SANS Summit

September 2010

Bas Kloet

Digital Forensic Investigator

Statistical Sampling
in Forensic Investigations

hoffmann

Trust is good, Hoffmann is better.

Who am I?

- Bas Kloet
  - Digital Forensic Investigator at Hoffmann Investigations since 2007
  - I also give a presentation on Advanced File Carving
- Hoffmann Investigations
  - Founded in 1962
  - Currently bout 80 employees, 1000 cases per year
  - Fraud, theft, industrial espionage

Why statistical sampling?

- Ever growing amount of (digital) data
  - Commercial software/conventional methods are often unworkable with multiple TB of data
- Triage on-site
- Solution: sampling

Different types of sampling statistics

- Prove the presence or absence of evidence
- Determine the scale of unwanted behaviour
- Note: this short presentation focuses on the way statistics can be used, not the underlying mathematics

Presence or absence of evidence

- Prove the presence or absence of evidence
  - Child pornography, incriminating e-mails, false invoices
- Example: 100.000 images, of which 500 (0.5%) are “illegal” (CP)
  - To prove the presence or absence of these images with 99.9% certainty, we only need to investigate a random sample of **1372** images!
  - This saves an investigator from having to investigate the other **98.628** images…

Scale of unwanted behaviour

- Determine the scale of unwanted behaviour
  - The number of CP images that are present on a computer
  - Visited websites/time spent online (employee productivity)
    - Internet history files
    - Proxy logs
  - Sending/receiving unwanted e-mails (computer abuse)
    - Illegal attachments
- Example: 100.000 internet history entries, determine how many are not work related…
  - With a random sample of just 656 entries, you can make predictions about the total set with 99% certainty and ±5% error rate.

Future possibilities

- Integrate this into the main forensic tools
  - Encase image overview
  - FTK search hits
  - Netanalysis internet history
- Find new ways to use statistics in forensic investigations
  - Bayesian logic
  - All the other statistics that financial auditors have been using for ages…

Contact Information

- Bas Kloet (b.kloet@hoffmannbv.nl)
- Robert-Jan Mora (r.mora@hoffmannbv.nl)

- The paper that Robert-Jan wrote on this subject:
  - http://www.forensicfocus.com/digital-forensic-sampling