Towards a Personal Identity Code Respecting Privacy

Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, 14000 Caen, France

Motivations

- Personal information might be collected during usage of Internet services;
- Such information are linked to the user, the browser, or the device;
- But collection of such information poses major privacy issues;
- **Our answer:** a non-cryptographic signature, in form of a binary code, to compare information while respecting user privacy.

Requirements

Non reversibility : the binary code must not give information about the collected personal information;

Confidentiality : the attribute values cannot be known, nor deducted, by the service;

Similarity conservation : if users' personnal information are similar, then their binary code must be too;

Non-usurpation : a third party cannot forge a code enabling him/her to usurp legitimate users' identity;

Revocation : legitimate user must be able to revoke an existing binary code.

Experiments

- **Target :** GREYC, ENSICAEN e-mail lists;
- **Dates :** March 2017;
- **Participants :** 22.

Participants were invited to answer 8 questions on privacy, then to copy an extract of the Universal Declaration of Human Rights. To prevent any influence for the keystroke dynamics, participants were informed of the keystroke acquisitions only from the lasts steps.
With only 22 participants, mostly located in Caen, the sample is not representative, but enables a first experiment of the personal identity code.

Data




Browser

Randomly generated key stored in browser.

Localisation data

IP adress, and locations deduced from it:

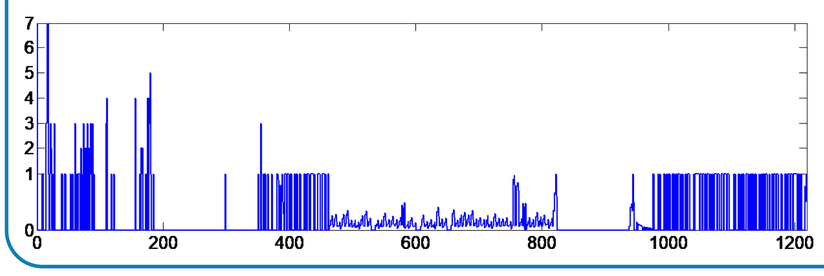
- administrative (e.g. country);
- physical (latitude, longitude);

Network data

HTTP Header fields;

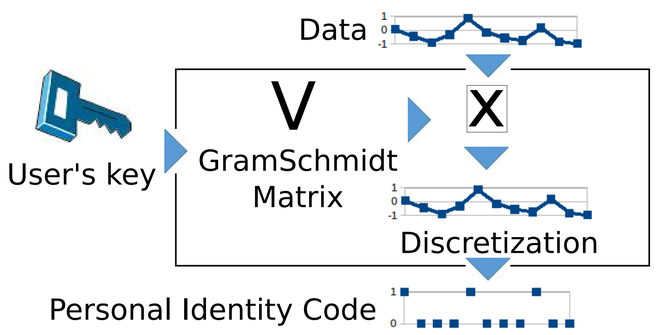
Biometric data

User keystroke;



After pre-processing, a vector of **1218 real values** is obtained from the collected data. Details about the pre-processing step can be found in the joigned paper.

Data protection

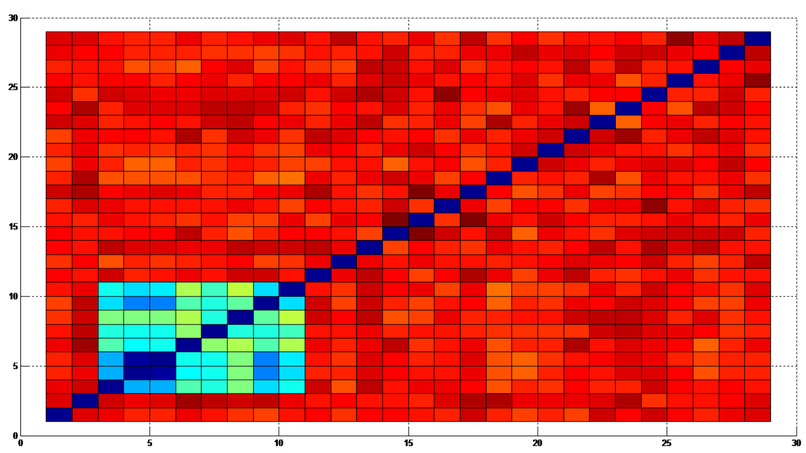


BioHashing (Teoh et al., 2004) allows generation of a binary model called BioCode (our Personal Idenfier Code) from a fixed-size float vector (our Data). This transformation is **non-reversible**, and keep input data **similarity** (in sense of their Hamming distance).
A secret (the user's key) required by this transformation enables BioCode **revocation**.

Personal Identity Code

Generated binary code fullfill our requirements, and might be used, e.g. for **strong user authentication**. However, communication channels, and service-side data still have to be protected in order to prevent **replay attacks**.

Results



Generated personal Identity Code from the experiment have been compared through their Hamming distances. Opposite figure indicates theses distances, blue for an high similarity, red for low.
Signatures **3 to 10** have been generated **by the same user** (i.e. with the same key), but **in different contexts**. Signatures **4 and 5** are judge very similar, this is in fact **the same user in the same context**.
This demonstrates the capacity of the proposed method to produce an exploitable code for personal information similarity computation.

Call for volunteers

Further experiments and collects will be conducted to pursue this work. If you wish to be notify once the collection website available, please send us an e-mail.

Authors

Denis Migdal
dmigdal@ensicaen.fr
GREYC Lab

Christhophe Rosenberger
crosenberger@ensicaen.fr
GREYC Lab

This poster has been presented at ICISSP 2018, Funchal. Authors would like to thank the Normandy Region for the financial support of this work.