Home Office

Fundamental principles for preparing the CAST Key Reference Hard Disk

Creating a reference drive to assist in the validation of a forensic disk imaging process

Publication number 063/16

Andrew Barnes, Paul Farr, Jade James, Pat Mason

November 2016

Summary

This document provides information about the underlying principles used by the Home Office Centre for Applied Science and Technology (CAST) in the preparation of each CAST Key Reference Hard Drive (CKRHD). It is intended to enable readers to understand the capability and utility of the prepared disk. In combination with the preparation procedure (see CAST publication 062/16), it will also assist competent users in designing and generating their own reference disks if desired. Whilst these documents outline the CAST approach to the preparation of reference drives, there may well be other appropriate methods. Organisations are responsible for verifying local deployments within their own quality system.

A CKRHD is a reference drive which can be used to assist in the validation of local procedures for the imaging or acquisition of digital storage media such as conventional hard disk drives. Specifically it will enable two key requirements (as found in ENFSI, 2015) to be tested:

- A complete copy of the persistently stored user-addressable data on the evidence item, as presented at the time of examination by the disk controller using the Logical Block Address (LBA) scheme, shall be acquired.
- The acquired image shall replicate the structure, order and contents of the user-addressable storage on the evidence item at the time of creation of the image.

The CKRHD is suitable to test these two requirements since the information which makes up the full extent of the user-addressable data content on the hard drive is created and characterised using a system which is independent of the imaging process under test. Hence the data content of the reference disk is explicitly known independently of any attempt to image the disk. This separation between the test data and the imaging process eliminates a potential source of bias in the validation exercise.

Note that it is anticipated that there will be a range of additional requirements for a local imaging process which will need to be validated using a range of other appropriate tests. The CKRHD is not intended to be suitable for all tests. For example, an imaging process may feature the requirement to acquire data from hidden or protected areas. The CKRHD would not be suitable for validating this requirement since it does not feature such areas.

Contents

1. Introduction

1.1. Background

As outlined in the Codes of Practice and Conduct for forensic science providers and practitioners in the Criminal Justice System Issue 3, the Forensic Science Regulator (FSR) and the National Police Chiefs' Council (NPCC) expect digital forensics units to gain accreditation to a recognised standard by October 2017 (see FSR, 2016). The digital forensic processes within the scope of accreditation are broad ranging, covering the acquisition and analysis of data from many different digital sources. The first milestone for providers of digital forensic services to achieve is the validation of methods to acquire data from conventional hard drives (known as *imaging*).

Deputy Chief Constable Nicholas Baker, the NPCC Digital Forensics Portfolio lead, requested assistance in addressing the requirements of the Regulator from the Home Office Centre for Applied Science and Technology (CAST). Following consultation with the Digital Forensics Portfolio Quality Standards Expert Network it was agreed that CAST would assist by defining standardised reference disks which local units can use as an element of testing and validating their local imaging process.

1.2. The reference disk and this document

The standardised reference disk is intended to help users to demonstrate that their local process can acquire a complete and accurate copy of the data stored on a conventional hard disk drive. It is expected that local users will make use of further test materials in order to fully validate their process against the complete set of local requirements.

Section 2 of this document includes definitions of key technical concepts, and covers the requirements for a typical imaging process and for a reference disk. Section 3 addresses the preparation principles including the selection of suitable equipment and suggested quality assurance checks. Recommendations regarding the use of a CAST reference disk, and consideration of other validation tests, are in section 4. References and a glossary are provided at the end of the document. The precise details of the preparation procedure used by CAST can be found in a separate document, publication 062/16.

2. Definitions and requirements

The intention of the CAST Key Reference Hard Drive (CKRHD) is to create a simple reference item which will address certain fundamental requirements of local imaging procedures. With the Forensic Science Regulator's initial focus being upon the imaging of conventional hard disk drives, the reference disk discussed here will be based upon the parameters relevant to a conventional hard disk drive. These terms are defined, and are followed by a discussion of the imaging process requirements included in ENFSI (2015), a subset of which contribute towards the specific requirements for the CKRHD.

2.1. Definitions

The definitions of the terms 'conventional hard disk drive', 'imaging' and 'reference disk' as used in this document are provided below. These definitions are consistent with those used in ENFSI (2015). Local definitions of the same terms may vary. If this is the case, readers need to understand how this will affect the capability and utility of the prepared disk.

2.1.1. Conventional hard disk drive

A 'conventional hard disk drive' is defined as a device which stores user-addressable data on rigid spinning platters coated in a ferromagnetic material. It communicates with a host device via an onboard disk controller over an ATA interface.

2.1.2. Imaging

'Imaging' is defined as the acquisition of an exact bit-stream copy of the persistently stored user-addressable data, as presented to the host by the disk controller using the Logical Block Address (LBA) scheme. The acquired image will replicate the structure and contents of the user-addressable data regardless of any file systems which may be present on the device at the time of creation of the image.

2.1.3. Reference disk

A 'reference disk' is a conventional hard disk drive which will provide a known and controlled data stream for capture when attached to an imaging system.

2.2. Typical imaging process requirements

An example set of requirements for imaging a conventional hard disk drive is included in Appendix E of ENFSI (2015). These requirements, illustrative of typical imaging process requirements, are reproduced in Table 1 along with an accompanying commentary. Local unit requirements may differ in detail according to operational demands. Care should be taken when undertaking an imaging process validation exercise to ensure that the tests and any reference disks which are used are suitable and sufficient for the local requirements.

Table 1. Typical imaging process requirements for a conventional hard disk drive, reproduced from ENFSI (2015). The accompanying comments are provided for additional context.

ID	Requirement	Comment
1	A complete copy of the persistently stored user-addressable data on the evidence item, as presented at the time of examination by the disk controller using the Logical Block Address (LBA) scheme, shall be acquired.	*The definition of the data which will be acquired by the imaging process. This definition encompasses the most commonly acquired data areas, colloquially known as a 'bit-for-bit physical image'.*
2	The acquired image shall replicate the structure, order and contents of the user-addressable storage on the evidence item at the time of creation of the image.	*The definition of the data which is the product of the process. Note that the requirement to replicate the data on the evidence item does not preclude the use of methods to increase storage efficiency or security, such as compression or encryption, so long as those methods are reversible and non-destructive. However, imaging issues such as repeated or swapped sectors will prevent this requirement from being satisfied.*
3	Areas hidden by the disk controller using widely recognised standard methods (Host Protected Area, Device Configuration Overlay) shall be acquired.	*The capture of (specific types of) hidden data areas is an operationally-driven requirement which is not common in detail across all units. Since there is not a standard approach to the acquisition of hidden data,* ***the CKRHD will be prepared without hidden areas.*** *If local units wish to acquire hidden areas, additional test material beyond the key reference disk will be necessary in order to demonstrate compliance with the local requirement.*
4	Vendor-specific storage areas such as reserved firmware addresses or service modules will not be acquired.	*Information stored in service areas, firmware or temporary caches is not frequently of value to an investigation and hence is not commonly acquired in standard processes.* ***The CKRHD will therefore not seek to standardise the information stored in vendor-specific storage areas****. If local units wish to acquire vendor-specific storage areas, additional test material beyond the key reference disk will be necessary in order to demonstrate compliance with the local requirement.*

ID	Requirement	Comment
5	The process shall interact with a conventional hard disk drive via an ATA interface.	*The standard interface for a conventional hard drive, as defined by 2.1.1. Devices sold as externally attached hard drives can be similar to conventional hard drives, but feature additional integrated electronics. These map the interface into a different format, such as USB. Since there is not an agreed method of mapping the ATA command set to other command sets, attempts to control such devices could produce unpredictable results due to manufacturers implementing the mapping in different ways. Local units should consider the impact of such issues on their imaging methods and mitigate them appropriately.*
6	All unresolved errors encountered during the acquisition of data from the evidence item shall be recorded.	*It is good practice to note problems encountered during an acquisition. However, limitations in available technology may affect error handling and prevent information such as error type and location being satisfactorily recorded. Due to the potential for differences in local unit's approaches to error handling,* ***the CKRHD will not feature deliberately introduced disk errors such as bad sectors****. Additional test material beyond the key reference disk will be necessary in order to demonstrate compliance with the local requirement.*
7	An auditable link shall be maintained between the acquired data and the original physical evidence item.	*The imaging process must maintain the continuity of the evidence. This is commonly achieved through procedural controls, and an appropriate means of demonstrating the effectiveness of these controls should be determined by local units.*
8	The integrity of the acquired data shall be maintained in a manner which is traceable back to the original acquisition from the physical evidence item.	*There should be controls built in to the imaging process and any subsequent evidence handling procedures to ensure and check that the acquired data continues to fairly represent the original physical evidence item. An appropriate means of demonstrating the effectiveness of these controls should be determined by local units.*

ID	Requirement	Comment
9	The imaging procedure shall not add to, remove or modify the original user-addressable data which is stored on the evidence item.	*There should be controls built in to the imaging process to protect the original evidence item. An appropriate means of demonstrating the effectiveness of these controls should be determined by local units. The CKRHD is not intended for this purpose.* *Despite any controls which are in place, there may still be a residual risk of the original evidence item being changed; in certain circumstances (e.g. a damaged disk or one which features a high degree of data fragmentation) the disk controller may cause changes to occur to the user-addressable data independently of any command or interaction from a host system or imaging process.* *Also, in certain circumstances it may be necessary to deliberately modify the behaviour of the disk controller in order to access hidden areas. Any such changes and the associated impact should be highlighted in the audit trail.*

2.3. CAST Key Reference Hard Disk requirements

Table 2 contains the requirements and constraints which dictate the design of the CKRHD. Whilst there can be considerable variety in the detail of local imaging requirements, it is reasonable to expect certain fundamental requirements to be common across a wide range of users. These requirements were determined in consultation with the NPCC expert network.

To ensure wide applicability, the CKRHD is targeted towards the testing of just two of the requirements noted in ENFSI (2015). Therefore, a number of the generic imaging process requirements are outwith the scope of the CKRHD. Alternative tests must be implemented, or a modification made to the basic CKRHD, in order for such requirements to be validated.

Table 2 Requirements and constraints for the CKRHD, derived in consultation with the NPCC expert network based upon the generic imaging process requirements listed in ENFSI (2015) and reproduced in Table 1.

ID	Requirement	Comment
A	The reference disk shall produce a known and controlled output when attached to an imaging system.	*To enable a local user to evidence a claim that their imaging process acquires all data, all data on the reference disk must be known and understood independently of the imaging process under test. This requirement is derived from entries 1 & 2 in* Table 1.
B	Each individual sector of data on the reference disk shall contain uniquely identifiable information.	*To evidence a claim that the imaging process has replicated the structure, order and contents of the target disk, it must be possible to discriminate between each logical segment of information on the reference disk. This requirement is derived from entry 2 in* Table 1.
C	The reference disk shall **not** contain areas which are hidden by the disk controller using widely recognised standard methods (Host Protected Area, Device Configuration Overlay).	*This requirement explicitly differs from entry 3 in* Table 1*. Each local unit will make their own decision as to whether capturing hidden data will form part of their process. With no standard national approach, the CKRHD will not include hidden areas. However, suitable hidden areas could be added to a standard CKRHD to match local requirements.*
D	The reference disk shall interact with a local imaging system via a SATA interface.	*A conventional hard drive is defined with an ATA interface. The SATA variant of the ATA interface is noted as the most frequently encountered in casework by the NPCC network.* ***The interface can be altered to match local requirements.*** *This constraint is derived from entry 5 in* Table 1.

3. Preparation principles

The information which makes up the full extent of the user-addressable data content on the reference disk should be created and characterised using a system which is independent of the imaging process under test. The separation between the preparation of the test data, and the deployment of the imaging process which is the subject of validation, eliminates a potential source of bias in the validation exercise.

The following sections outline the selection of suitable equipment, the preparation of a device which contains suitable test data, and suggested quality assurance checks to confirm that the reference disk is in the expected state. The details of the preparation procedure for the CKRHD are described in CAST publication 062/16.

3.1. Equipment

The core equipment to prepare and assure a reference disk is listed below.

- Device to act as the reference disk.
  - Conventional hard drive with a SATA interface.
- System to prepare the reference disk.
  - Computer system with spare SATA interface on the motherboard.
  - Appropriate operating system or software environment.
    - Possesses the capability to pass and handle relevant ATA commands.
  - Data generating software.
    - To create the correct quantity of known data for the reference disk.
  - Data checking software.
    - To view raw data contents.
    - To identify data using a hash value.
    - To inspect the complete set of known data for conformity on a byte-by-byte basis.
- Disk imaging equipment.
  - System of hardware and software to image the reference disk.

The selection of an appropriate device to act as a reference disk should be based upon a consideration of the constraints of the end user, and the requirements and associated risks of the imaging process which is to be validated. For example, to minimise the resource overhead of performing the validation tests, low capacity disks could be selected. Low capacity disks will minimise the data storage requirement and may also minimise the time required to complete a test. However, relying solely on low capacity disks could introduce risk or uncertainty relating to the capability of the imaging process to handle large capacity disks. This risk will need to be managed appropriately.

The system used to prepare the reference disk should be capable of communicating directly with the reference disk. It should have a physical interface which is compatible with the reference disk, and the hardware and software should be capable of passing and handling the necessary commands to characterise and control the parameters of the reference disk accurately. It will be necessary to generate and characterise the data contents to be placed on

the reference disk, and to physically transfer the data to the drive.

Once the reference disk has been prepared and set to a known state, equipment will be required to confirm that the disk is indeed in its expected state; this assurance could be provided by direct manual inspection of the data, by the use of summary identifiers such as hash values, by explicit byte-by-byte comparison of the data stream, or by a combination of these checks. Consideration should also be given to performing a verification check on the prepared reference disk using a typical disk imaging process to provide further confidence that the reference disk will behave as expected once applied to an imaging system which is the subject of validation.

3.2. Disk preparation

The basic stages of disk preparation are:

- Device characterisation
- Data generation
- Data characterisation
- Data transfer

3.2.1. Device characterisation

The device which is to act as the reference disk should be characterised. At a minimum, this will involve a determination of the precise storage capacity and establishing the status of user-addressable storage areas which are hidden or protected by the disk controller. Whilst the label on a conventional hard disk drive will provide limited information related to the capacity of the disk, this information should be supplemented by directly querying the onboard disk controller via the native interface. This will establish a baseline ground truth for the full extent of the user-addressable storage capacity as presented by the disk controller, and the presence or absence of hidden or protected areas.

3.2.2. Data generation

Once the full extent of the user-addressable storage area on the reference disk has been determined, a data stream can be generated to precisely match that extent. The data stream should avoid sparse or repeated content in order to enable the identification of and discrimination between individual logical sectors. This precludes the use of ‘realistic’ data contents, which characteristically feature large quantities of unused space in addition to segments of repeated information. Such data present an opportunity for any errors in the imaging process, such as anomalously repeated or incorrectly read data, to be missed when testing. A more effective approach is to ensure that each sector on the disk features unique, uncorrelated content. In order to be realistically verifiable, this unique, uncorrelated content needs to be produced in a deterministic manner. The use of a deterministic data generator will allow for direct verification of the data stream against the predicted output based upon the generating algorithm.

3.2.3. Data characterisation

The generated data stream can be characterised by passing it through an algorithm to calculate a deterministic summary value which identifies that particular data stream. This summary identifier may be used to verify the integrity of the data in the disk assurance process. It can also act as evidence that a disk imaging process has correctly acquired all of

the user-addressable data on the reference disk, as presented by the disk controller. Cryptographic hash functions such as MD5 and SHA1 are commonly used for verifying the integrity of data (Genoe, 2013).

3.2.4. Data transfer

Once the generated data stream has been characterised, the full extent of the generated data may be written to the reference disk. This operation should ensure that the complete contents of the user-addressable storage areas on the disk, as presented by the disk controller, are occupied by a known data structure and hence the reference disk is set to a known state.

3.3. Disk assurance

Quality assurance checks should be performed on the prepared reference disk to confirm that the disk is in the expected state and is therefore suitable for use in validating an imaging process. The checks should establish that the reference disk contains only the full extent of the known, generated data. There are a range of different potential checks which could include:

- Direct manual inspection of data.
- Summary identifier comparison.
- Explicit byte-by-byte comparison of the data.
- Imaging test

3.3.1. Direct manual inspection

Direct manual inspection of the data on the prepared reference disk and comparison against the generated data stream can be undertaken using software which is capable of viewing the raw data contents. Whilst it is not feasible to expect a human to be capable of thoroughly inspecting the complete data content of a typical storage system, sampling a subset of the data content can provide added confidence that the prepared reference disk is in the expected state. The extent of the dip sampling should be risk managed proportionately in the context of any other quality assurance checks which are performed.

3.3.2. Summary identifier comparison

Data which is acquired from the disk via the onboard disk controller may be passed through an algorithm to calculate a deterministic summary value which may be compared to that produced by the known generated data stream. The summary identifiers can be compared for similarity. A mismatch in the summary identifiers will indicate an issue which requires further investigation; the reference disk is not suitable for use until the issue has been identified and resolved.

The use of hash functions such as MD5 or SHA1 to generate the summary identifiers will provide a high level of confidence that a match between the summary identifiers means that the data streams match in terms of content and structure. However, note that it is possible (but highly improbable) that two different data stream inputs will lead to the same hash value output.

3.3.3. Byte-by-byte comparison

If the reference disk is filled with a data stream generated by a deterministic algorithm, the same algorithm can be used to verify on a byte-by-byte basis that the stream has been

successfully transferred onto the reference disk. A mismatch between the data which is acquired from the reference disk and the output predicted by the algorithm will require further investigation.

3.3.4. Imaging test

A verification check can be performed on the reference disk by examining it with a typical imaging system which is not the subject of validation. This may be used to confirm that the reference disk behaves in the expected manner, with the acquired data being of the correct extent and being captured within the expected timeframe. An acquisition which takes longer than expected might be an indication of a damaged or failing disk and such issues should be investigated prior to deployment of the reference disk in a validation procedure. Manual direct inspection, summary identifiers and explicit byte-by-byte comparison may all be used to further confirm that the acquired image matches with the known generated data.

4. Recommendations

4.1. Usage of the CKRHD

The CKRHD is intended for use to assess two specific requirements which are stated in ENFSI (2015) as:

- A complete copy of the persistently stored user-addressable data on the evidence item, as presented at the time of examination by the disk controller using the Logical Block Address (LBA) scheme, shall be acquired.
- The acquired image shall replicate the structure, order and contents of the user-addressable storage on the evidence item at the time of creation of the image.

The user-addressable storage area of the CKRHD contains a well-known set of data of precise length and with unique content contained within each different sector. Due to this characteristic, imaging errors such as incomplete data acquisition and misread or disordered data can be identified by simple comparison between the known data stream and the product of the imaging process. This comparison could be achieved by leveraging common summary identifiers such as hashing functions, by direct manual inspection of the data or by programmatically comparing the full extent of the captured data against the expected result.

Successfully matching the known data stream provides demonstrable evidence that an imaging process can fulfil the stated requirements. However, consideration must still be given to the extent to which the requirements have been demonstrated, and any additional risks within the local imaging procedure which may be relevant to these requirements.

In addition, there are anticipated to be a range of further requirements for an imaging process (see Table 1 in section 2.2) which must be evidenced in order to fully validate the process.

4.2. Consideration of other suitable validation tests

The CKRHD is not intended to be used as the sole validation test for a local imaging process. Local decisions and investigative needs will necessitate additional assessments to be performed in order to fully validate the local process against the complete set of stated requirements.

ENFSI (2015) lists nine separate requirements for a process to image a conventional hard disk drive, of which only two are intended for evaluation using the CKRHD. It may be possible to validate some of the remaining ENFSI requirements using the CKRHD, though some of these tests may introduce the risk of damage to the reference disk. If the CKRHD is damaged or altered, it should be removed from its intended service until it has been reset to a known and controlled state.

The third requirement in the ENFSI document (related to hidden or masked storage areas) is not addressed by the CKRHD. In order to test such a requirement, an area protected by the disk controller may be overlaid onto a prepared reference disk to obscure an arbitrary segment of data. An imaging system which is capable of overcoming the disk controller protection

should produce an image which is identical to the contents of the original reference disk, prior to the obscuration being introduced; the cause of any unexpected difference must be investigated. Consideration should be given to whether the method used to introduce or remove disk controller protection causes the relevant data already resident on the disk to change in terms of structure or content, and whether the impact of this change is acceptable.

Another potential requirement not addressed by the CKRHD is the capability of the local imaging process to handle error conditions. This could be tested through the use of a conventional hard disk drive which features known or deliberately introduced errors, or through the use of a protocol analyser which can deliberately inject error conditions in a controlled manner into the communication protocol.

When constructing validation exercises, thought must always be given to the purpose or requirements of the system under test, and how it may be demonstrated that the purpose or requirements have been met. Note that it is rarely possible to exhaustively assess the system under test within the scope of all potential operating conditions; as a result, the implications of limited testing and the risks which arise must be considered and handled appropriately according to local circumstances.

4.3. Alternative storage device types

By defining the imaging process as a transaction between the host imaging system and a target disk controller via an ATA interface, the precise nature of the storage system (whether it be a solid state drive or a conventional hard drive for example) can be considered a black box which is undefined. Therefore, a conventional hard disk drive prepared in accordance with the CKRHD methodology should be capable of testing the specific requirements noted in section 4.1 for any type of storage medium which communicates with the outside world via an onboard disk controller over an ATA interface; at this level of abstraction, a solid state drive and a conventional drive are functionally equivalent.

Mass storage media which use non-ATA interfaces (e.g. USB) will operate in a conceptually similar manner; the imaging system will attempt to recover the full contents of the user-addressable storage as presented by the disk controller at the time of imaging. Whilst there may be alternative risks engendered by the different interface protocol, the underlying principle of a reference disk featuring known data will still hold.

5. References and acknowledgements

5.1. References

CAST (2016) *CAST Key Reference Hard Disk preparation procedure*. Publication number 062/16.

ENFSI (2015) *Best practice manual for the forensic examination of digital technology*. European Network of Forensic Science Institutes, Forensic Information Technology Working Group, Version 01.

FSR (2016) *Codes of practice and conduct for forensic science providers and practitioners in the Criminal Justice System*. Issue 3.

Genoe, R. (2013) *Managing a digital investigation unit – a handbook for senior law enforcement officers*. University College Dublin Centre for Cybersecurity and Cybercrime Investigations, page 19.

5.2. Acknowledgements

Grateful thanks to the Information Security Research Group at the University of South Wales for peer reviewing the technical method.

6. Glossary

ATA	Advanced Technology Attachment. An interface standard featuring a common command set.
CAST	The Home Office Centre for Applied Science and Technology.
CKRHD	CAST Key Reference Hard Disk.
ENFSI	European Network of Forensic Science Institutes.
MD5	Message Digest 5, a cryptographic hashing function.
NPCC	National Police Chiefs' Council.
SATA	Serial ATA, an interface standard for the connection of storage devices.
SHA1	Secure Hashing Algorithm 1, a cryptographic hashing function.

ISBN: 978-1-78655-132-0