Demystifying ESI by Defining Key Terms

By now you have most likely heard of the recent federal rule changes addressing electronically stored information. If you haven't, fire up the computer and start Googling. There is a lot of information that has been written on ESI in the past 12months.

If you are familiar with ESI and its preservation and collection, you may have heard terms thrown around like "bit-stream image,""file slack" or "unallocated clusters." You may have even used the terms yourself without actually understanding what they mean.

Just like a file cabinet

Let's start with a brief, oversimplified overview of how computers store information on a hard drive. A hard drive resides inside a desktop or a laptop chassis, and it's about half the size of a brick but much lighter — in the case of a laptop, much smaller and lighter.

For the purpose f his not-too-technical discussion, we are going to assume that a hard drive contains all the ESI on a computer. Inside the hard drive are magnetized platters that can store billions of bits (binary digits) of information. A bit is either on or off, a one or a zero. These bits are then organized into larger segments called bytes. Eight bits equal one byte, and groups of 512 bytes are organized into one sector, physical space on a hard drive. The hard-drive file system then groups contiguous sectors into clusters.

Confused? Let's use the analogy of a filing cabinet and a table of contents or index. Our make-believe filing cabinet has several drawers, each containing 100 folders (clusters). Each folder can hold 100 pieces of paper (bytes), and we can take it one step further and think of words or letters on the paper as bits.

In the real world, if one wanted to retrieve a paper file located somewhere in the filing cabinet, one could consult the index or table of contents and pull out the file. A computer works in much the same way.   In addition to other complex tasks, an operating system (for example, Windows XP or 2000 and Mac SX) tracks and maintains all of the files in its virtual filing cabinet.

The difference comes when a file is deleted.   In the paper world, one would take a file out and throw it in the trash or shred it and remove the entry in the index. In the digital world, it is impractical for a hard drive to "wipe" or throw out every file when the delete key was pushed, so the computer simply crosses off the entry in the index. The computer now sees hat portion of the cluster or "folder" as available, but the old data, or bytes, still exist.

The free area here he old data resides is known as unallocated clusters. Because the old bytes are still there, computer forensic professionals are able to search unallocated clusters, retrieve and analyze data that a user has slated for deletion.

Cut the slack

File slack is a term that is often misused by the uninitiated. In its simplest terms, file slack is wasted

space in a cluster. If a cluster has the capacity of 4,096 bytes and a file that is ,000 bytes is written to that cluster, the remaining 1,096 bytes is file slack.

It's when that original file is deleted that things become more interesting. Another analogy: Go back a few years to when there was no TiVo. If you recorded a 60-minutetelevision program on a VHS tape and then someone recorded over the first 45minutes with another program, you would obviously lose those first 45 minutes of the60-minute program. However, 15 minutes of the original program still remains.

On a hard drive, that remaining data is file slack. Let's say that a computer file occupies an entire eight sectors, or exactly one cluster, for a total of 4,096 bytes of data. Think of bytes as words or letters. If that file is deleted, then that cluster is marked by the operating system as available for new data to be written. Remember, the data is not deleted; only the entry in the index is removed. Let's then assume that anew file is created that is 3,584 bytes in size, and it's looking for a place to call home. It cozily settles in the eight sectors that were originally occupied by our 4,096 byte file.

In this example, if we performed forensic analysis and looked at this one cluster, we would find 512 bytes of data from the original file. This may or may not yield ESI with evidentiary value, but none the less, it is residual data or, more accurately, file slack.

Following the bit stream

On paper, unallocated clusters and file slack may seem like a potential treasure trove of evidence. The proverbial smoking gun may be lurking in these areas, so how do we get it?

File slack is captured by a bit-stream image (the terms "forensic copy" or "forensic image," "hard drive clone" and "bitforbit copy" all have the same meaning). A bit-stream image is an exact copy of every bit that is found on a hard drive.

A bit-stream image also captures all of the metadata (data about data) associated with a file, but the imaging procedure has to be executed properly to ensure that no data is altered. Computer forensic professionals take great care to make certain that no data is altered during the imaging procedure. They utilize specialized equipment called write-blocking devices that halt any inadvertent alterations to a hard drive.

I have received more than a few calls from attorneys requesting that I create a bit-stream image of a hard drive because their client suspects that an employee has been pilfering trade secrets or violating usage policies. They then proceed to tell me that their client's IT guy was unsuccessful in his attempt to find anything useful, so they want to give me a crack at it.

This sends shivers down my spine, because I know that, while unintentional, the poking around by the IT guy has most likely altered or destroyed potential evidence.

I am not advocating barricading an employee's cubicle with yellow police tape when there is a suspicion of wrongdoing, but there are prudent measures that can be taken to mitigate the possibility of destroying evidence.

First and foremost is to limit access to the computer if it may contain any relevant evidence. One course of action is to have an IT representative or other qualified person remove the hard drive or the entire computer and put it in a safe, secure location until it can be properly imaged and examined, if necessary.

The flip side of creating a bit-stream image of a hard drive is an active-file collection. Creating a bit-stream image of a hard drive is not always necessary or prudent. Some matters only require the preservation and production of active files from hard drives or other sources of ESI. An active-file collection would only capture those files listed on the virtual index and would not capture any deleted files, unallocated space or file slack.

A fragile resource

These are matters that litigators often squabble over during the meet-and-confer stage, when preservation and production of ESI are discussed.

The preservation and collection of ESI is arguably the most important step in the e-discovery process. Data is fragile, and any mishandling of ESI can spell disaster for your client; the mere act of booting into Windows XP alters at least 50 files and creates a few new ones, as well.

Under the new federal rules, attorneys have an obligation to guide and advise their clients with respect to the complete and accurate preservation and production of all potentially relevant ESI in each proceeding. By familiarizing yourself with terms addressed here, you are one step closer to fulfilling that obligation — and perhaps on your way to becoming an e-discovery attorney.