Selective and Intelligent Imaging Using Digital Evidence Bags

By

Philip Turner

From the proceedings of

The Digital Forensic Research Conference

DFRWS 2006 USA

Lafayette, IN (Aug 14th - 16th)

DFRWS is dedicated to the sharing of knowledge and ideas about digital forensics research. Ever since it organized the first open workshop devoted to digital forensics in 2001, DFRWS continues to bring academics and practitioners together in an informal environment.

As a non-profit, volunteer organization, DFRWS sponsors technical working groups, annual conferences and challenges to help drive the direction of research and development.

http:/dfrws.org

available at www.sciencedirect.com

SCIENCE DIRECT®

journal homepage: www.elsevier.com/locate/diin

Digital Investigation

Selective and intelligent imaging using digital evidence bags

*Philip Turner**

QinetiQ, Digital Investigation Services, Trusted Information Management Department, St. Andrews Road, Malvern Worcestershire WR14 3PS, UK

Keywords:
Selective imaging
Intelligent imaging
Digital forensics
Digital evidence bags
Digital investigation

ABSTRACT

This paper defines what selective imaging is, and the types of selective imaging that can be performed. This is contrasted with intelligent imaging and the additional capabilities that have to be built into an imager for it to be 'intelligent'. A selective information capture scenario is demonstrated using the digital evidence bag (DEB) storage format. A DEB is a universal container for digital evidence from any source that allows the provenance to be recorded and continuity to be maintained throughout the life of the investigation. The paper concludes by defining the 'ultimate test' for an intelligent and selective imager approach.

1. Introduction

Traditionally the forensic acquisition process used for the capture of digital evidence had been quite an unsophisticated process. Once initiated the capture commenced at the beginning of the media and continued until the end of the media has been reached, with little error handling or recording capability. This approach, when the capacity of the source device is small, adequate time is available and a fully functional device used, is probably suitable in most situations. However, there are instances even when there are no constraints that you have to ask the question, 'Is it the best approach?'

For example, one of the actions when investigating the content of an image may be to sort the contents of the acquired evidence by file type and browse them. This allows the investigator to get a very quick idea of the types of application and work that has been performed on the system. Although this is a very quick and easy procedure to perform it is used to provide an initial pointer and may well influence the further tasks undertaken, for example, if a higher proportion of word processing documents are found then these may well initiate keyword searches. Alternatively, if a high proportion of graphics files were found then this may well initiate viewing the contents of those to categorise them. Based on the results of these initial findings, more advanced processes may then be commenced, for example, in the first instance data carving from drive free space for word documents, or in the second case data carving for particular graphics file formats.

2. Selective imaging

Selective imaging is a term that has been bandied about by many forensic practitioners for a while now. This is a term that is generally associated with the decision not to acquire all the possible information during the capture process. However this does not have to be so. At least in some official good practice guides (ACPO) it is now recognised that 'partial or selective file copying may be considered as an alternative' when it may not be practical to acquire everything. The usual reason for applying a selective approach is the quantity of

* Tel.: +44 1684 895777; fax: +44 1684 894365.
E-mail address: pbturner@qinetiq.com
1742-2876/$ – see front matter
doi:10.1016/j.diin.2006.06.003

information that may have to be acquired. Other reasons for performing a selective acquisition include, but are not limited to, forensic triage, intelligence gathering and legal requirements. There may be legal reasons why a selective approach should be adopted, for example, a case involving legal professional privilege (LPP) material. Adopting a selective approach has risks associated with it as highlighted in Kenneally and Brown's (June 2005, 2005) papers, but this in no way means the evidence should not be gathered in any less scientific or rigorous manner.

So how is selective imaging performed?

There are several types of selective imaging techniques that could be used. They are:

- Manual selective imaging;
- Semi-automatic selective imaging;
- Automatic selective imaging.

Manual selective imaging is where the forensic investigator chooses exactly which files are captured. For example, the investigator can use an interface similar to that of a file browser and is able to navigate the directory tree and choose which files to acquire.

Semi-automatic selective imaging is where the forensic investigator decides which file types or categories of information to capture. This may be based on file extension, file signature or file hash. When using a selective approach based on file hashes it is important to record which files are present and their provenance, even though the contents of each file may not be captured. It would also be prudent to record referential hash set information.

Automatic selective imaging is where the investigator selects the source and destination devices and the imager automatically acquires the evidence. This is accomplished in a selective manner according to pre-configured parameters or the particular circumstances pertaining to the case/investigation.

The different operating modes that a selective approach presents to the investigator, combined with the flexibility and many options for classifying and grouping information, potentially makes it very complex. One of the difficulties with selective imaging is recording the provenance of each item selected. This provides a number of options and there is often more than one metric that can be used to record the provenance of an item of information. For example, the location of a particular file on a disk could be recorded in the following ways:

- Physical sector locations (data runs);
- Logical cluster locations within a volume, with the addition of an offset from the beginning of the physical device;
- Folder location specified from the root folder. This must include partition reference information.

This leads to the questions; which of these is best? does it matter? does one method mean the evidence is 'better' or has more integrity than another. Should we just record as many provenential references as possible? One of the things to bear in mind with these is the attributes of provenance (Turner, February 2005):

- Unique;
- Unambiguous;
- Concise;
- Repeatable.

It can be argued that in its own way each method meets these criteria. It just depends upon the technical knowledge of the person trying to understand it. For example the general public, judge or legal professional is likely to be more familiar with a folder location than a more technical absolute disk sector or cluster reference. These other 'more technical' provenance descriptions may only complicate matters by introducing more technical vocabulary that actually detracts from and obscures the real information that is trying to be presented. In an ideal world we would record all provenential descriptions.

This would lead to multiple provenential definitions:

- Primary provenential key = physical sector locations;
- Secondary provenential key = logical cluster locations within a volume with the addition of an offset from the beginning of the physical device;
- Tertiary provenential key = folder location specified from the root folder.

3. Intelligent imaging

An intelligent imaging approach is the process of capturing the knowledge and experience of domain experts into an intelligent system. This enables the investigator who is not technically proficient, and who is aware only of the type of investigation they are conducting to use this type of imager. For example, they may be investigating a fraud, intellectual property theft, or possession and distribution of indecent material, and do not know what file types or locations that information may reside in that is pertinent to their case. They simply select the type of inquiry that is being conducted and the imager has the necessary intelligence built into it to acquire everything that would normally be relevant to the case. Intelligent imaging is not an attempt to use artificial intelligence methods and techniques in deciding what to capture. It could however alert the examiner to the presence of other categories of material outside the initial line of enquiry.

There are however risks and difficulties that have to be overcome in order to adopt this approach:

- How do you go about capturing the knowledge of the technical experts that are familiar with digital technical complexities and legal domain experts and combine them?
- How do you know that you have captured everything relevant to the case under investigation or have not missed evidence of other offences?

These two points are outside the scope of this paper but need to be considered in the future when tools and systems have been developed that are capable of performing intelligent imaging.

4. Digital evidence bags

Both selective and intelligent imaging techniques offer many more options and capabilities than current bit stream imaging. There are currently no commercial tools that perform selective imaging and adequately record the provenance of the selected information.

Furthermore, no method has existed that captured the criteria or method used by the examiner in deciding what to acquire. For example, was an arbitrary manual selection used or was information captured based on category of information, file extensions, file signature or hash set.

The proposed solution to these problems is by the use of the digital evidence bag (DEB) format (Turner, 2005, 2005–2006). A DEB is a universal container for digital information from any source. It allows the provenance of digital information to be recorded and continuity to be maintained throughout the life of the exhibit. Additionally, DEBs may be encapsulated within other DEBs. This feature differentiates the DEB structure from that used by current monolithic formats commonly in use.

Fig. 1 shows the environment that a selective/intelligent imager operates in. An imager of this kind maybe device specific but the data from any source device is captured into a homogenous DEB. This permits the examination and analysis application to be compatible with DEBs but independent of many disparate source devices.

The main components of a DEB are the tag, index and bag files (Fig. 2). The index and bag files together are known as an evidence unit (EU). It is the EU coupled with the customisable index definition that provides the tremendous flexibility afforded by the DEB framework. It is this flexibility that is required to implement selective and intelligent imaging methodologies.

A selective imager should be able to operate in the modes defined earlier and be able to create a DEB in either a manual, semi-automatic or fully automatic mode. To date a selective imager has been created which is capable of operating in both a manual and semi-automatic mode.

In the manual selective mode of operation the examiner is able to capture information to single or multiple EUs as required. In this mode the content type is recorded as manual with an arbitrarily defined label.

Imager configuration files (Fig. 3) and category definition files are used to automate the imaging process and can provide a consistent approach to acquiring evidence in a selective manner.

Fig. 4 shows a more detailed structure of the DEB components.

The DEB tag file shown is an example of that created once the acquisition process is completed and the DEB is

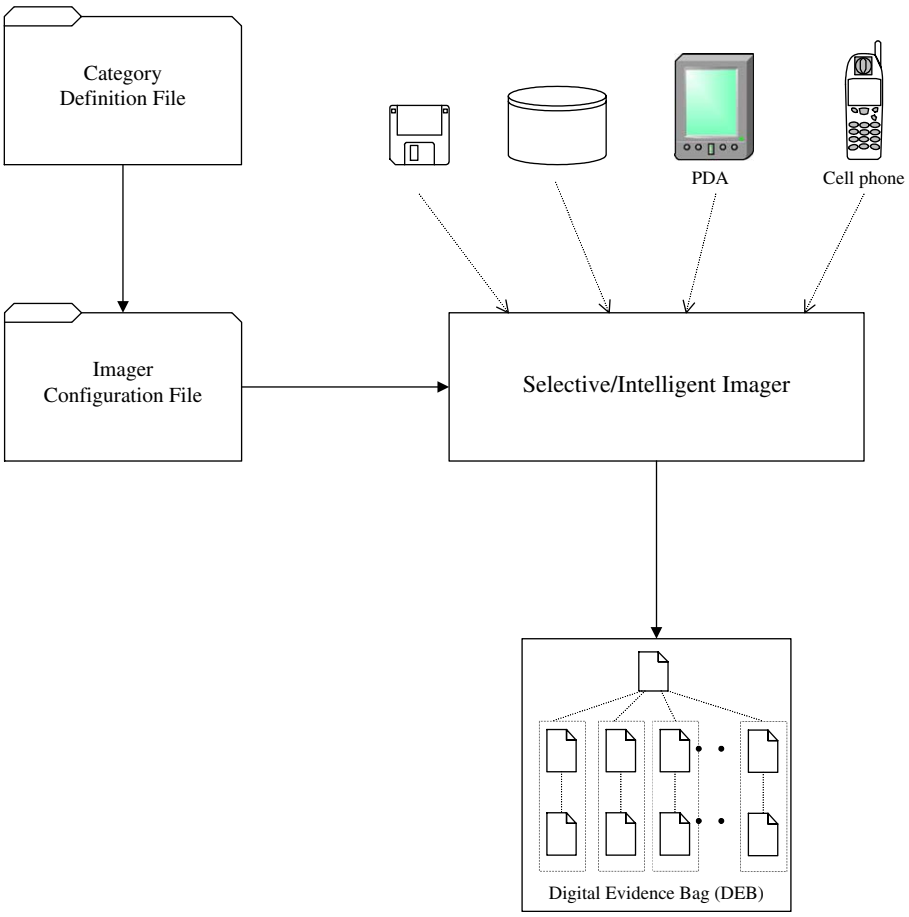

Fig. 1 – Selective/intelligent imager environment.

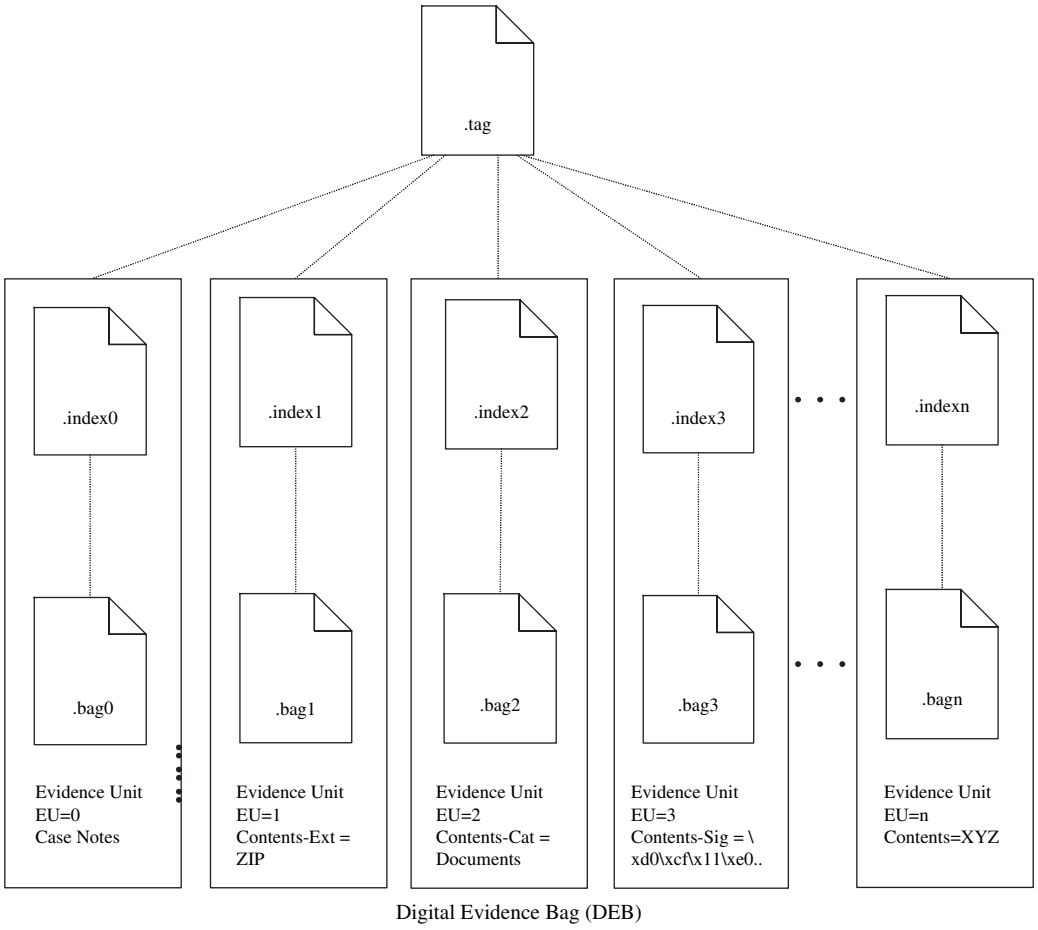

Fig. 2 – Digital evidence bag framework.

closed. The tag file is a plain text file comprising four main sections:

[DEB header];
[Evidence units];
[DEB footer];
[TCB].

The DEB header contains information such as investigating officer, timestamp of when the DEB was created, and description of what, where and when evidence was captured. Within the DEB header the 'Index Format' line specifies the default content sequence of the DEBs index files, it may also be specified per EU. The index file format is defined by a sequence of meta-tags. This allows each EU to be customisable within the DEB thus enabling a DEB to store information from a wide range of devices.

There are many index file meta-tags defined for use in DEBs and are a shorthand notation for the information they represent. They broadly fall into four categories, some examples are:

- Labels – file name and path (F), origin description (P), file attributes (Fa), command (C);
- Timestamps – last modified/completed (Tmod), accessed (Tacc), created/started/commenced (Tcre);

```
! Example Imager Configuration File        -- Comment lines
[EU]                                       -- EU definition selecting contents based
.ZIP                                       on File Extension
[EU]
=Documents                                 -- EU definition for category 'Documents'
.DOC                                       selecting contents based on File Extension
[EU]
\xD0\xCF\x11\xE0\xA1\xB1\x1A\xE1\x00\x00   -- EU definition selecting contents based
[EU]                                       on File Signature(header)
=Graphics Files                            -- EU definition for category 'Graphics
.BMP                                       Files' selecting contents based on multiple
.JPG                                       File Extensions
! Done
```

Fig. 3 – Example imager configuration file.

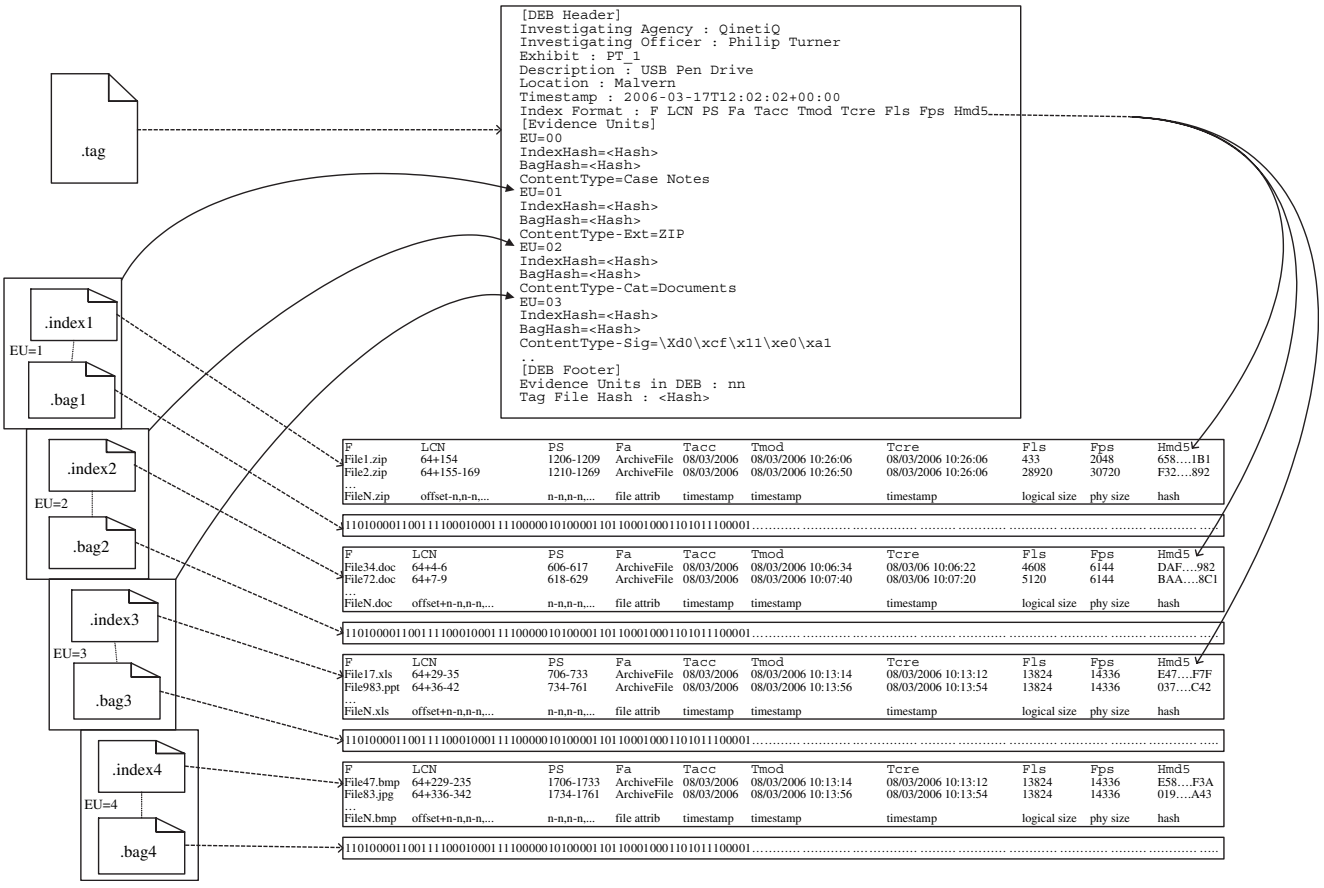

Fig. 4 – Example digital evidence bag.

- Numeric – physical sector (PS), logical cluster number (LCN), file logical size (Fls), file physical size (Fps);
- Integrity – MD5 hash (Hmd5), SHA hash (Hsha).

The evidence units' section of a DEB tag file is used to record all EUs created in the DEB. Each EU has its integrity hash of the index file and an integrity hash of the bag file.

Evidence Unit 0 is reserved for case notes and case associated metadata at the time of DEB creation. It contains information about the imager used to create the DEB, including revision number, imager application integrity hash, imager configuration file and details of capture selection criteria. Additional case information can be included in this EU, such as photographs and free-form text.

The content type of other evidence units is arbitrary and is determined by the examiner based upon the case requirements and/or configuration of the selective/intelligent imager acquisition tool. In Fig. 4, Evidence Unit 1 contains files with a defined signature (header). Alternatively, there are a number of options that may be used to identify the contents of a particular EU:

- `ContentType-Sig=<File Signature1>,<File Signature2>...`
- `ContentType-Ext=<File Extension1>,<File Extension2>...`
- `ContentType-Cat=<Category Type>`
- `ContentType-Manual=<label> {contents are Manually Selected}`
- `ContentType-CLI=<label> {contents are from a Command Line Interface}`

The DEB footer section is used to record the number of EUs within the DEB. The DEB is then sealed with a tag file integrity hash written at the end of the file.

Tag continuity blocks (TCB), although not shown in this example, are subsequently appended to the end of the DEB tag file when analysis of the evidence is undertaken. TCBs record the application function, signature and timestamp of when the DEB was accessed.

The bag files within each EU contain a concatenation of the binary information referenced by each entry in the corresponding index file.

5. The ultimate test

The 'ultimate test' for any imager and container that does not generate or store a standard bit stream copy is defined as follows.

The method and storage container used must be able to store sufficient information about the provenance of the information captured such that when the information is restored it is identical to that which would have been acquired should a bit stream image have been taken.

The practicalities of achieving this are slightly more involved. This requires an application that is capable of parsing DEB index file primary provenential key values in ascending order and generates a cryptographic hash over the corresponding bag file contents. This would form the basis of an application that was able to generate an image that is identical to that of the source device. The logical progression of this would be the restoration of the contents of the DEB to create a clone.

6. Conclusions

This paper has demonstrated the multitude of options that are available when capturing data in a selective manner. The importance of the container into which the information is placed cannot be overstated, as without a defined logical and structured approach the container could hinder the examination and analysis phases.

It is only by utilising such techniques as those described in this paper that we can better understand the demands that could be placed on this type of approach. Selective imaging has had much lip service over the past few years but very little work has been undertaken into the implementation of such an approach.

The methodology described and demonstrated in this paper is in sharp contrast to the archaic bit stream imaging methods currently used. It highlights the way forward to structuring the vast amounts of information in order to conduct an effective investigation.

REFERENCES

Association of Chief Police Officers (ACPO). Good practice guide for computer based electronic evidence. Version 3.

Kenneally Erin, Brown Chris. Risk sensitive digital evidence collection. Digital Investigation June 2005;2(2) (Elsevier).

Kenneally Erin, Brown Chris. Realizing risk sensitive evidence collection. In: Proceedings of Digital Forensic Research Workshop (DFRWS), http://www.dfrws.org/2005/proceedings/keneally_risk_slides.pdf; 2005.

Turner Philip. Digital provenance – interpretation, verification and corroboration. Digital Investigation February 2005;2(1) (Elsevier).

Turner Philip. Unification of evidence from disparate sources (digital evidence bags). In: Digital Forensic Research Workshop, http://www.dfrws.org/2005/proceedings/turner_evidencebags.pdf; 2005.

Turner Philip. Digital evidence bag format definition; 2005–2006.

Philip Turner has worked at QinetiQ (formerly known as the Defence Evaluation and Research Agency) for 21 years. He originally studied electronics and then moved into the area of information security and computer networking. He graduated from the Cheltenham and Gloucester College of Higher Education with a Bachelor of Science with Honors Degree in Computing and Real-Time Computer Systems in 1995. He is currently studying for a Ph.D at Oxford Brookes University. He has been working in the field of computer forensics and data recovery for over 8 years as Technical Manager in the Digital Investigation Services, Trusted Information Management department at QinetiQ.