File Structures and Forensics

Cybercrime Lab

U.S. Department of Justice

Computer Crime and intellectual Property Section

A Day in the Life of a Hard Drive

- Writing files
- Deleting files
- Overwriting files

- Recovering
- Recovering
- Recovering

- We’ll consider these different scenarios using a series of hypothetical events on a hard drive.

Definitions

- File System = table of contents for a hard drive
- Cluster = a unit of storage space
- Directory = maps each file name to its starting cluster, size, and other attributes
- File Allocation Table (FAT) = maps each cluster to its status

Directory

file name	start	size

FAT

cluster	value
1	0
2	0
3	0
4	0
5	0

Disk

cluster	data
1	
2	
3	
4	
5	

Writing a File

- The Directory entry is created pointing to the first available cluster.
- The relevant FAT entry is changed.
- “EOF” indicates the cluster is in use and is the last cluster of the file (“End Of File”).
- The file is written.

Directory

file name	start	size
accomplices.txt	1	1

FAT

cluster	value
1	EOF
2	0
3	0
4	0
5	0

Disk

cluster	data
1	Charles Manson, Leopold, and Loeb.
2	
3	
4	
5	

Writing a Bigger File

- A file takes up as many clusters as it needs.
- If a file continues past a cluster, the FAT entry for that cluster shows the next cluster used.
- The file is written.

Directory

file name	start	size
accomplices.txt	1	1
letter.doc	2	3

FAT

cluster	value
1	EOF
2	3
3	4
4	EOF
5	0

Disk

cluster	data
1	Charles Manson, Leopold, and Loeb.
2	Dear Mom, I know you don't approve of my new
3	friends, but we're having a lot of fun. Especially
4	plotting evil deeds. Love you.
5	

Reading a File

- Say the user wants to open letter.doc.
- The computer looks at the directory to see it starts in cluster 2.
- The computer follows the FAT entries to see which clusters hold the file's data (2 to 4).

Directory

file name	start	size
accomplices.txt	1	1
letter.doc	2	3

FAT

cluster	value
1	EOF
2	3
3	4
4	EOF
5	0

Disk

cluster	data
1	Charles Manson, Leopold, and Loeb.
2	Dear Mom, I know you don't approve of my new
3	friends, but we're having a lot of fun. Especially
4	plotting evil deeds. Love you.
5	

Deleting a File

- In the Directory, the first character is replaced with a special character indicating the file has been deleted.
- The file's FAT entries are set to indicate its clusters are available.
- Nothing happens to the data on the disk.

Directory

file name	start	size
?ccomplices.txt	1	1
letter.doc	2	3

FAT

cluster	value
1	EOF
2	3
3	4
4	EOF
5	0

Disk

cluster	data
1	Charles Manson, Leopold, and Loeb.
2	Dear Mom, I know you don't approve of my new
3	friends, but we're having a lot of fun. Especially
4	plotting evil deeds. Love you.
5	

Recovering a Deleted File

- Investigator can see most of the name from the Directory.
- FAT no longer shows chain of clusters.
- But an investigator can see the starting cluster and size in the Directory, and then examine the relevant span of clusters on the disk.

Directory

file name	start	size
?ccomplices.txt	1	1
letter.doc	2	3

FAT

cluster	value
1	0
2	3
3	4
4	EOF
5	0

Disk

cluster	data
1	Charles Manson, Leopold, and Loeb.
2	Dear Mom, I know you don't approve of my new
3	friends, but we're having a lot of fun. Especially
4	plotting evil deeds. Love you.
5	

Writing a File - Divided

- Say the user saves a file that takes up two clusters.
- The Directory and FAT entries are made starting at the first available cluster.
- Data is written to the disk.

Directory

file name	start	size
badguys.txt	1	2
letter.doc	2	3

FAT

cluster	value
1	5
2	3
3	4
4	EOF
5	EOF

Disk

cluster	data
1	Billy the Kid, Bonnie & Clyde, Al Capone, John
2	Dear Mom, I know you don't approve of my new
3	friends, but we're having a lot of fun. Especially
4	plotting evil deeds. Love you.
5	Dillinger, Jesse James, Frank James.

Reading a File - Divided

- Say the user wants to open badguys.txt.
- The computer looks at the directory to see it starts at cluster 1.
- The computer follows the FAT entries to see which clusters hold the data (1, 5).
- The computer retrieves the data.

Directory

file name	start	size
badguys.txt	1	2
letter.doc	2	3

FAT

cluster	value
1	5
2	3
3	4
4	EOF
5	EOF

Disk

cluster	data
1	Billy the Kid, Bonnie & Clyde, Al Capone, John
2	Dear Mom, I know you don't approve of my new
3	friends, but we're having a lot of fun. Especially
4	plotting evil deeds. Love you.
5	Dillinger, Jesse James, Frank James.

Deleting a File - Divided

- Same as when the file is contiguous.
- In the Directory, the first character is replaced with a special character indicating it's been deleted.
- The file's FAT entries are set to indicate its clusters are available.

Directory

file name	start	size
?adguys.txt	1	2
letter.doc	2	3

FAT

cluster	value
1	0
2	3
3	4
4	EOF
5	0

Disk

cluster	data
1	Billy the Kid, Bonnie & Clyde, Al Capone, John
2	Dear Mom, I know you don't approve of my new
3	friends, but we're having a lot of fun. Especially
4	plotting evil deeds. Love you.
5	Dillinger, Jesse James, Frank James.

Recovering a Deleted File - Divided

- Investigator can see most of the name and the starting cluster from the Directory.
- FAT no longer shows chain of clusters.
- But an investigator can see the starting cluster and size in the Directory, then hypothesize the first two available clusters constituted the file.

Directory

file name	start	size
?adguys.txt	1	2
letter.doc	2	3

FAT

cluster	value
1	0
2	3
3	4
4	EOF
5	0

Disk

cluster	data
1	Billy the Kid, Bonnie & Clyde, Al Capone, John
2	Dear Mom, I know you don't approve of my new
3	friends, but we're having a lot of fun. Especially
4	plotting evil deeds. Love you.
5	Dillinger, Jesse James, Frank James.

Problem 1: Intervening File Deleted

- The hypothesis that the first two available clusters belonged to the file is wrong.
- Results:
  - Recovery is partial
  - It's obvious that recovery is partial
  - Additional data may remain on the disk

Directory

file name	start	size
?adguys.txt	1	2
?etter.doc	2	3

FAT

cluster	value
1	0
2	0
3	0
4	0
5	0

Disk

cluster	data
1	Billy the Kid, Bonnie & Clyde, Al Capone, John
2	Dear Mom, I know you don't approve of my new
3	friends, but we're having a lot of fun. Especially
4	plotting evil deeds. Love you.
5	Dillinger, Jesse James, Frank James.

Problem 2: Partially Overwritten File

- Any part (or parts) of a deleted file may be overwritten by other files.
- Results:
  - Only partial recovery is possible
  - It's obvious that recovery is partial
  - Additional data may remain on the disk

Directory

file name	start	size
?adguys.txt	1	2
letter.doc	2	3
pw.txt	5	1

FAT

cluster	value
1	0
2	3
3	4
4	EOF
5	EOF

Disk

cluster	data
1	Billy the Kid, Bonnie & Clyde, Al Capone, John
2	Dear Mom, I know you don't approve of my new
3	friends, but we're having a lot of fun. Especially
4	plotting evil deeds. Love you.
5	My password is pw4321.

FAT/NTFS Comparison

	FAT (File Allocation Table)	**NTFS (New Technology File System)**
Use	Older computers, removable devices	Newer computers
Method of Organization	FAT	MFT (Master File Table), a database with more details
Benefits	Wider compatibility	More flexible, More reliable

Putting it in Perspective

- Our example had 5 clusters.
- Each cluster stored a small amount of text.
- A typical consumer-quality hard drive has about 61,000 clusters.
- Each cluster can contain about 1,000 printed pages of data.

Contact

Cybercrime Lab
Computer Crime and
Intellectual Property Section
United States Department of Justice

- Phone: 202-514-1026
- Web: www.cybercrime.gov