Chapter 14

REDACTING DIGITAL INFORMATION FROM ELECTRONIC DEVICES

A. Barclay, L. Watson, D. Greer, J. Hale and G. Manes

Abstract Redaction is the process of removing privileged information from a document before it is presented to other parties. This paper discusses the major issues associated with the redaction of digital information from electronic devices. A novel technique involving a tokenized representation is presented as a solution to digital redaction in legal proceedings.

Keywords: Electronic discovery, digital information, redaction, tokens

1. Introduction

With the introduction of digital evidence into the court system [8–12], the private sector and federal government must address a growing number of best practice issues in the field [1–4, 6]. This is especially true for digital redaction. Redaction is the process of removing privileged information from a document before it is presented to other parties. This paper focuses on digital redaction as it applies to the legal community.

During the discovery phase of court proceedings, it is necessary to provide information that is requested by opposing counsel. In general, an attorney's work on a case is protected by the work-product privilege, communications are protected between an attorney and their client, and other parties have no right to this information. The work-product privilege means that any documents prepared in anticipation of litigation or for trial by a party's representative enjoy qualified immunity from discovery. Similar privileges are involved in doctor-patient, priest-penitent and husband-wife relationships. To prove to the court that information is privileged, the party claiming privilege must show that the communication: (i) was made with an expectation of confidentiality, (ii) is essential

to a socially-approved relationship or purpose, and (iii) has not been waived by disclosure of the contents of the communications to persons outside the relationship.

The redaction process produces three items: an In Camera Copy of the privileged information, a Privilege Log and a Redacted Production Copy of the information. The In Camera Copy, which is presented to the judge in the case, contains all the items regarded as privileged. The Privilege Log and Redacted Production Copy are presented to opposing counsel. If a question arises as to whether a particular item in the Privilege Log meets the burden of privilege, the judge can review the material in the In Camera Copy and provide judgment.

Traditionally, the requested information has been presented in paper form. Currently, two methods are used to redact paper documents: "blackout" and physical removal. The blackout method involves using a black marker to conceal the portions of a document that are considered privileged. The physical removal method involves removing certain pages from a set of documents. Depending on the court's requirements, this may necessitate marking the exact locations from which the pages were removed.

The same concerns exist for privileged information residing on electronic storage devices, but no standard method of digital redaction has been adopted by the legal community [7]. Computerized methods that mimic the blackout process exist [5], as do those for mimicking the physical removal method. The former approach may engage a number of techniques to conceal text in a digital document. The latter typically involves collecting all the readable documents from a computer, placing them in a set, and selecting the items to redact. Yet, while electronic blackout and removal methods can sanitize a document found on an electronic device, they do not redact logical copies or copied fragments of the document that may remain.

This paper discusses the process of digital redaction and highlights the major issues associated with redacting digital information from electronic devices. A novel technique involving a tokenized representation is presented as a solution to digital redaction.

2. Data Redaction

With respect to redaction, it is increasingly necessary to produce the entire contents of computer disks and other electronic storage devices as evidence. This goes beyond simply selecting all readable documents on a drive. It involves producing information that exists in free or slack space, deleted items, document fragments and even data that may not be

in a readily identifiable format. The collection process produces what is referred to as a "forensics copy." This encumbers a data image redaction process to remove privileged information from the storage device.

2.1 Challenges and Considerations

The growing variety of electronic devices that integrate digital data storage components complicates the issue of data image redaction. Devices such as cell phones, digital cameras and digital music players, along with laptops and desktop computers store information using various file systems, media technologies and data formats. The sheer diversity of these storage options differentiates digital redaction from its physical pen-and-paper counterpart.

Aside from the variety of storage formats, other challenges to data image redaction in electronic devices include:

- The potential for encrypted data
- Deleted files that are recoverable in slack space or unoccupied regions of file systems
- Data fragmentation and replication
- Isolation of privilege by context for integrated data

A faithful digital redaction process must account for these subtleties in a systematic and comprehensive manner.

To completely redact digital information from an electronic device, it is imperative to determine the logical and physical locations of all pertinent documents and related data fragments that reside on the digital media. This is an issue because data is routinely stored in multiple locations on file systems in electronic devices. For example, Microsoft Word files are normally saved in a user-selected directory, but may also be automatically backed-up in a temporary folder as a part of normal operation; therefore, a Word document may logically exist in at least two separate locations on a computer system.

Deleting privileged information from digital media does not protect it from a thorough forensic examination. The only versions of a document that can be directly deleted are listed in file mapping tables. Other copies of the item are unaffected by the deletion of the original document and, therefore, could be recovered by a forensic examination.

Determining all the physical locations where digital information resides is also important due to the partitioning methods used in electronic media and devices. For example, suppose a user creates a file on a Linux system and subsequently saves it to a FAT partition of a hard drive. If

the drive is subsequently repartitioned, the file may fall out of the new logical partition size and be moved into the space on the hard drive reserved for the resized FAT partition. Thus, the file exists in the new location and in its original location.

To determine whether information is privileged, it is necessary to interpret the information rationally; if the information is unreadable, privilege cannot be determined. This presents a problem when the information stored on electronic devices is encoded, protected or encrypted. During the redaction process, digital data without rational interpretation may be produced because it contains no apparent privilege. In fact, the data may contain privileged information that is concealed by the encoding. Consequently, if a rational interpretation is discovered later, the data can be decoded. Thus, the possibility exists that privileged information could be revealed to opposing counsel.

The accuracy of the data image redaction process is also important. When producing a redacted copy, care should be taken to demonstrate that the integrity of the redacted copy is preserved as it relates to the source media. The redaction process should only remove the data segments marked for redaction and leave all other segments untouched. Thus, digital redaction methods should incorporate validation schemes that offer assurance regarding the integrity of the redaction process.

2.2 Foundational Methodology

There are requisite procedural elements for any system that aspires to meet the challenges of data image redaction. The first is to characterize privileged information. Subsequently, an investigation must be conducted on the Work Copy of the electronic device. This investigation should identify privileged information, complex and indeterminate data objects, and produce an index of redactable items. Finally, the data must be redacted to produce both a Redacted Production Copy with an associated Privilege Log, and an In Camera Copy.

Characterizing Privileged Information Redaction allows for the selective exclusion of information protected under privilege as defined by federal, state and local laws. These protections, e.g., based on attorney-client or doctor-patient relationships, provide different classes of privileged information.

The selection of privileged content is based on the current legal standards for such material. These standards involve communications between accepted members of an accepted privilege class acting in an accepted capacity. Additionally, the court may indicate that certain topics are off-limits and that related material is to be redacted as well.

Forensic investigations of digital media typically employ keyword or pattern-based searches to find relevant information. Such searches can also be used to identify redactable information belonging to a privilege class. By associating search criteria based on metadata and content with specific redaction classes, all the information on a source disk can be classified as privileged or non-privileged, with privileged information being additionally associated with a specific class.

Electronic Device Investigation The redaction process operates on a Work Copy of an electronic device, which is typically a forensic copy of the original media (it could be the original media if it is impractical or impossible to create a copy).

The investigation identifies known and unknown files by data carving the entire source media, finding deleted files in free space, hidden files, slack space, etc., via header and footer analysis. These files and hidden files are then keyword/pattern searched and each file object is labeled as being privileged/non-privileged, complex or indeterminate.

Data Objects A forensic investigation can reveal data that is not immediately interpretable; thus, the keyword/pattern identification will not be able to determine privilege status. Such data may be structured, compressed or otherwise encoded for interpretation by a special application or method (e.g., an Outlook PST file for an e-mail application). Encryption, data scrambling or fragmentation may also prevent immediate interpretation of data. Any data that is encoded or structured (and recognized as interpretable by a special filter or application) is treated as a "complex data object."

A metaphorical example of a complex data object is a sheet of used carbon paper containing interwoven, overlapping documents that are not easily interpreted. Initially, it is unclear if the carbon paper contains privileged information. However, further analysis could yield the individual documents that contain privileged information. Clearly, it would be irresponsible to release the carbon paper sheet to opposing counsel without performing an analysis and redacting privileged information.

Complex data objects are subject to an additional investigative process using appropriate tools and techniques that interpret the data and make it readable. The interpreted data can then be subjected to digital redaction. When no interpretation method is available, they can be regarded as "indeterminate data objects" and may be redacted until a method for interpretation presents itself (at which time the objects transition to complex data objects).

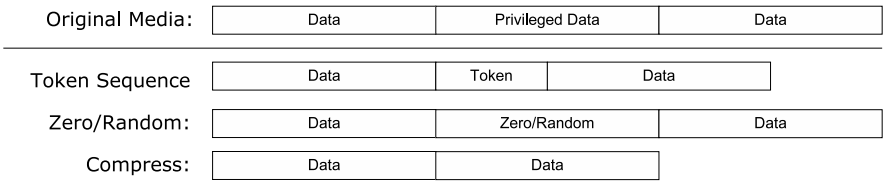

Figure 1. Data removal methods.

An example of an indeterminate data object is again a sheet of carbon paper, but one that had been used very extensively. Even if it is not possible to extract individual documents based on current process, it would be irresponsible to release the carbon paper sheet because a new process could be created to extract the privilege-bearing documents. Note also that complex data objects may themselves contain complex data objects, which would require redaction to be a recursive process.

Redaction A Redacted Production Copy of a data image is created by copying the Work Copy to sterile media after removing the redacted objects. This copy must contain no privileged information and, depending on the legal mandate, no complex/indeterminate information. Both the Redacted Production Copy and the Privilege Log are provided to the opposing counsel.

The privileged data and metadata are combined to create the In Camera Copy, which is passed to the judge. The judge uses the In Camera Copy to mediate disputes on the appropriateness of a redaction when there is a challenge to reveal redacted information.

Three approaches are available for removing redacted data from an image (Figure 1). The first approach, which is discussed in the next section, is to replace the redacted data with a bit sequence or token. This provides a replacement for the removed data, emulating current physical redaction processes and adding Privilege Log metadata to the Production Copy.

The second approach is to keep the Privilege Log separate, filling the space formerly occupied by the redacted object with zeroed or random data. However, this may not be appropriate in a legal setting due to the potential for inference analysis based on file size and location. It should also be noted that even when fill data is randomized, "pseudo" artifacts such as false file headers in the redacted copy are possible. However, the benefit of this approach is that it closely emulates the original media and is compatible with current investigative strategies.

The third approach is to remove the privileged data and compress the non-privileged data together. This Redacted Production Copy for-

mat is a bit sequence with no describing information intact. Inference analysis of the original data would not be readily apparent; however, some reconstruction is possible using allocation tables, etc. if they are not deemed privileged. Like the randomized method above, this technique may introduce "pseudo" artifacts. Many of the implementation constraints identified – such as minimizing inference analysis, mimicking physical redaction, adding value to software examination methods by exposing redaction points, and encoding non-privileged meta-information – are best addressed by implementing a token-based redaction system.

3. Redaction Tokens

Redaction tokens are bit sequences that replace or stand for private data, complex data objects or indeterminate data objects in the Redacted Forensic Copy. As such, they provide a method to describe the redaction process to the court and other examiners. Tokens help confirm the integrity of the redaction process and provide an accessible layer of abstraction for juries. Implementation requirements would, of course, depend on legal statutes and precedence. Nevertheless, redaction tokens have several inherent advantages:

- Tokens can create identifiers that bind redacted data objects to the Privilege Log.
- Tokens can act as markers for interoperability with other programs, making redacted data segments recognizable to external tools. Forensics suites could recognize markers and skip data carving or sliding window analysis on token data/metadata.
- Tokens can provide a basic audit log, with the token encoding information about the examiner, case, etc.
- Tokens can contain a digital signature of the examiner, supporting non-repudiation and chain of custody.
- Tokens can include a one-way hash of the redacted object to verify the integrity of the original object and the In Camera Copy.
- Tokens can emulate the pre-redaction environment; all non-redacted information will appear intact.
- Tokens mimic the paper redaction system familiar to courts, providing a conceptual understanding of digital redaction.

Bit sequences for redaction tokens may be generated in a variety of ways depending on the purpose of the tokens. A token can serve as a

method to represent redacted data, bind meta-information and provide accountability, or any combination thereof. The size of the smallest redacted object could also dictate the potential contents of a token, especially if a court requires that the original file sizes be maintained. For the UTF8 encoding format, a name that might be considered privileged could be as small as 6 bytes (thus, this becomes the maximum token size). On the other hand, redaction of large image files increases the prospective size of the token, potentially adding to its abilities.

Several issues must be considered when generating tokens. Tokens for each production must be consistent in format and agreed upon by all parties. Tokens should also be amenable to parsing. This issue is more complex than it might initially appear because tokens must avoid magic numbers and other bit sequences used in file headers and file system constructs. Additionally, tokens should be easily identifiable and generated in a reasonable amount of time. Finally, tokens must never reveal information about the contents of data objects represented in the Redacted Production Copy.

Figure 2. Basic redaction token schema.

3.1 Token Schema

The basic redaction token schema consists of a common start sequence, unique randomly-generated id, any number of optional enhancements, followed by a common closing sequence. A representative schema is presented in Figure 2, where the required elements are shown with solid lines and optional elements with dotted lines.

3.2 Token Methods

Redaction methods based on overwriting privileged data with tokens vary according to parsing speed, space needs, legal requirements and token schema/size. Token-based redaction methods include (Figure 3):

- Placing a single token in a data segment and compressing the redacted image to eliminate the remaining bytes of the segment.
- Replacing all the bytes in a data segment with a repeated token sequence.
- Placing a copy of the token in a data segment and replacing the rest of the segment with zeroed or random data.

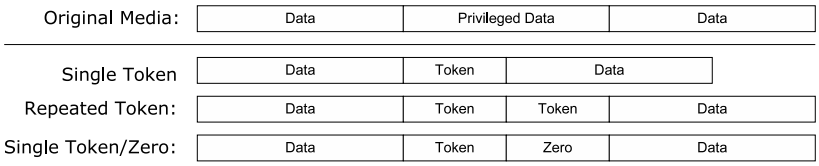

Figure 3. Token redaction methods.

The first method substitutes a single copy of the redaction token for the redacted data object, shrinking the image written to the Redacted Production Copy to eliminate the byte-level data storage occupied by the remainder of the object. This confutes inference analysis of the original size and location of the redacted object.

The second method creates a redacted forensic copy in which all the bytes that are selected as redactable are replaced with repeated copies of a token. Consecutive copies of the token are written until the redacted data segment is filled. The last copy of the token is a special case where it can either be written completely or compressed to include the closing stop sequence. This method preserves an accurate forensic copy with only the redacted information removed. It is analogous to blacking out entire documents and leaving them in place in a file box. However, this method permits inferences based on the size and spatial relationships of the redacted data.

The third method replaces all redactable bytes in a data segment with a single token followed by overwriting the remaining privileged bytes with zeroed or random data. This method is analogous to putting a placeholder and the same amount of blank pages as the original document into a file box. It closely models current paper-based redaction, but permits size and spatial inference analysis.

4. Conclusions

The growing volume of digital evidence introduced in legal proceedings makes digital redaction a major issue. Unfortunately, current redaction techniques do not address the complex issues related to evidence residing in electronic devices. Incomplete redaction may enable opposing counsel to access privileged information using digital forensic techniques and tools. Furthermore, criminal entities may be able to prevent electronic discovery based on the claim that privileged information is present and that no methods exist to guarantee privilege removal. The redaction technique presented in this paper addresses these issues by using tokens to systematically and comprehensively remove redactable information.

References

[1] M. Arkfeld, *Electronic Discovery and Evidence*, Law Partner Publishing, Phoenix, Arizona, 2005.

[2] B. Carrier, *File System Forensic Analysis*, Addison-Wesley, Crawfordsville, Indiana, 2005.

[3] A. Choudhri, L. Kagal, A. Joshi, T. Finin and Y. Yesha, PatientService: Electronic patient record redaction and delivery in pervasive environments, *Proceedings of the Fifth International Workshop on Enterprise Networking and Computing in the Healthcare Industry*, pp. 41–47, 2003.

[4] National Institute of Standards and Technology, Computer Forensics Tool Testing (www.cftt.nist.gov).

[5] National Security Agency, Redacting with Confidence: How to Safely Publish Sanitized Reports Converted from Word to PDF, Technical Report I333-TR-015R-2005, Fort Meade, Maryland, 2005.

[6] R. Nichols, D. Ryan and J. Ryan, *Defending Your Digital Assets Against Hackers, Crackers, Spies and Thieves*, McGraw-Hill, New York, 2000.

[7] G. Palmer, A Road Map for Digital Forensics Research, Technical Report DTR-T001-01, Digital Forensic Research Workshop (isis .poly.edu/kulesh/forensics/docs/DFRWS_RM_Final.pdf), 2001.

[8] U.S. District Court (District of Minnesota), Northwest Airlines v. IBT Local 2000, *Labor Relations Reference Manual*, vol. 163, pp. 2460–2461, 2000.

[9] U.S. District Court (Eastern District of Michigan), United States v. Alexander, *Westlaw* 2095701, 2004.

[10] U.S. District Court (Southern District of California), Playboy Enterprises, Inc. v. Welles, *Federal Supplement, Second Series*, vol. 7, pp. 1098–1105, 1998.

[11] U.S. District Court (Southern District of Indiana), Simon Property Group, L.P. v. mySimon, Inc., *Federal Rules Decisions*, vol. 194, pp. 639–644, 2000.

[12] U.S. District Court (Southern District of New York), Anti-Monopoly, Inc. v. Hasbro, Inc., *Westlaw* 649934, 1995.