Forming a Relationship between Artefacts identified in thumbnail caches and the remaining data on a storage device

Sarah Morris[1], Howard Chivers[2]

Centre for Forensic Computing, Cranfield University
Shrivenham, SN6 8LA

[1]S.L.Morris@Cranfield.ac.uk

[2]H.Chivers@Cranfield.ac.uk

Abstract

The primary function of a thumbnail cache is similar between different operating systems; however there is no consistent implementation; because the thumbnails are potentially interesting to forensic analysts it is important to understand the detail of how they are used in a particular operating system. Previous work has shown the importance of understanding the structure and the effect of user behaviour on various thumbnail caches. However, an analyst needs to demonstrate a relationship between artefacts identified in the thumbnail cache and those found elsewhere on the system in order to provide context and corroboration of any evidence derived from thumbnails. A relationship between artefacts can also assist in establishing possible event time lines, and understanding the user behaviour which led to the system being in its current state.

This paper establishes the relationships which are formed between user generated files and information stored in the thumbnail cache; this shows how a forensic analyser can infer relationships between the thumbnail cache and other artefacts identified on the system. This paper provides a description of each relationship between the thumbnail cache and other artefacts; these relationships allow the corroboration of evidence extracted from the thumbnail cache and provide an addition source of evidence of user behaviour. In addition to providing a useful reference for analysts when reconstructing a user's activity, this paper also uses the thumbnail cache as an example to discuss the importance of contextual analysis within forensic computing.

In addition to the relationships shown between standard image thumbnail cache records and the rest of the system, this research also identifies how relationships are formed between the thumbnail cache and system artefacts such as the icons present on the user's desktop, and also allows the identification of devices on the same network as the user.

1. Introduction

The primary function of the operating system thumbnail cache is to store visual thumbnails relating to user generated files in order to save system resources from unnecessarily rendering of the images each time they are requested. An analyst therefore would try to establish a relationship between a user file and an entry in the thumbnail cache which may provide an insight into the user's behaviour.

Morris [2011] identified the structure and behaviour of the Windows 7 thumbnail cache; the work identified the potential context of the records and sub-records within the thumbnail cache. Whilst the identified context of the thumbnail cache can provide an indication of the user's activity and the corresponding historical state of a file, it is necessary for an analyst to form a relationship between this information and a user file

This rest of this paper is structured as follows: Section 2 identifies related work; the methodology employed during this research is described in Section 3. The ways in which a relationship may be formed between the thumbnail cache and a user file are shown in Section 4; Section 5 shows the potential relationships between the remainder of the system and the thumbnail cache. Finally in Section 6 the results of this research are discussed and the paper is concluded in Section 7.

2. Related Work

In forensic computing, admissibility is the term used to describe evidence which is allowed to be presented in court. Generally it is necessary to show a relationship between the evidence being produced and the events that actually occurred in the case. Whilst it may not be possible for an analyst to know exactly what happened, it is their job to produce a report. The analyst's report on the case provides assistance to the court as it is their analysis of the evidence. This can assist in determining the admissibility of evidence and the facts of the case [Kennedy, 2006]. Initially an analyst collects artefacts from a system and it is necessary to form relationships between the artefacts in order to provide context to the evidence. Contextual analysis assists the analyst in understanding the behaviour of both the user and the events which led to the current state of the system. One common method of performing a contextual analysis is by completing an event timeline.

Digital investigations use the same five key phases as those used in traditional physical investigations; preservation, surveying, documenting, searching, reconstruction [Carrier, 2004]. Reconstruction in digital investigations involves analysing the information recovered from a system and constructing events from it which can be ordered to show how the information recovered came to exist in its present state. An event can be defined as an incident that changes the state of one or more objects; in a digital investigation, an event can be thought of as an action that changes one or more of the bits in which the information is stored.

In a case, event reconstruction can assist in proving or disproving guilt by establishing the order in which events occurred and the times within them. Once events have been reconstructed, characteristics can be used to determine whether the actions were committed with intent [Carney, 2004]. For example, analysing times can give an indication of whether an action was the result of a piece of software or the user, since the user will have slower reaction times than the machine, there would need to be larger gaps between events than a machine would require.

The importance of contextual analysis can be seen in the case of Vosburgh [Find Law, 2010], one of the significant pieces of evidence recovered in the case were some indecent images of children found in a Windows XP thumbnail cache (thumbs.db). During the case the prosecution presented the argument that the only reason for the existence of the images to be present in the cache was if they had existed in a directory that the user had viewed. However the defence performed a live demonstration in court showing alternative methods for the evidence being found in the thumbnail cache. Based solely on the evidence within the thumbnail cache both the scenarios suggested by the defence and prosecution may have occurred; however if the rest of the system was examined to put the thumbnail cache evidence in context an analyst can draw up a likely set of user actions which assist in establishing a likely scenario.

3. Methodology

This research was conducted using virtual machines; a baseline image of Windows 7 was created using a standard Windows ultimate ISO downloaded from MSDN. During the installation default options were selected and a single user account was created. After installation of the operating system the base line virtual machine was cloned for each experiment. The experiments were designed to mimic typical user behaviour to identify information which changed on the remainder on the system that could be related to changes within the thumbnail cache. During the experiments potential thumbnail cache source files were added, modified and deleted both on the main and external storage devices. The resulting changes to each virtual machine were analysed using a variety of tools, including Encase and WinHex.

4. Forming Relationships between the Thumbnail Cache and a User File

Establishing a relationship between a file created by a user and the thumbnail cache assists in both corroborating the information identified and in forming an overview of the user's interaction with a file. This section describes the ways in which a relationship can be formed between a thumbnail cache record and a user generated file.

4.1 Visual Thumbnail

A visual thumbnail provides a snapshot of at least part of a file at a point in the file's history; if the visual thumbnail relates to the current state of the file then a visual comparison could be performed to identify if there is any similarity in its content to a user file. If the visual thumbnail and a user file appear to contain the same data then there may be a relationship between the two sets of information; an analyst cannot be certain the two sets of data are related. For example, there may be multiple copies of a file on the user's system, in which case all the copies may visually match with the same sub-record however only one copy would have a relationship with it.

The visual thumbnail may be compared to part of a user file using computer vision techniques; if the two images are the same size then a one-to-one mapping could be applied [Dufournaud, 2000]. However given it is likely that the two images are different sizes and may also be in different image formats the problem involves a one-to-many mapping; the increased requirement for the calculation of mappings may mean the technique is too resource intensive for large data sets and would therefore require the data set to be pre-processed. Like the visual inspection by an analyst a matching algorithm can only narrow down the potentially related sub-records.

The visual thumbnail may not represent the last state of the file, in Windows 7 each record in the thumbnail cache may contain up to four different sized visual thumbnails [Morris, 2011]; it is possible that the visual thumbnails may represent different states of the file. The states represented by the visual thumbnail may not match the last state of a file which would prevent a visual match of the data. It would therefore be necessary to examine each visual thumbnail for a record to maximise the chance of identifying a visual match between the data in the record and the file; it may also assist in identifying previous states of the file which may provide useful information for an analyst.

4.2 Metadata

The removal of time stamps from the thumbnail cache in Windows 7 makes it difficult for an analyst to establish the time a record was last updated within the

thumbnail cache [Morris, 2011]. However it is possible to identify a time period during which a record was added to the thumbnail cache; previous versions of the thumbnail cache files are stored in Windows 7 by the Volume Shadow Service. By right clicking on the directory 'Explorer' where the thumbnail cache is located it is possible to identify previous versions of the directory. Selecting each previous version will show the thumbnail cache files in an earlier state; by comparing the information within the versions of the thumbnail cache it is possible to identify a basic order in which records were added.

Figure 1 shows the windows restore points for an Explorer directory. Since each previous version has a date attached it is possible to work out a time frame during which a record was added. It is also possible to identify the order user created files were added to the thumbnail cache by identifying their position in the individual thumbnail cache files. New sub-records are appended to the end of the file; sub-records relating to user generated content do not alter their relative position; however sub-records relating to other types of data tend to move closer to the top of the file. As a consequence of the movement of some sub-records may mean that the offset to user generated sub-records may alter slightly.

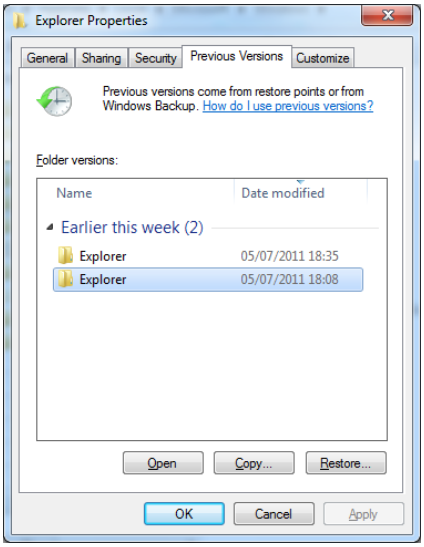

Figure 1: Restore points for the directory Explorer

The type of a file can be identified through the flags stored in the thumbnail cache; each type of file has a specific sub-set of flags; these flags are found in the main record for each file located in the thumbcache_idx file [Morris, 2011]. The flags can be used to identify the type of file which relates to the record, as the file type may not be obvious from a visual inspection.

4.3 Thumbnail cache ID

Each record contains a unique thumbnail cache ID which is made up of sixteen hexadecimal characters. Testing has shown this ID is comprised of data relating to the user file; it has been identified that the thumbnail cache ID is related to the volume, MFT data and the type of the file. In Section 5 the thumbnail cache ID is used to form a relationship through the Windows Desktop Search database.

4.4 Conclusion

This section has identified ways in which a relationship can directly be made between information in the thumbnail cache and a user generated file; whilst matching the visual thumbnail to a file can suggest a relationship it is only possible if the thumbnail represents the last state of a file. A match between a visual thumbnail and a file only provides an indication that the two data sets are related as there may be multiple files which match the thumbnail; however with corroboration from other sources it may be possible to strengthen the relationship. The removal of time stamps from the thumbnail cache makes it difficult for an analyst to provide an accurate time for when information is added to the thumbnail cache; however an event timeline can be constructed using the restore points for the thumbnail cache directory. Finally the thumbnail cache ID can assist in showing the volume, state of the MFT and the type of the file a record relates to.

5. Forming Relationships between the Thumbnail Cache and the Remainder of the System

In Section 4 methods were identified for defining direct relationships between a user generated source file and the thumbnail cache; this section looks at other relationships which can be formed between the thumbnail cache and system artefacts.

5.1 Windows.edb

The database for Windows Desktop Search stores a wide range of information including comprehensive metadata about user generated files [Chivers, 2011]; therefore the database can be used to provide extra and corroborating information on a user file which may be of interest to an analyst. Interestingly, the Windows Desktop Search database contains a field which holds the thumbnail cache ID; this field can be used to form a relationship between a file indexed in the database and a record in the thumbnail cache. However, there may not be a record in the database relating to the file or thumbnail cache ID the analyst is looking for; Chivers proposes several approaches to recovering deleted records in the database, which may recover the record the analyst requires.

5.2 Registry

There are records stored in the Windows 7 thumbnail cache that do not contain visual thumbnails [Morris, 2011]; some of these records contain a GUID. The GUID generally relates to system icons found on the user's desktop; the relationship can be identified by searching for the GUID within the registry. Figure 2 shows the results of searching for a GUID which relates to the Recycle Bin; by selecting the default icon directory shown in Figure 3 it is possible to identify the location of the icons being used. In this example the icons are stored in imageres.dll in positions 54 and 55 depending on the state of the recycle bin. In order to identify the icon used on the desktop it is necessary to identify the contents of file containing the icon as the result may be different to the standard icon, for example if the user has a customised theme. As identified in Section 4 the use of restore points can identify when the icons appeared on the desktop; generally system icons are added when a user first logs into their user area, however it is possible to add icons. The sub-records containing GUID's generally move to the top of the thumbnail cache file and do not always contain a reference in the index file. System icons may also be removed by a user, however the reference to the icon will remain in the thumbnail cache which may assist in identifying the behaviour of the user.

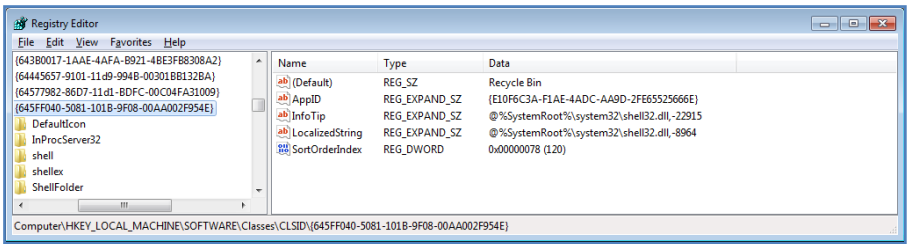

Figure 2: Searching for a GUID

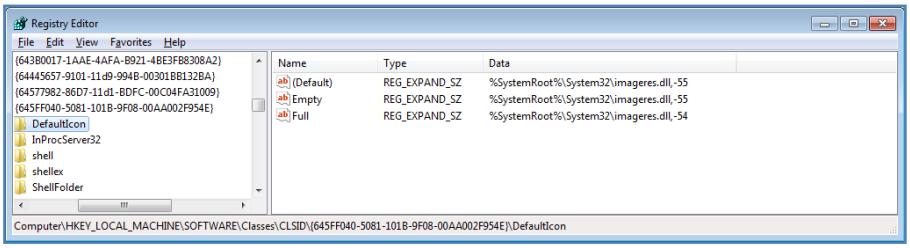

Figure 3: Identifying the Default Icon

Figure 4: An example contents from an imageres.dll

5.3 Shortcuts

Shortcuts on the system can also have visual thumbnails; these can be system, application or user created icons. Figure 5 shows an example of a shortcut Icon for the directory 'Forensic Software'; the icon is not stored in the thumbnail cache, however a reference to the icon is. Testing has shown that the reference is a sub-record without an image, the name field stores a 30-32 hexadecimal character string in Unicode. For software shortcuts, there is generally a registry key which stores the default icon to be used, which like the desktop icons could be extracted by the analyst.

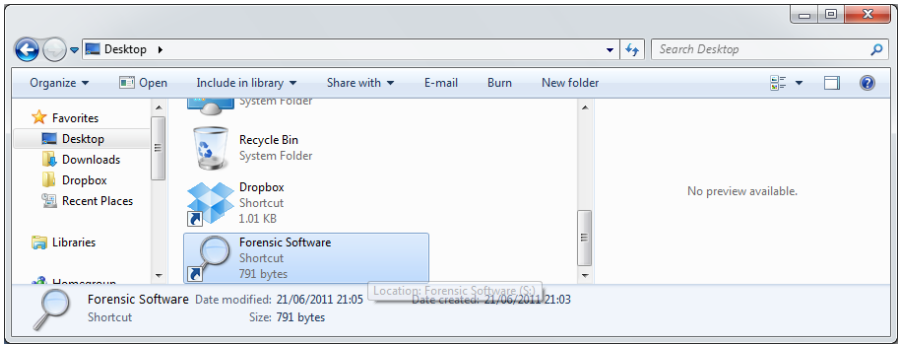

Figure 5: A Shortcut Icon in Windows Explorer

5.4 External Media

As discussed in Section 4.3 part of the algorithm for the generation of the thumbnail cache ID uses the volume ID; therefore generating an ID for a file could assist in showing the volume the file resided on. However, it would be time consuming to attempt to reverse engineer the ID given the number of variables

involved in its creation. Visual thumbnails may be stored in the centralised cache for any potential source file which is stored on an indexed volume; during testing external media were automatically indexed when connected to the system.

5.5 Encrypted Containers

Encrypted containers such as those used in TrueCrypt [2011] can be mounted as a volume by a user to enable the content to be accessed. Once the content is mounted it is possible to index the container the same as a normal volume. During testing it was found that the selection of records within a FAT32 TrueCrypt volume by a user results in the creation of corresponding records within the thumbnail cache.

5.6 Network Storage and Drive Allocation

There are sub-records stored in the thumbcache_256 file which store names in Unicode. Testing has shown these names to relate to network places and drive letters which have been allocated on the system. Figure 6 shows the drives allocated for the system; it is interesting to note that a list of drives is available in the central part of the explorer window as well as on the left hand side under the Computer icon. Both these areas of the explorer window will create records in the thumbnail cache; the drive letters stored in the thumbnail cache can be corroborated by checking the mounted devices stored in the registry.

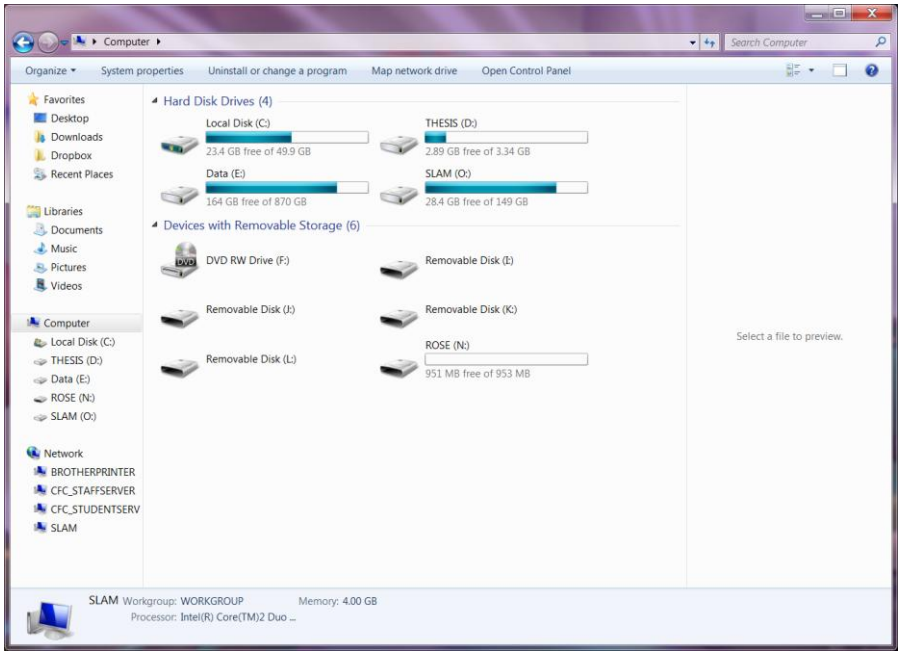

Figure 6: Drives currently available to the user

In Figure 7, a list of networked places can be seen; again the names for these places can be seen in both the main part of the explorer window and in the left hand side frame. Each networked place is stored in Unicode in the thumbnail cache; an analyst could identify the name of the machine being examined in the registry; the remaining networked device names may assist the examiner by providing a list of devices which may be of interest.

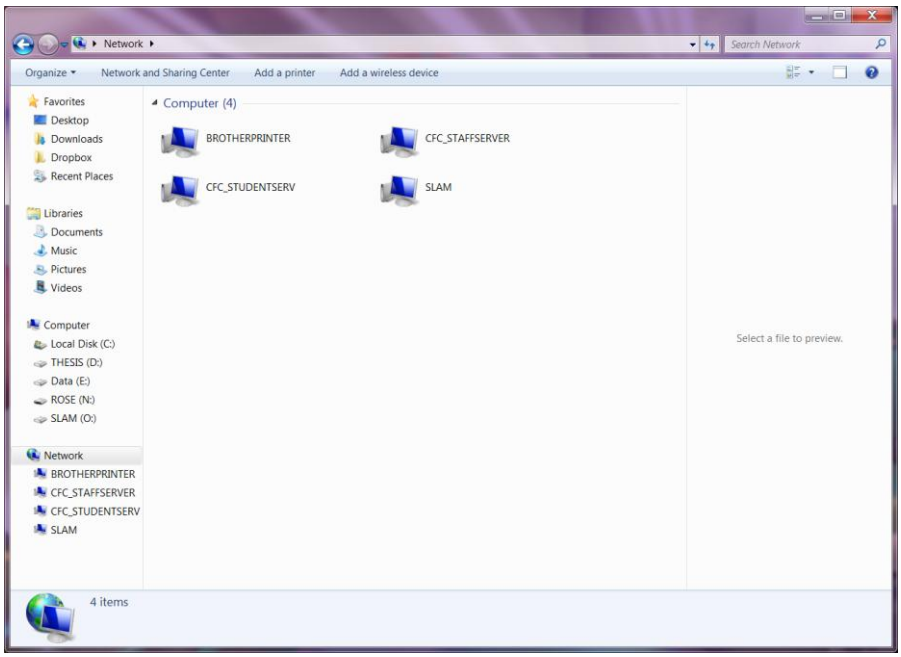

Figure 7: Networked places available to the user

5.7 Conclusion

This section has described the relationships between the thumbnail cache and other potential system artefacts. The Windows Desktop Search database can assist an analyst in forming a relationship between a file and an entry in the thumbnail cache as for user created files it stores a field for the thumbnail cache ID value. It is possible to carve out deleted records from the database, but the record an analyst requires may not be found. The GUID's stored in the thumbnail cache can be found by searching the registry; this will enable the type and default icon to be identified. Information on the icons used for shortcuts is also stored in the thumbnail cache; default icons for shortcuts may also be available in the registry. Thumbnail records for external storage media will only appear in the centralised thumbnail cache if the volume has been indexed by the system; if the volume has been indexed it behaves in the same way as internal volumes. Encrypted NTFS containers behave as a normal volume when mounted; therefore if they are indexed by the system records are created in the standard way. If the volume mounted is

FAT records tend to only be created for selected objects. Finally the names of network devices and allocated drive letters are available in the thumbnail cache, the name of the system being examined can be identified from the registry; the other device names stored may be further potential sources of evidence. The drive letters can be corroborated using the mounted devices key in the registry.

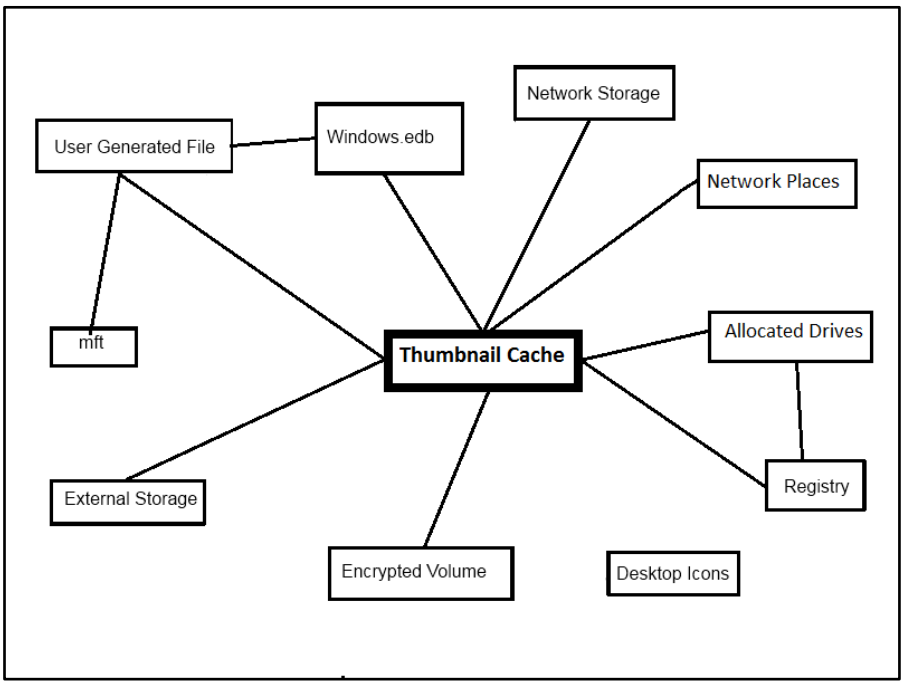

Figure 8: A Diagram of the relationships between the thumbnail cache and other system artefacts

6 Discussion

The relationships between the thumbnail cache and other artefacts on the system have been discussed in this paper; Figure 8 shows the relationships diagrammatically. Whilst the relationships vary in strength this paper has shown a variety of artefacts which can be combined to form corroborating evidence and assist in identifying user behaviour and understanding the context of the evidence. This research also shows relationships with the thumbnail cache can assist in showing the structure of the system through the drive letters allocated; the structure of the network can also be identified through the network places stored in the thumbnail cache records.

Understanding the context of these artefacts assists the analyst in establishing how the system came to be in its current state and what the evidence means. It is necessary to establish the user behaviour and this is commonly done through the creation of event timelines; the removal of times from the thumbnail cache in Windows 7 makes it challenging for an analyst to add the evidence from the cache to a timeline. However the use of system and file restore points allow an examiner to narrow down the time in which evidence was added to the thumbnail cache; thereby creating a basic order of when events occurred. Interestingly a greater understanding of the artefacts and user behaviour develops when the relationships are combined; the collection of evidence produces a greater insight into the system as a whole.

7. Conclusion

This paper has shown that relationships can be formed both directly and indirectly from the thumbnail cache to user generated files and other artefacts on the system. These relationships assist in providing corroboration of the evidence and provide further context of the artefacts identified. Contextual analysis of artefacts identified in a case can assist with understanding the user and system behaviour which led to the current state of the evidence. Encouraging examiners to view evidence contextually would lead to a greater understanding of the artefacts being analysed which allows a more complete portrayal of the evidence to the court to assist in ensuring justice.

References

Carney, M.; Rogers, M., *The Trojan made me do it: a first step in statistical based computer forensics event reconstruction*, International Journal of Digital Evidence Vol. 2 Issue 4. 2004.

Carrier, B. *File System Forensic Analysis*. Addison-Wesley, PA., March 2005.

Chivers, H.; Hargreaves, C., *Forensic data recovery from the Windows Search Database, Digital Investigation*, Vol.7 Issues 3-4. 2011

Dufournaud, Y., Schmid, C., and Horaud, R. 2000. Matching images with different resolutions. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, Hilton Head Island, South Carolina, USA

Find Law, "No. 08-4702. – UNITED STATES v. VOSBURGH – US 3rd Circuit". 2010. http://caselaw.findlaw.com/us-3rd-circuit/1522221.html

Kennedy, I. "Presenting digital evidence to court". 2006. http://www.bcs.org/content/ConWebDoc/7372

Morris, S.; Chivers, H. (2011) "An analysis of the structure and behaviour of the Windows 7 operating system thumbnail cache". *Proceedings from 1st Cyberforensics Conference.* University of Strathclyde, Glasgow, UK.

TrueCrypt, "TrueCrypt – Free Open-Source On-The-Fly Disk Encryption Software for Windows 7/Vista/XP, Mac OS". 2011. http://www.truecrypt.org