Reading Discussion

Blown to Bits

Chapter 3

Ghosts in the Machine

Secrets and Surprises of Electronic Documents

Notes for CSC 100 - The Beauty and Joy of Computing
The University of North Carolina at Greensboro

Question 1....

Describe the heart of the chapter in a few words

Seeing is Believing?

Editors like MS Word are called WYSIWYG:
"What You See Is What You Get"

With file formats, it's often WYGIMTWYS:
"What You Get Is More Than What You See"

Incorrect Redaction

Covering with a higher layer of black pixels does not remove information from the document!

Examples:

- News stories providing source materials
  (*NY Times* and *Washington Post*)
- Document from military on Italian journalist shooting
- President Obama's Tax Returns

How hard is this to uncover? Trivial - *let's see an example*!

Other Information in Documents

Metadata:

- Files often have information about who created it, when it was created, when it was last edited, etc.
  - See an example!
  - Often more metadata available than what is obvious!

Revision history/tracking :

- "Track Changes" in MS Word
  - Very useful for project management (remember "Versioning"?)
  - Embarrassing (or worse) if made public
- Revisions kept by Google docs, Dropbox, Wikipedia, ...
  - Even if they didn't show you revisions, they probably make backups!
  - Even "versions" of the full web! (http://www.archive.org/)

A Tie-In to Algorithms Discussion

From Blown to Bits, page 90:

> But more than electrical engineering is involved. At more than a megabyte per image, digital cameras and HD televisions would still be exotic rarities. A megabyte is about a million bytes, and that is just too much data per image. **The revolution also required better algorithms** — better computational methods, not just better hardware — and fast, cheap processing chips to carry out those algorithms.

Steganography

Idea: Hiding not just content of message, but the fact that there even is a secret message.

One way: Least significant bits of pixels or music samples look fairly random, and so can embed random-looking data.

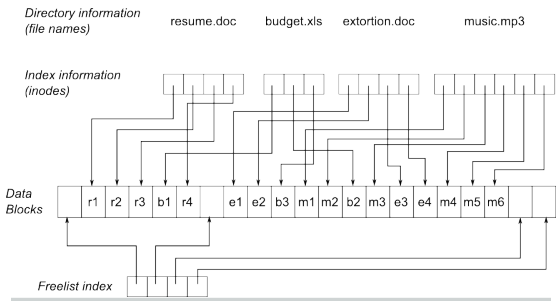
- And encrypted data is random-looking!

A good overview: http://www.garykessler.net/library/steganography.html

What happens when you delete a file?

(and really delete - not just move to trash can!)

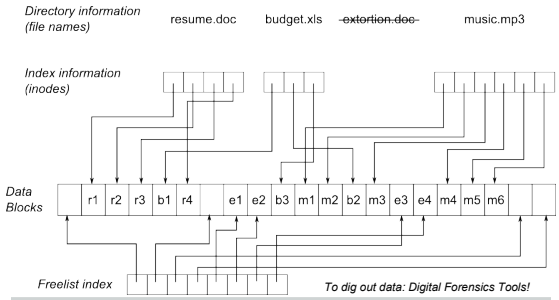
Typical filesystem - Uh Oh! Better delete extortion.doc!



What happens when you delete a file?

(and really delete - not just move to trash can!)

What really happened? Removed name and shifted data blocks to the free list.



To dig out data: Digital Forensics Tools!

Some points we'll return to later

Data representation for media

- Audio files
- Pictures
- Video

Compression: Lossless vs. Lossy

File formats: Standards, proprietary, etc.