Forensic Genomics: Kin Privacy, Driftnets and Other Open Questions

Frank Stajano, Lucia Bianchi,
Pietro Lio, Douwe Korff

Workshop on Privacy in the Electronic Society (WPES'08)

Motivation

- Looks at privacy from a technical, legal, and ethical point of view.
- Genetics and genomics.
- DNA analysis is commonly being used in forensic.
- DNA reveals a lot of information !!!
- To build a policy for regulating DNA collection, storage, and processing

Era of Genomics

- Genetic analysis is typically done for
  - Forensics
  - Medicine
- At present: DNA matching for few specific segments of genome
- New trend: Can get the entire DNA sequence
  - 3 billion base pairs
  - About 20,000 genes

- DNA sequencing is getting cheaper
  - Example of James Watson
  - Two months and less than $1 million
  - Your genetic sequence may reveal your ancestors !!
  - Watson: “Black Africans are not as intelligent as Whites”
    - Contradiction

Important Facts

- Predicts genome sequencing may cost 1000 USD in near future
- Difference of other biometric techniques and DNA matching
- Forensic analysis performed on small segment of the genome
  - Analyzing whole genome may reveal intimate psychological and medical condition
  - Blueprint for creating the organism!

Important Facts

- Capabilities of DNA analysis will be boosted by
  - Better acquisition
  - Efficient sequencing techniques
- Forensic analysis of DNA is error prone
  - How many matches before you say it is OK?
- Kin Privacy
  - You genome may be inferred from your relatives!!

A look into future

- A private investigator collects samples
  - Hairs, skin flakes
  - Obtain genome to learn about cancer, food allergies, other diseases
- Government holds genomic databases of their sequences
  - Clerks access data
  - Loses laptops/discs with genome data !!

In the future …

- Full genomic data revealed for medical research
  - Anonymized
  - Simulation tool predicting physical appearance
- Insurers demand genetic screening
- At birth, stem cells are preserved at a medical bank for future treatment
  - Security violation

- Like feticide
  - People may decide to kill babies rated “not good enough”
- Kin privacy

- Key Issues
  - For what purposes should DNA analysis be used?
  - How should it be protected?

Open Questions

- If police routinely acquire DNA samples
  - Should they be allowed to infer gender, race, diseases of the person
  - How to support legal prohibitions technically
    - Cryptography
    - Tamper proof analyzing machines that give answer to allowed queries
    - Correct use of such a machine

- Misuse of genome databases by the state
  - Political advantage
- Communist East Germany
  - Stasi collected sweat samples of dissidents
  - Stored in a “smell bank”
  - In 2007, similar techniques were used for preventive tracking of G8 demonstrators

- Phil Zimmermann, 1996
  - *This is unsettling because in a democracy, it is possible for bad people to occasionally get elected - sometimes very bad people. Normally, a well-functioning democracy has ways to remove these people from power.*
  - *But the wrong technology infrastructure could allow such a future government to watch every move anyone makes to oppose it. It could very well be the last government we ever elect.*
  - *When making public policy decisions about new technologies for the government, I think one should ask oneself which technologies would best strengthen the hand of a police state.*
  - *Then, do not allow the government to deploy those technologies. This is simply a matter of good civic hygiene.*

Fallibility of forensic DNA analysis

- Samples are anonymized using barcodes
- Intentional and unintentional errors in matching
- Insider fraud
- Contamination of samples
- DNA can be newly synthesized !!

Kin Privacy

- DNA of two persons matches 99.9%
- Much similar for persons having common ancestry
- How about twins?
- Your relative reveals much of your genome
  - Case study 1
    - 2 Murders in 1980
    - Solved in 2000
    - DNA match found with nephew/brother of killers

- Case study 2:
  - Big Pharma wants your genome
  - Should you share the profits with the kins?
  - What if someone opposes?

Cold and Hot Hits

- Hot hit: A preliminary set of suspects is prepared by lawful investigation
- Cold hit: Scene of crime samples are matched against the whole database
  - Should the law allow investigation of innocents?

Cold and Hot Hits

- Forensic genomics is done on statistical models
- Correctness varies
- In forensics, Likelihood Ratio (LR) is measured for finding a culprit
  - H_0: Sample found belongs to the suspect
  - H_1: Sample does not belong to suspect
- Measure ratio of max likelihoods of each hypothesis
- Cold hits gives better accuracy
- Is it a good idea to have a DNA database?

Regulatory Laws

- Genetic Information Nondiscrimination Act (GINA)
  - Prevent use of genomic materials for employment and insurance
- Ban on mass genetic screening

Challenges

- For sample acquisition
  - Consent of persons is necessary
  - Policy specifying purpose of acquisition
- Technical framework to prevent use of genomic data to specific purposes
- How much genomic data is necessary?
  - Technical framework to ensure that no more is revealed
- Can sequences be split and placed under the care of different organizations?

Challenges

- Build tamper-proof blackboxes
  - Allow only specific queries and give answers
  - Genome sequence not revealed