Data storage devices that may be examined for data recovery. Clockwise from top left: External 60 GB data store; Opened internal hard drive; Server rack; Floppy disk; 60 GB data store original by author; other images from everystockphoto.com
INTRODUCTION

Digital evidence is information stored or transmitted in binary format that is of potential probative value. Like many fields of forensic examination, the scope of digital evidence examination is very broad. Investigations may cover fraud, e-commerce transactions, Internet communications, identity theft, package tracking, online banking, and Internet searches.

Electronic devices may seem durable, but their data are not. Digital evidence is often quite fragile: It can be damaged, altered, or destroyed if handled improperly. As a consequence, unlike other types of investigation, the original evidence is preserved and the analysis carried out on a copy.

CENTRAL QUESTIONS

What can be answered:

- Can the file be accessed?
- Can the file be recovered?

What we cannot answer:

- How can data be recovered from overwritten data storage locations?
- Who performed a specific action at a computer?
What we are learning or researching:

- Can a program automatically reconstruct file fragment data into intelligible files for review?

**ANATOMY OF A COMPUTER**

Most people are familiar with computers at some level. *Hardware* is the physical equipment needed to run a computer, including the central processing unit, keyboard, monitor, and any storage devices. *Firmware* is a set of instructions or data programmed directly into the circuits of a machine for the purpose of controlling a hard drive or other electronic component. Of course, *software* consists of programs. To better understand the process of digital evidence analysis, it is important to have a basic understanding of computer architecture.

**Microprocessor–Central Processing Unit (CPU)**

The microprocessor of a personal computer consists of a single integrated circuit, an array of transistors, resistors, and other microelectronic components embedded on a silicon wafer. Each dime-sized piece of silicon may contain several million transistors. The type of CPU in a computer determines how fast it can execute instructions, often measured in floating point operations per second (*flops*). Most computers today can execute millions of instructions per second. (See Figure B.1.)

**Hard Disk (HD)**

While the microprocessor is excellent for performing calculations, it has almost no capacity for storing information. This job is performed by the hard disk. While not much larger than a floppy disk, a hard disk contains more information and spins faster. This makes it an important component for the forensic examiner. (See Figure B.1.)

**Floppy Disk**

A floppy disk contains less information than a hard disk. Also unlike the hard disk, the floppy disk can be removed from the computer. The term floppy refers to the disk of plastic on which information is magnetically recorded.

**Compact Disc (CD)**

As a storage media, the popular CD is already becoming obsolete. A compact disc may contain the information of as many as 450 floppy disks. (See Figure B.2.)

**Digital Video Disc (DVD)**

The larger capacity of DVDs makes them the new trend in media storage; a DVD can hold a minimum of 4.7 gigabytes of data.
**Flash Drives**

There are several types of external data storage that depend on data being stored on a printed circuit rather than written to the track of a disk. These range from small units about the size of your thumb (and so sometimes called *thumb drives*) with a capacity of 250 MB to a few GB. More recent variants offer extensive external data storage in compact units that interface with the host computer through a USB or FireWire port. (See Figure B.3).

**Hard Drive**

The hard drive stores and retrieves digital data on a magnetic surface. A *write head*, a special type of antenna, “writes” information to the hard disk by transmitting an electromagnetic flux that changes the polarization of the magnetic media. These changes can be “read” by a *read head* as the changes in the magnetic field are detected by a coil as it passes by. Often these functions are combined into a read–write head. The read–write head floats on a cushion of air nearly in contact but just above the surface of the platter. Because the hard drive is not airtight but requires a narrow range of air pressures in which to operate, it is equipped with a permeable filter that prevents dirt, dust, and other objects from getting between the head and the disk. If the head crashes into the disk, it almost always causes a data loss.

The hard disk is divided into subunits, creating a logical structure for the file operating system to access. The subunits, called *partitions*, are areas of hard disk that can have an independent file system. A *primary partition* serves as a container for itself, while an *extended partition* may contain more than one logical partition. Figure B.4 shows how physical sectors on a hard drive are logically grouped into clusters.

The hard drive is accessed by a bus (*bus bar*), an electrical conductor that serves as a common connection point between two or more electrical circuits. Several types exist.

**SMALL COMPUTER SYSTEM INTERFACE (SCSI)**

SCSI (pronounced “scuzzy”) is a standard interface and command set for the transfer of data between devices on a computer bus. It can be used to connect almost any type of device. (See Figure B.5.)
FIGURE B.6
FireWire is another interface system, capable of very rapid data transfer. This image shows firewire and USB ports on a laptop computer. From everystockphoto.

FIGURE B.7
External CD drive with SCSI connector and terminator. Storage devices come in many shapes and sizes. A feature that all external drives share is that they have to have some means of interfacing with the computer in order to transfer data back and forth. Here we have an external CD drive that uses a widely accepted format for the interface, namely the Small Computer System Interface, or SCSI. From http://www.everystockphoto.com/photo.php?imageid=2945111. License http://creativecommons.org/licenses/by/2.5/.

FIREWIRE
Developed by Apple Computer, this interface is similar to the IEEE standard 1394. It is commonly used with data storage devices and popular in digital cameras and audio systems. FireWire allows for peer-to-peer communication between devices without having to access the system memory or the CPU. It is faster than a universal serial bus (USB), can supply power (up to 45 watts), and allows for hot swapping. (See Figure B.6.)

UNIVERSAL SERIAL BUS (USB)
The USB system is an asymmetric design in which a hub is used to link multiple devices to a host controller. When a device is added to the hub, the controller loads the driver necessary to operate it. USB is employed in gaming devices and to connect peripherals to computers; however, monitors do not use USB because they require a higher rate of data transfer. As of January 2005, the USB specification is version 2.0.

WRITING TO CDs AND DVDs
A CD is a plastic plate covered with dye, a layer of aluminum, and, finally, the label. As the CD spins, a low-power laser is projected at it. The laser that “reads”
the CD looks for light to be reflected from the aluminum surface into an optical sensor. If the light reflects back to the sensor, the reflected light is detected as a 1. If the light is reflected away from the sensor, no light is detected and the machine records a 0. The sensor works with the motor to find the speed at which the laser should fire. To write on a CD requires a slightly higher power laser than that used for a reading laser. The frequency of the laser is such that, when the light shines on the dye, the dye is turned opaque.

A CD-RW uses slightly different technology. A phase shift compound is sandwiched between layers of dielectric materials. The laser writes to the CD-RW by heating that material to its melting point, at which it turns opaque (about 600°C). The material remains opaque as it cools. To make the material translucent again, the material is heated but to a lower temperature (perhaps 200°C). This time, as it cools, it will again allow the beam to reach the surface of the aluminum below. Unlike a hard drive, where space can be made available by labeling it as free, a CD-RW must be erased before it can be overwritten.

**TYPES OF EVIDENCE**

Digital evidence is processed in three categories: computer forensic, digital audio, and digital video examinations.

**Computer Forensics**

Evidence in computer forensics may take on a wide variety of forms, including computers, hard drives, and storage media. Today storage media can be presented as diskettes, tapes, compact discs (CDs), digital video discs (DVDs), or Zip discs. With the boom in consumer electronics, forensic examiners may also be called upon to investigate cell phones, two-way pagers, cameras, global positioning satellite (GPS) devices, fax machines, thumb drives, and computer gaming equipment. As new technologies emerge, the field must grow to encompass them.

**Digital Audio**

Digital audio files are most often found on tapes, CDs, DVDs, or drives.

**Digital Video**

Digital video files are often found on a variety of tape formats, CDs, DVDs, or drives.

Digital evidence has some things in common with questioned documents. Like evidence on paper, digital evidence may be altered or obliterated.

**Erasure**

Digital evidence that is “erased” is actually just placed in a folder on the computer, which marks it as “erased” by making the space available again. The data may not actually be overwritten for some time. However, the longer the data sit...
in an unallocated sector, the more likely they are to be overwritten or partially overwritten. If the data are partially overwritten, some of the original data may be recovered as a file fragment.

**Overwriting**

Overwriting is the digital equivalent of obliteration. Sometimes the original data may be able to be partially recovered. Unfortunately, like writing over a document with the same ink used to create it, this is usually a very difficult task.

**Wiping the Drive**

When done with professional software programs, overwriting can be used to “wipe a drive clean.” One way to accomplish this is to write a pattern of data over and over until the disk is full and then declare all the space free.

**COLLECTING AND PRESERVING DIGITAL EVIDENCE**

While many people think of computers as permanent, the data stored on computers can be as fragile as a tire print in snow. Further, unlike other forms of examination, one of the most important techniques for examining digital evidence—to ensure the evidentiary data are not corrupted—is to work with a copy. The process of collecting and preserving digital evidence is sometimes called acquisition.

**At the Scene**

Once the scene has been properly secured, the investigator may look for evidence. The number and type of computers present should be documented as well as whether or not they are connected by a network. Any removable media must be cataloged, packaged in crush-proof containers, and properly labeled. The investigator must also look for any off-site storage areas or remote computing locations where evidence may be hidden.

For most computers, the best practice is not to shut down the machine but rather to pull the plug and take it to the laboratory. The exception to this type of procedure is business servers, which must be shut down according to their protocols. Hard disks should be parked and padded so that they are not damaged during transport.

Data that may be lost if the battery is removed from an electronic device is called volatile memory. Depending on the type of device, evidence may be handled differently. Small electronic devices and electronic storage media should be wrapped in bubble wrap or otherwise secured from static electric discharge. The investigator should also ensure that the antenna is off on items such as Blackberries and cellular phones so that the device does not communicate out.

The evidence should be labeled on the outside. All seized materials should be stored in a cool, dry place away from magnetic fields. Any magnet has the potential to cause damage. Moving the magnet around will cause more damage.
Preserving Digital Evidence

Before the evidence may be submitted for analysis, several additional steps must be taken.

- **Write-block**: To prevent any of the information on the original evidence from being changed, the media will be write-blocked by placing it on a special machine (examination computer) or by linking it through a device that will prevent anything from being written on the evidence.

- **Imaging (copying)**: Imaging is the process of making a copy of the evidence. The image becomes a working copy master that is archived and copied for analysis.

- **Archiving**: The working copy is saved to stable media: data tape, hard disk, CD, or DVD. The purpose of the archive is so that the original evidence may be set aside. If the working copy is corrupted, a new working copy can be restored from the archive.
FIGURE B.8
Illustration of hashing. Hashing is the application of a mathematical function to convert data files into an integer that serves as an index for the data and that will change whenever the data are changed. From http://en.wikipedia.org/wiki/Image:Hash_function.svg

Operating a computer after seizure will overwrite some of the data and may change the date–time stamp information. In some cases, the investigator may also send a “preservation order” to the Internet service provider to obtain e-mail and other information about the account.

AUTHENTICATING DIGITAL EVIDENCE
To ensure that the original data are represented by the working copy, both are compared by an authentication process called hashing. The data are examined using a special program that takes in the complete, unique binary pattern of the evidence—trillions of bits!—and summarizes it into an output with a finite number of elements. This output is called the hash. The examiner must calculate the hash for the original data and for the copy. They should match. The bigger the hash, the greater the certainty that the original and the copy are the same when compared. If they do match, the copy is considered authenticated for additional work. (See Figure B.8.)

Several commercial programs are available:
- MD5: MD5 has a hash size of 128 bits. It is one of a number of message digest algorithms created by Massachusetts Institute of Technology’s Ronald Rivest. As cryptographers exposed weaknesses in MD5, its use was replaced in part by SHA-1.
- SHA-1: Developed by the National Security Agency (NSA) and published by the National Institute of Standards (NIST), SHA (Secure Hash Algorithm) is a family of hash programs. SHA-1 takes longer to calculate than MD5 but has a higher level of accuracy.
- SHA-2: NIST published four variants of SHA, which are collectively called SHA-2. The programs are named after the bit length of their digests: SHA-224, SHA-256, SHA-384, and SHA-512.

ANALYZING DIGITAL EVIDENCE
Once the working copy is authenticated, a variety of techniques may be used to search the evidence. The presumptive test, often called triage, is to write-block the original evidence and browse–key word search it to see if it is of potential probative value. The confirmatory tests on the working copy include a detailed or multiple keyword search, browsing, metadata recovery and analysis, file fragment recovery and analysis, and looking at the time line (user logs, system logs).

Techniques that may be used to search the evidence after the working copy is authenticated include the following:
Browsing: The examiner opens the files and looks at the content.
- Key word search: The examiner queries the drive with key words to identify documents with key words pertinent to the investigation. A running list of the words used will be documented. Because some criminals may use code words, the investigator may also look for numerical data.
- Metadata search: By examining the information about the data files, the metadata, investigators can see when the document was created, sent, or received. This may assist in the development of a time line for a crime or establish an alibi.
- Automatic log searches: These files, created by the computer programs as records of activity, show what people are doing on their accounts. For example, while DSL accounts are active all the time, files will be date–time stamped when moved or sent.
- Internal clock search: The setting of the computer’s internal clock is also reported as it weights the value of the date–time stamp. The examiner will note the time and time zone to which the computer was set on the day the investigation was performed.

Taken together, the analyses attempt to attribute ownership of the crime via software registration, e-mail accounts used, leads to credit cards, file privileges, user access, and profiles on the computer.

More specialized measures may need to be taken if there are hardware issues associated with digital evidence. Hard disks can be physically destroyed, but attempts can be made to secure whatever data are available. When a hard disk is crashed, the antenna may be sitting in contact with the silicon wafer. The controller card is take off the hard disk and replaced or adjusted until the disk can spin again. For burned disks, after swapping the controller card, the damaged evidence may be taken to a clean room for reconstruction. If a floppy disk has been cut up, the examiner will remove the jacket surrounding it and examine it under a microscope. The examiner will line up the tracks using a magnetic oxide solution, tape the pieces together, rejacket the media, and try to examine it again.

**Legal Issues in Searching Electronic Media (Privacy, Search and Seizure)**

Searching electronic media is not unlike searching a house. In order to conduct a search on the working copy, the forensic computer examiner needs a warrant for the working copy that sets the scope of the investigation. On a warrant looking for information about drugs, if information about child pornography were discovered, this evidence would require a new warrant. This can be very limiting in certain circumstances. In addition to state and local regulations, a complex set of laws governs electronic media. These include the Electronic Communications Privacy Act of 1986 (ECPA), the Cable Communications Policy Act (CCPA), and the USA Patriot Act of 2001.

Another important law governing digital evidence is the Digital Millennium Copyright Act (DMCA) of 1998. Congress passed this law as part of the copyright
act in response to pressure from creative content providers such as the motion picture industry, recording industry, gaming industry, and software publishers. The law has provisions both for violating copyrights and for creating or using technology that will crack protection measures on copyrighted material.

PROBLEMS

1. Give the word or phrase for the following definitions:
   a. a set of well-defined rules for problem solving; a computer program
   b. an operating system loaded in from media rather than from the hard drive
   c. using an algorithm to summarize the binary pattern of data in an object
   d. a file structure or operating system
   e. a portion of a file that may have been partially overwritten
   f. a file that contains data about an object
   g. a portion of the hard drive space
   h. a one-to-one correspondence copy
   i. the difference between the logical end of a file and the end of the last allocation unit for that file
   j. a part of the disk marked by the operating system that cannot be written to by the user

2. Define digital evidence and list examples.

3. Explain the difference between archive copy, image, physical copy, and duplicate.

4. Give three examples of storage media.

5. Describe the process by which the computer writes to the hard disk.

6. Describe the process by which a computer writes to a CD.

7. Define volatile memory and then list several devices that have it, as well as procedures an examiner should take to preserve the evidence when collected.

8. Describe a procedure for seizing evidence at the crime scene.


10. Describe a procedure for authenticating digital evidence using hashing.

11. List several methods for analyzing digital evidence.

12. Why must the antenna of a volatile memory device be turned off on seizure?

FURTHER READING

Electronic Communications Privacy Act (ECPA). 18 USC 2510 et seq.; 18 USC 2701 et seq.; 18 USC 3121 et seq.


Privacy Protection Act (PPA). 42 USC 2000aa et seq.


USA Patriot Act of 2001, Public Law 107-56, amended statutes relevant to computer investigations. Statutes amended include 18 USC 1030; 18 USC 2510 et seq.; 18 USC 2701 et seq.; 18 USC 3121 et seq.; and 47 USC 551.

GLOSSARY

Archive copy a copy of the original evidence preserved; the archive copy can produce working copies.

Algorithm a set of well-defined rules for problem solving; a computer program or set of computer programs designed to systematically solve a certain type of problem.

Best evidence evidence, such as that taken from a network, that may be only a “snapshot” of the information; used when the examiner does not physically have the computer or other evidence.

Control boot disk a disk that overrides a computer’s regular start-up programming by replacing it with a specific set of commands or instructions; an operating system loaded in from media rather than from the hard drive.

Copy a reproduction of information contained on an original evidentiary item, independent of electronic storage device. While a copy maintains contents, the attributes may change during the copying process.

Digital evidence information stored or transmitted in binary form that may have probative value.

Duplicate (n) a digital reproduction of all the data contained on a digital storage device, a one-to-one copy; (v) the process that maintains contents and attributes. The term duplicate is a legal standard meaning “exactly the same” and is more strict than the term copy.

Firmware a set of instructions or data programs placed directly into the circuits of a machine for the purpose of controlling a hard drive or other electronic component.

Fragment (file fragment) a portion of a file that may have been partially overwritten; also, this space is allocated to an amount of data too small to require a full block.

Hardware the actual components of the computer.

Hashing using an algorithm to look at the binary pattern of data in an object to produce a summary of what is there.

Image a file that contains data about an object.
Image file when making a copy of a hard disk, the image file contains all the data as well as the attributes (often including metadata); thus, images can be restored to make an exact duplicate of the evidence.

Imaging making up a copy of a hard disk or other media.

Log files files that record actions taken by a computer program or user’s account; experts may be able to determine what a user was trying to do by interpreting log files.

Logical copy a one-to-one copy of a partition, directory, or file on the hard drive; the copy made by “drop and drag” operations.

Logical data a file structure or operating system; it provides information such as the file name and content.

Logical image information about a partition, directory, or file on the hard drive.

Logical partition a partition within an extended partition; it exists only inside the logical structure of the software.

Lost cluster a portion of a file for which there is no descriptive linking information.

Malware programming designed to do unwanted operations on a computer.

Metadata information about a file or other document such as the date and time stamp, file access privileges, or other information.

Metafile a file describing specific information about another file.

Original digital evidence evidence collected from a crime scene.

Partition a portion of the hard drive space.

Physical copy a duplicate; a one-to-one correspondence copy; a bit copy.

Physical data fundamental storage units; it provides less information than logical data. The information given may be only that at sector X, some form of data exists.

Physical image file that contains all the necessary information to reconstruct a bit-for-bit file.

Reserved cluster a part of the disk marked by the operating system that cannot be written to by the user.

Slack space (file slack) when writing a file to a hard disk, the difference between the logical end of a file and the end of the last allocation unit for that file. For example, if a file contains 50 bytes but the allocation size is 512 bytes, there would be 462 bytes of slack space.

Software programs for a computer to execute.

Steganography communication that hides the existence of the communication; hiding one file within another file.
Unallocated space (free space) allocation units not assigned to files within a file system.

Virus an executable malware program.

Volatile memory memory that would be lost if the battery or power supply were removed from a device.

Working copy the copy on which analysis is performed; if the working copy becomes corrupted, a new copy of the evidence can be created from the archive.

Worm a type of malware that can change over time to evade detection.

Write blocker hardware or software that prevents writing to a media.

**STEGANOGRAPHY**

Steganography comes from the Greek meaning “covered writing.” It is the art and science of concealing information. Note the difference between concealing and disguising a message: Writing may be concealed by using lemon juice as ink, while writing may be disguised in a picture. In the realm of digital evidence, this process becomes hiding data in other data. For example, child pornography may be hidden in other pictures, audio files, or written documents.

**DISK OR DISC?**

When is a disk a compact disc? The spelling disk is most often used and preferred for storage media associated with computers. On the other hand, disc is the preferred term for round, flat objects like the ubiquitous optical storage media used for audio, data, and digital video.
**BOOBY TRAPS AND MALWARE**

Sometimes the investigator must ascertain if the user is aware of a program on a computer. In other cases, booby traps and malware—such as worms and viruses—may be left behind on discs or hard drives to hinder the investigator. These programs may engage in selective wiping of data operating on start-up or shutting down of the computer. To avoid such damage, the examiner will make a copy of the data and attach it to a forensic analysis program. The examiner’s machine will be the one running the operating system, so the offender’s programs will not be executed.

There are several ways to look for malware. The easiest is to use commercial software to scan a duplicate of the evidence. Because an image file cannot be scanned, the investigator may resort to scanning the findings recovered from it. One way is to look for the digital signature of known viruses as compared to those in a database. Other methods are more heuristic. The examiner must look at the machine-level calls that were made. Write calls to the hard drive are suspect. Because viruses are now increasingly communications programs, the examiner may also look at the computer code of executable programs, opening of computer modem or network connections, or calls to certain ports.

**CRASH**

In computing, a crash occurs when an application or part of the operating system stops performing its function or stops communicating with other parts of the computer system. Programs that crash may simply stop operating and appear to “freeze.” However, if the program is part of the operating system, the entire computer can be affected.