DATA RECOVERY

Daniel Boulos & Nicholas Zemljic

WHAT IS DATA RECOVERY?

Data recovery is the process of retrieving data from a storage medium that, for some reason, cannot be accessed normally. This process may be used to recover data from a variety of **storage media**, such as: hard disk drives, solid-state drives, other flash storage (such as USB drives, SD cards), or other disk storage (such as CDs, DVDs). The damage that causes data to be lost typically falls into one of two categories: **physical damage** (where the hardware is damaged or is malfunctioning), or **logical damage** (where part of the software and/or file system prevents the data from being accessed by the host operating system.) We'll discuss these different types of storage damage in greater depth a bit later.

The term "data recovery" can also be organized into two different contexts: **personal data recovery**, and **forensic data recovery**. Personal data recovery is what we normally associate with this topic. It simply refers to the retrieval of data that has been involuntarily lost or made inaccessible due to, for example, damaged storage media. By contrast, forensic data recovery often deals with retrieving data that has been purposely encrypted or hidden to prevent others (such as forensic investigators) from accessing the data.

While some of the content this presentation will apply to both contexts of the term, will focus primarily on the more common (and applicable) topic of personal data recovery.

COMMON SCENARIOS

Here, we've included some common scenarios where data recovery procedures would be necessary:

- There has been an operating system failure or some critical operating system files have been damaged, causing the device to not be able to boot up properly. In this case, a simple solution would be to use a **Live USB** to boot up from another operating system so that you can access the data from the storage medium.
- There has been a hard disk failure and there is physical damage to the storage medium. In this case, you may be able to repair the hardware, but the storage medium is often beyond repair and the focus is more on a one-time recovery in an attempt to salvage any data you can. This will often require the services of a **specialized data recovery company**.
- Files have been **deleted** from a storage medium. As we will discuss later in the presentation, when an operating system "deletes" files, often times the data is not immediately removed from the drive. This allows tools such as **file carvers** to recover this data.

TYPES OF STORAGE DAMAGE

Physical Damage

This type of storage damage occurs when the **physical hardware** of the storage medium is damaged. An example of this is when a hard disk drive suffers from a mechanical failure such as a **head crash**.

A **head crash** is a hard drive disk failure that occurs when a read-write head comes in contact with its rotating platter. This results in permanent, and usually irreparable, damage to the hard drive. It's typically caused by a sudden jolt of motion, for example: dropping a laptop while it's operating. To defend against this, modern laptops have what's known as **"active hard-drive protection"**, which usually consists of accelerometers that alert the system when excess acceleration is detected. The software will tell the hard drive to unload/park its heads to prevent them from coming in contact with the platter.

Physical damage will almost always cause **data loss**, and in many cases leads to damage of the supporting operating system and/or file system.

It is usually recommended that physical damage should not be repaired by end users, as these types of recovery situations may require the use of specific hardware or specialized technical expertise. It's also suggested that any physical repairing should be done in a dust-free and static-free environment to ensure that further damage is not done during the recovery process.

Data recovery companies may use various techniques to attempt to repair storage media, for example they may replace damaged or malfunctioning parts in the hard disk. However, even if this repairs the disk and makes the storage device usable, there may still be damage to the file system or operating system that prevents it from functioning properly. At this point, they may use a **disk-imaging** procedure to recover surviving data and transfer it to a reliable medium, so that the image can be analyzed and possibly reconstructed to a working state.

Logical Damage

This type damage refers to damage that has occurred at the **software-level**, causing errors or complete loss of access to data.

Damaged Filesystems

In some cases, data on a storage medium may be unreadable due to damage to the filesystem. In this case, using a data recovery tool (such as **Testdisk**) can repair the damaged filesystem to restore access to the data. If the filesystem can't be repaired, other tools (such as **dd rescue**) can be used to image storage media despite having filesystem damage. This type of data recovery can be performed by anyone, as it doesn't require any specialized physical equipment or access to inner hardware components.

If data has been damaged in such a way that it cannot be repaired, a procedure known as file carving can be used to recover parts of damaged files. **File carving** is the process of reassembling computer files from fragments in the absence of filesystem metadata.

Deleted Data

HDD:

For most operating systems, when a file is deleted on a Hard Disk Drive, the contents of the file are not immediately removed from the drive. Instead, to improve performance, they simply **remove references** to the file in the directory structure and mark the space that they occupy as available so that other data can be written to it later. This means that the original data remains on the disk and may be recoverable using specialized tools such as a **file carver**.

SSD:

Data recovery on an SSD is much different than that of a HDD due to a command and hardware supported process called **TRIM**.

The **TRIM** command is designed to enable the operating system to notify the SSD which pages no longer contain valid data due to erases either by the user or operating system itself. During a delete operation, the OS will mark the sectors as **free** for new data and send a TRIM command to the SSD to mark them as not containing valid data. After that the SSD knows not to preserve the contents of the block when writing a page, resulting in fewer writes to the flash, higher write speed, and increased drive life.

Similarly to both storage mediums when an item is deleted simply the meta-data is removed while the bulk of data is left untouched, however the space that data occupied is marked as free-space. On a traditional HDD this is a suitable method since when new data is sent to be written on the drive it can simply overwrite the sectors whether or not they are occupied.

SSD's use **NAND flash memory** to store and transfer information; free space is made up of blocks, within the blocks there are pages. Data is written to pages but only blocks may be erased. When a write request is sent to the SSD all the free pages are grouped together and cleared, which impacts the efficiency of the drive since the erasing is done at that time. To alleviate the slowed performance **TRIM** support was introduced. Pages that are marked for deletion are now handled before you need them. The SSD and OS are in communication about which pages have been cleared and therefore can be reused by the OS efficiently.

This now means that modern OS' and SSD's will be pre-emptively clearing the data on the drive upon deletion making it much harder if not impossible to recover. On the one hand it is harder for the attacker to break confidentiality; on the other it is harder as the defender to restore availability on accidentally deleted or lost data.

Garbage Collection using TRIM

Non-Queue method:
After each filesystem delete command the TRIM command is sent. This has a massive execution penalty.

Queued or Batch method:
To minimize any command penalty a batched trim occurs, rather than trimming upon every file deletion, by scheduling such batch jobs for when system utilization is minimal.

The shortcoming has been overcome in the Serial ATA revision 3.1.

DATA RECOVERY PROCESS

This outlines the general process for recovering data. Some steps, in particular Step 1, may not be necessary depending on the situation.

Step 1: *Repair any damaged hardware of storage device*
Repair the hard drive so that it is running in some form. This usually involves replacing malfunctioning or damaged parts.

Step 2*: Image the drive to a new drive or a disk image file*
It is important that no other data is written to the damaged drive before it can be imaged, to avoid overwriting the data you are trying to salvage.

Step 3: *Logical recovery of data*
After the drive has been imaged, the original data can be retrieved using software such as a file carver. You can also attempt to repair the file system using certain tools.

Step 4: *Repair the recovered files*
In the event that some files are damaged, you may need to use some software to try and reconstruct the data using, for example, a hex editor.

SOFTWARE DATA RECOVERY TOOLS AND TECHNIQUES

A simple technique to regain access to data that is on a storage media that won't boot (due to, for example, logical damage to the operating system) is to **mount** it to another computer, or use a **Live USB/CD** to boot another operating system on the machine.

Another solution is to use a tool such as **Testdisk** to try and recover lost partitions and/or make non-booting disks bootable again.

There are also a number of closed-source and/or commercial software solutions for data recovery, as well as some non-commercial and/or **open-source tools**.

A popular tool we're going to look at now is called **PhotoRec**.

PHOTOREC

PhotoRec is a free and open-source **file carving tool** designed to recover lost files.

As mentioned earlier, file carving is the process of reassembling computer files from fragments in the absence of filesystem metadata.

It can recover data from various **storage media**, including: SD cards, USB flash drives, hard drive disks, CDs, and DD disk image files.

It recovers most common photo formats (like JPEG images), audio files (such as MP3), videos (for example, MP4 files), document formats (including OpenDocument, Microsoft Office, PDF, and text files), and archive formats (such as ZIP). In all, it recognizes **over 440 file extensions**.

It's a **multi-platform** tool, so it's compatible with most Linux distributions and most versions of Windows and OS X.

PhotoRec ignores the file system, so it works even if the file system is severely damaged. It's able to recover data from several common filesystem formats including FAT, NTFS, and ext2/ext3/ext4.

How PhotoRec Works:

FAT, NTFS, ext2/ext3/ext4 filesystems store files in **data blocks** (called clusters in Windows), the size of which are constant. In general, most operating systems try to store data in a contiguous way so as to minimize **fragmentation**. When a file is deleted, the **meta-data** about this file is lost; *for example, in an ext3/ext4 filesystem, the names of deleted files are still present, but the location of the first data block is removed.*

To recover lost files, PhotoRec first tries to find the **data block size**. If the filesystem is not corrupted, this value can read from the **superblock** (for ext2/ext3/ext4) or **volume boot record** (for FAT, NTFS). Otherwise, the program tries to calculate the block size by reading the media and searching for files to compare with. Once the block size is known, the tool reads the media block by block, comparing each block against a **signature database**.

For example, PhotoRec identifies a JPEG file when a block begins with:
0xff, 0xd8, 0xff, 0xe0
or

0xff, 0xd8, 0xff, 0xe1
or
0xff, 0xd8, 0xff, 0xfe

Finally, once a file is recovered successfully, the program checks any previous data blocks to see if some fragmented files can also be recovered.

PHOTOREC DEMO

PhotoRec can be downloaded from their website (www.cgsecurity.org/wiki/PhotoRec) or through your Linux package manager.

Some of these commands will require root privileges (using *sudo*).

To run PhotoRec:
>> *photorec*

To check special device file of the SD card:
>> *fdisk -l*

To securely delete the contents of the SD card:
>> *shred -vzn 0 /dev/…*

To format the SD card back to FAT 32:
>> *mkfs.msdos -F 32 /dev/…*

GUTMANN METHOD

When deleting data from a hard disk (**HDD**), there are two common methods:

- **Fast method:** Meta data is deleted and the space is marked as free space. This is the most common way for operating systems to delete data, as it is much faster.
- **Gutmann method:** Commonly referred to as a [1-35] pass erase. It involves erasing a drive using a special algorithm to increase the difficulty of retrieving the original content on the drive. At this point only special extraction tools may have any chance of retrieving the data.

In our demonstration, we also showed how the **shred** command was used to securely delete (overwrite with 0s) data from the SD Card after it was “erased” by the digital camera.

RESTORING C.I.A. IN THE EVENT OF DATA LOSS

Here are some guidelines to follow when attempting to *mitigate the impact of data loss and/or protect against data loss*. The symbol next to each recommendation outlines which aspect of **C.I.A.** (**Confidentiality**, **Integrity**, and **Availability**) is either maintained or restored.

(**A**) *Use data recovery tools and techniques to retrieve lost data*

- In the case of an emergency, where data has already been lost, this allows the user to restore Availability since you regain access to lost data.

(**C**) *Encrypt sensitive data to maintain confidentiality*

- Always encrypting sensitive/confidential data beforehand ensures that Confidentiality is maintained in the event of any data loss, since an attacker cannot access your real data without decrypting it.

(**I**) *Use a RAID configuration for data redundancy*

- Using a RAID configuration with your system allows you to maintain Integrity by having multiple copies of data across multiple drives. This provides a way to check for data consistency and ensure that Integrity remains intact (or recover from any damage that affects Integrity).

(**I**) *Keep computer security in mind (and/or use antivirus software)*

- Maintain Integrity by avoiding suspicious links and programs that may be malicious. Such malicious software could cause logical damage to the storage medium, putting Integrity at risk.

(**A**) *Regularly backup your data to external storage media*

- Having recent offline/external backups of your data will allow you easily restore your data in the event of physical or logical damage, restoring Availability.

(**I**)(**A**) *Protect against physical damage*

- Many instances of physical damage can be prevented by using the correct hardware and/or software. For instance, using a surge protector can protect against power surges, and using a laptop that employs “active hard-drive protection” can prevent head crashes. Protecting against physical damage therefore maintains Availability, since you maintain access to the files, as well as Integrity, since you know they have not been damaged or modified.