How to Privately Find Double Acquisitions in Biometric Databases

Orr Dunkelman

Department of Computer Science, University of Haifa

January 15th, 2015

Joint works with Melissa Chase and Margarita Osadchy

Outline

The Israeli ID System

- Any Israeli citizen is assigned a 9-digit ID number.
- Actually, there are 8 digits (the ninth serves for error detection).
- Once reaching the age of 16 an ID card may be issued (by the age of 18 it is mandatory).
- This ID card is supposed to be carried at all times for identification.
- In real life, the ID number is sufficient for any practical purpose. . . (similarly to SSN in the US)

The Israeli ID System (cont.)

- The ID card is an identification form.
- One can usually get away with showing other credentials (driving license, passport, etc.)
- For some rare cases, only the ID card is valid as a form of identification.

The Israeli ID Card

- Each ID card is actually a laminated card.
- Contains a picture, the ID number, and some fixed identification information.
- The card has an appendix which contains some additional information which may change over time (current address, kids, etc.).
- The appendix is a simple piece of paper with no practical identification value.

So What's Wrong?

- ID card have no validity. This means that the picture can be 30 years old and the card still valid...
- Moreover, forging an Israeli ID card is not hard.
- Printing a new one from scratch is easy (requires a bit of practice).
- But actually taking someone else's ID card and replacing the picture is extremely easy.

So how come identity theft attacks are not a big issue in Israel?

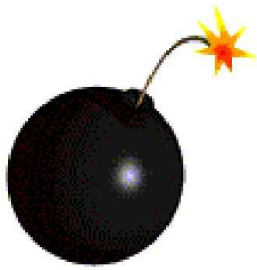
Forging ID Cards in Israel

- Most fake ID cards are actually legitimate cards that were modified or abused.
- The reason the person who had "lost" the identity card does not complain about identity theft is because most fake ID cards are using:
  - Dead people's ID cards.
  - ID cards of people who left Israel for good (with their consent).
- And usually the adversary is after the "government":
  - Collecting social security benefits (and similar support).
  - Voting in the elections...

  and not after the person whose identity was stolen.
- And sometimes, they actually do the "victim" a good service.

The Security Threat

- Besides these issues, Israeli IDs are extremely useful to people who try to enter Israel illegally.
- These people usually obtain a completely fake ID, and use it at security check points or when the police stops them for interrogation.
- These cases are usually easy to detect:
  - Sometimes the person is too young (or too old) for the claimed date of birth.
  - Sometimes the checksum is wrong.
  - The policeman/solider can query the database.



The Current ID Database

- Currently, the Israeli Ministry of Interior Affairs has a database about the entire Israeli population.
- This database contains all identification details, as well as family relations.
- It is composed of private and sensitive information (according to the Israeli law).
- Some parts of the database are given to various entities (banks, insurance companies, political parties).
- The database was supposed to be kept secret, but since several years now, it is possible to find the full database online. It is also updated every now and then.
- Recently, (one of the) responsibles for the 2005 leak was sentenced to prison.

The New Proposed ID Card System

- To overcome these issues, the Israeli parliament has discussed a new law concerning the ID System.
- The law suggests three methods to fight the counterfeit ID cards by offering three mechanisms:
  - The use of new ID cards which contain a smart card.
  - The use of biometric information for identification. The information will be stored on the smart card, signed by the state.
  - The establishment of a database containing the biometric information of all citizens.

The Good Parts

- Verifying the authenticity of ID cards will become simple.
- Identifying whether the person matches the ID will become simple.
- The new ID cards will have **an expiry date**.
- Hopefully, many fraudulent people will be purged from the database.

Purging "Nonexistent" Entities

- The idea is that the nonexistent entities will not be issued a new ID card, as they cannot arrive to the acquisition stations.
- However, it is very obvious that the holders of the forged cards will be able to arrive and claim a false new ID card.
- For that, during the first acquisition of a person, the identification will not be based solely on the information from the Ministry of Interior Affairs.
- In other words, they will be given access to more private data...
- At the end, if you have the cooperation of the person whose ID you are stealing, you can succeed in obtaining an additional ID card.

Biometric Database

- To solve the issue of one person holding multiple ID cards (up to people who change their declared identity once) a biometric database will be used.
- During the acquisition process, each citizen's biometrics will be measured and stored in the database.
- Then, collisions in the database will be found using simple forensics tools.
- Also useful to identify people once the ID card is lost (or stolen).

The Biometric Data

- The Israeli law dictates that the biometric data will be composed of:
  - High resolution photo of the face,
  - Fingerprints of both index fingers (left/right).

Privacy Concerns

- Private sensitive information will be kept by people who failed to safely keep other sensitive information.
- Many other entities will be given access to the database (the police will be granted full access, other entities may accept some restricted access rights).
- Database leakage means that everyone has access to your private biometric data.
- Finally, one could perform reverse queries — given biometric data, identify the person who owns it...

Privacy Concerns — Doom's Day Scenarios

- You use your fingerprint as the key for your hard disk encryption. Anyone with access to the database could decrypt your hard drive.
- Or the use of your biometric in access control mechanisms (what you are becomes what everybody can be).
- After leakage of database — someone plants your fingerprints in crime scenes (and fingerprints will mean nothing in court).
- After leakage of database — Israeli citizens can be easily identified.
- As the database contains high definition picture of all citizens, one can easily connect the database with CCTVs to obtain all-time surveillance.

Solving the Issue with no Database

- All kids, once reaching the age of 6 (or so), receive a kid's ID card (and have their biometrics sampled).
- At the age of 12, they are re-sampled and obtain a youth's ID card. They prove their identity using the kid's ID card.
- At the age of 18, they obtain the adult ID card, after proving their identity using the youth ID card.
- When an ID card is lost, in addition to **your attestation** and query, bring some **verified** others to attest that you who you claim to be.
- To subvert the system, the cheater will need to plan 12 years ahead.
- Note that the current cheaters will remain in the system, but as no new cheaters can join, over time there will be no cheaters.

Solving the Issue with no Database (cont.)

- To recover lost ID card — just issue two cards to each person.
- Or issue two cards, and store one in an official archive.
- Or perform a really lengthy check for people who lost their ID cards to overcome fraudulent behavior.
- But then no reverse queries will be possible, and sometimes these may be beneficial (e.g., an Alzheimer patient).

Formal Requirements from the Solution

- **Complete Privacy**: Given a person (and his biometric data), we could not determine whether he is in the database or not.
- **Fraud Detection**: Any person who holds two different ID cards will be detected.
- **Honest Detection**: It is impossible for anyone to forge fraudulent behavior of someone.
- **Controlled Reverse Query**: Given a court order, and only given such an order, a reverse query identifying the owner of a specific biometric data is done.

A Simple (Impractical) Solution

- We can easily distribute the data between various trusted entities using a secret sharing scheme.
- The other functionalities will be achieved by using standard secure multiparty computation techniques and protocols.
- Problems:
  - This solution is inefficient (large databases, communication overhead, etc.)
  - This solution is insecure (once a share is leaked it is lost, unless the entire database is re-shared).
  - Trusting the entities that have a lot to gain at any point of time to collude.

The Involved Entities

- Users — posses an ID number id and some biometric information $\mathcal{B}$.
- Database — collects the database entries, detects frauds.
- Blinding entities — generate blinded database entries, that can be opened only in case of a fraud.

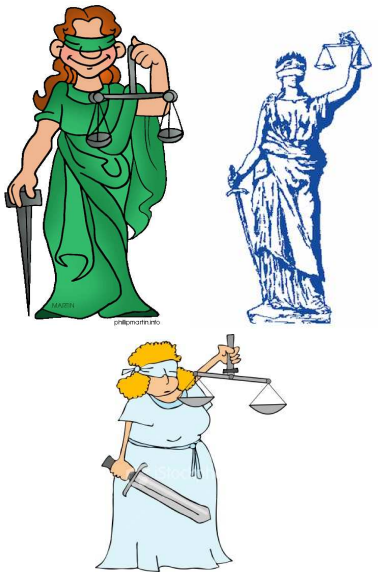




Acquisition of New Users

1. The user arrives to the acquisition station.
2. The user identifies himself, giving his ID number id.
3. The user's biometric data $\mathcal{B}$ is sampled.
4. The user and the blinding entities run a secure multiparty computation to compute the blinded entry.
5. The blinded entry is stored in the database.

id ,$\mathcal{B}$



The Blinded Entry

- Assume that there are ℓ blinding entities.
- Let s, r, t, z be secrets shared between the blinding entities, and let f be a PRF.
- The blinded entry is:

$$f_s(\mathcal{B}), f_r(\mathcal{B}) \cdot f_t(\mathcal{B})^{f_z(id)}, f_z(id), E_{f_r(\mathcal{B})}(id), \sigma_1, \ldots, \sigma_\ell$$

- $E_k(\cdot)$ is some (symmetric-key) encryption function.
- σ_i is a signature from blinding entity i (attestation that i "approves" the computation, entry, etc.)

Computing the Blinded Entry

- As there are ℓ blinding entities, we run a secure multiparty computation between $\ell + 1$ entities.
- The inputs of each blinding entity is s_i, r_i, t_i, and z_i the shares of $s, r, t,$, and z, respectively.
- The inputs of the user are $id, \mathcal{B}$.
- Both the user and the blinding entities use special hardware which is publicly verified.
- Hence, the secure multiparty computation is relatively efficient.
- We can also set the computation to be resistant to several malicious blinding entities.

Detecting Fraud

- The blinded entries are stored in the database.
- If two entries share the same $f_s(\mathcal{B})$, then they encode the same biometric data.
- Consider the two colliding entries:

$$f_r(\mathcal{B}) \cdot f_t(\mathcal{B})^{f_z(id)}, f_z(id), E_{f_r(\mathcal{B})}(id)$$

$$f_r(\mathcal{B}) \cdot f_t(\mathcal{B})^{f_z(id')}, f_z(id'), E_{f_r(\mathcal{B})}(id')$$

- As $f_z(id)$ and $f_z(id')$ are known, it is possible to extract $f_r(\mathcal{B})$, allowing the decryption of id and id'.

Lost ID Cards

- When a person loses his ID card, he has to arrive once again to an acquisition station, and provide his *id* number.
- The blinding entities just run the acquisition process again, generating the same

$$f_s(\mathcal{B}), f_r(\mathcal{B}) \cdot f_t(\mathcal{B})^{f_z(id)}, f_z(id), E_{f_r(\mathcal{B})}(id)$$

  which can be easily found in the database.
- Obviously, if the entry does not appear in the database, we have detected a fraudulent behavior.

Dealing with Honest Collisions

- There are about 8,000,000 Israeli citizens.
- It may happen that $\mathcal{B}_1 = \mathcal{B}_2$ for two different people (if $|\mathcal{B}| \approx 50$).
- When fraud investigation is initiated, two different people are actually found.
- In such a case, their entries are removed from the database and replaced by:

$$< f_s(\mathcal{B}), \text{“}collision\text{”} >$$

- The biometric data is sampled again (for both parties) using a second procedure (which produces more entropy).
- This new biometric data is used to compute again two blinded entries that go to a second database (of “enhanced” acquisition process).

Reverse Queries — Identifying the Owner of Biometrics

- The court issues an order that the identity corresponding to $\mathcal{B}$ is to be revealed.
- The blinding entities receive the order, and use the acquisition process to generate

$$f_s(\mathcal{B}), f_r(\mathcal{B}) \cdot f_t(\mathcal{B})^{f_z(0)}, f_z(0), E_{f_r(\mathcal{B})}(0)$$

- A collision in the database suggests the identity of the owner.
- This entry is not to be added permanently to the database.
- Note that if a threshold of the blinding entities refuses to participate, then there is no way for the query to complete.

Membership Queries

- Membership queries cannot be done without the collaboration of the blinding entities.
- Namely, even if we hold the blinded entry and the biometric information embedded in it, we cannot tell whether this is the case.
- Hence, even if the database is published (or leaked), the privacy of the citizens is kept.

Consistent Sampling

- The first solution assumes that one can recover $\mathcal{B}$ of a given person consistently.
- This is a very strong assumption. Which is not fully needed.
- To catch frauds, it is sufficient to have a high success rate (e.g., 90% chance).
- So we need to obtain consistent samples 90% of the time.

Consistent Sampling (cont.)

- The sampling is done in acquisition systems only.
- By providing these with high quality equipment we can increase the chance of consistency.
- We note that we can even use different biometrics than the one stored on the ID card itself.
- For example, an iris scanner, a 3D scanner, etc.
- From the obtained data, we just need to extract $\mathcal{B}$ in a relatively consistent manner.
- Unfortunately, it seems that this is not yet possible.

What Can We Do?

How to Handle Inconsistent Sampling?

- Recall Rita's talk from earlier.
- We can pick a random key K of 75-bit and "encrypt" it using a biometric $\mathcal{B}$.
- When you use the same biometric (with some noise), K will be revealed.

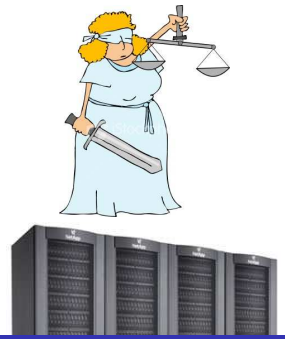
Second solution: Acquisition of New Users

1. The user arrives to the acquisition station.
2. The user identifies himself, giving his ID number id.
3. The user's biometric data $\mathcal{B}$ is sampled.
4. $\mathcal{B}$ is tried against any entry in the database. If an earlier entry "works", we found a double acquisition.
5. If all goes well, the blinding entities pick a random key K, and use the mentioned scheme with $\mathcal{B}$ to generate an entry of the form

$$Code(K) \oplus \mathcal{B}, E_K(0), E_s(id), \sigma_1, \ldots, \sigma_\ell$$

id ,$\mathcal{B}$

s



Quick Analysis

Advantages:

- Does not require consistency among samplings.
- Mostly implemented (if no SMP is needed).
- Frauds immediately detected.

Disadvantages:

- Requires comparing a lot of biometrics (about $(8,000,000)^2/2$ comparisons).
- Requires the cooperation of the blinding entities to recover the fraudulent IDs.
- One can try all possible keys K (2^{75} keys), find K, and extract $\mathcal{B}$. [Partial solution: use $E_{K||r}(0)$ where r is a short random string never stored]

Conclusion

- Presented two solutions for privately finding duplicates in biometric databases.
- In both cases, duplicates are found when adding new entries.
- By controlling the addition of new entries, we obtain privacy and security.
- In theory, one can even publish the entries online and still maintain security.*
- Shares the "trust" between more players, and maintaining an efficient solution.

*In the second solution — only with longer keys.

Additional Issues

- Hardware is to be verified (that it follows protocol).
- Good PRNGs to be used everywhere.
- Probably a good idea to not have a biometric database.

Current State of Affairs

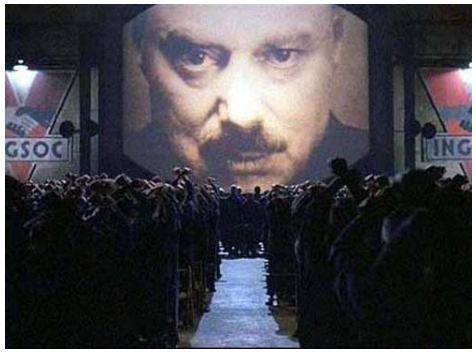
- Legally, we are currently in the pilot phase, which was extended.
- It is very unclear how the success of the pilot will be defined.
- And no real security is being used (despite the clear opposition).
- Joining the database is currently voluntary. Though, a lot of effort was put into promoting the database.
- So expect a catastrophe...

The End of the World is not that Near

- No one connected the high definition pictures to any CCTVs, yet.
- The database contains relatively few people.
- The database has not been leaked. yet.

The End of the World is not that Far Either

- The database is insecure.
- It gives a strong preference to the state over the privacy of citizens, showing lack of understanding of the concepts of democracy.
- Only dictatorships have started to implement such mandatory databases.
- Israel has a long history of “trust me, it’ll be OK” issues.



Questions?

Thank you for your Attention!

For more info:

http://www.cs.haifa.ac.il/~orrd/crypt/biometric.pdf