

CONDUCTING FORENSIC ANALYSIS UNDER THE NEW RULES

by

Gregory L. Fordham, CPA, CIA, CCE, Sec+, MCP



WWW.KNFCON.COM

1303 Hightower Trail, Suite 315
Atlanta, GA 30350
800-335-1188

CONDUCTING FORENSIC ANALYSIS UNDER THE NEW RULES

by

Gregory L. Fordham, CPA, CIA, CCE, Sec+, MCP

There is more to electronic discovery than digitizing documents or searching digital documents. Indeed, there is a great wealth of information that can be simply and easily discerned. In addition to the examination of content to prove or disprove claims and defenses, digital evidence can also be used to determine the respondent's compliance with a production request. This determination extends to both the completeness of the production as well as any efforts to "hide the ball".

Everyone has come to understand that deleting data does not necessarily delete the data and that deleted data can often be recovered. What may not be as widely appreciated, however, is that utilizing a computer system is like throwing pebbles in a pond. While the pebbles may nor may not be recoverable, the ripples they leave are detectable nonetheless.

The following sections review forensic analyses that requesters can use on the data produced in order to validate that the respondent's production is complete and that the smoking gun has neither been discarded nor disguised.

Assessing Compliance

One of the first things that a requester will want to know is whether the produced data fully complies with the production request. More specifically, does it match the scope of the production request or is there some data source that has been omitted? Similarly, is there any indication that the produced data has been altered or destroyed? Finally, is there anything that the respondent has done to hide relevant evidence or what did they do with the evidence that is the focus of the case? The following sections discuss registry analysis, link file analysis and various forensic processes that can be used to answer these and other questions.

Registry Analysis

The registry is a hierarchical database used on all 32 bit Microsoft Windows systems. The database is contained in a small number of files that correspond to users, security and system. The contents of those files and the database itself, can be viewed independently of the computer.

CONDUCTING FORENSIC ANALYSIS

Thus, if this kind of information is important then it can be included as part of the production request.

The registry is used to store settings and parameters for system operation. These parameters can involve users, software and hardware. Although the volume of information that one can find in the registry is immense, there are a few areas that are almost always of interest. Those areas are:

- system install dates,
- software and hardware installations,
- mapped drives, and
- users and last logons.

One of the items recorded in the registry is the system install date. In other words, when Windows was installed to the computer system. When building a conceptual model of the respondent's computer systems and a time line for what computers were used by what individuals at what times, the system installation date can help to validate that time line or alert the practitioner to re-installed systems or even re-built systems.

In addition to the system installation dates, the registry also records all hardware that is attached to the machine. This would include devices like the hard drive itself (in case that has been removed and replaced), external hard drives (whose contents may not be included in the produced items), memory sticks (also known as thumb drives, flash drives, jump drives and whose contents also may not have been included in the produced items), network adapters, wireless cards and any other kind of device that could be used to communicate with other computers on the network or data storage devices.

Software installations are also captured in the registry. Even when software is uninstalled it is often possible to still find some remnant of a previously installed program. Consequently, reviewing this information can be useful in determining whether users currently have or previously installed software that could be the focus of the case or was installed to destroy discoverable evidence. In at least one case, the mere presence of a wiping program installed on a computer hard drive resulted in an adverse inference jury instruction.¹

Another attribute found in the registry is the existence of mapped drives. Mapped drives are pointers to storage devices that are not attached to the local machine. In other words, they are accessed over a network. Their mapping, or reference to them, would generally indicate a storage location that is used frequently enough by the user that a permanent connection has been recorded on the user's system. Once again, the existence of mapped drives is a signal to the producer and/or the requester that there is another potential storage location what should be evaluated for the existence of discoverable data.

CONDUCTING FORENSIC ANALYSIS

A final feature commonly of interest is the user's logon data. This information identifies the users that have logged on to the machine as well as their last successful logon dates along with their last unsuccessful logon dates. Not only does this data help to determine the users of a particular computer, the time stamps can be useful in evaluating the respondents compliance with a preservation order. After all, the continued operation of a computer can destroy potential evidence.

Log files are the record of computer activity that are left by a particular software program. There are a number of log files that can be found on a computer system. They can be log files created by individual software application programs that record certain aspects of the program's operation. There are also operating system log files that record various operations of interest to the operating system. There are audit logs that capture activities identified by administrators to be of interest.

Event logs are but one type of log that a person can find on a Windows computer system from Windows 2000 and forward. The event logs contain recording of selected activities that the operating system has chosen to record. These activities can be system events, application events and security events and there are separate logs for each of the three event types—system, application, and security.

Although the number of events that are potentially recordable by the operating system are large, they are still a finite number and do not include every conceivable event that occurs. For example, a user pressing a key on the keyboard is an event but it is not one that is recorded in the event logs. Nonetheless, there are thousands of events that are recorded in the event logs.

An examination of these events can reveal a number of activities that are potentially interesting to practitioners. In some cases they could include each attachment of a USB device, such as thumb drive or external storage device, to the machine. So, if the practitioner had identified that a thumb drive had been excluded from production and wanted to try and determine when it had been used and how often it had been used then the event logs are one source that the practitioner could consult in order to make this determination.

Another useful purpose of event logs is to evaluate system clock reliability. Part of the data included with each logged event is a sequential identifier and the system clock date and time. If system users have wildly altered their system clocks (for example turning the clock back to delete data or alter documents) this activity could be detected by looking for anomalies between the sequential record identifier and the event date and time stamp.

Link Files

Link (LNK) files have a file extension and file signature of LNK. LNK files are shortcuts pointing to the location of a file. LNK files are essentially what one is viewing when one opens Microsoft Word, for example, and looks at the list of most recently used files. LNK files are useful in a forensic investigation to determine what files might have existed on other media. In addition, they are an artifact of a previously deleted file.

When examining link files the data of interest is typically their internal metadata. The internal metadata contains the file name and full path of the file to which it points. In other words, if a document existed on an external storage device, such as a thumb drive, then the full path within the LNK file will indicate the logical device letter and the full path and name of the target file. Practitioners reviewing a respondent's production would want to request link files and review their contents as a means to validate the completeness of the production in terms of documents produced and storage devices considered.

In addition to the full path and file name, LNK files also contain the date and time stamps for the file that it references at the time it was last referenced. A practitioner would want to consider these date stamps when evaluating a produced document or reasons for why a deleted document was not produced. If an evil doer had been successful in wiping an important document and removing all traces from the filing system, the LNK file could still provide evidence that the file had existed at the preservation was performed.

The importance of analyzing LNK files has carry-over to hyperlinks embedded in documents. A study of these hyper-links could also reveal not only other data storage locations that may or may not have been considered for production but they can also identify documents themselves that should have been produced.

Wiping Programs

Wiping and shredding programs are applications designed to protect private information contained in computer files by overwriting the contents of those files with other data thereby making the overwritten data unrecoverable. Since people have come to realize that deleting a file does not delete the data, evil doers have turned to wiping and shredding programs to truly delete data.

The legitimate purpose of these programs is to protect private data and not to commit a crime. Consequently, wiping programs tend to leave traces of their use and existence even though they may leave no traces of the original data. These traces of their use and existence are numerous and range from entries in the registry, log files in free space, indicators in the file system such as place holders and constant date values, as well as patterns in the characters used to overwrite the wiped and shredded data.

Detecting Data Hiding Techniques

There are any number of reasons that data hiding techniques could be important to a case. One could be that the respondent has or previously had hidden incriminating evidence of its misdeed among its other data holdings. Another could be that the entire focus of the case involves data that has been taken improperly by using data hiding techniques. The following sections discuss some fo the forensic processes that can be used to detect data hiding techniques and uncover the “smoking gun”.

Deleted Files

Perhaps one of the most obvious techniques for hiding data is to delete the file. As many have come to learn, however, deleting the file does not erase the data. Rather, it simply deletes the pointers to the various storage locations where the file is being kept on the storage media. The file, however, remains.

Computer storage media can be analogized to a library. The books are the individual files, the shelves are the various storage locations and the card catalog is the pointer to where those files are stored. When one deletes a file on computer storage media, what typically is done is that the card in the card catalog is changed to indicate that the file is no longer available. If one were to still go and look on the shelves where the card catalog says the book used to reside then it would still be found. The file for computer media will remain in its storage location until it is over written.

Recovering Deleted Files

There are several techniques for recovering deleted files. Searching for file signatures and data carving is one technique that has already been discussed. That technique is suitable when references to the file no longer reside in the file system

If references still exist in the file system then there are other techniques that rely on changing the file’s delete flag back to active. These days undeleting files has become routine. Numerous software applications exist for performing this process in short order or at least compiling data that would enable the recovery of the deleted data.

Beyond the entries in the existing file system, references to deleted files can also be found in file system remnants. Locating these remnants also relies on signature analysis. Essentially, the media is searched for the matching signatures. The related data is then compiled as if these remnants were part of the existing file system.

While data found in file system remnants may have a lower probability of recovery, the remnants can provide persuasive evidence of spoliation. The metadata that the remnants contain can be used to help identify when deleted documents last existed on the media and perhaps identify documents that were deleted after preservation efforts should have been initiated.

INFO2 File

The Recycle Bin is where files deleted by users are initially placed until they are deleted again by the user. Within the Recycle Bin is a file named INFO or INFO2 depending on the version of Windows. The INFO or INFO2 file contains the information about a deleted file such as its original location and its deletion date. The INFO and INFO2 files are the only place where a file's actual deletion date is typically recorded. For that reason, locating and reviewing the INFO and INFO2 files have significant interest for spoliation analyses.

Since system file deletions do not travel through the Recycle Bin its contents are limited to those files having been deleted by a user. When the Recycle Bin is emptied and the computer shut down, the INFO file is also deleted. Evil doers may at this point rest comfortably in thinking that their misdeeds will go undetected. If so, what they do not realize is that like other deleted files, the INFO or INFO2 files can be recovered as long as the file has not been overwritten. The recovery procedure requires searching for the file's signature in free space as well as matching its contents to the known layout of the INFO and INFO2 file's record format.

Disk Defragmentation

Every Windows system has installed a disk defragmentation utility. Through normal computer use, files can become fragmented as they are modified by users over time. As the contents of a computer media become fragmented, overall system performance can be degraded. The defragmentation utility provides a means of re-organizing the files so that they are in contiguous locations and system performance optimized.

The defragmentation utility also makes for an effective file wiping and shredding program. When the files are reorganized they are typically copied to the lesser used areas of the media before being returned to contiguous locations elsewhere on the drive. The process of moving these files two different times is an effective means of overwriting whatever data had been there.

If an evil doer deletes a number of files, then subsequently running the defragmentation utility could also end up overwriting those files and making them unrecoverable. Also, running the defragmentation utility can remove the references to deleted files from the filing system. Thus, the defragmentation process can overwrite a file and make it non-recoverable. In addition, it can clean the filing system and remove any artifacts that it ever existed.

CONDUCTING FORENSIC ANALYSIS

Determining if and when the disk defragmentation utility ran can be an important determination for a spoliation analysis. When the utility runs it will typically create a log file or update an existing log file. Finding that this has occurred after a preservation order is in place would be of great interest for a spoliation analysis.

Reformatting

Reformatting a storage media does not erase the data that it once stored. Rather reformatting simply resets all of the pointers that are used to reference the data stored in all of the storage locations. If the storage media is analogized to a library, reformatting would be like emptying the card catalog and replacing its contents with blank cards. There could still be all kinds of books on the shelf. All that would be required would be to go to the stacks and actually look at the books.

In the event that storage media has been reformatted by accident or purpose all that would be needed is to perform signature analysis on the media in order to identify file candidates. Essentially, one would search the drive using signature analysis to find instances of files having the requisite starting byte signatures and ending trailers. The data carving technique then takes everything in the middle and reconstructs the file.

If the recovered file was actually fragmented then its various components are not stored in contiguous storage locations. As a result, when all of the data between the starting signature and ending trailer are recovered and the file reconstructed, it likely will not function. In fact, when one views the file through a Word or Excel viewer, for example, the person is likely to only see a large volume of non-sensical special characters.

What needs to be done to facilitate the review process is that after the carved data is reconstructed that it be subjected to some kind of automated scoring process that evaluates the readability of the reconstructed documents. Those that receive a high enough score are passed for normal production. Those whose scores are not high enough can be produced in a separate batch for analysis using lower level review tools.

Steganography

The term steganography is of Greek origin and means covered writing. Steganography is different from cryptography in that cryptography does not hide the message. In other words, a password protected file is known to exist even though its contents are protected from disclosure by the encryption algorithm for which the password is the key. Steganography, however, hides the message in something that appears innocent. As a result, its very existence is unknown.

Steganography has a long history. One of its earliest uses involved a message about opposing troop movements that was tattooed onto the shaven head of a messenger. The message was hidden after his hair grew back and allowed him to travel to his destination and reveal his hidden message by shaving his head.

Since that time there have been a plethora of techniques developed to hiding messages. Some of the more pertinent methods are discussed as follows.

Changing File Extensions

Generally, people identify the file type by its extension. However, in a Windows environment the extension is not as reliable as in other systems. Consequently, one method of validating a file extension is through signature analysis.

In addition to be identified by their extensions, many types of files have a certain combination of bytes in the first few bytes of the file. As an example, the following signatures are found in the first bytes of the indicated file types.

SIGNATURE	FILE TYPE
PK	Zip file
ÿWPC	Word Perfect
MZ	Executable program file
Dĩ	Microsoft Office file

When attempting to hide data, evil doers often change the extension to a file and then mingle it in with other similar file types. For example, a Word document could have its extension changed from .DOC to .EXE and then the file moved to the Windows system directory where it will blend with other executable program files.

Signature analysis provides a means to examine all the files on a storage device and confirm that their file extensions match their file types.

Coloring

Coloring hides data by changing the font color to match the background, like white, or to have no coloring at all. Data that has been hidden by coloring can be found by changing the background to some other color and looking for a hidden message.

Text Box Sizing

Text box sizing is a technique where the message is placed in a text box and then the size of the text box is made so small that the message and even the box itself is imperceptible. Printing the document will not reveal the existence of the text box. Even performing a visual inspection on the screen will not reveal the existence of the text box if it is made sufficiently small.

Text Box Ordering

Text box ordering is a technique where the boxes are layered and one box with an innocuous message is positioned and sized to display over the top of the text box with the important message.

Hidden Data Format

Hidden data format is a technique that takes advantage of software features designed to hide data. Most word processors, spreadsheets and slide presentation software contain such features. Essentially, the message is inserted into a data file for one of these applications and then the option not to display the data is selected. This kind of hiding technique can be thwarted by selecting the entire document and removing the hidden text feature.

Embedded Images

Embedded images is where the picture of a message is embedded in a textual document. While the human eye will easily be able to read the message a search engine would likely not be able to detect the message.

Null Ciphers

Null ciphers is another technique with a long history. It involves embedding the characters of a message in another message that is constructed from the text of the hidden message. For example, each new sentence of the carrier message begins with a character from the hidden message.

Least Significant Bit

CONDUCTING FORENSIC ANALYSIS

The least significant bit technique is a newer method. It uses picture images, music and video files as carriers. Under this technique individual bytes or groups of bytes are used to represent individual pixels in picture and video or frequencies in audio. Small discrepancies in colors and sounds are not detectable by the human senses and thus go unnoticed.

For example, if a single byte has 256 different values then for an 8 bit (1 byte) color palette there are 256 color shades for the spectrum represented by that byte. It is unlikely that the human eye can distinguish between the shade represented by the 254th value versus the 255th value or the 23rd and 24th or the 100th and 101st, as an example. This technique uses the least significant bit of each byte, therefore, to carry the individual bit values of its message. Thus, it would take 8 bytes of the carrier message to carry a single byte of the hidden message.

The difference between the carrier file before and after it has been loaded with the hidden message is even less obvious when one considers that there is a 50/50 chance that the least significant bit is already set to the value needed for the corresponding bit in the payload message.

Appending

Appending is a technique where the secret file is appended to the end of the cover file. Often when software applications open files they read from the start of the file until they find the end of file marker. If two files have been appended then the software application will never read the entire file. The application will only read to the first end of file marker.

Assessment of Preservation and Production Techniques

After receiving digital data, one of the steps that a requester will want to take is an examination of the data to determine whether it has been altered since the preservation effort was performed. This can be accomplished by reviewing file system metadata that was produced as well as the metadata of other artifacts such as link files.

Another technique is to validate the MD5 or SHA-256 hash values of the preserved data to the valued computed for the produced data. One of the things that the requester should have asked to be produced was the hash values of the data when it was preserved.

Gaps in produced materials

Time line distributions are another way to exam the produced data for completeness. In addition to the data itself, file system metadata can also be used to create distributions over time of the files produced in order to validate that the production is complete.

Metadata Analysis: Finding, Interpreting and Evaluating

There has become an increased interest in metadata. According to Wikipedia, metadata (Greek meta "after", "about", "beyond" and Latin data "information") are data that describe other data. Generally, a set of metadata describes a single set of data, called a resource. Meta is a common English prefix, used to indicate a concept which is an abstraction from another concept, used to analyze the latter. For example "metaphysics" refers to things beyond physics, and "meta language" refers to a type of language or system which describes language. Metadata are of special interest in various fields of computer science and are used in features such as database, information warehousing, imaging, computer files systems, etc.

Metadata can involve data within the document files themselves as well as data within other system resources. For example, the filing system would contain various date stamps that would be of interest. With respect to backup tapes, the tape catalog, which is the tape's file system, would contain similar information.

There are still other resources that could contain important and relevant metadata. For example, the subjects discussed previously like the registry, event logs and link files are other sources of metadata that can have important and relevant metadata. The following sections discuss two other sources of metadata. They are the file system and the host file.

File System Metadata

If compute media can be analogized to a library then the file system is the equivalent of the card catalog. Just as within the card catalog there are specific pieces of information that readers expect to find, there are also specific pieces of data that are captured in the file system. The following sections discuss that data and how it can be used.

Last Written or Modified

The last written or modified date, also known as the last write date includes a time stamp. These values capture the date and time when the file was last modified. Unlike the creation date and time these values remain unchanged when a file is being copied from one drive to another.

Last Accessed

The last access date is also a feature found in Windows 98 and later systems. Notice that only the date value is captured. A time value has not been captured. The access date is changed whenever the file is used or its directory entry viewed in applications like Windows Explorer. Other system actions like virus checkers can also alter the last access date.

Created

The creation date and time identifies the date and time that the file was created on the drive. If the file is being copied from some other drive then its actual creation date and time could be much earlier.

Five bytes are required for the file creation date and time versus only four bytes for the file update date and time because the creation date and time is valid down within 10 milliseconds.

Master File Table Last Modified date

In NTFS systems there is also a date stamp for when the record in the MFT was changed.

Attributes

Each file is associated with several different attributes. The attribute features are bit mapped which means that individual bits of a byte are being used as semaphores to designate the attribute. The various types of attributes are Read-Only, a System file, a Hidden file, a Disk Drive volume name, a Subdirectory Name or Archive.

The archive attribute can provide information about whether a file has been changed since its last backup. First, when a full backup is performed then all files are backed up regardless of whether they have been changed. It is only when incremental or differential backups are performed that the status of the archive bit is noticed. If the file has been changed and the archive bit has been set then this is the signal to the backup system that the file should be backed up when the next differential or incremental backup is performed.

The archive bit is set to on by the operating system when changes are made to the file. The archive bit is set to off by the backup software after it backs up the file.

Locations

CONDUCTING FORENSIC ANALYSIS

Locations identify where the starting location of the data is on the media. Generally, this location is provided in terms of storage locations like clusters or sectors. It can also be presented in terms of absolute bytes, however.

Size

The size of a file is typically presented in bytes. Typically the file size can be a first indication of steganography. For example, the stated file size looks unusually large in relation to the data that it seems to contain such as an emptied e-mail file or a seemingly small Word document.

Host File Metadata

In addition to the filing system, the files themselves can contain metadata. This data can be used to validate the author, user, last written, created and accessed date stamps. In addition, the file might contain versioning information.

Author

Many programs capture the document's author name. Frequently, this is automatically filled by the application based on a value entered at the time of installation of the software or the installation of the operating system or when the user authenticated and logged on to the system.

User

Many times the name of the user of a data file is also captured. This value too is often automatically filled based on values entered at the time the application was installed, the operating system was installed or when the user authenticated and logged on to the system.

Date Stamps

In addition to the file system time and date stamps, the host file can often include a number of its own date and time stamps for file creation and modification.

Change Data

CONDUCTING FORENSIC ANALYSIS

When the document is changed, revision data can also be retained. Sometimes this data is maintained within the document such as with Word documents. In other times the revision data is retained in a separate archive file as engineering documents under formal configuration control and many database systems.

Conclusion

There is more to electronic discovery than digitizing documents or searching digital documents. Indeed, there is a great wealth of information that can be simply and easily discerned. In addition to the examination of content to prove or disprove claims and defenses, digital evidence can also be used to determine the respondent's compliance with a production request. This determination extends to both the completeness of the production as well as any efforts to "hide the ball".

Everyone has come to understand that deleting data does not necessarily delete the data and that deleted data can often be recovered. What may not be as widely appreciated, however, is that utilizing a computer system is like throwing pebbles in a pond. While the pebbles may nor may not be recoverable, the ripples they leave are detectable nonetheless.

Understanding the various forensic techniques enables requesters to avoid being duped by various data hiding techniques. In addition, it educates requesters about how to supplement their production requests with selected files and data in order to validate the respondent's production.

1. *Anderson v. Crossroads Capital Partners, L.L.C.*, 2004 WL 256512 (D.Minn. Feb. 10, 2004)