



THE JOURNAL OF
**DIGITAL FORENSICS,
SECURITY AND LAW**

**Journal of Digital Forensics,
Security and Law**

Volume 7 | Number 4

Article 1

2012

The Science of Digital Forensics: Recovery of Data from Overwritten Areas of Magnetic Media

Fred Cohen

Fred Cohen & Associates

Follow this and additional works at: <https://commons.erau.edu/jdfsl>

 Part of the [Computer Engineering Commons](#), [Computer Law Commons](#), [Electrical and Computer Engineering Commons](#), [Forensic Science and Technology Commons](#), and the [Information Security Commons](#)

Recommended Citation

Cohen, Fred (2012) "The Science of Digital Forensics: Recovery of Data from Overwritten Areas of Magnetic Media," *Journal of Digital Forensics, Security and Law*: Vol. 7 : No. 4 , Article 1.

DOI: <https://doi.org/10.15394/jdfsl.2012.1131>

Available at: <https://commons.erau.edu/jdfsl/vol7/iss4/1>

This Article is brought to you for free and open access by the Journals at Scholarly Commons. It has been accepted for inclusion in Journal of Digital Forensics, Security and Law by an authorized administrator of Scholarly Commons. For more information, please contact commons@erau.edu.

EMBRY-RIDDLE
Aeronautical University™
SCHOLARLY COMMONS

(c)ADFSL



THE SCIENCE OF DIGITAL FORENSICS: RECOVERY OF DATA FROM OVERWRITTEN AREAS OF MAGNETIC MEDIA

Fred Cohen

Seeking to understand the state of scientific consensus surrounding an area of forensics is often problematic. You cannot prove a claim about an unlimited set in the sense of testing every possibility. And yet there is a level of consensus surrounding the science of the day.

This is a slightly altered portion of an expert report I wrote recently, released with permission, that I thought might serve as an example of how to go about seeking the truth and presenting the state of the science when truly definitive statements based on first principles are not available. Of course, I look forward to the readership proving me wrong with real-world examples, but somehow, I doubt if I will find any.

INTRODUCTION

The first time I encountered data loss and recovery effects of magnetic memory was as a night and weekend computer operator for the computer science department of Carnegie-Mellon University in the 1973-1974 time frame. Part of my job involved dealing directly with outages and failures associated with magnetic memory components used in what, at the time, were large computer systems. On occasions, portions of magnetic core memory or disk drives would encounter various failure modes and the systems using these devices would have to be reconfigured to operate without the failed components until repair personnel could come in to repair them, typically during normal business hours on weekdays. In the early hours of one Sunday morning, I was having such problems with a magnetic core memory module (a cabinet about 6 ft. high and 3 ft. across), and after awakening the manager in charge was instructed to restart the memory and continue the operation of the computer, setting a particular value into a particular memory location to cause the system to continue operation. After several such incidents within a period of less than an hour, a more definitive outage was produced after a mechanical impulse was applied to the cabinet, the memory was reconfigured out of the system, the system operated at reduced memory until the next weekday, and no further outages were experienced.

The next time I encountered a similar incident involving magnetic memory loss and recovery was as a systems administrator in the early 1980s while I was a graduate student at the University of Southern California. A VAX computer I

was tasked with maintaining encountered a problem associated with an inability to restore from and access backup tapes after regular maintenance on the tape drives was completed. While newly written tapes were readable, tapes written from before the maintenance were not readable. I determined that the tape head alignment was different after the maintenance than before the maintenance, and set about to realign the tape head by mechanically adjusting a set screw while continuously reading from the tape and displaying the output to a screen until the output reflected output reasonably expected from the tape being tested. I then restored and rewrote the tapes as appropriate to reflect the proper alignment and, in doing so, recovered lost data associated with the difference in head alignment.

I continued to track progress in this area over the years by helping to create standardized approaches to dealing with the life cycle of data, including identifying and summarizing existing standards¹ like DoD 5200.28², 5200.28-M³, and so forth.

In the late 1990s, I increasingly worked on issues related to digital forensics, security, and countering security measures. Included in these issues was work in recovery of data involved in forensic investigations, including recovery from data and media disposed of according to different practices. This included writing software to recover data, identifying issues related to the destruction of data to various levels of surety, and recovery of data from such destruction processes. Experiments involving various media were undertaken as part of my research at Sandia National Laboratories, including rapid destruction techniques, and recovery from data destroyed with such techniques. I was also involved in various efforts to recover data in investigations for private concerns, in systems used for national security, in matters involving law enforcement, for private individuals, and in other similar situations. As part of that work, I identified the available methods for data destruction and recovery, life cycle issues associated with systems and data, and related issues involving media of various sorts. I have also studied, written a peer reviewed research paper, and given scientific presentations on methods of recovery of data from disk drives for forensic applications, including recoveries used in legal matters.

1 See: <http://all.net/books/standards/remnants/standards.html>

2 See: <http://csrc.nist.gov/groups/SMA/fasp/documents/c&a/DLABSP/d520028p.pdf>

3 See: <http://biotech.law.lsu.edu/blaw/dodd/corres/pdf2/p520028m.pdf> p520028m.pdf (1973)

HISTORICAL METHODS

According to information from 1996⁴, as of that time:

“In the 1980's some work was done on the recovery of erased data from magnetic media, but to date the main source of information is government standards covering the destruction of data. There are two main problems with these official guidelines for sanitizing media. The first is that they are often somewhat old and may predate newer techniques for both recording data on the media and for recovering the recorded data. For example most of the current guidelines on sanitizing magnetic media predate the early-90's jump in recording densities, the adoption of sophisticated channel coding techniques such as PRML, the use of magnetic force microscopy for the analysis of magnetic media, and recent studies of certain properties of magnetic media recording such as the behavior of erase bands. The second problem with official data destruction standards is that the information in them may be partially inaccurate in an attempt to fool opposing intelligence agencies (which is probably why a great many guidelines on sanitizing media are classified). By deliberately under-stating the requirements for media sanitization in publicly-available guides, intelligence agencies can preserve their information-gathering capabilities while at the same time protecting their own data using classified techniques.” [P4]

This paper described coding issues with overwriting of disk media and the notional methods of reading from areas imprecisely “seek”ed on disk and through magnetic force microscopy (MFM) and magnetic force Scanning Tunneling Microscopy (STM). The conclusion at that time was that multiple overwrites with different patterns were required to eliminate actual residual data physically present in the form of flux density variances detectable with analogue methods, and that because of variances, even this would be inadequate with MFM or STM.

In the same time frame, DoD practices⁵ indicated:

“Overwriting is a process whereby unclassified data are written to storage locations that previously held sensitive data. To satisfy the DoD clearing requirement, it is sufficient to write any character to all data locations in question. To purge the AIS storage media, the DoD requires overwriting with a pattern, then its complement, and finally with another pattern; e.g., overwrite first with 0011 0101, followed by 11001010, then 1001 0111. The number of

4 Peter Gutmann, “Secure Deletion of Data from Magnetic and Solid-State Memory”, Sixth USENIX Security Symposium Proceedings, San Jose, California, July 22-25, 1996

5 “A Guide to Understanding Data Remanence in Automated Information Systems”, NCSC-TG-025 - Library No. 5-236,082 – Version-2, Section 5: “Standards”.

times an overwrite must be accomplished depends on the storage media, sometimes on its sensitivity, and sometimes on differing DoD component requirements. In any case, a purge is not complete until a final overwrite is made using unclassified data.”

And below that:

“5.2.2 MAGNETIC HARD DISKS

The DoD has approved both overwriting and degaussing as methods to clear or purge this media. See Section 4, "Risk Considerations," and DoD 5200.28-M for additional information.”⁶[P17]

Section 4 indicates, in pertinent parts:

“4.4 STORAGE DEVICE SEGMENTS NOT RECEPTIVE TO OVERWRITE

A compromise of sensitive data may occur if media is released when an addressable segment of a storage device (such as unusable or "bad" tracks in a disk drive or inter-record gaps in tapes) is not receptive to an overwrite. As an example, a disk platter may develop unusable tracks or sectors; however, sensitive data may have been previously recorded in these areas. It may be difficult to overwrite these unusable tracks. Before sensitive information is written to a disk, all unusable tracks, sectors, or blocks should be identified (mapped). During the life cycle of a disk, additional unusable areas may be identified. If this occurs and these tracks cannot be overwritten, then sensitive information may remain on these tracks. In this case, overwriting is not an acceptable purging method and the media should be degaussed or destroyed.

4.5 OVERWRITE SOFTWARE AND CLEARING

Overwriting is an effective method of clearing data. In an operational system, an overwrite of unassigned system storage space can usually accomplish this, provided the system can be trusted to provide separation of system resources and unauthorized users. For example, a single overwrite of a file (or all system storage, if the circumstance warrants such an action) is adequate to ensure that previous information cannot be reconstructed through a keyboard attack. Note: Simply removing pointers to the file will not generally render the previous information unrecoverable. Software used for clearing should be under strict configuration controls. ...

4.6 OVERWRITE SOFTWARE AND PURGING

The DoD has approved overwriting and degaussing for purging data, although the effectiveness of overwriting cannot be guaranteed without examining each application. If overwriting is to be used in a specific application, software

6 “A Guide to Understanding Data Remanence in Automated Information Systems”, NCSC-TG-025 - Library No. 5-236,082 – Version-2, Section 5: “Standards”.

developers must design the software such that the software continues to write to all addressable locations on the media, in spite of intermediate errors. All such errors in usable sectors should be reported with a listing of current content. In addition, unusable sectors must be completely overwritten, because the unusable sector list will not show whether the sector ever contained any sensitive data. If any errors occur while overwriting or if any unusable sector could not be overwritten, then degaussing is required.” [P14]

METHODS FOR RECOVERY OF OVERWRITTEN HARD DRIVE DATA AFTER 2001

Between the late 1990s and the middle 2000s, I was unaware of any changes in the status of up-to-date hard drives with regard to data recovery. However, I also participated in many forums related to data recovery in forensics cases, and was aware of many of the methods in use. During the time starting from about 2000, I don't recall any instance of hearing or reading about recovery of data from overwritten areas of hard drives. While many technology changes were underway and methods for recovery were in use, none of these recovery methods involved recovering data from areas of a disk that had been previously overwritten.

In 2006, the National Institute of Standards and Technology stated the following in this regard in its guidelines for media sanitation (Page 6):⁷

“Advancing Technology has created a situation that has altered previously held best practices regarding magnetic disk type storage media. Basically the change in track density and the related changes in the storage medium have created a situation where the acts of clearing and purging the media have converged. That is, for ATA disk drives manufactured after 2001 (over 15GB) clearing by overwriting the media once is adequate to protect the media from both keyboard and laboratory attack.”

In other words, in 2006, the National Institute of Standards and Technology asserted that overwriting a disk drive of this sort once is adequate to make the overwritten data unrecoverable by any known methods. Keyboard attack in this context means any command issued from a computer, including those that actuate specific hardware mechanisms of specific disk drives. Laboratory attack ranges to the use of analog hardware, electron microscopy, special processing of

7 Richard Kissel Matthew Scholl Steven Skolochenko Xing, NIST Special Publication 800-88 "Guidelines for Media Sanitization: Recommendations of the National Institute of Standards and Technology", Computer Security Division Information Technology Laboratory National Institute of Standards and Technology Gaithersburg, MD 20899-8930, included by reference herein as iNISTSP800-88_rev1.pdf [Exhibit 13]

materials, and other similar techniques.

The best available technique I am aware of published since 1996 indicates only a 1 in 10,000 chance of being correct when trying to determine a previous value for any given overwritten bit (1 or 0 value) of data based on looking at residual content after only a single overwrite of the corresponding disk area.⁸ Quoting:

The basis of this belief is a presupposition is that when a one (1) is written to disk the actual effect is closer to obtaining a 0.95 when a zero (0) is overwritten with one (1), and a 1.05 when one (1) is overwritten with one (1). This we can show is false and that in fact, there is a distribution based on the density plots that supports the contention that the differential in write patterns is too great to allow for the recovery of overwritten data. ...

Consequently, we can categorically state that there is a minimal (less than a 0.01% chance) of recovering any data on a NEW and unused drive that has a single raw wipe pass (not even a low-level format). In the cases where a drive has been used (even being formatted for use) it is not possible to recover the information – there is a small chance of bit recovery, but the odds of obtaining a whole word are small.

To put this in context, in trying to extract a single byte of data (8 bits often associated with a single ASCII character in a document) using the identified techniques, the chances of being correct in such an extraction is 1 in 10^{32} better with this technique than with random guessing. In practice, no meaningful data recovery is feasible.

In order to seek additional information, I made requests for such information on a variety of online forums where members of various digital forensics communities communicate. I asked for any example refuting the results of the Wright et. al. Paper. To date, I have found no example of any instance in which digital data recorded on a hard-disk drive and subsequently overwritten was recovered from such a drive.

It is my opinion based on the information I have been able to discern, that any distinctions in terms of the ability to recover overwritten data between overwriting modern hard-disk drives one time, several times, while skipping tracks back and forth, and/or by other similar methods, are distinctions without any practical difference.

In 2007, as a result of work in a legal matter, and based on prior work including

8 Wright, Craig; Kleiman, Dave; Sundhar R.S., Shyaam, “Overwriting Hard Drive Data: The Great Wiping Controversy”, in “Information Systems Security: 4th International Conference, ICISS 2008”. Lecture Notes in Computer Science (Springer-Verlag New York, LLC) , 2008 December; 5352: 243-257.

prior work in recovery of data from a floppy disk with analog methods,⁹ I ended up trying to recover data from an old floppy disk through purely electronic means, not involving an electron microscope or similar special purpose equipment. In this case, I ended up co-authoring a paper about data recovery in such situations where the data was no longer available because of the loss of magnetic flux density over time and through various wear and degradation mechanisms.¹⁰

My approach involved a combination of adjusting the seek location of read heads through the use of set point screws on multiple floppy drives and multiple reread attempts until a read succeeded. The rereading approach essentially generates multiple values for data that is degraded to the point where it probabilistically yields a 1 or 0 after processing, and uses the cyclic redundancy (CRC) codes of the disk drive to ignore result until one comes out with content matching the CRC code read. This works for weak bits (cases where the residual data is very nearly adequate to trigger a proper read), but is of no use in cases of overwrite, where the data last written is essentially always a stronger signal than the previous data it overwrote. Analysis was then performed to determine the likelihood of a wrong read coming up as valid and identifying what, if any, changes could occur to produce a valid read based on loss of signal strength rather than overwrite.

Then, on or about 2008-12-05, I was contacted by a researcher I had known via communications in forensics-related Internet forums since at least 2005. He sent me a copy of a draft of an about to be published paper titled “Overwriting Hard Drive Data: The Great Wiping Controversy”¹¹ in a message indicating “I am presenting the attached paper in a couple weeks. I still find it difficult to believe that nobody decided to test Dr. Gutmann's supposition that you could recover data using an electron microscope.”¹²

After review, I asked several questions and got replies that satisfied me with regard to the issues in the present discussion. He replied to a request of mine

9 Hans-Joachim Leimkueller, “Computer Evidence Analysis and Recovery of Magnetic Storage Media Data”, 1995. Proceedings. Institute of Electrical and Electronics Engineers 29th Annual 1995 International Carnahan Conference on Security Technology, 18-20 Oct 1995IEEE, 1995.

10 F. Cohen, and Charles M. Preston, “A Method for Recovering Data From Failing Floppy Disks with a Practical Example”, IFIP TC11 presented Jan 2008, Published in “Advances in Digital Forensics IV”, Springer, ISBN 978-0-387-84926-3, pp29-42, 2008.

11 See Wright above for the final version.

12 Personal email not included in this editorial paper.

asking for a formal citation indicating that it would appear soon.¹³

“The publication is ICISS 2008, LNCS 5352 (Pp 243-257)

<http://www.springerlink.com/content/408263q111460147/?p=650ee5e3e45d4e1e845e>

2bfe8a959f1a&pi=20

Information Systems Security

4th International Conference, ICISS 2008, Hyderabad, India, December 16-20, 2008, Proceedings

Series: Lecture Notes in Computer Science

Subseries: Security and Cryptology , Vol. 5352

Sekar, R.; Pujari, Arun K. (Eds.)

2008, XIII, 307 p., Softcover

ISBN: 978-3-540-89861-0

I am presenting the paper in about 10 days. The conference paper is:

<http://www.seclab.cs.sunysb.edu/iciss08/>”

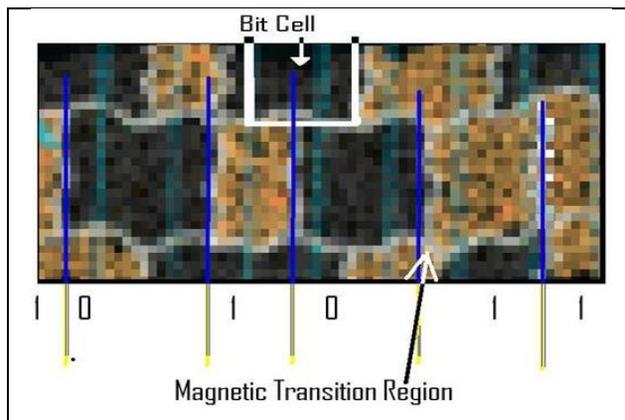
A copy of that publication is included herewith, and it is consistent with the previous copies sent to me and subsequent disclosures made to me and included herein as described below.¹⁴ Our subsequent correspondence later yielded additional data including:

“Hi

I have the data for the same regions already. In this case it is just the issue of processing this into the images. At present I have produced distribution plots when writing 1s then 1s, 0s then 0s, 0s then 1s, and 1s then 0s produce 0's and 1s...

¹³ Personal email not included in this editorial paper.

¹⁴ R. Sekar and A.K. Pujari (Eds.): ICISS 2008, LNCS 5352, pp. 243–257, 2008.



As for "depth", this comes more to the intensity of the media. When you look at figure 2 (above) the best you get is the intensity of each grain in the bit cell. This is as much as you can read.

Each grain varies within the bit cell. If you look at the image above, though the intensity varies, there is no information that tells you anything of value in data recovery. For instance, the 4th bit cell across from the left (a "0") has a comparatively high intensity region on both junctions of the cell (the left and right walls).

What you get is that the depth idea is magic rather than science. You can see the cumulative charge, not the underlying grains. In the next generation of drives this all becomes moot. As the write goes to a single particle domain, there is no off track data in existence, but for the time being, we still have issues coming from the nature of the bit cell composition and the grain structure that forms it. ..."

He included two other files¹⁵ with this email¹⁶ and those files supported the specific contentions made.

Among the results was an attempt to recover data using the best available methods, both from a disk which was overwritten 3 times with 0 byte values and from a disk that was written once with 0 values before writing only the desired data, and for which the location of the data on the disk was known in advance. The following extract from that paper shows the extent to which the best available methods in 1995 we were able to recover data from areas of a hard disk drive overwritten once.

The "Correct display" section represents the known good data originally placed

¹⁵ Personal correspondence not included in this editorial paper

¹⁶ Personal correspondence not included in this editorial paper

use of forensic evidence associated with reading overwritten areas of hard disk drives. As disk drives have continued to become higher density, newer methods used to write with more efficient low-level coding, and areas where bits are stored in physical form have become increasingly modularized on the media, such recovery has become infeasible by any known method.

In or about 2011, Peter Gutmann updated his paper from July 22-25, 1996 to reflect the changes in technology leading to the infeasibility of recovering data from modern disk drives once overwritten. In pertinent parts, it indicates:

“Looking at this from the other point of view, with the ever-increasing data density on disk platters and a corresponding reduction in feature size and use of exotic techniques to record data on the medium, it's unlikely that anything can be recovered from any recent drive except perhaps a single level via basic error-cancelling techniques. In particular the drives in use at the time that this paper was originally written are long since extinct, so the methods that applied specifically to the older, lower-density technology don't apply any more. Conversely, with modern high-density drives, even if you've got 10KB of sensitive data on a drive and can't erase it with 100% certainty, the chances of an adversary being able to find the erased traces of that 10KB in 200GB of other erased traces are close to zero.

Any modern drive will most likely be a hopeless task, what with ultra-high densities and use of perpendicular recording I don't see how MFM would even get a usable image, and then the use of EPRML will mean that even if you could magically transfer some sort of image into a file, the ability to decode that to recover the original data would be quite challenging.”¹⁷

Furthermore, other methods of recovery, such as reading with analogue devices or varying the alignment of disk heads all depend on the presence of the same underlying mechanisms as are identified by the aforementioned methods, and thus the methods identified here are the more definitive in terms of determining feasibility of recovery of overwritten data.

It is my view at this time that the open scientific community dealing with digital forensics now accepts these results in practical terms, that erasure of modern hard disk drives with a single overwrite is adequate to render the overwritten data unrecoverable by any known methods. This is also reflected in subsequent queries of members of the global digital forensics community.

QUERIES FOR INSTANCES REFUTING THE WRIGHT ET. AL.

17 Peter Gutmann, “Secure Deletion of Data from Magnetic and Solid-State Memory”
retrieved from http://www.cs.auckland.ac.nz/~pgut001/pubs/secure_del.html on
2012-05-27

RESULTS

In order to seek out refutations of these results from the historical record and personal knowledge of others, I sent the following message:

“I am looking for any actual cases where an overwritten portion of a hard disk drive was recovered using any method. Does anyone know of any such case? - Email me directly and I will summarize... FC

To be clear, I know of a case where data loss over time was recovered using analogue or reread techniques, and have heard of (but not seen cited) cases where this was done for floppy disks, and cases where non-overwritten portions of hard drives were recovered, but as far as I can tell, there have never been any cases where areas of hard drives actually overwritten were subsequently recovered.

If anyone knows of such a case, I would appreciate being contacted with details. Please email directly to fc@all.net - and I will summarize for the group.”

To the following online groups:

Computer Security and Forensics (5,130 members),

Current Topics in Digital Investigation Techniques (293 members),

Digital Forensic Certification Board (283 members), Digital Forensics Association (DFA) (4,423 members),

Digital Forensics Research Conference (DFRWS) (499 members),

International Information Systems Forensics Association (1535 members),

Techno Security & Digital Investigations Conference (1215 members), and

Digital Forensics in the Classroom (2450 members).

While there is substantial overlap between these groups, they represent a broad spectrum of individuals with expertise ranging from certified digital forensic practitioners with testifying experience over periods of years to educators who teach these issues in undergraduate and graduate classes, to investigators for law enforcement and private concerns who regularly do digital investigations, to speakers at international conferences from all manner of organizations.

To date I have found no example of any instance in which digital data recorded on a hard disk drive and subsequently overwritten was recovered from such a drive since 1985, when about 15% of the overwritten data was claimed to have been recovered from an modified frequency modulation (MFM) disk drive. Several people have had discussions in order to more clearly understand exactly what the question was, and of course there are data recovery methods that may work when the areas being examined have not been overwritten by other data, such as those used on floppies in the previously cited example, but none of the respondents had any relevant examples.

Based on these results and interactions, I believe that there is a consensus surrounding the irretrievability of overwritten data on modern hard disk drives in the identified communities. Indeed, there appears to be nobody in the identified community that disputes this result with any actual basis and no example of recovery of data from overwritten areas of modern disk drives. The only claims that there might be such a capability are based on notions surrounding possible capabilities in classified environments to which the individuals asserting such claims do not assert they have actual access and about which they claim no actual knowledge.

It is my opinion based on the information I have been able to discern, that any distinctions in terms of the ability to recover overwritten data between overwriting modern disk drives one time, several times, while skipping tracks back and forth, and/or by other similar methods, are distinctions without any practical difference. All of these overwriting methods render data on a modern disk drive unrecoverable by any known methods.

EDITORIAL SUMMARY AND DISCUSSION

It is not my intent to offer this as a peer reviewed paper, or I would have done so. It is not my claim that this represents a model of how to write an expert report either. My point is to seed the clouds of discussion about how to seek the truth from a scientific standpoint.

I recognize that the technology and surrounding science change over time. As a result, the time may come – perhaps very soon – when some new approach yields results allowing such recovery. Perhaps it will be nano-devices that slowly eat away the magnetic surface taking measurements of flux density as they go, and perhaps this will yield amazing results. Perhaps not. That's one of the interesting things about science. But I still have to report on what I can best determine from available data.

The consensus approach is hardly definitive scientific work – but it is a way to check my work against others in the field, and it did yield some clarity about the state of the art in recovery, even if that was not relevant to the specific question at hand.

I look forward to the opinions of others – sent to the editor, to me, or as part of published work. And I hope this will stimulate some discussion around what is “good enough” and what is not.

