A Study of Hard Drive Forensics on Consumers' PCs: Data Recovery and Exploitation

B. Dawn Medlin
Appalachian State University

Joseph A. Cazier
Appalachian State University

One of the first actions to take when getting rid of an old personal computer is to remove all of the files that contain identifying and personal information. Individuals can be surprisingly negligent in this effort. Many individuals may also believe that by simply moving their files to the recycle bin and then emptying that bin that all of their programs and files are permanently erased. If personal information is not totally deleted, acts of identity theft can easily occur. Our research study identified the types of information found and/or recovered from hard disk drives on computers that have been donated to charity, sold second-hand, or otherwise donated to other organizations for reuse. Of the fifty-five hard drives studied approximately 300,000 files contained identifiable information. Results showed the need for further training in relation to total file erasure from a hard drive as well as the negative results such as identity theft that can occur due to this lack of training or knowledge.

INTRODUCTION

Wiping a computer clean is not as easy as it may appear. Just deleting the personal files and emptying the recycle bin is essentially next to useless. The delete function only removes file names from a directory list and makes the sectors the files occupy on the hard drive available for future use. Meanwhile, these files actually continue to exist.

To positively prevent data from recovery, disks can be removed from disk drives and broken up, or even ground to microscopic pieces. But the question remains "how many individuals will go to this length to destroy data from their personal computers before donating them?

Although hard disks should only be disposed of after attempts of permanent erasure, many individuals may dispose of their computers without concern for their data believing that the data is simply not of value. Nevertheless, in today's digital age of computers where information technology has grown in substantial ways, the issue of securing information becomes even more imperative, especially given the increase in identity theft.

The scope of identity theft can be large and the levels of theft can range from the takeover or creation of bank accounts, credit cards, loans and/or utility services, to gaining employment and even using the victim's identification to purchase medicines and medical insurance. Experts have found that more than 10 million people were victims of identity theft as of 2004. Those numbers have declined over time, with 9.3 million people reporting in 2005 and 8.4 million in 2007. Yet, in 2007, there was more than $49.3 billion in fraud reported lost because of identity theft, with each victim averaging around $5,000 in losses

(FTC, 2010). Therefore, when individuals dispose of computers, they need to be aware how easy it can be to recover the data from the hard disk. In many cases it simply requires booting the computer and browsing the files on the disk if no passwords were required. In other situations it requires special data recovery programs and/or hardware. It is sufficient to say that it is a relatively easy task that most anyone can attempt.

Since a large number of individuals choose to donate, and their certainly is a need for these computers, it becomes even more imperative that individuals are made aware of the actions needed to completely erase their hard drives. If they do not complete the erasure, potential computer criminals will have no difficulties in obtaining through legitimate methods the information that they may need to commit identity theft and other computer crimes.

LITERATURE REVIEW

Recently, researchers have revealed that a large number of computers have been found in secondary markets that have contained information such as consumer's names, credit card numbers, and social security numbers (Jones, 2005). In fact, the most recently reported cases of identity theft seem to have originated more in offline situations than online. This is not a surprising fact in that most end users may believe that by simply deleting files from their recycle bins they are fully removing all of the files that could be potentially harmful.

It should also be noted that it is not just the data that contains personal information that can be exploited. Files containing video and audio footage, blogs, diaries, and instant messenger conversations can prove to be equally damaging and more easily exploited, especially if it can be linked to an individual. Calendars as well as address books provide routines and places that may be used to stalk an individual.

Identity Theft

Beyond the issue of stalking, identity theft of one of the largest negative results that can occur from information being left on a computer. A 2007 Identity Fraud Survey Report by Javelin states that it takes an average of 55 hours to repair any damage that might be done to a person's identity at a cost of $5,270 to the victim (Monahan, 2007). Further, the number of real hours spent recovering from identity theft is highly variable and lengthy, just as the time frame in which those hours are spent. It takes at least a year for 59% of victims to recover and it takes over two years for another 27% to recover (Foley, et. al 2006). In general, the timeframe in which the victim has the opportunity to make the phone calls and travel to clear their name is not during the weekend or in the evening, but during the normal work week. This creates a loss in employee productivity for victims, which in the end affects the bottom line of a business.

While the financial and time costs of being a victim of identity theft are somewhat concrete, there are other effects known as secondary wounding. When asked, 30% of respondents stated, "my ability to go on with my life is still being impacted" well over two years from the discovery of identity theft, and 7% stated that they were still affected ten or more years later. These secondary wounding effects include credit denial, increased rates on insurance and credit, collection harassment, card cancellations, inability to get employment and credit, and an inability to clear a false criminal record. The most common effects were difficulties gaining credit and loans (63%), credit denial (51%), and collection agency harassment (46%) (Foley, et al, 2006).

The emotional costs after the original crime are not taken into account when dealing with secondary wounding, but they should be considered when accounting for a loss in productivity (Gordon, et. al. 2007). Dr. Nelson, a noted victim psychologist, provides a section in the Identity Theft Resource Center (ITRC) Aftermath Report that sheds light on the extreme toll taken on victims as evident in the following quote:

> *"This study clearly proves the impact of identity theft on its victims leaves similar scars and long-term impact as demonstrated by victims of violent crime....It is disturbing to see*

> *how many people felt exhausted, too tired to continue to fight or even consider suicide instead of standing up for their rights…" (Gordon, et. al. 2007).*

The ITRC report includes a large table of emotional responses/symptoms describing an experience. Twenty-eight percent (28%) of respondents stated that they felt shame, and 29% felt an inability to trust others.

Proactive Erasure Methods

The literature review supporting the research topic surrounding hard drive data erasure encompasses the topics of data types, preservation, erasure, and recovery. Considering the fact that Microsoft Windows® dominates 91% of the market share, we have designed our study around the skills and equipment available to the average consumer (Garfinkel, S.L. & Shelat., A., 2003).

In order to completely understand why an individual needs to be aware of the risks involved with the proper disposition of their computer, one needs to first understand how a hard disk works. The average user probably does not know how files are written to the hard disk, how they are deleted, what types of files contain potentially harmful information, and how files can be recovered. To address these aforementioned topics, listed below are the steps that should be taken in order to ensure complete hard disk erasure.

How Data is Written to the HD

Data is written to the hard disk drive in clusters (the default size is 512 MB) by the drive's read/write head(s) that float on a cushion of air above the platters. A read/write head cannot move from a cluster in one sector (track) to a cluster in a track directly beside it without rotating the platter one full turn, thus clusters are written to the hard drive in a checkerboard fashion.

Because Microsoft Windows® products use a fixed cluster size; often the clusters themselves do not completely fill with data, creating what is commonly known as "slack space." When files are stored, the operating system physically writes the files to clusters on the platters, as well as logically writing a path in the operating system. This action occurs so that the computer will know which cluster is housing a specific file. (http://support.microsoft.com/kb/q211632/).

When files are "sent" or "deleted" to the recycle bin or trash, they are recoverable. The recycle bin or trash is just a hidden folder in the file system of the operating system to which files to be deleted are moved. File locations are also stored logically in the form of paths or pointers within the operating system (OS), so that the operating system knows where to find the file. This storage action is referred to as file allocation tables. When files are deleted using the operating system's "delete" function, the computer, to save time, only deletes the *path* or pointer to the file's location on the hard disk. Next, the operating system erases the path by labeling the entries for the appropriate clusters in the file allocation tables as "free space." Unbeknownst to most end users, the file, however, remains completely intact in the cluster(s) that house(s) it until the cluster(s) is (are) overwritten, allowing "off-the-shelf" programs to fully recover the files.

Residual Magnetism

A hard drive platter is like that of a dusty vinyl record, one could compare that the dust that may reside on the record is representative of the magnetically stored data on the platter, and the needle on the turntable on which the record would be played is representative of the read/write head. Additionally, the grooves on the record can represent or equate to the sectors (tracks) of the hard drive platters that the read/write head follows. As the record turns and the needle moves along the grooves, the needle pushes the dust to either side of the groove.

When files are overwritten, there is data that is spewed upwards by the read/write head, causing a kind of magnetic dust that piles up on either side of the path of the read/write head. This concept is called "Residual Magnetism Rosencrance." This data can be recovered with certain software and hardware scans.

Deleting Files

Many organizations publish what they consider acceptable standards for deleting files. The Department of Defense (DoD) and the National Security Agency (NSA) indicate that overwriting the entire hard drive seven times using a combination of passes with ones, zeros, and random data will effectively wipe out any information that may be present (United States Defense Security Service, 2006).

The Gutmann method uses three different algorithms to overwrite files with 35 passes of the hard disk. With this method, the algorithms used are generally accepted encoding algorithms used by hard drive manufacturers.

Data Recovery

Data can be recovered using a myriad of programs available at little to no cost. Most programs search slack space and "free space." Other programs can be set up to search the unallocated space. Unallocated space is the section of the hard disk that the operating system does not recognize as part of a partition.

There are other methods available for recovering data from the hard disks that require the removal of the platters, then using specialized hardware to scan the surface. These methods would be used to recover data from residual magnetism or from a read/head crash. Both are instances that require a microscope that can scan a large area, such as a hard disk platter.

Erasure programs can also be purchased "off the shelf" or online and range from $50.00 to $995.00. These programs can be effectively used for those individuals who are computer novices and may not be comfortable with recovering data and information through more technical methods. As seen in Table 1, several programs are quite inexpensive and can be easily purchased at most computer stores or through online sites such as Amazon.com. In addition, these programs offer the steps that make it quite simple for the average user to erase information or data that is stored on their computer's hard drive.

TABLE 1
SOFTWARE PACKAGES

Name of software package	Cost for single user license
Symantec's Norton Utilities	$50 - $80
Directory Snoop from Briggs Software	$39.95
Recover My Files from Get Data	$70
Win Hex	$50 - $200 depending on features
Easy Recovery from Ontrack	$200 - $325 depending on features
ProDiscover Basic from Technology Pathways	$995

As previously noted, individuals may not have either the computer or security knowledge to securely erase sensitive information from their donated computers. Although previous studies by researchers such as Garfunkel et al. (2003) and Jones (2005) have addressed security awareness levels of end users, their major metrics addressed only the number of recovered files on hard drives. This information does not truly convey the real threat that occurs if individuals leave sensitive information on their donated computers (Kwon, Lee, & Moon, 2006).

METHODOLOGY

Hard drives were collected from thrift stores and student donated hard drives during the summer and fall of 2007. Brand, model, size, and manufacture data were recorded for each drive. A pilot analysis of six hard drives was conducted to determine if there would be enough sensitive information found to justify a larger study. Interestingly, the pilot analysis resulted in information found that included such items as social security numbers, names, and passwords among other types of information.

After the pilot analysis was completed, fifty-five hard drives were examined for information. Of the fifty-five hard drives, eleven were un-bootable. Over 50,000 files were recovered on each hard drive with some hard drives containing over 300,000 files.

If the drives were password protected, access was attempted using a password-clearing program called "ntpsswrd". Those hard drives that were not bootable were hooked up as secondary hard disks (the primary hard disk being a clean Microsoft® Windows® XP load) and browsed as any other hard drive.

Several hard drives had other operating systems other than Microsoft® Windows® and a different file structure. These hard drives were booted with either a Knoppix Live CD or an Ubuntu Live CD. Once the Live CD was loaded, the files could be browsed. Because the primary purpose was to emulate what the average user would see using readily available resources, only those hard drives that either booted, on their own or with a Live CD, were examined in depth.

As previously, noted, sensitive information provides a prime target opportunity for identity thieves. As shown in Table 2 over half of the hard drives tested included at least a full name. Over 20% had addresses or phone numbers, with 8% that had at least one social security number available.

TABLE 2
INSTANCES OF INFORMATION

	1 > 10 Instances	11 > 100 Instances	100+ Instances	Total
Full Names	24%	7%	13%	44%
Phone Numbers	11%	4%	11%	26%
Addresses	5%	5%	11%	21%
Other Financial Documents	11%			11%
Social Security Numbers	5%	3%		8%
Bank Accounts	8%			8%
Tax Return/Information	8%			8%
Credit Card Information	5%			5%
Debit Card Information	5%			5%
PIN Numbers	5%			5%
Wills	3%			3%

Exploitable Examples

One hard drive contained the user's calendar and also contained an extensive address book in which the names, positions, addresses, and personal (cell) phone numbers of several high-ranking government employees were included. This user had a file that consisted of their last will and testament and other personal documents which had been saved and later sent to the Recycle Bin, all of which were recoverable without using special software. The same user owned two homes at the time the laptop was purchased at an estate sale. Left on the computer were files with directions to both properties, so it would be easy for a malicious individual to find out where the individual had homes and based on that fact possibly determine their income bracket, thus making them an easy target. Because of the contact information this individual had about government employees in various positions, the user could serve as a prime target for intimidation or blackmail.

Another hard drive contained information related to tax returns which contained individual client's names, addresses, phone numbers, social security numbers, and date of birth, almost everything one would need to assume another person's identity. The tax returns also found contained information regarding their annual income, what securities they held, and whether there were, any capital gains/losses based on those securities. Because of this information, activities such as the opening of bank accounts could occur as well as loans taken out in both this individual's name as well as the other employees of the organization. All of this information is part of what forms one's identity, and once in the wrong hands can be used to steal or even destroy that identity. If this personal information had been found with other potentially harmful data, the individuals connected with the data may have set themselves up for extortion and/or blackmail.

The next examined hard drive contained enough information to ascertain that the individual had homes in three different geographic locations, as well as several bank accounts in each of the three locations. Desktop software was found that contained notes about bank account numbers, credit/debit card numbers, PIN numbers, passports, driver's and pilot's license information for both the user and their spouse. The user had a file labeled "passwords" in which they listed access information to all of their online accounts.

Other information such as Internet records, multimedia files, and e-mail contents were also available. This information may not create the threat of serious harm, but it could be information that can be exploited, used to blackmail the user, or embarrass the user if the information were to become public. Some users possess enough knowledge to delete their temporary internet files, but they forget about other areas where files may be store such as cookies, which essentially tracks and individual's internet movement. As seen in Table 3, of the hard drives tested, 50% had cookies, 47% had temporary internet files, and 26% had mailboxes with accessible content.

TABLE 3
CONTENT INSTANCES

CONTENT	1 > 10 Instances	11 > 100 Instances	100+ Instances	Total
Cookies		11%	39%	50%
Temporary Internet Files		3%	45%	48%
Internet Favorites	8%	24%	11%	44%
Pictures	11%	16%	11%	48%
Mailbox Contents (Accessible & # of Messages)	3%	8%	16%	27%
Other Misc. Personal Docs.	13%	11%		24%
Music	5%	8%		13%
Videos	5%			5%
Other Misc. Multimedia		3%		3%

One of the hard drives purchased from a thrift store, apparently had been donated by a medical professional. On this computer there were e-mails between the husband, wife, and their son's school teacher in regards to their son's performance (or lack thereof) in school. That same e-mail account had e-mails between the dentist and her patients as well as other medical professionals, which at the very least is a violation of doctor-patient privilege, and potentially a violation of HIPPA.

Several of the hard drives as seen in Table 4 contained accounting software packages that included individual's names and social security numbers at 11%, followed by 8% of Napster files. With that information, government officials could file possibly file suit for downloading files from illegal P2P sites.

Of the drives tested, there were 5% of the file instances between 11 and 100 that contained employee and/or volunteer data.

In addition, and as noted in Table 4, when an individual donates their computer, information such as corporate tax returns, employee/volunteer information and memos and letters often contain information about multiple individuals and their private/sensitive information. The result of this information can create an opportunity for cyber criminals to gather information about multiple individuals.

TABLE 4
OTHER SECURITY RISKS

	1 > 10 Instances	11 > 100 Instances	100+ Instances	Total
Accounting Software	11%			11%
Napster or Napster Equivalent	8%			8%
Memos	3%	5%		8%
Letters		8%		8%

RECOMMENDATIONS

There are many sites on the Internet that recommend different methods for erasing a hard drive from simply placing the files into the recycle bin to purchasing software. Our recommendation is based on those hard drives that were built after 2001.

Many of the hard drives built after 2001 have a built-in program for securely erasing data, entitled Secure Erase. The program can be accessed through a series of commands embedded in the hard drive. However, before it can be used, the default must be changed from disable to enable within the motherboard BIOS. Additionally, this program works by overwriting every track on the hard drive. Most areas not touched by a simple deletion include bad blocks, directory structure, tracks not touched by the operating system, and unformatted sections of the disk, all of which can be touched by this embedded hard disk utility. In addition, most average computer users may not possess the knowledge of how to access the erasure program through these series of commands.

However, as previously mentioned there are several software packages that are available “off the shelf,” that end users can purchase at reasonable prices. In addition, there are external block overwriting software that may be purchased; however, there is now a free open source version called “Darik’s Boot and Nuke” (Jones, 2005). Whether purchased or downloaded at no cost, the program used should conform to a variety of different standards.

At minimum, it should offer choices of:

1) A single pass overwriting with either zeros, ones, or random data
2) The 1995 DoD standard of 3 passes, the first with either ones or zeros, the second with the opposite of the first, and the third pass should write random data
3) The current DoD standard of 7 passes
4) The Gutmann standard of 35 passes using different algorithms.

Degaussing will effectively render a working hard drive useless, unless a determined individual takes the platters out of the physical disk and attempts to recover the data using Magnetic Force Microscopy (MFM) (Minasi, 1999). The process of degaussing involves exposing the drive to a strong magnetic field. If degaussing is successful, it renders the drive unusable (Rosencrance, 2007).

CONCLUSION

In today's age of identity theft, data left on donated hard drives can lead to devastating results. As indicated in this paper, some methods of data elimination may also prove inadequate. Whichever method is chosen either software wiping or physical destruction, individuals must take "reasonable measures" to safeguard their personal data. Although thoroughly sanitizing or destroying a hard drives takes some effort, the potential costs associated with compromised data make it an important task.

REFERENCES

Fitzgerald, T. J. (2005). Deleted but Not Gone. *The New York*, November 2005. Retrieved on August 31, 2009 from http://www.nytimes.com.

Foley, L., Gordon, S., Barney, K., Rice, K., and Nelson, C. (2006) Identity Theft: The Aftermath 2006, *Identity Theft Resource Center*: San Diego, CA.

Garfinkel, S.L. & Shelat., A. (2003). Remembrance of Data Passed: A Study of Disk Sanitization Practices. *IEEE Security and Privacy*. 1, (1), 17-27.

Gordon, G., Rebovich, D., Choo, K., and Gordon, J., (2007) Identity Fraud Trends and Patterns: Build a Data-Based Foundation for Proactive Enforcement. *Center for Identity Management and Information Protection*: Utica, NY.

Gutmann, P. (2006). Secure Deletion of Data from Magnetic and Solid-State Memory. In Proceedings of the *6th USENIX Security Symposium*, San Jose, California, U.S.A.

Jones, A. & Valli, C. (2005). A UK and Austrailian Study of Hard Disk Disposal. *Edwin Cowan University School of Computer and Information Science Conference Proceedings*. 2005.

Jones, A. (2005). How Much Information Do Organizations Throw Away? *Computer Fraud and Security*. 5, (3), 4-9.

Kwon, Y.C., Lee, S. W., & Moon, S. (2006). Advances in information and computer security. *First International Workshop on Security*, IWSEC 2006, Kyoto, Japan.

Microsoft, Inc. (2007). WD 2000: How Word for Windows Uses Temporary Files. Retrieved on June 6, 2007 from <http://support.microsoft.com/kb/q211632/>.

Minasi, M. (1999). *The Complete PC Upgrade & Maintenance Guide, Tenth Edition*. San Francisco, CA: Sybex.

Rosencrance, L. (2007). Ebay Auction Yields Drive Holding Political Data. *Computer World*. Retrieved on October 4, 2007 from <http://www.computerworld.com>.

Schneier, B. (2005). Risks of Third Party Data. *Communications of the ACM*. 48, (5), 136.

TechWeb News. (2005). Seven in 10 Secondhand Hard Drives Still Have Data. *Information Week*. Retrieved on December 4, 2007 from <http://www.informationweek.com/story/showArticle.jhtml?articleID=163702381>

United States Defense Security Service. (2006). *National Industrial Security Program Operating Manual (NIPSOM)*. Washington: GPO.

United States. Department of the Navy. Naval Information Systems Management Center. *Remanence Security Guidebook*. (NAVSO P-5239-26). Washington: GPO 1993.

Weeks, K. (2007). Hospitals Protect Data By Erasing Old Hard Drives. *San Diego Business Journal*, 23.