PASSE-PARTOUT BIOMÉTRIQUES

Soutenance de thèse

Tanguy Gernot

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

tanguy.gernot@unicaen.fr
https://gernot.fr

GREYC
Laboratoire de recherche en sciences du numérique

Normandie Université

UNICAEN

ENSI CAEN
ÉCOLE PUBLIQUE D'INGÉNIEURS
CENTRE DE RECHERCHE

cnrs

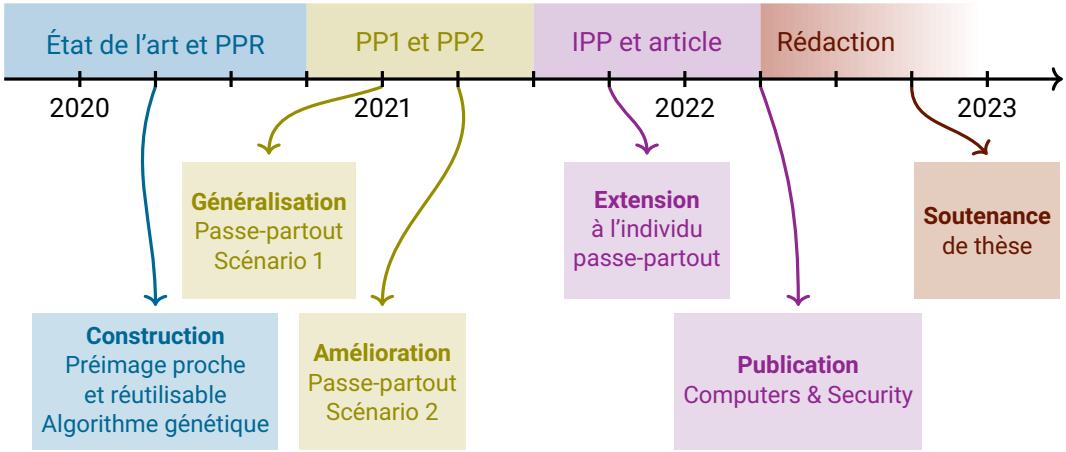
Plan

Plan

À l'Université de Caen Normandie.

- Licence Informatique 2014-2017.
- Maîtrise et Master Sécurité des Systèmes Informatiques (SSI / e-secure) 2017-2019.
- Thèse (bourse ministère) 2019-2022 :
  - Patrick Lacharme (MCF HDR ensicaen).
  - Laboratoire GREYC.
  - Équipe SAFE, activités de recherche en sécurité informatique.

ATER Ensicaen depuis septembre 2022.

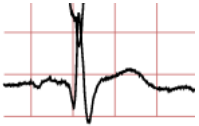

État de l'art et PPR
PP1 et PP2
IPP et article
Rédaction
2020
2021
2022
2023
Généralisation
Passe-partout
Scénario 1
Extension
à l'individu
passe-partout
Soutenance
de thèse
Construction
Préimage proche
et réutilisable
Algorithme génétique
Amélioration
Passe-partout
Scénario 2
Publication
Computers & Security

Plan

Définition de la CNIL

La biométrie regroupe l'ensemble des techniques informatiques permettant de reconnaître automatiquement un individu à partir de ses caractéristiques issues de différentes modalités : **physiques**, **biologiques**, ou **comportementales**.






Les données biométriques sont des données à caractère personnel, car elles permettent d'identifier une personne : ce sont des données sensibles !

Sensibilité des données biométriques

Contrairement à un mot de passe, les données biométriques ne sont pas modifiables à volonté (doigts, visages).

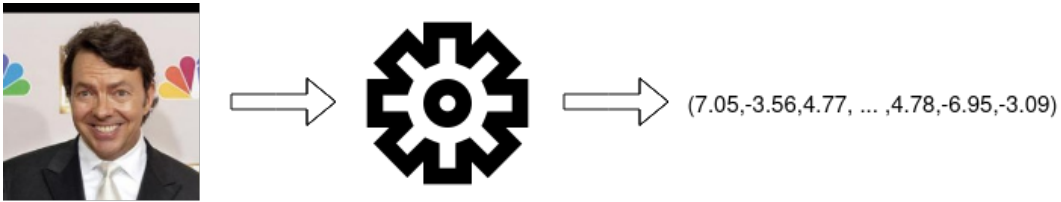
Variabilité des captures

Contrairement à la saisie d'un mot de passe qui ne tolère pas de fautes, chaque capture biométrique d'un individu est différente, mais espérée proche des autres.



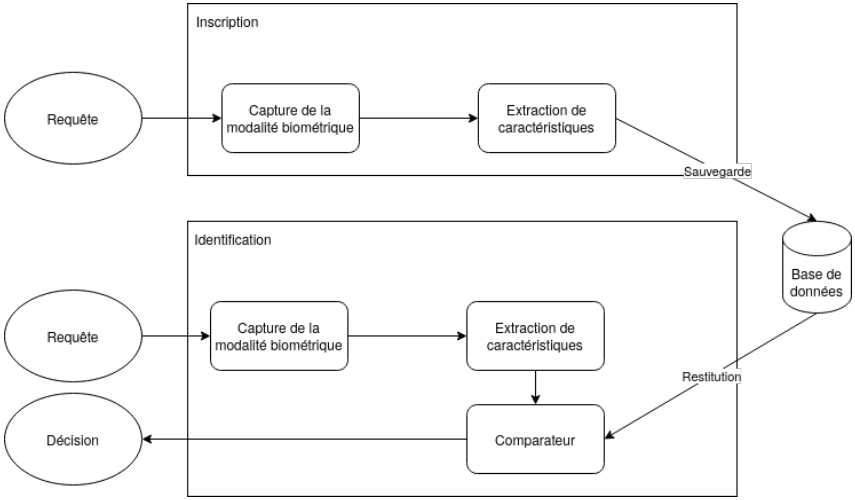
Reconnaître avec des données biométriques

Capture de la modalité $\Rightarrow$ Vecteur de caractéristiques
Réseau de neurones, filtres de Gabor, délimitation d'ondes



Biométrie : concepts fondamentaux et propriétés attendues

Identification

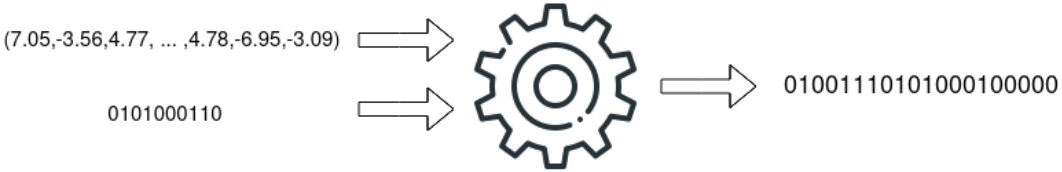


Impératif de protection

Les données biométriques sont des données à caractère personnel et elles ne sont pas modifiables.

Protection $\Rightarrow$ Transformation du vecteur en un gabarit (version protégée).

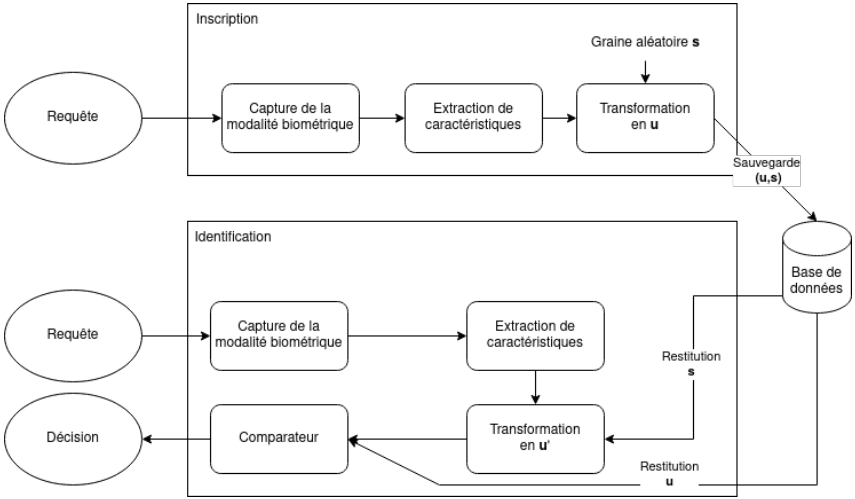
- Transformation paramétrée par une graine.
- Gabarit : petit vecteur (binaire).



- Révocable, non-inversible, performance, indistinguable.
- Comparaison dans le domaine transformé (comme pour les mots de passe).

Biométrie : concepts fondamentaux et propriétés attendues

Base de données biométriques révocables - Identification



Exemple de schéma de protection

Projection du vecteur de caractéristiques en un gabarit avec une matrice $N \times M$ obtenue depuis une graine, puis binarisation (seuillage).

- Biohashing
- Achlioptas $\begin{cases} \sqrt{3/M} \text{ avec probabilité 1/6} \\ 0 \text{ avec probabilité 2/3} \\ -\sqrt{3/M} \text{ avec probabilité 1/6} \end{cases}$

Quelques indicateurs

- FAR : taux de fausses acceptations / FRR : taux de faux rejets.
- Si FAR=FRR alors EER.
- Seuil τ : distance sous laquelle on considère que c'est le même individu.
- Le seuil τ est fixé en fonction de nos besoins en sécurité / utilisabilité (compromis en $\tau@EER$).

Base	Tailles			Originale		Protégée	
	#pers	#capt	vect	EER	$\tau@EER$	EER	$\tau@EER$
FVC (empreintes)	100	8	512	10%	240.7	16.5%	17
LFW (visages)	158	10	512	0.2%	1.227	1.9%	51
PTB (ECG)	158	7	990	10.8%	6321	17%	16

Problématiques

- Les données biométriques sont variables et non modifiables.
- Protection spécifique à cette propriété.
- Étudier la sécurité des transformations de données biométriques.
- Attaque par reconstruction.
- Une préimage d'un gabarit produit un gabarit identique si on la transforme.

Inversibilité des données biométriques

1. Construction de préimages proches et réutilisables.
2. Passe-partout (préimage universelle) depuis des données transformées.
3. Transformer des données pour un passe-partout (vulnérabilités).

Préimages proches et réutilisables.

Nagar et al., 2010

Attaque linéaire d'inversion ~~spécifique~~ au biohashing.
⇒ Construction générique en **boîte noire** (retour du score uniquement).

Lacharme et al., 2013

Construction d'une préimage depuis deux couples (gabarit, graine) issus d'un ~~même~~ vecteur de caractéristiques.
⇒ Construction d'une préimage depuis deux couples (gabarit, graine) issus de vecteurs de caractéristiques **différents**.

Feng et al., 2014

Perceptron pour construire un vecteur depuis un gabarit, puis escalade pour obtenir une capture depuis un vecteur (beaucoup de données).

⇒ Construction de préimages proches **et réutilisables**.

Dong et al., 2019

Construction de préimages proches avec un algorithme génétique pour LFW, 80% avec des gabarits de ~~500 bits~~ (fuite d'informations), 5% avec des gabarits de ~~16 bits~~ (force brute).

⇒ Gabarit de **128 bits**, 100% de préimages proches et réutilisables.

Nanwate and Sadhya, 2020

Construction de préimages ~~proches~~ avec un PSO (optimisation par essaims particulaires), ~~30%~~ pour LFW.
$\Rightarrow$ **100%** de préimages proches et **réutilisables**.

Gomez-Barrero and Galbally, 2020

Étude classifiant en 4 groupes de prérequis les méthodes d'inversion.

- **Connaître le format du gabarit**,
- **+ le score**,
- ~~+ fonction d'évaluation~~,
- ~~+ fonction d'extraction de caractéristiques~~

Existant faible concernant les passe-partout.

Masterprint : Roy et al., 2017, 2018, 2019

Construction de plusieurs empreintes digitales partielles usurpant de nombreux individus multienrôlés (cas des téléphones).
Basé sur les ~~captures~~ d'empreintes digitales ~~uniquement~~ : ajout de minuties.
~~Pas de protection~~.
$\Rightarrow$ Générique, protections utilisées.

Pas d'existant concernant le scénario 2 à notre connaissance.

Plan

Dans la continuité de Lacharme et al., 2013 + HDR

Objectifs

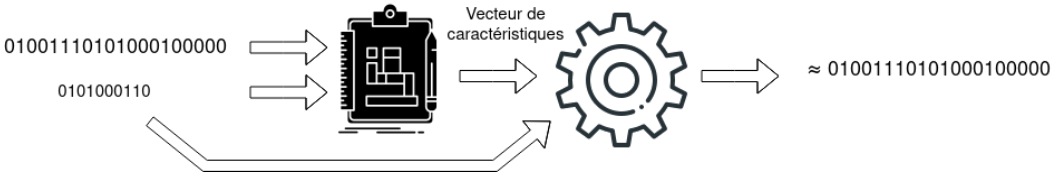
- Modéliser le concept de préimages proches et réutilisables.
- Déterminer les paramètres optimaux de construction avec un algorithme génétique.
- Comparer les performances.

Motivations

- Étudier la propriété de non-inversibilité de la transformation.
- Inverser 2 gabarits en une seule préimage.

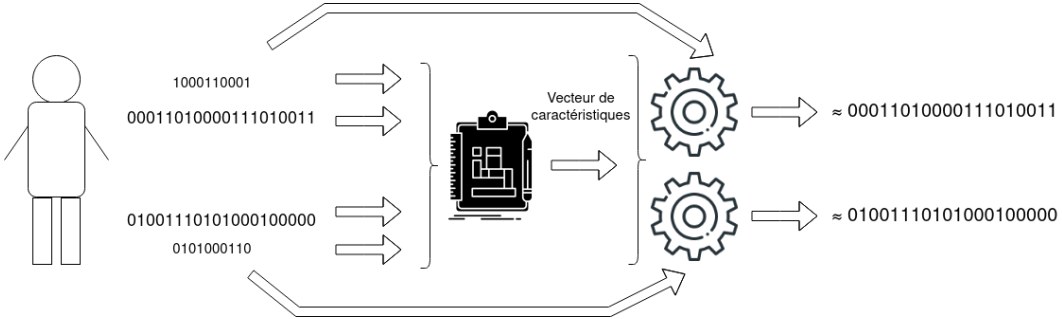
Description d'une préimage

Pour un couple (gabarit, graine), nous générons un vecteur qui, s'il est transformé avec cette graine, procure un gabarit proche.



Description d'une PPR

Pour 2 couples (gabarit, graine), nous générons un vecteur qui, s'il est transformé avec les graines, procure des gabarits proches.

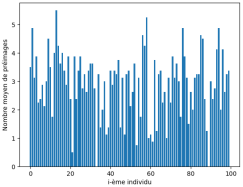


Stratégie

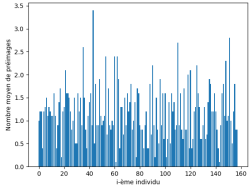
Pour 2 couples (gabarit, graine) d'un individu d'une base, nous testons si les vecteurs des **autres** individus de la base sont des PPR.

Choisie au sein de la base

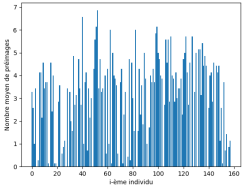
Base	Taux	Nombre moyen de PPR
FVC	96%	2.9
LFW	77%	1.2
PTB	78%	2.9



(a) FVC



(b) LFW



(c) PTB

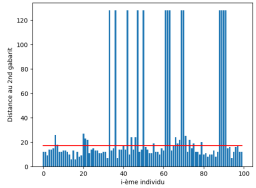
Figure: PPR choisies

Stratégie

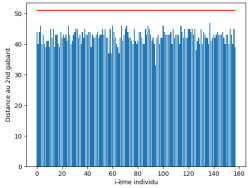
Pour 2 couples (gabarit, graine) d'un individu d'une base, nous construisons **aléatoirement** des vecteurs et nous testons s'ils sont des PPR.

Aléatoire

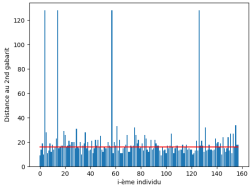
Base	Taux	Distance des candidats	Distance des PPR
FVC	68%	30	12
LFW	100%	42	42
PTB	42%	20	13



(a) FVC



(b) LFW



(c) PTB

Figure: PPR aléatoires

Stratégie

Pour 2 couples (gabarit, graine) d'un individu d'une base, nous construisons **avec un algorithme génétique** des vecteurs et nous testons s'ils sont des PPR.
$\Rightarrow$ Nous souhaitons améliorer les performances (proximité des gabarits) en optimisant son paramétrage.

Qu'est-ce qu'un algorithme génétique ?

- Algorithme d'optimisation : minimiser la valeur d'une fonction d'évaluation f en construisant son paramètre. Inspiré de la reproduction naturelle.
- Population de taille n chromosomes = vecteurs ...
- ... évoluant sur t générations = itérations ...
- dont $n/2$ parents se reproduisent pour donner $n/2$ enfants, avec croisement de leurs gènes et mutations aléatoires.

Fonction d'évaluation utilisée

Somme des distances du vecteur transformé avec les graines aux gabarits

Présentation des résultats

- Diagrammes en boîte (Q1, Q2, Q3, écart interquartile) pour la base PTB.
- Distances au second gabarit agrégées en fixant un paramètre.
- Comparaison avec les méthodes de construction aléatoire et de choix au sein des bases.

Distance au second gabarit

- Taille des gabarits si le vecteur transformé n'est pas proche du premier gabarit.
- Distance du vecteur transformé avec le second gabarit sinon.

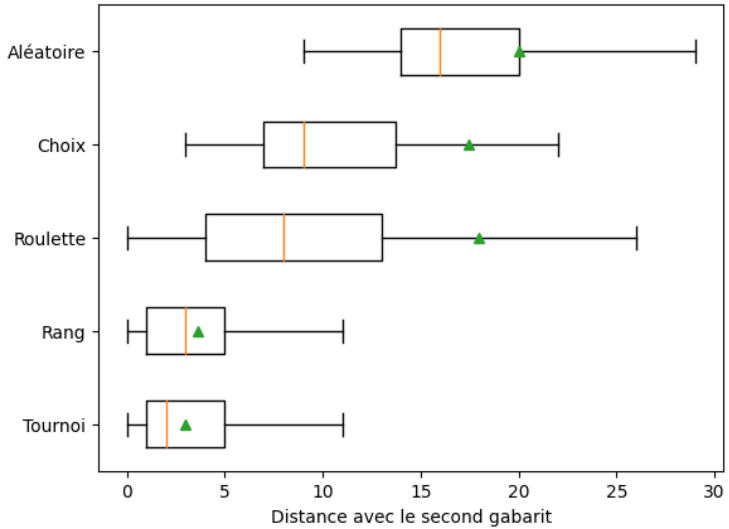
Exemple de sélection d'un paramètre

Mode de sélection : comment sélectionne-t-on les futurs parents reproducteurs ?

- Roulette : la probabilité d'être sélectionné est proportionnelle au score.
- Tournoi : nous piochons aléatoirement 2 vecteurs, et le meilleur est sélectionné.
- Rang : nous trions les vecteurs en fonction de leur score, et nous les sélectionnons dans l'ordre.

Construction optimisée de préimages par algorithme génétique

Génétique - Mode de sélection



Paramètres conservés (non ordonnés, en temps similaire)

- Sélection par rang.
- #population = 200 ; #itérations = 500.
- $P_{mutation} = 0.2$.
- Croisement double.

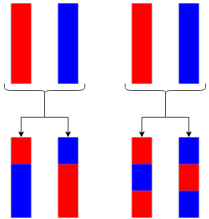
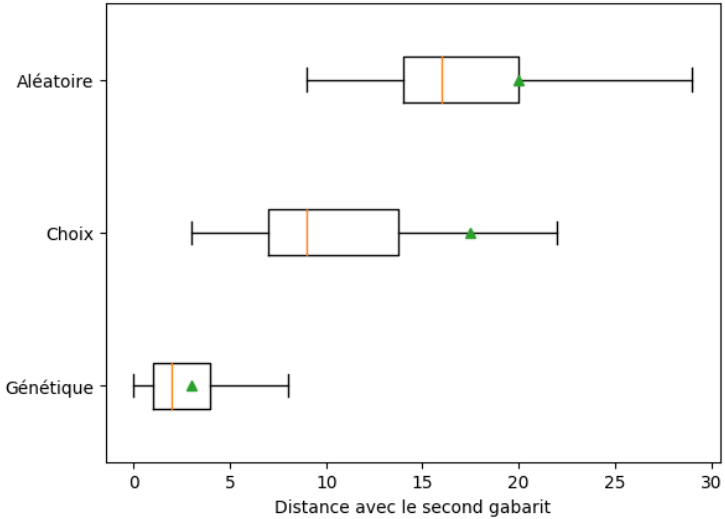


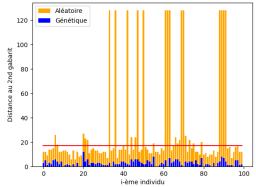
Figure: Croisement simple et double

Construction optimisée de préimages par algorithme génétique

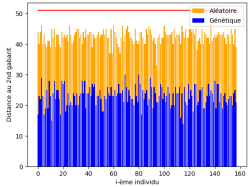
Comparaison des performances



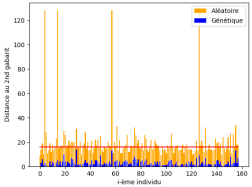
Base	Taux	Distance des PPR
FVC	100%	3
LFW	100%	23
PTB	100%	3



(a) FVC



(b) LFW



(c) PTB

Figure: PPR génétiques

Conclusion

- Nous obtenons des PPR dans 100% des cas pour les 3 bases.
- La distance au second gabarit est bien inférieure à la méthode aléatoire, en temps similaire.

Limites

- PPR : nécessite de connaître le score (distance de la comparaison).

Pouvons-nous construire de manière similaire des vecteurs proches de plus de 2 couples ?

Plan

Vecteur proche de tous les couples (gabarit, graine) : un passe-partout ?

Passe-partout depuis une base

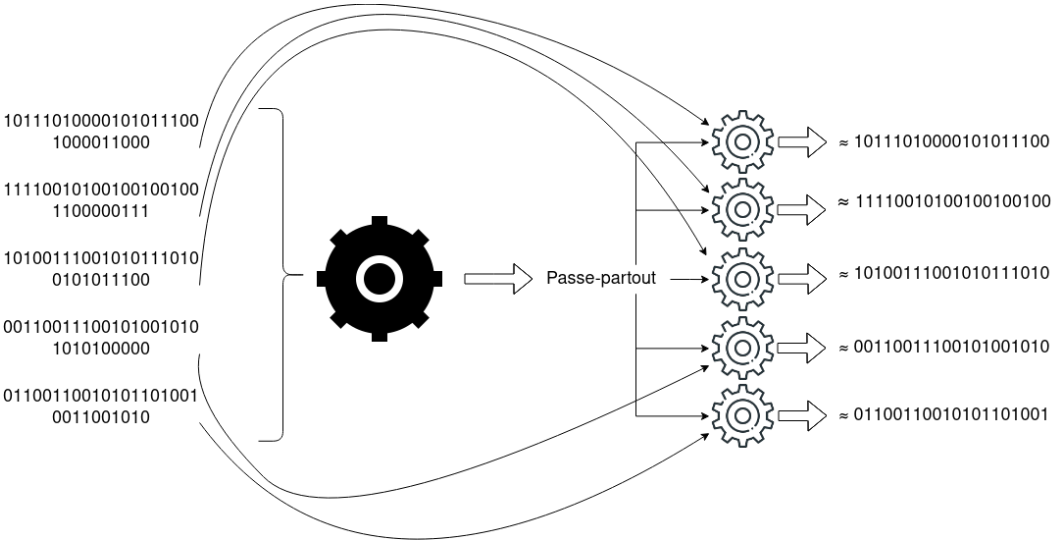
1. Soit $B = \{(u_i, s_i)\}_{i=1,\dots,n}$ une base de données biométriques révocables de n individus, avec u_i les gabarits et s_i les graines.
2. Construire un passe-partout x^* tel que $\forall i = 1, \dots, n$:
   - 2.1 u_i^* est le gabarit issu de la transformation de x^* avec la graine s_i.
   - 2.2 u_i^* est proche de u_i.

Motivation

Contourner un système de contrôle d'accès.

Du concept de préimage au concept de passe-partout

Description



Paramétré comme pour les PPR.

Fonction d'évaluation utilisée

Somme des distances de Hamming avec les vecteurs non usurpés.

Un vecteur passe-partout

Un vecteur x est un passe-partout pour B si pour chaque couple (gabarit, graine), le passe-partout est transformé avec la graine en un gabarit proche.

Taux de couverture optimale (TCO)

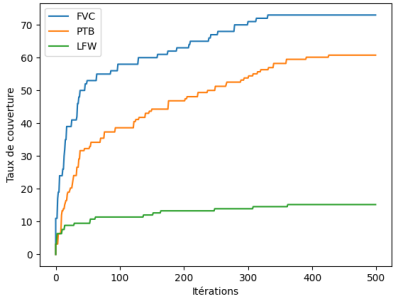
TCO ϵ : B est dite ϵ-couverte par le passe-partout x s'il existe ϵn couples (gabarit, graine) pour lesquels la transformation de x est proche.

Taille optimale de dictionnaire (TOD)

TOD r : B est dite partitionnée par un ensemble de r passe-partout $\{x^1, \ldots x^r\}$ si pour chaque couple (gabarit, graine), un des r passe-partout au moins est transformé en un gabarit proche.

Résultats

Base	TCO (%)	TOD	#pers
FVC	73	5	100
LFW	15.2	18	158
PTB	61	12	158



Description

Quid de la couverture sur des données biométriques non utilisées pour la construction du passe-partout ?
2 ensembles : un utilisé à la construction, l'autre non.

Motivation

Il est peu probable d'avoir toutes les données d'une base pour attaquer cette même base.

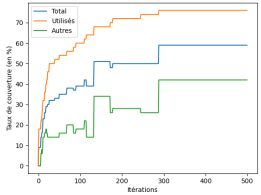
Résultat

Les passe-partout conservent environ la moitié de leur couverture.

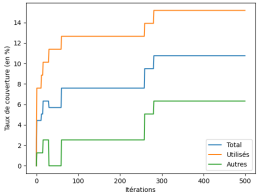
Base	TCO (%)	Report TCO (%)
FVC	42	73
LFW	6.3	15.2
PTB	44.3	61

Évolution de la couverture

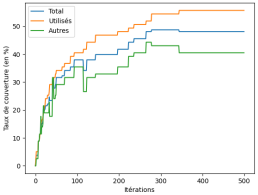
L'objectif de l'algorithme génétique est de maximiser la couverture de l'ensemble utilisé (courbe orange).
Les courbes bleue et verte sont informatives (et donc non croissantes).



(a) FVC



(b) LFW



(c) PTB

À quoi ça sert ?

1. Construire un passe-partout depuis un premier lot de données biométriques transformées.
2. Obtenir un taux de couverture important sur un autre lot de données transformées.
3. Pouvoir tromper un contrôle d'accès.

Conclusion

- Jusqu'à 73% de couverture.
- Réutilisabilité : conserve environ la moitié de la couverture.

Limites

Nécessite de connaître la base avec les couples (gabarit, graine) en une fois.
Algorithme hors ligne.

Pouvons-nous changer de prérequis pour augmenter le taux de couverture ?

Plan

Le vecteur passe-partout est fixé (réel ou synthétique).

Stratégie

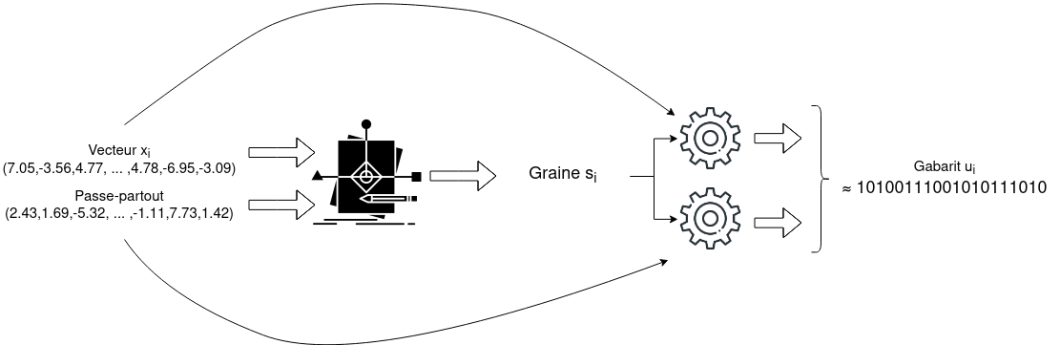
Nous partons de la base de données biométriques et nous allons **choisir** les graines (force brute) pour construire la base de données biométriques révocables.

Conséquences

- Être actif à la phase d'enrôlement.
- Cas d'usage éthique.

$\Rightarrow$ Objectif : augmenter le taux de couverture !

Pour chaque vecteur x_i, indépendamment des autres, nous choisissons une graine.
Algorithme en ligne / incrémental.



Transformation utilisée : Achlioptas (nous avons le choix).

Indicateurs

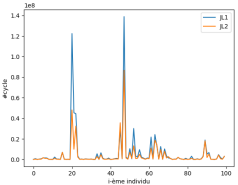
- TCO = 100%.
- TOD = 1.
- EER similaire graines choisies - graines aléatoires.

Analyse

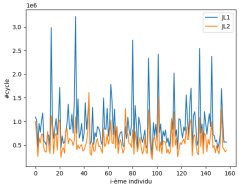
Nous avons largement augmenté le taux de couverture en un temps moindre.

Délégation de droits et passe-partout

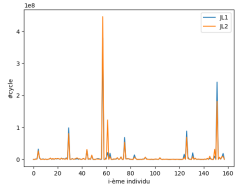
Résultats



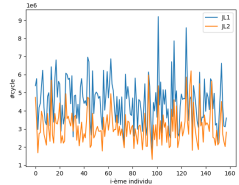
(d) FVC bin



(e) LFW bin



(f) PTB bin



(g) LFW euc

À quoi ça sert ?

- Éthique : possibilité de déléguer des droits sans modifier le circuit d'accès.
  $\rightarrow$ Au lieu d'ajouter une hiérarchie de droits sur les données biométriques, elles sont transformées en intégrant ces droits.
- Attaque : il faut être actif à la phase d'enrôlement et ajouter une porte dérobée sans code suspicieux et sans rendre la base suspecte (clé physique).
  $\rightarrow$ Elle est intégrée dans les données transformées.

L'individu dont est issu le vecteur passe-partout
souhaite couvrir la base plus tard !
$\rightarrow$ Étudier la couverture de ses futures captures (et les améliorer).

Individu passe-partout

Quid des autres captures de la personne dont est issu le vecteur passe-partout ?

Ensemble de recherche

Le premier sous-ensemble de vecteurs est utilisé pour la recherche de graines.

Ensemble de test

Le second sous-ensemble de vecteurs, non utilisé pour la recherche de graines, permet de vérifier la performance de ”l’individu passe-partout”.

Limite de 5 minutes par graine.

Présentation des résultats

Courbes cumulées décroissantes.
Objectif → Faire glisser la courbe rouge à droite de l'orange.

50% des vecteurs de test couvrent au moins 50% de la base ($T = 1$).
→ 50% des vecteurs de test couvrent au moins **80%** de la base ($T = 4$).

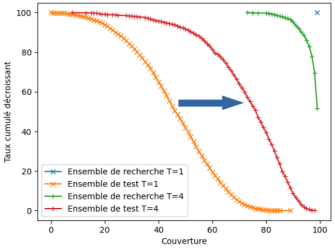
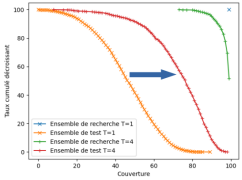


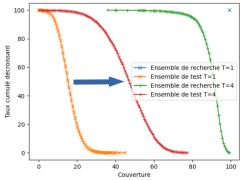
Figure: FVC bin

Délégation de droits et passe-partout

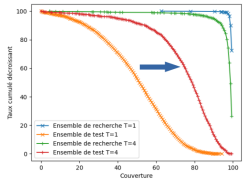
Extension - Résultats



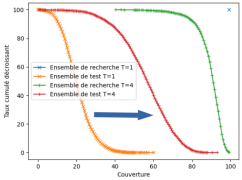
(a) FVC bin



(b) LFW bin



(c) PTB bin



(d) LFW euc

Où en est-on ?

- Nous souhaitons que de futures captures persistent à couvrir la base.
- Utiliser plusieurs vecteurs pour choisir la graine permet d'améliorer cette future couverture.

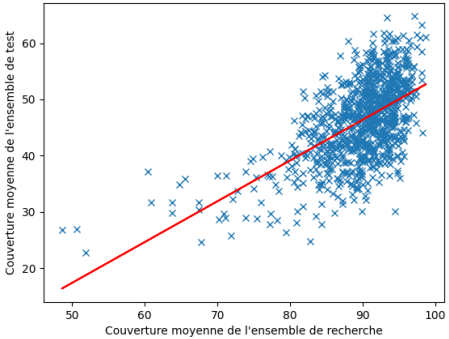
Pouvons-nous mieux choisir les vecteurs utilisés pour la recherche de graine et encore améliorer cette couverture ?

Corrélation

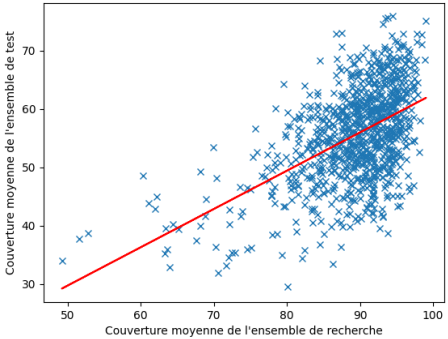
Corrélation bonne couverture de l'ensemble de recherche - bonne couverture de l'ensemble de test ?

Délégation de droits et passe-partout

Extension - Corrélation des performances



(a) bin



(b) euc

Figure: LFW

Conséquences

1. Utiliser plusieurs vecteurs pour la recherche de graines améliore la couverture d'autres vecteurs.
2. Il faut sélectionner les vecteurs utilisés pour la recherche de graines selon leur couverture pour maximiser la couverture d'autres captures.

Limite et perspective

- Performances moindres avec une base issue d'oreilles, et meilleures performances avec une base issue d'une fusion empreintes/visages.

Plan

Biométrie

- Reconnaître.
- Captures de modalités biométriques.
- Extraction de caractéristiques.
- Comparaisons.
- Données personnelles $\Rightarrow$ Protection, utilisée dans cette thèse !

Préimage ...

- ... proche (s'authentifiant avec un gabarit) ...
- ... et réutilisable (s'authentifiant pour un autre gabarit).

Outil

- Algorithme génétique.
- Minimiser la somme des distances.
- Optimisation du paramétrage.
- 100% de PPR.

Passe-partout : scénario 1

- Construction d'un passe-partout ...
- ... depuis une base de données biométriques révocables ...
- ... usurpant un maximum de gabarits !
- Partitionnement avec plusieurs passe-partout.

Résultats

- Jusqu'à 73% de couverture.
- Réutilisabilité : conserve environ la moitié de la couverture.

Passe-partout : scénario 2

- Construction d'une base de données biométriques révocables ...
- ... en choisissant les graines pour un passe-partout.
- Choix d'une transformation rapide.

100% de couverture (temps variables).

Individu passe-partout

- Utilisation de plusieurs vecteurs.
- Amélioration de la couverture de futurs vecteurs.
- Corrélation de couvertures : choisir de bons vecteurs.

Finalement

- Inverser 2 gabarits en une seule préimage.
- Contourner un système de contrôle d'accès en construisant un passe-partout à partir de données similaires.
- Déléguer des droits ou introduire une porte dérobée au sein des données transformées.

2 publications dans SECRYPT2022 suite aux travaux de cette thèse :

Durbet et al., 2022a

Authentication attacks on projection-based cancelable biometric schemes.
Modifier au minimum une capture d'empreinte digitale par rapport à un couple (gabarit, graine) pour qu'elle devienne un préimage proche. Ensuite, le faire pour de nombreux couples.

Durbet et al., 2022b

Near-collisions and their impact on biometric security.
Recommandation $\tau \leq \frac{M}{10}$.

- **Étudier l'impact de la taille de la base ou des données sur le TCO et le TOD (oreilles / fusion).**
- **Construction de bases de données biométriques synthétiques (expliquer l'impact).**
- Attaque par présentation : construction d'une capture de modalité produisant un vecteur.
- Choix de graine pour limiter les attaques (PPR et passe-partout scénario 1).
- Réutilisabilité d'une PPR sur de futures captures.
- Scénario 1 : construction d'un passe-partout sans données biométriques, à partir du type de modalités et de l'algorithme d'extraction.
- Scénario 1 : déterminer une fonction d'évaluation ne nécessitant pas le score.
- Scénario 2 : choix de graines pour plusieurs individus passe-partout.

Journal international

Biometric masterkeys, Computers & Security, Volume 116, 2022, 102642

École d'été internationale

Long-lived nearby-template preimages on biometric transformation with genetic algorithm, Biometrics, Forensics and Identity science for human-centered applications, Alghero Italy, 2020

[Merci]

Questions ?

tanguy.gernot@unicaen.fr
https://gernot.fr