

Challenges in Hardening Data against Recovery

By Jason Andress

ISSA member, Colorado Springs chapter

This article provides a methodology for analyzing means of permanent data removal as well as a look at the future of hardening data and relevant technologies.



While the IT industry devotes a tremendous amount of time and energy to storing data reliably, the problem of permanently deleting or hardening of data is not as easy as it seems. This article provides a methodology for analyzing means of permanent data removal as well as a look at the future of hardening data and relevant technologies.

One of the primary tenets of the security field is the requirement for data protection. Data protection is the ability to control access to data, thereby prohibiting unauthorized access and modification. While it is important to secure data to the appropriate level, this comes at a cost, both in the monetary sense and in the sense of impact to productivity.

When considering physical data security, items such as hardware redundancy, data backups, and additional security-related hardware such as firewall or proxy appliances all have a price tag attached to them. While this is the most obvious cost of security, other factors that may increase the long-term cost need to be factored in, in order to more accurately calculate the total cost of providing security.

Logical measures for securing data, such as encryption, also have a price tag associated with them, but generally less of

a monetary one. The cost of logical security measures is generally an inconvenience to the users of the data. Adding passwords, VPNs, and other similar security measures directly impact the ability of the user to function. When users are subject to levels of security that impact their ability to function, they tend to find ways to circumvent these security measures.

Deleting stored data is more difficult than it appears

One of the exacerbating factors in data being undesirably recovered is the ease with which data is apparently deleted. It is very easy for the average user to delete the files that he wishes to discard and assume that they are permanently destroyed. Unfortunately, this is far from being true. In many cases, information is unknowingly retained in documents and data is not as thoroughly deleted as the user might think.

Deleting files

The exact sequence of events that occurs when a file is deleted from digital storage media differs somewhat depending on the operating system and file system being used, but, for

the most part, the concepts are similar. When a file is deleted, the pointer to the file is removed and the area in which the file is stored in is marked as being available for use. In other words, the file still exists, almost entirely untouched, on the media. For example, the sequence of events for file deletion in the venerable NTFS file system works something like the following:

1. Read the data from the file record in the Master File Table (MFT)
2. Locate the entry that contains the directory in which the file is stored
3. Locate the address for the specific file being deleted
4. Remove the entry for the file from the file index
5. Remove the in-use flag for the entry in the MFT
6. Mark the clusters in which the file is stored as unallocated

As was previously mentioned, the file still exists, unscathed, on the media. Only the information that points to the file has been altered. Until the clusters in which the file was stored have been overwritten, the file will be recoverable in its entirety.

Sanitizing storage media

When digital storage media is being disposed of, the current user of the media may want to destroy all of the data recorded on it in order to prevent disclosure of data to future users of the media. This task is commonly undertaken by formatting the media using the utilities included in the operating system for performing such tasks. Unfortunately, considerably more heroic measures are required to safely destroy all data recorded on the media.

Depending on the specific tool used, less than one percent of the blocks on the drive may be changed by using a file system formatting utility to format the existing partitions. To make matters worse, performing such a formatting and subsequently loading another operating system on the media is likely to be insufficient as well. In the case of a specific machine being examined for data persistence purposes, data was recoverable, years afterward, from the original install of Solaris, even after the machine had been reinstalled with a Windows operating system, then subsequently with a Linux operating system.

In order to remove data from storage media properly and have some assurance that the process being followed will actually perform the function of data removal correctly and securely, some sort of standardization is clearly required.

Standards for secure data removal

The set of standards put in place for the U.S. Department of Defense (DoD) in DoD 5200.28-STD¹ have become the de

facto standard used to evaluate the tools and methods used for removing or destroying data. Many other standards exist as well:

- NAVSO P-5239-26 U.S. Navy²
- AFSSI-5020 – U.S. Air Force³
- AR380-19 – U.S. Army⁴
- GOST – Russian⁵
- VSITR – German⁶
- RCMP TSSIT OPS-II – Canadian⁷

Several products and methods that are available for data removal are compliant with several, if not all, of these standards.

Terminology

The common terminology used to discuss secure deletion of data is taken from the U.S. Department of Defense publication, the *National Industrial Security Program Operating Manual* (NISPOM).⁸ This manual introduces the concepts of cleaning and sanitization of digital storage media. In order to intelligently discuss these methods, a few other concepts need to be defined:

- **Clearing media** – refers to methods of removing data that will render the data unrecoverable by means less than but not including a laboratory attack
- **Sanitizing or purging media** – means that the media has had the information removed from it in such a way that the data cannot be recovered by any known means or method of analysis, including a laboratory attack

When dealing with classified data, either clearing or sanitization is deemed appropriate when the media is intended to be re-used at the same level of classification. Media that will be downgraded in classification level or declassified must be sanitized.

Standard(s) for sanitizing data

Looking specifically at the standard set forth by the DoD for types of media currently in common use, two main methods exist for removing data from media: degaussing and overwriting. The specific methods that must be used differ by media and are laid out in Table 1, as specified in NISPOM.

1 *National Industrial Security Program Operating Manual Supplement*, U.S. Department of Defense, 2004.

2 "VSO P-5239-26. Remanence Security Guidebook, Module 26," *Information Systems Security (INFOSEC) Program Guidelines*, Department of the Navy Naval Information Systems Management Center, 1993.

3 AFSSI-5020, 2008, <http://en.wikipedia.org/wiki/AFSSI-5020>.

4 Army Regulation 380-19, Department of the Army, 1998.

5 GOST, <http://en.wikipedia.org/wiki/GOST>.

6 VISTR, Bundesamt Fur Sicherheit in der Informationstechnik, 2004, [http://www.vd-bw.de/webvdbw/navigation_neu.nsf/22a49b54309493d2c12569700053b130/7dfa3ee7f009cf7ec1256fa2002ff158/\\$FILE/VSITR.pdf](http://www.vd-bw.de/webvdbw/navigation_neu.nsf/22a49b54309493d2c12569700053b130/7dfa3ee7f009cf7ec1256fa2002ff158/$FILE/VSITR.pdf).

7 *Information Technology Security Guide*, Royal Canadian Mounted Police, 2003, http://www.rcmp-rc.gc.ca/tsb/pubs/it_sec/g2-003_e.pdf.

8 *National Industrial Security Program Operating Manual Supplement*, U.S. Department of Defense, 2004.

Clearing and Sanitizing Data Storage

Media Type	Clear Media	Sanitize Media
Magnetic Tape		
Type I (!>350 Oe)	A or B	A, B, or Destroy
Type II (!>750 Oe)	A or B	B or Destroy
Type III (>750 Oe)	A or B	Destroy
Magnetic Disk Packs		
Type I (!>350 Oe)	N/A	A, B, or C
Type II (!>750 Oe)	N/A	B or C
Type III (>750 Oe)	N/A	Destroy
Magnetic Disk Packs		
Floppy	A, B, or C	Destroy
Bernoulli	A, B, or C	Destroy
Removable Hard Disk	A, B, or C	A, B, C, or Destroy
Non-Removable Hard Disk	C	A, B, C, or Destroy
Optical Disk		
Ready Only	N/A	Destroy
Write Once, Read Many	N/A	Destroy
Read Many, Write Many	C	Destroy

A – Degauss with Type I Degausser
B – Degauss with Type II Degausser
C – Overwrite all locations with a character,
the compliment of the character, then a random character

Table 1– Clearing and Sanitizing Data Storage

Degaussing media is accomplished by randomizing the magnetic field used to store the data. Some varieties of magnetic media require stronger types of degaussing equipment for erasure due to having higher levels of coercivity. The coercivity of the media, measured in Oersteds (Oe), defines the strength of the magnetic field required to reduce the magnetic field of the media to zero.⁹ The DoD considers degaussing to be a more reliable method of removing data than overwriting.

Overwriting the data stored on the media is accomplished by writing data three times to every byte on the disk. The first pass writes a character, the second pass writes the complement of the character used in the first pass, and the third pass writes a random character. Additionally, a random sampling must be conducted to ensure that all sectors have been overwritten, as well as a check performed to ensure that no new bad sectors have been created. If new bad sectors have been created, the media must be subsequently sanitized with the appropriate level of degauser.¹⁰

⁹ P. Wilkinson, "Protecting Military Data in a Flash," *COTS Journal*, 2005.

¹⁰ Note that DoD standards are updated with some regularity and that these methods may not be acceptable for some types of media and/or levels of data. It is always recommended to seek out the most recent update of the regulatory information governing the data and media in question.

Data recovery

Data recovery, which can be defined as the retrieval of information which has been put beyond the reach of the average computer user, is often considered to be one of the black arts of computing. Many operating systems have simple tools for recovering files that have accidentally been deleted, and many companies will happily sell you tools to help with these sorts of tasks. If these tools prove inadequate, there are companies that will attempt to recover data which is beyond the reach of these utilities or from media which has become damaged, for a price ranging from several hundred to several thousand dollars.

Data recovery tools

Data recovery utilities are generally most useful immediately after the file has been deleted, and then only if measures are not in place to prevent the recovery of the file. If the operating system does not have a built-in tool to recover deleted files, or if these tools prove not to be up to the task, many commercial file recovery tools can aid the user. These tools, although generally cosmetically more appealing and perhaps more user friendly, are generally not much more sophisticated in operation than some of the simpler tools. If these tools fail to return the needed data and the data is of sufficient importance to merit it, the next recourse is generally to turn to a commercial data recovery service.

Methods of data recovery

Data can be recovered from media in a variety of ways. It can be pulled from unallocated storage space, slack space, from storage artifacts, from areas of the disk which are not normally accessible to users and, if these methods fail, the data can be recovered from the media directly via magnetic microscopy.

Unallocated storage space

As previously discussed, deleted files are generally not removed from the storage media unless steps have been taken to do otherwise. The areas that the deleted files are stored in are marked as being available for use but, particularly given the large storage sizes of modern media, may not be overwritten for extended periods of time.

Slack space

Slack space is space that, due to the way that data is stored on digital media, is considered to be allocated to data, but may not be entirely filled by the data that it contains. When data is written to storage media, there is a minimum unit of data that can be written. Any remaining amount of space in the minimum unit that is not used is considered to be slack space.

A good analogy to explain this phenomenon is that of recording television programs onto a VCR tape. First, a one hour show is recorded and the tape is rewound. Next, a thirty min-

Factors in hardening data

Factor Number & Name	Description	Examples	Evaluation of Methods by Factor		
			Physical	Post-write	Pre-write
F1 - Speed	How quickly process can be executed	Minutes, hours, days	8	1	8
F2 - Completeness	How thorough the process is in removal of data	Media destroyed, % bits overwritten	10	3	8
F3 - Process cost	Monetary cost of executing the process	Equipment maintenance, software licensing	5	5	10
F4 - Effect on performance	Performance effect on the media before execution	CPU usage, login time, write speed	10	10	2
F5 - Training required	The level of training required to execute process	None, minimal, extensive	7	7	6
F6 - Equipment required	The level of equipment needed to execute process	Industrial shredder, degausser, blunt object	2	5	5
F7 - Reusability	How reusable the media is after process has been executed	Media destroyed, completely reusable	0	10	10
F8 - Administrative cost	Effort needed to reuse/replace the media after process has been executed	Replace media, recreate file system	5	10	10
F9 - Geographical requirements	Where the method is intended to be executed	Aircraft, ship, office	3	5	10

Table 2 – Factors in Hardening Data

ute show is recorded and the tape is rewound. Finally, fifteen minutes of a news broadcast are recorded. Upon examining the tape, the news broadcast is intact, fifteen minutes of the thirty minute show are present, and thirty minutes of the one hour show are present.

Storage artifacts

When a particular piece of media has been used to store data in more than one partition or file system over time, it is likely that storage artifacts will develop. An example of such a storage artifact is an area of the media which, using the present file system or partition is inaccessible, but was, at one time, used for storage. For example, a hard drive is originally configured with four partitions under operating system A. The drive is then loaded with operating system B. In the process of reconfiguring the drive to load operating system B, an eight megabyte portion of the drive is left as free space. This portion of the drive is never partitioned or formatted as part of the operating system load. It is possible that data from the original operating system exists in this unused portion of the media. Given the differing nature of storage systems used by various operating systems, it is quite possible that data stored in one of these areas may never be overwritten.

Non-user accessible storage areas

On many types of modern storage media, portions of the storage area are set aside for the use of the manufacturer. This reserved space may be used to compensate for bad blocks, to store data about the specifics and operation of the media, store the microcode that controls the operation of the device, and a variety of other items that vary by manufacturer. The other potentially large area of inaccessible storage may be cre-

ated when a manufacturer logically modifies a drive to reduce its capacity. Any of these places could, in theory be used to hide data.

Magnetic microscopy and edge effects

Data is written to a single disc platter in concentric circles called tracks. Due to the magnetic nature of the media and slight variances in the position of the drive head, the writing process leaves residual data outside the normal boundaries of the track; this residual data is referred to as the edge effect.

Magnetic microscopy is performed by using a device that can sense very slight magnetic fields to directly examine the surface of magnetic storage media. This allows the entire surface of the media, including the track edges to be examined instead of only the areas that are normally accessible.

The basic tools to do magnetic microscopy can be assembled relatively easily for around \$100.¹¹ This, of course, does not provide a full automated solution like those used by large-scale data recovery services, but makes a good starting point and exemplifies the level of technology required to recover apparently overwritten data.

Analysis of methods used to harden data

Data hardening classifies methods that may be used to prevent digitally recorded data from being recovered in a useful state from storage media. A wide variety of methods can be used in an attempt to harden data, some more successful than others. The determination of success for any given method is entirely situational; in some cases the speed of execution and

11 P. Wilkinson, "Protecting Military Data in a Flash," *COTS Journal*, 2005.

completeness of destruction might outweigh all other factors. In other cases, factors such as cost or training might weigh more heavily.

Factors in hardening data

When evaluating methods of hardening data, it is helpful to examine some of the factors that might make a given method more or less desirable in a given situation. Use of these factors enables rating of the various methods that might be used to harden data. In the sections below, each of the factors has been evaluated on a scale with a range from one to ten, one being the least desirable, and ten being the most desirable. These factors are intended as a general guideline for evaluation and may not rate the same, or even be applicable, in every situation. The evaluation methodology is intended to be flexible in this sense.

Some of the factors involved in hardening data are shown in Table 2.

For purposes of brevity, the various methods being examined have been classified into the following areas:

- Physical destruction of media
- Post-write logical alteration
- Pre-write logical alteration

Table 2 also shows the ratings assigned to the individual factors for each of the three categories, as well as the selected factor ratings for each method. Each of the methods is also displayed graphically below. Based on the rating of the selected factors, stronger methods, i.e., those with a greater set of more desirable characteristics, will cover more of the chart and weaker methods less.

Physical destruction

Physical destruction of the media is a generally accepted as a sure method of hardening data. This category includes such methods as shredding, incinerating, submerging in acid, blunt force, degaussing, and other similar

methods, and often equates to sanitizing, as described in NISPOM. As can be seen from the data, the primary benefits of physical destruction are speed, completeness, and no effect on performance. Unfortunately, executing this type of method generally requires the presence of some sort of equipment, an industrial shredder for instance, and entirely precludes reuse of the media. Figure 1 shows the total coverage for this method.

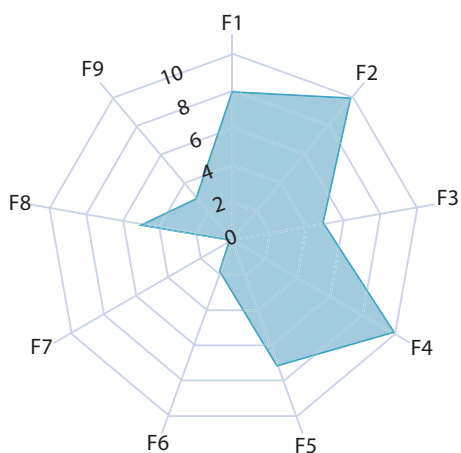


Figure 1 – Physical Destruction of Data

Post-write logical alteration

Post-write logical alteration, i.e., alterations made to the data after it is written to the media, is another accepted method of hardening data. The most common use of this method is overwriting data multiple times. Additionally, changing the partition structure, reformatting existing partitions, low-level formatting, changing the file system, or some combination of the above methods are commonly used. This method equates to clearing, and in some cases sanitizing, from NISPOM. This particular method has the advantages of having no effect on performance, complete reusability, and easy re-deployment of the media. The flaws in the use of this method lie in speed and completeness. In order to accomplish a relatively sure erasure of data, most variations of this method have to be executed repeatedly. Given media with a large capacity, this process could require several days to complete. Figure 2 shows the total coverage for this method.

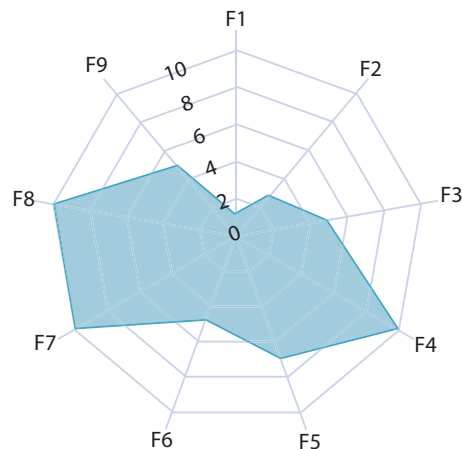


Figure 2 – Post-write Alteration of Data

Pre-write logical alteration is most commonly performed via encryption of data stored on the media. In order to provide a strong level of security, the use of a method that encrypts the media as a whole is desirable. Other methods, of course, are available, but tend toward the esoteric, such as the use of non-standard file systems. Pre-write alteration is generally fast, has a low process cost, enables the media to be reused, has a low administrative cost to re-deploy, and can be used without geographical restriction. The price of using this technique to protect data lies in hits to performance. This method, examined in a general sense, provides the best coverage across all categories, as can be seen in Figure 3.

Pre-write logical alteration

Pre-write logical alteration is most commonly performed via encryption of data stored on the media. In order to provide a strong level of security, the use of a method that encrypts the media as a whole is desirable. Other methods, of course, are available, but tend toward the esoteric, such as the use of non-standard file systems. Pre-write alteration is generally fast, has a low process cost, enables the media to be reused, has a low administrative cost to re-deploy, and can be used without geographical restriction. The price of using this technique to protect data lies in hits to performance. This method, examined in a general sense, provides the best coverage across all categories, as can be seen in Figure 3.

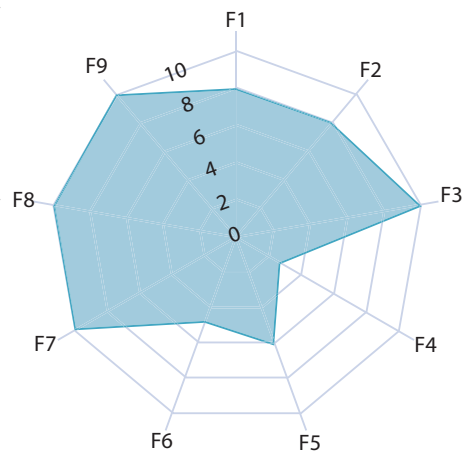


Figure 3 – Pre-write Alteration of Data

Future issues in data storage

The emergence of new technologies, as well as the political and social ramifications of their use, is sure to bring about a great deal of change in the areas of data and data storage. It is almost certain that, given the current global level of concern with security and the rising population of computer users, digital forensic science is a field that will see the devotion of a great deal of resources.

Many people now carry USB flash drives regularly that, as of this writing, can be easily found in capacities of 8 GB. Also common are MP3 players with capacities that seem to be as large as those found in many laptop devices. These devices can also usually be used for file storage. Storage has become enough of a concern that the IEEE began, in 2001, to sponsor an annual Security in Storage Workshop in order to discuss these and related issues. To address these issues, some companies have begun to design flash media specifically with the idea of data hardening in mind. An example of one such device is an 8 GB flash drive which can be entirely purged of data in under ten seconds.¹²

Conclusion

Given the enormous body of digitally stored data that exists today, ensuring the protection and proper disposal of this information is paramount. Current methods of hardening data against recovery can prove to be extremely challenging. Most

of these methods have one or more aspects in which they perform well, but none of the current ways of securely deleting data works well in every instance.

New techniques are sure to arise which will allow for more efficient and complete hardening of data, shortly followed by new methods of penetrating such techniques. Ultimately, if the data is important enough to merit the expense, any method of data hardening can be breached or circumvented entirely.

References

- D. Katz, “What You Don’t Know about Sarbanes-Oxley,” *cfo.com*, 2003, http://www.cfo.com/article.cfm/3008498/3/c_3037076.
- J. Mueller, “STM – A Project by Juergen Mueller,” 2008, http://www.e-basteln.de/index_r.htm.
- “Data Recovery Services,” Seagate Technology, LLC, <http://www.seagate.com/support/service/drs/faq.html>.

About the Author

Jason Andress is a tinkerer, rascalion, and all around geek. He works for a major software company, teaches graduate and undergraduate security courses, and enjoys a good game of Scrabble. He can be reached at jason.andress@gmail.com.



¹² *ibid.*