

# Digital File Fingerprinting for Computer Forensics Analysis

Dr John Haggerty



## Talk outline

- ? What is Computer Forensics?
- ? Computer Forensics vs. Computer Security
- ? Current practice
- ? Related literature
- ? Forsigs case study
  - Digital picture files
  - File fingerprint selection
  - Fingerprint matching
  - Results
  - Demo

## Some concerns

- ? Pervasiveness of computing devices had led to focus on computer forensics
- ? Computers are target of a crime, repository of information or tool for committing the crime
- ? Some concerns in the electronic age
  - paedophile activity, eg. spread of malicious images, 'grooming', planning to attack children, etc.
  - "phishing"
  - use of digital equipment for criminal activities, eg. encrypted e-mail, digital shredders, networks, etc.
  - monitoring networks to steal data, eg. PIN numbers, credit card information, personal information, etc.

## What is Computer Forensics?

- ? Computer Forensics – “*the application of computer investigation and analysis techniques*”
- ? Increased use of computer and network technologies has led to
  - new crimes
  - new ways of committing 'old' crimes
- ? Computer forensics aims to
  - identify root cause of an event
  - identify responsibility for an action

## Computer Forensics vs Security

- ? Distinct differences between the two fields
- ? Computer Forensics attributes culpability
  - identify the responsible individual or system
  - to process the evidence
  - to seek punishment/retribution
- ? Computer Security aims to maintain system integrity
  - confidentiality
  - integrity
  - availability

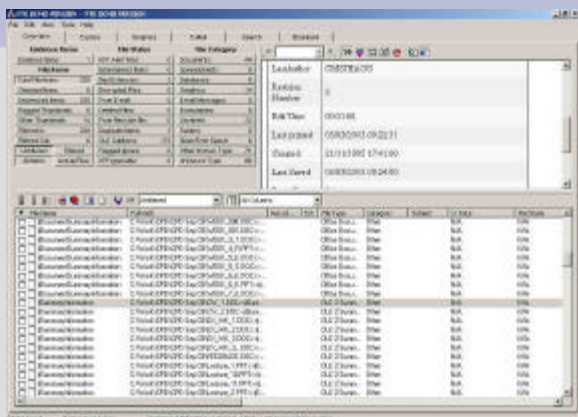
## Computer Forensics vs Security

- ? Degree of overlap between raw material or techniques used by both fields
  - eg. Forensics based on IDS approaches
- ? Forensics and Security sometimes have different and opposing aims
  - the user perspective
  - the investigator perspective
- ? Security functions only implement minimal logging
- ? Security countermeasures may work against computer forensics investigations

## Current practice

- ? Time consuming and laborious!
- ? Collect evidence and copy taken
- ? Forensics tool to recreate the logical structure of underlying OS
- ? View files
  - extant and deleted
- ? Report suspicious/malicious files with supporting evidence
  - eg. time/date/user created, accessed or modified
- ? Present evidence (to court, management, etc.)

## Example - FTK



## Current practice - problems

- ? Requirement for fast data searches in computer forensics
- ? Practitioners realise limitations with current tools
  - not designed for current hard drive capacities
  - time pressures from investigating team for results
- ? Currently, some practitioners improve speed of search through MD5 hash searches
  - MD5 reliability issues
  - change 1 byte to defeat search

## Related literature

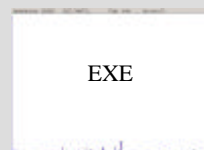
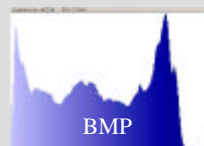
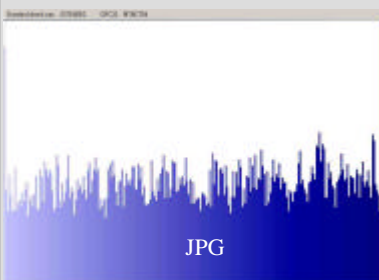
- ? File analysis approaches eg. *FTK* or *EnCase*
- ? Format specific approaches eg. *Jhead* or *Data Lifter*
- ? Statistical analysis of byte distribution eg. *Fileprint* or *Oscar*
- ? File fingerprinting for DRM eg. Schonberg & Kirovski, etc.
- ? IDS signature analysis eg. Haggerty *et al*, Carey *et al*, etc.

## Forsigs approach

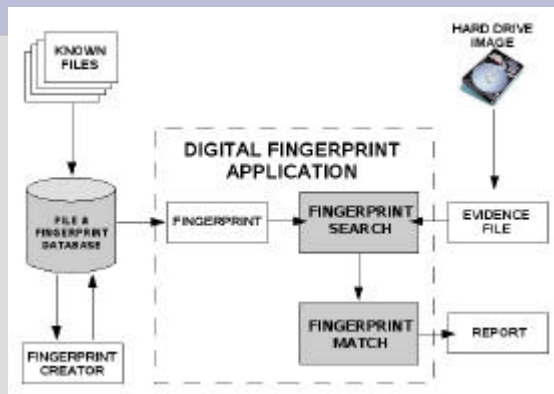
- ? Relies on searching data from byte 0 to byte  $n$  (where  $n$  is the last block of the hard drive)
- ? Makes use of underlying structure of data stored in memory locations
  - Typically 512-byte blocks in Linux, 4,096 bytes in Windows
- ? Design goals:
  - high-speed, large-volume analysis
  - scalability
  - digital viewpoint
  - keep false+ to a minimum

## JPEG searches

- ? JPEGs more suited to fingerprinting due to byte value distribution

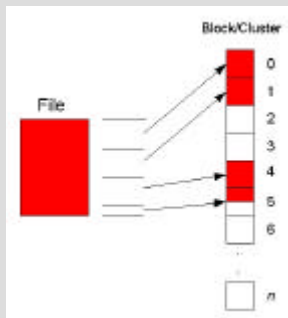


## Forsigs approach



## Fingerprint selection

- ? File written to hard drive is allocated to blocks/clusters

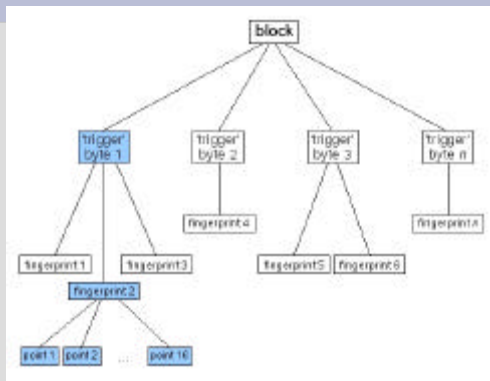


## Fingerprint selection

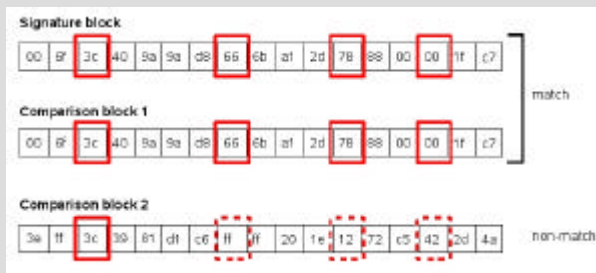
- ? *Ad hoc* block selected from original file through *siggrab* application

```
siggrab@linux:~/forsigs$ ./siggrab chb.dat chbfp.dat
Specifying the original image file size:
2221956
Specifying which block is to be used as the basis for the fingerprint:
600
Block 600 is between 387208 and 387712 bytes into the image file.
***** Fingerprint file created *****
```

## Fingerprint matching



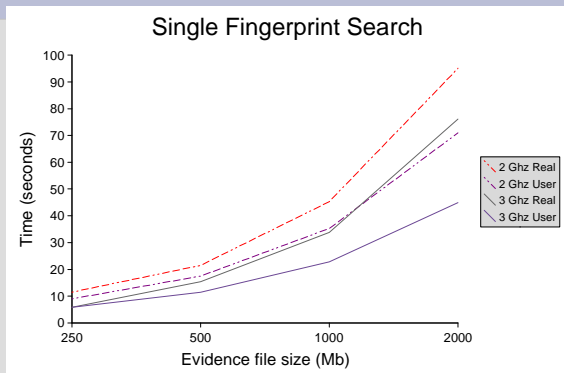
## Fingerprint matching



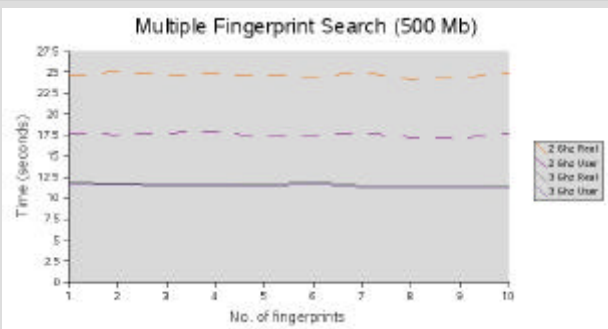
## Forsigs case study

- ? Tests on 2 hosts running Suse 10
  - 2GHz Athlon, 256Mb RAM
  - 3GHz Pentium 4, 1Gb RAM
- ? Speed of search measured in *real* and *user* time
- ? Files ranging from 250Mb to 2Gb
  - Files of various types, eg. MP3, system files (dat, swf, dll), exe, ogg pdf, ppt, doc, etc.
- ? 80 JPEG pictures in 250Mb file and 650 JPEG files in 2Gb search file
- ? No false positives returned!

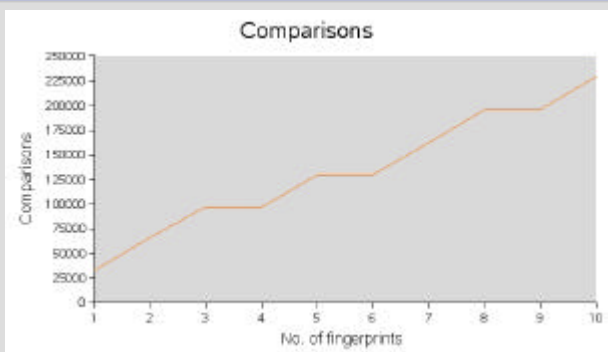
## Forsigs case study



## Forsigs case study



## Forsigs case study



## Forsigs case study - findings

- ? Search time on 3Ghz host faster – 34 secs to search 1Gb (versus 45 secs on 2Ghz host)
  - 55 minutes to search 100Gb on 3Ghz host (versus 75 minutes on 2Ghz host)
- ? Multiple fingerprints have no impact on the search time
- ? Number of trigger byte comparisons have no impact on search time
- ? Forsigs provides an efficient and effective forensic analysis for malicious digital pictures

## Summary

- ? Computer forensics is an emerging and important academic field
- ? Computer forensics and computer security are related but distinct processes
- ? Current practice has a number of issues
- ? File fingerprinting provides a reliable and efficient method for computer forensics analysis for malicious digital pictures

**Now for a demo of Forsigs...**