

Why overwrite a file more than once to securely delete all traces of a file?

Asked 16 years, 3 months ago Modified 9 years, 5 months ago

Viewed 12k times



Erasing programs such as Eraser recommend overwriting data maybe 36 times.

16



As I understand it all data is stored on a hard drive as 1s or 0s.



If an overwrite of random 1s and 0s is carried out once over the whole file then why isn't that enough to remove all traces of the original file?



security

disk

Share

Improve this question

Follow

edited Jul 1, 2015 at 10:48



Brian Tompsett - 汤莱恩

5,875 ● 72 ● 61 ● 133

asked Sep 12, 2008 at 18:11



kJack

2,014 ● 4 ● 27 ● 42

[How does a 7- or 35-pass erase work? Why would one use these methods?](#) – Graham Perrin Feb 5, 2013 at 3:48

Does it really possible to recover data accurately 90%- 100% from a magnetic drive which overwritten many time 20-30 or more at any cost? Yes or No? please help. – [Govind Totla](#)
Jan 3, 2014 at 13:30

13 Answers

Sorted by:

Highest score (default)



31



A hard drive bit which used to be a 0, and is then changed to a '1', has a slightly weaker magnetic field than one which used to be a 1 and was then written to 1 again. With sensitive equipment the previous contents of each bit can be discerned with a reasonable degree of accuracy, by measuring the slight variances in strength. The result won't be exactly correct and there will be errors, but a good portion of the previous contents can be retrieved.

By the time you've scribbled over the bits 35 times, it is effectively impossible to discern what used to be there.

Edit: [A modern analysis](#) shows that a single overwritten bit can be recovered with only 56% accuracy. Trying to recover an entire byte is only accurate 0.97% of the time. So I was just repeating an urban legend. Overwriting multiple times might have been necessary when working with floppy disks or some other medium, but hard disks do not need it.

answered Sep 12, 2008 at 18:15

**DGentry**

16.3k ● 8 ● 53 ● 66

Thanks, that's explained something that didn't seem to make sense to me before – [kjack](#) Sep 12, 2008 at 18:24

- 2 I'm actually not convinced this is correct information. There are no recorded instances of any recovery from an overwrite that yielded more than 1% of valid data. – [Bruce the Hoon](#) Sep 12, 2008 at 18:35

Government's with lots of money don't tend to publish their results. – [Adam Davis](#) Sep 12, 2008 at 18:41

- 4 Let me leave a better comment. *Researchers* have shown that it's possible to recover significant information from a magnetic disc that has been overwritten more than once. That doesn't mean it's easy, or that it happens regularly, but that *it's possible*, so it's worthwhile to take extra care. – [Adam Davis](#) Sep 13, 2008 at 1:46



4



Daniel Feenberg (an economist at the private National Bureau of Economic Research) claims that the chances of overwritten data being recovered from a modern hard drive amount to "urban legend":

[Can Intelligence Agencies Read Overwritten Data?](#)



So theoretically overwriting the file once with zeroes would be sufficient.

Share Improve this answer

answered Sep 15, 2008 at 21:20

Follow



Ron

41 ● 4



3



In conventional terms, when a one is written to disk the media records a one, and when a zero is written the media records a zero. However the actual effect is closer to obtaining a 0.95 when a zero is overwritten with a one, and a 1.05 when a one is overwritten with a one. Normal disk circuitry is set up so that both these values are read as ones, but using specialised circuitry it is possible to work out what previous "layers" contained. The recovery of at least one or two layers of overwritten data isn't too hard to perform by reading the signal from the analog head electronics with a high-quality digital sampling oscilloscope, downloading the sampled waveform to a PC, and analysing it in software to recover the previously recorded signal. What the software does is generate an "ideal" read signal and subtract it from what was actually read, leaving as the difference the remnant of the previous signal. Since the analog circuitry in a commercial hard drive is nowhere near the quality of the circuitry in the oscilloscope used to sample the signal, the ability exists to recover a lot of extra information which isn't exploited by the hard drive electronics (although with newer channel coding techniques such as PRML (explained further on) which require extensive amounts of signal processing, the use of simple tools such as an

oscilloscope to directly recover the data is no longer possible)

http://www.cs.auckland.ac.nz/~pgut001/pubs/secure_del.html

Share Improve this answer

answered Sep 12, 2008 at 18:26

Follow



dotmad

174 ● 1



2



Imagine a sector of data on the physical disk. Within this sector is a magnetic pattern (a strip) which encodes the bits of data stored in the sector. This pattern is written by a write head which is more or less stationary while the disk rotates beneath it. Now, in order for your hard drive to function properly as a data storage device each time a new magnetic pattern strip is written to a sector it has to reset the magnetic pattern in that sector enough to be readable later. However, it doesn't have to completely erase all evidence of the previous magnetic pattern, it just has to be good enough (and with the amount of error correction used today good enough doesn't have to be all that good). Consider that the write head will not always take the same track as the previous pass over a given sector (it could be skewed a little to the left or the right, it could pass over the sector at a slight angle one way or the other due to vibration, etc.)

What you get is a series of layers of magnetic patterns, with the strongest pattern corresponding to the last data write. With the right instrumentation it may be possible to

read this layering of patterns with enough detail to be able to determine some of the data in older layers.

It helps that the data is digital, because once you have extracted the data for a given layer you can determine exactly the magnetic pattern that would have been used to write it to disk and subtract that from the readings (and then do so on the next layer, and the next).

Share Improve this answer

answered Sep 12, 2008 at 23:09

Follow



Wedge

19.8k ● 7 ● 49 ● 71



2



The reason why you want this is *not* harddisks, but **SSDs**. They remap clusters without telling the OS or filesystem drivers. This is done for wear-leveling purposes. So, the chances are quite high that the 0 bit written goes to a different place than the previous 1. Removing the SSD controller and reading the raw flash chips is well within the reach of even corporate espionage. But with 36 full disk overwrites, the wear leveling will likely have cycled through all spare blocks a few times.

Share Improve this answer

answered Mar 11, 2009 at 12:17

Follow



MSalters

179k ● 11 ● 164 ● 368

Could you expand on this a little? I thought eraser just overwrote the file not the entire disk. 36 overwrites of the entire disk would take far longer than the time I've seen eraser take. – [kjack](#) Mar 13, 2009 at 11:33

- 1 Hmm, seems the developer of "eraser" doesn't understand SSDs then. If you "erase" a single file on SSD 36 times, you'll likely end up with $36 \times \text{sizeof}(\text{file})$ worth of zeroes, AND the original file in other blocks. – [MSalters](#) Mar 23, 2009 at 13:45
-



1



"Data Remanence" There's a pretty good set of references regarding possible attacks and their actual feasibility on [Wikipedia](#). There are DoD and NIST standards and recommendations cited there too. Bottom line, it's possible but becoming ever-harder to recover overwritten data from magnetic media. Nonetheless, some (US-government) standards still require at least multiple overwrites. Meanwhile, device internals continue to become more complex, and, even after overwriting, a drive or solid-state device may have copies in unexpected (think about bad block handling or flash wear leveling ([see Peter Gutmann](#))). So the truly worried still destroy drives.

Share Improve this answer

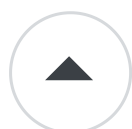
answered Sep 12, 2008 at 18:29

Follow



[Liudvikas Bukys](#)

5,870 ● 3 ● 27 ● 36



1



What we're looking at here is called "data remanence." In fact, most of the technologies that overwrite repeatedly are (harmlessly) doing more than what's actually necessary. There have been attempts to recover data from disks that have had data overwritten and with the



exception of a few lab cases, there are really no examples of such a technique being successful.

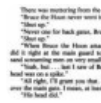
When we talk about recovery methods, primarily you will see magnetic force microscopy as the silver bullet to get around a casual overwrite but even this has no recorded successes and can be quashed in any case by writing a good pattern of binary data across the region on your magnetic media (as opposed to simple 0000000000s).

Lastly, the 36 (actually 35) overwrites that you are referring to are recognized as dated and unnecessary today as the technique (known as the Gutmann method) was designed to accommodate the various - and usually unknown to the user - encoding methods used in technologies like RLL and MFM which you're not likely to run into anyhow. Even the US government guidelines state the one overwrite is sufficient to delete data, though for administrative purposes they do not consider this acceptable for "sanitization". The suggested reason for this disparity is that "bad" sectors can be marked bad by the disk hardware and not properly overwritten when the time comes to do the overwrite, therefore leaving the possibility open that visual inspection of the disk will be able to recover these regions.

In the end - writing with a 1010101010101010 or fairly random pattern is enough to erase data to the point that known techniques cannot recover it.

Follow

answered Sep 12, 2008 at 18:26



Bruce the Hoon

1,686 ● 3 ● 19 ● 21



1

I've always wondered why the possibility that the file was previously stored in a different physical location on the disk isn't considered.



For example, if a defrag has just occurred there could easily be a copy of the file that's easily recoverable somewhere else on the disk.



Share Improve this answer

answered Sep 12, 2008 at 22:48

Follow



Sam Hasler

12.6k ● 10 ● 73 ● 106

Biggest reason why this is not considered is because these tools (when properly used) overwrite entire disks. Still, with SSDs this is a real issue as they remap erase blocks without telling the host, for wear leveling. – [MSalters](#) Mar 11, 2009 at 12:14



1

Here's a Gutmann erasing implementation I put together. It uses the cryptographic random number generator to produce a strong block of random data.



```
public static void DeleteGutmann(string
fileName)
{
    var fi = new FileInfo(fileName);
```



```
if (!fi.Exists)
{
    return;
}

const int GutmannPasses = 35;
var gutmanns = new byte[GutmannPasses][];

for (var i = 0; i < gutmanns.Length; i++)
{
    if ((i == 14) || (i == 19) || (i == 25)
    || (i == 26) || (i == 27))
    {
        continue;
    }

    gutmanns[i] = new byte[fi.Length];
}

using (var rnd = new
RNGCryptoServiceProvider())
{
    for (var i = 0L; i < 4; i++)
    {
        rnd.GetBytes(gutmanns[i]);
        rnd.GetBytes(gutmanns[31 + i]);
    }
}

for (var i = 0L; i < fi.Length;)
{
    gutmanns[4][i] = 0x55;
    gutmanns[5][i] = 0xAA;
```

Share Improve this answer

edited Nov 16, 2014 at 23:33

Follow


answered Sep 13, 2008 at 0:15



Jesse C. Slicer

20.1k ● 5 ● 72 ● 89

1 Well, sir, one question here. In above example, will not the OS allocate the free space to the new data you are writing to the file? Hence, it could leave the old / de-allocated clusters open to recovery tools. I may be wrong, but just wanted to be sure about it :-)

– [Nadeem Ullah](#) Jul 31, 2013 at 10:31 

2 No, it shouldn't as I've locked the entire file (and the contents) for writing during the overwrite process. Nothing is freed until the stream is closed and the file is deleted. At that point, the now-free clusters are available, but overwritten with Gutmann garbage.

– [Jesse C. Slicer](#) Jul 31, 2013 at 13:36

One issue with this implementation. If the file size is large, it uses a lot of memory (as it keeps an amount of data in memory, as large as 36 times the size of file). How about changing it not to hold the random bytes in memory, and keep flushing them to file as soon as we generate them?

– [Nadeem Ullah](#) Sep 3, 2013 at 10:19

You're certainly welcome to use whatever variation you'd like :) This is a classic computer science trade-off of speed versus space.

– [Jesse C. Slicer](#) Sep 3, 2013 at 12:18

I changed the code to write 16K garbage to the file at a time. But when a file is large (730M in my case), I get an `OutOfMemoryException` at `s.Write()` line. Not sure why I am getting this even when I am not holding a large number of bytes in memory (only 16K at a time). I did not change anything in your code, except the way I generate the random bytes.

– [Nadeem Ullah](#) Sep 24, 2013 at 11:03



0

There are "disk repair" type applications and services that can still read data off a hard drive even after it's been formatted, so simply overwriting with random 1s and 0s



one time isn't sufficient if you really need to securely erase something.



I would say that for the average user, this is more than sufficient, but if you are in a high-security environment (government, military, etc.) then you need a much higher level of "delete" that can pretty effectively guarantee that no data will be recoverable from the drive.

Share Improve this answer

answered Sep 12, 2008 at 18:15

Follow



[Scott Dorman](#)

42.5k ● 12 ● 81 ● 112

-
- 2 Formatting is different then wiping the first. Formatting just over write the file allocation table (sometimes only the first copy and not redundant copies.) This is why most people (and undelete/unformat) can recover files. – [Matthew Whited](#) Jan 14, 2010 at 16:43
-



0



The United States has requirements put out regarding the erasure of sensitive information (i.e. Top Secret info) is to destroy the drive. Basically the drives were put into a machine with a huge magnet and would also physically destroy the drive for disposal. This is because there is a possibility of reading information on a drive, even being overwritten many times.

Share Improve this answer

answered Sep 12, 2008 at 18:31

Follow



[Astra](#)

11.2k ● 3 ● 39 ● 41



See this: [Guttman's paper](#)

0

Share Improve this answer

answered Sep 13, 2008 at 0:22

Follow



[Will M](#)

2,491 ● 2 ● 18 ● 19



Just invert the bits so that 1's are written to all 0's and 0's are written to all 1's then zero it all out that should get rid of any variable in the magnetic field and only takes 2 passes.

0



Share Improve this answer

answered Oct 4, 2014 at 15:30

Follow



[Jim Mulder](#)

1