INSTITUT FÜR
RECHTSINFORMATIK
UNIVERSITÄT DES SAARLANDES

RECHTSWISSENSCHAFTLICHE
FAKULTÄT
UNIVERSITÄT DES SAARLANDES

Challenges of (IT) Forensics

Christoph Sorge
Saarland University

Chair of Legal Informatics
Faculty of Law

About the speaker

- Degree in Information Engineering and Management and PhD in Computer Science

- Professor of Legal Informatics at Saarland University, Faculty of Law; co-opted Professor of Computer Science; associated member of the CISPA Helmholtz Center of Information Security; Senior Fellow of the German Research Institute of Public Administration

- Member of the board, German Association for Computing in the Judiciary (EDV-Gerichtstag)

- > 150 publications of the research group at the intersection of computer science and law; several best paper awards and nominations

Forensic science

- Etymology:
  - derived from latin “forum” (public place)
  - pertaining to courts of law (court was held on a forum)

- Forensic Science uses methods from various scientific disciplines for investigation of crimes

This talk: 2 parts

- A case study on the responsibility of forensic scientists

- Encryption: A challenge of IT Forensics

Statistical considerations

- To some extent, the evaluation of evidence is a statistical consideration
- Incriminating evidence can be present coincidentally

→ Investigators should be able to assess the likelihood of a person‘s guilt given certain incriminating evidence (within their domain of expertise only)

Statistical Evidence: The Sally Clark Case

- December, 1996 and January, 1998: Infants aged 3 months and 8 weeks, respectively, die (both children of Sally Clark)
- In both cases, their mother was alone with the respective child
- First death initially considered as SIDS (sudden infant death syndrome), possibly connected to a respiratory infection
  - Injuries of the child were considered as effects of CPR
- According to autopsy, second death *may* have been caused by shaken impact syndrome (i.e., caused by the mother); therefore also re-evaluation of the first death by the coroner
- Statistical probability for an SIDS of a child of well-off, non smoking parents in a stable relationship (as in this case): 1/8543

Expert witness assessment

- Assessment by another paediatrician as expert witness in court: Probability for two cases of SIDS in a family is (1/8543)×(1/8543) = 1/72.982.849

- Comparison by the expert: This corresponds to the probability of betting on an 1:80 outsider in the “Grand National” horse race four years in a row and to win each time

- With 700,000 births per year in Great Britain, this corresponds to two “double SIDS” cases in a family per century

- Therefore, Sally Clark was convicted to life imprisonment (appeal was rejected)

Expert witness assessment

- Numerous mistakes made by the expert
- Independence assumption of SIDS deaths within a family – Hill (2004) assumes a risk increase by a factor of 10 if a sibling was previously affected
- Use of wrong probabilities (SIDS probability within the whole population was 1:1300
    - Factors decreasing the risk (stable relationship, non-smoking and well-off parents) were considered by the expert
    - Factors increasing the risk (e.g. sex) were not considered
    - Estimate for number of cases in Great Britain was computed using this low probability to achieve the result of two “double-SIDS” cases per century
- Actual number of “double-SIDS” cases according to Hill (2004): 4 to 5 cases per year

Core issue

- Relevance of the computed probability?
- Even if correctly computed: Probability of two children in a *randomly selected* family (with two children) dying of SIDS
- In the actual case
  - No random selection
  - Probability irrelevant—children were known to be dead

Remark

- Probability for winning the (German) lottery “6/49” and having the correct “super digit” with one try: 1:139.838.160 (about half the computed probability for double SIDS)
- In 2015: 26 lottery winners in Germany with 6 right numbers and the super digit
- Low probability → assume foul play and prosecute the lottery winners?
- (Note: Players can easily buy several lottery tickets, so the comparison does not work in all details)

Consequence

- Probability for two SIDS cases in a family *is* low
- Large number of families in the country → high probability for at least one family to be affected
- Problem: Looking for "lottery winners" (here: families with two dead children) in the whole population, then using the low probability as evidence against the selected "lottery winner" (or family)
- Probability of the Clark family being affected was not low—it was selected *because* it was affected *(prosecutor's fallacy)*

Actual goal

- We are actually looking for: Probability for a person being innocent despite the presence of incriminating leads (alternatively: Probability for the person‘s guilt given the same leads)

- Of course, this depends (*among others*) on the probability of those leads occuring randomly

Bayes in the Clark case

- According to Bayes

$$P(I|L) = P(L|I) \times \frac{P(I)}{P(L)}$$

Probability for a randomly selected person being innocent (in the Clark case: mother not being a child murderer) (Very high)

Probability for Sally Clark being innocent given the lead (in this case: two dead children)

Probability for the lead (two dead children) occuring despite the mother being innocent (very low)

Probability of the lead (in the Clark case: two dead children in the same randomly selected family) occuring (very low)

Only probability considered by the expert

Note

- Not all probabilities are known
- Alternative modelling possible
- Hill (2004) summarizes
  - Low probability for two SIDS deaths in one family
  - Probability for double murder by a mother is even lower, though
  - Estimate using existing data about infants:
    - 17 times as many SIDS deaths as murders
    - 4.5 to 9 times as many double SIDS deaths as double murders
    - Insufficient data about triple deaths (guess by the author: triple SIDS deaths slightly more likely than triple murder)

The Sally Clark case: Result

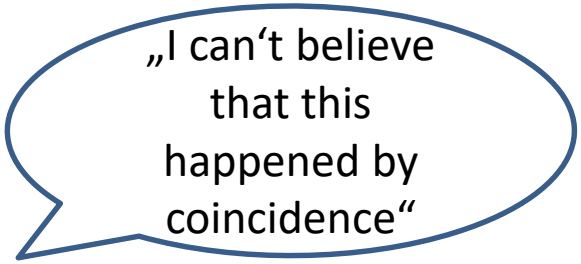
- Criminal Cases Review Commission investigated the case; acquittal and subsequent release of Sally Clark (in 2003, after about 5 years in prison)

- Two other mothers, both convicted of double murder, based on assessments by the same expert, were also released after new trials in the same year

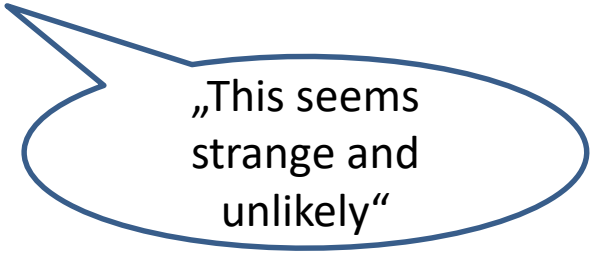
- Another woman, charged with murder of her three children, was also found not guilty despite the same statistical argument made by the same expert (also in 2003)

Lessons learned

- Be careful with statistical arguments



- This also holds for “implicit” use of statistics



- Be aware of the prosecutor’s fallacy
  - Do you consider a lottery winner as a fraudster
    because winning the lottery is unlikely
    or because you had evidence for manipulation of the lottery by the suspect and *then* saw him win the lottery?

Generalization

- Evaluating evidence is relevant to forensics in general, including IT forensics
- Examples of potential statistical “traps”
  - Face recognition
  - DNA matching
  - Writing style recognition
- Next: A non-statistical challenge of IT forensics

This talk: 2 parts

- A case study on the responsibility of forensic scientists

- Encryption: A challenge of IT Forensics

IT Forensics: Encryption as a challenge

- Data encryption used to be an exception – most data was stored and transmitted without encryption
- Nowadays: Widespread adoption of encryption
- Breaking state of the art encryption, based on published standards: Practically impossible *if properly implemented*

Common occurences of encrypted data

- Data in transit
  - Virtual Private Networks (IPsec, OpenVPN, …)
  - Transport Layer Security
  - Application-layer encryption (PGP, S/MIME etc.)

- Data at rest
  - Encrypted hard disks and other storage media (entire medium or individual partitions)
  - Encryption within the file system
  - Encryption by applications (e.g. encrypted Zip archives, GPG encryption etc.)

How to deal with encrypted data

- Your ideas?

How to deal with encrypted data

- Is access to the encrypted data even necessary?
  - Access to unencrypted copy possible (e.g. backups)?
  - Metadata available and sufficient?

- (Potentially) available metadata, data in transit
  - Addresses (sender/receiver): IP addresses, port numbers, host names and other application layer addresses
  - Overall transfer volume and distribution over time
  - Times at which communication occurs

How to deal with encrypted data

- Is access to the encrypted data even necessary?
  - Access to unencrypted copy possible?
  - Metadata available and sufficient?

- (Potentially) available metadata, data at rest
  - Size of encrypted volume or content
  - Timestamps, file names and other file system metadata (in case of individual file encryption)

Circumvention of encryption

- For data in transit: Circumvention of encryption
  - Access to data before encryption or after decryption
  - Requires access to the respective systems, usually by exploiting security vulnerabilities → installation of surveillance software
  - Note: Legal limitations to this approach

Decrypting encrypted data

- Decryption possible with access
  - to the decryption key
  - to the password if key is derived from a password

- Access to decryption key
  - DMA attacks or cold boot attacks if system is running and decryption key is in memory
  - Checking for hardware tokens

Identification of encryption keys

- How to find encryption keys in memory images?
  - Basic approach: Check for areas in memory with high entropy
  - For private keys in asymmetric encryption schemes: Low probability for random data to be a valid encryption key
  - Symmetric encryption keys: Consider additional information / redundancy like in AES key schedules
  - Cold boot attacks: Consider errors in memory images

Passwords

- Potential for password cracking
  - Brute force attacks
  - Problem (example, optimistically assuming 14 trillion attempts per second, achievable for single hashing with crypto miner hardware)

26 characes [a-z]

length	time
5	< 1 s
6	< 1 s
7	< 1 s
8	< 1 s
9	< 1 s
10	10 s
11	4 min.

26 characters [a-z] Slowdown by a factor of 1000

length	time
5	< 1 s
6	< 1 s
7	< 1 s
8	15 s
9	6,5 min.
10	2,8 h
11	3 d

52 characters [A-Z;a-z]

length	time
5	< 1 s
6	< 1 s
7	< 1 s
8	4 s
9	3 min.
10	2,9 h
11	6,2 d

52 characters [A-Z;a-z] Slowdown by a factor of 1000

length	time
5	< 1 s
6	1,4 s
7	1,2 min.
8	1 h
9	2,3 d
10	120 d
11	17 years

Passwords

- Potential for password cracking
  - Password reuse (from other services which are easier to attack – extreme case: password contained in password manager software)
  - Dictionary attacks, also taking into account knowledge about the user (name of pet etc.)
  - Current password cracking software allows setting rules, e.g. use dictionary passwords corresponding to password policy, “l33t” replacements etc.

Attacking hashed passwords: Precomputation

- Only hash value of password is stored → no direct access to password
- Attack 1: "Brute force"
  - Test all possible passwords and calculate hash value
  - Problem: very long runtime
- Attack 2: Look-up table
  - Pre-calculate hash values of all possible passwords → store in table
  - Look-up hash in table → get password
  - Problem: very large storage space required (e.g. 6 digits → 1.4 TByte)

Rainbow Tables

- Hellman Tables and the successor concept of Rainbow Tables allow a time/memory trade-off for the storage of pre-computed password hashes

- Problems with this approach
  - Most, but not all passwords of a given length will be contained in the table
    → success not guaranteed
  - False positive: Hash value found, but corresponding password not in the table

- Countermeasures getting more and more common

Usage of rainbow tables in forensics

- Usage for state-of-the-art encryption of data at rest: Limited due to use of proper key derivation functions with salt mechanism
- Usage for data in transit
  - e.g., 802.11 security with password authentication ("WPA" / "WPA2"): Use of SSID as a salt, rainbow tables exist for common SSIDs
  - Note: Change of key derivation in WPA3
- Usage for authentication
  - Stored hashed passwords if salts are not used (still common due to large number of legacy applications) → password reuse

General remark

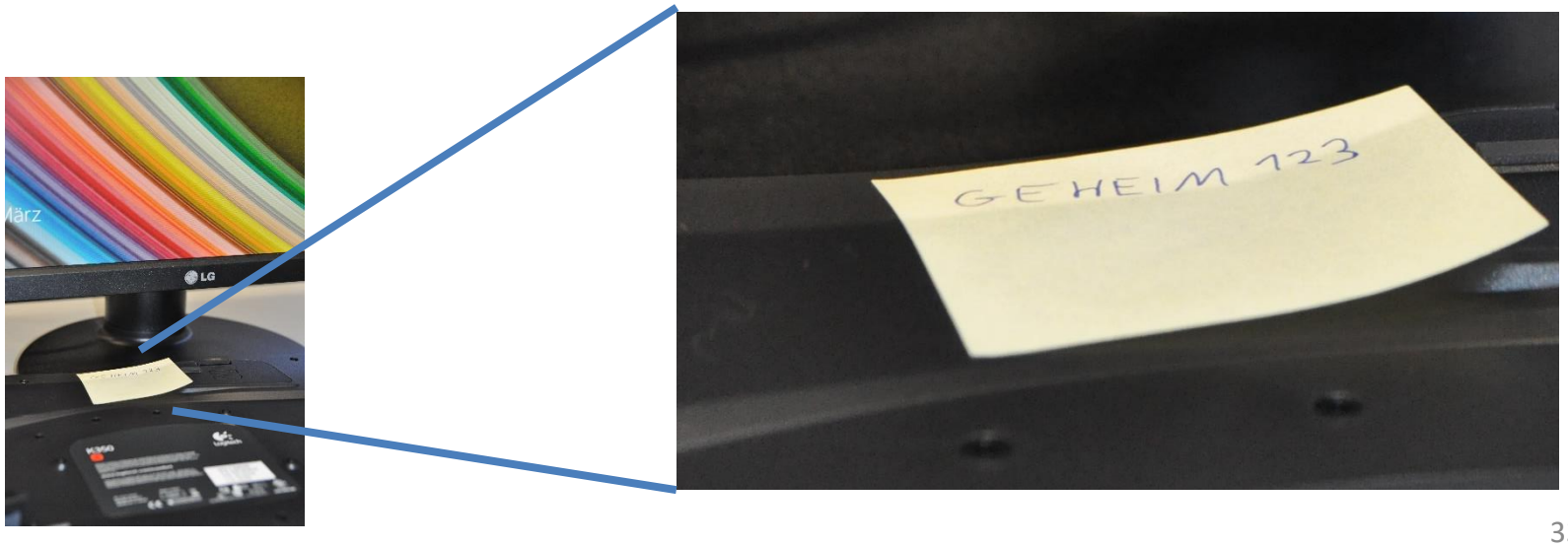
- Better understanding of IT security in the general public, by software developers and system administrators
  → basic attack modes like brute force attacks get increasingly difficult

- Example: Use of better key derivation functions
  - Generation of key from password using salt and multiple iterations of a hash function
  - Also: Use of specific hash functions that are hard to speed up (intentionally hard to parallelize, memory-intensive etc.)

- Individual programs may still use weak encryption (e.g., in proprietary data formats)

General remark

- Non-technical attacks may work as well
- Still a chance to find copies of encryption keys and passwords (backups etc.)



Conclusion of Part 2

- Little chance of breaking encryption
- Instead: Attempt circumvention
  - Exploit weak passwords
  - Exploit password re-use, e.g. by attacking passwords with rainbow tables
  - Get data before encryption or after decryption

Conclusion

- We have considered two key aspects of forensics
  - Appraising evidence
  - Finding evidence

- Lawyers working in the field should be aware of the challenges involved in both
  - Need for statistical and technical training
  - Your opinion?