Proceedings from 1st International Conference on Cybercrime, Security and Digital Forensics,
27-28 June 2011, University of Strathclyde, Glasgow, Scotland, UK.

An analysis of the structure and behaviour of the Windows 7 operating system thumbnail cache

Sarah Morris[1], Howard Chivers[2]

Centre for Forensic Computing and Security

Cranfield University, Shrivenham, SN6 8LA

S.L.Morris@cranfield.ac.uk

Abstract. Operating systems such as Windows 7 implement a thumbnail cache structure to store visual thumbnails and associated metadata. There is no standard implementation of a thumbnail cache or its functions, which has led developers to implement their own structures and behaviour. The artefacts present within a thumbnail cache are of interest to a forensic analyst as they can provide information on files within the system which may be of use to the investigation. This research investigates the structure and behaviour of the thumbnail cache implemented in Windows 7 and shows that as well as storing information relating to visual thumbnails the cache also stores the names of networked computers, GUIDs relating to system artefacts and allocated drive letter information. It also shows that due to the behaviour of the cache, information such as records relating to files which are no longer on the system may be available, proving interesting forensic evidence.

Keywords: thumbnail cache, windows 7, forensic computing,

1 Introduction

Microsoft operating systems are installed on a significant proportion of user machines. Due to the likelihood of encountering a Windows based operating system there is significant interest in finding and understanding relevant artefacts within the forensic community. Therefore experiments investigating changes to the thumbnail cache structure and behaviour are interesting and relevant.

This paper shows the structure and behaviour of the operating system thumbnail cache used in Windows 7; understanding the structure can assist in retrieving meaningful artefacts, whilst understanding the behaviour allows the artefacts to be placed in context. This paper begins with an overview of related work and the methodology employed during this research; followed by a description of the structures implemented within the thumbnail cache. Finally the effect user behaviour has on the thumbnail cache is investigated.

2 Related Work

In Windows XP, thumbnail caches occurred throughout the system, located in every directory which contains potential visual thumbnail source files and has been viewed in thumbnail mode by a user. These thumbnail caches were stored in hidden files named ‘thumbs.db’ and contain interesting forensic artefacts such as the file name of the original file [1].

The release of Windows Vista prompted speculation from Forensic analysts about the impact of the new Operating system on investigations [2]. One interesting change was the move away from the ‘thumbs.db’ files, which were present in individual directories, to each user having a single centralised thumbnail cache. The release of Windows 7 therefore prompts the question of identifying further changes; in fact, as shown below, these changes are limited.

Whilst the function of a thumbnail cache has been defined, the structure varies considerably between operating systems. This means that each thumbnail cache structure will have a different forensic value and the data stored within it may not represent the same series of events [3].

Users may not be aware of the thumbnail cache; the media openly discussed the potential privacy violations of its existence in the case of Vosburgh [4, 5]. During this case the defence put forward the argument that thumbs.db were hidden files, created without the defendant’s knowledge or consent. The prosecution argued that their existence supported the charges of possessing indecent images of children. This highlights the importance of understanding how such files are created and their relationship to any evidence found, in order to show what actions were taken by the computer’s user.

3 Methodology

In order to evaluate the structure and behaviour of the thumbnail cache used in Windows 7 a virtual machine was created with a 30GB hard drive. Windows 7 Ultimate was installed, using an ISO downloaded from the MSDN website. During the installation, the default settings were selected in order to mimic a typical user system.

Once the installation was complete the baseline machine was cloned and examined to determine the initial state of the thumbnail cache and related files. A clone of the baseline was created for each set of experiments to assist in determining the structure and behaviour of the thumbnail cache. Any experiment specific methodology is described in the appropriate sections in further detail.

4 Identifying the Structure

In order to extract meaningful information from data it is necessary to understand its context; in the Windows 7 thumbnail cache the structure provides a context for the data. This section describes the structure of the various components of the Windows 7 operating system thumbnail cache. It begins with an overview of the thumbnail cache directory structure; this is followed by an overview of the structures contained within the six files which make up the thumbnail cache.

4.1 The Thumbnail Cache Directory Structure

Whilst examining the baseline image, it was determined that each user has a single centralised thumbnail cache which is located at the path below.

[Drive]:/Users/[User_Name]/AppData/Local/Microsoft/Windows/Explorer

This directory contains six thumbnail cache files: an index and five further files. The index holds records, which provide pointers to the location of sub-records in the remaining files. The rest of this section identifies the structures present within the six thumbnail cache files and identifies the records and sub-records.

4.2 Thumbcache_idx.db

The file 'thumbcache_idx' provides an index to the information stored within the thumbnail cache; each source file is represented by a single record in the index. A record contains pointers to the locations of the associated sub-records. The file has a standardised structure for records with each one using 32 bytes. Each record begins with an eight byte unique identification string known as the thumbcache ID. The thumbcache ID can also be found in the record relating to the original source file within the windows desktop search file, Windows.edb. The flags identify the file type. The remaining twenty bytes store pointers to the sub-records for the original source file; each pointer identifies the starting position in bytes from the beginning of the relevant file.

The thumbcache_idx structure is different to the one used in Windows Vista, which used forty byte records. The last modification time of the source file has been removed, which accounts for the eight byte difference in record size.

4.3 Thumbcache_32, 96, 256 and 1024

The centralised thumbnail cache system implemented in Windows 7 allows for a maximum of four differently sized images to be stored for each source file. The images can be up to a maximum of 32x32, 96x96, 256x256 and 1024x1024 pixels respectively. Each of the four sizes is contained in its own file into which sub-records are stored; the structure of the sub-records can be divided into two separate sections: the header and the image. The sub-record header section contains the metadata for the sub-record and contains a clearly defined structure; the image section follows the standard structure for the image file type implemented, this is generally JPEG, BMP or PNG.

The 32 byte sub-record header structure has decreased by eight bytes from the 40 byte implementation used in Windows Vista; this is due to the removal of 8 bytes which represented the file extension in Unicode. The checksum values provide a method of authenticating the values in each section of the sub-record.

Whilst the thumbnail cache predominately stores visual thumbnail images, the thumbcache_256 also holds non-standard sub-records. Drive letters are stored for available drives, including portable storage devices; the network locations also include the name of the machine on which the thumbnail cache has been created.

4.4 Thumbcache_sr

Whilst this file is present in both the Windows Vista and 7 implementations of the operating system thumbnail cache, experimentation has yet to identify its purpose. The only recorded data for this file is contained in the standard file header.

4.5 Conclusion

This section has described the structure of the record and sub-record implementations used in the Windows 7 operating system thumbnail cache. The cache consists of six files: an index and five files to hold sub-record data. The records stored in the index store pointers to the sub-record data for each source file. The visual thumbnail sub-records consist of two sections, a header and an image. The following section will show the context in which the artefact structures identified in this section occur on an operating system.

5 Identifying the Behaviour

Three types of action can be observed on the information stored within the thumbnail cache: information can be added to files within the thumbnail cache, it can be modified or it can be deleted. This section identifies the behaviour the Windows 7 thumbnail cache exhibits when a user interacts with the operating system. For the experiments in this section both Microsoft Office and Adobe PDF viewer were installed on the baseline machine; the default settings were selected during their installation. The software was installed to allow common user file formats to be evaluated and manipulated during the experiments.

5.1 The Creation of Records and Sub-records

When a user interacts with a system, there are several ways in which they can create a new file: they can use an application, they can move a file from a storage device, or they can download a file from the internet. They can also view the files using a variety of methods such as in an application or in a file browser. This sub-section identifies the behaviour which results in records and sub-records being added to the thumbnail cache.

5.1.1 Using Different Methods for Adding Files

When a file is created in an application such as Microsoft Word it can be saved to a storage location, which permits the created file to be accessed again at a later date. As part of this research, several files were created using this method; it was noted that a record was not added to the thumbnail cache for these files until they were viewed in thumbnail mode. When a file is viewed in thumbnail mode in file explorer, or in a dialogue Window, like those used for opening and saving documents, a thumbnail record is created or modified as required for each potential thumbnail cache source file. In figure 5.1, several files are being viewed using the standard file browser. Each potential user created thumbnail cache source file will have records added into the thumbcache_256.db and thumbcache_96.db file; whilst system created images such as those used for the directory ‘Not Viewed’ will only be added to the thumbcache_96.db. If the size of the icons used is increased from medium to large icons, the resulting additions to the thumbnail cache would not change. However if the extra large icons are selected, the directory visual thumbnail would also be added to the thumbcache_256.db file. Sub-records are added to the thumbcache_1024.db file when they are previewed in the preview pane of the file browser; however this only occurs if the image being previewed is greater than the maximum size image permitted in the thumbcache_256.db file. Experimentation has found that image files, PDF files and directories viewed in thumbnail mode consistently add records to the thumbnail cache; however Microsoft Word files do not always create thumbnails.

When files are viewed on removable storage devices in thumbnail mode, entries are added to the thumbnail cache, experiments have shown that when a file is moved from a storage device to a hard disk drive a new record is added to the system when the file is subsequently viewed in its new location. This shows that thumbnail cache entries for removable storage devices would be present in the thumbnail cache and may provide corroboration that a device has an relationship with a particular machine or user. Network storage viewed in thumbnail cache mode does not create entries in the thumbnail cache until a file is moved to a local drive or device.

Any sub-records added to a thumbnail cache file are appended to the end of that file. However experiments have shown that the index file is a hash table, where entries are assigned a place in the table based on a calculation using the eight byte thumbnail cache ID number.

The index allocates one hundred and one spaces in the thumbnail cache when it is initialised. Experiments have shown that for index sizes fewer than five hundred and five records, each time a clash occurs in the hash table, another one hundred and one spaces are added to the table. For larger index sizes the number of new records added each time increases but is always a multiple of one hundred and one. All the existing records are then recalculated to account for the new table size.

Records added to the thumbnail cache may contain associated data in the Windows.edb file; the record within the windows.edb contains a thumbcache ID which allows a relationship between a source file and a thumbnail cache record to be shown. Figure 5.2 shows a file ‘Cake.JPG’, and shows the structure of the relationship between the files.

It is important to note that records within the thumbnail cache may not be directly associated with a visual thumbnail image; these records may instead hold information in Unicode in the identifier section of the sub-record header. Experiments have shown these records to contain GUIDs relating to entries in the Windows registry, the letters associated with allocated drives and networked places. The drive letter and network place records are generated when the file browser is used. As can be seen in figure 5.1, the right side of the file browser has icons for drives, folder locations and network places. As these items use default icons, only details are stored in the thumbnail cache. New records are added when locations are added to the right side panel, either by the system or a user and are visible in the file browser; sub-records relating to locations which are no longer present are not removed. This is a useful source of information for analysts as it provides a list of drives which have been allocated and machines which have been networked to the users machine; it also provides the name of the user's machine as this is shown in the network places list within the panel.

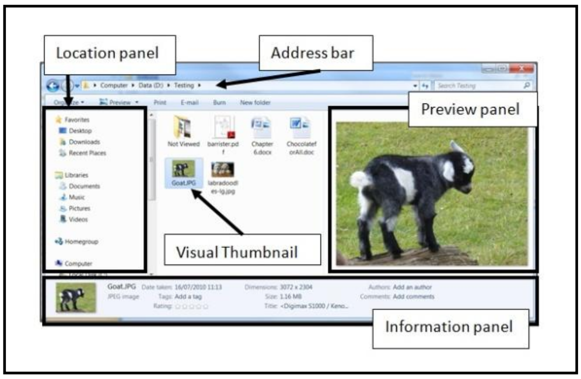

Figure 5.1: The creation of sub-records in thumbcache_1024.db

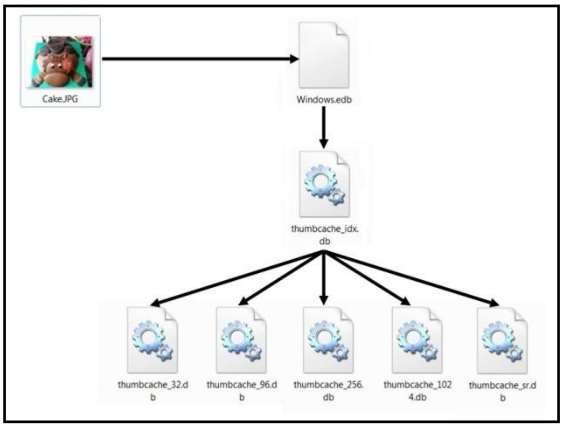

Figure 5.2: The relationship between a file and a record in the thumbnail cache

5.1.2 Circumstances where information is added to the thumbnail cache

A record is added to the thumbnail cache when no record currently exists for a given URI and the file at the URI is a potential thumbnail cache source file.

When a matching record cannot be found in the index a new record can be created under the following circumstance:

- A directory is viewed using the file browser, or the file browser within an application using thumbnail mode
- The directory being viewed contains file types which are compatible with the thumbnail cache
- A file exists within the directory which does not currently have a thumbnail cache record, or a sub-record of the appropriate size.

5.2 The Modification of Records and Sub-records

A user can modify a file in three distinct ways, by modifying the contents, the URI, or deleting the file. This sub-section discusses the experiments performed to ascertain the behaviour of the thumbnail cache when the file is modified; Section 5.3 discusses the deletion of files.

5.2.1 Modifying the Contents of a File

In order to ascertain the behaviour of the thumbnail cache when a user modified a file, several files were modified using their default applications. The files were all located in the file browser, and then selected to open, each file was then modified, saved and closed. The thumbnail cache was examined after the file was closed and no related changes to the thumbnail cache were observed.

The modified files were then viewed using a variety of methods, including through the open command in their default application and through the file browser. The results showed that when a file was viewed in thumbnail view, the record for the source file was updated to reflect the current state of the file. During the experiments it was noted that the thumbcache_1024.db image, where present was not consistently updated when the sub-record was viewed.

5.2.2 Modifying the Files URI

If a source file changes its corresponding URI, a new record is generated in the thumbnail cache and no alteration is made to the existing record. This effect was recorded during a series of experiments where a file was placed at a URI and viewed to ensure a record was added in the thumbnail cache. Once the record was stored in the thumbnail cache, the URI of the file was changed; this involved either altering the file’s name or the entire path. The file was then viewed at its new URI and the effect to the thumbnail cache was analysed. A new record was created in the thumbcache_idx, with sub-records also being created. Aside from the different thumbcache ID value, the record and sub-records were identical to the original test data. This suggests that it may be possible to create a false relationship between artefacts in the thumbnail cache and a source file.

In diagram 5.3 a directory is being viewed using Windows Explorer, the directory contains a sub-directory named ‘Not Viewed’; which is represented by a folder and a selection of thumbnails. Experimentation has shown that the thumbnails used to represent the contents of a directory are selected when the thumbnail is first generated. If the images used in the thumbnail are removed from the directory, the directory thumbnail is not always updated, which may lead to an inaccurate representation of the directory’s current contents. Only files which are used for the directories visual thumbnail record have sub-records created for them; therefore there may be records in the thumbnail cache for files which the user has not viewed. Since each user area has its own centralised thumbnail cache, the thumbnails used in the directory thumbnail image must represent source files which have been present in the directory at some point. It is also interesting to note that the standard visual thumbnail in the directory browser and the visual thumbnail in the information panel are not identical. This is because they are both being taken from different sub-records for that entry in the thumbnail cache and the sub-records were generated at different times and were invoked by different processes which were unaware of other sub-records for the thumbnail cache entry.

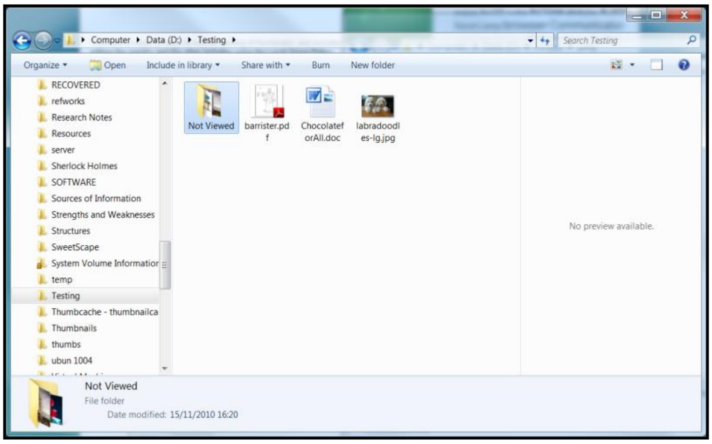

Figure 5.3: A screenshot from a test to see if records were created for files which had not been viewed

5.3 The deletion of records and sub-records

There are several options in Windows 7 which allow the removal of the thumbnail cache; for example the cache could be deleted by the user. However if the cache is simply deleted, the cache will reinitialise and start again, this may remove traces of a file which is no longer on the system from the current thumbnail cache. The disk clean up wizard provides an option to remove the thumbnail cache, this simply deletes the current thumbnail cache; a new cache would be created next time it was called.

The local group policy editor has an option to “turn off the caching of thumbnails in hidden thumbs.db files”, suggesting that Windows 7 can utilise both the directory specific and centralised forms of thumbnail cache. During the experiments conducted in this research, no evidence of a thumbs.db was recorded; however they have been identified on actual systems, suggesting further research into their behaviour is required.

Experiments showed that if a file is deleted from the system, it is not automatically removed from the thumbnail cache, instead the record remains; the cache also does not remove records which have not been accessed for some time. However it is possible to alter the default settings within Windows 7, to limit the size of the thumbnail cache, which experiments have shown causes the cache to remove inactive records when it reaches its maximum size. The cache has been shown to create a sub-directory within the explorer directory named ‘ThumbCacheToDelete’, this directory is where outdated versions of thumbnail cache files are placed prior to deletion. Moving the thumbnail cache files allows the system to regenerate new files; this is particularly used when the index requires extra space. During this research no method tested has led to the secure deletion of thumbnail cache files, therefore once deleted the files are available in unallocated space.

5.4 Conclusion

This section has described the behaviour of the thumbnail cache, it has highlighted the requirement for files to be viewed in thumbnail view for records to be created or modified. However there are situations, such as in the creation of visual thumbnails for directories, where a record may be created for a file which has not been viewed. It has shown the thumbnail cache contains records which do not have visual thumbnail images, these records provide information on which drive letters have been allocated and the names of networked machines. When a file is modified the thumbnail cache is only updated when the file is next viewed in thumbnail mode; however the old record is not deleted. There are various methods for deleting the thumbnail cache and preventing it from regenerating; however the deleted cache is not securely deleted. Finally the allocation of blocks of space to the thumbnail cache provides large fragments which generally contain several records or sub-records, which may assist with fragment classification.

6 Conclusion

This paper has examined the structure and behaviour of the operating system thumbnail cache implemented within Windows 7. Within Windows Vista, the thumbnail cache implementation changed significantly from Windows XP by using a centralised thumbnail cache for each user, this implementation was continued in Windows 7 with slight differences in the structure of records and sub-records. The thumbnail cache contains visual thumbnail records alongside GUID information, active drive letters and network places. The non-standard sub-records can provide assistance to an examiner by showing the structure of the device, associated devices and associated machines which may require analysis. Records are added to the cache, with a new sub-record being added to the bottom of the thumbcache files. Some records are deleted periodically, however this is application specific. The sub-records contain checksums to assist in authenticating the data; however it is still possible to tamper with parts of the thumbnail cache without it detecting an error.

References

1. Hurlbut, D., “Thumbs DB Files Forensic Issues”. AccessData Training Document. 2005. http://www.accessdata.com/media/en_US/print/papers/wp.Thumbs_DB_Files.en_us.pdf
2. Hargreaves, C., Chivers, H. and Titheridge, D. (2008), "Windows Vista and digital investigations", Digital Investigation, vol. 5, no. 1-2, pp. 34-48.
3. Morris, S. (2010) “A comparative study of the structure and behaviour of the operating system thumbnail caches used in Kubuntu and Ubuntu (9.10 and 10.04)”. Proceedings from 4th Cybercrime Forensics Education & Training. Canterbury Christ Church University, Canterbury, UK.
4. LiveLeak, “FBI agents posting bogus links and entrapping people”. 2009. http://www.liveleak.com/view?i=87d_1233031430
5. McCullagh, D., “FBI posts fake hyperlinks to snare child porn suspects”. 2008. http://news.cnet.com/8301-13578_3-9899151-38.html