

THE DESIGN AND IMPLEMENTATION OF AN EXTENSIBLE FORMAT FOR MEMORY DUMPS

XIAOLU ZHANG, LIANG HU, ZHENZHEN XIE, SHINAN SONG, XIANGYU MENG, *KUO
ZHAO

Department of Computer Science and Technology, Jilin University, Changchun 130012, P.R.China

E-mail: * zhaokuo@jlu.edu.cn

ABSTRACT

The preservation, collection, analysis and interpretation of the evidence of computer crime following the legal procedures has become a major problem on Computer Forensics, while current available memory dumps formats and technology have drawbacks. In this paper, we present a format of physical memory dumps applied to forensics. This new format of memory dumps has provided three major advantages. First, it is more flexible, based on the characteristics of real time changes in physical memory, our design supports an update of physical memory compression at any time and reduce its process time significantly. Secondly, it has a good extensibility, supporting the storage of metadata and image at the same time, which facilitates the management and control of memory image. Thirdly, using hash and digital signature mechanism protect the integrity and reliability storage of the evidence data. This paper has solved many practical problems in the storage and protection with existing physical memory image format.

Keywords: *Computer forensic, Memory dump, Metadata, hash, Digital signature*

1. INTRODUCTION

The term Computer Forensics [1] first appeared in 1991 at the premier IACIS (International Association of Computer Investigative Specialists) training session in Portland, Oregon. Since then, many specialists have conducted in-depth study on the basic problems of Computer Forensics, among which, a preliminary computer forensic framework put forward by DFRW [2] (Digital Forensics Research Workshop) is a more reasonable process model which enables the scientific circle to further develop and improve the basic theories and methods of Digital Forensics. The DFRW adopted the definition of Computer Forensics presented by G. Palmer [3], which was generally acknowledged by the experts in this field. G. Palmer said 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. As we can see, the preservation, collection, analysis and interpretation of the evidence of computer crime following the

legal procedures has become a major problem on Computer Forensics.

In the summer of 2005, a contest file called "memory analysis challenge" was issued by DFRW (Digit Forensic Research Workshop), which aroused wide public attention. Through analyzing the volatile data in physical memory, we can acquire masses of information which cannot be extracted from stable data storage media such as hard disk. Since then, the analysis, discussion, research and the development of related tools of physical memory has become a hot topic in Computer Forensics.

In the practical forensic process [4], to avoid corrupting the evidence information, we cannot directly operate on the hard disk of the target computer but to adopt the method that [5] analyzes the dump after extracting the whole memory dump and offsite storage. Traditional memory extraction tool called DD is short for Data Dumper in UNIX. DD format has long been a standard image format of forensics. Currently, most of the evidence acquisition and analysis software support DD format. DD's seemingly powerful memory dumps file actually had a lot of restrictions. First, the compressed DD image formats are not flexible so that we cannot read and decompress the data at the same time. Moreover, this kind of information such



as metadata like the image serial number, the identity of the implementation of the survey collected, or the date generated by memory dumps cannot be stored in this files. When the metadata are not stored in the image file, these metadata will be lost or separated from their corresponding image files, even mixed with the metadata of other memory dumps.

After evaluating the memory dumps produced by the available tools, we have designed a new memory dumps storage format to solve the problems above.

2. RELATED WORK

Because already know the memory dump which extracted by memory dump tools is similar to primitive binary format which extracted by DD tools, the reference value is not used. For that reason survey the format of disk image storage is used by disk information forensic, hoping to get a reference.

The raw format [6] is simply a file that contains the exact data that needs to be stored and the file could contain any type of data, including hard disk sectors, files, and network packets. Raw files can be easily created and read by any tool, but they do not store any metadata and are not compressed.

The Advanced Forensic Format (AFF) [6] is from Simson Garfinkel and Basis Technology. [7]The format is open and comes in three variants: AFF, AFD, and AFM. AFF stores all data and metadata in a single file, AFD stores the data and metadata in multiple small files, and AFM stores the data in a raw format and the metadata is stored in a separate file.

There are two independent formats that use the name Digital Evidence Bag (DEB) [6]. The first one we discuss is from Philip Turner and Qinetiq. The format is open and was first presented in a paper at DFRWS 2005. It uses a number of files to store the evidence and associated metadata. The metadata are stored in ASCII files. The second DEB format is from Wetstone Technology. This format uses XML to store the evidence and metadata.

The EnCase format [6] is a closed format that is defined by Guidance Software for use in their EnCase tool to store hard drive images and individual files. Its predecessor format is the Expert Witness format, which has been publicly documented. The EnCase format has added new metadata to the original Expert Witness format.

The Generic Forensic Zip (Gfzip) [6] format is from Rob J Meijer. Its design is open and uses data structures similar to AFF. The metadata and storage approach are different though. A ggzip file can be 'raw' compatible so that the metadata is stored after the evidence data and it also offers a 'packed' mode where redundant blocks of data are not stored.

The ProDiscover format [6] was defined by Technology Pathways for use in their products to store hard drive images. The format is open and has a published specification.

There are two SMART formats [6], which are defined by ASR Data for their products to store hard drive data. The default format stores the metadata in a separate text file where the contents can be easily viewed, but the exact layout has not been published. The second format, which we will call the SMART Expert Witness Compression format, is based on the original Expert Witness format.

The above survey can detect that, the success of disk image storage format in solving the problems of metadata, compression and evidence protection, then it worth to be reference for memory dump storage format designing.

3. DESIGN AND IMPLEMENTATION

3.1. Format Design

The memory dump format feature and advantages of our design is:

- Have ability to change the ordinary binary data form of the files in the memory to memory dump format of our design.

- Have ability to compress and store the original memory data to reduce the memory space size.

- Adding many necessary metadata to store it with source memory data to solve the problems of flexibility, security and management of the existing memory dump file.

- Any size (the default size of the block is 256K multiple) of original memory blocks can be compressed and processed, this solve the speed problem of updating memory dump according to the original memory by avoid repeating analysis of the unchanged memory block.

- Support internal consistency investigation, even if any parts loosed or corrupted the others parts still usable.

- Using Bad Flag and Credible flag can decide quickly which block of memory dump is corrupted

or incredible or not, furthermore, where is that block exactly, if we need.

Using a traditional hash function (MD5 or SHA-1) and based on RSA algorithm with advanced digital signature of X.509 certification to keep the integrity and reliability of the forensic data.

By measuring the time of compression task, consumed storage space and security those required parameters let us freely chose compression method, size of data block and decide whether the hash value and signature is for all dump file or not.

Then describe the differences between internal field structure of memory dump storage format and every data block contents, as shown in figure 1 and figure 2

Dump ID
Investigator ID
Dump Addition
Source Memory
Dump Date
Memory Size
Compression (Y/N)
Integrity (Y/N)
MD5/SHA-1(Y/N)
Hash(Y/N)
Splitting(Y/N)
Block Size
Splitting Size
Dump Size
Bad Flag
Credible Flag
Block Offset
BLOCK 0
BLOCK 1
.
.
.
Dump Head Hash
Dump Hash
X509
Head Hash SIG
Dump Hash SIG

Figure 1: Entire Format Of Memory Dump (Not To Scale).

Block Flag
Block Num
Block Bad Flag
Credible(Y/N)
Bad Block(Y/N)
Meta Data Hash
Forensic Data Hash
X509
Meta Data SIG
Forensic Data SIG
Forensic Data

Figure 2: Block format (not to scale).

3.2. Implementation

3.2.1. Metadata

We need to store the data related with computer forensic, such as completion time of memory map, digital signature or hash. All the metadata were included in a defining word, described in section 3.1 we will give a detailed description of each word and memory field. The metadata was classified in two types: one is the metadata of general map files, it contains a head and a tail; the other is in each data block in map files.

Table 1: All Fields' Description And Size Of Dump File Metadata

Segment name	Meaning	Size
Dump ID	Memory dump identifier	8 Byte
Investigator ID	Investigator identifier	8 Byte
Dump Addition	Memory dump additional information	32 Byte
Source Memory	Source memory mark	8 Byte
Dump Date	The date and time of making the dump	4 Byte
Memory Size		1 Bit
Compression(Y/N)		1 Bit
Integrity(Y/N)	The size of source memory	2 Bit
MD5/SHA-1(Y/N)	Compression/decompression	2 Bit
Hash(Y/N)	Integrity investigation	2 Bit
Splitting(Y/N)	Which type of (hash MD5/SHA-1)	2 Bit
Block Size	Hash or not the whole file	4 Byte
Splitting Size	Divide the file or not	8 Byte
Dump Size	Dump block size	1B-16kb
Bad Flag		



Credible Flag	File part size	1B-16kb
Block Offset	Dump file size	
Dump Head Hash	Corrupted block flag	32 Byte
Dump Hash	Credible block flag	32 Byte
X509	Forensic data block offset	
Head Hash SIG	Hash value of dump head	32 Byte
Dump Hash SIG	Hash value of entire dump	32 Byte
	X509 certificate (public key)	
	Digital signature of dump head hash value	
	Digital signature of entire dump hash value	

Table 2: All Fields Description And Size Of Block Metadata

Segment name	Meaning	Size
	Data block flag	
Block Flag	Data block identifier	
Block Num	Bad sector flag	8 Bit
Block Bad Flag	Credible/ Incredible	32 Bit
Credible(Y/N)	Have or not bad sector	64b-8mb
Bad Block(Y/N)	The hash value of metadata	1 Bit
Meta Data Hash	The hash value of forensic data	1 Bit
Forensic Data		32 Byte
Hash	X509 certificate (public key)	32 Byte
X509		
Meta Data SIG	Digital signature of metadata hash value	32 Byte
Forensic Data		32 Byte
SIG	Digital signature of forensic data hash value	
Forensic Data		
	Data used for forensic	

3.2.2. Storage compression

Storage

In order to reduce the space of data storage in many forensic data usually use obvious compression method. But the compression of memory dump file is time consumer. This is because that uncompressed data needs only for store time, but compressed data method needs for compression time and store time. Then the employer needs to find the optimal solution of time and space.

Data update

The main difference between memory dump and disk image is the memory dump take locard's exchange principle in consideration. when the tool make the memory dump and the memory which already dumped can be changed, for example, the contents of page number 11 can be changed, while the page number 30 is read. The speed of entire

memory dump system is determined by the speed of CPU, the speed of system bus and the speed of disk input/output. Due to disk image has static store property and the memory has dynamically change the volatile data, therefore the problem is that if the memory contents changed during the dumping process can effect or not on the analysis of dump data.

Therefore, the re-extraction and updating speed have to be fast as possible which the investigator needs it to analyses the memory dump, if the previous dump storage structure is used, the processing will be very slow. Because, sometimes the entire memory dump re-extraction and re-analysis is needed, but the amount of modified data is relatively small. If the new memory dump storage format has been applied, the process will be more simple. Using the segmental encryption and storing features, then can decide which block need to update by comparing the hash value of corresponding segment only, then avoid the updating of unmodified contents. Without using entire memory re-analysis will increase the speed of analysis. To describe this process in more detail and give the reference to practitioner who use this storage format, to update memory dump we present the simple algorithm which is based on this storage format like shown bellow.

Algorithm 1:

Let M0 is modified physical memory data, M1 is the memory dump which is needed to update.

Step1. Read Memory Size and Dump Source fields from metadata head of M1 and compare them with the name and size of M0(in most situation is same).

Step2. Read the Compression (Y/N) field from metadata head of M1, and decide before updating whether the data is compressed or not.

Step3. Read the Block Size field from metadata head of M1, and divide the M0 into blocks where the block size is same as the contents of Block Size field, and numbering and computing hash value for every block of M0, (using MD5/SHA-1(Y/N) field of M1 to determine hash method).

Step4. Compare the hash value of every data block of M0 with corresponding Forensic Data Hash field contents of M1 block data, if them is same then ignore else go to step 5.

Step5. Take the block data from M0 and add the metadata using the block format definition (if the Compression field is chosen, then firstly compute

hash value of the data block and compress it) and put it in corresponding data block in M1. Go back to step4, until storing all data blocks which have to change it.

Step6. Update metadata in whole file of M1 (needing reference for original values of every field in M1 file).

Integrity

If the data in any memory dump corrupted, so investigator have to check it, and quarantine the corrupted part, and delete the corrupted part only not the memory as all. In practical [10], Practitioners use hash value to check the data integrity. Memory dump storage format using MD5 and SHA-1 this two types of hash value is our new designing.

In this memory dump format also use 64B to 8MB field size (The default is 64B) of Block Bad Flag to identify which data block is corrupted. The size of flag field is depended on the size of chosen block.

The following algorithm describes checking of memory dump integrity, which provides the reference for the practitioners of memory dump storage format.

Algorithm 2:

Suppose: the needs of memory dump integrity investigation is M.

Step1. Extract the Compression (Y/N) and MD5/SHA-1(Y/N) fields from the head of M's metadata, to decide the Forensic Data block is Compression or not, also to know which type of hash used in that data.

Step2. Using the hash type to compute the hash value for Forensic Data in every M's blocks (if the compression field is Y then firstly decompressed Forensic Data), after that compute hash value for metadata.

Step3. Extract Data Hash and Meta Hash fields' value for every M's block.

Step4. Compare the obtain hash value with original hash value, If the Data Hash and Meta Hash value for the one block is same, so the Credible(Y/N) field will be N, this mean the data block or only the meta data is incredible.

Step5. If every M's block is integrated, then compute the integrity of whole file.

Reliability

Whenever, checking memory dump integrity is possible, but is difficult to keep it reliable. Suppose: A make memory dump, the attacker B, and the third party C. The attacker B can take forensic data field and change its contents also obtain hash value of the new data, and replace the original data and its value by the new data with its value, then A,C can not detect where is modified data which B changed it. Therefore, digital signature method is necessary to protect hash value. For this, when forensic data is used, the modification of hash value is possible to detect it, also the modification is Intentional or unintentional.

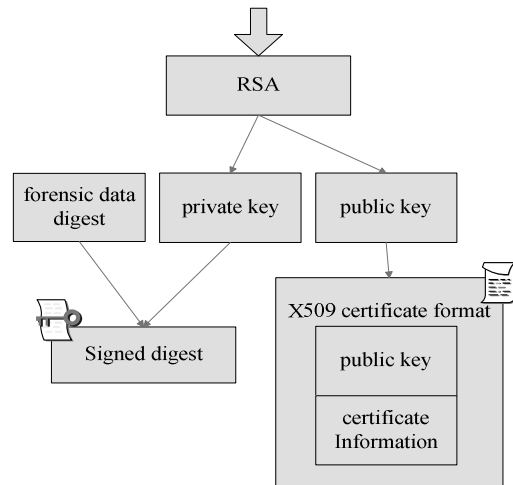


Figure 3: Process of hash value signature

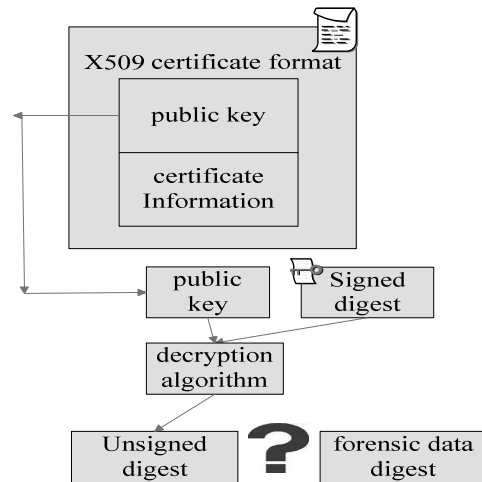


Figure 4: The Process Of Signature Verification

Particular process in detail as shown in figure 3:

Step1. Using RSA algorithm to obtain the public and private keys pairing by A, and store the private key which use in signature to private storage.

Step2. Take public key and put it in the X509 certification format, and complete the remaining information of that certification.

Step3. Using a private key to sign all hash values of Dump Head Hash and Dump Hash fields which is stored in head of memory dump or hash values of Meta Data Hash and Forensic Data Hash field which is stored in the data block. Then store the signature value in its corresponding XX_SIG field.

Signature verification process description:

Step1. C extract the public key from X509 field which exist in certification and decryption the contents of XX_SIG field by using public key.

Step2. If the decryption value is equal to corresponding XX_Hash value, this means all hash values which stored in XX_Hash field have not any intentional or unintentional modification, if not equal it means the hash value have already changed, nevertheless, the corresponding data contents is incredible, set Credible(Y/N) field N.

3.2.3. Experiment

In terms of the design in 3.2, we have developed a new program based on this memory dumps storage format, which can transform the given physical memory data of binary formats to our new memory dumps storage format. Figure.5 shows some metadata information and result information during our format storage to 133 MB memory data. This program has produced a satisfactory result both in compressed storage and the metadata proportion.

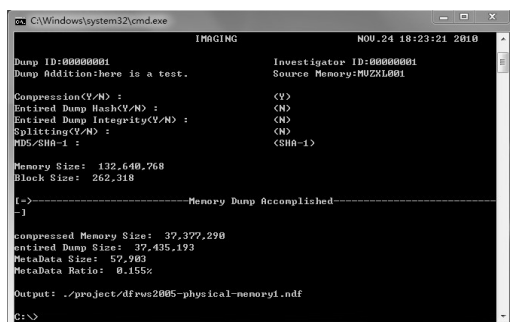


Figure 5: Screenshot Of The Tool In Action

4. CONCLUSIONS

This paper has given a detailed description of the design and implementation of a new extensible memory dumps format. This new memory dumps format has enabled the storage of physical memory data no longer limited to the format storage and solved the problems such as the storage space of binary formats, the data update model, and content

integrity and reliability. We have not only carried out a better plan on the storage mode of physical memory data, but also improved the effect of memory dumps on computer forensics, which has laid a foundation for the computer forensic mode, especially the development from the lonely physical memory forensic mode to the mass analysis and storage mode. Our current work is to continue optimizing the contents such as time, space and operability required in the application of physical memory forensic format and we expect to apply basic analysis ability which can better meet the needs of memory forensics.

5. ACKNOWLEDGMENT

This work is supported by the National Grand Fundamental Research 973 Program of China under Grant No. 2009CB320706, the National High Technology Research and Development Program of China under Grant No. 2011AA010101, the National Natural Science Foundation of China under Grant No. 61073009 and 60873235, Program of New Century Excellent Talents in University of Ministry of Education of China under Grant No. NCET-06-0300, the Youth Foundation of Jilin Province of China under Grant No. 201101035, and the Fundamental Research Funds for the Central Universities of China under Grant NO.200903179.

REFERENCES:

- [1] Yun Wang, James Cannady, James Rosenbluth, "Foundations of computer forensics:A technology for the fight against computer crime", Computer Law & Security Report, vol.21, no.2, 2005, pp.119-127.
- [2] Katrin Franke, Sargur N. Srihari, "Computational Forensics: An Overview", LNCS, vol.5158, 2008, pp.1-10.
- [3] Gary Palmer, "A Road Map for Digital Forensic Research, New York", the First Digital Forensic Research Workshop, 2001.
- [4] DING Li-Ping, WANG Yong-Ji, "Study on Relevant Law and Technology Issues about Computer Forensics", Journal of Software, vol. 16, no.2, 2005, pp. 260-275.
- [5] Richard M. Stevens, Eoghan Casey, "Extracting Windows command line details from physical memory", Digital Investigation, vol.7, supplement.1, 2010, pp.57-63.
- [6] DFRWS, CDESF, "Survey of Disk Image Storage Formats", <http://www.dfrws.org/CDESF/survey-dfrws-cdesf-diskimg-01.pdf>, 2006.



- [7] Michael Cohen, Bradley Schatz, "Hash based disk imaging using AFF4", Digital Investigation, vol.7, supplement.1, 2010, pp.121-128.
- [8] R.B.van Baar, W. Alink, A.R.van Ballegooij, "Forensic memory analysis: Files mapped in memory", Digital Investigation, vol.5, 2008, pp.52-57.
- [9] James Okolica, Gilbert L. Peterson, "Windows operating systems agnostic memory analysis", Digital Investigation, vol.7, supplement.1, 2010, pp.48-56.
- [10] Simson L. Garfinkel, "Providing Cryptographic Security and Evidentiary Chain-of-Custody with the Advanced Forensic Format, Library, and Tools", International Journal of Digital Crime and Forensics, vol.1, no.1, 2005, pp.1-28.
- [11] Emmanouil Vlastos, Ahmed Patel, "An open source forensic tool to visualize digital evidence", Computer Standards & Interfaces, vol.29, no.6, 2007, pp.614-625.
- [12] DING Li-Ping, ZHOU Bo-Wen, WANG Yong-Ji, "Capture and Storage of Digital Evidence Based on Security Operating System", Journal of Software, vol. 18, no.7, 2007, pp.1715-1729.
- [13] Simson Garfinkel, Paul Farrell, Vassil Roussev, George Dinolt, "Bringing science to digital forensics with standardized forensic corpora", 9th Annual DFRWA Conference 2009, vol.6, 2009, pp.2-11.
- [14] Simson L. Garfinkel, "Digital forensics research: The next 10 years", 10th Annual DFRWS Conference, vol.7, 2010, pp.64-73.
- [15] Karthikeyan.S, Sairam.N, Manikandan.G, Sivaguru.J, "A parallel approach for improving data security", Journal of Theoretical and Applied Information Technology, vol.39, no.2, 2012, pp.119-125.
- [16] Upadhyay.Sachin1, Singh.Yaspal, "Cryptanalysis of RSA through formal verification tools", Journal of Theoretical and Applied Information Technology, vol.39, no.1, 2012, pp.17-21.