Then the data length can be divided into 512 bits by adding one extension bit as 1 to end of the message. Balance of message is completed by 0 until completing 512 bits except 64 bits, which is real message. That message is added to the package with 2^{64} modes [20].

Main MD5 algorithm is divided into four 32 bits words called A,B, C and D, which are consist of 128 bits completely. Those values are started by constants. Then main algorithm is changed every 512 bits of block to generate the value bits. The processing of a message block consists of four similar stages called rounds, each round consists of a non-linear function, modular addition operation and bit-to-left scrolling, and 16 rounds exist to complete the Hash value in the process. F function, which is changed for every round, has four different eventuality as seen on Equation 1.-4..

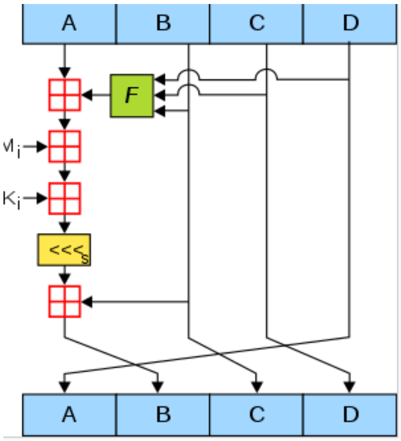

Figure 1: MD5 Hash function.

$$F\ (B,C,D) = (B \wedge C) \vee (B' \wedge D) \tag{1}$$

$$G\ (B,C,D) = (B \wedge D) \vee (C \wedge D') \tag{2}$$

$$H\ (B,C,D) = B \oplus C \oplus D \tag{3}$$

$$I\ (B,C,D) = C \oplus (B \vee D) \tag{4}$$

IV. Conclusion

In recent years, step up of information technologies and new generation microprocessors give an opportunity to solve heavy and confusing mathematical equations via computational analysis methods. Hash equations are also one of well-known confusing mathematical systems. It has new security solutions in a very unusual field of use. Forensic is also very important issue to solve the crimes and keep the evidences in secure. So, in this study we investigated use of Hash equations on forensic. We proposed detail procedures of actual active versions are known as MD5 and SHA1 for forensic informatics. Because of reproducibility risks of SHA1, we preferred MD5 for forensic applications.

Main MD5 algorithm consists of 32 characters in the Hash value and every character can be obtained a chain of specific solution of equations by using likelihood of four different function. Every character is generated by 512bits of block by using that algorithm.

In the next study is continuing to obtain a new algorithm. Using some of optimization technics may give an opportunity solve the equations more rapidly.

REFERENCES

[1] M. Ciampa, *Security+ 2008 in Depth*. Boston: Jenson Books Inc., 2009.

[2] C. R, Dougherty, "Vulnerability Note VU#836068 MD5 vulnerable to collision attacks," Ph.D. dissertation, Dept. Software Eng., CERT Carnegie Mellon Univ., 2008.

[3] J. Black, M. Cochran, T. Highland, "A Study of the MD5 Attacks: Insights and Improvements*," in Conf. Rec. 2006 13th international conference on Fast Software Encryption,* pp. 262-277.

[4] G. Hirshman. (2007,). Further Musings on the Wang et al. MD5 Collision: Improvements and Corrections on the Work of Hawkes, Paddon, and Rose. *IACR Cryptology*. [Online]. pp. 1-5. Available: https://pdfs.semanticscholar.org/7493/616b51c7e4fabeb7940a531c0b184e0eb1b0.pdf?_ga=2.242810370.1873242906.1544992439-1088478693.1544992439.

[5] B. Fox, *Fast MD5 and MD4 Collision Generators.* USA: Mc Graw Hill, 2013.

[6] A. Lenstra, X. Wang and B. Weger, "Colliding X.509 Certificates," *Cryptology ePrint Archive Report*. [Online]. *2005(067).* pp. 1-5. Available: https://eprint.iacr.org/2005/067.pdf

[7] V. Klima. (2005, March). Finding MD5 Collisions – a Toy For a Notebook. *Cryptology ePrint Archive Report*. [Online]. *2005(075)*. pp. 1-7. Available: https://eprint.iacr.org/2005/075.pdf

[8] V. Klima. (2006, April). Tunnels in Hash Functions: MD5 Collisions Within a Minute. *Cryptology ePrint Archive Report*. [Online]. *2006 (105).* pp. 1-17. Available: https://eprint.iacr.org/2006/105.pdf

[9] N. Shachtman, "Code Cracked! Cyber Command Logo Mystery Solved," *Wired News,* [Online]. Available: https://www.wired.com/2010/07/code-cracked-cyber-command-logos-mystery-solved/

[10] M. Stevens. (2012, January). Single-block collision attack on MD5. *Cryptology ePrint Archive Report.* [Online]. *2012 (040).* pp. 1-11. Available: https://eprint.iacr.org/2012/040.pdf

[11] R. L. Rivest. (1992, April). The MD5 Message-Digest Algorithm. *Internet Engineering Task Force of MIT Laboratory for Computer Science.* [Online]. *1992(1321).* pp. 1-21. Available: https://www.ietf.-org/rfc/rfc1321.txt

[12] D. Knuth, *The Art of Computer Programming, Volume 3.* Boston: Addison Westley, 1997.

[13] H. Dobbertin, (1996, June). The Status of MD5 After a Recent Attack. *The Technical Newsletter Of Rsa Laboratories*. [Online]. *3(2).* pp. 1-6. Available: ftp://ftp.arnes.si/packages/crypto-tools/rsa.com/cryptobytes-/crypto2n2.pdf.gz

[14] S. Turner. (2011, March) Updated Security Considerations for the MD5 Message-Digest and the HMAC-MD5 Algorithms. *Internet Engineering Task Force of MIT Laboratory for Computer Science.* [Online]. 2011(6151). pp. 1-7. Available: https://tools.ietf.org/html/rfc6151

[15] M. M. J. Stevens, "On Collisions for MD5," M. Sc. dissertation, Dept. Math. And Comp. Sci., Eindhoven Univ. of Tech., Eindhoven, 2007.

[16] M. Stevens, A. Lenstra and B. Weger. (2012, July). Chosen-prefix collisions for MD5 and applications. *International Journal of Applied Cryptography*. [Online]. 2(4). pp. 322-359. Available: https://dl.acm.-org/citation.cfm?id=2338853

[17] A. Banerjee, "Windows Enforcement of Authenticode Code Signing and Timestamping impact on SQL Server," *Microsoft Developer* [Online]. Available: https://blogs.msdn.microsoft.com/sql_server_team/windows-enforcement-of-authenticode-code-signing-and-timestamping-impact-on-sql-server/

[18] R. Sleevi, "Intent to Deprecate: SHA-1 certificates," *Google Developer* [Online]. Available: https://groups.google.com/a/chromium.org/-forum/#!topic/blink-dev/2-R4XziFc7A%5B1-25%5D

[19] *Adli Bilişim İhtisas Dairesi Görevleri,* Adli Tıp Kurumu, Ankara, 2018.