Annual ADFSL Conference on Digital Forensics, Security and Law

2014
Proceedings

May 28th, 2:40 PM

Generation and Handling of Hard Drive Duplicates as Piece of Evidence

T. Kemmerich
University College Gjøvik, thomas.kemmerich@hig.no

F. Junge
University of Bremen, fjunge@tzi.de

N. Kuntze
Fraunhofer SIT, nicolai.kuntze@sit.fraunhofer.de

C. Rudolph
Fraunhofer SIT, carsten.rudolph@sit.fraunhofer.de

Endicott-Popovsky
University of Washington, endicott@uw.edu

See next page for additional authors

Presenter Information
T. Kemmerich, F. Junge, N. Kuntze, C. Rudolph, B. Endicott-Popovsky, and L. Großkopf

GENERATION AND HANDLING OF HARD DRIVE DUPLICATES AS PIECE OF EVIDENCE

T. Kemmerich
University College Gjøvik
thomas.kemmerich@hig.no

F. Junge, University of Bremen
TZI, Bibliothekstr. 1, 28359 Bremen, Germany
fjunge@tzi.de

N. Kuntze
Fraunhofer SIT
nicolai.kuntze@sit.fraunhofer.de

C. Rudolph
Fraunhofer SIT
carsten.rudolph@sit.fraunhofer.de

B. Endicott-Popovsky
University of Washington
endicott@uw.edu

L. Großkopf
Kanzlei Prof. Dr. Lambert Grosskopf LL.M.Eur.
lawoffice@grosskopf.eu

ABSTRACT

An important area in digital forensics is images of hard disks. The correct production of the images as well as the integrity and authenticity of each hard disk image is essential for the probative force of the image to be used at court. Integrity and authenticity are under suspicion as digital evidence is stored and used by software based systems. Modifications to digital objects are hard or even impossible to track and can occur even accidentally. Even worse, vulnerabilities occur for all current computing systems. Therefore, it is difficult to guarantee a secure environment for forensic investigations. But intended deletions of dedicated data of disk images are often required because of legal issues in many countries.

This article provides a technical framework on the protection of the probative force of hard disk images by ensuring the integrity and authenticity using state of the art technology. It combines hardware-based security, cryptographic hash functions and digital signatures to achieve a continuous protection of the image together with a reliable documentation of the status of the device that was used for image creation. The framework presented allows to detect modifications and to pinpoint the exact area of the modification to the digital evidence protecting the probative force of the evidence at a whole. In addition, it also supports the deletion of parts of images without invalidating the retained data blocks.

Keywords: digital evidence, probative force hard disk image, verifiable deletion of image data, trusted imaging software

1. INTRODUCTION

To prepare and conduct court proceedings, more and more digital information is needed (Raghavan 2013; Saudi 2001). It is therefore common practice to generate images of entire drives (Garfinkel, 2006). In contrast to backups, an image also contains information about the file system structure of the original data carrier, including the master boot record, since raw data and not just individual files are copied. The use of images instead of physical hard disks for forensic investigations has the advantage that the owner of the hard disk can continue to use the disk after the image was taken. This is particularly relevant for companies that depend on the data on the disks. Long-term confiscation of computers and hard-drives can potentially ruin a company. Nevertheless, the forensic investigation on the basis of the image needs to preserve the integrity of the image and the process needs to ensure that no alterations and changes can be done. In many countries legal issues enforce the deletion of ”core area”–data that means private and intimate information in form of documents, photos, audio files and videos.

Obviously, disk images are just digital data and thus they are in principle easy to change. Exact manipulations are difficult or even impossible to notice on raw image data. Generating, storing and using images therefore require special care so that the images can serve as suitable digital evidence. However, all current tools provide no protection against malicious or deliberate changes to the images. Hash values, for example, from images don’t contain information about the processing state of the image or the time of generation.

The current process for forensic evaluation of hard disks assumes that all staff dealing with the image is trustworthy and has no motivation to maliciously change the image. Further, it also assumes that the computers used in the process are secure and only accessible by trusted staff. Both assumptions should be called into question. It might be true in most cases that the investigators are trusted and will not manipulate the image data. Nevertheless, with the current process and forensic tools they can easily change images without any chance of someone being able to prove that the image is not the correct one. All technical solutions (e.g., no-write during image creation, check-sums, hash values, no functions to change image in forensic software) only target accidental change. In general, one might assume that in some cases investigators have some incentive to manipulate data, either to get personal advantages or to harm the owner of the hard-drive. Security of the used devices is also critical. Investigators use standard computing platforms for the creation and evaluation of disk images. These devices can potentially be attacked in many different ways. Remote access, malware running on the device or combinations with social attacks can be used to maliciously change the image data.

Current regulations do not demand stronger security for digital evidence in forensic investigations. Guidelines established by the German BSI (2011) require integrity protection as realized by current forensic tools but technical solutions to preserve authenticity of the image are not considered. Reviews of current forensic software by NIST[1], NIST (2012) show that only cryptographic hash algorithms (e.g., SHA1) are available for integrity protection. No digital signatures, time-stamps or binding to status of the used devices are considered in any of the existing tools.

Clearly, various organizational issues need to be considered for the collection and use of digital evidence. Andrew (2007) defined such a process for images of storage data. In general, one can identify the following steps:

- Who came into contact, handled, and discovered the digital evidence?
- What were the procedures that were used to identify and collect the evidence?
- Where was the digital evidence discovered, collected, handled, stored, and examined?
- At what time was the digital evidence discovered, accessed, collected, examined, archived, or transferred?

[1] Computer forensics tool testing (cftt) project web site http://www.cftt.nist.gov/

- What was the reason to collect this particular evidence?
- Which technology was used to collect, examine, and store the digital evidence?
- How was the evidence protected from changes and manipulations?

All these different items are relevant. However, the contribution of this paper concentrates on the technology used to securely create images for storage devices and to protect them against accidental and malicious changes. The second aspect is to describe a procedure to ensure that deletion of core area data (private and intimate data) on the disk image is explicit comprehensible.

A significant discussion for the development of producing, securing, handling and maintaining digital and digitized evidence from the technical as well as from the legal side was discussed with experts from Europe, US, South Africa and Australia during a Dagstuhl Seminar in February 2014[2]. New requirements and next steps in research have been discussed and will be documented in a Dagstuhl Manifesto. Aspects of this paper will be taken into account for these developments.

2. DIGITAL DATA MODIFICATION

It is not only intentionally that digital data may be changed as Pinheiro et al. (2007) described. They may also vary randomly and spontaneously due to errors in the program that generates the disk images, or because of damage to the medium on which an image is stored (physical errors). Errors in the operating system may give the user the impression that image data altered (logical error) during the forensic investigation. When generating an image, it is important to distinguish between a physical image and a logical image. An image on the physical level duplicates data according to their actual storage on the respective hard drive or other relevant media. A logical image, on the other hand, provides the duplication of the data, as they are available to the operating system. Disks use error correction mechanisms to detect defective memory areas independently and exclude these areas when saving data. Data is stored in different physical conditions in different places. This process of mapping these different physical conditions to indistinguishable images cannot be retraced from the outside. To meet the storage requirements of today's applications and to improve reliability, modern operating systems use intelligent techniques to secure data on multiple disks. In UNIX and Linux operating systems, which are predominantly installed on servers, it is possible to create dynamically adjustable partitions (logical volumes), which may extend over several disk drives. The size of these virtual disks changes, even if data has already been stored in the logical volume. A redundant array of independent disks (RAID) serves to organize several physical disks of a computer into a logical drive. This provides for higher data availability in the event of individual hard drive failures and for better data access than a single physical drive would. In both technologies an image of such a disk array is therefore per se not a one to one copy, because the deposition on various storage media is not visible to image generation programs. The program is thus led to "believe" that the data is stored on one disk, even though the data is stored on multiple disks. The program does not generate an image then, but stores data anew on the backup medium instead.

The assumption that an image represents a realistic copy of a disk may therefore be deceptive because conventional programs only generate an image on the logical level. The physical details of the storage (the location on the hard disk or on a specific disk in a storage system), however, are not recognized by the program and will consequently not be logged either. Actually, a real copy at a physical level is in many cases not possible. Additional problems are introduced by technologies such as the integration of virtual drives over the Internet or the use of self-encrypting hard drives. In either case, the image available is a logical one. During the production of said image it needs to be documented that the image generated is a correct copy of the physically stored data in terms of content.

[2] http://www.dagstuhl.de/de/programm/kalender/semhp/?semnr=14092

3. IMAGE GENERATION TOOLS

Several programs and tools exist to generate images from given hard disks. One, that is commonly used tools by law enforcement institutions to generate an image, is the FTK Imager[3]. It allows generating images in various formats and supports the forensic analysis of the images. In the following, only the generation of an image using this program will be considered, the forensic analysis will be ignored.

The FTK Imager has no protection mechanism to recognize data changes during their transit to a new medium or to detect subsequent changes in the archive. The user can therefore only assume that he has established a 1:1 copy of the data carrier to be backed up. State Offices for Criminal Investigation argue that the image may still be trusted because the FTK Imager itself does not provide any means to change data at a later point in time, ignoring that a change is possible with only minimal effort. For example, it is possible to re-insert a previously generated image pretending to be its own drive and delete data without the FTK Imager realizing it, because the ASR Data's Expert Witness Compression format, (Knight, 2011), does not provide any effective techniques for detecting subsequent changes. Mechanisms used for integrity protection are CRC and the MD5 hash function. After a modification both CRC and MD5 hashes can easily be recalculated, so that changes cannot be detected by means of these values. The necessary technologies are already an integral part of any operating system. Such an attack could come from an investigator, but any administrator or other person with access to the evidence can perform such a change both easily and quickly.

Based on the grounds that the possibility to change secured Images with the FTK Imager does not really exist, State Offices for Criminal Investigation do not deem it necessary to document the image generation process, but claim further- more that data cannot be deleted selectively from images, not even evidentiary irrelevant or exempt data, such as highly personal information.

The AFF Format provides a mechanism to split the image into smaller segments called pages, (Garfinkel, 2006). Even though AFF supports page signing with digital signatures, it does not support further modification as well. This means that there is no built-in support for a later reproduction of undertaken steps, e.g., deletion of privacy sensible data. A lightweight tool for forensic purposes is the patch for GNU dd called dc3dd[4]. It is able to split an image into smaller pieces and to generate a hash value for each of these pieces, but it does not support a cryptographic signature by a single investigator nor a PKI infrastructure. It can be seen, that these mentioned tools fail to log an investigators action. Steps undertaken could be logged manually, but this is error-prone, arduous and depends upon the investigators expertise. Even worse, a malicious investigator can easily trick the programs and delete or modify data in the process of image creation. With regards to the ACPO guidelines[5], our goal is to automate the logging process and prevent errors and misuse in the creation and handling of forensic images.

4. REQUIREMENTS FOR THE PROBATIVE GENERATION OF IMAGES

Images of hard disks are just digital data that shall be used as digital evidence. Therefore, generation and storage of images shall follow the same rules as they are currently discussed for digital evidence in general. These requirements are concerned with the device that was used to create the image and the protection of the image itself. One possible definition is provided by Kuntze et al. (2012):

A data record can be considered secure if a device for which the following holds authentically created the digital evidence of it:

- The device is physically protected to ensure at least tamper-evidence.

[3] http://www.accessdata.com/products/digital-forensics/ftk
[4] http://sourceforge.net/projects/dc3dd/
[5] http://www.forensic-computing.ltd.uk/ACPO%20Guide%20v3.0.pdf

- The data record is securely bound to the identity and status of the device (including running software and configuration) and to all other relevant parameters (such as time, temperature, location, users in- volved, etc. [6]).
- The data record has not been changed after creation.

Consequently, integrity protection against unintentional changes is not sufficient. In contrast, the authenticity of the image creation process needs to be preserved and documented. Parameters to be documented and securely bound to the image include the software running on the device, persons controlling and authorizing the creation of the image or the creation time of the image. If the exact time is important, it might be required to use a trusted time source, e.g., an external time authority to time-stamp image data. In principle, images shall not be modified at all. Nevertheless, in some cases the law might prescribe modifications. Some parts of the image might contain private information. In this case, the information should not be stored and should be deleted. Current practice is to store complete images in contradiction to laws and privacy regulations. For example, the Constitutional Court has decided, that these kinds of private and intimate data are not allowed to use for investigation purposes and it is not allowed to collect such information. In case that such data is already stored, actions must be executed, do delete all of affected from any medium as well as from the court record. This has to be done under consideration of keeping the probative force of the hard disk image [7]. Thus, a proper image creation needs to support the documentation of the deletion of data within the image without invalidating the probative force of the remaining parts of the image.

5. TECHNOLOGICAL BUILDING BLOCKS

This section will introduce a set of building blocks that allows providing solutions fit for the requirements presented in the previous section. The main focus of this section lies within presenting a scheme protecting the integrity and authenticity of digital evidence with respect to the case of hard disk images. Additionally, a solution for a documented deletion of data is shown and the overall process documentation is discussed. Finally, some thoughts on the handling of images derived from multiple hard disks (e.g., for RAID arrays that combine multiple disk drive components into a logical unit) are introduced.

5.1 Integrity Protection

As discussed above, integrity protection always requires the protection of authenticity of the creation or of authorized changes to the data. Today, digital signatures are used to provide the authenticity of digital data. These signatures apply broadly to various scenarios such as long term archiving. To protect integrity and authenticity of digital evidence, first a digital hash is created using a hash algorithm like SHA-1, SHA-2 or other accredited standard, e.g., through NIST. This hash value representing ("finger-printing") the document is then digitally signed using algorithms like RSA, Jonsson, Kaliski (2003), in the PKCS standard or others accepted by a public authority like NIST[8].

The most basic approach to protect a hard disk image is to create one hash value for the complete image and then sign this hash value. In case of a modification the hash value of the modified image is different to the signed value. But as the hash applies to the whole image, a modification cannot be traced to a single part of the image. As a result, using only parts of the image during a forensic evaluations and deleting the unused or forbidden parts is not possible. The hash value would be changed which will result in a diminished probative force of it.

To allow for the tracing of modifications, the image can be regarded as a sequence of individual units (slices). For each unit an individual hash can be calculated. For a sequence of all these individual hash

[6] The actual set of parameters and the protection levels depend on the scenarios and on the type of data record
[7] http://heinrich.rewi.hu-berlin.de/doc/strpr/29_beweisverwertungsverbote_4.pdf
[8] http://csrc.nist.gov/groups/ST/toolkit/index.html

values a new hash can then be created and signed. This method is known as a hash tree and is used in several applications, for example in the area of long term archival, Kunz et al. (2008). Depending on the protection level the appropriate size for the individual units needs to be specified. As files are stored on the basis of clearly identified slices, a straightforward proposal is to use the block size also as the unit size for the hash values.

This hash tree now allows comparing the hash value for each individual slice with the reference as signed. Any modification can be tracked into an individual slice and therefore the file modified can be determined. If a finer granularity is required or techniques like block sub-allocation, Claar et al. (2000) are to be covered more precise strategies for the determination of the hashes need to be developed.

Digital signatures and the tree of hashes as used in the proposed scheme are stored independently of the image. Thus, any tool for the creation and validation of the protective digital signatures can be used in parallel with other existing tools for forensic investigations. This allows for an easy integration of additional protection schemes into existing forensic processes.

5.2 Data Deletion

The procedure described also permits to delete evidence-irrelevant or exempt data at a later point in time, i.e., to execute a deliberate data modification. To allow for the deletion of data, it is required that the deletion is documented and can be associated to the person who performed the deletion. Deleting information in an image is basically a modification to the image by writing zeros into the specific parts of the image that contain these information and thus destroying them. Such a modification can be clearly documented using the digital signature scheme as presented above.

Using the approach of a hash tree having an individual hash for each slice, in the deletion process the modified blocks can be documented using a list identifying each slice modified. This list shows for each modified slice the previous hash value and the new hash value. While verifying an image with deletions, every time a slice does not meet the expected hash value the verification process queries the list of modifications. If a slice was deleted, the original hash value for this slice is used for the verification. Thus, the original hash value can only be recreated if all other slices have not been changed. To document the deletion, a new hash value has to be formed and signed (deletion signature) after the deletion was carried out, so that the authenticity of other image data can be determined using the interaction between the original image signature and the erasure signature. It is documented that at the time of deletion a particular slice had a specific content, expressed by the hash value of the slice. In the context of this deletion, the slice is overwritten with a predetermined content. The deletion signature now says that a slice had a different content before the deletion, and who overwrote that content. During the examination of the image the altered slices stand out due to an erroneous hash value. This slice is then looked up in the deletion signature and the original value is used in the subsequent calculations, as provided by the deletion signature.

5.3 Process Documentation

In the process of the acquisition of a hard disk image, the core root of trust is the person creating the image in the first place. The correct execution of the process including a proper handling of the hard disk and the software involved in the data extraction depends on the people involved. Within this process the person creating the disk image vouch for the correct execution and normally express this by signing a written protocol.

To technically bind the image creation to a person, the person creating the image personally signs it by applying a personalized digital signature. This requires a specific key infrastructure and guidelines on signature creation on the side of the administration. The ESIGN act defines an electronic signature as follows for the United States (Knaus and Foley, 2001). Other countries have similar regulations on the definition of a non-handwritten signature.

The term 'electronic signature' means an electronic sound, symbol, or process, attached to or logically associated with a contract or other record and executed or adopted by a person with the intent to sign the record.

To legit the electronic resp. digital signature it must be clear that the signature belongs to the user signing the document. Typically a Public Key Infrastructure (PKI), see Maurer (1996) is used to establish on a technical and organizational level this relation. Additionally a signature creation device, according to CEN (2001) is required to produce the signed data set.

Additionally, the creation time of a specific image needs to be documented in the signature of the image complementing the information on the person involved. To allow for a time stamp a reference time is required and in most countries already available. A public time stamp authority provides signatures with a time embedded for data sent to the authority.

A second core root of trust is the software used in the process of image creation. The imaging software needs to be trustworthy to ensure that the image created is in accordance to the requirements towards digital evidence. To document the process proof on the software used is part of the information endorsed in the image signatures. Novel approaches towards security architectures as developed by the Trusted Computing Group (TCG)[9] allow for a trustworthy documentation of software running on a system and the reporting towards a verifier later.

An important aspect in the documentation of software used in the image creation process is the certification of the software as result of an evaluation process. In this evaluation process existing software is tested by an official authority to ensure the usability of the software. One example hereby is the work of the Computer Forensics Tool Testing (CFTT) work group of NIST[10] providing a list of software tested. To support the verification of images and the process performed the existing infrastructure like the CFTT would need to publish version specific digital signatures allowing for an automated process on the basis of the technology provided by the TCG.

The underlying idea in the application of TCG technology in the area of the creation of digital evidence is to document the software of the extraction process with each hard disk image, Herbert et al. (2006). Later an expert witness can determine from the documented software what possible modifications were possible during the creation. It is then also possible to determine the applicability of the software on the basis of assessments provided through the public evaluation (e.g., through NIST).

5.4 Handling of Multiple Hard Disks

In systems using different logical volumes (LVM), Hasenstein (2001) or different disks (RAID) Patterson et al. (1989), at first, a hash value has to be determined and digitally signed for each volume or each disk, respectively, followed by generating the new image. This image then must be assigned a hash value and signed digitally, making the connection between the individual data slices and the image re-constructible.

The generation of hash values and their signatures, however, do not solve the issue whether the image data is displayed actually unchanged (logical error) to the user in the forensic investigation. When evaluating the data, the image data will now be treated as a drive. Individual files are distributed over logical slices in the image and will be "assembled", just as they would on a hard drive. An expert must certify the software used for this process and its correct functioning must be proven.

If a single file is extracted from the image, the correlation to the image needs to be documented as well. As a minimum approach a signature can be used, stating who generated this file from which

[9] http://www.trustedcomputinggroup.org
[10] http://www.cftt.nist.gov/

image slices. In this case the signature would be formed with respect to the file, the signature of the image (or a hash value of these data) and the relevant image slices. Moreover, it is desirable that the software used be documented in the signature as well. Another issue is the proper presentation of the data. For example, documents from word processors are displayed with different content depending on the version of the program used. In case of doubt, an accurate analysis of the original data will have to be carried out and documented.

6. CONCLUSIONS

The integrity of digital evidence objects is central to the evaluation of the currently used process for the creation of images for storage devices. Current forensic tools do not use state-of-the art technology for the protection of images. Their weak protection mechanisms only cover accidental changes to image data.

However, the well-established state-of-the-art of technology provides solutions for the detection of deliberate or accidental alterations of digital evidence objects, the secure documentation of the state of the devices and software that was used to create the image, but also the deletion of irrelevant or exempt data at a later point in time, without affecting the protection of evidentiary data. Implementing these additional security measures is easy and straightforward. Furthermore, the technology can complement existing forensic tools without the necessity of a full integration into these tools. The described protection techniques can be implemented as separate software, since none of the protection techniques deals with actual creation process or evaluation process of the image itself.

Hash functions and digital signatures do not change the image. Further, they can be stored separated from the actual image data. Thus, the additional security measures will have no impact on the quality of the data itself during the backup process. Also existing solutions for presentation issues of image data and documentation are not affected. Interfaces can be easily defined to link extracted data to particular slices in order to also integrate deletion of private or unnecessary data from the stored image.

In summary, the law perspective should be aware of the technological state of the art and must create a clear demand for secure solutions and for solutions that are compliant with basic laws on people's privacy and on data retention.

REFERENCES

Andrew, M. W. (2007). Defining a process model for forensic analysis of digital devices and storage media. In Systematic Approaches to Digital Forensic Engineering, 2007. SADFE 2007. Second International Workshop, IEEE, 2007, 16–30.

Claar, J. F., Duvall, R. M., & Oliver, R.J. (2000). File system block sub-allocator, March 21 2000. US Patent 6,041,407.

European Committee for Standardisation (CEN). (2001). CWA 14169: Secure signature-creation devices "EAL 4+". CEN Workshop Agreement.

BSI (Bundesamt für Sicherheit in der Informationstechnik, Germany). (2011). (BSI). Leitfaden IT Forensik.

Garfinkel, S. L. (2006), Aff: A new format for storing hard drive images. *Commun. ACM*, *49*(2):85–87.

Hasenstein, M. (2001). The logical volume manager (lvm). White paper, 2001.

Herbert, H. C., David W Grawrock, D. W., Ellison, C. M., Golliver, R.A., Lin, D. C., McKeen, F.X., Neiger, G., Reneris, K., Sutton, J. A., Shreekant S., Thakkar, S.S., et al. (2006). Platform and method for remote attestation of a platform, January 24 2006. US Patent 6,990,579.

Jonsson, J., & Kaliski, B. (2003). Public-key cryptography standards (pkcs)# 1: RSA cryptography specifications version 2.1. Technical report, RFC 3447.

Knaus J. P., & Foley, T. E. (2001). Electronic records & signatures: The federal e-sign act and michigan ueta place them on legal par with their paper and ink counterparts. Mich. BJ, *80*, 39-40.

Knight (G. 2011). Forensic disk imaging report, Technical Report. JISC. (Unpublished)

Kuntze, N., Rudolph, C., Alva, A., Endicott-Popovsky, B., Christiansen, J., & Kemmerich, T. (2012). On the creation of reliable digital evidence. In G. Peterson and S. Shenoi, editors, *Advances in Digital Forensics VIII*. Springer, ISBN 978-3-642-33961-5.

Kunz, T., Okunick, S., & Pordesch, U. (2008). Data structure for security suitability's of cryptographic algorithms (dssc)-long-term archive and notary services (ltans). Technical report, IETF Internet-Draft.

Maurer, U. (1996). Modelling a public-key infrastructure. In *Computer Security-ESORICS*, *96*, 325-350. Springer.

NIST (2012), National Institute of Standards and Technology. Test Results for Digital Data Acquisition Tool: ASR Data SMART.

Patterson, D. A., Chen, P., Gibson, G., & Katz, R. H. (1989). Introduction to redundant arrays of inexpensive disks (raid). In COMPCON Spring'89. 34th IEEE Computer Society International Conference: Intellectual Leverage, Digest of Papers., IEEE, 112–117.

Pinheiro, E., Weber, W.D., & Barroso, L. A. (2007). Failure trends in a large disk drive population. In Proceedings of the 5th USENIX conference on File and Storage Technologies, 2.

Raghavan, S. (2013). Digital forensic research: current state of the art. *CSI Transactions on ICT*, *1*(1):91–114.

Saudi, M. M. (2001). An overview of disk imaging tool in computer forensics. SANS Institute.