Data Deletion
Challenges and Risks of Recovery

By:
Dr. Bruce Nikkel
Professor of Digital Forensics
Bern University of Applied Sciences

November 15, 2019

Introduction

This article discusses the challenges of deleting digital data. The word "delete" originates from the Latin *deletus* which means to destroy. The motivation to destroy data today is typically for security and privacy reasons, or to fulfill legal and regulatory obligations. It is helpful to understand what data deletion really means in a modern technological context. In many situations today, deleted data is not actually destroyed, but only removed from the computer user's visibility.

Early computer systems stored data on paper punch-cards and paper tape, and company archives were based on printed paper documents. Destroying data was a physical act involving shredding or burning. But even with physical destruction, advanced forensic techniques could still recover some data fragments under certain conditions. With the introduction of re-usable storage media, physical destruction was still a viable option, but over-writing old data provided a new method of destruction. This allowed data to be destroyed without losing the original investment purchase of the storage technology. The Enter computer museum in Solothurn (http://enter.ch) has an excellent historical collection of storage technologies (both analog and digital) and is worth visiting to see the evolution of computer storage.

Digital storage today typically uses one of three media technologies: magnetic media (disks and tapes), optical media (CDs, DVDs, Blu-ray), and solid state (SSDs, USB sticks). Magnetic and optical technologies are slowly becoming obsolete as solid state media becomes cheaper with much higher storage capacity. Cloud storage could also be mentioned here, but cloud providers use the same storage media and just make it available over a network.

The rest of this article will discuss what deletion really means with the different storage technologies. In many cases digital forensics tools and methods can be used to recover data from storage technologies after a user deletes data. In enterprise and cloud environments deleting data is more complicated because

backup, replication, and synchronization technologies are used to distribute data across multiple locations for resilience.

Magnetic Media

Recovering deleted data from magnetic hard disks is sometimes referred to as the golden age of digital forensics. Magnetic hard disks store data across physical platters in small data units called sectors. Operating systems like Windows, MacOS, or Linux use file systems to associate data on the disk with files visible to the user. When a user deletes a file, the OS simply removes the association, but the data still resides on the disk. Digital forensic tools are widely available to recover data deleted on magnetic disks. Secure delete programs are available to ensure files are really destroyed by writing random data or zeros to the entire file before instructing the OS to delete it. To securely erase an entire disk involves writing random data or zeros to every sector on the entire disk. On larger disks this process can many hours or even days.

Tapes are typically used for backups and archiving. Here files are stored sequentially on a magnetic strip of tape. Tapes don't have file systems like disks, and files can't be individually deleted in the same way. The entire tape, from beginning to end, must be overwritten with random data or zeros to delete data.

A further technique for magnetic data recovery is magnetic force microscopy (MFM). This advanced method involves special microscopes which detect magnetic strength on magnetic media and, in theory, can be used to deduce what data that was written to the disk in the past. The idea is similar to closely examining a piece of paper where pencil text was written and then erased. MFM recovery techniques require highly specialized equipment and considerable time and effort. Because of this (largely impractical) recovery method, secure wiping tools write random data multiple times to reduce the possibility of recovery. Many researchers question the practicality and feasibility of MFM recovery on modern hard disks[1][2].

Secure disk wiping software can destroy data without destroying the physical media (DBAN for example, http://dban.org) and is probably enough to satisfy most risk requirements. However, securely wiping magnetic disks is time consuming and requires knowledge of wiping tools. Organizations may find the cost of physical destruction is cheaper than securely wiping disks with software, especially with older disks that will not be reused. There are two common methods to physically destroy magnetic media. A method called degaussing exposes the entire disk or tape to an intense magnetic field, destroying the magnetic field containing digital data bits. Physical destruction is also achieved by shredding, burning, or melting the storage media.

Optical Media

Files on optical media (CD, DVD, Blu-ray) are written sequentially similar to tapes, but can be read randomly like hard disks. Digital bits are represented with "pits" and "lands" which absorb or reflect light from a laser.

Some optical media is only writable once and cannot be erased, other optical media can be erased and written over again. Some experimental research with optical microscopy has been done to attempt recovery of overwritten optical data but this has not emerged as a practical digital forensic data recovery method[3].

Optical media shredders are cheap and widely available. These mechanical devices punch a grid of small holes in the optical foil and also warp the disc. This prevents data recovery from common digital forensic methods. Shredding optical discs is the most cost effective method of data destruction.

Solid State Media

One of the characteristics of solid state media is that the memory cells can wear out over time. To address this problem, solid state drives are manufactured with a reserved portion of the drive (called over-provisioning) to replace data cells as they wear out. Over time more and more cells are replaced and a significant amount of data can be found in the over-provisioned area of the drive. This area is not accessible using software over the standard drive interface, and advanced hardware methods are required to recover data. A common forensic method called "chip-off" involves the removal (de-soldering) of the individual chips in the drive where the old data can be read from the replaced cells.

Because of the way SSD's work, cells need to be erased before they can be rewritten. Modern SSDs and operating systems will trigger this low level erase process when a file is deleted by a user (by sending a "TRIM" command to SSD firmware). This prevents deleted data from from being recoverable using traditional digital forenisic software. Older SSDs and USB sticks might not support the TRIM command, in which case recovery is possible (similar to magnetic disks).

Newer SSD drives have built-in encryption called OPAL, and securely wiping the disk simply involves resetting the encryption key[4]. In most cases this is a very effective method of data destruction. However, recent research shows some SSD manufacturers do not implement OPAL technology in a secure way and could allow encrypted data to be accessed using advanced techniques[5].

Here again, physical destruction will guarantee that data is destroyed and unrecoverable. However, the use of secure wiping software may be sufficient depending on the sensitivity of the data and the organization's risk appetite.

Cloud Storage

It is often said: "There is no cloud, its just other people's servers". This often forgotten truth is a challenge for data deletion. When data is stored with a cloud provider, all the associated data management responsibility is delegated to that provider, including deletion. But are cloud providers deleting data? Are they even able to confidently delete data? Cloud technologies typically involve distributing data across many servers in many countries. Data is replicated in multiple places and even dynamically moved around within the cloud. It is difficult for a cloud provider to even know exactly where data is stored or how many copies exist at any given time.

Most cloud providers focus on performance, reliability and availability of data. This means extensive replication and backups. When cloud providers talk about data security, it usually refers to authentication and encrypted access to the cloud over a network. But exactly how security of data storage is managed within the cloud is usually not revealed. If you delete data in a 3rd-party cloud, you have no guarantee that it is actually deleted. Data may appear deleted (no longer visible) to a cloud user, but copies are still likely to exist within a complex cloud.

Where else is your data?

One of the biggest challenges to deleting data is finding all the copies and making sure everything is securely destroyed. Data to be deleted may reside on backups, or replicated across resilient infrastructure. Data may reside on old servers or PCs that have been decommissioned in the past. Staff may have made copies of data on external media like USB sticks or external hard drives. Data may exist as file attachments in emails. Data may have been temporarily copied to external drives by contractors or out-sourcing partners during upgrades or data migration projects.

Traces of data may also be stored on mobile devices, or synchronized with other applications or devices. Intelligent printers and scanners with built-in storage may contain traces of documents from previous prints or scans. IoT devices used to help manage our digital lives also have increasing amounts of data storage. It is becoming difficult to know where data is located and in many cases difficult or even impossible to delete at all.

Encrypted Storage

With all the forensic recovery methods and cloud uncertainty what is a safe method of deleting data and ensuring it is really destroyed? Encryption can solve many problems with both data protection and data destruction. Destroying encrypted data can be achieved by destroying the decryption key (and all copies of the key). Even if copies of the encrypted data are scattered around the world in a cloud, destroying the keys will destroy access to that data. You

don't need to know exactly where the encrypted copies are, and you don't need to overwrite them with random data or zeros to destroy it.

An interesting practical example to illustrate the effectiveness of data destruction with encryption is the current trend of ransomware cyber attacks. Criminals encrypt the victim's data and hold the decryption key for ransom. This has been a very effective attack with many victims paying the ransom to avoid loss of data.

There is a new theoretical risk around destroying keys to delete encrypted data. In the distant future it is predicted that quantum computers will be able to decrypt traditional encryption algorithms with relative ease[6]. Encrypted data using traditional encryption today (i.e. not using quantum safe or quantum resistant cryptography) may be at risk of being recovered in the future[7].

Conclusion

It is interesting to think about data destruction. Properly destroying data is a challenge in our modern age of cloud computing and will change further as technology (including quantum computing) advances. Data deletion has similar goals as data protection and security, but completely opposite goals of data resilience and availability. Security is all about making risk decisions. What is the risk of data not being properly deleted? Has deletion activity satisfied due diligence expectations?

To make these risk decisions, it helps to understand who is able to recover or reconstruct an organization's deleted data:

- Staff who are keeping their own copies of data
- System administrators authorized to access backup or replicated copies
- Contractors and partner companies who made copies of data
- Cloud providers who won't or can't securely delete data
- Data recovery and digital forensic companies
- Academic research labs performing proof of concept data recovery under experimental conditions
- Government analysis labs using well funded equipment not available to the public

What is the motivation or likelihood that these parties will attempt to recover your deleted data?

There are many other questions to ask regarding the topic of data deletion and which methods to use. What is the data you are trying to destroy and why are you trying to destroy it? How many copies are there in existence? Where are the copies located? Can you delete them securely if needed? Clearly the topic of data deletion is complex and not as straightforward at is appears. Deleted data may no longer be visible to the user, but this doesn't mean data has been destroyed from existence. There is no way to prove the non-existence of data,

you can only prove the existence and effectiveness of methods and processes used to perform data destruction.

References

[1] J. Reardon, S. Capkun, and D. Basin, SoK: Secure data deletion, Proceedings of the 34th IEEE Symposium on Security & Privacy, 2013

[2] Nikolai Joukov, Harry Papaxenopoulos, and Erez Zadok, "Secure deletion myths, issues, and solutions", Proceedings of the 3rd International IEEE Security in Storage Workshop (SISW), 2005

[3] Greg Gogolin, James Jones, Derek Brower, "Maximizing Data Recovery Quality", Troy Vol. 53, Iss. 10, (Oct 2014): 46,48-49

[4] Trusted Computing Group, Opal Specification, https://trustedcomputinggroup.org/wp-content/uploads/TCG_Storage-Opal_SSC_v2.01_rev1.00.pdf

[5] Carlo Meijer, Bernard Van Gastel, "Self-Encrypting Deception, Weaknessess in the Encryption of of Solid State Drives", https://www.ieee-security.org/TC/SP2019/papers/310.pdf

[6] Bruce Schneier, "On Quantum Computing and Cryptography", https://www.schneier.com/blog/archives/2018/09/quantum_computi_2.html

[7] Daniel Bernstein, "Introduction to post-quantum cryptography", www.springer.com/cda/content/document/cda_downloaddocument/9783540887010-c1.pdf