

Secure Data Deletion and Recovery

By Jason Andress

Achievement of reliable storage for data consumes a tremendous amount of time and energy. The problem of permanently deleting data is not as simple as it appears.

Achievement of reliable storage for data consumes a tremendous amount of time and energy. The problem of permanently deleting data is not as simple as it appears. This article discusses the motivations for preventing data recovery; how data can be recovered; and how data can be deleted or destroyed in such a way as to make recovery impossible.

A wide variety of data is committed to digital storage, some of great value and other trivial. This data includes criminal records in justice departments, medical records at hospitals, mental health information, military secrets, and innumerable other data, distributed across countless systems. In most cases, concerns about private data are limited to preserving it, accessing it, and securing it against unauthorized access. In the case of trivial data, such as grocery purchase history, this is generally considered to be acceptable. In the case of financial records or medical information, concerns also focus on protecting the data to ensure that it does not fall into the wrong hands. In some of these cases, the ability to delete data reliably becomes paramount.

How data is stored

Throughout known history, data storage has been utilized for everything from record keeping for tribal societies to much more abstract, digitally stored data, such as financial transactions or radio noise from space. Older, more primitive methods include characters etched on tortoise shells around 6,600 BC, Incan quipu record keeping at 3,000 BC, and parchment in Roman times. Many older methods of data storage seem to have longer storage lifetimes than many modern methods. Artifacts have been found that are thousands of

years old and still well preserved. The expected storage life of a CD is estimated to be between five and ten years – but the amount of stored data is enormous in comparison with older methods.

Modern data storage methods include such items as the typeset book, the photograph, the magnetic disk drive, the compact disc, and flash memory. Data storage serves a myriad of purposes in the modern world. Data can be anything from the handwritten business records of gangs that deal crack cocaine, to records on the phone habits of millions of Americans kept by the National Security Agency (NSA). Data storage also keeps society functioning: Traffic patterns, grocery store inventories, medical records, financial transactions, and a host of other items yield data. Without the means to store data, maintaining a modern level of civilization would be a herculean endeavor.

Attempts to preserve data have been ongoing for thousands of years, from religious materials such as the Dead Sea scrolls hidden in caves, to classified data on troop movements that is stored on specially secured computer systems in locked and guarded buildings. It might be desirable to be able to destroy some data, however, such as the backup tapes of email documenting the Iran-Contra affair. More recently, the theft of Cisco Systems' IOS source code comes to mind¹.

The ultimate problem with such disclosures of information, intentional or otherwise, is that data is often impossible to secure again once it has been released. Particularly since the arrival of the information age, once data has been released on the Internet it is almost impossible to contain it. An excellent example of this is the case of the DeCSS source code. DeCSS is a tool, popularly believed to have

¹ "Cisco source code reportedly stolen." <http://www.eweek.com/article2/0,1895,1593862,00.asp>

been developed by one Jon Johansen, that could be used to defeat the encryption in place on DVD movies. So the DVD Copy Control Association, the Motion Picture Association of America and other organizations did their level best to litigate, threaten, and cajole the DeCSS source code out of existence. The chief reward for these efforts was the spread of the code far and wide, even onto the back of a t-shirt – an unfortunate effect of the seemingly endless fount of information the modern age enjoys.

Why be able to delete data?

Aside from attempts to destroy opposing ideas, data deletion can be vital in some situations – such as keeping technology and intelligence out of the hands of enemies. Once data is released, you can't put the genie back in the bottle. Whether it's the recipe for Kentucky Fried Chicken or detailed plans for nuclear weapons, resealing the information is almost impossible. Military cases are some of the more obvious ones where it might be better to destroy data rather than allow it to be exposed. For example, in April 2001 a US EP-3 reconnaissance plane, designated PR-32, collided over the South China Sea with a Chinese F-8 that was performing maneuvers at an extremely close range. The collision caused significant enough damage that the crew was told to prepare to ditch the aircraft. The crew then began jettisoning classified material and smashing equipment with an axe and other hard objects; upon landing they shredded more remaining papers. A later report found it highly probable that the People's Republic of China ended up having access to an amount of undestroyed classified material.

Ensuring that data can be securely deleted is important in any number of situations, some legitimate, some illicit. Several examples of failure to delete data have taken place in recent history. In May 2006, lawyers for AT&T filed a legal brief as part of a class-action lawsuit brought by the Electronic Frontier Foundation against AT&T for its part in the NSA communication surveillance program². The document, delivered in PDF format, contained sensitive information which had been electronically redacted by placement of black lines over sensitive portions. Unfortunately, the method used for redaction did not work equally well across all software used to read the PDF file, and in some cases allowed the redacted information to be copied and pasted into another area where it was readable.

To illustrate that this sort of thing is not restricted to PDF documents, another incident dealt with a failure to delete information properly from Microsoft Word. In February 2003 the British government produced a report on intelligence organizations in Iraq, which Colin Powell subsequently referred to while addressing the United Nations. Later examination of the document showed that ten of nineteen pages had been copied without permission, errors and all, from the work of a student at the Monterey Institute of International Studies³. The potential for failures such as these to wreak havoc is enormous, ranging from public embarrassment to loss of life. The advance of computing technology always seems to outpace the ability of its users to wield it properly.

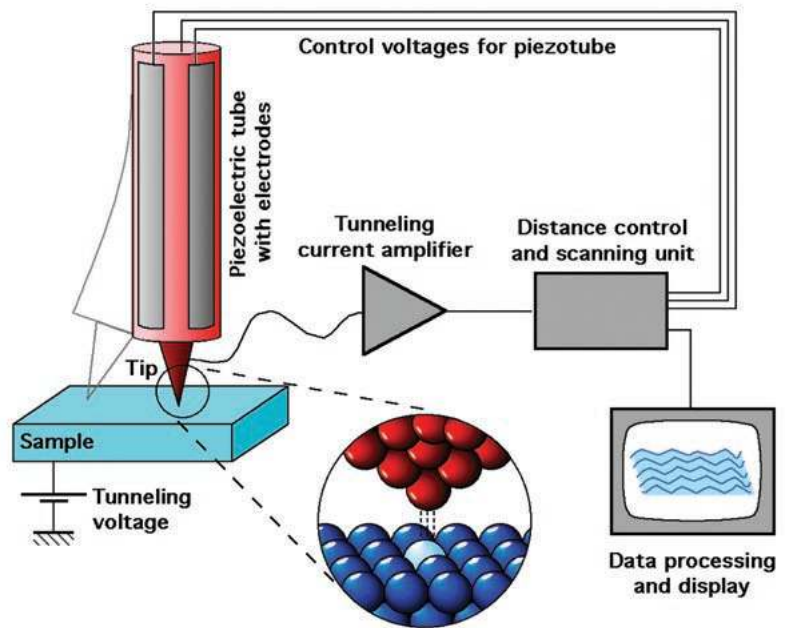


Figure 1 – Schematic view of a scanning tunneling microscope (STM) – Courtesy of Michael Schmid under Creative Commons Attribution ShareAlike 2.0 Austria License.

Early research

Work in digital data recovery and data deletion occurred in the 1980s to 1990s, eons ago in the age of digital computers. The primary motivation for it appears to have been just a glimmering of the data security needs that are commonplace today. Data storage was beginning to really take off, and perhaps a few prescient individuals began to see the need for research and exploration in these areas. Much of the early exploration was centered on magnetic data storage. A 1993 paper discussed the use of a scanning tunneling microscope (STM) to measure the fields on magnetic media in order to understand the conditions allowing overwritten data to remain readable⁴. Another paper discusses using laterally shifted signals in the recording process to overwrite previously written data⁵. The authors assert that this method almost entirely demagnetized portions of the media being tested. Around the same time, research was published on the use of an STM to display images of the actual bits written to a computer disk drive⁶; and the NSA patented a method to measure magnetic fields on magnetic media using an STM and magnetic probe. (See Figure 1.)

Peter Gutmann at the University of Auckland penned two of the major papers in the field of data deletion and data recovery. In 1996 he provided the basis for most of the modern work done in the field of secure data deletion, in his discussion of the (then unworkable) use of microscopy for data recovery from magnetic media⁷. He also discusses data remanence outside of the normal path used by the drive head. Another major portion of the paper covers erasing data. Further, Gutmann introduces the concepts of overwriting data with specific patterns to defeat recovery, and of using specific algorithms

2 "AT&T leaks sensitive info in NSA suit." *CNET News*. http://news.com.com/2100-1028_3-6077353.html?part=rs&tag=6077353&subj=new% s

3 "Microsoft Word bytes Tony Blair in the butt." *Computer Bytes Man* (2003). <http://www.computerbytesman.com/privacy/blair.htm>

4 R. D. Gomez et al., "Microscopic investigations of overwritten data." *Journal of Applied Physics*, Vol. 73, No. 10 (May 1993). 6001-6003

5 A. A. Adly et al., "Computation of magnetic fields in hysteretic media." *IEEE Transactions on Magnetics*, Vol. 29, No. 6 (1993). 2380-2382

6 P. Rice and J. Moreland, "Tunneling-stabilized magnetic force microscopy of bit tracks on a hard disk." *IEEE Transactions on Magnetics*, Vol. 27, No. 3 (1991). 3452-3454

7 P. Gutmann, "Secure deletion of data from magnetic and solid-state memory." *Sixth USENIX Security Symposium, Focusing on Applications of Cryptography*, 1996

to provide the most thorough magnetic interference on the existing data. He also goes over related topics, including degaussing, encryption, and data recovery and deletion from RAM.

In a 2001 paper Gutmann's focus changed in the direction of semiconductor storage devices such as RAM, EEPROM (electrically erasable programmable read-only memory) chips, and flash memory⁸. Portions of the paper not given over to background material on semiconductors and semiconductor physics do pertain directly to data deletion and recovery. Of particular interest is the section on minimizing the recoverability of data in RAM. The primary methods for doing so may be stated as:

- **Short-term retention** – Storing the data for as short a time as possible
- **Long-term retention** – Ensuring that a variety of data is stored, without any particular item being stored more often

Similar to the issues with RAM are issues with EEPROM and flash memory. Due to the limited number of write/erase cycles possible with these technologies, data can become trapped in storage areas no longer in use. Also, wear-leveling technologies designed to extend the life of such devices may cause multiple copies of a given file to exist in several locations. Such technologies may further defeat the technique of multiple overwriting of data storage locations, since each write may actually take place in a different area of storage.

It is unfortunate that the extreme data recovery measures described in Gutmann's papers are not often required in the quest for lost data, due largely to the method many operating systems use to delete files.

How operating systems delete 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 events are similar. When a file is deleted, the pointer to the file is removed and the area where it is stored is marked as being available for use. In other words, the file still exists, almost entirely untouched, on the medium. For example, the sequence of events for file deletion in the venerable NTFS file system works in the following way:

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. Remove the entry for the file from the file index
4. Mark the clusters in which the file is stored as unallocated⁹

The file still exists unscathed on the medium. 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.

Data recovery

Data recovery can be partitioned into two main categories: hardware and software. Most software-based methods involve manipulating the data stored on media though the use of the operating system itself or through the use of an application. The use of software-based

methods is limited by the capabilities of the operating system to read the drive in a normal fashion, thus limiting the data that can be recovered. Hardware-based methods usually involve physically manipulating the hardware in such a fashion as to recover the data it contains. These methods tend to require specialized hardware and/or software, but are capable of recovering considerably more data than software-based methods.

Software methods

Data recovery, which can be defined as the retrieval of information beyond the reach of the average computer user, is considered 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 such tasks. Other companies will attempt to recover data beyond the reach of these utilities, or from media that has become damaged, for much more than the price of software tools.

Data recovery utilities are generally most useful immediately after a file has been deleted, and then only if measures are not in place to prevent recovery of the file. If the operating system lacks 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. Although perhaps more cosmetically appealing and user-friendly, these tools are generally not much more sophisticated in operation than some of the simpler tools. If all of the above fails and the data is of sufficient importance, the next recourse is usually a commercial data recovery service.

Hardware methods

The hardware methods used to recover data are all based around microscopy of some variety. Their effectiveness has increased from theoretical to feasible in the last few years. First, several varieties of microscopy that provide a high level of detail can be used to examine the individual bits on magnetic media. Some of these are magnetic force microscopy, atomic force microscopy, and scanning tunneling microscopy. Although such devices could promise a much more thorough means for data recovery than could software-based methods, they were not useful for the recovery of data en masse. The equipment needed for such an undertaking eventually became inexpensive enough for hobbyists to build. But some difficulty lay in the coordination of what basically amounted to taking graphical images of the entire surface or surfaces of the storage medium, and translating these images back into usable bits.

Another option is spin stand imaging, a recently developed technology for recovering data from magnetic media, similar to magnetic microscopy techniques. Spin stand imaging involves removing the platters from a hard drive and mounting them on a spin stand. A digital sampling oscilloscope is then used to image the surface of the platter while the platter is simultaneously rotated on the spin stand, making it possible to image the entire surface of the platter¹⁰.

Data deletion standards

Many of the commercial tools intended for data deletion refer to one or more sets of standards. These standards, many of them governmental, specify the necessary steps for deleting data from a given medium, beyond recoverability. The final method of surety for many types of media is physical destruction.

8 P. Gutmann, "Data remanence in semiconductor devices." *10th USENIX Security Symposium*, 2001

9 Brian Carrier, *File System Forensic Analysis*. Addison-Wesley Professional, 2005

10 C. Tse et al., "Whole-track imaging and diagnostics of hard disk data using the spin-stand imaging technique." *Journal of Applied Physics*, 97 (2005). 104-107

Most of the standards that have been developed by the US Department of Defense (DoD) have their origins in NISPOM, which also provides the basis for the US Army, Navy and Air Force standards. These, and the commonly referenced sets of standards used in Canada and Germany, provide documentation on the data deletion processes the organizations have in place. See the References at the end of this article for more information.

Data deletion, as with data recovery, breaks down into two main categories: hardware and software. Most software-based methods involve manipulating the data stored on the media through use of the operating system itself or via an application. These methods generally leave the storage media intact and available for future use. Hardware-based methods usually involve physically manipulating the hardware in such a fashion as to erase the data that it contains, or render it unable to store data. Many hardware deletion methods, if they do not destroy the storage media entirely, do not leave them in a functional state.

Software methods

Software data deletion methods operate by either encryption or overwriting. While on the surface it is not actually secure data deletion, encryption becomes so to anyone without the decryption key. The ext3cow file system implements one such method, which amounts to encrypting each individual block of data and destroying the encryption key^{11 12}.

Overwriting data basically amounts to writing over existing data with new data in order to prevent the original data from being recovered. The DoD recommendation for this process is to overwrite a specific storage location at least three times. (Gutmann furthers this by proposing to overwrite each bit with a complementary bit three times, on the supposition that this will more efficiently distort the magnetic field.) There are any number of expansions and extensions of this type of technique, including building overwrites into the normal file system deletion process¹³, and further, built-in secure deletion on a block-by-block level¹⁴.

Hardware methods

Hardware data deletion methods break down into three major categories: degaussing, physical destruction, and the use of unusual media types. Degaussing, or severely disrupting the magnetic field of magnetic storage media, is a generally accepted means of securing data deletion. The NISPOM and other standards specify the exact type of degausser that must be used for each type of media. Additionally, the NSA Central Security Service has a degausser products list specifying manufacturers, models, and appropriate usage for secure deletion on specific media. A recently announced means of de-



Figure 2 – Researchers image a hard disk drive platter using magnetic force microscopy (MFM) – Georgia Tech Photo: Gary Meek

gaussing, more effective than previous methods, uses high-strength neodymium-iron-boron permanent magnets to produce magnetic fields comparable to those found in MRI machines¹⁵. (See Figure 2.)

Physical destruction of media is generally accepted as a sure method of securely deleting data. This category includes such methods as shredding, incinerating, submerging in acid, blunt force and degaussing. Some have even taken the task as far as to melt hard drives and pour them into ingots. Unfortunately, executing this type of method generally requires equipment like an industrial shredder, and entirely precludes reuse of the media, thus raising the cost of the process.

Due to concerns over secure data deletion in the last decade, certain types of storage media have been specifically designed with deletion in mind. Two examples of such unusual media types are those incorporating the secure erase feature, and flash memory-based media. Secure erase (SE) was added to the ATA and SCSI specifications in 2002. With this feature the drive can be issued a command to erase all areas of the drive capable of containing user data¹⁶. While this is convenient, it will reportedly still require an extended period of time to run, potentially several hours – as do the current software implementations of this function.

Flash media-based storage has the benefit of being able to securely delete data from flash memory, if not destroy the memory entirely, very quickly. Manufacturers claim to be able to perform either operation in a matter of seconds, as opposed to a matter of hours for traditional media¹⁷.

Solid state storage and emerging technologies

As digital storage technology continues to progress, many new avenues of research are sure to surface in the area of secure data deletion.

11 <http://www.ext3cow.com/>

12 Z. Peterson et al., "Limiting liability in a federally compliant file system." *Privacy, Obligations, and Rights in Technologies of Information Assessment (PORTIA)*, 2004

13 N. Joukov and E. Zadok, "Adding secure deletion to your favorite file system." *3rd International IEEE Security in Storage Workshop*, 2005

14 M. Sivathanu et al., "Life or death at block-level." *Sixth Symposium on Operating Systems Design and Implementation*, 2004

15 J. Toon, "Protecting Sensitive Data: Researchers develop fail-safe techniques for erasing magnetic storage media." *Georgia Tech Research News*, June 12, 2006. <http://gtresearchnews.gatech.edu/newsrelease/erase.htm>

16 G. Hughes and T. Coughlin, "Secure erase of disk drive data" (2002). <http://www.tomcoughlin.com/techpapers.htm>

17 P. Wilkinson, "Protecting military data in a flash." *COTS Journal*, 2005

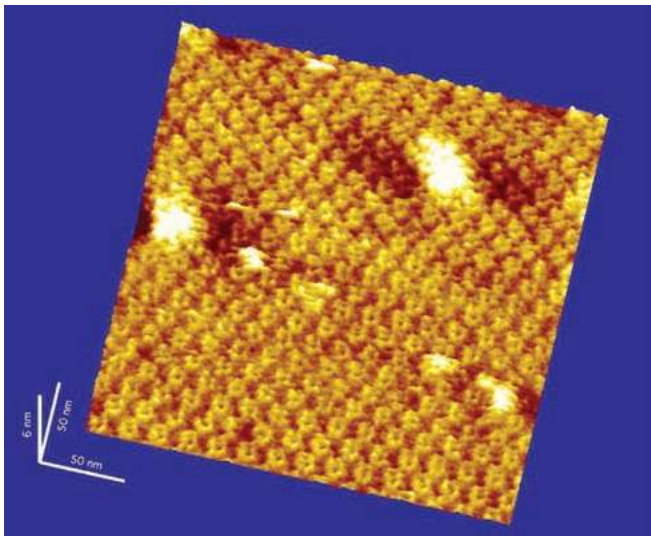


Figure 3 – Atomic force microscope (AFM) image of bacteriorhodopsin obtained in solution – Courtesy of Zhifeng Shao and D.M. Czajkowsky, University of Virginia

In the last few years alone, the large-scale adoption of flash media for data storage has changed the way the security industry looks at data portability. Any attempt to descry the future of computing technology can go wildly astray; however, some larger areas do present themselves for speculation.

Solid state storage, at present commonly appearing in the form of flash media, enjoys very heavy use by the computing public – and brings a variety of issues related to the security of the data it holds. Several opportunities arise here for research in the area of secure data deletion:

- Securely deleting data from flash media without undue wear on the media
- New methods of wear-leveling which do not leave extraneous copies of data behind
- Practical hardware data recovery from semiconductor storage

A host of new storage technologies appear and disappear on a seemingly daily basis. Some, such as Bacteriorhodopsin 192-based storage, a technology based on genetically engineered photosensitive bacteria, hold the promise of entirely new types of optical data storage¹⁸. (See Figure 3.) Whatever forms these new technologies may take, there are sure to be security problems to accompany them. How exactly does one securely erase data from bacteria, for instance?

Summary

Given the enormous body of digitally stored data that exists today, ensuring the protection and proper disposal of information is paramount. Current methods of secure data deletion can prove to be challenging. Most of these methods have one or more aspects in which they perform well, but none of the current methods for 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 to be followed by new methods of penetrating such techniques. Ultimately, if the data is impor-

tant enough to merit the expense, any method of data hardening can be breached or circumvented entirely.

References

Military and Governmental Data Deletion Standards:

US Department of Defense. *National Industrial Security Program Operating Manual (NISPOM)*, April 2004

US Department of the Air Force. *Air Force Instruction 33-202*, August 2002

US Department of the Army. *Army Regulation 380-19*, February 1998

US Department of the Navy. NAVSO P-5239-26. *Remanence Security Guidebook, Module 26, Information Systems Security (Infosec) Program Guidelines*, September 1993

Royal Canadian Mounted Police. *Information Technology Security Guide*, October 2003

Bundesamt für Sicherheit in der Informationstechnik. "Verwaltungsvorschrift des Innenministeriums zum Geheimschutz von Verschlusssachen beim Einsatz von Informationstechnik" (VSITR), Innenministeriums, 2004

About the Author

Jason Andress is a tinkerer, rascalion, and all around geek. He is working on his doctorate in Computer Science, teaches graduate and undergraduate CS courses, and enjoys a good game of Scrabble. He can be reached at jason.andress@gmail.com.

¹⁸ V. Renugopalakrishnan et al., "BioFold's high performance bacteriorhodopsin 192-based storage devices." *Insight Spring*, 2002