computers are bad

https://computer.rip - me@computer.rip - fax: +1 (505) 926-5492

2021-07-26 rip those bits to shreds

Programming note: you may have noticed that computer.rip has been up and down lately. My sincere apologies, one of the downsides of having a neo-luddite aversion to the same cloud services you work with professionally all day is that sometimes your “platform as a physical object” (PaaPO) starts exhibiting hardware problems that are tricky to diagnose, and you are not paid to do this so you are averse to spending a lot of your weekend on it. Some messing around and remote hands tickets later the situation seems to have stabilized, and this irritation has given me the impetus to get started on my plans to move this infrastructure back to Albuquerque.

Let’s talk a bit about something practical. Since my academic background is in computer security, it’s ego-inflating to act like some kind of expert from time to time. Although I have always focused primarily on networking, I also have a strong interest in the security and forensic concerns surrounding file systems and storage devices. Today, we’re going to look at storage devices.

It’s well known among computing professionals that hard disk drives pose a substantial risk of accidental data exposure. A common scenario is that a workstation or laptop is used by a person to process sensitive information and then discarded as surplus. Later, someone buys it at auction, intercepts it at a recycler, or similar and searches the drive for social security numbers. This kind of thing happens surprisingly frequently, perhaps mostly because the risk is *not* actually as common knowledge as you would think. I have a side hustle, hobby, and/or addiction of purchasing and refurbishing IT equipment at auction. I routinely purchase items that turn out to have intact storage, including from government agencies.

So, to give some obvious advice: pay attention to old devices. If your organization does not have a policy around device sanitization, it should. Unfortunately the issue is not always simple, and even organizations which require sanitization of all storage devices routinely screw it up. A prominent example is photocopiers, for years organizations with otherwise good practices were sending photocopiers back to leasing companies or to auction without realizing that most photocopiers these days have nonvolatile storage to which they cache documents. So having a policy isn’t really good enough on its own: you need to back it up with someone doing actual research on the devices in question. I have heard of a situation in which a server was “sanitized” and then surplussed with multiple disk drives intact because the person sanitizing it didn’t realize that the manufacturer had made the eccentric decision to put additional drive bays on the *rear* of the chassis!

But that’s all sort of besides the point. We all agree that storage devices need to be sanitized before they leave your control... but how?

Opinions on data sanitization tend to fall into two camps. Roughly, those are “an overwrite is good enough” and “the only way to be sure is to nuke it from orbit.” Neither of these positions are quite correct, and I will present an unusually academic review here of the current state of storage sanitization, along with my opinionated advice.

The black-marker overwrite

The most obvious way to sanitize a storage device, perhaps after burying it in a hole, is to overwrite the data with something else. It could be ones, it could be zeroes, it could be random data or some kind of systematic pattern. The general concept of overwriting data to destroy it presumably dates back to the genesis of magnetic storage, but for a long time it’s been common knowledge that merely overwriting data is not sufficient to prevent recovery.

A useful early illustration of the topic is Venugopal V. Veeravali’s 1987 master’s thesis, “Detection of Digital Information from Erased Magnetic Disks.” It’s exactly what it says on the tin. The paper is mostly formulae by mass, but the key takeaway is that Veeravali connected a spectrum analyzer to a magnetic read head. They showed that the data from the spectrum analyzer, once subjected to a great deal of math, could be used to reconstruct the original contents of an erased disk to a certain degree of confidence.

This is pretty much exactly the thing everyone was worried about, and various demonstrations of this potential lead to Peter Gutmann’s influential 1996 paper “Secure Deletion of Data from Magnetic and Solid-State Memory.” Gutmann looks at a lot of practical issues in the way storage devices work and, based on consideration of specific patterns that could remain considering different physical arrangements for data storage, proposes the perfect method of data erasure. The Gutmann Method, as it’s sometimes called, is a 35-pass scheme of overwriting with both random data and fixed patterns.

The reason for the large number of passes is partially Just To Be Sure, but the fixed pattern overwrites are targeted at specific forms of encoding. The process is longer than strictly needed just because Gutmann believes that a general approach to the problem requires use of multiple erasure methods, one of which ought to be appropriate for the specific device in question. This is to say that Gutmann never really thought 35 passes were necessary. Rather, to put it pithily, he figured eight random passes would do and then multiplied all the encoding schemes together to get 27 passes that ought to even out the encoding-related patterns on the drives of the time.

Another way to make my point is this: Gutmann’s paper is actually rather specific to the storage technology of the time, and the time was 1996. So there’s no reason to work off of his conclusions today. Fortunately few people do, because a Gutmann wipe takes truly forever.

Another influential “standard” for overwriting for erasure is the “DoD wipe,” which refers to 5220.22-M, also known as the National Industrial Security Program Operating Manual, also known as the NISPOM. I can say with a good degree of confidence that every single person who has ever invoked this standard has misunderstood it. It is not a standard, it is not applicable to you, and since 2006 it no longer makes any mention of a 3-pass wipe.

Practical data remanance

The concept of multi-pass overwrites for data sanitization is largely an obsolete one. This is true for several different reasons. Most prominently, the nature of storage devices has changed appreciably. The physical density of data recording has increased significantly. Drive heads now operate on magnetic coils and track dynamically rather than under absolute positioning (reducing error in tracking). And there are of course today many solid-state drives, which repeatedly overwrite data as a matter of normal operating procedure (but at the same time may leave a great deal of data available).

You don't need to take my word on this! Also in 2006, for example, the NIST issued new recommendations on sanitization stating that a single overwrite was sufficient. This may have been closely related to the 2006 NISPOM change. Gutmann himself published a note in 2011 that he no longer believes his famous method to be relevant and assumes a single overwrite to be sufficient.

Much of the discussion of recovery of overwritten data from magnetic media has long concentrated around various types of magnetic microscopes. Much like your elementary school friend who's uncle works for Nintendo, the matter is frequently discussed but seldom demonstrated. Without wanting to go too deep into review of literature and argumentative blog posts, I think it is a fairly safe assertion that *recovery of data by means of electron microscopy, force microscopy, magnetic probe microscopy, etc is infeasible for any meaningful quantity of data without enormous resources.*

The academic work that *has* demonstrated recovery of once-overwritten data by these techniques has generally consisted of extensive effort to recover a single bit at a low level of confidence. The error rate makes recovery of even a byte impractical. A useful discussion of this is in the ICISS 2008 conference paper "Overwriting Hard Drive Data: The Great Wiping Controversy," amusingly written in part by a man who would go on to claim (almost certainly falsely) to have invented Bitcoin. It's a strange world out there.

As far as summing up the issue, I enjoy the conclusion of a document written by litigation consultant Fred Cohen:

> 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... Indeed, there appears to be nobody in the [forensics and security litigation] 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.

Recovery of overwritten data by microscopy is, in practice, a scary story to tell in the dark.

The takeaway here is that, for practical purposes, a single overwrite of data on a magnetic platter seems to be quite sufficient to prevent recovery.

It’s not all platters

Here’s the problem: in practice, remanance on magnetic media is no longer the thing to worry about.

The obvious reason is the extensive use of SSDs and other forms of flash memory in modern workstations and portable devices. The forensic qualities of SSDs are, to put it briefly, tremendously more complicated and more poorly understood than those of HDDs. To even skim the surface of this topic would require its own post (perhaps it will get it one day), but the important thing to know is that SSDs throw out all of the concerns around HDDs and introduce a whole set of new concerns.

The second reason, though, and perhaps a more pervasive one, is that the forensic properties of the magnetic platters themselves are well understood, but those of the rest of the HDD are not.

The fundamental problem in the case of both HDDs and SSDs is that modern storage devices are increasingly complex and rely on significant onboard software in order to manage the physical storage of data. The behavior of that onboard software is not disclosed by the manufacturer and is not well understood by the forensics community. In short, when you send data to an HDD or SSD, we know that it puts that data somewhere but in most cases we really don’t know *where* it puts it. Even in HDDs there can be significant flash caching involved (especially on “fancier” drives). Extensive internal remapping in both HDDs and SSDs means that not all portions of the drive surface (or flash matrix, etc) are even exposed to the host system. In the case of SSDs, especially, large portions of the storage are not.

So that’s where we end up in the modern world: storage devices have become so complex that the recovery methods of the 1980s no longer apply. By the same token, storage devices have become so complex that we can no longer confidently make any assertions about their actual behavior with regards to erasure or overwriting. A one-pass overwrite is both good enough at the platter level and clearly not good enough at the device level, because caches, remapping, wear leveling, etc all mean that there is no guarantee that a full overwrite actually overwrites anything important.

Recommended sanitization methods

Various authorities for technical security recommendations exist in the US, but the major two are the NIST and the NSA.

NIST 800-88, summarized briefly, recommends that sanitization be performed by degaussing, overwriting, physical destruction of the device, or encryption (we will return to this point later). The NIST organizes these methods into three levels, which are to be selected based on risk analysis, and physical destruction is the recommended method for high risk material or material where no method of reliable overwriting or degaussing is known.

NSA PM 9-12 requires sanitization by degaussing, disintegration, or incineration for “hard drives.” Hard drives, in this context, are limited to devices with no non-volatile solid state memory. For any device with non-volatile solid state memory, disintegration or incineration is required. Disintegration is performed to a 2mm particle size, and incineration at 670 Celsius or better.

Degaussing, in practice, is surprisingly difficult. Effective degaussing of hard drives tends to require disassembly in order to individually degauss the platters, and so is difficult to perform at scale. Further, degaussing methods tend to be pretty sensitive to the exact way the degaussing is performed, making them hard to verify. The issue is big enough that the NSA requires that degaussing be followed by physical destruction of the drive, but to a lower standard than for disintegration (simple crushing is acceptable). For that reason, disintegration and incineration tend to be more common in government contexts.

It’s sort of funny that I tell you all about how multiple overwrite passes are unnecessary but then tell you that accepted standards require that you blend the drive until it resembles a coarse glitter. “Data sanitization is easy,” I say, chucking drives into a specialized machine with a 5-figure price tag.

The core of the issue is that the focus on magnetic remanance is missing the point. While research indicates that magnetic remanance is nowhere near the problem it is widely thought to be, in practice remanance is *not* the way that data is sneaking out. The problem is not the physics of the platters, it’s the complexity of the devices and the lack of reliable host access to the entire storage capacity.

ATA secure erase and self-encryption and who knows what else

The ATA command set, or rather some version of it, provides a low-level secure erase command that, in theory, causes the drive’s own firmware to initiate an overwrite of the entire storage surface. This is far preferable to overwriting from the host system, because the drive firmware is aware of the actual physical storage topology and can overwrite parts of the storage that are not normally accessible to the host.

The problem is that drive manufacturers have been found to implement ATA secure erase sloppily, or not at all. There is basically no external means of auditing that a secure erase was performed effectively. For that simple reason, ATA secure erase should not be relied upon.

Another approach is the self-encrypting drive or SED, which transparently encrypts data as it is written. These devices are convenient since simply commanding the drive to throw away the key is sufficient. SED features tend to be better implemented than ATA secure erase because of the fact that they are only implemented at all on high-end drives that are priced for the extra feature. That said, the external auditing problem still very much exists.

Another option is to encrypt at the host level, and then throw away the key at the host level. This is basically the same as the SED method but since the encryption is performed externally to the drive, the whole thing can be audited externally for assurance. In all reality this is a fine approach to data sanitization and should be implemented whenever possible. If you have ever been on the fence about whether or not to encrypt storage, consider this: if you are effective about encrypting your storage, you won’t need to sanitize it later! The mere absence of the key is effective sanitization, as recognized by the NIST.

The problem is that disk encryption features in real devices are inconsistent. Drive encryption may not be available at all, or it may only be partial. This makes encryption difficult to rely on in most practical scenarios.

The bottom line

When you dispose of old electronics, you should perform due diligence to identify all non-volatile storage devices. These storage devices should be physically destroyed prior to disposal.

DIY methods like drilling through platters and hitting things with hammers are not ideal, but should be perfectly sufficient for real scenarios. Recovering data from partially damaged hard drives and SSDs is possible but not easy, and the number of facilities that perform that type of recovery is small. There are lots of ways to achieve this type of significant damage, from low-cost hand-cranked crushing devices to the New Mexican tradition of taking things out to the desert and shooting at them. Await my academic work on the merits of FMJ vs hollow-point for data sanitization. My assumption is that FMJ will be more effective due to penetration in multi-platter drives, but I might be overestimating the hardness of the media, or underestimating the number of rounds I will feel like putting into it.

Ideally, storage devices should be disintegrated, shredded, or incinerated. Unless you are looking forward to making a large quantity of thermite, these methods are difficult without expensive specialized equipment. However, there are plenty of vendors that offer certified storage destruction as a service. Ask your local shredding truck company about their rates for storage devices.

Most conveniently, do what I do: chuck all your old storage devices in a drawer, tell yourself you’ll use them for something later, and forget about them. We’ll call it long-term retention, or the geologic repository junk drawer.