Informing Science Special Series: Expanding the Focus Volume 4 No 2, 2001

An Introduction to Computer Forensics: Gathering Evidence in a Computing Environment

Henry B. Wolfe
University of Otago, New Zealand

hwolfe@commerce.otago.ac.nz

Abstract

Business has become increasingly dependent on the Internet and computing to operate. It has become apparent that there are issues of evidence gathering in a computing environment, which by their nature are technical and different to other forms of evidence gathering, that must be addressed. This paper offers an introduction to some of the technical issues surrounding this new and specialized field of Computer Forensics. It attempts to identify and describe sources of evidence that can be found on disk data storage devices in the course of an investigation. It also considers sources of copies of email, which can be used in evidence, as well as case building.

Keywords: forensics, evidence, investigation, data recovery.

Introduction

Forensic evidence gathering techniques in criminal investigation has a long and established history. The personal computer has become one of the most important tools in the twentieth century. Computers play the vital role of information and data manipulation, storage and retrieval in just about every business sector. We use computers to store sensitive, proprietary and confidential information and data. We use computers to communicate over the Internet both locally and on a worldwide basis. We use computers to analyze complex relationships and for advanced (and simple) mathematics. As might be expected, this powerful tool can, in addition to its obvious positive uses, be used for illegal purposes. The age of the Internet has highlighted the importance of the development and use of forensic techniques as applied to the gathering of electronic evidence.

These techniques can and are being used in a prosecutorial or civil environment to obtain additional supporting evidence and in many cases, primary evidence to make the case for one side or the other.

The following material is intended to brief the reader on various techniques and tools that can be used in investigative and defensive modes and in some cases in both ways. Information technology has changed the way in which investigators are informed about evidence and some of the methods that are used to gather and analyze it. As the Computer Age progresses, this information and these techniques will play an increasingly more important role in the legal process and while some techniques may be replaced by others the focus on forensic computing will, never the less, increase.

Discovery/Offensive Computing

The phrase Offensive Computing sounds a bit ominous, however, as time goes on I believe that this area will surely gain in importance. In order for you to take advantage of advancing techniques you need to know about them. The term, in this context, refers to the use of computers in the discovery and investigative process (both overt and covert) and for gathering evidence in computing related legal cases.

Information stored on the computer(s) of a suspect, usually with the permission of the Court, can be viewed and analyzed. In a covert or hostile environment, however, there will be no authorization and cooperation that might otherwise be directed by the Court. There are a number of products currently available that enable the investigator to create evidentiary documentation from a target computer. This information is gathered using specially designed software and sometimes hardware as well. An example of an active surveillance covert tool is Data Interception by Remote Transmission (also known as ***D.I.R.T*** **,** produced by Codex Data Systems, Inc. of

Bardonia, New York (www.codexdatasystems.com). It is a controlled tool and regulated (in the US) by *Title 18 USC 2512*.). This program is much like *BackOriface* and *NetBus* (see Glossary), which are installed on a target machine from the Internet (usually from a Trojan) - no physical access to the machine is necessary. ***D.I.R.T.*** facilitates remote monitoring by the investigator. This tool can pick up keystrokes and captured data, which is then transmitted, via the Internet, to the monitoring station. It can transmit video and or sound as well - if a video camera or microphone is present on the target machine. It can capture screen images and access all files present on the target machine as well as run target machine resident programs. Presumably, this tool could be legally used when a law enforcement agency, with the required preliminary evidence gains an Order of the Court to place a target machine under surveillance for a specified period. In practice, however, it is probably more likely that the ***D.I.R.T.*** generated evidence may be discovered before the Court Order is issued. This kind of surveillance tool is easily abused and because of the public's lack of awareness that it even exists, few countermeasures have thus far been developed. That means that anyone can be the target of a trolling operation mounted by persons with such tools - be it law enforcement or private.

When it is possible to have physical access to the target system potentially damaging evidence may be obtained from the areas known as unallocated space (erased files), Windows work space or swap file, slack space, bad blocks (or sectors), extra tracks and other transient storage areas. When the target machine contains files that have been encrypted using programs like Microsoft Word or Excel, or WordPerfect and others the investigator uses readily available tools to derive the passwords for those files, documents those activities and decrypts the files that are encrypted. Directories, File Allocation Tables, and associated information (like names and types of files, dates and times those files are created or modified, etc.) are examined and analyzed to help build time-line profiles.

The information gathered from these techniques is stored on a separate medium and can be searched for key words or phrases relative to the case and relevant findings documented for later use. Most investigators recommend the creation of a CD-R containing a bit by bit mirror copy of the hard disk being scrutinized, however, that process may require several CD's to be created with the larger hard drives. CD's normally cannot be easily modified and can retain the exact information copied - this is important for evidentiary purposes and subsequent analysis. In the case of very large hard drives, tape backup may be more practical (using a Vogon Imager for example). The original target disk is NEVER used directly in any way that would enable ANY modification of it to take place. This can be accomplished by booting from a specialty disk and copying the hard drive through a parallel cable to the destination or evidentiary medium for later analysis.

Evidence Acquisition

Before any action is taken it is the responsibility of the investigator to observe the normal scene preservation and evidence gathering techniques. The location should be photographed from various angles and an inventory of relevant items should be created as the normal part of the case documentation. However, this paper is not intended to address that investigative process other than to remind the reader of its importance.

The investigator is generally faced with two possible scenarios. The first is that the target machine is currently switched on and active. If this is the case, then a DOS session should be opened and the backup process initiated. If possible, a copy of the current state of RAM should also be stored for future analysis. Passwords and other key information may be found here. The second scenario is that the target machine is switched off. In this case, the target machine should be booted up from a bootable floppy disk **after** a physical inspection to ensure that the floppy drive is connected **and** that it is the primary bootable drive (that information is found in the BIOS settings). To ensure that the floppy is the primary bootable device, the target hard drives are disconnected while ascertaining and possibly resetting the BIOS so that the floppy is the primary bootable device. The backup process will be initiated from the floppy booted DOS session. In either case the following are a few simple steps to the forensic process:

Evidence Gathering Sequence

The most important issue in forensic evidence gathering with respect to computers is the preservation of the original data. To that end, the first task at hand is to create a mirror copy of the hard disk somewhere. That may be on another hard disk, many floppy disks, one or more CD's, or tape, etc. Prior to doing this the target machine should be physically inspected (while turned off) to ensure that there are no physical devices in place that might damage information contained on the hard drive(s). To ensure that the mirror copy is accurate a CRC or hash value for all individual files or sub-directories and/or for the entire contents of the hard disk must be created from the original and matched to the corresponding CRC(s) or hash values created, in the same way, directly from the mirror copy - thus certifying its authenticity, accuracy and completeness. Most forensic backup tools have this as a feature (for example *EnCase*, a forensic product of Guidance Software, Inc., http://www.guidancesoftware.com/ and *SafeBack*, a forensic product of Sydex Corporation, Inc., http://www.sydex.com/sbqa.html, etc.). This entire process should be formally documented with times, dates, signatures of investigators and consulting experts involved, a detailed technical description of the target machine and any other information necessary to validate the activity. That documentation should become a part of the final case documentation.

Analysis of Target Storage Device(s)

A thorough analysis of open, unknown and hidden (as defined earlier in this document.) data residing on the target device (actually its certified copy - the original should not be used for analysis) should be made and the results of specific searches and search parameters should be formally recorded (for evidentiary purposes) and copies of relevant material printed for inclusion in the case file.

If specialist expertise is required, they should be supervised and their activities and results formally recorded (for evidentiary purposes). The written results of the analysis process will constitute the formal case documentation as it pertains to whatever may be found on the computing storage devices under investigation.

Case Preparation

It is the investigator's responsibility to maintain the chain of evidence as well as present the information gathered in a logical, professional and accurate format - as required by the specific jurisdiction. What is found may or may not support illegal activity. It is not the investigator's responsibility to suggest intent but rather to present the facts and certify their authenticity so that the truth may be found. It is, thereafter, up to the prosecutor, attorneys, judge and jury to consider any relevant evidence found as it applies to the specific case.

(***A useful reference for case preparation**:* Rosenblatt, Kenneth S., High-Technology Crime Investigating Cases Involving Computers, San Jose, KSK Publications, 1995, ISBN: 0-9648171-0-1.)

Investigating the Contents of a Hard Drive

There are three basic types of information normally or potentially stored on personal computers:

Open Information/Data (easily available to anyone with access).

Open information consists of operating systems executables and relevant data, configuration and temporary files; user applications software (like word processors, spreadsheets and so forth) and their associated configuration, data and work files; and user generated data (documents and files generated by and associated with the various applications residing on the user disk). This includes specialty files like graphic files, programming source code, etc. This may be a source of valuable evidence and would normally be scrutinized first.

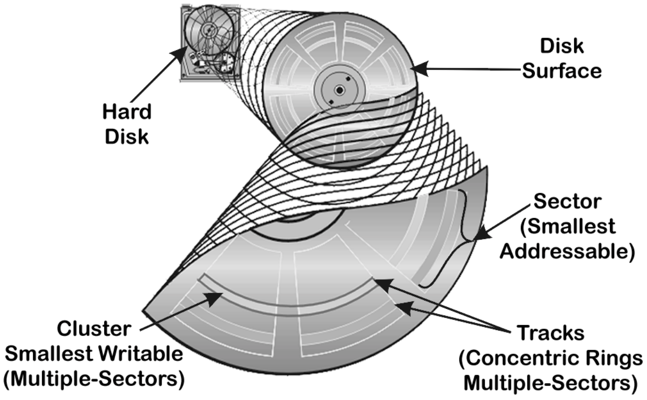

Figure 1: Disk Storage Illustration (Source: author)

Unknown or Potentially Unknown Information/Data

PC operating systems, in order to provide input/output (I/O) services for resident applications, perform their tasks in a well-documented, structured, and specific sequence. As a result, there are certain areas on a hard disk that may contain information that most users are or may not be aware of. Examples of these are deleted files, Windows temporary workspace, slack space, and RAM slack (see Glossary). Any or all of these areas may contain evidence relevant to an investigation.

It may be useful to discuss how PC's manage disk space here. The smallest piece of data ever written to a hard disk is not a bit nor is it a byte (see Figure 1). The smallest piece is actually a cluster. A cluster may be as small as a single sector (or 512 bytes) or it may be much larger in multiples of sector length in larger disks. The reason is that I/O is the slowest operation within computing and in order to manage this process efficiently, data are temporarily stored in buffers until the buffer is filled or until the last piece of data is placed there. At that point the entire cluster is written to hard disk. This technique reduces the actual number of physical writes to a more manageable level and thus speeds up the I/O process. When writing a cluster to disk the original cluster is loaded into RAM. The new information is moved into the original cluster beginning at the first byte. If the new information to be written does not fill the entire cluster, whatever was there before in the unused portion remains and the cluster with the new information and the residual data from what was there before is written to disk. That residual information or data is referred to as slack bytes (see Figure 2).

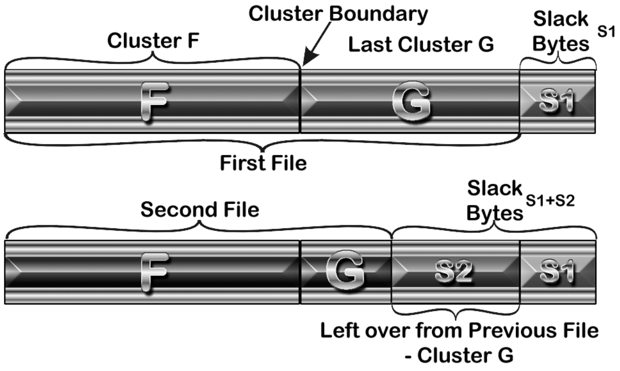

Figure 2: Slack Bytes Example (Unknown Data)
(source: author)

Hidden Information/Data (deliberately hidden or disguised)

Some information may be deliberately hidden in a number of places and in a number of ways in order to raise the level of difficulty required to find, access and/or interpret it. In and of itself, this activity does not indicate guilt or innocence, however, it should raise a red flag for the investigator and care should be taken in the handling, processing and documentation of such activity. A deliberate way to hide data/information is by the use of encryption. Files that are encrypted by certain software will have their own signature and may be detected by that signature. An example of this is the use of PGP (see Glossary) and other commercially available cryptographic software.

The first thing that many investigators want to do is find someone who can “break” the code. This usually ends up in frustration and wasted time. Few of the established cryptographic systems currently available in the marketplace can be “broken”. In a couple of recent cases that I was involved with both suspects made legal use of cryptography to protect their privacy (both were unwilling to provide their keys). In neither case was there any attempt made to “break” the code. Instead, in the one case, I analyzed the contents of the evidentiary hard disk and found six key candidates (the third was successful). In the second case, the suspect refused to hand over his keys stating that he had forgotten them. In that case I considered at least six alternatives for key acquisition (the use of a rubber hose was not amongst these alternatives) and one additional hardware alternative. The choice was made, a warrant was obtained and the plan was put in place. Within two hours, of returning the suspect’s computer, the keys were known to the authorities (subsequent indictments resulted). While cryptographic protections can be put in place by any user, in many cases (but not all) it is still possible to gain access to that encrypted information through the use of means other than cryptanalysis (this case was a police matter - the methods employed have not been disclosed for that reason).

If strong encryption is being used the one way to find out the contents of an encrypted file(s) may be for a Court of Law to order that the owner make his/her key available to the Court or its agent. It is important to mention that often, passwords and cryptographic keys are temporarily stored in the clear (in a readable format) within the Windows swap file. Therefore, the analysis of the Windows swap file may generate extremely important evidence as well as the tools to decrypt files that may otherwise be incapable of decryption without the suspect's cooperation.

Other examples of hiding places are partition waste space, bad sectors, extra tracks (see Glossary for definition of terms), changing the partition table such that a volume is hidden from the operating system and other spaces on disk not normally used. Information found in these areas should probably be investigated further since the user obviously felt that their data/information was important enough to take the trouble to hide it. An example of disguising data is changing the name of a file and its descriptive extension (.DOC, .JPG, .GIF, etc.). Most file types have a distinctive signature, however, and these can easily be checked (some forensic products provide a feature that highlights those files whose extension does not agree with its identified signature - an example is *EnCase*).

Information may also be routinely hidden within various graphics, audio and/or video files (.JPG, .WAV, .AVI, etc.). Steganography is the art or science of hiding data or information within other data or information. There are many software packages (*Invisible Secrets* for example) that accomplish this task freely available on the Internet. The information hidden may or may not be encrypted prior to being hidden. It is a standard forensic procedure to look for this specialized software on the target machine, as an indicator of this type of activity. If the target machine appears to have this type of activity present then further expertise might be required. Finding such applications would be an indication that they might have been used to protect the privacy of certain information/data.

Email

For those who make use of the email function today, it is easy to assume that once you have sent your message and deleted the file from your hard disk that the only other person who has a copy of that message is the recipient. That could not be further from the truth. When sending an email message, it will first travel to the server that the sender is connected to (where a copy is made) then through the Internet Service Provider (where a copy is made). From there it travels through a myriad of other computers (where, at each relay, a copy is made) before it arrives at the ISP of the receiver (where a copy is made) then to the receiver's server (where a copy is made) and finally to the receiver (where a copy is made). As you can plainly see, copies are made at many stops along the way and the message never really disappears. At any of the various junctures along the communications channel the copy may archived for permanent or semi-permanent retention.

What this means to us is that for purposes of the discovery process, it may be possible to retrieve old email messages even though none may reside on the accused's computer(s). This process is even now being refined and we may see in the near future the capability of retrieving such messages (for a price). So, there are two things to keep in mind about email. The first is that if you're going to use it to your benefit in the most secure way you should make use of encryption wherever the content warrants. If that is done using strong encryption then no matter how many copies exist they will be of no use to anyone but the person(s) who have the appropriate key to decrypt. The second thing to remember is that the accused may not have thought of this and you may be able to make use of the discovery process to retrieve pertinent evidence in the form of previous email messages.

Email Evidence

Local Machine

The target machine(s) may have files containing incoming email messages in addition to copies of the individual's outgoing message(s) that contain other possibly useful information. This includes dates, times, names of parties to the message, subject, and routing of the message. This routing information may be of particular interest and point to where other message traffic might reside (as a result of the archiving of servers along the various paths the messages take).

Local Servers

Local servers may have archives of email messages that originated from the target machine but have already been overwritten or destroyed by the person being investigated. Investigating the local servers and their archives may require specialist software tools (and appropriate Court Orders) to enable the search of their archive storage.

Hamilton, Ontario Case

Philips Services, Inc., an Ontario corporation was successful in obtaining, from an Ontario Court, an Order granting discovery that covered ISP's who may have relayed information in the form of Chat Room sessions and Email between certain parties. Whether this precedent setting case would have any influence in countries other than Canada remains to be seen but it is certainly worth knowing about.

Conclusion

Computers are an important part of society today and many people use them as routinely as phones were used in the past. The computer is, after all, just another tool. The Industrial Age brought us the opportunity to significantly extend our abilities to produce. Those things formerly produced by hand are now manufactured, in quantity, by machine. This net effect can be equated to an extension and enhancement of our body’s physical capabilities to be able to manipulate more, faster and with greater precision.

The Computer Age has brought with it the ability for each of us to extend our intellectual capabilities by enabling improved information storage and manipulative capacity. This can be equated to an extension of our individual intellect, memory and evaluation processes – our mind. However, what you think can be kept private by not exposing it to anyone – controlled by the individual. What you type into your computer, however, cannot be kept private in the same way. This fact makes it possible for the gathering of forensic computer evidence. This paper has introduced the reader to this new and important technique and outlined various topics that demonstrate ways and means to undertake an electronic evidence investigation.

Sources of Additional Information

(NOTE: This is a relatively new field. The number of publications available about the topic is severely limited, as would be the case for any new discipline).

Books

Anderson, Douglas T, Tribble, Mike. (1995). *The Hard Disk Technical Guide*. Eleventh Edition. Boulder, Colorado: Micro House. ISBN: 1-880252-28-7.

Bodo, Marlin. (1996). *Hard Drive Bible*. Eighth Edition. Sunnyvale, California: Corporate Systems Center. ISBN: 0-9641503-1-X.

Casey, Eoghan (2000). Digital Evidence and Computer Crime. London, Great Britain: Academic Press. ISBN: 0-12-162885-X.

Rosenblatt, Kenneth S. (1995). *High-Technology Crime Investigating Cases Involving Computers*. San Jose, California: KSK Publications. ISBN: 0-9648171-0-1.

Stephenson, Peter (2000). *Investigation Computer-Related Crime*. Boca Raton, Florida:CRC Press LLC. ISBN: 0-8493-2218-9.

Zaenglein, Norbert (1998), *Disk Detective, Boulder*, Colorado, Paladin Press, ISBN: 0-873640992-3.

Periodicals

Computers & Security, Oxford, England, Elsevier Advanced Technology, 8 issues per year, ISSN: 0167-4048.

Information Systems Auditor, International Newsletters, United Kingdom, monthly, ISSN: 1466-4569.

EnCase Legal Journal, South Pasadena, California, Guidance Software, Inc., monthly.

International Journal of Forensic Computing, West Sussex, United Kingdom, monthly, ISSN: 1363-6650.

Vogon Electronic Bulletin, Oxfordshire, United Kingdom, monthly, www.vogon-international.com.

Internet Sites of Interest

http://www.evidence-eliminator.com -*Evidence Eliminator* - hard drive protection (over-writes unwanted data after use automatically)

http://www.spnc.demon.co.uk/ilook/ilook.htm - *ILook Image Investigator* © is a forensic software product.

http://www.computer-forensics.com/ - Computer Forensics Limited

http://www.innovatools.com/software/isecrets/ - *Invisible Secrets* is a steganographic tool.

http://www.vogon-international.com/forensic_services-00.htm - Vogon International Limited

NOTE: The Internet is a fluid living thing. What is valid today may not be valid tomorrow. One or more of these addresses may no longer be active but give them a try.

Glossary

BackOriface and NetBus: tools created by the hacking community and freely available on the Internet, to attack Microsoft NT systems remotely, after being installed on the target system (usually via a Trojan or as a part of another executable) and gain the ability to monitor any activity on the target system.

Bad sectors: A sector is a group of bytes within a track and is the smallest group of bytes that can be addressed on a drive. There are normally tens or hundreds of sectors within each track. The number of bytes in a sector can vary, but it is almost always 512 on drives built in the U.S. Bad sectors reside in clusters that are flagged in the FAT (File Allocation Table) as bad and thereafter the flagged cluster(s) are no longer available to normal access, however, DIRECT reads and writes may still be possible. This feature is automatic based on the OS's ability to access any given cluster. If the OS fails a normal read for a given number of try's the cluster information is relocated (usually before the cluster becomes physically unreadable) and the cluster is flagged as being bad. This flagging can be achieved under user control as well.

Extra Tracks: Most hard disks have several more than the rated number of tracks. These extra tracks are used to make up for flaws that might occur during manufacture that would otherwise require that the entire disk be rejected for failing its quality control requirements. Most times they are not required or used but with DIRECT reads and writes they are accessible and provide a good place for hiding or storing sensitive data.

Partition waste space: After the boot sector of a partition, it is customary to skip the rest of the track and start the volume on the next track. This results in tens or even hundreds of sectors going to waste (not a big deal on a large drive). However, since this area is inaccessible to all but low-level disk viewers, it is an excellent hiding spot for information.

PGP (Pretty Good Privacy): **-** originally developed by Phil Zimmermann and made available in the public domain. Network Associates, Inc currently market this product in its commercial form. Both versions have the potential to provide strong encryption. Information encrypted by "strong" encryption may not be resolvable or made readable without the co-operation of the originator **-** either voluntarily or via Court Order.

RAM slack: is the space from the end of the file to the end of the containing sector and is called RAM slack. Before a sector is written to disk, it is stored in a buffer somewhere in RAM. If the buffer is only partially filled with information before being committed to disk, remnants from the end of the buffer will be written to disk. In this way, information that was never "saved" can be found in RAM slack on disk.

Slack space: is the space between the logical end and the physical end of file and is called the file slack. The logical end of a file comes before the physical end of the cluster in which it is stored. The remaining bytes in the cluster are remnants of previous files or directories stored in that cluster. See figure above.