RETHINKING THE STORAGE OF COMPUTER EVIDENCE

Computer Crime and Intellectual Property Section
Criminal Division, US Department of Justice

Tyler Newby, Trial Attorney, CCIPS
Joel M. Schwarz, Trial Attorney, CCIPS
Ovie L. Carroll, Director, CCIPS Cybercrime Lab

I. INTRODUCTION

When a federal criminal investigation involves computer evidence, prosecutors and investigators often rely on the services of investigators who have special training and accreditation in the field of computer forensics. These forensic examiners are typically responsible for the collection, processing and analysis of digital evidence acquired during an investigation. Primary among computer forensic examiners' duties is ensuring that the data seized during an investigation remains unaltered through trial.

The foundation of electronic evidence collection and analysis and the subsequent admissibility and use of that evidence at trial is the creation of a forensic image. Once a forensic image of the original data is created, it is typically copied to a hard disk drive, which is then stored in a locked evidence room. Chain of custody logs are maintained for anyone who accesses the hard drive image.

In complex cases, such as intrusion cases, a prosecutor or case agent may request full forensic analysis of an image to search for evidence to be used at trial. In less complex cases, a case agent or prosecutor may want to conduct a triage review of the image to search for easily identifiable evidence of a crime, such as pirated software and movies, chat logs and e-mails discussing the crimes, digital photographs and the like. In that case, the case agent may want to review a working copy of the forensic image, which requires putting in a request for a working copy image to be made.

In either situation, case agents and prosecutors are likely to confront a long queue when they put in a request for assistance from computer forensic specialists. As electronic storage of data has become increasingly common, the demands placed on a limited pool of computer forensic examiners have increased. For example, in the Federal Bureau of Investigation's *FY2008 Authorization and Budget Request to Congress*, it noted that its Computer Analysis and Response Team's (CART) case backlog increased 58% from 1,258 cases to 1,991 in just a one year period from FY2004 to FY2005 and is likely to increase in the future. As electronic communication devices, home networks and increasingly capacious hard drives become more prevalent, already thinly stretched investigative resources are likely to be in even more demand. Thus, it is not unlikely that a hard drive containing the evidence prosecutors need to prepare and try their cases will sit on a shelf for a period of several months, if not years.

This reality raises the basic question of whether storing an increasing number of hard drives – which like all things mechanical can break – for years on shelves in evidence rooms is the best way to store digital evidence. This article suggests an alternative evidence storage method for forensic images – storing them on secured Redundant Array of Independent (or Inexpensive) Disks (RAID) systems. This alternative may save space in evidence rooms and will better protect sensitive evidence from inadvertent destruction. Furthermore, storing images on a RAID, if done properly, will not affect authentication of the image as a duplicate of the original electronic media at trial.

This storage method most clearly applies to cases in which investigators make an image copy of the electronic media at the scene. Where investigators remove computers containing electronic evidence from the scene, use of RAID storage may also be appropriate, but prosecutors should consider the possibility of defence challenges before wiping the original computer hard drive or returning it to its owner. Of course, if the computer hardware is seized because it is contraband, the fruit of a crime, or an instrumentality it should be retained pending disposition of the case or forfeiture proceeding.

II. THE BASICS OF FORENSIC IMAGING

Forensic imaging is the process used to obtain a bit for bit copy of the data residing on the original electronic media obtained by law enforcement – regardless of whether that media is a single hard disk drive, flash memory card, DVD, compact disc or mobile phone SIM card. The imaging process entails the copying of all of the data present on the original storage media device, including system files, hidden and deleted data from allocated (partitioned), unallocated (un-partitioned), and free space (un-used space on a formatted partition).

Once the imaging procedure is completed, the image of the hard drive contains all logical files, erased files, and unused space which are available to the original hard disk drive. From there, the investigator can examine the image for relevant evidence, without accessing the original seized hard drive at all. This process allows investigators to review a duplicate of the original evidence while preserving that evidence in exactly the form it existed at the time of seizure.

III. EVIDENTIARY ISSUES RAISED BY FORENSIC IMAGING

Prosecutors and investigators must be mindful that the ultimate goal of any investigation is to acquire evidence that will be admissible at trial. The creation of a copy of original electronic evidence raises authentication, best evidence and reliability concerns. How can one be sure the forensic imaging process produced a true copy of the original evidence? Could the forensic image have been altered or corrupted in the time between its creation and offering into evidence at trial?

A. Best Evidence Issues

Federal Rule of Evidence 1002 requires the use of an original writing, recording or photograph to prove the contents of those items, unless provided otherwise by federal statute or the Federal Rules of Evidence. FED. R. EVID. 1002. The exception that proves the rule for forensic images is Rule 1003, which provides that a "duplicate" is admissible to the same extent as an original unless a genuine challenge is made to the authenticity of the original or it would be unfair to admit the duplicate instead of the original. FED. R. EVID. 1003. Rule 1001(4) defines a duplicate as a copy of the original made by, among other things, "mechanical or electronic re-recording . . . or by other equivalent techniques which accurately reproduces the original." FED. R. EVID. 1004. Thus, the focus must be on whether the image is an accurate and authentic reproduction of the original evidence.

B. Authentication of Forensic Images

Authentication is a predicate to the admissibility of any physical evidence. *See* FED. R. EVID. 901(a). To satisfy Rule 901, the proponent must produce "evidence sufficient to support a finding that the matter in question is what its proponent claims." *Id.; see, e.g., United States v. Simpson*, 152 F.3d 1241, 1250 (10th Cir. 1998). This requirement is typically easy to satisfy when the evidence is a single document and a co-operating witness, such as a recipient, author or custodian is available to authenticate it.

While the authentication requirements for computer data are no different than for other forms of evidence, authentication can appear more daunting when the data was extracted from a *copy* of the defendant's media that was made outside the defendant's presence. Furthermore, due to backlogs in obtaining forensic analysis of seized computer media, it is likely that the copies of the seized media sat in on a shelf in an evidence room for months or years before trial. These factors, combined with the ease (perceived or real) of altering computer data without notice, may tempt a particularly aggressive defence counsel to challenge authenticity of the proffered data.

Courts have generally looked askance at authenticity challenges to electronic evidence that are unsupported by anything other than speculation that the original data was altered by an unseen hand. *See, e.g., United States v. Whitaker*, 127 F.3d 595, 602 (7th Cir. 1997) (affirming admission of computer records where allegation of tampering was "almost wild-eyed speculation . . . [without] evidence to support such a scenario."); *United States v. Glasser*, 773 F.2d 1553, 1559 (11th Cir. 1985) ("The existence of an air-tight security system [to prevent tampering] is not, however, a prerequisite to the admissibility of computer printouts. If such a prerequisite did exist, it would become virtually impossible to admit computer-generated records.") In *Whitaker*, the Seventh Circuit upheld a district court's admission of print-outs of spreadsheets from the original computer seized, where the FBI agent involved in the seizure and the printing testified as to their authenticity. *Id*. Despite the permissive standard applied in *Whitaker*, good trial pre-strategy is to foreclose potential authenticity challenges before they are raised.

IV. HASH ALGORITHMS – AN ANSWER TO EVIDENTIARY ISSUES

To blunt potential authentication challenges to data extracted from a forensic image, it is useful to have a procedure to verify that the data on the image is an exact match of the original media. Computer forensic specialists have developed a procedure that guarantees just that. This process uses "hash" algorithms, which verify that the acquired image is the exact copy of the original media. The most commonly used hash algorithms – the Message Digest 5 (MD5) and Secure Hash Algorithm-1 (SHA-1) – take as input a message of arbitrary length and produces as output an n-bit "fingerprint" or "message digest" of the input. The algorithm then produces a digital signature which can be used to identify uniquely a given file, and therefore establish that the image is an authentic copy of the original evidence.

Verification using hash algorithms is highly reliable. The odds of two random files having the same hash are astronomically small – estimated to be approximately a 1 in 10^{38} chance. Moreover, the use of the hashing algorithm is a one way function, which means that it is easy to create a hash from a file but almost impossible to create a file matching a particular hash.

Hash validation, when combined with evidence of a chain of custody between the time the original computer media was seized and the image was created, is strong authenticating evidence that the forensic image is an exact duplicate of the original. Hash algorithms fit the examples listed in Rule 901(b)(4) of "distinctive characteristics" that can be used to authenticate evidence. FED. R. EVID. 901(b)(4). What are hashes if not indicators of "internal patterns, or other distinctive characteristics" of data?

Although published decisions addressing the use of hashing algorithms to authenticate forensic images are few, they are uniform in recognizing hashes as a proper means of establishing authenticity. *See, e.g., Williams v. Sprint/United Mgmt. Co.*, 230 F.R.D. 640, 655 (D. Kan.2005) (recognizing that hashing "allows a large amount of data to be self-authenticating with a rather small hash mark, efficiently assuring that the original image has not been manipulated.") In *Williams*, the district court rejected a civil litigant's purported concerns about producing electronic evidence in its native format by noting that the parties could detect any alteration by comparing hash values. The court found that a hash value is a "'digital fingerprint' akin to a tamper-evident seal . . . the file cannot be altered without a change also occurring in the hash mark." *Id.; see also Ohio v. Morris*, 2005 WL 356801, No. 04CA0036, (Ohio App. Feb. 16, 2005) (admitting forensic image even where testimony established that imaging software had validated the MD5 hashes of the original and image matched before forensic examiner erased the original hard drive); *Krause v. State*, 2007 WL 2004940, No. 01-05-01136-CR, (Tex. App. July 12, 2007) (forensic analyst's methodology was sufficiently reliable for purposes of expert testimony where analyst used forensic software that compared hashes on the image and the original media). Similarly, the Federal Judicial Center has identified MD5 and SHA hashes as commonly used algorithms to establish the authenticity of a forensic image. *See* FEDERAL JUDICIAL CENTER, MANAGING DISCOVERY OF ELECTRONIC INFORMATION: A POCKET GUIDE FOR JUDGES, FEDERAL JUDICIAL CENTER (2007) at 24, quoted with approval in *Lorraine v. Markel American Ins. Co.*, 241 F.R.D. 534, 536-37 (D. Md. 2007)

V. STORING FORENSIC IMAGES – AN ALTERNATIVE TO THE SHELF

As discussed above, provided that proper chain of custody is established between the times the original computer media are seized and forensic images are created, the hash verification process should eliminate any concerns over whether the forensic image was altered prior to trial. However, the practical concern of how and where to store the forensic images remains.

While the prevailing method of storing forensic images is certainly adequate and relatively simple, it has its shortcomings as well. First, as anyone who has dealt with electronic evidence likely knows, hard disk drives fail. A recent study of 100,000 different types of hard disk drives conducted by researchers at Carnegie Mellon University found that the actual reported failure rate of hard disk drives is much higher than stated in manufacturers' data sheets. Bianca Schroeder and Garth A. Gibson, *Disk Failures in the Real*

World: What Does an MTTF of 1,000,000 Hours Mean to You?, FAST07, 5TH USENIX CONFERENCE ON FILE AND STORAGE TECHNOLOGIES (2007). Although the observed real world failure rates were approximately 2%-4% (with some as high as 13%) are still relatively low, no one wants request a continuance of trial because the hard disk drive on which the forensic image was stored failed. Moreover, frequent handling and transportation of hard disk drives inevitably jostles the sensitive mechanical parts in the drives and can only increase the potential for drive failure.

A more advanced and safer method of maintaining forensic images is to upload or copy the forensic image and hash, to a fault tolerant RAID. A RAID is a category of disk drives that employ two or more drives in combination for fault tolerance and performance. The entire purpose of RAID storage is redundancy – if one disc in the array fails, the data remains secure on one of the other redundant discs. Also, unlike a powered-down hard disk drive, a running RAID system can be configured to conduct routine backups to tape archives, which can be stored off-site. This is a useful data recovery backstop in the event of a disaster, such as a flood or fire at evidence storage location. Indeed, the implementation of secure RAID evidence storage appears to adhere to the National Institute of Justice's Office of Justice Programs recommendation that investigators preserve evidence "in a manner designed to diminish degradation or loss." DEPARTMENT OF JUSTICE OFFICE OF JUSTICE PROGRAMS, NATIONAL INSTITUTE OF JUSTICE, CRIME SCENE INVESTIGATION: A GUIDE FOR LAW ENFORCEMENT (2000).

Moreover, a RAID storage system would save space in crowded evidence storage rooms and simplify the process of locating evidence when it is requested. A RAID system would reduce the necessity of having shelves stacked with numerous individual hard drives, each containing images of media seized from different subjects. Forensic images could be stored in folders corresponding to investigation name and number, subject name, and search location, making it easier to locate desired images when they are requested by prosecutors.

When the time comes to use the image at trial, forensic examiners would copy the image back to a hard drive and verify that the hash is unchanged. Hash validation after the image is transferred onto the RAID will ensure that the image stored on and ultimately recovered from the RAID is no different from the original data that was seized. Because it would rely on the already approved hash validation process, a RAID-based storage system should not undermine the authenticity or reliability of the forensic image that is eventually offered into evidence at trial.

Just like any piece of evidence, care would need to be taken to keep the RAID in a secure setting, such as in a locked, limited access server room with no Internet connections. Logging software could be added to the RAID to keep track of access to the virtual evidence lockers stored on it, and forensic images could be stored in password protected virtual lockers on the RAID. And of course, testing should be performed before a RAID–based evidence storage system is put into use.

Prosecutors interested in these and other computer forensic issues and techniques may register for the Computer Forensics for Prosecutors Course taught by CCIPS at the National Advocacy Center. The Computer Crime and Intellectual Property Section and the Cybercrime lab is also available to AUSA's for consultation on computer forensic and other technical investigative matters by calling (202) 514-1026. Many other resources are available on our section's public website, www.cybercrime.gov. In addition, anyone in the Criminal Division or US Attorneys' Offices can find additional resources on our new intranet site, CCIPS Online. Just go to DOJ Net and click on the "CCIPS Online" link.