

## **Geeks with Guns, or How I Stopped Worrying and Learned to Love Computer Evidence**

**Richard L. Hardy<sup>i</sup> & Susan S. Kreston<sup>ii</sup>**

**Introduction.** When testifying about computer forensic evidence, it is often necessary for the forensic examiner to explain technical terms and concepts to the judge or jury, or to rebut technical testimony and evidence offered by the defense. As is so often the case with a “battle of the experts,” what the jury chooses to believe depends upon whom they accept as their teacher. The difficulties in testifying about highly technical evidence encompass a need to be technically accurate while also being clear. To accomplish this, complex concepts must be explained in easily understood terms while conveying a true picture of the underlying technical information. Over time, experienced forensic examiners have developed analogies that assist with understanding several of the most commonly addressed issues. The best of these analogies have been developed in concert with prosecutors, investigators, and other examiners, and have been refined and extended over time. Following are some concepts that need to be addressed in court, and analogies that could prove useful in their explanation.

**File Allocation Table.** The File Allocation Table (FAT) can be likened to a library card catalogue that discloses the location of books within a library. The title of the book is contained in the Allocation Table, and the book is the File itself. To find a book, one looks up the name in the card catalogue, which then points to the book’s location.

**Deleted Files.** Deleted files are files whose reference has been removed from the file system, and the area of the electronic media they occupy is released for reuse. Until overwritten with new characters, these files may be recovered.

The use of the term “deleted” is, in many ways, misleading. Deleting a file can be analogized to putting household garbage in a garbage bag, but keeping the bag in the house. While the garbage has, technically, been thrown away, it can still be readily retrieved. Therefore, finding a deleted file is simply a matter of finding the file that remains in the FAT, but with a different first character.

Using the library card analogy discussed above, when a book (file) is deleted, the library card referring to that book is replaced with one containing the book’s name, minus the first character, and no reference information to the location of the book. On the FAT, the first character of the file’s name is replaced with a *sigma* or “σ”. For example, a file named “Criminalevidencehere” would become “σriminalevidencehere.” This effectively makes it impossible to find the file simply by searching for it under its title. However, the book itself is not touched until the space it occupies on the shelf is reused by another book. Prior to that, recovering the deleted book is a matter of finding the replaced card, and re-referencing it to the location of the original book, which can then be read in its entirety.

**Wiping.** “Wiping” a file entails not only deleting the reference to the file in the FAT, but also overwriting its contents, usually with “0”s, preventing later recovery of its contents. Unallocated space and file slack are also common targets of this practice. This process requires specialized software. An analogy for this procedure could be the act of wiping fingerprints from surfaces at a crime scene.

**Unallocated Space.** This is area on the media that is not currently referred to by the file system. If this area has been previously used, and not “wiped,” it will contain remnants from that prior use. Deleted files are one type of unallocated space. The other type of unallocated space would be space that has never been allocated (made part of any file). Returning to the library card analogy, unallocated space would be those areas of the library with no corresponding library card. If books or parts of books are still on the shelves (not wiped) however, they may still be read, even though they are not recognized by the card catalogue system.

**File Slack.** File slack is the area allocated to a file that the file does not actually use. File slack may or may not contain data, depending on whether it has been previously used by other files. A useful analogy for file slack is videotape. If you record a half hour show to a fresh hour long videotape, the remaining thirty minutes is analogous to file slack, even though it will be empty. If you had previously recorded a one hour program to that tape, once you finished watching the half hour show, you would see thirty minutes of the originally recorded one hour program. This would be analogous to file slack containing the contents of a previously deleted file.

**Sectors and Clusters.** A sector is 512 bytes of information. Sectors are then grouped into clusters. A cluster will contain several sectors, commonly 8 or 16. A cluster is the smallest area that can be written to by a file, although the file need not fill the entire cluster. An analogy for this would be sectors as pages in a manila file folder, with the file folder itself being the cluster. This analogy can be extended to include a filing cabinet that holds the folder representing the hard drive.<sup>iii</sup>

**RAM Slack.** RAM slack is the area in the last sector used by a file that contains information taken from whatever is in the computer's memory at the time that the file is written. Using the videotape analogy again, RAM slack could be likened to the commercial at the end of a recorded program, filling out the final portion of the recording time programmed into the VCR.

**Internet Protocol Addresses and Domain Name Servers.** Internet Protocol (IP) addresses are numbers used to locate computers on a network. Every computer has an IP address when it is in use. Phone numbers are analogous to IP addresses in that for one phone to communicate with another, both must have unique numbers in the phone system.

This analogy can be expanded to explain Domain Name Servers (DNS) by using the analogy of a phone book, where the DNS uses a typed in name to find the IP address number, just as a phone book uses a name to find a phone number. Domain Name Servers (DNS) are computers that translate plain text names into Internet Protocol (IP) addresses by using a table cross-referencing the two. For instance, typing "[www.cnn.com](http://www.cnn.com)" into a browser, such as Internet Explorer or Netscape, will cause a check to a DNS that will return the IP address "[64.236.24.20](http://64.236.24.20)," pointing the browser to the computer that contains the requested web page.

**E-mail headers and "Spoofing."** E-mail headers contain routing information from the e-mail programs used to forward the message through to its final destination. These are assembled by the servers and are attached at the top of the e-mail. They can be used to trace the origin of the e-mail being examined. IP addresses would be one type of information found in the e-mail header. Two analogies for e-mail headers could be stamps placed on a passport as an individual passes through countries, or forwarding addresses placed on a piece of mail by each post office as the letter passes through the postal system to its final destination. Spoofing e-mail is the act of forging information contained within the header of the e-mail to attempt to mask its origin. One possible analogy for e-mail spoofing would be check fraud, in which the name and address information on the check is changed prior to the check being passed.

**Encryption and Steganography.** Encrypting a file or image is the deliberate scrambling of information, so that the original information may only be unscrambled with a key. It is an attempt to hide the content of the message and may be likened to a secret code that needs the "decoder ring" to discover the meaning of the message.<sup>iv</sup> Steganography differs from encryption in that it is an attempt to hide the very existence of a file. Steganography hides one file inside another, so that the second file is completely concealed. Steganography can best be summarized as "Hide in plain sight."<sup>v</sup> The hidden file will usually be encrypted and/or password protected. An analogy for steganography could be a book with a hollow area, used to hide something within it. A slightly more apt analogy might be a picture created from very small words that only reveals its text content upon close inspection. Another might be a hide-a-key rock. A final analogy is the hidden stairway behind the wall in murder mystery movies.

**MD5 hash values.** These are extremely exact measures of the size and structure of an object, such as a disk, a file, or a folder. An analogy for this value would be an electronic fingerprint for the object. The slightest alteration in an object, such as minutely cropping an image, or changing the shading on even one pixel, will result in a completely different hash value. Hiding an image or a file within an object, i.e., steganography, will cause a drastic change in hash value. The scientific possibility of two different objects having the same MD5 hash value is more than 1 in 340 undecillion.<sup>vi</sup> This is a higher level of certainty than even DNA enjoys.

**Conclusion.** The development of analogies for computer forensic topics is an ongoing collaborative process. Sharing and discussion of these analogies and explanations can help prosecutors and forensic examiners in their quest to find methods to explain complex technical concepts in an understa

---

<sup>i</sup> Detective/Computer Forensic Examiner, Regional Computer Forensic Laboratory, San Diego, California. The author may be contacted at [rhardy@rcfl.org](mailto:rhardy@rcfl.org).

<sup>ii</sup> Counsel, National Center for Justice and the Rule of Law. The author may be contacted at [skreston@olemiss.edu](mailto:skreston@olemiss.edu) or at [susankreston@cyberforensicassociates.com](mailto:susankreston@cyberforensicassociates.com)

<sup>iii</sup> The authors would like to thank Mark Menz of My Key Technology for this analogy.

<sup>iv</sup> The mathematical probabilities of breaking an encrypted file are daunting. Odds of winning the state lottery today = approximately  $2^{28}$ . Chances of being struck by lightning today = approximately  $2^{33}$ . Chances of being struck by lightning and winning the top lottery prize = approximately  $2^{61}$ . Possible permutations of a modern encryption key = approximately  $2^{1024}$  – (there are fewer atoms in the known universe).

<sup>v</sup> See Brad Astrowsky, STEGANOGRAPHY: Hidden Images, A New Challenge in the Fight Against Child Porn, UPDATE, Vol. 13, No. 2 (2000). This publication is also available on-line at [http://www.ndaa-apri.org/publications/newsletters/update\\_volume\\_13\\_number\\_2\\_2000.html](http://www.ndaa-apri.org/publications/newsletters/update_volume_13_number_2_2000.html)

<sup>vi</sup> Undecillion would be written as a 1 followed by 36 zeros.