Where Data Hides and Resides

Understanding Hidden Data in Windows

Submitted by

Keith R Gittings

GSEC Practical Assignment
Version 1.4b
Research Topic in Information Security

April 30, 2004

Table of Contents

Abstract

Any good Information Security Professional knows that simply deleting a file does not render a file irretrievable; to securely delete a file, the file must be wiped using a wiping utility. However, even this step does not ensure the removal of all data contained in the file. Modern Operating systems store data in multiple locations and make use of the registry, temp files, cookies, metadata, and other forms of data storage to perform tasks on behalf of the user. If this data is left in tact it provides a trail of information that will render the file wiping an obstacle, but not a barrier, to reading that data. This paper focuses on the types of data that the Windows Operating System leaves behind and the software utilities designed to remove that data.

1 Introduction

Several recent stories about high ranking public officials have shown the need for concern about the data that we do not see. The Danish Prime Minister, Anders Fogh Rasmussen, had to switch to disseminating his speeches as PDF files after he accidentally disclosed the origins of his speech (*The Age* 2004). Rasmussen did not want public that his speech originated from Chrisopher Arzroun, an ultra-liberalist from whom Rasmussen had publicly dissociated himself. The disclosure occurred because Rasmussen had published his speech as a Microsoft Word document that contained Arzouni name in the metadata.

Another example comes from researchers at the Massachusetts Institute of Technology who performed an experiment to determine the commonality of the sale of un-sanitized drives. The researchers purchased 158 hard drives on the secondary marker (Garfinkel AND Shelat 2003) from computer stores, small businesses, and the eBay online auction service. They found a frightening majority of these drives to contain credit card numbers, medical records, financial records, and other personal information. This information was available on the drives due to lack of proper sanitation and wiping prior to their release into the market. These examples stress the importance of understanding what data is on a user drive and how to remove it.

2 Deleting Files

As many common utilities prove (i.e. Norton Undelete, etc.), data is not permanently erased by deleting a file. Rather, it can easily be retrieved by using one of the undelete utilities or a simple hex editor. The reason for this is the way that files are commonly deleted.

2.1 Common Deletion Method

The most common file deletion scenario occurs when a user selects a file for deletion. The file then moves to the Recycle Bin, where it will be deleted at a later date when the user empties the Recycle Bin. This method does not actually remove the file from the hard drive, but rather hides it from the Operating System. The space that those files once occupied is put back into the pool of unused space where it can be allocated for reuse. The files can still be accessed via an undelete utility or hex editor, which ignores the operating system's file system tables. Other programs, such as Encase from Guidance Software, can perform a forensic examination of a hard drive to view its contents not as a series of files but as fragments of information that are still remaining on the drive after files have been deleted.

2.2 Wiping Utilities

Since the common deletion method still renders files recoverable, software has been developed to make recovering files nearly impossible. The wiping utilities follow methods that have been developed and approved by organizations such as the National Security Agency (NSA) and the Defense Security Service (DSS) and are used to protect sensitive information from being unwillingly disclosed.

2.2.1 General Use of Wiping Utilities

To prevent a deleted file from being recoverable the file should be wiped. Wiping a file is a broad term used to identify the method for overwriting a file until that file is rendered unrecoverable. There are many methods of wiping a file, such as BCWipe and WipePro+, that are compliant with the Department of Defense's U.S. DoD 5200.28-STD and the Peter Gutmann wiping scheme. These methods will overwrite a file a specified number of times using a specified method. Besides wiping a particular file, wiping utilities also provide methods to remove other information contained within your system.

2.2.2 Additional Features of Wiping Utilities

Since slack space, swap space, and free disk space (discussed in Section 4.2) all contain potentially sensitive information, they must also be wiped. For this reason, wiping utilities offers options to wipe free disk space, swap space, file slack, and directory entries as well. Wiping utilities can also be used to wipe the contents of an entire hard drive. Some utilities, such as WipeDrive from AccessData, are bootable and are capable of wiping an entire drive.

2.3 Clearing Utilities

Wiping utilities like BCWipe also have utilities that can clear recently used files. More specialized programs such as Privacy Eraser are able to clear much of the hidden data residing on a system. These clearing utilities can clear swap files, temp files, histories, cookies, and data files that most users have never even thought of removing (these types of files are discussed in Section 3.4). Using a clearing utility is the easiest way to clear most of these files.

3 Types of Hidden Data

While it is beyond the scope of this paper to list the numerous types of data present on a system (due to each program's different data files and registry entries), the following sections will discuss several of the most common types. By understanding some of the different types of data, the reader can understand how difficult it truly is to delete all personal information from a live system.

3.1 Web Browser Sources

When a user searches the Internet, many files are created and updated that leave a complete picture of where users have been and what they have done. This section covers the different types of information stored and the most common method of clearing this information for Microsoft Windows and Internet Explorer (IE).

3.1.1 Cookies

Websites store information about a visitor's online session activity on their site. The browser stores this data in a text file known as a cookie; the web server receives the cookie from the web browser every time the visitor returns to the site. A cookie stores information that allows a web server to customize a page for a particular user. The cookie may store submitted information in a form such as "airport" and "travel dates" when a user searches for an airline ticket. Also, it could store personal information such as a user's name, address, email address, phone number, or other information that was once entered. Cookies can be cleared by opening Internet Explorer and selecting Tools>Internet Options>Delete Cookies.

3.1.2 Cache (Temporary Internet Files)

When a user visits a website, the browser typically downloads the source and images from that website to the local drive and stores them in a cache file or Temporary Internet Files directory. The browser stores this information to enable quick access the next time the user visits the website, forming a history of the sites a user has visited.

Temporary Internet Files and cache files can be cleared through common methods in the web browser. For example, you can clear the Temporary Internet Files in Internet Explorer by selecting Tools > Internet Options > Delete Files. Both online and offline files must be deleted.

3.1.3 Location Bar History

Common Web Browsers such as Internet Explorer keep a history of the websites that a user visits in the location bar. The location bar history can be accessed by scrolling through the web addresses listed in the location bar. It can be erased by clearing the browser history in IE, Tools>Internet Options>Clear History.

3.1.4 Browser History

Internet Browsers (i.e., Internet Explorer, Netscape, or Mozilla) store information about websites visited, an action that can take a great deal of space and leaves a trail of websites visited. This information can be cleared from your browser. For example, the information in Internet Explorer can be cleared by selecting Tools > Internet Options > Clear History.

3.1.5 Autocomplete Memory

When a user completes an online form and then goes back to that form, much information may already have been completed. While this feature saves a user from having to retype information, it also saves potentially sensitive information, such as the user's name, address, and anything else that a user may have typed.

3.1.6 Downloaded Program Files

Various websites have plug-ins available for download that allow a user to perform some action at that website. These plug-ins can leave a trail of visited websites.

3.1.7 Index.dat

Index.dat files are hidden on a Windows computer and contain information about visited websites. This information includes URLs of web pages that were visited and links to cookies on your system. Additionally, Index.dat files also store information about emails sent and received through Microsoft Outlook and Outlook Express.

The information contained in the Index.dat files remains even after the Temporary Internet Files (Cache) and the History is cleared in Internet Explorer. Microsoft claims that Index.dat files speed up the loading of web pages by caching them; however, unlike the cache the index.dat files are locked by windows and cannot normally be cleared or deleted. Index.dat Viewer allows a user to see what data is held in this file and allows the user to clear Index.dat files. Other clearing programs, such as Privacy Eraser, also perform this task.

3.2 Common Email Programs

3.2.1 PST Files (Microsoft Outlook)

When Microsoft Outlooks is set to store emails in personal folders the emails are stored locally in a PST file. The PST file is a single file that holds the information store, which contains all of the email and attachments that the user stores under personal folders. When a user deletes an email it is moved into a deleted items folder which the user can then empty. Though the email is then no longer present to the user, it is still present and can be retrieved because of the set up of Microsoft Outlook and the PST file. A simple test of the email's presence is to check the size of your PST file, delete several megabytes of email, and then re-check the size of your PST file. The size of the PST file remains the same, although the space once occupied by the email is now made available to Microsoft Outlook for storage of future emails and attachments. A simple method for recovering deleted email is available in the *High Technology Crime Investigation Association* June 2002 newsletter (Shane 2002).

A user may permanently delete an email from Microsoft Outlook by compacting the PST file (which removes the unallocated space) and then clearing the slack space associated with that file. Compaction is accomplished in Microsoft Outlook either manually (right click on Personal Folder > Properties for “Personal Folder”>Advanced>Compact Now) or by setting Outlook to compact automatically when the CPU is not being used. After the PST file has been compacted slack space is still associated with that file. This slack space can be wiped with a wiping utility.

3.2.2 Microsoft Exchange Server

It is important to remember that once a file has been deleted from a user’s email and the PST file has been compacted, the email is still not necessarily gone. Microsoft Exchange Server sometimes stores deleted email so that email administrators can undelete it in the event of user error. This permanently deleted email must be deleted along with the local copy to prevent an email from reappearing after it has been deleted.

3.3 Word Processors (Microsoft Office)

Word Processors and programs in that family (i.e., Microsoft Word) have a lot of data embedded into the file that is not generally seen by the user. As the story of Rasmussen in the introduction illustrates, this hidden data can prove embarrassing.

3.3.1 Temp files

Microsoft Word produces hidden temp files that can maintain copies of or information about the file on which the user is working. Many of these temp files can be seen when “show hidden files” is active (Tool>Folder Options>View>Show Hidden Files and Folders). A search on a user’s Windows system for files with *.tmp extension will show hundreds of temps files that contain data from nearly every application on which that user has been working. Additionally, Microsoft uses OLK directories that may maintain a complete copy of files that have been viewed on your system.

3.3.2 Recent Documents

Microsoft Office programs (Word, PowerPoint, etc.) all have an option for a “Recent Documents” section. This section allows a user to quickly view a document recently used. Recent Documents are organized numerically by file. The Recent documents can be cleared from Microsoft Word (Tools>Options>General>Recently Used Files) by changing the value from 4 (default) to 0 and then applying the change.

Microsoft Office also stores links to documents in the recent directory (C:/Documents and Settings/*username/*Recent*).* These links can allow others to know the most recent documents on which a user has been working.

3.3.3 Metadata

Metadata is data about data. Microsoft Office stores metadata with the actual file. To illustrate how much metadata a Microsoft Word document stores, consider that a one-character Word document is 19 kilobyte compared to 1 byte for a one-character Notepad document. This metadata lists potentially sensitive information about the user and the user's company.

To respond to this and other reported issues, Microsoft published a tool, rhdtool.exe (Microsoft 2003). This tool removes information that a user may not want to share publicly. The tool can be run from Office XP or Office 2003 on individual files or from the command line as a batch job on multiple files.

3.4 Miscellaneous Microsoft Windows Files

Microsoft stores extensive information about a user's actions on the system. Some of this information is stored intentionally (such as Windows' search history) to increase user efficiency.

3.4.1 Windows Search History

Windows and Internet Explorer store information on files, data, etc., for which a user has searched. This convenient feature enhances usability by saving a user from needing to retype searches; however, history also leaves behind potentially sensitive information of that user's activities.

3.4.2 Open/Save History

The Windows registry stores much of the information we discuss in this paper. One example of this saved information is the Open/Save History. The registry stores this information when a user opens and saves a file.

3.4.3 Windows Swap File

Perhaps the most important files for storing hidden data is the Windows Swap file. The Windows Swap file can potentially store any information that has ever been used on a system in plain text—even encrypted information. The swap file is used with Virtual Memory as a back store for pages that are swapped in and out of memory. Since everything processed on a system must reside in its memory at one time or another, the memory can contain anything from personal information to passwords and identifiers. It is named Win386.swp in Windows 9x and pagefile.sys in Windows NT/2000/XP.

Like Index.dat, the Swap file is locked and cannot be deleted. Programs such as BCWipe have options to wipe the swap file. While the system is running, BCWipe has an option to wipe unused space in the Swap File. Since the Swap File is locked while the system is running, a complete wipe can only be performed at logon. A user can perform this task by setting up a task in BCWIpe. The task can be set up selecting BCWipe TaskManger>Tasks>Create New Task>Special Folders>Swap File. The user then schedules a time for the swap file to be wiped. The list of special folders contains many other files as well that should be wiped on a regular basis.

3.4.4 Hibernation file

When a computer transitions from an active state to a hibernation state, the contents of RAM is immediately copied into a hibernation file. This process allows the system to hibernate and therefore use less power since power to the RAM is no longer required. Since the contents of this file includes the contents of RAM at a particular state, it could potentially hold information about anything the user had opened at any given time. Wipe programs have options to wipe the hibernation file for the user.

3.4.5 Start Menu Run History

Windows stores programs that are run in the start menu. This feature prevents users from needing to type the program name each time they run it; however, it can also leave information for others regarding the users actions.

3.4.6 MediaPlayer file list

Multimedia programs such as Windows MediaPlayer and RealOne Player store information about the Internet locations of files that have been viewed by the players. This information can give people a history of a user’s Internet use.

3.4.7 Windows NT Data Sreams

The Windows file System NTFS allows for the storage of data on alternate data streams (De Clerq 2001). This allows for multiple forms of data to be associated with a single file. Microsoft added the data streams so that Windows NT could act as a file server for Macintosh clients. With the use of data streams, data can be contained in a file that is completely invisible to the Windows Explorer user. To view and remove data contained in data streams, you can either use the DOS command line (De Clerq 2001) or a program like NTI computer forensic tool.

4 File System Files

Because of the design of most modern file systems, not all of the space can be utilized. This unutilized space contains data from files that previously resided in that location. This data can come in the form of slack space and free space.

4.1 Free Space

Free space, or unallocated space, is a series of clusters not currently allocated by the File System. When a file is deleted, the data is not deleted but rather the clusters for that file are added back to the pool of available clusters. These clusters still contain the same data but are made invisible to the user. Hex editors, or programs such as Norton Undelete and Encase, are able to view the clusters without using the file system and can read the data contained in the unallocated clusters. Once the clusters are reallocated, the amount of the cluster that is not used by the new file becomes slack space, which is explained in greater detail below.

4.2 Slack Space

Windows File Systems, such as FAT32 and NTFS, use fixed cluster size to store files. Since the data contained in a file does not always match the length of the cluster size, an extra space is present in the cluster after the end of file (EOF) marker for that file. This space after the EOF marker is known as slack space. The slack space is usually comprised of remnants of the last file that resided in that cluster. The following example illustrates how unassociated data remains in a cluster.

For example, **Figure 1** represents the cluster size, and **Figure 2** represents the data contained within the cluster.

|---|

Figure 1. Cluster Size

|--John’s social security number is 999-85-9999EOF-|

Figure 2. Data Contained within the Cluster

Now, assume that the data in **Figure 2** was deleted and then the data “Willy Wonka” was saved as a file in the same cluster. The resulting cluster is shown in **Figure 3**.

|--Willy WonkaEOFsecurity number is 999-85-9999--|

Figure 3. New Data Contained within the Cluster

The user only sees “Willa Wonka” in the file, but an investigation of the cluster with a HEX editor would show a random social security number after the EOF marker.

To prevent this problem, Wiping Programs such as BCWipe or WipePro+ will wipe the slack space so that the file appears as it does in **Figure 4**.

|--Willa WonkaEOF-------------------------------------|-

Figure 4. Data After Wiping

5 Intentionally Hidden Data

Not all data that is hidden on a system is produced by the Operating System. Some data is intentionally hidden by the user. Users can hide data in many simple ways. One simple way to hide data would to be change the file extension. For example, to hide a .jpg file, simply change the file extension to .doc. Since Windows uses the file extension to identify the type of file, the image file will not be recognized as an image file. Programs such as QuickView Plus allow a user to view files without relying on the file extension. The sections below discuss more complex methods of hiding data—Steganography and Encryption.

5.1 Steganography

Steganography, which literally means "covered writing", dates back to ancient Greece. The Ancient Greeks practiced Steganography by shaving a messenger's head, tattooing a message onto the head, and then allowing the hair to grow back to cover the message. The messenger then carried a fake message in hand. Steganography was also practiced by cutting a message into a piece of wood and then covering the wood tablet with wax. The practice of Steganography or "Stego" survives today with modern day Stego software.

5.2 Steganography Software

Steganography software, such as S-Tools, works by taking a message and embedding it into a file. Stego software can be used with almost any type of file, but it is most prevalent in Media files (Images, Music, Video, etc.). The software embeds the file into the unused or useless bits of the files that are able to hide the message. Stego is intended to be unseen to the naked eye; however, some stegoed image files may leave a noticeable effect on the image, particularly when compared to the original. Stego Software may alter parts of the file to embed its message. For example, an image file may use a slightly different blue than the original blue of the image. While the effect on the image may be unnoticeable, that bit change may hide valuable data. Steganography also most be reversible; therefore, the procedure used to Stego a file must also be reversible to retrieve that message.

5.2.1 Detecting Steganography

Of course, since software has been written to hide messages, other software has been written to detect those hidden message. Since Stego leaves a pattern that allows the file to be unstegoed, that pattern can be detected by software (a link to Stego Software is listed). It has been rumored that terrorist and other nefarious groups have been using Steganography software to communicate by placing stegoed images on public sites such as Ebay. Neils Provos, a Ph.D. student in Computer Science at the University of Michigan in Ann Arbor, made this claim the topic of his dissertation (Carvin 2001). Provos developed detection tools that analyzed more than 2 million images posted on Ebay for signs of stego; however, he found no cases of stego on any of the images.

5.3 Encryption

Encryption is not used to hide data the way Steganography does, but instead makes viewing data as difficult as possible. This can make knowing what data resides on your system difficult, even when the user knows where it is. It is also important to be aware that encrypted data can appear as plaintext in memory both before and after it has been encrypted. Therefore, while a user may have the file encrypted on the system, it may be in plaintext form in the hibernation or swap file. For this reason, it may be beneficial to use Swap File Encryption.

5.3.1 Swap File Encryption

Since the swap file can contain data from anything that was viewed on a system in plain text, it is a good candidate for encryption. BCWipe offers an option to encrypt the swap file this way, ensuring that an encrypted document does not have an unencrypted document sitting in the swap file.

6 Conclusion

It is nearly impossible to completely remove all the private information stored on a system. Forensic Labs, such as the Defense Computer Forensics Lab, specialize in finding information that users thought they had deleted and removed (Radcliff 2004). One must be aware of the data stored on a system and where it hides and resides. The only true way to remove all data on a system is to use an approved wiping method on the entire hard drive or use a highly rated degausser. Of course, after performing the actions, the drive should be shredded, burned, and scattered in the ocean.

Bibliography

Articles

Associated Press. "Microsoft zaps your dirty laundry" CNN.com. February 9, 2004. http://www.cnn.com/2004/TECH/biztech/02/09/erase.microsoft.ap/

Carvin, Andy. "When a Picture Is Worth a Thousand Secrets: The Debate Over Online Steganography" *The Digital Beat.* 2001 http://www.benton.org/publibrary/digitalbeat/db103101.html

De Clercq, Jan. "NT Gatekeeper: Understand Alternate Data Streams." *Windows NT Magazine* 2001. http://www.winnetmag.com/articles/print.cfm?ArticleID=20915

Garfinkel, Simon and Shelat, Abhi. "Remembrance of Data Passed: A Study of Disk Sanitization Practices" *Published in IEEE Security & Privacy, Vol. 1 No. 1, 2003.* http://www.computer.org/security/v1n1/garfinkel.htm

Kuepper, Brian. "What you don't see on your hard drive." *SANS Reading Room. 2002* http://www.sans.org/rr/papers/27/653.pdf

Lane, Sarah. "Sarah's Office Tweak: Remove Hidden Data" *TechTV*. 2004. http://www.techtv.com/screensavers/print/0,23102,3621553,00.html

Mallery, John. "Secure File Deletion: Fact or Fiction" *SANS Reading Room. 2001* http://www.sans.org/rr/papers/27/631.pdf

Microsoft. "Office 2003 /XP Add-in: Remove Hidden Data." 2003. http://www.microsoft.com/downloads/details.aspx?FamilyID=144E54ED-D43E-42CA-BC7B-5446D34E5360&displaylang=en

Radcliff, Deborah. "Insides the DOD's crime lab" *Network World Fusion.* 2004 http://www.nwfusion.com/research/2004/0308dod.html

Radcliff, Deborah. "Steganography: Hidden Data". *Computerworld.* June 10, 2002. http://www.computerworld.com/printthis/2002/0,4814,71726,00.html

Shane, Randall. "Techie Tips: Mining Clues from Email (Part 1 of 2)". *High Technology Crime Investigation Newsletter* June 2002. http://www.htcia.org/online_newsframe.htm

The Age Online Staff. “Hidden Data in Word” *The Age. February 4,* 2004.
http://www.theage.com.au/articles/2004/02/04/1075853909521.html

The Age Online Staff. “Danish prime minister gets bitten by Word” *The Age.* 2004.
http://theage.com.au/articles/2004/01/13/1073877800625.html

Varghese, Sam. “UK government gets bitten by Microsoft Word” *The Age.* 2003
http://theage.com.au/articles/2003/07/02/1056825430340.html

Webopedia. “Steganography”. *Webopedia Encylopedia.*
http://www.webopedia.com/TERM/s/steganography.html

Software

BCWipe
http://www.jetico.com/

Encase from Guidance Software
http://www.encase.com/

Index.dat Viewer
http://www.exits.ro/

Norton Undelete
http://www.symantec.com/sabu/sysworks/basic/

NTI Computer Forensic Tools
http://www.forensics-intl.com/tools.html

Privacy Eraser.
http://www.privacyeraser.com/

QuickView Plus
www.jasc.com

S-Tools
Steganography Software Listing
http://www.jjtc.com/stegoarchive/stego/software.html

WipePro+
http://www.marcompress.com/Eval-Download.htm

WipeDrive
www.accessdata.com