

ADOPTION OF CHAIN OF CUSTODY IMPROVES DIGITAL FORENSIC INVESTIGATION PROCESS

Talib M. Jawad Abbas

College of Information Engineering, Al-Nahrain University, Baghdad, Iraq
talib_altalib@yahoo.com

Received:28/01/2018, Accepted:13/5/2018

Abstract-Chain of custody plays an important role to determine integrity of digital evidence, because the chain of custody works on a proof that evidence has not been altered or changed through all phases, and must include documentation on how evidence is gathered, transported, analyzed and presented. The aims of this work is first to find out how the chain of custody has been applied to a wide range of models of the digital forensic investigation process for more than ten years. Second, a review of the methods on digitally signing an evidence that achieves the successful implementation of chain of custody through answering a few questions "who, when, where, why, what and how", and thus providing digital evidence to be accepted by the court. Based on the defined aims, an experimental environment is being setup to outline practically an acceptable method in chain of custody procedure. Therefore, we have adopted SHA512 for hashing and regarding encryption RSA and GnuGP is applied where according to the defined requirement a combination of these algorithms could be adopted as a practical method.

Keywords: Chain of custody, Digital integrity, Digital forensics, Digital evidence, Digital investigation, Digital Forensics investigation.

I. INTRODUCTION

For the past decade, the number of offenses which relate to computers and other devices has grown and products are needed that can help the application of the law to using computer-based evidence to determine the, who, what, where, when, and how of crimes. As a result, computer and network forensics have evolved to present findings, it is necessary to explain how the evidence is handled and analyzed to demonstrate chain of custody and thoroughness of methods. The forensics aim during the legal process is to detect original and main sources through the chain of custody [1].

This means control of the names of the individuals involved gathering evidence and every person or entity has in custody subsequently, and return the items which were collected or moved, and declaring the agencies, issues, name of suspect or victim related to and a brief description of each item.

We will review a set of digital forensic investigation models/frameworks that have been produced during the past years and then identify the chain of custody processes. Also, this paper focuses on methods for digitally signed an evidence.

The paper is structured as follows: the next section outlines the main terminology used in the field of forensics and review briefly some digital forensic investigation models; in section three we will explain the role of chain of custody in digital forensic investigation process; section four will discuss the various available methods and a practical implementation in a Linux Ubuntu 16.04 environment considering cryptography libraries for digitally signed an evidence. Conclusions and recommendations are given in section five.

II. EXISTING DIGITAL FORENSIC INVESTIGATION MODELS

There have been a number of definitions for digital forensics in the last decade, one of them is "the digital forensics can be defined as the application of science and engineering to addressing legal issues of digital evidence" [2]. The term

digital evidence, plays a central part in digital forensics, A reliable digital data from any incident that outstands and claims the incidents hypothesis defined by Carrier and Spafford [3].

In digital forensics, the preservation, extraction and documentation of digital evidence are closely related to two essential forensic principles; the Chain of Custody and the Order of Volatility. This means that various electronic and computerized devices could be a component that contains digital forensics evidence.

The process of answering to questions regarding digital states and events is called a digital investigation. A series of methods and procedures are adopted by the court of law which eventually leads a special case of digital investigation "digital forensics investigation" to be considered in the court procedures [4]. In Table I we present an evaluation on the implemented methods and frameworks through phases and stages in the chain of custody in a duration of ten years, [5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15]. In this table we review a number of models adopted through digital investigation procedure by different authors. It is obvious that some of these models have implemented the principles of chain of custody specifically and in details while others implemented it partially and undetailed.

TABLE I
 EVALUATION OF CHAIN OF CUSTODY REGARDING FORENSIC MODELS

Code of Model	Name of Digital Forensic Investigation Framework/Model	Places /Phases	,Evaluation, appearance
M2001	,DFRWS Investigative Model	preservation	Implicit
M2002	Abstract Digital Forensic Model	preservation	Implicit
M2003	,End-to-End Digital Investigation (EEDI)	preservation	Explicit
M2004	An Extended Model of Cybercrime Investigations	All phases	Implicit
M2005	Case-Relevance Information Investigation	All,phases	Implicit
M2006	Visual Network Forensic Techniques and Processes	Data validation	Explicit
M2007	Common Process Model and Computer Forensics	Analysis (Duplication Step)	Explicit
M2008	New Digital Forensics Investigation Procedure Model	Disk forensics	Explicit
M2009	Digital Forensic Model based on Malaysian Investigation Process (DFMMIP)	Data acquisition & archive storage	Implicit
M2010	Network Forensic Generic Process Model	preservation	Explicit
M2011	Systematic Digital Forensic Investigation Model	securing the scene & preservatio	Explicit

The results of the study show that ripe chain of custody framework suited does not exist for digital forensic investigation process, mainly because of lack of a real standard.

III. CHAIN OF CUSTODY ROLE IN DIGITAL FORENSICS INVESTIGATION PROCESS

The digital forensics investigation process is managed by wide range of models and frameworks that have been reviewed in section two of this paper. The issues of chain of custody is not concentrated on in lots of designed models, obviously the process flow of information is not presented. The processes of data collection and feeble evidence are two components

not focused on in most of present models. The process for any model includes phases of any digital forensics investigation designed to get the digital evidence.

The term "digital evidence" means a digital data that supports or refutes a hypothesis about digital events or the state of digital data [16]. While the digital data is any type of information e.g. image, voice/video file, text existing on an electronic device such as laptop, desktop, cellular phone, etc., where any data that reveals a strong relation between the victim and cause of the crime is evidence.

Achieving the previous specifications requires a framework and tools to ensure the survival of digital evidence on its nature preserving without changing or altering when dealing with it, these tools or processes are called chain of custody. Securing the evidence from contamination and unauthorized access to the crime scene are handled by the defined protocols from chain of custody, (See Fig.1). In the next section, we will study the methods of identifying the, who, what, when, where, how, and why of each piece of evidence or material that investigator collected during the investigation.

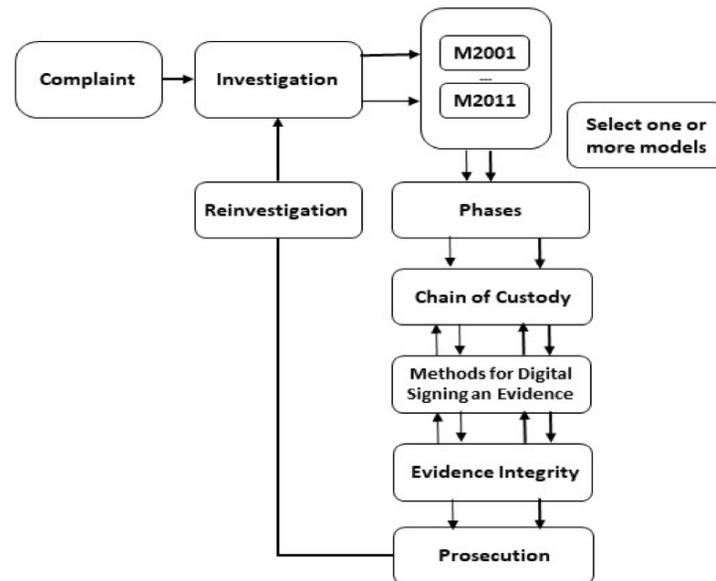


Figure 1: Digital forensic investigation process

IV. METHODS FOR INTEGRITY OF DIGITAL EVIDENCE

To establish chain of custody, a detailed documentation of the evidence treatment is mandatory to be reported from all the individuals who are in contact with it. The aim from this documentation is to answer a number of questions as:

What is the evidence?

How did you get it?

When was it collected?

Who has handled it?

Why did that person handle it?

Where has it travelled, and where was it ultimately stored? It is very important to find out the answer to the above questions. Therefore, the following lines will try to answer each question through the use of appropriate methods.

What/Why. The acceptance and validity of an evidence in the court should be conditional to the chain of custody protocol assuring precisely what was the evidence and why that person handled it. To achieve this, we must calculate a fingerprint of evidence. The algorithm for calculating a fingerprint of this work will be a message digest which is a combination of digits and characters generated for the given input. It is supposed to be the same as many time the procedure is repeated for the same input and changed due to any single bit of change in the input file. Which means for any different file a good message digest algorithm generates a unique output. Depending on the requirements of the organization, number of evidences, files, etc. to be hashed any of the hashing algorithms could be adopted.

There are message digest algorithms such as MD4 (old), MD5, Secure Hash Algorithm (SHA) SHA1, SHA2 family (SHA-224/256, SHA-384/512, HAVAL, and SNEFRU). Now we have a hash value (the combination of characters and numeric value is referred to as the hash value, while hash function means establishing mathematical calculation that generates a numerical value based on the input data) of digital evidence (fingerprint), (see Fig.2).

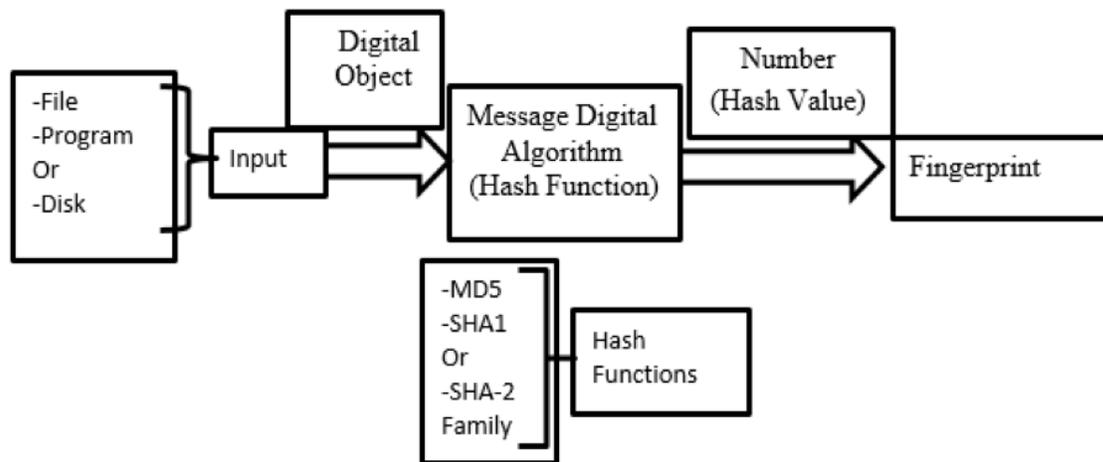


Figure 2: Process of answering the questions of what was the evidence and why that person handled it

According to the diagram shown in Fig.2 we have set an experimental environment on a Linux Ubuntu 16.04 LTS using the cryptography libraries testing the MD5, SHA1, SHA224, SHA256, SHA384, SHA512 hashing algorithms on a set of data samples consisted of 34 files from different types of document and media (.docx, .pdf, .txt, .wav, .mp4, .JPG, .PNG) where the data size ranged between 12680 to 146943468 Byte (Main test results table). The hashing algorithm output structure mentioned in Table II and Table III presents a comparison on the smallest and largest data sample hash output.

TABLE II
 HASHING ALGORITHM OUTPUT STRUCTURE

No.	Hashing Algorithm	Block size / Bit	Output Character
1	MD5	128	32
2	SHA-1	160	40
3	SHA-224	224	56
4	SHA-256	256	64
5	SHA-384	384	96
6	SHA-512	512	128

TABLE III
 DATA SET HASHING SAMPLE

No	File Type	File Size / Byte	MD5	SHA-1	SHA-224	SHA-256	SHA-384	SHA-512
1	.png	12680	335f5 8faee4129 985812c54 52bb3d64a	3edbaaa0 56bd3f49e0a c81a7a88ada 57f62c07d3e 7d5f6e09ad79d7b	3edbaaa0 56bd3f49e0a c81a7a88a da57f62c07d3e 7d5f6e09ad7 9d7b	998d3bd f077737eb4e 9665985d288 4bfac49aef30 33f0833471c 2a0d47bce992	7222517eba9 463d5be304acd4 8f85a38bdbb45f0 1946899ad3b815f ed1e73baa79d592 03942549b1c521 cf8cbde1db96	a9c4d48dbc075c1b10 1efb09ca6c8752ac6f9539 db86922fc707c36683eab3 ed958dba461a1f7ba3a03b 02c0752fae31929c6b330f 739f71e34f87bc6f99c5d2
2	.mp4	14 694346 8	ddf19 5e5d38bbd f11f765cf9 aae13db1	2005d e56e8fd24a 11c7438c3 33b88777e 26b8181	49426a5 52b820f475f 798451969b4 0148519c8e2 3f3817a16d2 95491	3ee4a4d 3020e623afc 896d3aba2f0 2a07309548e cf14ae82082 d4578239007 0b	1c2dccea701 4ab9d3f5f12a09a 03270c8d259bbb c82112f52300e46 6e0b85cddce2500 9504d82672349a 1cbd15857ca9	4df4e5a4b971ede04f8 34202e47e81d316e0e71a5 b5c217350a44e49949d98a 3b3955320294bbc4558cd 92c45c0044db7bf49745c3 6283cf537e9864f8047c75

MD5 is generating a random combination of digits and characters for the input file. SHA family are considering a better method in generating the hash value which makes SHA able to have a larger range of IDs. As MD5 outputs 32 character where sha1, 224, 256, 384 and 512 respectively 40, 56, 64, 96 and 128 characters. The main fact in having less collisions in SHA512 is more due to the wide range of bit space rather than the method.

The aim of the study from considering the hash function is to guarantee an accurate checksum during checking the evidence files from manipulation at any level, therefore the security solution procedures such as attempt limitations, cyclic time, salting is not mandatory due to encryption methods to be implemented in the next section. The main point is in having a unique ID free of collisions or rainbow tables predictions which is achieved by SHA512, mainly due to that SHA512 space is 4times more as it has 2^{384} range of IDs than MD5 which is 128 bits (HEX 32byte and Binary 16 byte). Therefore, we consider the sha512 method in this paper.

Who. We must perform an authentication and give an answer to the question regarding the procedure of investigations at each level and stage who had access to the digital evidence and any manipulations of evidence [18]. For this purpose, the good method is use a biometric identification and authentication for digital signing (Who).

Digital signatures provide a method of documenting digital evidence by combining a message digest (MD5, SHA1, SHA2 , etc.) of a digital object with additional information such as the current time. The resulting hash is encrypted by (RSA, PGP, DSA) using a signing key that is associated with an individual, the resulting encrypted block is the signature.

Later, anyone can verifying signature using hash value and the public key verify thus, prevents unauthorized regeneration of signature unless it is to penetrate the private key (see Fig.3)

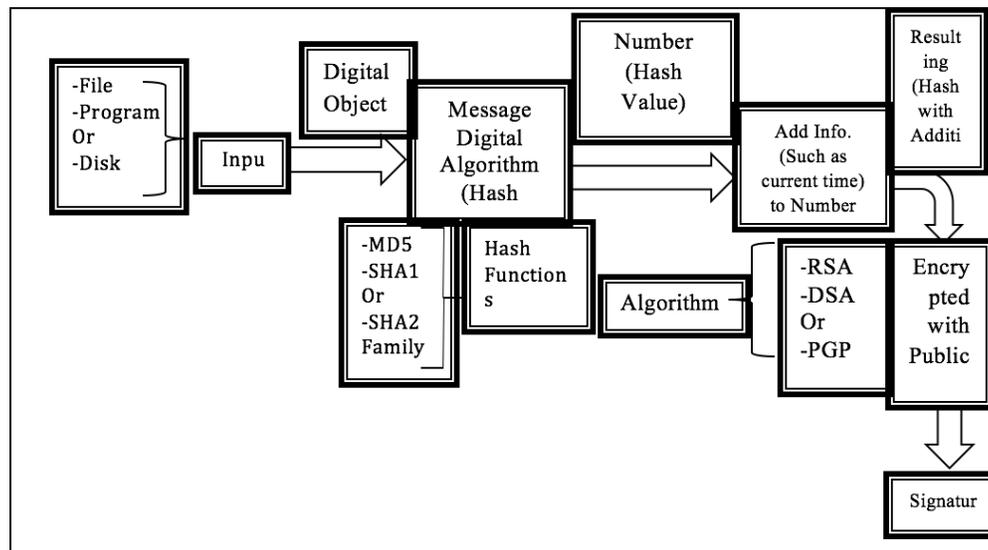


Figure 3: Process of answering to questions who handles the evidence

When. Preparing answers for questions that might be asked at the court is very important. When and for how long was the digital evidence accessed by the authorized individuals? The other question might be: A signed digital evidence integrity for How long is considered approved? [19]. This question makes the Time a main factor to prove the digital evidence integrity. Therefore, a trustable source of time will be mandatory. Consequently, in the procedure of investigation there are certain moments that digital forensic depends mainly on the Time Stamp due to detect a particular state.

A Time Stamp Authority (TSA) as a Trusted Third Party supports the user with an "existence proof" for a certain data and time [20]. The method to bind a data with timestamp is based on digital signatures and hash functions and consists of several steps. First, a hash is calculated by a message digest (MD5, SHA1, SHA2, etc.) of a digital object. Second, this hash is sent to the TSA. Third, the TSA concatenates a timestamp to the hash and calculates the hash of this concatenation. Finally, this hash is in turn digitally signed with the private key of the TSA and the result is signature (see Fig.4).

Due to the diagram in Fig.2 experiment results we have considered the SHA512 for hashing the files therefore based on the Fig.3 diagram to guarantee the accuracy of the sent file from a crime scene to the secondary destination

(Laboratory, Head Quarter office, etc.) we have proposed to encrypt the data by using RSA cryptosystem or Gnu Privacy Guard (GnuPG) in continuous the paper will present the most suitable method of encryption due to the defined requirements. To achieve this aim we have set an experiment based on the functionality of each encryption algorithm considering the same machine environment for Fig.2, as follow.

The first step is to define the data that is required to be encrypted which according to Fig.3 diagram is the hash of the evidence file + the timestamp of the evidence (file creation time) therefore we have dumped the mentioned data into a .txt file to encrypt that file as well other method is possible where attach the hash + timestamp as a string and encrypt it directly.

1) *RSA Cryptosystem* : RSA is an asymmetric encryption algorithm which the size of encrypted data is limited to the key size and regarding the Padding mode, where for Optimal Asymmetric Encryption Padding OAEP is as[21]:

1 Byte = 8 Bit Key size – 42 = Maximum file/data size	$(1024/8) - 42 =$ $128 - 42 = 86 \text{ Byte}$
	$(2048/8) - 42 =$ $256 - 42 = 214 \text{ Byte}$
	$(4096/8) - 42 =$ $512 - 42 = 470 \text{ Byte}$

It is obvious that if the evidence file is larger than for example 512 byte the RSA algorithm is not able to encrypt the file therefore as a solution it is suggested to generate/select a strong password, encrypt and send it using RSA and use the password for the file encryption considering a symmetric algorithm such as Advance Encryption Standard (AES).

In the implemented experiment the RSA key is 4096 bit (512 byte), maximum data file size is 204 byte and minimum is 147 bytes, all the cipher files are a 512 byte binary file. The following Figures presents one of the data files used for encryption/decryption.

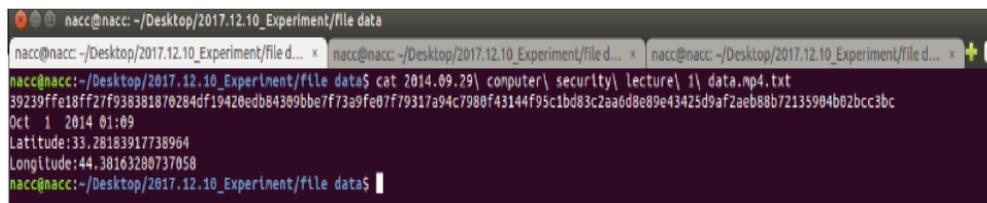


Figure 4: Evidence data file (plaintext) 1

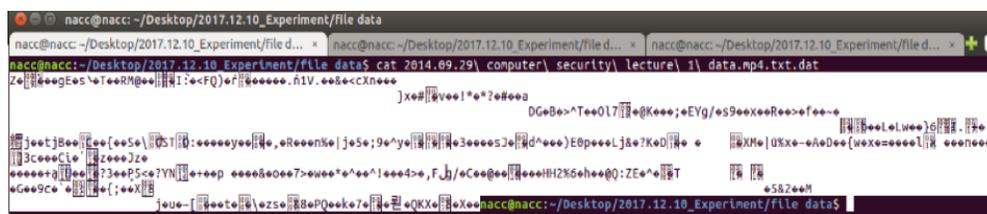


Figure 5: Asymmetric openssl RSA encryption (cipher text)

2) *Gnu Privacy Guard (GnuPG)*: Gpg uses either symmetric or asymmetric encryption for encrypting or decrypting data files/folders in other words it supports both symmetric encryption and public key encryption where it could be adopted in a wide range of applications due to the high flexibility which it provides.

We have implement both methods of encryption where using Symmetric Key requires to provide the receiver with the key. The best approach is to use the asymmetric (public key) which satisfies the key exchange protocols.

The gpg cryptosystem is able to encrypt/decrypt data file and evidence file with any size which is an advantage in comparison to the limitations of Openssl RSA, the gpg considers RSA, DSA (Digital Signature Algorithm) in it protocol options therefore gpg is a powerful tool to be adopted. In this method sensitive information is not revealed and also by publishing/broadcasting the public key a variety of encryptors would be able to encrypt data using gpg. The user is supported with a wide range of encryption algorithms to select from. These reasons make it a very useful security tool for encrypting data as files or folders. The following Figures presents the same data file (Fig.1) encrypted by symmetric and asymmetric gpg encryption



Figure 6: Gpg symmetric encryption (Cipher text)

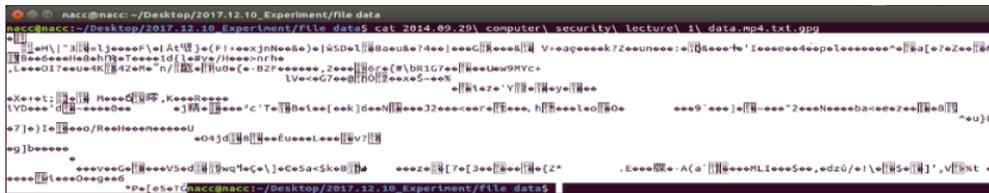


Figure 7: Gpg asymmetric encryption (Cipher text)

Where. Next, according to "chain of custody" it is important to find out the place where is handled with digital evidence. For this aim to obtain the requirements of investigation regarding the collection of evidence the Global Positioning System (GPS) is proposed.

The GPS system provides the investigator with an accurate constant longitude and latitude of the evidence collection position based on the GPS technical standards therefore this procedure had led to a new era in the evidence collection in the investigations process [22].

Other information such as GPS coordinates or a secure timestamp could be included in the signature as well. For this, we have to deal with the same procedures given in Fig.8 with the addition of features to identify the place into a final result, which means that the signature contains a right location of the digital evidences.

Regarding Fig.8 diagram for extra security accuracy the GPS is considered therefore the latitude and longitude are added to the data file, the presented figure for Fig.3 diagram experiment implementation is the complete data file considering all details (hash, time stamp, latitude, longitude).

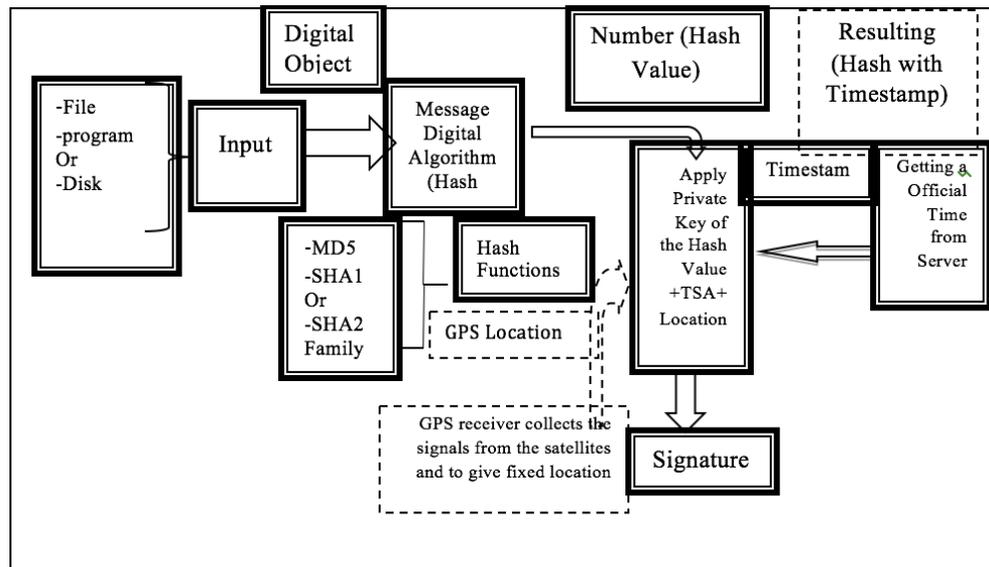


Figure 8: Process of answering to question of when or where was digital evidence collected & used?

How. As mentioned earlier, the attention of the investigator at the crime scene is more about the circumstances of the crime how happened instead of distinguishing and analyzing the evidence at the scene, which reveals "how" of the crime scene is more important than the "who". Also, the complete range of the investigation takes the form of Methods (What) vs. Procedures (How). It is meant to outline the requirements and procedures of analysis and data collection.

It is clear that the methods which are presented above is concentrate on dealing with the real world devices, such as GPS for the right location of where each piece of digital evidence is handled, timestamp generators for the right time access the evidence, biometrics devices on the electronic signature of each object, and hash code of digital files (digital fingerprint), leaving chain of custody closer to real world and ensuring that digital evidence will be accepted by court. That all the functions mentioned previously respond to the group Five "W", as for the "H" there are a two options, The first is in the [23] while the second one is explained in detail above.

V. CONCLUSIONS

The goal of this work is first, to find out how the application of chain of custody to a wide range of models of the digital forensic investigation process is to show a weaknesses in the larger number of these models that is a consequence of the lack of answers to a few questions (who, when, where, why, what and how). Second, a review is made of the methods for digitally signing an evidence that achieve the successful implementation of chain of custody to answer the questions mentioned above and provide evidence of a digital sound accepted for presentation to the court. We believe that the methods which have been presented above are focused on interaction with the real world equipment, such as GPS coordinates for determining a location of evidence, time stamp is to provide an accurate data and time for the evidence collection or access, biometrics devices for authentication and identification of a person who handled evidence, and hash function for digital

fingerprint of evidence, a correct implementation of these methods could be a robust chain of custody and guarantees that the court will accept the collected evidence. Although, each function (method) has advantages and disadvantages, these methods can be used in combination. Due to a careful implementation of the mentioned methods above and considering a various cryptograph algorithm that achieves the security requirements at different levels the outcome would cover any scenarios of data encryption and decryption.

REFERENCES

- [1] Palmer, G., "A Road Map for Digital Forensic Research", Digital Forensic Research Workshop (DFRWS): Utica, New York, 2001.
- [2] Sammes, A. & Jenkinson, B., "Forensic Computing A Practitioners Guide", Springer-Verlag, New York; 2000.
- [3] Brian, B. & Spafford, "Getting Physical with the Digital Investigation Process", International Journal of Digital Evidence, Volume 2, Issue 2, 2003.
- [4] Selamat, S.R. & Yusof, R. & Sahib, S., "Traceability in Digital Forensic Investigation Process", 2011 IEEE Conference on Open Systems (ICOS2011), September 25-28, Langkawi, Malaysia, pp. 101, 2011.
- [5] Agrawal, A. Gupta, M. Gupta, S. Gupta, C., "Systematic Digital Forensic Investigation Model", International Journal of Computer Science and Security (IJCSS), Volume (5) : Issue (1), pp. 118-126, 2011.
- [6] Ciardhuain, S. O., " An Extended Model of Cybercrime Investigations", International Journal of Digital Evidence, Summer 2004, Volume 3, Issue 1, pp. 11 2004.
- [7] Digital Forensics Research Workshop, " A Road Map for Digital Forensics Research", DTR - T001-01 FINAL, Utica, New York, Document authored from the collective work of all DFRWS attendees by: Gary Palmer, The MITRE Corporation, 2001. Available at:www.dfrws.org.
- [8] Emmanuel, S., Pilli, J. R. C., Rajdeep, N., "Network forensic Frameworks: Survey and Research Challenges", pp. 14-27, October 2010 Available at: DigitalInvestigationVolume 7, Issues 1-2.
- [9] Freiling, F. C., Schwittay, B.,"A Common Process Model for Incident Response and Computer Forensics", Proceedings of Conference on IT Incident Management and IT Forensics, Germany, pp. 10-15, 2007.
- [10] Peter, S., "A Comprehensive Approach to Digital Incident Investigation", An Article Appearing in Elsevier Information Security Technical Report, pp. 9, 2003.
- [11] Robert F. Erbacher, Kim Christensen and Amanda Sundberg, "Visual Forensic Techniques and Processes," Proceedings of the 9th Annual NYS Cyber Security Conference Symposium on Information Assurance, Albany, NY, pp. 72-80, June, 2006.
- [12] Ruibin, G. Garrtner, M.,"Case-Relevance Information Investigation: Binding Computer Intelligence to the Current Computer Forensic Framework", International Journal of Digital, Evidence, Vol. (4), Issue 1, spring, pp. 4, 2005.
- [13] Selamat, S. R., Yusof, R. and Sahib, "Mapping Process of Digital Forensic Investigation Framework", Faculty of Information Technology and Communication, Universiti Teknikal Malaysia Melaka, Ayer Keroh, Melaka, Malaysia, IJCSNS International Journal of Computer Science and Network Security, Vol.8 No.10, pp. 164, October 2008.
- [14] Sundresan, P. ,"Digital Forensic Model based on Malaysian Investigation Process", International Journal of Computer Science and Network Security (IJCSNS), Vol. 9, No. 8, pp. 38-44, 2009. citeseerx.ist.psu.edu/viewdoc/download?...
- [15] Yong-Dal Shin, "New Digital Forensics Investigation Procedure Model", Fourth International Conference Networked Computing and Advanced Information Management, pp. 528-531, 2008.
- [16] Inikpi O. Ademu, Chris O. Imafidon, David S. Preston, "A New Approach of Digital Forensic Model for Digital Forensic Investigation", International Journal of Advanced Computer Science and Applications, Vol. 2, No.12, pp. 175, 2011.
- [17] COSIC, J., BACA, M. "A Framework to Improve Chain of Custody in Digital Investigation Process", Proceedings of the 21st Central European Conference on Information and Intelligent Systems. pp. 435-438, 2010.
- [18] Hosmer, C., "Proving the Integrity of Digital Evidence with Time", International Journal of Digital Evidence Spring, Vol.1, Issue 1, pp. 3, 2002.
- [19] Internet X.509 PKI, "Time Stamp Protocol (TSP)". Available at: <http://tools.ietf.org/html/rfc3161>.
- [20] Ohmart, P., "Townsend Security Data Privacy Blog", (2011): <https://info.townsendsecurity.com/bid/29195/how-much-data-can-you-encrypt-with-rsa-keys>
- [21] Strawn, C., "Expanding the Potential for GPS Evidence Acquisition", Small Scale digital evidence Forensic Journal, Vol.3, No1., ISSN# 1941-6164, pp. 1, June 2009.
- [22] Giova, G., "Improving Chain of Custody in Forensic Investigation of Electronic Digital Systems", IJCSNS International Journal of Computer Science and Network Security, Vol. 11, No. 1, January, 2011.
- [23] Jasmin, O., Miroslav, B., "Proving Chain of Custody and Digital Evidence Integrity with Time Stamp", MIPRO, 33rd International Convention on Information and Communication Technology, Electronics and Microelectronics, Opatija, 2010.
- [24] Kruse, W.G. & Heiser, J.G., "Computer Forensics: Incident Response Essentials, Addison-Wesly, Boston, MA, 2002.
- [25] Patzakis, J. M., " Maintaining the Digital Chain of Custody", Password - The ISSA Journal, Oak Creek/U.S., pp. 14-15, February 2003.
- [26] Vanstone, P., Oorschot, V., & Menezes, A., "Handbook of Applied Cryptography, CRC", Press, 1997.