Unification of Digital Evidence from Disparate Sources

By

Philip Turner

From the proceedings of

The Digital Forensic Research Conference

DFRWS 2005 USA

New Orleans, LA (Aug 17th - 19th)

Unification of Digital Evidence from Disparate Sources (Digital Evidence Bags)

Philip Turner

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

1. Abstract

This paper outlines a new approach to the acquisition and processing of digital evidence obtained from disparate digital devices and sources. To date the capture of digital based evidence has always been in its entirety from the source device and different methods and containers (file types) are used for different types of digital device (e.g. computer, PDA, mobile phone). This paper defines a new approach called a Digital Evidence Bag (DEB) that is a universal container for the capture of digital evidence. Furthermore, the Digital Evidence Bag concept could be used to permit the streamlining of data capture and allow multiple sources of evidence to be processed in a multiprocessor distributed environment and thereby maximizing the use of available processing power. The approach described in this paper allows for the first time the forensic process to be extended beyond the traditional static forensic capture of evidence into the real-time 'live' capture of evidence. In addition to this the Digital Evidence Bag can be used to provide an audit trail of processes performed upon the evidence as well as integrated integrity checking.

2. Introduction

Traditional computer forensics is on the edge of a precipice. Some practitioners might actually say that we have gone over the edge and are plummeting into the depths of oblivion with no end in sight. The reason for this imminent doomsday is the sheer volume of data that has to be processed during the course of a digital forensic investigation. For example, it is not uncommon to see single hard disk drives in excess of 350Gb. This is compounded by the fact that current forensic tools are being stretched past the limit of what they were designed to do thus resulting in the whole forensic process becoming problematic.

Furthermore, the situation still doesn't get any better when you take into account the diverse number of devices that process digital information and are capable of having digital information extracted from them. This just means that even more specialised applications have to be learnt, understood and used by the forensic practitioners in order to capture the information.

So why are we in this woeful state of affairs? Well, there is not a single reason, but perhaps the main one is the fact that when digital forensics is undertaken the only actual forensic task is the capture of the image from the original media. This is the source of our troubles, well, to be precise the actual containers that we capture the information into is the problem. The reason for this is that to process the captured information, the forensic image has to be processed as a single entity by the analysis tool.

Reference: QinetiQ/S&DU/TIM/WP050322

The second reason why we are having difficulty is that each forensic capture utility for the diverse range of digital devices, captures the information into differing format containers. That is not to say that a single format container should be used to capture data from a computer as from a PDA or live network packet capture, just that the wrapper used should be consistent.

3. Traditional Evidence Capture

In the world of law enforcement when a crime scene is visited in the course of an enquiry or investigation, the law enforcement officers use bags and seals to store the items of evidence that are found which are considered relevant at the time. Note that not everything can be captured, for example they very rarely dismantle a building brick by brick and take it away for analysis. In the digital forensic world however, we have the advantage that, should we should wish, it is possible to capture everything.
The item would then be placed into a bag that is sealed at the scene. The seal number is recorded and a tag is attached which may include details such as:

Investigating Agency / Police Force;
Exhibit reference number;
Property reference number;
Case/Suspect name;
Brief description of the item;
Date and time the item was seized/produced;
Location of where the item was seized/produced;
Name of the person that is producing the item as evidence;
Signature of the person that is producing the item;
Incident/Crime reference number;
Laboratory reference number.

The tag also contains sections for continuity purposes that can be signed when other people take custody of the item. This is used to provide continuity and assure provenance of the item from the time the item was seized to the time the item is used as evidence in court, restored to the owner or destroyed.

The continuity sections usually show the following details:

Name/Rank and number of person taking custody of the item;
Signature of the person that is taking custody of the item;
Date and time the person takes custody of the item.

It is not uncommon for many bags of evidence to be seized when a crime scene is visited and the size, shape and type of those bags varies depending upon the contents and type of article. For this reason different capacity bags are used.

This individual wrapping also permits various articles to be distributed between the various specialist laboratories that can process that item. For example some items may require fingerprint analysis, others may require DNA analysis, whilst others may just require interpretation of their contents by the investigating officer.

So how can the tried and tested method of evidence capture described above be undertaken in the rapidly changing digital world?

4. Digital Evidence Capture

Currently in the digital world the closest equivalent to the physical evidence capture process is either the plain ‘dd’ image file [1] or the

proprietary format produced by the forensic tool vendors.

The 'dd' raw file capture contains no method of attaching details such as the date and time of capture, the person performing the capture process, or any mechanism to help assure the integrity of what has been captured. These features can be generated after the capture but usually require additional actions of the person carrying out the process as separate distinct functions.

In contrast to this, some proprietary formats [2] allow some of these details to be entered when the capture commences. In addition to this, values are written at regular intervals to permit the identification of errors that could occur within that stream of data. A digital fingerprint (hash) is also generated to give assurance at a later date that the contents are the same as when the data was first captured.

These methods tend to attempt to capture the whole of the evidence into a single one-size fits all bag 'entity'. If the item being captured is too large to fit into one file, as it often is with the capacity of modern hard disk drives, then the file is fragmented into 'chunks'. This is to allow the file to be backed up later, or to be split between multiple smaller media if insufficient capacity single storage devices are not available at the time of capture.

However, in order to be able to process the contents of either of these types of data capture output, the totality of the fragmented files usually has to be made available again to the single application that is going to be used to process that evidence.

The above scenario gets even more complicated if data is being captured in real-time, for example as a network packet capture. This type of application is similar in principle to the 'dd' capture process with the difference that the amount of data to be captured is unknown when the process is commenced.

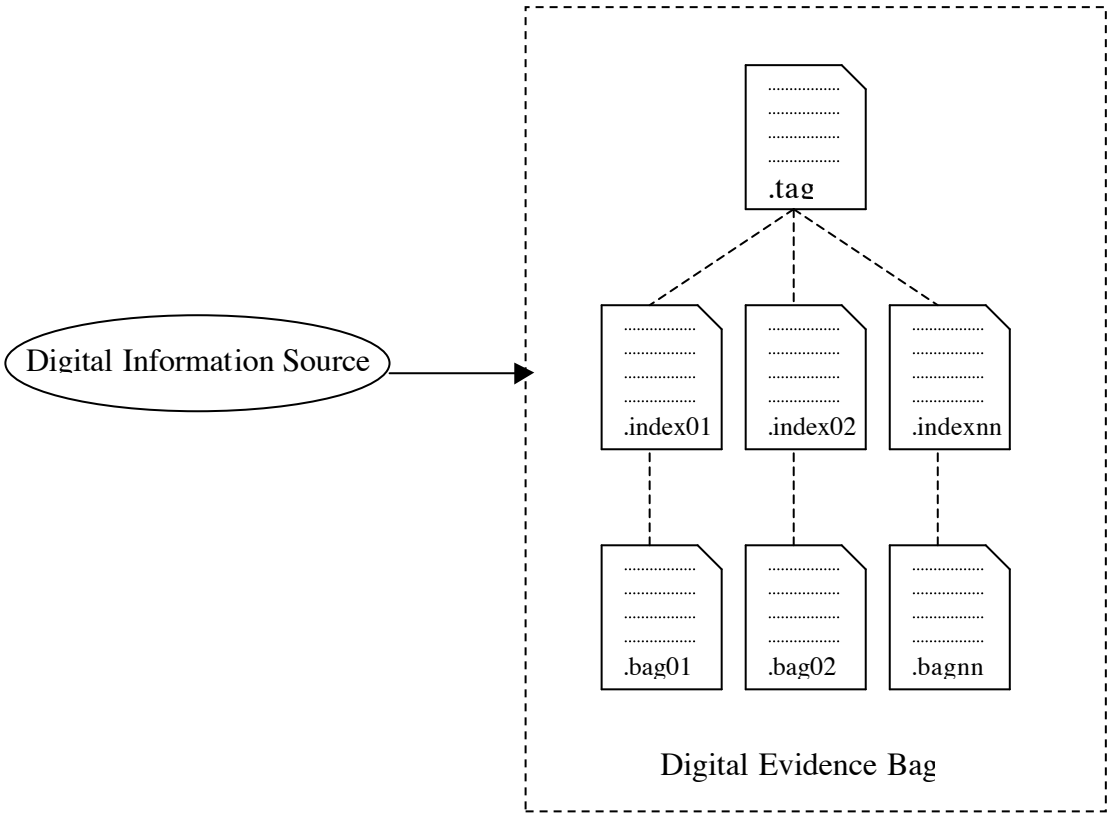
5. Digital Evidence Bags (DEBs)

5.1 A New Approach

What is required to help solve these problems is a new approach. That is not to say that we should immediately throw away all of our current tools and use something else, but the current tools and techniques should be adapted to work in a more flexible way. This may eventually lead to a new way of capturing and processing digital information in a forensically sound manner.

To help solve some of these problems the concept of a Digital Evidence Bag (DEB) is demonstrated. A digital evidence bag is a wrapper for any type of digital based evidence or information. The bag has potentially an infinite capacity (although in practice the size will be limited) and depending upon the user requirements can store information that could be captured in both a static or real-time environment. Furthermore each bag contains its own tag information, complete with integrity assurance information and continuity sections.

5.2 Basic DEB structure



The following lists the DEB files that are created as part of the capture process, hence for a single evidence capture three types of files are created:-

- .tag file;
- .indexnn file;
- .bagnn file.

The tag file is a plain text file containing the following information:

- DEB reference identifier;
- details of the evidence contained in the DEB;
- the name and organisation of person capturing the information;
- the date and time the capture process started;
- a list of Evidence Units (EU's) contained in the DEB. An EU is the name given to an .indexnn file and its corresponding .bagnn file;
- a hash of the captured information contained in the DEB;
- tag seal number comprised of a hash of the tag file to date, this is equivalent to the traditional seal number;
- Tag Continuity Blocks (TCB's) containing continuity information of when any DEB application accesses the DEB;

- the format definition of the .index file.

DEB applications update the tag file with a Tag Continuity Block (TCB) so that its contents reflect the history of operations performed on the bagnn files. Such information includes the date and time the application was used against the bagnn, include an application signature so that it is known what category of application and what version of application was used. The DEB application also updates the tag seal number.

The index file is a text based tab delimited file detailing the contents of the corresponding bagnn file. The index may contain details such as a list of filenames, folder paths, and timestamp information relating to the contents of the digital information in the bag. Alternatively it may contain details of the physical device, for example the make, model and serial number of the device captured. The exact contents and format of this file vary depending upon the content type of the EU. Its format is specified in the format definition in the tag file and is comprised of a series of Meta Tag labels that define its contents.

The bag file is the file containing the actual evidence captured. The contents of this bag file may be either raw binary information (e.g. from raw device capture), files (e.g from logical volume acquisition), structured text (e.g. from network packet capture) or categorised files (e.g. one bag containing all txt files, another containing all MS word docs, another containing all JPEG files etc.).

5.4 DEB Evidence Capture Scenarios

Traditionally digital forensic evidence capture is a static single process that is used to acquire the evidence material from beginning to end (dumb full capture) of the media concerned. With the implementation of a DEB structure a number of alternative scenarios become possible.

The traditional static full evidence capture (dumb full capture) is still catered for by implementing an index file with the following columns:

```
DeviceDesc. Manuf. Model Serial# Capacity Hash
```

In addition to the static full evidence capture an intelligent approach can be adopted. This would be implemented using an index file with the following columns:

```
Filename Extn. Attrib. TimeMod TimeAcc TimeCre LogSz PhySz Provenance Hash
```

This allows every file found in the original evidence to have an entry in the index file. A further enhancement of this is to create an EU containing only files of one particular type. This allows a more streamlined evidence analysis process to be undertaken.

Another scenario that DEBs bring to the evidence capture process is that of selective evidence capture. This involves only capturing files of a specific type or with a particular filename. This type of approach permits a forensic triage process to be performed and yet still allows the output to be captured in a forensically sound manner. This would be implemented using an index of the same form as that used for an intelligent mode full evidence capture.

One of the most important additions to the digital forensic capture process that DEBs permit is that of real-time evidence capture. DEBs allow the output of real-time commands and processes to be captured in a format that is compatible with that used in a static environment. In addition to this it can be performed with all the integrity checking mechanisms normally only associated with evidence capture in a static environment.

5.5 DEB Prototype Implementations

A number of prototype implementations of the DEB scenarios have been produced to demonstrate the concept and assist in the validation of this approach to evidence capture.
These applications have been written in Delphi and operate on the Microsoft Windows platform. They demonstrate the mechanics of creating and using Digital Evidence Bags. The applications written to date include:

Application Wrapper – a windows command line application wrapper. It accepts windows command line input and then processes and captures the command entered and the output generated from that command and creates a DEB. Multiple commands can be entered and all the output is contained in the DEB.

Digital Evidence Bag Viewer – allows the contents of a DEB to be viewed by the user. It demonstrates all the mechanisms required of DEB analysis applications and how they should interact with the DEB and update the tag file accordingly.

Selective Imager – allows the logical file structure of a disk to be viewed and the selected files to be captured to a DEB.

5.6 DEB Experience

The implementations created to date have allowed the concepts presented in this paper to be implemented, albeit in a prototype environment. They have proven that the concepts of being able to create a common structure that is capable of containing evidence sourced from both static and real-time environments can be achieved.

In addition to this it is possible to create applications that can maintain continuity of information in a similar fashion to that used in the physical world of evidence preservation outside of the digital environment.

5.7 DEB Future Work

Having created DEBs from various sources, the next phase of the work will be to create more advanced DEB processing and analysis applications. These will not only be able to process the information contained in a DEB but allow distribution of the Evidence Units in a multiprocessor distributed environment. This could be achieved in a similar way to that already implemented [3].

The capture process can also be refined further by integration of more intelligence. This can be done by directing the capture process with the aims of the investigation in mind at the outset of the case.

5.8 DEBs – Enhancing Best Practice

There is no manual on how to be an investigator, and passing on the knowledge of best practice is very difficult. In the digital world just keeping up with the current technological advances is exceeding difficult and time consuming.

The DEB approach of recording what applications and processes have been carried out on the evidence, and the sequence those analysis tasks are performed in an automated process, allows for the ability to learn the most effective order to undertake tasks. It also allows us to more quickly identify the shortcomings of the current application sets and recognize what additional applications would be most useful in the investigators toolbox.

Furthermore, this type of mechanism could also assist in the testing and certification of investigators, as it would permit trainees to undertake test cases and automatically record how investigators tackled them.

In addition, this type of approach allows us to bridge the gulf between the digital forensic practitioner and academia by giving the digital forensic world a metric that can be used to show how, where and when evidence was found.

7. Conclusions

The digital forensic community is in need of a new approach to the way in which the information from digital devices is gathered and processed. The Digital Evidence Bag concept demonstrated in this paper meets both the current and future needs of that community, and is capable of handling the large volumes of data associated with such enquiries.

The following advantages are immediately evident from such an implementation:

- Scaleable approach to evidence acquisition;
- Scaleable approach to forensic processing, allowing for the first time the digital evidence to be processed across multi-processor and distributed systems;
- Increased evidential material throughput, directing the most applicable techniques at the appropriate types of evidence;
- Incorporate some of the current evidence capture and analysis methods thus not negating the financial investment in current tools and methods;
- The ability to process evidence from a diverse range of digital devices;
- Allow the integration of real-time data acquisition into a sound forensic framework;
- Permit a selective and/or intelligent data acquisition approach to be implemented as opposed to the current collect everything approach;
- The ability to automatically create an audit trail of processes carried out on an item of evidence. The metric from which could allow analysis of the most effective way to process digital evidence and be used to educate new practitioners in the best way to undertake a forensic investigation.

References

[1] The Linux Kernel and the Forensic Acquisition of Hard Disks with an Odd Number of Sectors – Jesse D. Kornblum, International Journal of

Digital Evidence, Fall 2004, Volume 3, Issue 2.

[2] Encase® Legal Journal – Guidance Software Inc. http://www.encase.com/corporate/whitepapers/downloads/LegalJournal.pdf.

[3] Breaking the Performance Wall: The Case for Distributed Digital Forensics, Vassil Roussev, Golden G. Richard III, Department of Computer Science, University of New Orleans, LA70148. {vassil, golden}@cs.uno.edu